Modeling the operation of CMOS primitive circuits and MOSFET devices

Labros Bisdounis

Associate Professor
Electrical Engineering Department
Technological Educational Institute of Patras

bisdounis@teipat.gr

Motivation

- Since performance of a digital circuit is directly affected by the transient response and the energy dissipation of the constituent logic gates, the corresponding topic is covered in most textbooks on digital integrated circuits.
- Estimation of CMOS primitive circuits' behavior, in terms of analysis and computation of their dynamic characteristics (transient response, energy dissipation) is today a standard part of digital circuit design.
- Using transistor-level simulators (SPICE-like) with continuous-time modeling of the devices, are very expensive in terms of storage and computation time.
- Hence, much of research has addressed the development of analytical timing and energy dissipation models for basic structures, without the necessity of expensive numerical iterations.
- This talk mainly regards the methodology for the derivation of closed-form, accurate expressions for the afore-mentioned parameters.
- The operational conditions of basic CMOS structures are determined and the differential equations describing their operation are solved analytically by using appropriate approximations in order to simplify the modeling procedure without significant accuracy degradation.

Motivation

- As a case study the CMOS inverter will be used.
- Following a detailed analysis of the inverter operation, accurate expressions for its output response are derived for the different operation regions.
- Based on this analysis, analytical models for the calculation of design parameters, such as propagation delay and energy dissipation, are produced.
- The models are parametric, taking into account basic device technology parameters (threshold voltages, small-geometry effects), design parameters (device sizes, parasitics, output loads), as well as operational parameters (supply voltage, input voltage transition time).
- The inverter model can be extended to multi-input CMOS gates by using reduction techniques of series- and parallel-connected transistors.
- The accuracy of the used MOS device I-V model determines to a large extent the accuracy of CMOS structures' timing and energy models.
- Because of that, the talk will emphasize on the adoption of an accurate and compact device model that takes into account the influences of predominant effects in modern nanometer device technologies.

Modeling methodology

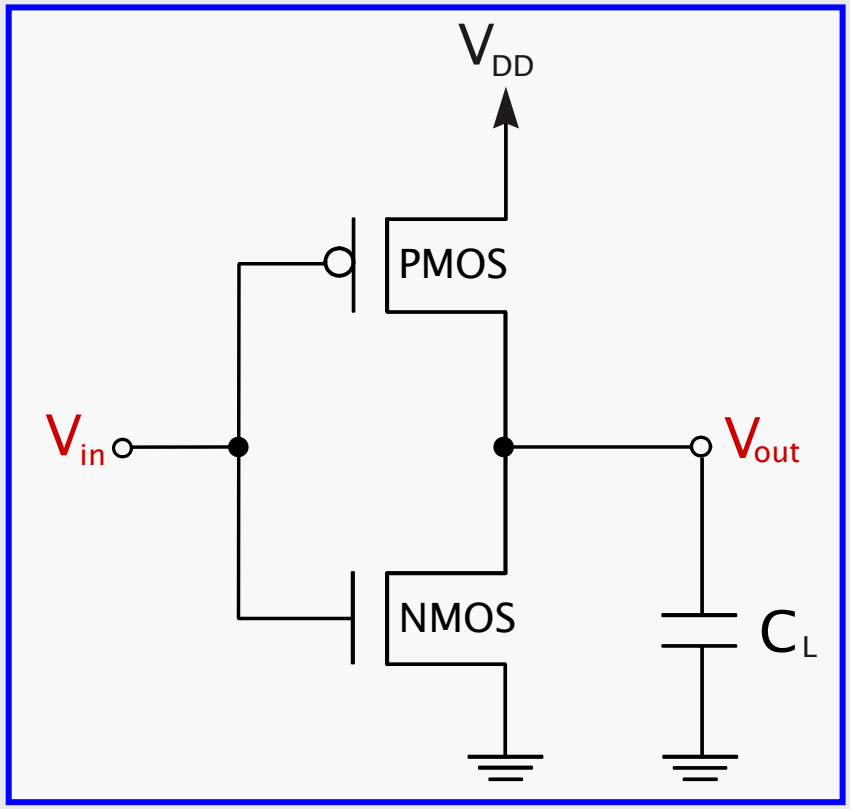
- Understand the circuit operation.
- Create a model of the circuit including parasitics to be considered.
- Write the differential equation that describes the circuit operation, i.e. apply the Kirchhoff's current law at the output node of the primitive circuit.
- Define the operation regions of the circuit's transistors in every time interval of the circuit operation.
- Adopt an appropriate current model for the transistors in all operation regions (cutoff, linear, saturation).
- The adopted transistor model should match to the I-V characteristics of the transistors by accounting first and main second order effects, according to the used technology process.
- The transistor current model should combine simplicity to provide the ability for explicit expressions derivation for design parameters, and accuracy to account for the influences of main device physical mechanisms.

Modeling methodology

- Define operational conditions which determines the bounds of different operating regions.
- Solve the differential equation in each region of operation to determine the output voltage waveform.
- Output voltage waveform is determined as an integration of the device currents in contrast to the average current method where the current is assumed equal to the average of its values at the limits of the time interval of interest.
- Since the target is to combine simplicity and accuracy, make creative and reasonable assumptions in case that the differential equation cannot be solved analytically.
- Use circuit simulations to define candidate regions for approximations.
- Avoid the use of numerical approaches by using smart techniques such as Taylor series expansion to solve boundary equations.

$$f(x) \approx f(\alpha) + f'(\alpha)(x-\alpha) + \frac{f''(\alpha)}{2}(x-\alpha)^2 + \frac{f^{(3)}(\alpha)}{6}(x-\alpha)^3 + \dots$$

Case study: CMOS inverter

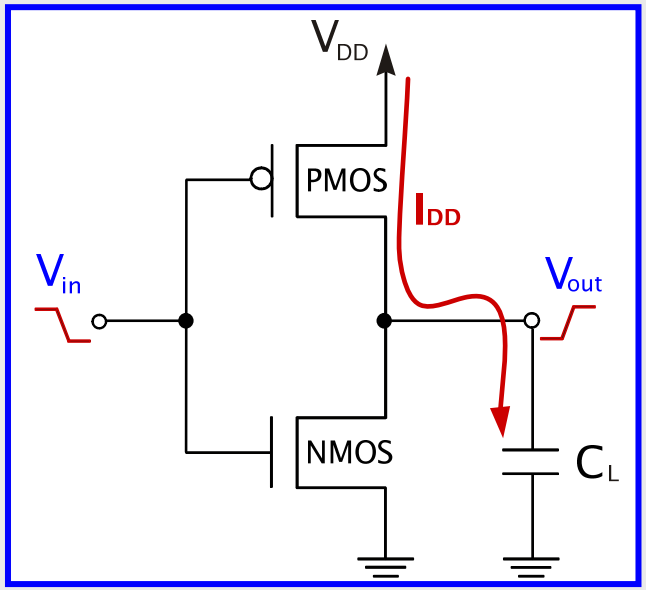


Inverter energy dissipation

$$E = E_D + E_L + E_{SC}$$

- E_D: dynamic energy dissipation due to charge and discharge of the capacitive load, during the inverter switching.
- E_L: leakage energy dissipation.
- E_{SC}: short-circuit energy dissipation due to the direct current path from power supply to ground, during the inverter switching.

Dynamic energy dissipation



- During the rising transition of the output node the supplied energy by the power supply is:

$$E_D = V_{DD}\int_0^{t_{tr}} I_{DD}\,dt = V_{DD}\int_0^{t_{tr}} C_L\frac{dV_{out}}{dt}dt = V_{DD}\int_0^{V_{DD}} C_L\,dV_{out} = C_L V_{DD}^2$$

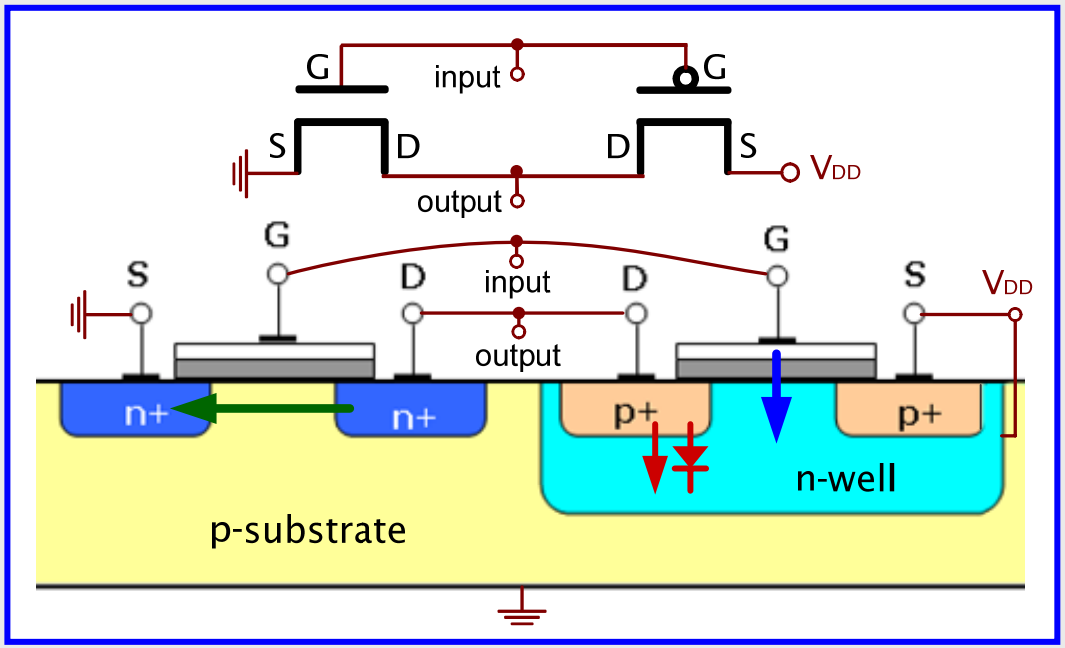
- The supplied energy is independent from the output waveform of the inverter.
- During the same transition, the energy stored to the output load capacitance is:

$$E_{C_L} = \frac{1}{2}C_L V_{DD}^2$$

- So, half of the supplied energy is stored to the load capacitance and will be lost to the ground during the falling output transition, and the rest is lost in the form of heat at the PMOS device.

Leakage energy dissipation

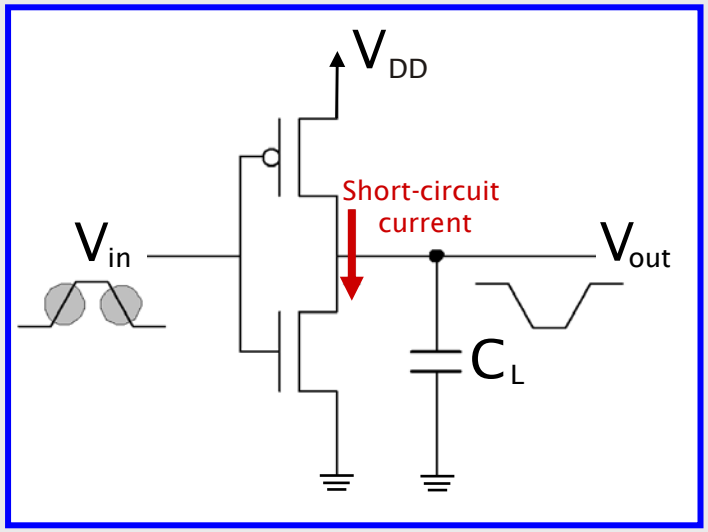
- The leakage energy dissipation is caused by three types of leakage currents:
  - the reverse-bias diode leakage current at the transistor diffusion areas,
  - the subthreshold current through a turned-off transistor channel, and
  - the tunnelling current through very thin gate oxide.

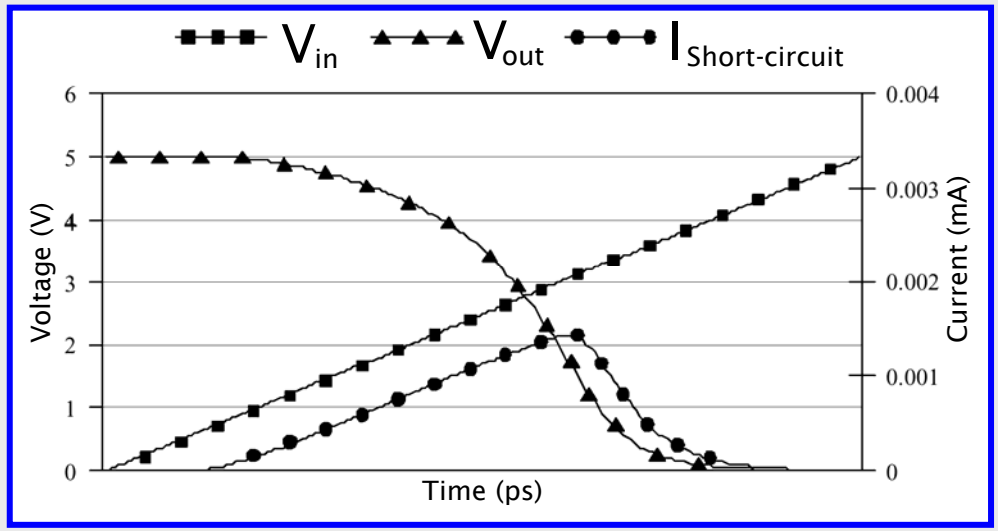


- The leakage energy dissipation is mainly dependent to technology process parameters and it is significant for nanometer transistors.

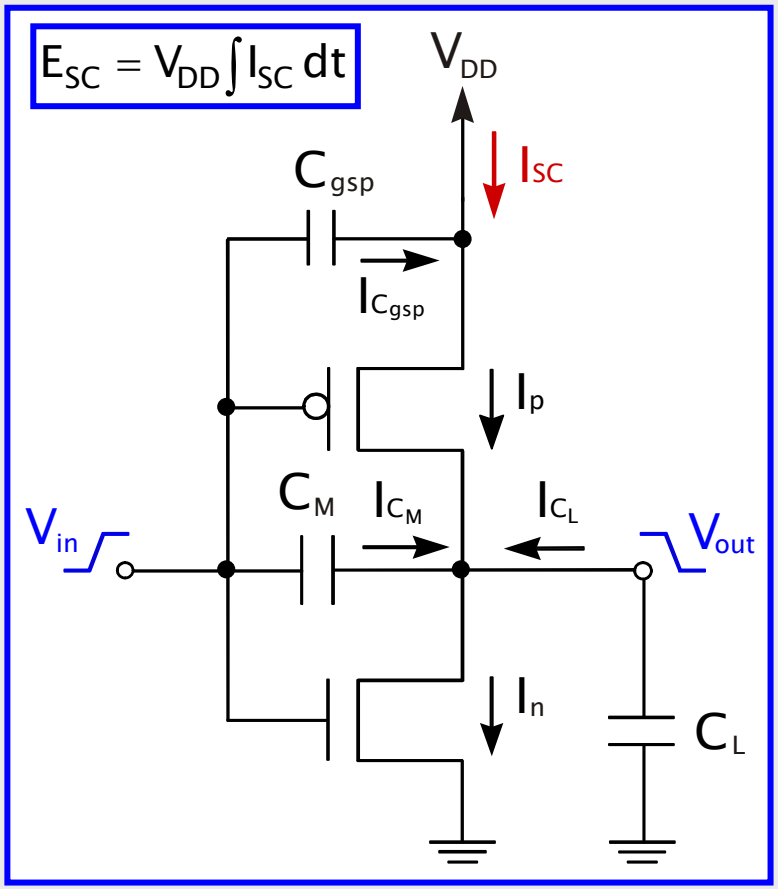
Short-circuit energy dissipation

- Short-circuit energy dissipation is due to the direct path between power supply and ground, which is formed during the inverter switching.
- It depends on the input voltage transition time, the output load, the supply voltage, and the internal design characteristics of the inverter.
- Because of that, the computation of this type of energy dissipation is complex.
- Accurate analysis of the output voltage waveform is required, and the afore-mentioned methodology will be followed.



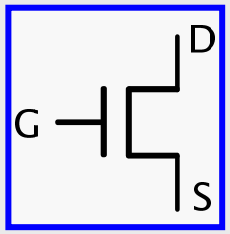


Short-circuit energy dissipation



- The model is based on closed-form expressions of the CMOS inverter output waveform, which include the influences of both transistor currents and the gate-drain coupling (C_M).
- A version of the alpha-power law MOSFET model is adopted to relate the terminal voltages to the drain current in deep-submicrometer devices.
- The resulting energy model accounts for the influences of input voltage transition time, transistors' sizes, device carrier velocity saturation, gate-drain and short-circuiting transistor's gate-source capacitances, and output load.

Adopted MOSFET current model



$$I_D = B\left(V_{GS} - V_T\right)^{\alpha}$$

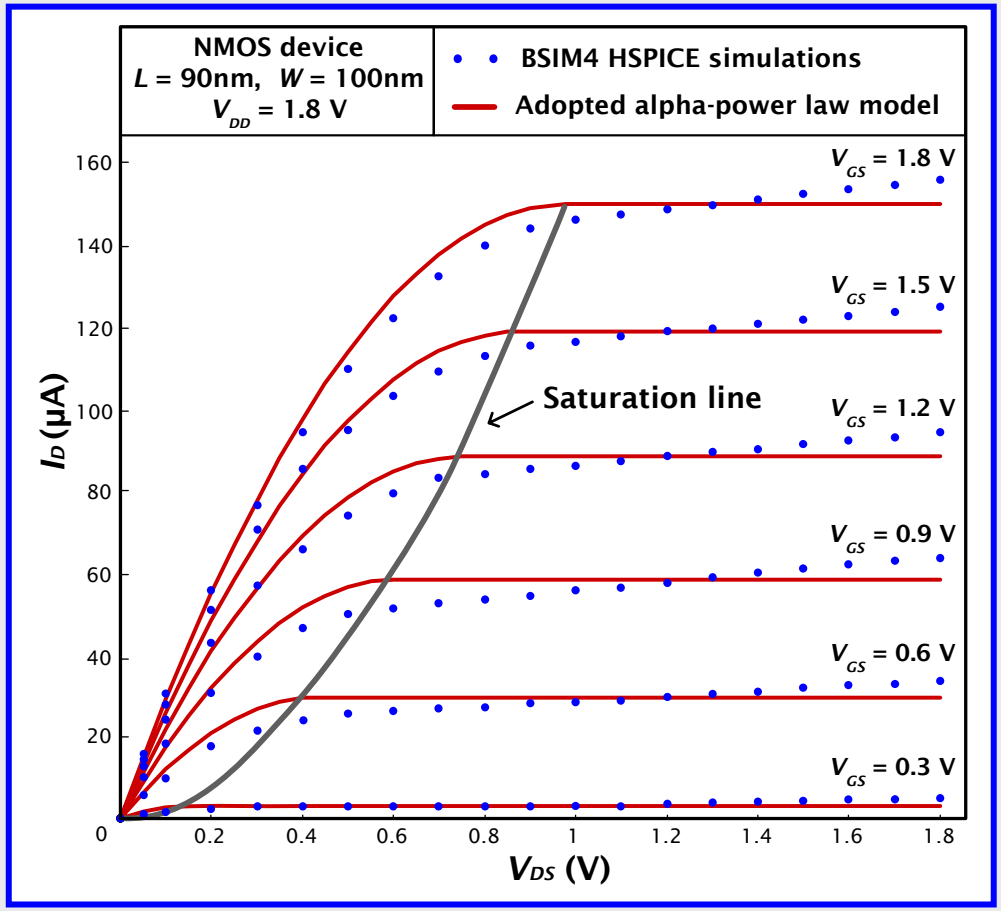
$V_{DS} > V'_{DO}$ Saturation region

$$I_D = B\left(V_{GS} - V_T\right)^{\alpha}\left(2 - \frac{V_{DS}}{V'_{DO}}\right)\frac{V_{DS}}{V'_{DO}}$$

$V_{DS} \le V'_{DO}$ Triode or linear region

$$V'_{DO} = K\left(V_{GS} - V_T\right)^{\alpha/2}$$

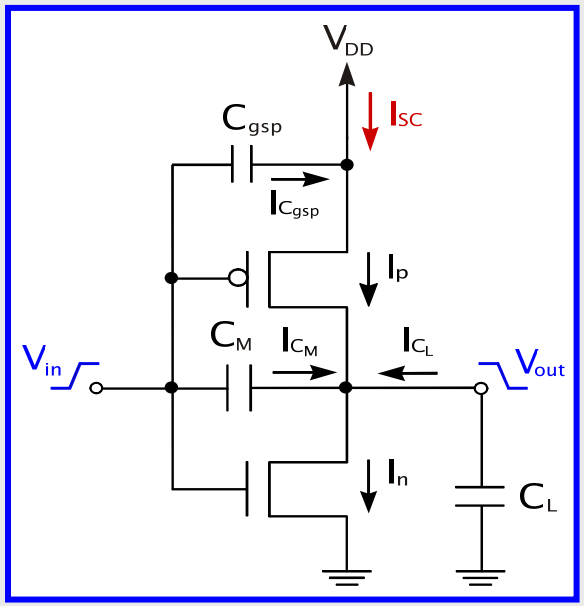
Saturation voltage



The model parameters are determined by the used technology process or extracted from the MOSFET device output characteristics

Output voltage waveform analysis

- The output voltage waveform of the CMOS inverter is obtained for a rising input ramp: $V_{in} = V_{DD} \cdot (t/\tau)$ for $0 \le t \le \tau$, $V_{in} = 0$ for $t \le 0$ and $V_{in} = V_{DD}$ for $t \ge \tau$, where τ is the input voltage rise time.
- The analysis for a falling input is symmetrical.
- The differential equation which describes the discharge of the load capacitance C_L for a CMOS inverter, taking into account the current through the gate-drain coupling capacitance (C_M) is written as:



$$C_L \frac{dV_{out}}{dt} = C_M \left(\frac{dV_{in}}{dt} - \frac{dV_{out}}{dt} \right) + I_p - I_n$$

The gate-drain capacitance (C_M) is the sum of the gate-drain capacitances of both transistors

Output voltage waveform analysis

After normalizing voltages with respect to V_{DD} and using the variable **x = t / τ**, the device currents become:

$$I_p = \begin{cases} k_{lp1}\,(1-x-p)^{\alpha_p/2}(1-u_{out}) - k_{lp2}\,(1-u_{out})^2, & 1-u_{out} < u'_{dop} \\ k_{sp}\,(1-x-p)^{\alpha_p}, & 1-u_{out} \geq u'_{dop} \end{cases}$$

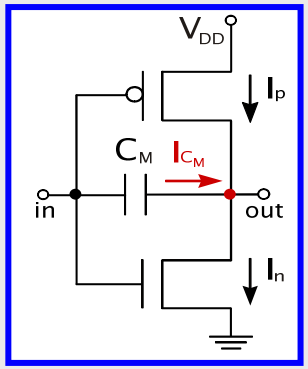
$$p = \frac{|V_{TP}|}{V_{DD}} \qquad u'_{dop} = k_{vp}(1-x-p)^{\alpha_p/2}$$

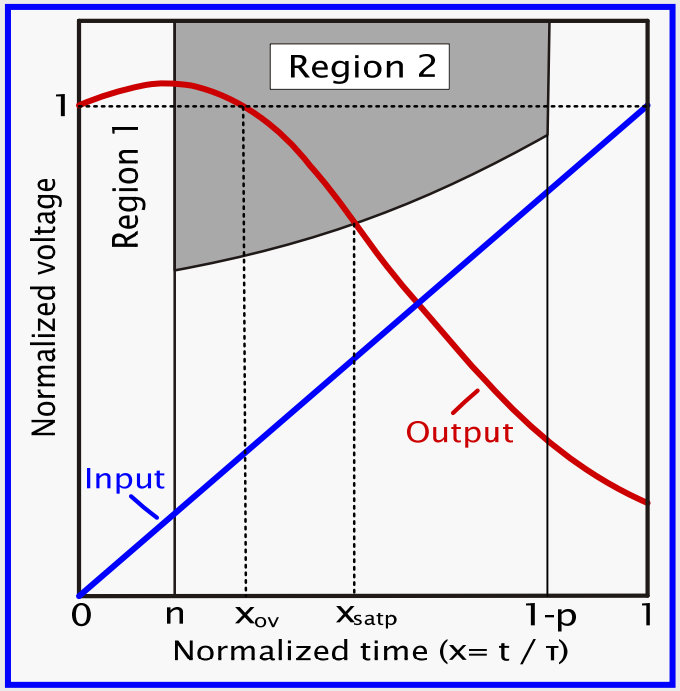
$$I_n = \begin{cases} k_{ln1}\,(x-n)^{\alpha_n/2}\,u_{out} - k_{ln2}\,u_{out}^2, & u_{out} < u'_{don} \\ k_{sn}\,(x-n)^{\alpha_n}, & u_{out} \geq u'_{don} \end{cases}$$

$$n = \frac{V_{TN}}{V_{DD}} \qquad u'_{don} = k_{vn}(x-n)^{\alpha_n/2}$$

where k_{l1}, k_{l2}, k_s, and k_v are constants depended on K, B, V_{DD}, and α.

Output voltage waveform analysis





- In region 1 ($0 \leq x \leq n$) the NMOS transistor is off and the PMOS transistor is in the linear region.
- In region 2 ($n \leq x \leq x_{satp}$) the NMOS transistor is saturated and the PMOS transistor is still in the linear region.
- x_{satp} is the normalized time value when the PMOS transistor is entering the saturation region.
- Part of the charge from the input which injected through the C_M causes an overshoot at the beginning of the output signal transition ($0 \leq x \leq x_{ov}$).
- During the overshoot there is no current from power supply to ground because the output voltage is greater than the supply voltage.

Output voltage waveform analysis

$$C_L \frac{dV_{out}}{dt} = C_M \left(\frac{dV_{in}}{dt} - \frac{dV_{out}}{dt} \right) + I_p - I_n \Rightarrow \frac{d u_{out}}{d x} = c_m + \frac{(I_p - I_n)\,\tau}{(C_L + C_M)\,V_{DD}}$$

$$I_p = k_{lp1}\,(1 - x - p)^{\alpha_p/2}\,(1 - u_{out}) - k_{lp2}\,(1 - u_{out})^2 \qquad c_m = \frac{C_M}{C_L + C_M}$$

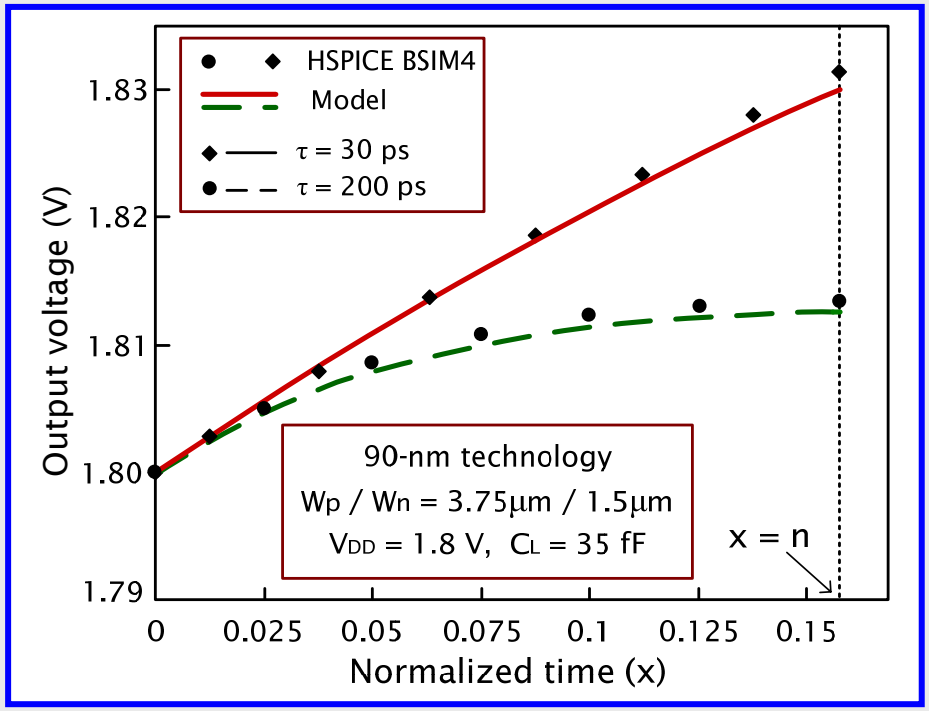
- In region 1 ($0 \le x \le n$), the differential equation cannot be solved analytically.
- The charge injected through C_M causes the main influence on the output signal in this region 1.
- Approximations:
  - An average value of $x = n / 2$ is used in the first term of PMOS current.
  - An approximated expression $u_{out} = 1 + c_m x$ (that is the output signal if only the charge through C_M is taken into account) is used in the quadratic term of the PMOS current. The charge contributed by this term of the PMOS current is very small due to the small values of the PMOS normalized drain-source voltage in this operating region.

Output voltage waveform analysis

$$u_{out} = 1 + \frac{c_m}{C^3 A_{lp1}^3}\left[2A_{lp2}c_m(e^{-xCA_{lp1}} - 1) + 2C A_{lp1} A_{lp2} c_m x + C^2 A_{lp1}^2 (1 - e^{-xCA_{lp1}} - A_{lp2} c_m x^2)\right]$$

$$A_{lpi} = \frac{\tau\, k_{lpi}}{V_{DD}\,(C_L + C_M)} \qquad C = \left(1 - p - \frac{n}{2}\right)^{\frac{\alpha_p}{2}} \qquad c_m = \frac{C_M}{C_L + C_M}$$

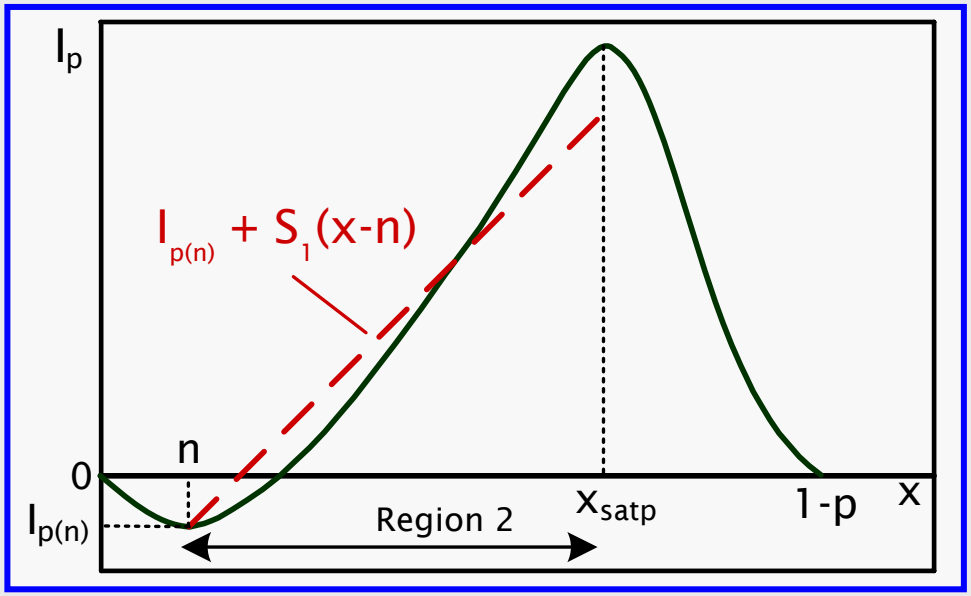
Output voltage expression for Region 1



Output voltage waveform analysis

- In region 2, the NMOS device is saturated, while the PMOS device remains in the linear region.
- The output node differential equation cannot be solved analytically, and because of that the PMOS current is approximated by a linear function of x:

$$I_p = I_{p(n)} + S_1(x-n) \qquad I_{p(n)} = k_{lp1}(1-n-p)^{\alpha_p/2}(1-u_n)$$



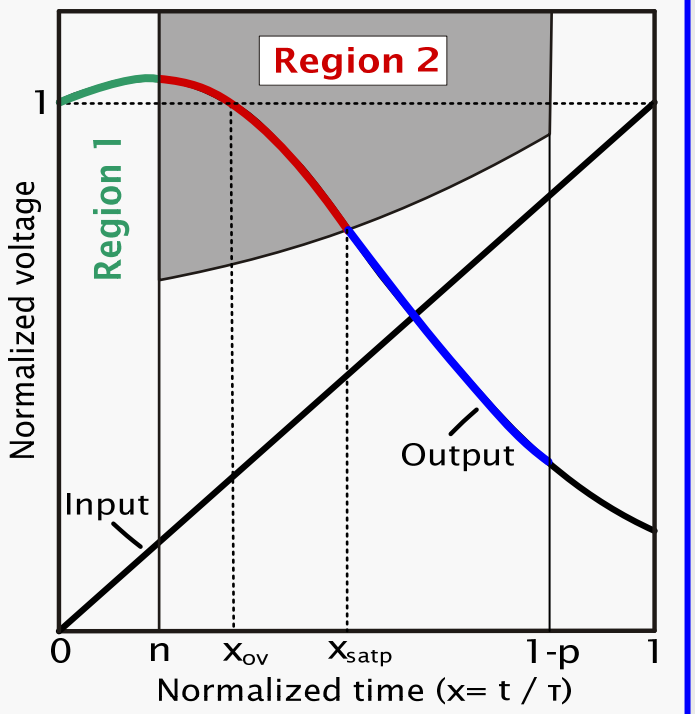
The current slope S_1 is computed by equating the exact PMOS current in the linear region with the approximated one at x = (1 – p) /2

Output voltage waveform analysis

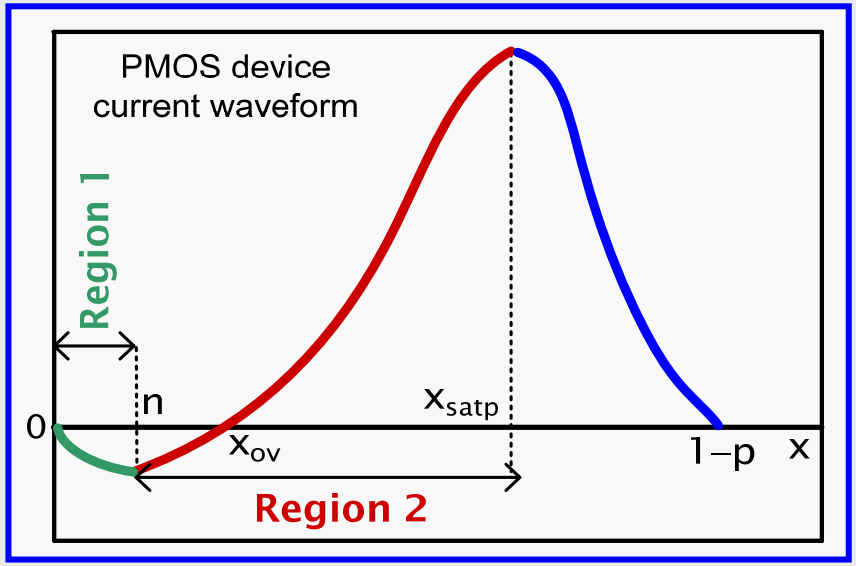
Output voltage expression for Region 2

$$u_{out} = u_n + c_m (x-n) + I_{p(n)}\, d\, (x-n) + \frac{S_1\, d\, (x-n)^2}{2} - \frac{A_{sn}\, (x-n)^{\alpha_n+1}}{\alpha_n + 1}$$

$$A_{sn} = \frac{k_{sn}\, \tau}{V_{DD}(C_L + C_M)} \qquad d = \frac{\tau}{V_{DD}(C_L + C_M)}$$



$$I_p = \begin{cases} k_{lp1}\, (1-x-p)^{\alpha_p/2}\, (1-u_{out}) - k_{lp2}\, (1-u_{out})^2 \\ k_{sp}\, (1-x-p)^{\alpha_p} \end{cases}$$



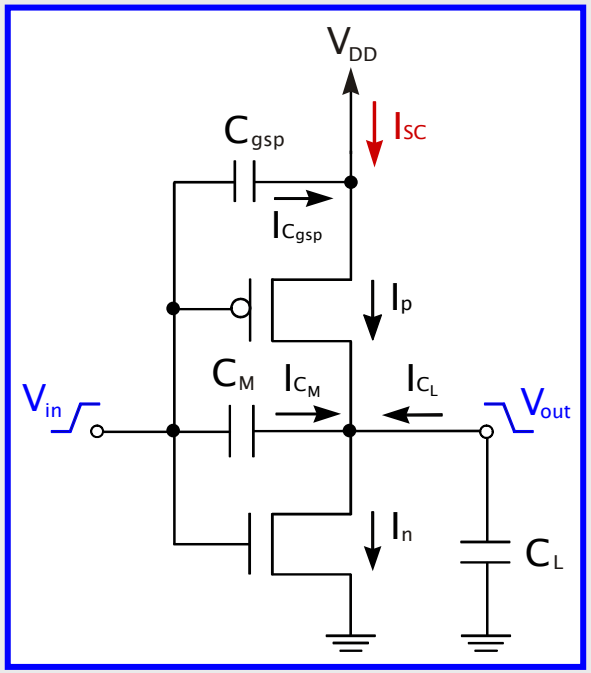
Short-circuit energy dissipation analysis

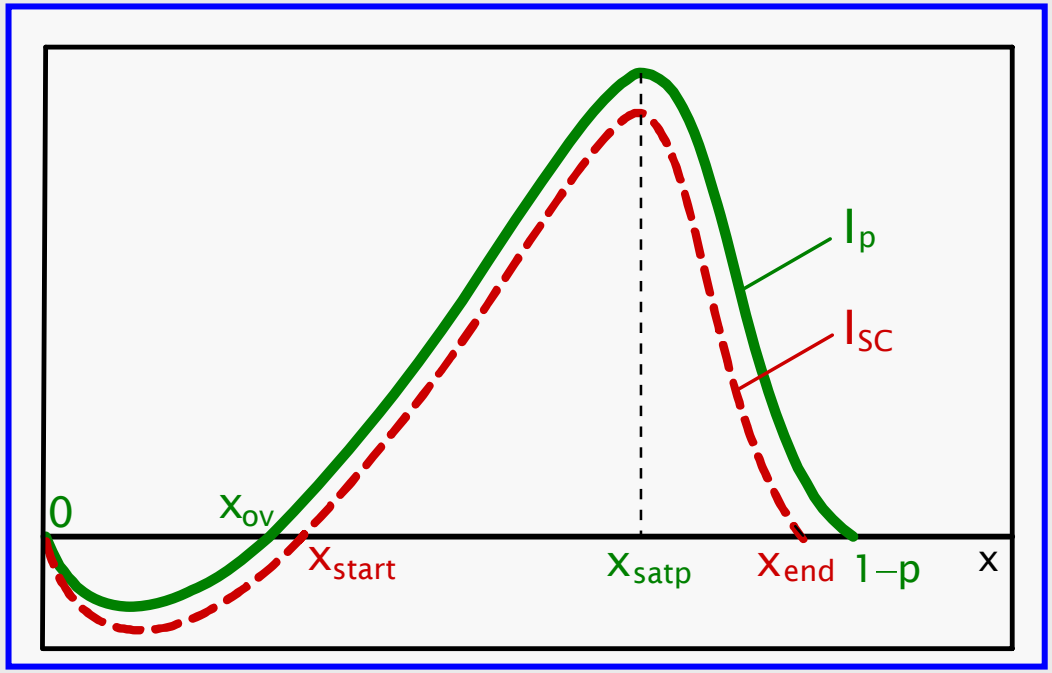
The current through the PMOS device includes two non-short-circuit current components: the current flowing through C_{gsp} and the current flowing from the output to the supply node during the overshoot of the output signal.

$$I_{SC} = I_p - I_{C_{gsp}}$$

$$I_{C_{gsp}} = C_{gsp}(V_{DD}/\tau)$$

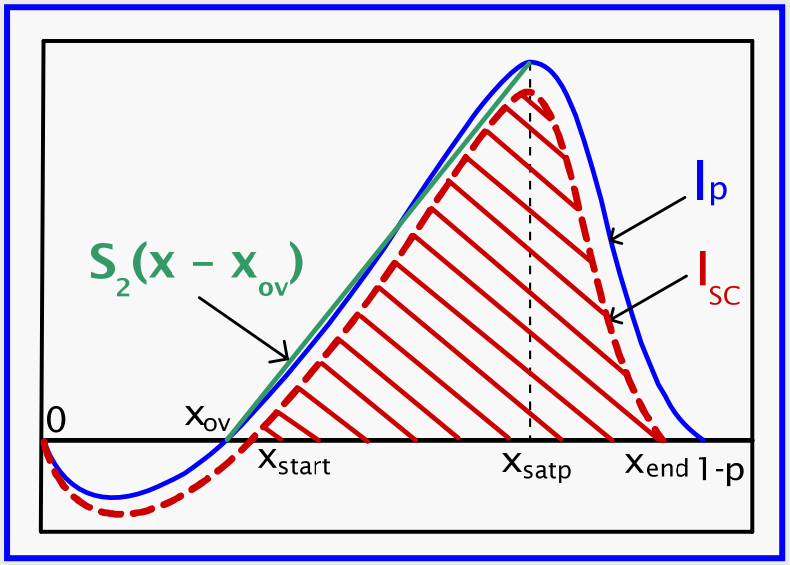
$$E_{SC} = V_{DD}\int_{x_{start}}^{x_{end}} I_{SC}\,\tau\,dx = V_{DD}\left(\int_{x_{start}}^{x_{satp}} I_{SC}\,\tau\,dx + \int_{x_{satp}}^{x_{end}} I_{SC}\,\tau\,dx\right)$$





Short-circuit energy dissipation analysis

- A linear approximation of the PMOS transistor current is used in the first integral, while in the second integral the PMOS saturation current is substituted.
- S_2 is the slope of I_p and is calculated by equating the exact PMOS current in the linear region with the approximated one, at the middle of the interval $[x_{ov}, x_{satp}]$.
- x_{ov}, x_{satp} are computed by using 2nd order Taylor series expansions of u_{out} & u'_{dop}.



$$E_{SC} = V_{DD}\int_{x_{start}}^{x_{satp}} (I_p - I_{C_{gsp}})\,\tau\,dx + V_{DD}\int_{x_{satp}}^{x_{end}} (I_p - I_{C_{gsp}})\,\tau\,dx$$

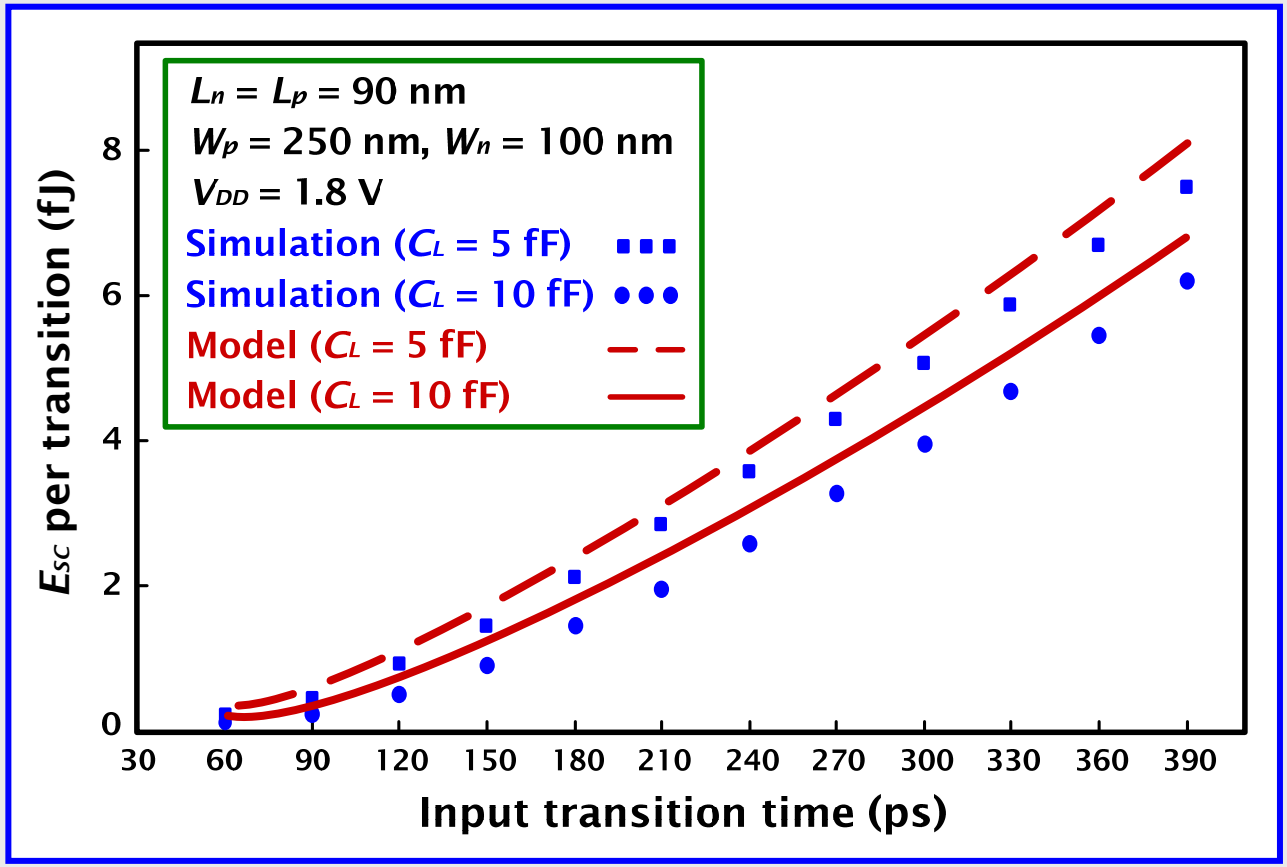
$$E_{SC} = \frac{V_{DD}}{2}(x_{satp} - x_{start})[(x_{satp} + x_{start} - 2x_{ov})S_2 - \frac{2C_{gsp}V_{DD}}{\tau}] + \frac{V_{DD}k_{sp}\tau}{\alpha_p + 1}[(1-p-x_{satp})^{\alpha_p+1} - (1-p-x_{end})^{\alpha_p+1}] - C_{gsp}V_{DD}^2(x_{end} - x_{satp})$$

$$S_2(x_{start} - x_{ov}) - C_{gsp}(V_{DD}/\tau) = 0 \Rightarrow x_{start}$$

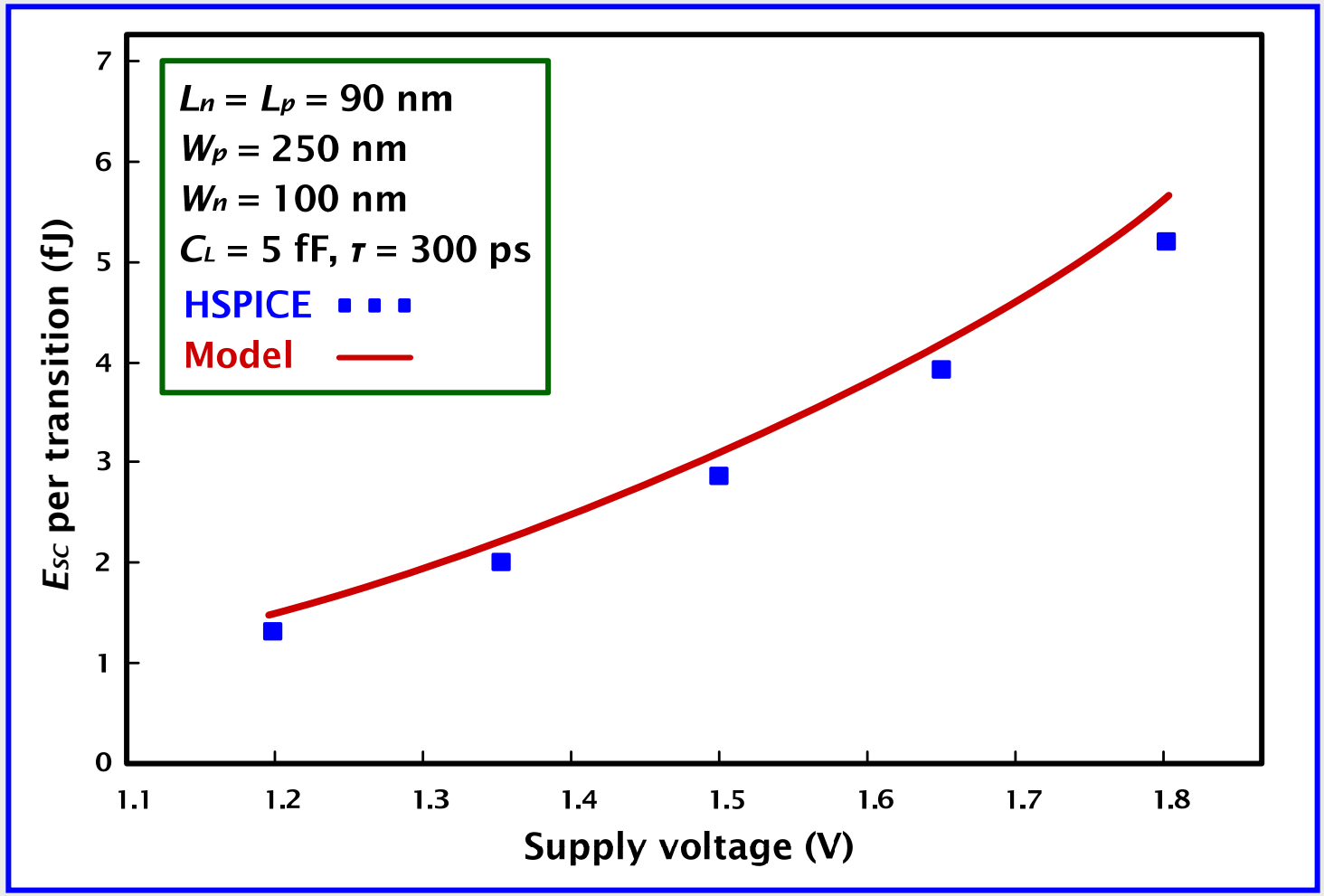
$$k_{sp}(1 - x_{end} - p)^{\alpha_p} - C_{gsp}(V_{DD}/\tau) = 0 \Rightarrow x_{end}$$

Experimental results

The accuracy of the short-circuit energy model has been validated through comparisons with BSIM4 HSPICE simulations for various input transition times, output loads, and supply voltages.

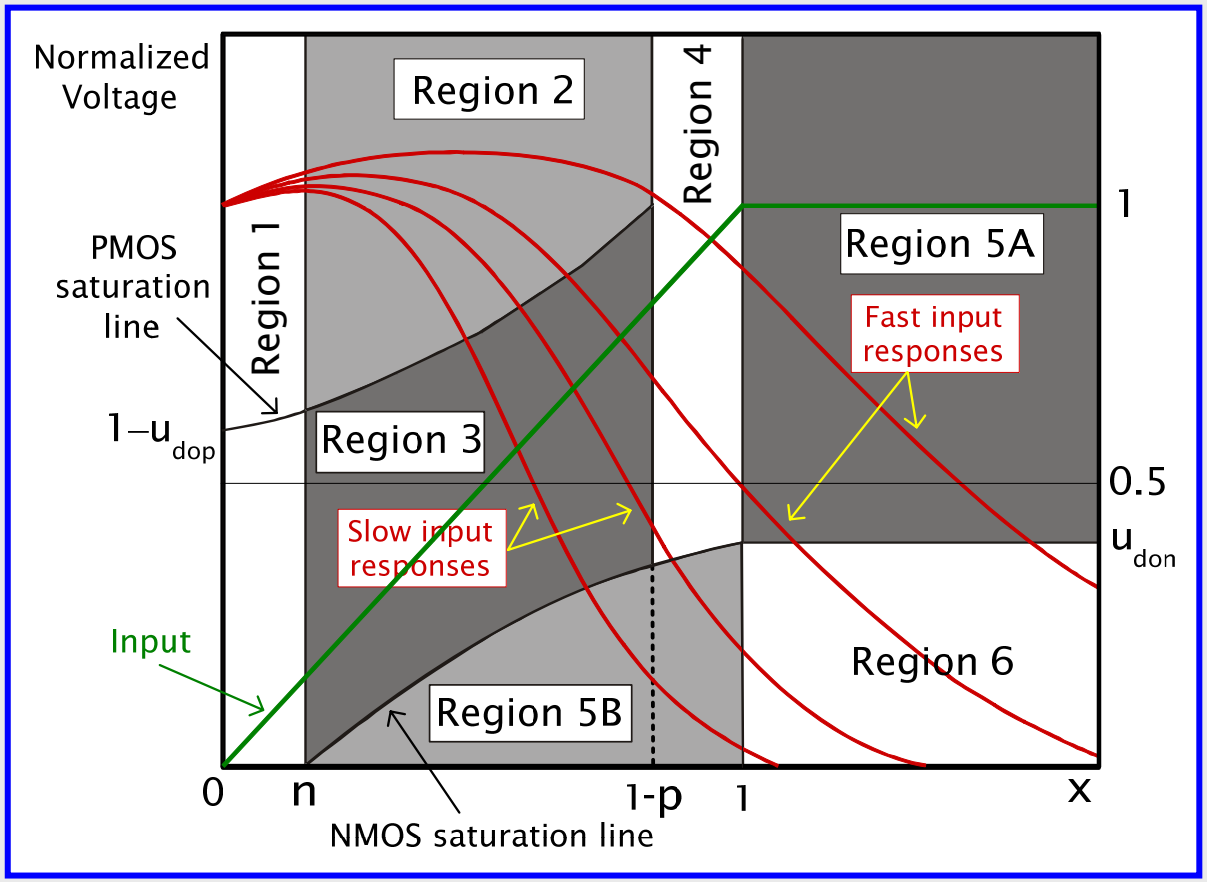


Experimental results

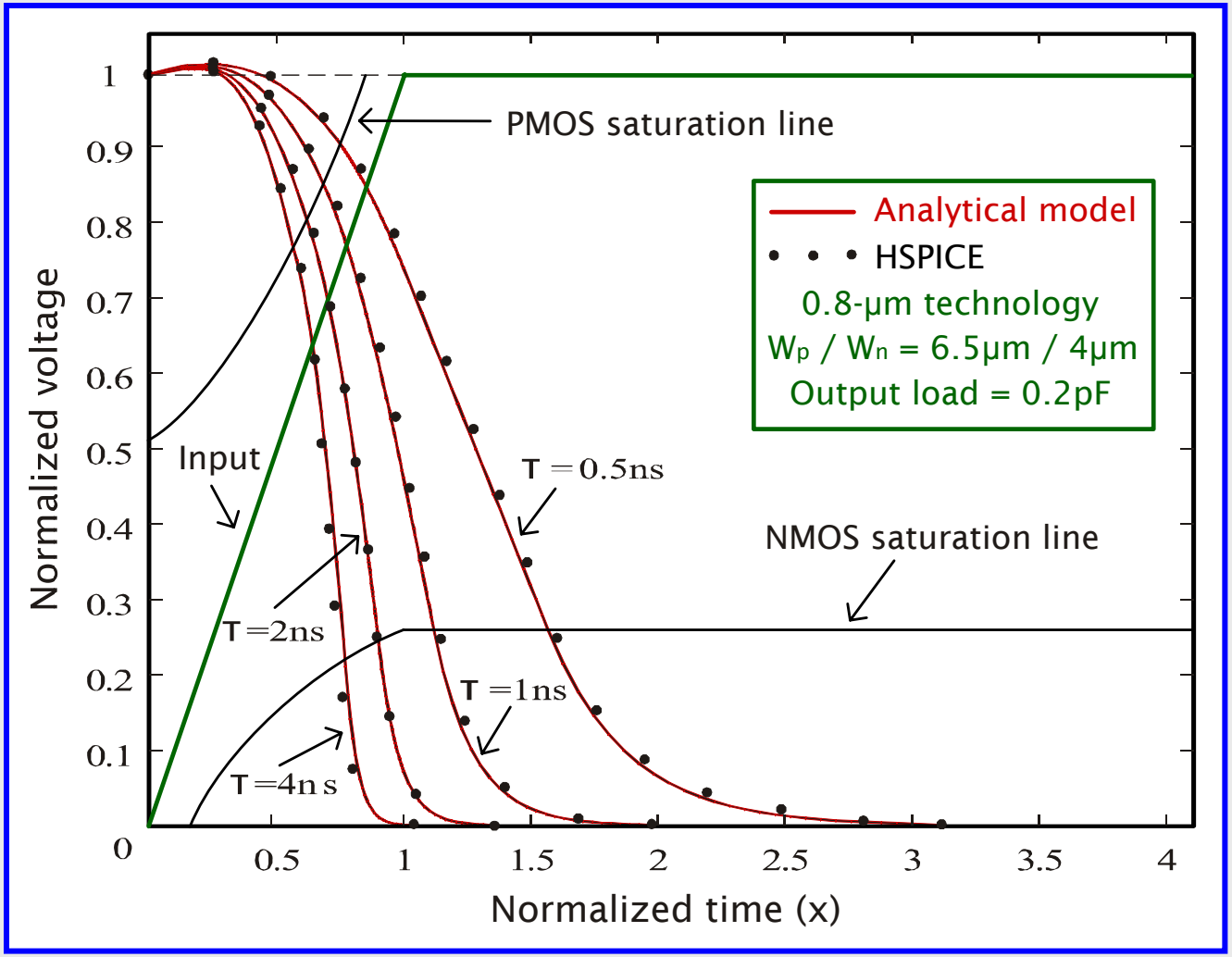


Complete inverter transient response

- The combination of saturation, linear & cut-off transistors' operation modes defines different operating regions of the CMOS inverter.
- The input voltage slope is also considered.
- The differential equation is defined and solved for all operating regions.
- Appropriate approximations are needed (e.g. region 2).

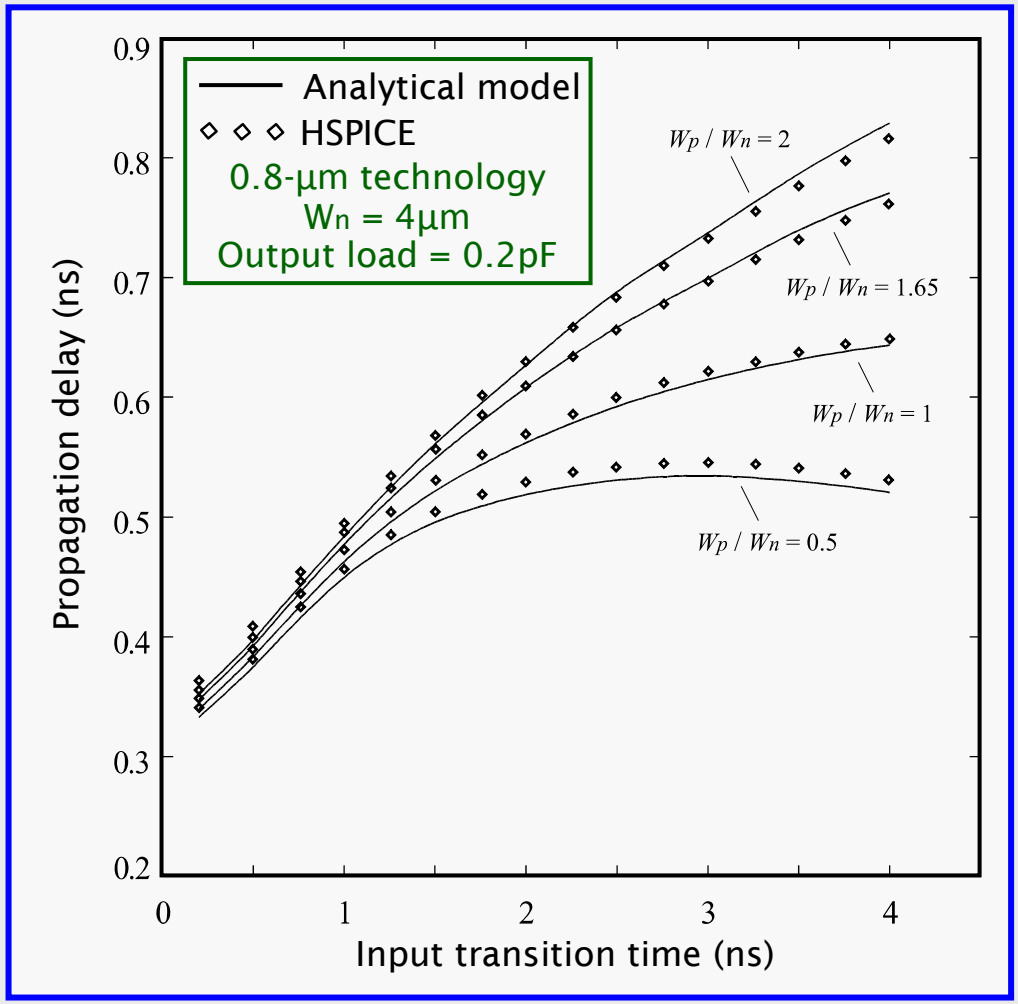


Complete inverter transient response



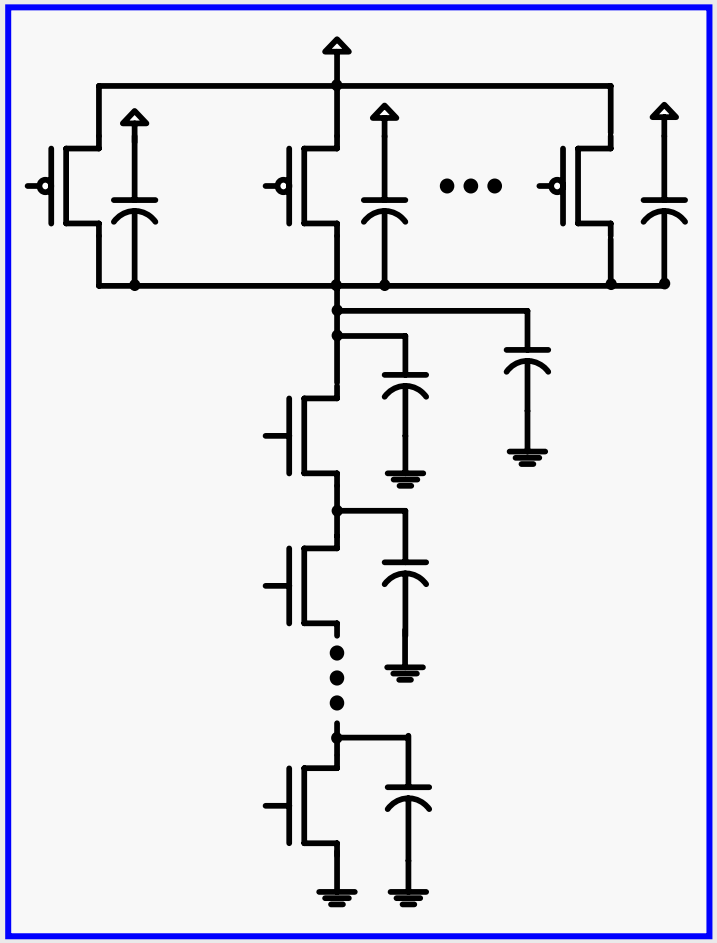
Inverter propagation delay

The inverter delay at the 50% voltage level is derived directly from the analytical expression of the inverter output voltage at the proper operating region

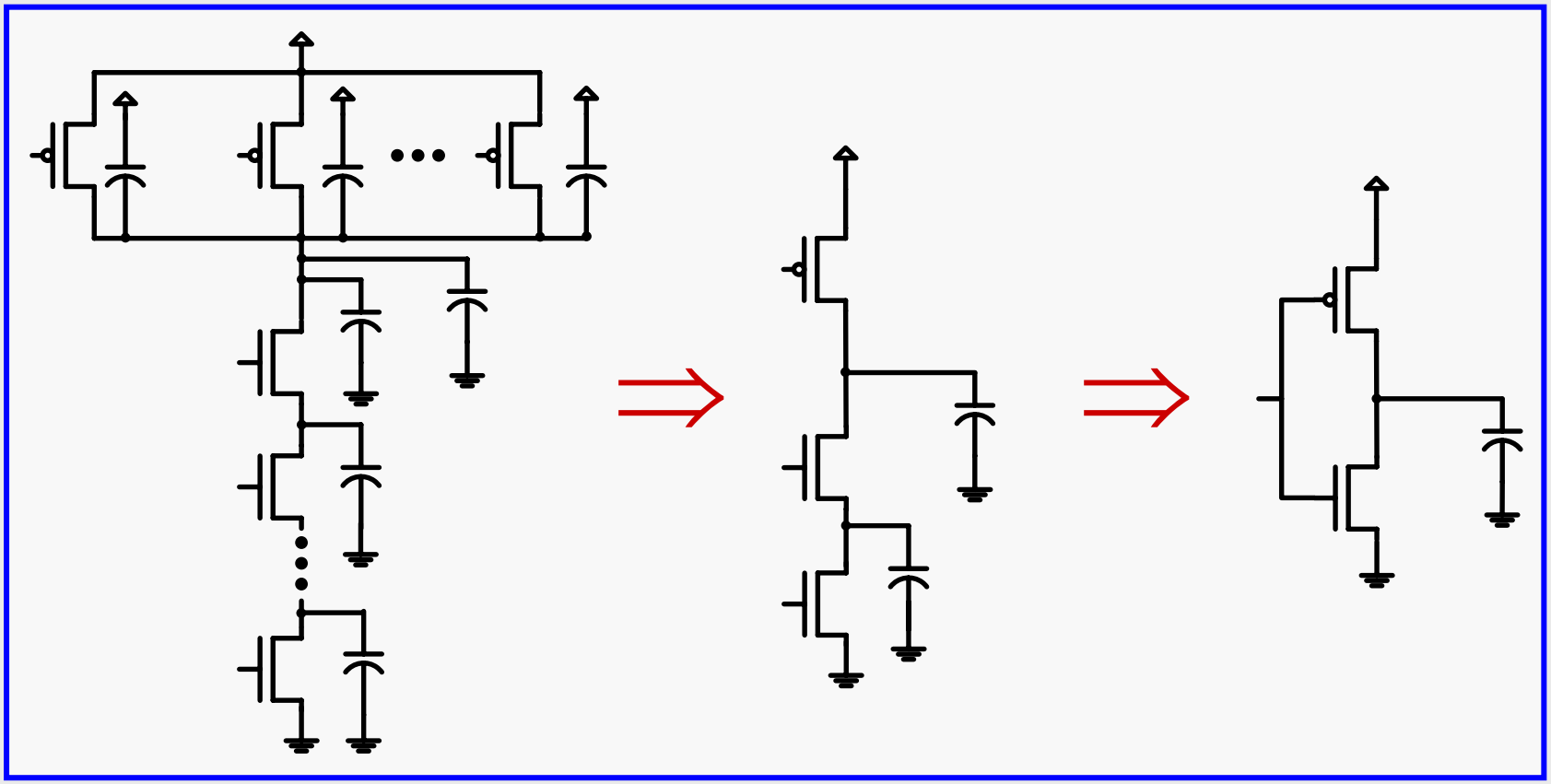


Extension to multi-input CMOS gates

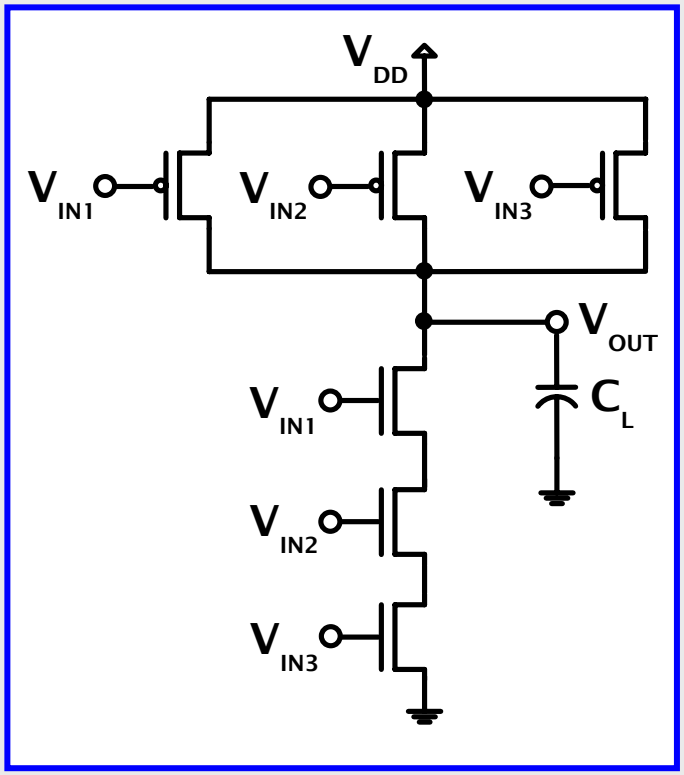
- The extension of the inverter model to multi-input CMOS gates is performed by reducing each gate to an equivalent inverter.
- This procedure requires the modeling of:
  - the series- and parallel-connected transistors
    ⇒ reduction to single equivalent transistors,
  - the case of overlapping inputs
    ⇒ reduction to a single effective input signal.
- For a successful reduction of a gate, one should take into account:
  - the transition time of the gate inputs,
  - the number of switching inputs of the gate,
  - the position of the switching inputs,
  - the body effect,
  - the output load of the gate, and
  - the internal node capacitances.

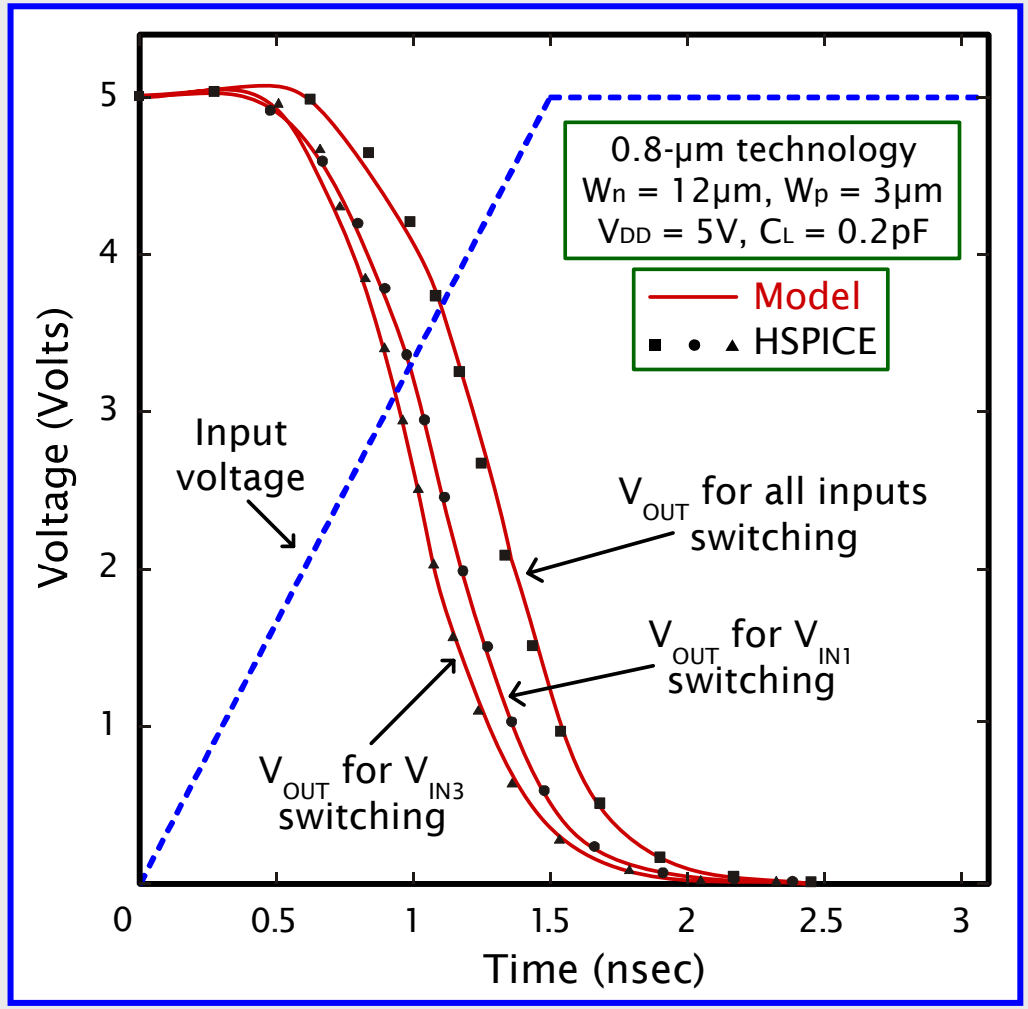


Extension to multi-input CMOS gates



Experimental results





MOSFET modeling for circuit analysis

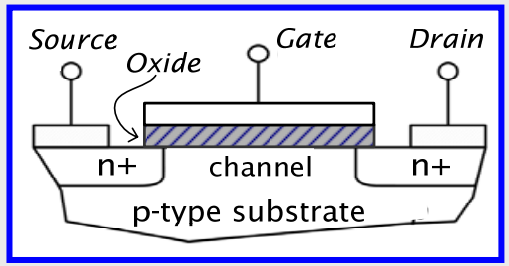
- The accuracy and computational efficiency of circuit analysis models is directly affected by the accuracy and the simplicity of the used MOSFET model.
- Such a model must meet two equally important requirements:
  - accurate transistor drain current prediction, and
  - simplicity of the device model to obtain explicit expressions for design parameters (i.e. transient response, energy dissipation).
- In most existing MOSFET I-V models, the effects that determine the device behavior are accounted for through physical and empirical parameters.
- With the growing complexity of physical mechanisms in nanometer devices, MOSFET I-V models become very complex and employ a large number of parameters to provide the highest accuracy.
- Although these complex but accurate models can be handled by circuit simulators, they do not satisfy the requirement of computational efficiency.
- Hence, compact MOSFET models are needed, as simple as possible to take into account the influences of essential physical mechanisms in nanometer devices by using few parameters extracted through measurements or simulations.

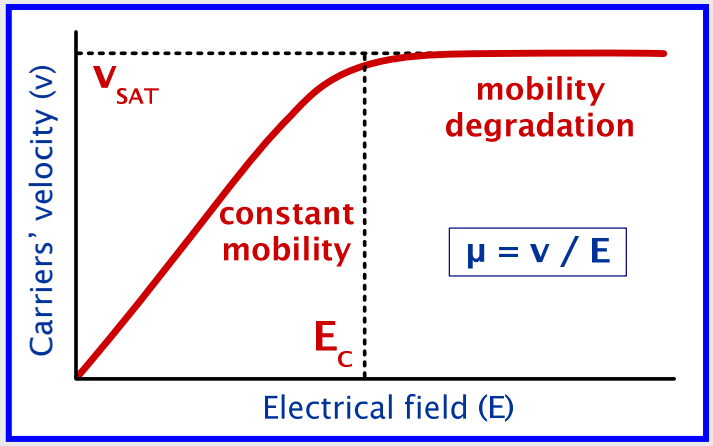
MOSFET modeling for circuit analysis

- In order to support the modeling methodology of circuit primitives, an accurate and compact I-V model for nanometer MOSFETs is needed.
- The I-V equations of such model may use empirical parameters to match measured or simulated device characteristics.
- The influence of predominant effects in nanometer devices should be taken into account:
  - mobility degradation and velocity saturation,
  - channel-length modulation,
  - drain-induced barrier lowering (DIBL),
  - body effect,
  - narrow channel width,
  - source-drain parasitic resistance.

Mobility degradation

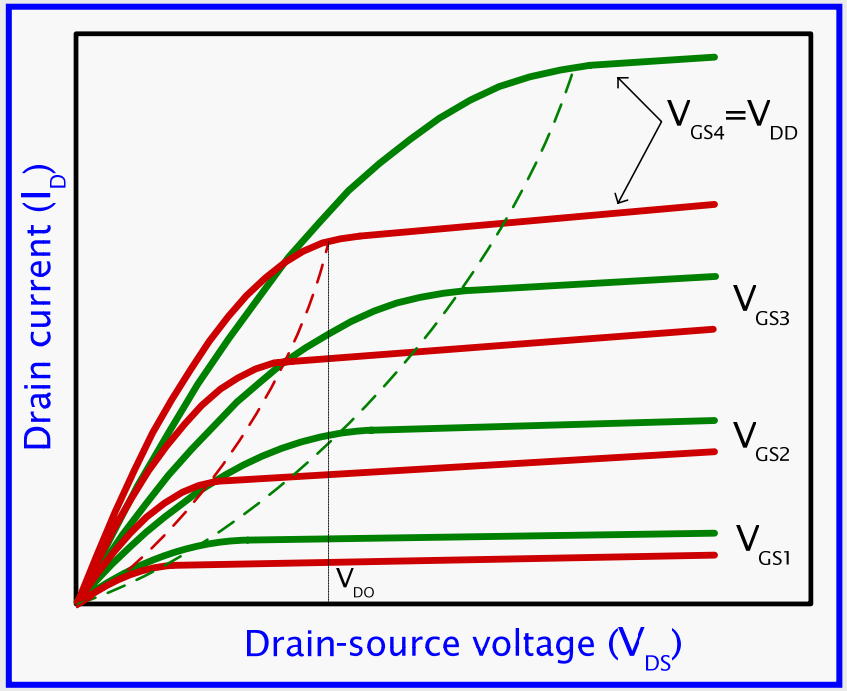
- For small electric fields, the carriers' mobility is constant and independent of the applied electric field.
- When the horizontal electrical field (moving the channel carriers) reaches a critical value, the carriers velocity tends to saturate due to scattering effect (i.e. electrons moving in semiconductor material collide with silicon atoms).
- This effect is more pronounced for reduced channel length that implies higher horizontal electric fields for equivalent drain–source voltages.
- The vertical electric field originating from the gate voltage further inhibits channel carrier mobility.
- This field pushes carriers toward the gate oxide and the carriers' mobility is reduced due to carrier collisions with the oxide-channel interface.





Mobility degradation

- Influence on the device output characteristic curves:
  - Saturation is achieved at smaller drain-source voltages,
  - The spacing of the I-V curves in saturation is not according to square law, but becomes nearly proportional to gate-source voltage increment.
- The modeling of mobility degradation and velocity saturation effects is achieved by employing the velocity saturation index (α) to describe the power laws featuring the drain current and the drain-source saturation voltage.

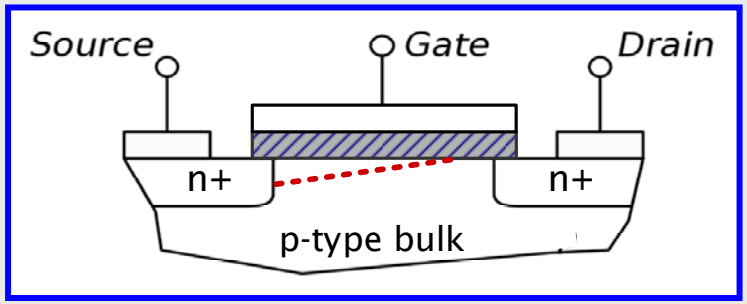


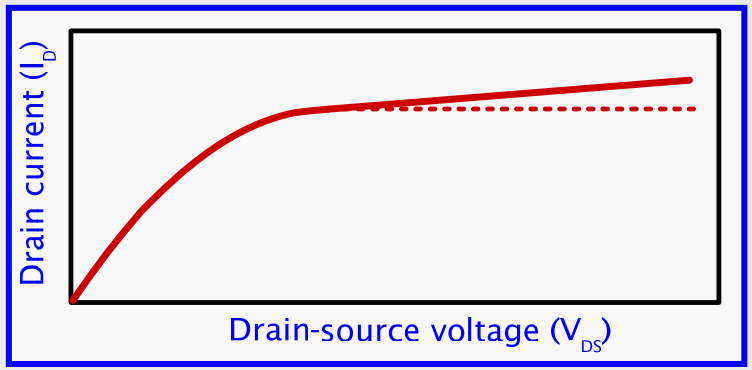
$$I_D = B(V_{GS} - V_{TH})^2 \Rightarrow I_D = B(V_{GS} - V_{TH})^{\alpha}$$

$$V'_{DO} = V_{GS} - V_{TH} \Rightarrow V'_{DO} = V_{DO}\left(\frac{V_{GS} - V_{TH}}{V_{DD} - V_{TH}}\right)^{\frac{\alpha}{2}}$$

Channel length modulation and DIBL

- Channel length modulation (CLM) refers to the shortening of the length of the inverted channel region with increase in drain bias.

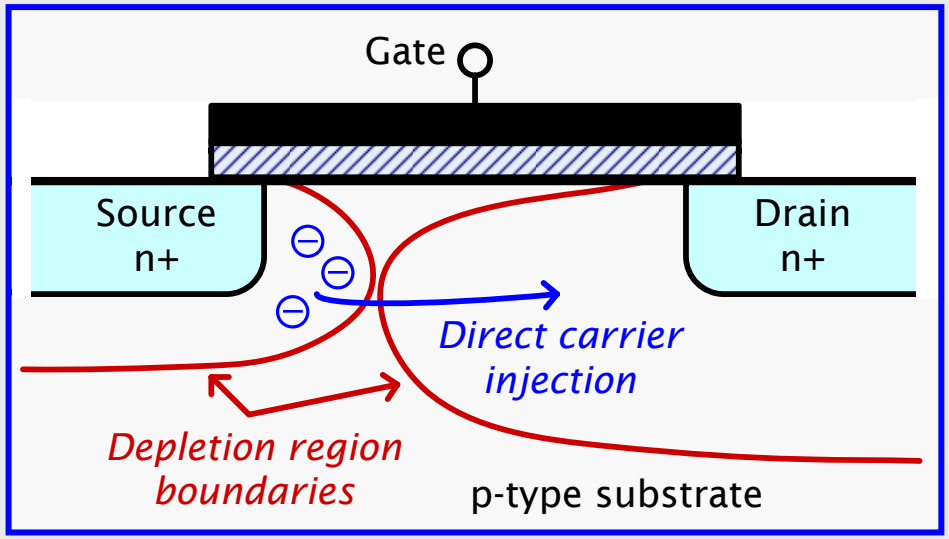




- When the device operates in saturation and the drain voltage increases, the uninverted region at the vicinity of the drain (pinch-off region) expands toward the source, shortening the length of the channel region.
- Due to the fact that resistance is proportional to length, shortening the channel decreases its resistance, causing an increase in current with increase in drain bias for a device operating in saturation.
- The effect is more pronounced when the source-to-drain separation is short (i.e. in deep-submicrometer and nanometer devices).

Channel length modulation and DIBL

- In a MOSFET device, a potential barrier exists between the source and the channel, which is controlled by the gate voltage.
- When the gate voltage is increased the barrier between the source and the channel is decreased, increasing the carriers injection from the source to the channel over the lowered barrier.
- In very short-channel devices, as drain voltage increases, more depletion is performed by the drain bias, and the electric field at the drain penetrates to the source region caused an additional decrease of the barrier at source.
- This is referred as Drain Induced Barrier Lowering (DIBL) effect.



- As a result, the device can conduct significant drain current due to an increase of carriers injected from the source.

Channel length modulation and DIBL

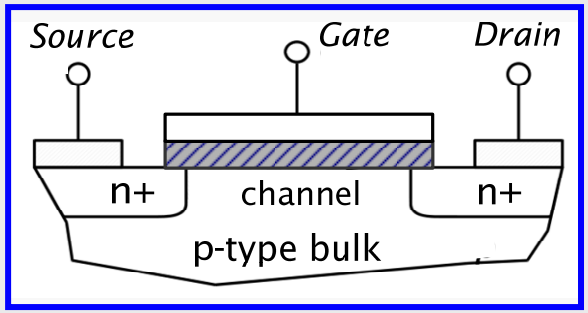
- DIBL affects the drain current vs drain bias curve, causing the current to increase with drain bias in the saturation region of operation (i.e. at high drain-to-source voltages).
- This current increase is additional to that caused by the CLM effect.
- The dependence between the drain current and the drain-source voltage in the saturation region, which is due to CLM and DIBL effects, is modeled through the inclusion of two parameters (A, D):

$$I_D = B(V_{GS} - V_{TH})^{\alpha} \Rightarrow I_D = B(V_{GS} - V_{TH})^{\alpha}[A + D(V_{DS} - V_{DO})]$$

- Given the practical target of the model, this linear dependence maintains the simplicity, while provides the required accuracy by including the influence of both effects.

Body effect

- When a positive V_{SB} is applied, the bulk is at a negative potential with respect to the source, and this increases the depletion between the source and the bulk. The minority electrons attracted from the p-type bulk have to overcome this increase in depletion, and therefore the gate voltage required to form and maintain an inversion layer or channel (i.e. threshold voltage) becomes higher.



- For the determination of the device threshold voltage when $V_{SB} > 0$, a linear approximation of the BSIM4 model expression describing the body effect is used:

$$V_{TH} = V_{TO} + K_1\left(\sqrt{\varphi_s + V_{SB}} - \sqrt{\varphi_s}\right) + K_2 V_{SB} \quad \Rightarrow \quad V_{TH} = V_{TO} + \gamma V_{SB}$$

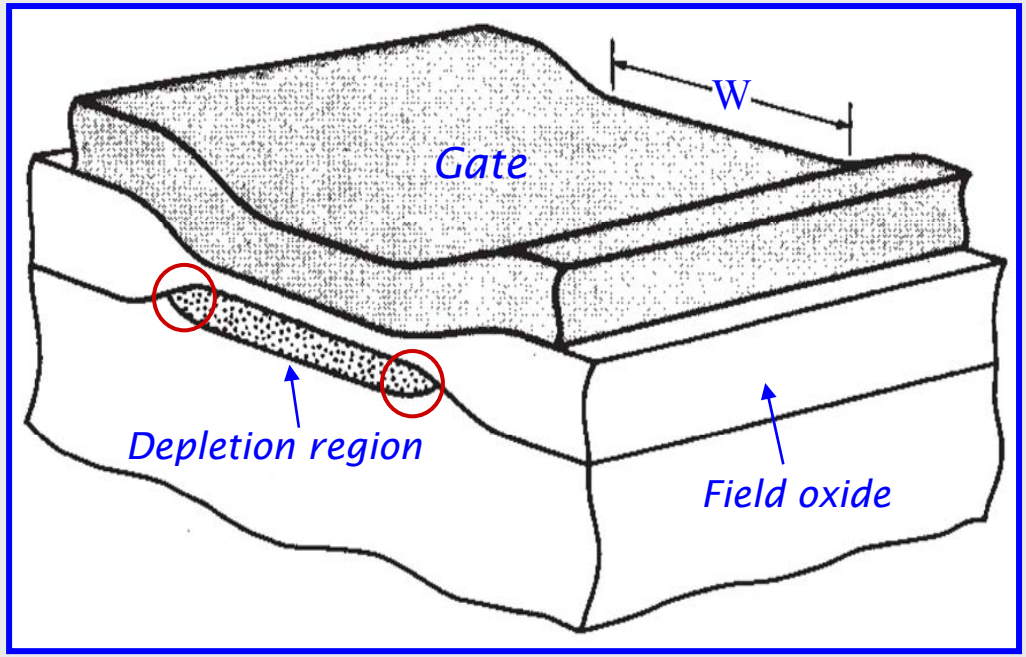
V_{TO}: threshold voltage for $V_{SB} = 0$, φ_s: inversion surface potential, K_1, K_2: BSIM4 body effect coefficients, γ: simplified body effect coefficient.

Narrow channel width effect

- In MOSFET devices, the depletion region is not limited to just the area below the thin oxide, since the polysilicon gate overlaps the field oxide on both sides of the channel region (along the width direction of the device).

- For large device widths the part of the depletion region on the sides is a small percentage of the total depletion region.

- As the device width is scaled down, the depletion charge under the gate is reduced but the fringing charge remains relatively unchanged, constituting a significant proportion.



- The gate is responsible for depleting a larger region, and because of that higher gate voltage is required, resulting in increased threshold voltage.

Narrow channel width effect

- In effect, this results in lower driving capability per width unit of narrow width devices in comparison with the driving capability per width unit of wide width devices.

$$I_D \propto (V_{GS} - V_{TH})^{\alpha}$$

- The prediction of the drain current for varying device widths (i.e. the inclusion of narrow channel width effects) is obtained by computing the transconductance parameter B of the device as a quadratic function of the device channel width:

$$B = \beta_1 + \beta_2 W + \beta_3 W^2$$

- The coefficients β_i are determined by fitting the quadratic plot to the B vs W plot, respectively (once for a given nanometer technology).

Source-drain parasitic resistance

- In long channel devices the source-drain parasitic resistance is negligible compared with the channel resistance.
- However, in very short-channel devices, it can be an appreciable fraction of the channel resistance and can therefore cause significant current degradation.
- The most severe current degradation occurs in the triode region, i.e. for low values of drain-source voltage.
- This is because the channel resistance is low (the slope of the drain current vs. drain-source voltage curve is high) under such bias conditions.

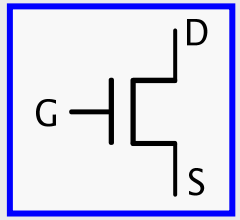
$$I_D = \frac{V_{DS}}{R_{ch} + R_{sd}} \qquad R_{ch} = \frac{V_{DS}}{I_{D-without\,R_{sd}}} \qquad I_D = \frac{I_{D-without\,R_{sd}}}{1 + (R_{sd}\, I_{D-without\,R_{sd}}) / V_{DS}}$$

- Since, the drain current dependence of the drain-source voltage is small in saturation, the current in this region is least affected by the parasitic resistance.
- The source-drain parasitic resistance effect can be taken into account model by using lower transconductance parameter (B_{tri}) in the triode region than that in the saturation region (B_{sat}).

Putting all together

- Triode or linear region:

$$I_D = B_{tri}(V_{GS} - V_{TH})^{\alpha}\left(2 - \frac{V_{DS}}{V'_{DO}}\right)\frac{V_{DS}}{V'_{DO}} \qquad V_{DS} \le V'_{DO} \qquad (1)$$



- Saturation region:

$$I_D = B_{sat}(V_{GS} - V_{TH})^{\alpha}[A + D(V_{DS} - V_{DO})] \qquad V_{DS} > V'_{DO} \qquad (2)$$

where:

$$V'_{DO} = V_{DO}\left(\frac{V_{GS} - V_{TH}}{V_{DD} - V_{TH}}\right)^{\frac{\alpha}{2}} \qquad (3) \qquad V_{TH} = V_{TO} + \gamma V_{SB} \qquad (4)$$

$$B_{tri} = \frac{I'_{DO}}{(V_{DD} - V_{TH})^{\alpha}}, \quad B_{sat} = \frac{I_{DO}}{(V_{DD} - V_{TH})^{\alpha}}, \quad A = \frac{I'_{DO}}{I_{DO}}, \quad D = \frac{1 - A}{V_{DD} - V_{DO}}$$

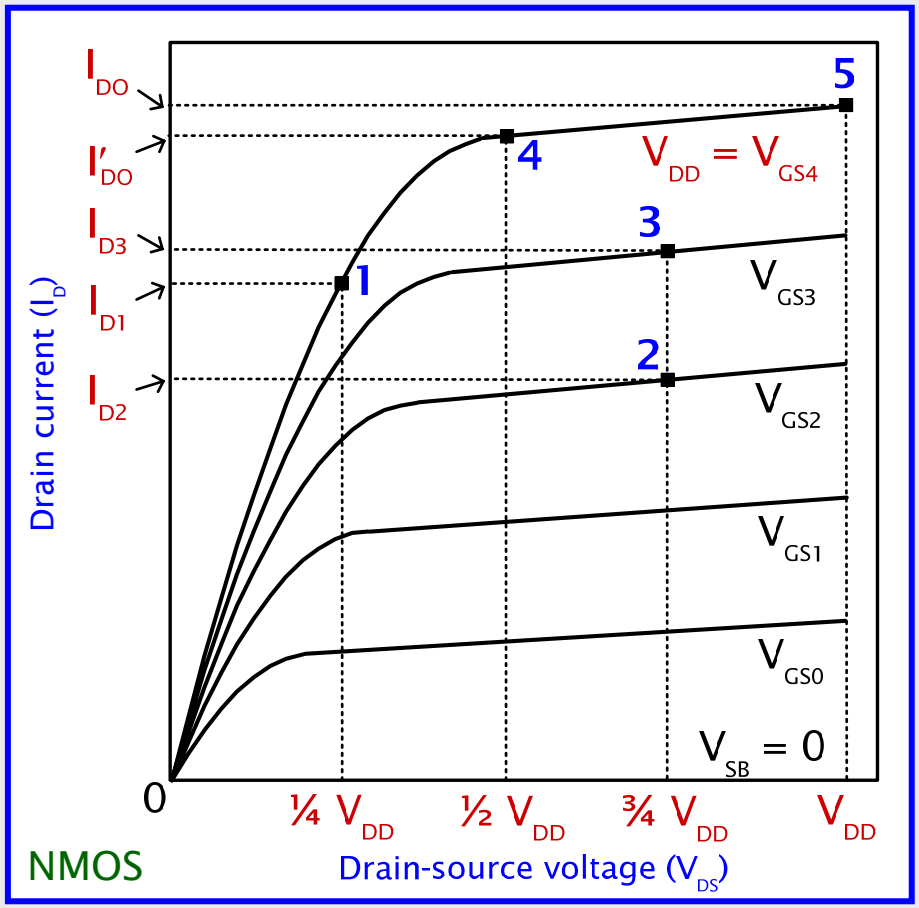
Model parameters extraction

- I_{DO}: drain current at $V_{GS} = V_{DS} = V_{DD}$
- I'_{DO}: drain current at $V_{GS} = V_{DD}$, and $V_{DS} = 1/2 \cdot V_{DD}$ (for the NMOS device), $V_{DS} = 2/3 \cdot V_{DD}$ (for the PMOS device)
- By simulating the device, we obtain I_{DO} and I'_{DO} and consequently B_{tri} and B_{sat} for the minimum & the maximum used device width, as well as for few intermediate width values, in order to fit B_{tri} vs W and B_{sat} vs W plots to quadratic plots, such as:

$$B_{tri} = \beta_{t1} + \beta_{t2}W + \beta_{t3}W^2$$

$$B_{sat} = \beta_{s1} + \beta_{s2}W + \beta_{s3}W^2$$

once for a given nanometer technology.



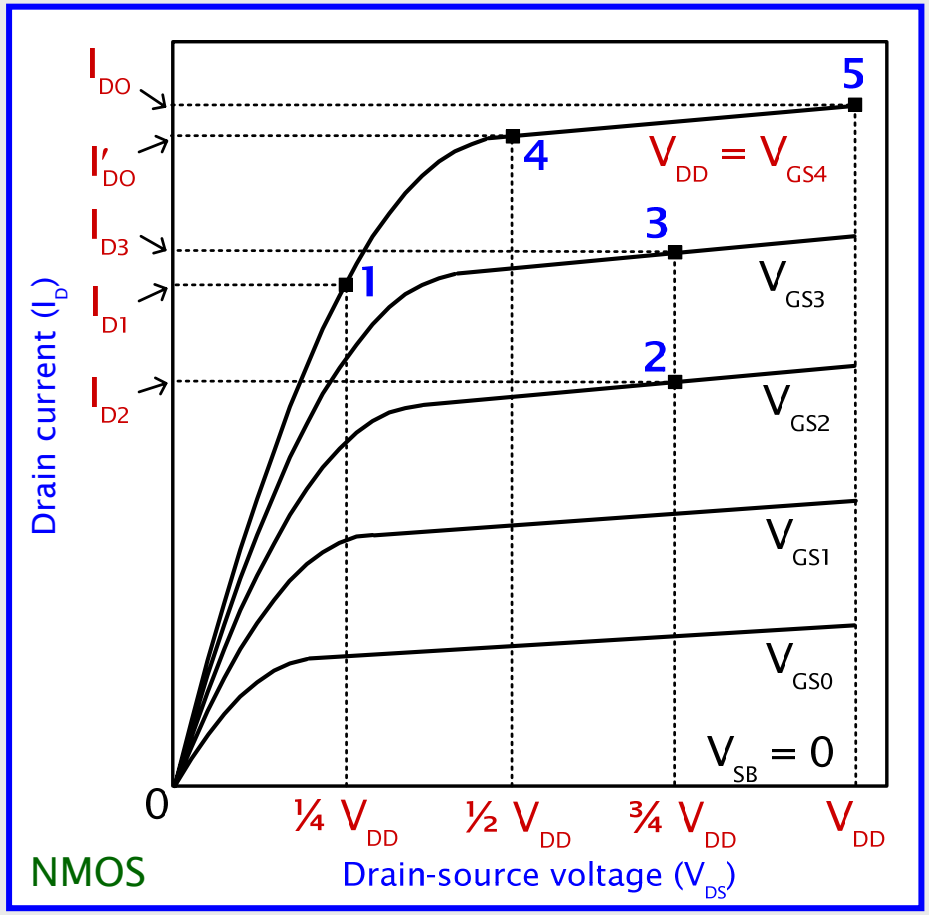
Model parameters extraction

- α is extracted from (2) by using fitting points 2 and 3:

$$\alpha = \frac{\ln\left(\frac{I_{D3}}{I_{D2}}\right)}{\ln\left(\frac{V_{GS3} - V_{TO}}{V_{GS2} - V_{TO}}\right)}$$

- V_{DO} (saturation voltage at $V_{GS}=V_{DD}$) is computed by combining (1) and (3) and using fitting point 1:

$$V_{DO} = \frac{I'_{DO}\, V_{DD} + V_{DD}\sqrt{I'_{DO}(I'_{DO} - I_{D1})}}{4 I_{D1}}$$



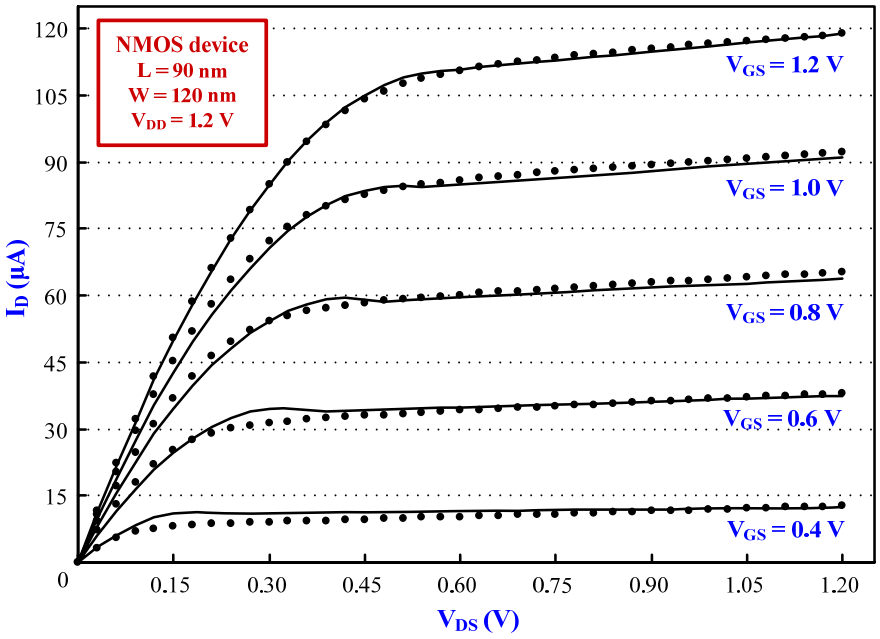
Model parameters (90-nm technology)

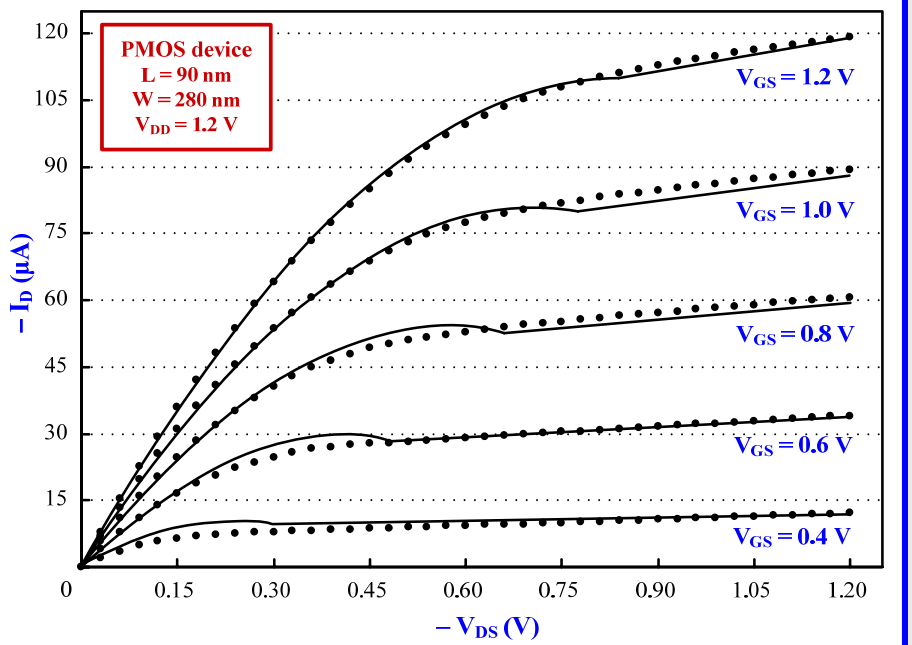
Model parameters	W [nm]	α [ ]	I_{DO} [μA]	I'_{DO} [μA]	V_{DO} [V]	V_{TO} [V]	γ [ ]	φ_s [V]	K_1 [$V^{1/2}$]	K_2 [ ]
NMOS device	120	1.072	118.78	110.55	0.579	0.290	0.196	0.873	0.383	0.01
PMOS device	280	1.298	119.10	109.86	−0.847	−0.236	0.201	0.865	0.340	−0.01

Transconductance coefficients	b_{t1} [A / V^{α}]	b_{t2} [$A / m \cdot V^{\alpha}$]	b_{t3} [$A / m^2 \cdot V^{\alpha}$]	b_{s1} [A / V^{α}]	b_{s2} [$A / m \cdot V^{\alpha}$]	b_{s3} [$A / m^2 \cdot V^{\alpha}$]
NMOS device	-13.393×10^{-6}	1131.84	-8.2901×10^{6}	-8.420×10^{-6}	1165.49	-3.4340×10^{5}
PMOS device	-22.373×10^{-6}	1177.54	-2.5811×10^{8}	-24.280×10^{-6}	1276.82	-2.8003×10^{8}

Experimental results

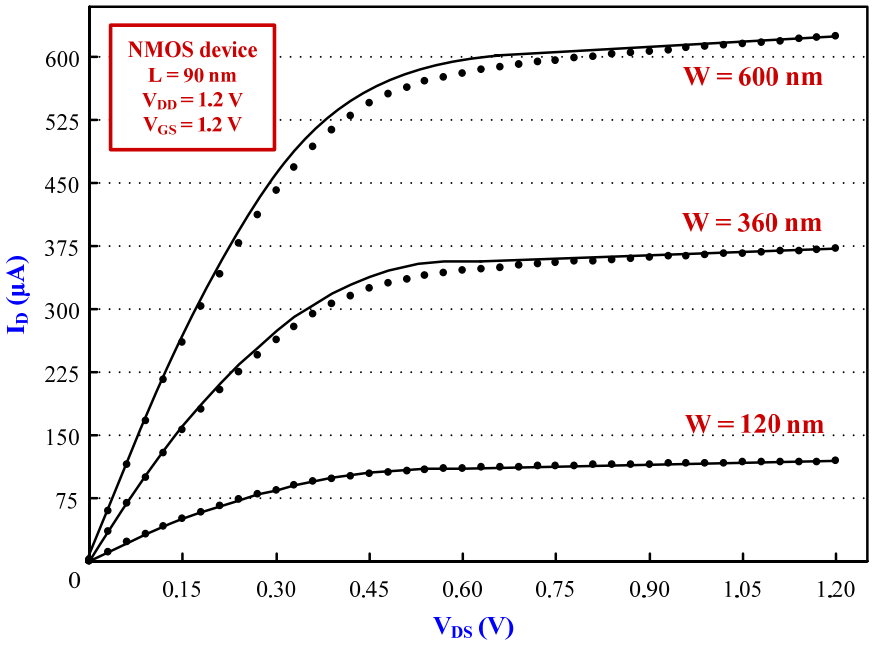
The I-V results obtained from the model (continuous lines) show very good agreement with those produced by BSIM4 HSPICE simulations (dots)

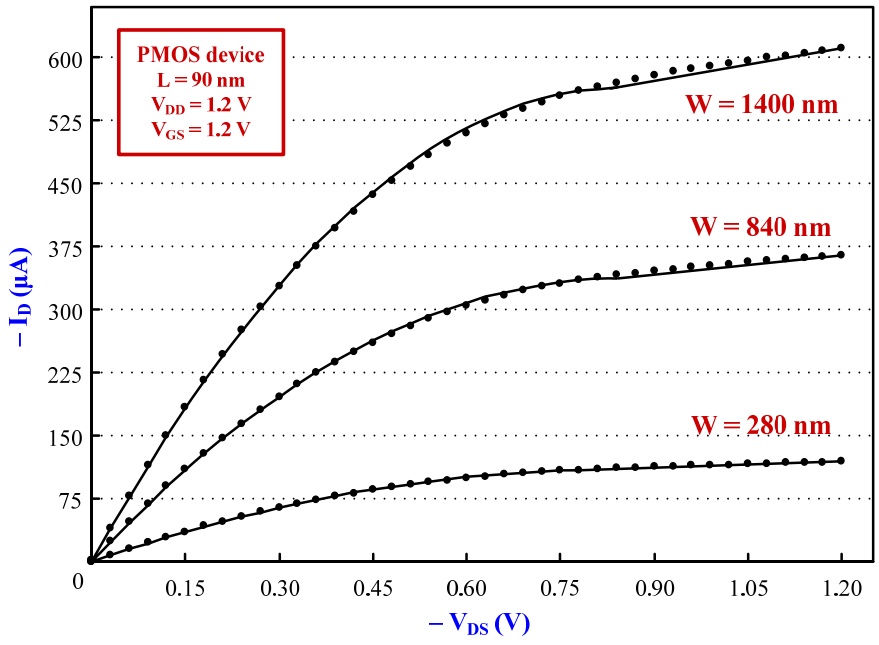




Experimental results

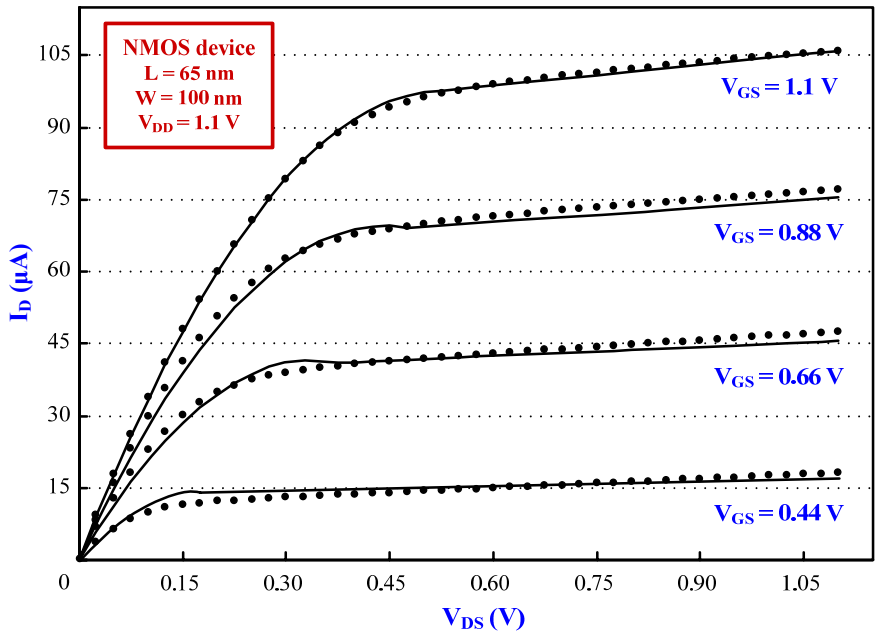
The presented model has been validated for varying device widths and the results are similarly accurate

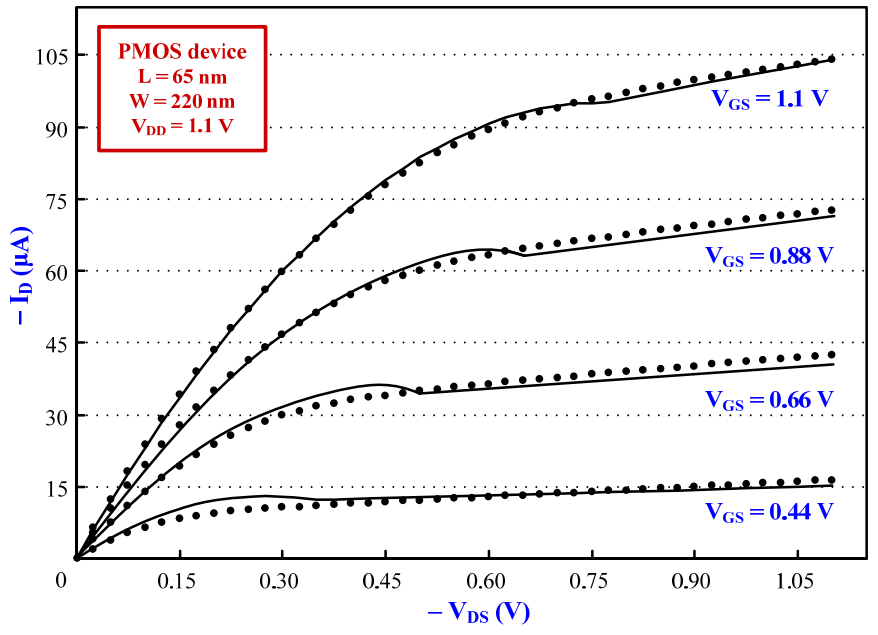




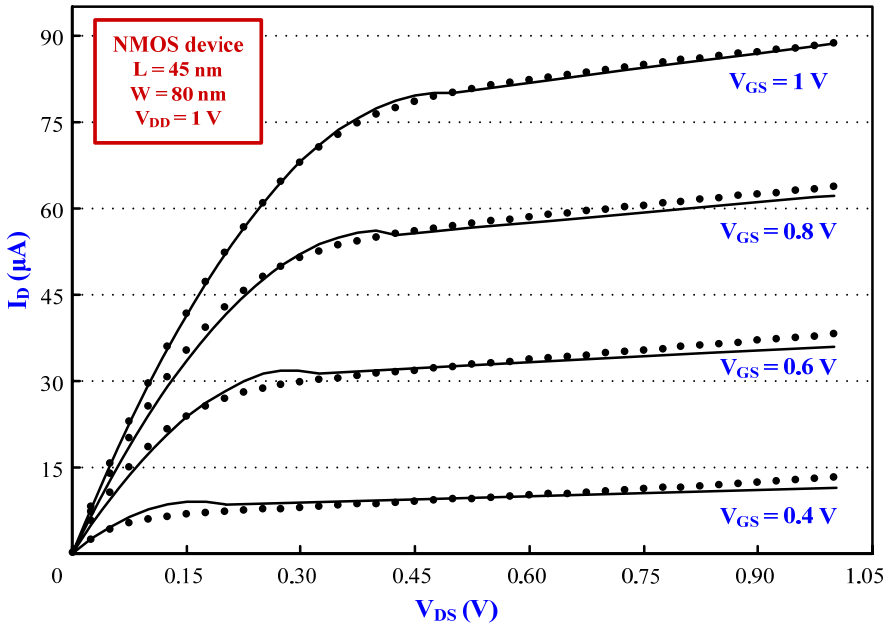
Experimental results

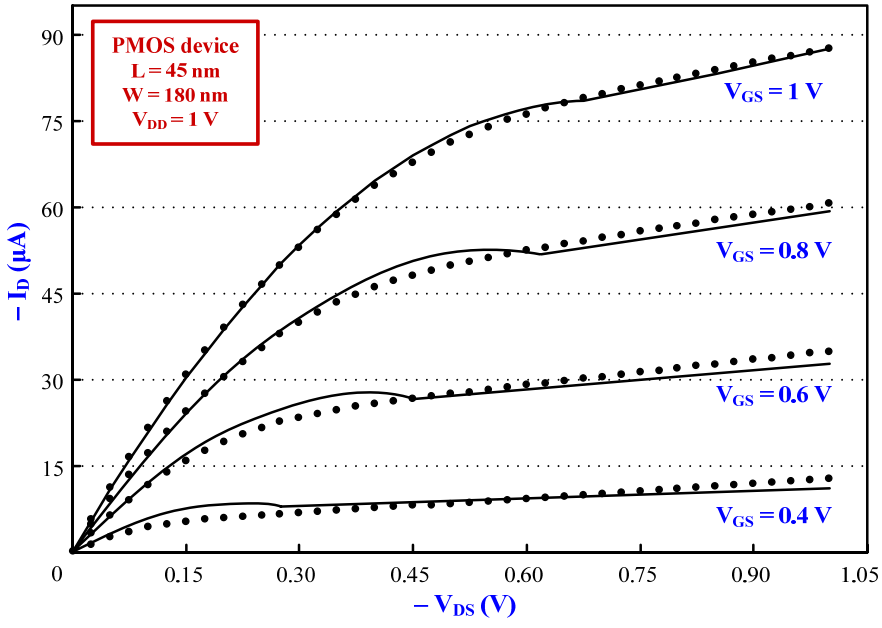
The I-V results of the model has been compared with BSIM4 HSPICE simulations for two additional sub-100-nm CMOS technologies (65-nm, 45-nm)





Experimental results

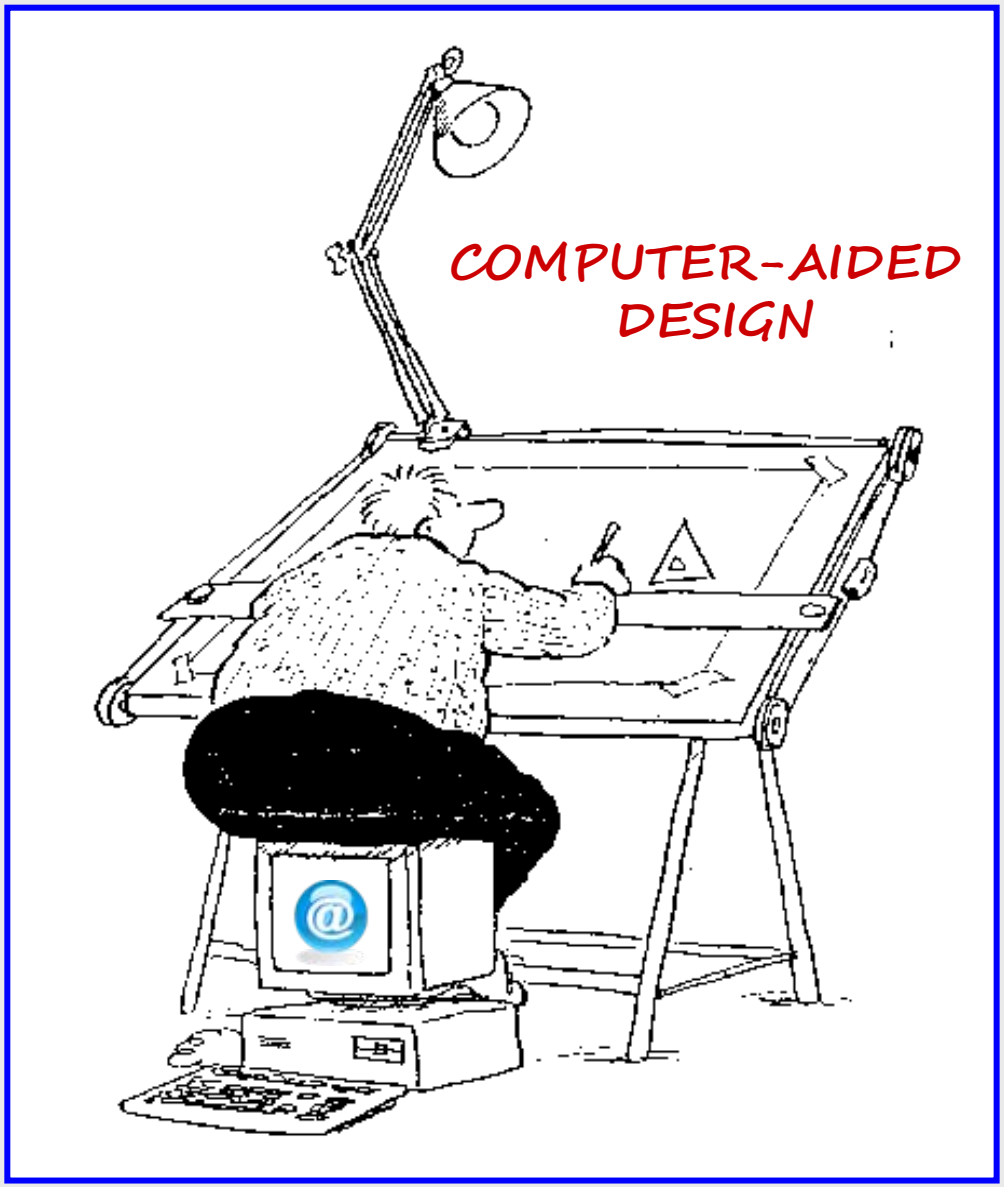




- The choice to avoid complex current dependences on drain-source voltage led to a discontinuity at the boundary between triode and saturation regions.
- Such discontinuity could cause problems in circuit simulators, which require continuity of the functions as well as of the derivatives, but should not cause problems in case of use for analytical computation of design parameters.

Conclusion

- The modeling of CMOS primitive circuits' behavior, in terms of analysis and computation of their transient response and energy dissipation, requires:
  - Deep understanding of the circuits' operational characteristics as well as of the devices' operational characteristics.
  - Inventiveness in deciding about the device and circuit modeling approach.
  - Efficient and reasonable assumptions and approximations.
  - Good background in required mathematics.
  - Efficient use of the simulation tools (e.g. HSPICE, Mathematica, …).
  - Making continuous tradeoffs between accuracy and computational efficiency.
- …and of course the eventual target of such modeling effort is to help in creating more efficient Computer-Aided Design (CAD) tools.



Thanks for your attention

3rd IEEE Greece GOLD A.G. Summer School

Athena

1-6 July, Pyrgos, Greece 2012

Algorithms, Telecom, Hardware, Encryption, Networks, Applications