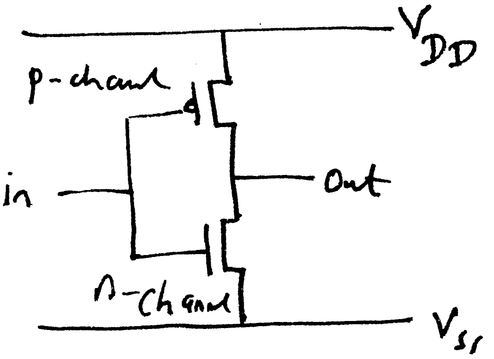
2004

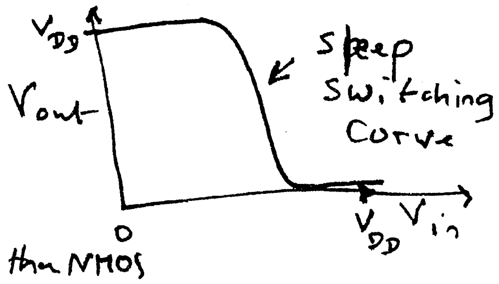
4B7 Q1

DHm 1.2.2004

a) complementary MOS technology has the decisive advantage over NMOS that either the p channel or the n-channel transistor is off at any time and the current drawn is small (except at the switch transient)

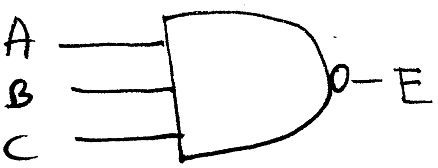
In addition to LOW POWER the technology also has a wide margin of device operation. Hence VLSI can be successful.





The fabrication technology is more complex than NMOS but now it is well established and dominant.

b)



three input NAND $E = \overline{A.B.C}$

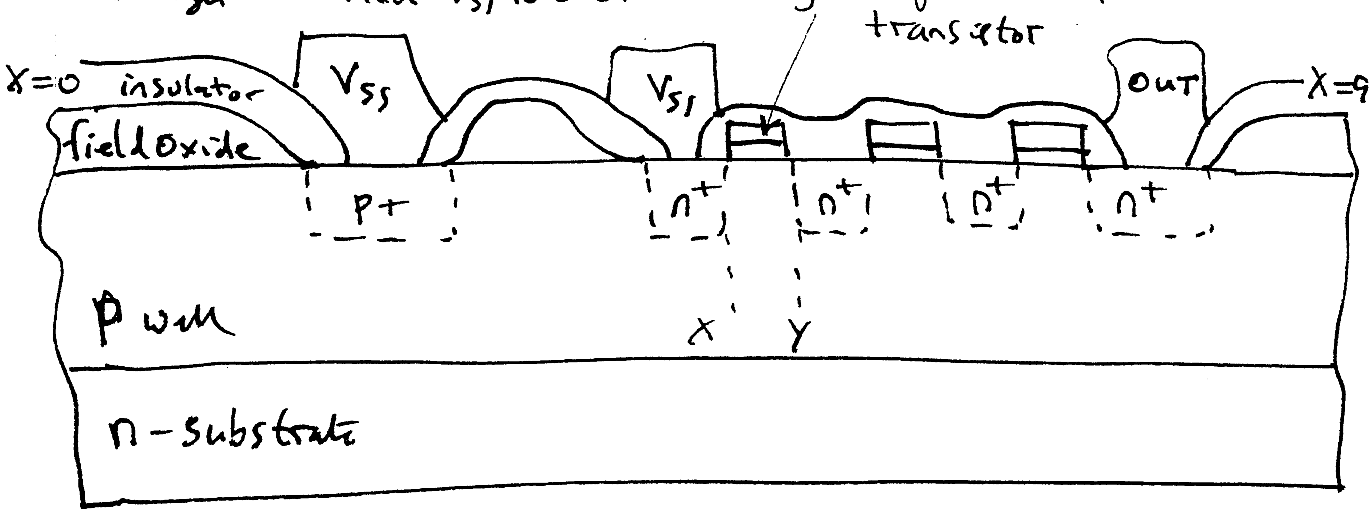
upper right p-type for PMOS

lower right n-type for NMOS

upper left n-type for ohmic to substrate

lower left p-type for ohmic to p-well

All three n channel gates have to be high to connect V_{SS} to out.



In self aligned technology the polysilicon gate is in place when the n^+ source and drain implants are performed with a result that the undoped channel is only the region under the gate i.e. self aligned. The lithographic step for the gate therefore defines the distance XY which determines the switching speed from the carrier transit time

2004 4B7 Q1 (continued)

The electrical width is determined by the active area. DPm 1·2·2004

In this case $\frac{\text{width p-channel}}{\text{width n-channel}} = \frac{2}{3}$

Worst case rise time is with 1 p-channel device conducting.

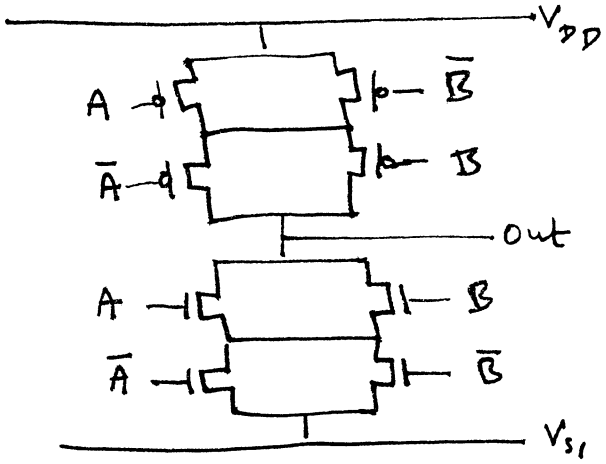
" " fall time is the series connection of 3 n-channel devices.

But $\frac{\text{electron mobility}}{\text{hole mobility}} = 2$ Hence $\frac{\text{rise time}}{\text{fall time}} = \frac{3}{2} \times \frac{1}{3} \times 2$

$= 1$ $\boxed{\text{Equal}}$

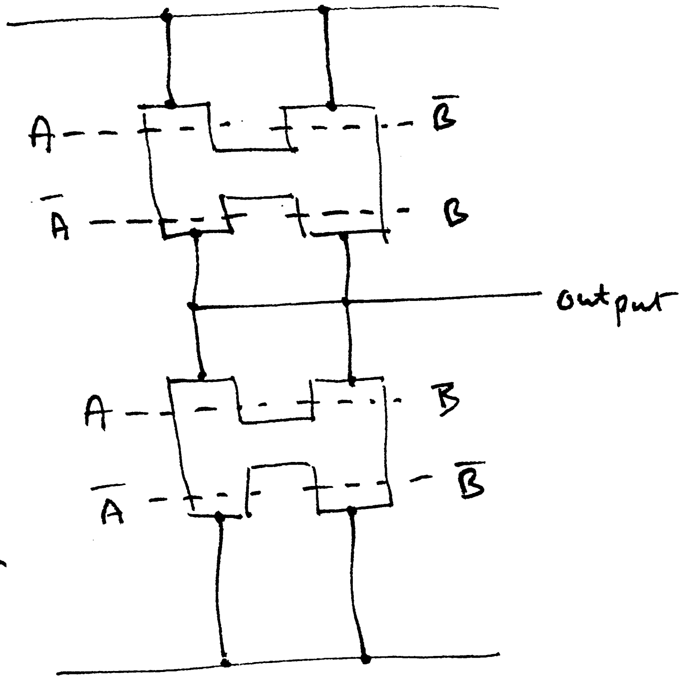
d) Truth table

B \ A	0	1
0	1	0
1	0	1



confirm that output is only low when both inputs are different

confirm that no input combination gives a conduction path right through



The layout can be simply two active regions. (shown as outline)

The polysilicon gates are shown dotted.

Worst case output rise time is through 2 p devices in series.

Worst case fall time is through 2 n devices in series.

ie. equalize the worst case times by choosing transistor widths in inverse proportion to the electron and hole carrier mobilities.

2004 4B7 Q1 (continued) Dfm 1.2.2004

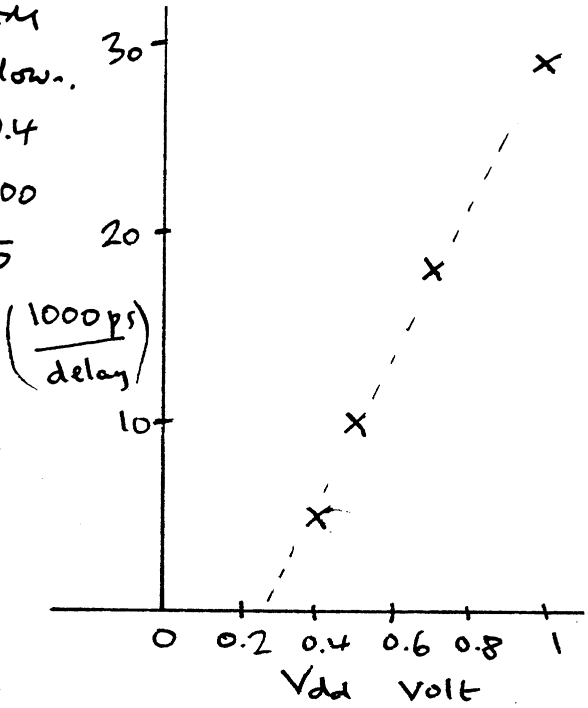
An inverting device is required in the ring and the 11 member example has a period corresponding to $2 \times 11 \times \tau = 22\tau$ where τ is the single gate delay. As seen with the 0.1.010 sequence an odd numbered ring is unstable giving oscillators which are observed through a buffer device. Practical devices are slower than unloaded ring oscillators.

c) The power/stage goes from 30 μW at 2V to 0.2 μW at 0.4V.

As the power supply is reduced towards threshold the ring frequency goes linearly down.

Vdd	1V	0.7	0.5	0.4
Delay (ps)	34ps	55	100	200
1/Delay (1000/ps)	29	18	10	5

Extrapolated threshold $\boxed{0.25V}$



d) Let C be input capacitance of n+p gates
n number of gates in the ring
V power supply V_{DD}
P the power per gate in the ring

The energy dissipated when a gate is either charged or discharged is $\sim CV^2$

the frequency at which the voltage is either switched up or down is $\frac{1}{n\tau}$

Hence $P = \frac{CV^2}{n\tau}$ modelling as a simple capacitor

$$\therefore C = \frac{P n \tau}{V^2} = \frac{1 \times 10^{-6} \times 101 \times 55 \times 10^{-12}}{0.7 \times 0.7} = 11 \times 10^{-15} \text{ F}$$

But total area of p+n channel transistor gates is $(3 \times 10^{-6} + 4 \times 10^{-6}) \times 100 \times 10^{-9}$
$= 7 \times 10^{-13} \text{ m}^2$

Simple capacitor $C = \frac{A \varepsilon_0 \varepsilon_r}{d}$ take $\varepsilon_r = 4$ for SiO_2
$\varepsilon_0 = 8.9 \times 10^{-12}$

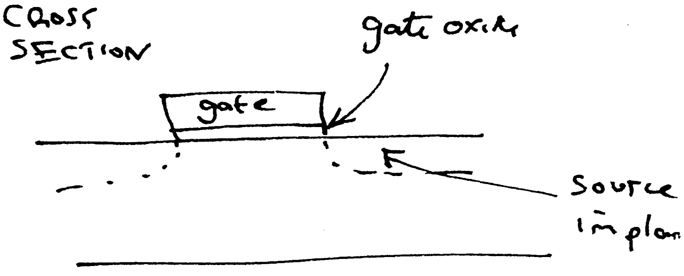
$$\therefore \text{oxide thickness } d = \frac{A \varepsilon_0 \varepsilon_r}{C} = \frac{7 \times 10^{-13} \times 8.9 \times 10^{-12} \times 4}{11 \times 10^{-15}} = 2.3 \times 10^{-9},$$

Assuming no stray capacitance and ignoring the interconnect line capacitance the gate oxide thickness is $\boxed{2.3 \text{ nm}}$ which is physically reasonable.

4B7 ^{2004} Q2

DBm 1.2.2004

a) Scaling CMOS devices involves reducing all the device dimensions the most important of which is the source drain distance.



This brings the benefits of reduced device switching speed and increased circuit density allowing more devices on a chip and fewer interconnections.

The figure of power supply voltage vs. FET channel length shows how V_{DD} is of order 1 volt for 0.05 μm devices. There has to be a corresponding scaling down of the threshold voltage to ensure successful device operation.

(i) Gate oxide thickness is reduced to main device operating currents.

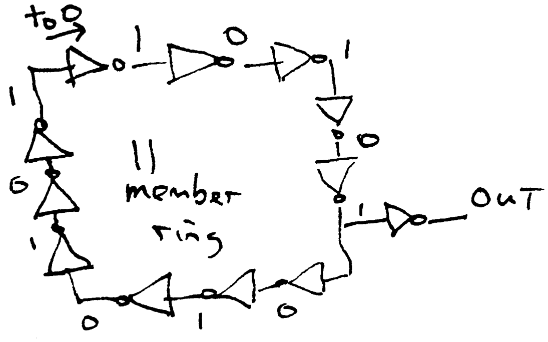
In memory devices the leakage current per device is important for determining system power in the off state. Here the most aggressively small ~~gate~~ source-drain distance is not necessarily used.

(ii) For high performance devices in microprocessors short switching speed is more important than low leakage current, and the shortest feasible source-drain distances are used.

One physical limit is tunnelling through the gate insulator which can lead to an unacceptable mixing of the controlling fluid (electrons) with the controlled fluid (channel charge). Using higher dielectric materials than SiO_2 can alleviate this problem to enable high specific capacitance and low tunnel currents at the same time.

In the 5 year time scale CMOS technology is going to be governed by the cost and practicality of deep UV photolithography rather than physical device limits and line width will be reduced towards 50 nm.

b) Ring oscillator circuits have a lower frequency output than single gates and measurement via a low frequency output pad to a frequency counter is possible



2004 4B7 Qn3 (i) (Rather more detailed than expected from candidates)

The gain stage in a typical OA is a differential amplifier with high voltage gain but low power capability. An output stage is required to provide:-

- high current gain (limited voltage gain acceptable)
- distortion-free amplification
- efficiency
- protection from loading effects including short circuit

In CMOS there are various ways to accomplish this, eg.

- source follower (< unity voltage gain)
- class A amplifier (various forms including class B "push-pull")

In this (class A, single-ended) design, power consumption is to be minimised by sizing the devices to deliver the smallest standing currents necessary to deliver the required slew rate

Slew rate is determined from available o/p current:

$$i = C_L \, dV_{out}/dt$$

The current needed to charge/discharge C_L at the given rate:-

$$i_s = \pm 40 \times 10^{-12} \times 20 \times 10^{6} \text{ A} = 0.8 \text{ mA}$$

M2 is provided with fixed gate bias $V_{GG} = 0$v. Hence V_{GS_2} is -3v. We assume all the current available from M2, i_{d_2} can be available to charge C_L, ie, that M1 draws negligible current while this happens. We also assume that $V_{out} = 0$ (the middle of the required range, ± 1v). Hence $V_{DS} = -3$v.

M2 is always in saturation mode provided $V_{out} < +1$v. Rearranging the S-H eqn:

2004 4B7 Qn 3 (2)

$$W_2 = L_2 . \frac{i_{D2}}{\frac{1}{2}(V_{GS}-V_T)^2(1+\lambda V_{DS})\,\mu\epsilon/t_{ox}}$$

$$\frac{W_2}{L_2} = \frac{0.8\times10^{-3}}{\frac{1}{2}(-3+1)^2(1+0.02\times(-3)\times10^{-5}} = \frac{160}{4\times .94} = 42.5$$

Under quescent conditions, M1 draws the same current; however, since M2 delivers 0.8mA over the whole of the range while in saturation, M1 must be sized for 1.6mA to allow C_L to be discharged at the required rate. When Vout = 0, $V_{DS_1} = 3v$. M1 will stay in saturation provided $V_{GS_1} < V_{DS_1} + V_T$, or +4v. Considering the other key points:

not required in answer

Vout (V)	V_{DS_1} (V)	Limit V_{GS_1} for sat (V)	Vin (V)
+1	4	5	+2
0	3	4	+1
-1	2	3	+0

In the worst case, if $V_{GS_1} = 3v$ while Vout is 0v, the size for M1 should be chosen as:-

$$\frac{W_1}{L_1} = \frac{1.6\times10^{-3}}{\frac{1}{2}(3-1)^2(1+(0.01)\times3)\times1.8\times10^{-5}} = \frac{320}{4\times1.03\times1.8} = 172.6$$

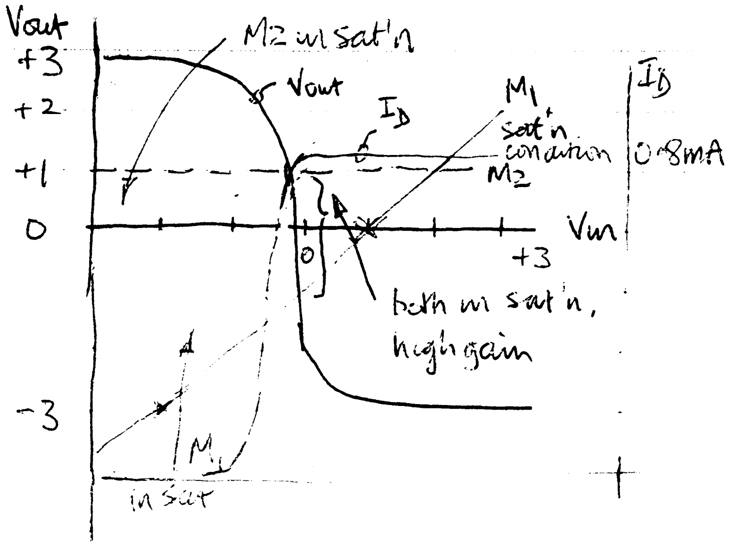
not req. in ans

Other assumptions, eg. $V_{GS_1} = 4, 5$ can be justified
Note that if V_{GS} is +5v, $i_{d_1} = (5-1)^2/(3-1)^2 = 4 \times$ as great

NB DC output voltage is highly dependent on precise control of V_{GS_1}
The biasing arrangements of M1 are not shown but are critical

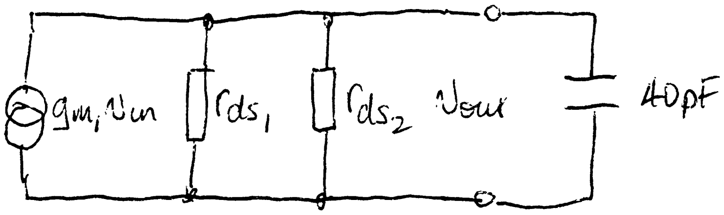
A real cct needs a self-bias arrangement to define quescent values for Vin and Vout. Current negative feedback can achieve this

2004 Qn 3(3)



Note that the req. range of outputs ±1V can be accommodated while both devices stay in saturation - just!

To calculate ~~sf~~ gain, note the ssec for the o/p



$g_{ds} \doteq I_D \lambda$

$g_m = \sqrt{2 \frac{\mu \epsilon}{t_{ox}} \frac{W}{L} I_D}$

We will assume I_D = 0.8mA quiescent

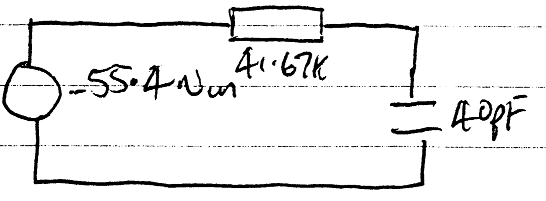
$$\text{SS gain} = -\frac{g_{m_1}}{g_{ds_1} + g_{ds_2}} = -\sqrt{2\times10^{-5}\times172.6\times0.8\times10^{-3}} \times \frac{1}{0.01+0.02}$$

$$= -\sqrt{2\overset{10^{-2}\times}{\wedge}172.6\times0.8} \times \frac{1}{0.03} = -55.4$$

$$r_{out} = r_{ds_1} \| r_{ds_2} = \frac{1}{0.8\times10^{-3}} \times \frac{1}{0.01+0.02}$$

$$= 41.67\,k\Omega$$

Considering the Thevenin eqvt of the output



For -3dB ~~pt~~, Real = Imag

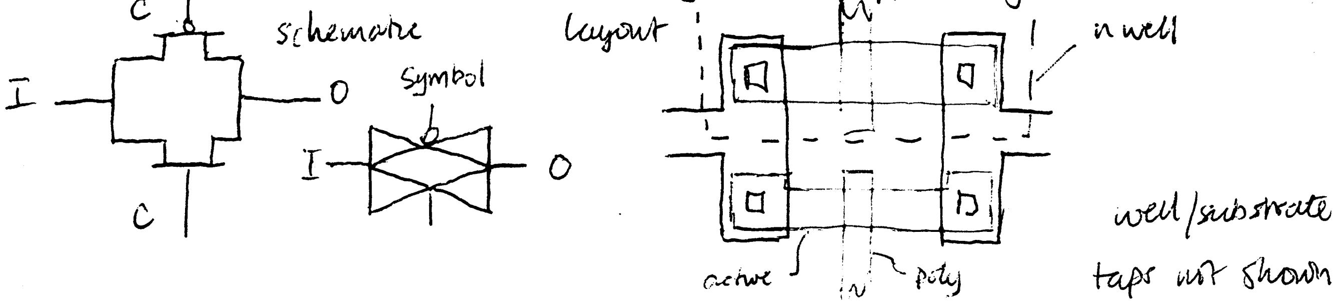
$j\omega_{-3} CR = 1$

$\omega_{-3} = 1/(40\times10^{-12}\times41.7\times10^{3}) \rightarrow 95\,\text{kHz}$

This can be improved by reducing rout - can be achieved by application of -ve current feedback, which requires a resistor between output & input. This also has benefit of stabilising the bias point, but the voltage gain is reduced.

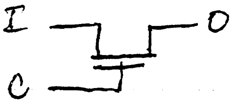
2004 4B7 Qu 4 (1) (More detailed than expected from candidates)

A transmission gate in CMOS comprises a pair of complementary transistors with source & drain regions strapped together



The gates are driven with complementary signals C, $\bar{C}$. When C=1, $\bar{C}$=0, both the p & n device conduct. In the opposite state both are non-conductive. NB the device is bilateral when seen from I or O – it can conduct in either direction.

To understand the performance issue, consider a single n-channel pass-transistor used as a switch. The n-type device conducts when



C is high, non-conductive when C is low. To allow a channel to form V_{GS} must be greater than V_T. If C is set to logic 1 = V_{DD}, then if I is also driven to V_{DD}, O cannot rise above $V_{DD} - V_T$, typically a drop of ~1 volt. Thus O would be a 'weak' version of the high at I. Note that a logic low is transferred reliably to O without o/set. As a result of the logic 'drop' it is impracticable to cascade single transistor switches. A similar state of affairs is found for the p-type device, which can transfer a logic 1 without o/set, but a logic 0 is transferred only weakly, with 'rise' of $|V_T|$.

By combining the two devices in parallel a switch can be made which has neither problem.

TT for T/gate

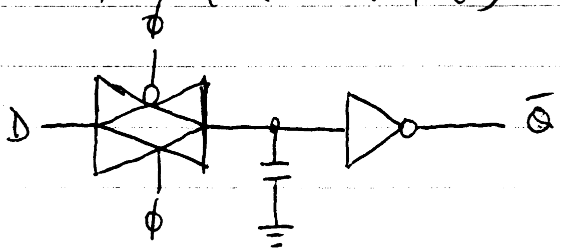
I	C	O
0	0	Z
1	0	Z
0	1	0
1	1	1

Z means high impedance

Hence a near perfect digital switch can be achieved with only two devices.

In digital cct a t-gate may be used to realise a multiplexer. They are commonly used to control feedback paths & signal paths in sequential gates e.g. dynamic D-type FF

2004 - 4B7 Qn 4 (2)



~~Ad~~ Advantages:

- low device count for MUX / FF etc
- bilateral characteristic
- high performance, no losses
- can be cascaded

Disadvantages:

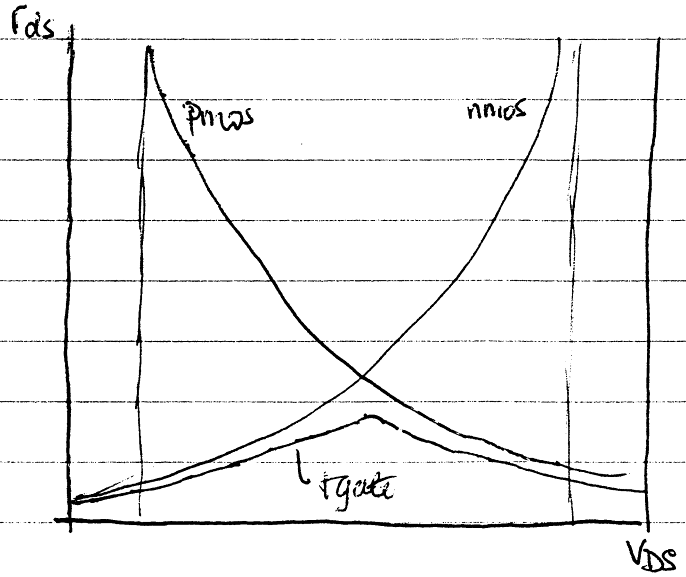
- A passive device - does not regenerate logic levels
- requires complementary control signals (extra logic)
- May be sensitive to clock dispersion or skew, charge sharing

In analogue ccts t-gates may be used in switches, multiplexers for linear signals, sample holds, DA converters.

Advantages

- Efficient switch with low offset voltage
- good frequency response
- Good ratio Roff / Ron
- Compact structure

Disadvantages



Notwithstanding above comments, the effective resistance of the paralleled devices in a tgate is significant and varies with applied voltage. In sensitive linear applications this can cause distortion

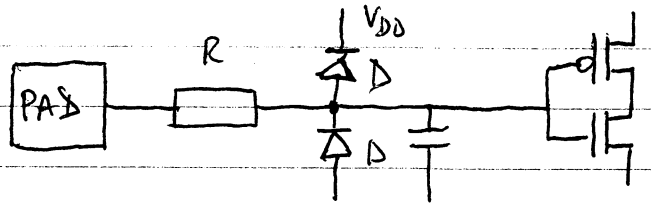
Effective channel resistance of p, n devices and t-gate

4B7 2004 Qn 4 (3)

Input pad structures are required to protect MOSFET inputs from:

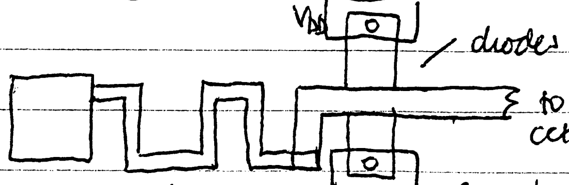
- over & undervoltages
- consequential latchup conditions
- electrostatic discharge

Gate oxide thicknesses in modern processes are o(30nm) such with breakdown voltages around 10v. Input resistances may exceed $10^{12}\Omega$. Since the gate electrode typically has capacitance of a fraction of a pF, only a small packet of charge is ~~to~~ required to generate voltages far in excess of $V_{breakdown}$.
The human being is often modelled (for evaluation of static 'risk') as a capacitance ~ 100pF charged to ~ 1.5kV in series with a resistance of a few kΩ. The energy available is sufficient to vaporise a considerable volume of Si.
Protection can be achieved with the cct below:



R may be implemented as a strip of poly Si 20–100 sq. long

Diodes become fwd biased as Vpad exceeds $V_{DD}+0.7$, $V_{SS}-0.7$ sinking excess current to the rails



Diodes ~~D~~ are provided with guard rings & generous well/substrate taps to minimise carrier injection

C is parasitic capacitance due to D & R

Selection of values is necessarily a compromise. Excessive RC will give good protection, but will delay legitimate digital edges and cause slower rise/fall

Often punch-thru devices are used in place of the Diodes (very short, closely spaced S&D, no gate, which avalanche at ~ 10v)

2004 4B7 Qn 5 (i) (More detailed answer than exp. from candidates)

The threshold voltage V_T of a MOSFET is that potential which must be applied between gate and source in order to bring about strong inversion within the channel. There are three main components to this potential:

- ϕ_{GC}, the difference in work functions between the gate material and the Si substrate the channel side
- a negative potential arising from the existence of undesired positive charge within the gate oxide and at the oxide/substrate interface. - referred to as Q_{ox}, and assumed to reside entirely at the interface
- a voltage $-2\phi_F - Q_B/C_{ox}$ needed
  (a) to bring the surface potential to the strong inversion conditio
  (b) to offset the induced depletion layer charge, Q_B
  ie, to 'unbend' the energy bands that result when the MOS system is first brought together, and to bring the surface potential ϕ_s to be equal to ϕ_F

In essence, the originally p-type S/C becomes n-type with this gate potential applied. Further increases in V_{GS} produce only slight change in surface potential ϕ_s

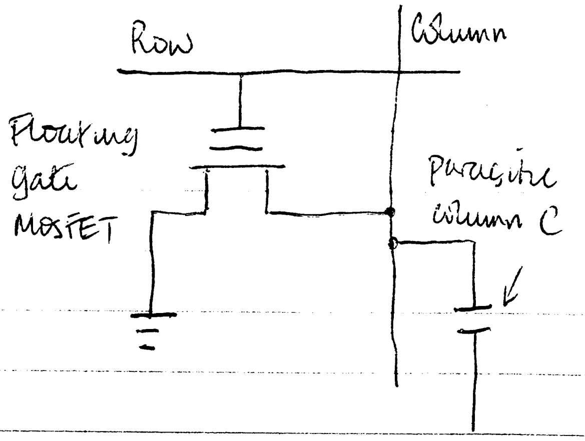
Hence the main factors determining V_T are:

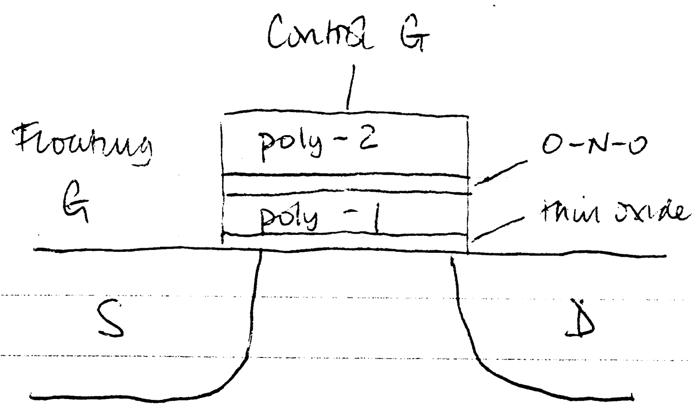
- Materials used for the gate electrode (Al or polySi) determining its work function
- properties of the dielectric used for the gate insulator, fixing the capacitance C_{ox}; its thickness t_{ox}
- channel dopant density
- impurities, defects, dangling bonds etc at Si - SiO2 interface
- potential between source and substrate - which acts as a second, or "back"-gate
- temperature

2004 4B7 Qn 5 (2)

The 'flash' memory has a very simple structure, akin to that of the one-transistor DRAM cell, except that no storage capacitance is required. This leads to a compact layout. The design stores data through use of an unusual MOS structure which allows the threshold voltage of the FET to be modified electrically, in a non-volatile manner.

In the transistor structure an extra floating gate is interposed between the gate & channel.



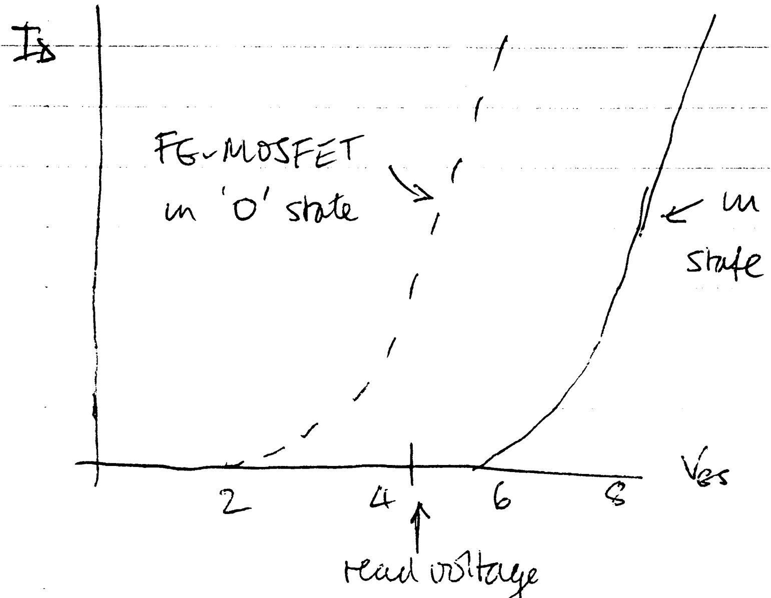


The dielectric that separates the upper (control) gate from the floating gate is typically an oxide-nitride-oxide sandwich. Ie. the floating gate is electrically isolated, but is capacitively coupled to the control gate and to the substrate.

Writing data involves 'programming' the floating gate and hence adjusting the FET to have two or more different threshold voltages V_T relative to the control gate. If the floating gate contains a large electronic charge, the device has a higher V_T. This is regarded as the '1' state.

If the charge is removed from the floating gate, the MOSFET has a lower V_T - the device is in the '0' state

2004 4B7 Qn 5 (3)



Advantages :-
- Compact, high density
- Non-volatile
- No capacitor needed
- Less sensitive to charge sharing and noise

Disadvantages
- More complex process.
- Slower write operation
- Need for higher voltages

Writing.

Transferring charge into the FG by 'hot carrier' effects. A high field is applied to the drain and gate so the device is in saturation. The carriers in the pinch-off region are then 'hot' - highly energetic - and a proportion of them are scattered into the floating gate. This action can be enhanced by thinning the gate oxide in the vicinity of the drain. Once in the floating gate, electrons are trapped in a potential well and remain so indefinitely or until the cell is erased.

Erasure involves removing charge. This is achieved by inducing Fowler-Nordheim tunnelling between FG & source. The gate is grounded and the source taken to a high voltage. This allows electrons to tunnel through the oxide barrier from the FG to the Source.

Reading

This is done by observing the channel current with a suitably chosen gate voltage (read voltage) to discriminate the state. For the device above a suitable value would be c. 4.5V. A moderate voltage, say 3V is applied to the gate.
- If the device is in the '1' state negligible current flows
- If " '0' state, V_{GS} exceeds V_T and current flows

This concept can be extended to give devices with e.g. 4 or even 8 different V_Ts, allowing 2 or 3 bits to be stored in one cell