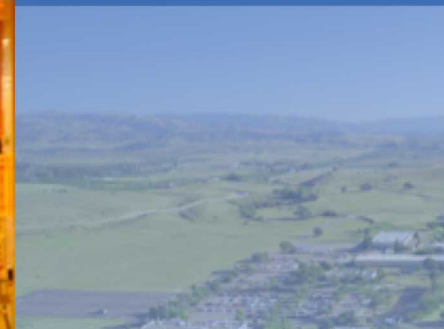
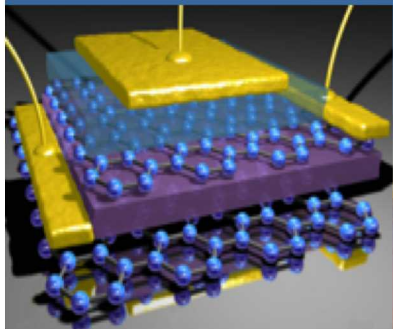




Energy-Efficient Edge Computing: Challenges and Opportunities Beyond Moore's Law

October 8, 2018



PRESENTED BY

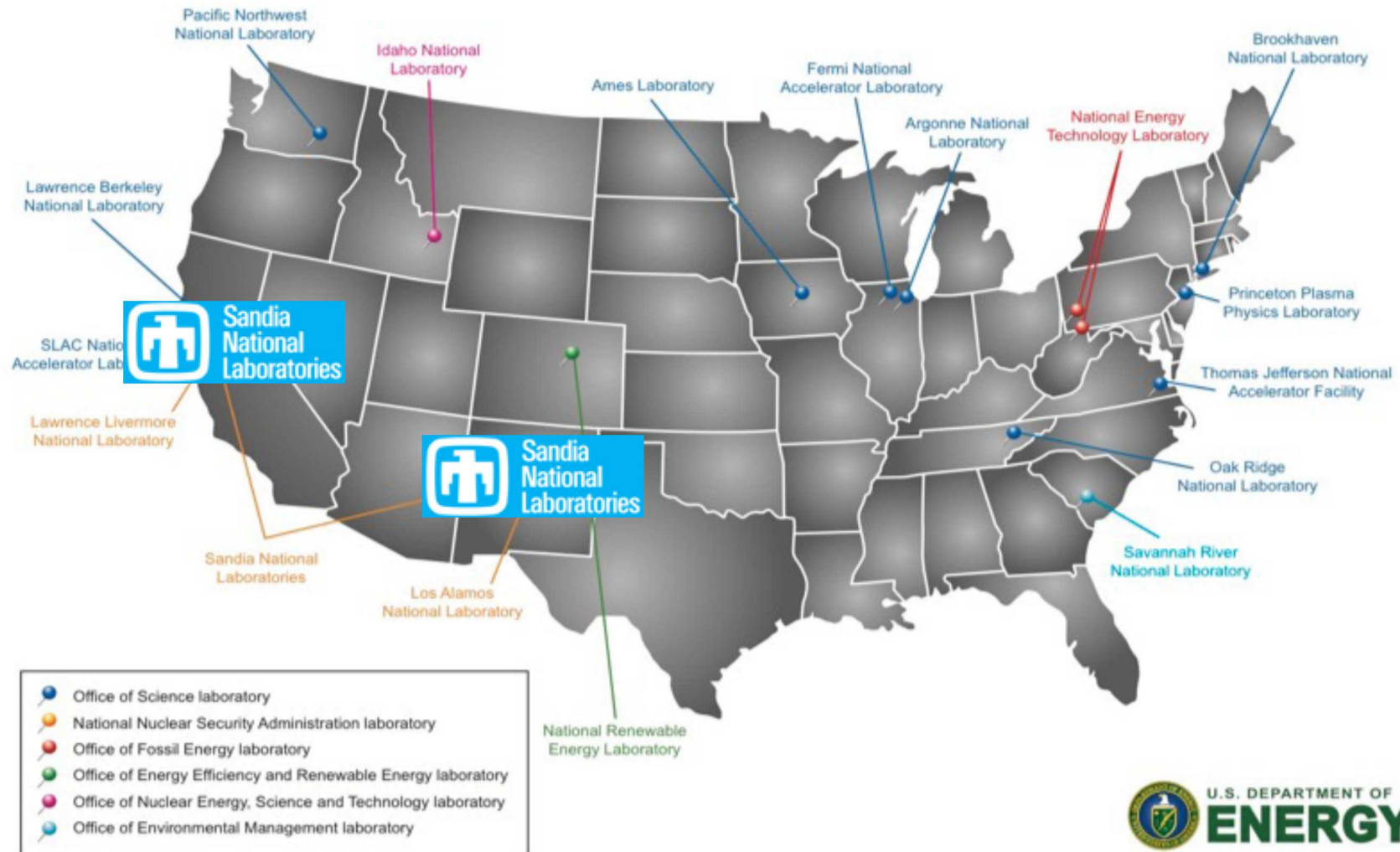
Christian Mailhot – Senior Manager, Materials Research
Sandia National Laboratories, Livermore, CA 94551

Outline

- Overview of US Department of Energy (DOE) national laboratories complex
- Introduction to Sandia National Laboratories
- Information Technology challenges beyond Moore's Law:
 - Core computing: High-performance computing
 - Edge computing: Connected autonomous vehicles (CAVs)
- Technical approach: Multiscale Codesign Framework (MCF)
 - Algorithms and Software Environments
 - Hardware and Circuit Architectures
 - Devices and Communications
 - Materials
 - Advanced Manufacturing
- Partnerships:
 - DOE national laboratories
 - Industry
 - Academia

The DOE national laboratories support the agency's broad missions in national security, energy and environment, and discovery science

Department of Energy National Laboratories

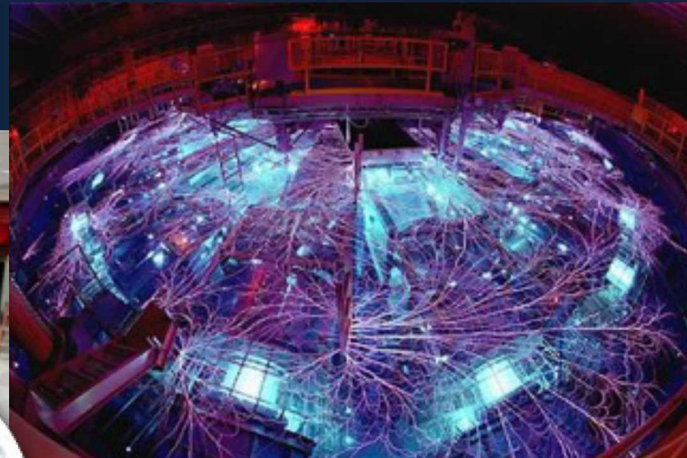


Sandia National Laboratories core scientific capabilities

4

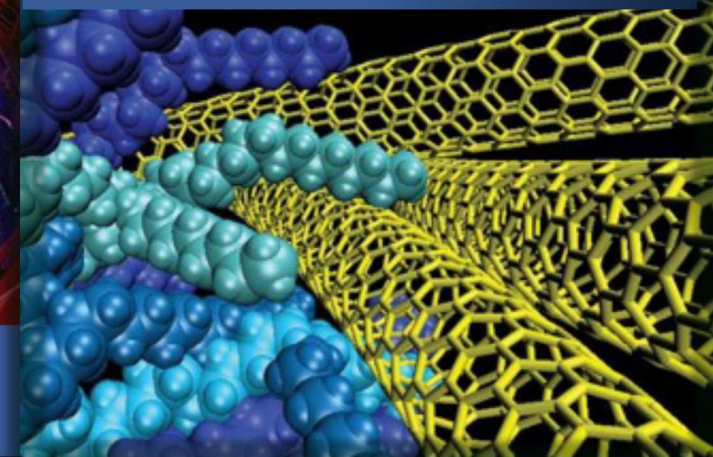


Computing & Information Sciences

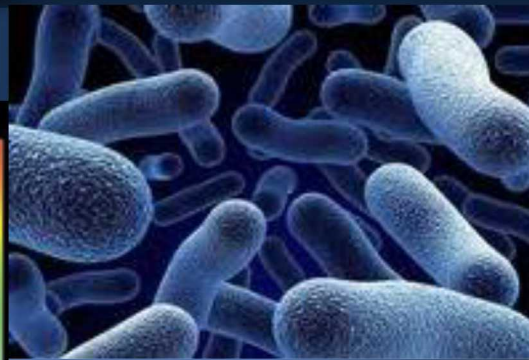
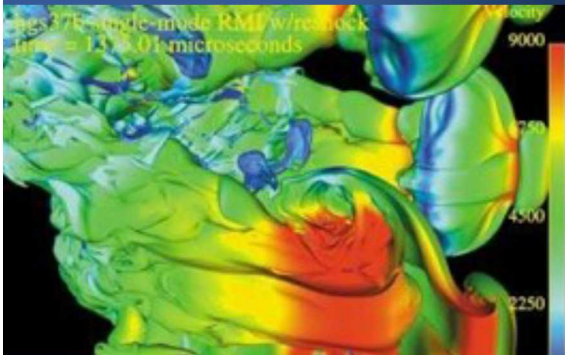


Radiation Effects & High Energy Density Science

Materials Science and Engineering

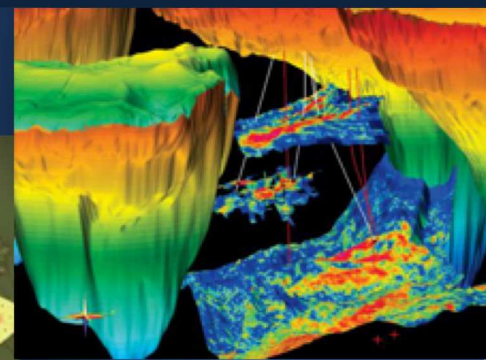


Engineering Science



Bioscience

Nanodevices & Microsystems



Geoscience

Materials science capabilities and user facilities at Sandia



Microsystems and Engineering Science Applications (MESA)



- 60+ years as DOE/NNSA mission lead in electronics
- Silicon and III-V Materials

Advanced Materials Laboratory (AML)



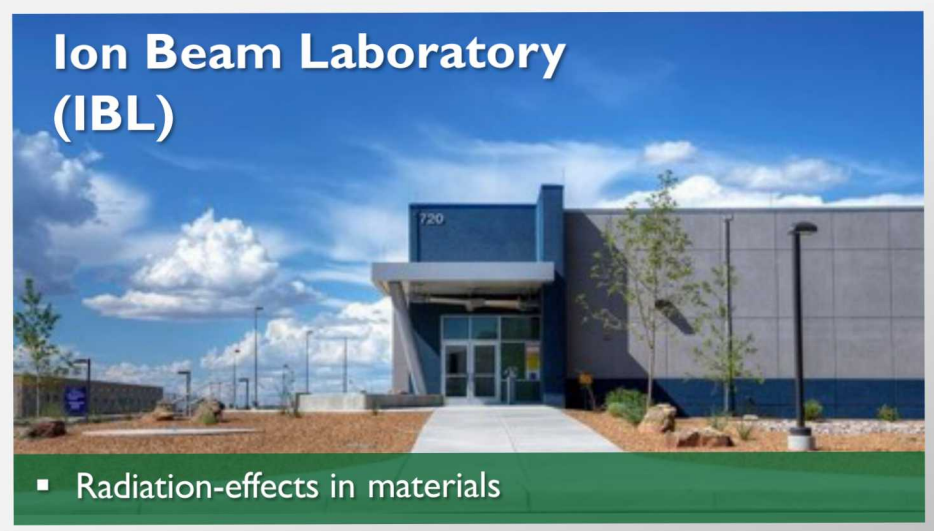
- Unique materials synthesis efforts

Center for Integrated Nanotechnologies (CINT)



- DOE BES-SUF NSRC
- Celebrating 10th Anniversary
- Focus on Integration of Nanoscience and Technology

Ion Beam Laboratory (IBL)



- Radiation-effects in materials

Moore's Law (1965): The observation that the number of transistors in a dense integrated circuit doubles every 2-3 years.



Emerging challenges:

1. predicted **tapering of scaling technologies**
2. **leveling of performance-per-watt** in the coming decade will result in greater energy needs to sustain an equivalent increase in information processing performance

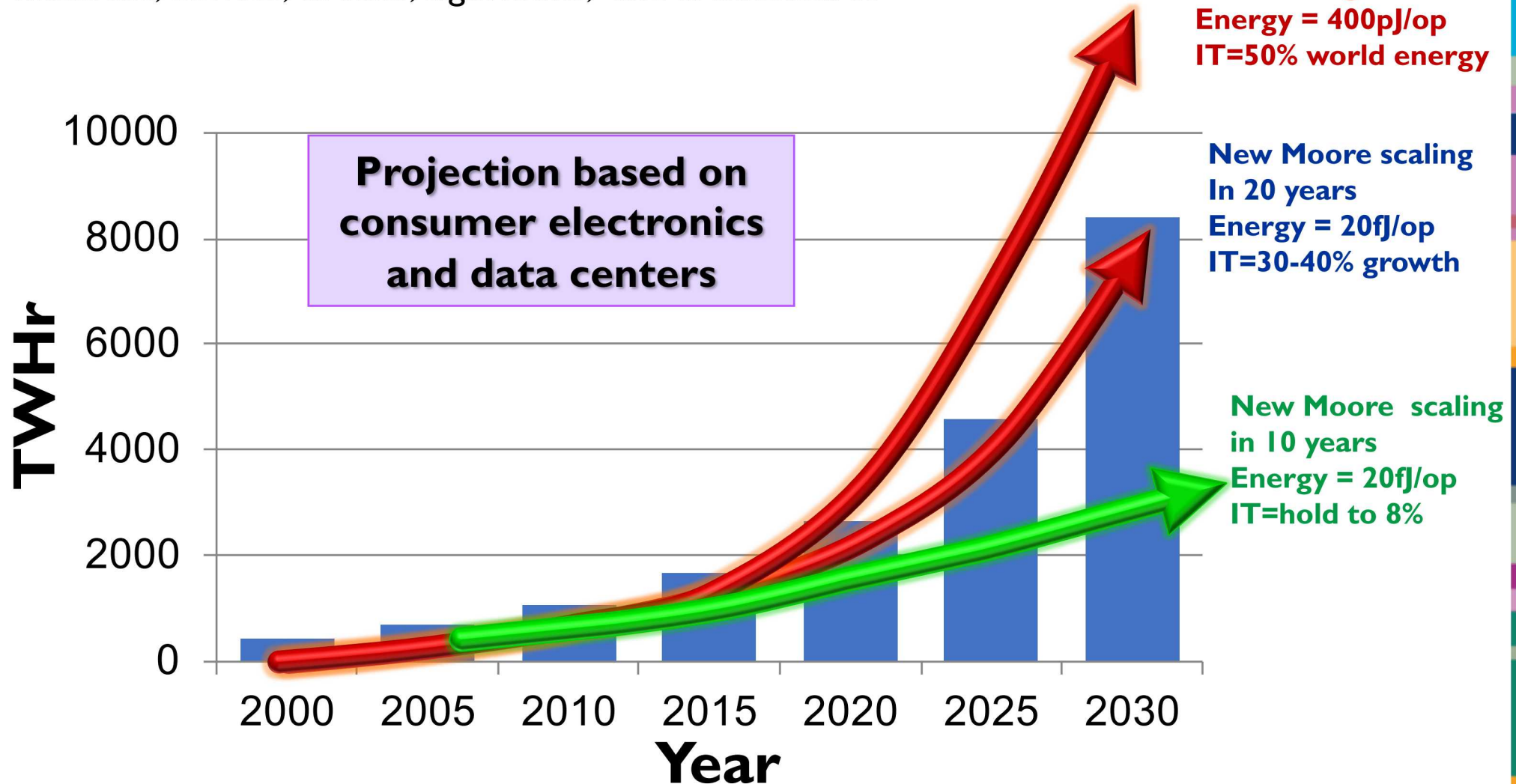
Beyond Moore's Law Information Processing: Attacking the IT energy consumption challenge.

7

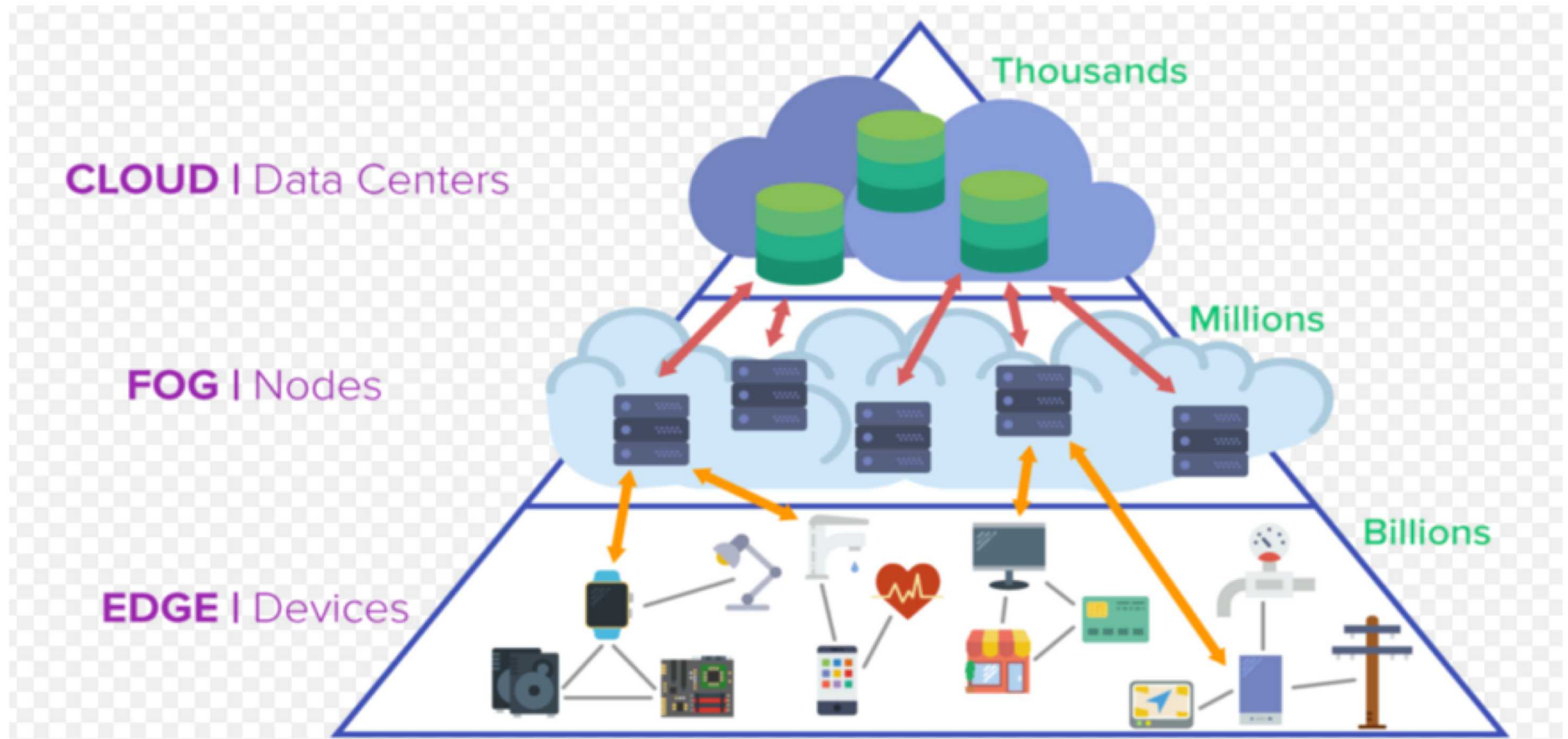


The challenge: IT will challenge future electricity supply

The opportunity: Advances at virtually every level of the technology hierarchy offer promise: materials, devices, circuits, algorithms, and architectures



Information processing at the “edge” of a network



- Edge computing **pushes applications, data and computing power** away from centralized points (the “core” or “cloud”) **to the extremes of a network** (the “edge”) which makes contact with the physical world or end users.
- Edge computing **de-emphasizes the core computing environment**, limiting or removing a major bottleneck and a potential single point of failure.
- By performing analytics and knowledge generation at the edge, **communications bandwidth between systems under control and the central data center is reduced.**

Edge processing requirements

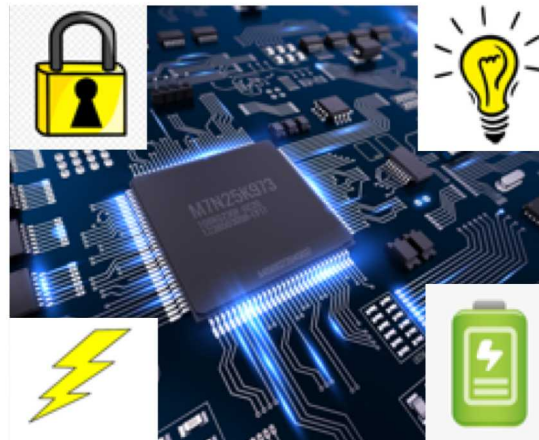


Mission driver: several Sandia missions require information processing at the point-of-sensing for real-time analysis without data transmission to centralized computer – **information processing at the “edge”**



Requirements: moving processing power to the edge

1. **Low SWaP (size, weight, and power)**
2. Radiation environments
3. Trusted and secure
4. Re-configurable circuits for evolving threats and challenges

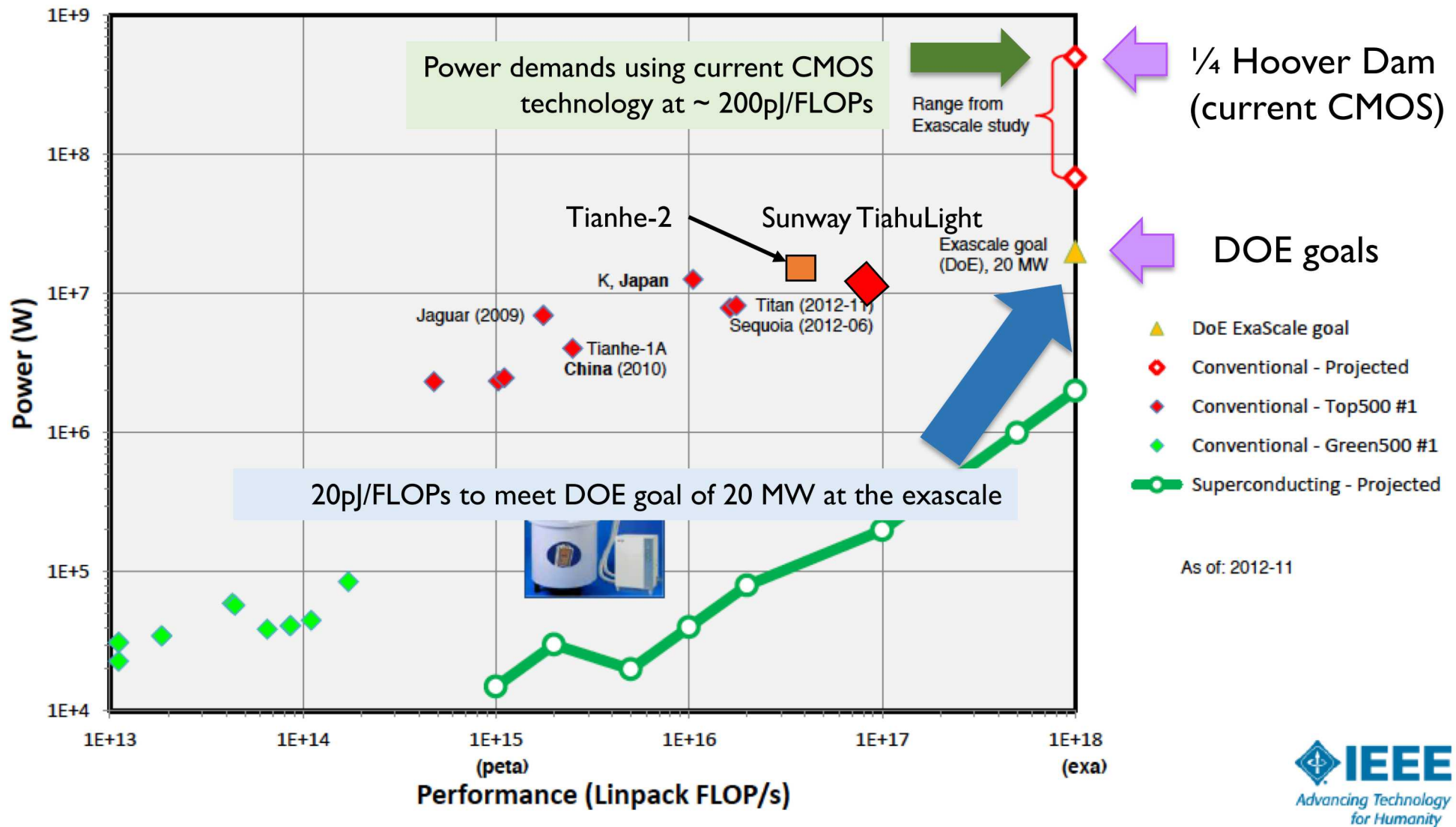


Energy challenge:

1. The performance-to-power ratio is leveling
2. Too much power is required to process too much data

Tenfold energy efficiency gain is needed to meet DOE goal of HPC at the Exascale

10



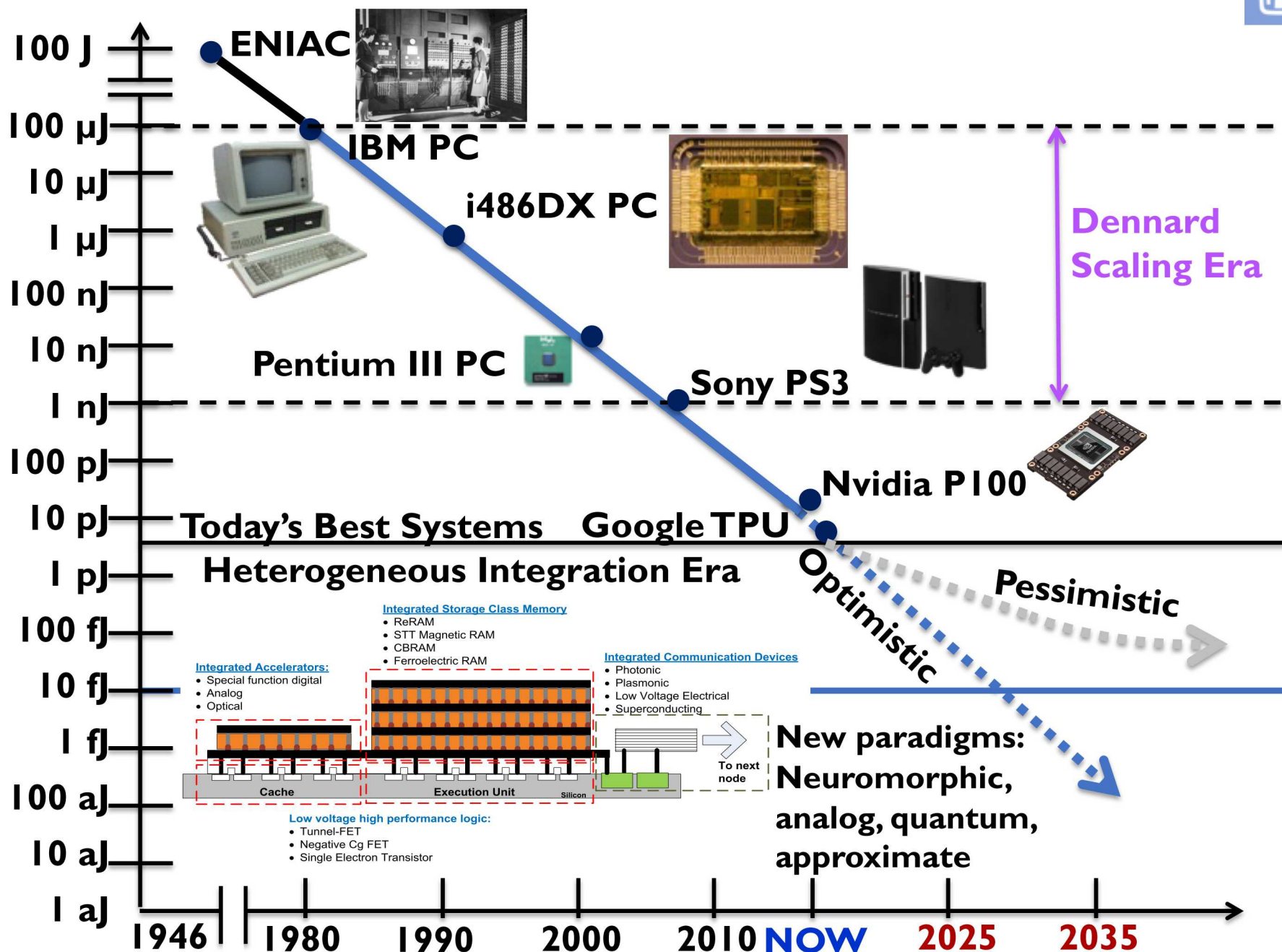
Scott Holmes (2013), data added from <http://top500.org>

Evolution of Computing Machinery



11

Energy Per Mathematical Computation



Dennard (MOSFET) scaling: As transistors get smaller their [power density](#) stays constant, so that the power use stays in proportion with area: Both [voltage](#) and [current](#) scale (downward) with length

Sensing and computing for autonomous vehicles



Autonomous vehicle computing has to:

1. process large amounts of information simultaneously
2. extract essential features from multiple sensor data stream
3. make decisions (brake, turn, accelerate, etc.)

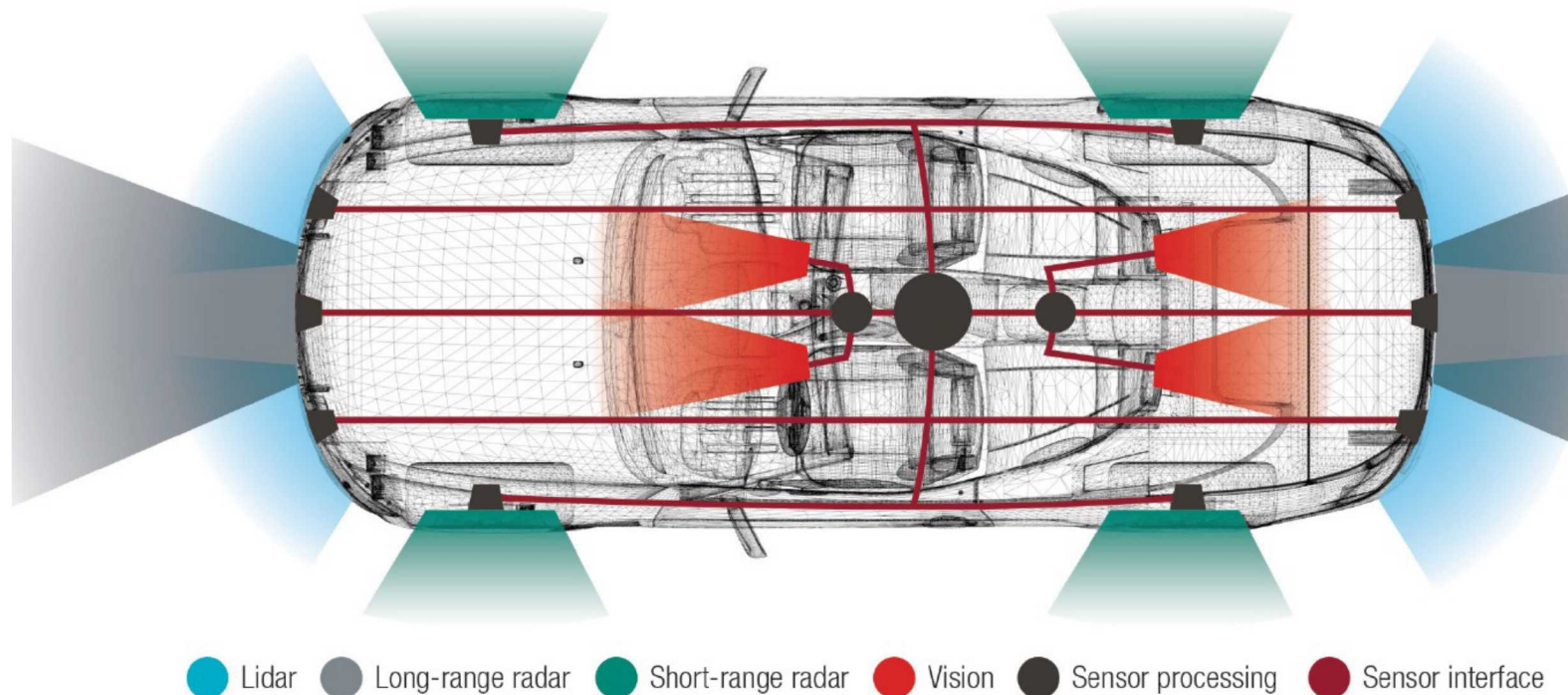


Figure 2. In the future autonomous car, sensor processing will occur throughout the car's periphery to send signals and commands to the car's mechanical parts.

F. Mujica, "Scalable electronics driving autonomous vehicle technologies"

R. Sagar, "Making cars safer through technology innovation" www.ti.com/corp-ino-auto-wp-lp, 2017

2–4 kW of electrical power is consumed by on-board computers and sensors

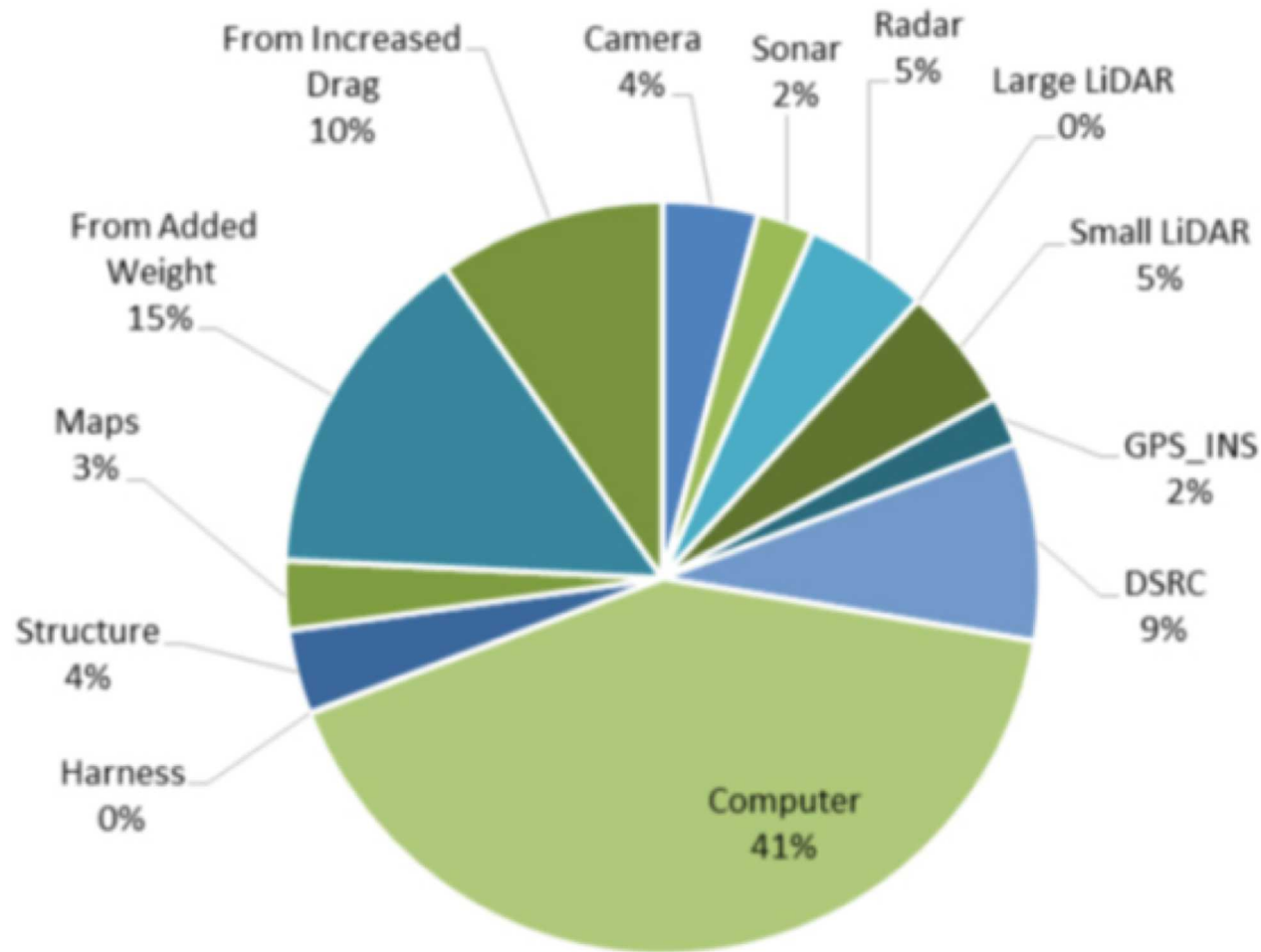


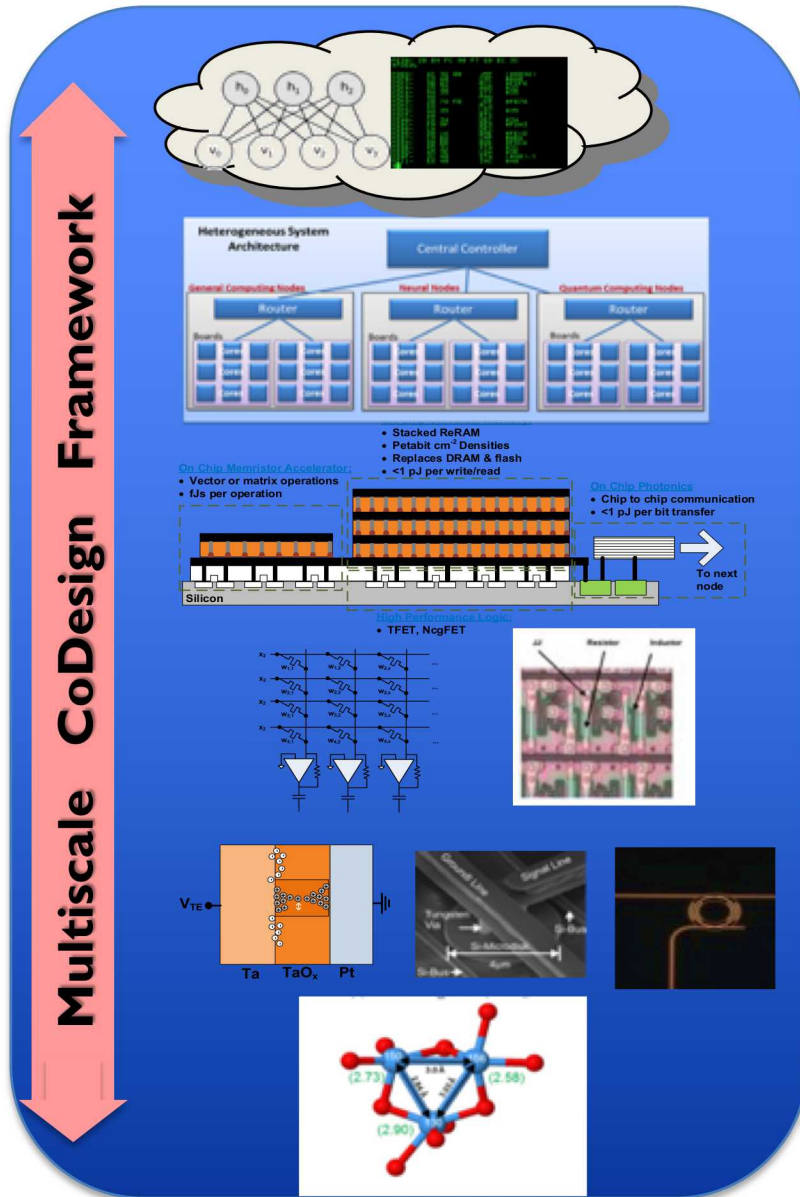
Image: University of Michigan

Life Cycle Assessment of Connected and Automated Vehicles: Sensing and Computing Subsystem and Vehicle Level Effects

James H. Gawron, Gregory A. Keoleian, Robert D. De Kleine, Timothy J. Wallington, and Hyung Chul Kim

Environmental Science & Technology **2018** 52 (5), 3249-3256

Technical strategy: Multiscale Codesign Framework: Bridging length scales for energy-efficient information processing



Programming paradigms

System architecture modeling

Component hetero-integration

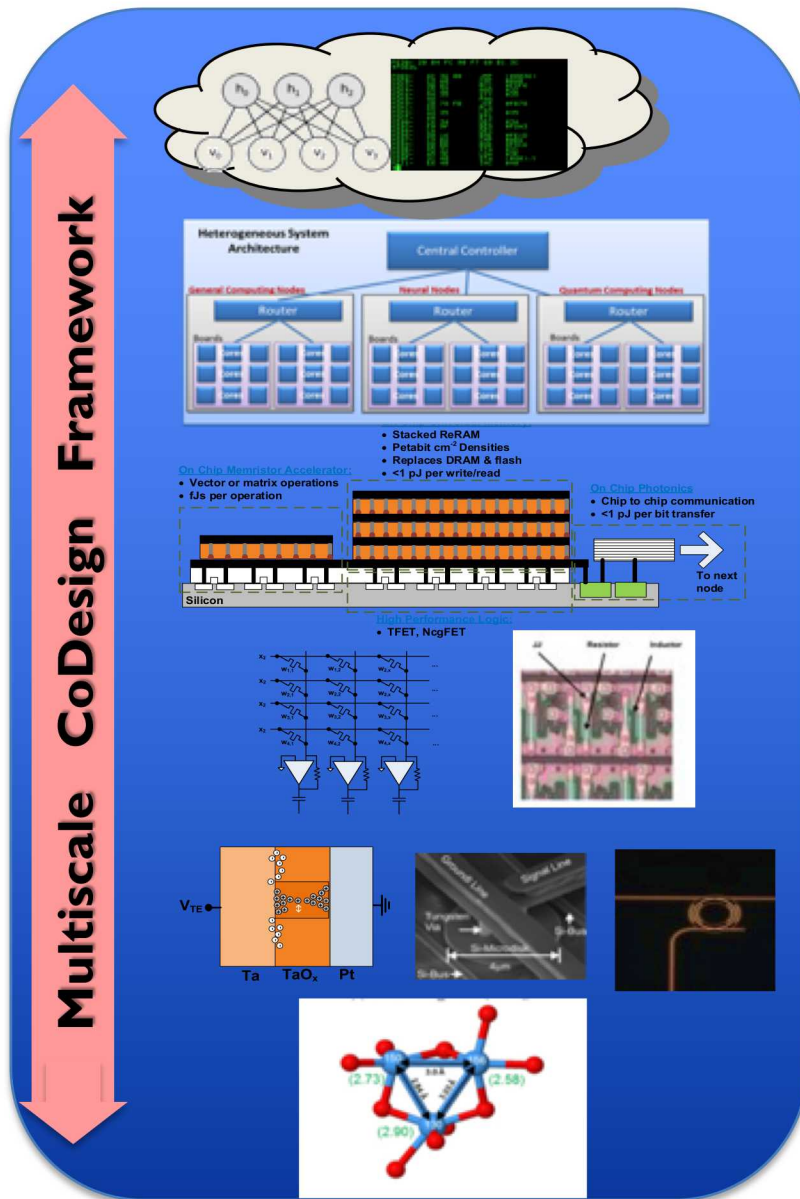
Device demonstration

Device physics

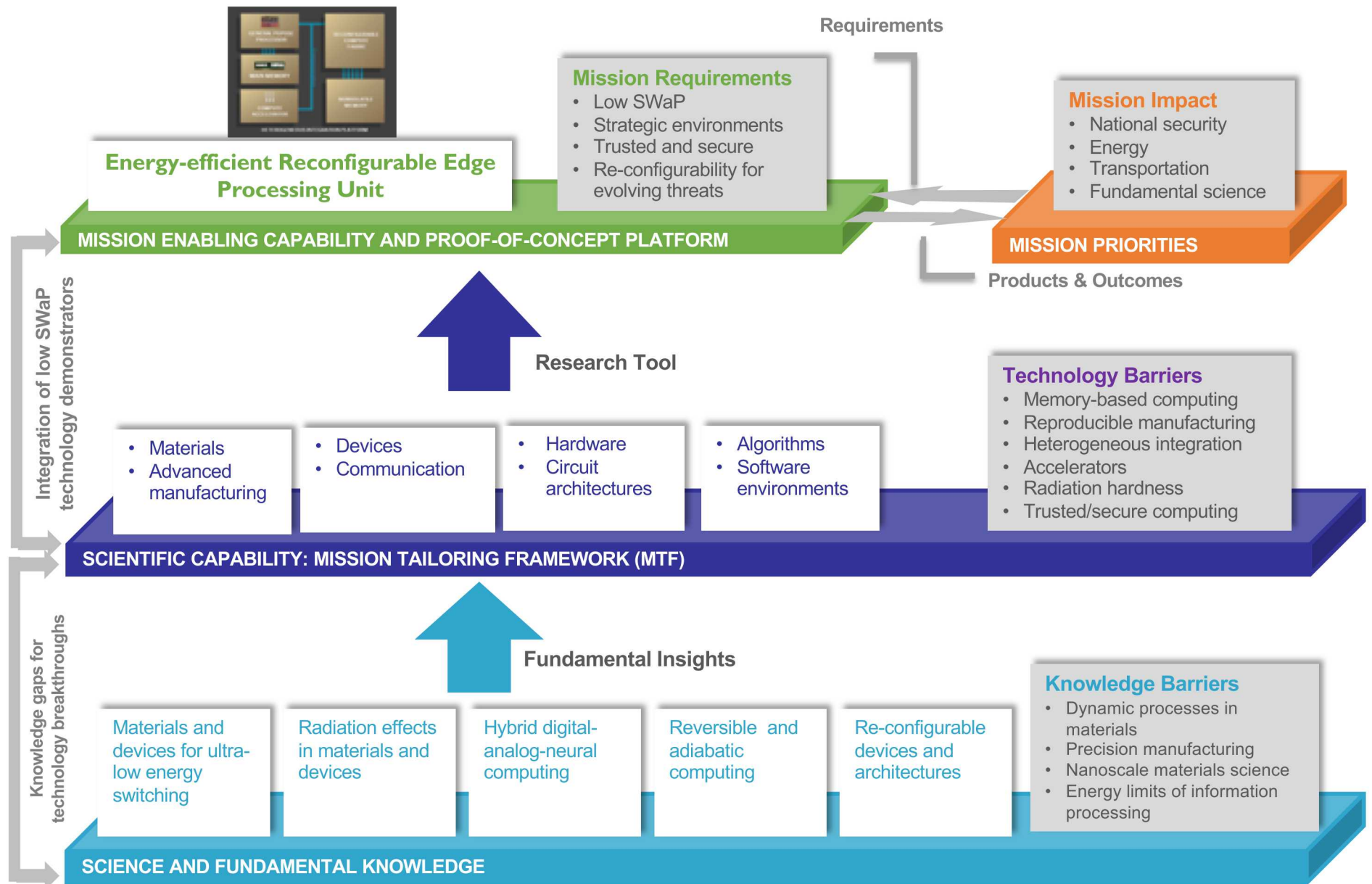
Fundamental materials science

Multiscale CoDesign Framework: Building partnerships across the national laboratories complex

15



16 **Vision:** Develop S&T capabilities for energy-efficient edge information processing and sensing over the horizon of present-day technologies to enable future missions



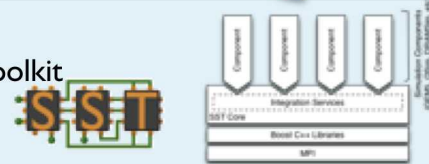
Algorithms and Software Environments

- Application Performance Modeling



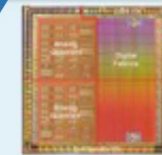
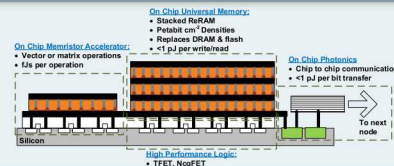
Computer System Architecture Modeling

- Next generation of Structural Simulation Toolkit
- Heterogeneous systems HPC models



Microarchitecture Models

- McPAT, CACTI, NVSIM, gem5

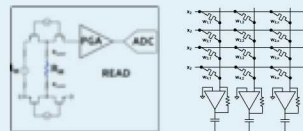


Component Fabrication

- Processors, ASICs
- Photonics
- Memory

Circuit/IP Block Design and Modeling

- SPICE/Xyce model

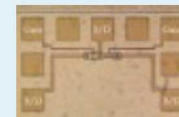
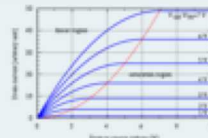


Test Circuit Fab and Measurement

- Subcircuit measurement

Compact Device Models

- Single device electrical models
- Variability and corner models

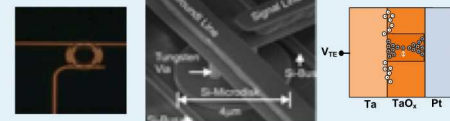


Device Measurements

- Single device electrical behavior
- Parametric variability

Device Physics Modeling

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

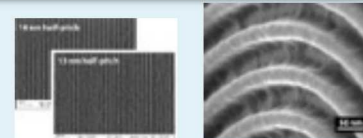
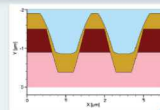


Device Structure Integration and Demonstration

- Novel device structure demonstration

Process Module Modeling

- Diffusion, etch, implant
- Simulation
- EUV and novel lithography models

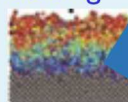


Process Module Demonstrations

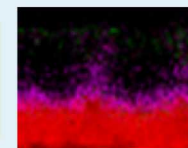
- EUV and novel lithography
- Diffusion, etch, implant simulation

Atomistic and Ab-Initio Modeling

- DFT – VASP, Socorro
- MD – LAMMPS



**10,000x improvement:
20 fJ per instruction**



Fundamental Materials Science

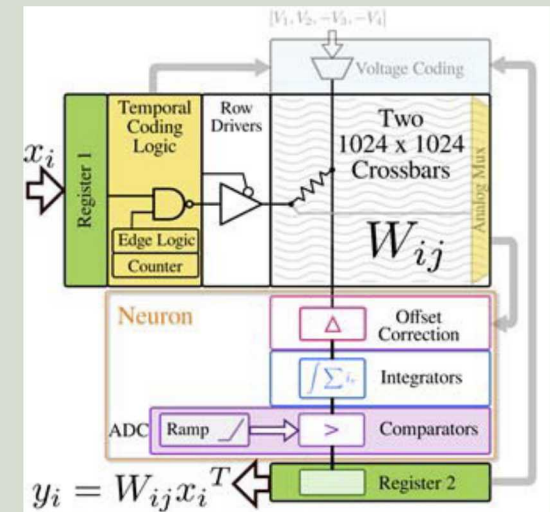
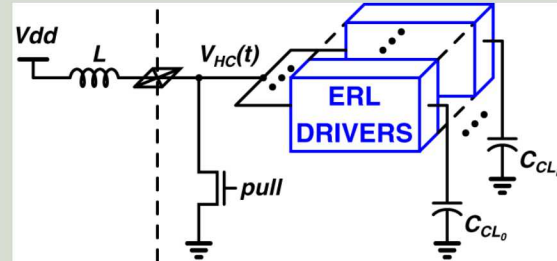
- Properties/Defects via Electron, Photon, & Scanning Probes
- Novel Materials Synthesis

Research topics for low-energy computing devices



Architectures

- Analog computing
- Neuromorphic accelerators
- Approximate or stochastic architectures



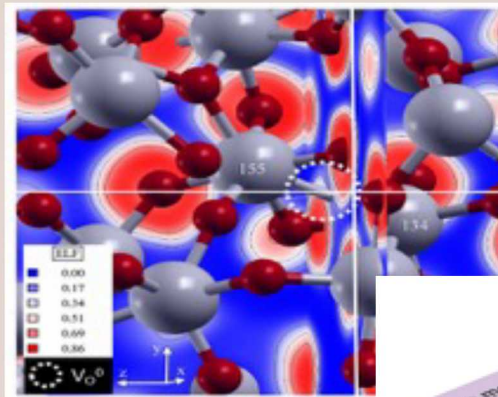
Materials and Devices

- Analog memories
- Novel materials and transistors
- Interconnect materials
- Diffusion barrier materials
- Thermal management

Patents pending:

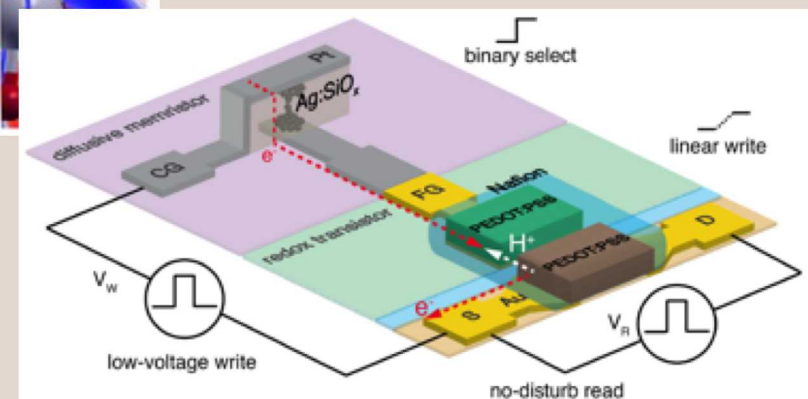
“Tunable Ionic electronic transistor”, Talin et al., filed 02/09/2017

“Ionic floating gate memory”, Fuller et al., filed 06/19/2018

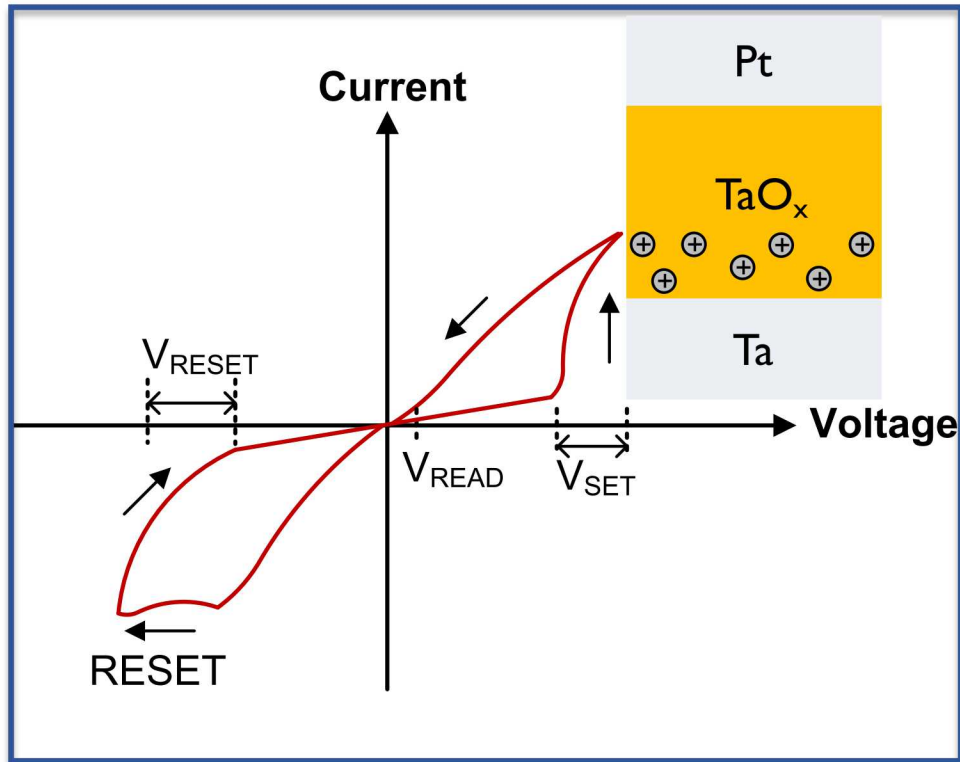


Nature Materials 16, 414 (2017)

Advanced Materials 29, 1604310 (2017)



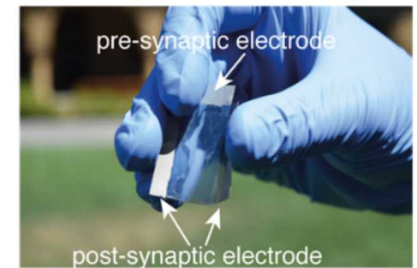
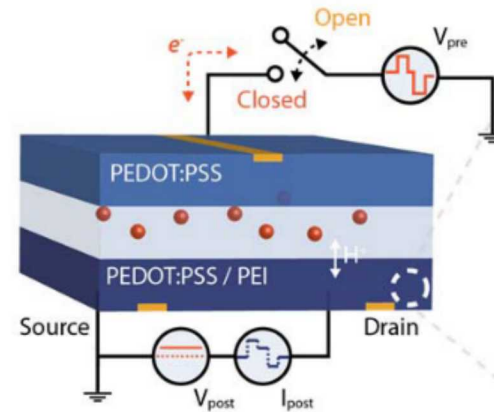
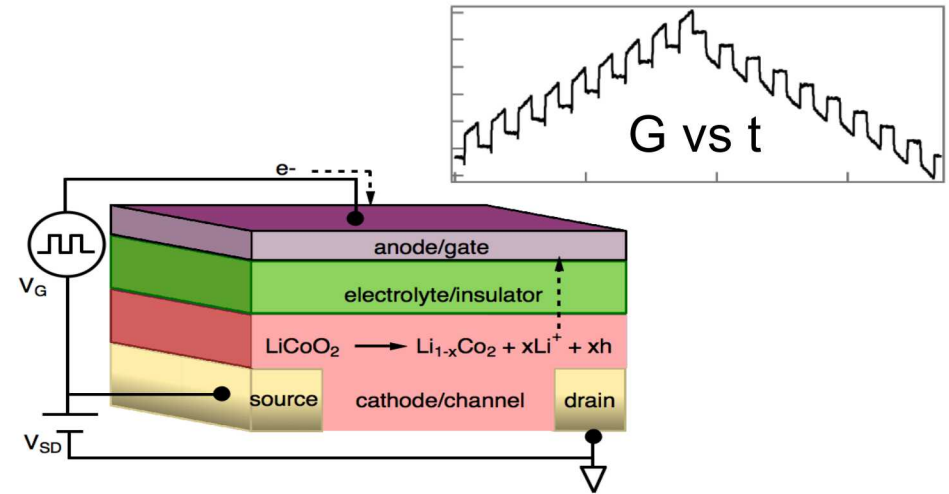
Designing, modeling, and fabricating devices with improved neural computing characteristics



Filament surface temperature (T_s):

$$T_s = T_{RT} + \sigma V^2 \frac{d_E}{2k_E d_o} \left[1 - \frac{k_E}{k_F} \frac{r_F^2}{4d_E d_o} \right]$$

Mickel et al., Adv Mater, 26, 4486, 2014
Landon et al., APL 2015, 107, 023108



Fuller et al., Adv Mater 2016, 10.1002/adma.201604310
van de Brugt et al., Nat Materials 2017, 10.1038/nmat4856

Partnership strategy: Leveraging community efforts and tailoring to Sandia's mission needs

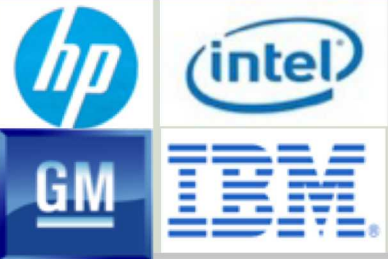


Academia



1. Engagement with Sandia's Academic Alliances partners
2. Establishment of regional alliances with Bay Area universities and national laboratories
3. Engagement with Semiconductor Research Corporation (SRC) to leverage academic investments in ultra-low energy electronics

Industry Leaders



1. Interactions with industry leaders
 - a. HP Enterprise: CAV edge computing infrastructure / NDA being put in place
 - b. Automotive OEMs: GM
2. On-going working relationships with several companies
3. Roadmap development with industry consortia
 - a. IEEE International Roadmap for Devices and Systems ([IRDS](#))

DOE National Laboratories



1. Build on multi-lab partnership established in 2016 (Sandia lead)
2. On-going interactions with DOE national laboratories

Federal sponsors



1. Engagement with EERE/Vehicle Technologies Office (VTO) and Advanced Manufacturing Office (AMO)
2. Department of Defense and other government agencies
3. Participation in DARPA Electronics Resurgence Initiative (ERI)
4. Anticipate a national "whole-of-government" initiative in establishing private-public partnership in next-generation semiconductor technologies

Implementing the Multiscale Codesign Framework (MCF) within a public-private partnership

Take the best ideas

Materials

- Low-D materials
- Multiferroics
- Non-linear optical

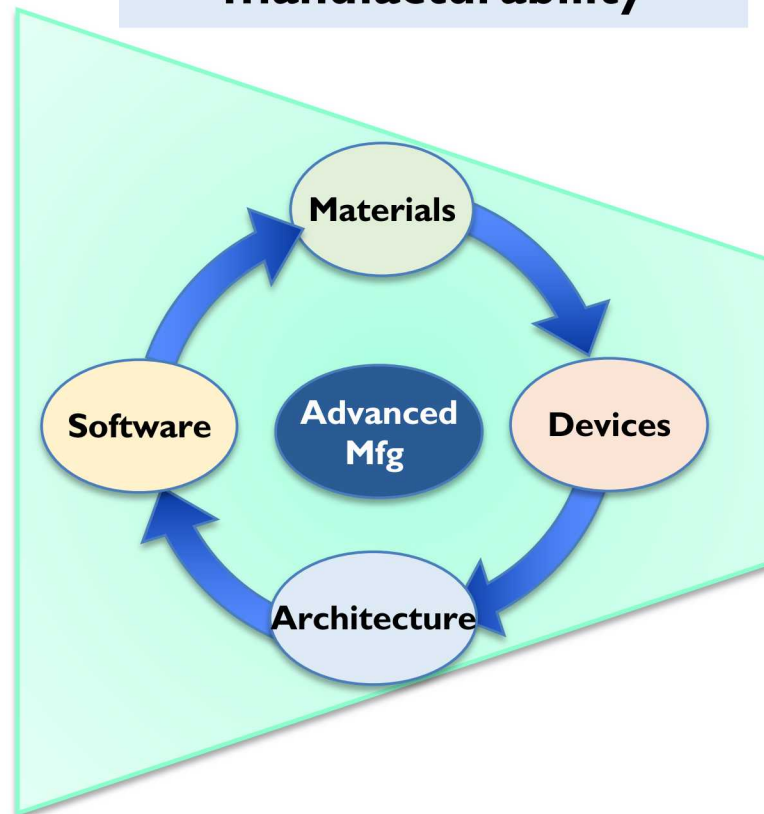
Devices

- TFETs
- Memristors
- Optical logic

Systems

- Embedded memory
- Optical interconnects
- Multi-level logic
- ...

Rapidly evaluate for performance, integration, and manufacturability



Develop and deploy

Chip Foundries

- Intel
- Global Foundries
- TSMC

Equipment Makers

- AMAT
- LAM
- Lithography

System

- IBM
- Dell

New startups

- Multibeam e-
- ALE

Community

DOE National Laboratories

Industry

OFFICIAL USE ONLY



Backup Slides

Opportunities/Partnerships



Government Agencies

NIST

DARPA



Academia

MIT

Berkeley
UNIVERSITY OF CALIFORNIA



Industry

intel

hp

SRC

IBM

AMD

Micron

National Laboratories



Sandia
National
Laboratories



Beyond Moore's Law Key Capabilities & Facilities:

Post-Exascale Scientific Computing

24



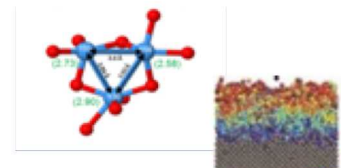
Target Application:
Physics Codes



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

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

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

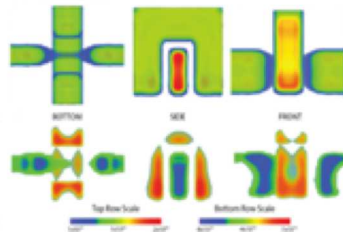


Ab Initio Models



Devices

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

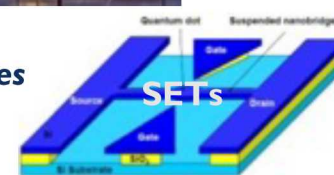


Universal Charge Device Simulator



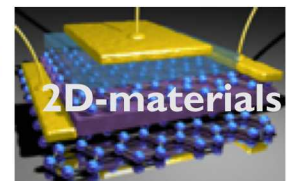
MESAFab (SNL)

CHISEL



Quantum dot

Suspended nanobridge



2D-materials



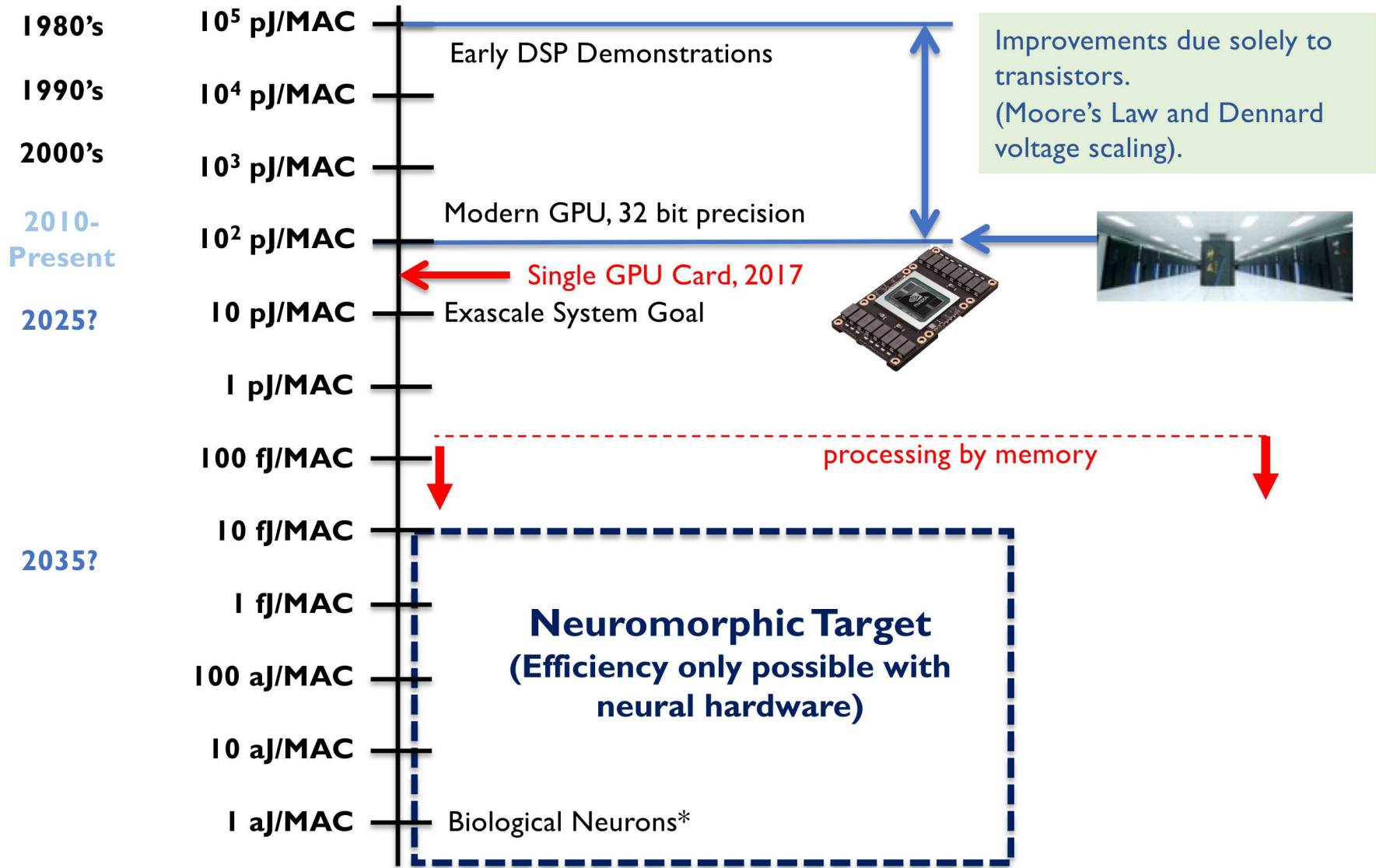
CNMS

MOLECULAR FOUNDRY



Setting energy efficiency performance goals

Energy efficiency of multiply and accumulate operations (MACs)



modified from Hasler et. al. *Frontiers in Neuroscience*, 2013