Lecture 26: Device Characterization

- **Announcements:**
- Lab 2 Report will be due Friday, Dec. 12 (during the RRR week), 7 p.m.
- HW#11 due Tuesday, next week, at 8 a.m.
- Final Exam will be Friday, Dec. 19, 8-11 a.m. in 3106 Etcheverry
- ---
- **Lecture Topics:**
  - Threshold Implant
    - Threshold voltage
    - Needed ΔV_t
    - V_t Implant Cases
  - Review of MOS Device Modeling
  - Device Characterization
    - Practical problems and solutions
    - Extraction of parameters
  - Go through Final Exam Info Sheet
    - Pass out example Final Exam
  - Course Wrap Up
  - Course Evaluations
- ---
- **Last Time:**

Question: How is V_t best adjusted?
What factors most impact V_t?

$$V_t = \phi_{ms} - \psi_s - \frac{Q_{ss}}{C_{ox}} - \frac{Q_B}{C_{ox}}$$

Parameters to Adjust:

① $\psi_s = 2\phi_f$: $\phi_f = \frac{kT}{q}\ln\left(\frac{N_D}{n_i}\right)$ for n-substrate (n-doping conc.; intrinsic conc.)

$\phi_f = \frac{kT}{q}\ln\left(\frac{n_i}{N_A}\right)$ for p-substrate (p-doping conc.; intrinsic conc. for undoped Si)

These are logarithmic w/ doping conc.!

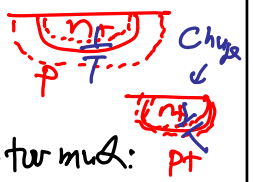
i.e., 10X increase in $N_B \rightarrow \underbrace{2.3\frac{kT}{q} \sim 60\,mV}_{\text{very small change}}$

∴ ϕ_f not a good way to adjust V_t

② $\phi_{ms} = \phi_{f(sub)} - \phi_{f(gate)} \rightarrow$ ineffective for the same reason as

③ $|Q_B| = \sqrt{2q\epsilon_s N_B\,(2|\phi_f| + |V_{SB}|)}$ ☆

can increase $|Q_B|$ w/ $N_B\uparrow$
(can get significant ΔV_t here)



...but if you must increase N_B too much:
⇒ problems: ① lower carrier mobility, $\mu\downarrow$
② S/D capacitance ↑
③ lower junction breakdown voltage

✱ → can also $\Delta V_{FB} \rightarrow \Delta V_t$
↳ impractical → many devices would need to have their own well — too much area!

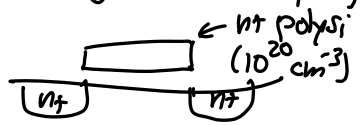
④ $C_{ox}\downarrow$: but then lose drive

⑤ $\frac{Q_{ss}}{C_{ox}}$: Q_{ss} due to oxide-Si interface charge
not controllable (easily)
↑ want to minimize
↳ but if we could introduce a controlled amount of Q_{ss} → best way to get ΔV_t

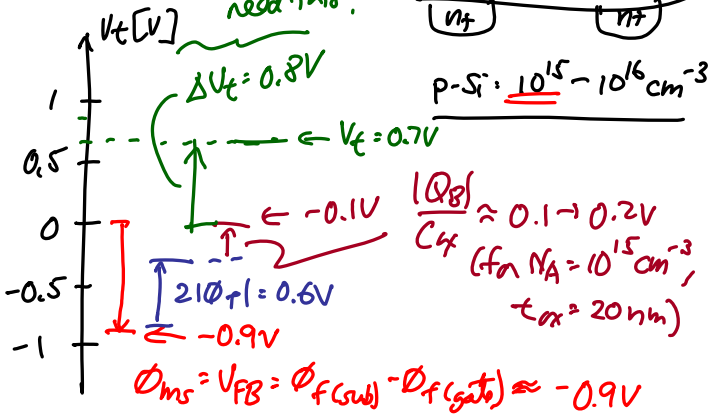
Ex. Threshold implant for NMOS
↳ enhancement implant

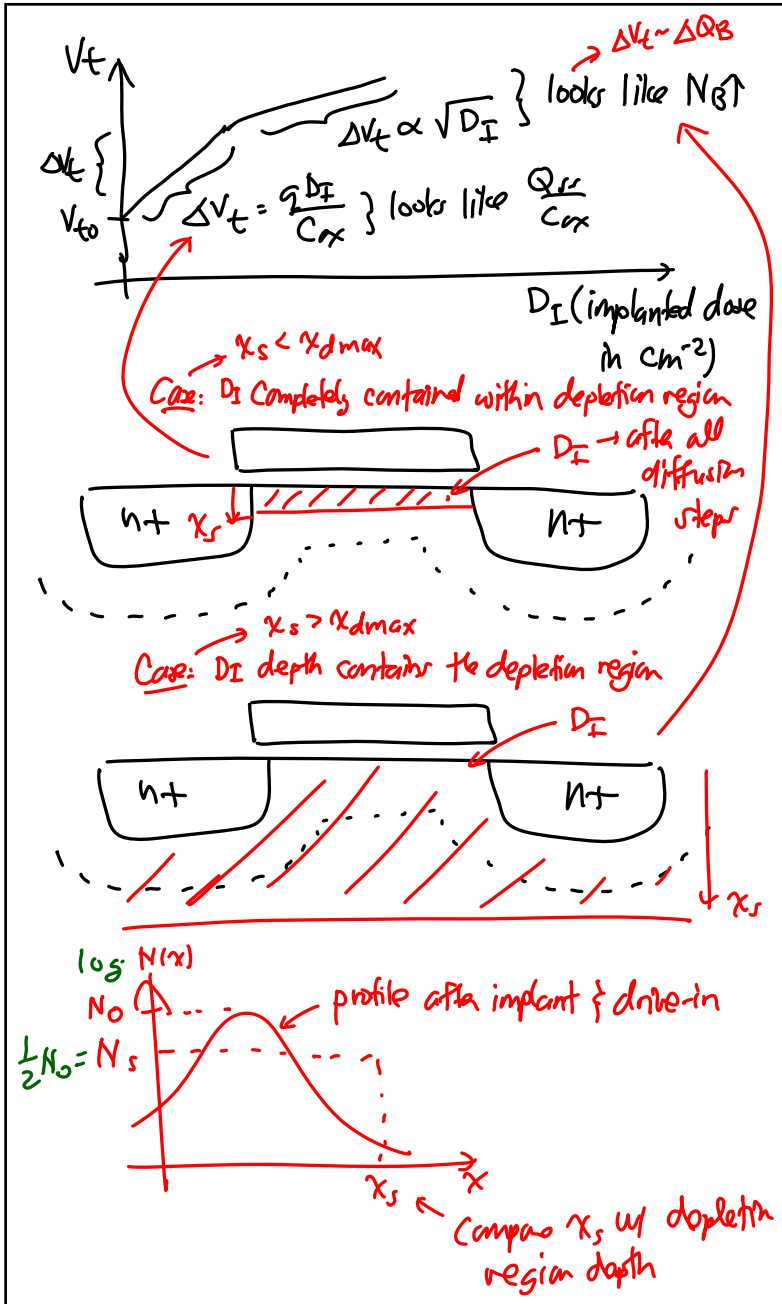
$$V_t = V_{FB} - 2\phi_f - \frac{Q_B}{C_{ox}}$$ } starting V_t (before implant)

(−) (−) (−)



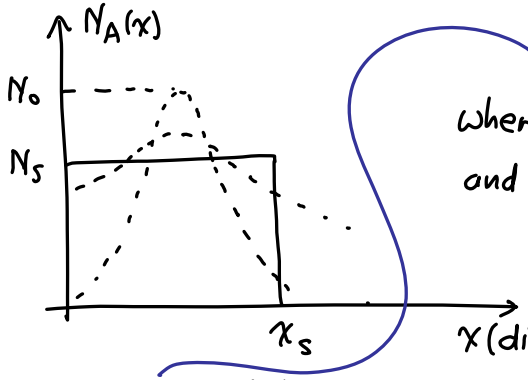
p-Si: $10^{15} - 10^{16}$ cm^{-3}





V_t Implant Design Procedure

Several cases → to distinguish between cases, approx. the threshold implant by a step function:



where $N_s = \frac{1}{2} N_0$

and $N_s x_s = Q' \propto D_I$

↑ want this

total dose in the Si

This just turns out to be a good approx.

↳ actual value can be determined experimentally

*

Case ①: δ function of (−) charge (ionized B acceptors) @ the Si-SiO_2 interface ↳ (i.e., $x_s = 0$)

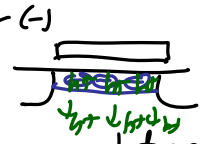
$$\boxed{\Delta V_t = -\frac{qD_I}{C_{ox}}} \propto D_I$$

⇒ equivalent to a reduction of fixed oxide charge Q_{ss} by qD_I:

$$V_t = V_{FB} - 2\phi_f - \frac{Q_B}{C_{ox}} - \frac{Q_{ss}}{C_{ox}} - \frac{qD_I}{C_{ox}}$$

(−) (−) (+) ← (−)

$\Delta V_t = (+)$ $\Delta V_t = (+)$ $\Delta V_t = (-)$ $\Delta V_t = (+)$ ⇒ what we need!



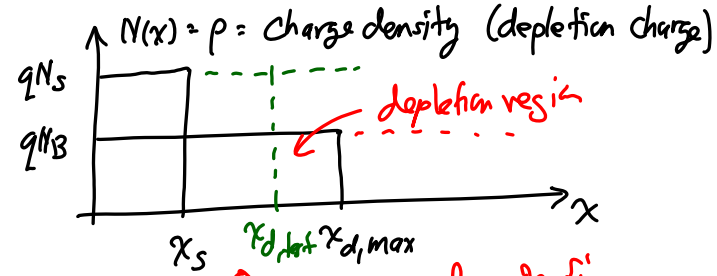
Case ②: x_s = finite < $x_{d,test}$

where $x_{d,test} = \sqrt{\frac{2\epsilon_s}{q}\frac{1}{N_s}(2\phi_{f,test})}$ not N_B!

where $\phi_{f,test} = V_T \ln\left(\frac{N_s}{n_i}\right)$

} Note that these are just test conditions!

If true, then all the implanted dose is within the depletion region w/ $x_{d,max}$ in the N_B-doped area: i.e.,



Q Q → dose in Si

$$\boxed{\Delta V_t = -\frac{qD_I}{C_{ox}}} \propto D_I \text{ (still)}$$

Case ③: $x_s > x_{d,test}$

⇒ depletion region completely contained within the implanted x_s region

$$V_t = V_{FB} + |\psi_s| + \frac{\sqrt{2q\epsilon_s N_s}}{C_{ox}}\left(|\psi_s| + V_{SB}\right)^{1/2}$$

$$|\psi_s| = 2\left|\frac{kT}{q}\ln\frac{n_i}{N_s}\right|$$

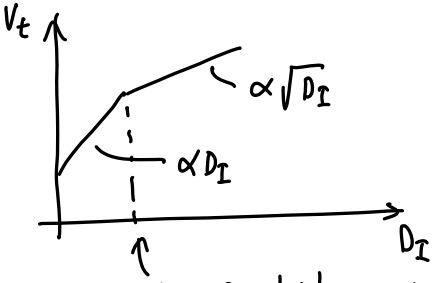
this term Δ's the most!

Thus, for $N_s \gg N_B$:

$$\Delta V_t = \frac{\sqrt{2q\epsilon_s N_s}}{C_{ox}}\left(|\psi_s| + V_{SB}\right)^{1/2} \propto \sqrt{N_s} \Rightarrow \Delta V_t \propto \sqrt{D_I}$$

$$\left[N_s = \tfrac{1}{2}N_0 = \frac{1}{2}\frac{D_I/2}{\sqrt{\pi(Dt)}} \Rightarrow N_s \propto D_I\right]$$

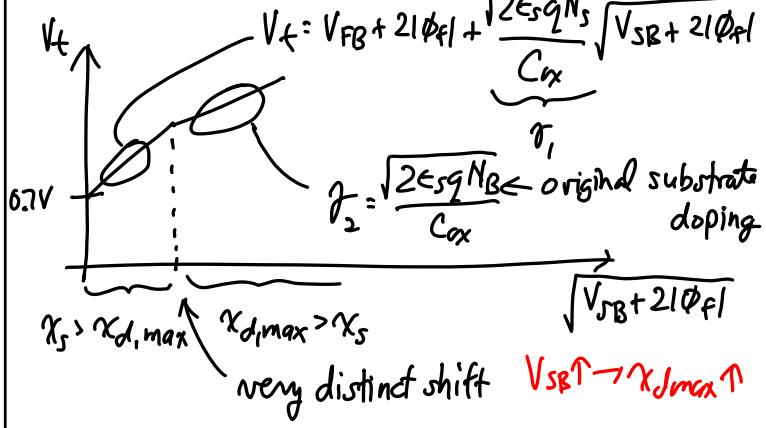
and



$N_s \propto D_I$ high enough that $x_s > x_{d,test}$

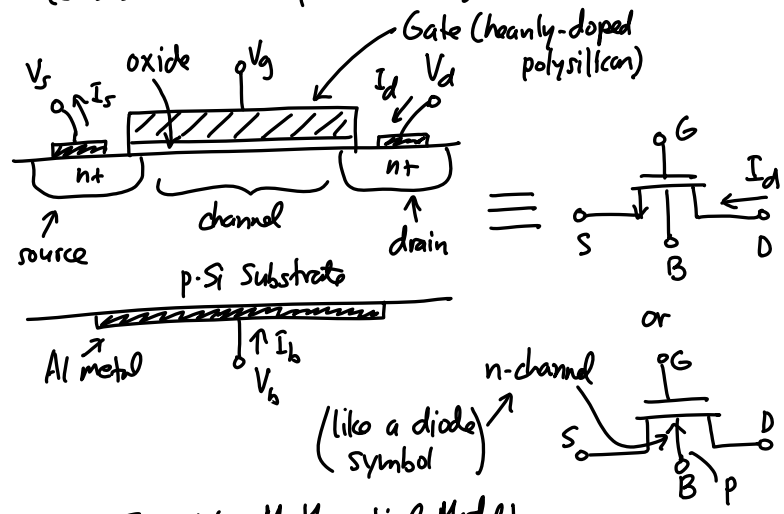
When x_s passes through $x_{d,max}$ (or vice versa) → can see a distinct change in body effect parameter γ as $V_{SB}\uparrow$:

$$V_t = V_{FB} + 2|\phi_f| + \underbrace{\frac{\sqrt{2\epsilon_s q N_s}}{C_{ox}}}_{\gamma_1}\sqrt{V_{SB} + 2|\phi_f|}$$

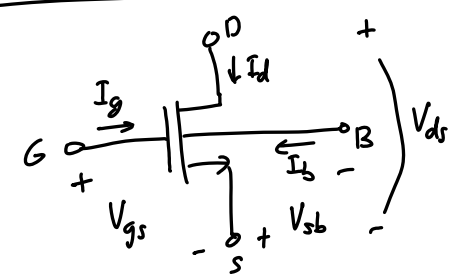


MOS Device Model Summary

(cross-section view of NMOS device)



NMOS Transistor Mathematical Model



① Cut-Off Region: ($V_{gs} \le V_t$)

$I_g = I_b = 0$; $I_d = 0$

② Linear (or Triode) Region: $(V_{gs}-V_{tn} \geq V_{ds} \geq 0)$

$$I_g = I_b = 0; \quad I_d = \mu_n C_{ox} \frac{W}{L}\left(V_{gs}-V_{tn}-\frac{V_{ds}}{2}\right)V_{ds}$$

$$= k_n\left(V_{gs}-V_{tn}-\frac{V_{ds}}{2}\right)V_{ds}$$

③ Saturation Region: $(V_{ds} \geq V_{gs}-V_{tn} \geq 0)$

$$I_g = I_b = 0; \quad I_d = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{gs}-V_{tn})^2(1+\lambda V_{ds})$$

$$= \frac{1}{2}k_n(V_{gs}-V_{tn})^2(1+\lambda V_{ds})$$

where:

$\mu_n \triangleq$ e⁻ mobility in the channel

$C_{ox} \triangleq$ gate oxide capacitance per unit area

$k_n = k_n' \frac{W}{L} = \mu_n C_{ox}\frac{W}{L}$

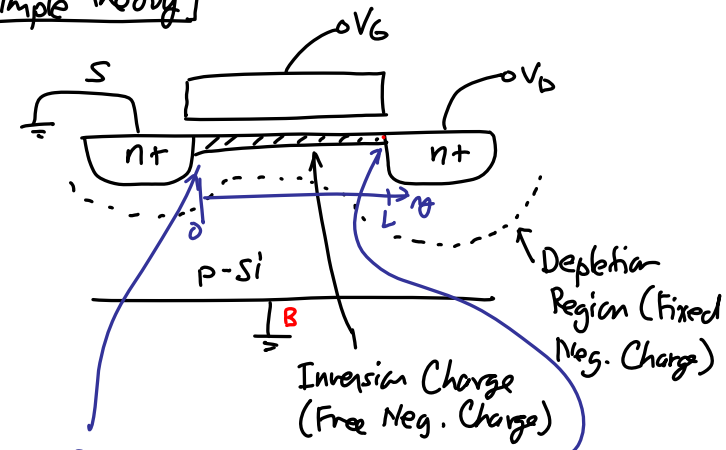
$I_g = I_b = 0$ for all regions (at least for dc)

$$V_{tn} = f(V_{SB}) = V_{t0} + \gamma\left(\sqrt{V_{SB}+2|\phi_f|}-\sqrt{2|\phi_f|}\right)$$

Body Factor → $\gamma = \frac{1}{C_{ox}}\sqrt{2q\epsilon_s N_{sub}}$ ← substrate doping conc.

MOSFET IV Curves

Simple Theory

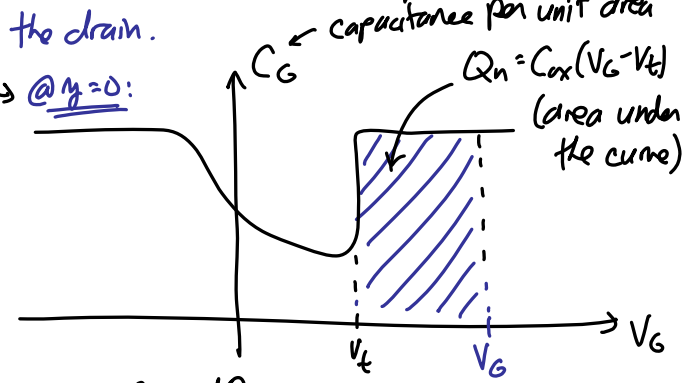


$Q_n(y=0) = C_{ox}(V_G - V_t)$

$Q_n(y=L) = C_{ox}(V_G - V_D - V_t)$

Thus, the inversion charge thins down towards the drain.

@ $y=0$:



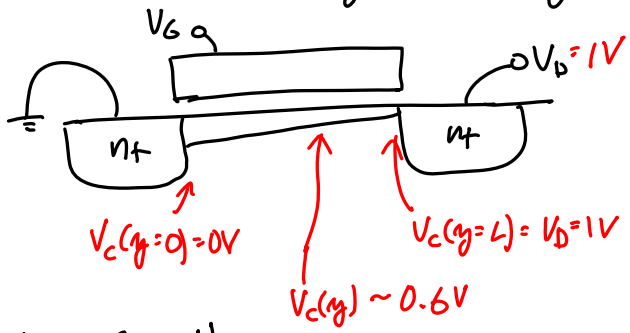
$C_G = \frac{dQ}{dV} = \frac{dQ_n}{dV}$ ← when inversion charge can keep up w/ the frequency

$$Q_n = \int dQ_n = \int_{V_t}^{V_G} C_G\,dV = C_{ox}(V_G - V_t) \quad @\ y=0$$

For y = somewhere between 0 & L:

$$Q_n = \int_{V_t}^{V_G - V_c(y)} C_G\,dV = C_{ox}(V_G - V_c(y) - V_t)$$

channel voltage @ location y



Get Drain Current

⇒ I_D must be the same @ every channel location (by current continuity)

⇒ Thus:

$$I_D = I_D(y) = W Q_n(y) v(y)$$

width; carrier velocity, i.e., velocity of Q_n

$$v(y) = \mu_n \cdot E_y(y) = \mu_n \cdot \frac{dV_c(y)}{dy}$$

mobility; channel directed electric field @ y

$$I_D(y) = W \mu_n Q_n(y)\, dE_y(y)$$
$$= W \mu_n C_{ox}(V_G - V_c(y) - V_t) \cdot \frac{dV_c(y)}{dy}$$

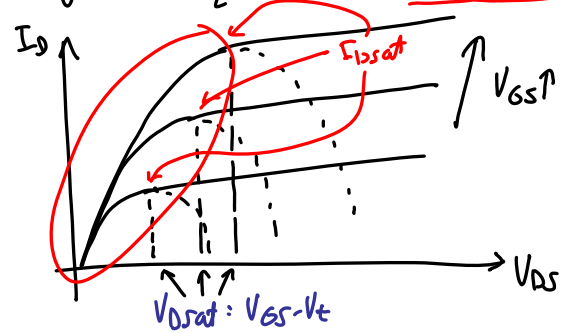
$$\therefore \int_0^L I_D\,dy = W \mu_n C_{ox} \underbrace{\int_{V_s}^{V_D} (V_G - V_c(y) - V_t)\, dV_c(y)}$$

$$(V_G - V_t)V_c - \tfrac{1}{2}V_c^2 \Big|_{V_s}^{V_D}$$
$$= (V_G - V_t)V_D - \tfrac{1}{2}V_D^2 - (V_G - V_t)V_s + \tfrac{1}{2}V_s^2$$
$$= (V_G - V_t)V_{DS} - \tfrac{1}{2}(V_D + V_s)(V_D - V_s)$$
$$= (V_G - V_t - \tfrac{1}{2}V_D - \tfrac{1}{2}V_s)V_{DS}$$
$$= (V_G - V_s - V_t - \tfrac{1}{2}V_D + \tfrac{1}{2}V_s)V_{DS}$$
$$= (V_{GS} - V_t)V_{DS} - \tfrac{1}{2}V_{DS}^2$$

$$I_D L = W \mu_n C_{ox}\left[(V_{GS} - V_t)V_{DS} - \tfrac{1}{2}V_{DS}^2\right]$$

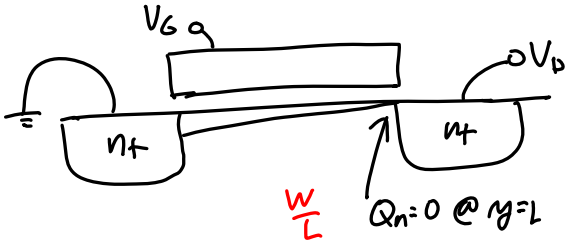
$$\therefore \boxed{I_D = \mu_n C_{ox} \frac{W}{L}\left[(V_{GS} - V_t)V_{DS} - \tfrac{1}{2}V_{DS}^2\right]}$$

↳ This is the classic linear (or triode) region I_D equation for $V_{DS} < V_{GS} - V_t$!



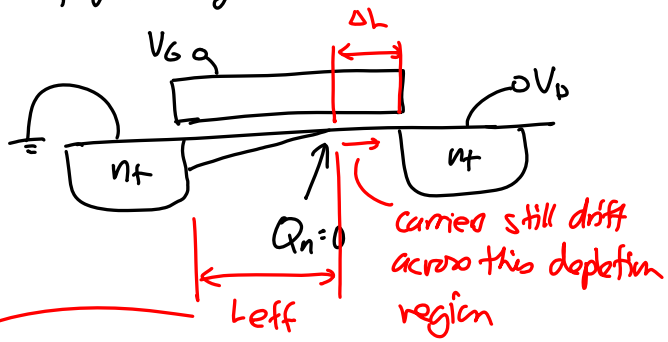
When $V_{DS} = V_{GS} - V_t = V_{Dsat} \rightarrow Q_n = C_{ox}(V_G - V_D - V_t) = 0$

↳ get this:



and $I_D = I_{Dsat} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_t)^2$

When $V_{DS} > V_{GS} - V_t$, then $V_c(y) = V_{GS} - V_t$ @ a y between $y = 0$ & $y = L$: ⇒ get:



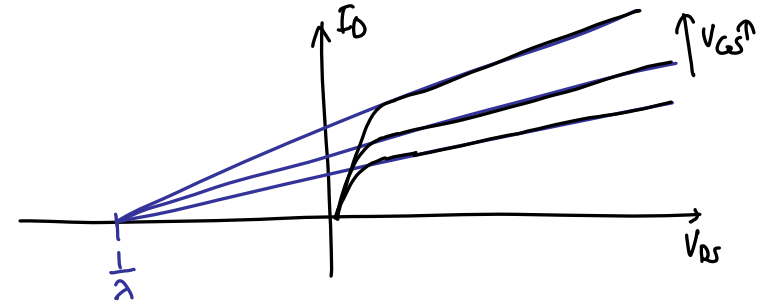
Now, the effective channel length is

$L_{eff} = L - \Delta L \rightarrow I_D$ will rise

For $V_{DS} > V_{GS} - V_t$ (i.e., $V_{DS} > V_{Dsat}$):

$$\boxed{I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_t)^2(1 + \lambda V_{DS})}$$

where λ models the "channel length modulation" phenomenon & can be measured from:



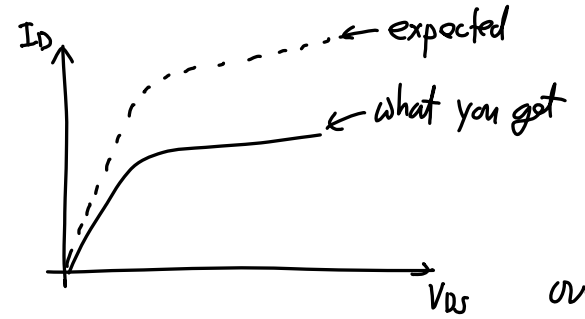
The above models "long-channel" devices.
For more advanced "short-channel" devices, the models must be adjusted.
↳ topics covered in EE 231 & EE 141

↓ move on to talk about:

Device Characterization (i.e., what you're doing in your lab right now)

Potential Problems During Measurement

(a)



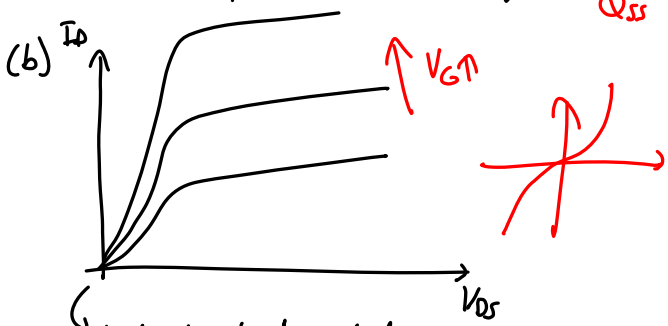
over ⇄

⇒ can be caused by:

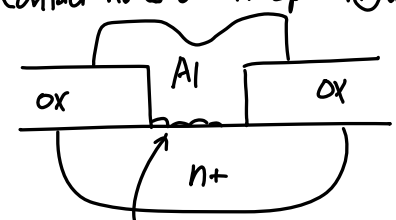
1. Series resistance → check probe contact
2. Excessive surface states → check CV curve & subthreshold slope

Q_{ss}

(b)

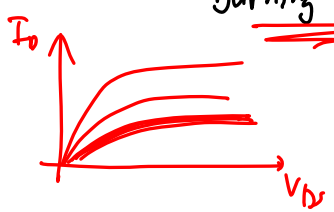


Indicates bad contact resistance.
Contact holes didn't open right:

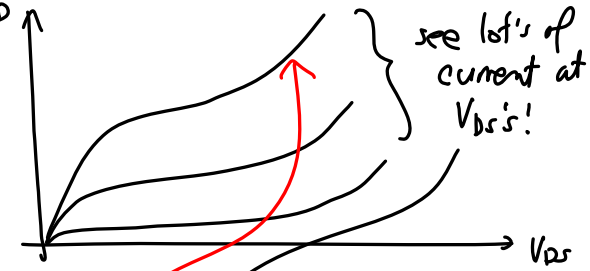


Some residual oxide left.

Can be fixed by sending a lot of current through the contact and "burning through"!

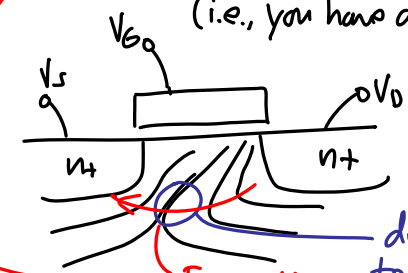


(c)



see lot's of current at higher V_{DS}'s!

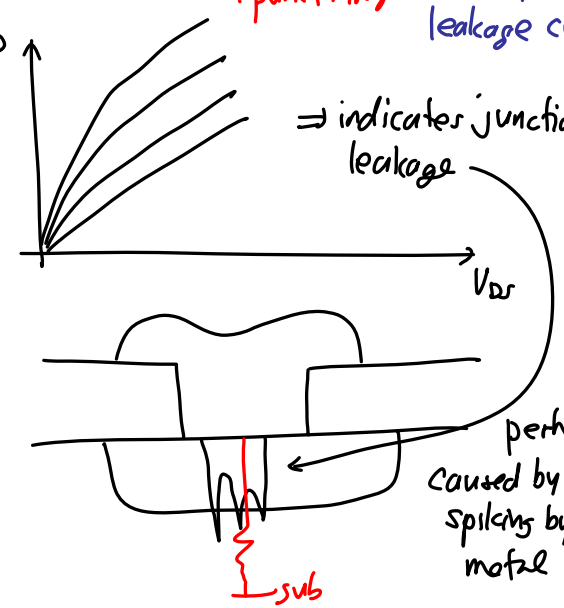
caused by punchthrough (i.e., you have a short-channel device)



$I_{punchthrough}$

depletion regions touch & allow leakage current

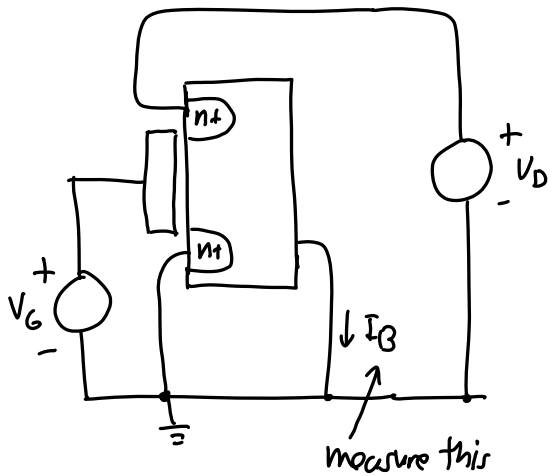
(d)



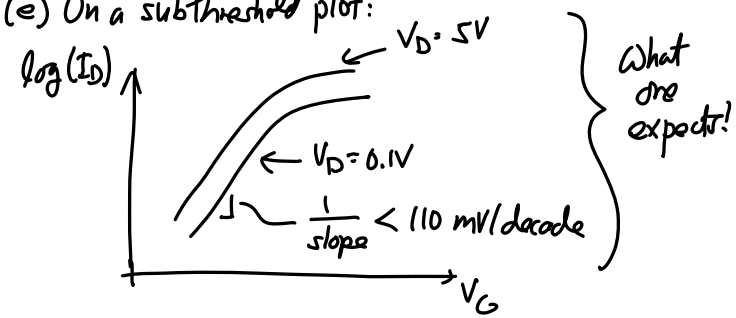
⇒ indicates junction leakage

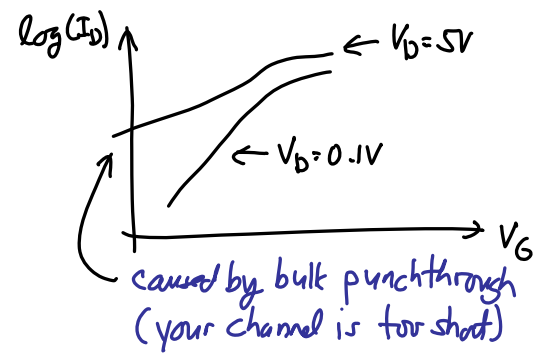
perhaps caused by junction spiking by the metal

To check for this, can hook up as follows:

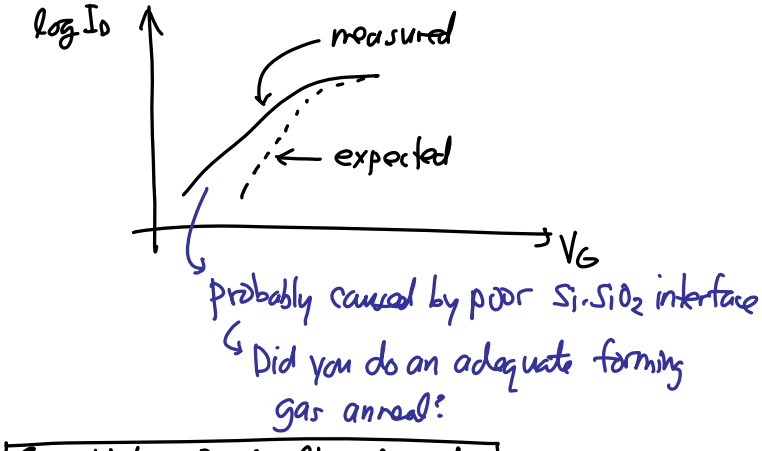


(e) On a subthreshold plot:





What if you get this?

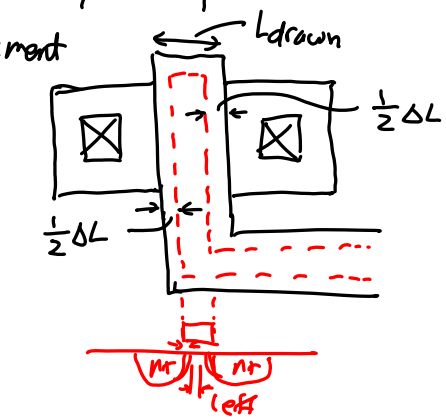


Quantitative Device Characterization

⇒ How can we measure the following:

(a) ΔL ← gate length deviation
(b) ΔW ← gate width deviation
(c) μ ← mobility
(d) γ ← body effect parameter

(a) ΔL measurement



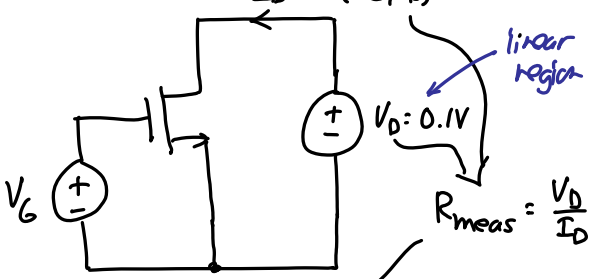
ΔL is due to:

1. Lithography
2. Poly gate etch undercut
3. Lateral diffusion of S/D junctions

$L_{eff} = L_{drawn} - \Delta L$

$\Rightarrow$ How does one measure ΔL?

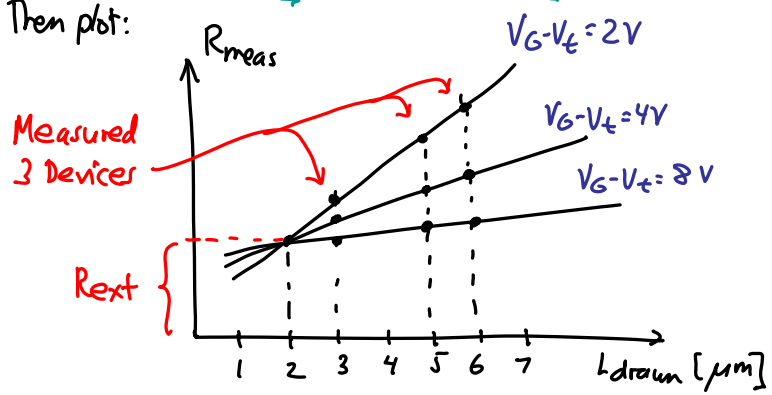
$I_D = f(V_G, V_D)$



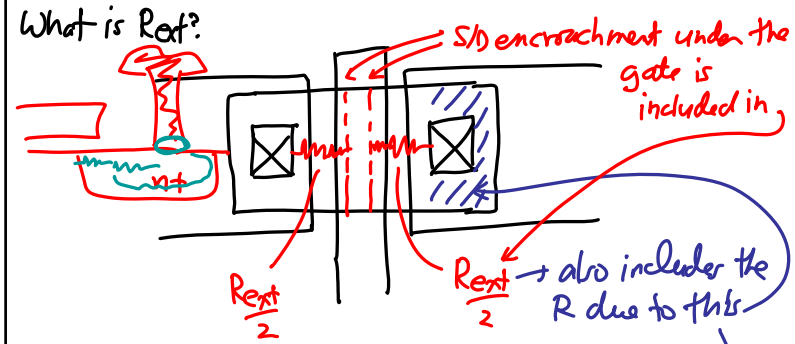
$R_{meas} = \frac{V_D}{I_D}$

Measure R_{meas} for several devices w/ the same W_{drawn} & different L_{drawn}.

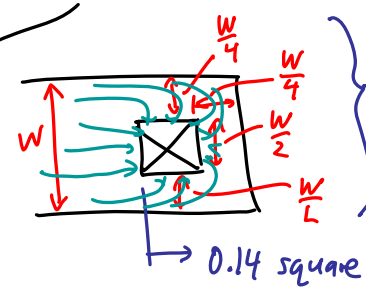
Then plot:



What is R_{ext}?



Need to eliminate R_{ext} from R_{meas} to find the actual MOS channel resistance R_c.



Assume metal R is much smaller than any other

$$I_D = \mu C_{ox} \frac{W}{L_{eff}} \left[(V_{GS} - V_t) V_{DS} - \cancel{\tfrac{1}{2} V_{DS}^2} \right]$$

$[V_{DS} = 0.1V = \text{small}] \Rightarrow I_D \approx \mu C_{ox} \frac{W}{L} (V_{GS} - V_t) V_{DS}$

This includes R_c only; no R_{ext}.

$R_c \triangleq$ channel resistance

$$= \frac{V_{DS}}{I_D} = \frac{\cancel{V_{DS}}}{\mu C_{ox} \frac{W}{L} (V_{GS} - V_t) \cancel{V_{DS}}} = \frac{L_{eff}}{\mu C_{ox} W (V_{GS} - V_t)}$$

$L_{eff} = L_{drawn} - \Delta L$

$$\therefore \boxed{R_c = \frac{L_{drawn} - \Delta L}{\mu W C_{ox}(V_{GS} - V_t)}}$$

and

$$\boxed{R_{meas} = R_c + R_{ext} = \frac{L_{drawn} - \Delta L}{\mu W C_{ox}(V_{GS} - V_t)} + R_{ext}}$$

can solve for ΔL!

(b) ΔW measurement

⇒ pick devices w/ different W_{drawn} but same L_{drawn}

⇒ but if W_{drawn} is changing, then R_{ext} isn't a constant → must suppress R_{ext} in order to use this equation

must use a device w/ a large L_{drawn} (~ 50 μm)

[L_{drawn} = large] ⇒ $R_{meas} = R_c + \cancel{R_{ext}}$

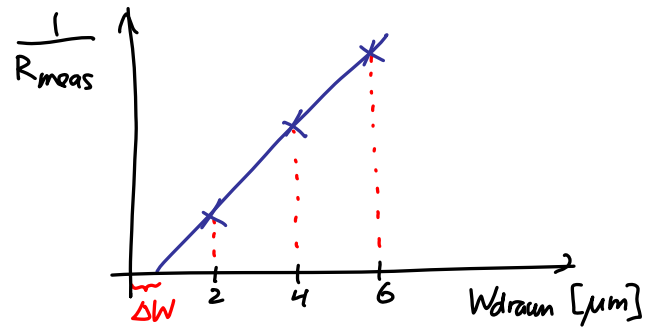
$$\therefore R_{meas} \approx \frac{L_{eff}}{W_{eff}\,\mu C_{ox}(V_{GS} - V_t)} \gg R_{ext}$$

$W_{eff} = W_{drawn} - \Delta W$

$$\frac{1}{R_{meas}} = K(W_{drawn} - \Delta W)$$

$$\Rightarrow \frac{1}{R_{meas}} = 0 \text{ when } W_{drawn} = \Delta W!$$

Thus:



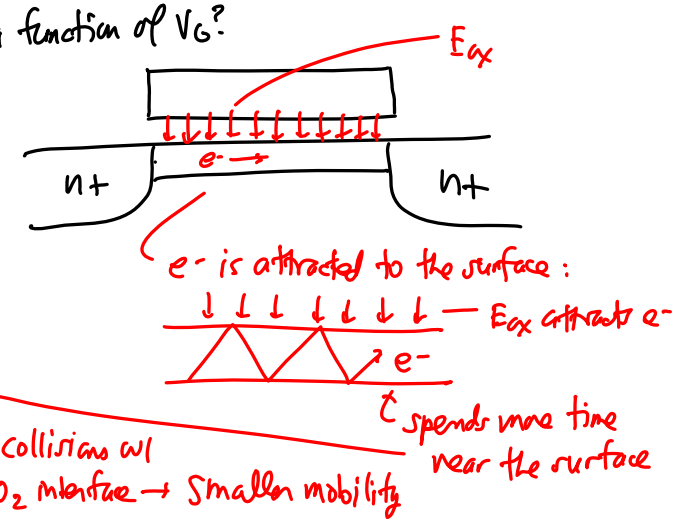
(c) $\mu(V_G)$ → mobility as a function of V_G

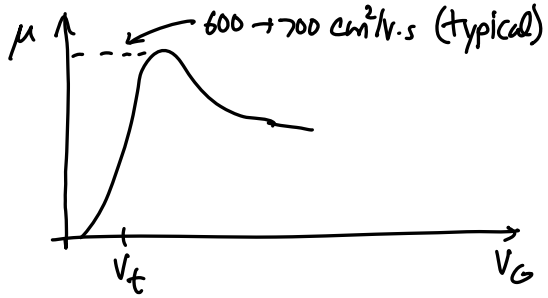
⇒ use large devices (e.g., W = 50 μm, L = 50 μm)

⇒ V_{DS} = small ≈ 0.1V : $I_D = \mu C_{ox} \frac{W}{L}(V_{GS} - V_t)V_{DS}$

$$\mu(V_G) = \frac{I_D}{\frac{W}{L} C_{ox}(V_{GS} - V_t)V_{DS}}$$

Why a function of V_G?

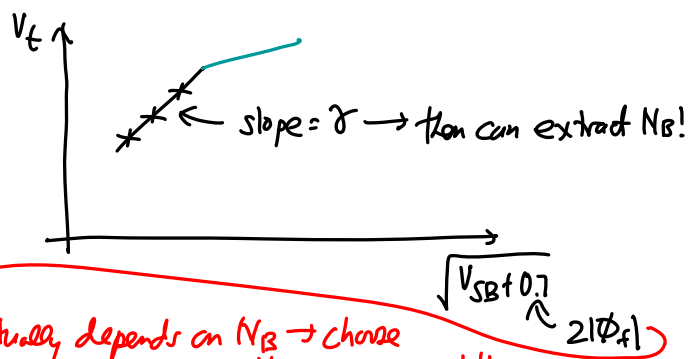




(d) γ ⇒ again, use large devices, e.g., W = 50 μm, L = 50 μm

$$V_t = V_{t0} + \gamma\left(\sqrt{2|\phi_f| + V_{SB}} - \sqrt{2|\phi_f|}\right)$$

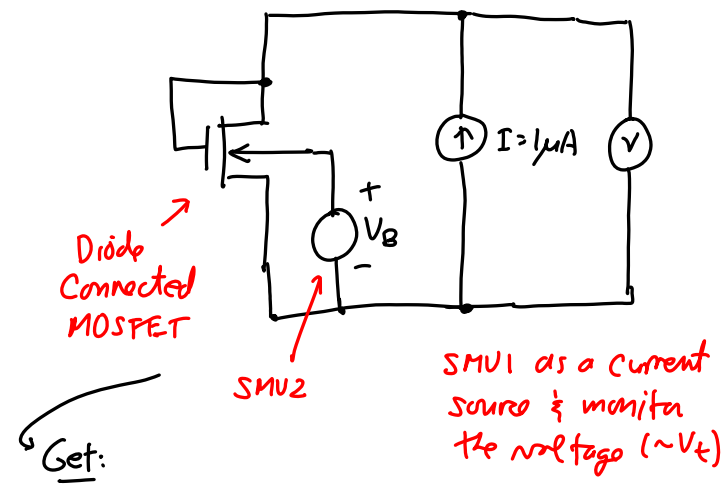
$$\gamma = \frac{\sqrt{2\epsilon_s q N_B}}{C_{ox}}$$ } for uniformly doped channel



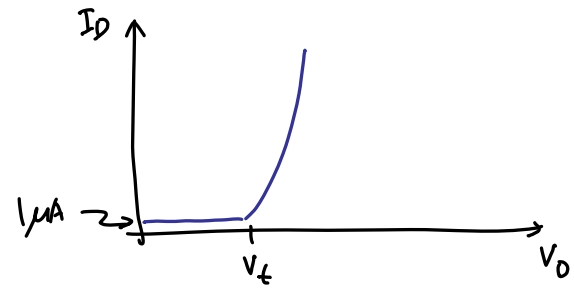
Actually depends on N_B → choose a starting $|\phi_f|$ then iterate w/ N_B till convergence → then get ϕ_f + N_B ...

How does one measure V_t?
↳ several methods ...
(we'll discuss just one ... you'll use another in your lab ...)

Using the 4155 Semiconductor Parameter Analyzer:

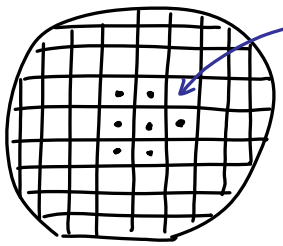


Get:



Yield

- Yield Loss Mechanisms:
  - Clusters: bad process control
    - Alignment out of spec.
    - Nonuniform implantation



when all in the center are dead → usually due to an implantation problem

    - Wafer warpage
    - Thin-film thickness & etching out of control
  - Circuit sensitivity
    - Circuit design problem
    - Need to train the circuit designers
  - Point defects (defect area << chip area)
    - Dust particles
    - Contaminants, e.g., heavy metal, sodium, etc.
    - Pin-holes in PR & in thin-films
    - Crystal defects
    - These are very small defects, but they can still kill your devices
    - Often, only one dead device is all that's needed to kill the entire circuit

Yield Models

Definition. Yield $= Y = \dfrac{\text{No. Functional Dice}}{\text{Total No. of Dice}}$ on the wafer

Mathematically,

$$Y = Y_0 \cdot Y_1$$

where $Y_0 \triangleq$ yield loss due to cluster & circuit design problems

$Y_1 \triangleq$ yield loss in point defects

Useful Quantities:

① Die Size: $s = \sqrt{A_c}$

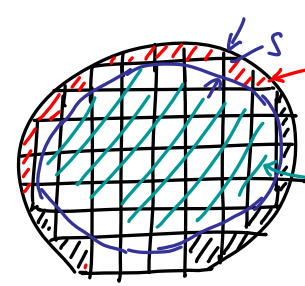
s: die size; A_c: chip area

② # Available Chips in a Wafer $= N_{avail}$, where

$$N_{avail} = \frac{\text{Wafer Area}}{A_c} = \frac{\text{Wafer Area}}{s^2}$$

$$N_{avail} \approx \frac{\pi\left(\frac{d}{2} - s\right)^2}{A_c}$$

d: wafer diameter



③ # good chips/wafer: $N_G = Y \cdot N_{avail}$

④ $D_0 \triangleq$ defect density [no. defects/cm^2]

Poisson Model

→ assumes that a chip is bad if it contains one defect

Ex. $D_0 = 5$ defects/cm^2

$A_c = 1$ cm^2

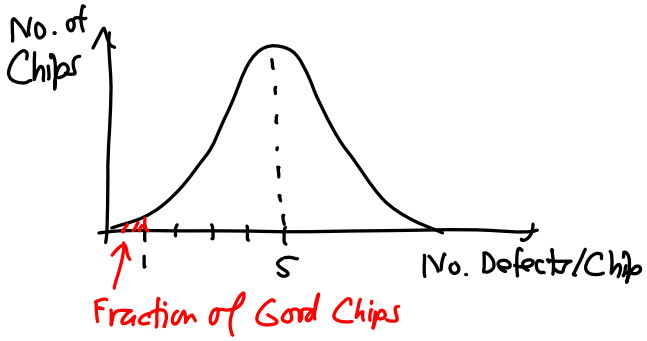
∴ No. of defects/chip $= A_c \times D_0 = 5$

↳ every chip is dead according to this model

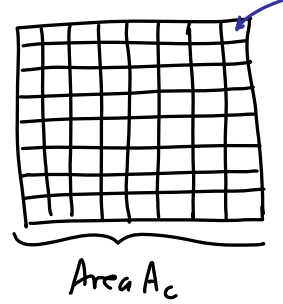
↳ but this actually not true!

↳ $A_c D_0$ is actually an avg. value

↳ thus, we actually have a probability distribution



Split the chip into smaller pieces:

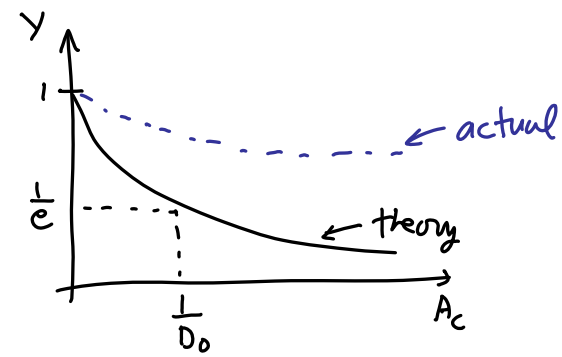


$\delta A_c \rightarrow D_0 \delta A_c \ll 1$

The probability that δA_c does not contain a defect $= 1 - D_0 \delta A_c$.

Probability that A_c is good:

$$\prod_{\text{all } A_c} (1 - D_0 \delta A_c) = \boxed{\exp(-D_0 A_c) = Y}$$



⇒ Thus, in practice the yield is much better than predicted by the Poisson model.

↳ not a very good model!

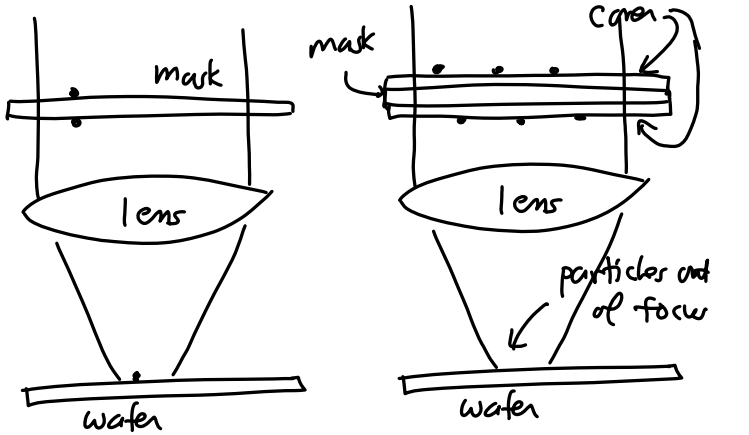
↳ too pessimistic!

- Problems with the Poisson model:
  - No consideration of process induced yield loss, e.g., misalignment
  - No consideration of clusters
  - No consideration of defects in non-critical areas (such as at the edges of a wafer, where the dies are incomplete, and where the highest concentration of defects exists, since the wafers are handled by their edges)

Price Model

$$Y = \frac{1}{(1+D_0A_c)^n}$$, where n = # critical steps (e.g., masking steps)

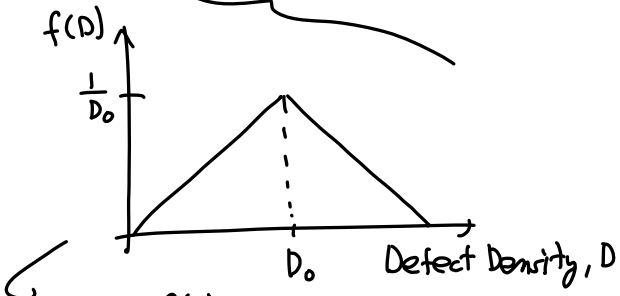
However, w/ stepper lithography, particle defects due to litho no longer an issue, since particles on the mask are defocused:



⇒ this and other similar phenomena invalidate the Price Model

Nonuniform Distribution

$$Y = \int_0^\infty \exp(-DA_c) f(D)\, dD$$



w/ this f(D):

$$Y = \left[\frac{1-\exp(-D_0A_c)}{D_0A_c}\right]^2$$

⇒ other f(D)'s yield other formulations (see Jaeger)

⇒ The ITRS uses:

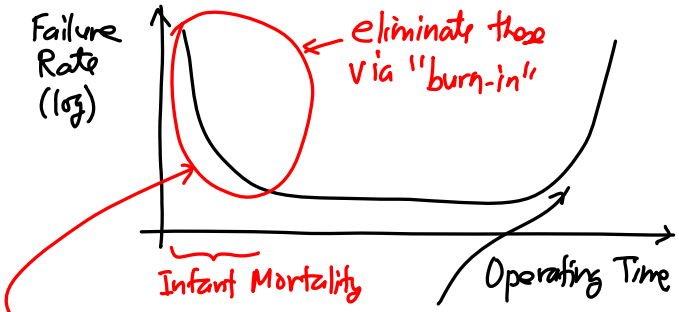
$$Y = \left[1+\frac{D_0A_c}{\alpha}\right]^{-\alpha}$$

where $\alpha \triangleq$ clustering parameter (0.5–10)

ITRS uses "5".

Reliability & Burn-in

⇒ typical failure vs. time plot for MOS ckt.



But why this?

This makes sense, when operated for too long, it fails.

Two Major Mechanisms for Chip Failure:

1. Metal failure (due to bad contact to Si or breakage caused by electromigration)
2. Hot e^- effects → fields are so high that the active area & gate oxide near the drains gets damaged → V_t shifts over time
   ↳ learn in EE 231

To eliminate infant mortality

↳ Burn-In:

1. Power up @ $1.5 \times V_{DD}$.
2. Temperature ↑ (>100°C).
3. Stress (under these conditions) for a long time.

Result: kill unhealthy chips; healthy ones survive!