EE141-Spring 2012 Digital Integrated Circuits

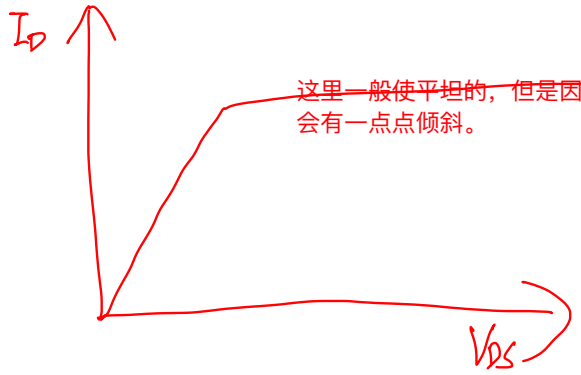
Lecture 10
capacitors and inverters

Administrativia

- Midterm to be graded early next week on Tuesday – Scores will be available next Wednesday
- Project to be launched next week

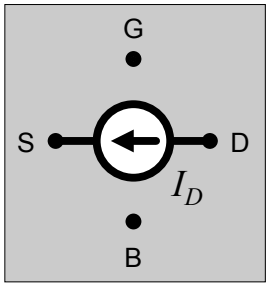
Transistor Summary

小的晶体管更容易达到SAT



A Unified Model for Manual Analysis

define $V_{GT} = V_{GS} - V_T$



for $V_{GT} \leq 0$: $I_D = 0$

for $V_{GT} \geq 0$:

$$I_D = k' \cdot \frac{W}{L} \cdot \left(V_{GT} \cdot V_{DS,eff} - \frac{V_{DS,eff}^2}{2} \right) \cdot (1 + \lambda \cdot V_{DS})$$

with $V_{DS,eff} = \min (V_{GT}, V_{DS}, V_{D,VSAT})$

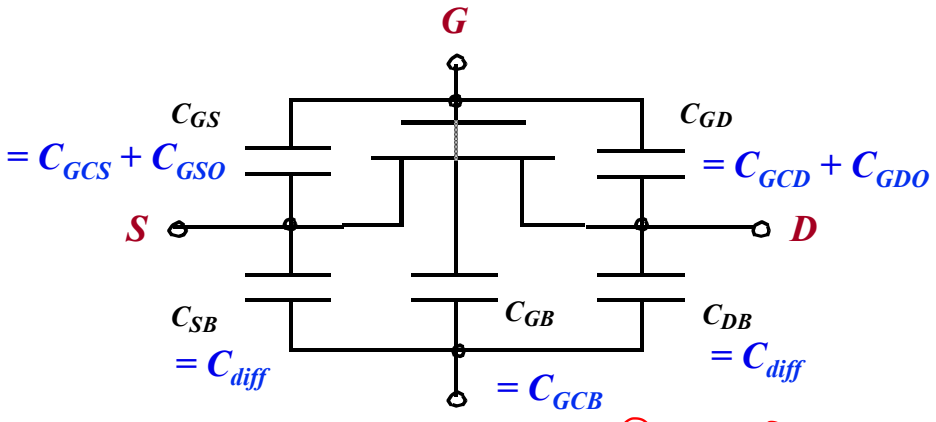
这一项就是为什么SAT了，电流仍然增加的原因。

与 V_{GS} 无关

The dynamic response of a MOSFET transistor is a sole function of the time it takes to (dis)charge the parasitic capacitances that are intrinsic to the device, and the extra capacitance introduced by the interconnecting lines (and are the subject of Chapter 4). They originate from three sources: the basic MOS structure, the channel charge, and the depletion regions of the reverse-biased pn-junctions of drain and source. Aside from the MOS structure capacitances, all capacitors are nonlinear and vary with the applied voltage.

MOS Capacitance

MOS Capacitances



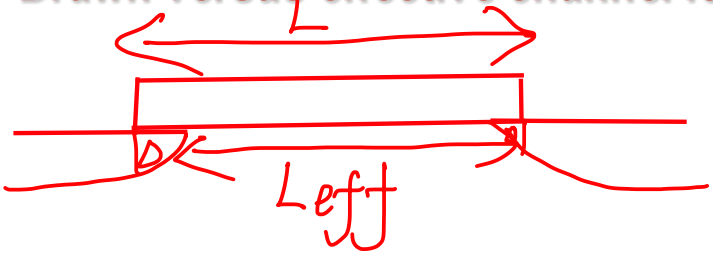
Gate Capacitance

- Capacitance (per area) from gate across the oxide is $W \cdot L \cdot C_{ox}$, where $C_{ox} = \varepsilon_{ox}/t_{ox}$

平行板电容

ox=oxide,t=thickness

Drawn versus effective channel length



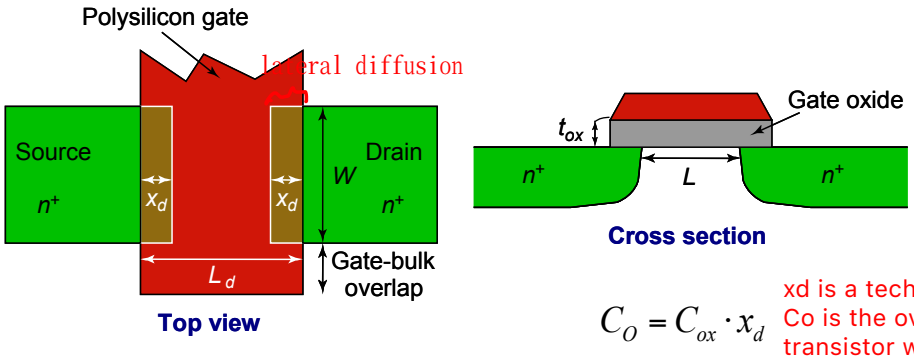
$I_D = \frac{W}{L_{eff}} \cdots$

$L_{eff} = L - 2D$

L effect 因为存在重叠，导致沟道的长度比门的长度要短。

The source and drain diffusion should end right at the edge of the gate oxide. In reality, both source and drain tend to extend somewhat below the oxide by an amount xd, called the lateral diffusion.Hence, the effective channel of the transistor L becomes shorter than the drawn length Ld (or the length the transistor was originally designed for) by a factor of ΔL = 2xd. It also gives rise to a parasitic capacitance between gate and source (drain) that is called the overlap capacitance. This capacitance is strictly linear and has a fixed value.

Gate Overlap Capacitance



$$C_O = C_{ox} \cdot x_d$$

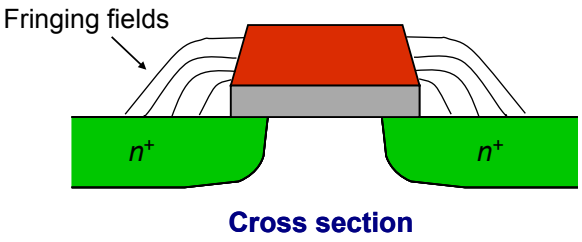
xd is a technology-determined parameter, Co is the overlap capacitance per unit transistor width.

Off/Lin/Sat → $C_{GSO} = C_{GDO} = C_O \cdot W$

The gate of the MOS transistor is isolated from the conducting channel by the gate oxide that has a capacitance per unit area equal to Cox = εox / tox. The total value of this capacitance is called the gate capacitance Cg and can be decomposed into two elements one part of Cg contributes to the channel charge, another part is solely due to the topological structure of the transistor

Gate Fringe Capacitance

between gate and drain exists fringing capacitance,一般会被包含在Coverlap中

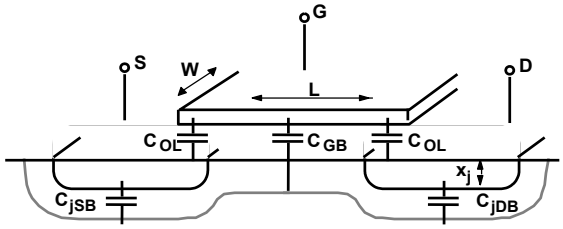


- C_{OV} not just from metallurgic overlap – get fringing fields too
- Typical value: ~0.2fF·W(in μm)/edge

Perhaps the most significant MOS parasitic circuit element, the gate-to-channel capacitance C_{GC} varies in both magnitude and in its division into three components (C_{GCS}, C_{GCD} and C_{GCB}) (being the gate-to-source, gate-to-drain, and gate-to-body capacitances, respectively), depending upon the operation region and terminal voltages.

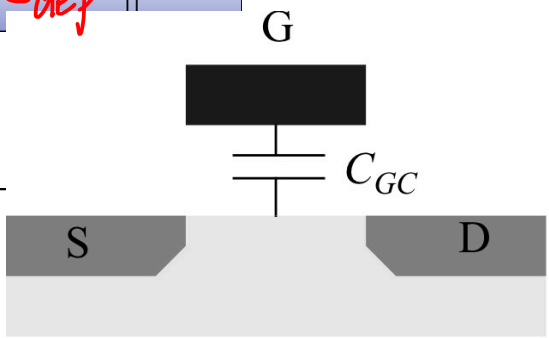
ebook: 96

Transistor In Cutoff



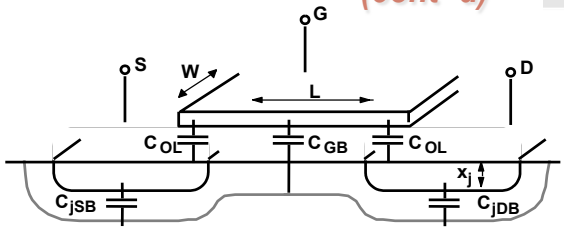
- When the transistor is off, no carriers in channel to form the other side of the capacitor.
  - Substrate acts as the other capacitor terminal
  - Capacitance becomes series combination of gate oxide and depletion capacitance

G — C_{ox} — C_{dep} — B

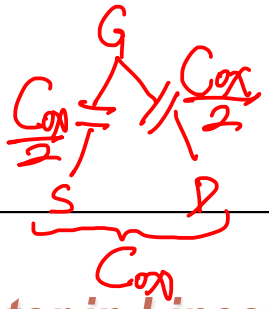


(a) cut-off

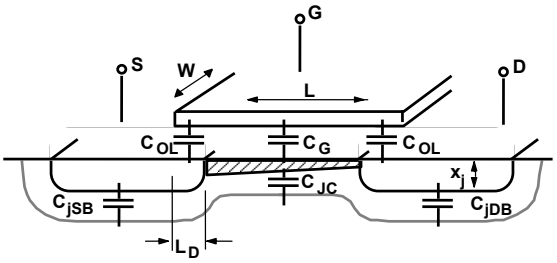
Transistor In Cutoff (cont'd)

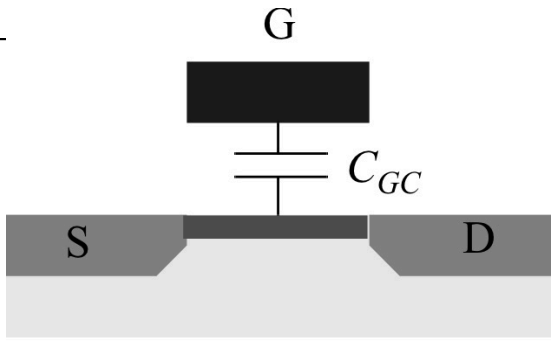


- When $|V_{GS}| < |V_T|$, total C_{GB} much smaller than $W \cdot L \cdot C_{ox}$
  - Usually just approximate with $C_{GB} = 0$ in this region.
- (If V_{GS} is "very" negative (for NMOS), depletion region shrinks and C_{GB} goes back to ~$W \cdot L \cdot C_{ox}$)



Transistor in Linear Region

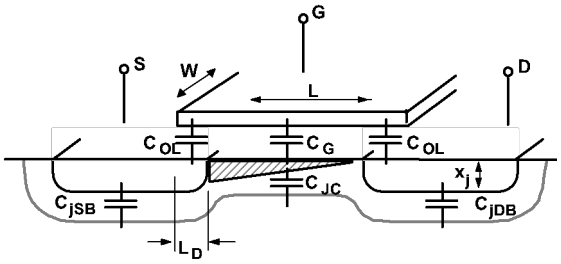




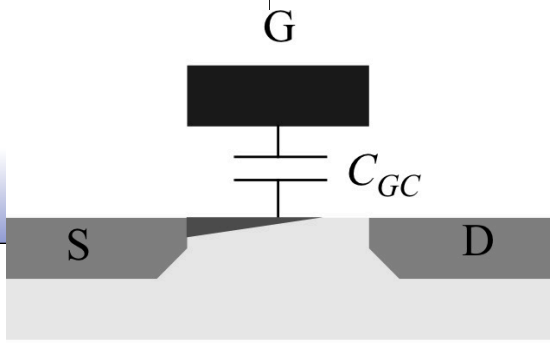
(b) resistive

- Channel is formed and acts as the other terminal
  - C_{GB} drops to zero (shielded by channel)
- Model by splitting oxide cap equally between source and drain
  - Changing either voltage changes the channel charge

Transistor in Saturation Region

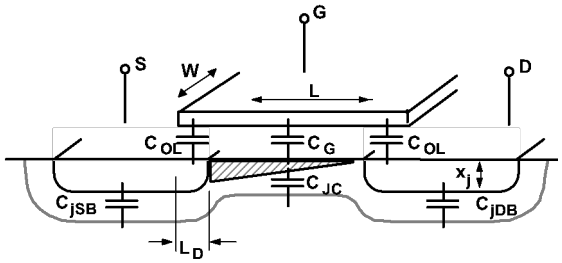


- Changing source voltage doesn't change V_{GC} uniformly
  - E.g. V_{GC} at pinch off point still V_{TH}
- Bottom line: $C_{GCS} \approx 2/3 \cdot W \cdot L \cdot C_{ox}$



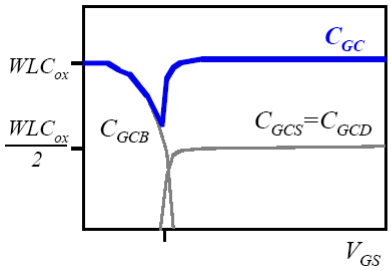
(c) saturation

Transistor in Saturation Region (cont'd)

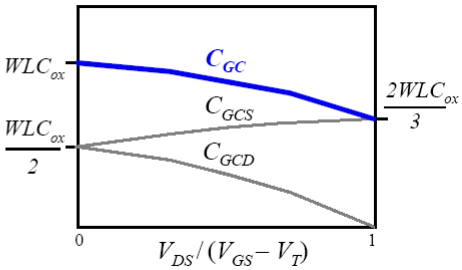


- Drain voltage no longer affects channel charge
  - Set by source and V_{DS_sat}
- If change in charge is 0, $C_{GCD} = 0$

Gate Capacitance



C_{gate} vs. V_{GS} (with $V_{DS} = 0$)



C_{gate} vs. operating region

Average distribution of channel capacitance of MOS transistor for different operation regions

Operation Region	C_{GCB}	C_{GCS}	C_{GCD}	C_{GC}	C_G
Cutoff	$C_{ox}WL$	0	0	$C_{ox}WL$	$C_{ox}WL+2C_oW$
Resistive	0	$C_{ox}WL/2$	$C_{ox}WL/2$	$C_{ox}WL$	$C_{ox}WL+2C_oW$
Saturation	0	$(2/3)C_{ox}WL$	0	$(2/3)C_{ox}WL$	$(2/3)C_{ox}WL+2C_oW$

$C_j = \frac{C_{j0}}{(1 - V_D/\phi_0)^m}$

Cj is the junction capacitance per unit area as given by ebook:72 Eq. (3.9).
As the bottom-plate junction is typically of the abrupt type, the grading coefficient m approaches 0.5.

Diffusion Capacitance

Junction Capacitances

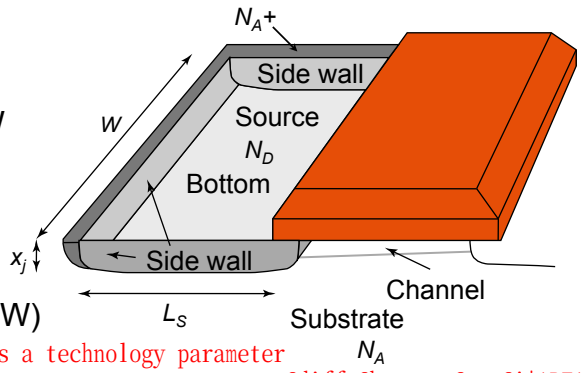
The detailed below picture shows that the junction consists of two components:

The bottom-plate junction,which is formed by the source region (with doping ND) and the substrate with doping NA.The total depletion region capacitance for this component equals

- Bottom
  - Area cap
  - $C_{bottom} = C_j \cdot L_S \cdot W$

The side-wall junction, formed by the source region with doping ND and the p+ channel-stop implant with doping level NA+.

- Sidewalls
  - Perimeter cap
  - $C_{sw} = C_{jsw} \cdot (2L_S + W)$



x_j, the junction depth, is a technology parameter

$C_{jsw} = C'_{jsw} x_j$ is a capacitance per unit perimeter

Cdiff=Cbottom+Csw=Cj*AREA+Cjsw*PERIMETER

- GateEdge
  - $C_{ge} = C_{jgate} \cdot W$
  - Usually automatically included in the SPICE model

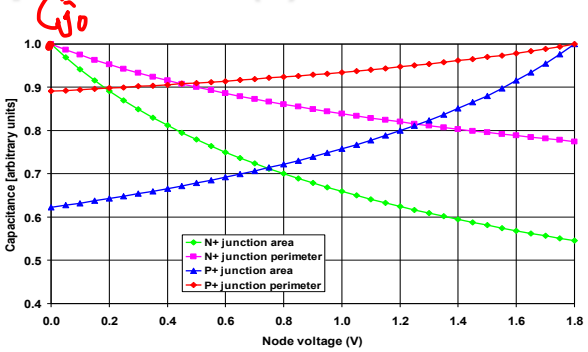
In general, it can be stated that the contribution of diffusion capacitances is at most equal, and very often substantially smaller than the gate capacitance.

SPICE (Simulation program with integrated circuit emphasis)

Junction Capacitance (2)

C_{j0}

- Junction caps are nonlinear
  - C_J is a function of junction bias



- SPICE model equations:
  - Area C_J = area × C_{J0} / (1+ $|V_{DB}|/\phi_B)^{mj}$
  - Perimeter C_J = perim × C_{JSW} / (1 + $|V_{DB}|/\phi_B)^{mjsw}$
  - Gate edge C_J = W × C_{Jgate} / (1 + $|V_{DB}|/\phi_B)^{mjswg}$
- How do we deal with nonlinear capacitance?

Linearizing the Junction Capacitance

Since all these capacitances are small-signal capacitances, we normally linearize them and use average capacitances along the lines of ebook:73 Eq. (3.10).

Replace non-linear capacitance by
large-signal equivalent linear capacitance
which displaces equal charge over voltage swing of interest

$$C_{eq} = \frac{\Delta Q_j}{\Delta V_D} = \frac{Q_j(V_{high}) - Q_j(V_{low})}{V_{high} - V_{low}} = K_{eq} C_{j0}$$

K_{eq}<1因为C_{j0}是最大的电容.

$$K_{eq} = \frac{-\phi_0^m}{(V_{high} - V_{low})(1-m)}\left[(\phi_0 - V_{high})^{1-m} - (\phi_0 - V_{low})^{1-m}\right]$$

V_{DD} $\frac{V_{DD}}{2}$ (1→0)

0 $\frac{V_{DD}}{2}$ (0→1)

Capacitance Model Summary

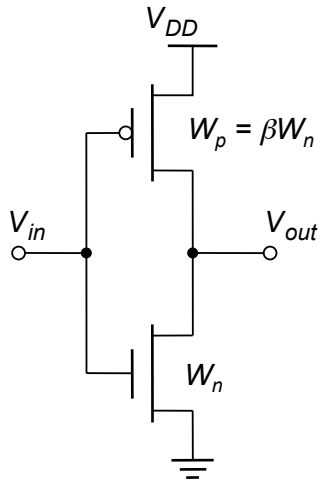
- **Gate-Channel Capacitance**
  - $C_{GC} \approx 0$ ($C_{ox}WL_{eff}$) ($|V_{GS}| < |V_T|$)
  - $C_{GC} = C_{ox} \cdot W \cdot L_{eff}$ (Linear)
    - 50% G to S, 50% G to D
  - $C_{GC} = (2/3) \cdot C_{ox} \cdot W \cdot L_{eff}$ (Saturation)
    - 100% G to S
- **Gate Overlap Capacitance**
  - $C_{GSO} = C_{GDO} = C_O \cdot W$ (Always)
- **Junction/Diffusion Capacitance**
  - $C_{diff} = C_j \cdot L_S \cdot W + C_{jsw} \cdot (2L_S + W) + C_{jg} W$ (Always)

Capacitances in 0.25 μm CMOS Process

	C_{ox} (fF/μm^2)	C_O (fF/μm)	C_j (fF/μm^2)	m_j	ϕ_b (V)	C_{jsw} (fF/μm)	m_{jsw}	ϕ_{bsw} (V)
NMOS	6	0.31	2	0.5	0.9	0.28	0.44	0.9
PMOS	6	0.27	1.9	0.48	0.9	0.22	0.32	0.9

CMOS Inverter VTC

The CMOS Inverter



Transient response is dominated mainly by the output capacitance of the gate, CL, which is composed of the drain diffusion capacitances of the NMOS and PMOS transistors, the capacitance of the connecting wires, and the input capacitance of the fan-out gates.
The gate response time(RpCL) is simply determined by the time it takes to charge the capacitor CL through the resistor Rp.Hence, a fast gate is built either by keeping the output capacitance small or by decreasing the on-resistance of the transistor. The latter is achieved by increasing the W/L ratio of the device. The reader should be aware that the on-resistance of the NMOS and PMOS transistor is not constant, but is a nonlinear function of the voltage across the transistor.

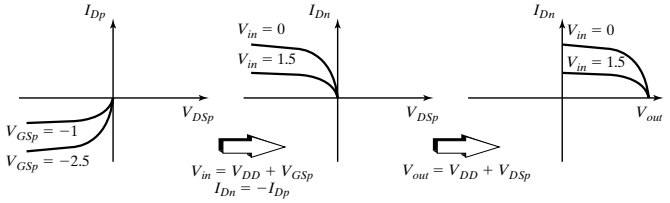
PMOS Load Lines

The nature and the form of the voltage-transfer characteristic (VTC) can be graphically deduced by superimposing the current characteristics of the NMOS and the PMOS devices. Such a graphical construction is traditionally called a load-line plot. It requiresthat the I-V curves of the NMOS and PMOS devices are transformed onto a common coordinate set.

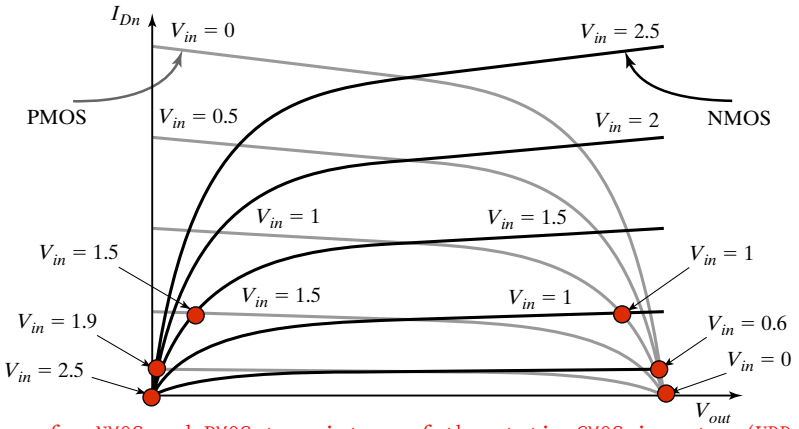
PMOS Load Lines

ebook:163

- For DC VTC, $I_{Dn} = I_{Dp}$ the subscripts n and p denote the NMOS and PMOS devices, respectively
  - Graphically, looking for intersections of NMOS and PMOS IV characteristics
- To put IV curves on the same plot, PMOS IV is "flipped" since $|V_{DSp}| = V_{DD} - V_{out}$
  - Also, $|V_{GSp}| = V_{dd} - V_{in}$



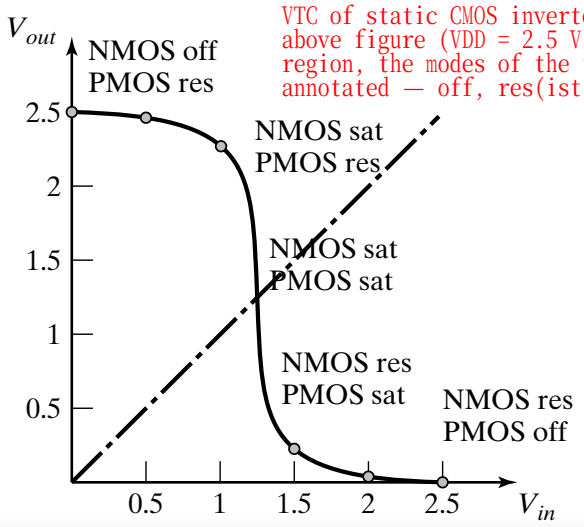
CMOS Inverter Load Characteristics



Load curves for NMOS and PMOS transistors of the static CMOS inverter (VDD = 2.5 V). The dots represent the dc operation points for various input voltages

For a dc operating points to be valid, the currents through the NMOS and PMOS devices must be equal. Graphically, this means that the dc points must be located at the intersection of corresponding load lines. A number of those points (for Vin = 0, 0.5, 1, 1.5, 2, and 2.5 V) are marked on the graph. All operating points are located either at the high or low output levels.The VTC of the inverter hence exhibits a very narrow transition zone. This results from the high gain during the switching transient, when both NMOS and PMOS are simultaneously on, and in saturation. In that operation region, a small change in the input voltage results in a large output variation. This result has been shown below VTC figure.

CMOS Inverter VTC

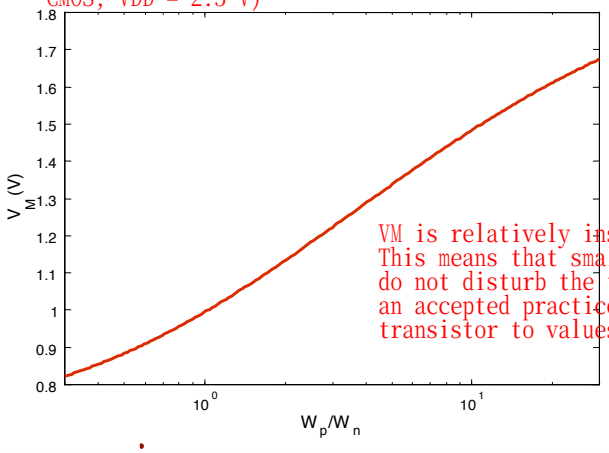


VTC of static CMOS inverter, derived from above figure (VDD = 2.5 V). For each operation region, the modes of the transistors are annotated — off, res(istive), or sat(urated).

The switching threshold, VM, is defined as the point where Vin = Vout. Its value can be obtained graphically from the intersection of the VTC with the line given by Vin = Vout. In this region, both PMOS and NMOS are always saturated, since VDS =VGS.

Switching Threshold as a Function of Transistor Ratio

Simulated inverter switching threshold versus PMOS/NMOS ratio (0.25 µm CMOS, VDD = 2.5 V)

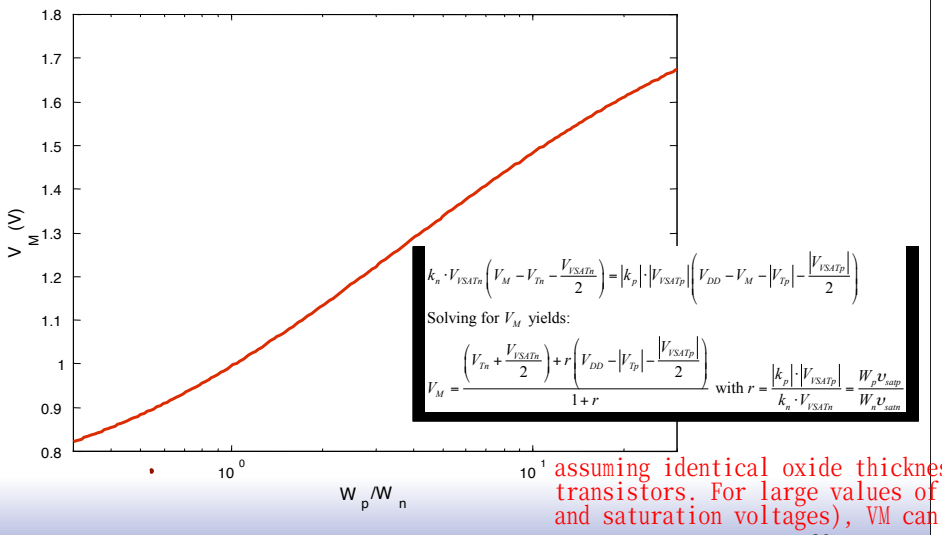


$$I_{dn}(V_M) = I_{dp}(V_M)$$

VM is relatively insensitive to variations in the device ratio. This means that small variations of the ratio (e.g., making it 3 or 2.5) do not disturb the transfer characteristic that much. It is therefore an accepted practice in industrial designs to set the width of the PMOS transistor to values smaller than those required for exact symmetry.

An analytical expression for VM is obtained by equating the currents through the transistors. We solve the case where the supply voltage is high so that the devices can be assumed to be velocity-saturated (or VDSAT < VM - VT). We furthermore ignore the channel length modulation effects

Switching Threshold as a Function of Transistor Ratio



$$k_n \cdot V_{VSATn}\left(V_M - V_{Tn} - \frac{V_{VSATn}}{2}\right) = |k_P| \cdot |V_{VSATp}|\left(V_{DD} - V_M - |V_{Tp}| - \frac{|V_{VSATp}|}{2}\right)$$

Solving for V_M yields:

$$V_M = \frac{\left(V_{Tn} + \frac{V_{VSATn}}{2}\right) + r\left(V_{DD} - |V_{Tp}| - \frac{|V_{VSATp}|}{2}\right)}{1+r} \quad \text{with } r = \frac{|k_P| \cdot |V_{VSATp}|}{k_n \cdot V_{VSATn}} = \frac{W_p v_{satp}}{W_n v_{satn}}$$

assuming identical oxide thicknesses for PMOS and NMOS transistors. For large values of VDD (compared to threshold and saturation voltages), VM can be simplified:

$$V_M \approx \frac{rV_{DD}}{1+r}$$

VM为1/2VDD时最好，此时噪声容限最大。此时r约等于1。
此时也要求：

$$(W/L)_p = (W/L)_n \times (V_{DSAT_n} k'_n)/(V_{DSAT_p} k'_p)$$

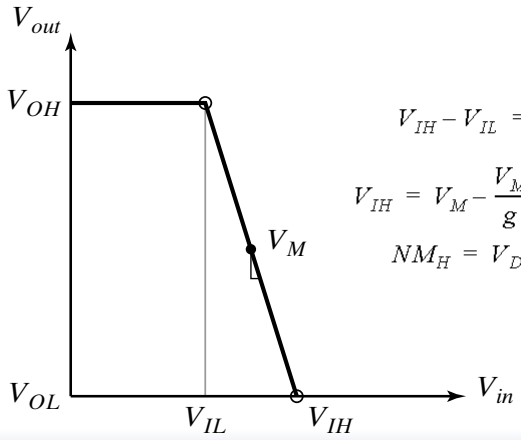
V_{IH}和V_{IL}的定义为 $\frac{dV_{out}}{dV_{in}} = -1$（增益g=-1）的operational points.

利用定义来求解VIH和VIL太过复杂，因此使用如下简化方法：

Determining V_{IH} and V_{IL}

引入了一点点错误，但这点错误在初始设计中是允许的。

A simplified approach



$$V_{IH} - V_{IL} = -\frac{(V_{OH} - V_{OL})}{g} = \frac{-V_{DD}}{g}$$

$$V_{IH} = V_M - \frac{V_M}{g} \qquad V_{IL} = V_M + \frac{V_{DD} - V_M}{g}$$

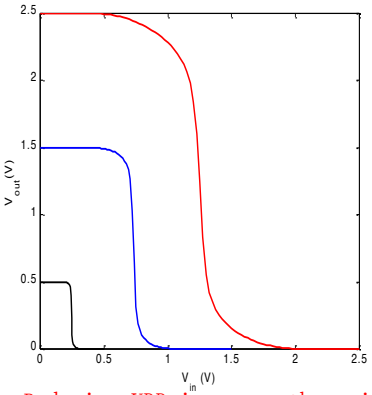
$$NM_H = V_{DD} - V_{IH} \qquad NM_L = V_{IL}$$

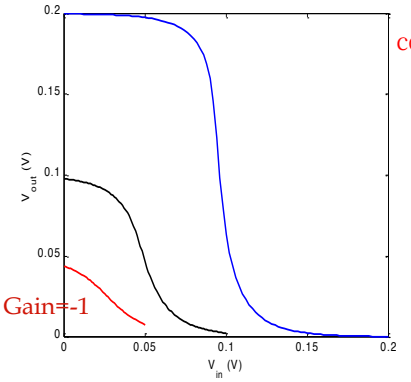
To achieving sufficient gain for use in a digital circuit, it is necessary that the supply must be at least a couple times ϕ_T=kT/q (=25 mV at room temperature), the thermal voltage introduced in Chapter 3

工作电压变低，能对VTC产生一定的正面影响，但是可能会导致电路对器件的参数和外部噪声变得敏感。

The gain of the inverter in the transition region actually increases with a reduction of the supply voltage!

Gain as a function of VDD





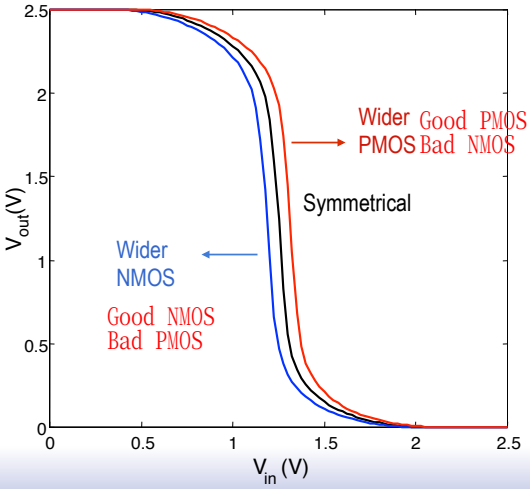
consider about the power

Reducing VDD improves the gain..

but it detoriates for very-low supply voltages.

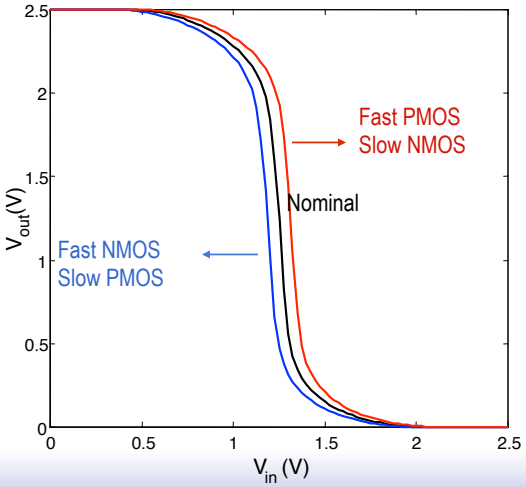
静态互补金属氧化物半导体反相器的直流特性对这些变化相当不敏感，栅极在很宽的工作条件范围内保持功能。这些变化指工作温度范围比较大，器件实际制造出来的参数和标称值有差距。例如器件的大小等参数变化。

Impact of Sizing



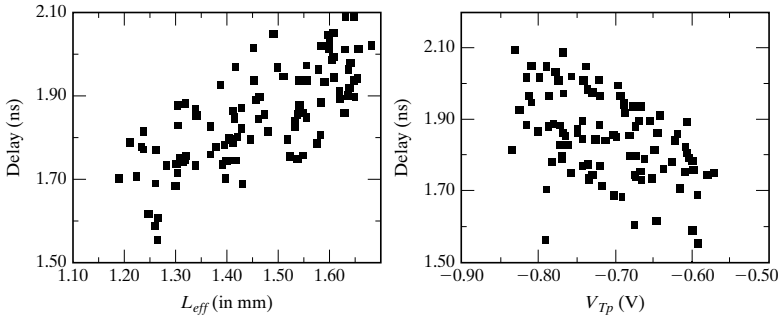
Impact of device variations on static CMOS inverter VTC. The "good" device has a smaller oxide thickness (- 3nm), a smaller length (-25 nm), a higher width (+30 nm), and a smaller threshold (-60 mV). The opposite is true for the "bad" transistor.

Impact of Process Variations



Process Variations

Not all transistors are alike
Impacts parameters such as reliability and performance



Define process corners: SS, FF, SF, FS