Physics and Engineering of CMOS Devices

Ken Uchida
Department of Physical Electronics
Tokyo Institute of Technology

Scaling

Scaling of MOS Transistors

- By scaling down the transistor dimensions, the number of transistors per chip (area) can be increased.
- Furthermore, the performance of transistors is also improved by the scaling. The performance of transistors (LSIs) is measured in terms of drain current (driving capability of transistors) or/and speed.

Scaling Strategy: Constant Field (1)

For MOSFETs, the gate controllability is the key. However, in shorter channel MOSFETs, the gate controllability easily gets lost. The scaling theory of MOSFETs is the strategy to keep the gate controllability in shorter channel MOSFETs. The point is to keep the electric field inside MOSFETs be the same in shorter channel MOSFETs to avoid the short channel effects (SCE). Suppose the transistor dimensions are scaled by a factor of $1/k$ ($k>1$).

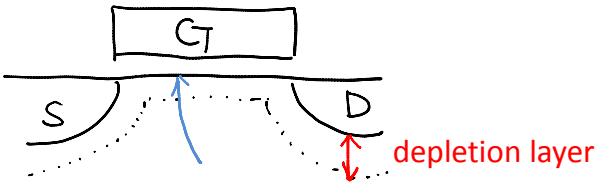
$$x \rightarrow x' = \frac{x}{k}$$

$$\phi \rightarrow \phi' = \frac{\phi}{k}$$

$$E \rightarrow E' = \frac{\partial \phi'}{\partial x'} = E$$

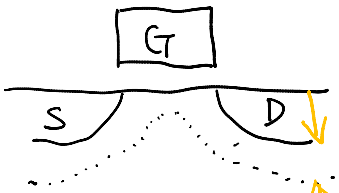
Schematic of MOS Transistors

Longer Channel



The channel potential is controlled by the gate voltage.

Shorter Channel

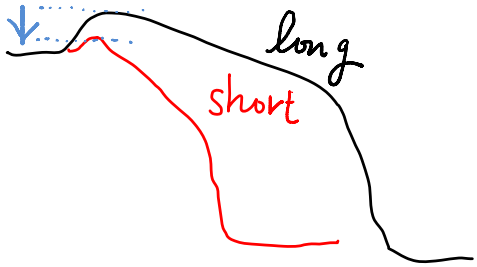


This should be scaled!

The channel potential is controlled by the gate voltage as well as the source / drain voltage.

Not only device dimensions but also depletion layer width should be scaled!

Electric field in the channel



The potential is lowered by the drain field, which contributes to reduce the threshold voltage.

Scaling Strategy: Constant Field (2)

In order to keep the electric field in MOS transistors constant, substrate impurity should be increased by k.

$$d\left(\frac{d\phi}{dx}\right) = \frac{qN_{\text{sub}}}{\kappa_s \varepsilon_0} dx$$

$$\frac{d\phi}{dx} = \int_0^{x_d/k} \frac{qN_{\text{sub}}}{\kappa_s \varepsilon_0} dx = \frac{x_d}{k} \frac{qN_{\text{sub}}}{\kappa_s \varepsilon_0}$$

$$N_{\text{sub}} \rightarrow N_{\text{sub}}' = kN_{\text{sub}}$$

The substrate impurity concentration should be increased by the factor of k.

Impact of Scaling on Device Performance

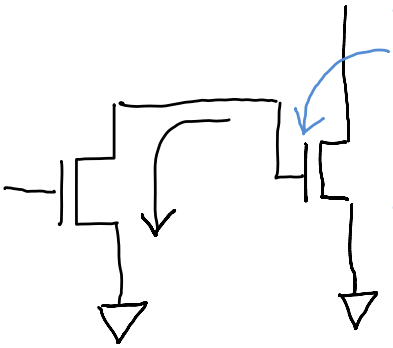
Capacitance per unit area

$$I_d = \mu C_{ox} \frac{W}{L}\left(V_g - V_{th}\right)^2$$

$\times k$ $\times \frac{1}{k^2}$

$$I_d \propto \frac{1}{k}$$

I_d is reduced by $1/k$. I_d per unit width stays constant.



The gate capacitance of MOS capacitor, which is the load capacitor for the transistors, is expressed as

$$C_g = \frac{LW\kappa_s\varepsilon_0}{t_{ox}} \propto \frac{1}{k}$$ (decreases)

The delay time, which is the time needed to charge up or discharge the load capacitor (MOS capacitor) is given by

$$\tau = \frac{C_g V_d}{I_d} = \propto \frac{1/k \cdot 1/k}{1/k} = \frac{1}{k}$$

The delay time is decreased as we scale down the transistor dimensions!

Scaling Strategy: Constant Field (3)

Device/Circuit Parameter	Scaling Factor
Device Dimension (L, W, t_{ox})	$1/k$
Doping Concentration (N_{sub})	k
Voltage (V)	$1/k$
Current (I)	$1/k$
Capacitance per unit area (C_{ox})	k
Delay time (1/Speed)	$1/k$
Power dissipation per circuit	$1/k^2$
Power density	1

The voltage has NOT been scaled as was expected.

Generalized Scaling (1)

However, because of non-scalable factors, voltages (power supply voltage) has not been lowered as was expected from the constant-field scaling theory. We will introduce a new scaling factor α ($\alpha < k$).

$$x \rightarrow x' = \frac{x}{k}$$

$$\phi \rightarrow \phi' = \frac{\phi}{\alpha}$$

$$E \rightarrow E' = \frac{\partial \phi'}{\partial x'} = \frac{k}{\alpha} E$$

>1

$$\frac{k}{\alpha}\frac{d\phi}{dx} = \int_0^{x_d/k} \frac{qN_{\mathrm{sub}}}{\kappa_s \varepsilon_0} dx = \frac{x_d}{k}\frac{qN_{\mathrm{sub}}}{\kappa_s \varepsilon_0}$$

$$N_{\mathrm{sub}} \rightarrow N_{\mathrm{sub}}' = \frac{k^2}{\alpha} N_{\mathrm{sub}}$$

>1

The electric field in MOSFETs increases with a decrease in dimensions!

Generalized Scaling (2)

Device/Circuit Parameter	Scaling Factor
Device Dimension (L, W, t_{ox})	$1/k$
Doping Concentration (N_{sub})	k^2/α
Voltage (V)	$1/\alpha$
Current (I)	k/α^2
Capacitance per unit area (C_{ox})	k
Delay time (1/Speed)	α/k^2
Power dissipation per circuit	k/α^3
Power density	$(k/\alpha)^3$

The power consumption increases dramatically in LSIs with scaled devices!