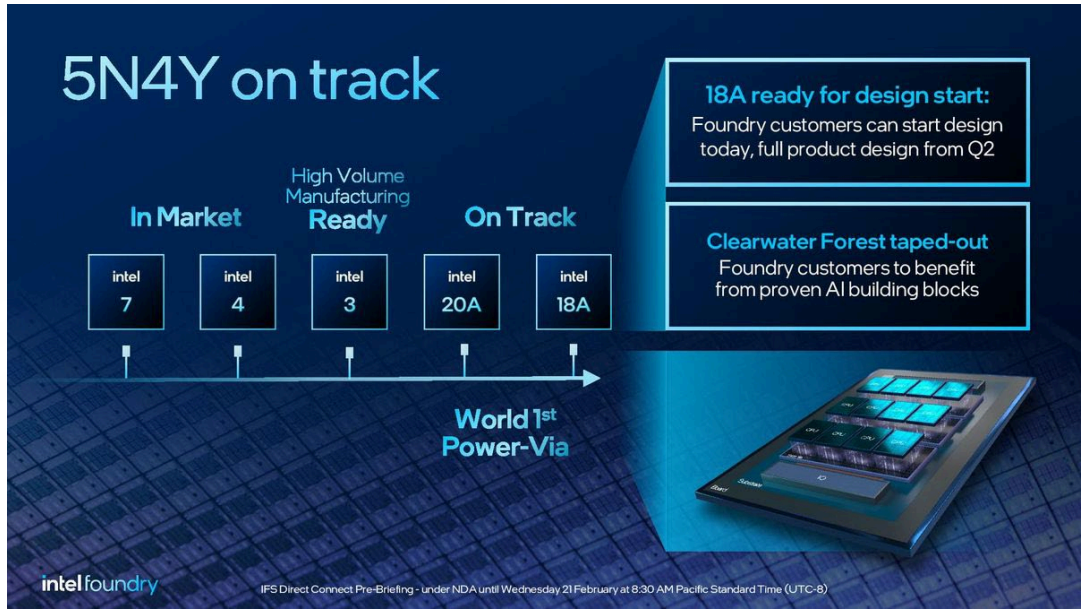
EECS251B : Advanced Digital Circuits and Systems

Lecture 11 – Transistor Models

Borivoje Nikolić

February 21, 2024, TomsHardware: Intel announces new roadmap at IFS Direct Connect 2024: New 14A node, Clearwater Forest taped-in, five nodes in four years remains on track.

Intel unveiled a new roadmap that includes a new 14A node, the industry's first to use High-NA EUV, here at its Intel Foundry Services (IFS) Direct Connect 2024 event. IFS also added new extensions to its existing lineup of process nodes and announced that it is now producing wafers of its Clearwater Forest processors on its 18A node.



https://www.tomshardware.com/pc-components/cpus/intel-announces-new-roadmap-at-ifs-direct-connect-2024-new-14a-node-clearwater-forest-taped-in-five-nodes-in-four-years-remains-on-track

Announcements

- Lab 4b slightly delayed
- Homework 2 due nextweek

MOS Transistor and Gate Delay Models

Modeling Goals

- Models that traverse design hierarchy
- Start with transistor models
- Gate delay models
- Use models to time the design
- Modeling variability

- Based on 251A, approach
  - Start simple
  - Increase accuracy, when needed

Device Models

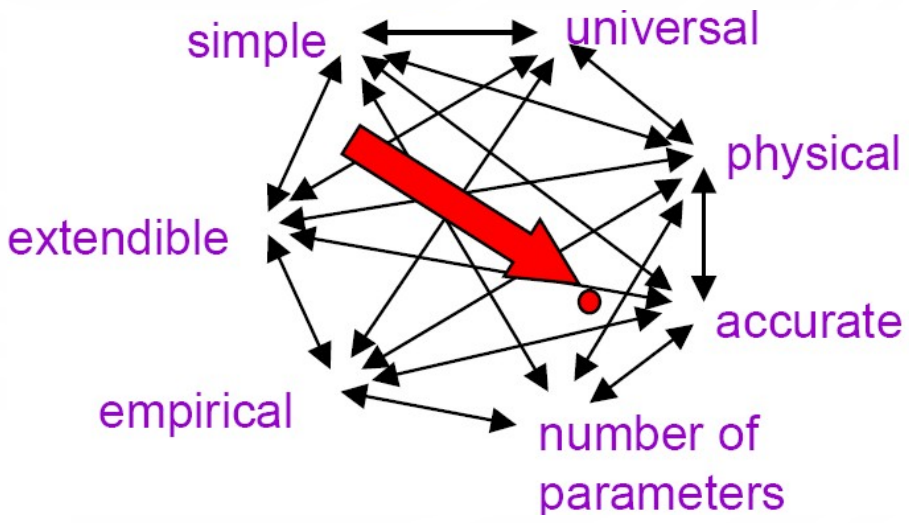
- Transistor models
  - I-V characteristics
  - C-V characteristics
- Interconnect models
  - R, C, L
  - Covered in EE240A

Transistor Modeling

- Different levels:
  - Hand analysis
  - Computer-aided analysis (e.g. Matlab, Python, Excel,…)
  - Switch-level simulation (some flavors of ‘fast Spice’)
  - Circuit simulation (Hspice)
- These levels have different requirements in complexity, accuracy and speed of computation
- We are primarily interested in delay and energy modeling, rather than current modeling
- But we have to start from the currents…

Transistor Modeling

- DC
  - Accurate I-V equations
  - Well behaved conductance for convergence (not necessarily accurate)
- Transient
  - Accurate I-V and Q-V equations
  - Accurate first derivatives for convergence
  - Conductance, as in DC
- Physical vs. empirical



from BSIM group

Transistor I-V Modeling

- BSIM
  - Superthreshold and subthreshold models
  - Need smoothening between two regions
- EKV/PSP
  - One continuous model based on channel surface potential

Long-Channel MOS On-Current

MOS I-V (BSIM)

Start with the basics:

$$I_{DS} = W C_{ox}(V_{GS} - V_{Th} - V_C(x))\ \mu E(x)$$

v: velocity at x

Q'(x): Charge density in channel at x

MOS I-V (BSIM)

Start with the basics:

$$I_{DS} = \mathrm{W} C_{ox}(V_{GS} - V_{Th} - V_C(x))\ \mu E(x)$$

$$I_{DS} = \mathrm{W} C_{ox}(V_{GS} - V_{Th} - V_C(x))\ \mu(\mathrm{d}V_C(x)/\mathrm{dx})$$

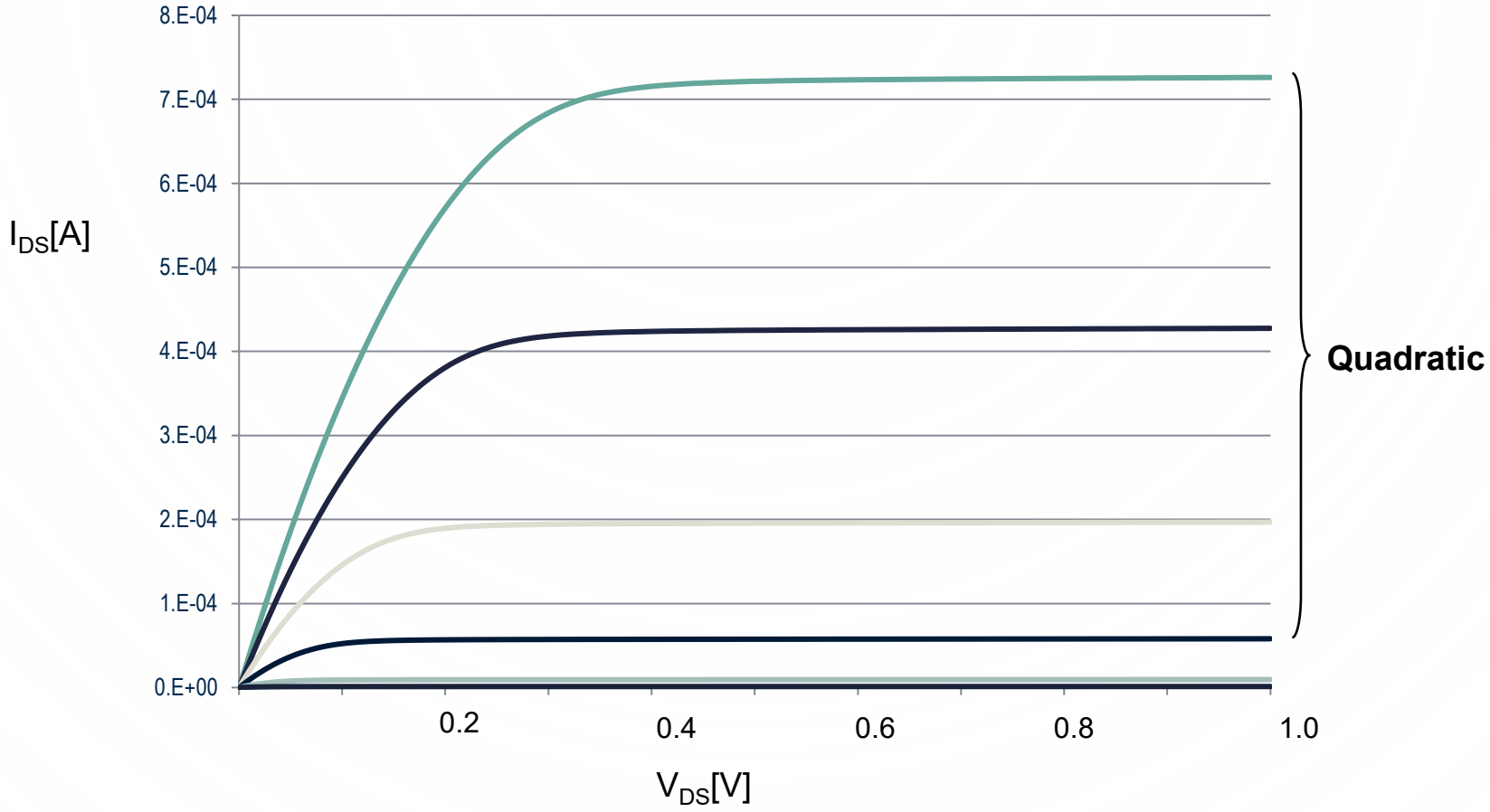
- When integrated over the channel:

$$I_{DS} = \frac{W}{L}\mu C_{ox}\left(V_{GS} - V_{Th} - \frac{V_{DS}}{2}\right)V_{DS}$$

Transistor saturates when $V_{GD} = V_{Th}$, - the channel pinches off at drain's side.

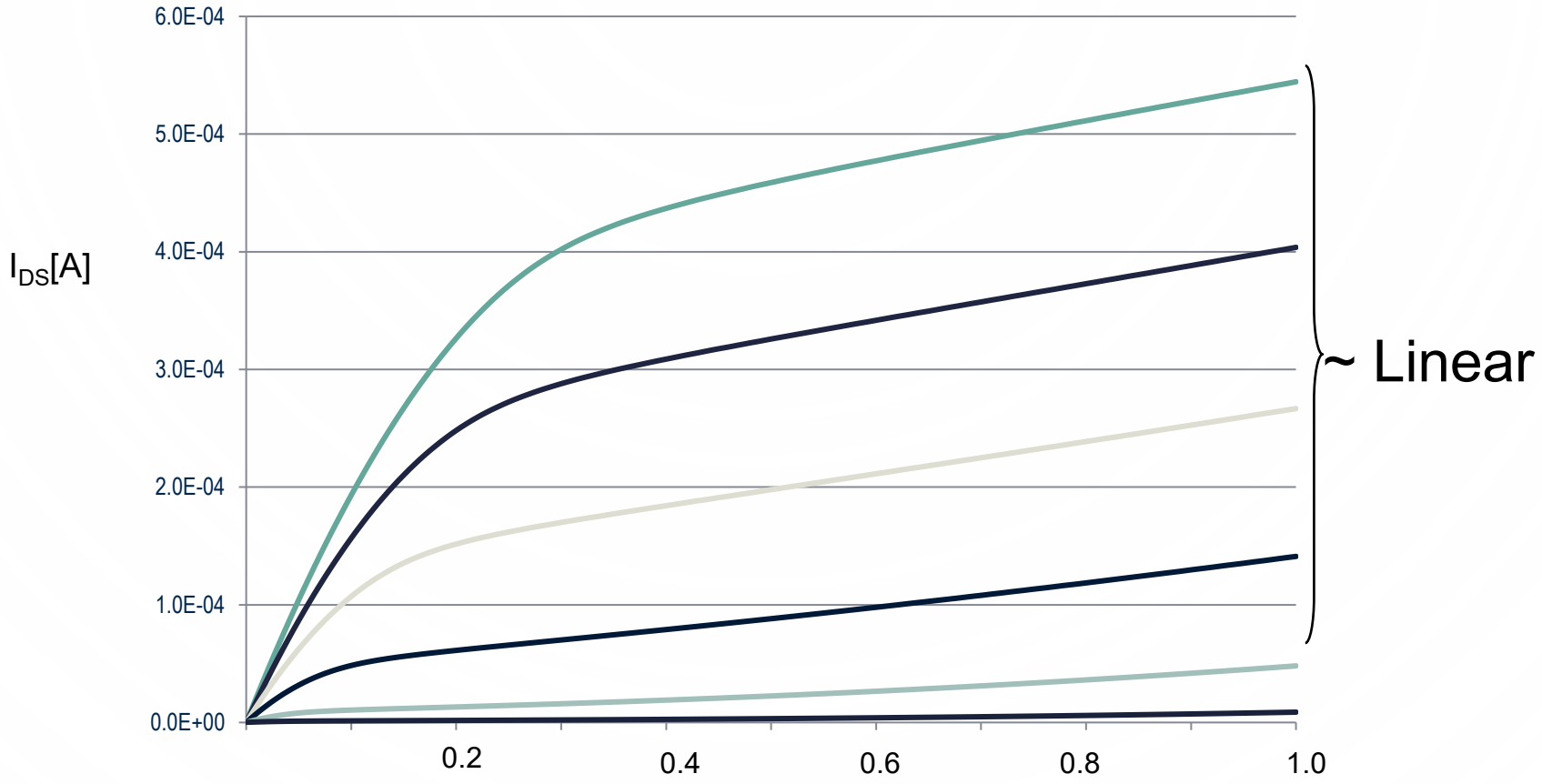
$$I_{DS} = \frac{W}{2L}\mu C_{ox}(V_{GS} - V_{Th})^2$$

MOS Currents (32nm CMOS with L>>1µm)



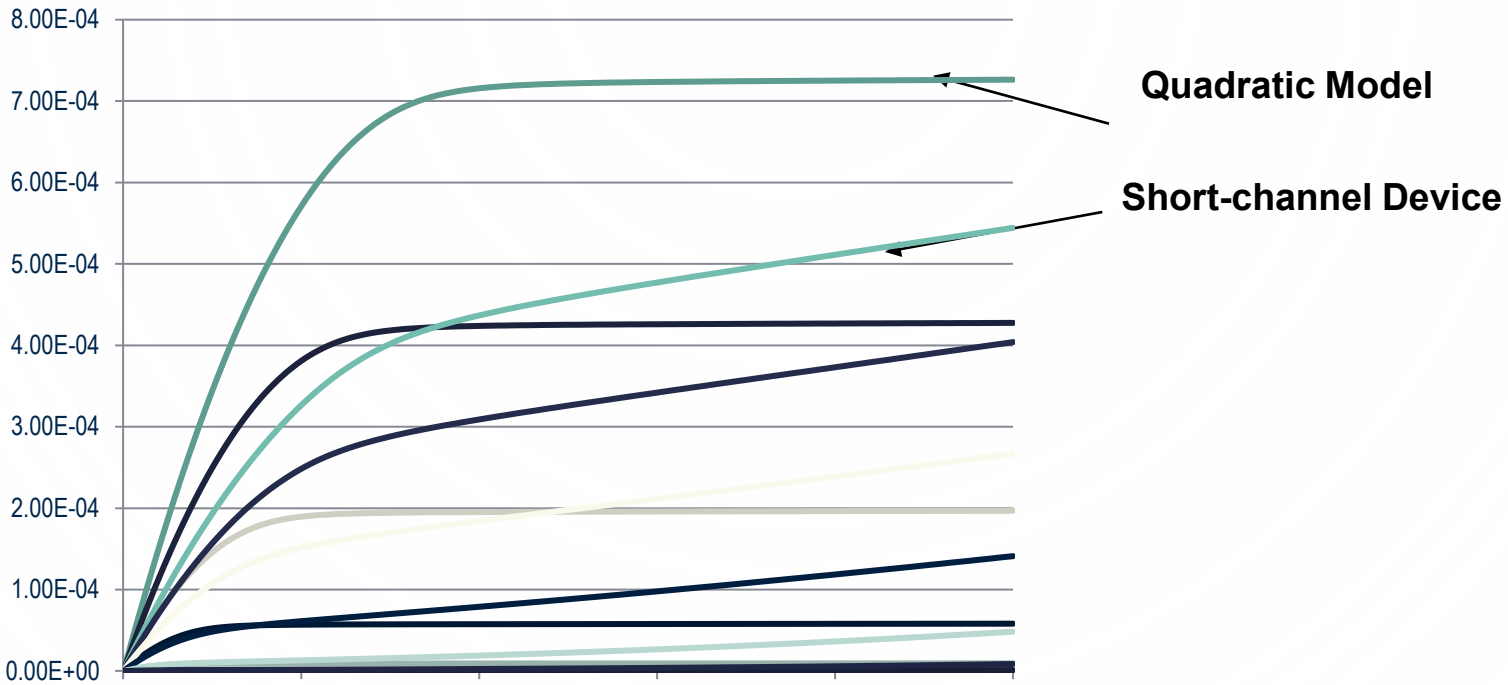
Currents according to the quadratic model
Correct for long channel devices (L ~ µm)

Simulated 32nm Transistor



L = 32nm

Simulation vs. Model



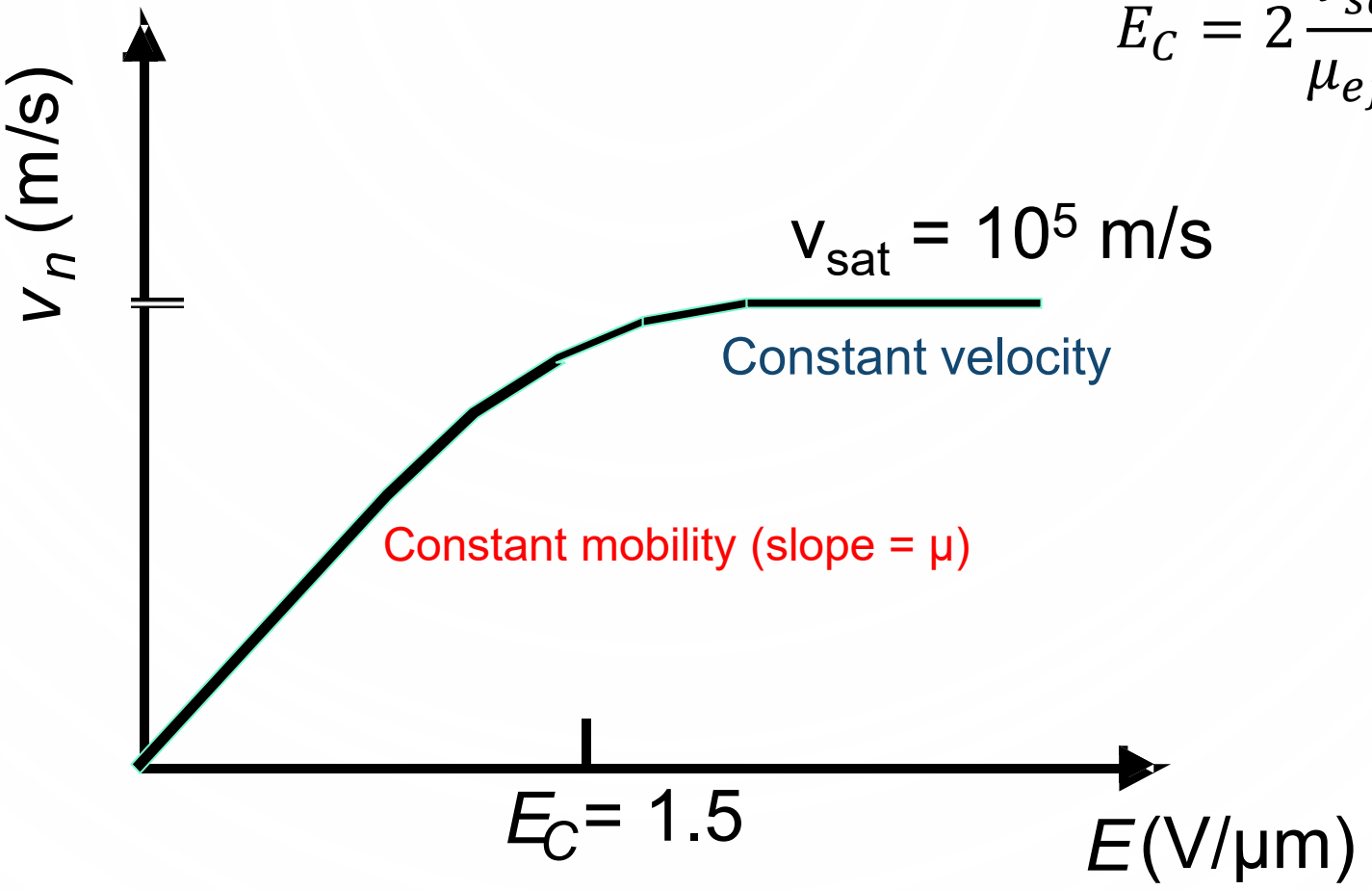
Major discrepancies:

- shape
- saturation points
- output resistances

Velocity Saturation

Velocity Saturation

$$E_C = 2\frac{v_{sat}}{\mu_{eff}}$$

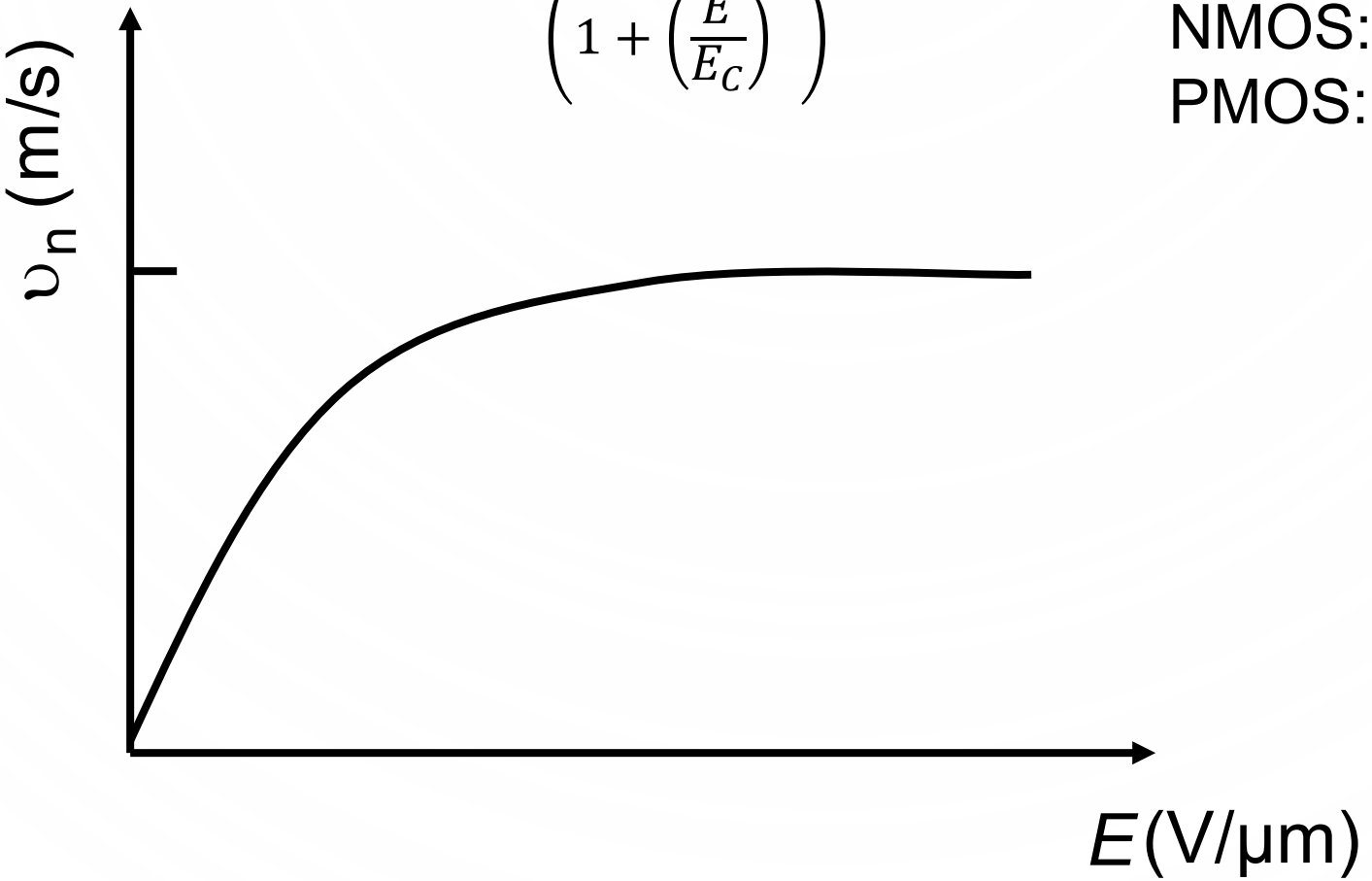


Modeling Velocity Saturation

- Fit the velocity-dependence curve

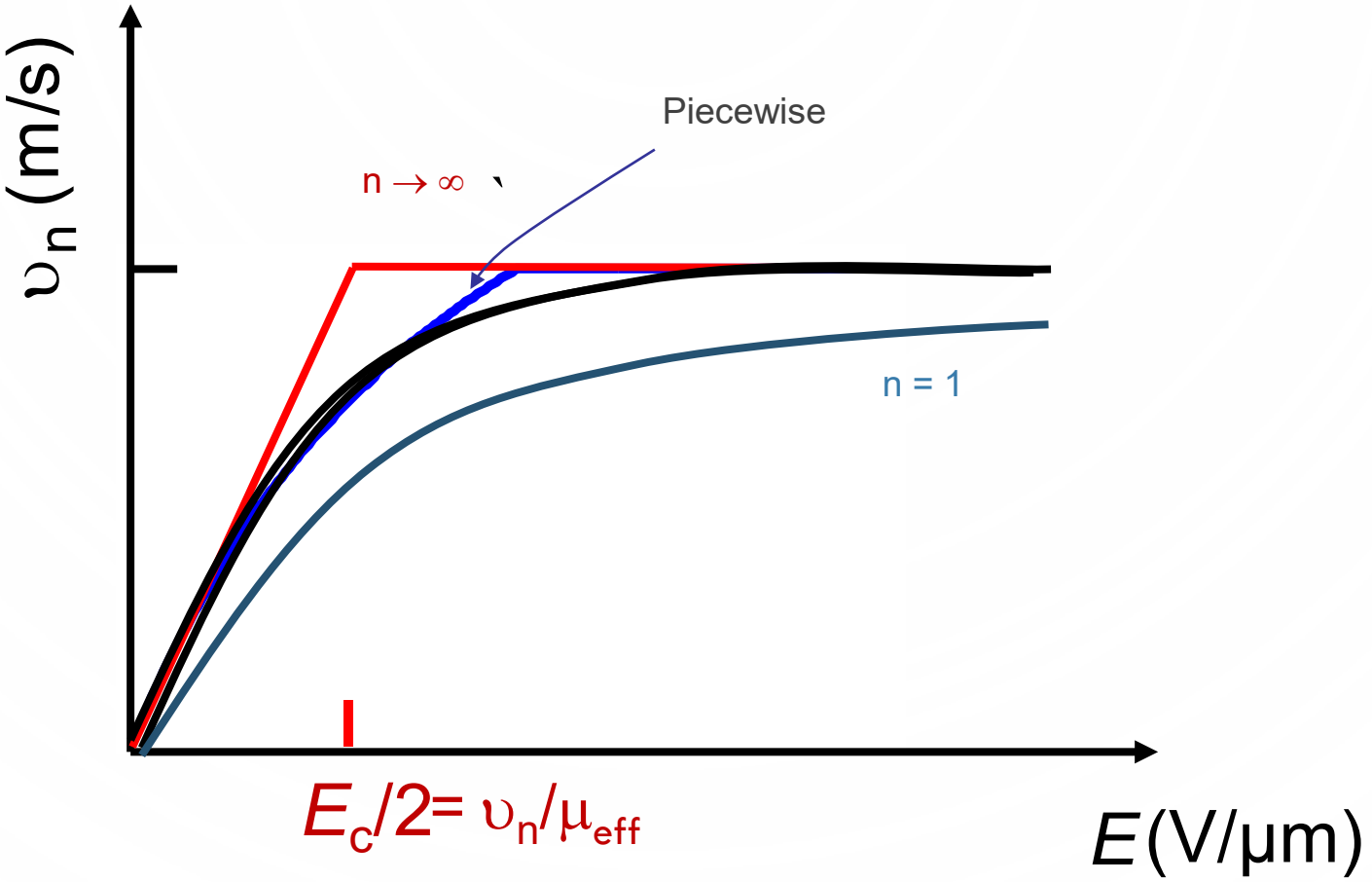
$$v = \frac{\mu_{eff} E}{\left(1 + \left(\frac{E}{E_C}\right)^n\right)^{1/n}}$$

NMOS: $n = 2$
PMOS: $n = 1$



Modeling Velocity Saturation

- A few approximations: (a) $n \to \infty$, (b) $n = 1$, (c) piecewise



Short-Channel MOS On-Current

Approximation $n \to \infty$

1) $v = \mu_{eff} E, \quad E < E_C$

$$I_{DS} = \mu C_{ox} \frac{W}{L} \left((V_{GS} - V_{Th}) V_{DS} - \frac{{V_{DS}}^2}{2} \right)$$

2) $v = v_{sat}, \quad E > E_c$

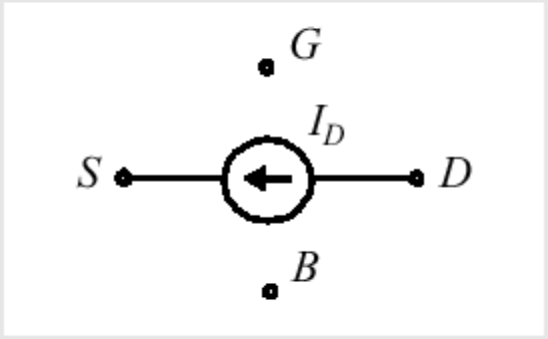
$$I_{Dsat} = \mu C_{ox} \frac{W}{L} \left((V_{GS} - V_{Th}) V_{Dsat} - \frac{{V_{Dsat}}^2}{2} \right)$$

$V_{Dsat} = ?$

Can be reduced to Rabaey DIC model by making V_{Dsat} = const

Is this physically justified?

MOS Model from DIC, 2nd ed.



$$I_D = 0 \text{ for } V_{GT} \le 0$$

$$I_D = k'\frac{W}{L}\left(V_{GT}V_{min} - \frac{V_{min}^2}{2}\right)(1 + \lambda V_{DS}) \text{ for } V_{GT} \ge 0$$

with $V_{min} = \min(V_{GT}, V_{DS}, V_{DSAT})$,

$V_{GT} = V_{GS} - V_T$,

and $V_T = V_{T0} + \gamma(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|})$

γ - body effect parameter

From Rabaey, 2nd ed.

Unified MOS Model

- Model presented is compact and suitable for hand analysis.
- Still have to keep in mind the main approximation: that V_{DSat} is constant . When is it going to cause largest errors?
  - When does *E* scale? – Transistor stacks.
- But the model still works fairly well.
  - Except for stacks

Approximation $n = 1$, piecewise

- $n = 1$ is solvable, piecewise closely approximates

Velocity, *v*:

$$v = \begin{cases} \dfrac{\mu_{eff} E}{1 + E/E_C}, & E < E_C = \dfrac{2 v_{sat}}{\mu_{eff}} \\ v_{sat}, & E > E_C \end{cases}$$

Sodini, Ko, Moll, TED'84
Toh, Ko, Meyer, JSSC'88
BSIM model

Drain Current

- We can find the drain current by integrating I_{DS}

$$I_{DS} = \mathrm{W}\, C_{ox}(\, V_{GS} - \, V_{Th} - \, V_C(x))\; v$$

Linear:

$$I_{DS} = \frac{\mu C_{ox}}{1 + (V_{DS}/E_C L)} \frac{W}{L} \left((V_{GS} - V_{Th}) V_{DS} - \frac{{V_{DS}}^2}{2} \right)$$

In saturation:

$$I_{DSat} = C_{ox} W v_{sat} (V_{GS} - V_{Th} - V_{Dsat})$$

$$I_{Dsat} = \frac{\mu C_{ox}}{1 + (V_{Dsat}/E_C L)} \frac{W}{L} \left((V_{GS} - V_{Th}) V_{Dsat} - \frac{{V_{Dsat}}^2}{2} \right)$$

Drain Current in Velocity Saturation

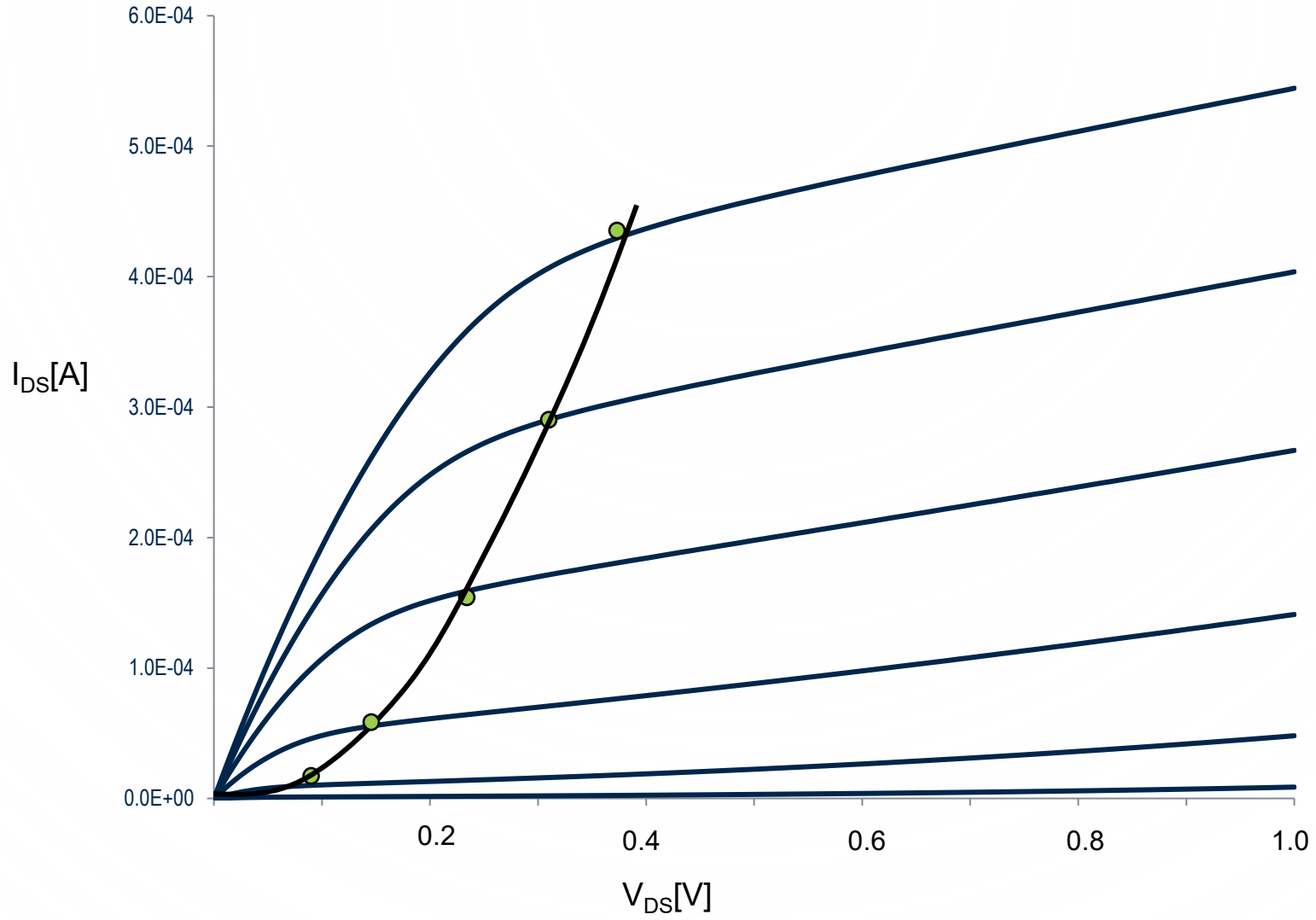
- Solving for V_{Dsat}

$$V_{DSat} = \frac{(V_{GS} - V_{Th})E_C L}{(V_{GS} - V_{Th}) + E_C L}$$

- And saturation current

$$I_{DSat} = \frac{W}{L}\frac{\mu_{eff} C_{ox} E_C L}{2}\frac{(V_{GS} - V_{Th})^2}{(V_{GS} - V_{Th}) + E_C L}$$

Velocity Saturation



Velocity Saturation

- $E_C L$ is V_{GS}-dependent
- Can calculate V_{DSat} (V_{Th} ~ 0.4V in 28nm)

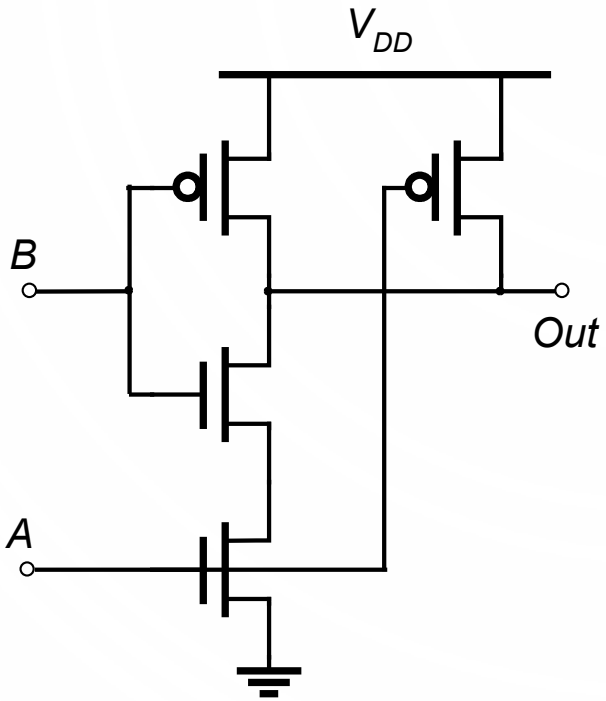
V_{GS} [V]	0.5	0.6	0.7	0.8	0.9	1.0
V_{DSat} [V]	0	0.05	0.11	0.18	0.25	0.33

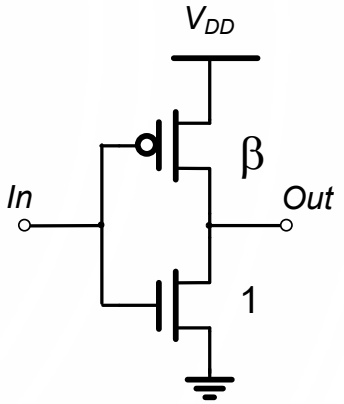
- For $V_{GS} - V_{Th} << E_C L$, V_{DSat} is close to $V_{GS} - V_{Th}$
- For large V_{GS}, V_{DSat} bends upwards toward $E_C L$
- Therefore, $E_C L$ can be sometimes approximated with a constant term
  - But also need to understand the limitation of the approximation

Application of I-V Models

Application of Models: NAND Gate

- 2-input NAND gate



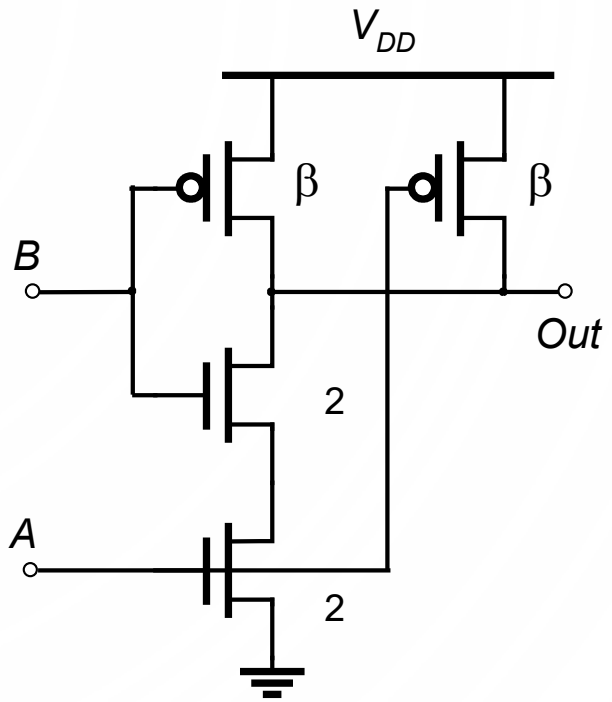


Sizing for equal transistions:

- P/N ratio (β-ratio): 1 for L< 20nm, 1.6 for 20nm <L< ~65nm, 2 for L > 90nm
- Upsizing stacks by a factor proportional to the stack height

Transistor Stacks

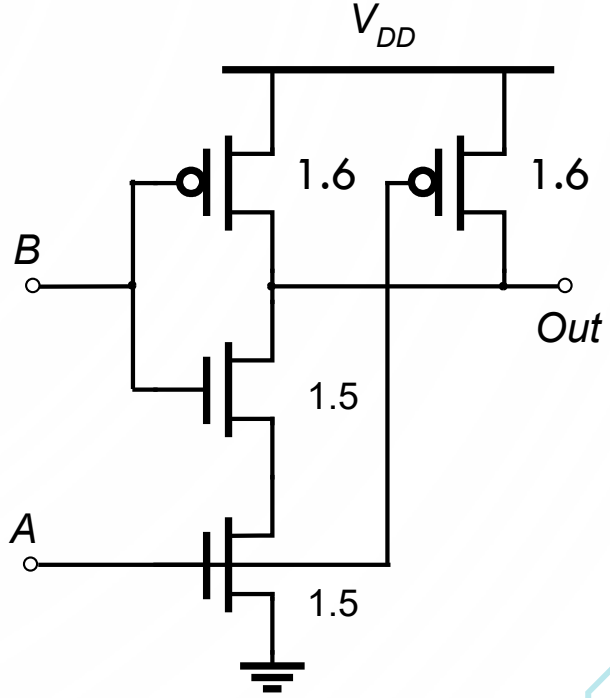
- With transistor stacks, V_{DS}, V_{GS} reduce.
- Unified model assumes V_{DSat} = const.
- For a stack of two, appears that both have exactly double R_{ekv} of an inverter with the same width
- Therefore, doubling the size of each, should make the pull-down R equivalent to an inverter



Velocity Saturation

- As $(V_{GS}-V_{Th})/E_CL$ changes, the depth of saturation changes

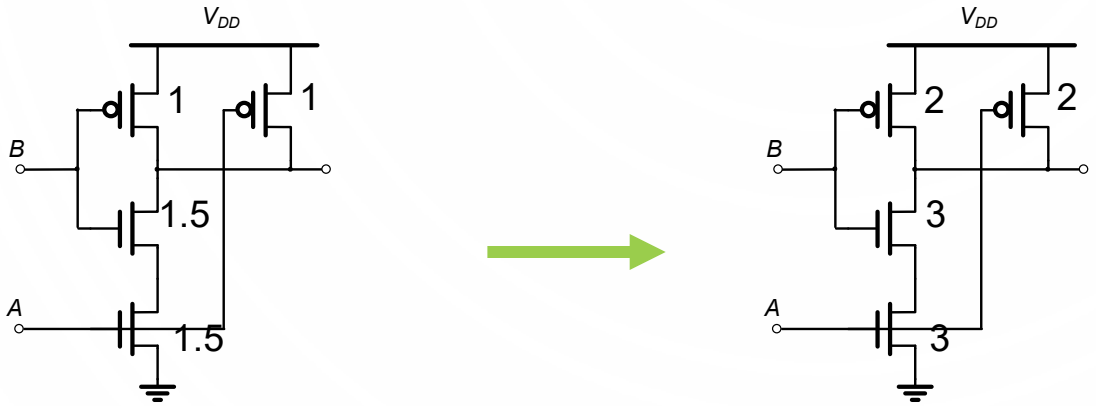
$$I_{DSat} = \frac{W}{L}\frac{\mu_{eff}C_{ox}E_CL}{2}\frac{(V_{GS}-V_{Th})^2}{(V_{GS}-V_{Th})+E_CL}$$



- For V_{GS}, V_{DS} = 1.0V, E_CL is ~0.75V
- With double length, E_CL is 1.5V (in this model in 28nm)
- Stacked transistors are less saturated
- V_{GS}-V_{Th} = 0.6V, I_{DSat} ~ 2/3 of inverter I_{DSat} (64%)
- Therefore NAND2 should have pull-down sized 1.5X
- Check any library NAND2's
- Current halved in a stack of 3
- PMOS less impacted by velocity saturation since lower mobility

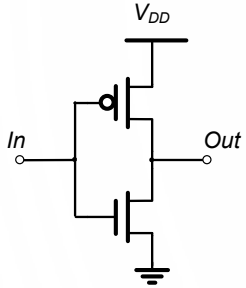
Note about FinFETs

- Widths are quantized

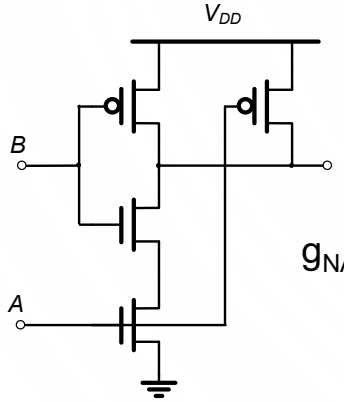


Example: Logical Effort

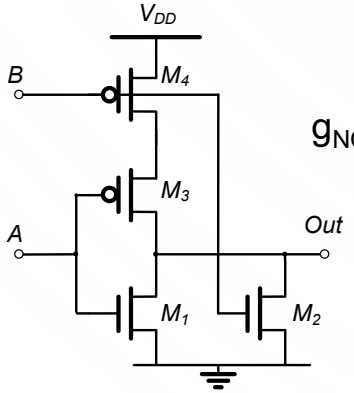
- Older CMOS (130nm)



$g_{inv} = 1$

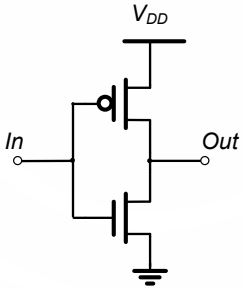


$g_{NAND2} =$

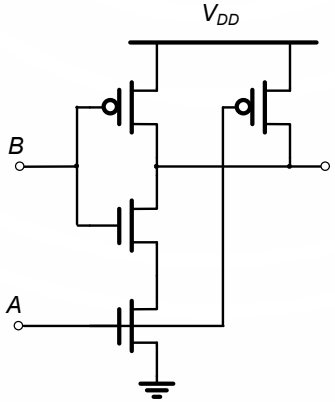


$g_{NOR2} =$

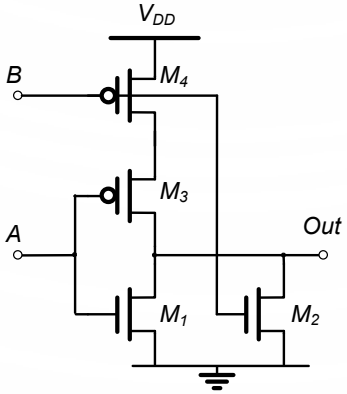
- Planar CMOS (~28nm, bulk, FDSOI)



$g_{inv} = 1$

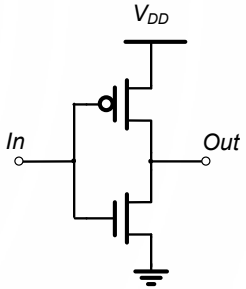


$g_{NAND2} =$

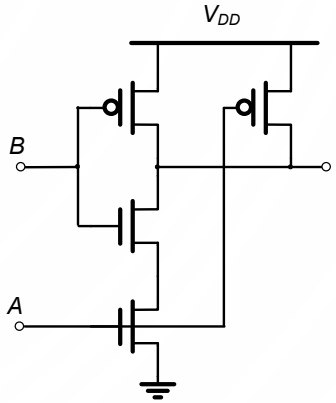


$g_{NOR2} =$

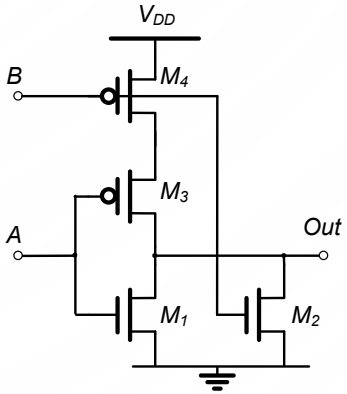
- FinFET (7nm)



$g_{inv} = 1$



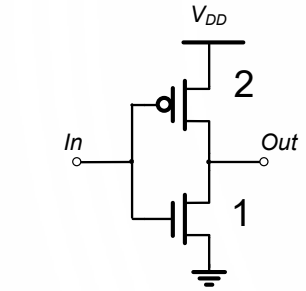
$g_{NAND2} =$



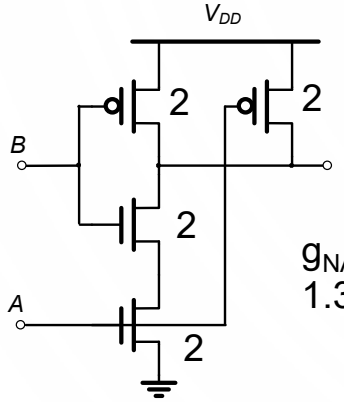
$g_{NOR2} =$

Example: Logical Effort

- Older CMOS (130nm)

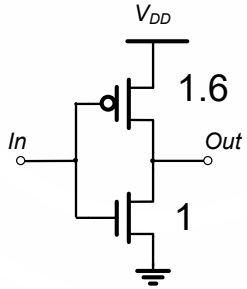


$g_{inv} = 1$

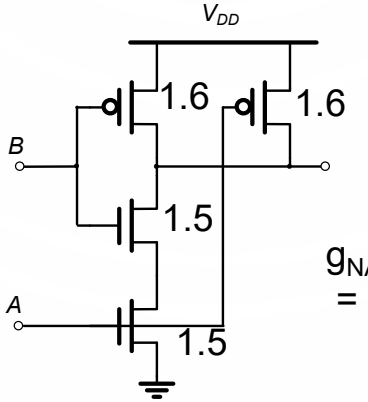


$g_{NAND2} = 4/3 = 1.33$

- Planar CMOS (~28nm, bulk, FDSOI)

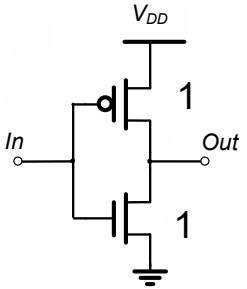


$g_{inv} = 1$

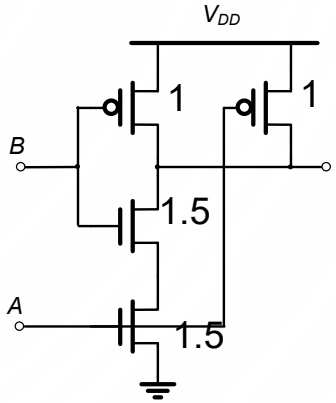


$g_{NAND2} = 3.1/2.6 = 1.2$

- FinFET (7nm)



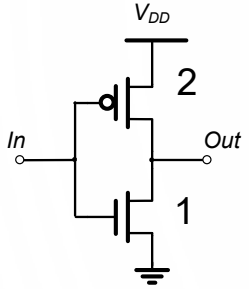
$g_{inv} = 1$



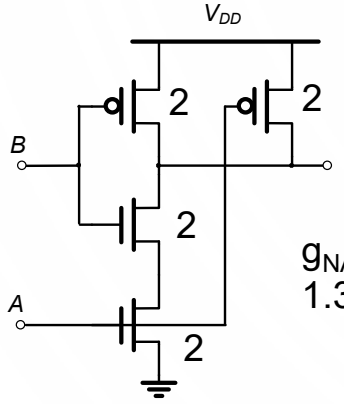
$g_{NAND2} = 2.5/2 = 1.25$

Example: Logical Effort

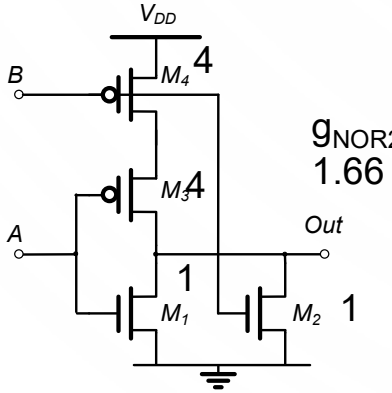
- Older CMOS (130nm)



$g_{inv} = 1$

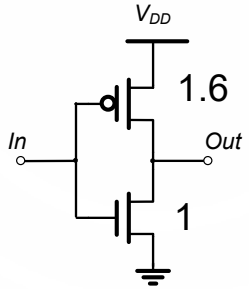


$g_{NAND2} = 4/3 = 1.33$

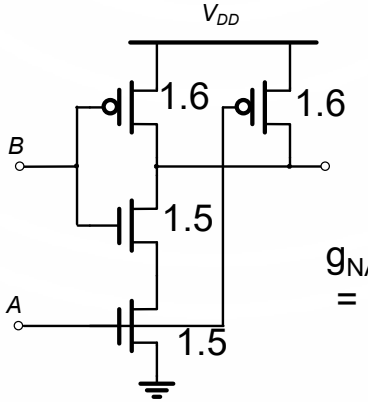


$g_{NOR2} = 5/3 = 1.66$

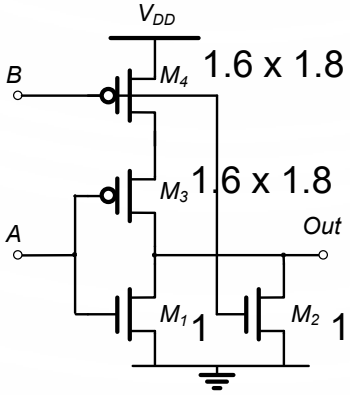
- Planar CMOS (~28nm, bulk, FDSOI)



$g_{inv} = 1$

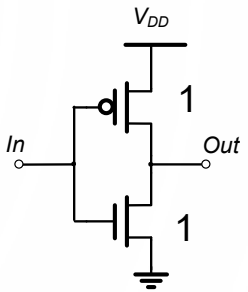


$g_{NAND2} = 3.1/2.6 = 1.2$

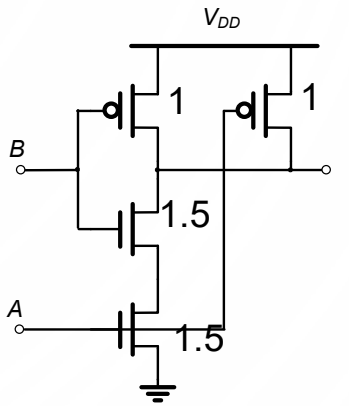


$g_{NOR2} = 3.9/2.6 = 1.5$

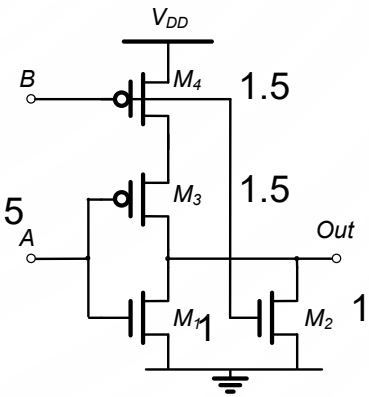
- FinFET (7nm)



$g_{inv} = 1$



$g_{NAND2} = 2.5/2 = 1.25$



$g_{NOR2} = 2.5/2 = 1.25$

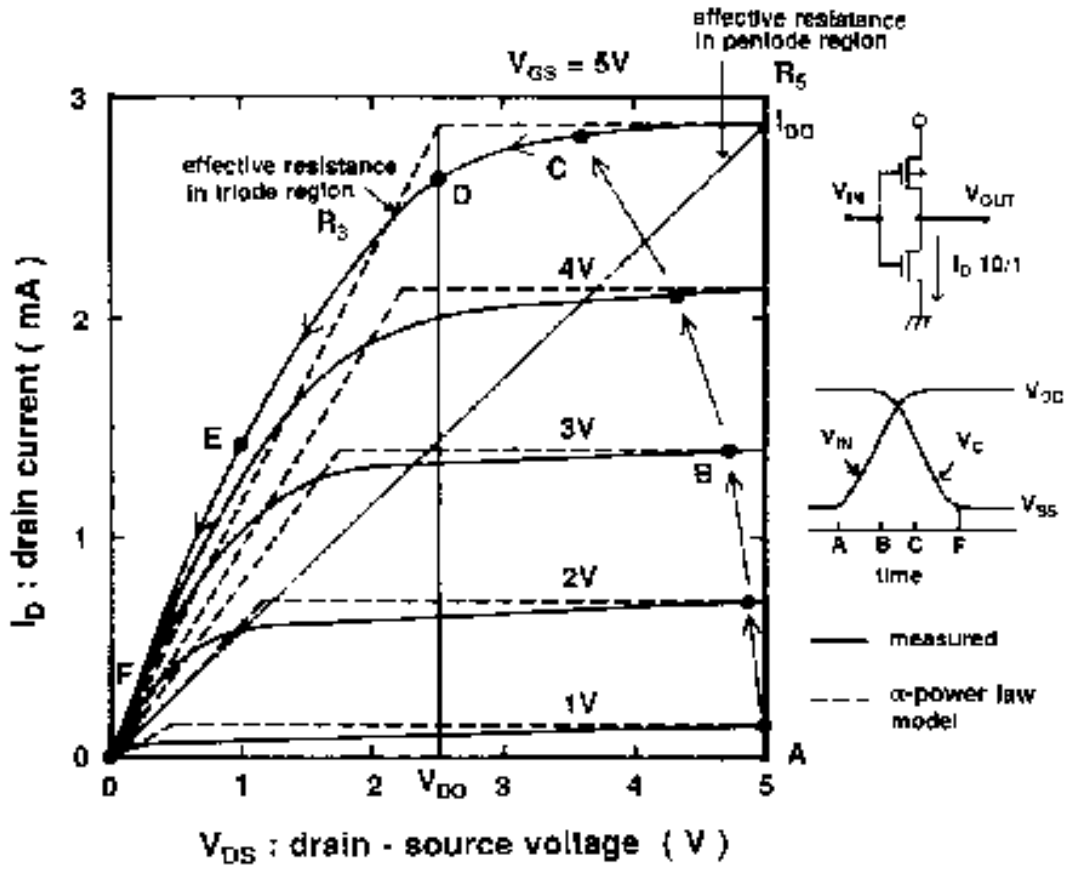
Other Velocity Saturation Models

Other Models: Alpha Power Law Model

- Simple model, sometimes useful for hand analysis

$$I_{DS} = \frac{W}{2L}\mu C_{ox}\left(V_{GS} - V_{Th}\right)^{\alpha}$$

Parameter a is between 1 and 2.



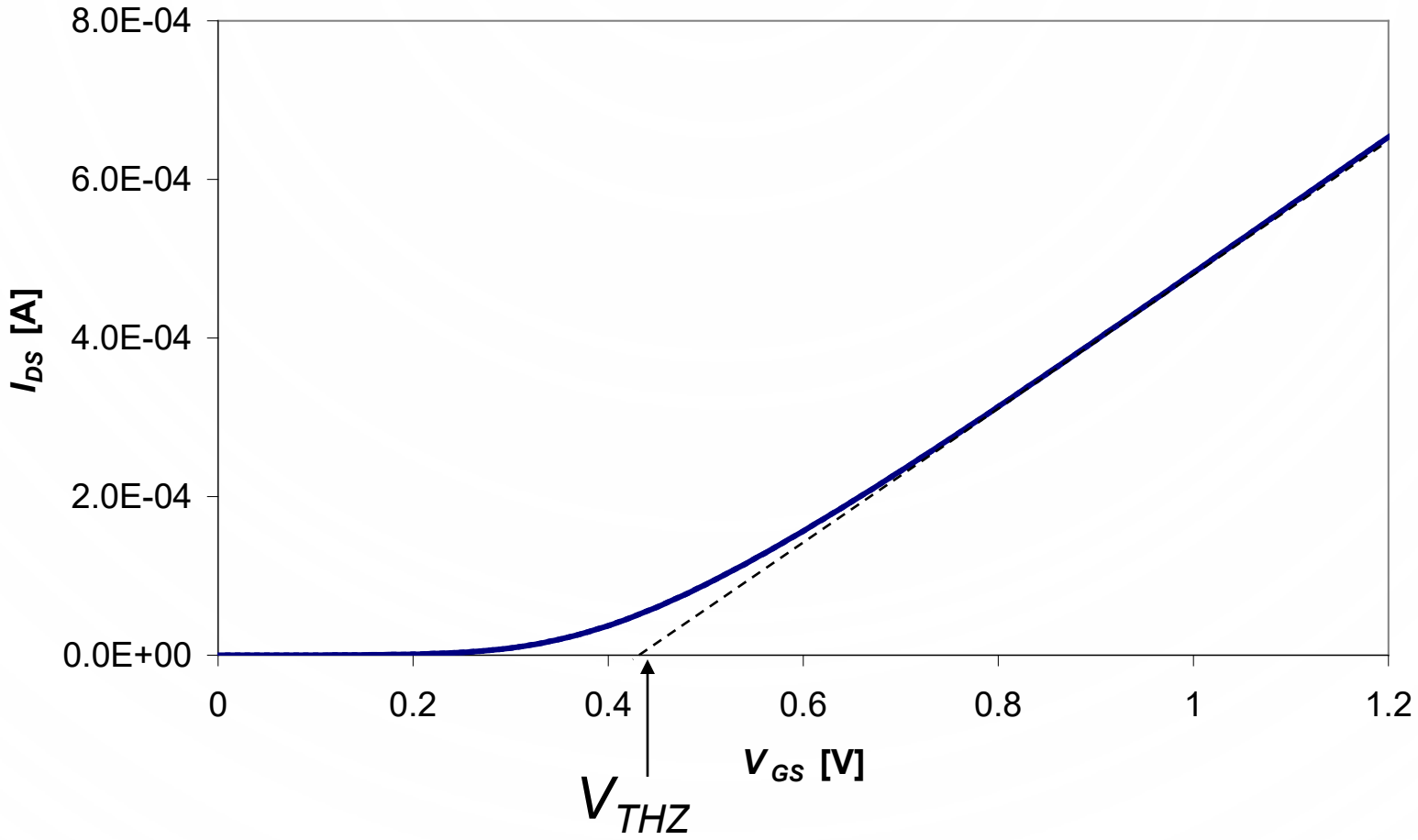
Sakurai, Newton, JSSC 4/90

Alpha Power Law Model

- This is not a physical model
- Simply empirical:
  - Can fit (in minimum mean squares sense) to variety of α's, V_{Th}
  - Need to find one with minimum square error – fitted V_{Th} can be different from physical
  - Can also fit to $\alpha = 1$
    - What is V_{Th}?

$K(V_{GS} - V_{THZ})$ Model ($\alpha = 1$)

Drain current vs. gate-source voltage



Saturation Current Models

Model	Usage
$I_{DS} = K\frac{W}{L}(V_{GS} - V_{THZ})$	Delay estimates with $V_{DD} >> V_{TH}$
$I_{DS} = \frac{W}{L}\frac{\mu C_{ox}}{2}(V_{GS} - V_{TH})^2$	Long channel devices (rare in digital)
$I_{DS} = \frac{W}{L}\frac{\mu C_{ox}}{2}(V_{GS} - V_{TH})^{\alpha}$	Delay estimates in a wider range of V_{DD}'s
$I_{DS} = \frac{W}{L}\mu C_{ox}\left((V_{GS} - V_{TH})V_{Dsat} - \frac{V_{Dsat}^2}{2}\right)$	Easy to remember, does not handle stacks correctly
$I_{DS} = \frac{W}{L}\frac{\mu C_{ox}}{2}\frac{E_C L(V_{GS} - V_{TH})^2}{(V_{GS} - V_{TH}) + E_C L}$	Handles stacks correctly, sizing

Next Lecture

- Gate models