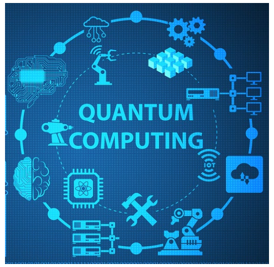
Center for Semiconductor Manufacturing

Challenges and opportunities at the confluence of semiconductor manufacturing (SM) and applied math (AM): The SM-AM Agora



Krishna Muralidharan (MSE)

Xiaodong Yan (MSE)

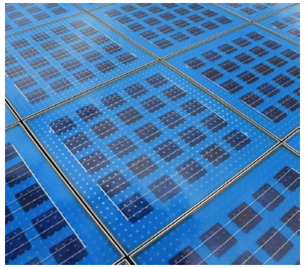
Monica Titus (CHEE)

Soheil Salehi (ECE)

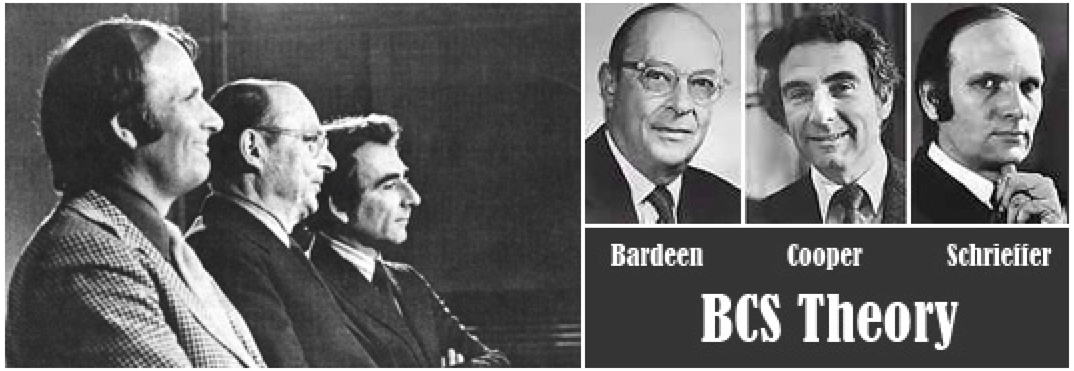
Brian Kim (MSE)

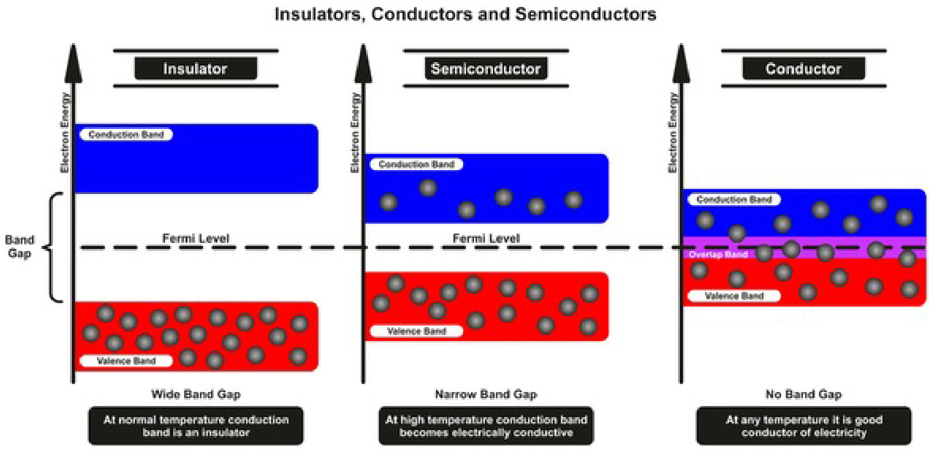
Ryan Biggie* (MSE)

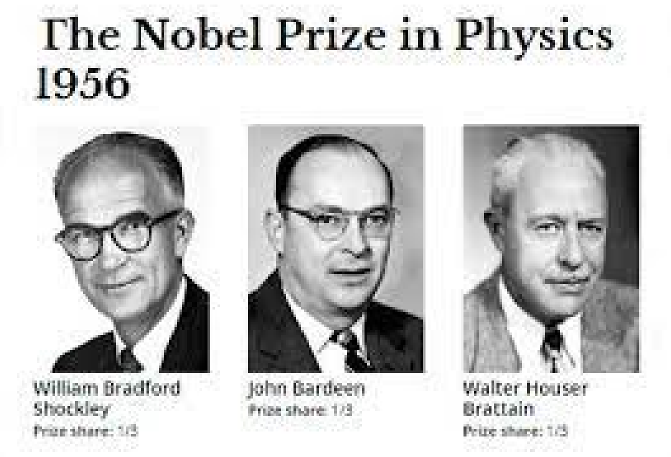
Patrick Lohr* (CHEE)



Are You Smarter than a 5th Grader?




Insulators, Conductors and Semiconductors
Insulator
Semiconductor
Conductor
Electron Energy
Conduction Band
Valence Band
Overlap Band
Band Gap
Fermi Level
Wide Band Gap
Narrow Band Gap
No Band Gap
At normal temperature conduction band is an insulator
At high temperature conduction band becomes electrically conductive
At any temperature it is good conductor of electricity

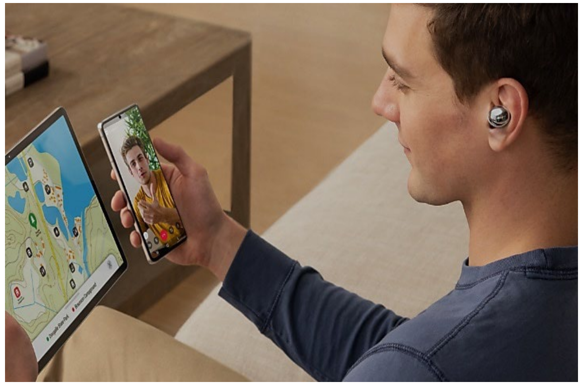


Three American physicists, John Bardeen, Walter H. Brattain, and William Shockley, were honored last month in Stockholm, where they were jointly awarded the 1956 Nobel Prize in Physics for “their investigations on semiconductors and the discovery of the transistor effect”.

Semiconductors power the world

"semiconductor industry is an irreplaceable enabler of **tens of trillions of dollars** of annual economic activity worldwide"

Mobile



Wearable

Medical

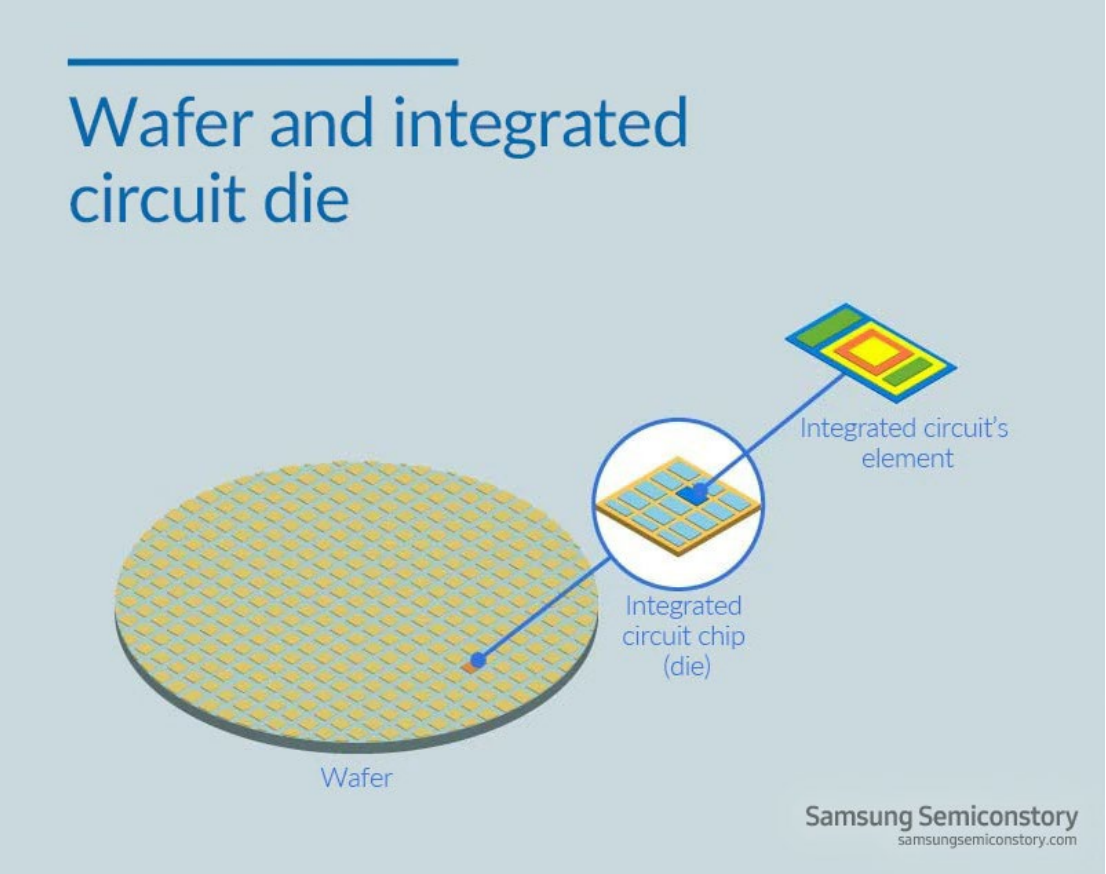
Automotive

Energy

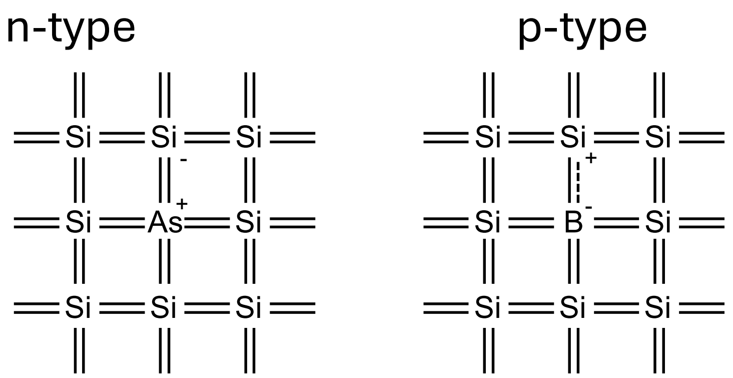
Manufacturing

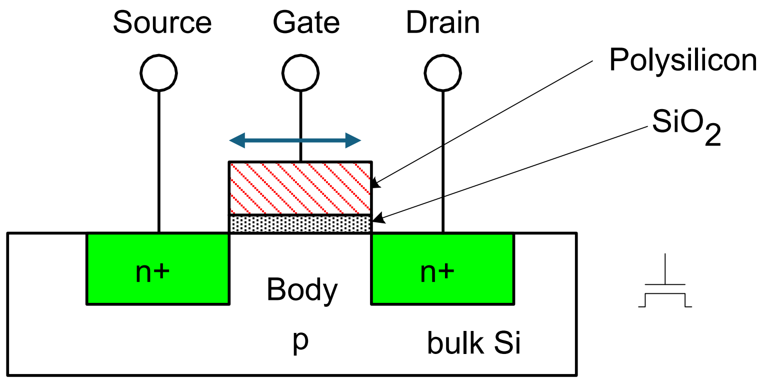
Silicon is the building block of Integrated Circuits/semiconductor chips

ICs contain billions of components: resistors, capacitors, ***transistors***

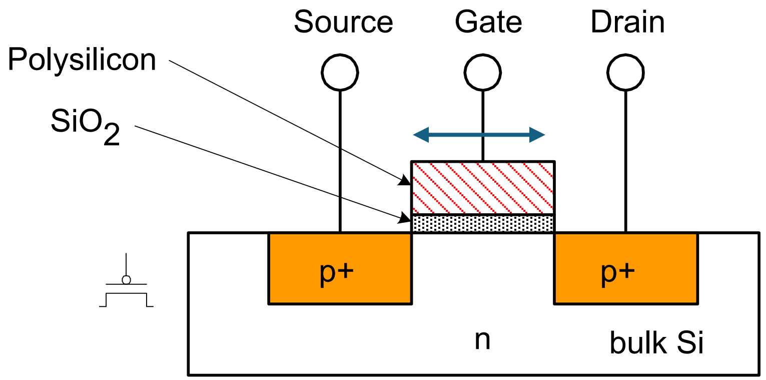

Wafer and integrated circuit die
Integrated circuit's element
Integrated circuit chip (die)
Wafer
Samsung Semiconstory
samsungsemiconstory.com

Transistors: logic gateway





n-mos

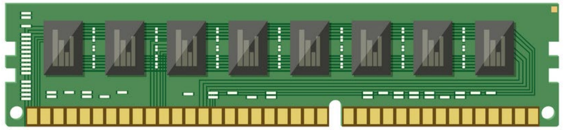


p-mos

Four terminals: gate, source, drain, body

IC classification

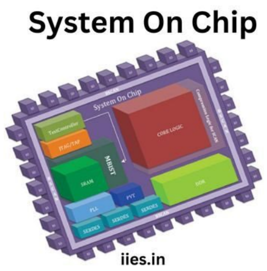
Memory chips



Microprocessors (CPU, GPU)

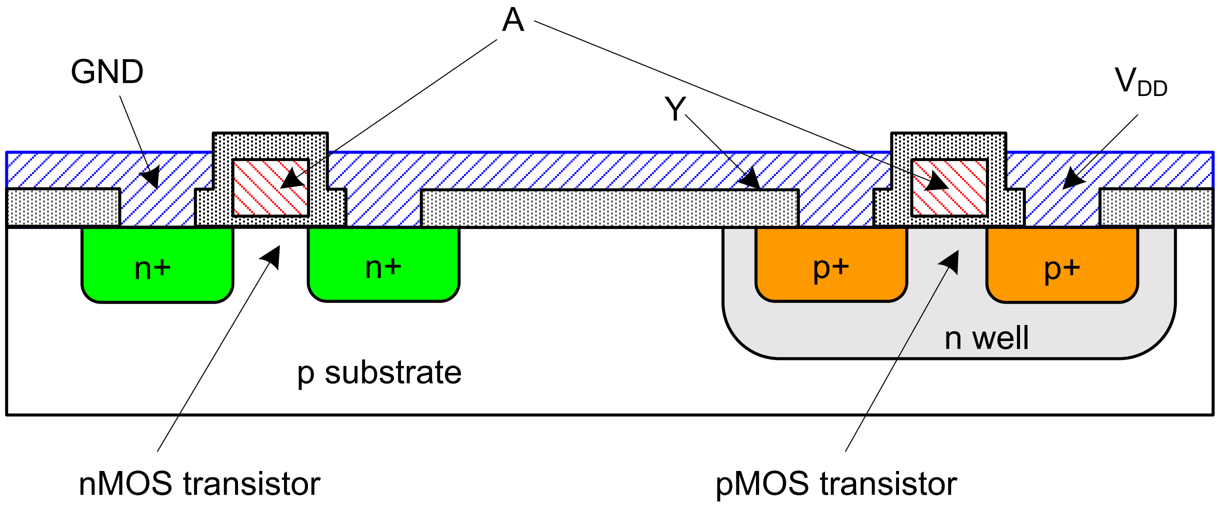
ASIC

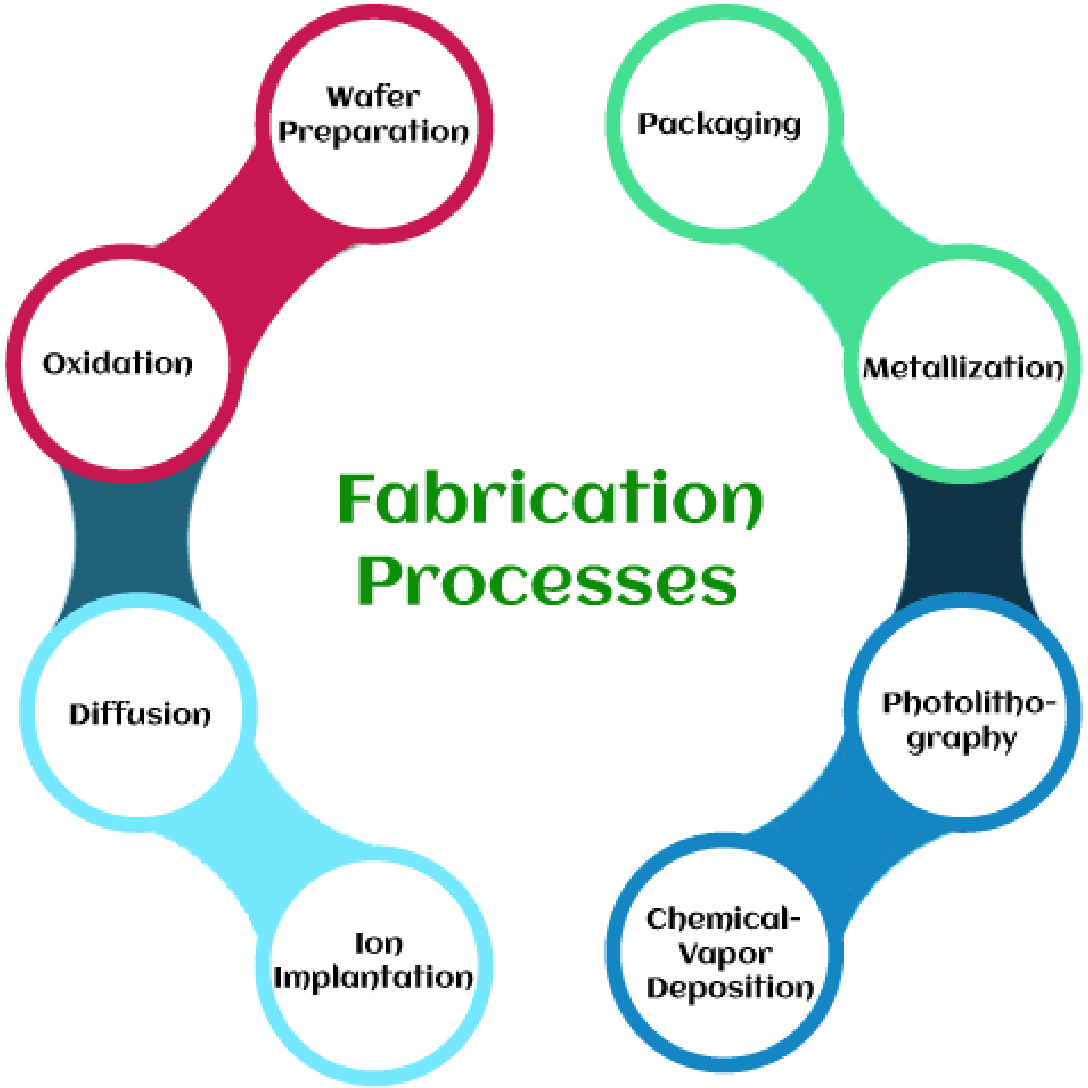
System on Chip



Analog chips

C-MOS: dominant technology for constructing IC



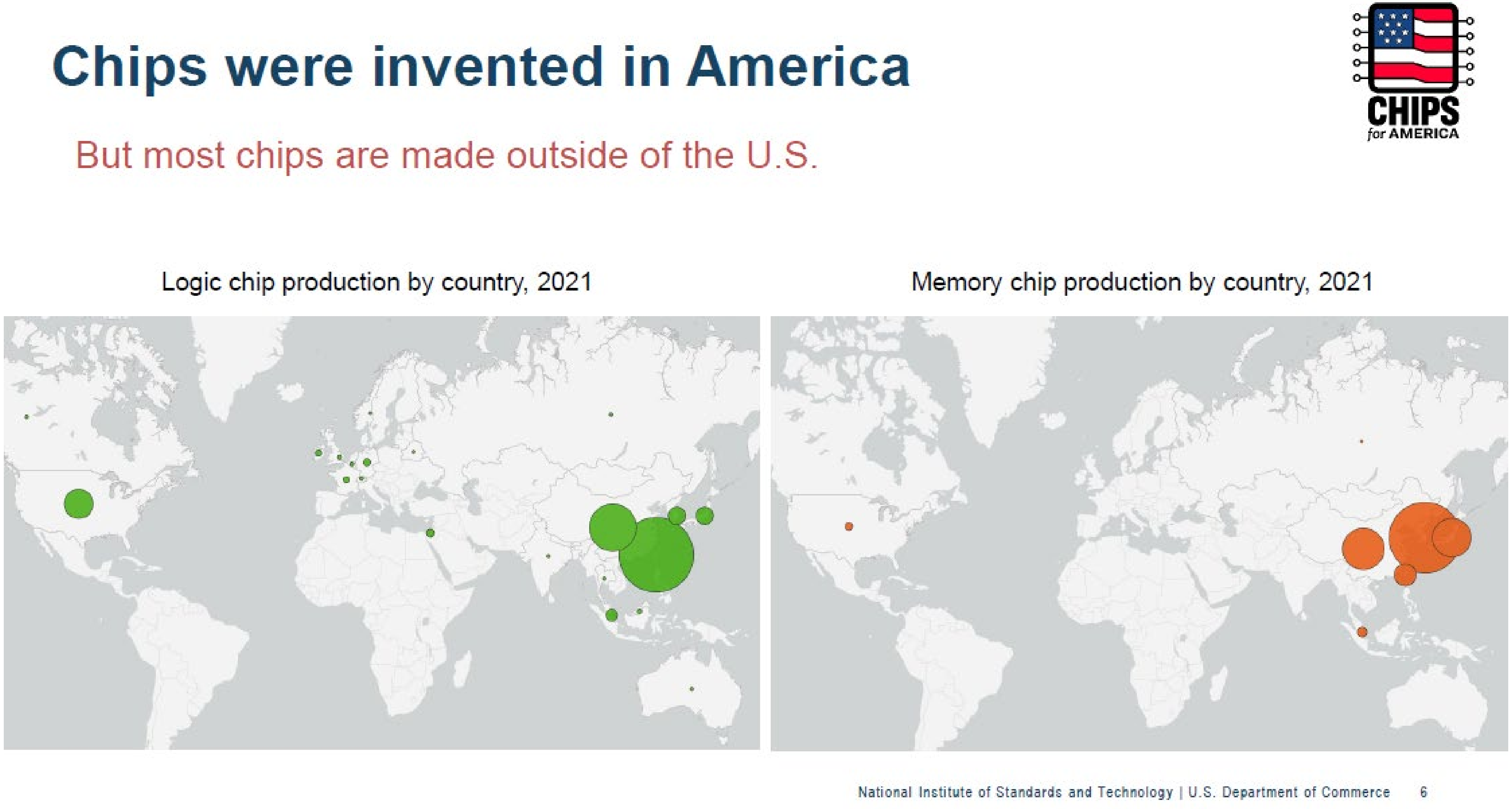


Fabrication Processes

The CHIPs Act

~$50 Billion for workforce, reshoring, R&D

CHIPS
for AMERICA



Chips were invented in America

But most chips are made outside of the U.S.

Logic chip production by country, 2021

Memory chip production by country, 2021

National Institute of Standards and Technology | U.S. Department of Commerce 6

The CHIPs Act

CHIPS for America Vision

CHIPS for America

Economic Security

This act enables us to build more resilient supply chains for important components.

National Security

This act enables us to bring the most sophisticated technologies back to the U.S.

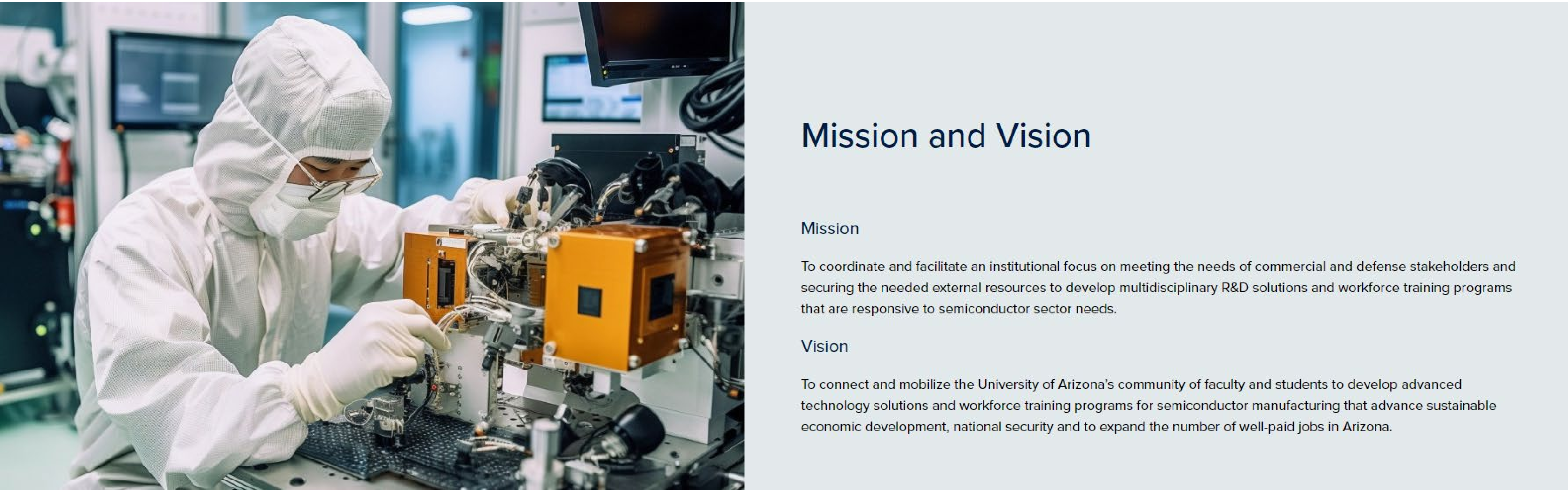
Future Innovation

Chips are key to the technologies and industries of the future, so we need to be at the forefront. This act will ensure long-term U.S. leadership in the sector.

National Institute of Standards and Technology | U.S. Department of Commerce 7

The Center for Semiconductor Manufacturing

https://csm.arizona.edu/



Mission and Vision

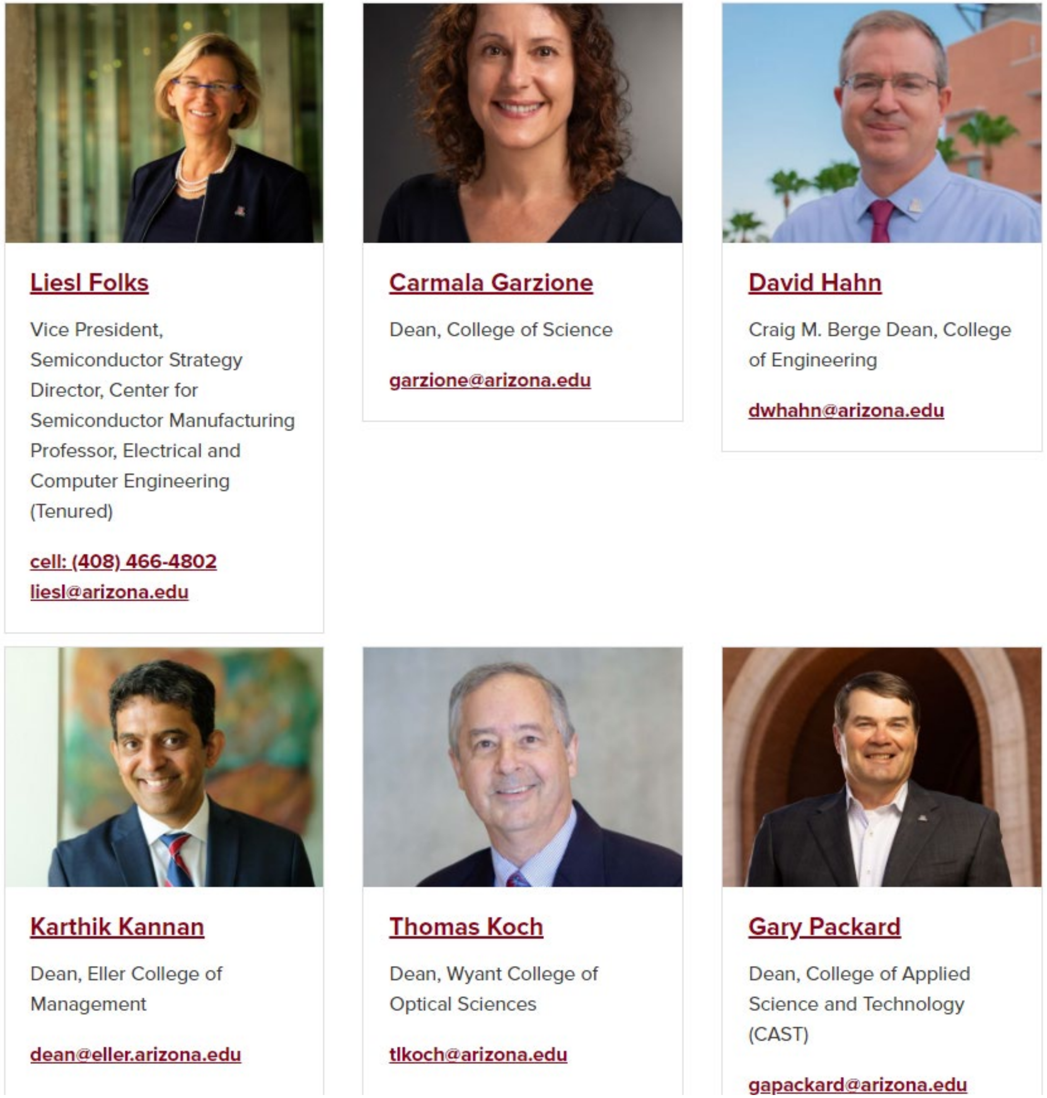
Mission

To coordinate and facilitate an institutional focus on meeting the needs of commercial and defense stakeholders and securing the needed external resources to develop multidisciplinary R&D solutions and workforce training programs that are responsive to semiconductor sector needs.

Vision

To connect and mobilize the University of Arizona's community of faculty and students to develop advanced technology solutions and workforce training programs for semiconductor manufacturing that advance sustainable economic development, national security and to expand the number of well-paid jobs in Arizona.

The Center for Semiconductor Manufacturing



Liesl Folks

Vice President, Semiconductor Strategy
Director, Center for Semiconductor Manufacturing
Professor, Electrical and Computer Engineering (Tenured)

cell: (408) 466-4802
liesl@arizona.edu

Carmala Garzione

Dean, College of Science

garzione@arizona.edu

David Hahn

Craig M. Berge Dean, College of Engineering

dwhahn@arizona.edu

Karthik Kannan

Dean, Eller College of Management

dean@eller.arizona.edu

Thomas Koch

Dean, Wyant College of Optical Sciences

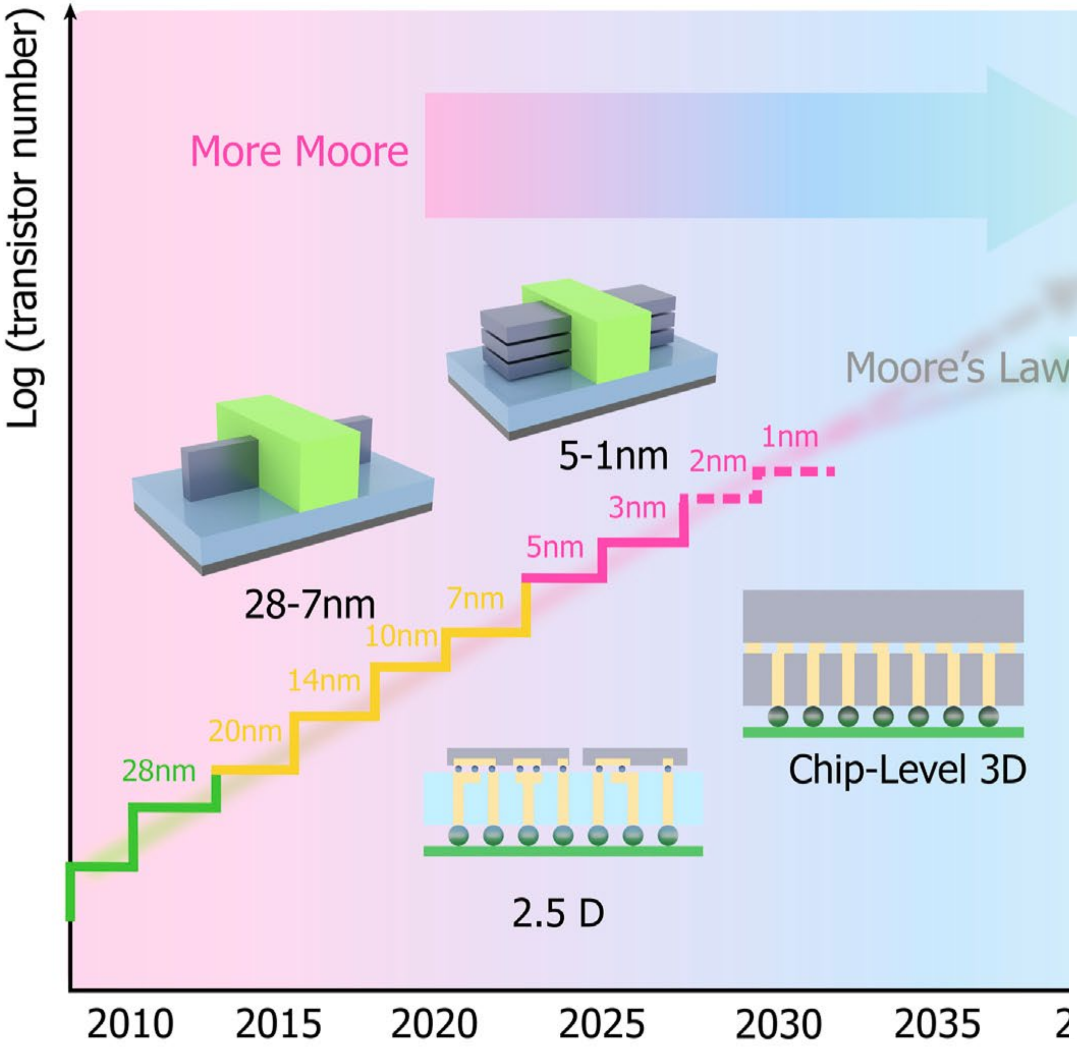
tlkoch@arizona.edu

Gary Packard

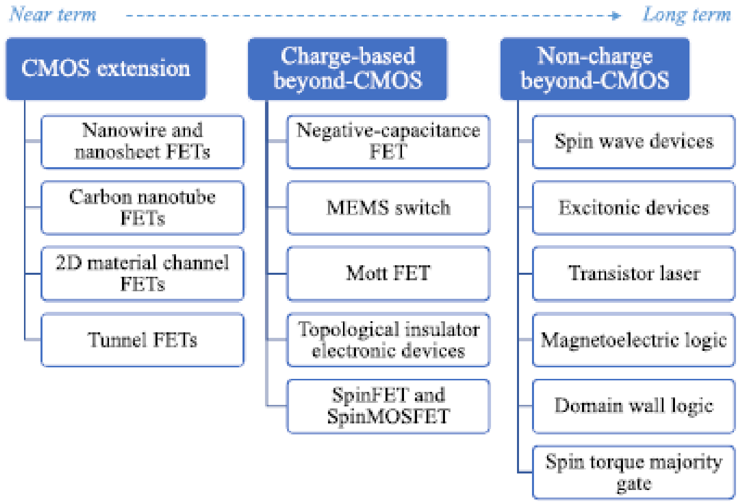
Dean, College of Applied Science and Technology (CAST)

gapackard@arizona.edu

The post-Moore world



https://doi.org/10.3389/fmats.2023.1224537

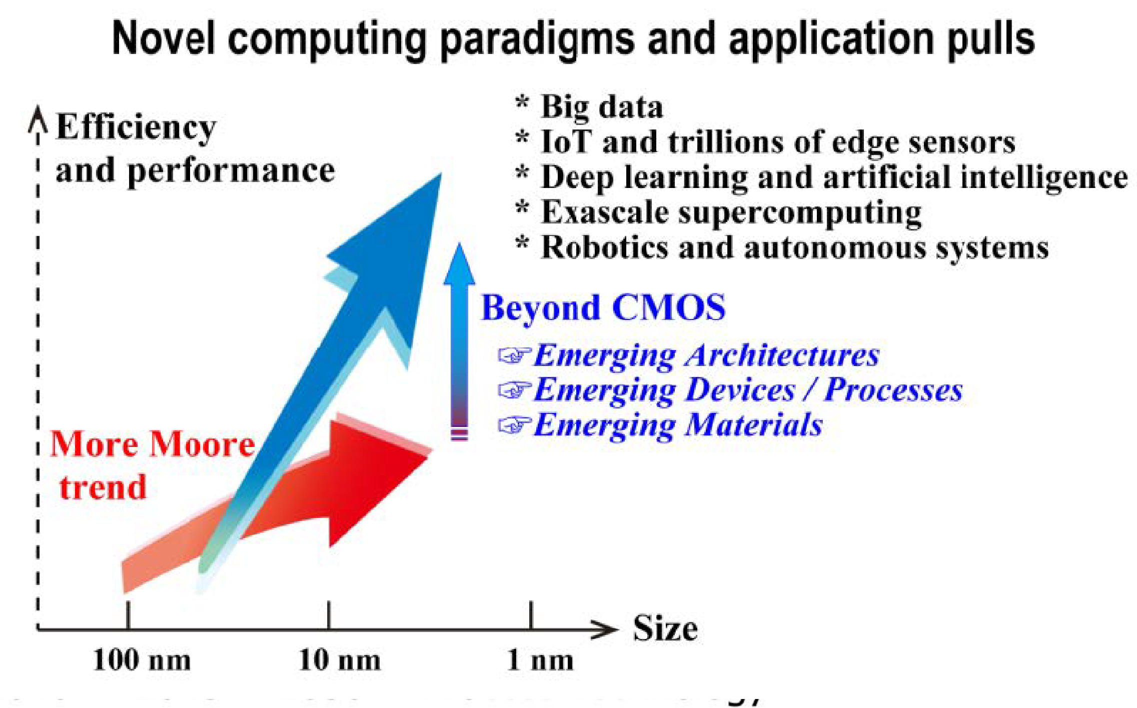


CMOS extension
- Nanowire and nanosheet FETs
- Carbon nanotube FETs
- 2D material channel FETs
- Tunnel FETs

Charge-based beyond-CMOS
- Negative-capacitance FET
- MEMS switch
- Mott FET
- Topological insulator electronic devices
- SpinFET and SpinMOSFET

Non-charge beyond-CMOS
- Spin wave devices
- Excitonic devices
- Transistor laser
- Magnetoelectric logic
- Domain wall logic
- Spin torque majority gate

Jpn. J. Appl. Phys. 61, SM1003 (2022)


Novel computing paradigms and application pulls

* Big data
* IoT and trillions of edge sensors
* Deep learning and artificial intelligence
* Exascale supercomputing
* Robotics and autonomous systems

THE INTERNATIONAL ROADMAP FOR DEVICES AND SYSTEMS: 2022

Challenges and Opportunities

- Chiplets, SoC and heterogenous integration

"Companies are now talking of "chiplets" comprising smaller "building blocks" to be mixed and matched much like **Lego Blocks** creating new flexibility in design and production. Intel, TSMC, Samsung, Arm, Qualcomm, and others have come together to establish standards for building these chips. Mediatek and Nvidia recently announced a collaboration on chiplets."

- Post-silicon technologies: low-D materials, high band-gap semiconductors
- Beyond CMOS: new avenues using photonic , spintronic and bio ICs
- AI-specific chips: "AI chips include graphics processing units (GPUs), field-programmable gate arrays (FPGAs), and application-specific integrated circuits (ASICs) "
- Hardware security

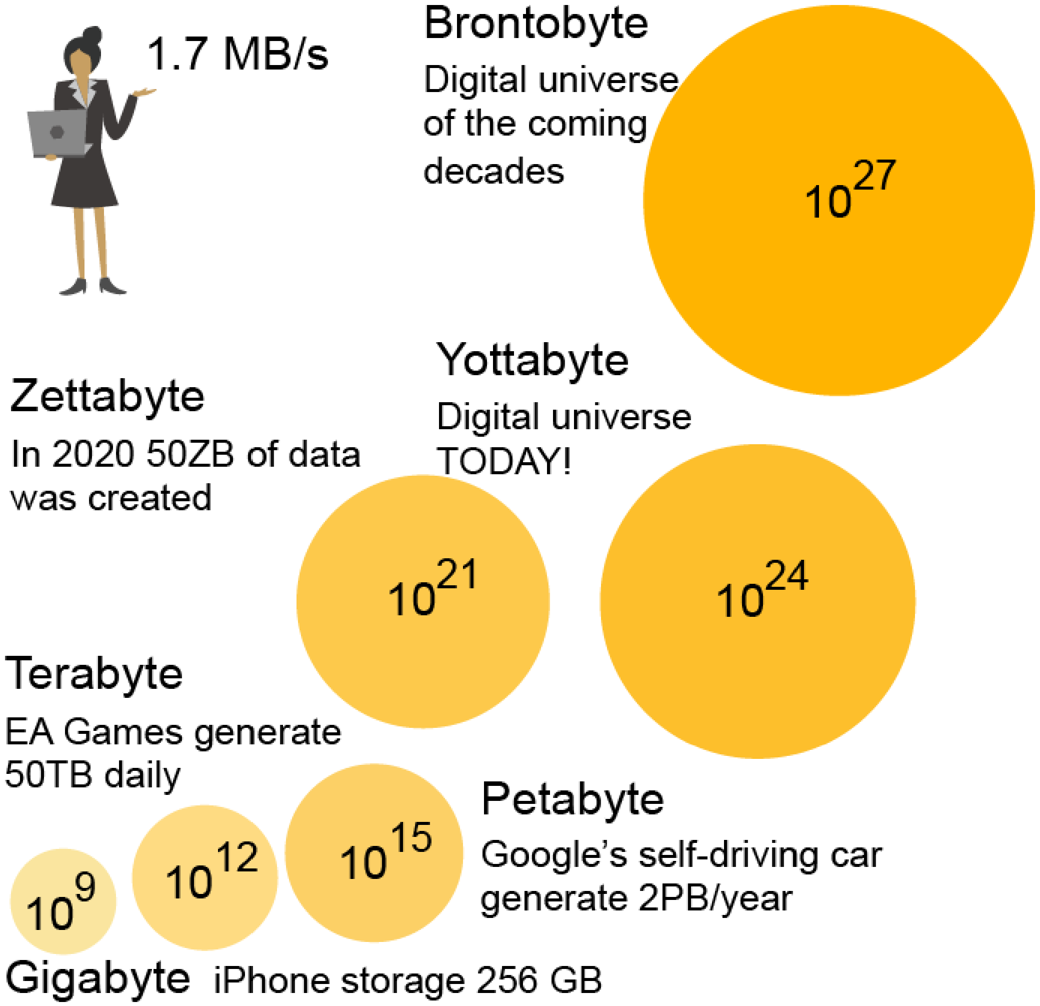
- electron to devices (e2d): Multiphysics, multiscale algorithms (materials & manufacturing)
- AI for generative design of ICs
- Computer vision for defect detection
- New computing paradigms (neuromorphic, quantum and quantum analog)
- New approaches for hardware security

https://www.linkedin.com/pulse/gordon-moores-law-hits-technical-limits-lego-inspires-kumud-goel-fgozc?trk=articles_directory

‘Neu’ computing paradigms in the Post-Moore era: Xiaodong Yan

Challenges in traditional computing arise in leveraging big data applications

Ever growing pressure for data size



1.7 MB/s

Brontobyte
Digital universe of the coming decades

Yottabyte
Digital universe TODAY!

Zettabyte
In 2020 50ZB of data was created

Terabyte
EA Games generate 50TB daily

Petabyte
Google's self-driving car generate 2PB/year

Gigabyte iPhone storage 256 GB

Seagate and IDC Global Datasphere

Energy cost in data manipulation

- Data transmission
- Data processing

5%-15% of Energy

x 300 years

Vidal J, Climate Home News, 2017

Supercomputer vs. Human brain on energy and area efficiency

von-Neumann Computing

Human brain

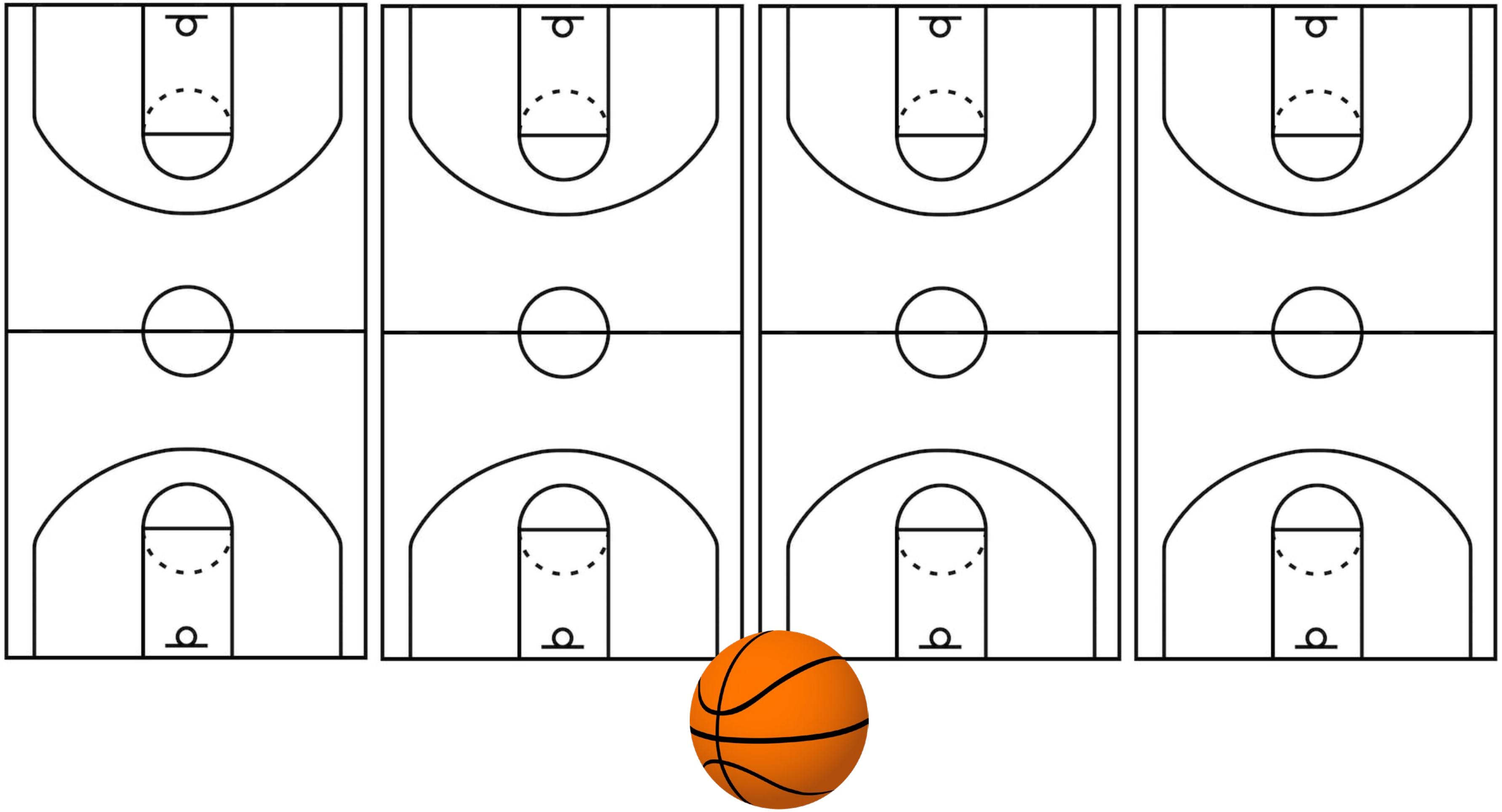
v.s.

20 Watts, 10^{15} synapses,
10^{11} neurons,
10^{16} synaptic events/sec

1.4 M Watts (>500 times average household power consumption), 8192 processors,
10^9 synaptic events/sec (10^9 flops)

Area efficiency…

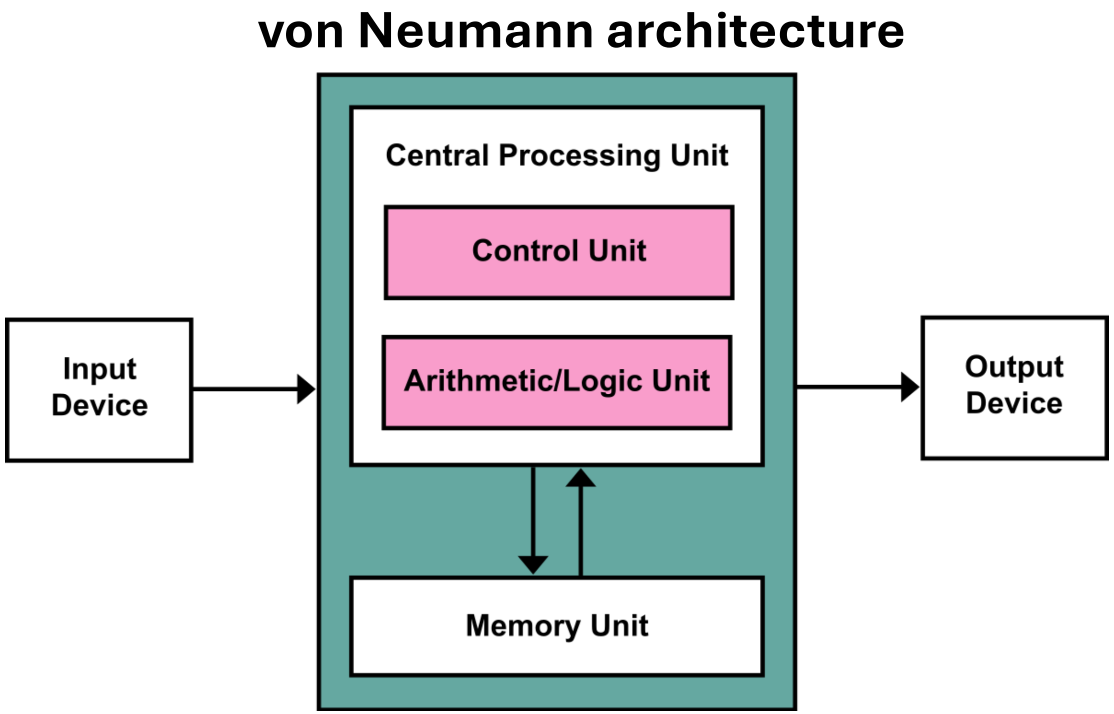
Supercomputer **Human brain**

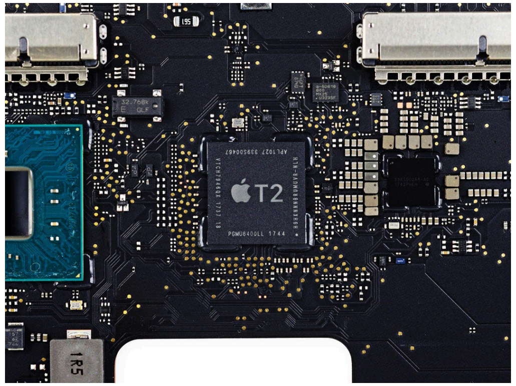


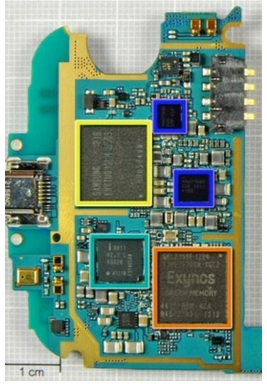
Motivation

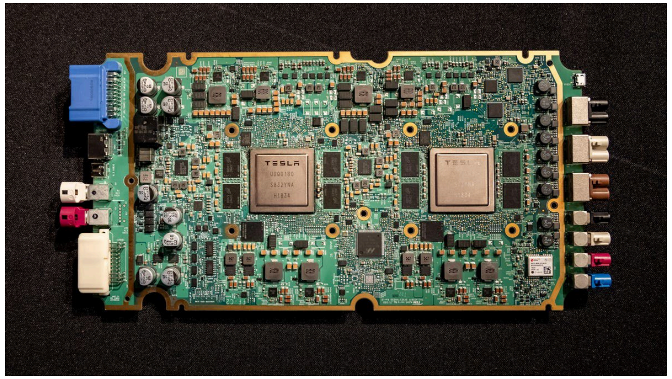
Can we make our computing hardware as efficient as the human brain?

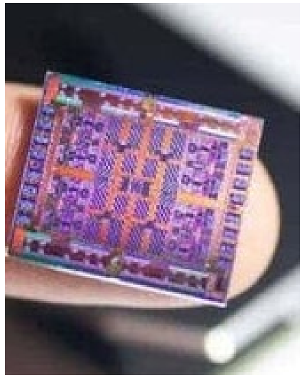
von-Neumann architecture and the “von-Neumann bottleneck”



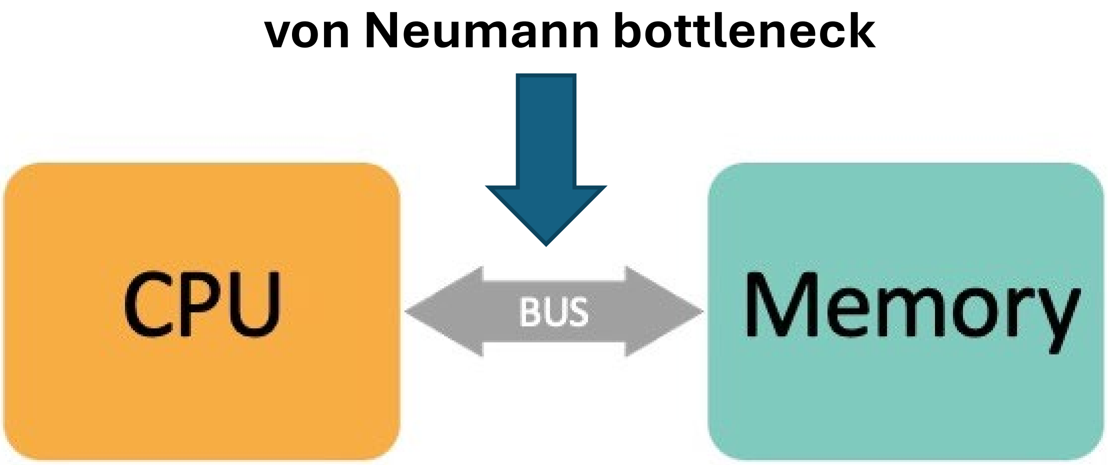


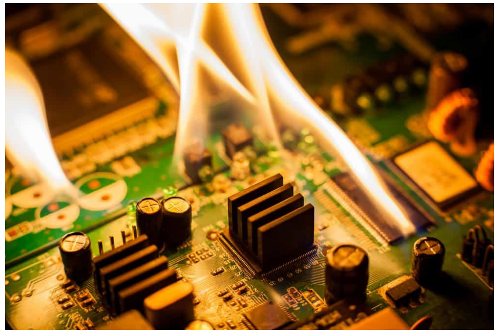






von-Neumann architecture and the “von-Neumann bottleneck”

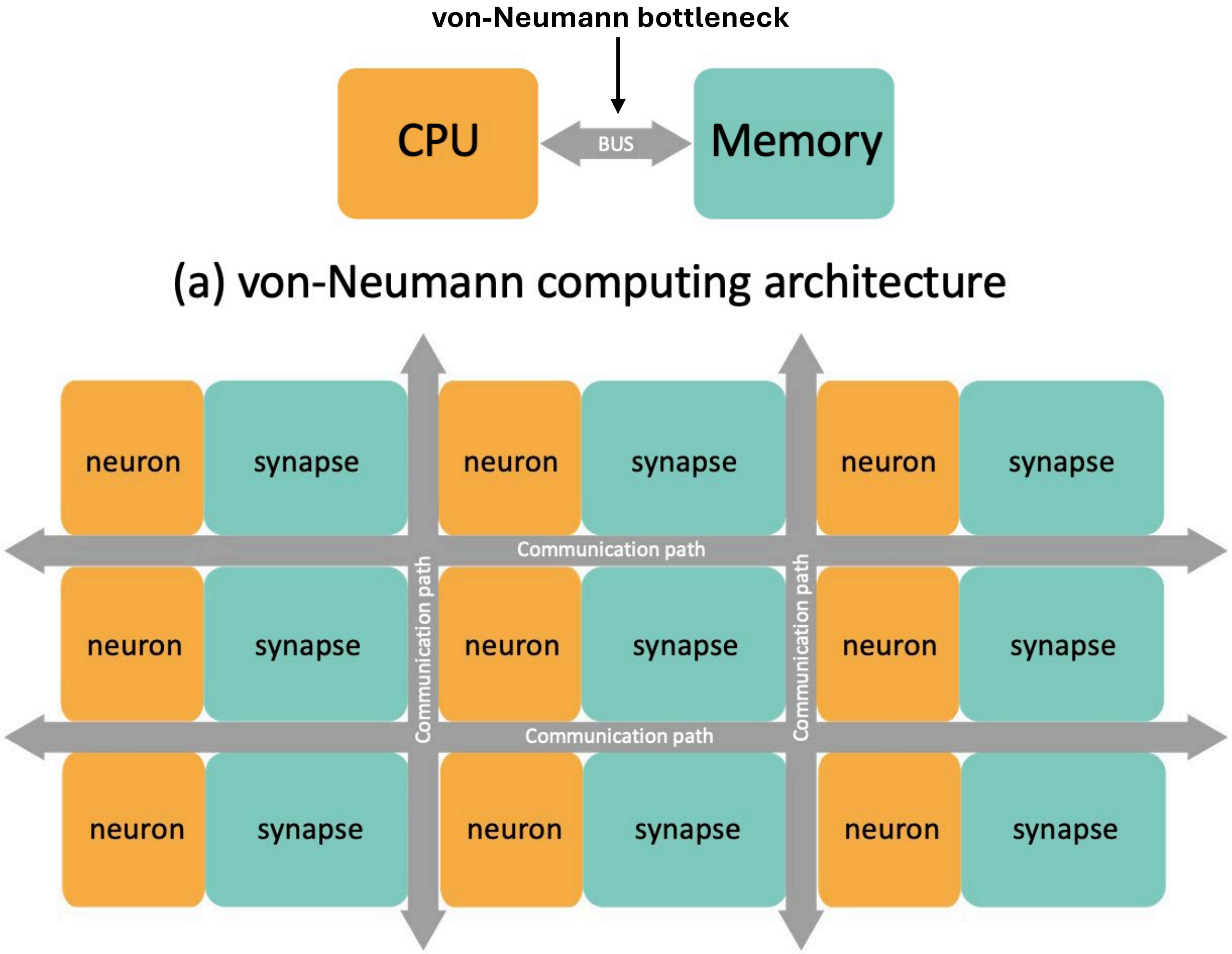

von Neumann bottleneck



a human: 3.5 gallons a day

350,000 gallons of water a day

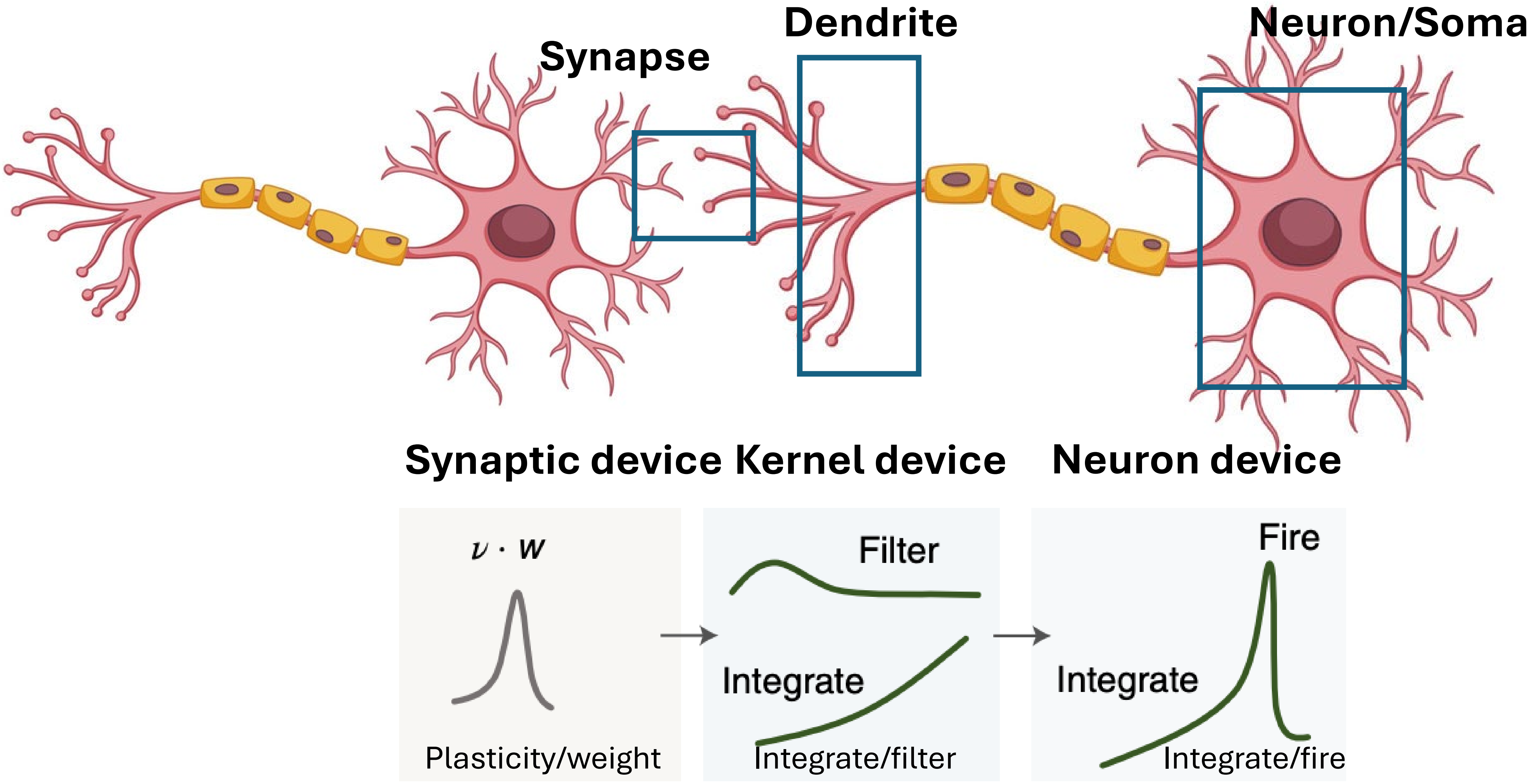
Neuromorphic computing: an efficient computing paradigm inspired by the brain



(a) von-Neumann computing architecture

(b) Neuromorphic computing architecture

Neuromorphic computing: an efficient computing paradigm inspired by the brain and realized by emerging semiconductor devices

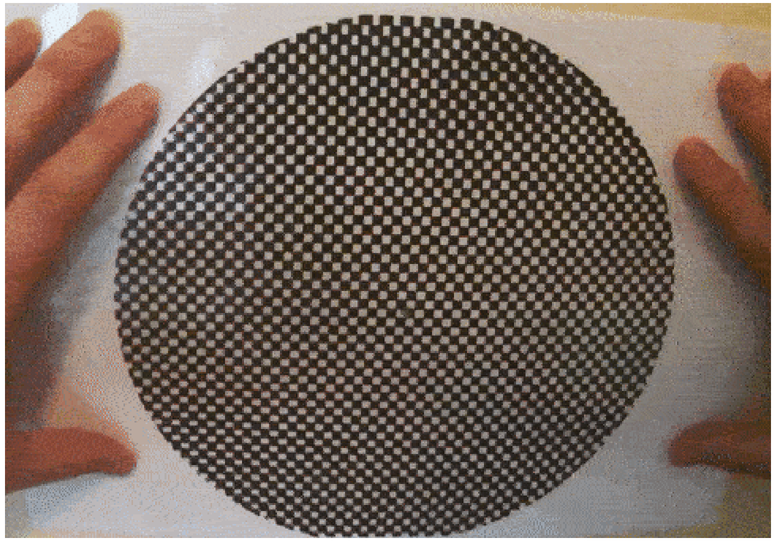


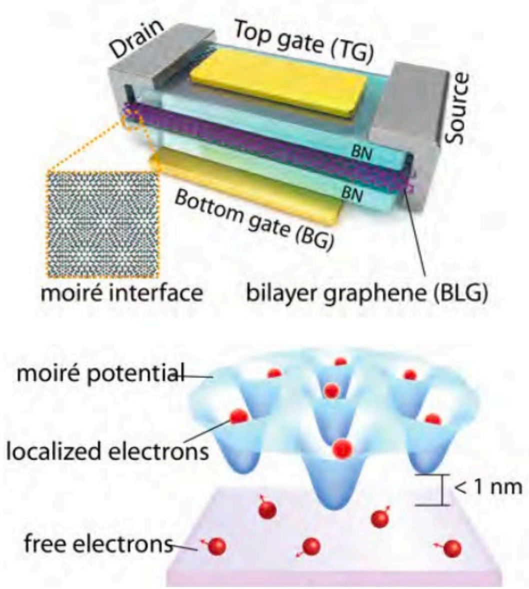
An example: moiré synaptic transistors

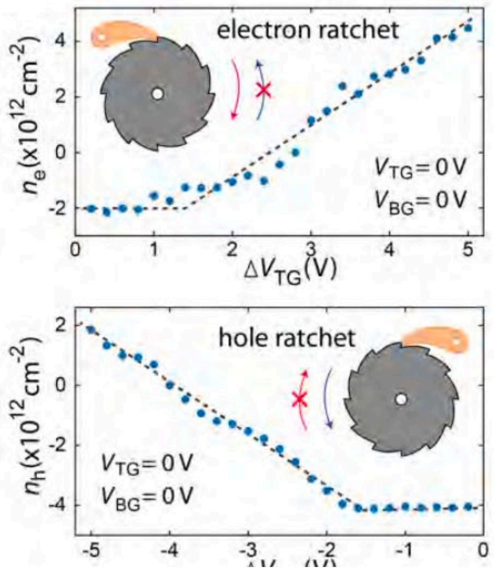
Moiré Synaptic Transistor with Room-Temperature Neuromorphic Functionality

X. Yan, Z. Zheng, V. K. Sangwan, J. H. Qian, *et al.*, *Nature*, **624**, 551 (2023).

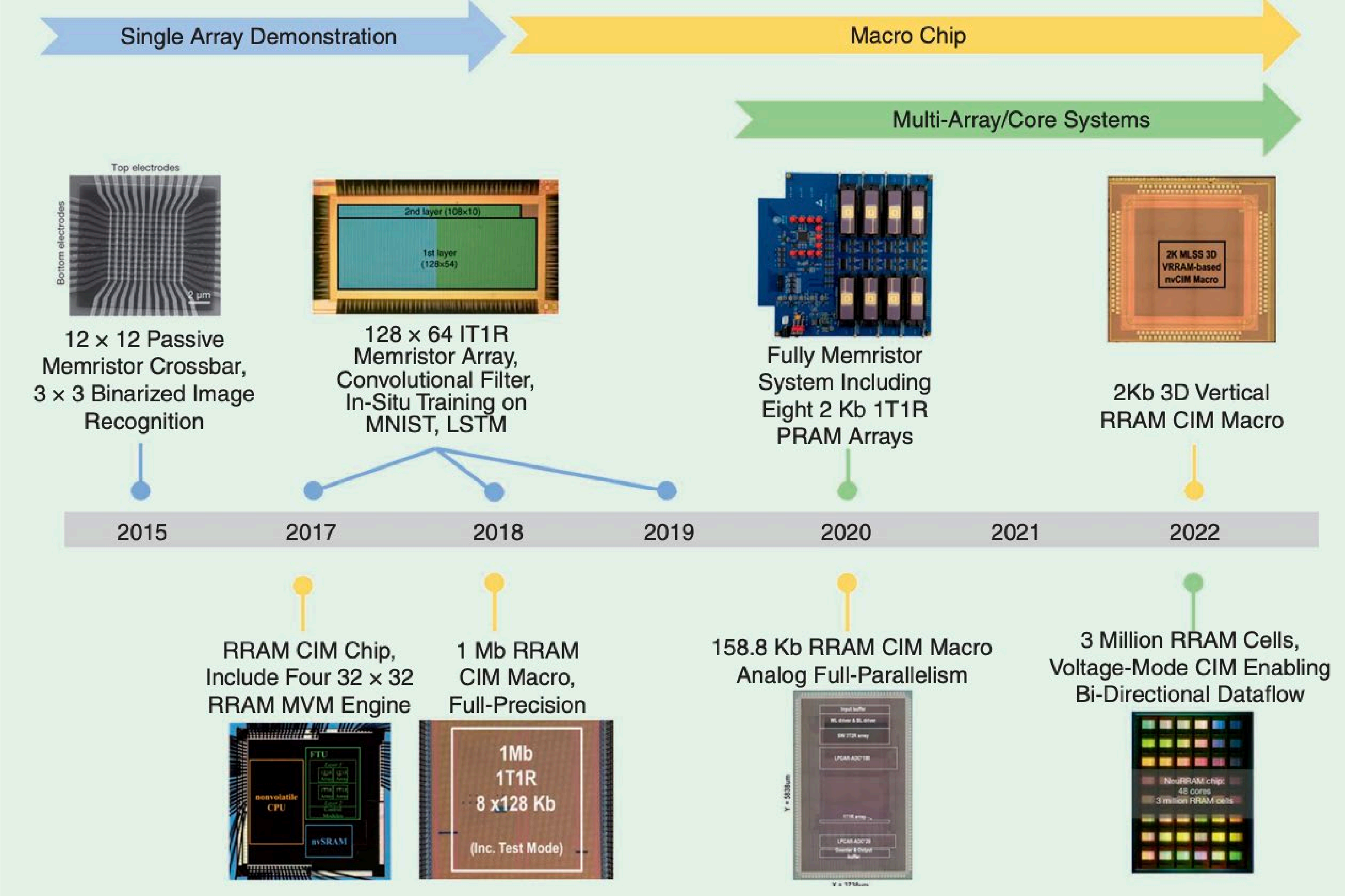
Collaborators: S. E. Liu, K. Watanabe, T. Taniguchi, S. Y. Xu, P. Jarillo-Herrero, Q. Ma







Semiconductor Chips for Neuromorphic Computing: Computing-In-Memory


Top electrodes
Bottom electrodes
2 µm
2nd layer (108×10)
1st layer (128×54)
2K MLSS 3D VRRAM-based nvCIM Macro
FTU
nonvolatile CPU
nvSRAM
1Mb 1T1R 8 x128 Kb (Inc. Test Mode)
NeuRRAM chip: 48 cores 3 million RRAM cells

Single Array Demonstration

Macro Chip

Multi-Array/Core Systems

12 × 12 Passive Memristor Crossbar, 3 × 3 Binarized Image Recognition

128 × 64 IT1R Memristor Array, Convolutional Filter, In-Situ Training on MNIST, LSTM

Fully Memristor System Including Eight 2 Kb 1T1R PRAM Arrays

2Kb 3D Vertical RRAM CIM Macro

2015 2017 2018 2019 2020 2021 2022

RRAM CIM Chip, Include Four 32 × 32 RRAM MVM Engine

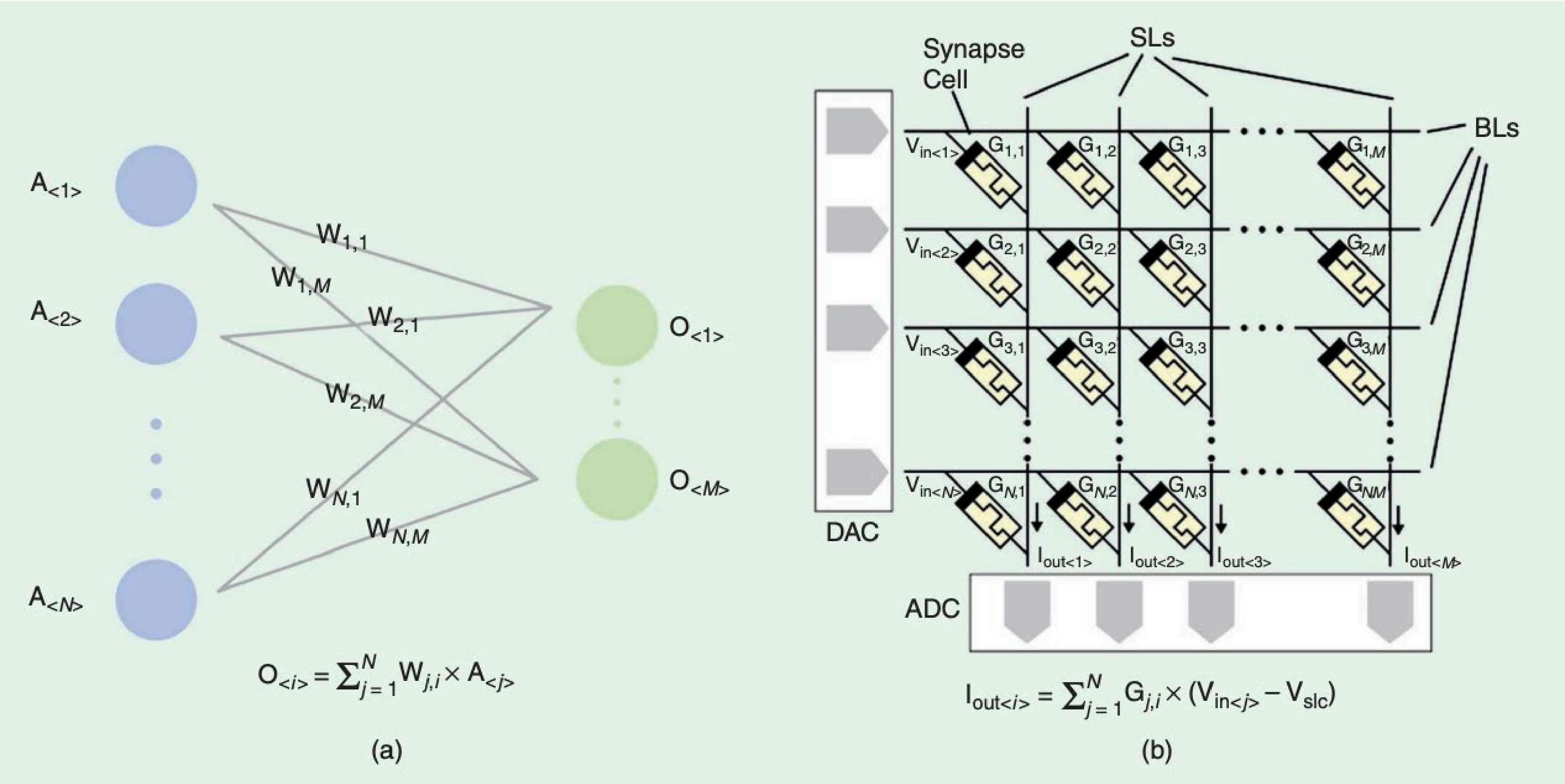
1 Mb RRAM CIM Macro, Full-Precision

158.8 Kb RRAM CIM Macro Analog Full-Parallelism

3 Million RRAM Cells, Voltage-Mode CIM Enabling Bi-Directional Dataflow

Co-design of software and hardware

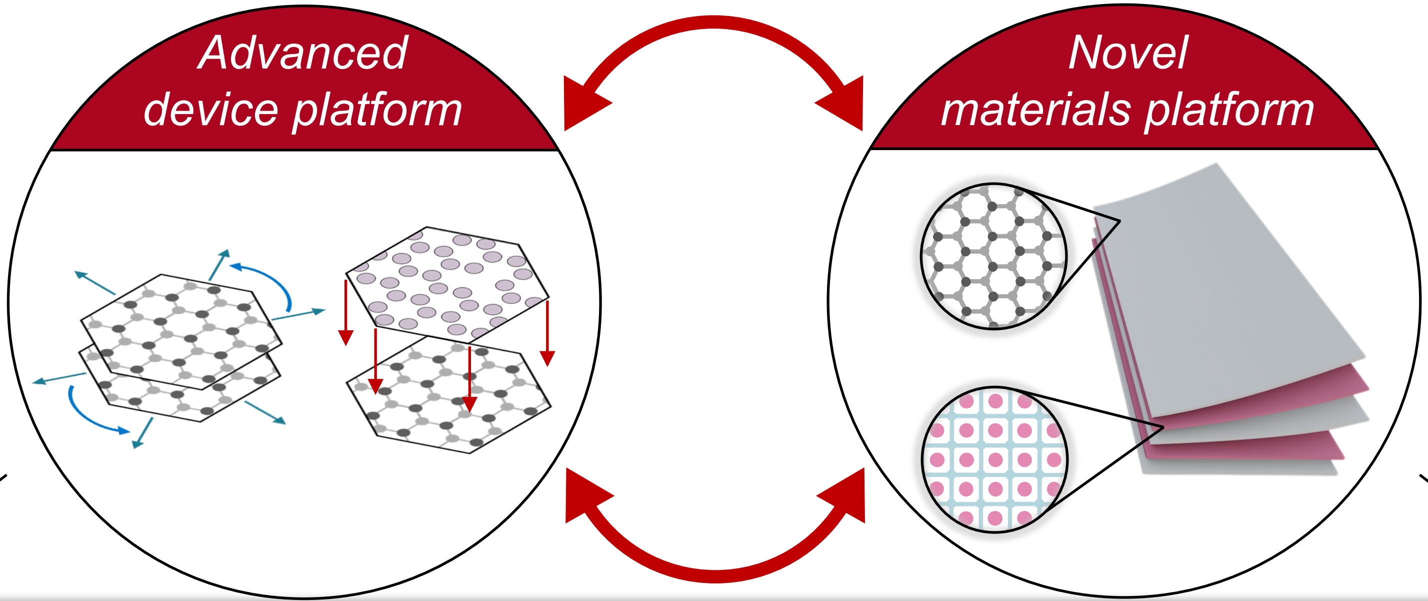
Mathematical Neural Network **Hardware Neural Network**



$A_{<1>}$
$A_{<2>}$
$A_{<N>}$
$W_{1,1}$
$W_{1,M}$
$W_{2,1}$
$W_{2,M}$
$W_{N,1}$
$W_{N,M}$
$O_{<1>}$
$O_{<M>}$
$O_{<i>} = \sum_{j=1}^{N} W_{j,i} \times A_{<j>}$
(a)
Synapse Cell
SLs
BLs
DAC
$V_{in<1>}$
$V_{in<2>}$
$V_{in<3>}$
$V_{in<N>}$
$G_{1,1}$ $G_{1,2}$ $G_{1,3}$ $G_{1,M}$
$G_{2,1}$ $G_{2,2}$ $G_{2,3}$ $G_{2,M}$
$G_{3,1}$ $G_{3,2}$ $G_{3,3}$ $G_{3,M}$
$G_{N,1}$ $G_{N,2}$ $G_{N,3}$ G_{NM}
$I_{out<1>}$ $I_{out<2>}$ $I_{out<3>}$ $I_{out<M>}$
ADC
$I_{out<i>} = \sum_{j=1}^{N} G_{j,i} \times (V_{in<j>} - V_{slc})$
(b)

‘New materials in the Post-Moore era:
Brian Kim

Next-generation 2D/quantum materials and devices

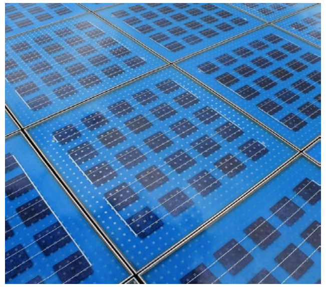


Advanced device platform

Novel materials platform

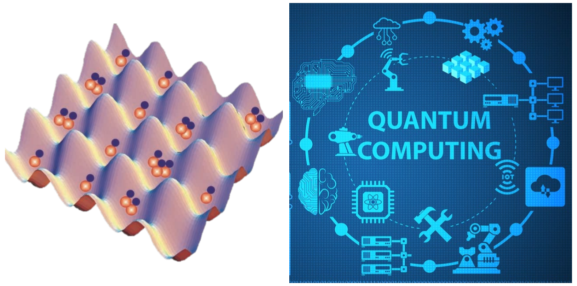
Emerging nanophotonics

Next-generation photovoltaics

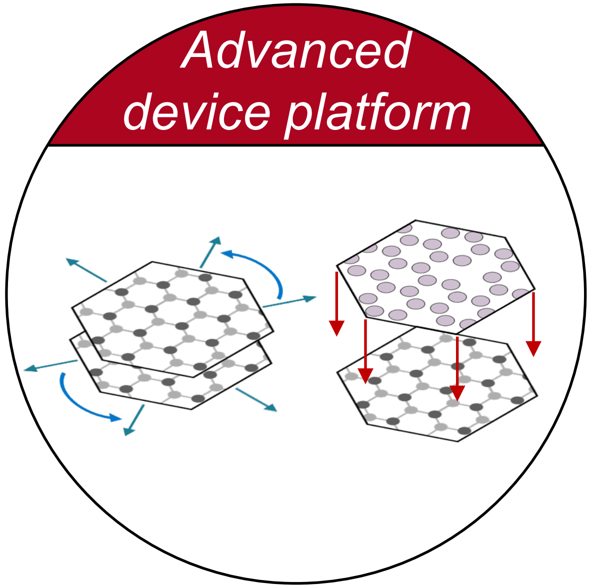


Neuromorphic computing

Quantum technologies

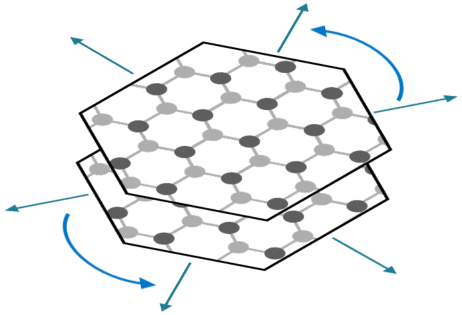


Advanced device platform

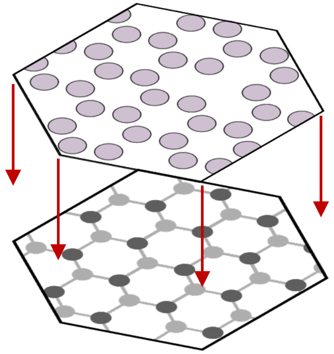


Advanced device platform

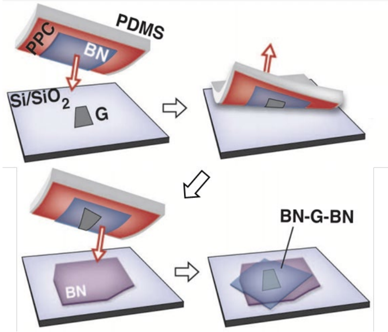
In-situ straintronics & twistronics



Programmable metasurfaces

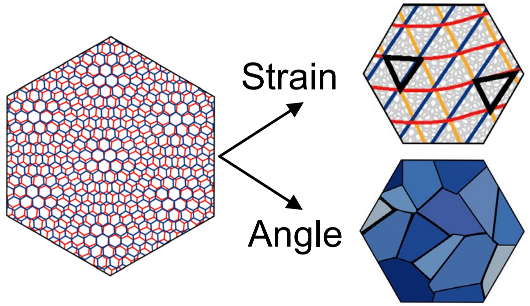


“Fixed” once stacked



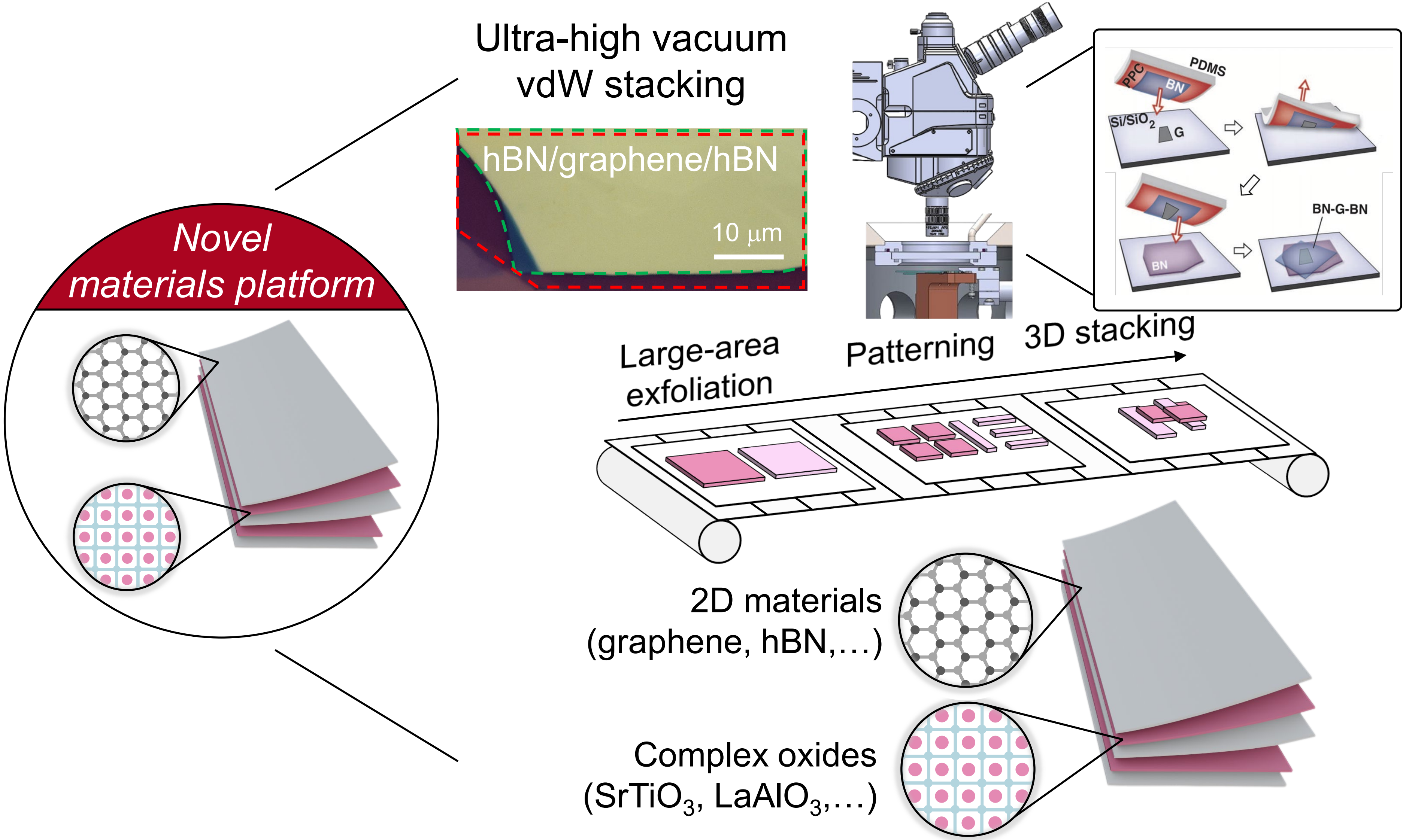
L. Wang et al., Science 342, 614–617 (2013).

Spatial inhomogeneity



C. N. Lau et al., Nature 602, 41–50 (2022).

Novel materials platform



Novel materials platform

Ultra-high vacuum vdW stacking

hBN/graphene/hBN

10 μm

Large-area exfoliation

Patterning

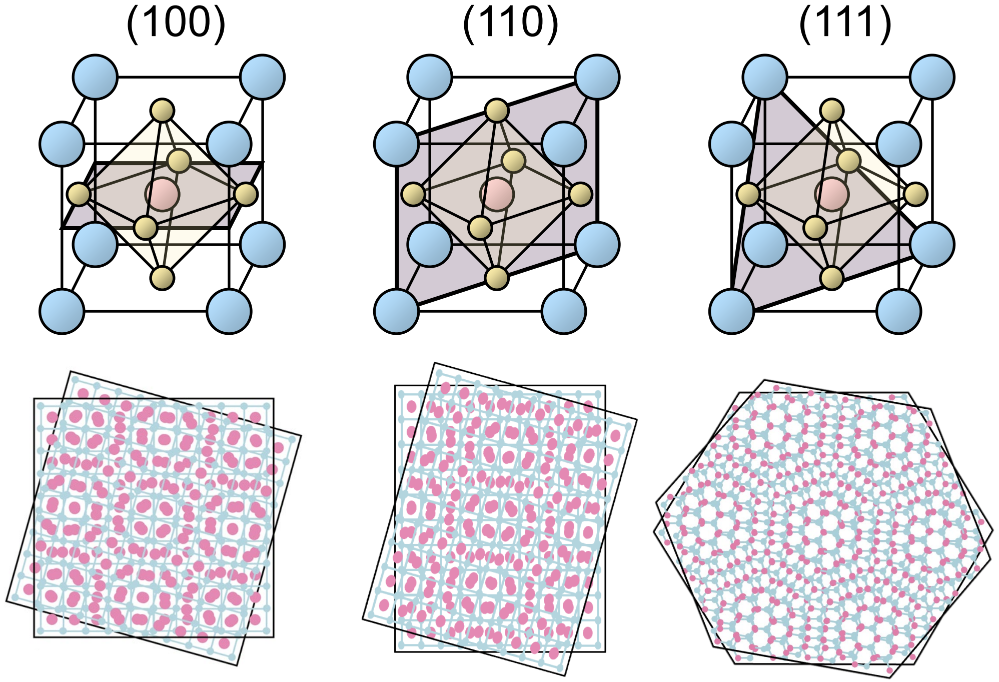
3D stacking

2D materials (graphene, hBN,…)

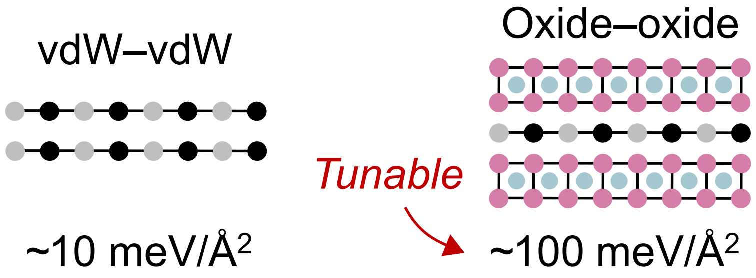
Complex oxides ($SrTiO_3$, $LaAlO_3$,…)

Novel moiré systems

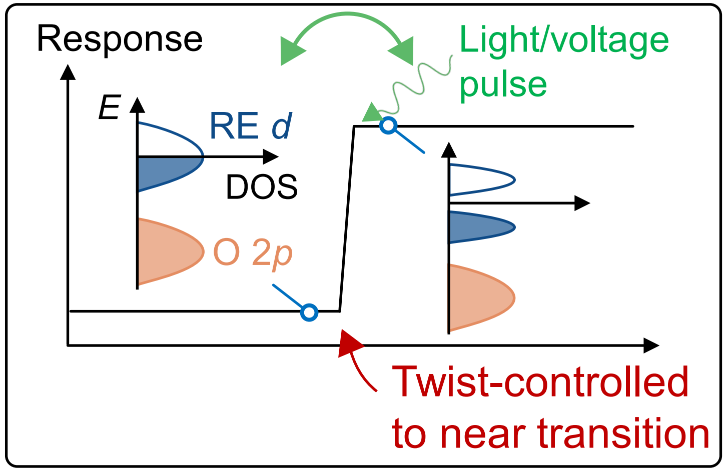
New moiré symmetries



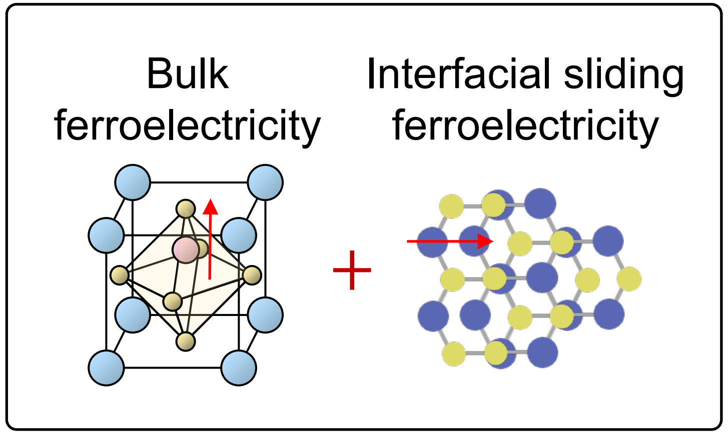
Stronger & tunable potential modulation



Reconfigurable twisted Mott memristors

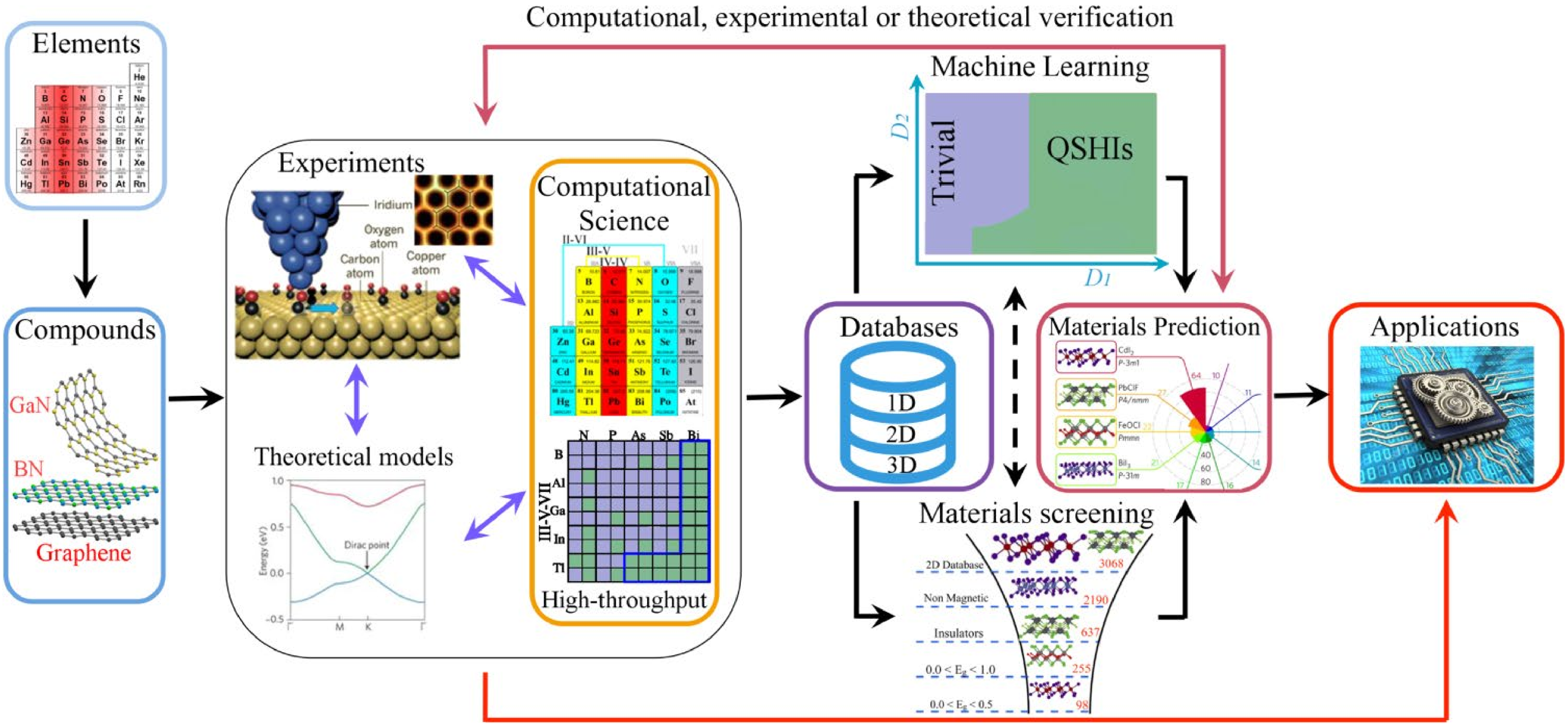


Twisted (multi-)ferroics



AI–Driven Materials Discovery and Design
Patrick Lohr

Accelerating Materials Design



Computational, experimental or theoretical verification

Elements

Compounds

Experiments

Theoretical models

Computational Science

High-throughput

Databases

Machine Learning

Materials Prediction

Materials screening

Applications

Schleder, G. R. *et al.* From DFT to machine learning: recent approaches to materials science–a review. *J. Phys. Mater.* **2**, 032001 (2019)

The Materials Project Database

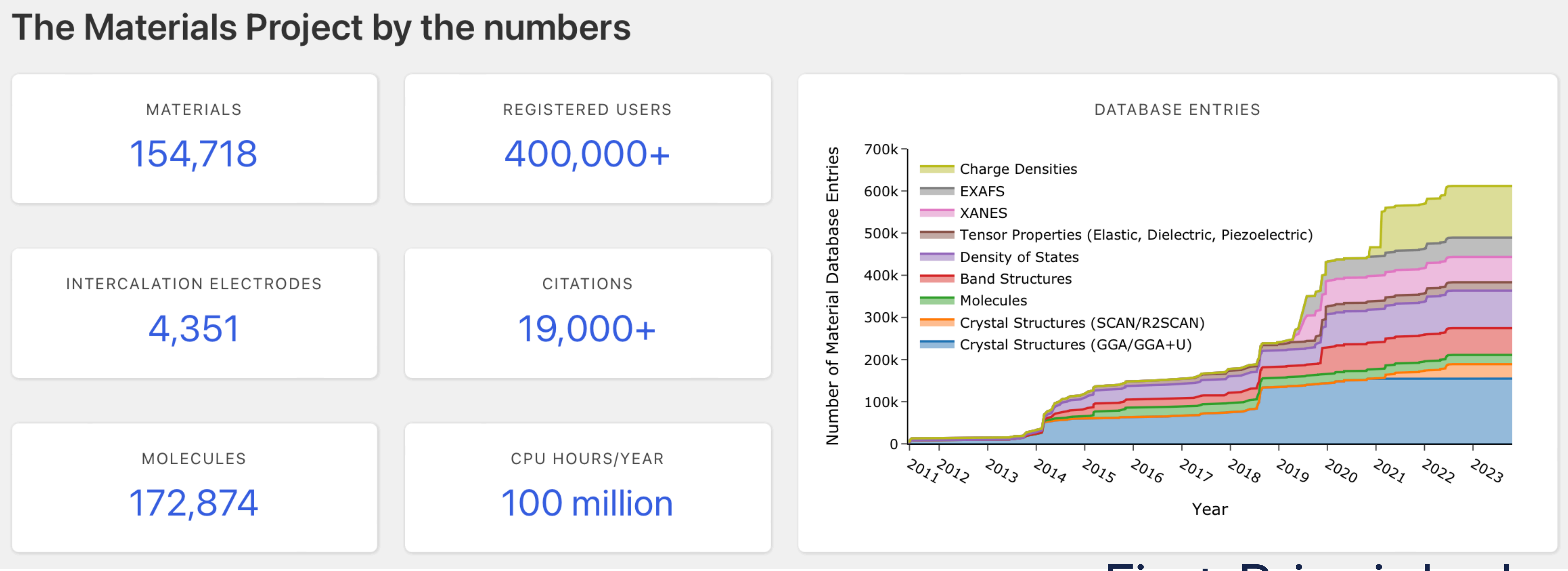
MATERIALS PROJECT

BERKELEY LAB

Lawrence Berkeley
National Laboratory

https://next-gen.materialsproject.org

The Materials Project Database



The Materials Project by the numbers

MATERIALS
154,718

REGISTERED USERS
400,000+

INTERCALATION ELECTRODES
4,351

CITATIONS
19,000+

MOLECULES
172,874

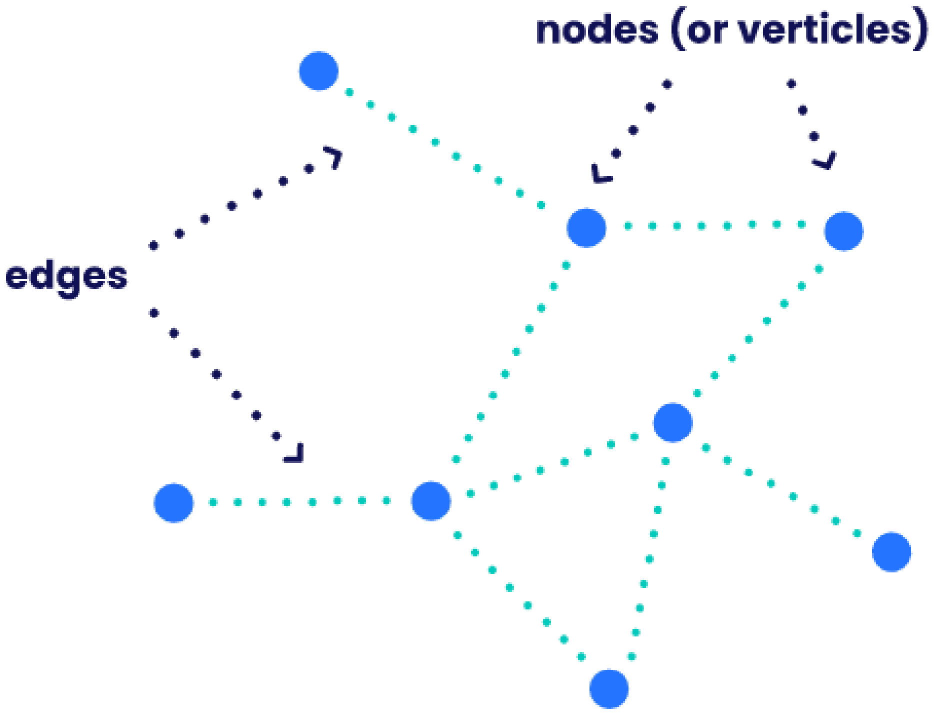
CPU HOURS/YEAR
100 million

DATABASE ENTRIES

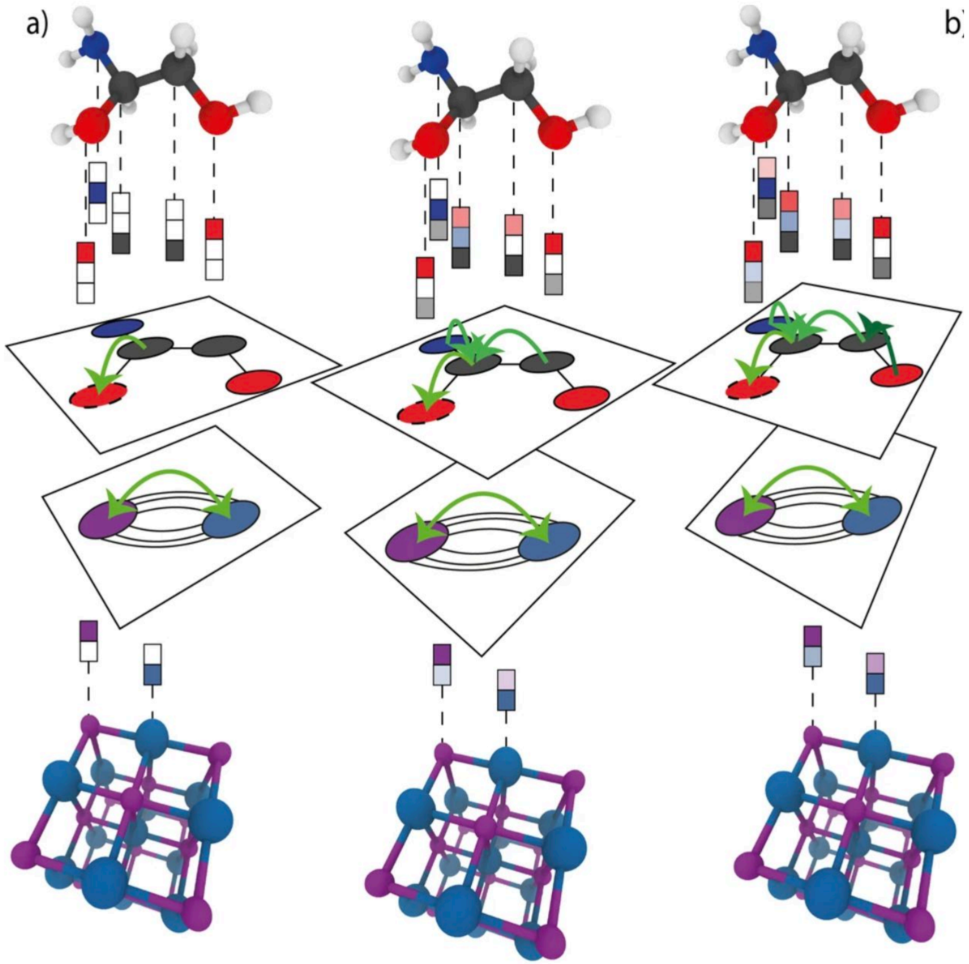
First-Principles!

https://next-gen.materialsproject.org

Graph Neural Networks (GNNs)

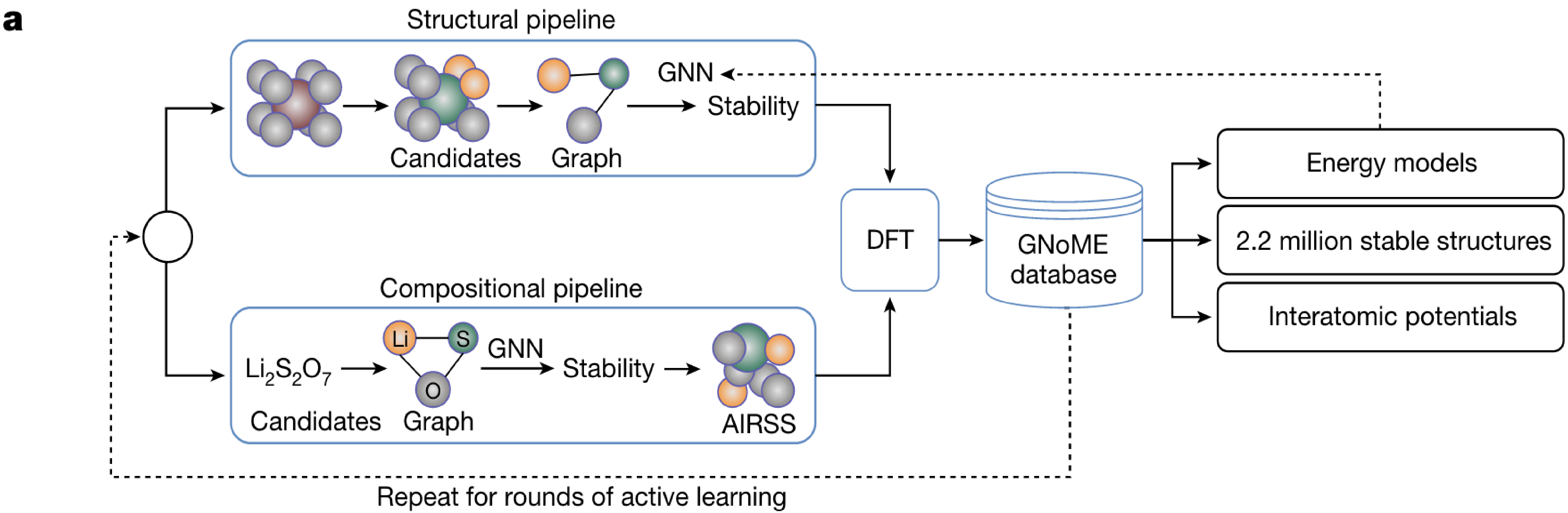


Structure–Property Relationships



Merchant, A. *et al.* Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023)

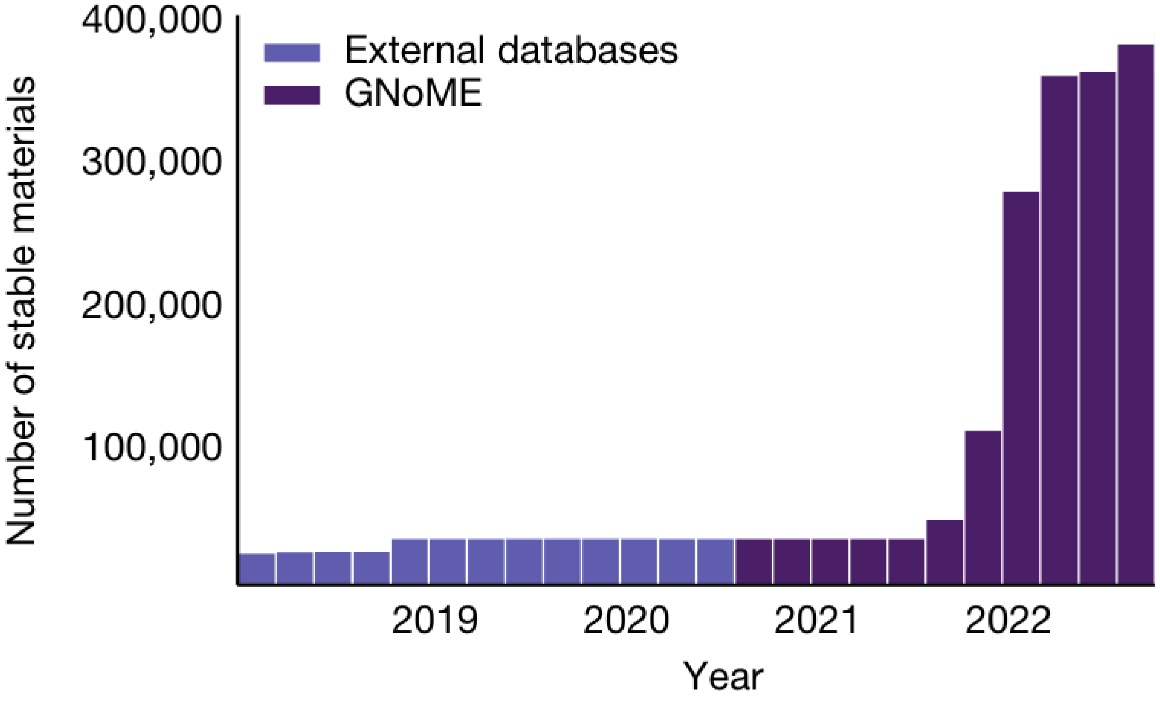
GNNs for Materials Exploration (GNoME)

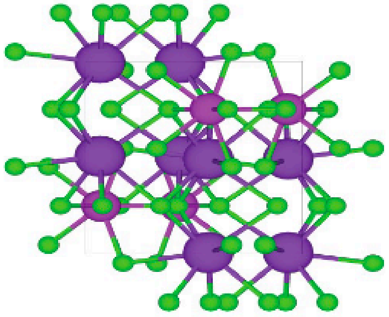


Google DeepMind

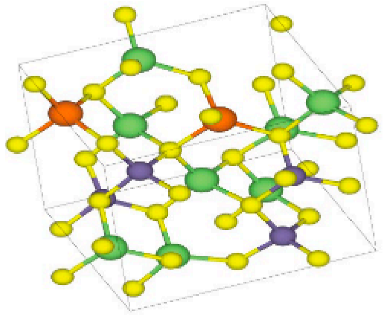
Merchant, A. *et al.* Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023)

Material Stability

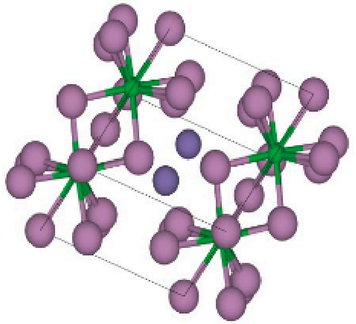




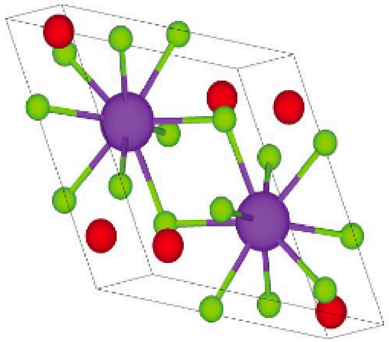
K_2BiCl_5



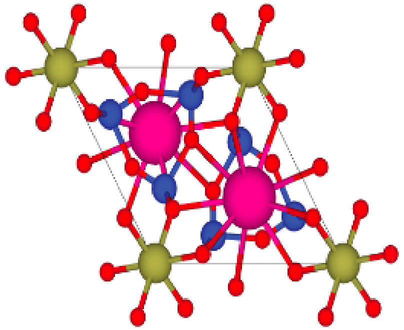
$Li_4MgGe_2S_7$



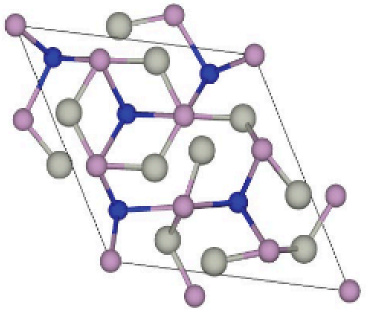
Mo_5GeB_2



KV_3Se_3



$Rb_2HfSi_3O_9$

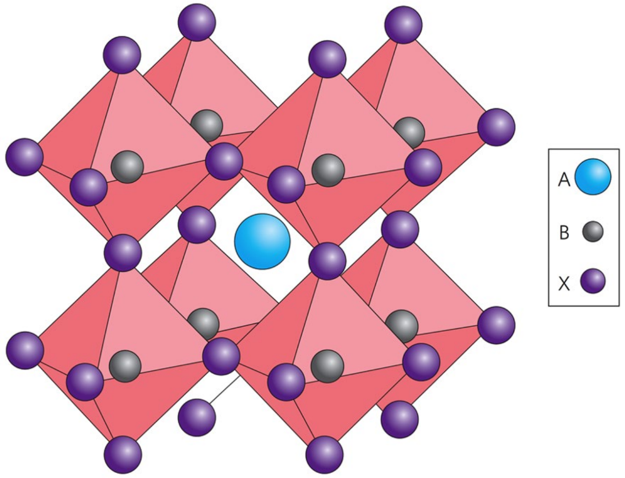


$Tm_5Pd_9P_7$

Merchant, A. *et al.* Scaling deep learning for materials discovery. *Nature* **624**, 80–85 (2023)

Thin-Film Semiconductors (My Work)

Metal Halide Perovskites

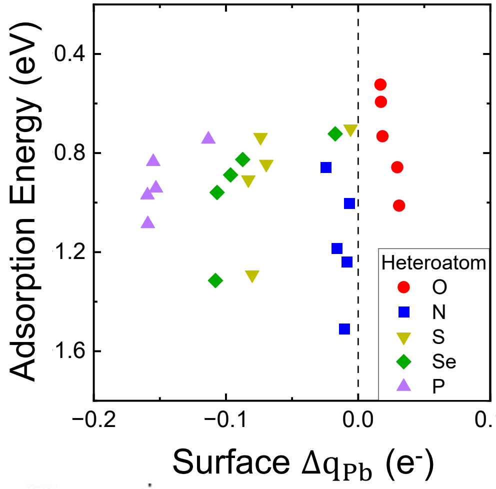


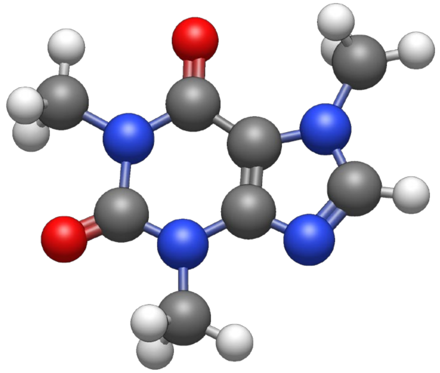
High-Throughput DFT

MATERIALS PROJECT

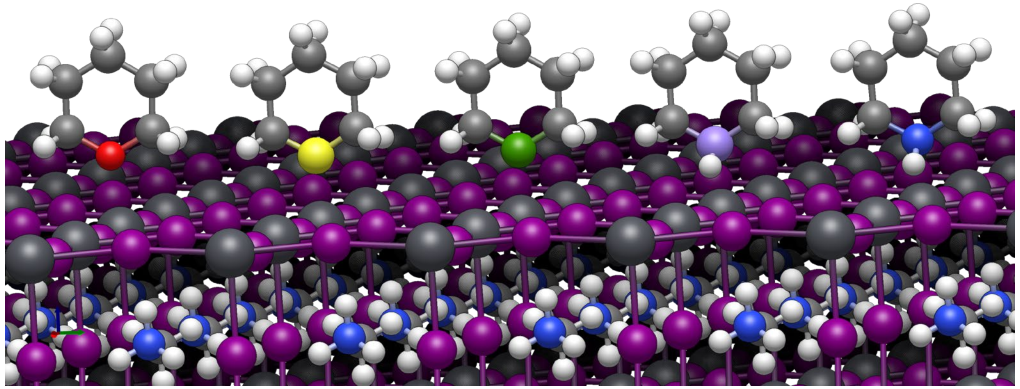
atomate

Surface Properties

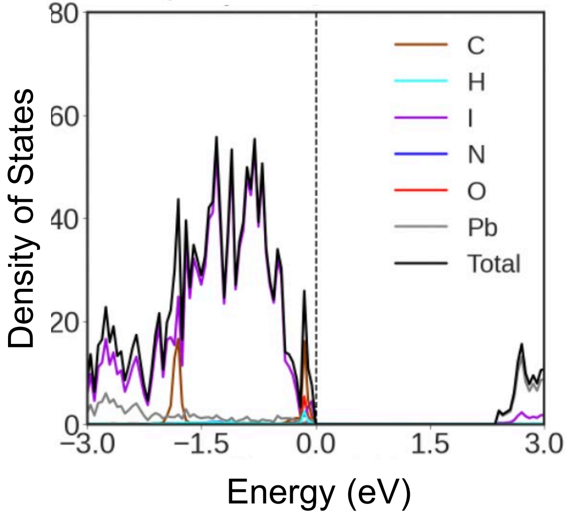




Small Organic Molecules



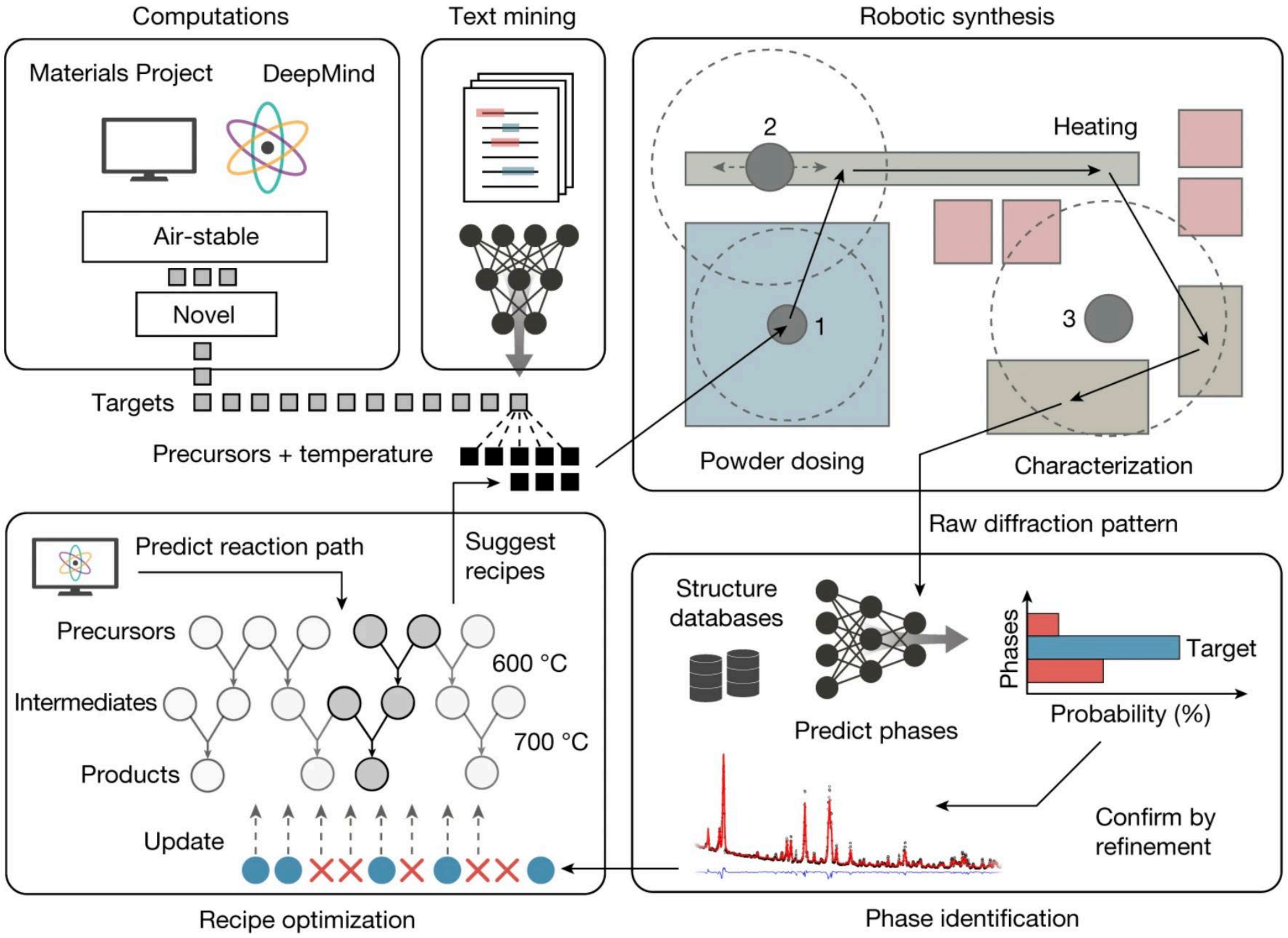
Surface Adsorption



Automated Synthesis (A-LAB)

Szymanski, N. J. *et al.* An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**, 86–91 (2023)

Automated Synthesis (A-LAB)

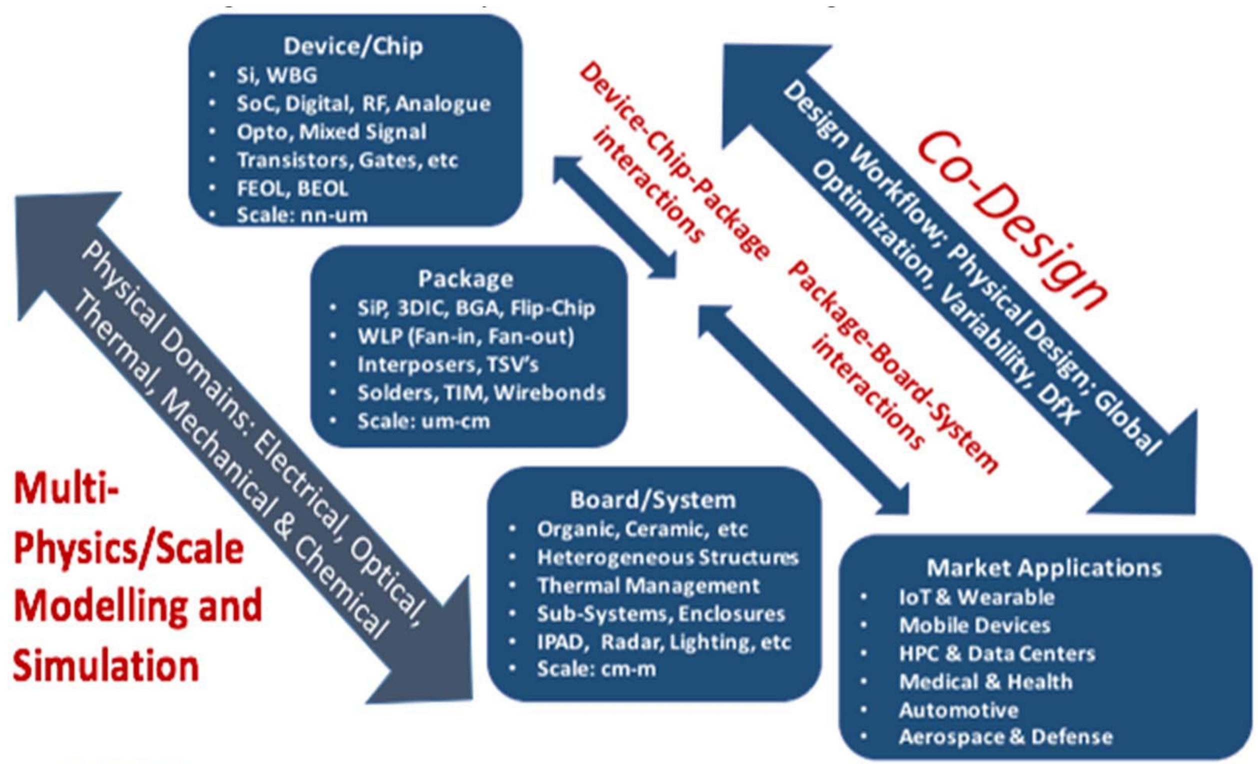


1

2

3

Back to Basics



Device/Chip
- Si, WBG
- SoC, Digital, RF, Analogue
- Opto, Mixed Signal
- Transistors, Gates, etc
- FEOL, BEOL
- Scale: nn-um

Package
- SiP, 3DIC, BGA, Flip-Chip
- WLP (Fan-in, Fan-out)
- Interposers, TSV's
- Solders, TIM, Wirebonds
- Scale: um-cm

Board/System
- Organic, Ceramic, etc
- Heterogeneous Structures
- Thermal Management
- Sub-Systems, Enclosures
- IPAD, Radar, Lighting, etc
- Scale: cm-m

Market Applications
- IoT & Wearable
- Mobile Devices
- HPC & Data Centers
- Medical & Health
- Automotive
- Aerospace & Defense

Device-Chip-Package interactions

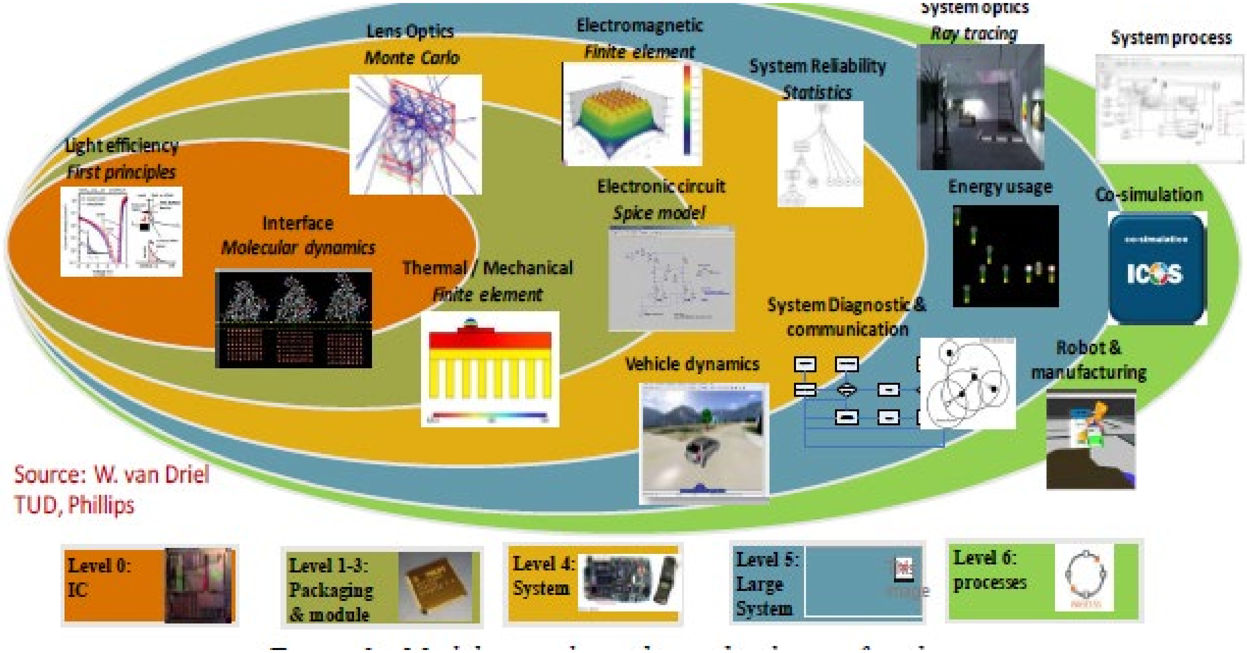
Package-Board-System interactions

Co-Design

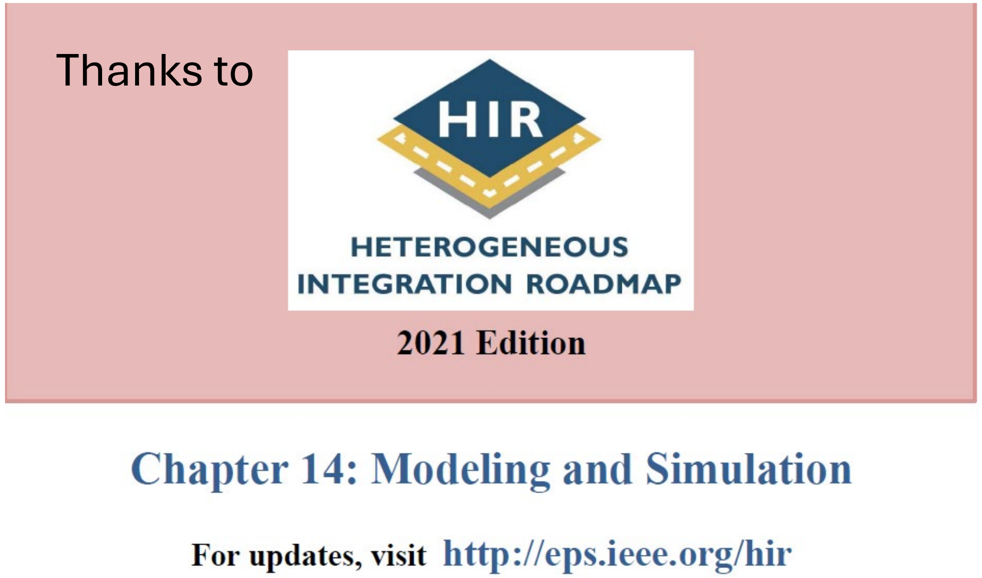
Design Workflow; Physical Design; Global Optimization, Variability, DfX

Physical Domains: Electrical, Optical, Thermal, Mechanical & Chemical

Multi-Physics/Scale Modelling and Simulation



Source: W. van Driel
TUD, Phillips



Thanks to

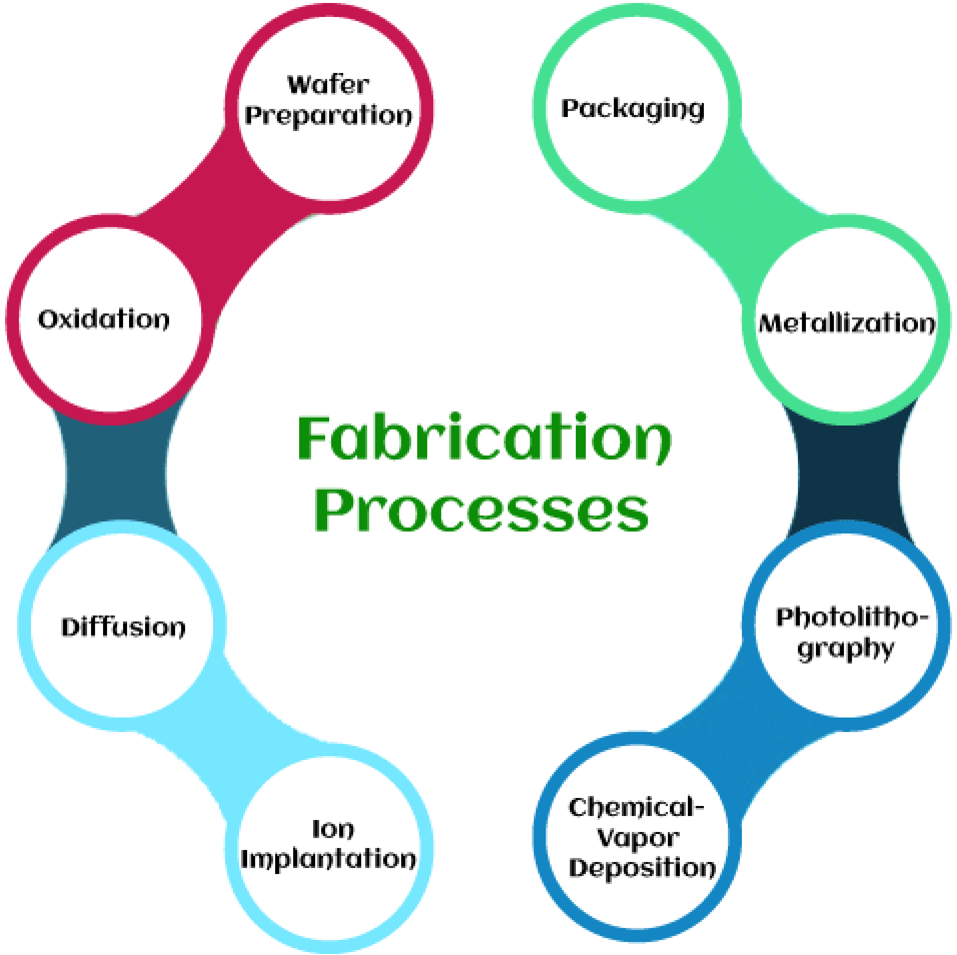
HETEROGENEOUS INTEGRATION ROADMAP

2021 Edition

Chapter 14: Modeling and Simulation

For updates, visit http://eps.ieee.org/hir

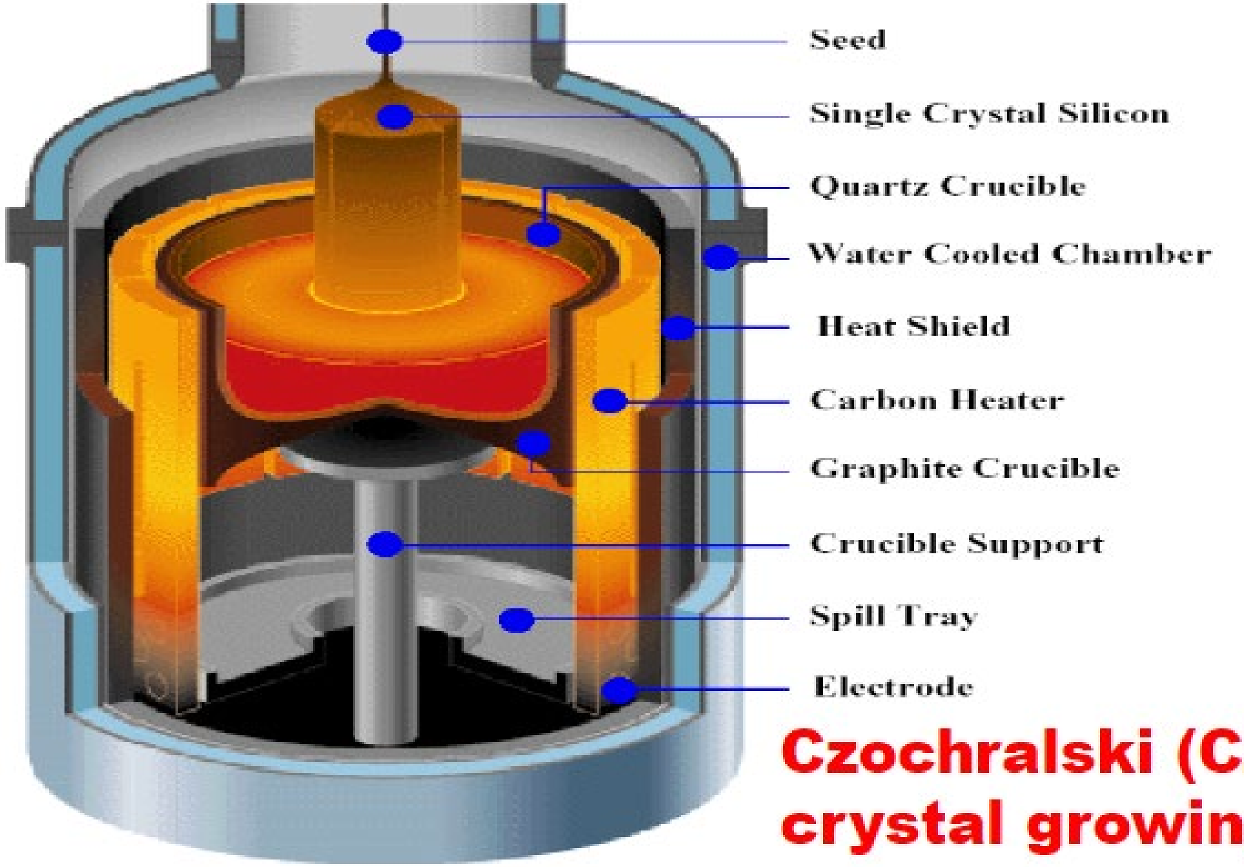
Semiconductor Manufacturing: A deep dive



- Typically several hundred to thousands of steps to complete a modern device!

Wafer Manufacturing

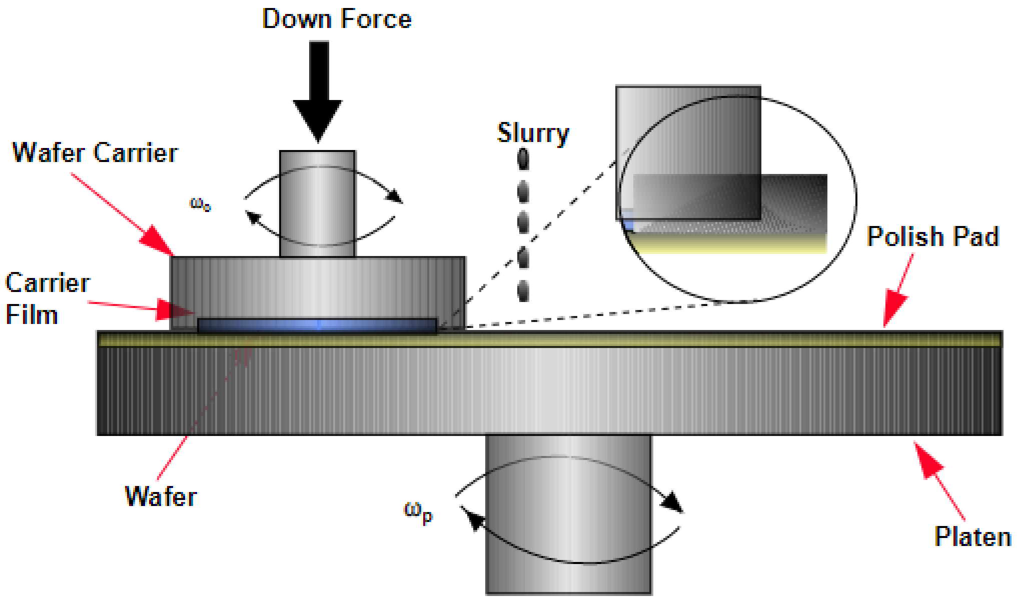
- Czochralski (CZ) Method
  - High purity silicon and selected dopant are melted in large crucible
  - Seed crystal of known crystal orientation is dipped into melt pool
  - Seed and crucible are counter-rotated while seed is slowly drawn upward from melt surface
  - Silicon solidifies into large cylindrical ingot which is sliced into wafers which are polished and processed further



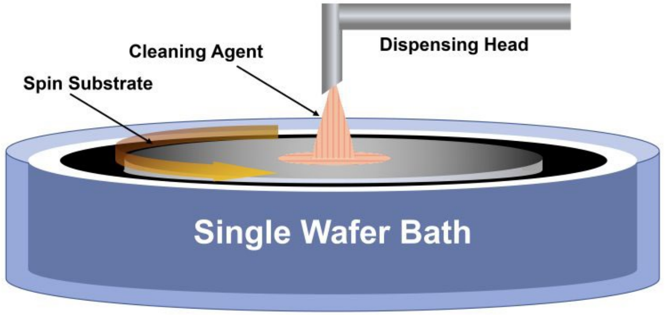
Czochralski (CZ) crystal growing

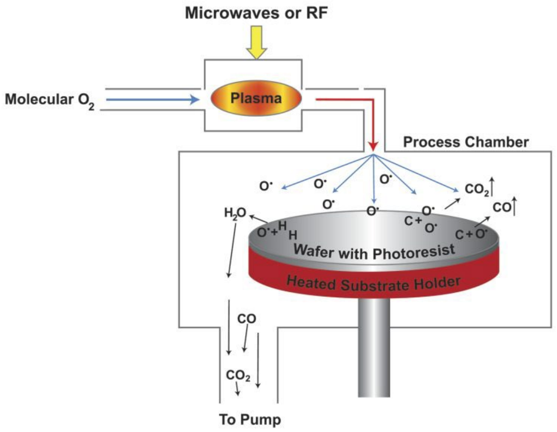
Chemical-Mechanical Planarization (CMP): Need for multiphysics modeling for process optimization

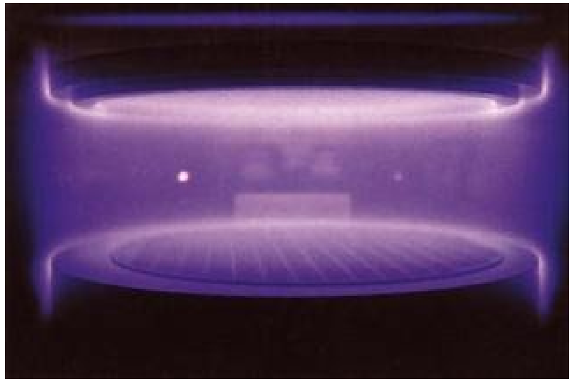
- CMP is used to remove excess material and planarize/polish wafer surface after different processing steps
- Uses both mechanical force (pressure, abrasives) and chemical means (etch chemistry) to remove material
- Wafer is placed into carrier and pressed into polishing pad while slurry (etchant + abrasive) is dispensed onto pad
- Both the wafer carrier and pad are rotated during planarization
- Slurry chemistry and abrasives are tailored to material being removed



Wafer Cleaning: accounts for 30 % of processing steps



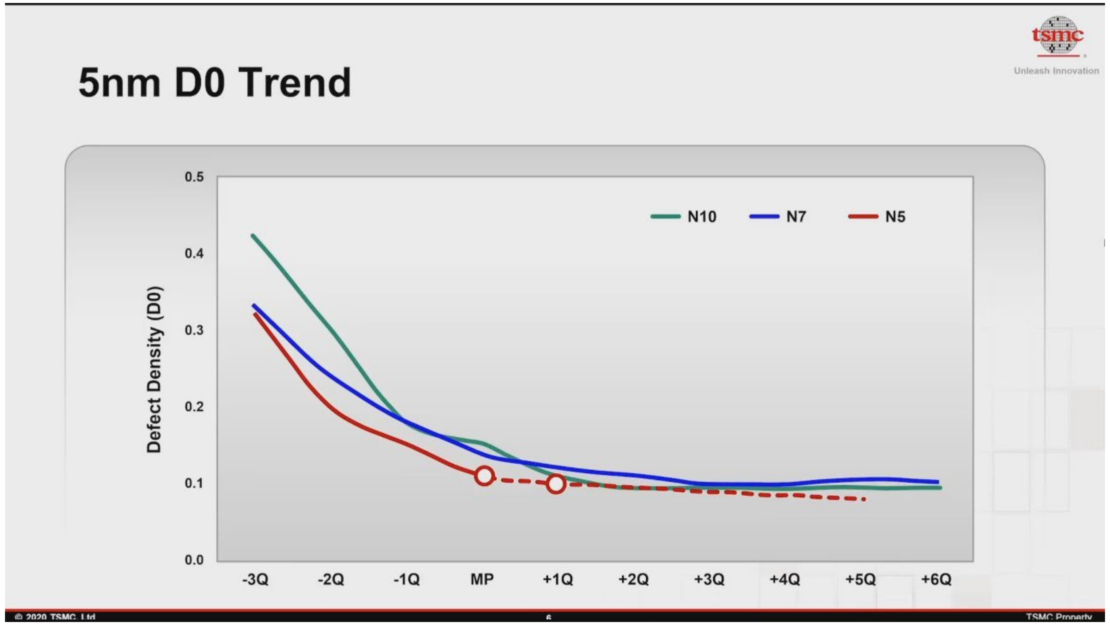




- **<u>high yield (>> 95%) required</u>** for manufacturers to remain competetive
- Cleaning steps used to remove contamination and prepare surface for next step
  - Form native oxide (hydrophilic), Si-H termination (hydrophobic), etc.
- Wet Cleaning
  - Single wafer or batch process using aqueous chemistry
- Dry Cleaning
  - **Plasma-based** clean steps to remove contamination
- Supercritical CO_2 Clean
  - Good for high aspect ratio structures

Wafer Cleaning: accounts for 30 % of processing steps

Defect density D0 (< 0.09 /cm^2)



5nm D0 Trend

© 2020 TSMC, Ltd 6 TSMC Property

AI-leading the way

1. **Advanced Process Control (APC):**
Data-drive real-time monitoring & control

2. **Machine Learning for Predictive Maintenance:**
Downtime minimization via predicting potential failures

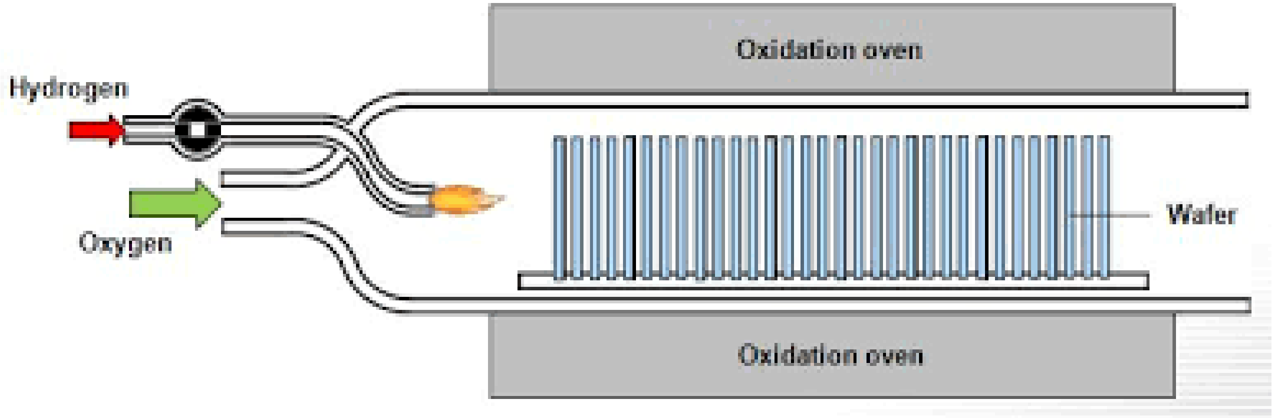
3. **Smart Manufacturing and Industry 4.0:**
Data-driven decisions for operational efficiency

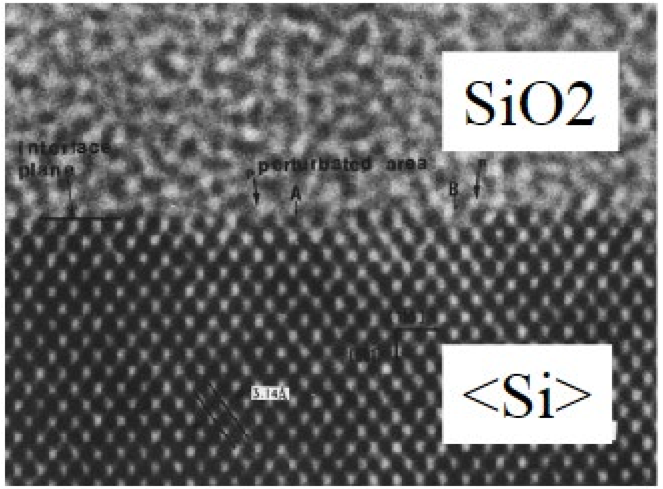
4. **Defect Detection with Computer Vision:**
Identifying defects during manufacturing

5. **Statistical Process Control (SPC):**
maintains consistency and reduces variability

https://www.linkedin.com/pulse/enhancing-semiconductor-yield-innovative-solutions-optimal-performance-gx1ef

Thermal Oxidation

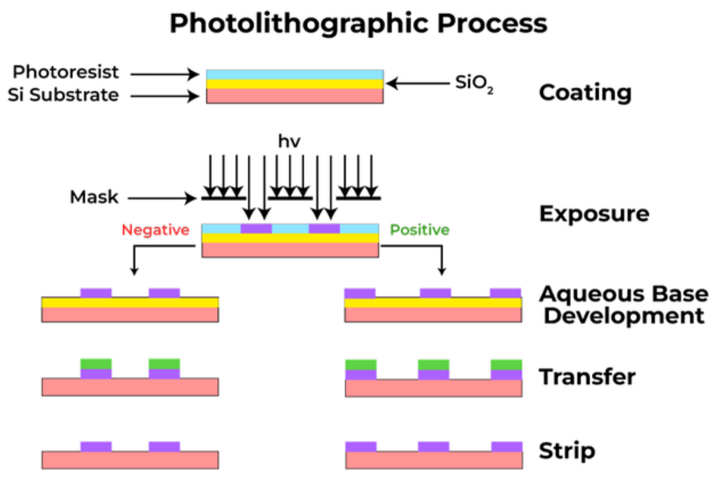


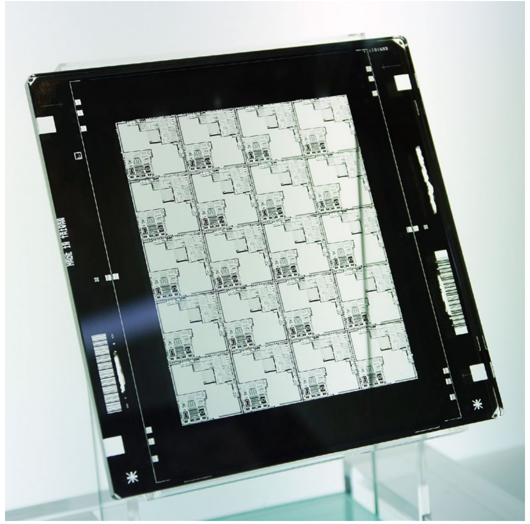


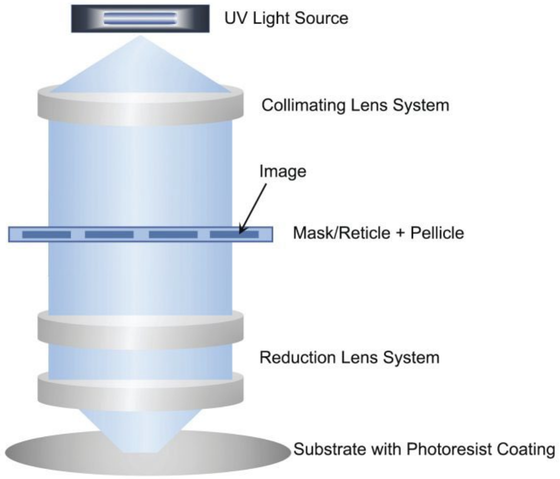
- Forms insulating regions between devices and can protect wafer surface during other processes
- Wafers are loaded into a “boat” and placed into furnace
- Furnace is heated in “dry” (O_2-only) or “wet” (O_2 + steam) conditions and exposed Si oxidizes
- Temperature, time, and conditions are adjusted to achieve desired oxide thickness

Photolithography

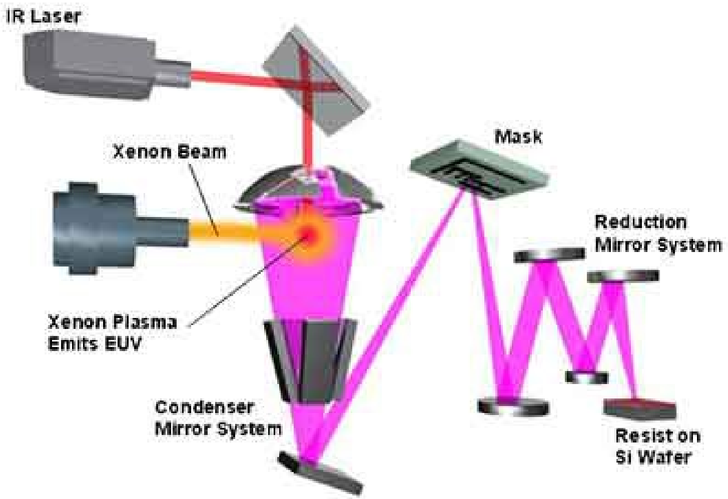
- Wafer is coated in a photoactive film that reacts to particular wavelengths of light
- Wafer is exposed with light passed through a patterned photomask to transfer the image to the photoresist to create a positive or negative image
- Resist layer is “developed” to remove unwanted resist and leave desired image structure to be used in subsequent processing step (etch, deposition, etc.)







Traditional Photolithography



EUV Photolithography

ML in Lithography Hotspot Prediction

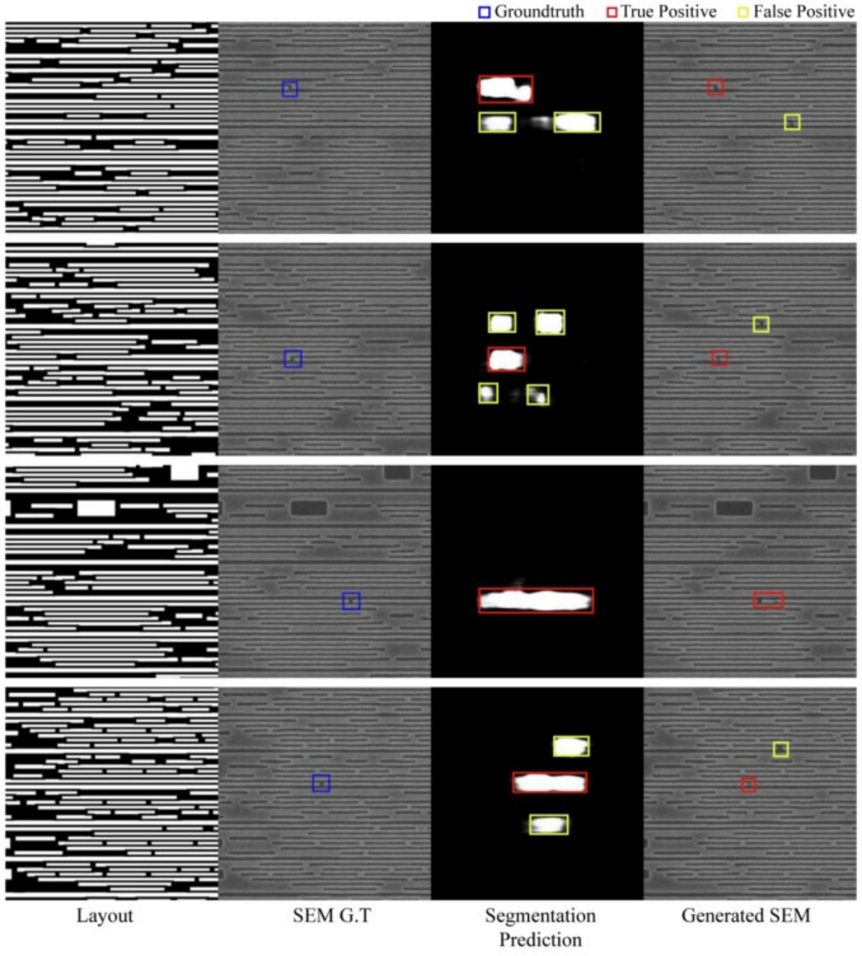
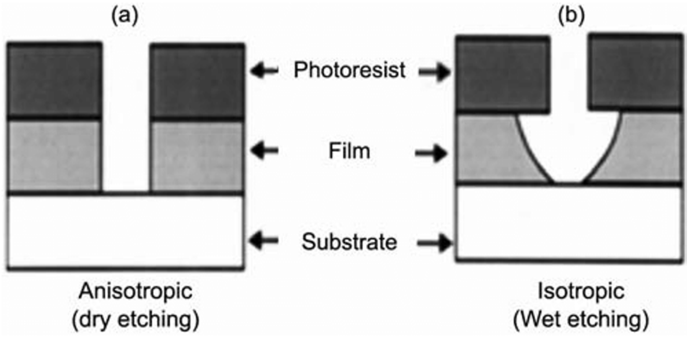


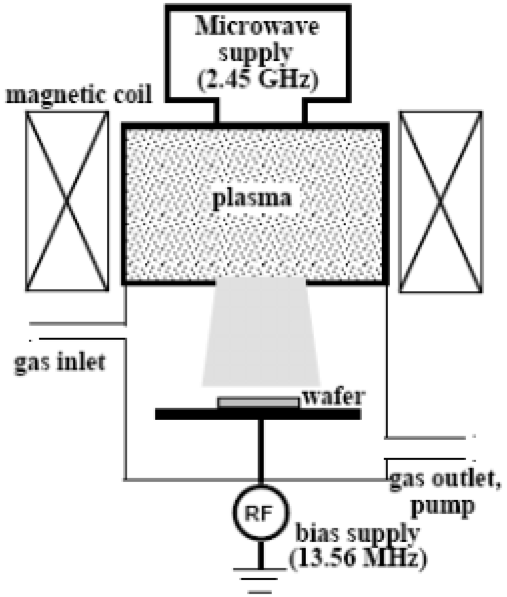
Fig. 5. Example results for the test dataset. Real and predicted hotspots are indicated by blue bounding boxes in SEM G.T images and by red and yellow bounding boxes in generated SEM images which are final outputs of model, respectively. G.T represents ground truth. In segmentation prediction maps, yellow bounding boxes represent false positives (potential hotspot regions) in segmentation prediction map.

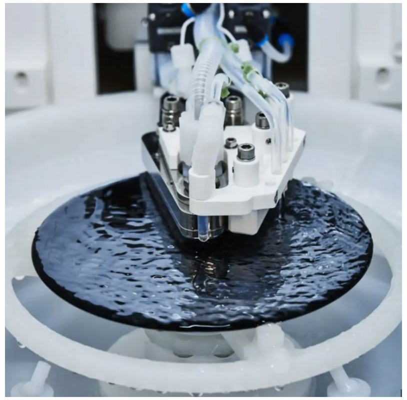
Kim, Jaehoon, et al. "Hotspot Prediction: SEM Image Generation With Potential Lithography Hotspots." *IEEE Transactions on Semiconductor Manufacturing* (2023).

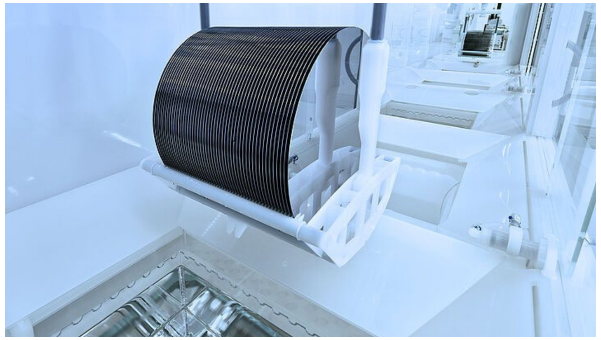
- High feature density can lead to defects in lithography, or "hotspots", that can transfer to subsequent process steps
- ML algorithms takes in design level layout diagrams and corresponding SEM images and learns to predict where real hotspots or potential hotspots will occur and generate a predicted SEM image

Etching









- Material is removed from areas of a wafer using “wet” or “dry” etching methods
- **Wet etching** introduces a wafer to a chemical solution formulated to attack the desired material
  - Isotropic in non-crystalline materials
- **Dry etching utilizes plasma chemistry to attack a specific material** (allows for directional etching)
  - Anisotropic etching, good for small features
- Prior lithography steps allow for specific patterns to be etched into layers on the wafer

Finite Element Analysis in Wet Etch Processing

Derek Bassett et al 2015 ECS Trans. 69 159

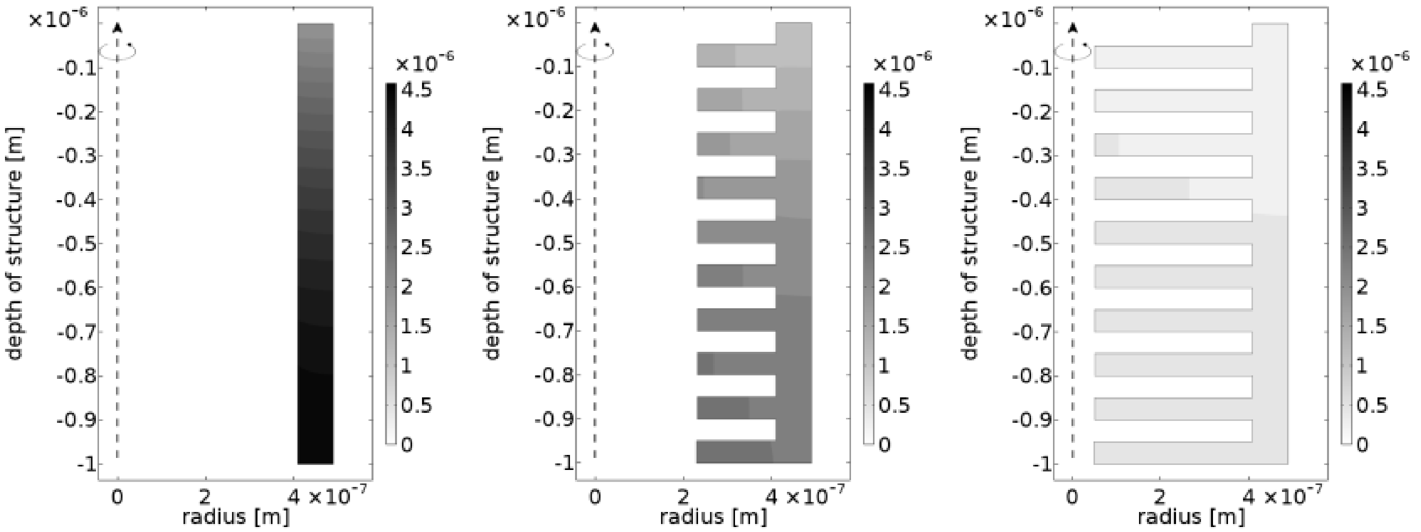


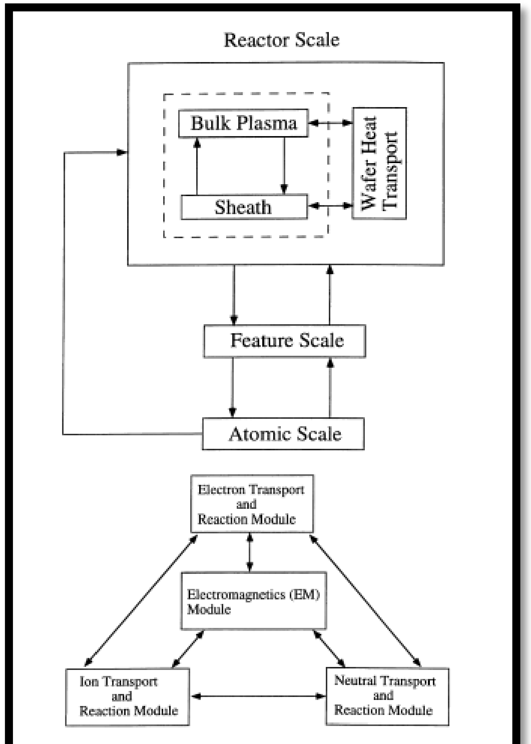
Figure 6: Etch profiles for simulations with varying outer liquid flow velocity showing mole fraction of $Si(OH)_4$. The left image is at t = 0 s, the middle image is at t = 830 s, and the right image is at t = 1660 s when etching of the SiN is complete. This is an example case for all the SiN layers etching at the same rate, indicating a reaction-limited etching process.

- FEM analysis allows for predicting etch profile progression and concentration profiles of byproduct under various conditions
- **Predictability can help avoid structural defects** and make the process engineer’s life much easier

Modeling Plasmas & Etch Equipment (Monica Titus)

Hybrid models are utilized to capture macro-scale plasma behavior in equipment and feature-scale behavior at plasma-surface boundaries

Example of hybrid-model structure

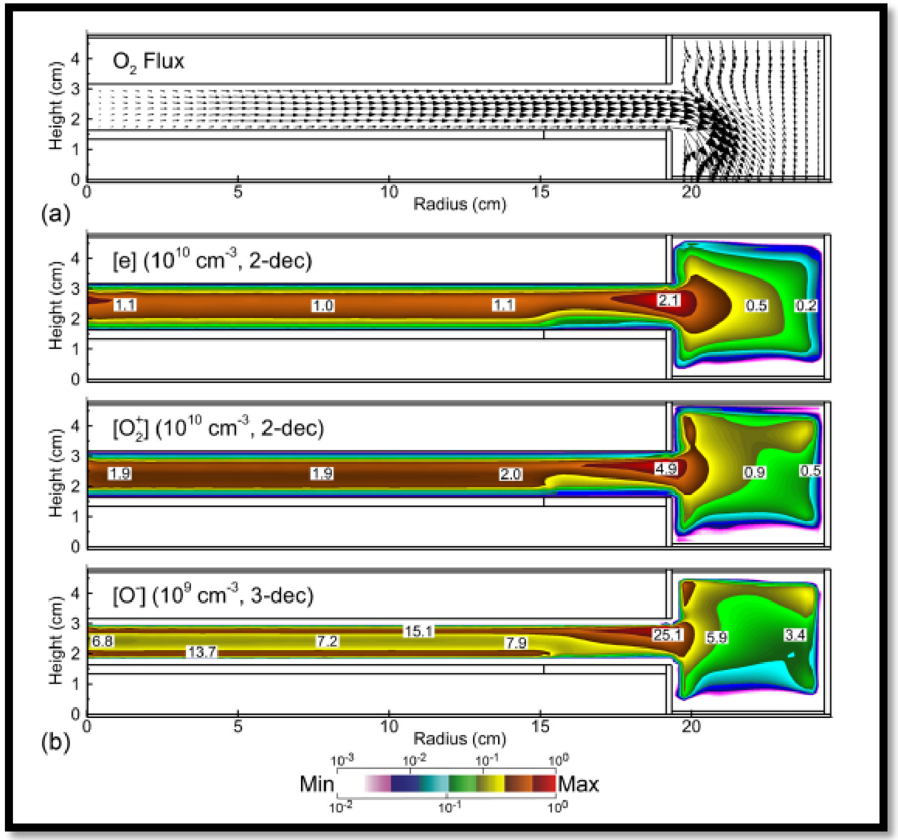


Reactor Scale
Bulk Plasma
Sheath
Wafer Heat Transport
Feature Scale
Atomic Scale
Electron Transport and Reaction Module
Electromagnetics (EM) Module
Ion Transport and Reaction Module
Neutral Transport and Reaction Module

Fig. 6. A methodology for modeling plasma reactors by braking the problem into smaller parts. (Top) From the reactor, to the feature, to the atomic scale. (Bottom) Modules used for reactor scale model.

https://www.chee.uh.edu/sites/chbe/files/faculty/economou/tsf_00_review.pdf

Example of 2D global plasma model & plasma specie densities



https://cpseg.eecs.umich.edu/pub/articles/JVSTA_39_052403_2021.pdf

Example of Energy & Angle Distributions from Bulk Plasma

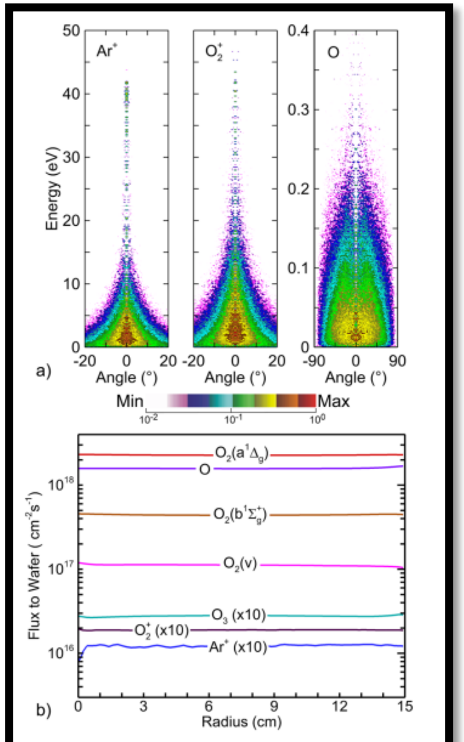


FIG. 3. Plasma properties during the oxidation step for the base case conditions (500 W, 10 MHz, 2 Torr, Ar/O_2 = 4/1 at 5 slm). (a) EADs for Ar^+, O_2^+, and O atoms. (b) Fluxes of reactants to the wafer. The noise in the flux of Ar^+ is a result of the statistics of the Monte Carlo simulation used for secondary electrons.

https://cpseg.eecs.umich.edu/pub/articles/JVSTA_39_052403_2021.pdf

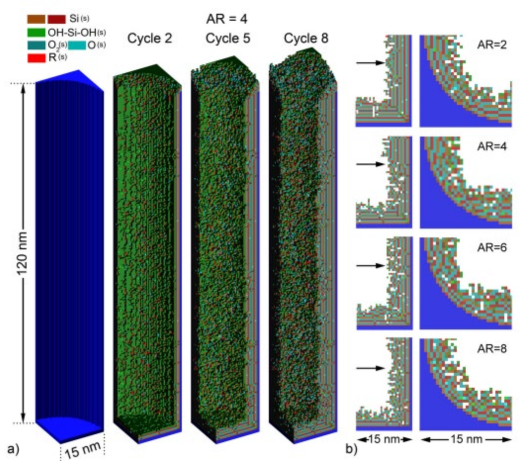
Plasma & hardware simulation results help process engineers estimate properties that affect feature profiles

- Reaction probability
- Incident Angle
- Reflection probability
- Sticking coefficients
- Flux uniformity
- Specie uniformity

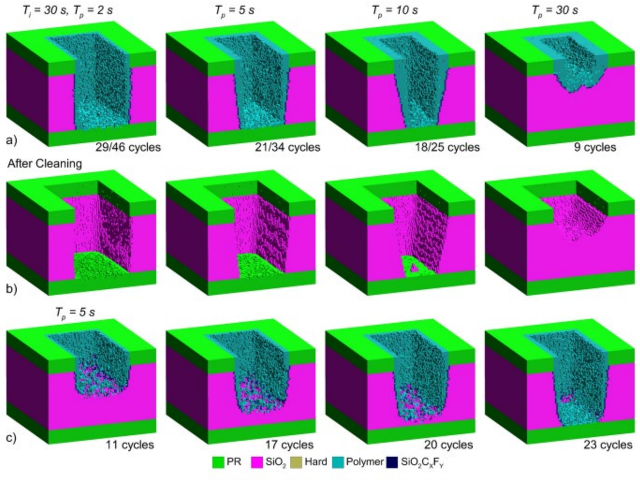
Modeling material evolution (Monica Titus)

Molecular Dynamics (MD) and Hybrid Models illustrate etch, mixing, deposition, adsorption, implantation, and desorption affects, among others, on substrate materials. The generation of mixing layers, contaminants, and surface roughness, help process engineers mitigate undesired effects or target desirable results by appropriate hardware and chemistry selection.

Hybrid Simulations:
Global Plasma Model + Particle-in-Cell Monte Carlo

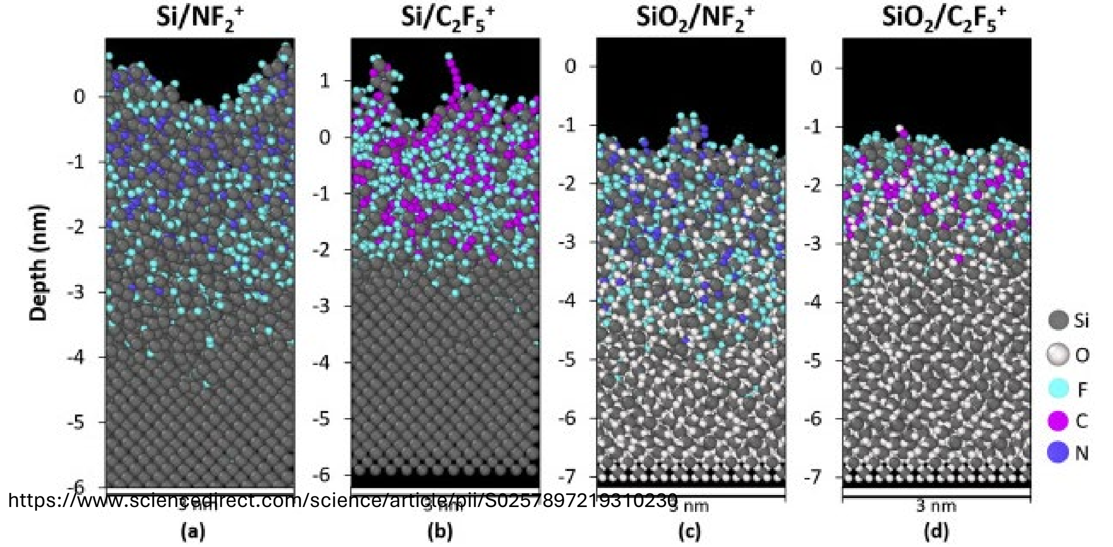


https://cpseg.eecs.umich.edu/pub/articles/JVSTA_39_052403_2021.pdf



https://cpseg.eecs.umich.edu/pub/articles/JVSTA_39_033003_2021.pdf

MD Simulations



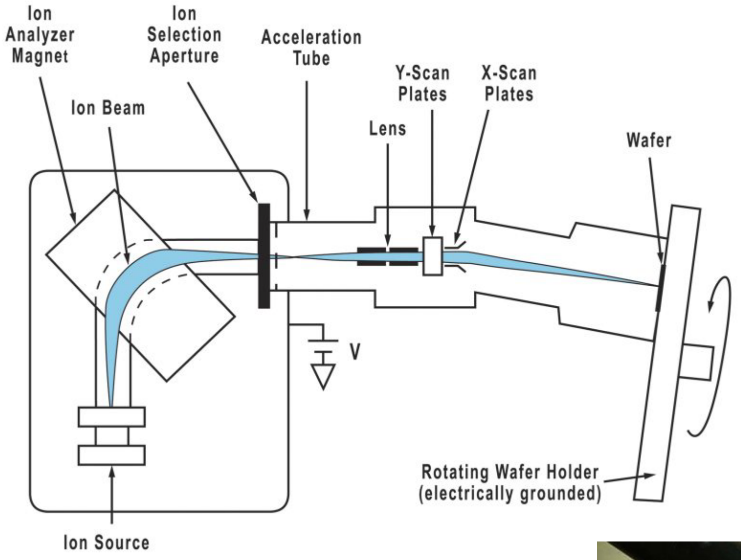
https://www.sciencedirect.com/science/article/pii/S0257897219310230

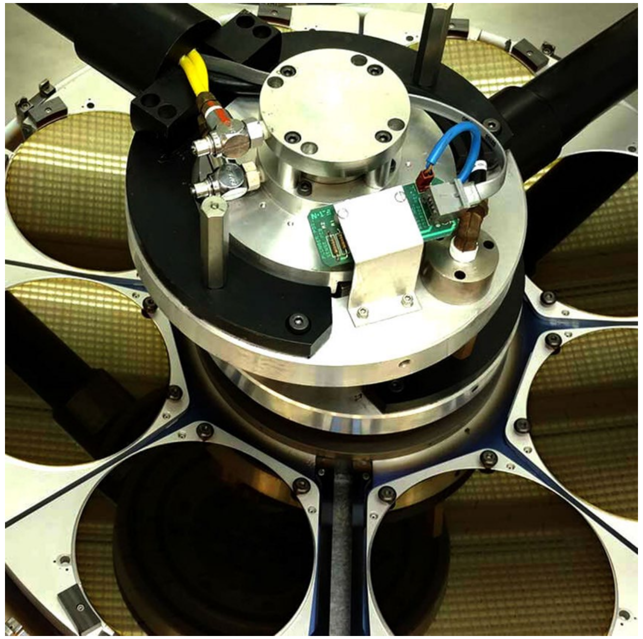
Example of Hybrid simulations incorporating global plasma models with PIC-MC simulations to demonstrate specie distribution, uniformity, and profile evolution during atomic layer deposition and etching conditions.

Example of MD simulation incorporating particle physics and chemistry to demonstrate etch of Silicon under various chemistries and the thickness of the mixing and damage layer that results.

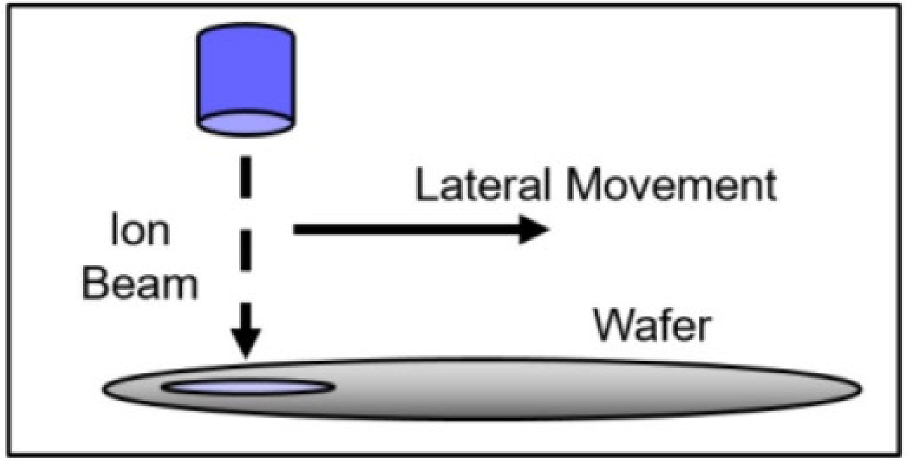
Doping

- Group III or Group V elements are added to silicon to enhance conductivity by introducing excess electrons or holes (forming source/drain regions, n- or p-wells)
- Ion implantation
  - Dopant ions are introduced from an ion source and are accelerated through an E-field toward wafer surface to deposit a dose of dopant atoms
  - Depth profile of dopant determined by current, accelerating voltage, and time
- Diffusion Doping
  - Wafers are loaded into furnace similar to that in oxidation and dopant gases are introduced to chamber
  - Dopant diffuses into exposed areas under high temperature conditions

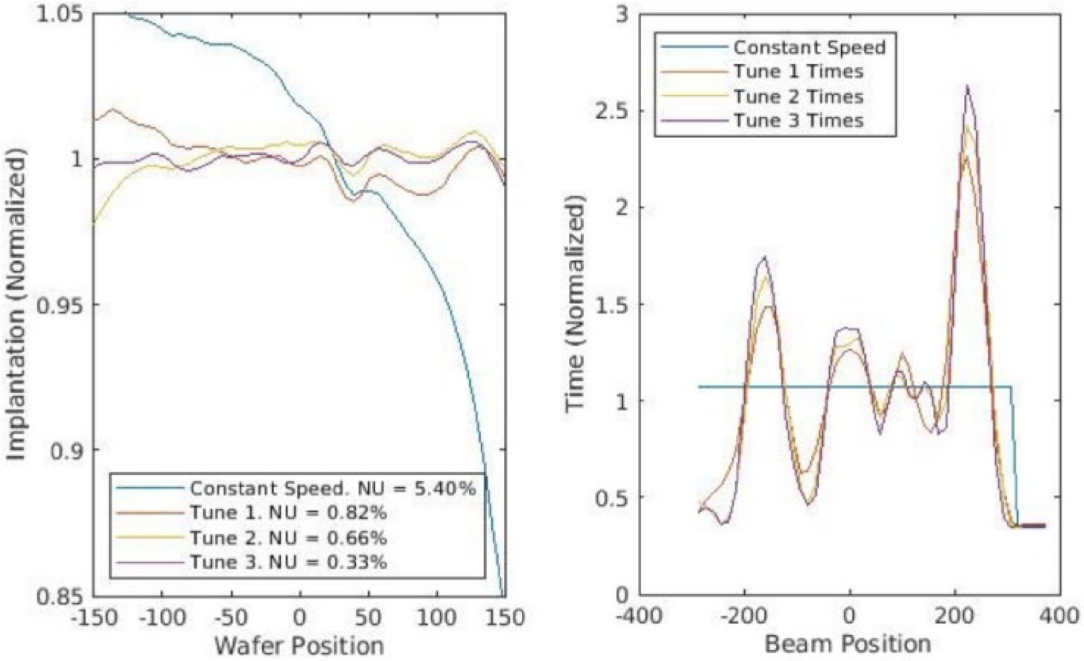




Ion Implantation Optimization via Machine Learning

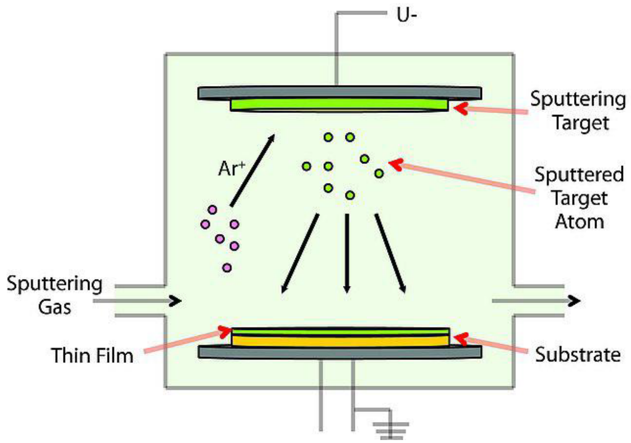


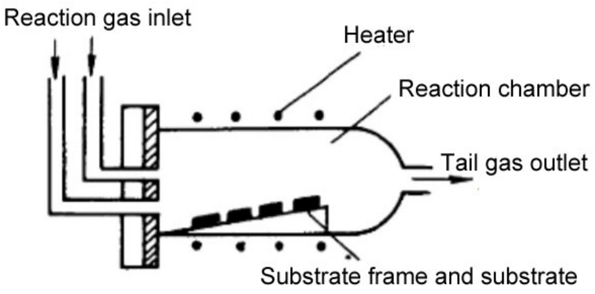
Lang, Christopher I., et al. "Intelligent Optimization of Dosing Uniformity in Ion Implantation Systems." *IEEE Transactions on Semiconductor Manufacturing* 35.3 (2022): 580-584.

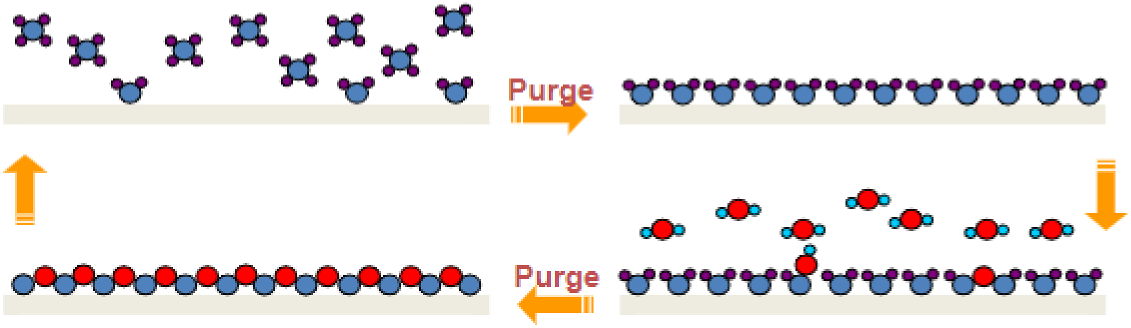


- Using ML to account for beam shape and intensity and ensure uniform spatial dosage
- ML increased uniformity include spatial variation of implant time across wafer

Deposition

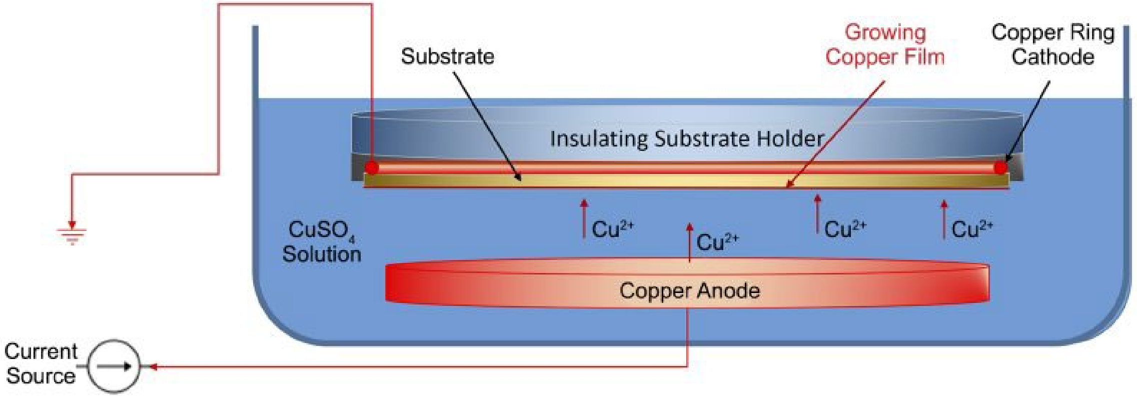


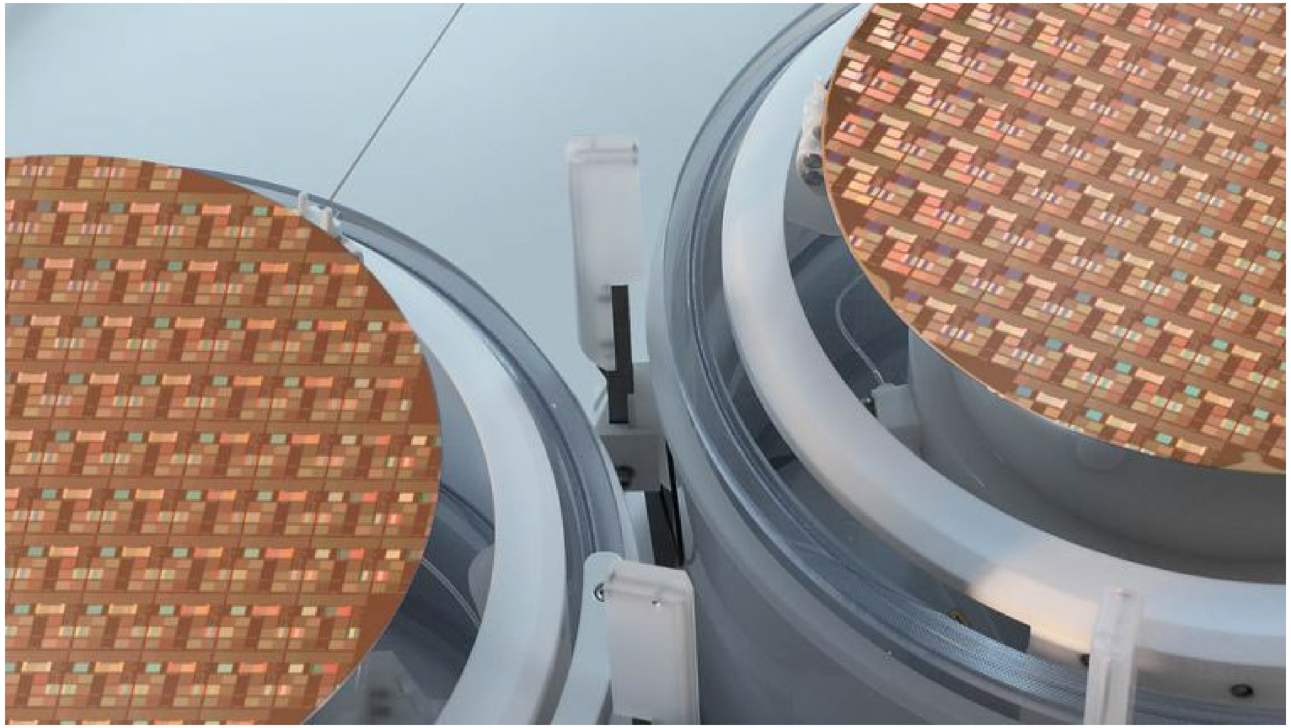




- Physical Vapor Deposition
  - A solid source of material is thermally evaporated or sputtered and the resulting vapor condenses on the wafer to form a thin film
- Chemical Vapor Deposition
  - Wafers in a reaction chamber are introduced to precursor gases that react and deposit a solid film on the surface
- Atomic Layer Deposition
  - Special-case cyclic CVD process using typically two precursors
  - Precursor A introduced to form monolayer on wafer, then precursor B introduced to form another monolayer on top of first (A,B,A,B…)
  - Used for very thin layers of dielectrics (like high-k gate oxides) and some metals

Metallization

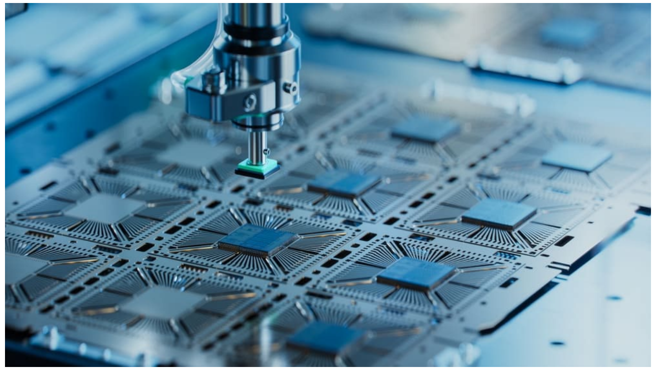


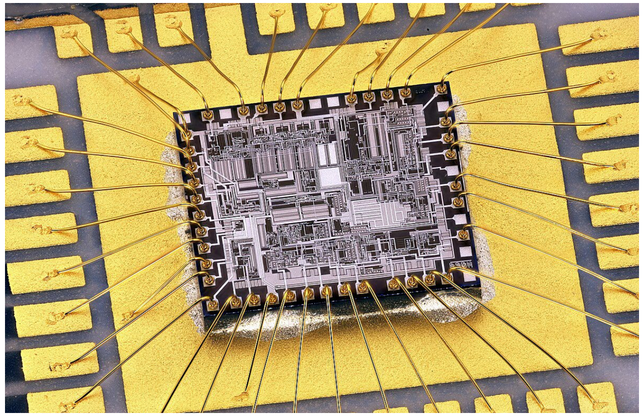


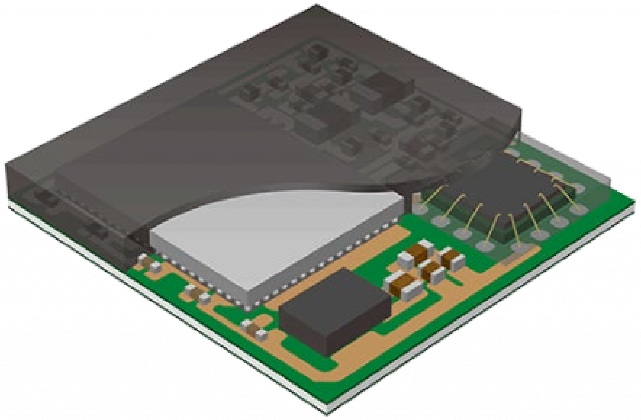
- Electrochemical deposition is used to deposit copper in metal layers on wafer (other metals can also be deposited with ECD)
- Low-k dielectric is deposited onto wafer and patterned to form areas for metal lines between devices
- A barrier layer (TaN/Ta) and a Cu seed layer are deposited, then wafer is introduced to Cu-containing solution
- As current is passed through the electrochemical cell, wafer acts as the cathode and Cu is deposited

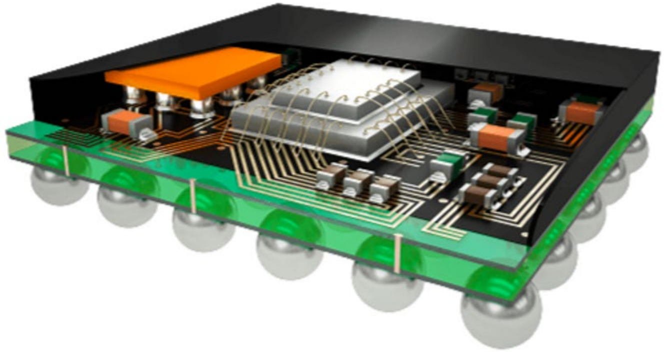
Packaging

- Packaging is important to optimize thermal/electrical/ mechanical environments for given application
- Individual die are cut from wafer and attached to substrate
- Electrical connections between substrate and die are made by wirebonding
- Die is encapsulated and electrical terminations on package are formed
- More sophisticated packages can include multiple stacked die attached together through vias and BGAs or wirebonds (System-in-Package)

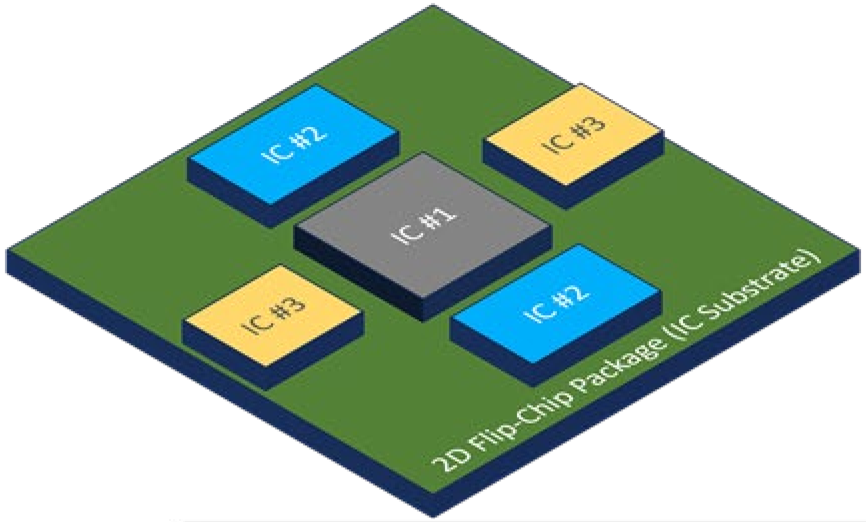


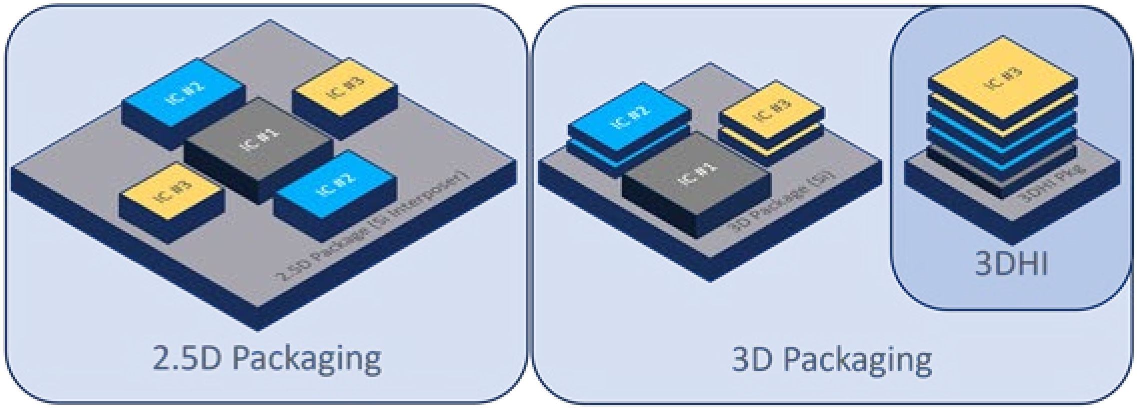


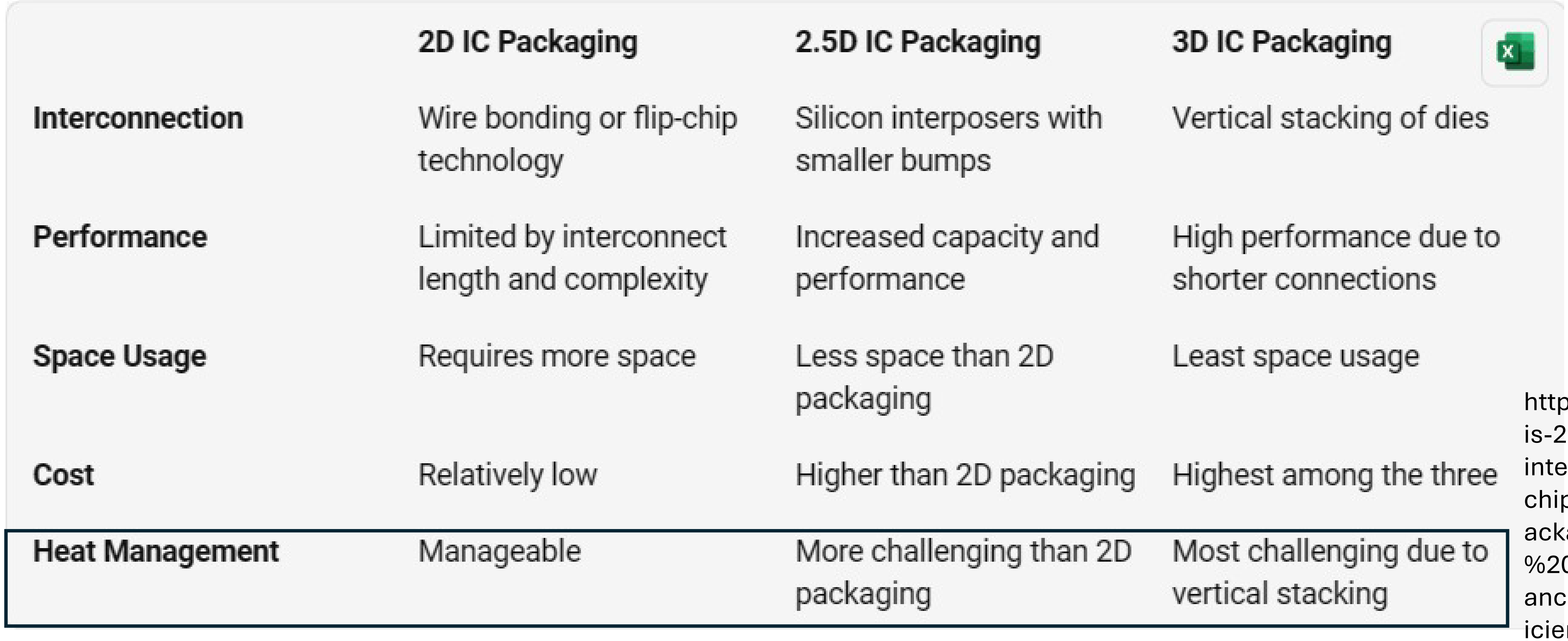




Packaging: from 2D-2.5D-3D



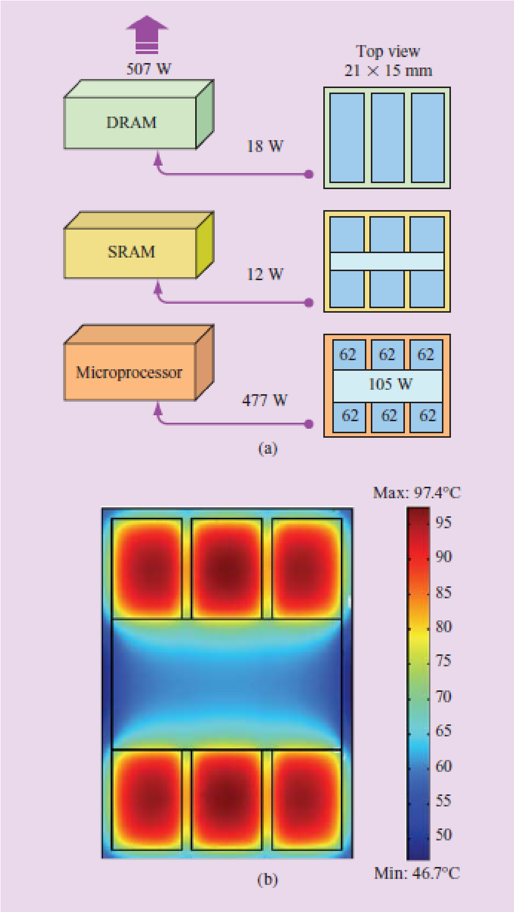




	2D IC Packaging	2.5D IC Packaging	3D IC Packaging
Interconnection	Wire bonding or flip-chip technology	Silicon interposers with smaller bumps	Vertical stacking of dies
Performance	Limited by interconnect length and complexity	Increased capacity and performance	High performance due to shorter connections
Space Usage	Requires more space	Less space than 2D packaging	Least space usage
Cost	Relatively low	Higher than 2D packaging	Highest among the three
Heat Management	Manageable	More challenging than 2D packaging	Most challenging due to vertical stacking

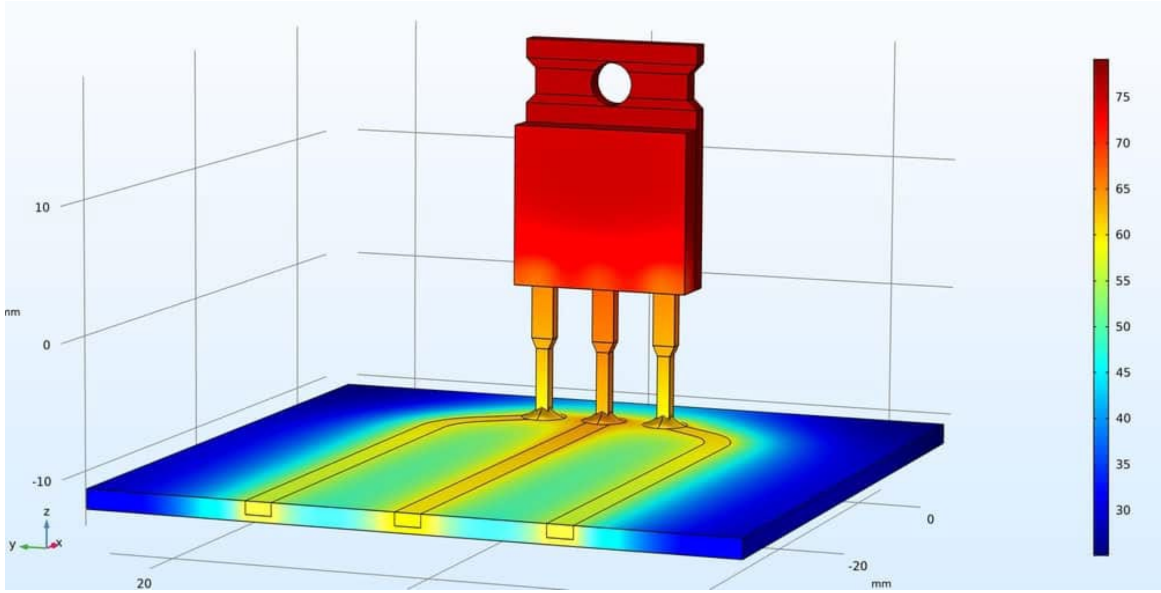
https://techovedas.com/what-is-2d-2-5d-3d-packaging-of-integrated-chips/#:~:text=2D%20IC%20Packaging%3A%20Components%20are,improved%20performance%20and%20power%20efficiency.

Thermo-mechanical models for ‘packaged systems’



FEM model of a 3D stack

IBM J. RES. & DEV. VOL. 52 NO. 6
NOVEMBER 2008

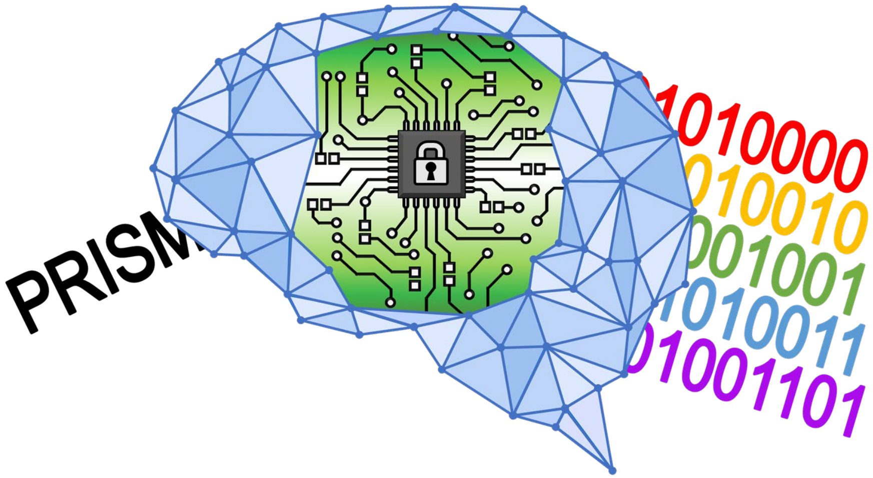


Thermal profile of a ‘single’ transistor (cadence.com)

How do we dissipate the heat?

Hardware Security in the Age of AI and Emerging Semiconductor Devices

Soheil Salehi
Assistant Professor of Electrical and Computer Engineering (ECE), The University of Arizona
Director of Privacy-preserving, Intelligent, and Secure Computing Laboratory (PRISM Lab)
Email: ssalehi@arizona.edu; Website: https://soheilsalehi.com/



THE UNIVERSITY OF ARIZONA

Threats on Hardware Supply Chain and AI Algorithm

❑ Hardware security is questioned due to:

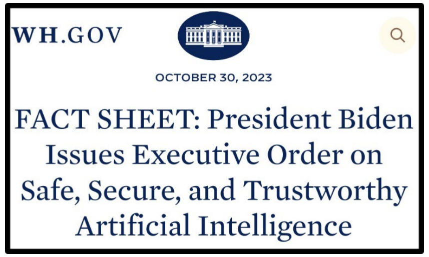
- Emerging attacks and globalized fabrication supply chain.
- Trojan, IP theft, IC Cloning, Counterfeiting, etc.

❑ Vulnerability of Transformer models:

- Dependency on large-scale training datasets
- Vulnerability to adversarial attacks
- Complexity of the hardware accelerator
- Lack of transparency

❑ Threats affecting the integrity of AI accelerator chips:

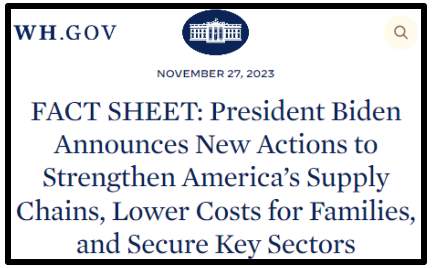
- Side-Channel and Probing Attacks
- Reverse-Engineering Attacks
- Fault Injection and Focused Ion Beam Attacks
- Hardware Trojan Attacks



WH.GOV

OCTOBER 30, 2023

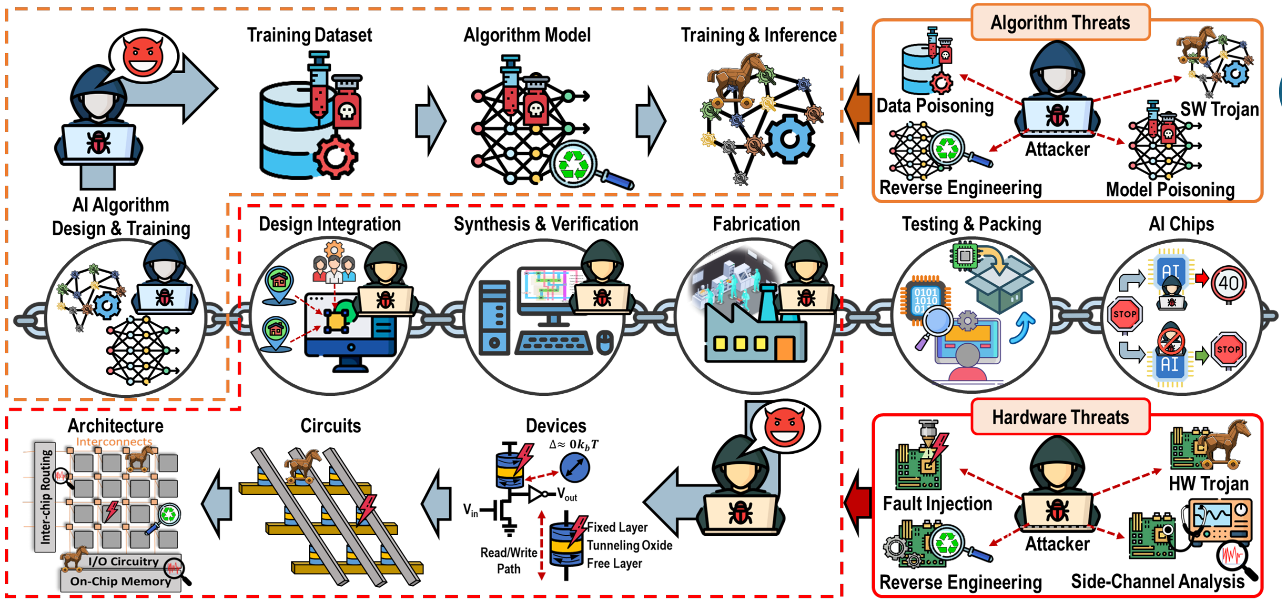
FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence



WH.GOV

NOVEMBER 27, 2023

FACT SHEET: President Biden Announces New Actions to Strengthen America's Supply Chains, Lower Costs for Families, and Secure Key Sectors



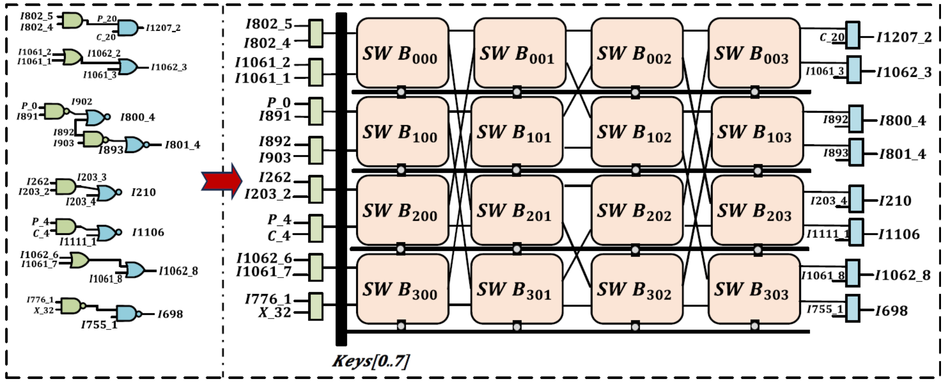
Source of Threats: Almost Everywhere!

Design Engineer

OASIC: Optimized and Automated Secure IC Design Flow

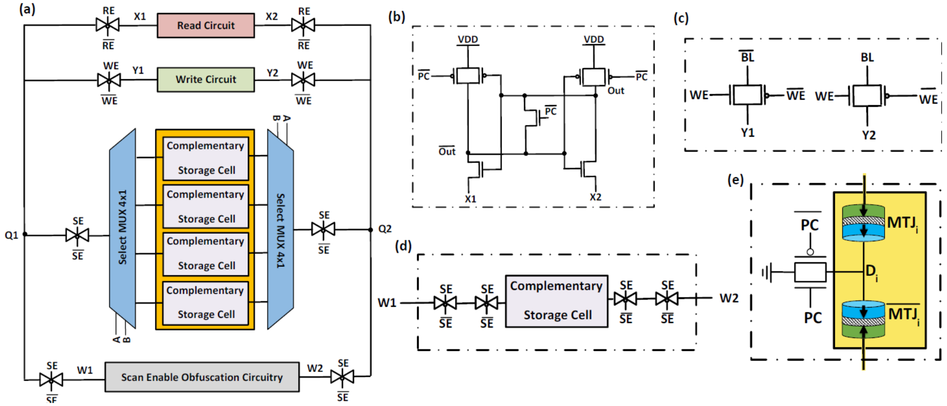
Reconfigurable Interconnect and Logic (RIL) Blocks:

✓ High output corruptibility and dynamic morphing mitigate reverse-engineering attacks.

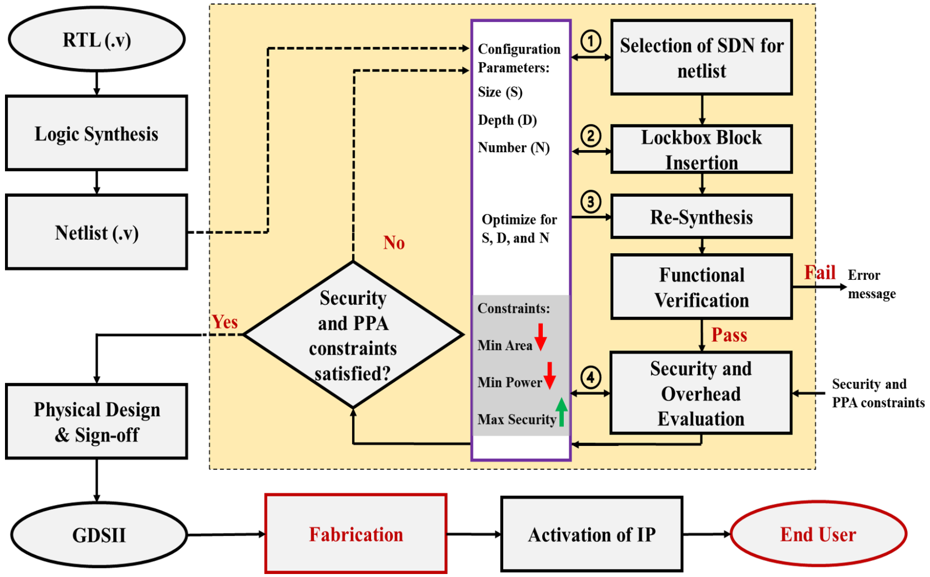


Symmetrical MRAM LUT (SyM-LUT):

✓ Low power variation mitigates AI-assisted power side-channel attacks.



OASIC: Optimized and Automated Secure IC Design Flow:



RTL (.v)

Logic Synthesis

Netlist (.v)

Configuration Parameters:
Size (S)
Depth (D)
Number (N)

Optimize for S, D, and N

Constraints:
Min Area
Min Power
Max Security

1 Selection of SDN for netlist

2 Lockbox Block Insertion

3 Re-Synthesis

Functional Verification

Fail Error message

Pass

4 Security and Overhead Evaluation

Security and PPA constraints

Security and PPA constraints satisfied?

No

Yes

Physical Design & Sign-off

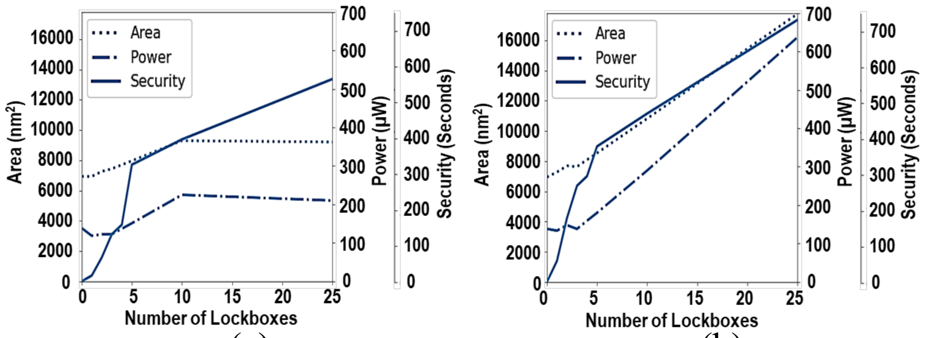
GDSII

Fabrication

Activation of IP

End User

Simulation Results:



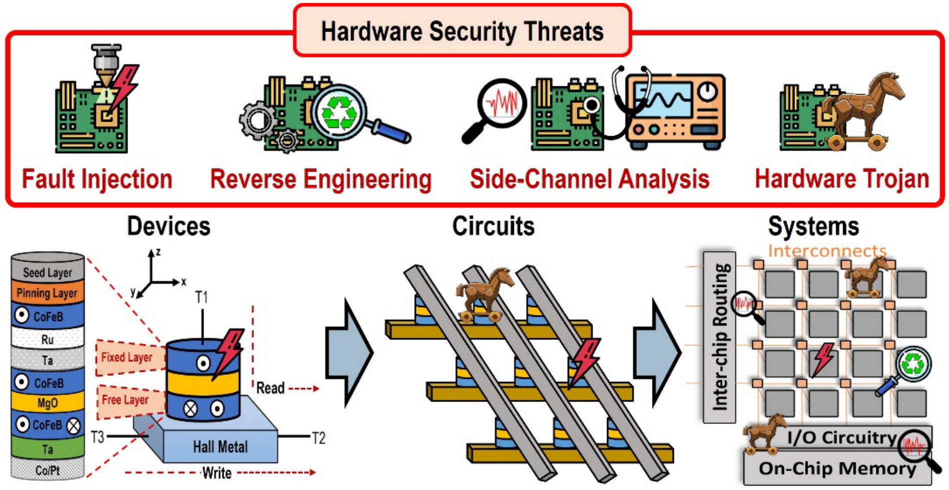
(a) (b)

Power-Area-Security trade-off of obfuscating an AES core with the proposed (a) 8x8 (b) 8x8x8 lockboxes.

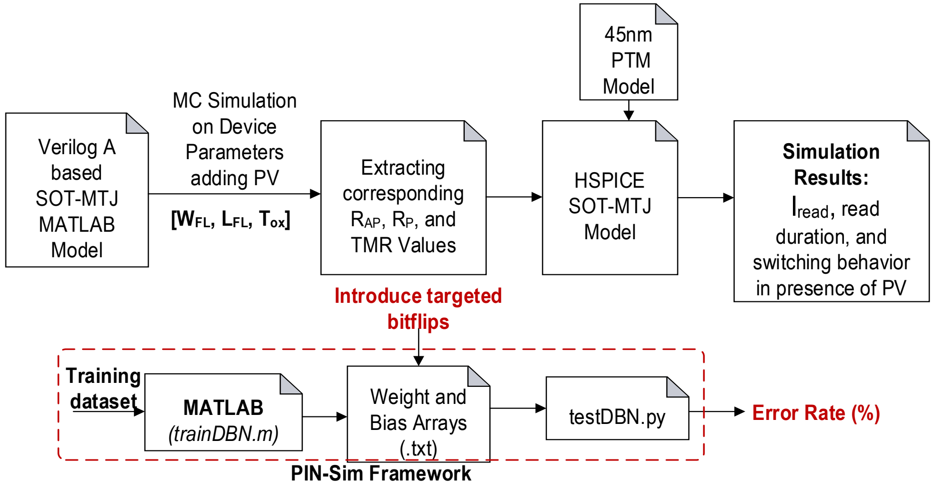
- Optimization: Number (N) = {2, 3}, Size (S) = {4, 8}, and Depth (D) = {2, 3}.
- Scalability: Design Size (↑) = Overhead (↓)
- Overhead: Area =<1.15 and Power =<1

Secure AI Hardware Accelerator Design using Post-CMOS Devices

Semiconductor Supply Chain Threats on AI Accelerators:



Simulation framework and process flow:


Verilog A based SOT-MTJ MATLAB Model

MC Simulation on Device Parameters adding PV

$[W_{FL}, L_{FL}, T_{ox}]$

Extracting corresponding R_{AP}, R_P, and TMR Values

45nm PTM Model

HSPICE SOT-MTJ Model

Simulation Results: I_{read}, read duration, and switching behavior in presence of PV

Introduce targeted bitflips

Training dataset

MATLAB *(trainDBN.m)*

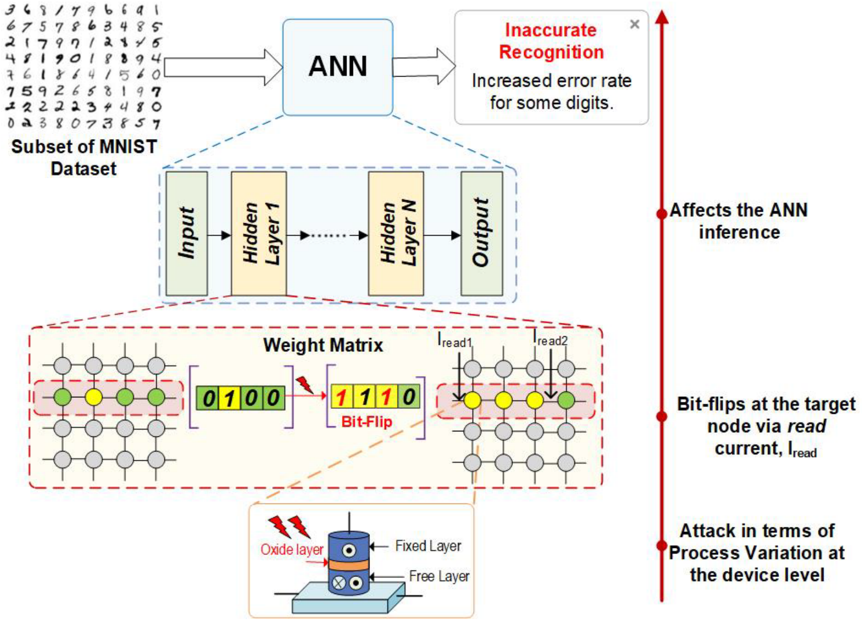
Weight and Bias Arrays (.txt)

testDBN.py

Error Rate (%)

PIN-Sim Framework

Flow diagram of the sensitivity and threat analysis approach:


Subset of MNIST Dataset

ANN

Inaccurate Recognition

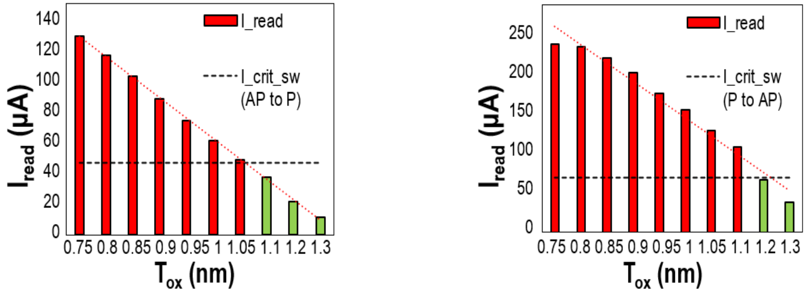
Increased error rate for some digits.

Affects the ANN inference

Bit-flips at the target node via *read* current, I_{read}

Attack in terms of Process Variation at the device level

Simulation Results:



(a)

(b)

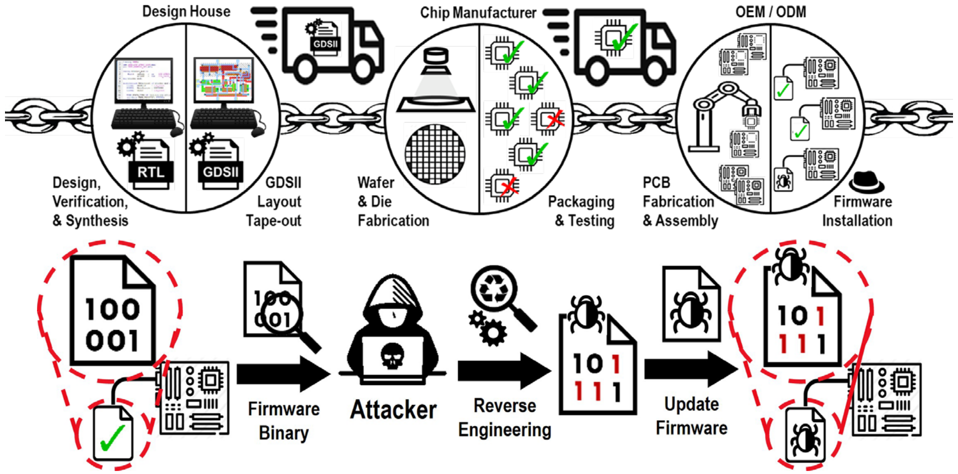
SOT-MTJ device read current (I_{read}) variation with change in oxide layer thickness during (a) ‘AP’ to ‘P’, and (b) ‘P’ to ‘AP’ switching.

Uniformly change t_{ox} from 0.75nm to 1.2nm without modifying the netlist nor even having access to the circuit design itself. MNIST dataset with 0.05% of bit-flips in crossbar:

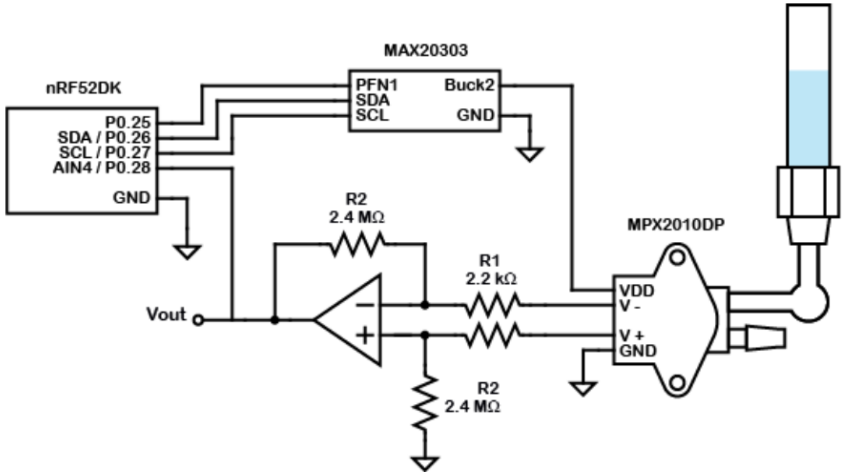
- Digits ‘4’ and ‘5’ show highest overall error rates of 85.71%.
- Digit ‘9’ has the lowest impact and ‘2’, ‘3’, and ‘8’ unaffected.

FANDEMIC: Firmware Attack Construction and Deployment on Power Management Integrated Circuit and Impacts on IoT Applications

Semiconductor Supply Chain Threats on firmware:



Design framework for attack implementation flow:



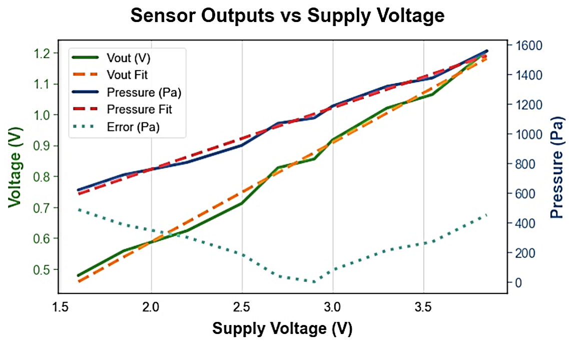
Flow diagram of the sensitivity and threat analysis approach:

- Identify addresses associated with Two-Wire Interface (TWI) peripheral that nRF52 uses for I2C.
- Reverse engineer binary to locate Buck2 configuration.
- Modify Buck2 configuration bytes and checksum.

```
original_temp_1.8V_09.hex:
0003 8D20: 30 30 34 32 32 32 31 41  37 30 39 44 46 38 30 46 0042221A 709DF80F
0003 8D30: 33 30 30 30 32 42 30 33  44 30 31 36 34 42 30 31 30002B03 D0164B01
0003 8D40: 32 32 43 39 0D 0A 3A 31  30 42 33 34 30 30 30 31 22C9..:1 0B340001
0003 8D50: 41 37 30 30 32 45 30 31  34 34 42 30 30 32 32 31 A7002E01 44B00221
0003 8D60: 41 37 30 31 33 34 42 30  39 32 32 35 41 37 30 33 A70134B0 9225A703
0003 8D70: 33 0D 0A 3A 31 30 42 33  35 30 30 30 30 44 46 31 3..:10B3 50000DF1
0003 8D80: 31 33 30 33 31 39 34 36  30 32 32 30 46 46 46 37 13031946 0220FFF7
0003 8D90: 31 32 46 45 30 35 39 30  30 35 39 42 31 44 0D 0A 12FE0590 059B1D..
0003 8DA0: 3A 31 30 42 33 36 30 30  30 30 30 32 42 30 38 44 :10B3600 0002B08D

modified_temp_3.8V_1d.hex:
0003 8D20: 30 30 34 32 32 32 31 41  37 30 39 44 46 38 30 46 0042221A 709DF80F
0003 8D30: 33 30 30 30 32 42 30 33  44 30 31 36 34 42 30 31 30002B03 D0164B01
0003 8D40: 32 32 43 39 0D 0A 3A 31  30 42 33 34 30 30 30 31 22C9..:1 0B340001
0003 8D50: 41 37 30 30 32 45 30 31  34 34 42 30 30 32 32 31 A7002E01 44B00221
0003 8D60: 41 37 30 31 33 34 42 31  44 32 32 35 41 37 30 31 A70134B1 D225A701
0003 8D70: 46 0D 0A 3A 31 30 42 33  35 30 30 30 30 44 46 31 F..:10B3 50000DF1
0003 8D80: 31 33 30 33 31 39 34 36  30 32 32 30 46 46 46 37 13031946 0220FFF7
0003 8D90: 31 32 46 45 30 35 39 30  30 35 39 42 31 44 0D 0A 12FE0590 059B1D..
0003 8DA0: 3A 31 30 42 33 36 30 30  30 30 30 32 42 30 38 44 :10B3600 0002B08D
```

Simulation Results:



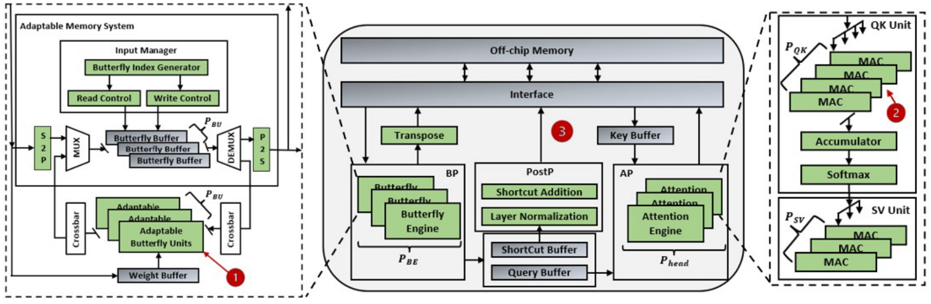
- Data corruption by altering sensor power supplies.
- Error rate of ~3.75% relative to pressure reading at 2.9V for every change in 0.1V of supply voltage.

Outputs averaged over 100 samples with 100ms period

Config	V_{Supply}	$D[V_{Out}]$	V_{Out}	Pressure (Pa)	% Error
0x10	1.60	545	0.48	618.44	44.04
0x15	1.85	635	0.56	720.99	34.77
0x1C	2.20	710	0.62	805.54	27.12
0x22	2.50	810	0.71	919.40	16.81
0x26	2.70	941	0.83	1068.09	3.36
0x2A	2.90	974	0.86	1105.23	0.00
0x2C	3.00	1043	0.92	1184.12	7.14
0x32	3.30	1161	1.02	1317.43	19.20
0x37	3.55	1211	1.06	1374.38	24.35
0x3D	3.85	1371	1.20	1556.19	40.80

Secure Generative AI Hardware Accelerator Design

Hardware overview of adaptable butterfly accelerator for Generative AI algorithm and Hardware Trojan (HT) vulnerabilities:

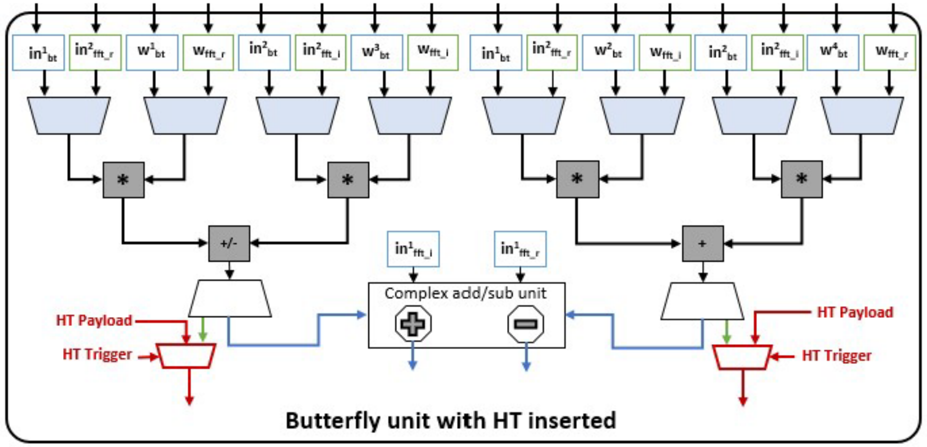


(b) Butterfly Engine

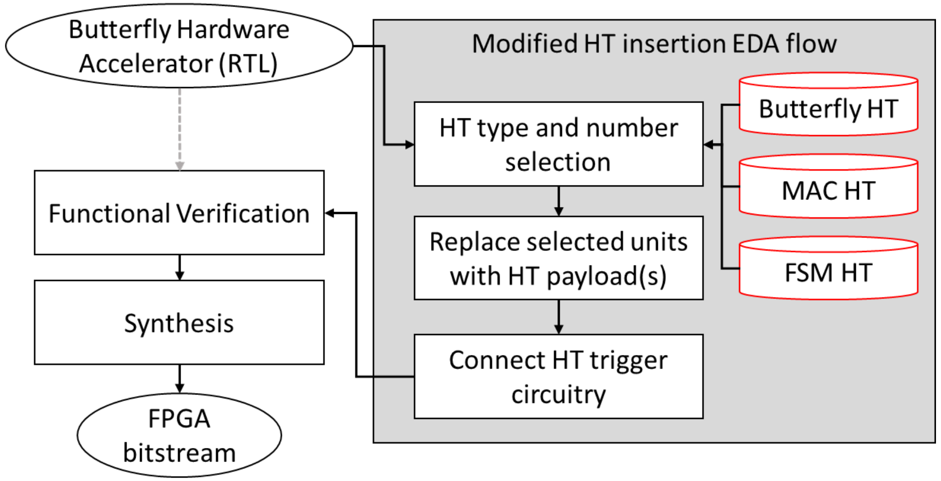
(a) Hardware Overview

(c) Attention Engine

Sample HT-modified architecture of the butterfly unit with:



EDA flow for inserting HTs in the hardware accelerator:



Butterfly Hardware Accelerator (RTL)

Modified HT insertion EDA flow

HT type and number selection

Butterfly HT

MAC HT

FSM HT

Replace selected units with HT payload(s)

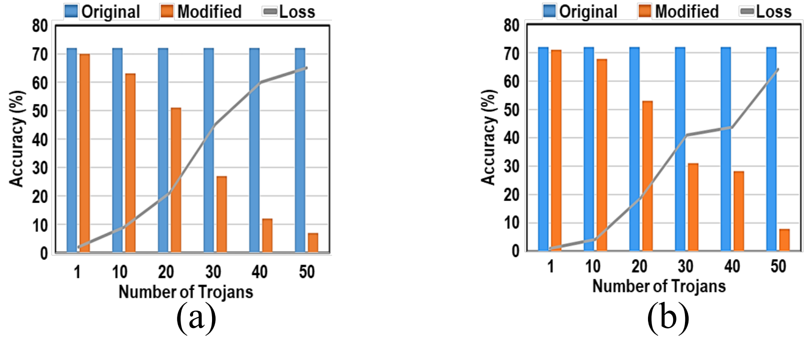
Connect HT trigger circuitry

Functional Verification

Synthesis

FPGA bitstream

Simulation Results:

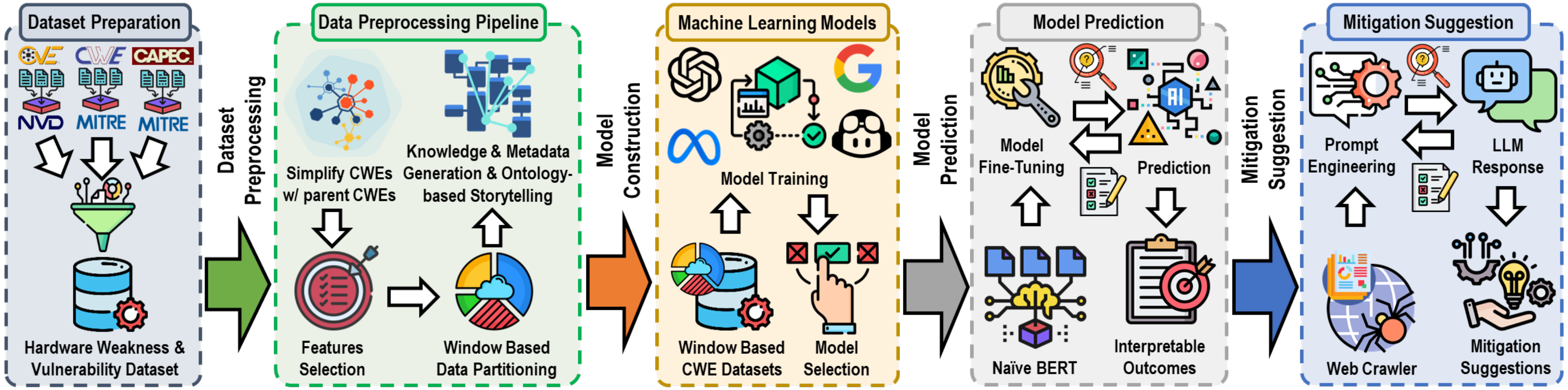


Accuracy degradation for different numbers HTs within: (a) MAC accelerator unit, and (b) Butterfly accelerator unit.

- Accuracy degradation of 5%-65% for varying numbers of HTs.
- Butterfly HT and MAC HT have negligible area overhead of 0.28% and 0.29%, respectively.
- 0.0421% and 0.052% of test vectors triggered
- Butterfly HT and MAC HT, respectively.

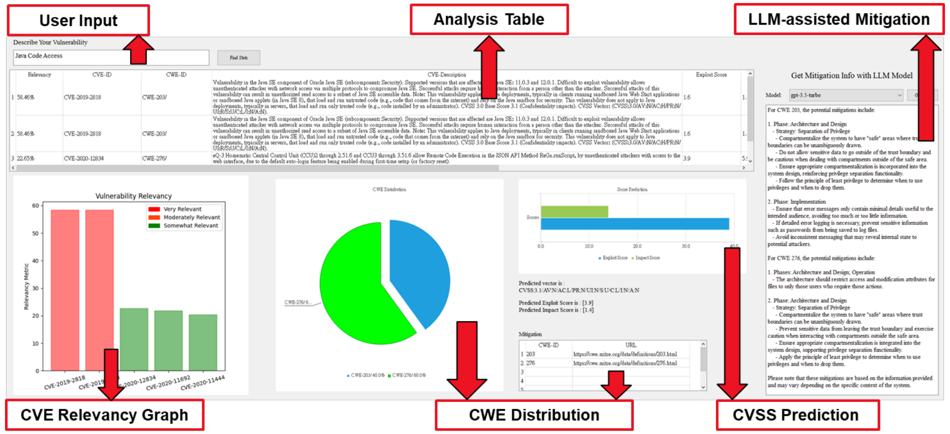
HW-V2W-Map for Hardware Vulnerability and Risk Assessment

Flow diagram of the Hardware Vulnerability to Weakness Mapping framework for risk assessment and root cause analysis:



Graphical User Interface for the HW-V2W-Map

- ✓ **Demo: https://youtu.be/rdejfpFcqXk**
- ✓ **Repository: https://gitlab.com/yuzhenglin/HW-V2W-Map**



Describe Your Vulnerability: Java Code Access

	Relevancy	CVE-ID	CWE-ID	CVE-Description	Exploit Score
1	58.46%	CVE-2019-2818	CWE-203/	Vulnerability in the Java SE component of Oracle Java SE (subcomponent: Security). Supported versions that are affected are Java SE: 11.0.3 and 12.0.1. Difficult to exploit vulnerability allows unauthenticated attacker with network access via multiple protocols to compromise Java SE. Successful attacks require human interaction from a person other than the attacker. Successful attacks of this vulnerability can result in unauthorized read access to a subset of Java SE accessible data. Note: This vulnerability applies to Java deployments, typically in clients running sandboxed Java Web Start applications or sandboxed Java applets (in Java SE 8), that load and run untrusted code (e.g., code that comes from the internet) and rely on the Java sandbox for security. This vulnerability does not apply to Java deployments, typically in servers, that load and run only trusted code (e.g., code installed by an administrator). CVSS 3.0 Base Score 3.1 (Confidentiality impacts). CVSS Vector: (CVSS:3.0/AV:N/AC:H/PR:N/UI:R/S:U/C:L/I:N/A:N).	1.6
2	58.46%	CVE-2019-2818	CWE-203/	Vulnerability in the Java SE component of Oracle Java SE (subcomponent: Security). Supported versions that are affected are Java SE: 11.0.3 and 12.0.1. Difficult to exploit vulnerability allows unauthenticated attacker with network access via multiple protocols to compromise Java SE. Successful attacks require human interaction from a person other than the attacker. Successful attacks of this vulnerability can result in unauthorized read access to a subset of Java SE accessible data. Note: This vulnerability applies to Java deployments, typically in clients running sandboxed Java Web Start applications or sandboxed Java applets (in Java SE 8), that load and run untrusted code (e.g., code that comes from the internet) and rely on the Java sandbox for security. This vulnerability does not apply to Java deployments, typically in servers, that load and run only trusted code (e.g., code installed by an administrator). CVSS 3.0 Base Score 3.1 (Confidentiality impacts). CVSS Vector: (CVSS:3.0/AV:N/AC:H/PR:N/UI:R/S:U/C:L/I:N/A:N).	1.6
3	22.65%	CVE-2020-12834	CWE-276/	eQ-3 Homematic Central Control Unit (CCU)2 through 2.51.6 and CCU3 through 3.51.6 allow Remote Code Execution in the JSON API Method ReGa.runScript, by unauthenticated attackers with access to the web interface, due to the default auto-login feature being enabled during first-time setup (or factory reset).	3.9

Predicted vector is :
CVSS:3.1/AV:N/AC:L/PR:N/UI:N/S:U/C:L/I:N/A:N

Predicted Exploit Score is : [3.9]
Predicted Impact Score is : [1.4]

Mitigation

	CWE-ID	URL
1	203	https://cwe.mitre.org/data/definitions/203.html
2	276	https://cwe.mitre.org/data/definitions/276.html
3		
4		

Get Mitigation Info with LLM Model

Model: gpt-3.5-turbo

For CWE 203, the potential mitigations include:

1. Phase: Architecture and Design
   - Strategy: Separation of Privilege
     - Compartmentalize the system to have "safe" areas where trust boundaries can be unambiguously drawn.
     - Do not allow sensitive data to go outside of the trust boundary and be cautious when dealing with compartments outside of the safe area.
     - Ensure appropriate compartmentalization is incorporated into the system design, reinforcing privilege separation functionality.
     - Follow the principle of least privilege to determine when to use privileges and when to drop them.

2. Phase: Implementation
   - Ensure that error messages only contain minimal details useful to the intended audience, avoiding too much or too little information.
   - If detailed error logging is necessary, prevent sensitive information such as passwords from being saved to log files.
   - Avoid inconsistent messaging that may reveal internal state to potential attackers.

For CWE 276, the potential mitigations include:

1. Phases: Architecture and Design; Operation
   - The architecture should restrict access and modification attributes for files to only those users who require those actions.

2. Phase: Architecture and Design
   - Strategy: Separation of Privilege
     - Compartmentalize the system to have "safe" areas where trust boundaries can be unambiguously drawn.
     - Prevent sensitive data from leaving the trust boundary and exercise caution when interacting with compartments outside the safe area.
     - Ensure appropriate compartmentalization is integrated into the system design, supporting privilege separation functionality.
     - Apply the principle of least privilege to determine when to use privileges and when to drop them.

Please note that these mitigations are based on the information provided and may vary depending on the specific context of the system.

Simulation Results:

Window-based ML-assisted Important HW CWE Prediction

Time Window	Training R^2 Score	Testing R^2 Score	Testing Mean Squared Log Err	Testing Mean Absolute Err	Testing Median Absolute Err
3-year	0.98	0.92	1.95	21.98	6.37
2-year	0.97	0.86	1.78	24.31	6.90
1-year	0.98	0.91	2.14	23.28	6.94

Window-based ML-assisted CVE CVSS Prediction

Model Name	Evaluation Loss	Evaluation $F1$ Score	Evaluation ROC AUC	Evaluation Accuracy	Evaluation Precision	Evaluation Recall
CWE-CWE	0.18	0.66	0.80	0.62	0.73	0.86
CWE-CWE-Binary	0.44	0.81	0.81	0.80	0.68	0.79
CWE-CVE	0.46	0.84	0.84	0.83	0.84	0.85
CWE-CAPEC	0.51	0.79	0.79	0.77	0.77	0.83

- Predicting the relationship between CWE to CVE and CWE to CAPEC is more accurate than predicting CWE to CWE due to fewer labels involved.

In conclusion

- Enormous R&D and workforce opportunities for ‘math’ inclined folks
- UA-CSM is poised to spearhead the growth of the semiconductor footprint on campus and in AZ
- More than 30 faculty across campus who would greatly benefit from the AM GIDP program
- Start talking to the faculty if you want to be part of the bandwagon

Acknowledgement:
Thanks to the internet for all the nice figures and blurbs