TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Institut für Computertechnik
Institute of Computer Technology

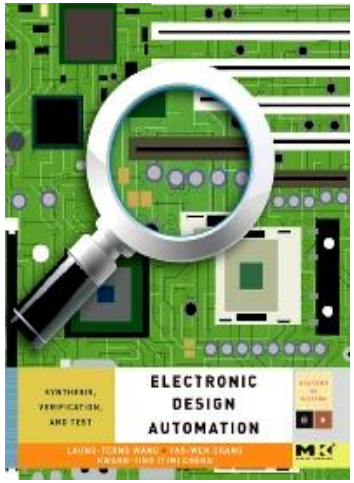
Digitale Integrierte Schaltungen 2018 WS

Prof. Axel Jantsch

Verwendete Literatur

Lehrbuch Digitaltechnik
Jürgen Reichardt
3. Auflage, Oldenbourg 2013
https://www.degruyter.com/viewbooktoc/product/228761

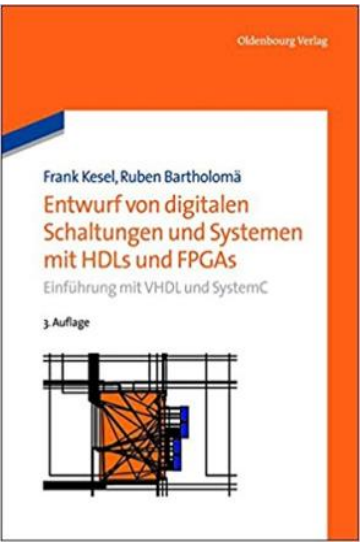
Kapitel: 2,4, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17



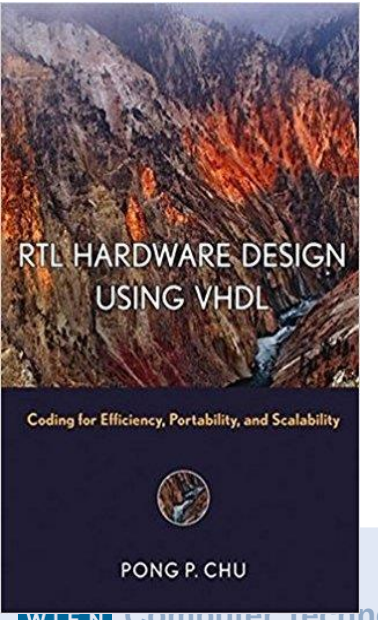
„Logic Synthesis in a Nutshell“,
Jie-Hong Jiang and Srinivas Devadas,
http://www.sciencedirect.com/science/article/pii/B9780123743640500138

Kapitel: 61, 6.2, 6.3, 6.4

Weiterführende Literatur

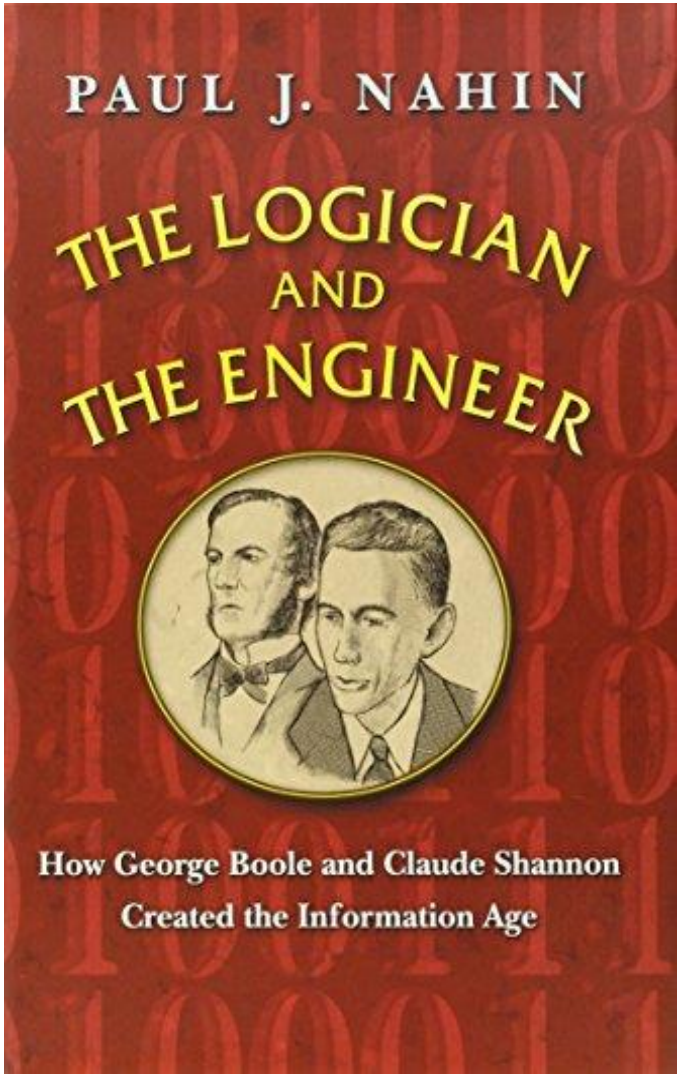


Entwurf von digitalen Schaltungen und Systemen mit HDLs und FPGAs : Einführung mit VHDL und SystemC
Frank Kesel, Ruben Bartholomä
München : Oldenbourg 2013



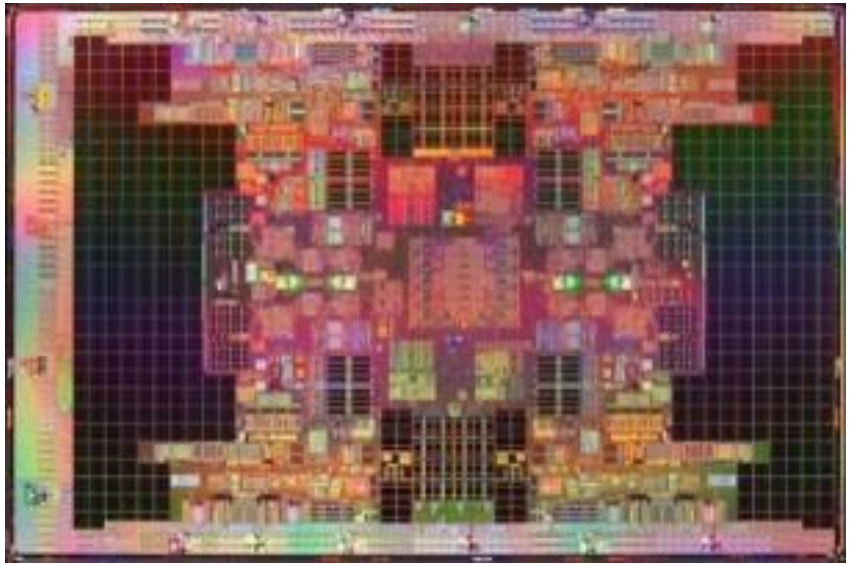
RTL Hardware Design Using VHDL
P. Pong Chu
Wiley, 2006
http://ieeexplore.ieee.org/xpl/bkabstractplus.jsp?bkn=5237648

Empfohlene Lektüre für den Abend:



Vorspann

- Was sind integrierte Schaltungen?
- Was ist VLSI?
- Was sind ASICs?
- Was sind SoCs?

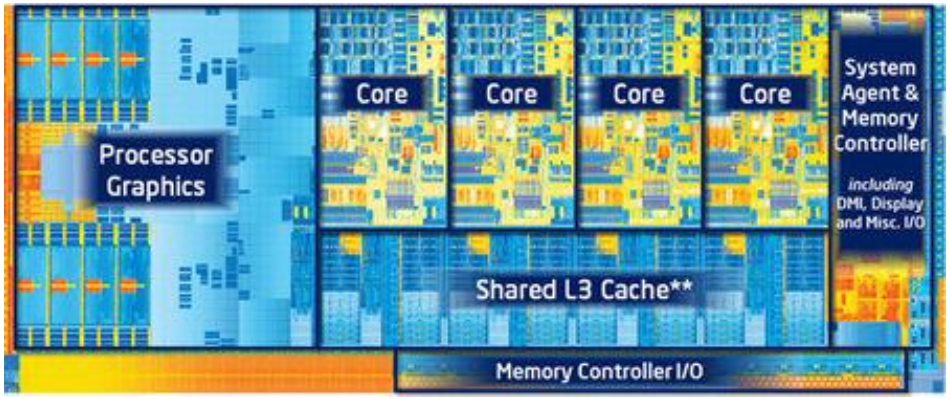


Intel Tukwila

2.046 Billion Transistors

21.5 x 32.5 mm^2, 65nm

170W, 2GHz

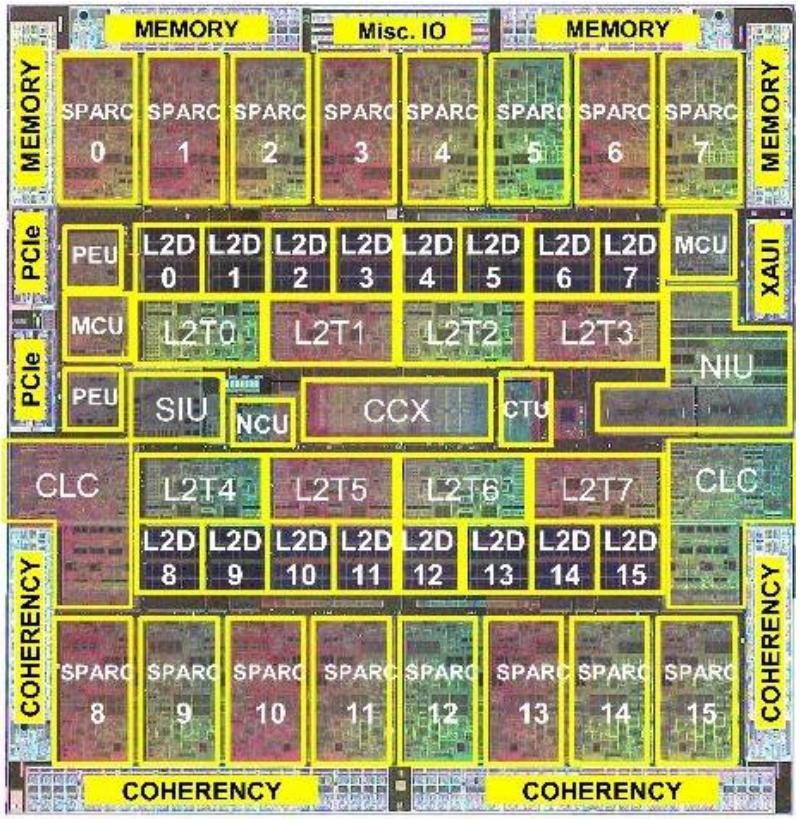


Intel Ivy Bridge

1.4 Billion Transistors

160mm^2, 22nm

77W, 3.9 GHz

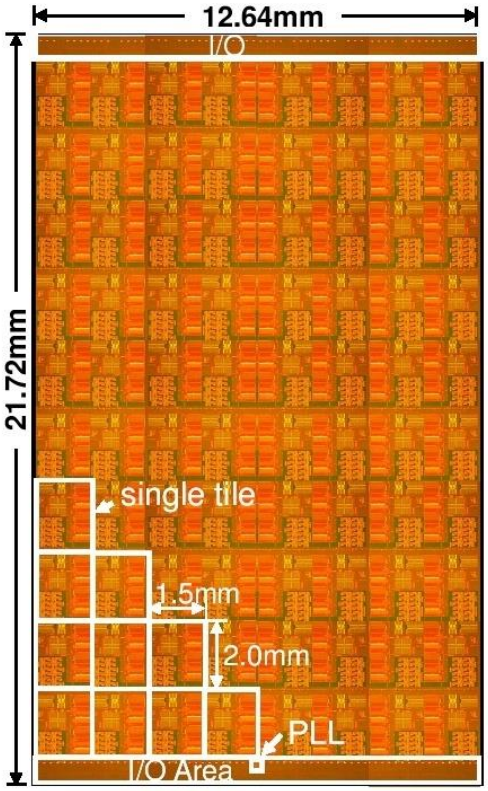


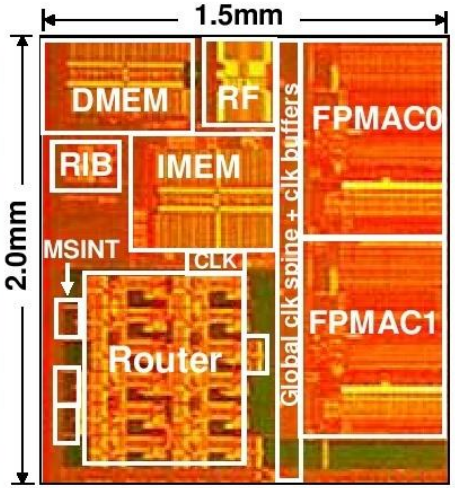
Sun Niagara 3

1 Billion Transistors

16 cores, 377mm^2, 40nm

1.67 GHz, 60W





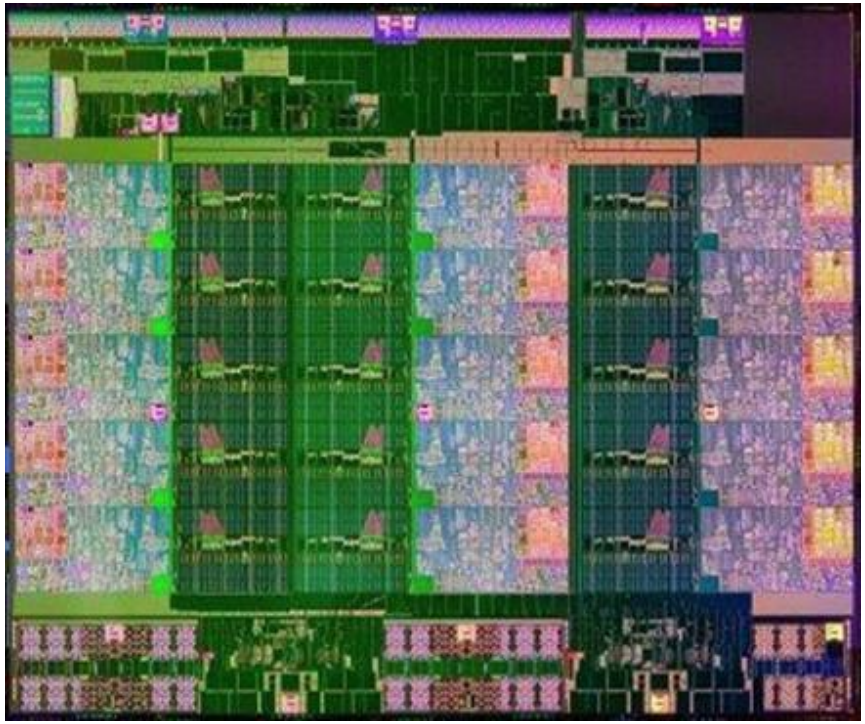
Technology	65nm CMOS Process
Interconnect	1 poly, 8 metal (Cu)
Transistors	100 Million
Die Area	275mm^2
Tile area	3mm^2
Package	1248 pin LGA, 14 layers, 343 signal pins

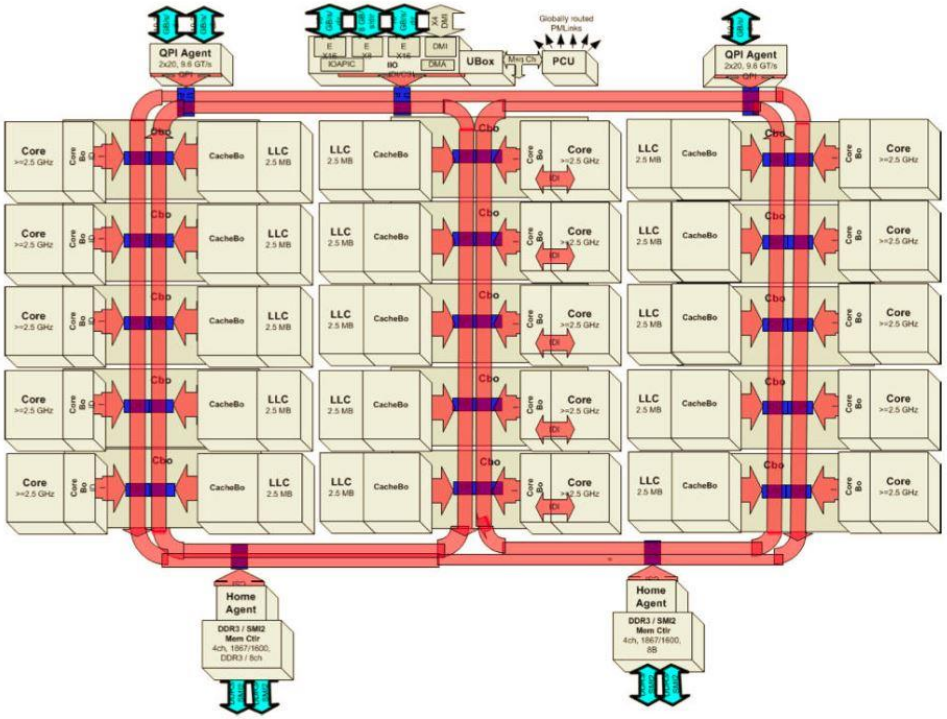
Intel Teraflop

100 Million Transistors

80 cores, 275mm^2, 65nm

3.67GHz, 62W





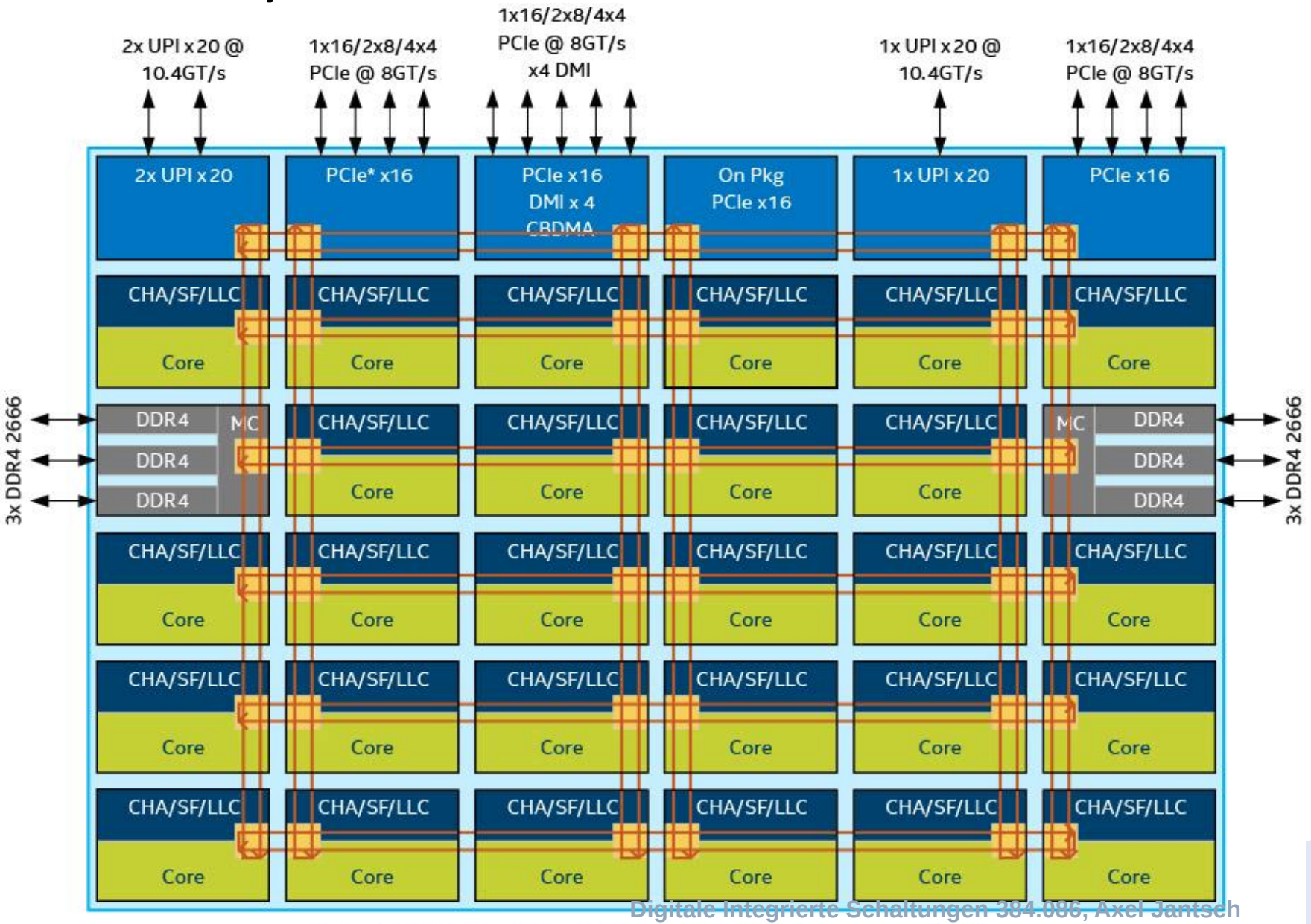
Intel Westmere Xeon E2

4.31 Billion Transistors

15 cores, 541mm^2, 22nm

2.8GHz, 155W

Intel Skylake Server Architecture

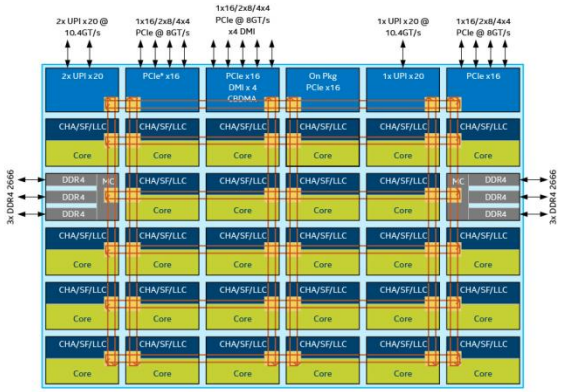


Intel Skylake Server Architecture

10-28 cores (12-30 tiles), 14nm, 694 mm2 (30 tiles)

Cache

- L0 µOP cache:
  - 1536 µOPs/core,
  - 8-way set associative
  - 32 sets, 6-µOP line size
- L1 I Cache:
  - 32 KiB/core,
  - 8-way set associative
  - 64 sets, 64 B line size
- L1D Cache:
  - 32 KiB/core,
  - 8-way set associative
  - 64 sets, 64 B line size
  - 4 - 5 cycles latency
  - Write-back policy



- L2 Cache:
  - 1 MiB/core,
  - 16-way set associative
  - 64 B line size
  - Write-back policy
  - 14 cycles latency
- L3 Cache:
  - 1.375 MiB/core,
  - 11-way set associative,
  - shared across all cores
  - 2,048 sets, 64 B line size
  - Write-back policy
  - 50-70 cycles latency

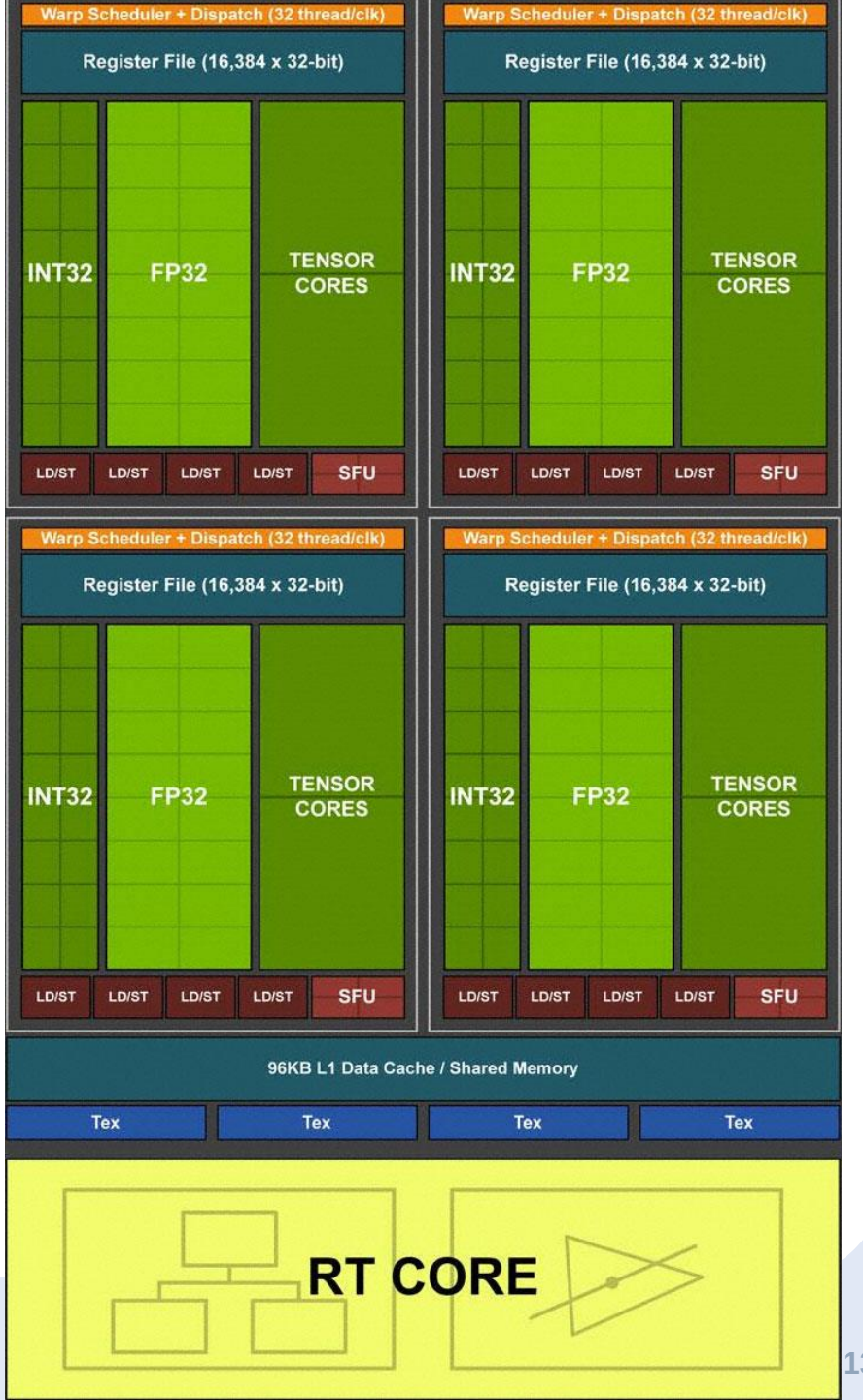
Nvidia Turin TU102

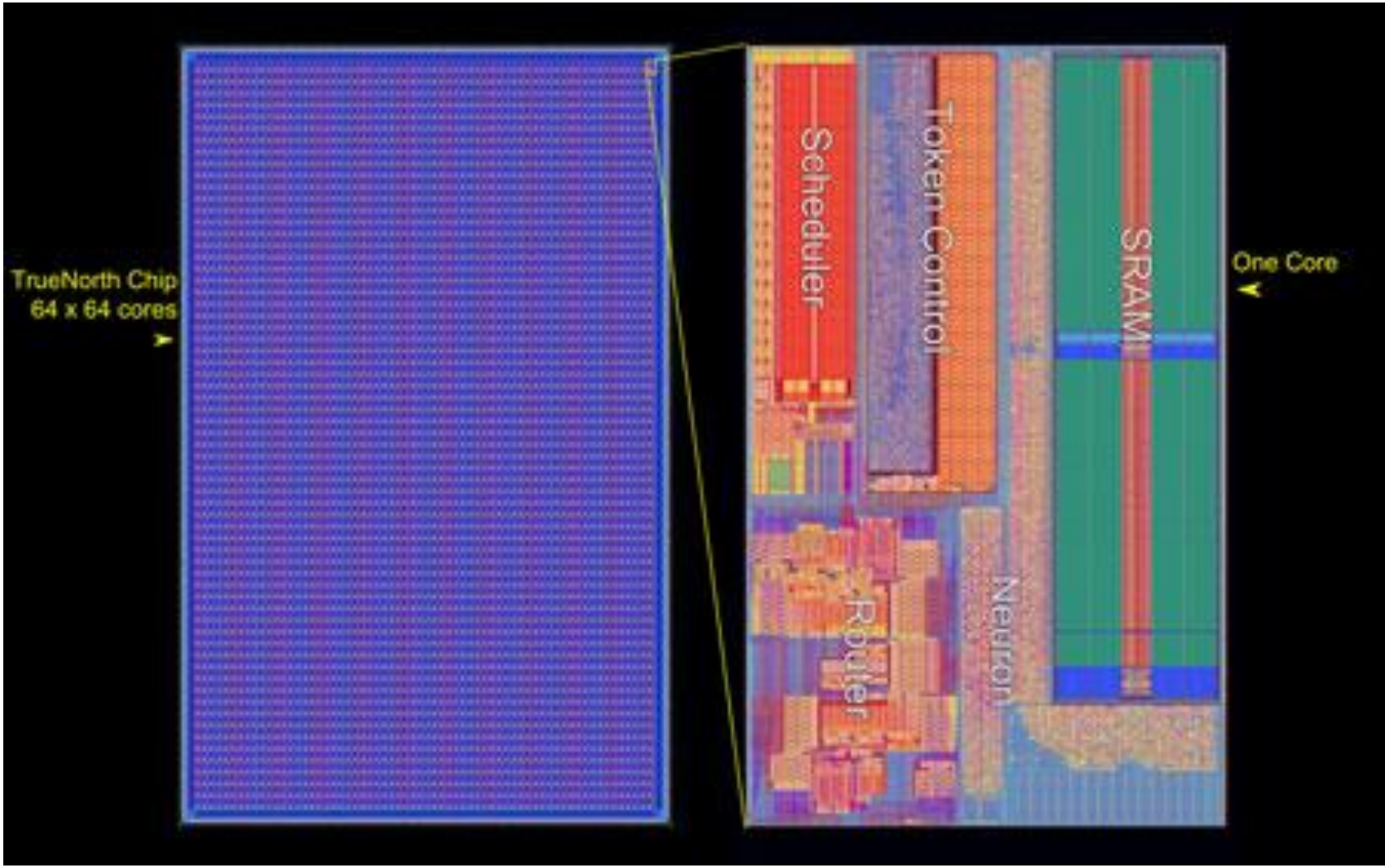
	Full TU201
Process (nm)	12
Transistors (billion)	18.6
Sie size (mm^2)	754
Streaming Multiprocessors (SM)	72
CUDA Cores	4608 (64/SM)
Tensor Cores	576 (8/SM)
RT Cores	72
Clock (MHz)	≤ 1500
CUDA TFlops (FP32)	13.8
L1 Cache (MB)	6.912
L2 Cache (MB)	6
Bus width	384
Power (W)	200-250
Bandwidth (GB/s)	672

N


PCI Express 3.0 Host Interface
GigaThread Engine
Memory Controller
Memory Controller
Memory Controller
Memory Controller
Memory Controller
Memory Controller
Memory Controller
Memory Controller
Memory Controller
Memory Controller
Memory Controller
Memory Controller
GPC
Raster Engine
TPC
SM
RT CORE
L2 Cache
High-Speed Hub
NVLink – Two x8 Links

Nvidia Turin TU102 Streaming Multiprocessor





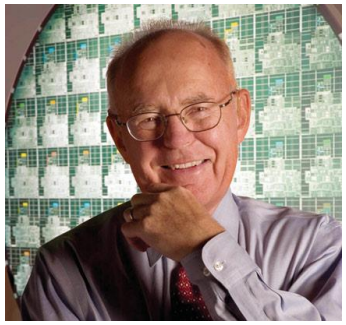
IBM TrueNorth

5.4 Billion Transistors

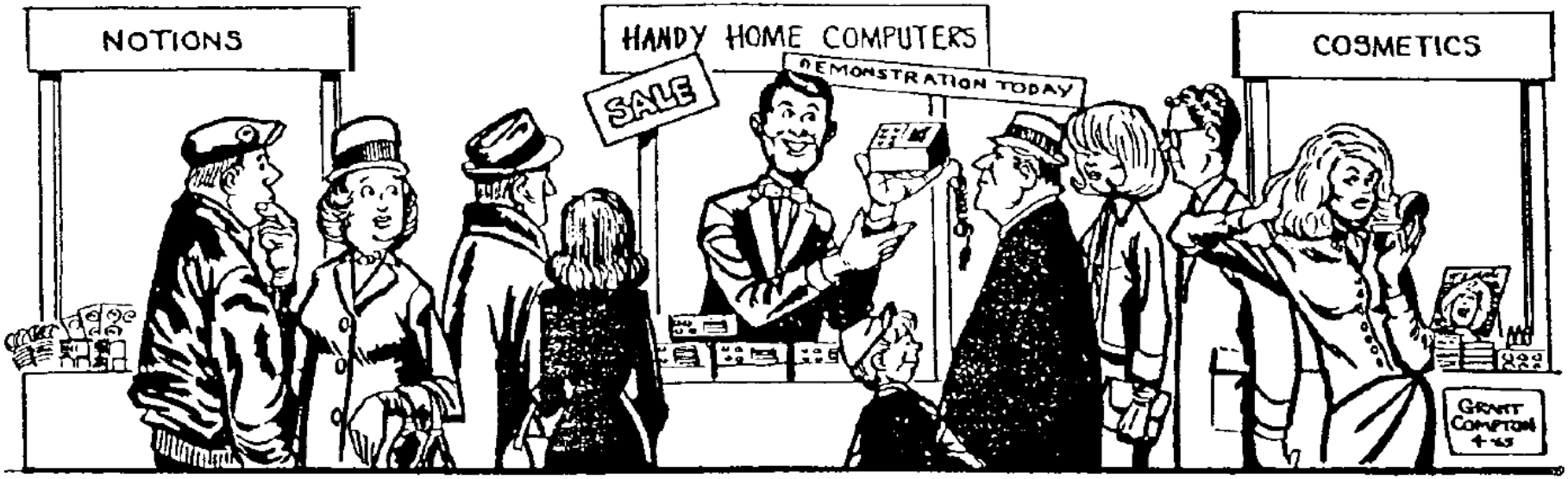
4096 cores, 350mm^2, 28nm

70mW

1965

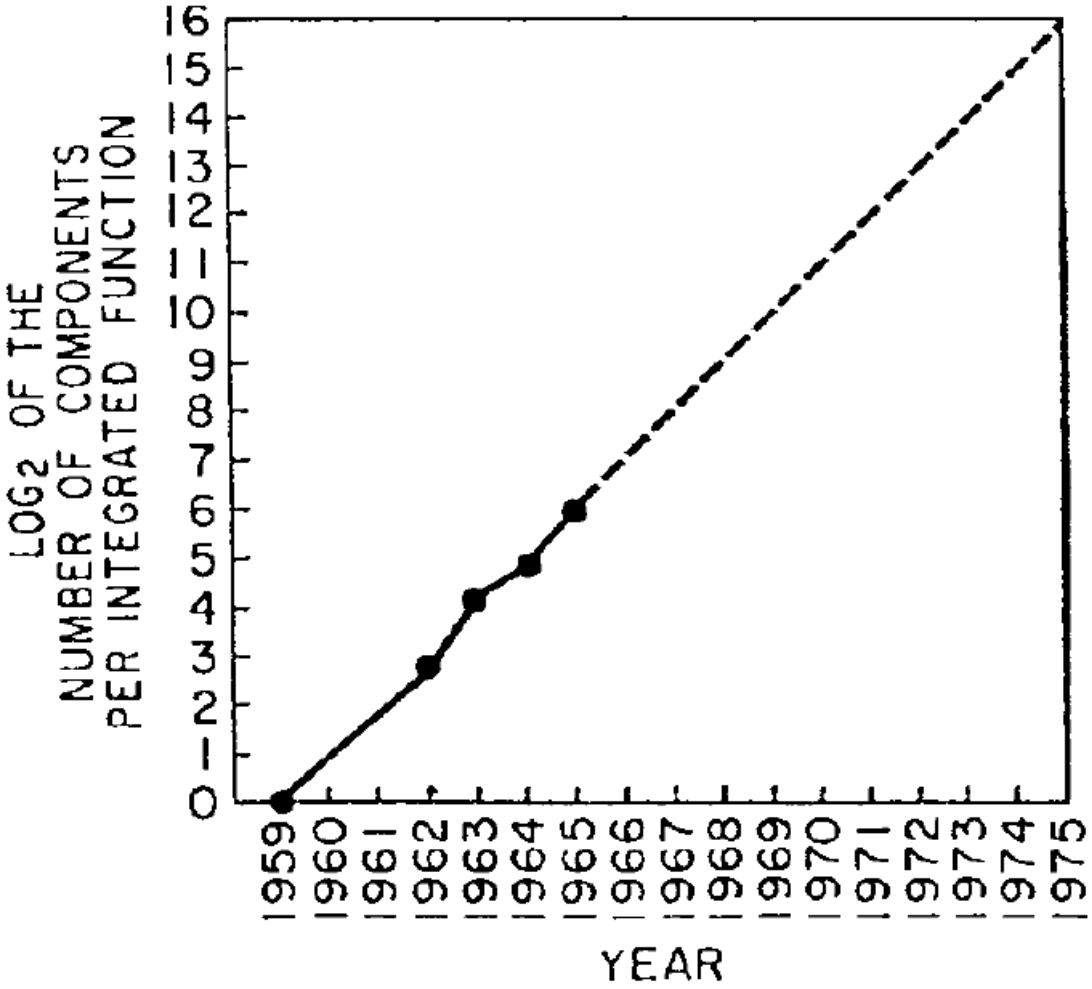


born 1929



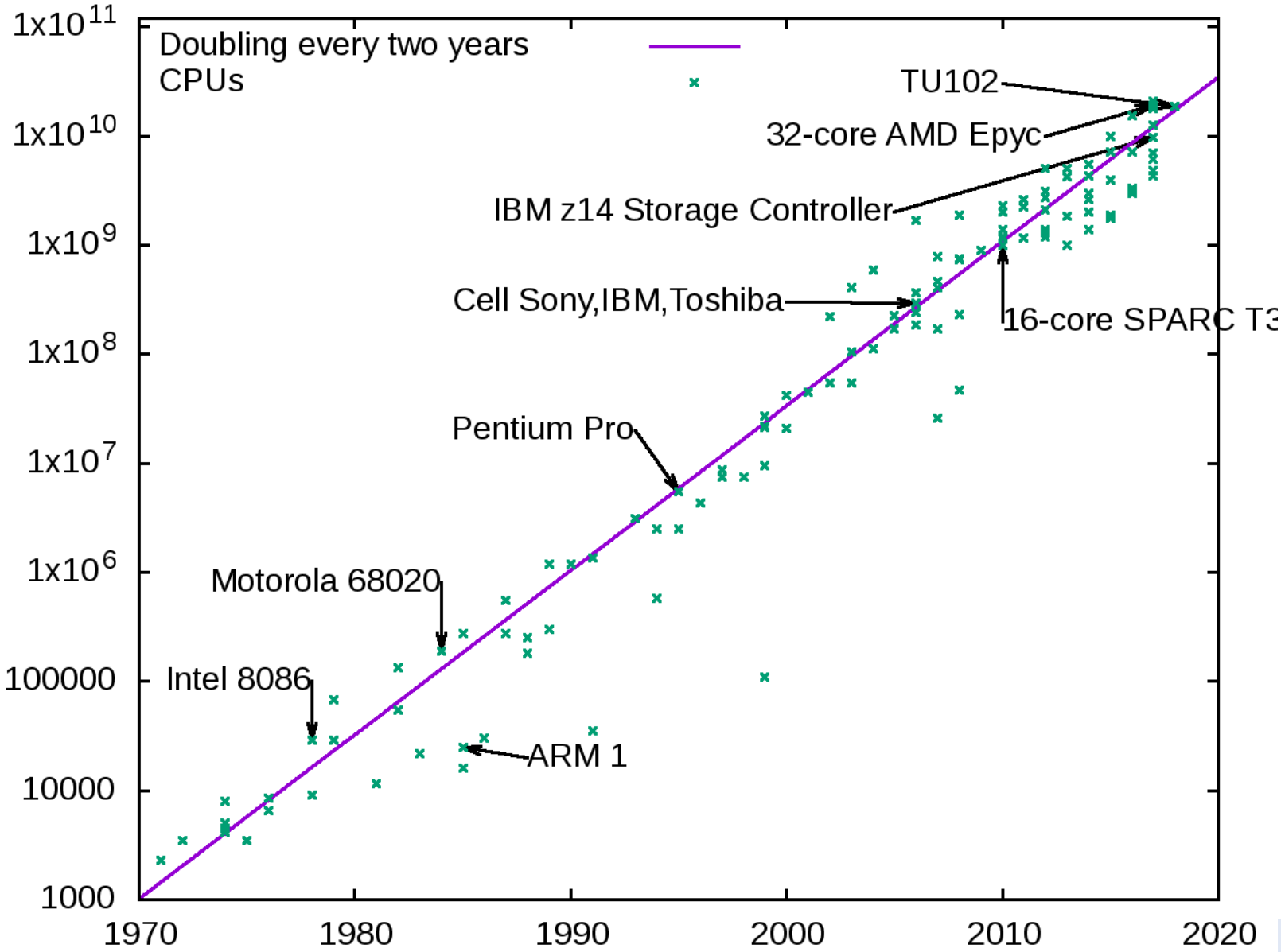
Gordon E. Moore, "Cramming More Components onto Integrated Circuits", *Electronics*, pp. 114-117, April 1965.

Moore‘s Law

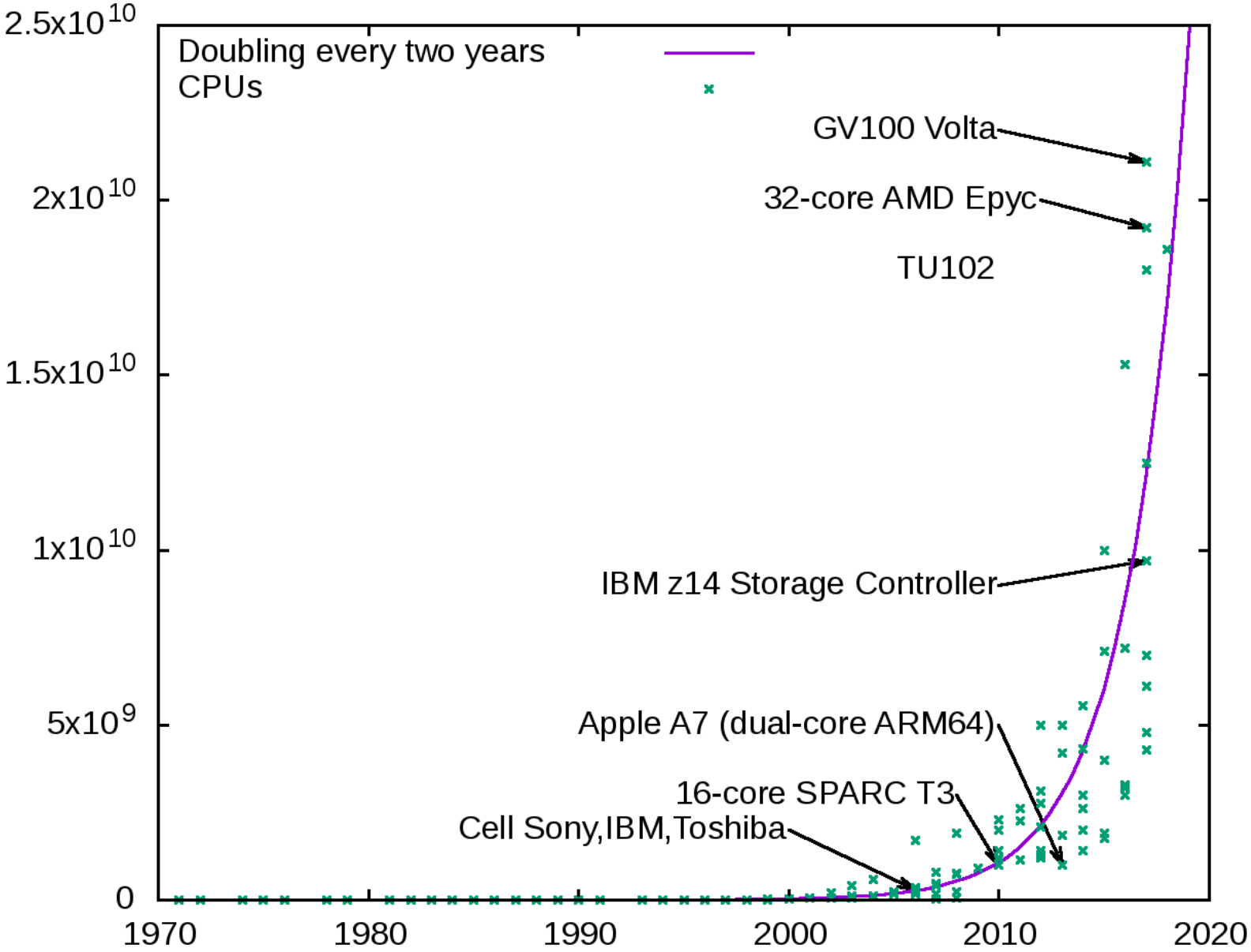


Gordon E. Moore, "Cramming More Components onto Integrated Circuits", *Electronics*, pp. 114-117, April 1965.

CPUs Transistor Count



CPUs Transistor Count



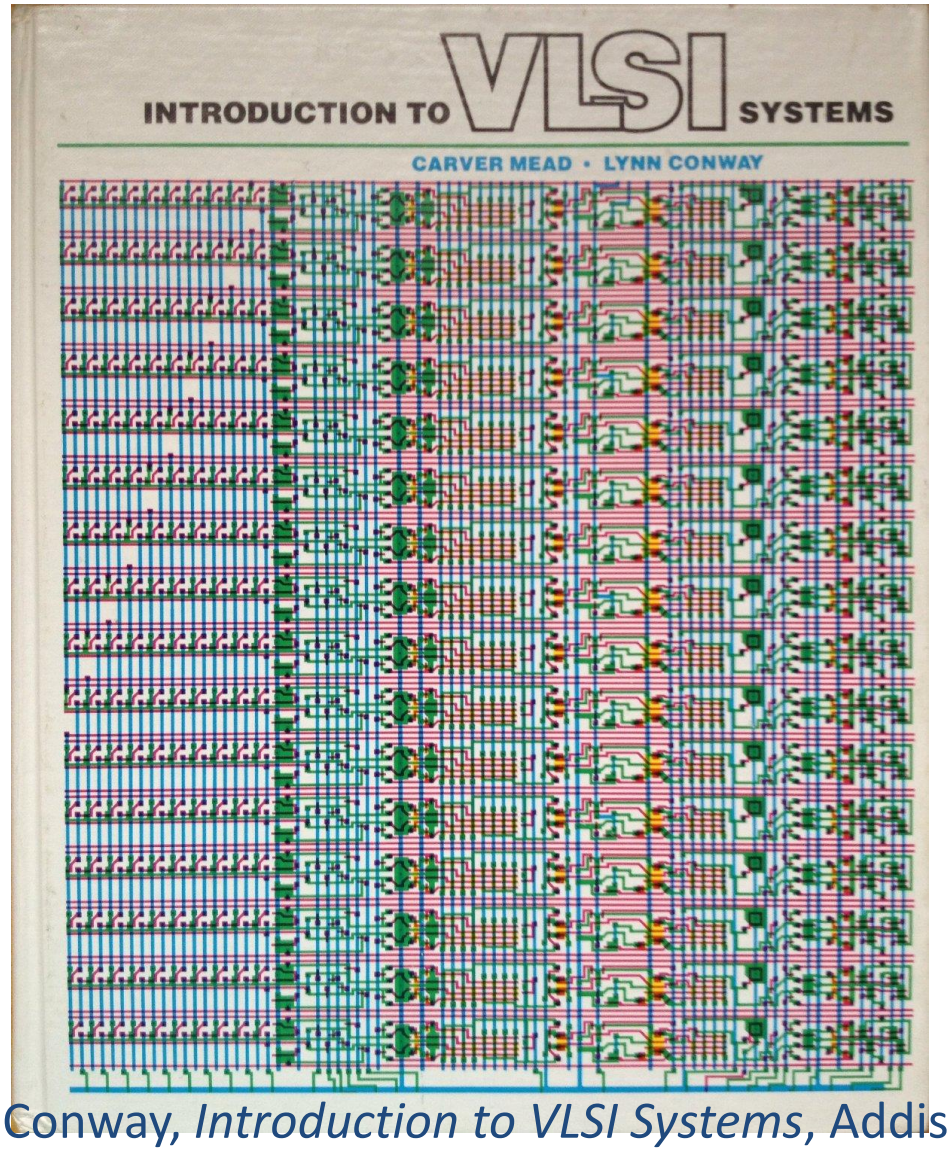
Three Main Deflection Points

- Mead-Conway Revolution in 1980
- Systems-on-Chip in 2000
- Multi-Core Revolution in 2010

1980
20 000 Transistoren

Mead-Conway Revolution

born 1934

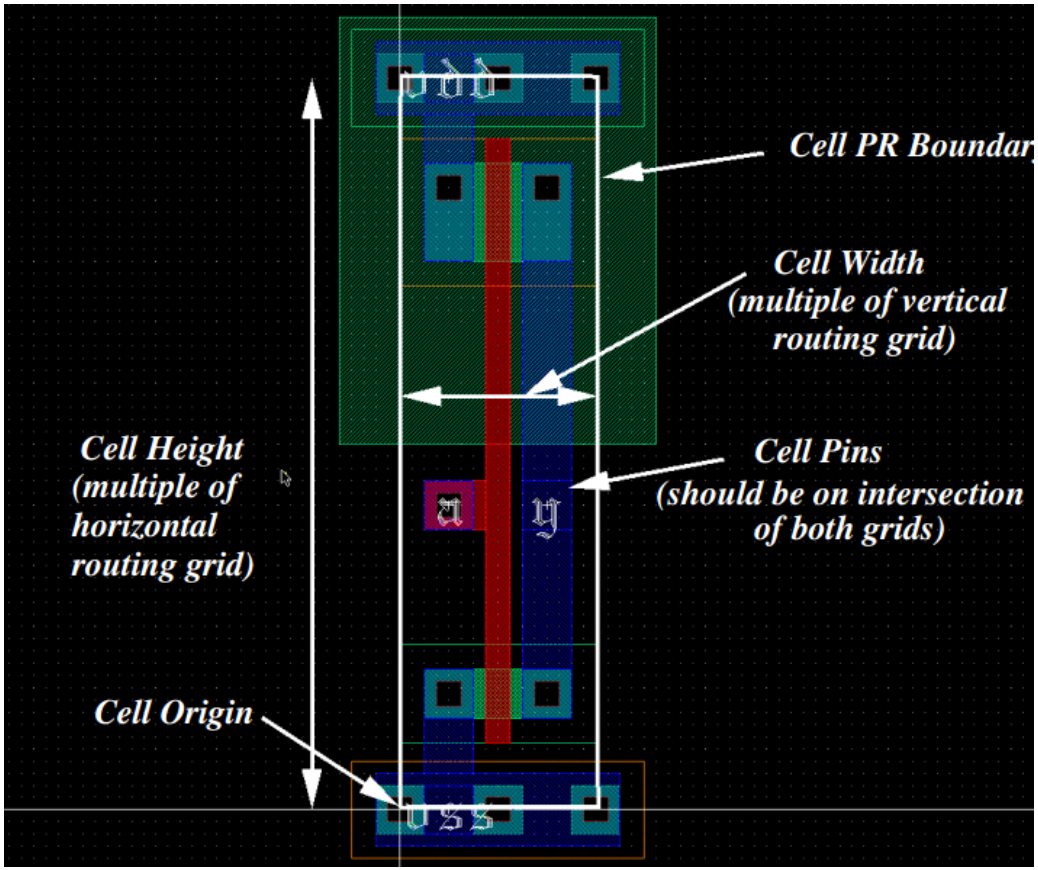


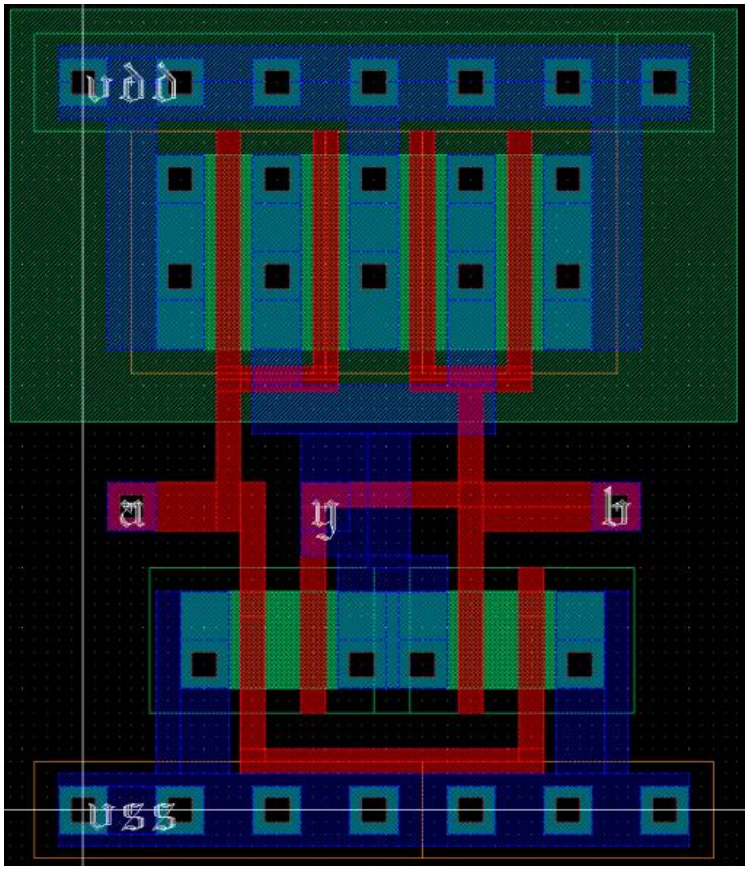
born 1938

Carver Mead and Lynn Conway, *Introduction to VLSI Systems*, Addison-Wesley, 1980.

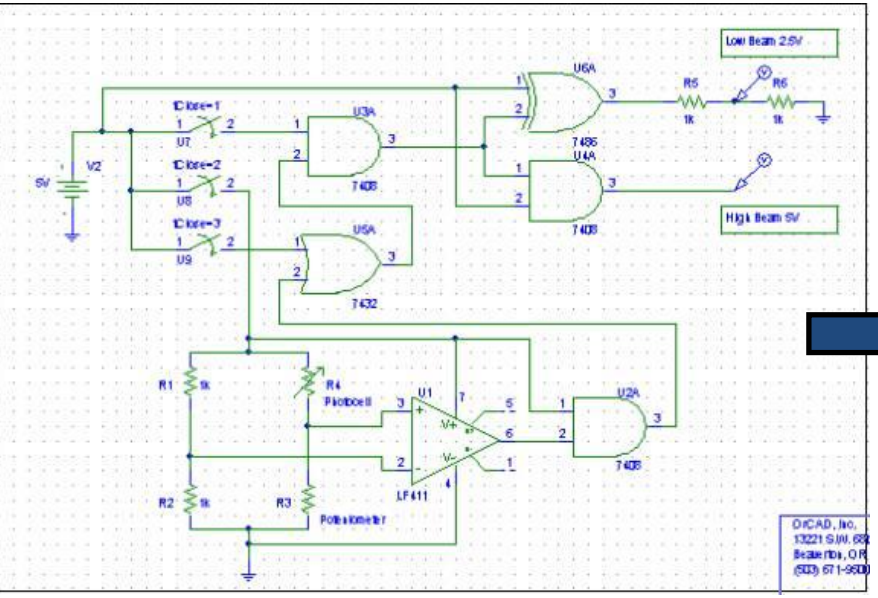
1980 – 20 000 Transistors

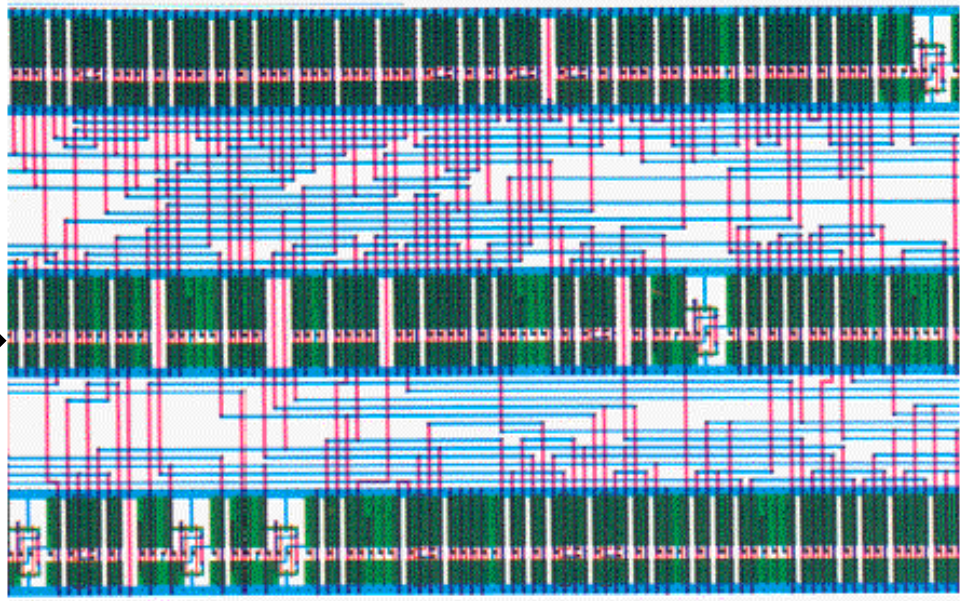
Standard Cells



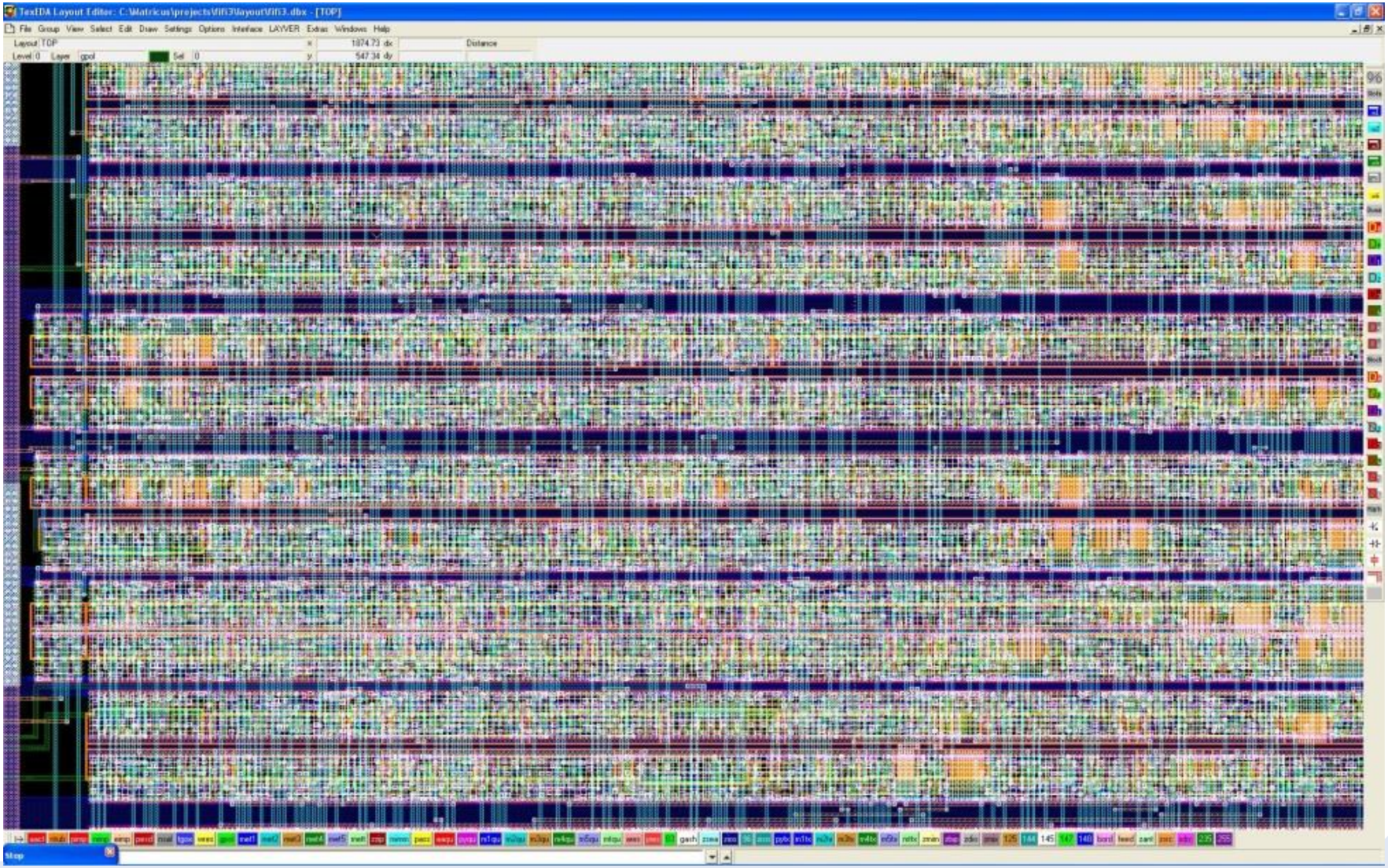


Standard Cells





Standard Cells



Logic Synthesis

x \ y	∧ 0	∧ 1
0	0	0
1	0	1

x \ y	∨ 0	∨ 1
0	0	1
1	1	1

x \ y	→ 0	→ 1
0	1	1
1	0	1

x \ y	⊕ 0	⊕ 1
0	0	1
1	1	0

Figure 1. Truth tables

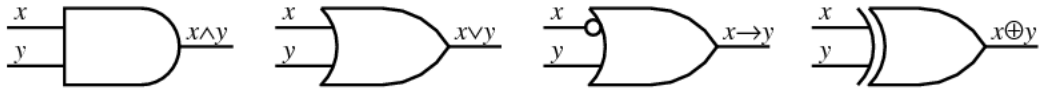


Figure 2. Logic gates

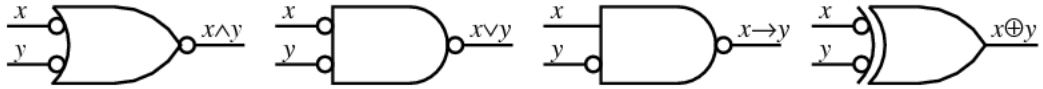
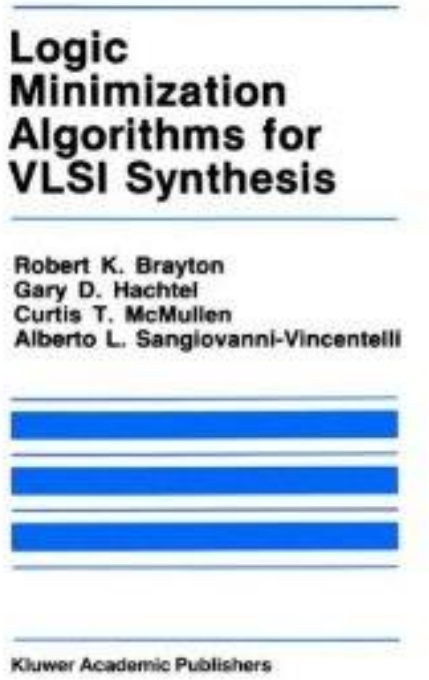


Figure 3. De Morgan equivalents

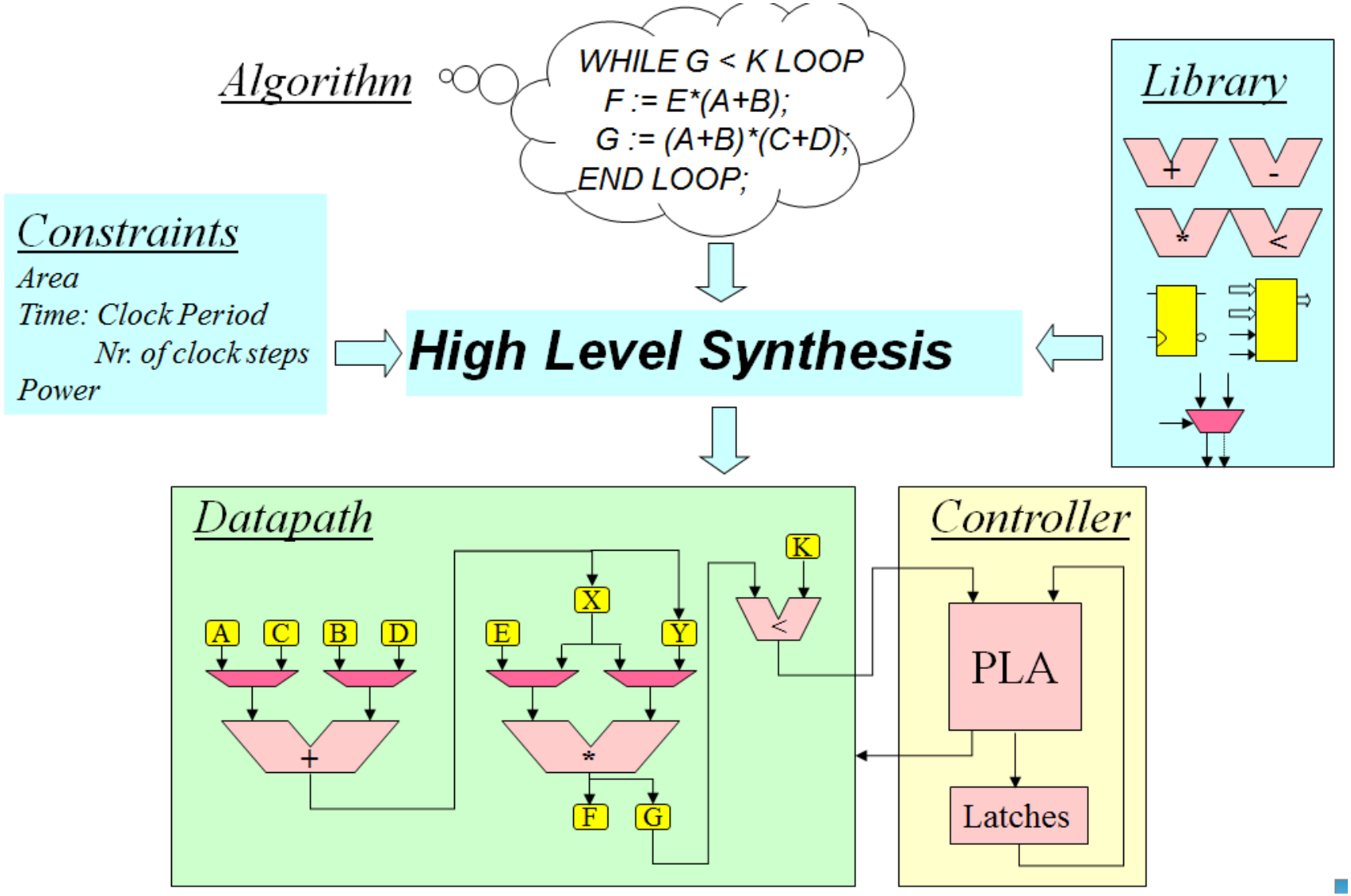
Figure 4. Venn diagrams



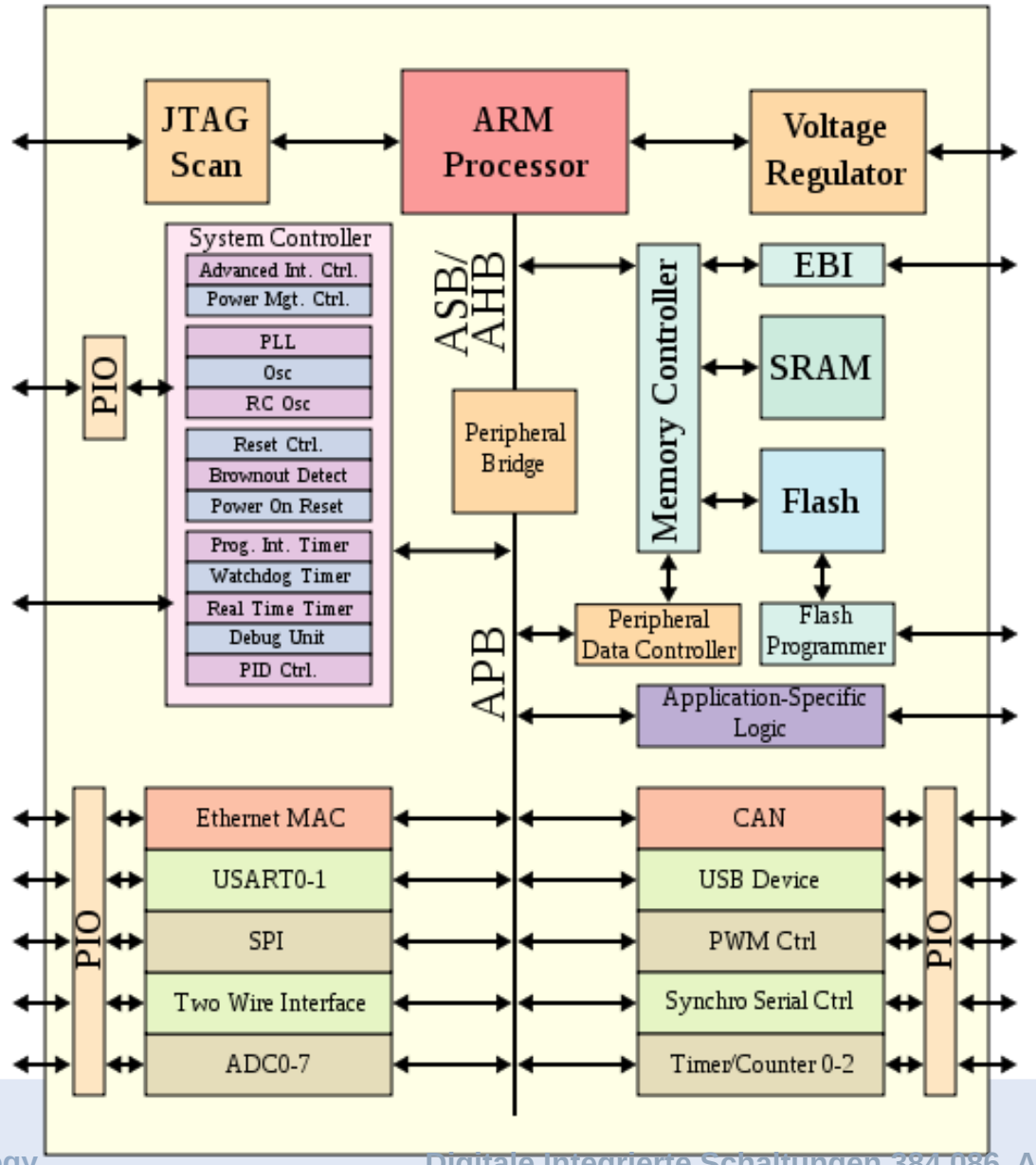
Logic Minimization Algorithms for VLSI Synthesis, Kluver Academic Publisher, Robert K. Brayton, Gary D. Hachtel, Curtis T. McMullen, and Alberto L. Sangiovanni-Vincentelli, 1984.

1990
1 000 000 Transistoren

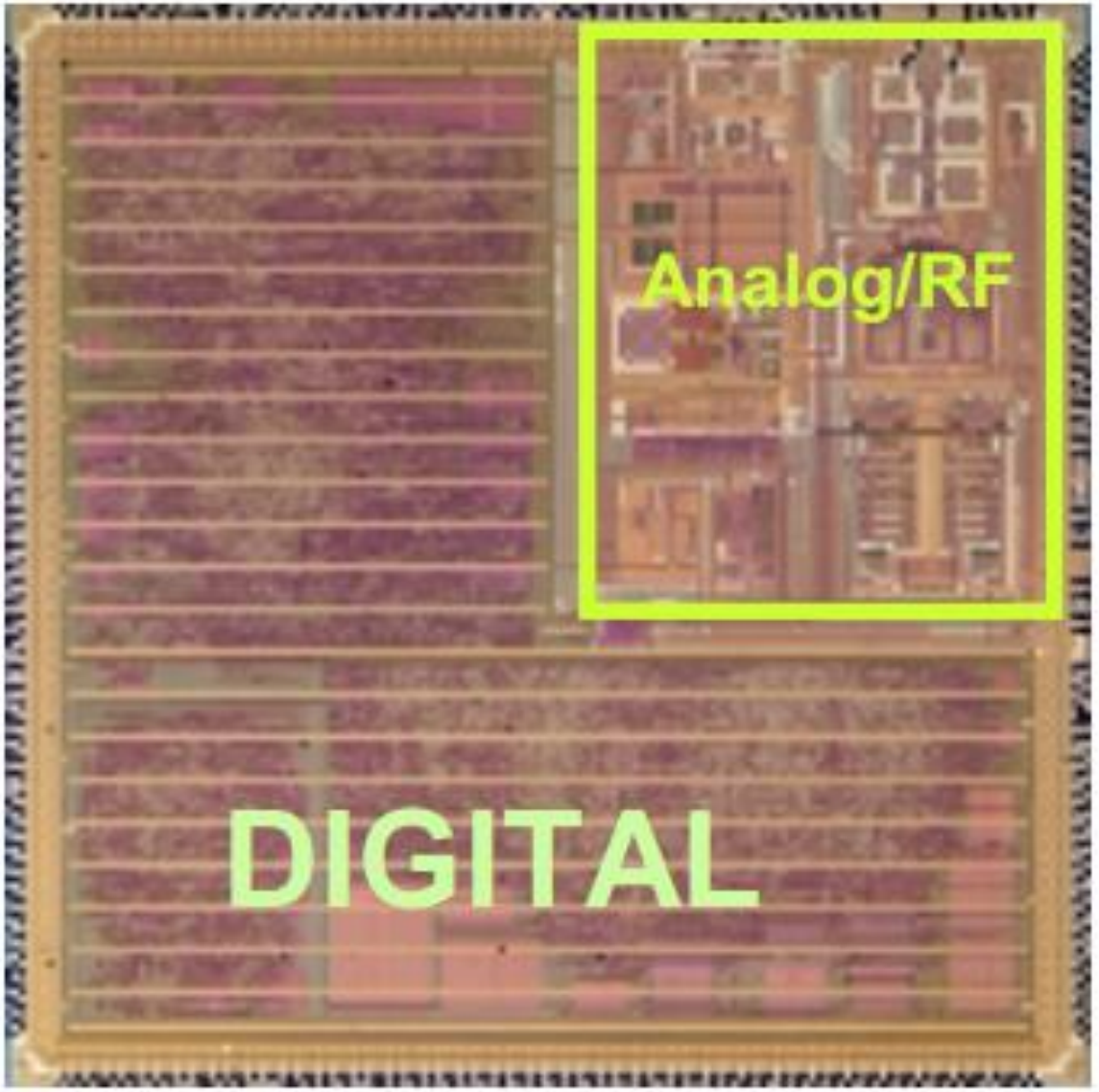
High Level Synthesis



Emergence of Systems on Chip



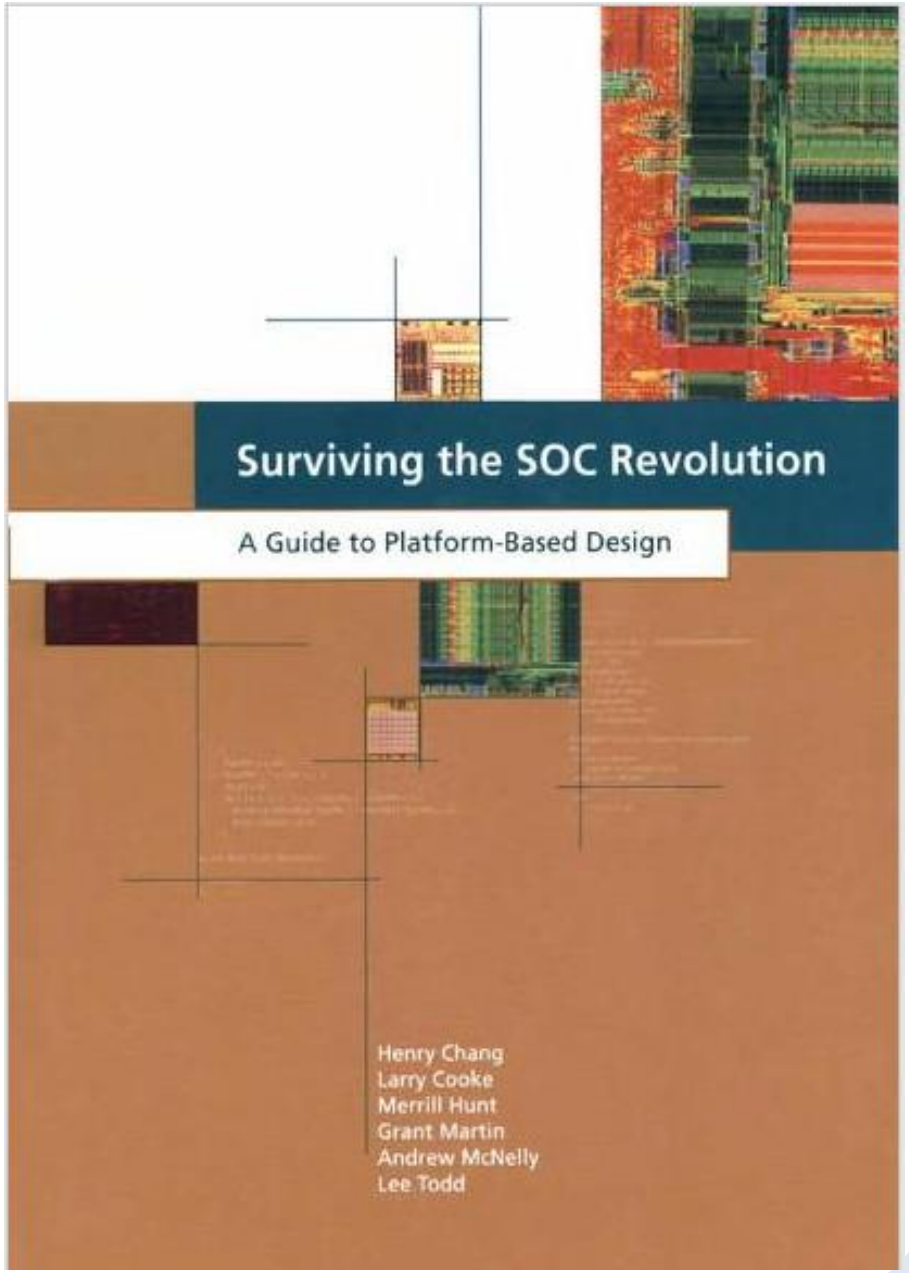
Emergence of Systems on Chip



2000

20 000 000 Transistoren

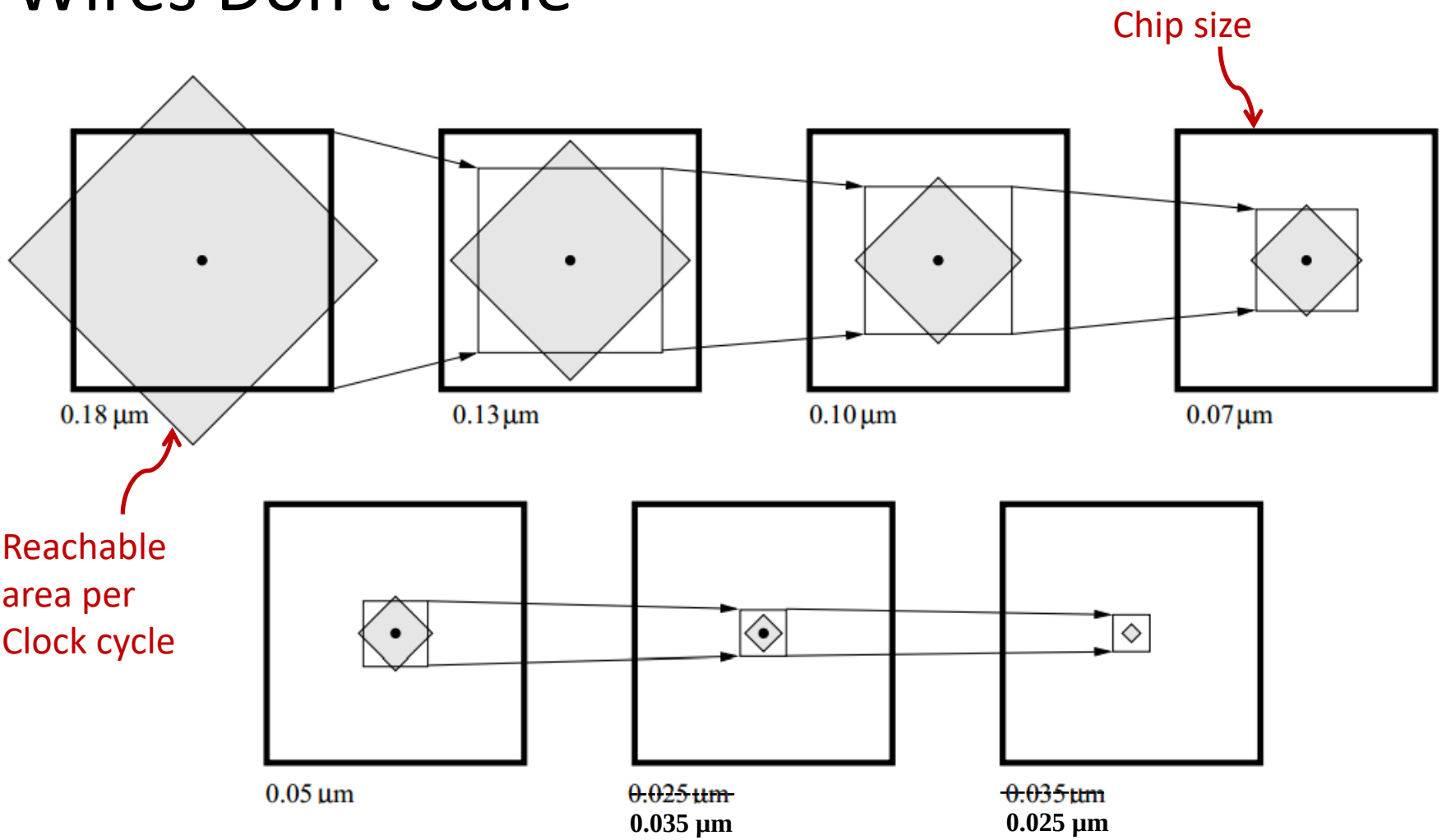
The SoC Revolution



Surviving the SOC Revolution - A Guide to Platform-Based Design, Kluwer Academic Publishers, Henry Chang, Larry Cooke, Merrill Hunt, Grant Martin, Andrew McNelly, and Lee Todd, 1999.

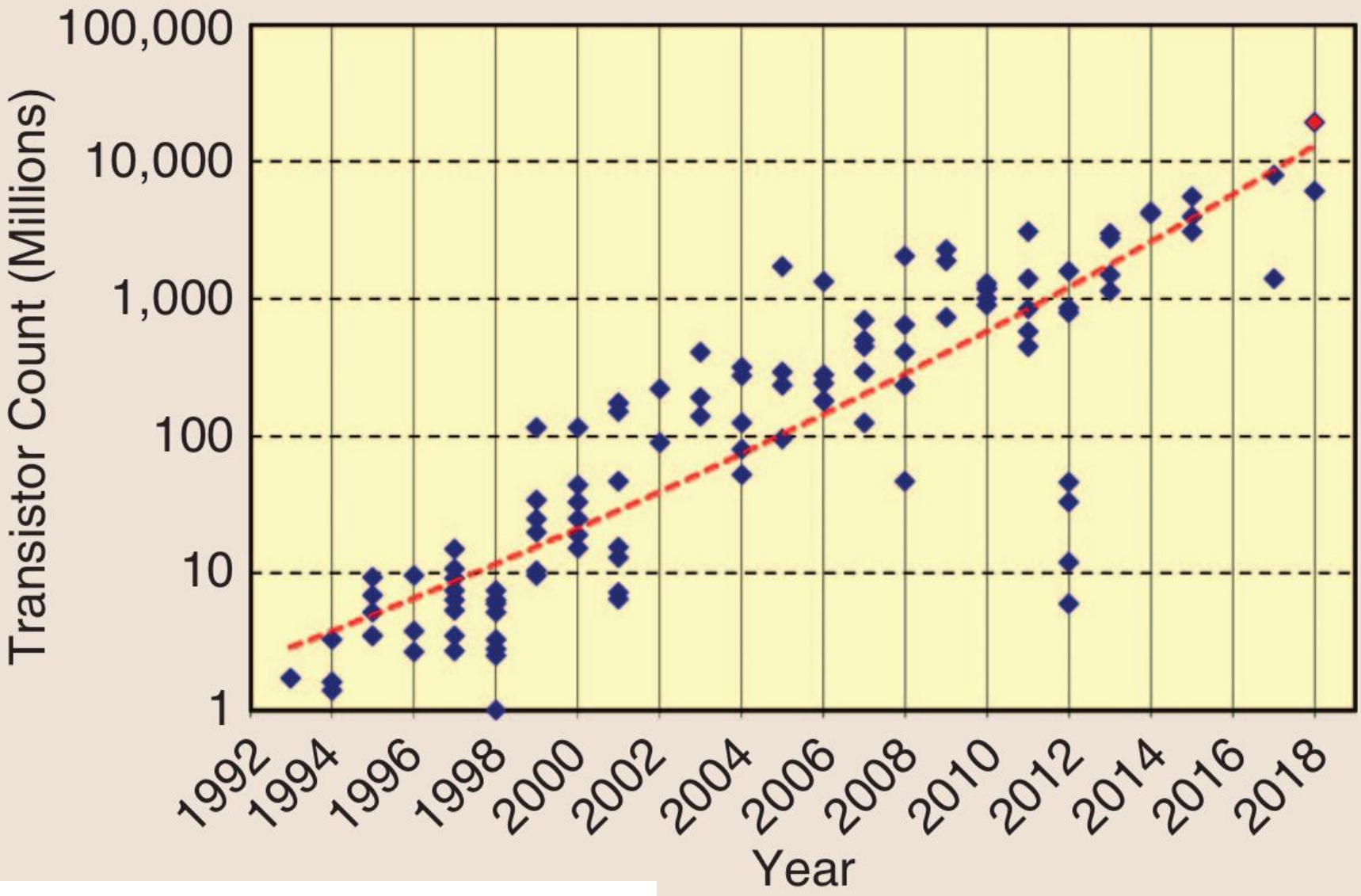
2000 – 20 000 000 Transistors

Wires Don‘t Scale



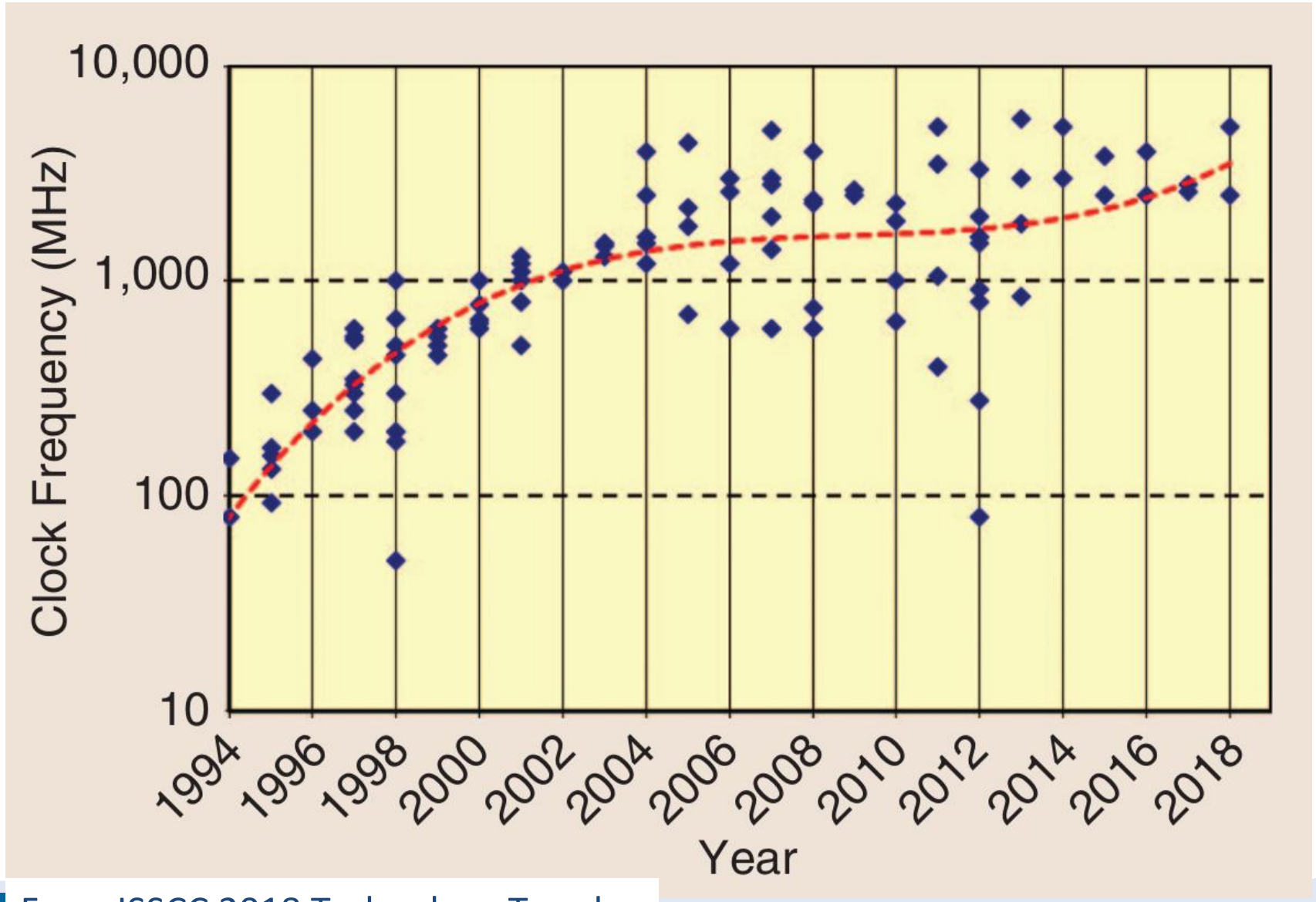
Ron Ho, *On-chip wires: Scaling and efficiency*, PhD thesis, Stanford University, 2003.

Capacity Trend



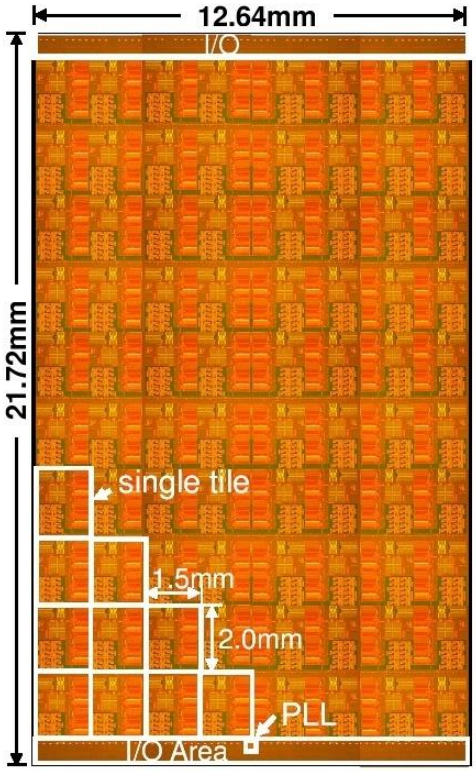
From ISSCC 2018 Technology Trends

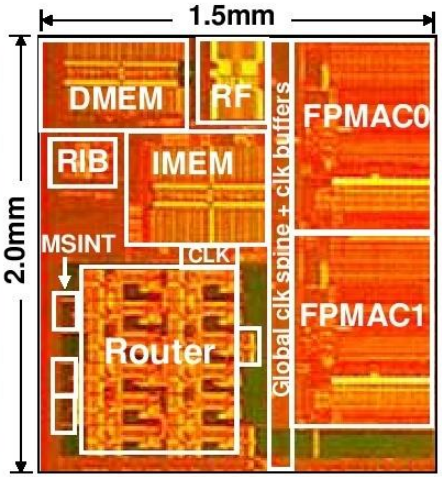
End of Frequency Scaling



From ISSCC 2018 Technology Trends

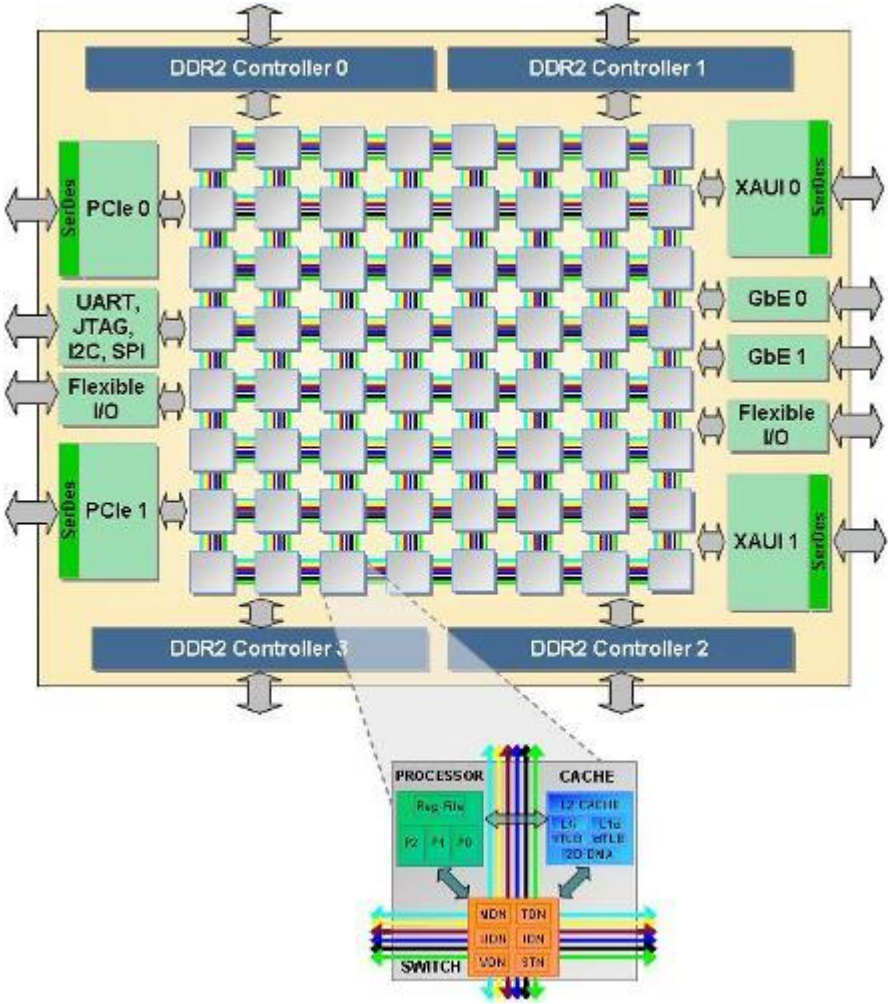
Emergence of Networks-on-Chip





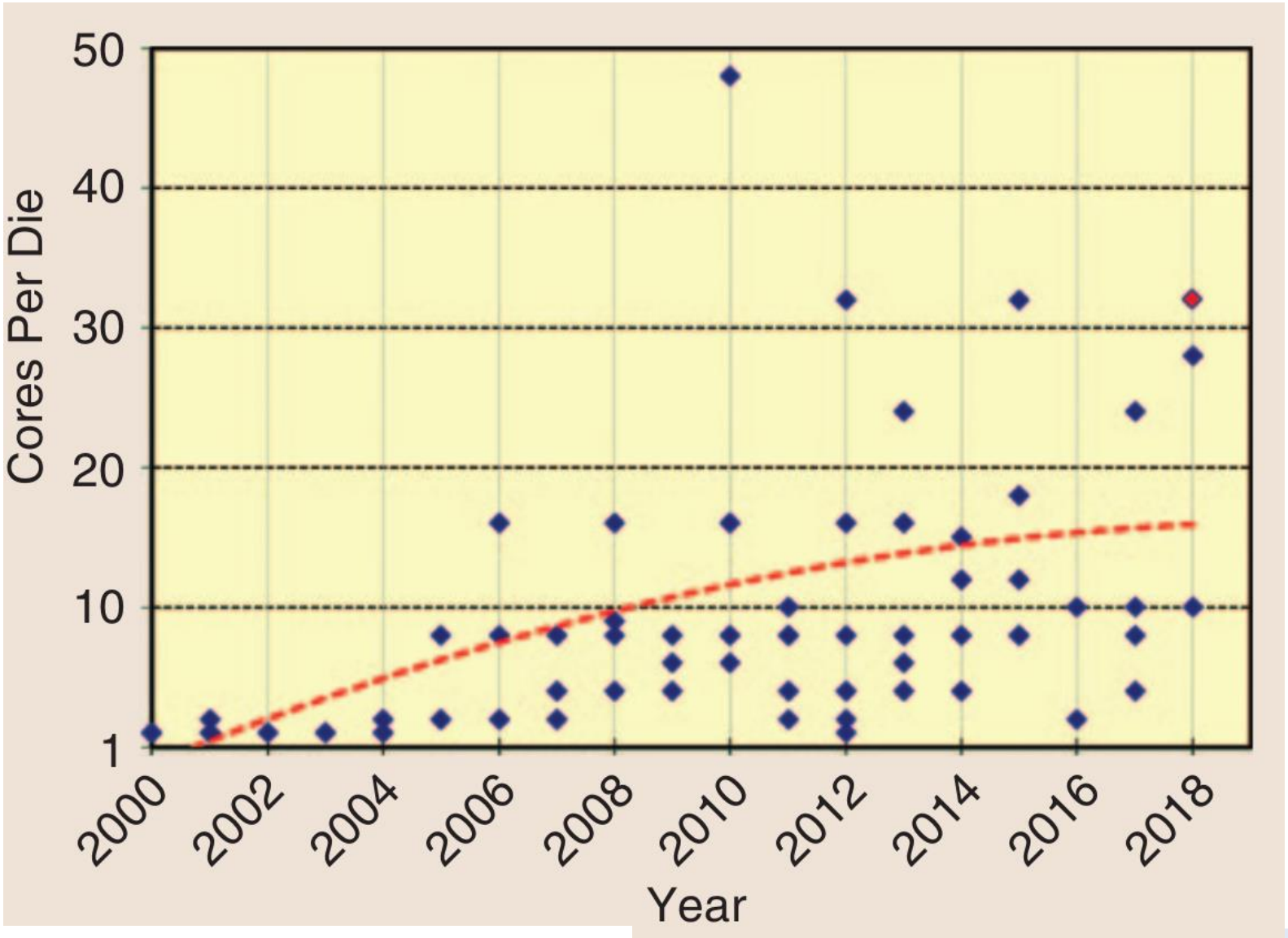
Technology	65nm CMOS Process
Interconnect	1 poly, 8 metal (Cu)
Transistors	100 Million
Die Area	275mm²
Tile area	3mm²
Package	1248 pin LGA, 14 layers, 343 signal pins

Intel Teraflop



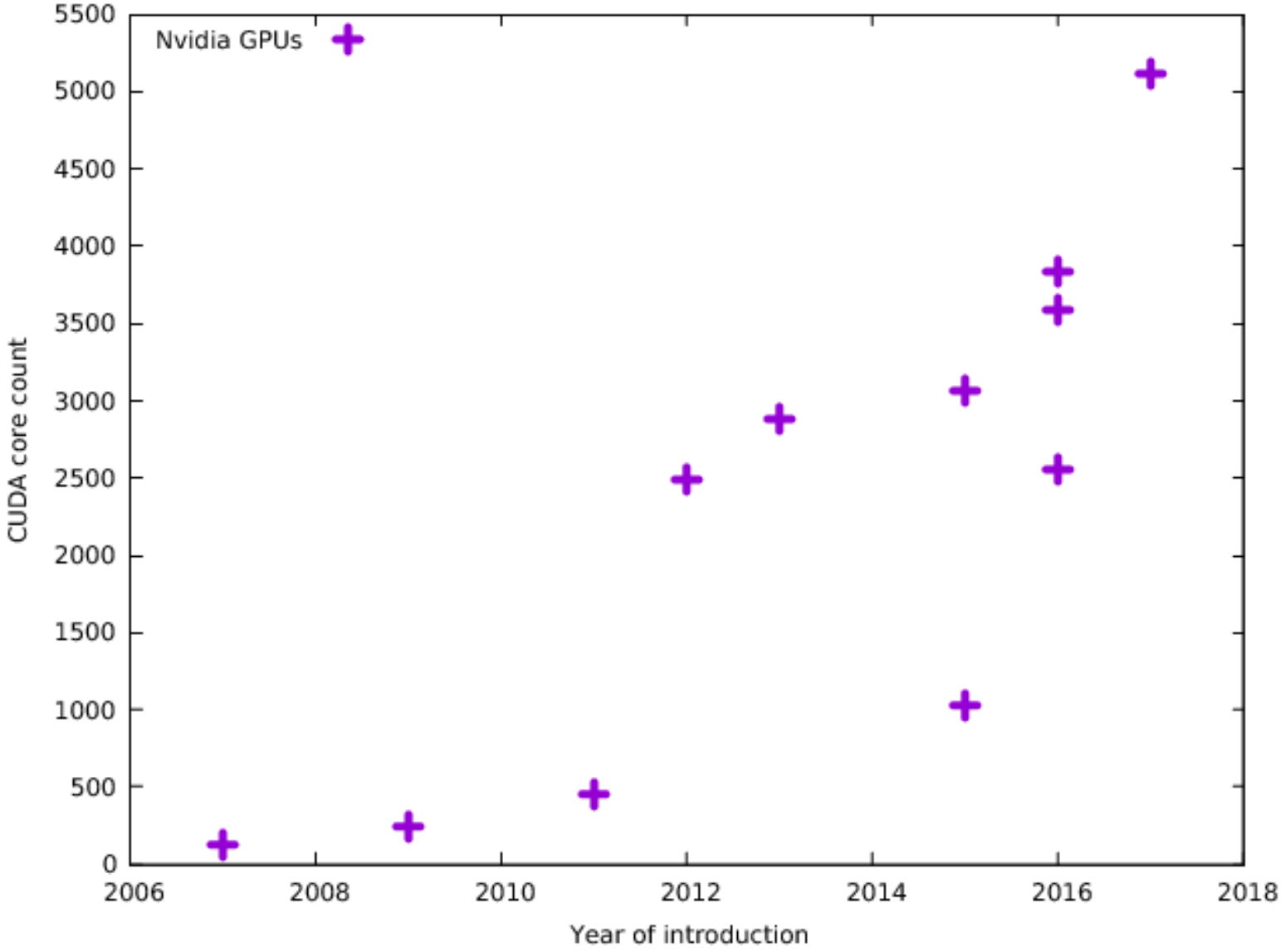
Tilera TilePro 64

The Multicore Revolution - Core Count



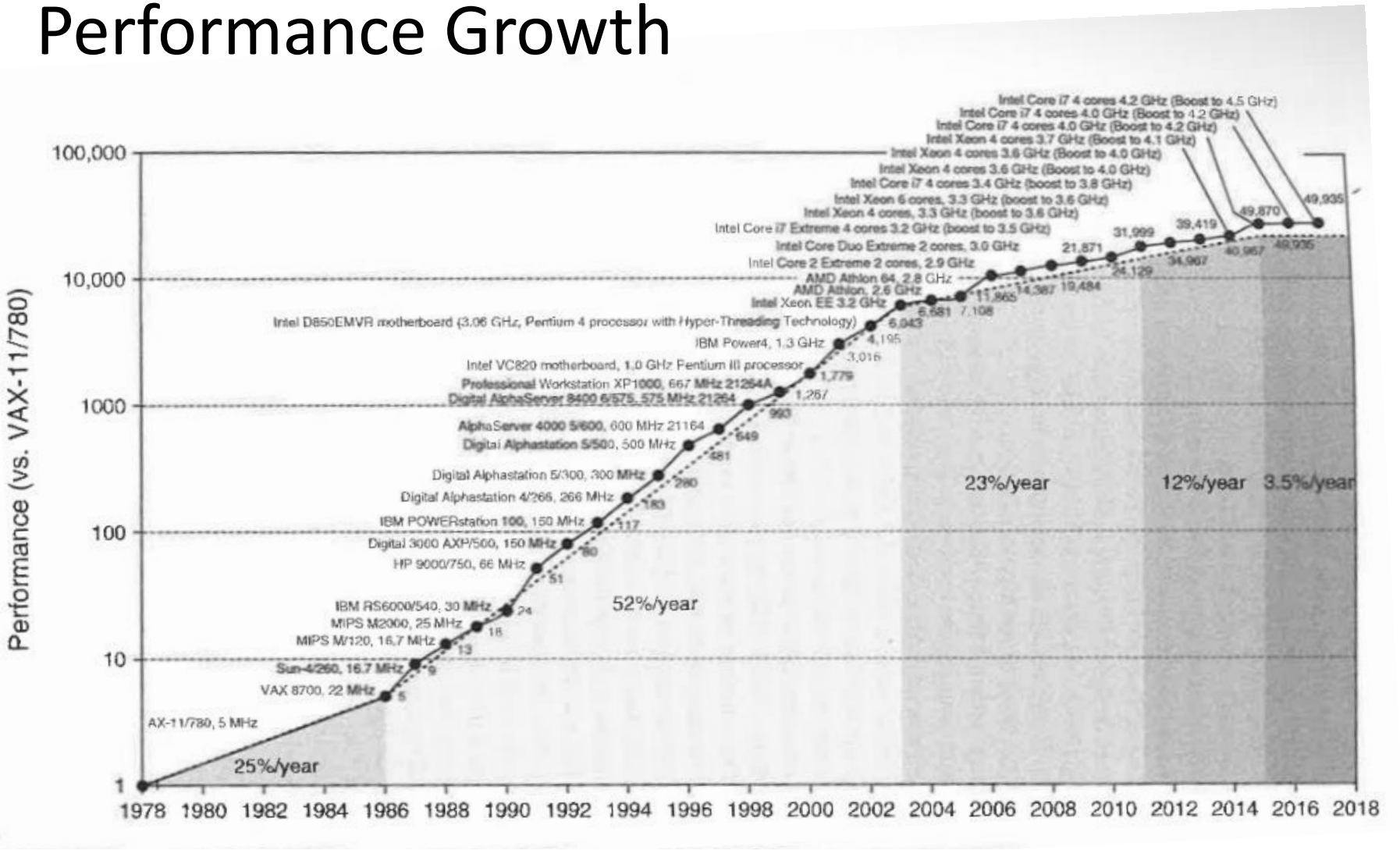
From ISSCC 2018 Technology Trends

Graphics Processors



The Multicore Revolution

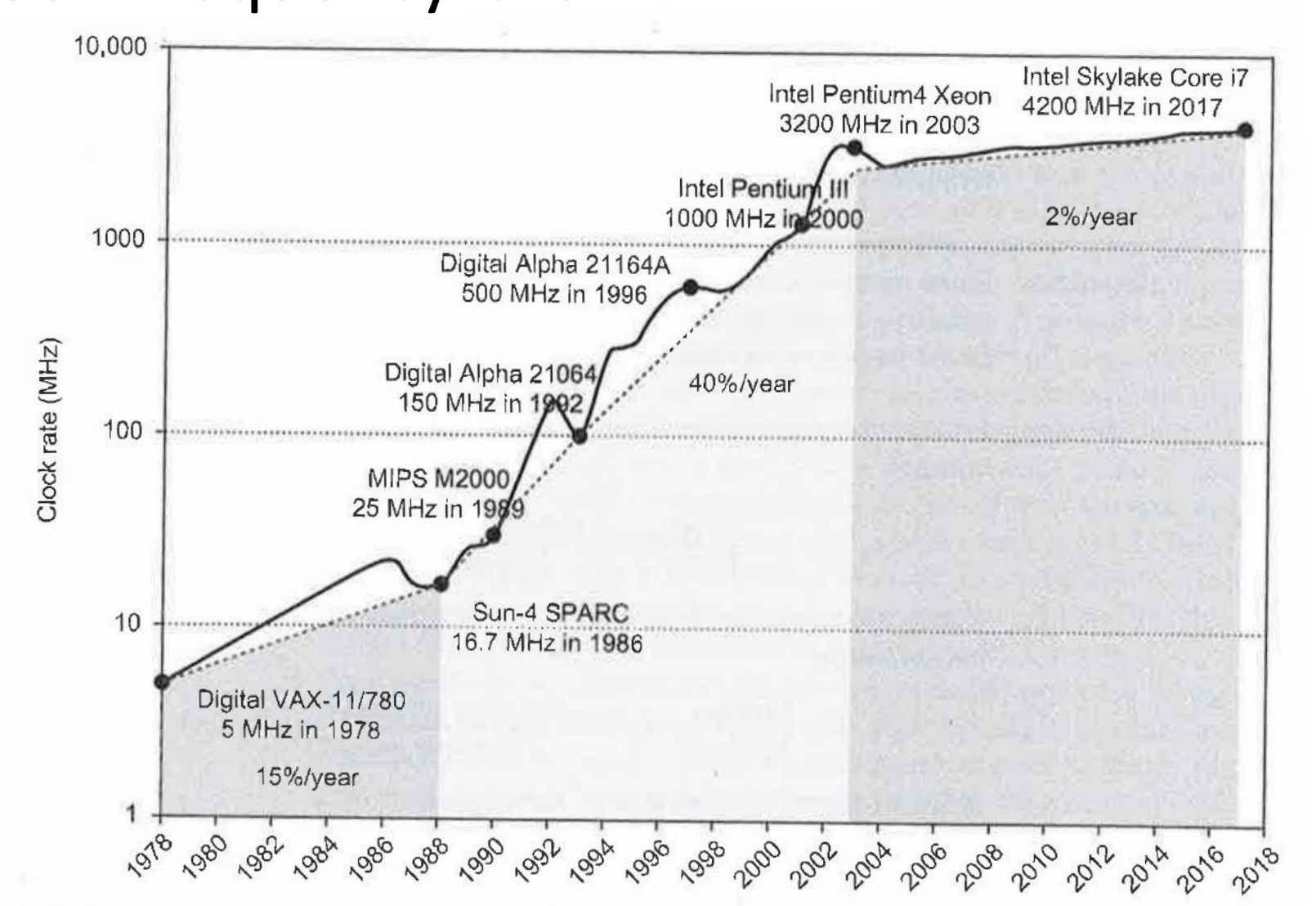
Performance Growth



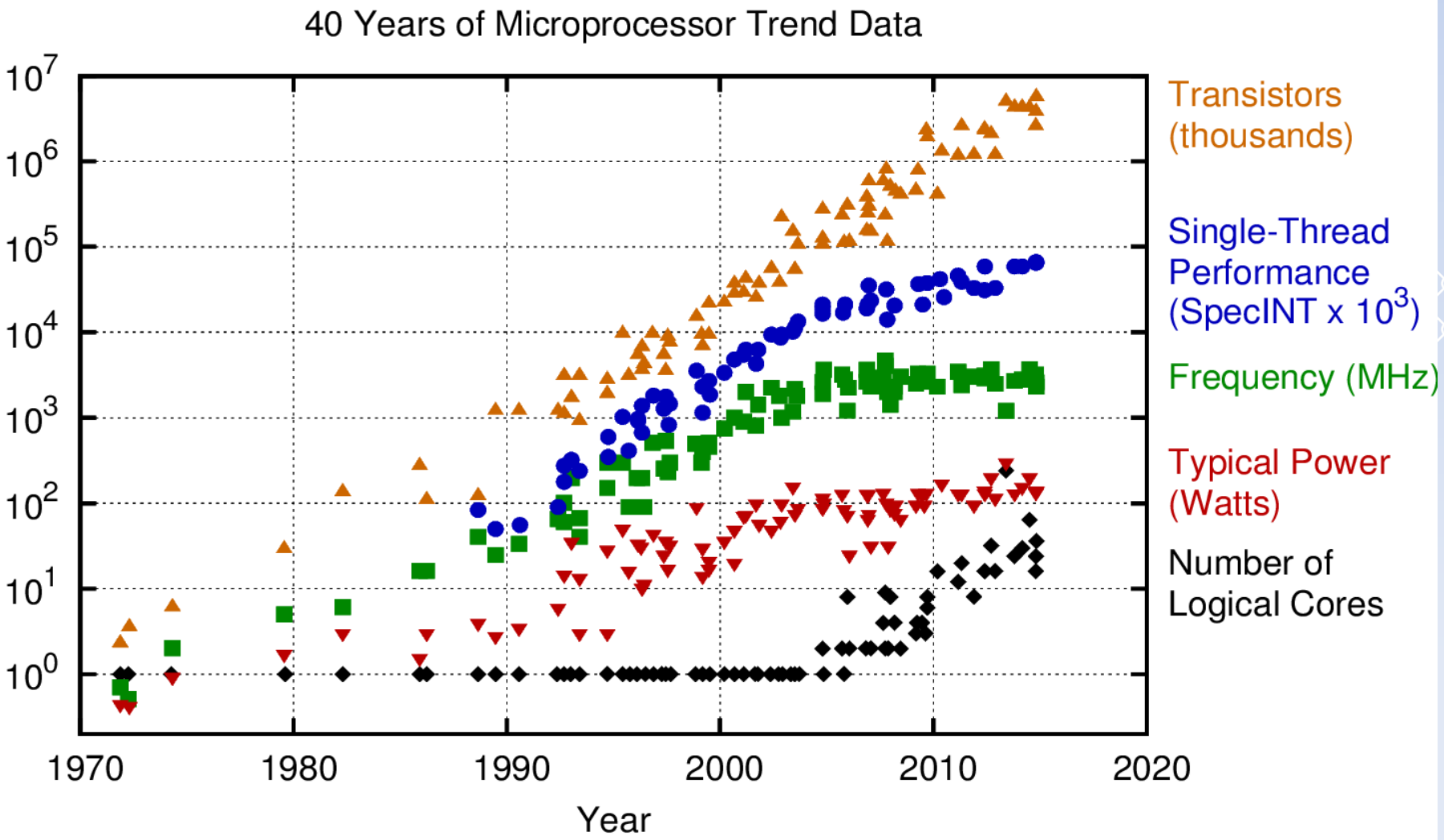
Performance as measured by the SPEC INT Benchmark

Computer Architecture, J. Hennessy and D. Patterson, 6th edition, 2018; Figure 1.1 page 3.

Clock Frequency Growth



Computer Architecture, J. Hennessy and D. Patterson, 6th edition, 2018; Figure 1.11 page 26.



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 2010-2015 by K. Rupp

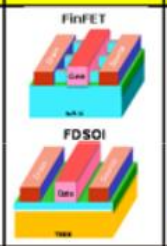
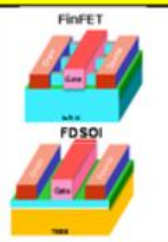
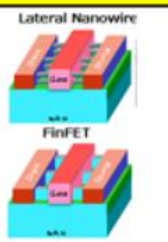
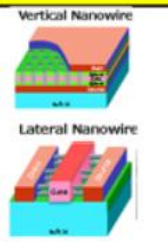
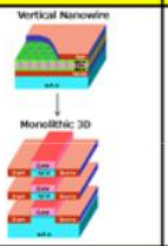
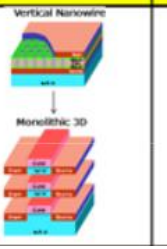
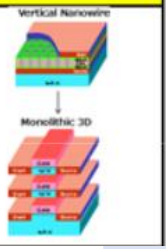
Technology Roadmap

Technology node (2-3 Years) scaling -2030:

- Performance: >30% more maximum operating frequency at constant energy
- Power: >50% less energy per switching at a given performance
- Area: >50% area reduction
- Cost: <30% wafer cost – 30-35% less die cost for scaled die.

International Roadmap for Devices and Systems
2016 Edition
More Moore White Paper
https://irds.ieee.org

Technology Roadmap

YEAR OF PRODUCTION	2015	2017	2019	2021	2024	2027	2030
Logic device technology naming	P70M56	P54M36	P42M24	P32M20	P24M12G1	P24M12G2	P24M12G3
Logic industry "Node Range" Labeling (nm)	"16/14"	"11/10"	"8/7"	"6/5"	"4/3"	"3/2.5"	"2/1.5"
Logic device structure options	finFET FDSOI	finFET FDSOI	finFET LGAA	finFET LGAA VGAA	VGAA, M3D	VGAA, M3D	VGAA, M3D
							

PxxMxx notation refers to Pxx: contacted poly pitch and Mxx: metalx pitch in nm.

Acronyms used:

FDSOI: Fully-Depleted Silicon-On-Insulator (FDSOI),

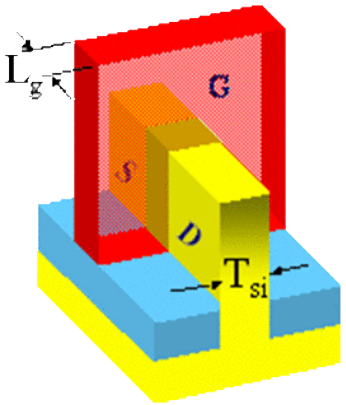
LGAA: Lateral Gate-All-Around-Device (GAA),

VGAA: Vertical GAA,

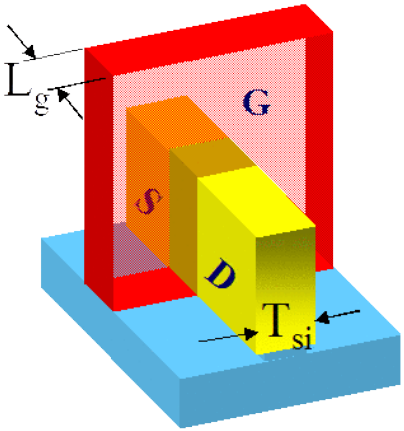
M3D: Monolithic-3D.

International Roadmap for Devices and Systems, 2016 Edition
More Moore White Paper, https://irds.ieee.org

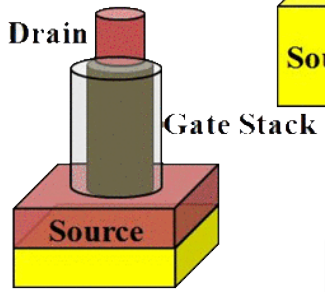
FinFET und Gate All Around



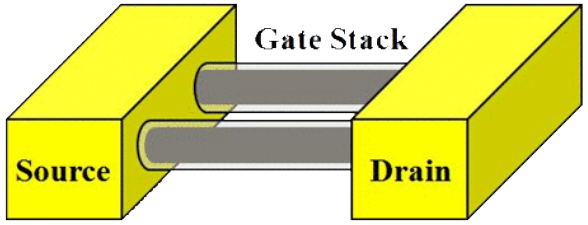
FinFET on Bulk



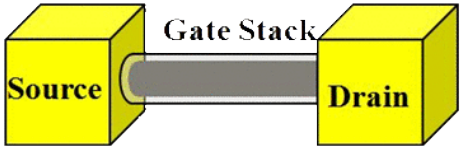
FinFET on SOI



Vertical CG FET



Twin Silicon Nanowire FET



Horizontal Nanowire FET

Technology Roadmap

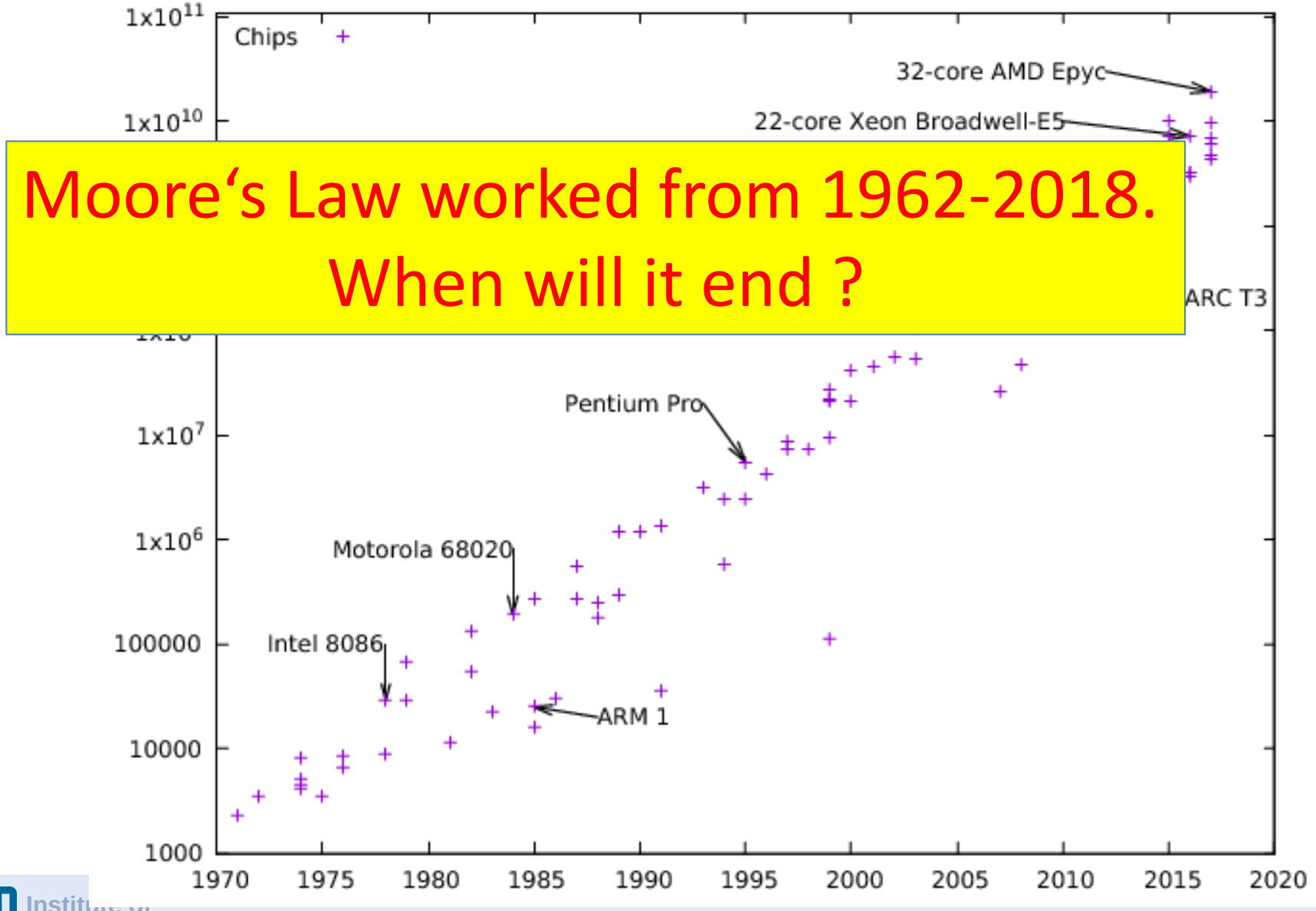
	2015	2017	2019	2021	2024	2027	2030
Industry labeling	„16/14“	„11/10“	„8/7“	„6/5“	„4/3“	„3/2.5“	„2/1.5“
Metal pitch (nm)	56	36	24	20	12	12	12
Contacted Poly pitch (nm)	70	48	42	32	24	24	24
L_g: HP Logic (nm)	24	18	14	10	10	10	10
L_g: LP Logc (nm)	26	20	16	12	12	12	12
Vdd (V)	0.8	0.75	0.7	0.65	0.55	0.45	0.4
V_t : HP (V)	0.129	0.129	0.133	0.136	0.084	0.052	0.052
V_t : LP (V)	0.351	0.336	0.333	0.326	0.201	0.125	0.125
Inversion layer thickness (nm)	1.1	1.0	0.9	0.85	0.8	0.8	0.8
Energy/switching (fJ)	3.47	2.52	1.89	1.24	0.94	0.63	0.50

International Roadmap for Devices and Systems, 2016 Edition
More Moore White Paper, https://irds.ieee.org

Technology Roadmap

	2015	2017	2019	2021	2024	2027	2030
Industry labeling	„16/14“	„11/10“	„8/7“	„6/5“	„4/3“	„3/2.5“	„2/1.5“
SRAM density (Mbit/mm2)	17	29	50	78	217	217	217
NAND2 density (Mgates/mm2)	9	19	33	78	87	87	87
FO3 delay(ps)	14.80	11.28	11.03	8.69	9.43	8.67	8.33
FO3 dynamic power @1GHz (µW)	3.49	1.93	1.07	0.56	0.3	0.2	0.16
FO3 leakage power (µW)	0.118	0.081	0.048	0.029	0.017	0.014	0.012
No of wiring layers	13	14	15	16	18	22	30

International Roadmap for Devices and Systems, 2016 Edition
More Moore White Paper, https://irds.ieee.org



In 2604

Universal Limits on Computation

Lawrence M. Krauss[1] and Glenn D. Starkman[1,2]
[1] *Center for Education and Research in Cosmology and Astrophysics, Department of Physics, and Department of Astronomy, Case Western Reserve University, Cleveland, OH 44106-7079*
[2] *Department of Physics, CERN, Theory Division, 1211 Geneva 23, Switzerland*

The physical limits to computation have been under active scrutiny over the past decade or two, as theoretical investigations of the possible impact of quantum mechanical processes on computing have begun to make contact with realizable experimental configurations. We demonstrate here that the observed acceleration of the Universe can produce a universal limit on the total amount of information that can be stored and processed in the future, putting an ultimate limit on future technology for any civilization, including a time-limit on Moore's Law. The limits we derive are stringent, and include the possibilities that the computing performed is either distributed or local. A careful consideration of the effect of horizons on information processing is necessary for this analysis, which suggests that the total amount of information that can be processed by any observer is significantly less than the Hawking-Bekenstein entropy associated with the existence of an event horizon in an accelerating universe.

y 2004