U.S. DEPARTMENT OF ENERGY

Office of ENERGY EFFICIENCY & RENEWABLE ENERGY

ADVANCED MATERIALS & MANUFACTURING TECHNOLOGIES OFFICE

EES2 ENERGY EFFICIENCY SCALING

Increasing Computing Energy Efficiency is Key Requirement for Sustainability

Tina Kaarsberg, PhD
EES2 Lead and Acting Program Manager

17 April 2024

March/ April 2024 Headlines

The Washington Post
Democracy Dies in Darkness

Amid explosive demand, America is running out of power

AI and the boom in clean-tech manufacturing are pushing America's power grid to the brink. Utilities can't keep up

By Evan Halper

March 7, 2024 at 6:05 a.m. EST

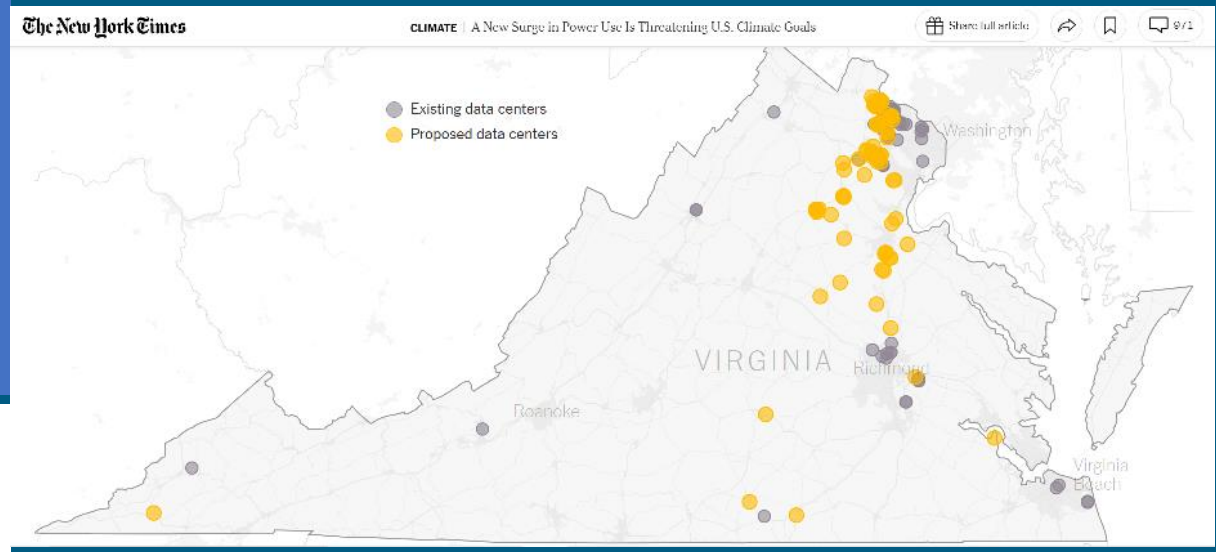
APRIL 2024 | PHYSICS TODAY

Will AI's growth create an explosion of energy consumption?

AI could gobble up a quarter of all electricity in the U.S. by 2030 if it doesn't break its energy addiction, says Arm Holdings exec Fortune. April 16, 2024 at 5:36 AM EDT

A New Surge in Power Use Is Threatening U.S. Climate Goals

A boom in data centers and factories is straining ele ctric grids and propping up fossil fuels By Brad Plumer and Nadja Popovich March 14, 2024

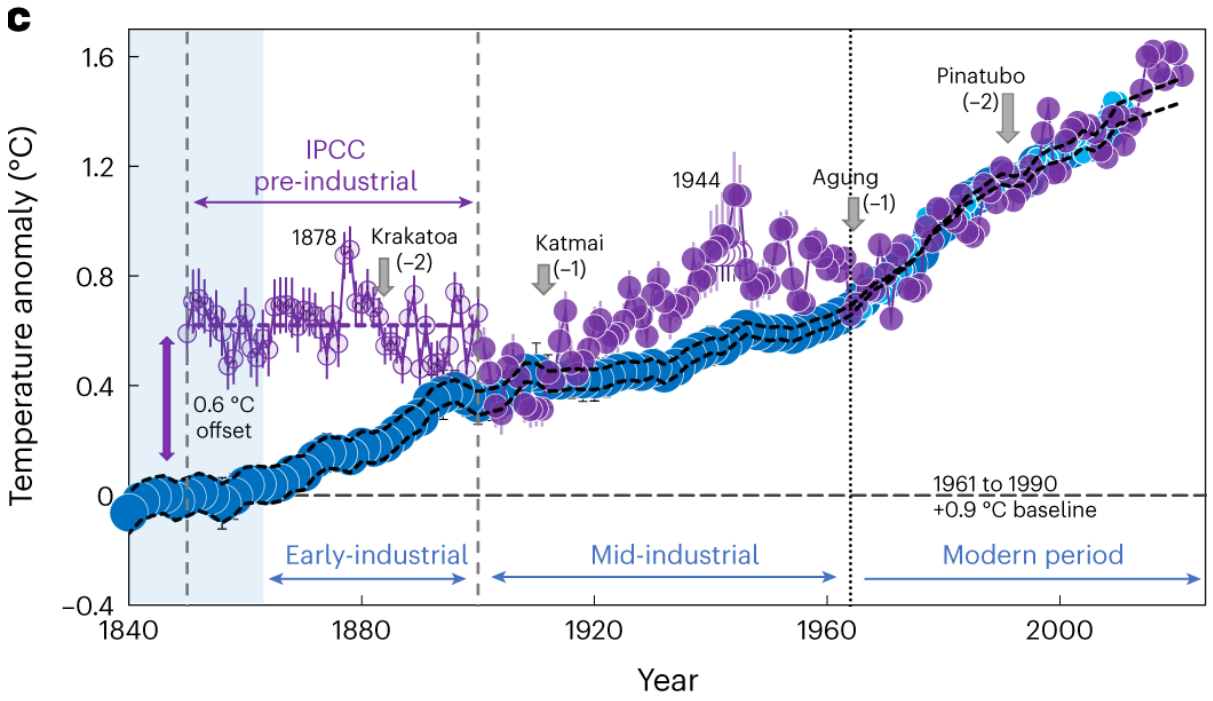


Big Tech's Latest Obsession is Finding Enough Energy

The AI boom is fueling an insatiable appetite for electricity, which is creating risks to the grid and the transition to cleaner energy sources

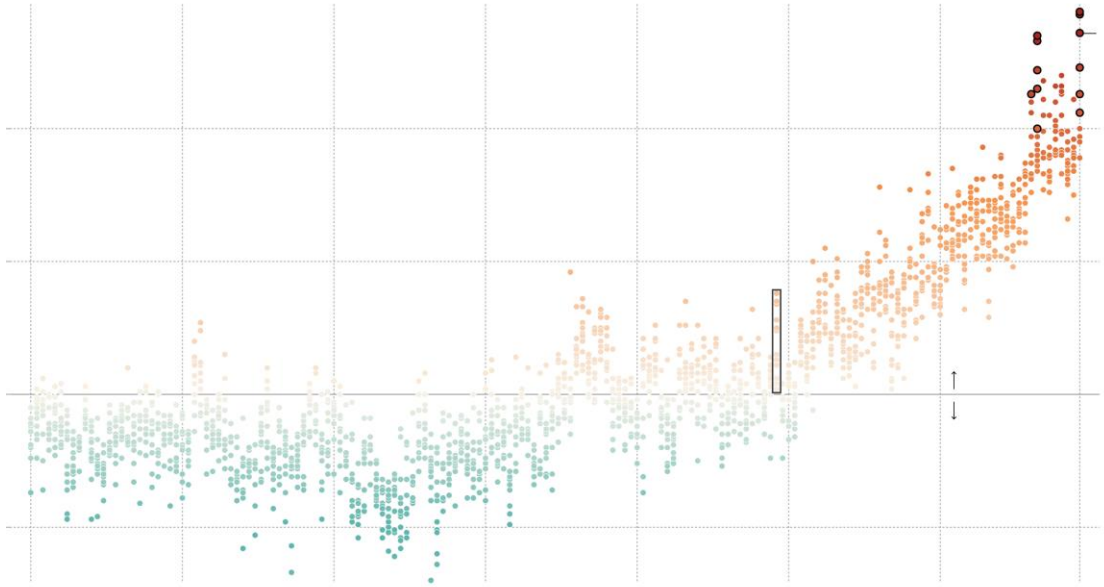
Katherine Blunt & Jennifer Hiller March 24, 2024

New Urgency I: Dangerous Threshold Crossed→ Energy Use



Malcolm T. McCulloch et al, Nature Climate, 2024

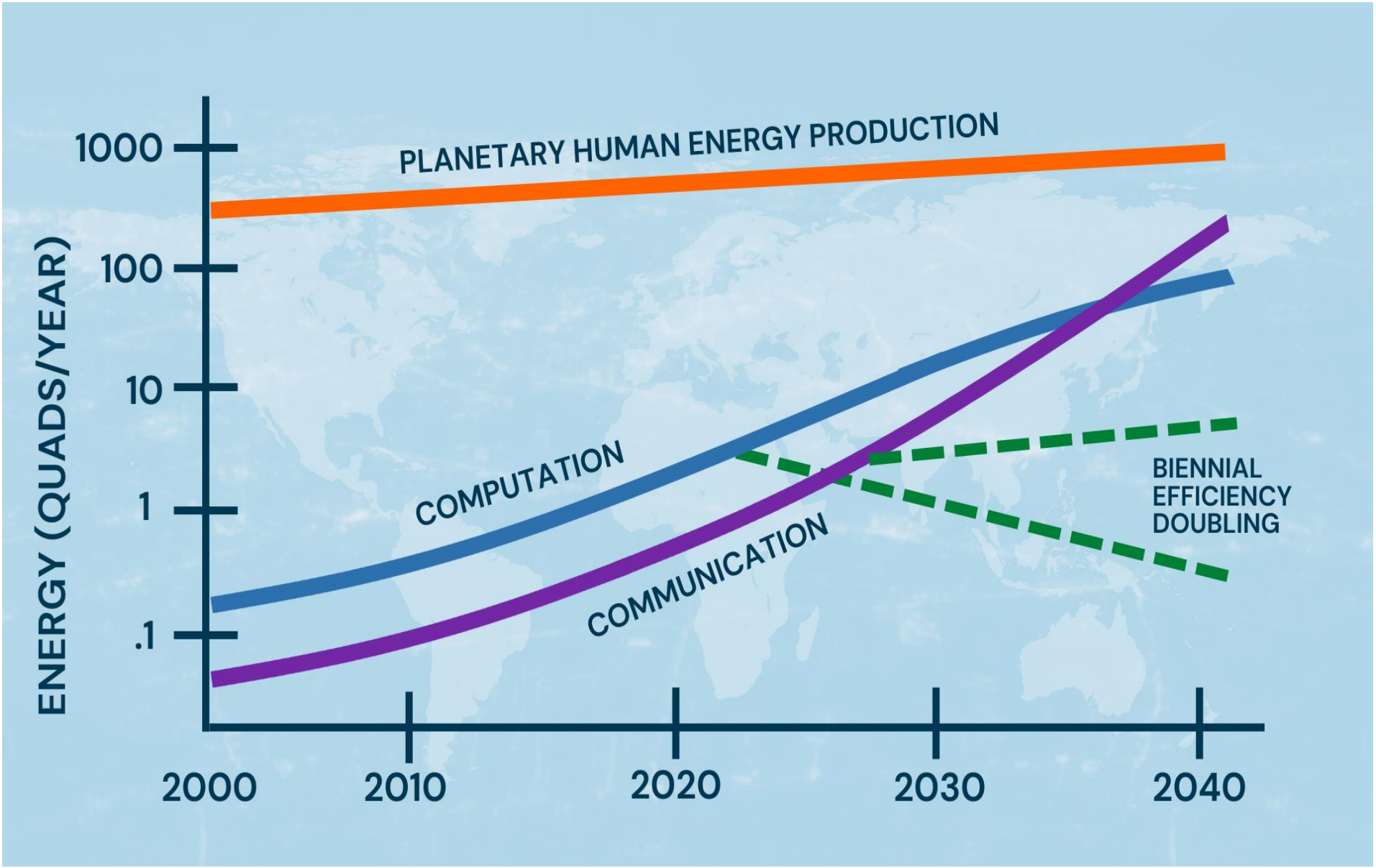
300 years of sclerosponge thermometry shows global warming has exceeded 1.5 °C



NY Times, December 26, 2023

Scientists are already busy trying to understand whether 2023's off-the-charts heat is a sign that global warming is accelerating.

Unsustainable Microelectronics Energy Use /EES2 Goals



EES2 ENERGY EFFICIENCY SCALING

Semiconductor Research Corporation (SRC) projected microelectronics could consume nearly 25% of planetary energy production

Energy Efficiency Scaling for 2 Decades (EES2)

ANNOUNCED by AMO in JANUARY 12 2022:

Shift from R&D Roadmaps based on biennial length-based scaling (e.g., Moore's law) to ultra-energy-efficiency scaling and ensure all R&D includes some energy-efficiency focus

- Specifically, develop in partnership with U.S. and Allied Country Semiconductor Industry an RDD&D roadmap to ensure
  - Doubling* of microelectronics' energy efficiency every two years *or faster* for the coming decades
  - In two decades, increase energy efficiency of next generation microelectronics by >1000X

https://ees2.SLAC.stanford.edu

The EES2 Cooperation Pledge

We the undersigned agree to cooperate

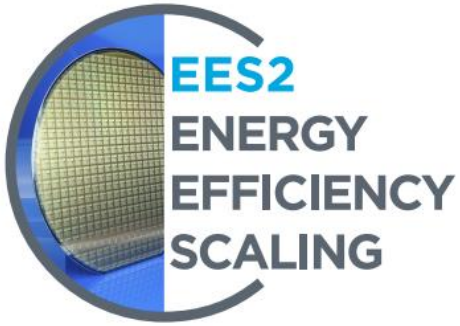
- To document and learn from the extraordinary record of microelectronics', including power electronics', energy efficiency such as increases greater than 1,000,000x in energy efficiency since the invention of the transistor nearly 75 years ago;
- To document and learn from microelectronics' past and forecasted future ability to enable all sectors of the economy to become more energy efficient and sustainable;
- To identify and publicize problems solved and opportunities offered by microelectronics' Energy Efficiency Scaling over 2 Decades (EES2);
- To participate in the AMMTO-led EES2 2022-2023 R&D Roadmap;
- To explore formation of a partnership, perhaps "EES2 Allies" that enable the EES2 1000X efficiency goal by leading EES2 R&D Roadmapping after 2025 and by catalyzing the deployment of cost-effective technologies, including power electronics, needed to stay on the EES2 path of doubling microelectronics' energy efficiency every two years.

We do this because

- Microelectronics' life-cycle energy use is rapidly becoming unsustainable as microelectronics demand begins to outpace continuing efficiency improvements due to burgeoning computing, communication, and electrification demands
- EES2 is a key organizing principle that aims to help meet new energy demands
- The EES2 is a technology leadership path that provides economic and other public benefits.

Recruited 65 pledge signing organizations for the DOE's Energy Efficiency Scaling for 2 Decades (EES2) Goal since 2022

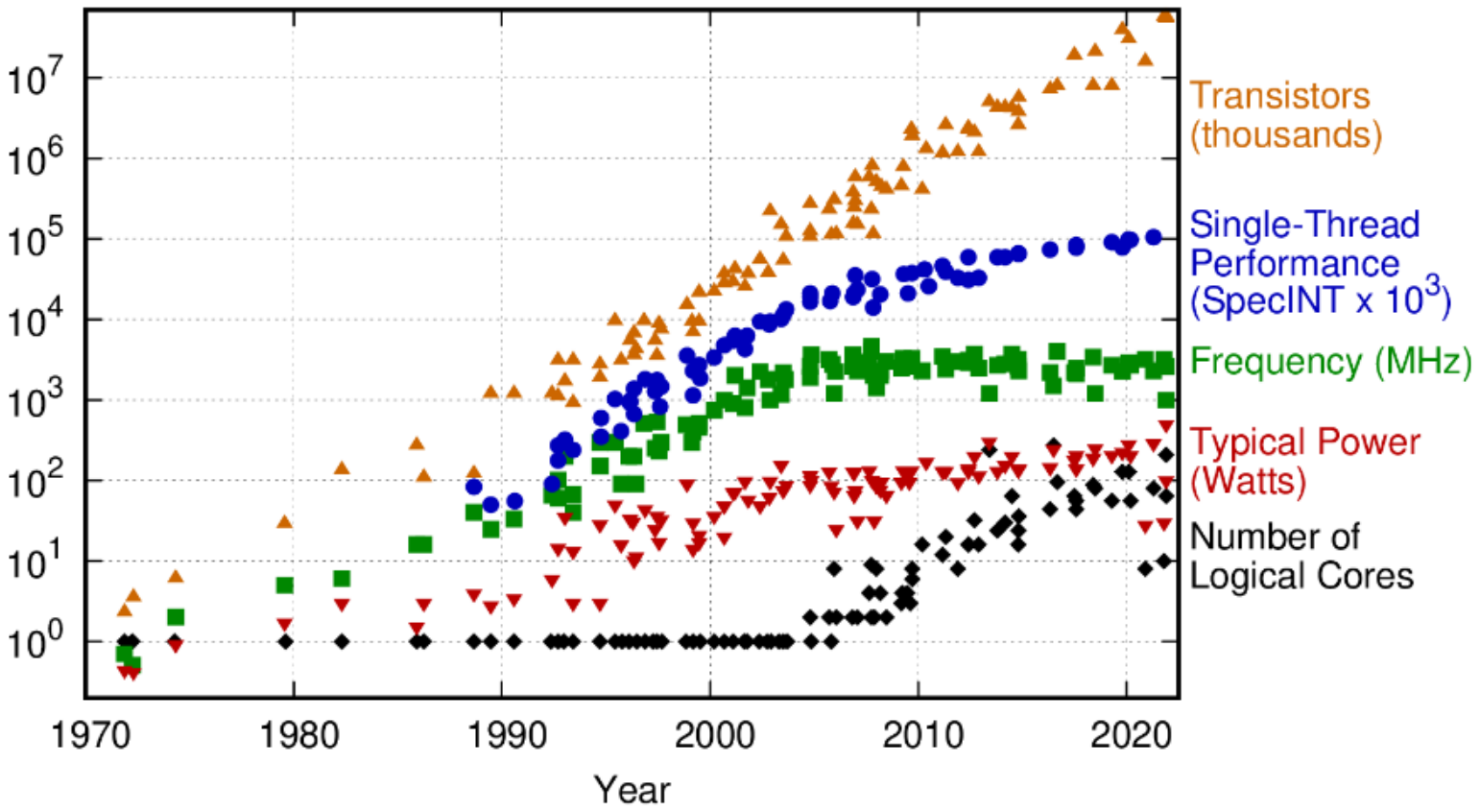
Roadmap Version 1.0: Energy Inefficiency Factors



Semiconductor energy use doubled every three years 2010—2020 due to 3 factors:

1. End of Dennard Scaling
2. Memory Access Bottlenecks
3. AI/ML Model Complexity & Other Energy-Intensive Emerging Applications

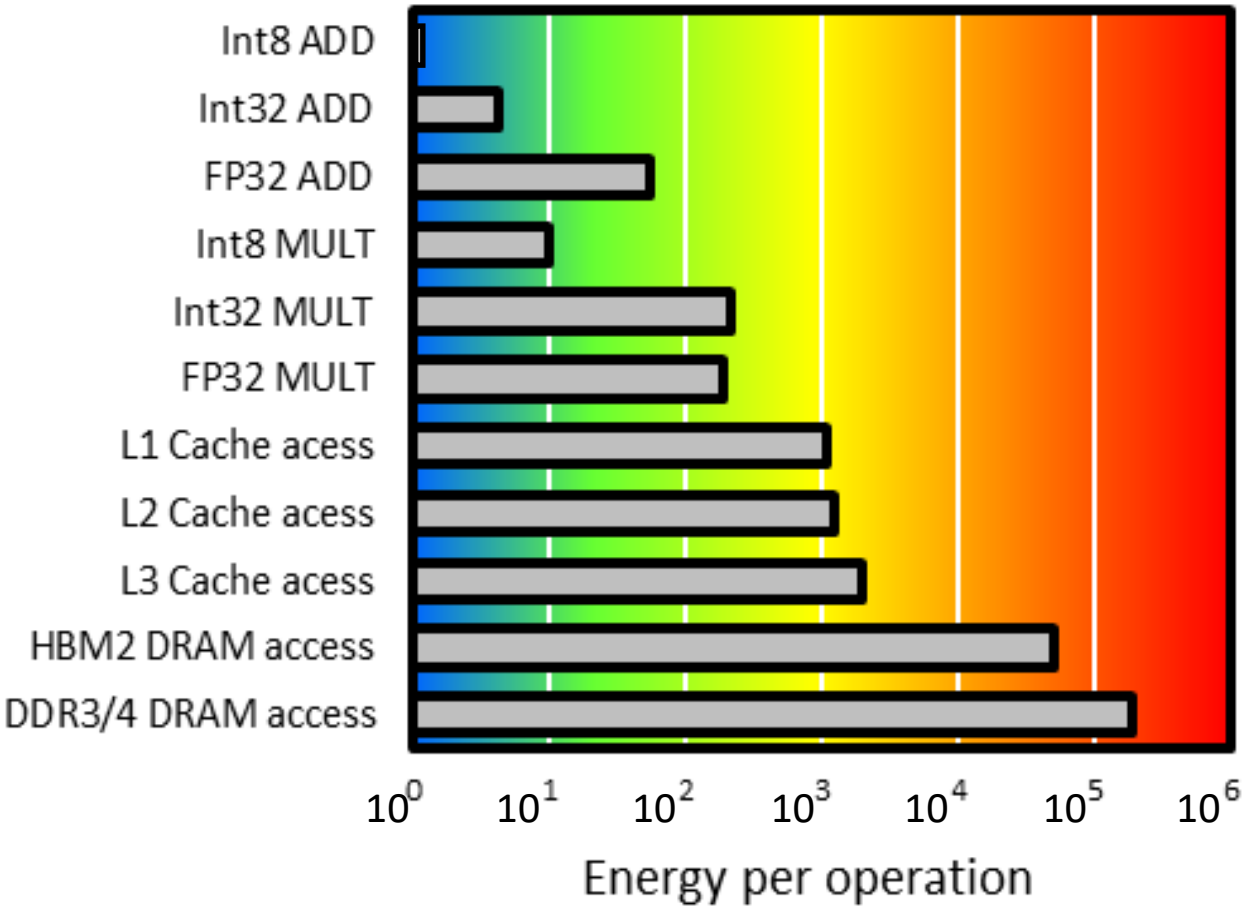
1. End of Dennard Scaling (& other Scaling Changes)



Moore's Law Modification: Continued transistor density growth needs 3D

Dennard Scaling Ended: automatic efficiency improvements stopped around 2005

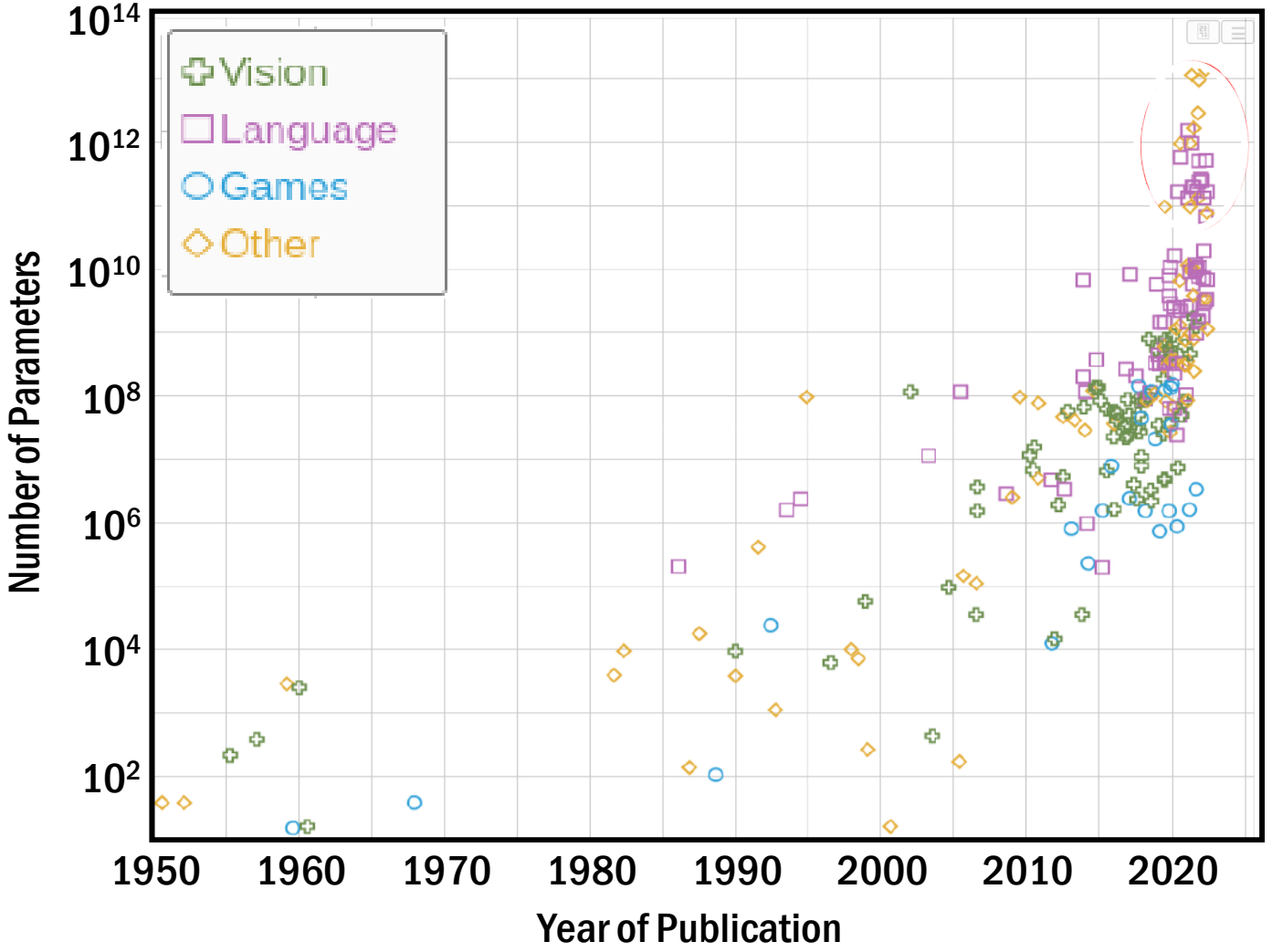
2. Memory Access Bottlenecks



Source: Jouppi, Norman P., et al. 2021. “Ten Lessons from Three Generations Shaped Google’s TPUv4i: Industrial Product.”

Memory Wall problem:
Relative energy cost of logic and memory operations is too high

3. AI/ML Model Complexity



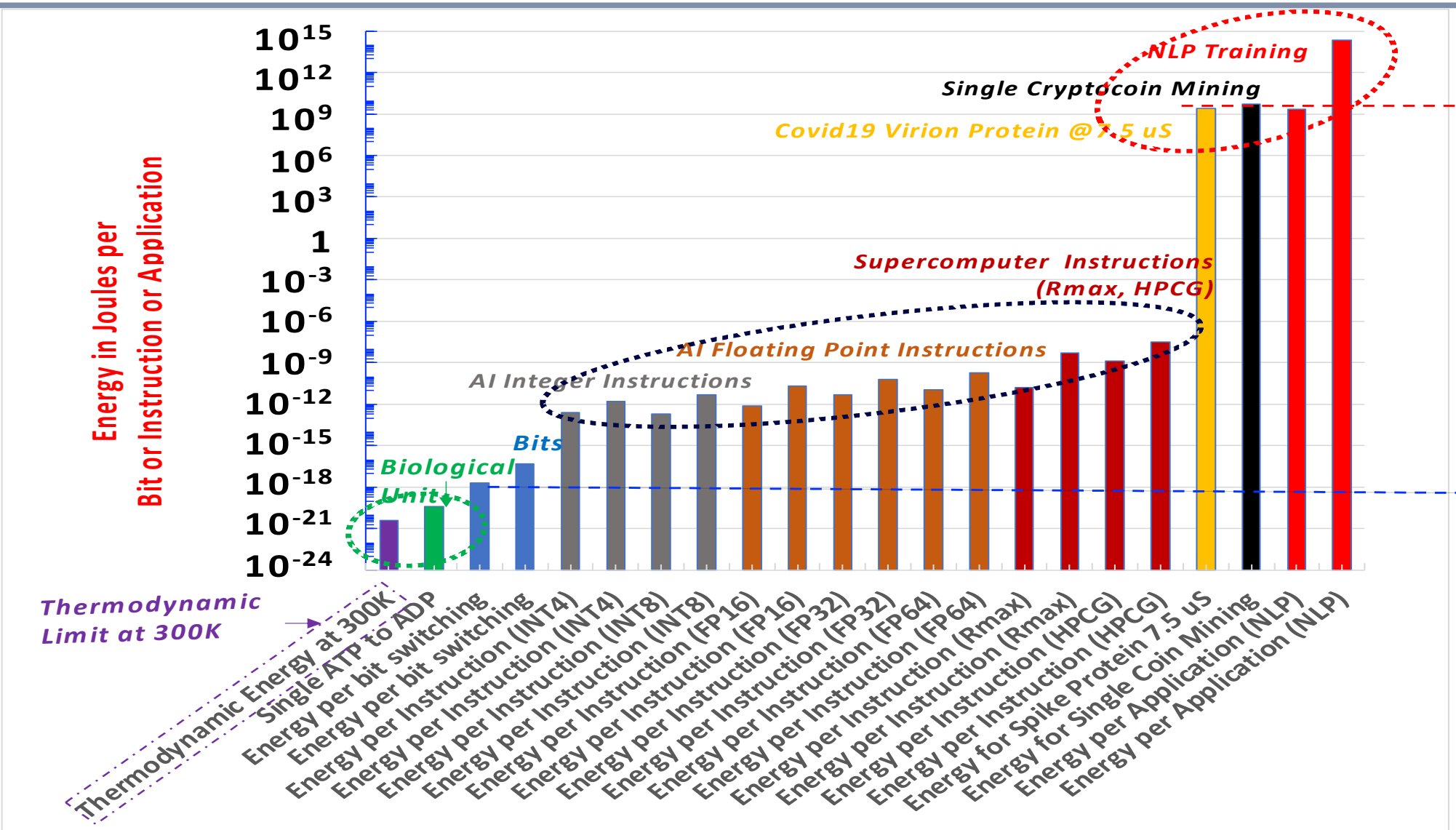
Complexity of machine learning models

(Source: Villalobos et.al. 2022)

EES2 ENERGY EFFICIENCY SCALING

The last decade has witnessed a surge in AI/ML model complexity (by more than 100,000) — a major driver of escalating energy demands

3) cont.... Seminal Shankar Analysis: Bits to Bitcoin

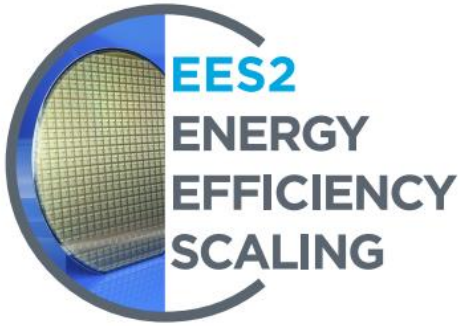


10^{9} to 10^{15}

10^{-18} to 10^{-21}

27 to 36 orders of magnitude
(From Algorithms to Atoms)

Three Metrics for Energy Efficiency



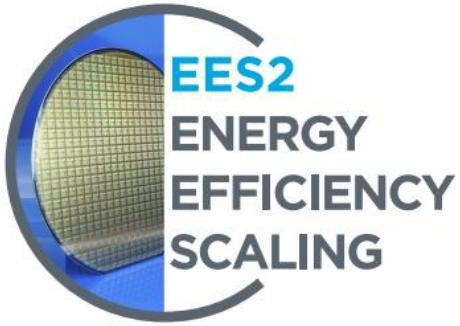
	Near Term: Energy per Instruction	*Near-Med-Term* Energy per Bit	*Med-Long Term*: Energy per Application
Materials		x	x
Devices		x	x
Circuits	x	x	x
Architecture	x		x
Heterogeneous Integration	x		x
Advanced Packaging	x		x
Algorithms			x
Software			x

(Key: light blue= near-term; dark blue = long-term)

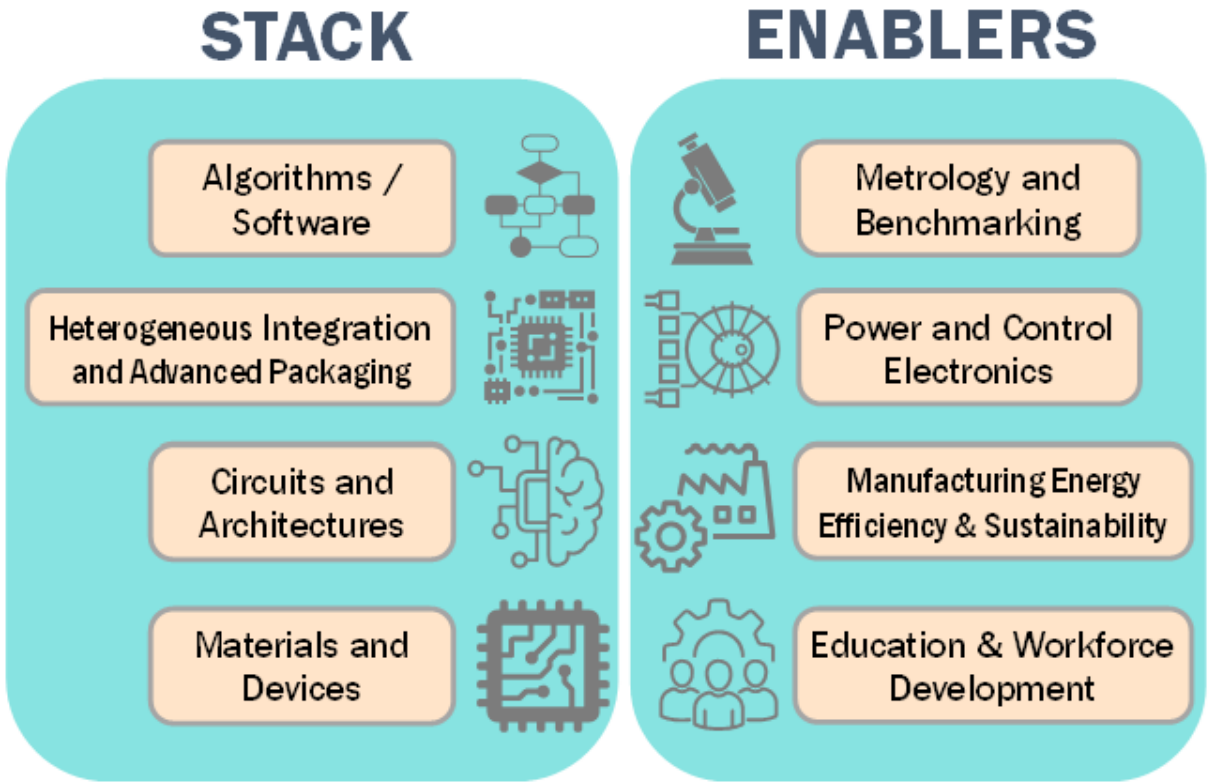
Roadmap Recommendations Divided into Three Categories:

1) Instructions (hardware and software codesign),
2) Bits (hardware), and
3) Applications (software-driven full-stack co-design).

EES2 Roadmap V1.0 (Compute) Working Groups



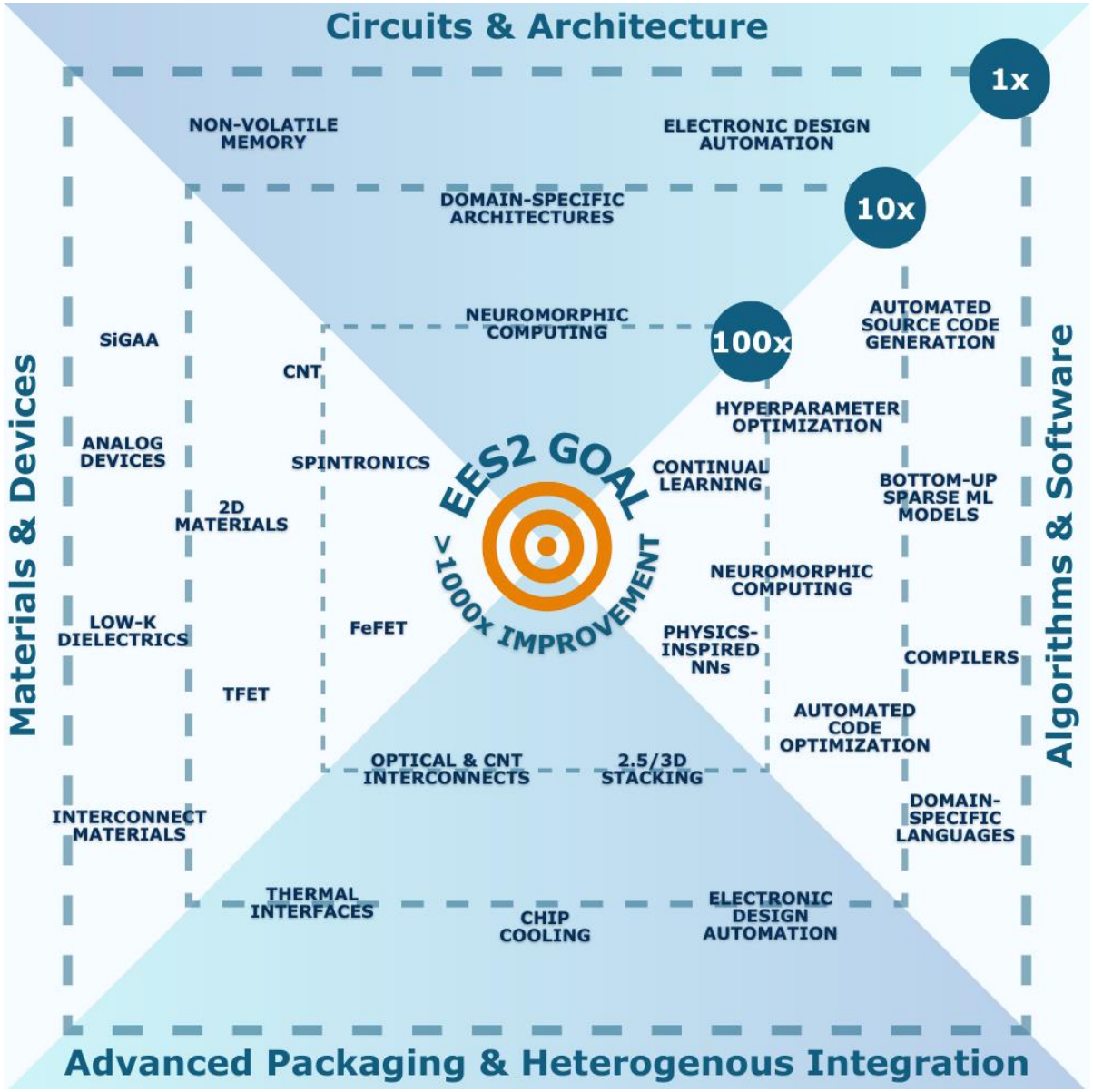
Compute stack Working Groups assembled in co-design pairs



Enablers Working Groups supporting stacks to enable

1000X efficiency goals

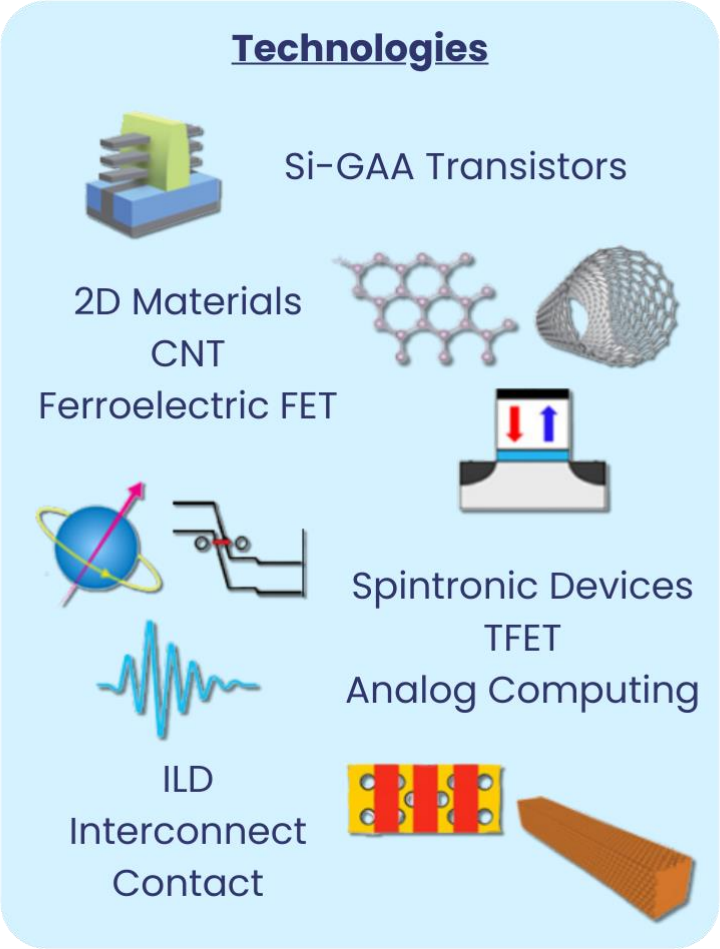
Technologies of the EES2 Roadmap



EES2 ENERGY EFFICIENCY SCALING

Technologies of the compute stack with potential energy saving opportunities

Materials and Devices

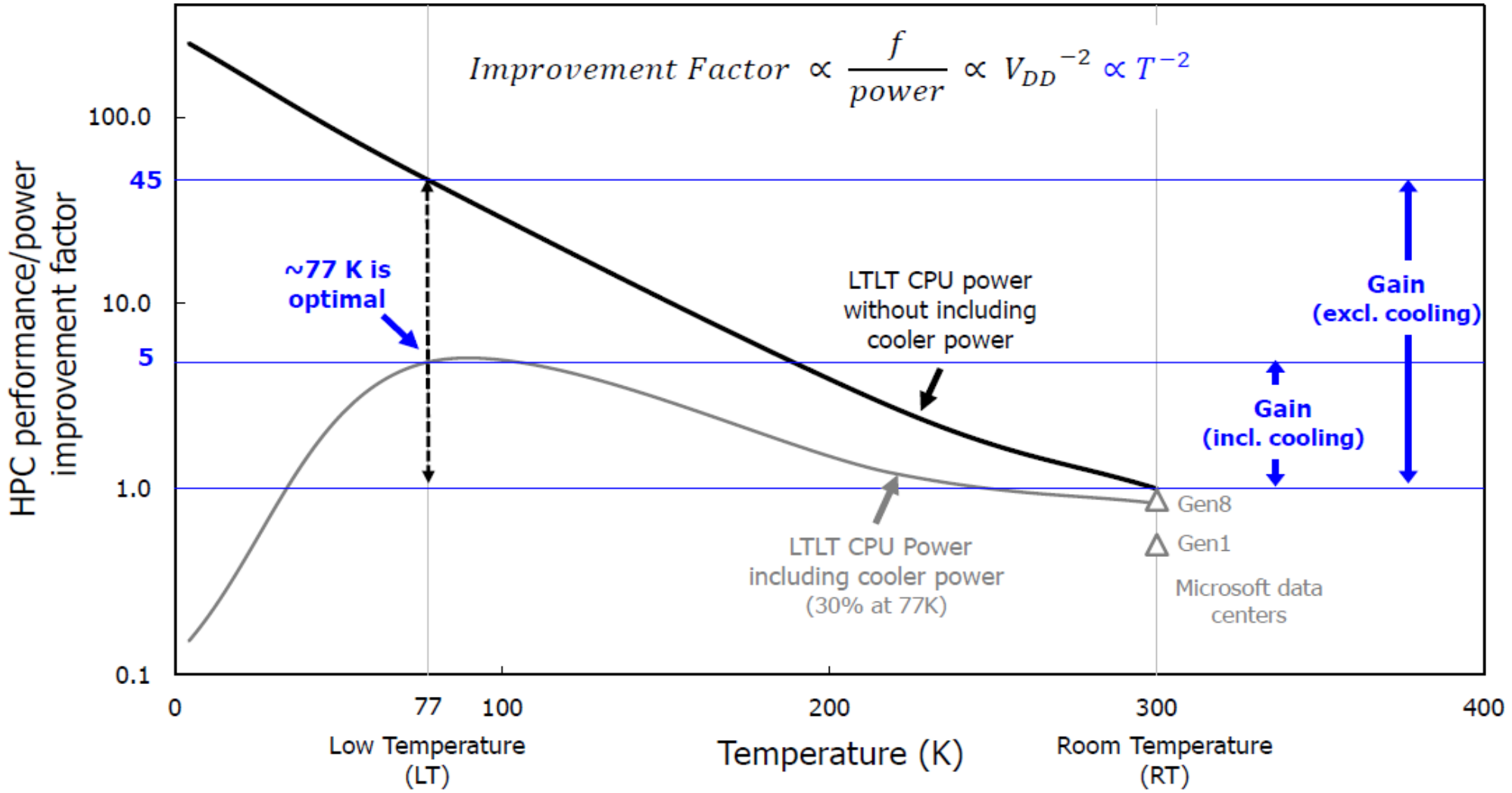


Key Challenges

- Pursue high-quality material manufacturing for energy-efficient devices.
- Develop integration methods for novel material use.
- Analyze fundamental properties for emerging materials and understand their implications on device behavior.
- Bridge the knowledge gap by fostering cross-disciplinary collaboration for device efficiency.
- Utilize advanced modeling for R&D acceleration.

Materials/Devices Example: 25X Efficient Low T Logic

Low Temperature (LT) Operation Offers a Breakthrough



Microelectronic Devices Enabled by Atomic Layer Deposition for Ultra-low Power Architectures

Jeffrey Elam (Lead), Moinuddin Ahmad, Angel Yanguas-Gil, Xingfu Wu, Sandeep Madireddy, and Anil Mane (Argonne National Laboratory – Prime Recipient) + Mark Hersam (Northwestern University), Eric Pop (Stanford University), Elton Graugnard (Boise State University)

FY24$:AMM TO EES2 Lab Call for 10X Energy Efficient Devices

…Team also leads SC Threadwork from Codesign FOA project (BES/ASCR $)

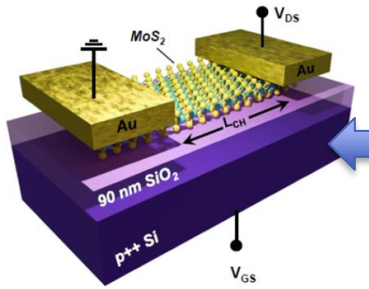
Summary: Revolutionary devices, materials, and manufacturing processes are needed to address the dramatic rise in energy consumption by microelectronics. In this project, we will scale-up the atomic layer deposition (ALD) of two-dimensional (2D) molybdenum disulfide (MoS_2) to accelerate the commercialization 2D-field effect transistors (2D-FETs) and memtransistors for ultra-low energy microelectronics

Impact: 2D-FETs will enable 3-dimensional integrated circuits (3D-ICs) with up to **50x energy savings** and memtransistors will enable (through **co-design**) neuromorphic circuits with up to 10^6x lower energy compared to conventional CMOS.

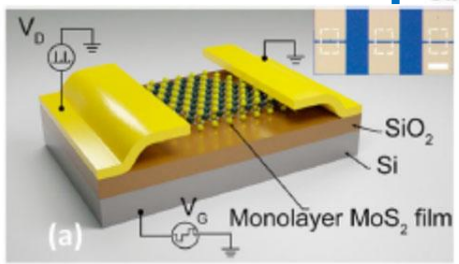
Period of performance: 36 months

Key Device Technology (Both need 2D MoS2 thin films

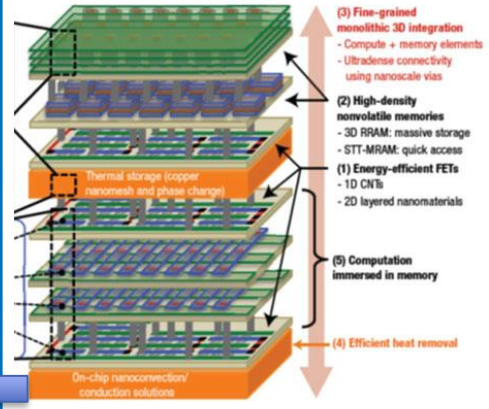
2D Field Effect Transistor (2DFET)



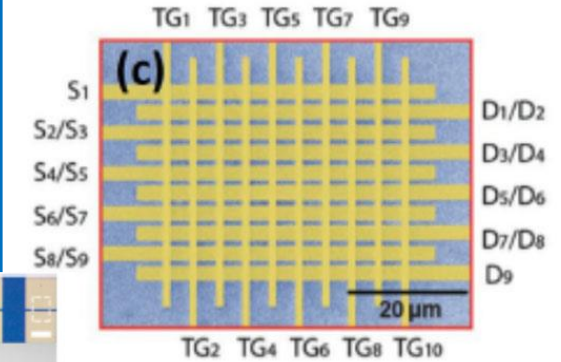
Memtransistor (Memristor)



Ultra-low Power Architectures include:

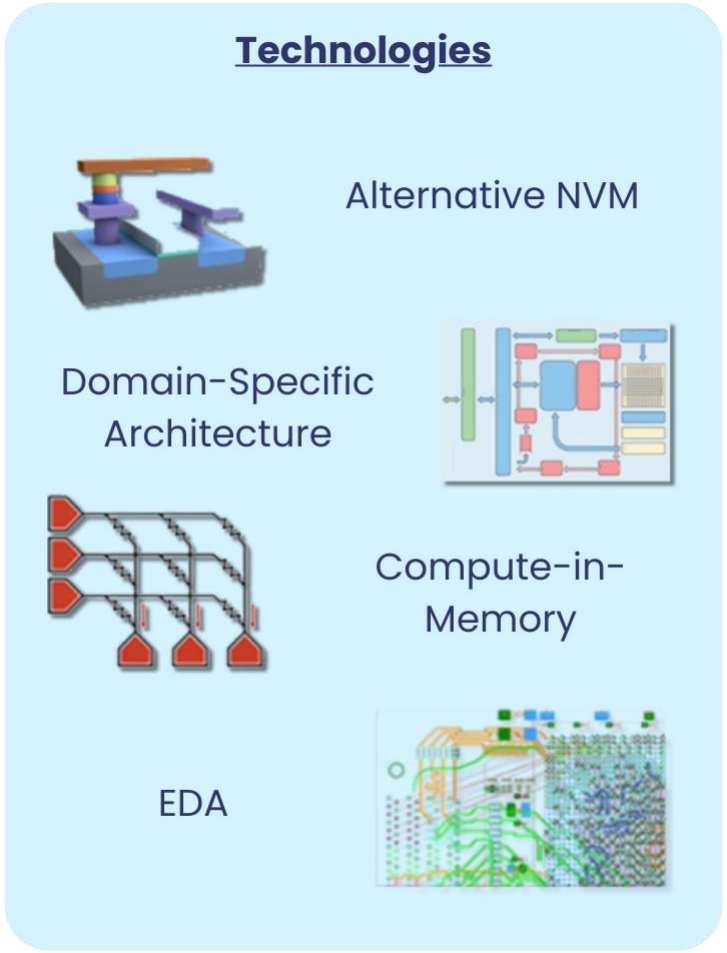


3D integration of memory & logic 50X energy efficiency)



Neuro-morphic circuits (10^6X energy efficiency

Circuits and Architectures



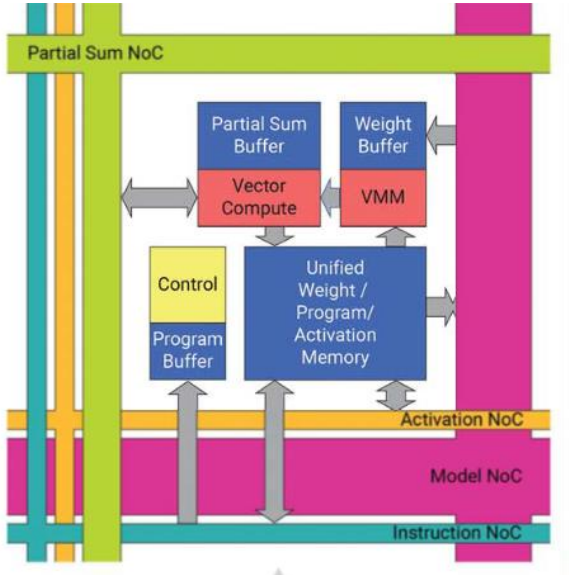
Key Challenges

- Advance compute-in-memory technology through innovative design and co-design EDA tools.
- Demonstrate novel NVM technology alternatives to traditional DRAM/NAND.
- Utilize advanced EDA for the creation of novel device architectures.
- Reduce ownership costs via new interconnect fabrics.

NorthPole Neural Inference Accelerator

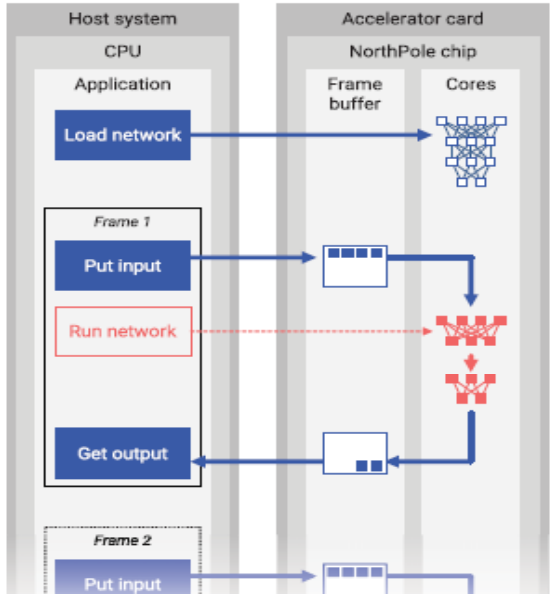
Dharmendra S. Modha/ IBM Research Lab, et al.

EES2 ENERGY EFFICIENCY SCALING



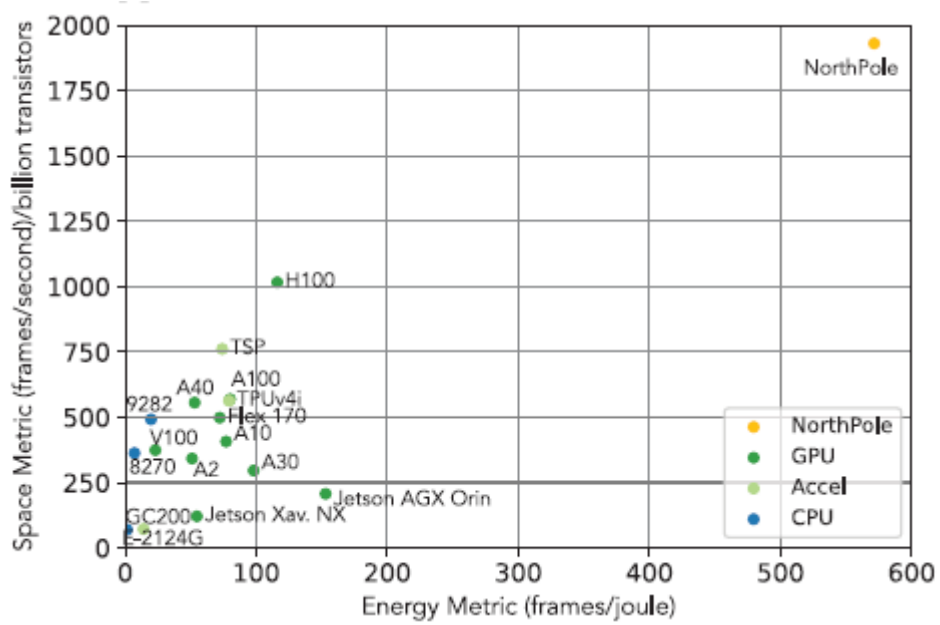
- Chip design optimized for neural inference (no conditional branching)
- Low data precision only (8-, 4-, and 2-bit), configured per layer
- 256-core array, each capable of massively parallel operations (e.g., 8192 2-bit ops/cycle)
- 224MB on-chip memory is intertwined with compute in the layout of each core

- NOCs integrate inter-core computation and memory and enable reconfigured weights and instructions per layer
- Co-designed software orchestrates resources for high utilization
- Simple use model—self-contained network execution with only 3 commands (write input, run network, read results) — facilitates integration with sensors and IT infrastructure



Summary:
The IBM NorthPole chip is a brain-inspired, all-digital inference engine exploiting compute-in-memory specifically optimized for neural network performance.

Energy Impact: Compared to GPU, NorthPole achieves 25x greater energy efficiency (as frames/second per watt) for ResNet50 image classification benchmark.

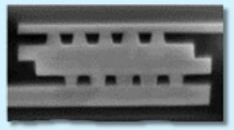


Source: https://www.science.org/doi/full/10.1126/science.adh1174

Advanced Packaging and Heterogeneous Integration

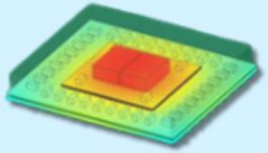
EES2 ENERGY EFFICIENCY SCALING

Technologies



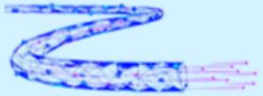
Interconnects for 2.5/3D Stacking

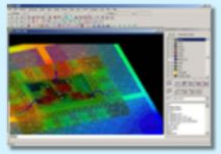
Thermal Management



Advanced System Cooling

Alternative Interconnect





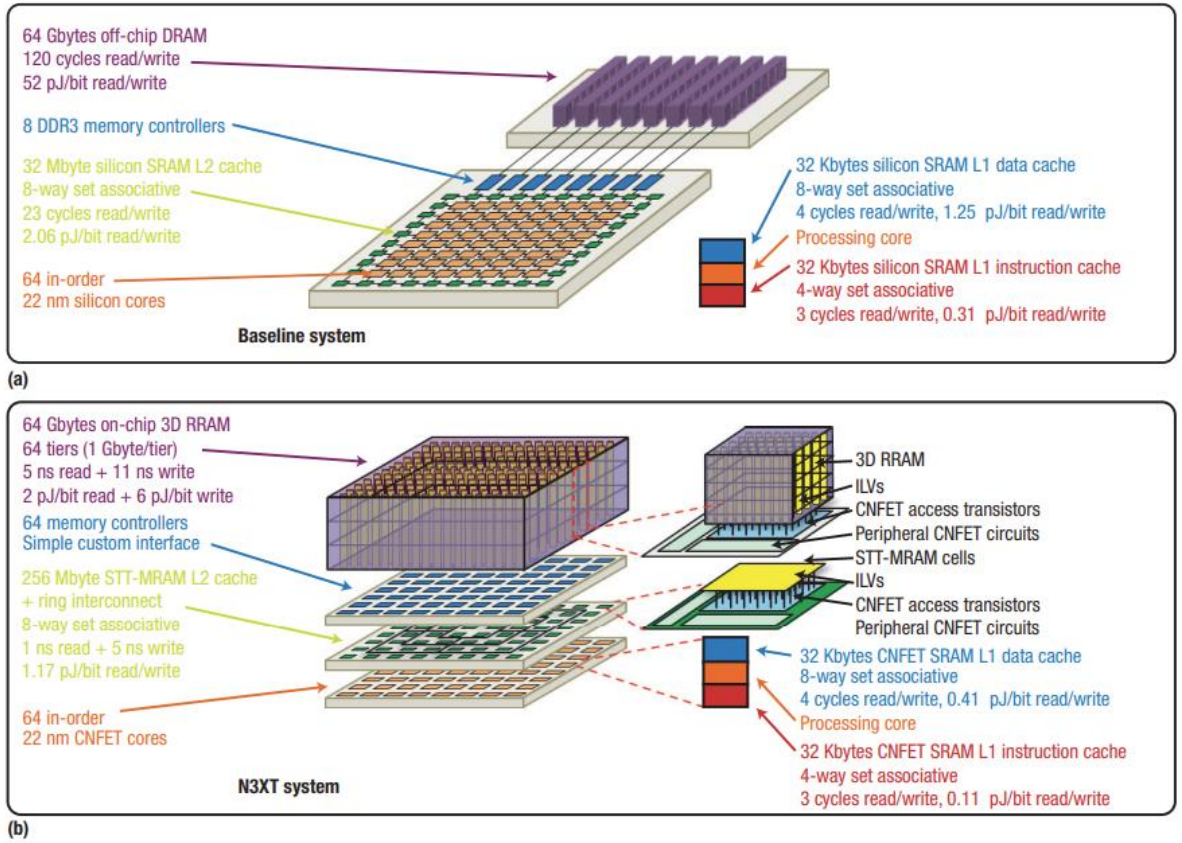
Packaging EDA

Key Challenges

- Standardize UCIe.
- Develop non-CMOS compatible devices towards monolithic integration.
- Explore novel materials/designs in thermal interfaces.
- Optimize memory access costs and heat management with advanced interconnect technologies.

Nano-Engineered Computing System Technology (N3XT)

Mohamed M. Sabry Aly/ Stanford University, et al.



a) Baseline Conventional System
b) Monolithically integrated 3D system enabled by N3XT

Summary:
The project represents a revolutionary approach to electronics that strives to break away from traditional two-dimensional semiconductor architecture. Instead, it focuses on building integrated systems using a three-dimensional design, where components such as processors, memory, and sensors are stacked atop one another. This 3D integration is achieved using advanced technologies such as carbon nanotubes and nanoscale interconnects. The project aims to create high-density, energy-efficient chips that overcome the scaling challenges of existing silicon-based technologies, potentially leading to faster and more capable computational devices.

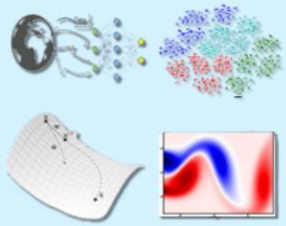
Energy Impact:
For the specific energy savings, the project reports achieving up to **1000x reduction** in required energy per function compared to conventional two-dimensional electronics.

Source: https://poplab.stanford.edu/pdfs/Aly-N3XT1000-computer15.pdf

Algorithms and Software

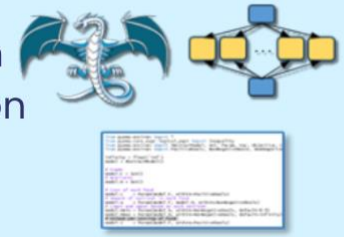
EES2 ENERGY EFFICIENCY SCALING

Technologies



Continual Learning
Sparse Networks
Meta-Learning PINNS

Compiled Python
Auto-Parallelization
AI-Driven Code
Generation



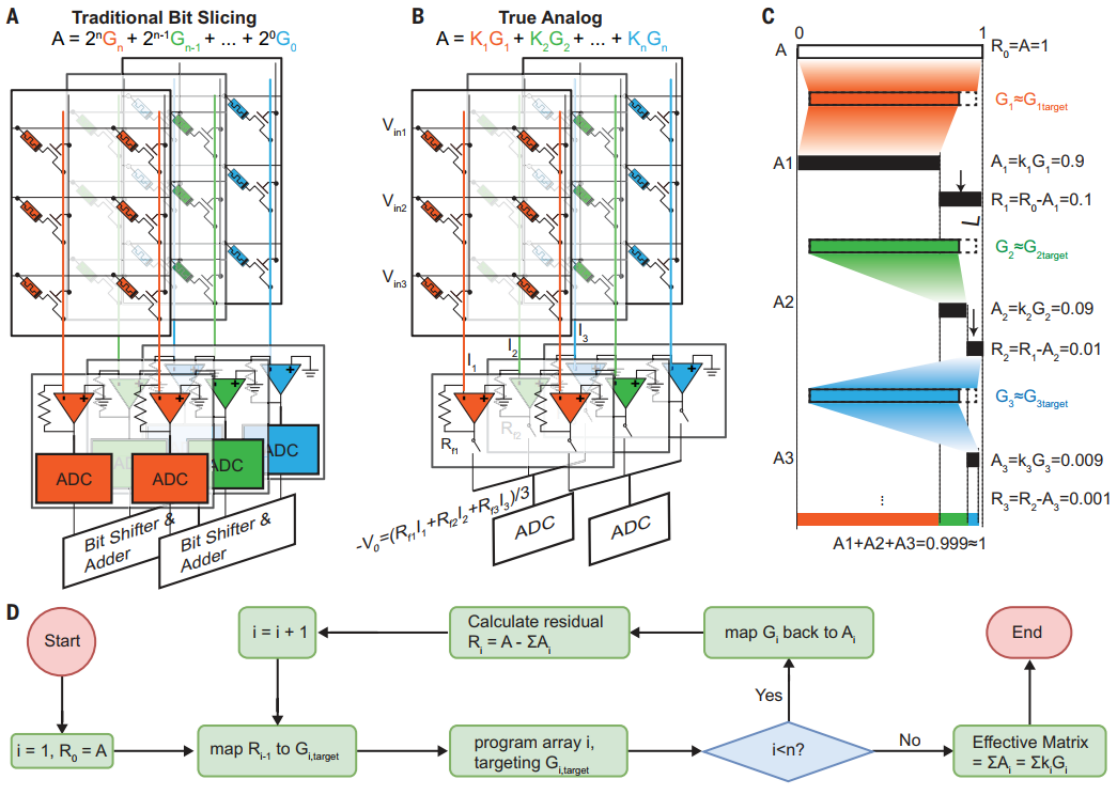
Domain-Specific
Languages and Tools
Neuro-Inspired
Computing

Key Challenges

- Enhance energy-saving algorithms through efficient utilization of memory and parallelism.
- Improve profiling tools for software energy tracking.
- Align new energy-efficient hardware for use within existing codebase and infrastructure.
- Reduce ML energy consumption through new strategies in ML training and inference.
- Improve fundamental understanding of learning and intelligence for machine intelligence applications.

Programming memristor arrays with arbitrarily high precision for analog computing

Wenhao Song/ TetraMem, et al.



a) Traditional crossbar arrays with ADCs and additional postprocessing circuits
b) Proposed arbitrary precision programming circuit with shared ADCs
c) Example of programming a numerical value A=1 into multiple memristor devices step by step
d) Flowchart of the arbitrarily high-precision programming algorithms

Summary:

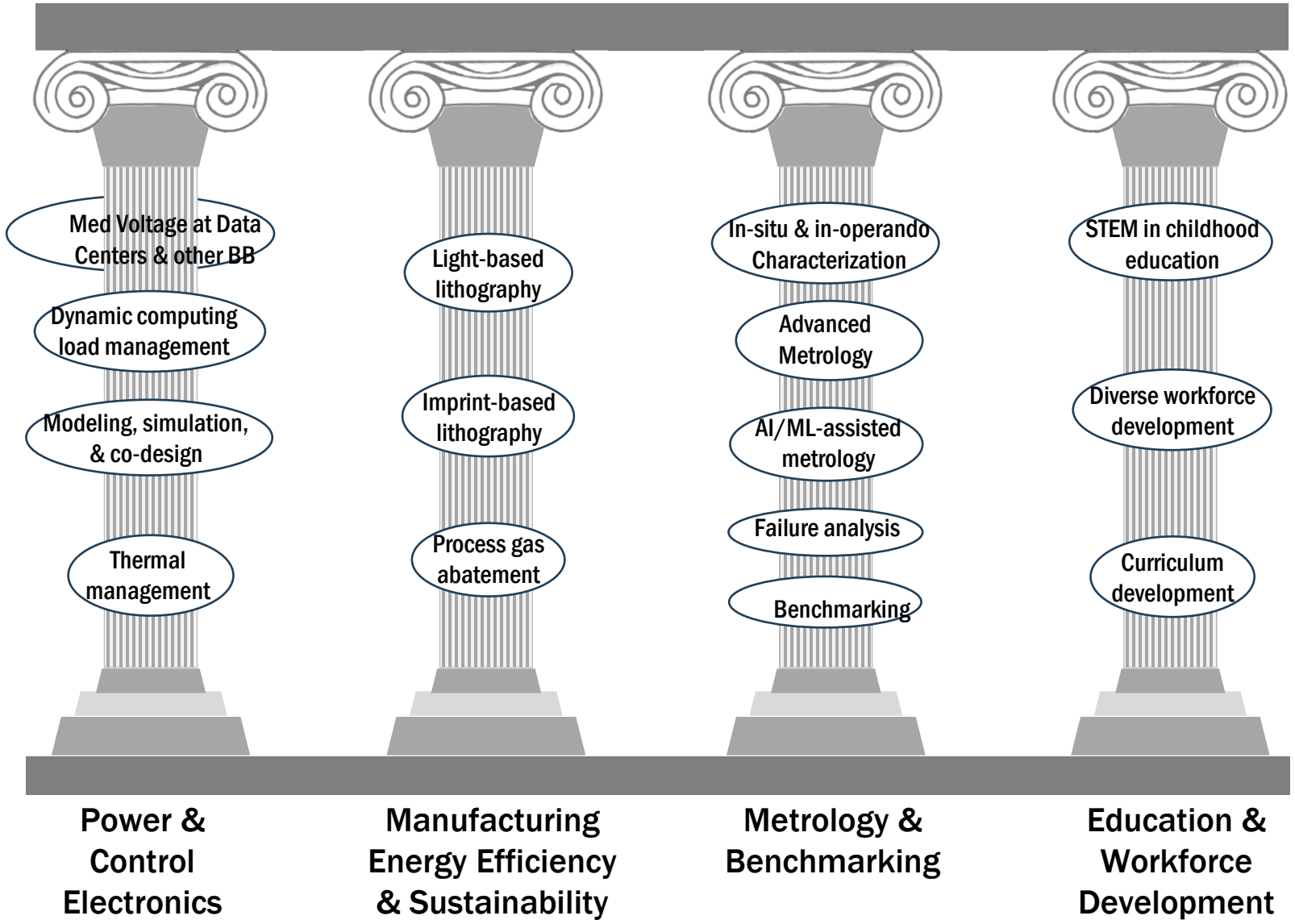
The study presents an innovative protocol for programming memristor arrays to perform high precision analog matrix multiplication. It utilizes a collective device approach for numerical representation and dynamic error adjustment to ensure precise calculations. This method showcases significant power efficiency benefits compared to conventional digital computing techniques.

Energy Impact:

By dynamically compensating for device imperfections, the algorithm ensures accurate computational results with less energy than traditional digital systems, highlighting a strategic move towards algorithm-driven energy savings in computing.

Source:
https://www.science.org/doi/10.1126/science.adi9405

Enablers: Preliminary Results



Enabling technologies and approaches to support the compute stack.

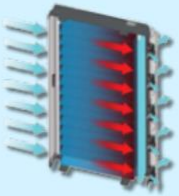
Power and Control Electronics

EES2 ENERGY EFFICIENCY SCALING

Technologies

Dynamic Computing Load Management

Advanced Thermal Management



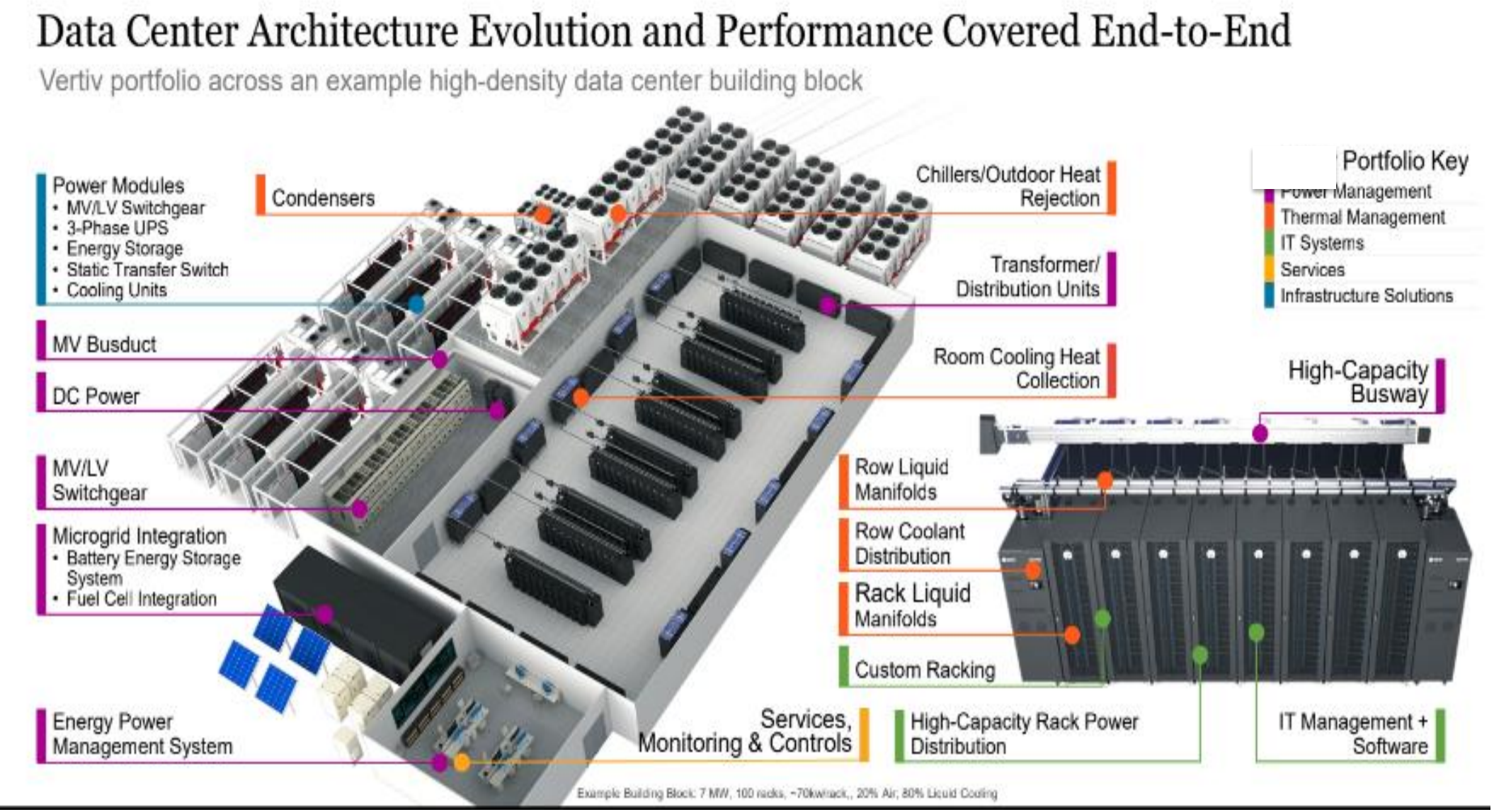
Enhancement of Modeling, Simulation, and Co-Design Capabilities

Key Challenges

- Optimize power delivery for new med/high power demands in data centers and new super factories with wide-bandgap power electronics..
- Advance co-design to integrate power delivery with circuits and architectures.
- Innovate in on-chip thermal management
- Extend energy efficiency to legacy computing and non-datacenter contexts.

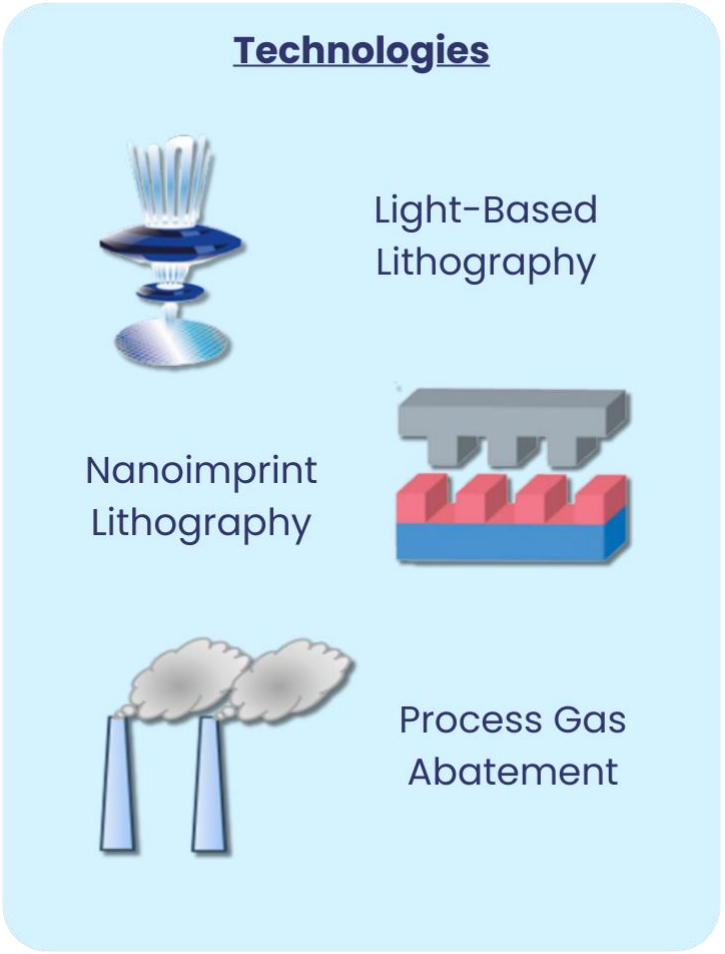
Power and Control Electronics

EES2 ENERGY EFFICIENCY SCALING



Save Energy by Connecting at Med Voltage w/ SiC and GaN

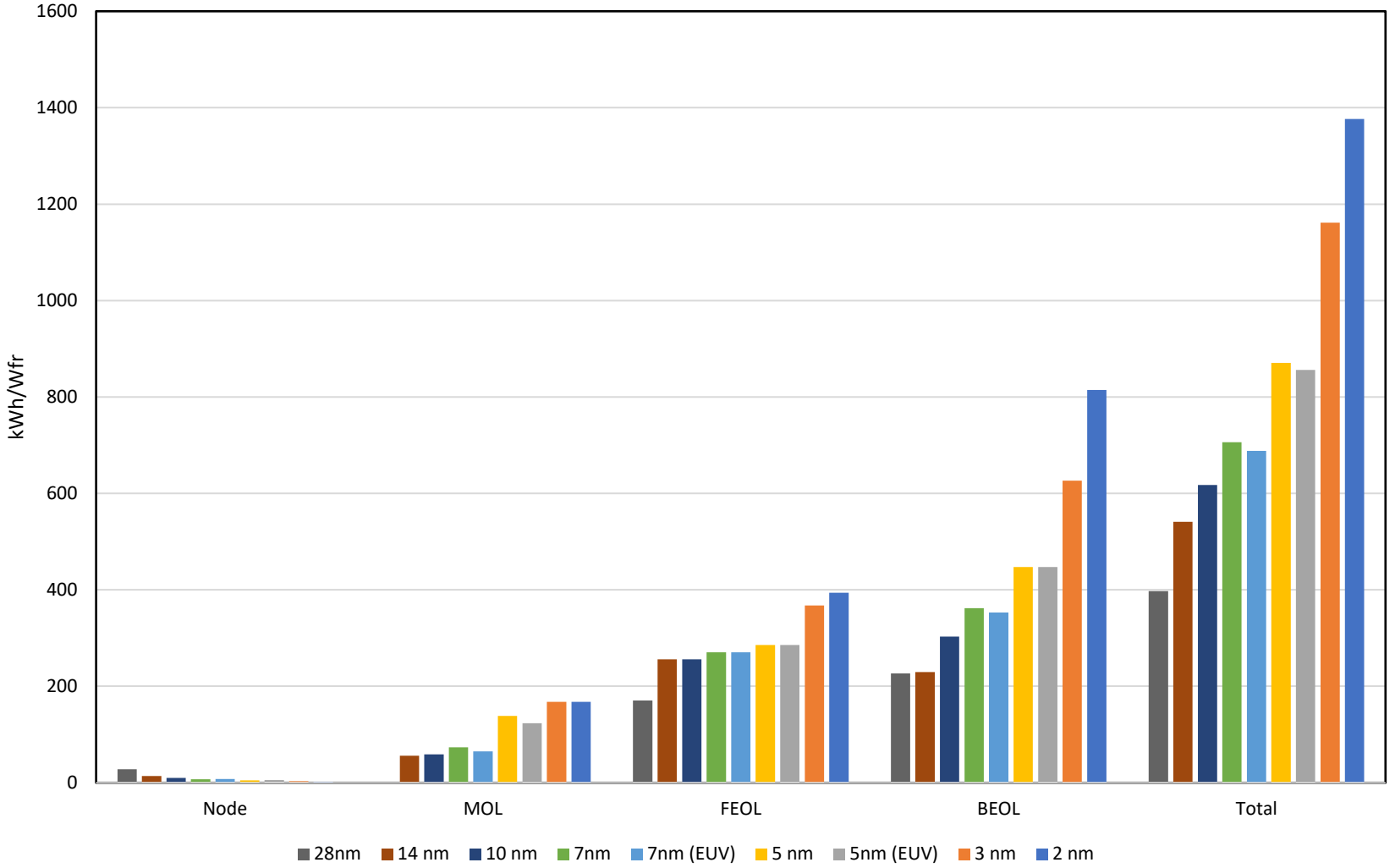
Manufacturing Energy Efficiency and Sustainability



Key Challenges

- Enhance EUV lithography energy efficiency.
- Refine nanoimprint lithography.
- Design compact, advanced abatement systems.
- Select lower-GWP gases for processing.
- Set comprehensive energy metrics for lithographic processes
- Create effective recycling methods for process gases

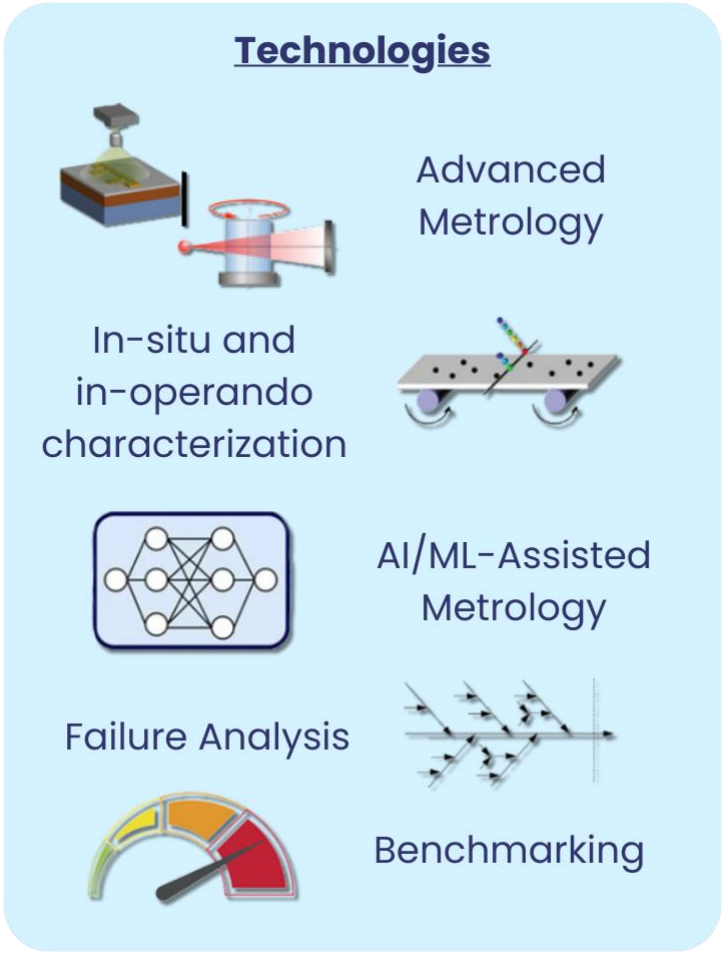
Manufacturing Energy Costs per Wafer



Substantial energy increase per wafer with implementation of EUV lithography

MOL = middle of line; FEOL = front end of line; BEOL = back end of line; EUV = extreme ultraviolet. (Bardon and Parvais 2023)

Metrology and Benchmark



Key Challenges

- Address the disparity between expected and actual chip performance.
- Advance 3D metrology technology.
- Enhance thermal measurement methods for complex chip architectures.
- Integrate advanced characterization methods into manufacturing lines.

Methodology for Compute Energy Assessment

Major Industry Segments

Major Use Cases

Select and Define representative benchmarks

Bottom-up Energy Use Evaluation (Direct Energy Costs)
- Circuits
- Packaging
- Devices

- Laboratory measurements
- Simulations

Benchmark

TOOLING
- Profiling
- Simulation

Total Energy Performance

Education and Workforce Development

Technologies

Diversity

Early Childhood Education

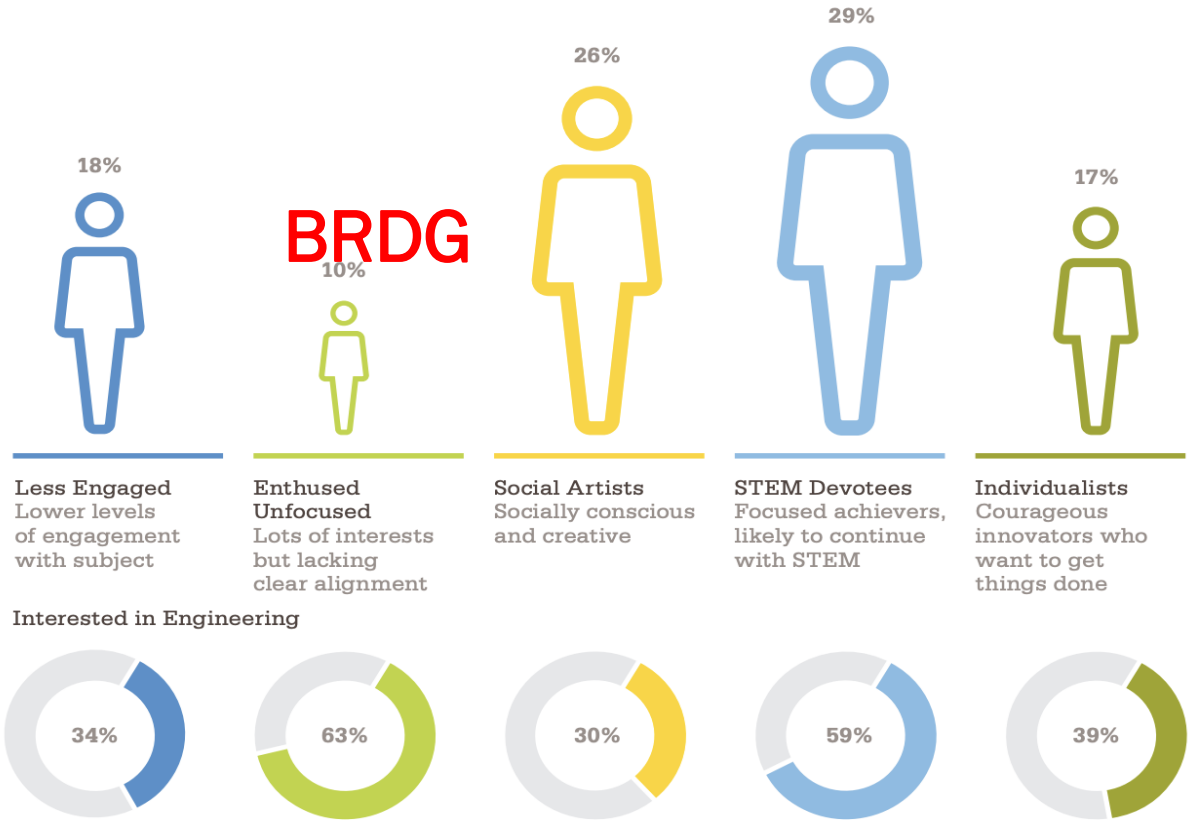
Curriculum Development

Key Challenges

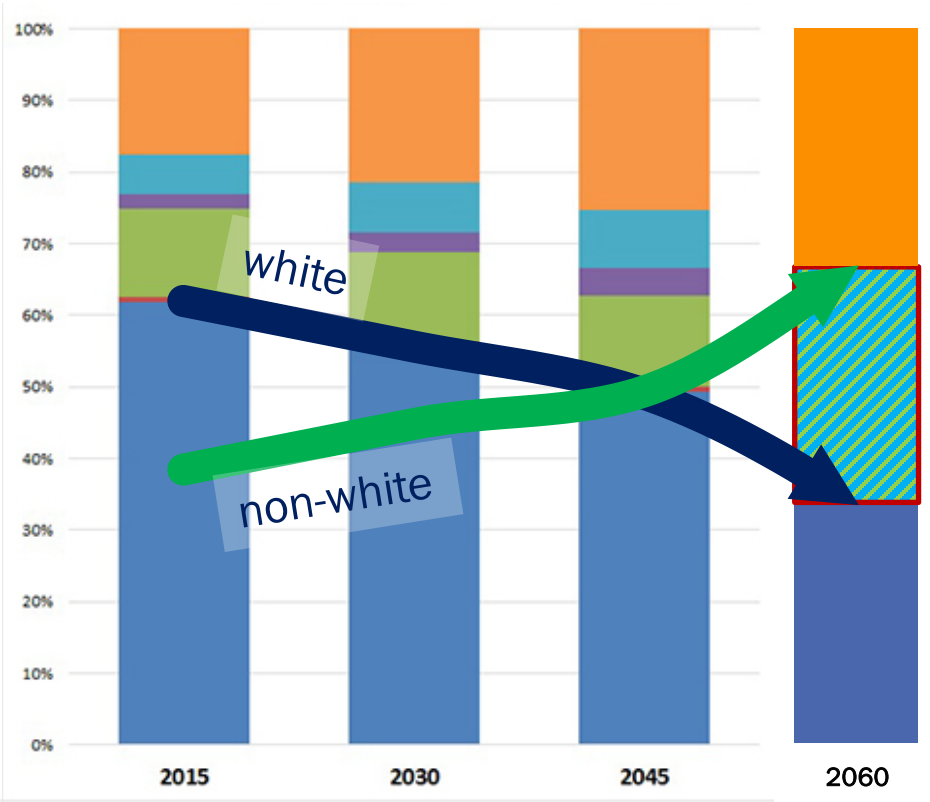
- Attracting a workforce attuned to environmental and social impacts may simultaneously increase workforce diversity.
- Increase diversity with emphasis on first in family
- Advance software driven co-design in curricula.
- Merge hands-on experience with theoretical learning.
- Support ongoing learning and professional growth.
- Foster collaboration between academia, industry, and government to align educational programs.

Education and Workforce Development

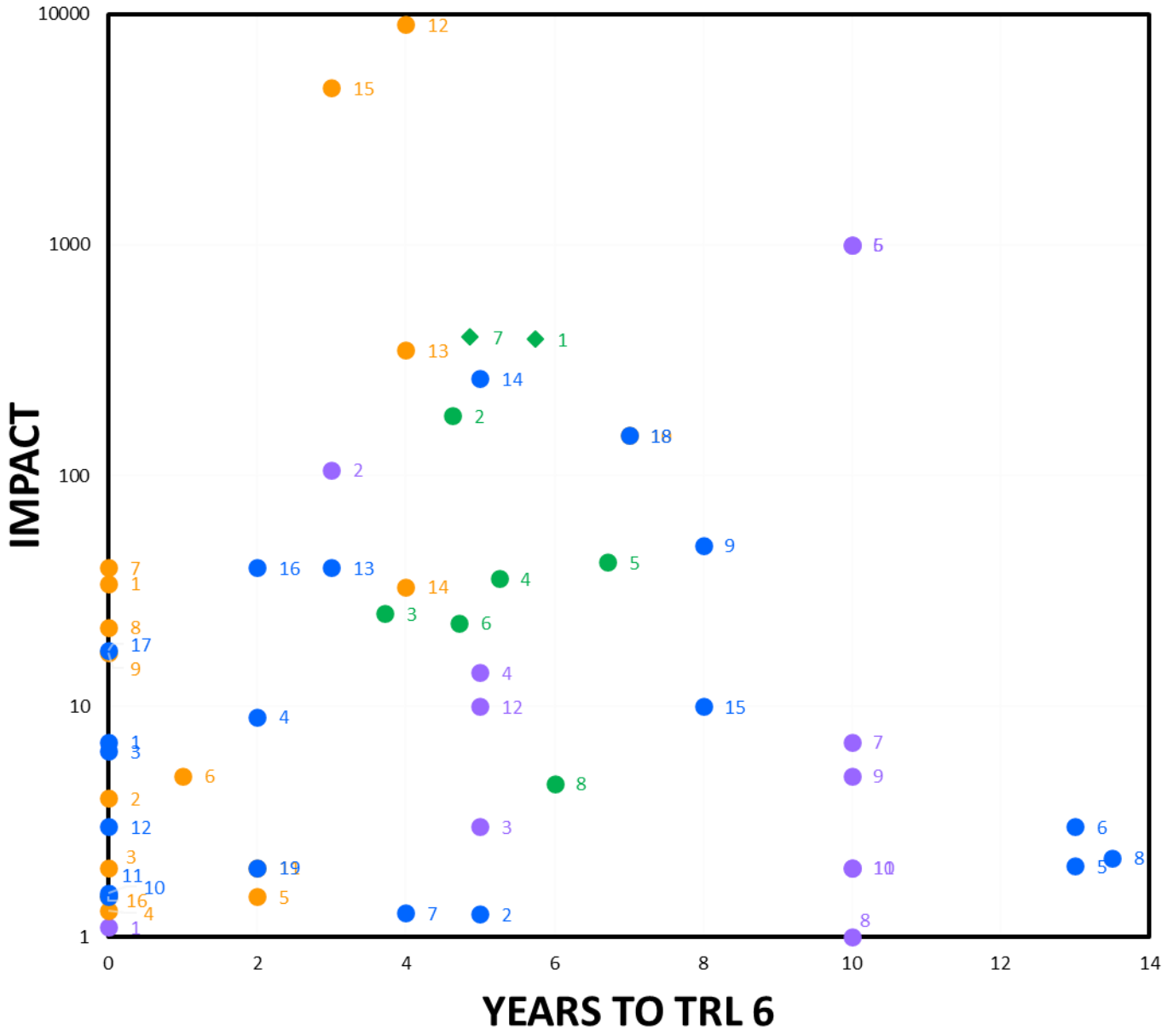
- Tim Wei estimated that US “traditional” STEM people supply will decline by as much as 50% by 2045 (e.g. need to at least double supply)
- Analysis of “Tribes” of potential Students



https://www.imeche.org/policy-and-press/reports/detail/five-tribes-personalising-engineering-education



Results by WG: Estimated Energy Saving Potential



LEGEND:

Materials and Devices

1 Si-GAA
2 CNT Memory
3 CNTFET (Logic)
4 TFET
5 Spintronic memory
6 FeFET (Flash)
7 Analog devices for neuromorphic computing
8 FeFET (SRAM)
9 Contact & interconnect
10 Novel ILD
11 Spintronic logic
12 2D materials

Circuits and Architectures

1 ReRAM vs NAND
2 STTRAM vs NAND
3 NRAM vs DRAM
4 ReRAM vs DRAM
5 CNT NVM
6 Metis SRAM
7 Molecular dynamics ASIC
8 FPGAs for machine vision
9 SRAM stacked 3D DNN accelerator
10 MIV stacked ReRAM
11 HBM Cache
12 Neuromorphic memcapacitive devices
13 Neuromorphic memristor matrix multiplier
14 Neuromorphic asynchronous computing
15 CMOS SRAM CIM
16 CXL optimized DDR5

Advanced Packaging & Heterogeneous Integration

1 LMP solder with polymer
2 Nanostructured thermal interface surface
3 CNT TIM
4 Graphene TIM
5 Graphene interconnects
6 CNT interconnects
7 Rh/Ir interconnect
8 CNT for 3D ICs
9 3D IC MIVs
10 Feveros
11 TSV for 3D IC
12 Hybrid bonding (Cu-Cu)
13 Optical off-chip interconnect
14 Optical on-chip interconnect
15 Optical bus
16 UCIe chiplet standard
17 3D stacked SRAM
18 MIV stacked ReRAM
19 HBM on logic

Algorithms and Software

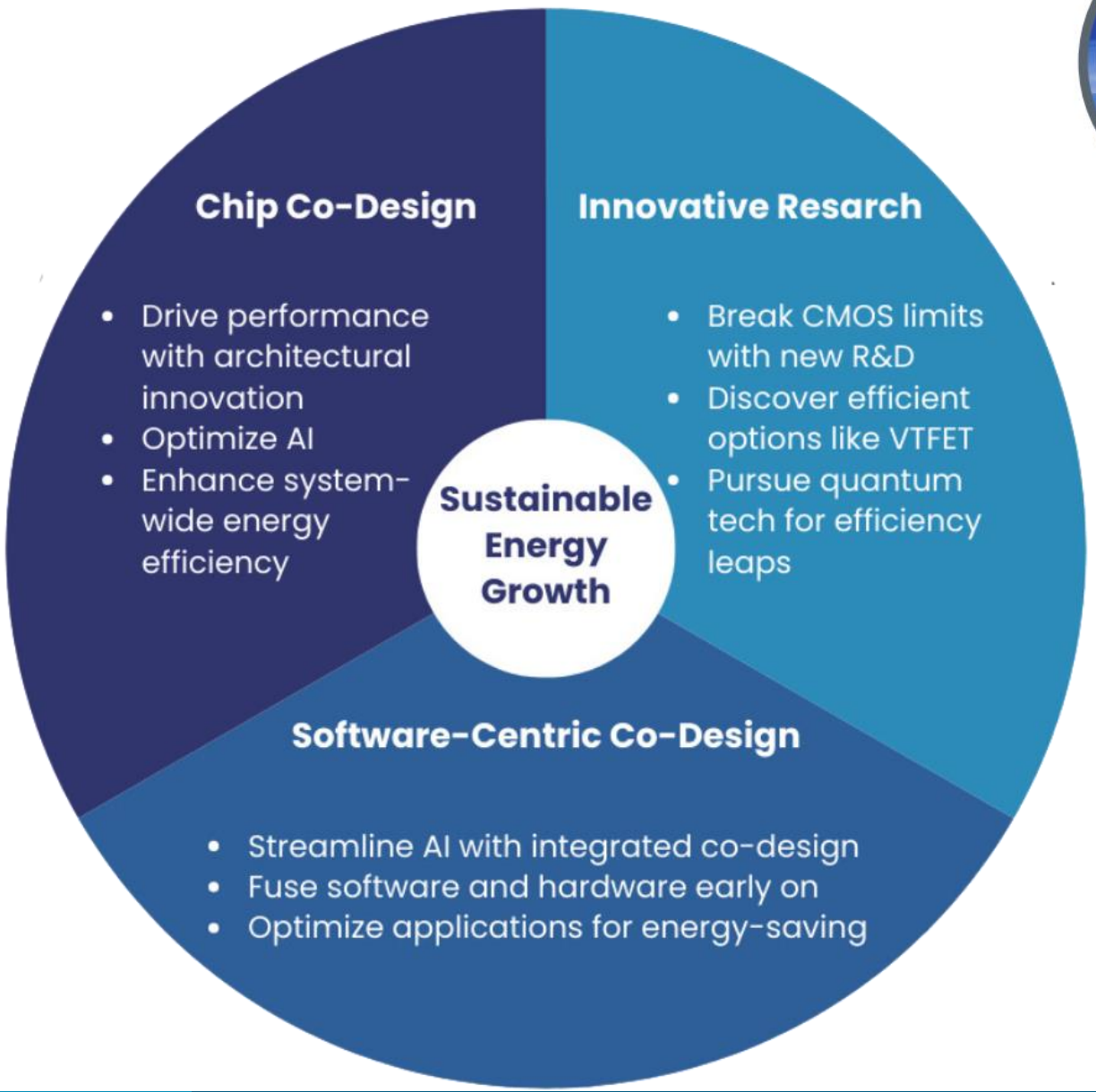
1 Reduced energy for ML algorithms
2 Algorithm-specific energy (tooling)
3 Algorithm-specific energy (benchmarking)
4 Languages, compilers, and runtime systems
5 Communication protocols
6 Privacy and security
7 Software for emerging architectures
8 Computational reliability

Three Solutions to Unsustainable Energy Growth

EES2 ENERGY EFFICIENCY SCALING

Roadmap Recommendations Divided into Three Categories:

1) Instructions (hardware and software codesign),
2) Bits (hardware), and
3) Applications (software-driven full-stack co-design).

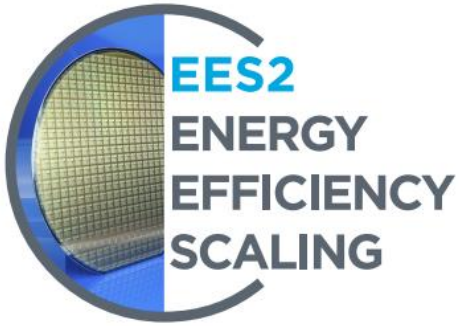


Conclusion: We can DO This!!—but need your help!

-By 2030 (when we plan to reach 10X with EES2 RD&D) we may need to accelerate to 20-40X! Deploy, Deploy Deploy!!

-By 2037 (when we plan to reach 100X with EES2) we must have a good sense of how we can get to 1000X (quantum? Nature-inspired? Distributed? So we need to engage basic researcher, inventors, international!!

Key Strategies for Deploying EES2



Mechanisms from Energetics' working group process:

1. Public-Private Partnerships –especially R&D!!
2. We Can Do IT—EES2 Roadmap Outreach
3. Comprehensive Benchmarking across all layers of computing
4. Software Driven Co-design Interdisciplinary Collaboration and Curriculum Creation

The Future

- **Enhanced Microelectronic Efficiency:** EES2's active investment in research for energy-efficient microelectronics with Labcall awards.
- **AI Energy Efficiency Synergy:** DOE's ASCR FOA aligns with EES2, focusing on energy-efficient AI with neuromorphic and statistical techniques.
- **Expanding Technology Horizon:** Inclusion of quantum and optical computing in roadmap's future vision.
- **Community-Driven Progress:** EES2's RFI seeks public opinion to further refine the roadmap.

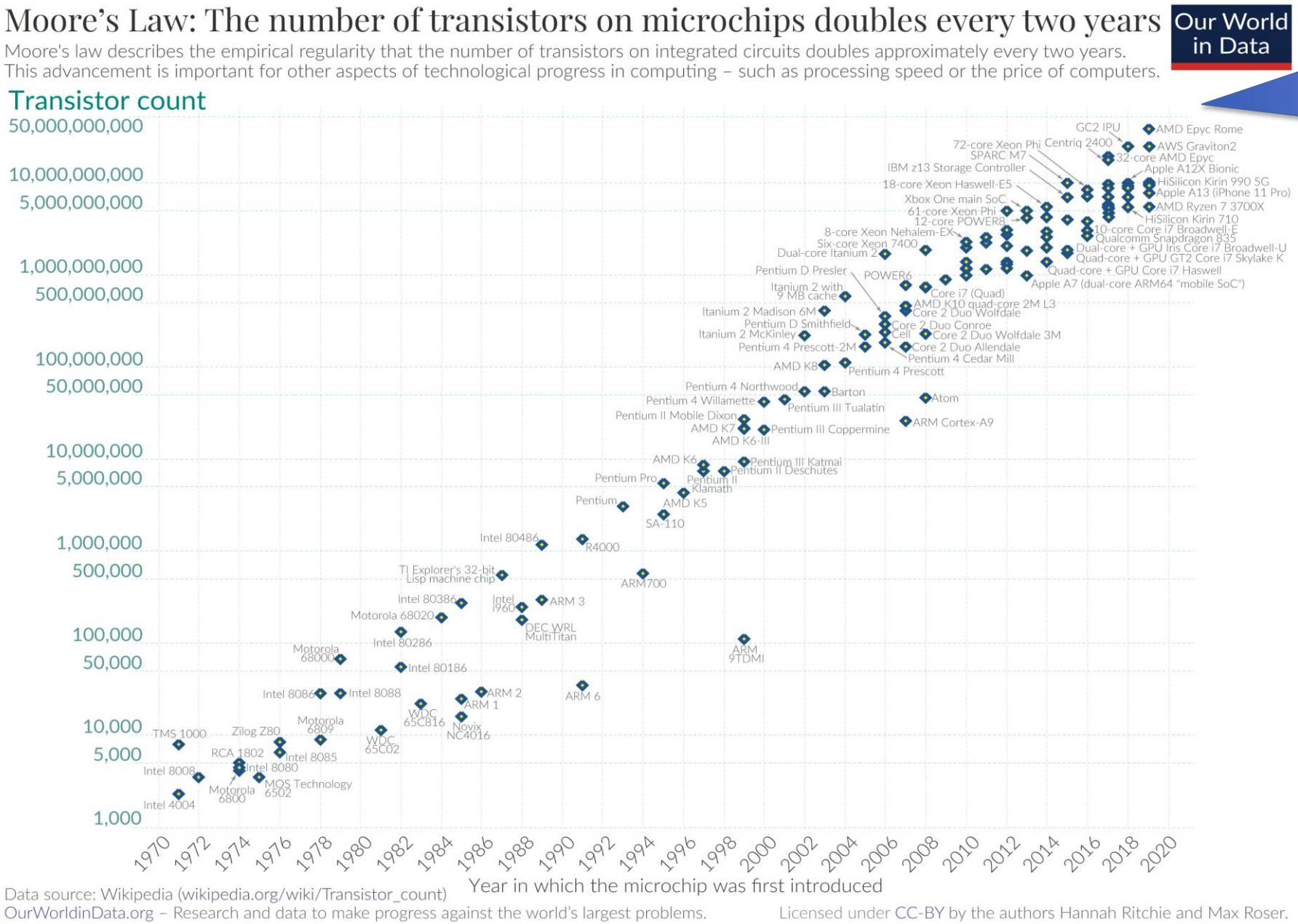
Source: Generated with OpenAI's DALL-E

Q&A

Extra After this

Backup

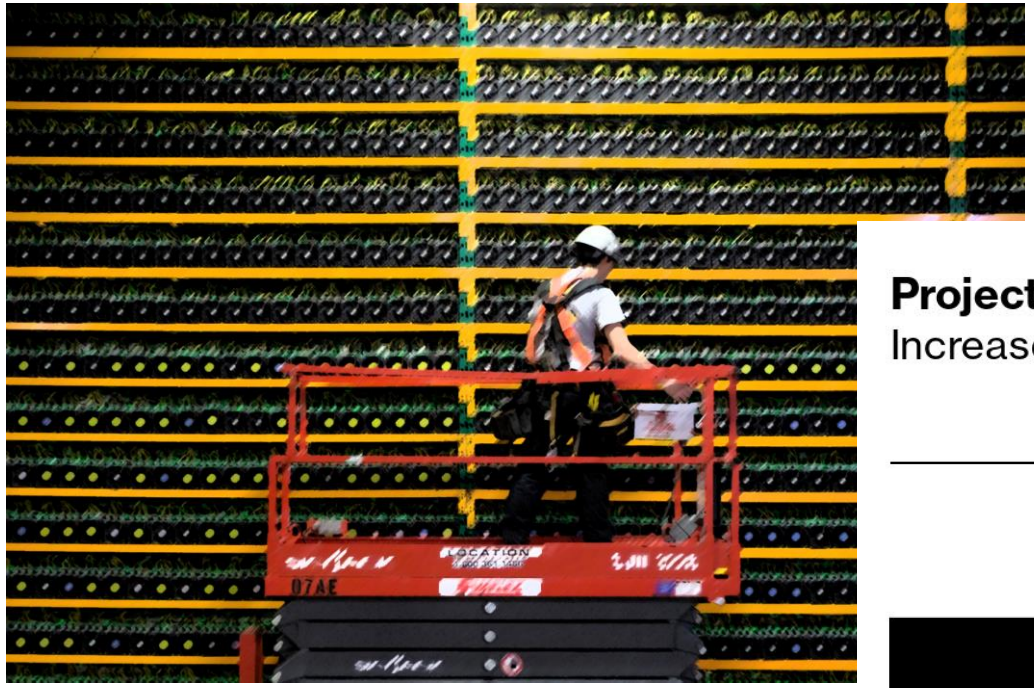
Moore's Law: Biennial doubling requires semilog



2023 most efficient chip by AMD still on track.

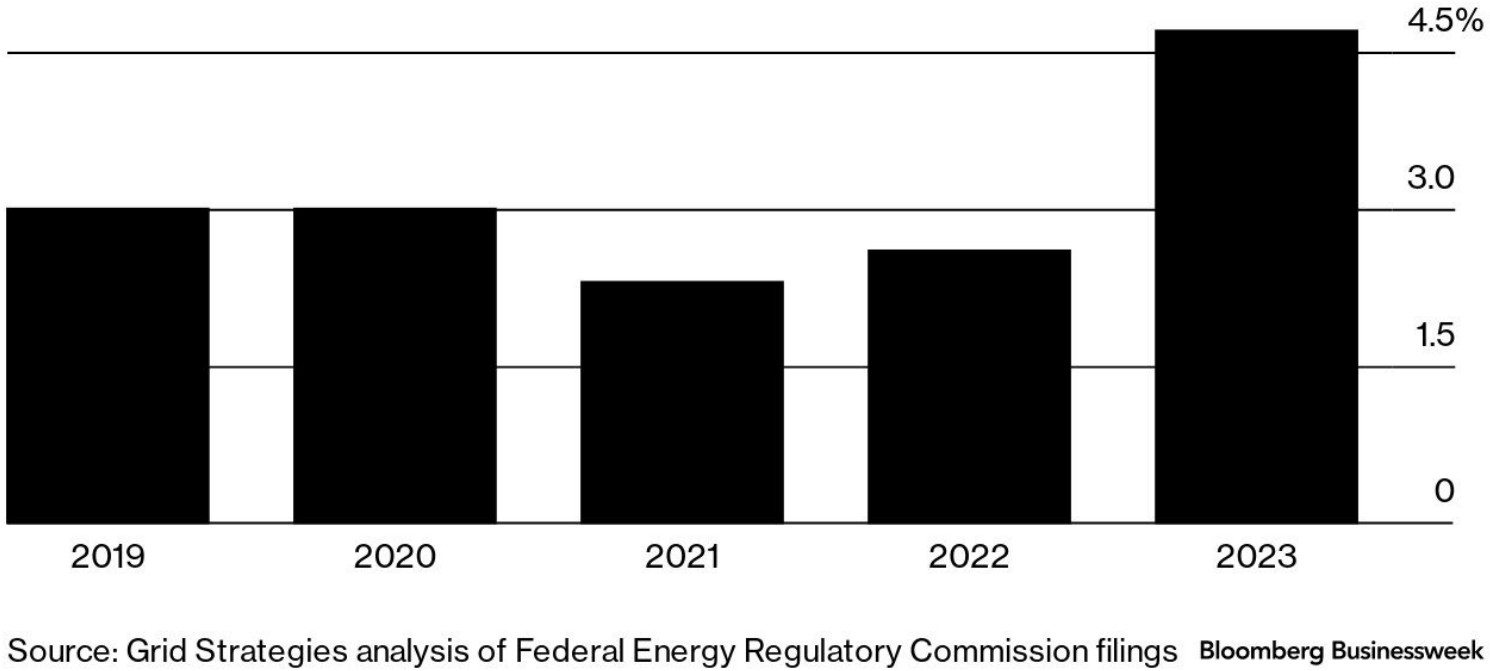
EES2 ENERGY EFFICIENCY SCALING

New Urgency II: AI- and Bitcoin- Driven Electricity-use in Data Center



Bitcoin mining rack. *Time,* Feb, 2024https://apple.news/AA5nb3bRNTIWvcp1yBl5v5Q

Highest electricity use ever for Data Centers drives major increase (+4.5%) from prior stable electricity forecasts.



January 2024 *Bloomberg News*: “Hungry Hungry AI”

Preliminary* RD&D Recommendations for EES2

(~MAPT table)

***does not include SLAC analysis & Plenary input including Highlights**

EES2 ENERGY EFFICIENCY SCALING

EES2 Working Group	Near Term (0-5yrs)	Mid Term (5-10 yrs.)	Long Term (10+ yrs.)
Materials and Devices	Build upon CMOS technology-compatible transistors, interconnects, memory, with innovations in materials/processes.	Beyond CMOS including devices reliant on alternative carriers or carrier transport, and device structures.	Exploratory materials and revolutionary devices.
Advanced Packaging & Heterogeneous Integration (HI)	Thermal engineering, reliability, and interface materials.	New methods for integrating devices, interconnects, and memories.	3D HI at the package level for different architectures and hardware across multiple technology nodes. **help SC
Circuits and Architectures	Domain-specific architectures. AI-driven EDA, PDK design for efficiency	Co-optimization of hardware system components for distinct applications, with energy efficiency as a design variable.	Architectures to enable new QC & NI computation for distinct applications by moving memory to or in compute, and through advanced interconnect, with energy as a design variable.
Software and Algorithms	Energy-aware algorithms and software.	Improved ML algorithms and exploitation of ML in software development	Software and algorithms co-optimization with advanced architectures. **help SC

Roadmap 1 Overview: Energy Efficiency-based Solutions for HW

Innovation Metric Group	Headroom in Energy compared to minimum at thermodynamic limit at 300K (**Orders of Magnitude**)	Potential for **10X** efficiency in < 10 years	Potential for **100X** efficiency in < 15 years	Technical Potential for **1000X** efficiency (alone) in < 20 years	Technical Barriers to **1000x**	Cultural, Educational Barriers to **1000X** alone	How Fast is Energy Use currently changing	How Fast could Energy Efficiency improve **compared to EES2**
Bits (Transistors or Switching)	1 to 10^{2}**	Low-Medium	Low	High	High	High	Decreasing	In-line
Instructions (INT4 to HPCG)	2 to 9*, **	Medium-High	High	High+	Medium-High	Medium	Increasing	Faster

*Shankar, S. and Reuther, A., 2022, September. IEEE High Performance Extreme Computing Conference (HPEC) (pp. 1-8). IEEE.

**Shankar, S., 2023, September. IEEE High Performance Extreme Computing Conference (HPEC) (pp. 1-6). IEEE.

Energy Efficiency-based Solutions for Software

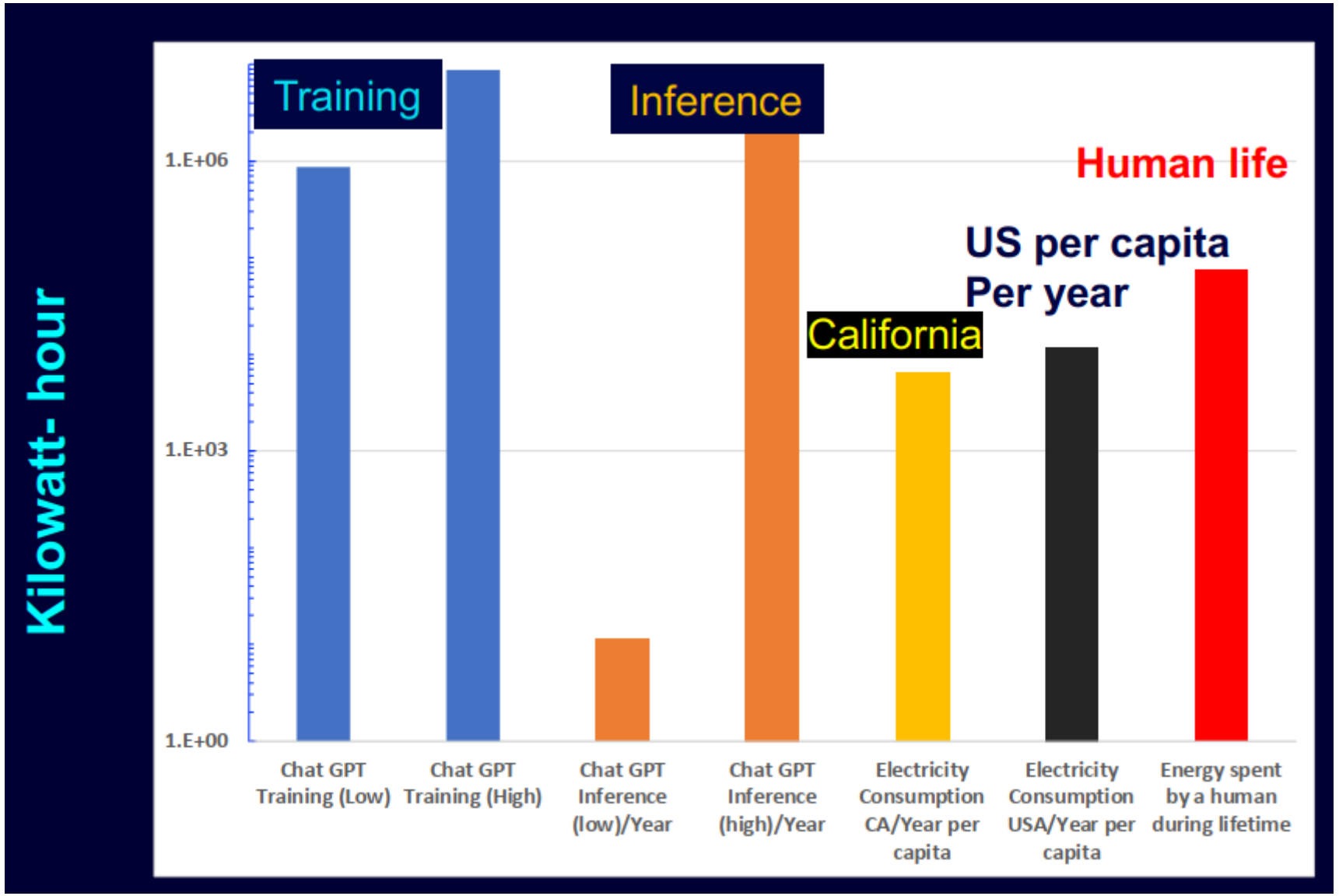
Innovation Metric Group	Headroom in Energy compared to minimum at thermodynamic limit at 300K (**Orders of Magnitude**)	Potential for **10X** efficiency in < 10 years	Potential for **100X** efficiency in < 15 years	Technical Potential for **1000X** efficiency (alone) in < 20 years	Technical Barriers to **1000x**	Cultural, Educational Barriers to **1000X** alone	How Fast is Energy Use currently changing	How Fast could Energy Efficiency improve **compared to EES2**
Applications (for complete Simulations; E.g. Scientific Simulation, AI Training)	20 to 36*,**	High	High+	High++	Low-Medium	Low	Rapidly Increasing	Faster+

Need for energy efficiency for SW

- Applications-specific algorithms need focus in addition to hardware and transistors
- As computer systems are hierarchically based on building blocks transistors, devices, and components overall energy use is determined by the efficiency of the components

**Shankar, S et al., 2022 and 2023*

AI/ML Model Complexity



EES2 ENERGY EFFICIENCY SCALING

Training and inference energy rival annual per capita consumption

Source: Shankar 2023.

Key URLs and Videos of 9-20-22 Inaugural Pledge Signing

Department of Energy Announces Pledges from 21 Organizations to Increase the Energy Efficiency of Semiconductors and Bolster American Manufacturing: https://www.energy.gov/eere/articles/department-energy-announces-pledges-21-organizations-increase-energy-efficiency

Undersecretary For Science and Innovation Geri Richmond is first to sign video Here→ https://vimeo.com/769659745/f054bc19d5

See also **EERE Commitment to support EES2** from Deputy Assistant Secretary for Energy Efficiency Carolyn Snyder → https://vimeo.com/769659330/2673be6fbd

Energy Efficiency Scaling for 2 Decades (EES2)

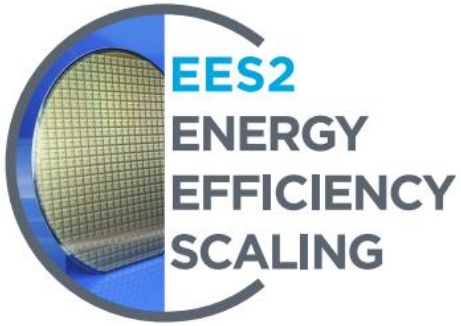
ANNOUNCED by AMO in JANUARY 12 2022:

Shift from R&D Roadmaps based on biennial length-based scaling (e.g., Moore's law) to ultra-energy-efficiency scaling and ensure all R&D includes some energy-efficiency focus

- Specifically, develop in partnership with U.S. and Allied Country Semiconductor Industry an RDD&D roadmap to ensure
  - Doubling* of microelectronics' energy efficiency every two years *or faster* for the coming decades
  - In two decades, increase energy efficiency of next generation microelectronics by >1000X

https://ees2.SLAC.stanford.edu

EES2 Roadmap V1.0 Working Groups



Computing 'STACK'

- Algorithms / Software
- Heterogeneous Integration and Advanced Packaging
- Circuits and Architectures
- Materials and Devices

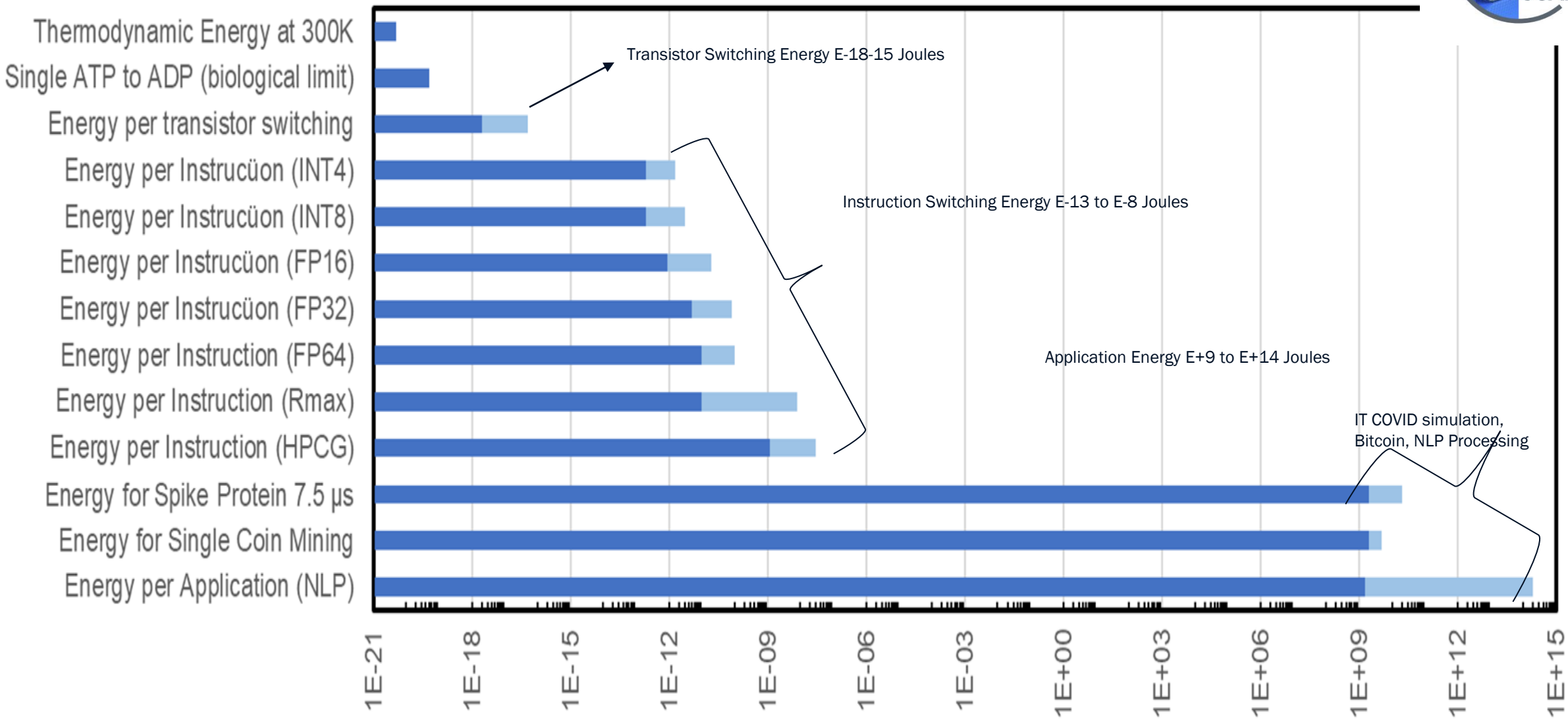
ENABLERS

- Metrology and Benchmarking
- Power and Control Electronics
- Manufacturing Energy Efficiency and Sustainability
- Education and Workforce Development

Added in August

But We Can't Succeed without help from "IT"

EES2 ENERGY EFFICIENCY SCALING

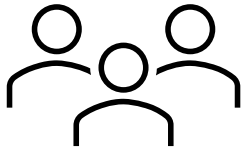


Energy in Joules Per Bit/ Instruction/ Application—1000X per tic mark, 36 OOM total

EES2 WG Highlight Talks by Month

Working Group	Materials and Devices	Circuits and Architectures	Het Intg Adv Pkg	Metrology & Benchmark	Power & Control	Software Algorithms	Mfg Energy Efficiency
Speaker (s)	John B, SLAC Steffen, CTI & J.P. Aligned Carbon	Azeez, Metis & Emre, Stonybrook	Na Li, Carbice & Moinuddin, ANL	Jim B, NIST	Paul S. Sandia	Brian H Micron	Prashant N ORNL
Working Group Highlight Month (s)	March June	April	May	August	July	July	June

Microelectronics Education and Workforce, Tim W,



Northwestern

...in addition to facilitated WG sessions

Office of Science Advanced Scientific Computing Research (ASCR)

Long-term investments in applied mathematics and computer science enabled exascale.

MPICH 2

R&D 100

2005

TOP500 # 1

GREEN500 # 2

HPL-MxP # 1

Frontier, #1 on the Top500, **leads the world in computational capability**, and is also **#2 in the world in energy efficiency**, and is **#1 in the world for AI capability**.

The exascale and AI-enabled science era will lead to dramatic capabilities to predict extreme events and their impacts on the electric grid across weather and climate time scales...

and will accelerate the design and deployment of clean-energy technologies to create a better future.

ASCR Co-design Partnerships' Long History

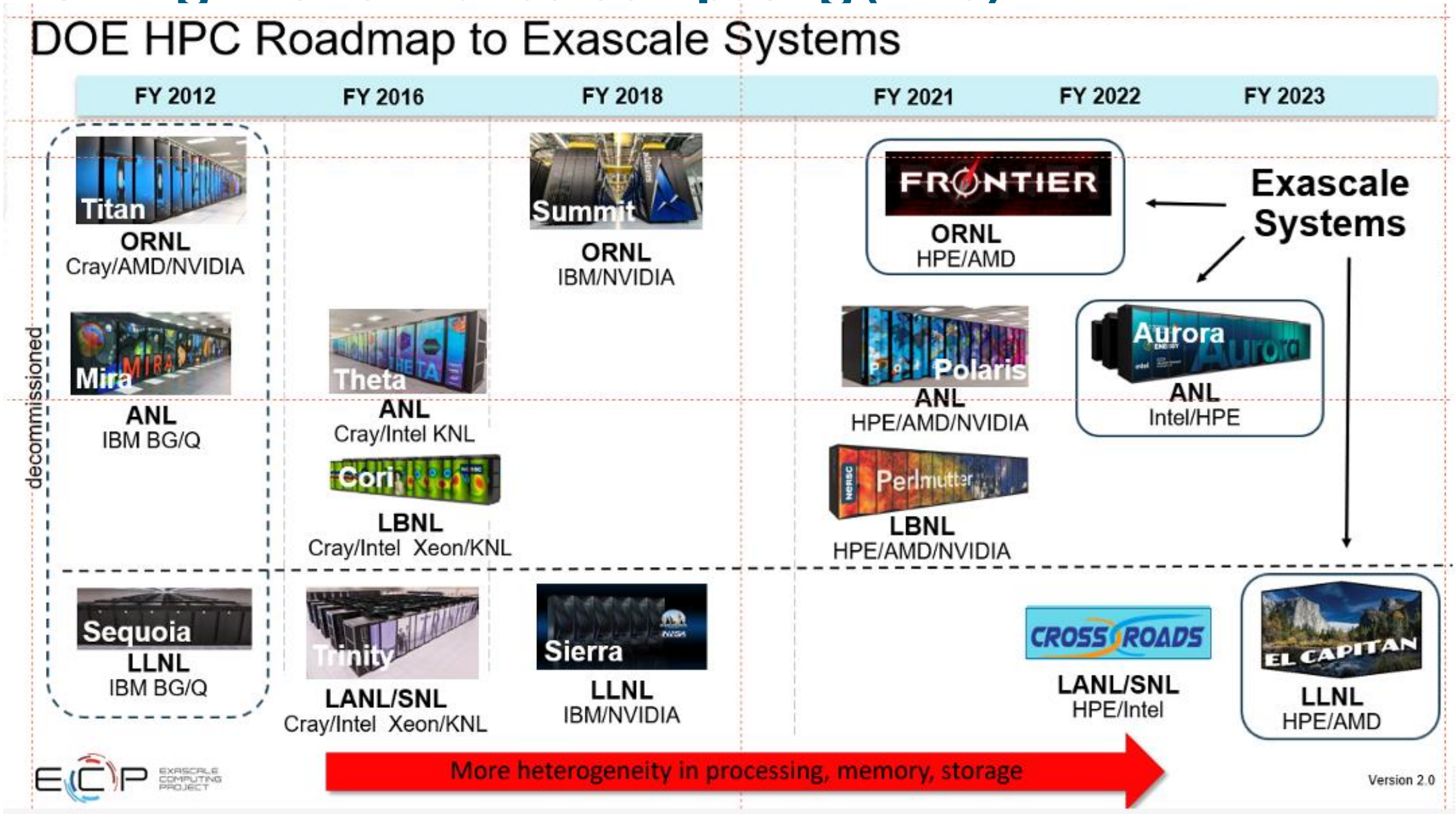
ASCR Workshop on Reimagining Codesign

Sponsored by the U.S. Department of Energy, Office of Advanced Scientific Computing Research - March 16-18, 2021 - https://www.orau.gov/ASCR-CoDesign/

- ASCR's recent past, in partnership with NNSA and other stakeholders, has invested significantly in codesign, including:
  - Vendor Collaborations:
    - Fast Forward 1, July 2012 – Sept 2014
    - Fast Forward 2, Nov 2014 – 2016
    - Design Forward, Oct 2013 – 2015
    - Design Forward 2, 2015 – 2017
    - ECP PathForward, June 2017 – *ending now*
    - Partnered with AMD, Cray, HPE, IBM, Intel, NVIDIA
    - Technologies related to compute, memory, networking, and storage.
    - For more history, see https://www.osti.gov/servlets/purl/1513941
  - Codesign Centers:
    - Pre-ECP Codesign Centers (starting 2012 - 2016):
      - Exascale Co-Design Center for Materials in Extreme Environments (ExMatEx)
      - Center for Exascale Simulation of Advanced Reactors (CESAR)
      - Center for Exascale Simulation of Combustion in Turbulence (EXaCT)
      - For more history, see https://www.csm.ornl.gov/workshops/applmath11/documents/talks/Pao_CoDesign.pdf.pdf
    - ECP Codesign, including:
      - Center for Efficient Exascale Discretizations (CEED)
      - Co-design Center for Online Data Analysis and Reduction at the Exascale (CODAR)
      - Co-design Center for Particle Applications (CoPA)
      - For more information, see https://www.exascaleproject.org/research-group/co-design-centers/

This is just some of the *recent* past. ASCR has a long history with codesign.

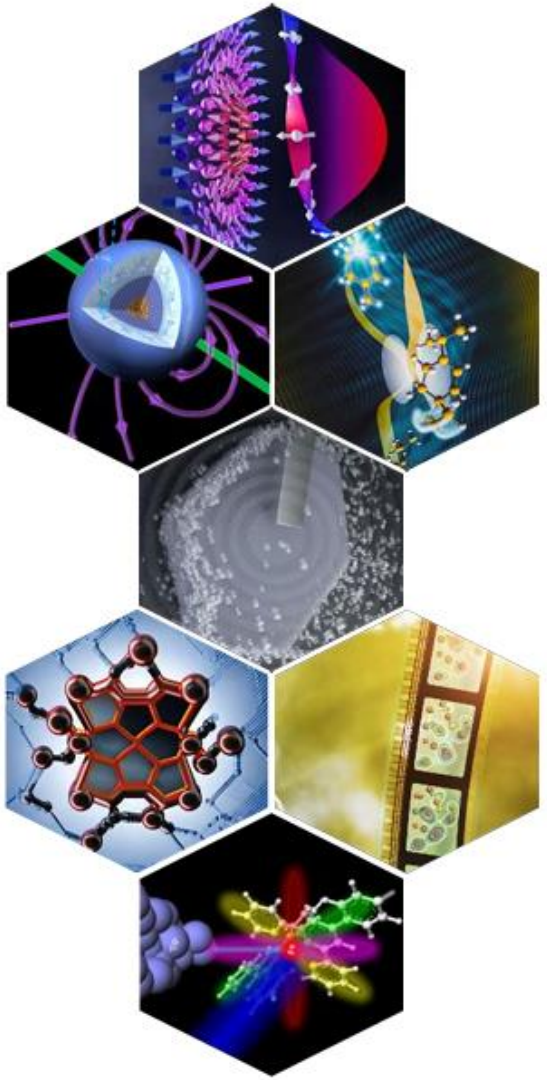
DOE High Performance Computing (HPC)



Basic Energy Sciences: Understanding Matter and Energy at Electronic, Atomic, and Molecular Levels

BES fulfills its mission through:

- Supporting **basic research**
  - "Grand Challenge" science
  - Discovery and design of materials and chemical processes that underpin a broad range of energy technologies
- Operating **world-class scientific user facilities** in X-ray, neutron, and nanoscale science
- Managing **construction and upgrade projects** to maintain **world-leading** scientific user facilities
- Ensuring **broad participation** in the research portfolio and user communities



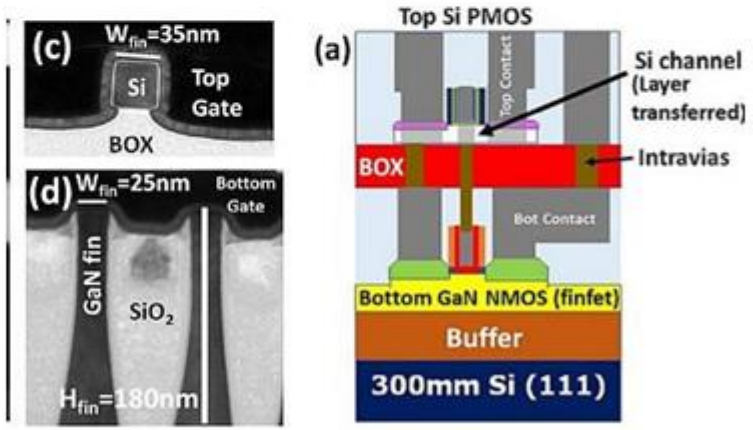
BES—2018 Workshop, 2021 Co-design FOA; 2023 EFRC

Power electronics: infrastructure for 2050

- Next generation power electronics are critical to the future grid and a sustainable energy system; smaller size, higher temperature, higher voltages
- New UWBG materials, heterogenous integration, interface and degradation science
- Power electronics will be the "BIL" of 2050; do the science and microelectronics now

Pittsburgh

Washington



GaN FinFet on Si CMOS – low loss power inverter – Intel 2021