

Agenda

- Tutorial 3:
 - Digital Design Flow
- Tutorial 4:
 - ECE1388 Project 3 - 4-bit array multiplier block
- Case Study:
 - Why all this is useful – 2 recent design examples
- Tutorial 5: More Advanced Design Techniques and the Future
 - ML-Driven CAD: Google Keynote

Agenda

- Tutorial 3:
 - Digital Design Flow
- Tutorial 4:
 - ECE1388 Project 3 - 4-bit array multiplier block
- Case Study:
 - Why all this is useful – 2 recent design examples
- Tutorial 5: More Advanced Design Techniques and the Future
 - ML-Driven CAD: Google Keynote

Tutorial 3

Digital Design Flow

ECE1388

Mustafa Kanchwala

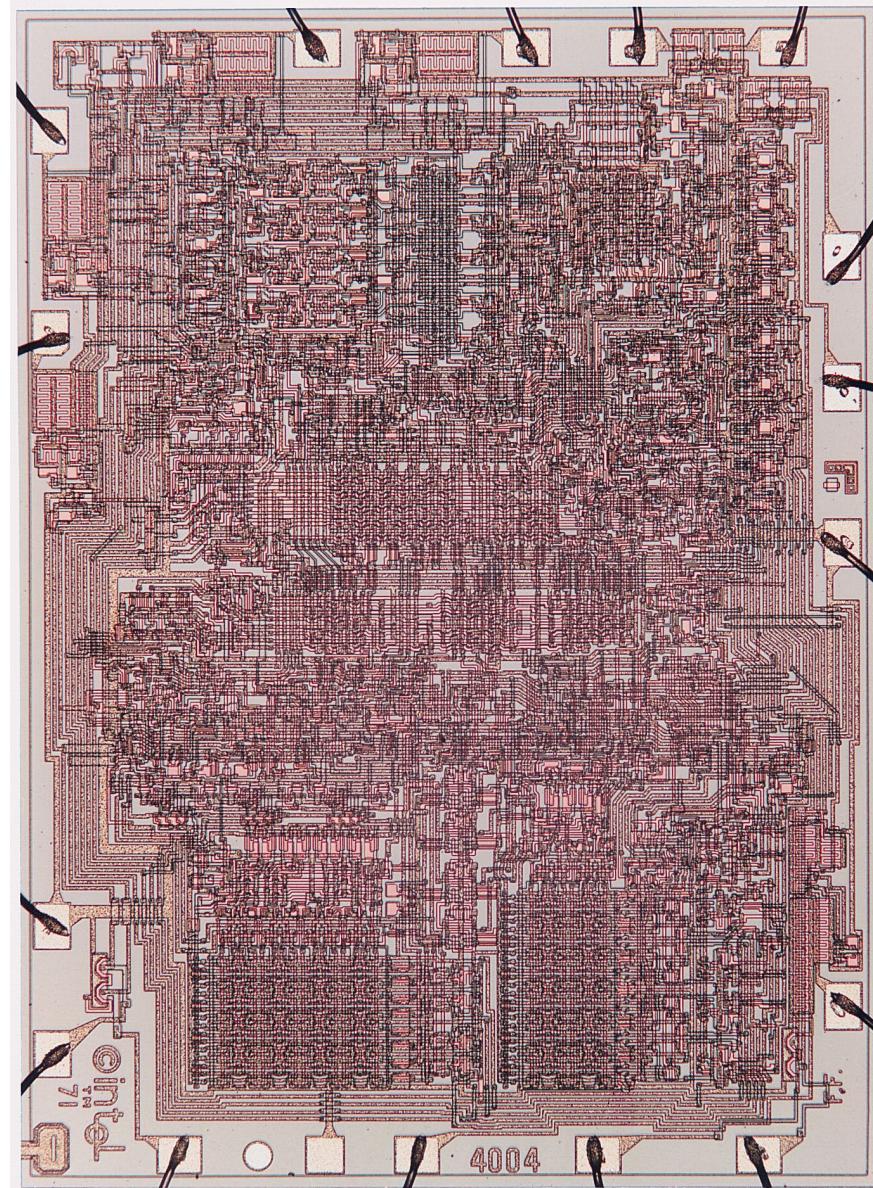
Adapted from slides by Gerard O'Leary

Outline

- What is “digital implementation”?
 - Complexity Case Study:
 - Evolution of Intel Microprocessors from 4004 to Core i9
 - Slides courtesy: www.cmosvlsi.com
- Overview: Digital Design Flow

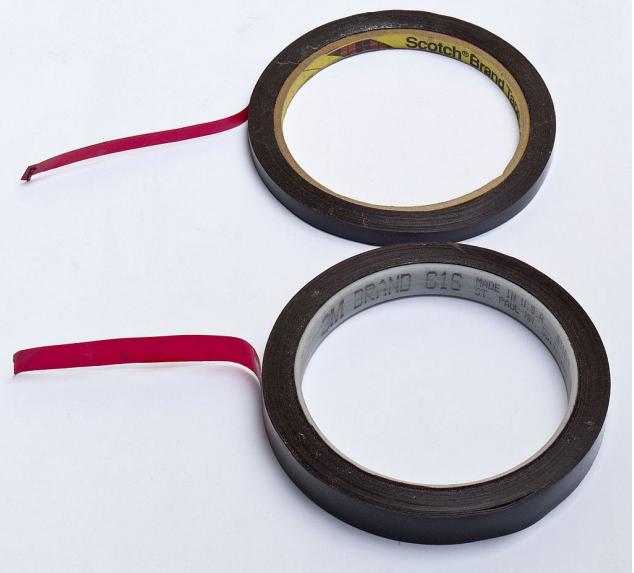
4004

- First microprocessor (1971)
 - For Busicom calculator
- Characteristics
 - **10 µm process**
 - **2300 transistors**
 - 400 – 800 kHz
 - **4-bit word size**
 - 16-pin DIP package
- Masks hand cut from Rubylith
 - Drawn with color pencils
 - 1 metal, 1 poly (jumpers)
 - Diagonal lines (!)



Fun Facts!

Lithographers Tape:
aka Rubylith

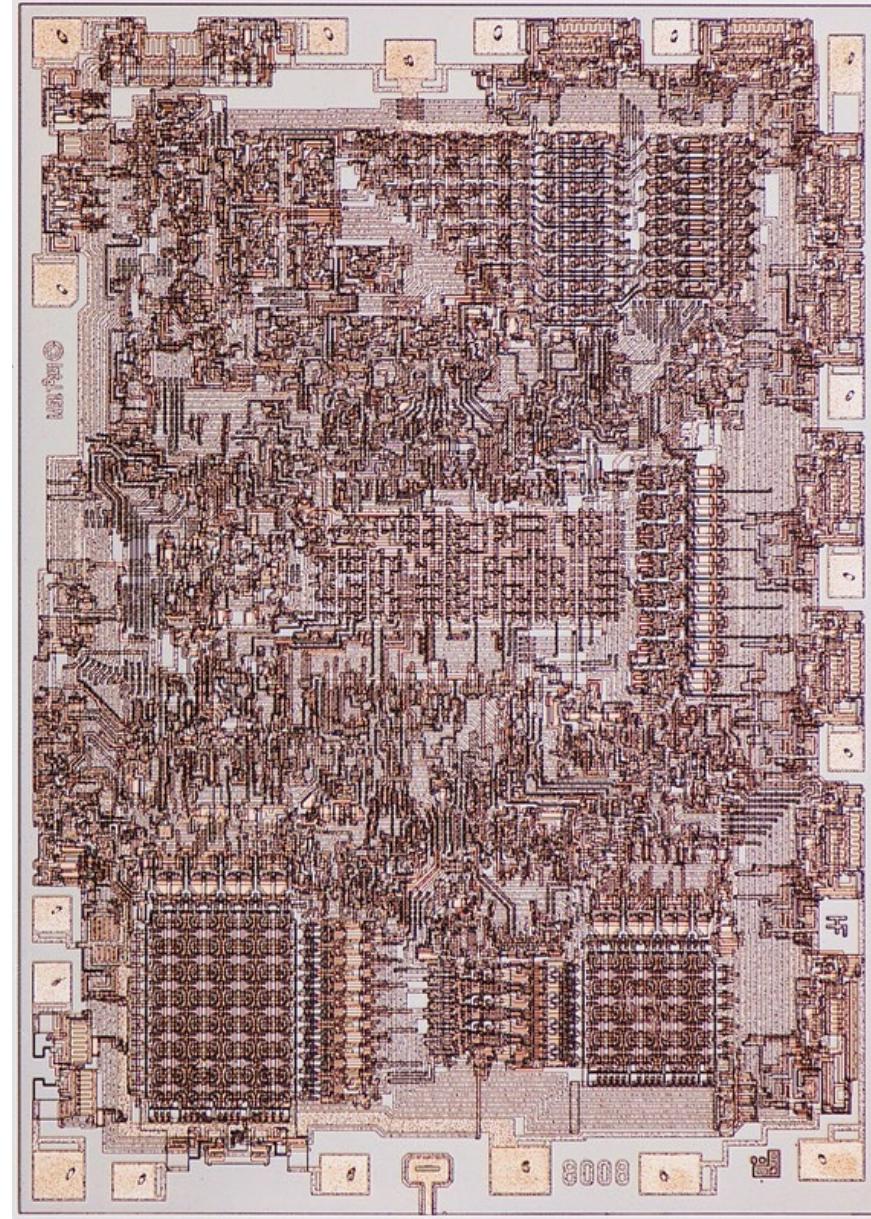


Founders of Intel
Gordon Moore
Robert Noyce
Andy Grove

Rubylith Cutout

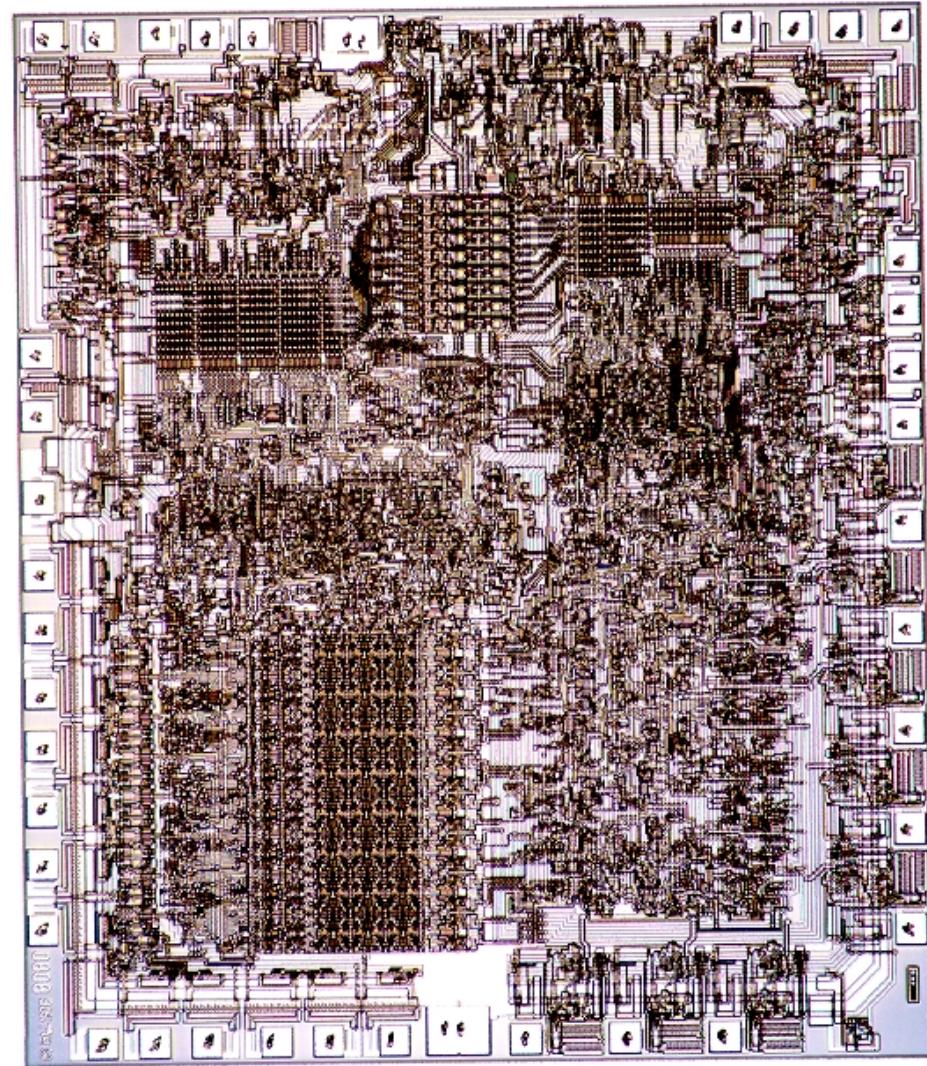
8008

- 8-bit follow-on (1972)
 - Dumb terminals
- Characteristics
 - **10 µm process**
 - **3500 transistors**
 - 500 – 800 kHz
 - 8-bit word size
 - 18-pin DIP package
- Note 8-bit datapaths
 - Individual transistors visible



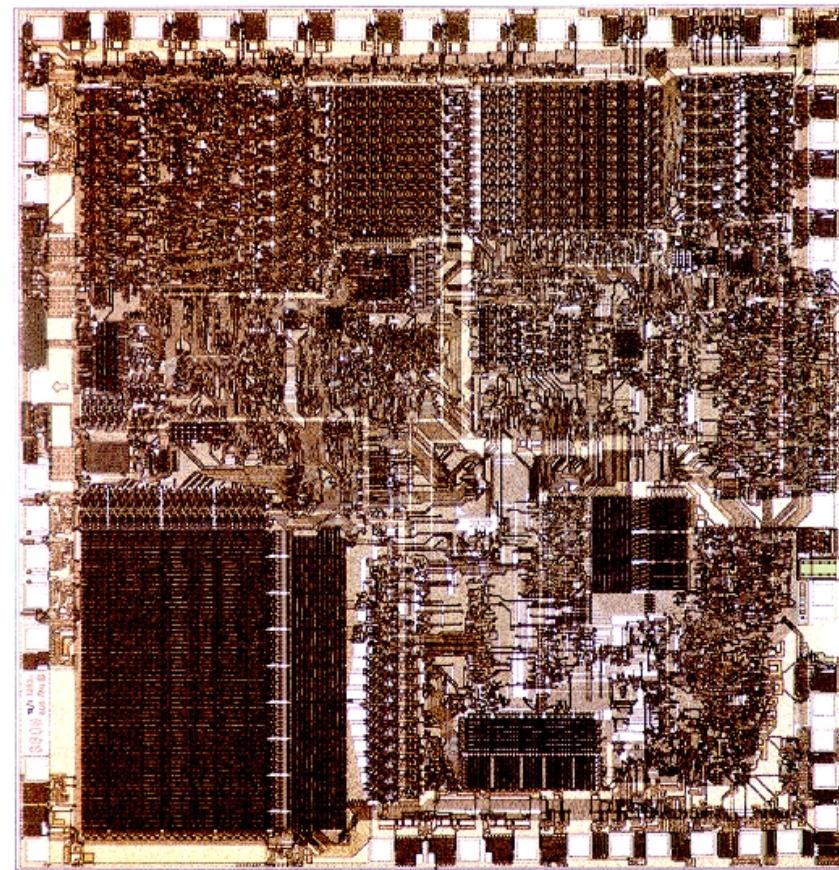
8080

- 16-bit address bus (1974)
 - Used in Altair computer
 - (early hobbyist PC)
 - Characteristics
 - **6 µm process**
 - **4500 transistors**
 - 2 MHz
 - 8-bit word size
 - 40-pin DIP package



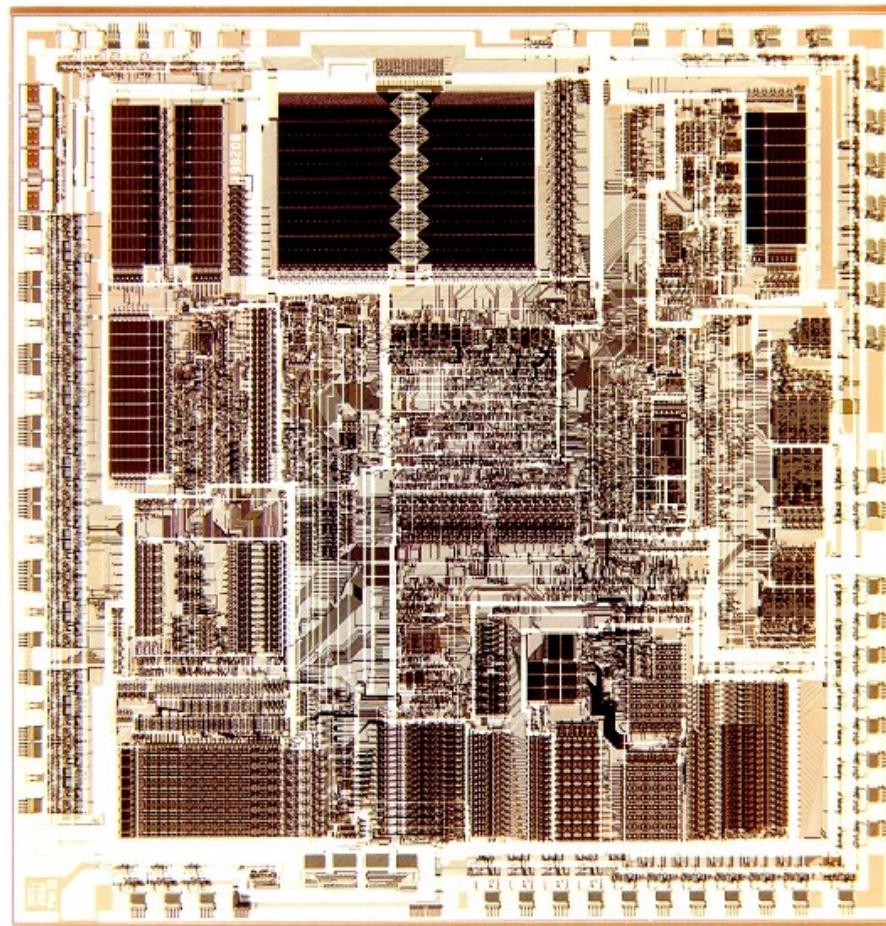
8086 / 8088

- 16-bit processor (1978-9)
 - IBM PC and PC XT
 - Revolutionary products
 - Introduced x86 ISA
- Characteristics
 - **3 μm process**
 - **29k transistors**
 - 5-10 MHz
 - 16-bit word size
 - 40-pin DIP package
- Microcode ROM



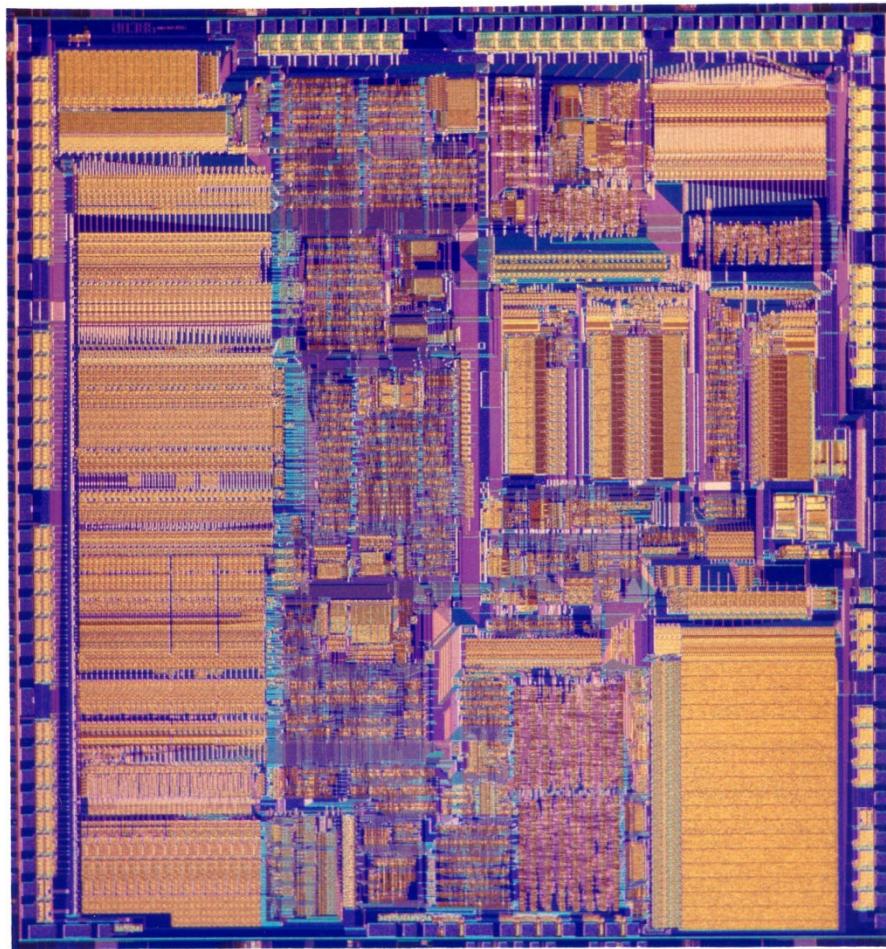
80286

- Virtual memory (1982)
 - IBM PC AT
- Characteristics
 - **1.5 µm process**
 - **134k transistors**
 - 6-12 MHz
 - 16-bit word size
 - 68-pin PGA
- Regular datapaths and ROMs
Bitslices clearly visible



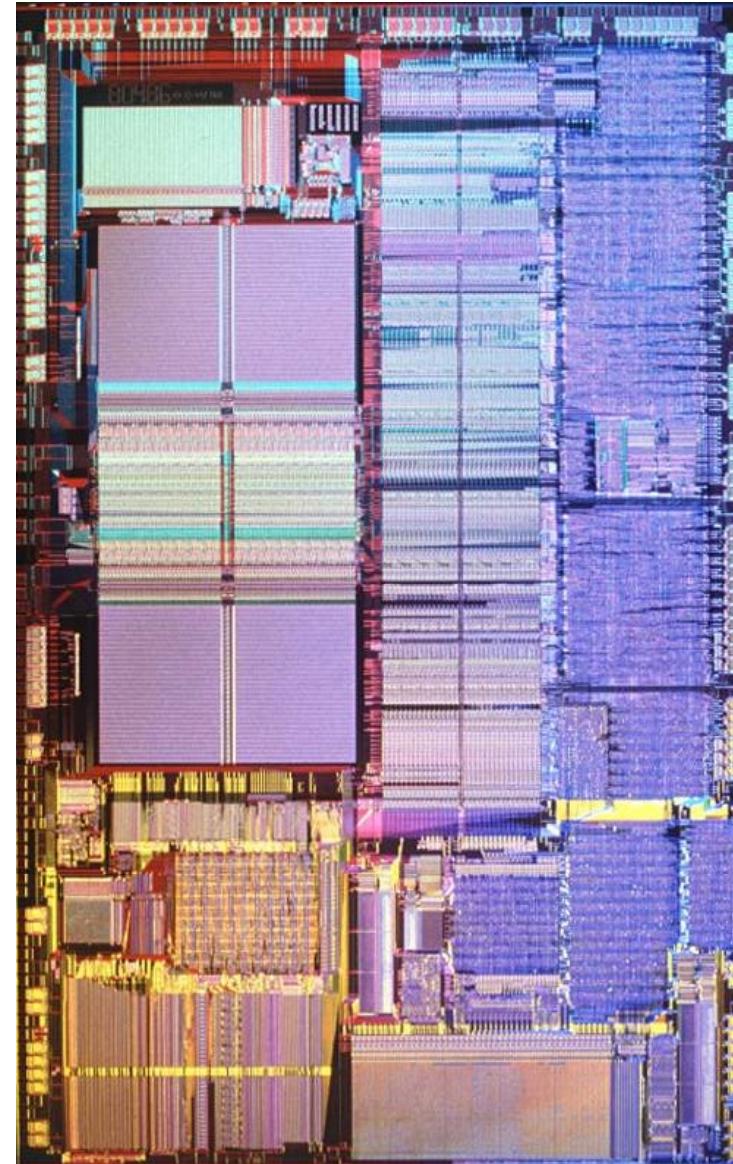
80386

- 32-bit processor (1985)
 - Modern x86 ISA
- Characteristics
 - **1.5-1 µm process**
 - **275k transistors**
 - 16-33 MHz
 - 32-bit word size
 - 100-pin PGA
- 32-bit datapath,
microcode ROM,
synthesized control



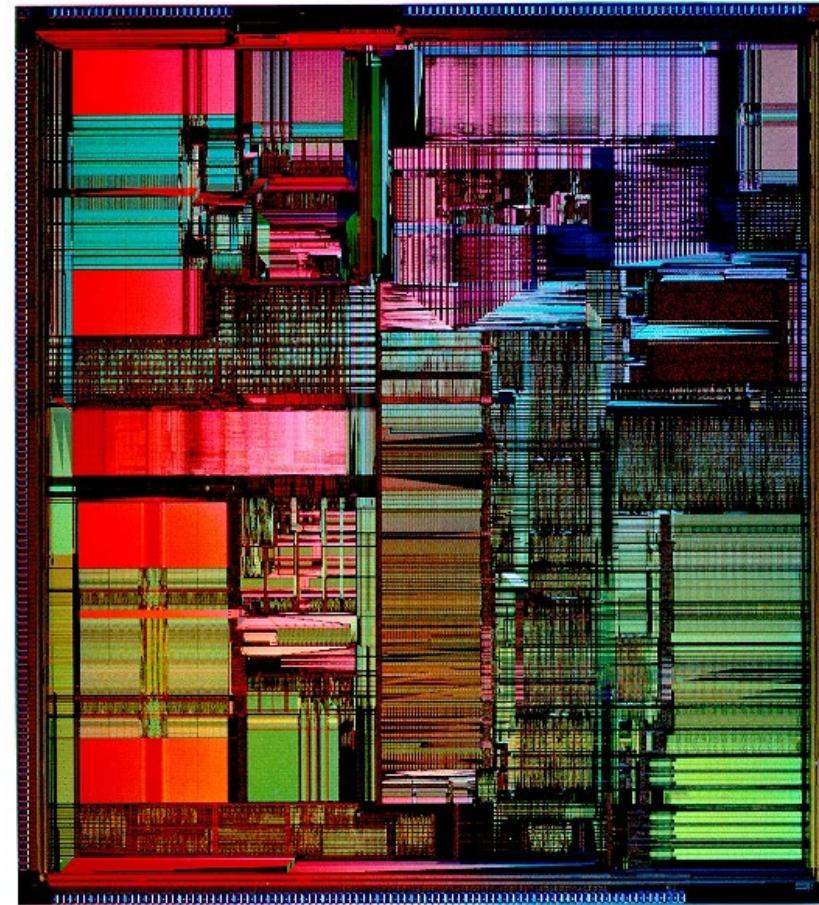
80486

- Pipelining (1989)
 - Floating point unit
 - 8 KB cache
- Characteristics
 - **1-0.6 μm process**
 - **1.2M transistors**
 - 25-100 MHz
 - 32-bit word size
 - 168-pin PGA
- Cache, Integer datapath, FPU, microcode, synthesized control



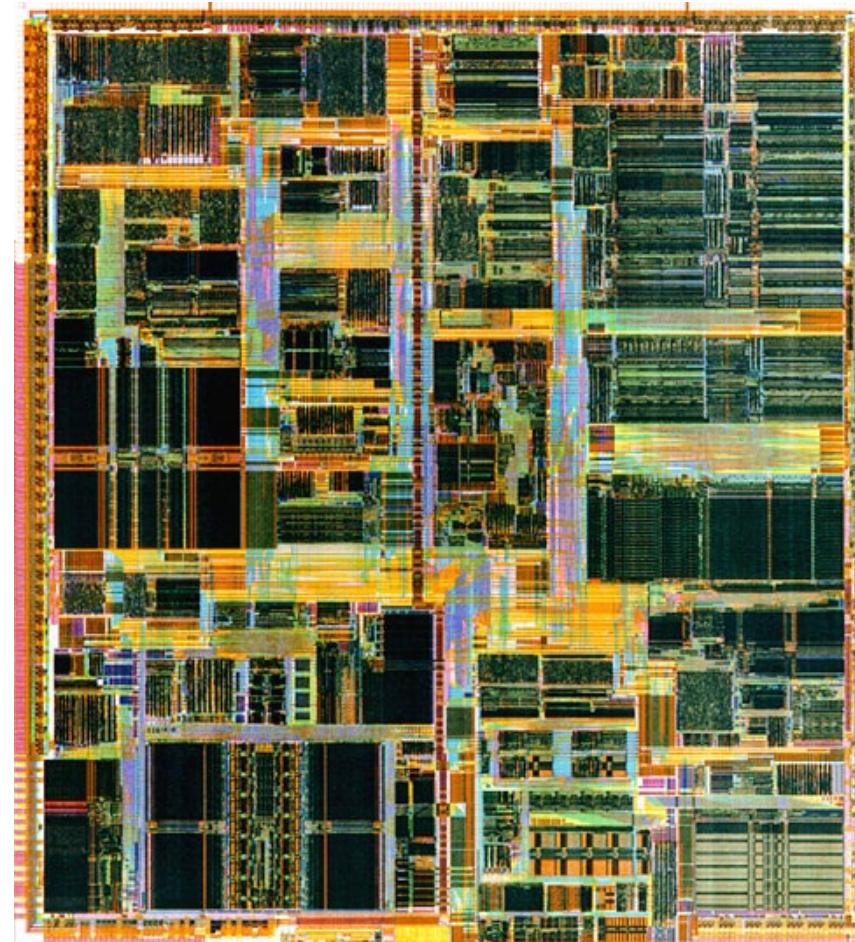
Pentium

- Superscalar (1993)
 - 2 instructions per cycle
 - Separate 8KB I\$ & D\$
- Characteristics
 - **0.8-0.35 μm process**
 - **3.2M transistors**
 - 60-300 MHz
 - 32-bit word size
 - 296-pin PGA
- Caches, datapath,
FPU, control



Pentium Pro / II / III

- Dynamic execution (1995-9)
 - 3 micro-ops / cycle
 - Out of order execution
 - 16-32 KB I\$ & D\$
 - Multimedia instructions
 - PIII adds 256+ KB L2\$
- Characteristics
 - **0.6-0.18 μm process**
 - **5.5M-28M transistors**
 - 166-1000 MHz
 - 32-bit word size



Pentium 4

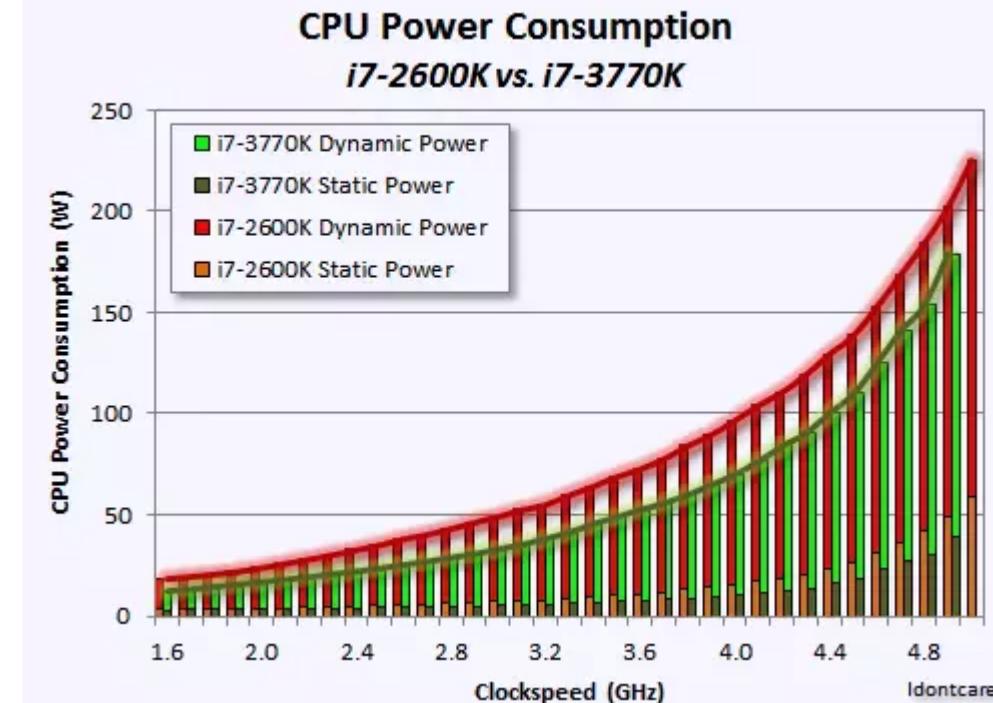
- Deep pipeline (2001)
 - Very fast clock
 - 256-1024 KB L2\$
- Characteristics
 - **180 – 65 nm process**
 - **42-125M transistors**
 - 1.4-3.4 GHz
 - Up to 160 W
 - 32/64-bit word size
 - 478-pin PGA
- Units start to become invisible on this scale



Pentium 4

- Deep pipeline (2001)
 - Very fast clock
 - 256-1024 KB L2\$
- Characteristics
 - **180 – 65 nm process**
 - **42-125M transistors**
 - **1.4-3.4 GHz**
 - Up to 160 W
 - 32/64-bit word size
 - 478-pin PGA
- Units start to become invisible on this scale

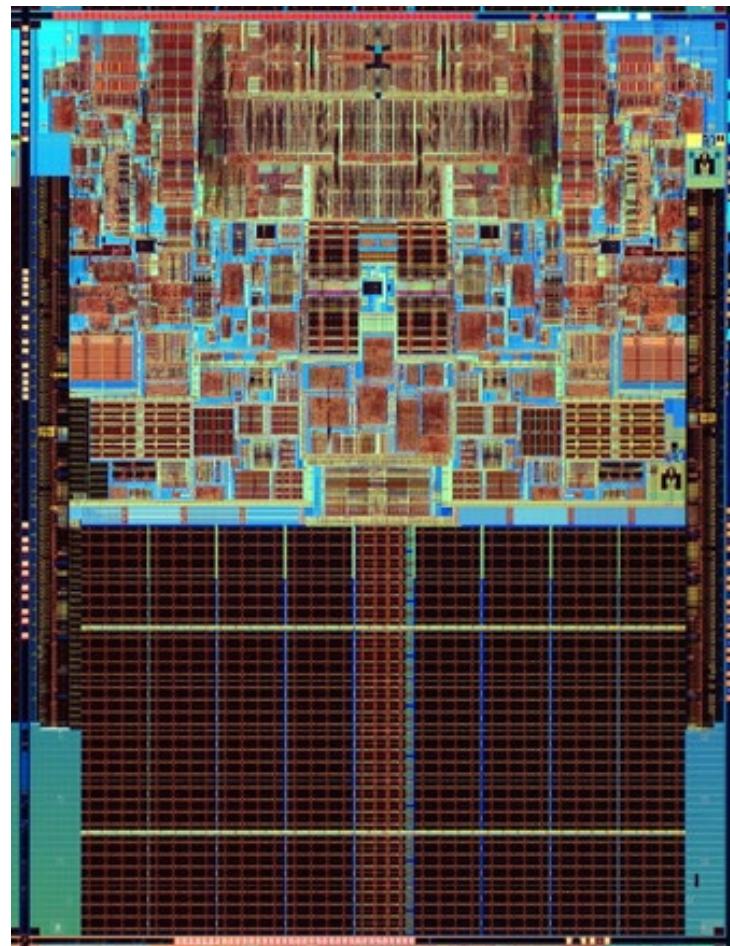
Dennard Scaling: As a transistor's size goes down, the amount of power it consumes also goes down roughly proportionally by area (i.e. the power density of the chip remains constant).



Single thread performance capped

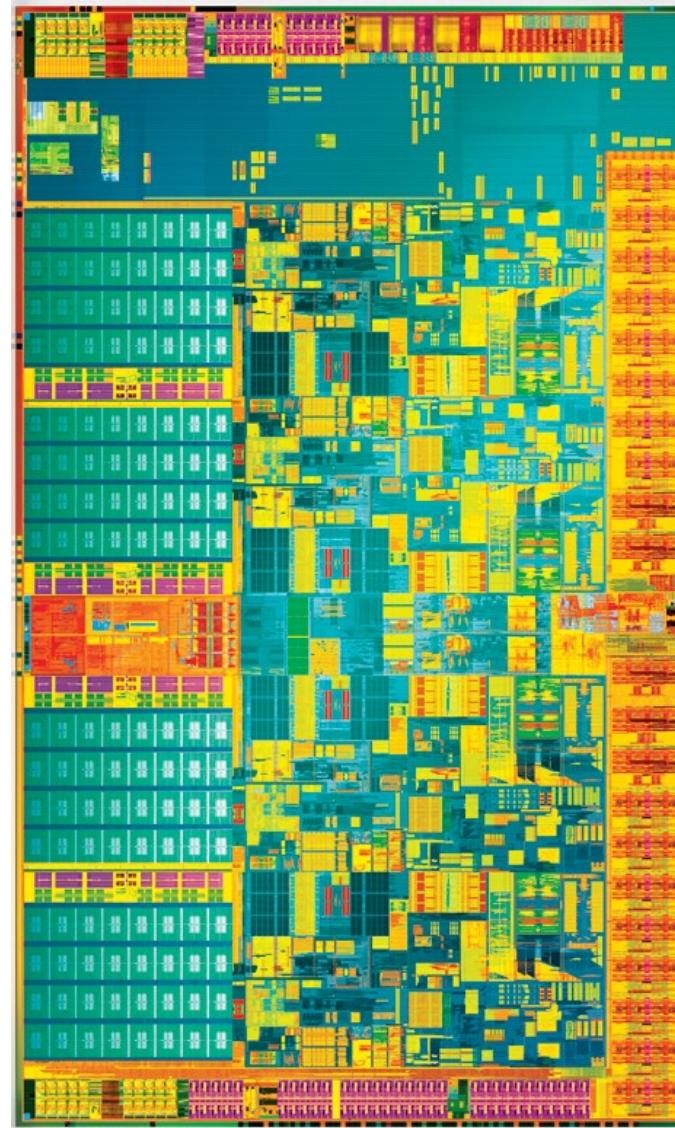
Core2 Duo

- Dual core (2006)
 - 1-2 MB L2\$ / core
- Characteristics
 - **65-45 nm process**
 - **291M transistors**
 - 1.6-3+ GHz
 - 65 W
 - 32/64 bit word size
 - 775 pin LGA
- Much better performance/power efficiency



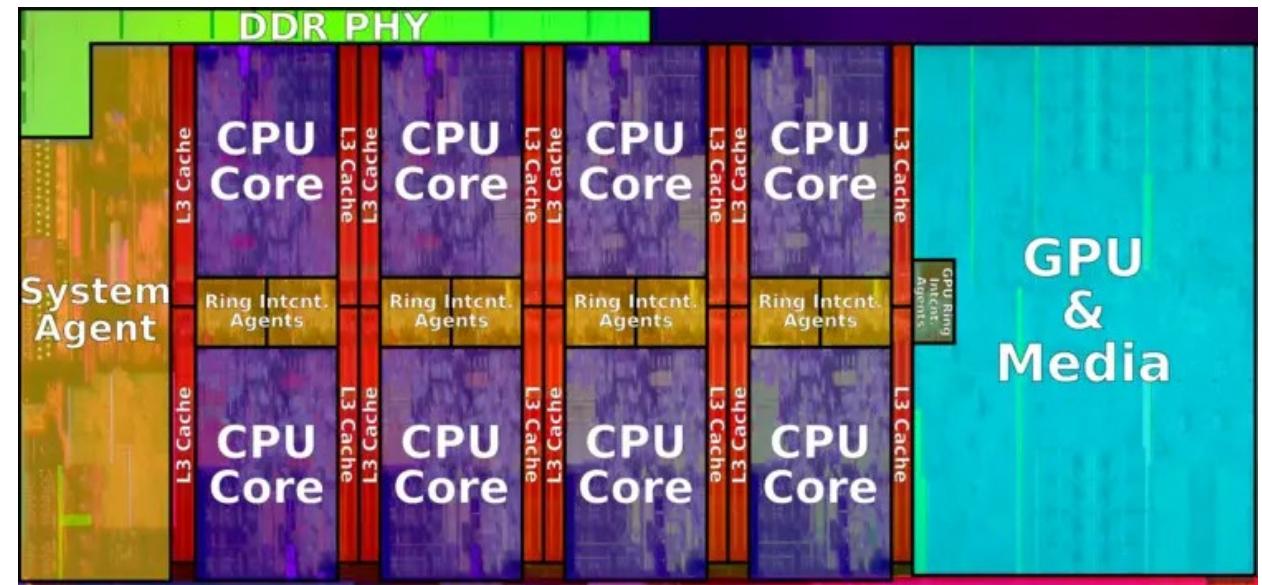
Core i7

- Quad core (2008)
 - Refinement of Core architecture
 - 2 MB L3\$ / core
- Characteristics
 - **45-14 nm process**
 - **~3B transistors**
 - 2.66-4+ GHz
 - Up to 130 W
 - 32/64 bit word size
 - 1366-pin LGA
 - Multithreading
- On-die memory controller



Core i9

- 8-18 Core (2019)
 - 36 threads
 - 9900K-10980XE
 - **14 nm process**
 - **~7B transistors**
 - 5 GHz
 - 22M Cache
 - 256 GB DDR4

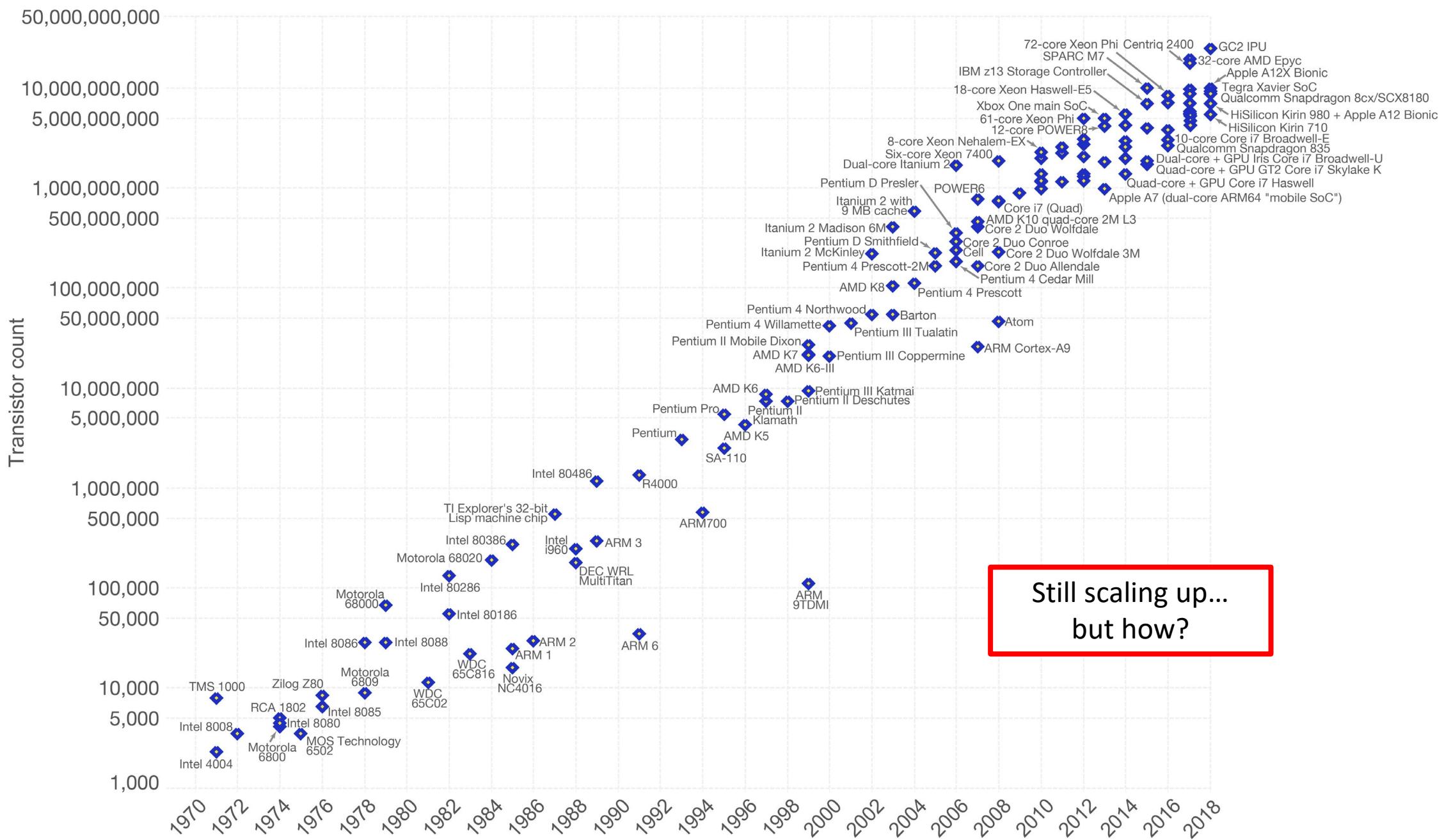


Scaling

- “This incredible growth rate could not be achieved by hiring an exponentially-growing number of design engineers. It was fulfilled by adopting new design methodologies and by introducing innovative design automation software at every processor generation.”
 - Pat Gelsinger et. al. (Intel) - IEEE Solid-State Circuits Magazine, 2010.

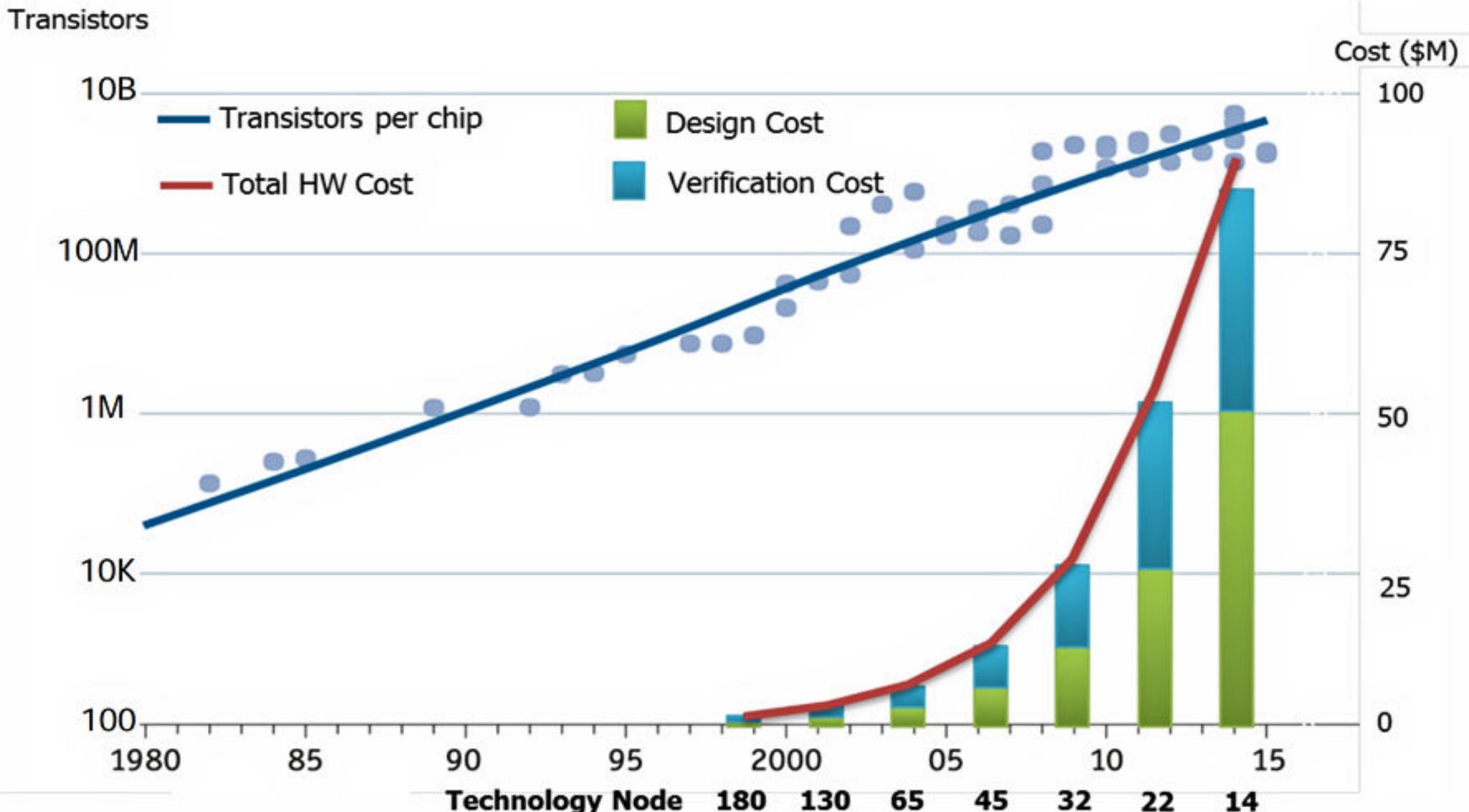
Processor	Year	Feature Size (μm)	Transistors	Frequency (MHz)	Word Size	Power (W)	Cache (L1 / L2 / L3)	Package
4004	1971	10	2.3k	0.75	4	0.5	none	16-pin DIP
8008	1972	10	3.5k	0.5–0.8	8	0.5	none	18-pin DIP
8080	1974	6	6k	2	8	0.5	none	40-pin DIP
8086	1978	3	29k	5–10	16	2	none	40-pin DIP
80286	1982	1.5	134k	6–12	16	3	none	68-pin PGA
Intel386	1985	1.5–1.0	275k	16–25	32	1–1.5	none	100-pin PGA
Intel486	1989	1–0.6	1.2M	25–100	32	0.3–2.5	8K	168-pin PGA
Pentium	1993	0.8–0.35	3.2–4.5M	60–300	32	8–17	16K	296-pin PGA
Pentium Pro	1995	0.6–0.35	5.5M	166–200	32	29–47	16K / 256K+	387-pin MCM PGA
Pentium II	1997	0.35–0.25	7.5M	233–450	32	17–43	32K / 256K+	242-pin SECC
Pentium III	1999	0.25–0.18	9.5–28M	450–1000	32	14–44	32K / 512K	330-pin SECC2
Pentium 4	2000	180–65 nm	42–178M	1400–3800	32/64	21–115	20K+ / 256K+	478-pin PGA
Pentium M	2003	130–90 nm	77–140M	1300–2130	32	5–27	64K / 1M	479-pin FCBGA
Core	2006	65 nm	152M	1000–1860	32	6–31	64K / 2M	479-pin FCBGA
Core 2 Duo	2006	65–45 nm	167–410M	1060–3160	32/64	10–65	64K / 4M+	775-pin LGA
Core i7	2008	45 nm	731M	2660–3330	32/64	45–130	64K / 256K / 8M	1366-pin LGA
Atom	2008	45 nm	47M	800–1860	32/64	1.4–13	56K / 512K+	441-pin FCBGA

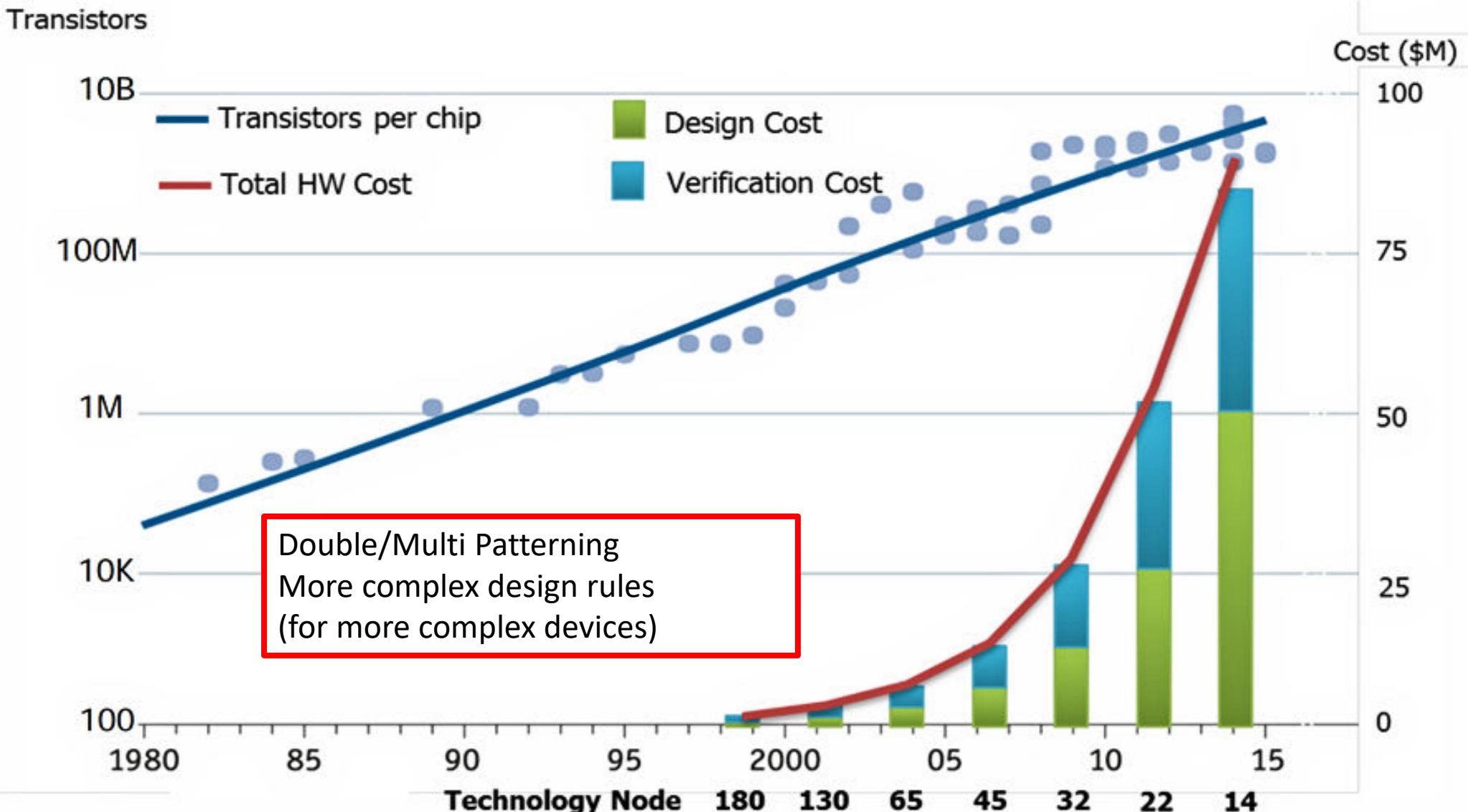
Beyond Moore's Law



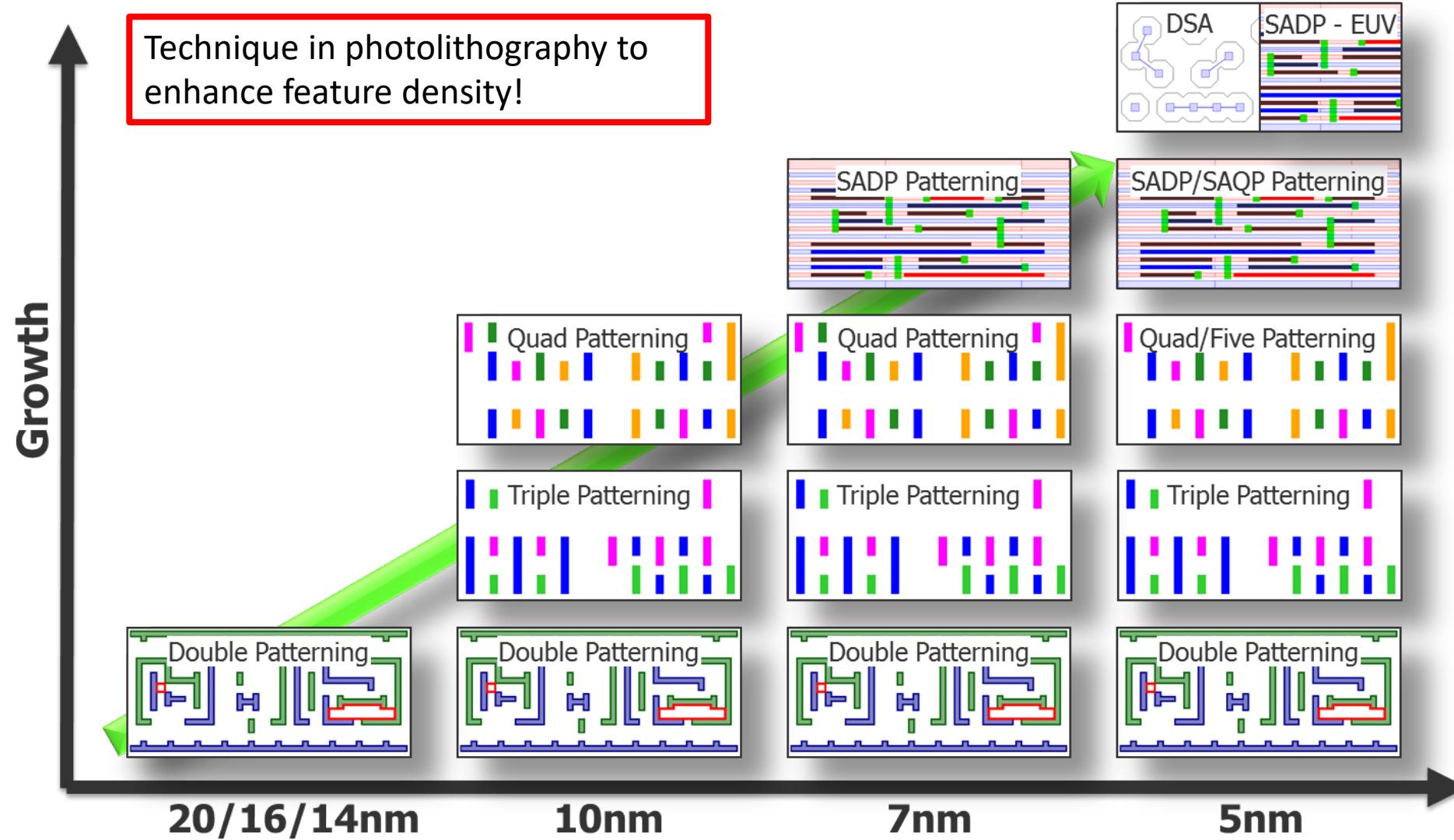


Has EDA failed to keep up with Moore's Law?

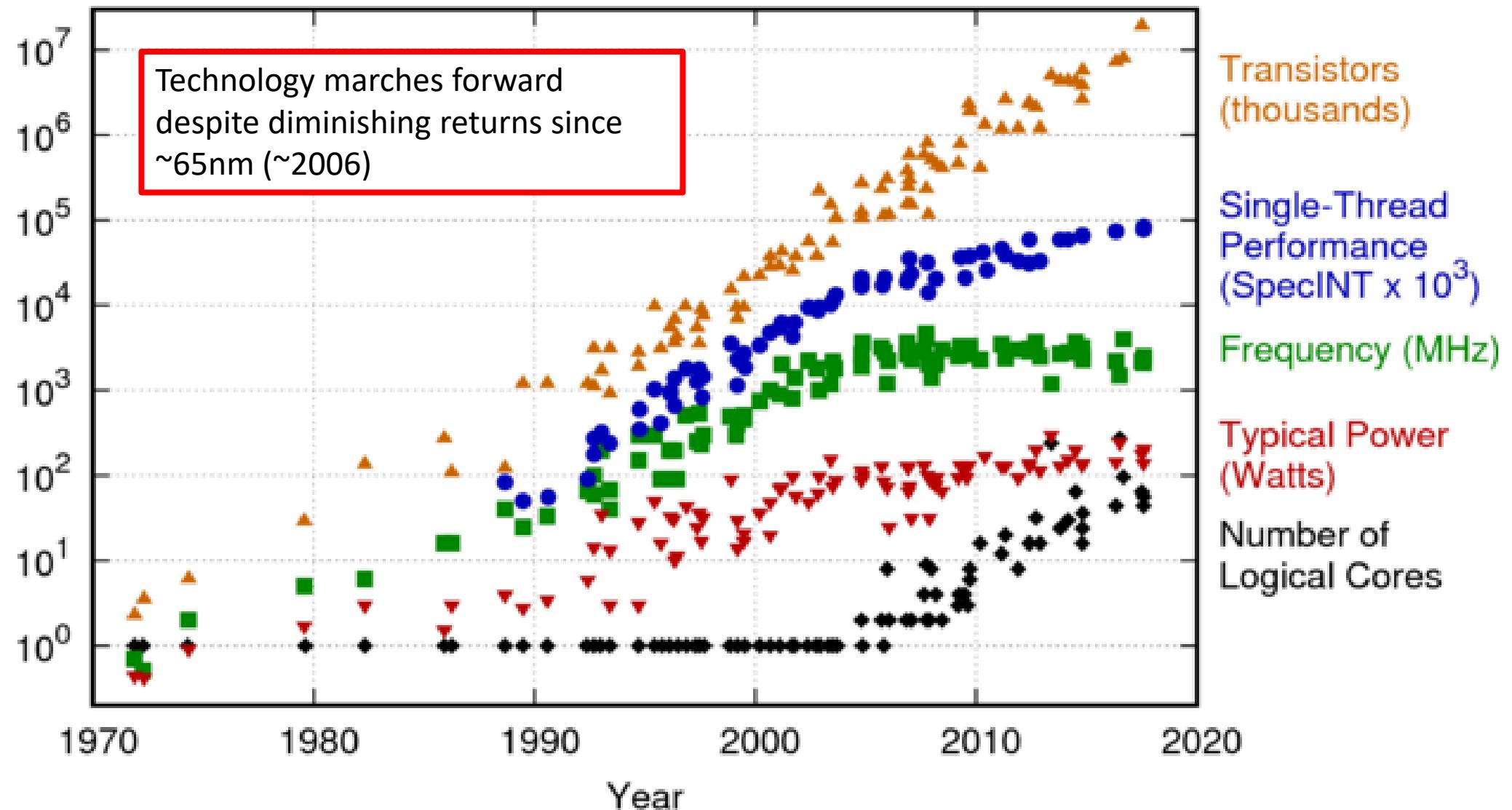




Multi Patterning

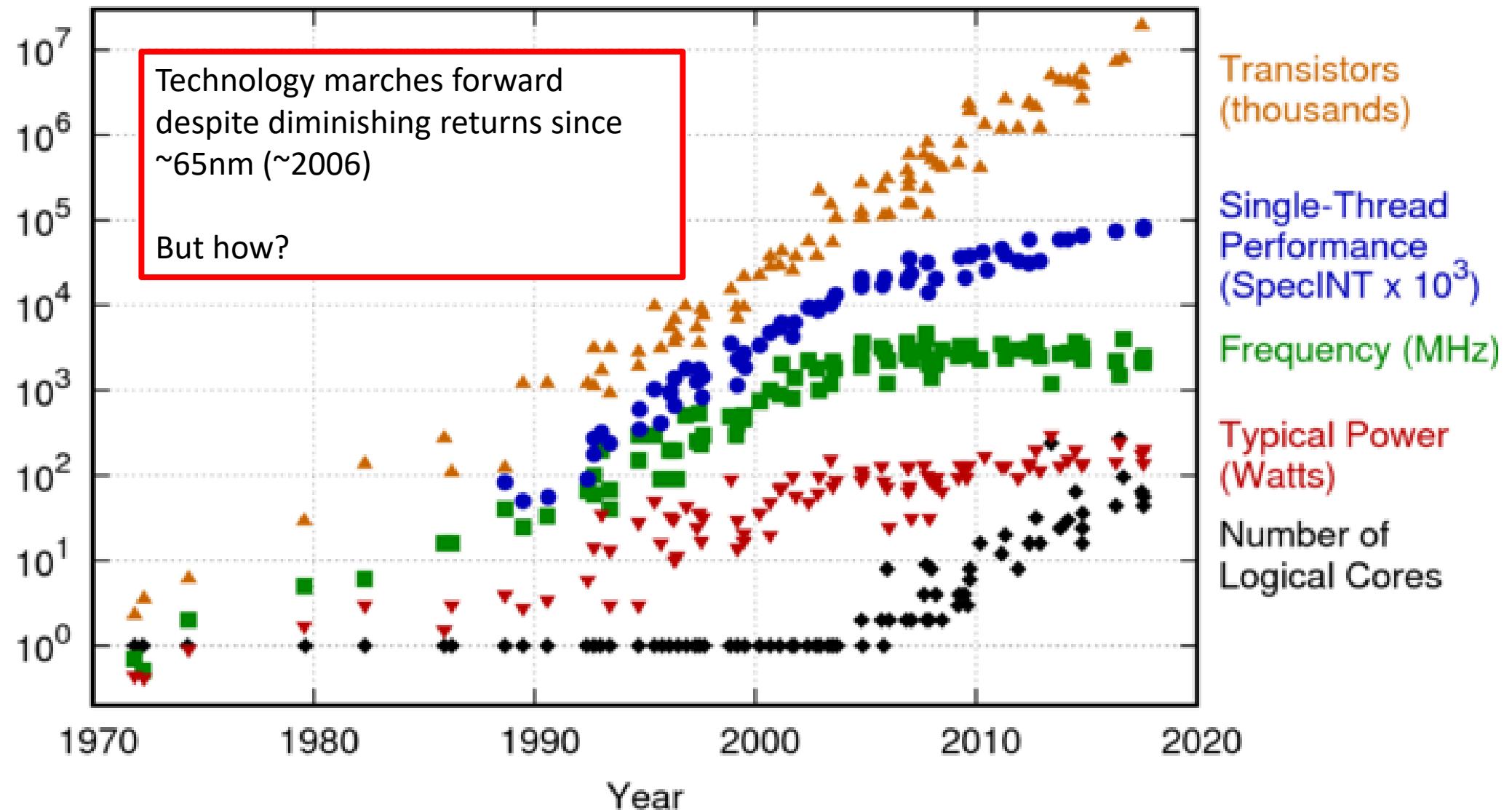


42 Years of Microprocessor Trend Data



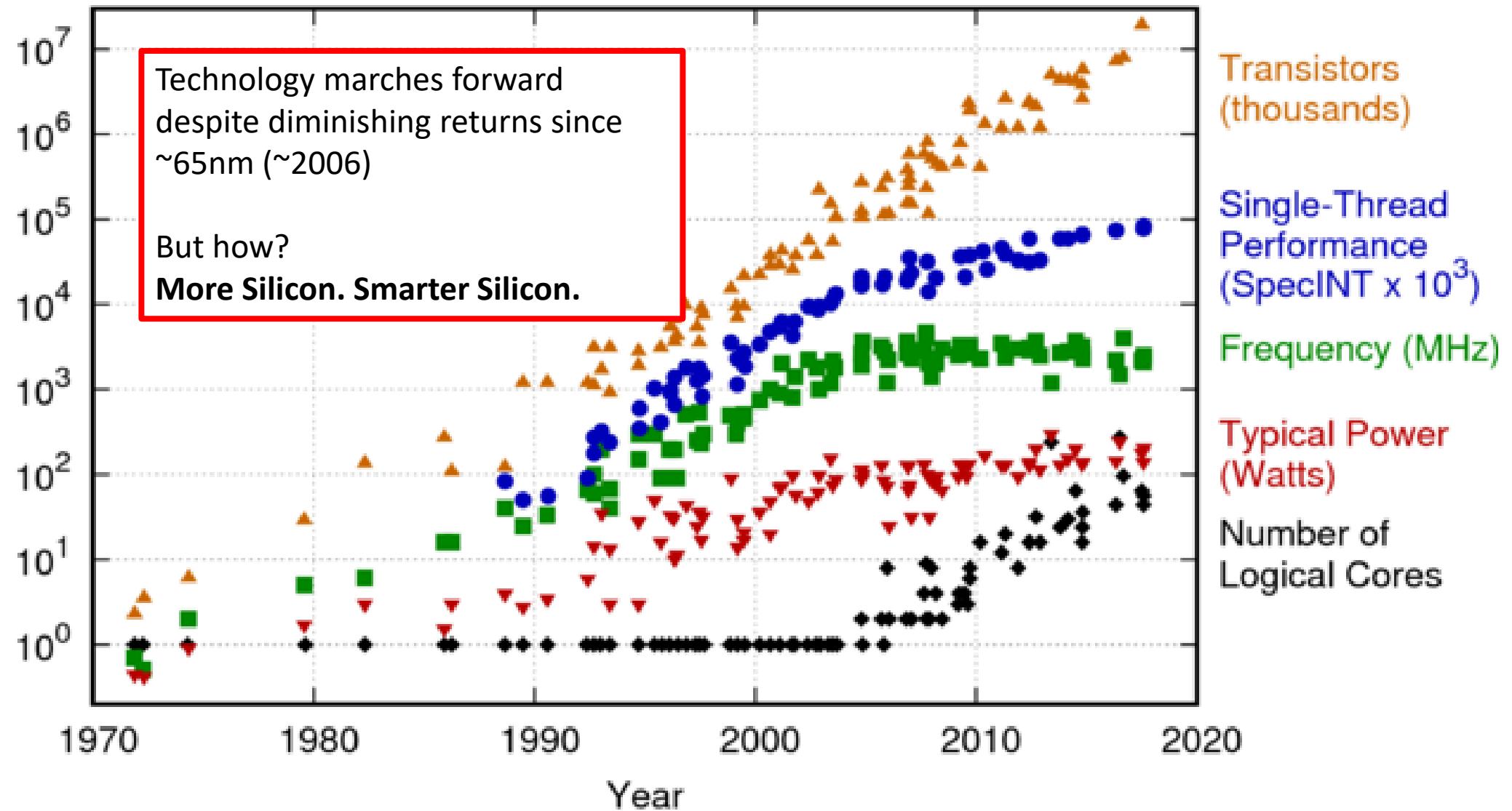
Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

42 Years of Microprocessor Trend Data



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

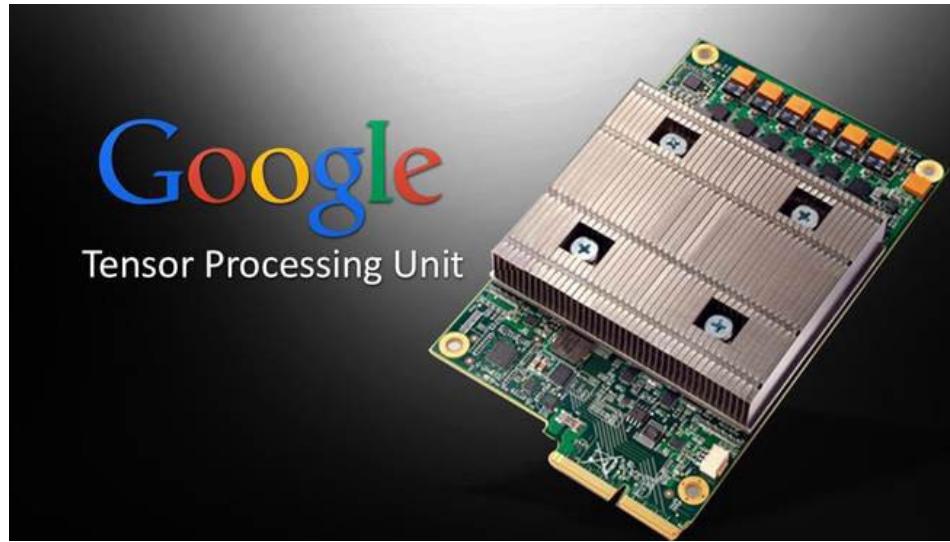
42 Years of Microprocessor Trend Data



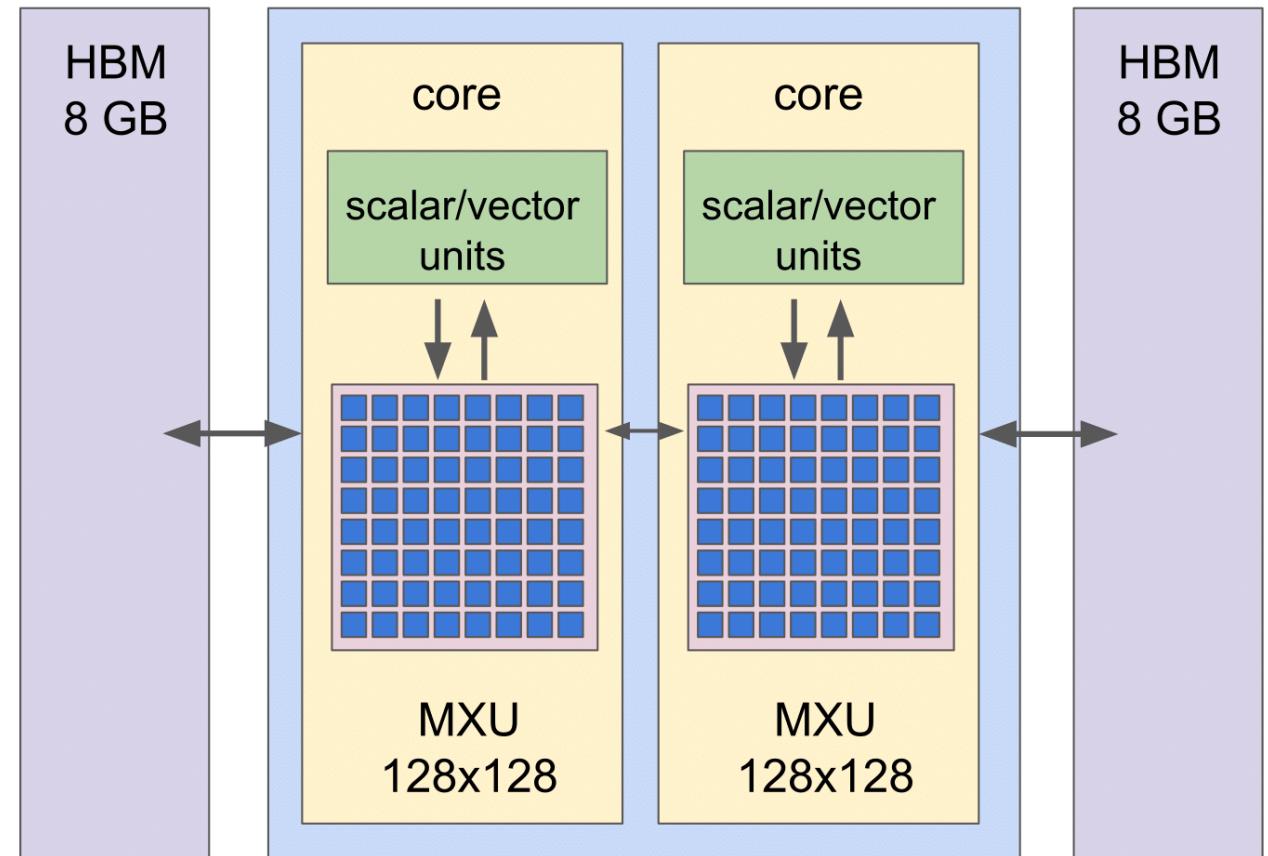
Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

Beyond Moore's Law

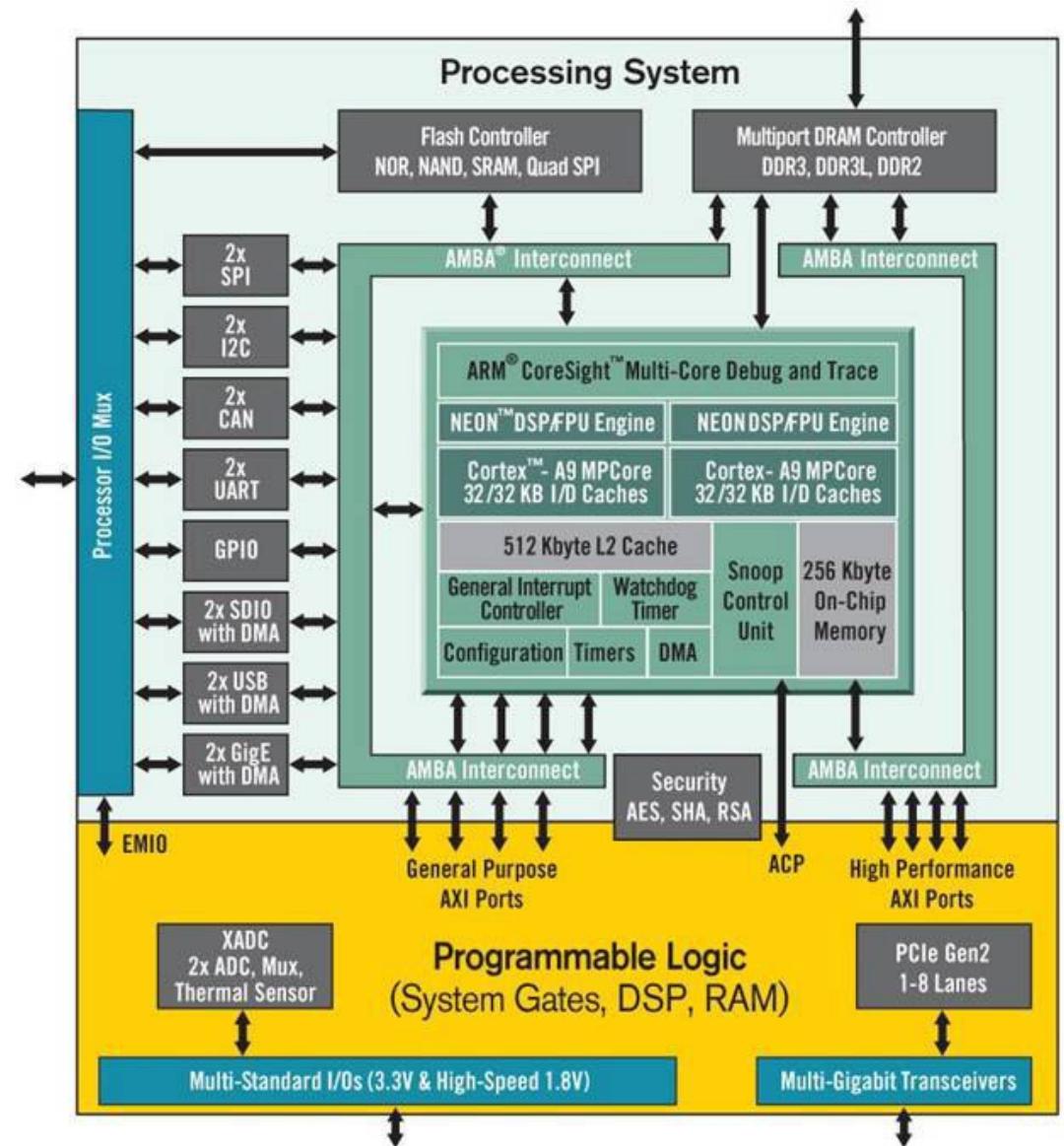
Application-Specific Acceleration



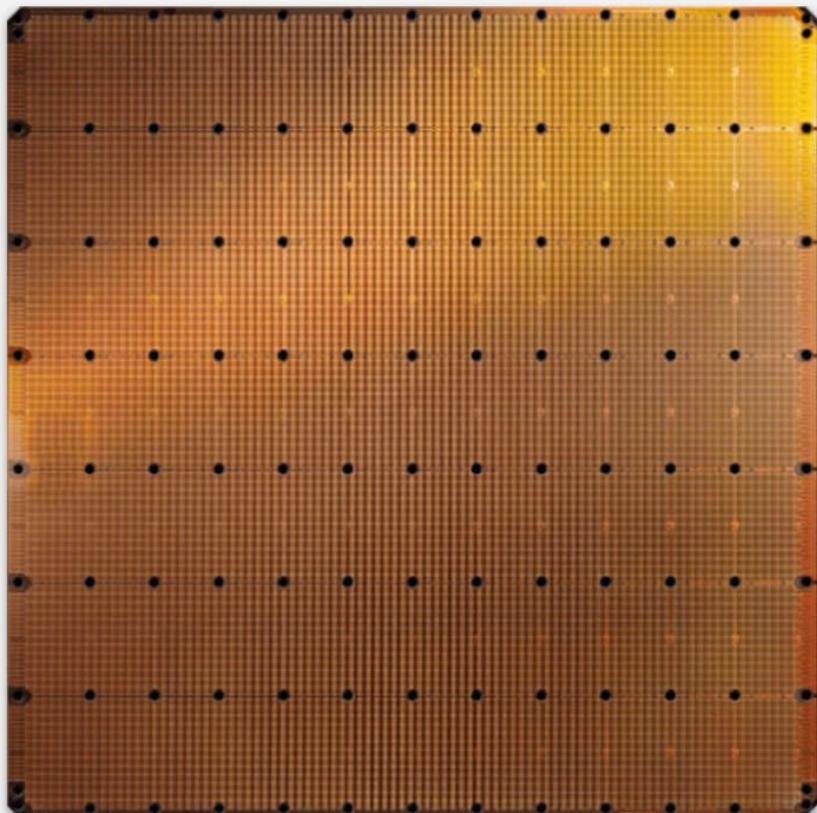
- 16 GB of HBM
- 600 GB/s mem BW
- Scalar/vector units:
32b float
- MXU: 32b float
accumulation but
reduced precision for
multipliers
- 45 TFLOPS



Heterogeneous Computing



WSI: Wafer scale integration



Cerebras WSE-2
46,225mm² Silicon
2.6 Trillion transistors



Largest GPU
826mm² Silicon
54.2 Billion transistors

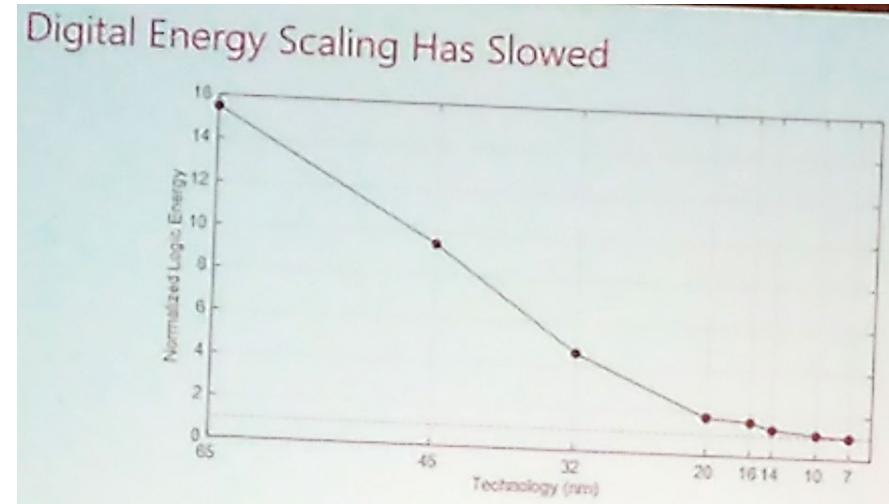
	WSE-2	A100	Cerebras Advantage
Chip Size	46,225 mm ²	826 mm ²	56 X
Cores	850,000	6912 + 432	123X
On-chip memory	40 Gigabytes	40 Megabytes	1,000 X
Memory bandwidth	20 Petabytes/sec	1.6 Terabytes/sec	12,733 X
Fabric bandwidth	220 Petabits/sec	4.8 Terabits/sec	45,833 X

Outline

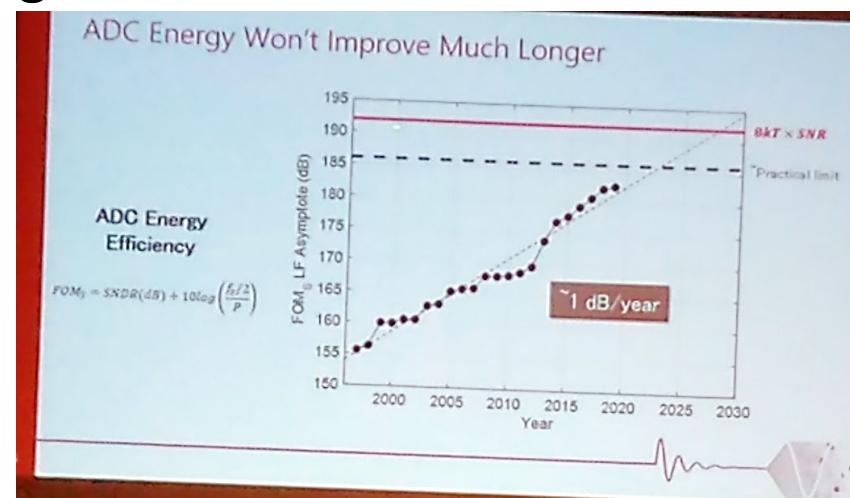
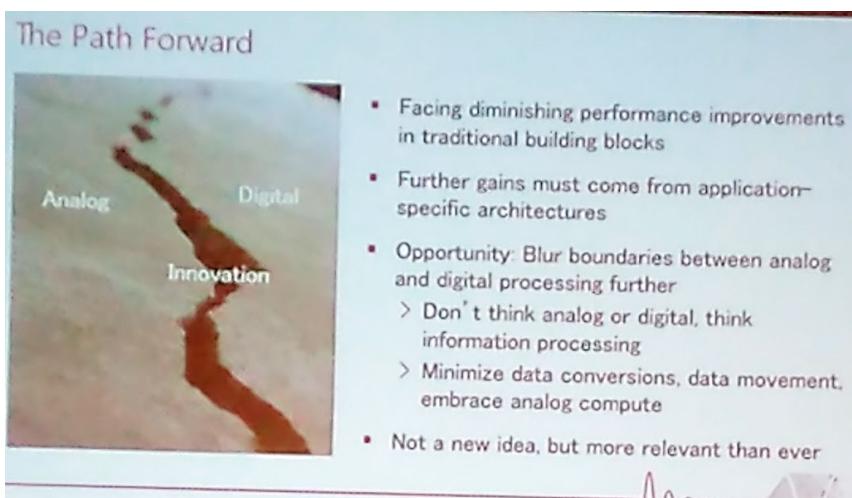
- Why do we need a methodology for digital implementation?
 - Complexity Case Study:
 - Evolution of Intel Microprocessors from 4004 to Core i7
 - Slides courtesy: www.cmosvlsi.com
- Digital Design Flow Overview
- Low-power Design

Why should you care about digital implementation?

- Digital designers:
 - Dennard Scaling/Moore's law ending
 - Next-gen digital processing relies on:
 - Heterogeneous computing /custom accelerators
 - Physical architecture innovation
- Analog designers
 - ADC Energy Ceiling
 - “Think Information Processing”

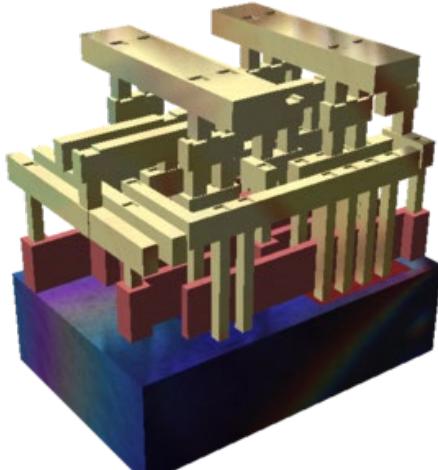


Slides: B. Murmann, BioCAS 2019

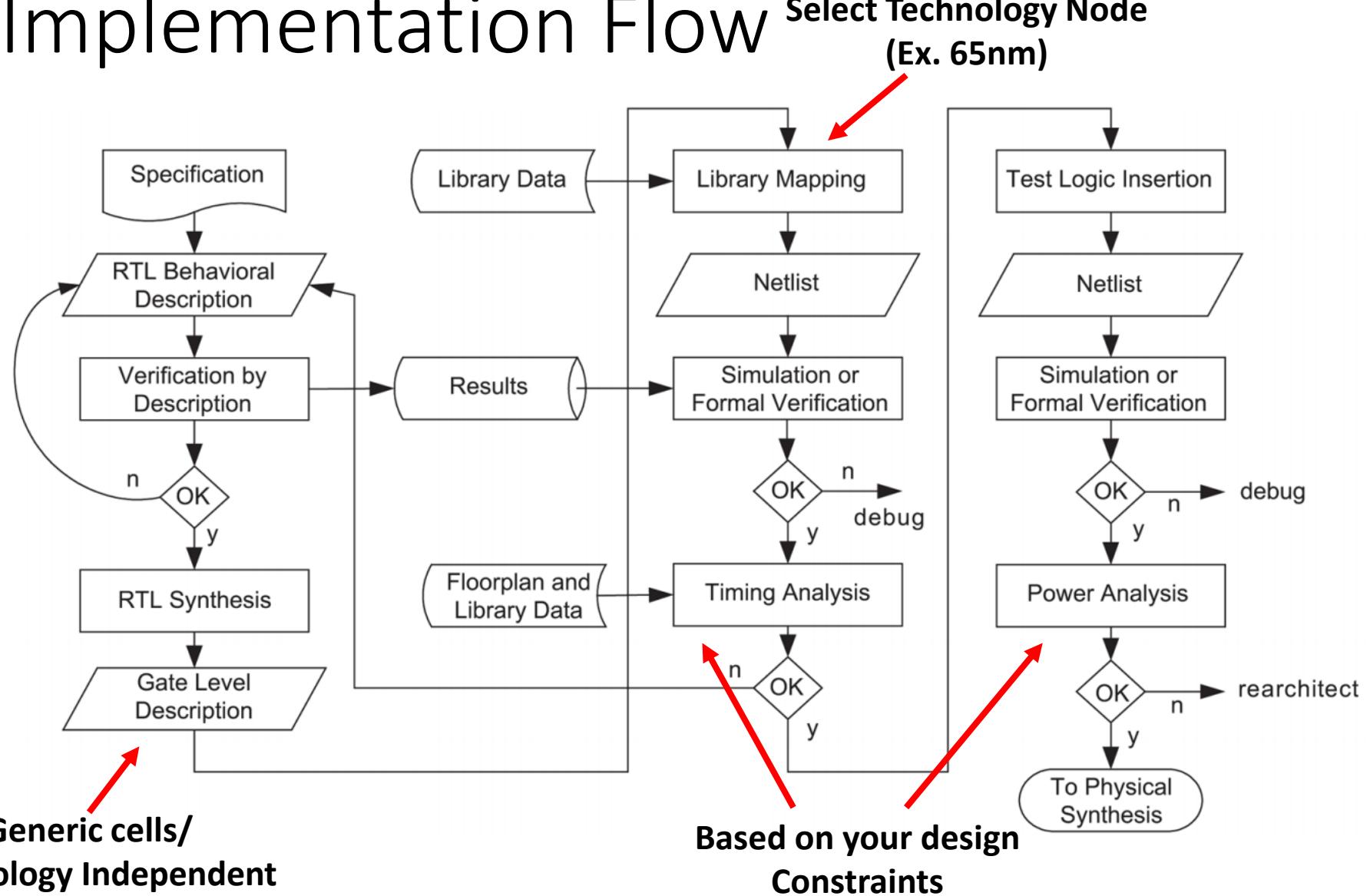


VLSI Digital Implementation Flow

```
1 module adder
2   ( input [7:0] A,
3     input [7:0] B,
4     input clk,
5     input rst_n,
6     output reg [7:0] sum,
7     output reg carry);
8
9   always @(posedge clk or negedge rst_n)
10    if (!rst_n)
11      {carry,sum} <= 0;
12    else
13      {carry,sum} <= A + B;
14 endmodule
```



Generic cells/
Technology Independent

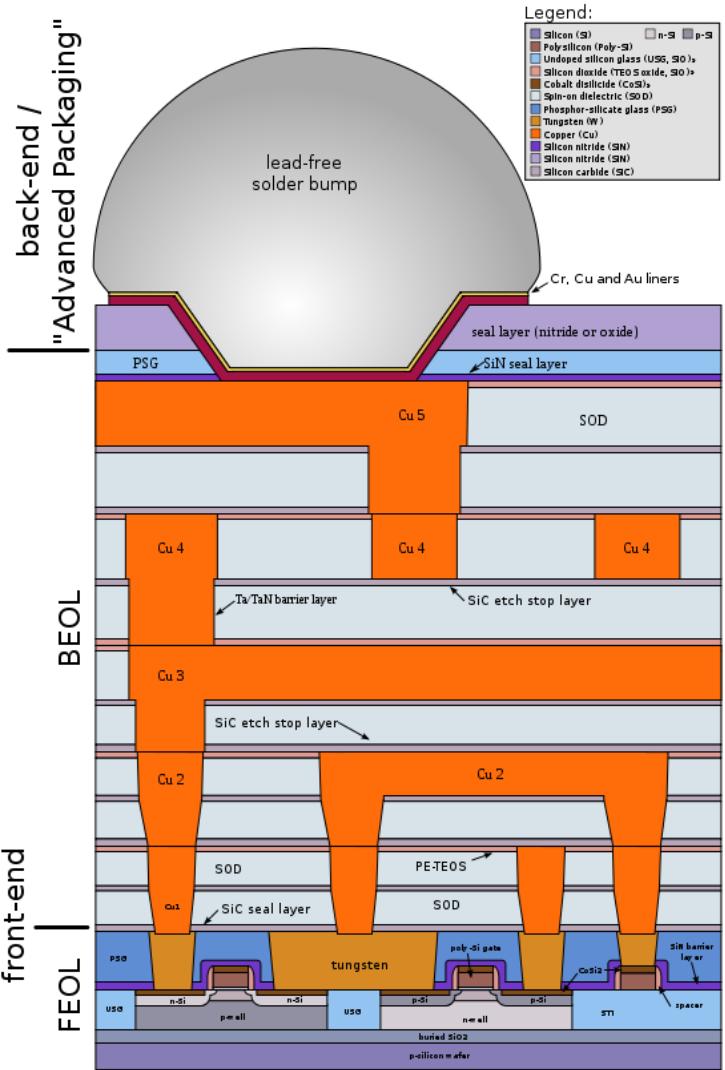
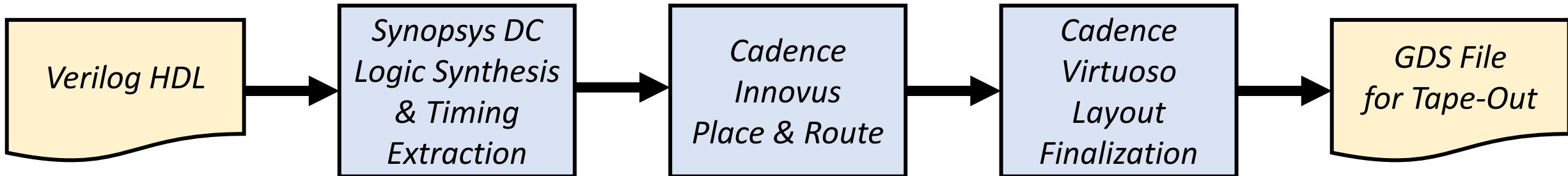


Logic Design and Verification

- Overview:
- Design starts with a specification
 - Text description or system specification language
 - Example: Matlab, Python, C, SystemC, SystemVerilog
- RTL Description
 - Most often, designer manually converts to Verilog or VHDL
 - Automated conversion from system specification to RTL possible
 - Example: Cadence C-to-Silicon Compiler
- Verification
 - Generate test-benches and run simulations to verify functionality

VLSI Digital Implementation Flow

- The purpose of the front/back-end (FEOL, BEOL) digital design flow is to convert a Verilog HDL description of a design into a GDS file for silicon fabrication.
- The Verilog HDL must first be synthesized using a set of standard cells, placed and routed within a constrained silicon die area, and lastly exported as a GDS file.



Agenda

- Tutorial 3:
 - Digital Design Flow
- Tutorial 4:
 - ECE1388 Project 3 - 4-bit array multiplier block
- Case Study:
 - Why all this is useful – 2 recent design examples
- Tutorial 5: More Advanced Design Techniques and the Future
 - ML-Driven CAD: Google Keynote