SAND2023-02968C

Materials Discovery for Energy-Efficient Computing

Sandia/UC Davis Spring 2023 Symposium

Ashfia Huq

(Christian Mailhiot, Jefferey Nelson and Many other co-workers)

Sandia National Laboratories

Tracking Number:
Type: FORMAL - Conference - Presentation
Title: Materials Discovery for Energy-Efficient Computing
SAND Number:
Contact: Ashfia Huq

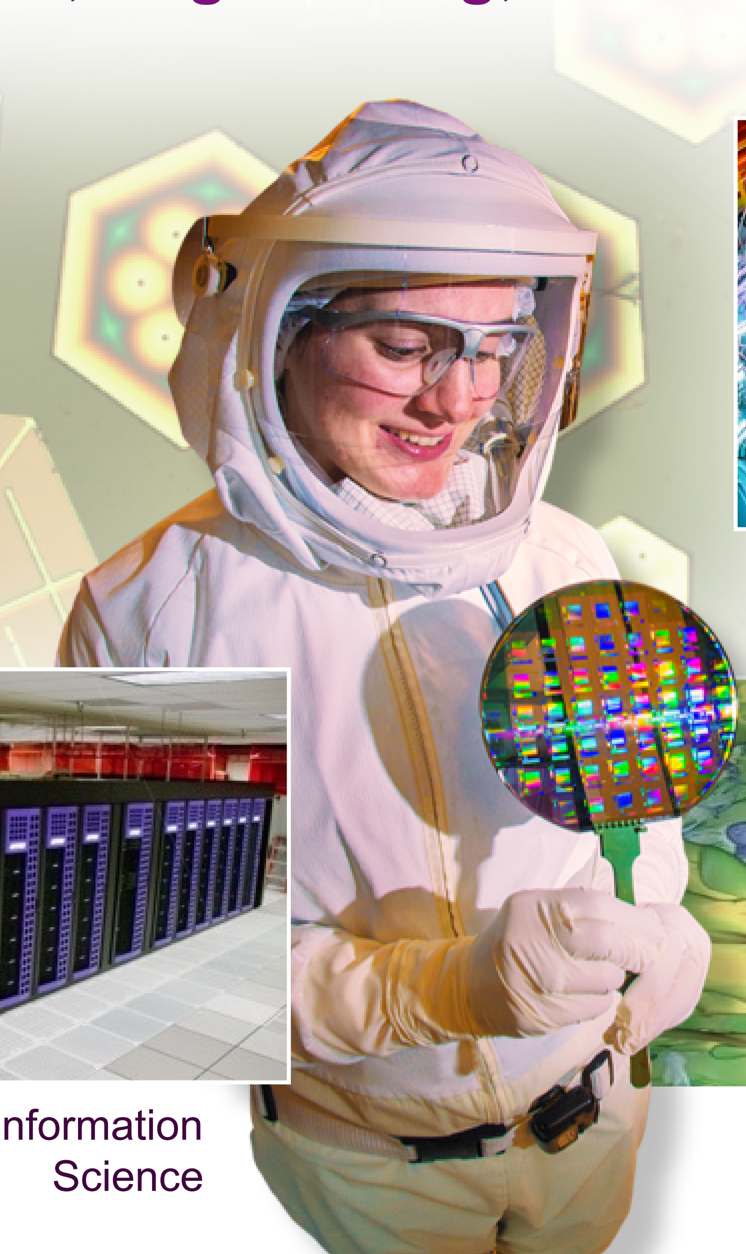
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

National Nuclear Security Administration under contract DE-NA0003525.SAND2021-4091 O

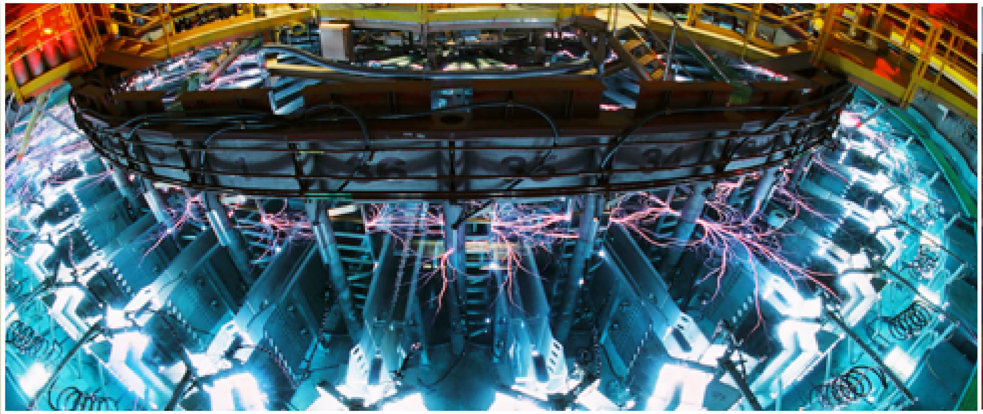
Outline

- Overall Materials Science at SNL
- Why Energy Efficiency in Computing ?
- Materials discovery and Engineering for energy-efficient computing
- Chips + Science Act
  - Boost Domestic Research & Manufacturing in Microelectronics
  - Why Sandia?

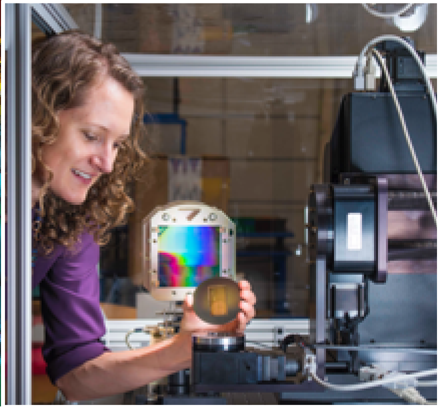
Materials Research Foundations advance the frontier-of-knowledge in the physical, engineering, and computer/information science



Nanodevices & Microsystems

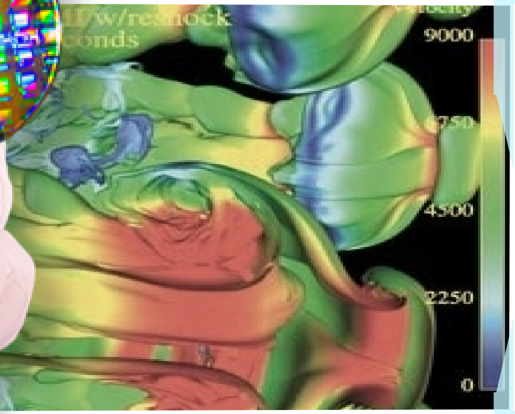


Radiation Effects & High-Energy Density Science



Materials Science

Computing & Information Science



Engineering Science

Earth Science

Bioscience

Global Energy Consumption in Information & Computing Technology

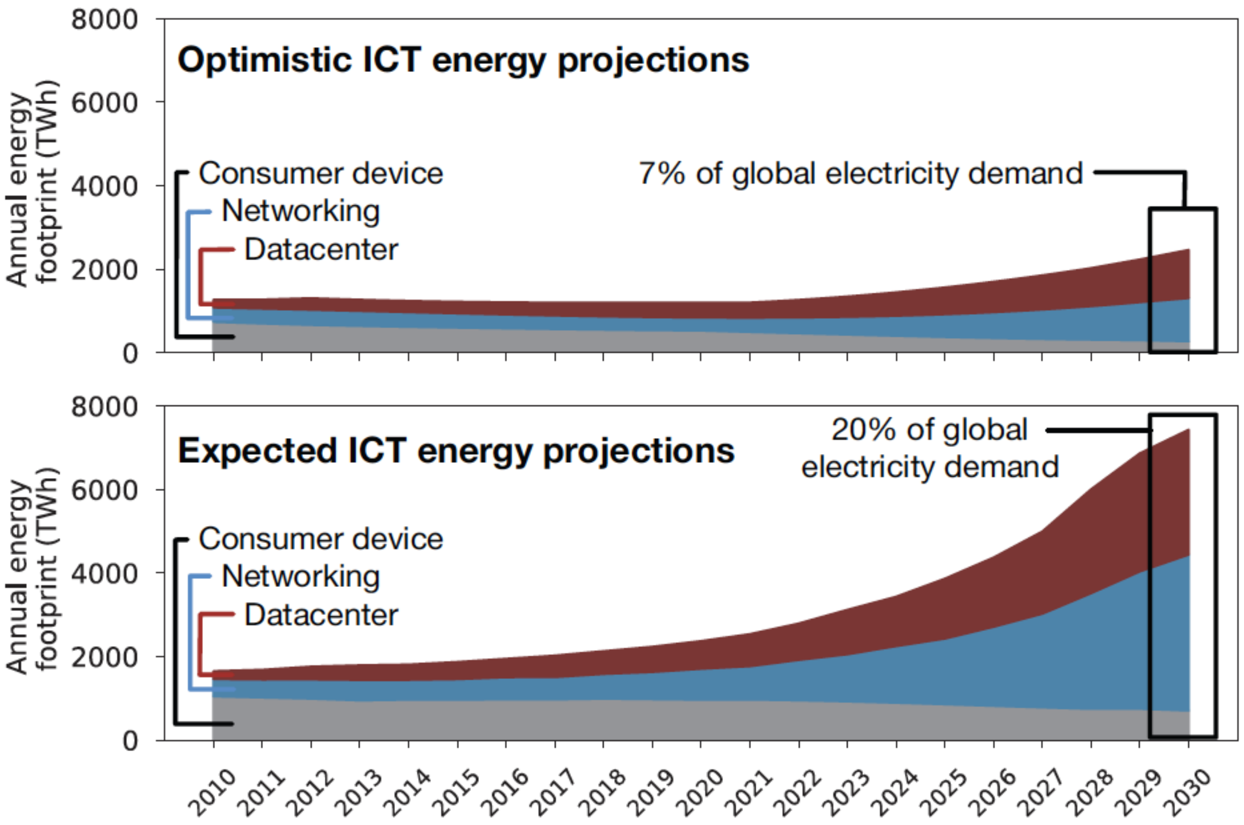
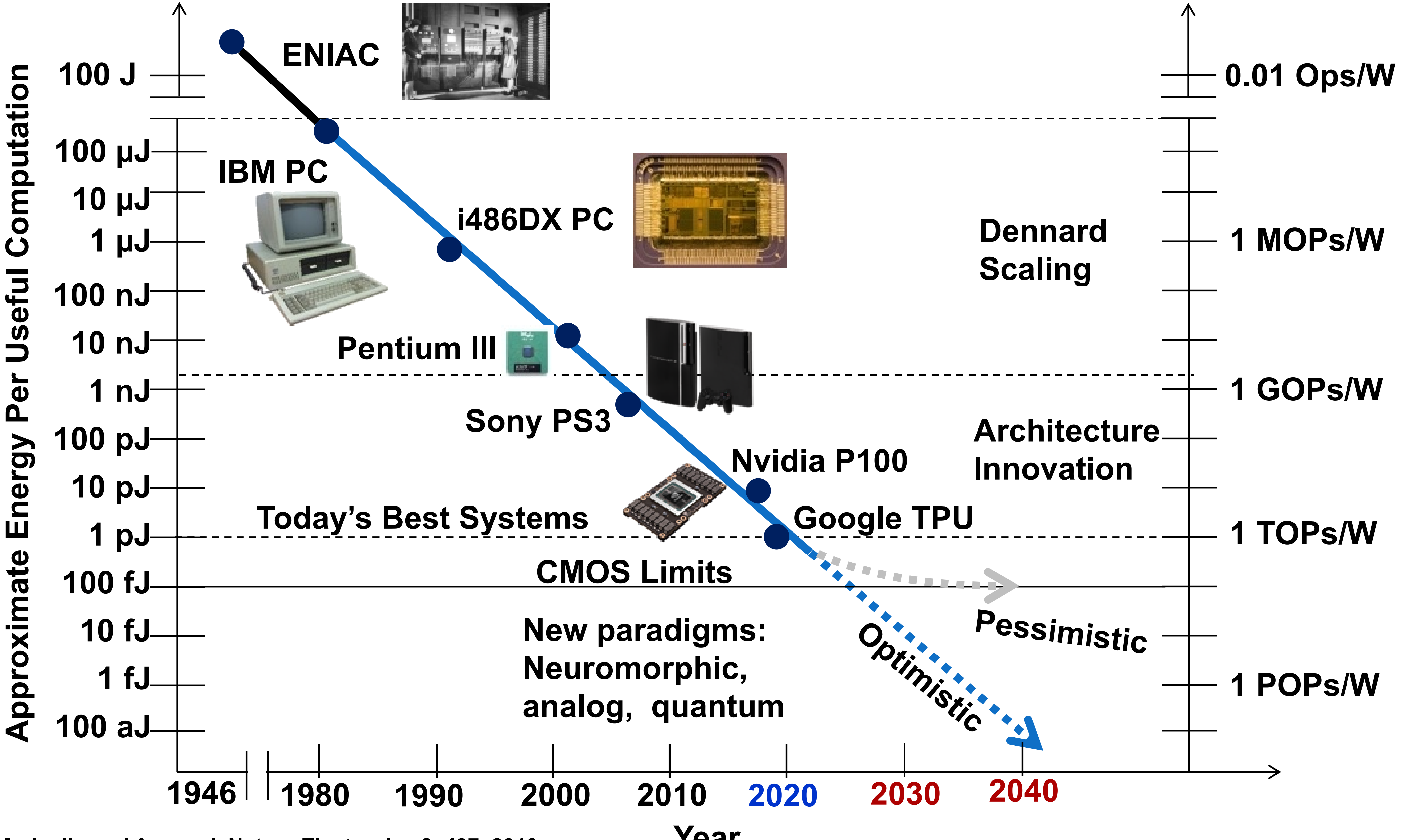


Fig. 1. Projected growth of global energy consumption by information and computing technology (ICT). On the basis of optimistic (top) and expected (bottom) estimates, ICT will by 2030 account for 7% and 20% of global demand, respectively [1].

Chasing Carbon: The Elusive Environmental Footprint of Computing (2021)
Harvard University, Facebook and Arizona State University
DOI 10.1109/HPCA51647.2021.00076

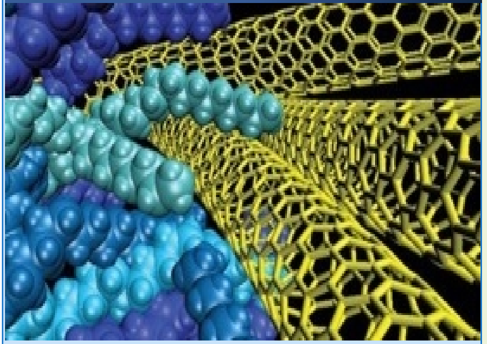
Energy-efficiency of computing platforms



Adapted from: **Marinella and Agarwal, Nature Electronics 2, 437, 2019**

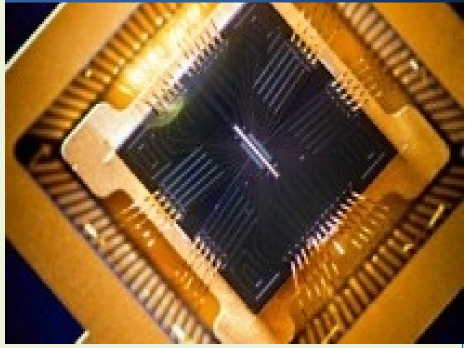
Research areas needed to advance low-energy information processing

Materials Science and Engineering



- Quantum materials
- Nanoscale materials
- Optoelectronic materials
- Organic synapses
- Superconductors
- Topological materials
- Atomic manufacturing

Nanodevices and Microsystems

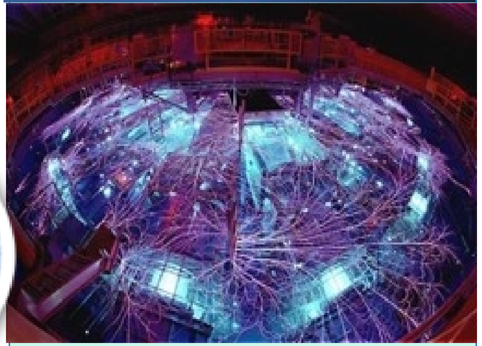


- Spintronics
- Memristors
- Mottronics
- Superconducting single quantum flux (SFQ)
- Single-electron transistors
- Magnomics
- Skyrmionics
- Piezotronics
- Optical interconnects

Computing & Information Science

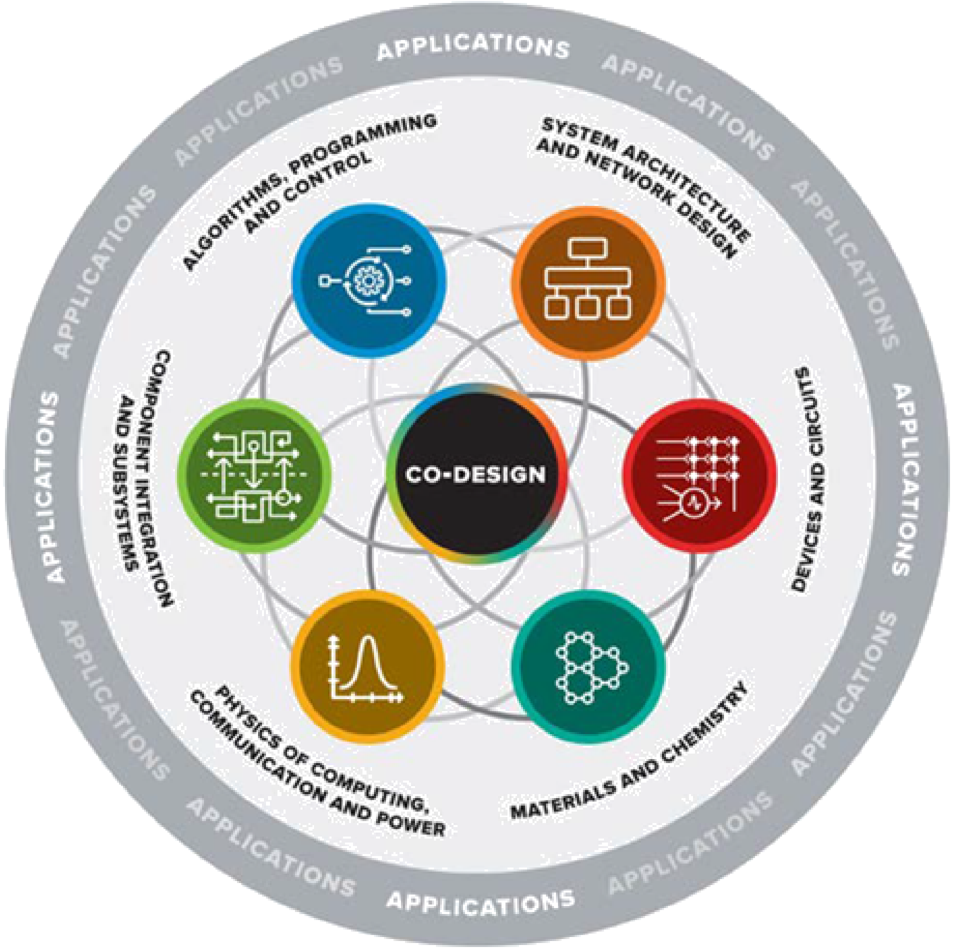
- Neuromorphic computing
- Reversible computing
- Approximate computing
- Specialized accelerators
- Quantum information science
- Low-power architectures

Radiation Environments



- Strategic radiation hardening
- Circuit performance in strategic environments

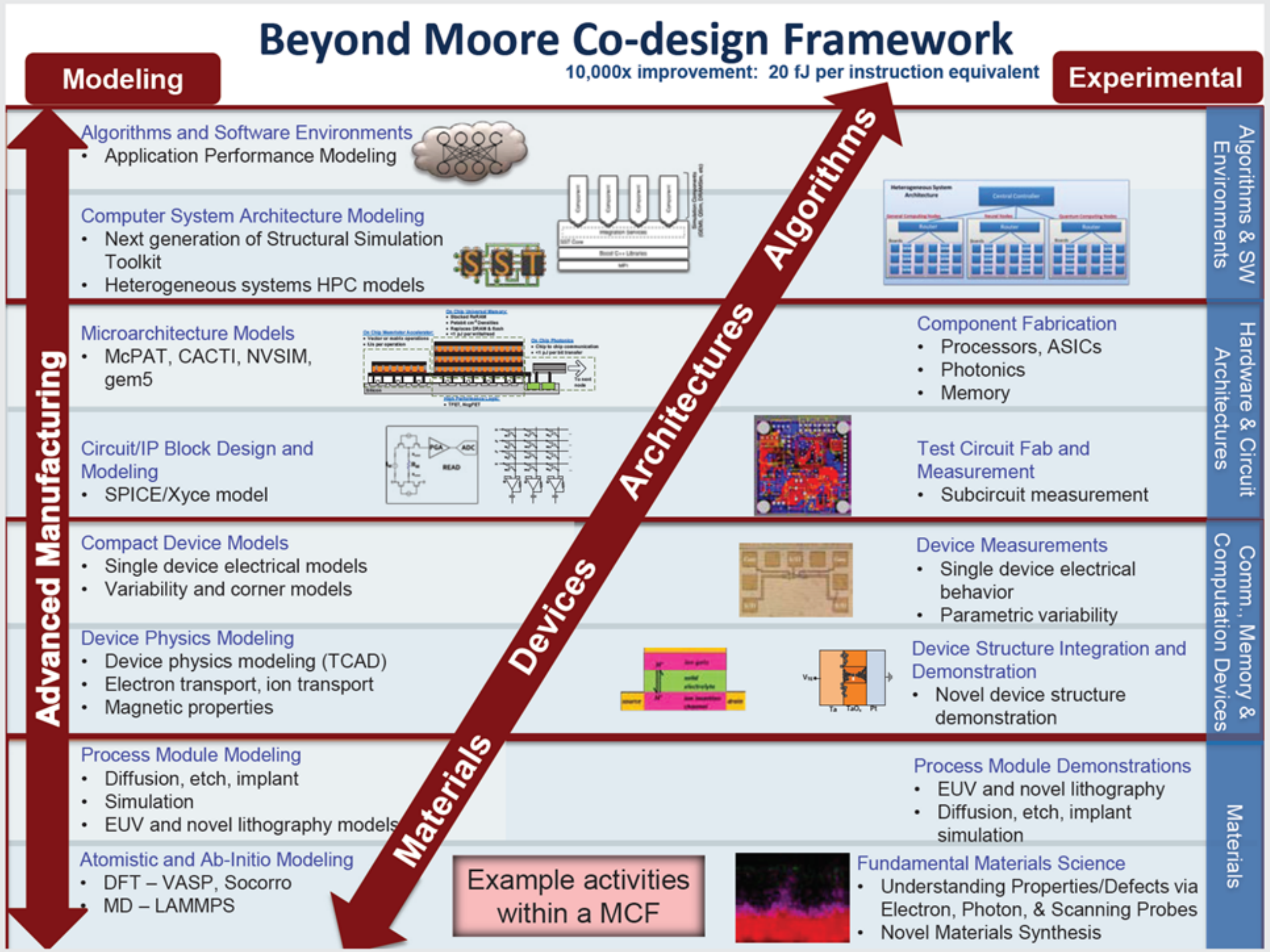
Multi-scale and interdisciplinary co-design approach is required to meet requirements for energy-efficient computing



DOE Basic Research Needs for Microelectronics, 2028.

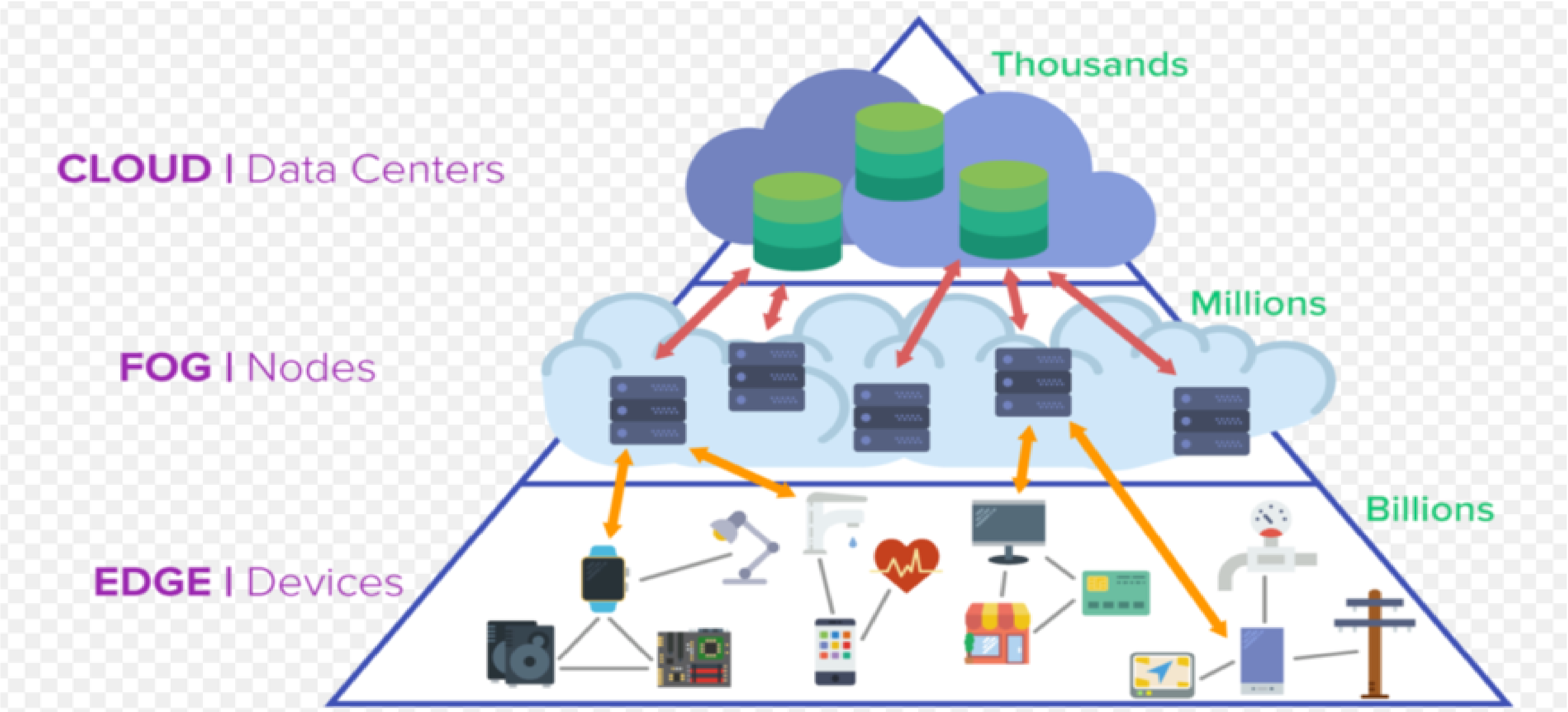
BEYOND MOORE CO-DESIGN FRAMEWORK

10,000x improvement: 20 fJ per instruction equivalent



Comprehensive view of the many computational and experimental techniques needed in the rational design of new microelectronics. Shows the need for strong integration across many levels during the design phase. McPAT = multicore power, area, and timing modeling; NVSim = nonvolatile memory simulator; TCAD = technology computer-aided design; EUV = extreme ultraviolet lithography; DFT = density functional theory; VASP = Vienna ab initio simulator package; MD = molecular dynamics; LAMMPS = large-scale atomic/molecular massively parallel simulator; MCF = Moore co-design framework; ASIC = application-specific integrated circuit. Courtesy of Matt Marinella, Sandia National Laboratories.

Energy-efficient 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.

Meeting energy-efficiency and performance requirements for edge information processing and sensing: Computing at the point-of-sensing

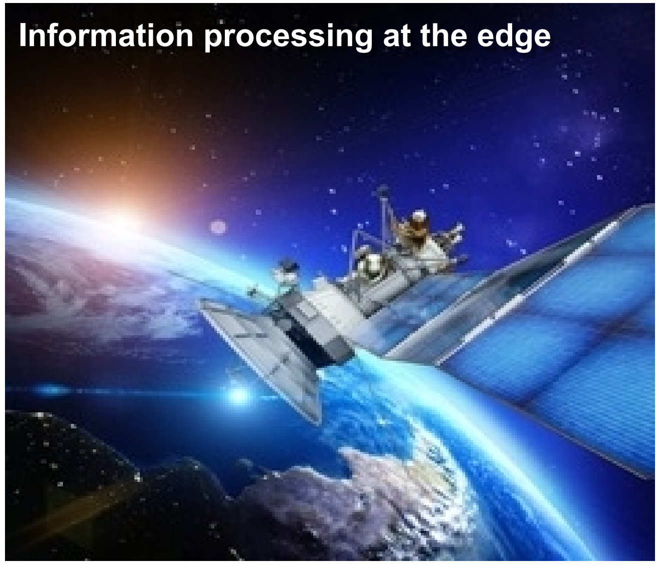
Drivers: Several applications require information processing at the point-of-sensing for real-time analysis without data transmission to centralized computer: Information processing at the "edge"

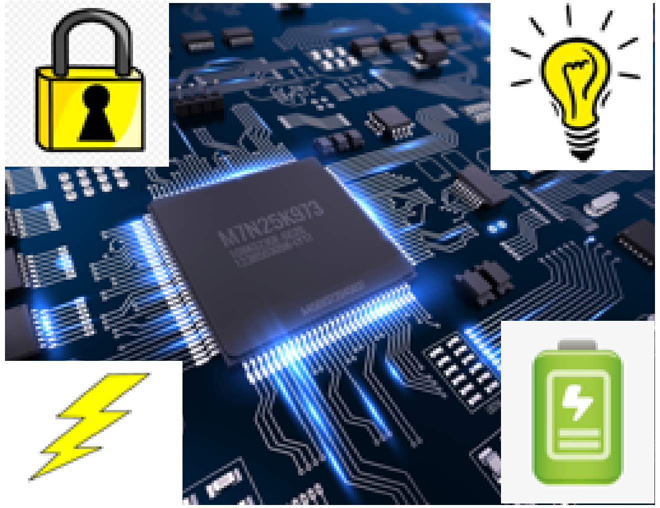
Requirements: Moving processing power to the edge

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

Energy challenge:

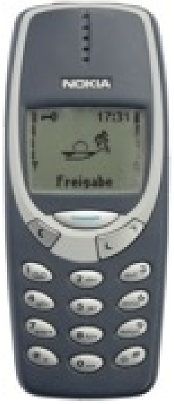
1. The (performance/power) ratio is levelling
2. Too much power is required to process too much data





Energy-efficient ultra-scale-class computing at the edge will open new applications that cannot be anticipated with today's technology

Enabled by Mega/Gigascale

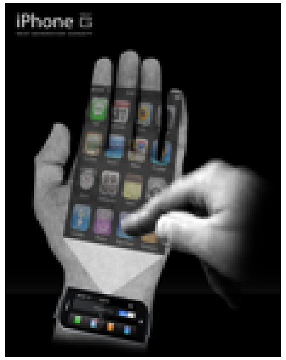


>1000x improvement

Enabled by Terascale

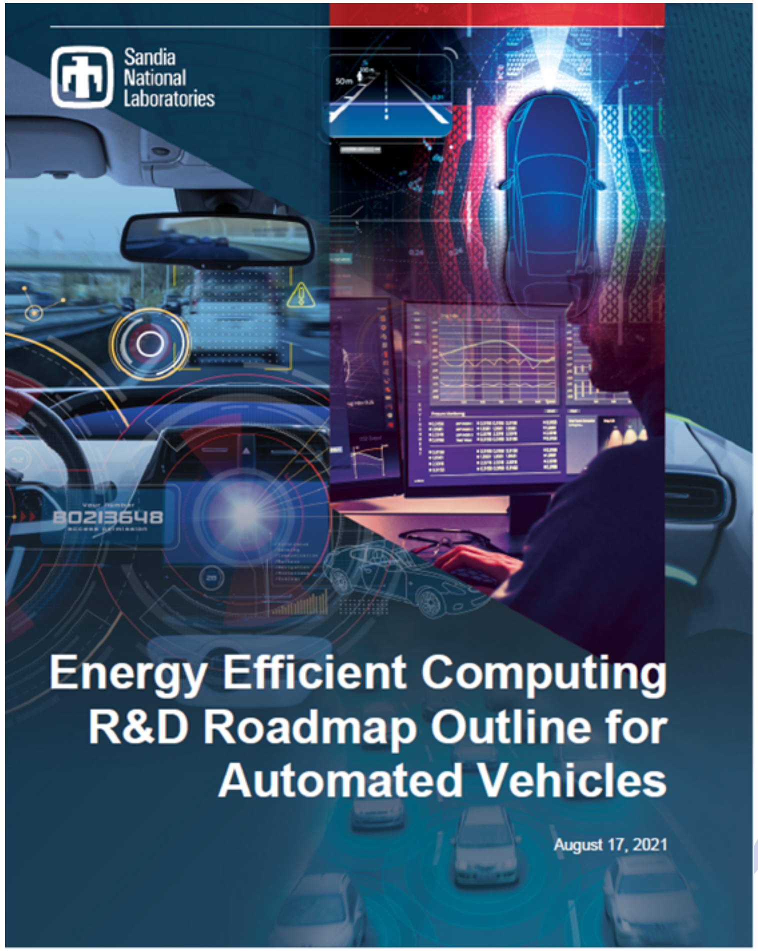
>1000x improvement

Enabled by Peta/Exascale



1990 2000 2010 2020 2030 2040

An example: Hypothetical power performance needs for highly automated vehicles (2019)



computing must meet SWaP constraints

~1 petaflops
~100 W (system)
~100 TOPS/watt

Highly automated driving

>10x less power

>10x compute

>100x power performance

TOPS = trillion (tera) operations

~100 teraflops
~1000 W (system)
~1 TOPS/watt

Significant innovation will be required in microelectronic materials and devices, sensing and computing architectures, and computer algorithms.

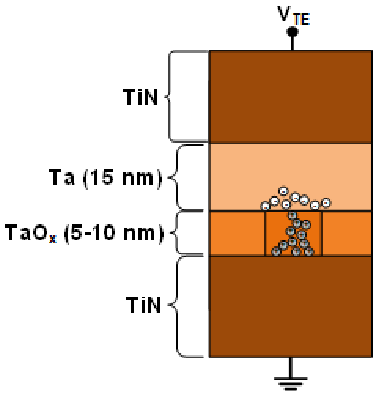
https://www.wired.com/story/self-driving-cars-power-consumption-nvidia-chip/

Commercial Leap-Ahead Technologies

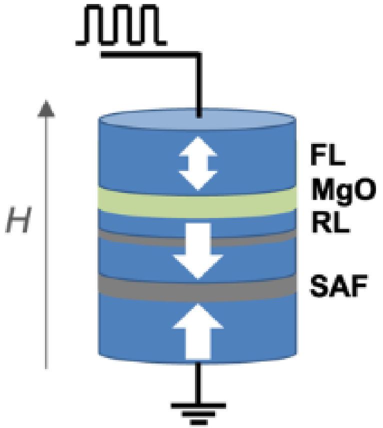
Neuromorphic, probabilistic and reversible Computing

Devices and circuits

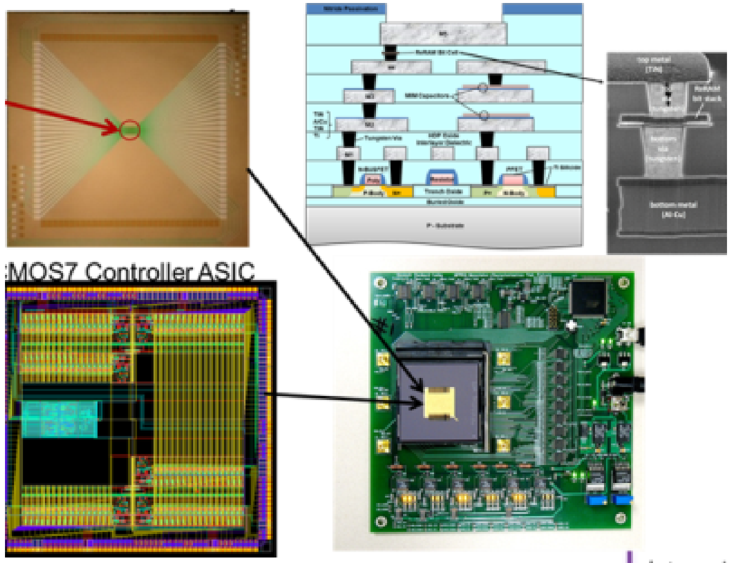
CMOS Integration



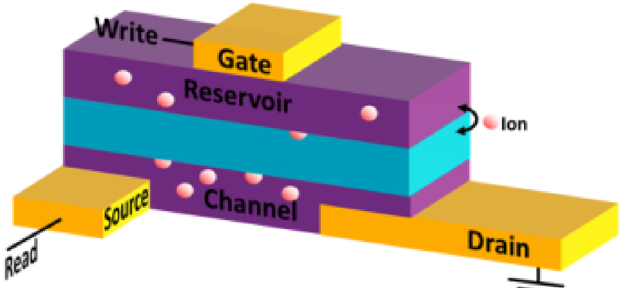
ReRAM memristor



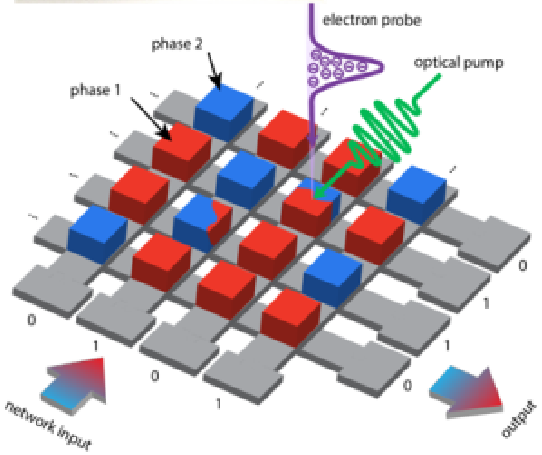
Magnetic tunnel junction



Reversible computing

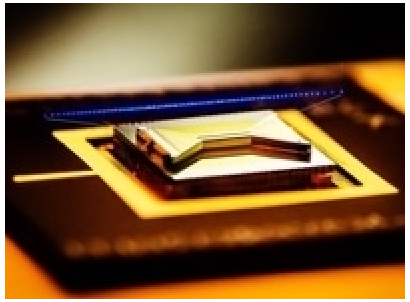


ECRAM



Computing Discovery Platform

Quantum Technologies



Peregrine ion trap

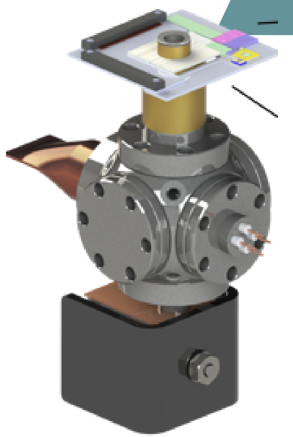
QSCOUT

Facilities

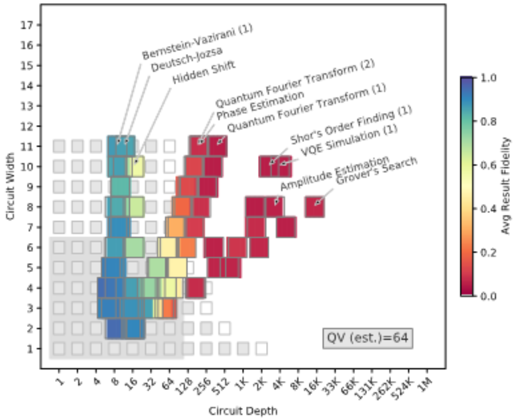
- MESA: low volume, high mix fabrication
- QSCOUT: open ion trap testbed
- IBL: counted ion implantation
- CINT: advanced fabrication, quantum materials
- QPL: benchmarking tools, like gate-set tomography
- CDC: science and technology of HW trust and security
- PRF: plasma research, including processing

Activities

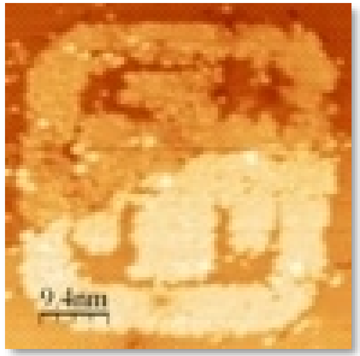
- Working devices in multiple technologies: trapped ions, semiconductor spins, superconducting circuits, neutral atoms, and optomechanical technologies.
- Applications ranging from quantum computing, quantum sensing, and networks
- Application-oriented benchmarking of noisy, intermediate scale quantum systems
- Quantum Systems Accelerator (DOE quantum hub): co-design algorithms, devices, and engineering solutions needed to deliver certified quantum advantage
- Re-application of technologies developed for quantum computing to traditional microelectronics, including computing, failure analysis, and trust/security



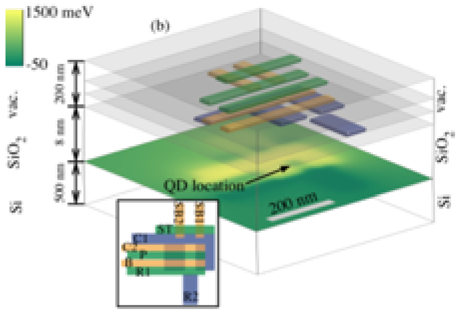
TICTOC - integrated trapped ion clock



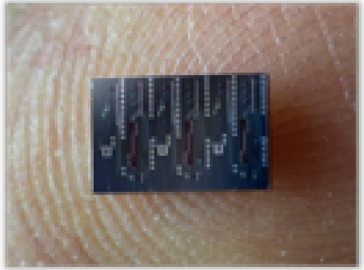
Application benchmarking



Advanced fabrication

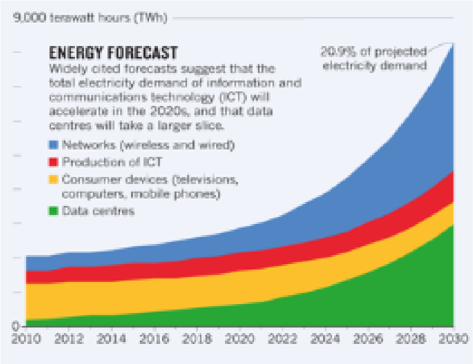


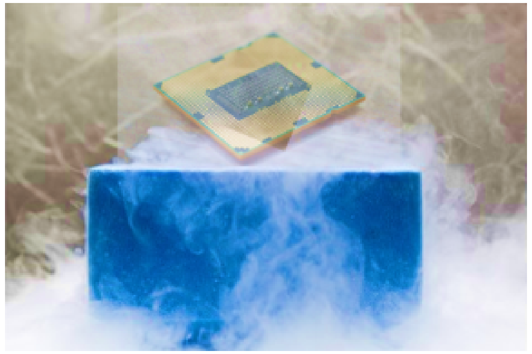
Device Modeling

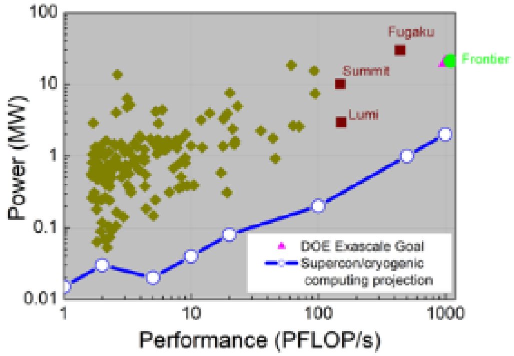


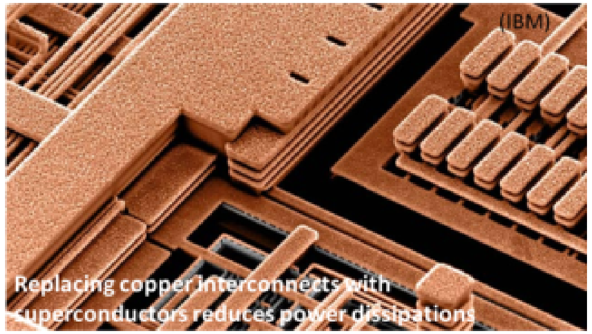
Chip-scale quantum transceiver

Cryogenic Computing



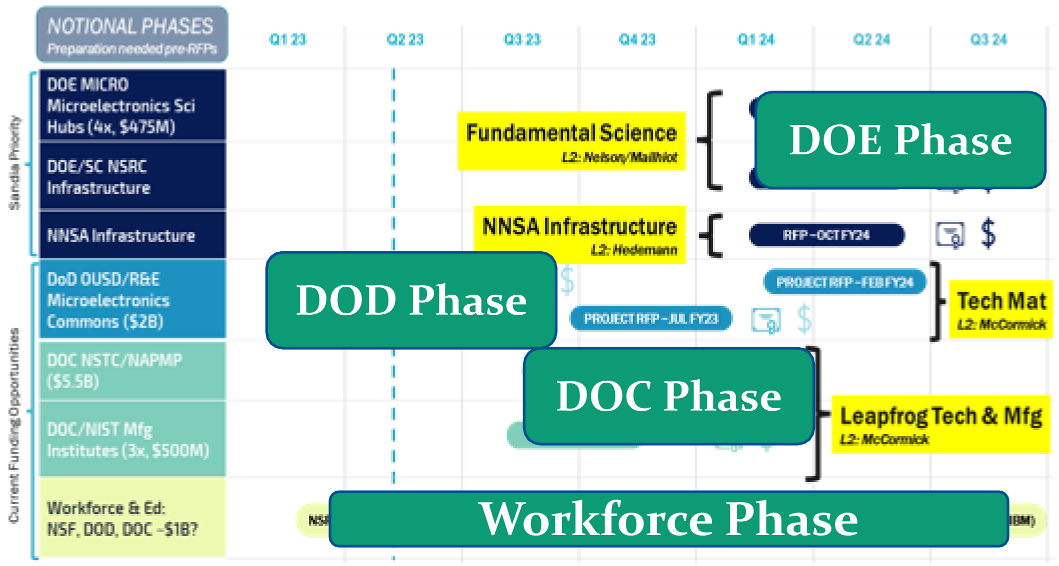


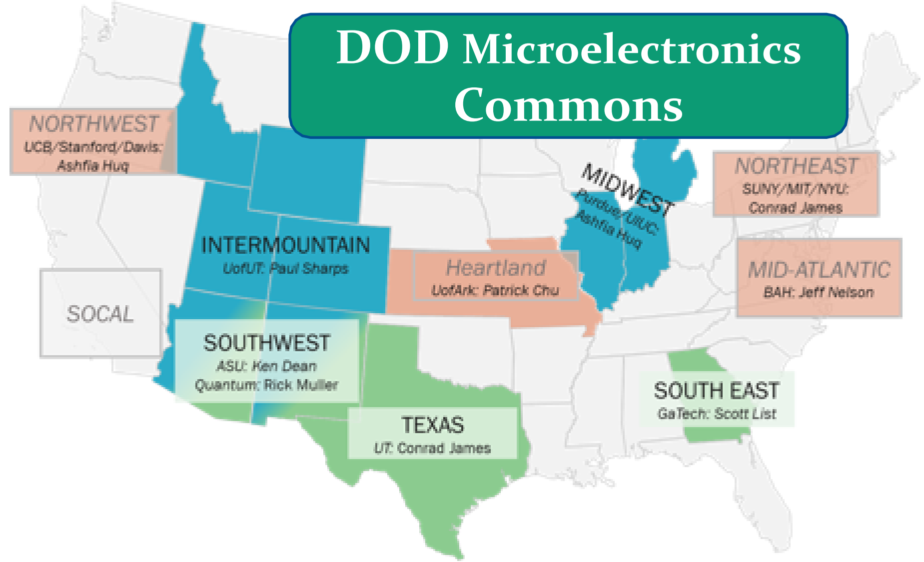




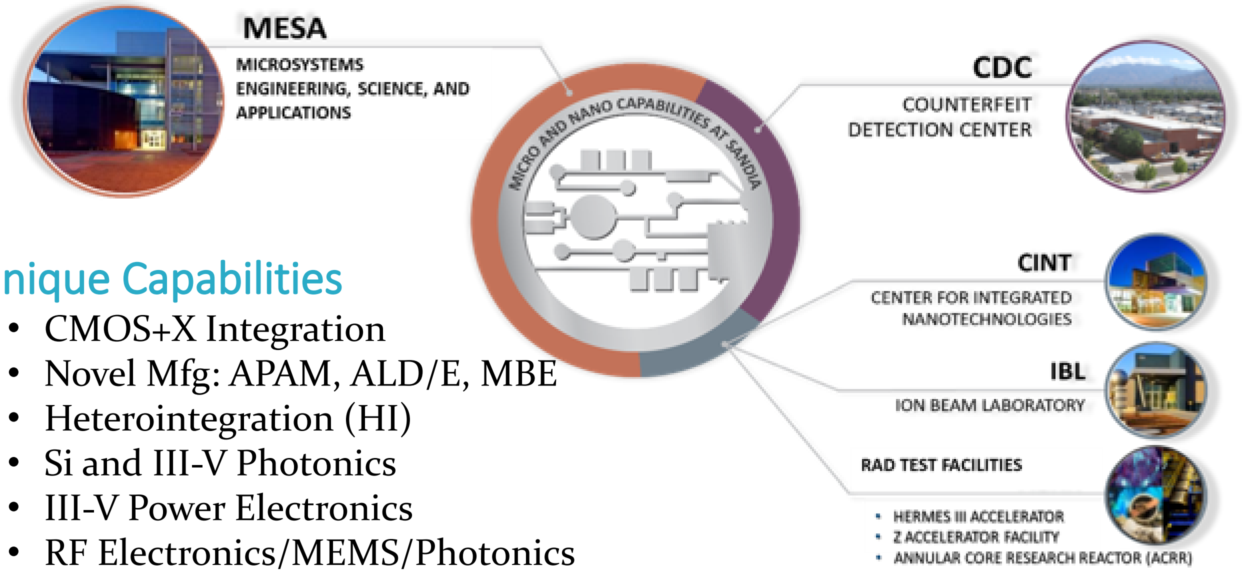
- Electricity use of information and communications technology may exceed 20% of all global electricity by 2030.
- Need to increase energy efficiency of microelectronics.
- Cryogenic computing as a promising approach for low-energy, power efficient circuit applications.
- Two examples:
  - Josephson junction (JJ) field-effect-transistors (FFTs) for Boolean operations.
  - Superconducting neuromorphic computing for beyond CMOS computing
- Cryogenic computing scheme is compatible with superconducting interconnects, huge reduction of power dissipation.

Sandia Priorities: Serve the nation & prepare for leadership in the DOE phase ➔ Engage, advise, & influence during DOD, DOC, & DOE phases





Unique Facilities



Unique Capabilities

- CMOS+X Integration
- Novel Mfg: APAM, ALD/E, MBE
- Heterointegration (HI)
- Si and III-V Photonics
- III-V Power Electronics
- RF Electronics/MEMS/Photonics
- Neuromorphic HW/SW/Algorithms
- Quantum Sensing, Info, & Networking
- RadFx and Extreme Environs
- Trust/Security & Novel Metrology
- Workforce Development (Intern/Coop Institutes, HBCU/MSI)

DOC/DOS/Workforce/EDA

- DOC: Mostly same players as DOD, $39B Factory Incentives , $11B R&D
- DOC: Factories➔ 3 Mfg. Institutes➔ R&D Programs (NSTC/NAPMP)
- DOS: Export Control, etc., $100M
- NSF & DOC Workforce: SRC, ASA/SEMI, Hubs, Industry, Labs, $200M+
- Econ. Dev. Admin (EDA): Tech Hubs Program coordination w/CHIPS, $500

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

Courtesy of J. Nelson, Sandia National Laboratories

Capabilities for security and trust, extreme environments, and the science of semiconductors amplify our mission impact for NNSA, DOE, and the Nation

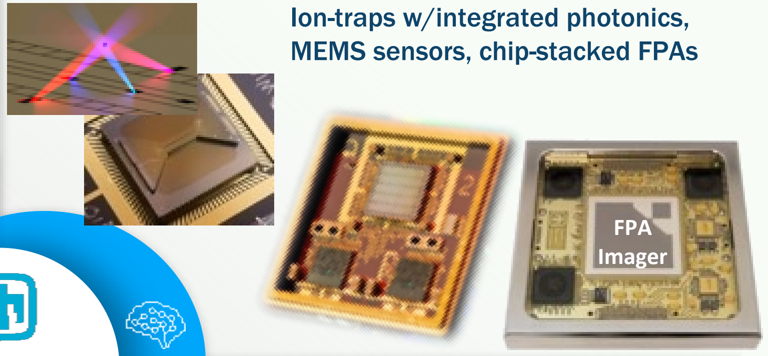
Sandia's Priority S&T Drivers

Flexible, low-volume, high-mix fabrication capabilities for agile delivery of innovative low-volume technologies: Needs span > 7 technology platforms

~300,000 parts across 44 products:

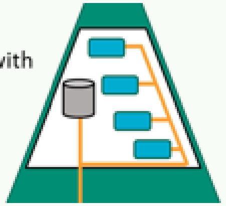
- 13 Si CMOS7 ASICs
- 8 III-V HBT SSICs
- 1 multi-kV GaN Diode
- 1 MEMS Sensor
- 2 Photonic Arrays
- 1 Optoelectronic Device
- 2 Focal Plane Arrays
- 16 RFICs

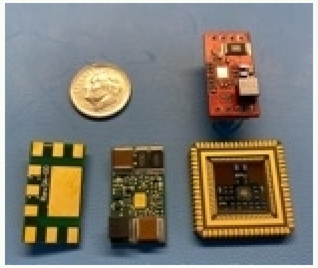
Heterointegration technologies to enable rapid realization of novel/reliable functionality for emergent applications: Ion-traps w/integrated photonics, MEMS sensors, chip-stacked FPAs



Modernized power electronics systems using wide bandgap semiconductors and magnetic materials for agile ND systems, power grid, and electric drive trains with resiliency, radiation hardness, and energy-efficiency: modular ND Power Bus architectures and modules

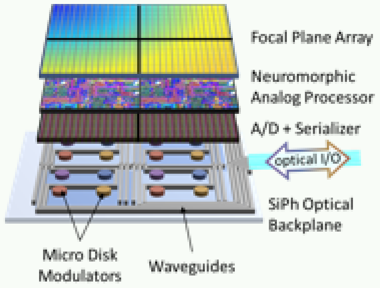
Bus-based approach with PoL power converters





Specialized computational accelerators and architectures that use quantum, neuromorphic, and other bespoke accelerators for revolutionary performance and efficiency in Edge and HPC applications.

Neuromorphic FPA



Quantum Accelerator

Science Frontiers

Novel metrology, materials, and nanoscience **innovations** that accelerate technology development

Atomic-scale manufacturing capabilities that surpass limitations in current fabrication

A ***laboratory of the future paradigm*** **exploits** AI-enhanced co-design for semiconductor-based technologies