Lilienfeld's FET with 3D CMOS and Enhanced Moore's Law by Heterogeneous Integration toward the Bicentennial

Nicky Lu PhD, Stanford; National Academy of Engineering (NAE) & Inventors (NAI), USA

Chair & Founder, Etron Tech, eYs3D Microelectronics, DeCloak & eEver Tech
Chair (2019~), AI-on-Chip Taiwan Alliance (AITA)
Distinguished Chair Professor & Outstanding Alumnus, National Taiwan University

Chair (2014-15), WSC (World Semiconductor Council)
Chair (2013-17) & Board Member (1996~), TSIA (Taiwan Semiconductor Industry Association)
Chair (2009-11) & Board Member (2004~), GSA (Global Semiconductor Alliance)
Managing Director (2002~), TPIA (Taiwan Science-Based Industrial Park Industrial Alliance)
Managing Director, TEEMA (Taiwan Electrical & Electronic Manufacturers' Association)

IBM Corporate Award for DRAM Inventions & Contributions to World's First 8" Wafer Fab by IBM
Architected/Co-Developed Taiwan-Self-Owned 0.5 μm Logic/SRAM/DRAM in Her First 8" Wafer Fab
Awarded 304 Patents (US 84), IEEE Fellow ('91) & Solid-State Circuits D. O. Pederson Award ('98)

June 18, 2026

Celebrating A Century of Discovery & Invention

Quantum Science & Tech

On June 7, 2024, the United Nations General Assembly officially declared 2025 as the "International Year of Quantum Science and Technology" (YQ), marking the 100th anniversary of the birth of quantum mechanics.

Field-Effect Transistor (FET)

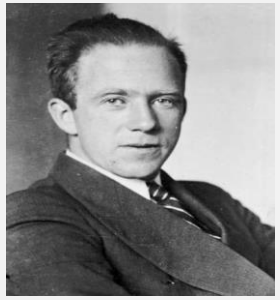
In 2025, the IEEE Electronics Devices Society (EDS) will commemorate the 100th anniversary of Julius Lilienfeld's patent application for the first field-effect transistor (FET) on October 22, 1925 (Canada), on October 8, 1926 (USA). This groundbreaking invention laid the foundation for modern integrated circuits. To celebrate this significant milestone, EDS is hosting a series of events titled "FET100."

After CY Lu, 2025; Nicky & CY Lu, 2026

As Physicists Redraw The Universe on Paper, Engineers Quietly Plant The Seeds for A Silicon-Based Civilization

Quantum Mechanics to Computing

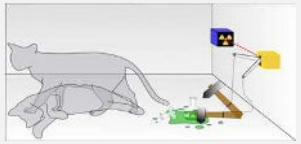
Werner Kari Heisenberg



$$\frac{\partial}{\partial t}A_{\mathcal{H}} = \frac{1}{i\hbar}[A_{\mathcal{H}}, H]$$

Erwin Schrödinger

$$\hat{H}\Psi = i\hbar\frac{\partial}{\partial t}\Psi$$

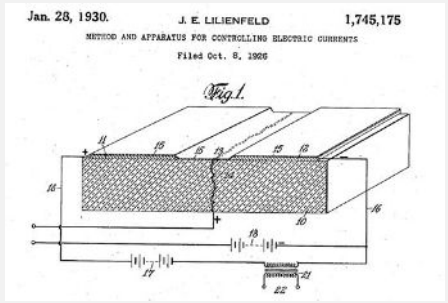


1925~1926

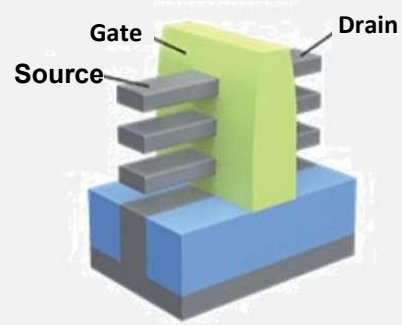
Field Effect Transistor to CMOS Si4.0/HI

US 1745175 "Method and apparatus for controlling electric current" first filed in Canada on 1925-10-22 / 1926-10-8 (US), describing a field-effect transistor

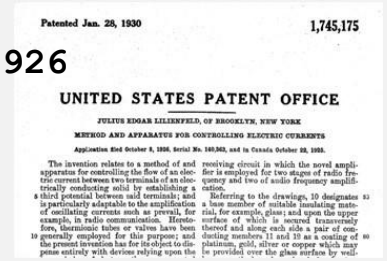
Julius Edgar Lilienfeld



Today: GAA (Gate-All-Around) in 2nm



1925~1926


Patented Jan. 28, 1930 1,745,175

UNITED STATES PATENT OFFICE

JULIUS EDGAR LILIENFELD, OF BROOKLYN, NEW YORK

METHOD AND APPARATUS FOR CONTROLLING ELECTRIC CURRENTS

The invention relates to a method of and apparatus for controlling the flow of an electric current between two terminals of an electrically conducting solid by establishing a third potential between said terminals; and is particularly adaptable to the amplification of oscillating currents such as prevail, for example, in radio communication. Heretofore, thermionic tubes or valves have been generally employed for this purpose; and the present invention has for its object to dispense entirely with devices relying upon the

receiving circuit in which the novel amplifier is employed for two stages of radio frequency and two of audio frequency amplification.

Referring to the drawings, 10 designates a base member of suitable insulating material, for example, glass; and upon the upper surface of which is secured transversely thereof and along each side a pair of conducting members 11 and 12 as a coating of platinum, gold, silver or copper which may be provided over the glass surface by well-

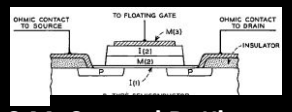
After CY Lu, 2025; Nicky & CY Lu, 2026

From Lilienfeld's FET to Today's 3D CMOS approaching 1nm

William Shockley
John Bardeen Walter
Brattain Point contact
transistor (Ge) (Bell
Labs)

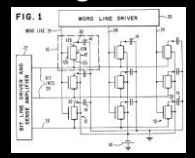
Geoffrey Dummer
presented Integrated
Circuits work at
conference

Jack Kilby's Original
Integrated Circuit
(Ge) (TI)

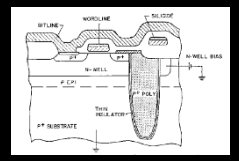


S.M. Sze and D. Khang
(Bell Labs)
Floating-Gate MOSFET

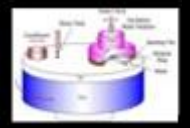
First Si
Transistor
(TI)



R. H Dennard
1T1C DRAM
(IBM)



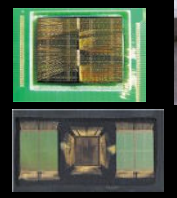
Highest Density
3D DRAM Cell &
World's 1st 8' Wafer



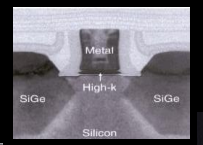
CMP(IBM)

Asia's 1st
8' Fab

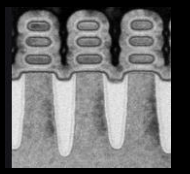
Strained
Silicon, SiGe



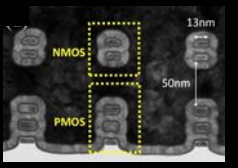
World's 1st KGDM,
Known Good Die
Memory (Etron)



Hi-k Metal
Gate (Intel)



Nanosheet / GAA

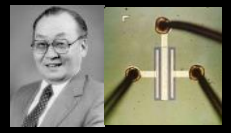


CFET

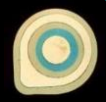
1925 1947 1950 1952 1958 1960 1967 1970 1984 1991 1998 2001 2002 2007 2011 2013 2020 2023 2026

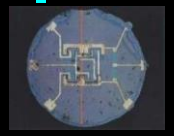
Julius Edgar
Lilienfeld File his
idea of first
transistor

William Shockley
Junction Transistor
(Bell Labs)



First field-effect
Transistor J Atalla
D Khang (Bell Labs)



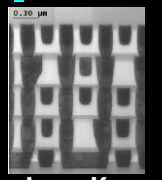


Jean Hoemi Planar
Transistor(Si) SiO2
introduced
(Fairchild)

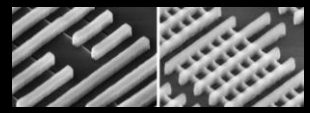
1st CMOS
device
(Fairchild)

Cu Interconnect
(IBM)

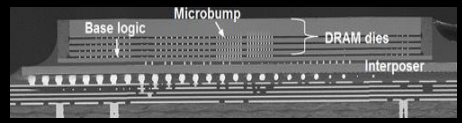
Heterogeneous
Integration
of Dice in KGD
Si/SiO_2 FETs
(Etron)



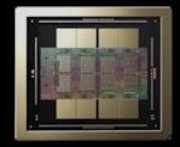
Low-K



3D FinFET/Trigate device
production
(Intel)



HBM 2013
(SK Hynix)



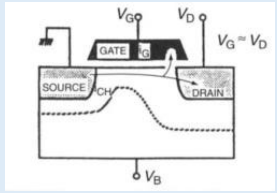
GPU + HBM4
(Nvidia)

After CY Lu, 2025; Nicky & CY Lu, 2026

Device Scaling Methodology from Silicon1.0 (Line Scaling) to Si2.0 (Area Scaling), Accelerating Moore's Economics at an Exponential Rate

(Si1.0) Line Scaling

(2D MOSFET 0.7X scaling)



Detailed Cross Sections for Scaled-Down Device Structure
Robert Dennard and IBM, 1974

SCALING RESULTS for CIRSUIT PERFORMANCE	
Device or Circuit Parameter	Scaling Factor
Device dimension t_{ox}, L, W	$1/\kappa$ **(0.7X)**
Doping concentration N_g	κ (1.43X)
Voltage V	$1/\kappa$
Current I	$1/\kappa$
Capacitance $\varepsilon A/t$	$1/\kappa$
Delay time/circuit CV/I	**$1/\kappa$(0.7X)**
Power dissipation/circuit VI	$1/\kappa^2$
Power density VI/A	1

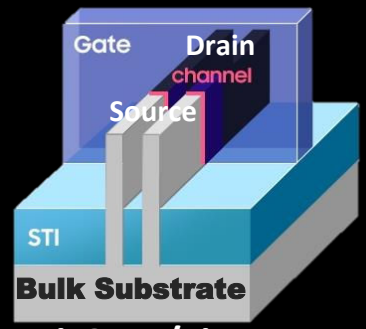
Power Density approaching 1 ← **Target Set by Dennard's Scaling (Badly Needed for Today's AI)**

(Si2.0) Area Scaling

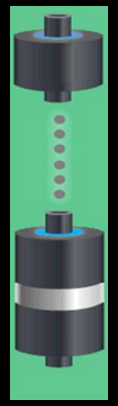
(3D FET+ >0.8X Line-Scaling)

PPACT Can Not Meet Moore's Law Demands

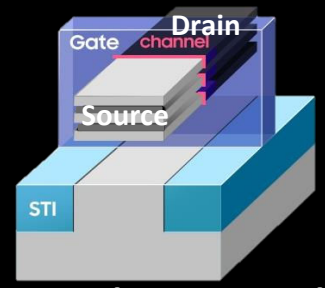
TSMC Leading Foundry 3D Areal Density Scaling:
10 nm Production in 2016;
7 nm in 2018;
5 nm in 2020;
3 nm in 2022;
2 nm in 2025; Intel 1.8nm
Outlook: 1.6/1.4 nm production in 2028?



Tri-Gate/FinFET



3D Flash

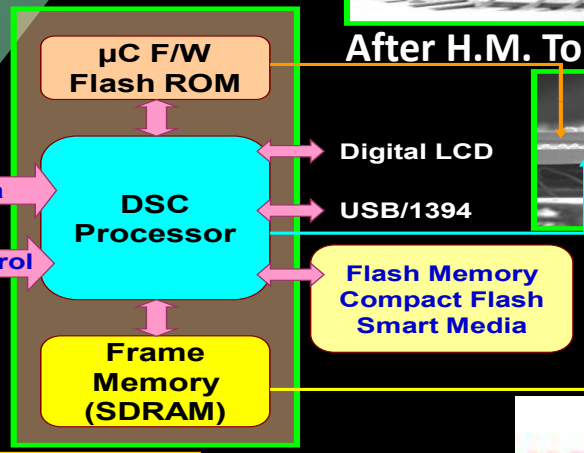


GAAFET/RibbonFET/ MBCFET

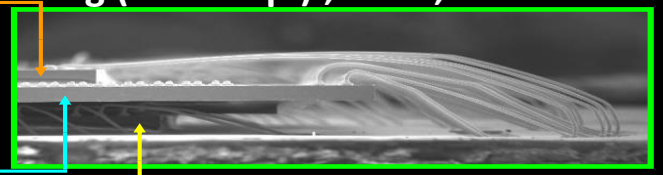
3D Transistor

After Nicky Lu, Plenary Talks @ A-SSCC2016; ISSM 2020, DRC 2021 http://www.etron.com.tw/manager/ISSM2020_NickyLu_Proprietary.pdf

A Breakthrough of Si3.0 in Y2000 – Proven Known-Good-Die Technologies by KGD-DRAM Production with Si/SiO_2 FETs Enabling Intelligent 2.5/3D System Chips



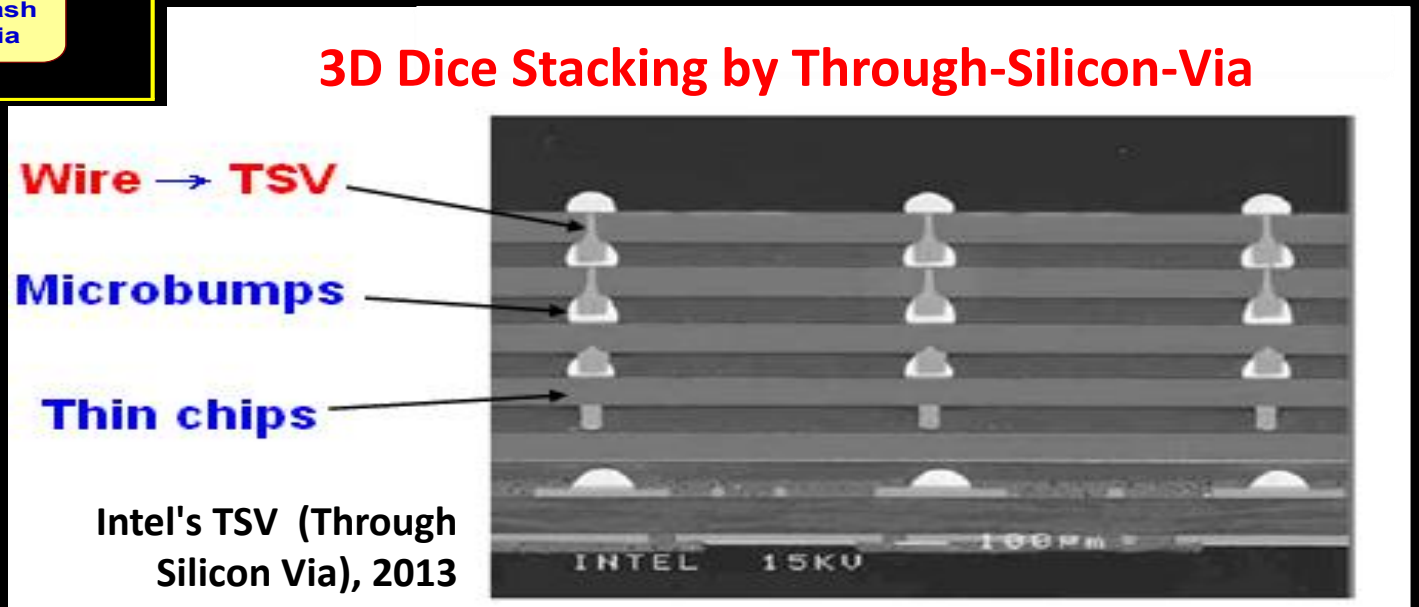
After H.M. Tong (ASE Corp.) ; N. Lu, ISSCC 2004



Etron Has Been Shipping KGD DRAM Since 2002. Shipments Have Totaled Over 3.1 Billion Units to-date. This Has Made Etron a Recognized Leading Contributor in 3D Technology

After Nicky Lu, KGD Products 2000, ISSCC 2004

3D Dice Stacking by Through-Silicon-Via

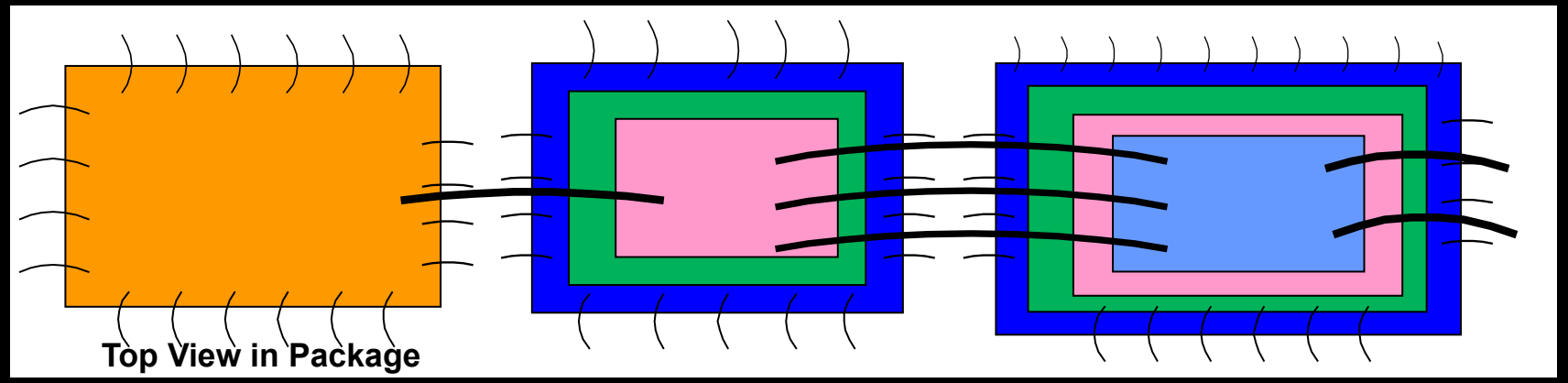


Intel's TSV (Through Silicon Via), 2013

Heterogeneous Integration (HI) Impacts Silicon3.0

- New System Architecture by Dice in a Package: Multi-dimensional Layouts in Chips/Subsystems to Increase Integration in addition to Device Scaling
- mDIC (m-Dimensional Dice Integration Chip); m= 2, 2.5, 3, 4…

Top View of Package or Module

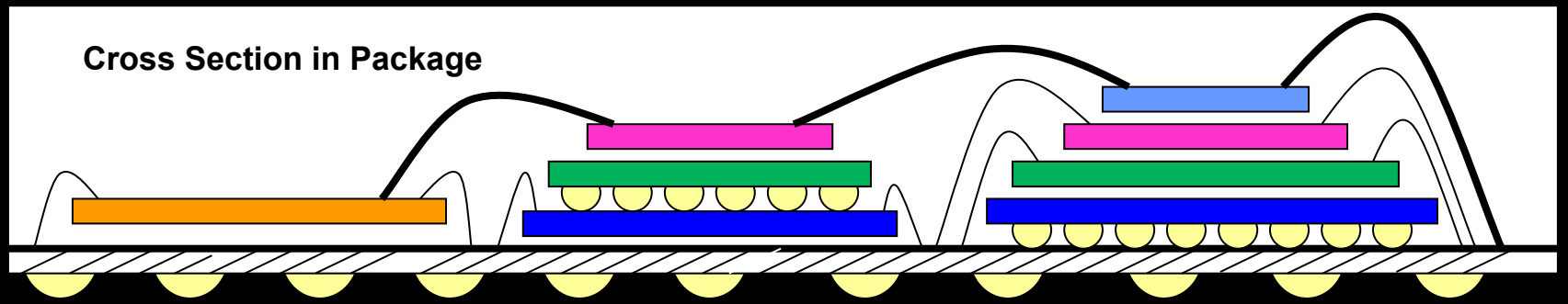


異質整合

（异构集成）

For Example: RRF or Power | Analog or Cache over SoC | DRAM + Analog + Cache + Logic

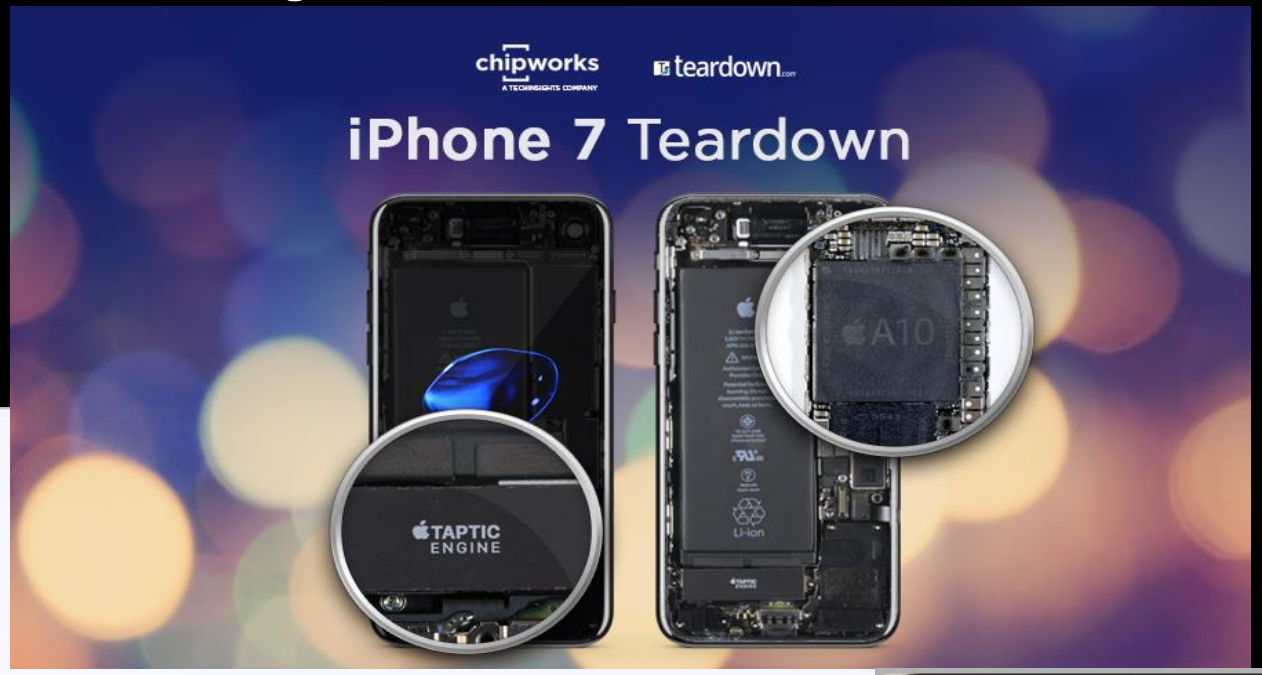
Cross Section



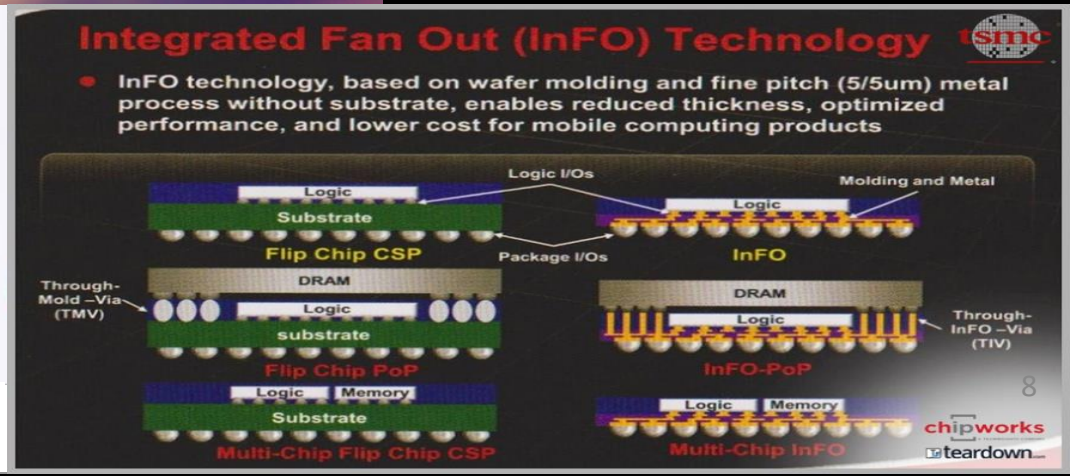
After Nicky Lu, ISSCC 2004 Plenary Talk (Published Heterogeneous Integration of Si-Centric Technology Era Emerging in 21st Century)

An Example: a Large-Volume Phone Maker Used HI

3D Bare-Die+Chip+InFO inside iPhone 7



2016



Source: Chipworks.com; September, 2016

Silicon 4.0 Era – Elevate Moore/Dennard's Scaling to A New "Function X Value" Holistic Design Architecture for Intelligent Foundational Microsystems to Facilitate AI Economic Boom – Both Volume & ASP Surge in 2026

(Si1.0)

Line Scaling

2D MOSFET

(Si2.0)

Area Scaling

(3D FinFET+ >0.7XLine-Scaling)

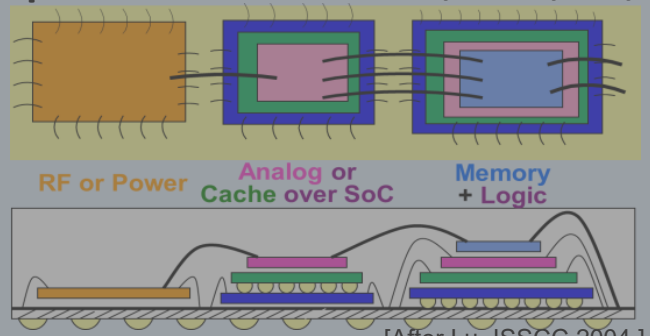
Continues Moore's Law

(Si3.0)

Volume or Volumetric Scaling

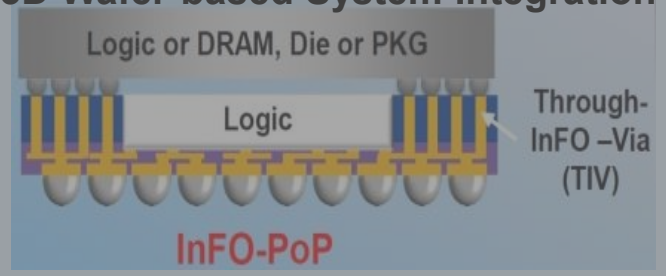
Heterogeneous Integration

Multiple Known-Good-Dice, MCP, SiP, MCM, Wirebond,Flipchip, SMT, PoP, etc.



[After Lu, ISSCC 2004.]

3D Wafer-based System Integration



[After Yu et al, IEDM 2014, Semicon West 2016]

(Si 4.0)

Function X Value Scaling

Optimizing Monolithic & Heterogeneous Integration (OMHI)of Si & Non-Si Materials/ Devices to Create Self-Smart MicroSystems Empowered by Nano-Si-Centric & AI Technologies

HETEROGENEOUS INTEGRATION ROADMAP IEEE

Refining Monolithic Devices/Technologies

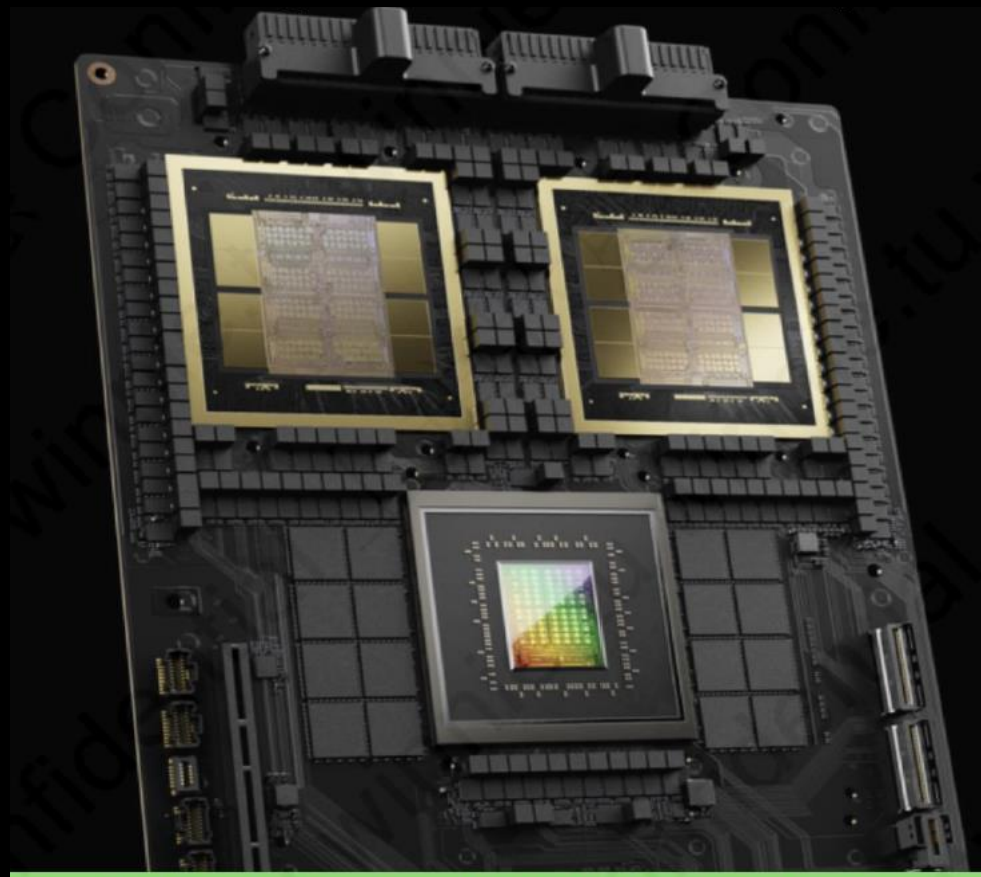
HIR

AI-Driven Self-Smart Microsystems & Networks for Super-Intelligent Systems

After Nicky Lu, Plenary Talks @ A-SSCC2016; ISSM 2020; DRC 2021; DRC 2025; ES Week ACM 2025

VLSI, Integrate at the Lowest Level Possible

VLSI → MCM → Board → Rack → Data Center



Si4.0
Booming !

One PCB = 2 Blackwell GPU B200 CoWoS + 1 Grace CPU chip + HBM's (High Bandwidth Memories). PCB:Printed Circuit Board

Made by TSMC Si technology and CoWoS
(J. Huang, Computex Taipei, 2024)

Nvidia's rack links 36 PCB's for generative AI training and inference

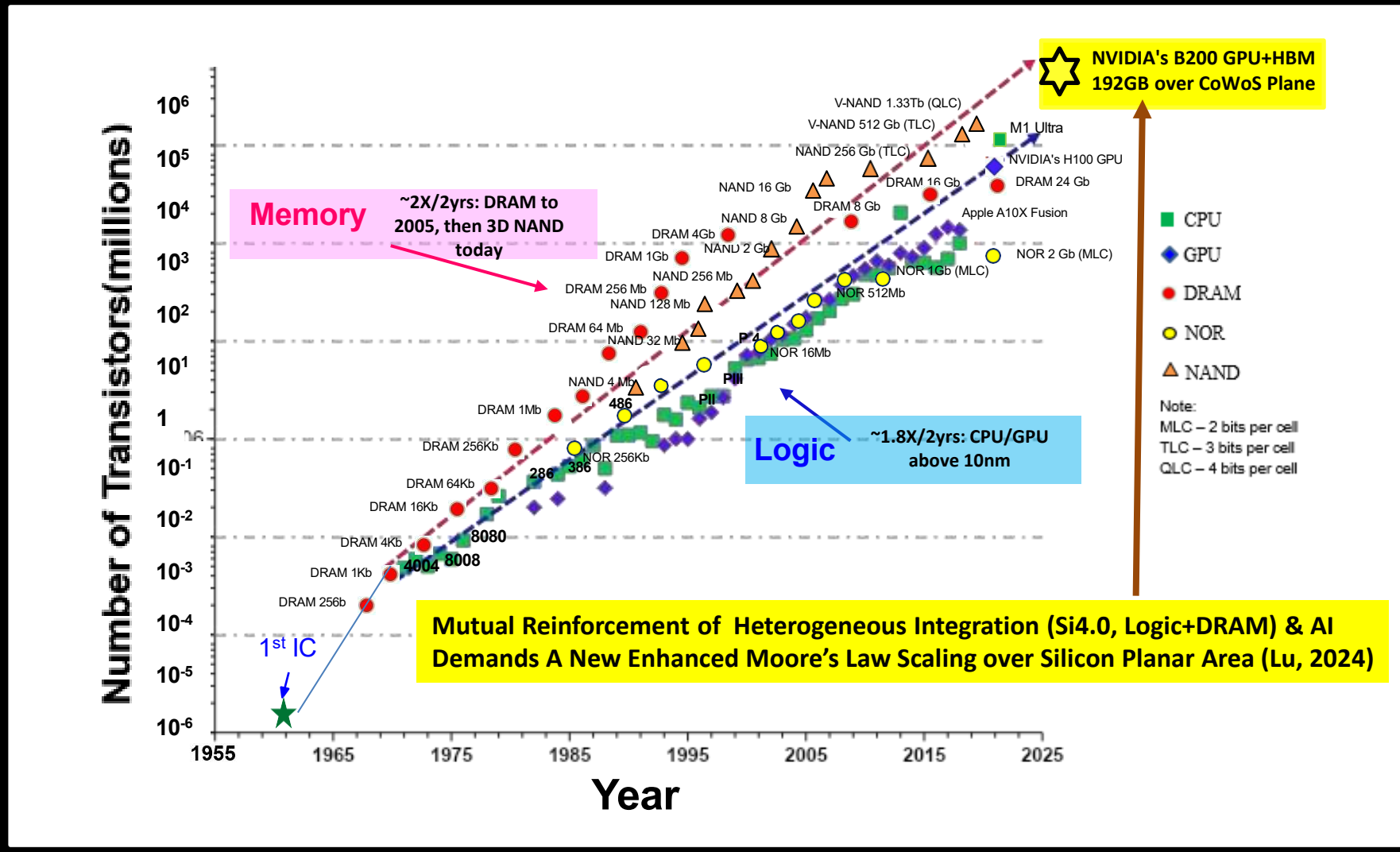
All connected w/ NVLink at 1.8 TB/sec

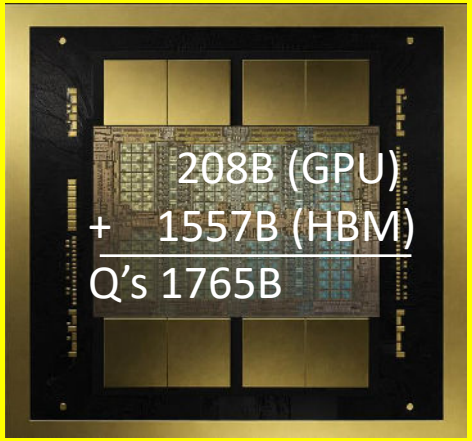
AI Compute Perf: 2.5x Hopper (old product)

Sources: John Chen, VLSI Symposium of Technology and Circuit, June 2025

A New Si4.0 Enhanced Moore's Law by Heterogeneous Integration

(Moore's Law : The Number of Transistors per Monolithic Chip Doubles Each 24 Months)





NVIDIA's B200

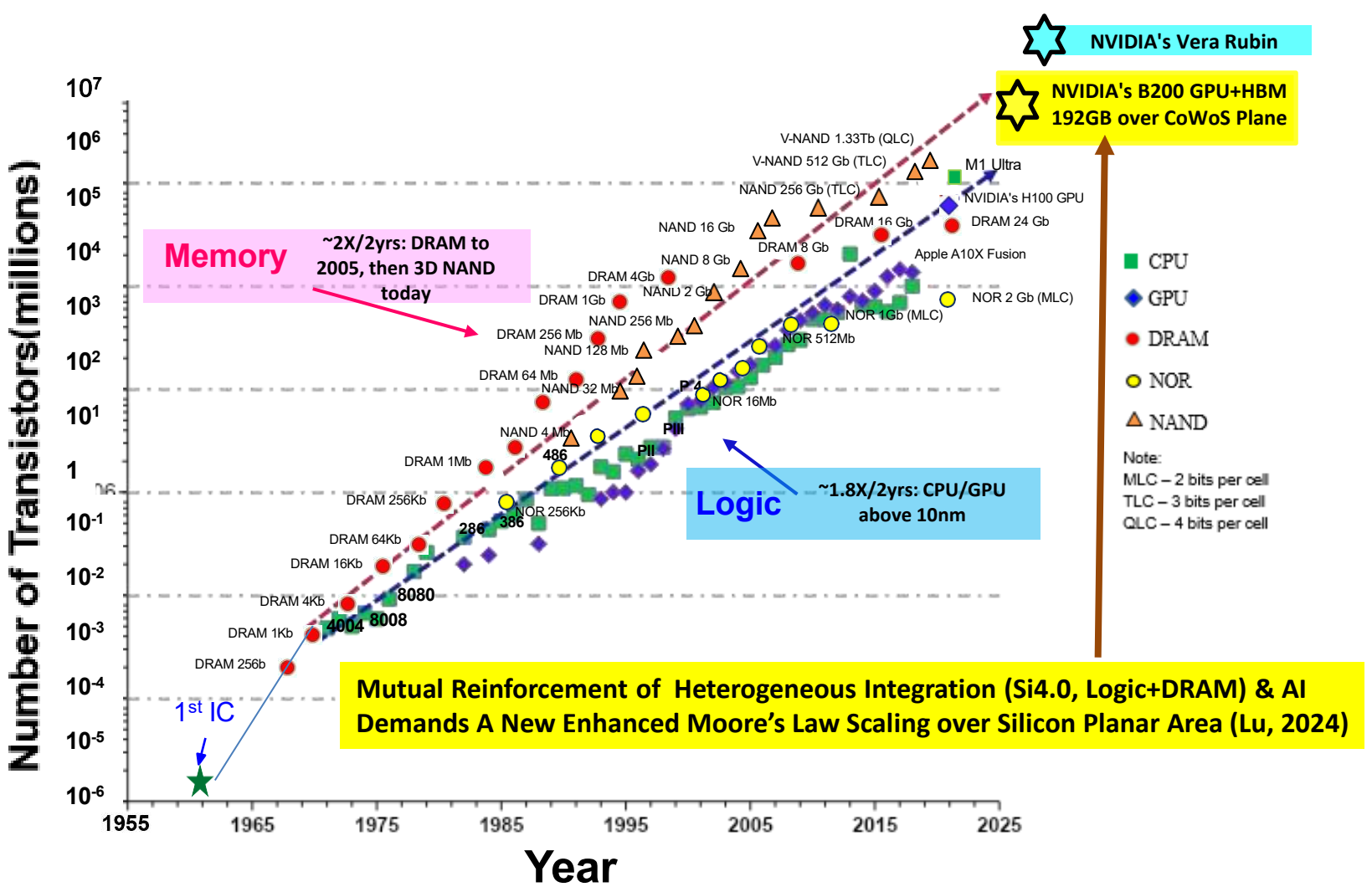
https://en.wikipedia.org/wiki/Transistor_count
https://en.wikipedia.org/wiki/Random-access_memory
https://en.wikipedia.org/wiki/Flash_memory#NAND_flash

* https://en.wikipedia.org/wiki/Moore%27s_law#cite_note-26

Referenced: C.Y. Lu, Etron

Vera Rubin Reinforces & Validates Si 4.0: The Function x Value Era

System-Level Integration of GPU, CPU, HBM and LPDDR5X Extends Enhanced Moore's Law from Transistor Count to Value Creation



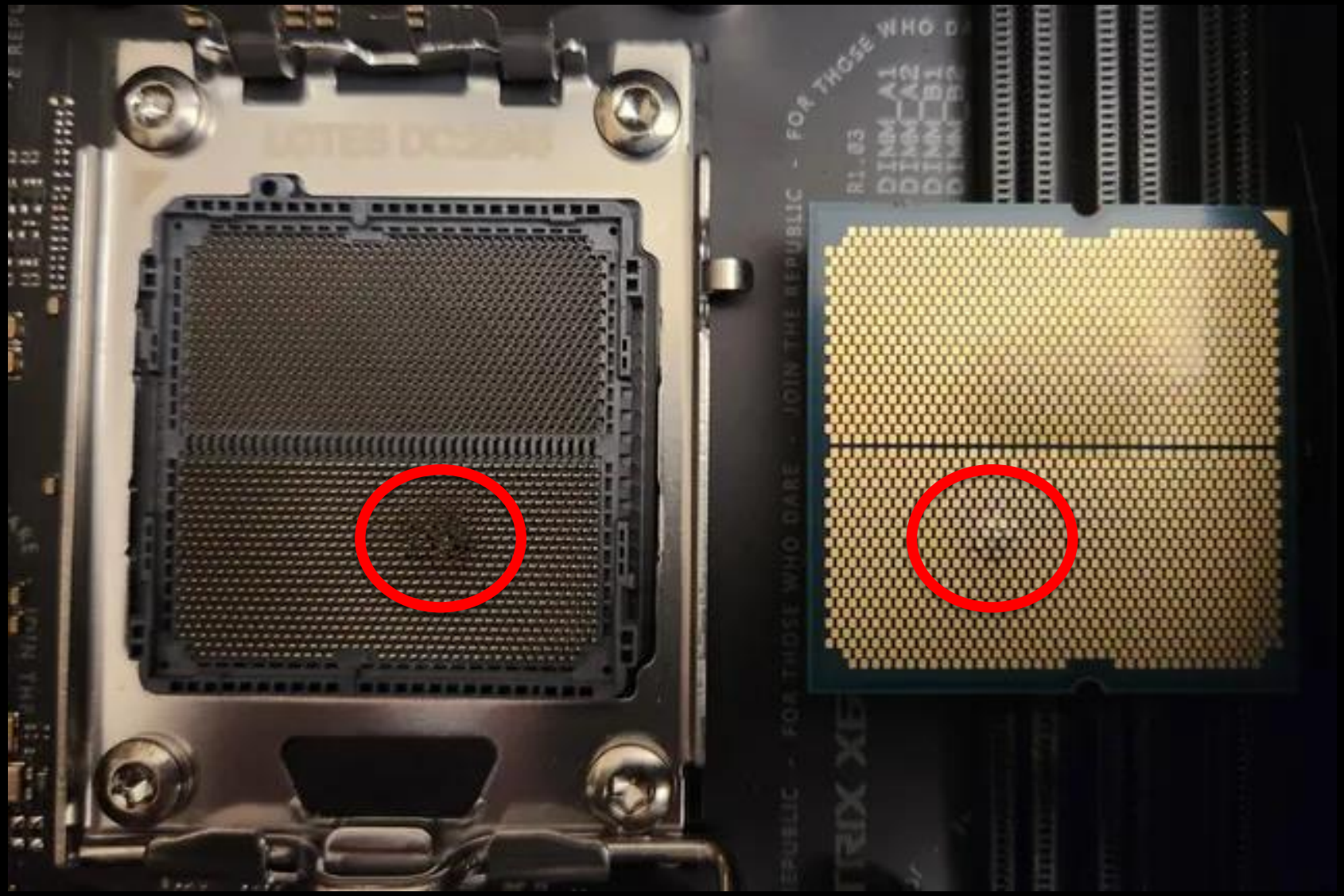
NVIDIA's Vera Rubin Proved

https://en.wikipedia.org/wiki/Transistor_count
https://en.wikipedia.org/wiki/Random-access_memory
https://en.wikipedia.org/wiki/Flash_memory#NAND_flash

* https://en.wikipedia.org/wiki/Moore%27s_law#cite_note-26

Referenced: C.Y. Lu, Etron

Heat Dissipation Challenges in H.I.: XPU+Chiplets Burning Out



Current Difficulties of The 1st Path Moore's Law Scaling (1PMLS)*

FET Structures Cannot Meet Area-Density Requirements – 2X per Generation

Peak Quoted Transistor Density For Standard Cell Optimization (MTr/mm^2 :Million Transistor per Area)

Technology Node	Company A	Company B	Company C	If obey Moore's Law
22nm		18.8		
16nm	28.9	Calculation Base on 16nm →		(28.88)
14nm		44.7	33.3	37.72
12nm	33.8			51.30
10nm	52.51		51.8	73.93
8nm (~12nm x 0.7 by Authors)			95.1	115.52
7nm	91.2(N7)/113.9(N7+)	100.8		150.88
5nm	138.2			295.73
4nm	143.7	123.4		462.08
3nm	216(N3E)/224(N3P)	148	150 MBCFET	821.48
2nm	313 (GAAFET)	238 (Ribbon FET)		1848.32
A14	391 (GAAFET)			3771.86

Ref:
TSMC & Intel papers, etc., IEEE IEDM & ISSCC, 2017-23; Wikichip & Anandtech & Tech Centurion, 2019-23; Techinsights, 2023-25

Referred and Estimated by Etron

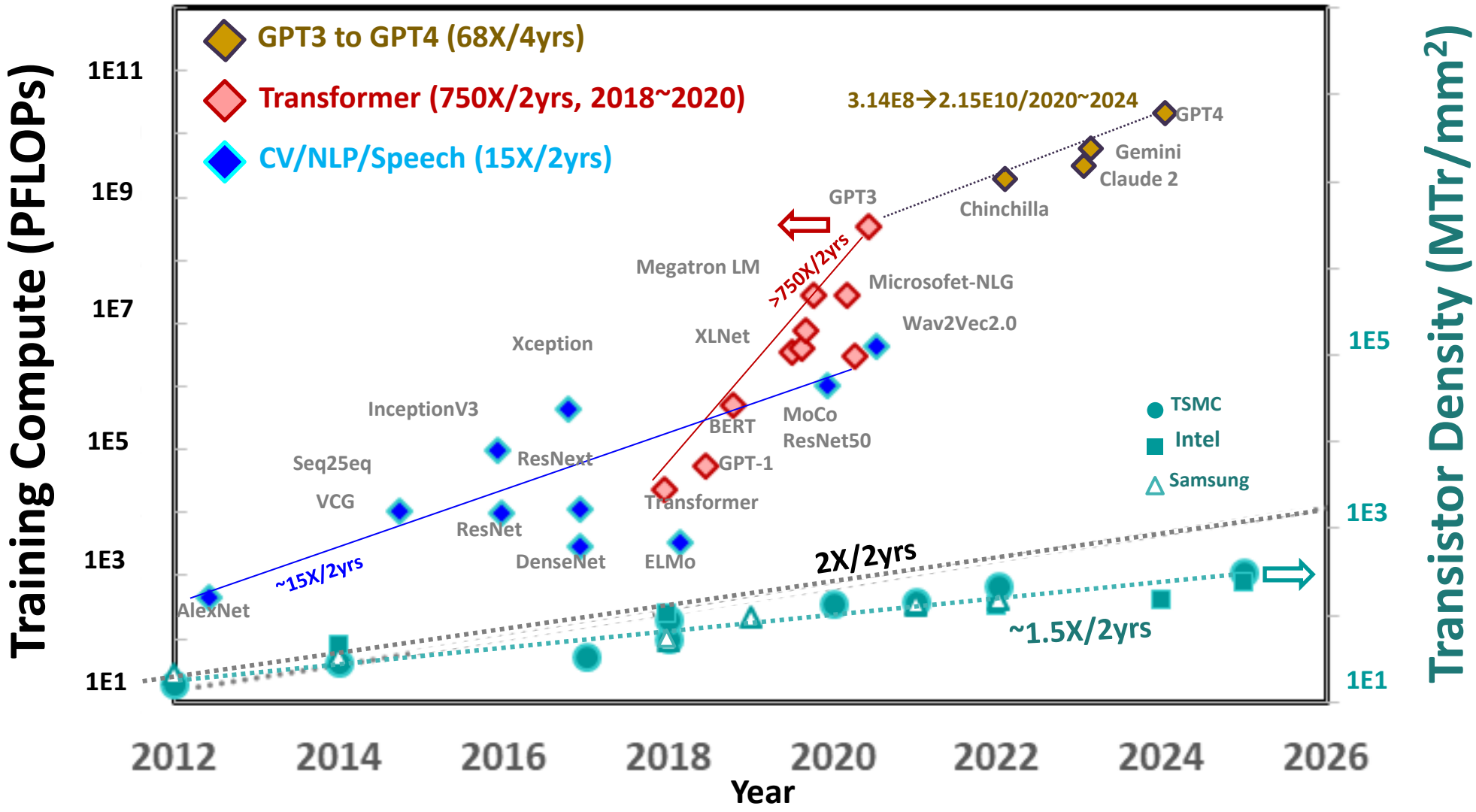
*Invited Paper in 2025 International VLSI Symposium on Technology, Systems and Applications, (VLSI-TSA) https://ieeexplore.ieee.org/document/11046689

The 2nd Path Moore's Law Scaling* Improved Part I: Optimized PPACT with Relaxed Litho by Device Design, eg. A CMOS Invention Can Deliver Better Scaling Results (Use 3D Sentaurus Modeling)

Peak Quoted Transistor Density For Standard Cell Optimization (MTr/mm^2 :Million Transistor per Area)						
Technology Node	**Company A**	**Company B**	**Company C**	**If obey Moore's Law**	**Node By Authors**	**NuFET**
22nm		18.8				
16nm	28.9 →			(28.88)		
14nm		44.7	33.3	37.72		
12nm	33.8			51.30	12nm (Nu12)	~67
10nm	52.51		51.8	73.93		
8nm (~12nm x 0.7 by Authors)			95.1	115.52	8nm (Nu8)	150
7nm	91.2(N7)/113.9(N7+)	100.8		150.88		
5nm	138.2			295.73	5.6nm (Nu5.6)	300
4nm	143.7	123.4		462.08	4nm (Nu4)	376
3nm	216(N3E)/224(N3P)	148	150 (MBCFET)	821.48	2.8nm (Nu2.8)	450
2nm	313 (GAAFET)	238 (Ribbon FET)		1848.32	2nm (Nu2)	639 (NuGAA)
1.4nm	391 (GAAFET)			3771.86		

*Invited Paper in 2025 International VLSI Symposium on Technology, Systems and Applications, (VLSI-TSA) https://ieeexplore.ieee.org/document/11046689

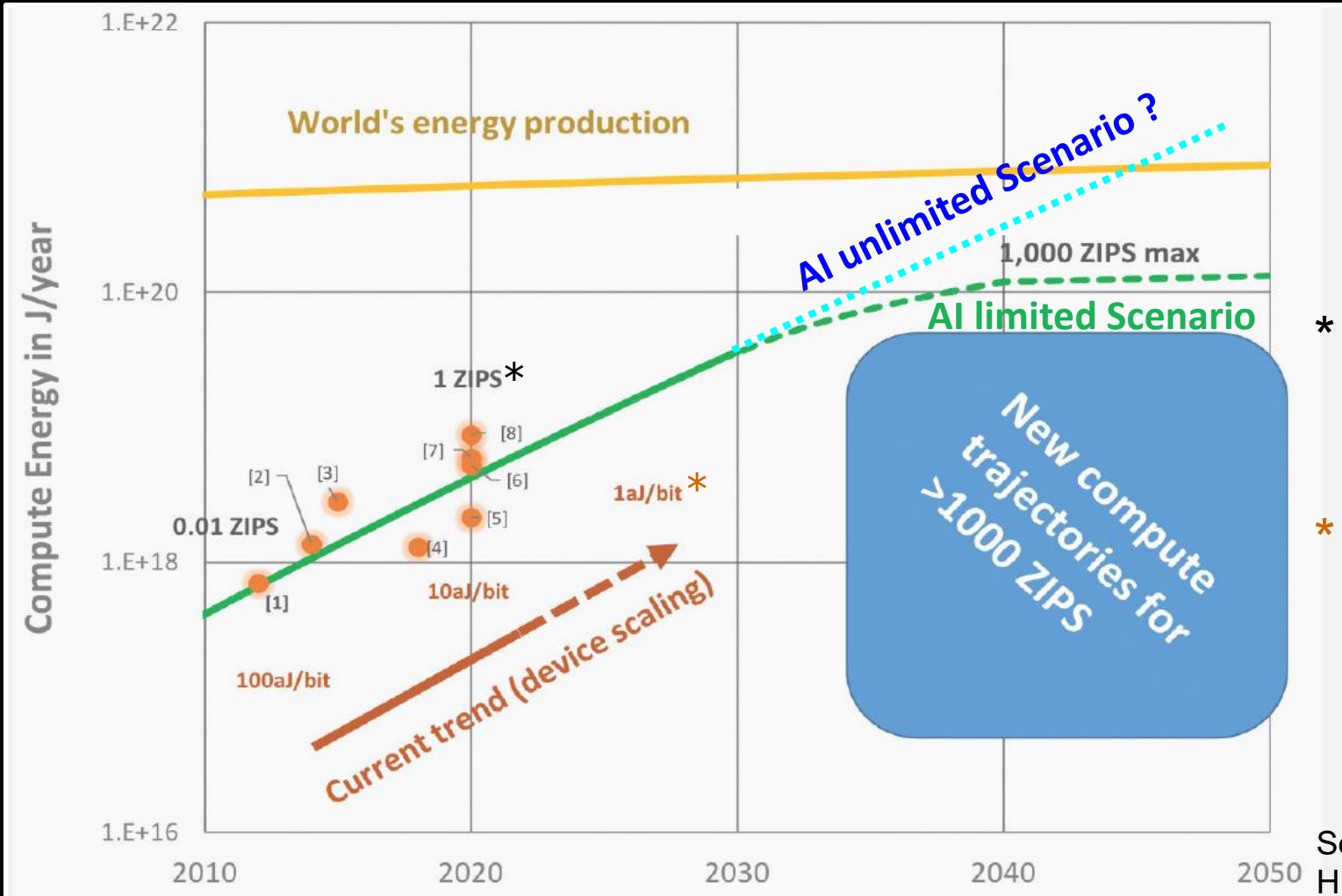
AI Training Compute Rate is Outpacing Silicon Transistor Scaling Rate



*Ref: "AI and Memory Wall", Berkeley AI Research (BAIR) Lab, arXiv:2403.14123v1 Mar 2024; Etron 2025**

*Invited Paper in 2025 International VLSI Symposium on Technology, Systems and Applications, (VLSI-TSA) https://ieeexplore.ieee.org/document/11046689

Global Energy Production Can NOT Meet Ever Rising Energy Demands for Computing



* ZIPS (10^{21} compute instructions per second)

* aJ: attojoule (10^{-18} Joules)

Source: SRC Jan 2021; After Nicky Lu, HIR 5th Annual Conference 2022*

Source: Invited Paper IEEE The 83rd DRC (Device Research Conference), June 23, 2025 (Plenary Talk) https://ieeexplore.ieee.org/document/11105734

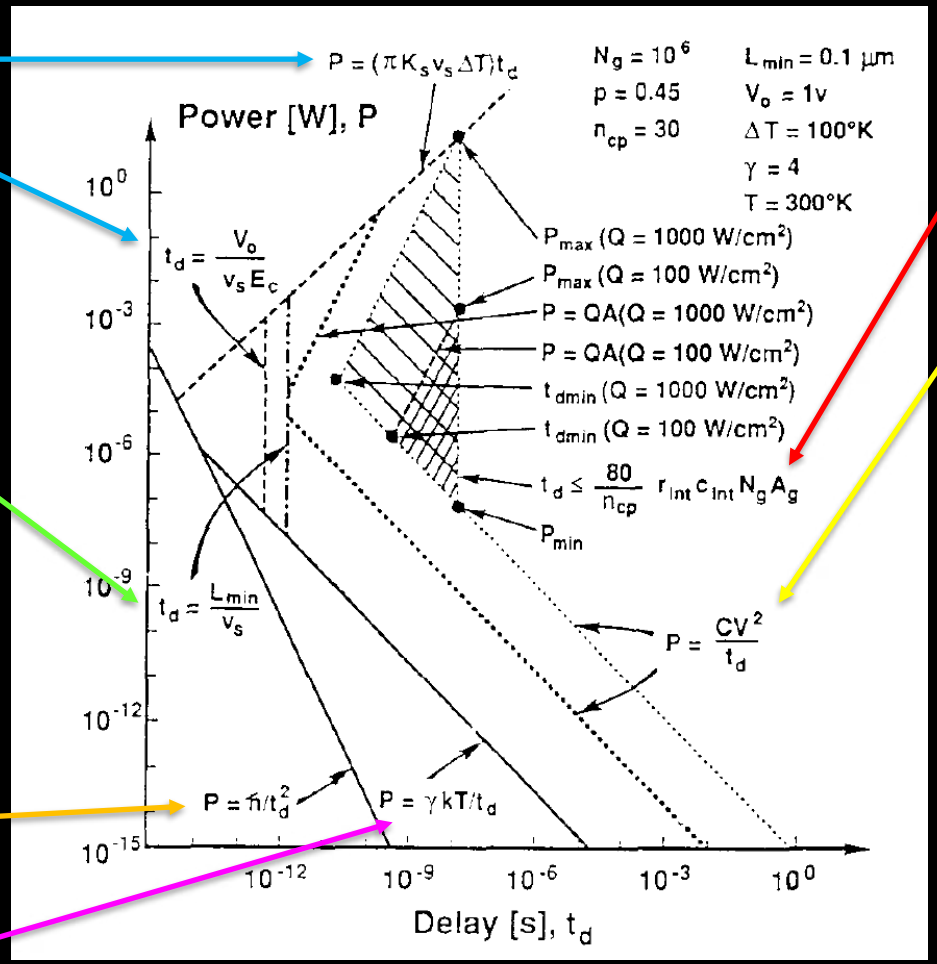
Study of Ultimate Limits of Semiconductor and IC

Meindl's Analyses* : Holistic and Hierarchical Methodology

2. Material limit

3. Device limit

1. Fundamental Limit
   - Quantum limit
   - Thermodynamic limit



5. System limit

4. Circuit limit

6. For future AI, application constraints may guide the selection of the most suitable technologies for specific needs — eg. 2D CMOS for certain special machines, or quantum computing at extremely low temperatures, etc.. So **Application Limit** must be studied. --- (Lu's Projection)

* James Meindl, Stanford, RPI, Georgia Tech, study from 1978 to 2018 ; ISSCC 1984, 1993, 2001

Source: Invited Paper IEEE The 83rd DRC (Device Research Conference), June 23, 2025 (Plenary Talk) https://ieeexplore.ieee.org/document/11105734

A fundamental limit on signal energy transfer during an irreversible binary switching is

$$kT(\ln 2) \sim 2.87\times10^{-21}\ Joule@300k$$
$$\sim 3.83\times10^{-24}\ Joule@0.4k$$

k Boltzmann's constant; *T* absolute temperature.

Today a 10nm/3nm switching device consumes

$$\sim 1E^{-15}\ /\ \sim 5E^{-16}\ Joule@0.75V$$
$$\sim 2E^{-17}\ /\ \sim 1E^{-17}\ Joule@0.104V$$

*After von Neumann (1957), Landauer (1961)/Buret (2012), Meindl (2000)

With FET Plus Si4.0 Heterogeneous Integration and AI

The Bicentennial Awaits New Triode Architecture & Invention

Two paths can carry computing & AI through another century - extend Silicon, SiO_2 or break beyond it

ON ONE HAND, A Sure Growing IC Industry,

Si4.0 Enhanced Moore's Law

Symbiotic Growth of AI & Heterogenous Integration, Silicon 4.0, Is Generating An Intelligence^N-Driven Huge Economic Boom

ON THE OTHER HAND, An Explosive Industry Boom

Beyond CMOS FET

Inventions Desperately Required — Derive A New Class of Solid-State Devices, Grounded in Quantum Physics

> **"When can humanity invent ultra–power-efficient, magically swift solid-state devices based on quantum physics, that can operate at room temperature on Earth and in the freezing environment of Space?"**

Physics already permits it — reversible computation that dissipates little energy, a foundation we now build on:

von Neumann	Landauer	Buret
1957	1961	2012
the energy bound	*cost = irreversibility*	*reversibility, energy-free*

Thank You!

Through Semiconductor ICs
We Connect People to
Realize Tech Dreams !