

EE241B : Advanced Digital Circuits

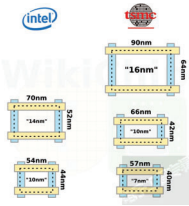
Lecture 2 – Scaling Trends

Borivoje Nikolić



Apple, Huawei Use TSMC, But Their 7nm SoCs Are Different. When talking about the most advanced semiconductor manufacturing processes, it seems that most of the SoCs in 2019 can be collectively classified as 7nm. But not all 7nm is equal.

EE Times, January 22, 2020.



Announcements

- Sign up for Piazza if you haven't already

Assigned Reading

- R.H. Dennard et al, "Design of ion-implanted MOSFET's with very small physical dimensions" IEEE Journal of Solid-State Circuits, April 1974.
 - Just the scaling principles
- C.G. Sodini, P.-K. Ko, J.L. Moll, "The effect of high fields on MOS device and circuit performance," IEEE Trans. on Electron Devices, vol. 31, no. 10, pp. 1386 - 1393, Oct. 1984.
- K.-Y. Toh, P.-K. Ko, R.G. Meyer, "An engineering model for short-channel MOS devices" IEEE Journal of Solid-State Circuits, vol. 23, no. 4, pp. 950-958, Aug. 1988.
- T. Sakurai, A.R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," IEEE Journal of Solid-State Circuits, vol. 25, no. 2, pp. 584 - 594, April 1990.

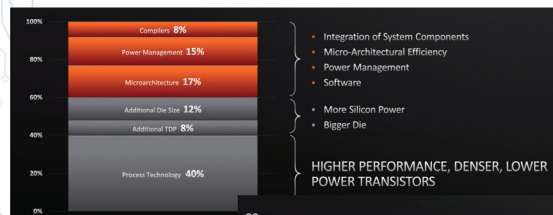
Outline

- Scaling issues
- Technology scaling trends
- Features of modern technologies
 - Lithography
 - Process technologies

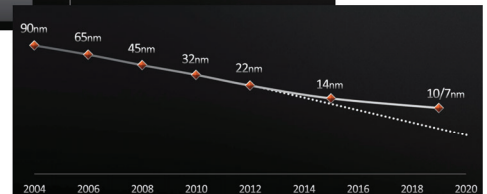
Trends and Challenges in Digital Integrated Circuit Design



Putting Scaling in Perspective

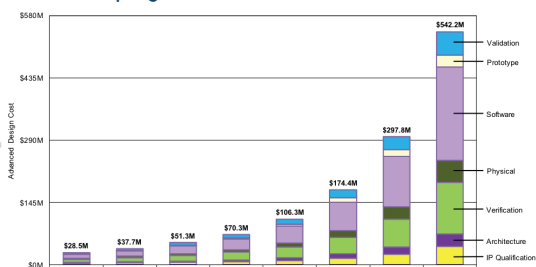


Performance gains over the past decade



Lisa Su, HotChips'19 keynote

Cost Of Developing New Products



- These are non-recurring (NRE) costs, need to be amortized over the lifetime of a product
- We will attempt to dismantle this...

Major Roadblocks

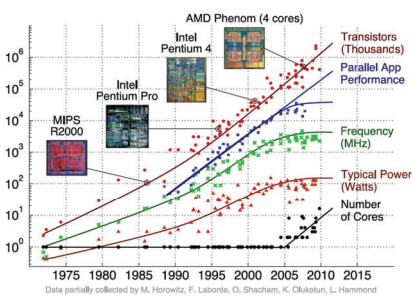
1. Managing complexity
How to design a 10 billion (100 billion) transistor chip?
And what to use all these transistors for?
2. Cost of integrated circuits is increasing
It takes >>\$10M to design a chip
Mask costs are many \$M in 16nm technology
3. Power as a limiting factor
End of frequency scaling
Dealing with power, leakages
4. Robustness issues
Variations, SRAM, memory, soft errors, signal integrity
5. The interconnect problem

Part 1: Technology

1.A Scaling Trends

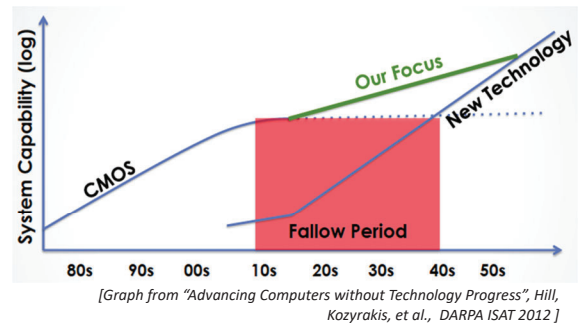


Power and Performance Trends



- What do we do next?

What Should We Do?

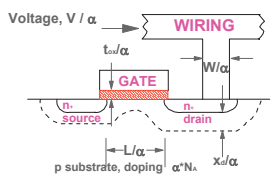


1.B Scaling Issues

Key Points

- Technology scaling (Moore's law) is slowing down
 - But logic and memory are continuing to scale, possibly at different pace
 - We anyway can't power up all transistors we can put on a chip
- Dennard scaling has ended >10 years ago
 - We cover it for understanding of current issues
- There are many technology flavors available at the moment
 - We need to know what each one brings to us, so we can choose the right one for your project
(may have to wait until the end of the class to figure out all options)

CMOS Scaling Rules (Dennard)



SCALING:

Voltage: V/α

Oxide: t_{ox}/α

Wire width: W/α

Gate length: L/α

Diffusion: x_d/α

Substrate: $\alpha \cdot N_A$

RESULTS:

Higher Density: $\sim \alpha^2$

Higher Speed: $\sim \alpha$

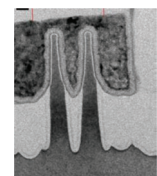
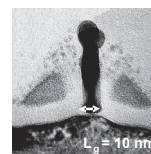
Power/ckt: $\sim 1/\alpha^2$

Power Density: ~~Constant~~

R. H. Dennard et al.,
IEEE J. Solid State Circuits, (1974).

Research vs. Production Devices

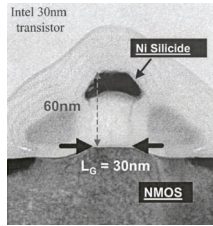
10nm device (Intel), circa 2003



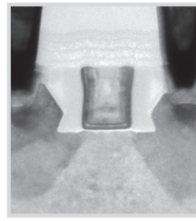
10nm node (Intel), IEDM'2017

CMOS Scaling

- Two 30nm transistors (research and production)



~2000 (research)

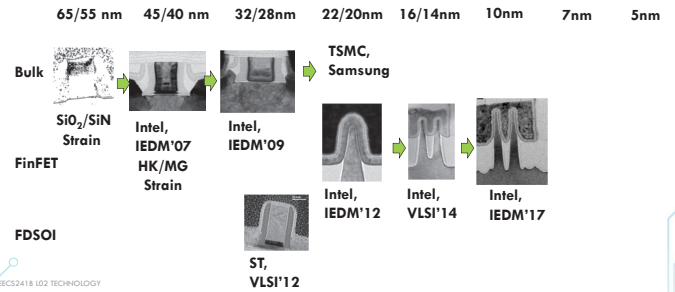


~2010 (production)

EECS32418: U2 TECHNOLOGY

Transistors are Changing

- From bulk to finFET and FDSOI

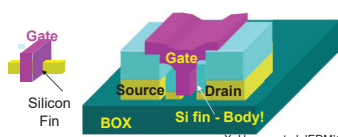


17

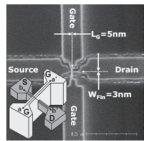
EECS32418: U2 TECHNOLOGY

18

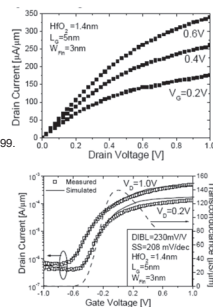
Sub-5nm FinFET



X. Huang, et al, IEDM'1999.



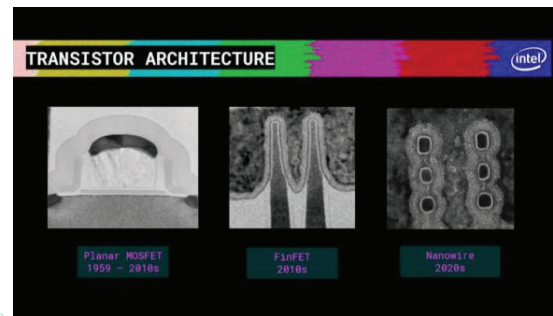
Lee, VLSI Technology, 2006



EECS32418: U2 TECHNOLOGY

Beyond 5nm

- Gate-all-around transistors/nanowires



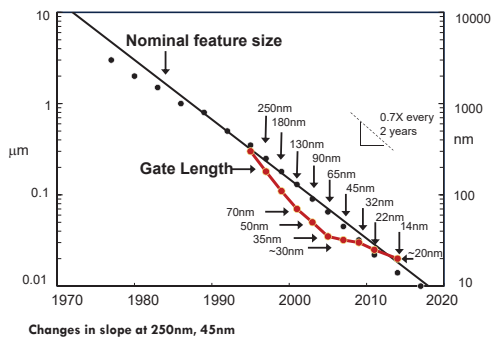
J. Keller, EECS Colloquium, UC Berkeley, Sept 2019.

19

EECS32418: U2 TECHNOLOGY

20

Physical Gate Scaling

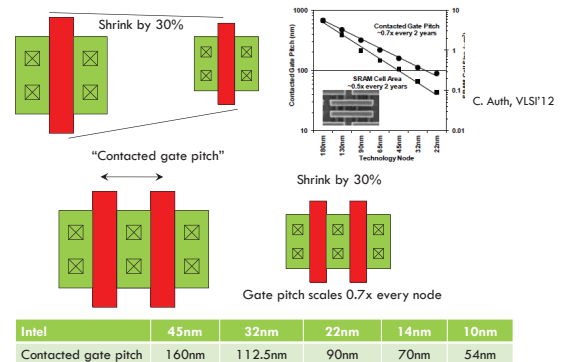


Changes in slope at 250nm, 45nm

Source: Intel, IEDM presentations

EECS32418: U2 TECHNOLOGY

Transistor Scaling



C. Auth, VLSI'12

21

EECS32418: U2 TECHNOLOGY

22

Pitch Scaling

Intel	45nm	32nm	22nm	14nm	10nm
Contacted gate pitch	160nm	112.5nm	90nm	70nm	54nm
Shrink		0.7	0.8	0.78	0.77

- Clearly not 0.7 anymore...

- But (Intel 14nm, Natarajan, IEDM'14)

- Fin pitch: 42nm (0.7x shrink)
- Metal 0: 56nm (-)
- Metal 1: 70nm (0.78x)
- Metal 2: 52nm (0.65x)

- Intel's metric:

- CPP x MXP
- 0.78 x 0.65 = 0.51

- CPP & FP matter more

EECS32418: U2 TECHNOLOGY

Various Technology flavors

- Intel 14nm

Options	High Speed Logic (HPL)	High Speed Logic (HPL)	High Speed Logic (HPL)	High Speed Logic (HPL)	High Speed Logic (HPL)
Vdd (V)	0.7	0.7	0.7	1.5-3.3	1.5-3.3
Gate Pitch (nm)	70	70	84	Min. 140	Min. 140
Fin Pitch (nm)	42	42	42	42	42
NMOS/PMOS Idsat (mA/μm)	1.3/1.2 @ 0.7V, 100ns	85/72 @ 0.7V, 100ns	50/32 @ 0.7V, 100ns	115/111 @ 1.8V, 10pA	115/111 @ 1.8V, 10pA

- Different foundries

Feature	Samsung 14 nm	Intel 14 nm	TSMC 16 nm
Fin pitch (nm)	48	42	48
1/3 fin pitch	16	14	16
Gate length (nm)	~30	~24	~33
Contacted gate pitch (nm)	78	70	90
Minimum metal pitch (nm)	64	52	64
6T SRAM cell area (μm²)	0.08	0.059	0.074

Source: Tech Insights EETimes

23

EECS32418: U2 TECHNOLOGY

24

Not All Technologies are Equal

Intel

Node	CPP	MxP	FP
65nm	230	230	
45nm	160	160	
32nm	112.5	112.5	
22nm	90	80	60
14nm	70	52	42
10nm	54	36	34
7nm	37	32	

Samsung

Node	CPP	MxP	FP
45nm	180	140	
32nm	130	100	
28nm	115	90	
20LPE	90	80	60
14LPE	78	64	48
10LPE	68	48	42
7LPP	54	36	27

TSMC

Node	CPP	MxP	FP
45nm	190	140	
40nm	170	130	
28nm	120	90	
20SoC	90	64	
16FF	90	64	48
16FFC	96	64	48
10FF	66	44	36
7FF	57	40	30
5FF	50	28	

- **CPP = Contacted poly pitch**
- **MxP = Minimum metal pitch**
- **FP = Fin pitch**

Source:
A. Wei, Technlights
IEDM'17, IEDM'19, WikiChip, SemiWiki'20

BEC32418: U22 TECHNOLOGY

25

Source:
A. Wei, TechInsights
IEDM'17, IEDM'19, WikiChip, SemiWiki'20

- EECS241B: L02: TECHNOLOGY 25

ASAP²

- Predictive technology kit used in this class
- None of the above processes, but close

Layer	Lithography	Width/drawn (nm)	Pitch (nm)
Fin	SAQP ^a	6.5/7	27
Active (horizontal)	EUV	54/16	108
Gate	SADP	21/20	54
SDT/LUSD	EUV	25/24	54 ^b
LIG	EUV	16/16	54
VIA0-VIA3	EUV	18/18	25 ^a
M1-M3	EUV	18/18	36
M4 and M5	SADP	24/24	48
VIA4 and VIA5	LELE	24/24	34 ^a
M6 and M7	SADP	32/32	64
VIA6 and VIA7	LELE	32/32	45 ^a
M8 and M9	SE	40/40	80
VIA8	SE	40/40	57 ^a

^a Corner to corner spacing as drawn.
^b Horizontal only.

BSC2418 L02 TECHNOLOGY

26

-
- The diagram illustrates the device structure, showing a cross-section of a substrate with multiple layers. From top to bottom, the layers are labeled: LISO, PMD3, SDT, PMD2, PMD1, and PMD0. Below the PMD0 layer, there are three diamond-shaped regions labeled 'fin' and 'SD'. The bottom layer is a substrate.

^a Corner to corner spacing as drawn.
^b Horizontal only.

Ideal vs. Real Scaling

- Leakage slows down V_{Th} , V_{DD} scaling

The plot shows the scaling of various semiconductor parameters versus technology node size (L_g in nm). The y-axis is logarithmic, ranging from 0.1 to 10000. The x-axis is logarithmic, ranging from 10 to 1000 nm. The parameters and their ideal scaling lines are:

- I_{DSAT} [$\mu A/\mu m$]: Black circles, ideal line is horizontal at 1000.
- V_{DD} : Black triangles, ideal line is a straight line with a slope of 1.
- T_{inv} [ps]: Black squares, ideal line is a straight line with a slope of 1.
- V_{th} [V]: Black circles, ideal line is a straight line with a slope of 1.
- I_{Dsat} : Black circles, ideal line is a straight line with a slope of 1.

The plot illustrates how leakage current (I_{Dsat}) and threshold voltage (V_{th}) scaling are affected by technology node size, leading to a slowdown in V_{Th} and V_{DD} scaling.

BEC52418 UO2 TECHNOLOGY

27

-

Technology Flavors

- LP keeps drain leakage constant

The graph plots Normalized current (Y-axis) against Technology [nm] (X-axis, ranging from 180 to 40 nm). It shows several current components: $I_{DNL,LP}$ (a constant horizontal line at the top), $I_{DNL,DP}$ (a line decreasing as technology scales down), I_{DNL} (a dashed line decreasing as technology scales down), $I_{DNL,LP}$ (a line increasing as technology scales down), and I_{DNL} (a dashed line increasing as technology scales down). The LP technology is highlighted by a shaded region between the constant $I_{DNL,LP}$ line and the decreasing $I_{DNL,DP}$ line, indicating that drain leakage is kept constant across technology nodes.

EECS241B (02) TECHNOLOGY

28

-

32nm Technology Flavors (Intel)

Logic Transistor

(HP or SP)

Low Power Transistor

(LP)

HV I/O Transistor

(1.8 V or 3.3 V)

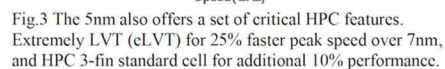
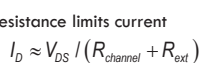
Transistor Type	Logic (option for HP or SP)		Low Power	HV I/O (option for 1.8 or 3.3 V)	
	HP	SP	LP	1.8V	3.3V
EOT(nm)	0.95	0.95	0.95	~ 4	~ 7
Vdd (V)	.75/ 1	.75/ 1	0.75/1.2	1.5/ 1.8	1.5/ 3.3
Pitch(nm)	112.5	112.5	126	min. 338	min. 675
Lgate (nm)	30	34	46	>140	>320
NMOS Idsat (mA/μm)	1.53 @ 1V	1.12 @ 1V	0.71 @ 1V	0.68 1.8 V	0.7 3.3 V
PMOS Idsat (mA/μm)	1.23 @ 1V	0.87 @ 1V	0.55 @ 1V	0.59 1.8 V	.34 @3.3 V
Ioff (nA/μm)	100	1	0.03	0.1	<0.01

C.-H. Jan, IEDM'09, P. VanDerVoorn, VLSI Tech'10

C.-H. Jan, IEDM'09, P. VanDerVoorn, VLSI Tech'10

The graph plots Stand-by Power Pdb (uW) against Speed (GHz). It shows four main data series: SVT (blue squares), LVT (blue circles), N5 (red circles), and uLVT (black circles). The uLVT series is the fastest, followed by N5, LVT, and SVT. Annotations indicate that uLVT is 25% faster than N5 and 10% faster than N5 HPC.

Feature	Approx. Speed (GHz)	Approx. Stand-by Power Pdb (uW)
SVT	0.8	0.5
LVT	1.2	1.0
N5	1.6	1.5
uLVT	2.0	2.0

[illegible]

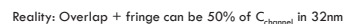
Parasitic Capacitance Scaling

The graph plots Capacitance (F/μm) on the y-axis (log scale from 1 to 8) against Technology node (nm) on the x-axis (linear scale from 120 to 0). A blue curve labeled 'Intrinsic channel capacitance' decreases as technology node decreases. An orange curve labeled 'Total parasitic capacitance' increases as technology node decreases. The two curves intersect at approximately 32 nm, where a dashed circle highlights the region and a label 'Parasitics dominate!' points to the orange curve. An inset diagram shows a cross-section of a transistor with labels for '50 Control', 'Parasitic capacitance', '50 Control', 'gate', 'drain', and 'source'.

Technology node (nm)	Intrinsic channel capacitance (F/μm)	Total parasitic capacitance (F/μm)
120	~6.5	~1.5
90	~4.5	~1.8
60	~3.0	~2.2
32	~2.0	~2.0
20	~1.5	~3.0

Reality: Overlap + fringe can be 50% of C_{channel} in 32nm

S. Thompson, *Materials Today*, 2006.



S. Thompson, *Materials Today*, 2006.