Lecture 32 - The "Short" Metal-Oxide-Semiconductor Field-Effect Transistor *(cont.)*

April 27, 2007

Contents:

1. MOSFET scaling

Reading assignment:

P. K. Ko, "*Approaches to Scaling.*"

Key questions

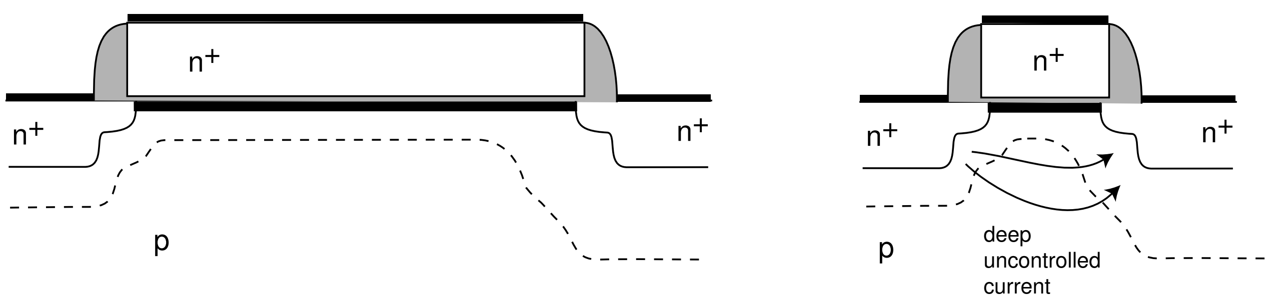
- What happens if a MOSFET gate length is simply shrunk in size without changing anything else?
- How should the MOSFET design change as it shrinks down in size?

1. MOSFET scaling

Several driving forces for scaling down size of MOSFET:

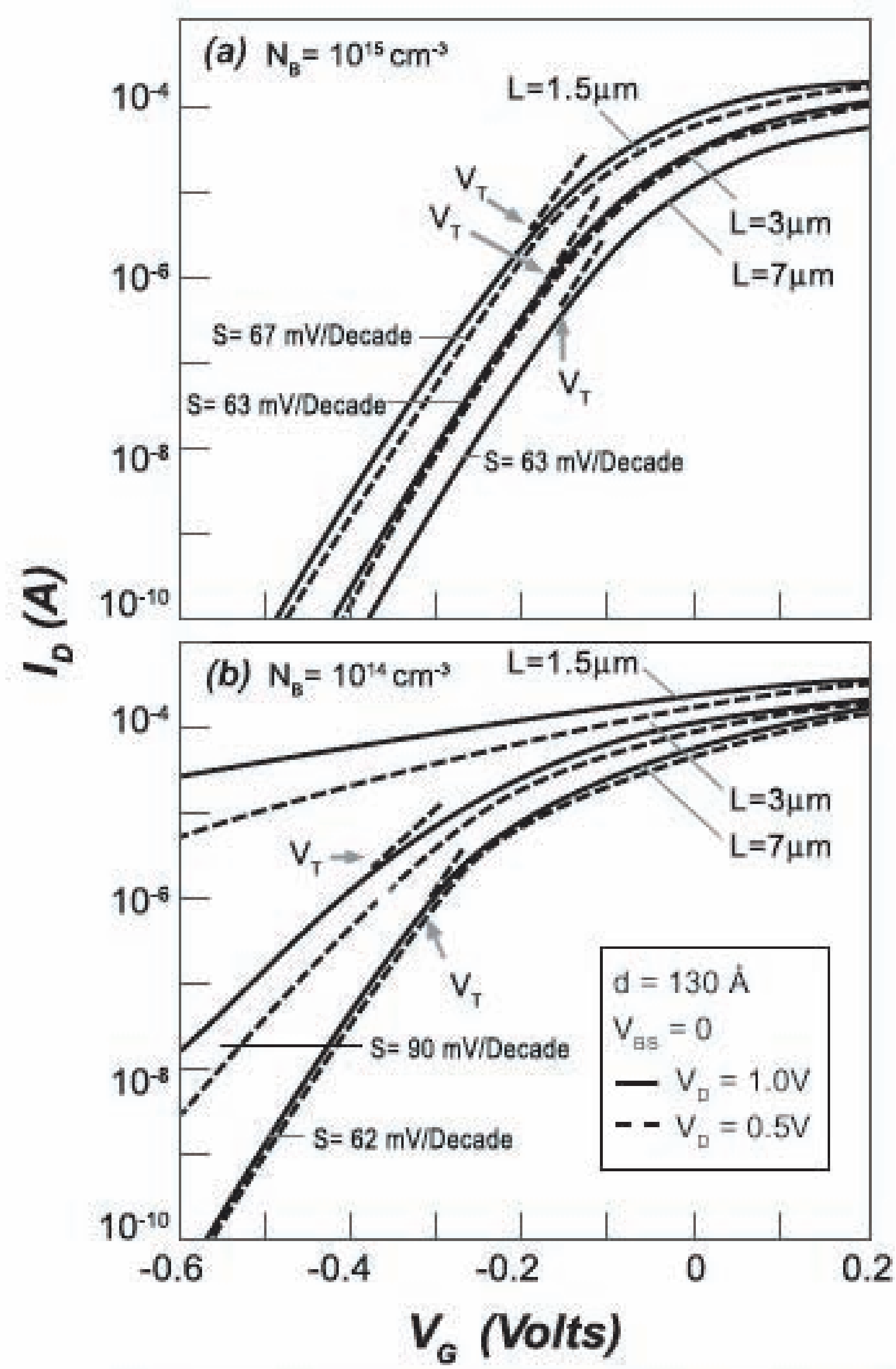
- higher density circuits: SSI, MSI, LSI, VLSI, ULSI, RLSI, ...
- higher performance: $L \downarrow \Rightarrow I_D \uparrow \Rightarrow \tau_{switch} \downarrow$
- lower power consumption: $L \downarrow \Rightarrow V_{DD} \downarrow$

Simple L scaling compromises *electrostatic integrity* and produces *punchthrough* (extreme case of short-channel effects):



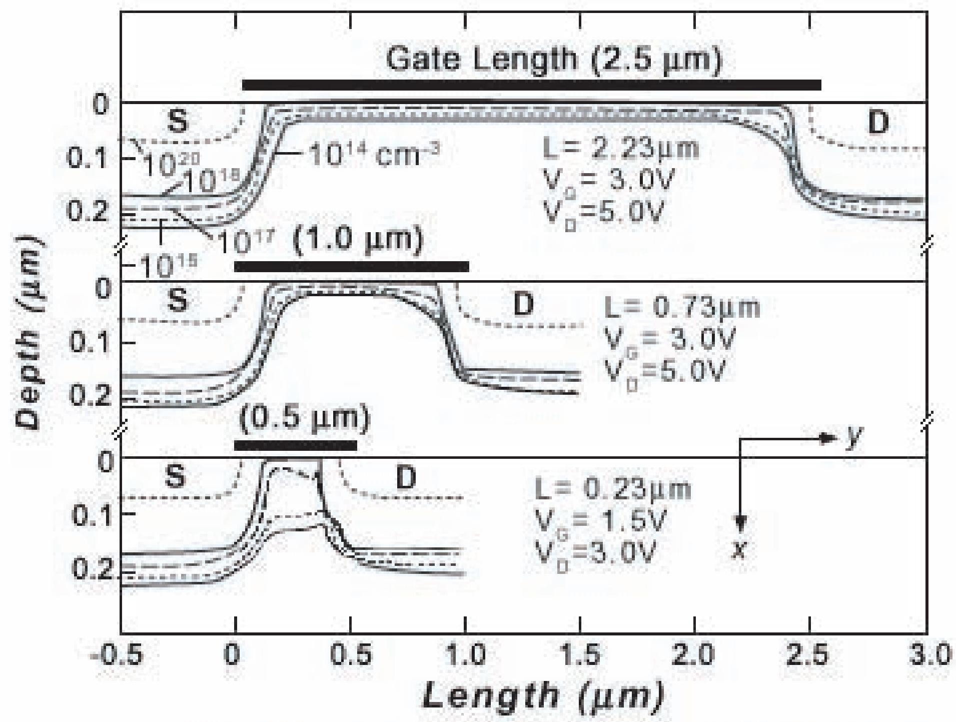
To avoid punchthrough:

- $N_A \uparrow \Rightarrow V_T \uparrow \Rightarrow I_D \downarrow$
- $V_{DD} \downarrow \Rightarrow I_D \downarrow$
- $x_{ox} \downarrow \Rightarrow V_T \downarrow \Rightarrow I_D \uparrow$



Subthreshold characteristics for various channel lengths.
(a) $N_A=10^{15}$ cm^{-3}, (b) $N_A=10^{14}$ cm^{-3}.
Adapted from S. M. Sze,
Physics of Semiconductor Devices, 2nd ed., Wiley, 1981 (470)

Image by MIT OpenCourseWare.
Adapted from Sze, S. M. *Physics of Semiconductor Devices*. 2nd ed.
New York, NY: John Wiley & Sons, 1981, p. 470. ISBN: 9780471056614.



Constant electron density contours for 3 MOSFETs with channel lengths 2.23, 0.73, and 0.23 µm.

Adapted from S. M. Sze,
Physics of Semiconductor Devices, 2nd ed., Wiley, 1981 (481).

Image by MIT OpenCourseWare.
Adapted from Sze, S. M. *Physics of Semiconductor Devices*. 2nd ed.
New York, NY: John Wiley & Sons, 1981, p. 481. ISBN: 9780471056614.

Need smart way of scaling:

- constant field scaling
- constant voltage scaling
- generalized scaling

□ Constant field scaling

Scale keeping vertical and horizontal electric fields constant.

Define: *scaling factor* $S > 1$

parameter	**scaling factor**
device dimensions (L, W, x_{ox})	$1/S$
doping level (N_A)	S
supply voltage (V_{DD})	$1/S$

Consequences (use simple long-channel theory):

- gate capacitance:

$$C'_{gs} = C'_{ox}L'W' = SC_{ox}\frac{L}{S}\frac{W}{S} = \frac{C_{gs}}{S} \downarrow$$

- threshold voltage:

$$V'_T = V_{FB} + \phi_{sth} + \gamma\sqrt{\phi_{sth}} \simeq \frac{1}{C'_{ox}}\sqrt{2\epsilon_s q N'_A \phi_{sth}} \sim \frac{V_T}{\sqrt{S}} \downarrow$$

- drive current:

$$I'_D = \frac{W'}{2L'}\mu_e C'_{ox}(V'_{DD} - V'_T)^2 = \frac{\frac{W}{S}}{2\frac{L}{S}}\mu_e S C_{ox}(\frac{V_{DD}}{S} - \frac{V_T}{\sqrt{S}})^2 = \frac{I_D}{S} \downarrow$$

- gate delay:

$$\tau' = \frac{C'_{gs}V'_{DD}}{I'_D} = \frac{\frac{C_{gs}}{S}\frac{V_{DD}}{S}}{\frac{I_D}{S}} = \frac{\tau}{S} \downarrow$$

- power-delay product or *switching energy*:

$$C'_{gs}{V'_{DD}}^2 = \frac{C_{gs}}{S}(\frac{V_{DD}}{S})^2 = \frac{C_{gs}V_{DD}^2}{S^3} \downarrow\downarrow\downarrow$$

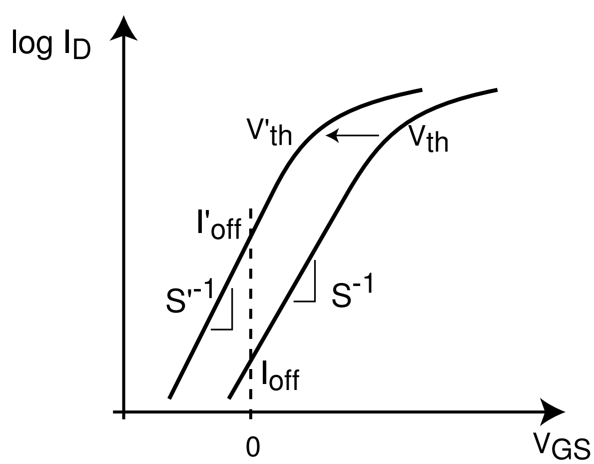
- switching energy density:

$$\frac{C'_{gs}{V'_{DD}}^2}{L'W'} = \frac{\frac{C_{gs}V_{DD}^2}{S^3}}{\frac{L}{S}\frac{W}{S}} = \frac{1}{S}\frac{C_{gs}V_{DD}^2}{LW} \downarrow$$

- inverse subthreshold slope:

$$n' = 1 + \frac{C'_{sth}}{C'_{ox}} = 1 + \frac{\sqrt{S}C_{sth}}{SC_{ox}} = 1 + \frac{C_{sth}}{\sqrt{S}C_{ox}} \downarrow$$

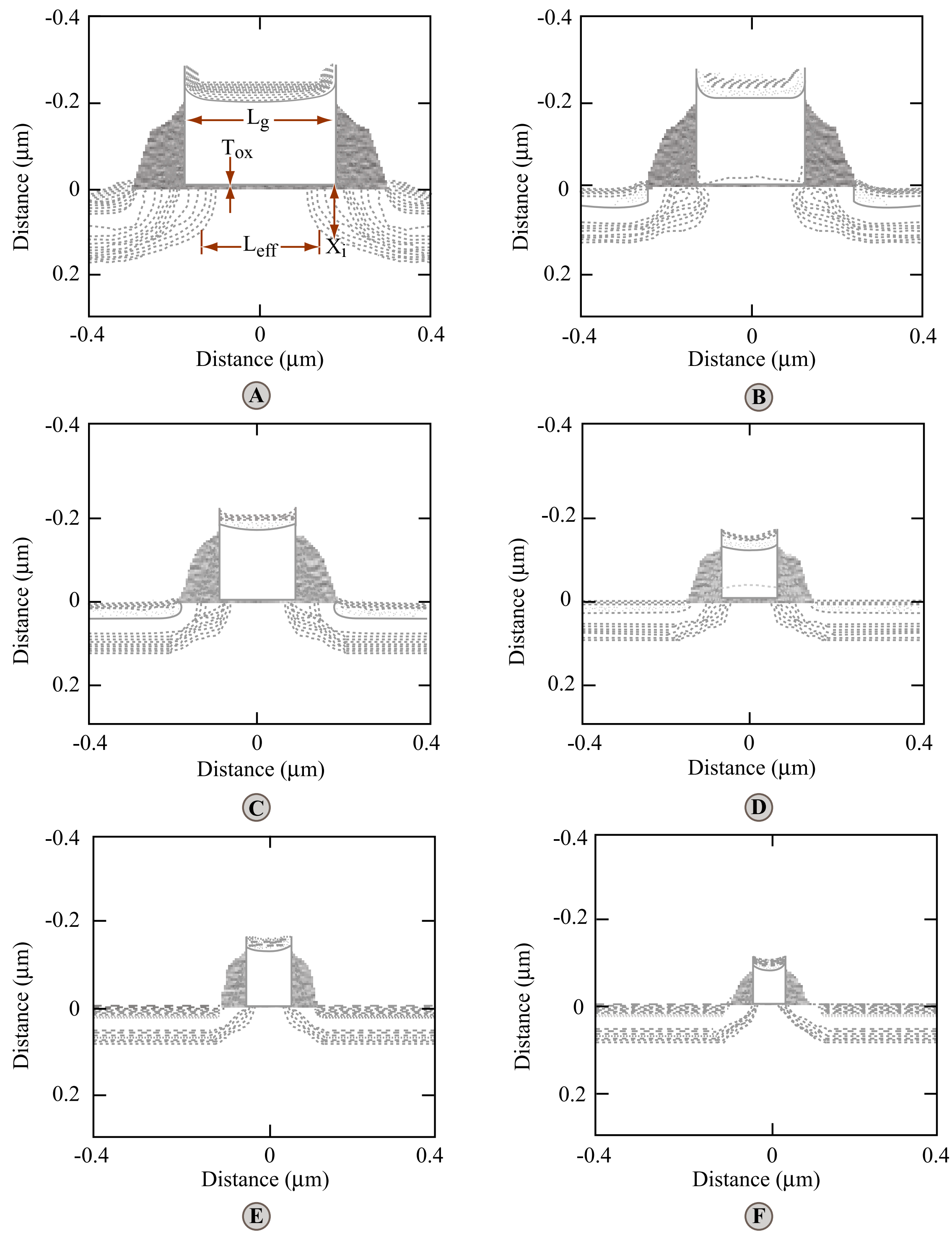
but since $V_T \downarrow$, $I_{off} \uparrow\uparrow$.



Two key problems with constant field scaling:

- system designers don't want to scale V_{DD}
- $I_{off} \uparrow\uparrow \Rightarrow$ more static power

□ More rigorous study of constant field scaling using 2D simulations [P. Vande Voorde, HP Journal, 1997]



Simulated device structures. Dark shading is Oxide. Lighter shading is Silicide. Dashed line are doping contours.
(a) L_g = 0.35 μm, T_{ox} = 8.0 nm. (b) L_g = 0.25 μm, T_{ox} = 6.0 nm. (c) Lg = 0.18 μm, T_{ox} = 4.5 nm. (d) Lg = 0.13 μm, T_{ox} = 3.4 nm. (e) Lg = 0.10 μm, T_{ox} = 2.5 nm. (f) Lg = 0.07 μm, T_{ox} = 1.9 nm.

Image by MIT OpenCourseWare. Adapted from Figure 1 on p. 97 in Vande Voorde, Paul. "MOSFET Scaling into the Future." *Hewlett-Packard Journal* (August 1997): 96-100.

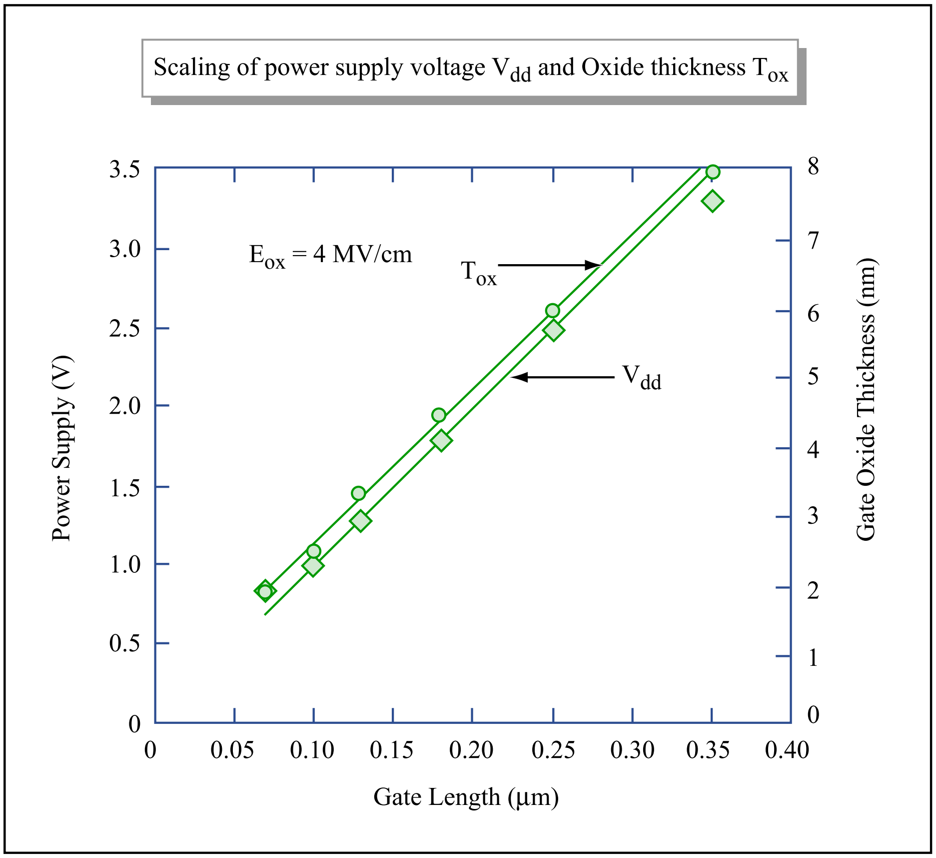


Image by MIT OpenCourseWare. Adapted from Figure 2 on p. 98 in Vande Voorde, Paul. "MOSFET Scaling into the Future." *Hewlett-Packard Journal* (August 1997): 96-100.

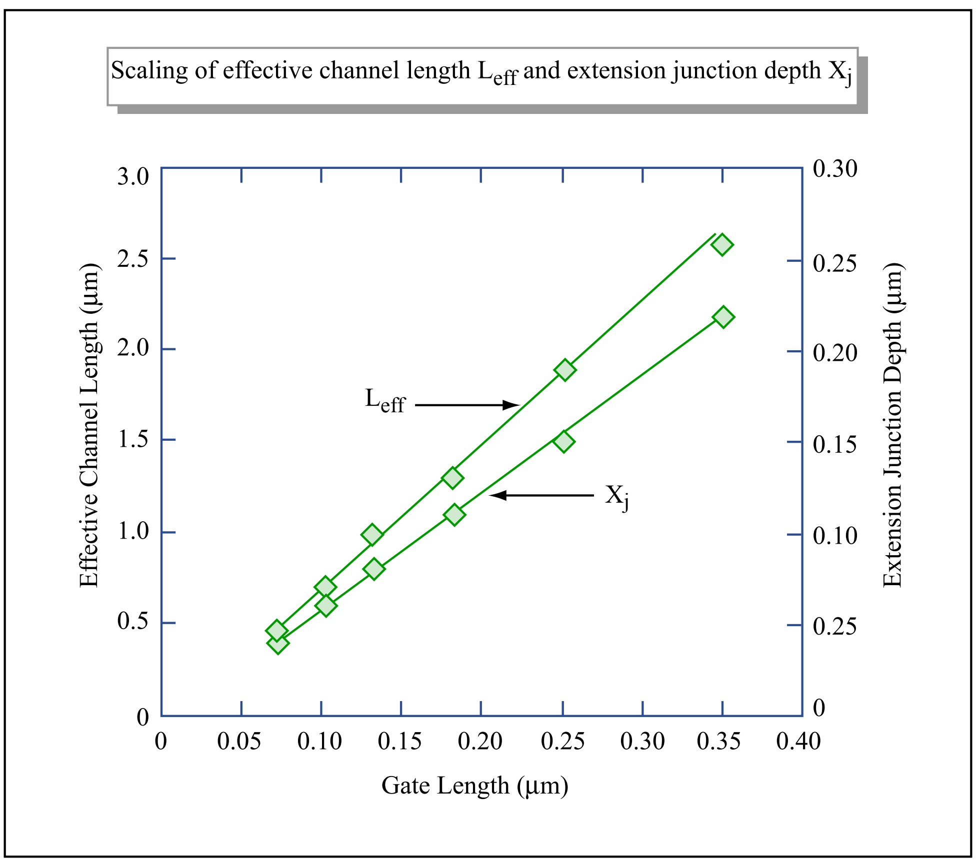


Image by MIT OpenCourseWare. Adapted from Figure 3 on p. 98 in Vande Voorde, Paul. "MOSFET Scaling into the Future." *Hewlett-Packard Journal* (August 1997): 96-100.

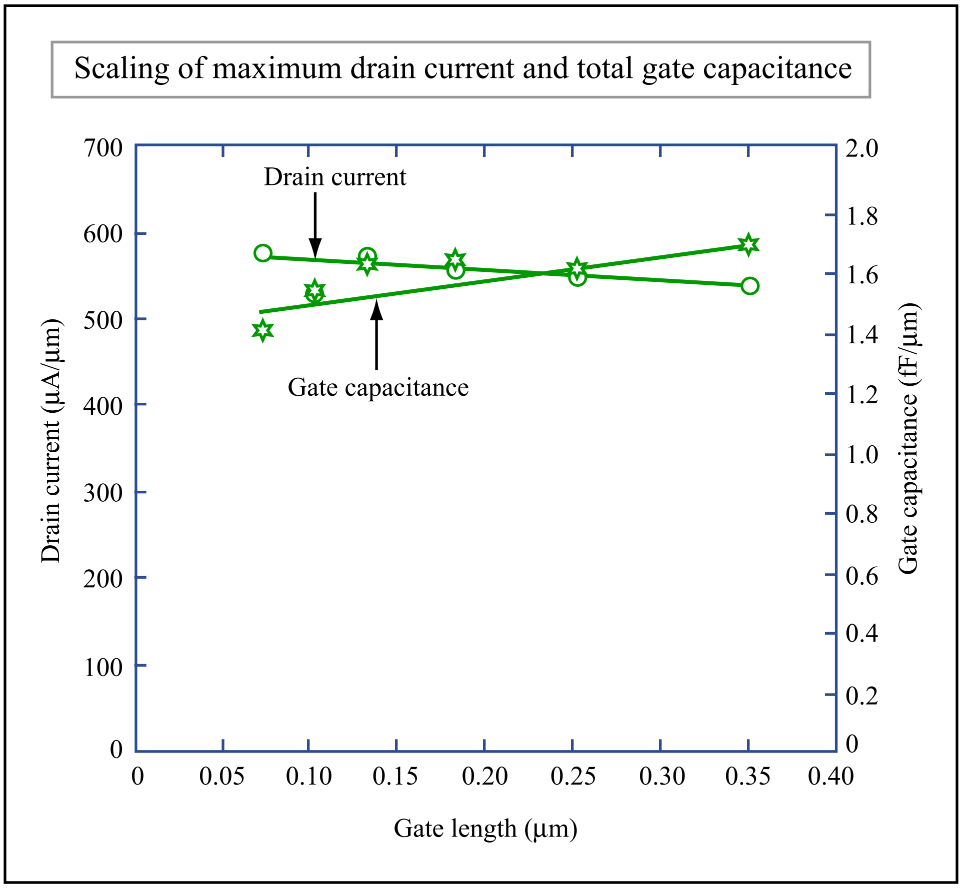


Image by MIT OpenCourseWare. Adapted from Figure 5 on p. 98 in Vande Voorde, Paul. "MOSFET Scaling into the Future." *Hewlett-Packard Journal* (August 1997): 96-100.

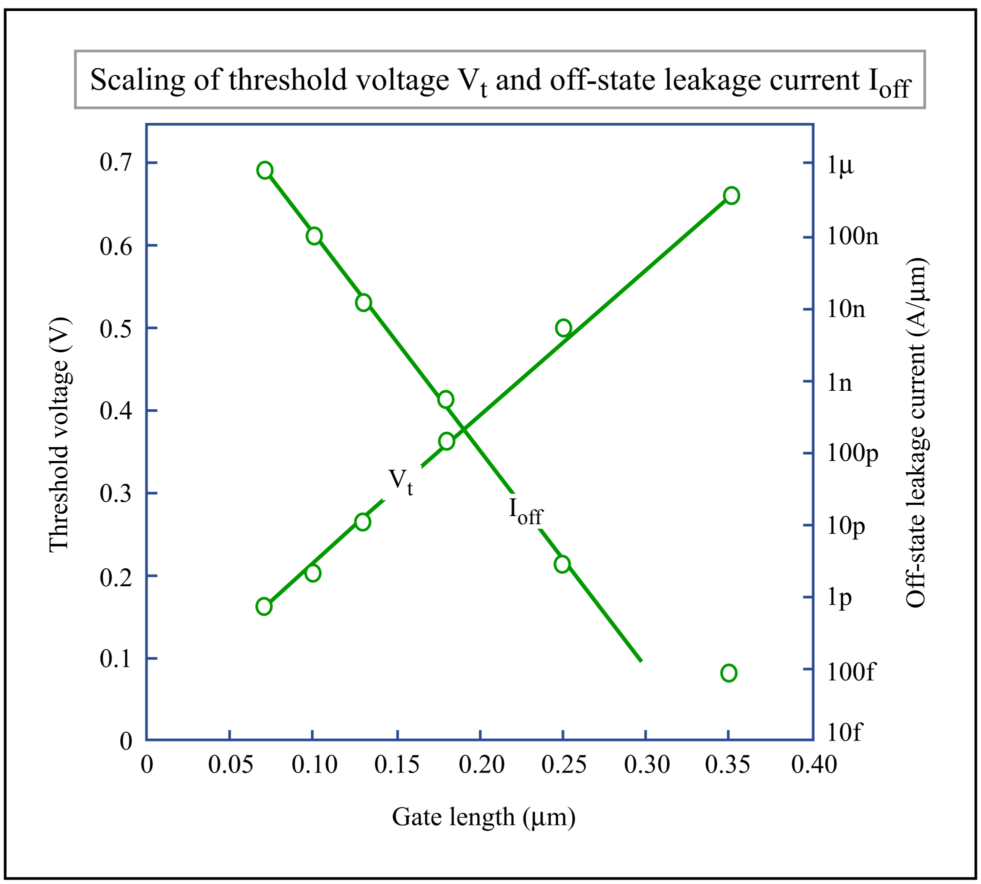


Image by MIT OpenCourseWare. Adapted from Figure 4 on p. 98 in Vande Voorde, Paul. "MOSFET Scaling into the Future." *Hewlett-Packard Journal* (August 1997): 96-100.

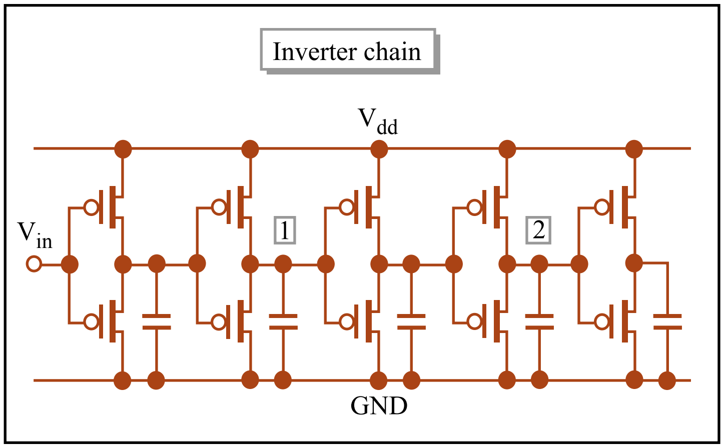


Image by MIT OpenCourseWare. Adapted from Figure 8 on p. 99 in Vande Voorde, Paul. "MOSFET Scaling into the Future." *Hewlett-Packard Journal* (August 1997): 96-100.

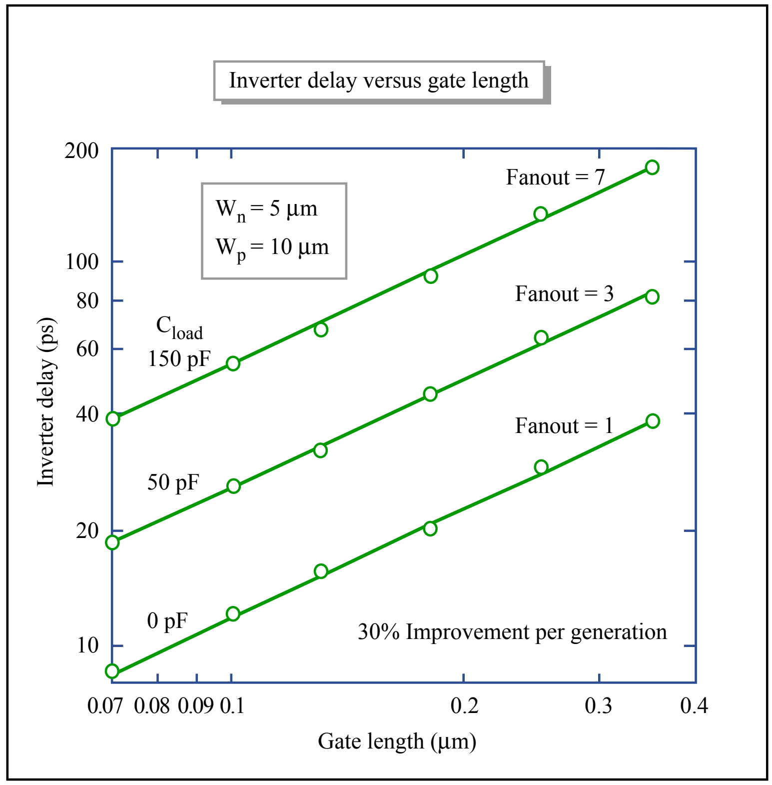


Image by MIT OpenCourseWare. Adapted from Figure 10 on p. 99 in Vande Voorde, Paul. "MOSFET Scaling into the Future." *Hewlett-Packard Journal* (August 1997): 96-100.

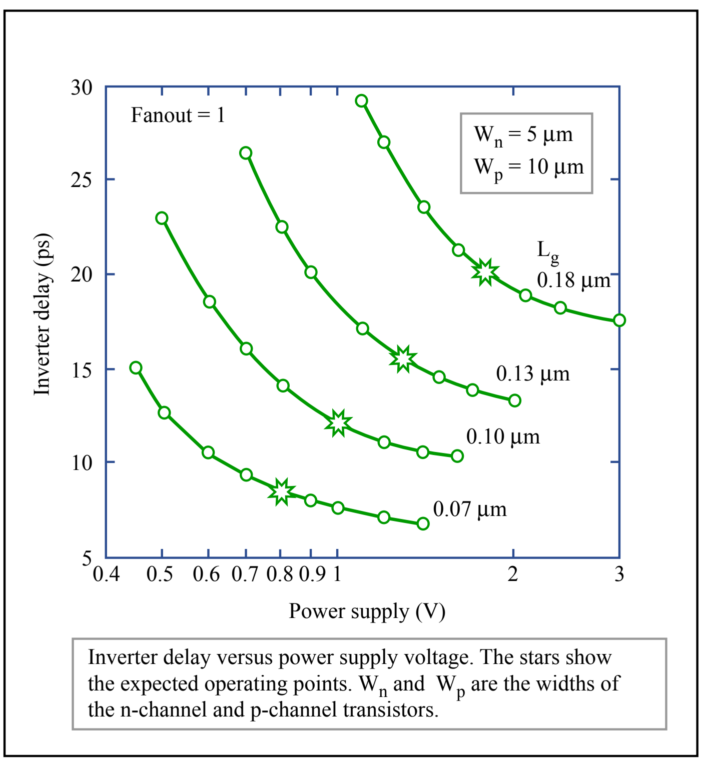


Image by MIT OpenCourseWare. Adapted from Figure 11 on p. 99 in Vande Voorde, Paul. "MOSFET Scaling into the Future." *Hewlett-Packard Journal* (August 1997): 96-100.

□ Constant voltage scaling

Scale all device dimensions but do not scale V_{DD}.

parameter	**scaling factor**
device dimensions (L, W, x_{ox})	$1/S$
doping level (N_A)	S
supply voltage (V_{DD})	1

Consequences (using long-channel theory):

figure of merit	**scaling factor**
C_{gs}	$1/S$
V_{th}	$1/\sqrt{S}$
I_D	S
τ	$1/S^2$
$C_{gs}V_{DD}^2$	$1/S$
$C_{gs}V_{DD}^2/LW$	S

Features of constant voltage scaling:

- Performance $\uparrow\uparrow$
- But:
  - It does not address I_{off} problem.
  - Electric field across oxide $\uparrow$:

$$\mathcal{E}_{ox} = \frac{V_{DD}}{x_{ox}} \propto S \uparrow$$

Reliability problems when $\mathcal{E}_{ox} \simeq 4\ MV/cm$.

  - Electric field in semiconductor (at drain end of channel) $\uparrow$:

$$\mathcal{E}_m = \sqrt{\frac{(V_{DS} - V_{DSsat})^2}{l^2} + \mathcal{E}_{sat}^2} \propto S \uparrow$$

with

$$l^2 = \frac{\epsilon_s}{\epsilon_{ox}} x_{ox} x_j \propto S^{-2}$$

Reliability problems when $\mathcal{E}_m \simeq 0.5\ MV/cm$.

  - Power density $\uparrow \Rightarrow$ system power $\uparrow$

□ Generalized scaling

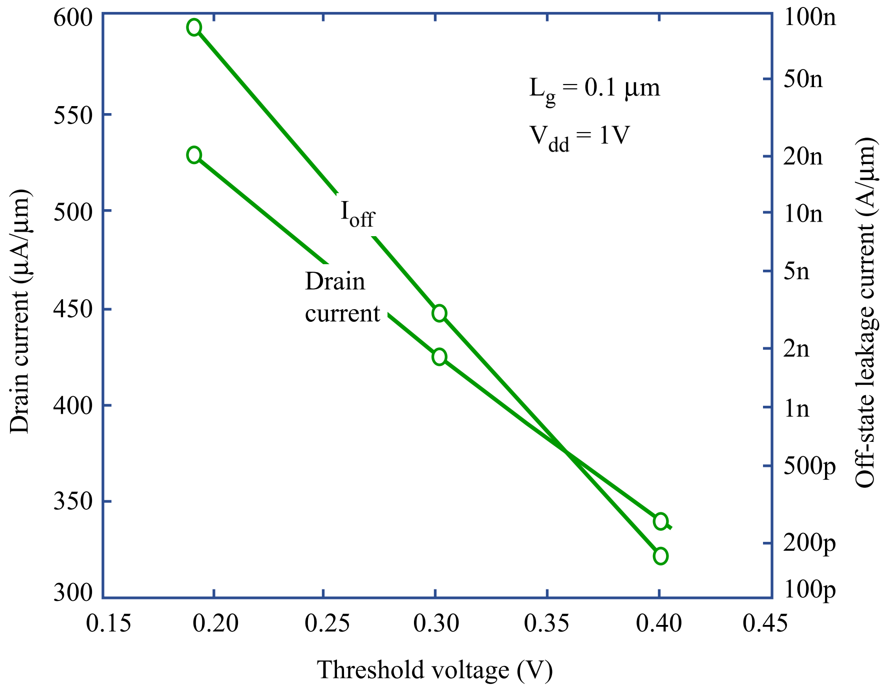
- scale oxide thickness more slowly than other device dimensions
- scale V_{DD} keeping $\mathcal{E}_{ox}$ constant

parameter	**scaling factor**
L, W	$1/S$
x_{ox}	$1/R$
N_A	S
V_{DD}	$1/R$

with $1 < R < S$.

In generalized scaling:

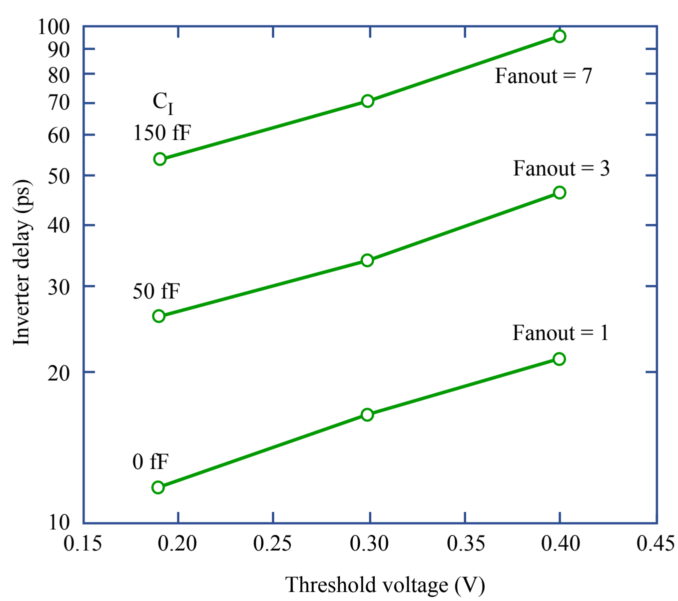
- I_{off} problem alleviated by not scaling V_T so aggresively;
  trade-off: performance
- V_{DD} scales;
  trade-off: performance



Drain current and off-state leakage current I_{off} versus threshold voltage V_t for the 0.1-µm generation.

Image by MIT OpenCourseWare. Adapted from Figure 12 on p. 100 in Vande Voorde, Paul. "MOSFET Scaling into the Future." *Hewlett-Packard Journal* (August 1997): 96-100.

Inverter delay versus threshold voltage V_t for the 0.1-µm generation



L_g = 0.1 µm V_{dd} = 1V W_n = 5 µm W_p = 10 µm

Image by MIT OpenCourseWare. Adapted from Figure 13 on p. 100 in Vande Voorde, Paul. "MOSFET Scaling into the Future." *Hewlett-Packard Journal* (August 1997): 96-100.

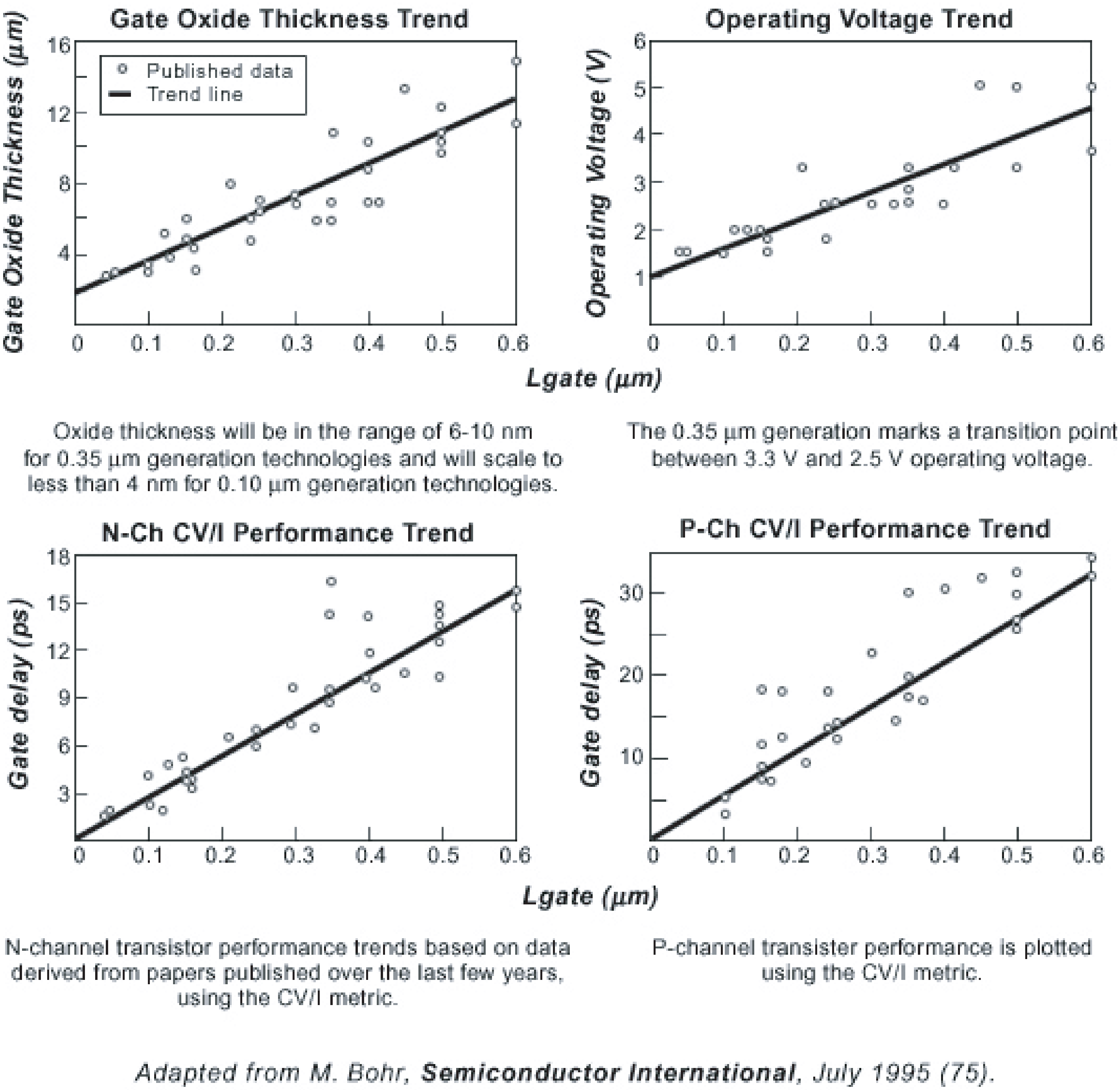
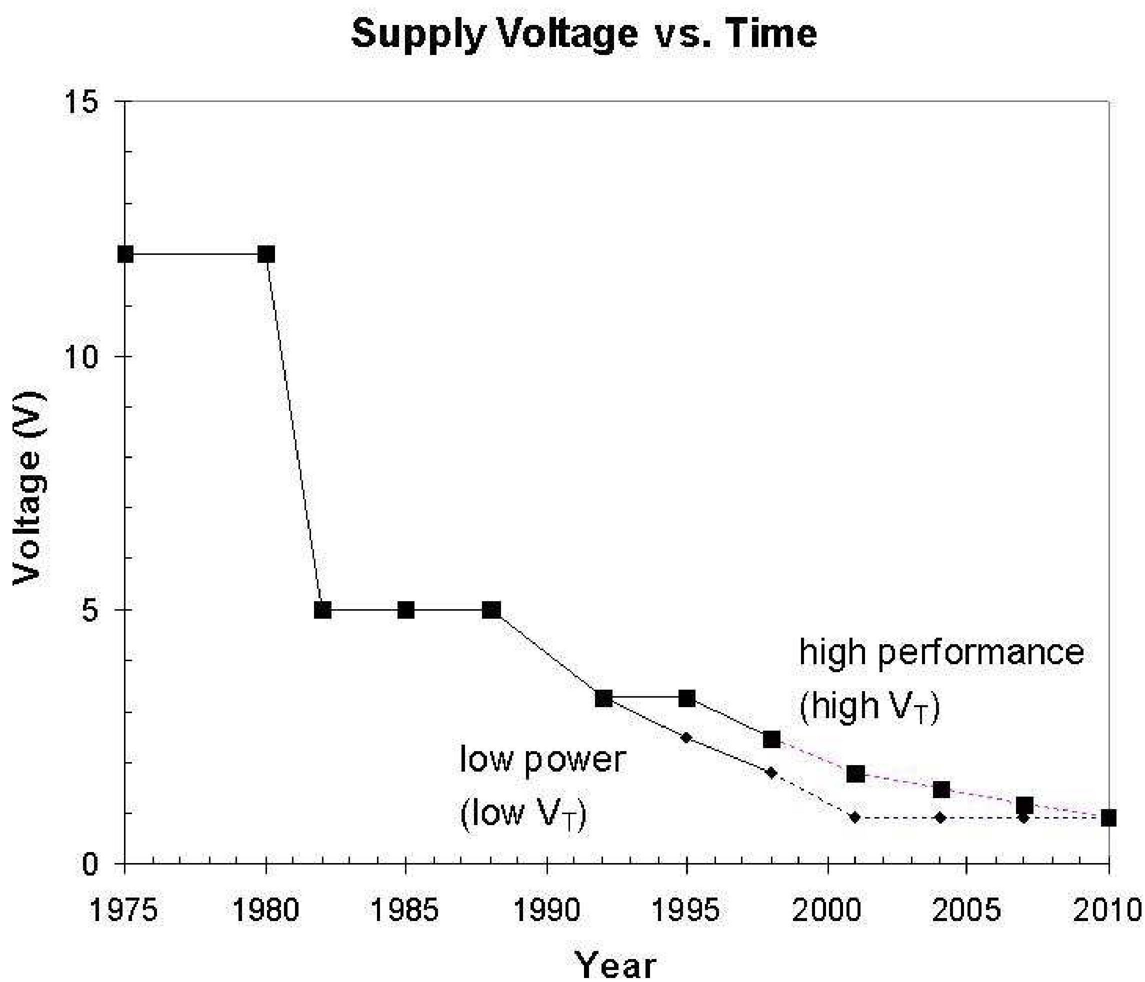


Image by MIT OpenCourseWare. Adapted from Bohr, M. *Semiconductor International* (July 1995): 75.

□ Modern generalized scaling

- Concept of *generation*: every 2 years, new technology is deployed with 30% reduced transistor delay and twice as high transistor density (microprocessor performance doubling every 2 years).
- Everything scales: L (↓), W (↓), x_{ox} (↓), N_A (↑), x_j (↓), and V_{DD} (↓).
- Scaling goal: *extract maximum performance from each generation* (maximize I_{on}), for a given amount of:
  - short-channel effects (DIBL), *and*
  - off-current
- Currently two technology flavors:
  - *high-performance*: high V_{DD} (high I_D, low τ), low V_T (high I_{off});
  - *low-power*: low V_{DD} (low I_D, high τ), high V_T (low I_{off}).


Supply Voltage vs. Time
15
10
5
0
Voltage (V)
1975
1980
1985
1990
1995
2000
2005
2010
Year
high performance
(high V_T)
low power
(low V_T)

Key conclusions

- *Constant field scaling*: scale all device dimensions keeping vertical and horizontal electric fields constant.

  Consequences:

  - $I_{off} \uparrow$
  - system designers don't want to scale V_{DD}

- *Constant voltage scaling*: scale all device dimensions keeping voltage constant.

  Consequences:

  - $I_{off} \uparrow$
  - fields everywhere $\uparrow \Rightarrow$ reliability compromised

- For a long time scaling proceeded through constant V_{DD} path with abrupt drops in V_{DD}.

- Scaling goal: *extract maximum performance from each generation* (maximize I_{on}), for a given amount of:

  - short-channel effects (DIBL), *and*
  - off-current

- *Generalized scaling* demands simultaneous scaling of L_g, x_{ox}, x_j, N_A, and V_{DD}.