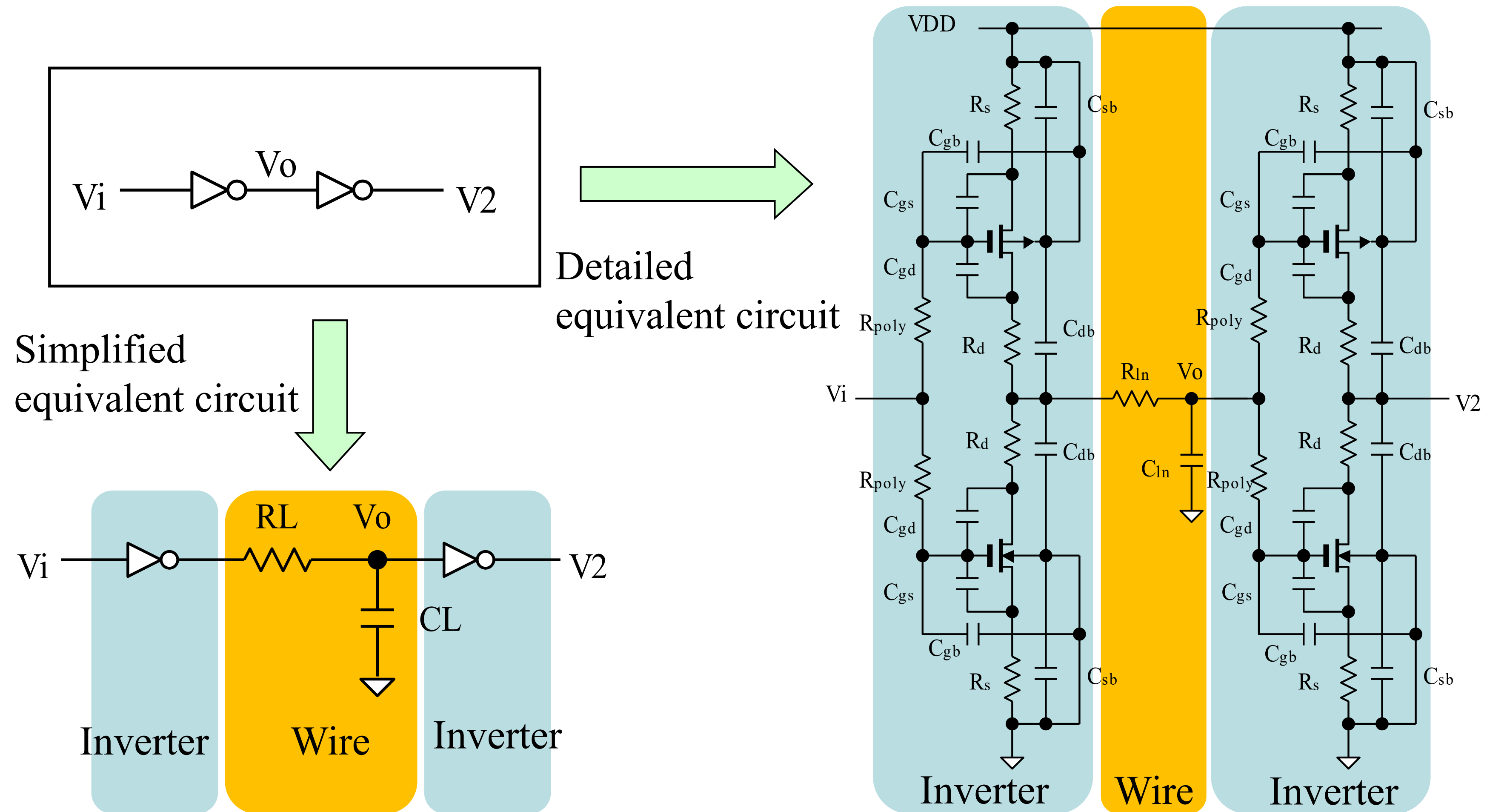
6.1 Operating speed

Analysis of delay time and clock frequency

6.1.1 Circuit characteristics considering parasitic elements

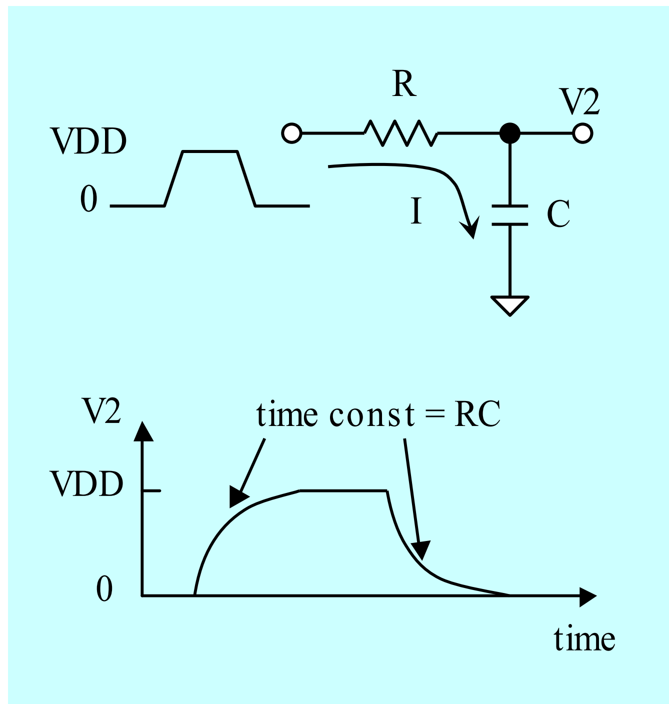
Parasitic elements in LSI

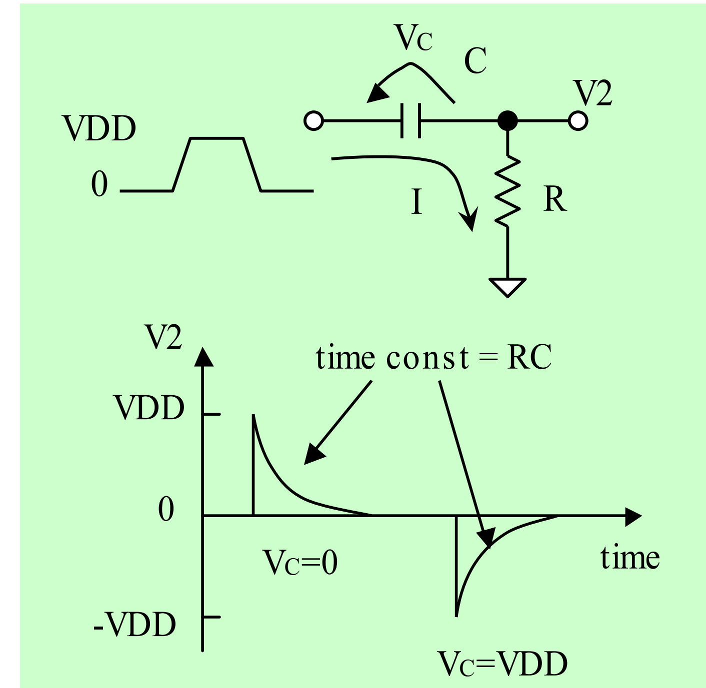


Note: A MOSFET model includes the parasitic elements in the layout pattern.

Transient response of parasitic elements

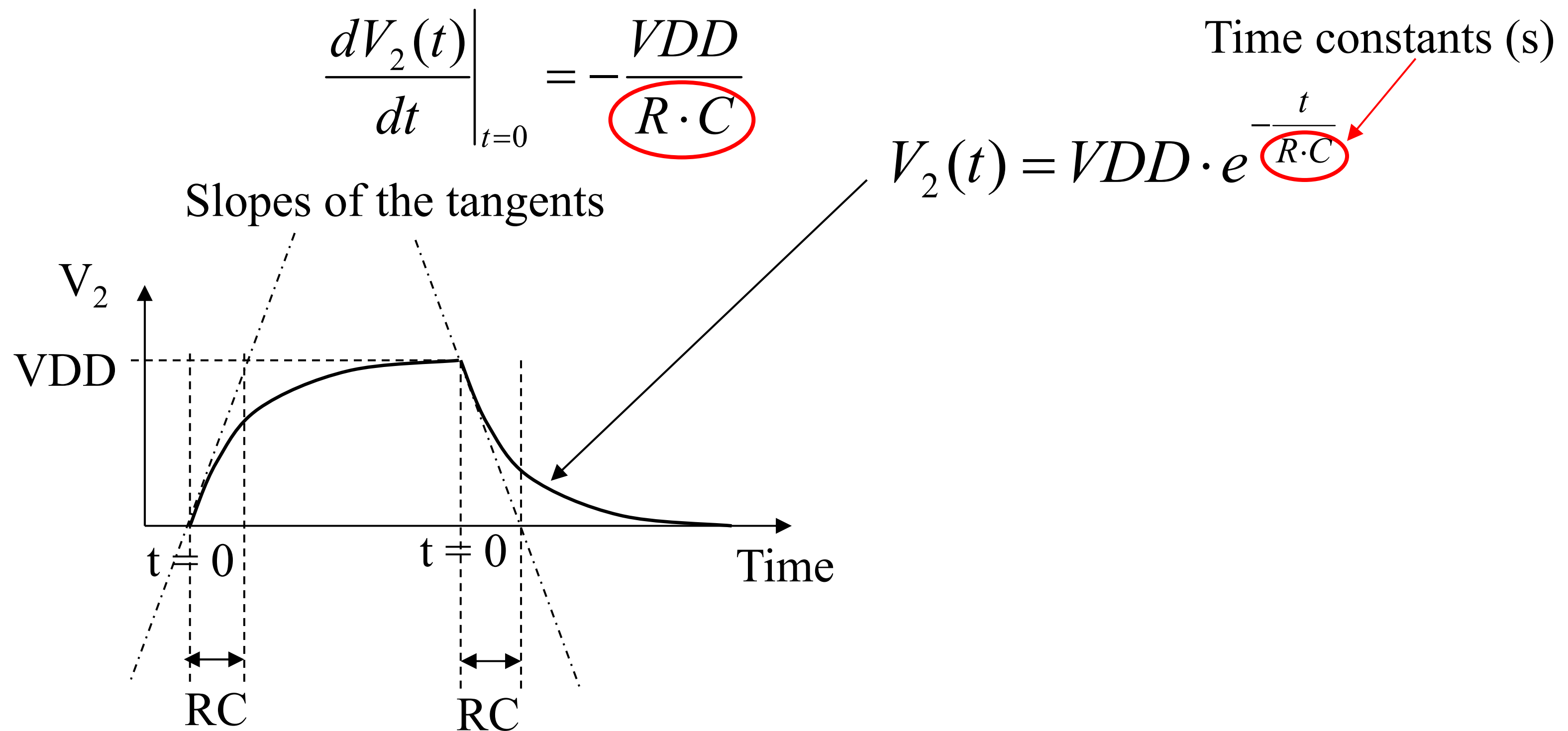
- A wire in LSI is modeled by parasitic resistances and parasitic capacitances.
  - Parasitic R and C depend on the length and the width of the wire.
  - A parasitic L must be considered in RF circuits.





Time constant

- The slopes of the tangents are increased with increasing the time constant RC. The R and C are the resistance and capacitance of the parasitic elements in the logic circuits.

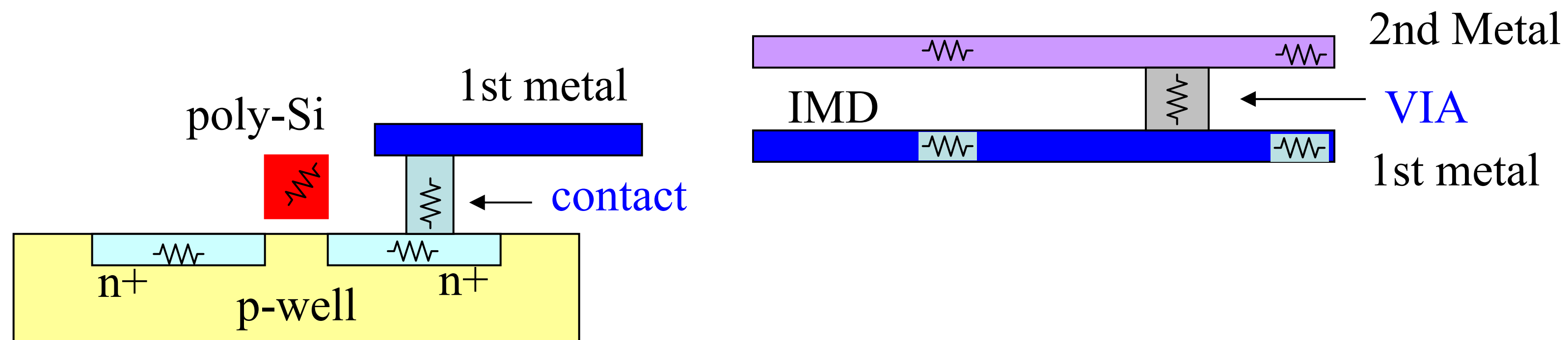


6.1.2 Extraction of a parasitic resistance

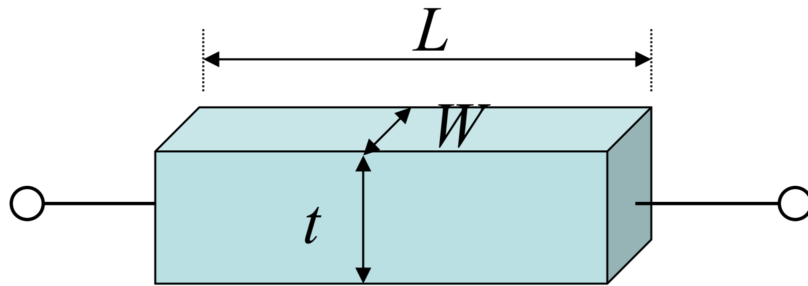
The cause of the parasitic resistance

Parasitic resistance

- Metal layer
- VIA
- Poly Si
- n-type Si layer and p-type Si layer
- Contact



Estimation of resistance



$$R = \frac{\rho}{t}\frac{L}{W} = R_S \frac{L}{W}$$

ρ (Wm) Resistivity

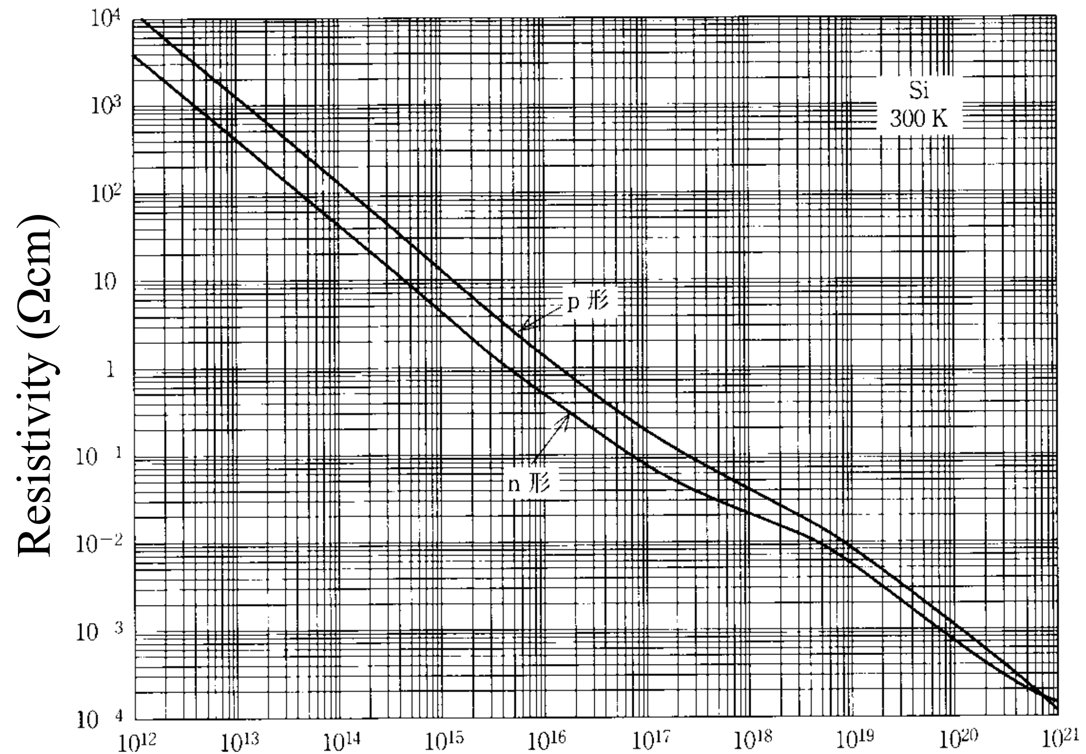
R_S (Ω/□) Sheet resistance

- The sheet resistance R_S of each layer is disclosed by the manufacture.
- The resistance R is estimated from the values of L and W in layout pattern.
- The measurement of Rs is easier than the measurement of ρ.

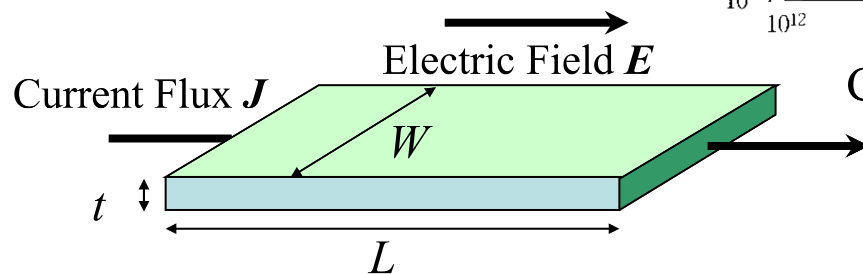
Resistivity of Si

$$E(\mathrm{V/m}) = \rho(\Omega\mathrm{m})J(\mathrm{A/m^2})$$

The resistivity of Si depends on the donor or acceptor concentration.
The concentration dependence is called "Irvin curve".

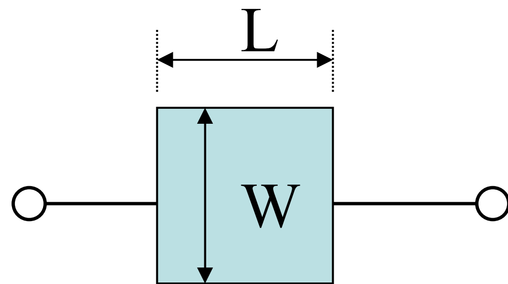


Concentration of donors or acceptors (cm^{-3})



Measurement of sheet resistance

Top view



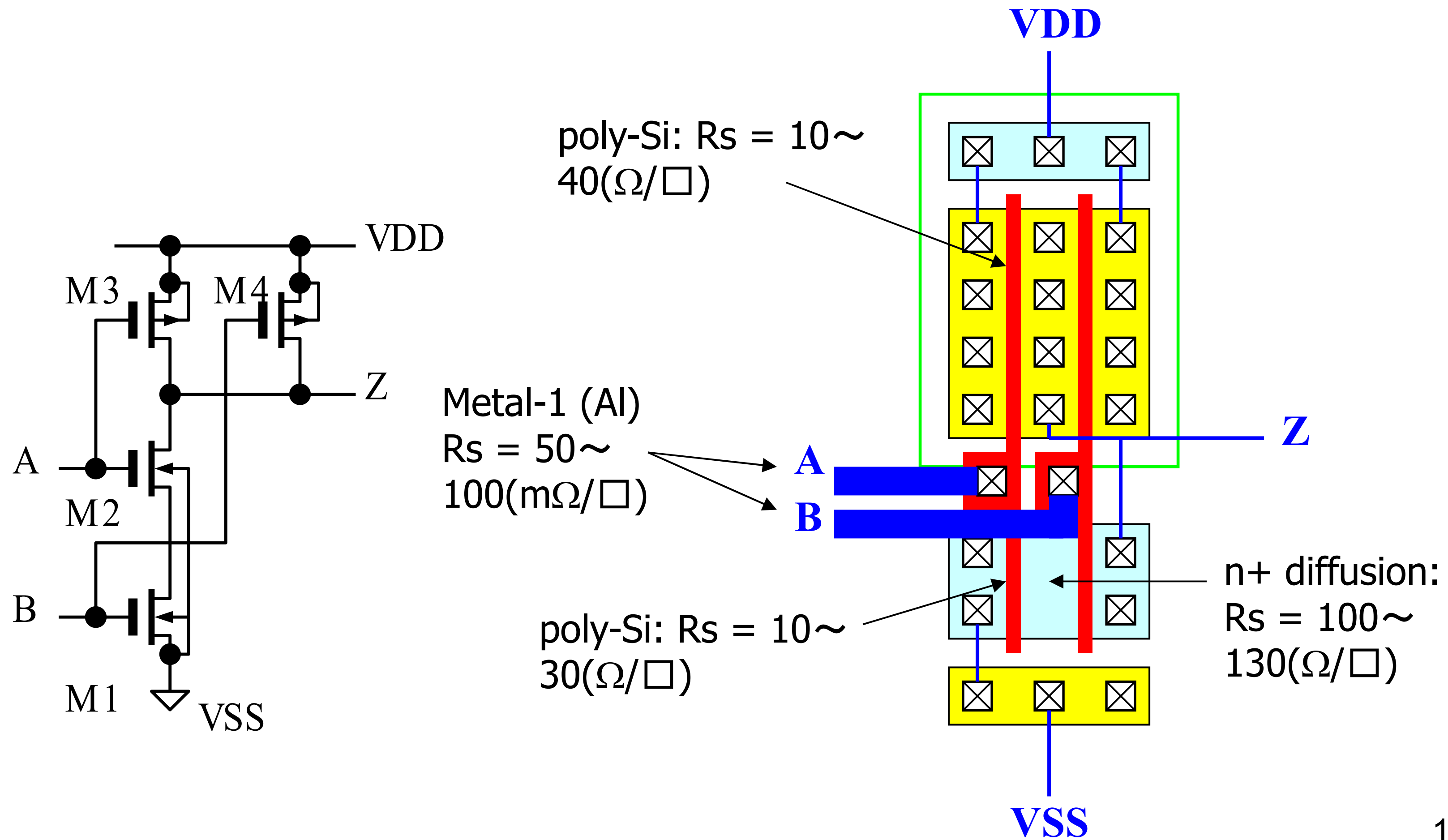
$$R = R_S \frac{L}{W} = R_S \text{ , when } L = W$$

The sheet resistance is measured by measuring the resistance at both sides.

[Note] The definition of the sheet resistance is different from the resistance, but the unit of Rs is the same as the unit of resistance (Ω). An unit Ω/☐ is used in order to distinguish the sheet resistance from the resistance.

Examples of sheet resistance

- Parasitic resistance in 2-input NAND



6.1.3 Extraction of a parasitic capacitance

The cause of the parasitic capacitance

Parasitic capacitances in MOSFET

1. Gate oxide capacitance

$$C_{gc} = (L - L_{OV})WC_{OX}$$

$$C_{OX} = \varepsilon_{SiO_s}\varepsilon_0 \frac{1}{t_{ox}}$$

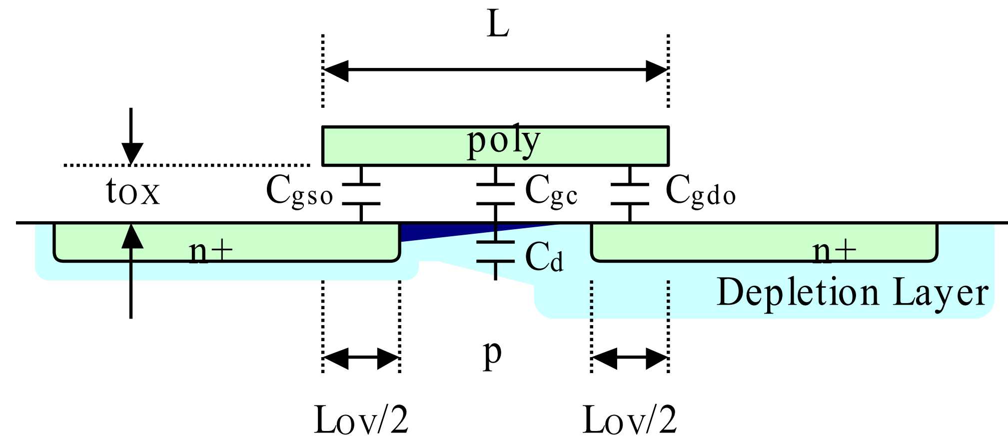
2. Overlap capacitance

$$C_{gso} = C_{gdo} = \frac{L_{OV}}{2}WC_{OX}$$

3. Depletion layer capacitance

$$C_d = \sqrt{\frac{qN_A\varepsilon_{SiO_s}\varepsilon_0}{2\phi_S}}$$

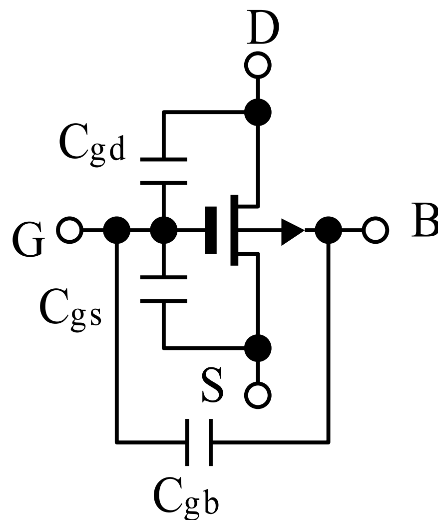
ϕ_S and Cd depend on the V_{gs}.



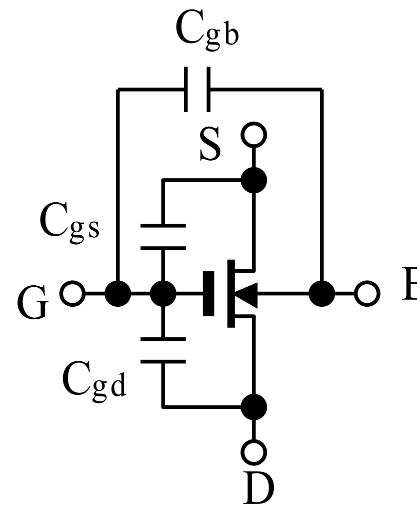
N_A: Impurity concentration in channel
ϕ_S: Surface potential of MOS interface

Inter-electrode capacitance

The parasitic capacitances C_{gc}, C_{gso}, C_{gdo}, C_d
cause the equivalent inter-electrode capacitances $C_{gs}(V)$, $C_{gd}(V)$, $C_{gb}(V)$.



n-ch MOSFET

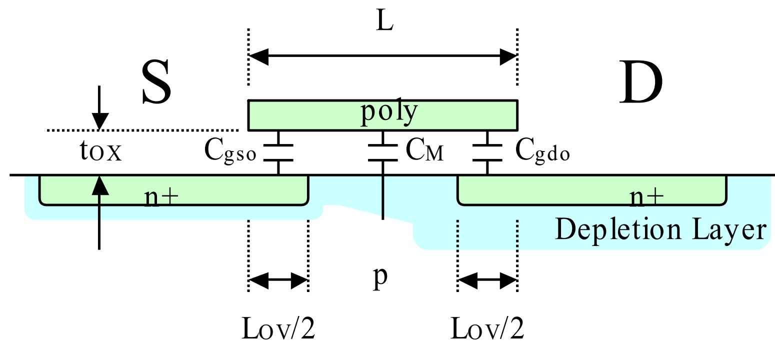


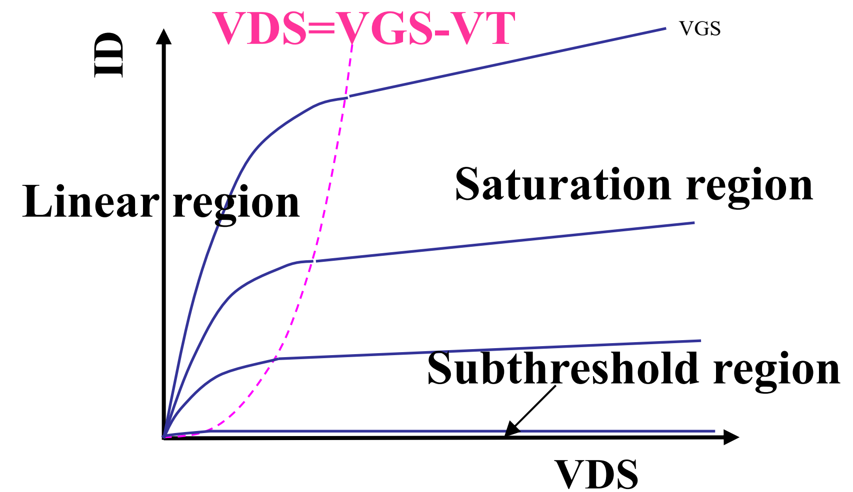
p-ch MOSFET

[Note] The inter-electrode capacitances are modeled in MOSFET device model of circuit simulator.

Capacitance model in subthreshold region

When $V_{gsn} < V_{tn0}$,





There is no channel.

$$C_{gs} = \frac{L_{OV}}{2} W \cdot C_{OX}$$

$$C_{gd} = \frac{L_{OV}}{2} W \cdot C_{OX}$$

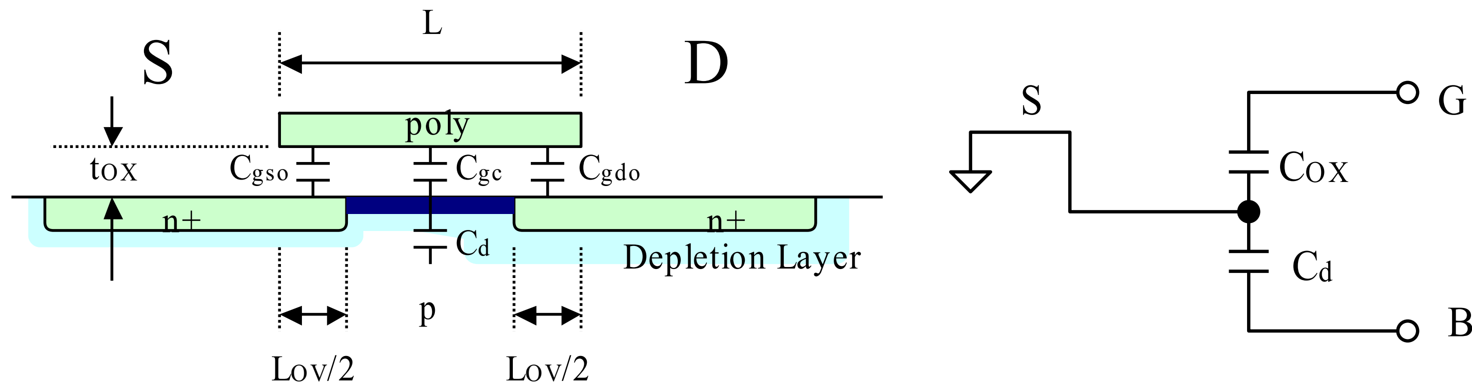
Overlap capacitance

$$C_{gb} = C_M = \frac{C_{OX} \cdot Cd}{C_{OX} + Cd}$$

Series connection of a gate oxide capacitance and a depletion layer capacitance

Capacitance model in linear region

When $V_{gsn} - V_{tn0} > V_{dsn}$,



The channel is distributed in evenly.

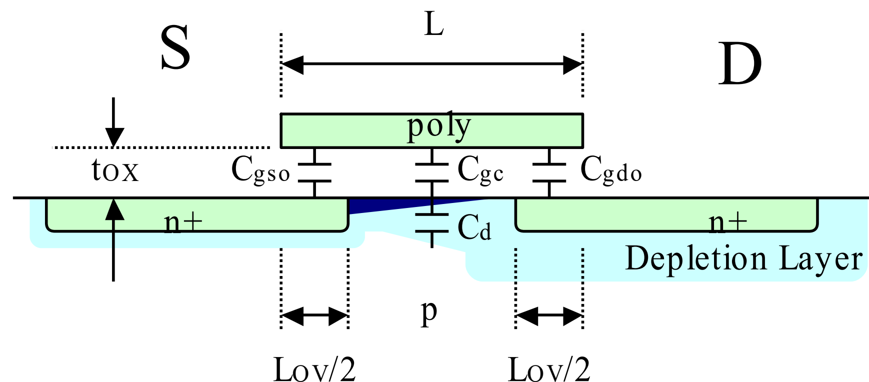
$$C_{gs} = C_{gd} = C_{gso} + \frac{C_{gc}}{2} = \frac{L_{OV}}{2} W \cdot C_{OX} + \frac{1}{2}(L - L_{OV}) \cdot W \cdot C_{OX}$$

$$C_{gb} = \frac{C_{OX} \cdot Cd}{C_{OX} + Cd} \approx 0$$

Each one half of C_{gc} is assigned to the source electrode and the drain electrode.

Capacitance model in saturation region

When $V_{gsn} - V_{tn0} < V_{dsn}$,



- The $3/2C_{gc}$ is assigned to the source electrode.
- The C_{gc} is not connected to the drain electrode due to the pinch-off effect.

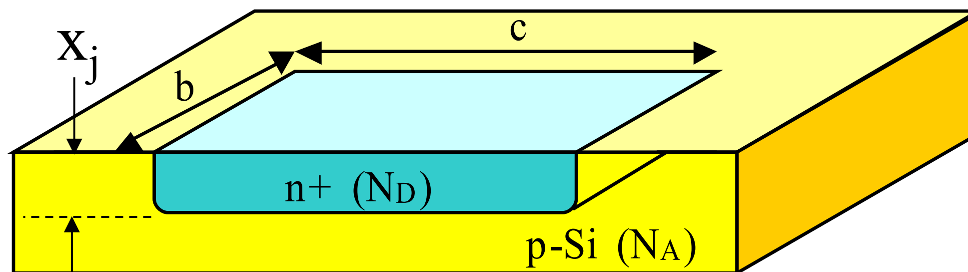
The channel is distributed in evenly.

$$C_{gs} = C_{gso} + \frac{2}{3}C_{gc} = \frac{L_{OV}}{2}W \cdot C_{OX} + \frac{2}{3}(L - L_{OV}) \cdot W \cdot C_{OX}$$

$$C_{gd} = C_{gdo} = \frac{L_{OV}}{2}W \cdot C_{OX}$$

$$C_{gb} \approx 0$$

Junction capacitance



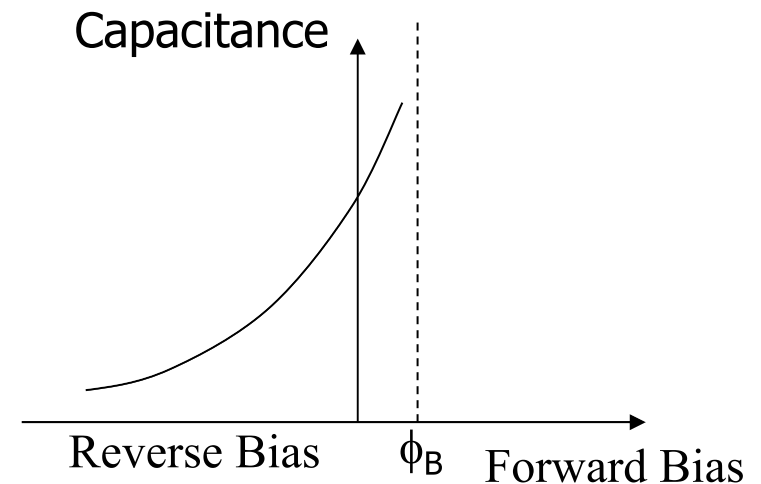
Junction capacitance

$$C_j = \{b \cdot c + 2(b+c) \cdot x_j\} \cdot C_{jo}$$

Junction capacitance per unit area

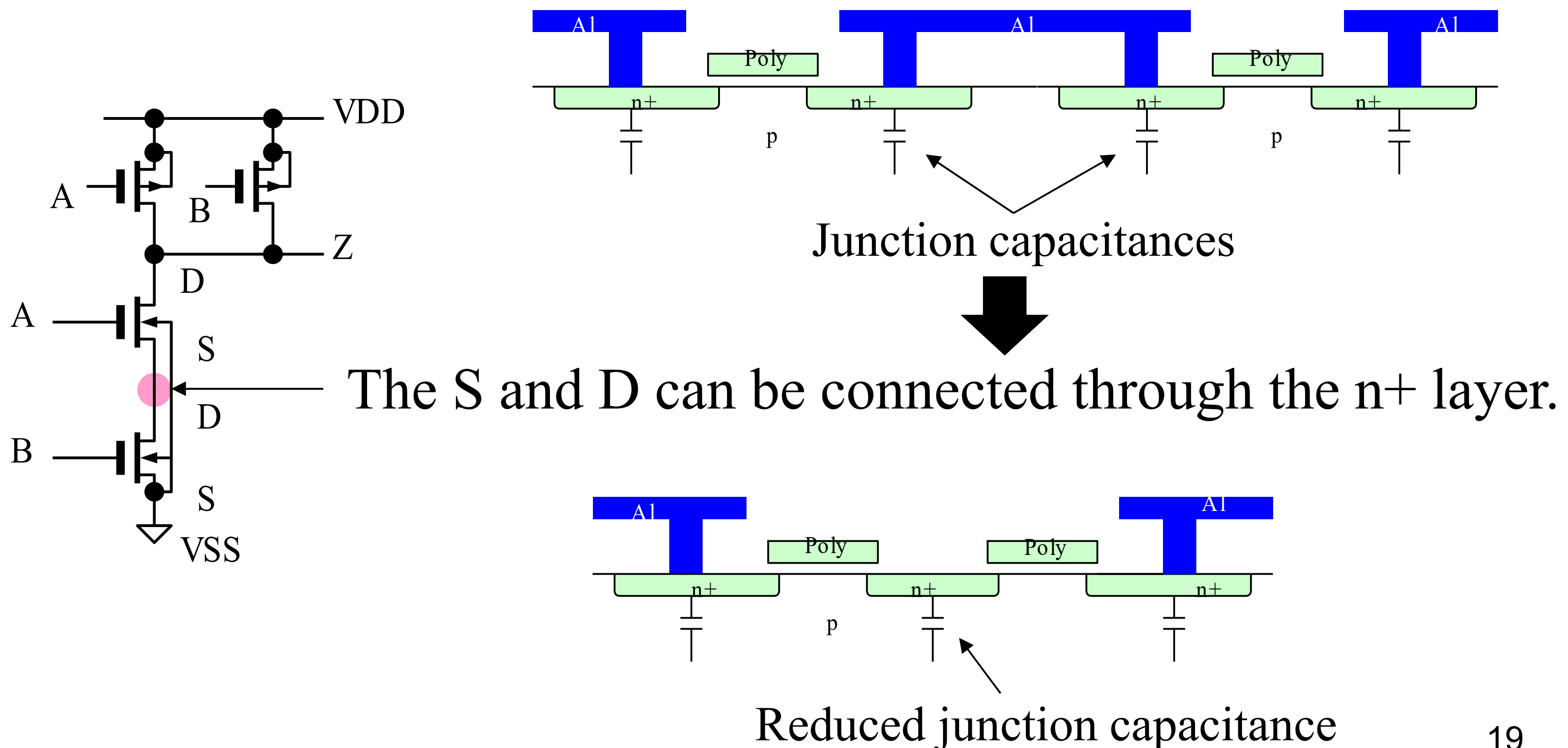
$$C_{jo} = \sqrt{\frac{q\varepsilon_0\varepsilon_{Si}N_A N_D}{2\phi_B(N_A+N_D)} \frac{1}{(1-V_{pn}/\phi_B)}}$$

ϕ_B: Built-in Potential ≒ 0.6V

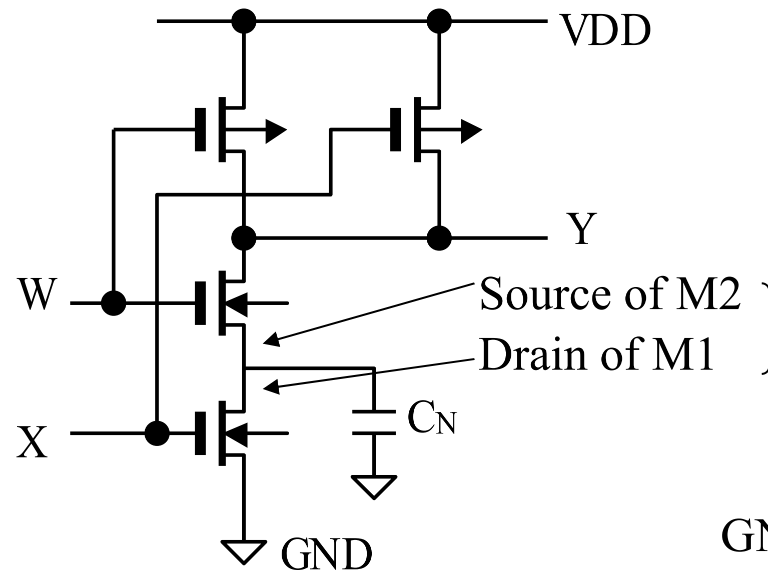


Reduction of junction capacitance

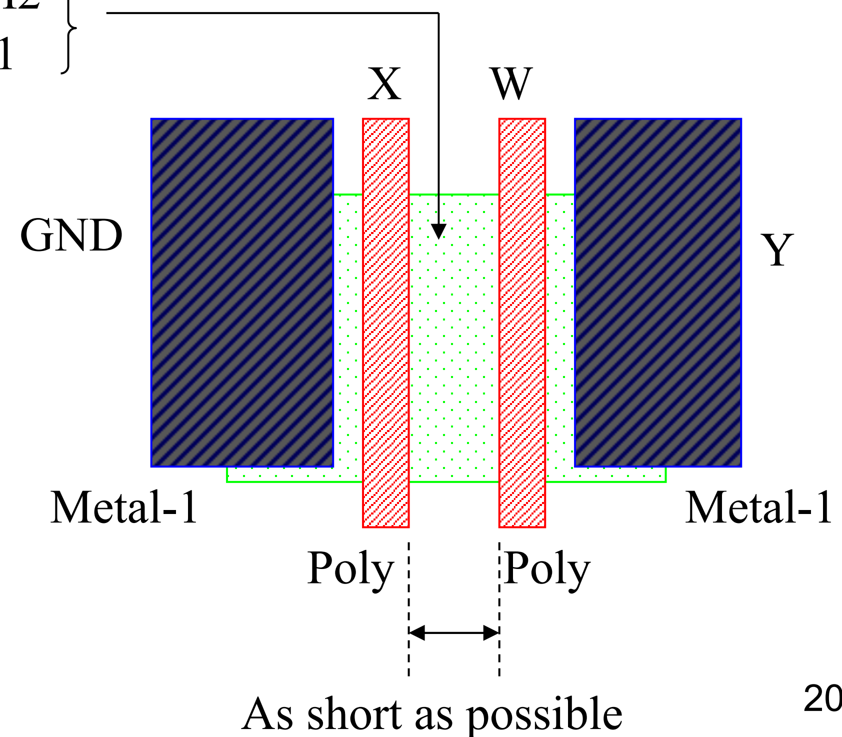
The junction capacitance causes the load of MOSFET. The junction capacitance can be reduced by the direct connection of MOSFETs.



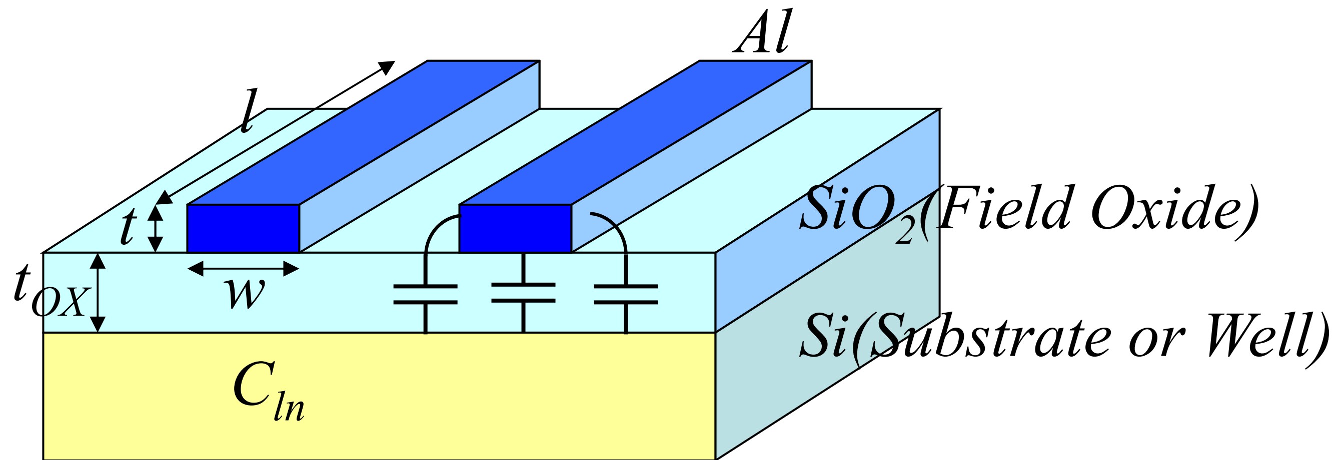
Layout example of direct connection



The parasitic capacitance CN is reduced by reducing the active area between the gates X and W.



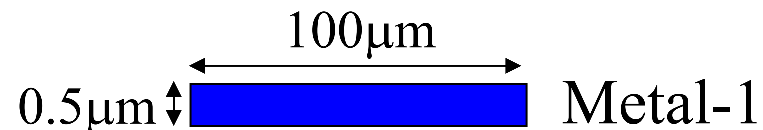
Parasitic capacitance in wiring layer



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

C_{FOX}: Parasitic capacitance per unit area

Numerical example

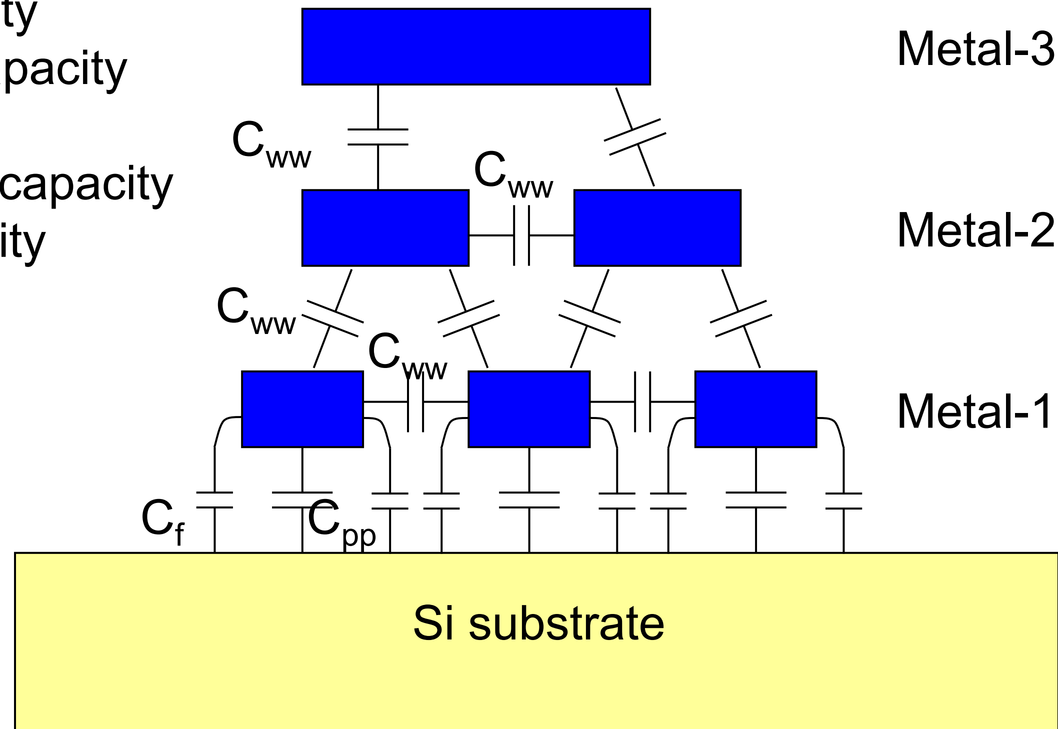


Parasitic resistance $R_{ln} = 100\text{m}\Omega \cdot \frac{100\mu\text{m}}{0.5\mu\text{m}} = 20\Omega$

Parasitic capacitance $C_{ln} = 200\text{fF/mm} \cdot 0.1\text{mm} = 20\text{fF}$

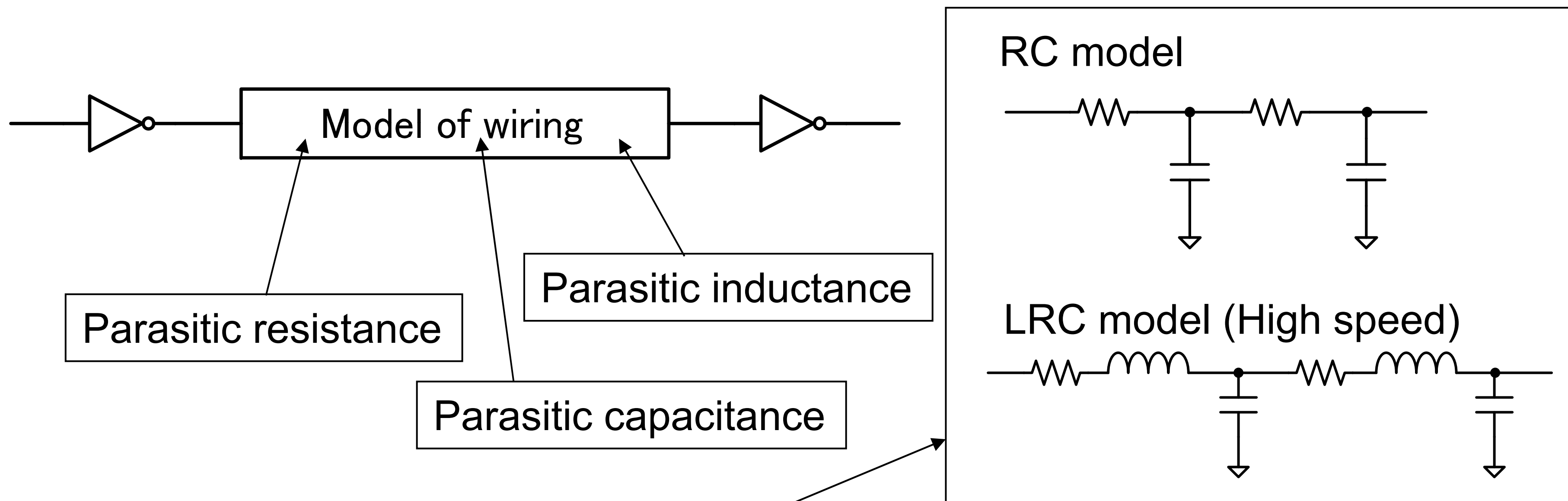
Parasitic capacitance in multilayer wiring

C_{ww}: Metal-metal capacity
C_s : Metal-substrate capacity
$C_s = C_{pp} + \Sigma C_f$
C_{pp} : Parallel plate capacity
C_f : Fringe capacity



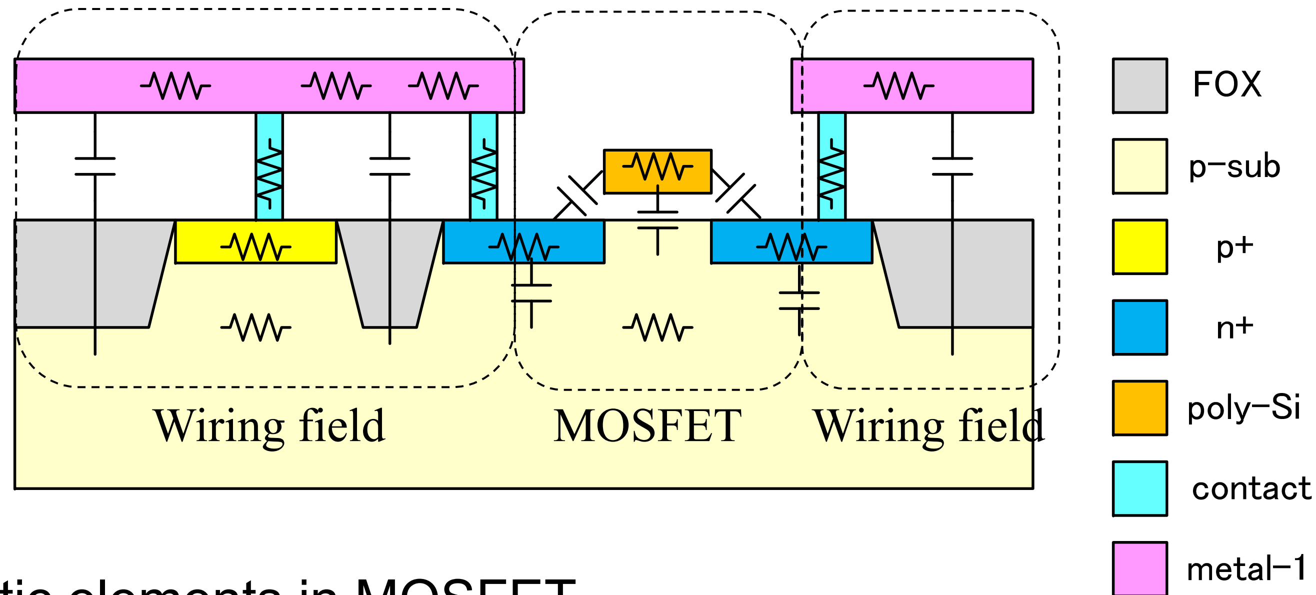
Equivalent circuit of wiring

The equivalent circuit model of wiring is used for the detailed estimation of the delay time requires



The long wiring, such as BUS, clock line, Word line and Bit line in memory array, can be modeled by using the RC or LRC ladder network.

Summary of parasitic devices

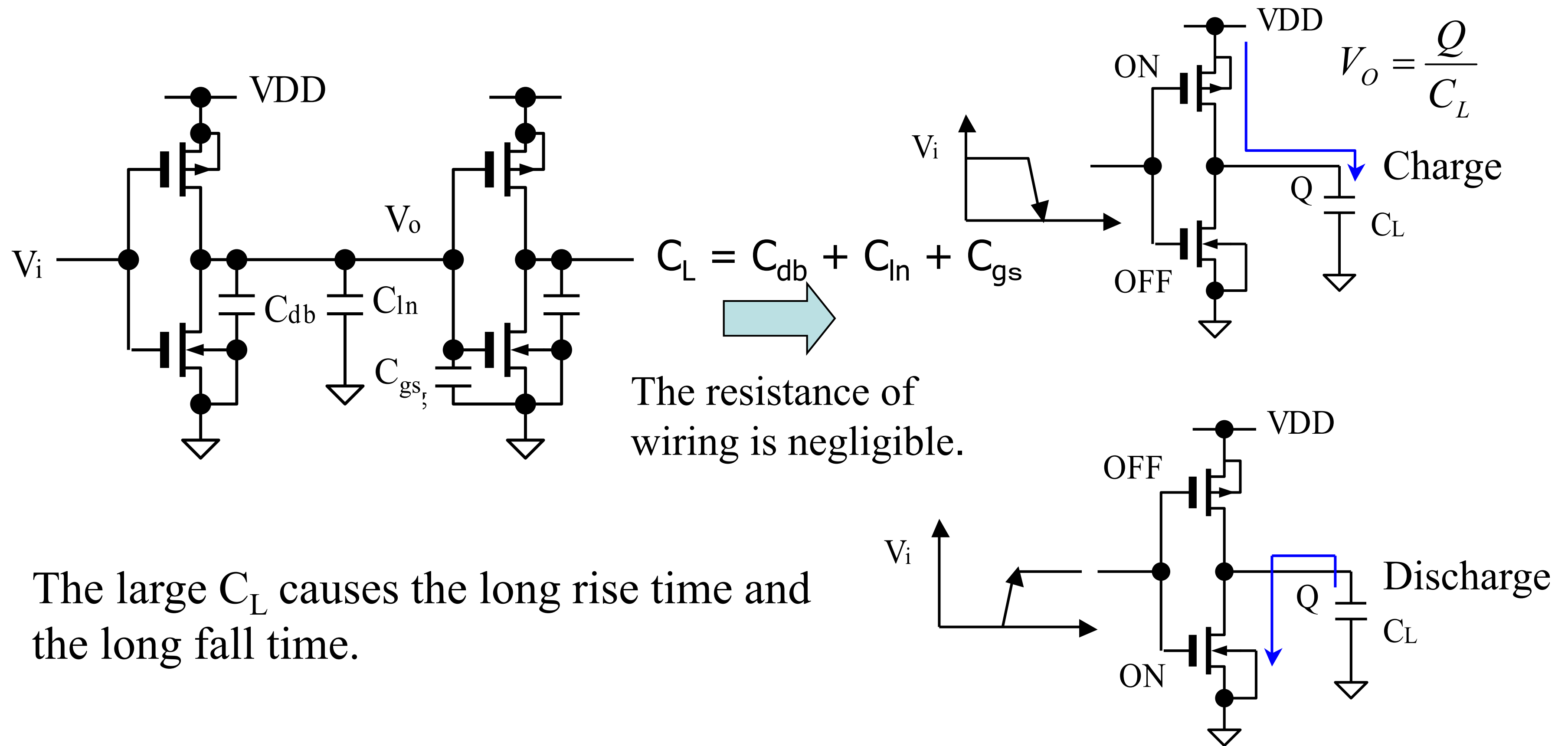


- Parasitic elements in MOSFET
  - is modeled in the MOSFET model. The parasitic capacitances and the parasitic resistances are reflected on the circuit simulation by using the size parameters of MOSFET.
- Parasitic elements in wiring
  - is distributed in the wiring field. The equivalent circuit model is manually or automatically generated by extracted by LPE (Layout Parasitic Extraction or Layout Parameter Extraction).

6.1.4 Transient response of CMOS logic

Delay time analysis of CMOS logic and driver circuit for speeding up

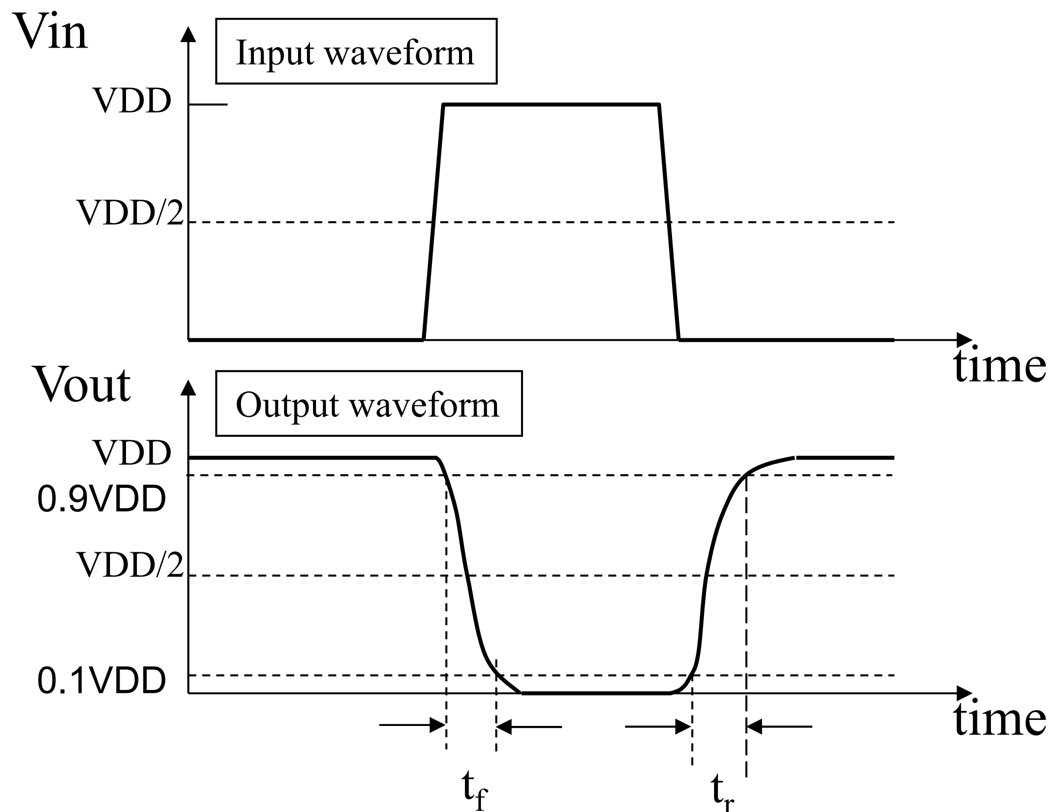
Cause of rise time and fall time



The large C_L causes the long rise time and the long fall time.

Definition of rise time and fall time

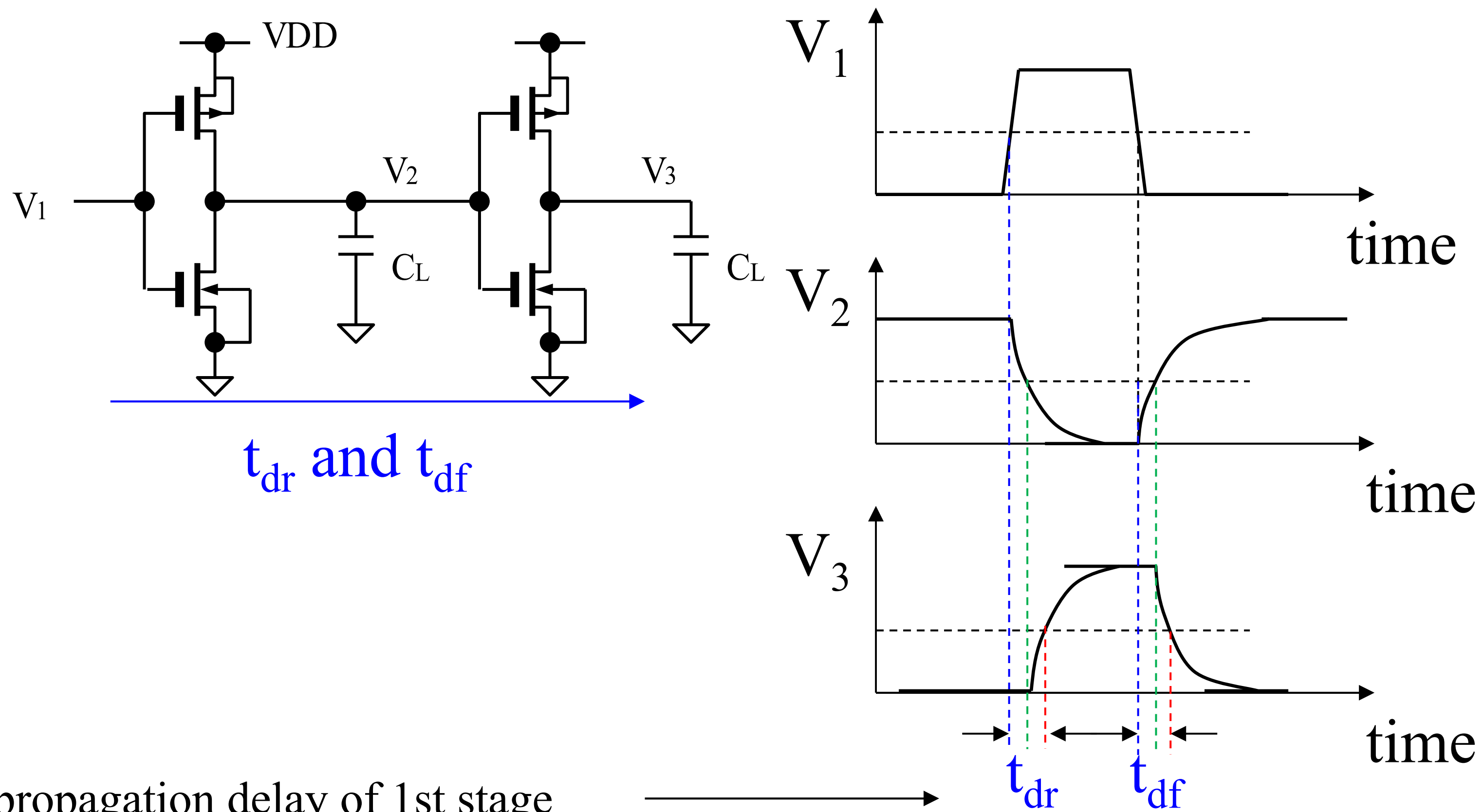
Transient response of inverter



t_f : Fall time

t_r : Rise time

Cause of propagation delay

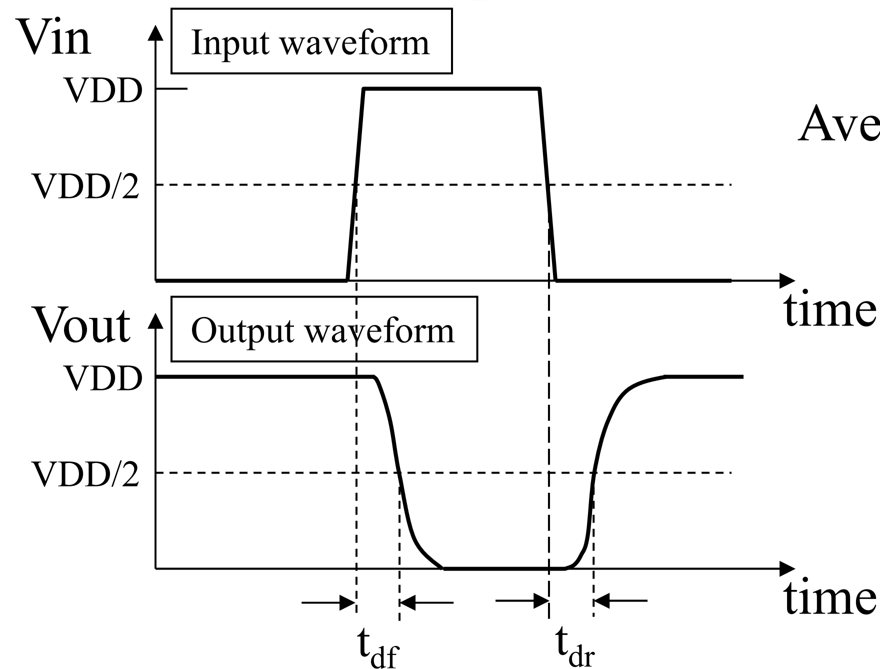


The propagation delay of 1st stage

A multi-stage logic circuit accumulates the delay time of gates. On the other hand, the rise time and fall time do not depend on the number of stages.

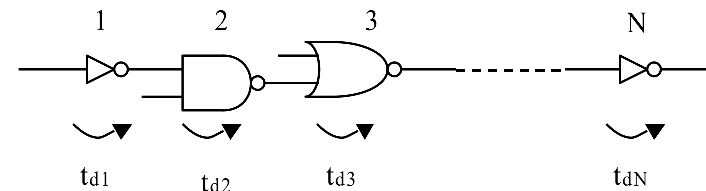
Definition of propagation delay time

Transient response of inverter



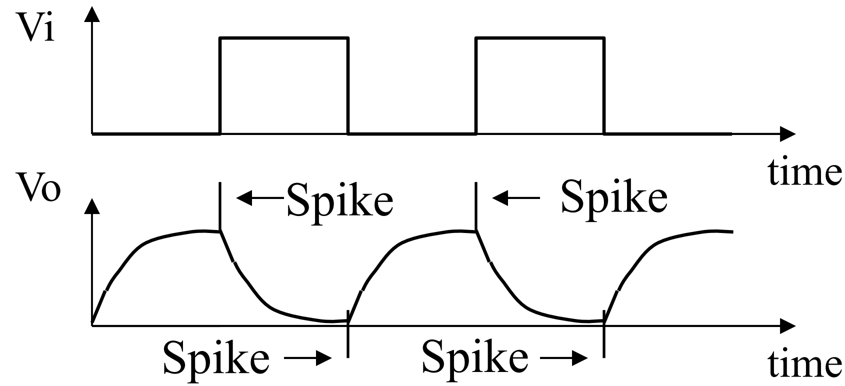
Average delay time $t_d = \dfrac{t_{df} + t_{dr}}{2}$

In the case of N-stage gate



Total delay time $t_{total} = \sum_{n=1}^{N} t_{d_n}$

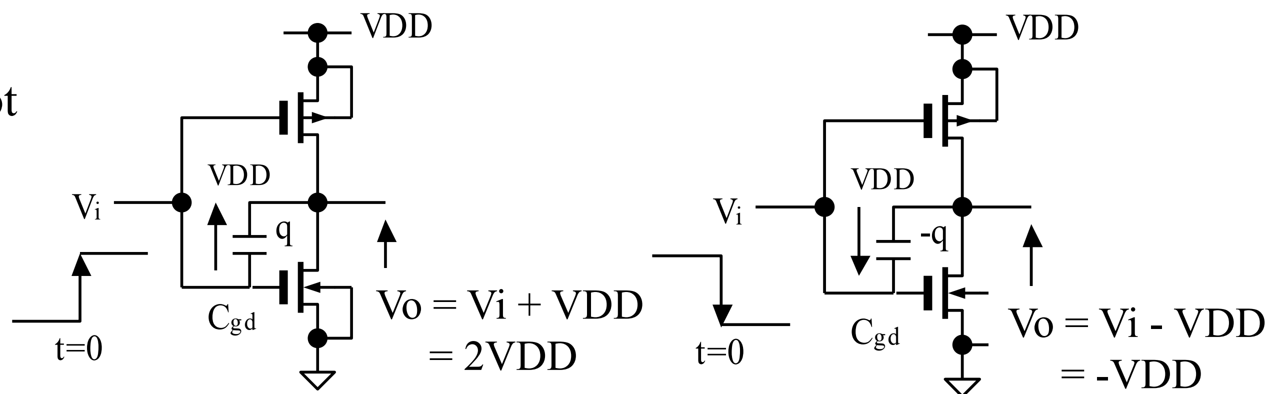
Spikes



A parasitic capacitance in CMOS logic causes spikes. The spikes may be the cause of the malfunction of the circuits.

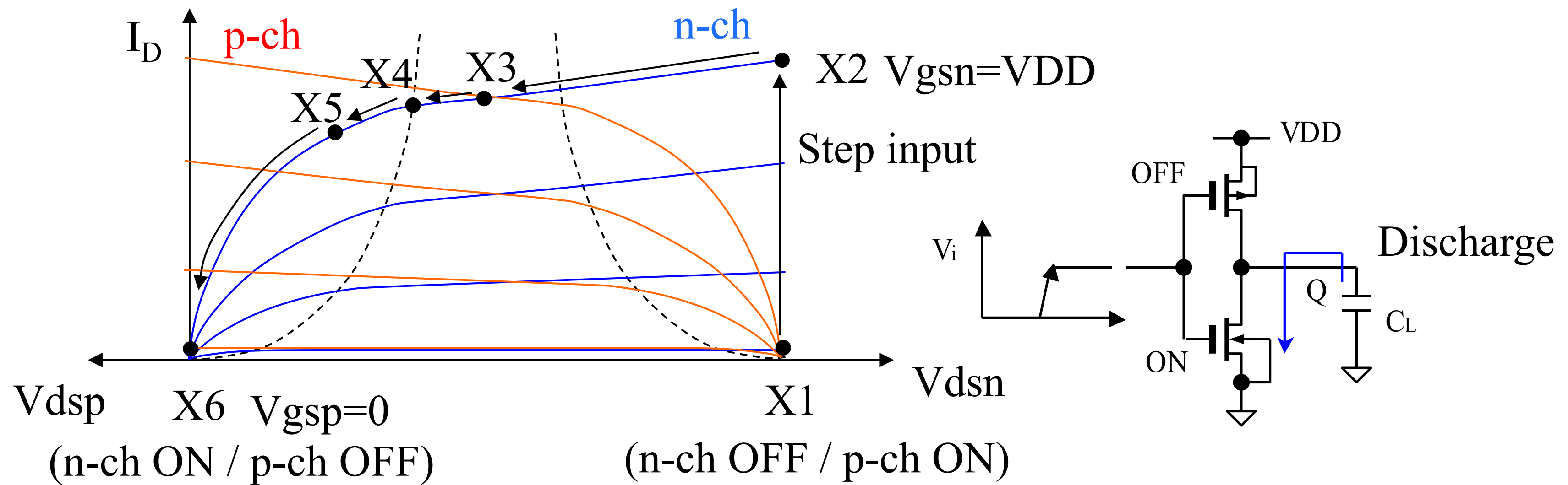
Clock Feedthrough

The charge q on cannot instantly escape from C_{gd} and the output voltage Vo is doubled momently.



Transient response of MOSFETs

States of MOSFETs X1 - X2 - X3 - X4 - X5



State	X1	X2,X3	X4,X5	X6
n-ch	Subthreshold	Saturation	Linear	Linear
p-ch	Linear	Subthreshold	Subthreshold	Subthreshold

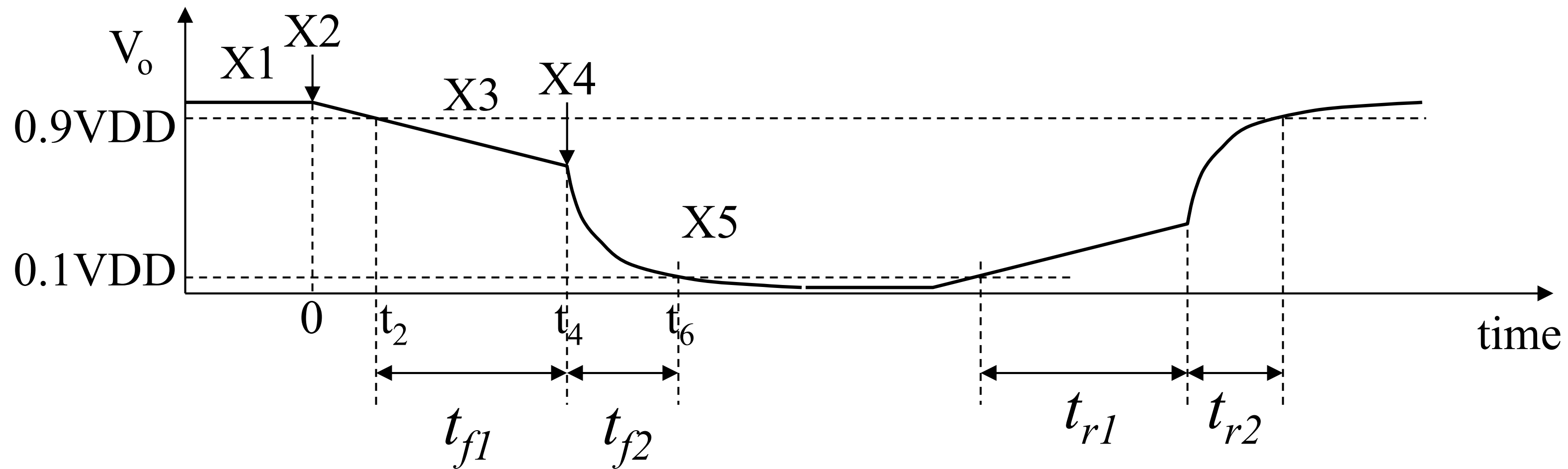
Current in MOSFET in the transient response

States	Current of n-ch MOSFET	Output voltage
X1→X2 (Subthreshold to Saturation)	Momentary transition (Direct path short current, See the latter chapter)	VDD (Constant)
X2→X3→X4 (Saturation)	$\frac{\beta_n}{2}(VDD - V_{tn0})^2$	$-\frac{\beta_n}{2C_L}(VDD - V_{tn0})^2 t + VDD$
X4→X5→X6 (Liniear)	$\beta_n\{(VDD - V_{tn0}) \cdot V_o - \frac{1}{2}V_o{}^2\}$	$\frac{2(VDD - V_{tn0})}{\exp\{\frac{\beta_n}{C_L}(VDD - V_{tn0})t\} + 1}$

[Note] cf. Appendix 4 for the detailed calculation.

Calculation of rise and fall time

Output wave form (but practically, there is no angle at X4.)



$$\begin{cases} t_f = t_{f1} + t_{f2} = \dfrac{2C_L}{\beta_n(VDD - V_{tn0})}\left\{\dfrac{V_{tn0} - 0.1\cdot VDD}{VDD - V_{tn0}} + \dfrac{1}{2}\ln\left(\dfrac{1.9\cdot VDD - 2\cdot V_{tn0}}{0.1\cdot VDD}\right)\right\} \\ t_r = t_{r1} + t_{r2} = \dfrac{2C_L}{\beta_p(VDD + V_{tp0})}\left\{\dfrac{-V_{tp0} - 0.1\cdot VDD}{VDD + V_{tp0}} + \dfrac{1}{2}\ln\left(\dfrac{1.9\cdot VDD + 2\cdot V_{tp0}}{0.1\cdot VDD}\right)\right\} \end{cases}$$

Methods for reducing the rise time and fall time

C_L	Small
$\beta_{n/p}$ or $W_{n/p}/L_{n/p}$	Large
VDD	Large

[Note] The load capacitance C_L depends in the W/L of MOSFET. The large $\beta_{n/p}$ and the small C_L is incompatible. According to the further analysis,

$$\frac{\left(\frac{W_n}{L_n}\right)}{\left(\frac{W_p}{L_p}\right)} = \sqrt{\frac{\mu_p}{\mu_n}} \quad \text{for minimum time of } t_r + t_f$$

FYI, for maximum noise margin

$$\begin{cases} V_{tn0} = |V_{tp0}| \\ \beta_n = \beta_p \end{cases} \qquad \beta_n = \mu_n C_O \frac{W_n}{L_n}, \beta_p = \mu_p C_O \frac{W_p}{L_p} \rightarrow \frac{\left(\frac{W_n}{L_n}\right)}{\left(\frac{W_p}{L_p}\right)} = \frac{\mu_p}{\mu_n}$$

Maximum operating frequency of gate

Definition of the maximum operating frequency f_{max}

$$f_{\max} \approx \frac{1}{T_{\min}} = \frac{1}{t_r + t_f} \ (Hz)$$

When the pulse width is lower than t_r + t_f, the logic swing is decreased from V_{OH} - V_{OL}.

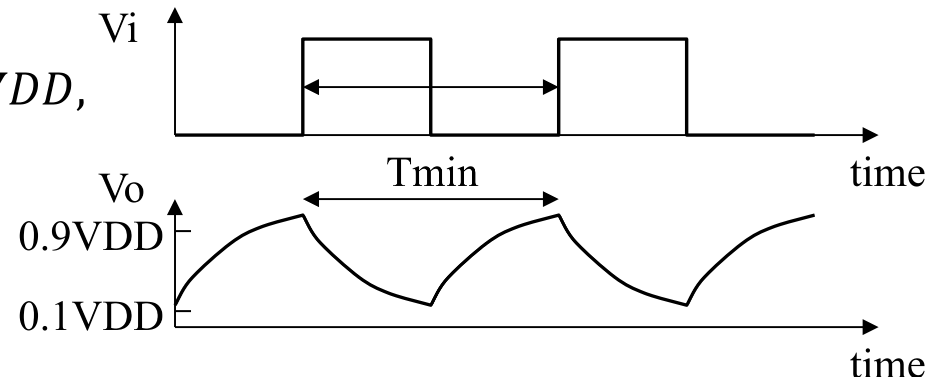
Example

$$When\ V_{tn0} = 0.2VDD\ and\ V_{tp0} = -0.2VDD,$$

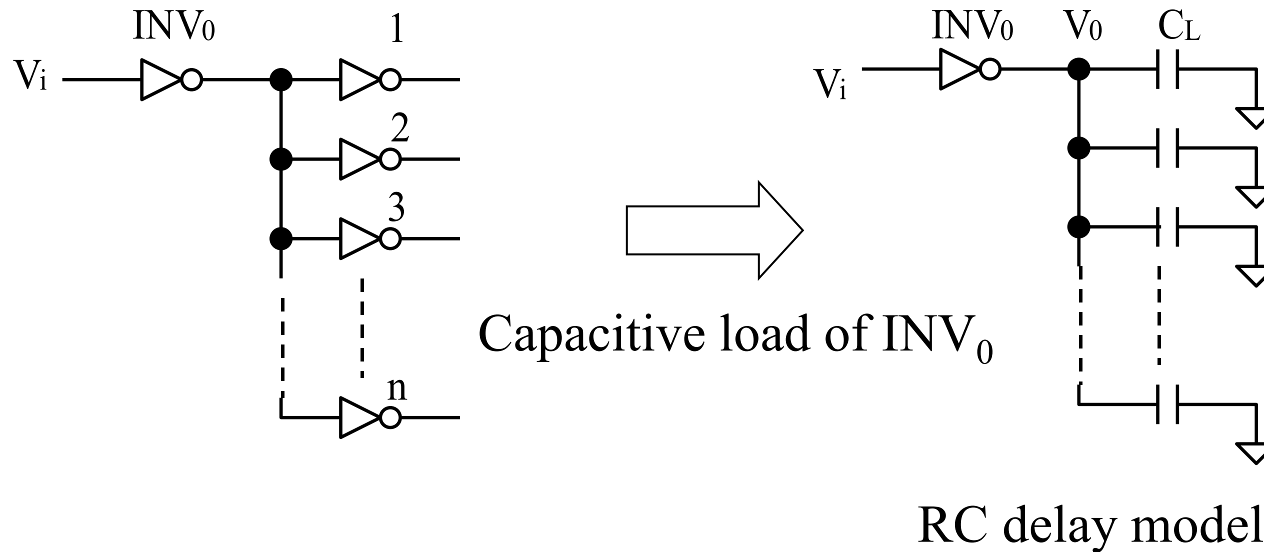
$$t_f = \frac{3.70C_L}{\beta_n VDD}$$

$$t_r = \frac{3.70C_L}{\beta_p VDD}$$

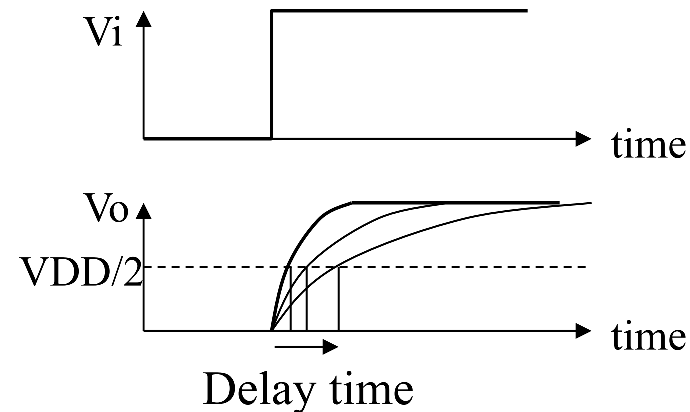
$$f_{max} = \frac{VDD}{3.70C_L}\frac{\beta_n \beta_p}{\beta_n + \beta_p}$$



Fan-out 1



Number of fan-out
n_{fo}= Number of gate in the next stage

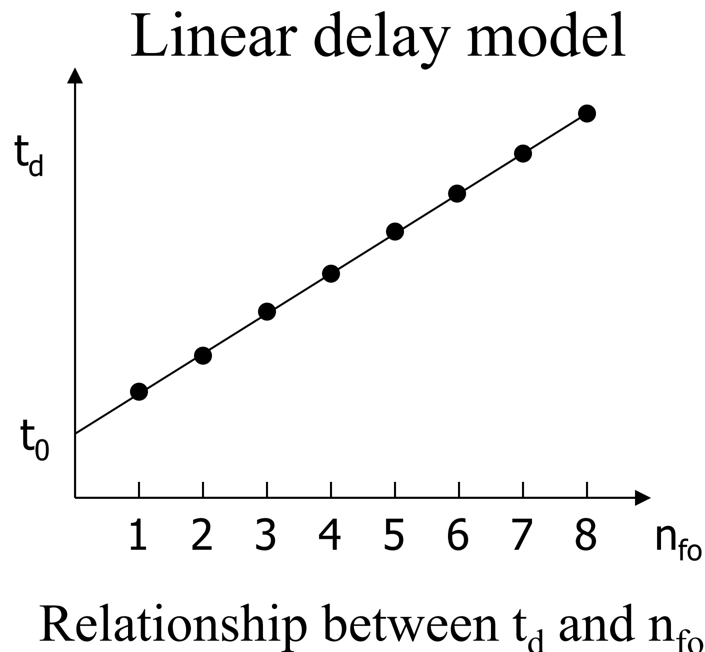


The total load of $INV_0 = n_{fo}C_L$.
The large n_{fo} causes the long charging time.

Fan-out 2

A delay time is proportional to on the number of stages and fan-out. However, it should be noted that the number of stages and fan-out is incompatible. The small number of stages is not always optimum for the high-speed operation.

Linear delay model



Relationship between t_d and n_{fo}

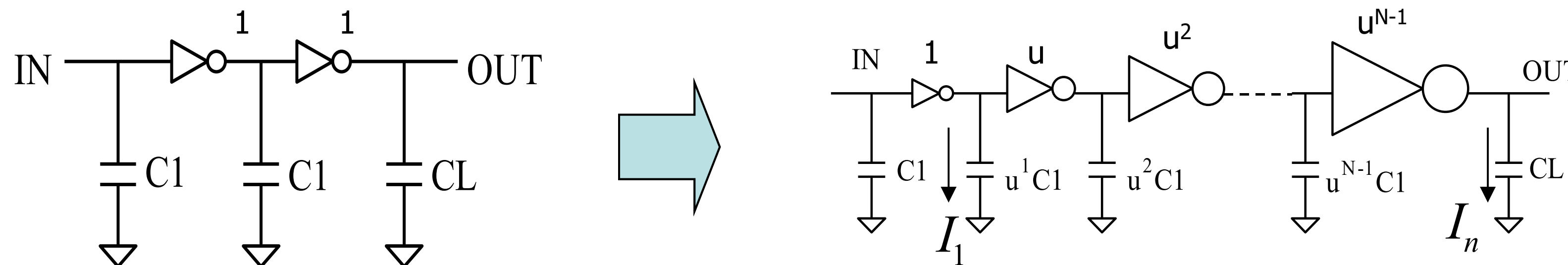
The intrinsic delay time t_0 (Parasitic delay) is observed even if $n_{fo} = 0$. The parasitic capacitance on the output node of the gate causes t_0.

[Note] The parasitic capacitance of wiring is added to the total load of the gate. This disturbs the proportional relationship between t_d and n_{fo}.

Multistage driver

A large capacitive load, e.g. data bus, cannot drive at a high speed by the small inverter (with small W/L of MOSFET). A buffer consisted cascaded inverters is applied to decrease the delay time.

The size ratio W_{N+1}/W_N of MOSFET per 1-stage is u.



For input capacitance of gate C1, Load CL, effort x,

$$CL = x \cdot C1$$

The input capacitance and the output current are proportional to the size of MOSFET.

$$\begin{cases} C_n = u^n C1 \\ I_n = u^{n-1} I1 \end{cases}$$

Delay time of nth-stage $t_{dn} = \frac{C_n}{I_n} = \frac{u^n}{u^{n-1}}\frac{C1}{I1} = u \cdot t_{d1}$

Optimization of MOSFET size

When the load capacitance CL can be drive by N-stage driver.

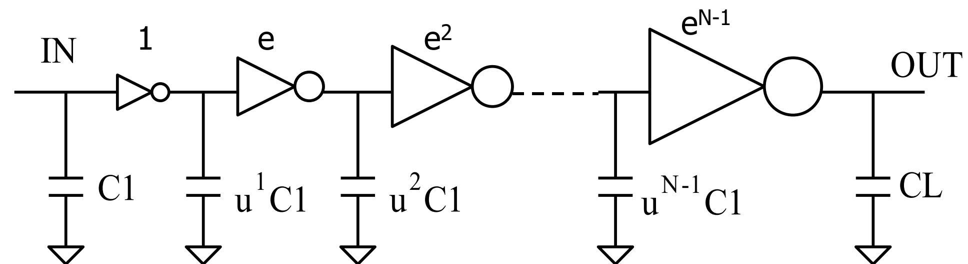
$$CL = u^N \cdot C1 = x \cdot C1 \quad \Rightarrow \quad N = \frac{\ln x}{\ln u}$$

Total delay time: $$t_d = \sum_{n=1}^{N-1} t_{dn} = N \cdot u \cdot t_{d1} = u \frac{\ln x}{\ln u} t_{d1}$$

The total delay time is minimized by the optimization of u.

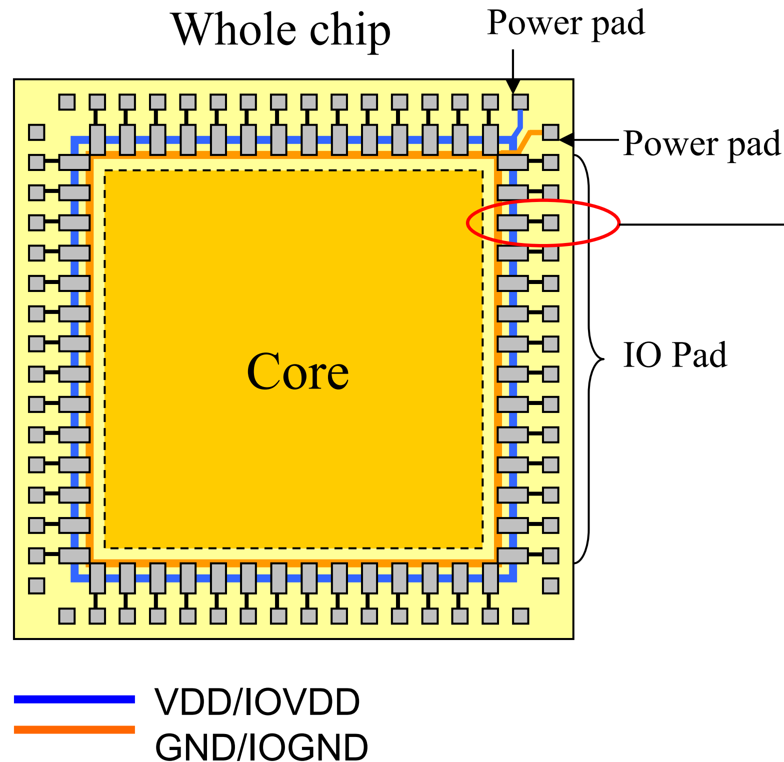
$$\frac{\partial t_d}{\partial u} = \frac{\ln x}{\ln u} t_{d1} \left(1 - \frac{1}{\ln u}\right) = 0$$

$$\therefore u = e = 2.7182 \cdots$$



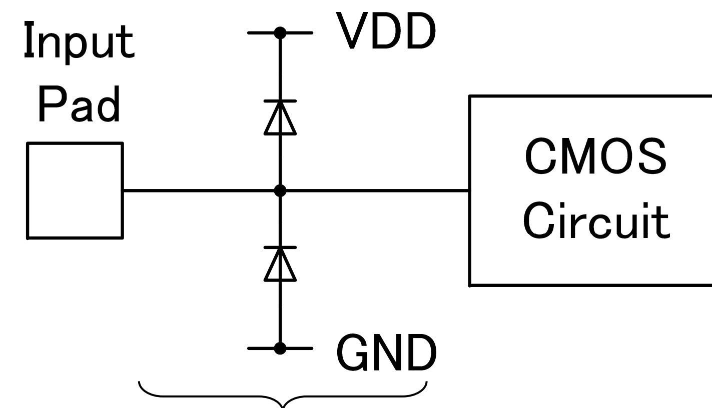
The output capacitance is ignored in this calculation. The drain capacitance of (n-1)th stage should be added to the input capacitance of nth stage as a load capacitance of the (n-1)th stage. When the input capacitance and output capacitance is considered, the optimum value of u is 3.59.

Input buffer



IO buffer + Power ring + Pad

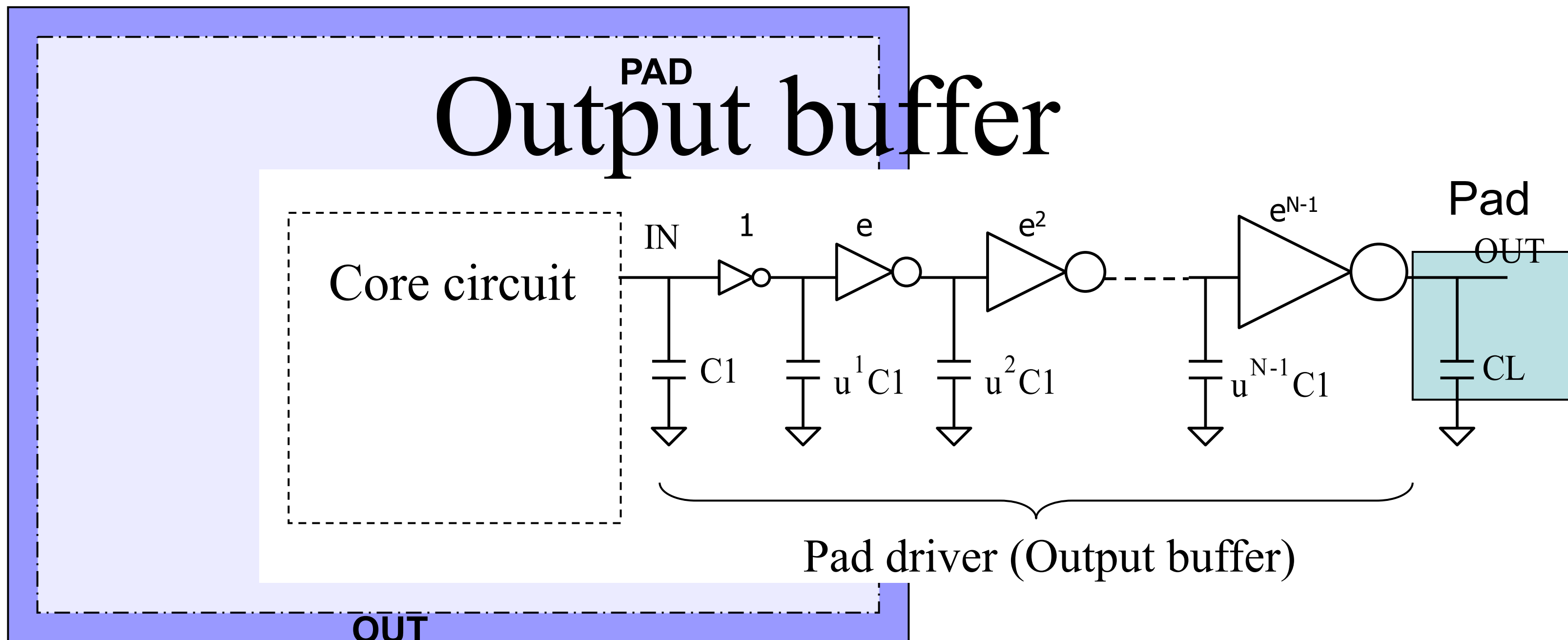
- Input buffer = ESD protection
- Output buffer = Multistage driver

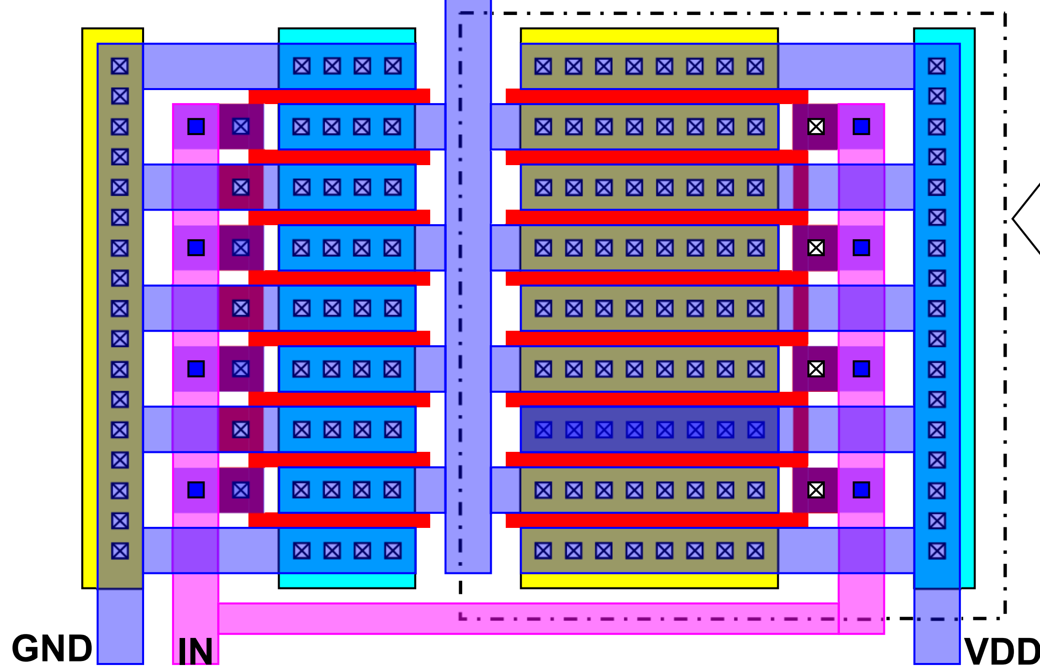


ESD (Electrostatic Discharge)

[Note] Normally, the power pads and the power lines for IO circuit are independently provide from the power line of core circuit.

Output buffer





Layout example of output buffer

If you want to implement a bidirectional buffer, replace a last stage of a cascaded inverter with a tristate buffer.