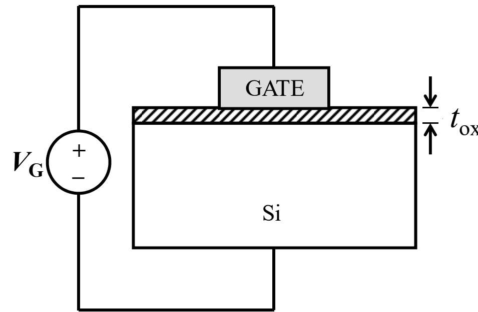
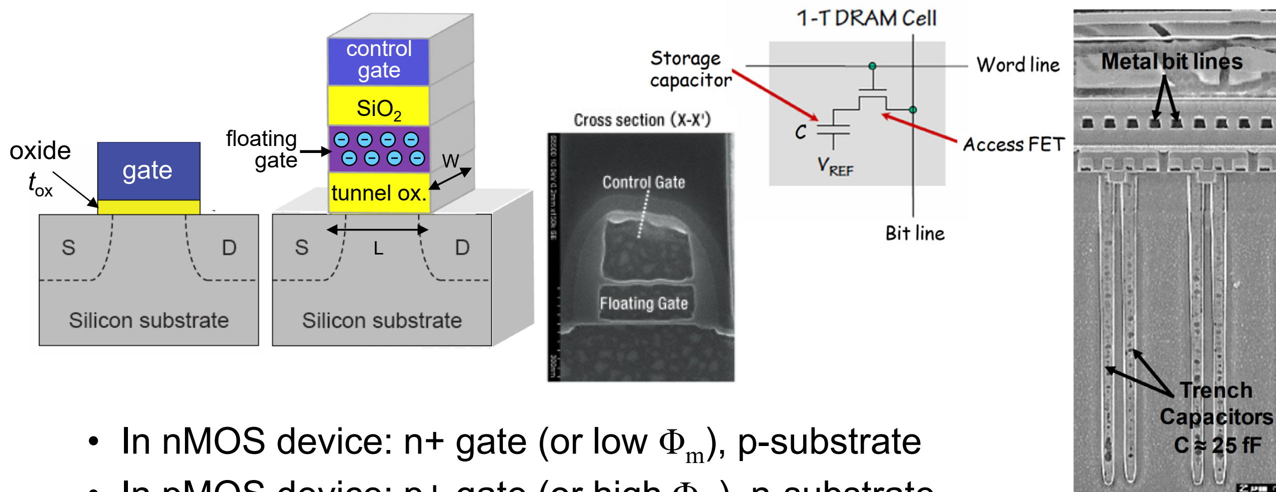
EE 116 Lecture 30

Metal-Oxide-Semiconductor (MOS) Capacitor



- Read: Ch. 6.1, 6.2, 6.3.1-6.3.2, 6.5.1, 6.5.4
- Skim: 6.3.3, 6.3.4.1, 6.5.5, 6.6.1
- https://truenano.com/PSD20/contents/toc6.htm

- MOS capacitor needed for MOSFETs, Flash, DRAM:



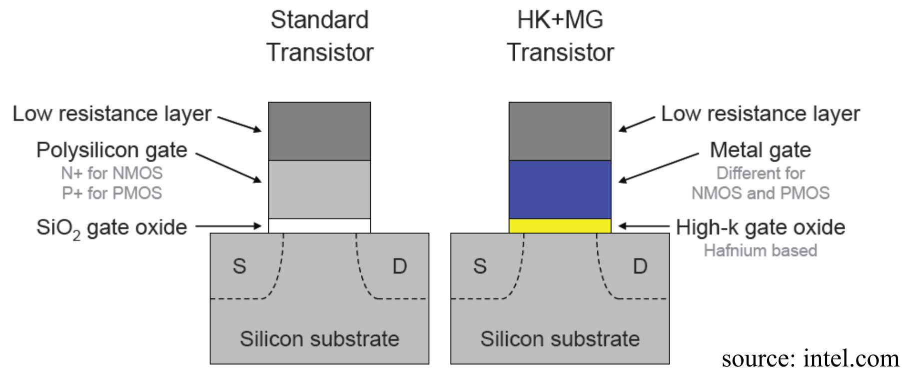
- In nMOS device: n+ gate (or low Φ_m), p-substrate
- In pMOS device: p+ gate (or high Φ_m), n-substrate

Note gate = metal by Intel at 45nm tech node, since ~2008. Why?

- SiO_2 most common gate insulator (E_G = 9 eV, ε_r = 3.9)

Intel switched to bilayer HfO_2 ($E_G \approx 5$ eV, $\varepsilon_r \approx 20$) with SiO_2. Why?

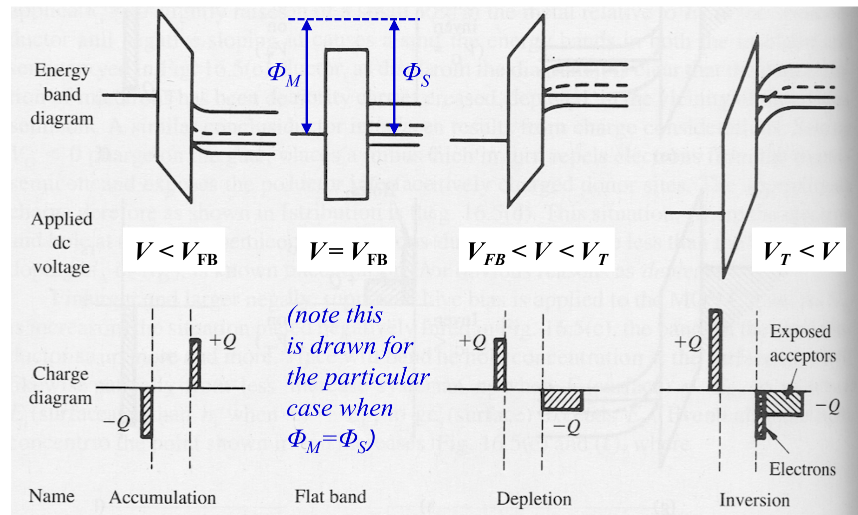
- Metal/high-K MOSFET (we'll come back to it later):



source: intel.com

- Draw band diagram of MOS capacitor with n+ gate and p-substrate.

- What happens if we apply a gate voltage V?



- There are two important reference voltages here:
  1) Flat-band voltage, V_{FB} = voltage needed on gate to get E-field = 0 everywhere (flat bands). Note, this can be zero ("ideal" MOS), but generally depends on gate Φ_M or doping, $V_{FB} = \Phi_M - \Phi_S$
  2) Threshold voltage, V_T = voltage needed on gate to get electron concentration at Si/SiO_2 surface same as that of (majority) holes in the bulk. Si surface is "inverted".

Example 6.1	Calculate the flatband voltage of a silicon nMOS capacitor with a substrate doping $N_a = 10^{17}$ cm^{-3} and an aluminum gate (Φ_M = 4.1 V). Assume there is no fixed charge in the oxide or at the oxide-silicon interface.
Solution	The flatband voltage equals the work function difference since there is no charge in the oxide or at the oxide-semiconductor interface. $$V_{FB} = \Phi_{MS} = \Phi_M - \chi - \frac{E_g}{2q} - V_t \ln\frac{N_a}{n_i}$$ $$= 4.1 - 4.05 - 0.56 - 0.026 \times \ln\frac{10^{17}}{10^{10}} = -0.93\text{ V}$$ The flatband voltages for nMOS and pMOS capacitors with an aluminum or a poly-silicon gate are listed in the table below.

	Aluminum	p$^+$ poly	n$^+$ poly
nMOS	-0.93 eV	0.14 eV	-0.98 eV
pMOS	-0.09 eV	0.98 eV	-0.14 eV

Note:

$$\Phi_{Sp,n} = \chi + \frac{E_G}{2} + (E_i - E_F)$$

$$\approx \chi + \frac{E_G}{2} \pm kT \ln\left(\frac{N_{A,D}}{n_i}\right)$$

Φ_M → look it up (typical range 3-5 eV)

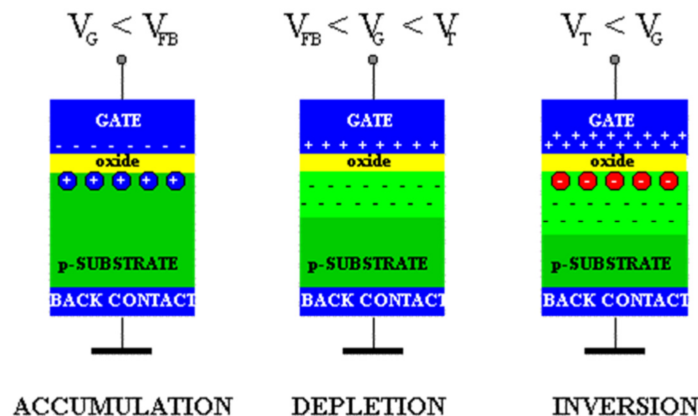
(minor typo online: eV vs. V)

Note: we will NOT derive expression for V_T in EE 116, but it is done in EE 216.

Example 6.2	Calculate the threshold voltage of a silicon nMOS capacitor with a substrate doping $N_a = 10^{17}$ cm^{-3}, a 20 nm thick oxide ($\varepsilon_{ox} = 3.9\,\varepsilon_0$) and an aluminum gate ($\Phi_M$ = 4.1 V). Assume there is no fixed charge in the oxide or at the oxide-silicon interface.
Solution	The threshold voltage equals: $$V_T = V_{FB} + 2\phi_F + \frac{\sqrt{4\varepsilon_s q N_a \phi_F}}{C_{ox}}$$ $$= -0.93 + 2 \times 0.42 + \frac{\sqrt{4 \times 11.9 \times 8.85 \times 10^{-14} \times 1.6 \times 10^{-19} \times 10^{17} \times 0.42}}{3.9 \times 8.85 \times 10^{-14} / 20 \times 10^{-7}}$$ $$= -0.09\text{ V}$$ Where the flatband voltage was already calculated in example 6.1. The threshold voltage voltages for nMOS and pMOS capacitors with an aluminum or a poly-silicon gate are listed in the table below.

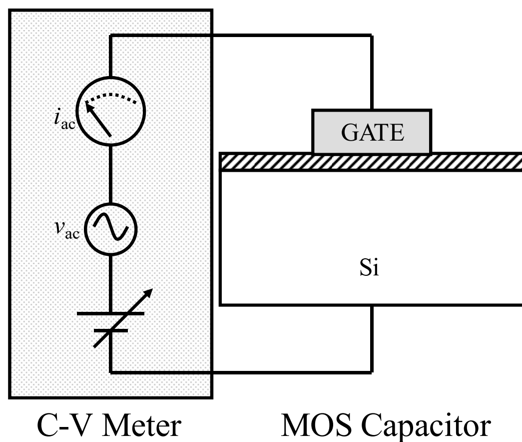
	Aluminum	p$^+$ poly	n$^+$ poly
nMOS	-0.09 eV	0.98 eV	-0.14 eV
pMOS	-0.93 eV	0.14 eV	-0.98 eV

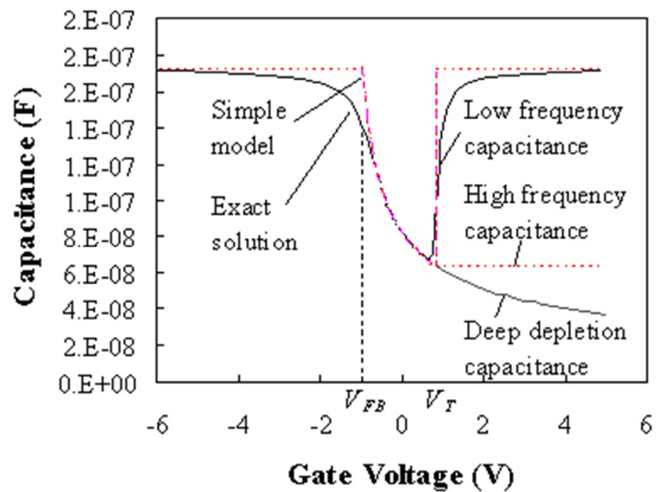
- Three interesting regions of MOS operation:
  - Accumulation ($V < V_{FB}$ for p-substrate)
  - Depletion ($V_{FB} < V < V_T$)
  - Inversion ($V_T < V$)



- Note the signs will change for n-substrate

- How does the measured C-V curve look like?





$$i_{ac} = C\frac{dv_{ac}}{dt}$$

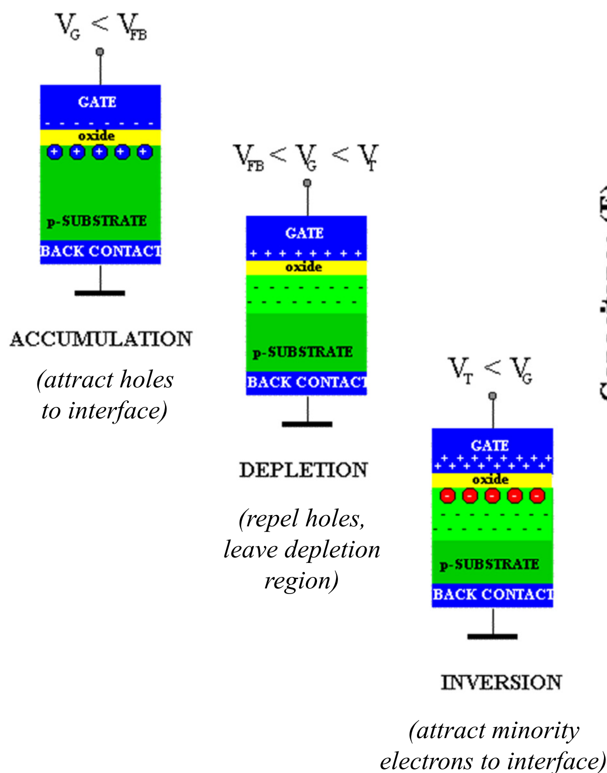
$$C = \left|\frac{dQ_{GATE}}{dV_G}\right| = \left|\frac{dQ_s}{dV_G}\right|$$

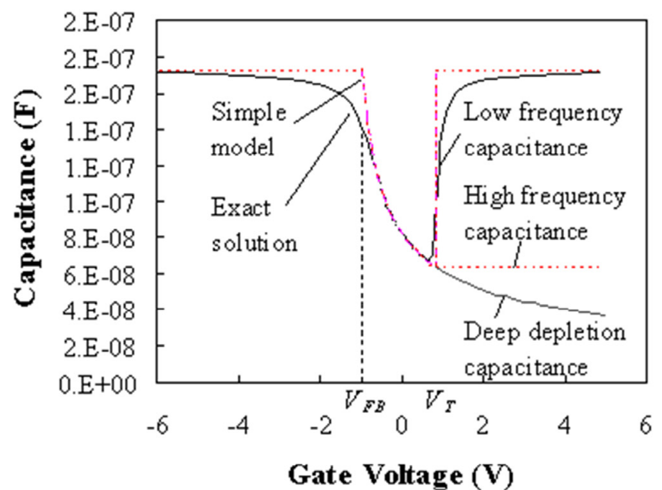
$$\max C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$$ (typically per area, in F/cm²)

- Why does the measured C-V look like that?

$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$

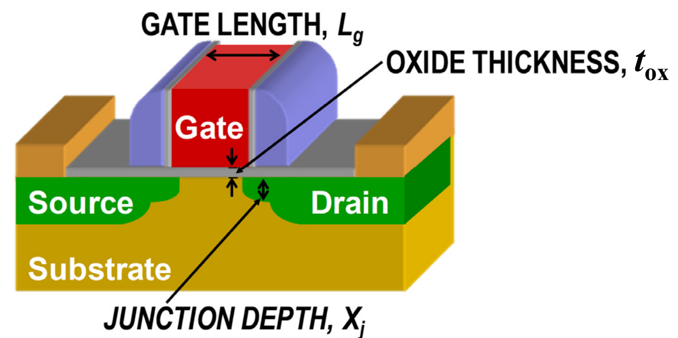
C_d





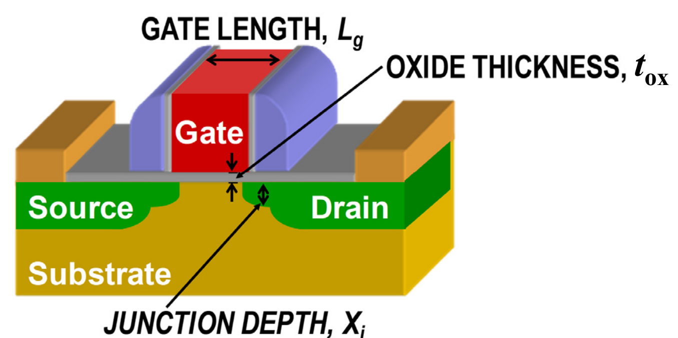
EE 116 Lecture 31

MOS Field-Effect Transistor (MOSFET)



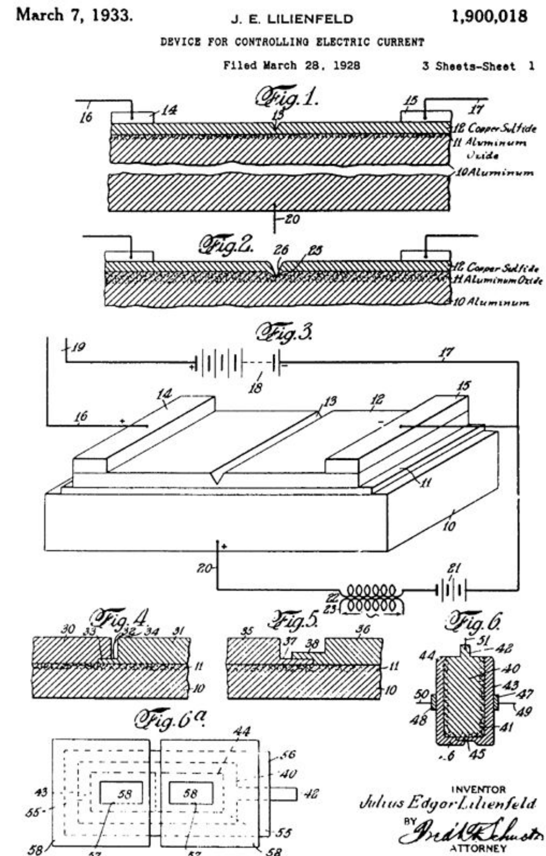
- Read: Ch. 7.1, 7.2, 7.3.1-2
- Skim Ch. 7.5, 7.6.3-4, 7.7
- https://truenano.com/PSD20/contents/toc7.htm

- The MOSFET is a MOS capacitor with Source/Drain terminals
- How does it work?
  - Gate voltage (V_{GS}) controls mobile charge sheet under oxide (in "channel")



  - Source-drain voltage (V_{DS}) sweeps the mobile charge away, creating current (I_D)
- Desired characteristics (water faucet analogy):
  - "On" current ________________
  - "Off" current_________________

- First MOSFET patents: Julius Lilienfeld (early 1930s)
- No experimental demonstration



- This invalidated many of Bardeen, Brattain and Shockley's transistor patent claims in the late 1940s!

- But the MOSFET did not work in practice until the 1960s. Why?

- 1960s – MOSFET Demonstrated

Dawon Kahng

John Atalla

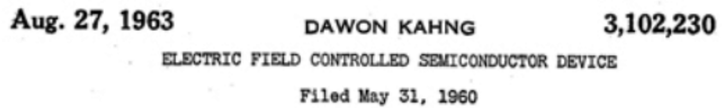
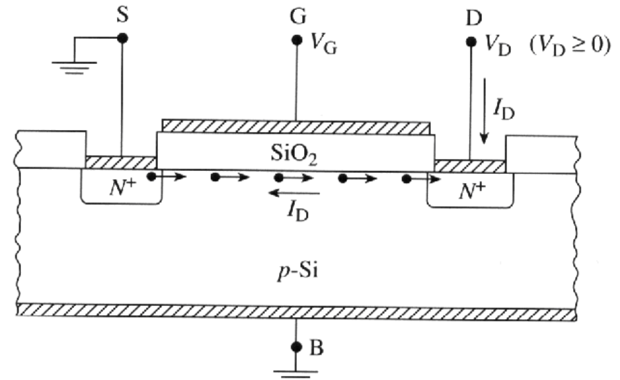


FIG. 1A

FIG. 1B

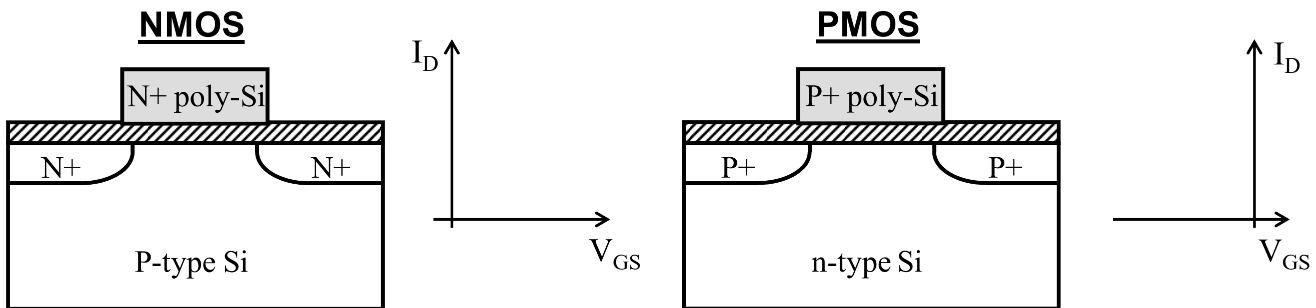
- John Atalla and Dawon Kahng at Bell Labs demonstrate the first successful MOS field-effect amplifier

- Typical 2-D cross-section view of the N-MOSFET:

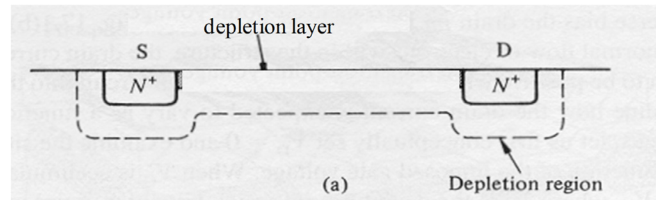


- Gate voltage (V_{GS}) controls Source-to-Drain current (I_D)
- Note direction of carrier flow and of current flow
- "Source" terminal refers to source of carriers (not current)

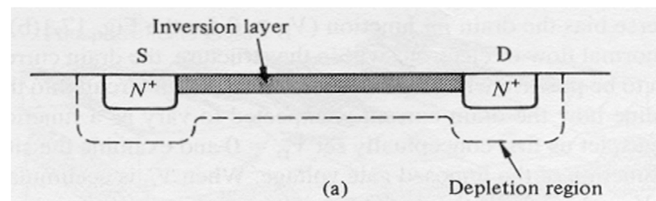
- Two types of MOSFETs: **NMOS** (n-channel) and **PMOS** (p-channel)



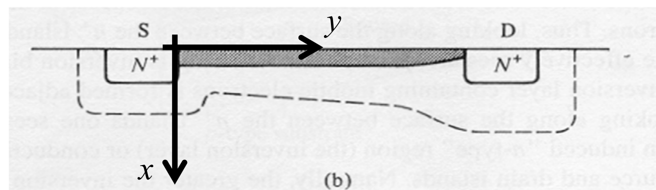
- Theory of the MOSFET (*here N-MOSFET):
  - When $V_{GS} < V_T$ the channel is ________________



  - When $V_{GS} > V_T$ the channel is ________________



  - If small drain voltage ($V_{DS} > 0$) is applied __________



- Will charge sheet move *primarily* by drift or diffusion?

  Current ≈ width X charge sheet X velocity

- What is the inversion charge: $|Q_{inv}| \approx C_{ox}(V_{GS} - V_T)$
- What is the drift velocity: $v_d \approx \mu E \approx \mu(V_{DS}/L)$

- At low V_{DS}, the inversion channel essentially acts like a resistor!

$$I_D = WQ_{inv}v_d \approx WC_{ox}(V_{GS}-V_T)\mu E \approx \mu WC_{ox}(V_{GS}-V_T)\frac{V_{DS}}{L}$$

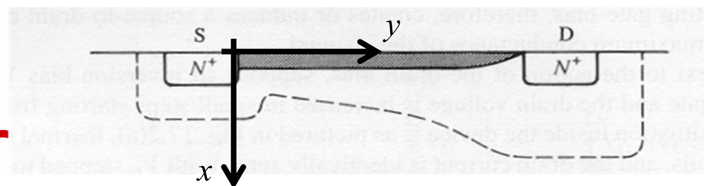
- What about higher drain voltages V_{DS}?
- Must take into account variation of potential along channel, $0 < V_y < V_{DS}$. So inversion layer charge at any point is

$$|Q_{inv}(y)| = C_{ox}(V_{GS}-V_T-V_y)$$

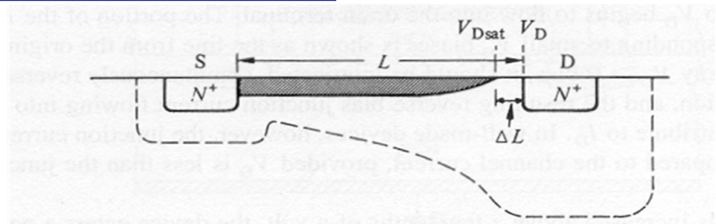
- And the current is:

$$\int_0^L I_D dy = W\int_0^L Q_{inv}(y)\mu\frac{dV_y}{dy}dy \longrightarrow I_D = \mu C_{ox}\frac{W}{L}\left(V_{GS}-V_T-\frac{V_{DS}}{2}\right)V_{DS}$$

- Still linear in V_{GS} voltage! This is the <u>linear</u> region.
- When $V_{DS} = V_{GS} – V_T$ the channel becomes _____________

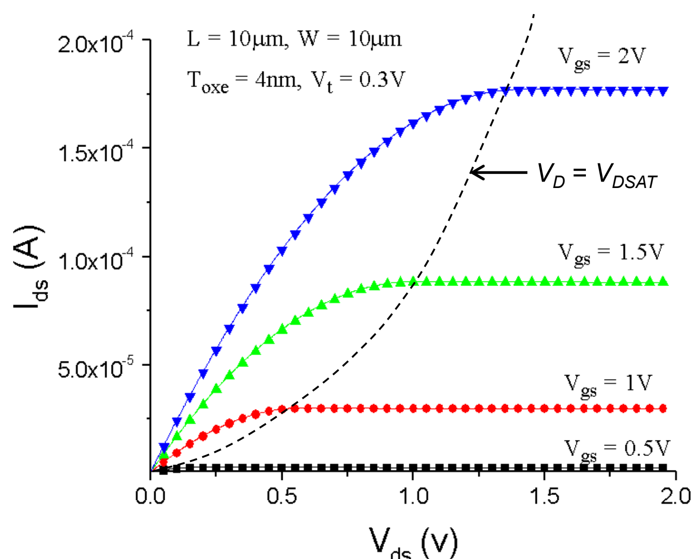


- When $V_{DS} > V_{GS} - V_T$ the un-inverted (drain depletion) region increases, as does the _____________________



- Any increase in V_{DS}:
  - Reduces the amount of inversion charge, but…
  - Increases the lateral field (drift velocity)
- The two effects cancel each other out, so at $V_{DS} > V_{GS} - V_T$ the drain current is no longer a function of V_{DS}! The current <u>saturates</u> to a value only dependent on V_{GS} (i.e. charge).
- Putting in $V_{DS} = V_{GS} – V_T$ (the pinch-off, i.e. saturation condition) in the previous equation:

$$I_D = \mu C_{ox}\frac{W}{2L}(V_{GS}-V_T)^2$$

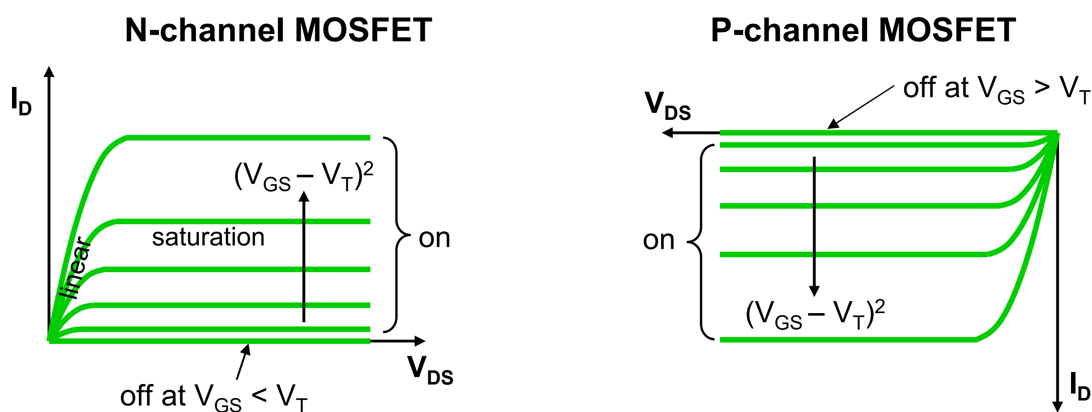


Further increase in V_{DS} does not change I_D (to first order), $I_D \approx$ constant for $V_D > V_{DSAT} = V_{GS} - V_T$

SATURATION REGION.

- This is the simple, so-called quadratic or "long"-channel FET
- Recall what I_D vs. V_{GS} looks like:

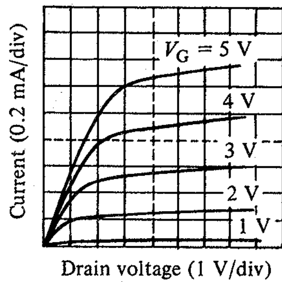
- Recap: increase in $|V_{GS}|$ results in increase of carrier density (inversion charge) in the channel and increase in I_D
- At high $|V_{DS}|$, after pinch-off the I_D saturates



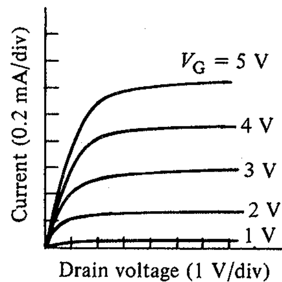
- For P-channel MOSFET (PMOS) all the polarities are reversed and the inversion layer exists when $V_{GS} < V_T < 0$

- Beyond the quadratic MOSFET model (see EE 216, EE 316)
- What happens in "short" channel devices? High-field and v_{sat}
- Ex: modern device L ~ 30 nm, V ~ 1 V → Field ~ 3×10^5 V/cm!

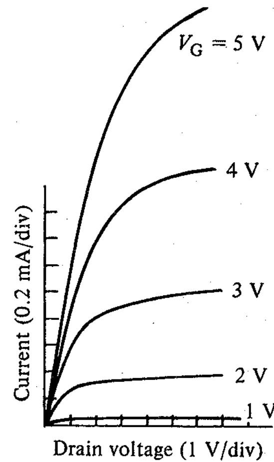
$$I_D \approx WQ_{inv}v_{sat} \approx WC_{ox}(V_{GS}-V_T)v_{sat}$$



Experimental



Computed including velocity saturation



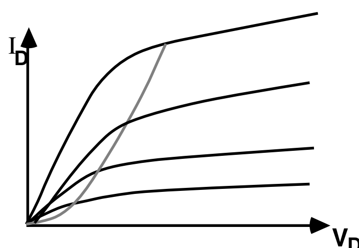
Computed ignoring velocity saturation

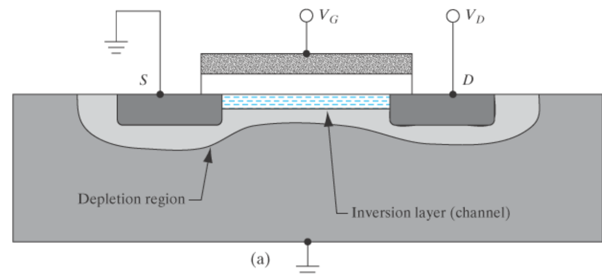
- Back to the physical picture, why does I_D vs. V_{DS} saturate?

- Why is this desirable?
  - Voltage gain, dV_{DS}/dI_D because small changes in I_D cause large swings in V_{DS}

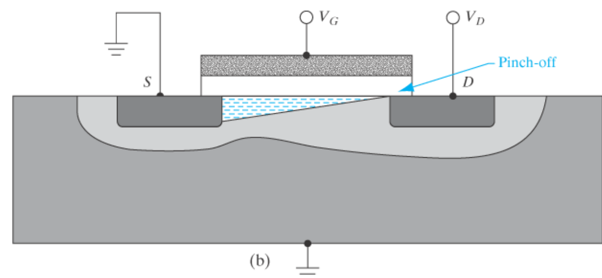
- Note there is some channel length modulation, so empirically we write:

$$I_{D_{SAT}} = \frac{W}{2L}\mu C_{ox}(V_{GS}-V_T)^2(1+\lambda V_{DS})$$

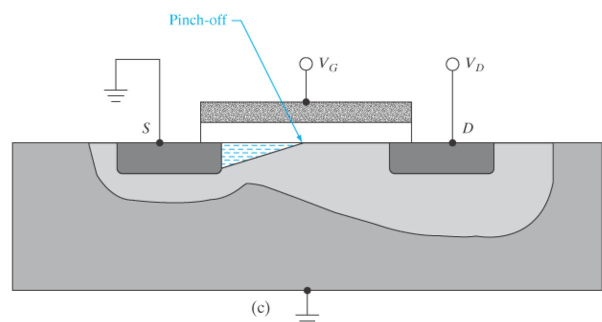




(a)

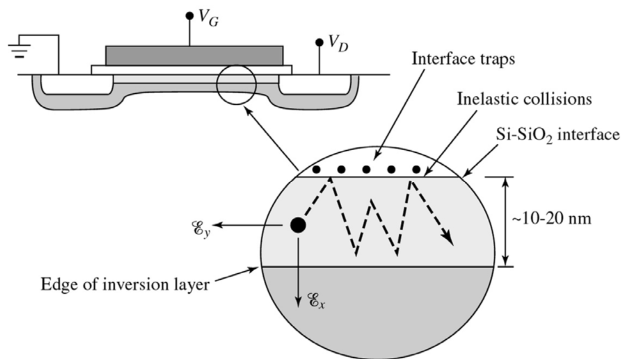


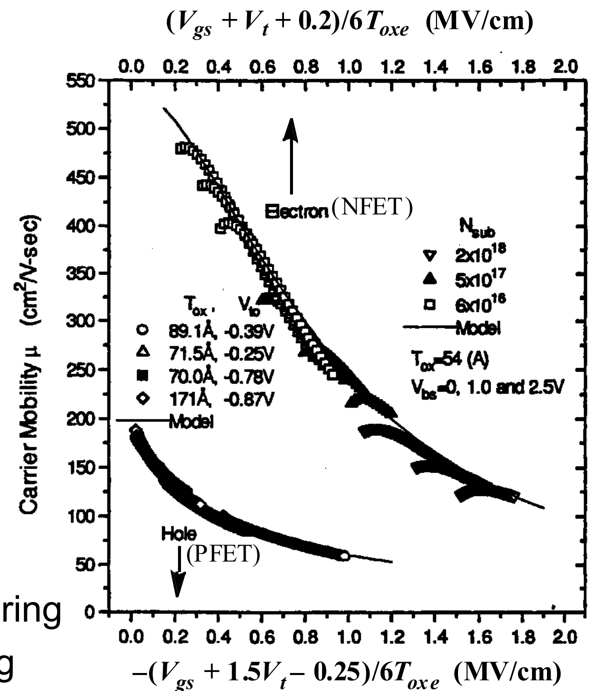
(b)



(c)

- What is the "effective mobility" μ_{eff} in the MOSFET channel?
- Can we look it up in the bulk-silicon charts?



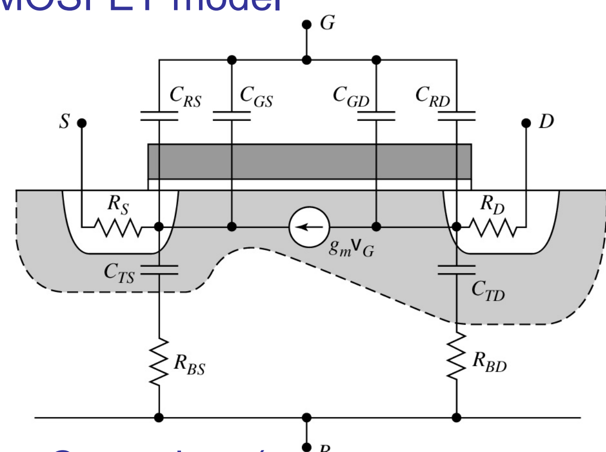


- Scattering mechanisms affecting mobility in channel:
  - Charged impurity (Coulomb) scattering
  - Lattice vibration (phonon) scattering
  - Surface roughness scattering

EE 116 Lecture 32

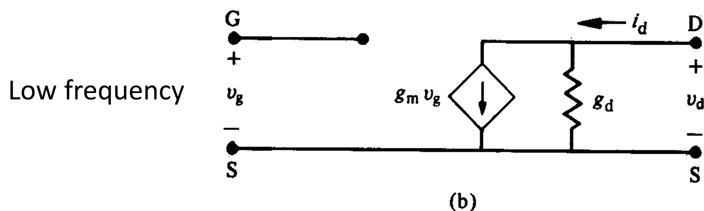
MOSFET Analog Amplifier and Digital Inverter

- Analog applications: Small-Signal MOSFET model

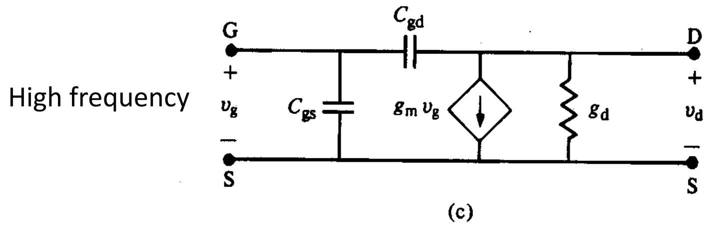


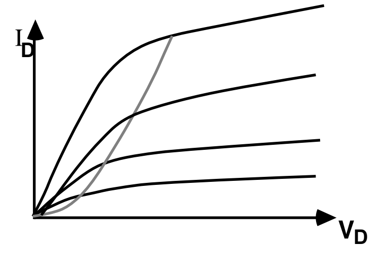
- Of all elements in the model… $C_{GS} \sim C_{ox}$ and g_m (= transconductance dI_D/dV_{GS}) are essential, the rest are parasitics which must be reduced
- Note that a lot of elements are voltage-dependent, e.g. depletion capacitances vary with depletion widths and voltage

Low frequency



High frequency





- Drain current: $i_d = g_d v_d + g_m v_g$

$$I_{D_{SAT}} = \frac{W}{2L}\mu C_{ox}(V_{GS}-V_T)^2(1+\lambda V_{DS})$$

- Conductance parameters:

$$g_d = \left.\frac{\partial I_D}{\partial V_D}\right|_{V_G=const} = \lambda I_{Dsat0}$$ output conductance

$$g_m = \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D=const} \approx \frac{W}{L}\mu C_{\text{ox}}(V_{GS}-V_T)$$ transconductance (here in saturation, using simple quadratic model)

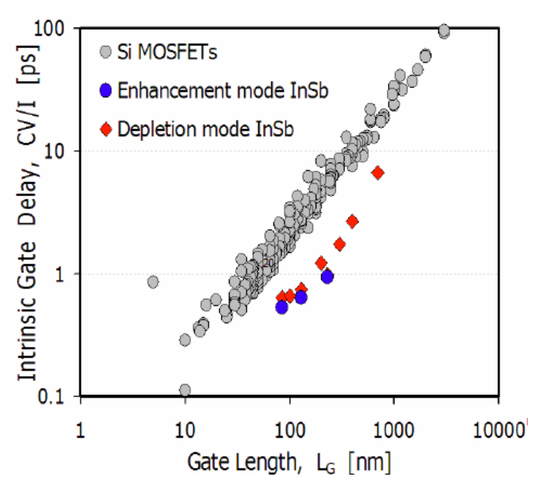
- See EE 216 and tie-ins to circuits classes

- Cutoff frequency f_T = frequency where MOSFET no longer amplifies input (gate) signal
- Obtained by considering high-freq. small-signal model with output shorted, finding freq. where $|i_{out}/i_{in}| = 1$

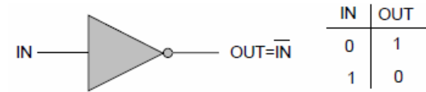
$$f_T = \frac{g_m}{2\pi C_{ox}} = \frac{\mu_{eff}}{2\pi L^2}(V_{GS}-V_T) \propto \frac{1}{L^2}$$

- Something we already knew qualitatively → higher MOSFET operating frequency achieved by decreasing channel length L, increasing mobility μ_{eff}

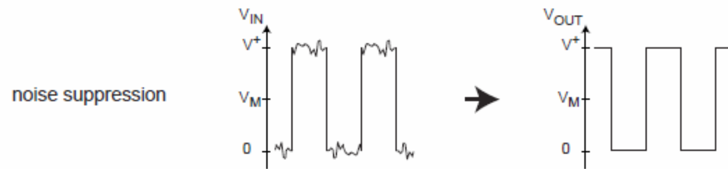
- Smaller = faster for devices (though parasitics play a big role in realistic circuits)



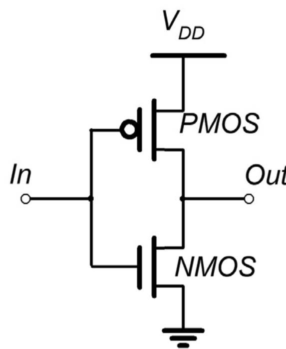
- Logic applications: CMOS inverter

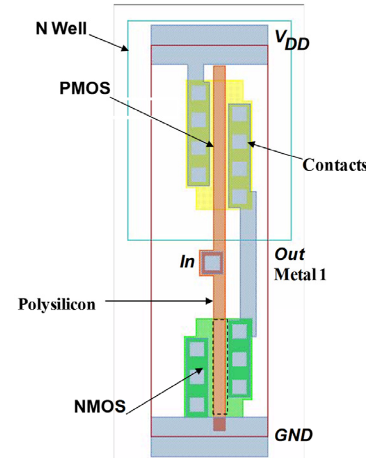


- Key property: signal regeneration – returns logic outputs (0 or 1=V^+=V_{DD}) even in presence of noise



- Complementary MOS (CMOS) inverter

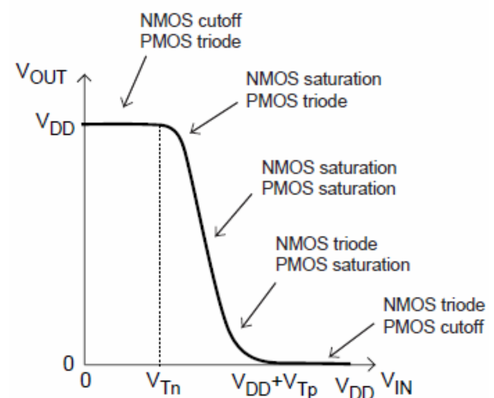




- Qualitative operation:
  - When V_{in} = 0 → V_{out} = V_{DD}
    NFET is off PFET is on
  - When V_{in} = V_{DD} → V_{out} = 0
    NFET is on PFET is off

- Other key property of CMOS inverter: no power consumption while idling in either logic state (only while switching)

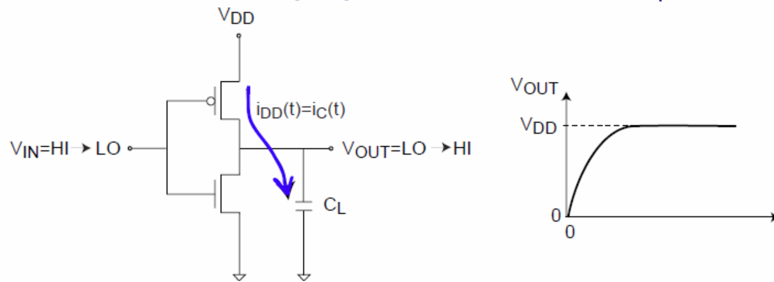
- Consider PFET as "load" to NFET:
- Note "rail-to-rail" logic levels 0 and V_{DD}



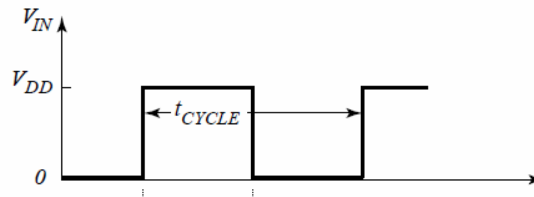
- Want transition voltage $V_{DD}/2$, but usually $L_p = L_n$ which means choose $W_p/W_n \approx 2$ because $\mu_n \approx 2\mu_p$ (for Si)*

*what about other materials?

- A quick look at CMOS power dissipation
- Energy consumed while charging capacitive load: E_P = ________

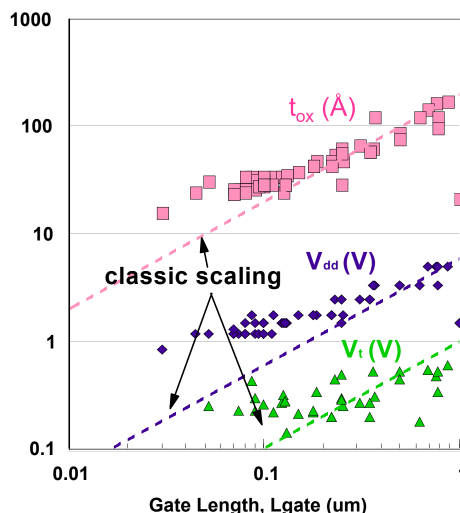


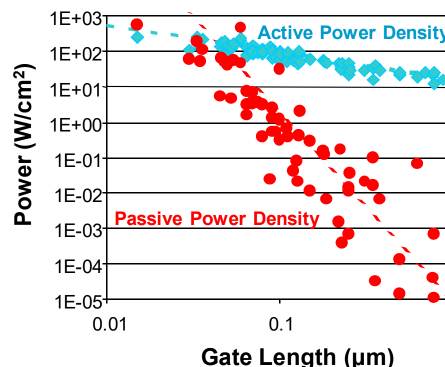
- C_L is discharged through NFET → E_N = __________



- Total energy dissipated per clock cycle: $E = C_L V_{DD}^2$
- Frequency *f* cycles per second → active power $P = fC_L V_{DD}^2$
- This is very important: fundamental trade-off between speed (f) and power dissipation. Reducing voltage and parasitic C's is a must to keep power low at higher speeds.

- In reality, there is also passive power (leakage) dissipated by the FETs supposed to be "off": $P_{off} = I_{leak} V_{DD}$
- $I_{off} \sim I_{on}/10^4$ in modern technology per transistor
- But this can become a headache when you have 100s of millions of "sleeping" transistors (i.e. "passive power" vs. "active power")!





Ex: see IBM journal of Research & Dev.
http://www.research.ibm.com/journal/rd/504/tocpdf.html

Bonus Slides

Modern Transistors

and

How Do We Shrink Them Below 10 nm?

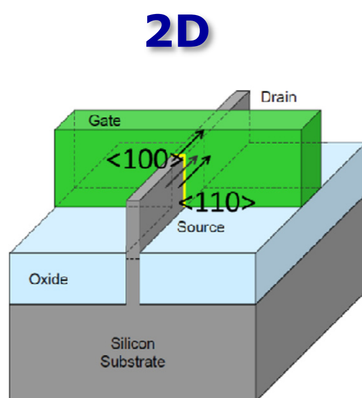
Transistor Scaling Limits

– Better gate control if channel is made thinner!

– Gate lengths L_G scale with channel thickness t_{ch}

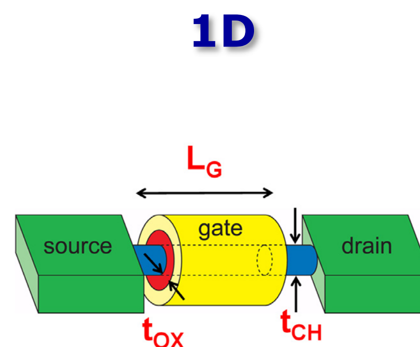
$L_G \sim (t_{ch} t_{ox})^{1/2}$

2D



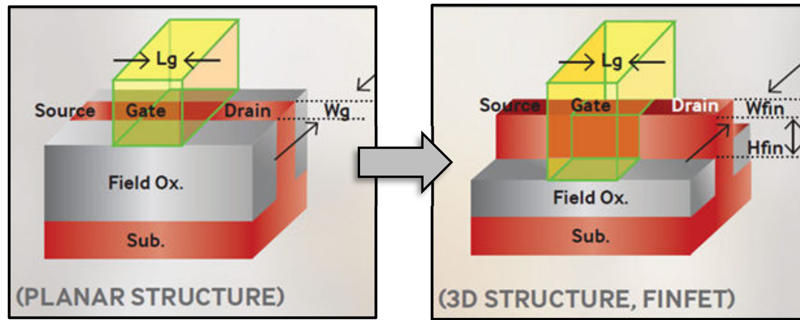
"Fin"FET or tri-gate

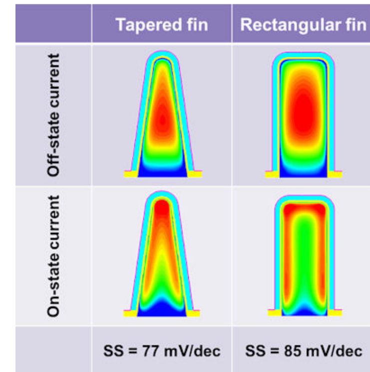
1D



wrap-gate nanowire or nanotube

Transition from Planar FETs to FinFETs

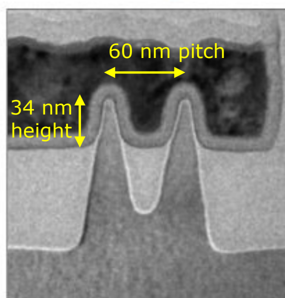




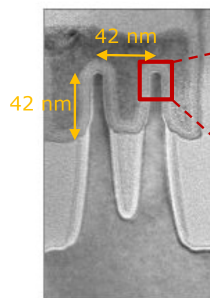
- FinFET: gate controls channel "fin" from 3 sides
- Much better gate control over channel and reduced leakage current
- Transition from planar to FinFETs:
  - Intel at "22 nm" technology node in 2012
  - Samsung and TSMC at "14 nm" technology node in 2014

http://www.eetindia.co.in/ART_8800677161_1800000_TA_721cad8f.HTM

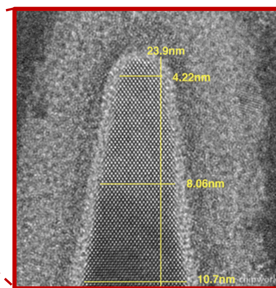
Intel 22 nm and 14 nm Technology FinFETs



22 nm 1st Generation Tri-gate Transistor



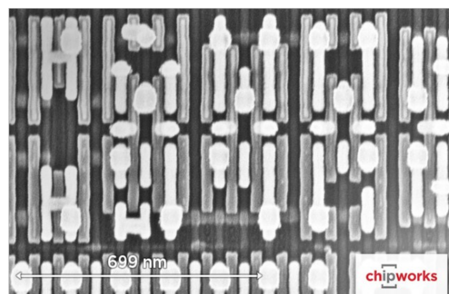
14 nm 2nd Generation Tri-gate Transistor



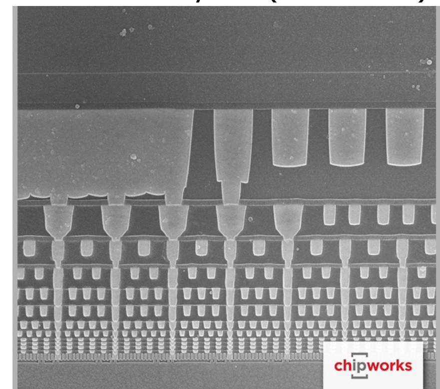
tapered fin
average width = 8 nm

	22 nm Node	14 nm Node
Fin Pitch	60	42
Gate Pitch	90	70

Intel 14 nm SRAM (top view)



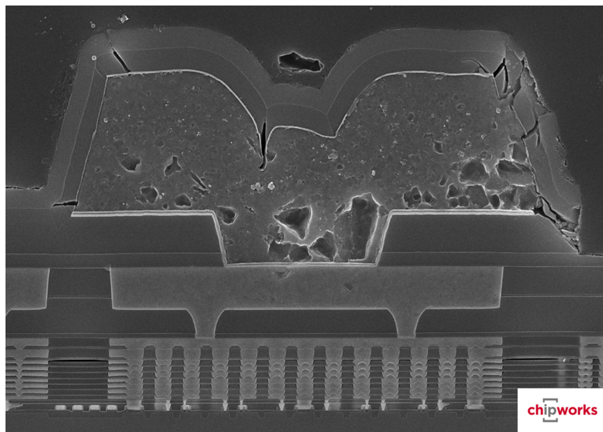
13 metal layers (side view)

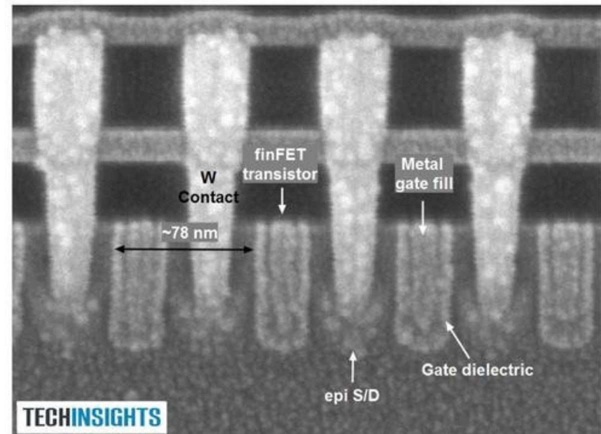


source: Intel and ChipWorks

Samsung 14 nm Technology FinFETs

- Samsung “14 nm” node FinFETs
- Contacted gate pitch (CGP) of ~78 nm
- Available in Galaxy S6 and iPhone 6s processors (A9 chip)

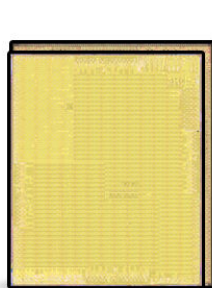




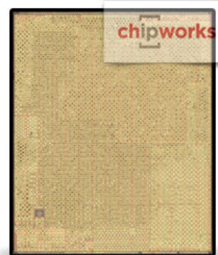
https://www.chipworks.com/about-chipworks/overview/blog/inside-the-samsung-galaxy-s6
http://m.eetasia.com/ART_8800712420_480200_NT_323cab49_3.HTM#.VkuAqL_9nEY

TSMC FinFET Technology for A9 Chip

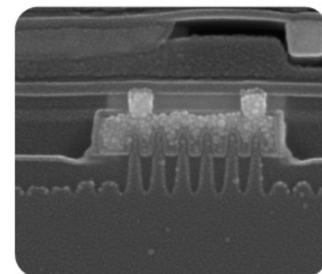
- TSMC makes about half of A9 chips using their “16 nm” FinFETs
- Apple probably set up Samsung and TSMC to compete, and worried that one company alone could not meet demand for the iPhone 6s
- Chip die sizes are slightly different



APL0898
Samsung
96mm²



APL1022
TSMC
104.5mm²

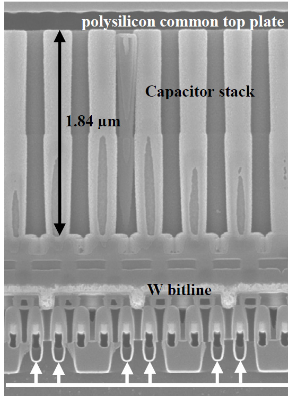


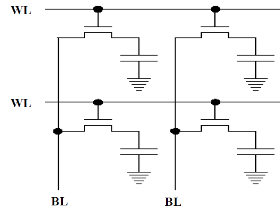
TSMC 16 nm FinFET

https://www.chipworks.com/about-chipworks/overview/blog/a9-is-tsmc-16nm-finfet-and-samsung-fabbed

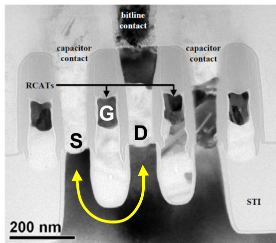
"3D" Transistors in DRAM and Flash

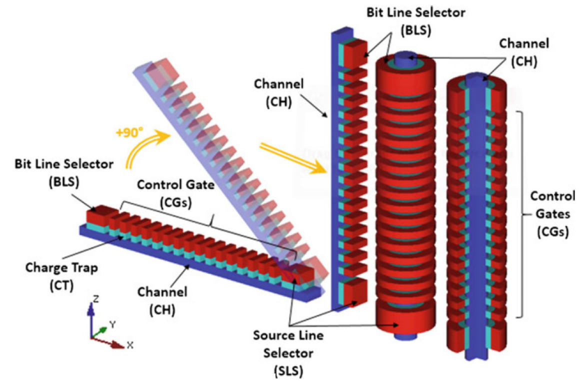
- DRAM access transistors became U-shaped ~2005
- Flash memory went "vertical" ~2013

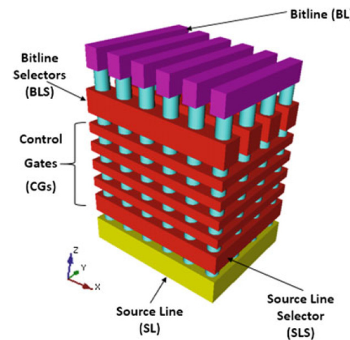


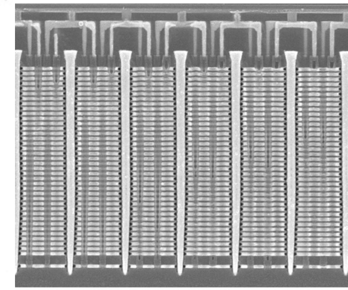


U-shaped DRAM transistors





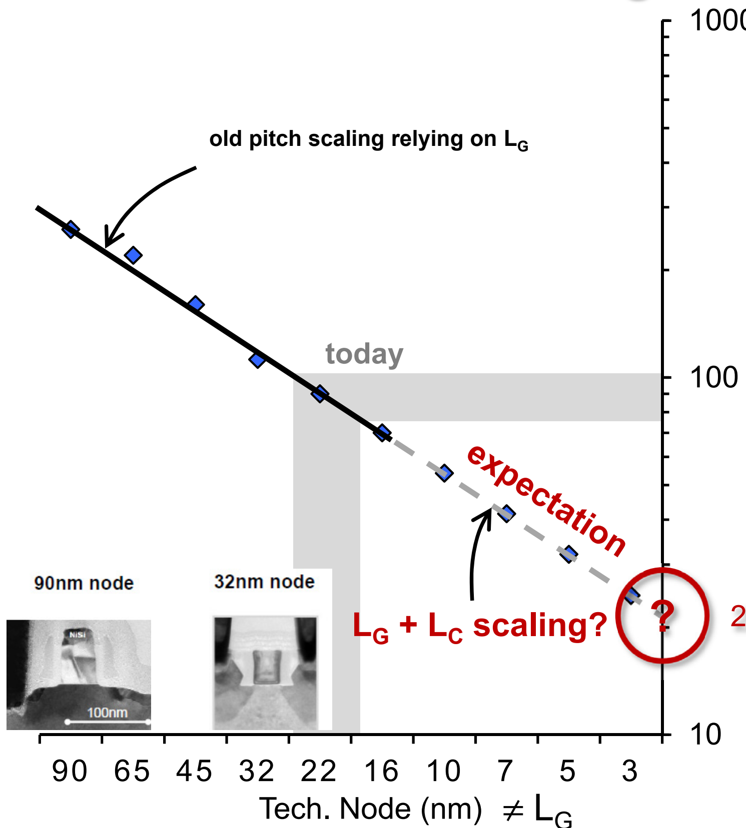


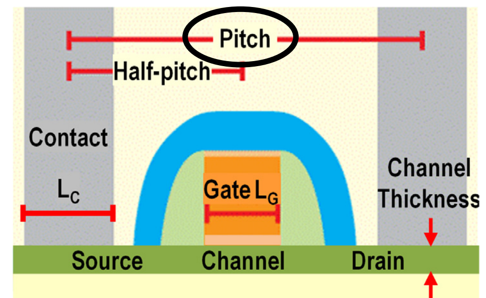


Close-up image of V-NAND flash array (32-layers)

 Source: ChipWorks

Transistor Scaling: Gate + Contacts





Gate Pitch is key metric of device scaling

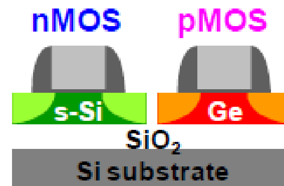
Must scale both gate and contacts (L_G and L_C)

Contacting atomically thin materials is difficult!

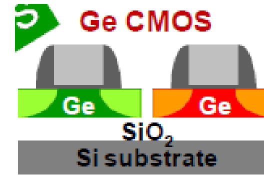
source: IEEE Spectrum, Intel

Future CMOS Choices?

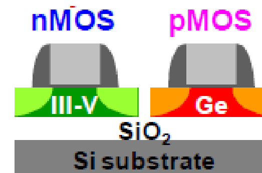
- Si NMOS and Ge PMOS
  - Technologies well developed
  - Marginal benefit



- Ge NMOS and PMOS
  - High μ_n and μ_p demonstrated
  - NMOS needs more work



- III-V NMOS and Ge PMOS
  - Several efforts underway on III-V NMOS
  - Integration of III-V and Ge on Si ???



- III-V NMOS and PMOS
  - Single material needed for both NMOS and PMOS
  - Very little work on III-V PMOS