

FÖRELÄSNING 7

Fördröjningsminimering vid buffring

ON-resistansen

Energiåtgång och effektförbrukning i CMOS

RAM-minnet

Fördröjningsminimering vid buffring

(S&S4 5.8/S&S5: 4.10.4, 10.1.2, 10.2.4)

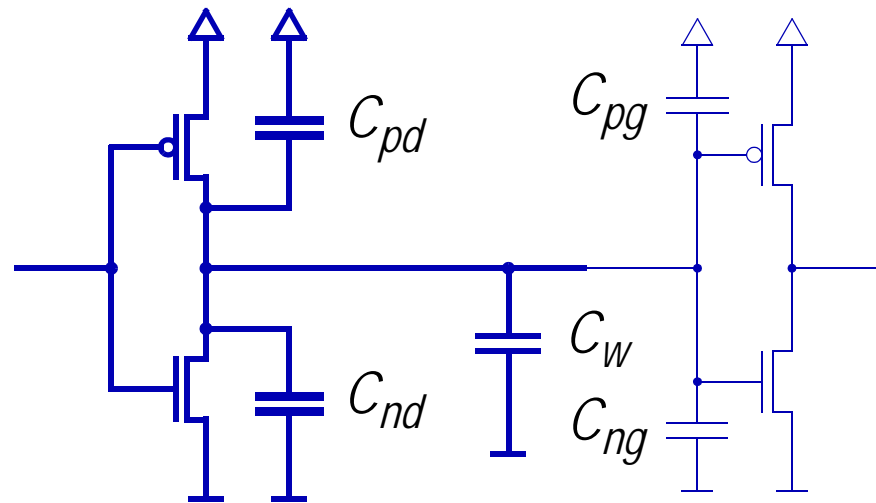
DIGITALT FINLIR - ATT SAMOPTIMERA TVÅ GRINDAR

- ◆ På den förra föreläsningen kom vi fram till att fördröjningen för en CMOS-inverterare berodde på såväl transistorns bredd som lastkapacitans:

$$t_d \propto \frac{C_L}{\mu C_{ox} \cdot \frac{W}{L} \cdot V_{DD}}$$

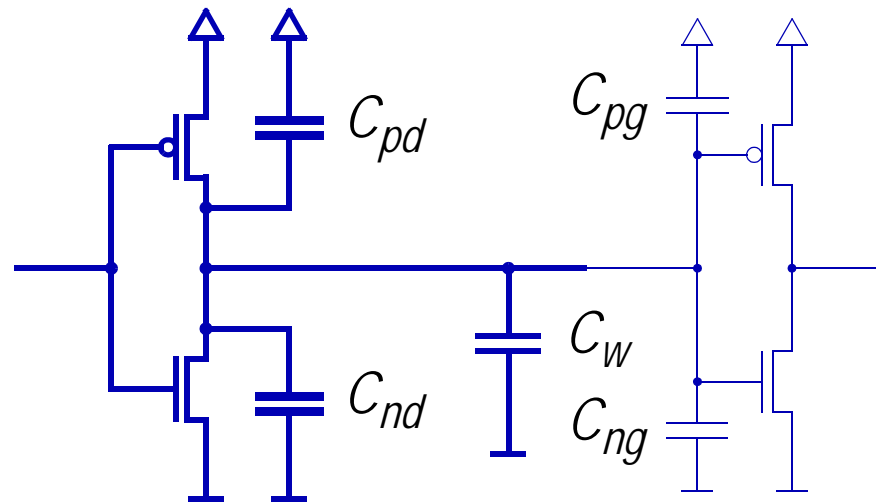
- ◆ Fördröjningen kan alltså minskas genom att man dels håller ned lastkapacitansen C_L och dels ökar bredden W på transistorerna.
- ◆ Fast förstås ... att öka W medför att man ökar kapacitansen också!

EN BELASTAD INVERTERARES FÖRDRÖJNING 1(4)



- ◆ Vår inverterare ska driva en andra inverterare och vill minimera den totala fördröjningen, genom båda kretsarna. En vanlig utmaning — kallas buffring!
- ◆ Vi vill nu optimera förhållandet mellan bredderna på PMOS och NMOS.
- ◆ På den första inverterarens utgång är $C_L = C_{pd} + C_{nd} + C_{pg} + C_{ng} + C_w$ (terminologi: pd = PMOS drain, w = wire).

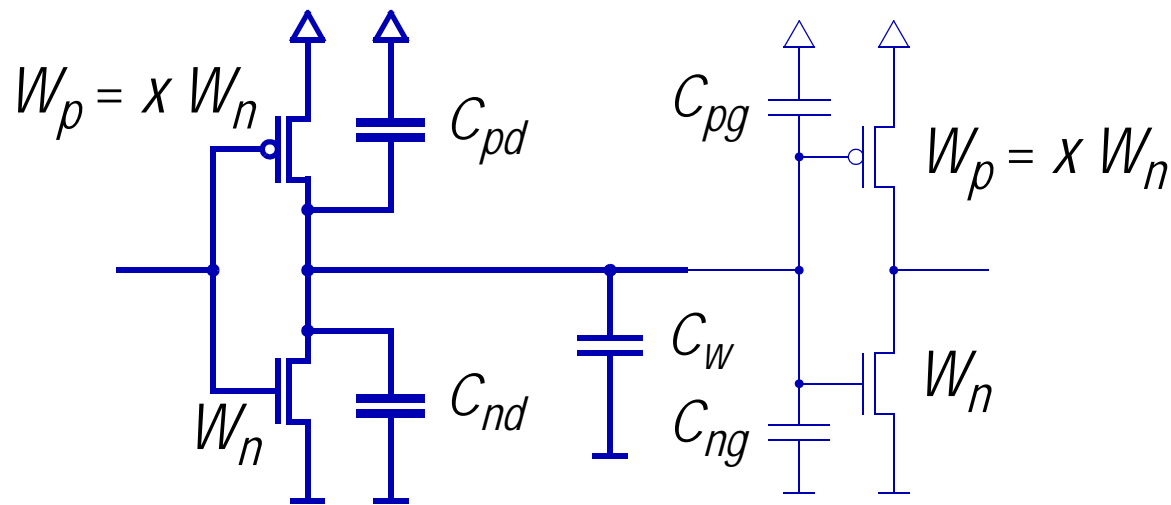
EN BELASTAD INVERTERARES FÖRDRÖJNING 2(4)



1. Vi vill bredda PMOS:en i den första inverteraren, för att minska fördröjningen, men för att ha konsekventa designregler genomför vi även breddökningen på den belastande inverteraren:

$$W_p = x \cdot W_n, \text{ d.v.s. } C_L = (1 + x)(C_{nd} + C_{ng}) + C_w$$

EN BELASTAD INVERTERARES FÖRDRÖJNING 3(4)

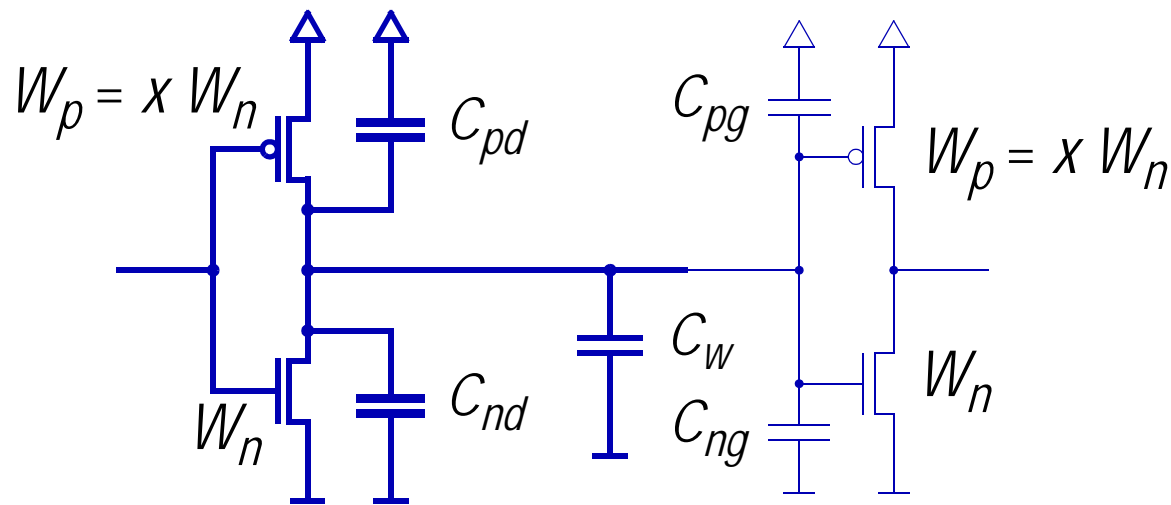


$$t_d \propto \frac{C_L}{\mu C_{ox} \cdot \frac{W}{L} \cdot V_{DD}}$$

2. Fördröjningen i första inverteraren: $t_d = \frac{L C_L}{2 C_{ox} V_{DD}} \left(\frac{1}{W_p \mu_p} + \frac{1}{W_n \mu_n} \right) \Rightarrow$

$$t_d = \frac{L [(1 + x)(C_{nd} + C_{ng}) + C_w]}{2 C_{ox} W_n \mu_n V_{DD}} \left(1 + \frac{\mu_n}{x \cdot \mu_p} \right).$$

EN BELASTAD INVERTERARES FÖRDRÖJNING 4(4)

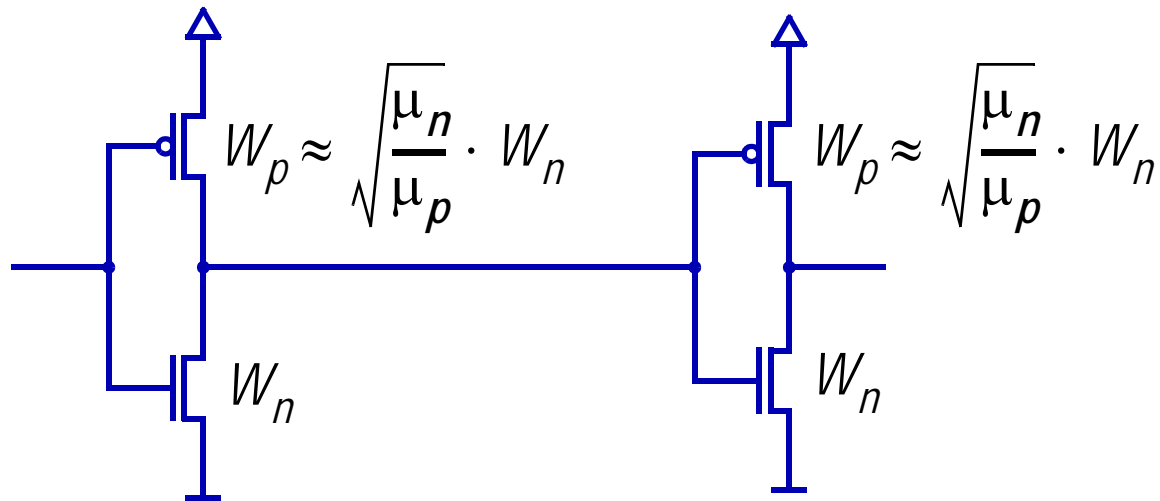


$$t_d \propto \frac{C_L}{\mu C_{ox} \cdot \frac{W}{L} \cdot V_{DD}}$$

3. Minimum: $\frac{\partial t_d}{\partial x} = 0 \Rightarrow x = \sqrt{\frac{\mu_n}{\mu_p} \left(1 + \frac{C_w}{C_{nd} + C_{ng}} \right)}$

(C_w är liten endast när ledningen sitter mellan två lokala grindar).

DEN OPTIMALA “SIZING’EN”



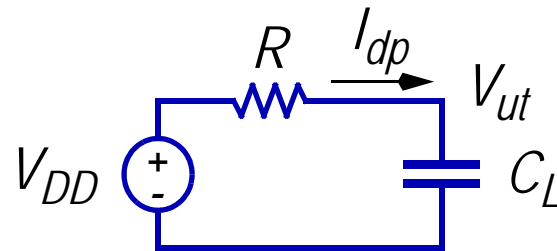
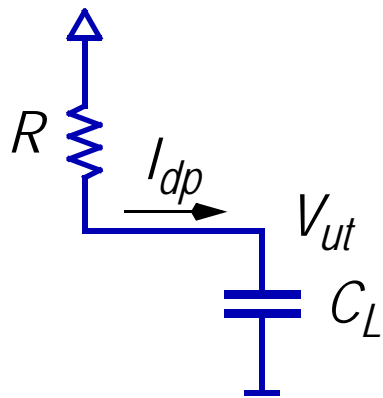
- ◆ Dessa förhållanden mellan bredderna är giltiga när man tittar på två kaskadkopplade inverterare som analyseras fristående från övriga kretsar.
- ◆ I verkligheten tillkommer fler kapacitanser, bl.a. från kretsar på ingången av den första inverteraren och på utgången av andra inverteraren.

ON-resistansen

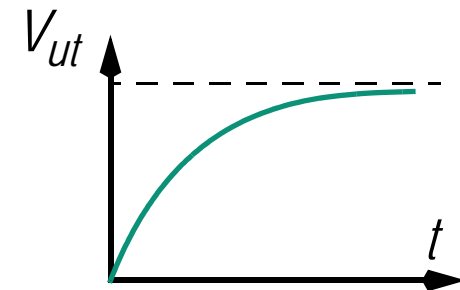
RC-APPROXIMATION

- ◆ På den förra föreläsningen härledde vi ett uttryck för hur uppladdningen av utgången på en inverterare går till.

Genom att använda en resistans som (approximativ!) modell för PMOS-transistorn fick vi en relativt lättanalyserad RC-krets:

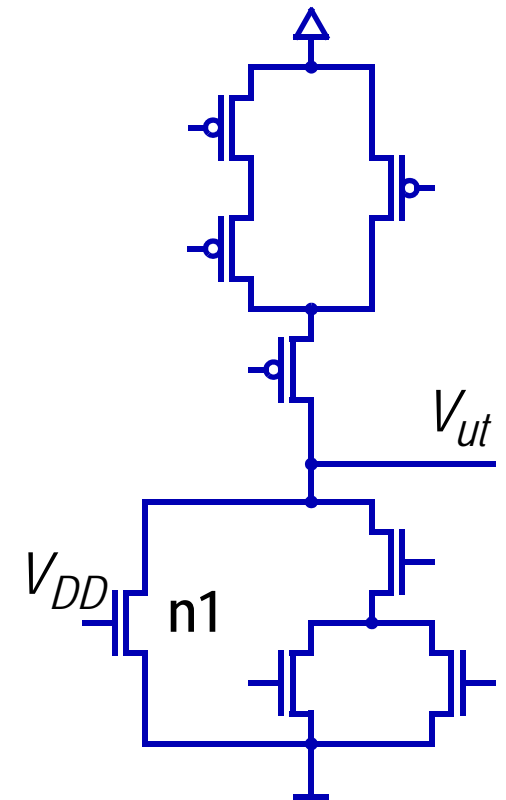


$$V_{ut}(t) = V_{DD} \cdot \left(1 - e^{-\frac{t}{RC_L}} \right)$$



MOS-TRANSISTORN SOM STRÖMBRYTARE

- ◆ Ta t.ex. NMOS-transistor **n1**, i kretsen till höger vilken vi byggde under den förra föreläsningen:
- ◆ När ingången till **n1** blir logiskt hög och utgången då laddas ur genom **n1**, vilken resistans representerar då n1:s kanal?
- ◆ Låt oss studera omslaget:
 1. Omslaget påbörjas och V_{ut} börjar falla från V_{DD} ...
 2. ... snart har V_{ut} fallit ned till $V_{DD} - V_T$. Men $V_{GS} = V_{DD}$, vilket betyder att $V_{ut} < V_{GS} - V_T$!
- ◆ Alltså, **n1** går in i sitt linjära område ganska så snart efter att omslaget påbörjats. Det linjära området verkar intressant.



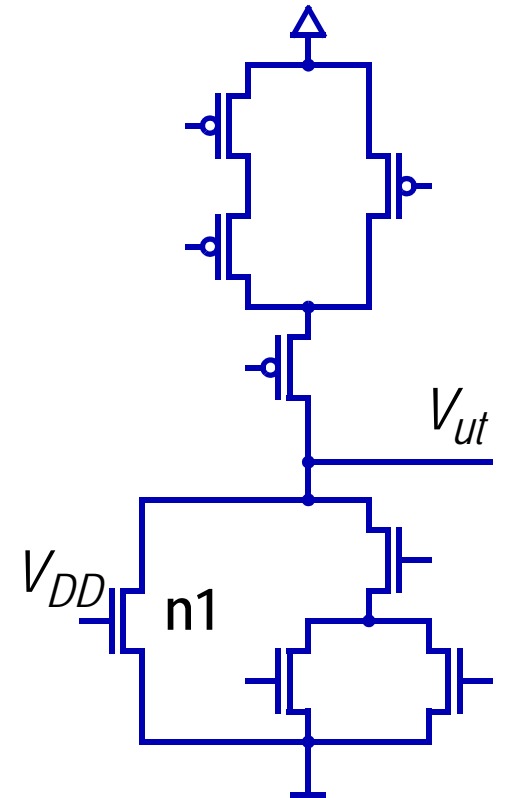
ON-RESISTANSEN HOS MOS-TRANSISTORN 1(5)

- ◆ Låt oss anta att V_{ut} (= V_{DS} för n1) till sist har fallit till en mycket låg spänning.
- ◆ Vi vet redan att n1 är i sitt linjära område. Men dessutom, ifall V_{DS} är mycket litet, blir V_{DS}^2 försumbart i uttrycket

$$I_D = k \left[(V_{GS} - V_T) V_{DS} - \frac{1}{2} V_{DS}^2 \right].$$

- ◆ Med ett uttryck som $I_D = k (V_{GS} - V_T) V_{DS}$ kan vi helt plötsligt skapa oss ett mått på kanalens ON-resistans:

$$R_{ON} = \frac{V_{DS}}{I_D} = \frac{1}{k (V_{GS} - V_T)} = \frac{1}{k (V_{DD} - V_T)}$$

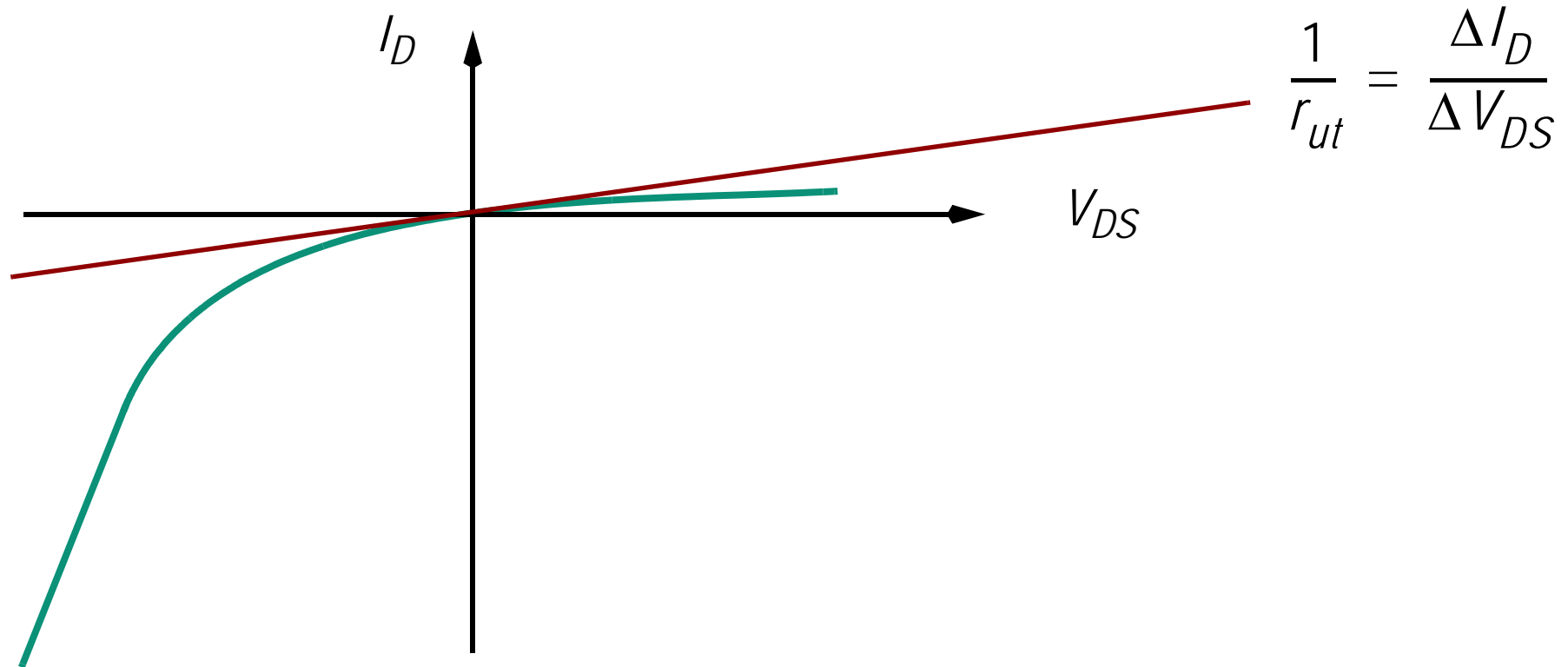


ON-RESISTANSEN HOS MOS-TRANSISTORN 2(5)

- ◆ När man räknar på ON-resistansen för en MOS-transistor, tänker man sig att kanalspänningen är nästan 0 V.
- ◆ Om man protesterar och säger att ON-resistansen inte kan vara särskilt representativ för dess egenskaper genom hela omslaget ... ja då har man en poäng.
- Men vi är rationella ingenjörer och vi approximerar när vi måste (*men vi håller samtidigt, genom vår grundmurade förståelse, ordning på ungefär hur stort fel vår gissning ger*).
- ◆ Ur en annan synvinkel är ON-resistansen helt logisk:
Hur mycket spänning ligger över en riktig strömbrytare som är påslagen?
Jo, cirka 0 V.
Kanalspänningen ska vara mycket liten, annars är transistorn inte påkopplad!

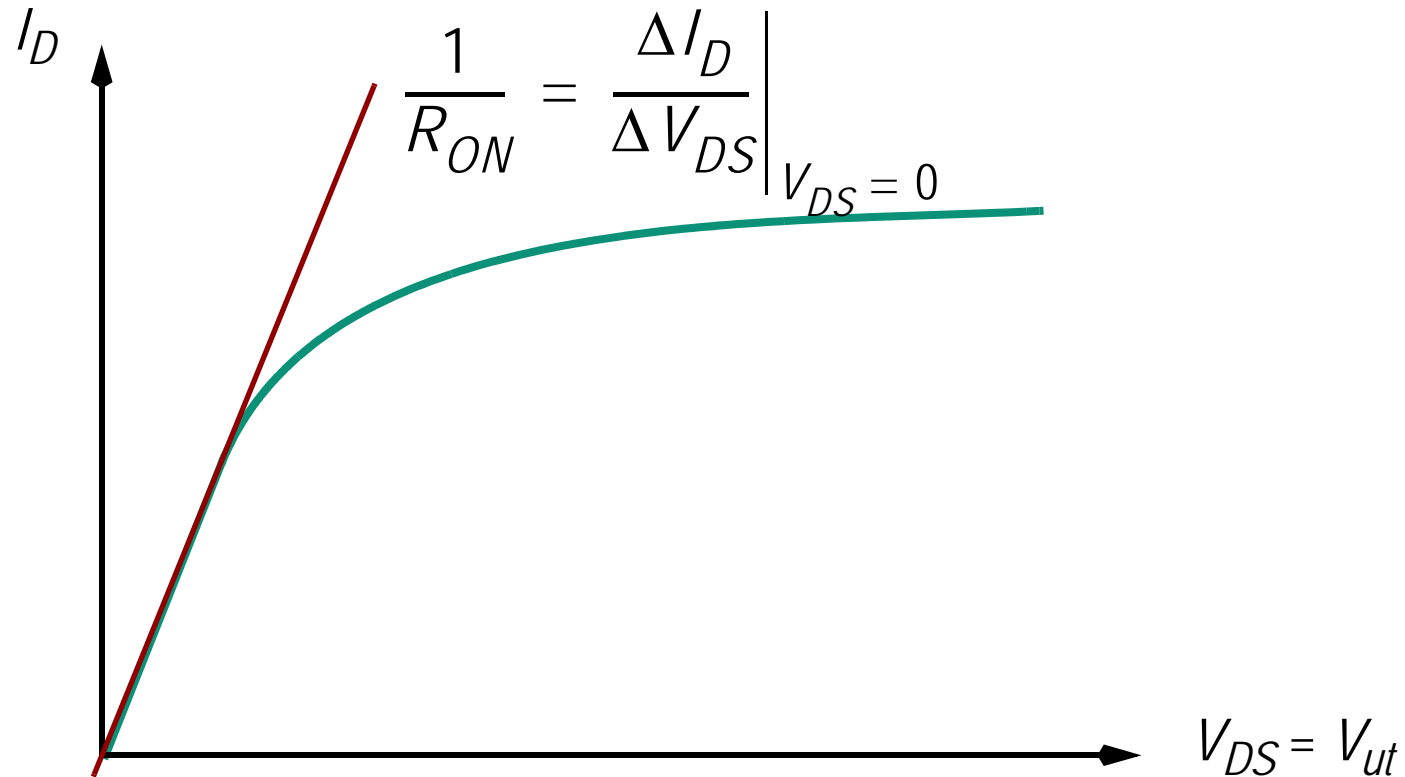
ON-RESISTANSEN HOS MOS-TRANSISTORN 3(5)

- ◆ Minns ni småsignalsresistansen för den mättade transistorn, i Föreläsning 4?
- ◆ Vi flyttade så att säga origo på grafen till arbetspunkten och mätte upp lutningen.



ON-RESISTANSEN HOS MOS-TRANSISTORN 4(5)

- ◆ Gör samma manöver som för småsignalsresistansen, fast nu med $V_{DS} = 0$.
- ◆ Denna resistans är en vanlig "storsignalsresistans" som approximerar en icke-linjär komponent. För $V = 0$ är $I = 0$ och Ohms lag $V = R I$ är uppfylld.

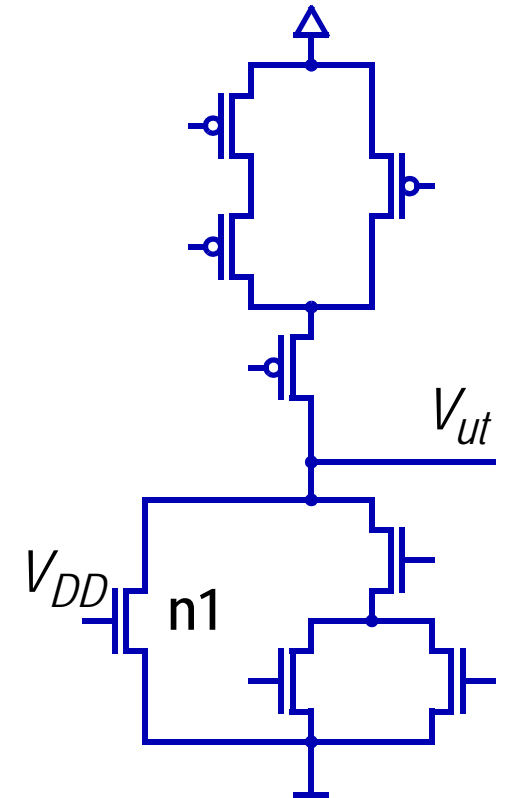


ON-RESISTANSEN HOS MOS-TRANSISTORN 5(5)

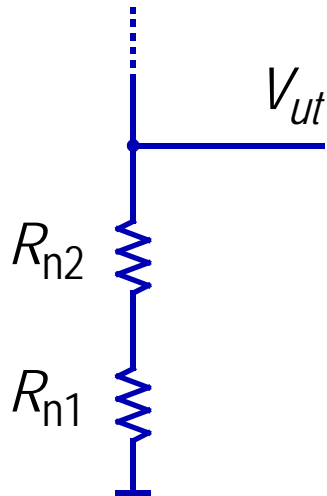
- ◆ ON-resistansen i kanalen kan tydligen skrivas

$$R_{ON} = \frac{1}{k(V_{DD} - V_T)} = \frac{L}{W} \cdot \frac{1}{\mu C_{ox}(V_{DD} - V_T)}$$

- ◆ Här kan W , L och V_T påverkas av kretskonstruktören. I digitala kretsar använder man oftast det minsta L som finns och V_T ändrar man inte i en handvändning.
- ◆ För att ta konstruktionsbeslut kan man alltså använda sig av ON-resistansen: $R_{ON} = \frac{1}{W} \cdot \text{konstant}$.
- ◆ *Så hur använder man egentligen måttet ON-resistans?*

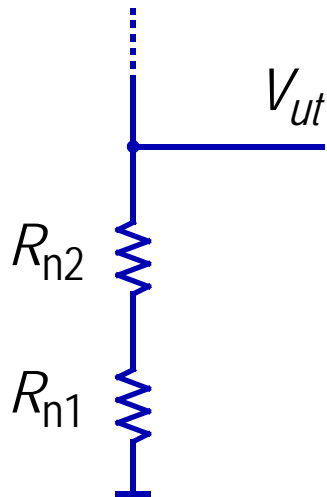


EKVIVALENT ON-RESISTANS 1(2)



- ◆ Vi tänker oss två seriekopplade NMOS:ar som representeras av två ON-resistanser.
- ◆ Det värsta som kan hända, fördröjningsmässigt, är att de två transistorerna slås på exakt samtidigt.
- ◆ *Kan man approximera hur de två transistorerna samverkar i ett sådant omslag?*

EKVIVALENT ON-RESISTANS 2(2)



- ◆ Utnyttjar vi begreppet ON-resistans kan vi ta fram en RC-tidskonstant $R_{eq} C_L$, där den ekvivalenta ON-resistansen R_{eq} är $R_{n1} + R_{n2}$.
- ◆ Skulle vi ha seriekopplat två 4- μm breda NMOS:ar, beter sig seriekopplingen som vore den en enda 2- μm bred NMOS-transistor.
- ◆ För det generella fallet kan vi formulera en ekvivalent bredd:

$$\frac{1}{W_{eq}} = \frac{1}{W_{n1}} + \frac{1}{W_{n2}}.$$

ILLUSTRATION AV "SIZING" 1(2)

1. Anta att NMOS-nätet till höger (en NOR-grind) klarar sitt fördröjningskrav med minimal bredd (med s.k. "minimum-size": $W_n = 2 \mu\text{m}$)
2. Sök fallet för "worst-case" fördröjning i NMOS-nätet!
Värsta fördröjningen uppstår om endast en av de två NMOS-transistorerna slår om (den andra är avstängd).
3. Bestäm den ekvivalenta bredden på PMOS-nätet!
PMOS-nätet som helhet ska vara "roten av mobilitetskvoten" (vilken råkar vara 2) gånger så stor som NMOS-nätet som helhet:

$$W_{p, eq} = 2 \cdot 2 = 4 \mu\text{m}$$

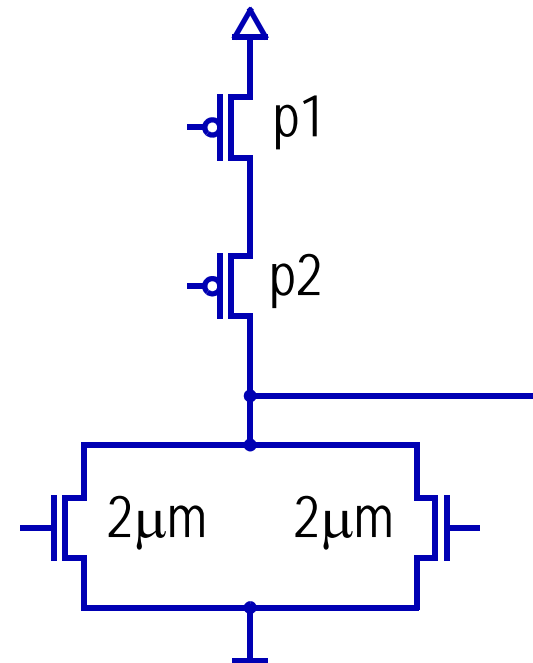


ILLUSTRATION AV "SIZING" 2(2)

4. Bredda PMOS:ar ("siza") för att uppfylla $W_{p,eq}$

Vi har alltså $\frac{1}{W_{p,eq}} = \frac{1}{W_{p1}} + \frac{1}{W_{p2}}$ och $W_{p1} = W_{p2}$,

vilket ger

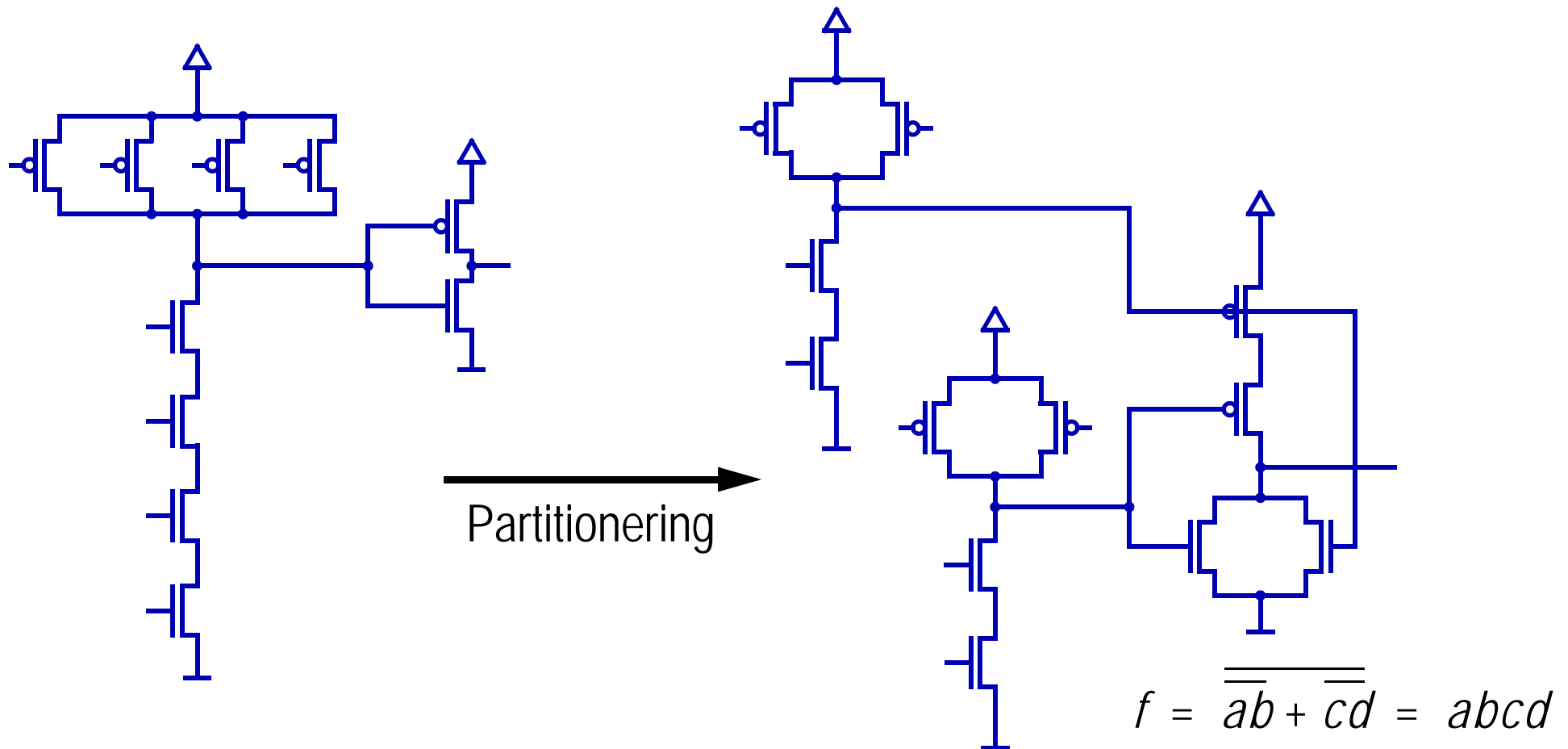
$$W_{p1} = W_{p2} = 8 \mu\text{m}!$$

- ◆ *Vad har vi gjort?* Jo, vi har "sizat" så att den värsta stigtiden (på utgång) som kan inträffa är lika stor som den värsta falltiden!

ATT PARTITIONERA GRINDAR

- ◆ Man kan skriva fördröjningen genom en grind som en funktion av både "fan-in" (FI) och "fan-out" (FO):
 - Fan-out anger hur många grindar som drivs av grinden som vi anger fan-out för.
 - Fan-in handlar om hur många ingångar grinden i fråga har, och därför anger fan-in i CMOS hur många seriekopplade MOS:ar vi återfinner i NMOS- eller PMOS-nätet.
- ◆ $t_d = a_1 FI + a_2 FI^2 + a_3 FO$.
- ◆ Som man kan se har fan-in större inflytande på fördröjningen än fan-out har. Tumregeln för grindar i CMOS brukar vara att undvika fler än tre (3) MOS-transistorer i serie.
Man ordnar detta genom partitionering.

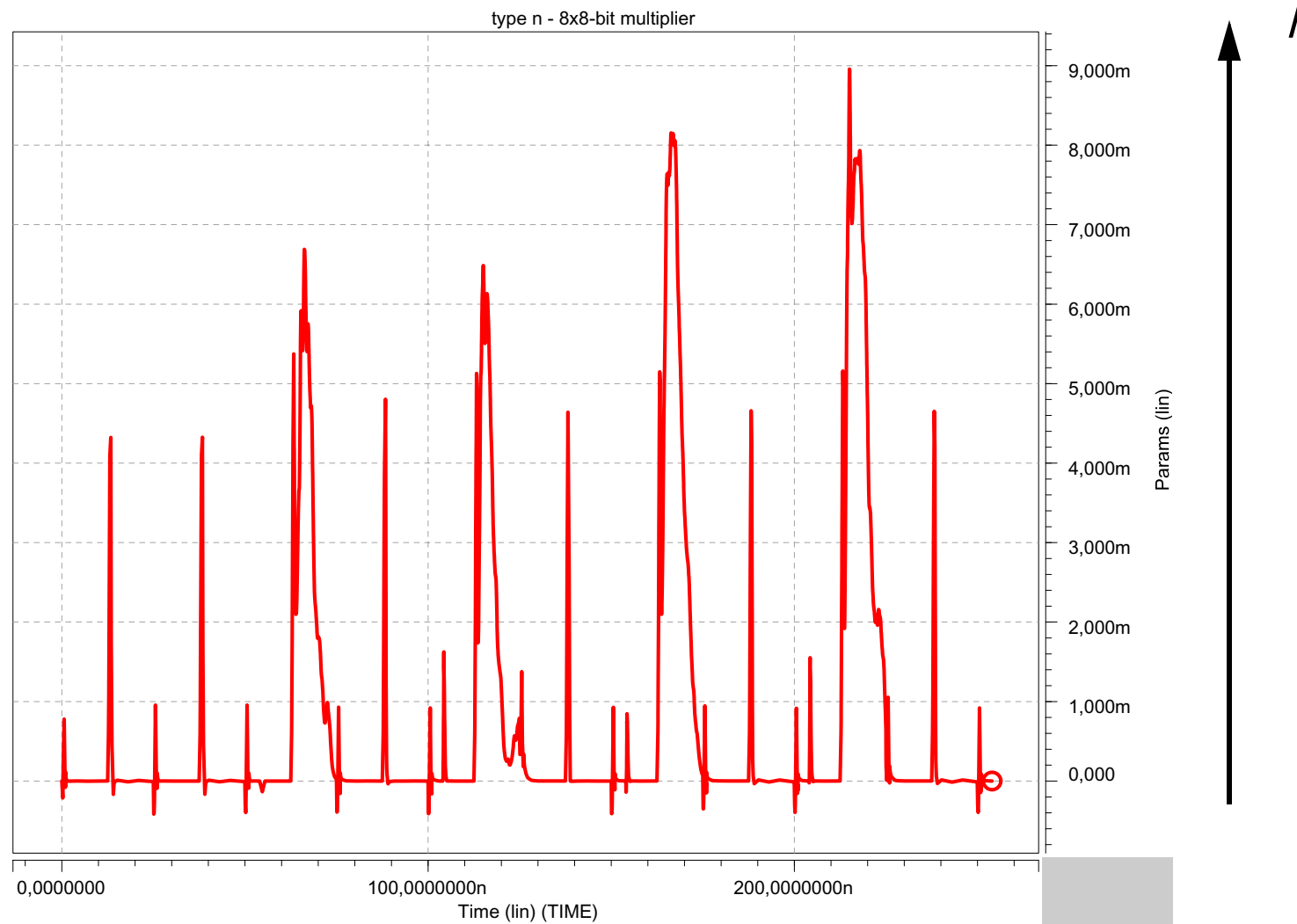
4-IN AND-GRIND - PARTITONERAD ELLER INTE?



Energiåtgång och effektförbrukning i CMOS

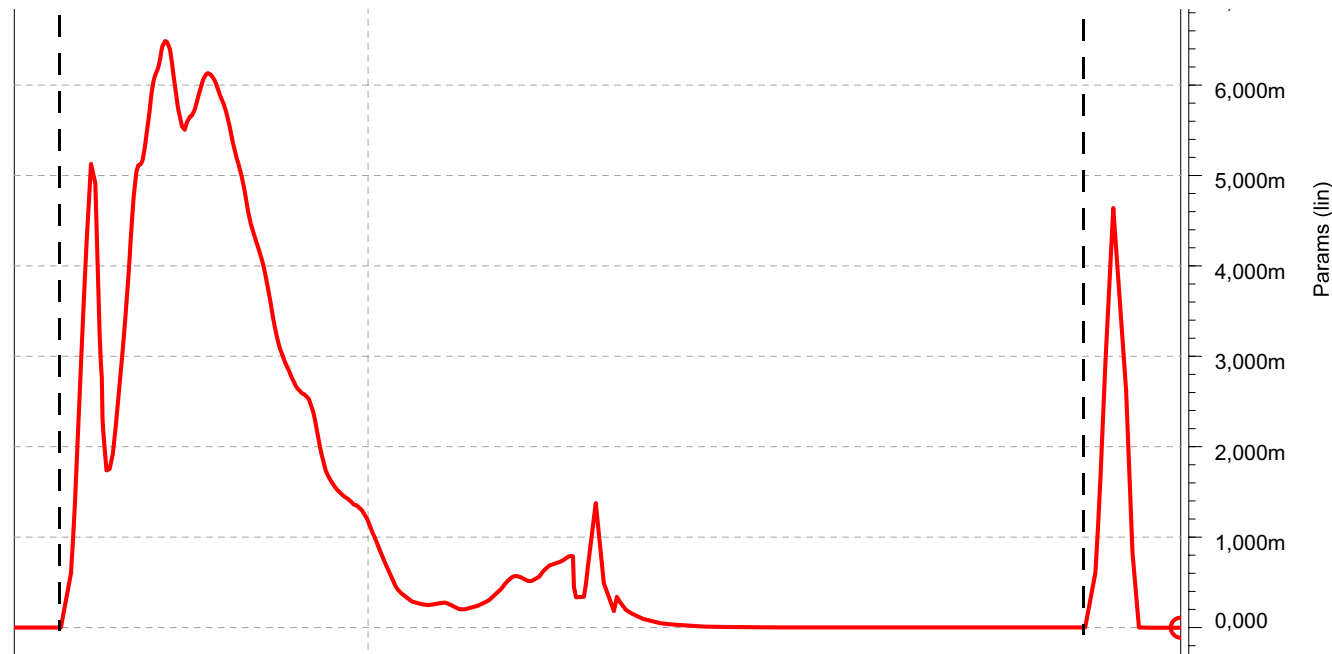
(S&S4 5.8/S&S5: 4.10.4, 10.1.2, 10.2.4)

STRÖM DRAGEN FRÅN V_{DD} UNDER 10 KLOCKPERIODER

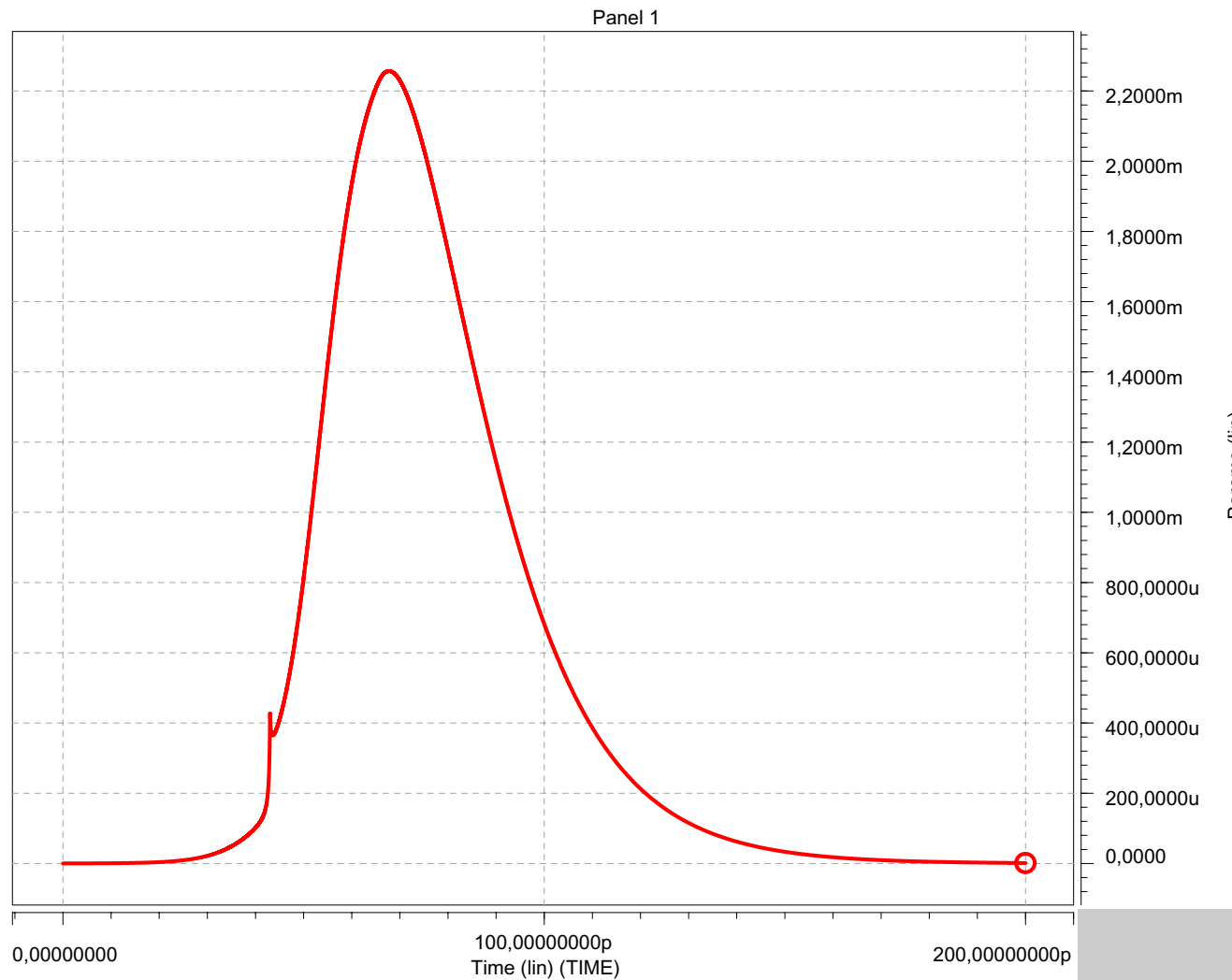


NÄRBILD PÅ EN DRYG KLOCKPERIOD

- ◆ I en krets med många grindar och stort logikdjup, varierar strömmen från cykel till cykel. Skälet är att olika s.k. indatavektorer, de logiska signaler som läggs på ingångarna, skapar logiska omslag på olika platser, i olika grindar.

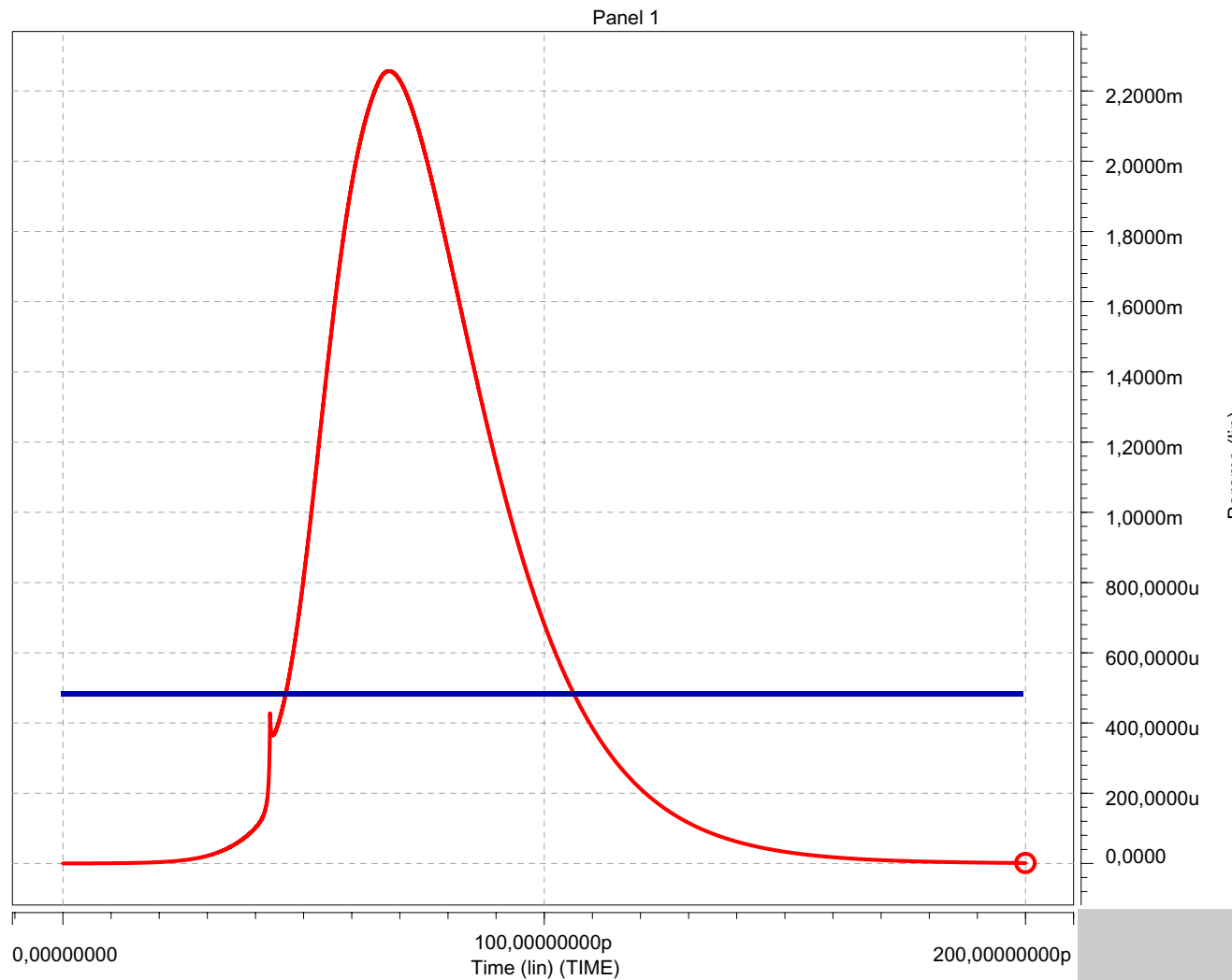


UPPLADDNING AV EN ENDA INVERTERARES UTGÅNG



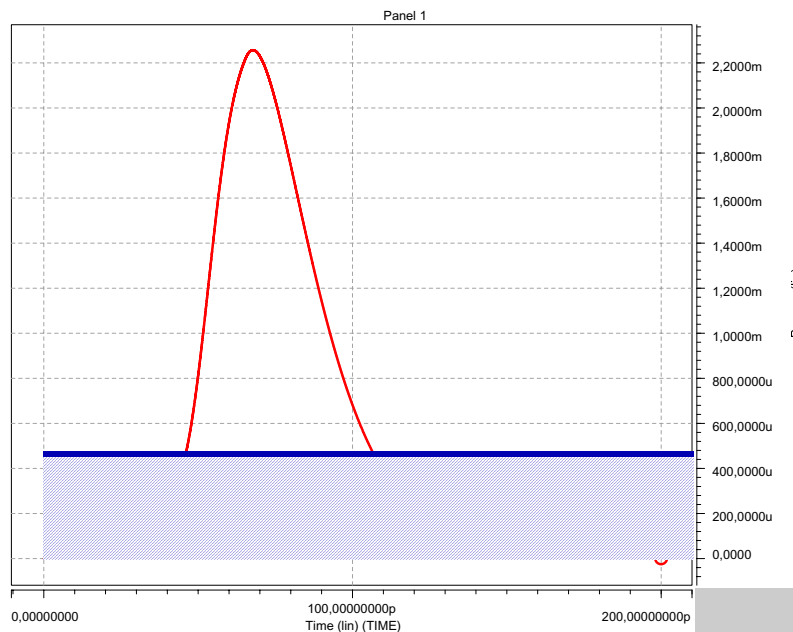
$$Q = \int I dt$$

MEDELSTRÖM UNDER TIDSINTERVALLET



$$I_{\text{medel}} = \frac{\int I dt}{T}$$

MEDELEFFEKT OCH ENERGI UNDER TIDSINTERVALLET



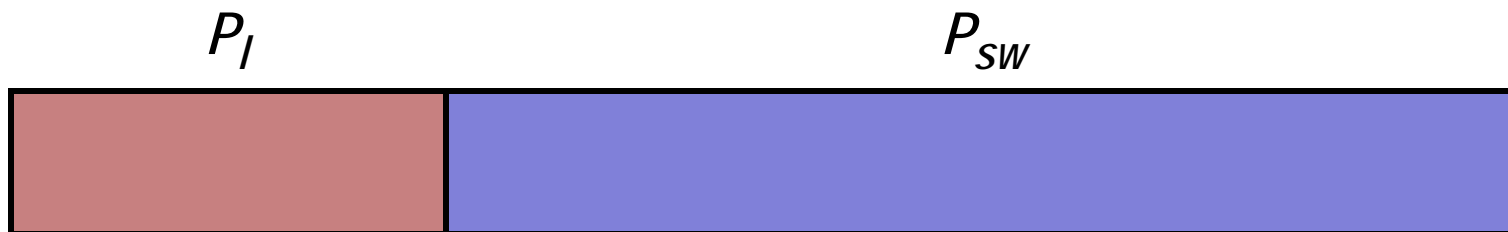
1. Laddningar motsvarande $Q = I_{\text{medel}} \cdot T$ har åkt från V_{DD} till utnoden.
2. Den elektriska energin $E = Q \cdot V_{DD}$ har dragits från V_{DD} .
3. Den totala effekten $P = \frac{Q \cdot V_{DD}}{T} = I_{\text{medel}} \cdot V_{DD}$ har utvecklats (som värme, i PMOS och NMOS:arna) när utsignalen senare har slagit om till en 0:a och Q har nått jorden!

NÅGRA REFLEKTIONER KRING EFFEKT OCH ENERGI

- ◆ Elektrisk energi säger något om hur många laddningar som förbrukats under en viss tid och syftar indirekt till en medeleffekt över tiden.
- ◆ Man använder ofta energibegreppet när man jämför inbyggda system; en applikation (kanske att packa upp en MP3-låt?) man kör på en inbyggd processor kommer kräva en viss energi från det att den startar till det att den är klar. Ju mer energi applikationen kräver, desto kortare batterilivslängd.
- ◆ Momentan effekt $p = i \cdot V_{DD}$, å andra sidan, ger en vink om hur mycket ström som dras i ett givet ögonblick. Denna information är viktig för hur man konstruerar kretsarna. Till exempel: Hur breda ledare krävs för att forsla ström? Hur ska IC-kapseln dimensioneras?

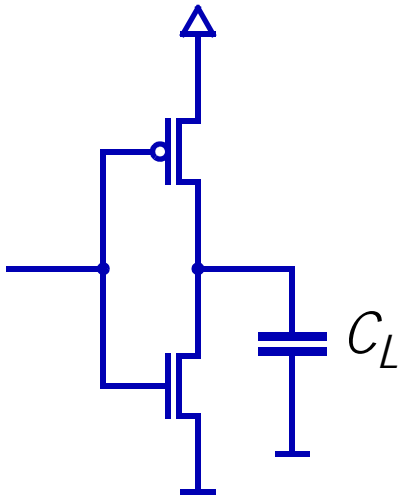
FLERA SKÄL TILL EFFEKT/ENERGI

- ◆ Statisk effektförbrukning
 - Läckageeffekt P_I främst p.g.a. subtröskelströmmar ($V_{DS} = V_{DD}$ och $V_{GS} = 0$)
- ◆ Dynamisk effektförbrukning
 - Omslagseffekt (switching) P_{SW}



Storleksförhållanden mellan olika "effektslag"

OMSLAGSEFFEKT (SWITCHING POWER)



- ◆ Det krävs energin

$$E = Q \cdot V_{DD} = (C_L \cdot \Delta V) \cdot V_{DD}$$

från matningsspänningen V_{DD} för att ladda upp utgångsnoden, som initialt är tom på laddning, till ΔV . Denna energi omvandlas till värme i transistorerna under upp- och urladdning.

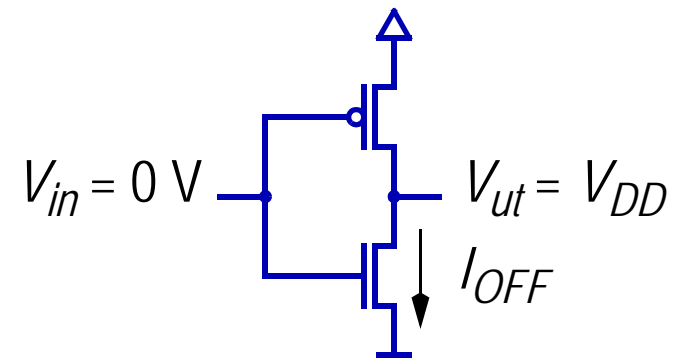
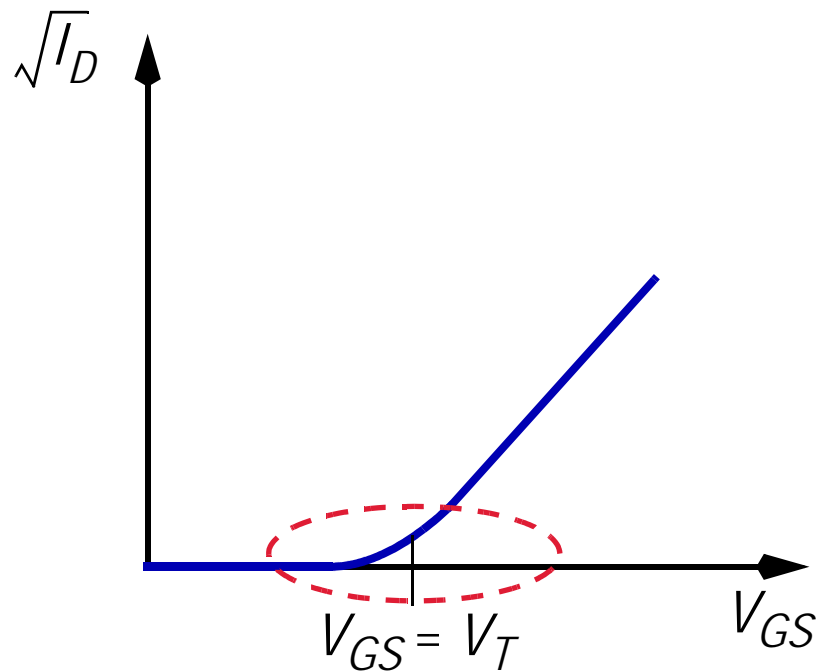
- ◆ Energi är effekten utvecklad under en viss tid, så effekten skrivs som

$$P_{SW} = \frac{1}{T} \cdot V_{DD}^2 \cdot \sum_i A_i \cdot C_i = f \cdot V_{DD}^2 \cdot \sum_i A_i \cdot C_i$$

där A_i = sannolikheten att noden i slår $0 \rightarrow 1$ (eller $1 \rightarrow 0$).

MOS-TRANSISTORNS SUBTRÖSKELLÄCKAGE

◆ $I_{OFF} \propto e^{\frac{q(V_{GS} - V_T)}{kT}} \left(1 - e^{-\frac{qV_{DS}}{kT}} \right)$ när V_{GS} är kring V_T

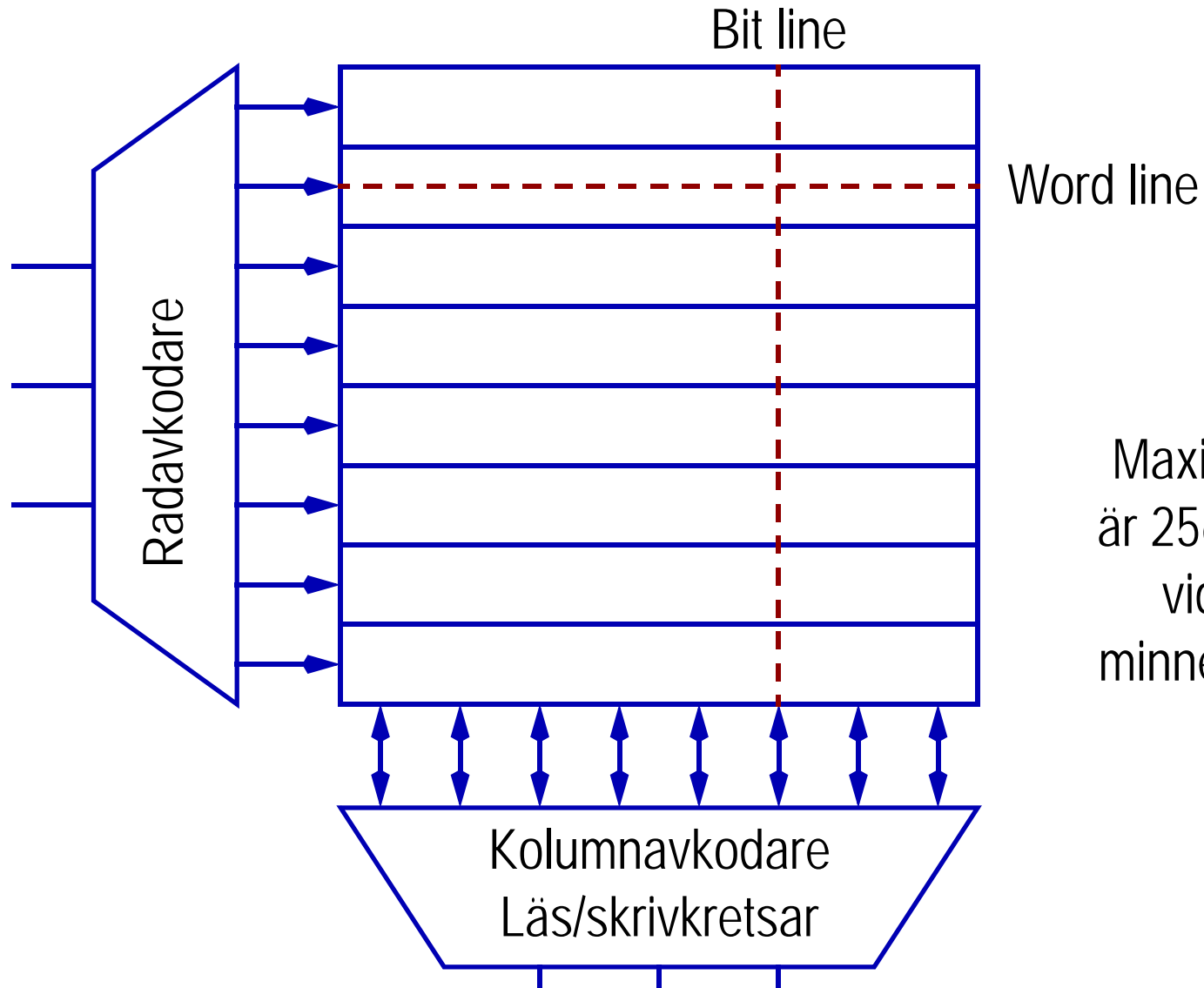


För detta fall är det NMOS:en som läcker

RAM-minnet

(S&S4: 13.9-13.11 / S&S5: 11.3-11.5)

RAM:ETS ARKITEKTUR



Maximalt antal bitar
är 256 kbits. Därefter
vidtar man s.k.
minnespartitionering.

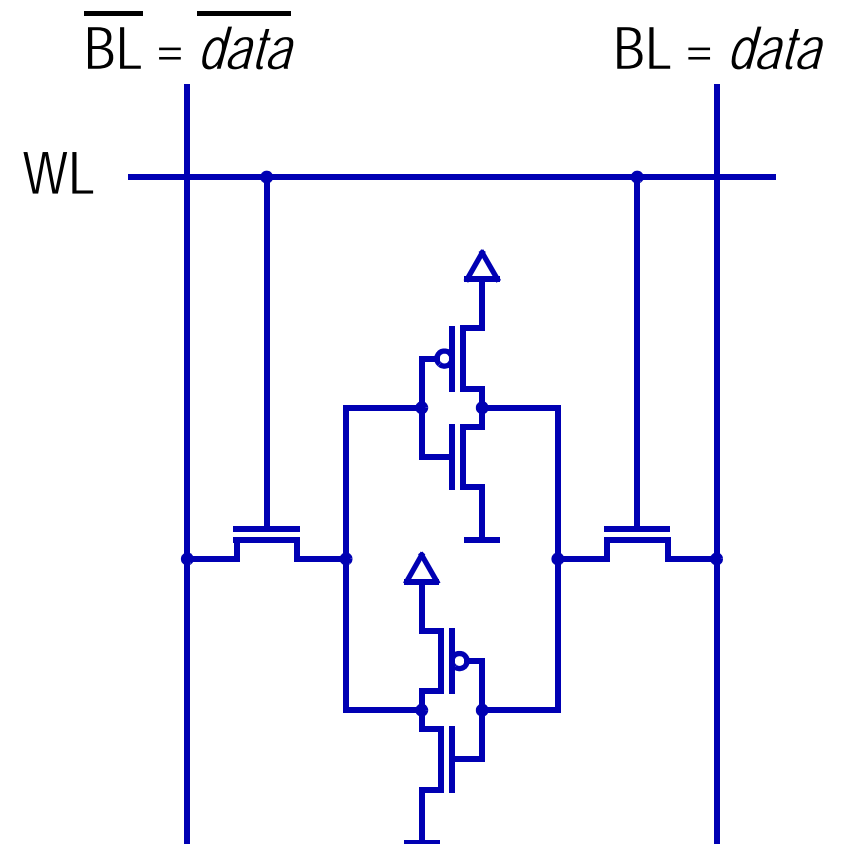
RAM:ETS DELAR

- ◆ Minnesceller ...
 - måste hållas extremt små.
- ◆ Läsförstärkare för utläsning och drivare för skrivning ...
 - måste ha extremt bra prestanda, för dessa måste kompensera för alla tricks man gör i minnescellerna. Dessa block ska t.ex. återvinna stabila signaler, från mycket svaga.
- ◆ Rad- och kolumnavkodare ...
 - konstrueras mer rättframt.

EN STANDARD 6-TRANSISTORS SRAM CELL 1(2)

Skrivning till cell:

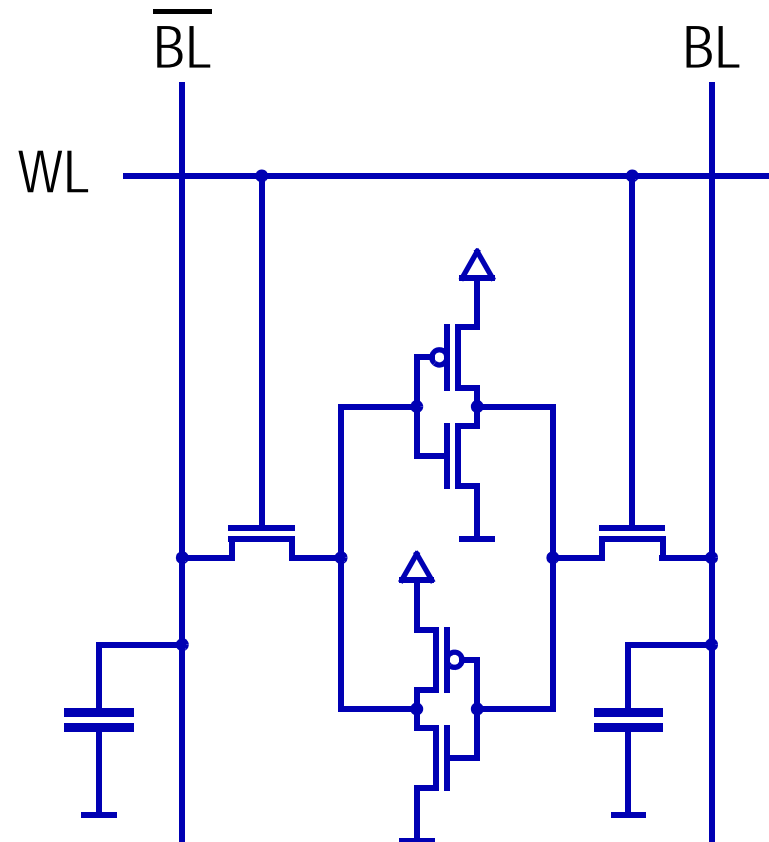
1. Sätt "bit-linorna" till $data$ respektive dess invers (\overline{data}).
2. Ge adressordet (öppna passtransistorer).
3. Nu tvingas inverterarna att "lagra" det nya datat.



EN STANDARD 6-TRANSISTORS SRAM CELL 2(2)

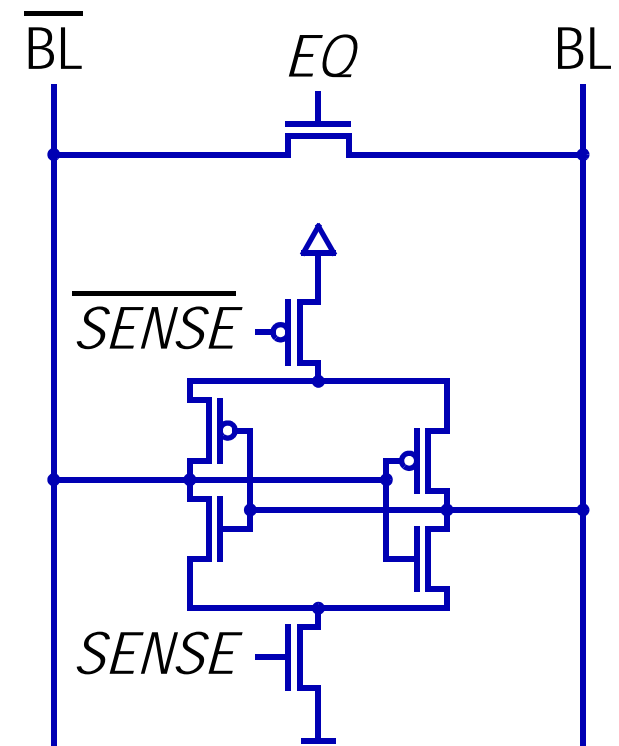
Läsning av cell:

1. Förladda bit-linorna till V_{DD} alternativt $V_{DD}/2$ och lämna dem i ett s.k. dynamiskt tillstånd.
2. Ge adressordet (öppna passtransistorer).
3. Inverterarna kommer nu att ändra spänningen på bit-linorna enligt lagrat data.



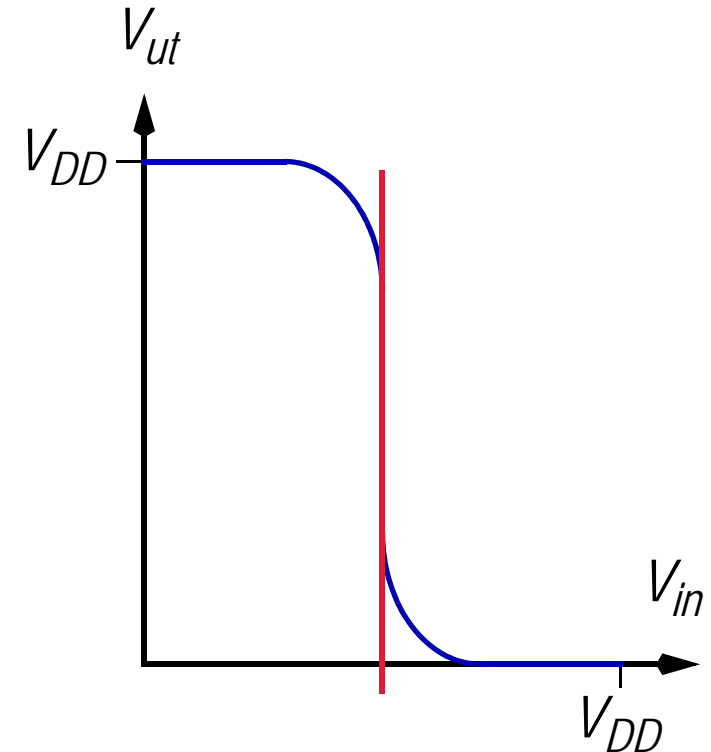
LÄSFÖRSTÄRKARE

- ◆ Läsförstärkaren gör så att man kan ha ett lågt ΔV på bit-linorna, vilket ger möjlighet att bygga små RAM. Den leder också till lägre effektförbrukning och högre hastighet, då den hjälper till att utföra omslaget på bit-linorna.
1. Utjämna BL:s (om förladdning till $V_{DD}/2$ används så står inverterarna mitt på sina VTC:er, vilket ger maximal förstärkning). EQ är alltså satt till 1.
 2. Ge lässignal, d.v.s. $SENSE = 1$.
 3. Minsta lilla skillnad mellan BL och \overline{BL} upptäcks av läsförstärkaren som slår om en lina till 0.



LÄSFÖRSTÄRKARENS ARBETSPUNKT

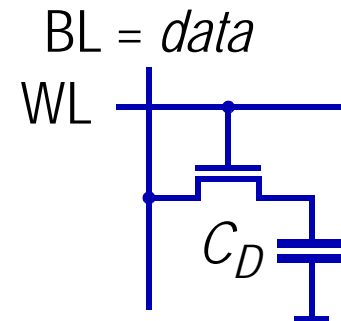
- ◆ Används förladdning till $V_{DD}/2$ så ställer man alltså förstärkaren i ett läge så båda inverterarna befinner sig på sina V_{TRIP} (då V_{in} ungefär är lika med $V_{DD}/2$).
- ◆ Idealt sett är förstärkningen oändligt hög i mitten av VTC:n.
- ◆ I verkligheten är förstärkningen inte oändlig, men ändå mycket stor och minsta lilla ändring av V_{in} kring V_{TRIP} får stora följdverkningar på V_{ut} .



EN KLASSISK 1-TRANSISTORS DRAM CELL

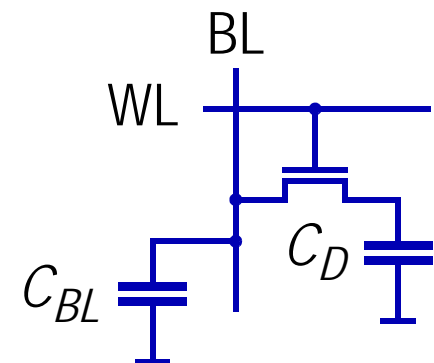
Skrivning till cell:

1. Sätt BL till *data*.
2. Ge adress på WL.
3. Nu lagras det nya datat dynamiskt på C_D .



Läsning av cell:

1. Förladda BL till någon spänning.
2. Ge adress på WL.
3. Spänningen på BL kommer förändras då laddningen i C_D kommer distribueras ut på både C_D och C_{BL} .



En läsförstärkare används för att avkänna förändringen!

LITE MER OM DRAM-CELLER

- ◆ Utläsning av DRAM-cell är destruktiv, eftersom laddningen på C_D förändras. Utläsningen måste alltså förstärka upp bit-linorna och återföra detta värde till DRAM-cellerna — detta kallas refresh.
- ◆ DRAM:et bygger alltså på s.k. laddningsdelning mellan C_D och C_{BL} . Sålunda måste C_D ha en ganska stor kapacitans, för att C_D 's laddning ska kunna inverka på C_{BL} 's spänning. Kapacitansen kan vara på några tiotals fF, och detta ska ordnas på en ytterst begränsad yta (tricks krävs!).