دانشگاه صنعتی امیرکبیر

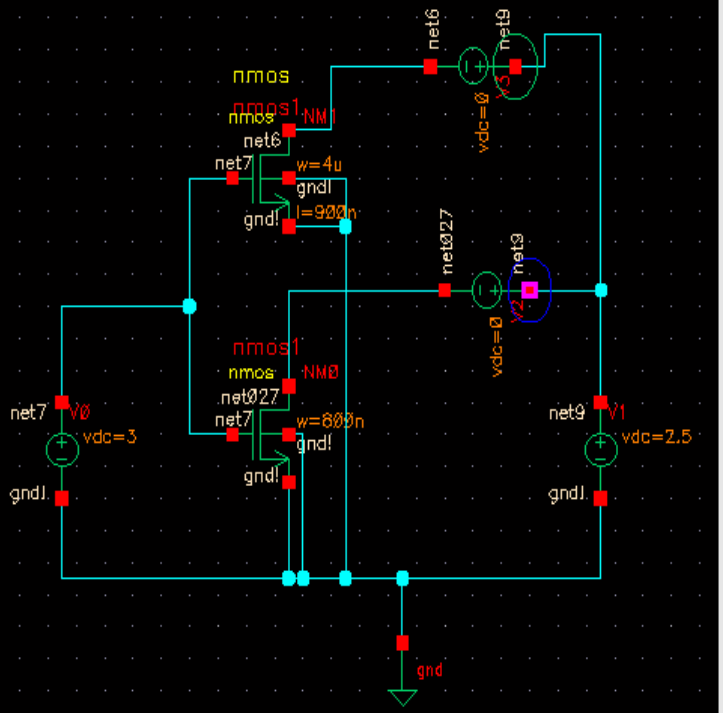
دانشکده مهندسی برق

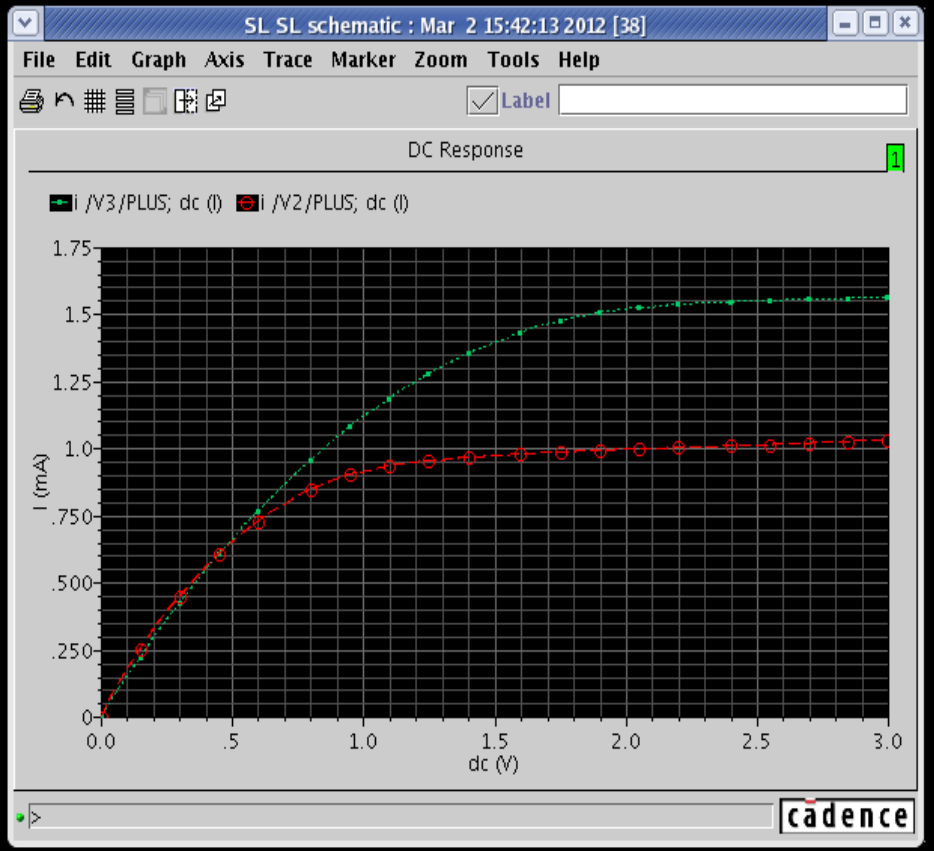
طراحی مدارهای VLSI

فصل چهارم: واورنگر CMOS

مجید شالچیان

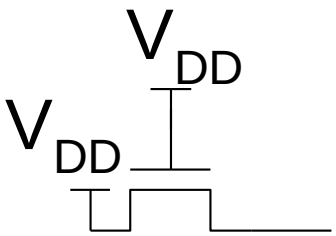
majid.shalchian@gmail.com


nmos
nmos1
nmos
NM1
net6
net7
w=4u
gnd!
l=900n
gnd!
net6
net9
vdc=0
V3
net027
net9
vdc=0
V2
nmos1
nmos
NM0
net027
net7
w=600n
gnd!
gnd!
net7
V0
vdc=3
gnd!
net9
V1
vdc=2.5
gnd!
gnd


SL SL schematic : Mar 2 15:42:13 2012 [38]
File Edit Graph Axis Trace Marker Zoom Tools Help
Label
DC Response
1
i /V3/PLUS; dc (I)
i /V2/PLUS; dc (I)
1.75
1.5
1.25
1.0
.750
.500
.250
0
I (mA)
0.0
.5
1.0
1.5
2.0
2.5
3.0
dc (V)
cadence

ترانزیستور انتقالی

- در ترانزیستور انتقالی مقدار ولتاژ خروجی چقدر است؟
- فرض کنیم خروجی برابر VDD باشد. در این حالت چون



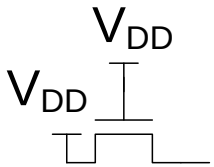
VGS=0<Vt است ترانزیستور قطع می شود.

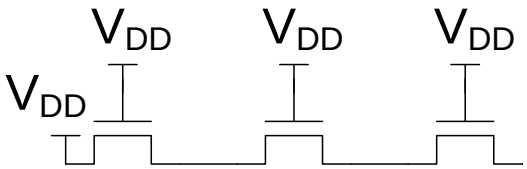
پس برای آنکه ترانزیستور روشن بماند حداکثر ولتاژ خروجی آن

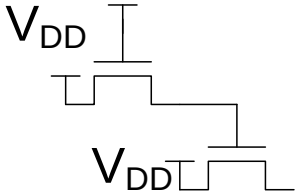
برابر VDD-Vt خواهد شد.

به همین دلیل می گوییم ترانزیستور nMOS مقدار منطقی ۱ را خوب عبور نمی دهد.

مثال



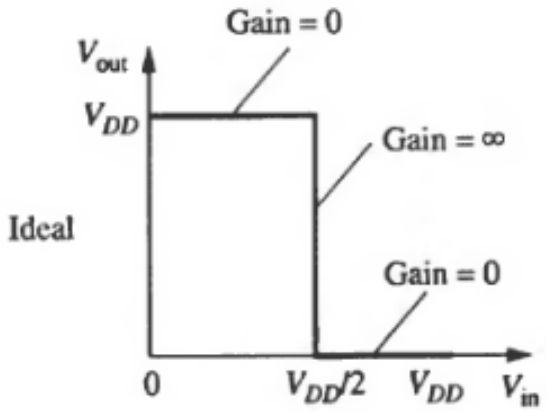




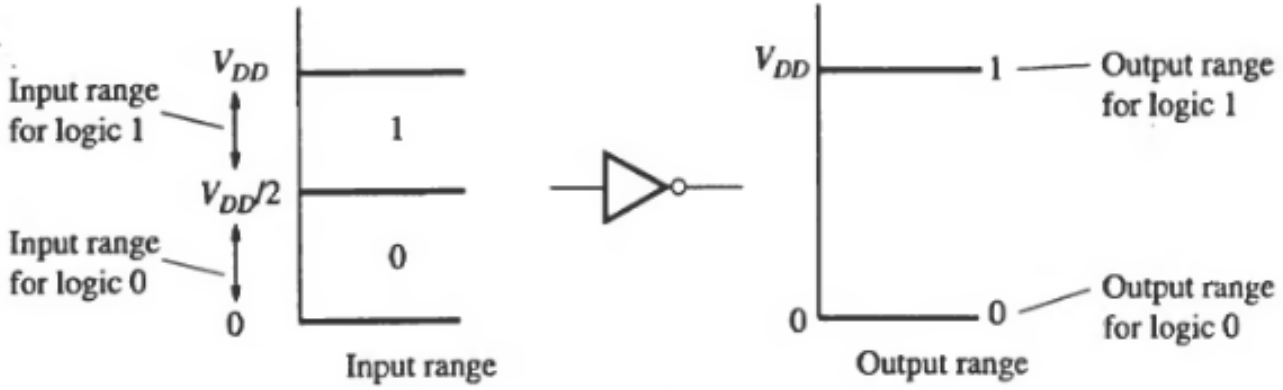
وارونگر ایستا

- مقدمه
- مشخصه انتقالی
- رفتار ایستا
- رفتار پویا
- تحلیل توان، انرژی و طراحی برای مصرف کم

مقدمه: وارونگر ایده ال

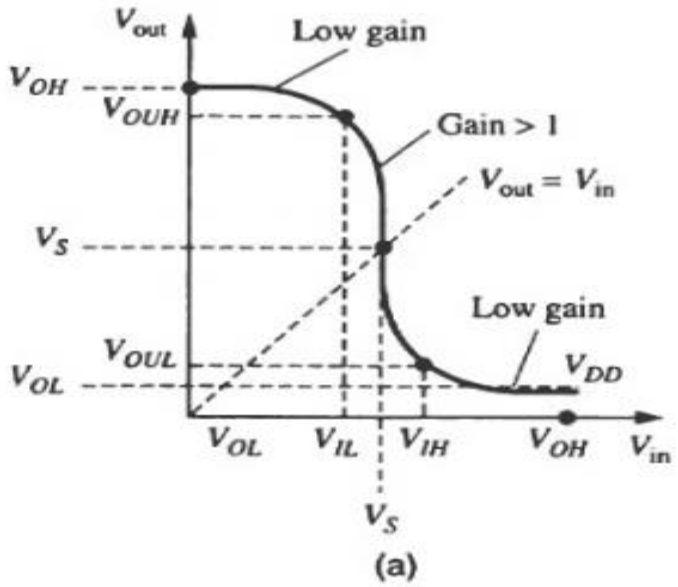


(a)

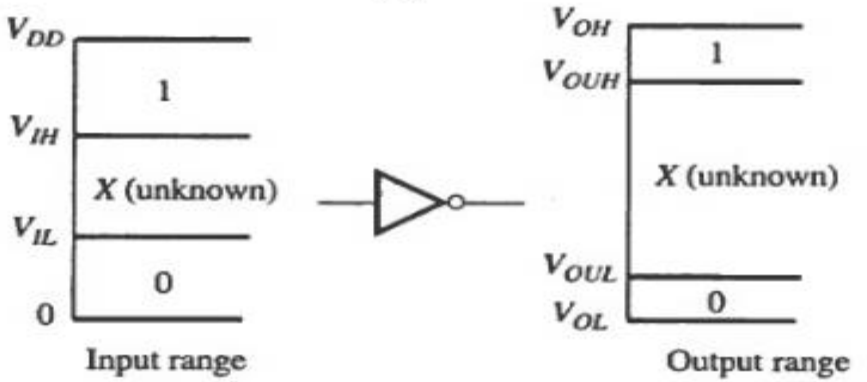


(b)

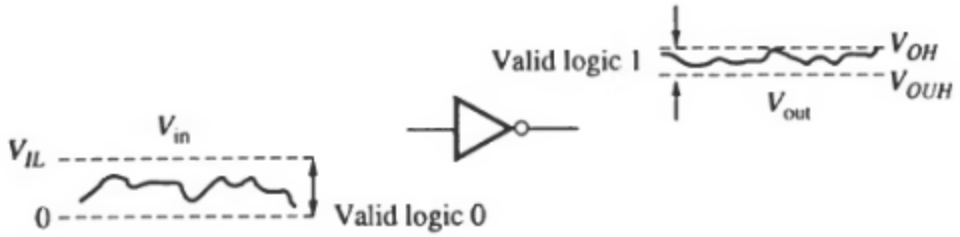
مشخصه واقعی تر وارونگر



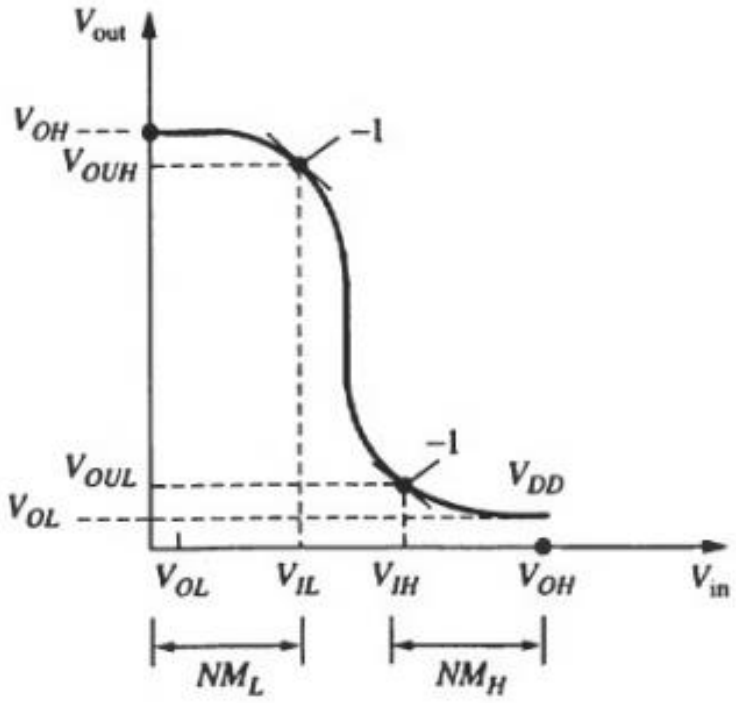
(a)



(b)



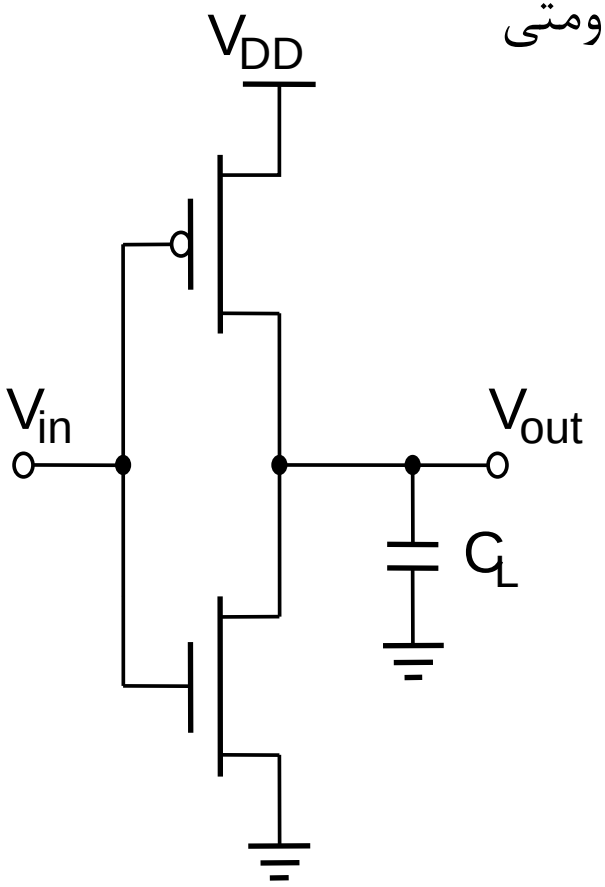
یادآوری حاشیه نویز



$$NM_H = V_{OH} - V_{IH}$$

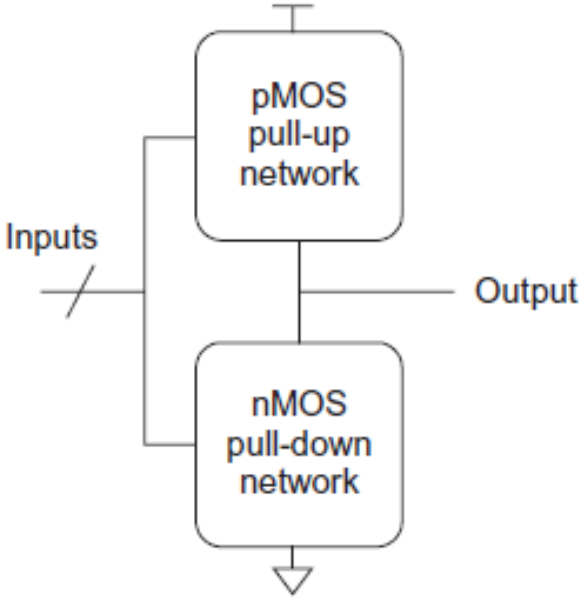
$$NM_L = V_{IL} - V_{OL}$$

پیاده سازی با منطق CMOS ایستا



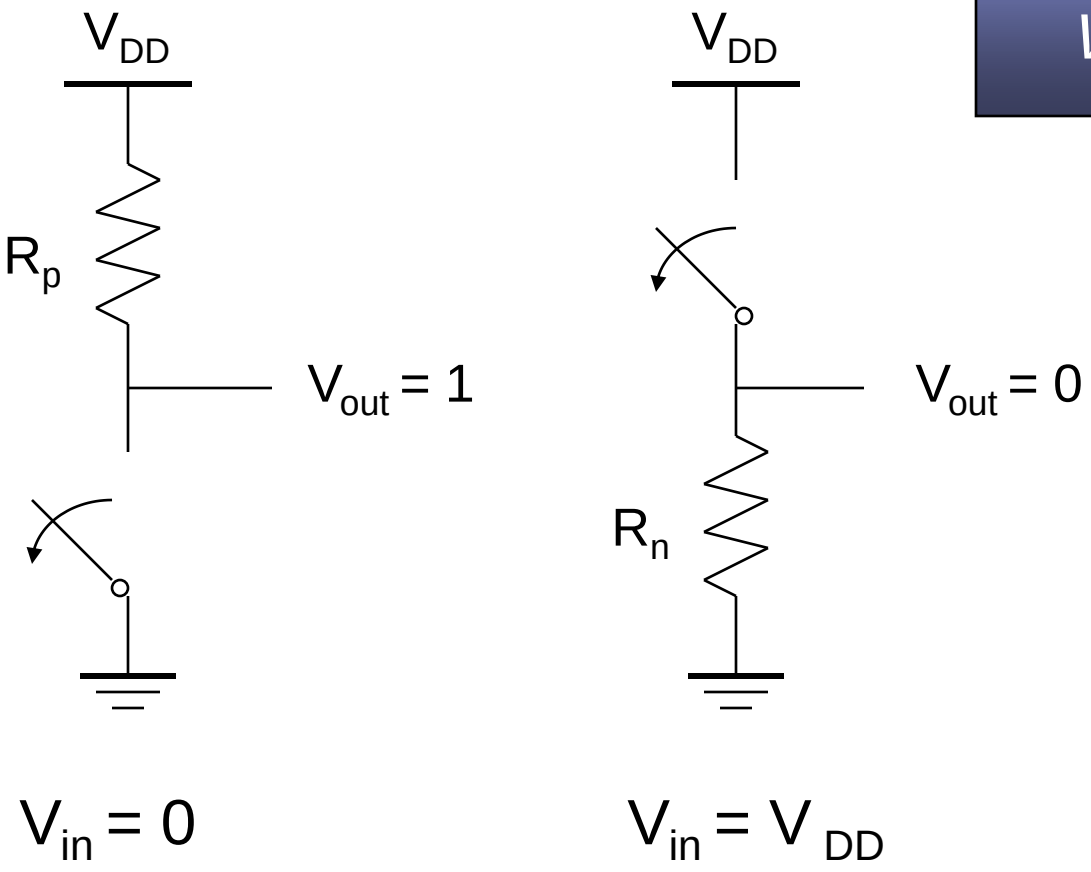
- در منطق CMOS ایستا خروجی همواره یک مسیر مقاومتی به VDD یا زمین دارد.

- وارونگر پایه مدار های منطقی ترکیبی CMOS ایستا است.



آنالیز مرتبه اول حالت ایستا

$V_{OL} = 0$
$V_{OH} = V_{DD}$



ویژگیهای CMOS

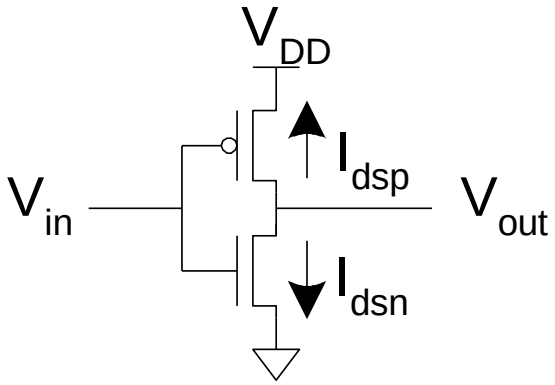
- Full rail-to-rail swing $\Rightarrow$ high noise margins
  - Logic levels not dependent upon the relative device sizes $\Rightarrow$ transistors can be minimum size $\Rightarrow$ ratioless
- Always a path to V_{dd} or GND in steady state $\Rightarrow$ low output impedance (output resistance in kΩ range) $\Rightarrow$ large fan-out (albeit with degraded performance)
- Extremely high input resistance (gate of MOS transistor is near perfect insulator) $\Rightarrow$ nearly zero steady-state input current
- No direct path steady-state between power and ground $\Rightarrow$ no static power dissipation
- Propagation delay function of load capacitance and resistance of transistors

وارونگر ایستا

- مقدمه
- مشخصه انتقالی
- رفتار ایستا
- رفتار پویا
- تحلیل توان، انرژی و طراحی برای مصرف کم

تحلیل عملکرد وارونگر

- DC Response: V_{out} vs. V_{in} for a gate
- Ex: Inverter
  - When $V_{in} = 0$ -> $V_{out} = V_{DD}$
  - When $V_{in} = V_{DD}$ -> $V_{out} = 0$
  - In between, V_{out} depends on transistor size and current
  - By KCL, must settle such that $I_{dsn} = |I_{dsp}|$
  - We could solve equations
  - But graphical solution gives more insight

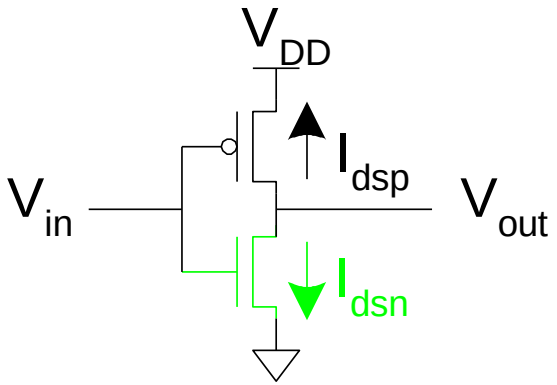


ناحیه عملکرد ترانزیستور ها

- Current depends on region of transistor behavior
- For what V_{in} and V_{out} are nMOS and pMOS in
  - Cutoff?
  - Linear?
  - Saturation?

nMOS Operation

Cutoff	Linear	Saturated
V_{gsn} <	V_{gsn} >	V_{gsn} >
	V_{dsn} <	V_{dsn} >



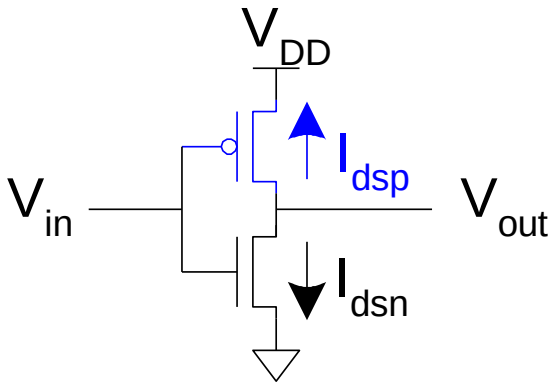
pMOS Operation

Cutoff	Linear	Saturated
$V_{gsp} > V_{tp}$ $V_{in} > V_{DD} + V_{tp}$	$V_{gsp} < V_{tp}$ $V_{in} < V_{DD} + V_{tp}$ $V_{dsp} > V_{gsp} - V_{tp}$ $V_{out} > V_{in} - V_{tp}$	$V_{gsp} < V_{tp}$ $V_{in} < V_{DD} + V_{tp}$ $V_{dsp} < V_{gsp} - V_{tp}$ $V_{out} < V_{in} - V_{tp}$

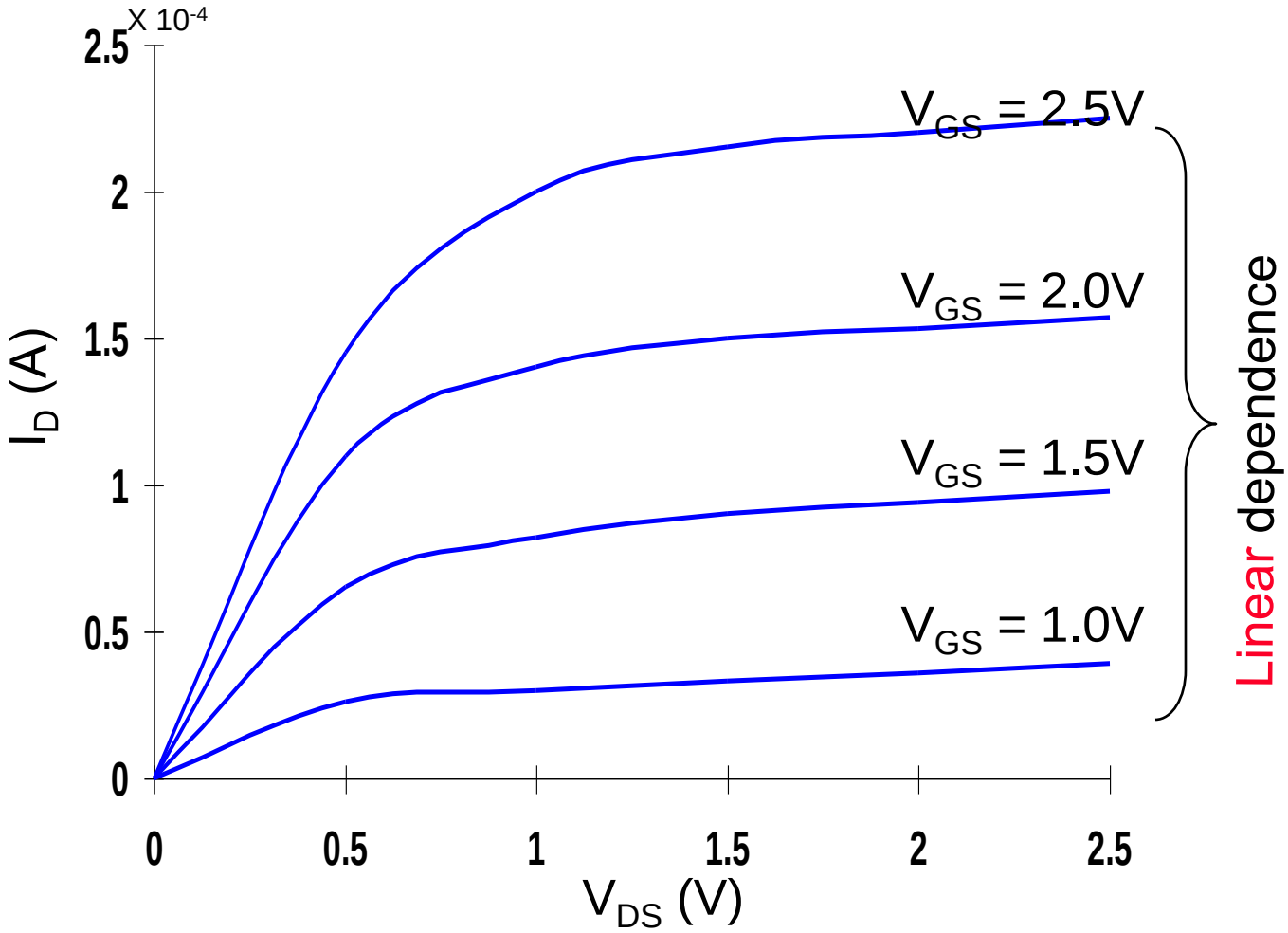
$V_{gsp} = V_{in} - V_{DD}$

$V_{dsp} = V_{out} - V_{DD}$

$V_{tp} < 0$



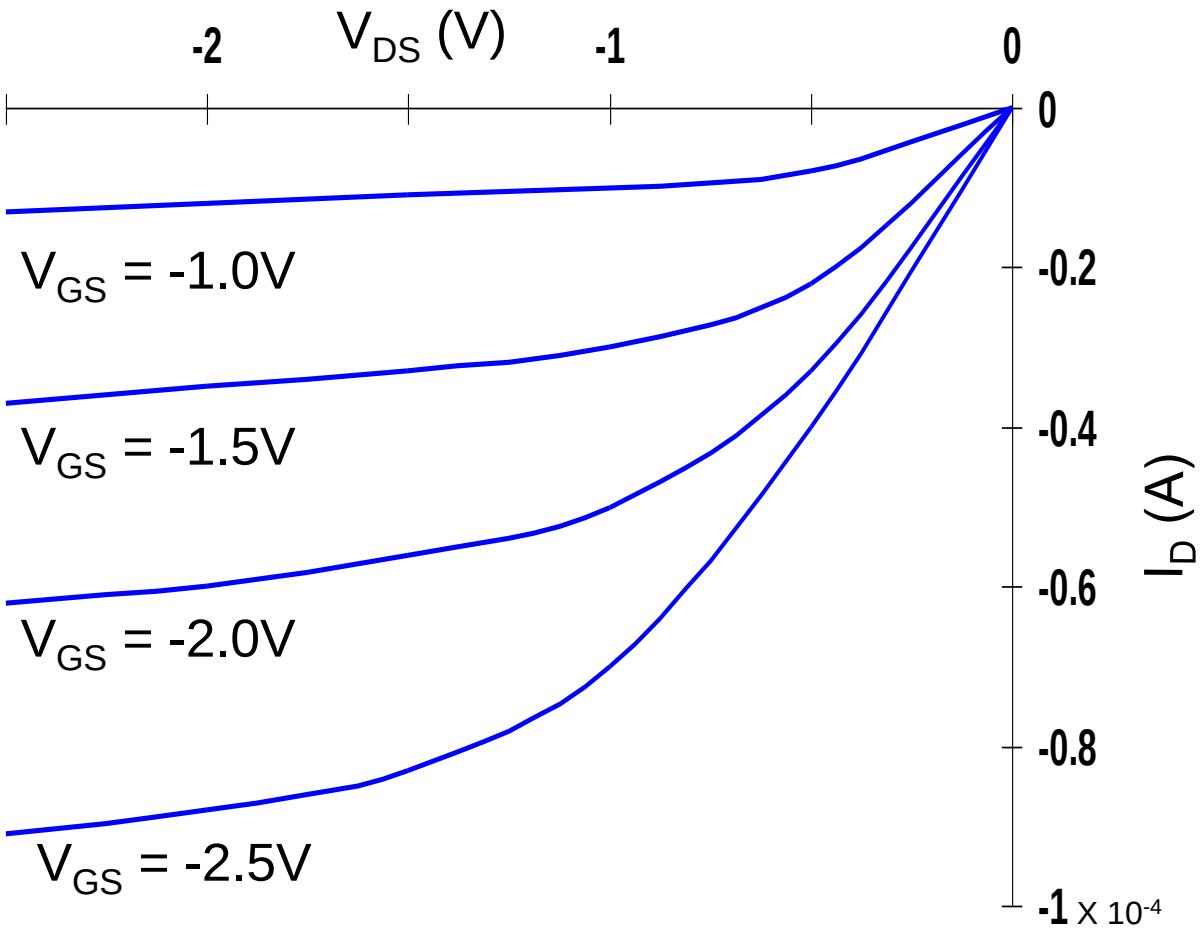
Review: Short Channel I-V Plot (NMOS)



NMOS transistor, 0.25um, L_d = 0.25um, W/L = 1.5, V_{DD} = 2.5V, V_T = 0.4V

Review: Short Channel I-V Plot (PMOS)

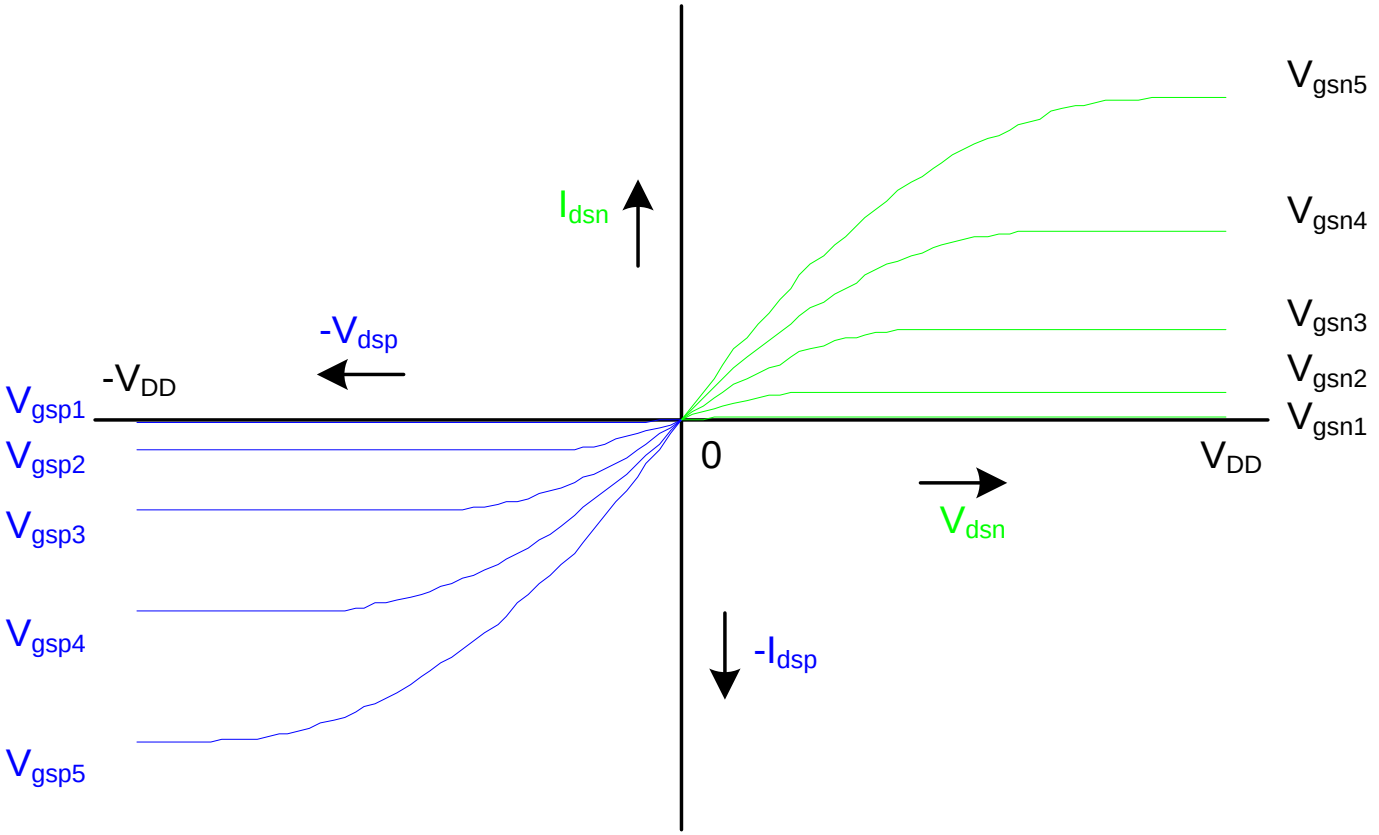
- All polarities of all voltages and currents are reversed



PMOS transistor, 0.25um, L_d = 0.25um, W/L = 1.5, V_{DD} = 2.5V, V_T = -0.4V

مشخصه جریان – ولتاژ ترانزیستور ها در یک محور

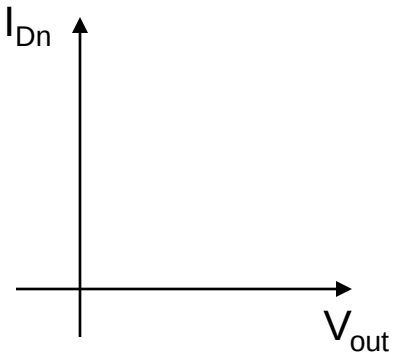
- Make pMOS is wider than nMOS such that $\beta_n = \beta_p$

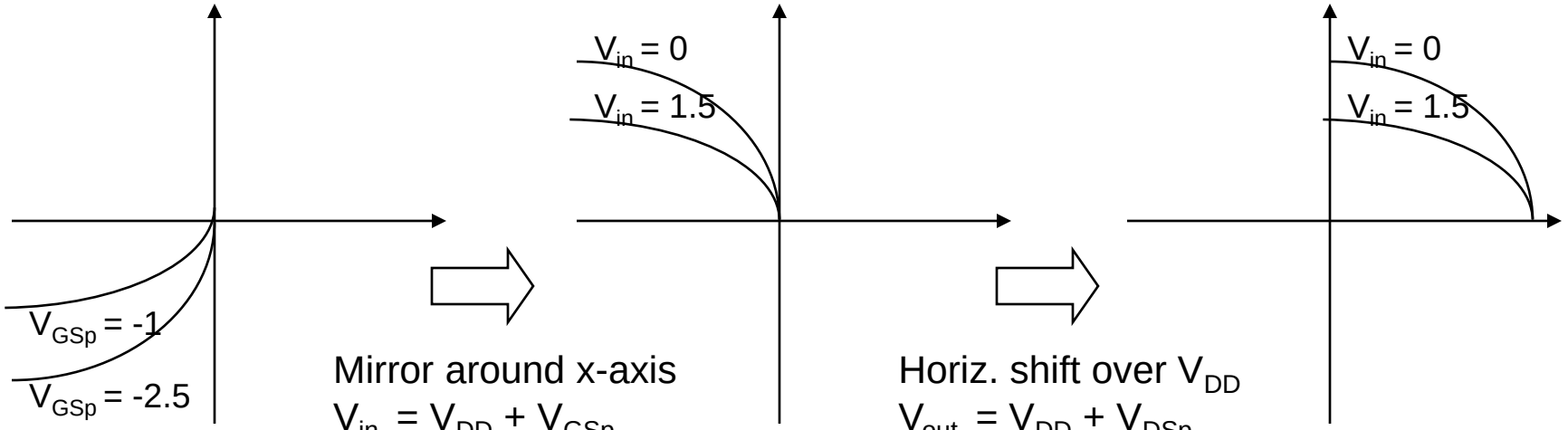


تبدیل مختصات مشخصه pMOS I-V

- Want common coordinate set V_{in}, V_{out}, and I_{Dn}

$$I_{DSp} = -I_{DSn}$$
$$V_{GSn} = V_{in} \; ; V_{GSp} = V_{in} - V_{DD}$$
$$V_{DSn} = V_{out} \; ; V_{DSp} = V_{out} - V_{DD}$$





Mirror around x-axis

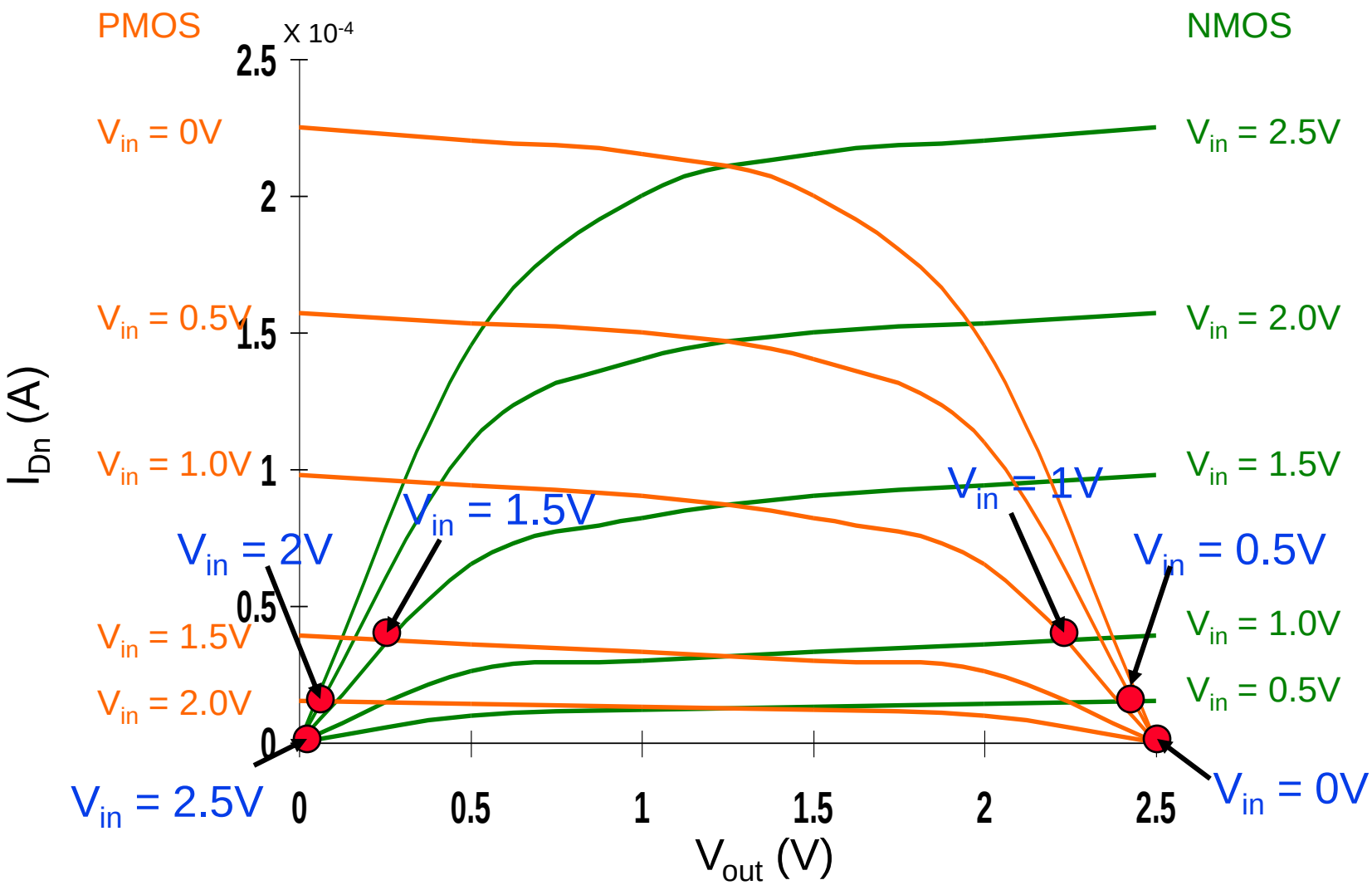
$V_{in} = V_{DD} + V_{GSp}$

$I_{Dn} = -I_{Dp}$

Horiz. shift over V_{DD}

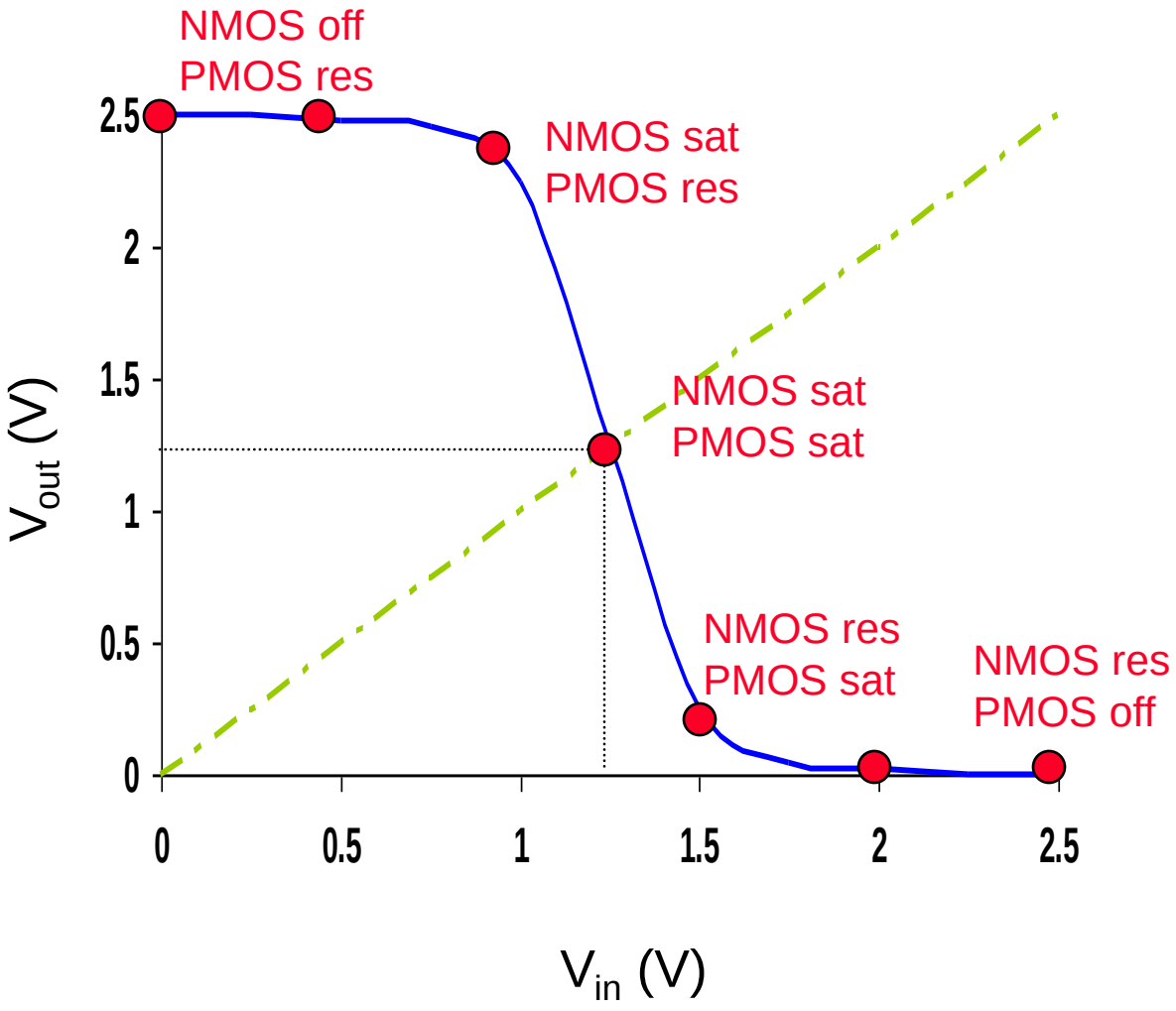
$V_{out} = V_{DD} + V_{DSp}$

CMOS Inverter Load Lines



0.25um, $W/L_n = 1.5$, $W/L_p = 4.5$, $V_{DD} = 2.5V$, $V_{Tn} = 0.4V$, $V_{Tp} = -0.4V$

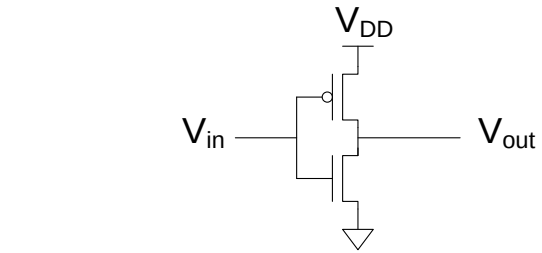
CMOS Inverter VTC

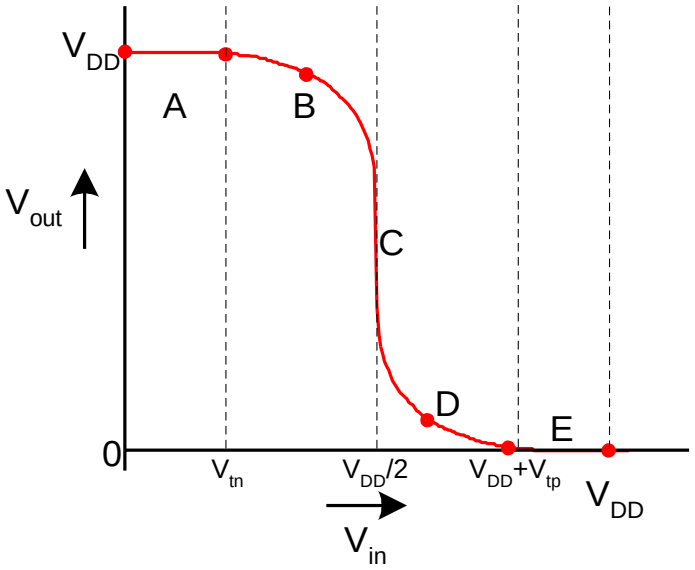


Operating Regions

- Revisit transistor operating regions

Region	nMOS	pMOS
A		
B		
C		
D		
E		



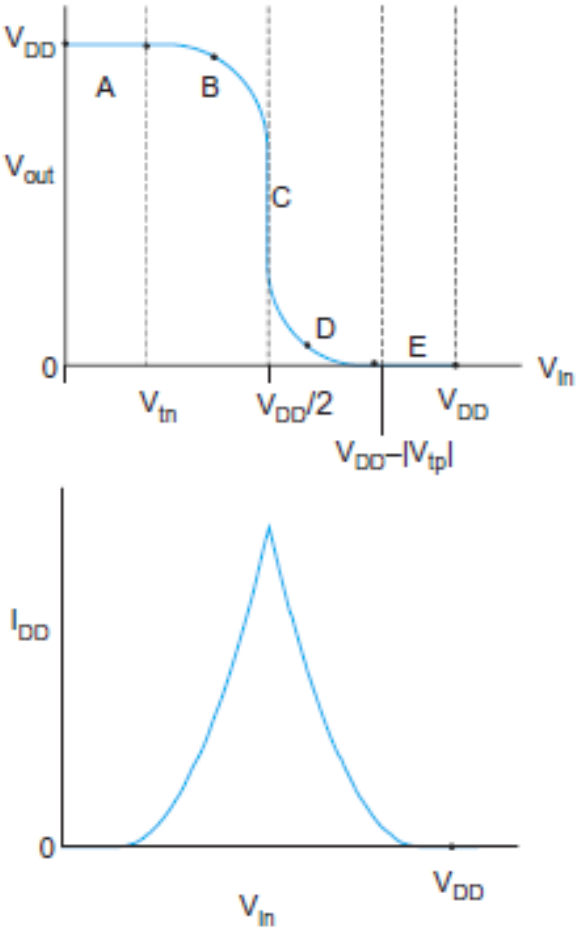


نواحی کار وارونگر با ساختار متقارن

$$k_n^{'} = k_p^{'}$$

$$V_{tn} = -V_{tp}$$

Region	Condition	p-device	n-device	Output
A	$0 \le V_{in} < V_{tn}$	linear	cutoff	$V_{out} = V_{DD}$
B	$V_{tn} \le V_{in} < V_{DD}/2$	linear	saturated	$V_{out} > V_{DD}/2$
C	$V_{in} = V_{DD}/2$	saturated	saturated	V_{out} drops sharply
D	$V_{DD}/2 < V_{in} \le V_{DD} - \lvert V_{tp}\rvert$	saturated	linear	$V_{out} < V_{DD}/2$
E	$V_{in} > V_{DD} - \lvert V_{tp}\rvert$	cutoff	linear	$V_{out} = 0$



وارونگر ایستا

- مقدمه
- مشخصه انتقالی
- رفتار ایستا
- رفتار پویا
- تحلیل توان، انرژی و طراحی برای مصرف کم

هدف

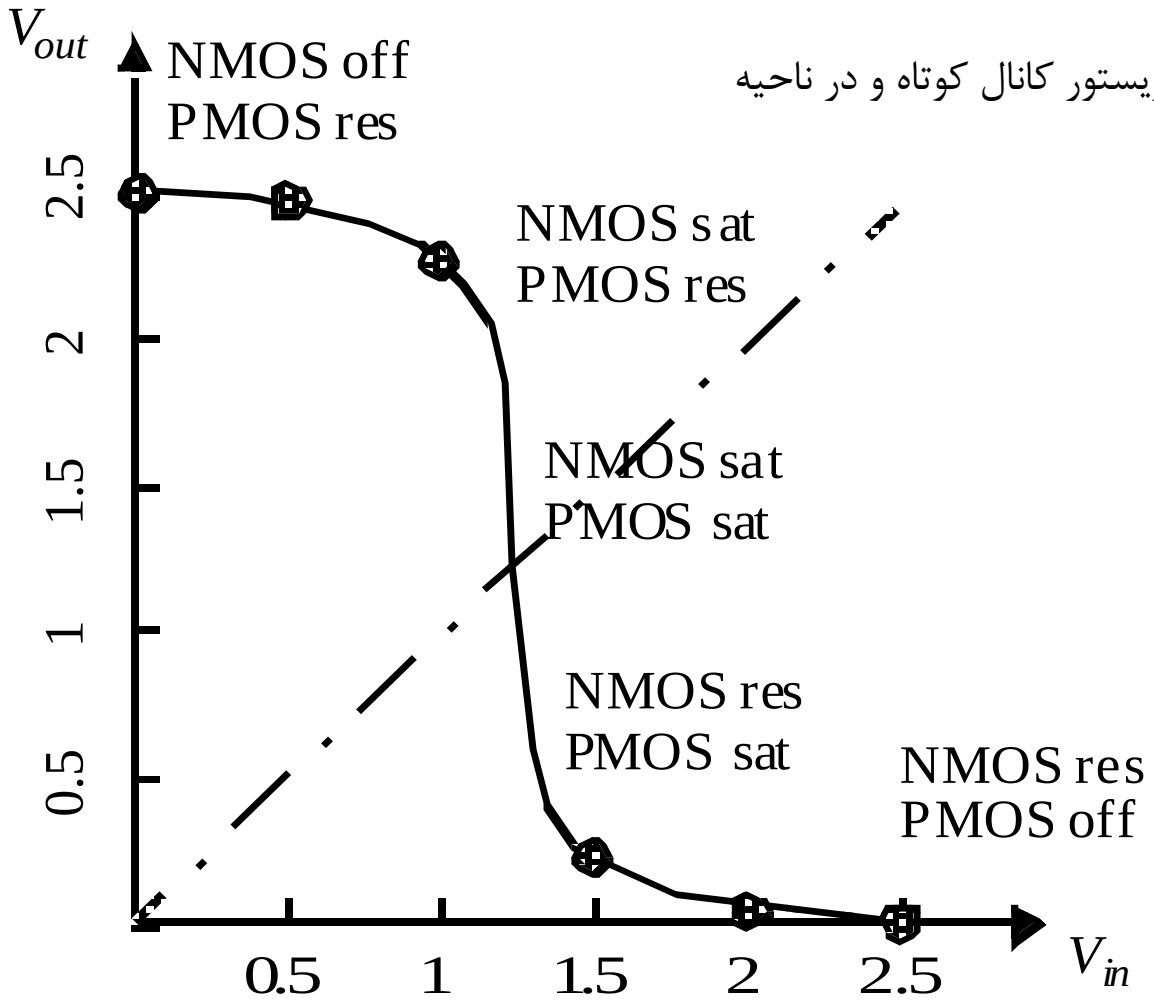
هدف از تحلیل ایستایی تحلیل پارامتر های استحکام وارونگر نظیر حاشبه نویز و بررسی اثر تغییر پارامتر های ابعادی روی مشخصه انتقالی ترانزیستور می باشد.

معمولا می خواهیم حاشیه نویز بیشینه و مشخصه انتقالی متقارن داشته باشیم

تحلیل رفتار ایستایی

۱– تحلیل نقطه آستانه سوییچینگ VM (Vin=Vout)

فرض می کنیم هر دو ترانزیستور کانال کوتاه و در ناحیه اشباع سرعت هستند.



تحلیل آستانه سوییچینگ

$$k_n V_{DSATn}\left(V_M - V_{Tn} - \frac{V_{DSATn}}{2}\right) + k_p V_{DSATp}\left(V_M - V_{DD} - V_{Tp} - \frac{V_{DSATp}}{2}\right) = 0$$

Solving for V_M yields

$$V_M = \frac{\left(V_{Tn} + \frac{V_{DSATn}}{2}\right) + r\left(V_{DD} + V_{Tp} + \frac{V_{DSATp}}{2}\right)}{1 + r} \quad \text{with} \quad r = \frac{k_p V_{DSATp}}{k_n V_{DSATn}} = \frac{\upsilon_{satp} W_p}{\upsilon_{satn} W_n}$$

اگر فرض کنیم مقدار VDD نسبت به ولتاژ آستانه و ولتاژ اشباع سرعت ترانزیستور ها بحد کافی بزرگ باشد.

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

R بیانگر قدرت درایو نسبی pmos نسبت به nmos است. اگر pMOS قویتر باشد VM از VDD/2زیادتر می شود و اگر NMOS قویتر باشد کمتر می شود.

شرط تقارن

$V_M = V_{DD}/2$ (to have comparable high and low noise margins), so want $r \approx 1$

$$\frac{(W/L)_p}{(W/L)_n} = \frac{k'_n V_{DSATn}(V_M - V_{Tn} - V_{DSATn}/2)}{k'_p V_{DSATp}(V_{DD} - V_M + V_{Tp} + V_{DSATp}/2)}$$

مثال آستانه سوییچینگ

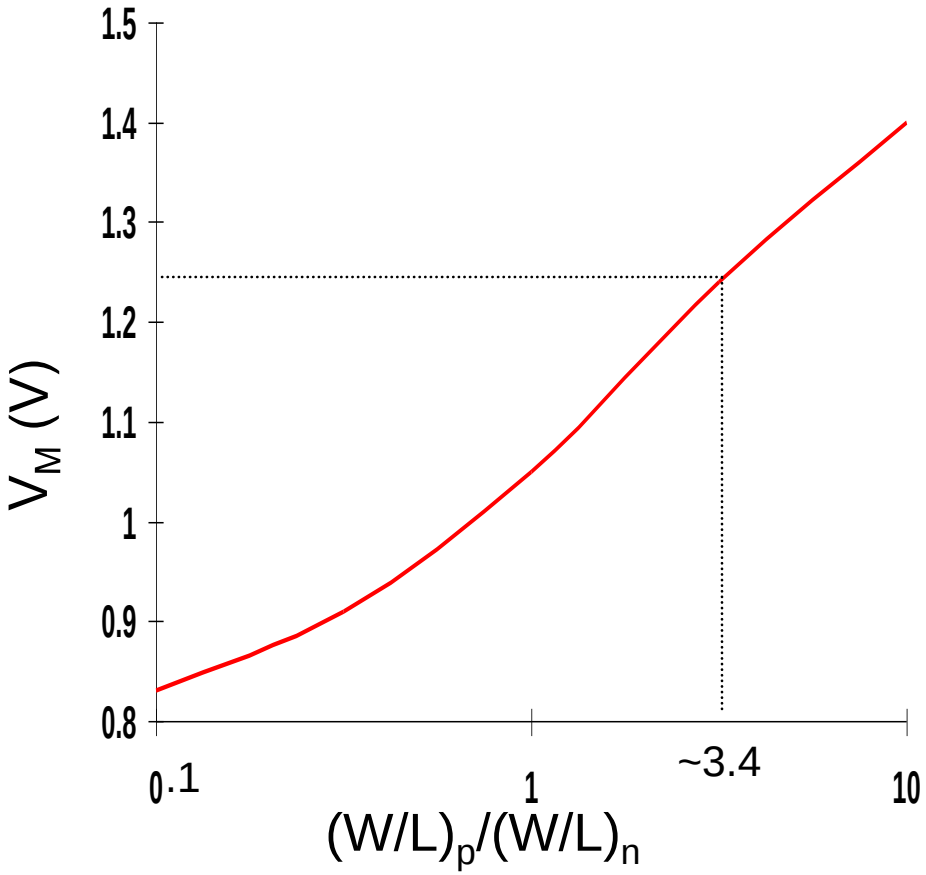
- In our generic 0.25 micron CMOS process, using the process parameters from slide L03.25, a V_{DD} = 2.5V, and a minimum size NMOS device ($(W/L)_n$ of 1.5)

	$V_{T0}(V)$	$\gamma(V^{0.5})$	$V_{DSAT}(V)$	$k'(A/V^2)$	$\lambda(V^{-1})$
NMOS	0.43	0.4	0.63	115×10^{-6}	0.06
PMOS	-0.4	-0.4	-1	-30×10^{-6}	-0.1

$$\frac{(W/L)_p}{(W/L)_n} = \frac{115 \times 10^{-6}}{-30 \times 10^{-6}} \times \frac{0.63}{-1.0} \times \frac{(1.25 - 0.43 - 0.63/2)}{(1.25 - 0.4 - 1.0/2)} = 3.5$$

$$(W/L)_p = 3.5 \times 1.5 = 5.25 \text{ for a } V_M \text{ of } 1.25V$$

شبیه سازی بستگی آستانه سوییچ کردن به ابعاد

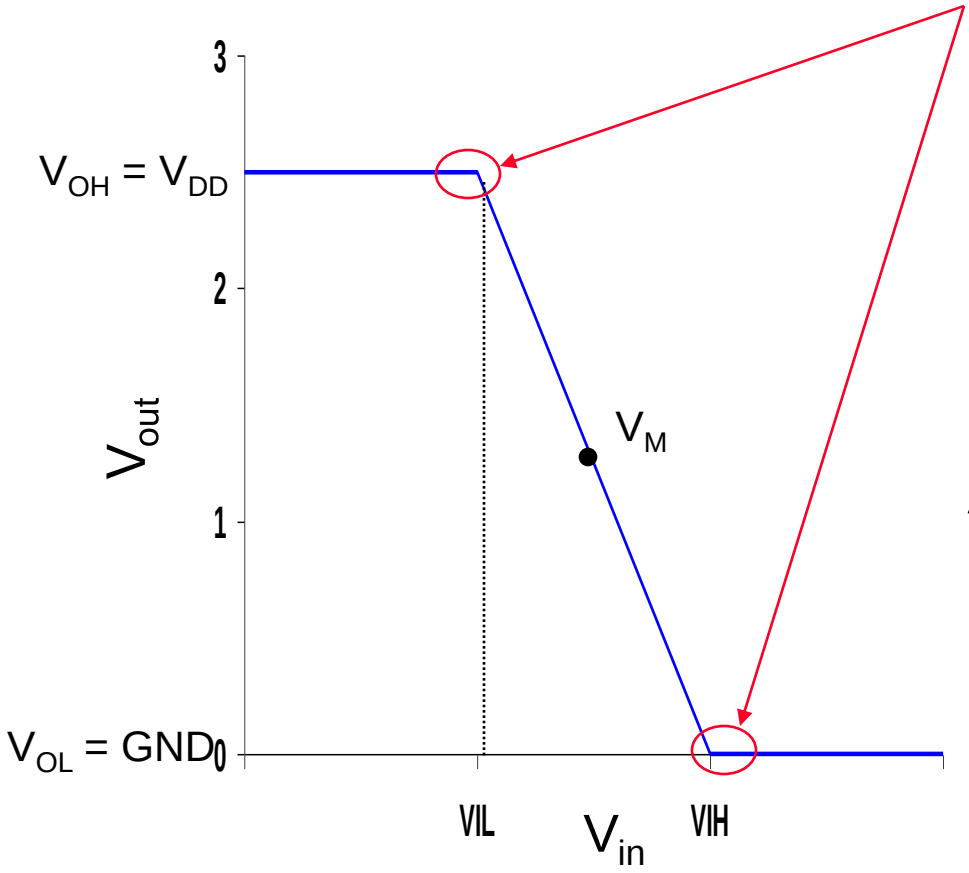


Note: x-axis is semilog

- V_M is relatively insensitive to variations in device ratio
  - setting the ratio to 3, 2.5 and 2 gives V_M's of 1.22V, 1.18V, and 1.13V
- Increasing the width of the PMOS moves V_M towards V_{DD}
- Increasing the width of the NMOS moves V_M toward GND

تحلیل رفتار ایستایی

۲– تحلیل حاشیه نویز



A piece-wise linear approximation of VTC

By definition, V_{IH} and V_{IL} are where $dV_{out}/dV_{in} = -1$ (= gain)

$NM_H = V_{DD} - V_{IH}$
$NM_L = V_{IL} - GND$

Approximating:

$V_{IH} = V_M - V_M/g$
$V_{IL} = V_M + (V_{DD} - V_M)/g$

So high gain in the transition region is very desirable

تحلیل با تقریب

محاسبه گین در نقطه VM

$$k_n V_{DSATn}\left(V_{in} - V_{Tn} - \frac{V_{DSATn}}{2}\right)(1 + \lambda_n V_{out}) +$$

$$k_p V_{DSATp}\left(V_{in} - V_{DD} - V_{Tp} - \frac{V_{DSATp}}{2}\right)(1 + \lambda_p V_{out} - \lambda_p V_{DD}) = 0$$

با مشتق گرفتن از معادله فوق نسبت به Vin

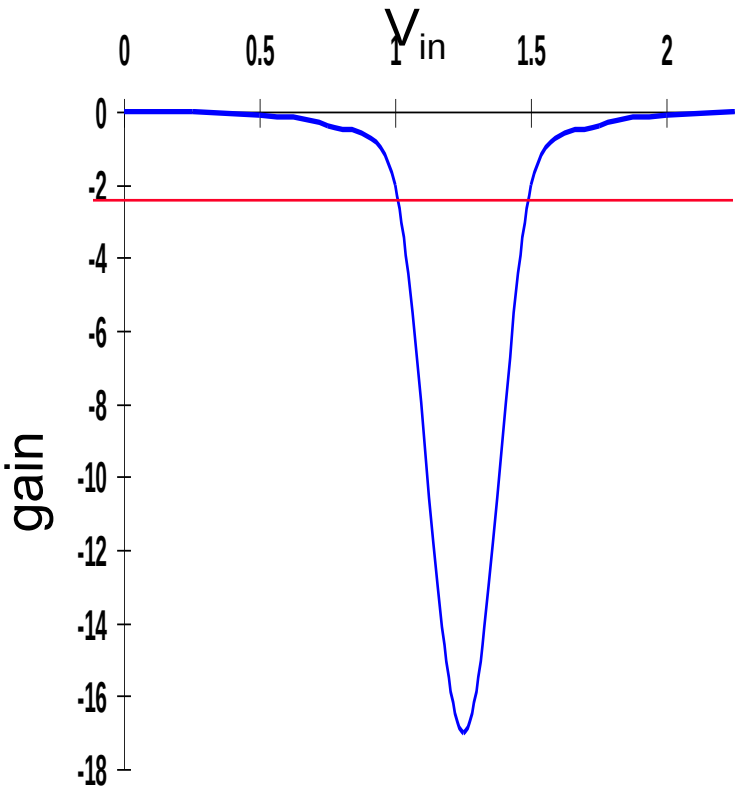
$$\frac{dV_{out}}{dV_{in}} = -\frac{k_n V_{DSATn}(1 + \lambda_n V_{out}) + k_p V_{DSATp}(1 + \lambda_p V_{out} - \lambda_p V_{DD})}{\lambda_n k_n V_{DSATn}(V_{in} - V_{Tn} - V_{DSATn}/2) + \lambda_p k_p V_{DSATp}(V_{in} - V_{DD} - V_{Tp} - V_{DSATp}/2)}$$

تحلیل با تقریب

با جایگذاری $V_{in} = V_M$.

$$g = -\frac{1}{I_D(V_M)}\frac{k_n V_{DSATn} + k_p V_{DSATp}}{\lambda_n - \lambda_p}$$

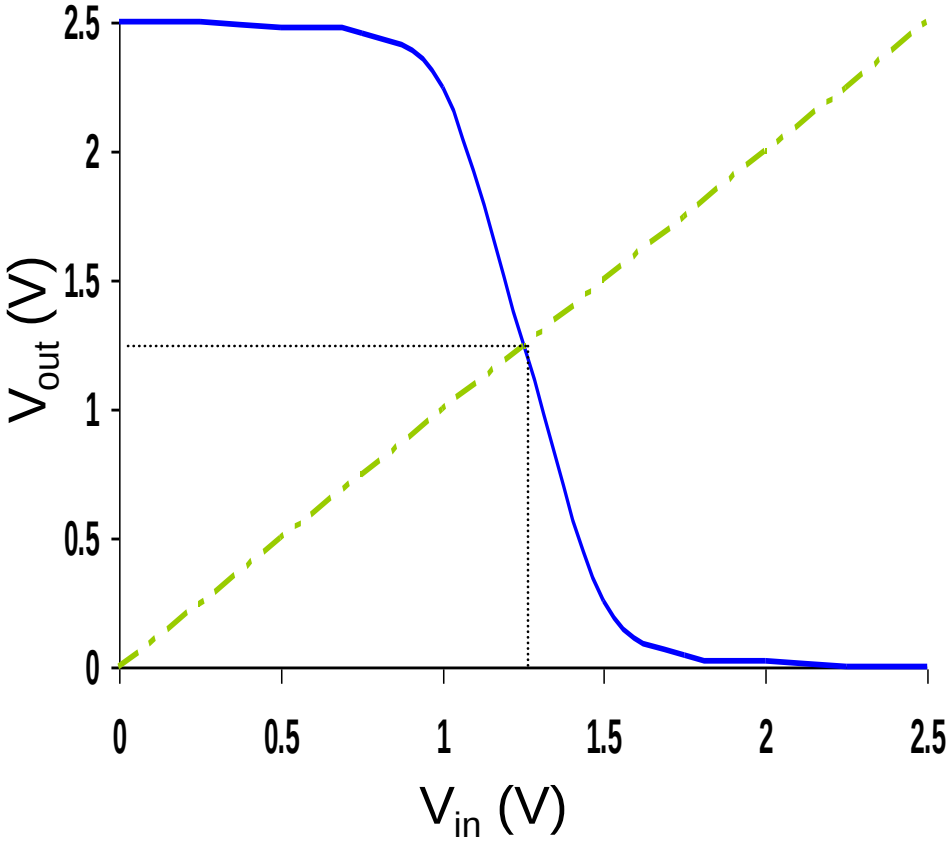
$$\approx \frac{1+r}{(V_M - V_{Tn} - V_{DSATn}/2)(\lambda_n - \lambda_p)}$$



بستگی شدید به پارامتر های تکنولوژی نظیر ضریب مدولاسیون طول کانال، ولتاژ آستانه و ولتاژ اشباع دارد.

مقش طراح انتخاب V_M با سایز بندی مناسب ترانزیستور ها است.

مثال محاسبه حاشیه نویز با استفاده از گین



0.25um, $(W/L)_p/(W/L)_n = 3.4$

$(W/L)_n = 1.5$ (min size)

$V_{DD} = 2.5V$

$V_M \approx 1.25V$, g = -27.5

$V_{IL} = 1.2V$, $V_{IH} = 1.3V$

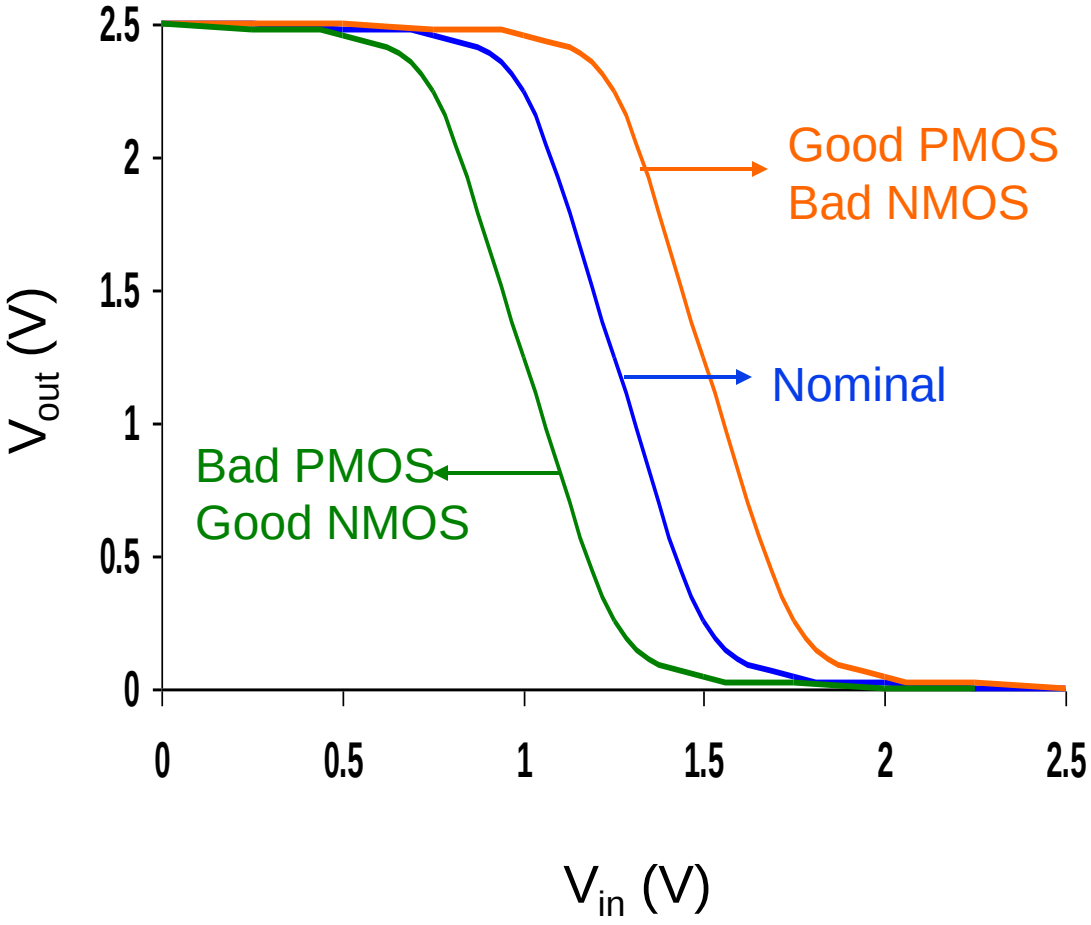
$NM_L = NM_H = 1.2$

(actual values are

$V_{IL} = 1.03V$, $V_{IH} = 1.45V$

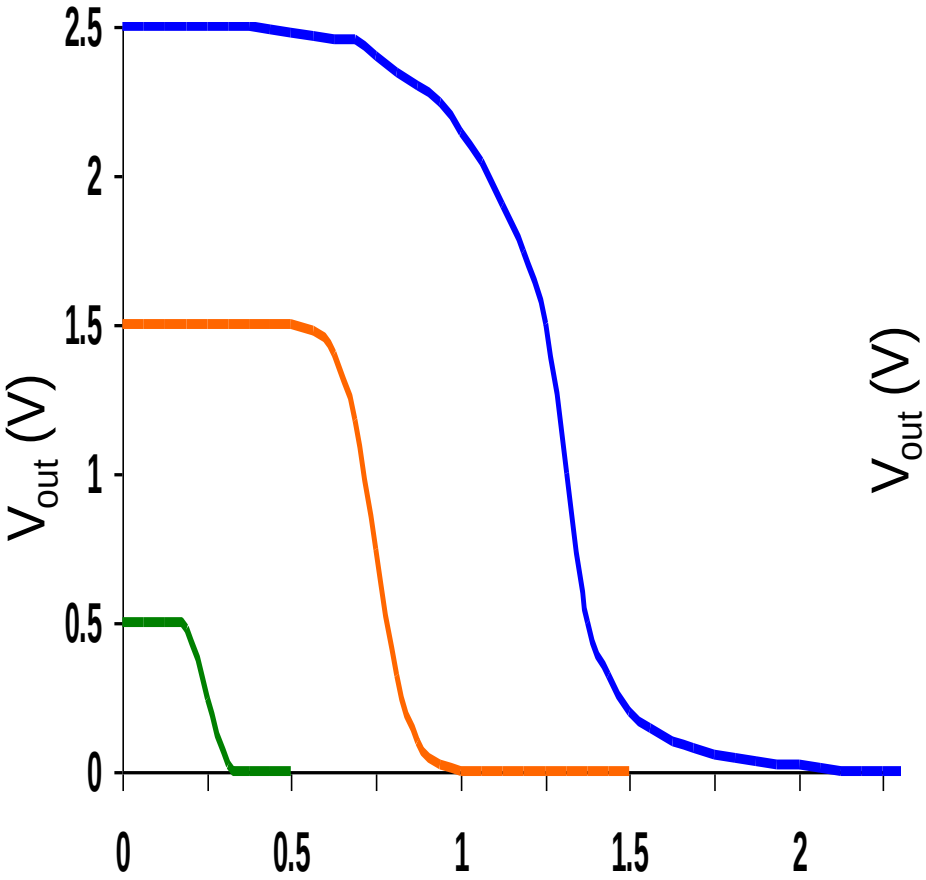
$NM_L = 1.03V$ & $NM_H = 1.05V$)

Impact of Process Variation on VTC Curve

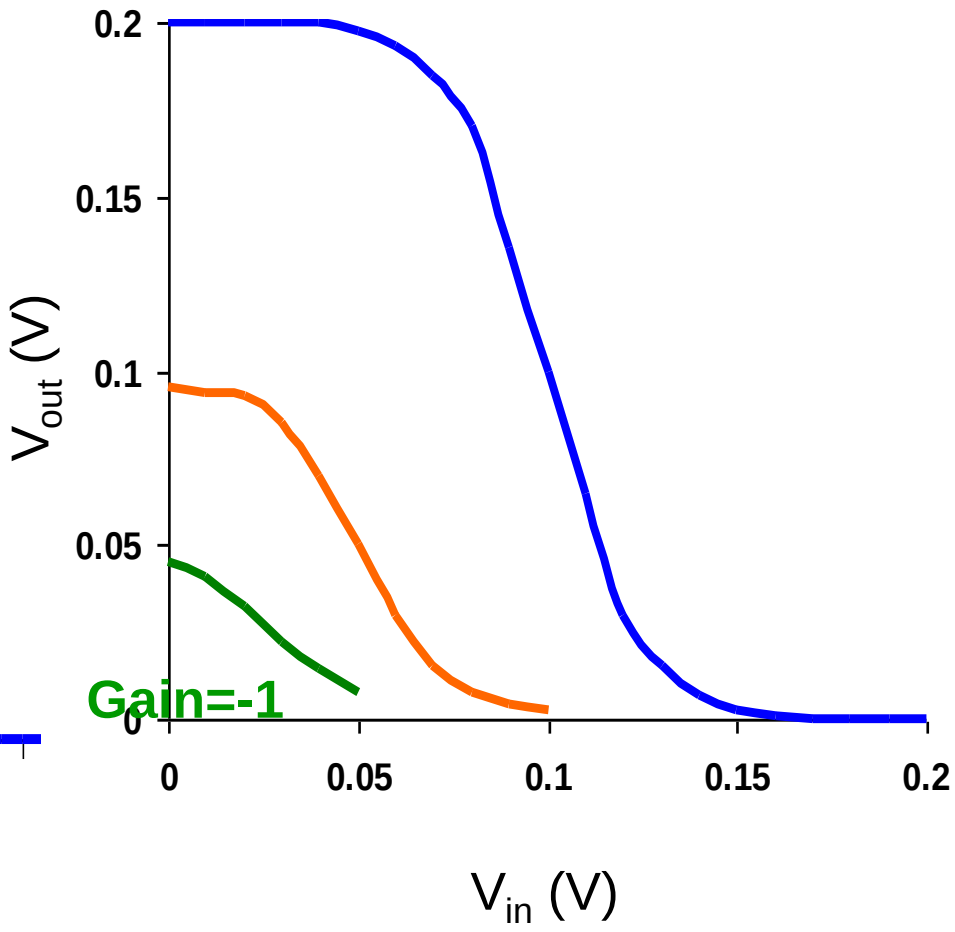


- Process variations (mostly) cause a shift in the switching threshold

Scaling the Supply Voltage



Device threshold voltages are kept (virtually) constant

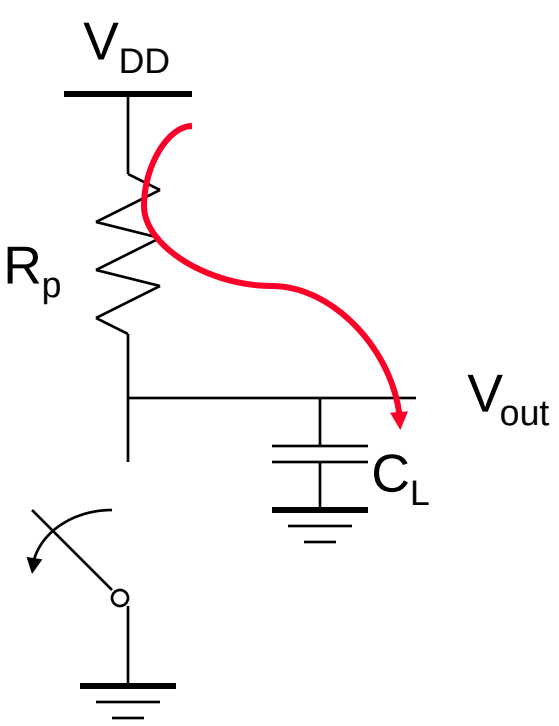


Device threshold voltages are kept (virtually) constant

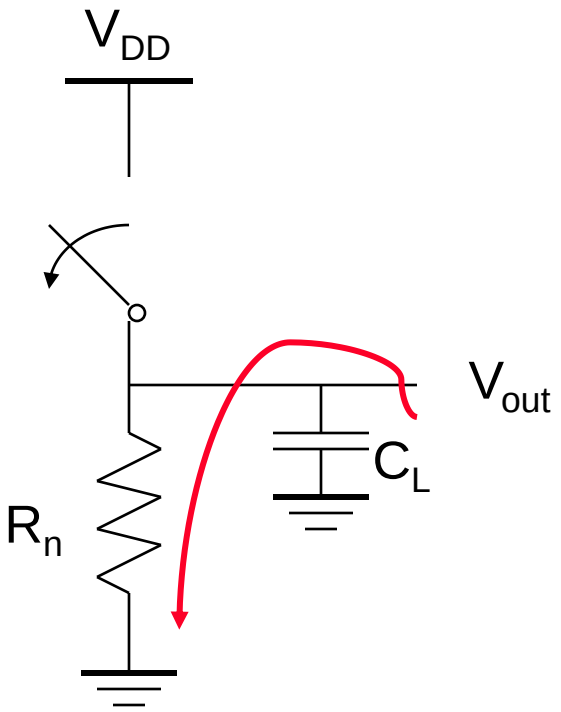
وارونگر ایستا

- مقدمه
- مشخصه انتقالی
- رفتار ایستا
- رفتار پویا
- تحلیل توان، انرژی و طراحی برای مصرف کم

تحلیل رفتار سوییچینگ



$V_{in} = 0$

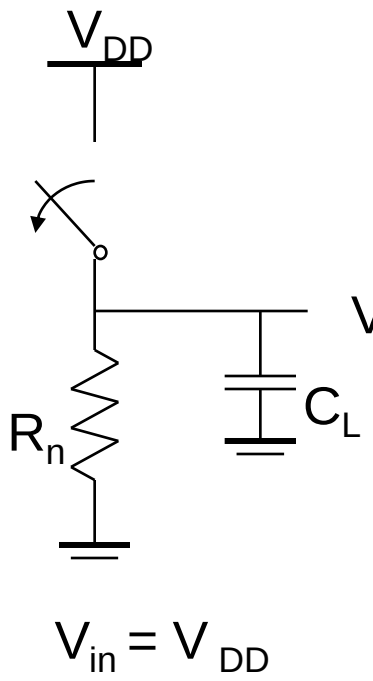


$V_{in} = V_{DD}$

- Gate response time is determined by the time to charge C_L through R_p (discharge C_L through R_n)

تحلیل تاخیر وارونگر

- Propagation delay is proportional to the time-constant of the network formed by the pull-down resistor and the load capacitance



$$t_{pHL} = f(R_n, C_L)$$

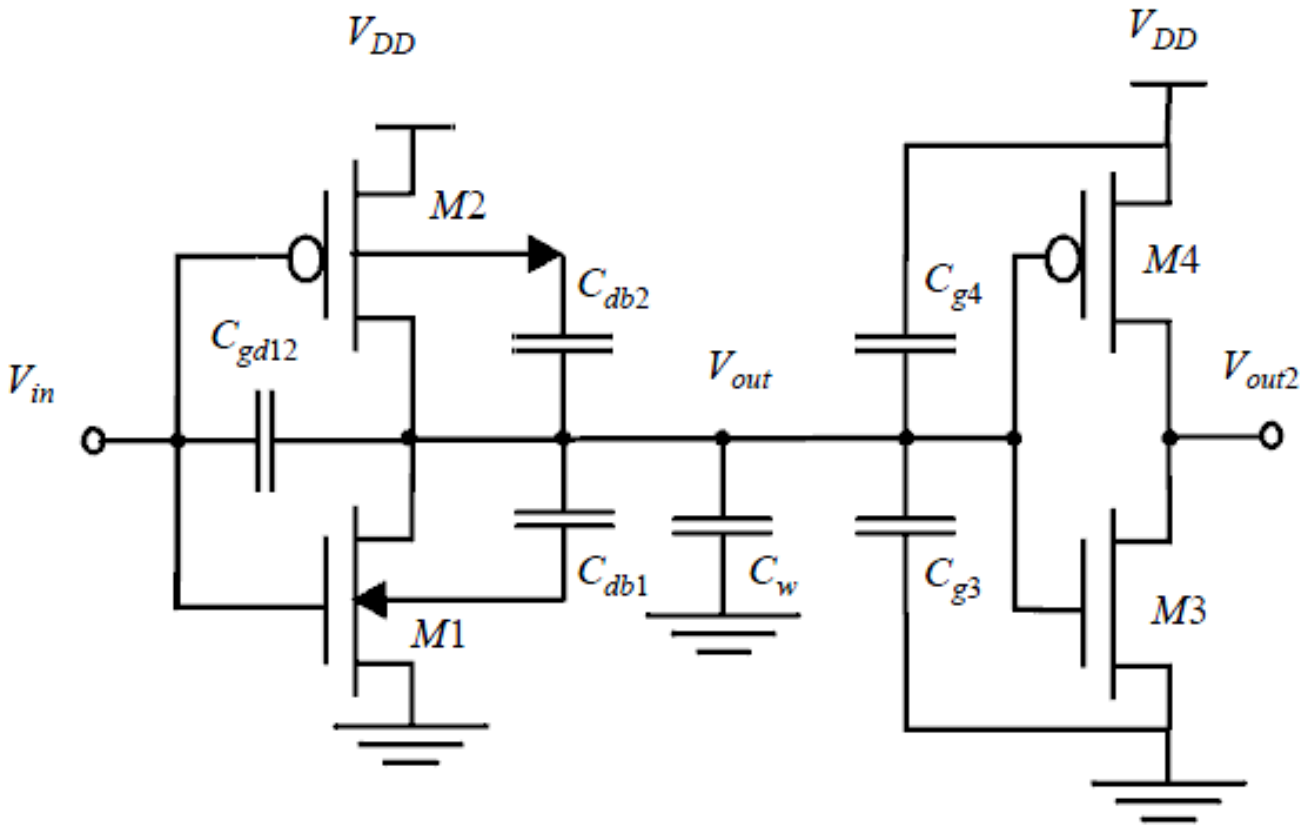
$$t_{pHL} = \ln(2)\, R_{eqn}\, C_L = 0.69\, R_{eqn}\, C_L$$

$$t_{pLH} = \ln(2)\, R_{eqp}\, C_L = 0.69\, R_{eqp}\, C_L$$

$$t_p = (t_{pHL} + t_{pLH})/2 = 0.69\, C_L(R_{eqn} + R_{eqp})/2$$

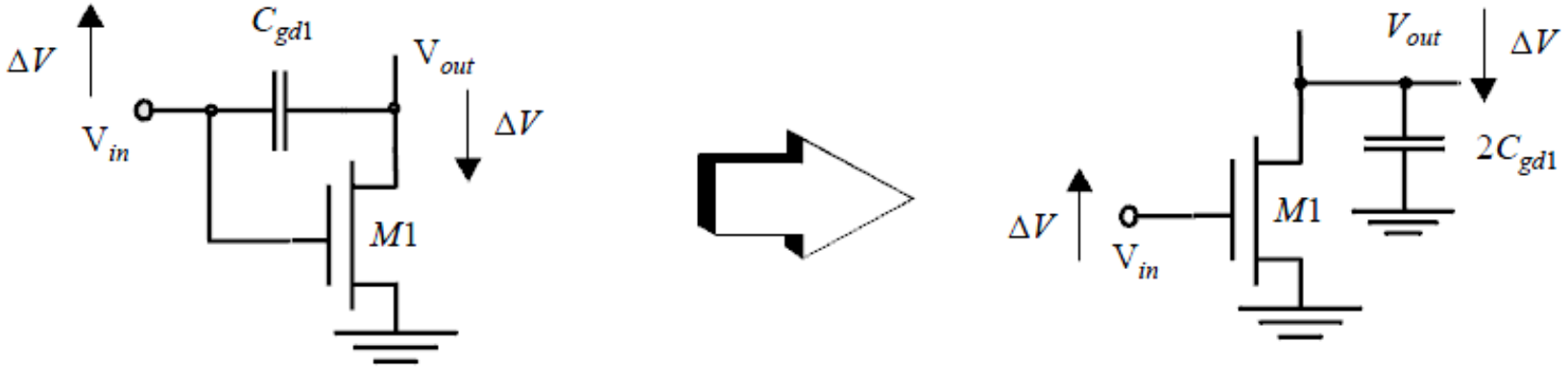
- To equalize rise and fall times make the on-resistance of the NMOS and PMOS approximately equal.

مولفه های خازن



خازن گیت درین

چون ترانزیستور M1 در ولتاژ های ورودی VDD/2 > 0 فقط در ناحیه قطع یا اشباع است فقط خازن Overlap داریم که آنهم با استفاده از اثر میلر می توان بین درین و زمین قرار داد.



$$C_{gd} = 2\,C_{GD0}W$$

خازن درین – بدنه

۲-برای دو ترانزیستور M1 و M2 خازن بین درین تا زمین وجود دارد که غیر خطی است و می توان یک خازم معادل با خطی سازی و تقریب در نظر گرفت. مقدار این خازن بستگی به سطح و محیط جانبی پیوند درین بدنه دارد.

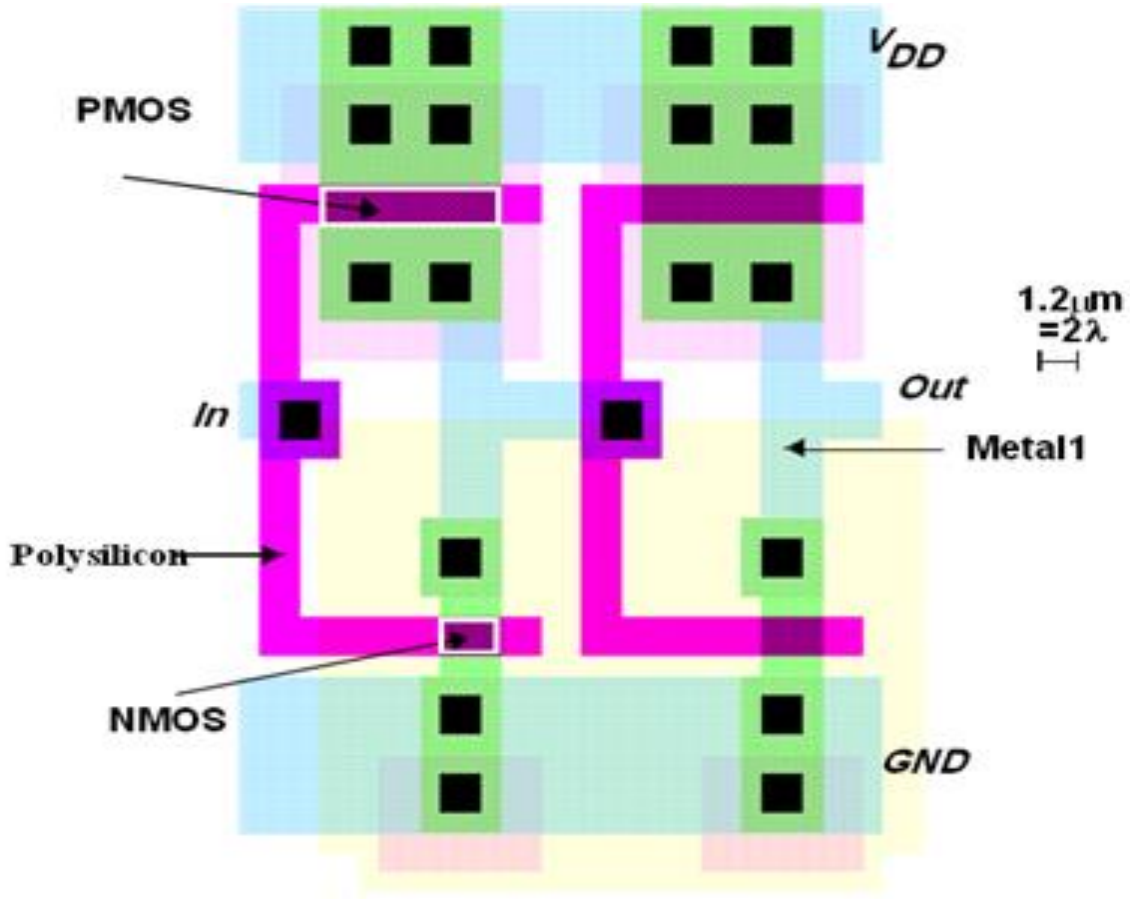
$$C_{eq} = K_{eq} C_{j0}$$

۳-خازن ورودی طبقه بعد که بار Fanout این طبقه محسوب می شود.

$$\begin{aligned} C_{fanout} &= C_{gate}(\text{NMOS}) + C_{gate}(\text{PMOS}) \\ &= (C_{GSOn} + C_{GDOn} + W_n L_n C_{ox}) + (C_{GSOp} + C_{GDOp} + W_p L_p C_{ox}) \end{aligned}$$

۴- خازن اتصالات میانی CW

مثال محاسبه خازن های دو وارونگر cascade



تکنولوژی 0.25 μm

λ=0.125 μm

برای ترانزیستور NMOS

$Ad=4x4\ \lambda^2+3x1\ \lambda^2$

$Pd=(5+4+4+1+1)\ \lambda$

برای ترانزیستور PMOS

$Ad=9x4\ \lambda^2+9x1\ \lambda^2$

$Pd=(5+9+5)\ \lambda$

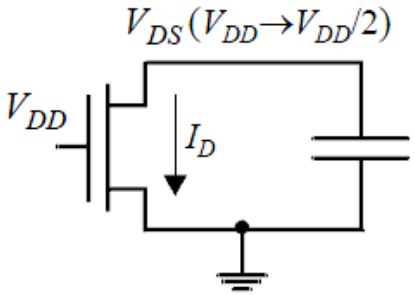
ادامه مثال

	W/L	AD (μm^2)	PD (μm)	AS (μm^2)	PS (μm)
NMOS	0.375/0.25	0.3 (19 λ^2)	1.875 (15λ)	0.3 (19 λ^2)	1.875 (15λ)
PMOS	1.125/0.25	0.7 (45 λ^2)	2.375 (19λ)	0.7 (45 λ^2)	2.375 (19λ)

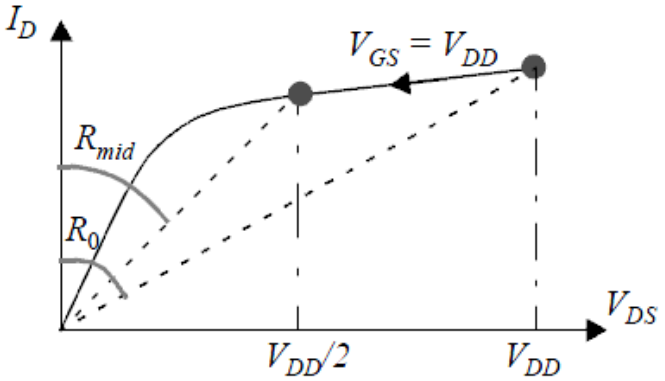
Capacitor	Expression	Value (fF) (H→L)	Value (fF) (L→H)
C_{gd1}	$2\ CGD0_n\ W_n$	0.23	0.23
C_{gd2}	$2\ CGD0_p\ W_p$	0.61	0.61
C_{db1}	$K_{eqn}\ AD_n\ CJ + K_{eqswn}\ PD_n\ CJSW$	0.66	0.90
C_{db2}	$K_{eqp}\ AD_p\ CJ + K_{eqswp}\ PD_p\ CJSW)$	1.5	1.15
C_{g3}	$(CGD0_n+CGSO_n)\ W_n + C_{ox}\ W_n\ L_n$	0.76	0.76
C_{g4}	$(CGD0_p+CGSO_p)\ W_p + C_{ox}\ W_p\ L_p$	2.28	2.28
C_w	From Extraction	0.12	0.12
C_L	Σ	6.1	6.0

مقدار مقاومت هنگام پرشدن (خالی شدن) خازن

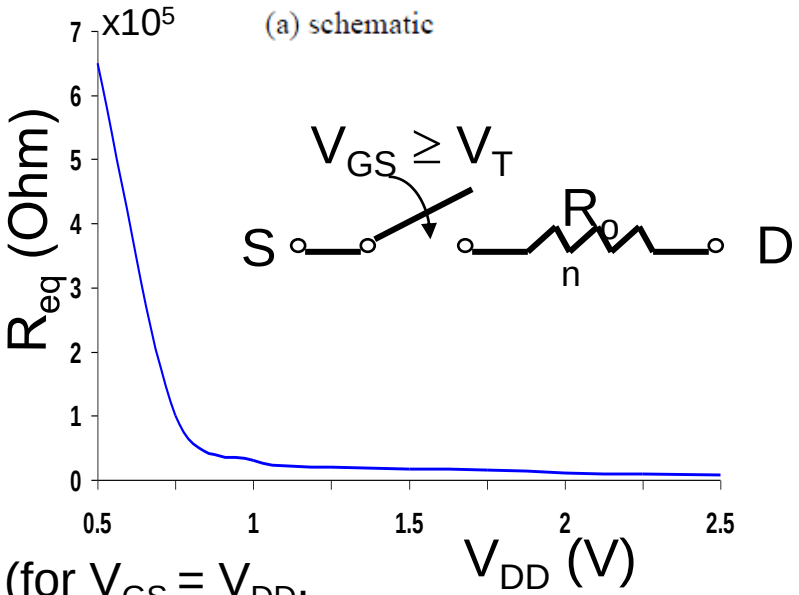
با توجه به اینکه تاخیر انتشار با رسیدن به ۵۰٪ مقدار نهایی تعریف می شود.



(a) schematic



(b) trajectory traversed on ID-VDS curve.



(for $V_{GS} = V_{DD}$, $V_{DS} = V_{DD} \rightarrow V_{DD}/2$)

از مدل ترانزستور به عنوان یک سوییچ با تقریب معادله زیر استفاده می کنیم.

$$R_{eqn} = 3/4\ V_{DD}/I_{DSATn}$$

محاسبه تاخیر انتشار برای وارونگر دو طبقه قبلی

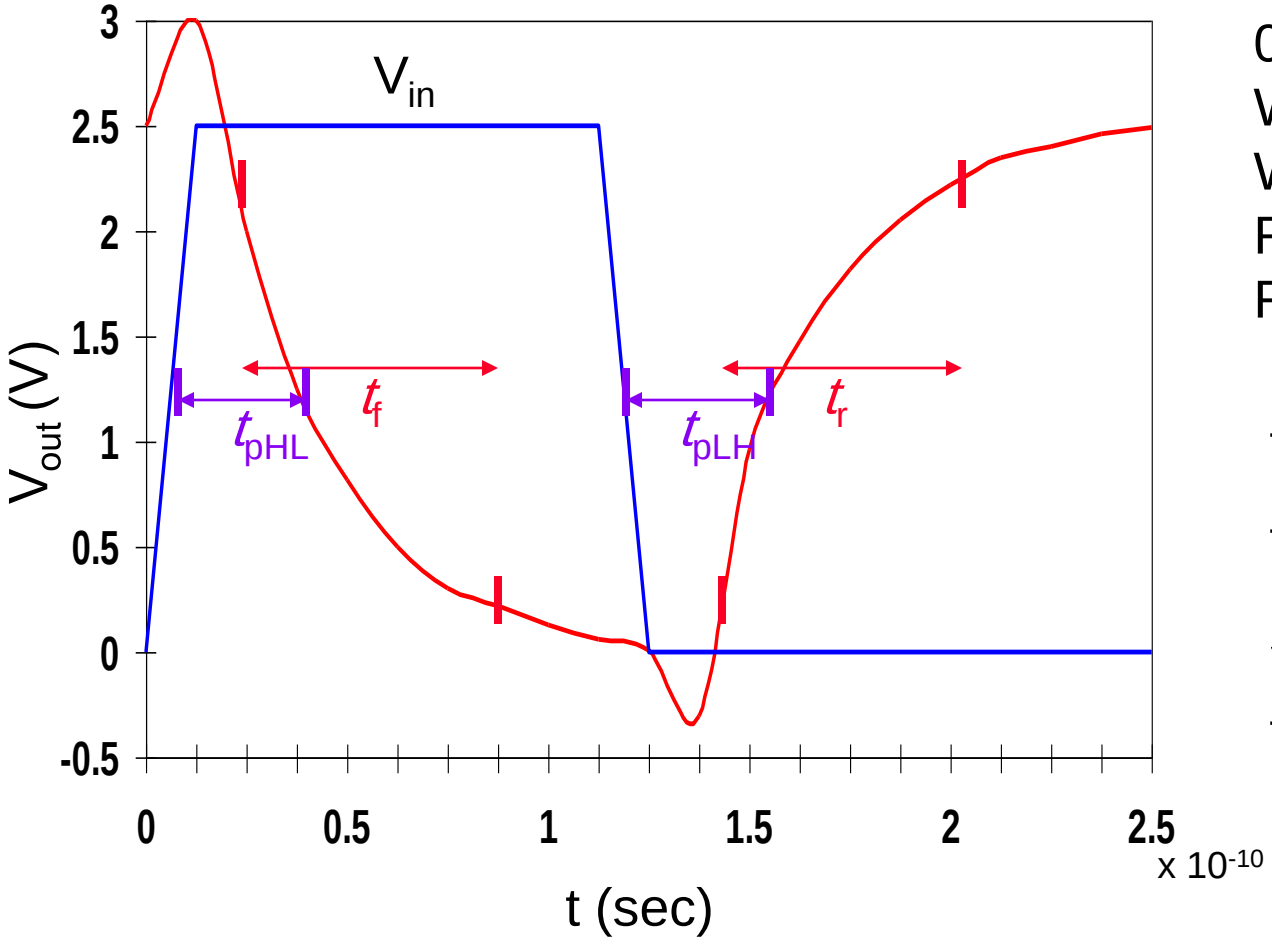
با در نظر گرفتن مقاومت های on ترانزیستور های nMOS و pMOSبا حداقل ابعاد معادل 13 KΩ و 31 KΩ (با توجه به اینکه عرض ترانزیستور های nMOS و pMOS فوق بترتیب ۱.۵ و ۴.۵ برابر حداقل ابعاد هستند. مقدار مقاومت های آنها با تقسیم مقاومت ترانزیستور مرجع به ضرایب فوق بدست می آید و لذا داریم .

$$t_{pHL} = 0.69 \times \left(\frac{13\text{k}\Omega}{1.5}\right) \times 6.1\text{fF} = 36\ \text{psec}$$

$$t_{pLH} = 0.69 \times \left(\frac{31\text{k}\Omega}{4.5}\right) \times 6.0\text{fF} = 29\ \text{psec}$$

$$t_p = \left(\frac{36+29}{2}\right) = 32.5\ \text{psec}$$

Inverter Transient Response



V_{DD}=2.5V
0.25μm
W/L_n = 1.5
W/L_p = 4.5
R_{eqn}= 13 kΩ (÷ 1.5)
R_{eqp}= 31 kΩ (÷ 4.5)

t_{pHL} = 36 psec

t_{pLH} = 29 psec

so

t_p = 32.5 psec

From simulation: t_{pHL} = 39.9 psec and t_{pLH} = 31.7 psec

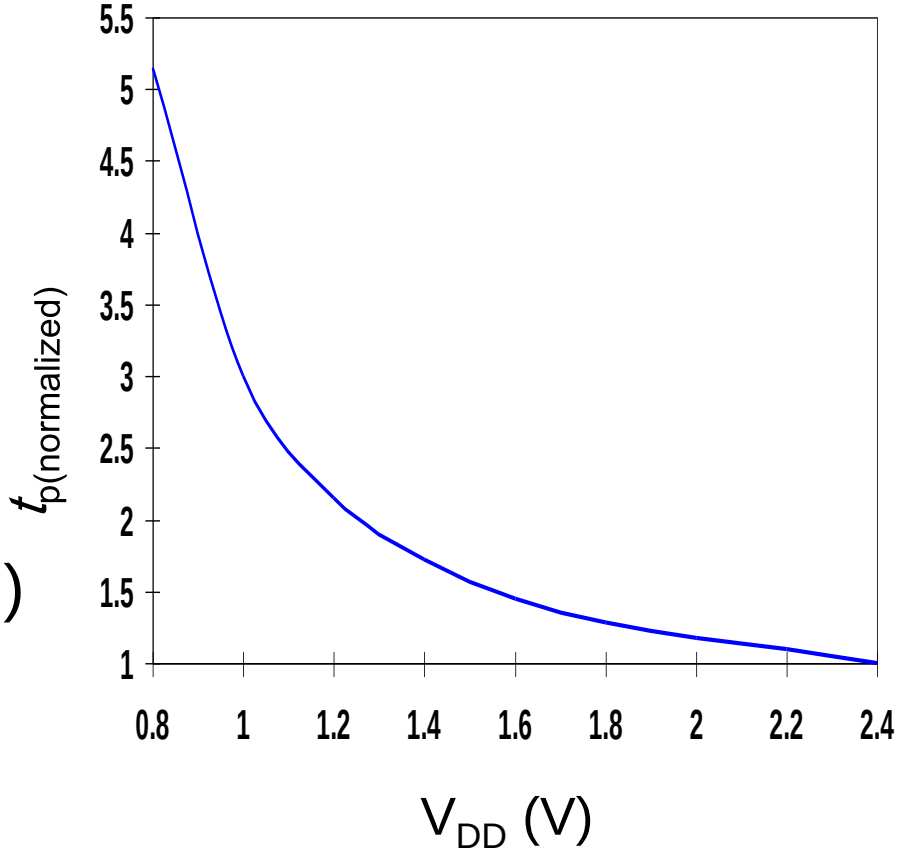
پارامتر های موثر روی تاخیر انتشار وارونگر

- با جایگزاری مقاومت معادل تقریبی

$$t_{pHL} = 0.69\, R_{eqn}\, C_L$$

$$= 0.69\,(3/4\,(C_L\, V_{DD})/I_{DSATn})$$

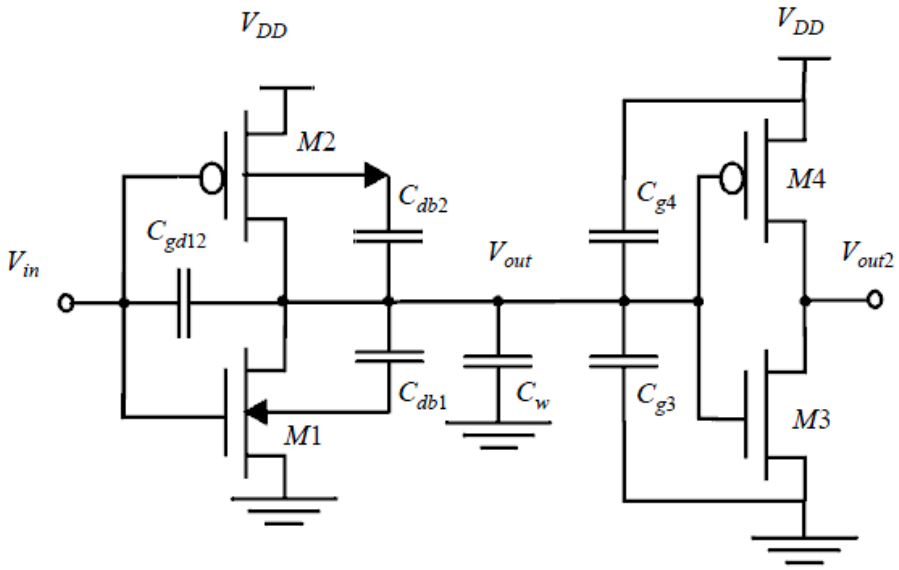
$$\approx 0.52\, C_L / (W/L_n\, k'_n\, V_{DSATn})$$



$$t_{pHL} = 0.69\frac{3}{4}\frac{C_L V_{DD}}{I_{DSATn}} = 0.52\frac{C_L V_{DD}}{(W/L)_n k'_n V_{DSATn}(V_{DD} - V_{Tn} - V_{DSATn}/2)}$$

پارامتر های موثر در سرعت

- Reduce C_L
  - internal diffusion capacitance of the gate itself
    - keep the drain diffusion as small as possible
  - interconnect capacitance
  - Fanout (Cg Next Level)



پارامتر های موثر در سرعت

- Increase W/L ratio of the transistor
  - the most powerful and effective performance optimization tool in the hands of the designer
  - watch out for self-loading! – when the intrinsic capacitance dominates the extrinsic load
- Increase V_{DD}
  - can trade-off energy for performance
  - increasing V_{DD} above a certain level yields only very minimal improvements
  - reliability concerns enforce a firm upper bound on V_{DD}

نسبت NMOS/PMOS

- So far have sized the PMOS and NMOS so that the R_{eq}'s match (ratio of 3 to 3.5)
  - symmetrical VTC
  - equal high-to-low and low-to-high propagation delays

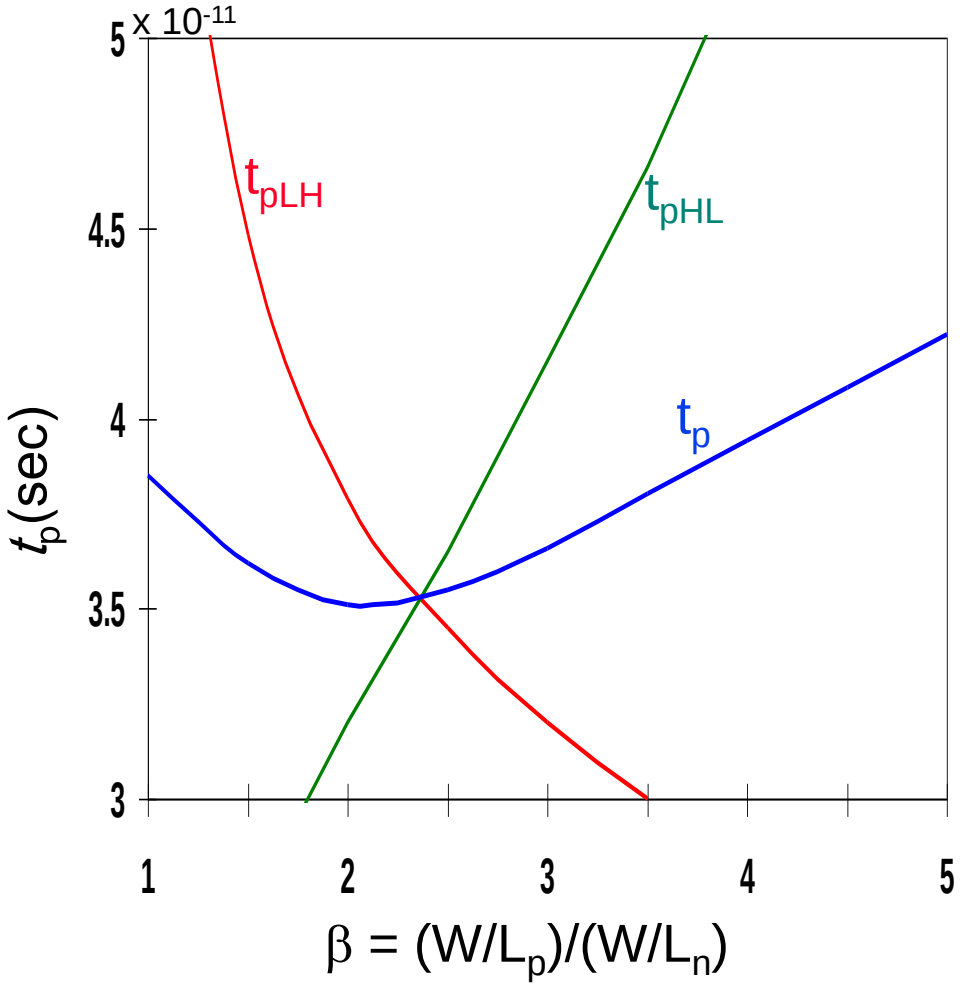
- If speed is the only concern, reduce the width of the PMOS device!
  - widening the PMOS degrades the t_{pHL} due to larger parasitic capacitance

$$\beta = (W/L_p)/(W/L_n)$$

$r = R_{eqp}/R_{eqn}$ (resistance ratio of identically-sized PMOS and NMOS)

$\beta_{opt} = \sqrt{r}$ when wiring capacitance is negligible

PMOS/NMOS Ratio Effects



β of 2.4 (= 31 kΩ/13 kΩ) gives symmetrical response

β of 1.6 to 1.9 gives optimal performance

Device Sizing for Performance

- Divide capacitive load, C_L, into
  - C_{int} : intrinsic - diffusion and Miller effect
  - C_{ext} : extrinsic - wiring and fanout

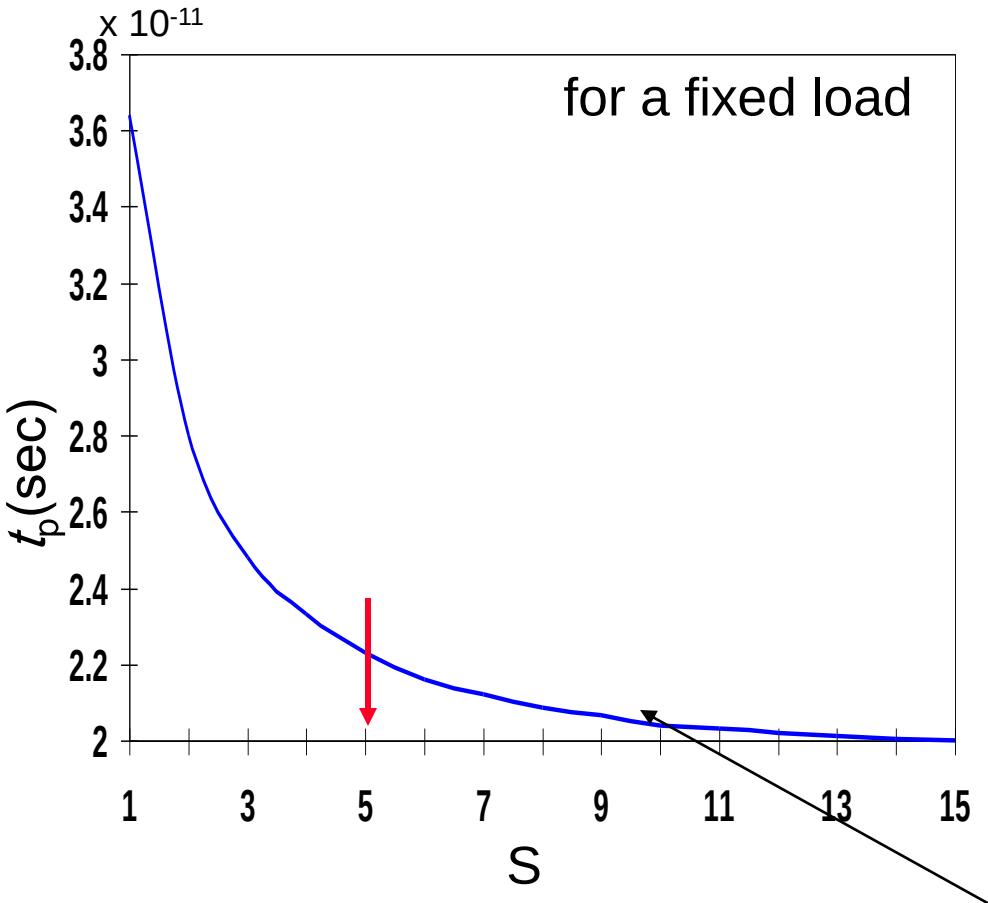
$$t_p = 0.69\, R_{eq}\, C_{int}\, (1 + C_{ext}/C_{int}) = t_{p0}\, (1 + C_{ext}/C_{int})$$

  - where $t_{p0} = 0.69\, R_{eq}\, C_{int}$ is the intrinsic (unloaded) delay of the gate
- Widening both PMOS and NMOS by a factor S reduces R_{eq} by an identical factor ($R_{eq} = R_{ref}/S$), but raises the intrinsic capacitance by the same factor ($C_{int} = SC_{iref}$)

$$t_p = 0.69\, R_{ref}\, C_{iref}\, (1 + C_{ext}/(SC_{iref})) = t_{p0}(1 + C_{ext}/(SC_{iref}))$$

  - t_{p0} is independent of the sizing of the gate; *with no load the drive of the gate is totally offset by the increased capacitance*
  - any S sufficiently larger than (C_{ext}/C_{int}) yields the best performance gains with least area impact

Sizing Impacts on Delay



The majority of the improvement is already obtained for S = 5. Sizing factors larger than 10 barely yield any extra gain (and cost significantly more area).

self-loading effect (intrinsic capacitance dominates)

Impact of Fanout on Delay

- Extrinsic capacitance, C_{ext}, is a function of the fanout of the gate - the larger the fanout, the larger the external load.

- First determine the input loading effect of the inverter. Both C_g and C_{int} are proportional to the gate sizing, so $C_{int} = \gamma C_g$ is independent of gate sizing and

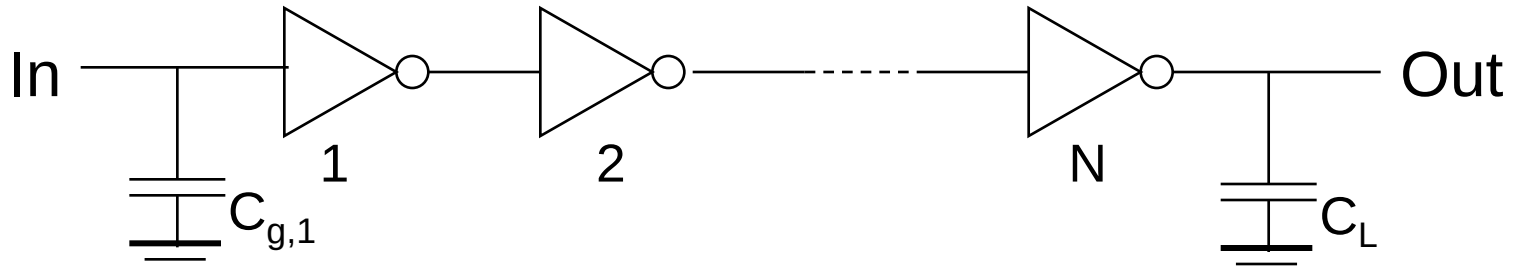
$$t_p = t_{p0}\,(1 + C_{ext}/\,\gamma C_g) = t_{p0}\,(1 + f/\gamma)$$

i.e., the delay of an inverter is a function of the ratio between its external load capacitance and its input gate capacitance: the effective fan-out f

$$f = C_{ext}/C_g$$

زنجیره وارونگر ها

- حداقل کردن تاخیر با انتخاب مناسب تعداد بهینه و ابعاد بهینه وارونگر ها



the delay of the j-th inverter stage is

$$t_{p,j} = t_{p0}\,(1 + C_{g,j+1}/(\gamma C_{g,j})) = t_{p0}(1 + f_j/\gamma)$$

and $$t_p = t_{p1} + t_{p2} + \ldots + t_{pN}$$

$$t_p = \Sigma t_{p,j} = t_{p0}\,\Sigma\,(1 + C_{g,j+1}/(\gamma C_{g,j}))$$

Sizing the Inverters in the Chain

- می توان نشان داد حداقل میزان تاخیر وقتی اتفاق می افتد که $C_{g,j+1}/C_{g,j} = C_{g,j}/C_{g,j-1}$
- به عبارت دیگر افزایش ابعاد طبقات نسبت به یکدیگر با نسبت ثابتی اتفاق بیافتد که در آن ضریب تناسب همان f است از رابطه زیر بدست می آید.

$$f = \sqrt[N]{C_L / C_{g,1}} = \sqrt[N]{F}$$

که در آن F نسبت خازن بار خروجی به خازن ورودی طبقه اول است.

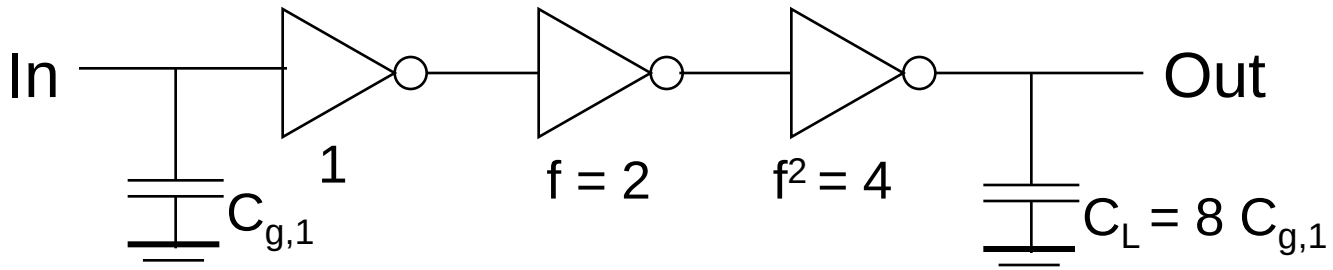
$(F = C_L/C_{g,1})$

بنابراین حداقل تاخیر از رابطه زیر بدست می آید.

$$t_p = N t_{p0}(1 + \sqrt[N]{F})$$

N تعداد طبقات است بنابراین می توان N مناسب را انتخاب نمود تا تاخیر بهینه شود.

Example of Inverter Chain Sizing



$C_L/C_{g,1}$ has to be evenly distributed over N = 3 inverters ❑

$$C_L/C_{g,1} = 8/1$$

$$f = \sqrt[3]{8} = 2$$

انتخاب تعداد طبقات بهینه N

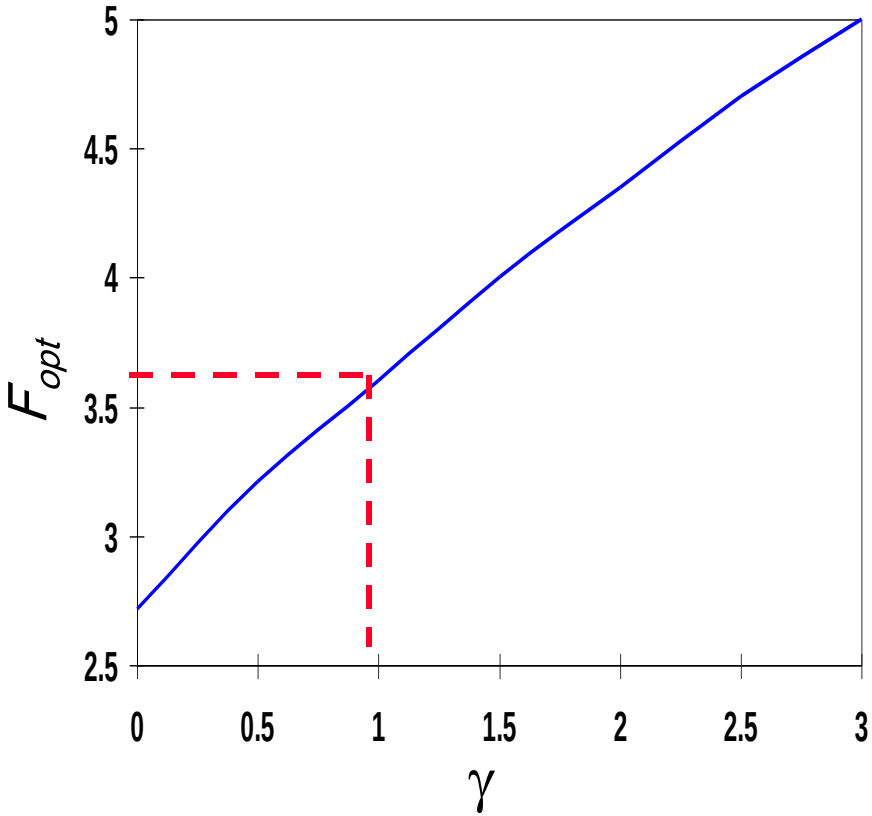
- اگر تعداد طبقات را زیاد کنیم بدلیل تاخیر ذاتی که در N ضرب می شود تاخیر زیاد می شود
- اگر تعداد طبقات را کم بگیریم تاخیر ناشی از Fanout غالب می شود.

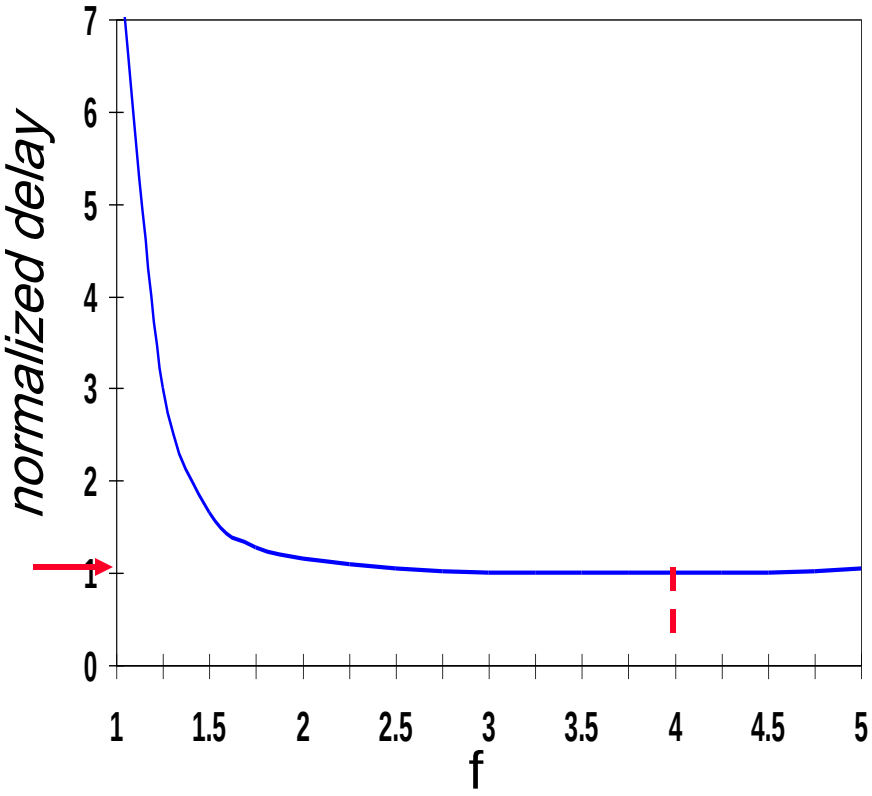
- مقدار بهینه از حل معادله زیر بدست می آید.

$$\gamma + \sqrt{F} - (\sqrt{F} \ln F)/N = 0$$

- For $\gamma = 0$ $N = \ln(F)$ and the *effective-fan out becomes* $f = e = 2.71828$
- For $\gamma = 1$ (the typical case) the optimum effective fan-out (tapering factor) turns out to be close to 3.6

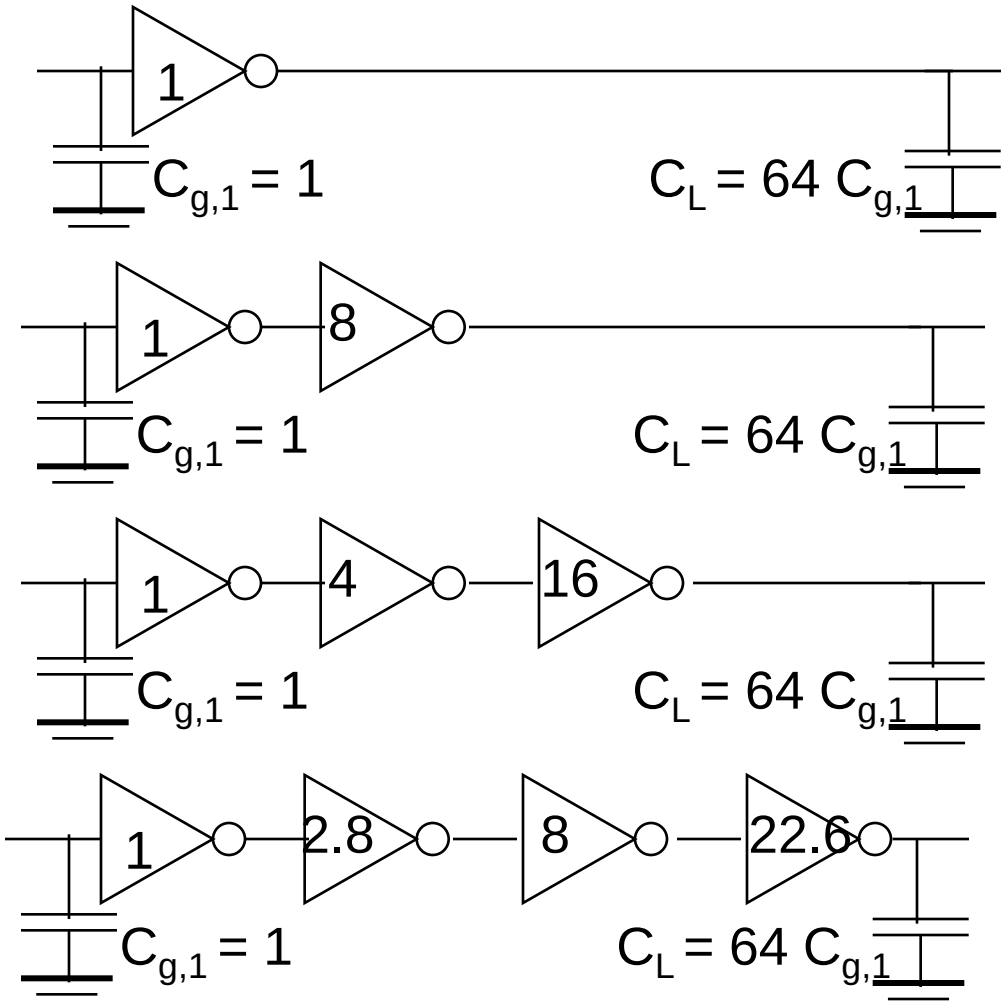
Optimum Effective Fan-Out





- Choosing f larger than optimum has little effect on delay and reduces the number of stages (and area).
  - Common practice to use f = 4 (for $\gamma = 1$)
  - But too many stages has a substantial negative impact on delay

Example of Inverter (Buffer) Staging



N	f	t_p
1	64	65
2	8	18
3	4	15
4	2.8	15.3

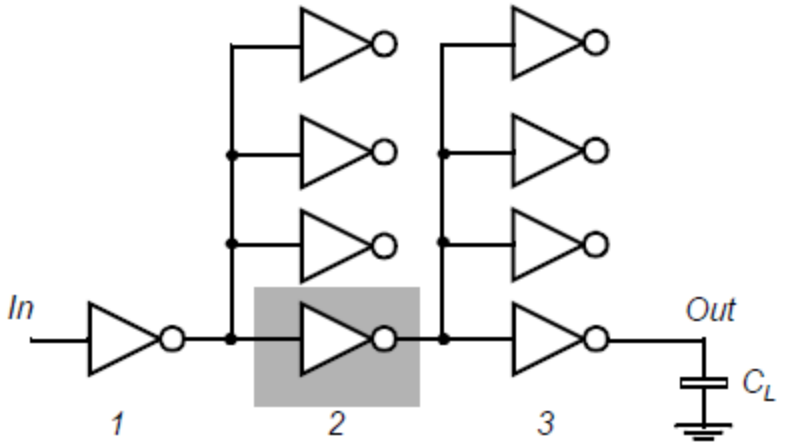
Impact of Buffer Staging for Large C_L

F ($\gamma = 1$)	Unbuffered	Two Stage Chain	Opt. Inverter Chain
10	11	8.3	8.3
100	101	22	16.5
1,000	1001	65	24.8
10,000	10,001	202	33.1

- Impressive speed-ups with optimized cascaded inverter chain for very large capacitive loads.

مثال : تاخیر بهینه

$C_L = 64\, C_{g,1}$.



Hints: Determine first the ratio's between the devices that minimize the delay. You should find that the following must hold,

$$\frac{4C_{g,2}}{C_{g,1}} = \frac{4C_{g,3}}{C_{g,2}} = \frac{C_L}{C_{g,3}}$$

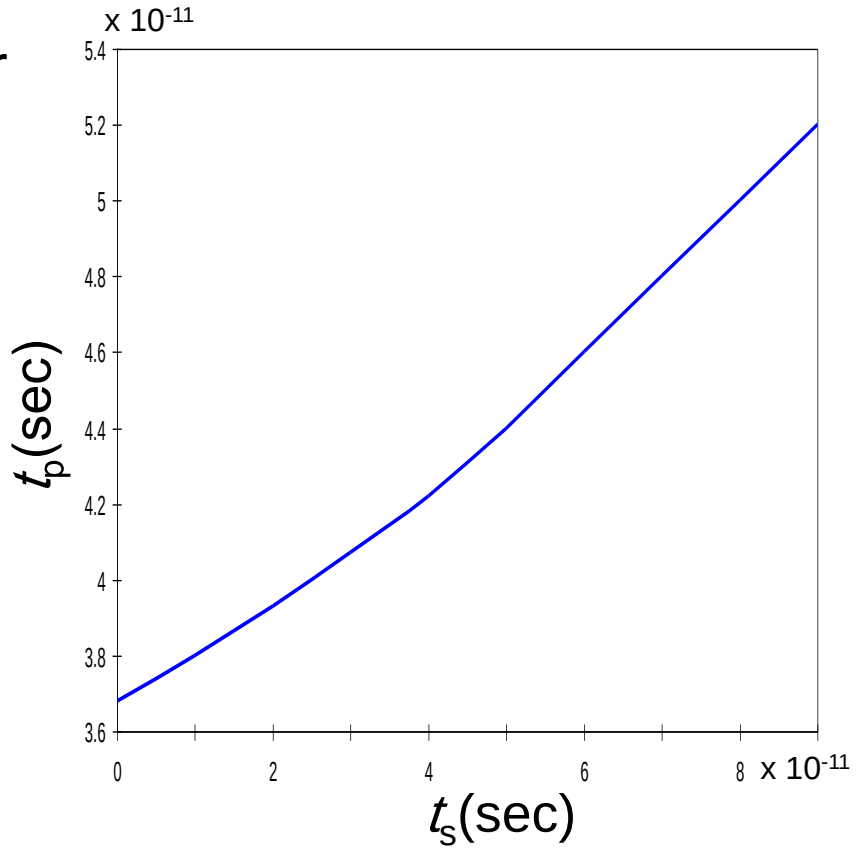
Finding the actual gate sizes ($C_{g,3} = 2.52C_{g,2} = 6.35C_{g,1}$) is a relatively straightforward task. Straightforward sizing of the inverter chain, without taking the fanout into account, would have led to a sizing factor of 4 instead of 2.52.

Input Signal Rise/Fall Time

- In reality, the input signal changes gradually (and both PMOS and NMOS conduct for a brief time). This affects the current available for charging/discharging C_L and impacts propagation delay.

- t_p increases linearly with increasing input slope, t_s, once $t_s > t_p$

- t_s is due to the limited driving capability of the preceding gate



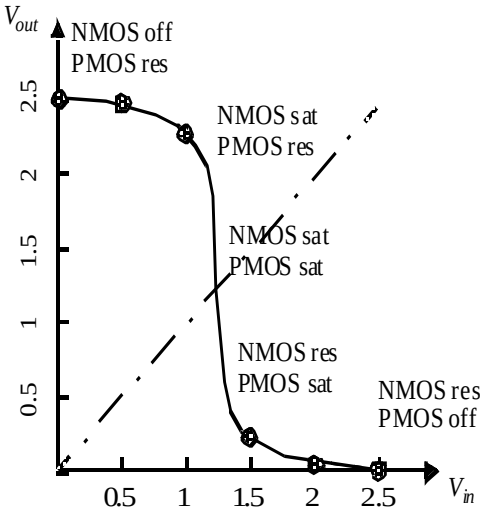
for a minimum-size inverter
with a fan-out of a single gate

وارونگر ایستا

- مقدمه
- مشخصه انتقالی
- رفتار ایستا
- رفتار پویا
- تحلیل توان، انرژی و طراحی برای مصرف کم

مولفه های تلفات

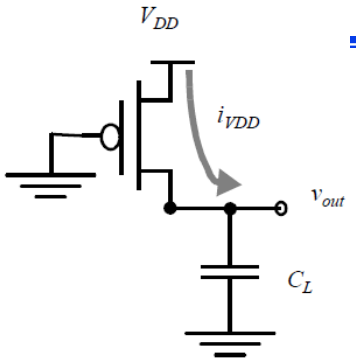
- **تلفات دینامیکی : در اثر سوییچینگ و مصرف انرژی برای شارژ و دشارژ خازن بار**



- **تلفات ناشی از مسیر جریان مستقیم**

- **تلفات استاتیک (جریان نشتی و زیر آستانه)**

تلفات دینامیکی



ساختار وارونگر CMOS استاتیک بگونه ای است که تلفات توان استاتیک در آن ناچیز است و عمده تلفات در حالت تغییر وضعیت از 0 به 1 یا بر عکس اتفاق می افتد. هر بار که خازن بار از مسیر pMOS شارژ می شود ولتاژ آن از 0 به VDD می رسد.

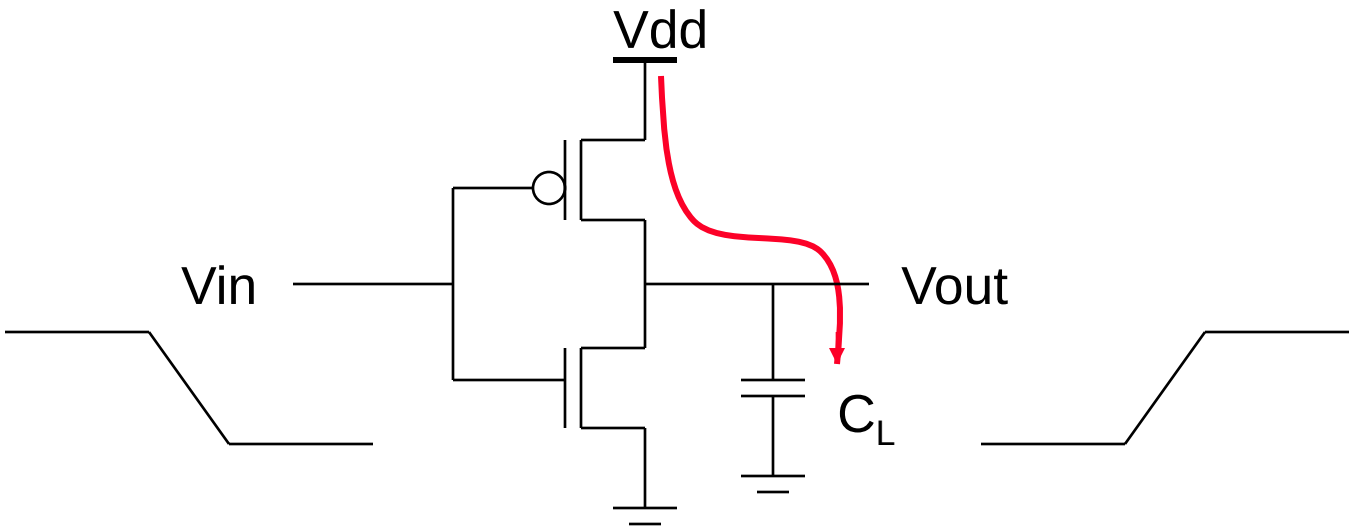
انرژی صرف شده توسط منبع VDD و انرژی ذخیره شده در خازن را می توان با روابط زیر بترتیب نشان داد.

$$E_{VDD} = \int_0^\infty i_{VDD}(t) V_{DD} dt = V_{DD} \int_0^\infty C_L \frac{dv_{out}}{dt} dt = C_L V_{DD} \int_0^{V_{DD}} dv_{out} = C_L V_{DD}^2$$

$$E_C = \int_0^\infty i_{VDD}(t) v_{out} dt = \int_0^\infty C_L \frac{dv_{out}}{dt} v_{out} dt = C_L \int_0^{V_{DD}} v_{out} dv_{out} = \frac{C_L V_{DD}^2}{2}$$

فقط نیمی از انرژی گرفته شده از منبع در خازن بار ذخیره می شود و نیم دیگر آن در ترانزیستور pMOS تلف می شود. میزان تلفات بستکی به ابعاد (مقاومت) ترانزیستور pMOS ندارد. در سیکل دشارژ انرژی ذخیره شده در خازن در ترانزیستور nMOS تلف می شود. بنابر این انرژی تلف شده در هر سیکل برابر $C_L VDD^2$ می شود.

Dynamic Power Consumption



$$\text{Energy/transition} = C_L * V_{DD}^2 * P_{0\to1}$$

$f_{0\to1}$

$$P_{dyn} = \text{Energy/transition} * f = C_L * V_{DD}^2 * P_{0\to1} * f$$

$$P_{dyn} = C_{EFF} * V_{DD}^2 * f \quad \text{where } C_{EFF} = P_{0\to1}\, C_L$$

Not a function of transistor sizes!

Data dependent - a function of switching activity!

طراحی برای مصرف کم

با توجه به معادله توان دینامیکی ۳ عامل در کاهش توان موثر هستند

کاهش ولتاژ تغذیه : چون توان مصرفی با توان دوم ولتاژ متناسب است این عامل مهمی است با کاهش ولتاژ تاخیر گیت زیاد می شود. بنابراین یک trade off بین توان و تاخیر وجود دارد

کاهش خازن بار: چون کاهش خازن موثر هم تلفات و هم تاخیر را کاهش می دهد کاهش آن مطلوب است. اما در عمل چون بخش عمده ای از خازن بار خازنهای داخلی ترانزیستور مثل خازن گیت و نفوذ هستند کاهش خازن ها به معنی کوچک سازی ترانزیستور است که به نوبه خود چون مقاومت معادل آنها را افزایش می دهد منجر به افزایش تاخیر می شود.

کاهش احتمال سوییچینگ این موضوع را با معماری مناسب مدار تا حدی می توان انجام داد و در فصل های آتی به آن اشاره می شود

کاهش فرکانس: معمولا بر خلاف میل طراح است. (گیت کردن کلاک سیگنال های غیر فعال)
در عمل ترکیبی از کاهش ولتاژ تغذیه و انتخاب مناسب ابعاد گیت جهت بهینه سازی توان و همزمان جلوگیری از افزایش تاخیر بیش از حد مطلوب مورد استفاده قرار می گیرد. در شبکه های بزرگ توجه به ابعاد موثر و کاهش ولتاژ تغذیه می تواند مصرف توان را تا ۱۰ برابر کاهش دهد.

Lowering Dynamic Power

Capacitance:
Function of fan-out, wire length, transistor sizes

Supply Voltage:
Has been dropping with successive generations

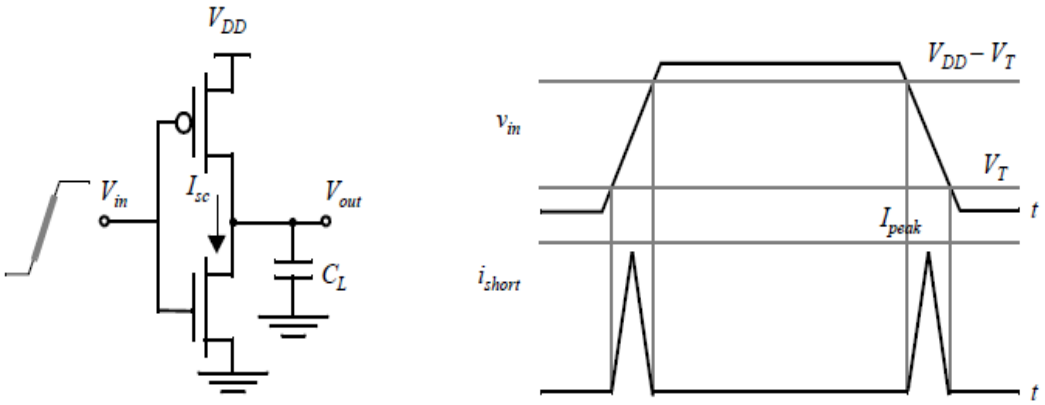
$$P_{dyn} = C_L V_{DD}^2 P_{0\rightarrow1} f$$

Activity factor:
How often, on average, do wires switch?

Clock frequency:
Increasing…

تلفات ناشی از مسیر جریان مستقیم

در حالت گذرا چون شیب سیگنال ورودی محدود است برای مدت کمی یک مسیر جریان مستقیم بین VDD و زمین ایجاد می شود.



$$E_{dp} = V_{DD}\frac{I_{peak}t_{sc}}{2} + V_{DD}\frac{I_{peak}t_{sc}}{2} = t_{sc}V_{DD}I_{peak} \qquad t_{sc} = \frac{V_{DD}-2V_T}{V_{DD}}t_s \approx \frac{V_{DD}-2V_T}{V_{DD}} \times \frac{t_{r(f)}}{0.8}$$

$$E_{sc} = t_{sc}\, V_{DD}\, I_{peak}\, P_{0\rightarrow1}$$

$$P_{sc} = t_{sc}\, V_{DD}\, I_{peak}\, f_{0\rightarrow1}$$

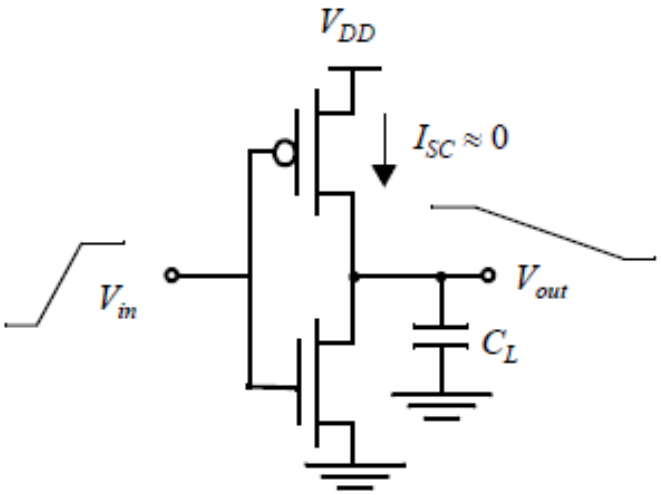
جریان پیک اتصال کوتاه

I_{peak} determined by

- the saturation current of the P and N transistors
  - which depend on their sizes, process technology, temperature, etc.
- strong function of the ratio between input and output slopes
  - a function of C_L

اثر خازن بار روی تلفات مسیر جریان مستقیم

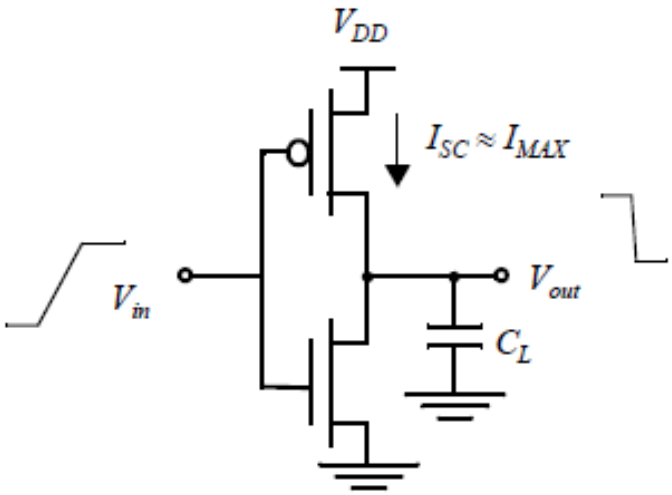
در حالتیکه تاخیر گیت زیاد باشد، وقتی سیگنال ورودی تغییر می کند ولتاژ خروجی هنوز بالاست، لذا اختلاف پتانسیل سورس و درین pMOS ناچیز است و جریان اتصال کوتاه ناچیز خواهد بود. در عوض اگر خازن بار کوچک باشد و تاخیر گیت کم باشد. ولتاژ درین سورس سریع تغییر می کند و لذا جریان اتصال کوتاه زیاد خواهد بود



(a) Large capacitive load

Large capacitive load

Output fall time significantly larger than input rise time.

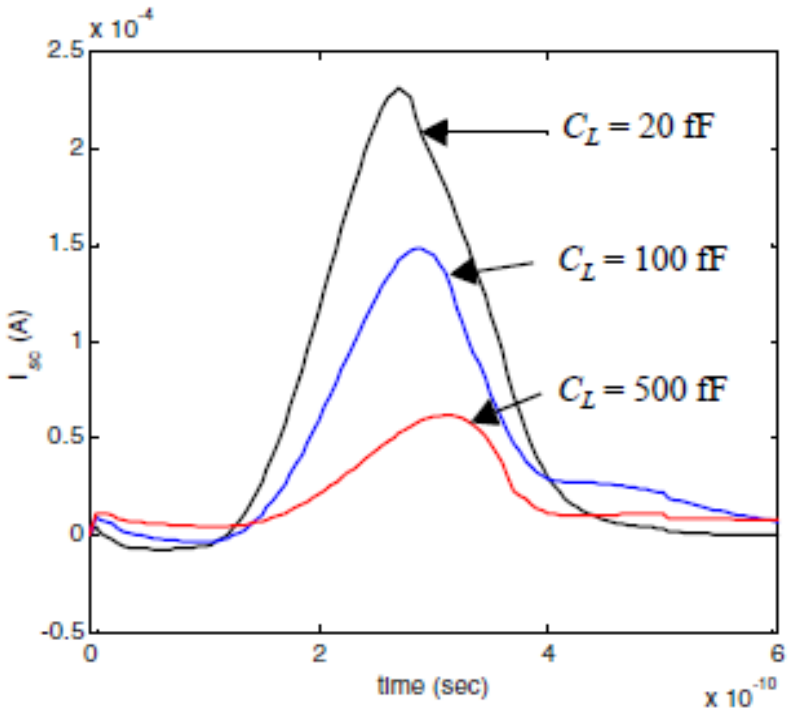


(b) Small capacitive load

Small capacitive load

Output fall time substantially smaller than the input rise time.

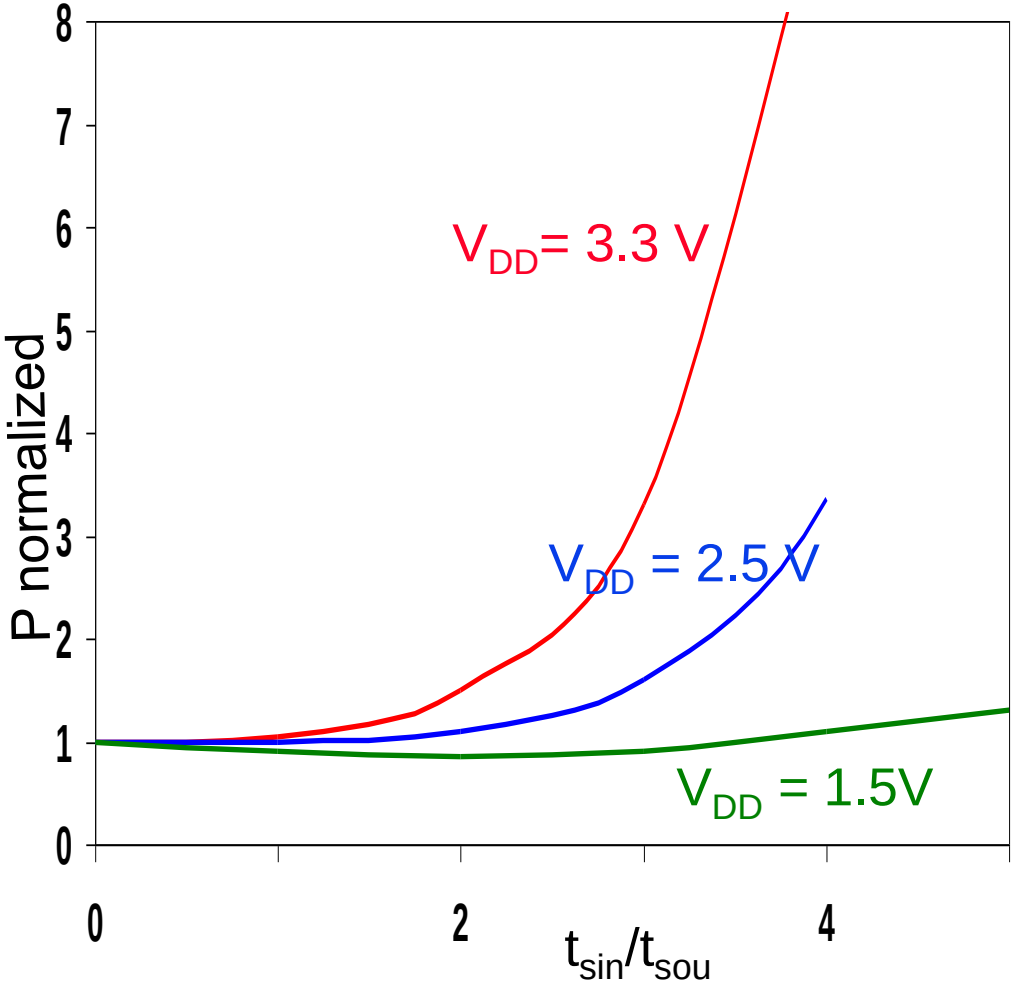
جریان اتصال کوتاه بر حسب خازن بار



When load capacitance is small, I_{peak} is large.

Short circuit dissipation is minimized by matching the rise/fall times of the input and output signals - slope engineering.

P_{sc} as a Function of Rise/Fall Times



When load capacitance is small ($t_{sin}/t_{sout} > 2$ for $V_{DD} > 2V$) the power is dominated by P_{sc}

If $V_{DD} < V_{Tn} + |V_{Tp}|$ then P_{sc} is eliminated since both devices are never on at the same time.

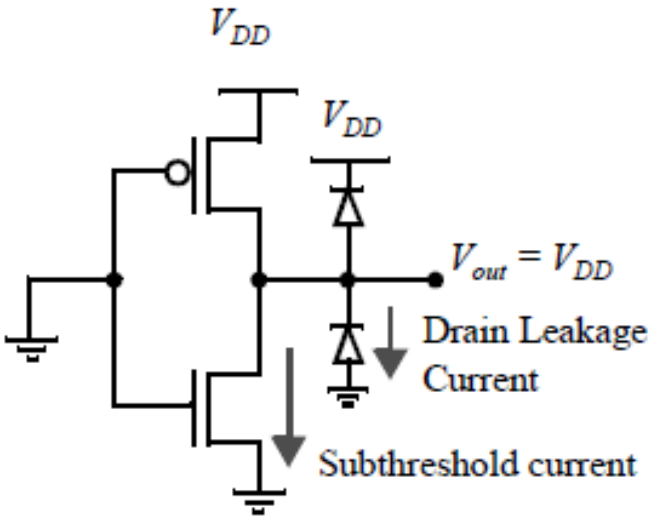
W/L_p = 1.125 μm/0.25 μm
W/L_n = 0.375 μm/0.25 μm

C_L = 30 fF

تلفات استاتیک

در حالت استاتیک به دلیل وجود جریان نشتی پیوند های بایاس معکوس درین و سورس با بدنه تلفات ایستایی داریم. در عمل این تلفات ناچیز است و در حد $10\text{-}100\ pA/\mu m^2$ است. اما این جریان نشتی با افزایش دما بصورت نمایی زیاد می شود و به عنوان نمونه در دمای $85^\circ C$ حدود ۶۰ برابر دمای اطاق می شود.

همینطور تلفات ناشی از جریان زیر آستانه برای ترانزیستور ها در حالت خاموش وجود دارد.

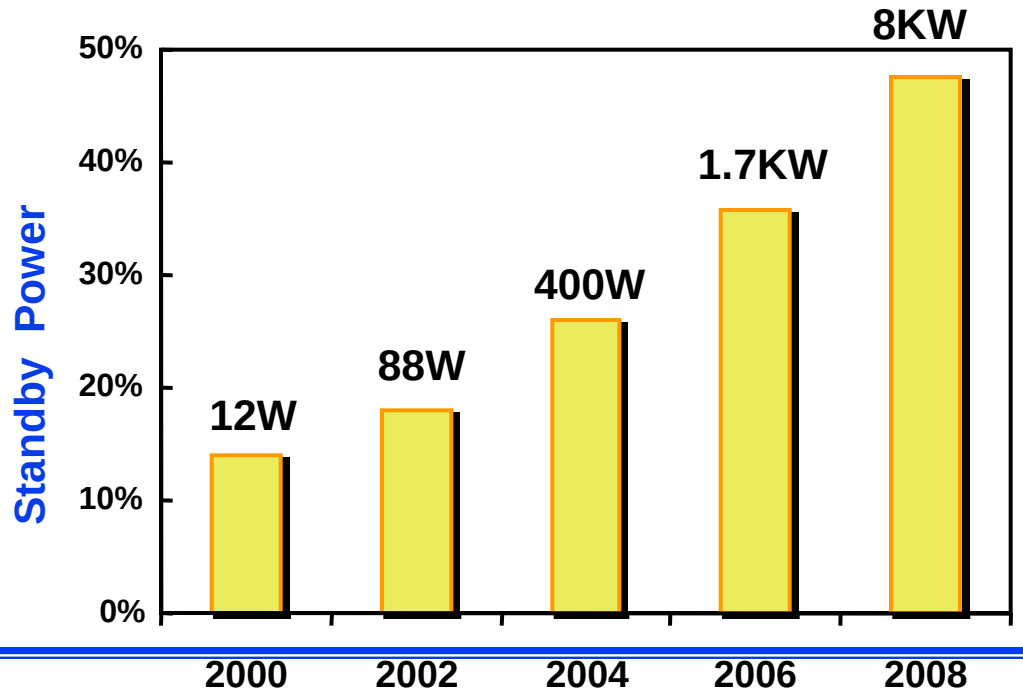


$$P_{stat} = I_{stat} V_{DD}$$

Why worry about power? -- Standby Power

Year	2002	2005	2008	2011	2014
Power supply V_{dd} (V)	1.5	1.2	0.9	0.7	0.6
Threshold V_T (V)	0.4	0.4	0.35	0.3	0.25

- Drain leakage will increase as V_T decreases to maintain noise margins and meet frequency demands, leading to excessive battery draining standby power consumption.

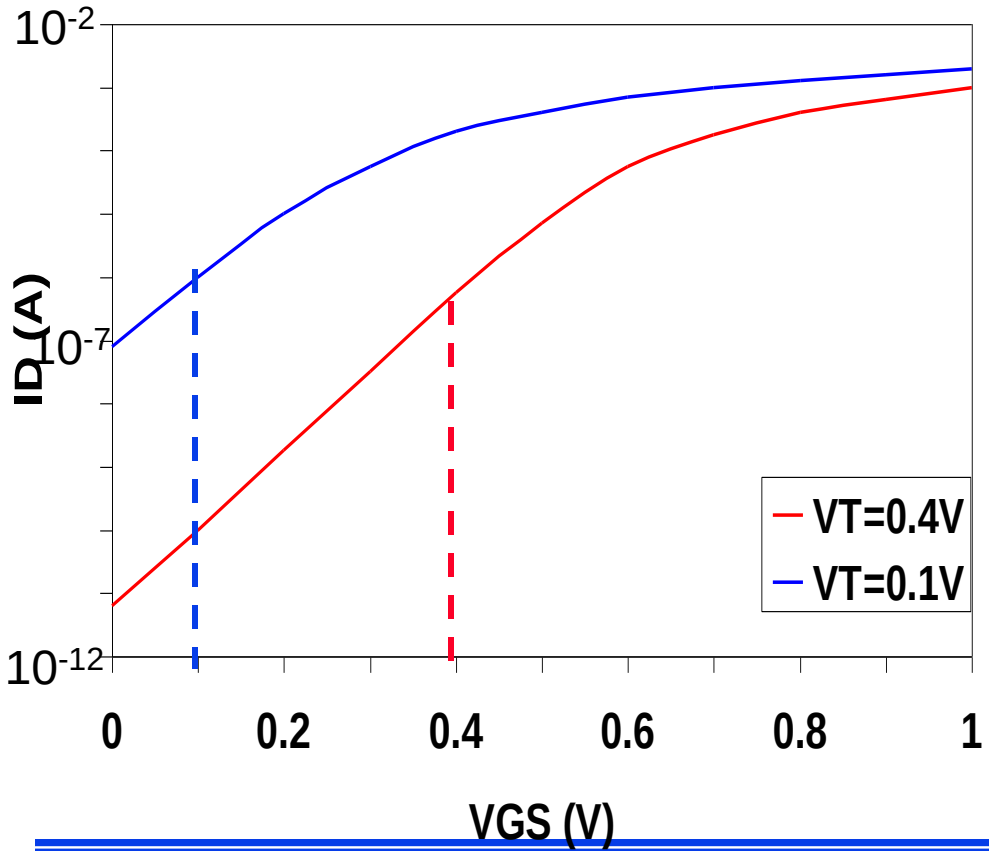


…and phones leaky!

Source: Borkar, De Intel®

Leakage as a Function of V_T

- Continued scaling of supply voltage and the subsequent scaling of threshold voltage will make subthreshold conduction a dominate component of power dissipation.

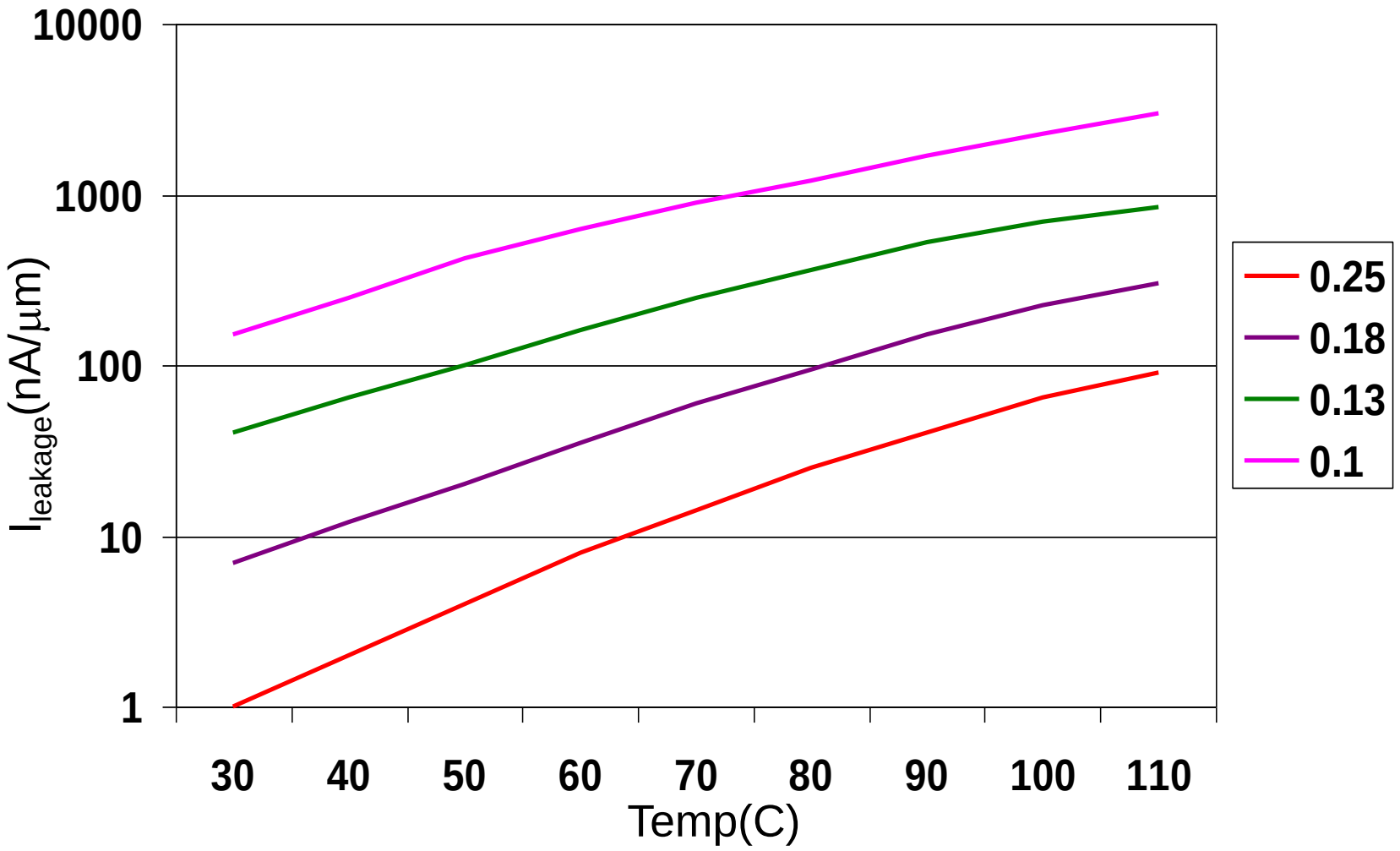


- An 90mV/decade V_T roll-off - so each 255mV increase in V_T gives 3 orders of magnitude reduction in leakage (but adversely affects performance)

TSMC Processes Leakage and V_T

	CL018 G	CL018 LP	CL018 ULP	CL018 HS	CL015 HS	CL013 HS
V_{dd}	1.8 V	1.8 V	1.8 V	2 V	1.5 V	1.2 V
T_{ox} (effective)	42 Å	42 Å	42 Å	42 Å	29 Å	24 Å
L_{gate}	0.16 μm	0.16 μm	0.18 μm	0.13 μm	0.11 μm	0.08 μm
I_{DSat} (n/p) (μA/μm)	600/260	500/180	320/130	780/360	860/370	920/400
I_{off} (leakage) (ρA/μm)	20	1.60	0.15	300	1,800	13,000
V_{Tn}	0.42 V	0.63 V	0.73 V	0.40 V	0.29 V	0.25 V
FET Perf. (GHz)	30	22	14	43	52	80

Exponential Increase in Leakage Currents



From Dc,1999

Energy & Power Equations

$$E = C_L V_{DD}^2 P_{0\to1} + t_{sc} V_{DD} I_{peak} P_{0\to1} + V_{DD} I_{leakage}$$

$$f_{0\to1} = P_{0\to1} * f_{clock}$$

$$P = C_L V_{DD}^2 f_{0\to1} + t_{sc} V_{DD} I_{peak} f_{0\to1} + V_{DD} I_{leakage}$$

Dynamic power (~90% today and decreasing relatively)

Short-circuit power (~8% today and decreasing absolutely)

Leakage power (~2% today and increasing)

کل تلفات

$$P_{tot} = P_{dyn} + P_{dp} + P_{stat} = (C_L V_{DD}^2 + V_{DD} I_{peak} t_s) f_{0\to 1} + V_{DD} I_{leak}$$

حاصلضرب توان در تاخیر و انرژی درتاخیر

پارامتر PDP که حاصلضرب توان در تاخیر است. این پارامتر معیاری از میزان انرژی مصرفی می باشد. اگر از تلفات استاتیک و تلفات مسیر جریان مستقیم صرفنظر کنیم می توان دید PDP معادل انرژی مصرفی گیت بازای یک سوییچینگ می باشد.

$$PDP = C_L V_{DD}^2 f_{max} t_p = \frac{C_L V_{DD}^2}{2}$$

پارامتر قابل توجه دیگر حاصلضرب انرژی در تاخیر می باشد که معیاری از سرعت و انرژی گیت است.

$$EDP = PDP \times t_p = P_{av} t_p^2 = \frac{C_L V_{DD}^2}{2} t_p \qquad t_p \approx \frac{\alpha C_L V_{DD}}{V_{DD} - V_{Te}} \qquad V_{Te} = V_T + V_{DSAT}/2$$

$$EDP = \frac{\alpha C_L^2 V_{DD}^3}{2(V_{DD} - V_{TE})}$$

مثال: ولتاژ تغذیه بهینه

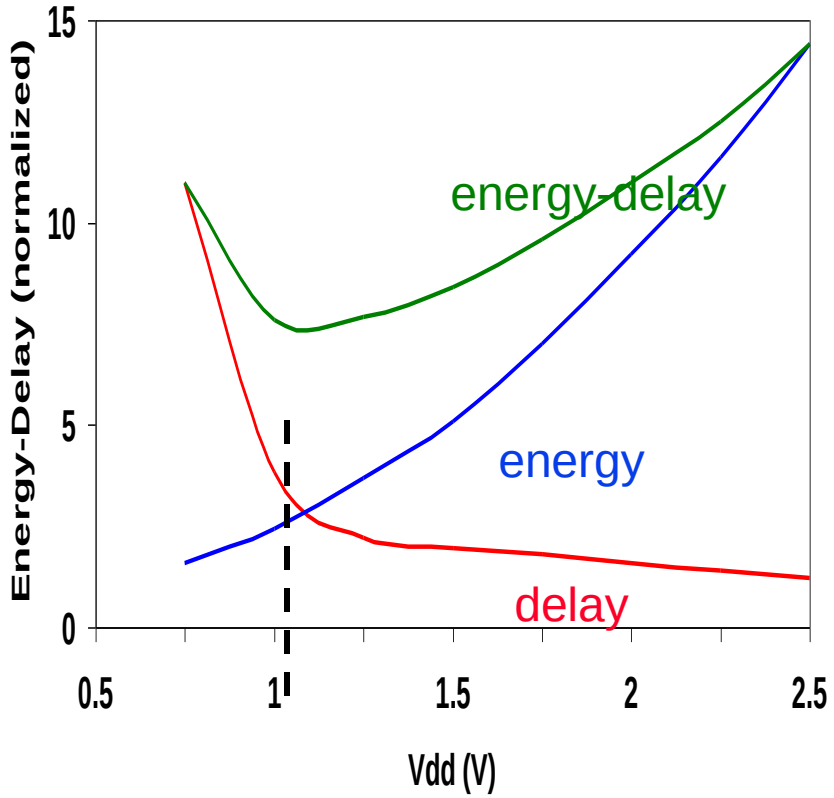
محاسبه ولتاژ بهینه برای حداقل کردن EDP

$$V_{DDopt} = \frac{3}{2} V_{TE}$$

$$V_{Tn} = 0.43 \text{ V}, V_{Dsatn} = 0.63 \text{ V}, V_{TEn} = 0.74 \text{ V}.$$
$$V_{Tp} = -0.4 \text{ V}, V_{Dsatp} = -1 \text{ V}, V_{TEp} = -0.9 \text{ V}.$$
$$V_{TE} \approx (V_{TEn} + |V_{TEp}|)/2 = 0.8 \text{ V}$$

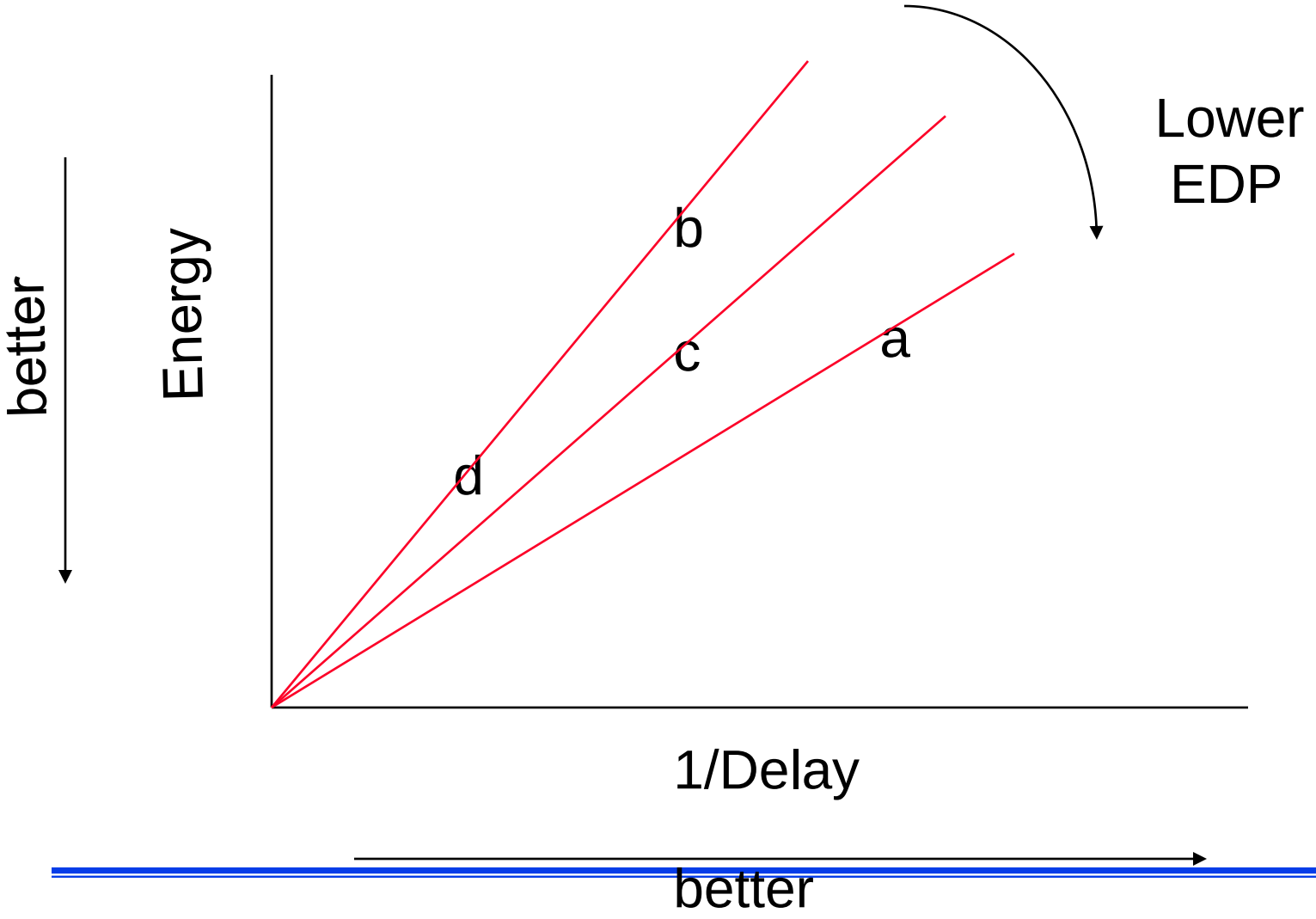
$$V_{DDopt} = (3/2) \times 0.8 \text{ V} = 1.2 \text{ V}.$$

- EDP is the average energy consumed multiplied by the computation time required
- takes into account that one can trade increased delay for lower energy/operation (e.g., via supply voltage scaling that increases delay, but decreases energy consumption)



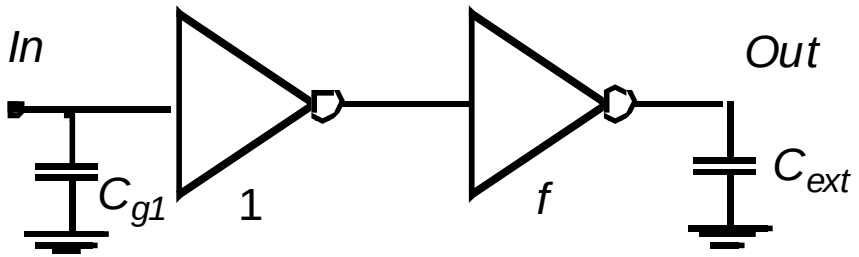
Understanding Tradeoffs

- Which design is the “best” (fastest, coolest, both) ?



Power and Energy Design Space

	Constant Throughput/Latency		Variable Throughput/Latency
Energy	Design Time	Non-active Modules	Run Time
Active	Logic Design Reduced V_{dd} Sizing Multi-V_{dd}	Clock Gating	DFS, DVS (Dynamic Freq, Voltage Scaling)
Leakage	+ Multi-V_T	Sleep Transistors Multi-V_{dd} Variable V_T	+ Variable V_T



- Goal: Minimize Energy of whole circuit
  - Design parameters: f and V_{DD}
  - $tp \leq tpref$ of circuit with f=1 and $V_{DD} = V_{ref}$

$$t_p = t_{p0}\left(\left(1+\frac{f}{\gamma}\right)+\left(1+\frac{F}{f\gamma}\right)\right)$$

$$t_{p0} \propto \frac{V_{DD}}{V_{DD}-V_{TE}}$$

Transistor Sizing

- Performance Constraint (γ=1)

$$\frac{t_p}{t_{pref}} = \frac{t_{p0}}{t_{p0ref}} \frac{\left(2+f+\frac{F}{f}\right)}{(3+F)} = \frac{V_{DD}}{V_{ref}} \frac{V_{ref}-V_{TE}}{V_{DD}-V_{TE}} \frac{\left(2+f+\frac{F}{f}\right)}{(3+F)} = 1$$

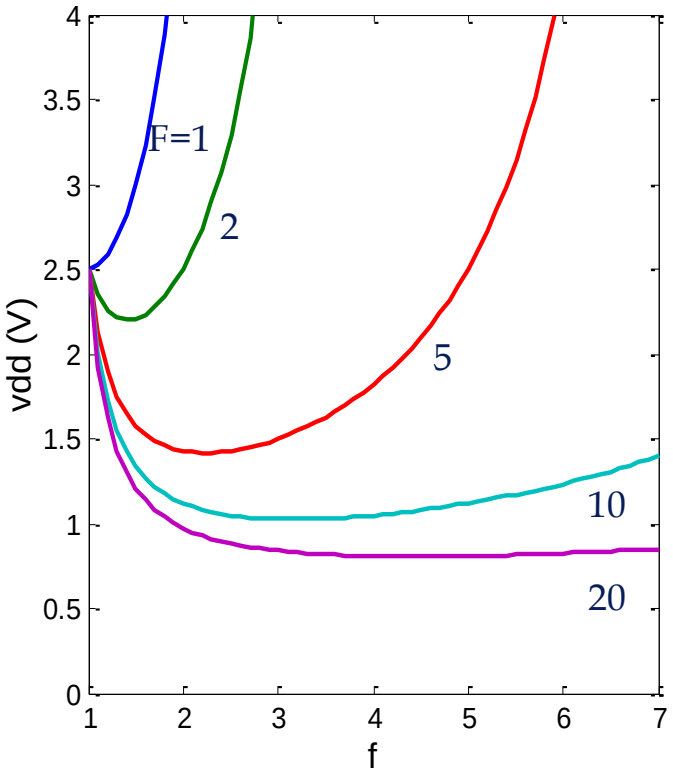
- Energy for single Transition

$$E = V_{DD}^2 C_{g1}\left[(1+\gamma)(1+f)+F\right]$$

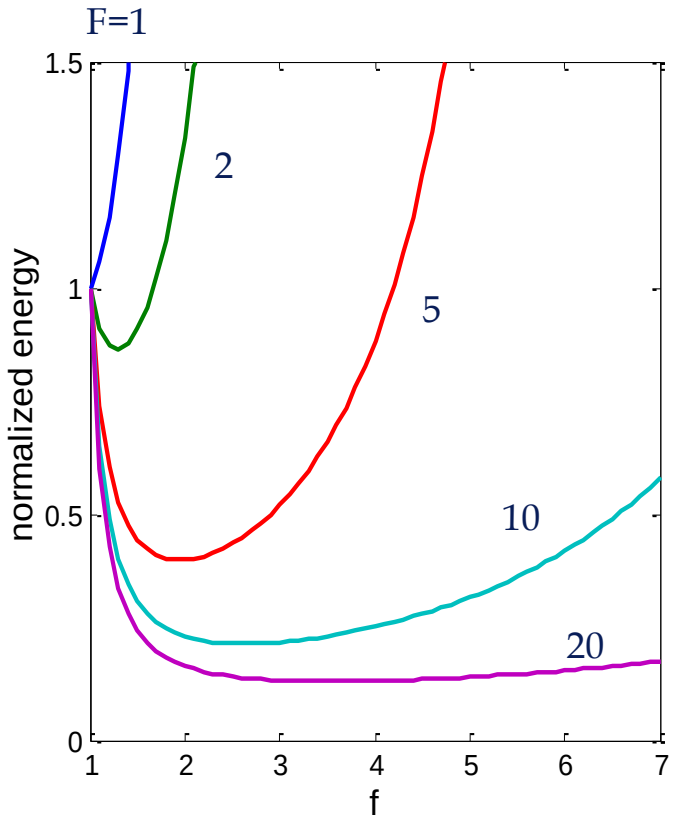
$$\frac{E}{E_{ref}} = \left(\frac{V_{DD}}{V_{ref}}\right)^2 \left(\frac{2+2f+F}{4+F}\right)$$

Transistor Sizing

$V_{DD}=f(f)$

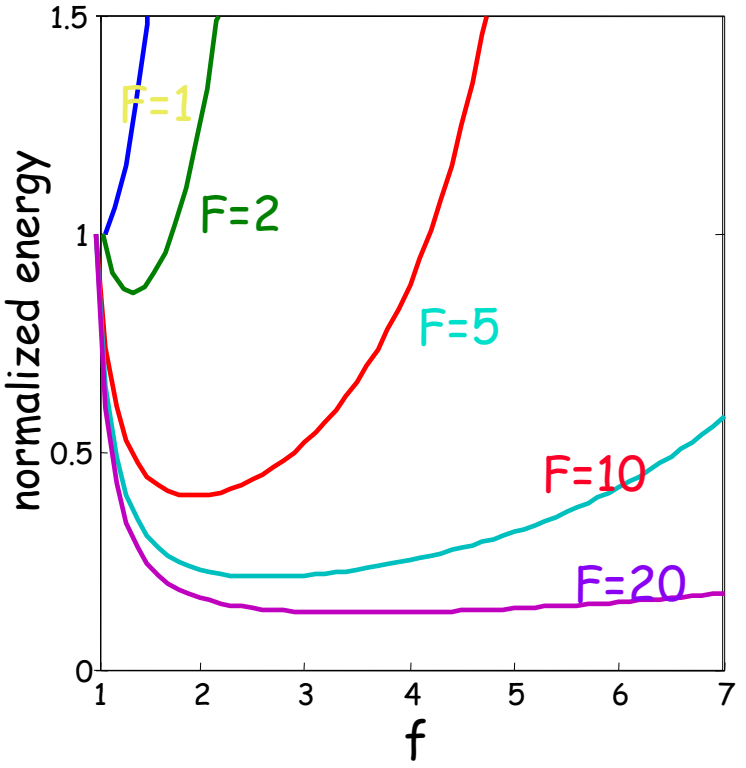


$E/E_{ref}=f(f)$



Dynamic Power as a Function of Device Size

- Device sizing affects dynamic energy consumption
  - gain is largest for networks with large overall effective fan-outs ($F = C_L/C_{g,1}$)
- The optimal gate sizing factor (f) for dynamic energy is smaller than the one for performance, especially for large F's
  - e.g., for F=20, f_{opt}(energy) = 3.53 while f_{opt}(performance) = 4.47
- If energy is a concern avoid oversizing beyond the optimal

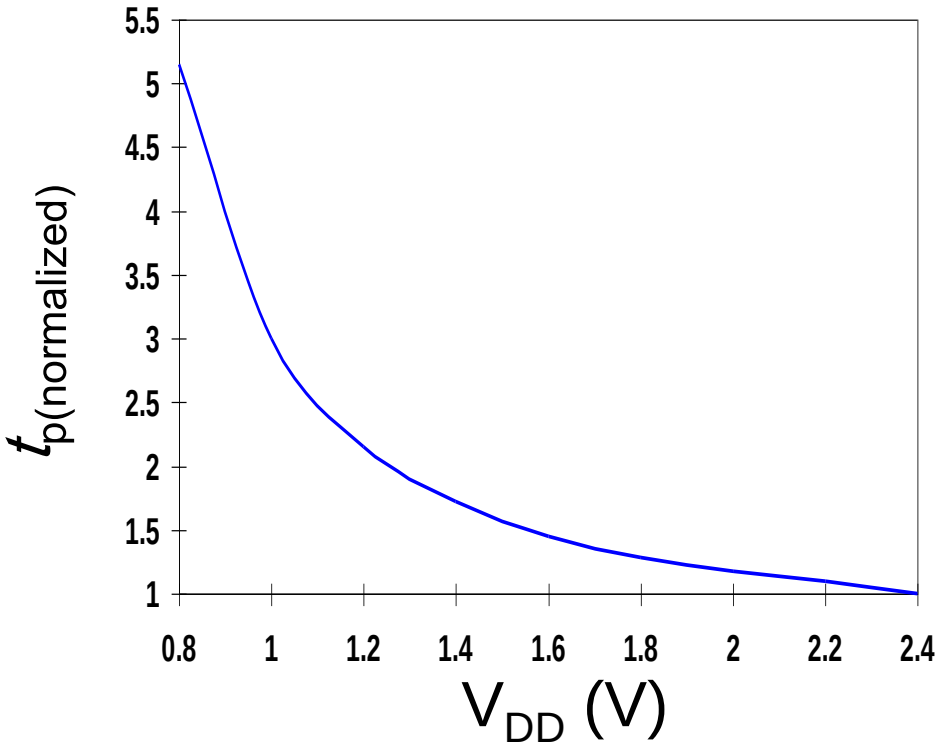


Multi VDD

- Prime choice: Reduce voltage!
  - Recent years have seen an acceleration in supply voltage reduction (USE Multi VDD)
- Reduce switching activity (How?)
- Reduce physical capacitance
  - Device Sizing: for F=20
    - f_{opt}(energy)=3.53, f_{opt}(performance)=4.47

Dynamic Power as a Function of V_{DD}

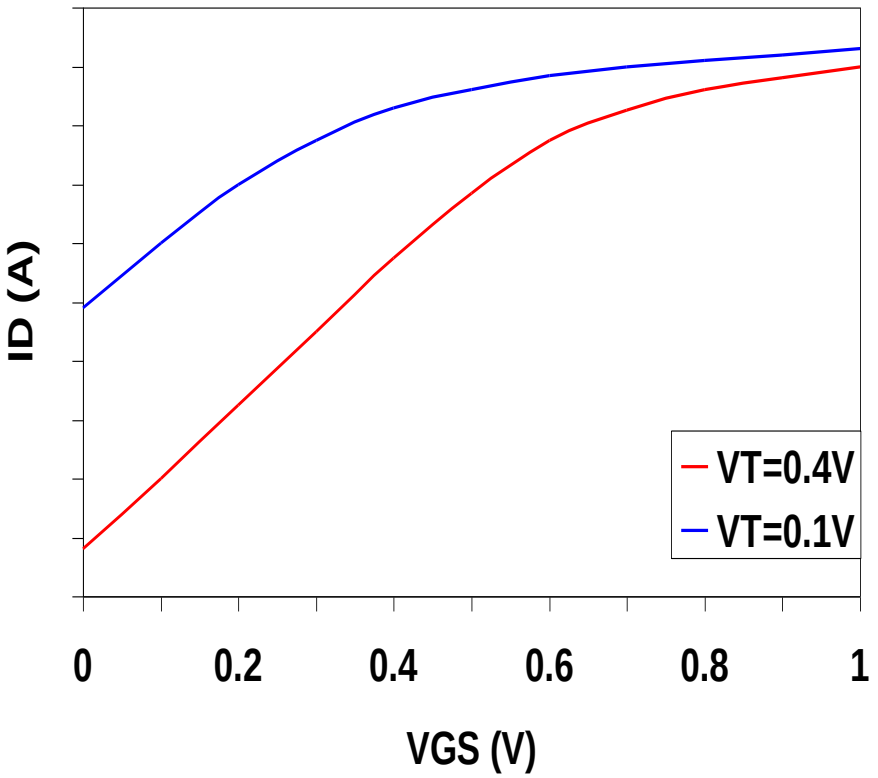
- Decreasing the V_{DD} decreases dynamic energy consumption (quadratically)
- But, increases gate delay (decreases performance)



- Determine the critical path(s) at design time and use high V_{DD} for the transistors on those paths for speed. Use a lower V_{DD} on the other gates, especially those that drive large capacitances (as this yields the largest energy benefits).

Multi V_T

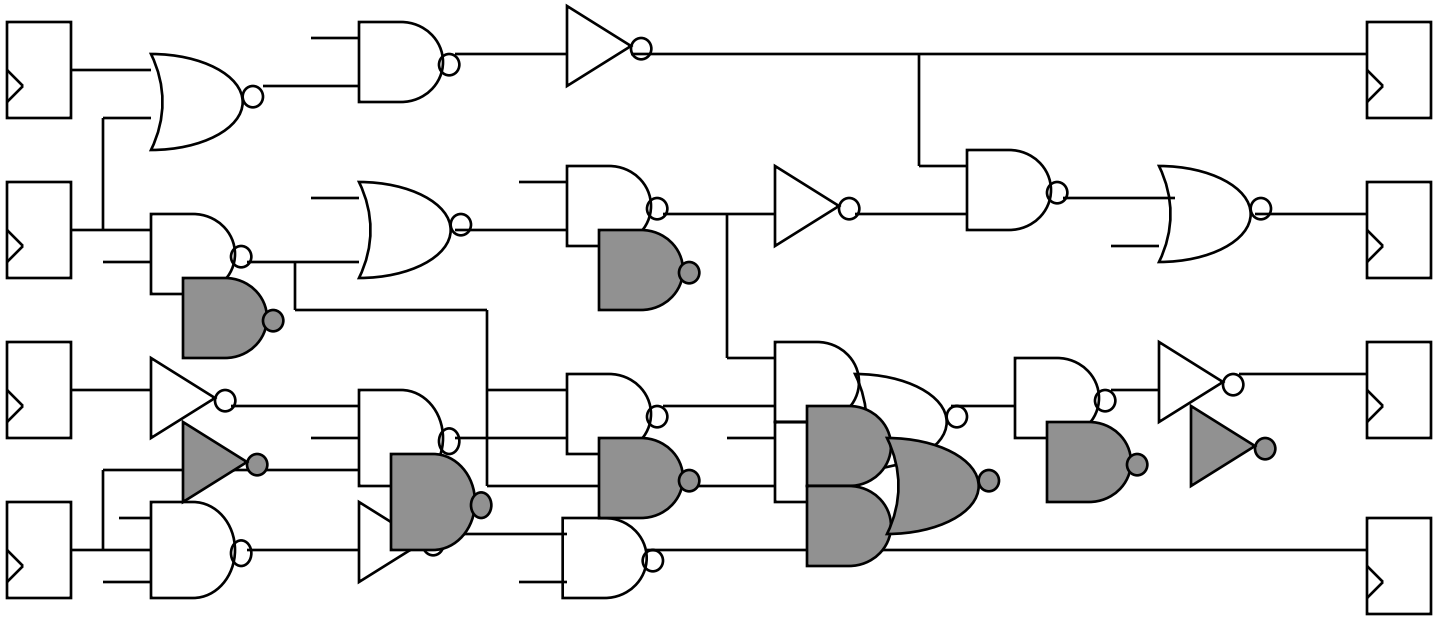
- Reducing the V_T increases the sub-threshold leakage current (exponentially)
  - 90mV reduction in V_T increases leakage by an order of magnitude
- But, reducing V_T decreases gate delay (increases performance)



- Determine the critical path(s) at design time and use low V_T devices on the transistors on those paths for speed. Use a high V_T on the other logic for leakage control.
  - A careful assignment of V_T's can reduce the leakage by as much as 80%

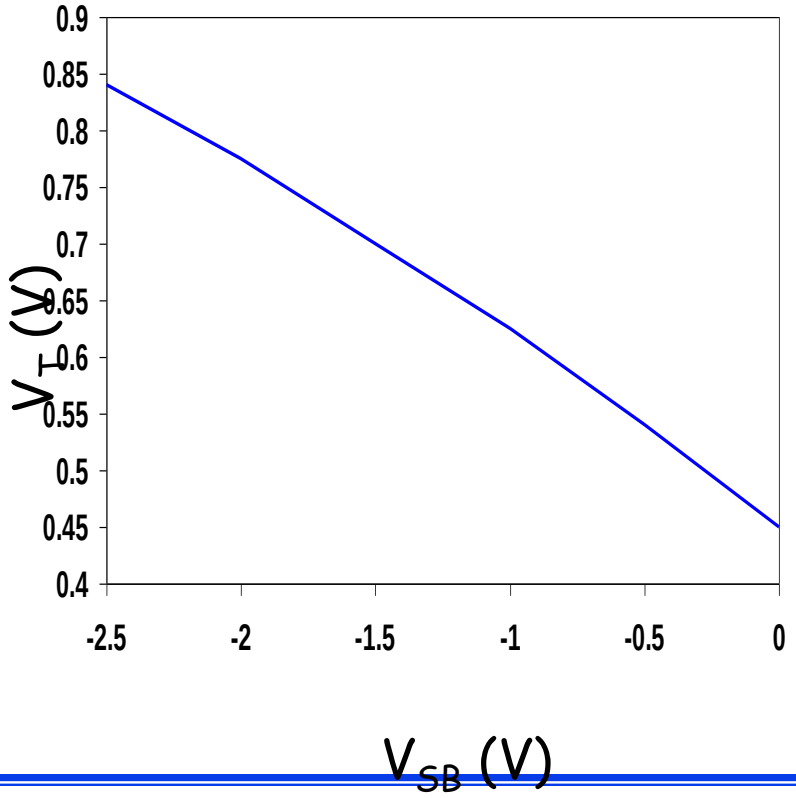
Dual-Thresholds Inside a Logic Block

- Minimum energy consumption is achieved if all logic paths are critical (have the same delay)
- Use lower threshold on timing-critical paths
  - Assignment can be done on a per gate or transistor basis;



Variable V_T (ABB)

- $V_T = V_{T0} + \gamma(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|})$
- For an n-channel device, the substrate is normally tied to ground ($V_{SB} = 0$)
- A negative bias on V_{SB} causes V_T to increase
- Adjusting the substrate bias at run time is called adaptive body-biasing (ABB)
  - Requires a dual well fab process



Logic Design Optimization

- Switching activity, $P_{0\rightarrow1}$, has two components
  - A static component – function of the logic topology
  - A dynamic component – function of the timing behavior (glitching)

2-input NOR Gate

A	B	Out
0	0	1
0	1	0
1	0	0
1	1	0

Static transition probability

$$P_{0\rightarrow1} = P_{out=0} \times P_{out=1}$$
$$= P_0 \times (1-P_0)$$

With input signal probabilities

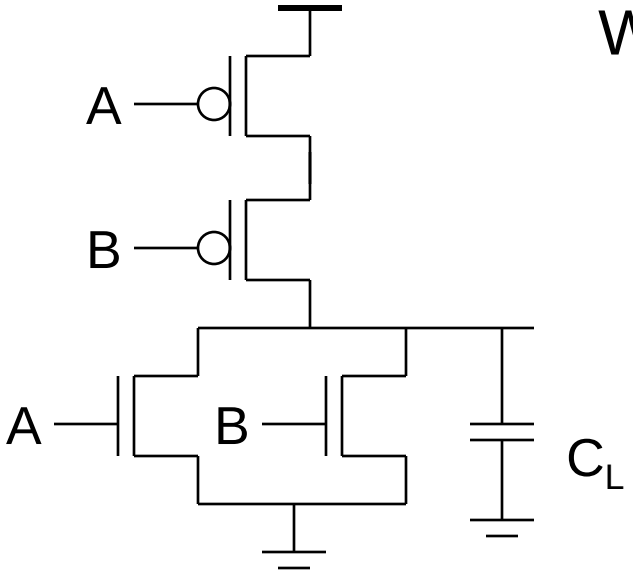
$$P_{A=1} = 1/2$$
$$P_{B=1} = 1/2$$

NOR static transition probability

$$= 3/4 \times 1/4 = 3/16$$

NOR Gate Transition Probabilities

- Switching activity is a strong function of the input signal statistics
  - P_A and P_B are the probabilities that inputs A and B are one



With input signal probabilities

$P_{A=1} = 1/2$

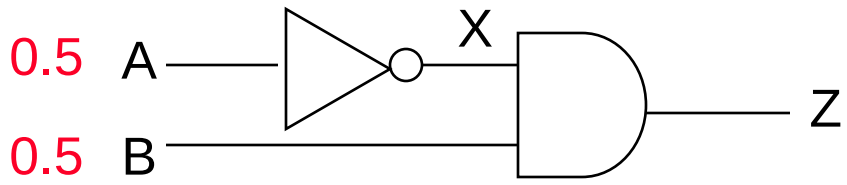
$P_{B=1} = 1/2$

A	B	Out
0	0	1
0	1	0
1	0	0
1	1	0

$$P_{0\rightarrow1} = P_0 \times P_1 = (1-(1-P_A)(1-P_B))\,(1-P_A)(1-P_B)$$

Transition Probabilities for Some Basic Gates

	$P_{0\to1} = P_{out=0} \times P_{out=1}$
NOR	$(1 - (1 - P_A)(1 - P_B)) \times (1 - P_A)(1 - P_B)$
OR	$(1 - P_A)(1 - P_B) \times (1 - (1 - P_A)(1 - P_B))$
NAND	$P_A P_B \times (1 - P_A P_B)$
AND	$(1 - P_A P_B) \times P_A P_B$
XOR	$(1 - (P_A + P_B - 2P_A P_B)) \times (P_A + P_B - 2P_A P_B)$



For X: $P_{0\to1} = P_0 \times P_1 = (1-P_A) P_A$

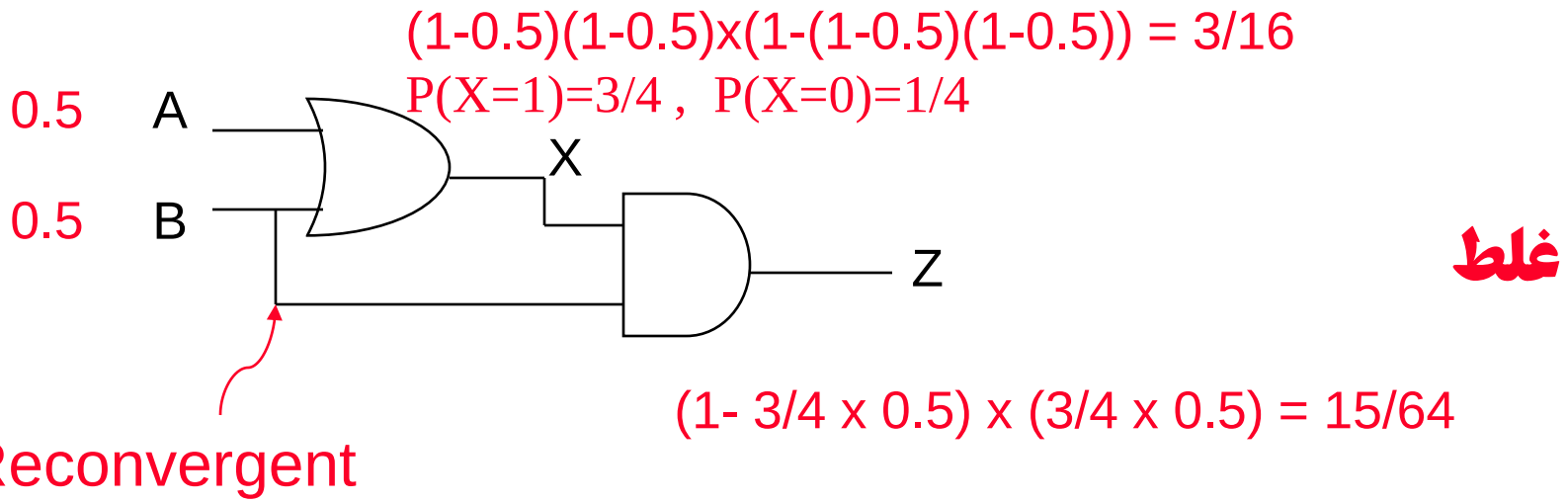
$= 0.5 \times 0.5 = 0.25$

For Z: $P_{0\to1} = P_0 \times P_1 = (1-P_X P_B) P_X P_B$

$= (1 - (0.5 \times 0.5)) \times (0.5 \times 0.5) = 3/16$

Inter-signal Correlations

- Determining switching activity is complicated by the fact that signals exhibit correlation in space and time
  - reconvergent fan-out



$$P(Z=1) = P(B=1) \;\&\; P(A=1 \text{ OR } B=1)$$

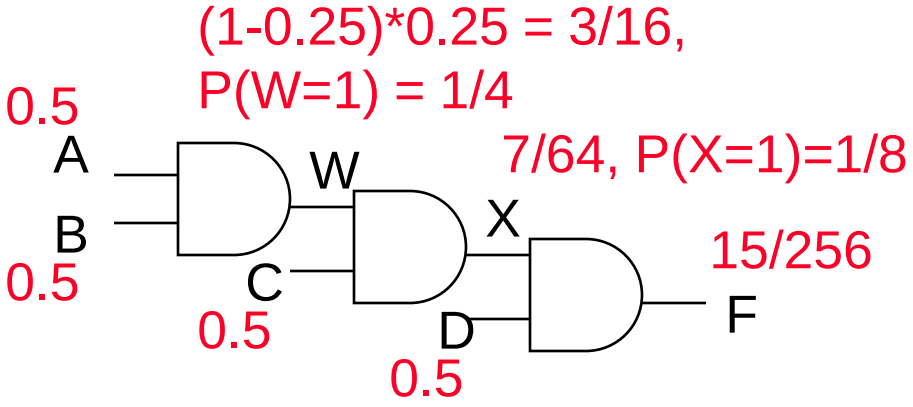
$Z=(A+B)B=AB+B=B(1+A)=B$ **P(Z=1)=1/2** P(0>1)=1/4

- Have to use conditional probabilities

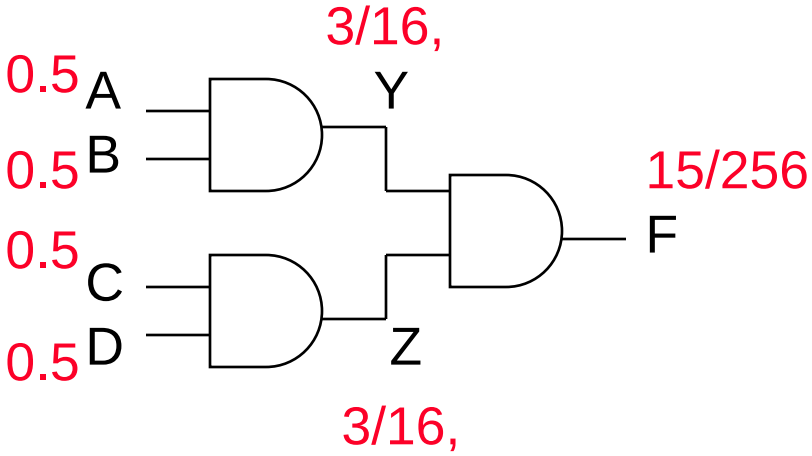
Logic Restructuring

- Logic restructuring: changing the topology of a logic network to reduce transitions

AND: $P_{0\to1} = P_0 \times P_1 = (1 - P_A P_B) \times P_A P_B$



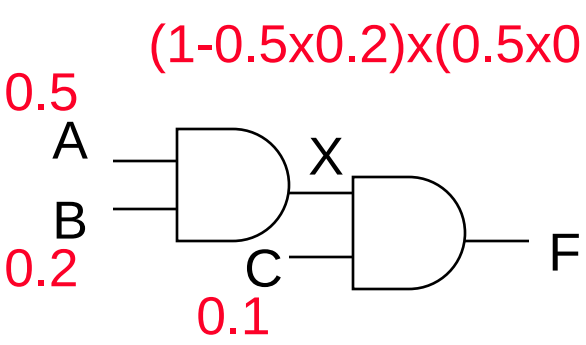
Σ=0.355

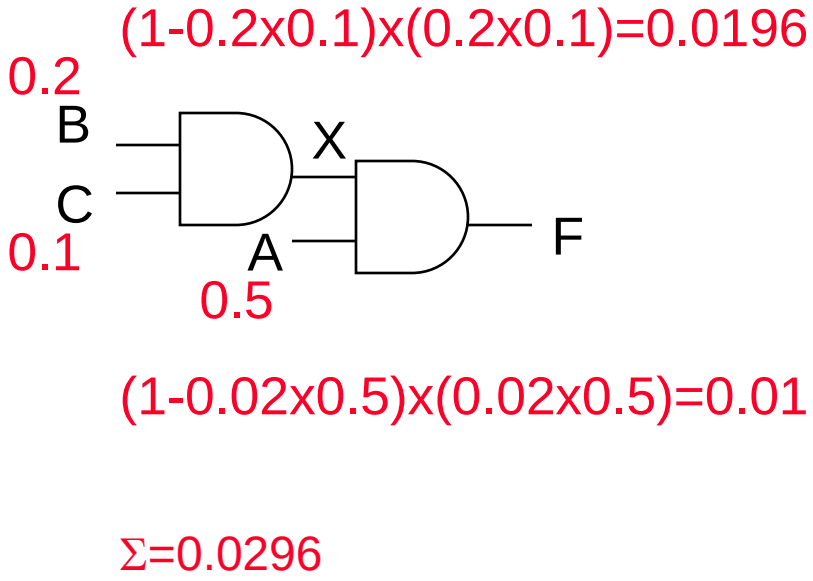


Σ=0.433

Chain implementation has a lower overall switching activity than the tree implementation for random inputs

Input Ordering





Beneficial to postpone the introduction of signals with a high transition rate (signals with signal probability close to 0.5)

جلسه آینده

- درس بعدی
  - گیت های ترکیبی