MICRO-461
Low-power Radio Design for the IoT

5. Modeling of active and passive devices at RF

The MOS Transistor at RF

Christian Enz

Integrated Circuits Lab (ICLAB), Institute of Microengineering (IMT), School of Engineering (STI)

Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland

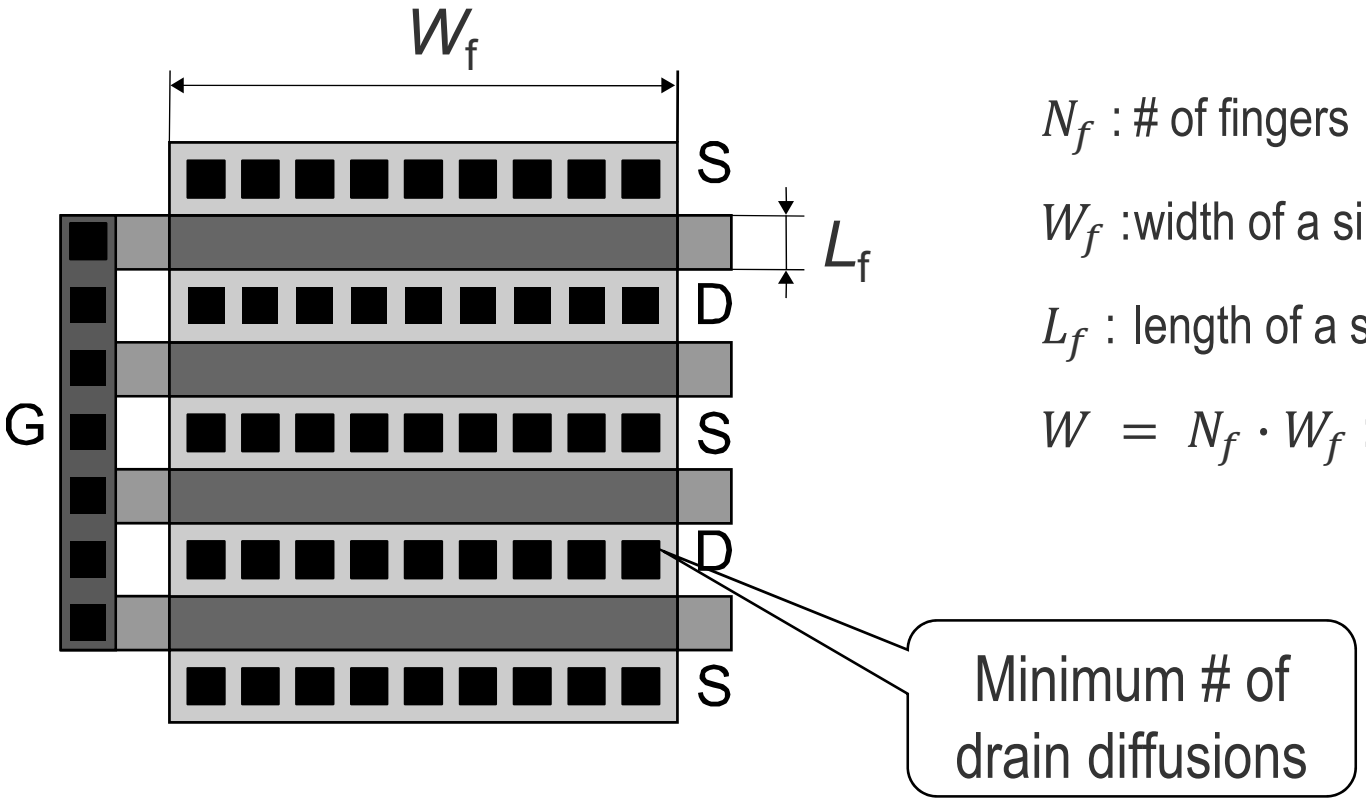
EPFL

Outline

- **Introduction**
- Transistor Figures-of-Merit (FoM)
- Equivalent Circuit at RF
- Large-signal Model at RF

RF MOS Transistor Structure and Layout

- RF MOS Transistors are usually **large** devices
- Implemented as **multi-finger** devices due to the “narrow-line effect” limiting the transistor width



N_f : # of fingers

W_f :width of a single finger

L_f : length of a single finger

$W = N_f \cdot W_f$: total width

What Changes at RF?

- Transistor characteristics such as gain and transconductance start to degrade due to **intrinsic** frequency limitations and **extrinsic** parasitics
- Frequency limit of **intrinsic part** set by frequency ω_{qs} delimiting **quasi-static** (QS) and **non-quasi-static** (NQS) operation

$$\omega_{qs} = \underbrace{\omega_{spec}}_{\substack{\text{technology}\\ \text{and geometry}\\ \text{dependent}}} \cdot \underbrace{\Omega_{qs}(q_s, q_d)}_{\substack{\text{unitless and}\\ \text{only bias}\\ \text{dependent}}}$$

with $\omega_{spec} \triangleq \frac{2\mu \cdot U_T}{L^2}$ and $\Omega_{qs} = 15\frac{(q_s + q_d + 1)^3}{4q_s^2 + 4q_d^2 + 12q_s q_d + 10q_s + 10q_d + 5}$

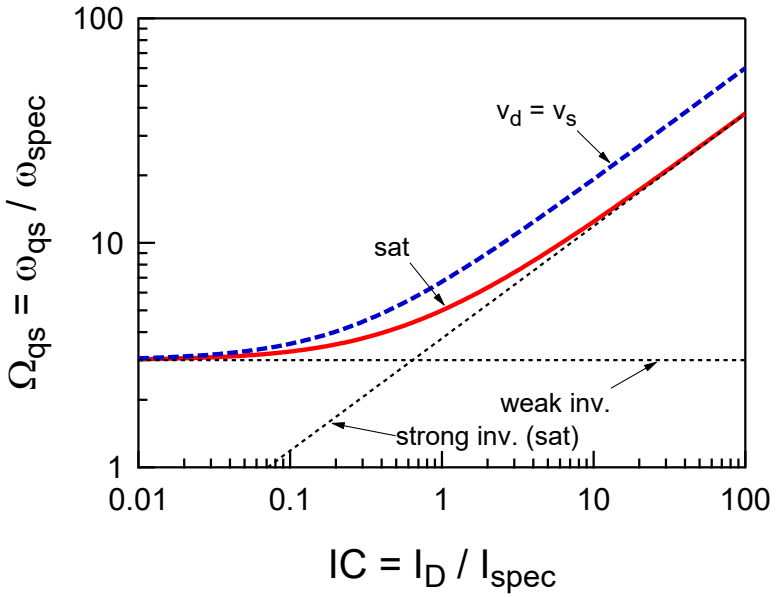
- In saturation ($q_s \gg q_d$)

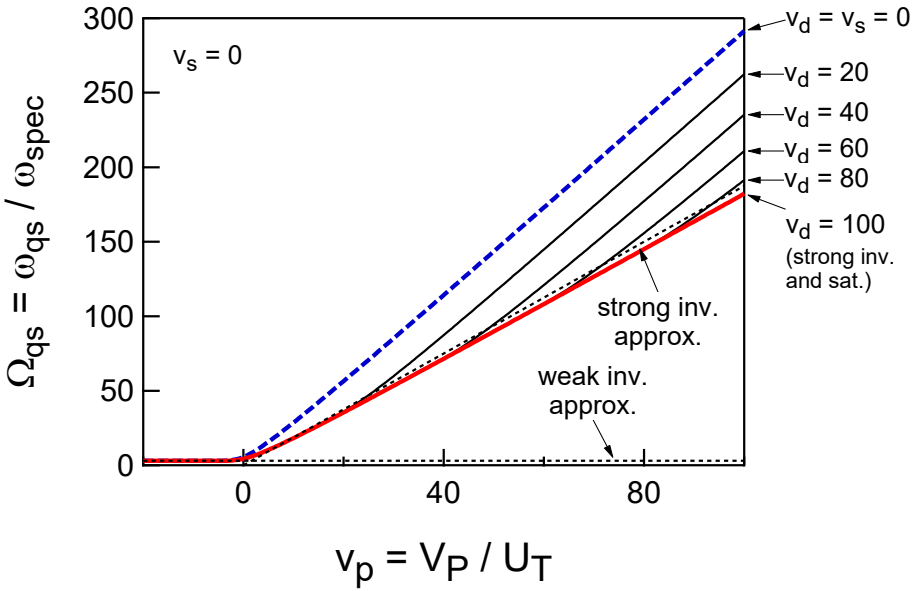
$$\Omega_{qs} \cong 15\frac{(q_s+1)^3}{4q_s^2 + 10q_s + 5} = \begin{cases} 3 & \text{WI} \\ \frac{15}{4}q_s = \frac{15}{4}\sqrt{i_f} = \frac{15}{8}\frac{V_P - V_S}{U_T} & \text{SI} \end{cases}$$

- To avoid any degradation due to NQS, $\omega_{qs} \approx 7x \; to \; 10x$ operating frequency
- Achieved by sufficiently large bias and/or reduced length

What Changes at RF?

$$\Omega_{qs} \triangleq \frac{\omega_{qs}}{\omega_{spec}} \cong 15\frac{(q_s+1)^3}{4q_s^2+10q_s+5} = \begin{cases} 3 & \text{WI (sat)} \\ \frac{15}{4}q_s = \frac{15}{4}\sqrt{i_f} = \frac{15}{8}\frac{V_P - V_S}{U_T} & \text{SI (sat)} \end{cases}$$





📖 C. C. Enz and E. A. Vittoz, *Charge-Based MOS Transistor Modeling - The EKV Model for Low-Power and RF IC Design*, John Wiley, 2006.

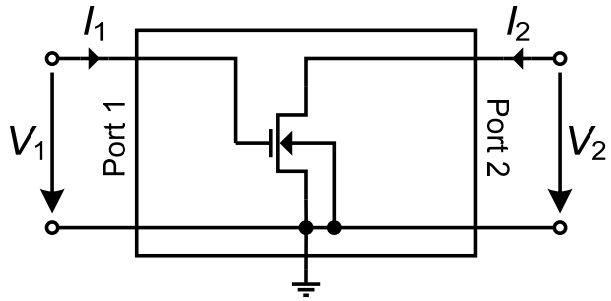
What Changes at RF?

- Limitations due to **extrinsic parasitics** are strongly **dependent on the layout**
- Usually frequency limitations are due to **extrinsic capacitances** and particularly the capacitance at the drain (C_{BDj} and C_{GDe})
- Some limitations are characterized by several **figures-of-merit (FoM)** such as:
  - Transit frequency F_t
  - Maximum frequency of oscillation F_{max}
  - Minimum noise figure NF_{min}

Outline

- Introduction
- **Transistor Figures-of-Merit (FoM)**
- Equivalent Circuit at RF
- Large-signal Model at RF

Transit Frequency – Definition



- The small-signal current gain h_{21} is defined as

$$h_{21} \triangleq \left.\frac{I_2}{I_1}\right|_{V_2=0} = \frac{Y_{21}}{Y_{11}} = \frac{G_m - j\omega(C_m + C_{GD})}{j\omega C_G} \cong \frac{G_m}{j\omega C_G} = \frac{\omega_t}{j\omega}$$

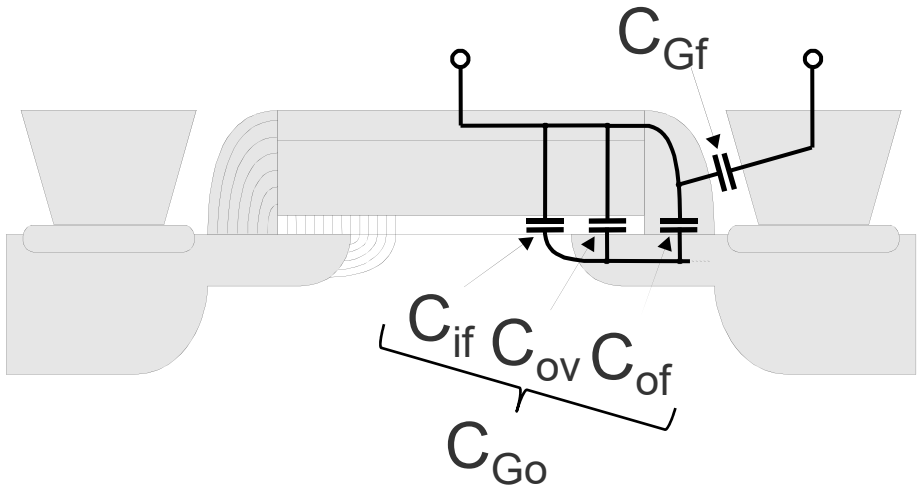
- The **transit frequency** is then defined as the frequency for which the current gain magnitude becomes unity

$$F_t = \frac{1}{2\pi} \cdot \frac{G_m}{C_G}$$

- where $C_G = C_{Gi} + C_{Ge}$ is the **total gate capacitance** made of an **intrinsic part** C_{Gi} and **extrinsic part** C_{Ge}
- The intrinsic gate capacitance is proportional to $W \cdot L \cdot C_{ox}$ and bias dependent

📖 C. C. Enz and E. A. Vittoz, *Charge-Based MOS Transistor Modeling - The EKV Model for Low-Power and RF IC Design*, John Wiley, 2006.

Transit Frequency and Extrinsic Gate Capacitance



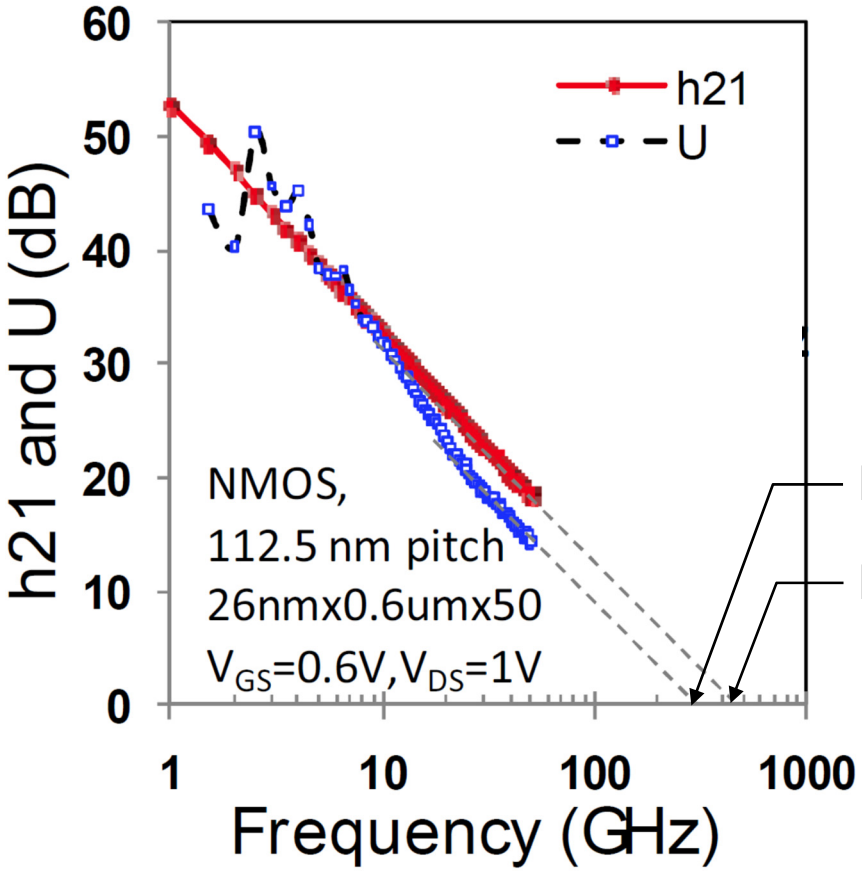
- The **extrinsic gate capacitance** C_{Ge} is made of the **overlap capacitance** C_{Go} and the **fringing capacitance** C_{Gf}

$$C_{Ge} = C_{Go} + C_{Gf} = W \cdot \left(C_{GoW} + C_{GfW}\right) = W \cdot C_{GeW}$$

- where C_{GeW} is the total extrinsic capacitance per unit width
- Note that C_{Go} and C_{Gf} scale with W but not with L

Measured Current and Unilateral Power Gains

- F_t and F_{max} are obtained from measurements for one operating point by simple extrapolation (32nm bulk CMOS process in the example below)

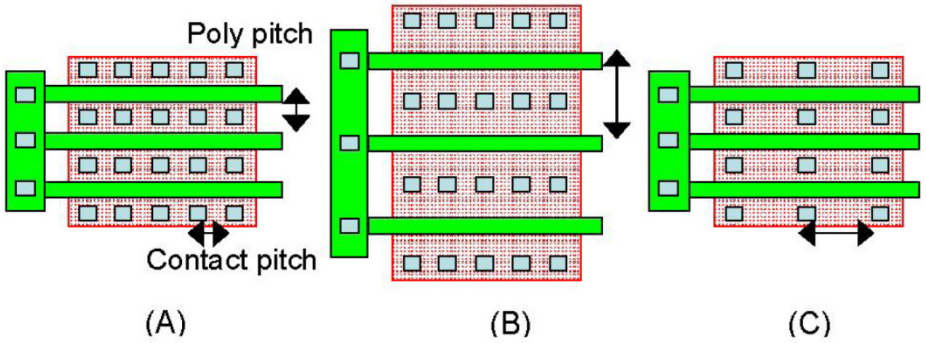


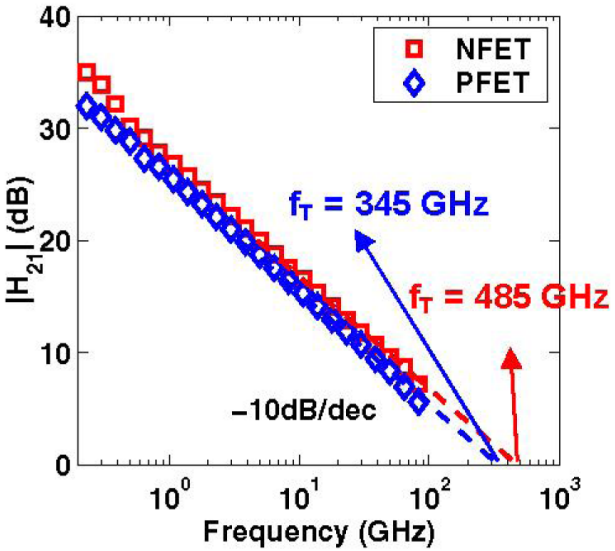
$$\omega_t = -\Im\{h_{21}\} \cdot \omega_{spot}$$

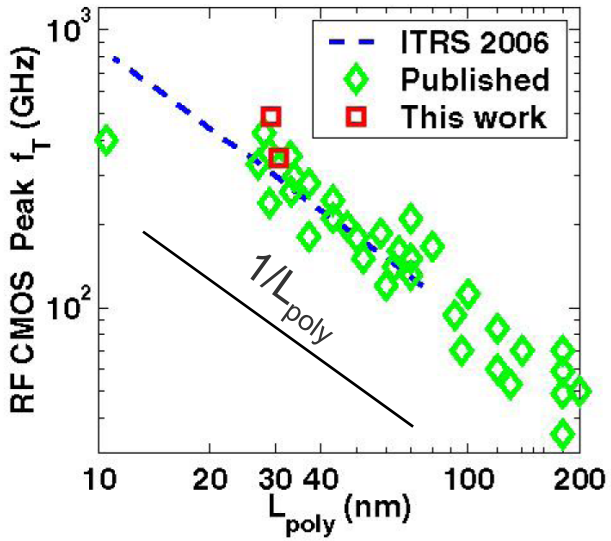
Where ω_{spot} is a sufficiently small frequency (typically 1GHz) at which the current gain shows a −20dB/dec slope

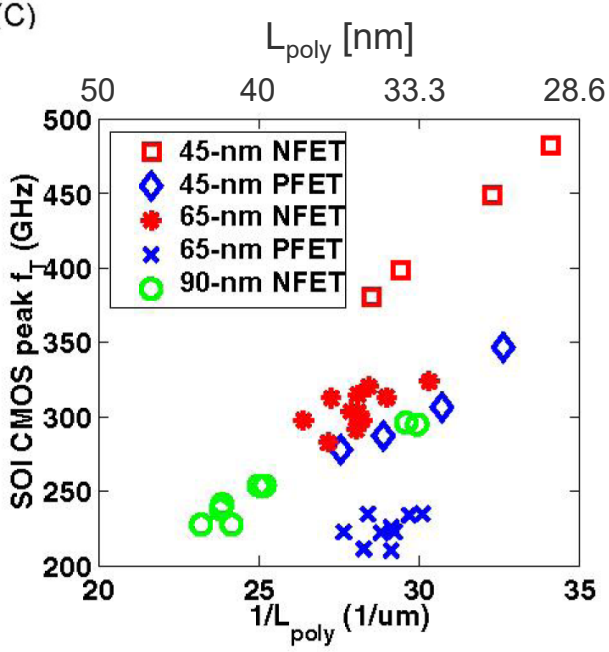
P. VanDerVoorn, *et al.*, Symposium on VLSI Technology, 2010

Transit Frequency of a 45 nm SOI CMOS Process



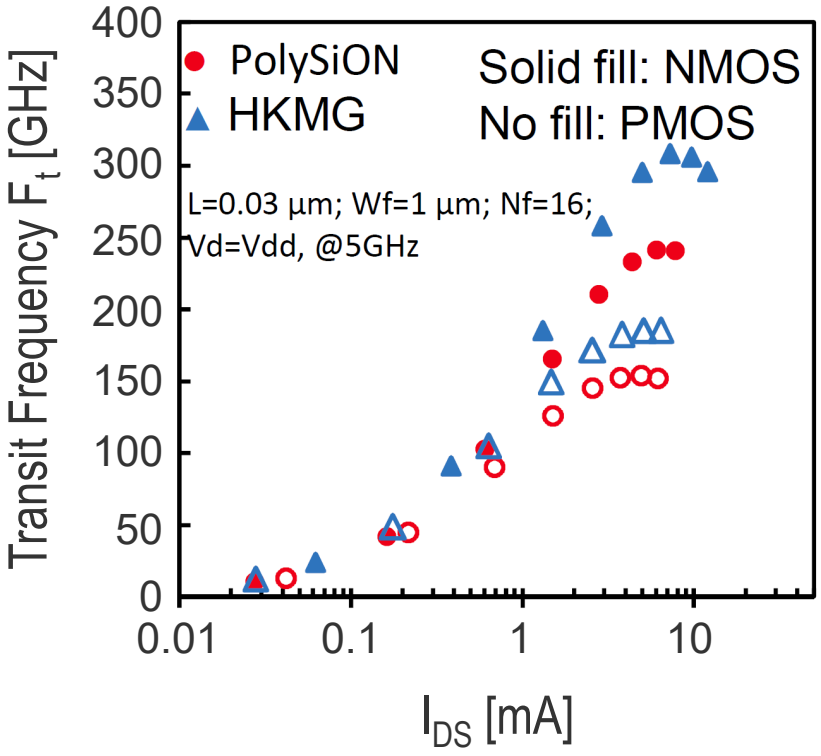


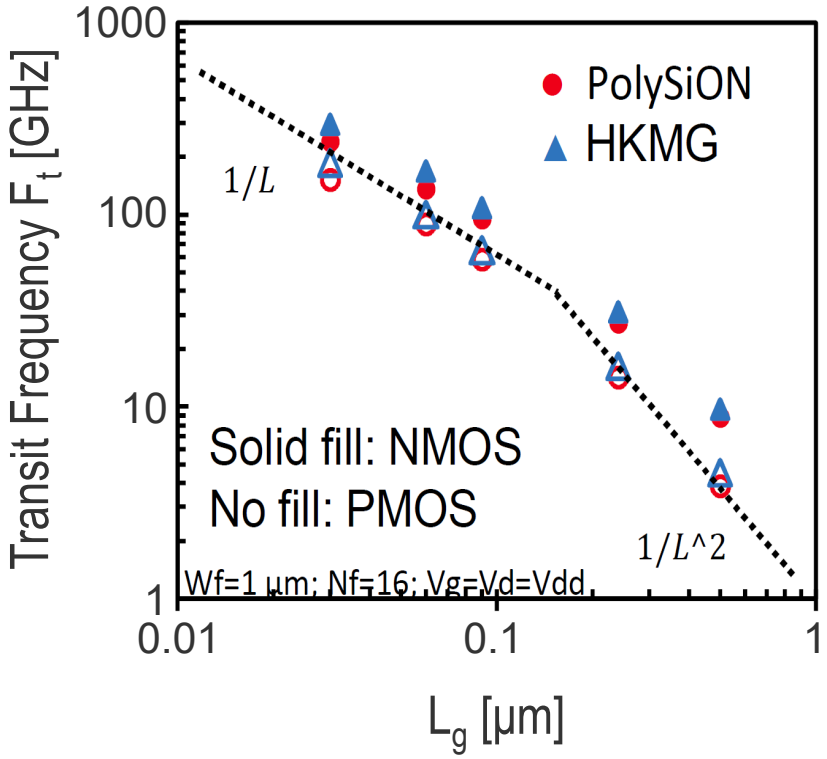




📖 Sungjae Lee, *et al.*, IEDM 2007

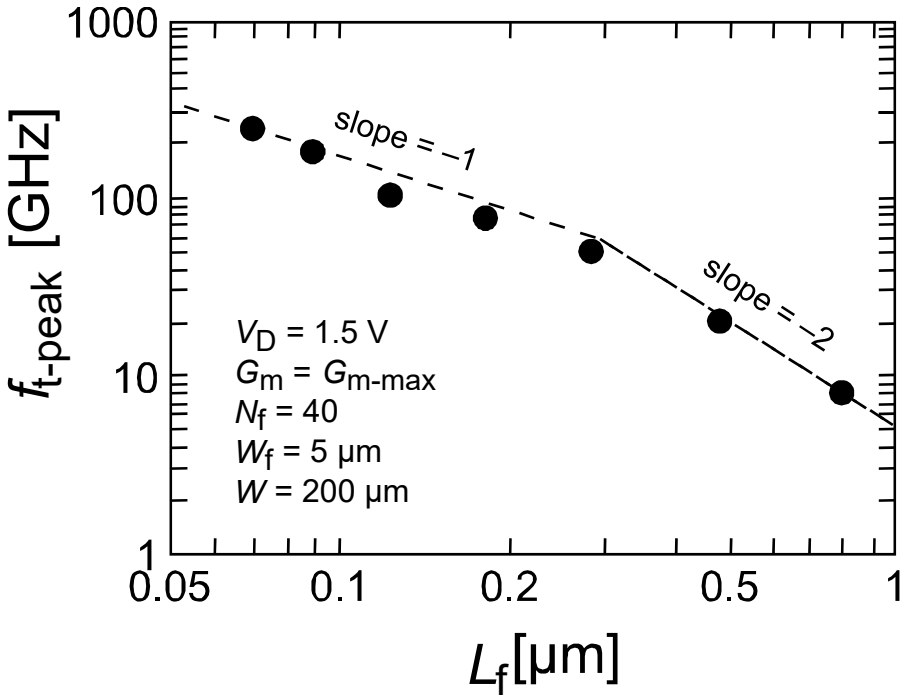
Transit Frequency of a 28nm Bulk CMOS Process





📖 K. W. J. Chew, *et al.*, RFIC, 2015

Transit Frequency Scaling



- Scaling of ω_t is affected by short-channel effects such as **velocity saturation**

$$\left.\begin{aligned} G_{msat} &\cong W \cdot C_{ox} \cdot v_{sat} \\ C_G &\cong W \cdot L_f \cdot C_{ox} \end{aligned}\right\} \Longrightarrow \omega_t \triangleq \frac{G_{msat}}{C_G} \cong \frac{v_{sat}}{L_f}$$

- Scales only as $1/L_f$ when velocity saturation is present instead of $1/L_f^2$

📖 H. S. Momose *et al.*, IEDM 1996.

MOSFET Model Valid for RF from Weak to Strong Inversion

- To take advantage of the highest transit frequency reached at minimum length MOSFET operating at RF have usually a minimum length
- The high transit frequency achieved with advanced CMOS technologies can be traded-off with power consumption by shifting the operating point from strong inversion to moderate or eventually even weak inversion
- It is therefore crucial to have a MOSFET model that accounts for velocity saturation and is valid from weak to strong inversion

Inversion Coefficient Definition

- Overdrive voltage $V_G - V_{T0}$ or $V_{GS} - V_T$ **not convenient for weak inversion**
- Replaced by the **inversion coefficient** IC characterizing the global level of inversion of the transistor in **saturation**

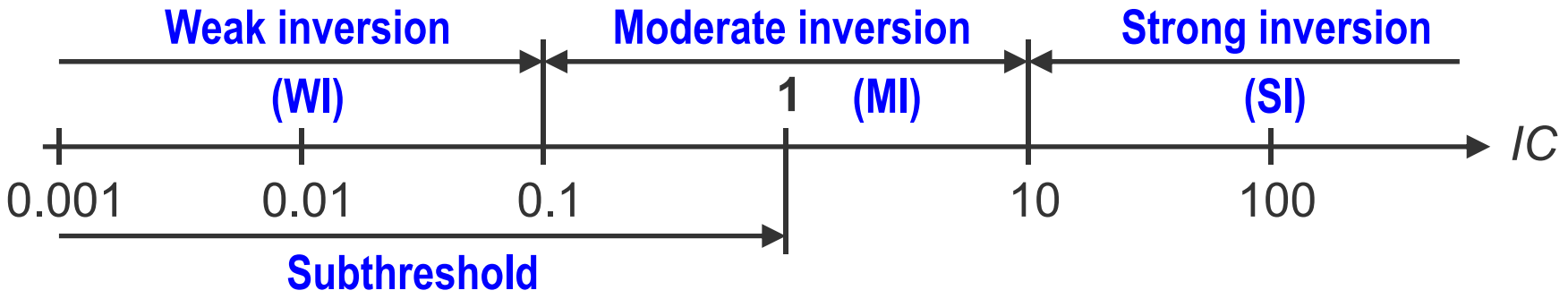
$$IC \triangleq \frac{I_D|_{saturation}}{I_{spec}}$$

Typical values of $I_{spec\square}$ for 28-nm:
750 nA for NMOS
200 nA for PMOS

- Where the specific current I_{spec} is defined as

$$I_{spec} \triangleq I_{spec\square} \cdot \frac{W}{L} \text{ with } I_{spec\square} \triangleq 2n\mu C_{ox} U_T^2 \text{ and } U_T \triangleq \frac{kT}{q}$$

- The different regions of operation in **saturation** can then be defined as



C. C. Enz, F. Krummenacher, and E. A. Vittoz (EKV), Analog Integrated Circuits and Signal Processing Journal, vol. 8, pp. 83-114, July 1995.

Simplified EKV Charge-based Model (in saturation)

- The **normalized drain current** in saturation or **inversion coefficient** is given by

$$IC = \frac{I_D|_{\text{saturation}}}{I_{spec}} = \frac{4(q_s^2 + q_s)}{2 + \lambda_c + \sqrt{\lambda_c^2(2q_s + 1)^2 + 4(1 + \lambda_c)}}$$

- $q_s \triangleq Q_i(x = 0)/Q_{spec}$ is the **normalized inversion charge** at the source where $Q_{spec} = -2nC_{ox}U_T$
- λ_c is the **velocity saturation** (VS) parameter corresponding to the fraction of the channel under full VS

$$\lambda_c = \frac{L_{sat}}{L} \text{ with } L_{sat} = \frac{2\mu_0 U_T}{v_{sat}} = \frac{2U_T}{E_c}$$

- q_s is related to the gate and source voltage according to

$$v_p - v_s = \ln(q_s) + 2q_s \text{ with } v_p = \frac{V_P}{U_T} = \frac{V_G - V_{T0}}{nU_T}\, v_s = \frac{V_S}{U_T}\, U_T = \frac{kT}{q}$$

- Only **requires the following 4 parameters**: n, $I_{spec\square}$, V_{T0}, L_{sat}

C. Enz, F. Chicco, and A. Pezzotta, IEEE Solid-State Circuits Magazine, vol. 9, no. 3, pp. 26-35, Summer 2017.

Effect of VS on the Drain Current in SI

- In SI and saturation, the model reduces to

$$i_d = \frac{2q_s^2}{1+\sqrt{1+(\lambda_c q_s)^2}}$$

- L_{sat} represents the **portion of the channel that is under full VS**
- For very short channel and/or high overdrive voltage

$$\lambda_c q_s = \frac{\mu_0}{v_{sat}} \cdot \frac{V_P - V_S}{L} \gg 1 \implies i_d \cong \frac{2q_s}{\lambda_c}$$

- Remembering that in SI $q_s \cong (V_P - V_S)/(2U_T)$ leads to the denormalized drain current given by

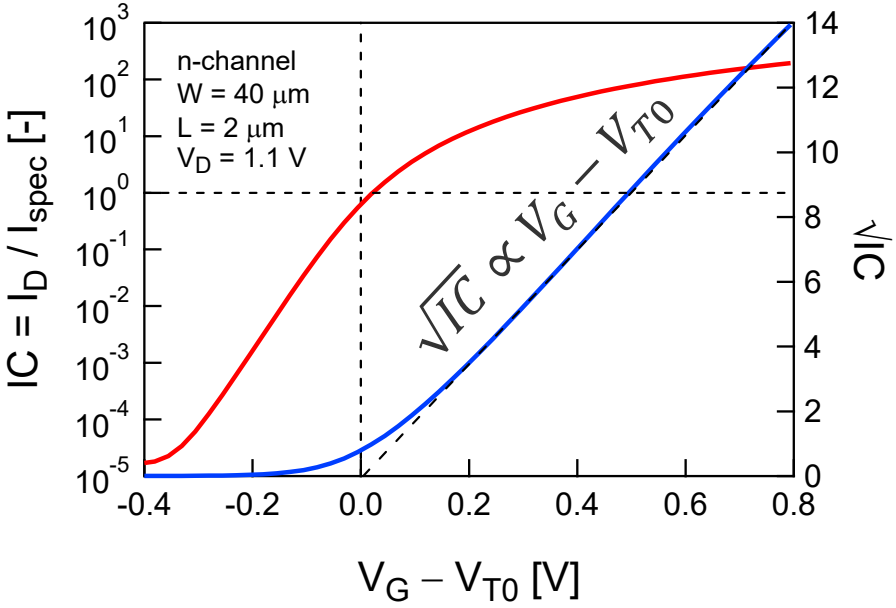
$$I_D \cong WnC_{ox}v_{sat}(V_P - V_S) = WC_{ox}v_{sat}(V_G - V_{T0} - nV_S)$$

- The current becomes a **linear function** of the charge and therefore of the overdrive voltage and also **independent of the length**

Effect of VS on the Drain Current (40nm Process)

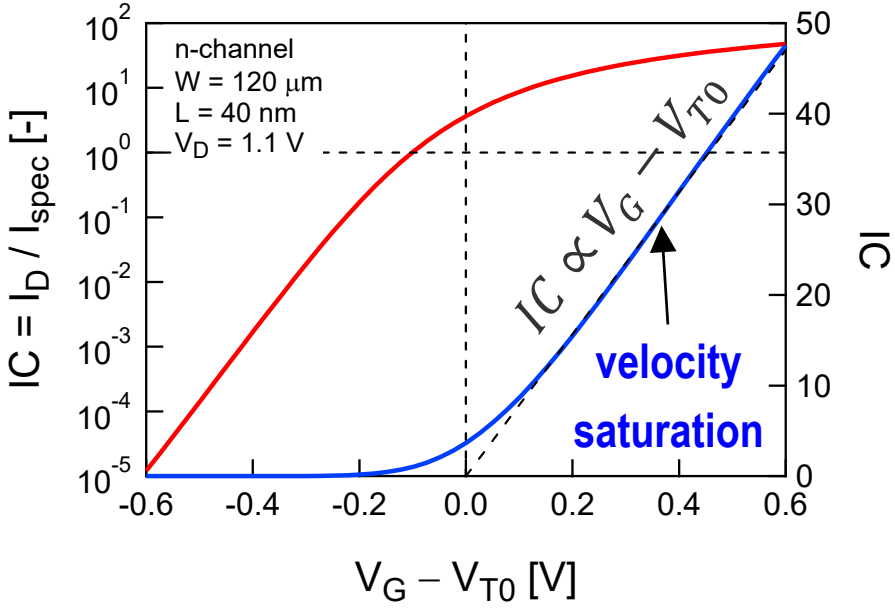
Long-channel

$$I_D \cong \frac{\beta}{2n}(V_G - V_{T0})^2$$



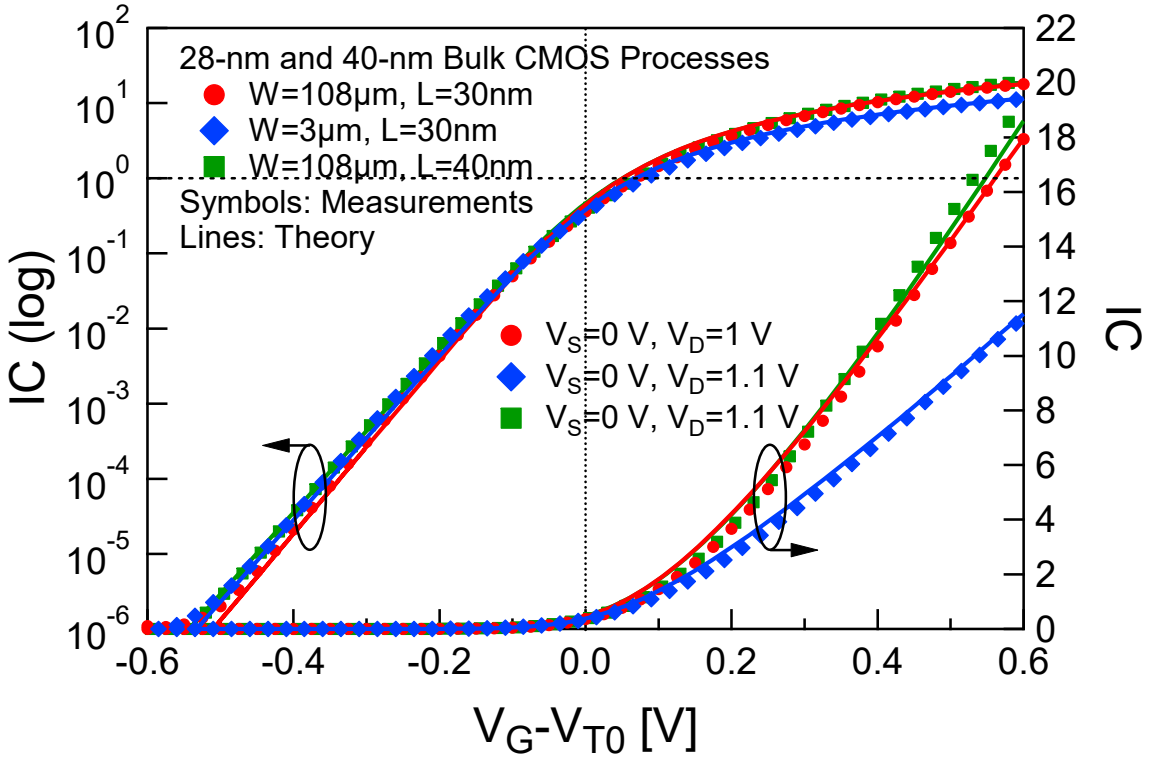
Short-channel

$$I_D \cong WC_{ox}v_{sat}(V_G - V_{T0})$$



- Velocity saturation has a **strong impact** on the drain current in **strong inversion**
- The current becomes proportional to $V_G - V_{T0}$
- Hence the gate and source transconductances become **independent of the current** (and **independent of the length**)

Drain Current for 28 and 40-nm Bulk CMOS Processes



- Simple model validated on **28-nm and 40-nm bulk CMOS processes** over **more than 6 decades of current** despite only requiring **few parameters, namely**:

$$n, I_{spec\square}, V_{T0}, L_{sat}$$

C. Enz, F. Chicco, and A. Pezzotta, IEEE Solid-State Circuits Magazine, vol. 9, no. 3, pp. 26-35, Summer 2017.

Effect of VS on the Transconductance in SI

- The effect of VS on the **source transconductance** in SI is given by

$$g_{ms} \triangleq \frac{G_{ms}}{G_{spec}} = \frac{q_s}{\sqrt{1+(\lambda_c q_s)^2}}$$

- where $G_{ms} = n \cdot G_m$ is the source transconductance and $G_{spec} \triangleq I_{spec}/U_T = 2n\beta U_T$
- For $\lambda_c \cdot q_s \gg 1$, g_{ms} saturates to $1/\lambda_c$

$$g_{ms} \cong \frac{1}{\lambda_c} = \frac{L}{L_{sat}} \text{ in SI and saturation}$$

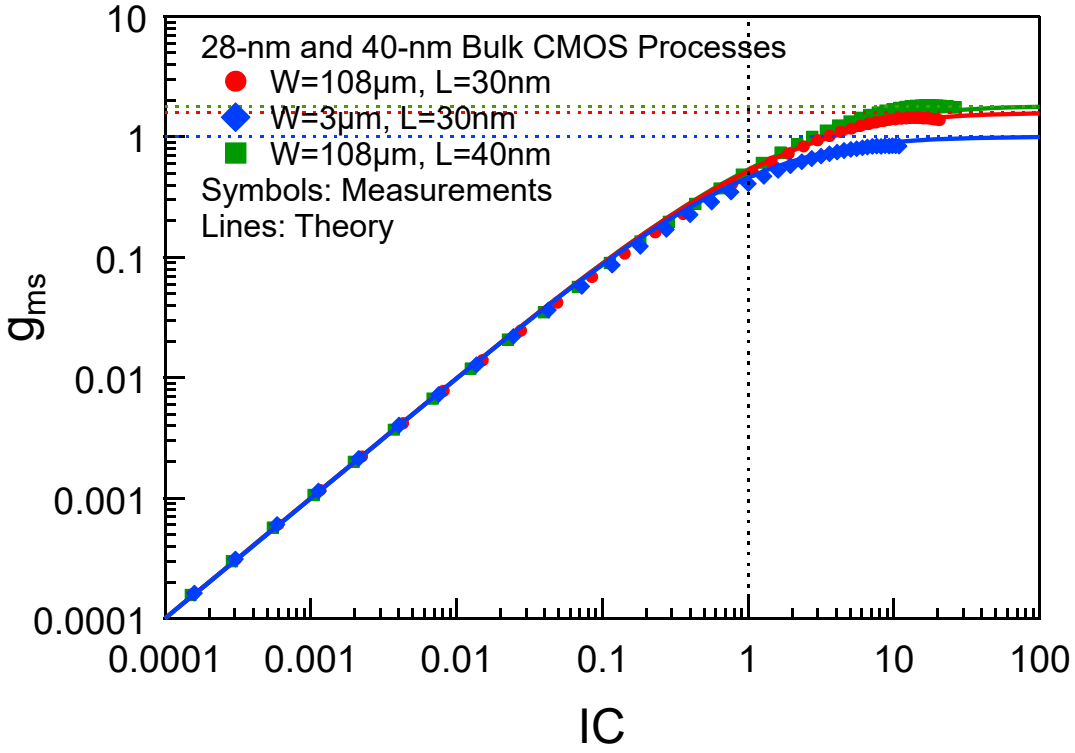
- or in denormalized form

$$G_{ms} \cong \frac{G_{spec}}{\lambda_c} = nW C_{ox} v_{sat}$$

- G_{ms} becomes **independent of the length and of the current**
- It only depends on v_{sat} and increases with W

C. C. Enz and E. A. Vittoz, *Charge-Based MOS Transistor Modeling - The EKV Model for Low-Power and RF IC Design*, John Wiley, 2006.

G_m vs. IC for 28 and 40-nm Bulk CMOS Process



$$g_{ms}(IC) \triangleq \frac{G_{ms}}{G_{spec}} = \frac{nG_m}{G_{spec}}$$

$$= \frac{\sqrt{(\lambda_c IC + 1)^2 + 4IC} - 1}{\lambda_c(\lambda_c IC + 1) + 2}$$

$$\cong \frac{1}{\lambda_c} \text{ for } IC \gg 1$$

with $\lambda_c = \frac{L_{sat}}{L}$

and $L_{sat} = \frac{2\mu_0 U_T}{v_{sat}} = \frac{2U_T}{E_c}$

- Simple model of **transconductance** validated on 28-nm and 40-nm bulk CMOS processes over more 5 decades of current despite only requiring few parameters, namely:

$$n, I_{spec\square}, V_{T0}, L_{sat}$$

C. Enz, F. Chicco, and A. Pezzotta, IEEE Solid-State Circuits Magazine, vol. 9, no. 3, pp. 26-35, Summer 2017.

Effect of VS on the Drain Current in WI

- Velocity saturation also affects the current in weak inversion (in saturation)

$$i_d = \frac{q_s}{1 + \frac{\lambda_c}{2}}$$

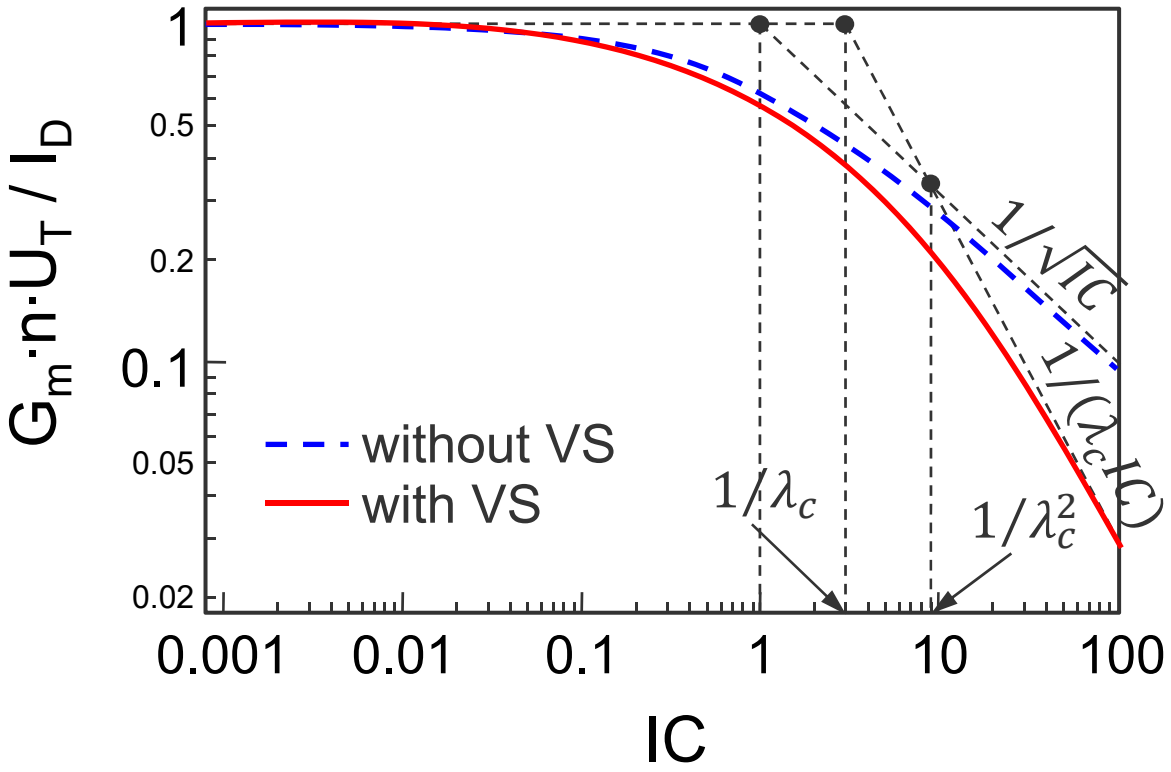
- The source transconductance is then given by

$$g_{ms} = \frac{q_s}{1 + \frac{\lambda_c}{2}} = i_d$$

- The source (gate) transconductances **remain proportional to the current**
- The G_{ms}/I_D ratio **remains equal to unity** as for the long channel case

📖 A. Mangla, J.-M. Sallese and C. Enz, MIXDES 2011

The Current Efficiency G_m/I_D FoM



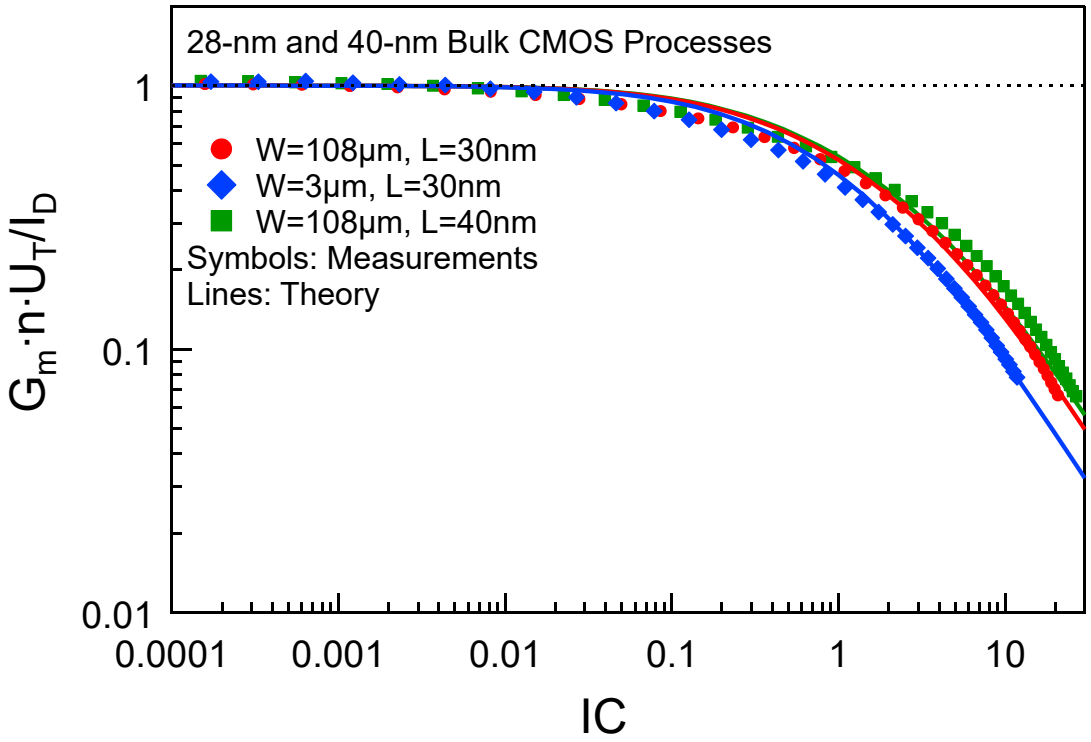
$$\lambda_c \triangleq \frac{L_{sat}}{L}$$

$$L_{sat} \triangleq \frac{2\mu_0 U_T}{v_{sat}} = \frac{2U_T}{E_c}$$

$$\frac{g_{ms}}{i_d} = \frac{G_{ms}U_T}{I_D} = \frac{G_m n U_T}{I_D} = \frac{\sqrt{(\lambda_c IC+1)^2+4IC}-1}{IC(\lambda_c(\lambda_c IC+1)+2)} = \begin{cases} 1 & \text{WI and sat.} \\ \dfrac{1}{\lambda_c IC} & \text{SI and sat.} \end{cases}$$

A. Mangla, M. A. Chalkiadaki, F. Fadhuile, T. Taris, Y. Deval, and C. C. Enz, Microelectronics Journal, vol. 44, pp. 570-575, July 2013.

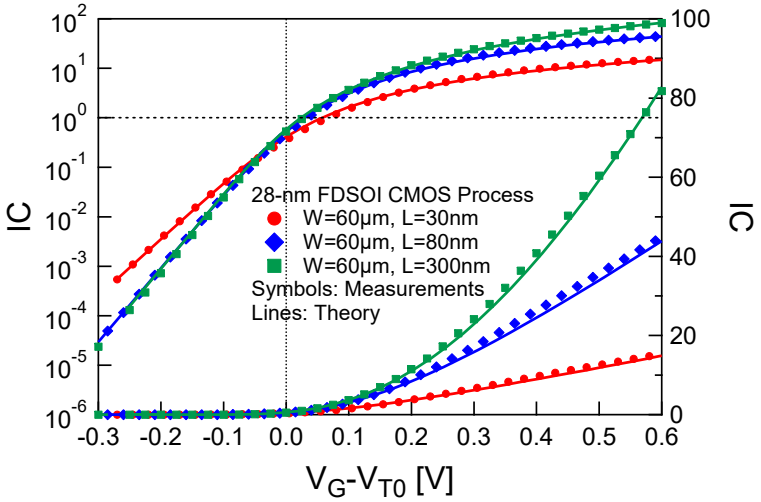
G_m/I_D vs. IC for 28 and 40-nm Bulk CMOS

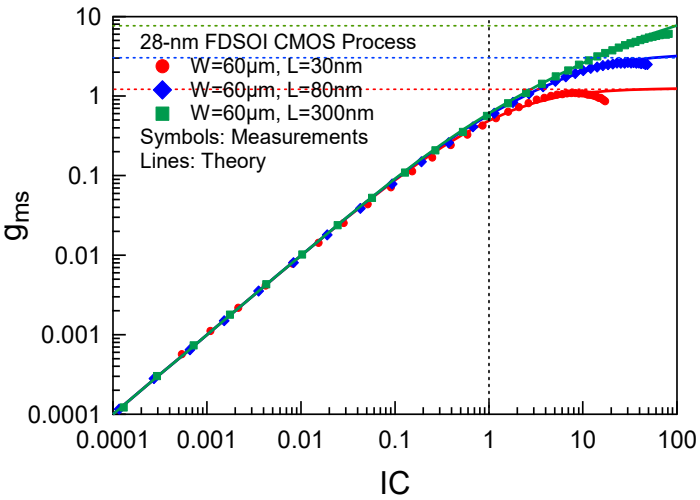


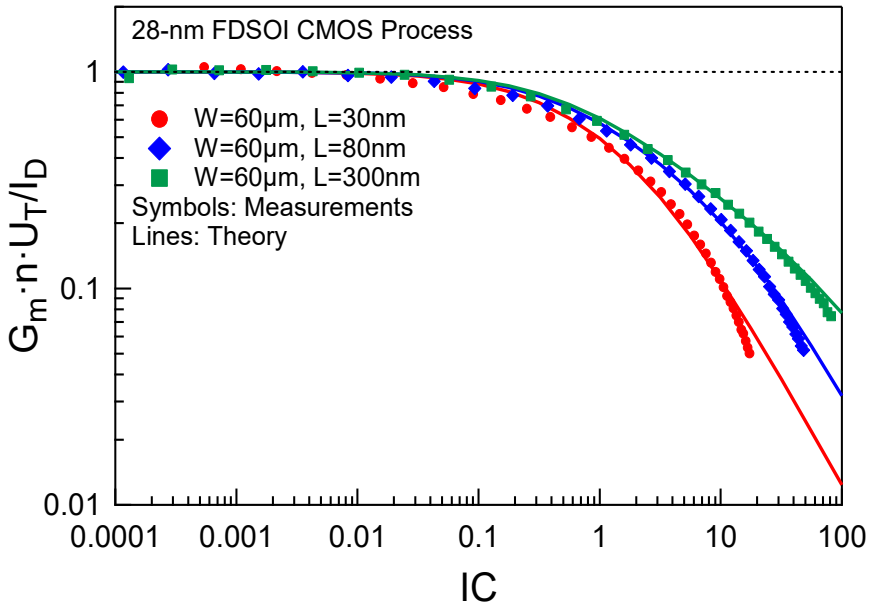
$$\frac{g_{ms}}{i_d} = \frac{G_{ms}U_T}{I_D} = \frac{G_{ms}nU_T}{I_D} = \frac{\sqrt{(\lambda_c IC + 1)^2 + 4IC} - 1}{IC(\lambda_c(\lambda_c IC + 1) + 2)}$$

C. Enz, F. Chicco, and A. Pezzotta, IEEE Solid-State Circuits Magazine, vol. 9, no. 3, pp. 26-35, Summer 2017.

IC, G_m and G_m/I_D for 28-nm FDSOI Process

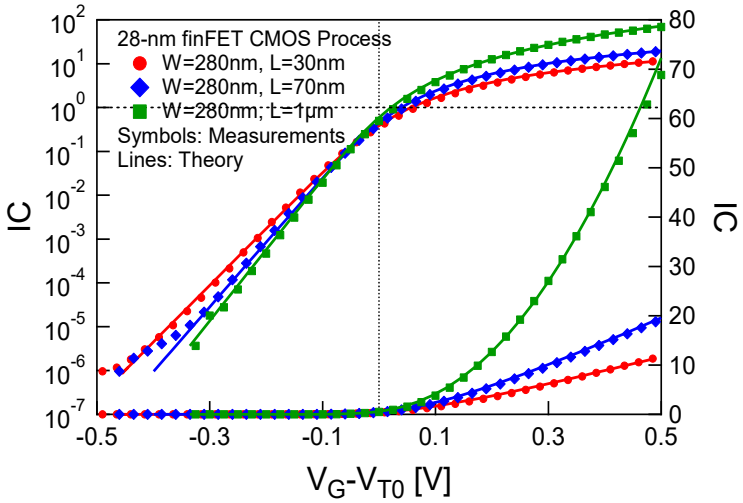


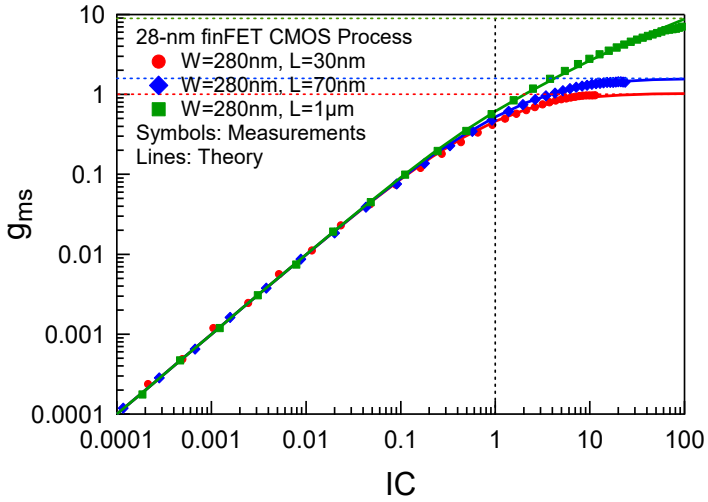


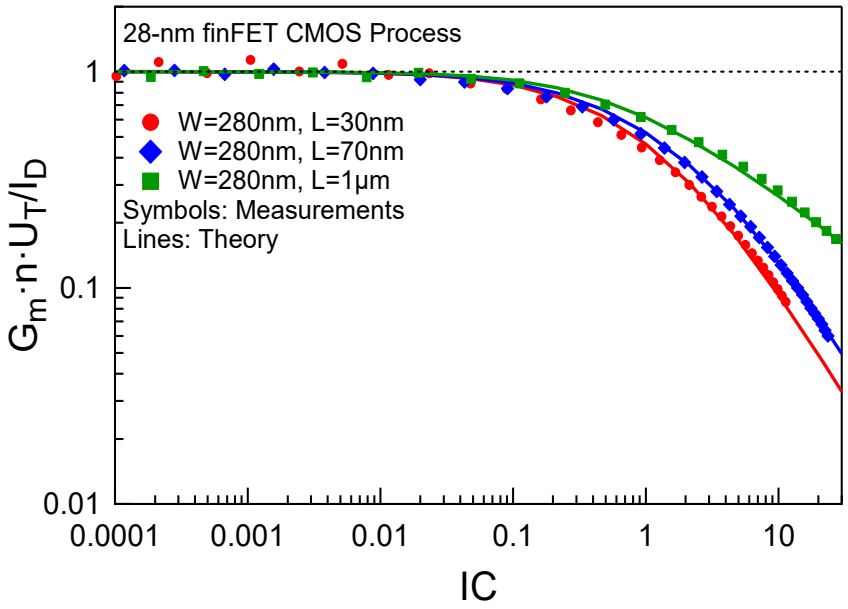


C. Enz, F. Chicco, and A. Pezzotta, IEEE Solid-State Circuits Magazine, vol. 9, no. 3, pp. 26-35, Summer 2017.

IC, G_m and G_m/I_D for 28-nm FinFET Process

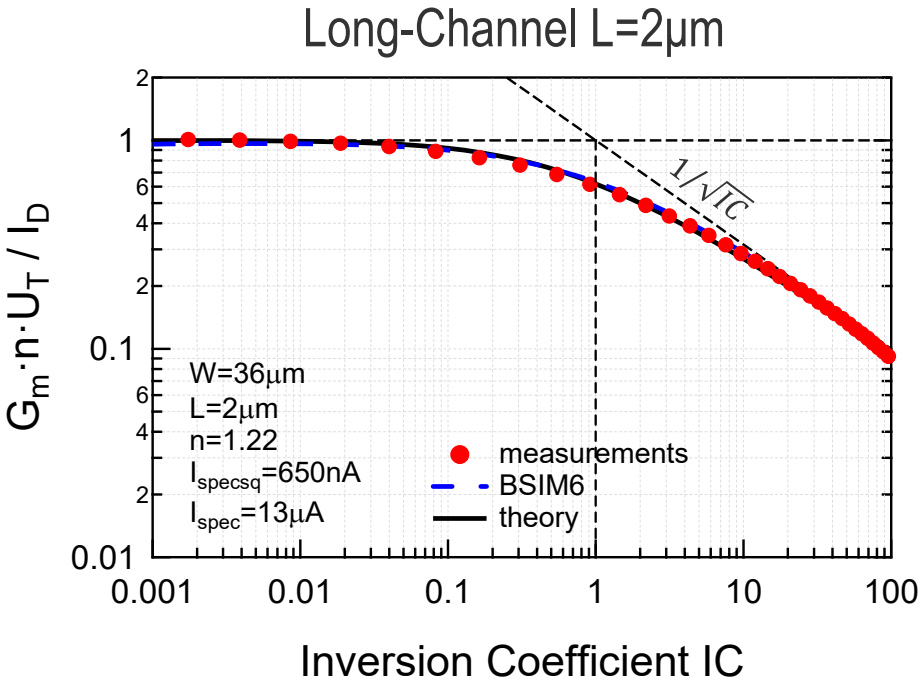


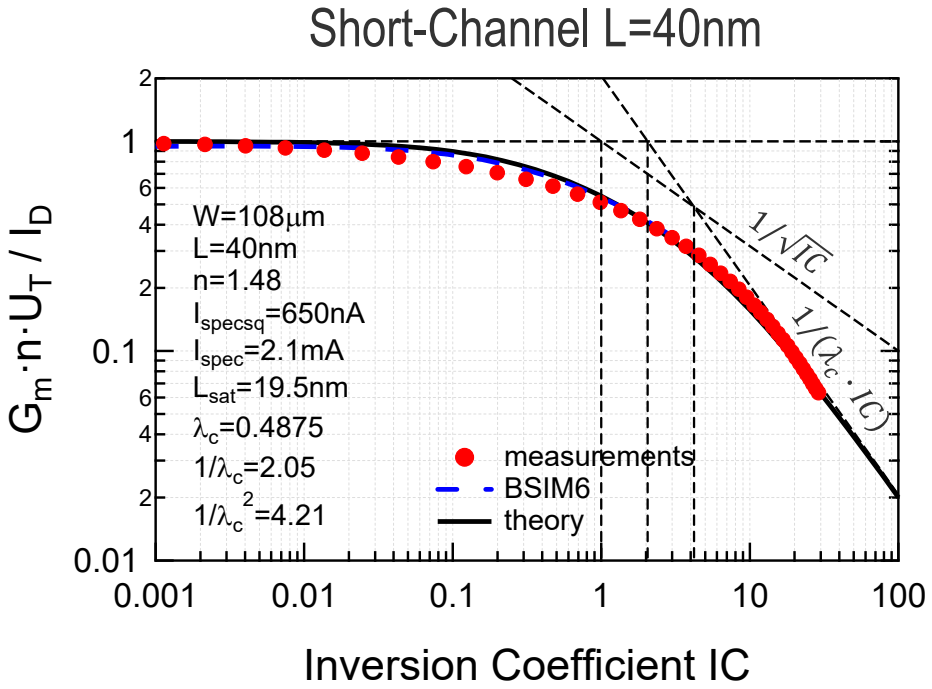




📖 C. Enz, F. Chicco, and A. Pezzotta, IEEE Solid-State Circuits Magazine, vol. 9, no. 3, pp. 26-35, Summer 2017.

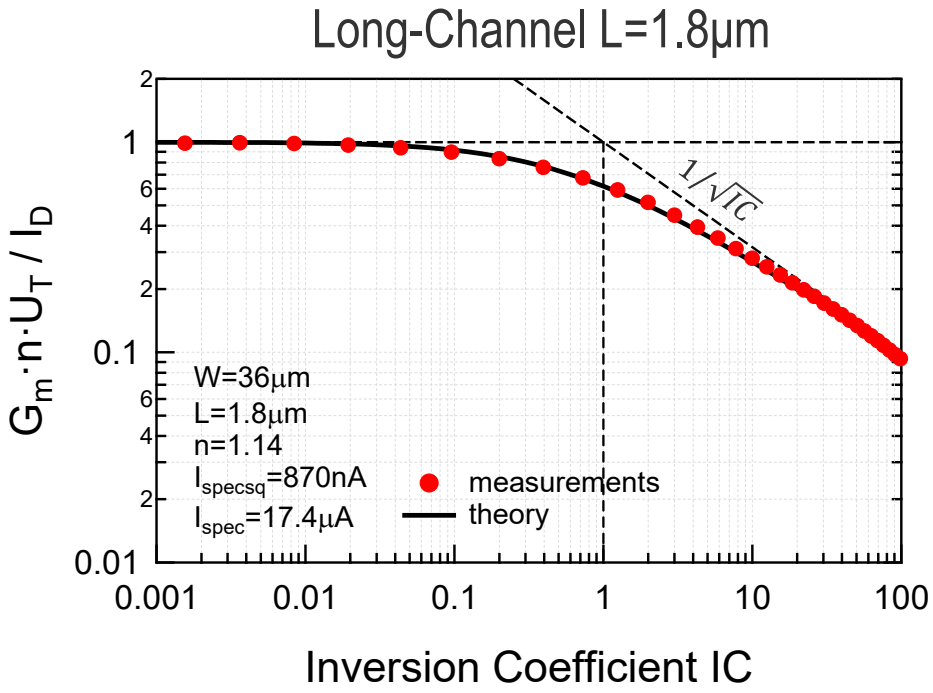
G_m/I_D vs. IC for 40nm Bulk CMOS Process

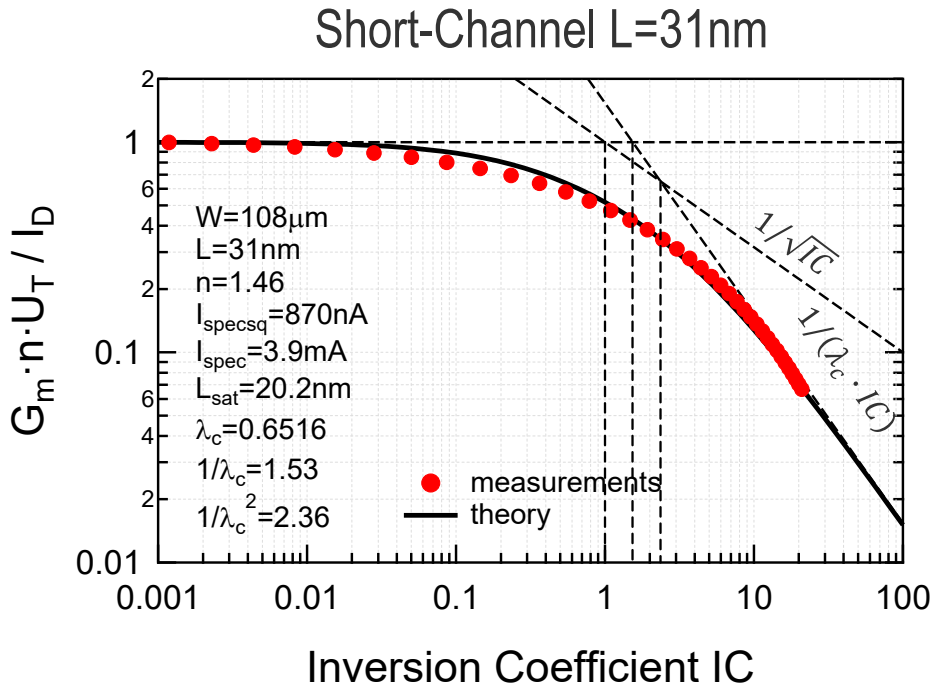




$$\frac{g_{ms}}{i_d} = \frac{G_{ms}U_T}{I_D} = \frac{G_m n U_T}{I_D} = \frac{\sqrt{(\lambda_c IC+1)^2+4IC}-1}{IC(\lambda_c(\lambda_c IC+1)+2)} = \begin{cases} 1 & \text{WI and sat.} \\ \dfrac{1}{\lambda_c IC} & \text{SI and sat.} \end{cases}$$

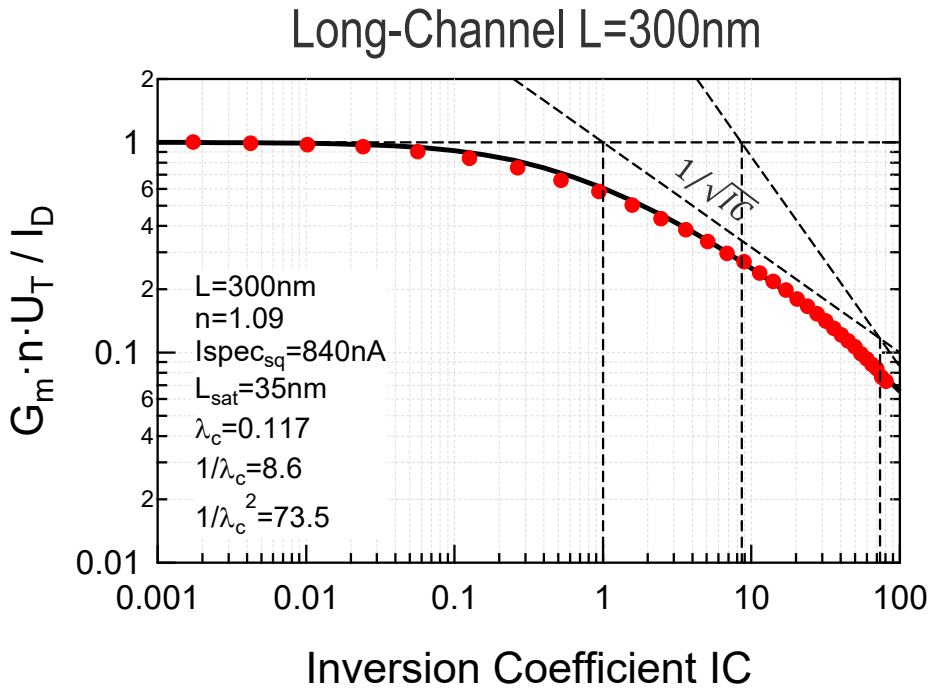
G_m/I_D vs. IC for 28nm Bulk CMOS Process

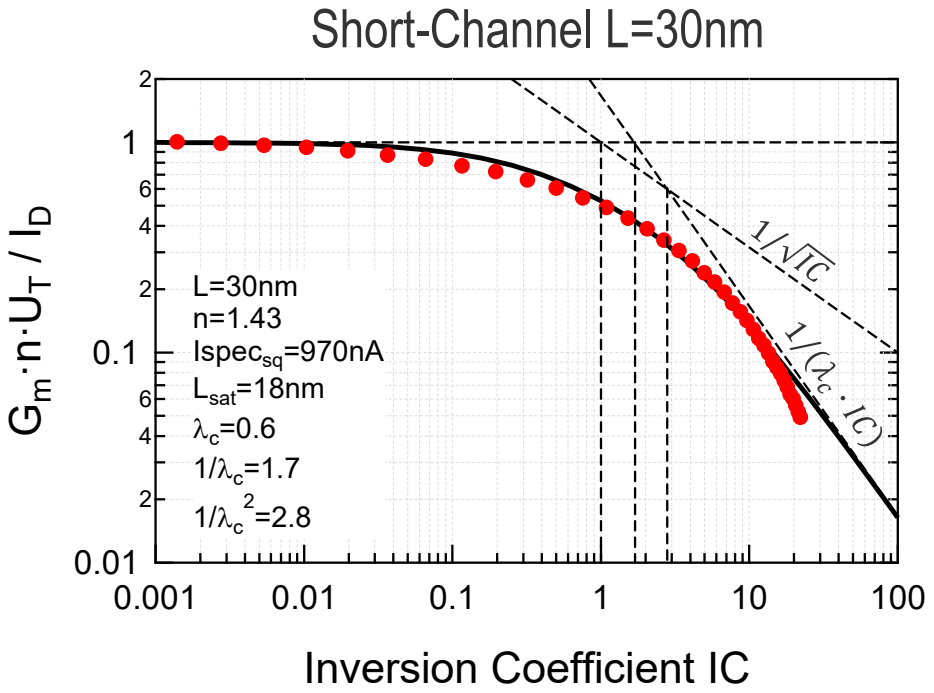




$$\frac{g_{ms}}{i_d} = \frac{G_{ms}U_T}{I_D} = \frac{G_m n U_T}{I_D} = \frac{\sqrt{(\lambda_c IC + 1)^2 + 4IC} - 1}{IC(\lambda_c(\lambda_c IC + 1) + 2)} = \begin{cases} 1 & \text{WI and sat.} \\ \dfrac{1}{\lambda_c IC} & \text{SI and sat.} \end{cases}$$

G_m/I_D vs. IC for 28nm FDSOI CMOS Process





$$\frac{g_{ms}}{i_d} = \frac{G_{ms}U_T}{I_D} = \frac{G_m n U_T}{I_D} = \frac{\sqrt{(\lambda_c IC + 1)^2 + 4IC} - 1}{IC(\lambda_c(\lambda_c IC + 1) + 2)} = \begin{cases} 1 & \text{WI and sat.} \\ \dfrac{1}{\lambda_c IC} & \text{SI and sat.} \end{cases}$$

Transit Frequency versus Inversion Coefficient

- The **transit frequency** can be written in terms of the **inversion coefficient** as

$$\omega_t \triangleq \frac{G_m}{C_G} = \frac{G_{spec}}{n \cdot W \cdot L \cdot C_{ox}} \cdot \frac{g_{ms}(IC)}{c_{Gi} + \frac{C_{Ge}}{W \cdot L \cdot C_{ox}}} \quad \text{where} \quad G_{spec} \triangleq 2n\mu C_{ox} \frac{W}{L} U_T$$

- where the normalized source transconductance is given by

$$g_{ms} = \frac{\sqrt{(\lambda_c \cdot IC + 1)^2 + 4IC} - 1}{\lambda_c \cdot (\lambda_c \cdot IC + 1) + 2} = \begin{cases} IC & \text{WI} & IC \ll 1 \\ \frac{\sqrt{4IC+1} - 1}{2} & \text{without VS} & \lambda_c = 0 \\ \sqrt{IC} & \text{SI without VS} & IC \gg 1 \text{ and } \lambda_c = 0 \\ \frac{1}{\lambda_c} & \text{SI with VS} & \lambda_c \cdot IC \gg 1 \end{cases}$$

- And the normalized capacitances are given by

$$c_{Gi} \cong \begin{cases} 1 - \frac{1}{n} & \text{WI and saturation} \\ 1 - \frac{1}{3n} & \text{SI and saturation} \end{cases} \qquad C_{Ge} = C_{Go} + C_{Gf} = W \cdot (C_{GoW} + C_{GfW}) = W \cdot C_{GeW}$$

Transit Frequency versus Inversion Coefficient

- Replacing G_{spec} results in

$$\omega_t = \frac{2\mu \cdot U_T}{L^2} \cdot \frac{g_{ms}(IC)}{c_{Gi} + \frac{C_{Ge}}{W \cdot L \cdot C_{ox}}} = \omega_{spec} \cdot \frac{g_{ms}(IC)}{c_{Gi} + \frac{C_{Ge}}{W \cdot L \cdot C_{ox}}} \quad \text{where} \quad \omega_{spec} \triangleq \frac{2\mu \cdot U_T}{L^2}$$

- For short channel devices, c_G is usually dominated by the extrinsic part (and hence independent of IC)

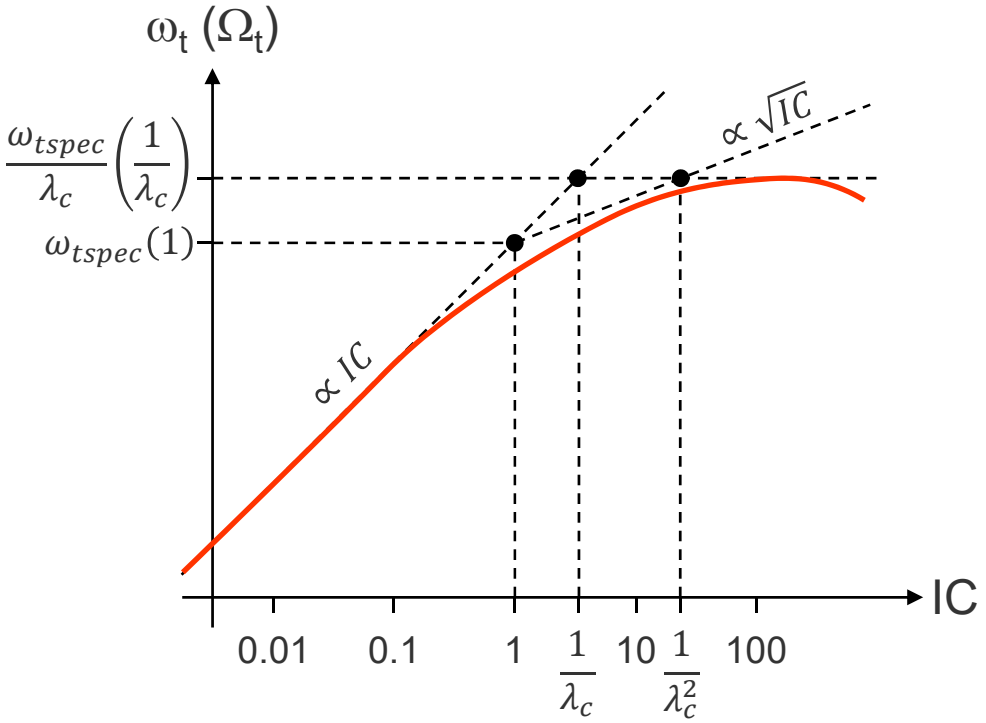
$$c_G \triangleq \frac{C_G}{W \cdot L \cdot C_{ox}} = c_{Gi} + \frac{C_{Ge}}{W \cdot L \cdot C_{ox}} \cong \frac{C_{Ge}}{W \cdot L \cdot C_{ox}} = \frac{C_{GeW}}{L \cdot C_{ox}}$$

- Resulting in

$$\omega_t \cong \omega_{spec} \cdot \frac{L \cdot C_{ox}}{C_{GeW}} \cdot g_{ms}(IC) = \frac{2\mu \cdot U_T}{L} \cdot \frac{C_{ox}}{C_{GeW}} \cdot g_{ms}(IC)$$

- which only scales as $1/L$ compared to ω_{spec} which scales as $1/L^2$

Specific Transit Frequency ω_{tspec}



$$\omega_t \cong \omega_{tspec} \cdot \frac{\sqrt{(\lambda_c \cdot IC + 1)^2 + 4IC} - 1}{\lambda_c \cdot (\lambda_c \cdot IC + 1) + 2} =$$

$$= \begin{cases} \omega_{tspec} \cdot IC & \text{WI} & IC \ll 1 \\ \omega_{tspec} \cdot \sqrt{IC} & \text{SI without VS} & IC \gg 1 \text{ and } \lambda_c = 0 \\ \dfrac{\omega_{tspec}}{\lambda_c} & \text{SI with VS} & IC \gg 1 \end{cases}$$

$$\Omega_t \triangleq \frac{\omega_t}{\omega_{tspec}} = g_{ms}(IC) = \frac{\sqrt{(\lambda_c \cdot IC + 1)^2 + 4IC} - 1}{\lambda_c \cdot (\lambda_c \cdot IC + 1) + 2}$$

- ω_{tspec} is the transit frequency obtained for $IC = 1$ assuming WI (obtained from the WI asymptote)

$$\omega_{tspec} \triangleq \omega_t\big|_{\text{WI and } IC=1} = \frac{\omega_{spec}}{c_{Gi} + \dfrac{C_{Ge}}{W \cdot L \cdot C_{ox}}} \cong \frac{\omega_{spec}}{\dfrac{C_{Ge}}{W \cdot L \cdot C_{ox}}}$$

📖 C. Enz and M. Chalkiadaki, APMC 2015

Maximum (or Peak) Transit Frequency

- For short-channel devices $C_{Gi} \ll C_{Ge}$
- Hence, the specific transit frequency ω_{tspec} roughly scales as $1/L$

$$\omega_{tspec} \cong \frac{I_{spec\square}}{nU_T \cdot C_{GeW} \cdot L}$$

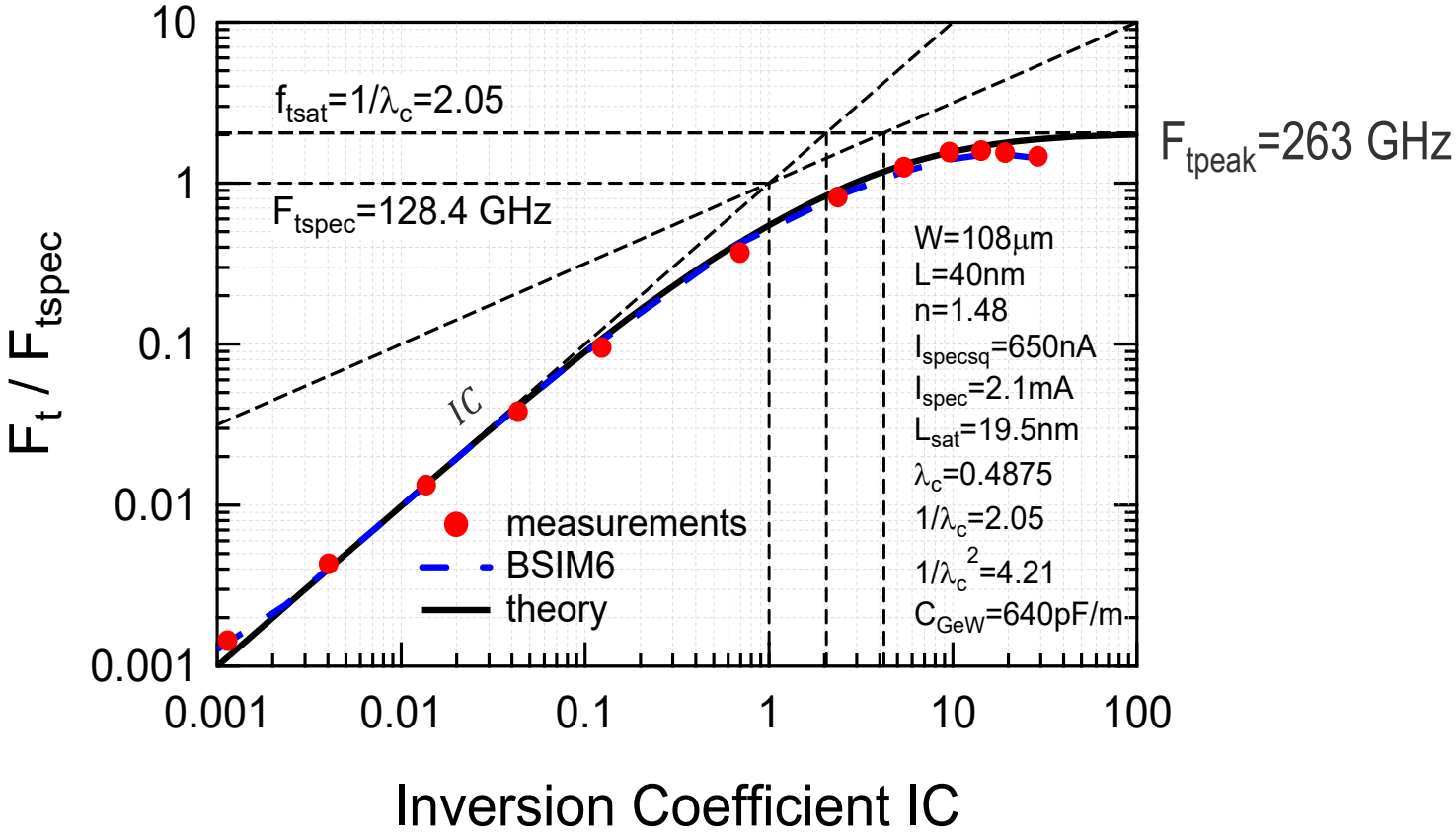
- The transit frequency saturates in SI due to velocity saturation to

$$\omega_{tpeak} = \frac{\omega_{tspec}}{\lambda_c} \cong v_{sat} \cdot \frac{C_{ox}}{C_{GeW}}$$

- Since C_{GeW} does not scale with L, ω_{tpeak} **does not scale with L** either
- ω_{tpeak} can therefore only take advantage of scaling through the increase of C_{ox} mitigated by the possible increase of C_{GeW}

📖 C. Enz and M. Chalkiadaki, APMC 2015

F_t vs. IC for 40nm Bulk CMOS Process

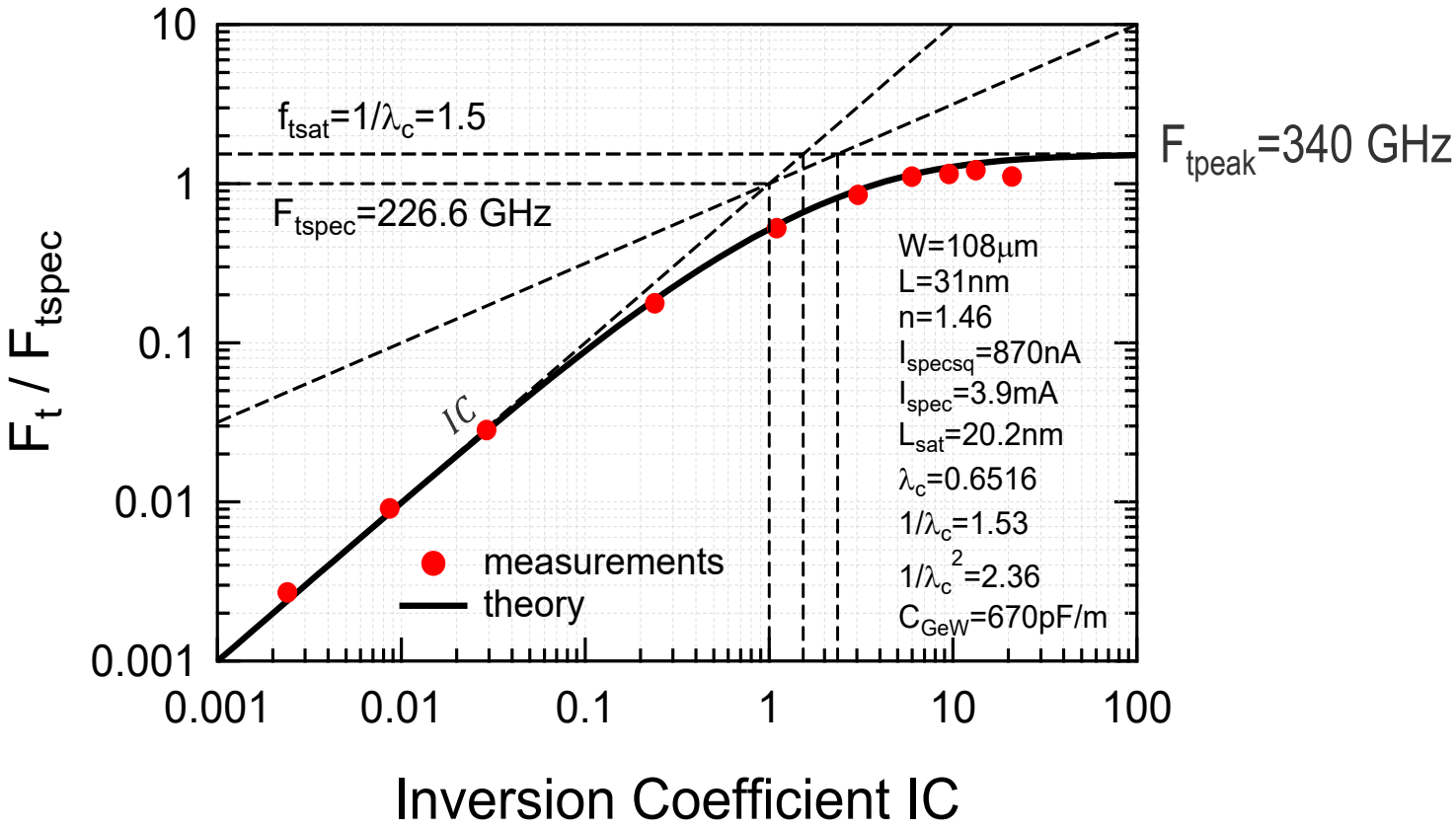


$$f_t \triangleq \frac{F_t}{F_{tspec}} = g_{ms} = \frac{\sqrt{(\lambda_c IC+1)^2+4IC}-1}{\lambda_c(\lambda_c IC+1)+2} = \begin{cases} IC & \text{WI and sat.} \\ \dfrac{1}{\lambda_c} & \text{SI and sat.} \end{cases}$$

C. Enz and M. Chalkiadaki, APMC 2015

A. Mangla, M. A. Chalkiadaki, F. Fadhuile, T. Taris, Y. Deval, and C. C. Enz, Microelectronics Journal, vol. 44, pp. 570-575, July 2013.

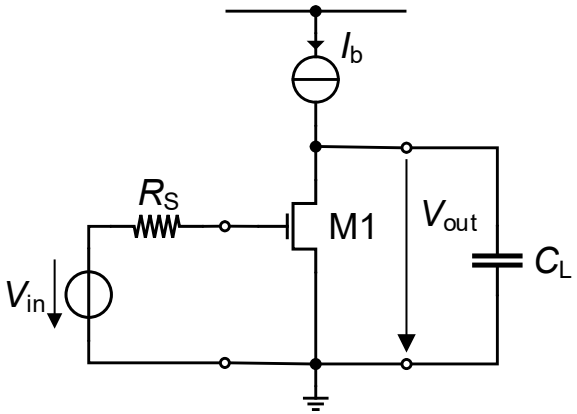
F_t vs. IC for 28nm Bulk CMOS Process



$$f_t \triangleq \frac{F_t}{F_{tspec}} = g_{ms} = \frac{\sqrt{(\lambda_c IC+1)^2+4IC}-1}{\lambda_c(\lambda_c IC+1)+2} = \begin{cases} IC & \text{WI and sat.} \\ \dfrac{1}{\lambda_c} & \text{SI and sat.} \end{cases}$$

📖 C. Enz and M. Chalkiadaki, APMC 2015

Figure-of-Merit for Low Power RF



- A. Shameli and P. Heydari, *ISLPED* 2006
- T. Taris, *et al.*, RFIC 2011
- A. Mangla, J.-M. Sallese and C. Enz, MIXDES 2011

- The **voltage gain** and **noise factor** of common-source stage loaded by similar stage (i.e. having a fan-out FO equal to 1 and hence $C_L = C_{GS}$) are given by

$$A_v \triangleq \frac{\Delta V_{out}}{\Delta V_{in}} = -\frac{G_m}{G_{ds}+j\omega C_L} \cong -\frac{G_m}{j\omega C_L} = j\frac{\omega_u}{\omega} \text{ with } \omega_u = \frac{G_m}{C_L} = \frac{G_m}{C_{GS}} \cong \omega_t$$

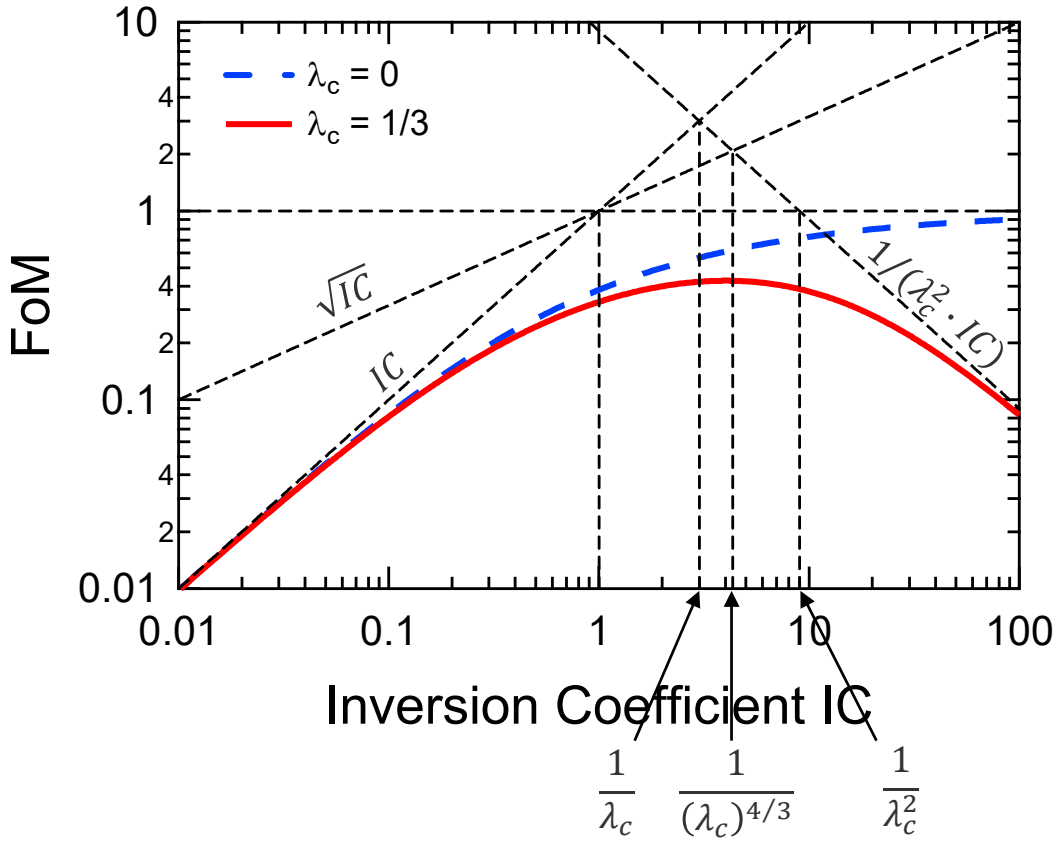
$$F = 1 + \frac{\gamma_{nD}}{G_m \cdot R_S} \text{ (assuming thermal noise from M1 and resistance } R_S \text{ only)}$$

- A FoM can be defined in order to **maximize the gain-bandwidth product** and **minimize the noise factor** at a **given current**

$$FoM \triangleq \frac{\omega_u}{(F-1)\cdot I_b} \cong \frac{R_S}{\gamma_{nD}} \cdot \boxed{\frac{G_m \cdot \omega_t}{I_b}}$$

- This FoM is proportional to the $G_m/I_b \cdot \omega_t$ ratio, which is an important FoM for low-power RF IC design

The $G_m/I_D \cdot F_t$ FoM is Maximum in Moderate Inversion



$$\lambda_c = \frac{1}{3}$$

$$\frac{1}{\lambda_c} = 3$$

$$\frac{1}{\lambda_c^2} = 9$$

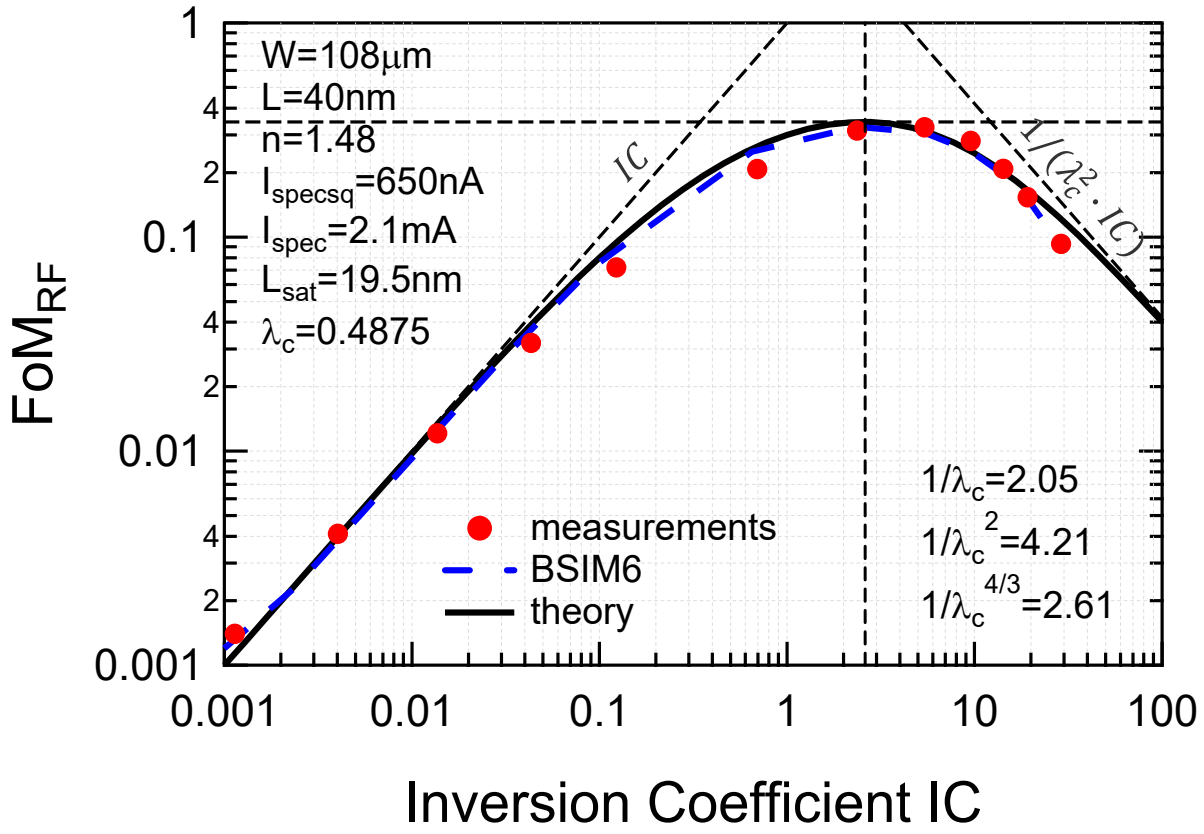
$$IC_{opt} \approx \frac{1}{\lambda_c^{4/3}} = 4.33$$

$$FoM_{opt} \approx FoM(IC_{opt}) = 0.43$$

$$FoM_{RF} = \frac{g_{ms} \cdot \Omega_t}{IC} = \frac{g_{ms}^2}{IC} = \frac{1}{IC} \cdot \left(\frac{\sqrt{(\lambda_c \cdot IC + 1)^2 + 4IC} - 1}{\lambda_c \cdot (\lambda_c \cdot IC + 1) + 2} \right)^2 \cong \begin{cases} IC & \text{WI} \quad IC \ll 1 \\ 1 & \text{SI without VS} \quad IC \gg 1 \text{ and } \lambda_c = 0 \\ \dfrac{1}{\lambda_c^2 \cdot IC} & \text{SI with VS} \quad IC \gg 1 \end{cases}$$

C. Enz and M. Chalkiadaki, APMC 2015

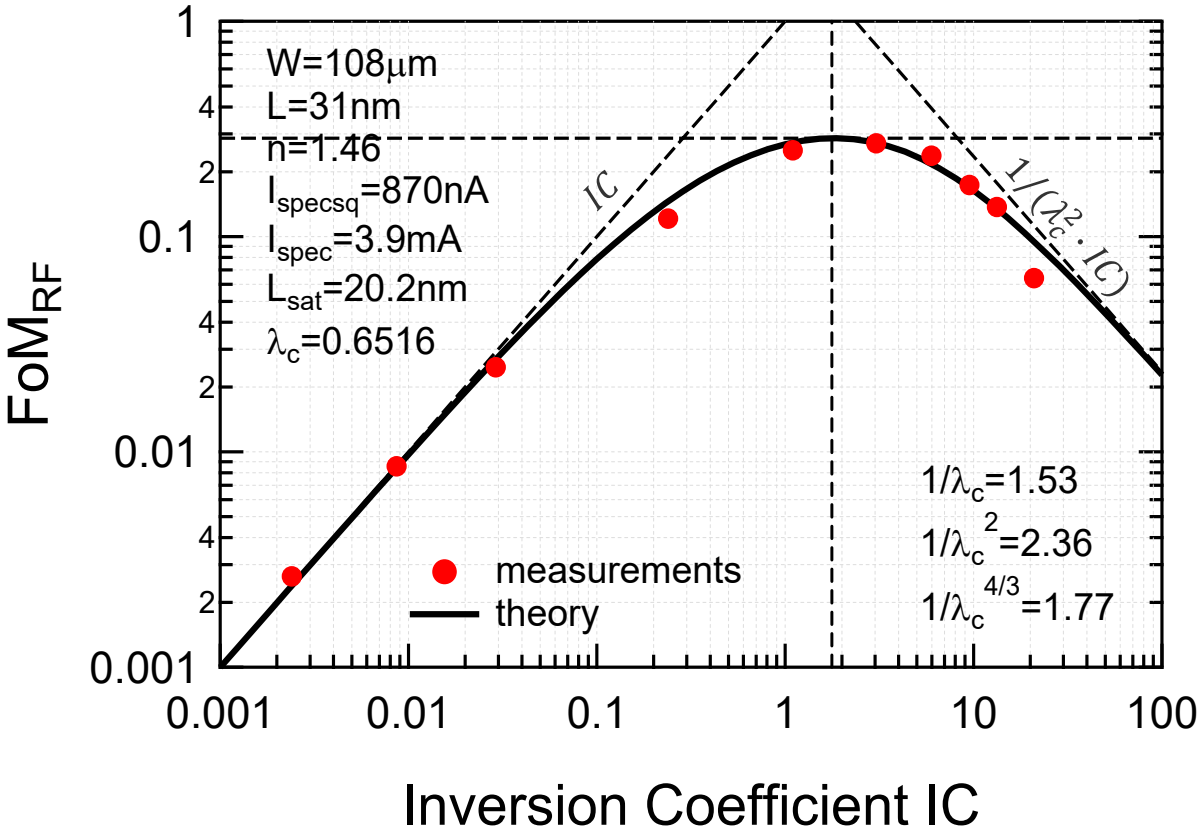
$G_m/I_D \cdot F_t$ vs. IC for 40nm Bulk CMOS Process



$$FoM_{RF} = \frac{g_{ms} \cdot \Omega_t}{IC} = \frac{g_{ms}^2}{IC} = \frac{1}{IC} \cdot \left(\frac{\sqrt{(\lambda_c \cdot IC + 1)^2 + 4IC} - 1}{\lambda_c \cdot (\lambda_c \cdot IC + 1) + 2} \right)^2$$

📖 C. Enz and M. Chalkiadaki, APMC 2015

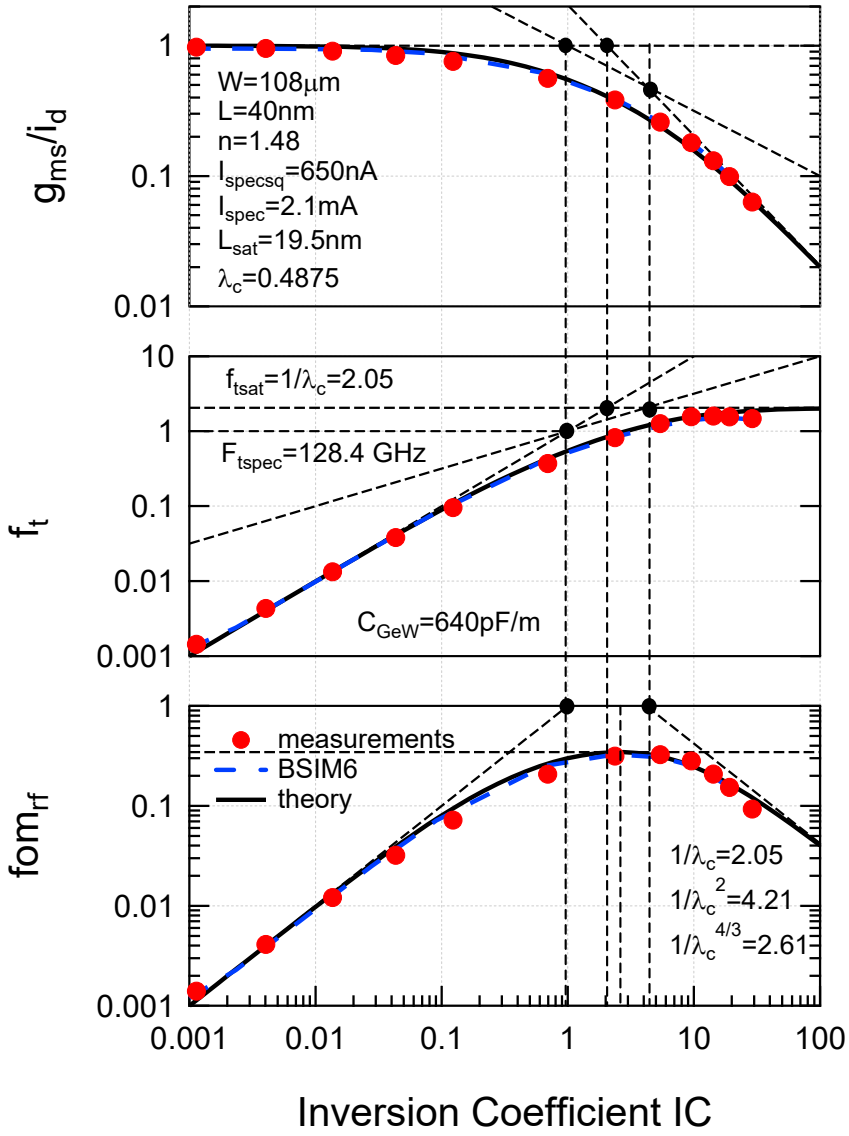
$G_m/I_D \cdot F_t$ vs. IC for 28nm Bulk CMOS Process

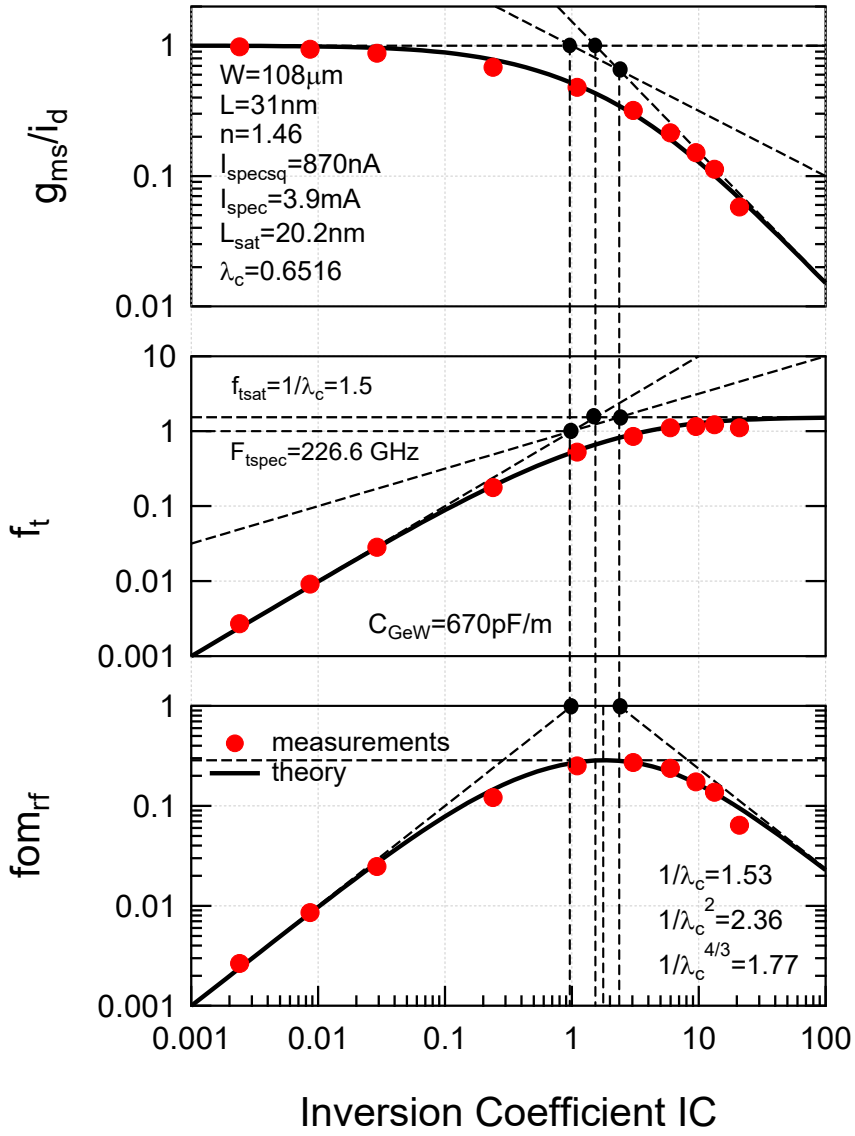


$$FoM_{RF} = \frac{g_{ms} \cdot \Omega_t}{IC} = \frac{g_{ms}^2}{IC} = \frac{1}{IC} \cdot \left(\frac{\sqrt{(\lambda_c \cdot IC + 1)^2 + 4IC} - 1}{\lambda_c \cdot (\lambda_c \cdot IC + 1) + 2} \right)^2$$

C. Enz and M. Chalkiadaki, APMC 2015

Combined FoMs vs. IC for 40nm and 28nm Bulk CMOS





📖 C. Enz and M. Chalkiadaki, APMC 2015

Maximum Frequency of Oscillation

- F_t is only a narrow way to characterize the ability of a device to operate at RF
- Another figure of merit that also accounts for the R_G and C_{GD} can be defined from the unilateral power gain U which corresponds to the maximum available gain (corresponding to the transducer gain with matched source and load impedance $Y_G = Y_{11}^*$ and $Y_L = Y_{22}^*$) with its feedback transadmittance neutralized ($Y_{12} = 0$)

$$U = \frac{\left|Y_{21}\right|^2}{4\left(G_{11}G_{22} - G_{12}G_{21}\right)} \quad \text{where} \quad G_{kl} \triangleq \Re\left\{Y_{kl}\right\} \quad \text{with} \quad k,l \in \left\{1,2\right\}$$

- From simple QS model one obtains

$$U \cong \frac{G_m^2}{4R_G C_G\left(G_{ds}C_G + G_m C_{GD}\right)\omega^2} \cong \frac{G_m}{4R_G C_G C_{GD}\omega^2} = \left(\frac{\omega_{\max}}{\omega}\right)^2$$

$$\omega_{\max} \cong \frac{G_m}{2\sqrt{R_G C_G\left(G_{ds}C_G + G_m C_{GD}\right)}} \cong \frac{1}{2}\sqrt{\frac{G_m}{R_G C_G C_{GD}}} = \frac{1}{2}\sqrt{\frac{\omega_t}{R_G C_{GD}}}$$

- The smaller the $R_G \cdot C_{GD}$ product the higher the F_{max} (it is therefore also used as another figure of merit)

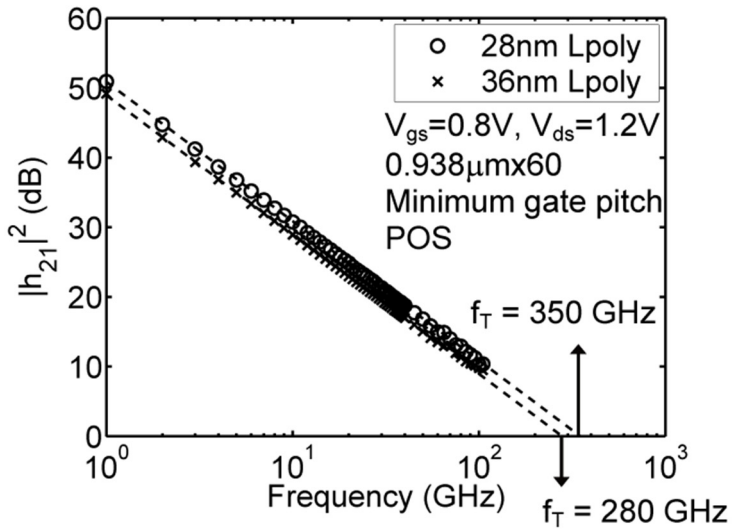
Minimum Noise Figure

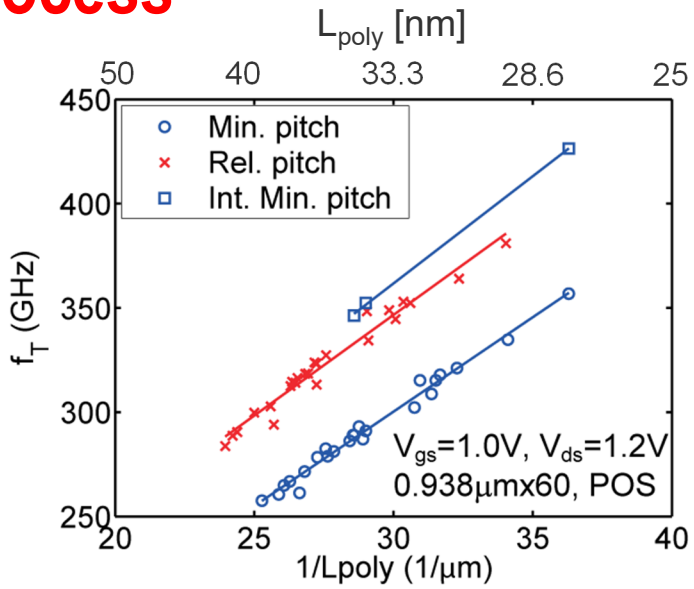
- Having high F_t and F_{max} is not sufficient, **low noise** is also required
- This feature is measured by the **noise factor** F or the **noise figure** NF

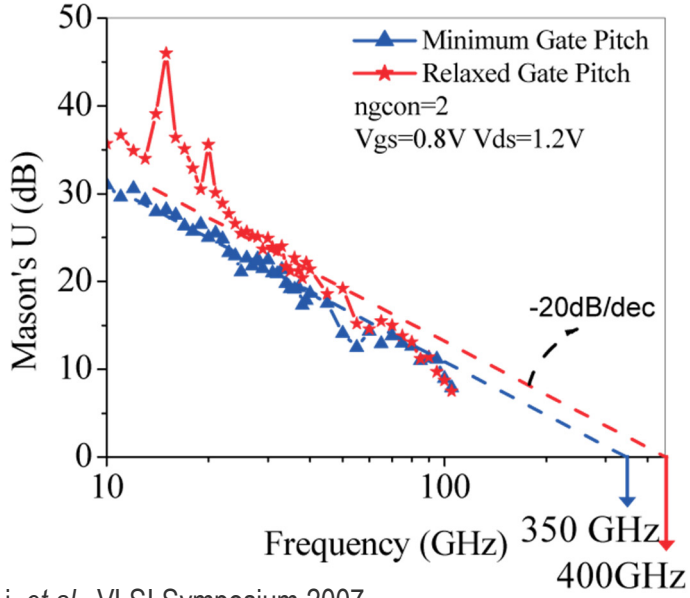
$$NF \triangleq 10 \cdot \log F$$

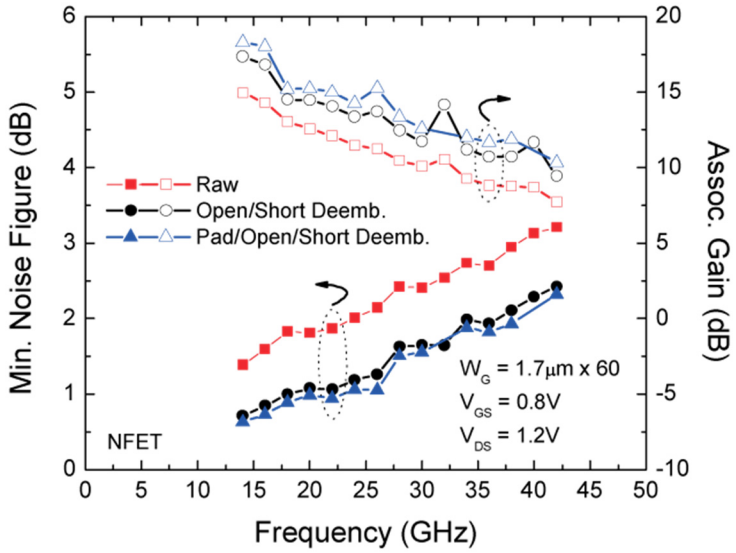
- The noise factor F is defined as the ratio of the total noise power measured at some point along the amplification chain (usually at the output) to the noise due to the generator only measured at that same point
- The noise factor depends thus on the **generator admittance** and becomes minimum for a particular value of this generator admittance
- The **minimum** value of the **noise factor** F_{min} (or **noise figure** NF_{min}) represents what a device can ultimately achieve in terms of minimum thermal noise contribution and is therefore used as a figure-of-merit
- For a MOST biased in strong inversion it is approximated by $F_{min} \cong 1 + \frac{\omega}{\omega_t}$
- The higher the F_t the smaller F_{min} for a given operating frequency

FoM of a 45 nm Bulk CMOS Process





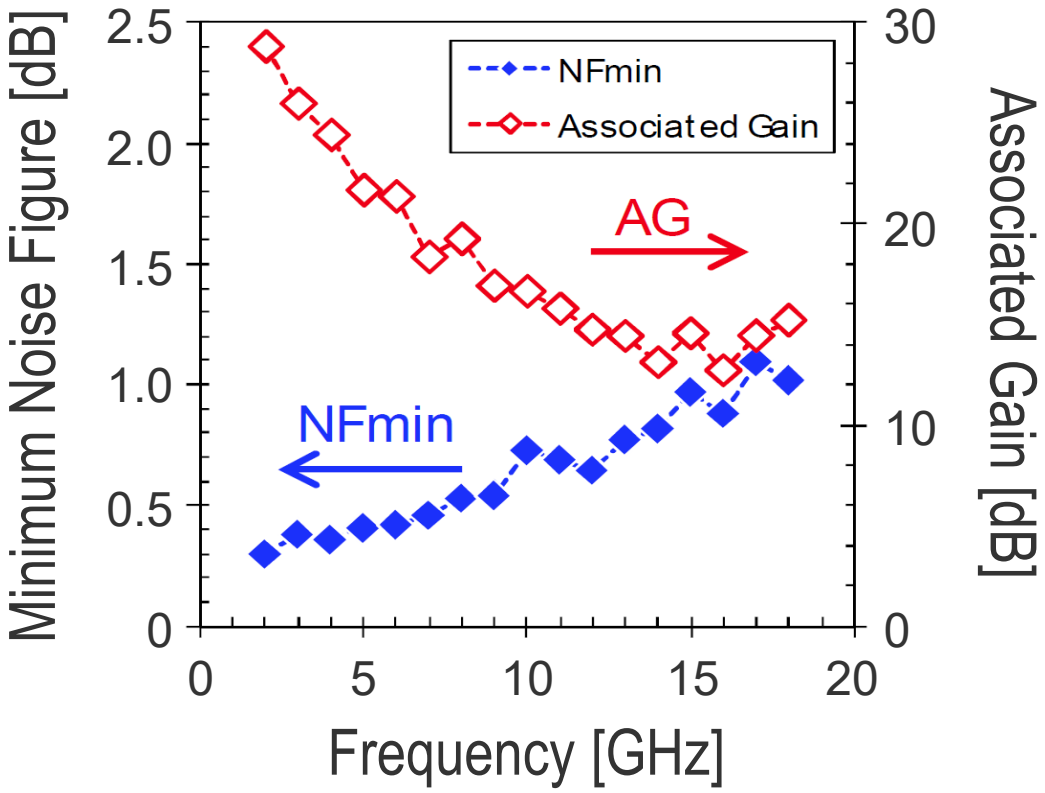




📖 Hongmei Li, *et al.*, VLSI Symposium 2007

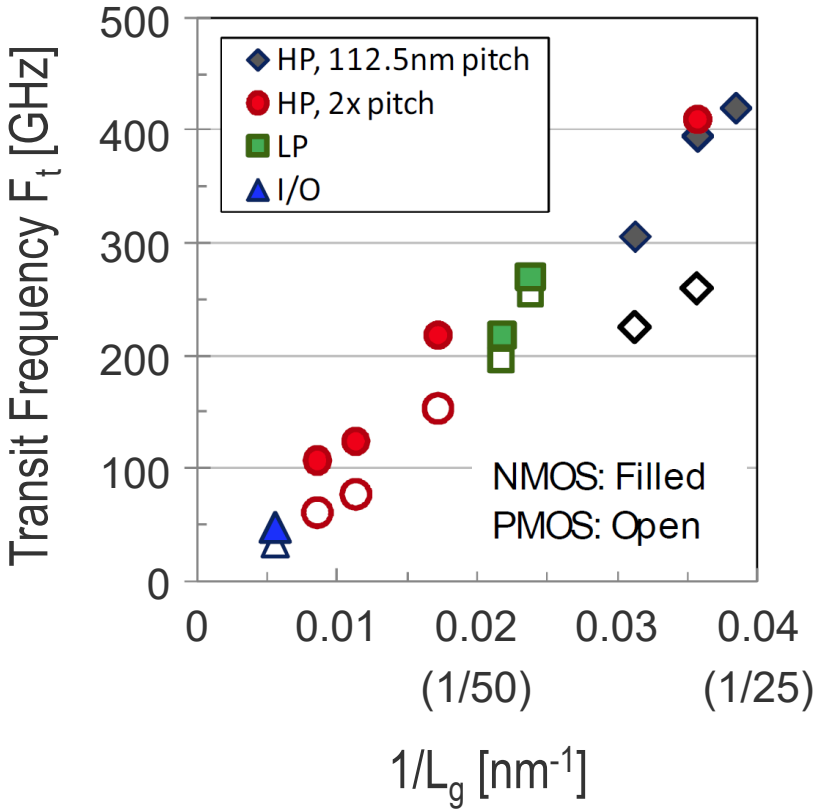
Minimum Noise Figure

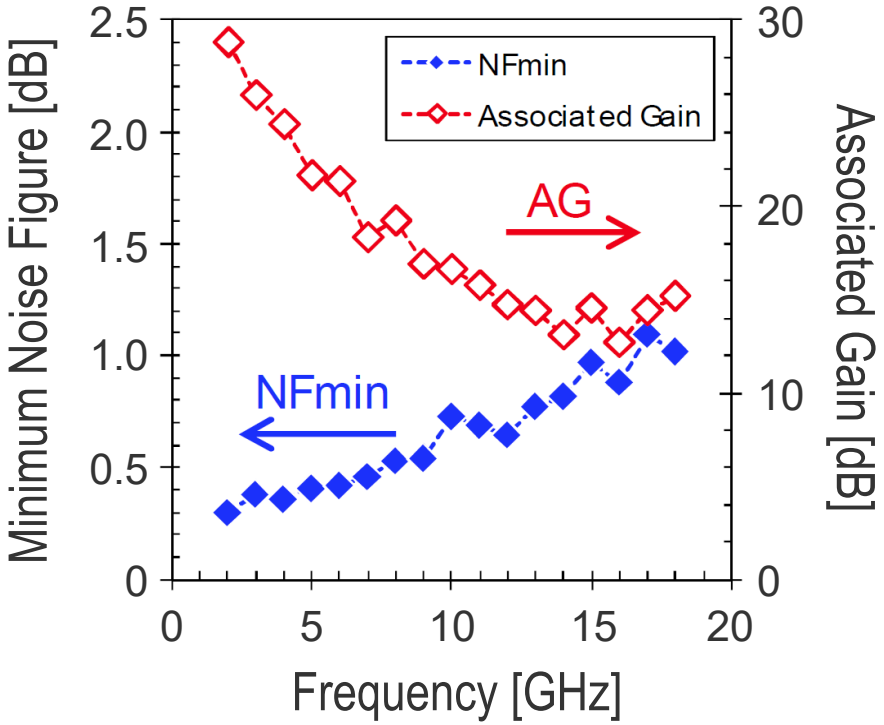
- Advanced processes can achieve a NF_{min} smaller 0.5 dB below 7 GHz as shown below for a 32nm bulk CMOS process



P. VanDerVoorn, *et al.*, Symposium on VLSI Technology, 2010

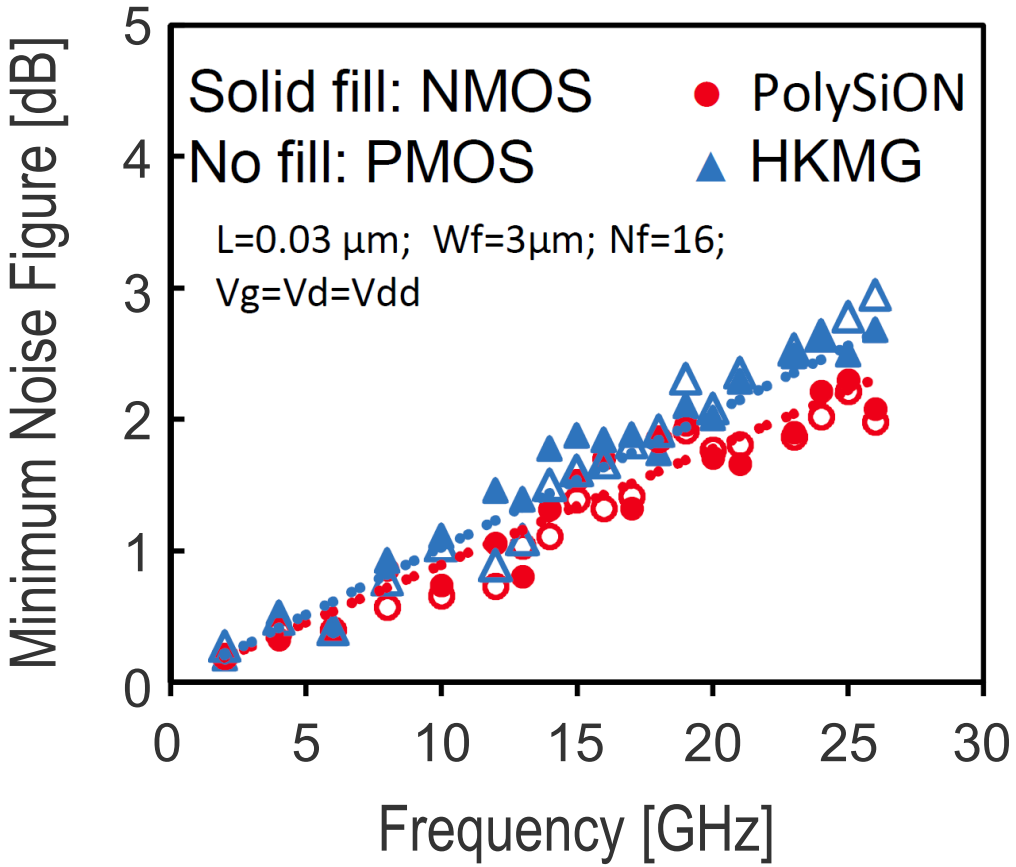
FoM of a 32nm Bulk CMOS Process





P. VanDerVoorn, *et al.*, Symposium on VLSI Technology, 2010

Minimum Noise Figure for a 28nm Bulk CMOS Process



K. W. J. Chew, *et al.*, RFIC, 2015

Noise Factor

- The **noise factor** of a single transistor is given by

$$F = F_{\min} + \frac{R_n}{G_s} \cdot \left[\left(G_s - G_{opt} \right)^2 + \left(B_s - B_{opt} \right)^2 \right]$$

- Where $G_s \triangleq \Re\{Y_s\}$ and $B_s \triangleq \Im\{Y_s\}$ are the real and imaginary part of the source admittance Y_s
- F requires **four noise parameters** $F_{min}, R_n, G_{opt}, B_{opt}$
- **Noise matching** corresponds to $F = F_{min}$ for $G_s = G_{opt}$ and $B_s = B_{opt}$

$$R_n \cong \frac{\gamma_{nD}}{G_m} + R_G$$

$$F_{\min} \cong 1 + 2\omega C_{GS} \cdot \frac{\gamma_{nD}}{G_m} \cdot \sqrt{\frac{\beta_{nG}}{\gamma_{nD}} \cdot \left(1 - c_g^2\right)}$$

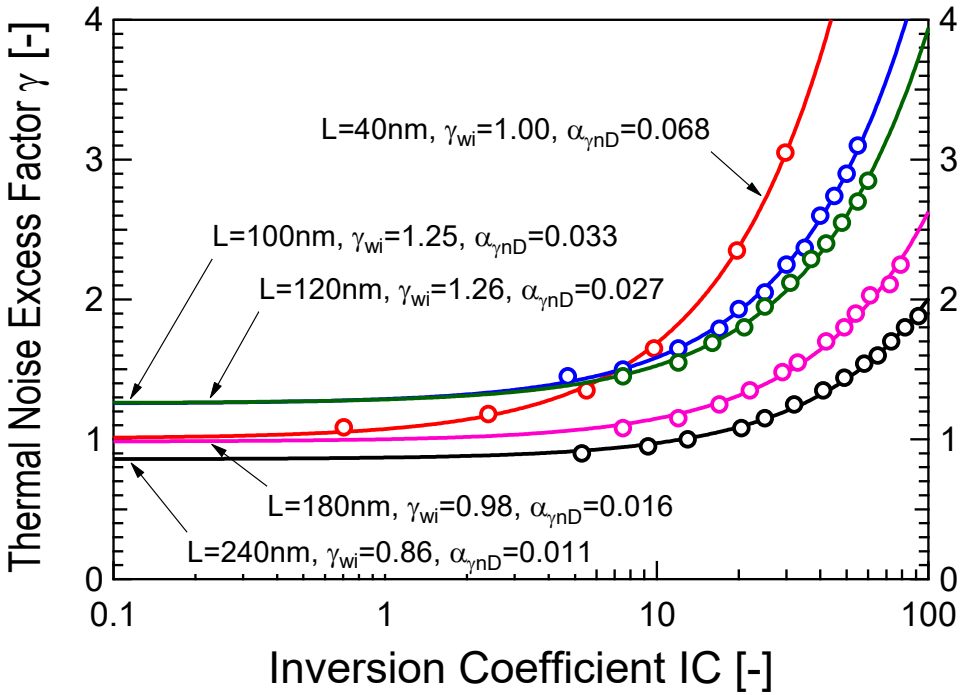
For long-channel:

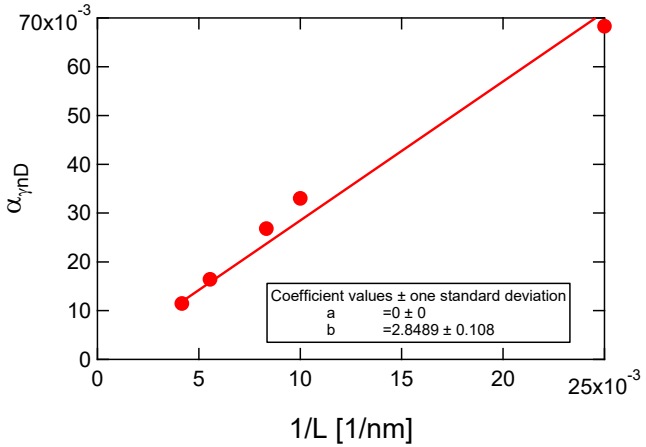
$$\gamma_{nD} = \begin{cases} \frac{n}{2} & \text{WI} \\ \frac{2n}{3} & \text{SI} \end{cases} \qquad \beta_{nG} = \begin{cases} \frac{1}{5n} & \text{WI} \\ \frac{4}{15n} & \text{SI} \end{cases}$$

$$\frac{\beta_{nG}}{\gamma_{nD}} = \frac{2}{5n^2} \qquad c_g \cong 0.4 \ldots 0.6$$

C. C. Enz and E. A. Vittoz, *Charge-Based MOS Transistor Modeling - The EKV Model for Low-Power and RF IC Design*, John Wiley, 2006.

Short-channel Effects on γ_{nD} (in saturation)





- The noise excess factor γ_{nD} can be modelled versus IC as

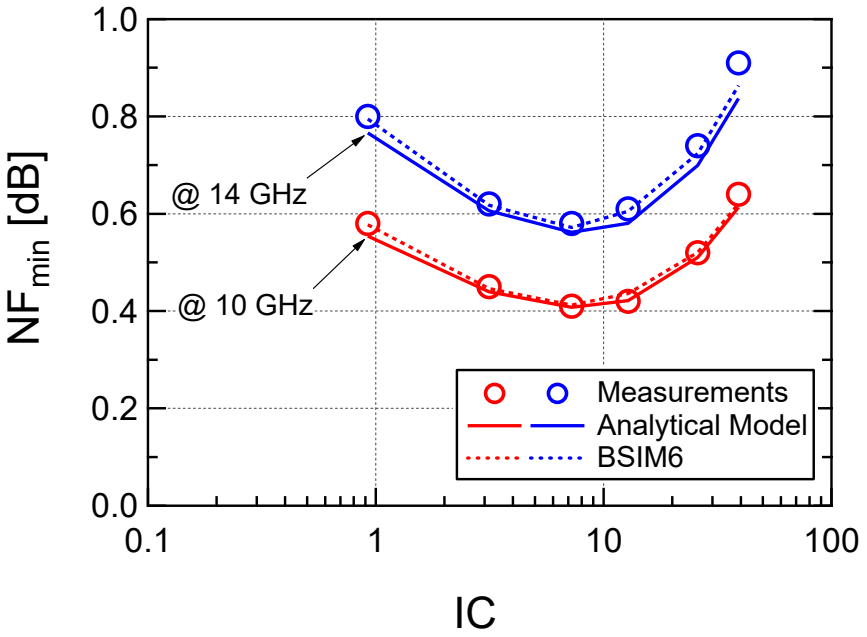
$$\gamma_{nD} \cong \gamma_{wi} + \alpha_{\gamma_{nD}} \cdot IC$$

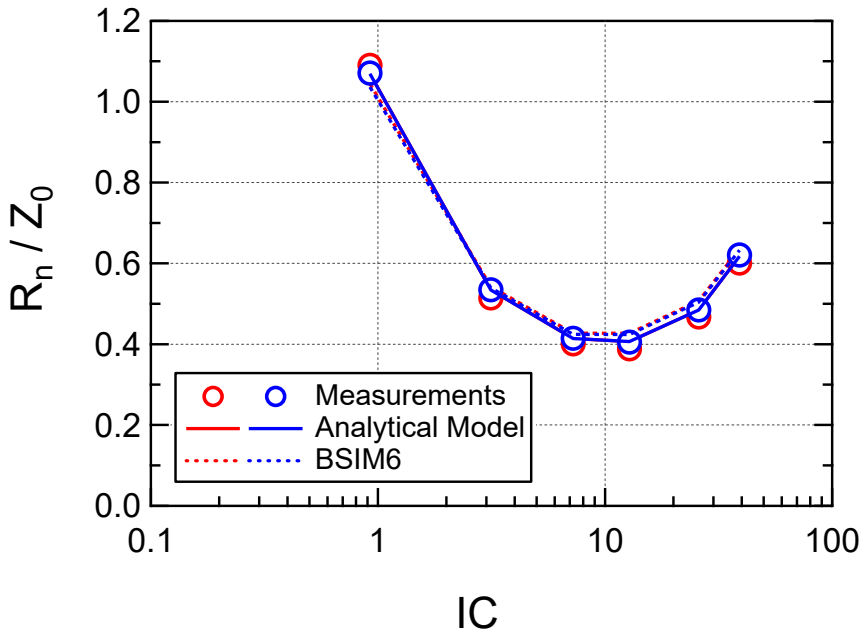
- Where γ_{wi} and $\alpha_{\gamma_{nD}}$ are empirical factors
- $\alpha_{\gamma_{nD}}$ scales approximatively as $\alpha_{\gamma_{nD}} \cong 2.85/L$ where L is in nm

📖 A. Antonopoulos et al., "CMOS Small-Signal and Thermal Noise Modeling at High Frequencies," TED, vol. 60, No. 11, Nov. 2013.
📖 M. Chalkiadaki, PhD Thesis 2016.

NF_{min} and R_n versus IC for 40nm Bulk CMOS Process

- The minimum noise figure NF_{min} and input-referred noise resistance R_n show a minimum in MI due to the sharp increase of γ_{nG} at high IC





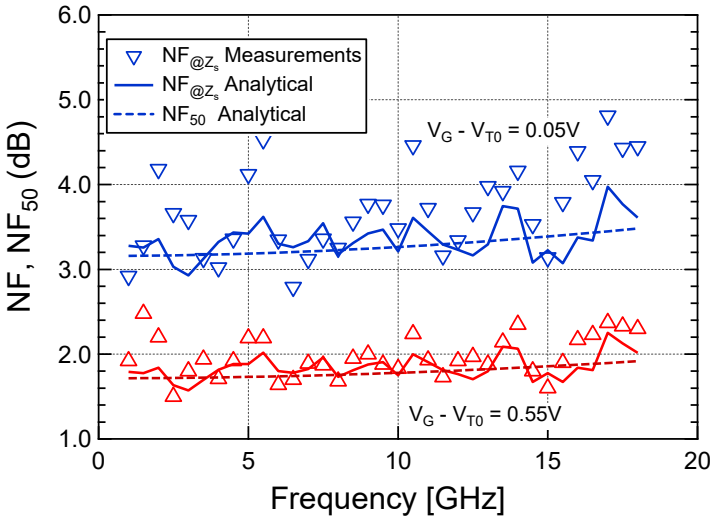
$$F_{\min} \cong 1 + 2\omega C_{GS} \cdot \frac{\gamma_{nD}}{G_m} \cdot \sqrt{\frac{\beta_{nG}}{\gamma_{nD}} \cdot \left(1 - c_g^2\right)}$$

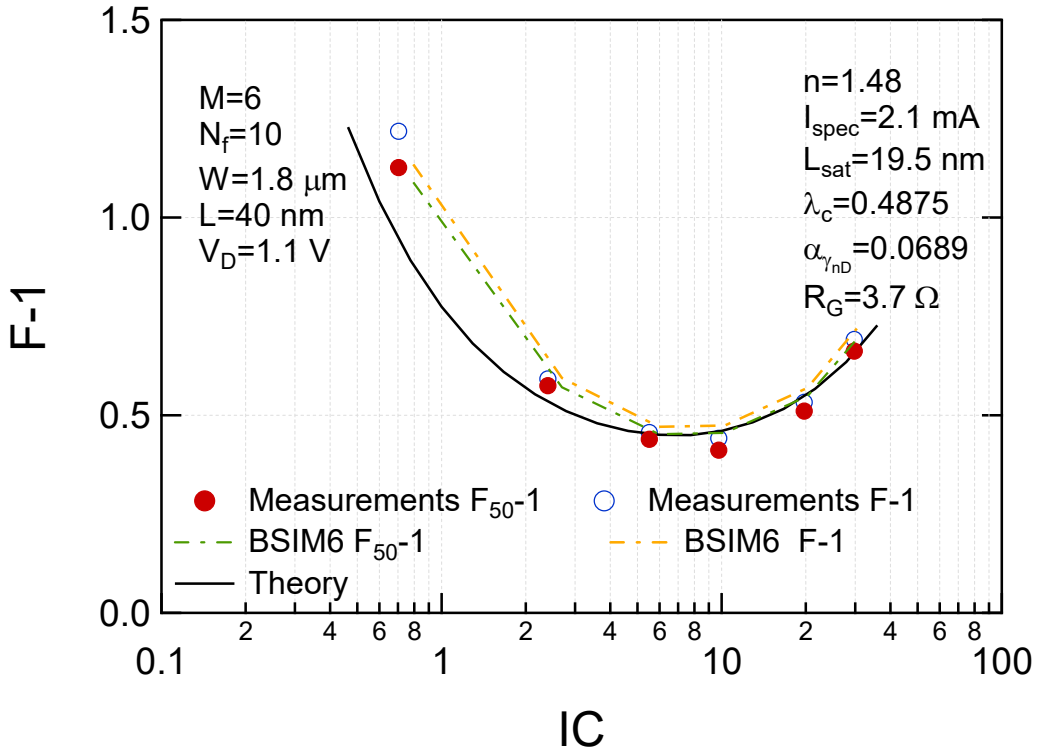
$$R_n \cong \frac{\gamma_{nD}}{G_m} + R_G$$

M. Chalkiadaki and C. Enz, TMTT, July 2015.

Actual Noise Figure

- The actual noise figure also shows a minimum in MI





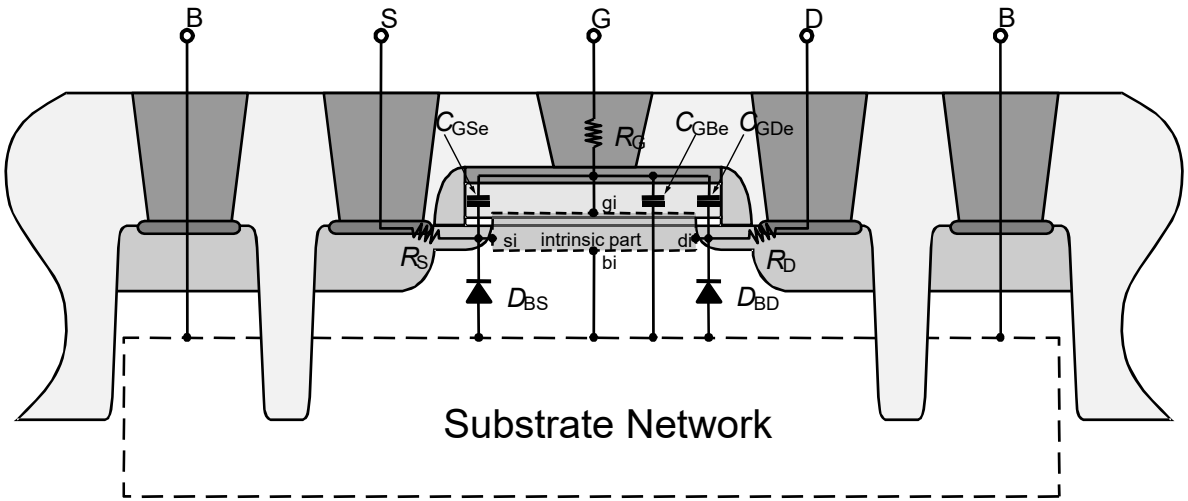
$$F-1=\frac{1}{50\Omega}\cdot\left(\frac{\gamma_{nD}}{G_m}+R_G\right)$$

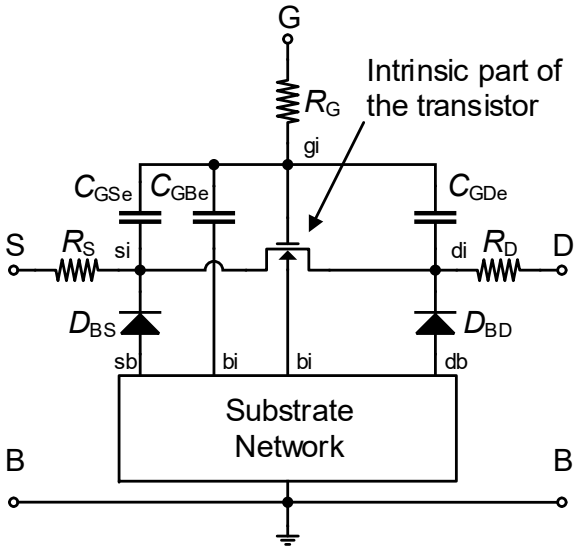
M. Chalkiadaki, PhD Thesis 2016
M. Chalkiadaki and C. Enz, TMTT, July 2015.

Outline

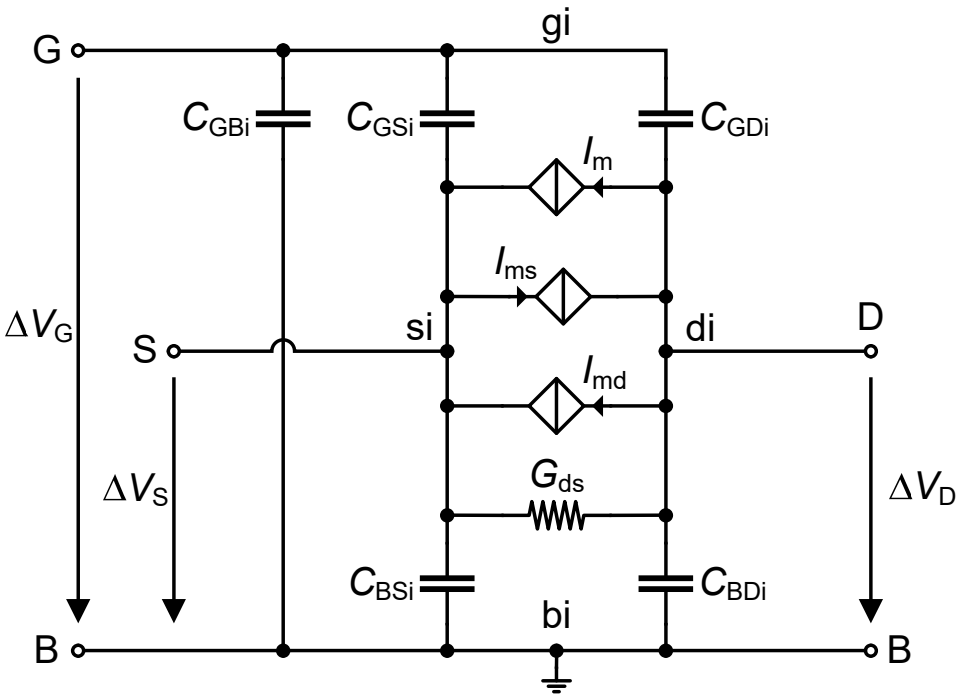
- Introduction
- Transistor Figures-of-Merit (FoM)
- **Equivalent Circuit at RF**
- Large-signal Model at RF

Equivalent Circuit at RF





Intrinsic Quasi-Static Small-signal Model



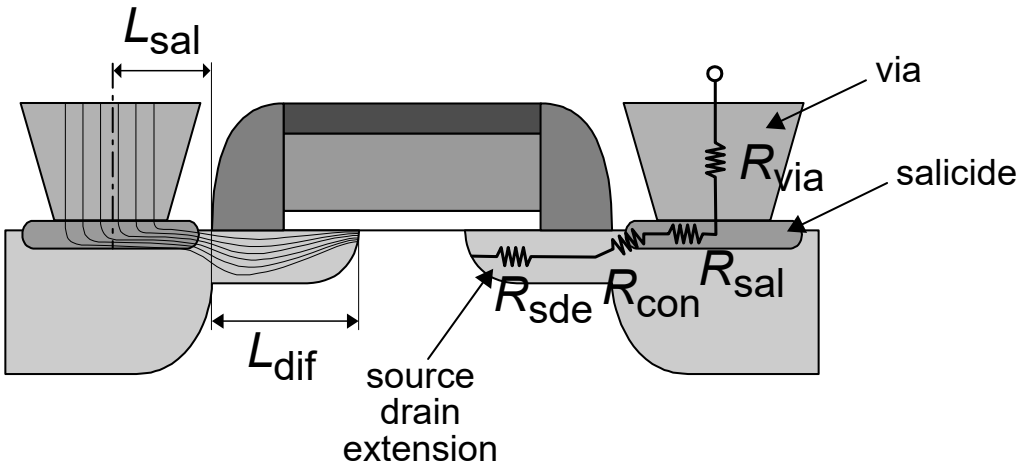
Channel time constant τ_{qs} defined as the propagation time along the channel and equal to the inverse of the quasi-static frequency ω_{qs}

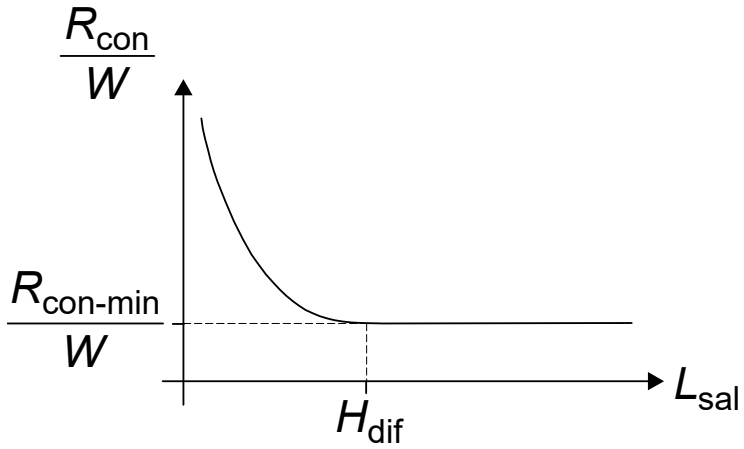
$$\tau_{qs} = \frac{1}{\omega_{qs}} = \frac{C_{ms}}{G_{ms}} = \frac{C_{md}}{G_{md}} = \frac{C_m}{G_m}$$

$$I_m = Y_m \cdot \Delta V_G$$
$$I_{ms} = Y_{ms} \cdot \Delta V_S$$
$$I_{md} = Y_{md} \cdot \Delta V_D$$

$$Y_m = G_m \cdot \left(1 - j\omega \cdot \tau_{qs}\right) = G_m - j\omega \cdot C_m$$
$$Y_{ms} = G_{ms} \cdot \left(1 - j\omega \cdot \tau_{qs}\right) = G_{ms} - j\omega \cdot C_{ms}$$
$$Y_{md} = G_{md} \cdot \left(1 - j\omega \cdot \tau_{qs}\right) = G_{md} - j\omega \cdot C_{md}$$

Source and Drain Resistances Scaling

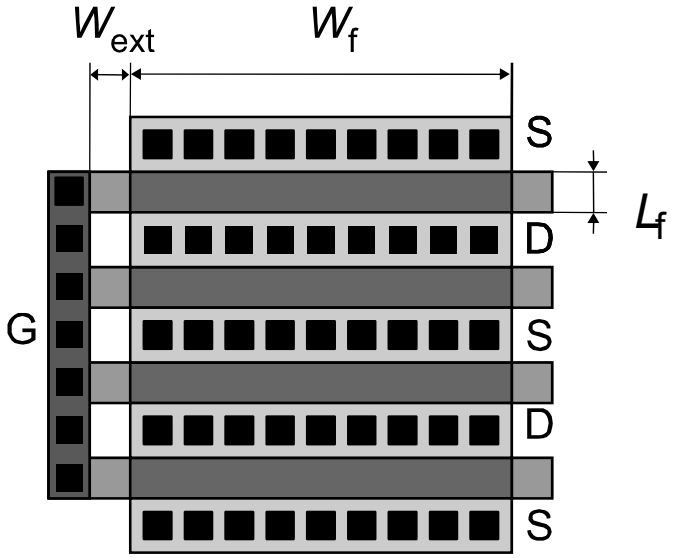


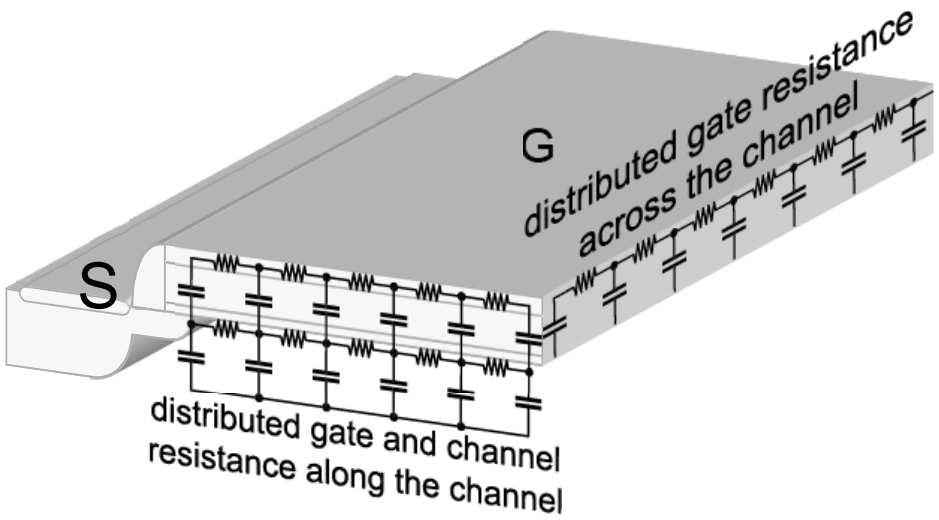


$$R_{S(D)} = R_{sde} + R_{con} + R_{sal} + R_{via} \cong R_{sde} + R_{con} \propto \frac{1}{W} \qquad W \triangleq N_f \cdot W_f$$

- R_S and R_D dominated by contact and source/drain extensions (SDE) resistances

Gate Resistance





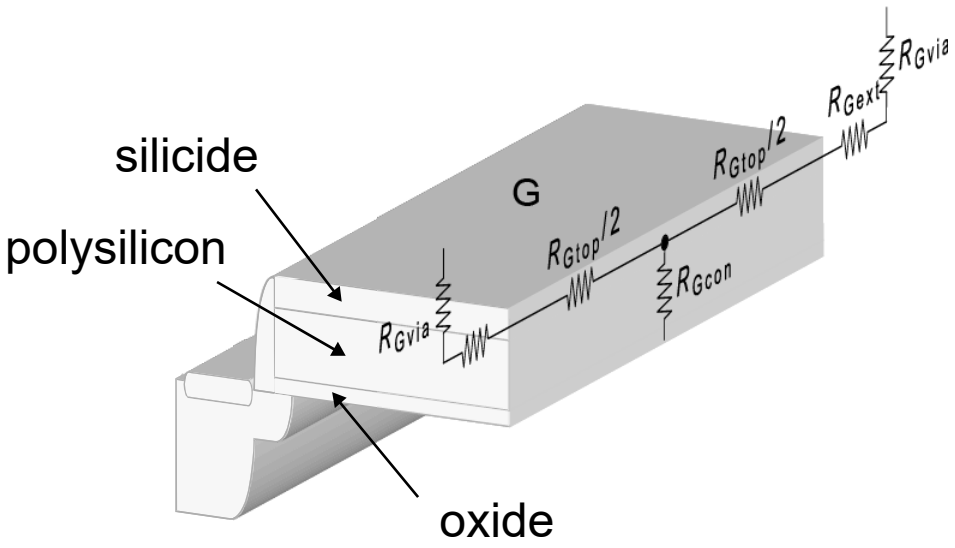
$$R_G = R_{Gtop} + R_{Gext} + R_{Gvia} + R_{Gcon}$$

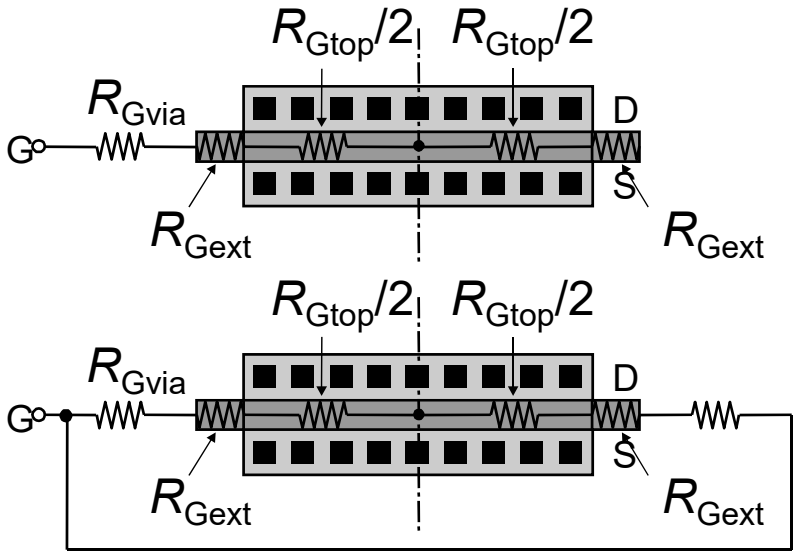
Salicide resistance: $R_{Gtop} = \frac{1}{3} \cdot \frac{W_f}{N_f \cdot L_f} \cdot R_{Gsq}$ $R_{Gext} = \frac{W_{ext}}{N_f \cdot L_f} \cdot R_{Gsq}$

Via resistance: $R_{Gvia} = \frac{R_{via}}{N_{via}}$ Silicide to poly contact resistance: $R_{Gcon} = \frac{\rho_{con}}{N_f \cdot W_f \cdot L_f}$

Where R_{Gsq} is the gate salicide resistance per squares (typically 3 Ω/sq) ρ_{con} is the silicide to poly contact resistance per area (typically 20 Ω/µm^2)

Gate Resistance





- Connecting the gate at both ends and assuming the metal has negligible resistance

$$R_G \cong \frac{R_{Gtop}}{4} + \frac{R_{Gext}}{2} + \frac{R_{Gvia}}{2} + R_{Gcon}$$

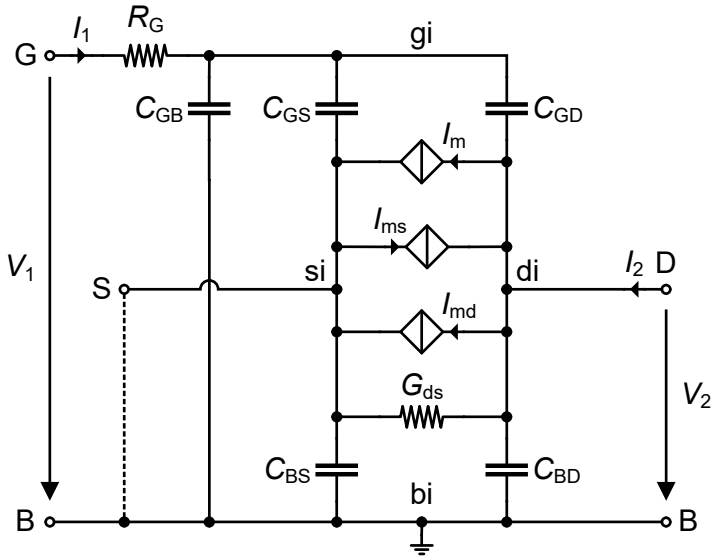
Which is about **4 times smaller** than the resistance of a gate contacted only on one side

Capacitances Scaling

$$C \propto W \qquad R_S, R_D \propto \frac{1}{W} \qquad W \triangleq N_f \cdot W_f$$

- The different RC time constants due to R_S and R_D do not depend on W but only on the gate length L_f and overlap length L_{ov}
- For a minimum length device, the poles due to R_S and R_D are at a much higher frequency than the transit frequency F_t and can therefore be neglected when calculating the Y-parameters
- Neglecting the substrate network for the moment leads to the following small-signal schematic which will be used for deriving the Y-parameters

$$C_{GS} = C_{GSi} + C_{GSe}$$
$$C_{GD} = C_{GDi} + C_{GDe}$$
$$C_{GB} = C_{GBi} + C_{GBe}$$
$$C_{BS} = C_{BSi} + C_{BSj}$$
$$C_{BD} = C_{BDi} + C_{BDj}$$



$$I_m = Y_m \cdot \left(V(gi) - V(bi)\right)$$
$$I_{ms} = Y_{ms} \cdot \left(V(si) - V(bi)\right)$$
$$I_{md} = Y_{md} \cdot \left(V(di) - V(bi)\right)$$
$$Y_m = G_m - j\omega C_m$$
$$Y_{ms} = G_{ms} - j\omega C_{ms}$$
$$Y_{md} = G_{md} - j\omega C_{md}$$

Approximate Y-parameters

- Neglecting R_S and R_D and the substrate network

$$Y_{11} \cong \frac{j\omega C_G}{1+j\omega R_G C_G}$$

$$Y_{12} \cong \frac{-j\omega C_{GD}}{1+j\omega R_G C_G}$$

$$Y_{21} \cong \frac{G_m - j\omega\cdot(C_{GD}+C_m)}{1+j\omega R_G C_G}$$

$$Y_{22} \cong \frac{G_{ds} + \omega^2 R_G C_{GD} C_m + j\omega\cdot(C_{GD}+C_{BD})}{1+j\omega R_G C_G}$$

$$C_G \triangleq C_{GS} + C_{GD} + C_{GB}$$

- Assuming $\omega R_G C_G \ll 1$

$$\frac{1}{1+j\omega R_G C_G} \cong 1 - j\omega R_G C_G \quad \text{for} \quad \omega R_G C_G \ll 1$$

$$Y_{11} \cong \omega^2 R_G C_G^2 + j\omega C_G$$

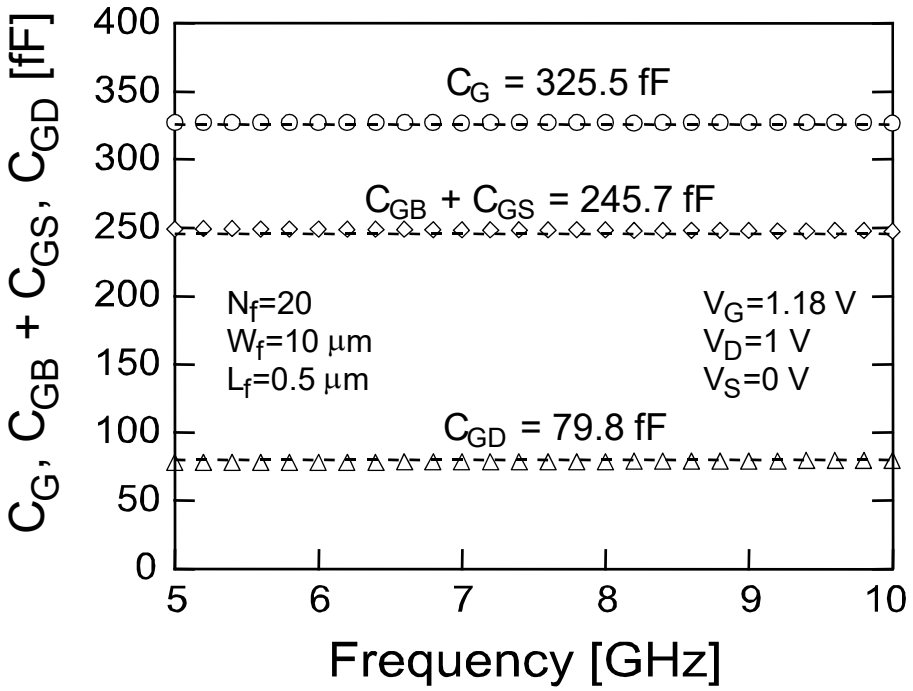
$$Y_{12} \cong -\omega^2 R_G C_G C_{GD} - j\omega C_{GD}$$

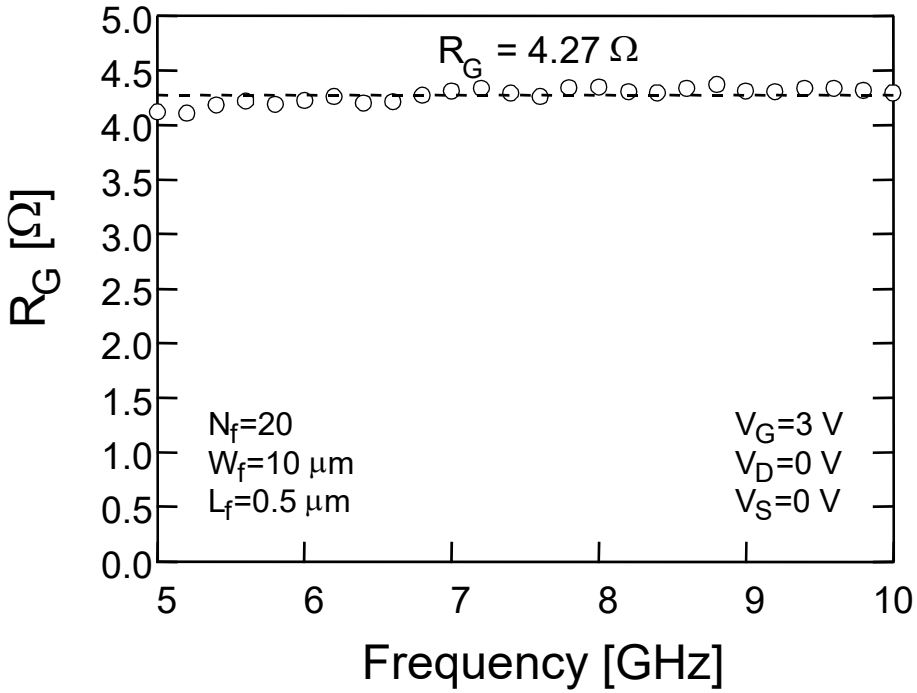
$$Y_{21} \cong G_m - \omega^2 R_G C_G\cdot(C_m + C_{GD}) - j\omega\cdot(C_m + C_{GD})$$

$$Y_{22} \cong G_{ds} + \omega^2 R_G\cdot(C_G C_{BD} + C_G C_{GD} + C_{GD} C_m) + j\omega\cdot(C_{BD} + C_{GD})$$

- Can be used for direct extraction of components from measured data

Direct Extraction of Small-Signal Circuit Components



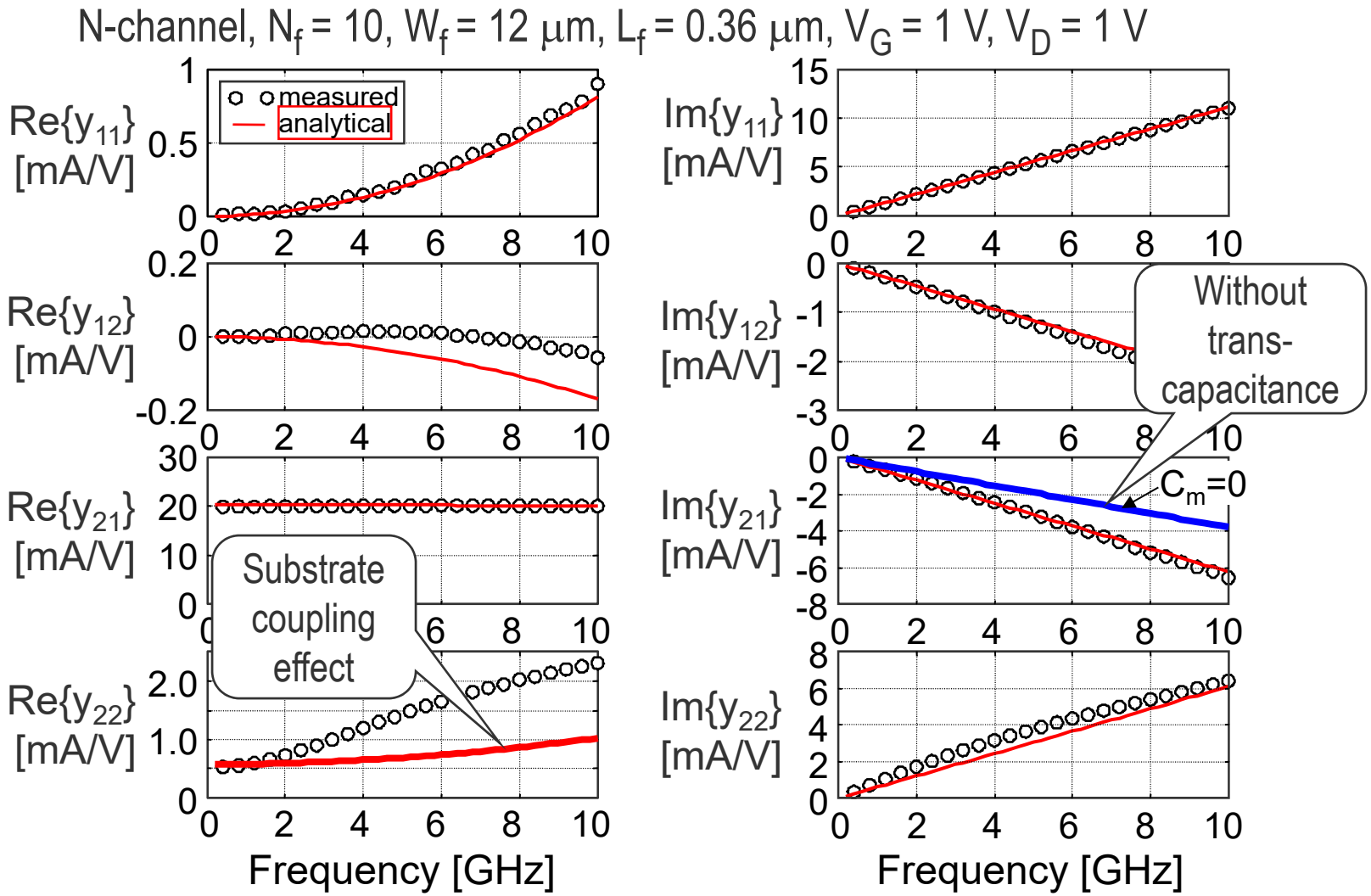


$$C_G = \frac{\Im\{Y_{11}\}}{\omega}$$

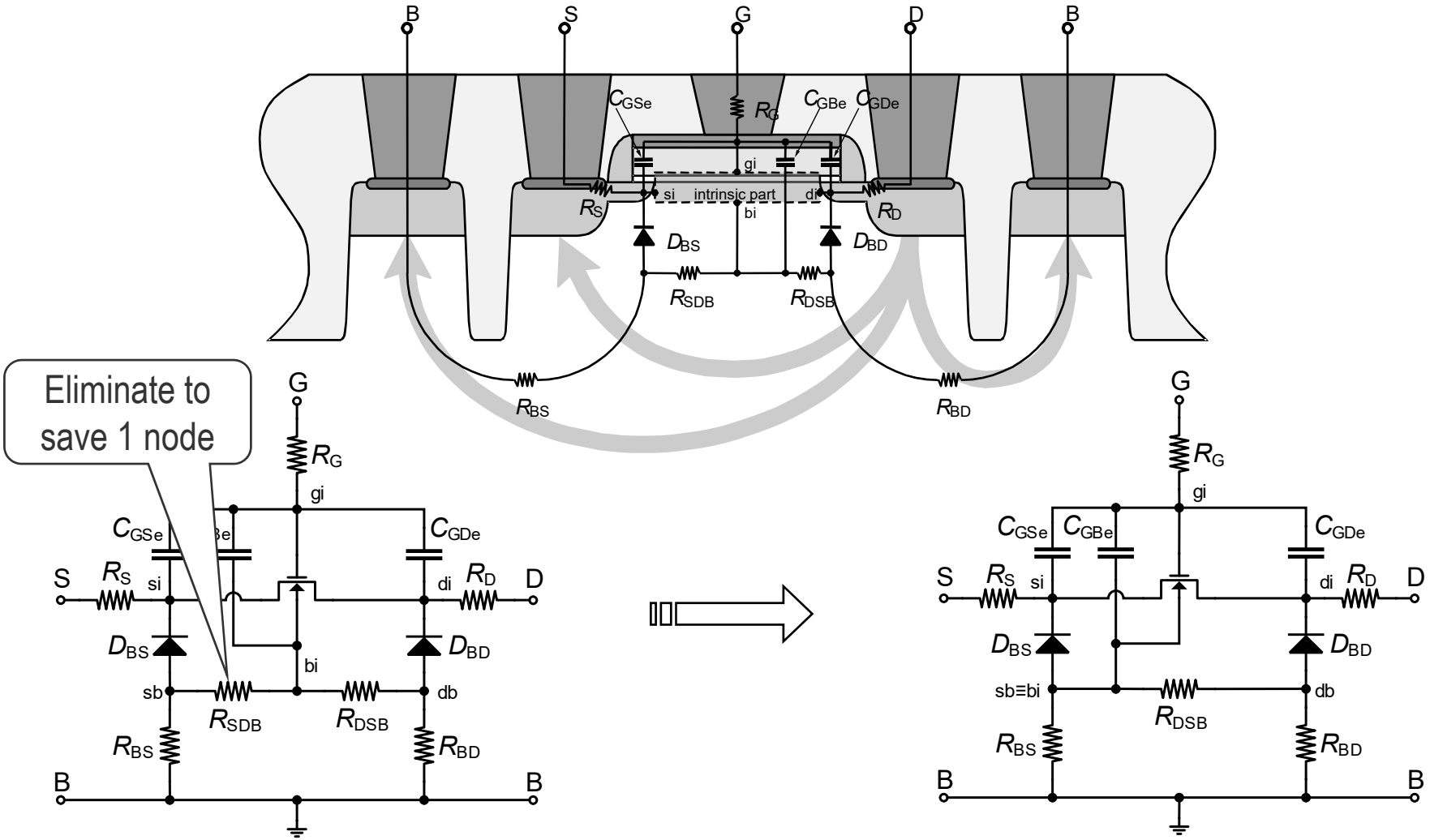
$$C_{GD} = \frac{\Im\{Y_{12}\}}{\omega}$$

$$R_G = \frac{\Re\{Y_{11}\}}{\Im\{Y_{11}\}^2}$$

Measured versus Analytical Y-parameters

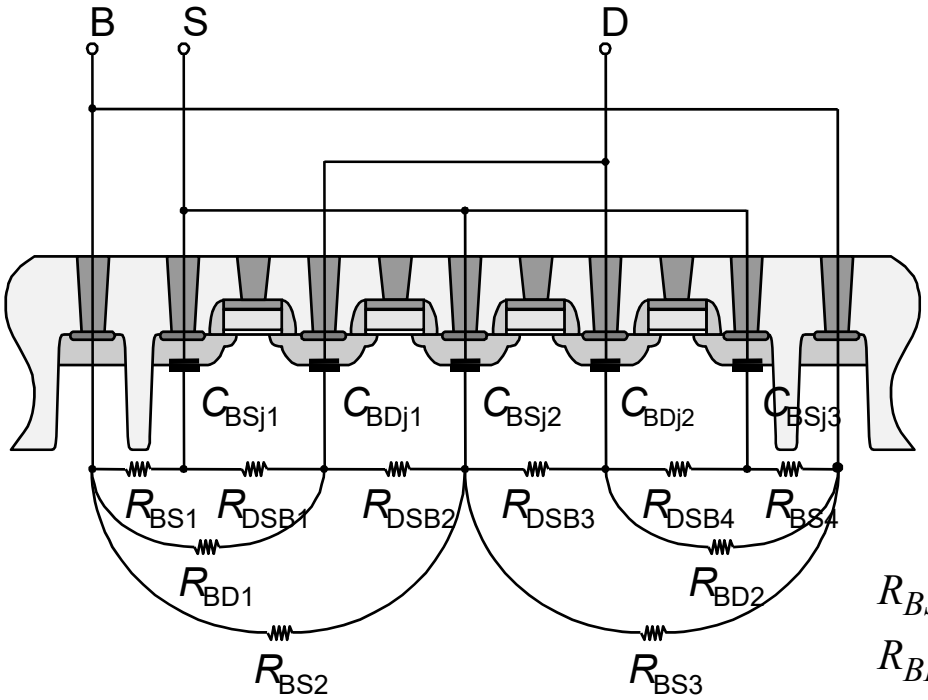


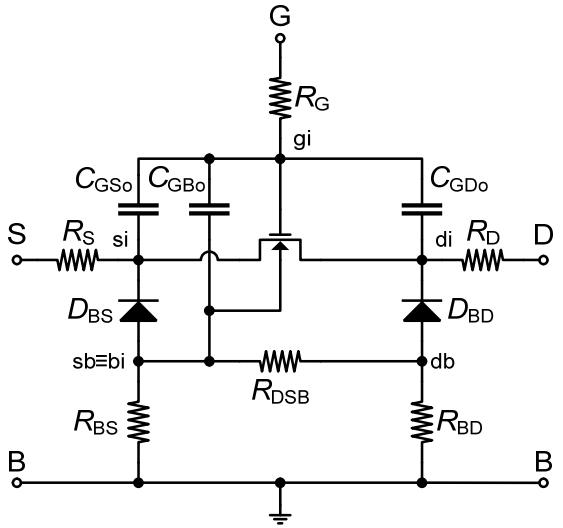
Intra-device Substrate Coupling



- Saves one component and one node, but makes the circuit asymmetric

Substrate Resistive Network – Even Number of Fingers





$$C_{BSj} = \sum_{k=1}^{N_s} C_{BSjk} \cong N_s \cdot C_{BSjf}$$

$$C_{BDj} = \sum_{k=1}^{N_d} C_{BDjk} \cong N_d \cdot C_{BSjf}$$

$$R_{BS1} \cong R_{BS4}$$

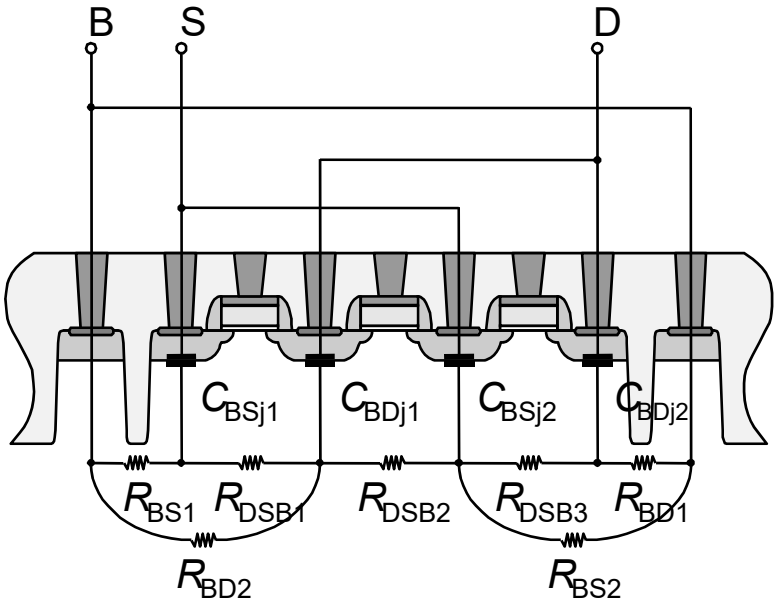
$$R_{BD1} \cong R_{BD2}$$

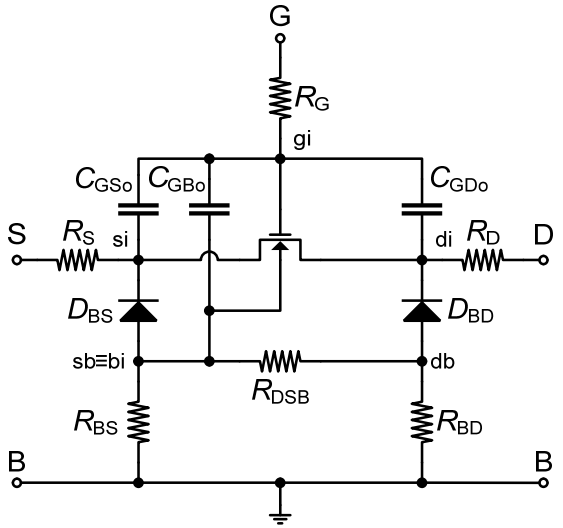
$$\frac{1}{R_{DSB}} = \sum_{k=1}^{N_f} \frac{1}{R_{DSBk}} \cong \left(\frac{L_f}{N_f \cdot W_f} \cdot R_{DSB-sh} \right)^{-1}$$

$$\frac{1}{R_{BS}} = \sum_{k=1}^{N_s} \frac{1}{R_{BSk}} \cong \frac{1}{R_{BS1}} + \frac{1}{R_{BS4}} \cong \frac{2}{R_{BS1}} \cong \frac{2W_f}{r_{BS-end}}$$

$$\frac{1}{R_{BD}} = \sum_{k=1}^{N_d} \frac{1}{R_{BDk}} \cong \frac{1}{R_{BD1}} + \frac{1}{R_{BD2}} \cong \frac{2}{R_{BD1}} \cong \frac{2W_f}{r_{BD-end}}$$

Substrate Resistive Network – Odd Number of Fingers





$$R_{BS1} \ll R_{BS2}$$

$$R_{BD1} \ll R_{BD2}$$

$$R_{BS1} \cong R_{BD1}$$

$$C_{BSj} = \sum_{k=1}^{N_s} C_{BSjk} \cong N_s \cdot C_{BSjf}$$

$$C_{BDj} = \sum_{k=1}^{N_d} C_{BDjk} \cong N_d \cdot C_{BSjf}$$

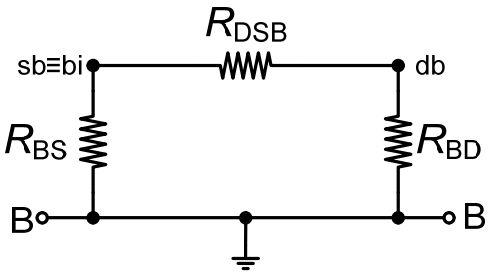
$$\frac{1}{R_{DSB}} = \sum_{k=1}^{N_f} \frac{1}{R_{DSBk}} \cong \left(\frac{L_f}{N_f \cdot W_f} \cdot R_{DSB-sh} \right)^{-1}$$

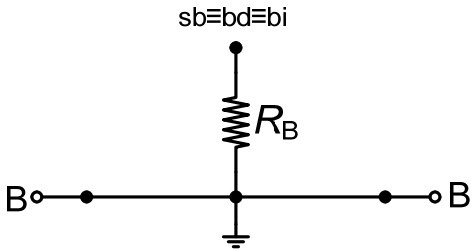
$$R_{BS} \cong R_{BD} \cong R_{BS1}$$

$$\frac{1}{R_{BS}} \cong \frac{W_f}{r_{BS-end}}$$

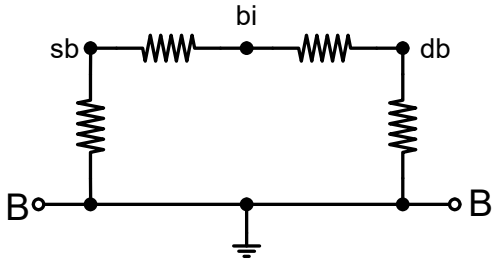
Simplified Substrate Networks

- Since R_{DSB} is inversely proportional to N_f, for RF MOS transistors with many fingers ($N_f > 4$) $R_{DSB} \ll R_{BS}, R_{DSB} \ll R_{BD}$ hence R_{DSB} can be neglected
- R_{BS} and R_{BD} are then connected in parallel and result in a single substrate resistance R_B which is often enough for capturing first-order intra-device substrate coupling effects and additional substrate thermal noise

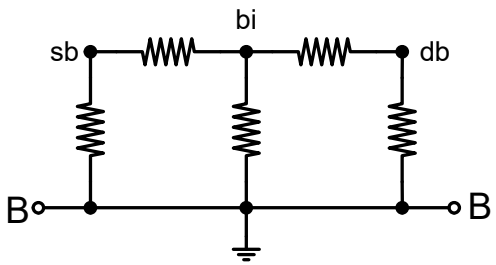




- Other substrate networks have been published

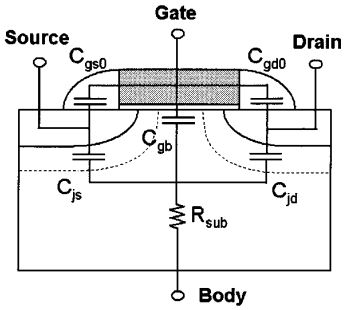


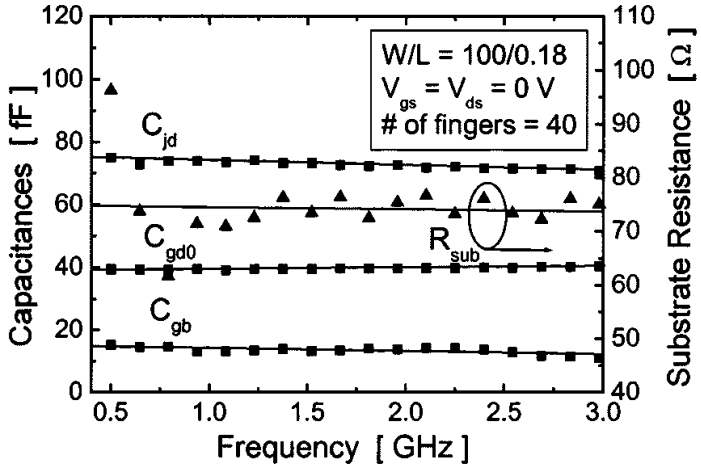
[Liu, IEDM 97]

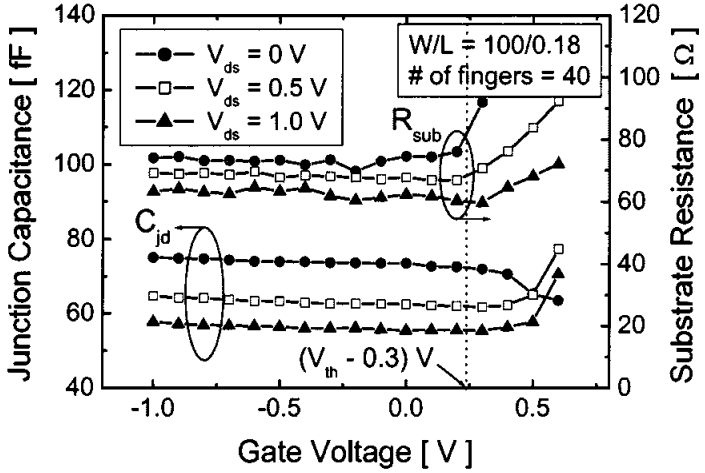


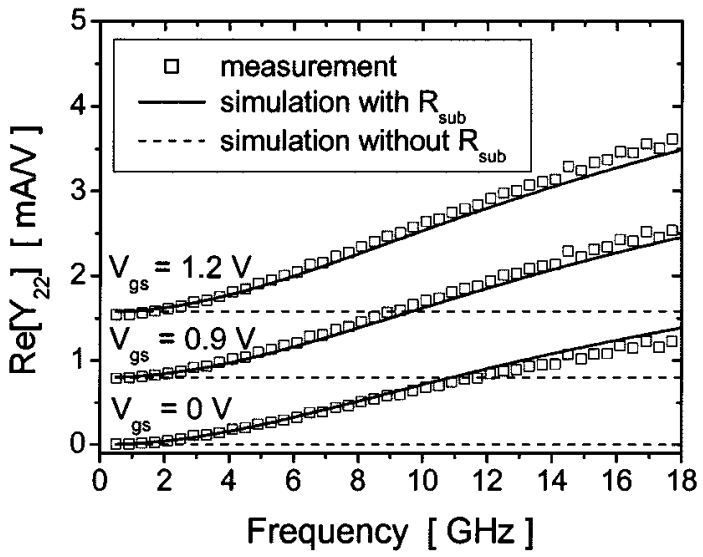
Epitaxial process [Tiemeijer, ESSDERC 98]

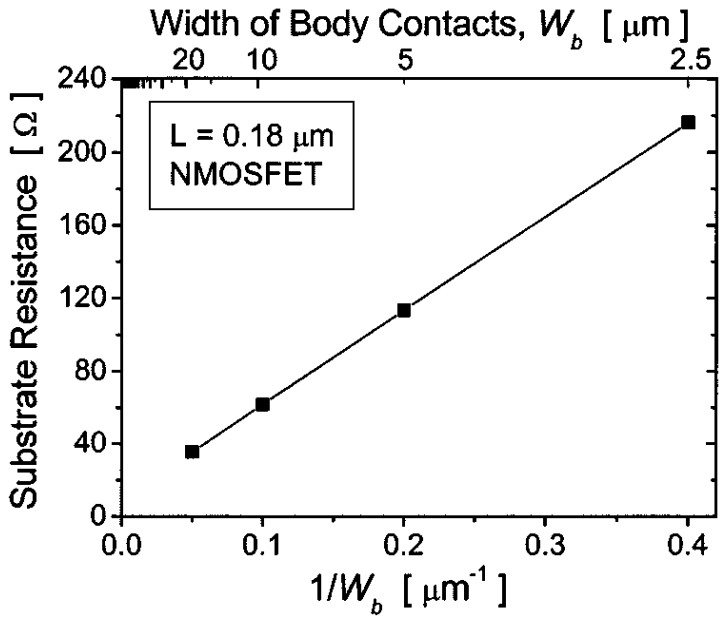
Substrate Resistance Extraction





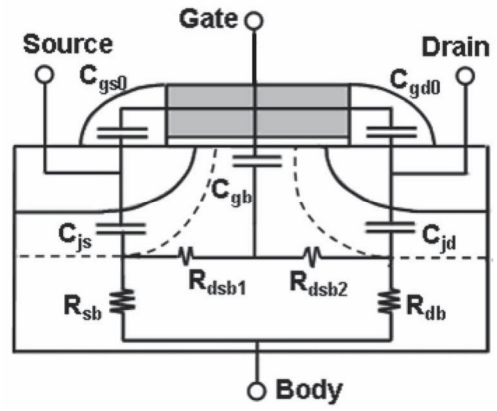


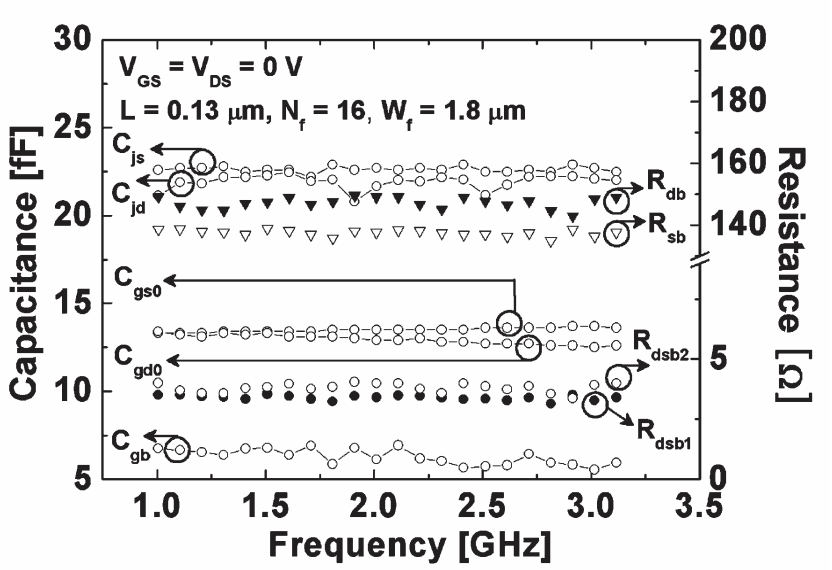


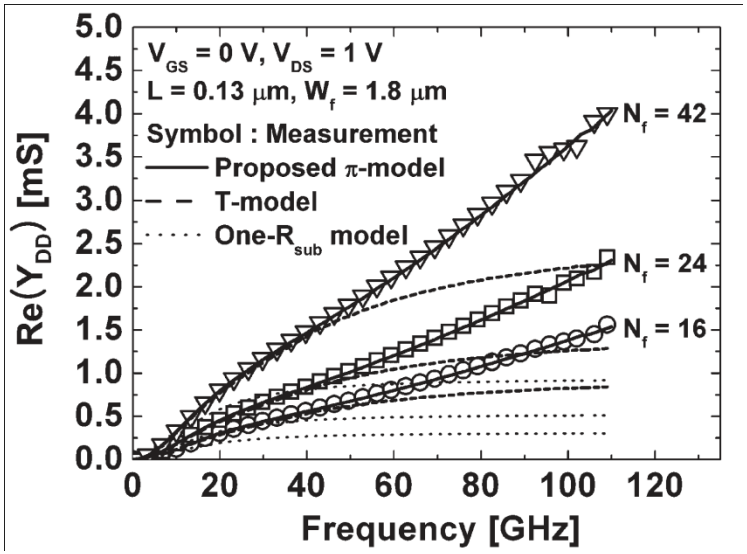


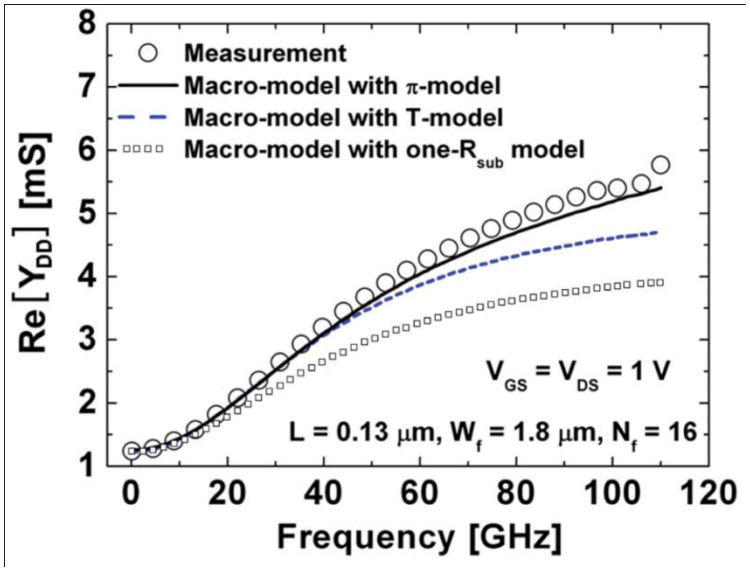
J. Han, M. Je, and H. Shin, EDL, July 2002.

Extraction of π-Type Substrate Resistance





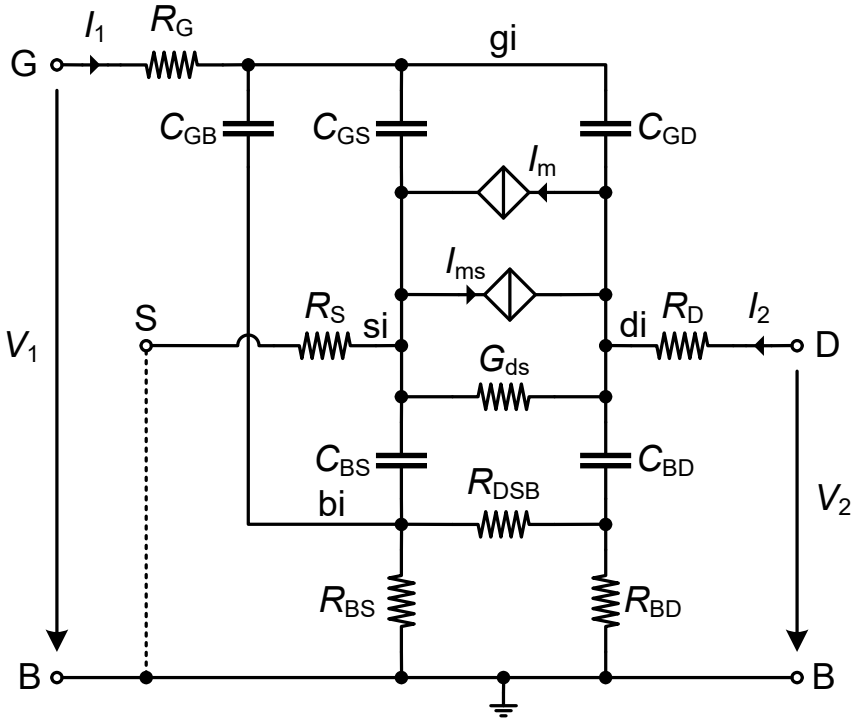




I. M. Kang, J. D. Lee.and H. Shin, EDL, May 2007.

Complete Equivalent Small-signal Circuit (Saturation)

$$C_{GS} = C_{GSi} + C_{GSe}$$
$$C_{GD} = C_{GDi} + C_{GDe}$$
$$C_{GB} = C_{GBi} + C_{GBe}$$
$$C_{BS} = C_{BSi} + C_{BSj}$$
$$C_{BD} = C_{BDi} + C_{BDj}$$



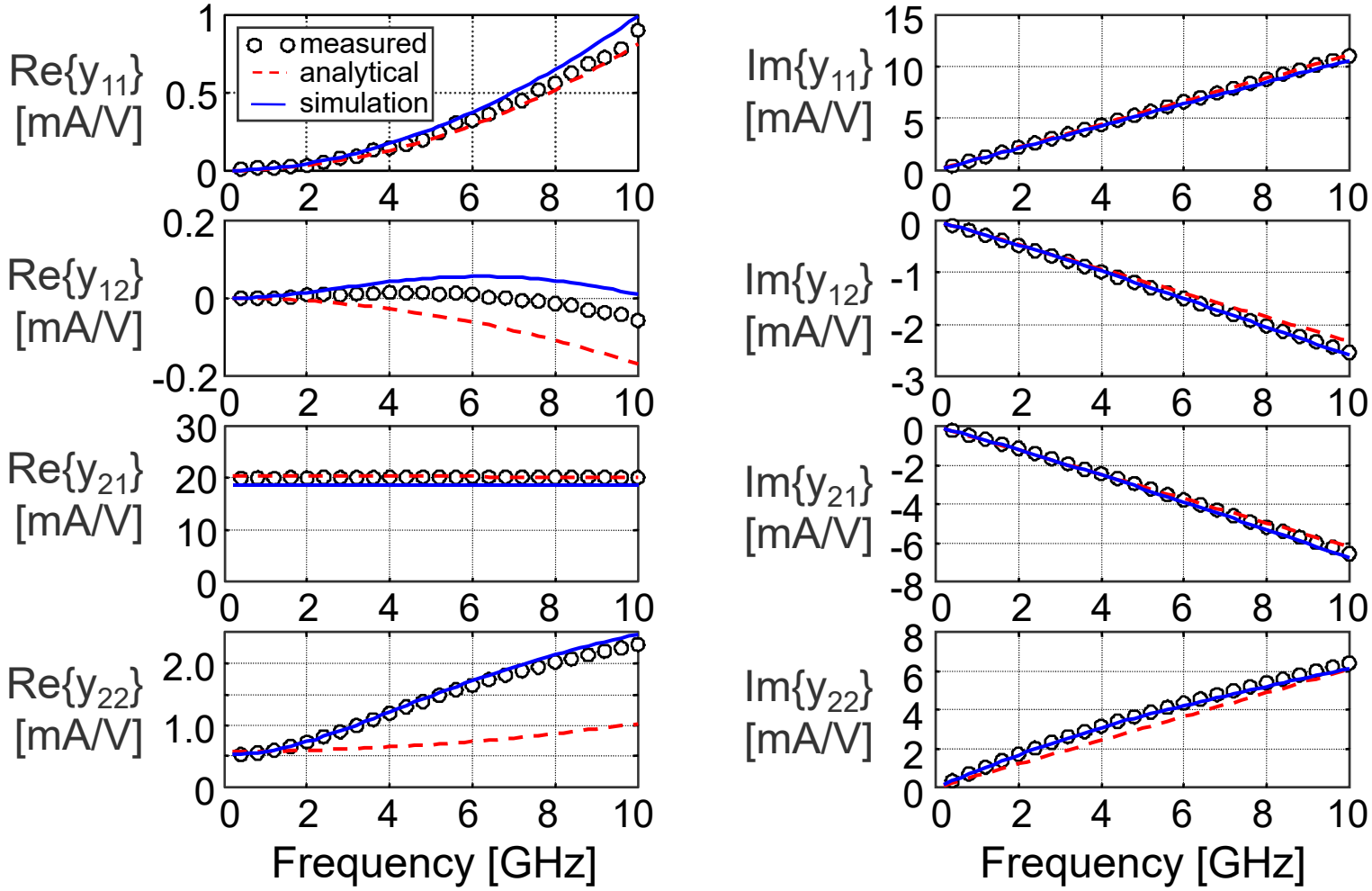
$$I_m = Y_m \cdot \left(V(gi) - V(bi)\right)$$
$$I_{ms} = Y_{ms} \cdot \left(V(si) - V(bi)\right)$$

$$Y_m = \frac{Y_{ms}}{n}$$
$$G_m = \frac{G_{ms}}{n}$$
$$C_m = \frac{C_{ms}}{n}$$

$$Y_m = G_m \cdot \left(1 - j\omega\tau_{qs}\right) = G_m - j\omega C_m$$
$$Y_{ms} = G_{ms} \cdot \left(1 - j\omega\tau_{qs}\right) = G_{ms} - j\omega C_{ms}$$

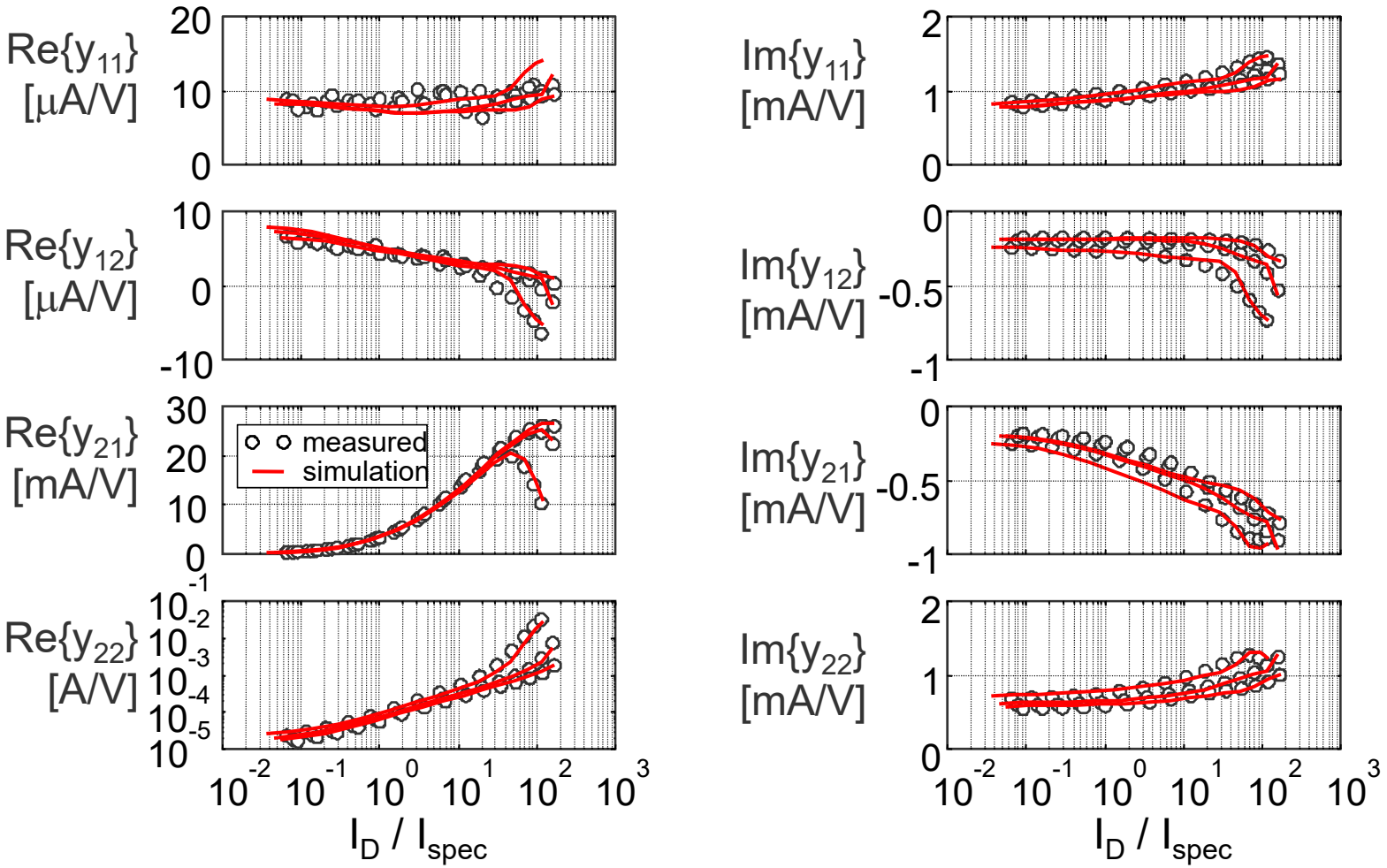
Measured versus Simulation for 0.35µm CMOS Process

N-channel, $N_f = 10$, $W_f = 12\ \mu m$, $L_f = 0.36\ \mu m$, $V_G = 1\ V$, $V_D = 1\ V$, EKV v2.6



Y-parameters versus Bias

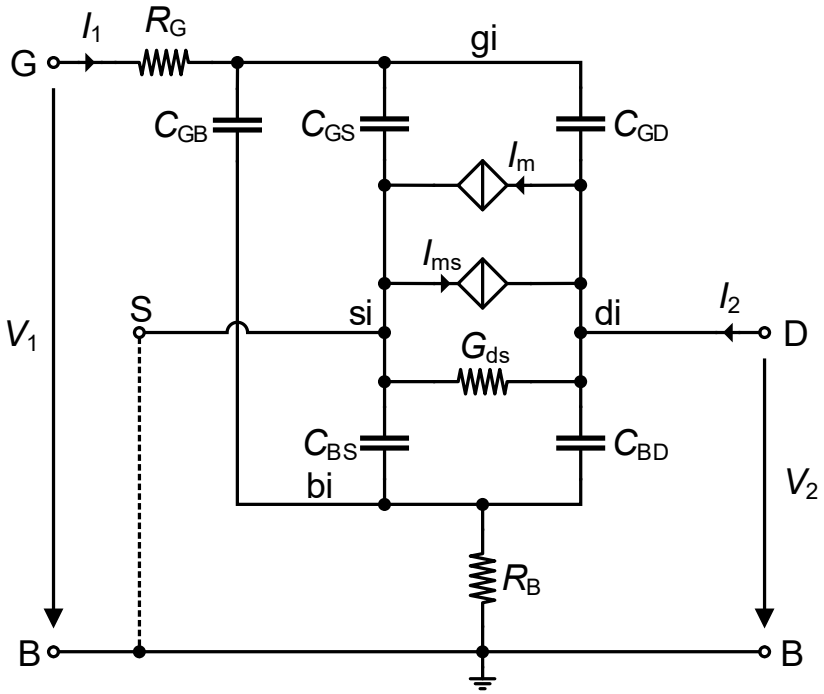
N-channel, N_f = 10, W_f = 12 µm, L_f = 0.36 µm, f = 1 GHz, V_D = 0.5, 1, 1.5 V, EKV v2.6



Simplified Equivalent Small-signal Circuit

- Neglecting again the poles due to R_S and R_D assuming that they are at a much higher frequency than the transit frequency F_t
- For large number of fingers ($N_f > 4$), the substrate can be replaced by a **single substrate resistance** R_B, leading to the following small-signal common-source schematic in saturation which will be used for deriving the Y-parameters

$$C_{GS} = C_{GSi} + C_{GSe}$$
$$C_{GD} = C_{GDi} + C_{GDe}$$
$$C_{GB} = C_{GBi} + C_{GBe}$$
$$C_{BS} = C_{BSi} + C_{BSj}$$
$$C_{BD} = C_{BDi} + C_{BDj}$$



$$I_m = Y_m \cdot \left(V(gi) - V(bi)\right)$$
$$I_{ms} = Y_{ms} \cdot \left(V(si) - V(bi)\right)$$
$$Y_m = G_m - j\omega C_m$$
$$Y_{ms} = G_{ms} - j\omega C_{ms}$$

Approximate Y-parameters

- Assuming that

$$\omega^2\left[R_B^2C_B^2+R_G\left(R_GC_G^2+2R_BC_{GB}^2\right)\right]+\omega^4R_G^2R_B^2\left(C_{GB}^2-C_BC_G\right)^2\ll 1$$

- which can be valid for operating frequencies up to the low THz range
- Neglecting: (i) all the higher than second order terms and (ii) the least dominant terms, the simplified expressions for the Y-parameters in saturation can be derived as

$$Y_{11}\cong\omega^2\left(R_GC_G^2+R_BC_{GB}^2\right)+j\omega C_G$$

$$Y_{12}\cong\omega^2\left(R_BC_{BD}C_{GB}-R_GC_{GD}C_G\right)-j\omega C_{GD}$$

$$Y_{21}\cong G_{m\text{-}eff}+\omega^2\left[R_BC_{GB}\left(C_{BD}-C_m-C_{ms}\right)-R_GC_G\left(C_{GD}+C_m\right)\right]-j\omega\cdot\left(C_{GD}+C_m\right)$$

$$Y_{22}\cong G_{ds}+\omega^2\left[R_BC_{BD}\left(C_{BD}-C_m+C_{ms}\right)+R_GC_{GD}\left(C_{GD}+C_m\right)\right]+j\omega\cdot\left(C_{GD}+C_{BD}\right)$$

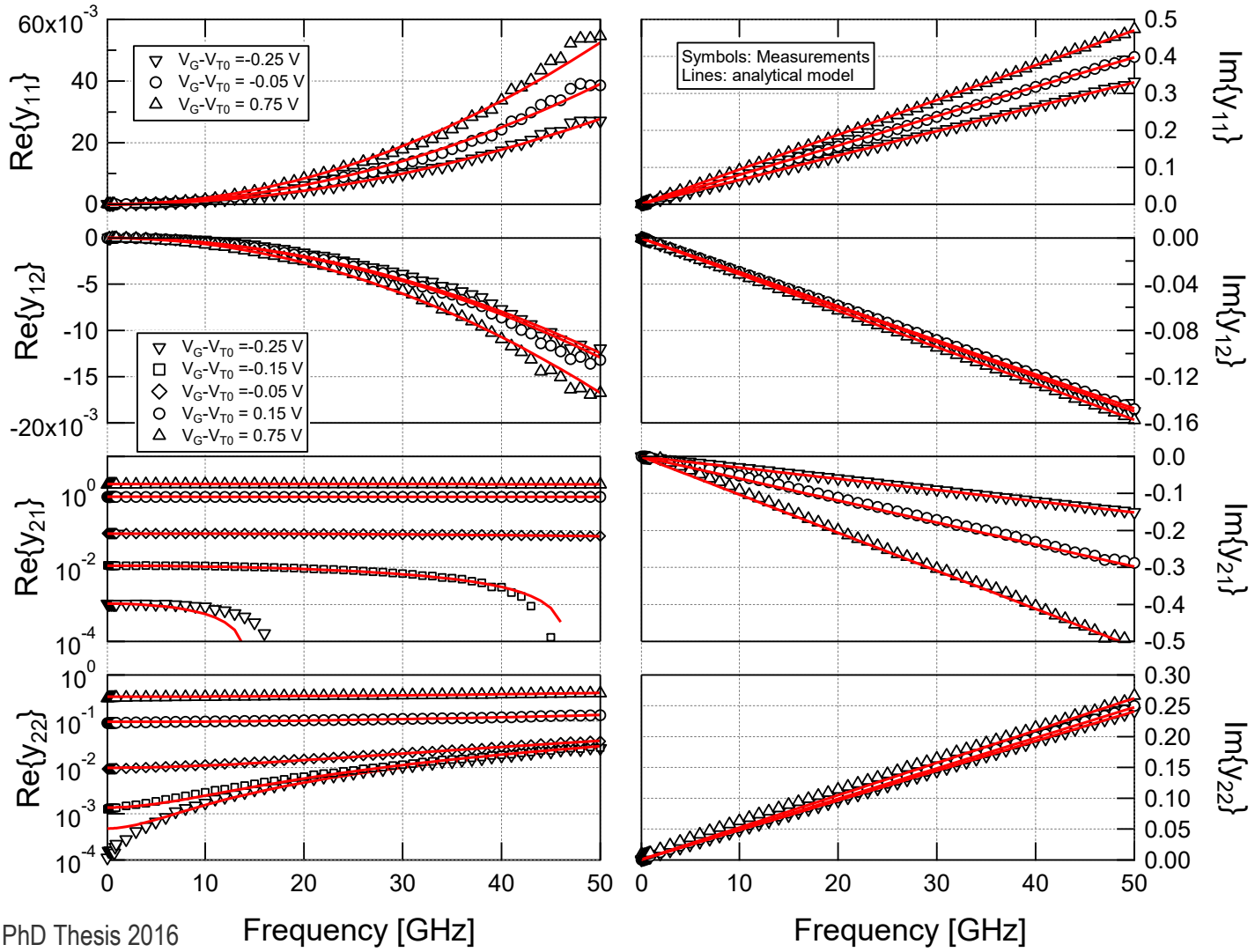
- Where G_{m-eff} is the effective gate transconductance accounting for the degradation due to the source resistance

$$G_{m\text{-}eff}\triangleq\frac{G_m}{1+G_{ms}\cdot R_S}$$

📖 M. Chalkiadaki and C. Enz, TMTT, 2015.

Measured versus Analytical for 40nm CMOS Process

N-channel, M=10, N_f = 10, W_f = 2 μm, L_f = 40 nm, V_S = 0 V, V_D = 1.1 V

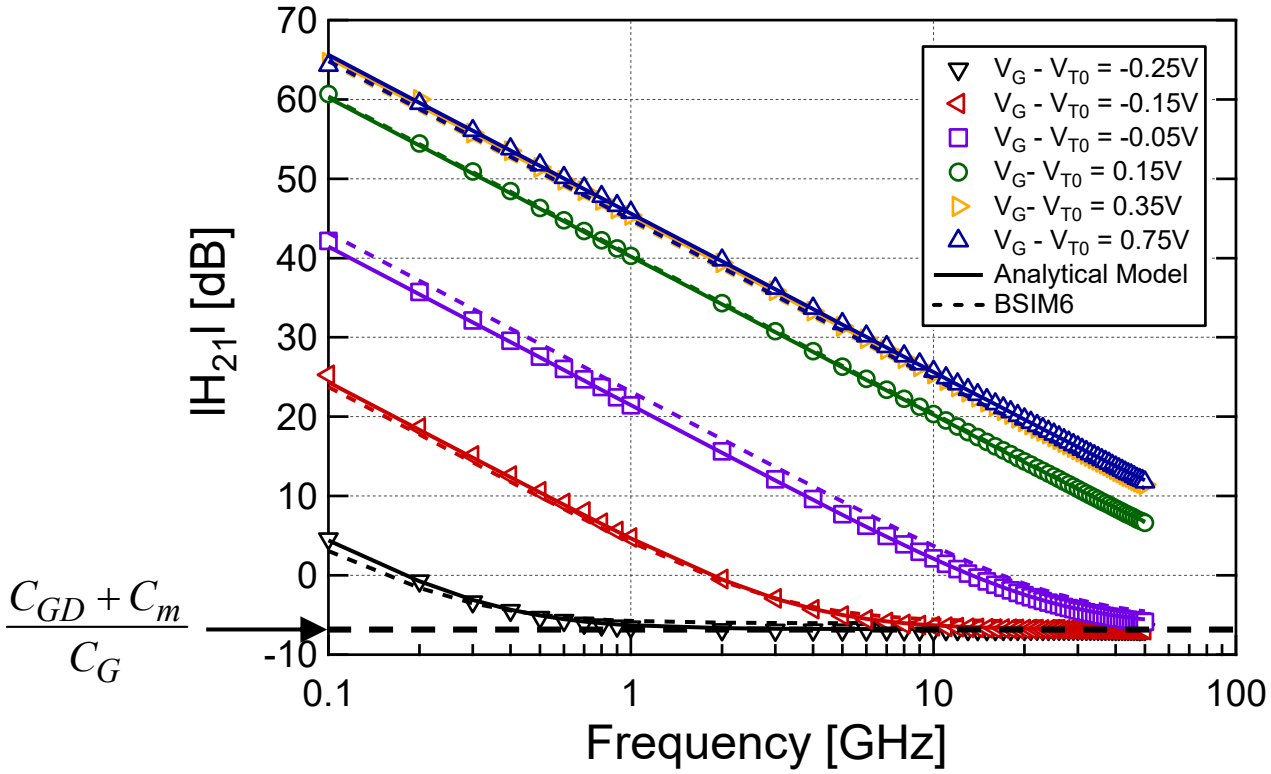


M. Chalkiadaki, PhD Thesis 2016

M. Chalkiadaki and C. Enz, TMTT, July 2015.

Current Gain vs Frequency for 40nm CMOS Process

$$H_{21}(\omega) \cong \frac{G_{m\text{-}eff} - j\omega\left(C_{GD} + C_m\right)}{j\omega C_G}$$

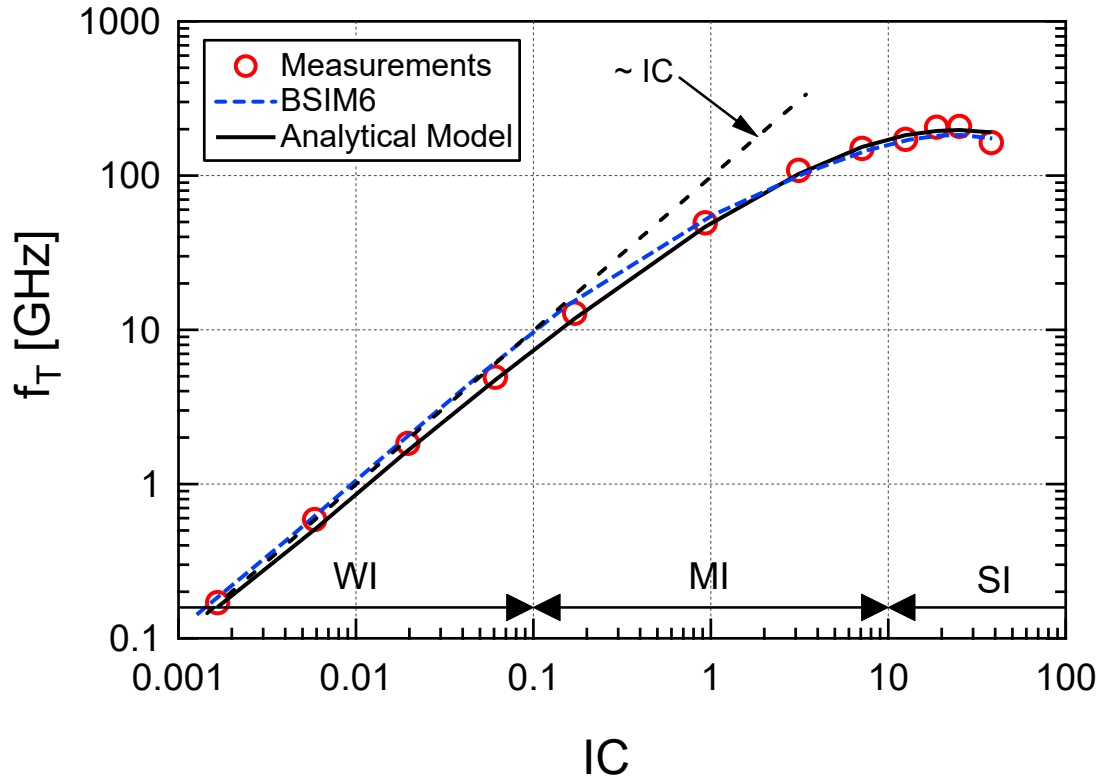


M. Chalkiadaki, PhD Thesis 2016

M. Chalkiadaki and C. Enz, TMTT, July 2015.

Transit Frequency vs IC for 40nm CMOS Process

$$f_t = \frac{G_{m\text{-}eff}}{2\pi\sqrt{C_G^2 - \left(C_{GD} + C_m\right)^2}} \cong \frac{G_{m\text{-}eff}}{2\pi C_G}$$



M. Chalkiadaki, PhD Thesis 2016

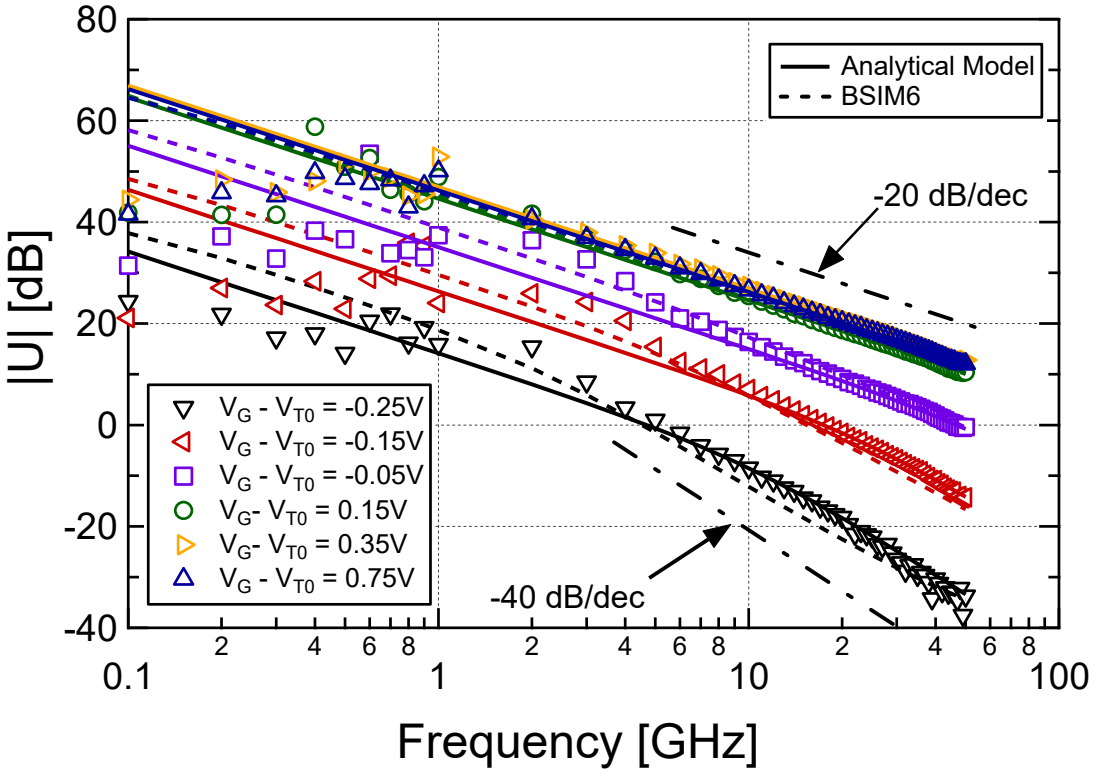
M. Chalkiadaki and C. Enz, TMTT, July 2015.

Unilateral Gain vs Frequency for 40nm CMOS Process

$$U = \frac{\omega_m^2}{\omega^2 \cdot \left(1 + \left(\frac{\omega}{\omega_p}\right)^2\right)}$$

$$\omega_m^2 \triangleq \frac{G_{m\text{-}eff}^2}{4\left[R_B C_{GB}\left(G_{ds}C_{GB} - G_{m\text{-}eff}C_{BD}\right) + R_G C_G\left(G_{ds}C_G + G_{m\text{-}eff}C_{GD}\right)\right]} \cong \frac{G_{m\text{-}eff}}{4\left[R_G C_G C_{GD} - R_B C_{GB} C_{BD}\right]}$$

$$\omega_p^2 \triangleq \frac{R_B C_{GB}\left(G_{ds}C_{GB} - G_{m\text{-}eff}C_{BD}\right) + R_G C_G\left(G_{ds}C_G + G_{m\text{-}eff}C_{GD}\right)}{R_B R_G\left(C_{BD}C_G + C_{GB}C_{GD}\right)\left(C_G\left(C_{BD} - C_m + C_{ms}\right) + C_{GB}\left(C_{GD} + C_m\right)\right)}$$

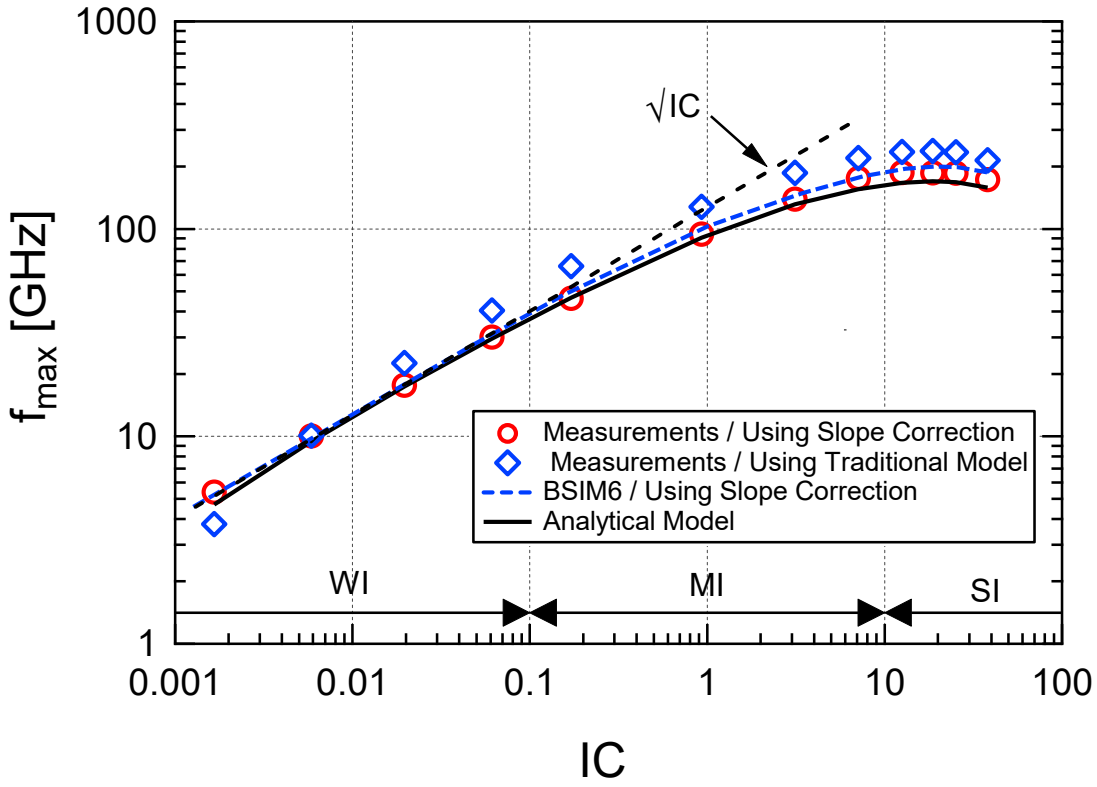


M. Chalkiadaki, PhD Thesis 2016

M. Chalkiadaki and C. Enz, TMTT, July 2015.

f_{max} versus IC for 40nm CMOS Process

$$f_{\max} \cong \frac{\sqrt{\omega_p \cdot \left(\sqrt{4K + \omega_p^2} - \omega_p\right)}}{2\sqrt{2}\pi}$$



M. Chalkiadaki, PhD Thesis 2016

M. Chalkiadaki and C. Enz, TMTT, July 2015.

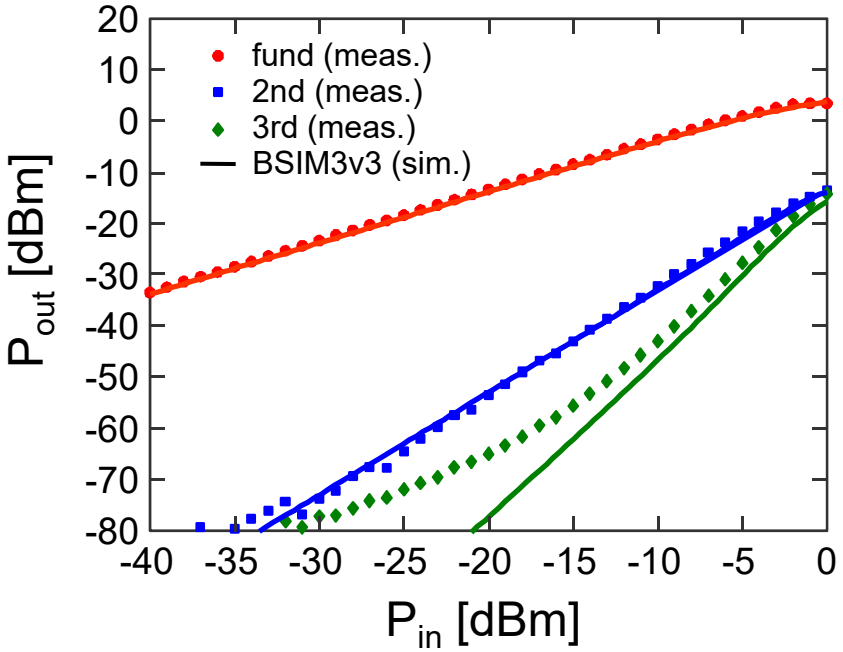
Outline

- Introduction
- Transistor Figures-of-Merit (FoM)
- Equivalent Circuit at RF
- **Large-signal Model at RF**

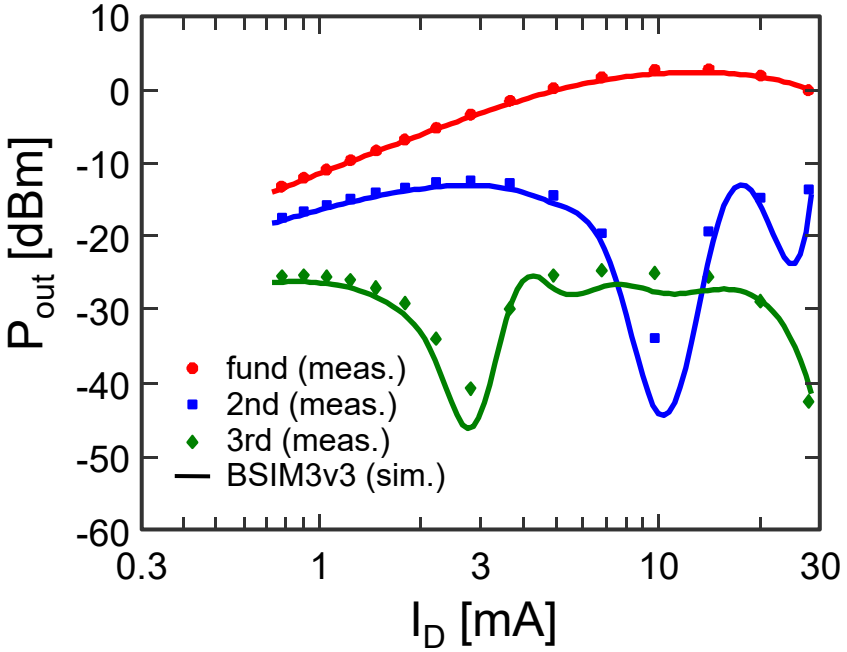
Large-signal Model at RF

- Dominated by static (DC) I-V nonlinearity (i.e. nonlinearities coming from the capacitors seem not to play a significant role)

$N_f = 10$, $W_f = 12\ \mu m$, $L_f = 0.36\ \mu m$
$V_G = 1.055$ V, $V_D = 1$ V, $V_S = 0$ V, $f_{spot} = 900$MHz



$N_f = 10$, $W_f = 12\ \mu m$, $L_f = 0.36\ \mu m$
$P_{in} = -4$ dBm, $V_D = 1$ V, $V_S = 0$ V, $f_{spot} = 900$ MHz



Drain Current Linearity

- While noise defines the minimum signal level, linearity is the main specification setting the largest signal that can be handled for a certain linearity requirement
- To evaluate the linearity of a MOS transistor, the normalized drain current in saturation can be approximated by the following Taylor series

$$i_{dsat} \cong i_{dsat0} + g_{m1} \cdot \Delta v_g + \frac{g_{m2}}{2} \cdot \Delta v_g^2 + \frac{g_{m3}}{6} \cdot \Delta v_g^3$$

- The evaluation of the normalized transconductances g_{mk} become very complex when calculated as derivatives with respect to the gate voltage. It is much easier to evaluate them by taking advantage of the charge formulation of the current according to

$$g_{mk} \triangleq \frac{G_{mk} \cdot n^k U_T^{k-1}}{G_{spec}} = \frac{\partial^k i_{dsat}}{\partial v_g^k} = \frac{\partial^k i_{dsat}}{\partial q_s^k} \cdot \left(\frac{\partial^k v_g}{\partial q_s^k} \right)^{-1}$$

F. Chicco, A. Pezzotta and C. Enz, “Charge-Based Distortion Analysis of Nanoscale MOSFETs,” TCAS1, Vol. 66 , No. 2, pp. 453-462, Feb. 2019.

Evaluation of the Transconductances

- The derivatives $\partial^k i_{dsat}/\partial q_s^k$ are calculated from the normalized drain current expression in saturation

$$i_{dsat}(q_s) = \frac{4\left(q_s^2 + q_s\right)}{2 + \lambda_c + \sqrt{4\left(1+\lambda_c\right) + \lambda_c^2 \cdot \left(1+2q_s\right)^2}}$$

- whereas the derivatives $\partial^k v_g/\partial q_s^k$ are calculated from the pinch-off voltage versus charge expression evaluated at the source

$$v_p - v_s \cong \frac{v_g - v_{t0}}{n} - v_s = 2q_s + \ln q_s$$

- This leads to the expressions of g_{mk} in terms of q_s given in the next slide
- They can be expressed in terms of the inversion coefficient IC replacing q_s by

$$q_s = \frac{\sqrt{\left(1+\lambda_c IC\right)^2 + 4IC} - 1}{2}$$

F. Chicco, A. Pezzotta and C. Enz, “Charge-Based Distortion Analysis of Nanoscale MOSFETs,” TCAS1, Vol. 66 , No. 2, pp. 453-462, Feb. 2019.

Normalized Transconductances

- The normalized transconductances g_{mk} and coefficients a, b, c and d are given by

$$g_{m1} = \frac{a-1}{\sqrt{4+4\lambda_c+a^2\lambda_c^2}},$$

$$g_{m2} = \frac{g_{m1}}{a}\cdot\frac{4+4\lambda_c+a\lambda_c^2}{4+4\lambda_c+a^2\lambda_c^2},$$

$$g_{m3} = \frac{g_{m1}}{a^3}\cdot\frac{16+32\lambda_c+8b\lambda_c^2+8a^2c\lambda_c^3+a^3d\lambda_c^4}{\left(4+4\lambda_c+a^2\lambda_c^2\right)^2}$$

$$a = 1+2q_s,$$

$$b = (1-2q_s)\left(3+7q_s+6q_s^2\right),$$

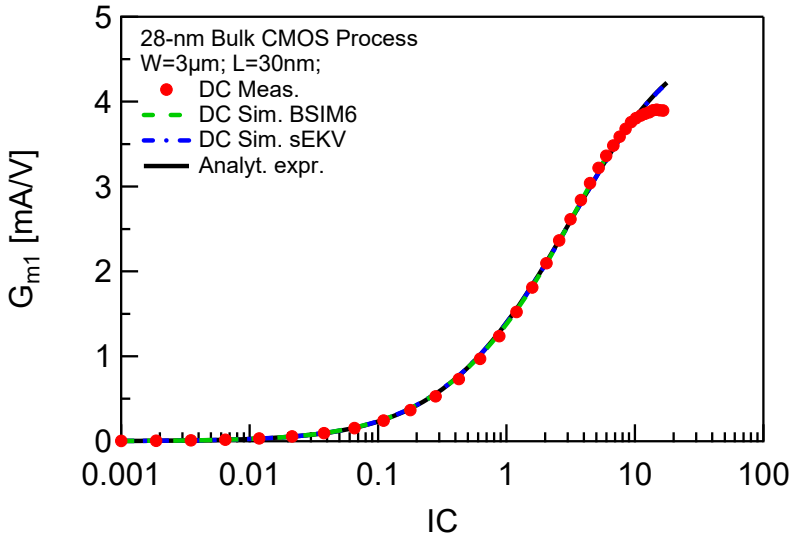
$$c = 1-3q_s,$$

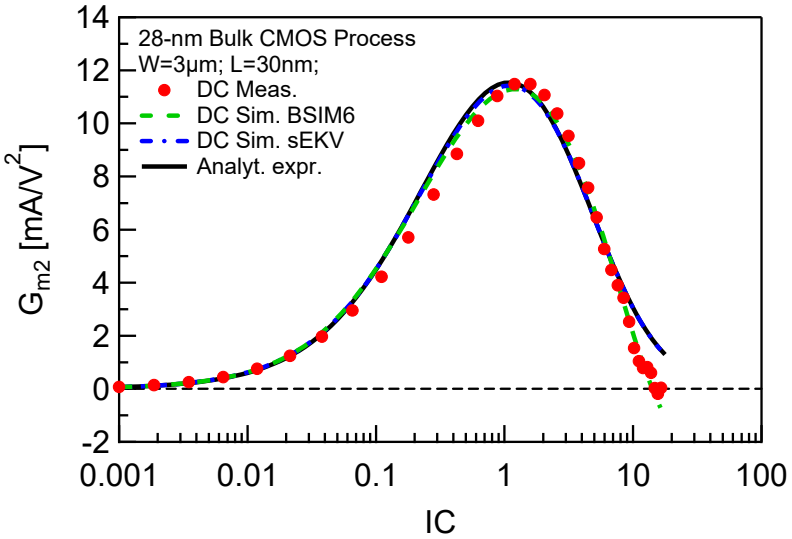
$$d = 1-4q_s$$

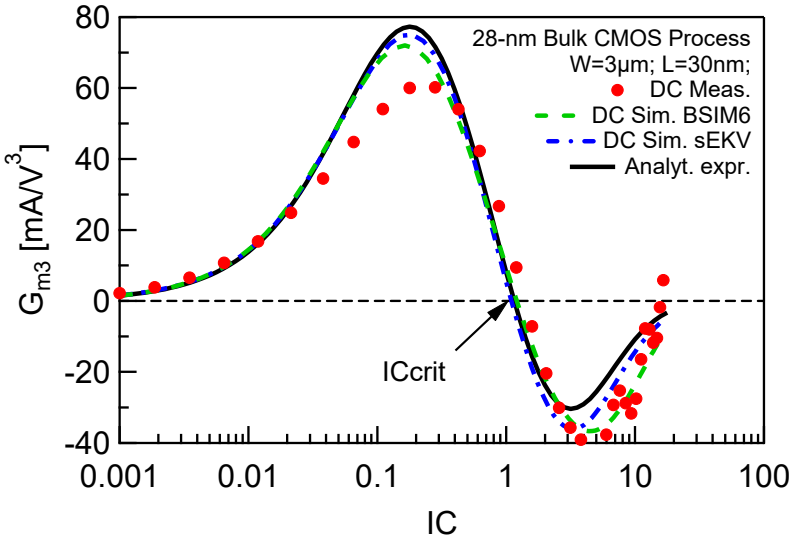
- As shown in the next slide, because of velocity saturation, $g_{m3} = 0$ for an inversion coefficient IC_{crit} that for short-channel devices lies in the middle of the moderate inversion
- Linearity is thus improved within a "sweet spot" around $IC = IC_{crit}$

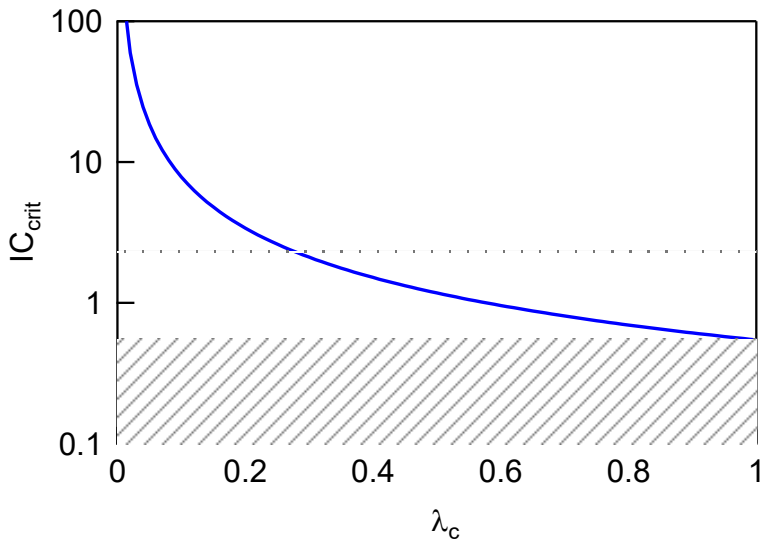
F. Chicco, A. Pezzotta and C. Enz, "Charge-Based Distortion Analysis of Nanoscale MOSFETs," TCAS1, Vol. 66 , No. 2, pp. 453-462, Feb. 2019.

Transconductances G_{mk} versus IC









📖 F. Chicco, A. Pezzotta and C. Enz, "Charge-Based Distortion Analysis of Nanoscale MOSFETs," TCAS1, Vol. 66 , No. 2, pp. 453-462, Feb. 2019.

Single-tone Analysis

- In single-tone analysis, the gate voltage variation is a sinewave $\Delta V_G = A \cdot cos(\omega t)$ and the output current is given by

$$I_D(t) \cong I_{D(0)} + I_{D(1)} \cdot \cos(\omega t) + I_{D(2)} \cdot \cos(2\omega t) + I_{D(3)} \cdot \cos(3\omega t)$$

- where $I_{D(k)}$ are the harmonics amplitude given by

$$I_{D(0)} \cong I_{D0} + \frac{G_{m2} \cdot A^2}{4},$$

$$I_{D(1)} \cong G_{m1} \cdot A + \frac{G_{m3} \cdot A^3}{8},$$

$$I_{D(2)} \cong \frac{G_{m2} \cdot A^2}{4},$$

$$I_{D(3)} \cong \frac{G_{m3} \cdot A^3}{24}$$

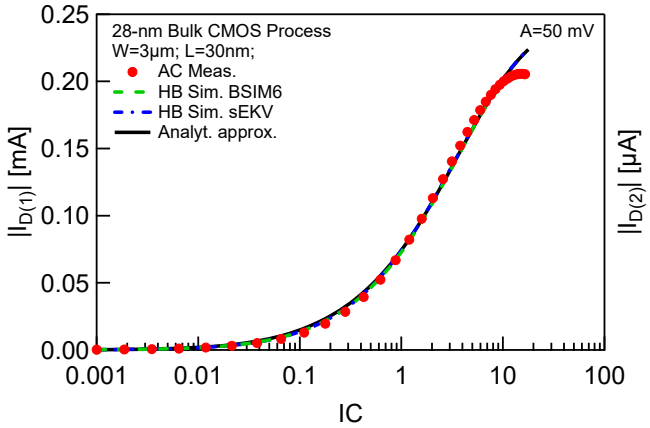
$$HD_2 \triangleq \left|\frac{I_{D(2)}}{I_{D(1)}}\right| = \frac{2G_{m2} \cdot A}{8G_{m1} + G_{m3} \cdot A^2},$$

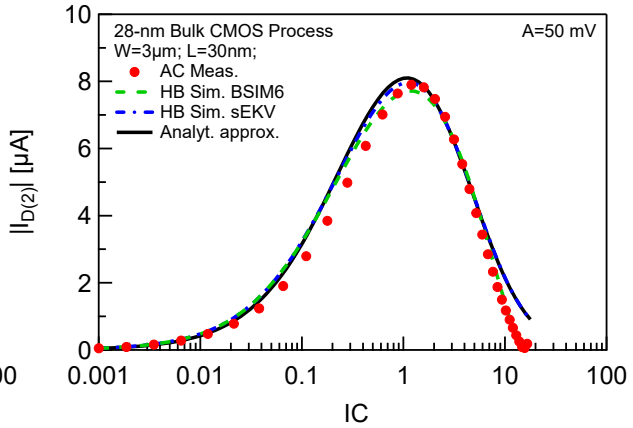
$$HD_3 \triangleq \left|\frac{I_{D(3)}}{I_{D(1)}}\right| = \frac{G_{m3} \cdot A^2}{3\left(8G_{m1} + G_{m3} \cdot A^2\right)}$$

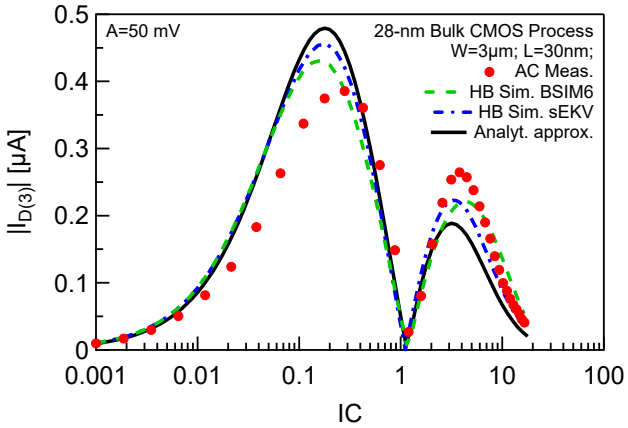
- Note that $HD_3 = 0$ for $IC = IC_{crit}$ due to $G_{m3} = 0$ as shown in the previous slides

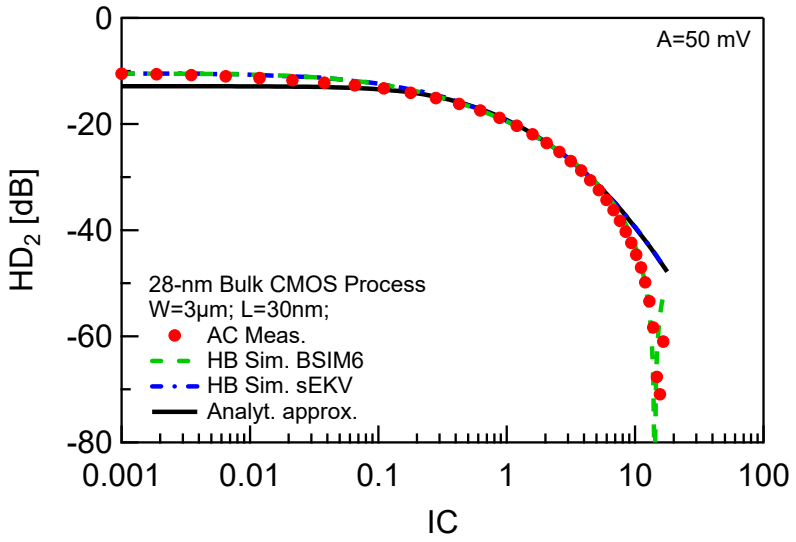
F. Chicco, A. Pezzotta and C. Enz, "Charge-Based Distortion Analysis of Nanoscale MOSFETs," TCAS1, Vol. 66 , No. 2, pp. 453-462, Feb. 2019.

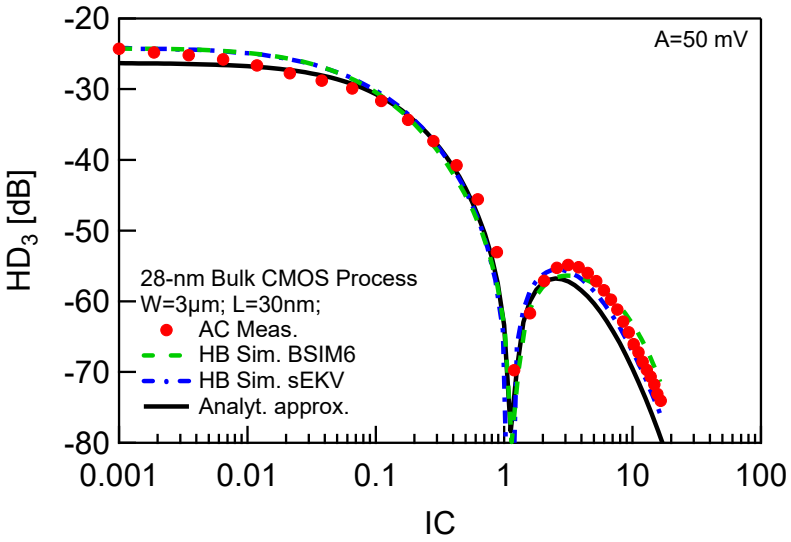
Single-tone Harmonics Model versus Measurements











F. Chicco, A. Pezzotta and C. Enz, "Charge-Based Distortion Analysis of Nanoscale MOSFETs," TCAS1, Vol. 66 , No. 2, pp. 453-462, Feb. 2019.

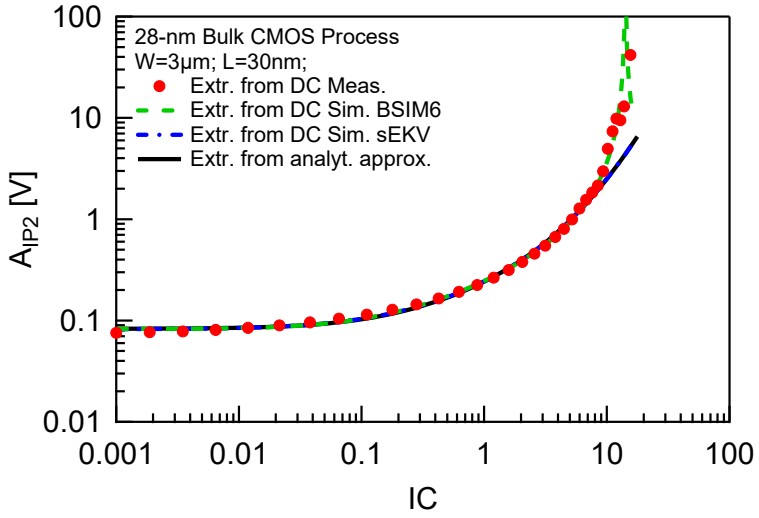
Two-tone Model

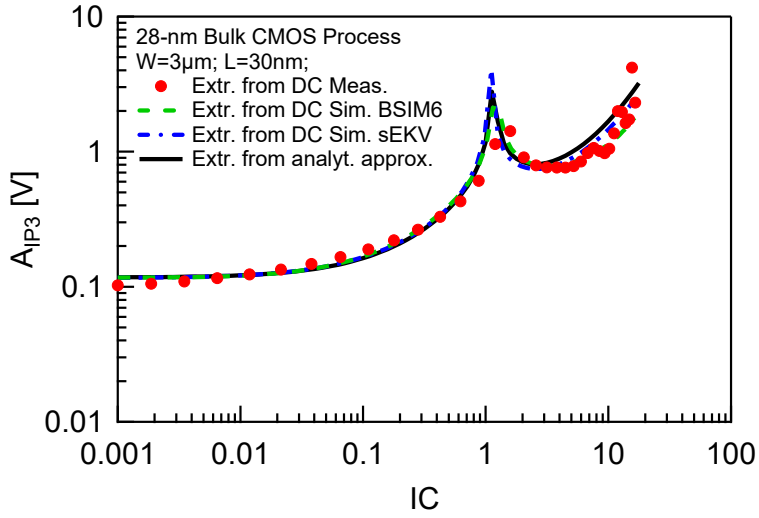
- In two-tone analysis, the gate voltage variation is made of two sinewaves $\Delta V_G = A_1 \cdot cos(\omega_1 t) + A_2 \cdot cos(\omega_2 t)$ the harmonics and intermodulation products are then given by

Term	Frequency	Harmonics or IM Amplitude
DC	$\omega = 0$	$I_{D(0)} = I_{D0} + \frac{G_{m2}}{4} \cdot (A_1 + A_2)$
1st-harmonic	$\omega = \omega_1$	$I_{D(1,1)} = G_{m1} \cdot A_1 + \frac{G_{m3}}{4} \cdot \left(A_1 A_2^2 + \frac{A_1^3}{2} \right)$
	$\omega = \omega_2$	$I_{D(1,2)} = G_{m1} \cdot A_2 + \frac{G_{m3}}{4} \cdot \left(A_2 A_1^2 + \frac{A_2^3}{2} \right)$
2nd-harmonic	$\omega = 2\omega_1$	$I_{D(2,1)} = \frac{G_{m2} \cdot A_1^2}{4}$
	$\omega = 2\omega_2$	$I_{D(2,2)} = \frac{G_{m2} \cdot A_2^2}{4}$
3rd-harmonic	$\omega = 3\omega_1$	$I_{D(3,1)} = \frac{G_{m3} \cdot A_1^3}{24}$
	$\omega = 3\omega_2$	$I_{D(3,2)} = \frac{G_{m3} \cdot A_2^3}{24}$
IM2	$\omega = \omega_1 \pm \omega_2$	$I_{D(\mathrm{IM2})} = \frac{G_{m2} \cdot A_1 \cdot A_2}{2}$
IM3	$\omega = 2\omega_1 \pm \omega_2$	$I_{D(\mathrm{IM3},1)} = \frac{G_{m3} \cdot A_1^2 \cdot A_2}{8}$
	$\omega = \omega_1 \pm 2\omega_2$	$I_{D(\mathrm{IM3},2)} = \frac{G_{m3} \cdot A_1 \cdot A_2^2}{8}$

F. Chicco, A. Pezzotta and C. Enz, “Charge-Based Distortion Analysis of Nanoscale MOSFETs,” TCAS1, Vol. 66 , No. 2, pp. 453-462, Feb. 2019.

IP2 and IP3





- The 2nd and 3rd-order input-referred intercept points are given by

$$A_{IP2} = 2\left|\frac{G_{m1}}{G_{m2}}\right| \quad \text{and} \quad A_{IP3} = \sqrt{8\left|\frac{G_{m1}}{G_{m3}}\right|}$$

- Again A_{IP3} becomes infinite in the "sweet spot" around $IC = IC_{crit}$ which lies in the middle of the moderate inversion ($IC = 1$ in the particular case shown here)

F. Chicco, A. Pezzotta and C. Enz, "Charge-Based Distortion Analysis of Nanoscale MOSFETs," TCAS1, Vol. 66 , No. 2, pp. 453-462, Feb. 2019.

References

Most of this Chapter is based on Chapter 11 to 13 of Reference [1]

[1] C. Enz and E. A. Vittoz, *Charge-Based MOS Transistor Modeling - The EKV Model for Low-Power and RF IC Design*, 1st ed: Wiley, 2006.

[2] G. Gonzalez, *Microwave Transistor Amplifiers – Analysis and Design*, 2nd ed. Prentice-Hall, 1996.

[3] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed. Cambridge University Press, 2004.

[4] G. D. Vendelin, A. M. Pavio and U. L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, 2nd ed. Wiley, 2006.

[5] B. Razavi, *RF Microelectronics*, 2nd ed. Pearson, 2012.