6.11.3 The International Technology Roadmaps for Semiconductors (ITRS)

Although the transistor scaling theories presented in the last two sections do present the framework for device miniaturization, in recent years, much more detailed roadmaps, called the International Technology Roadmap for Semiconductors (ITRS) [6.23], have been compiled by a technology groups whose members are drawn from a number of international semiconductor companies and organizations. The ITRS presents the way forward, seen in any particular year, and highlights the expected achievements as well as possible hurdles.

Each year's roadmap is a detailed document, typically extending 15 years into the future. An example of some of the entries from the 2005 ITRS roadmap for high-performance logic technology are given in Table 6.3 [6.24]. It can be seen from the roadmap that the physical gate length is expected to continue scaling down quite significantly. On the other hand, as expected, supply voltage scales down only slightly. Gate oxide thickness will scale down initially, finally saturating at about 5 nm, below which further scaling is going to be difficult. This implies that although lateral fields will continue to increase, vertical fields (which determine oxide reliability) will saturate at about 14 MV/cm. Of course, it should be emphasized that as the years progress, different technologies will be used; for example, Table 6.3 includes transitions from bulk CMOS to FD SOI CMOS to dual-gate (DG) CMOS, and also from SiO_2 to high-k dielectrics around 2008. The roadmap predicts an approximately 17 % yearly improvement in intrinsic switching speed every year (not that the "intrinsic" device speed is much higher than the actual speed of, say, a microprocessor chip made using this technology. The ITRS has similar predictions for other technologies, such as low stand-by power technology, low operating power technology, DRAM technology and NV memory technology.

Scanned by CamScanner

Table 6.3

Year of Production	2005	2008	2011	2014	2017	2020
MPU Physical Gate Length (nm)	32	22	16	11	8	5
Equivalent Oxide Thickness (Å)	12	9	6	5	5	5
Supply Voltage (V)	1.1	1	1	0.9	0.7	0.7
Threshold Voltage (mV)	195	169	170	190	200	208
Off Leakage Current (μA/μm)	0.06	0.17	0.22	0.11	0.11	0.11
On Drive Current (μA/μm)	1020	1570	2015	2290	2533	2981
Total Gate Capacitance (aF/μm)	0.81	0.80	0.69	0.56	0.49	0.36
Intrinsic Gate Delay (ps)	0.87	0.54	0.34	0.21	0.13	0.08
Intrinsic Switching Speed (GHz)	1149	1852	2941	4762	7692	12500

Scanned by CamScanner

2/2/2020

Observations from Table 6.3

1. Physical gate length are <u>shrinking</u> (32nm in 2005 to 8 nm in 2017)

   eg. 1. 32nm in 2005 to 22nm in 2008

   Shrinking factor is $22nm \times (\underbrace{1.455}) = 32nm$

   2. 22nm in 2008 to 16nm in 2011

   Shrinking factor is $16nm \times (\underbrace{1.375}) = 22nm$

2. Oxide thickness (t_{ox}) are also <u>shrinking</u> (12°A in 2005 to 5°A in 2017)

3. Supply voltage (V) are also <u>shrinking</u> (1.1V in 2005 to 0.7 in 2017)

4. Intrinsic switching speed (GHz) is increasing (1149 GHz in 2005 to 7692 GHz in 2017)

So, clearly some parameter's of MOS transistor are shrinking while some are ↑sing by same <u>factor</u>.

This is explained by 'Scaling theory'

"<u>Scaling factor</u>"

Scanned by CamScanner

"Transistor Scaling"

- helped the device designers to "shrink" the devices.

- Scaling theory laid down - **broad guidelines** of how the various parameters in a scaled transistor should change & **what the consequences** would be.
- Is less useful today (general scaling theory)

- Scaling of MOS transistor is **systematic reduction** of overall dimension's of the devices (as allowed by the available tech^n), while **preserving** the geometric ratio's found in the larger devices.

- Extent of scaling ie achievable is determined by the "fabricat^n technology" & by "min. feature size"

Advantages

- Feature size /Area ↓ses
  (Channel length L)
- Package density ↑ses

SSI	MSI	LSI	VLSI	ULSI	GSI	TSI
logic gates	Mux, demux	FSM registers	ASIC embedded design	gen. purpose processor	P-IV	v. high speed memory.

- More functionality
- Robust design (v. good perf^n ch^r)
- Better yield (failure rate v. low)
- Better Reliability in some systems

Scanned by CamScanner

- Transistor scaling factor 'S' ; $S > 1$
- Industry (Semiconductor fabrication) generally make use of $S = 1.25$ to 1.5
- <u>Types of Scaling:</u>
  1. Constant Electric field scaling (Full Scaling)
  2. Constant voltage scaling
- <u>Constant field scaling</u> (Full scaling):
  - Proposed by Dennard (1974)
  - As the device dimensions (W, L, t_{ox} & x_j) are scaled down by factor $\left(\frac{1}{S}\right)$, where $S > 1$, the <u>electric field</u> in the device should remain <u>constant</u>.
  - In order to keep electric fields constant, all voltages are also scaled by a factor $\left(\frac{1}{S}\right)$
  - Short form: CF scaling

Scanned by CamScanner

- Poisson's equation describing the relationship between charge densities and electric fields dictates that in CF scaling, the charge densities must be increased by a factor S in order to maintain the charge-field conditions.

- Constant voltage scaling:

- In CV scaling, all dimensions (L, W, t_{ox}, x_j) of the MOSFET are reduced by a factor 'S'
- The power supply voltage and terminal voltages remains unchanged.
- In CV scaling, doping densities are increased by a factor of S^2 in order to preserve charge-field relations (as per Poisson's equation).
- CV scaling allows the electric field to increase but requires that both vertical and Lateral fields scale by the same amount.

Scanned by CamScanner

Important MOSFET Parameters:

	MOSFET Parameters
Dimensions	Physical dimensions (W, L, t_{ox}, x_j)
	Depletion width (x_d)
Transistor electrical parameters	Doping densities (N_A, N_D)
	Electric field (E)
	Supply & terminal voltage ($V_{DD}, V_{GS}, V_{DS}, V_{SB}$)
Transistor / Chip Performance	On drive current (I_{ON})
	Capacitance (C)
	Delay time (τ)
	Power dissipation ($P_D = VI$)
	Chip complexity (Number of transistor on a chip)
	Functional throughput (Complexity × Speed) How powerful a chip made with such a technology is
	Power density

Scanned by CamScanner

Parameter	Before scaling	After Scaling: Constant-field scaling	After Scaling: Constant-voltage scaling	Comment
1. Channel Length	L	$\frac{L}{S}$	$\frac{L}{S}$	Dimension's reduces, device area ↓ses, → Overall package density ↑ses (can fit more transistor in chip area)
2. Channel width	W	$\frac{W}{S}$	$\frac{W}{S}$	
3. Device Area (Single transistor area)	$A = L \times W$	$\frac{W}{S} \times \frac{L}{S} = \frac{A}{S^2}$	$\frac{A}{S^2}$	
4. Gate oxide thickness	t_{ox}	$\frac{t_{ox}}{S}$	$\frac{t_{ox}}{S}$	Oxide thickness reduces by factor 'S', need of quality thin oxide materials ↳ ultra-thin
5. Source & Drain diffusion region depth	x_j	$\frac{x_j}{S}$	$\frac{x_j}{S}$	Junction depth reduces by factor 'S'

Scanned by CamScanner

Parameter	Before Scaling	After scaling CF scaling	CV scaling	Comment
6. Gate oxide capacitance	$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$ (F/cm^2)	$\frac{\epsilon_{ox}}{t_{ox}/S} = S.C_{ox}$	$S.C_{ox}$	Parasitic capacitance C_g is scaled down by factor 'S' ↓ we can predict that the transient characteristics (ie charge-up & charge-down times) of the scaled device will improve accordingly.
7. Parasitic Gate Capacitance	$C_g \propto C_{ox} W L$	$S C_{ox} \frac{W}{S} \frac{L}{S} = \frac{C_g}{S}$	$\frac{C_g}{S}$	• Proportional reduction of all dimensions on-chip will lead to a <u>reduction</u> of various parasitic capacitances & resistances $\left(R = \frac{\rho L}{A}\right)$ as well, RC delay ↓ses, contributing to the overall performance improvement.

Scanned by CamScanner

Parameter	Before scaling	After scaling CF scaling	CV scaling	Comment
8. Transconductance parameter	$K_n' = \mu_n C_{ox}$	$\mu_n S C_{ox} = S K_n'$	$S.K_n'$	Device gain ↑ses by factor S. Hence, sensitivity (ie ability to amplify/detect weak signal) of transistor increases.
9. Device Gain K_n, K_p	$K_n = K_n'\left(\frac{W}{L}\right)$ $K_p = K_p'\left(\frac{W}{L}\right)$	$S K_n$ $S K_p$	$S K_n$ $S K_p$	
10. Doping densities N_A, N_D	N_A $(/cm^3)$ N_D $(/cm^3)$	$S N_A$ $S N_D$	$S^2 N_A$ $S^2 N_D$	Based on Poisson's equation $\left(\frac{dV^2}{dx^2} = \frac{-\rho}{\epsilon};\ E = \frac{dV}{dx}\right)$, to maintain charge-field relationship, doping densities must be ↑sed by S & S^2 respectively.
11. Electric field (E)	E $\left(\frac{V}{cm}\right)$	E	SE (αE) ↑ alpha	In CV scaling, E-field ↑ses slightly by α $(1 < \alpha < S)$, but requires that both vertical & lateral fields scale by the same amount.

Scanned by CamScanner

Parameter	Before Scaling	After scaling CF scaling	CV scaling	Comment
12. Supply voltage	V_{DD} V_{SS}	V_{DD}/S V_{SS}/S	V_{DD} V_{SS}	Threshold voltage (V_{TO}) does not scale well, ie it is not possible to scale down V_{TO} easily.
13. Terminal voltages	V_{GS} V_{DS}	V_{GS}/S V_{DS}/S	V_{GS} V_{DS}	- The scaling of voltages may not be very practical in many cases - Peripheral & interface circuitry may require certain voltage levels for all input & output voltages, which in turn would necessitate multiple power supply voltages & complicated level-shifter arrangements For these reasons, CV scaling is usually preferred over CF scaling.
14. On drive current (I_{ON})	$I_{D(Lin)} = \frac{K_n}{2}\left[2(V_{GS}-V_{TO})V_{DS} - V_{DS}^2\right]$	$\frac{S K_n}{2}\frac{1}{S^2}$ I_{Dsat}/S	$S I_{Dsat}$	
Drain current (I_D)	$I_{D(sat)} = \frac{K_n}{2}(V_{GS}-V_{TO})^2$	I_{DLin}/S	$S I_{DLin}$	

Scanned by CamScanner

Parameter	Before scaling	After scaling CF scaling	CV scaling	Comment
15. Current density	$J = \frac{I_D}{A}$	$\frac{I_D/S}{A/S^2} = SJ$	$\frac{SI_D}{A/S^2} = S^3J$	Drain current density ↑ses by a factor of S^3. This large ↑se in current densities may eventually cause serious reliability problems for the scaled transistor, such as "electromigration".
16. Power dissipation	$P_D \propto V_{DD} I_D$	$\frac{V_{DD}}{S}\frac{I_D}{S} = \frac{P_D}{S^2}$	$V_{DD} I_D S = SP_D$	Power density by a factor of S^3. This large ↑se in power densities may eventually cause serious reliability problems for the scaled transistor, such as "electromigration", "hot-carrier degradation" which leads to ↑se in device temperature ↓ device can melt. ∴ Need for a proper heat sink required.
17. Power density (ψ)	$\psi = \frac{P_D}{A}$	$\frac{P_D/S^2}{A/S^2} = \psi$	$\frac{SP_D}{A/S^2} = S^3\psi$	

Scanned by CamScanner

Parameter	Before Scaling	After scaling CF scaling	After scaling CV scaling	Comment
18. Delay time (τ)	$\tau_{delay} \propto \frac{C\Delta V}{I_{DD}}$	τ/S	τ/S^2	In CV scaling, delay ↓ses by a factor of S^2
19. Switching Speed	S_s	$S \times S_s$	$S^2 \times S_s$	Speed ↑ses by a factor of S^2 in CV scaling
20. Chip Complexity	No. of transistor on a chip	S^2	S^2	Package density ↑ses → more functionality
21. Functional throughput (Complexity x Speed)	F.T	$F.T \times S^3$	$F.T \times S^4$	–
22. Power-delay Product	P_τ	$\frac{P_\tau}{S^3}$	$S P_\tau$	–

Scanned by CamScanner

A] Advantages of Constant field scaling:

1. Complexity of the circuit (number of transistors on a chip) increases by a factor S^2.
2. Speed improves by a factor S
3. The functional throughput, which measures how powerful a chip made with such a technology is, improves as S^3, all while keeping the (dynamic) chip power [power density] constant.
4. Power dissipation decreases by a factor S^2.
5. Parasitic capacitances & resistance reduces, RC delay ↓ses, contributing to overall performance improvement.

B] Limitations of Constant field Scaling:

1. Threshold voltage does not scale well, ie, it is not possible to scale down V_{TO} easily
2. The subthreshold current does not scale as $\left(\frac{1}{S}\right)$ and becomes larger in relation to the on drive current.

Scanned by CamScanner

3. It is difficult to scale voltages down as drastically as dimensions, since circuit voltages would soon reach the order of $\left(\frac{KT}{q}\right)$.

4. In CF scaling, there is a need of multiple power supply, need of complex interfacing & level translator circuit & hence complexity in design.

C] <u>Advantages of Constant voltage Scaling</u>:-

1. The complexity of the circuit (number of transistors on a chip) increases by a factor of S^2.
2. Delay time ↓ses by a factor of $\frac{1}{S^2}$, so that speed ↑ses rapidly with scaling
3. Functional throughput improves even further, going as S^4
4. CV scaling can operate with single power supply ie there is no need for complex interface and level translator circuit & hence design is simple
5. Parasitic capacitances & resistances reduces, RC delay ↓ses, contributing to overall performance improvement

Scanned by CamScanner

D] Limitations of Constant voltage scaling:

1. Power density ↑ses by S^3, which implies that the total chip power will also ↑se by S^3.
2. Drain current density ↑ses by a factor of S^3, this large increase in current and power densities may eventually cause serious reliability problems for the scaled transistor, such as electromigration, hot-carrier degradation, oxide breakdown, and electrical over-stress.
3. Power dissipation ↑ses by a factor 'S', which heats up the device quickly. Thus, there is a need for proper heat sinks.

—— × ——

Scanned by CamScanner

3.8 Compare the two technology scaling methods, namely, (i) the constant electric-field scaling and (ii) the constant power-supply voltage scaling. In particular, show analytically by using equations how the delay time, power dissipation, and power density are affected in terms of the scaling factor, S.

To be more specific, what would happen if the design rules change from, say, 1 μm to $1/S$ μm ($S > 1$) ?

Scanned by CamScanner

Solution:-

a) Constant electric-field scaling

$\rightarrow I'_{D(Lin)} = \frac{K_n}{2}\left[2(V_{GS}-V_{TO})V_{DS} - V_{DS}^2\right]$

$V_{TO} << V_{GS}$ & V_{DS} & it does scale down well

$\rightarrow I'_{D(Lin)} = \frac{SK_n}{2}\frac{1}{S^2}\left[2(V_{GS}-V_{TO})V_{DS} - V_{DS}^2\right]$

$$\boxed{I'_{D(Lin)} = \frac{I_{D(Lin)}}{S}}$$

Also, $I'_{D(sat)} = \frac{K_n}{2}(V_{GS}-V_{TO})^2 = \frac{SK_n}{2}\frac{1}{S^2}(V_{GS}-V_{TO})^2$

$$\boxed{I'_{D(sat)} = \frac{I_{D(sat)}}{S}}$$

i) Power dissipation (P_D):

$P_D = I_D V_{DS}$

$P'_D = \frac{I_D}{S} \cdot \frac{V_{DS}}{S} = \frac{P_D}{S^2}$

$$\boxed{P'_D = \frac{P_D}{S^2}}$$

Scanned by CamScanner

ii) Power density ψ:

$$\psi' = \frac{P_D'}{A} = \frac{P_D/S^2}{A/S^2} = \psi$$

$$\boxed{\psi' = \psi}$$ --- remain same

iii) Delay time (τ):

$$\tau' \approx \frac{C\Delta V}{I_{DD}} \approx \frac{C/S \times \Delta V/S}{I_{DD}/S} = \frac{\tau}{S}$$

$$\boxed{\tau' = \frac{\tau}{S}}$$ --- reduces by S

b) Constant Power-supply voltage scaling:-

$$\rightarrow I_{D(lin)}' = \frac{K_n}{2}\left[2(V_{GS} - V_{TO})V_{DS} - V_{DS}^2\right]$$

$$= \frac{SK_n}{2}\left[2(V_{GS} - V_{TO})V_{DS} - V_{DS}^2\right]$$

$$\boxed{I_{D(lin)}' = S I_{D(lin)}}$$

$$\rightarrow I_{D(sat)}' = \frac{SK_n}{2}\left[V_{GS} - V_{TO}\right]^2$$

$$\boxed{I_{D(sat)}' = S I_{D(sat)}}$$

Scanned by CamScanner

i) <u>Power dissipation P_D :</u>

$$P_D' = I_D V_{DS}$$

$$= S I_D V_{DS}$$

$$\boxed{P_D' = S . P_D}$$ ---- ↑ses by S

ii) <u>Power density φ :</u>

$$\varphi' = \frac{P_D'}{A} = \frac{S P_D}{A/S^2}$$

$$\boxed{\varphi' = S^3 \varphi}$$ ---- ↑ses by a factor S^3

iii) <u>Delay time τ :</u>

$$\tau' = \frac{C \Delta V}{I_{DD}} = \frac{C/S \times \Delta V}{S I_{DD}} = \frac{\tau}{S^2}$$

$$\boxed{\tau' = \frac{\tau}{S^2}}$$ ---- ↓ses by a factor S^2

— x —

Scanned by CamScanner

Numerical 01: (Scaling)

29/1/2020

Mr Snape is planning to scale down device X whose parameters are:

$W = 180nm$, $L = 90nm$, $V_{DD} = 5V$, $I_{ON} = 1500\mu A$

$P_D = 500\mu W$, $t_{ox} = 18°A$, doping densities $N_A = 10^{16}/cm^3$

Intrinsic gate delay = 0.95 ps, and switching speed = 1052.6 GHz

What will be the new values of above parameters when the device X is scaled down by using

a) Constant Electric-field scaling

b) Constant supply voltage scaling

Scaling factor S to be considered as $\sqrt{2}$

Sol^n:

	Before scaling	Const field scaling		Const voltage scaling	
W	180nm	$\frac{W}{S} = \frac{180nm}{\sqrt{2}}$	$\boxed{127.28\ nm}$	$\frac{W}{S}$	$\boxed{127.28nm}$
L	90nm	$\frac{L}{S} = \frac{90nm}{\sqrt{2}}$	63.64 nm	$\frac{L}{S}$	63.64 nm
V_{DD}	5V	$\frac{V_{DD}}{S} = \frac{5V}{\sqrt{2}}$	3.535 V	V_{DD}	5V
I_{ON}	1500μA	$\frac{I_{ON}}{S} = \frac{1500\mu A}{\sqrt{2}}$	1060 μA	$S I_{ON}$	$1500\mu A \times \sqrt{2}$ 2121 μA
P_D	500μW	$\frac{P_D}{S^2} = \frac{500\mu W}{(\sqrt{2})^2}$	250μW	$S P_D$	707 μW
t_{ox}	18°A	$\frac{t_{ox}}{S} = \frac{18°A}{\sqrt{2}}$	12.72°A	$\frac{t_{ox}}{S}$	12.72°A

Scanned by CamScanner

		Before Scaling	CF scaling	CV scaling
Doping densities	N_A	$10^{16}/cm^3$	$sN_A = \sqrt{2}\times10^{16}$ $1.414\times10^{16}/cm^3$	$s^2N_A = 2\times10^{16}/cm^2$
Intrinsic	Gate delay τ	0.95ps	$\frac{\tau}{S} = \frac{0.95ps}{\sqrt{2}}$ 0.672ps	$\frac{\tau}{S^2} = \frac{0.95ps}{2}$ 0.475ps
Intrinsic	Switching speed S_f	1052.6 GHz	$S_f \times S = 1052.6GHz\times\sqrt{2}$ 1488.6 GHz	$S_f^2 \times S_f = 1052.6GHz \times 2$ 2105 GHz

Scanned by CamScanner