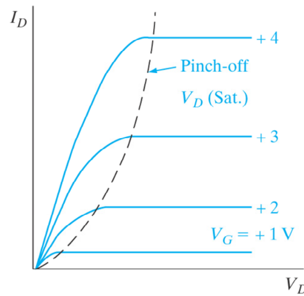
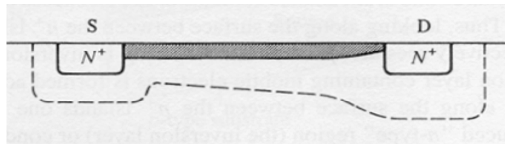
ECE 340: Supplementary Lecture (L42)

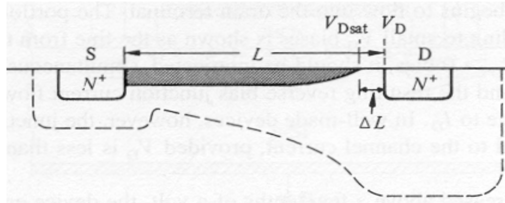
Short-Channel MOSFET and Non-Ideal Behavior

Recap the "long channel" MOSFET on one page:





$$I_{D,lin} = \frac{Z}{L}\mu_{eff}C_i\left(V_{GS} - V_T - \frac{V_{DS}}{2}\right)V_{DS}$$

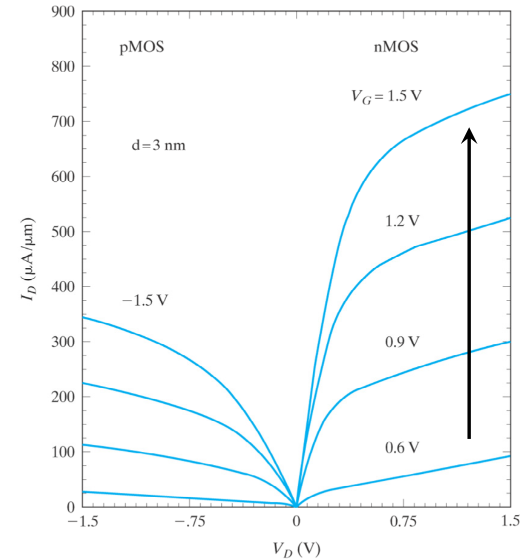


$$I_{D,sat} = \frac{1}{2}\frac{Z}{L}\mu_{eff}C_i(V_{GS} - V_T)^2$$

$I_D \approx Z \times$ charge sheet $\times$ velocity (cm $\times$ C/cm^2 $\times$ cm/s = C/s = A)

For $V_D \geq V_{GS} - V_T$ the channel gets "disconnected" from V_D, and only a function of $V_{D,sat} = V_{GS} - V_T$.

The "short channel" MOSFET.

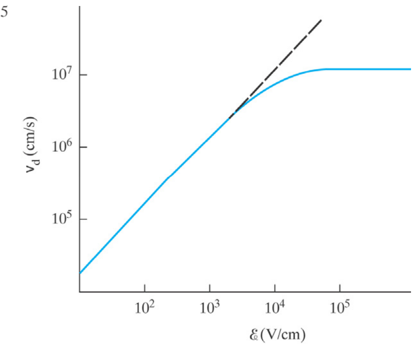


Note, in "saturation":

- **I_D does not increase as V_{GS}^2**
- **I_D remains a (weak) increasing function of V_{DS} (remember B.I.G…. mo' voltage mo' current).**

Modern devices L ≈ 35 nm, so at $V_D \approx 1$ V → E ≈ 3×10^5 V/cm

Carrier velocity is almost always saturated! $v_{sat} \approx 10^7$ cm/s



So if the drain current $I_D \approx Z \times$ charge sheet $\times$ velocity

Then approximately: $I_D \approx ZC_i(V_{GS}-V_T)v_{sat}$

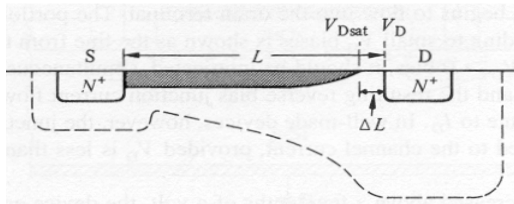
Last result is a bit counterintuitive, but pretty important!

- **$I_{D,sat}$ proportional to $(V_{GS} - V_T)$ rather than $(V_{GS} - V_T)^2$**
- **$I_{D,sat}$ is practically independent of L!**
- **So to get more current, scale <u>V_T and/or C_i</u> instead of L**

Luckily, we do get one break as devices get smaller = velocity overshoot:

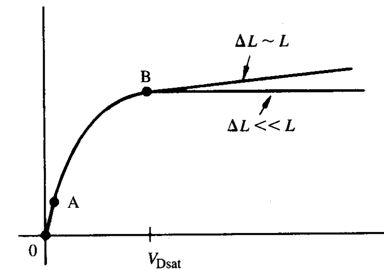
- **For very short L (~ 10–50 nm) some electrons travel across the channel without ___scattering_____**
- **Those "ballistic" electrons can achieve v > v_{sat}**
- **This effectively raises the *collective* $v_{sat} > 10^7$ cm/s (~30% increase for 100 nm NFET, more if shorter)**

What about the weak dependence of $I_{D,sat}$ on V_{DS}?



This is due to "channel length modulation" → L' = L – ΔL

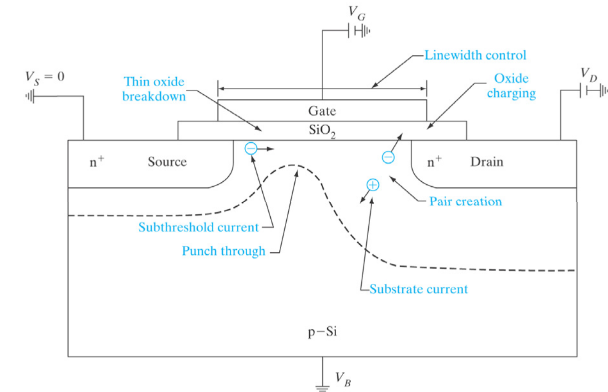
Where ΔL α $(V_{DS} - V_{DS,sat})$

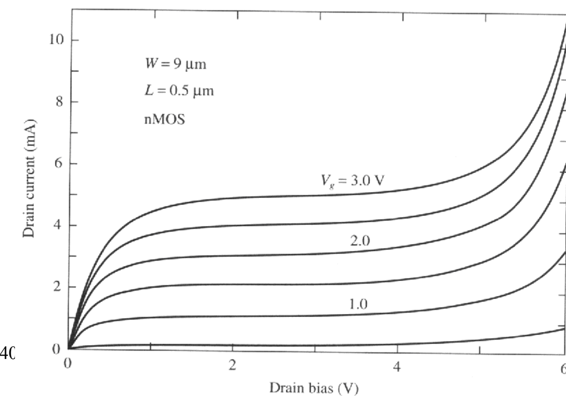


$$\frac{\Delta L}{L} = \lambda(V_{DS} - V_{DSsat})$$

$$I'_{D,sat} = I_{D,sat}[1 + \lambda(V_{DS} - V_{DSsat})]$$

What if I keep increasing V_D and/or decreasing L?





Avalanche (impact ionization)

<u>A note on sub-threshold behavior</u>. So far, we've assumed no channel current (I_D=0) if $V_{GS} < V_T$. This is incorrect.

If $\phi_S > \phi_F$, there is some (weak) inversion charge at the surface, which gives rise to sub-threshold current flowing between the source and drain:

$$I_D = \mu_{eff} C_{ox} \frac{Z}{L}(m-1)\left(\frac{kT}{q}\right)^2 e^{q(V_G - V_T)/mkT}(1 - e^{-qV_{DS}/kT})$$

Where m is ___________________.

Note exponential dependence on voltages, suggesting I_D below threshold is in fact limited by __________________.

Define sub-threshold slope, S:

$$S \equiv \left(\frac{d(\log_{10} I_D)}{dV_{GS}}\right)^{-1} = \frac{kT}{q}\ln(10)(1 + \frac{C_d}{C_{ox}})$$

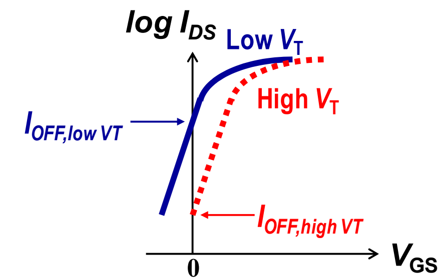
Typical value: 70–100 mV/decade. Want it small! (why?)

Transistor constant-field scaling (Dennard's Law)

	MOSFET Device and Circuit Parameters	Multiplicative Factor ($\kappa > 1$)
Scaling assumptions	Device dimensions (t_{ox}, L, W, x_j)	$1/\kappa$
	Doping concentration (N_a, N_d)	κ
	Voltage (V)	$1/\kappa$
Derived scaling behavior of device parameters	Electric field ($\mathscr{E}$)	1
	Carrier velocity (v)	1
	Depletion-layer width (W_d)	$1/\kappa$
	Capacitance ($C = \varepsilon A/t$)	$1/\kappa$
	Inversion-layer charge density (Q_i)	1
	Current, drift (I)	$1/\kappa$
	Channel resistance (R_{ch})	1
Derived scaling behavior of circuit parameters	Circuit delay time ($\tau \sim CV/I$)	$1/\kappa$
	Power dissipation per circuit ($P \sim VI$)	$1/\kappa^2$
	Power–delay product per circuit ($P\tau$)	$1/\kappa^3$
	Circuit density ($\propto 1/A$)	κ^2
	Power density (P/A)	1

V_T design trade-off:

- **High ON current I_D α $(V_{DS} - V_T)^n$ requires ________ V_T**
- **Low OFF current requires _________ V_T**



Since V_T cannot be scaled down aggressively, the power-supply voltage (V_{DD} = max V_{DS}) has not been scaled down in proportion to the MOSFET dimensions:

Feature Size (μm)	Power-Supply Voltage (V)	Gate Oxide Thickness (Å)	Oxide Field (MV/cm)
2	5	350	1.4
1.2	5	250	2.0
0.8	5	180	2.8
0.5	3.3	120	2.8
0.35	3.3	100	3.3
0.25	2.5	70	3.6

Today (if using SiO_2 gate oxide):

~35-45 nm ~1-1.25 V 10-20 Å 5 MV/cm

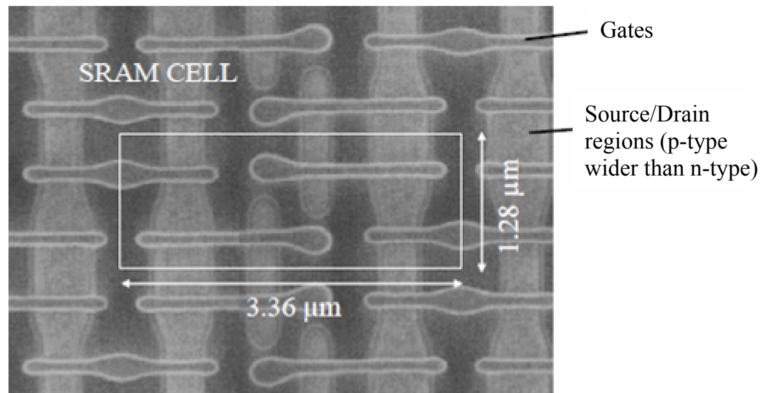
…with HfO_2 gate oxide: _______ ________

ECE 340: Supplementary Lecture (L43)

MOSFET Scaling Limits

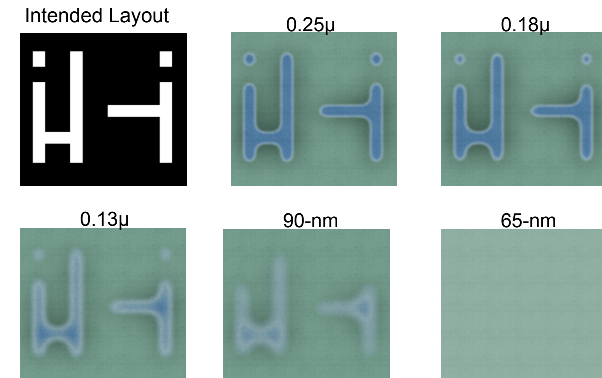
Some modern CMOS limitations:

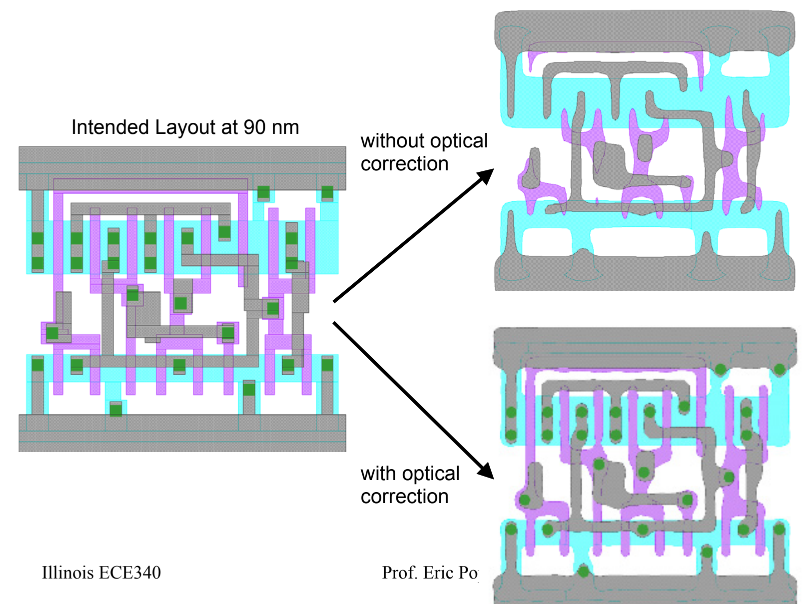
1) **Power and leakage were shown in L42.**
2) **Memory is often a speed bottleneck. So modern CPUs are full of on-chip cache, which occupies 50% or more of floor space, leaving less room for logic. Why? An SRAM cell is BIG (six transistors), area ~ 20x7 = 140F^2 where F is the transistor technology node (e.g. 65 nm).**



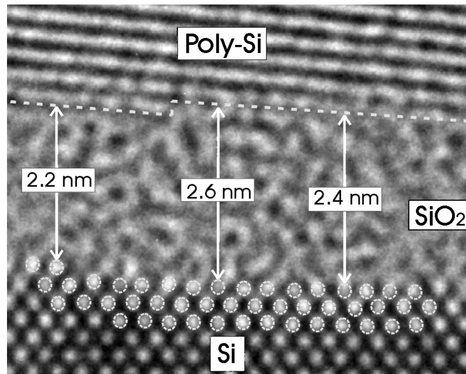
Note: Invention of *electrically accessible* (preferably *non-volatile*) memory with *high speed* and *high density* would lead to a revolution in computer architectures.

3) **Variability in circuit printing (lithography) as features are well below the light wavelength used.**

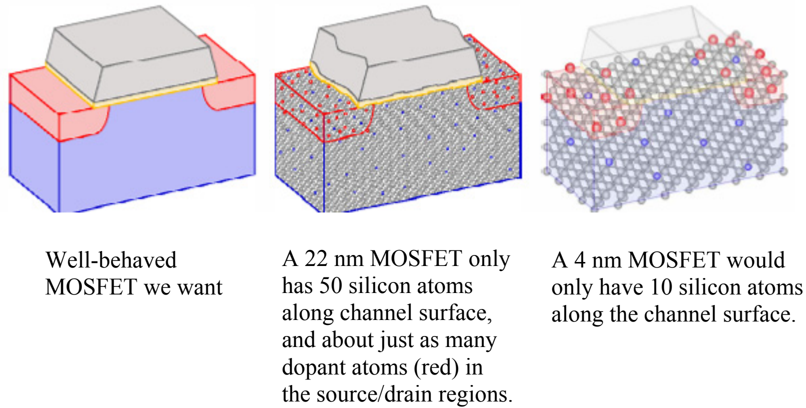




4) Variability in gate oxide thickness (moving to HfO_2 helps, as it gives us more room for atomic-layer error)



5) Variability in dopant atom numbers



Well-behaved MOSFET we want

A 22 nm MOSFET only has 50 silicon atoms along channel surface, and about just as many dopant atoms (red) in the source/drain regions.

A 4 nm MOSFET would only have 10 silicon atoms along the channel surface.