UCLA Electrical and Computer Engineering

Fall 2023: ECE M216A

Design of VLSI Circuits and Systems

Prof. Dejan Marković

ee216a@gmail.com

Teaching Staff, Office Hours

Prof. Dejan

(phonetic: Deyan)

Marković

Office hours

56-147E Eng-IV Bldg.

- Campus: Tue, Thu 1:15-2:30pm
- MSOL: Tue, Thu 12:45-1:30pm

TA:

Gokul

Kuppuswamy

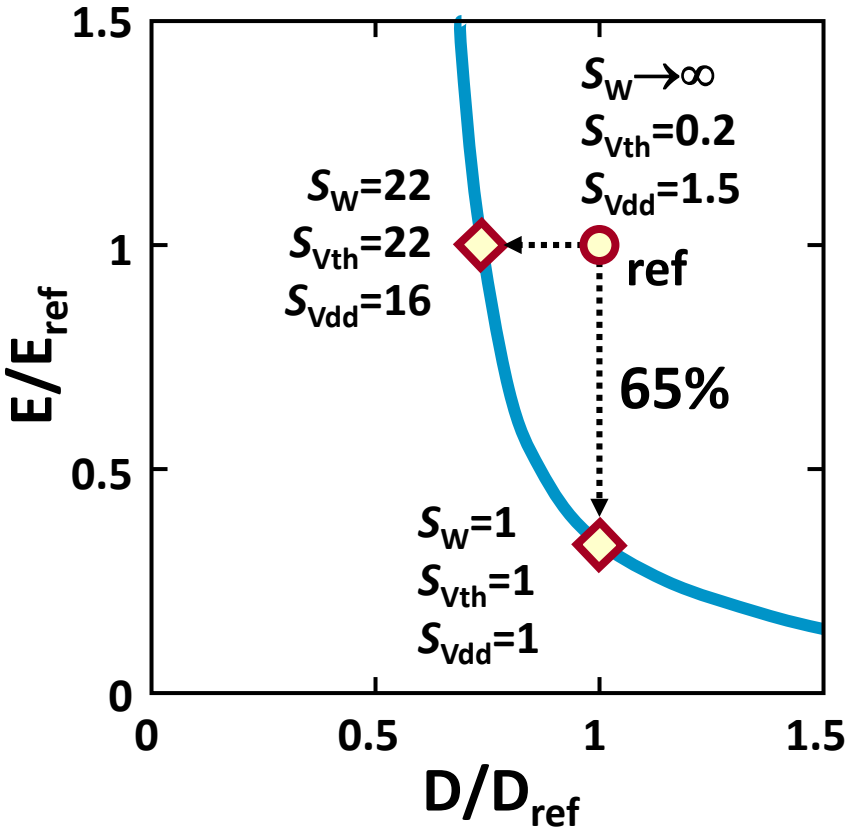
Office hours

Graduate Lounge,

5th Floor, Eng-IV Bldg.

- Campus: TBD
- MSOL: TBD

Elevator Pitch



Modeling and design

of energy-delay optimal

VLSI circuits and systems

Background

Familiarity with

- **Digital ICs**
- **VLSI design**
- **CAD tools**

ECE 115C vs. ECE M216A

circuits

115C (intro)

- Simple transistor and circuit models
- Circuit design styles
- Logic gate design
- Custom blocks (adders)

circuits + systems

216A (advanced)

- Several transistor and circuit models
- Constrained design (Power, Area, Speed)
- RTL, chip synthesis
- Test, packaging

Background: ECE 115C Material

ECE 115C

Lectures 2-5

(2) MOS IV Model
(3) MOS RC Model
(4) Inverter VTC
(5) Propagation Delay

ECE M216A Goals (1/2)

Understanding the basic building blocks of VLSI

- Transistors/Wires
- Logic Gates and Layout
- Datapath Blocks

Be able to conceptually model a system

- Logic Optimization
- State Machine Design (RTL)

ECE M216A Goals (2/2)

Be able to build a system (using a subset of the tools)

- Verilog Modeling
- Synthesis, Place and Route

Understanding the constraints and tradeoffs

- Delay analysis (gates and interconnects)
- Clocking methodology
- System integration issues
  (Power/Ground routing, Noise)

Course Objective and Key Outcomes

Energy-performance optimal design:

- **Outcome 1:** energy and delay **models**
- **Outcome 2:** circuit energy-delay **optimization**
- **Outcome 3:** high-level description, **chip synthesis**

VLSI Design Challenges

- **Power-limited** performance
- Limited technology improvements
- Methods for energy efficient design
- Flexibility (multi-mode, multi-standard)

Course Outcomes

1. CMOS scaling
2. RC transistor model
3. Static CMOS logic gate design
4. Design with HDL (Verilog)
5. Dynamic and leakage power model
6. Power and delay calculation
7. Logical effort and gate sizing
8. Energy-delay tradeoff analysis
9. Clocking methodologies and timing analysis
10. Design automation using logic synthesis
11. State machine design (ASM or FSM)

Online Resources

Bruinlearn is Your Class Portal

- **Joint for “on campus” and MSOL**
- **Links to Piazza, Gradescope, etc.**

ee216a@gmail.com

- **Submission of assignments**
- **Personal queries**
  - Before you email, think of **WHY** can't you post the question on Piazza

Course Material

- Lecture notes
- Homework
- CAD tutorials
- Class project
- Selected papers from IEEExplore (http://ieeexplore.ieee.org)

Books (Optional)

ECE 115C textbook

- J. Rabaey, A. Chandrakasan, B. Nikolić, *Digital Integrated Circuits: A Design Perspective*, (2nd Edition), Prentice Hall, 2003.

Another popular VLSI textbook

- N. Weste, D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, (3rd Edition), Addison Wesley, 2004.

Journals and Conferences

Circuits

- IEEE Journal of Solid-State Circuits (JSSC)
- IEEE International Solid-State Circuits Conference (ISSCC)
- European Solid-State Circuits Conference (ESSCIRC)
- Symposium on VLSI Circuits (VLSI)
- Custom Integrated Circuits Conference (CICC)
- Other conferences and journals

CAD

- IEEE Transactions on Computer-Aided Design (TCAD)
- International Conference on Computer Aided Design (ICCAD)
- Design Automation Conference (DAC)

Schedule and Syllabus

Weeks 1-5: Circuits	
1	Intro, Scaling
	MOS, Delay Models
2	Logic Design
	Logical Effort
3	Adders
	Verilog 1
4	Latches and FFs
	Verilog 2
5	Clocking Methods
	Midterm Review
	MSOL Midterm 11/4-6

Weeks 6-10: Systems	
6	Midterm (in-class) 11/6
	Timing Analysis
7	Logic Synthesis
	Power Model
8	E-D Optimization
	Physical Synthesis
9	Packaging
	Test
10	CPU, GPU, FPGA, ASIC
	Future SoCs
11	Final: 12/15, 8-11am MSOL Final: 12/15-17

Fall Quarter 2023 Schedule

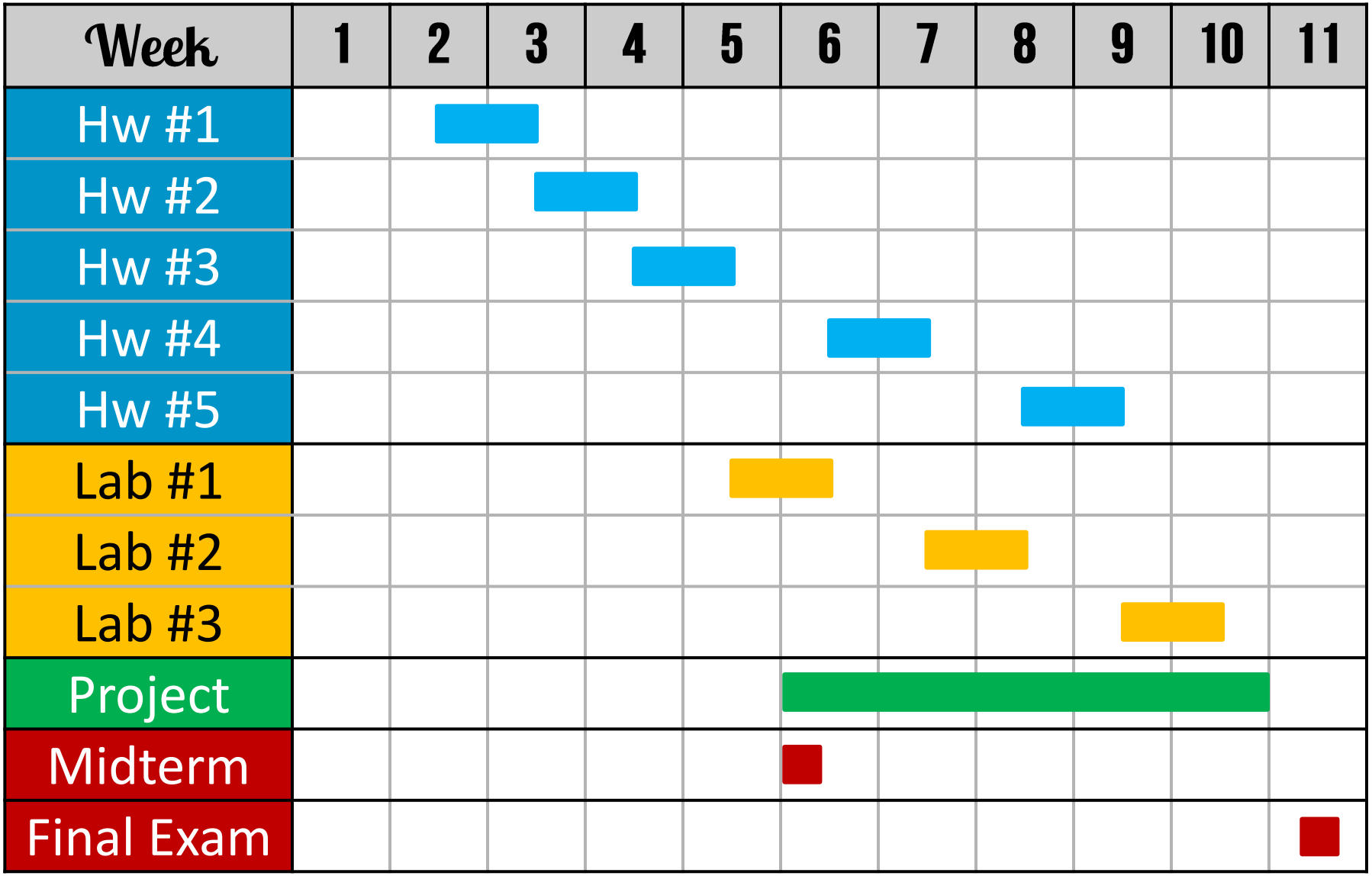
- **Our final exam is on Friday, December 15, 2023 (at 8am)**

FALL QUARTER 2023	
Quarter begins	Monday, September 25
Instruction begins	Thursday, September 28
Study List deadline (becomes official)	Friday, October 13
Veterans Day holiday	Friday, November 10
Thanksgiving holiday	Thursday-Friday, November 23-24
Instruction ends	Friday, December 8
Common final exams	Saturday-Sunday, December 9-10
Final examinations	Monday-Friday, December 11-15
Quarter ends	Friday, December 15
Christmas holiday	Monday-Tuesday, December 25-26
New Year's holiday	Monday-Tuesday, January 1-2, 2024
Winter campus closure	TBD

Grading Policy & Organization

15%	• 5 homework sets
6%	• 3 CAD labs
30%	• Project
24%	• Midterm
25%	• Final

Gantt Chart


Week
1
2
3
4
5
6
7
8
9
10
11
Hw #1
Hw #2
Hw #3
Hw #4
Hw #5
Lab #1
Lab #2
Lab #3
Project
Midterm
Final Exam

Homework Topics

1. • Scaling, Models, Logical Effort
2. • Logic Design, Verilog ALU
3. • FFs, Verilog FSM
4. • Clk, Timing Analysis
5. • Power, E-D Optimization

CAD Labs

1. Verilog testbench
2. PrimeTime, PrimePower
3. UPF (Multi-V_{DD} and Clk)

Class Project

- **Team project (3 partners)**
  - Start teaming up
- **Topic TBD**
  - Details in Week 5

Generic Technologies

Cadence

45nm PDKs + library

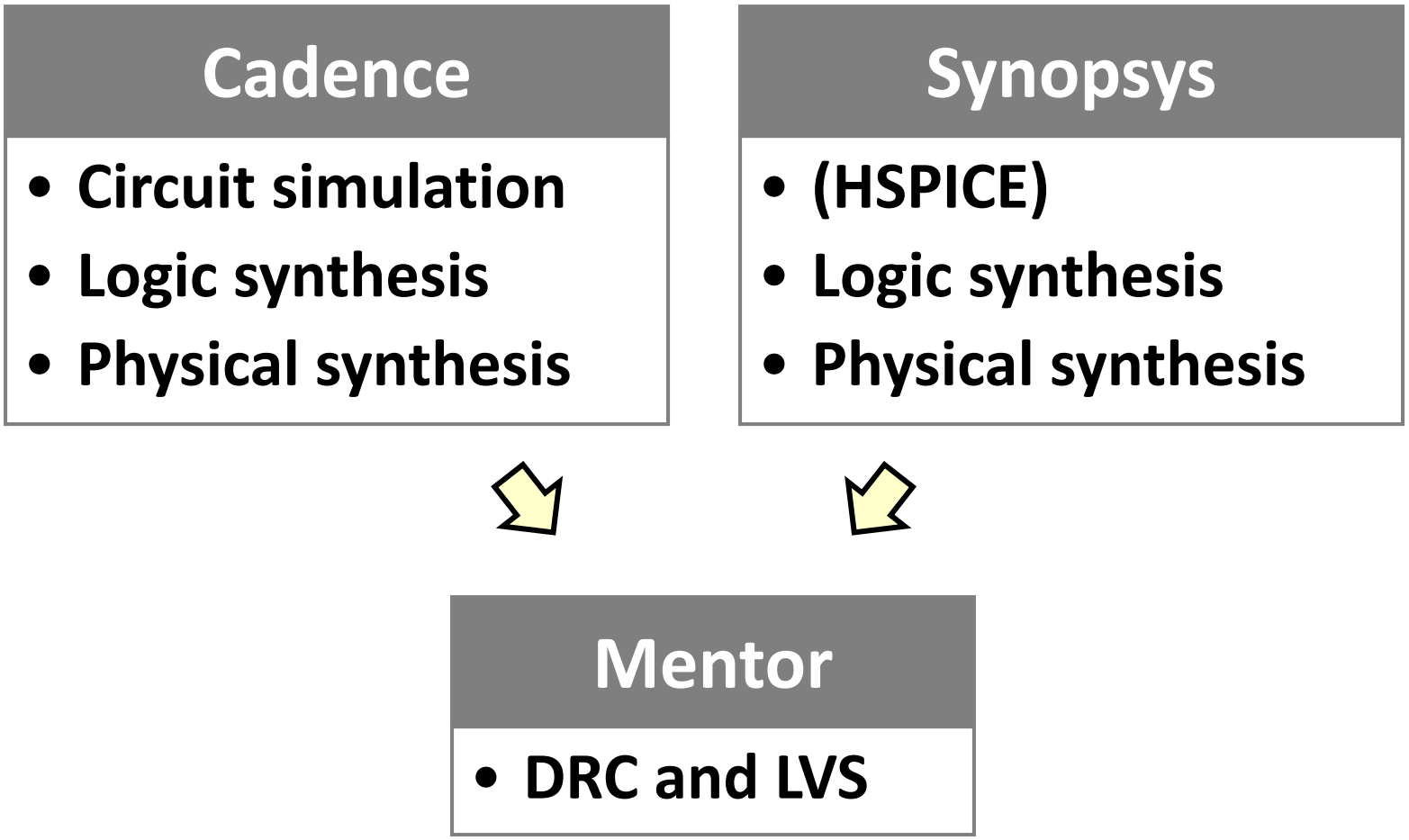
- **PDK: Cadence 45nm GPDK**
- **PDK + lib: Nangate open cell library (NCSU FreePDK, ASU PTM)**

Synopsys

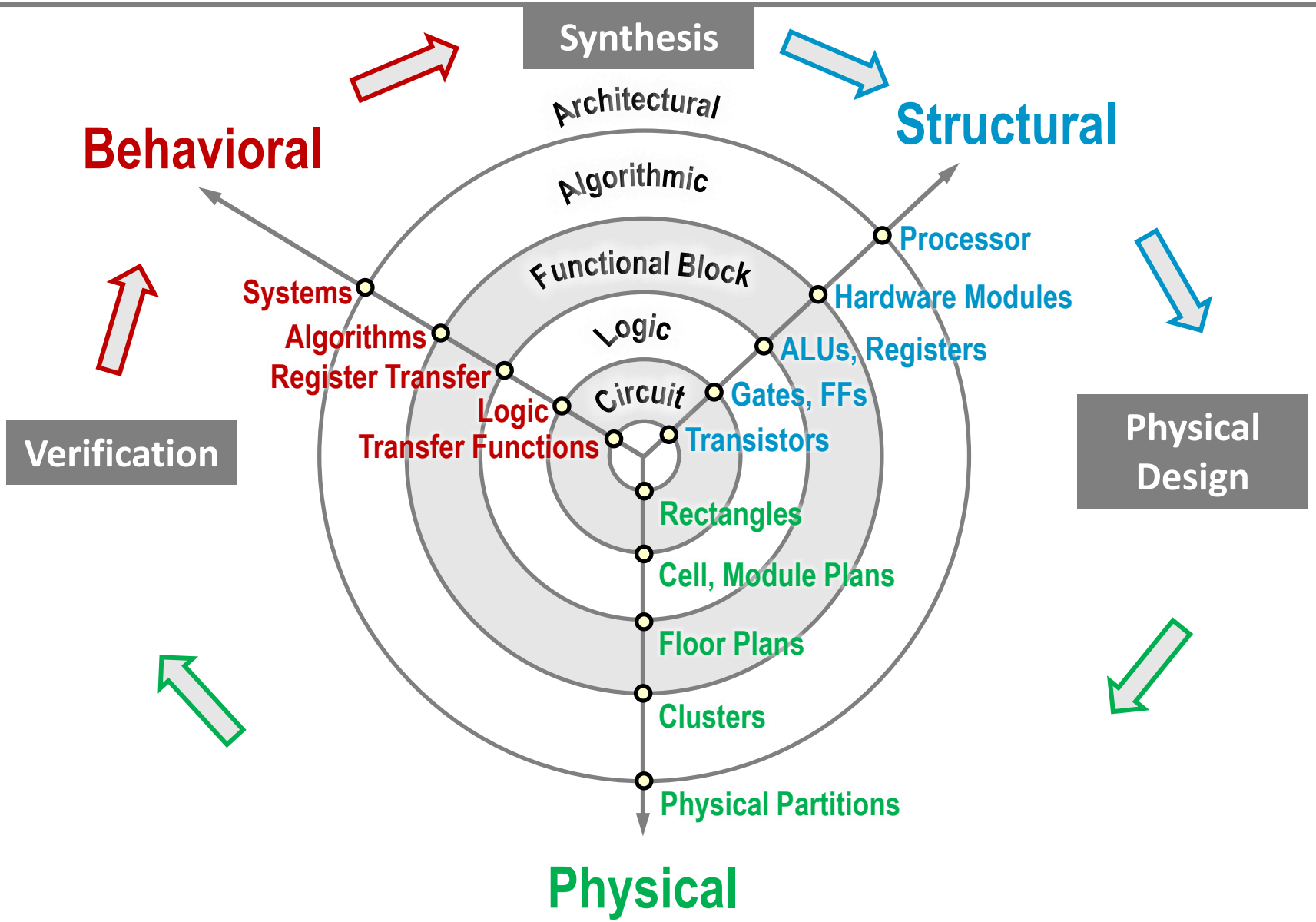
32/28nm EDK + libraries

- **EDK + libs: Synopsys kit and libs**
  - Std cell, I/O, mem, PLL, ref. designs

CAD Tools


Cadence
• Circuit simulation
• Logic synthesis
• Physical synthesis
Synopsys
• (HSPICE)
• Logic synthesis
• Physical synthesis
Mentor
• DRC and LVS

Design Description: Gajski-Kuhn Y Chart



Module

1

ECE M216A
Fall 2023

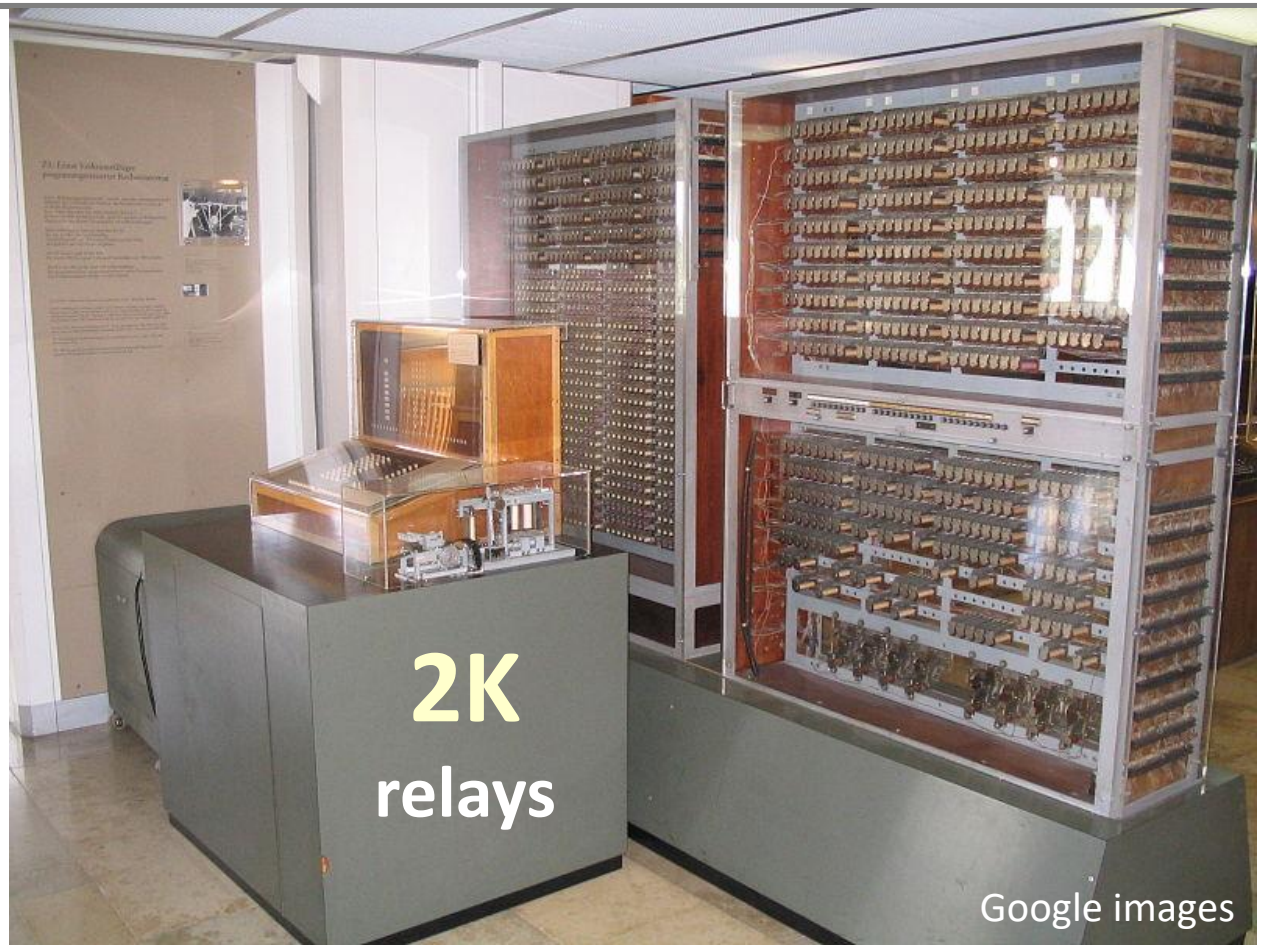
Introduction, CMOS Scaling

Prof. Dejan Marković

ee216a@gmail.com

The First Digital Electronic Computer

Zuse Z3
(1941)

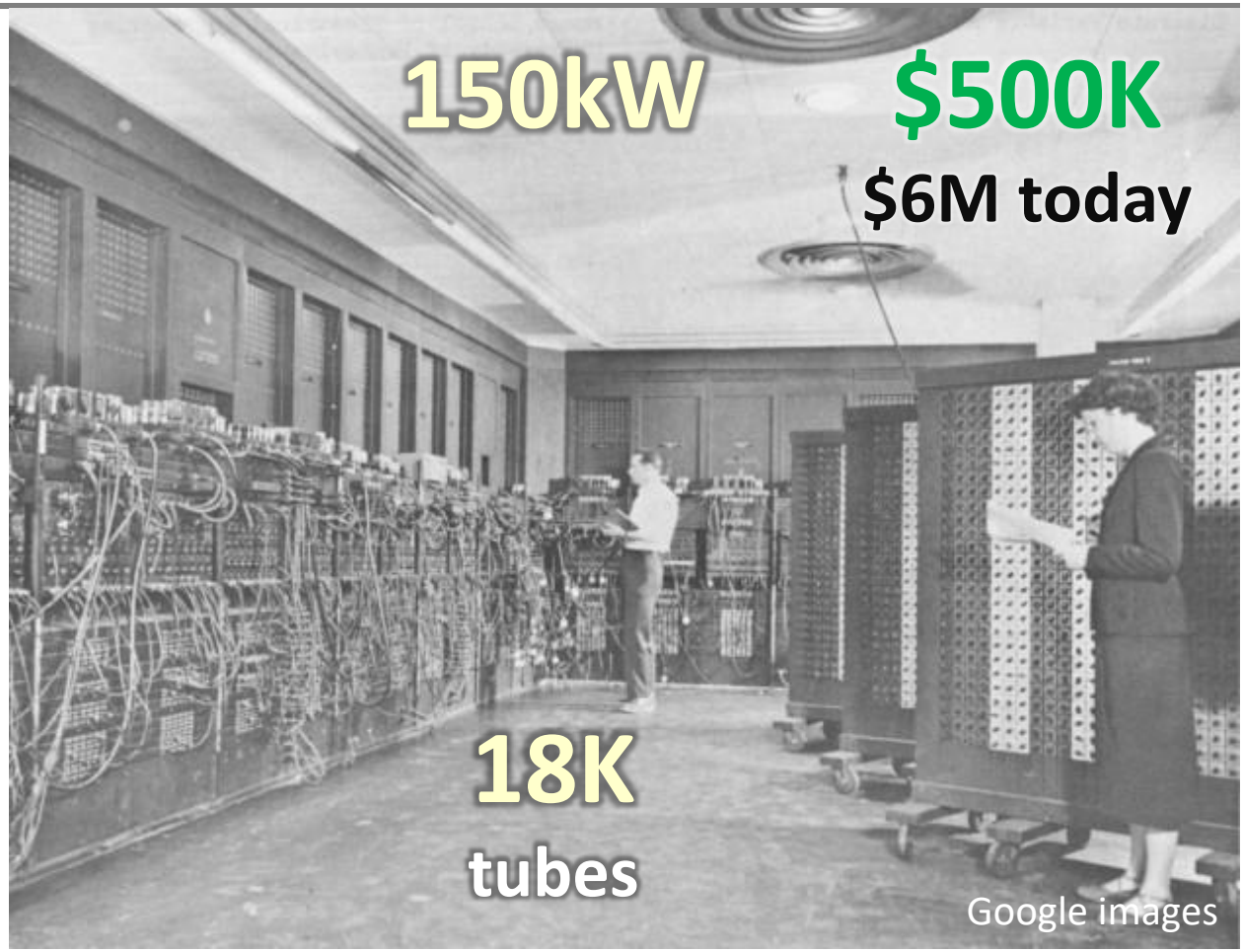


Binary
5 – 10 Hz
22b words

Device: Electromechanical relay

Five Years Later

ENIAC
(1946)

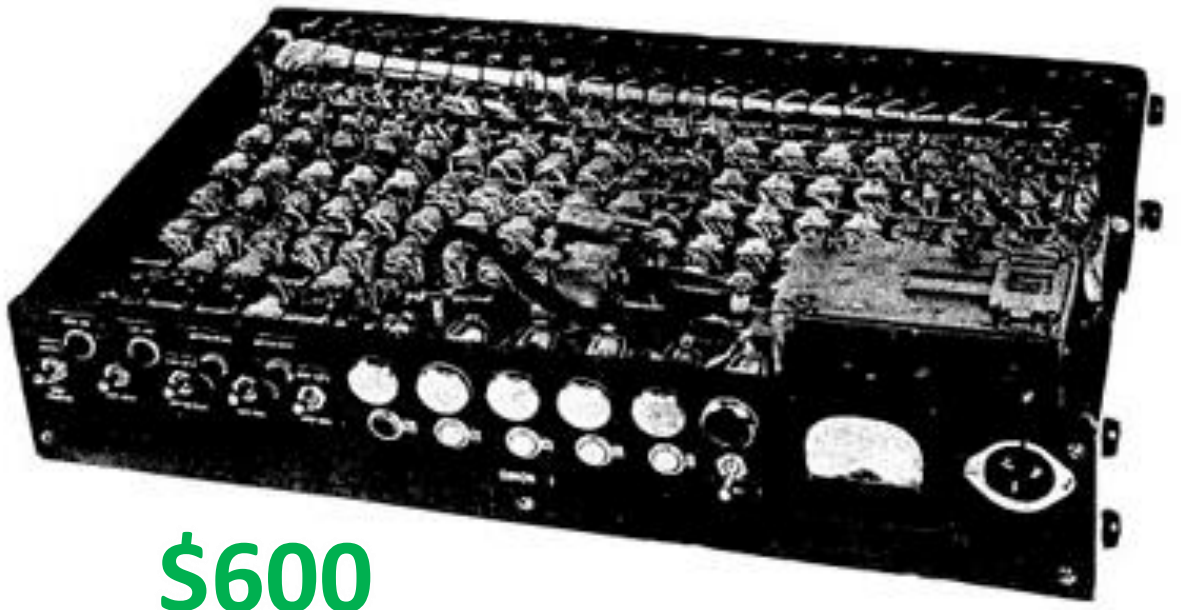


Decimal
5M joints
hand-soldered

Device: Vacuum tube

The First PC

Simon
(1950)



$600

Google images

4 ops:

+, −x, >, S

2b Reg/ALU

Device: Electromechanical relay

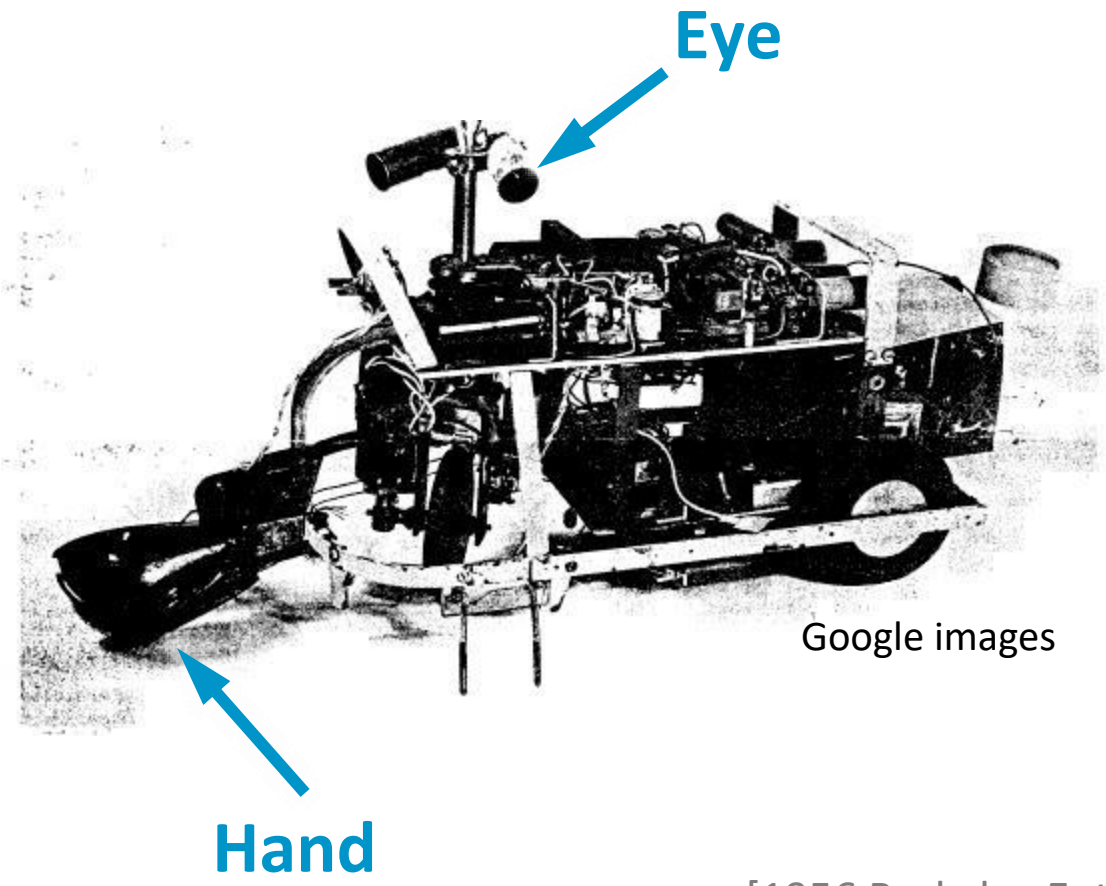
What is the Machine's Future?

Mr. Berkeley's answer:

"Simon has two futures. In first place Simon can grow. With another chassis and some wiring and engineering, the machine will be able to compute decimally. Perhaps in six months more, we may be able to have it working on real problems. In the second place, Simon may start a fad of building baby mechanical brains, similar to the hobby of building crystal radio sets that swept the country in the 1920's."

[1956 Berkeley Enterprises Report]

Squee (The Electronic Robot Squirrel)

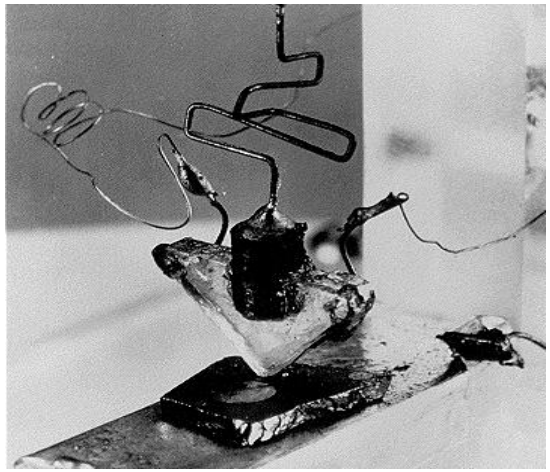


Google images

[1956 Berkeley Enterprises Report]

Integrated Electronics

Transistor

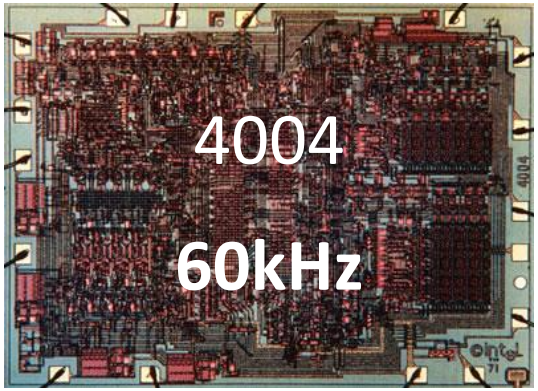


1948 (Bell Labs)

IC

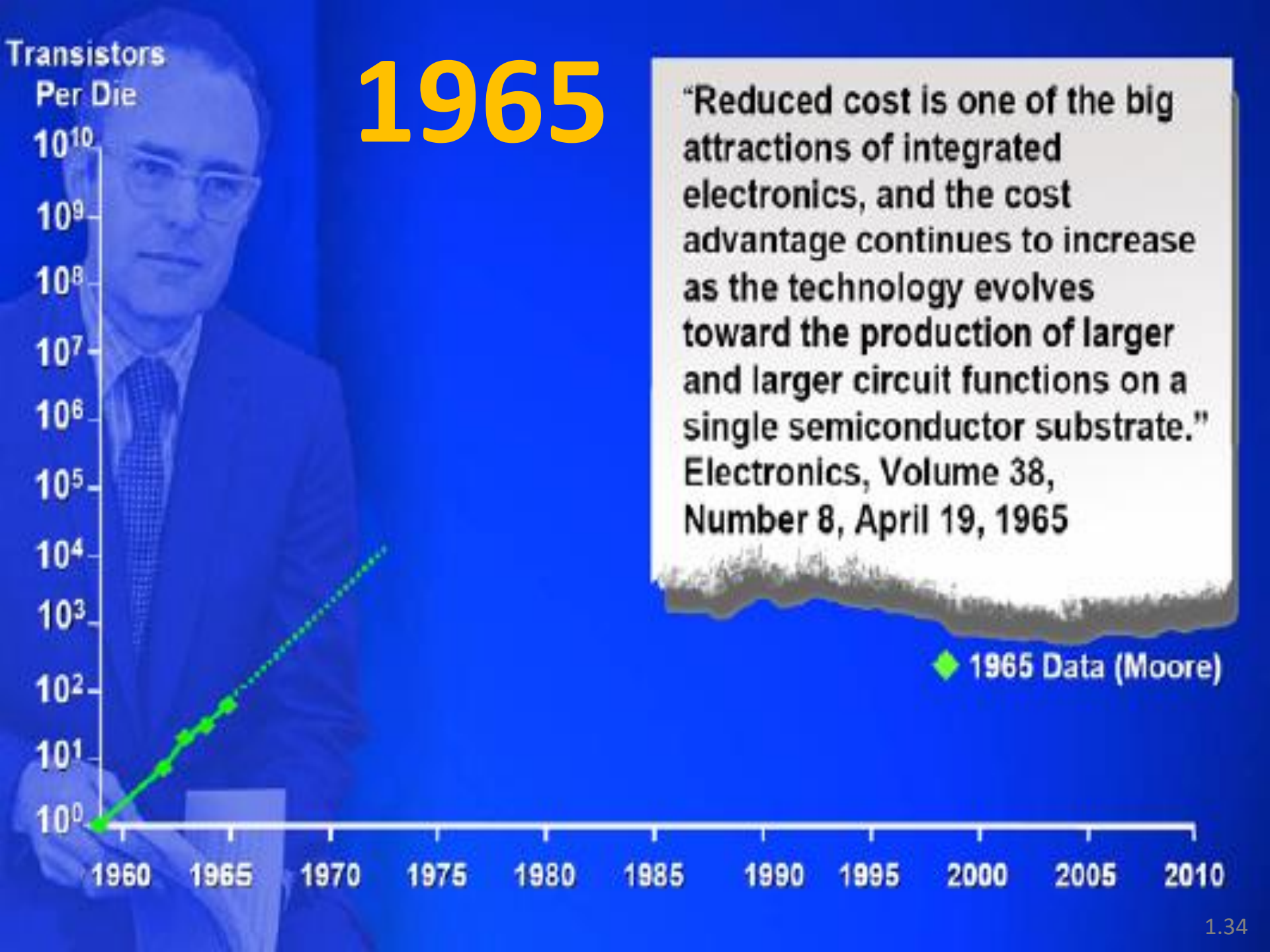
1958 (TI)

µP



1971 (Intel)

1965



"Reduced cost is one of the big attractions of integrated electronics, and the cost advantage continues to increase as the technology evolves toward the production of larger and larger circuit functions on a single semiconductor substrate."
Electronics, Volume 38, Number 8, April 19, 1965

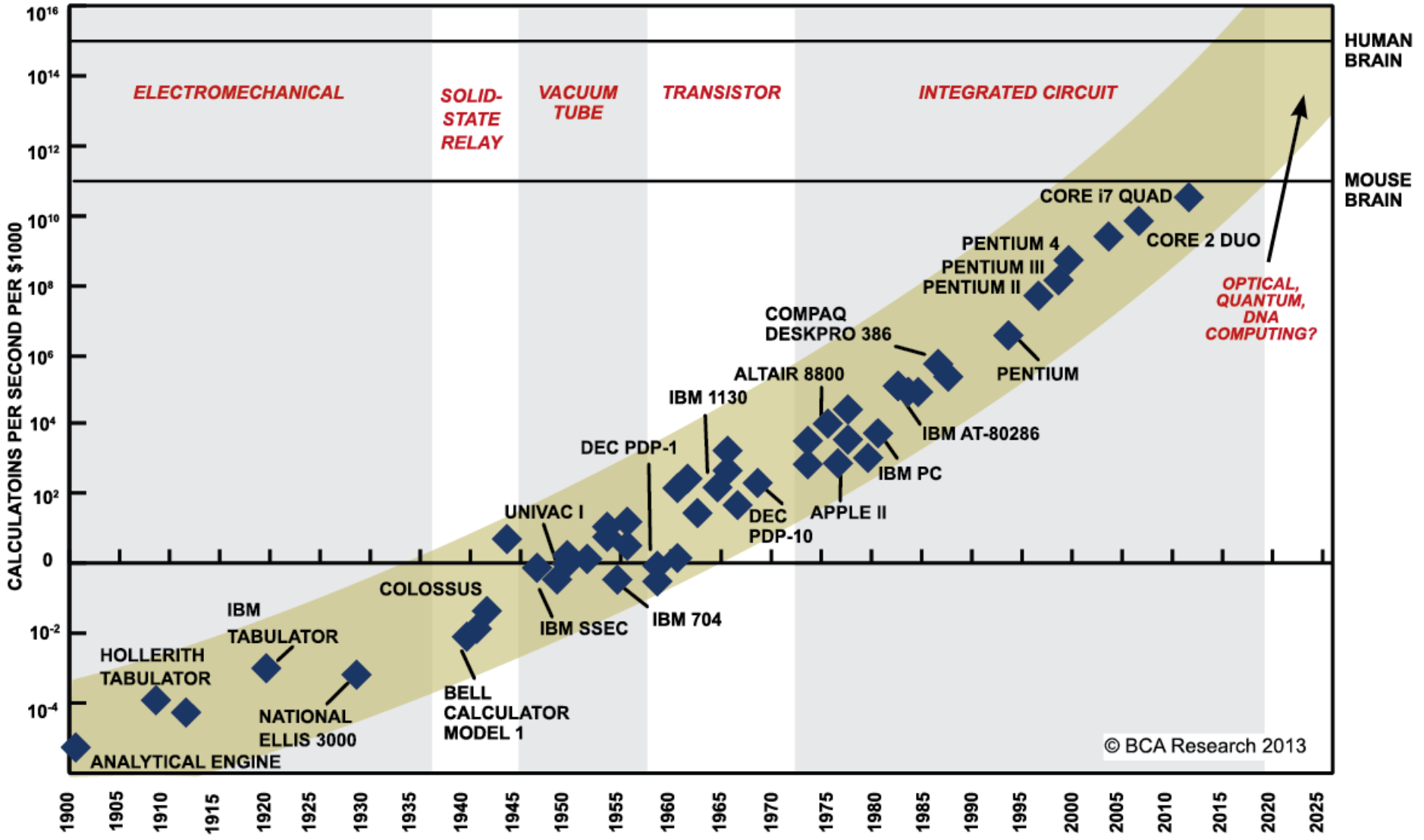
Moore's Law

- **In 1965, Gordon Moore noted that the number of transistors on a chip doubled every 18 to 24 months**

"The complexity for minimum component costs has increased at a rate of roughly a factor of two per year. Certainly over the short term, this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years. That means by 1975, the number of components per integrated circuit for minimum cost will be 65,000."

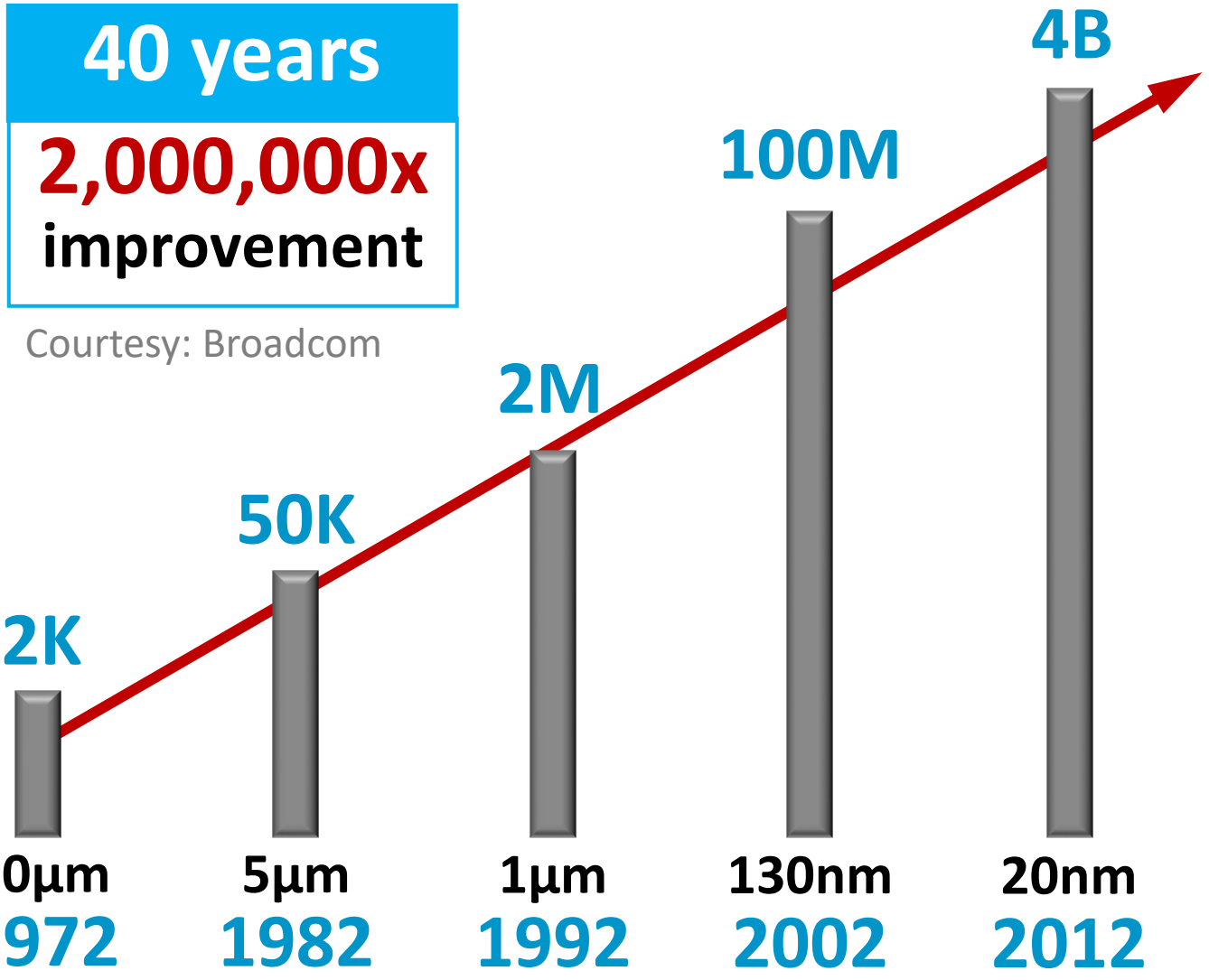
[G. Moore, Electronics, 1965]

120 Years of "Moore's Law"



SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, *THE VIKING PRESS*, 2006. DATAPOINTS BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

Transistors / cm^2



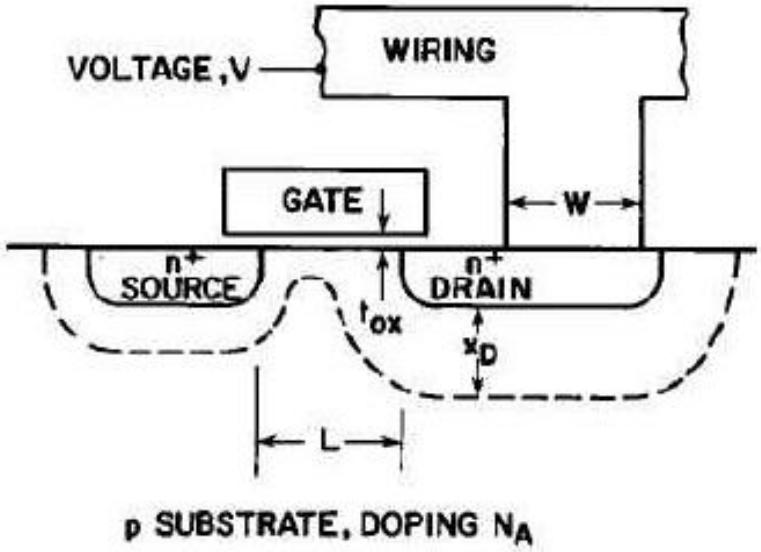
Scaling

Voltage: V_{DD}, V_T

Size: W, L, t_{ox}

Dennard's Classical MOSFET Scaling (1974)

Scaling Factor	Device or Circuit Parameter
$1/\kappa$:	Device dimension t_{ox}, L, W
κ :	Doping concentration Na
$1/\kappa$:	Voltage V
$1/\kappa$:	Current I
$1/\kappa$:	Capacitance $\varepsilon A/t_{ox}$
$1/\kappa$:	Delay time/circuit VC/I
$1/\kappa^2$:	Power dissipation/circuit VI
1 :	Power density VI/A



R. Dennard, JSSC, Oct 1974.

Constant E-field Scaling

Voltage and size scale by the same factor, S (S > 1)

- **E = V/L = constant**

Outcomes:

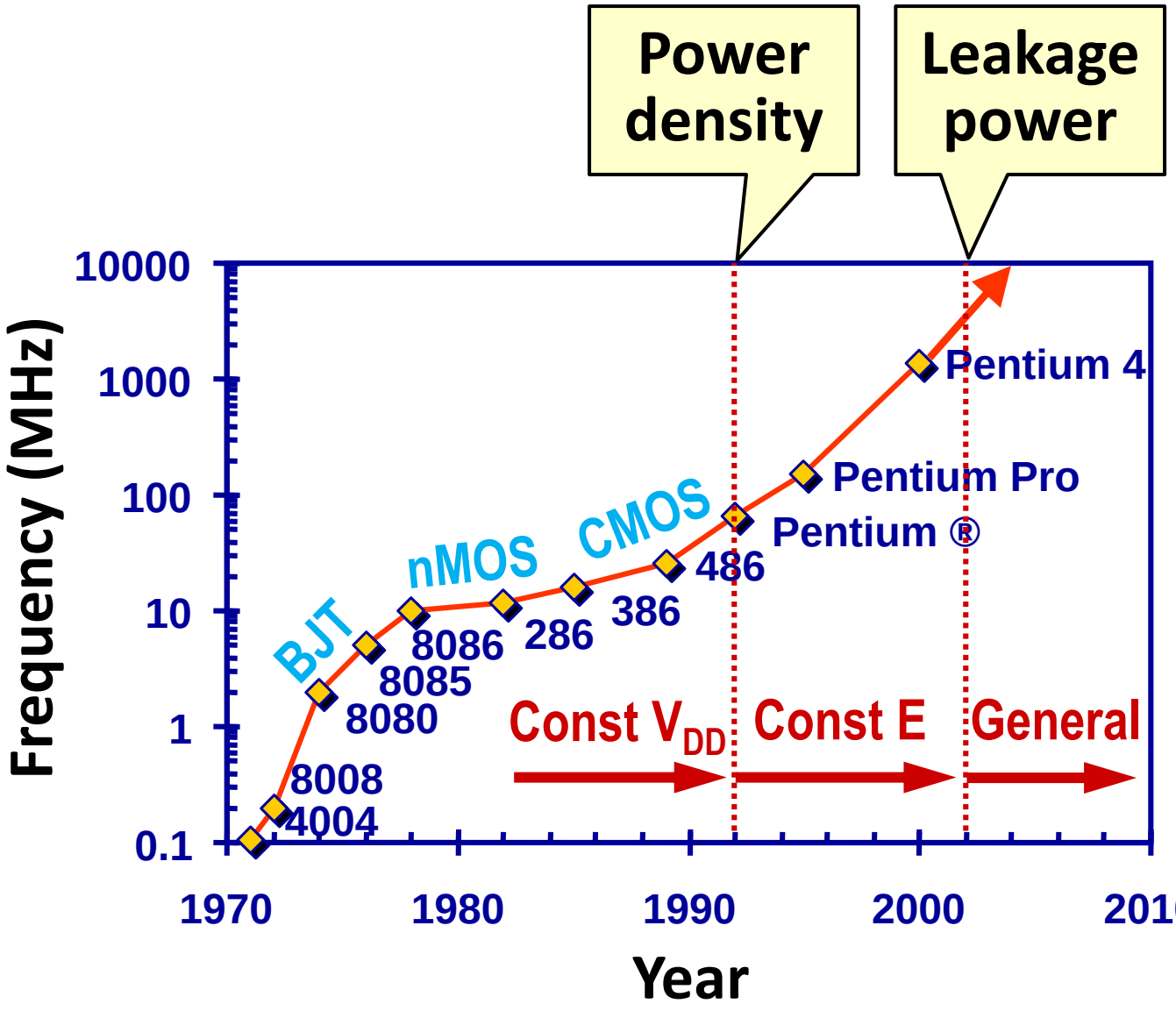
• More transistors/area	$1/S^2$
• Faster delay	$1/S$
• Lower energy/op	$1/S^3$

Problem: V_T scaling (exponential leakage)

Constant E-field Scaling

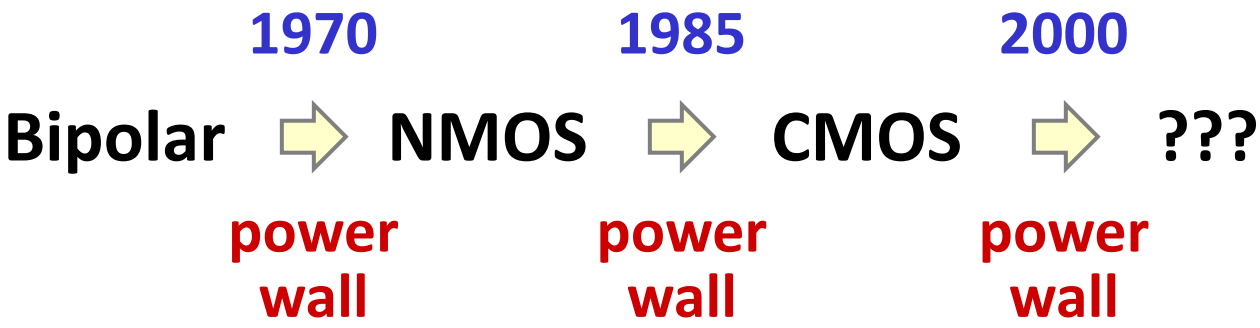
Ended at the 130nm node

Historical Scaling Trends



Courtesy:
S. Borkar
(Intel)

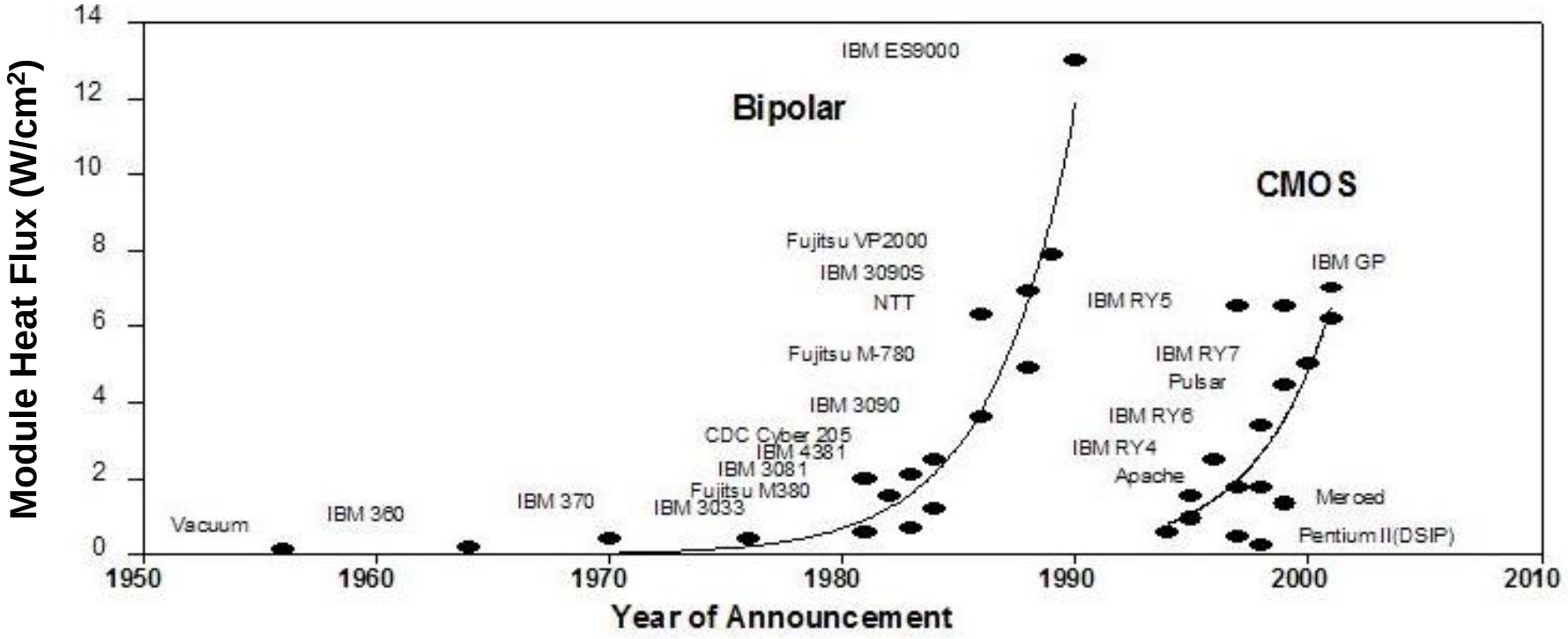
Technology Scaling is Power Driven



- **CMOS delivered better cost performance**
  - It was more energy efficient
  - It improved the integration level

Bipolar → Power Wall → CMOS

- **Technologies:** bipolar, nMOS, CMOS
- **Constant voltage scaling:** increasing power



Courtesy: Roger Schmidt (IBM)

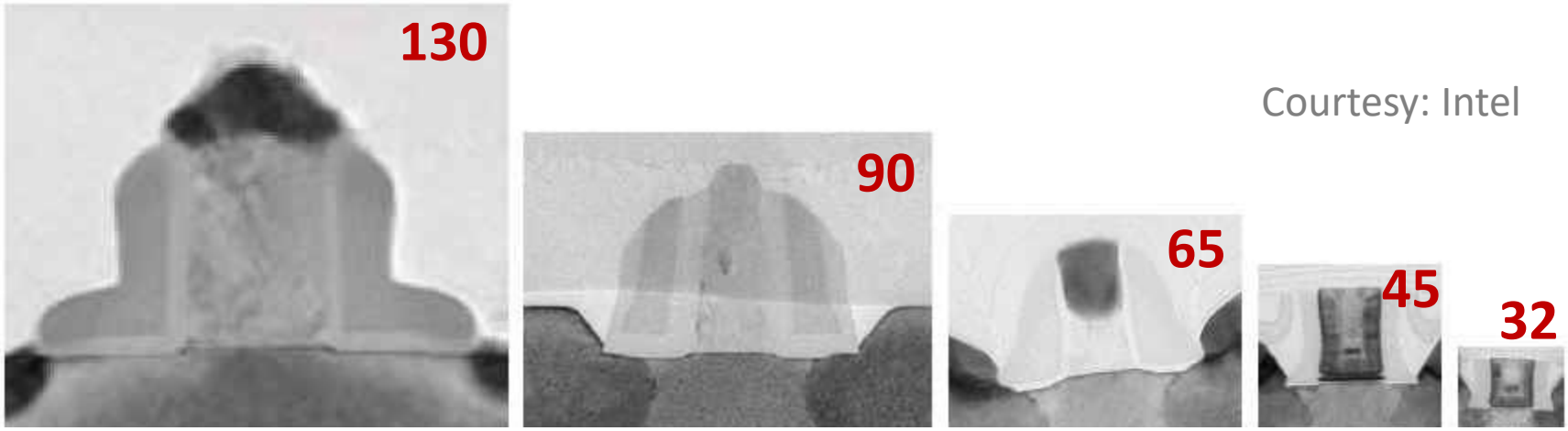
Scaling Scenarios: Fixed V, Fixed E, General

Parameter	Relation	Fixed V	Fixed E	General
W, L, t_{ox}		***1/S***	***1/S***	***1/S***
V_{DD}, V_T		***1***	***1/S***	***1/U***
Area/Device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	**S**	**S**	**S**
C_{gate}	C_{ox} WL	**1/S**	**1/S**	**1/S**
k_n, k_p	C_{ox} W/L	**S**	**S**	**S**
I_{sat} (*)	C_{ox} WV	**1**	**1/S**	**1/U**
Current Density	I_{sat} / Area	**S^2**	**S**	**S^2/U**
R_{on}	V / I_{sat}	**1**	**1**	**1**
Intr. Delay	R_{on} C_{gate}	**1/S**	**1/S**	**1/S**
Power	I_{sat} V	**1**	**$1/S^2$**	**$1/U^2$**
P Density	Power/Area	**S^2**	**1**	**S^2/U^2**

(*) I_{sat} model (derive from 115c/lec-2/#13)

Challenges in Scaling

Fixed E (Past)	General (Now)
• **Scaling reduced cost**	• Scaling reduces cost
• **Scaling increased performance**	• **Materials & devices increase performance**
• **Performance constrained**	• **Power constrained**
• **Active power dominates**	• **Standby power matters**



Courtesy: Intel

General Scaling

Size scaling S > Voltage scaling U

Voltage scaling slowing down

- **V_T determined by leakage**
- **t_{ox} also set by leakage**

Some techniques to mitigate U < S

- **Strained silicon (90nm)**
- **High-K metal gate (45nm)**
- **3D transistors (22nm)**

Chip Utilization Drops Every Generation

Power-limited scaling = DARK silicon

Parameter	Classical const-E scaling	Leakage-limited scaling
Threshold, V_T	**1/S**	**1/U**
Supply, V_{DD}	**1/S**	**1/U**
Quantity, Q	S^2	S^2
Frequency, F	S	S
Capacitance, C	1/S	1/S
Power, P	**1**	**S^2/U^2**
Utilization = 1/P	**1**	**U^2/S^2**

$P \propto Q \cdot F \cdot C \cdot V_{DD}^2$

Utilization drop: **S^2/U^2**

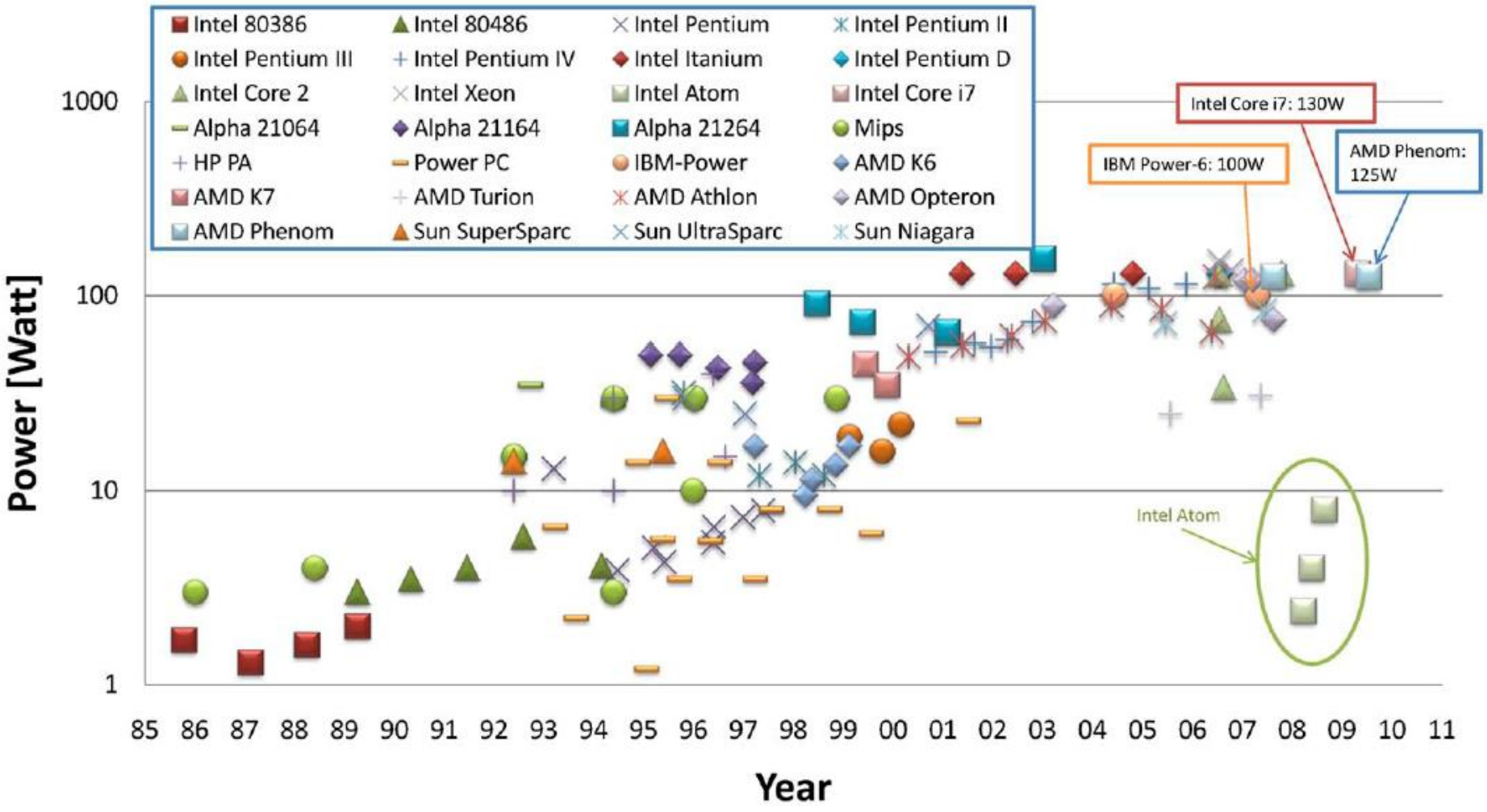
The Utilization Wall (Assume U = 1)

- **Voltage scaling is broken (U < S)**
- **Improvements for a fixed chip size**
  - Computing capability: **2.8x**
    - Transistor count: 2x
    - Operating frequency: 1.4x
  - Energy efficiency: **1.4x**

2x shortfall ⇨ **DARK** silicon

Power-limited scaling

CPUs Have Reached the 130W Power Limit

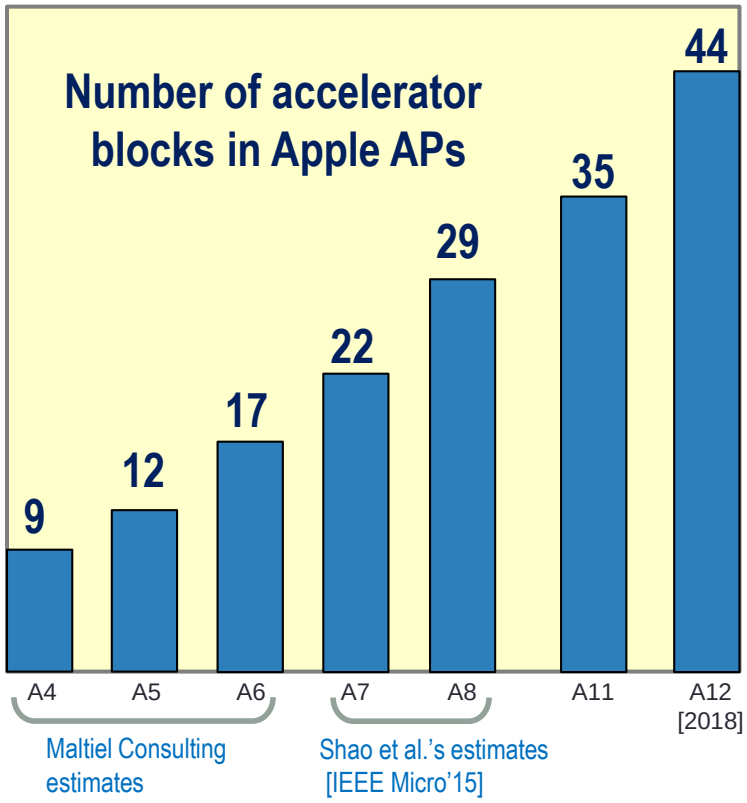


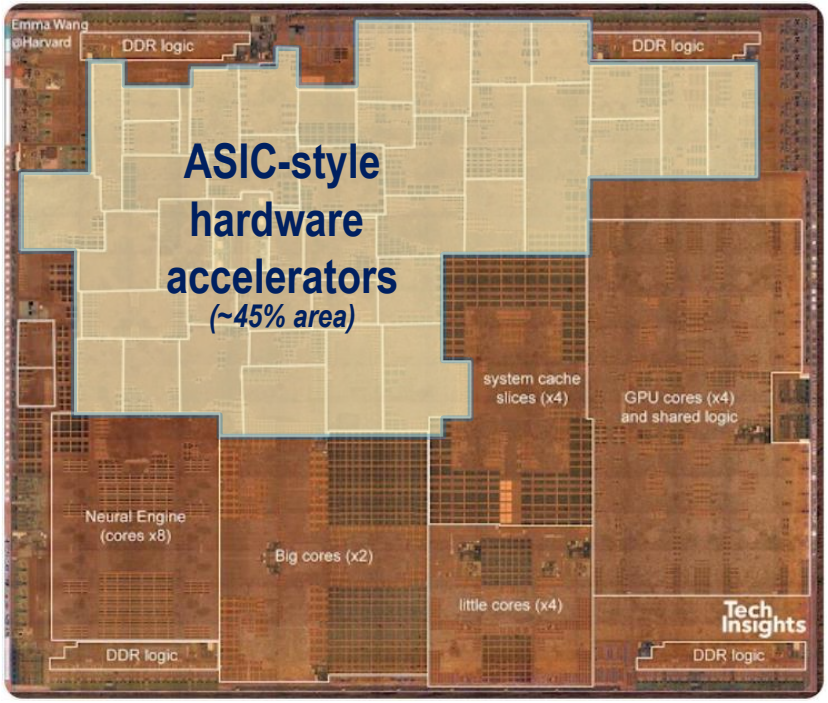
N.B. Asadi, PhD Thesis, Stanford 2010.

SoC Today: The Apple Approach

Linear growth in the "UnCore" units, exceeding ½ of SoC chip area today (A15)

Increasing "dark silicon" area (A12: ~45%, A15: ~55%), <10% chip is active





Apple A12 die photo

Semiconductor Scaling

- **Integration density continues to grow**
- **RC delay did not scale**
  - RC delay started to overtake gate delay
- **V_T did not scale**
  - To cope with leakage
- **V_{DD} did not scale**
  - To sustain performance growth

The Limits

Theoretical
(Physics)

Practical
(Physical + manufacturing cost)

- **System**
- **Circuit**
- **Device**
- **Material**
- **Fundamental**

[J. Meindl, Proc. IEEE, 1995]

Circuit Limits

1. Logic levels (gain)
2. Energy/transition
3. Delay
4. Global interconnect

1 Logic Levels (Gain)

- **Distinguish logic 0's from 1's**
- **Restore logic levels → |Gain| >1**

$$V_{DD} \geq \frac{2kT}{q}\left(1+\frac{C_{fs}}{C_d}\right)\ln\left(2+\frac{C_0}{C_d}\right)$$

$$\geq \beta\frac{kT}{q} \approx 0.1V \qquad \text{(T = 300K)}$$

$$\beta \approx 4$$

[J. Meindl, Proc. IEEE, 1995]

Circuit Limits (Cont.)

2 • **Energy/transition**
- Neglecting E_{static}

$$E_{tran} = \frac{1}{2} C_L {V_{DD}}^2$$

3 • **Delay**
- Limited by I_{DSat}

$$I_{DSat} \propto W C_{ox} v_{sat} (V_{GS} - V_T)$$

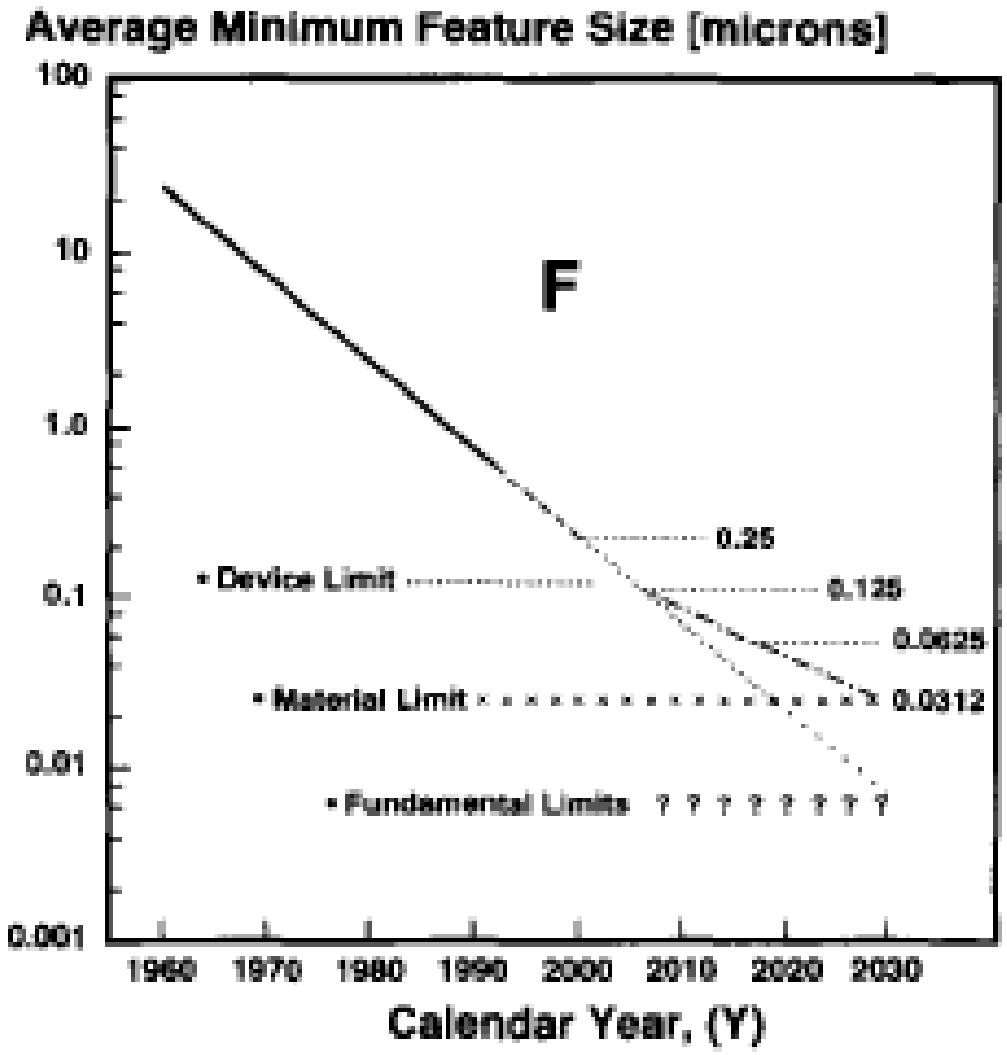
$$t_p = \frac{1}{2} \frac{C_L V_{DD}}{I_{DSat}}$$

4 • **Global interconnect**
- Interconnect delay should not exceed gate delay

$$\tau \propto (2.3 R_{gate} + R_{wire}) C_{wire}$$

$$R_{wire} < 2.3 R_{gate}$$

Practical Limits: Minimum Feature Size

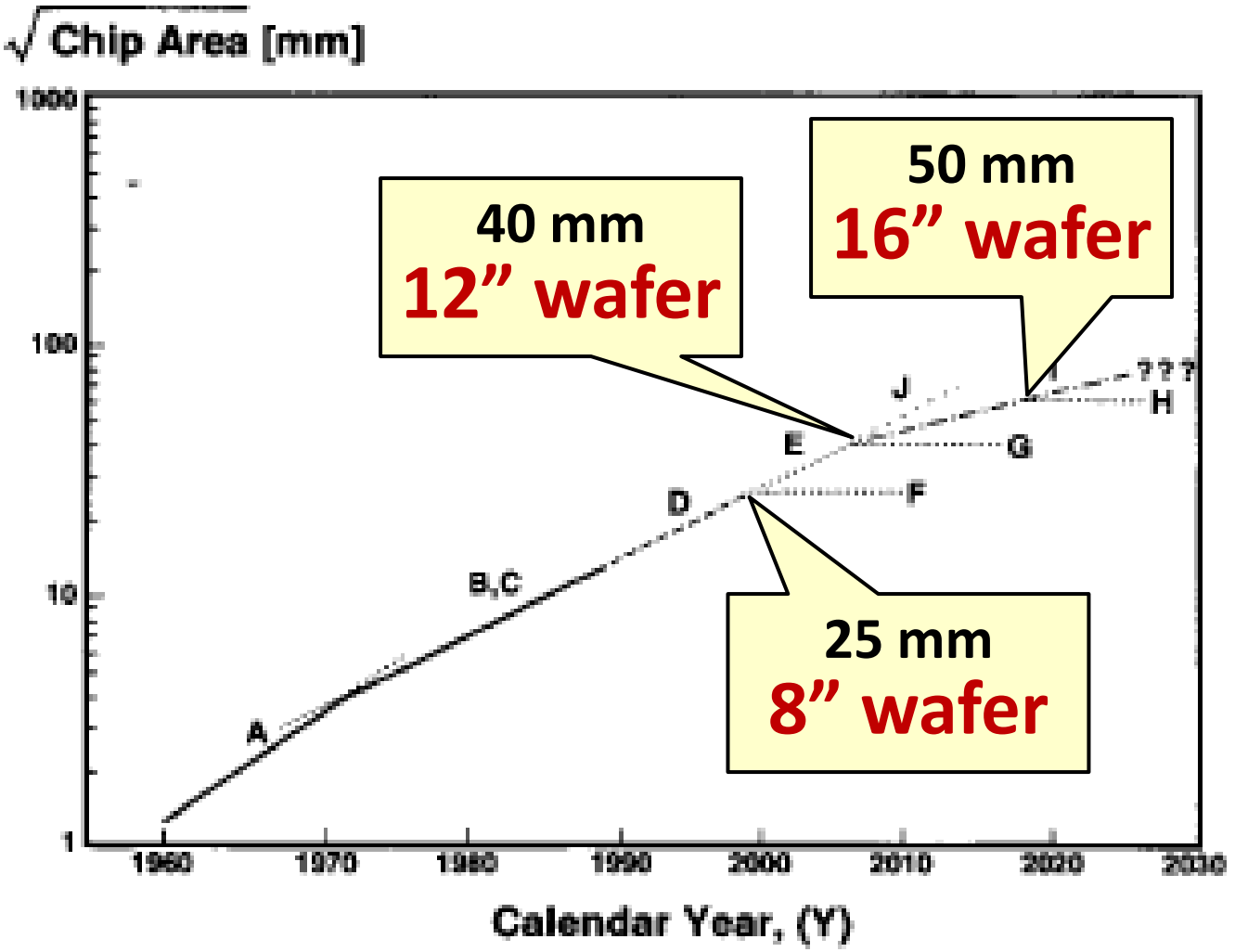


~130nm is the most cost-effective technology

(the last generation for which deep UV microlithography will suffice)

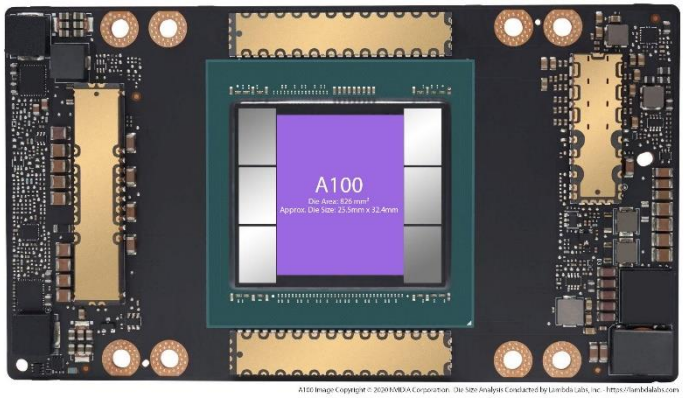
[J. Meindl, Proc. IEEE, 1995]

Practical Limits: Die Size



[J. Meindl, Proc. IEEE, 1995]

Die Size Olympics

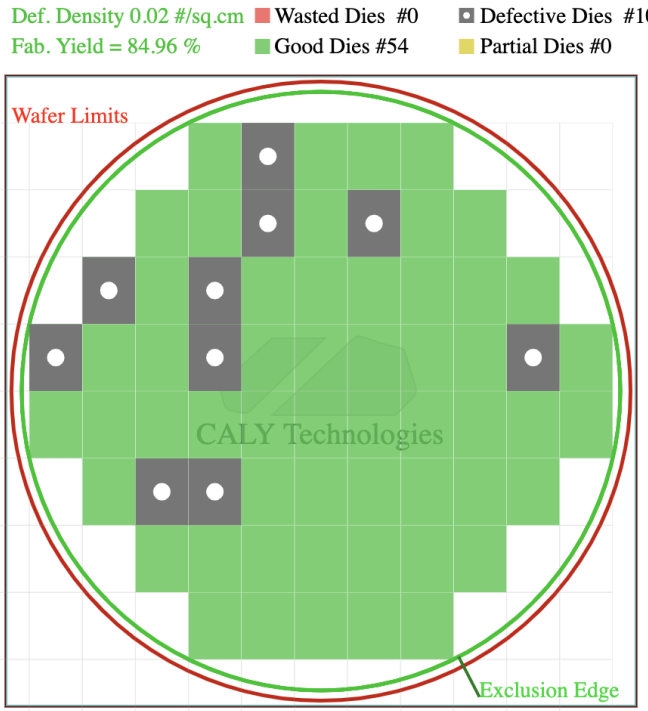


Nvidia A100

826mm^2

7nm

54B Tr.



Max Dies Per Wafer (without defect) #64

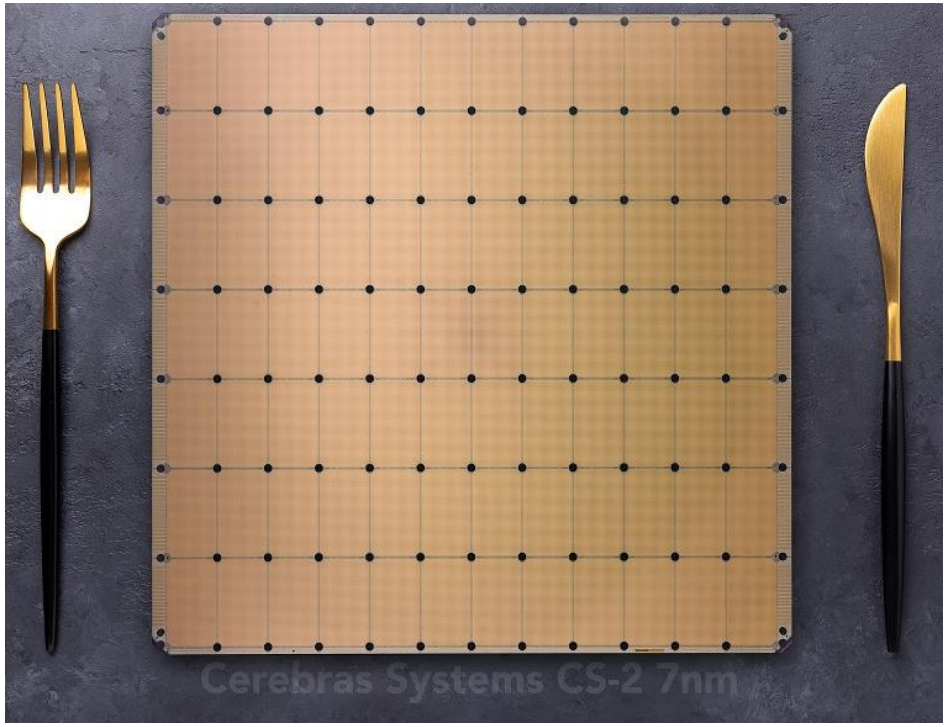
12” wafer

- **Reticle size limit 858mm^2**

Cerebras CS-2

Wafer scale **7**nm **2.6**T Tr.

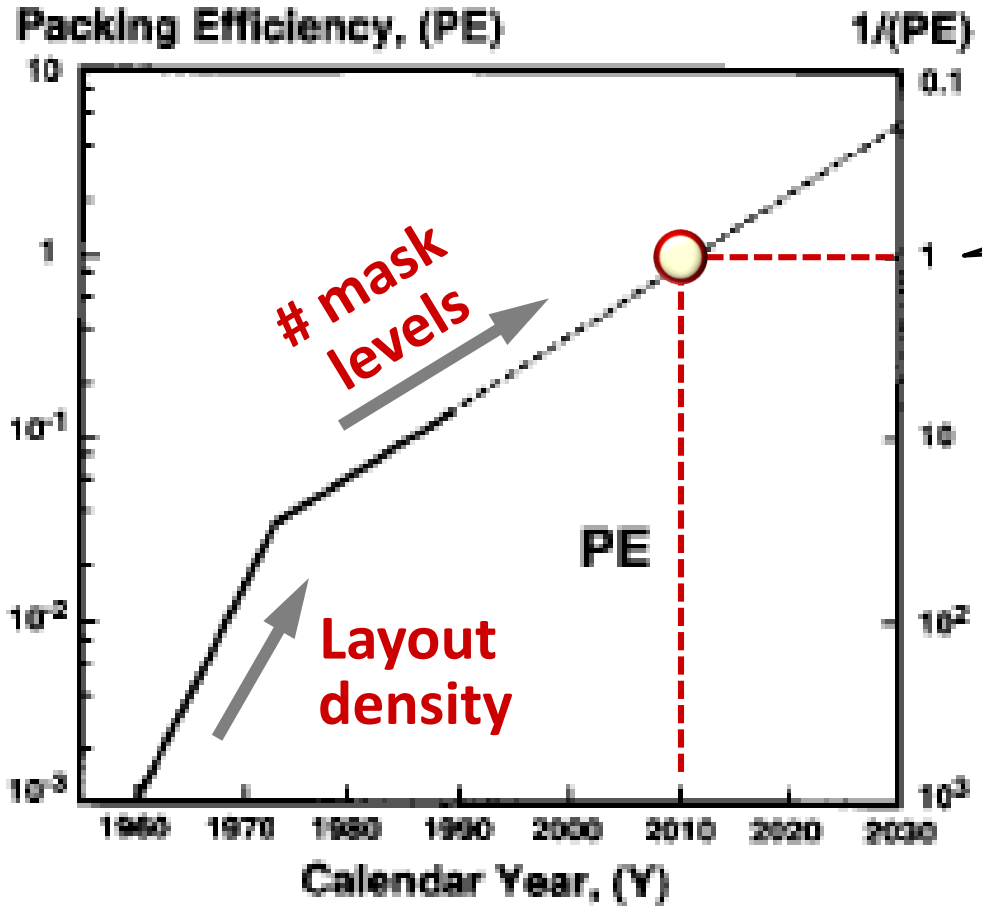
850,000 cores



Images from: https://lambdalabs.com and https://www.extremetech.com

Practical Limits: Packing Efficiency

Packing efficiency = # transistors / min feature area

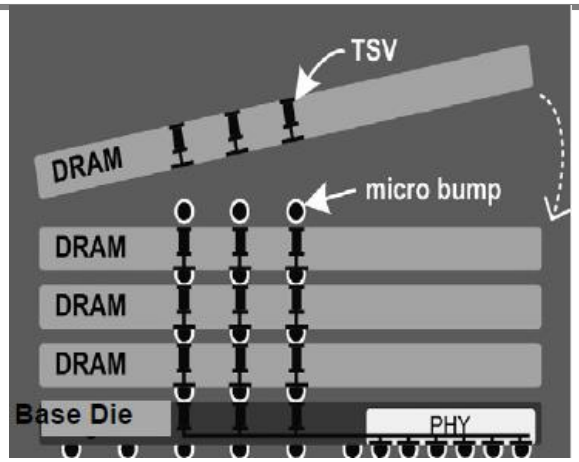


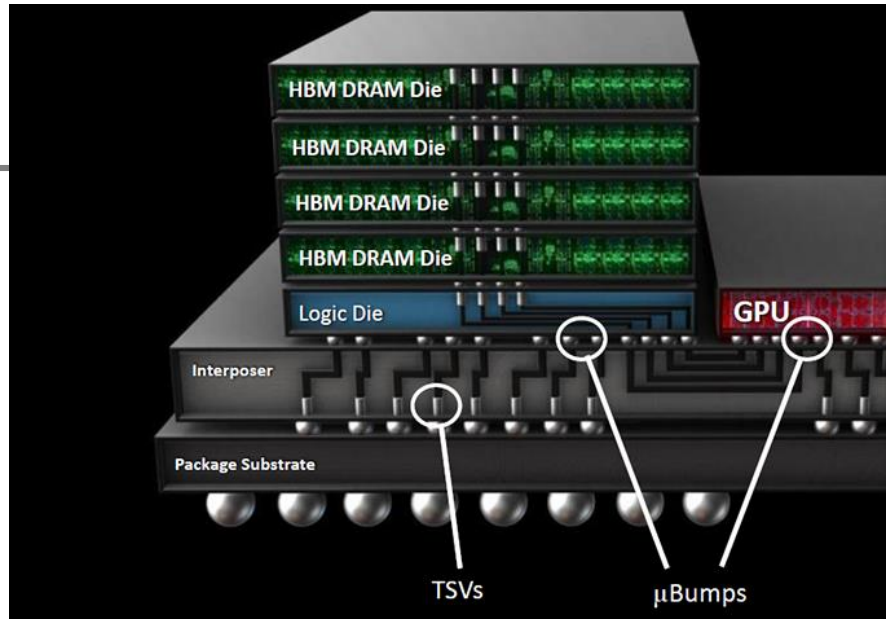
3D integration

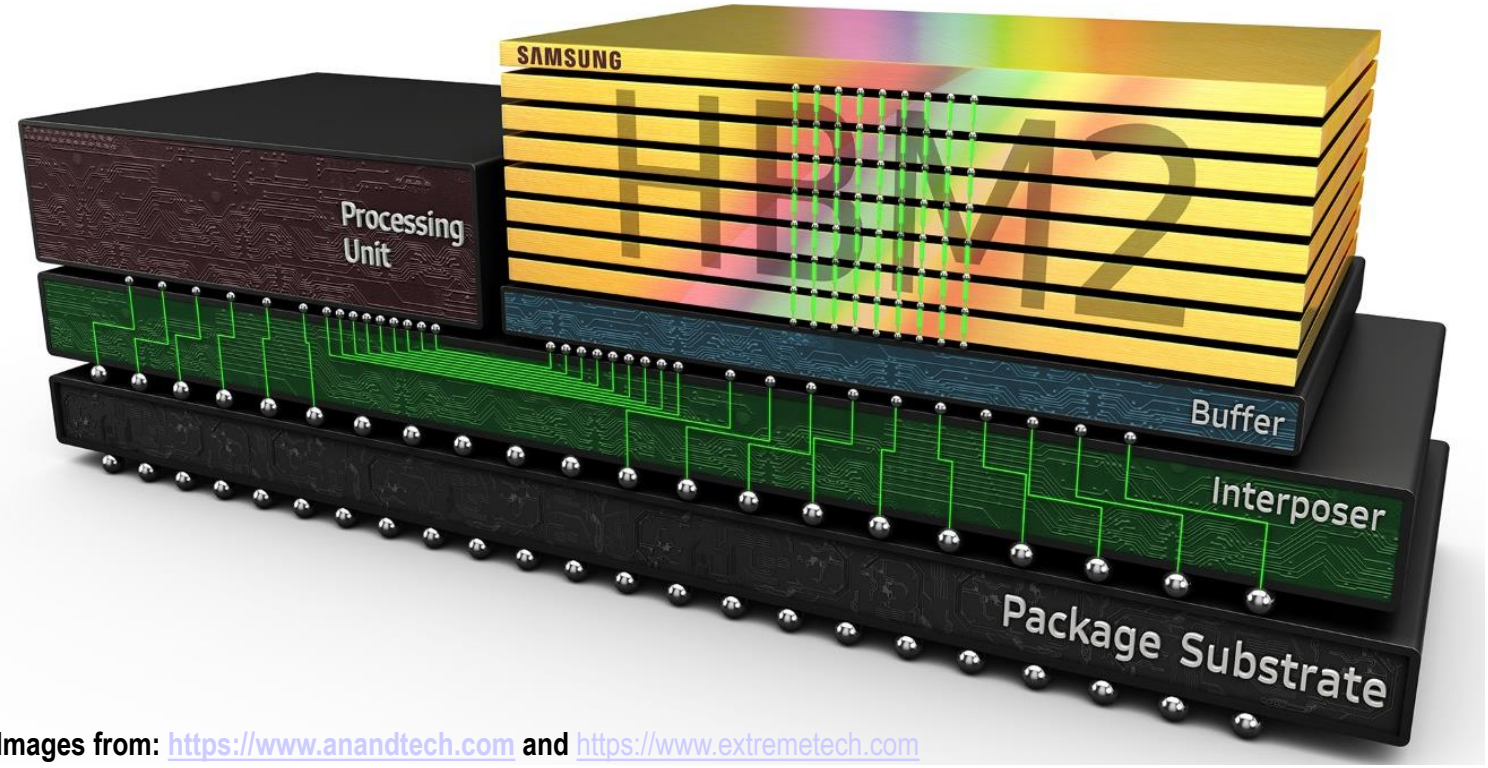
[J. Meindl, Proc. IEEE, 1995]

3D Integration

- **E.g. High-Bandwidth Memory (HBM)**



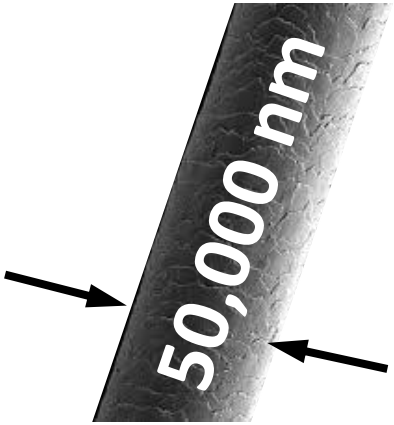




Images from: https://www.anandtech.com **and** https://www.extremetech.com

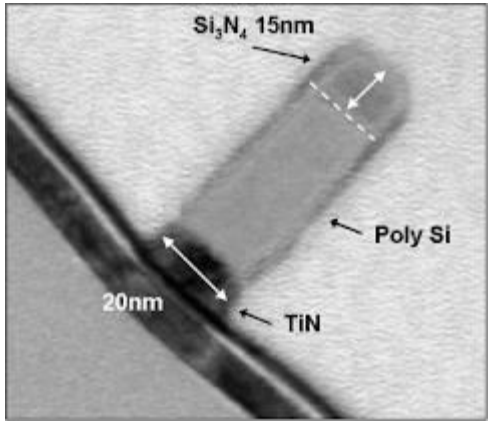
Approaching Atomic Limits

Hair



200,000x

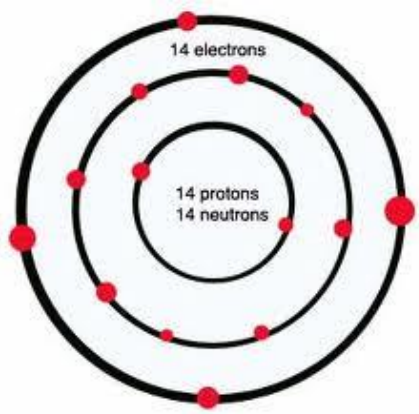
20nm FET



80x

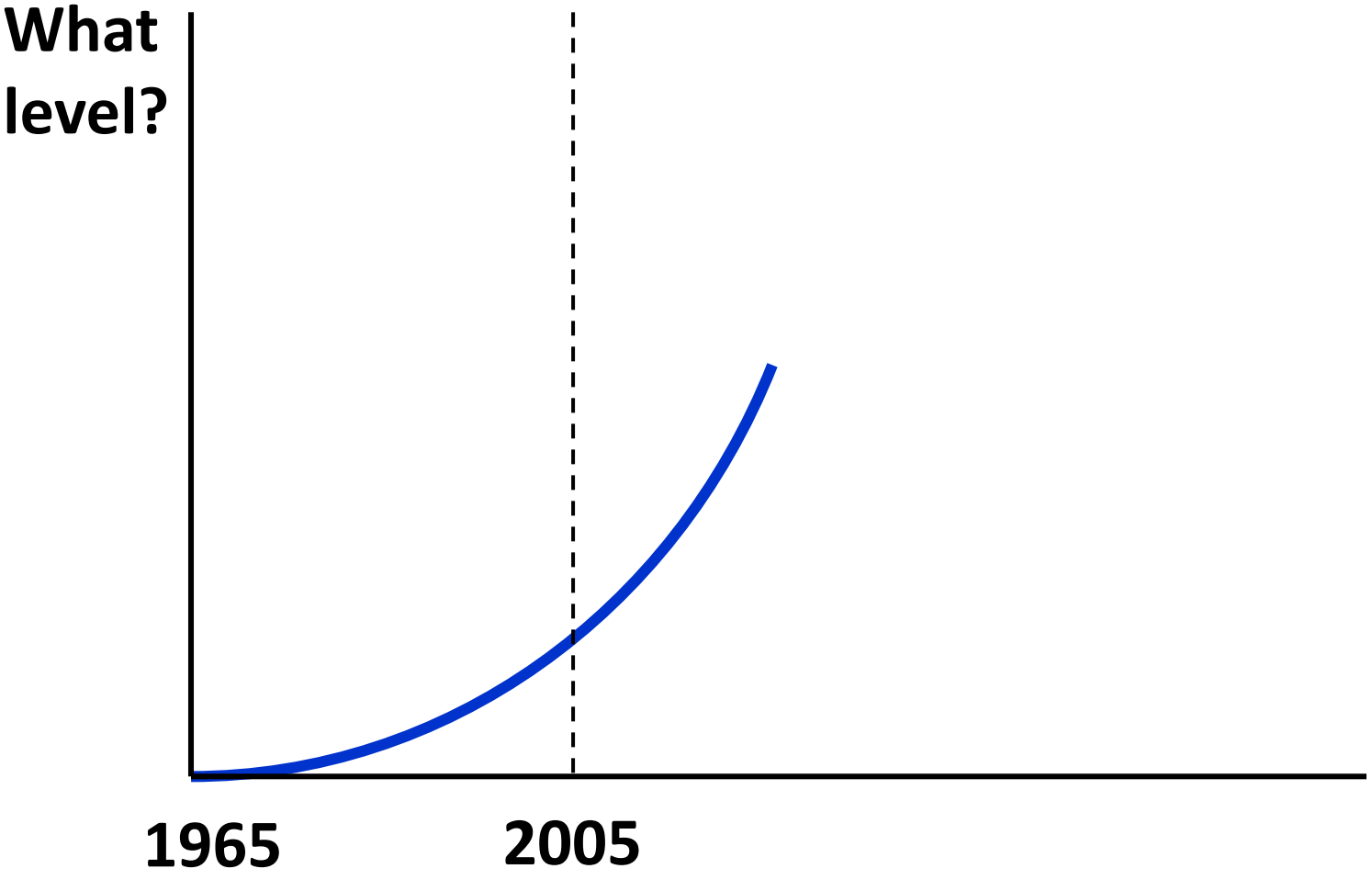
Si atom

0.25nm

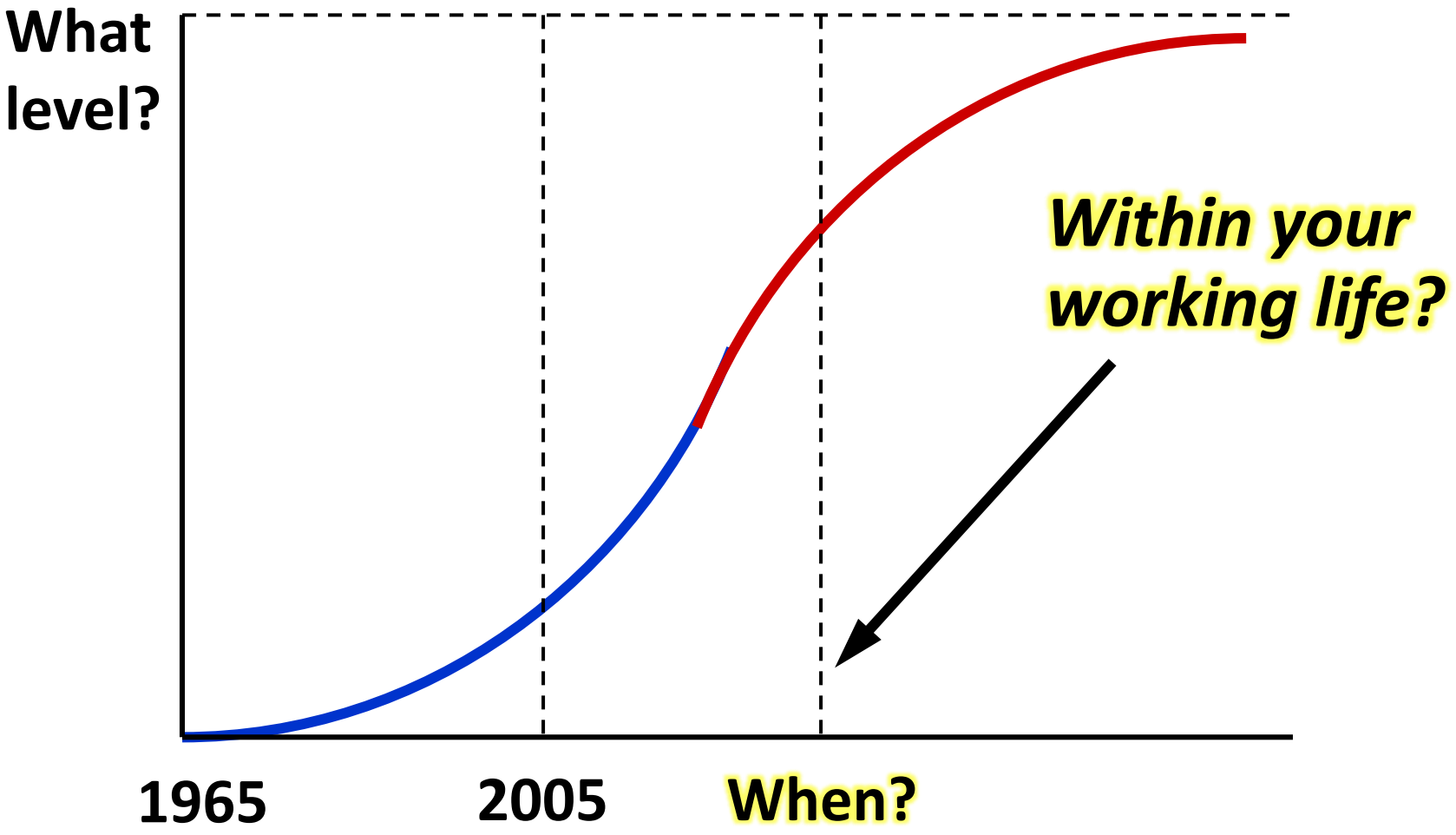


1x

Moore's Law and the Long Term



Moore's Law and the Long Term



CMOS Replacement?

- Replacing CMOS by another more energy efficient technology is a distant prospect now
- Low-power high-speed CMOS technology is becoming an indispensable, rather than desirable, technology
- Power is the main challenge we need to address