Electrical and I/O Design

Conditions leading to malfunctioning in CMOS gates

1) Incorrect supply levels or power noise
2) Input noise
3) Faulty transistors / connections to transistors
4) Incorrect ratios
5) Charge sharing in clocked schemes

- CMOS will tolerate limited levels of power supply noise when powered by a well designed supply.

- It was shown that

$$t_{HL} \& t_{LH} \approx K \cdot \frac{C_L}{\beta V_{DD}}$$

Recall $t_{r,f} = \frac{C_L V_{DD}}{\beta_{EFF}(V_{DD} - V_T)^2}$

β_{EFF} { Transistor geometry; # of transistors in series or parallel

C_L { Driver size; No. of load gates; Wiring

Fan-In and Fan-out:

Fan-in: Number of inputs in a single gate

Fan-out: Number of inputs being fed by a single output

Stage Ratio: Increase in transistor size in successive logic gates

The larger the fan-in the larger the stage ratio and the worse the delay because of the larger number of series transistors.

$$t_r = \frac{R_P}{n}(mn C_d + C_w + K C_a)$$

R_P = Effective pull-up resistance

n = Width multiplier for p-devices

k = fan-out

m = fan-in

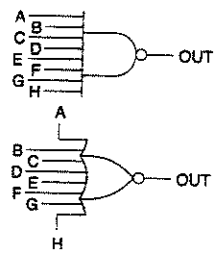
$C_G + C_D + C_W = C_L$

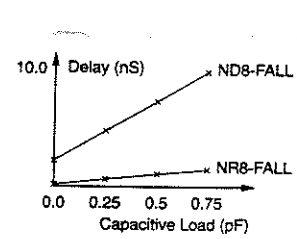
Excluding C_W

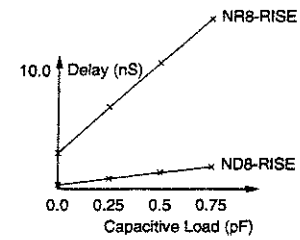
$$t_r = R_P C_G m r + \frac{R_P C_G}{n} f(k) + \frac{R_P C_G}{n} k$$

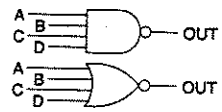
with

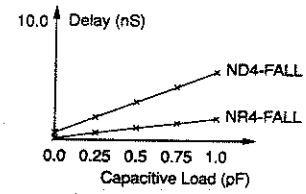
$r = \frac{C_d}{C_G}$ $f(k)$ = a function of the fan-out

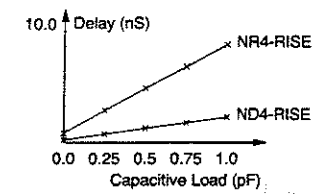


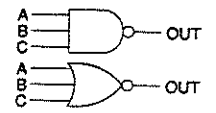


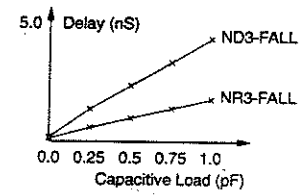


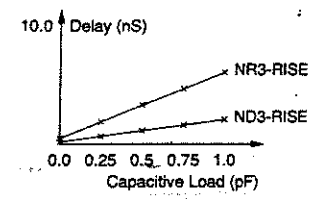


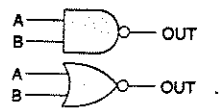


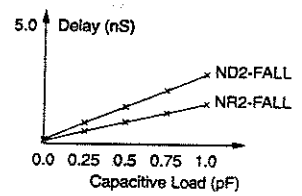


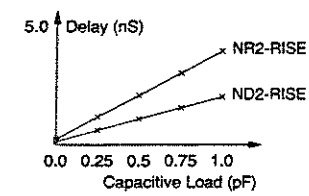


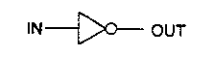


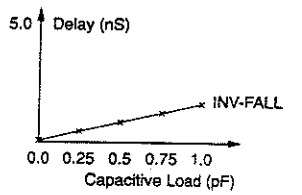


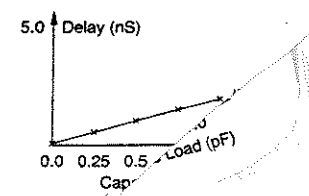








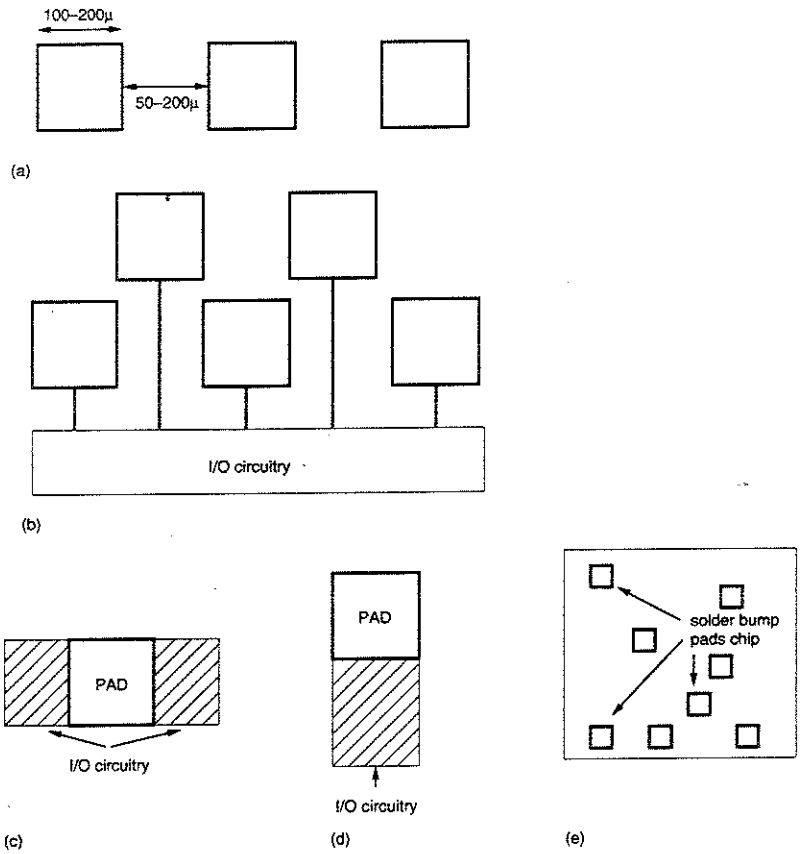




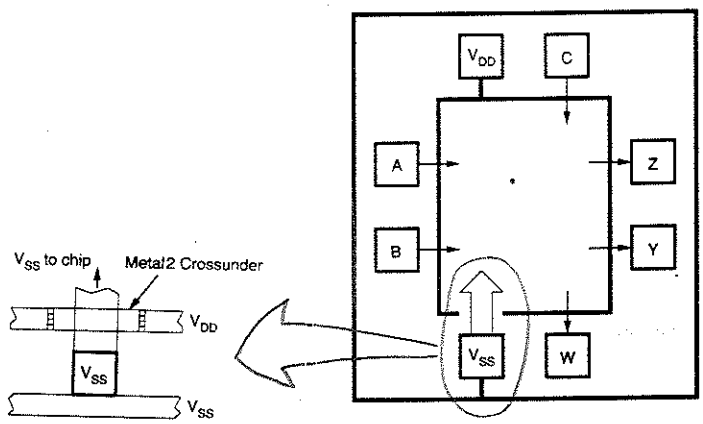
I/O Structures

- The design of I/O pads is a critical stage due to the large off-chip capacitances it typically drives.
- In most cases, pads are selected from a library, according to the process being used.
- Pad size is a function of the bonding wire size to be attached Typical sizes 100 to 150μ sq.
- Pad spacing determined by the minimum pitch at which the bonding machine can operate. Typical spaces 150 to 200μ
- High pad counts achieved through interdigitation
- Pad styles

CORE LIMITED: Associated I/O circuitry on pad sides
PAD LIMITED: " " " under pad towards IC center
ANYWHERE : Placed anywhere on the chip



- The widths of VDD and ground rails are computed from a worst case power dissipation analysis
- Multiple power supply pads provided to improve noise immunity
- Multiple vias on the interconnect reduces contact resistance
- In the I/O frame it is common practice to place the ground rail in the outermost trace



Sample I/O pad config with

- Power rails
  - V_{DD}
  - GND
- Input Output signals
  - A
  - B
  - C
  - W
  - Y
  - Z

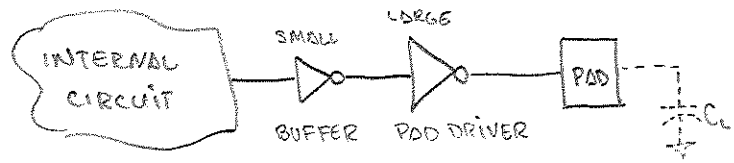
- Output Pads must have sufficient drive capability to achieve adequate rise and fall times.

Recall $t_p = \frac{C_L V_{SWING}}{I_{av}}$ $\Rightarrow$ The larger C_L the larger I_{av} to maintain the same t_p

- Proper I_{av} achieved through transistor sizing

Recall $I_{av} \propto \beta = C_{ox} \mu \frac{W}{L}$ → transistor size

- The large size of the pad driver usually requires an additional buffer to present a small capacitance to the internal circuit



- Due to large I/O currents, guard rings nessary around transistors to prevent latchup
- Large I/O transistors built from parallel smaller transistors
  Why? - Shorter gate (poly) lines.
  - Parallel metal connections to avoid electromigration

Example layout of an output pad driver.

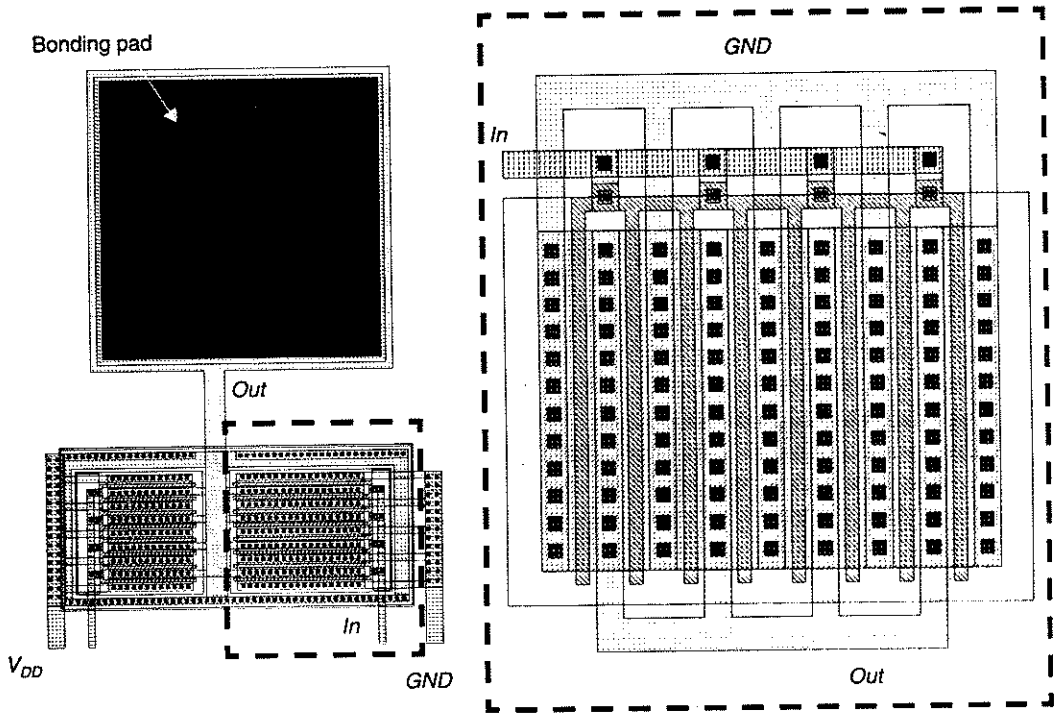


Figure 8.17 Layout of final stage of bonding-pad driver. The plot on the right side is a magnification of the NMOS transistor connected between *GND* and *Out*. See also Colorplate 13.

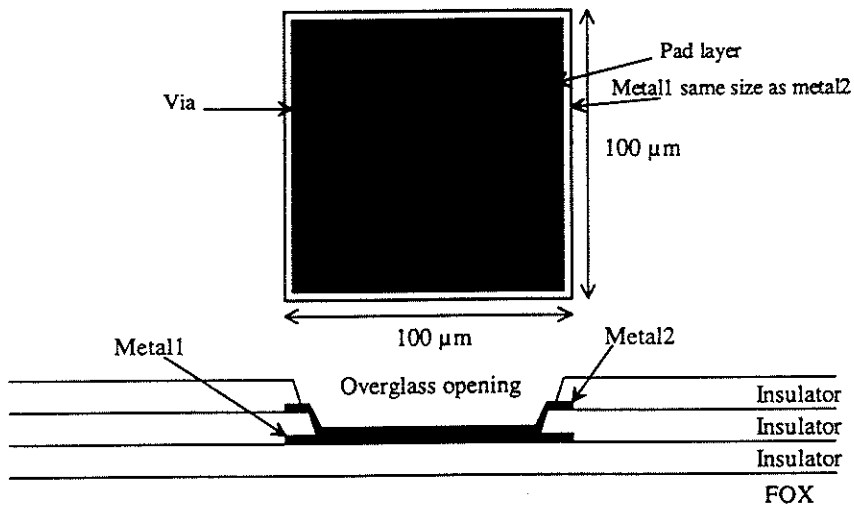


Figure 3.3 A bonding pad using metal1 and metal2.

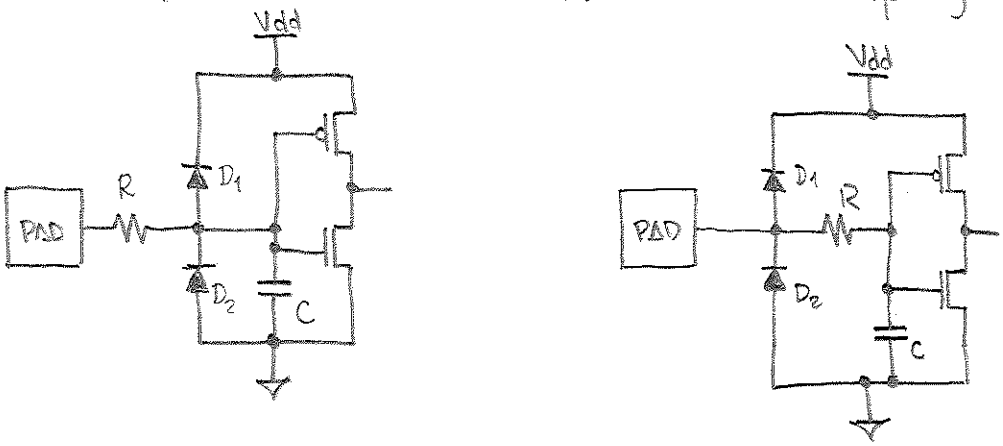
Input Pads

- Input buffers required to withstand/provide
  - High voltages
  - Static discharges
  } protection required
  - High input impedance ($10^{12} - 10^{13}\ \Omega$)
- Large voltage accumulation due to input capacitance

$$V = \frac{I \cdot \Delta t}{C_g}$$

Example: $I = 10\mu A$ $C_g = 0.03\,pF$, $\Delta t = 1\mu s \Rightarrow V = 330\,V$!!!

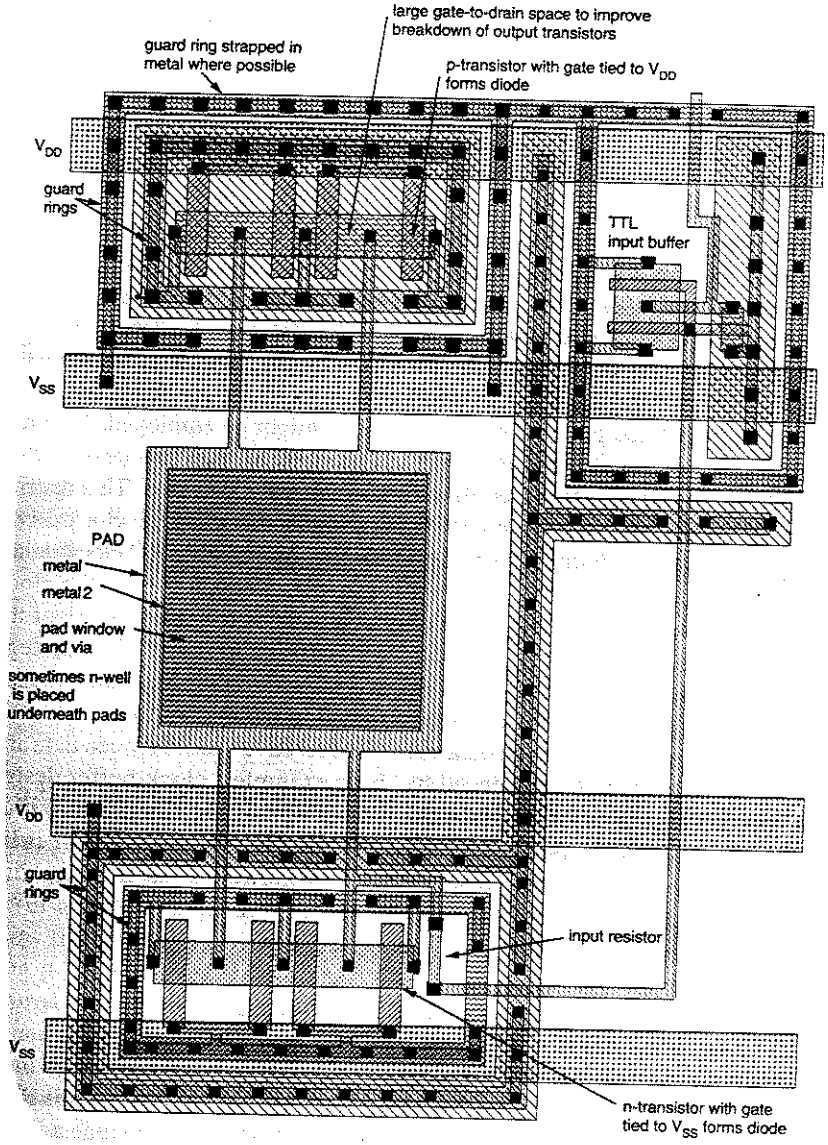
- Protective circuit based on resistors & clamping diodes



- Diodes clamp the input voltage above Vdd and below Vss
- Resistor limits the peak current
  - $R \simeq 200\,\Omega$ to $3K\,\Omega$
- Capacitor symbolizes input capacitance. Note that
  - RC time constant induced at input
  - Must be considered in the design of high-speed inputs
- Construction:
  - R: Polysilicon resistor
  - Diodes formed with p^+ diff on an n^+ substrate
  - Double guard rings used around input transistors

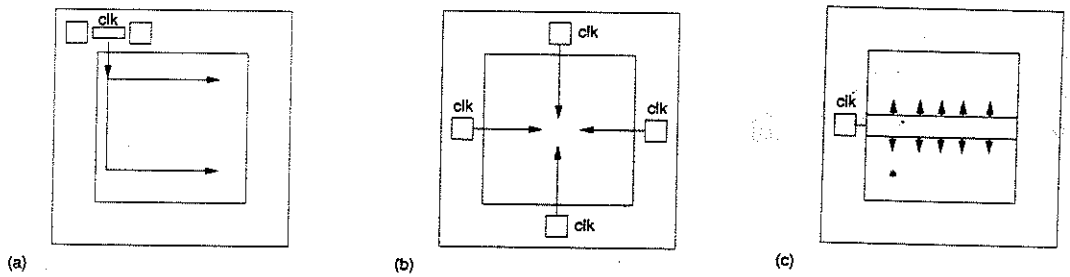
← For diode formation

- Gate length (L) longer than usual to control breakdown (~50V)
- A symbolic input pad layout



- Input pads for clock signals require special considerations:
  - Large internal loading
  - Fast rise and fall time requirements
- Techniques like
  - PLL-based clock re-synchronization & early clock inputs
  - Tri-state buffered multi clock inputs
  - Centralized, chip-wide clock driver

Approaches

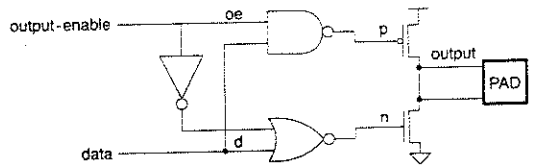


Single driver — Four-sided (3-state) — down the center

- In a layout, the usual order is:
  1) Power rails (V_{DD} and GND)
  2) Clock distribution networks
  3) Internal logic.

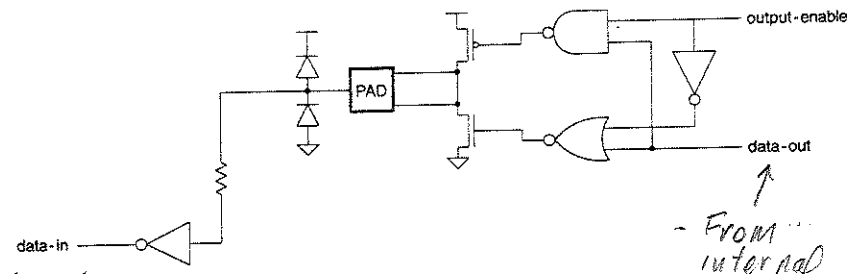
Tri-state & Bidirectional pads

- Necessary for bus contending signals



OE	D	out
1	0	0
1	1	1
0	X	Z

(a) Tri-state pad



TRUTH TABLE

OE	D	N	P	OUT	
0	X	0	1	Z	(high impedance)
1	0	1	1	0	
1	1	1	0	0	1

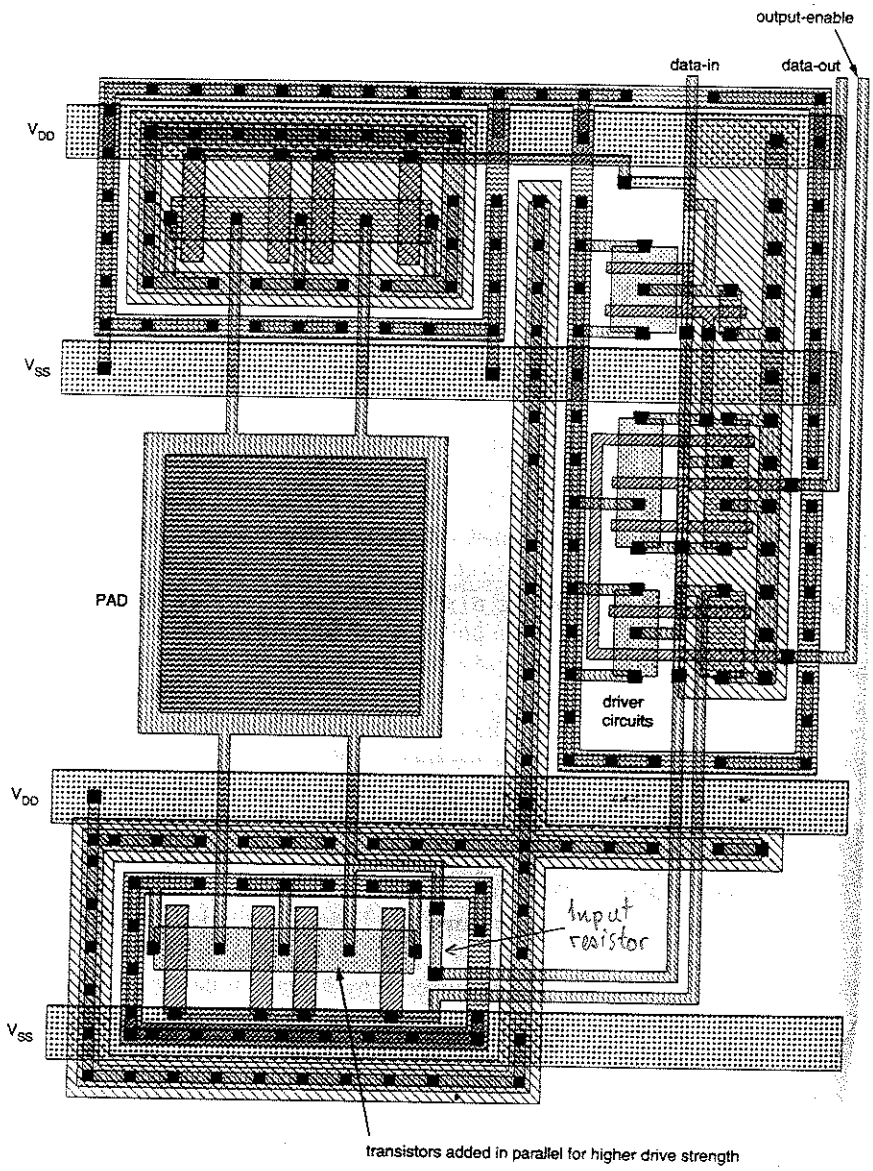
(b) Bidirectional pad

- Tri-state configuration makes out to float when OE = 0
- Bi-directional buffer combines input and tri-state pads

- Tri-state pads can be conveniently used as
  - Input pads
  - Output pads
  - Tri-state pads

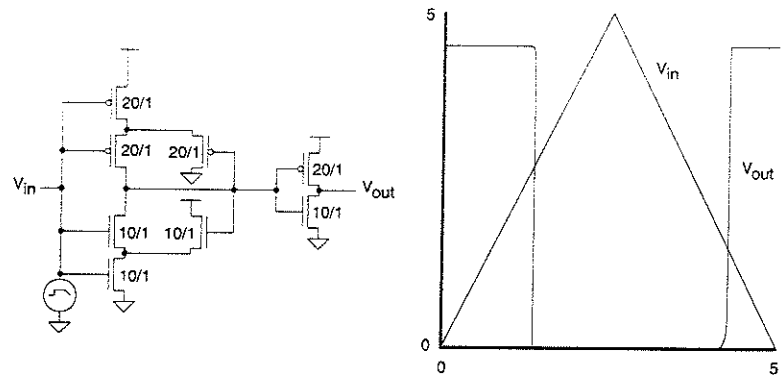
  } To reduce library size / maintanance

- Symbolic layout of a tri-state pad

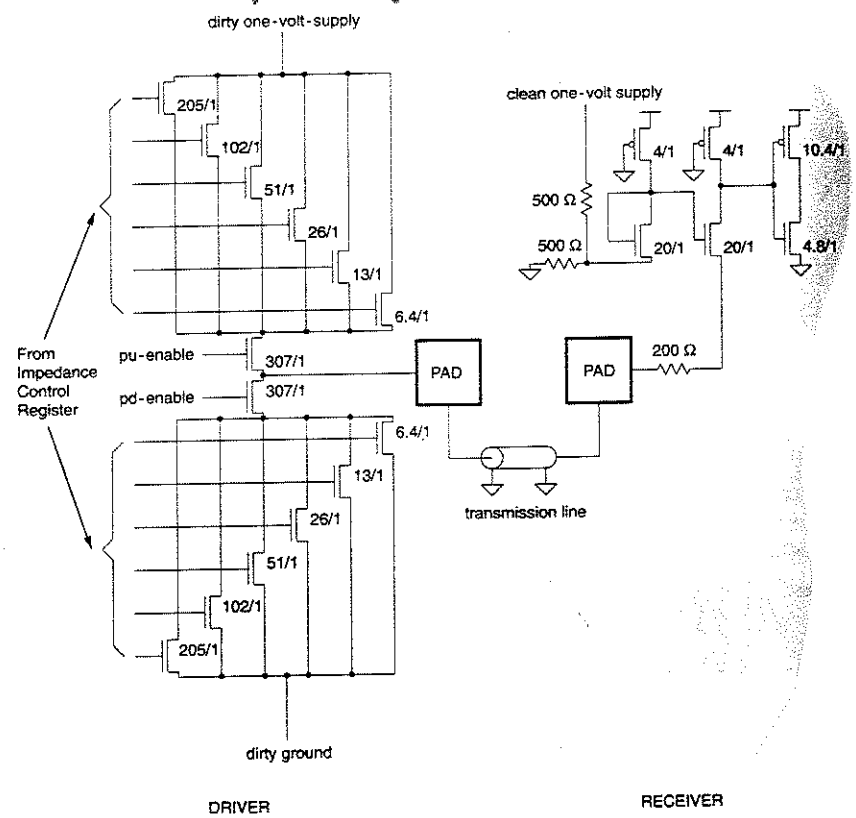


Other types of pads

- Discretionary wiring pads
  - Pull-up or Pull-down based on long gate MOS
- Low set-up and hold times
  - Latch integrated in input pad
- Pads with hysteresis
  - Based on schmitt trigger circuit



- Operation based on different threshold voltages for PUN & PDN
- Reduced Voltage-swing pads used for fast transitions (ECL pads)



Stage Ratio Calculation

Consider the propagation delay of a CMOS inverter

$$t_p = \frac{t_{PLH} + t_{PHL}}{2}$$

$$= \frac{C_L V_{DD}}{2}\left(\frac{1}{K_n(V_{DD}-V_{Tn})^2} + \frac{1}{K_p(V_{DD}-|V_{Tp}|)^2}\right)$$

Assuming $V_{DD} \gg |V_T|$

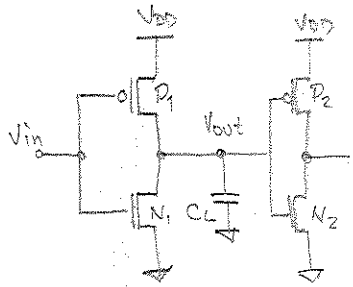
$$t_p \approx \frac{C_L}{2V_{DD}}\left(\frac{1}{K_n} + \frac{1}{K_p}\right) \Rightarrow$$ Driver circuit required to handle large loads (large C_L)

Cascading Inverter Stages:

t_p optimization ($t_{PLH} = t_{PHL}$) achieved sizing K_p & K_n according to the

$\varepsilon = \mu_n/\mu_p$ ratio. ← True only for a single gate

When multiple inverters are cascaded:



$$C_L = (C_{dp_1} + C_{dn_1}) + (C_{gp_2} + C_{on_2}) + C_W$$

Assuming PMOS α times wider than NMOS

$$\alpha = \frac{W_P/L_P}{W_n/L_n} \Rightarrow C_{dp} = \alpha C_{dn},\ C_{gp} = \alpha C_{gn}$$

$$\Rightarrow C_L = (1+\alpha)(C_{dn} + C_{gn}) + C_W = (1+\alpha)C_n + C_W$$

where $C_n = C_{dn} + C_{gn}$

$$\Rightarrow t_p = \frac{(1+\alpha)C_n + C_W}{2V_{DD}}\left(\frac{1}{K_n} + \frac{1}{K_p}\right) = \frac{(1+\alpha)C_n + C_W}{2V_{DD}K_n}\left(1 + \frac{\varepsilon}{\alpha}\right)$$

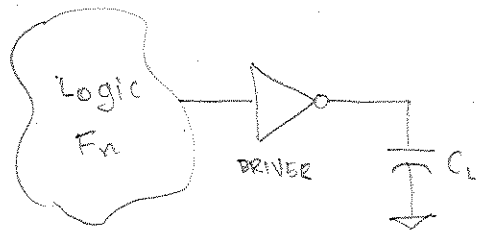
Using $\frac{W_P}{L_P} = \alpha\frac{W_n}{L_n}$ and $\mu_p = \mu_n/\varepsilon$, $K_p = \frac{\alpha}{\varepsilon}K_n$

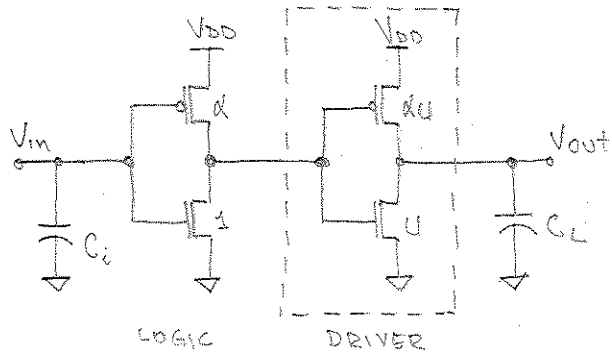
taking $\partial t_p/\partial\alpha = 0$ and isolating α^* $\quad \alpha^* = \sqrt{\varepsilon\left(1+\frac{C_W}{C_n}\right)}$

When $C_W \ll C_n$, $\alpha^* \to \sqrt{\varepsilon}$ ← True in inverter chains.

Note that for $\varepsilon = 2.5$, $\alpha^* = 1.6$ ⇒ Optimal t_p does not occur for the symmetrical case.

Single Inverter as Driver





Assume: $\begin{cases} C_L = x\,C_i, \quad W_p = \alpha\, W_n \\ \text{Driver transistors are } u \text{ times larger than logic} \end{cases}$

Find the value of u that optimizes $t_p = t_{p\,inv} + t_{p\,drv}$

Let $t_{p\alpha}$ be the propagation delay of a minimum-size gate with a unit fan-out. Then

$$t_p \approx \underbrace{u t_{p\alpha}}_{\text{logic}} + \underbrace{\frac{x}{u} t_{p\alpha}}_{\text{driver}} = \left(u + \frac{x}{u}\right) t_{p\alpha}$$

- The logic stage will be u times slower because of the increased driver size
- The driver itself benefits from u to drive the x-times larger load.

To find the optimal size make $\frac{\partial t_p}{\partial u} = 0$ and solve for u^*

$$\frac{\partial t_p}{\partial u} = \left(1 - \frac{x}{u^2}\right) t_{p\alpha} \quad \therefore \quad u^* = \sqrt{x} \quad \text{making} \quad t_p^* = 2\sqrt{x}\, t_{p\alpha}$$

When deciding whether or not to introduce a buffer-driver, just compare the single delay case versus the logic-driver case.

So it'll be worth only if $x\, t_{p\alpha} > 2\sqrt{x}\, t_{p\alpha}$ $\therefore$ $x > 4$ justifies the buffer insertion

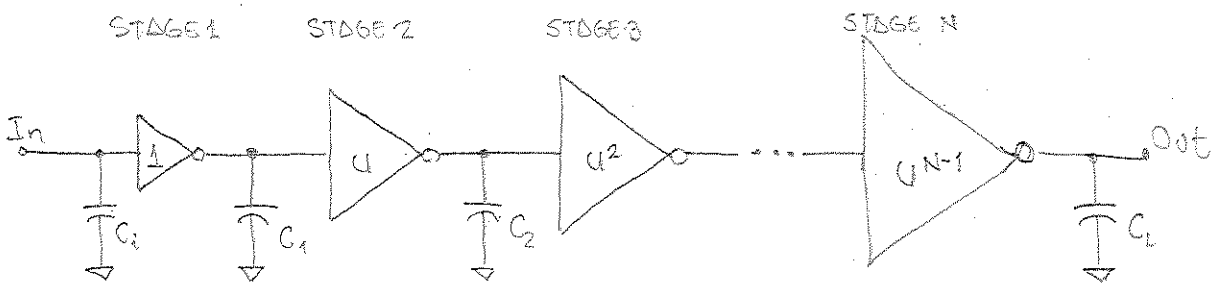
Multi-stage Buffer-driver

When C_L is too large, a single buffer might still yield poor t_p

Example: $x = 1000$; $t_p^* = 2\sqrt{10^3}\, t_{p\alpha} = \underline{64\, t_{p\alpha}} < 1000\, t_{p\alpha}$

↑ still large

Solution: Insert N stages.



- Divide the delay equally by N stages. Achieved by scaling-up all stages by a constant factor u.

Thus for each stage $t_{p_{stg}} = u \cdot t_{p0}$, where $u = C_{out}/C_{in}$

- For the last stage, $C_L = X C_{in} = u^N C_{in}$, yielding

$$t_p = N \cdot u \cdot t_{p0}, \text{ and } N = \ln(X)/\ln(u)$$

- Make

$$\frac{\partial t}{\partial u} = \frac{\partial}{\partial u}\left[u \cdot \frac{\ln(X)}{\ln(u)}\right] t_{p0} = 0 \quad \therefore \quad u^* = e = 2.7182$$

and

$$t_p^* = e \ln(X) t_{p0} = e \ln\left(C_y/C_i\right) t_{p0}$$

Optimal Buffer Design: Consecutive stages scale-up exponentially. In practice, $u = 3$ or $u = 4$ are rather accepted.

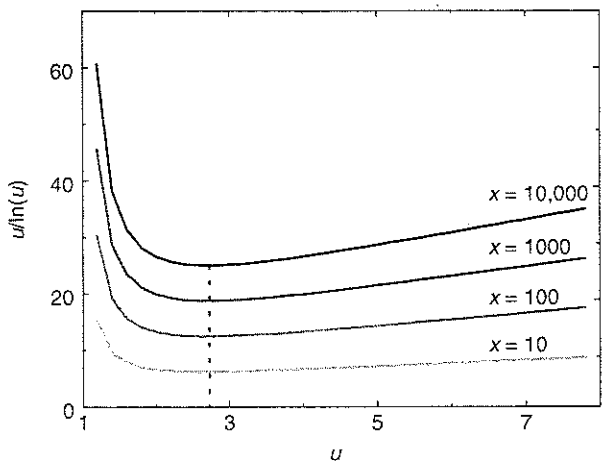


Figure 8.13 t_p/t_{p0} versus u for various values of x.

Table 8.4 t_{opt}/t_{p0} versus x for various driver configurations.

x	Unbuffered	Single Buffer	Cascaded Buffers
10	10	6.3	6.3
100	100	20	12.5
1000	1000	63	18.8
10,000	10,000	200	25.0