6.3 Scaling rule

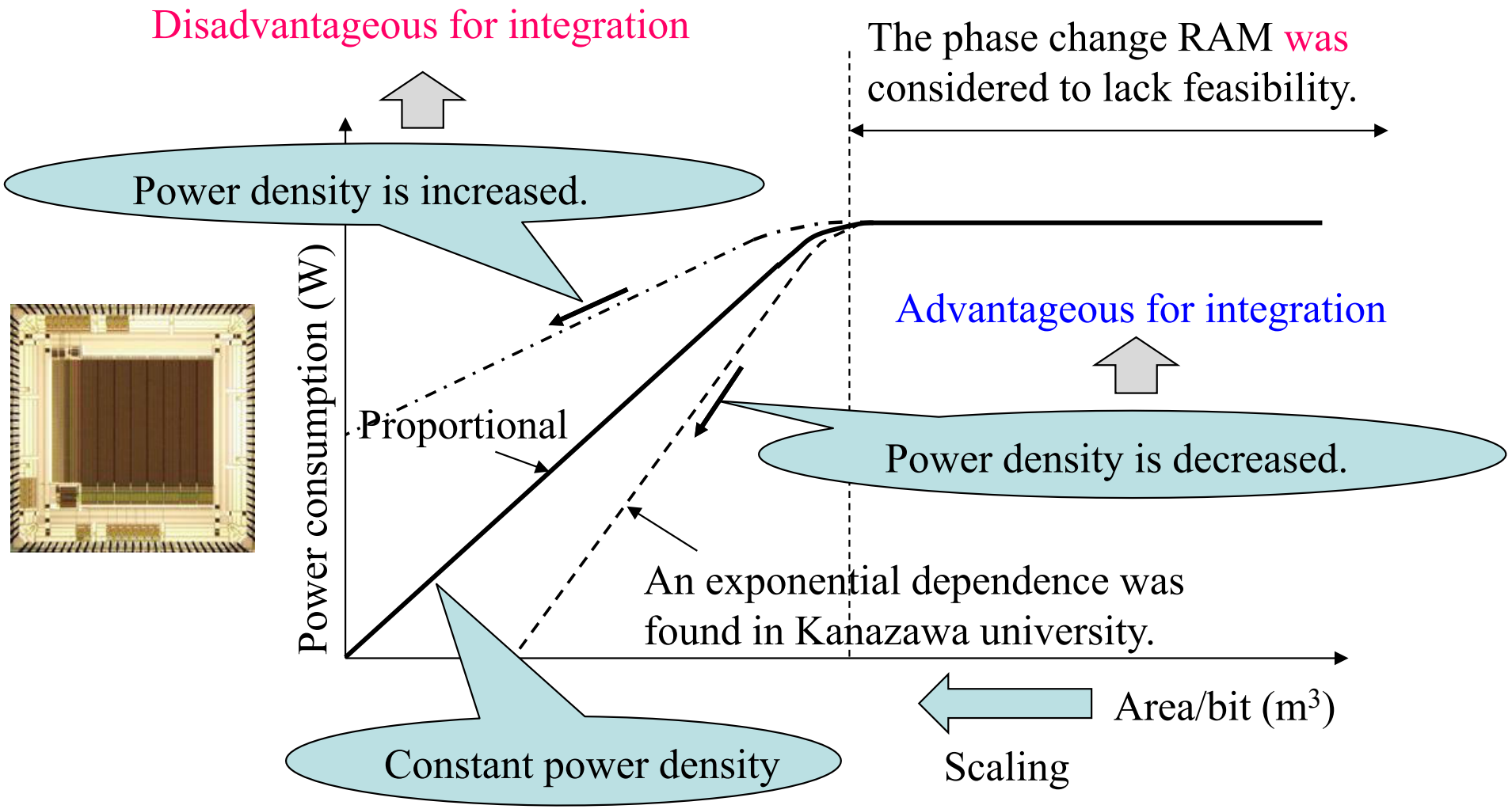
Scaling effect and performance prediction

Scaling rule (Scaling law)

By Robert Dennard, et al.

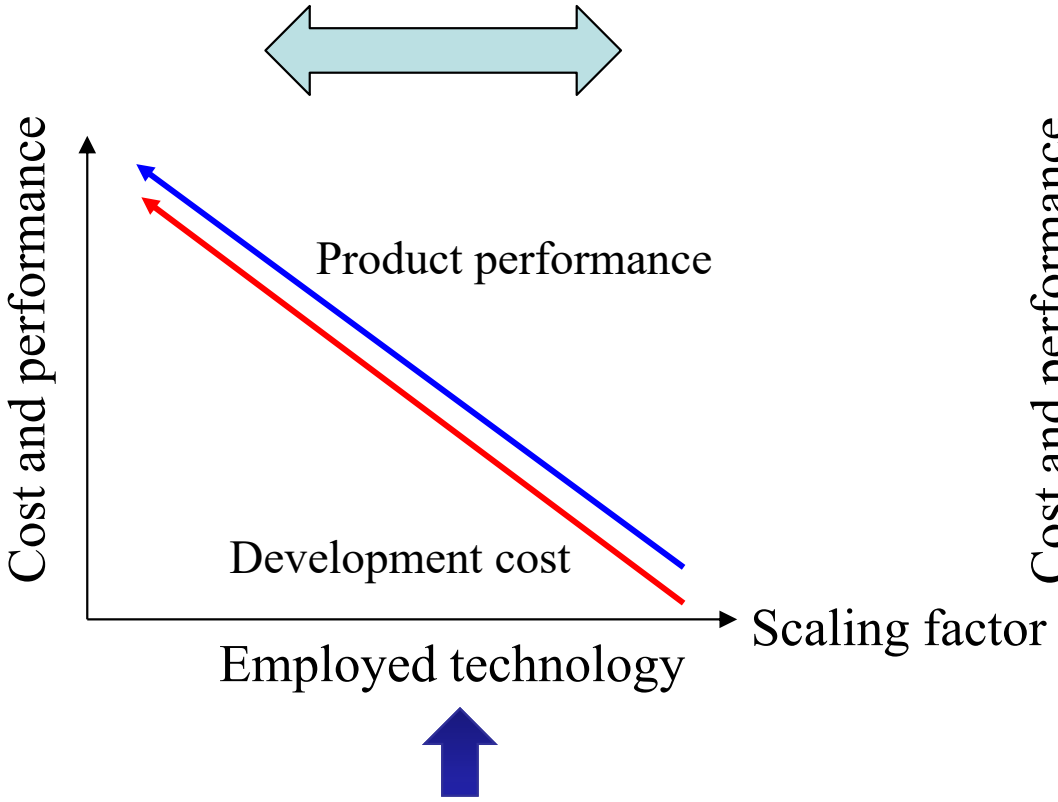
- This rule states, roughly, that as MOSFETs get smaller, the power density of MOSFET stays constant, so that the power consumption stays in proportion with area, and both voltage and current scale with length.
  1. This law can be used to estimate the circuit performance changes due to scaling.
  2. The LSI designers can choose the manufacturing technology in consideration of the balance between the circuit performance and the manufacturing cost. A state-of-the-art manufacturing technology is disadvantageous in cost.

Scaling effect of phase change RAM



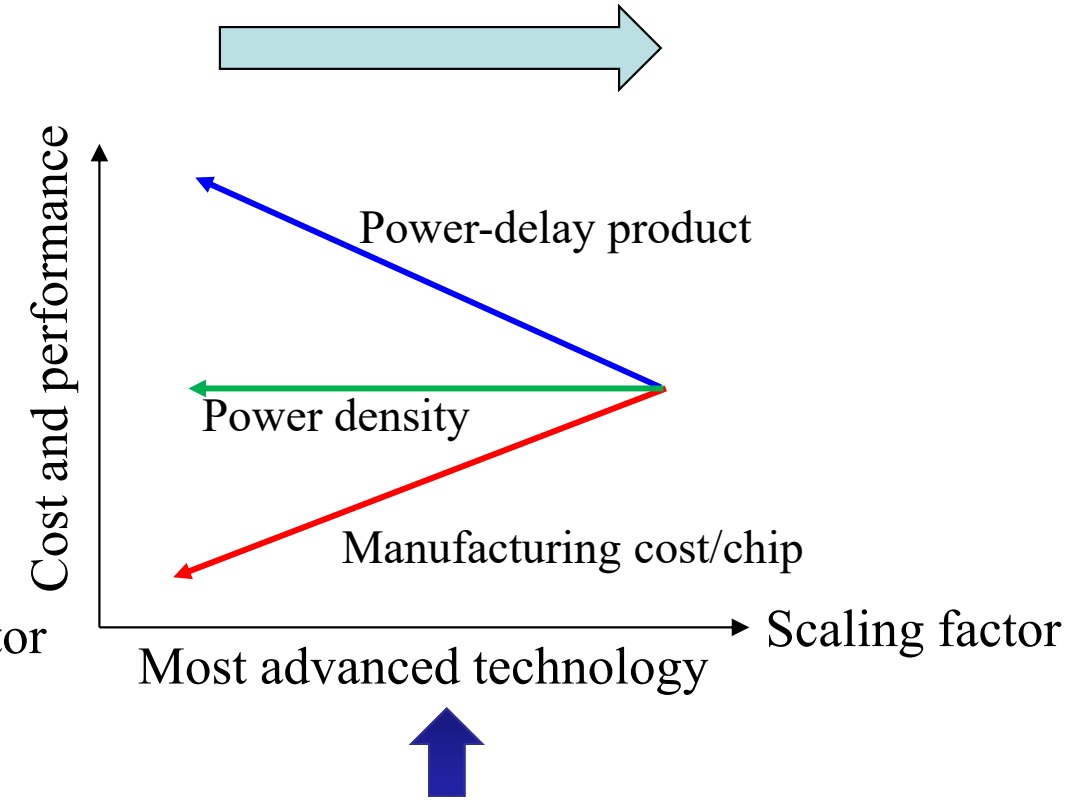
The significance of scaling law

Perfect option in current



Scaling low for designers

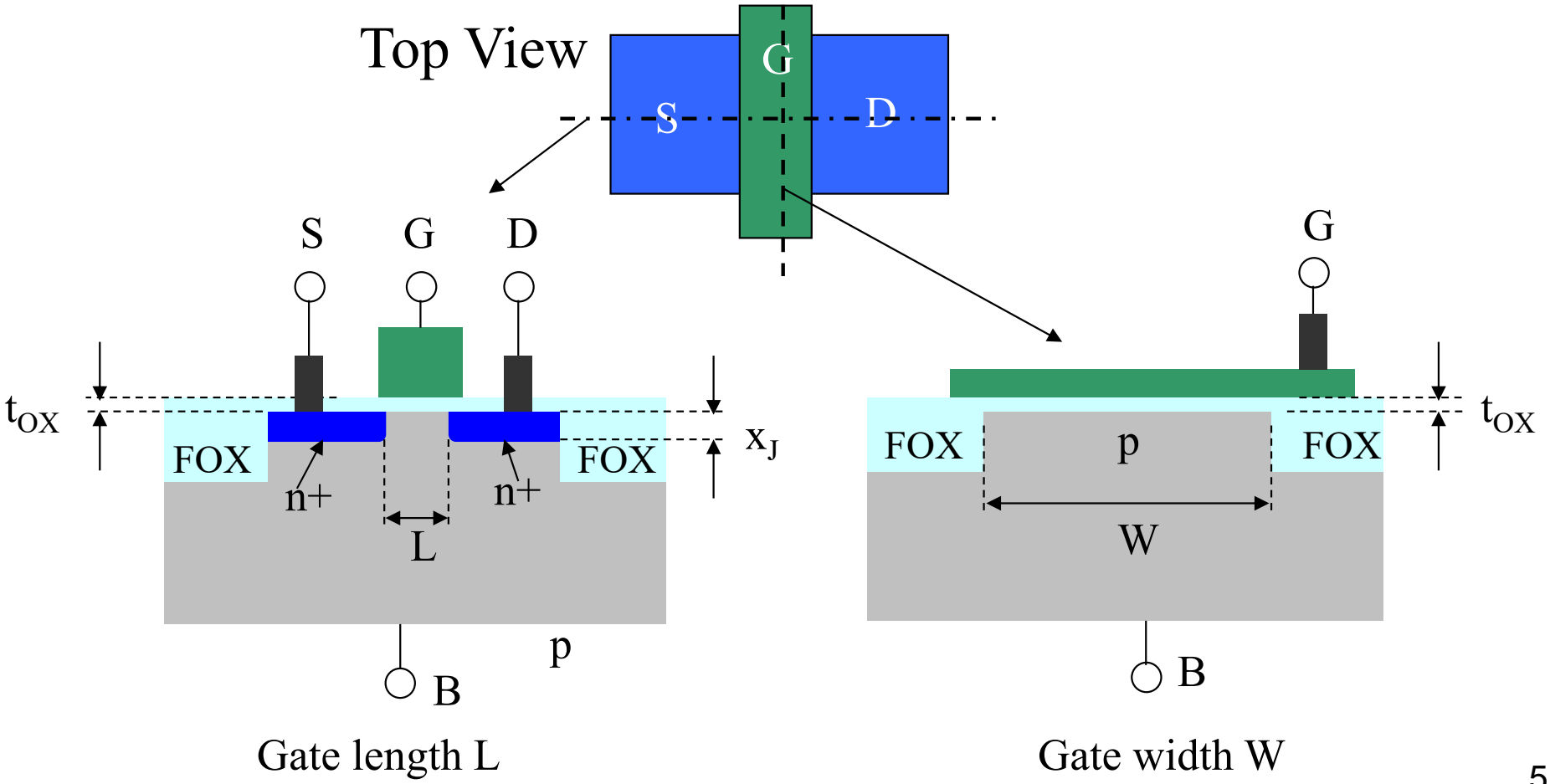
Driving force of technology development



Scaling low for manufactures

Scaling law (比例縮小則)

- The size of MOSFET is scaled down without changing the dimensional ratio.
- The electric field in the MOSFET is kept constant.



Scaling method

Physical parameters	Scaling factor	Remarks
Gate length L	$1/\alpha$	Horizontal direction (Design parameter)
Gate width W	$1/\alpha$	Horizontal direction (Design parameter)
Gate oxide thickness t_{ox}	$1/\alpha$	Vertical direction
Junction depth x_j	$1/\alpha$	Vertical direction
Impurity concentration N	α	Strictly speaking, N/(Depletion layer thickness)
Power supply voltage VDD	$1/\alpha$	To keep the electric field constant

Dennard, et al., IEEE SC-9, 256, 1974

Scaling effects 1

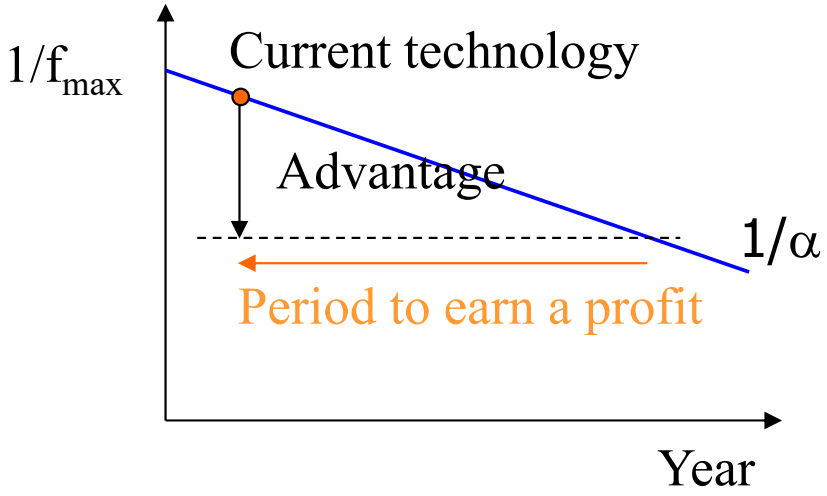
Scaling effect	Scaling factor	Remarks
Device area A	$1/\alpha^2$	
Electric field E	1	Constant electric field
Parasitic capacitance C	$1/\alpha$	$C = \varepsilon_0 \varepsilon_r \frac{(L/\alpha)\cdot(W/\alpha)}{t_{OX}/\alpha}$
Drain current I	$1/\alpha$	The details are to be mentioned later.
Threshold voltage V_T	$1/\alpha$	Adjustment required

Scaling effects 2

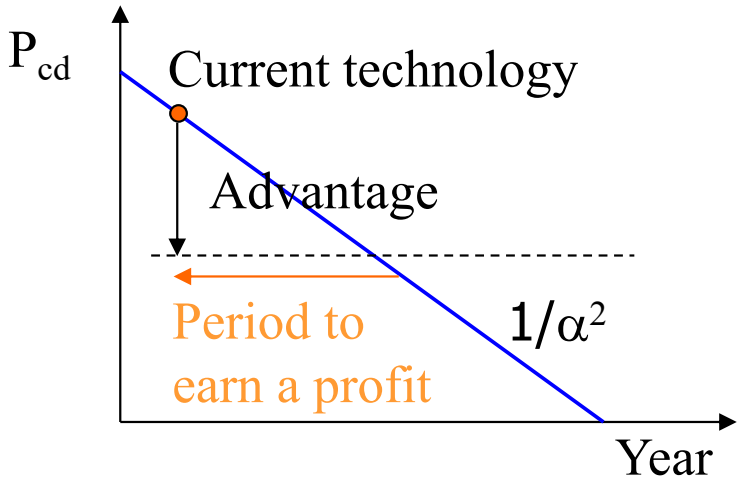
Scaling effect	Scaling factor	Remarks
Delay Time $T \propto CR=CV/I$	$1/\alpha$	Increasing the operating speed
Active power dissipation $P \propto Cf_cV^2$ (P_{cd})	$1/\alpha^2$ (max f_c) $1/\alpha^3$ (const. f_c)	Reduction of the power consumption
Static power dissipation $P \propto IV$ (P_l and P_{dp})	$1/\alpha^2$	Reduction of the power consumption
Power-Delay Product $PT \propto CV^2$	$1/\alpha^3$	
Power Density P/A Temperature rise by Joule heating	1	For maximum operating frequency

Performance prediction by scaling law

1/maximum operating frequency



Charge-discharge power consumption



Alternation rate of technology generations ＝ 0.7 times/3 years

- The performance trends of the digital systems can be estimated by the scaling law.
- The scaling down of MOSFET does not necessarily enhance the overall performance of the digital systems. The noise margin is decreased, and the static power consumption is increased.

Scaling law of drain current 1

Linear region

$$I_{dsn} = \beta_n \{(V_{gsn} - V_{tn0}) \cdot V_{dsn} - \frac{1}{2} V_{dsn}^2\}$$

⇩ Reduction

$$I_{dsn}^{'} = \beta_n^{'} \{(\frac{V_{gsn}}{\alpha} - \frac{V_{tn0}}{\alpha}) \cdot \frac{V_{dsn}}{\alpha} - \frac{1}{2} \cdot \frac{V_{dsn}^2}{\alpha^2}\}$$

$$= \frac{\beta_n^{'}}{\alpha^2} \{(V_{gsn} - V_{tn0}) \cdot V_{dsn} - \frac{1}{2} \cdot V_{dsn}^2\}$$

Capacitance of GOX per area

$$\beta_n = \mu_n C_O \frac{W_n}{L_n}$$

⇩ Reduction

$$\beta_n^{'} = \mu_n \alpha \cdot C_O \frac{W_n / \alpha}{L_n / \alpha}$$

$$= \alpha \beta_n$$

$$\therefore I_{dsn}^{'} = \frac{1}{\alpha^2} \cdot \alpha \beta_n \{(V_{gsn} - V_{tn0}) \cdot V_{dsn} - \frac{1}{2} \cdot V_{dsn}^2\} = \frac{1}{\alpha} I_{dsn}$$

Scaling law of drain current 2

Saturation region

$$I_{dsn} = \frac{\beta_n}{2}(V_{gsn} - V_{tn0})^2$$

⇩ Reduction

$$I_{dsn}^{'} = \frac{\beta_n^{'}}{2}\left(\frac{V_{gsn}}{\alpha} - \frac{V_{tn0}}{\alpha}\right)^2$$

$$= \frac{\beta_n^{'}}{2\alpha^2}(V_{gsn} - V_{tn0})^2$$

$$\beta_n = \mu_n C_O \frac{W_n}{L_n}$$

⇩ Reduction

$$\beta_n^{'} = \mu_n \alpha \cdot C_O \frac{W_n / \alpha}{L_n / \alpha}$$

$$= \alpha\beta_n$$

$$\therefore I_{dsn}^{'} = \frac{1}{2\alpha^2} \cdot \alpha\beta_n (V_{gsn} - V_{tn0})^2 = \frac{1}{\alpha} I_{dsn}$$

Scaling law of threshold voltage

The scaling law of threshold voltage is approximately equivalent to that of voltages in the circuit.

$$V_{tn}(V_{sub}) = V_{FB} + 2\phi_B + \frac{1}{C_O}\sqrt{2\varepsilon_0\varepsilon_{Si}qN_A \cdot (2\phi_B - V_{sub})}$$

V_{FB} (Flat-band voltage) - A physical property of metal and semiconductor
ϕ_B (Bulk potential) - A function of impurity concentration
The 3rd term is a function of V_{sub} (= V_{BS})

⇩ Reduction

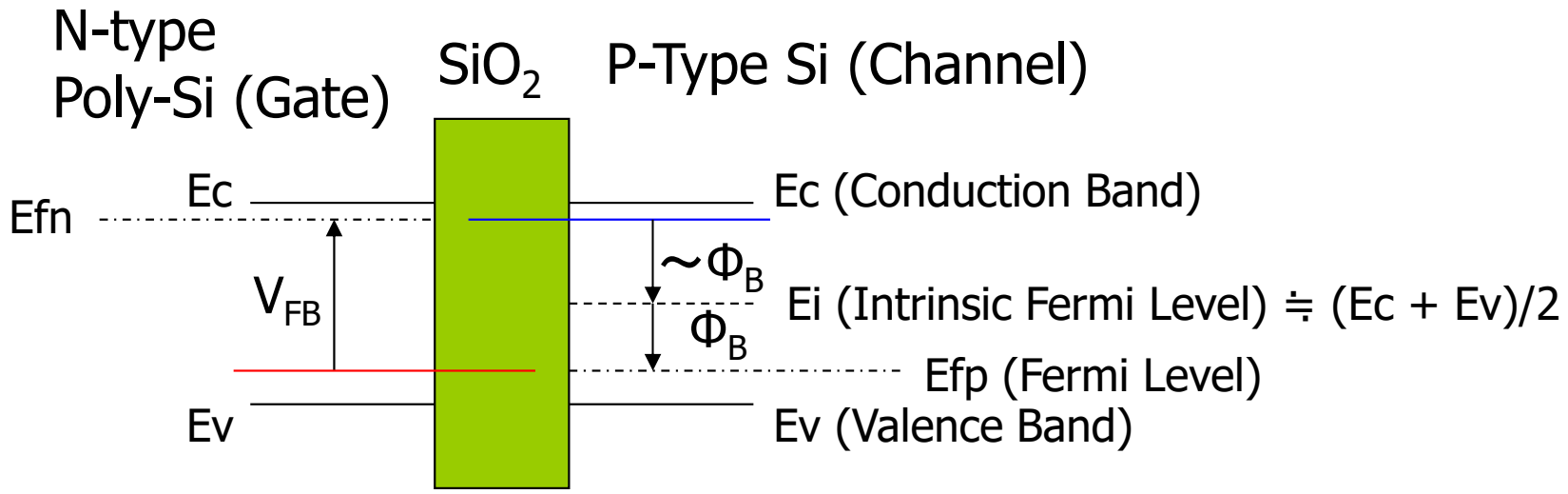
$$V_{tn}'(V_{sub}) = V_{FB} + 2\phi_B + \frac{1}{\alpha \cdot C_O}\sqrt{2\varepsilon_0\varepsilon_{Si}q(\alpha \cdot N_A) \cdot (2\phi_B - \frac{V_{sub}}{\alpha})}$$

$$= V_{FB} + 2\phi_B + \frac{1}{\alpha \cdot C_O}\sqrt{2\varepsilon_0\varepsilon_{Si}qN_A \cdot (2\alpha\phi_B - V_{sub})}$$

Approximations

$$\begin{cases} V_{FB} + 2\phi_B \cong 0 \\ \phi_B' \cong \phi_B \end{cases} \rightarrow \boxed{V_{tn0}' \cong \frac{V_{tn0}}{\sqrt{\alpha}} \qquad (V_{sub} = 0)}$$

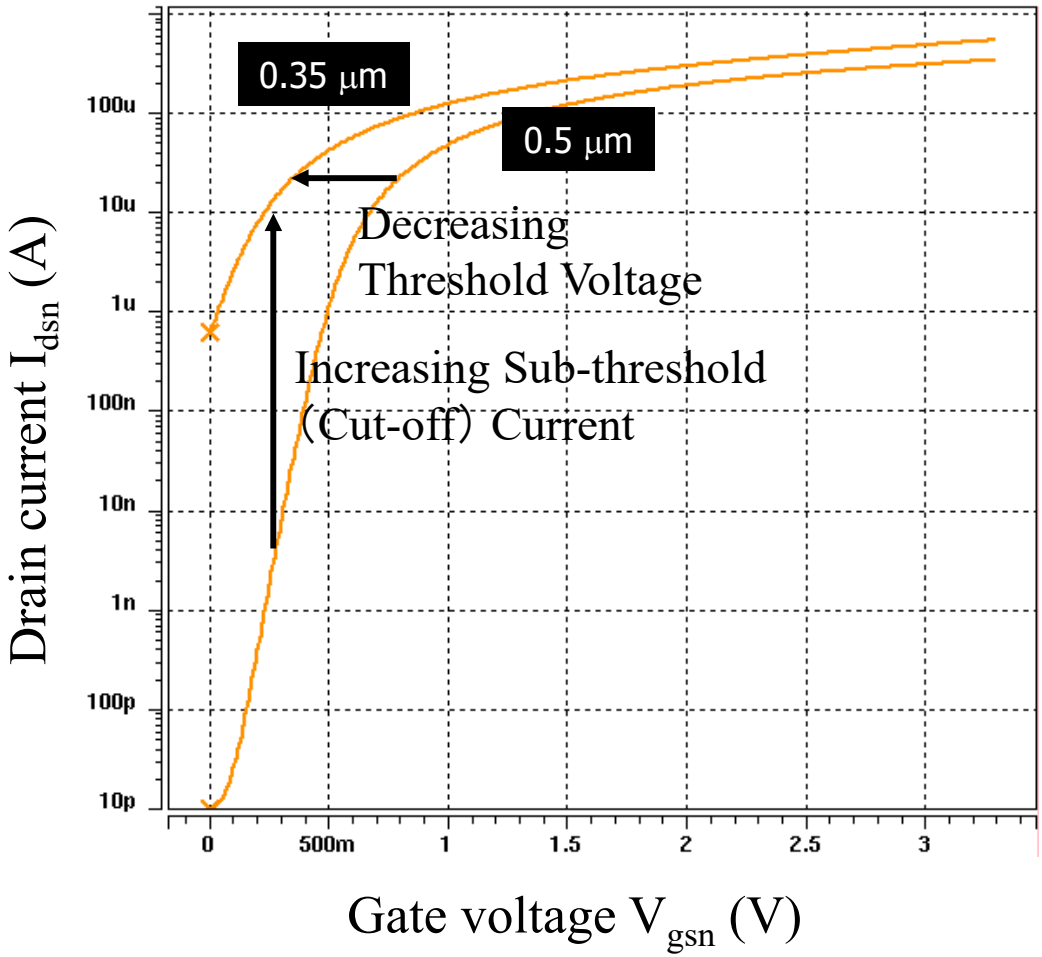
Band-structure lineup at poly-Si gate or polycide gate structure.



When N_A (in the channel) = N_D (in the gate electrode), $V_{FB} \cong -2\phi_B$

Increased subthreshold current

- The subthreshold current of MOSFET increases with scaling down.
- The reduction of static power consumption is becoming an important issue.
- The threshold voltage is proportional to $\alpha^{-0.5}$, but the power supply voltage is proportional to α^{-1}.



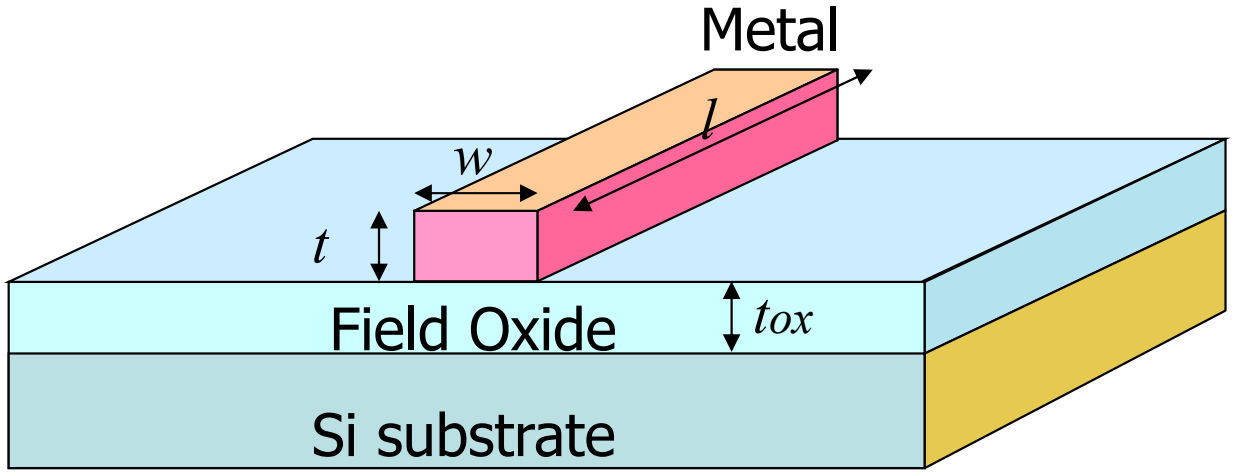
Scaling method of interconnects

Physical parameters	Scaling factor	Remarks
Metal Length l	1/α	Horizontal direction (Design parameter)
Metal Width w	1/α	Horizontal direction (Design parameter)
Metal Thickness t	1/α	Vertical direction

Figure of merit

- Transmission speed
- Power consumption
- Durability for large current

Scaling effect of interconnect 1



Resistance

$$R_{\ln} = \rho \frac{l}{t \cdot w}$$

⇩ Reduction

$$R'_{\ln} = \rho \frac{\frac{l}{\alpha}}{\frac{t}{\alpha} \cdot \frac{w}{\alpha}} = \alpha \cdot R_{\ln}$$

Capacitance

$$C_{\ln} = \varepsilon_0 \varepsilon_{ox} \frac{l \cdot w}{t_{ox}}$$

⇩ Reduction

$$C'_{\ln} = \varepsilon_0 \varepsilon_{ox} \frac{\frac{l}{\alpha} \cdot \frac{w}{\alpha}}{\frac{t_{ox}}{\alpha}} = \frac{C_{\ln}}{\alpha}$$

Scaling effect of interconnect 2

Scaling effect	Scaling factor	Remarks
Resistance R_{ln}	α	
Capacitance C_{ln}	$1/\alpha$	
Delay Time $T_d = C_{ln} \cdot R_{ln}$	1	
Voltage Drop V_{ln}	1	
Current Density J_{ln}	α	

Note: The scaling of interconnects does not theoretically affect the delay time, however, the actual delay time of the interconnects is increased, because the fringe capacitance increases with increasing the number of wires.

Constant voltage scaling

Another scaling factor κ is introduced for the constant voltage scaling. The reduction of the power supply voltage and the threshold voltage degrade the noise margin and subthreshold leakage.

Physical parameters	Scaling factor	Remarks
Power Supply Voltage VDD	$1/\kappa$	Horizontal direction (Design parameter)
L, W, tox, xj	$1/\alpha$	Horizontal direction (Design parameter)
Impurity Concentration N_D, N_A	α^2/κ	Vertical direction

Scaling of potential and electric field

Poisson's equation

$$\frac{d^2V}{dx^2} = -\frac{qN_{A,D}}{\varepsilon_0\varepsilon_{Si}}$$

⇩ Reduction

$$\frac{d^2V'/\kappa}{dx^2/\alpha^2} = -\frac{q}{\varepsilon_0\varepsilon_{Si}} \cdot \frac{\alpha^2}{\kappa} \cdot N_{D,A}$$

$$\frac{d^2V'}{dx^2} = -\frac{q}{\varepsilon_0\varepsilon_{Si}} \cdot N_{D,A} = \frac{d^2V}{dx^2}$$

Constant voltage

Gauss's law

$$\frac{dE}{dx} = \frac{qN_{A,D}}{\varepsilon_0\varepsilon_{Si}}$$

⇩ Reduction

$$\frac{dE'}{dx/\alpha} = \frac{q}{\varepsilon_0\varepsilon_{Si}} \cdot \frac{\alpha^2}{\kappa} \cdot N_{A,D}$$

$$\frac{dE'}{dx} = \frac{q}{\varepsilon_0\varepsilon_{Si}} \cdot \frac{\alpha}{\kappa} \cdot N_{A,D} = \frac{\alpha}{\kappa}\frac{dE}{dx}$$

The electric field is proportional to α/κ.

As LSI technology becomes more advanced…

1. R&D Platform in medical engineering, biotechnology, environmental engineering, space development, etc.
2. Autonomous system, dynamic reconfigurable system, AI processor over 100 million gates
3. Sensors, Rader and ultra-high speed communication systems with millimetric and submillimetric waves
4. Micro-total analysis systems and metamaterials with nanostructures
5. Ultra-low power system that can operate with environmental energy and energy harvesting circuits