E3-238
Analog and Mixed Signal VLSI Circuit Design

Dr. Navakanta Bhat

Assistant Professor, ECE Department
Indian Institute of Science, Bangalore-560012

Email: navakant@ece.iisc.ernet.in
URL: http://ece.iisc.ernet.in/~navakant/Navakant_Bhat.html

Version 1, 26th July 2003

Scope of the Course-ware

"This course material has been developed to supplement the the discussions during the lectures in class. You can use this as the principal reference material. However, this course material is not a text book. You may still want to read up some of the books listed in the reference list to gain more insight or to get alternate explanation for a given topic."

Your feedback is wel-come in terms of any corrections or any additions to be done to the course-ware to improve its utility.

Note: The emphasis in this course is to designs analog circuits on a digital CMOS technology. Some of the discussions should be viewed and appreciated with this context in mind.

List of Reference books

Due to the advent of mixed signal SOCs, numerous books have been published on Analog Design. A partial list :

1. Analog CMOS Design
Razavi, McGraw Hill Publication

2. CMOS: Circuit Design, Layout , and Simulation
Boise, Baker, Lee, Prentice Hall Publication

3. Analog VLSI : Signal and Information Processing
Ismail and Feiz, McGraw Hill Publication

4. Analysis and Design of Analog Integrated Circuits
Gray and Meyer, Wiley Publication

Topics covered

- CMOS versus Bipolar Implementation
- The Sub-micron MOS Transistor for Analog Design
- Small signal parameters for MOSFET
  Cut-off frequency, Concept of Poles and Zeros, Miller approximation
- Single stage Amplifiers
  Common source, Common gate, Source follower, Cascode
- Current Mirror
  Cascode Current Mirror, Wilson Current Mirror, Regulated Cascode
- Layout issues
- Bandgap Voltage Reference

Topics covered

- Differential Amplifier
- Gilbert Cell
- Design of 2 stage CMOS OPAMP
  - Differential to Single ended conversion, DC and AC response
  - Frequency Compensation, Pole Splitting, Zero Cancellation
- OPAMP Performance Metrics
  - Slew rate, CMRR, Offset, Noise …
- Output Stage
- OTA and OPAMP Circuits
- Sample and Hold
- Switched Capacitor Circuits

Topics covered

- Comparator

- Sense Amplifier
  Voltage SA, Current SA, Latch type SA, Gain bandwidth analysis

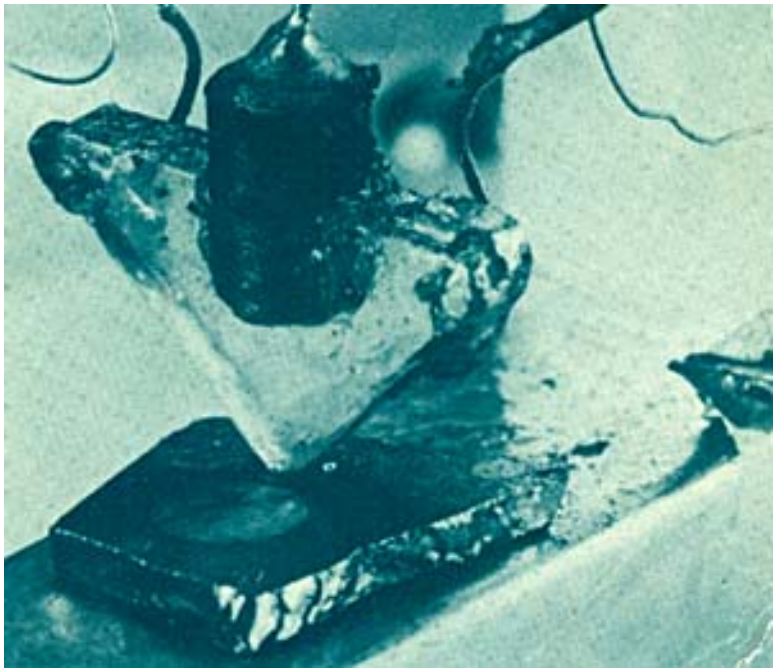
- Impact of mismatch on Analog design
  Offset effects in Sense Amplifier

- Statistical Design and Simulation

- Alternate Device Gallery for Low Voltage Low Power Analog Design
  Wide common range OPAMP
  Bulk driven OPAMP
  Lateral BJT in CMOS Technology
  Sub-threshold operation of MOSFET : Neural Network Application
  Floating Gate transistor as analog memory

The First Transistor : 1947

The baby is born!



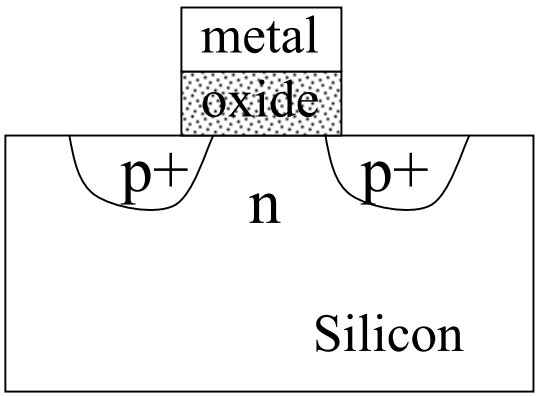
Bardeen, Brattain, Schokley @ Bell Labs

- First transistor was point contact Ge bipolar junction transistor, whereas the VLSI is neither based on Ge nor on BJT!
- The Bell Labs team was in fact trying to make MOS transistor, but got stuck with surface states and ended up with BJT!

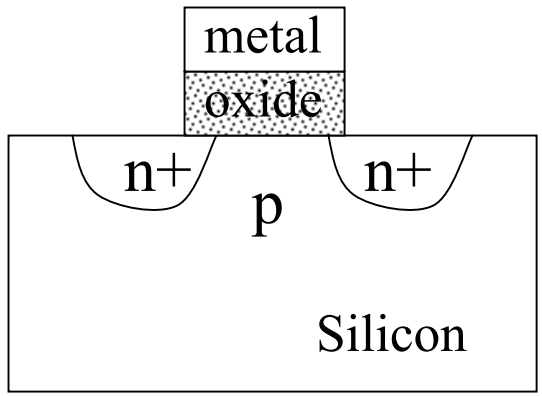
Metal Oxide Semiconductor Field Effect Transistor (MOSFET)

- Field effect transistor concept proposed in 1930s by Lilienfeld
- First MOSFET fabricated in 1960 by Kahng and Atalla

PMOSFET



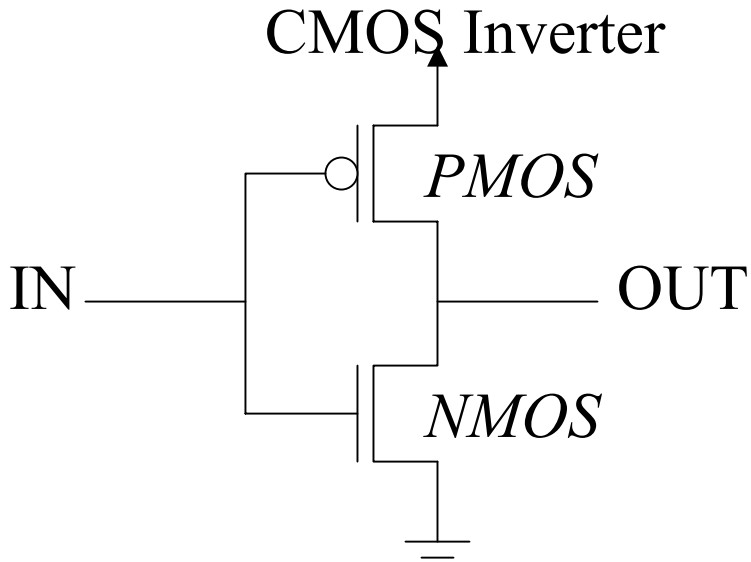
NMOSFET



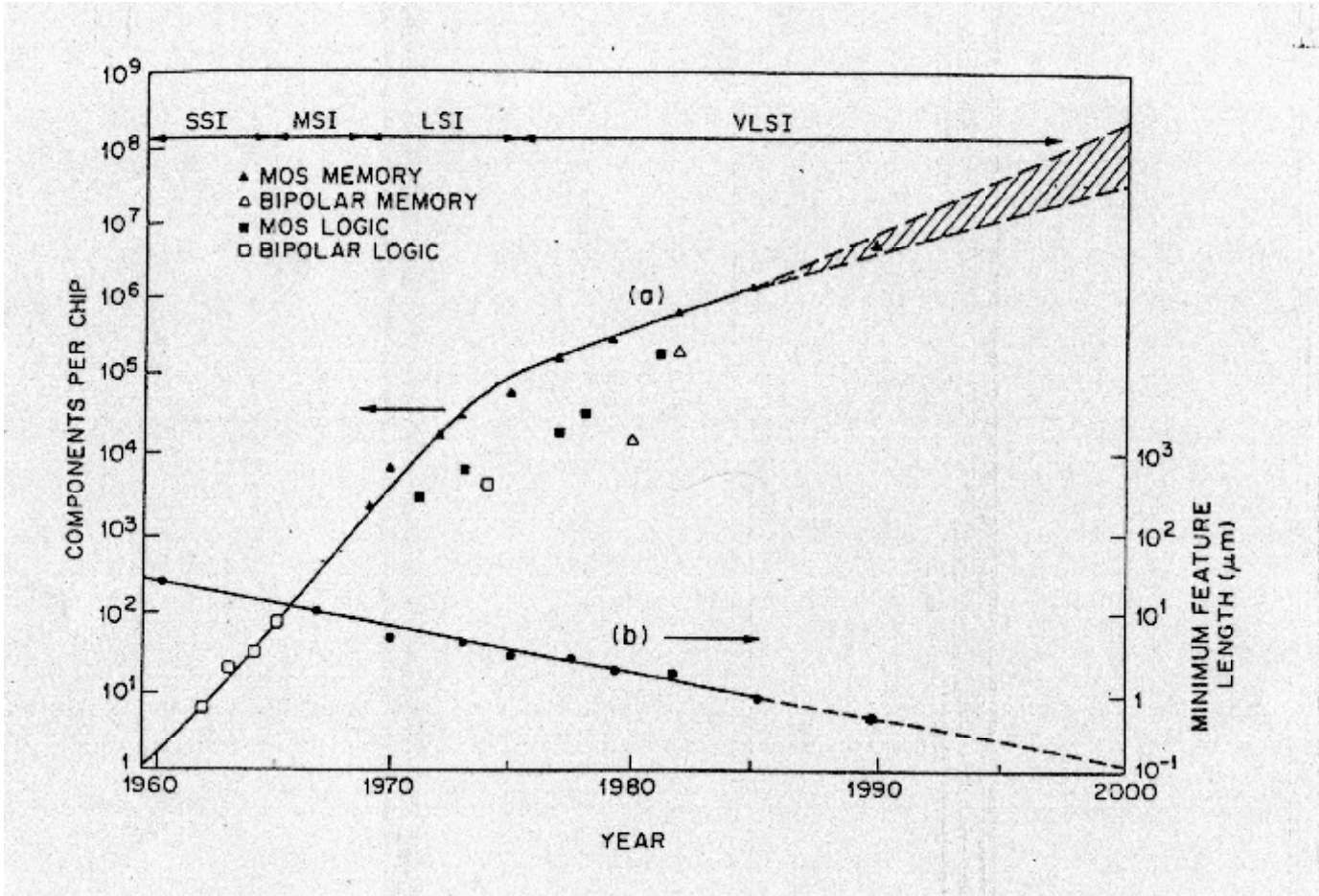
- Early MOS technology was based on PMOSFETs

CMOS (Complementary MOS) technology

- Both NMOSFETs & PMOSFETs are used
- No static power consumption
- Very high integration density
- Very good isolation
- Very low cost

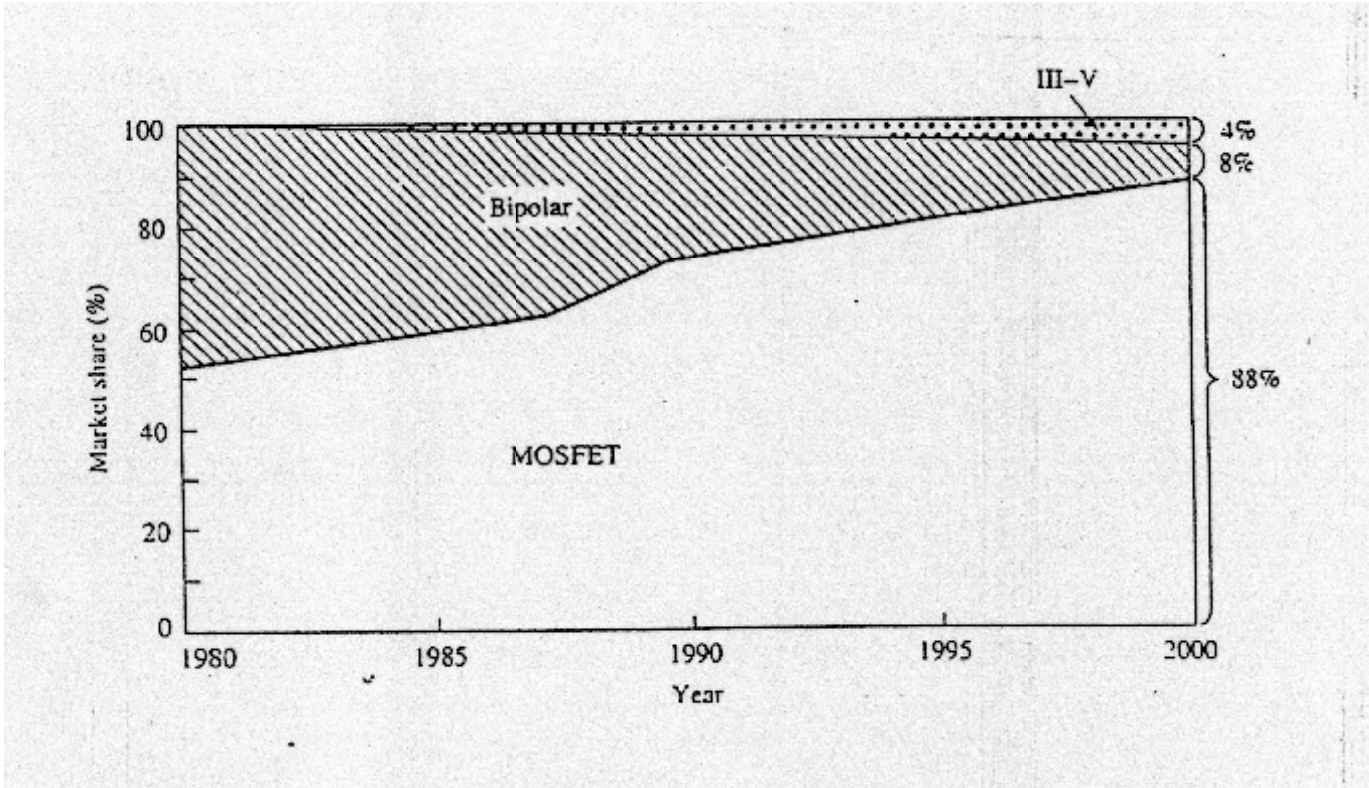


Evolution from SSI to VLSI



- In the the Digital world, MOSFET completely displaced BJT due to all the advantages offered by CMOS

CMOS market share



- CMOS captures more than 90% of electronics market share

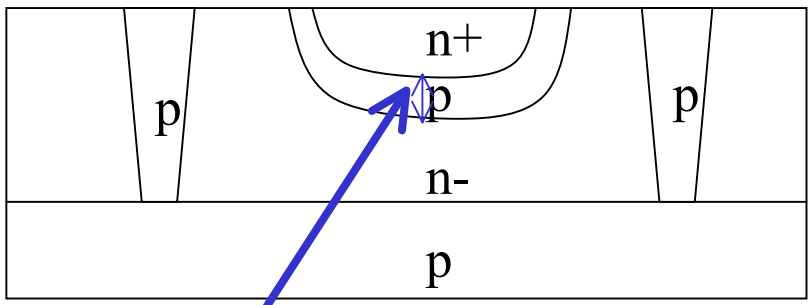
Process Technology Today (2003)

- 0.13μm digital technology in volume production
- Number of transistors per chip is ~ 1 billion(DRAMs), ~ 100 million (microprocessors)
- Technology scaling for future more challenging and expensive
- State of the art fab set-up costs more than US$2 billion
- Recovering the fab cost requires a modular process technology approach capable of producing diverse products

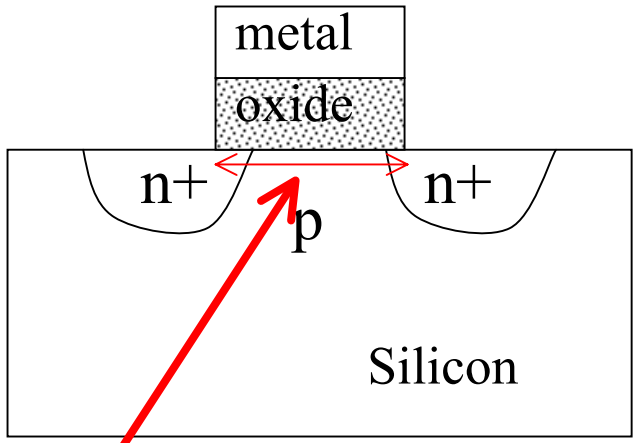
What do we do with the technology capable of making millions of transistor on a tiny area in Si? :

Mixed Signal Systems On Chip (SOC)

BJT versus MOSFET speed



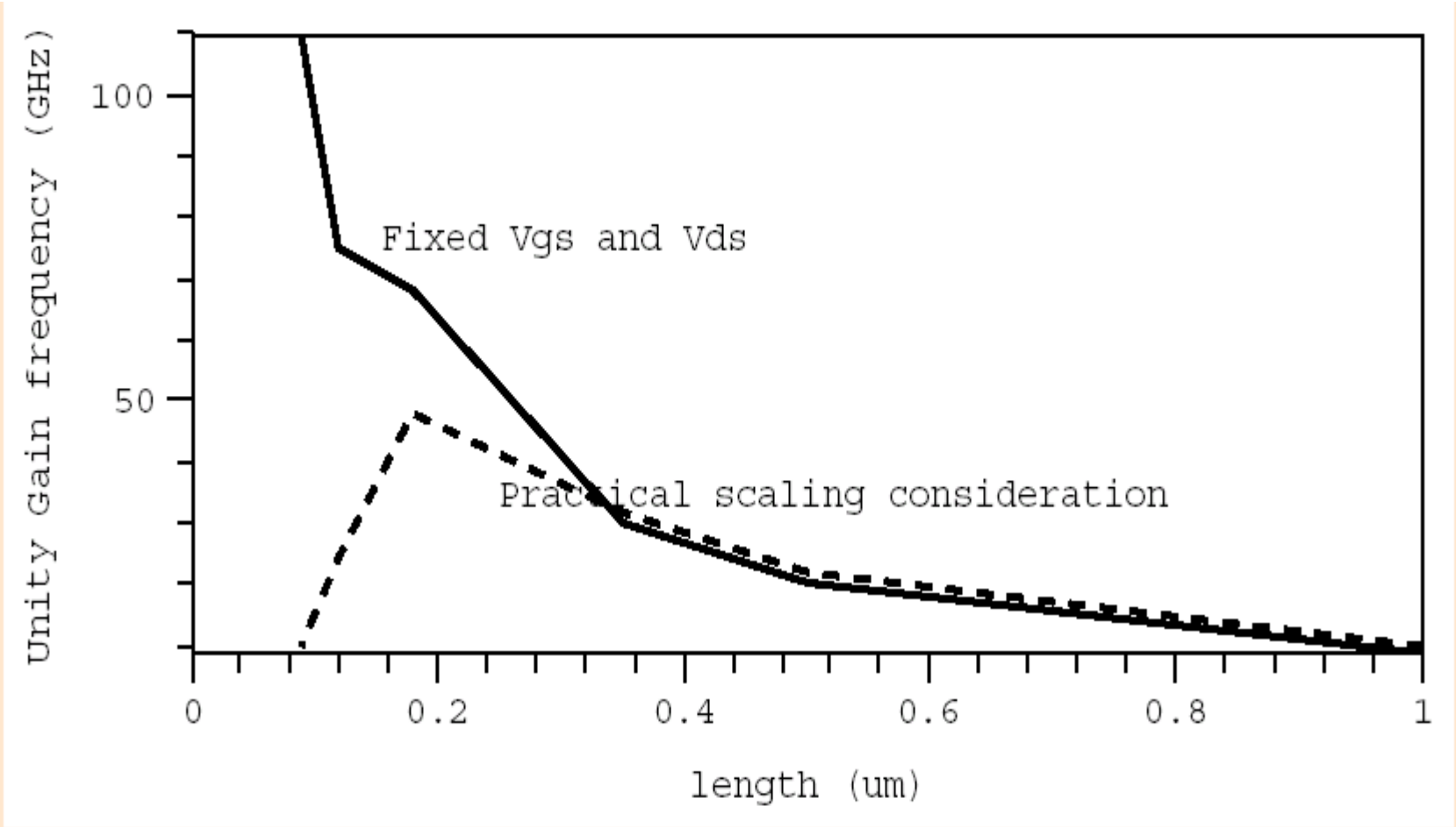
Base width defined by diffusion process



Channel length defined by Photolythography process

- Historically BJT used to be faster than MOSFET
- CMOS scaling has brought MOSFET on par with BJT

Cut-off frequency, f_T



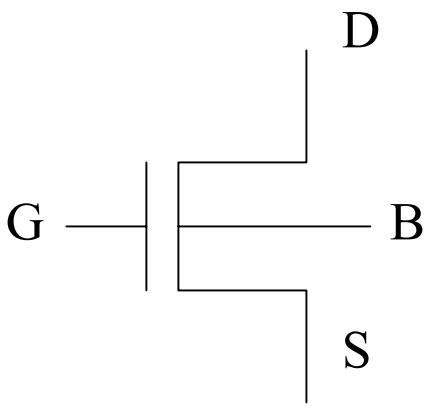
- MOSFET f_T has increased with scaling

Analog Design on Digital Technology

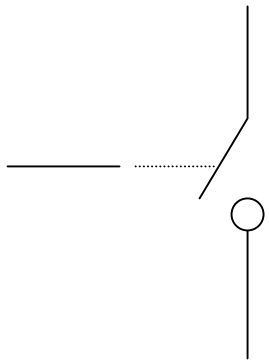
- Microprocessors are today's technology drivers
- The most elegant analog designs make use of the existing digital technology
- Every modification to the baseline technology adds on to the manufacturing cost
- Design For Manufacturability (DFM)
- CMOS analog circuits are logical choice

The Sub-micron MOS Transistor for Analog Design

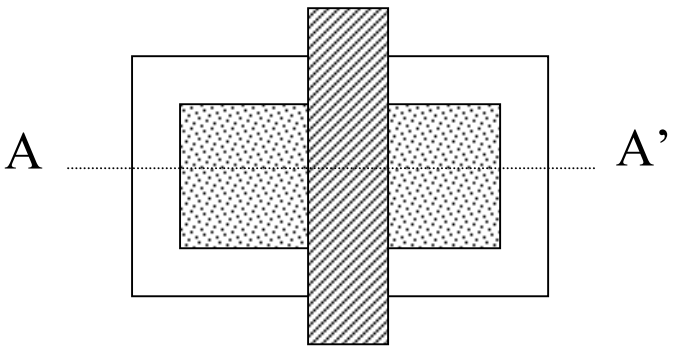
Transistor abstraction



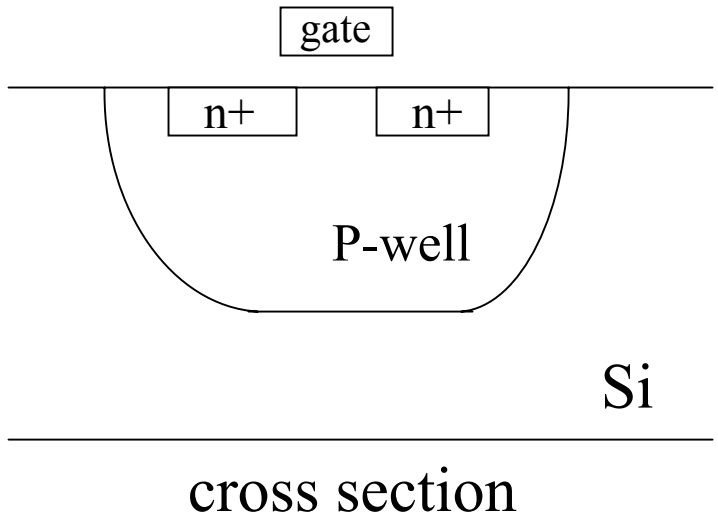
schematic



switch model

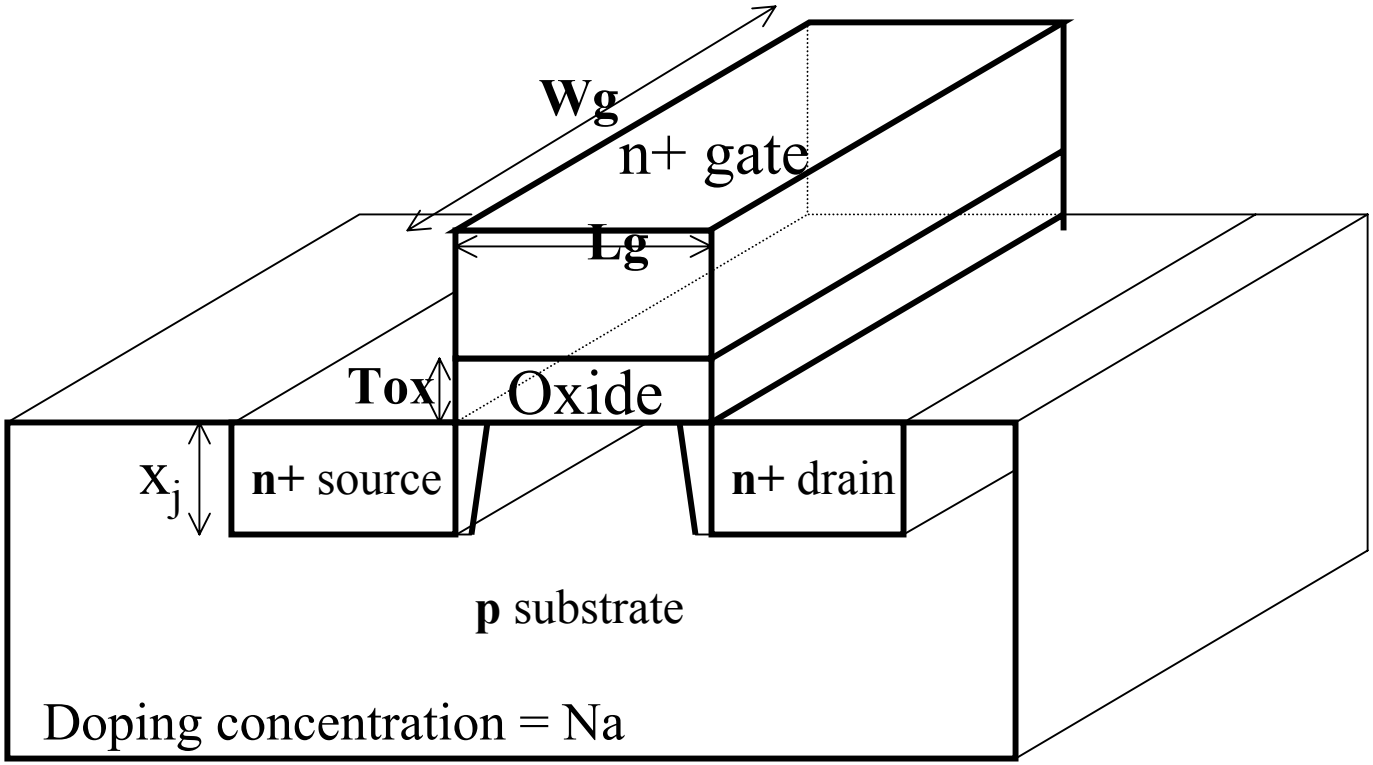


lay out



cross section

Simple 3-D picture of MOSFET



The 2 important dimensional parameters of MOSFET under circuit designer's control are:

Lg = Length of the gate
Wg = Width of the gate

Simple MOS Theory

Vgs < *Vt*, MOSFET is in cut off region

$$Ids = 0$$

Vgs > *Vt*, *Vds* < *Vgs*-*Vt*, MOSFET is in linear region

$$Ids = \frac{\mu \varepsilon_{ox} W}{T_{ox} L}\left[(Vgs - Vt)Vds - \frac{Vds^2}{2}\right]$$

Vgs > *Vt*, *Vds* > *Vgs*-*Vt*, MOSFET is in saturation region

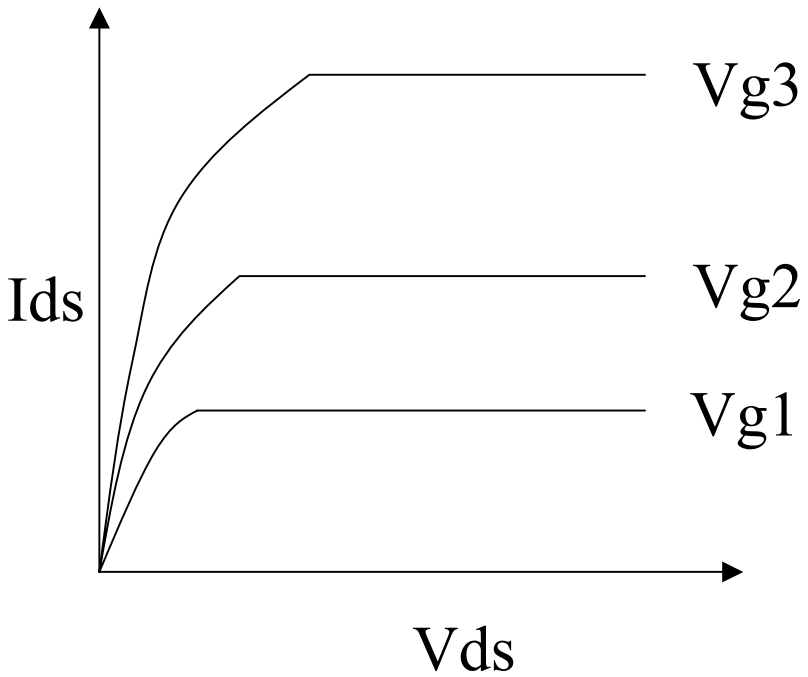
$$Ids = \frac{\mu \varepsilon_{ox} W}{T_{ox} L}\frac{(Vgs - Vt)^2}{2}$$

where μ is mobility, ε_{ox} is permittivity of the oxide, and *Vt* is the threshold voltage of the MOSFET

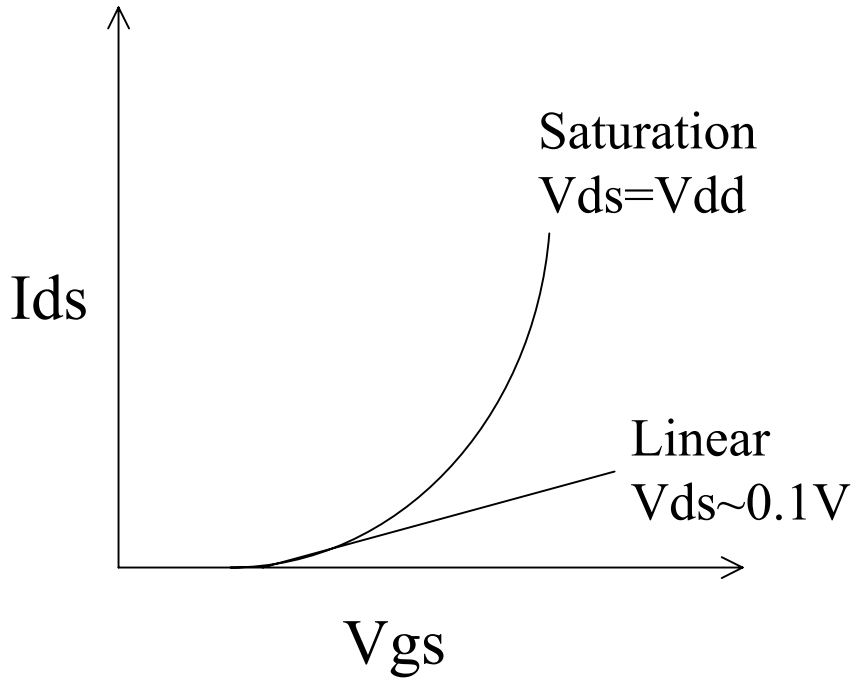
$$V_t = V_{fb} + 2\phi_b + \frac{T_{ox}\sqrt{4\varepsilon_s q N_a \phi_b}}{\varepsilon_{ox}}$$

I-V characteristics

Output Characteristics



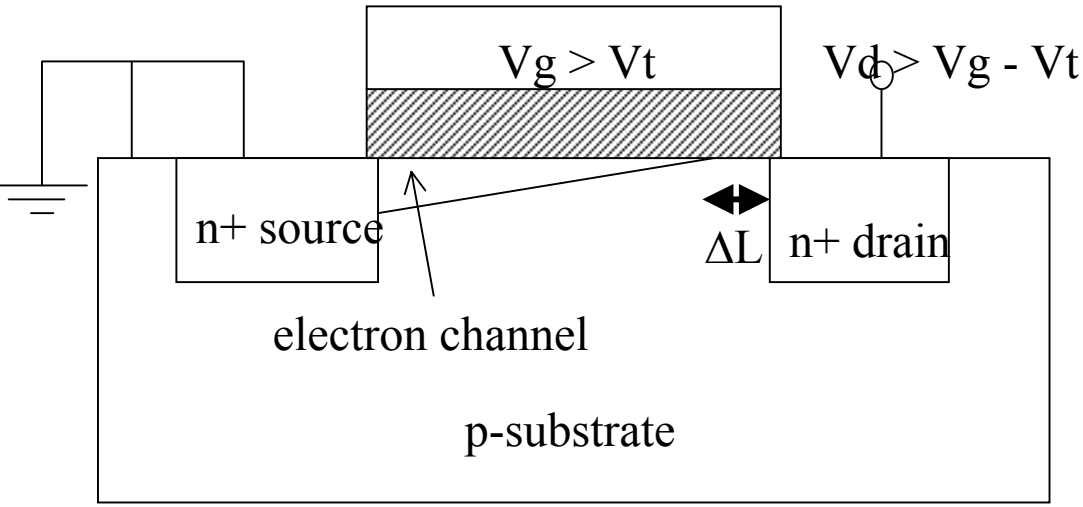
Transfer Characteristics

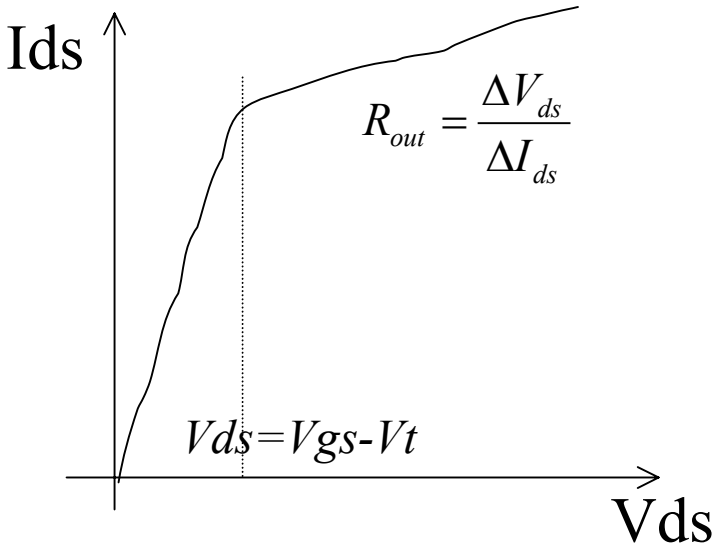


- Ids is constant and independent of Vds in saturation
- Ids is zero in sub-threshold region

Both of these idealities are incorrect especially for the sub-micron MOS transistor

Channel length modulation





Effective channel length is *Leff = L - ΔL, where ΔL=f(Vds)*

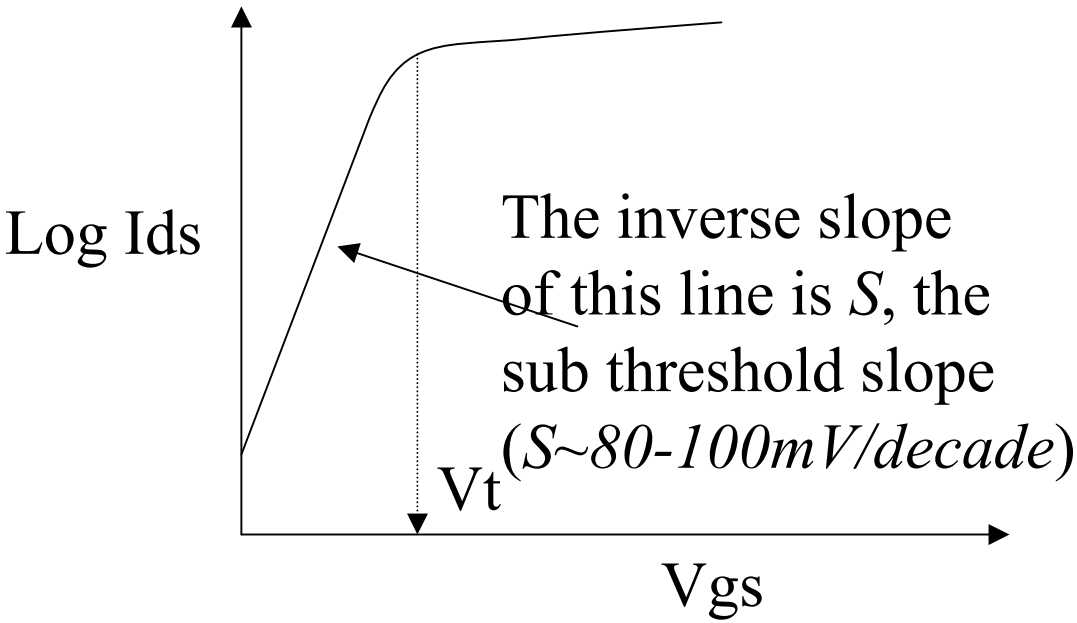
$$I_{ds} = \frac{\mu \varepsilon_{ox} W}{T_{ox} L_{eff}} \frac{(Vgs - Vt)^2}{2} \qquad I_{ds} = \frac{\mu \varepsilon_{ox} W}{T_{ox} L} \frac{(Vgs - Vt)^2}{2} (1 + \lambda V_{ds})$$

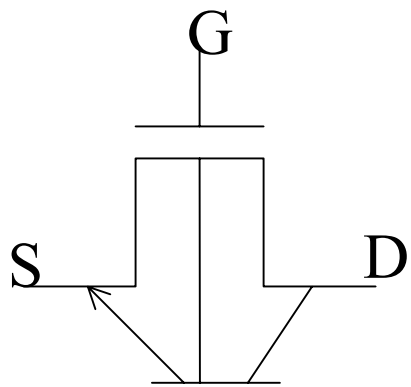
Ids increases slightly in saturation region with increasing Vds

This limits the AC output resistance for analog applications

λ is channel length modulation parameter in SPICE

Sub threshold conduction





- For Vg < Vt, current is non zero and is exponential function of Vg
- S = 2.3kT/q (1 + Csi/Cox) mV/decade
  Csi=depletion capacitance in Si, Cox=oxide capacitance,kT/q=thermal voltage
- MOSFET should be designed to have minimum possible S
- Sub threshold analog circuits work based on this principle

Sub threshold limitation on Vt scaling

Suppose S = 100mV/decade

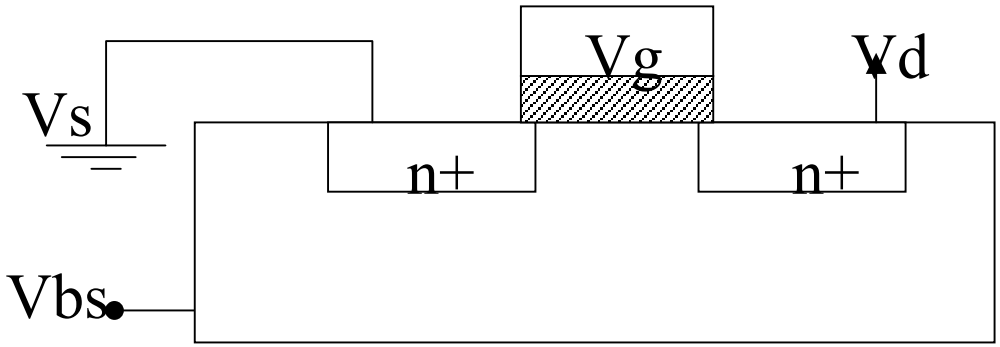
Suppose MOSFET should have Ion/Ioff = 10^6

Then for Vds=Vsupply,
when Vgs is changed from 0V (off state) to supply (on state)
Ids should change by ~ 6 decades

Vtmin = 100 * 6 = 0.6V

=> The value of S will impose the lower limit on Vt scaling

Body effect



$$V_{t0} = V_{fb} + 2\phi_b + \frac{T_{ox}\sqrt{4\varepsilon_s q N_a \phi_b}}{\varepsilon_{ox}}$$

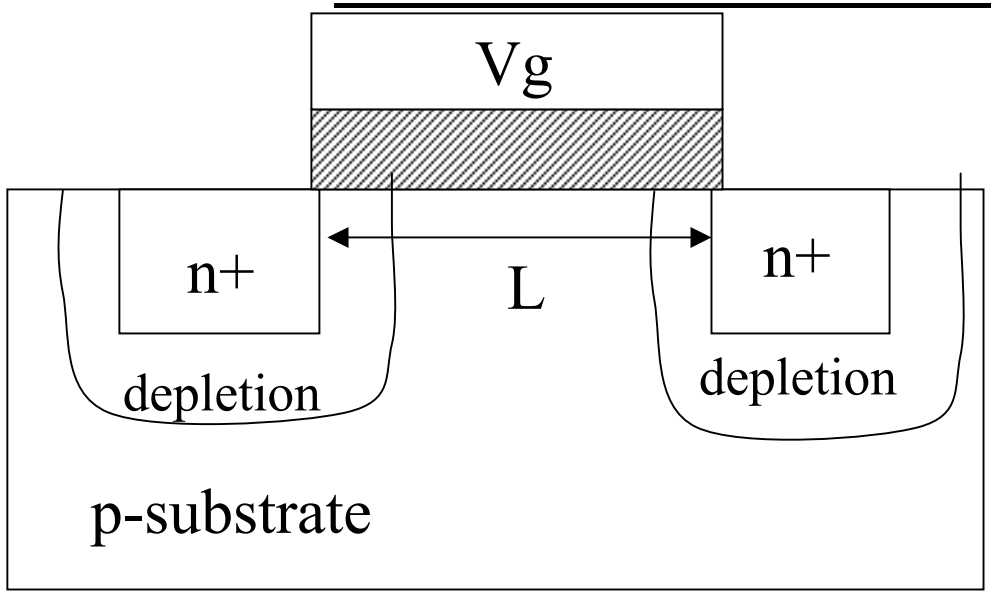
$$V_t = V_{t0} + \gamma\left(\sqrt{|V_{bs}| + 2\phi_b} - \sqrt{2\phi_b}\right)$$

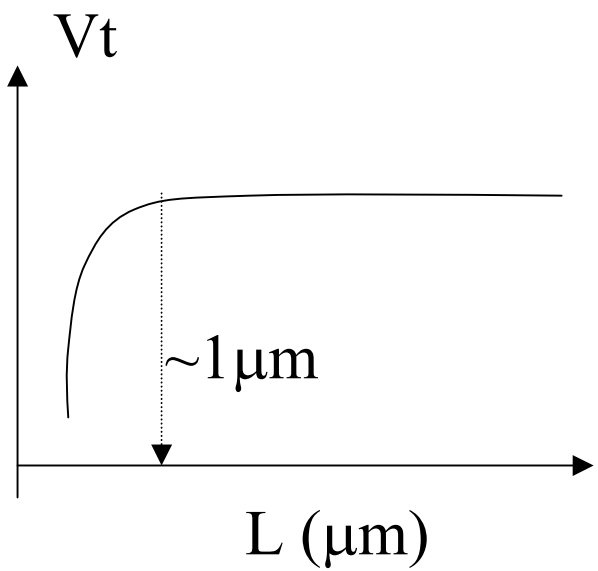
$$\gamma = \frac{T_{ox}\sqrt{2q\varepsilon_s N_a}}{\varepsilon_{ox}}$$

γ = body effect factor (γ = 0.3-0.7)

- Vt increases due to body effect
- This results in a transconductance term

Short Channel Effect (SCE)

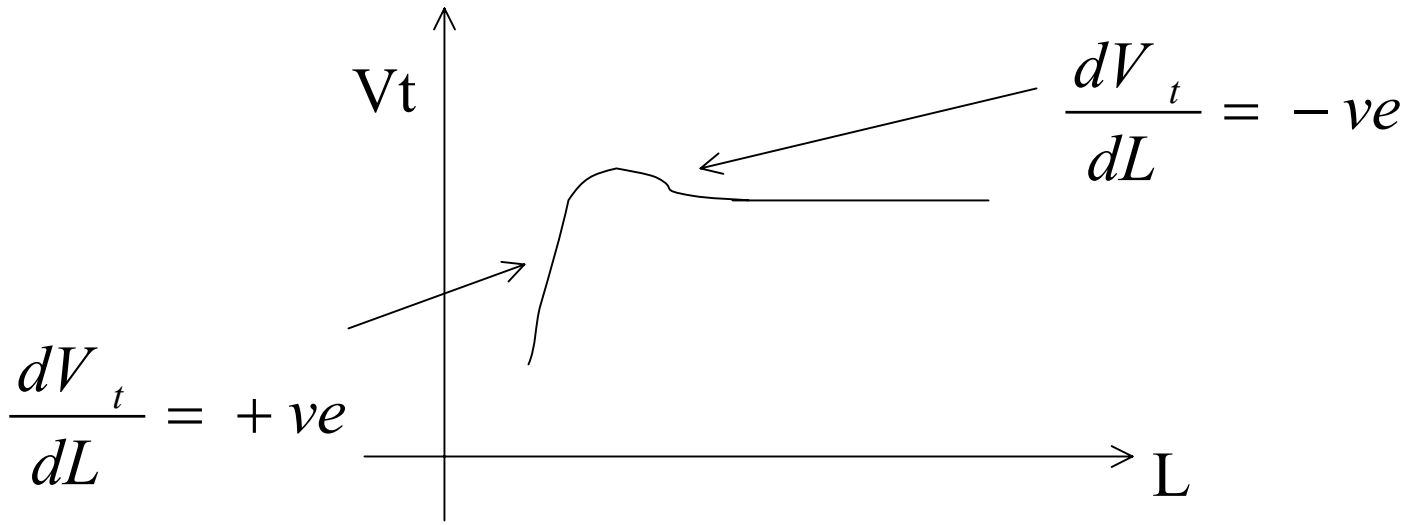




- Fraction of the depletion charge (Qd in Vt equation) is supported by the source and drain junctions and hence Vg need not support this
- When L is very small (~ 1μm) this charge becomes significant fraction of the total depletion charge and can not be neglected

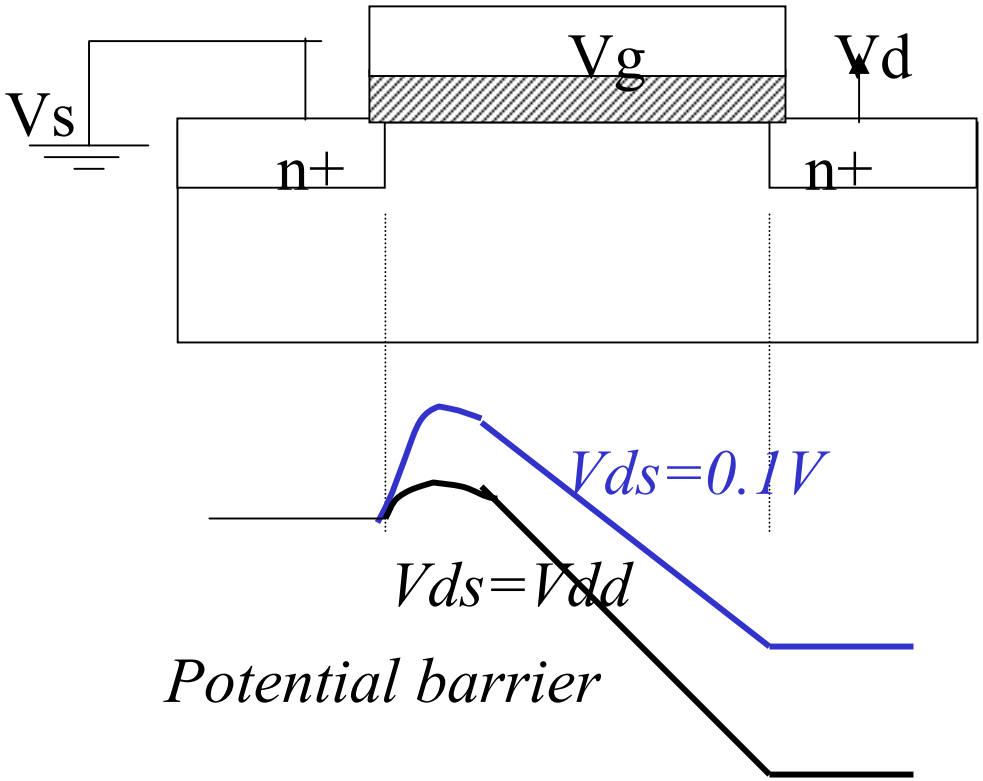
  => Vt decreases with decreasing L
- Impacts matching of transistors in analog applications

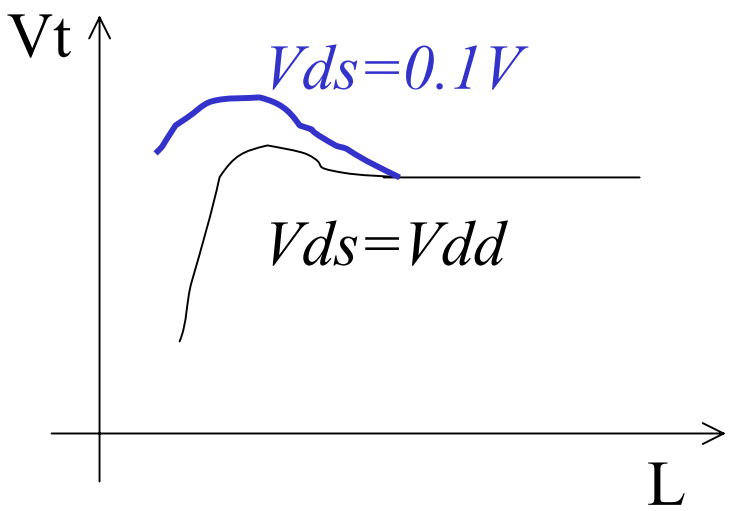
Reverse Short Channel Effect



- Invariably exists in almost all the sub-micron technologies
- The techniques used to suppress SCE are responsible for RSCE
- Vt becomes very sensitive function of L

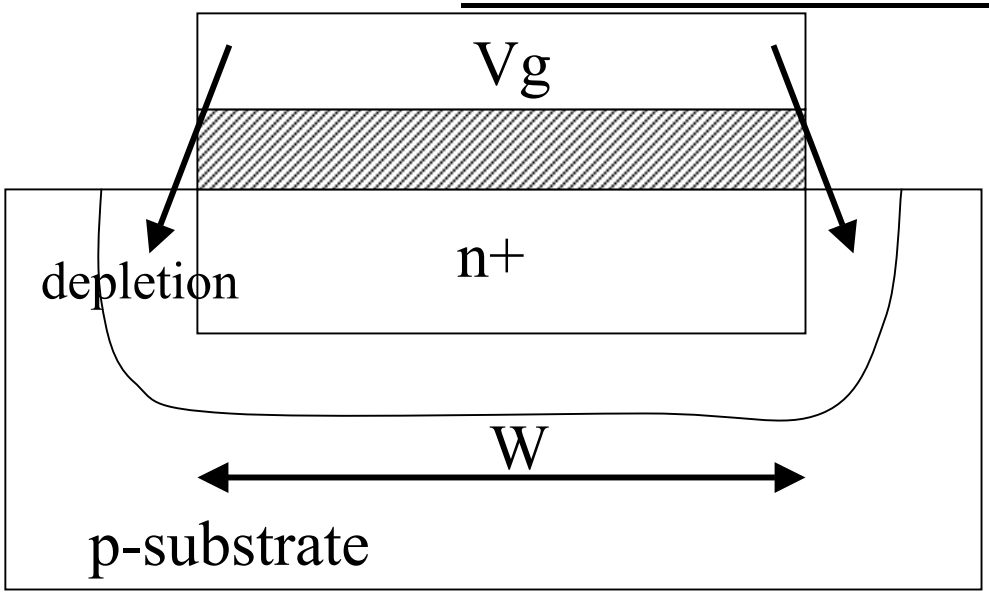
Drain Induced Barrier Lowering (DIBL)

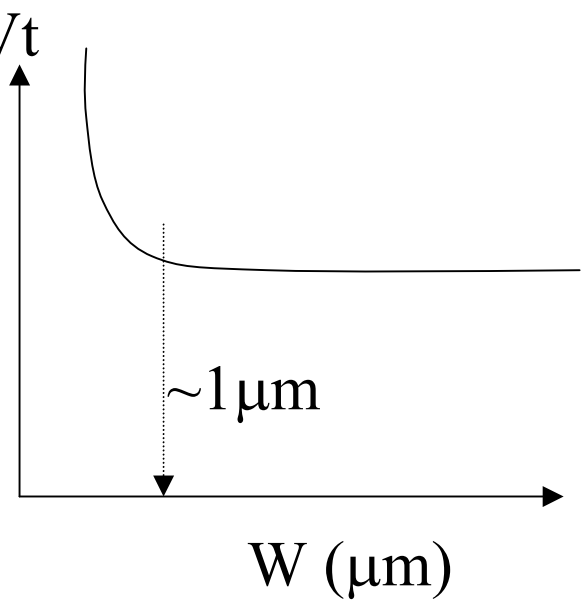




- Vt is also a function of drain voltage in sub-micron transistors
- DIBL effect is negligible in the long channel regime

Narrow Width Effect

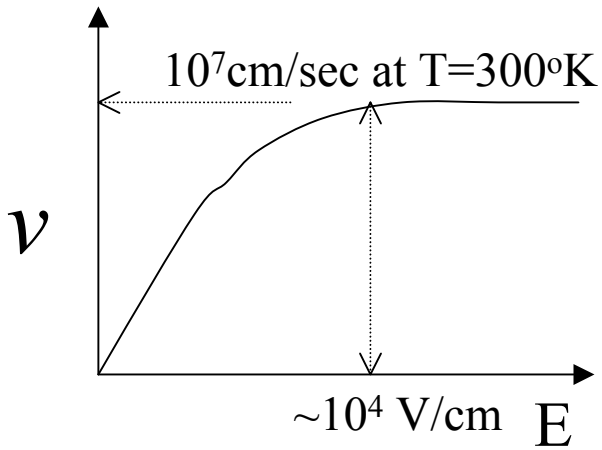


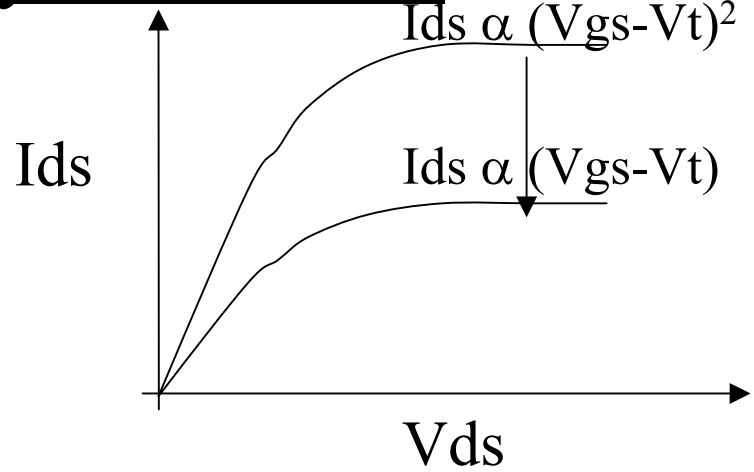


- Additional depletion charge at the edge of source & drain should be supported by the Vg before inverting the channel
- When W is very small (~ 1μm) this charge becomes significant fraction of the total depletion charge and can not be neglected

=> Vt increases with decreasing W

Velocity saturation





$v = \mu E$ *valid only at low electric fields (E)*

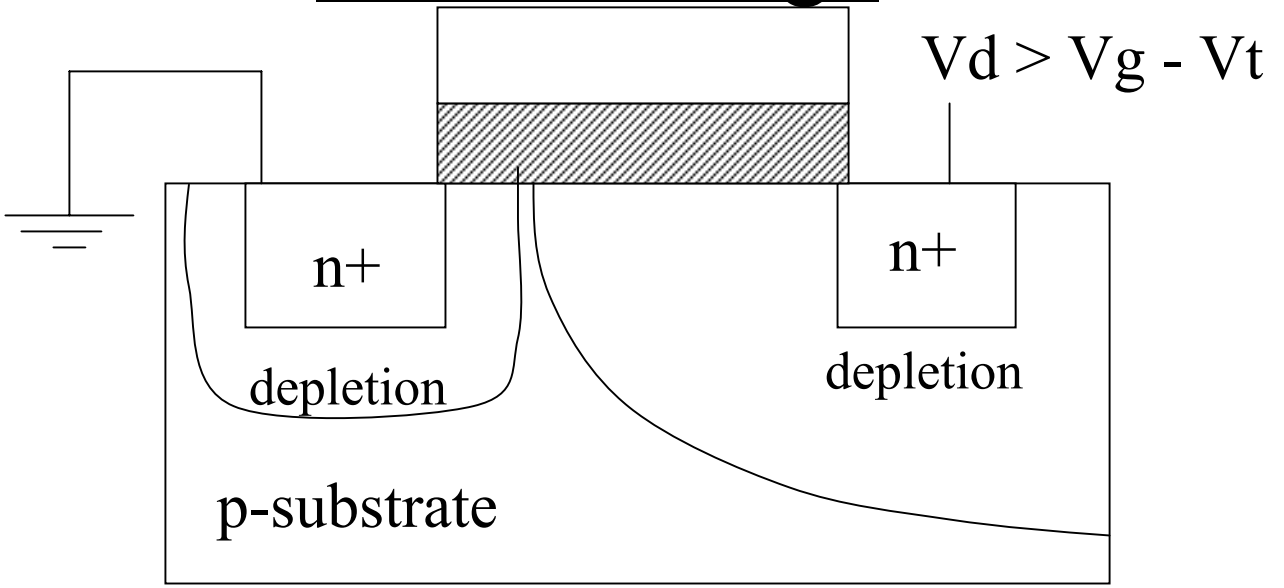
Ids will be less than expected due to velocity saturation

- For velocity saturated transistor, the saturation drive current is

$$Ids = \frac{\varepsilon W (V_{gs} - V_t) v_{sat}}{T_{ox}}$$

- Transconductance will be independent of L
- For L=0.1μm transistor operating at Vd=1V:
  E=10^5 V/cm => transistor is velocity saturated

Punch through



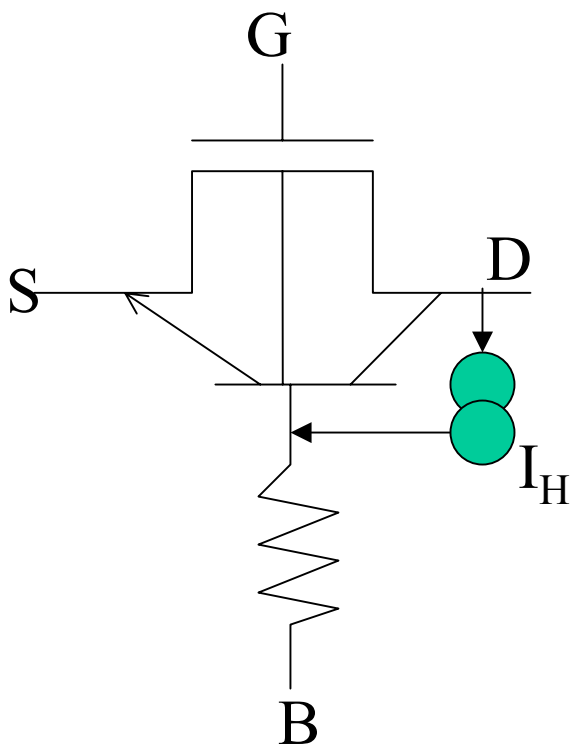
For short channel length, the drain depletion region merges with source depletion region for Vd > Vpt, punch through voltage

This results in large transistor current resulting in breakdown

Punch trough voltage, Vpt, is one of the limiting factors on Vdsmax

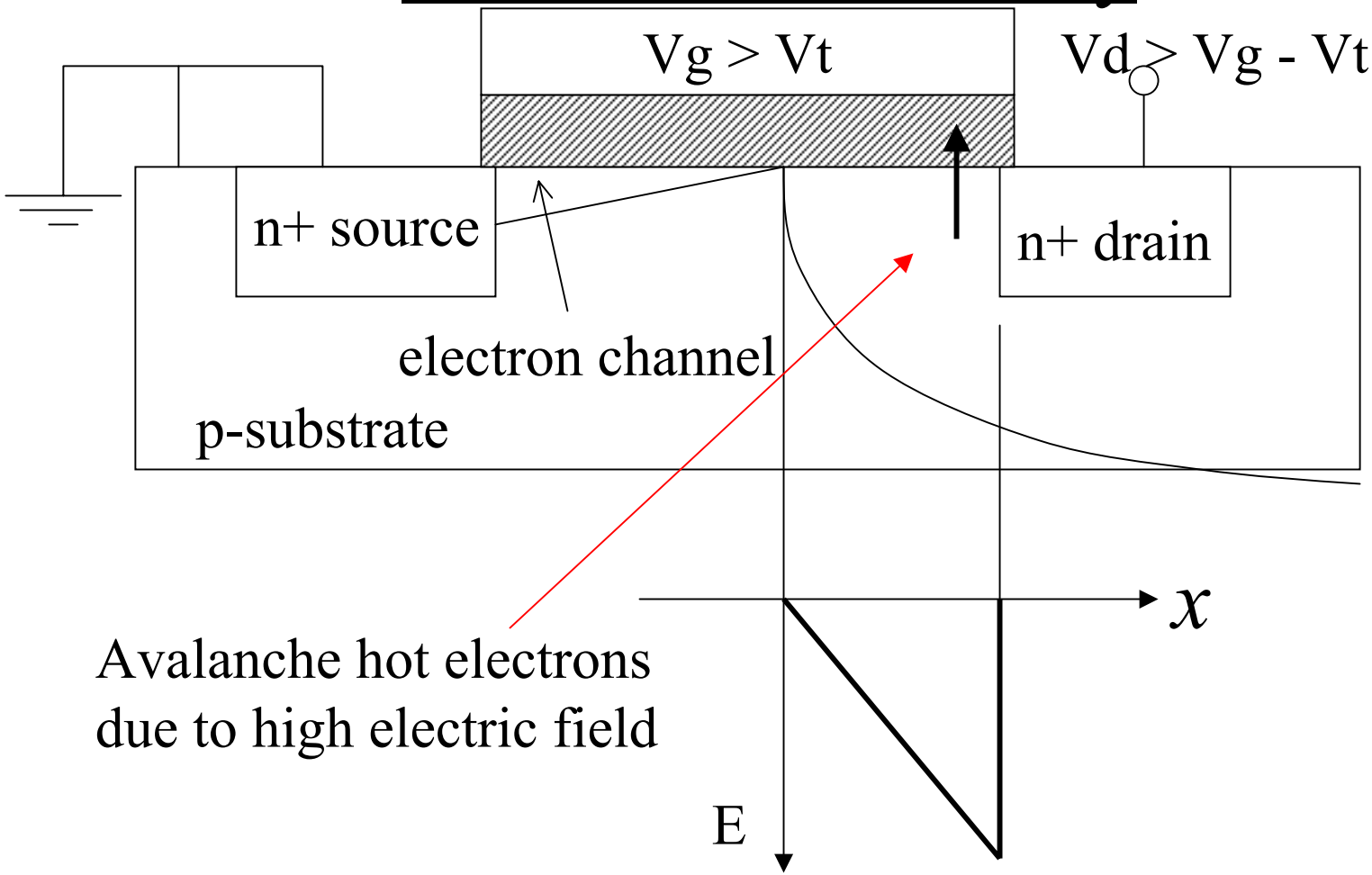
Bipolar induced breakdown

Parasitic Bipolar effect can trigger breakdown even before the punch through voltage is reached.



Avalanche effect at the drain generates e-h pairs.
The hole current going into the bulk can turn on BJT

Hot carrier reliability



Hot electron generation is maximum when Vgs~Vds/2

Transistor degradation due to hot carriers

Some of the high energy electrons are injected into the gate oxide by surmounting the barrier at Si-SiO_2 interface

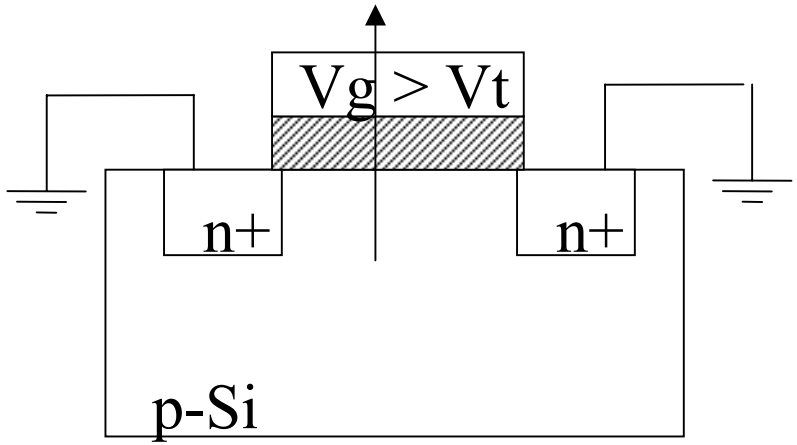
Injected electrons get trapped in the oxide

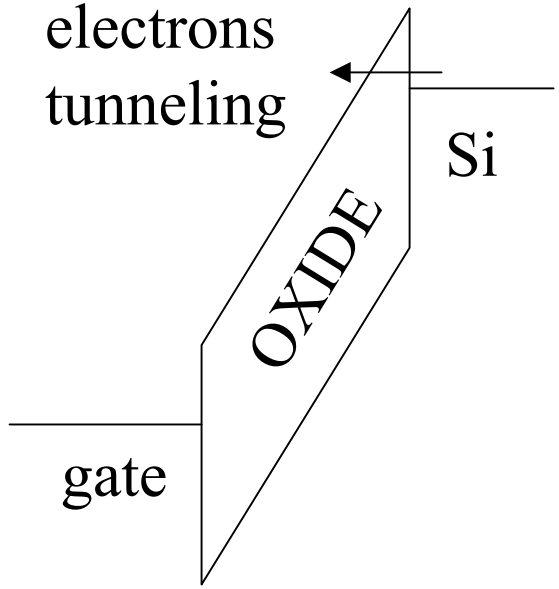
The trapped electrons increase Vt, decrease mobility and decrease drain current

Transistor degradation in turn reflects in circuit behavior resulting in decreased speed and functional failure under extreme conditions

Hot carrier degradation can be significant in analog transistor biased such that Vgs~Vds/2

Gate oxide reliability





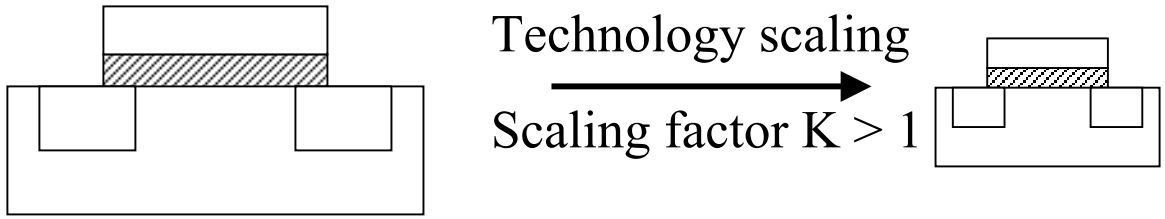
Eox = Vox/Tox

Under high electric fields electrons tunnel through the oxide

Tunneling electrons create damage in the oxide and hence affect the transistor performance

Gate oxide reliability worsens with decreasing oxide thickness

Constant field scaling



Primary scaling factors:

Tox, L, W, Xj (all linear dimensions)	1/K
Na, Nd (doping concentration)	K
Vdd (supply voltage)	1/K

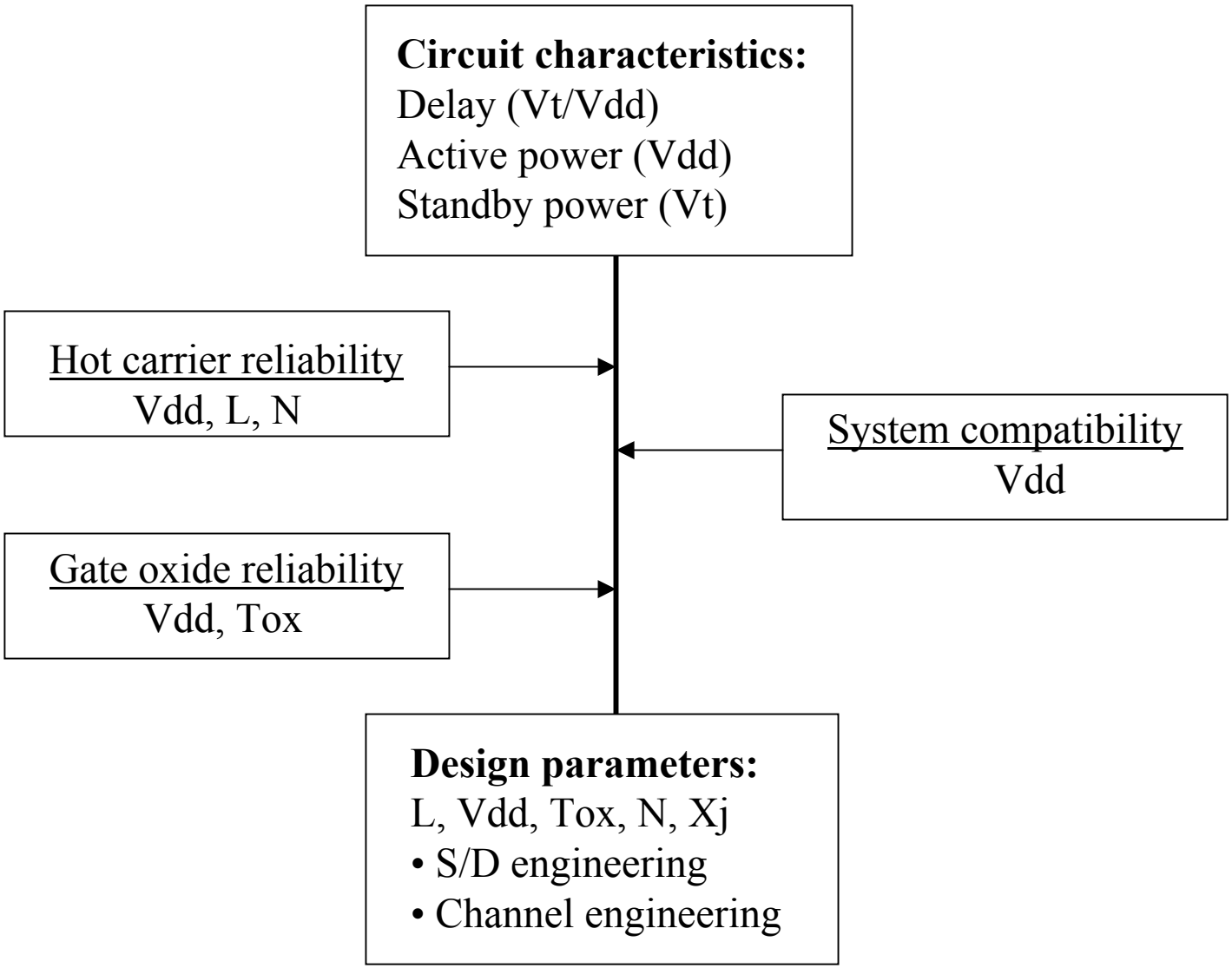
Derived scaling behavior of transistor:

Electric field	1
Ids	1/K
Capacitance	1/K

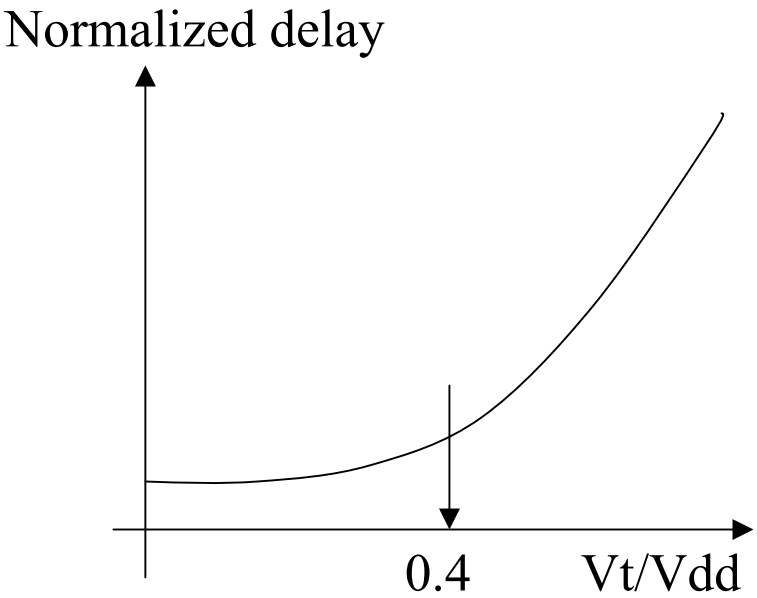
Derived scaling behavior of circuit:

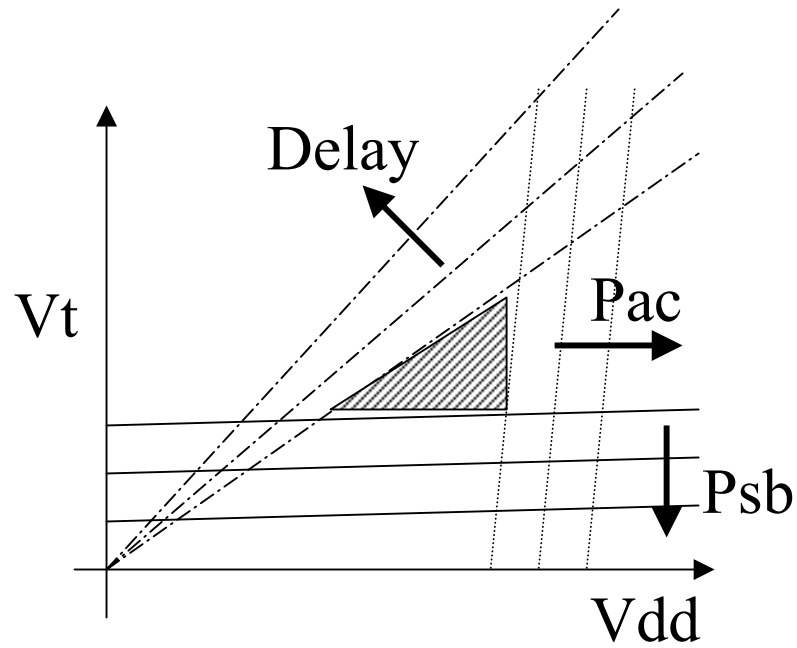
Delay (CV/I)	1/K
Power (VI)	$1/K^2$
Power-delay product	$1/K^3$
Circuit density (α 1/A)	K^2

Transistor design methodology for Digital Technology



Vt-Vdd design plane





Delay increases significantly for Vt/Vdd > 0.4

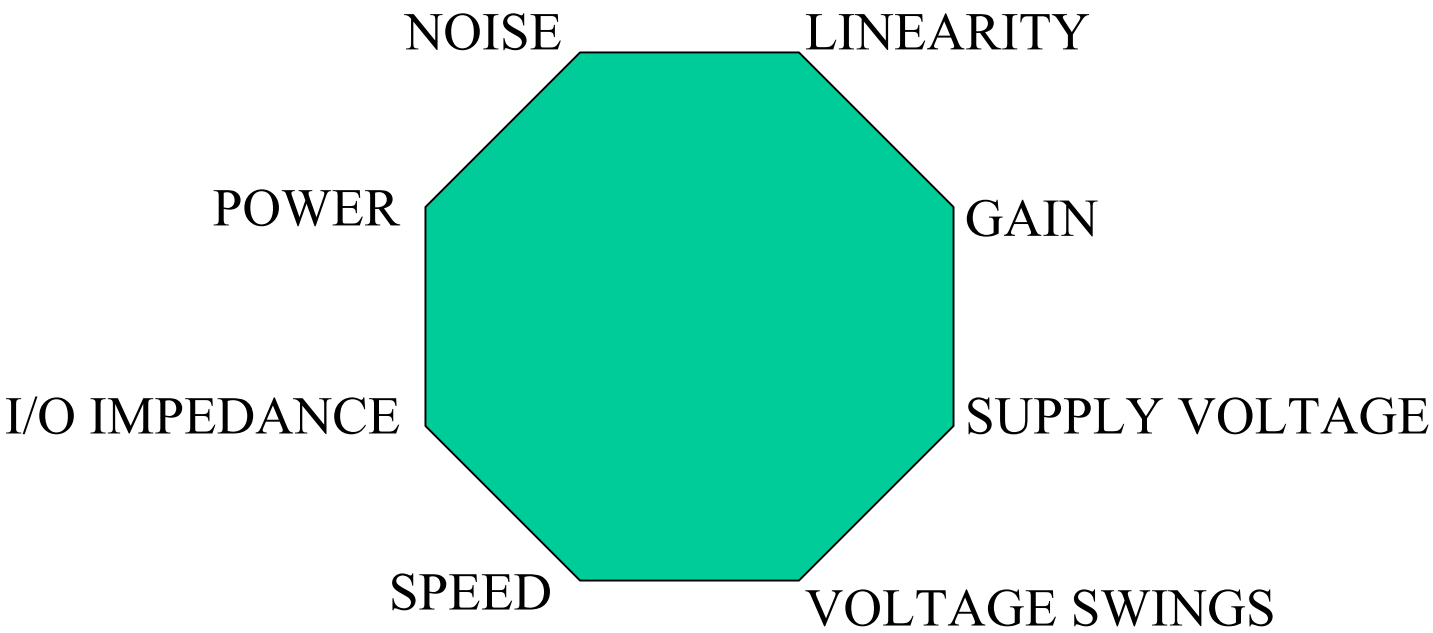
Pactive (Pac) = CV_{dd}^2f

Pstandby (Psb) = $WV_{dd}I_{off}$

Delay and Power are the only trade-off points for digital design

Analog Circuit Performance Metrics

The Analog Octagon:



B. Razavi

Multiple trade-offs involved in Analog Design make it very complex

Small Signal parameters for MOSFET

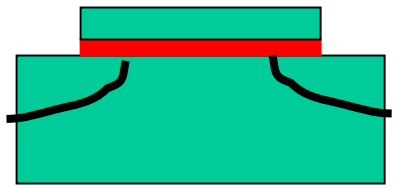
AC Small Signal Parameters

Small signal parameters are derived from DC equations

The value of the small signal parameters is a function of the DC bias point

MOSFET is assumed to be square law device under saturation, i.e. no velocity saturation effect

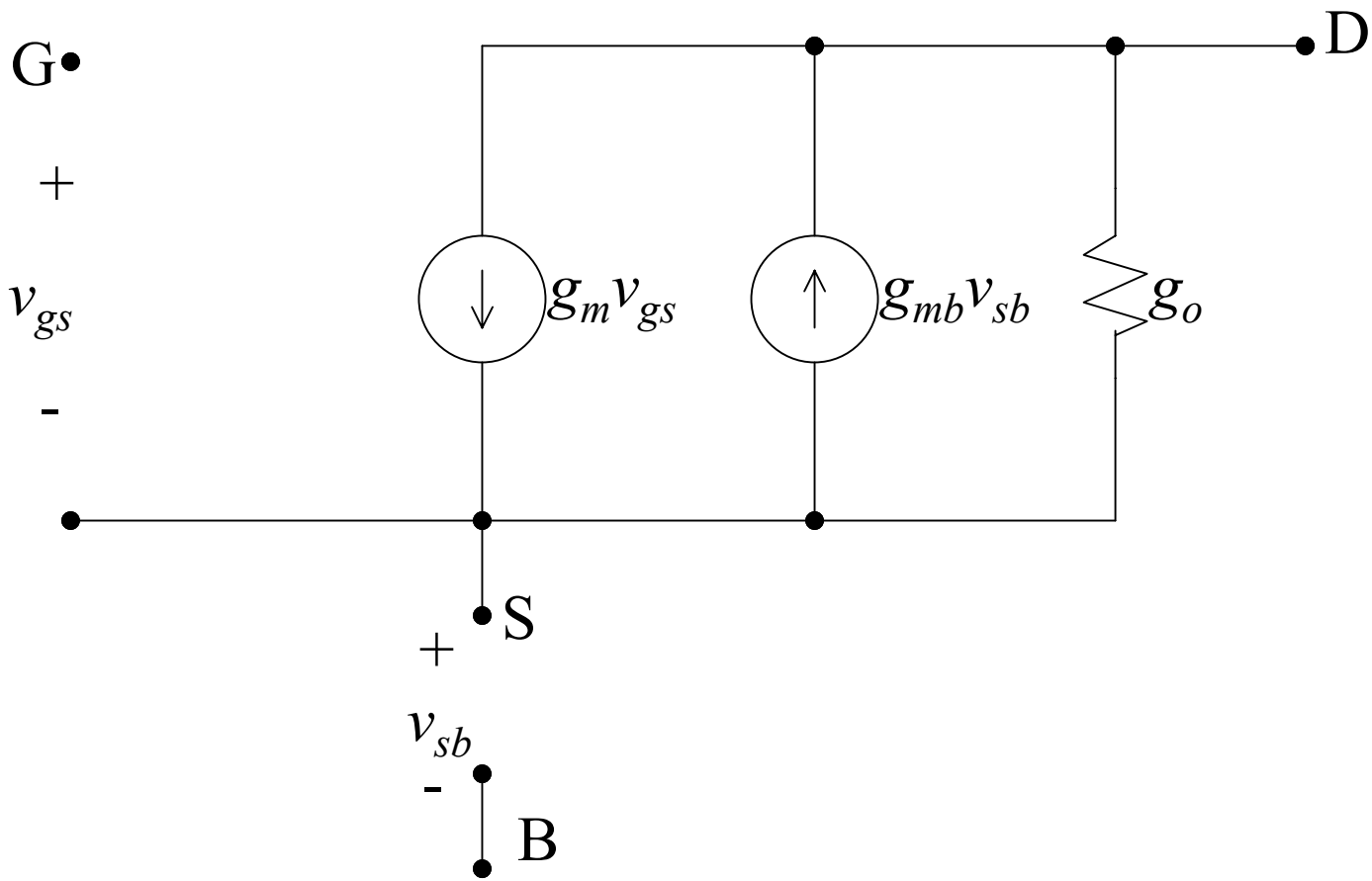
L is metallurgic channel length, $L=Lg-2L_D$



L entered in SPICE is Lg. Internally SPICE will subtract $2L_D$ to model the transistor

Low Frequency Model

NEGLECT CAPACITANCES IN LOW FREQUENCY MODEL



g_m,g_{mb} and g_o are the 3 conductance parameters

Transconductance, g_m

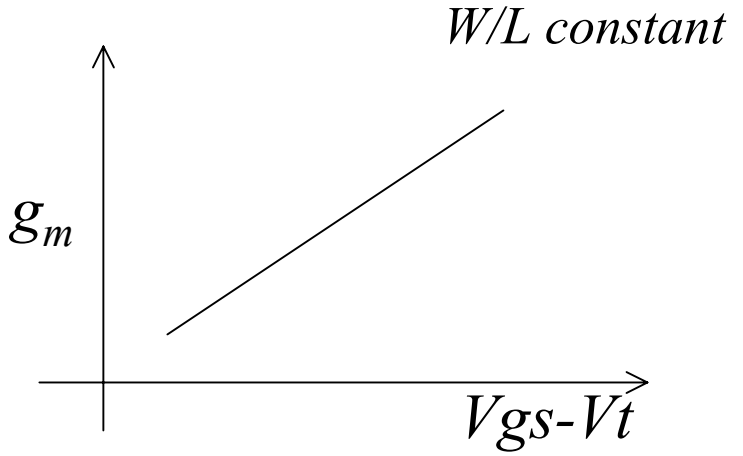
$$g_m = \frac{i_{ds}}{v_{gs}} = \frac{\partial I_{ds}}{\partial V_{gs}} \quad \text{, with constant Vds, Vsb}$$

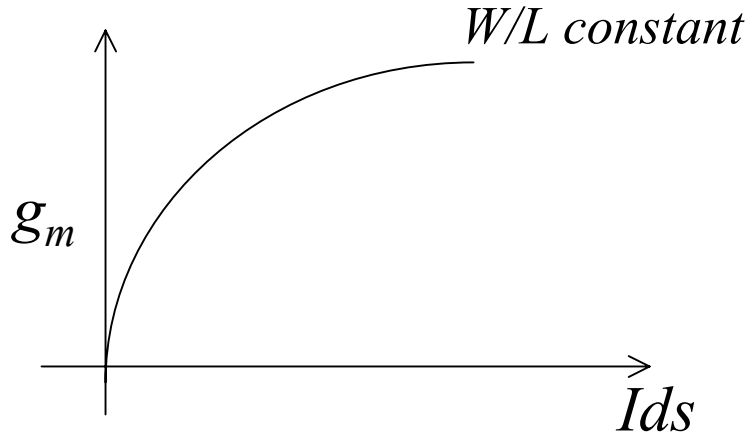
$$g_m = \frac{\mu \varepsilon_{ox} W \left(V_{gs} - V_t\right)\left(1 + \lambda V_{ds}\right)}{L T_{ox}}$$

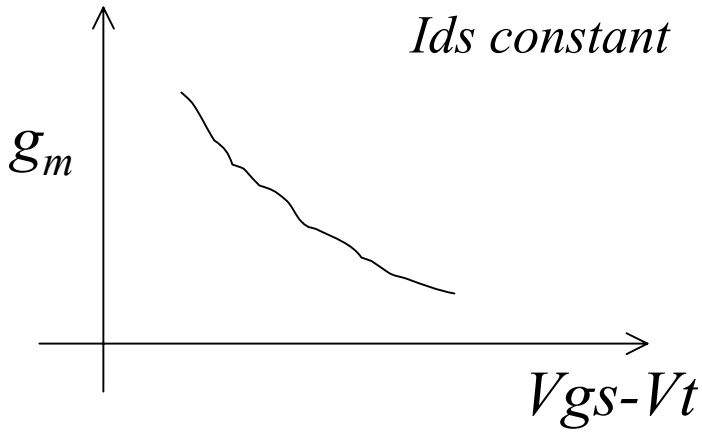
$$g_m = \sqrt{\frac{2 \mu \varepsilon_{ox} W I_{ds} \left(1 + \lambda V_{ds}\right)}{L T_{ox}}}$$

$$g_m = \frac{2 I_{ds}}{V_{gs} - V_t}$$

g_m versus device dimension and bias point







In order to obtain large g_m,
the input/output swing
trades-off with the transistor size

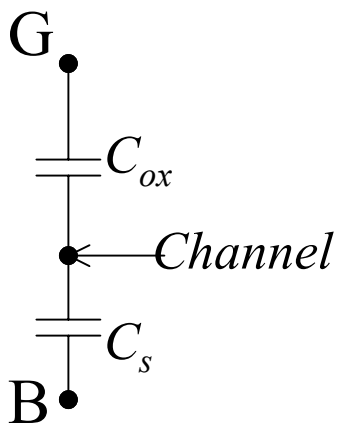
Body effect transconductance, g_{mb}

$$g_{mb} = -\frac{i_{ds}}{v_{sb}} = -\frac{\partial I_{ds}}{\partial V_{sb}} = -\frac{\partial I_{ds}}{\partial V_t}\frac{\partial V_t}{\partial V_{sb}} \quad , \quad \textit{with constant Vds, Vgs}$$

$$g_{mb} = \frac{\gamma}{2\sqrt{2|\phi_b| + V_{sb}}} g_m = \frac{C_s}{C_{ox}} g_m = \eta g_m$$

η Is typically between 0.1 to 0.3

The back gate effect



- The body terminal acts as the back gate
- However, the front gate is more efficient in controlling the channel compared to back gate
- This is achieved through the proper design of the MOS transistor

Output conductance, g_o

$$g_o = \frac{i_{ds}}{v_{ds}} = \frac{\partial I_{ds}}{\partial V_{ds}} = \frac{\partial I_{ds}}{\partial L_{eff}} \frac{\partial L_{eff}}{\partial V_{ds}} \quad , \quad \textit{with constant Vgs, Vsb}$$

$$g_o = \lambda \frac{\mu \varepsilon_{ox} W \left(V_{gs} - V_t\right)^2}{2 L T_{ox}}$$

$$g_o = \frac{\Delta}{2\left(L - \Delta\right)\left(V_{ds} - V_{dsat}\right)} I_{ds} \quad , \quad \textit{more rigorous analysis}$$

To simplify modeling, λ is assumed to be independent of *Vds*

$$g_o \approx 0.01 g_m$$

The output conductance is inversely proportional to L

Typical values for 0.35μm Technology

3.3V, L=0.35μm, W=0.7μm T_{ox}*=7nm,* V_{t0}*=0.6V,*

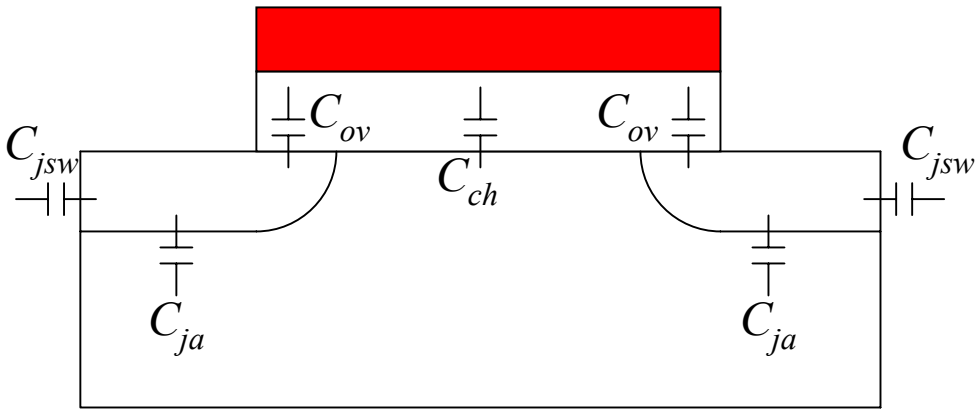
Vgs-Vt=1V

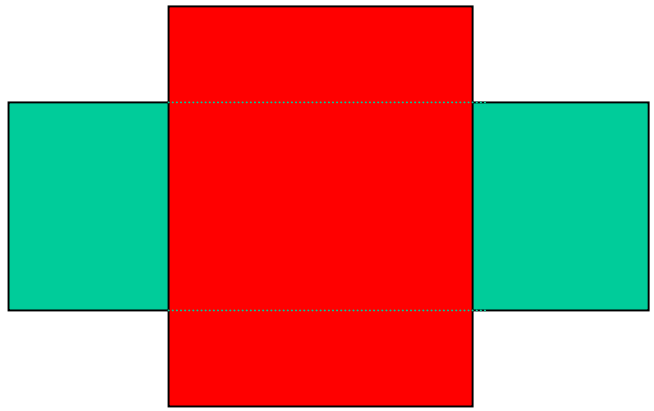
$g_m \sim 10^{-4}$ A/V $\qquad$ $1 / g_m \sim 10$ kΩ

$g_{mb} \sim 0.1\, g_m \sim 10^{-5}$ A/V $\qquad$ $1 / g_{mb} \sim 100$ kΩ

$g_o \sim 0.01 g_m \sim 10^{-6}$ A/V $\qquad$ $1 / g_o \sim 1000$ kΩ

MOS Capacitances





C_{ch} is split between source/drain depending on biasing condition

$C_{ov}=nWL_DC_{ox}$ *where* $1<n<2$ *due to fringing and* $C_{ox}=\varepsilon_{ox}/T_{ox}$

C_{ja} = area component of junction depletion capacitance

C_{jsw} = side wall component of junction depletion capacitance

MOS capacitances

$C_{sb}=AS*C_{ja}(V_{sb}) + PS*C_{jsw}(V_{sb})$

$C_{db}=AD*C_{ja}(V_{sb}) + PD*C_{jsw}(V_{sb})$

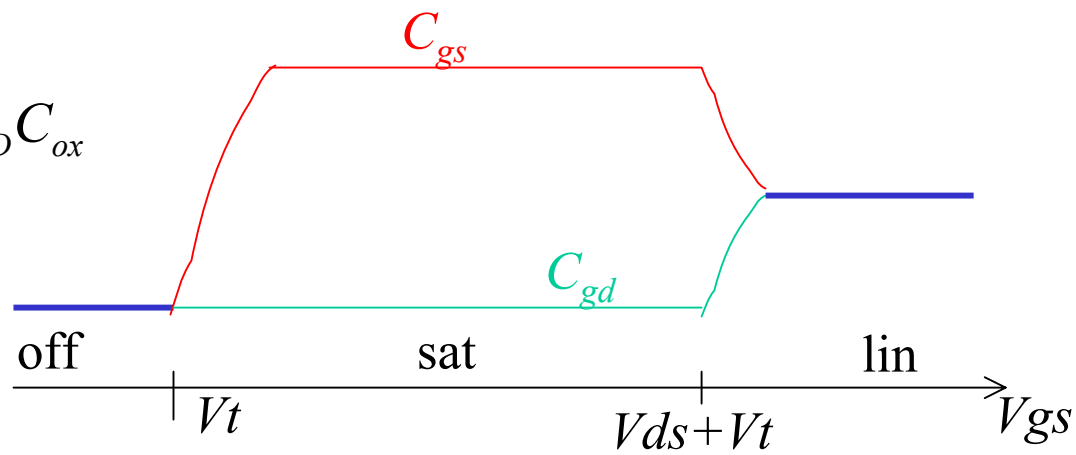
AS,PS source area, perimeter; AD, PD drain area, perimeter

In saturation:

$$C_{gs} = \frac{2}{3}WL_{eff}C_{ox} + nWL_DC_{ox} \qquad C_{gd} = nWL_DC_{ox}$$

In linear region:

$$C_{gs} = C_{gd} = \frac{WLC_{ox}}{2} + nWL_DC_{ox}$$



Typical capacitance values

For L=0.35mm and W=1mm, Junction width = 0.7mm

C_{ox}=4.3fF/μm^2 and C_{ja}=1fF/ μm^2

C_{db}=C_{sb}=0.7fF

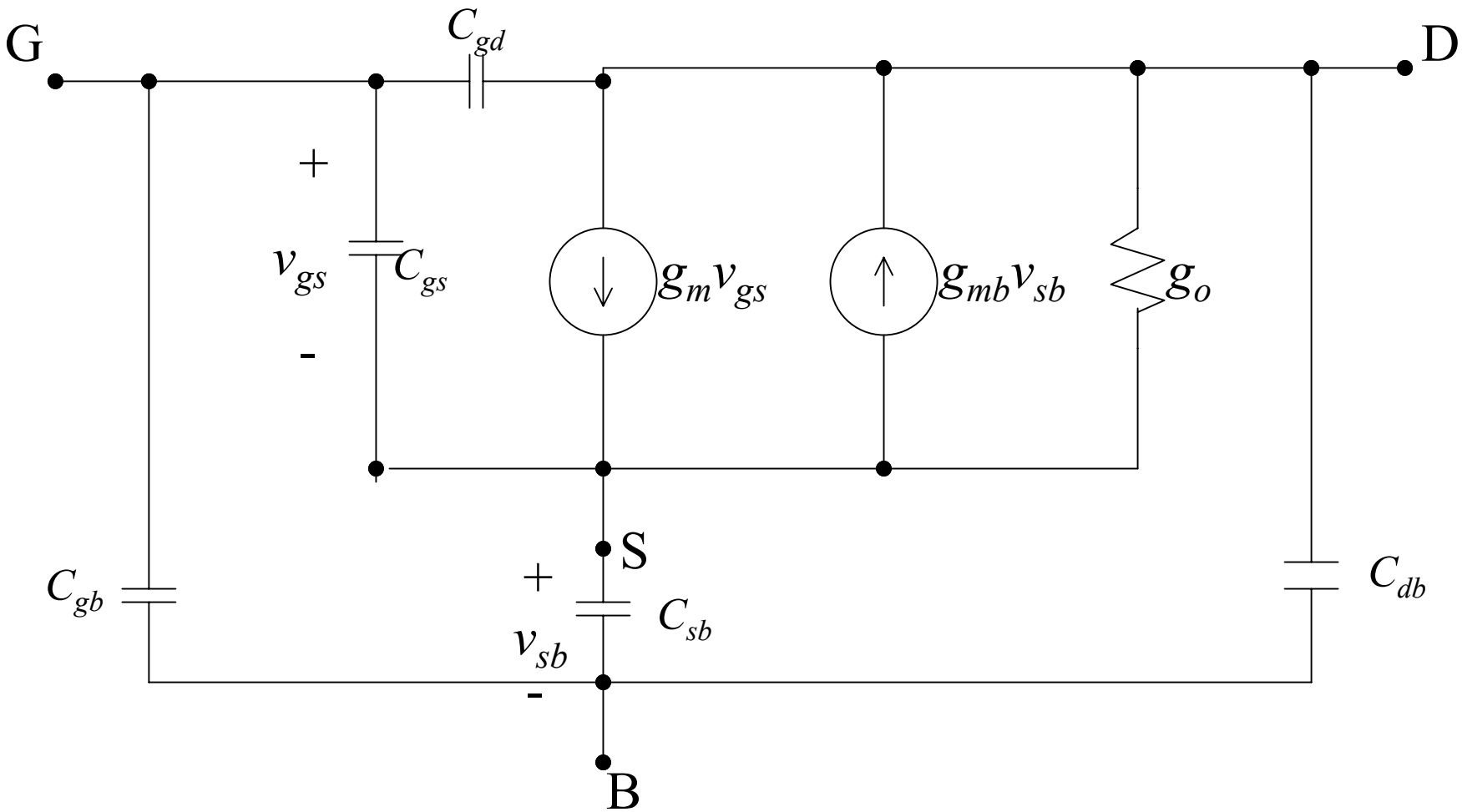
In saturation:

C_{gs}=1.1fF C_{gd}=0.2fF

In linear region:

C_{gs}=C_{gd}=0.8fF

Small signal equivalent circuit



Gate to bulk capacitance C_{gb} is insignificant and hence neglected

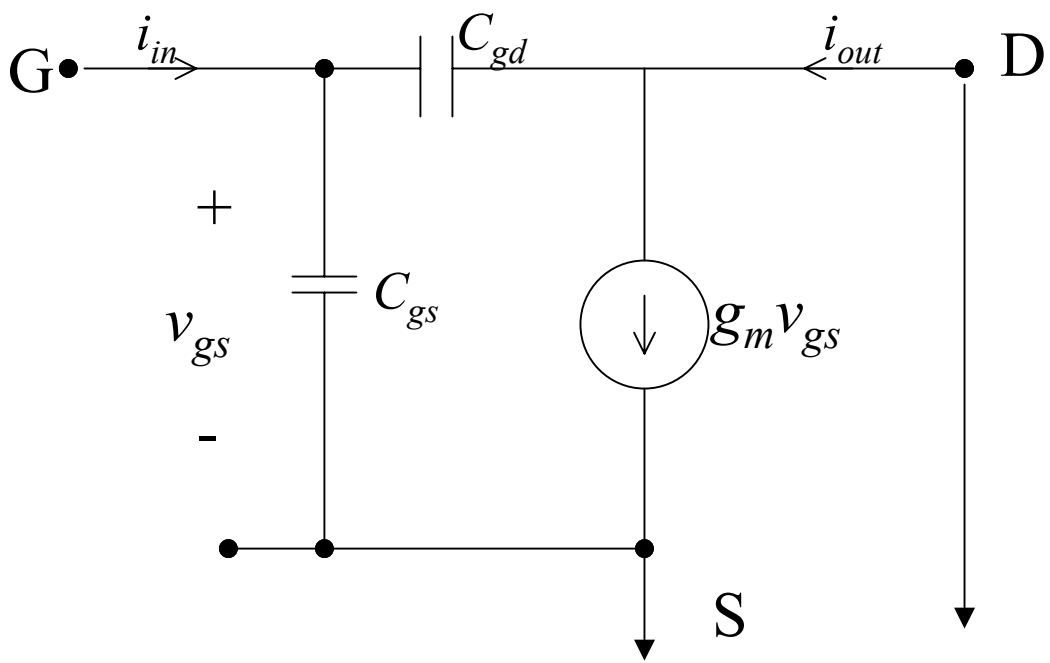
Cut-off frequency

Cut off frequency (Transition frequency)

$$f_t = f\,|_{i_{out}=i_{in}} \quad , \text{ with } Z_L=0$$

At f_t, the out put short circuit current gain is unity

f_t is the performance metric of transistor for high frequency operation



$$i_{in} = sC_{gs}v_{gs} + sC_{gd}v_{gs}$$

$$i_{out} = g_m v_{gs} - sC_{gd}v_{gs}$$

$$At\ s=j\omega_t=2\pi f_t,\ |i_{out}|=|i_{in}|$$

$$w_t = \frac{g_m}{C_{gs}\sqrt{1+{2C_{gd}}/{C_{gs}}}}$$

$$f_t = \frac{1}{2\pi} w_t$$

Scaling trend for f_t

$$f_t \cong \frac{1}{2\pi}\frac{g_m}{C_{gs}}$$

Technology scaling factor k > 1

Ideally gm does not scale and C scales as 1/k
hence f_t increases by a factor k

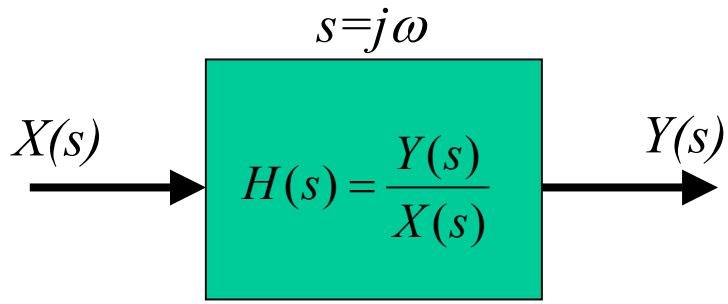
However, if Vt is not scaled due to leakage constraints then g_m will decrease in DSM regime, thus affecting the scaling trend

In cascaded voltage gain stages f_t forms the upper limit for the *unity gain-bandwidth*

Concept of Poles and Zeros

Poles and Zeros

The transfer function $H(s)$ of any system can be represented as



$$H(s) = \frac{\left(1 - s/z_1\right)\left(1 - s/z_2\right)\ldots}{\left(1 - s/p_1\right)\left(1 - s/p_2\right)\ldots}$$

$z_1, z_2, \ldots$ are zeros and $p_1, p_2, \ldots$ are the poles of the system

Physical Significance of Poles and Zeros

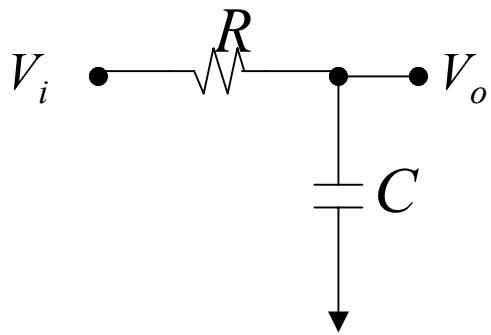
Poles

- Any capacitance which shunts the input-to-output signal path to ground results in a pole for the circuit
- Theoretically every node in the I/O signal path introduces a pole
- The poles degrade the high frequency response of the circuit

Zeros

- Any capacitance which appears in the input-to-output signal path results in a zero for the circuit
- The zeros enhance the high frequency response of the circuit
- The zeros degrade the low frequency response of the circuit

Impact of Pole

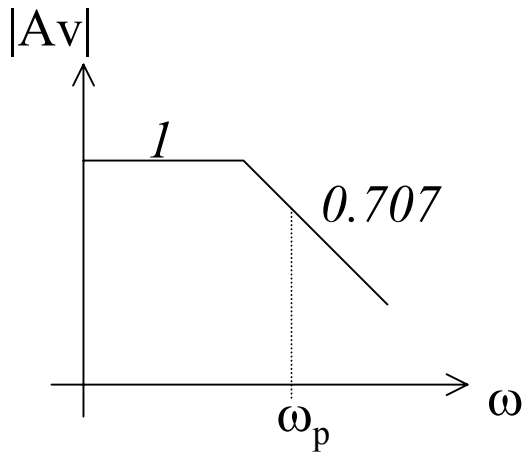


$$\frac{Vo}{Vi} = \frac{1/sC}{R + 1/sC}$$

$$\frac{Vo}{Vi} = \frac{1}{1 - s/(-1/RC)}$$

The pole is on the left half of *s* plane (i.e. *s=-1/RC* $\Rightarrow$ *stable system*)

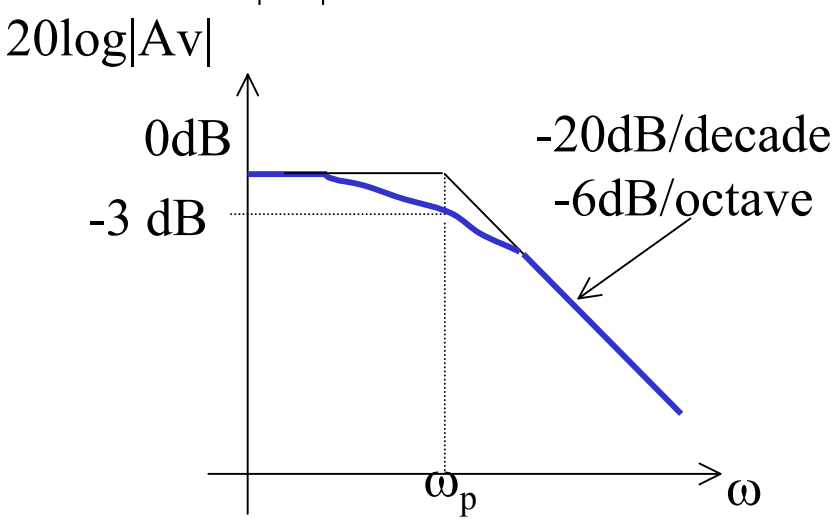
The pole frequency ω_p=1/RC

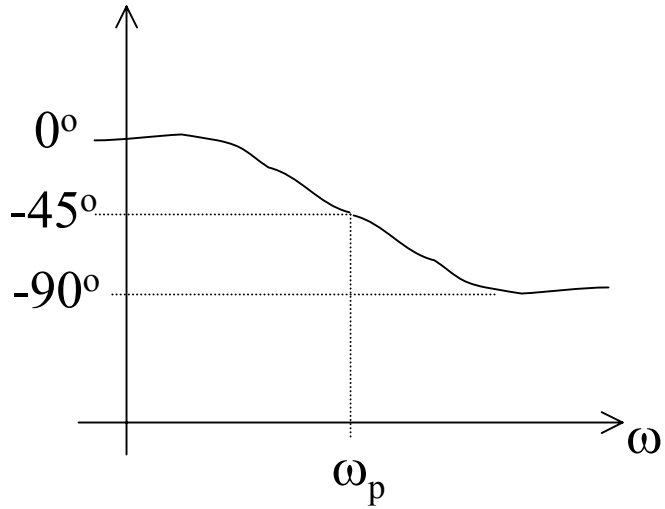


Gain and Phase response due to pole

$$G = 20\log\left|\frac{V_o}{V_i}\right| = -20\log\left[\sqrt{1+\left(w/w_p\right)^2}\right]$$

$$\phi = -\tan^{-1}\left[\frac{w}{w_p}\right]$$





At ω_p the gain is 3dB lower

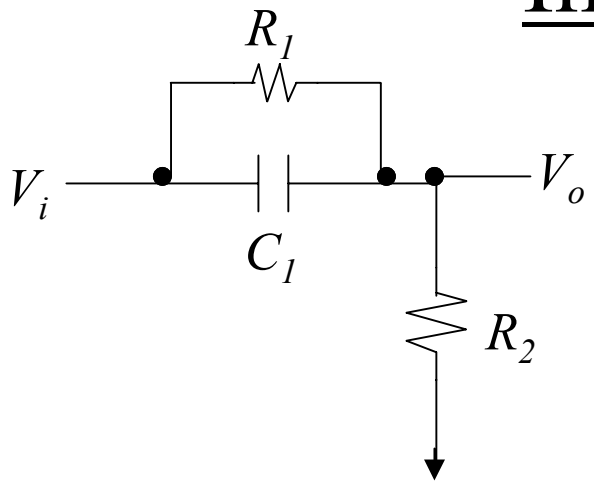
Beyond ω_p, the gain decreases at a rate of –20dB/decade

At ω_p the the phase shift is -45^o

At $0.1\omega_p$ the the phase shift saturates to 0^o

At $10\omega_p$ the the phase shift saturates to -90^o

Impact of zero

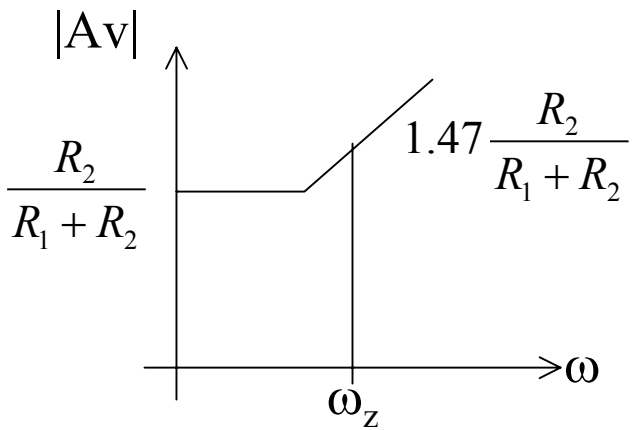


$$\frac{V_o}{V_i} = \frac{R_2}{R_1 + R_2} \frac{1 + sR_1C_1}{1 + \dfrac{sR_1R_2C_1}{R_1 + R_2}}$$

$$\frac{V_o}{V_i} = \frac{R_2}{R_1 + R_2} \frac{1 - \dfrac{s}{(-1/R_1C_1)}}{1 - \dfrac{s}{-(R_1+R_2)/R_1R_2C_1}}$$

$z=-1/R_1C_1$ and $p=-(R_1+R_2)/R_1R_2C_1$

The zero frequency $\omega_z=1/R_1C_1$

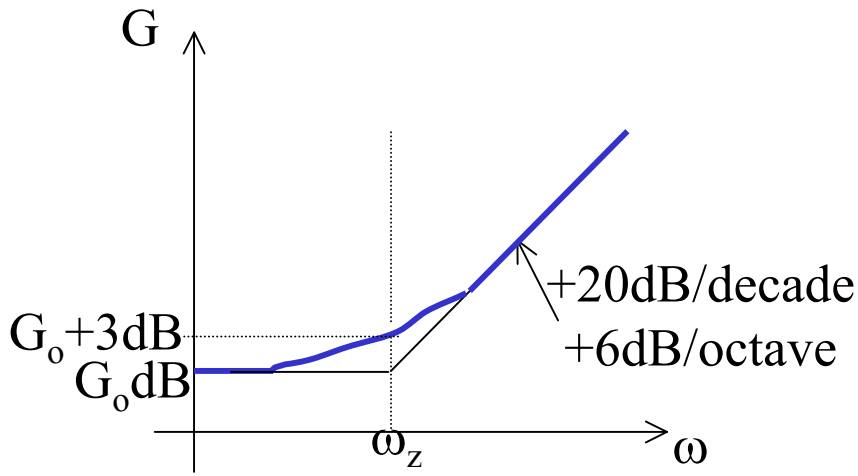


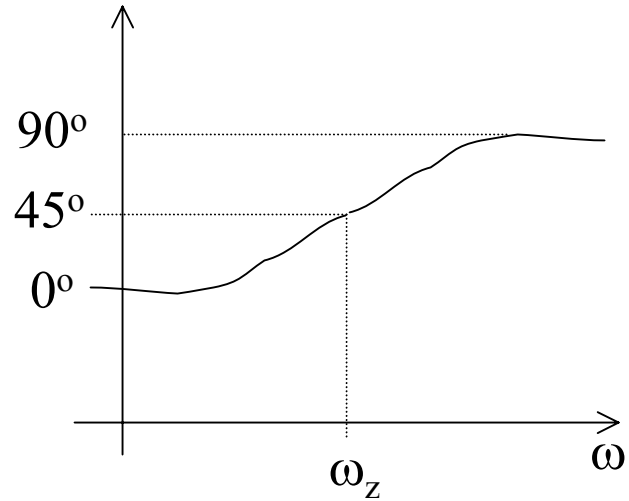
Gain and Phase response due to zero

$$G = 20\log\left|\frac{V_o}{V_i}\right| = G_o + 20\log\left[\sqrt{1+\left(w/w_p\right)^2}\right]$$

$$\phi = \tan^{-1}\left[\frac{w}{w_p}\right]$$

$$G_o = 20\log\left[\frac{R_2}{R_1+R_2}\right]$$





At ω_z the gain is 3dB higher

Beyond ω_z, the gain increases at a rate of 20dB/decade

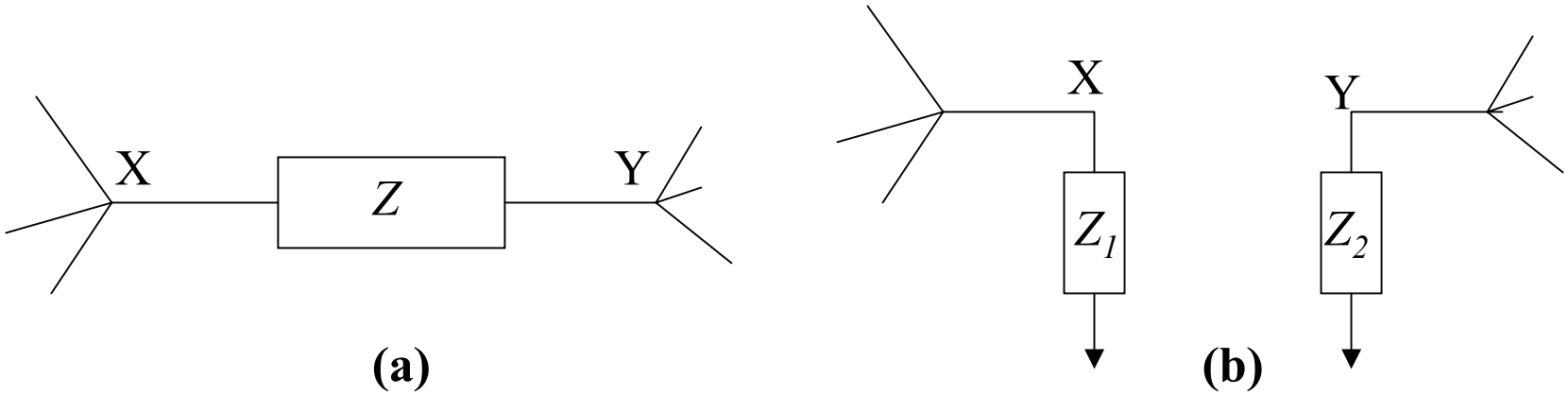
At ω_z the the phase shift is 45o

At $0.1\omega_z$ the the phase shift saturates to 0o

At $10\omega_z$ the the phase shift saturates to 90o

Miller Approximation

Miller's Theorem



(a) **(b)**

If the circuit in (a) can be converted into that of (b) then

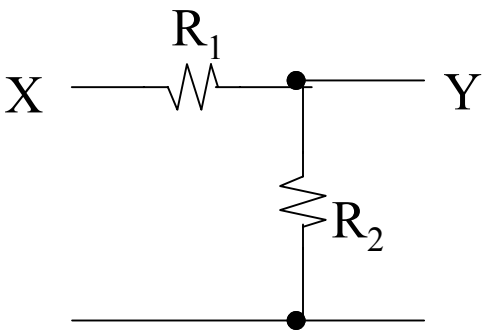
$$Z_1 = \frac{Z}{1 - A_v} \quad \textit{and} \quad Z_2 = \frac{Z}{1 - A_v^{-1}} \quad , \quad \textit{where} \quad A_v = \frac{V_y}{V_x}$$

Proof: For the two circuits to be equivalent, the current flowing through Z from X to Y in (a) should be same as current flowing through Z1 in (b)

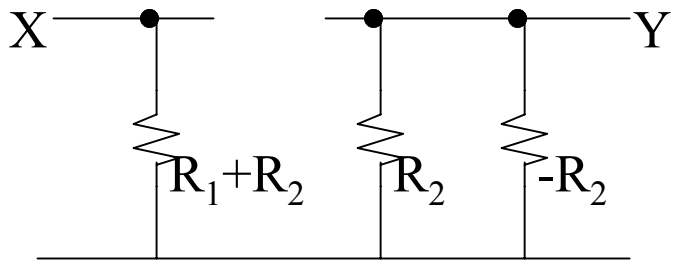
$$\frac{V_x - V_y}{Z} = \frac{V_x}{Z_1} \Rightarrow Z_1 = \frac{Z}{1 - V_y / V_x} \qquad \text{Similarly,} \quad Z_2 = \frac{Z}{1 - V_x / V_y}$$

Caveats of Miller's Theorem

If the impedance Z forms the only signal path between X and Y,
Then the conversion is not valid



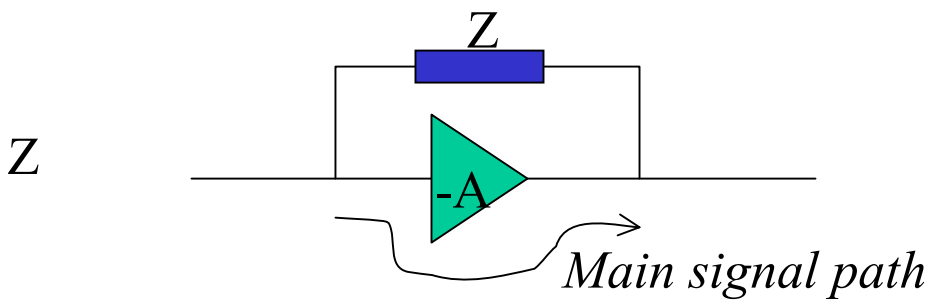
$$A_v = \frac{R_2}{R_1 + R_2}$$



The output gain is no longer preserved in the modified circuit

The zeros in the transfer functions are discarded by Miller's transformation

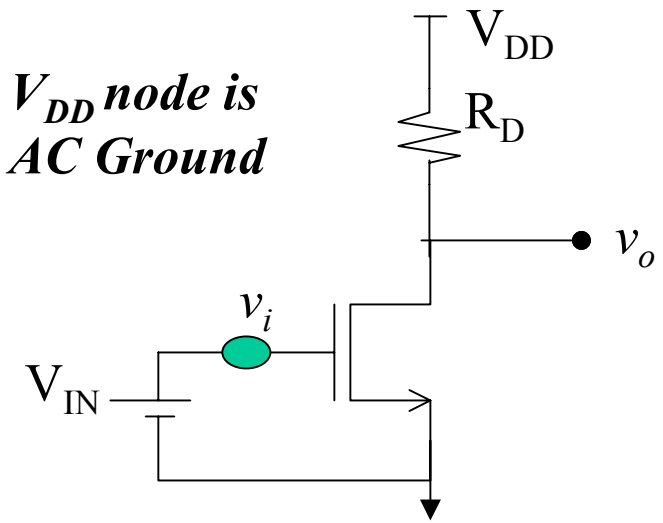
The theorem proves useful when the impedance is in parallel with main signal path



Single stage amplifier

- Common Source
  - Resistive load
  - Diode connected load
  - PMOS current source load
  - Source degeneration
- Source Follower
- Common Gate
- Cascode
- Folded Cascode

Common Source with Resistive load



V_{DD} node is AC Ground

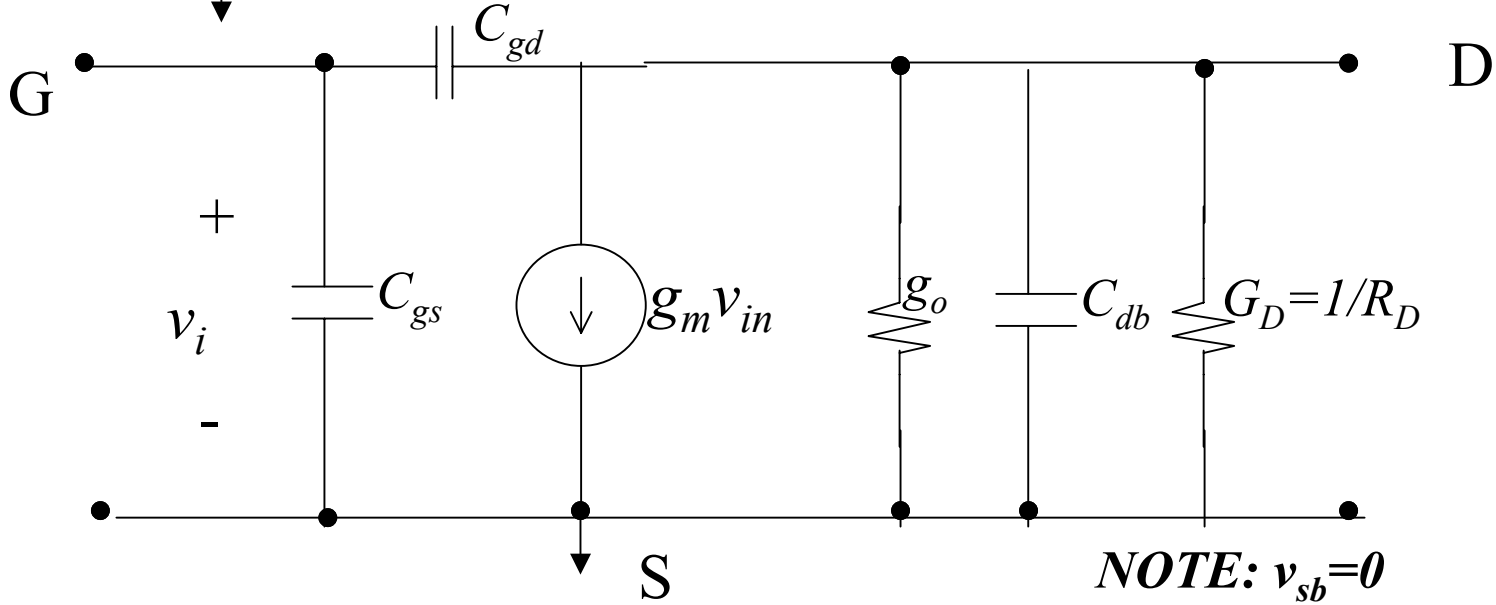
DC bias point

Transistor is kept in saturation in order to maximise gm and ro and hence the gain

$$V_{in} = V_{gs}$$

$$I_{ds} = \frac{\mu\varepsilon_{ox} W}{T_{ox} L} \frac{(Vgs - Vt)^2}{2} (1 + \lambda V_{ds})$$

$$V_{DD} = I_{ds} R_D + V_{ds}$$



NOTE: v_{sb}=0

If there is any output load capacitance, that should be added with C_{db}

AC analysis

Applying KCL at the output node:

$$(v_o - v_i)sC_{gd} + g_m v_i + (g_o + G_D + sC_{db})v_o = 0$$

DC gain at s=0:

$$A_v(0) = -\frac{g_m}{g_o + G_D} = -g_m \frac{r_o R_D}{r_0 + R_D}$$

The gain at high frequency:

$$A_v(0) = -\frac{g_m}{g_o + G_D} \frac{1 - sC_{gd} / g_m}{1 + s(C_{db} + C_{gd}) / (g_o + g_L)}$$

$z_1 = g_m/C_{gd}$ *and* $p_1 = -(g_o + G_L)/(C_{db} + C_{gd})$

NOTE: Miller's theorem also gives the same pole frequency

Important Trade-offs

The maximum DC gain can never exceed the intrinsic gain of the transistor which is given by $g_m r_o$ (For sub micron transistor, the typical intrinsic gain is around 30

In the limit $R_D < r_o$, The gain increases with increasing R_D and is Approximately equal to $g_m R_D$

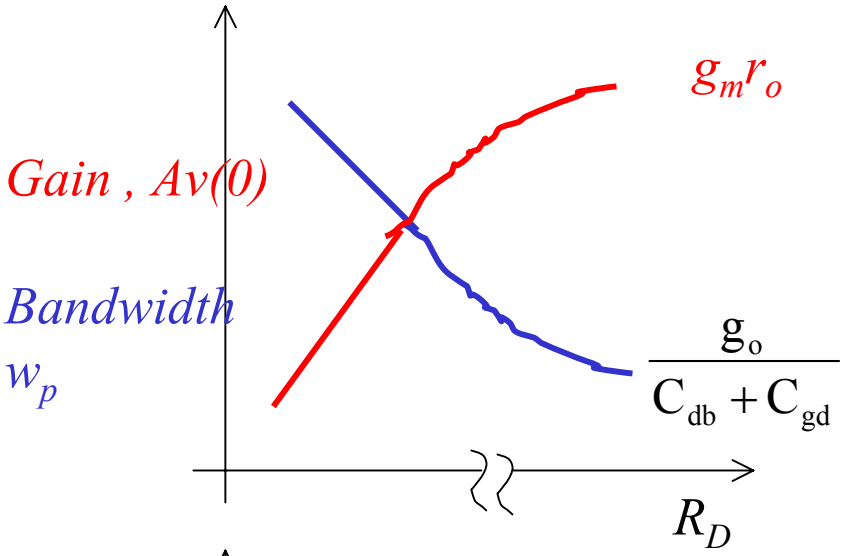
High R_D also results in lower pole frequency and hence the band width

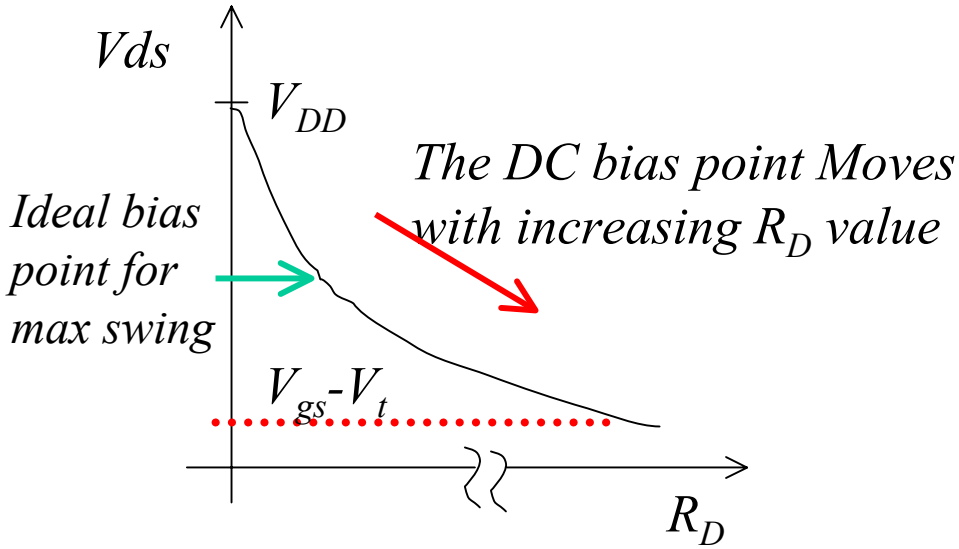
High R_D also results in smaller allowable output voltage swings because Transistor come out of saturation when $V_{DD}-V_{RD} < V_i - V_t$

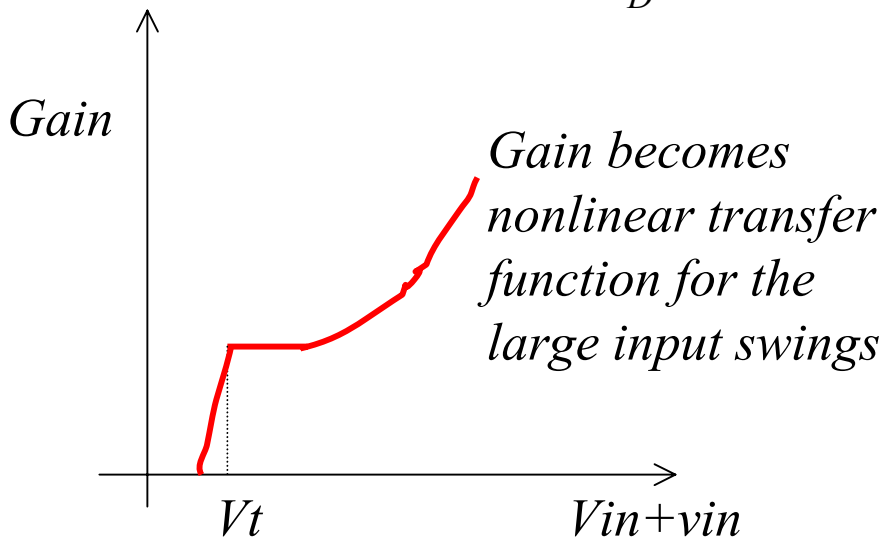
If attempt to decrease V_{RD} by decreasing Vgs-Vt and increasing W/L (thereby keeping g_m constant and decreasing I_d), the input pole can become Dominant, the input swing decreases, large area overhead on Silicon

The gain is a strong function of g_m, which in turn depends on V_{gs}. This results is large non linearity when the input swings are large

Trade-off







WISHLIST

Need to make gain insensitive to the input swings

Need to decouple the DC bias point from the AC requirements

Diode connected load

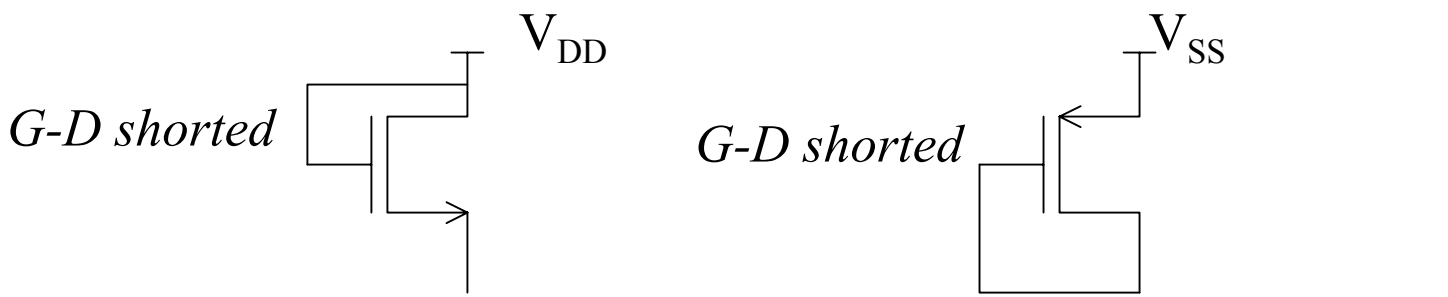
Very inefficient to implement a large resistance in CMOS technology

Hence CS with resistive load is never implemented on CMOS

Diode connected transistor can act as a resistance

This configuration can make the gain a little more insensitive to input

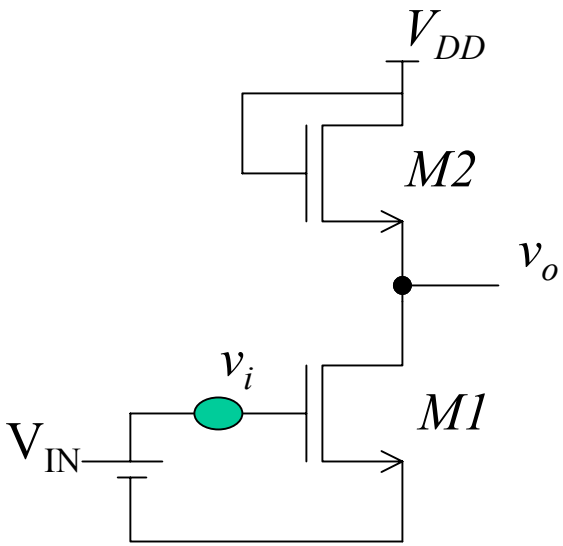
NMOS or PMOS diode connected load can be used

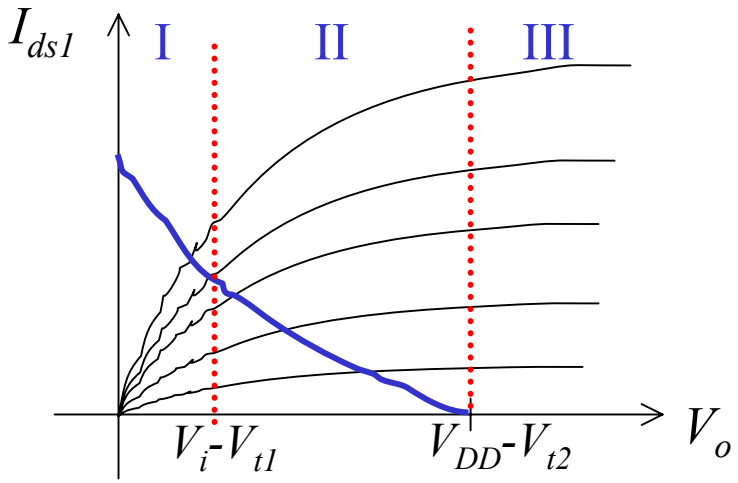


Diode connected NMOS (sensitive to body effect)

Diode connected PMOS (insensitive to body effect)

CS with diode connected load

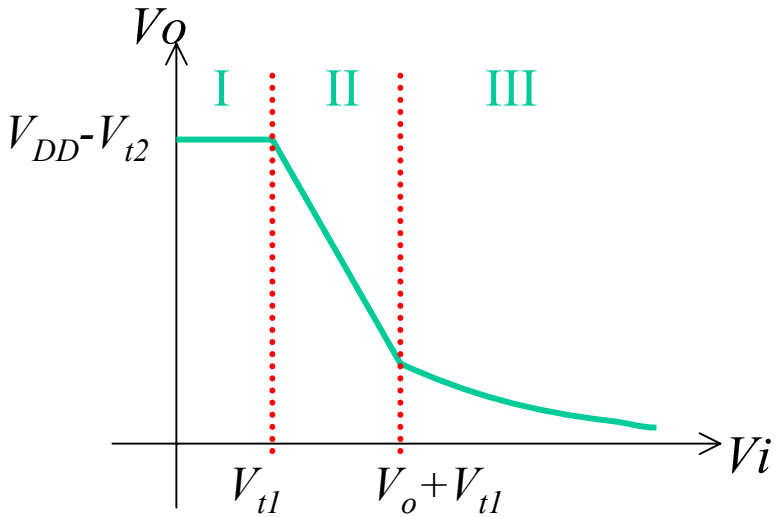




Region 1: M1 linear, M2 saturation

Region 2: M1 saturation, M2 saturation

Region 3: M1 saturation, M2 cutoff



Region 1: M1 cutoff, M2 saturation

Region 2: M1 saturation, M2 saturation

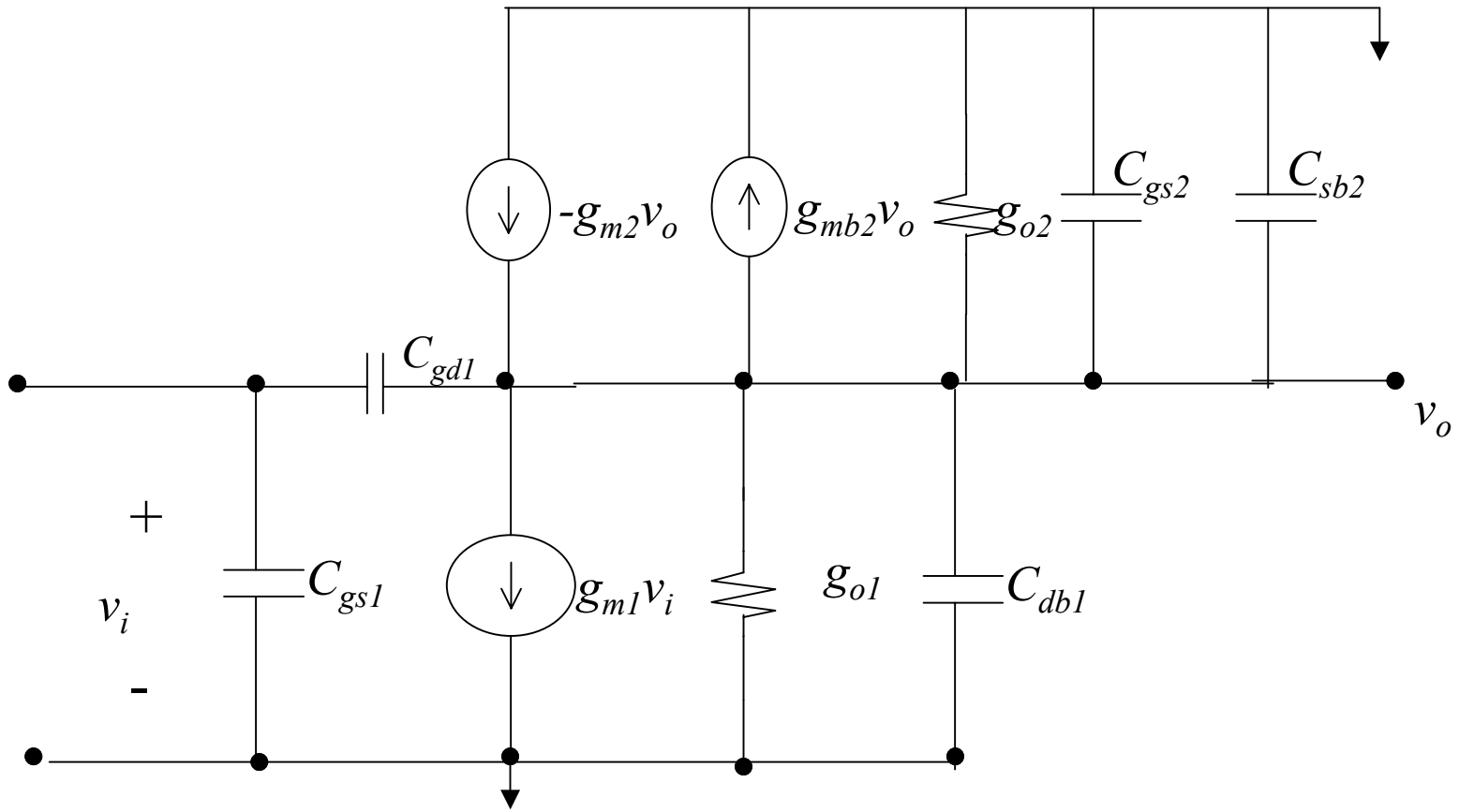
Region 3: M1 linear, M2 saturation

Region 2 is useful operating region

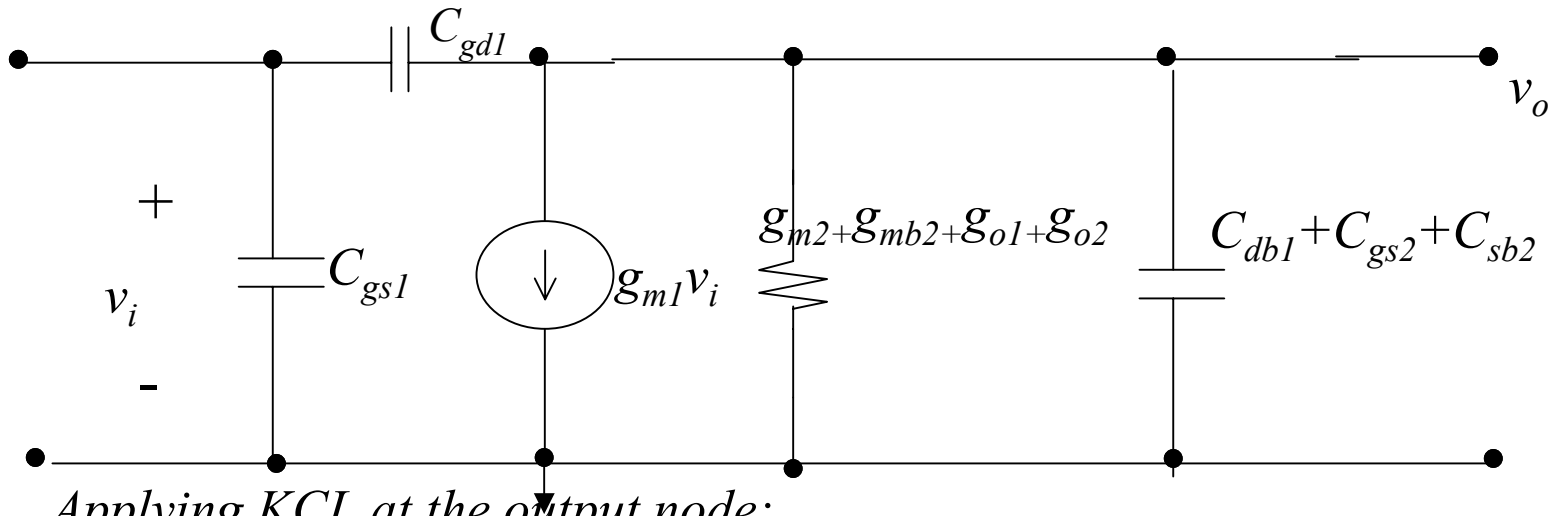
AC analysis

For M2, C_{DB2} is shorted. V_{gs2}=-V_o and V_{sb}=V_0

If the controlling voltage for a voltage controlled current source, is the voltage across the VCS itself, then it reduces to a conductance



Low frequency gain



Applying KCL at the output node:

$$(v_o - v_i)sC_{gd1} + g_{m1}v_i + (g_T + sC_T)v_o = 0$$

DC gain at s=0:

$$A_v(0) = -\frac{g_{m1}}{g_T} = -\frac{g_{m1}}{g_{m2} + g_{mb2} + g_{o1} + g_{o2}} \approx -\frac{g_{m1}}{g_{m2}(1+\eta_2)}$$

The impedance seen looking into the source of the diode connected MOS transistor is ~$1/g_m$

Insensitivity of gain to input swing

The DC bias current $I_{ds1}=I_{ds2}$,

$$\frac{g_{m1}}{g_{m2}} = \frac{V_{gs2} - V_{t2}}{V_{gs1} - V_{t1}} \quad , \quad since \quad g_m = \frac{2I_{ds}}{V_{gs} - V_t}$$

Replacing V_{gs}-V_t in terms of gm and W/L

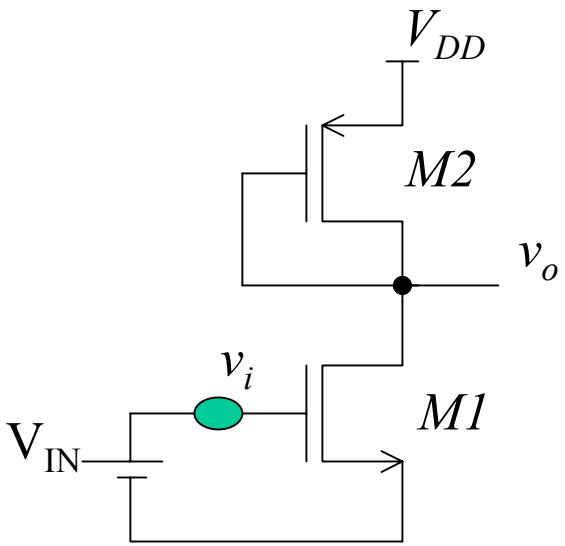
$$\frac{g_{m1}}{g_{m2}} = \sqrt{\frac{(W/L)_1}{(W/L)_2}}$$

$$A_v(0) = -\sqrt{\frac{(W/L)_1}{(W/L)_2}} \frac{1}{(1+\eta_2)}$$

The gain is essentially controlled by device dimension which are the design parameters under the control of the designer

There is some sensitivity on body bias

Body bias insensitivity using PMOS load



Since the body and the source of PMOS are connected to V_{DD} node, V_{sb}=0 and hence the body bias trans-conductance term becomes zero

$$A_v(0) = -\sqrt{\frac{(W/L)_1 \mu_n}{(W/L)_2 \mu_p}}$$

Where μ_n/μ_p is the ratio of electron and hole mobility
μ_n/μ_p ~2 to 3

Trade off between gain and output swing

This trade off is not eliminated even in diode connected load

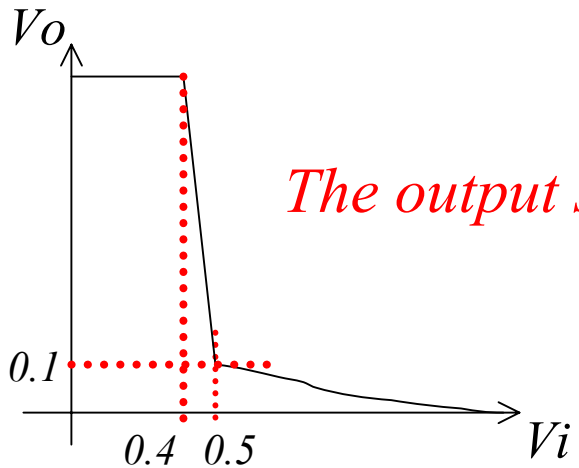
$$A_v(0) \approx \frac{V_{gs2} - V_{t2}}{V_{gs1} - V_{t1}}$$

Suppose V_{DD}=3.5V, V_{t1}= V_{t1}=0.4V, and gain is required to be 30

Let V_{gs1}-V_t=0.1V in order to maximise output swing (i.e. V_{IN}=0.5V)

Then, V_{gs2}-V_t=3V i.e. V_{gs2}=V_{ds2}=3.4V

Hence V_o=V_{ds2}=0.1V which is just at the verge of saturation.



The output swing is virtually zero

The necessity for higher gain makes the bias point unfavourable to output swing

High Frequency gain

$$A_v(s) = -\frac{g_{m1}}{g_{m2}+g_{mb2}+g_{o1}+g_{o2}} \frac{1-\frac{sC_{gd1}}{g_{m1}}}{1+\frac{s(C_{gd1}+C_{db1}+C_{gs2}+C_{sb2})}{g_{m2}+g_{mb2}+g_{o1}+g_{o2}}}$$

$$z = \frac{g_{m1}}{C_{gd1}}$$

$$p = -\frac{g_{m2}+g_{mb2}+g_{o1}+g_{o2}}{C_{gd1}+C_{db1}+C_{gs2}+C_{sb2}}$$

Gain bandwidth product:

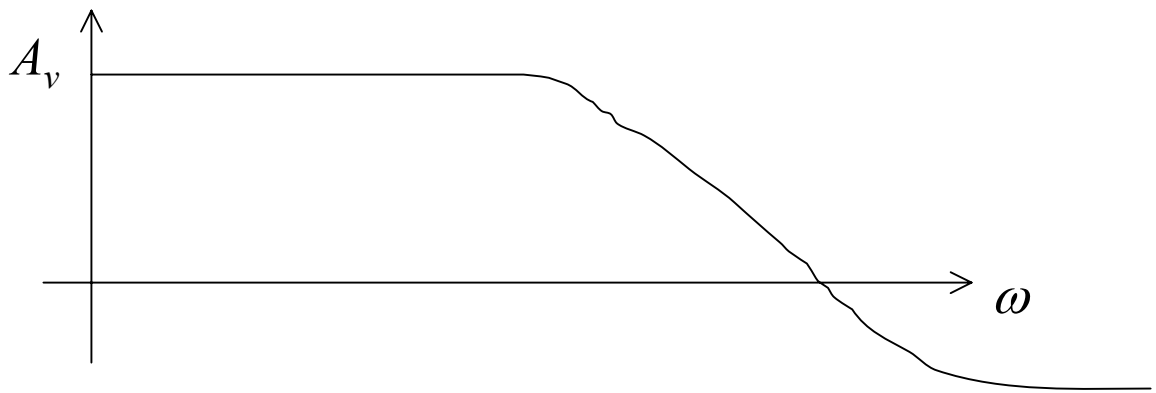
$$\frac{1}{2\pi}\left|A_v(0)\right|\omega_p = \frac{1}{2\pi}\frac{g_{m1}}{C_{gd1}+C_{db1}+C_{gs2}+C_{sb2}}$$

The relative location of pole and zero

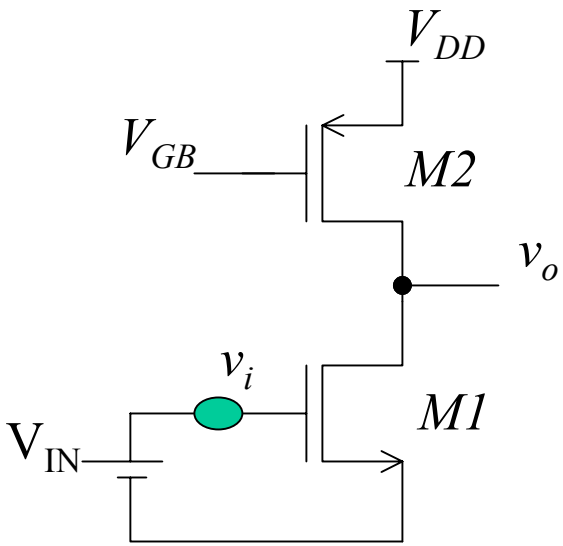
$$\left|\frac{p}{z}\right| = \frac{C_{gd1}(g_{m2} + g_{mb2} + g_{o1} + g_{o2})}{g_{m1}(C_{gd1} + C_{db1} + C_{gs2} + C_{sb2})}$$

$$\left|\frac{p}{z}\right| = \frac{1}{A_v(0)} \frac{1}{1 + \dfrac{C_{db1} + C_{gs2} + C_{sb2}}{C_{gd1}}} << 1$$

i.e. The amplifier works as a single pole transfer function with the dominant pole and insignificant zero location



CS amplifier with current source load



In order to remove the trade-off between the gain versus the output swing, the DC resistance should be decoupled from AC impedance of the load!

i.e. use a constant current source load

Remove the trans-conductance contribution to AC resistance by fixing G & S voltage

M2 in the above circuit acts as a constant current source as long as it is in saturation condition $|V_{ds}| > |V_{gs}\text{-}V_t|$

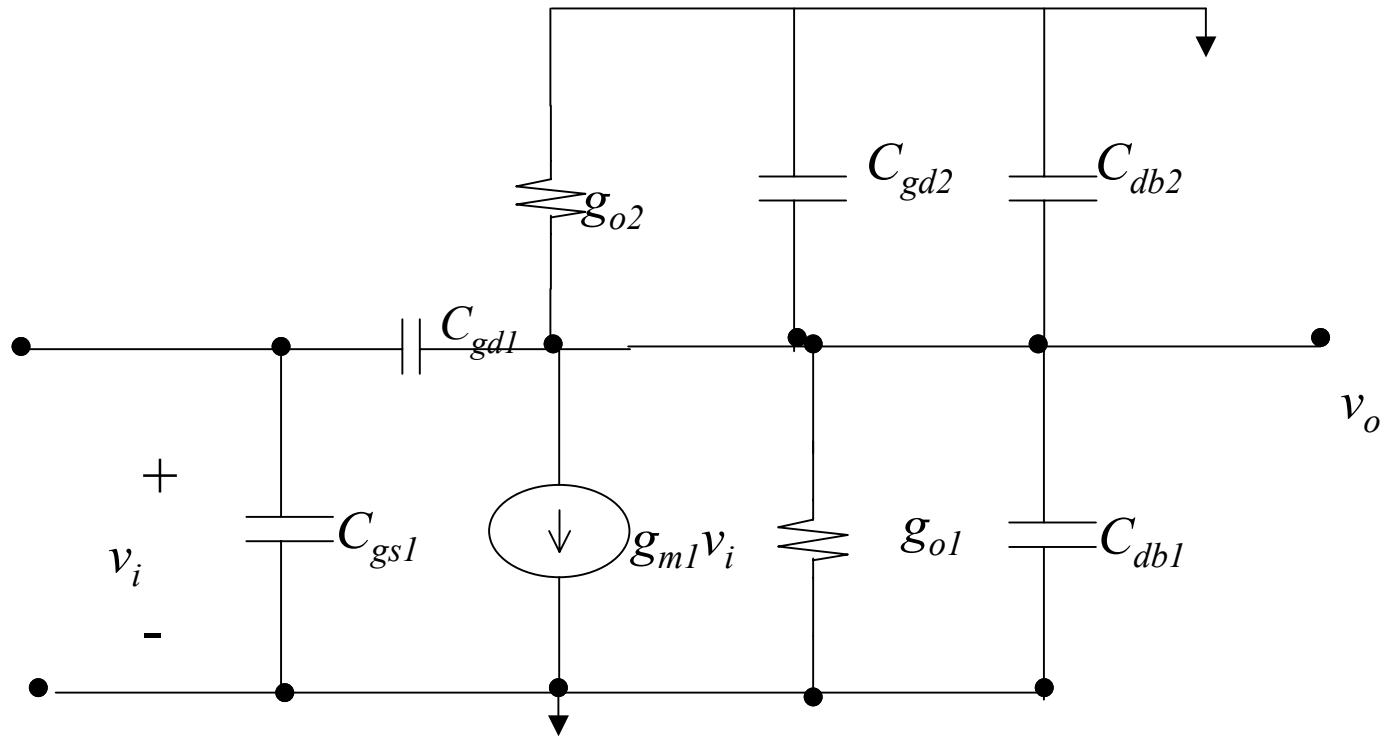
$$I_{ds} = \frac{\mu\varepsilon_{ox} W}{T_{ox} L_{eff}} \frac{(V_{DD} - V_{GB} - V_t)^2}{2} \qquad R_{ac} = \frac{1}{g_{o2}}$$

Small signal equivalent circuit

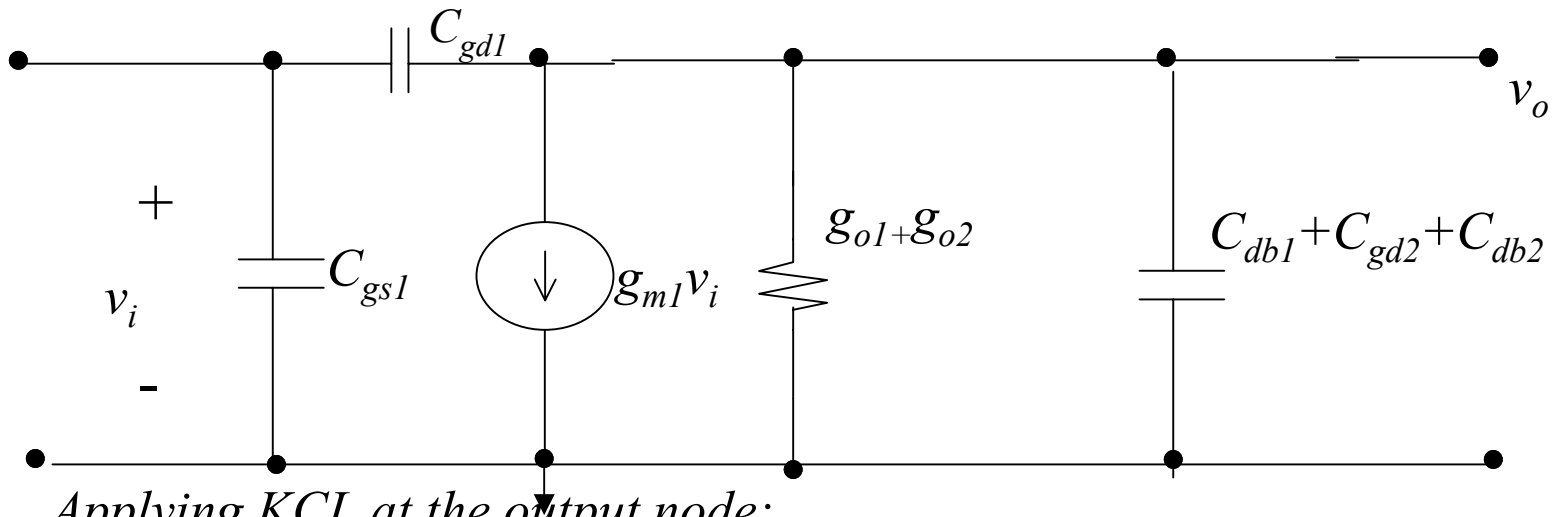
For M2,

C_{gs} and C_{sb} are shorted

v_{gs2} and v_{sb2} are zero and hence the corresponding g_m terms are absent



Low frequency gain



Applying KCL at the output node:

$$(v_o - v_i)sC_{gd1} + g_{m1}v_i + (g_T + sC_T)v_o = 0$$

DC gain at s=0:

$$A_v(0) = -\frac{g_{m1}}{g_T} = -\frac{g_{m1}}{g_{o1} + g_{o2}}$$

The impedance seen looking into the current source MOSFET load transistor is ~$1/g_o$

High Frequency gain

$$A_v(s) = -\frac{g_{m1}}{g_{o1}+g_{o2}} \frac{1-\frac{sC_{gd1}}{g_{m1}}}{1+\frac{s(C_{gd1}+C_{db1}+C_{gd2}+C_{db2})}{g_{o1}+g_{o2}}}$$

$$z = \frac{g_{m1}}{C_{gd1}}$$

$$p = -\frac{g_{o1}+g_{o2}}{C_{gd1}+C_{db1}+C_{gd2}+C_{db2}}$$

Gain bandwidth product:

$$\frac{1}{2\pi}\left|A_v(0)\right|\omega_p = \frac{1}{2\pi}\frac{g_{m1}}{C_{gd1}+C_{db1}+C_{gd2}+C_{db2}}$$

Comparison with diode connected load

↑ Low frequency gain $A_v(0)$ is higher

↑ 3 dB bandwidth is lower

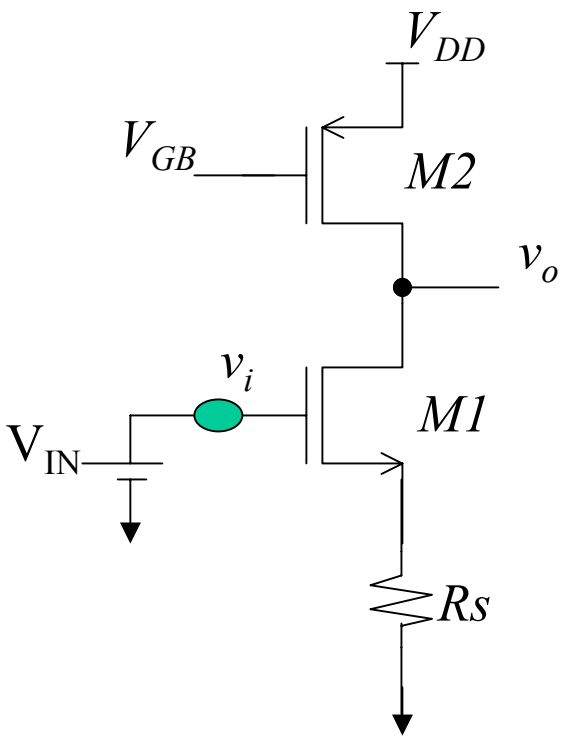
↑ Gain-bandwidth product may be slightly higher

↑ Larger output swing without sacrificing gain

— Dominant pole transfer function similar to diode connected load

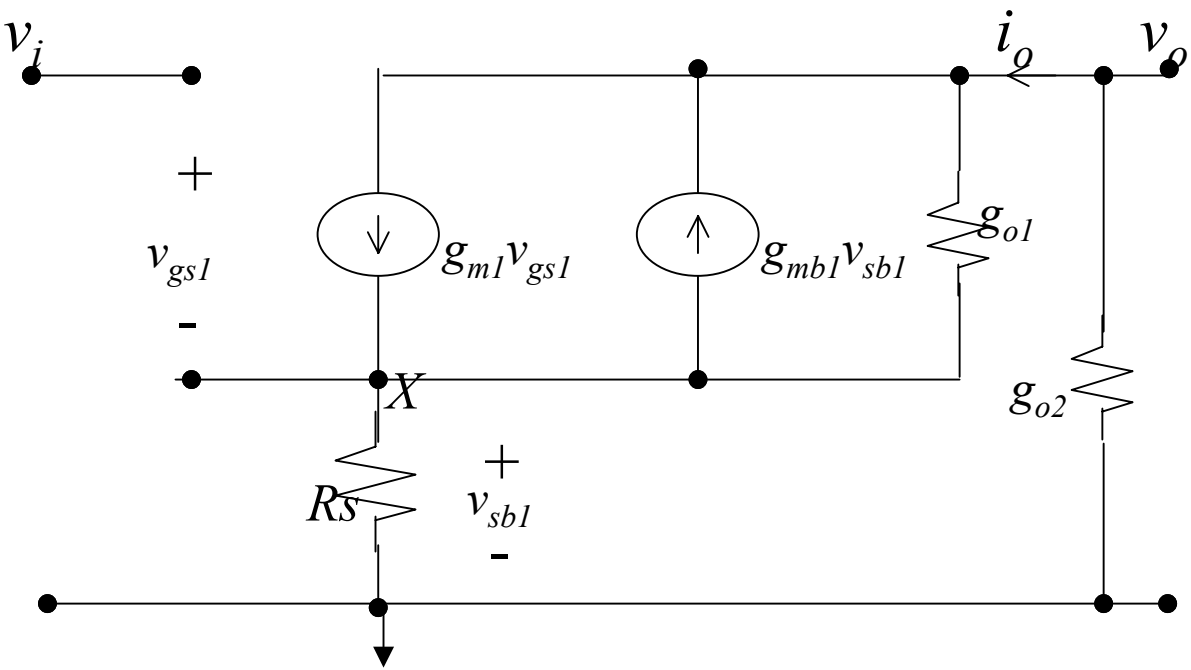
↓ Gain is dependent on DC bias condition (g_m)

CS with source degeneration



The low frequency behaviour is affected

The small signal equivalent at low frequency:



Low Frequency gain

$$i_o = -v_o g_{o2}$$

$$v_{sb} = i_o R_s$$

$$v_{gs1} = v_i - i_o R_s$$

Applying KCL at node X

$$i_o = g_{m1}(v_i - i_o R_s) - g_{mb1} i_o R_s + (v_o - i_o R_s) g_{o1}$$

$$A_v(0) = -\frac{g_{m1}}{g_{o1} + g_{o2}\left[1 + R_s(g_{m1} + g_{mb1} + g_{o1})\right]}$$

$$A_v(0) \approx -\frac{g_{m1}}{g_{o1} + g_{o2} g_{m1} R_s} = -\frac{1}{g_{o2} R_s}$$

Effect of source degeneration

The transconductance of M1 is de-rated from g_{m1} to $1/R_s$

The gain becomes insensitive to bias and input swing

The output resistance of M1 is increased by a factor $g_{m1}R_s$

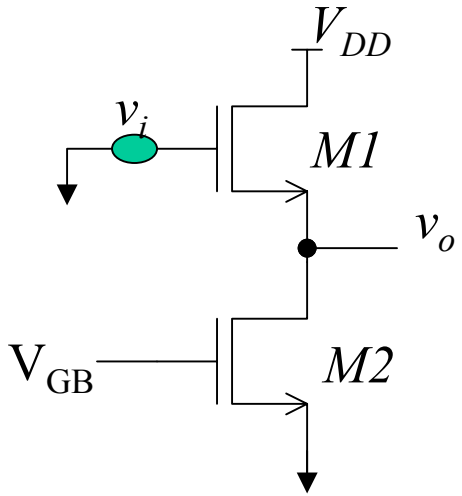
This concept is used in several analog circuits to enhance the out put impedance

$$A_v(0) \approx -\frac{g_{m1}}{g_{m1}R_s\left(g_{o2}+\frac{g_{o1}}{g_{m1}R_s}\right)}$$

$$A_v(0) \approx -\frac{1}{R_s}\frac{1}{\left(1/r_{o2}+1/r_{o1}g_{m1}R_s\right)}$$

The pole frequency (high frequency response) is not affected

Source Follower (Common Drain)

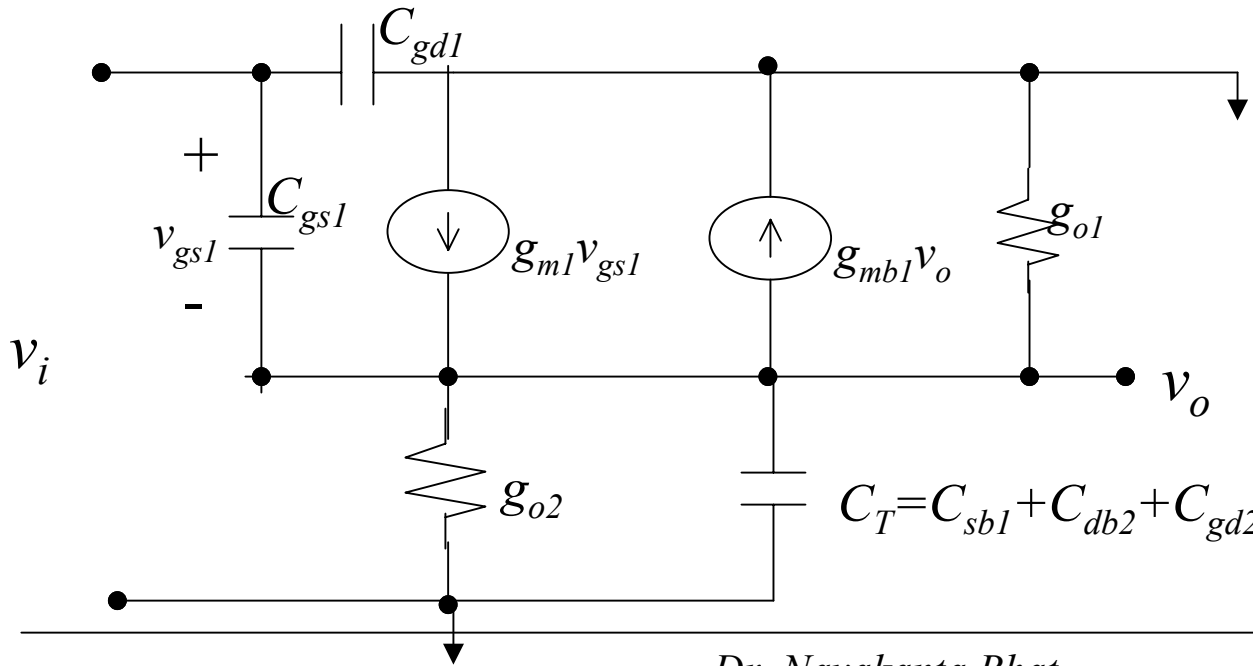


Used as buffer device or level shifter

Provides current gain

Provides low output impedance

Voltage gain is almost unity



$v_{gs1}=v_i-v_o$

g_{m2} *and* g_{mb2} *do not appear since* $v_{gs2}=v_{sb2}=0$

C_{gs2} *and* C_{sb2} *are shorted*

Low frequency gain

Applying KCL at the output node

$$(g_{mb1} + g_{o1} + g_{o2} + sC_T)v_o = g_{m1}(v_i - v_o) + sC_{gs1}(v_i - v_o)$$

$$Av(0) = \frac{g_{m1}}{g_{m1} + g_{mb1} + g_{o1} + g_{o2}}$$

$$Av(0) \approx \frac{g_{m1}}{g_{m1} + g_{mb1}} = \frac{1}{1+\eta}$$

Transistor M1 suffers from body effect, I.e. *Vt=f(Vo),* which results in significant non linearity in gain

In a twin well technology the gain can be made independent of η by connecting the body of M1 to the source of M1 and putting M1 in an isolated p-well

High frequency response

$$Av(0) = \frac{g_{m1}}{g_{m1} + g_{mb1} + g_{o1} + g_{o2}} \frac{1 + \frac{sC_{gs1}}{g_{m1}}}{1 + \frac{s(C_{gs1} + C_{sb1} + C_{gd2} + C_{db2})}{g_{m1} + g_{mb1} + g_{o1} + g_{o2}}}$$

$$z = -\frac{g_{m1}}{C_{gs1}} \qquad p = -\frac{g_{m1} + g_{mb1} + g_{o1} + g_{o2}}{C_{gs1} + C_{sb1} + C_{gd2} + C_{db2}}$$

Both pole and zero are on the left half of S plane

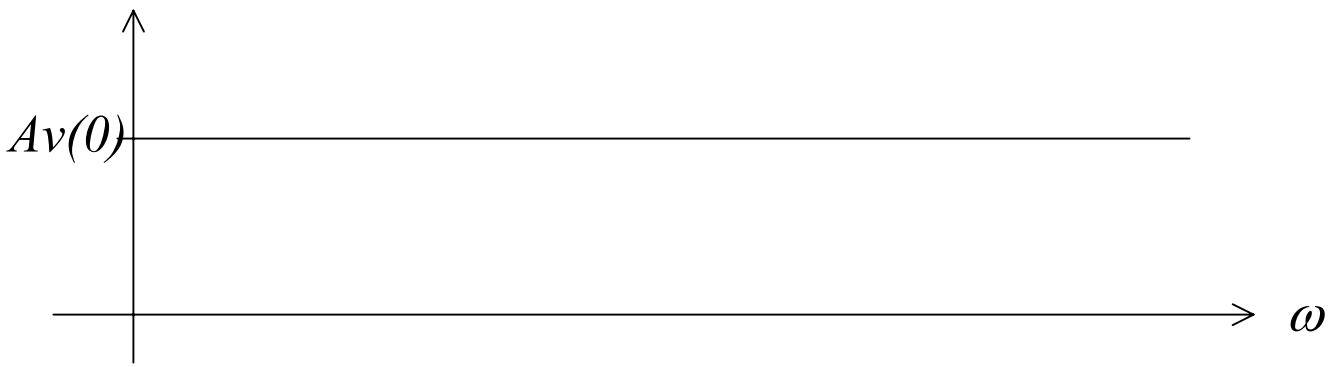
The relative location depends on the values of different parameters

Broadband condition

$$|p| = |z|$$

$$\frac{g_{m1} + g_{mb1} + g_{o1} + g_{o2}}{C_{gs1} + C_{sb1} + C_{gd2} + C_{db2}} = \frac{g_{m1}}{C_{gs1}}$$

$$\frac{C_{gs1}}{C_{gs1} + C_{sb1} + C_{gd2} + C_{db2}} = A_v(0)$$

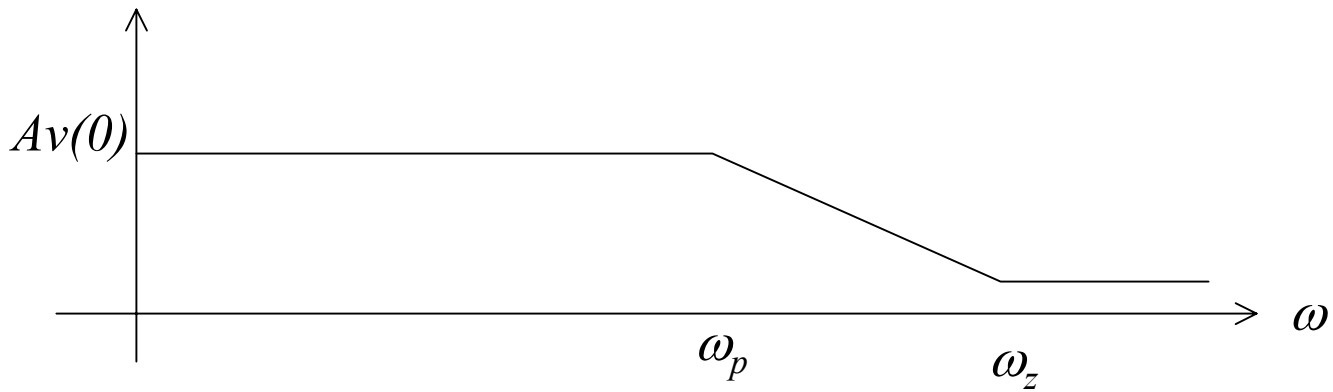


Relative location of pole and zero

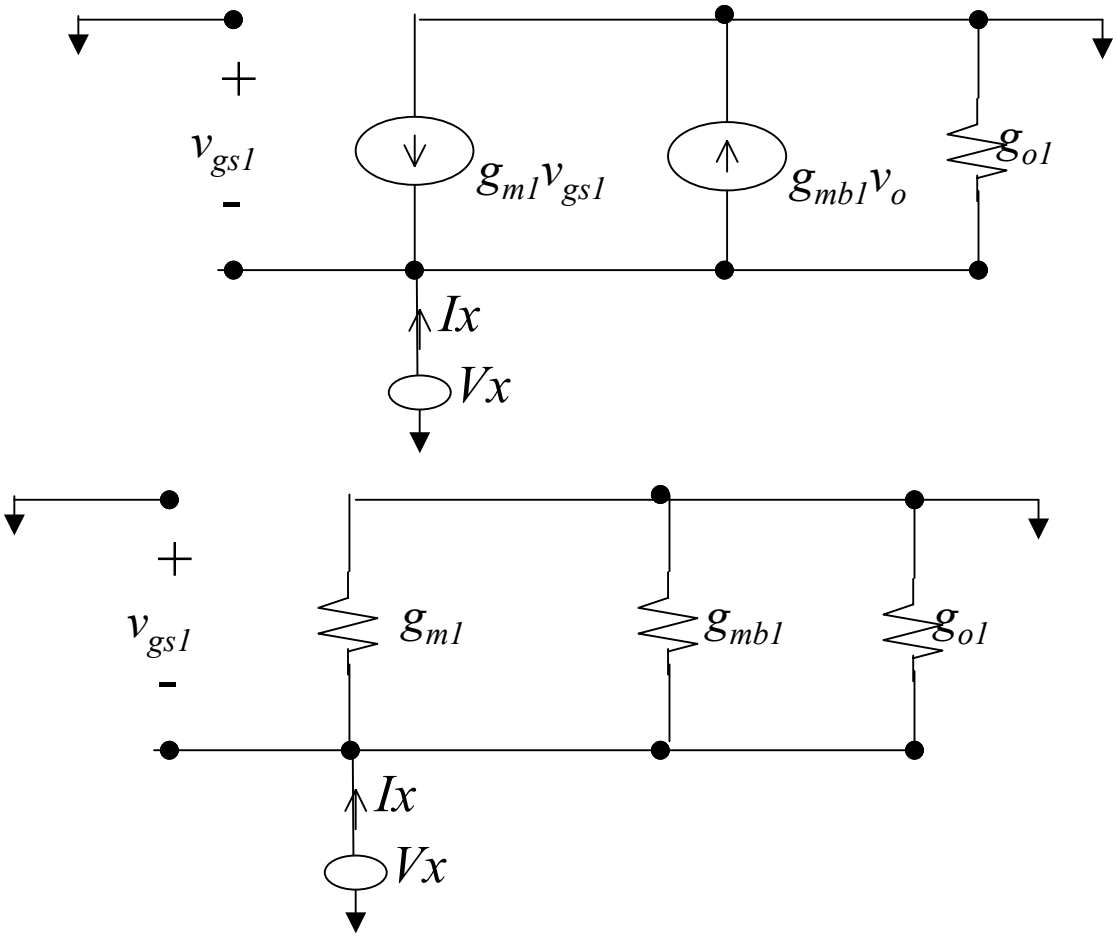
In most of the cases p<z

since

$$C_{gs1} + C_{sb1} + C_{gd2} + C_{db2} \rangle \frac{C_{gs1}}{A_v(0)}$$



Low frequency output resistance



Input source is shorted for output Resistance computation

A test voltage source Vx is applied at output node

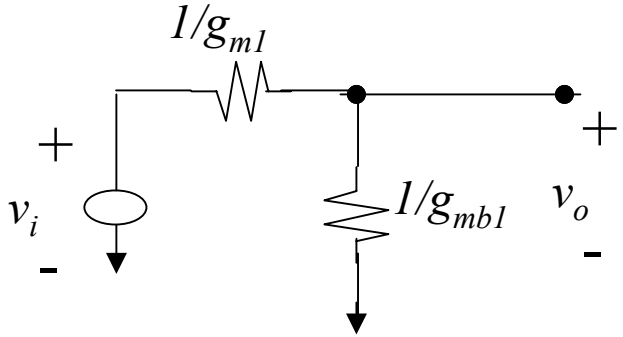
Ro=Vx/Ix

$$I_x = V_x g_{m1} + V_x g_{mb1} + V_x g_{o1}$$

$$R_o = \frac{Vx}{Ix} = \frac{1}{g_{m1} + g_{mb1} + g_{o1}}$$

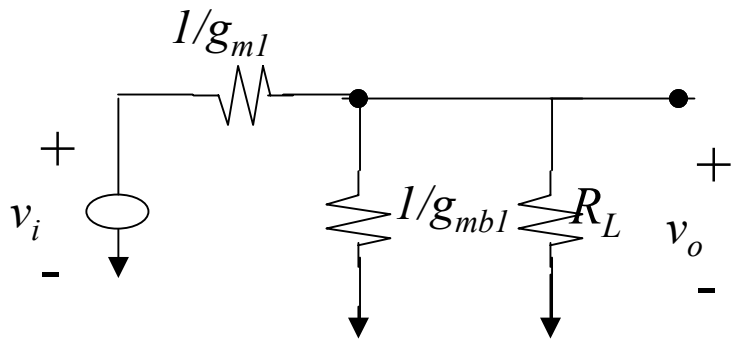
$$R_o \approx \frac{1}{g_{m1} + g_{mb1}}$$

Thevenin equivalent



$$A_v = \frac{v_o}{v_i} = \frac{1/g_{mb1}}{1/g_{m1} + 1/g_{mb1}} = \frac{g_m}{g_m + g_{mb}}$$

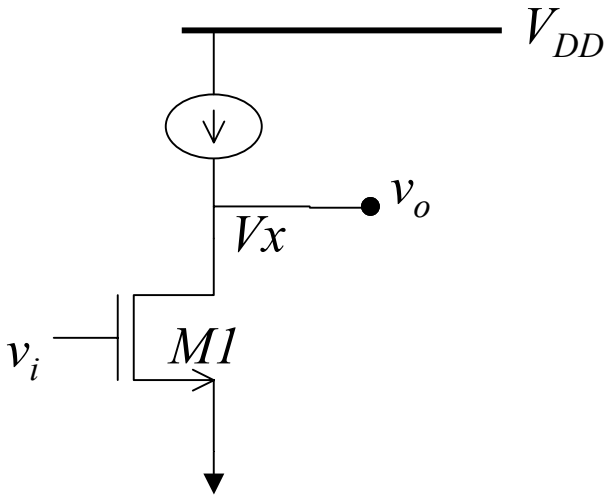
If the source follower is loaded by R_L such that $R_L < 1/g_{mb1}$



$$A_v = \frac{R_L}{1/g_{m1} + R_L}$$

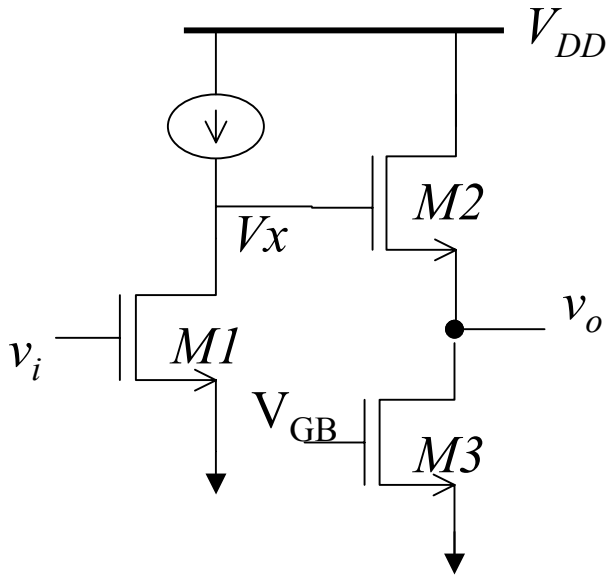
Limitation of SF on previous stage

CS with current source load



$$Vx_{min}=V_i-V_t$$

CS with current source load, driving SF

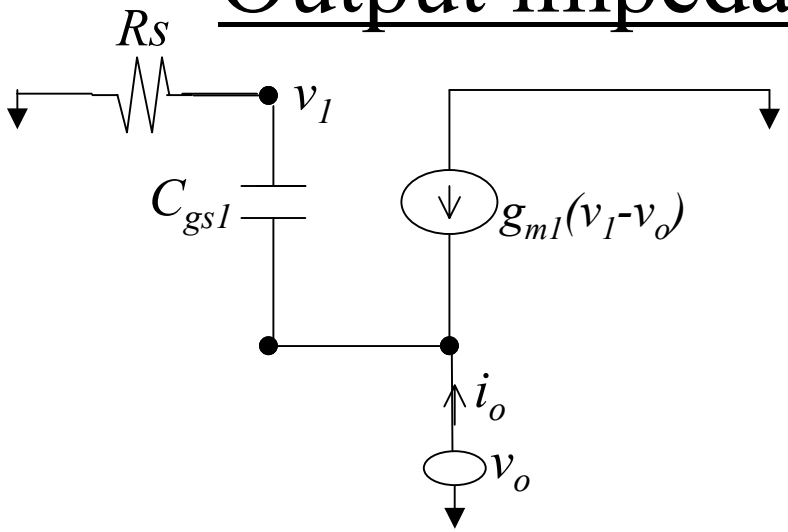


$$Vx_{min}=V_{gs2}+V_{gs3}-V_t$$

Otherwise M3 comes out of saturation

The voltage swing at node X is reduced

Output impedance in presence of *Rs*



Assumptions:

$g_{mb1}+g_{o1}+g_{o2} << g_{m1}$

Rs is small compared to $1/sC_{gd1}$

Output shunt capacitance excluded

Applying KCL at output node

$$i_o + g_{m1}(v_1 - v_o) + sC_{gs1}(v_1 - v_o) = 0$$

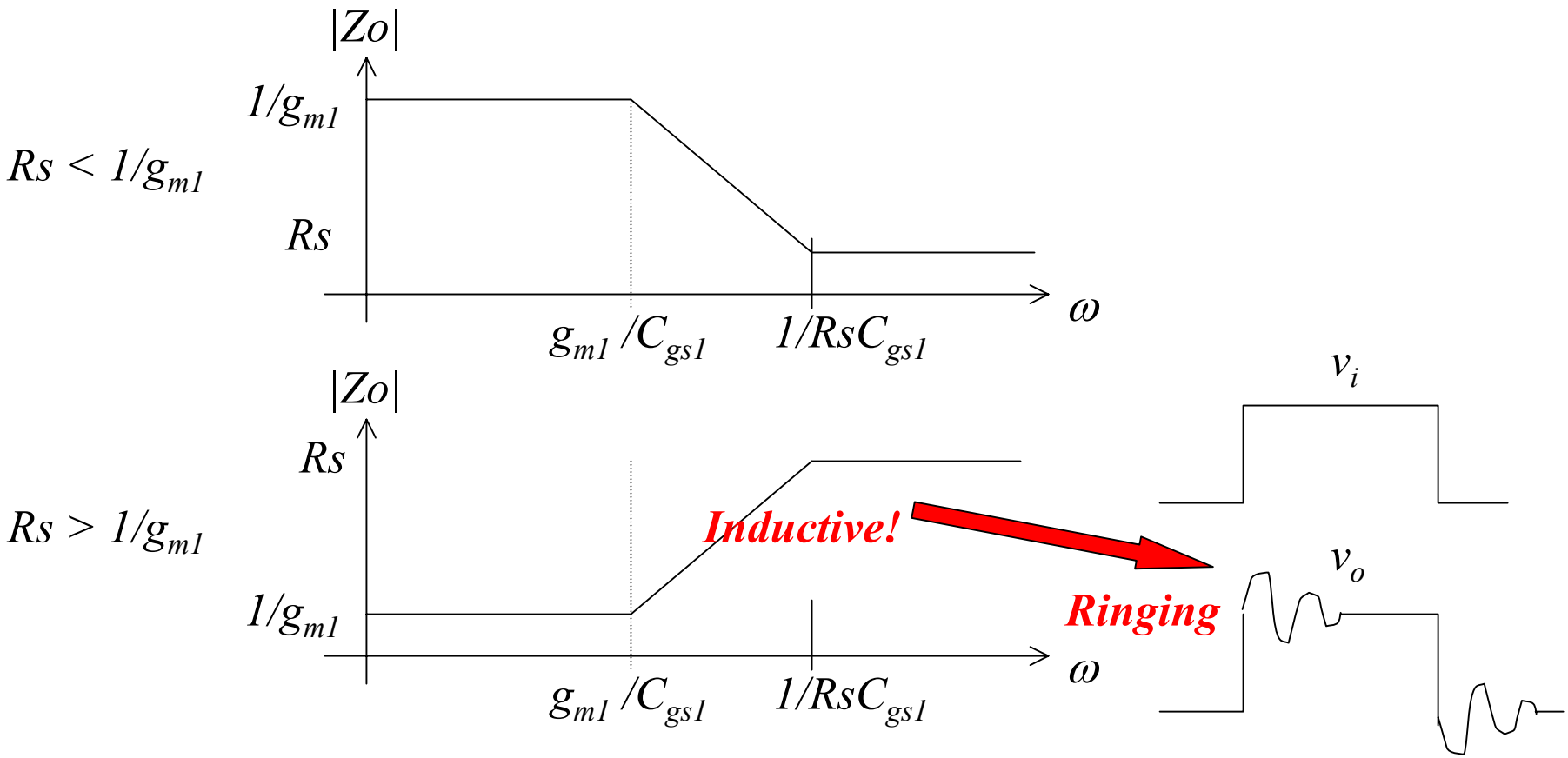
using $$\frac{v_1}{v_o} = \frac{R_s}{R_s + 1/sC_{gs1}}$$

$$Z_o = \frac{v_o}{i_o} = \frac{1}{g_{m1}} \frac{1 + sR_sC_{gs1}}{1 + s\,C_{gs1}/g_{m1}}$$

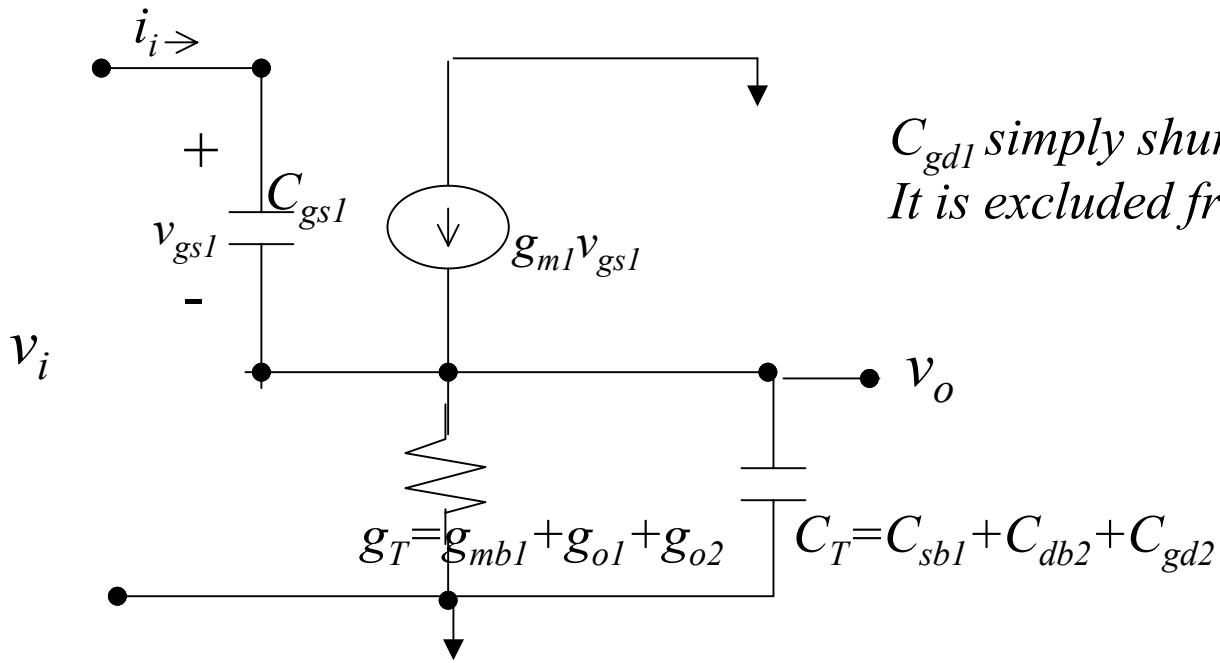
Zo(0)=1/g_{m1} *and Zo(∞)=Rs*

Variation of Z_o with ω

$$Z_o = \frac{v_o}{i_o} = \frac{1}{g_{m1}} \frac{1 + sR_s C_{gs1}}{1 + s \dfrac{C_{gs1}}{g_{m1}}}$$



Input Impedance *Zin*



C_{gd1} simply shunts the input and hence It is excluded from equivalent circuit

$$v_i = \frac{i_i}{sC_{gs1}} + \left(i_i + \frac{g_{m1} i_i}{sC_{gs1}} \right) \frac{1}{(g_T + sC_T)}$$

$$Z_{in} = \frac{v_i}{i_i} = \frac{1}{sC_{gs1}} + \left(1 + \frac{g_{m1}}{sC_{gs1}} \right) \frac{1}{(g_T + sC_T)}$$

Low Frequency dependence of Zi

For relatively low frequency, $sC_T << g_T$

$$Z_{in} = \frac{v_i}{i_i} = \frac{1}{sC_{gs1}} + \left(1 + \frac{g_{m1}}{sC_{gs1}}\right)\frac{1}{g_T}$$

$$Z_{in} = \frac{1}{g_T} + \frac{1}{sC_{gs1}}\left(1 + \frac{g_{m1}}{g_T}\right)$$

$$Z_{in} \approx \frac{1}{g_{mb1}} + \frac{1}{sC_{gs1}}\left(\frac{g_{m1} + g_{mb1}}{g_{mb1}}\right)$$

i.e. a fraction of C_{gs1} is felt at the input along with $1/g_{mb1}$

High Frequency dependence of Zi

For relatively low frequency, $sC_T >> g_T$

$$Z_{in} = \frac{v_i}{i_i} = \frac{1}{sC_{gs1}} + \left(1 + \frac{g_{m1}}{sC_{gs1}}\right)\frac{1}{sC_T}$$

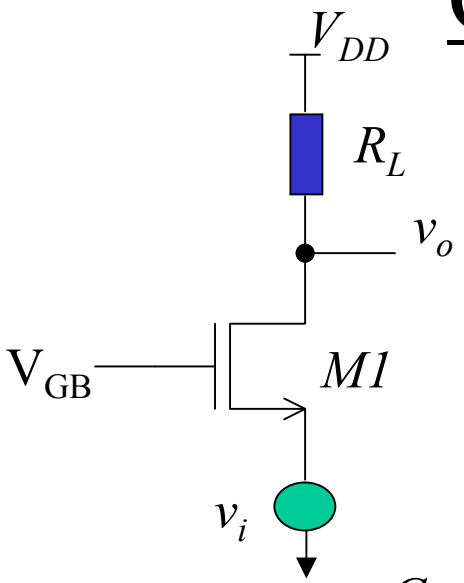
$$Z_{in} = \frac{1}{sC_{gs1}} + \frac{1}{sC_T} - \frac{g_{m1}}{\omega^2 C_{gs1} C_T}$$

Negative Resistance!

At the input a series combination of C_{gs1}, C_T and a negative resistance is seen

This negative resistance could be exploited to build oscillators

Common Gate Amplifier



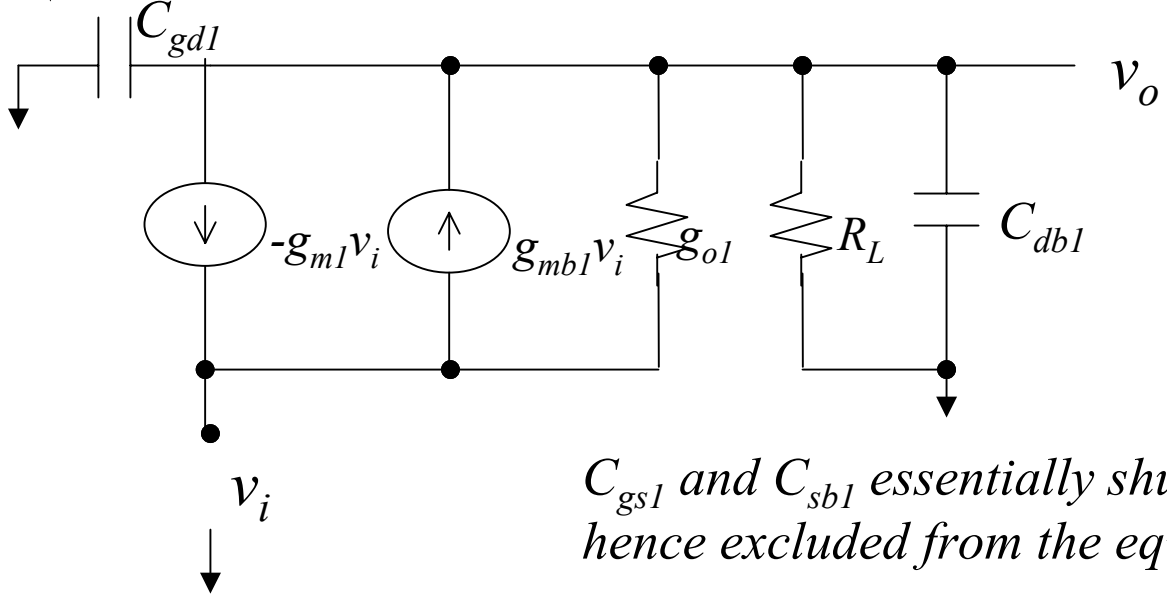
Provides large voltage gain

Current gain is about unity

Provides low input impedance

Provides high output impedance

Often used for impedance transformation



C_{gs1} and C_{sb1} essentially shunt input and hence excluded from the equivalent ckt

Gain Expression

Applying KCL at the output node

$$(g_{m1} + g_{mb1})v_i + (v_i - v_o)g_{o1} = (G_L + sC_{db1} + sC_{gd1})v_o$$

$$Av = \frac{vo}{vi} = \frac{g_{m1} + g_{mb1} + g_{o1}}{g_{o1} + G_L} \frac{1}{1 + \frac{s(C_{db1} + C_{gd1})}{g_{o1} + G_L}}$$

$$A_v(0) = \frac{g_{m1} + g_{mb1} + g_{o1}}{g_{o1} + G_L}$$

$$p = -\frac{g_{o1} + G_L}{C_{db1} + C_{gd1}}$$

The transfer function has single pole and no zero

Low frequency Input impedance

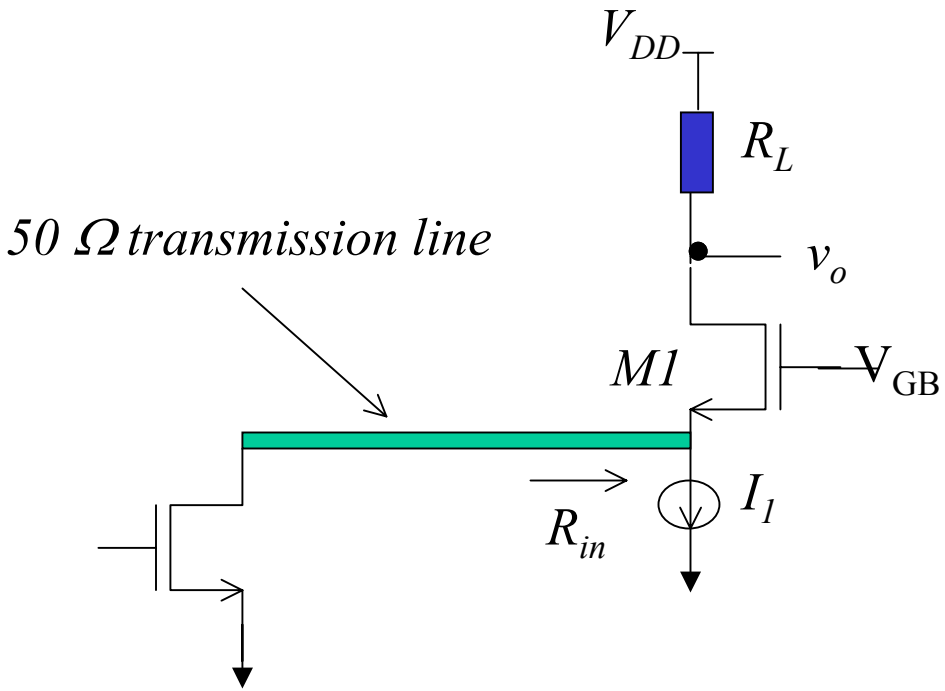
$$Z_{in} = \frac{v_i}{i_i} = \frac{1 + g_{o1} R_L}{g_{m1} + g_{mb1} + g_{o1}}$$

$$Z_{in} = \frac{1 + R_L / r_{o1}}{g_{m1} + g_{mb1} + 1 / r_{o1}}$$

If $r_{o1} >> R_L$ *and* $1/r_{o1} << g_{m1} + g_{mb1}$, *then*

$$Z_{in} = \frac{1}{g_{m1} + g_{mb1}}$$

Impedance Transformation

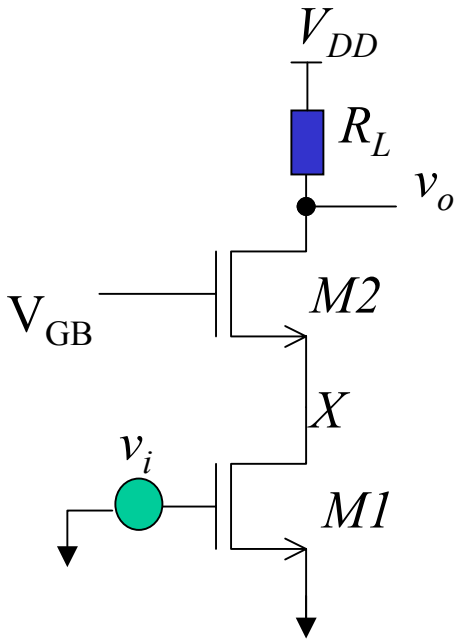


The impedance R_L can be transformed into 50 Ω at the input by making $1/(g_{m1}+g_{mb1})=50\ \Omega$

Cascode Amplifier

The term “cascode” is believed to be abbreviation of “cascaded triodes”

Cascode is a combination of common source and common gate stage



M1 : Input Device

M2 : Cascode Device

Features of Cascode Amplifier

Output impedance increases

Intrinsic gain is squared

Shielding property :

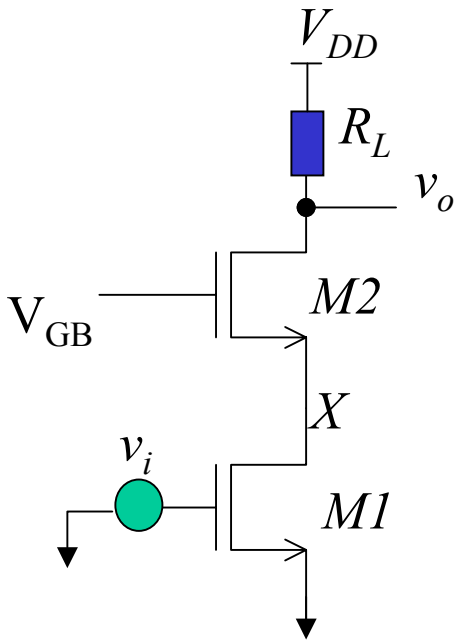
Node X is desensitized w.r.t. o/p

Input pole, in presence of *Rs,* is pushed away

Output swing impacted due to stacking

$Vo_{min} = V_{ds2} + V_{ds1}$

Low frequency gain for low R_L



$R_L << r_{o2eff}$

The current flowing through the node X due to an input voltage v_i is $i_x = g_{m1} v_i$

The current through output node is same as i_x

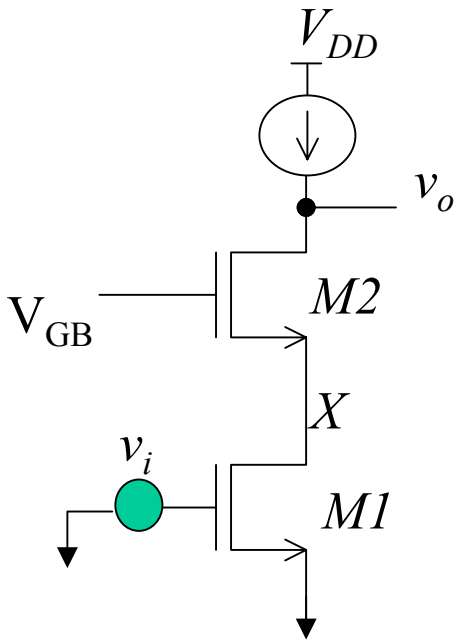
$i_o = i_x \qquad v_o = i_x R_L$

$v_o = g_{m1} v_i R_L$

$Av(0) = g_{m1} R_L$

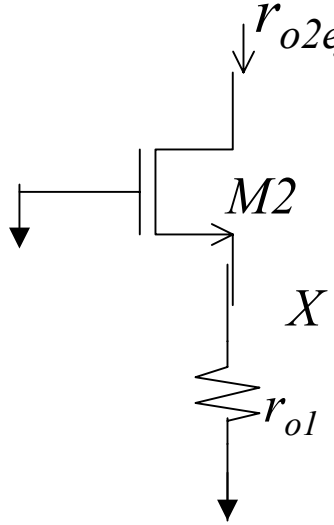
This result is identical to common source stage

Low frequency gain for larger R_L



For an ideal current source load, the gain is dependent on the output resistance seen looking into drain of M2

For computation of r_{o2}, M2 can be viewed as common Source stage with source degeneration of ro1



$$r_{o2eff} = (g_{m2} r_{o1}) r_{o2}$$

The output impedance enhanced!

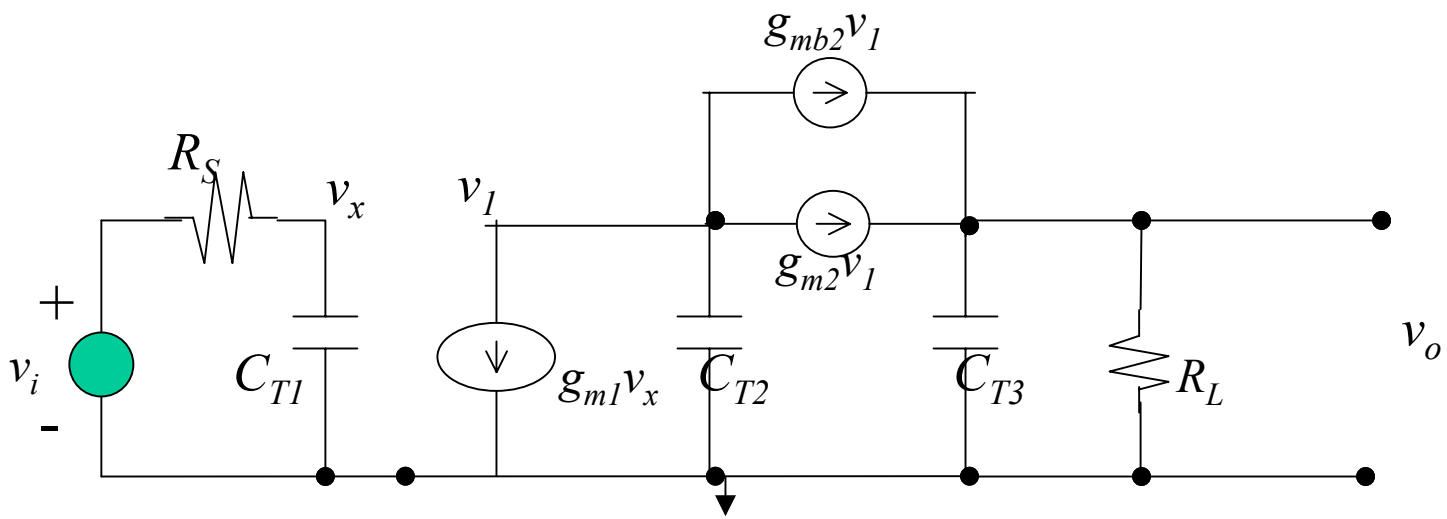
$$A_v(0) = g_{m1} g_{m2} r_{o1} r_{o2}$$

The intrinsic gain of cascode is the square of the CS stage

The increased output impedance is exploited in current mirror design

Input pole desensitisation

The small signal equivalent, neglecting g_{o1}, g_{o2} and neglecting zero due to C_{gd1}



$C_{T1}=C_{gs1}+C_{gd1}(1+v_1/v_x)$, *using Miller's theorem*

$C_{T2}=C_{gs2}+C_{gd1}+C_{db1}+C_{sb2}$

$C_{T3}=C_{db2}+C_{gd2}+C_L$

Gain Expression

$$Av = -g_{m1}R_L \frac{1}{1+sR_sC_{T1}} \frac{1}{1+\dfrac{sC_{T2}}{(g_{m2}+g_{mb2})}} \frac{1}{1+sR_LC_{T3}}$$

$$A_v(0) = -g_{m1}R_L$$

$$p_1 = -\frac{1}{R_sC_{T1}}$$

$$p_2 = -\frac{g_{m2}+g_{mb2}}{C_{T2}}$$

$$p_3 = -\frac{1}{R_LC_{T3}}$$

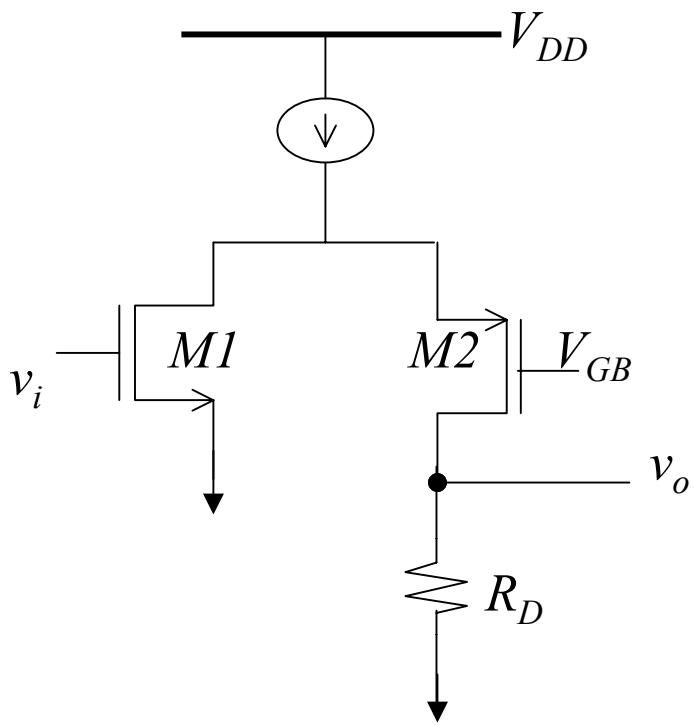
$C_{T1}=C_{gs1}+2C_{gd1}$ since $v_1/v_x=1$,
i.e. Miller capacitance at I/p is drastically reduced

Folded Cascode

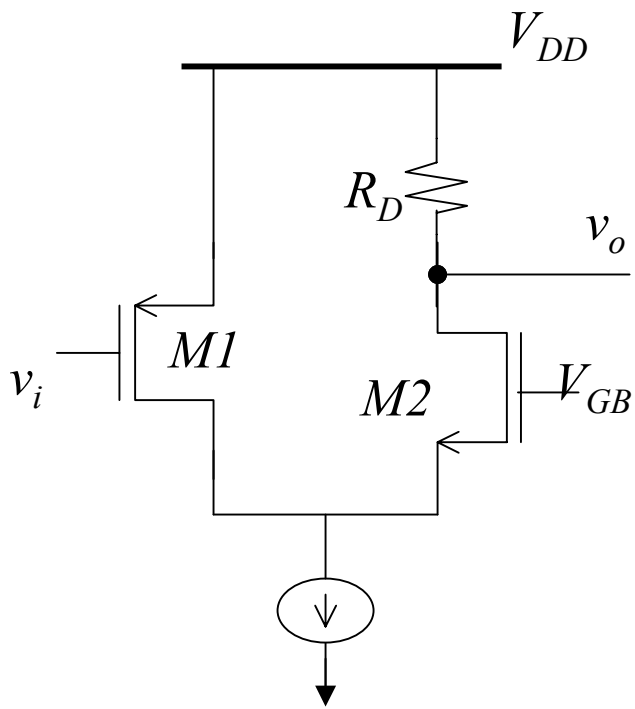
Input and Cascode devices are complementary

Hence the current is either folded up or down

Avoids stacking of transistors



Current is folded down



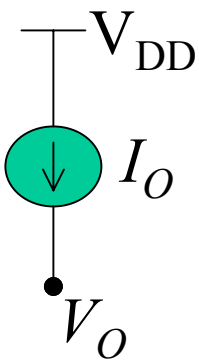
Current is folded up

Current Mirror

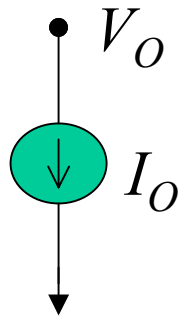
Basic configuration
Cascode Current Mirror
Wilson Current Mirror
Regulated Cascode

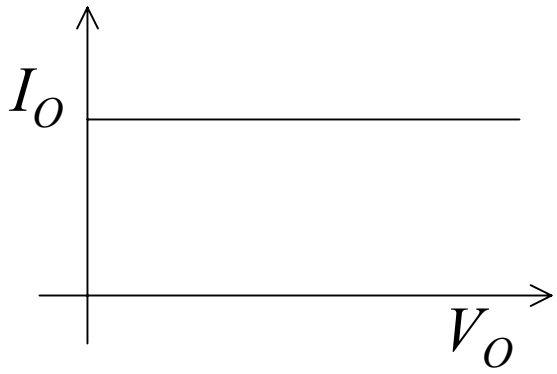
Basic Definitions

Current Source

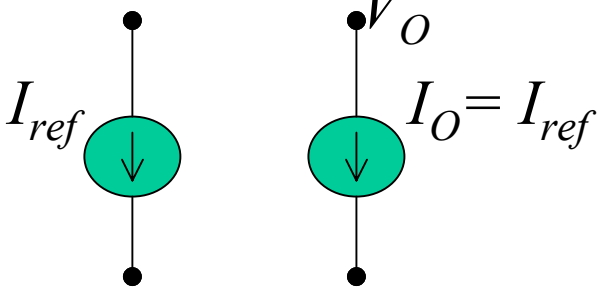


Current Sink

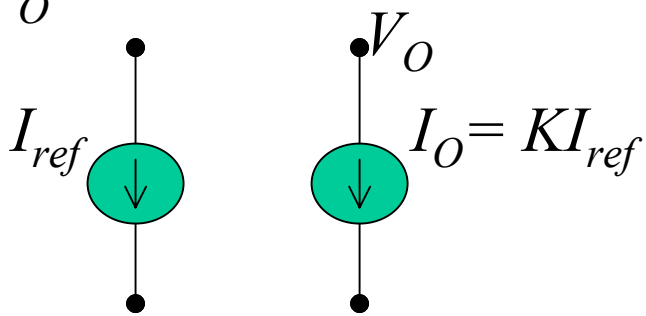




Current Mirror



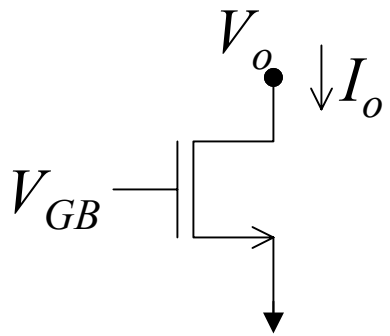
Current Lens



Often times, the term current mirror is used to include current lens

Simplest current source/sink

NMOS/PMOS transistor in saturation

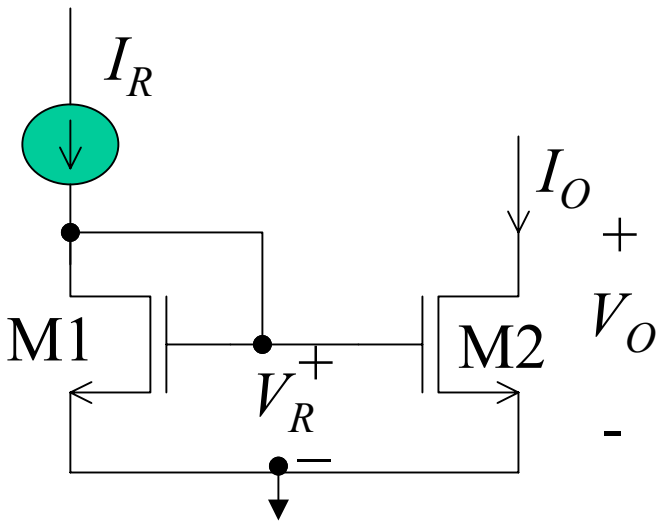


$$I_o = \frac{\mu \varepsilon_{ox} W}{T_{ox} L} \frac{(V_{GB} - Vt)^2}{2} (1 + \lambda V_o)$$

- Sensitive to variation in V_o (I.e.*Ro* is not infinity)
- Sensitive to variation in V_{GB}
- Sensitive to temperature variation (V_t, μ_n, μ_p)
- Sensitive to process variation (V_t, W, L, T_{ox})

Strategy: Create one very well defined current reference using complex temperature compensation and V_{DD} insensitivity. Then use current mirrors (copy) to generate others

Basic Current Mirror



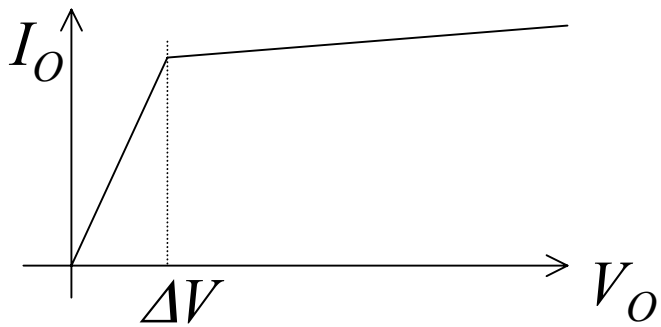
M1 is diode connected and is always in saturation

I_R sets a unique bias voltage V_R

$$I_R = \frac{\mu \varepsilon_{ox} W_1}{T_{ox} L_1} \frac{(V_R - V_t)^2}{2} (1 + \lambda V_R)$$

M2 will mirror this current provided $V_O > V_R - V_t$

Since $V_R = V_t + \Delta V$ where ΔV *is gate over drive ;* $V_O > \Delta V$



Current ratio

Neglecting channel length modulation effect,

$$I_O = \frac{(W/L)_2}{(W/L)_1} I_R$$

However it is advisable to choose constant L for
Both M1 and M2 and ratio based on Ws only

$$L = L_g - 2L_D$$

L_D is constant for all $L_g \Rightarrow$ Ratio in L_g does not translate to ratio in L

$$I_O = \frac{W_2}{W_1} I_R$$

Channel width effect

Strictly speaking

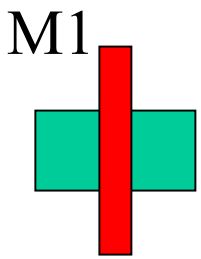
$$W = W_g - 2\Delta W$$

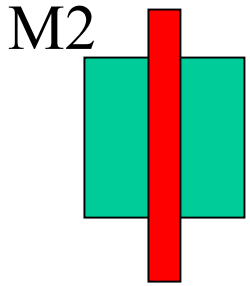
where ΔW is due to field oxide encroachment

Ratio in W_g does not translate to ratio in W

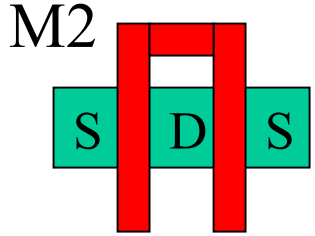
Parallel transistor layout can be used to overcome this problem

Ex: $I_o = 2I_R$, then $W_2 = 2W_1$





$$I_o = \frac{2W_{g1} - 2\Delta W}{W_{g1} - 2\Delta W} I_R$$



$$I_o = \frac{2W_{g1} - 4\Delta W}{W_{g1} - 2\Delta W} I_R = 2I_R$$

Issues with basic current mirror

Vds effect results in incorrect mirroring

$$\frac{I_O}{I_R} = \frac{W_2}{W_1}\frac{1+\lambda V_o}{1+\lambda V_R}$$ For $L_1=L_2$, $\lambda_1=\lambda_2=\lambda$

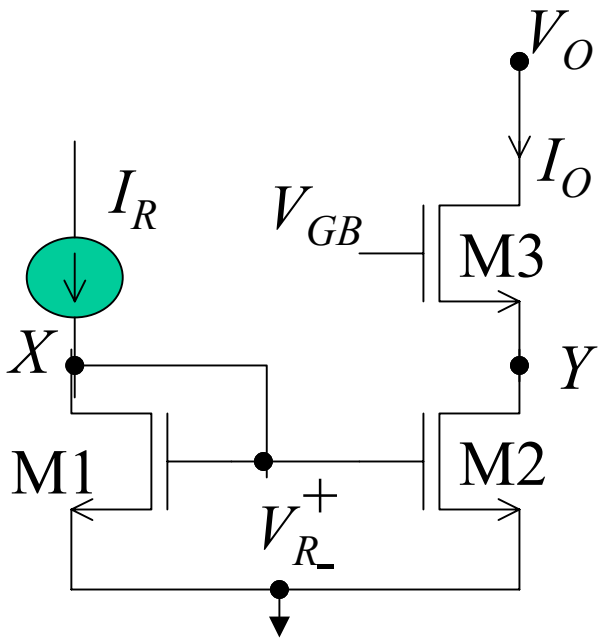
For $\lambda V_R << 1$, *and neglecting* λ^2 *term:*

$$\frac{I_O}{I_R} = \frac{W_2}{W_1}\left[1+\lambda(V_o - V_R)\right]$$

The output resistance is finite

$R_o = r_{o2}$

Cascode current mirror



M3 is in common gate configuration

Hence M2 and M3 form cascode pair

M3 ***shields*** node Y from variations in V_o

$$\frac{\Delta v_o}{\Delta v_y} = \frac{g_{m3} + g_{mb3}}{g_{o3}}$$

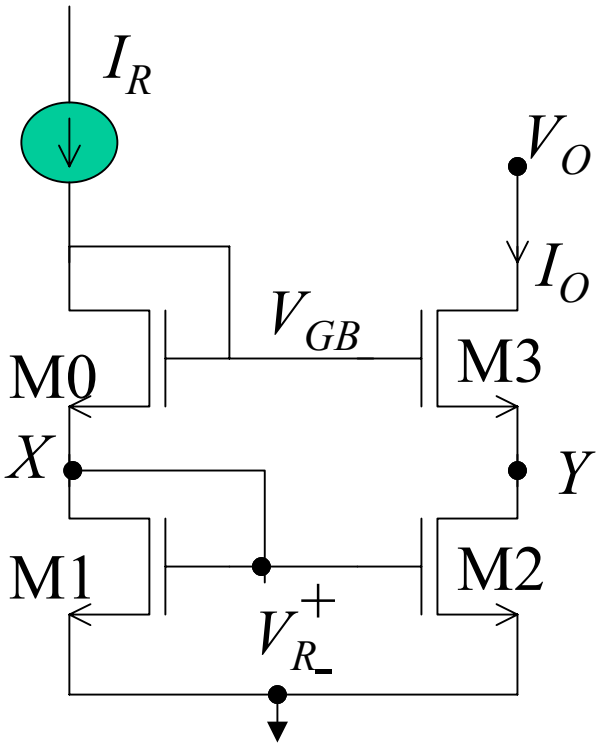
$$\Delta v_y = \frac{\Delta v_o}{A_v(0)}$$

In order for I_o=I_R, V_{GB} should be chosen such that Vx=Vy

$$V_{GB} = Vx + V_{GS3}$$

This is achieved by introducing another diode connected transistor in series with M1

Cascode current mirror with matching



$$V_{GB} = V_x + V_{gs0}$$

For $V_{gs0}=V_{gs3}$, we need

$$\frac{(W/L)_3}{(W/L)_0} = \frac{(W/L)_2}{(W/L)_1}$$

Then $V_x=V_y$ and $I_O=I_R W_2/W_1$

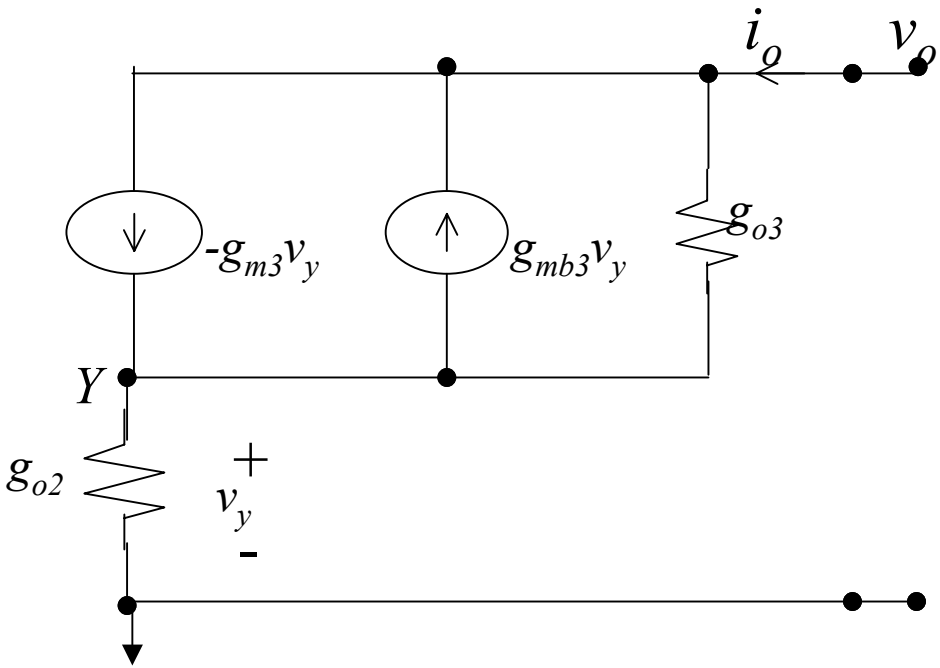
The minimum allowed Vo:

$V_{GB}= V_{gs0}+V_{gs1} = 2\Delta V+2V_t$
assuming similar overdrive and Vt

$\therefore V_{Omin} = 2\Delta V+V_t$

Beyond V_{Omin}, M3 comes out of saturation

Output resistance of cascode mirror



Applying KCL at output node

$$i_o = -(g_{m3} + g_{mb3})v_y + g_{o3}(v_o - v_y)$$

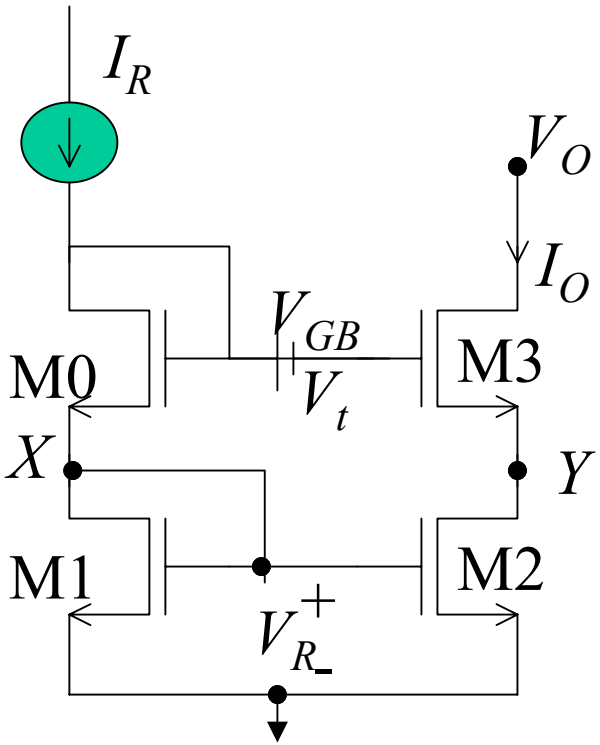
$$i_o = v_y g_{o2}$$

$$R_o = v_o / i_o$$

$$R_o = r_{o3} \bullet g_{m3} r_{o2}(1+\eta) \approx r_{o3} \bullet g_{m3} r_{o2}$$

The output resistance of M3 (r_{o3}) is enhanced by a factor $g_{m3}r_{o2}$

Cascode mirror with improved o/p swing



The voltage at G_o is $V_{G0}=2\Delta V+2Vt$

Hence the voltage at G3 is

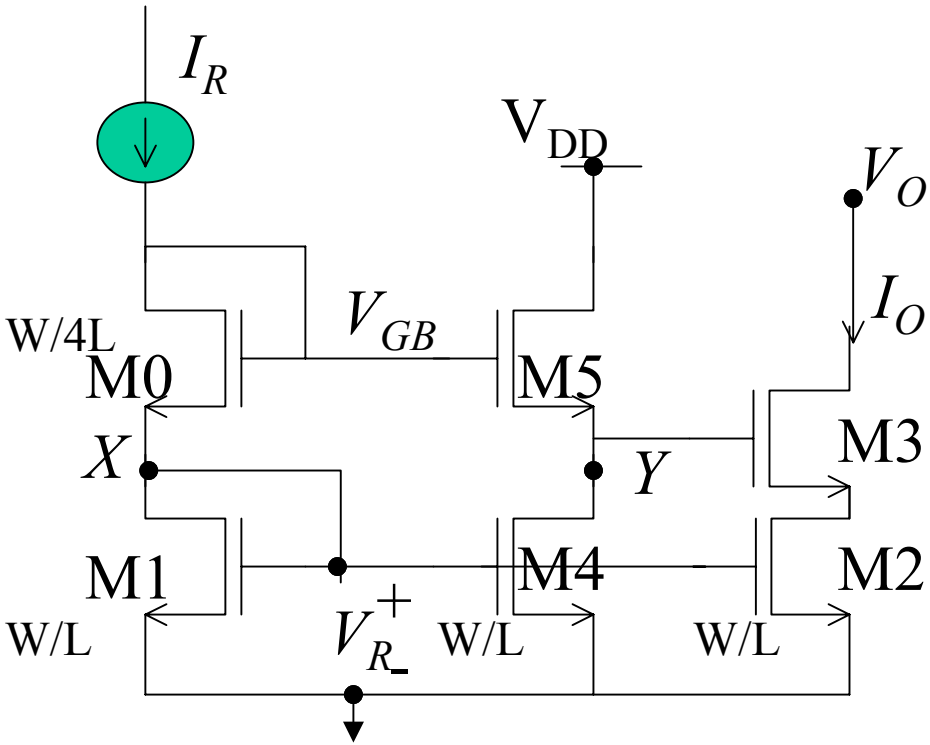
$V_{G3}=2\Delta V+Vt$

$\therefore V_{Omin}=2\Delta V$

Although output swing is increased, it should be noted that Vx≠Vy
Hence the improvement has come at the expense of current matching

Cascode mirror with improved o/p swing

Vt can be implemented using following configuration



The size of M0 is 4 times smaller than that of M1

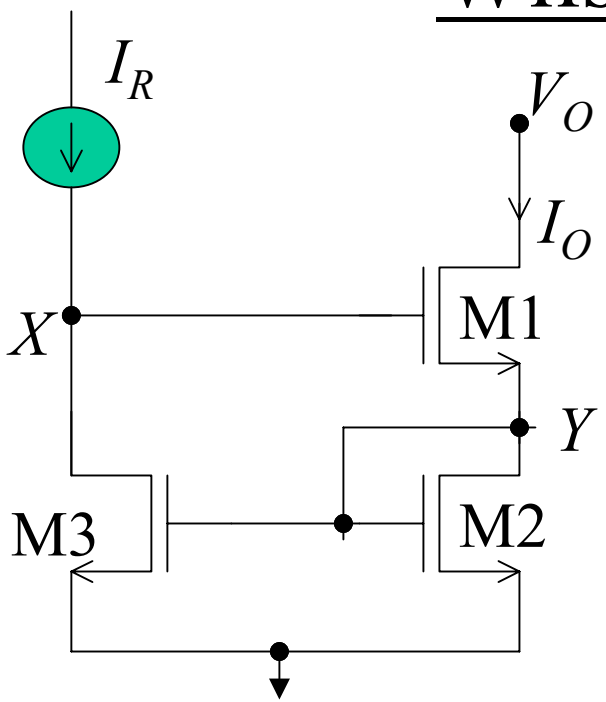
For the currents to be identical the gate overdrive of M0 be twice that of M1 (2ΔV and ΔV)

$$V_{gs0}=3\Delta V+2V_t$$

$$V_{gs3}=2\Delta V+V_t$$

$$V_{Omin}=2\Delta V$$

Wilson Current source



Negative feedback arrangement through M1, M2 and M3.

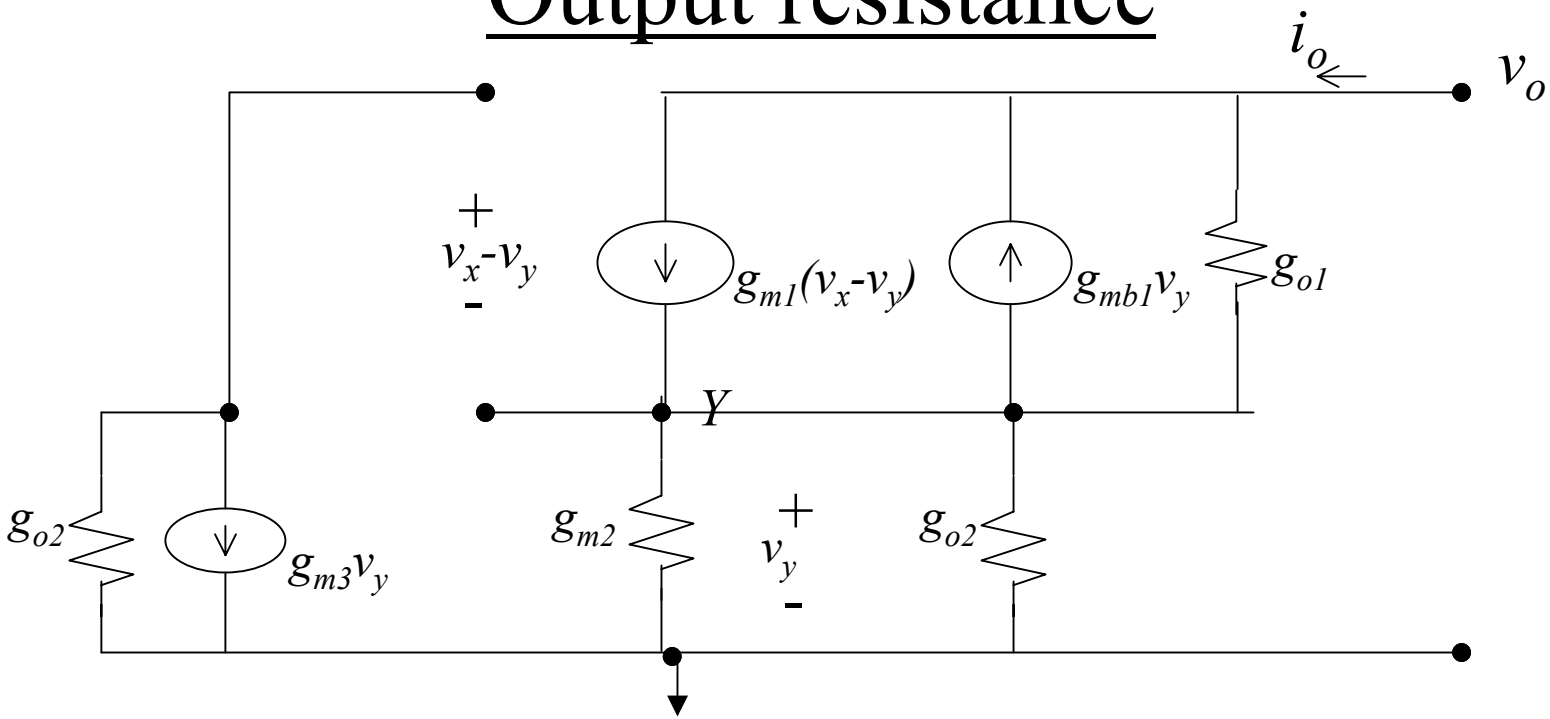
- If *Vo* increase, then Io tends to increase *Id2* and hence *Vy* increase
- Since I_R is constant, *Vx* decreases thus decreasing the gate drive for M1
- This will restore Io to its initial value

i.e. any change in *Vo* is absorbed as an appropriate change in *Vx*

$$\frac{I_O}{I_R} = \frac{W_2}{W_3}\left[1 + \lambda_2 V_x - \lambda_3 V_y)\right]$$

$$V_{omin} = 2\Delta V + Vt$$

Output resistance

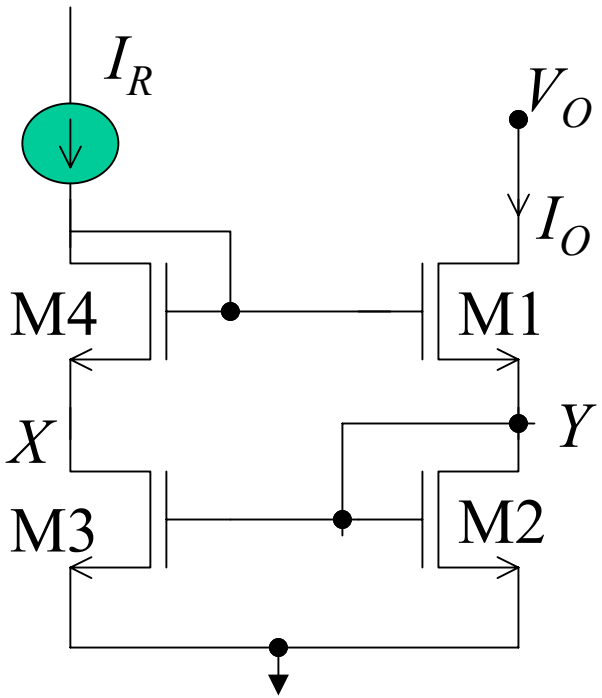


$$R_o = v_o / i_o$$

$$R_o = r_{o1}\left[1 + \frac{g_{m1}g_{m3}}{g_{m2}g_{o3}}\right]$$

$$R_o \approx r_{o1}(g_{m1}r_{o3})$$

Modified Wilson



The current matching is improved

$Vx=Vy$ can be ensured if

$$\frac{(W/L)_1}{(W/L)_4}=\frac{W_2}{W_3}$$

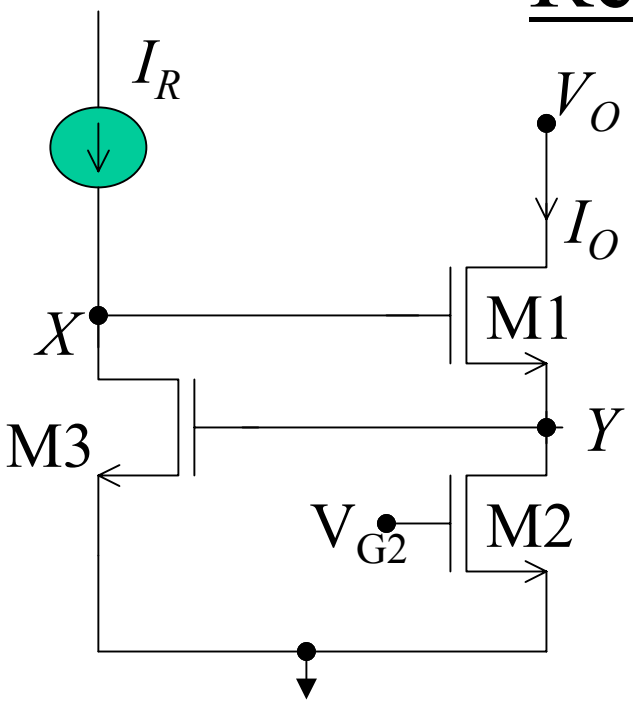
Under this condition,

$I_O=I_R W_2/W_3$

The output resistance is similar to Wilson

$Ro=r_{o1}(g_{m1}r_{o3})$

Regulated Cascode



The gate of M3 is connected to drain of M2

The gate of M2 is connected to fixed V_{G2}

The negative feedback is provided through M1 and M3

Change in *Vo* is absorbed at *Vx*

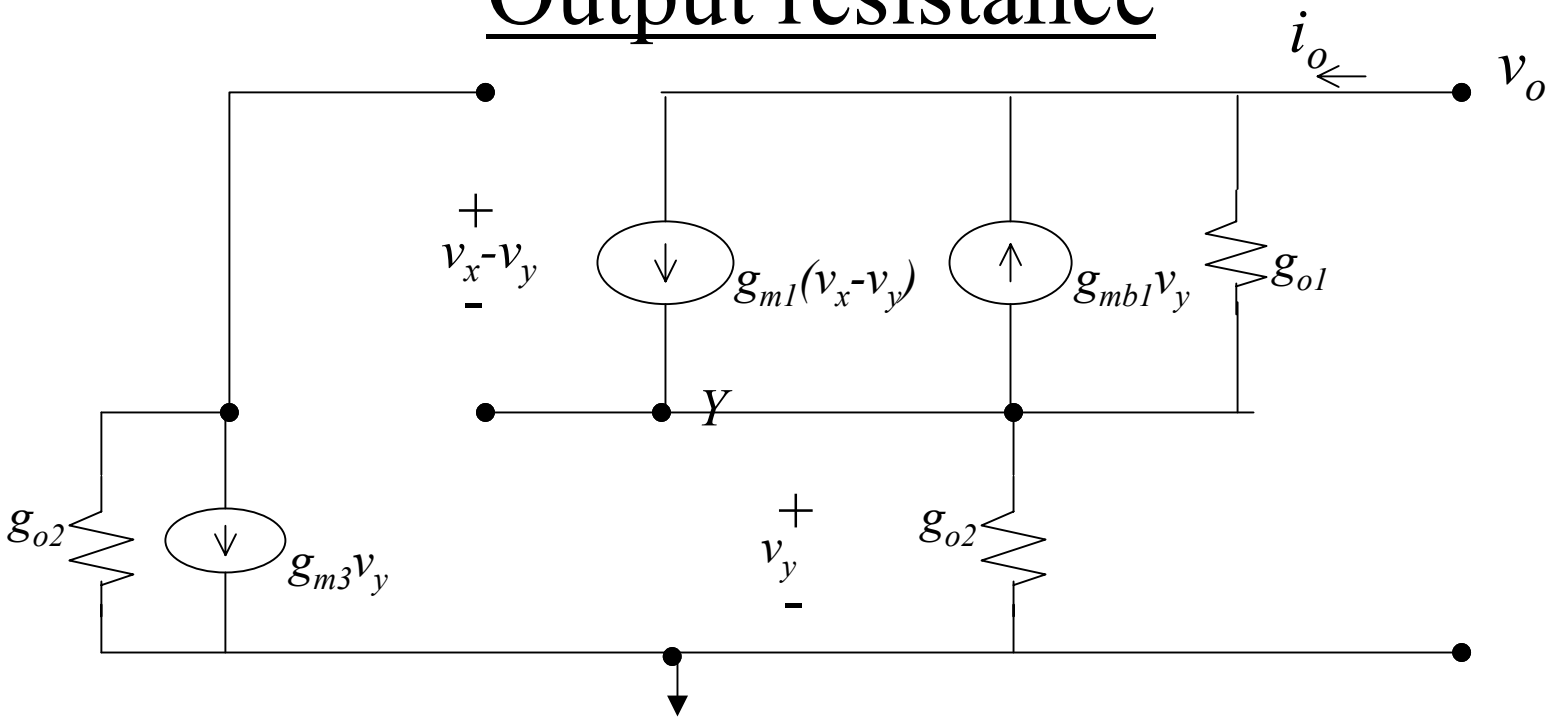
No explicit conventional mirror connection of transistors
The mirroring is entirely due to the negative feedback

The output resistance is enhanced significantly

The minimum allowed output voltage is lowered $V_{omin}=2\Delta V$

*The circuit works reasonably well even if V_{omin} drops to ΔV

Output resistance

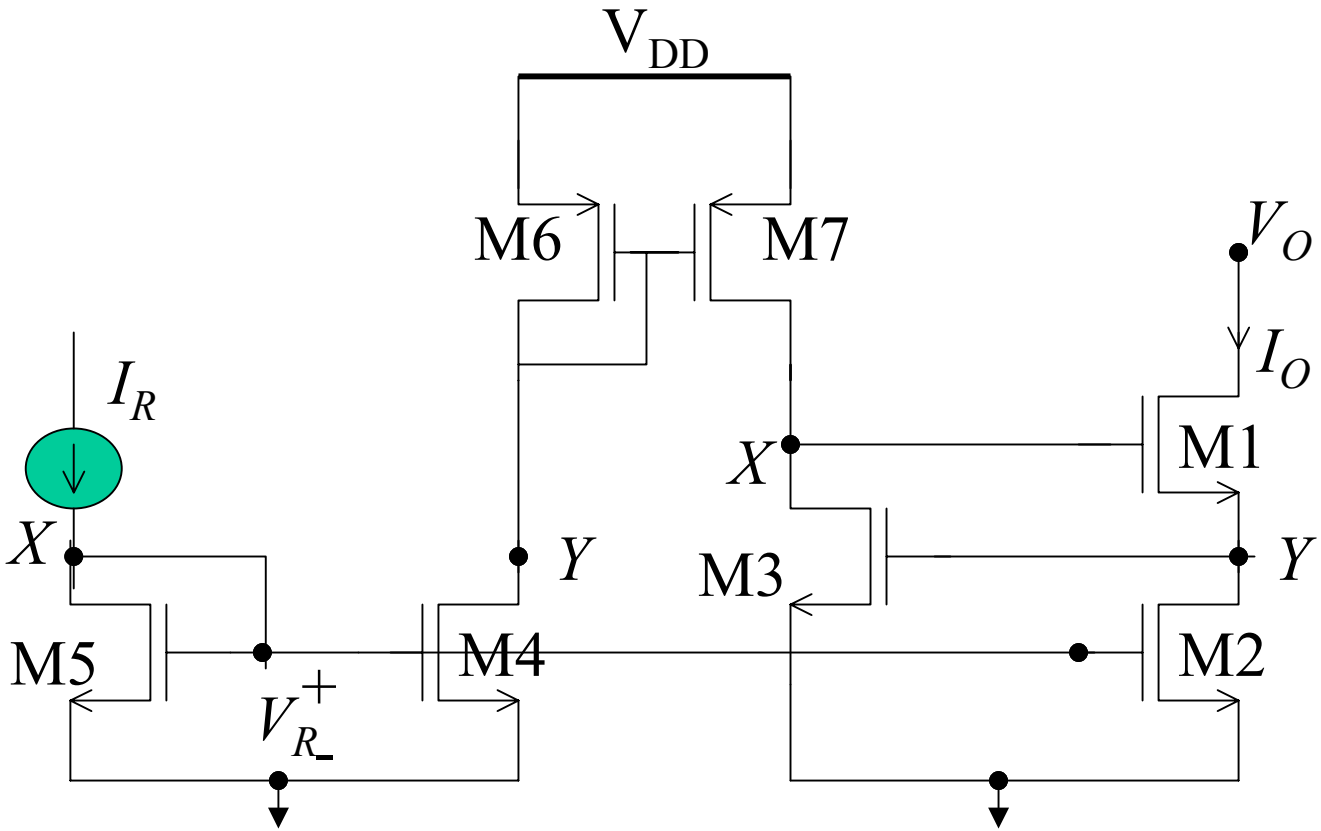


$$R_o = r_{o1}\left[1+\frac{g_{m1}g_{m3}}{g_{o2}g_{o3}}\right]$$

$$R_o \approx r_{o1}(g_m r_o)^2$$

Very high output impedance

Regulated cascode with bias generation



I_R generates fixed bias for VG2

I_R is mirrored on to M3 through the NMOS (M4-M5) and the PMOS (M6-M7) current mirrors

Layout Issues

Orientation

Symmetry

Adding dummy layers

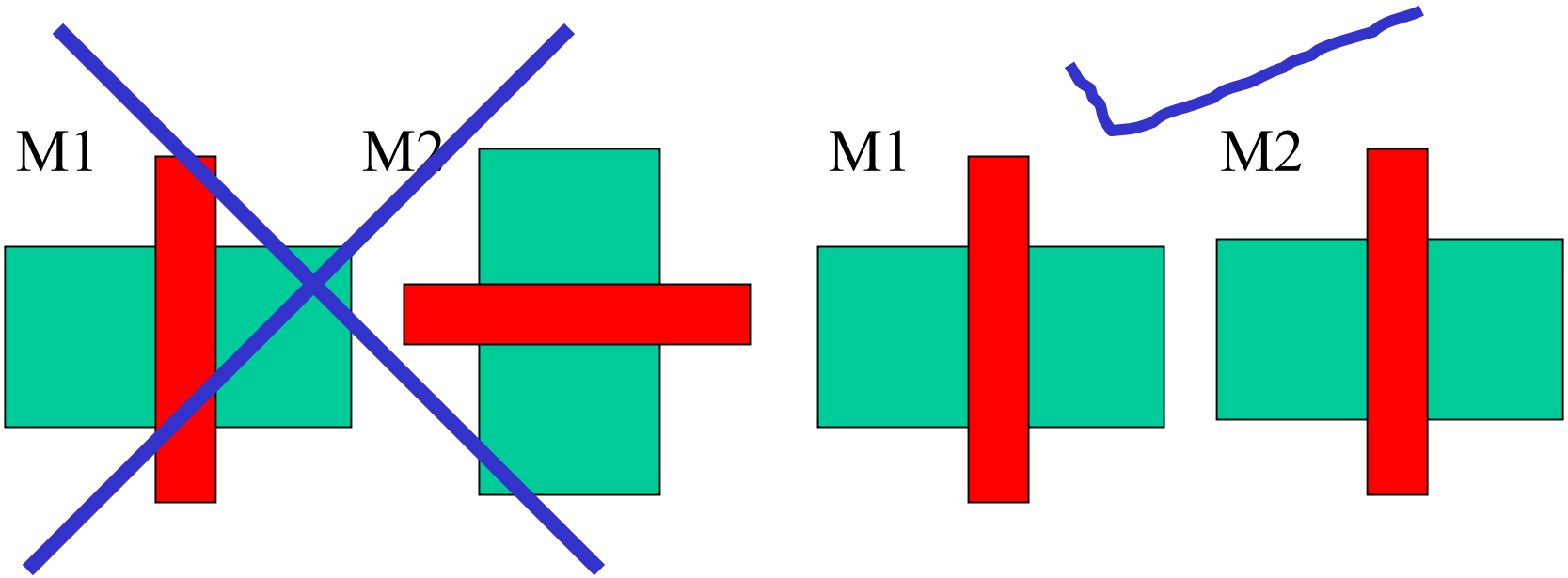
Unit cell repetition

Common centroid

Avoiding interconnect resistance

Orientation

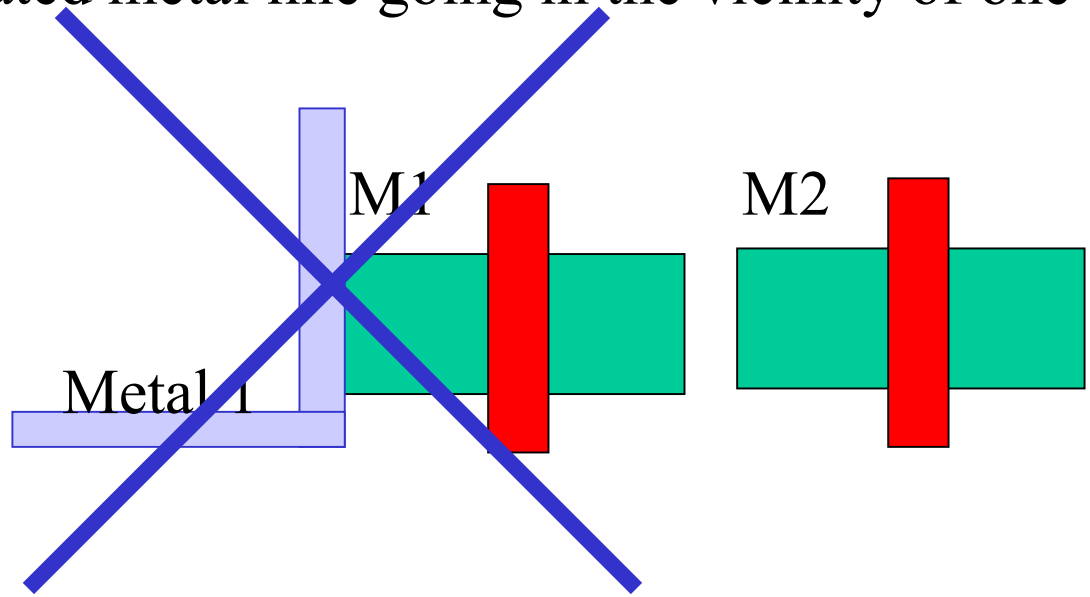
Matched transistors should be oriented in same direction



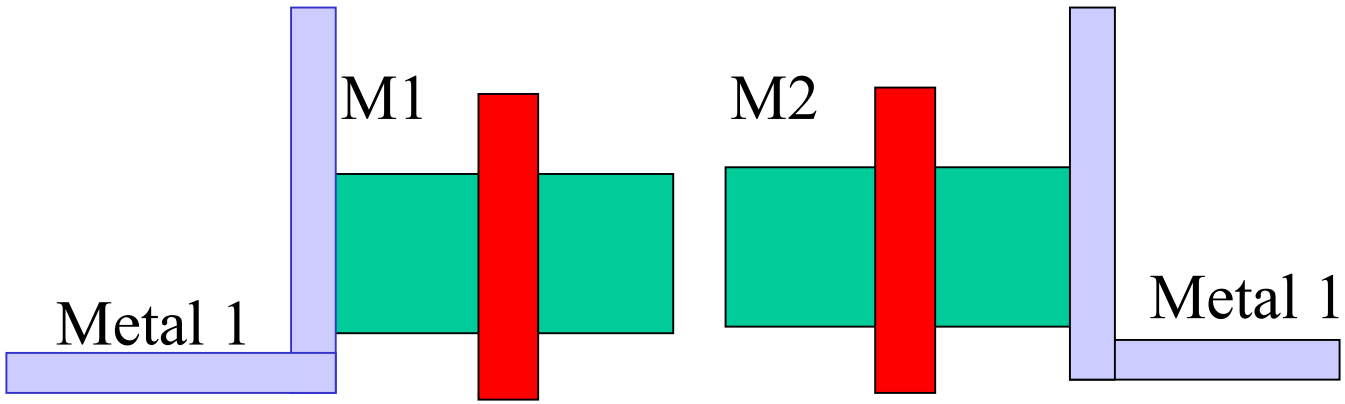
Photolythography process has different biases in different axes, hence the requirement

Symmetry

An unrelated metal line going in the vicinity of one of the transistor



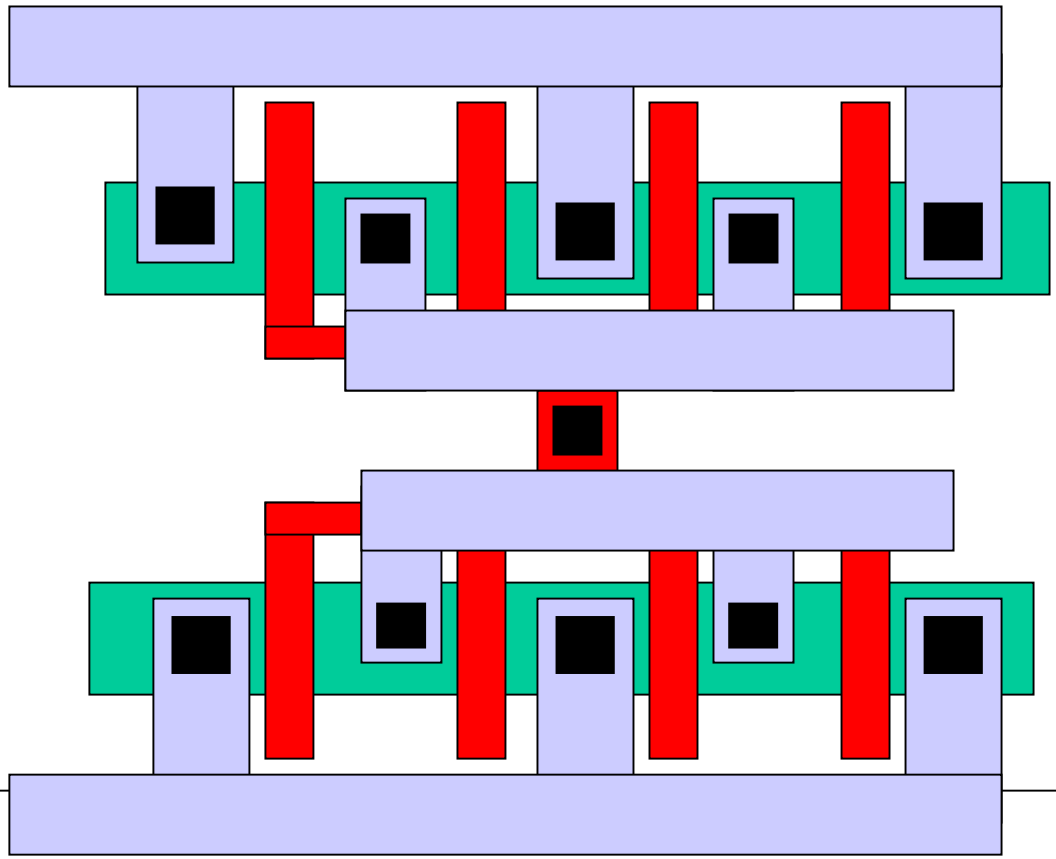
Symmetry should be preserved by adding another similar line



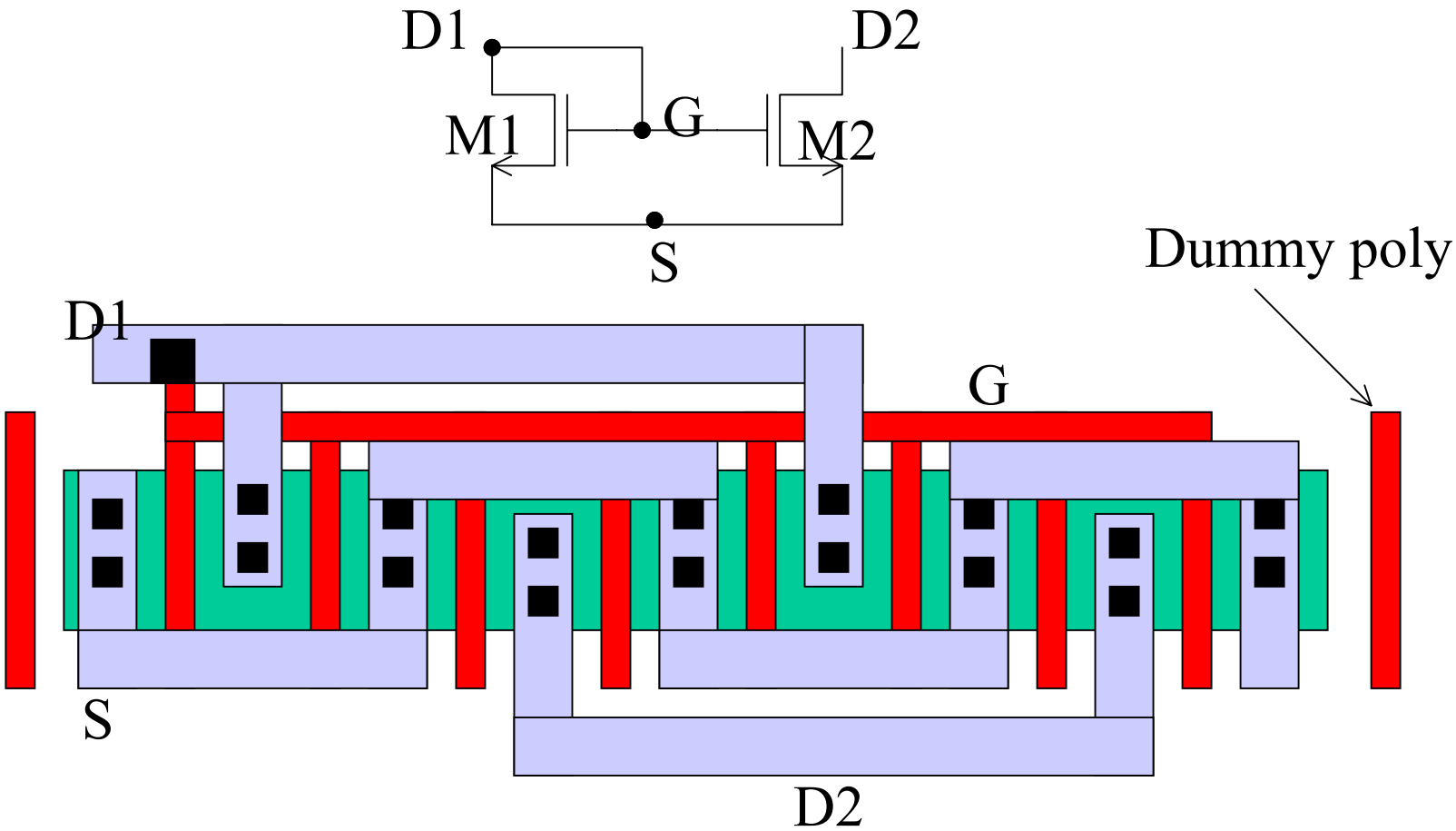
Unit cell repetition

Wide transistor should be laid out as parallel transistors of unit width to decrease gate resistance, s/d area capacitance as well as to counter ΔW effect

Disproportionate aspect ratio can be managed as below:



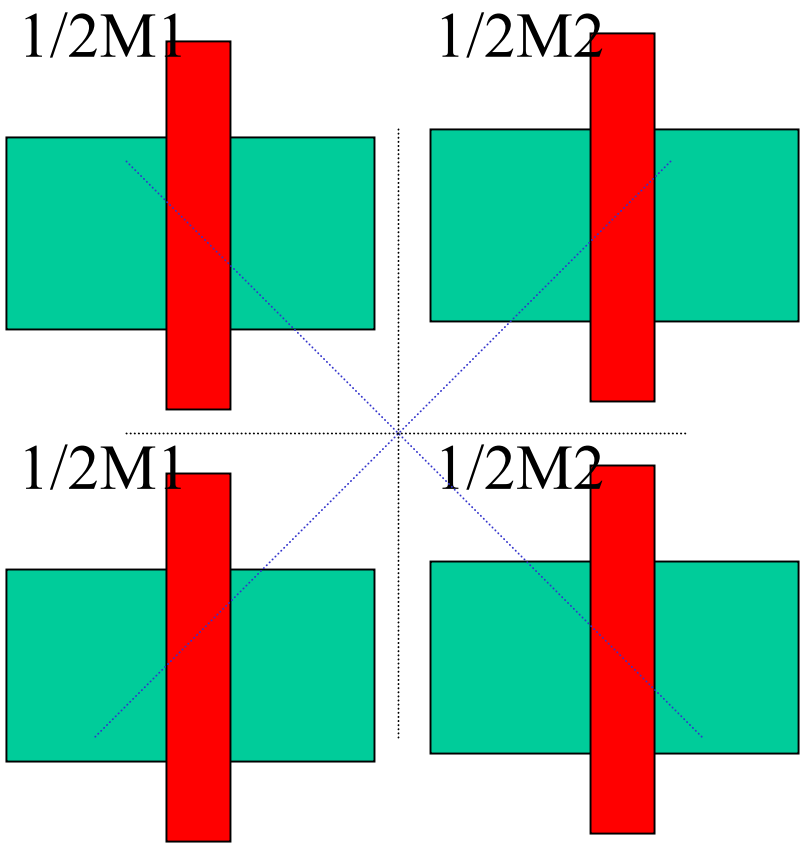
Interdigitation and dummy layer



Interdigitation distributes the transistors uniformly

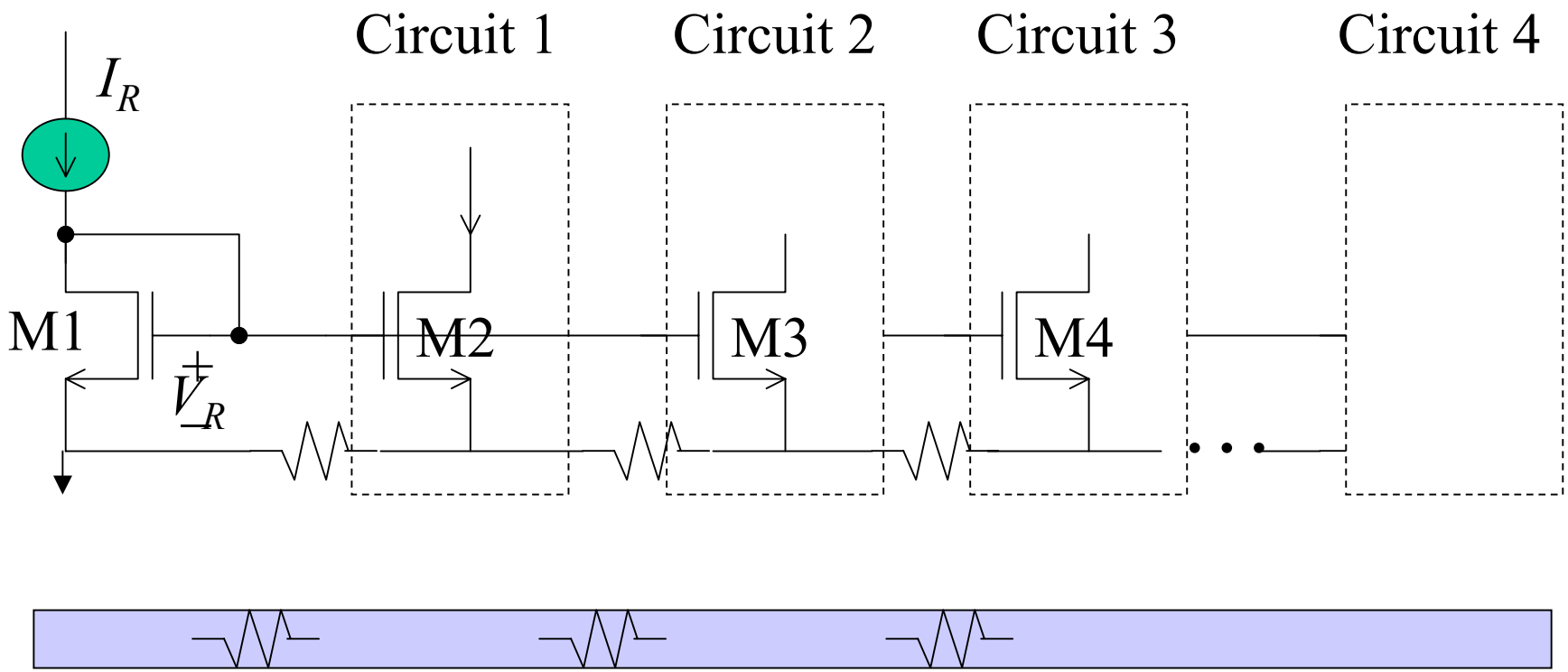
Dummy poly line eliminates loading effect in photo and etch

Common centroid



Common centroid configuration eliminates the first order gradient effects of parameters along both the axes

Interconnect routing



To distribute I_R in a large circuit, the resistance of ground bus makes $V_{gsn} \neq V_{gs1}$, thus affecting the current mirroring significantly

Interconnect routing

Decrease the ground bus resistance

Provide multiple ground node connections if possible
And use short span ground bus

Keep several reference distributed in a large circuit
and mirror the reference locally

Bandgap Voltage Reference

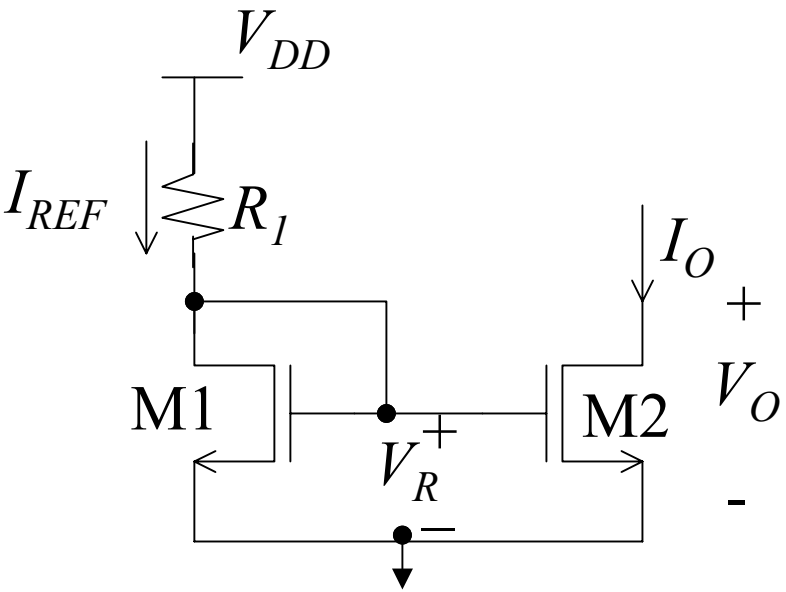
Bandgap Voltage reference

Design Task to set the DC bias of any circuits

1. Power Supply Independent Biasing

2. Temperature Independent Biasing
   Bandgap Voltage Reference

Resistance biasing



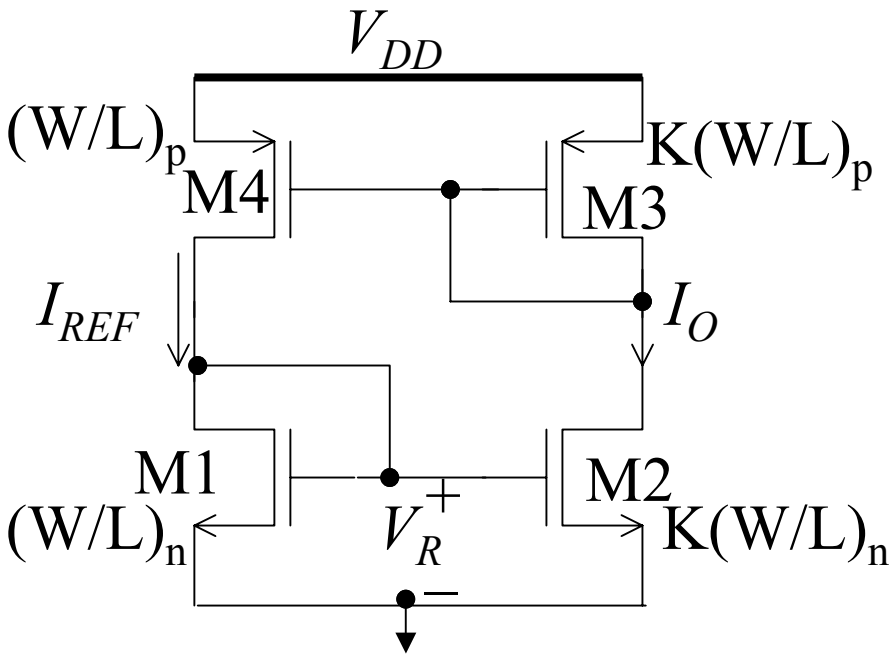
$$I_o = \frac{V_{DD}}{R_1 + 1/g_{m1}} \frac{W_2}{W_1}$$

$$\Delta I_o = \frac{\Delta V_{DD}}{R_1 + 1/g_{m1}} \frac{W_2}{W_1}$$

I_o is very sensitive to variation in V_{DD}

In order to have low sensitivity, the circuit must bias itself i.e. self biasing

Self Biasing circuit



I_O is bootstrapped to I_{REF}

$I_O = K\, I_{REF}$

But how do we fix I_O

Because as long as all transistors are saturated, any current is a valid solution for the circuit!

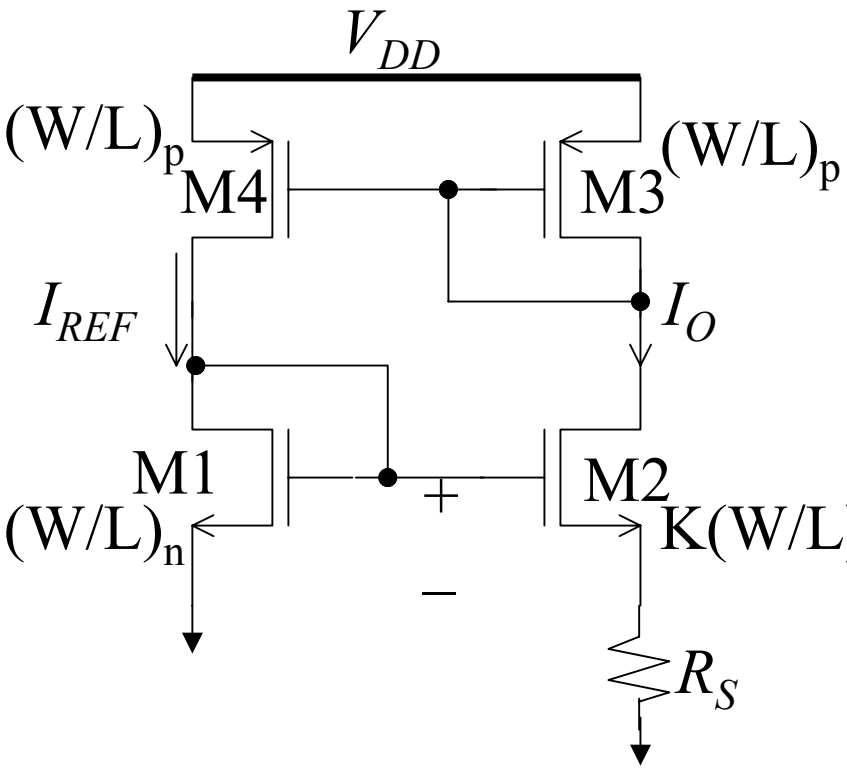
In order to uniquely define the current another constraint should be added to the circuit

Self Biasing

$$V_{gs1} = V_{gs2} + I_O R_S$$

$$\sqrt{\frac{2I_O}{\mu C_{OX}(W/L)_n}} + V_t = \sqrt{\frac{2I_O}{\mu C_{OX}(W/L)_n}} + I_O R_S$$

$$I_O = \frac{2}{\mu C_{OX}(W/L)_n} \frac{1}{R_S{}^2} \left(1 - \frac{1}{\sqrt{K}}\right)^2$$



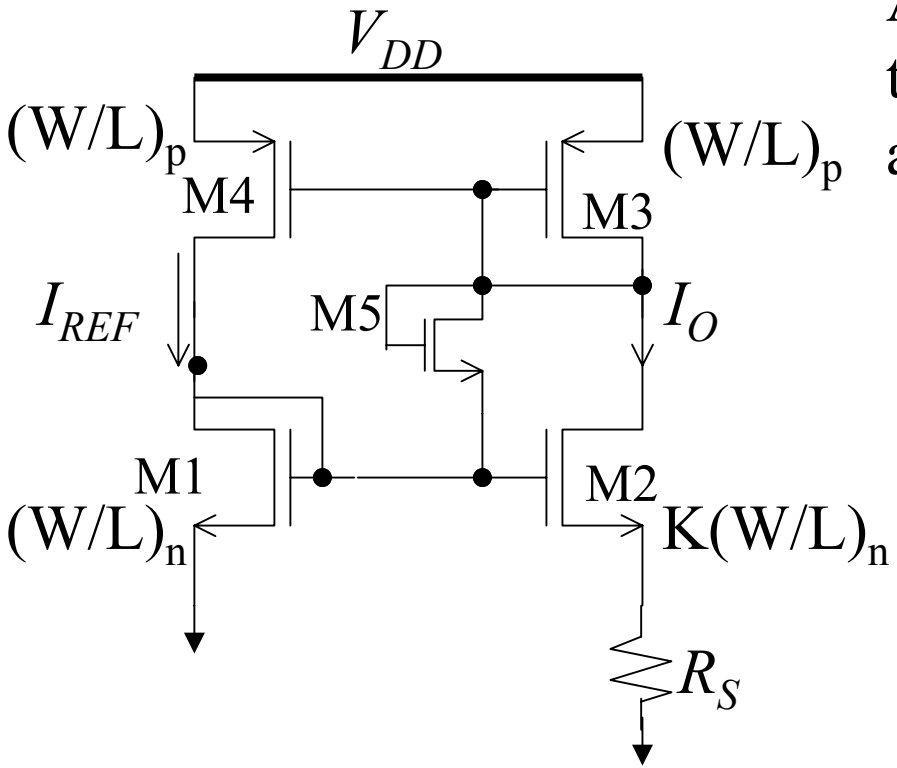
The current is uniquely defined and it is independent of supply voltage

The current is still a function of Process parameters and temperature

The Start-up problem

The previous circuit can support zero current as well!

At the start up it should be ensured that the circuit does not enter this degenerate situation



At the start up the 3 diode connected transistors M1, M5 and M3 provide a path to the ground give non zero *Io*

$$V_{t1} + V_{t5} + |V_{t3}| < V_{DD}$$

$$3V_t < V_{DD}$$

After start up, M2, M4 should turn on and M5 should be turned off

$$V_{gs1} + V_{t5} + |V_{gs3}| > V_{DD}$$

Temperature Independent Reference

The concept:

Generate the reference by combining two voltages of which one has negative temperature coefficient and the other one has a positive temperature coefficient

$$V_{REF} = \alpha_1 V_1 + \alpha_2 V_2$$

$$\frac{\partial V_1}{\partial T} = -ve$$

$$\frac{\partial V_2}{\partial T} = +ve$$

$$\alpha_1 \frac{\partial V_1}{\partial T} + \alpha_2 \frac{\partial V_2}{\partial T} = 0$$

V_{BE} of a BJT (diode) is a good candidate for negative TC

Difference between 2 different V_{BE}s is a good candidate for +ve TC

Other candidates such as resistor etc. also exist

Negative TC

The BJT collector current is given by

$$I_C = I_S \exp\left(\frac{V_{BE}}{V_T}\right) \rightarrow V_{BE} = V_T \ln\frac{I_C}{I_S}$$ where $V_T=kT/q$, is thermal voltage

$I_S=C\mu kTn_i^2$ *C is proportionality constant*

$\mu \alpha \mu_O T^m$ *where* $m\approx -3/2$

$n_i^2 \alpha T^3 exp(-Eg/kT)$ *where Eg is bandgap of Si, Eg* $\approx 1.12eV$

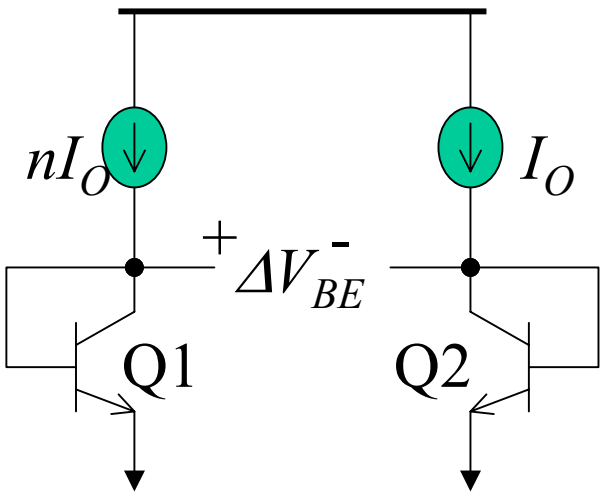
At constant collector current, using above equations,

$$\frac{\partial V_{BE}}{\partial V_T} = \frac{V_{BE} - (4+m)V_T - E_g/q}{T}$$

Note that the TC is a function of V_{BE} and T itself

For V_{BE}=0.75V and T=300^{o}K, $\frac{\partial V_{BE}}{\partial V_T} = -1.5mV/^oK$

Positive TC



Suppose that the two collector currents are nI_O and I_O (typically done by adjusting the device dimensions/layout)

$$\Delta V_{BE} = V_{BE1} - V_{BE2}$$

$$\Delta V_{BE} = V_T \ln \frac{nI_O}{I_S} - V_T \ln \frac{I_O}{I_S}$$

$$\Delta V_{BE} = V_T \ln n$$

$$\frac{\partial \Delta V_{BE}}{\partial T} = \frac{k}{q} \ln n$$

Note that the TC is independent of I_c and Temperature

Bandgap Reference

$$V_{REF} = \alpha_1 V_1 + \alpha_2 V_2$$

Choose $V_1 = V_{BE}$ and $V_2 = \Delta V_{BE}$

Set $\alpha_1 = 1$ and choose α_2 such that the TC is zero at 300°K

Since dV_1/dT = -1.5mV/°K and $d\Delta V_{BE}/dT$ = +0.087mV/°K,
Choose α_2 so that $(\alpha_2 \ln n)(0.087mV/^oK) = +1.5mV/^oK$

$$V_{REF} \approx V_{BE} + 17.2V_T$$

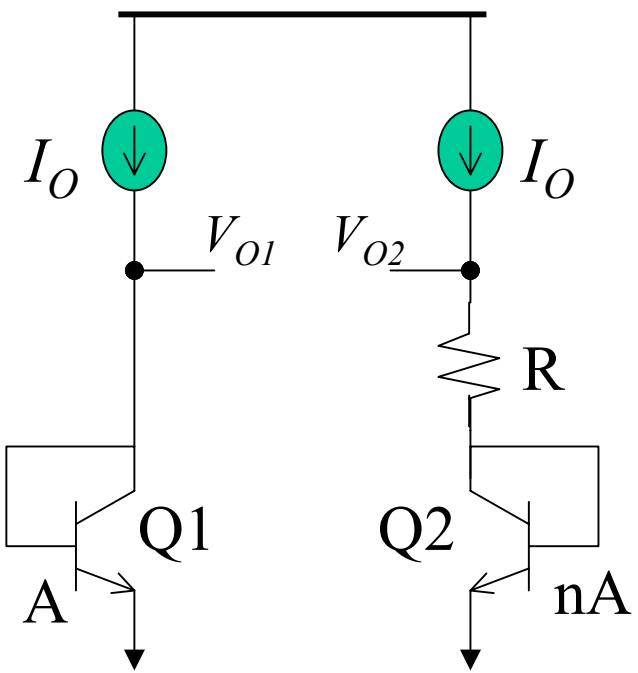
$$V_{REF} \approx 1.19V$$

Note: VREF can also be expressed as

$$V_{REF} \approx \frac{E_g}{q} + (4+m)V_T$$

In the limit as $T \to 0$, $V_{REF} \to E_g/q$
Hence the name bandgap voltage reference

The circuit to add V_{BE} and $17.2V_T$



Q1 is unit transistor with area A
Q2 has n unit transistors in parallel
The current in one unit of Q2 is I_O/n

Suppose that V_{O1} and V_{O2} are made equal by some external means

Then,

$$V_{BE1}=RI_O+V_{BE2}$$

$$RI_O = V_{BE1} - V_{BE2} = V_T \ln n$$

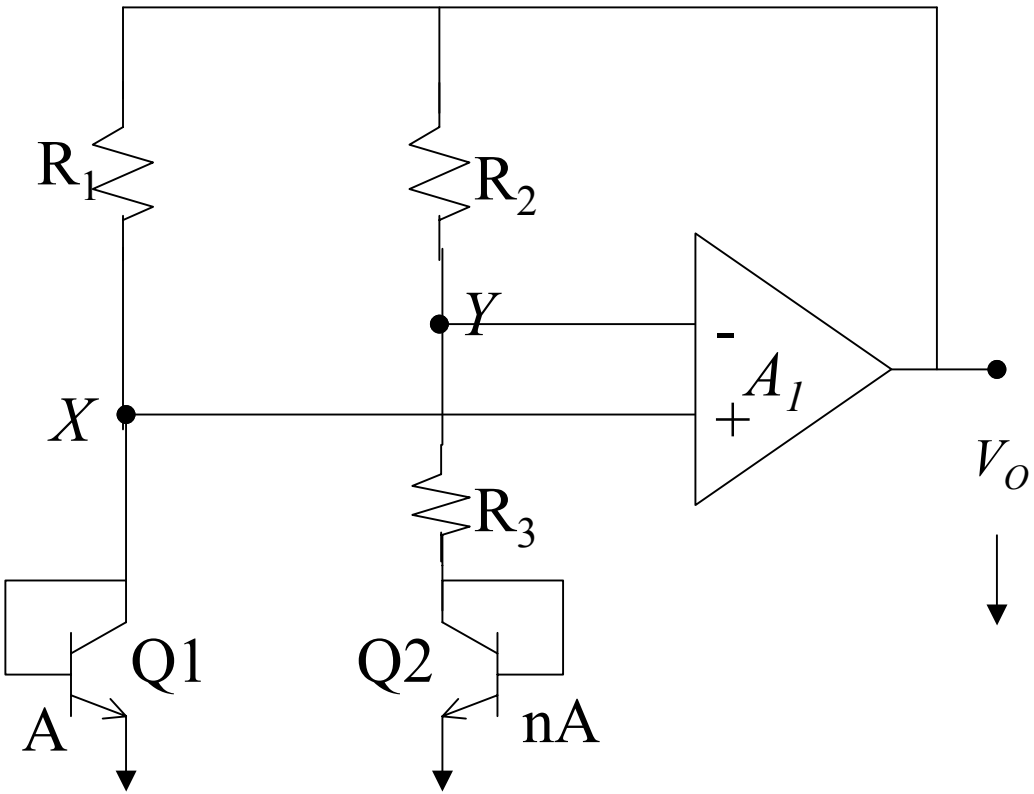
$$V_{O2}=V_{BE2} + V_T \ln n$$

which is the required reference

Need a mechanism for $V_{O1} = V_{O2}$

$\ln n$ =17.2 results in impractical n and hence should some how scaled properly

Circuit Implementation



The OPAMP forces $V_X = V_Y$

$$V_{BE1} - V_{BE2} = V_T \ln n$$

This results in a current through the right branch

$$I_{R3} = V_T \ln n / R_3$$

$$Vo = V_{BE2} + \frac{V_T \ln n}{R_3}(R_2 + R_3)$$

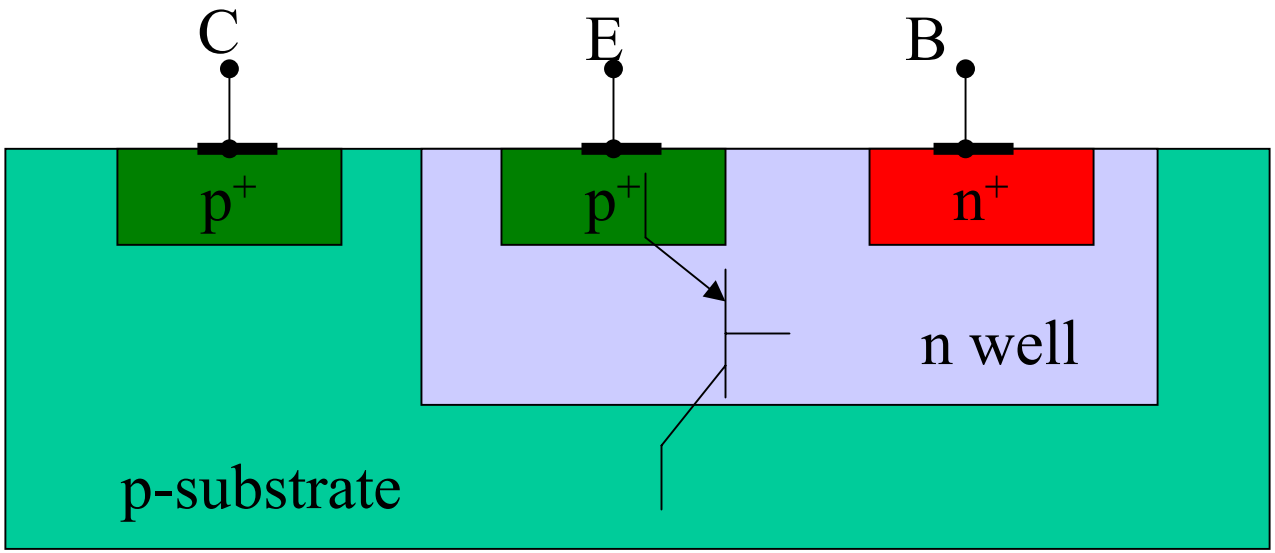
$$Vo = V_{BE2} + V_T \ln n\left(1 + \frac{R_2}{R_3}\right)$$

If R2/R3 = 10, then n=5

The output *Vo* gives the required reference voltage

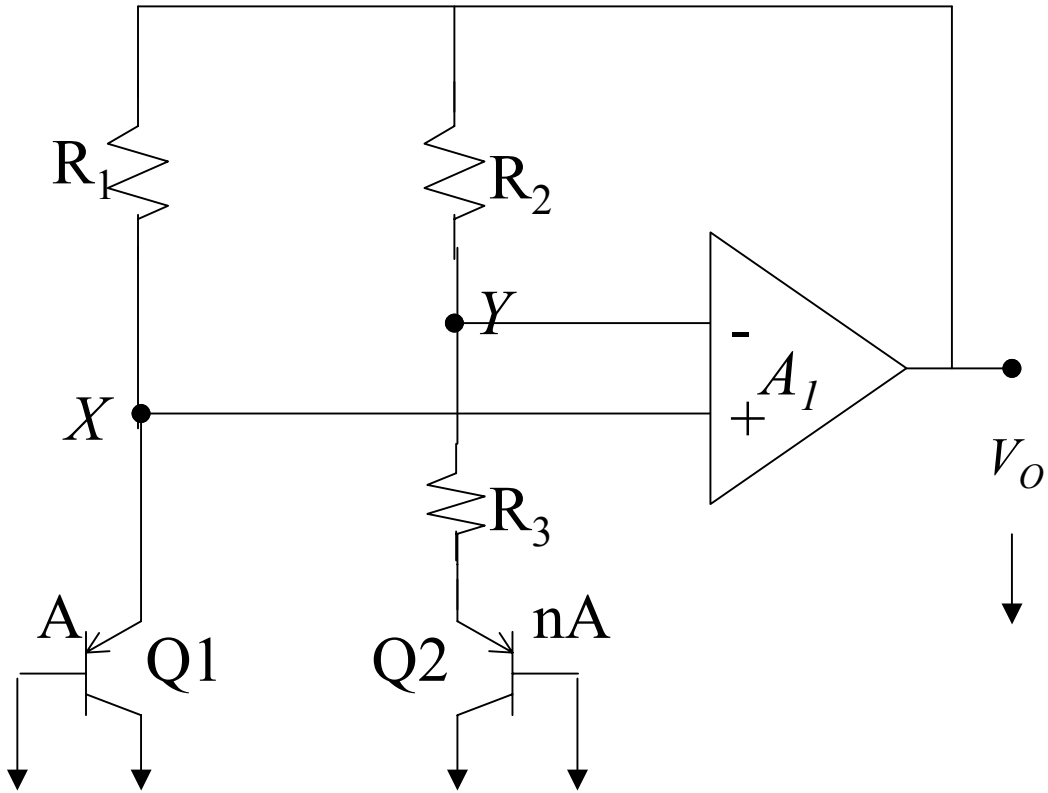
Compatibility with CMOS Technology

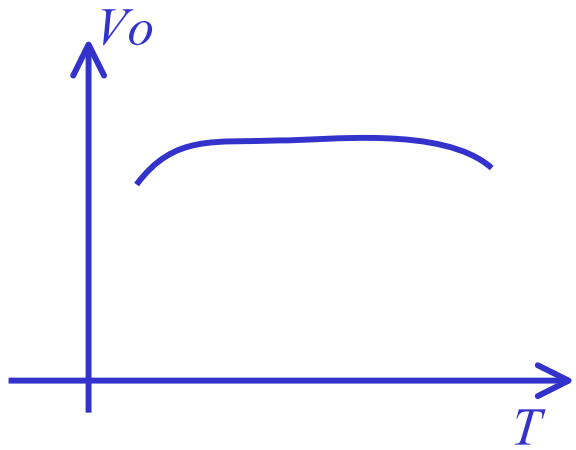
In an n-well technology the vertical PNP BJT can be realised



The p-substrate, connected to most negative potential (Gnd) acts as a collector whereas n-well and p^+ region act as base & emitter

Modified circuit for CMOS Technology

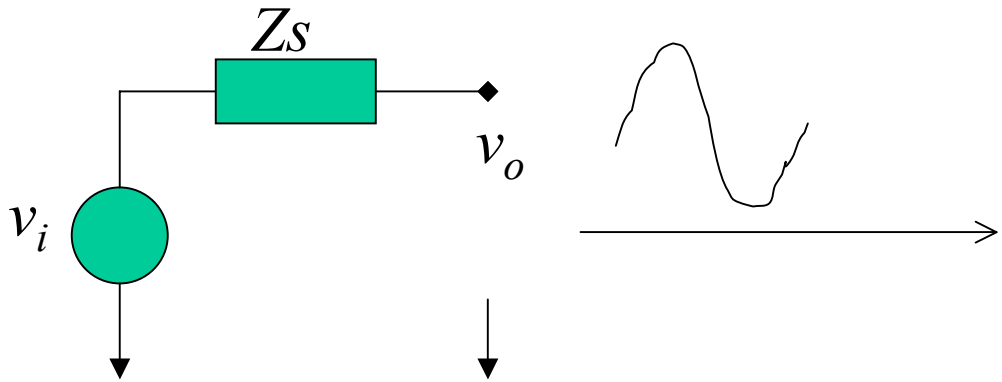




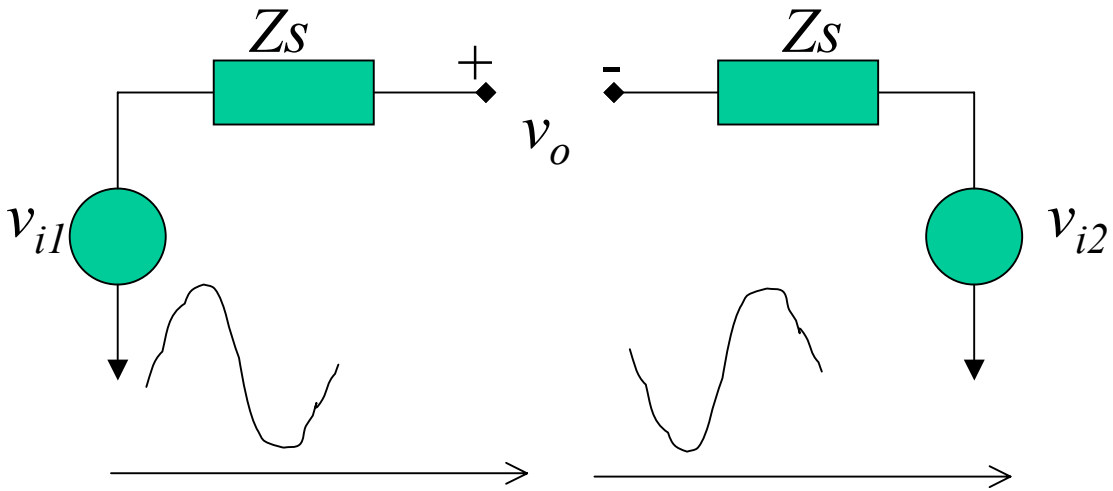
The output voltage Vo is the bandgap voltage reference

Differential Amplifier

Single ended vs. differential signaling



Single ended signal is measured with respect to a fixed potential (gnd)



Differential signal is measured with respect to 2 nodes which make equal and opposite excursion

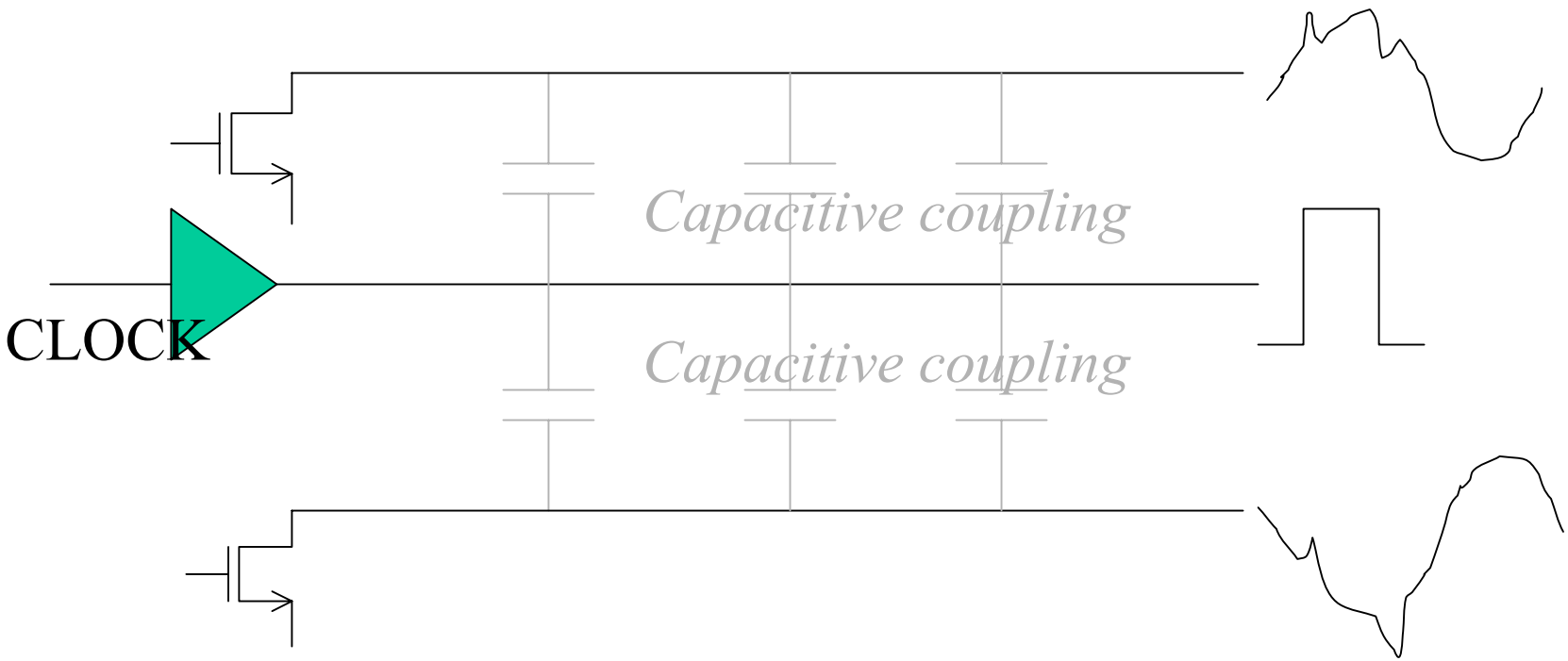
Differential Voltage $v_d = \Delta v_i = v_{i1} - v_{i2}$

Common mode Voltage $V_c = \frac{v_{i1} + v_{i2}}{2}$

Advantages of differential signaling

High immunity to environmental noise

i.e. common mode noise is rejected

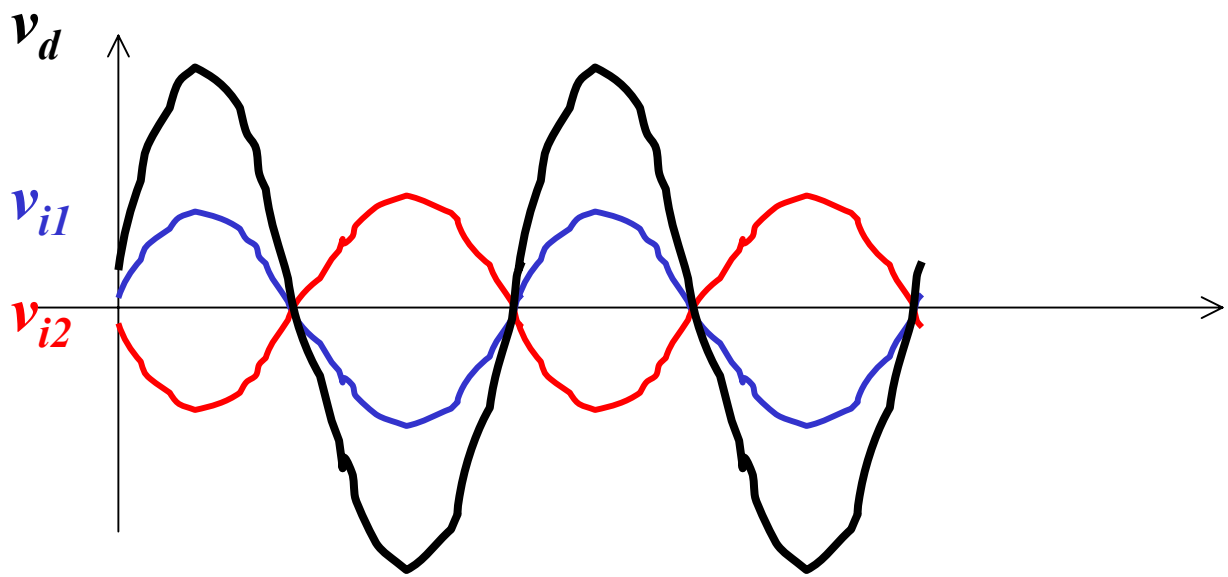


The capacitive coupling noise in cancelled

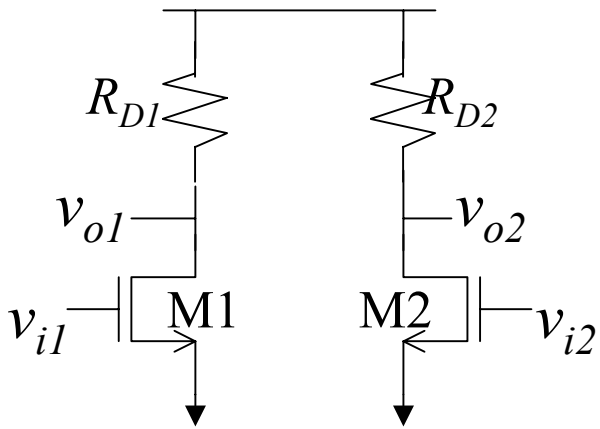
Advantages of differential signaling

Similar to noise reception, the noise injected by differential signal lines is also very low

The output swing is doubled in differential signaling



Differential signal amplification



Combine two single ended amplifiers to form a differential amplifier

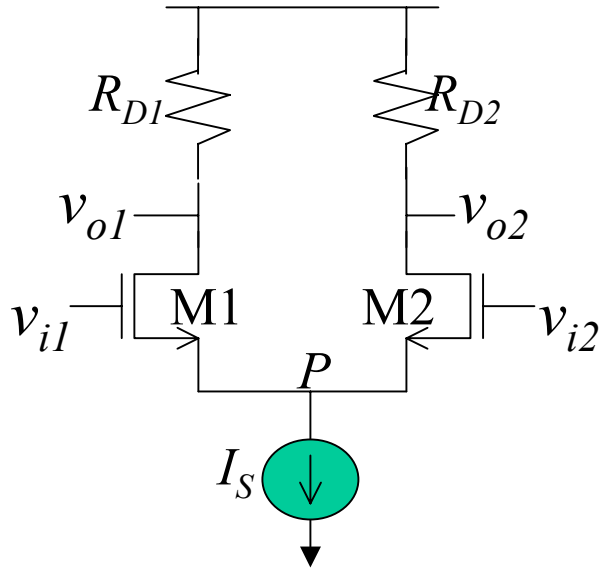
The common mode input level determines the differential gain, which is undesirable

Requirement of an ideal Differential Amplifier:

Differential gain should be independent of common mode input

$\Rightarrow$ A configuration in which bias current is independent of *Vc*

Differential Amplifier with current source bias



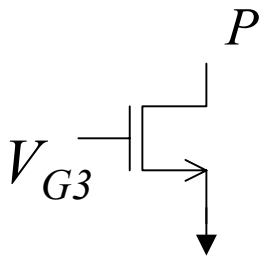
If *Is* is constant then I_{d1} and I_{d2} are independent of *Vc*

$I_{d1} = I_{d2} = Is/2$

Does it mean *Vc* can be between 0 to ∞?

The lower limit of *Vc*:

Is is typically realised using current mirror transistor



When *Vc=0*, M1 and M2 are off, M3 is in deep triode region and *Vp*=0 and *Is*=0

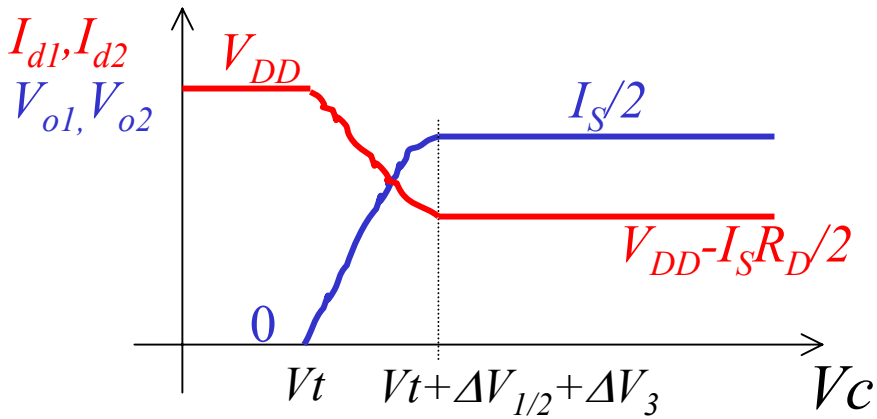
Hence the circuit does not act as an amplifier

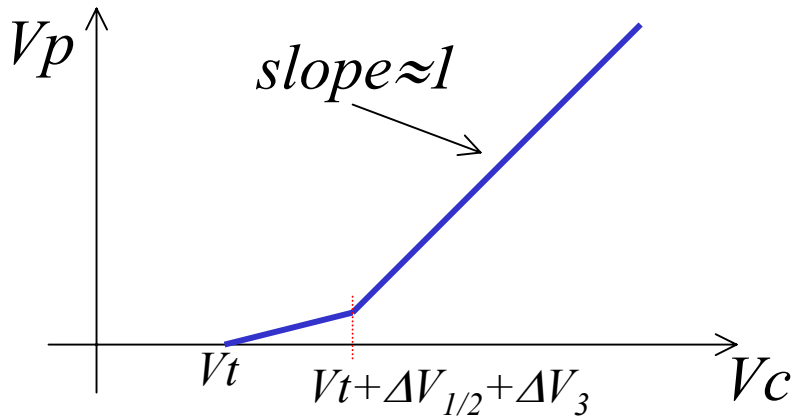
Lower limit for *Vc*

When *Vc* reaches *Vt*, M1 and M2 start turning on and hence *Vp* starts following *Vc* (source follower)

The current Is starts increasing and hence *Id1* and *Id2*

When *Vp* reaches ΔV_3 of M3, then M3 comes into saturation *Is* remains constant and so do *Id1* and *Id2*





$$V_{c\min} \approx V_t + 2\Delta V$$

Upper limit for *Vc*

As *Vc* starts approaching V_{DD} at certain value of *Vc*, M1 and M2 come out of saturation

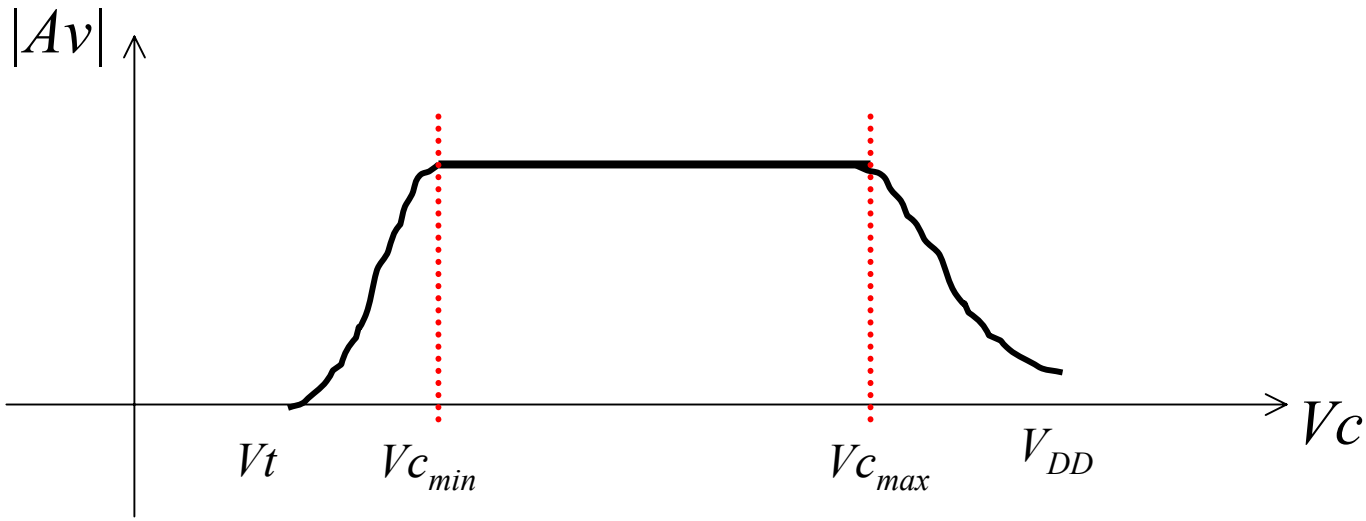
This happens at $V_{c\max} = V_{DD} - \frac{I_S}{2} R_D + V_t$

Beyond this point circuit is not usable since gm and ro drop

Further after this point, *Vp* starts lagging behind *Vc* in order to maintain high gate overdrive to conduct the current $I_S/2$

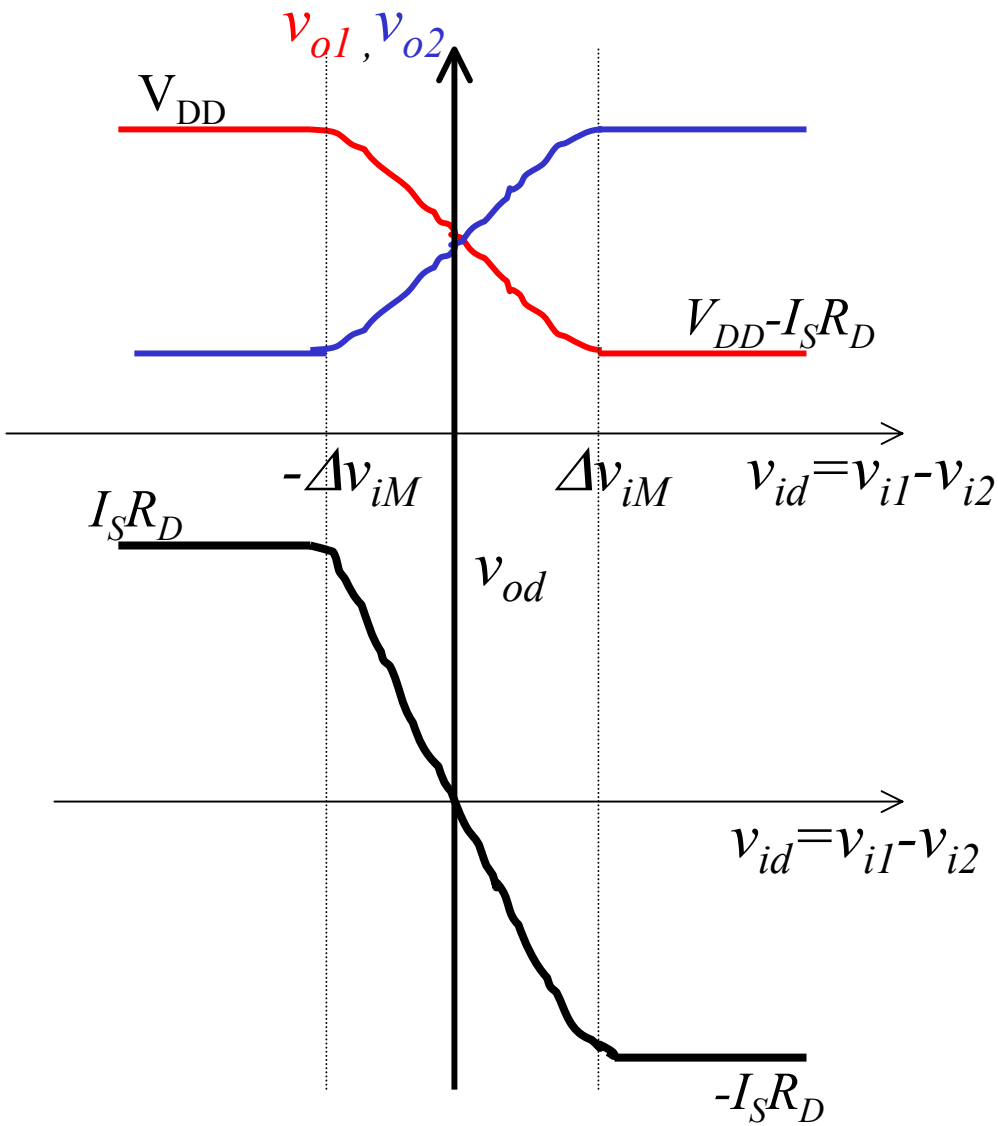
We will revisit this point when we discuss low voltage wide common mode range OPAMPs

The useful range of operation w.r.t. *Vc*



Useful voltage gain can be obtained for $Vcmin < Vc < Vcmax$

Differential Response

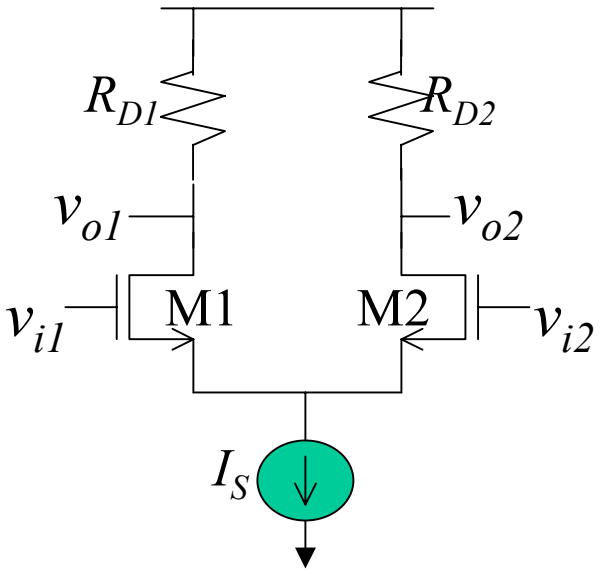


When $v_{id} = \pm \Delta v_{iM}$

The entire current I_S flows through only one of the two branches of differential pair (i.e. either through M1 or M2)

This limits the maximum input swing for v_{id}

Input range and Transconductance



$$I_{d1} = \frac{\mu\varepsilon_{ox} W_1}{T_{ox} L_1} \frac{\left(V_{gs1} - V_t\right)^2}{2} = \frac{k_1}{2}\left(V_{gs1} - V_t\right)^2$$

$$I_{d2} = \frac{\mu\varepsilon_{ox} W_2}{T_{ox} L_2} \frac{\left(V_{gs2} - V_t\right)^2}{2} = \frac{k_2}{2}\left(V_{gs2} - V_t\right)^2$$

$\Delta V_i = V_{i1} - V_{i2}$ *and* $\Delta I_d = I_{d1} - I_{d2}$

The objective is to get $\Delta I_d = f(\Delta V_i)$

$$I_{d1} = \frac{I_S}{2} + \frac{\Delta I_d}{2} \qquad I_{d2} = \frac{I_S}{2} - \frac{\Delta I_d}{2}$$

Let $k_1 = k_2 = k$, *then*

$$V_{gs1} = V_t + \sqrt{\frac{2I_{d1}}{k}} \qquad V_{gs2} = V_t + \sqrt{\frac{2I_{d2}}{k}}$$

Input range and Transconductance

$$\Delta V_i = V_{i1} - V_{i2} = V_{gs1} - V_{gs2}$$

$$\Delta V_i = \sqrt{\frac{2}{k}}\left[\sqrt{I_{d1}} - \sqrt{I_{d2}}\right]$$

$$\Delta V_i^{\,2} = \frac{2}{k}\left[I_{d1} + I_{d2} - 2\sqrt{I_{d1} I_{d2}}\right]$$

$$\Delta I_d = \frac{k}{2}\Delta V_i \sqrt{\frac{4I_S}{k} - \Delta V_i^{\,2}}$$

This is used to get the input range and transconductance

Input range

$$\Delta I_d = \frac{k}{2}\Delta V_i \sqrt{\frac{4I_S}{k} - \Delta V_i^2}$$

At $\Delta V_i = V_{iM}$, $\Delta I_d = I_{SS}$, *Hence*

$$\Delta V_{iM} = \sqrt{\frac{2I_S}{k}}$$

Trade-off

ΔV_{iM} can be increased by decreasing k

but lower k $\Rightarrow$ higher Vgs-Vt

this results in increased Vcmin

ΔV_{iM} can be increased by increasing I_S

but higher I_S $\Rightarrow$ lower V_{DD}-$I_S R_D$/2+Vt

this results in reduced Vcmax

also higher static power

Typical input range

If the bias current I_S=100μA

$\mu C_{ox} = 50\mu A/V^2$

$W/L = 100$

$$\Delta V_{iM} = \sqrt{\frac{2\times 100}{50\times 100}} = \sqrt{0.04} = 0.2V$$

Differential Transconductance

$$G_m = \frac{d\Delta I_d}{d\Delta V_i}$$

$$G_m = \frac{k}{2}\sqrt{\frac{4I_S}{k} - \Delta V_i^2} - \frac{k}{2}\frac{\Delta V_i^2}{\sqrt{4I_S/k - \Delta V_i^2}}$$

G_m is maximum when $\Delta V_i = 0$

$$G_{m0} = \sqrt{kI_S}$$

Trade-off

G_m can be increased by increasing k

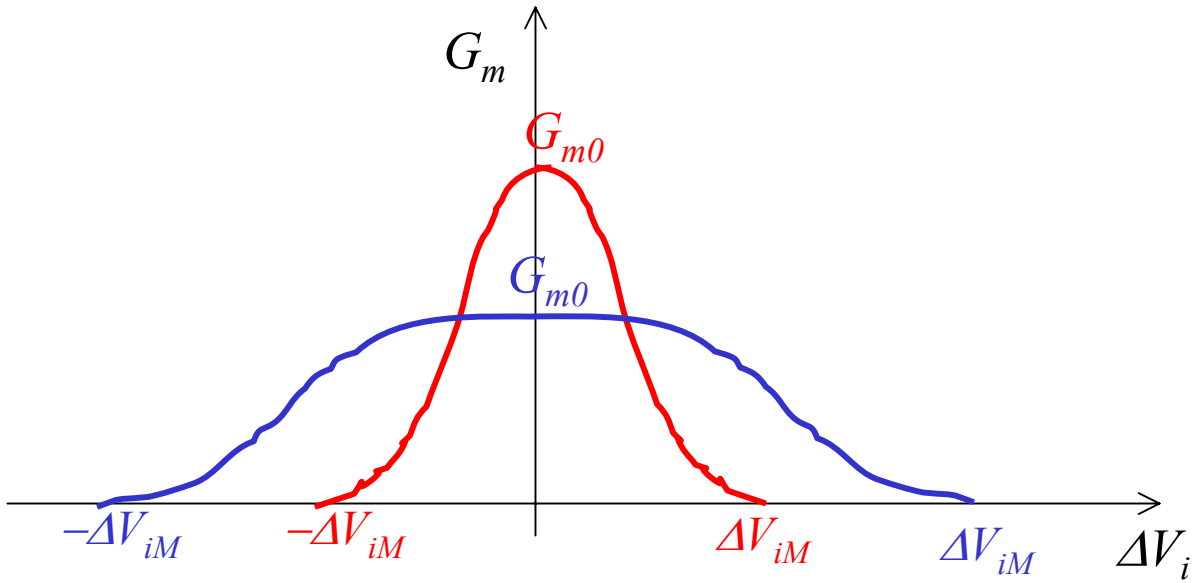
but higher k $\Rightarrow$ lower input swing ΔV_{iM}

G_m can be increased by increasing I_S

but higher I $\Rightarrow$ lower Vcmax

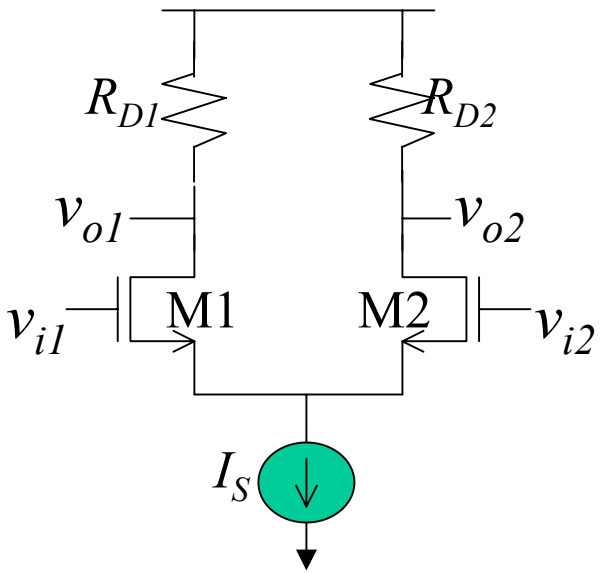
also higher static power

G_m versus ΔV_i



Trade-off between Gain and linearity!

Differential and Common mode response



Any given inputs v_{i1} and v_{i2} can be decomposed into differential and common mode signals

$$v_{id}=v_{i1}-v_{i2} \qquad v_{ic}=\frac{v_{i1}+v_{i2}}{2}$$

$$v_{od}=v_{o1}-v_{o2} \qquad v_{oc}=\frac{v_{o1}+v_{o2}}{2}$$

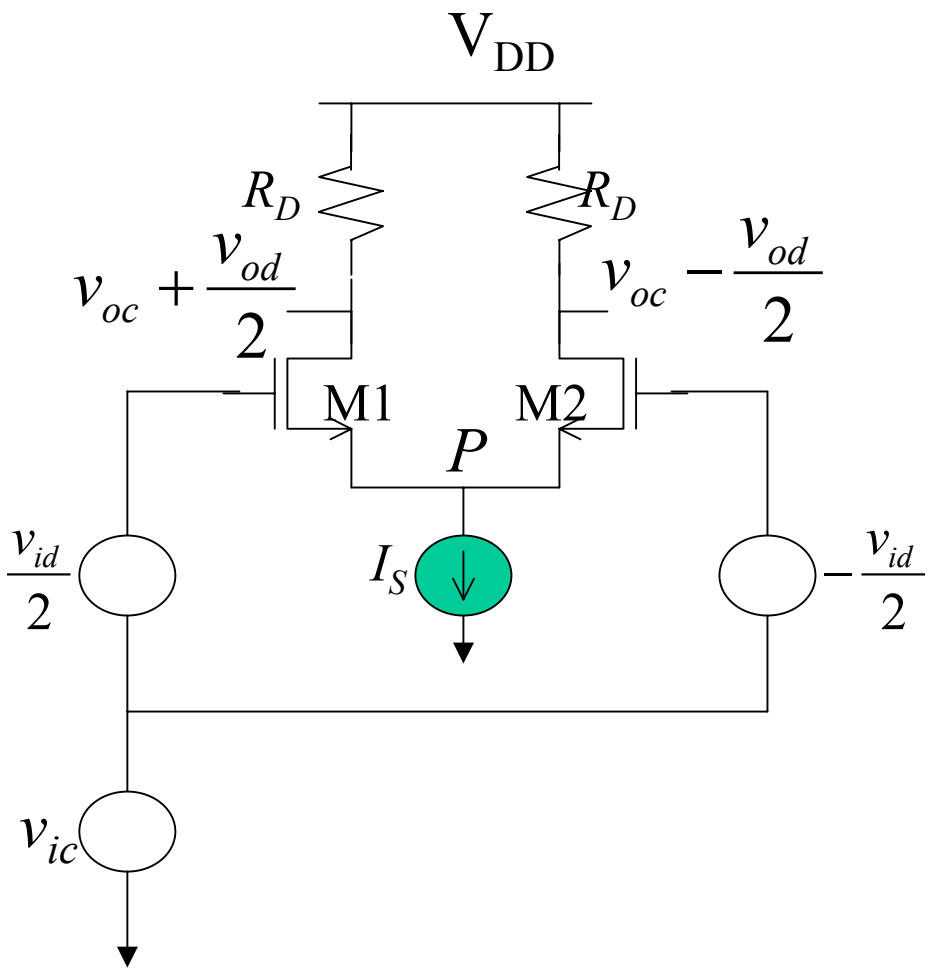
$$v_{i1}=v_{ic}+\frac{v_{id}}{2} \qquad v_{i2}=v_{ic}-\frac{v_{id}}{2}$$

$$v_{o1}=v_{oc}+\frac{v_{od}}{2} \qquad v_{o2}=v_{oc}-\frac{v_{od}}{2}$$

Differential mode gain $A_d=\dfrac{v_{od}}{v_{id}}$ with $v_{ic}=0$

Common mode gain $A_c=\dfrac{v_{oc}}{v_{ic}}$ with $v_{id}=0$

Modified circuit in terms v_{id} and v_{ic}



For the symmetric differential pair:

$g_{m1}=g_{m2}$,

$r_{o1}=r_{o2}$,

$R_{D1}=R_{D2}$

The differential response is obtained by setting $v_{ic} = 0$

The common mode response is obtained by setting $v_{id} = 0$

Differential response

What happens at node P for a pure differential input?

1. If v_{i1} changes by $+\Delta v$ and v_{i2} by $-\Delta v$ and
if the circuit is linear then Vp does not change

2. Since the current flowing out of node P is constant (IS),
the change in currents in two arms should cancel
i.e. $g_m(\Delta v_{gs1} + \Delta v_{gs2})=0$ or $\Delta v_{gs1} = -\Delta v_{gs2} = \Delta v_{gs}$

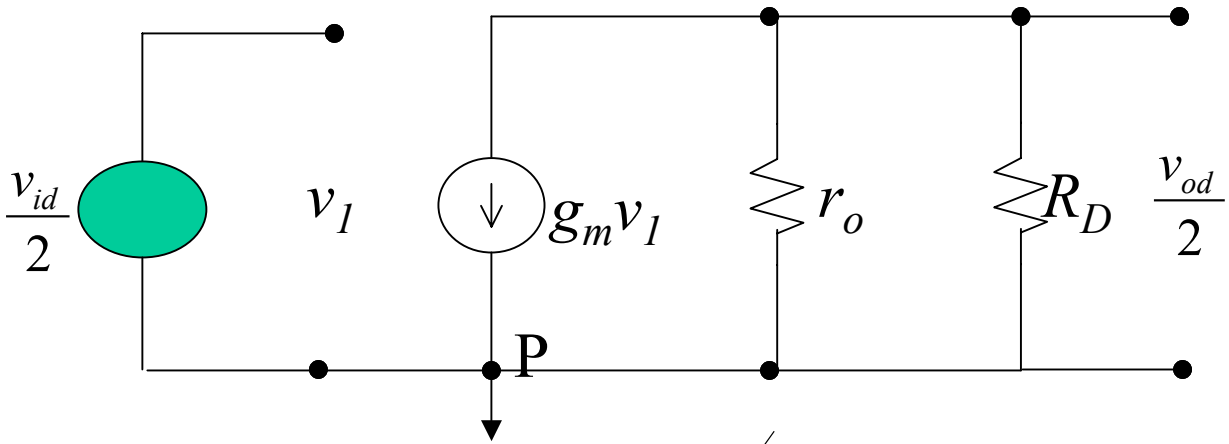
3. Also $V_{gs1}=V_{i1}-Vp$ and $V_{gs2}=V_{i2}-Vp$

From 1, 2, 3, it follows that Vp can not change and
the entire change in input is absorbed by gate overdrive

Vp is at AC ground

Half circuit concept

A fully differential circuit can be analysed by looking at only one half of the circuit



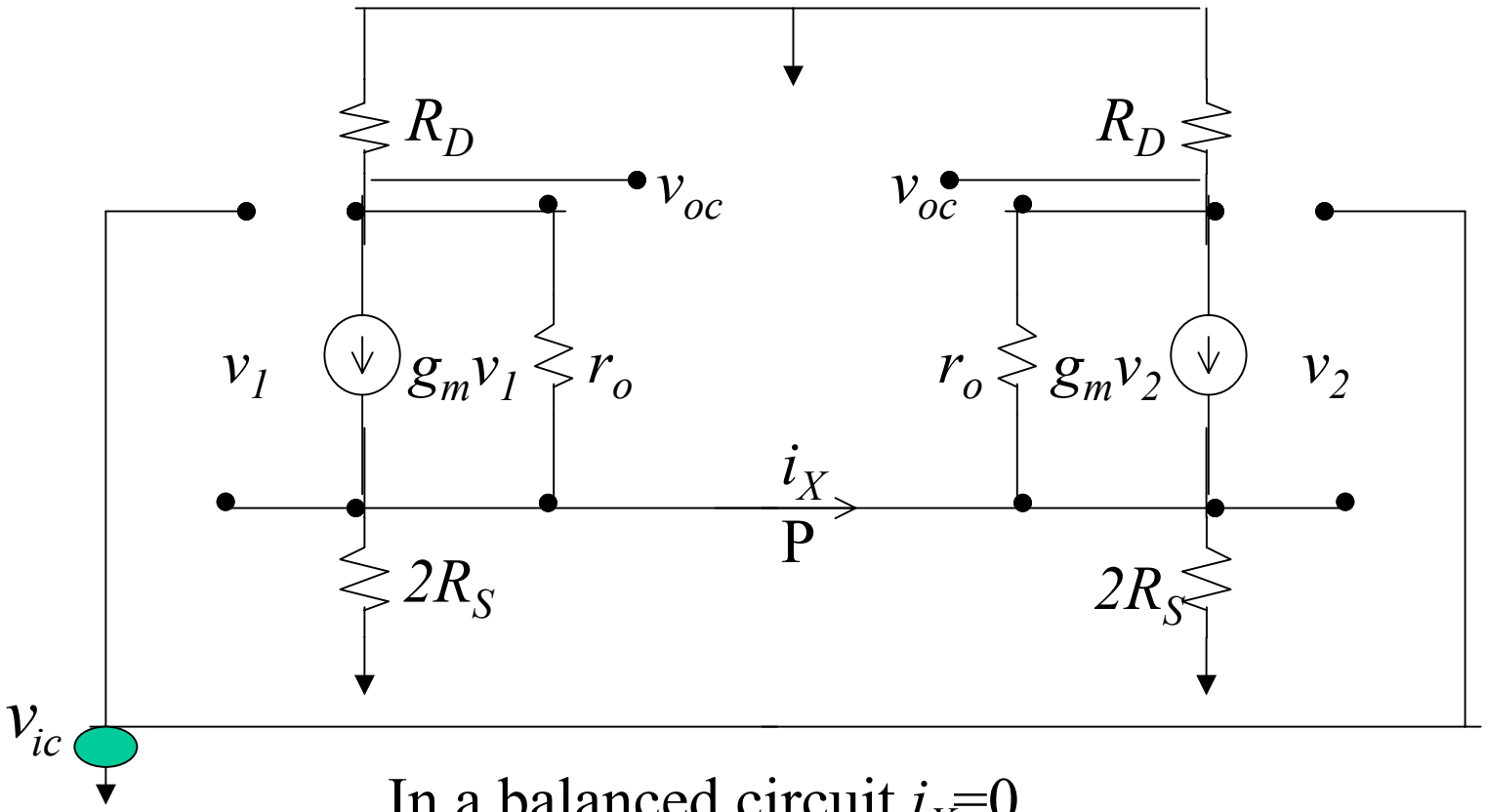
$$A_d = \frac{v_{od}}{v_{id}} = \frac{v_{od}/2}{v_{id}/2} = -g_m \frac{r_o R_D}{r_o + R_D}$$

$$A_d \approx -g_m R_D \qquad \text{If } R_D << r_o$$

Note: If a single stage CS amplifier is biased with I_S, then the gain would be twice the differential gain

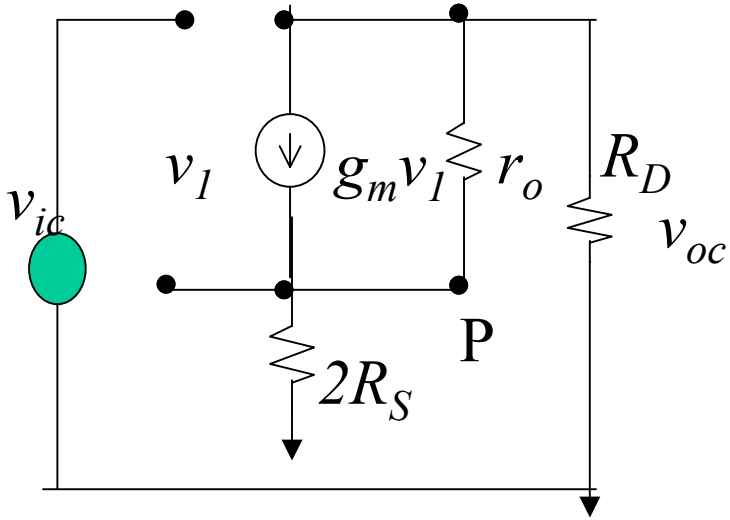
Common mode equivalent circuit

For a pure common mode signal v_{ic} is translated equally in both branches, thus changing v_{oc}
Vp is NOT at AC ground



In a balanced circuit i_X=0

Common mode response



The circuit can be broken into two parts

The equivalent circuit looks like common Source amplifier with source degeneration

The degenerated transconductance $-1/2R_S$

$$\therefore A_c = -\frac{R_D}{2R_S}$$

The common mode gain affects the DC bias point and also affects the output differential mode swing

Common Mode Rejection Ratio

CMRR is a measure of differential amplifier which indicates its ability to suppress common mode gain and enhance the differential mode gain

$$CMRR = \frac{A_d}{A_c}$$

$$CMRR = \frac{-g_m R_D r_o / R_D + r_0}{-R_D / 2R_S}$$

$$CMRR = \frac{2 g_m r_o R_S}{r_o + R_D}$$

For high CMRR, R_S should be as large as possible
Hence the need for good current source at source of M1/M2

Common mode to differential conversion

Another important problem is the conversion of CM signal to DM output in presence of device mismatch

We define A_{cd} as $A_{cd} = \frac{v_{od}}{v_{ic}}$ with v_{id} =0

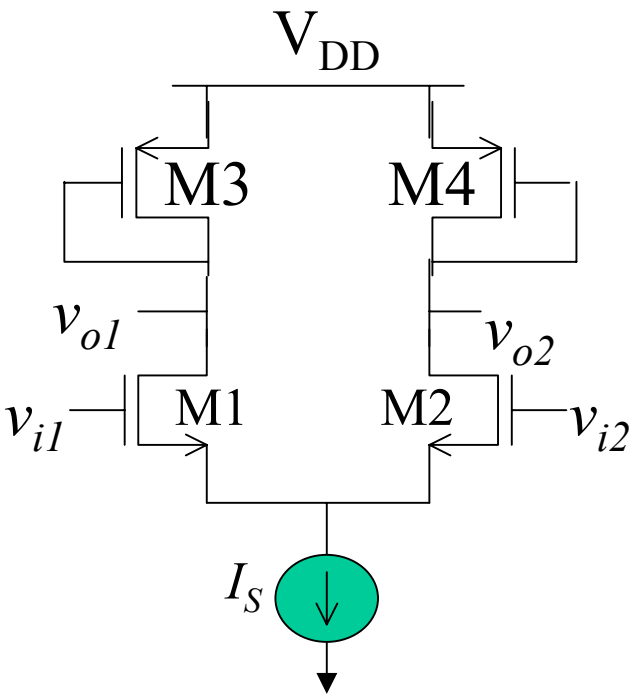
$$CMRR = \frac{A_d}{A_{cd}}$$

For mismatch in R_D, $A_{cd} = -\frac{\Delta R_D}{2R_S}$

For mismatch in gm, $A_{cd} = -\frac{\Delta g_m R_D}{(g_{m1} + g_{m2})R_S + 1}$

The problem becomes serious at high frequencies, since R_S gets shunted by the capacitances

Differential pair with diode connected load



Advantages:

Resistance is eliminated

Output common mode voltage is well defined

Disadvantages:

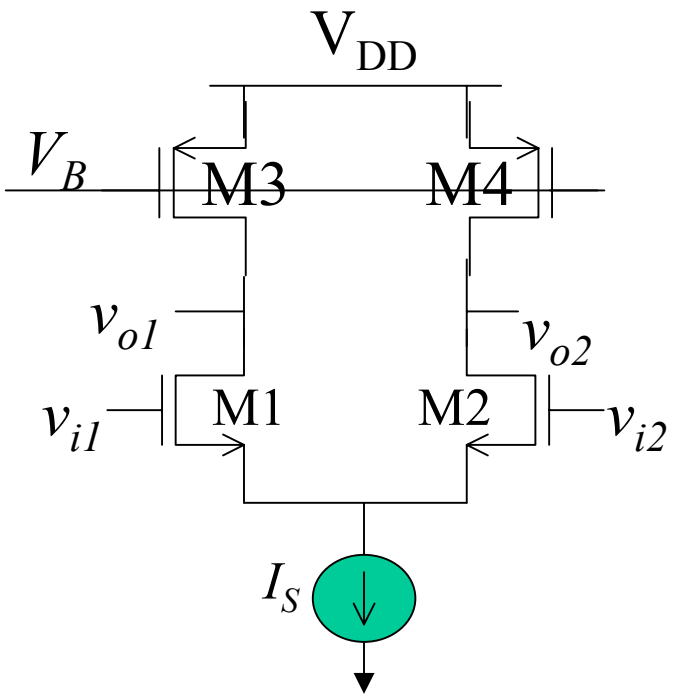
The gain is limited by g_{mp} since

$$Ad=-g_{mn}/g_{mp}$$

Stringent trade off between gain and swing

$$\because Av = -\sqrt{\frac{\mu_n (W/L)_n}{\mu_p (W/L)_p}} = \frac{(V_{gs}-V_t)_p}{(V_{gs}-V_t)_n}$$

Differential pair with current source load



Advantages:

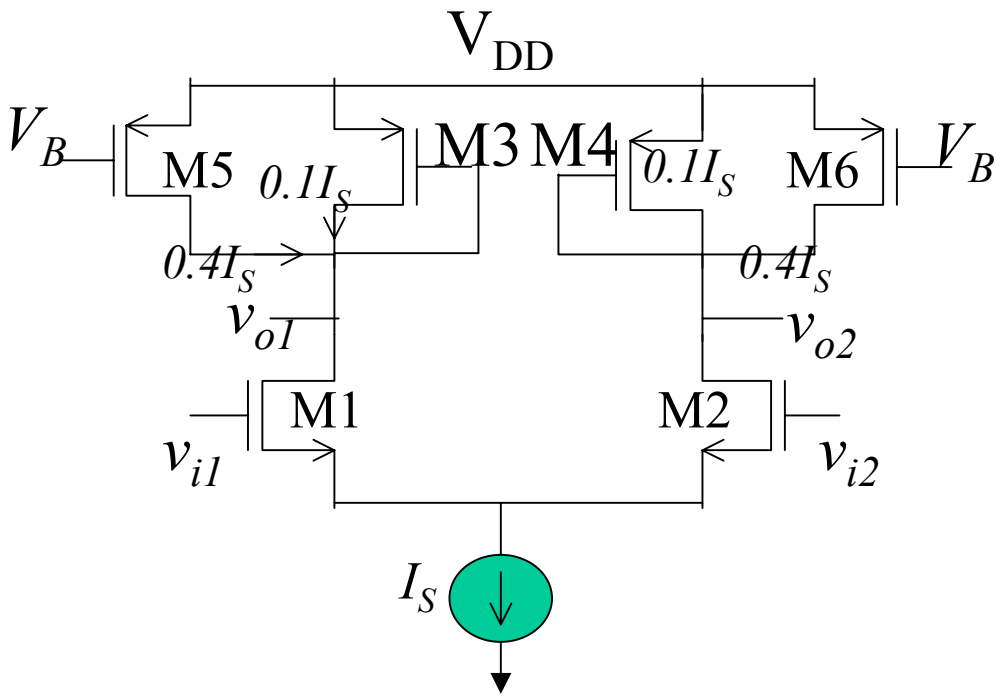
Ad=-$g_{mn}(r_{on}||r_{op})$

Gain and swing are not very strongly coupled as in the earlier case

Disadvantages:

The output common mode voltage is not very well defined

Combined load



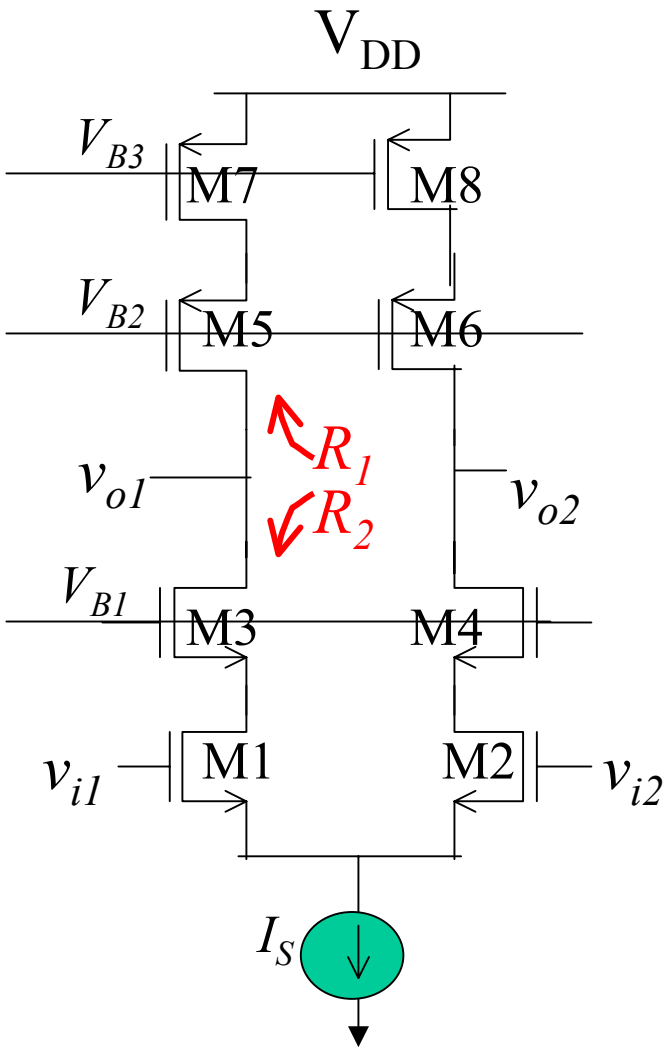
The load consists of diode as well as constant current source to exploit advantages of both the configuration

Only a small fraction of I_S is routed through diode

The output gain is better than diode load

The output common mode voltage is also fixed

Cascode configuration



TELESCOPIC CASCODE

Cascoding increases the output resistance significantly

$$R_1=(g_{m5}r_{o7})r_{o5}$$

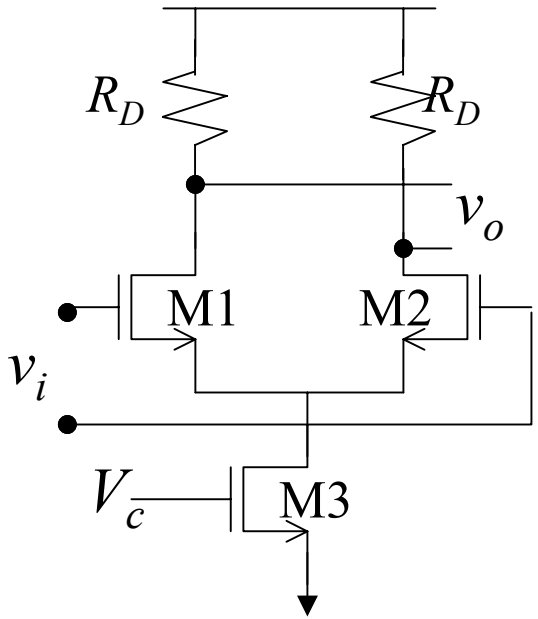
$$R_2=(g_{m3}r_{o1})r_{o3}$$

$$A_v=-\frac{g_{m1}}{(g_{m5}r_{o7})r_{o5}+(g_{m3}r_{o1})r_{o3}}$$

Stacking transistors reduces the voltage swing

Gilbert Cell

2 Quadrant Multiplier



$$\frac{v_o}{v_i} = -g_m R_D$$

$$v_o = -g_m R_D v_i$$

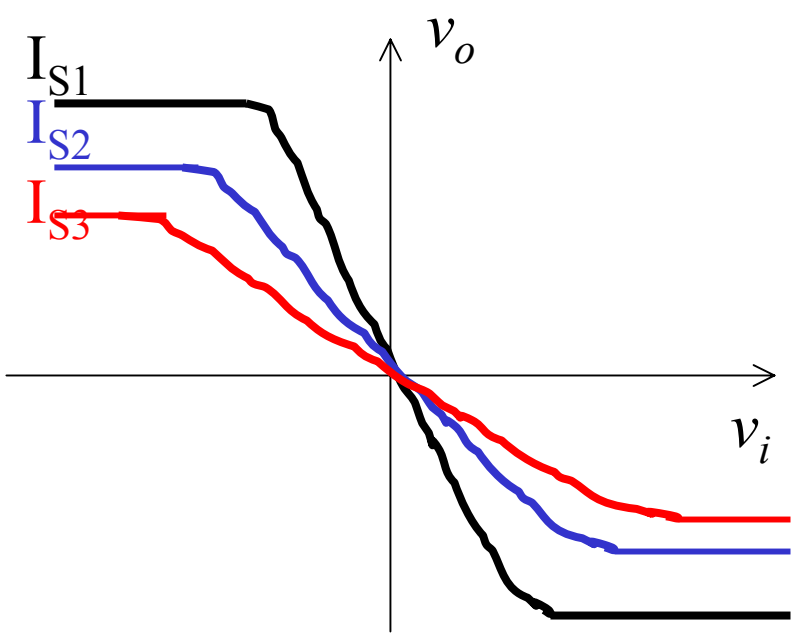
$$v_o = -f(V_c) R_D v_i$$

$$v_o = -\alpha V_c R_D v_i$$

Assuming $g_m = \alpha V_c$

Also functions as Variable Gain Amplifier (VGA)

The transfer function



$$v_o = -\alpha V_c R_D v_i$$

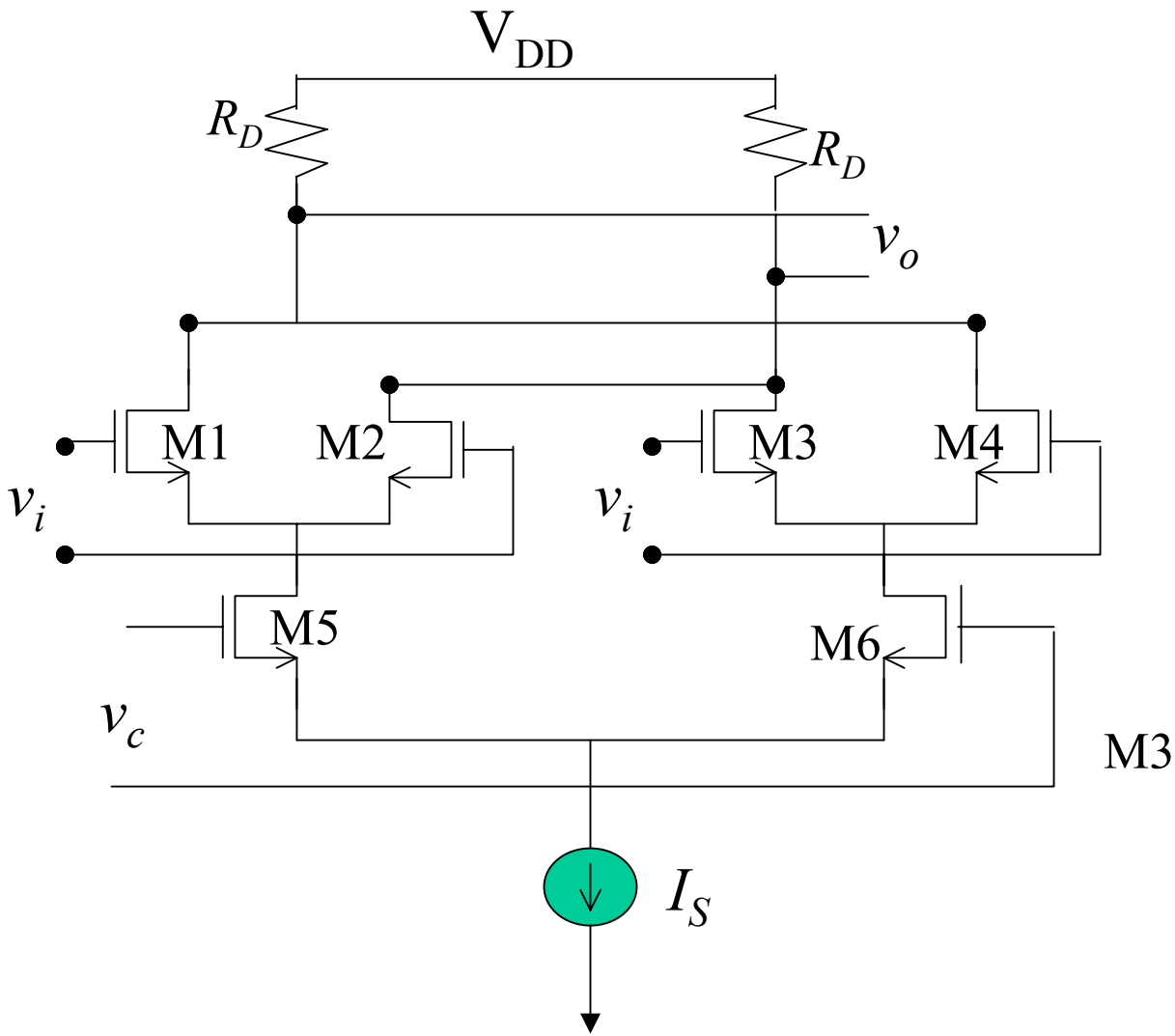
It is two quadrant multiplier
Because Vc can't be negative

Note that Vc is the DC voltage and not the AC small signal voltage

In order to build the 4 quadrant multiplier v_c should be AC voltage

Construct two VGAs and combine the output!

Gilbert Multiplier



The circuit analysis

$$v_o = R_D(i_{d1} + i_{d4}) - R_D(i_{d2} + i_{d3})$$

Case 1: $v_c = 0$, $v_i \neq 0$; *then* $i_{d1} = -i_{d4}$ *and* $i_{d2} = -i_{d3}$ *Hence* $v_o = 0$

Case 2: $v_i = 0$, $v_c \neq 0$; *then* $i_{d1} = i_{d2}$ *and* $i_{d3} = i_{d4}$ *Hence* $v_o = 0$

Case 3: $v_i \neq 0$, $v_c \neq 0$;

$$\Delta g_{m1} = \Delta g_{m2} = \alpha \frac{v_c}{2} \qquad \Delta g_{m3} = \Delta g_{m4} = -\alpha \frac{v_c}{2}$$

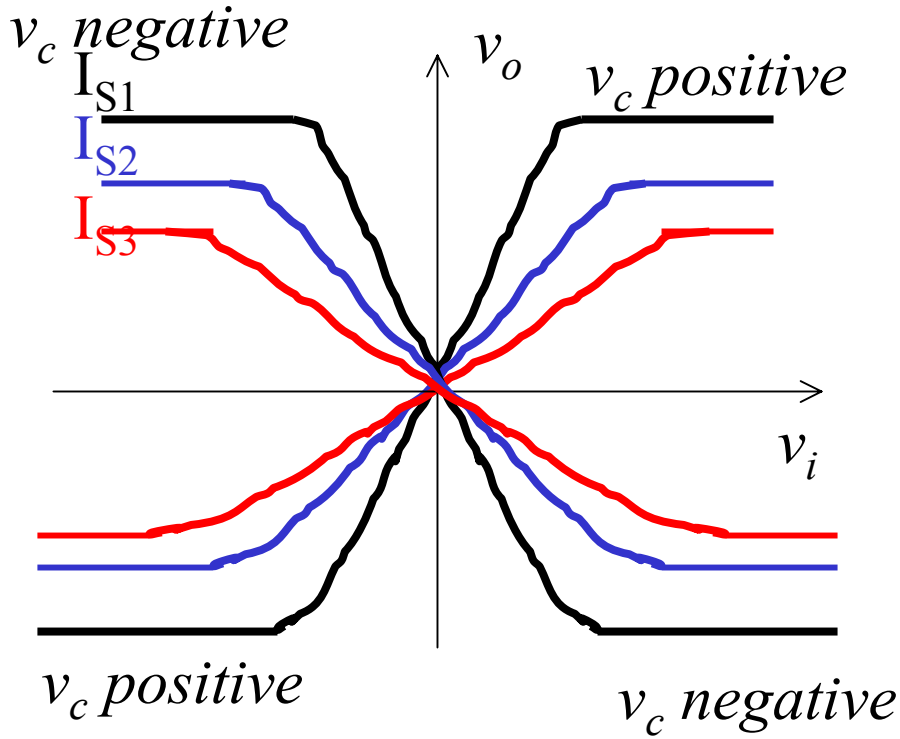
$$i_{d1} + i_{d4} = \frac{v_i}{2}(\Delta g_{m1} - \Delta g_{m4}) = \frac{v_i}{2}\left(\frac{\alpha v_c}{2} + \frac{\alpha v_c}{2}\right) = \frac{\alpha v_i v_c}{2}$$

$$i_{d2} + i_{d3} = -\frac{v_i}{2}(\Delta g_{m2} - \Delta g_{m3}) = -\frac{v_i}{2}\left(\frac{\alpha v_c}{2} + \frac{\alpha v_c}{2}\right) = -\frac{\alpha v_i v_c}{2}$$

$$v_o = \alpha R_D v_i v_c$$

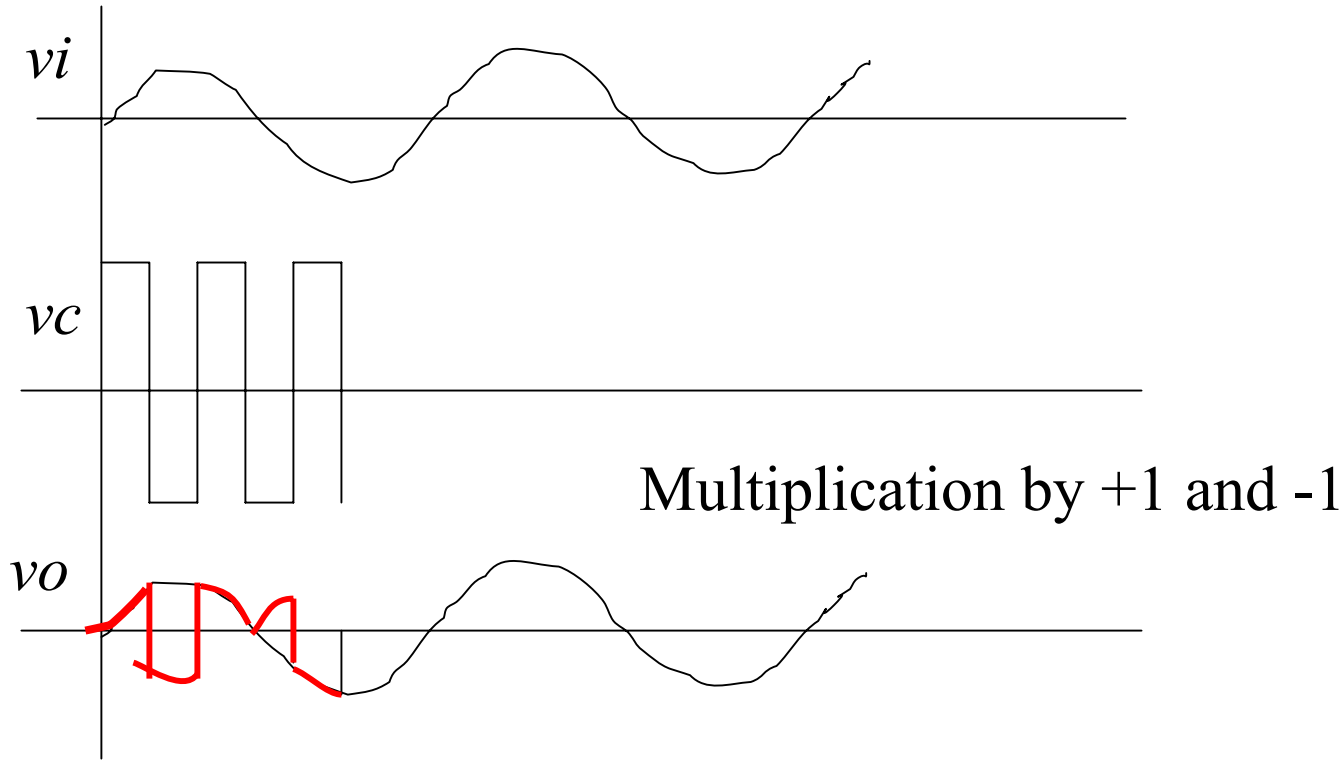
The circuit acts as 4 quadrant multiplier for small signal v_i, v_c

Transfer curves



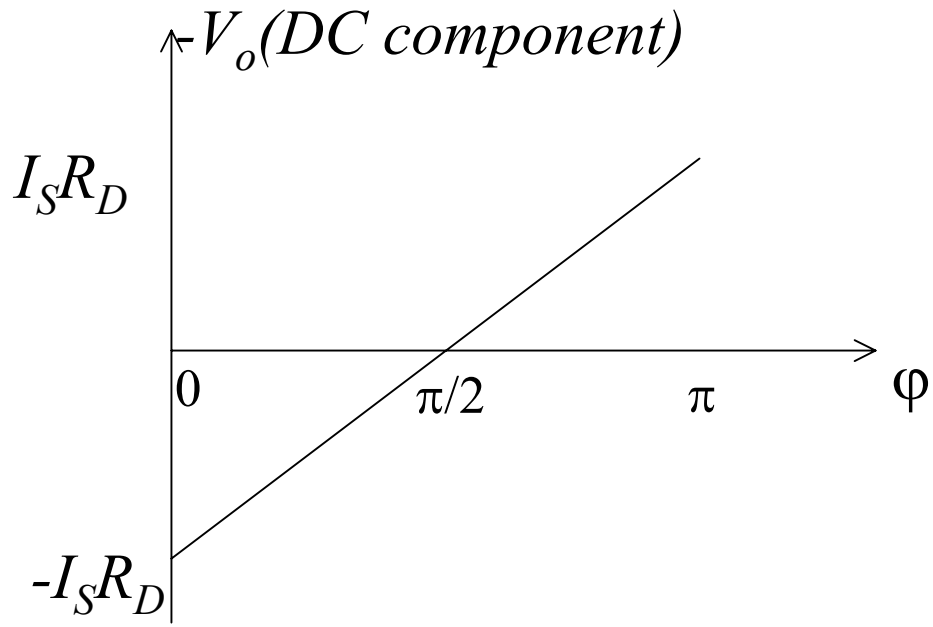
Balanced Modulator

Among *vc* and *vi*, when one of them is small signal and the other is large signal square wave, the circuit acts like a modulator



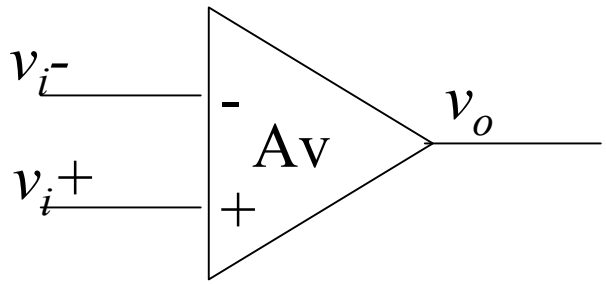
Phase Detector

When both *vi* and *vc* are large signal square waves, the circuit functions as phase detector with the DC component of the output voltage proportional to the phase difference



Design of 2 stage OPAMP

Ideal OPAMP



Infinite differential gain

Infinite input impedance

Zero output impedance

Zero input current

Zero common mode gain

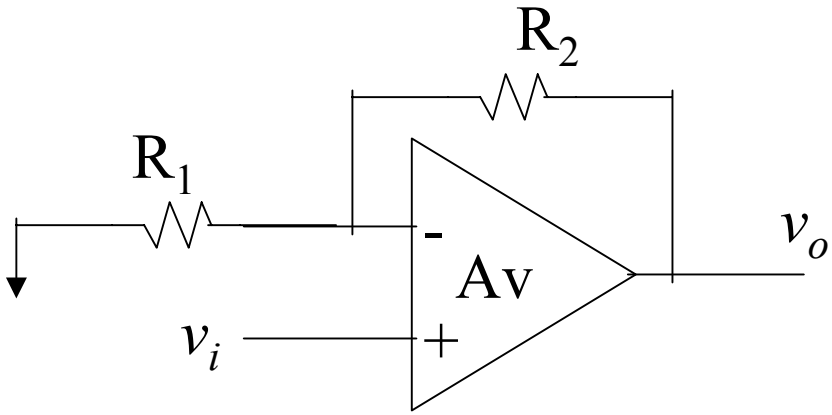
Unfortunately Ideal OPAMP does not exist in reality!

Further attempts to reach ideality with these parameters will have trade off with respect to speed, power, voltage swings etc

We will treat OPAMP as a "high gain differential amplifier" designed with an adequate performance metrics for a given application at hand

Parameters of interest-Open loop gain

OPAMPs are invariably used with closed loop negative feedback



For an ideal OPAMP (Av=∞), the closed loop gain can be set by only resistance ratio

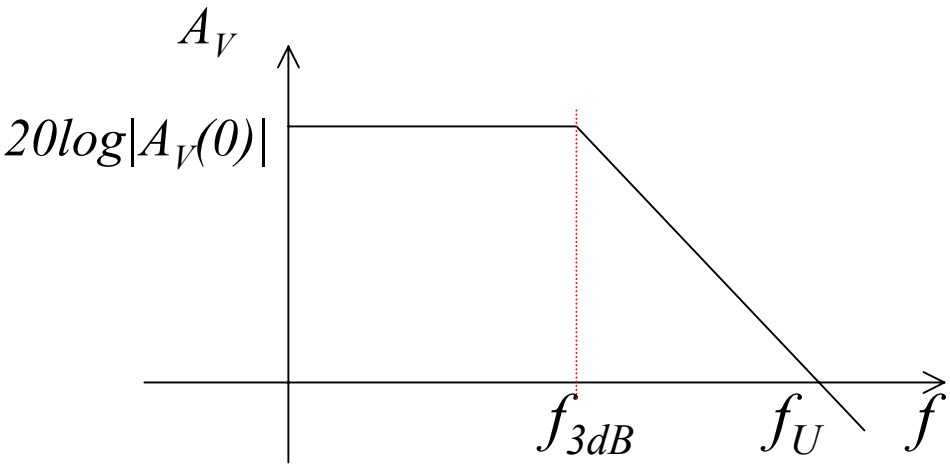
$$\frac{v_o}{v_i} = 1 + \frac{R_2}{R_1}$$

Suppose *R2/R1*=9 and *Av*≠∞ and it is required to have less than 0.1% error in the gain. Then what is the minimum Av required?

$$\frac{v_o}{v_i} = 1 + \frac{R_2}{R_1}\left(1 - \frac{R_1 + R_2}{R_1}\frac{1}{Av}\right)$$

For gain error < 0.1%, the open loop gain $Av > 10{,}000$

Parameters of interest-small signal bandwidth



The open loop gain drops at higher frequency resulting in an increased error for the closed loop feedback system

Also the large signal settling time for the closed loop system depends on the open loop unity gain frequency

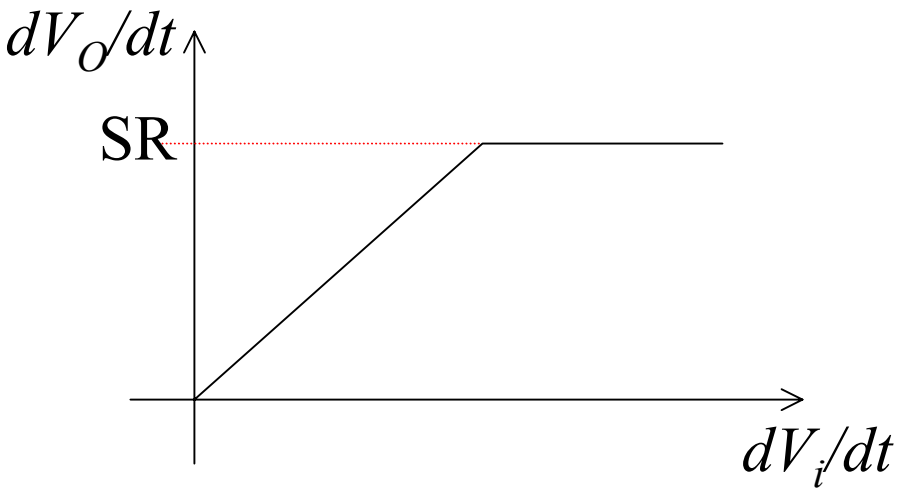
$$\tau = \frac{A_{CL}}{A_v(0)\omega_{3dB}} = \frac{A_{CL}}{\omega_U}$$

Parameters of interest-slew rate

Determines the large signal behaviour

It gives the highest rate of change of input beyond which the output does not respond instantaneously

Need very large slew rate for linearity



Parameters of interest

Output swing: Trade off between O/P swing, bias current and gain

Linearity:

Differential implementation to suppress even harmonics
Allow significant open loop gain so that closed loop feedback system achieves required linearity

Noise : Thermal noise and 1/f noise

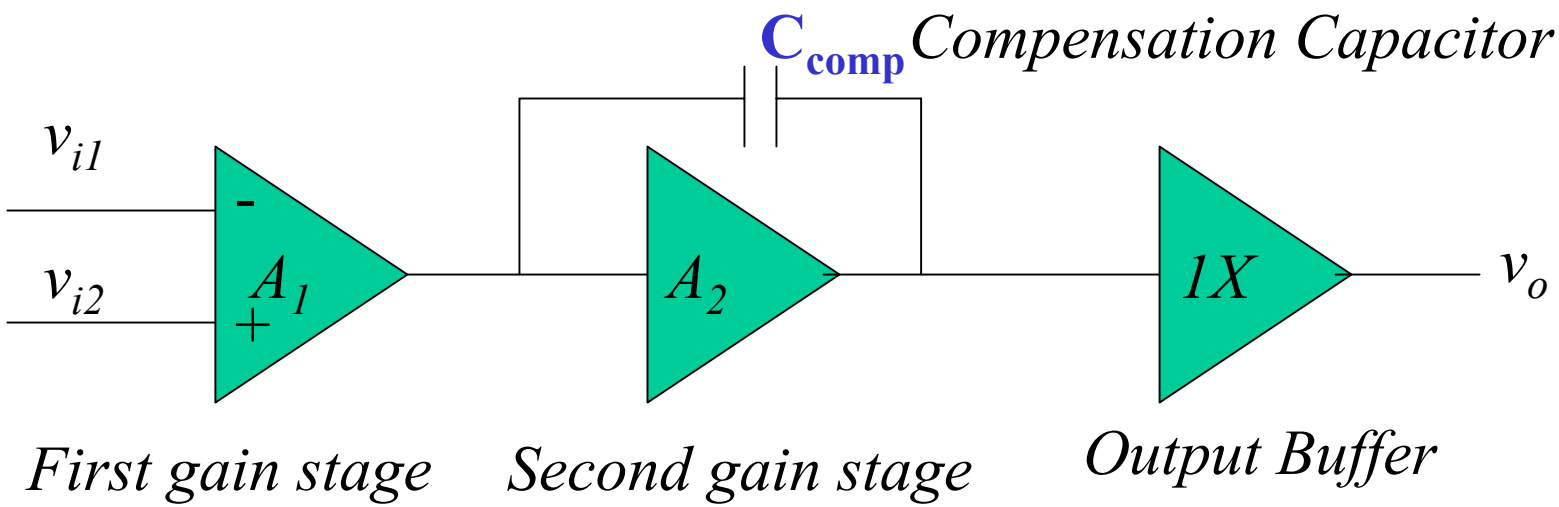
Offset : Systematic and Random offset

Output load : Typical on-chip OPAMP applications mostly have very low capacitive load < 1pF
Stand alone OPAMPs may have to drive high capacitive and low resistance loads

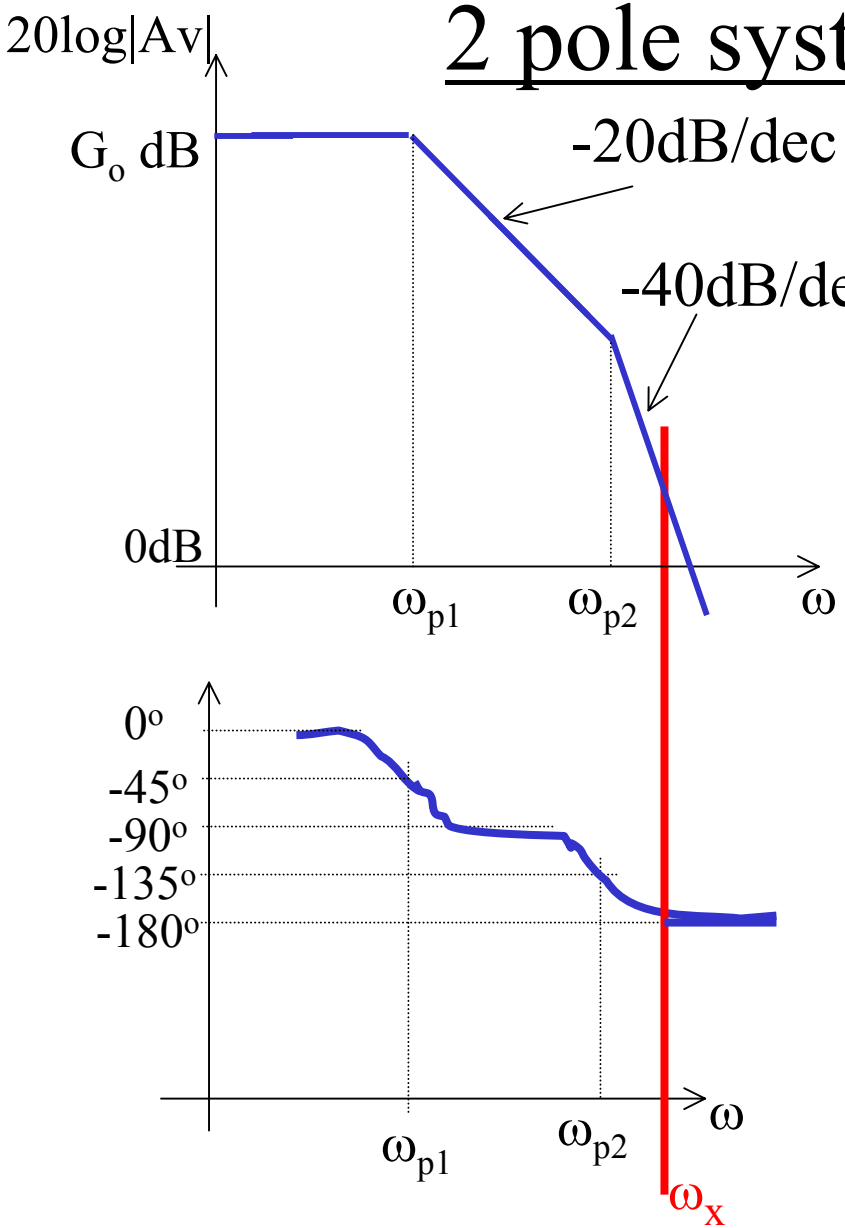
Basic 2 stage OPAMP

Most of the OPAMP designs have two gain stages

Unless absolutely desired, more gain stages should be avoided
Since the frequency compensation becomes complex due to
Multiple dominant poles



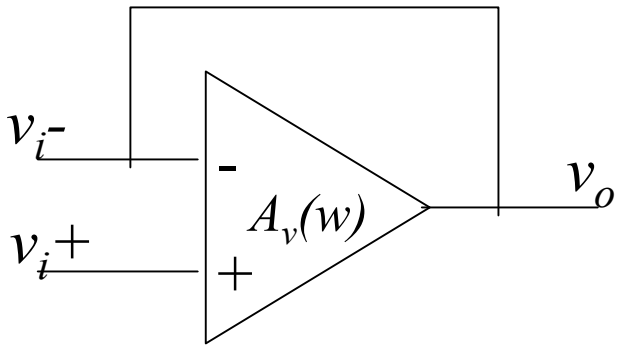
2 pole system response



$$H(s) = \frac{1}{\left(1 - s/p_1\right)\left(1 - s/p_2\right)}$$

If Av(0) is very large and if ω_{p2} is close to ω_{p1}, Then it is likely that at at some frequency ω_x, the phase shift will be -180^o but the gain will still be greater than unity

Implication in closed loop negative feedback system



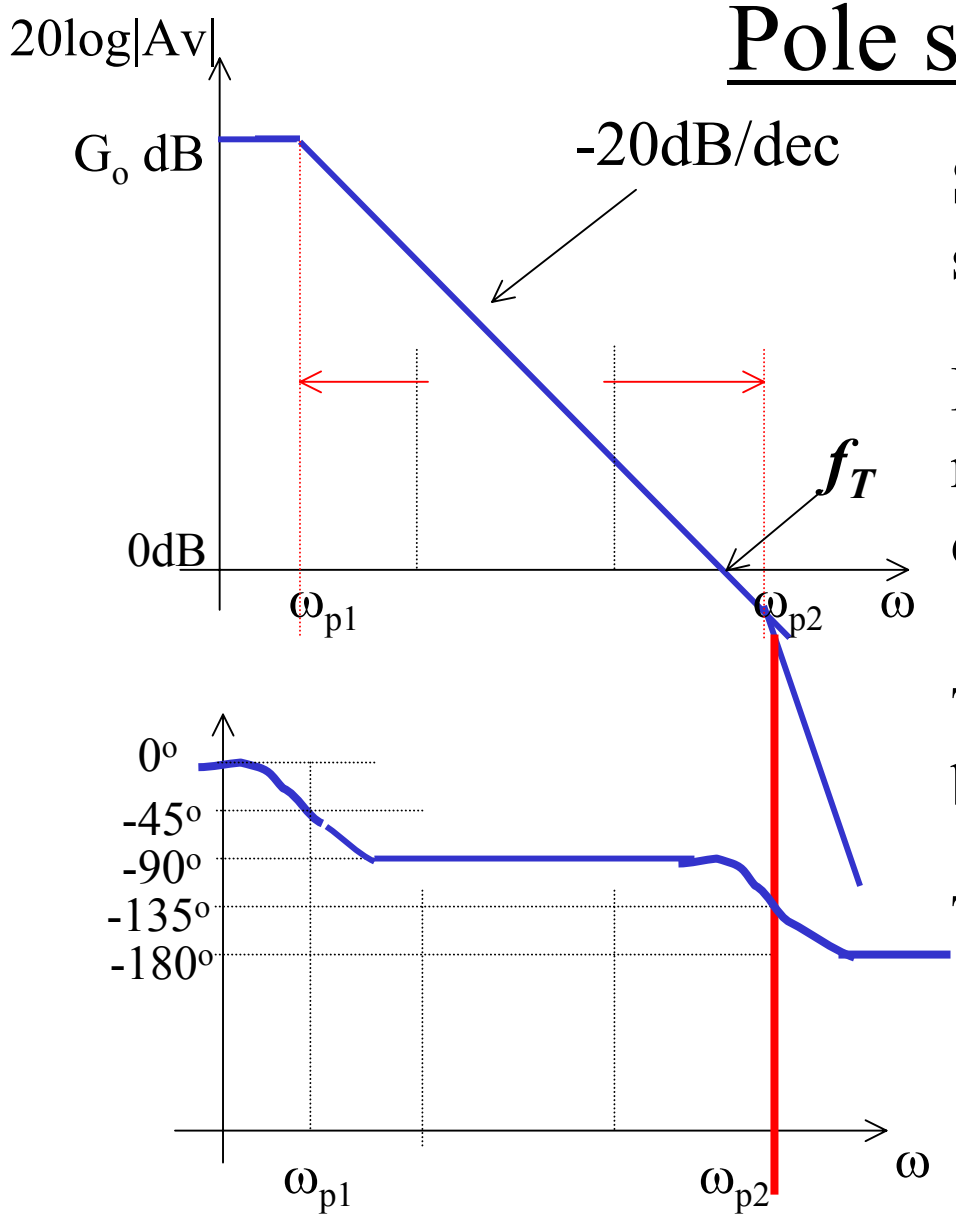
At DC and low frequency there is a phase shift –180^o between the Input v_i and the output v_o
(This is due to the inversion between Gate and Drain voltage of Transistor)

If $Av(\omega)$ is a two pole transfer function, the poles introduce and additional phase shift of –180^o at ω_x

The negative feedback system gets converted to a positive feedback system!!

The system becomes unstable and oscillatory

Pole splitting



Split the nearby poles far apart by some technique

By the time the second pole is reached the gain has already dropped below unity

The closed loop feedback system becomes a stable system

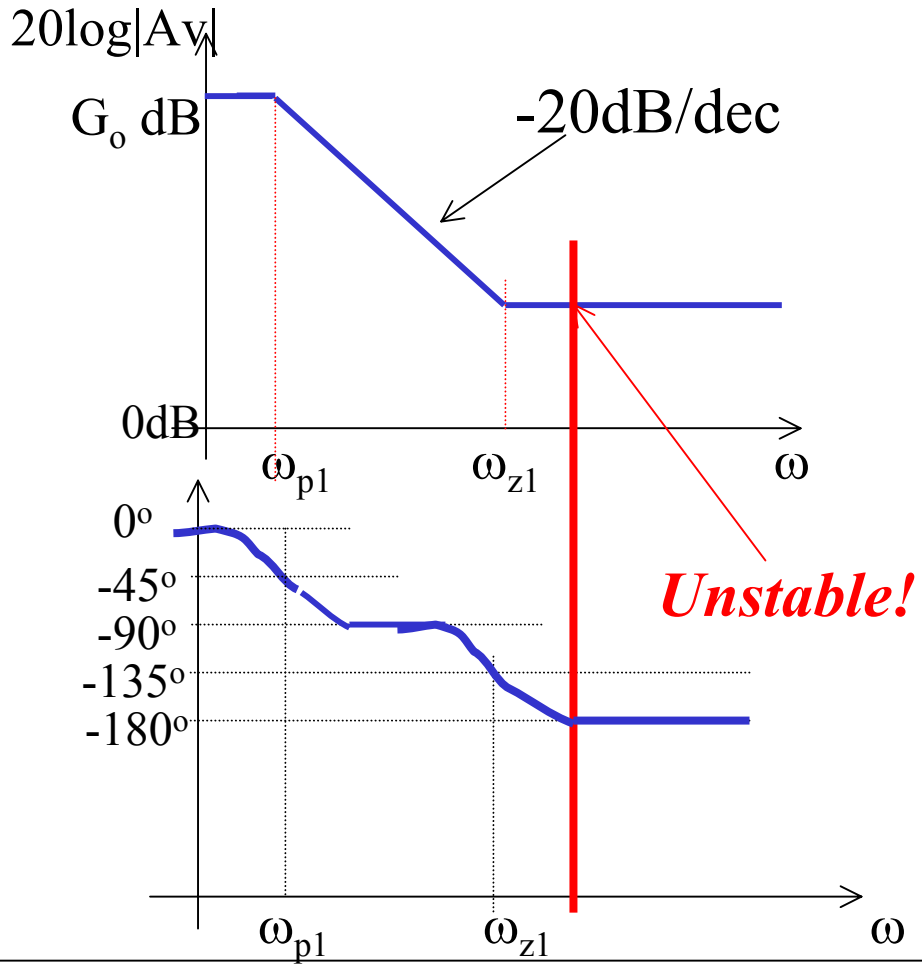
The phase margin is defined as

$$PM = \phi(\omega_t) + 180^o$$

PM should be positive for stability

Positive Zero can be more dangerous

$$H(s) = \frac{(s - z_1)}{(s + p_1)} \qquad \phi = -\tan^{-1}\left[\frac{w}{w_{z1}}\right] - \tan^{-1}\left[\frac{w}{w_{p1}}\right]$$

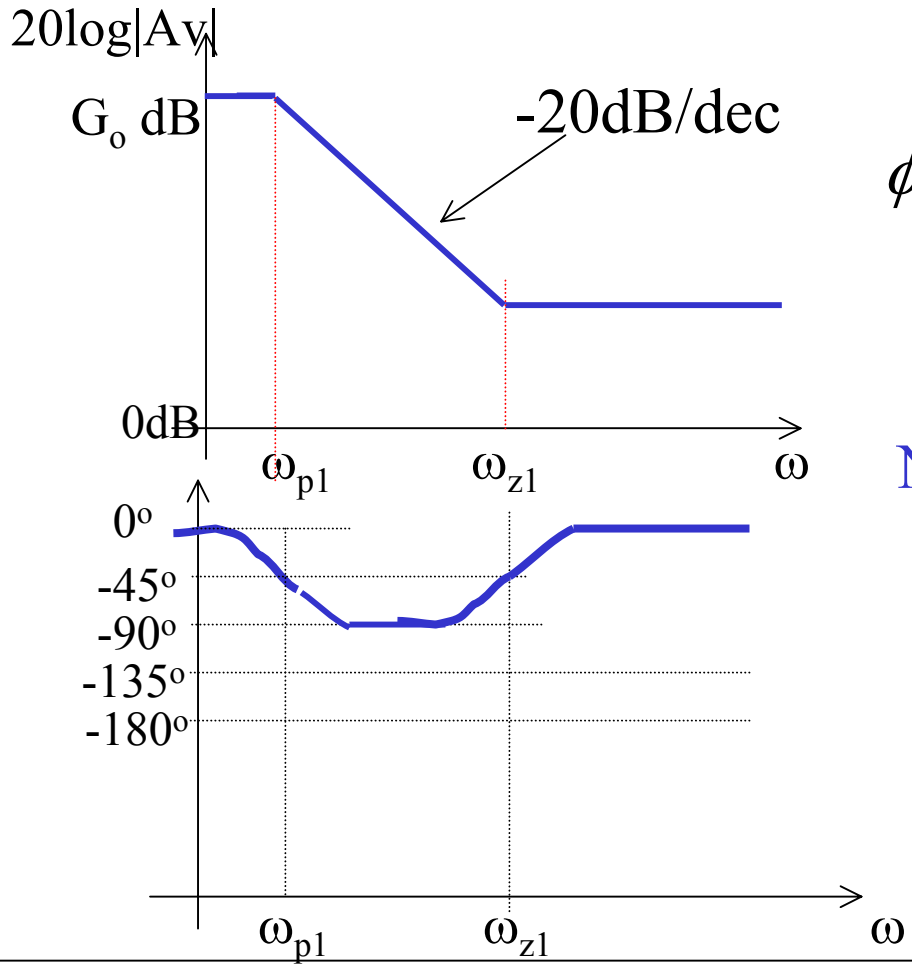


Positive zero can make the system unstable

If the positive zero exists nearby dominant pole, then cancel the zero

Negative zero is good

$$H(s) = \frac{(s + z_1)}{(s + p_1)}$$



$$\phi = \tan^{-1}\left[\frac{w}{w_{z1}}\right] - \tan^{-1}\left[\frac{w}{w_{p1}}\right]$$

Negative zero can increase PM

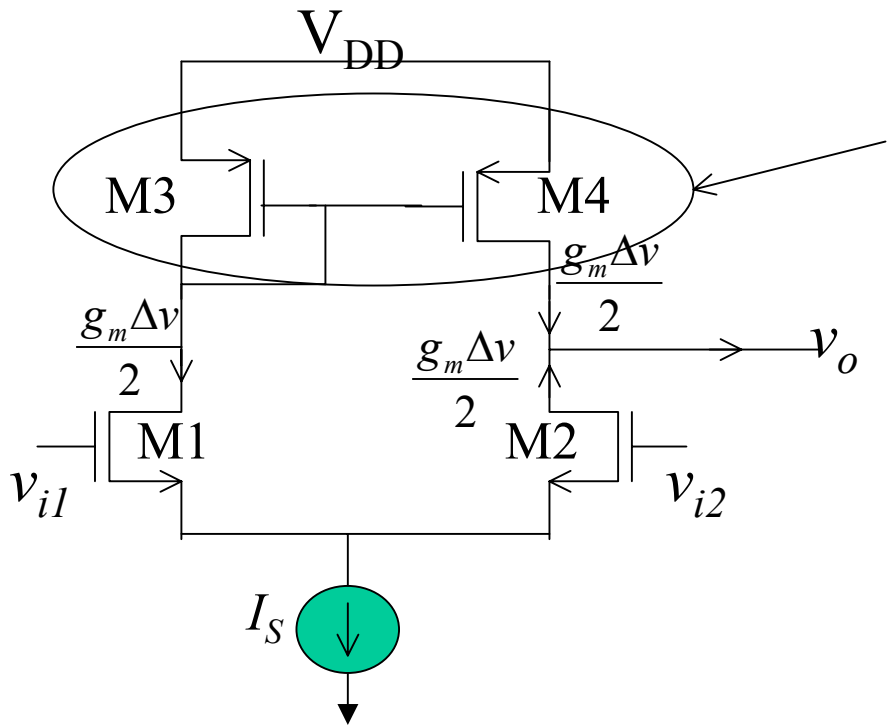
First stage for the 2 stage OPAMP

The first stage should do two tasks

Produce reasonably large gain

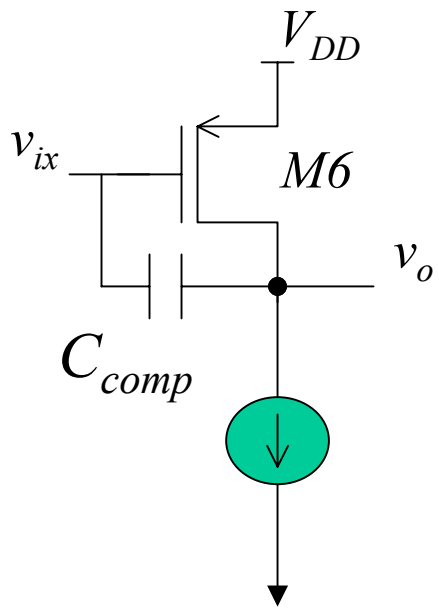
Perform differential to single ended conversion

This is done using differential amplifier with active current mirror load to enhance the gain



Active current mirror adds currents in two branches and doubles the gain while performing differential to single ended conversion

Second stage for the 2 stage OPAMP

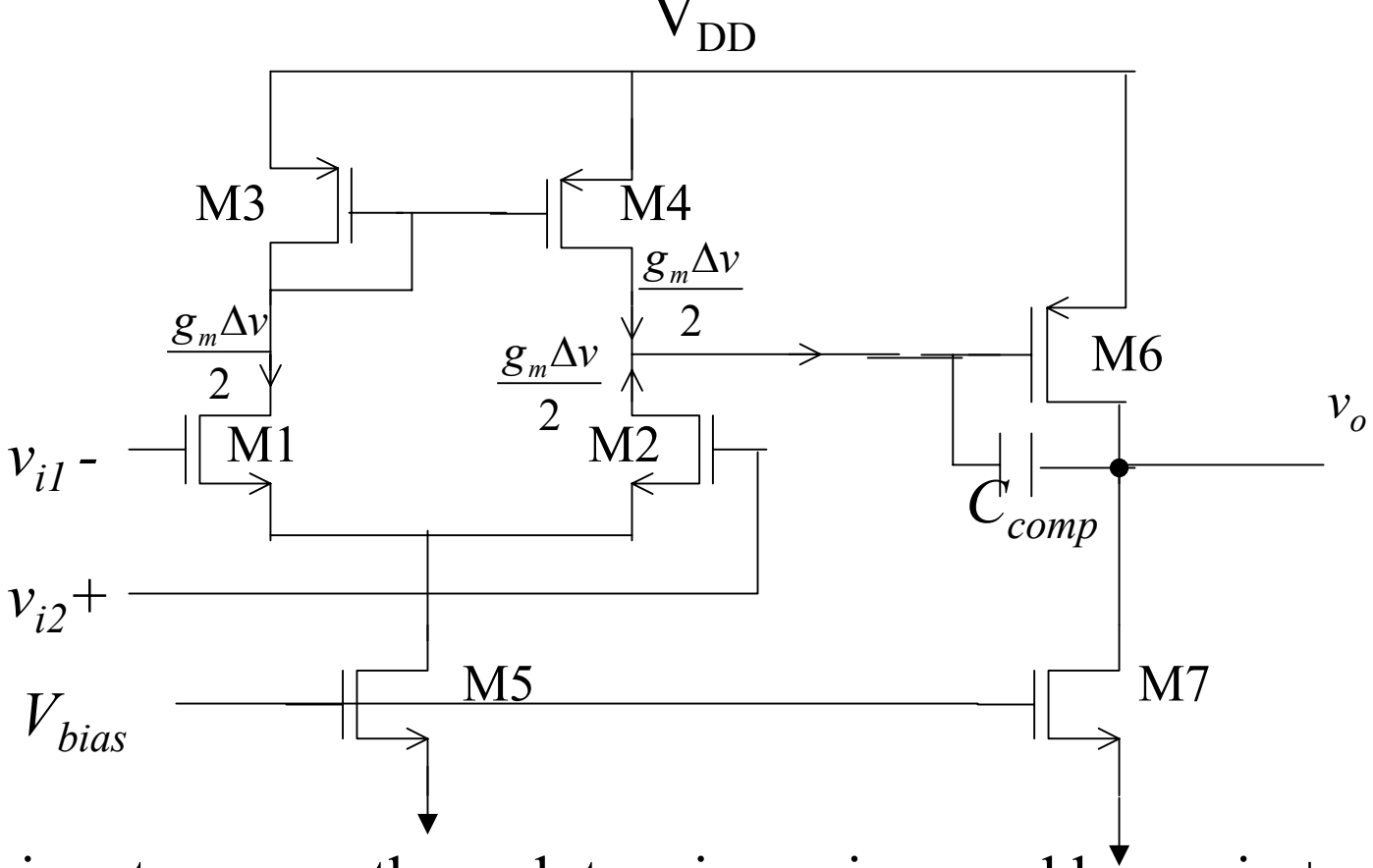


To 1X buffer

PMOS common source amplifier
with the current source load

C_{comp} performs frequency compensation

Two stage OPAMP circuit

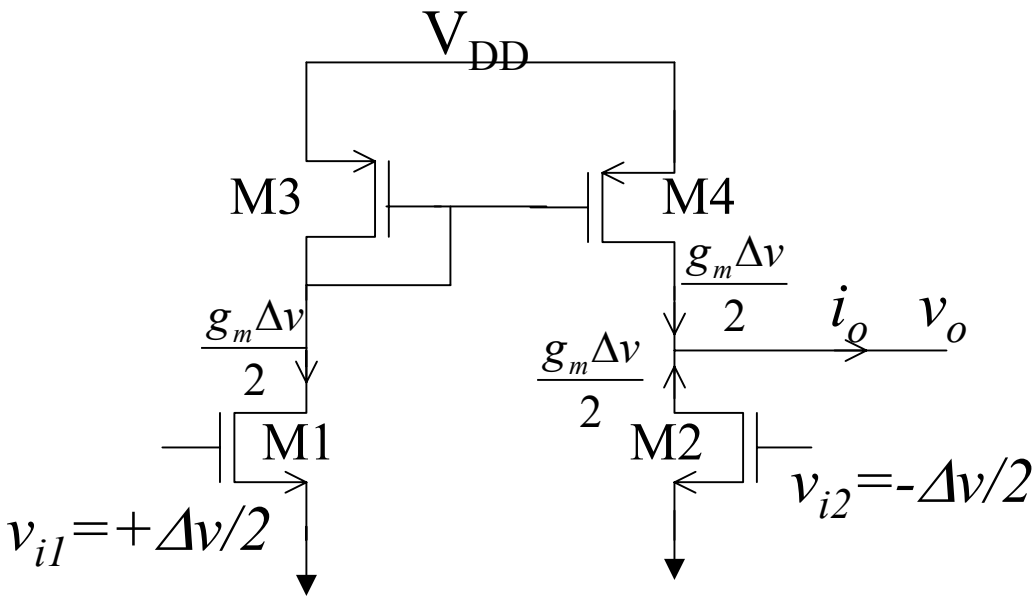


The input v_{i2} goes through two inversions and hence is +ve *i/p*

The input v_{i1} goes through one inversions and hence is -ve *i/p*

Note that the AC voltages at drain of M1 and M2 will be quite different

First stage Low Frequency Differential Gain



$$g_{m1}=g_{m2}=g_m$$

$$i_o = g_m \Delta v$$

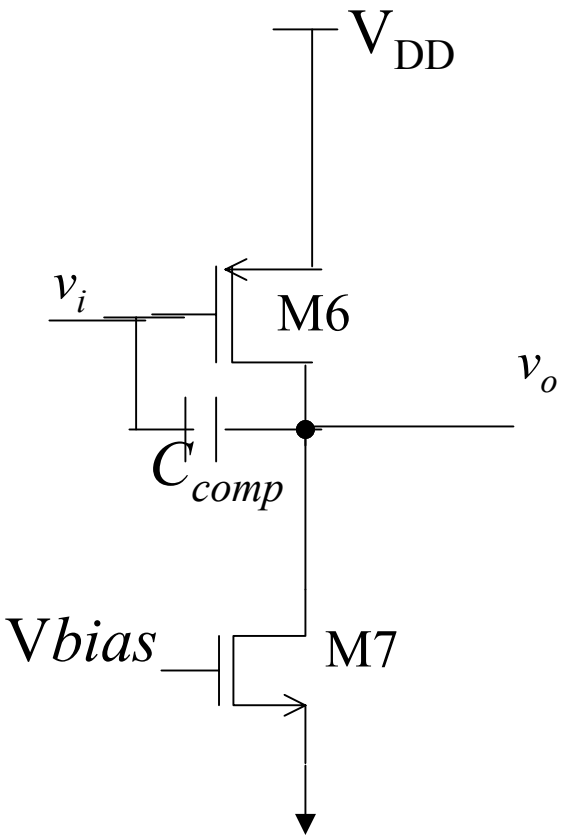
$$R_o = \frac{1}{Go}$$

$$R_o = \frac{1}{g_{o2} + g_{o4}}$$

$$v_o = i_o R_o = \frac{g_m \Delta v}{g_{o2} + g_{o4}}$$

$$A_1 = \frac{v_o}{\Delta v} = \frac{g_m}{g_{o2} + g_{o4}}$$

Second stage Low Frequency Differential Gain



Common source PMOS amplifier with current source load

$$A_2 = \frac{v_o}{v_i} = -\frac{g_{m6}}{g_{o6} + g_{o7}}$$

Combined two stage differential response

$$A_v(0) = -\frac{g_m}{g_{o2} + g_{o4}} \frac{g_{m6}}{g_{o6} + g_{o7}}$$

$$A_v(0) = -\sqrt{\mu_n C_{ox} \frac{W}{L} I_{d1}} \frac{1}{I_{d1}(\lambda_2 + \lambda_4)} \sqrt{\mu_p C_{ox} \frac{W_6}{L_6} I_{d6}} \frac{1}{I_{d6}(\lambda_6 + \lambda_7)}$$

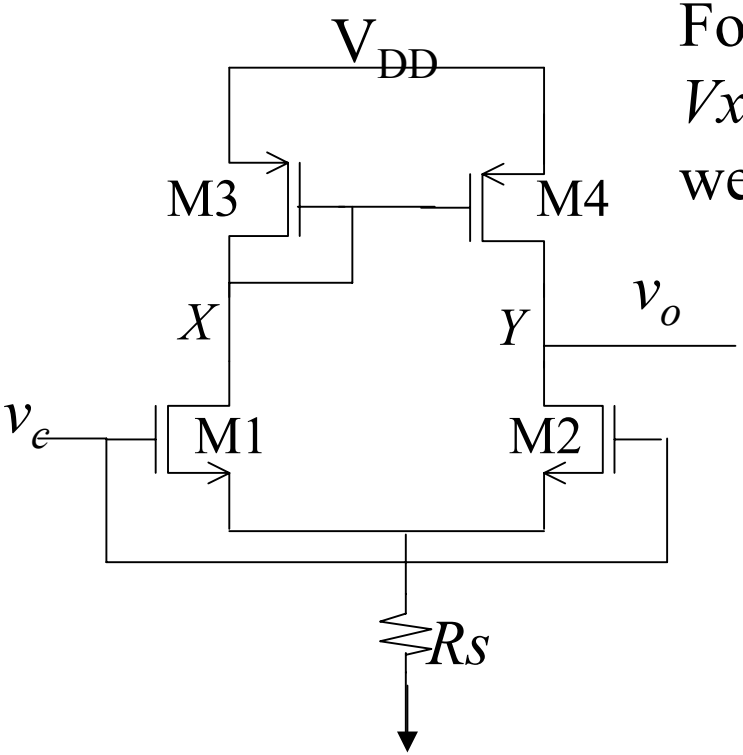
$$A_v(0) = -\frac{1}{I_d} C_{ox} \frac{W}{L} \sqrt{\mu_n \mu_n} \frac{1}{(\lambda_2 + \lambda_4)(\lambda_6 + \lambda_7)} \qquad \textit{For Id1=Id6}$$

$$A_v(0) = -\frac{2}{(\lambda_2 + \lambda_4)(V_{gsn} - V_{tn})} \frac{2}{(\lambda_6 + \lambda_7)(V_{gsp} - V_{tp})}$$

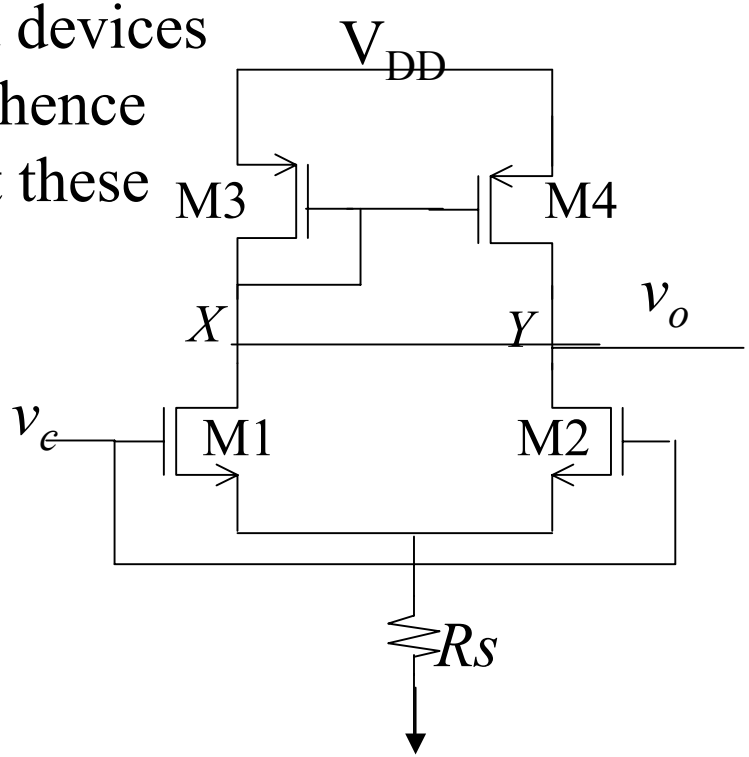
For high gain either use small *Id* or large device dimension

Small *Id* impacts slew rate, large *W/L* impacts area and input capacitance

First stage Common mode Gain



For matched devices
$Vx = Vy$, and hence
we can short these

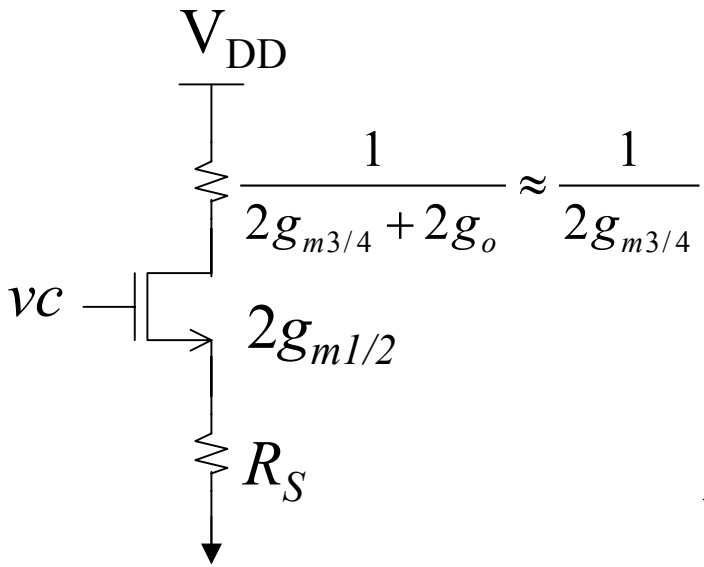


$g_{m1}=g_{m2}=g_m$

$g_{m1}=g_{m2}=g_m$

This brings M1 in parallel M2 and M3 in parallel M4

Equivalent circuit for CM response



This configuration now looks like common source amplifier with source degeneration *Rs* and drain Resistance $R_D = 1/2g_{m3/4}$

$$Ac_1 = -\frac{R_D}{R_s} = -\frac{1}{2g_{m3/4}R_S}$$

The second stage simply amplifies this further
The two stage common mode gain is

$$Ac = Ac_1 Ac_2 = -\frac{1}{2g_{m3/4}R_S}\frac{g_{m6}}{g_{o6} + g_{o6}}$$

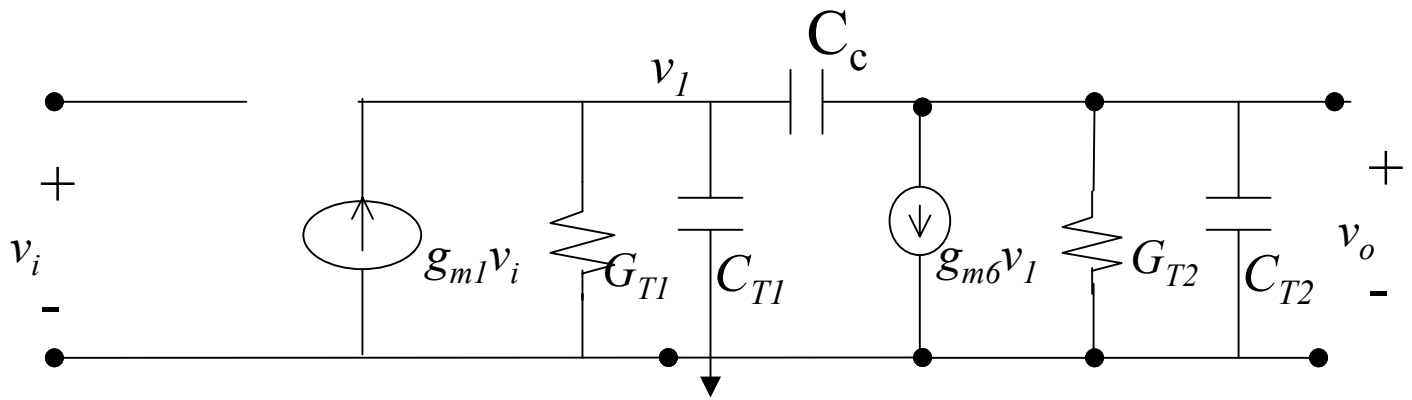
Common Mode Rejection Ratio

$$CMRR = \frac{A_d}{A_c} = \frac{g_{m1} \big/ (g_{o1} + g_{o2})}{1 \big/ 2g_{m3/4}R_s}$$

$$CMRR = \frac{2g_{m1}g_{m3}R_s}{g_{o1} + g_{o2}}$$

CMRR is essentially determined by the first stage

High frequency equivalent circuit



$$G_{T1} = 1/R_{T1} = g_{o2} + g_o$$

$$G_{T2} = 1/R_{T2} = g_{o6} + g_{o7}$$

$$C_{T1} = C_{db2} + C_{db4} + C_{gs6}$$

$$C_{T2} = C_{db6} + C_{db7} + C_L$$

$$C_C = C_{gd6} + C_{comp}$$

High frequency Response

Writing the nodal equations for equivalent circuit and solving for the gain, we obtain poles and one zero

$$p_1 = -\frac{1}{g_{m6} R_{T1} R_{T2} C_c}$$

$$p_1 = -\frac{g_{m6} C_c}{C_{T1} C_c + C_{T2} C_c + C_{T1} C_{T2}} \approx \frac{g_{m6}}{C_{T1} + C_{T2}}$$

$$z_1 = \frac{g_{m6}}{C_c}$$

p_1 needs to be made a dominant pole by appropriately choosing the compensation capacitance

z_1 is a positive zero that can impact stability

Pole splitting and choice of Cc

$$p_1 = -\frac{g_{m1}}{g_{m1} g_{m6} R_{T1} R_{T2} C_c} = -\frac{1}{|Av(0)|}\frac{g_{m1}}{C_c}$$

Cc should be chosen such that the unity gain frequency $\omega_u << p_2$ to get adequate Phase Margin

Also for the single pole response (with p_1), the unity gain frequency is given by

$$\omega_u = \frac{g_{m1}}{C_c}$$

$$C_c = \frac{g_{m1}}{\omega_u}$$

Feed forward zero

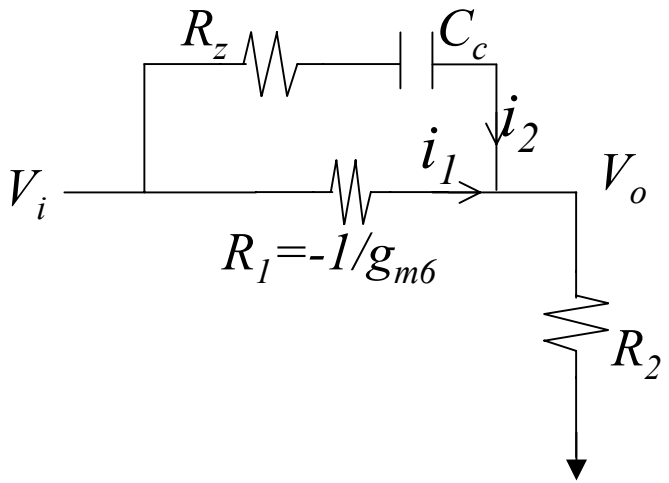
The positive zero location is very close to ω_u and this will degrade the phase margin

$$z_1 = \frac{g_{m6}}{C_c}$$

If $g_{m6} \leq g_{m1}$ then $z_1 \leq \omega_u$

This zero should be cancelled, otherwise the system becomes unstable

Zero cancellation with nulling resistance



$$\frac{V_o}{V_i} = \frac{R_2\left[1 + sC_c(R_1 + R_z)\right]}{R_1 + R_2 + sC_c(R_1R_2 + R_1R_z + R_2R_z)}$$

$$z_1 = \frac{1}{C_c(-R_1 - R_z)}$$

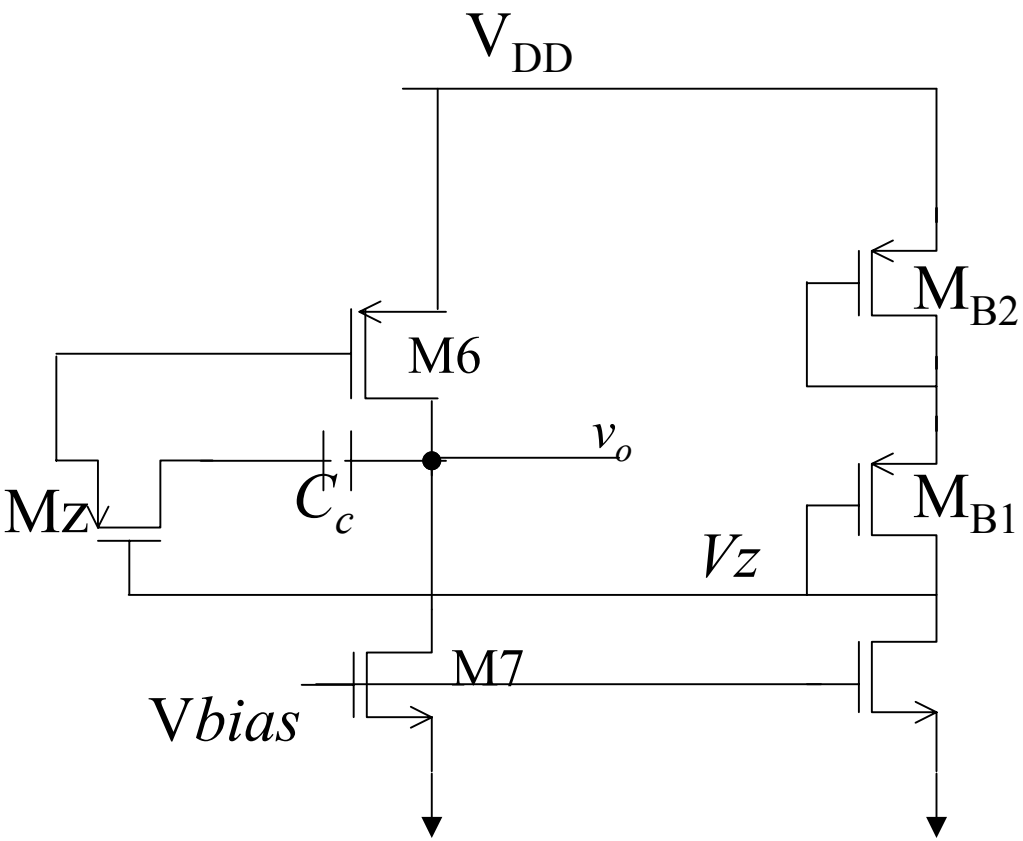
$$z_1 = \frac{1}{C_c(1/g_{m6} - R_z)}$$

Qualitatively, adding R_z makes i_2 weaker at any given frequency compared to the value of i_1 , i.e. the effect of feed forward zero is suppressed

When $R_z = 1/g_{m6}$, *the zero is at infinity*

For $R_z > 1/g_{m6}$, *the zero moves to the left half plane improving PM (lead compensation)*

Zero cancellation with PMOS



The PMOS transistor Mz is biased such that it is in linear region

Vgs-Vt > Vds

The linear region *Rz* is

$$R_Z = \frac{1}{\mu_n Cox \, W_z / L_z \, (V_{gsz} - V_{tz})}$$

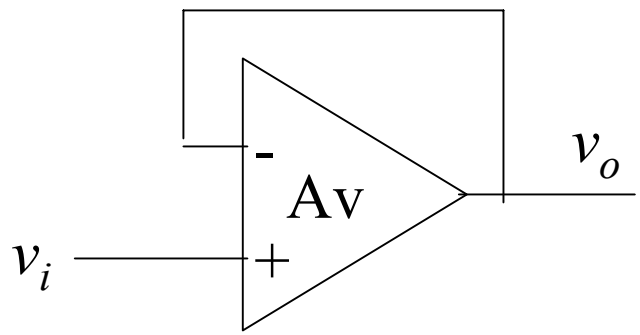
For M6,

$$\frac{1}{g_{m6}} = \frac{1}{\mu_n Cox \, W_6 / L_6 \, (V_{gs6} - V_{t6})}$$

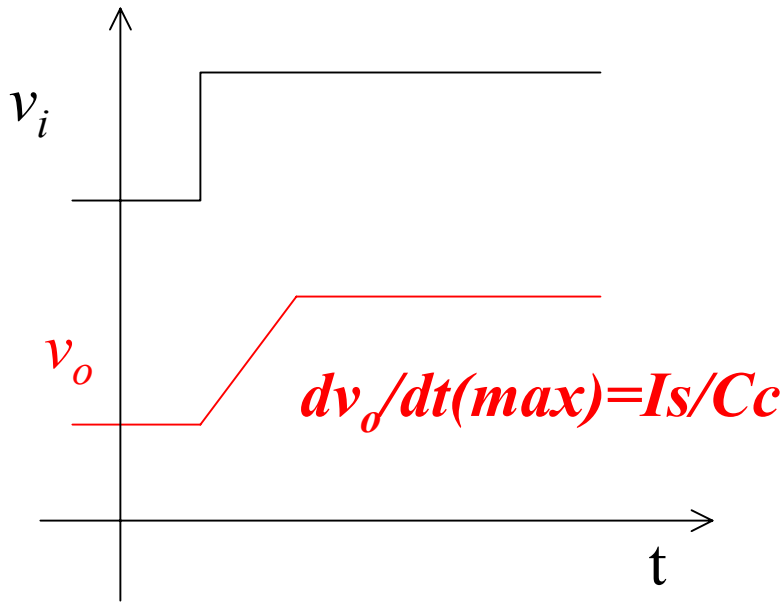
The R_Z is able track $1/g_{m6}$ very well in spite of through process variations

OPAMP Performance Metrics

Slew rate



When a step input is applied, in order for the output to follow the input the capacitance *Cc* should be charged to the new value



The maximum current that is available to charge this capacitor is the bias current *Is*

Hence the slew rate, i.e. the maximum rate of change of output is = *Is/Cc*

Random Input offset

Random offset arises due to transistor mismatch in the supposedly matched differential pair (first stage)

The effect of offset is modeled as an input referred offset voltage in series with the input terminal of ideal OPAMP

$$Vos = \Delta V_{t1,2} + \Delta V_{t3,4}\frac{g_{m3}}{g_{m1}} + \frac{(Vgs - Vt)_{1,2}}{2}\left(-\frac{\Delta W/L_{1,2}}{W/L_{1,2}} - \frac{\Delta W/L_{3,4}}{W/L_{3,4}}\right)$$

$$\Delta V_{t1,2} = V_{t1} - V_{t2}$$

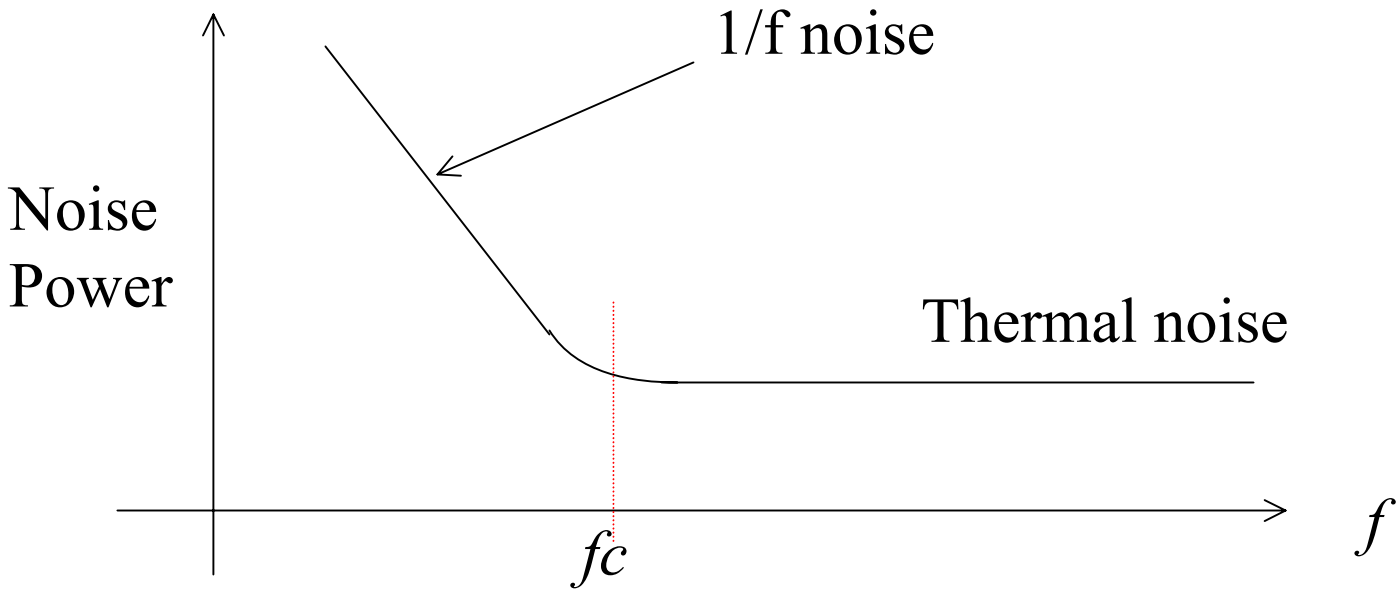
Offset voltage in series with the gate of M1

$$V_t = \frac{V_{t1} + V_{t2}}{2}$$

$$\Delta(W/L) = (W/L)_1 - (W/L)_2$$

$$W/L = \frac{W/L_1 + W/L2}{2}$$

Thermal and *1/f* noise



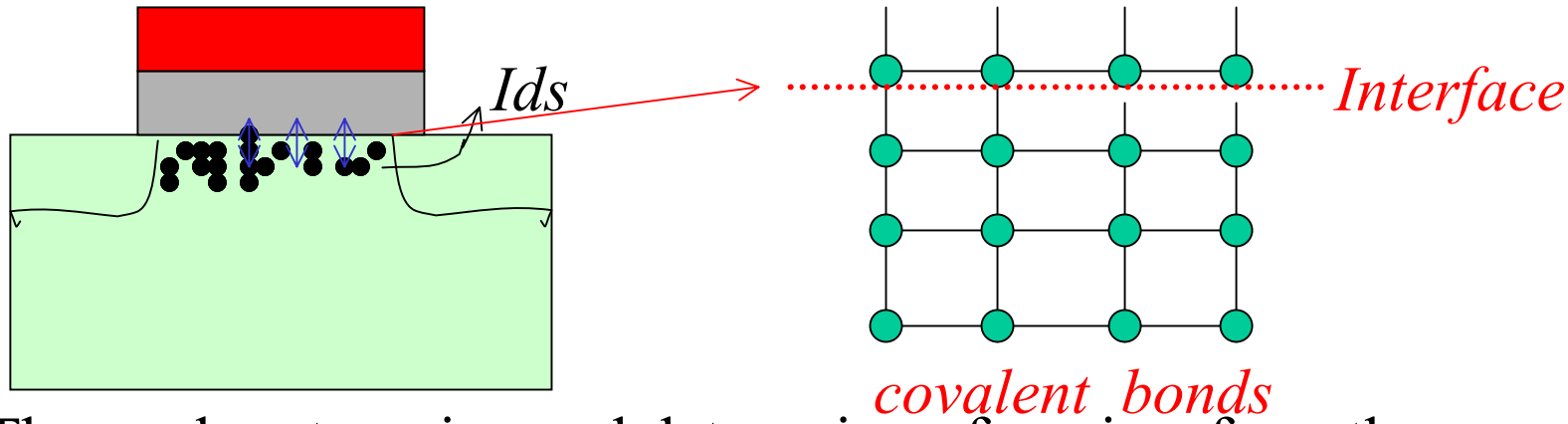
Thermal noise is due to random fluctuation of carriers in a resistor
Higher bias current helps decrease thermal noise

The 1/f noise or flicker noise is due to interface states in MOSFET

fc is 1/f corner frequency in the range of 500KHz

Flicker (*1/f*) noise due to interface states

Flicker noise is due to trapping and detrapping of carriers from the interface states



The random trapping and detrapping of carriers from the channel creates fluctuations in drain current

$$\overline{v_n^{\,2}} = \frac{K}{C_{ox}WL}\frac{1}{f}$$

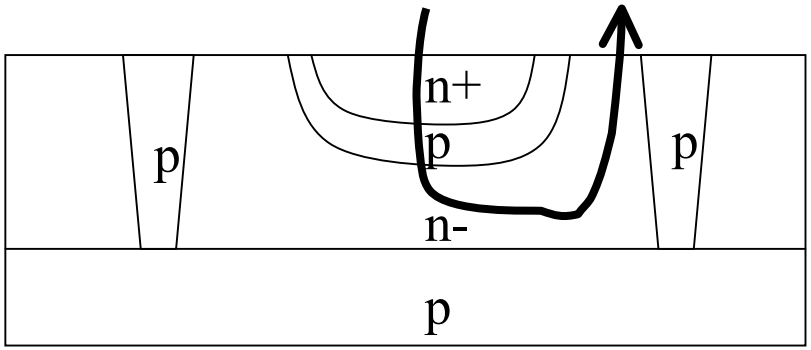
Noise varies as $1/A$ due to averaging effect

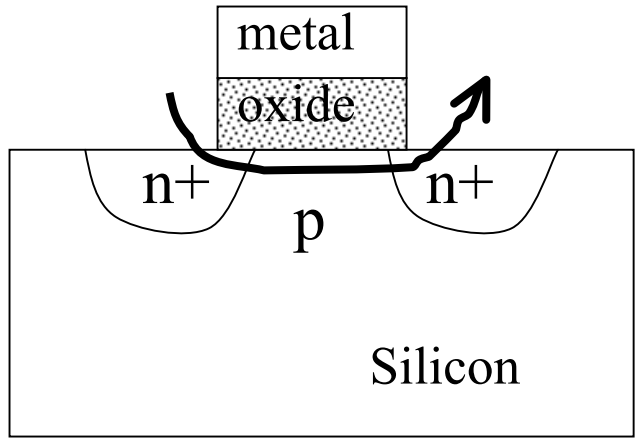
Noise varies as $1/C_{ox}$ since the fraction of charge is less

Noise varies as $1/f$ since traps have certain time constant

MOSFET versus BJT

The BJT circuits are not affected by flicker noise!





The current flow path does not encounter any kinds of defects since the current flow is entirely in the bulk of Silicon

The current flow path is abutting The interface defects region

Dual gate vs Single gate technology

- Historically CMOS technology had a single gate type
- Both NMOS and PMOS had n^+ poly-Si gate
- This was because the poly thickness was fairly large (more than 0.5μm) and it was difficult to activate such poly using implant and annealing
- Hence in-situ doping (i.e. doping during deposition process itself) was invariably used
- As a result both NMOSFET and PMOSFET had n^+ gate
- Almost all the recent technologies use dual poly gate i.e. n^+ gate for NMOS and p^+ PMOS
- The poly is fairly thin (0.2μm or less) and hence the activation is done during s/d implant and anneal step itself

The problem of PMOS V_t setting

$$V_t = V_{fb} + 2\phi_b + \frac{T_{ox}\sqrt{4\varepsilon_s q N_a \phi_b}}{\varepsilon_{ox}}$$

The Vfb term depends on gate material

For NMOS with n^+ gate, Vfb ~ -0.9V, 2ϕb=0.7V, the third term is positive for p- well And hence Vt can be set to low value

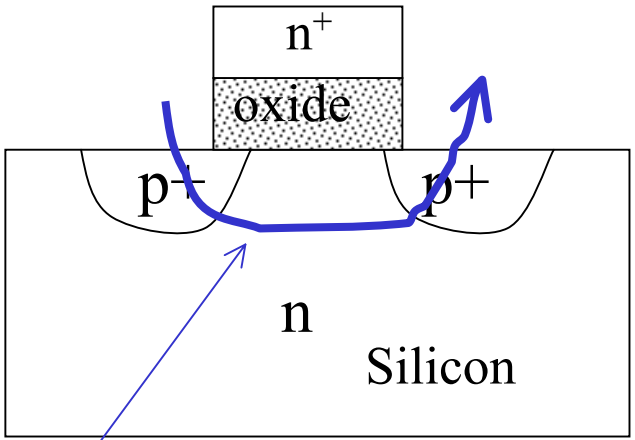
For PMOS with p^+ gate, Vfb ~ +0.9V, 2ϕb= -0.7V, the third term is negative for n- well And hence Vt can be set to low value

For PMOS with n^+ gate, Vfb ~ -0.1V, 2ϕb= -0.7V
hence Vt setting on n-well becomes very difficult

Hence in a single gate technologies the PMOS well was typically counter doped to bring Vt to manageable levels. This process in turn pushes the inversion layer away from interface

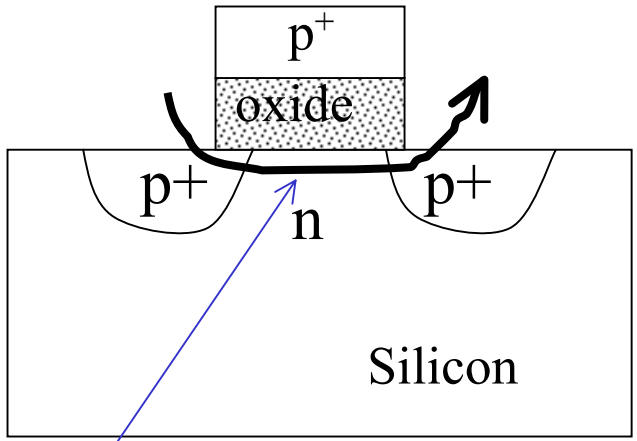
Buried vs Surface Channel PMOSFET

Buried channel PMOSFET



Current flows away from interface

Surface channel PMOSFET



Current flows at the interface

The flicker noise performance of buried channel PMOS is similar to BJT with *almost* zero flicker noise

However the surface channel PMOS is no better than the surface channel NMOSFET

Output Stage

Output stage requirement

Capable of providing high output current to drive large loads

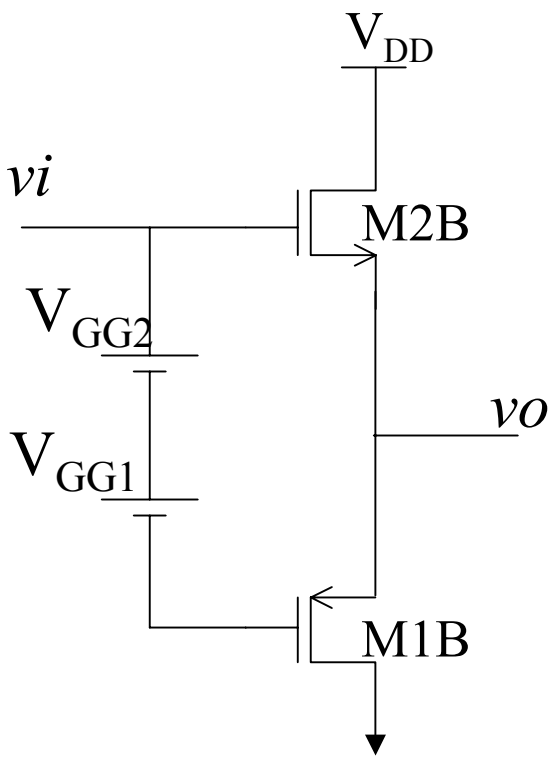
However, the DC bias current should be low to avoid Static power dissipation

The output impedance should be very low

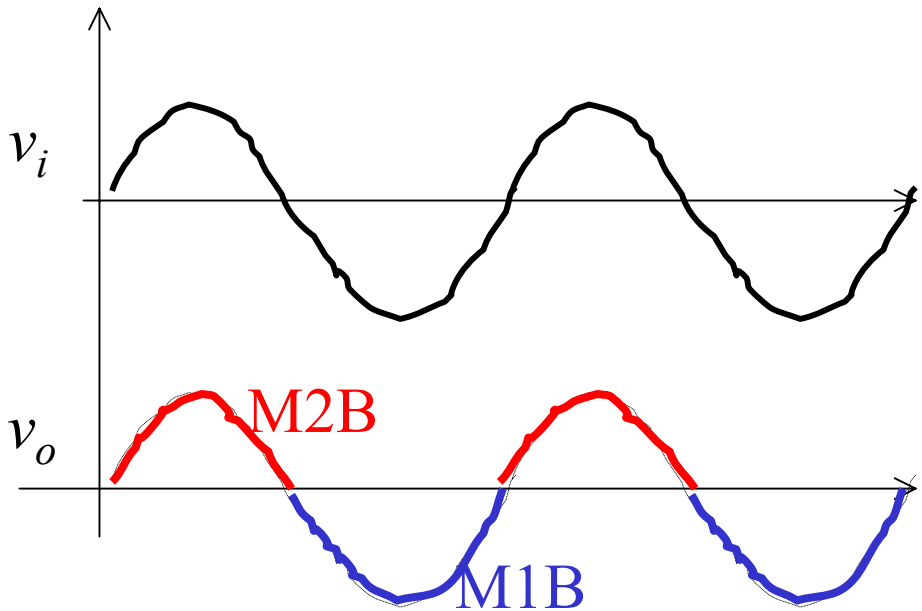
Source follower can serve the purpose

Class AB NMOS and PMOS source follower (push-pull) stage is a preferred configuration

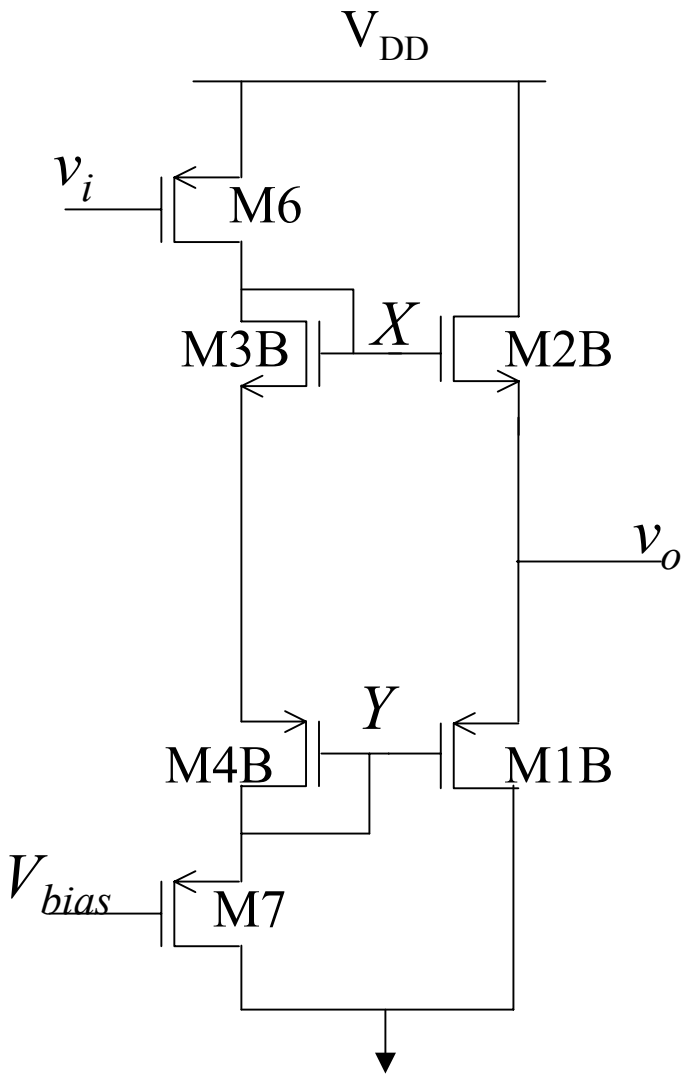
Output stage class AB



V_{GG1} and V_{GG2} are set such that M1B and M2B are biased just above *Vt* to avoid cross over distortion



Biasing the output stage



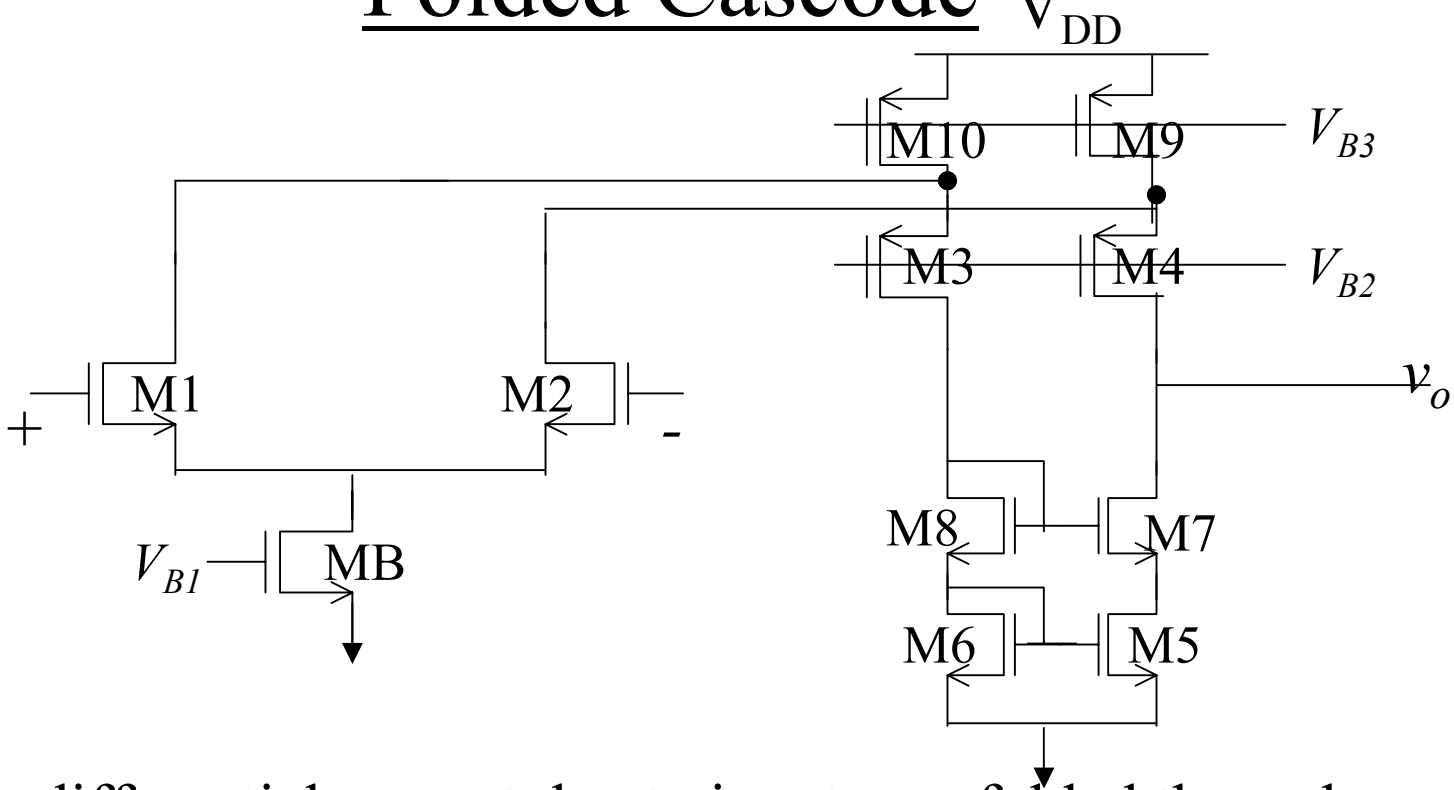
M3B and M4B are diode connected NMOS and PMOS respectivel

$$V_{xy} = V_{gs3b} + V_{gs4b}$$

$$V_{xy} = V_{tn3b} + V_{tp4b} + 2\Delta V$$

Choose the sizes of M3B and M4B such that *Vxy* is just above the the two $V_t s$ of M1B and M2B to avoid cross over distortion

Folded Cascode



The differential current due to inputs are folded through M1-M3 and M2-M4 pairs

The gain will be comparable to 2 stage OPAMP due to cascoding of the load transistor

The load capacitance itself acts as compensation capacitance

Folded cascode gain

$$A_v(0) = \frac{g_m}{\frac{g_{o2} + g_{o9}}{g_{m4} r_{o4}} + \frac{g_{o5}}{g_{m7} r_{o7}}}$$

$$A_v(0) = \frac{\left(g_m r_o\right)^2}{3}$$ Assuming all *gm* and *ro* are identical

The dominant pole is associated with the output

C_L provides frequency compensation

Increasing C_L improves phase margin

OTA and OPAMP Circuits

Operational Transconductance Amplifier

OTA is essentially an OPAMP without an output buffer

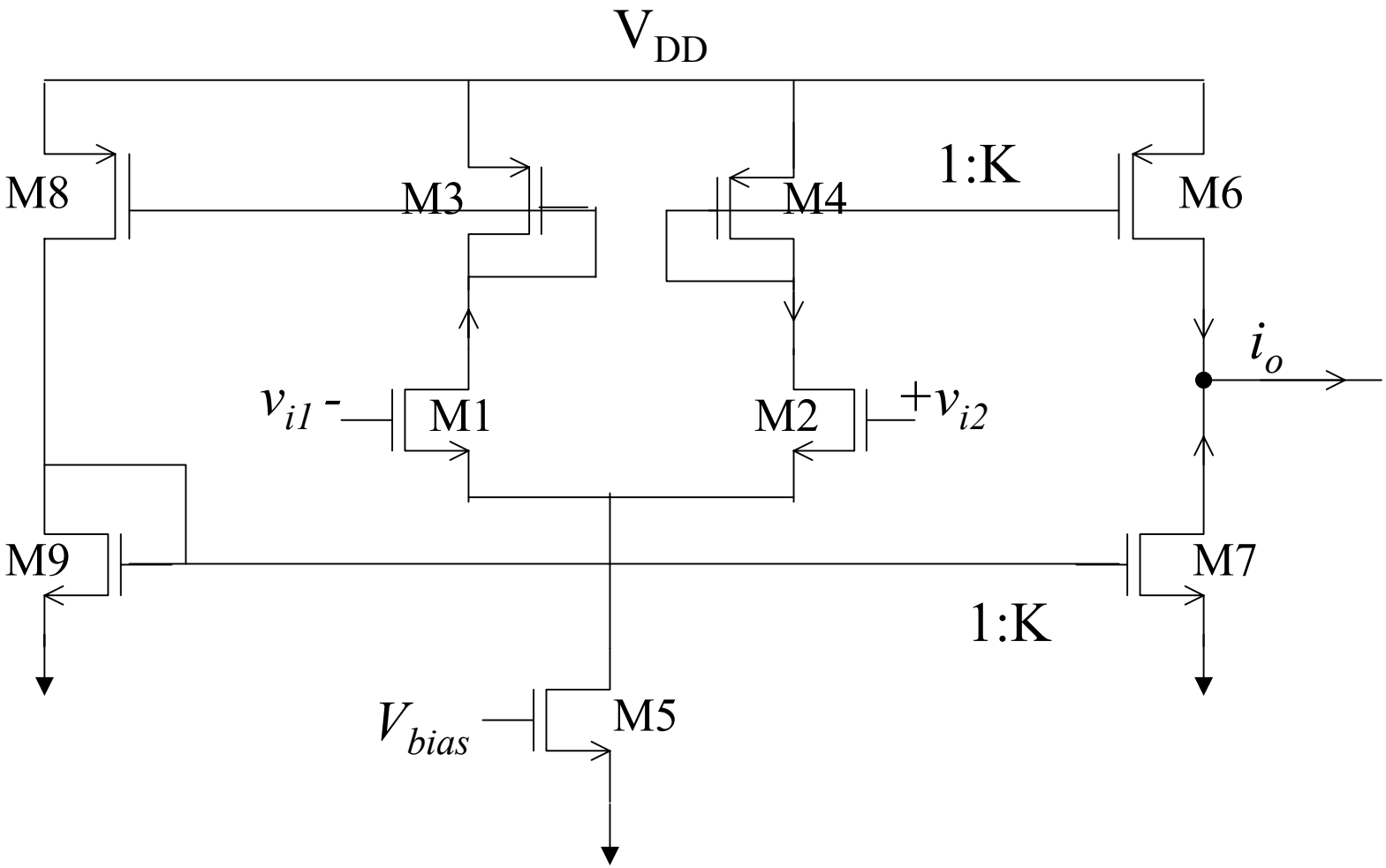
An OTA without output buffer can drive only capacitive loads

OTA is an amplifier where all nodes except I/O are low impedance nodes. Hence the two stage OPAMP configuration minus buffer is NOT an OTA since the drain of M4 is high impedance node

As the name suggests, the quantity of interest in OTA is not the voltage gain, but it is

$$Gm = \frac{i_{out}}{v_{i2} - v_{i1}} = \frac{i_{out}}{v_i}$$

The basic OTA circuit configuration



Gm expression

Assumptions

$g_{m1}=g_{m2}$ and $(W/L)_3=(W/L)_4=(W/L)_8$

$(W/L)_6=K(W/L)_4$ and $(W/L)_7=K(W/L)_9$

Then

$$i_o = i_{d6} - i_{d7} = K(i_{d4} - i_{d9}) = K(i_{d2} - i_{d1})$$

$$i_o = K\left(g_m \frac{v_i}{2} + g_m \frac{v_i}{2} \right) = K g_m v_i$$

$$G_m = \frac{i_o}{v_i} = K g_m$$

Transconductance Gm

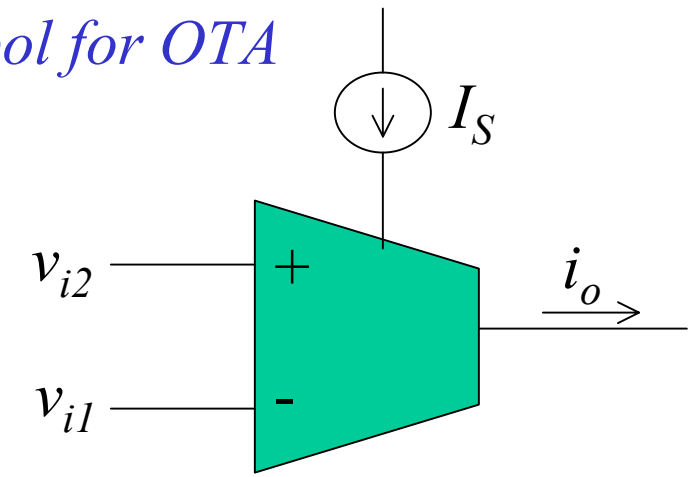
Gm can be set by appropriate K

For a given K (i.e. after design) Gm can still be varied by setting an appropriate bias current, I_S

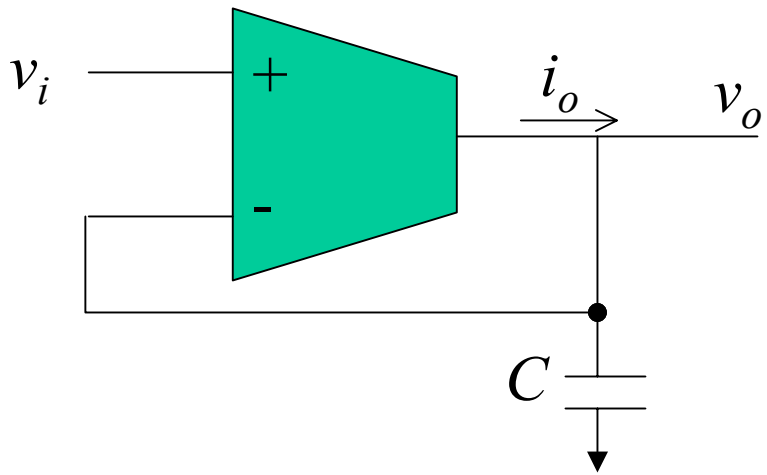
i.e. Filters made using OTA can be tuned by changing I_S

Output pole is the only dominant pole!
i.e. capacitive loads improve the phase margin

The symbol for OTA



Simple Low pass filter



$$v_o = i_o \frac{1}{j\omega C}$$

$$v_o = G_m(v_i - v_o)\frac{1}{j\omega C}$$

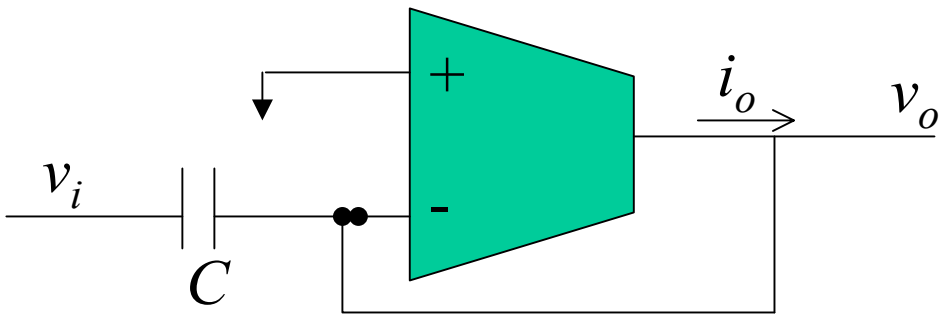
$$\frac{v_o}{v_i} = \frac{G_m / j\omega C}{1 + G_m / j\omega C}$$

$$\frac{v_o}{v_i} = \frac{1}{1 + j\omega\left(\frac{C}{G_m}\right)}$$

Single pole low pass filter with a cut off frequency of

$$\omega_p = G_m/C$$

Simple High pass filter



$$v_o - v_i = \frac{i_o}{j\omega C} = \frac{-G_m v_o}{j\omega C}$$

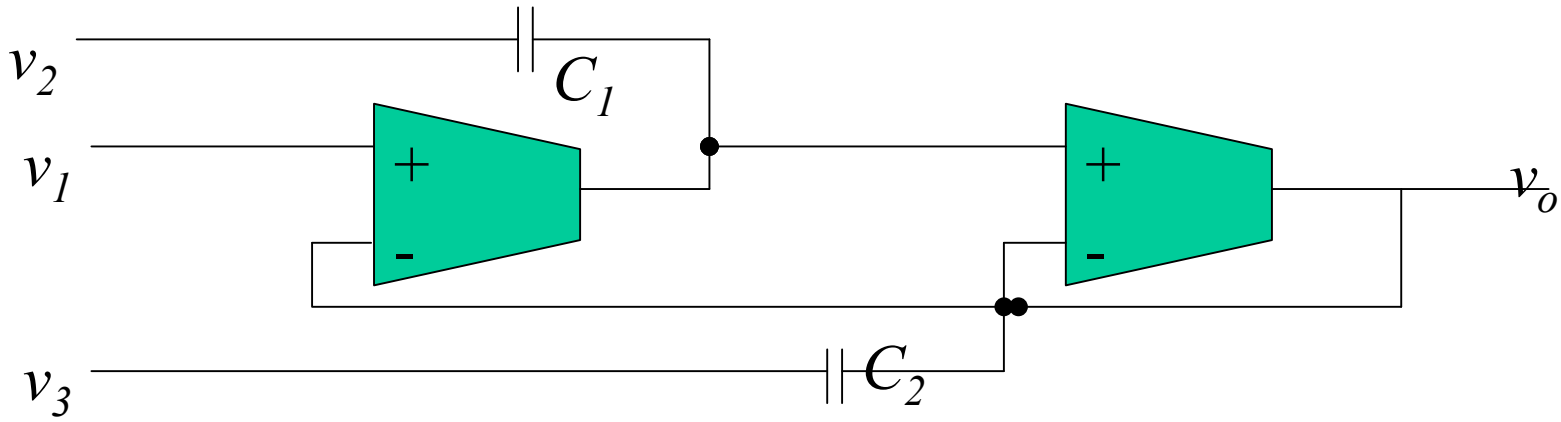
$$\frac{v_o}{v_i} = \frac{1}{1 + G_m / j\omega C}$$

$$\frac{v_o}{v_i} = \frac{j\omega\left(C/G_m\right)}{1 + j\omega\left(C/G_m\right)}$$

High pass filter with cut off frequency of

$$\omega_p = G_m/C$$

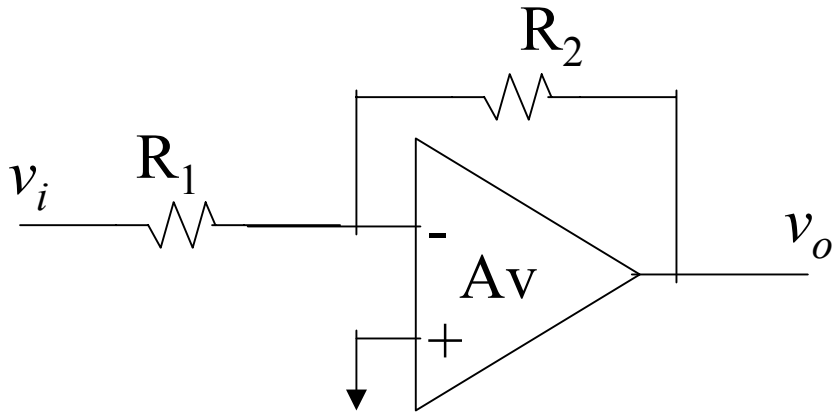
General biquadratic (biquad) configuration



Filter	Input Condition	Transfer function
Low-pass	$v_1 = v_i,\ v_2 = 0,\ v_3 = 0$	$g_m^2 \Big/ s^2 C_1 C_2 + sC_1 g_m + g_m^2$
High-pass	$v_1 = 0,\ v_2 = 0,\ v_3 = v_i$	$s^2 C_1 C_2 \Big/ s^2 C_1 C_2 + sC_1 g_m + g_m^2$
Band-pass	$v_1 = 0,\ v_2 = v_i,\ v_3 = 0$	$sC_1 g_m \Big/ s^2 C_1 C_2 + sC_1 g_m + g_m^2$
Band-reject	$v_1 = v_i,\ v_2 = 0,\ v_3 = v_i$	$s^2 C_1 C_2 + g_m^2 \Big/ s^2 C_1 C_2 + sC_1 g_m + g_m^2$

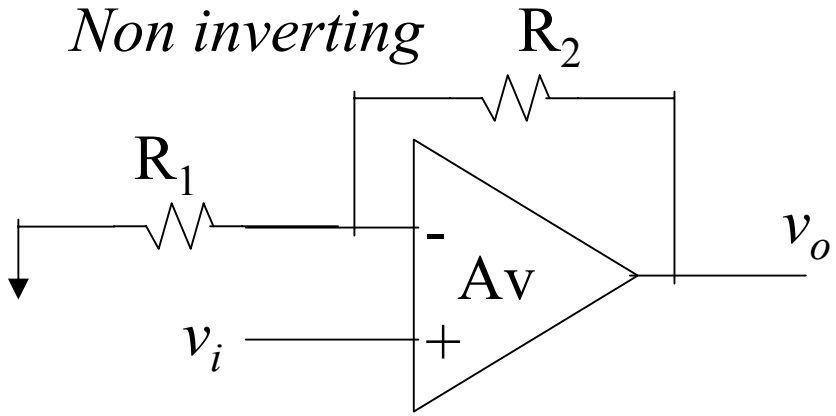
Inverting and Noninverting amplifier

Inverting



$$\frac{v_o}{v_i} = -\frac{R_2}{R_1}$$

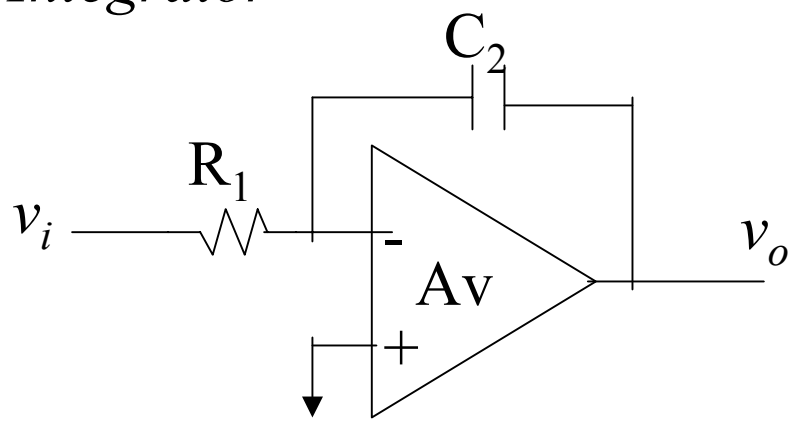
Non inverting



$$\frac{v_o}{v_i} = 1 + \frac{R_2}{R_1}$$

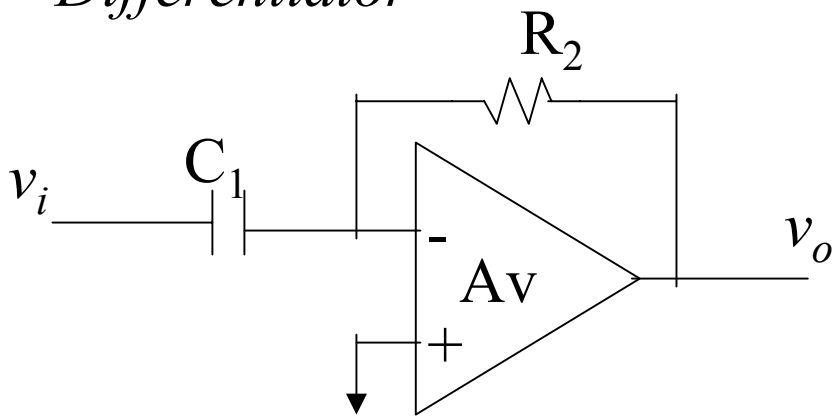
Integrator and Differentiator

Integrator



$$v_o = -\frac{1}{R_1 C_2}\int v_i dt$$

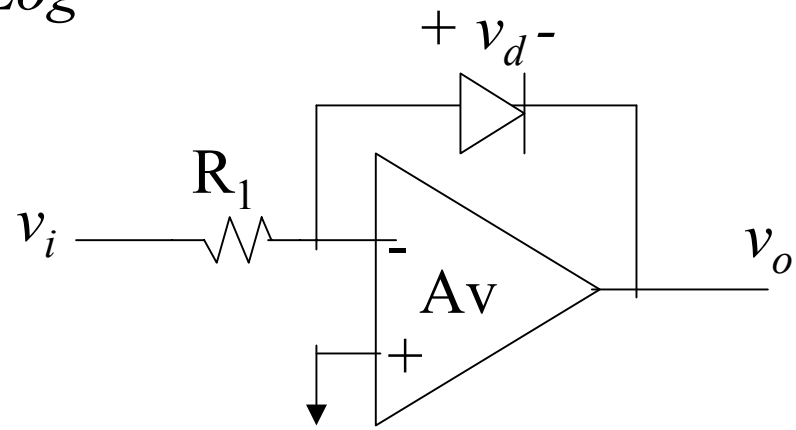
Differentiator



$$v_o = -R_2 C_1 \frac{dv_i}{dt}$$

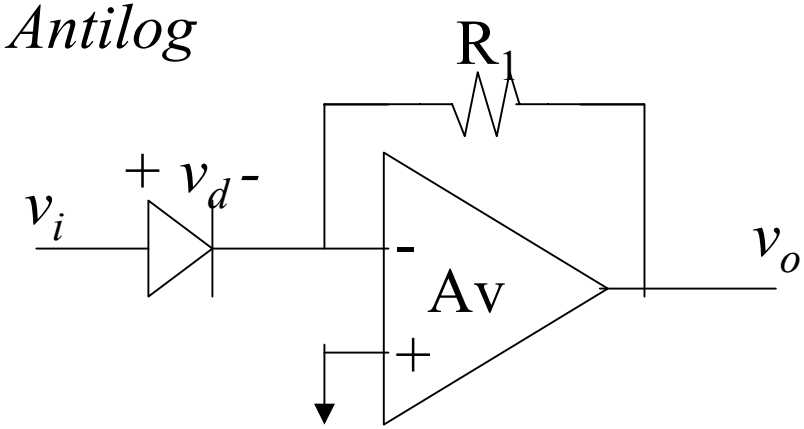
Log and Antilog Amplifier

Log



$$v_o = -V_T \ln \frac{v_i}{R_1 I_o}$$

Antilog



$$v_o = -I_o R_1 e^{\frac{v_i}{v_T}}$$

Sample and Hold Circuit

Sample and Hold Circuit

This is an essential requirement for discrete time systems (sampled data systems)

Applications:

ADCs,

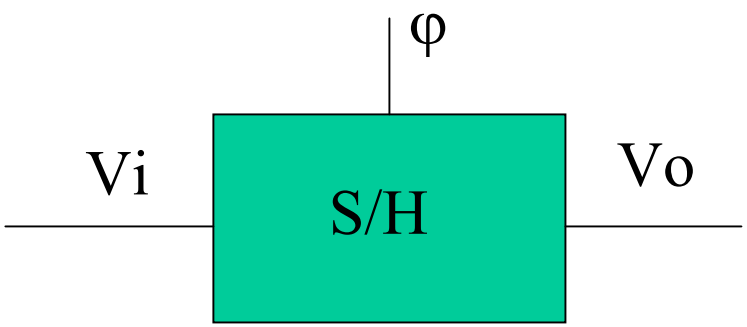
Switched capacitor filters

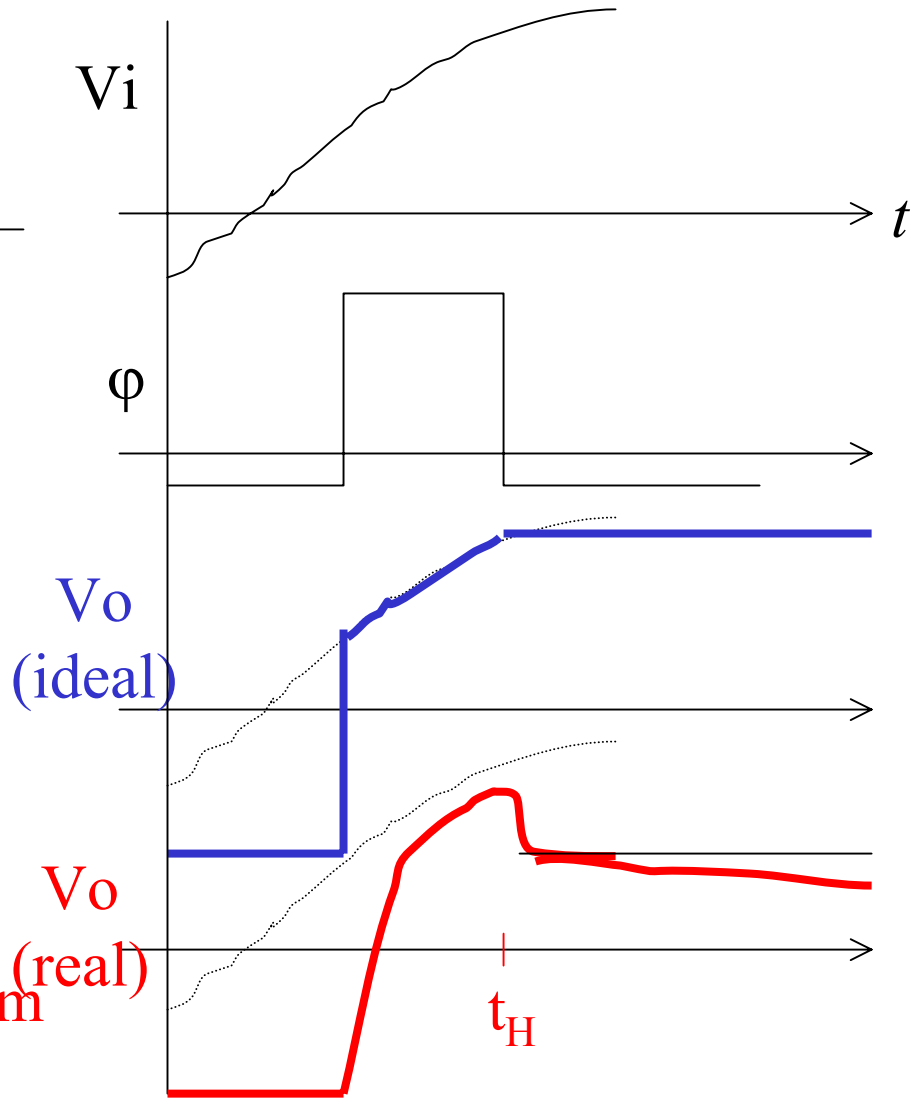
Comparators etc.

Requirement of discrete time operation:

1. Switches to perform sampling
2. High input impedance to sense the charge without corrupting (ideally suited for CMOS and not for BJT)

Requirements for S/H circuit





Advantages of MOSFET over BJT as a switch

1. ON but zero current
2. S/D voltages are not pinned to gate voltage
3. Conducts well in both the directions

But it is still does NOT perform Ideal Sampling function!

MOSFET switch issues

Finite acquisition time

Finite bandwidth in sample mode

DC offset in sample mode (Vos1)

Finite aperture delay (Δt)

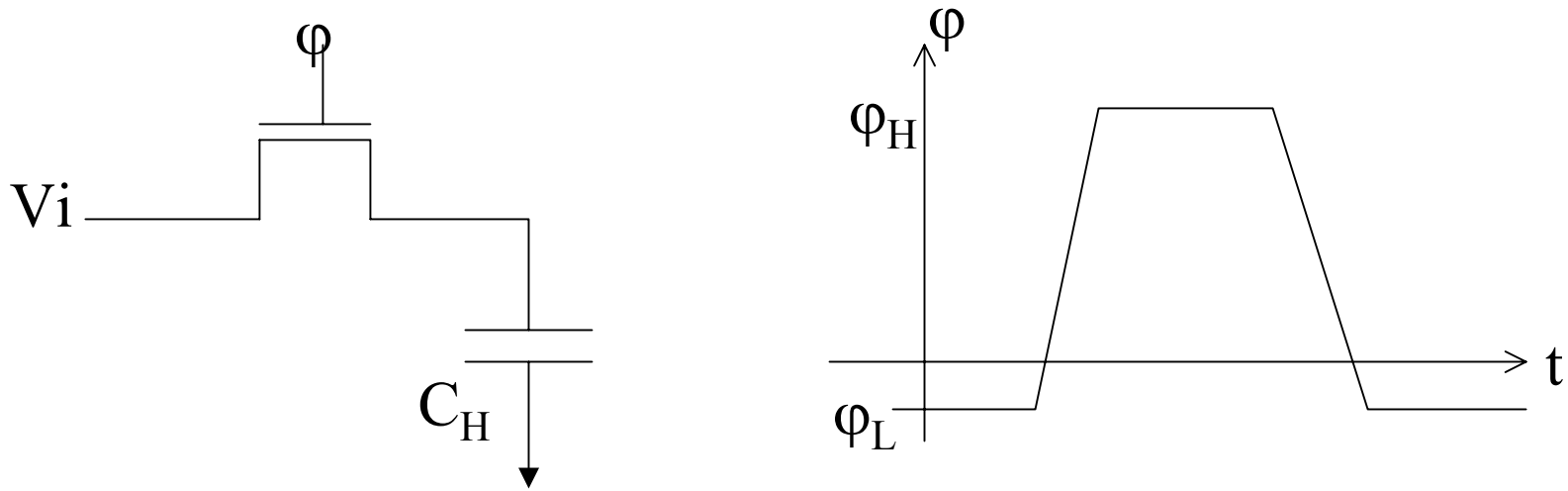
Pedestal error (Vos2) : (Charge injection and Clock feed through)

Droop in Hold mode

$$For\ t > t_H + \Delta T$$

$$V_o(t) = V_i(t_H + \Delta t) + V_{os1} + V_{os2} + \Delta V(t)$$

MOSFET sampling



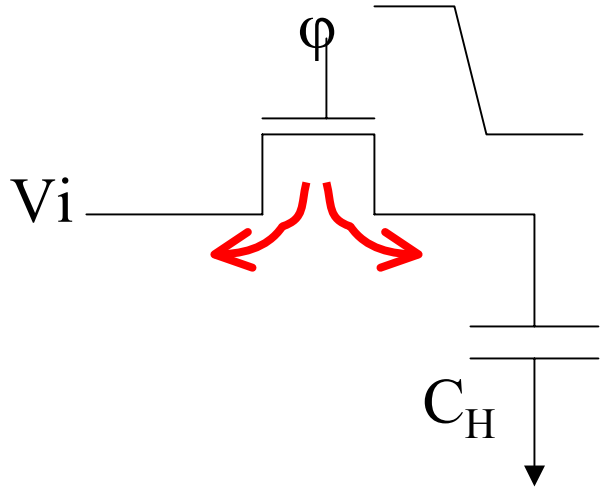
Acquisition time $\tau = R_{on}C_H$ (RC time constant of channel)

$$R_{on} = \frac{1}{\mu Cox \frac{W}{L}(V_{gs} - V_t)}$$ In linear region

Bandwidth in sample mode = 1/τ

$V_{os1}=0$ provided MOSFET is in linear region (i.e. $Vin < \varphi_H - Vt$)
Otherwise $Vo \neq Vin$ instead $Vo = \varphi_H - Vt$

Channel charge injection



When switch is ON, channel charge is

$$Qc=WLCox(\varphi_H-Vi-Vt)$$

When φ goes low, the switch turns off and the channel charge must exit out

An approximation is, 50% of this charge Goes to the out put node

The fraction that goes to output node is a complex function of parameters such as impedance seen at each node to the ground, clock transition time etc.
(ex: if clock makes slow transition all the charge could be absorbed at input)

$$\Delta V = -\frac{Q_c}{2C_H}$$

$$\Delta V = -\frac{WLC_{ox}(\phi_H - V_i - V_t)}{2C_H}$$

Effect of charge injection

$$V_o = V_i - \frac{WLC_{ox}(\phi_H - V_i - V_t)}{2C_H}$$

Vt is impacted by body effect

$$Vo = Vi\left(1 + \frac{WLC_{ox}}{2C_H}\right) + \gamma\frac{WLC_{ox}}{2C_H}\sqrt{2\varphi_b + V_i} - \frac{WLC_{ox}}{2C_H}\left(\phi_H - V_{t0} + \gamma\sqrt{2\phi_b}\right)$$

Gain error *Nonlinearity* *DC offset*

Speed-Precision product : $\tau\Delta V$

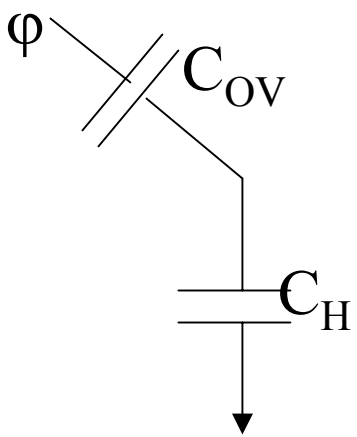
$$\tau \cdot \Delta V = \frac{L^2}{2\mu}$$

Interesting trade-off!

Depends only on L and is independent of transistor width and the value of the sampling capacitance

Clock feedthrough

When the switch is being turned off, the clock transition capacitively couples to the output



$$\Delta V = (\phi_H - \phi_L)\frac{C_{ov}}{C_{ov} + C_H}$$

$$C_{ov} = nC_{ox}WL_d$$

Note: If clock makes slow transition (quasi static)
Then the clock feed through error is significantly less

Example for error values

W=10μm, L=2μm, Vt=0.7V, Cox=1.38fF/μm^2,
C_{ov}=3fF, C_H=1pF, φ_H=5V, φ_H=0V

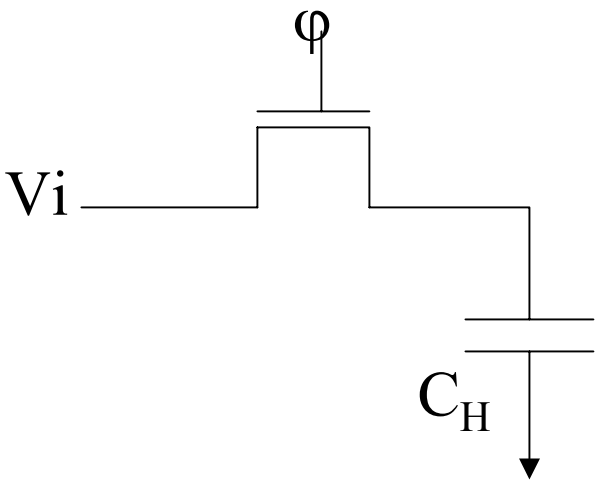
Gain error = 1.1%

Charge injection offset = 47mV

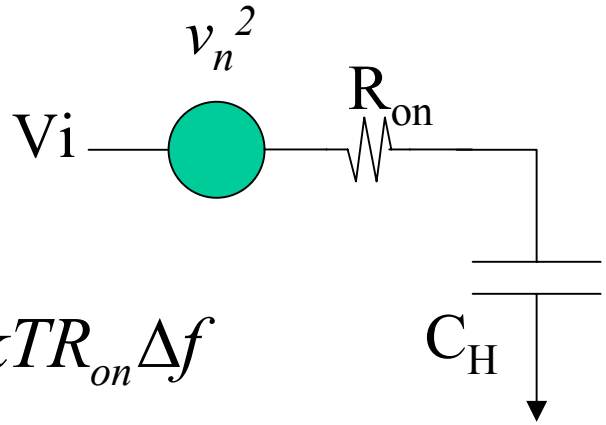
Clock feed through offset = 15mV

Total offset = 62mV

kT/C noise



Equivalent ckt in sample mode



$$\overline{v_n^{\,2}} = 4kTR_{on}\Delta f$$

$$H(f) = \frac{1}{1 + j2\pi f R_{on} C_H}$$

$$\overline{v_{oT}^{\,2}} = 4kTR_{on}\int_0^{\infty} \frac{1}{\left|1 + j2\pi f R_{on} C_H\right|^2} df$$

$$\overline{v_{oT}^{\,2}} = \frac{kT}{C_H}$$

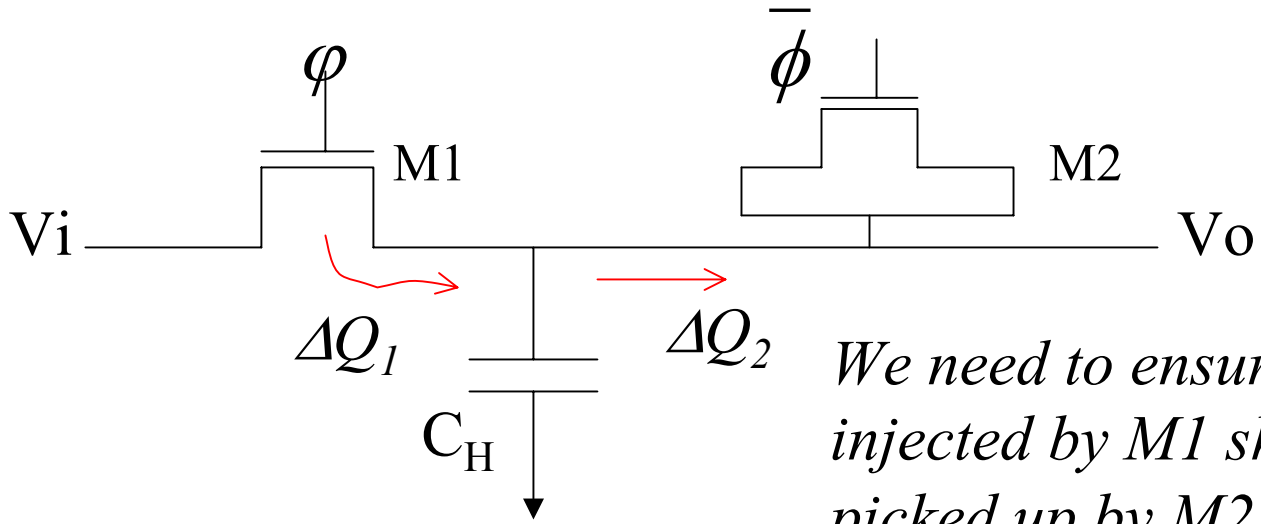
$$v_{oT}(rms) = \sqrt{\frac{kT}{C_H}}$$

For C_H=1pF, T=300oK,

$V_{oT}(rms) = 64.3\,\mu V$

The lowest limit!

Offset cancellation with dummy switch



We need to ensure that ΔQ_1 injected by M1 should be picked up by M2 (ΔQ_2)

$\Delta Q_1 = 0.5 W_1 L_1 C_{ox}(V_{DD}-Vi-Vt)$ $\Delta Q_2 = W_2 L_2 Cox(V_{DD}-Vi-Vt)$

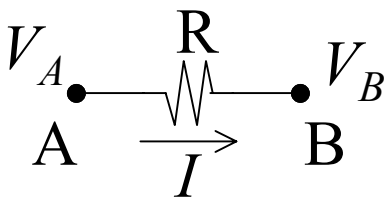
Choose $L_1=L_2$ and $W_2= W_1/2$ to cancel the charge injection

Note that the clock feed through error is also cancelled

$$\Delta V_{cft} = -\Delta\phi \frac{C_{ov}}{C_{ov}+C_H} + \Delta\phi \frac{C_{ov}}{C_{ov}+C_H} = 0$$

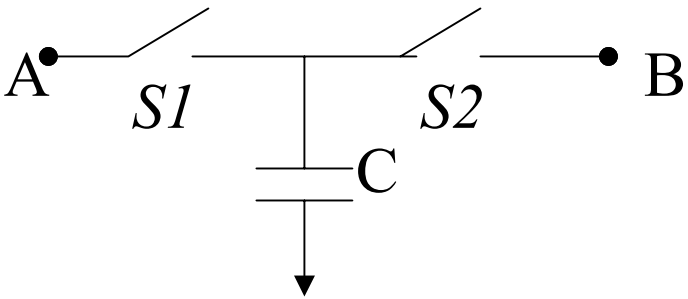
Switched Capacitor Circuits

Switched capacitor as a resistor

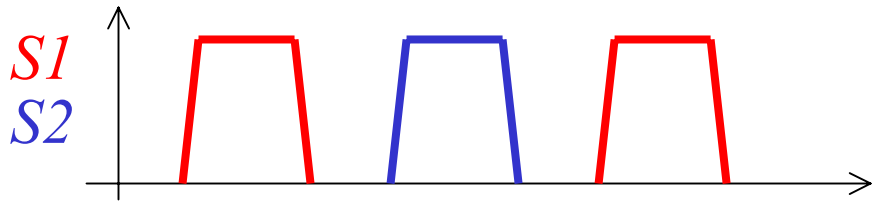


$$I = \frac{V_A - V_B}{R}$$

The Resistor takes certain amount of charge per second from node A to B resulting in a current I



S1 and S2 are non overlapping



In one cycle a charge $q = VaC - VbC$ is transferred from A to B

i.e. charge transferred per second from A to B is

$$I_{avg} = f_{clk} C (V_A - V_B) \qquad I_{avg} = \frac{V_A - V_B}{1/f_{clk} C} \qquad R_{eq} = \frac{1}{f_{clk} C}$$

Switched capacitor acts like an equivalent resistance!
Provided f_{clk} is higher than signal bandwidth

Motivation for switched capacitor circuits

It is very easy to build a capacitor compared to a resistor in the CMOS Technology

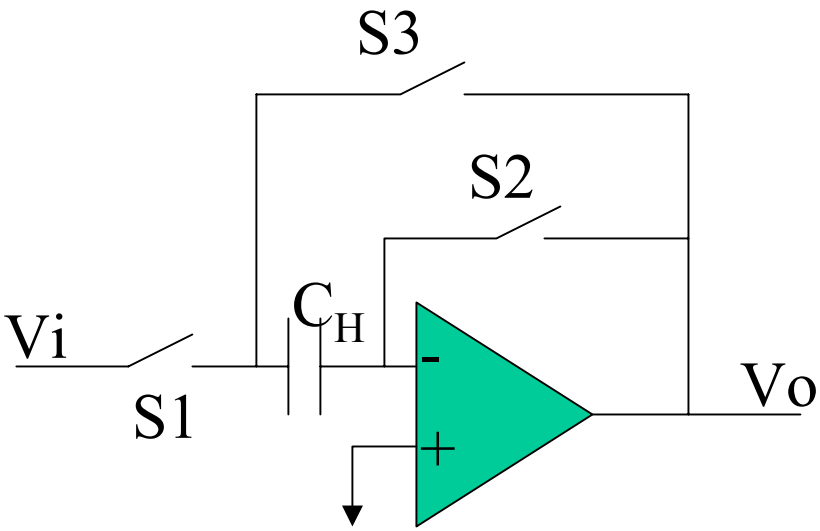
The value of the capacitance does not depend on temperature

The capacitor connected to the output of an OPAMP does not impact the resistance and hence the open loop gain

Monolithic active RC filters can be built using switched capacitor circuits

In most of the switched capacitor circuits, the poles and zeros are governed by capacitance ratios rather than absolute values. The precision on capacitance ratios is significantly better compared to precision on absolute C and R

Unity gain buffer/Sampler



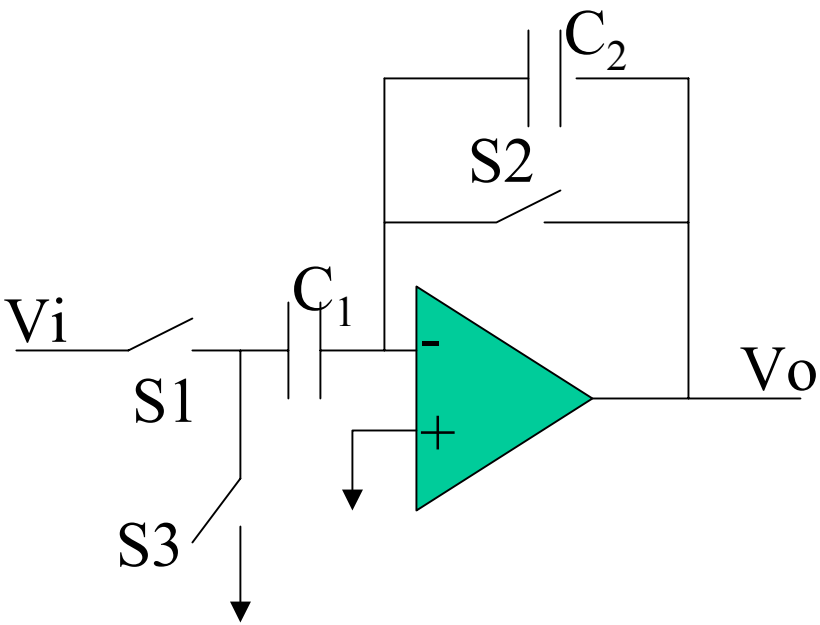
S3 open and S1, S2 closed

One plate of the capacitor is at virtual ground and hence *Vi* is sampled on the other plate

S1, S2 open and S3 is closed

C_H is connected to *Vo* and hence *Vo* is sampled input voltage
Further the circuit enters in holding phase with constant voltage on C_H

Inverting Amplifier



Initially S3 open and S1,S2 closed

Negative feedback is enabled and Hence – input of OPAMP is at Virtual ground. C_H samples Vi

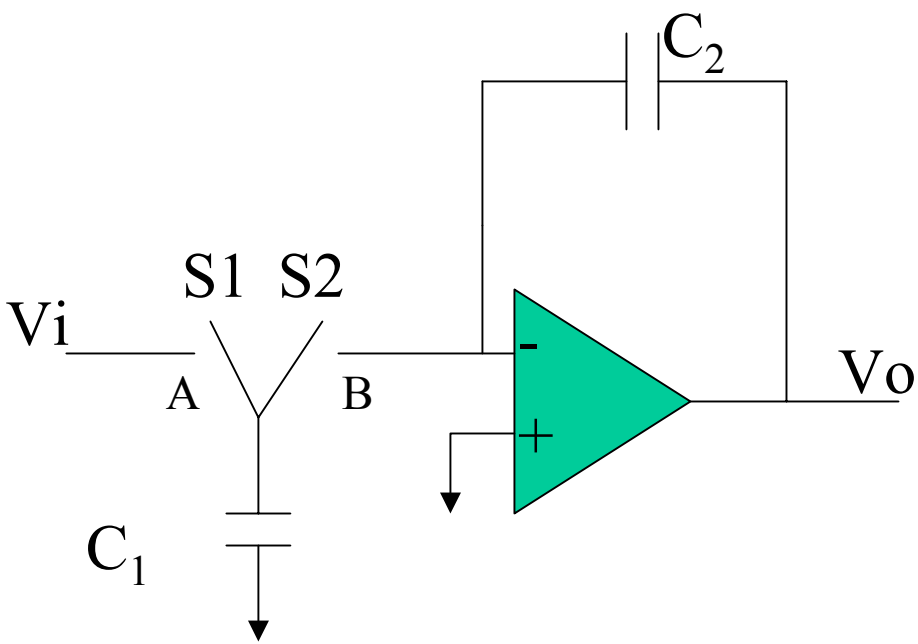
S1 and S2 open and S3 closed

Capacitance discharges and the charge will be transferred to C2

$$V_o = -V_{in}\frac{C_1}{C_2}$$

From conservation of charge

Integrator



S1, S2 are non overlapping and are switched at a frequency of *fc*

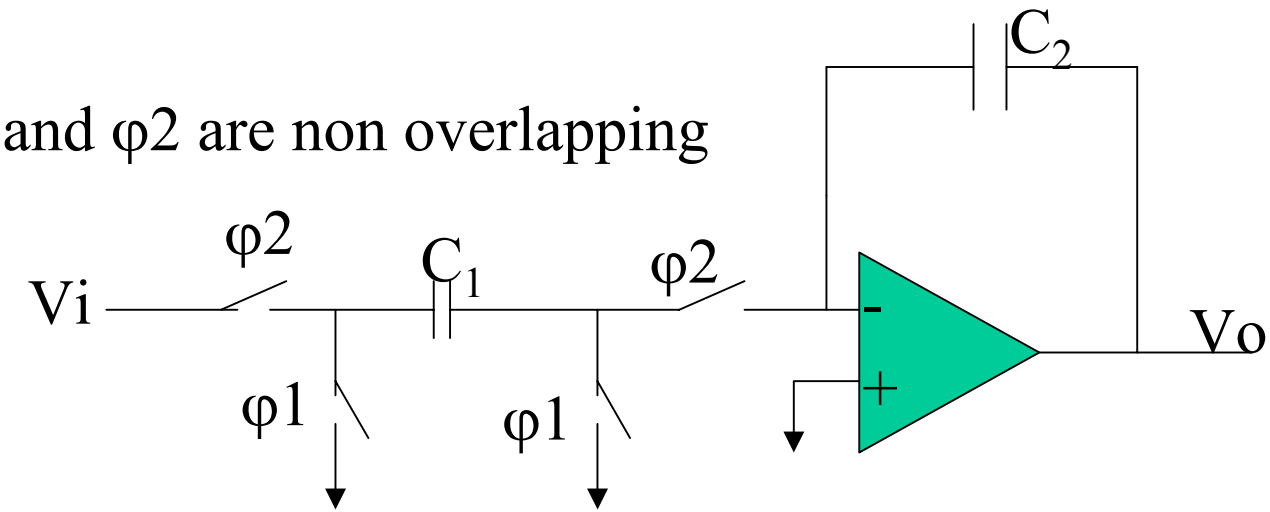
Hence there is an effective Resistance between nodes A and B

$$R_{eq} = \frac{1}{C_1 f_c}$$

$$v_o = -\frac{C_1 f_c}{C_2} \int v_i dt$$

Stray insensitive inverting integrator

φ1 and φ2 are non overlapping



φ1 is closed and φ2 is open , C1 is discharged to 0V

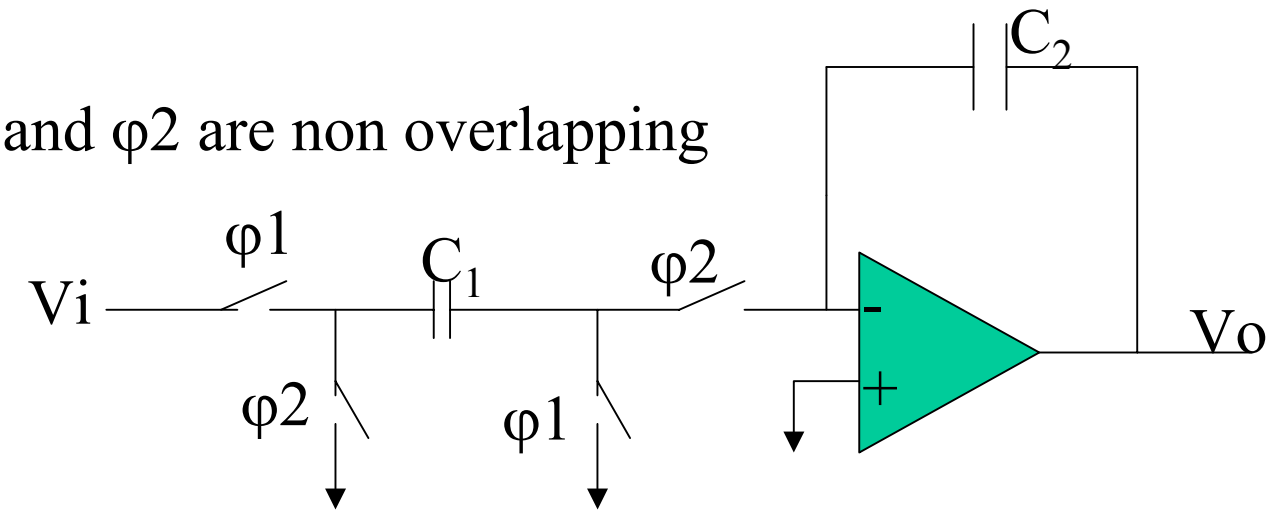
φ1 is open and φ2 is closed , a charging current flows through C1 and C2

At the n^{th} sampling instance

$$v_o(n) = V_o(n-1) - \frac{C_1}{C_2} v_i(n)$$

Stray insensitive noninverting integrator

φ1 and φ2 are non overlapping



φ1 is closed and φ2 is open , C1 is charged to Vi

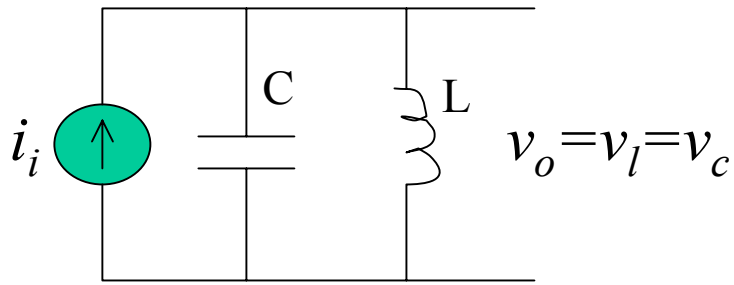
φ1 is open and φ2 is closed , a charging current flows through C2 from Vo and the charge is transferred to C2

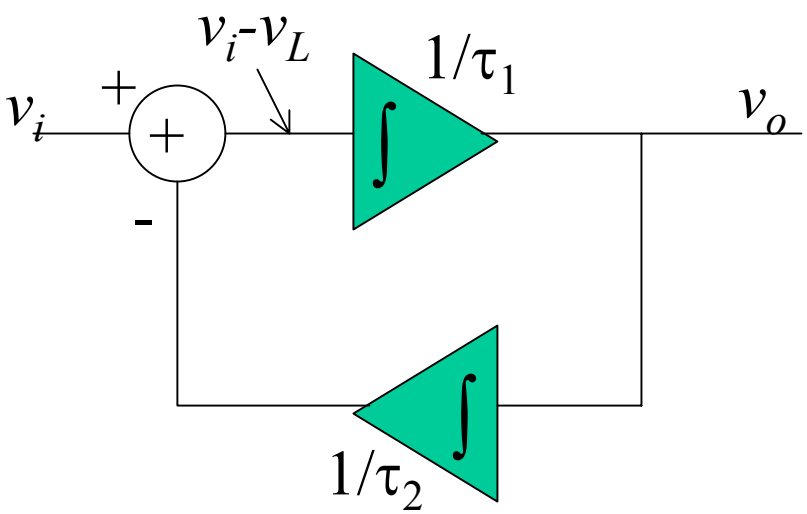
At the n^{th} sampling instance

$$v_o(n) = V_o(n-1) + \frac{C_1}{C_2} v_i(n)$$

Integrator as versatile building block

Lossless resonator





$$i_L = \frac{1}{L}\int v_o dt$$

$$v_o = \frac{1}{C}\int (i_i - i_L)dt$$

Change variables

$v_i=i_iR^*$, $v_L=i_LR^*$

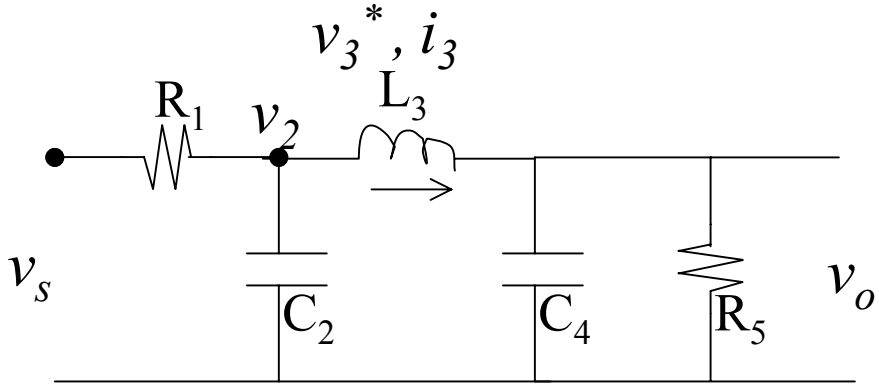
$$v_L = \frac{R^*}{L}\int v_o dt$$

$$v_o = \frac{1}{R^*C}\int (v_i - v_L)dt$$

$$\tau_1 = R^*C$$

$$\tau_2 = \frac{L}{R^*}$$

Ladder filter : Low pass



Define $v_3^* = i_3$ x R^* where v_3^* is scaled inductor voltage

$$v_2 = \frac{v_s}{j\omega C_2 R_1} - \frac{v_2}{j\omega C_2 R_1} - \frac{v_3^*}{j\omega C_2 R^*}$$

$$v_3^* = \frac{v_2}{j\omega L_3 / R^*} - \frac{v_o}{j\omega L_3 / R^*}$$

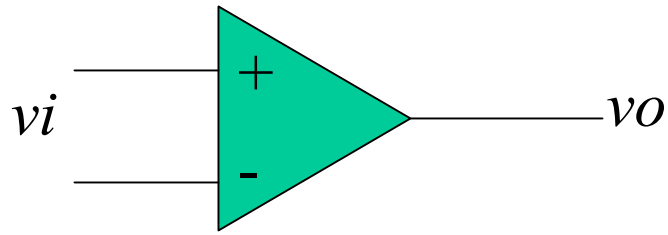
$$v_0 = \frac{v_3^*}{j\omega C_4 R^*} - \frac{v_o}{j\omega C_4 R_5}$$

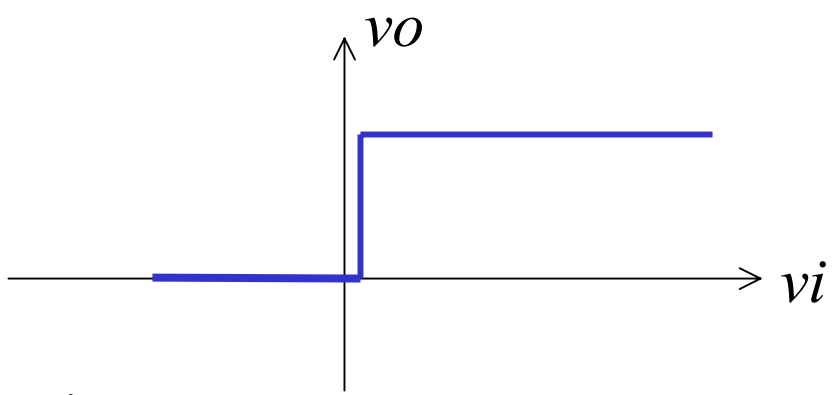
These can be realized using summing integrator

Comparator

Characteristics of comparator

Comparator is a nonlinear circuit which generates rail to rail output for small differential input signal





Configurations:

High gain amplifier without latch

Latched comparator

Selection of Av

The desired resolution and hence v_{imin} sets up the required gain

Suppose it needs to be used for 12 bit Flash ADC application

Let Full scale I/p=4V , 12 bit ADC $\Rightarrow$ 4K levels

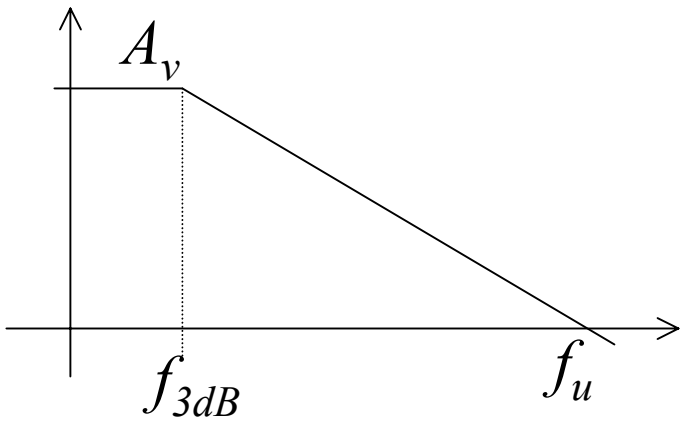
i.e. 1 LSB = 1mV and half LSB=0.5mV

If the output rail voltage is 5V

$$A_V = \frac{5}{0.5 \times 10^{-3}} = 10000$$

High gain realized in a single stage affects speed

Suppose the gain bandwidth product is 10MHz



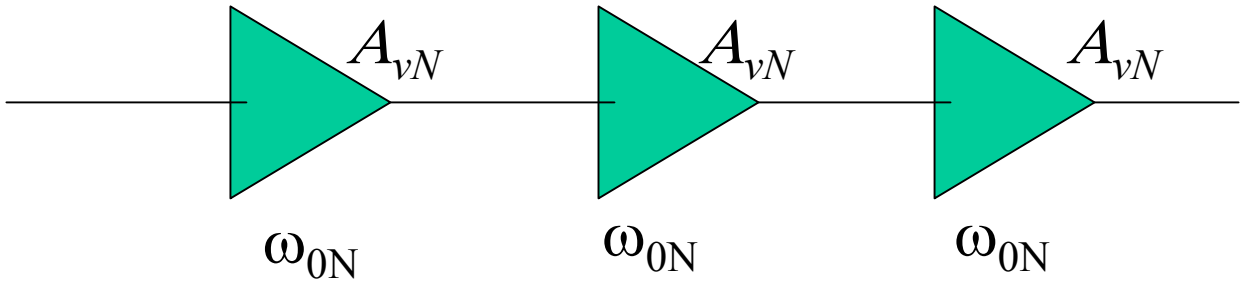
Then $f_{3dB} = f_u/A_v = 1KHz$

$\tau = 1/2\pi f_{3dB}$

$\tau = 160\ \mu sec$

Instead of using a single stage, cascaded stages with lower individual gain but constant overall gain can be used to improve the speed

Time constants of cascaded stage



$$A_T(j\omega) = \frac{A_{vn}(0)^N}{\left(1 + \frac{j\omega}{\omega_{oN}}\right)^N}$$

For an n stage cascade, each stage can have significantly lower gain in conjunction with higher bandwidth

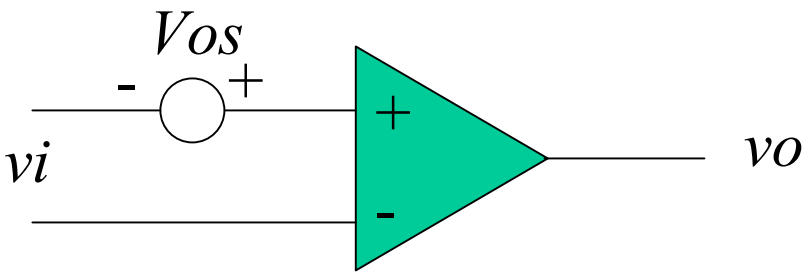
The gains get multiplied, whereas the time constants add up

$$\frac{\omega_{oN}}{\omega_{o1}} = A_T(0)^{\frac{N-1}{N}} \sqrt{2^{1/N} - 1}$$

ω_{oN} is 3dB bandwidth

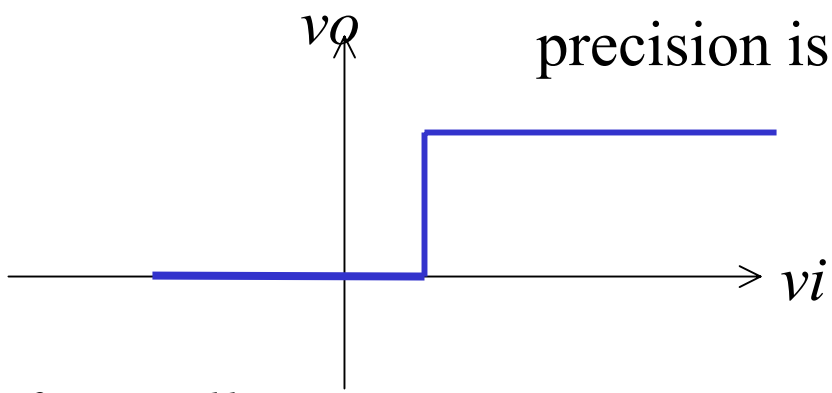
For N=3, $\omega_{0N}/\omega_{01} = 236$ and Av(0)=21.5

Offset cancellation



The offset impacts performance

If offset is not cancelled then the precision is affected

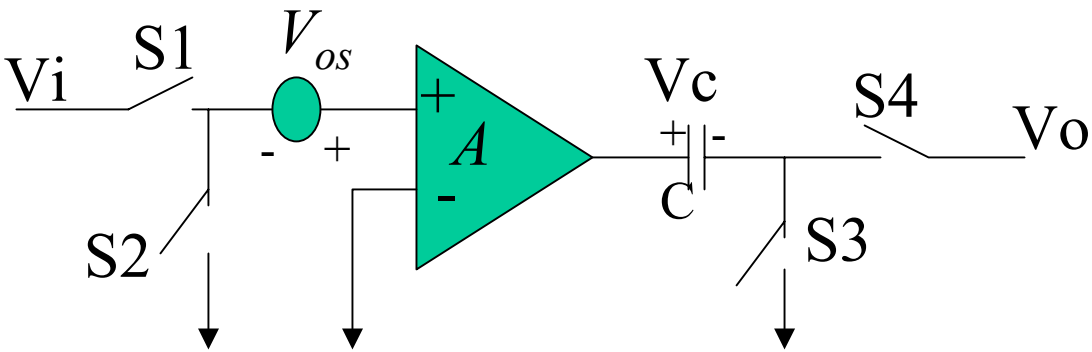


The concept of cancellation:

Sample the offset voltage

Store the offset voltage either at the input or output

Output series cancellation



Offset store: S1, S4 open and S2, S3 closed

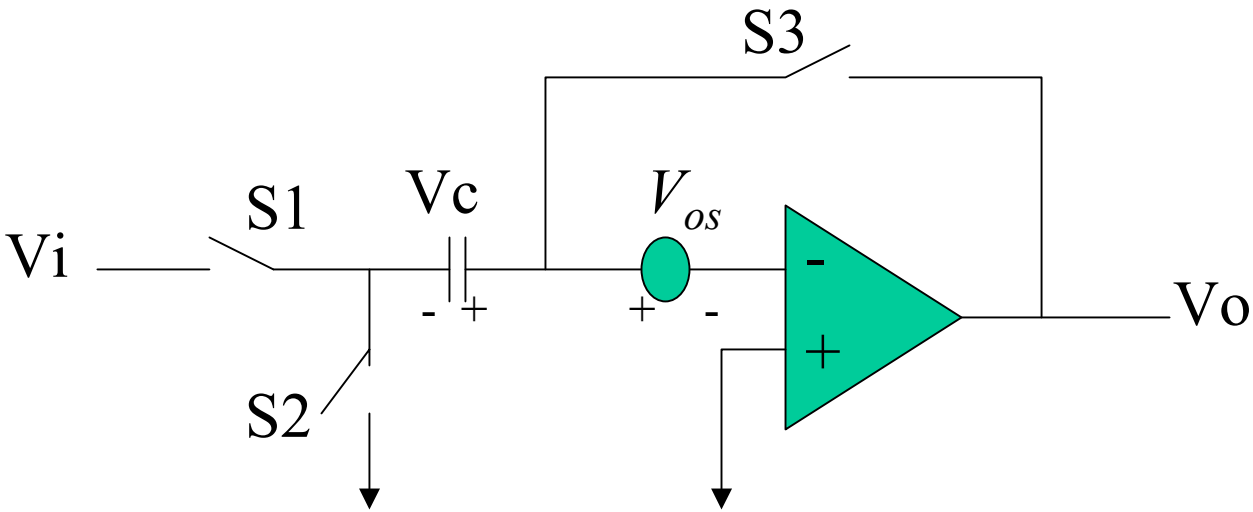
$Vc = AVos$, *The amplified offset voltage is stored on C*

Amplify: S2, S3 open and S1, S4 closed

$Vo = A(Vi + Vos) - Vc = AVi$

Note: If the gain is very large the OPAMP may saturate during the offset store phase

Input series cancellation



Offset store: S1 open and S2, S3 closed

The OPAMP is in negative feedback mode

$Vo=Vc=-A(Vc - Vos)$

$$V_C = \frac{A}{1+A} V_{os} \approx V_{os}$$

Amplify: S2, S3 open and S1 closed

$$V_o = -A\left(V_i - \frac{V_{os}}{1+A}\right)$$

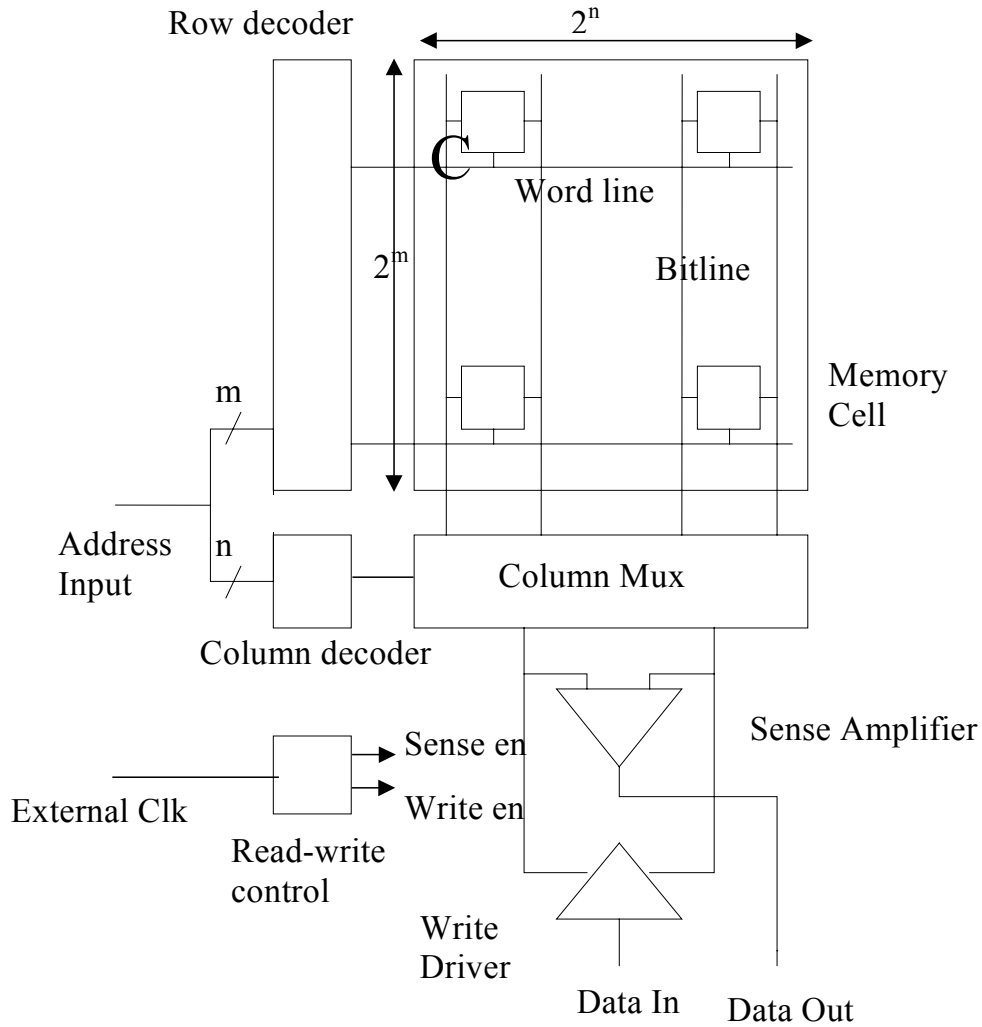
i.e. input referred offset is reduced by a factor 1/(1+A)

Sense Amplifier

Voltage Sense Amplifier
Current sense Amplifier
Latch type Sense Amplifier
Gain bandwidth analysis

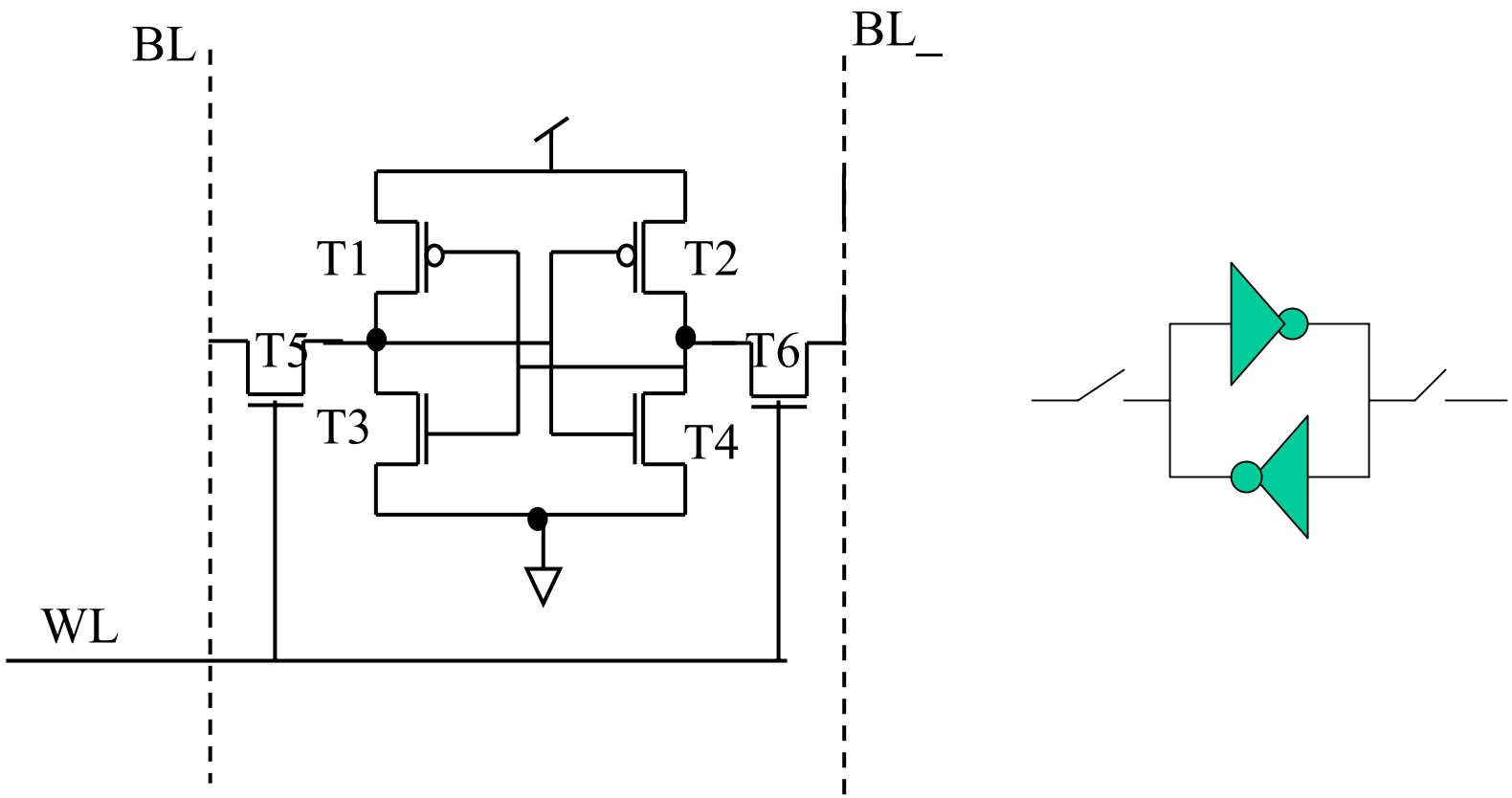
Acknowledge the contributions made by Sugato Mukherjee

SRAM structure



SRAM is 2 dimensional array of memory cells (C)

6T SRAM cell



SRAM cell consists of two cross coupled inverters

SRAM Read operation

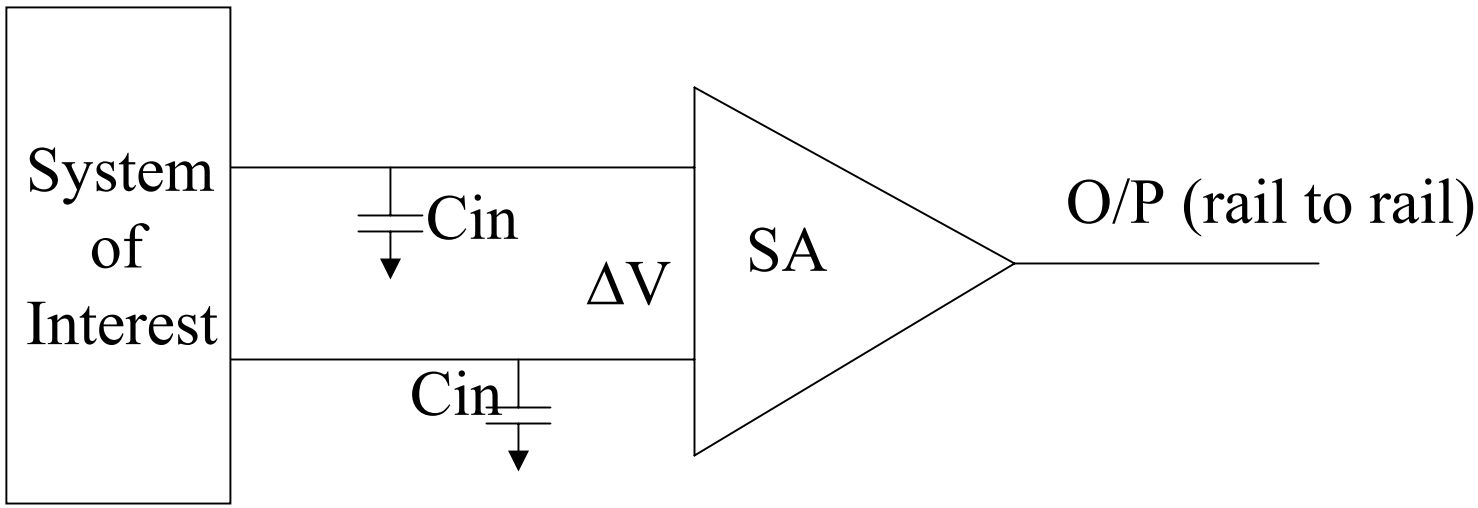
BL and BL_ are pre-charged and equalized to Vdd

WL selects the memory cell

Either BL or BL_ starts discharging through cell

Differential voltage between BL and BL_ is amplified by the sense amplifier

Role of sense amplifier



Sluggish inputs (i.e. large input capacitance)

In the absence of SA,

$$\Delta V = Iin * \Delta T / Cin$$

Sense Amplifier in SRAM

• Sense amplifier is one of the most critical elements in the design of a high speed SRAM

•Sense amplifier is the most important analog block in an otherwise digital memory

• Sense amplifier amplifies small voltage swing on bit-lines to CMOS voltage levels.

• Data sensing delay comprises about 50% of the total access time.

Current Sense and Voltage Sense

Voltage Sense

- Input signal comes to gate of MOS transistors
- Input impedance tends to be very large

Current Sense

- Input signal comes to drain/source of MOS transistors
- Ideally zero input impedance
- Low input differential swing lowers interconnect delay

Current sense vs. Voltage Sense

Current	Voltage
Cell current is sensed	Voltage on bit lines is sensed
Speed is independent of C_{BL}	Speed is a function of C_{BL}
Low input impedance	High input impedance
Low power for small swings	High power for large swings
More sensitive to offset voltage	Less sensitive to offset voltage

SA Implementations

Voltage Sensing

- Current mirror based SA
- PMOS cross-coupled SA

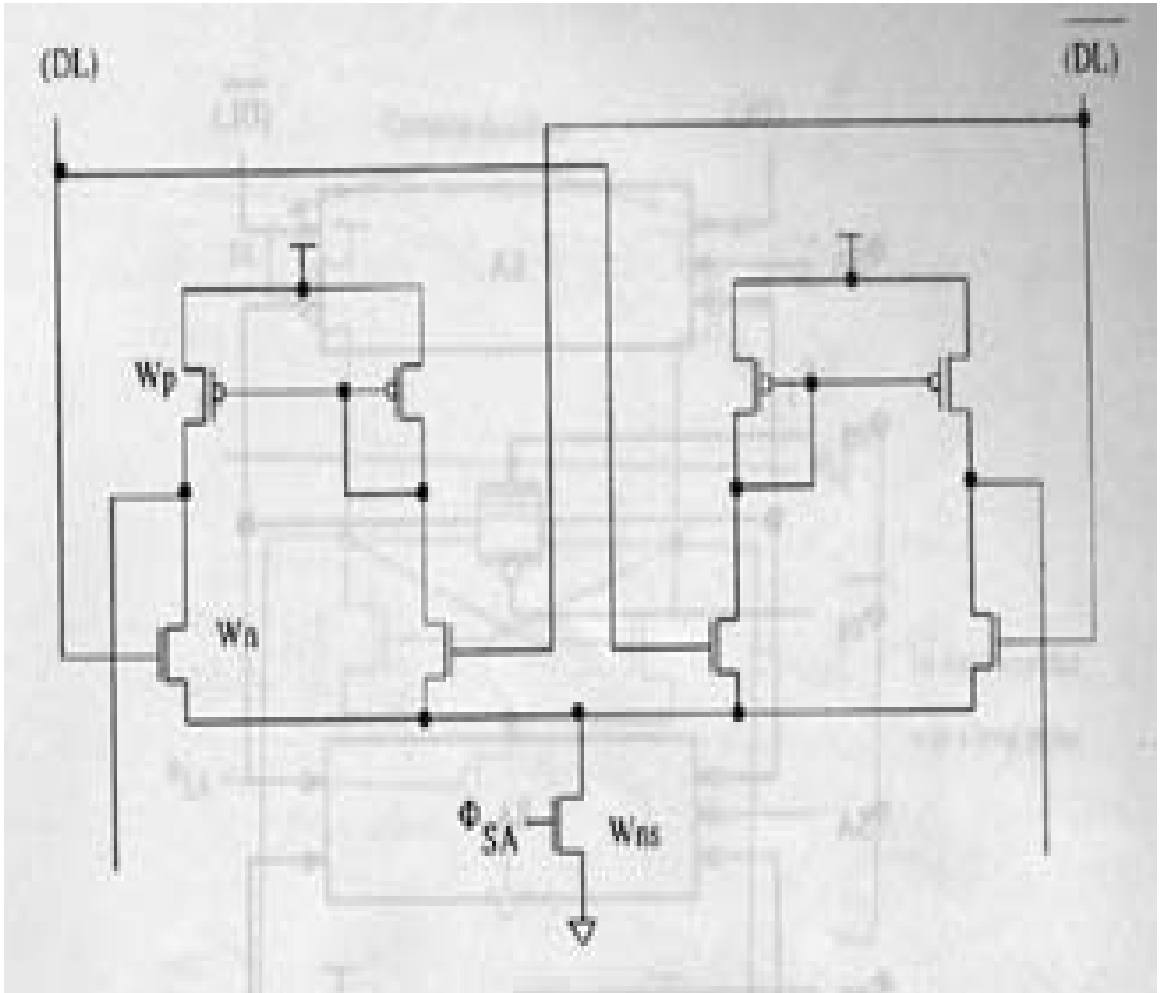
Current Sensing

- Current conveyor based SA

Other Schemes

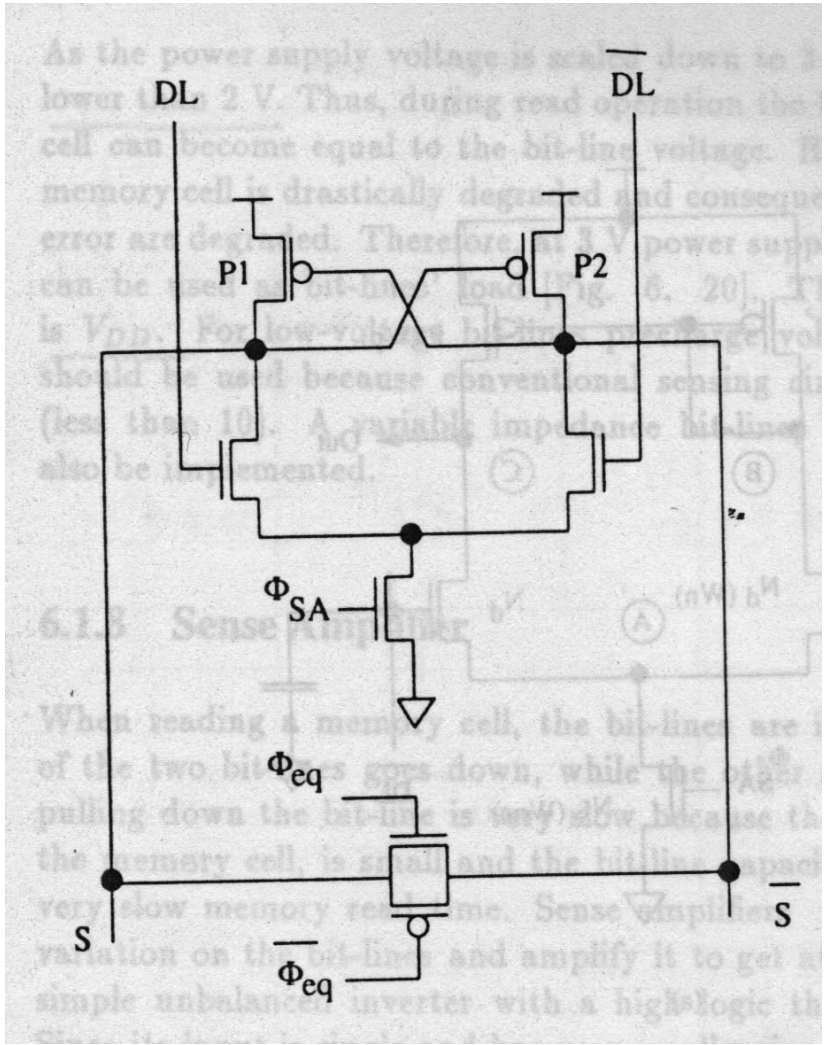
- Half latch based SA
- Input decoupled latch SA

Current-mirror based SA

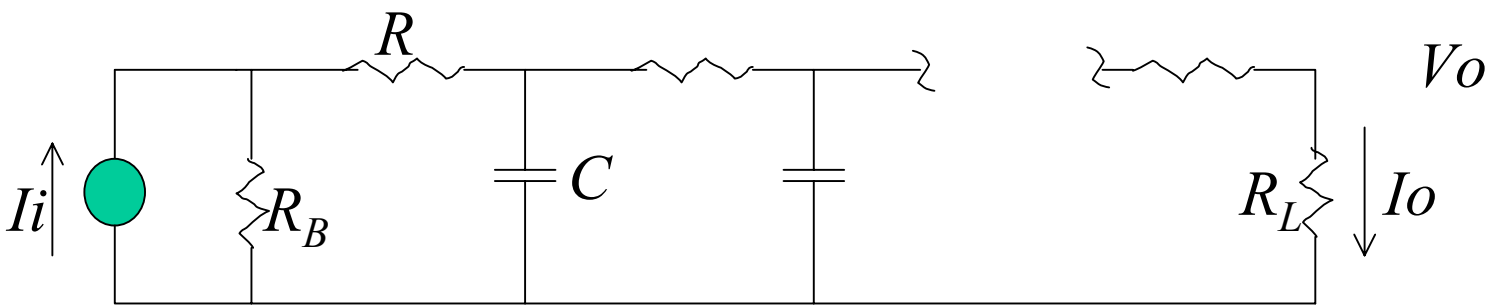


All differential voltages should be equalized before sensing

PMOS cross-coupled SA



Current mode vs Voltage mode



Voltage mode signaling : $R_L = \infty$

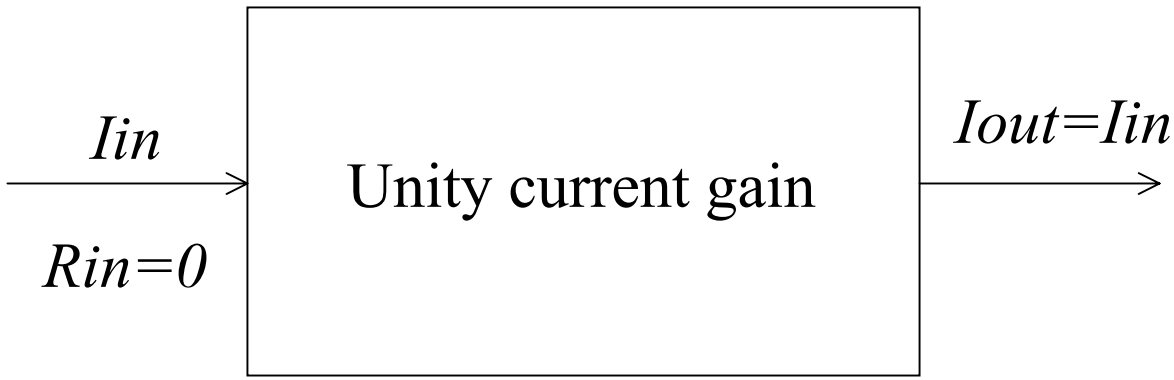
$$\tau = R_B C_T$$

Current mode signaling : $R_L = 0$

$$\tau = \frac{R_T C_T}{2}$$

Ref: IEEE JSSC vol. 26, no. 4, April 1991

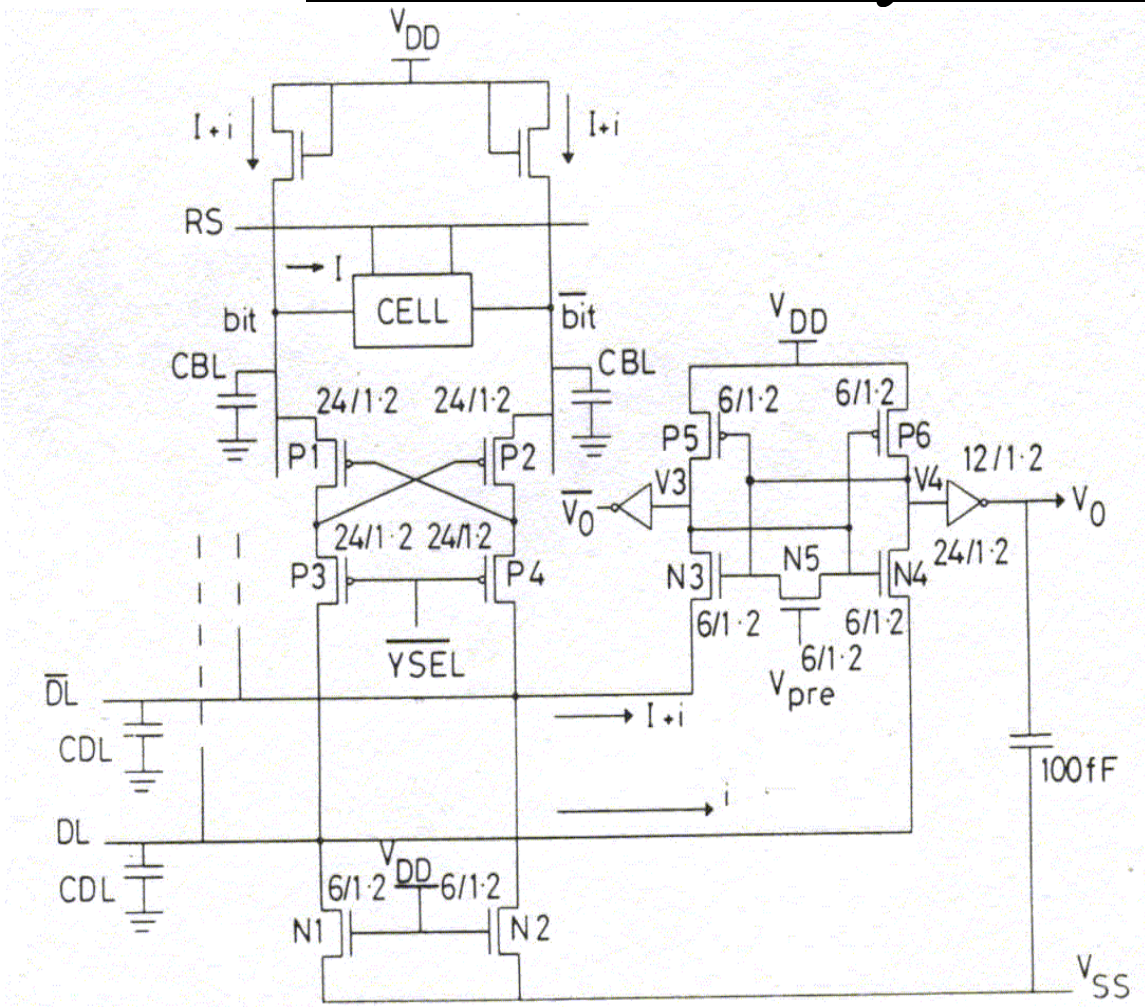
Current conveyor



For ideal current conveyor

Rin=0 and Iout=Iin

Current Conveyor based SA



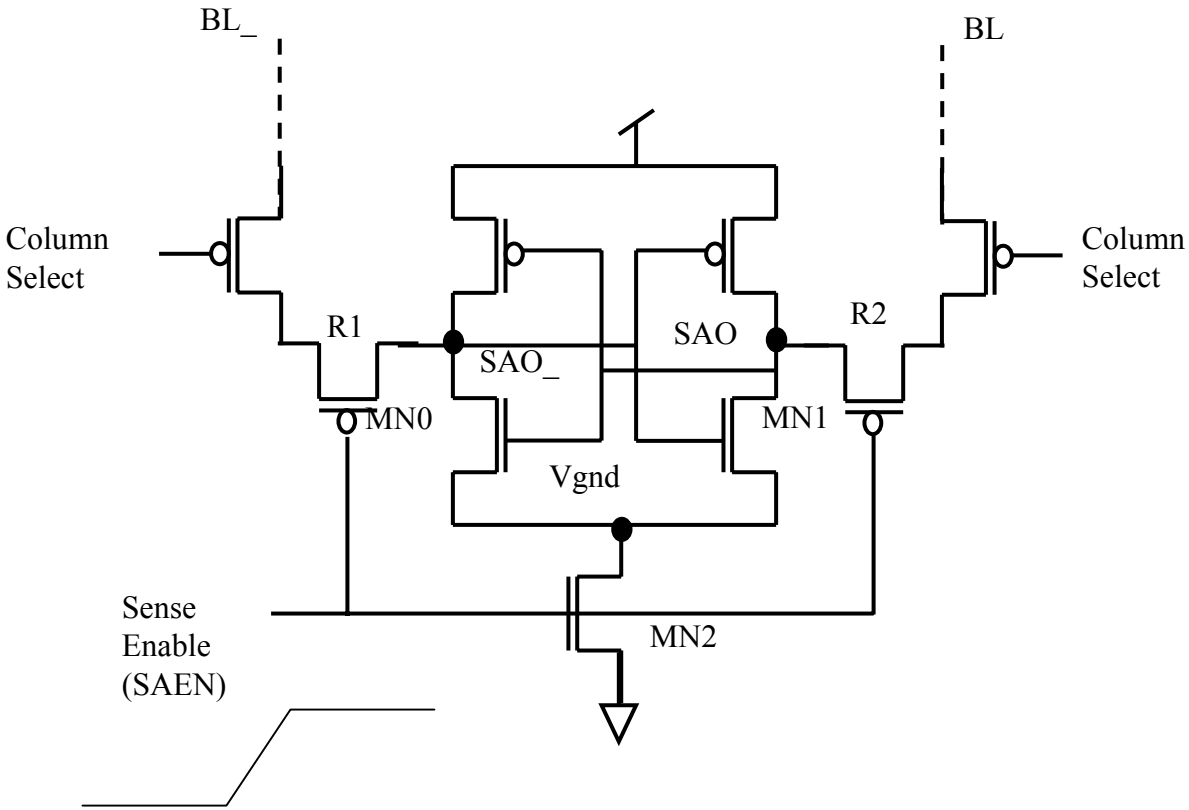
P1 and P2 provide negative feedback

P1 and P2 can be viewed as resistances whose value is modulated to suppress change in input voltage

P1 and P2 are in sat

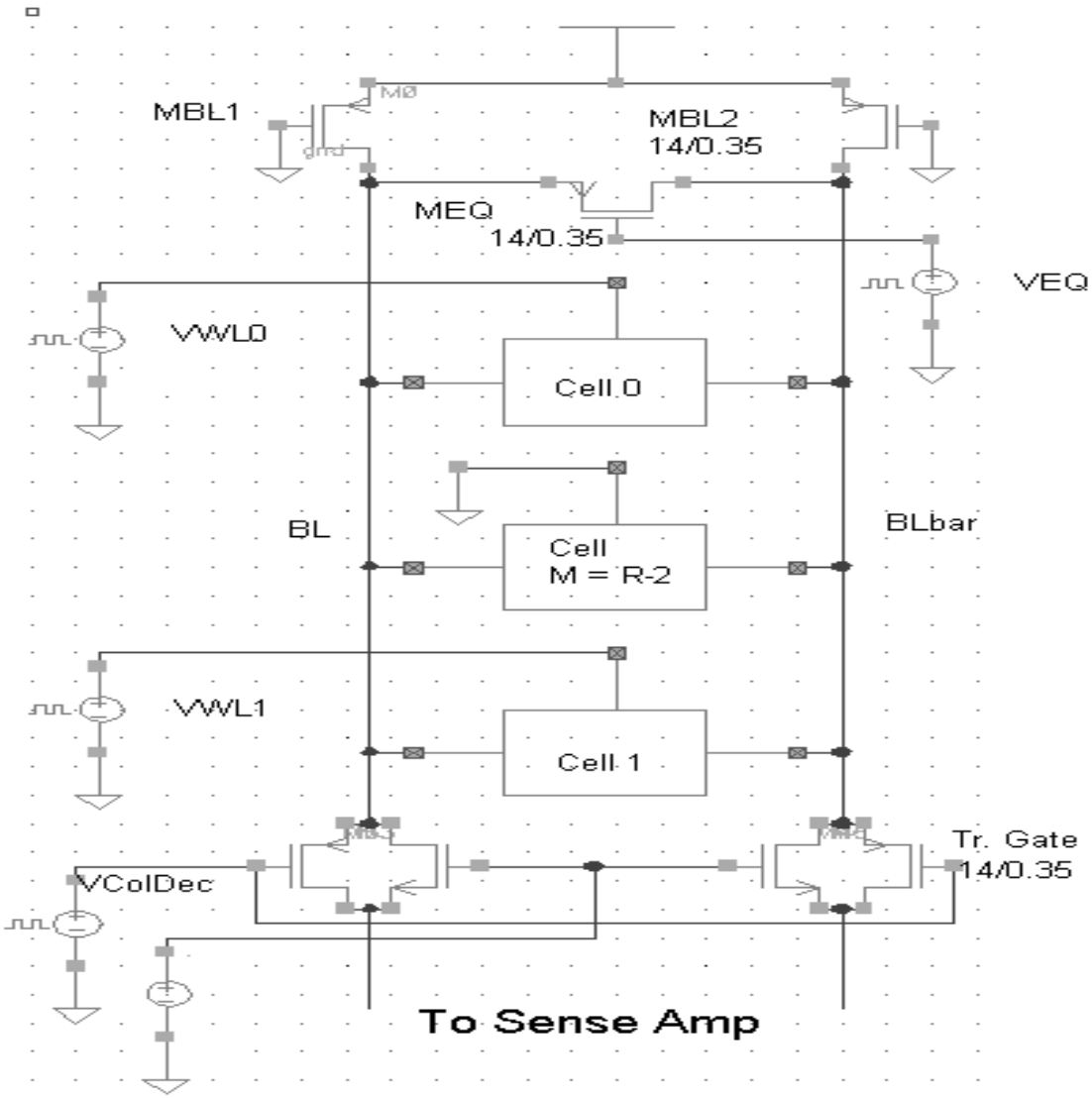
Latch type sense amplifier

Positive feedback Latch



Sense enable controls the latching operation

0.35µm Simulation Environment



The loading effect of other cells in the memory is taken into account

Alternate 1 and 0 is read from cell 0/1 to verify correct read operation

0.35µm Simulation Environment(contd.)

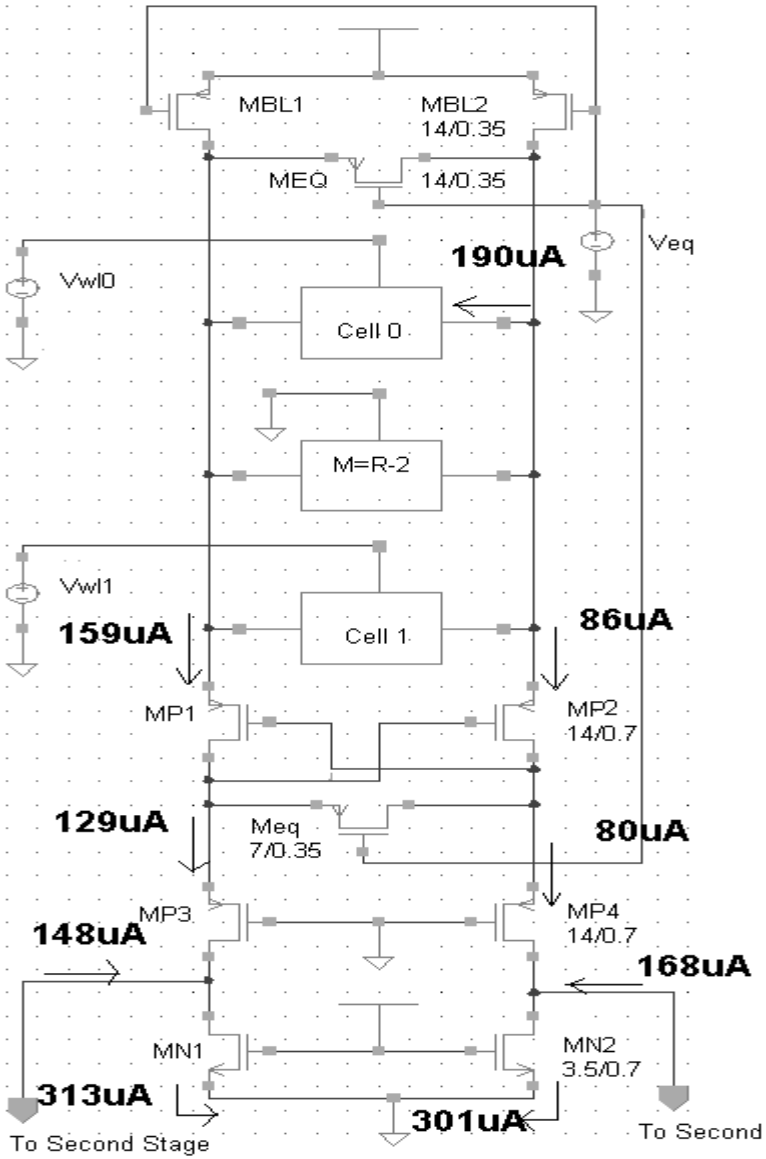
- Cycle time = 1.55 ns
- Supply voltage = 3.0 Volts
- Transistor models = Typical
- Temperature = 25˚ C
- Rise/fall time for pulsed sources = 200 ps
- Bit-line pre-charge turned OFF during sensing
- Amplifier transistor length = 0.7µm

Current Conveyor Simulation

- Circuit nodes do not reach steady state value for high-speed
- Divergence of cross-coupled nodes for high-speed operation
- Solved by equalization transistor
- Second stage activation with memory – wrong operation
- Second stage activation delayed for correct functioning
- Memory cell current not transported – attenuation

➢ *Performance degradation for high operating speed*

Current Attenuation in Current Conveyor



- Measured at end of cycle
- 2nd Stage deactivated

Memory cell differential current is 190μA

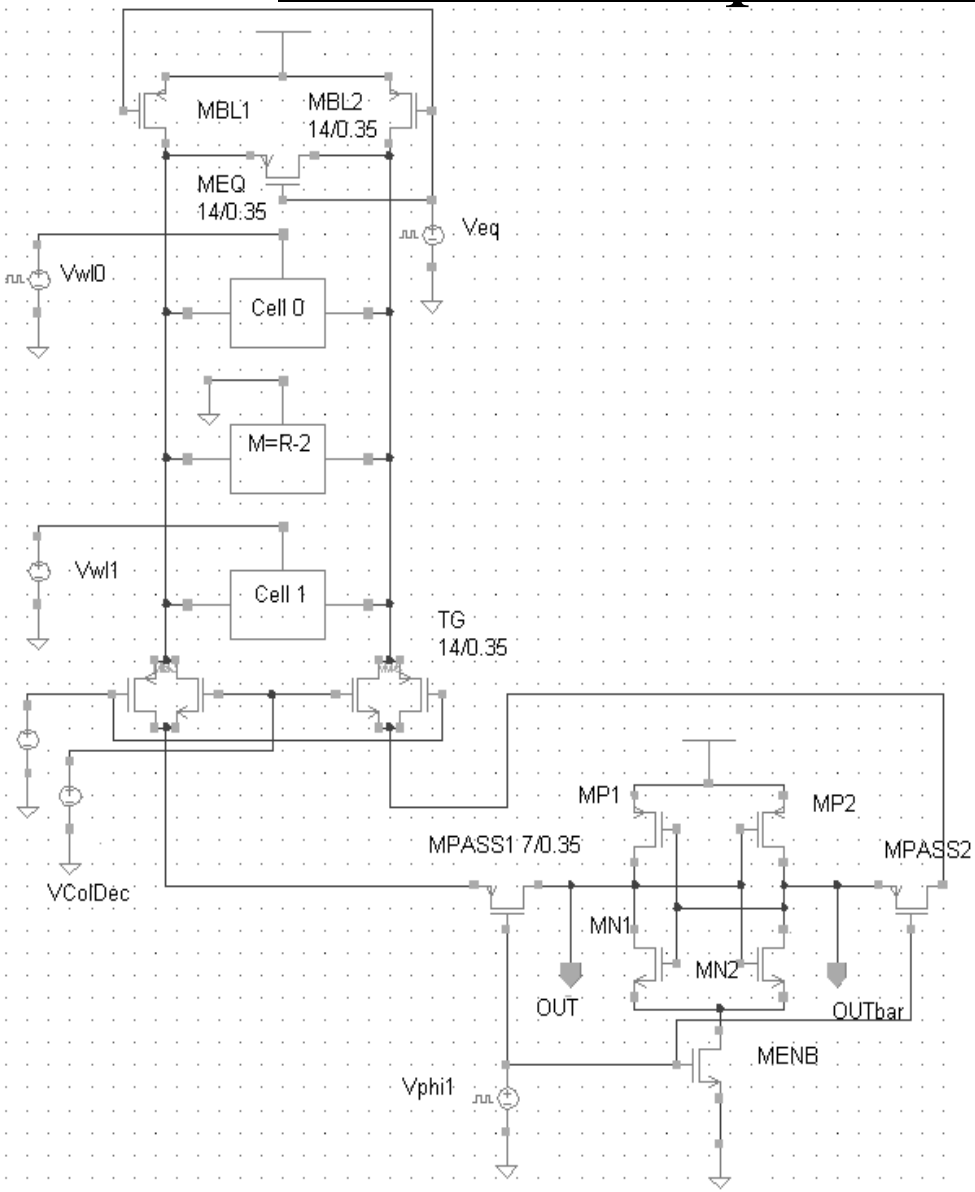
Differential current transported to the second stage is 20μA

Current attenuation due to capacitive effects at high speed

Current Conveyor Performance

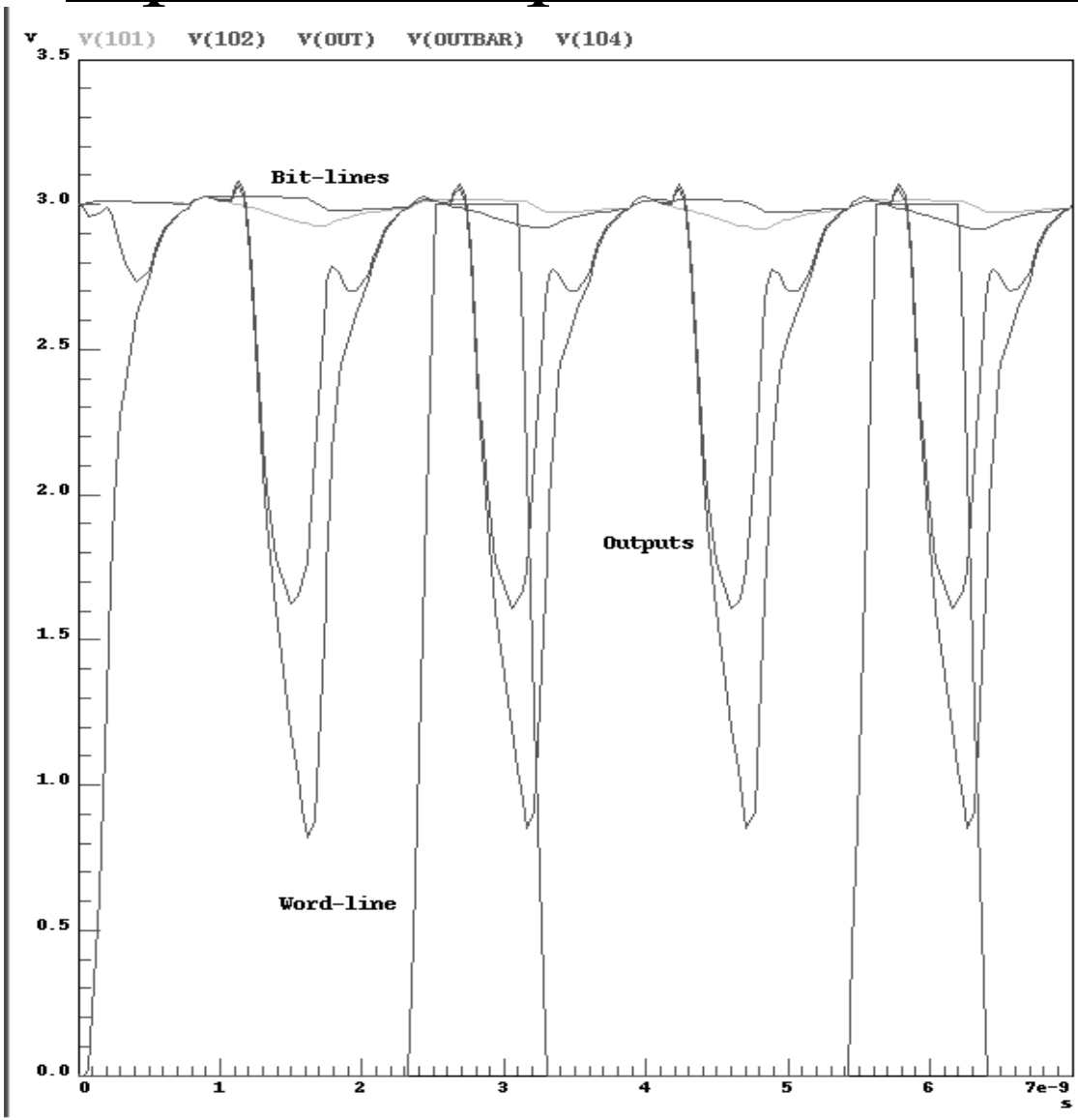
No. of Rows	Sense Delay	Bit-line differential	Power
256	500ps	0.11 Volts	3.52mW
512	605ps	0.07 Volts	3.53mW

Vdd bias Input Decoupled Latch SA



- Single stage
- Input decoupled by MPASS1,2
- Latching occurs after decoupling
- MP1,MP2,MN1,MN2: 2.1/0.7 nominal size

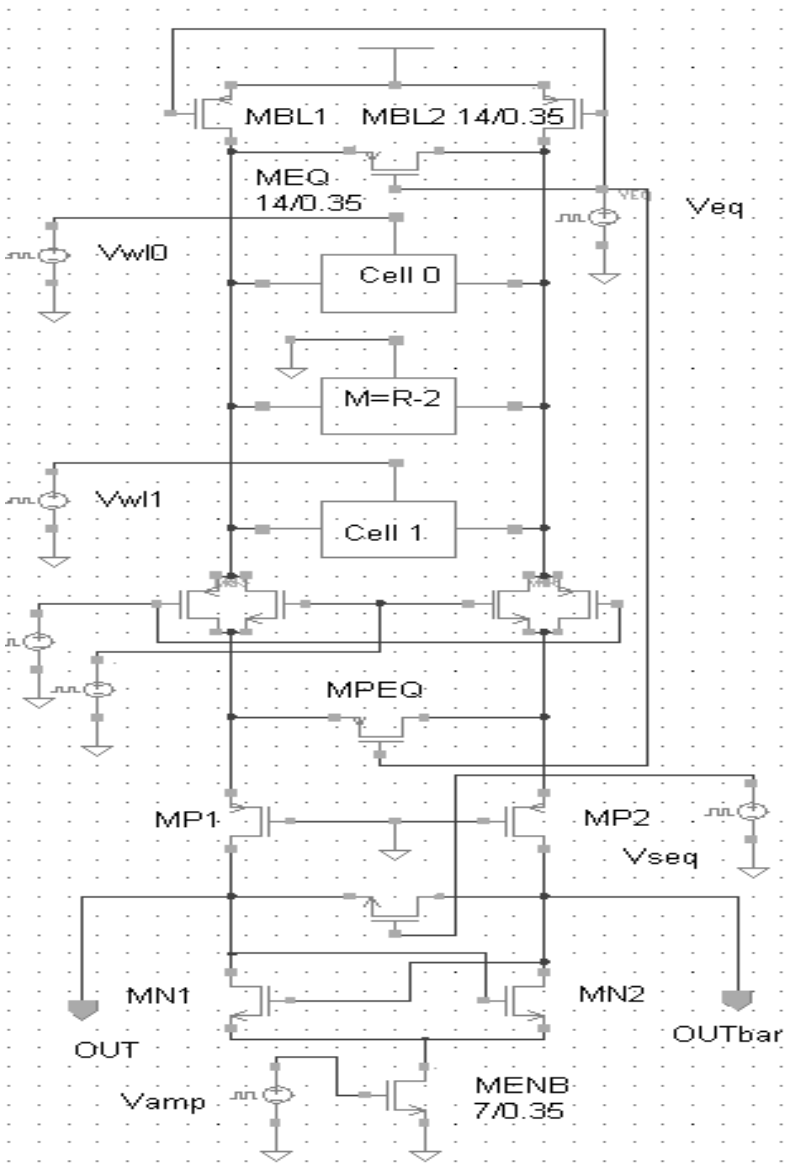
Input Decoupled Latch SA Waveforms



Initially both the outputs are biased at Vdd

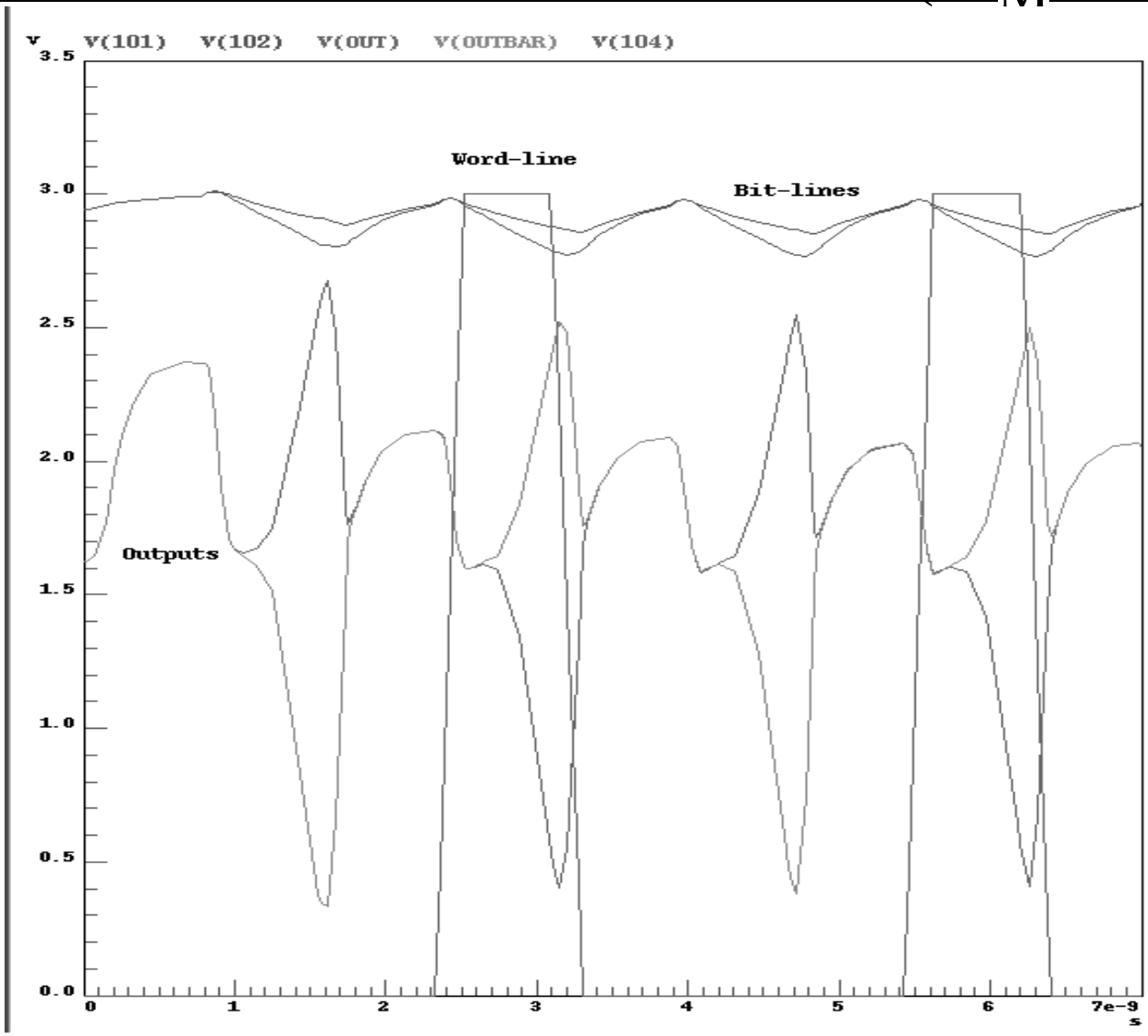
During sensing one of the output falls faster than the other output

Half Latch SA

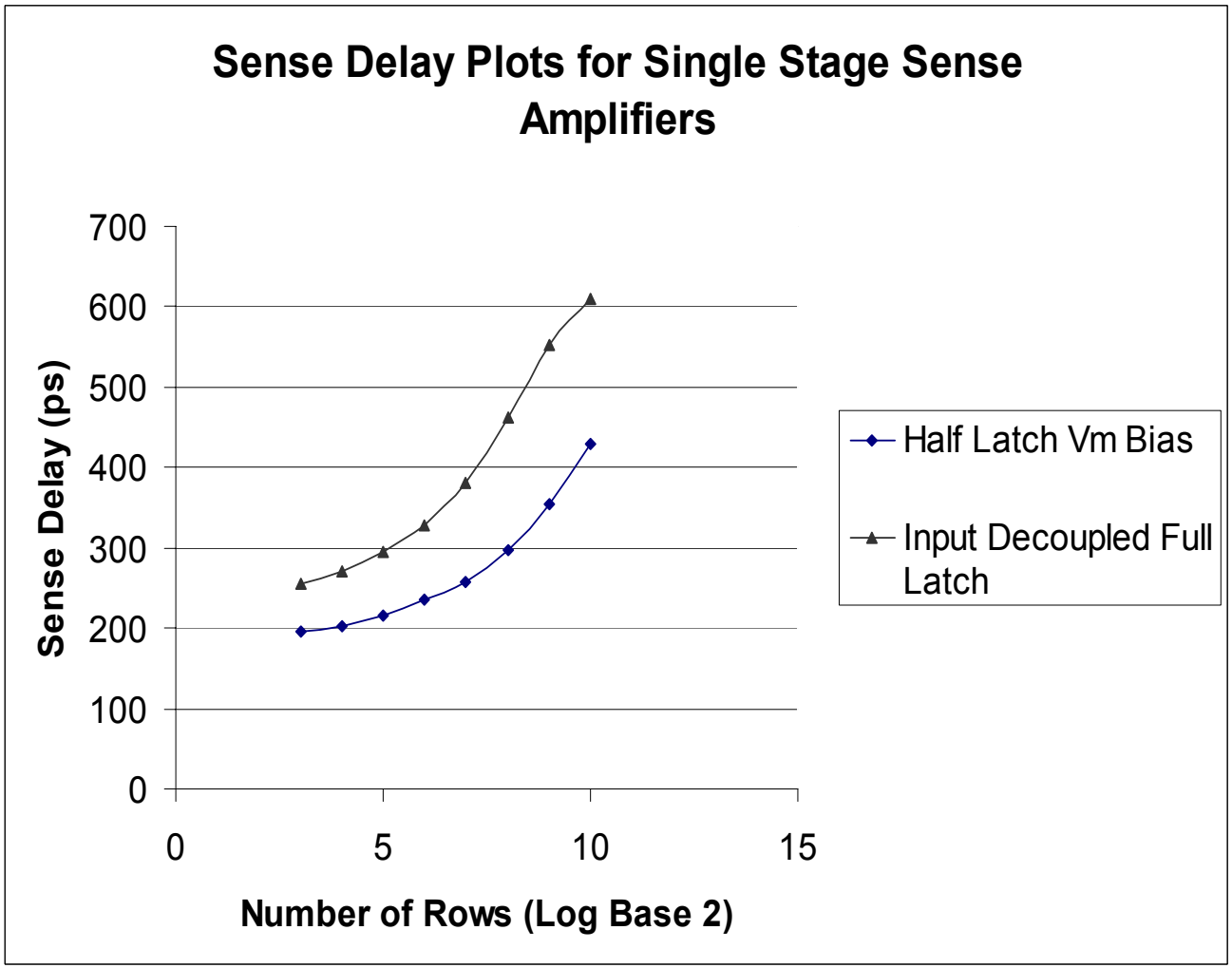


- Single stage
- Powered from the bit-lines
- V_M bias gives best performance
- PMOS transistors operate in linear zone
- NMOS transistors in saturation at start
- MP1,MP2,MN1,MN2: 2.1/0.7 nominal size

Half Latch SA Waveforms (V_M Bias)



Comparison of Single Stage Latch Style SA



SA Comparison for 512 rows

SA Type	Sense Delay	Bit-line Differential	Power
Curr.Conv.	605ps	0.07Volts	3.53mW
PMOS cc	624ps	0.09Volts	3.51mW
Half Latch	355ps	0.09Volts	2.54mW
I/P Dec. Lat.	552ps	0.09Volts	3.00mW

Frequency Domain Analysis

- Open loop gain-bandwidth analysis for single stage positive feedback latch style amplifiers
- Small signal input impedance analysis for "current sensing" schemes

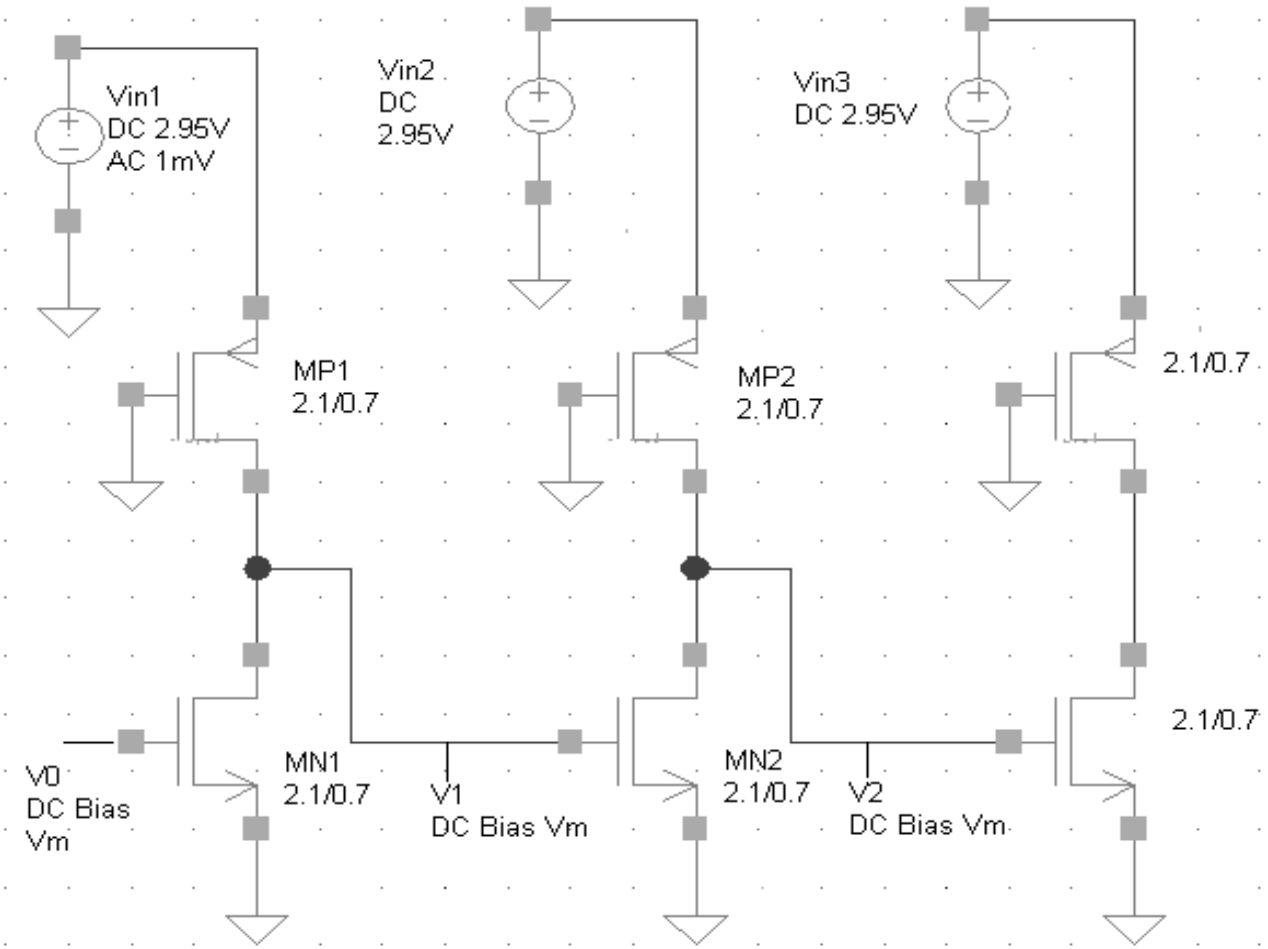
Advantages

- Minimum number of components needed
- Compare different amplifier structures
- Perform a preliminary sizing of amplifier transistors
- AC simulations will be much faster than the transient simulation

Open Loop Gain-Bandwidth Analysis

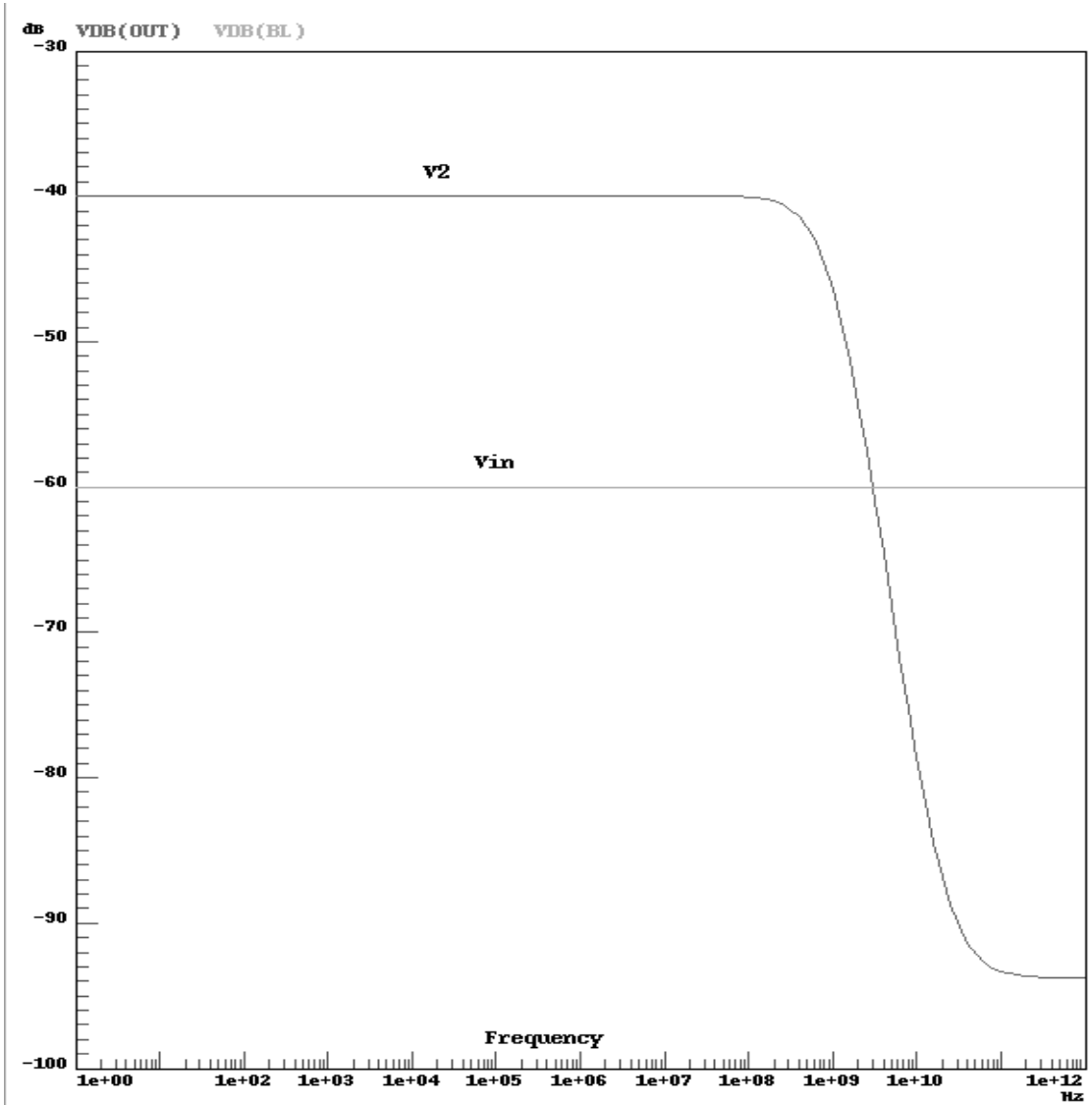
- Open loop GBW indicator of response speed of positive feedback amplifiers
- Cross-coupled inverter amplifier core of the single stage latch style SA
- Feedback loop opened with appropriate loading
- Input signals mimic the bit-line differential signals
- Frequency analysis done for open loop gain
- GBW studied for various amplifier transistor sizes

Open Loop Circuit for Half Latch



Dummy transistors are added to simulate the loading

Gain Characteristic for Half Latch



3-pole 3-zero response up to 1 THz

For figure shown

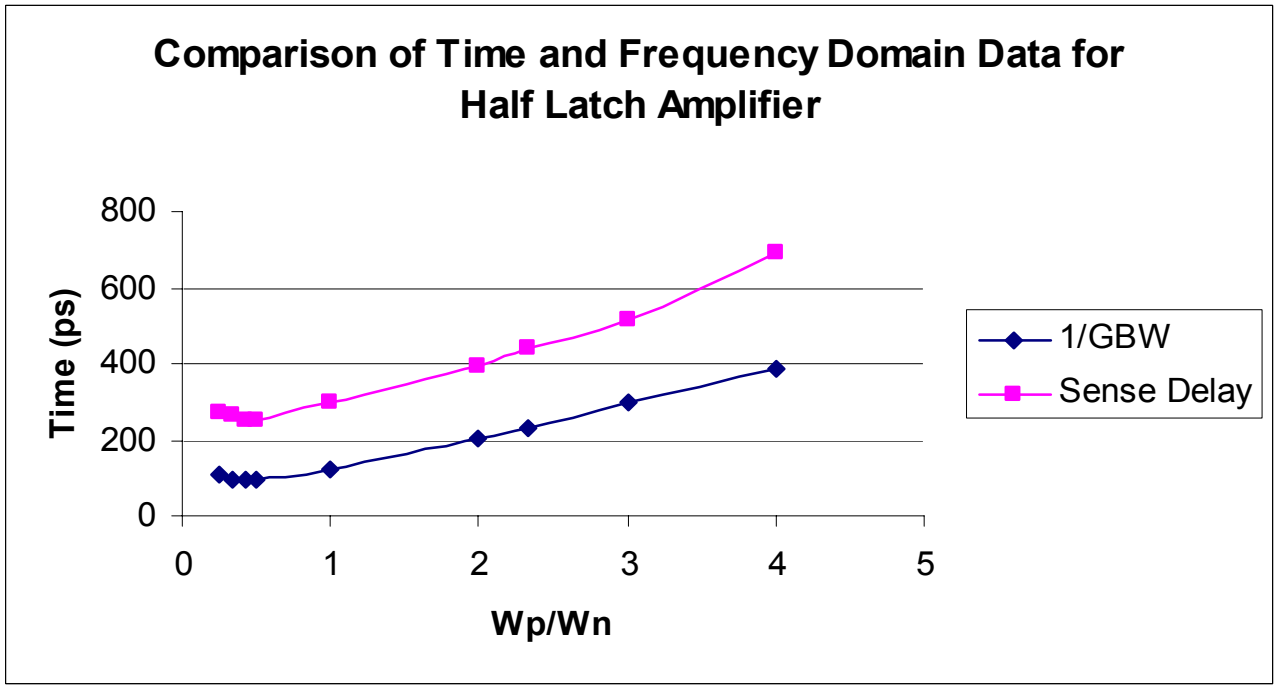
Gain=20dB

Bandwidth=802MHz

GBW=8.02GHz

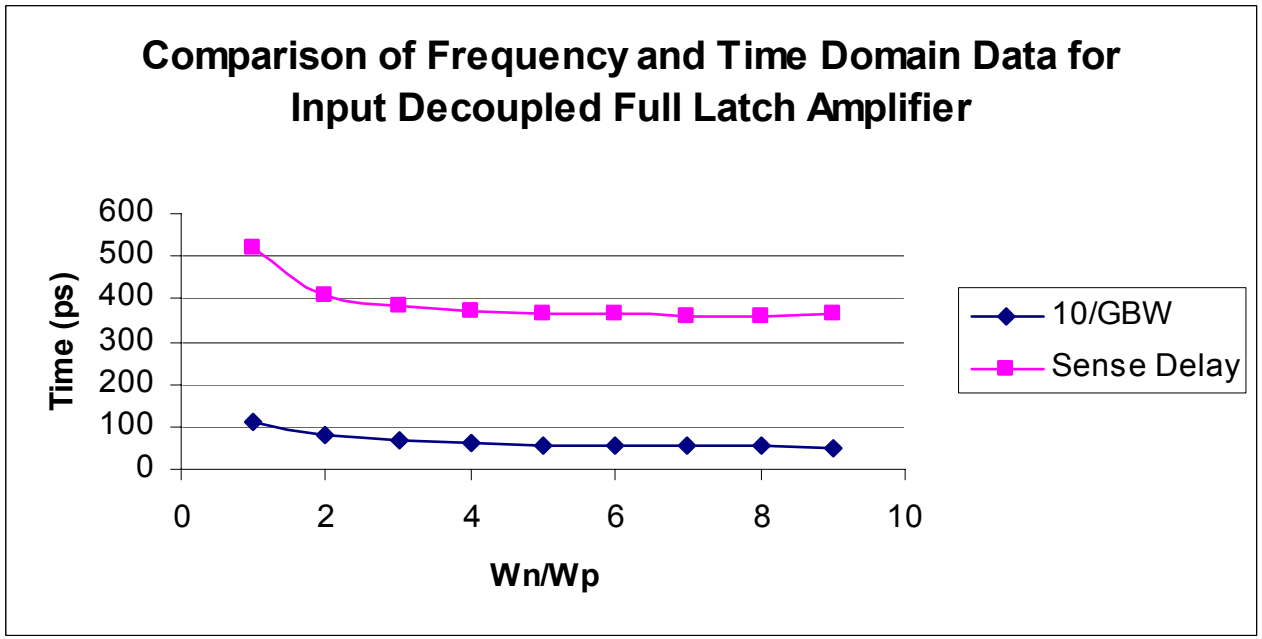
Similar responses for other two amplifiers

Time and Frequency Domain Results-Half Latch Amplifier



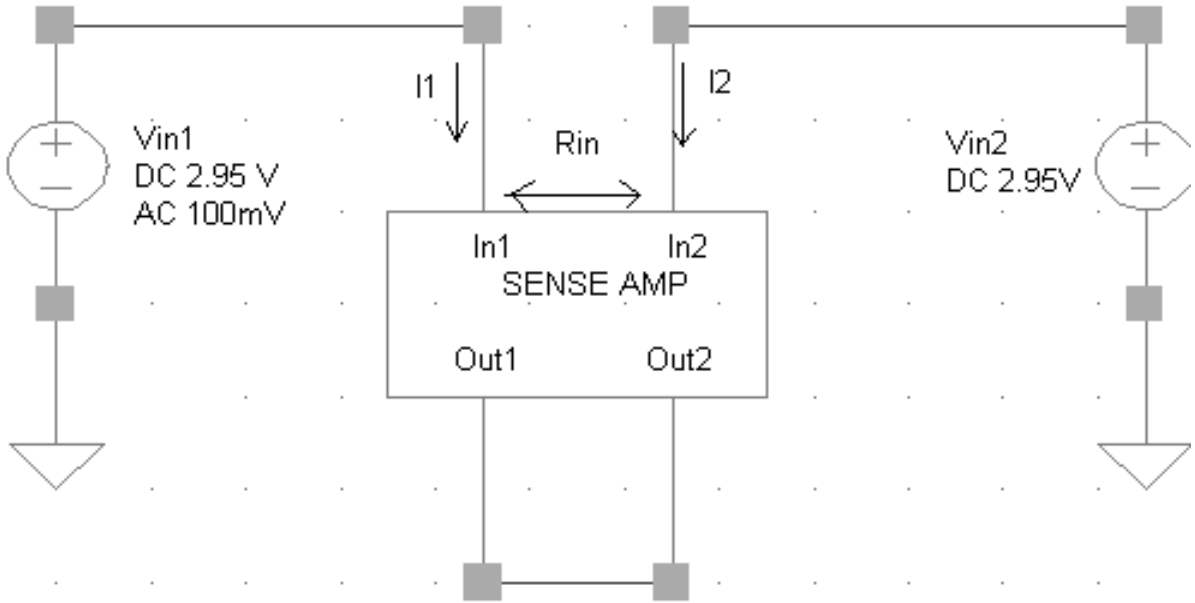
$(GBW)^{-1}$ compared with time domain sense delay for varying transistor sizes

Time and Frequency Domain Results: I/P Decoupled Latch Amplifier



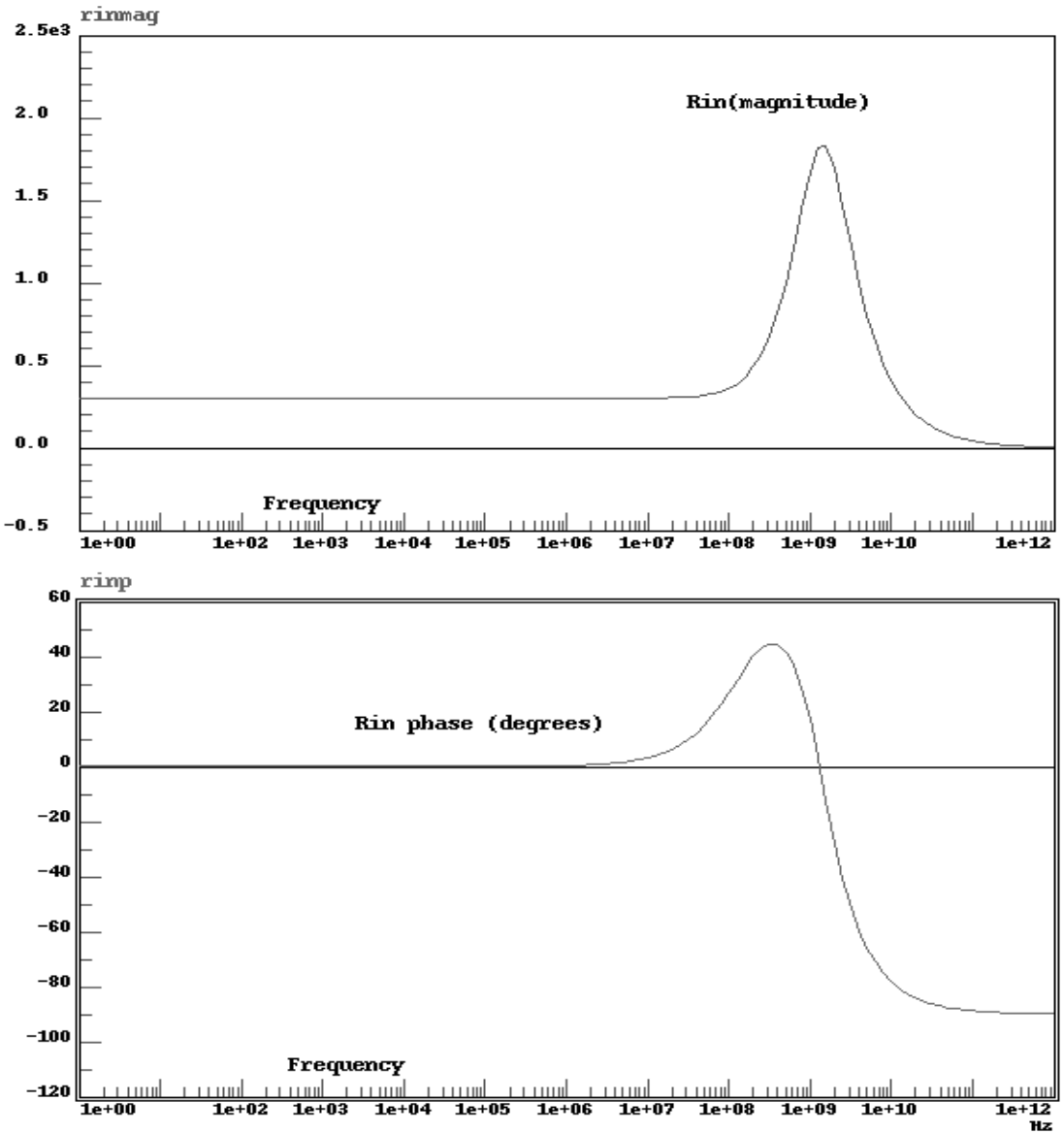
Scaled $(GBW)^{-1}$ compared with time domain sense delay for varying transistor sizes

Small Signal Input Impedance Study



- Circuit to measure input impedance
- Differential output = 0
- Differential input impedance measured as a ratio of differential input voltage and current

Input Impedance of Current Conveyor



2-pole 1-zero response up to 1THz

DC value =302Ω

Peak value = 1.83kΩ at 1.29 GHz

For 1.55 ns cycle time 1.3GHz is twice fundamental frequency

INPUT IMPEDANCE NOT LOW FOR HIGH SPEED OPERATION.

Summary of SA architecture

- Half latch and input decoupled latch amplifiers found to give best speed performance
- Current conveyor based schemes are not well suited for high-speed applications
- Frequency domain analysis useful tool to study positive feedback amplifiers
- *Access time limitations expected to increase with increasing densities*

Impact of Mismatch on Analog Design

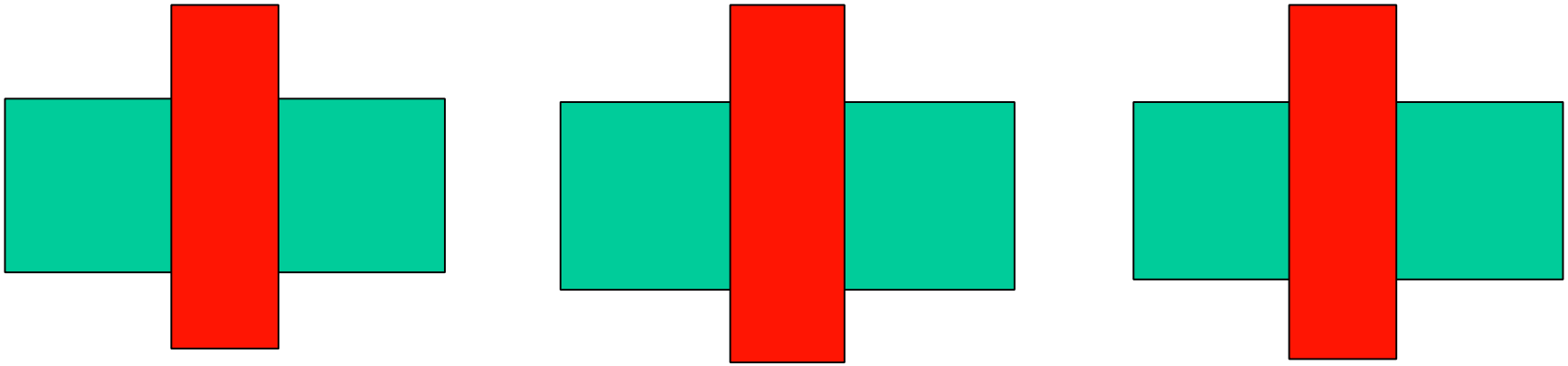
Acknowledge the contributions made by Ravpreet Singh and Srinivasaiah

Transistor mismatch in deep sub-micron technology

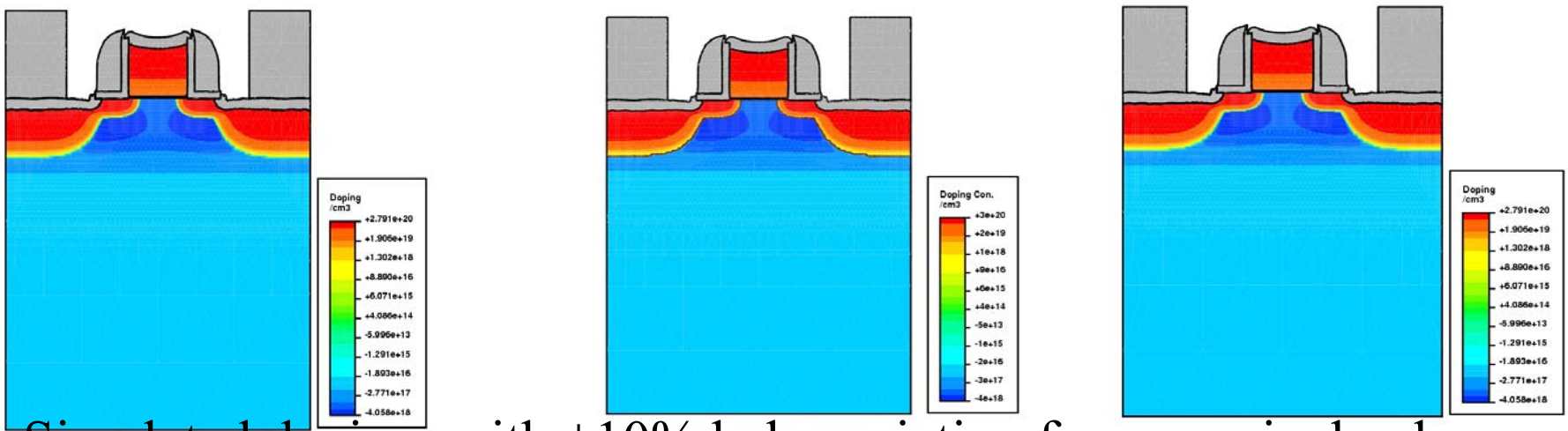
- Factors causing transistor mismatch
- Modeling the transistor mismatch
- Controlling mismatch effect at process/device level
- Impact of transistor mismatch in sense-amplifier design
- Controlling mismatch effect at circuit level

Transistor Mismatch Effects

3 identical transistors in a chip at the circuit design phase

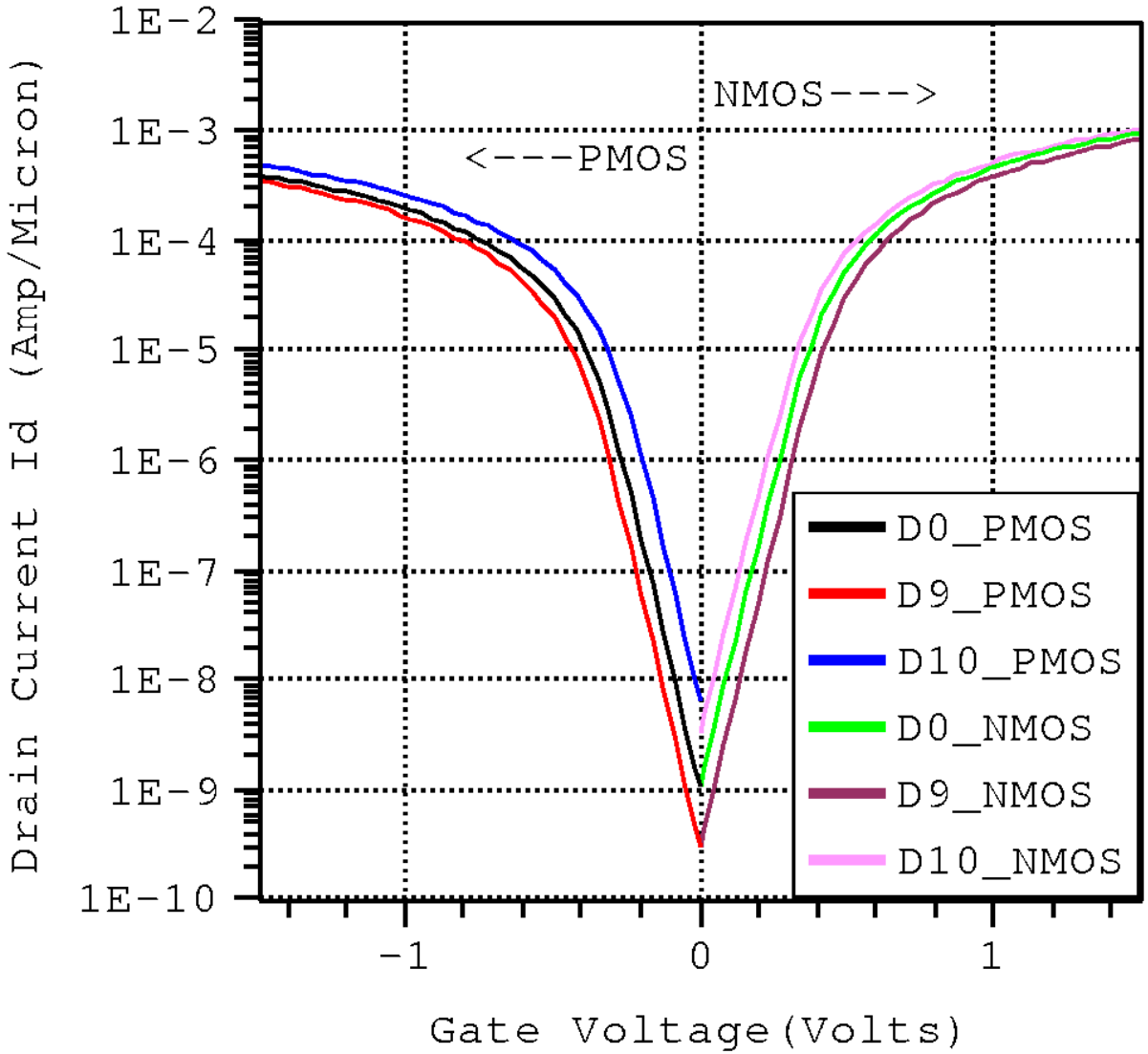


The structure of 3 transistors after the completion of IC processing



Simulated devices with ±10% halo variation from nominal value

Impact of process variation on drain current



Factors Causing Mismatch

1.*Intrinsic type*

- Discrete dopant effect
- Interface state density fluctuations

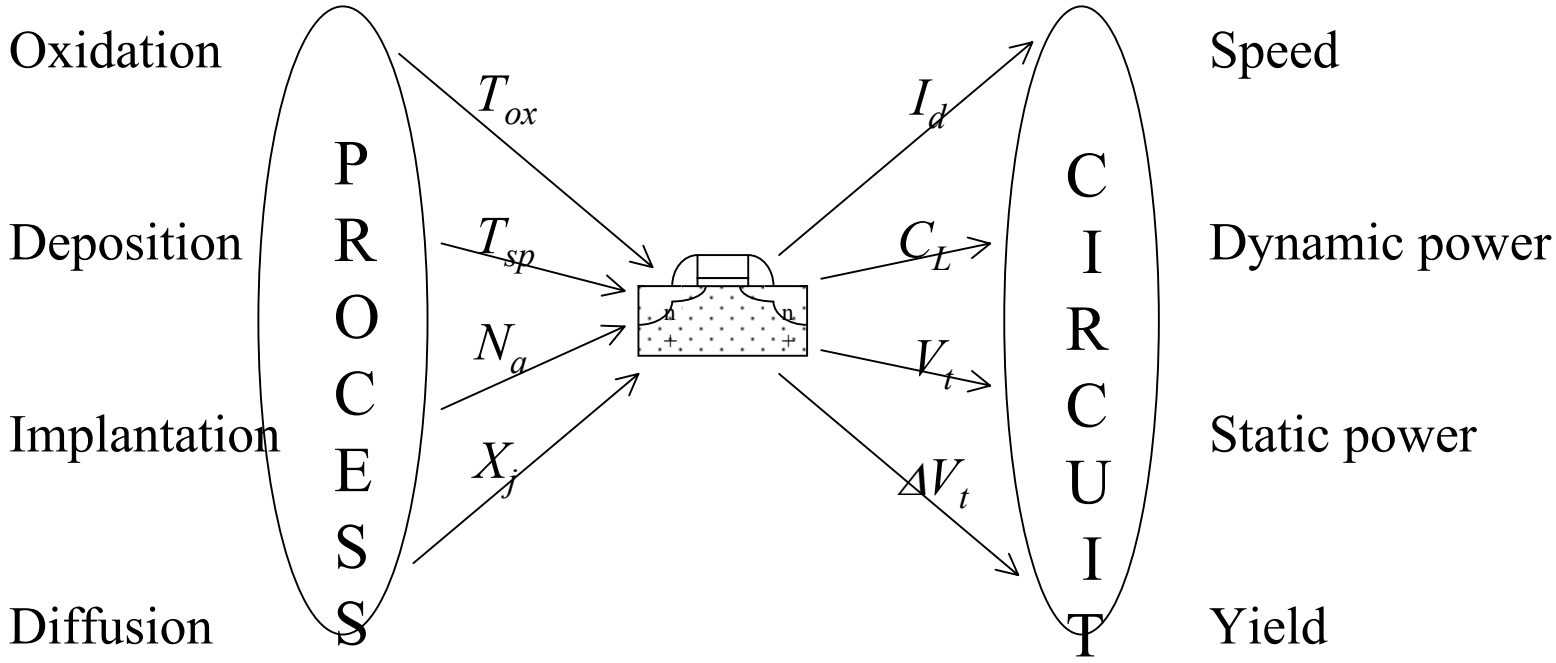
2.*Extrinsic type due to random variation in:*

- Gate length and width
- Oxide thickness
- Implant dose
- Implant energy
- Anneal temperature
- Gate & S/D overlap
- Spacer thickness

Device parameters affected by process parameters

- I_{off}, the leakage current
- I_{on}, the saturation current
- V_t, the threshold voltage
- S, the Sub threshold slope
- g_m, the Tranconductance.
- Various R s, C s and parasitics

Impact of process parameters on circuit parameters



- Circuit performance has a direct relation on process in a complex way.
- The relation between circuit parameter to process parameter is highly nonlinear.
- Some of the Process level parameters are statistically correlated.

Short range and long range order

Inter die, Inter wafer and Inter lot variations have long range order

Intra die variations can be medium and short range

The variation along a clock tree has medium range order

The variation in matched differential pair is short range order

SPICE corner models are derived from long range order and they will be very pessimistic for short range order

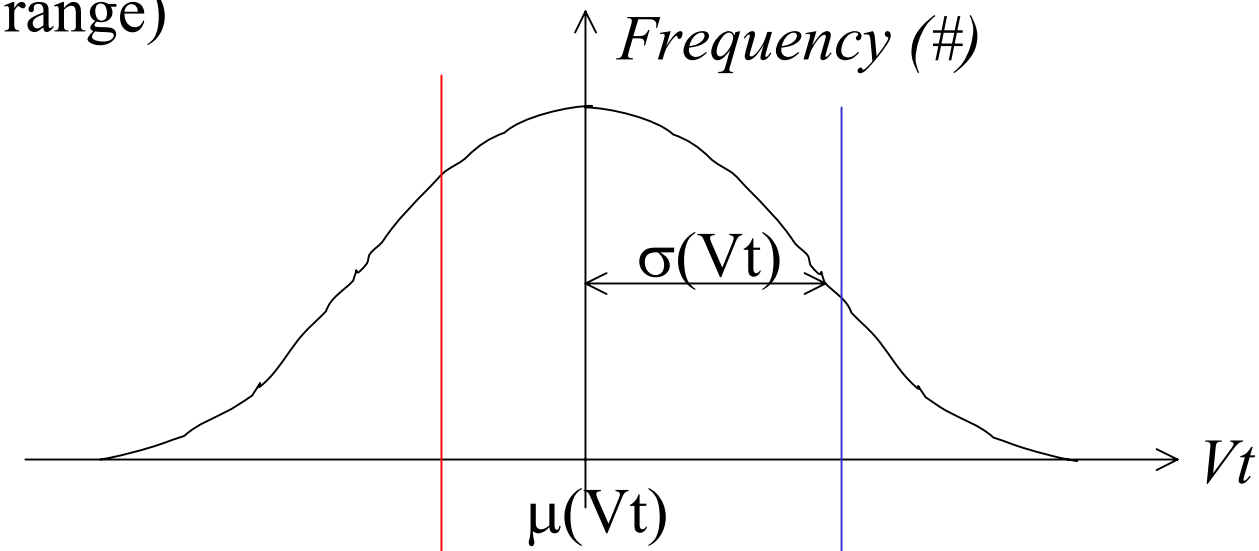
Typically the intra die variation of a parameter P between two transistors M1 and M2 is given by

$$\sigma^2(P_1 - P_2) = \frac{a_p}{2W_1L_1} + \frac{a_p}{2W_2L_2} + s_p{}^2 D_{12}{}^2$$

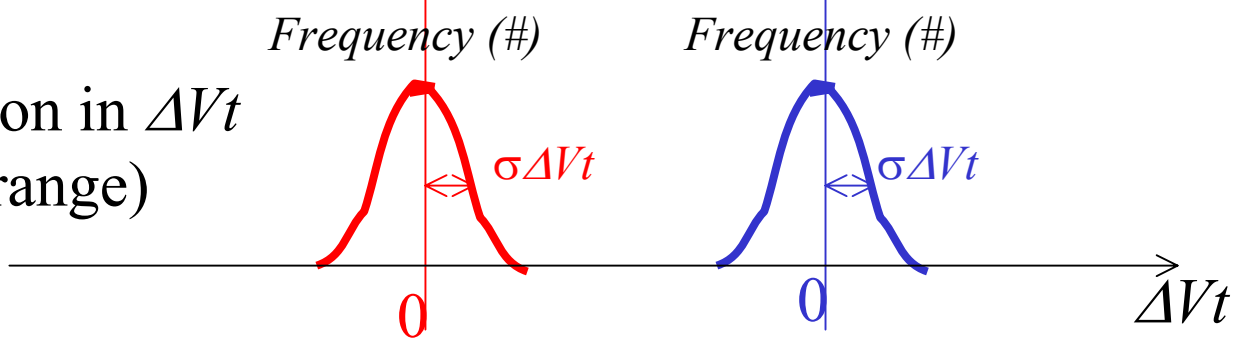
σ^2 is variance, D_{12} is distance between M1-M2, a_p and s_p are Process technology dependent

Distinction in variation of a parameter vs. variation in matching of parameter

Variation in Vt
(long range)

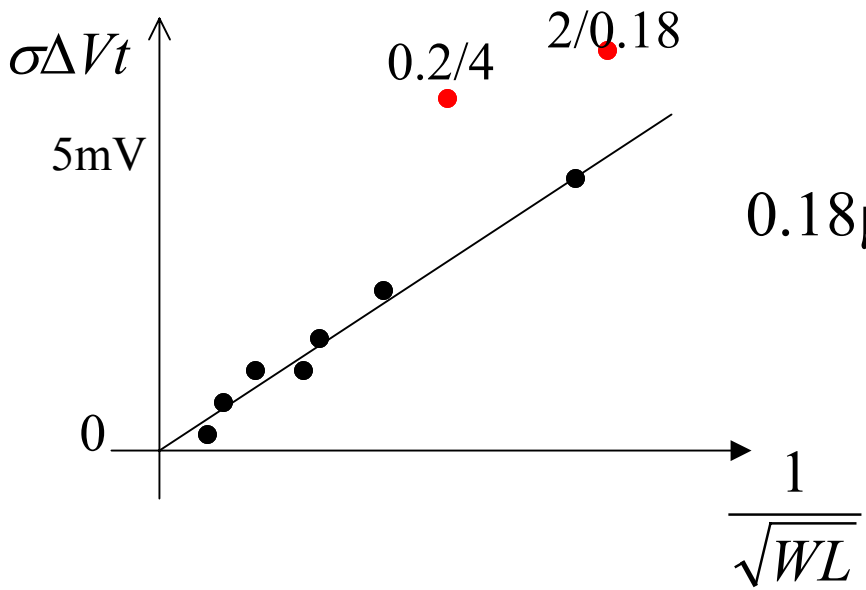


Variation in *ΔVt*
(short range)



Mismatch coefficient

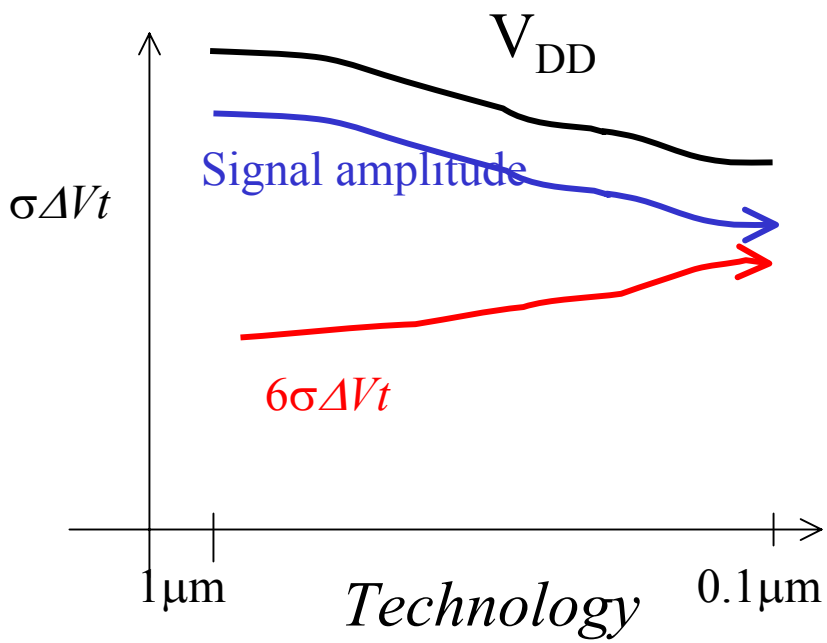
$$\sigma\Delta Vt = \frac{A_{VT}}{\sqrt{WL}} = \frac{qT_{ox}\sqrt{2NW_d}}{\varepsilon_{ox}\sqrt{WL}}$$



0.18μm technology

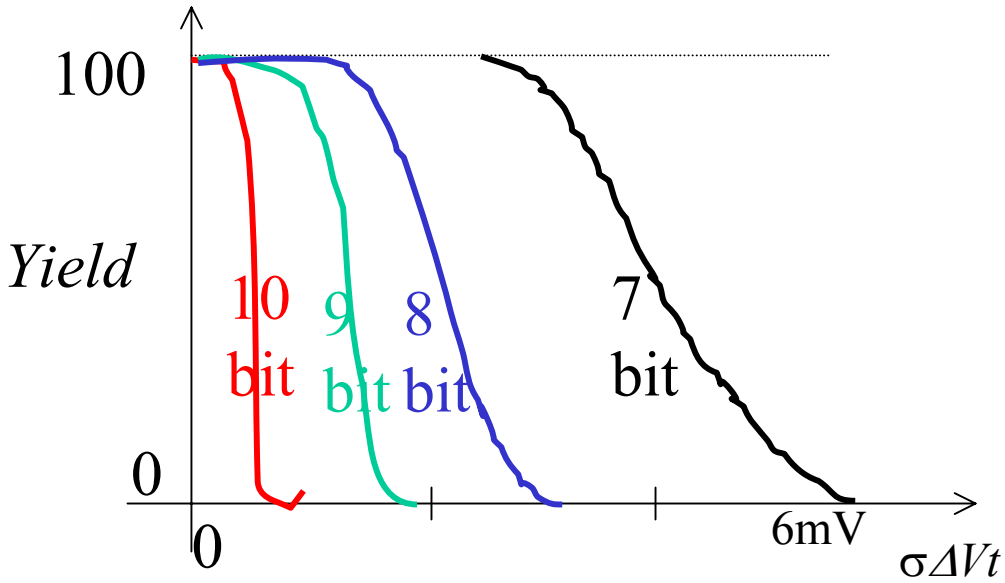
The equation is obeyed very well except at the edge of the Process technology

Scaling Trend



The parameter matching is becoming difficult with scaling

ADC Yield



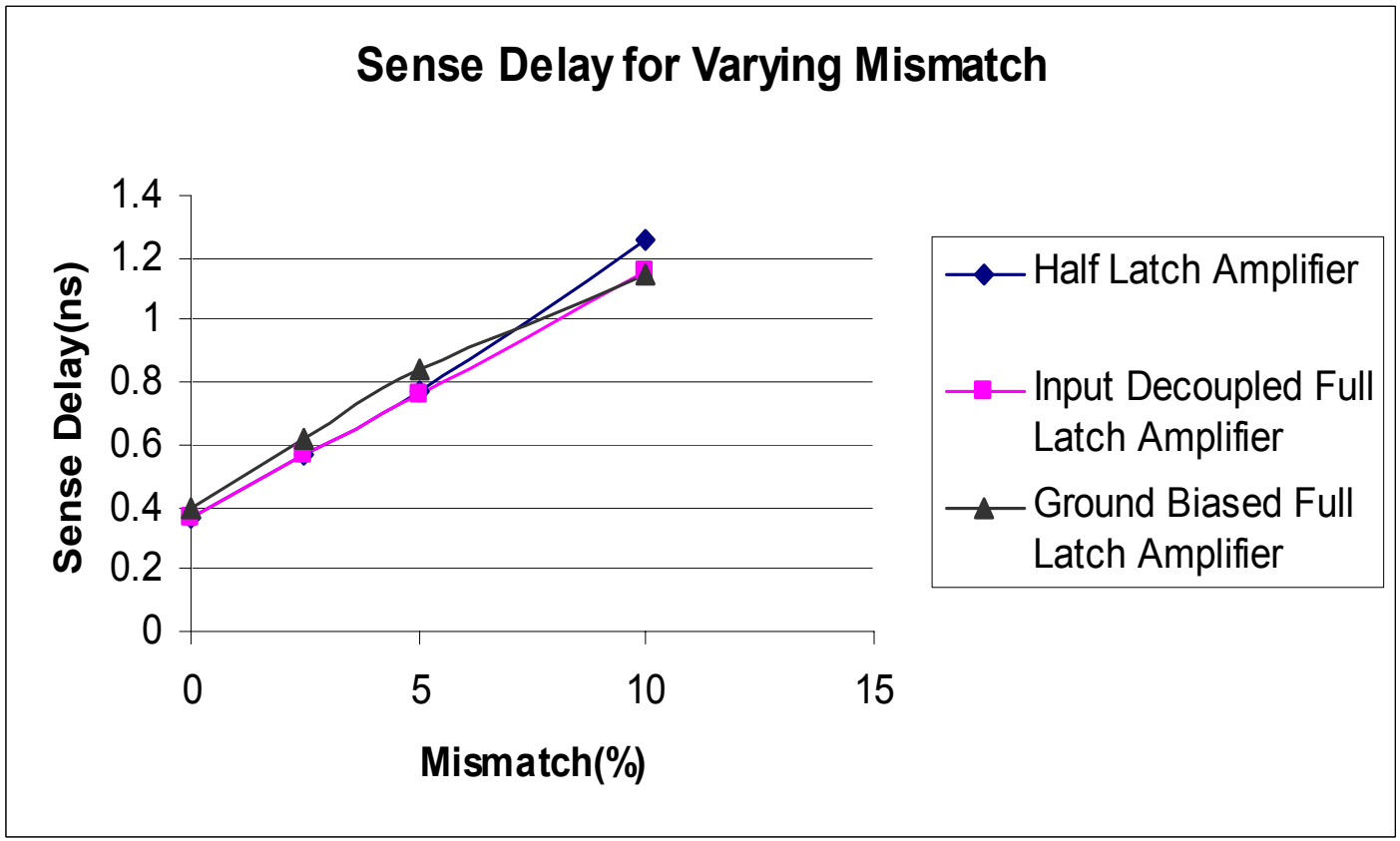
The higher precision requires very low mismatch

The yield for high precision drops off very fast

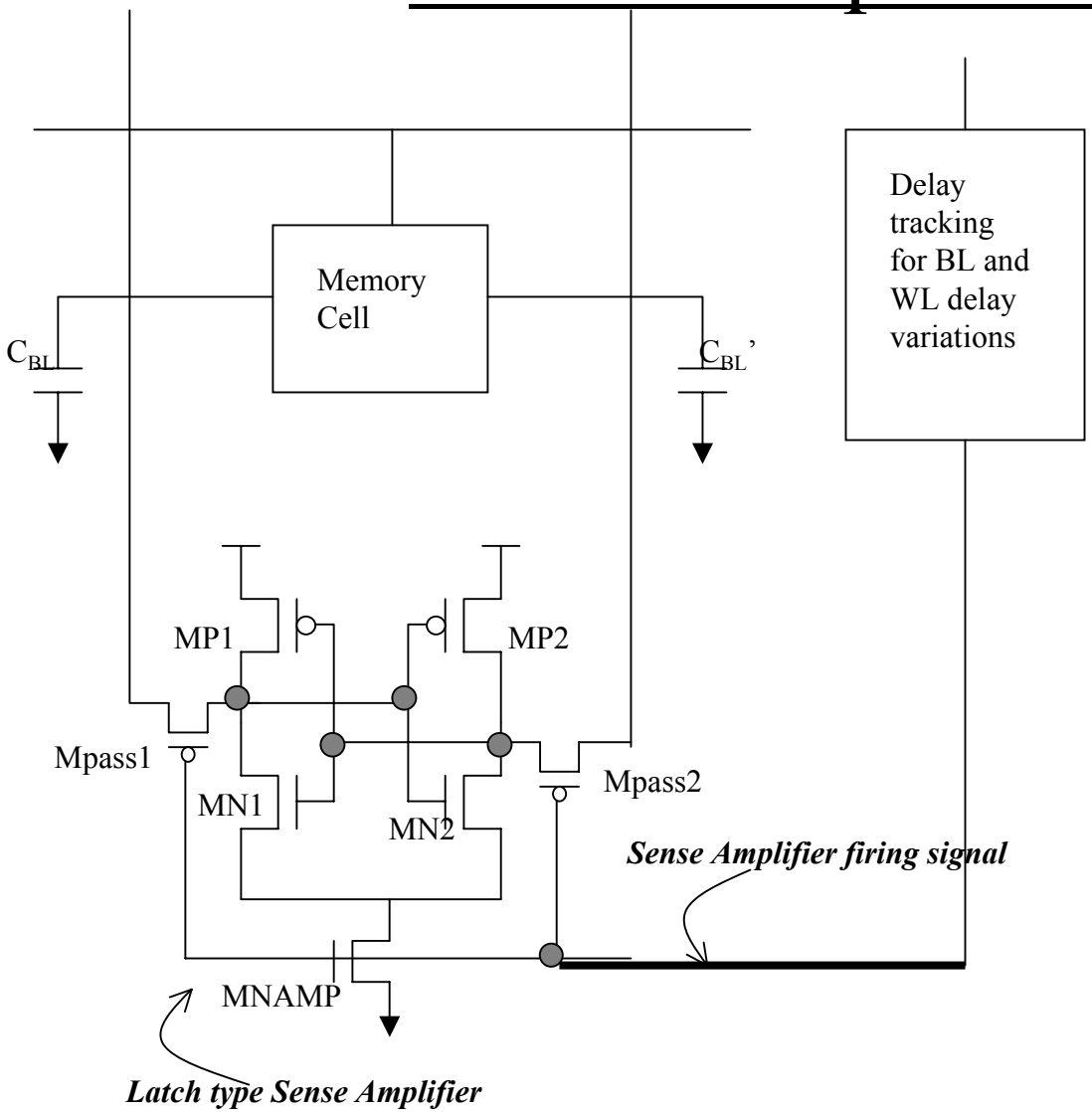
Offset Effects

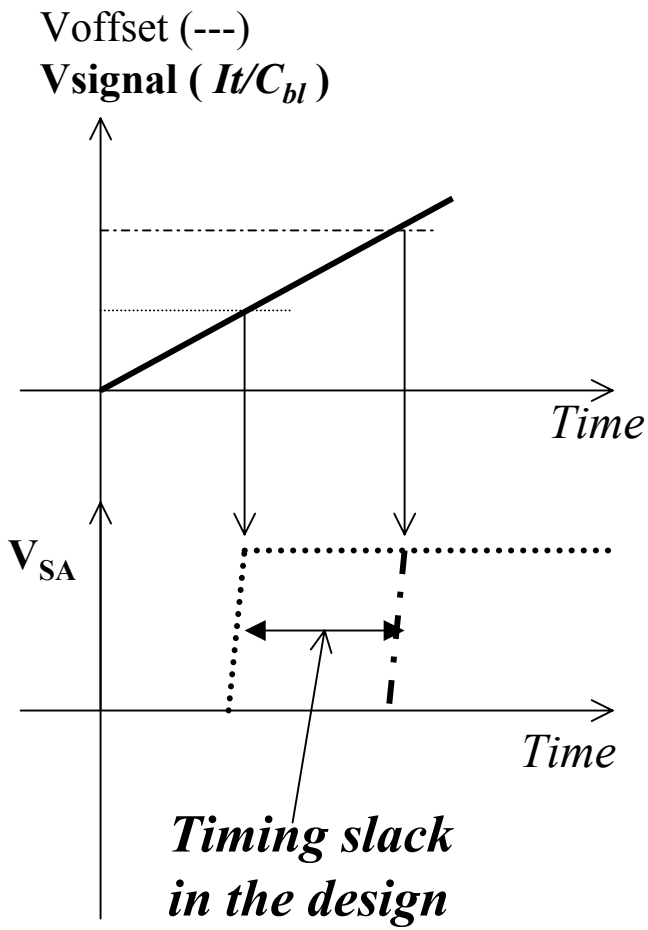
- Mismatch effect modeled by variation in β and V_T
- Sense amplifiers can be activated only after input signal voltage (bit-lines) compensates offset voltage
- Different degrees of mismatch – 10%,5%,2.5%
- Worst case offset voltage generated for each sense amp
- Matched pair of transistors – (max V_T, min β) and (min V_T, max β)

Comparison of Latch Amplifiers with Offset (contd)



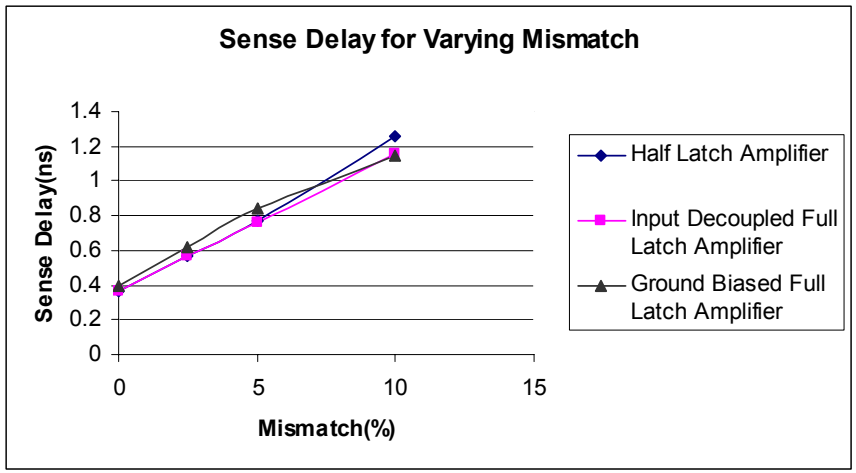
Effect of input offset on SA

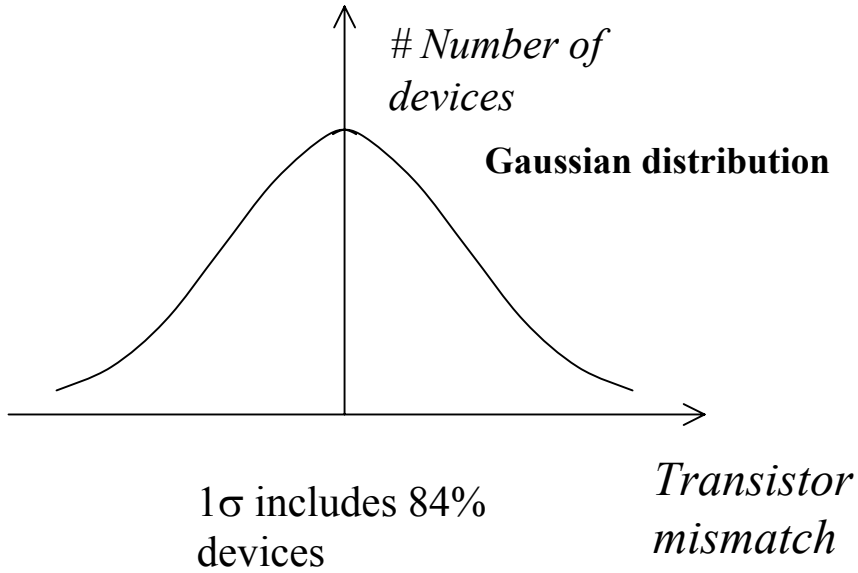




- Higher the offset, longer the delay for proper functionality

Impact of transistor mismatch on SA Delay

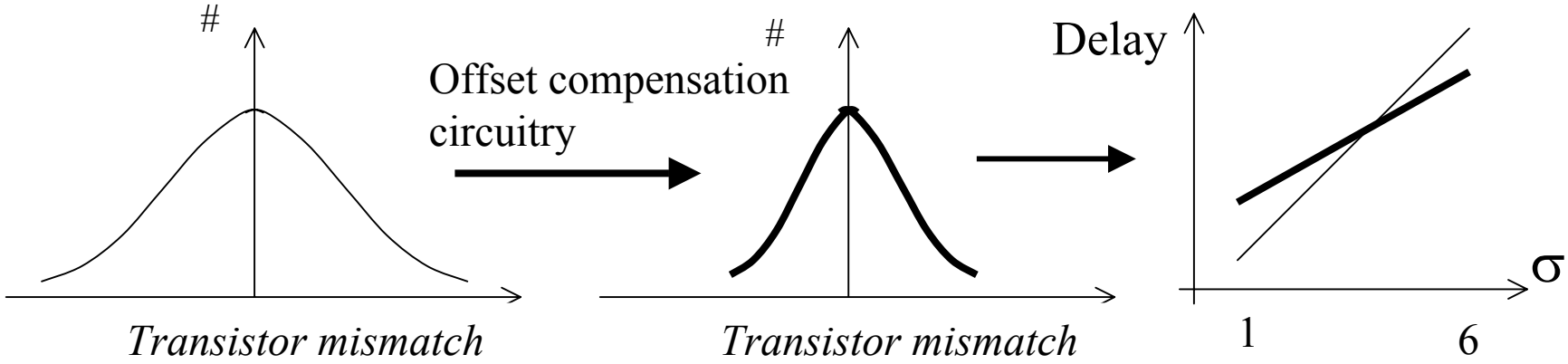




Design Index (σ)	Percentage of functional devices
1	84.1344740
2	97.7249938
3	99.8650033
4	99.9968314
5	99.9999713
6	99.9999999

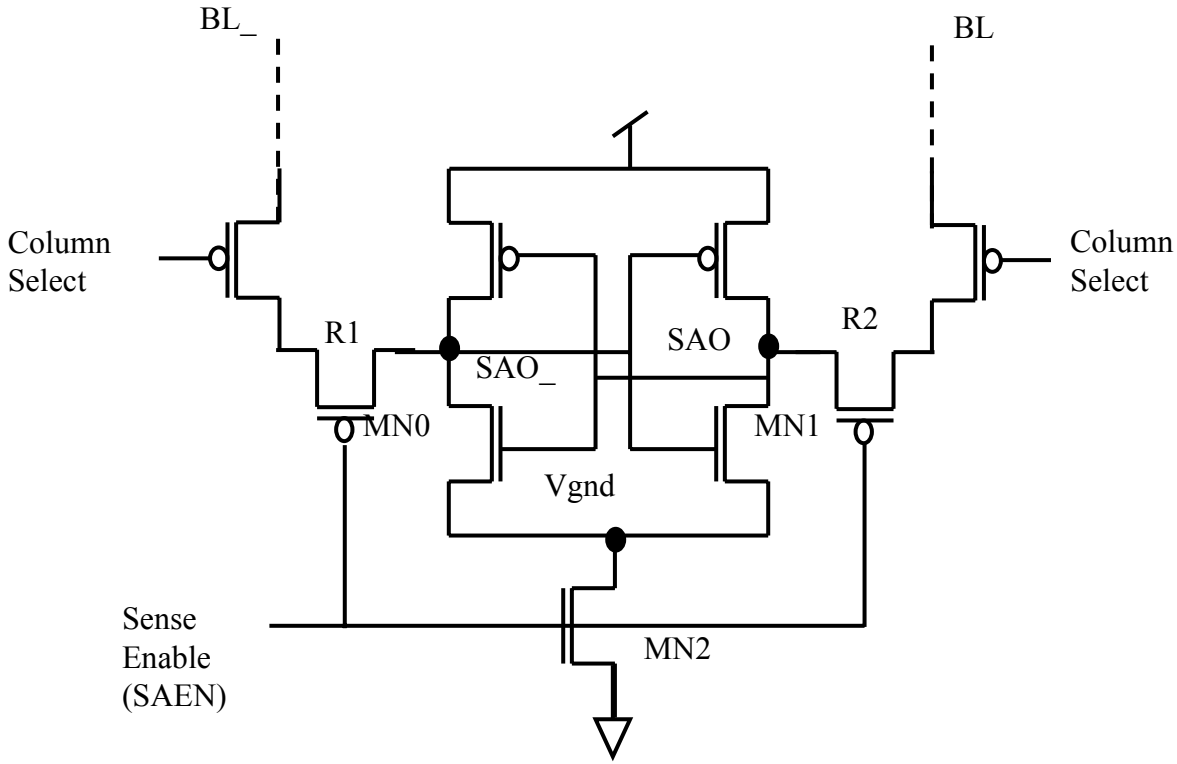
- 6σ design index required from circuit yield consideration
- Conservative design compromises the circuit speed (factor of 2 or more!)
- This is a typical speed versus yield trade-off which exists universally

Conventional solution



- Adding offset compensation circuit to Sense-Amplifier
- Increased number of transistors connected to SA output
- Added circuitry increases SA output node capacitance
- Intrinsic SA response becomes sluggish

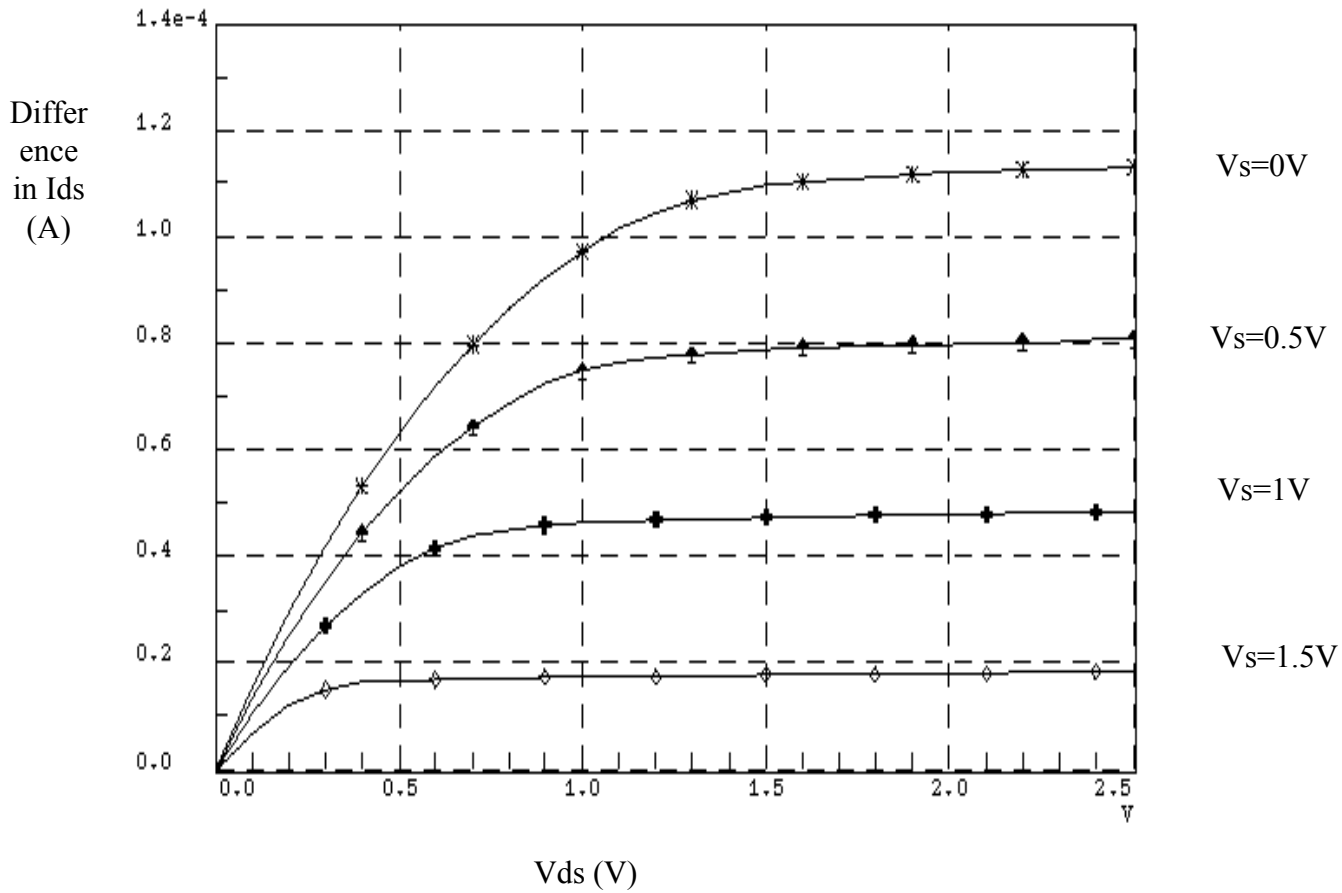
Offset Analysis of Latch-type sense amplifier



$$T_{datapath} = \frac{C_{BL} \times \Delta V_{BL}}{I_{cell}} + \tau_{SA} \qquad \frac{P_{dynamic}}{f} = C_{BL} \Delta V_{BL} V_{dd}$$

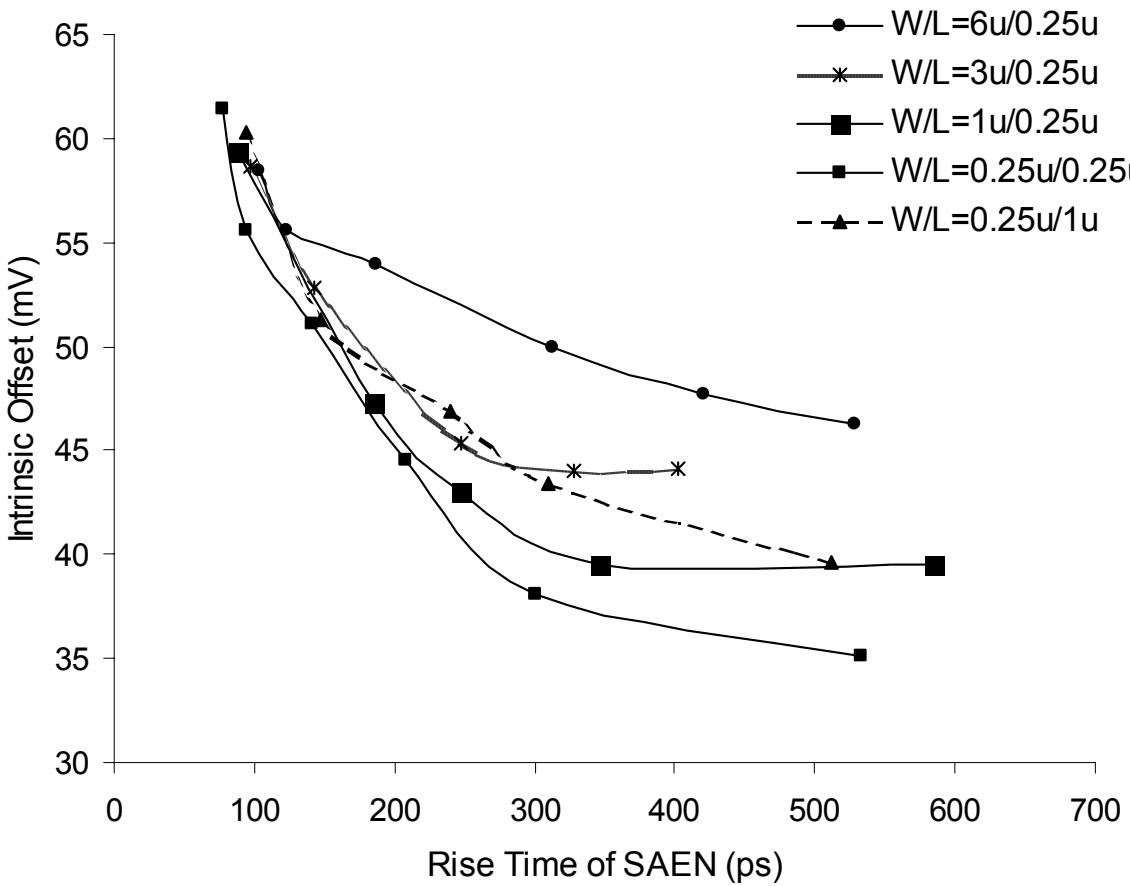
Latch type sense amplifier is preferred configuration for high speed

Effect of source-substrate bias

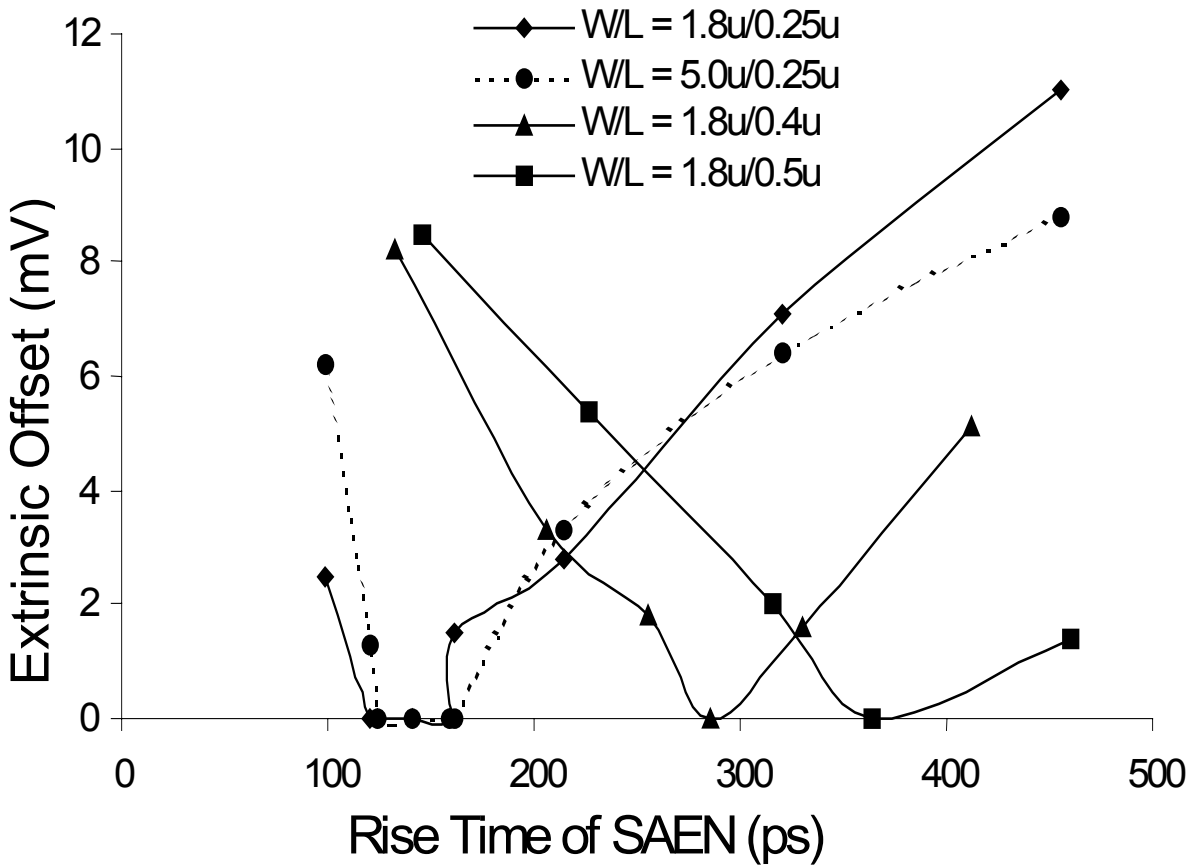


Applying body bias increases common mode Vt and hence the difference in drain current decreases

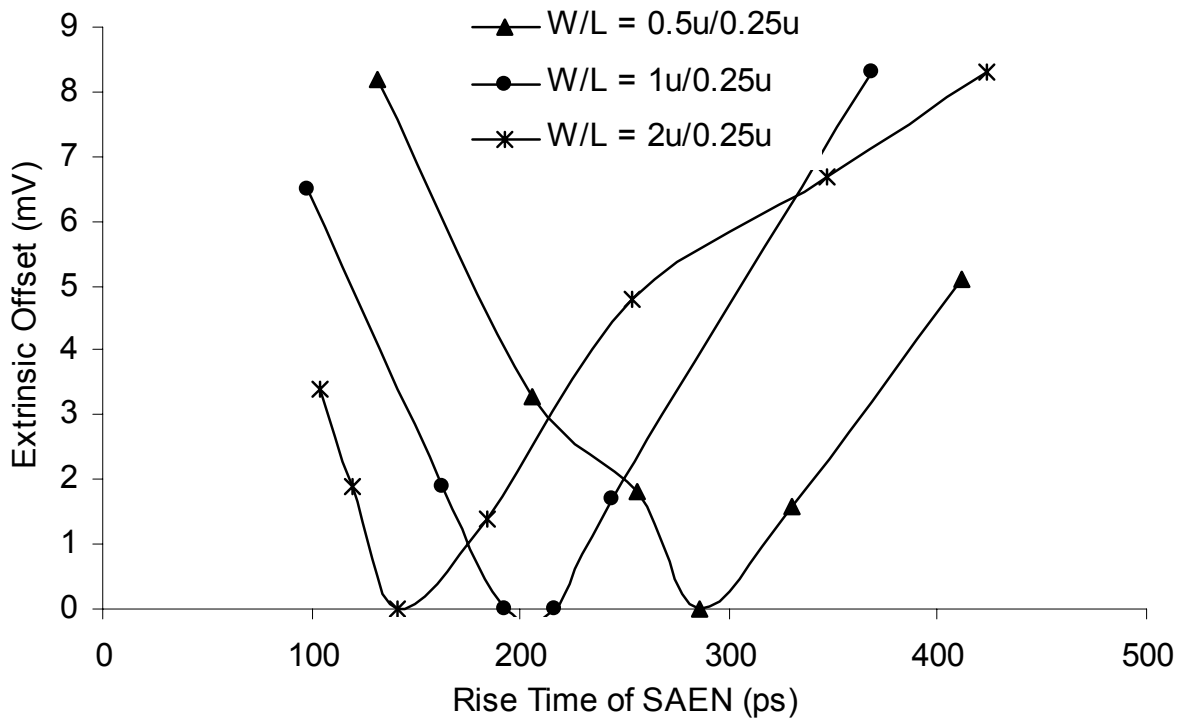
Intrinsic offset vs tail transistor size



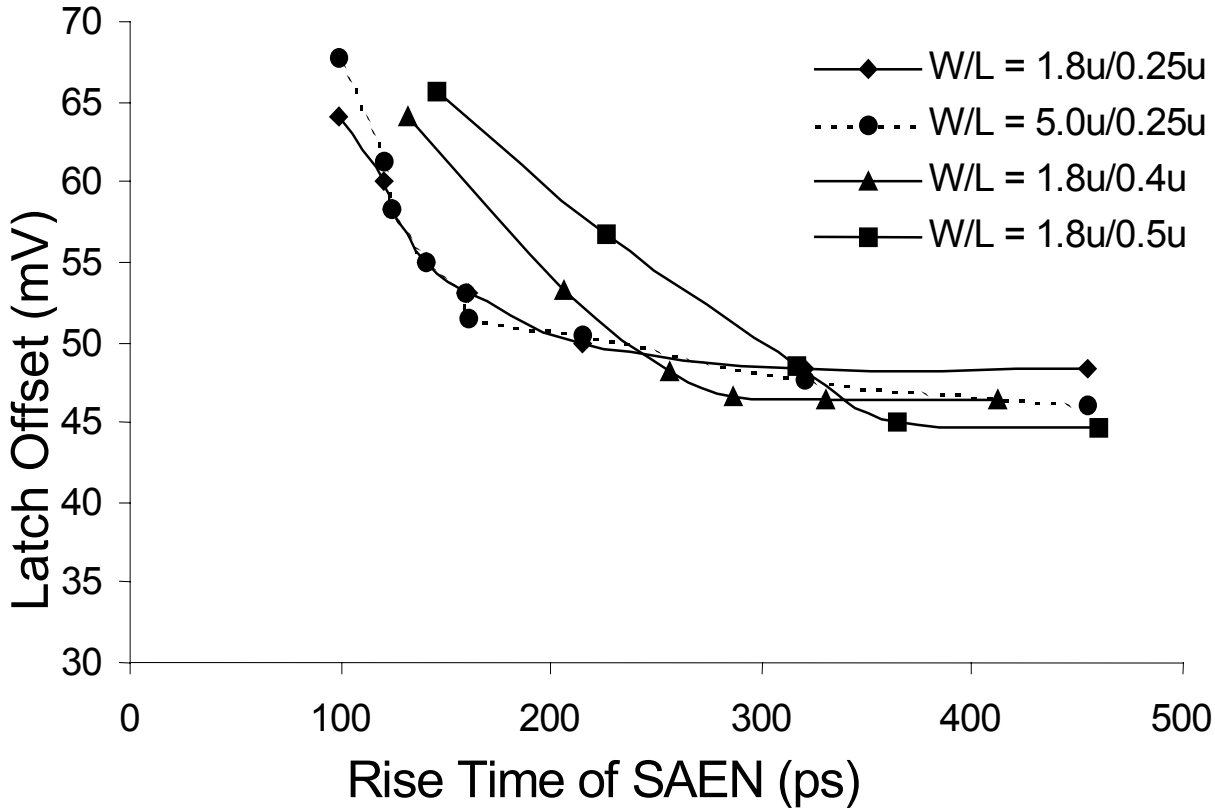
Extrinsic offset vs pass transistor size



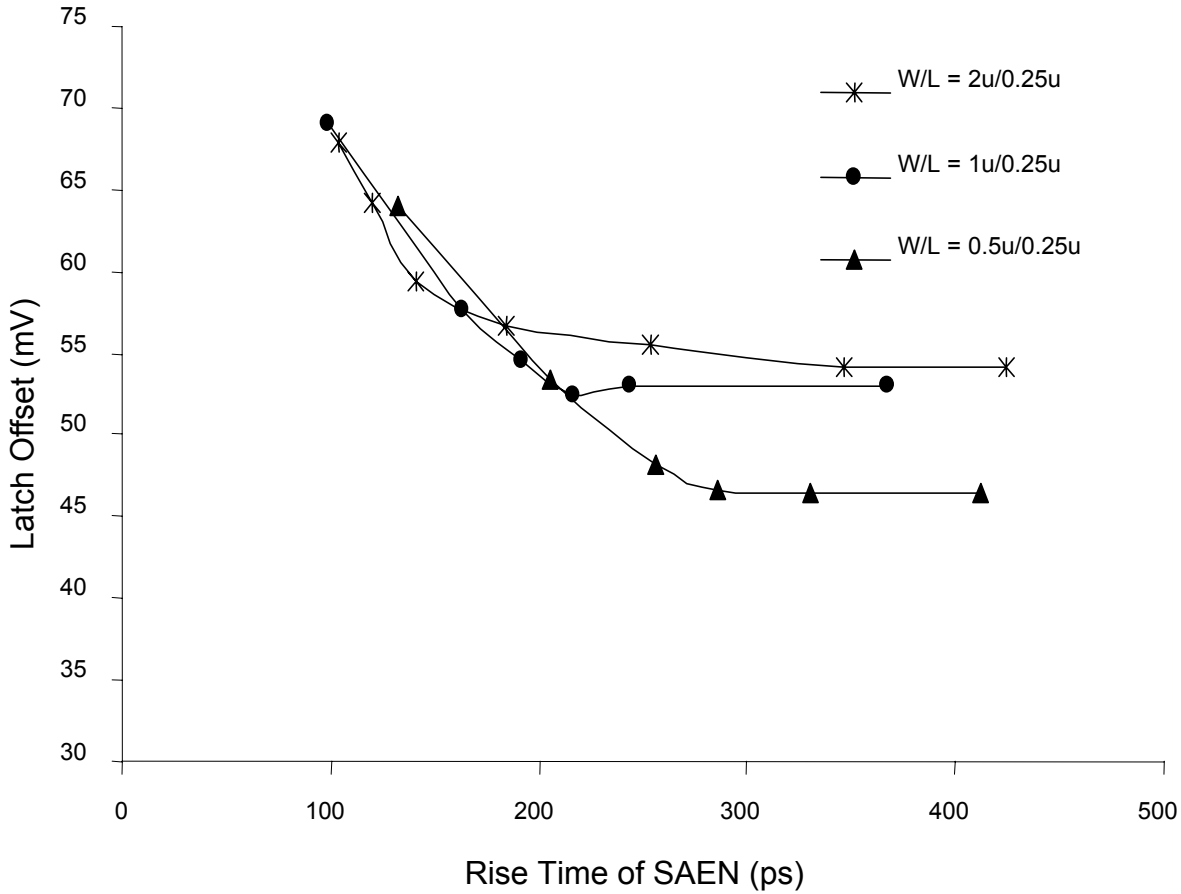
Extrinsic offset vs tail transistor size



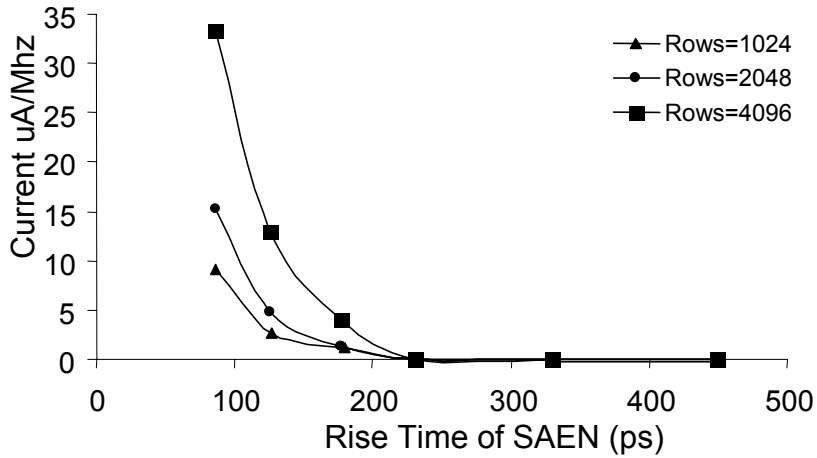
Total latch offset vs pass transistor size

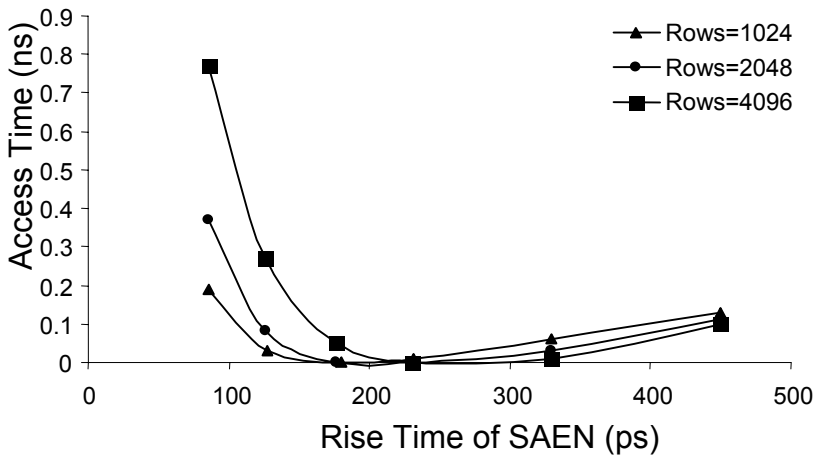


Total latch offset vs tail transistor size



Speed and Power versus SAEN Rise Time





Low power option also corresponds to high speed option!

Effect of rise time on speed and power

Rise Time (ns)	Access time (ns)		
	1024 Rows	2048 Rows	4096 Rows
86	2.66	3.44	5
126	2.5	3.15	4.5
177	2.47	3.07	4.28
230	2.48	3.07	4.23
330	2.53	3.1	4.24
450	2.6	3.18	4.33

Rise Time (ns)	Current consumption (uA/MHz)		
	1024 Rows	2048 Rows	4096 Rows
86	78.2	102	165.3
126	71.6	91.5	145
177	70.2	88	136
230	69	86.7	132
330	69	86.7	132
450	69	86.7	132

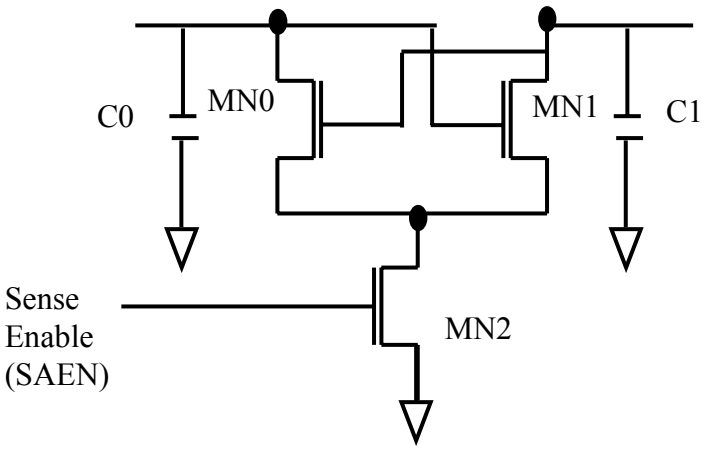
Analytical modeling

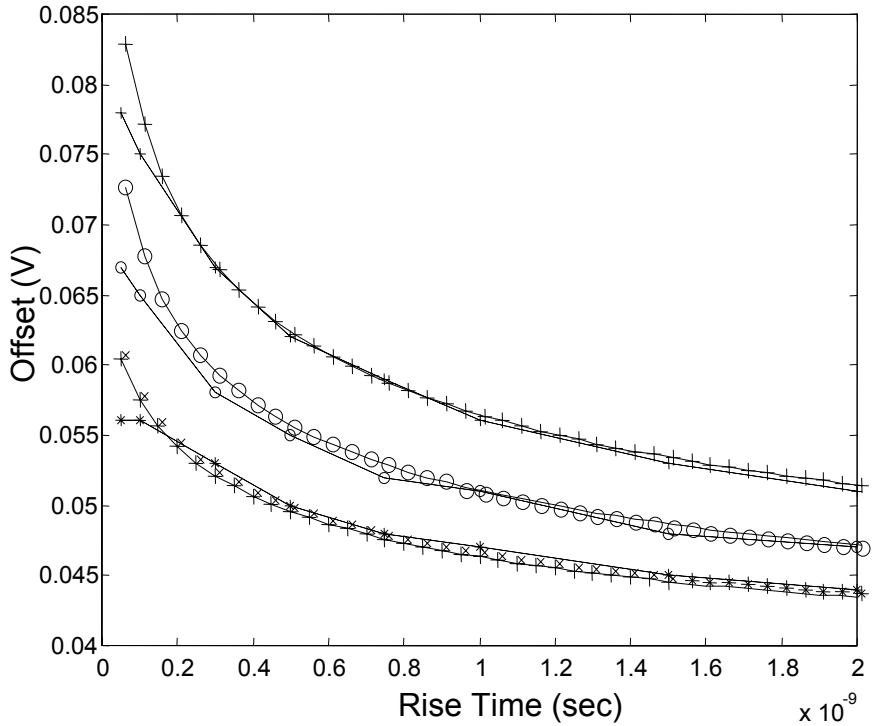
The input offset voltage is the minimum required input signal
In order to produce correct latching

$$V_{SIG} > \Delta V_T + \frac{1}{2}\left\{\left(\frac{\Delta\beta}{\beta}+\frac{\Delta C}{C}\right)+\frac{\Delta C}{C+C_{GS}}\frac{KC_{GS}}{\beta}\frac{3a_1}{a_2{}^2}\right\}(V-V_S-V_T)_{max}$$

$$+\,mdec\frac{\Delta C\,C_{GS}}{(C+C_{GS})^2}(V-V_S-V_T)_{max}$$

Analytical Modeling





Excellent matching with simulation results are obtained

Summary of transistor mismatch effects

- Random variations in IC process parameters result in mismatch among identically designed transistors
- Transistor mismatch limits performance and yield of sense amplifiers in memory application
- The mismatch effects will become worse with technology scaling
- Analog and mixed signal circuit design should be able to overcome the transistor mismatch effects

Need for Statistical Design and Simulation

Pessimistic design with Worst Case Process corners SPICE models

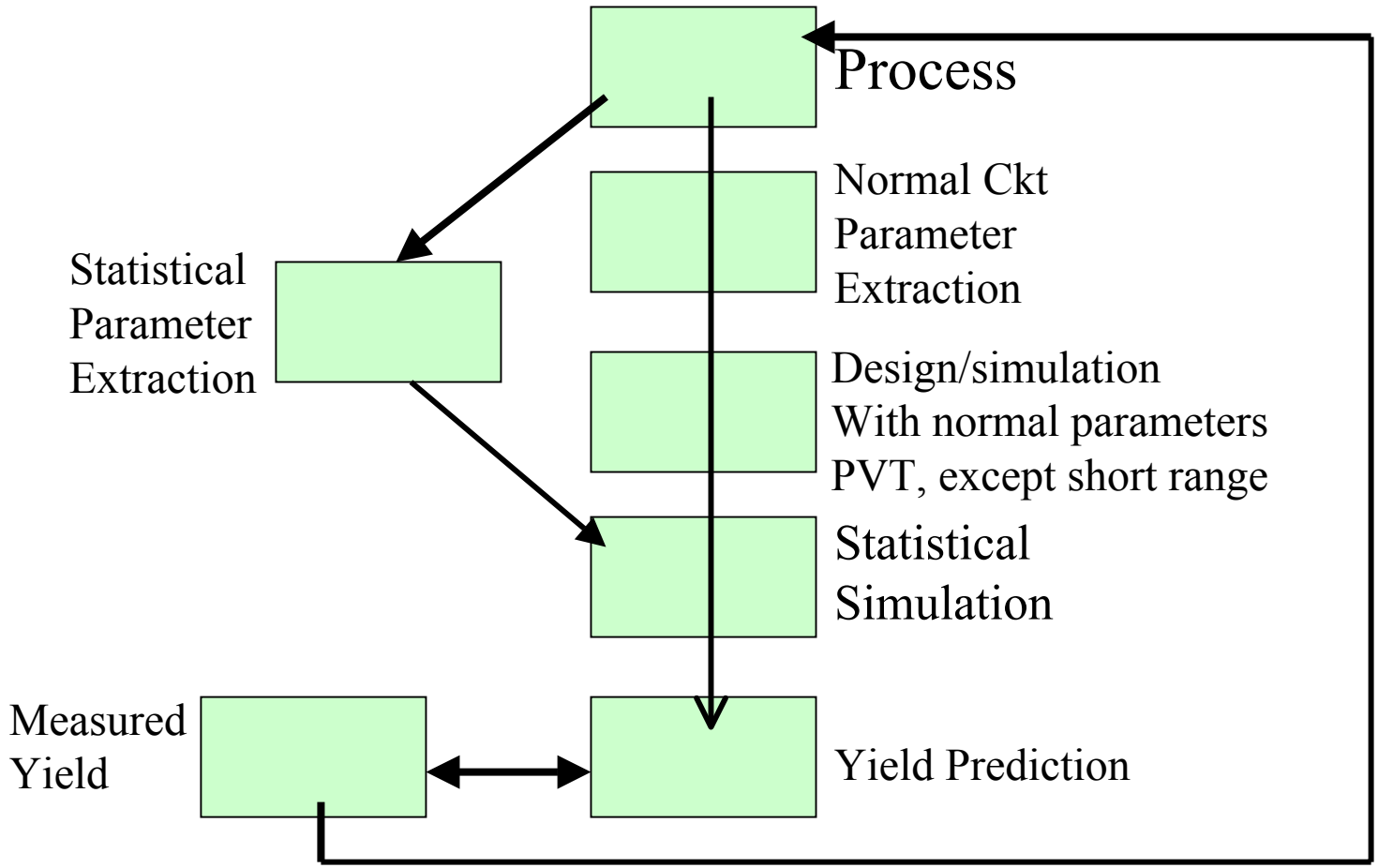
If the corner parameters are used to simulate the worst case mismatch effect, the design will be pessimistic

As a designer, in order to do a reliable as well as high performance design, obtain the matching data as well from the fab

It does take quite a bit of effort for the fab to generate mismatch data, but it would be worth the effort from designer's viewpoint

Statistical circuit simulation

Ideally the statistical design/simulation should be part of the design flow for analog circuits



Monte Carlo Technique

Statistical technique used to predict the output distribution when there is no closed form expression relating output distribution to input

Random number generation is used to randomly assign a value to input variables and walk through the input to output transformation

By transforming a large set of inputs to the output the output distribution is obtained

Response surface methodology using DOE

For a complex circuit the SPICE simulations in the Monte Carlo loop become computationally inefficient

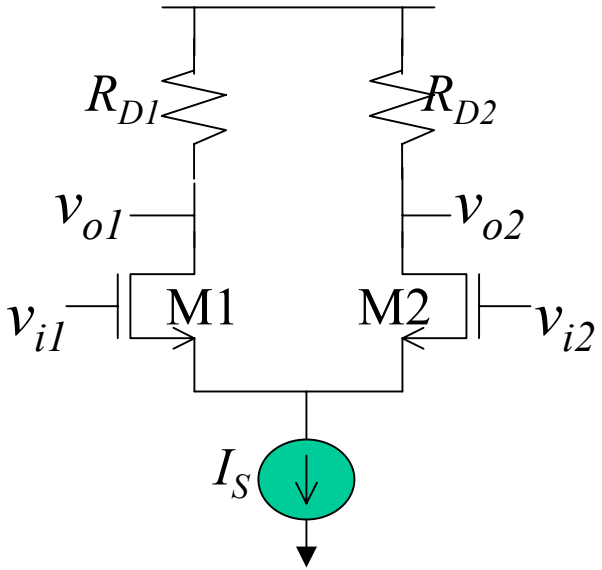
By performing very few input to output transformation, a mathematical model could be fit to relate the output quantity to the input (linear, quadratic or some other function)

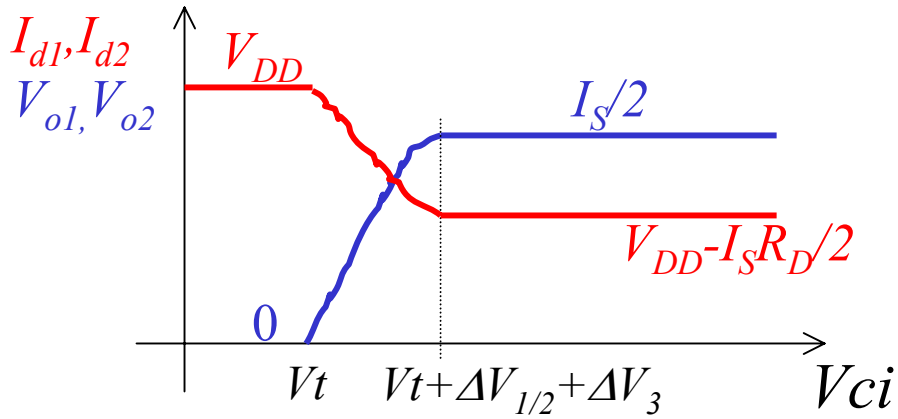
This model replaces SPICE simulations from MC loop

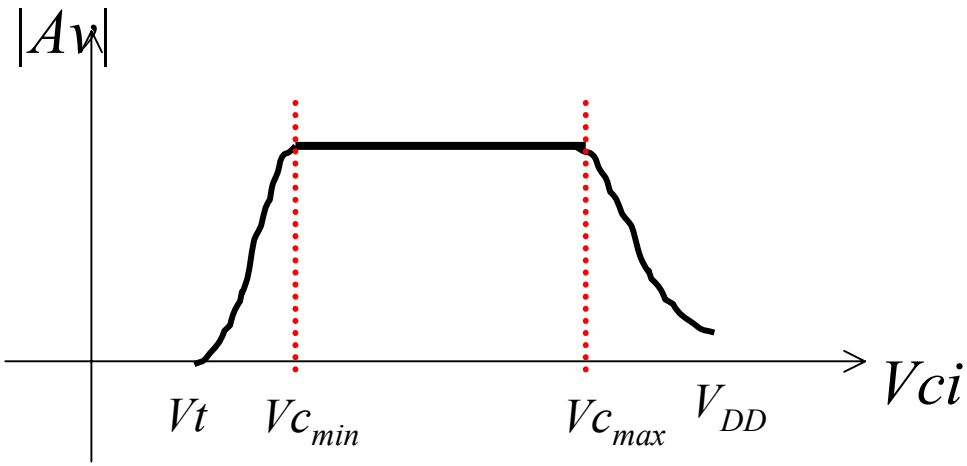
Computational efficiency is enhanced significantly

Wide Common Range OPAMP

Common mode range for differential amplifier





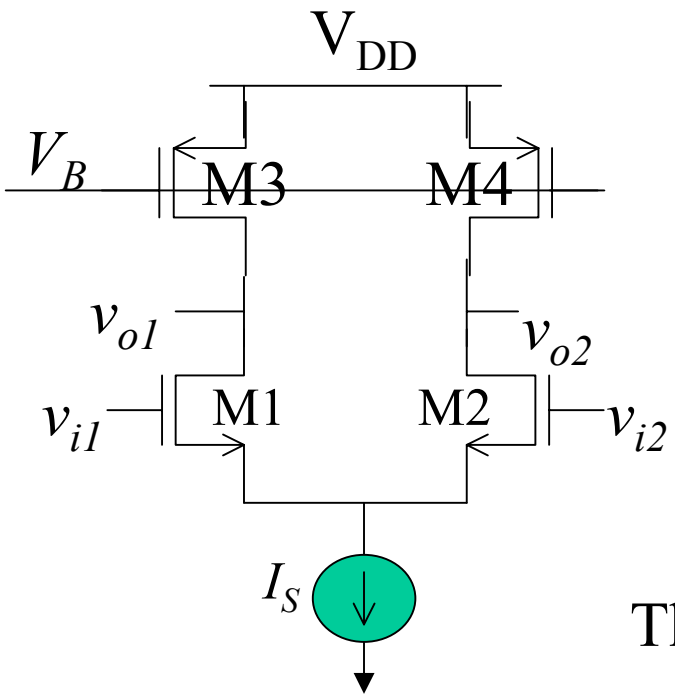


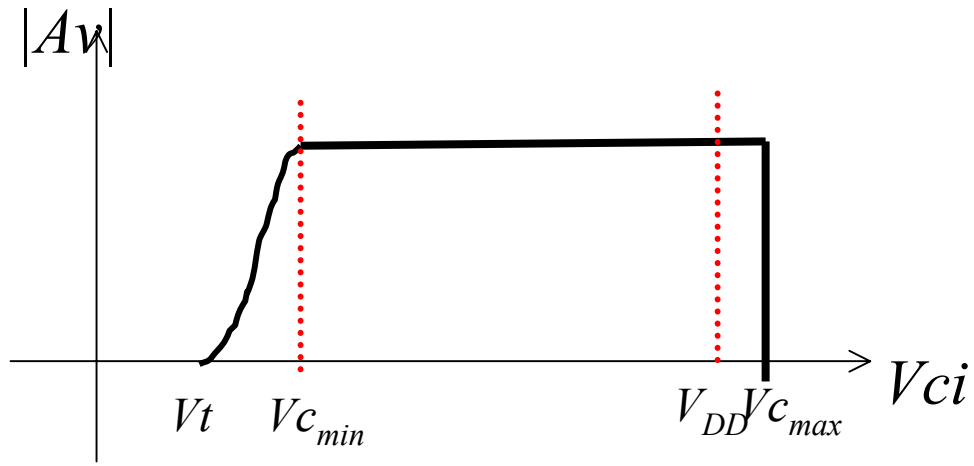
Only the lower limit of *Vci* is the hard limit and is above negative rail (0V)

We presumed the upper limit also to be much below positive rail (VDD) due to *IsRd/2* drop

What if you replace the load RD with current source load?

The common mode range with current source load

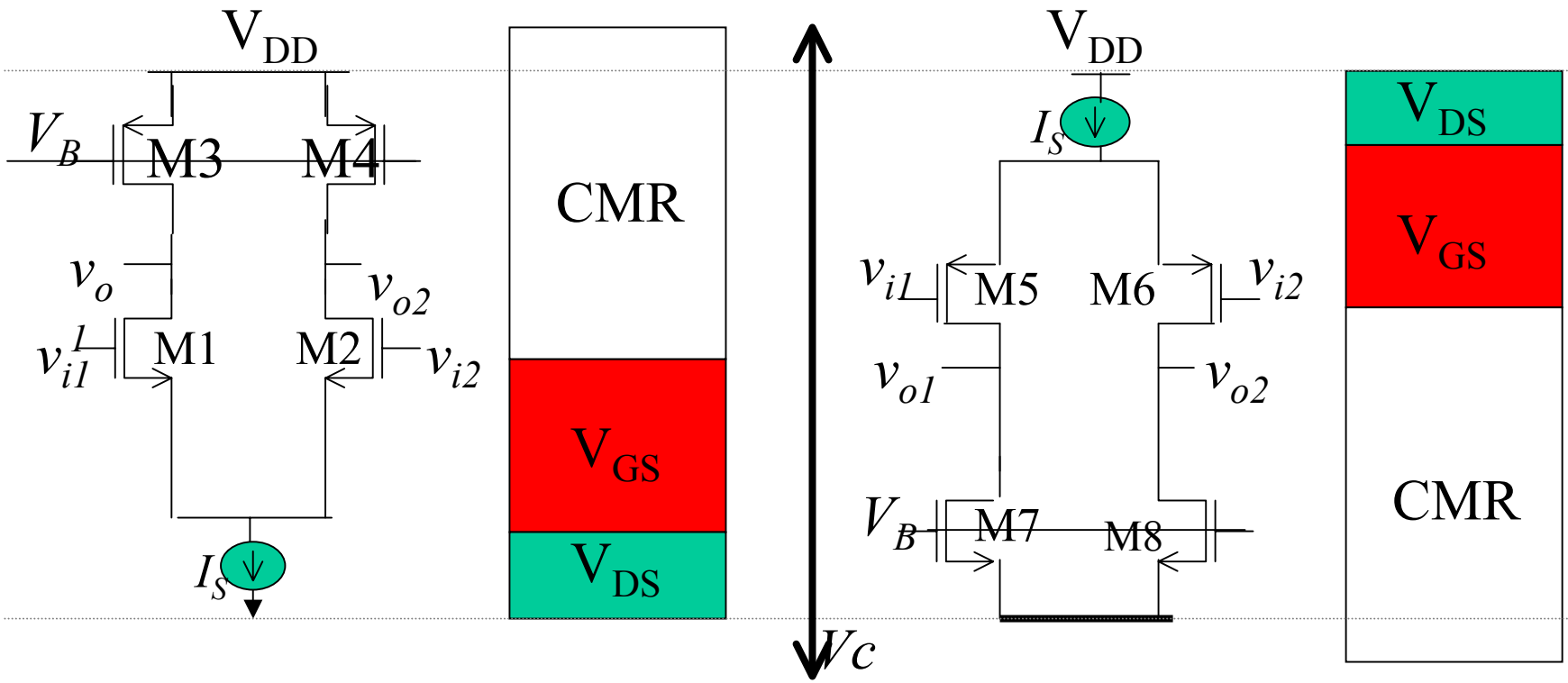




The V_{cimax} will be even greater VDD!

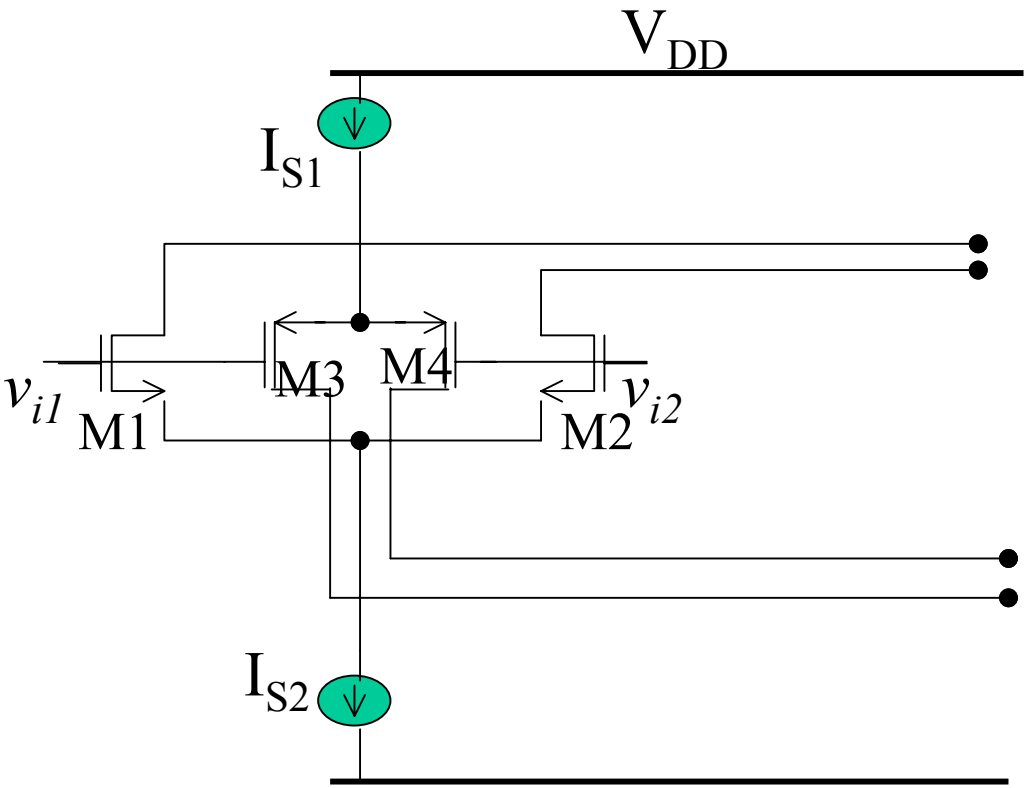
However the V_{cimin} is the hard limit

Common Mode Range for NMOS and PMOS



Combine NMOS and PMOS input differential pair to obtain rail to rail common mode range, *Vci* !
(In fact the CMR Could be even beyond rail to rail)

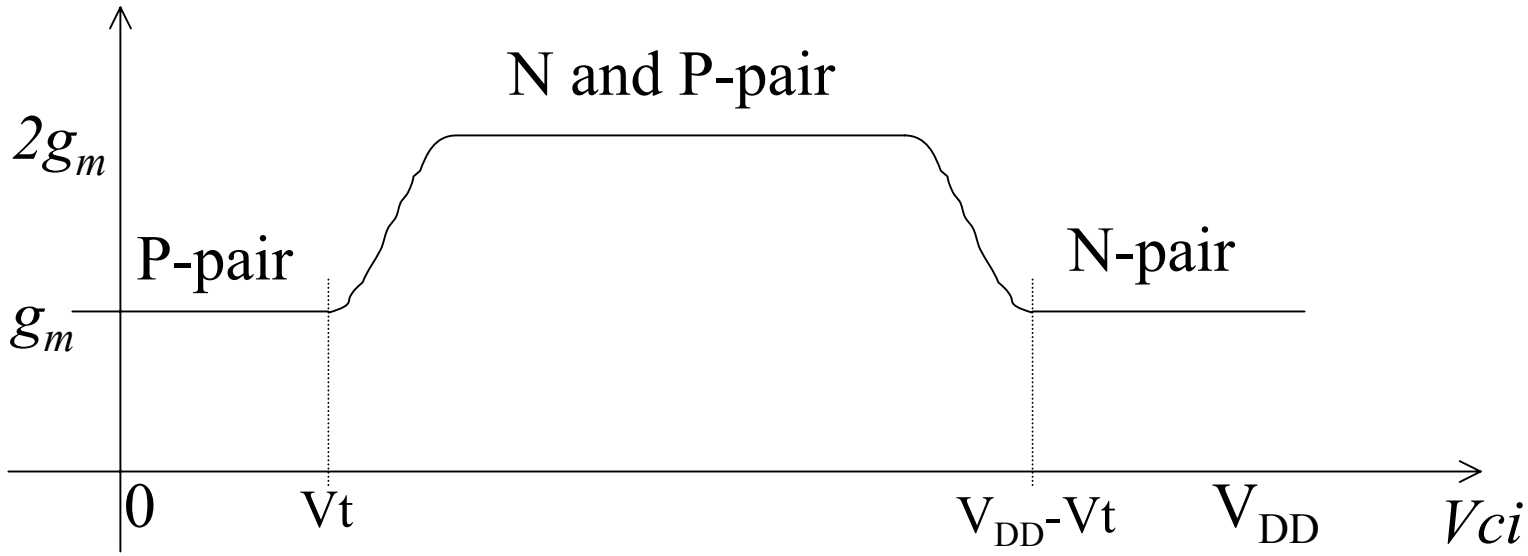
The rail to rail input stage



} To summing circuit using current mirror

Low common mode input : Only P-type diff pair operates
Intermediate common mode input: Both N and P-type operate
High common mode input : Only N-type diff pair operates

The variation of the first stage gain



For the simple circuit, the first stage gain is not constant over the common mode input range

Additional circuit is required to maintain the constant gain

The concept of gain control

Suppose that N and P pair bias current is constant and N and P transistors are sized to match the transconductance

$$\sqrt{k_n I_S} = \sqrt{k_p I_S} = \sqrt{k I_S}$$

When both pairs are active the transconductance gets added

$$G_m = 2\sqrt{k I_S}$$

If we like to maintain the same transconductance when either N or P differential pair is switched off, then we need to change the active pair bias current using some circuitry

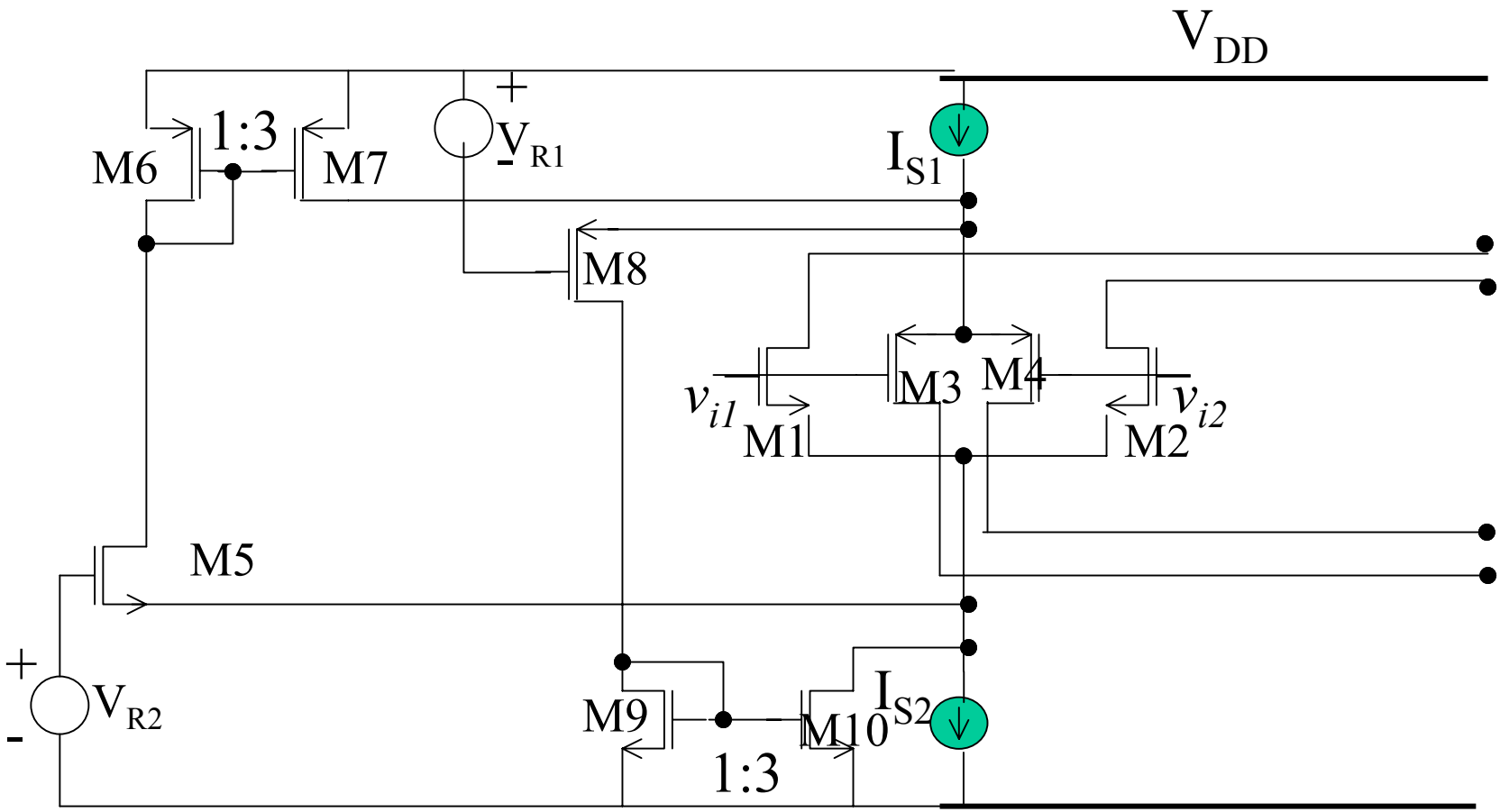
$$G_m = \sqrt{k_n 4 I_S} = 2\sqrt{k I_S}$$

Increase N bias current by a factor of 4 when only N is active

$$G_m = \sqrt{k_p 4 I_S} = 2\sqrt{k I_S}$$

Increase P bias current by a factor of 4 when only P is active

The first stage with gain control



V_{R1} and V_{R2} are chosen slightly above Vt of M5 and M8

The current switching

For low V_{ci} i.e $V_{ci} < V_t$

M8 is OFF and M5 is ON $\Rightarrow I_{S2}$ flows through M5
The current mirror M6-M7 multiplies I_{S2} by a factor of 3 onto the drain of M7. This is added to I_{S1}, thereby increasing The P bias current by a factor of 4

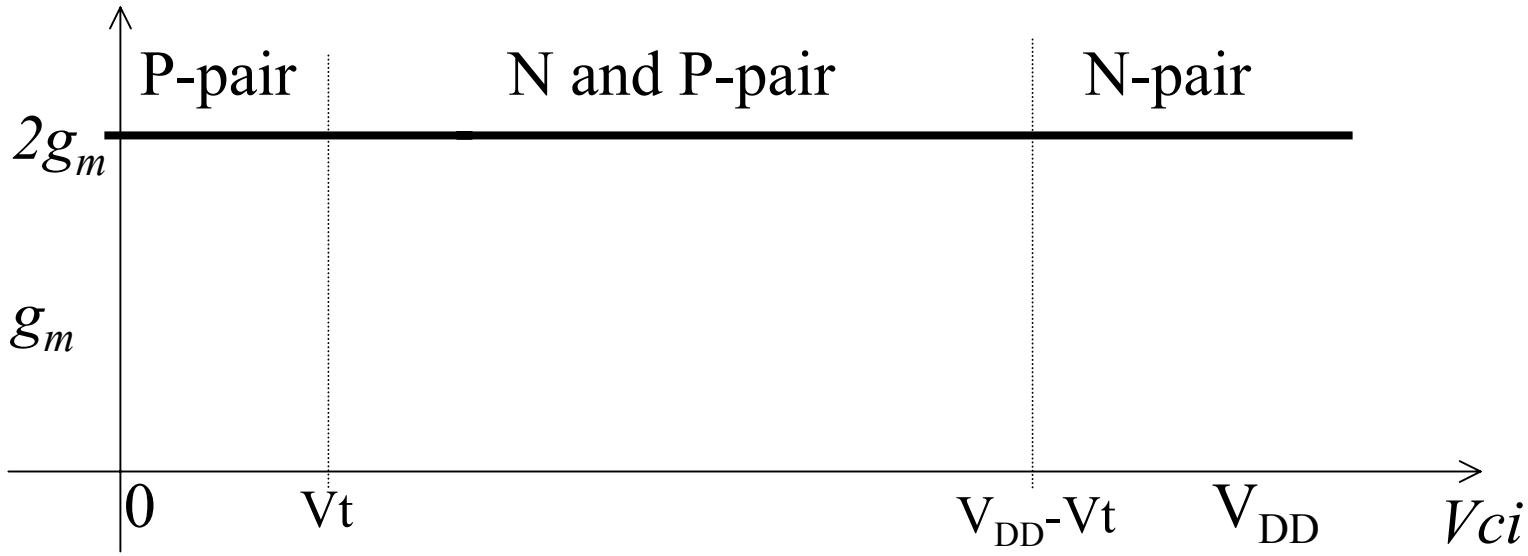
For high V_{ci} i.e $V_{ci} > V_{DD} - V_t$

M5 is OFF and M8 is ON $\Rightarrow I_{S1}$ flows through M8
The current mirror M9-M10 multiplies I_{S1} by a factor of 3 onto the drain of M10. This is added to I_{S1}, thereby increasing the N bias current by a factor of 4

For intermediate V_{ci} i.e $V_t < V_{ci} < V_{DD} - V_t$

M5 is OFF and M8 is OFF $\Rightarrow$ Current mirrors are disabled
N bias current = P bias current = Is

The corrected gain

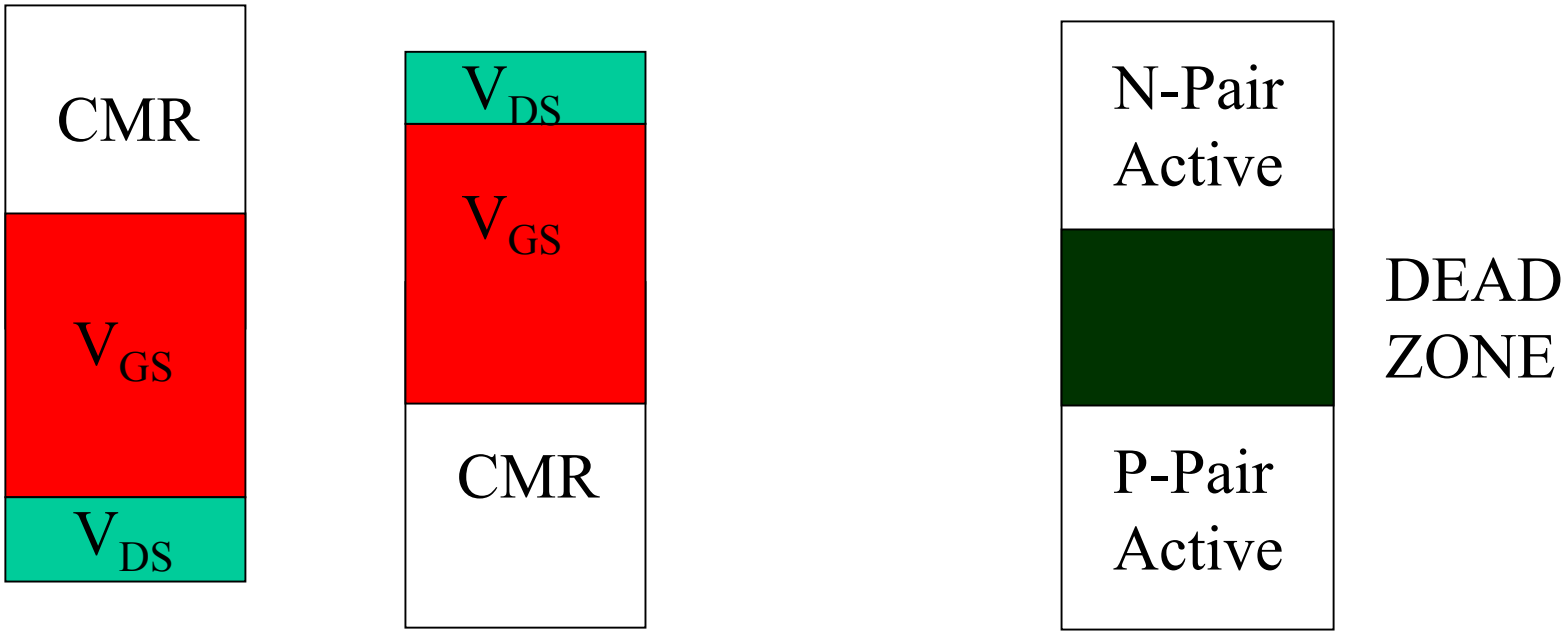


Adding current switching makes the differential gain almost flat over the common mode range

Bulk Driven OPAMP

The dead zone problem in complementary input stage

Suppose that supply voltage is very small $V_{DD} < V_{GSn}+V_{GSp}+2\Delta V$



Then there is a dead zone in Vc range!

Both the pairs are deactivated and no useful operation

Overcoming the Vt problem in single stage

N and P stages are inherently limited by the fact that the gate voltage should be more than *Vt* to turn on the transistor

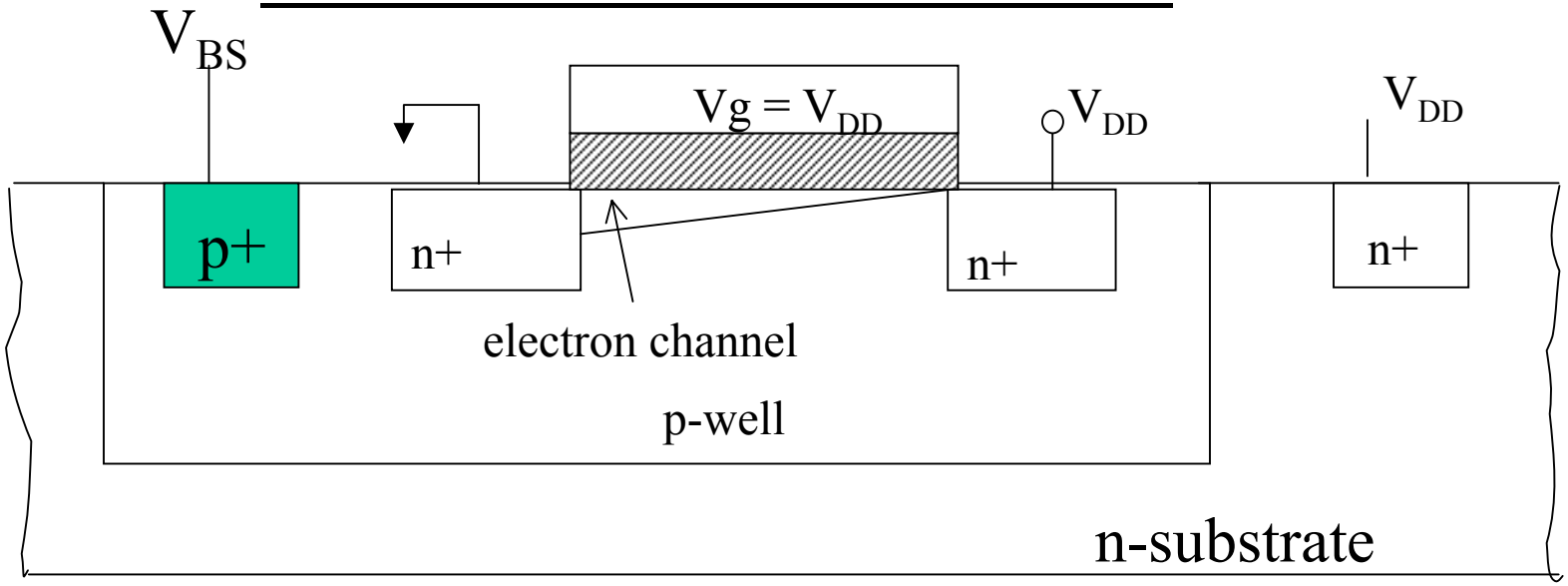
Possible remedy:

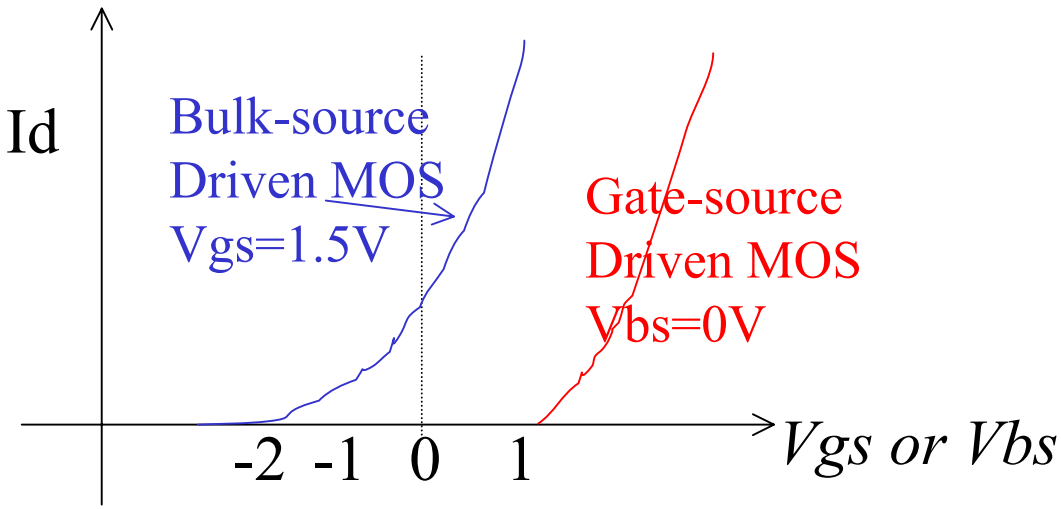
Fix the gate voltage above Vt and apply the inputs to the body of the transistor (bulk driven)

Modulating the body voltage results in change in Vt and hence gives rise to body effect transconductance Which is used to operate on the input signals

Note: At the higher range of input common mode voltage significant junction leakage may result

The bulk driven NMOS



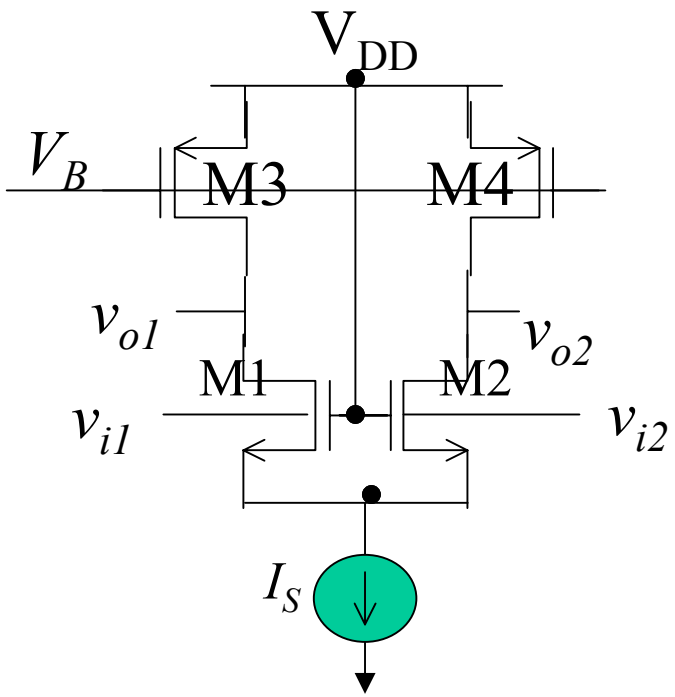


Bulk driven MOSFET can be viewed like a JFET

Parasitic BJTs are not turned on

The transistor is ON even when the input V_{BS}=0

Bulk driven NMOS first stage



Rail to rail *Vci* is possible

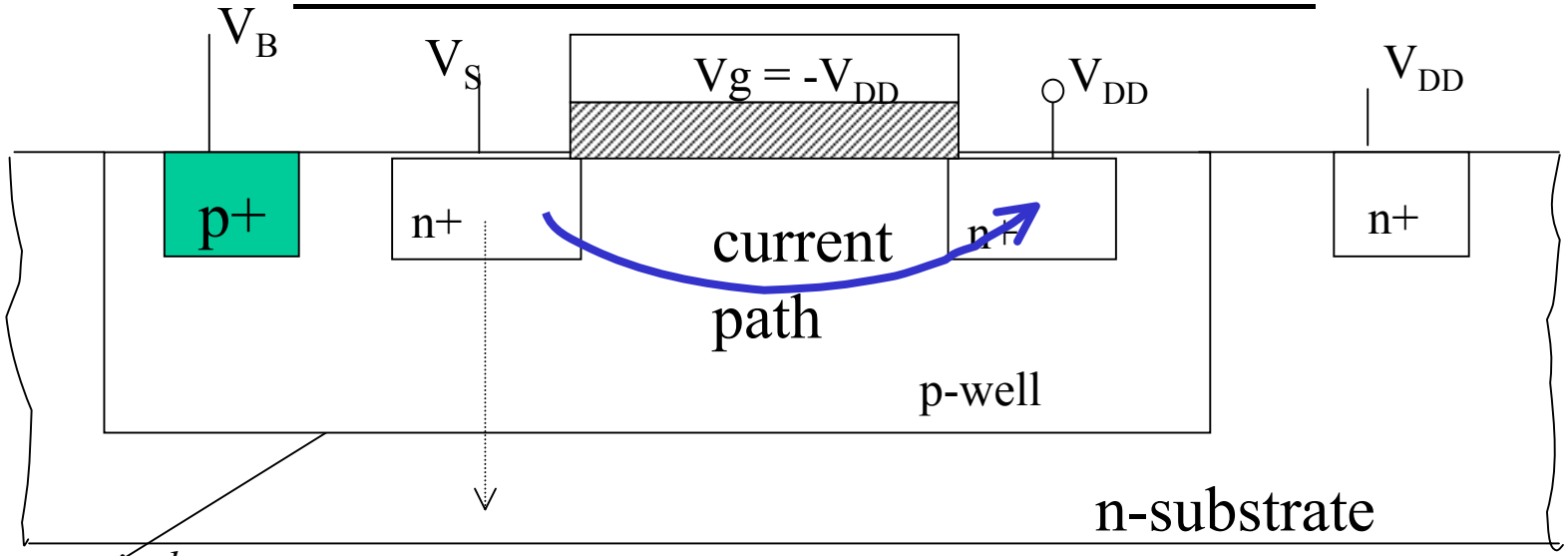
Gm is not constant over the *Vci* range

If the supply voltage is less than 1V then the maximum forward bias of junction would be about 0.5V

Note that body effect is active even when body to source voltage is forward biasing the junction

Lateral BJT in CMOS Technology

The Lateral BJT in CMOS



Suppress vertical component through layout technique

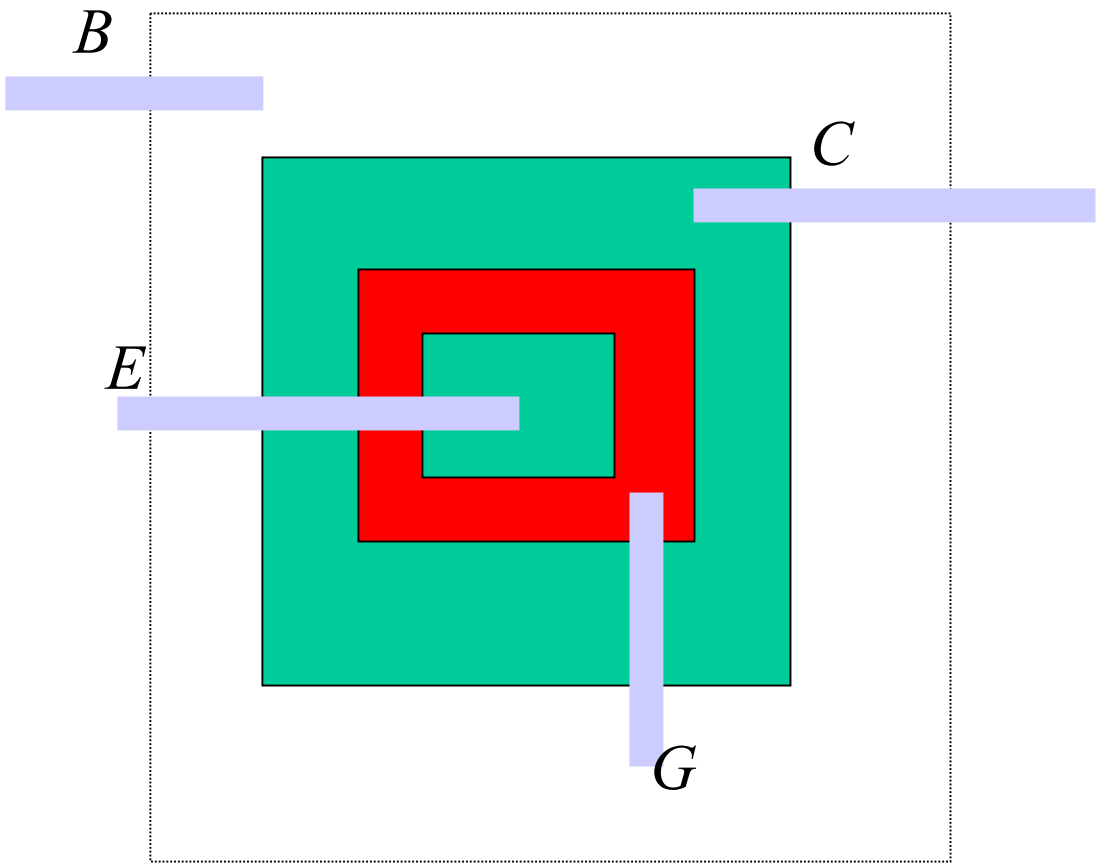
Use the lateral BJT between source and drain

Intentionally forward bias well-source (base-emitter) junction

Gate voltage is negative to turn off the MOSFET

For small channel lengths, BJT has reasonable gain

Surround drain layout



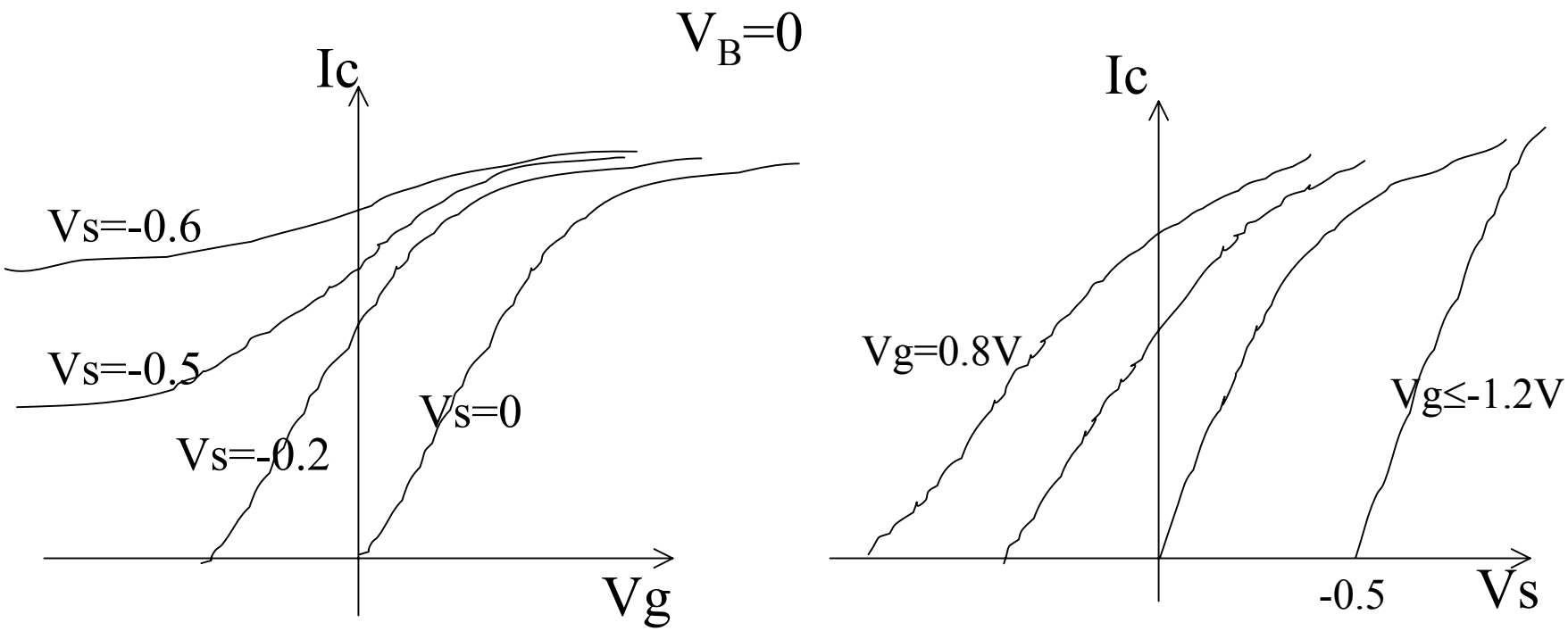
The edgeless gate layout Improves the lateral BJT performance over the vertical one

The lateral current collection has increased

Does not require any modifications to CMOS technology

1/f noise improves

The device characteristics



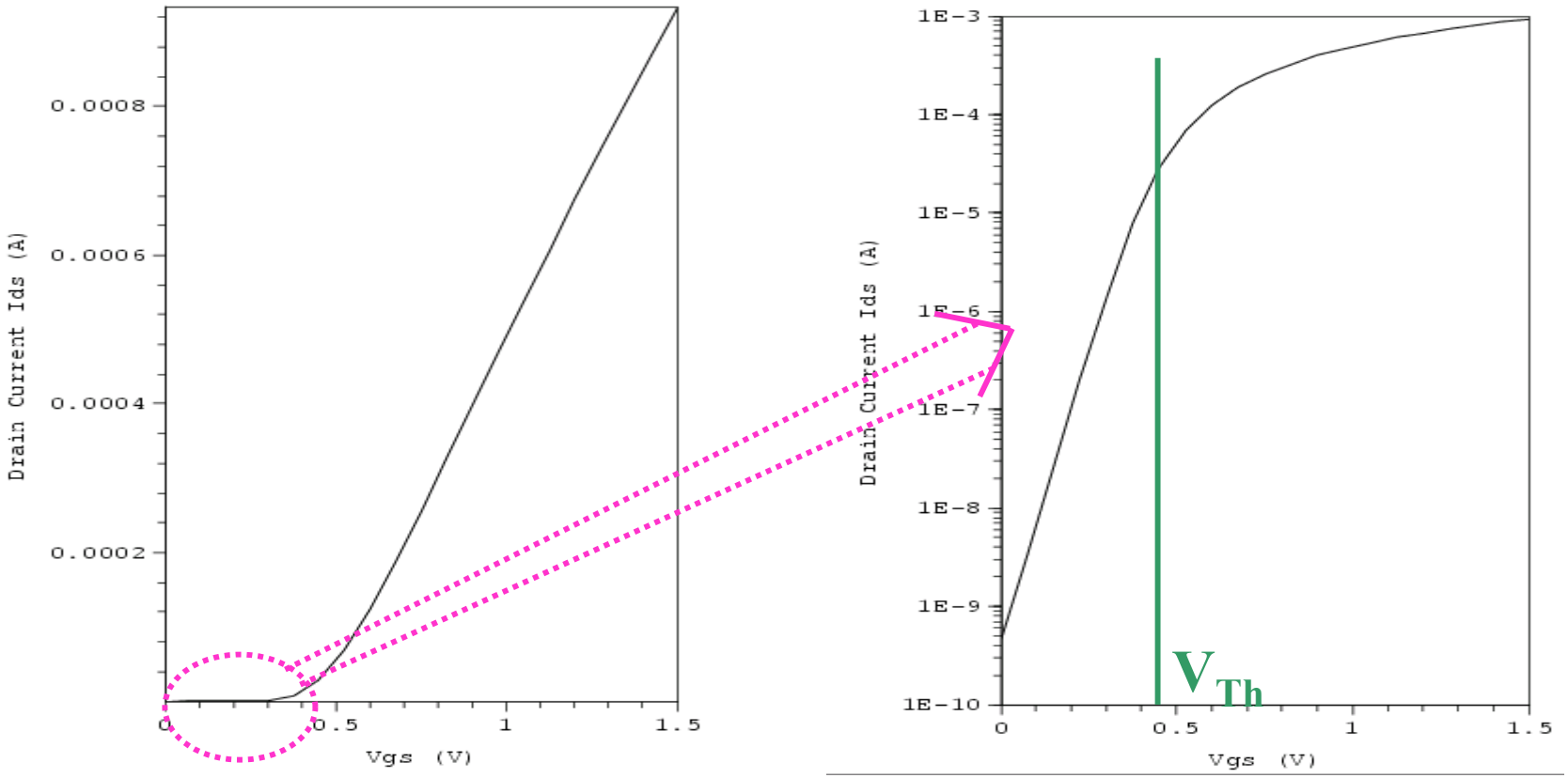
When the gate voltage is negative,
the *Ic-Vs* characteristics look like ideal BJT

This device can also overcome the Vt related problem

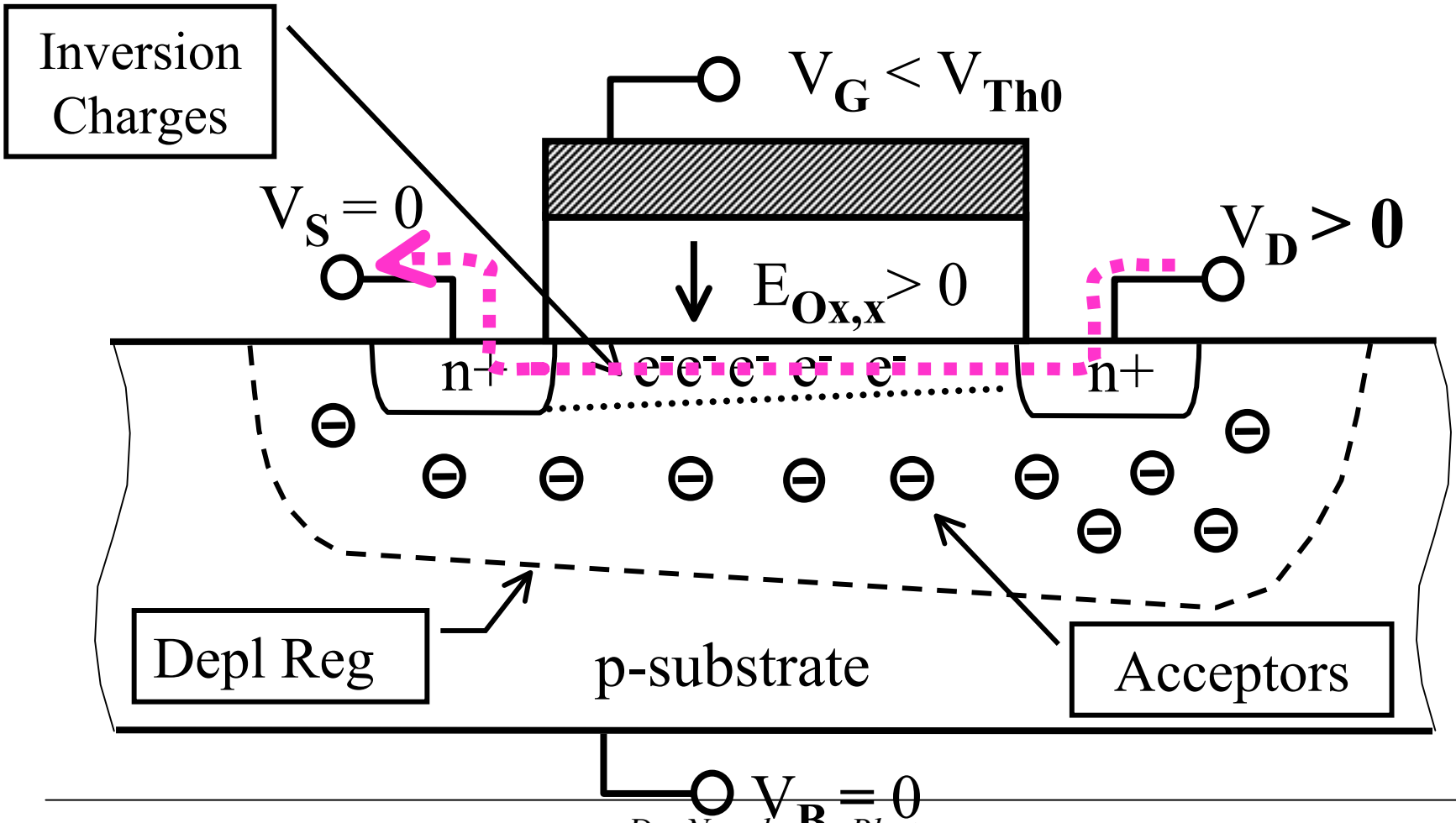
Subthreshold Operation : Neural Networks

Acknowledge the contributions made by Amit Gupta

Subthreshold Operation



Subthreshold Operation



Subthreshold Operation

$$I_{DS} = I_0 e^{q\kappa V_{GS}/kT}\left(1 - e^{-qV_{DS}/kT}\right)$$

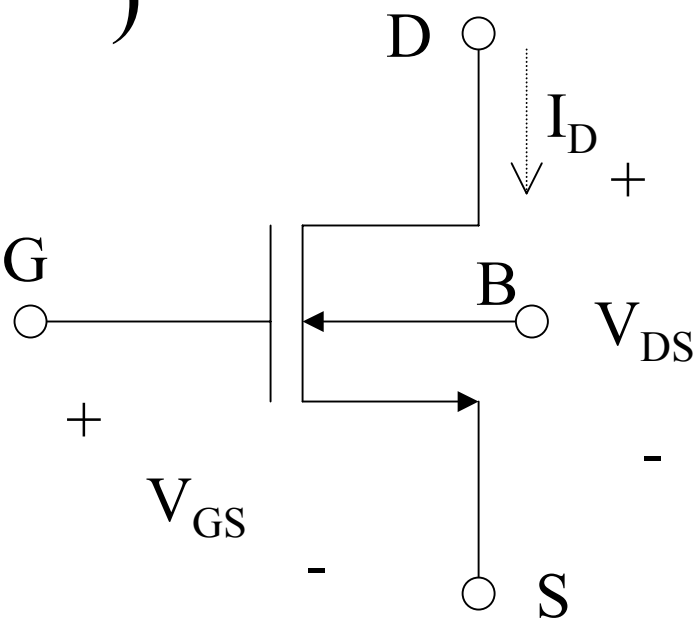
Where

$$\kappa = \left(1 + C_D / C_{OX}\right)^{-1}$$

$$I_0 = k_x \frac{W}{L} e^{q[(1-\kappa)V_{BS} - V_{Th0}]/kT}$$

For $V_{DS} > 50mV$

$$I_{DS} = I_0 e^{q\kappa V_{GS}/kT}$$



$V_{GS} < V_{Th0}$

Subthreshold Operation

- Exponential non-linearity
- Extremely low power dissipation
- Highest processing rate per unit power
- Saturation of drain current in few kT/q

Vt limitation is not present to build rail to rail OPAMPs

Motivation for Neuromorphic circuits

- Sequential Processors Vs Neural Networks
- Hardware Implementation
- Analog VLSI
- Low Power Networks

Motivation

- Modern Computer
  - Pre Programmability
  - Repetitive Computation
- Human Brain
  - Speech Recognition
  - Pattern Recognition

Human Brain

- A powerful information processor
- Massively parallel complex network of neurons
- Neuron
  - weak computation unit
  - 7-8 orders of magnitude slower than current Si gates
- Knowledge acquired through learning
- Synaptic weights are used to store the acquired knowledge

NEURAL NETWORKS:
Parallel Distributed Processors

Machines inspired from the brain's performance!!!

Analog VLSI

- Analog systems carry more information per wire and fewer transistors per operation
- Analog computing primitives, multiplication and addition, are much smaller
- Redundant h/w to ensure fault tolerance

- Analog's cost is low

- Real world interfacing: Analog systems eliminates the need of ADC/DAC.

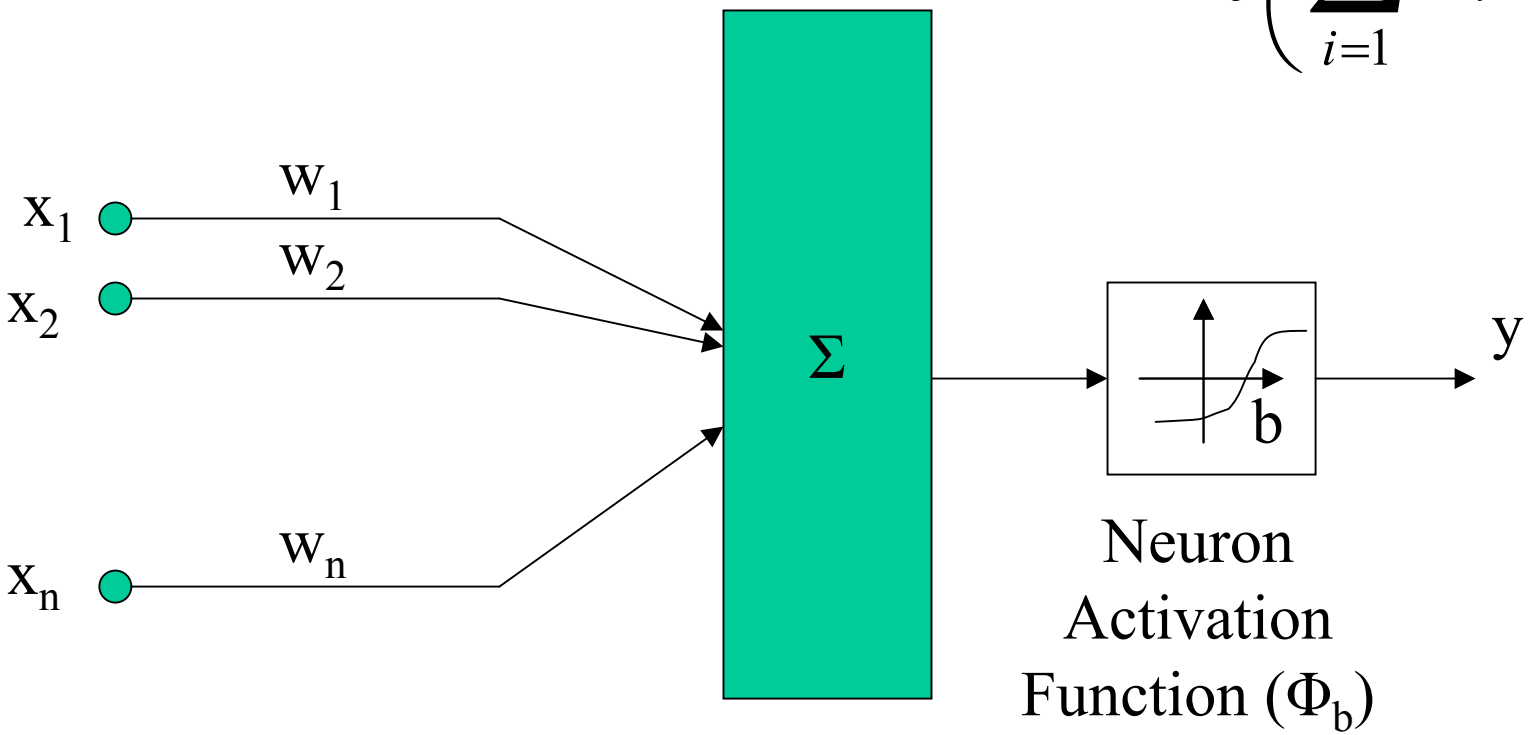
Low Power Networks

- Battery driven portable systems
- High circuit density

SUBTHRESHOLD OPERATION
IS THE NATURAL CHOICE

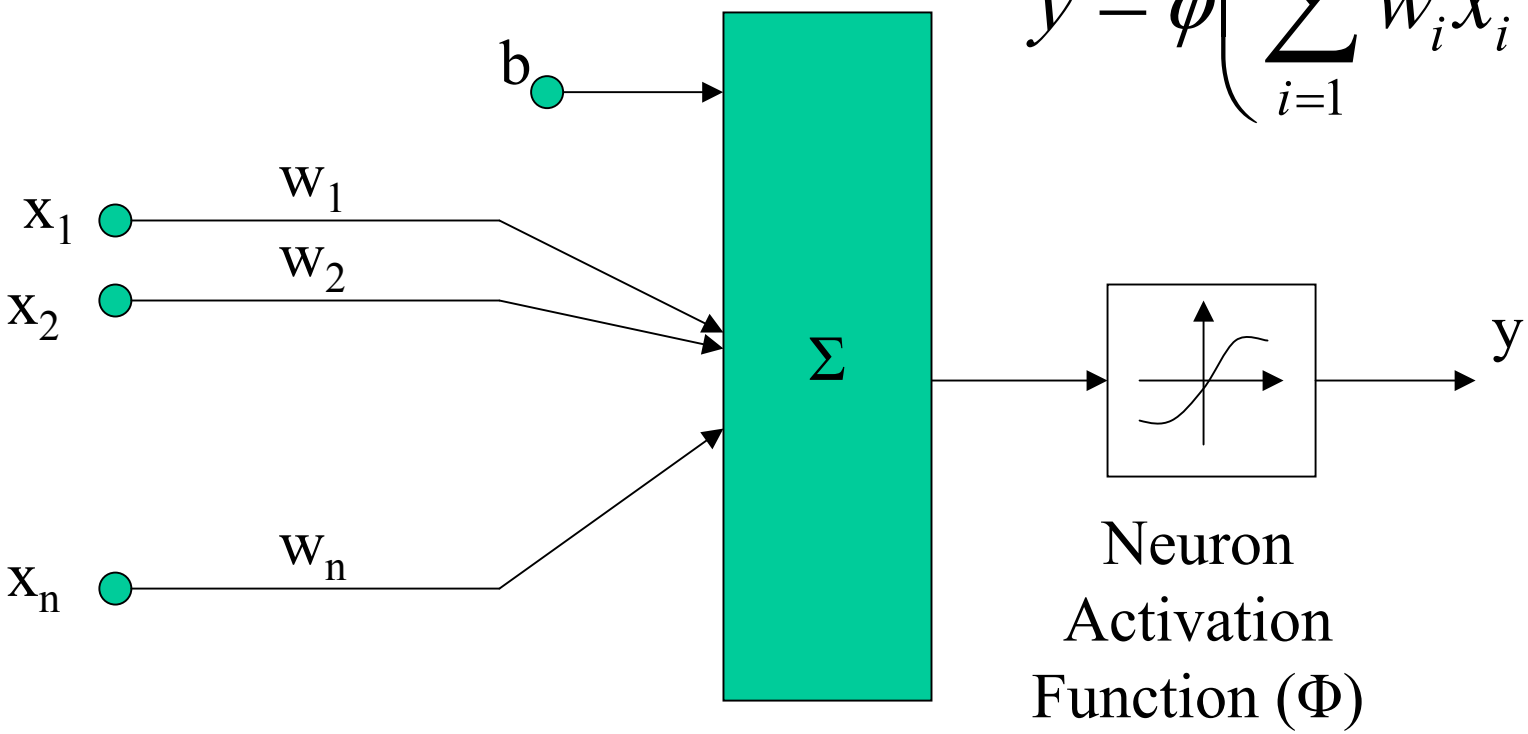
Model of a Neuron

$$y = \phi_b\left(\sum_{i=1}^{n} w_i x_i\right)$$



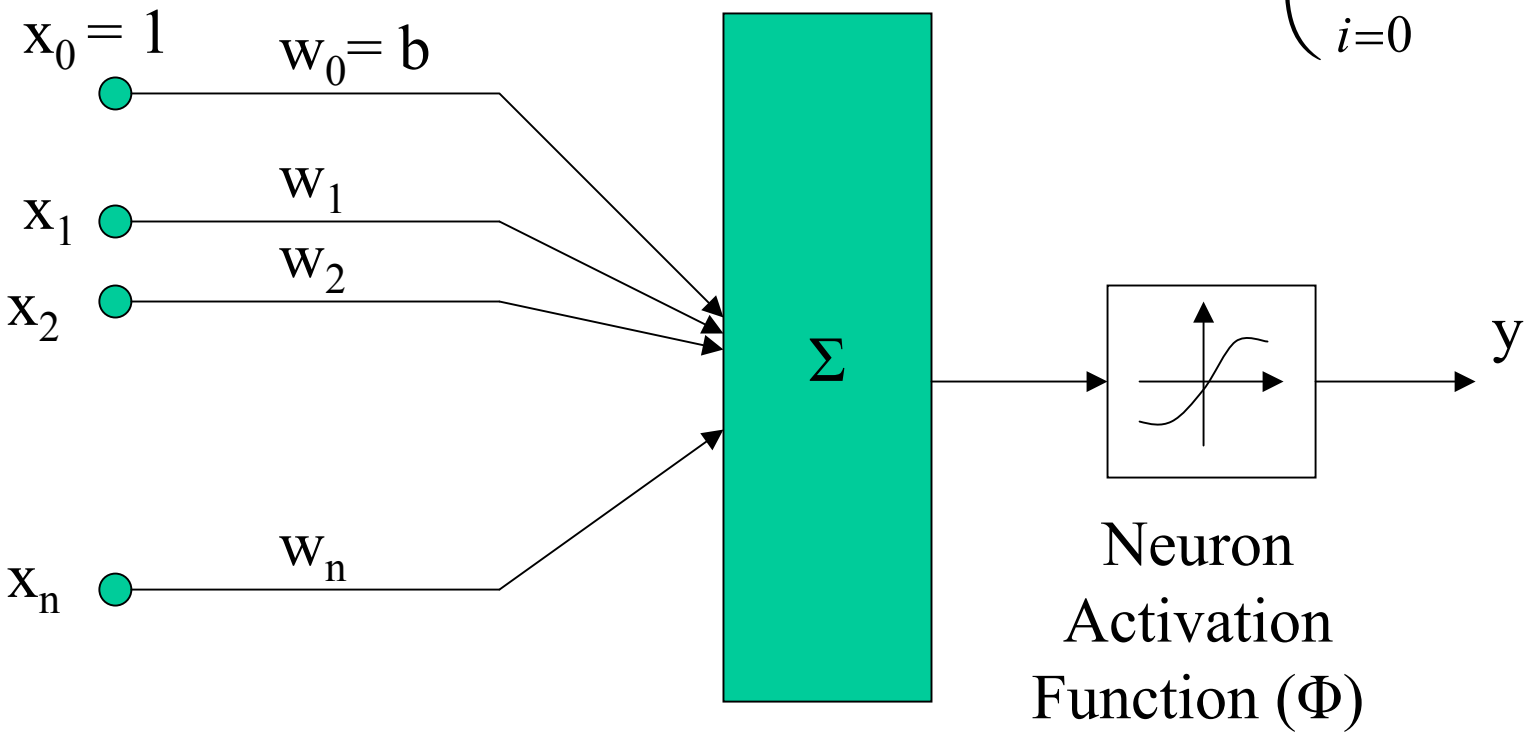
Model of a Neuron

$$y = \phi\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

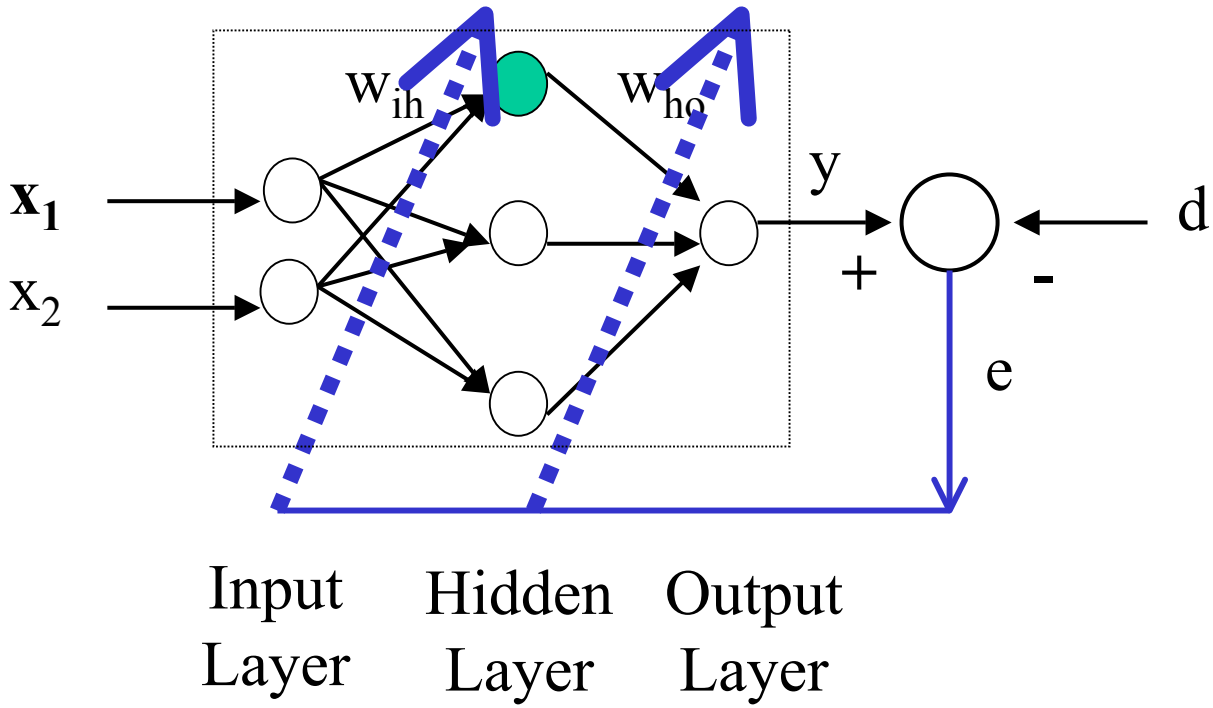


Model of a Neuron

$$y = \phi\left(\sum_{i=0}^{n} w_i x_i\right)$$



Feedforward Neural Networks



- Back Propagation Algorithms: Highly popular for training
- Perturbation Algorithms: Hardware friendly

Neuron Activation Function (NAF)

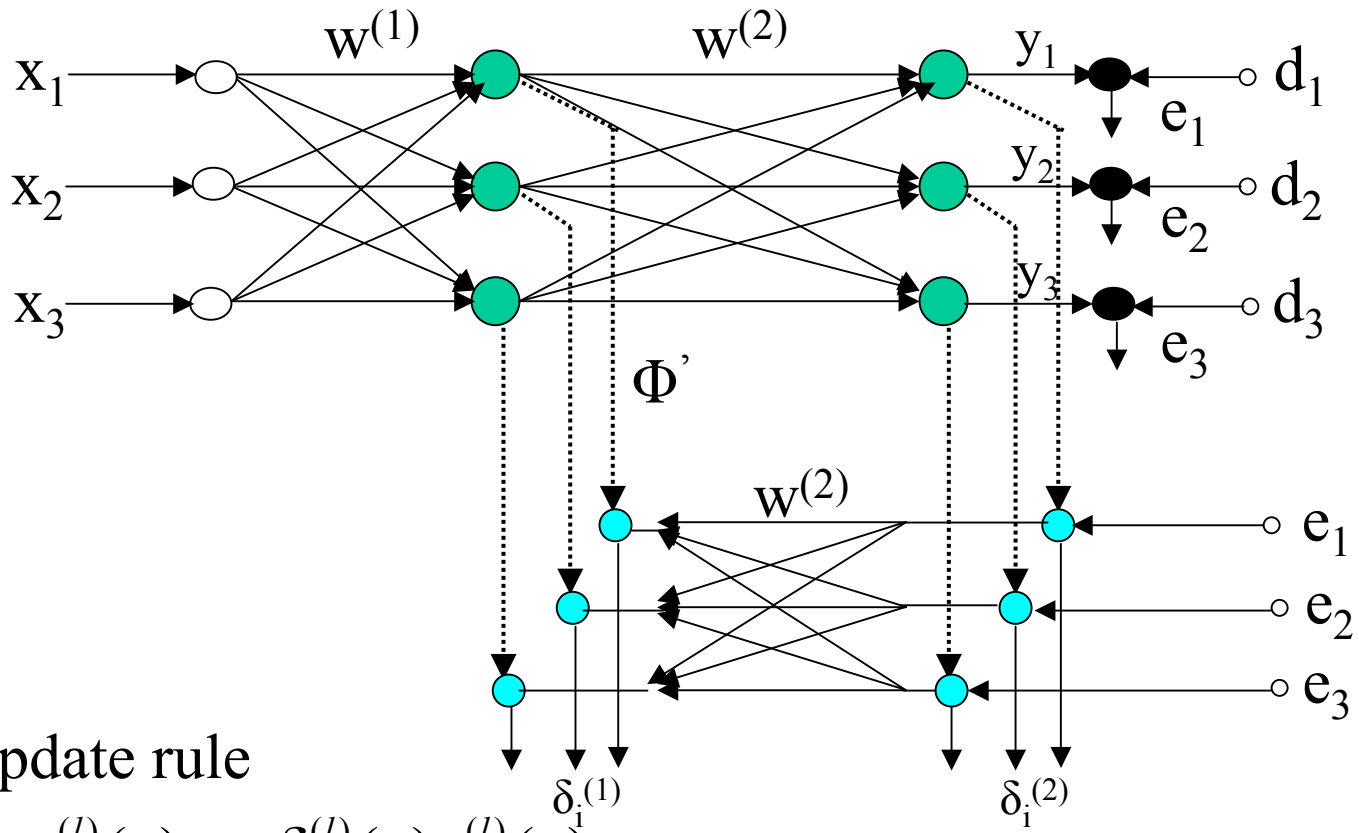
- Common choices
  - Logsigmoid

$$y_j = \frac{1}{1+e^{-\lambda s_j}} \qquad , y_j \in \{0,1\}$$

  - Tansigmoid

$$y_j = \frac{e^{\lambda s_j} - e^{-\lambda s_j}}{e^{\lambda s_j} + e^{-\lambda s_j}} \qquad , y_j \in \{-1,1\}$$

BP Algorithm



- Weight update rule

$$w_{ij}^{(l)}(n+1) = w_{ij}^{(l)}(n) + \eta \delta_i^{(l)}(n) z_j^{(l)}(n)$$

BP Algorithm

- Derivative computation

$$y' = 1 - y^2$$ For tansigmoidal transfer fn

$$y' = y(1-y)$$ For logsigmoidal transfer fn

•Implementation needs multiplier/squarer and adder/subtractor

$$y' \approx \frac{y(x+\delta x) - y(x)}{\delta x}$$ Forward difference approx

$$y' \approx \frac{y(x+\delta x/2) - y(x-\delta x/2)}{\delta x}$$ Central difference approx

•Implementation needs switched capacitor to introduce small voltage.

•Asymmetry eliminates need of switched capacitor and supporting circuitry

Parallel Perturbative Algorithm

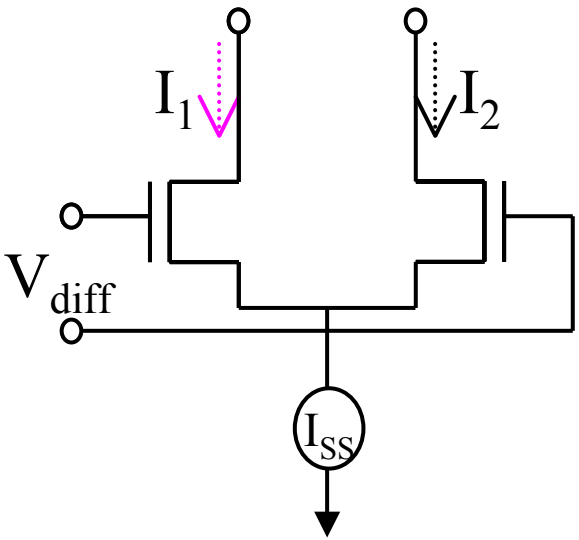
- If (Error(k) < Error(k-1)) then
  - SmallRandomPerturb(Weights(k))
- Else Restore(Weights(k-1)) and
  - SmallRandomPerturb(Weights(k-1))

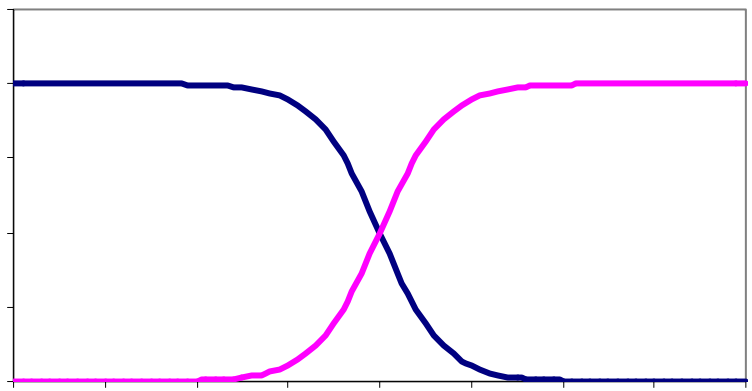
Low Power Neurons

Differential Transconductance Amplifier

$$I_1 = I_{SS}\frac{1}{1+e^{-q\kappa V_{diff}/kT}}$$

$$I_2 = I_{SS}\frac{1}{1+e^{q\kappa V_{diff}/kT}}$$

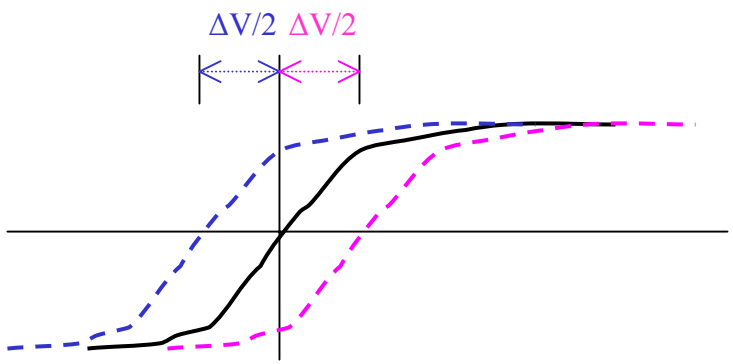




Differentiation

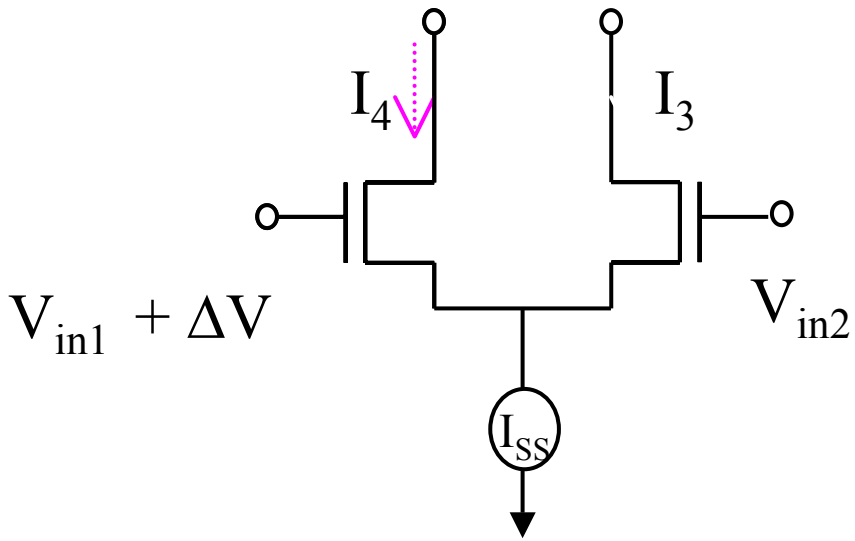
- Central Difference Approximation

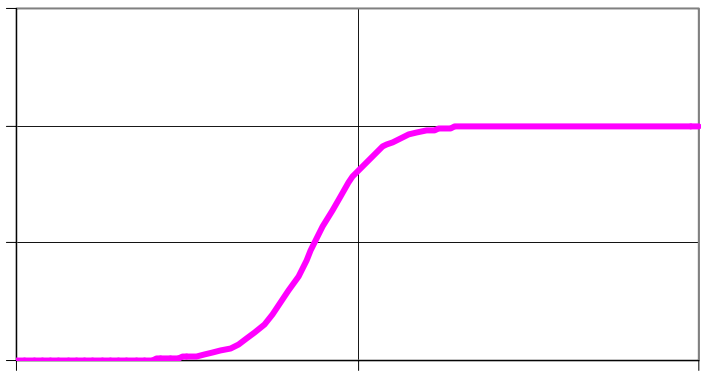
$$I'(V_{diff}) = \lim_{\Delta V \to 0} \frac{I(V_{diff} + \Delta V) - I(V_{diff} - \Delta V)}{2\Delta V}$$



Differentiation

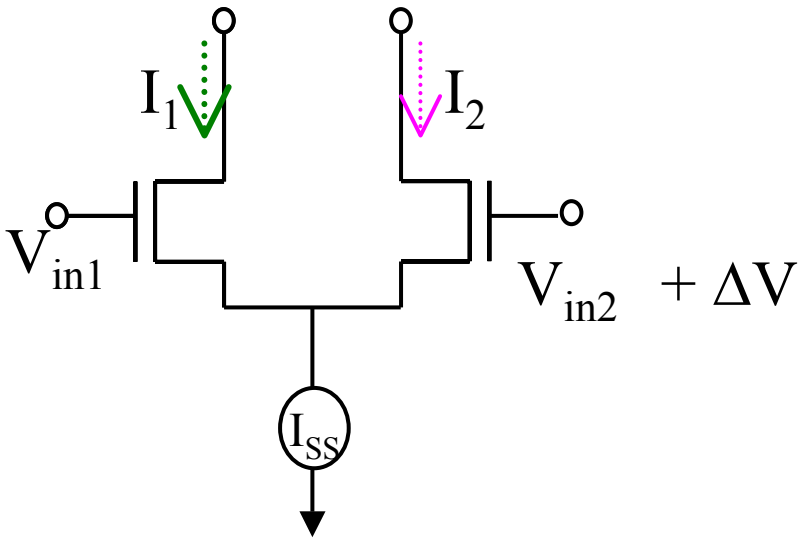
$$I = I(V_{diff} + \Delta V)$$





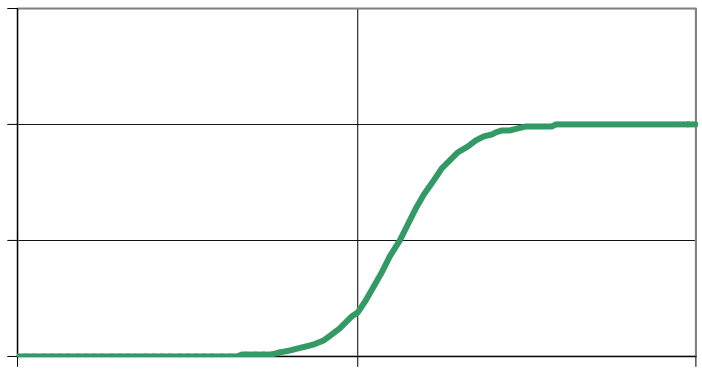
$$V_{diff} = V_{in1} - V_{in2}$$

Differentiation

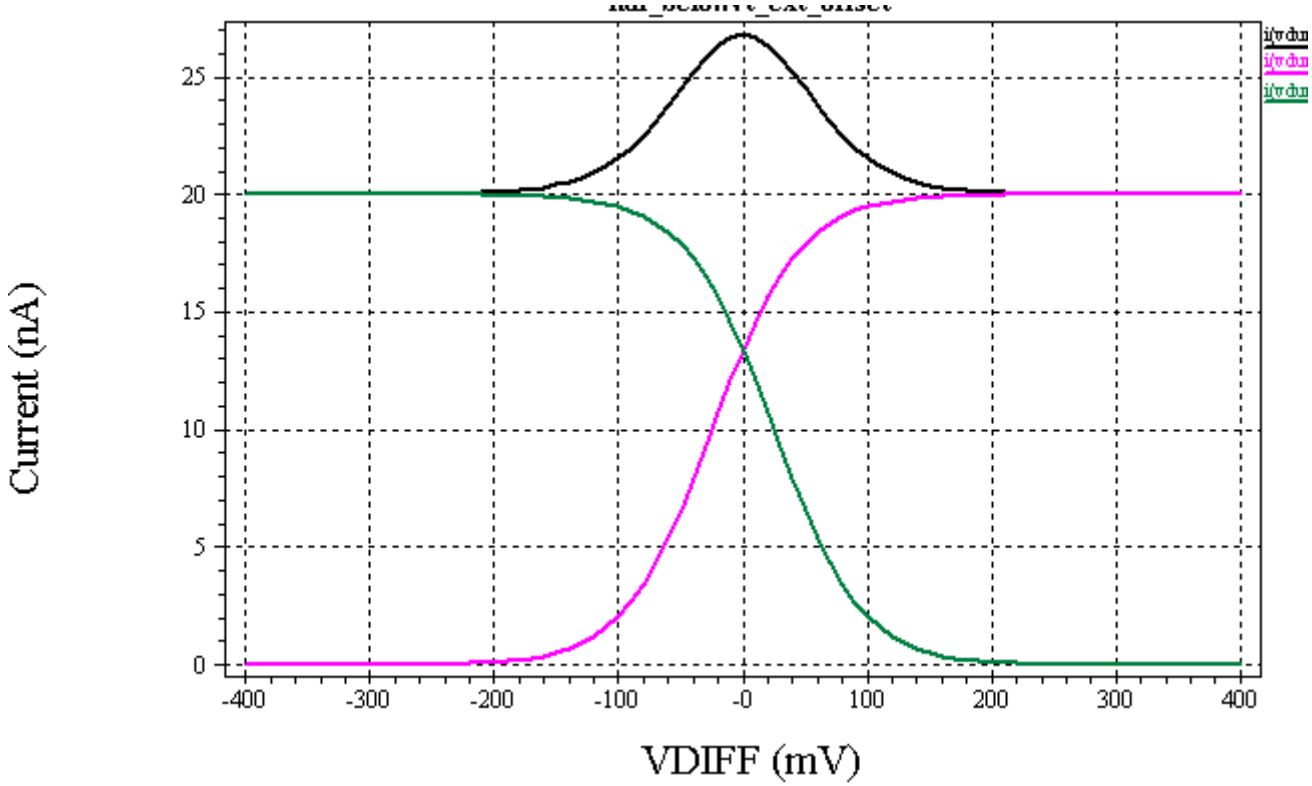


$$I = I(V_{diff} - \Delta V)$$

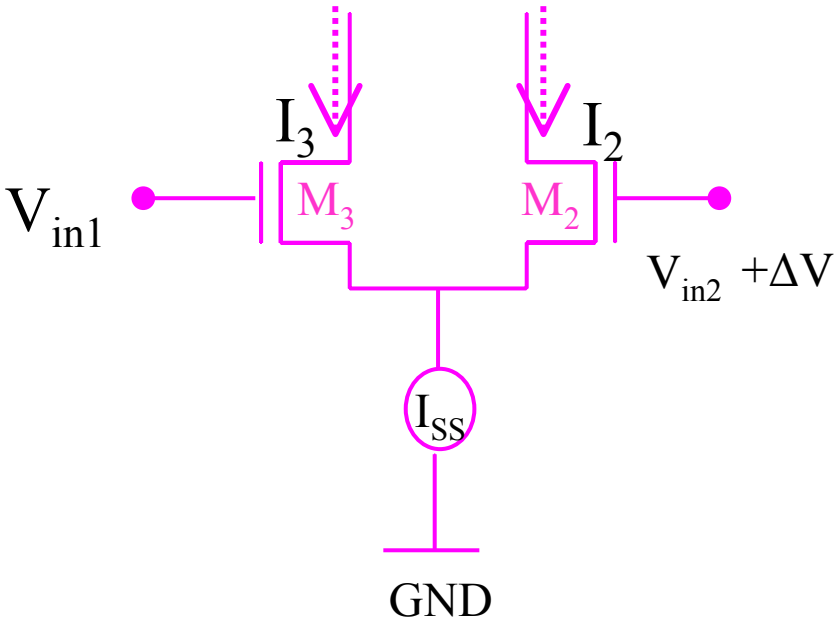
$$V_{diff} = V_{in1} - V_{in2}$$

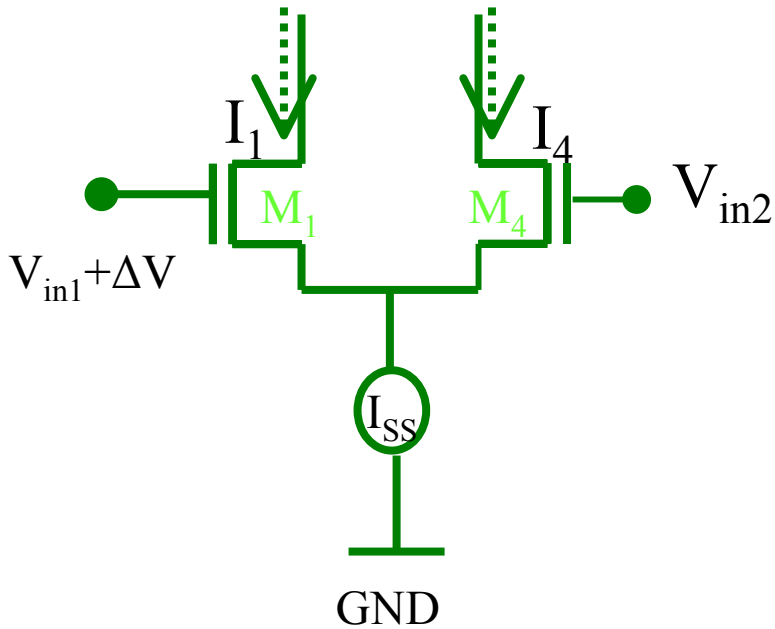


Differentiation



Derivative of Neuron Activation Function (DNAF)

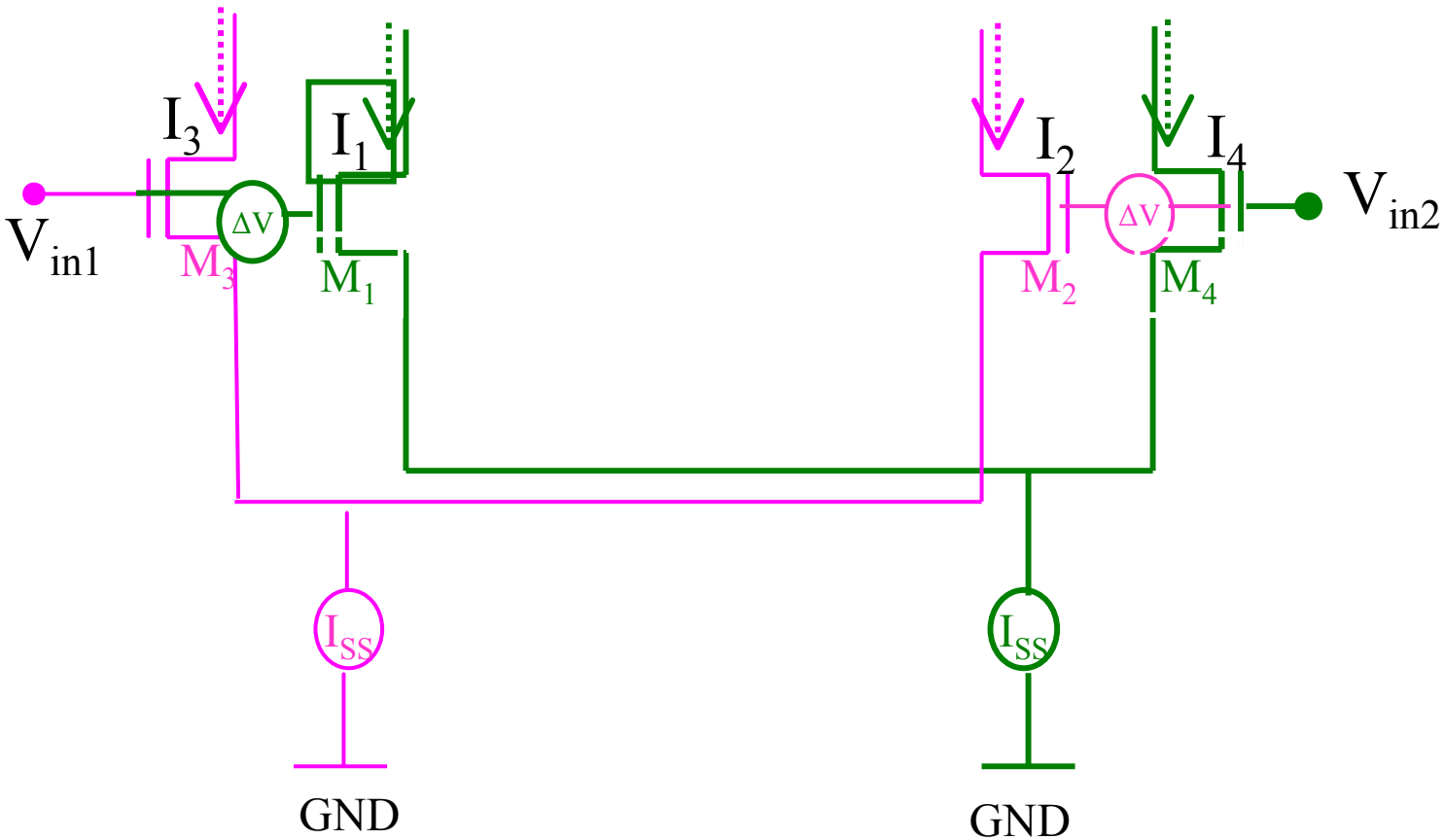




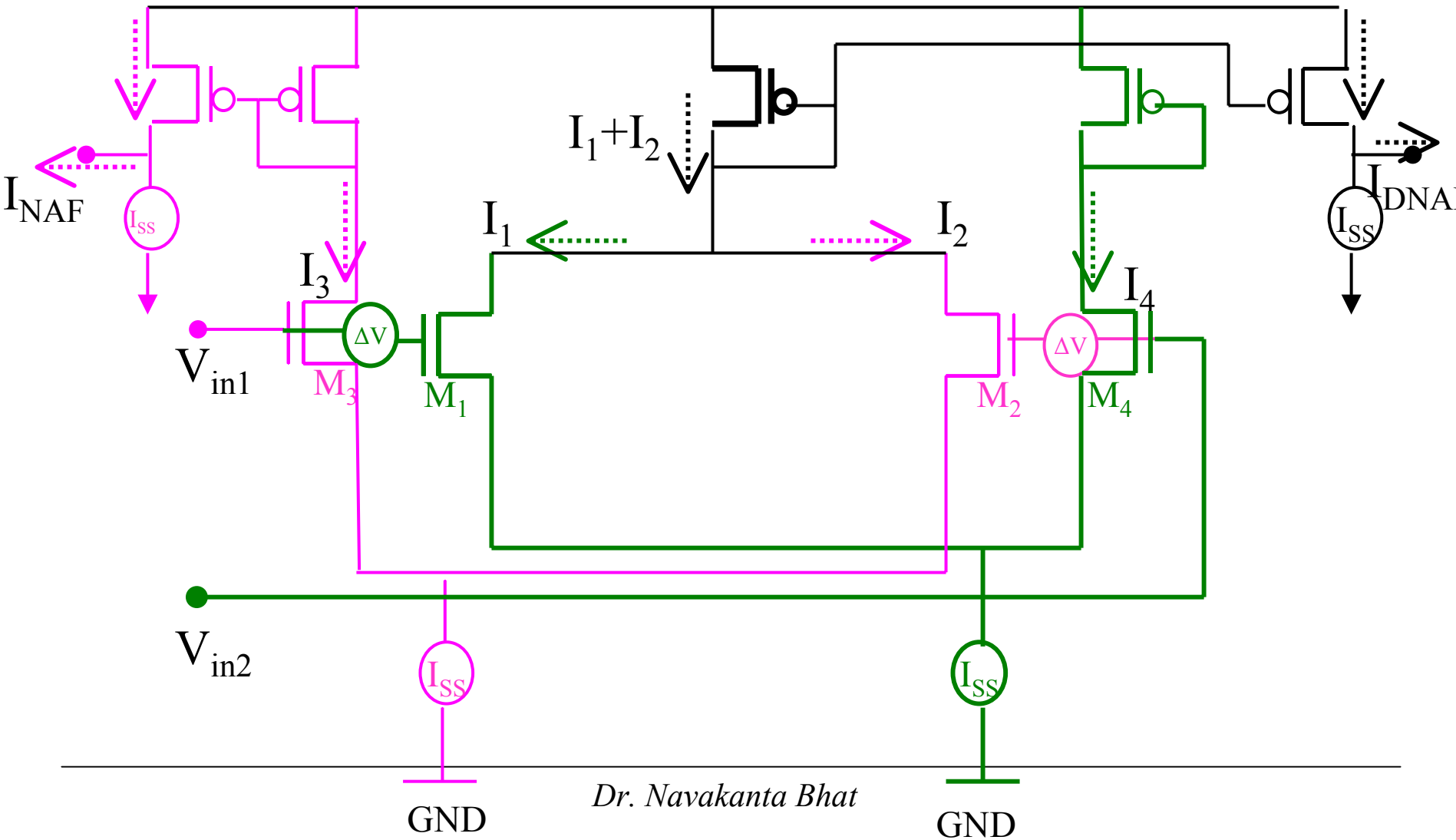
$$I = I(V_{diff} - \Delta V)$$

$$I = I(V_{diff} + \Delta V)$$

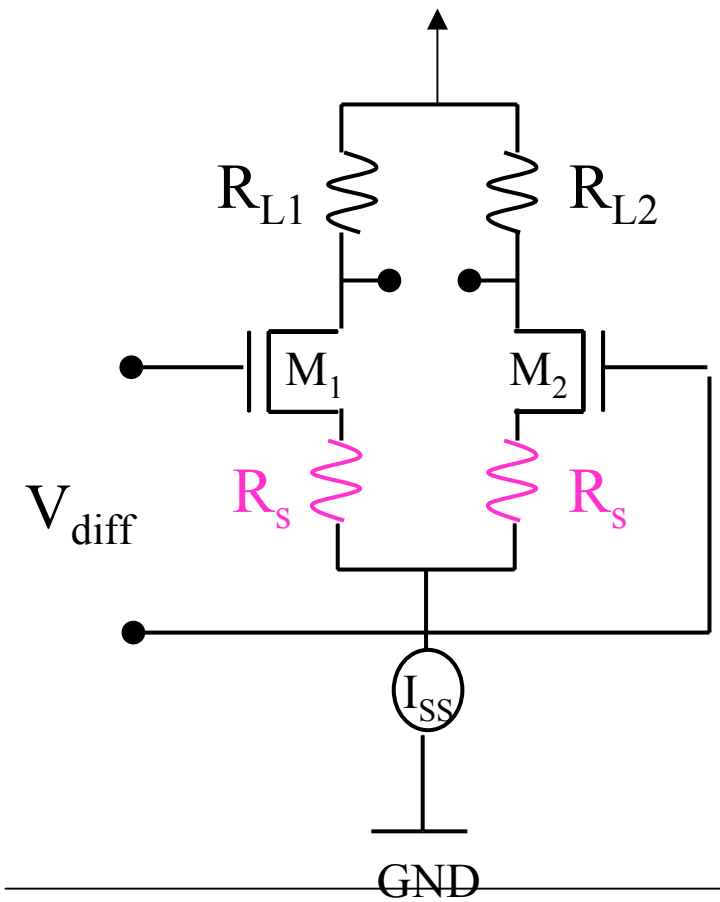
Derivative of Neuron Activation Function (DNAF)



Neuron Circuit (with External Offset Voltage)

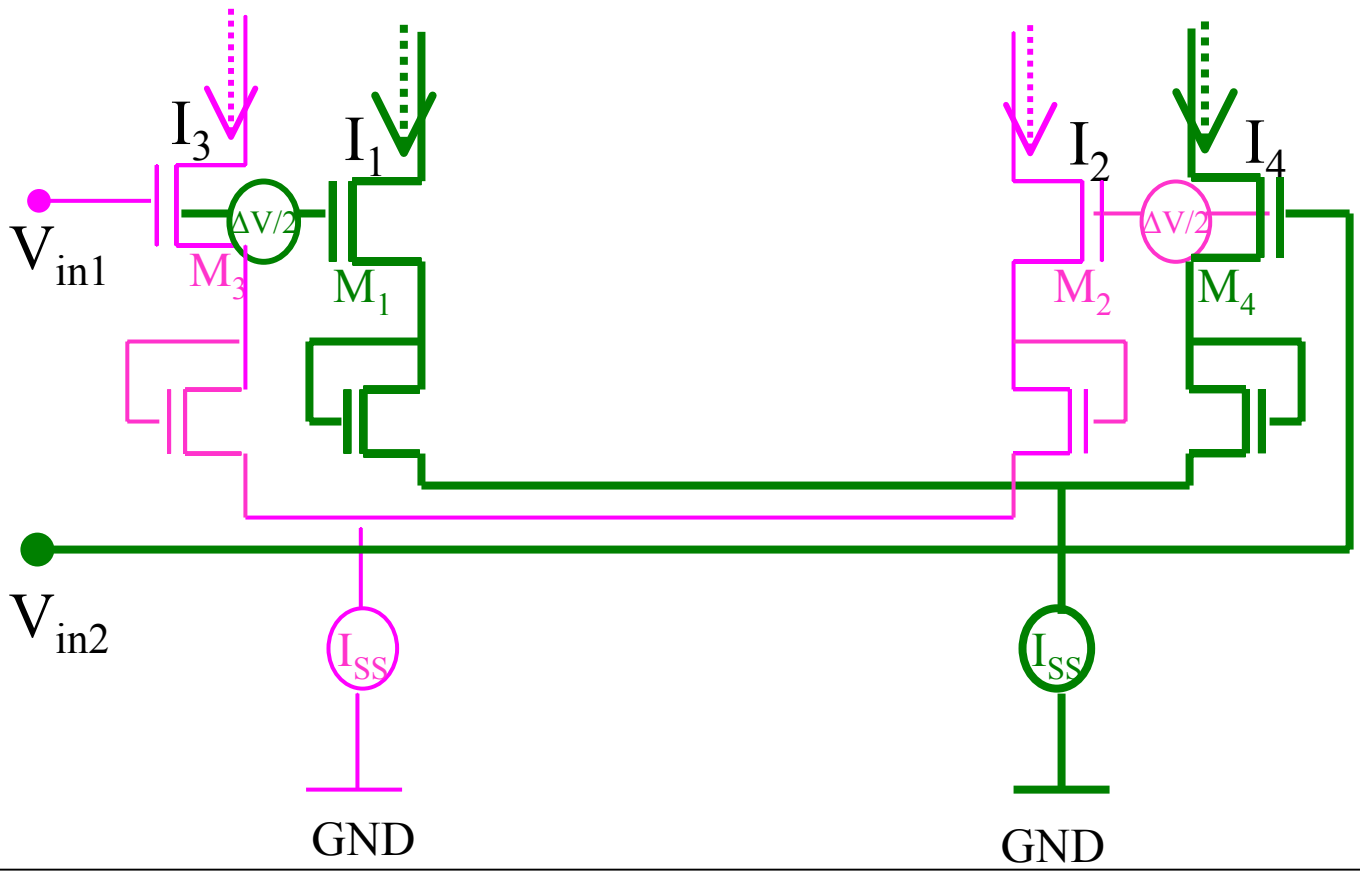


Source Degeneration

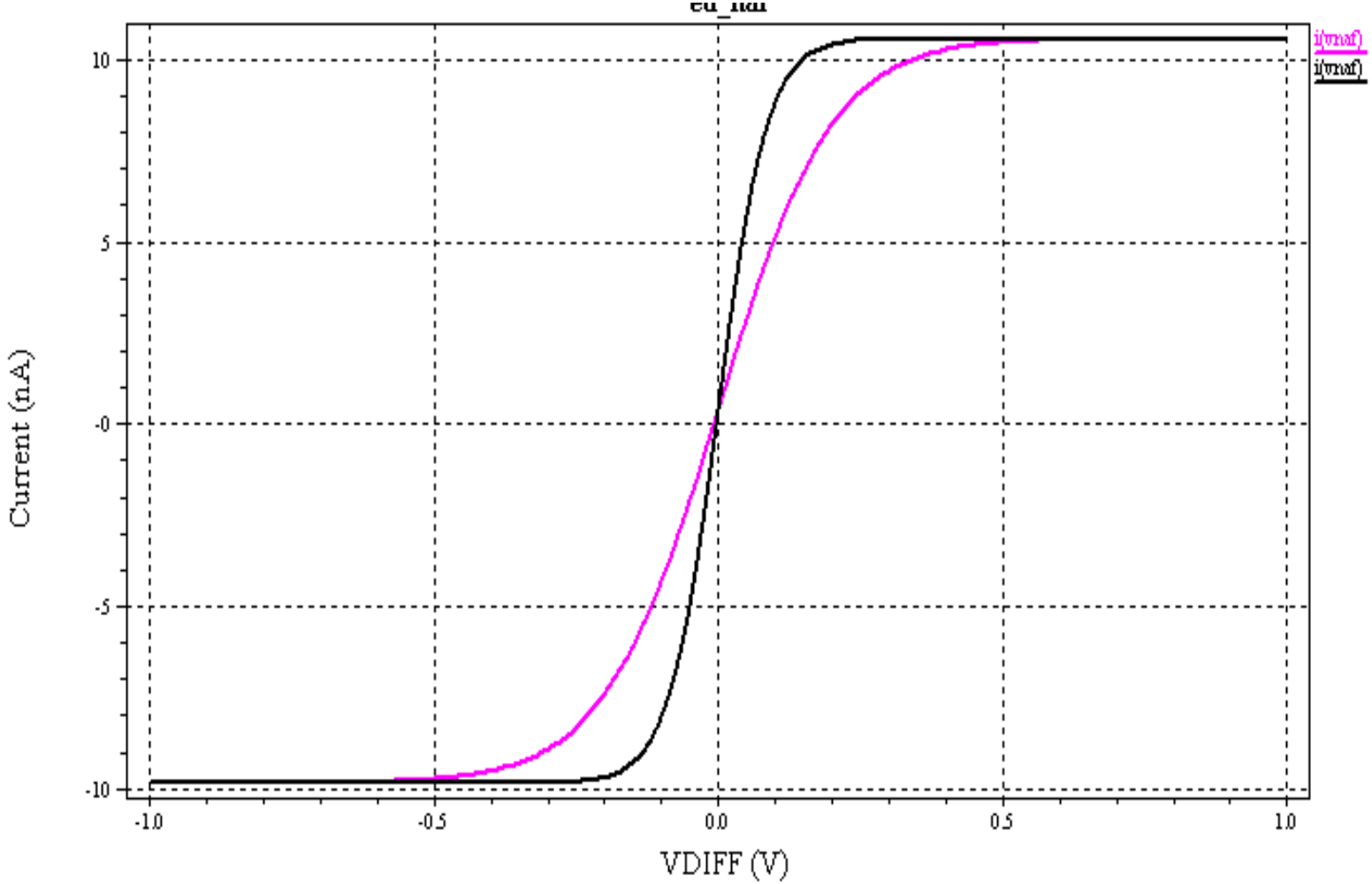


- To increase the range of input voltage over which the diff pair behaves approximately as a linear amplifier.
- To increase the input impedance.
- To stabilize the gain (negative feedback)

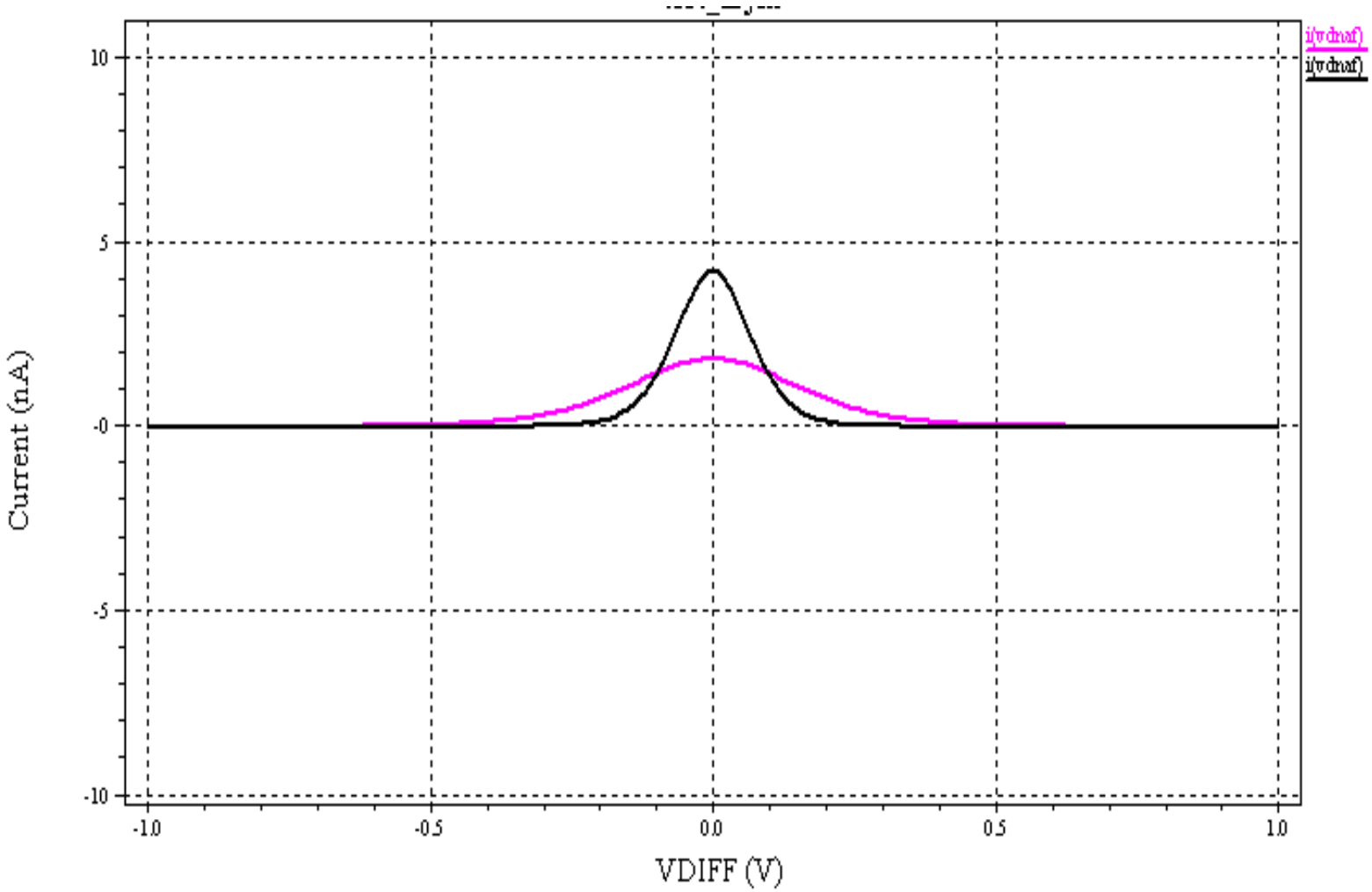
Effect of Source Degeneration on NAF/DNAF



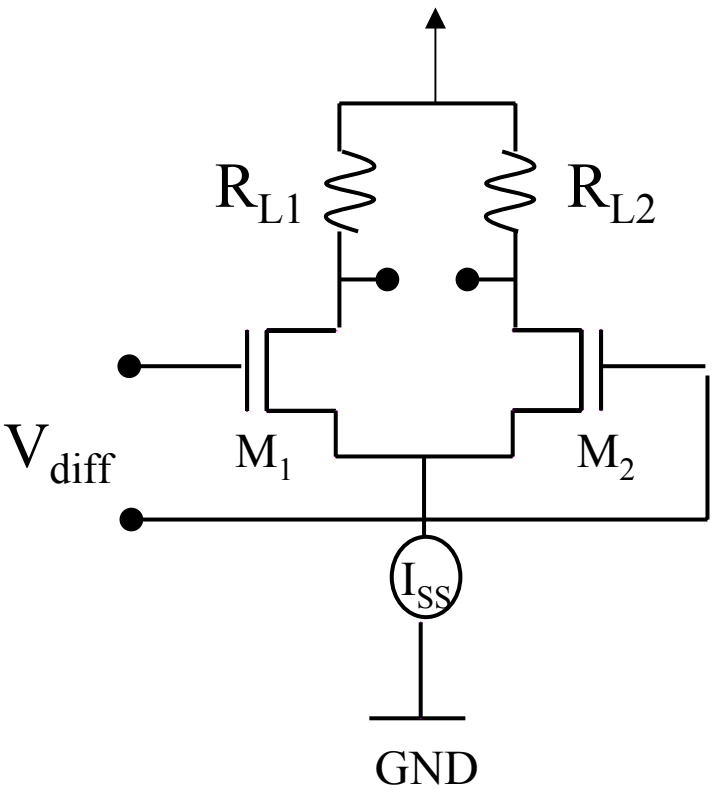
NAF



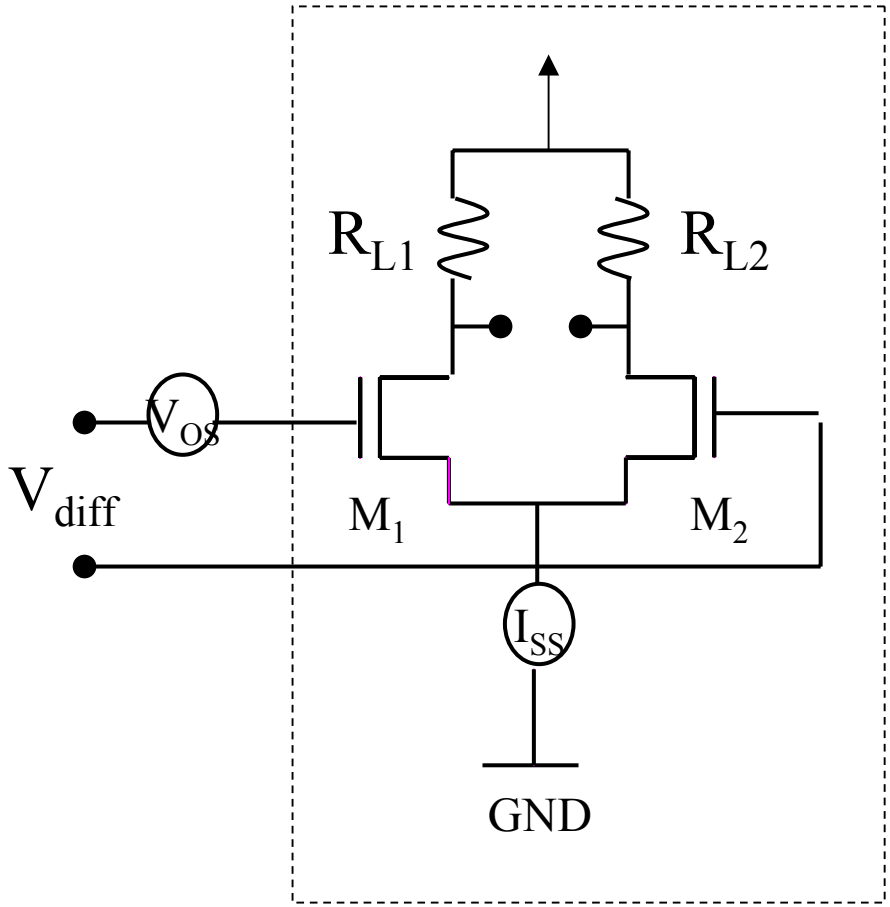
DNAF



Offset Voltage

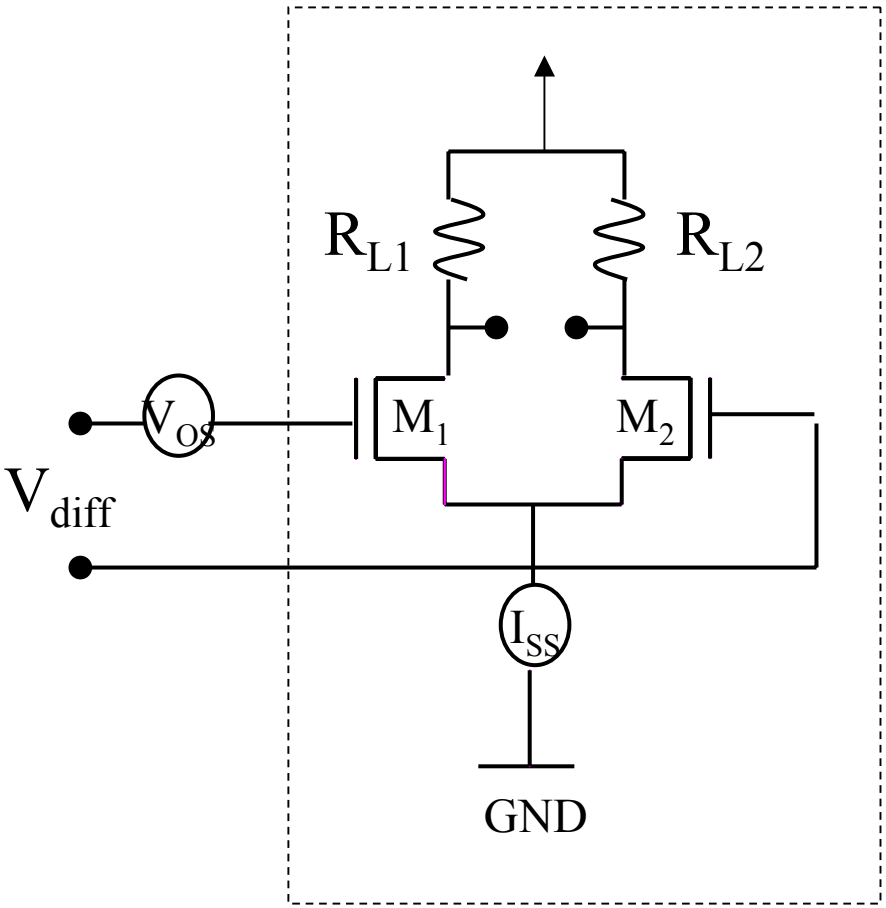


Asymmetric differential pair



Symmetric differential pair

Offset Voltage



$$V_{OS} = V_{GS1} - V_{GS2}$$

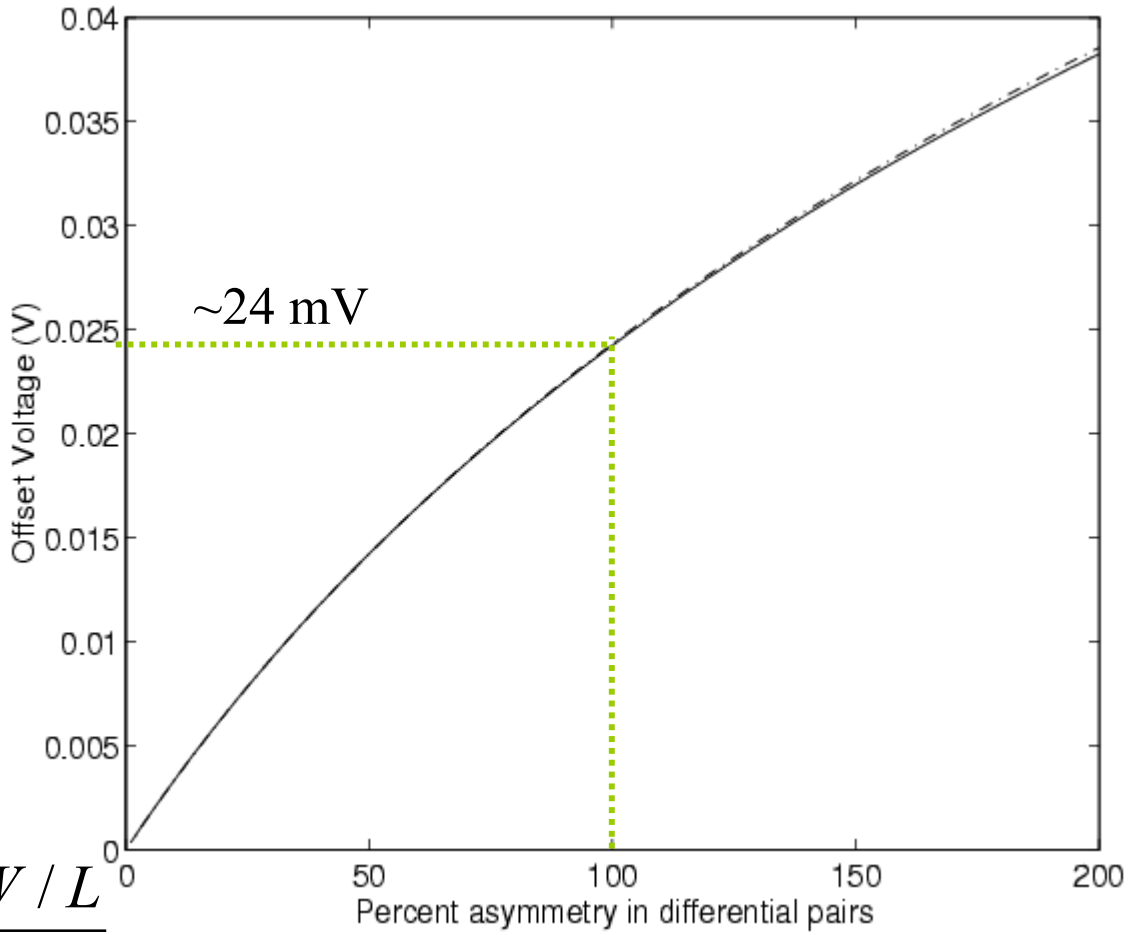
$$V_{GS} = \frac{kT}{\kappa q} \ln\left(\frac{I}{I_0}\right)$$

$$I_1 R_{L1} = I_2 R_{L2}$$

$$V_{OS} = \frac{kT}{\kappa q} \ln\left[\frac{R_{L2}(W/L)_2}{R_{L1}(W/L)_1}\right] + \Delta V_{Th0}$$

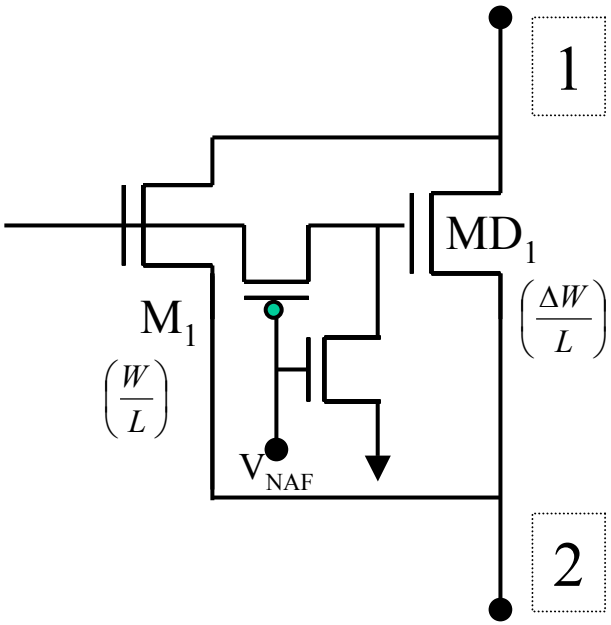
Symmetric differential pair

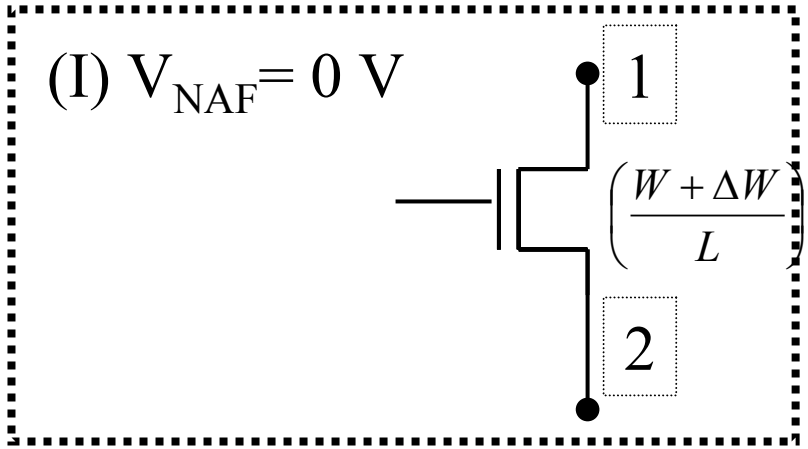
Offset Voltage Vs Asymmetry

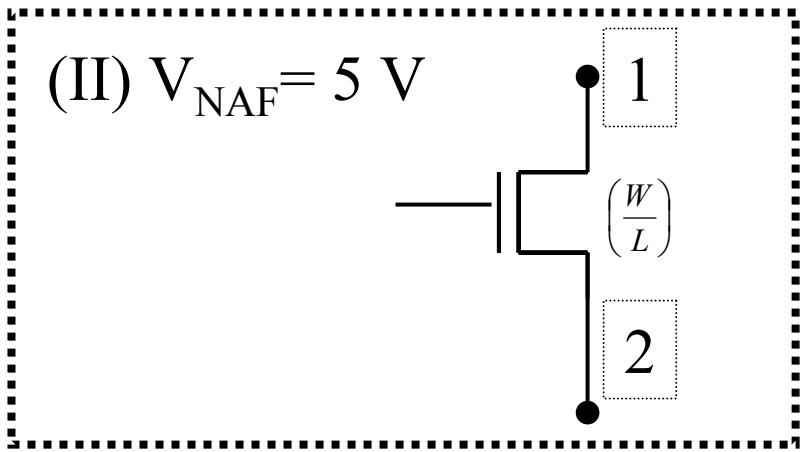


$$Asymmetry = \frac{\Delta W / L}{W / L}$$

Asymmetry

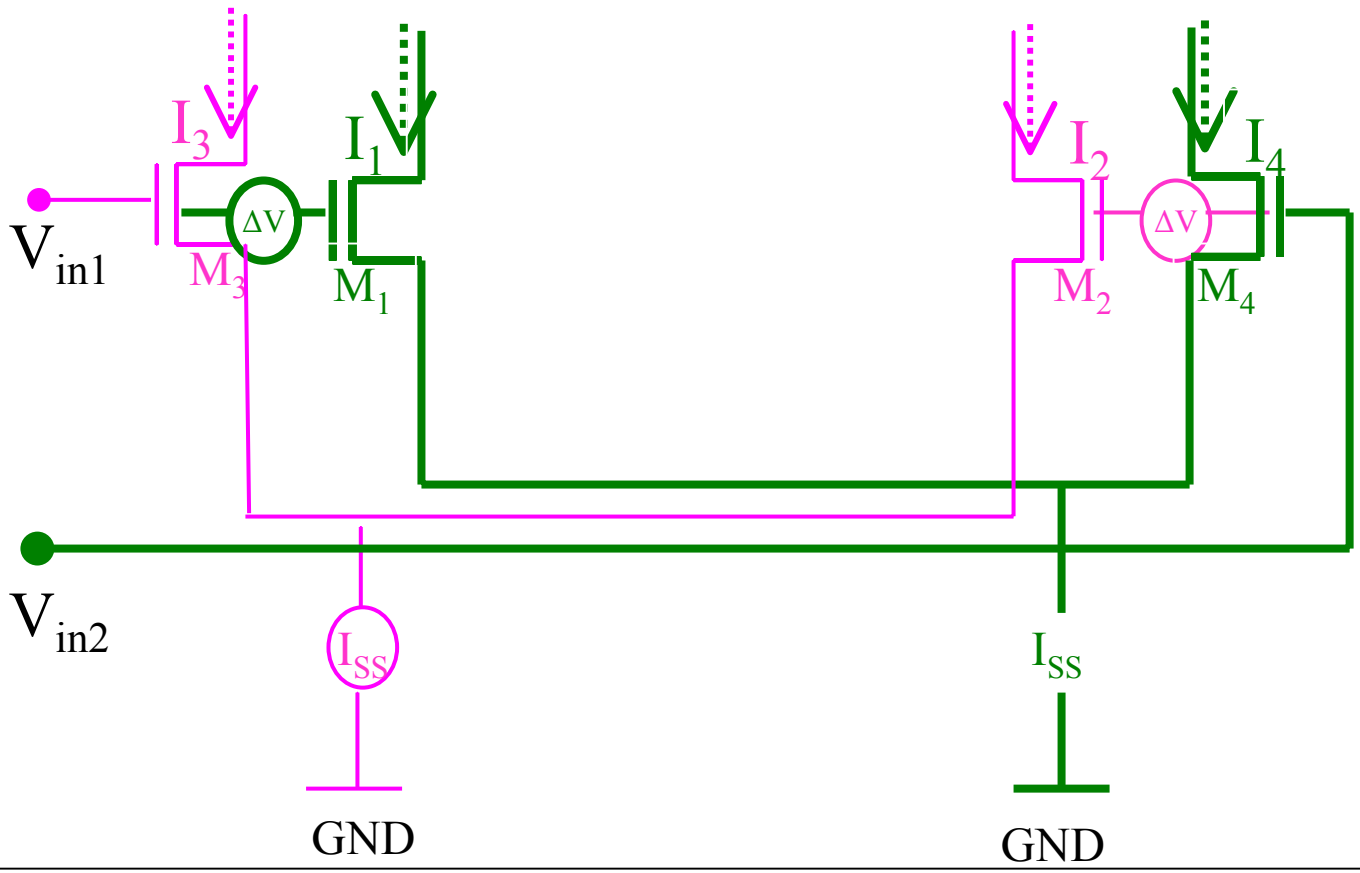




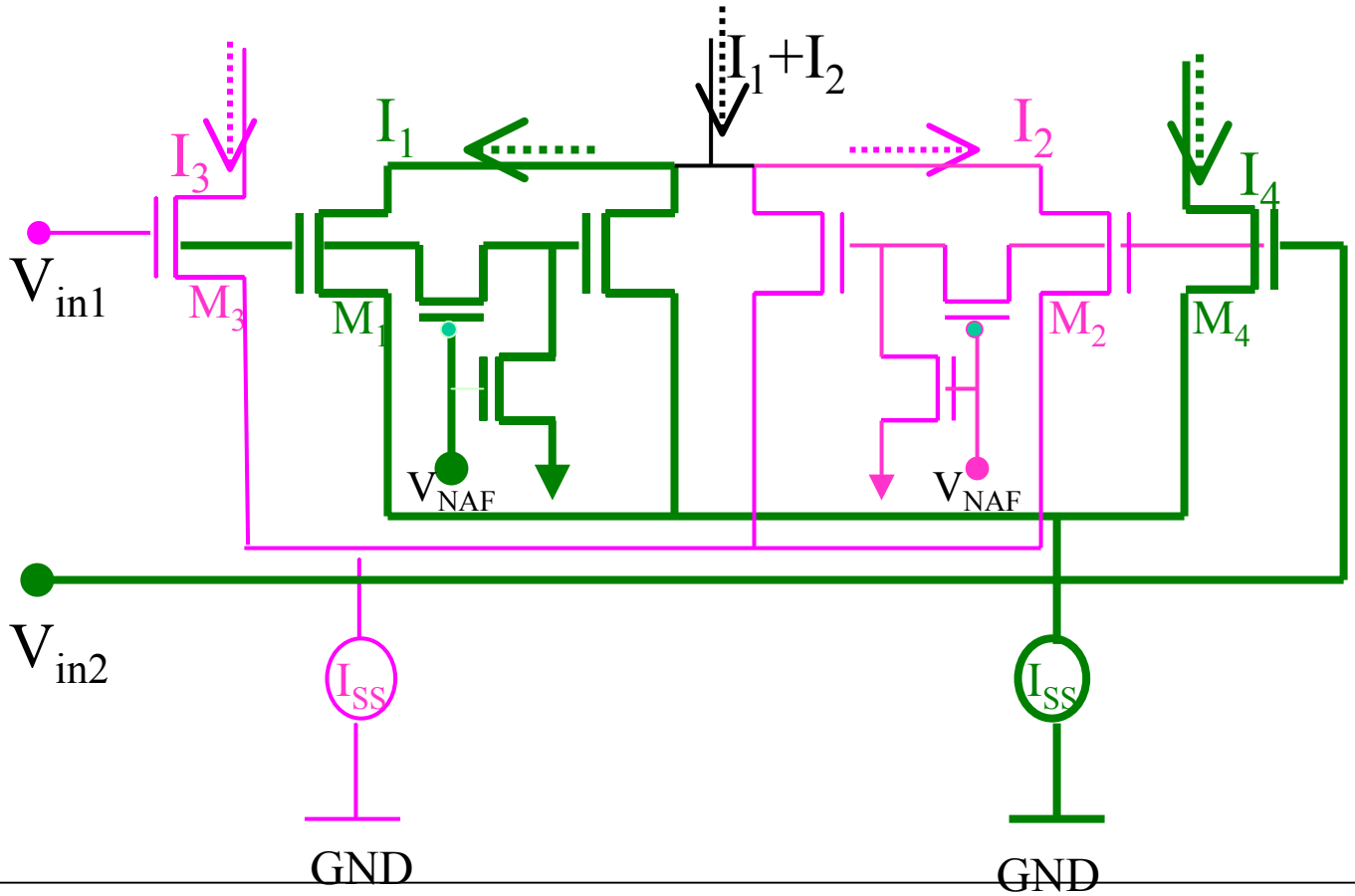


$$Asymmetry = \frac{\Delta W / L}{W / L}$$

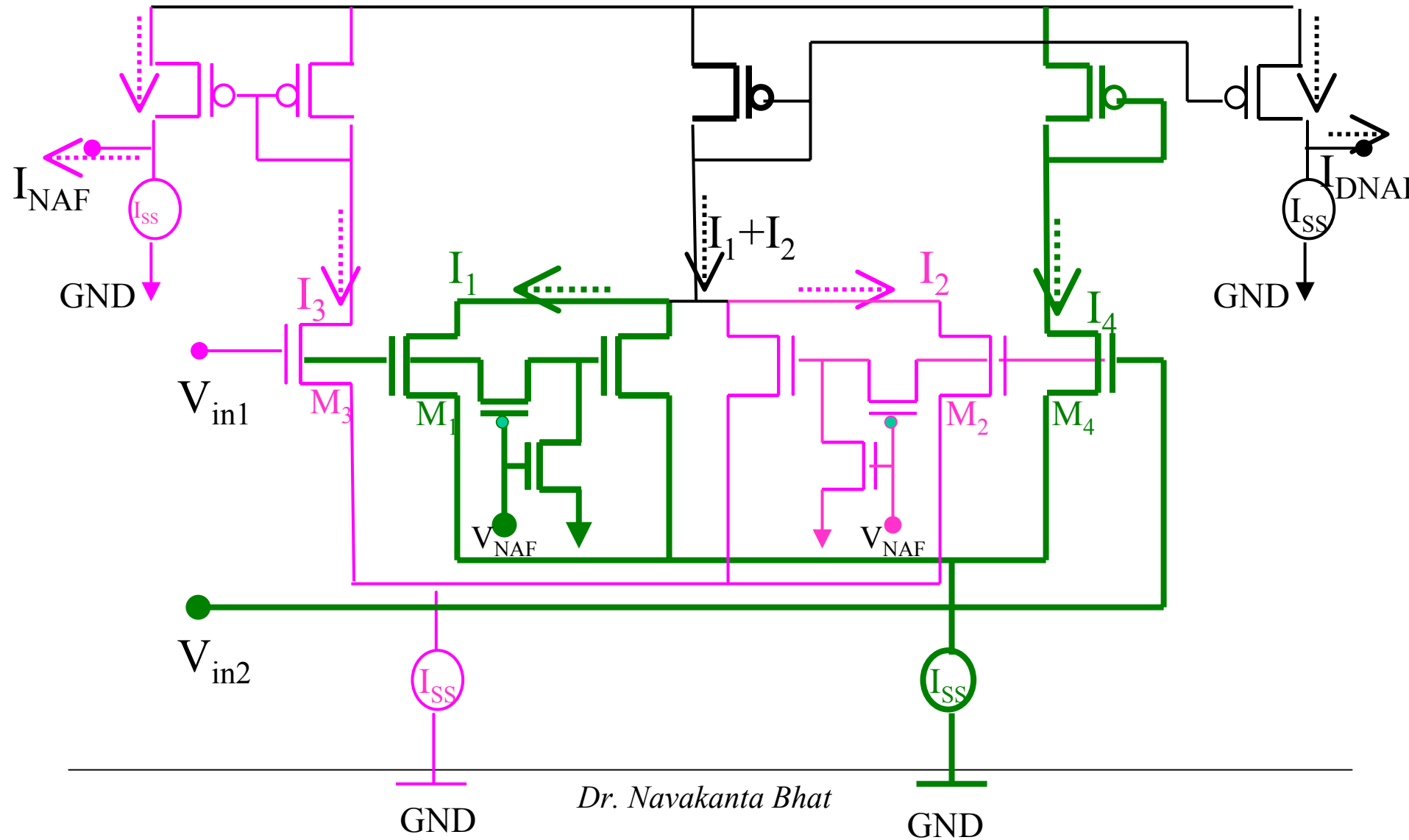
Derivative of Neuron Activation Function (DNAF)



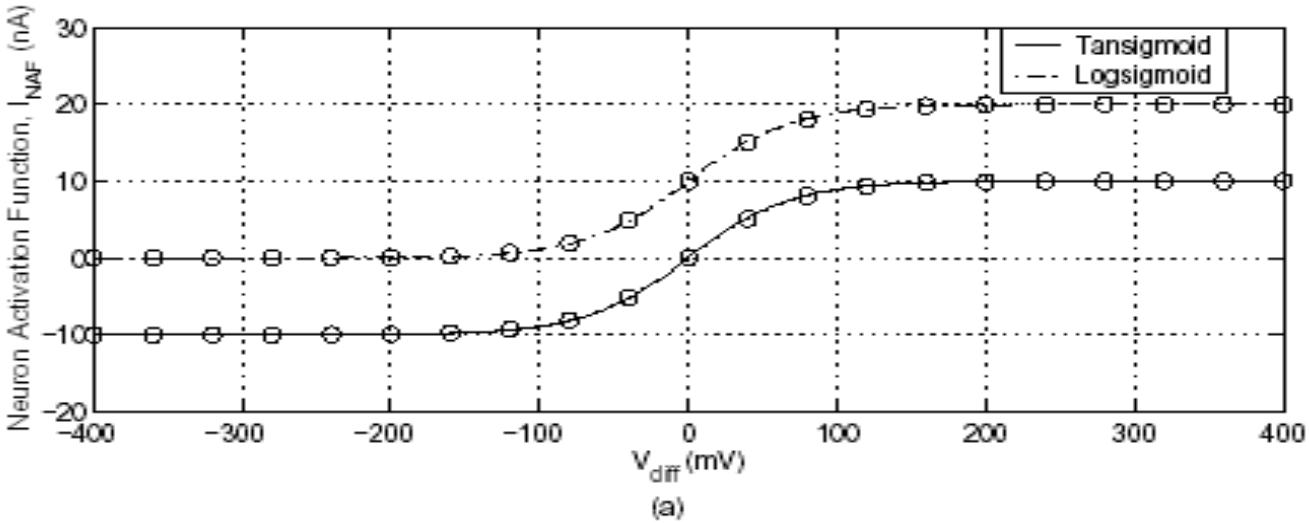
Derivative of Neuron Activation Function (DNAF)



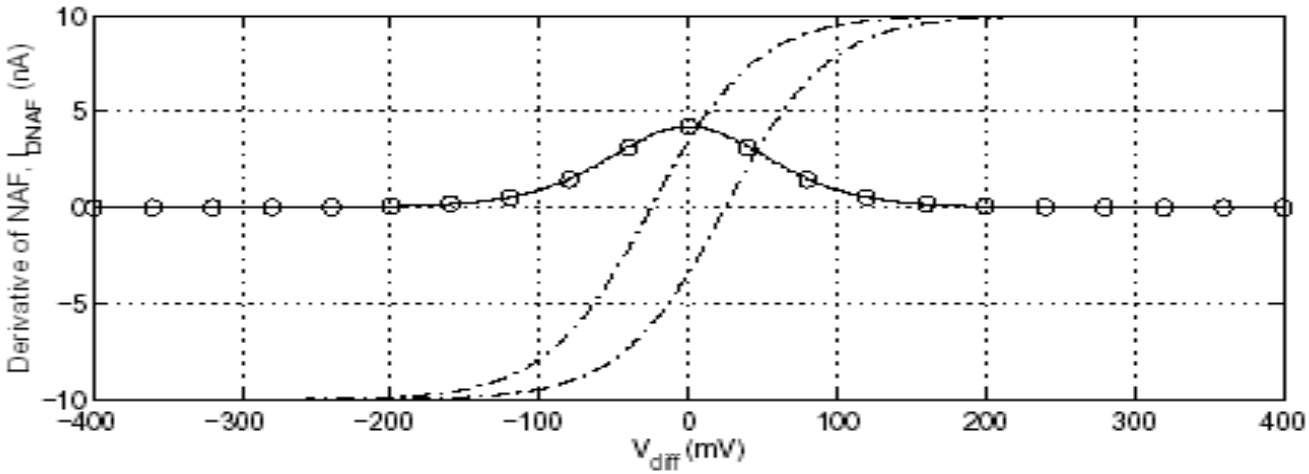
Neuron Circuit (with Asymmetry)



NAF & DNAF

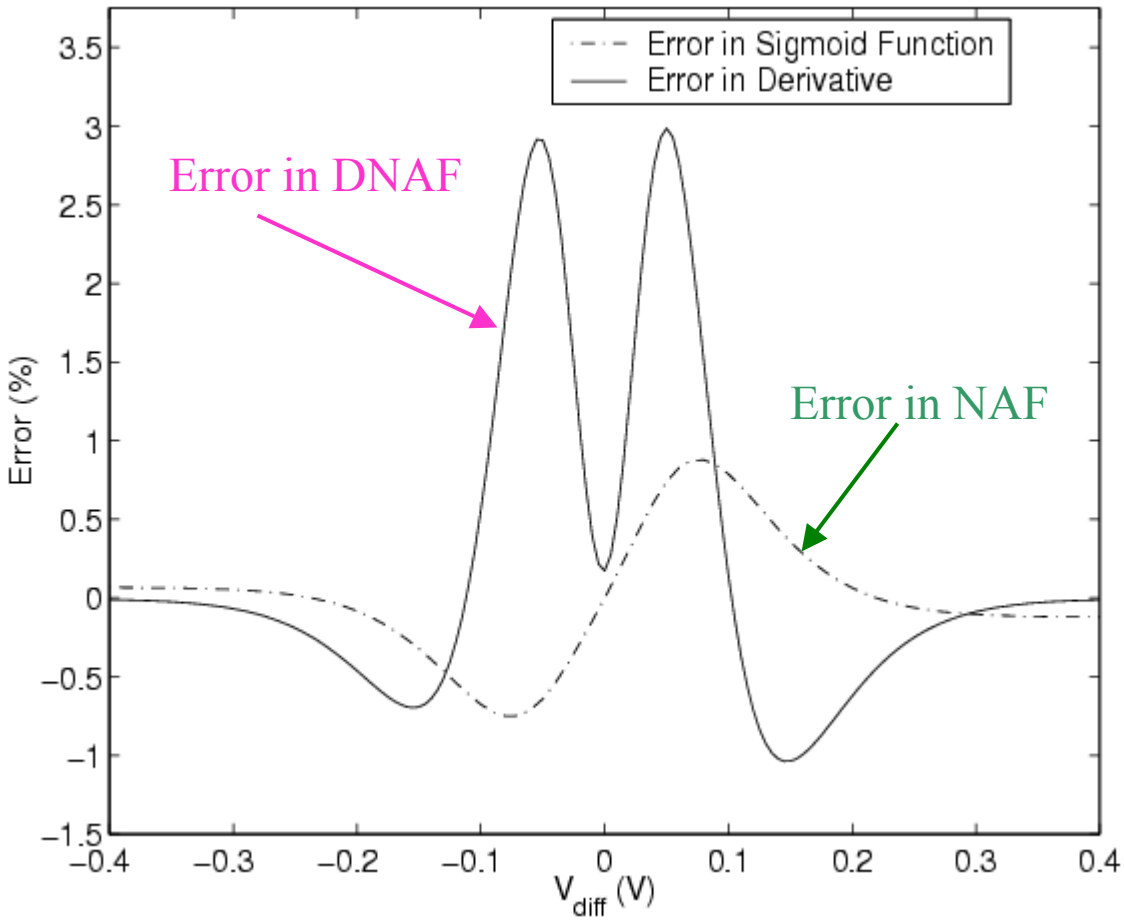

Tansigmoid
Logsigmoid
Neuron Activation Function, I_{NAF} (nA)
30
20
10
0
-10
-20
-400
-300
-200
-100
0
100
200
300
400
V_{diff} (mV)

(a)

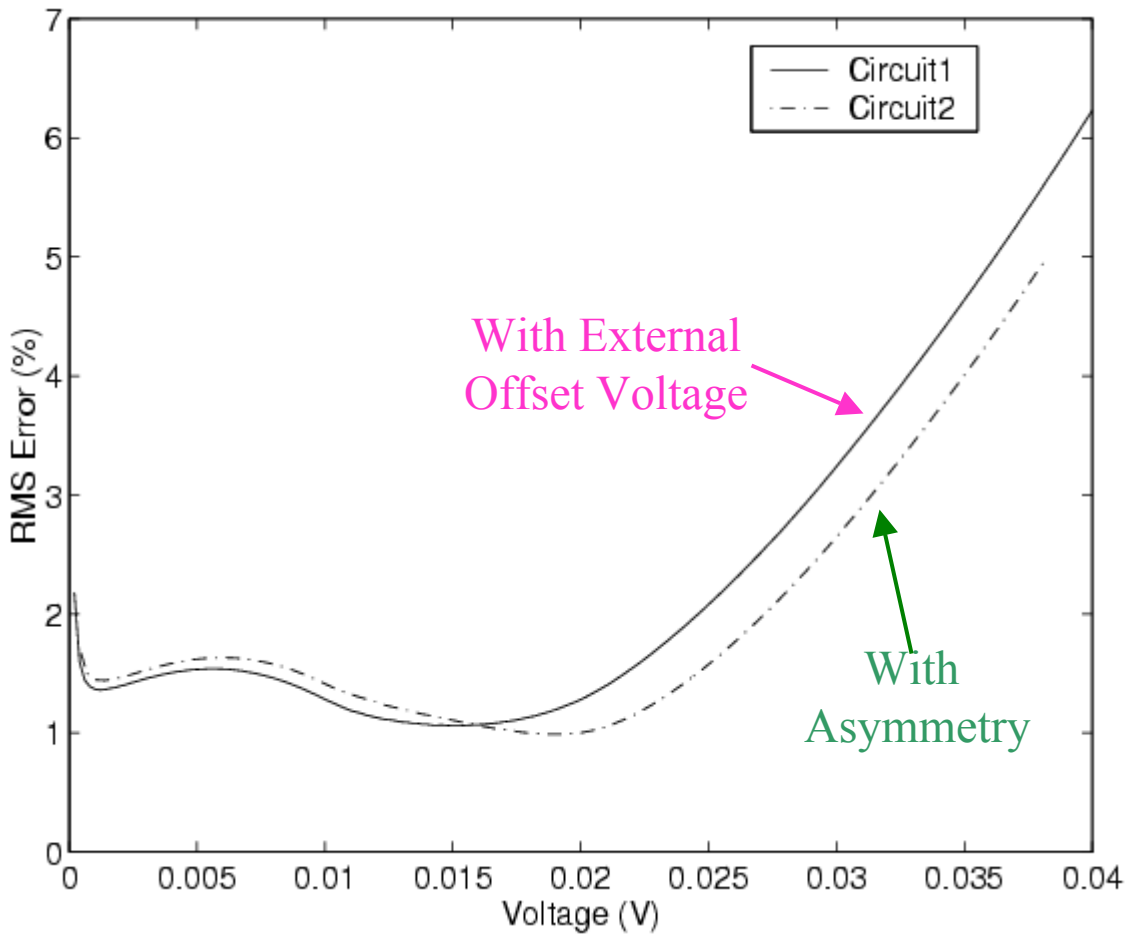

Derivative of NAF, I_{DNAF} (nA)
10
5
0
-5
-10
-400
-300
-200
-100
0
100
200
300
400
V_{diff} (mV)

(b)

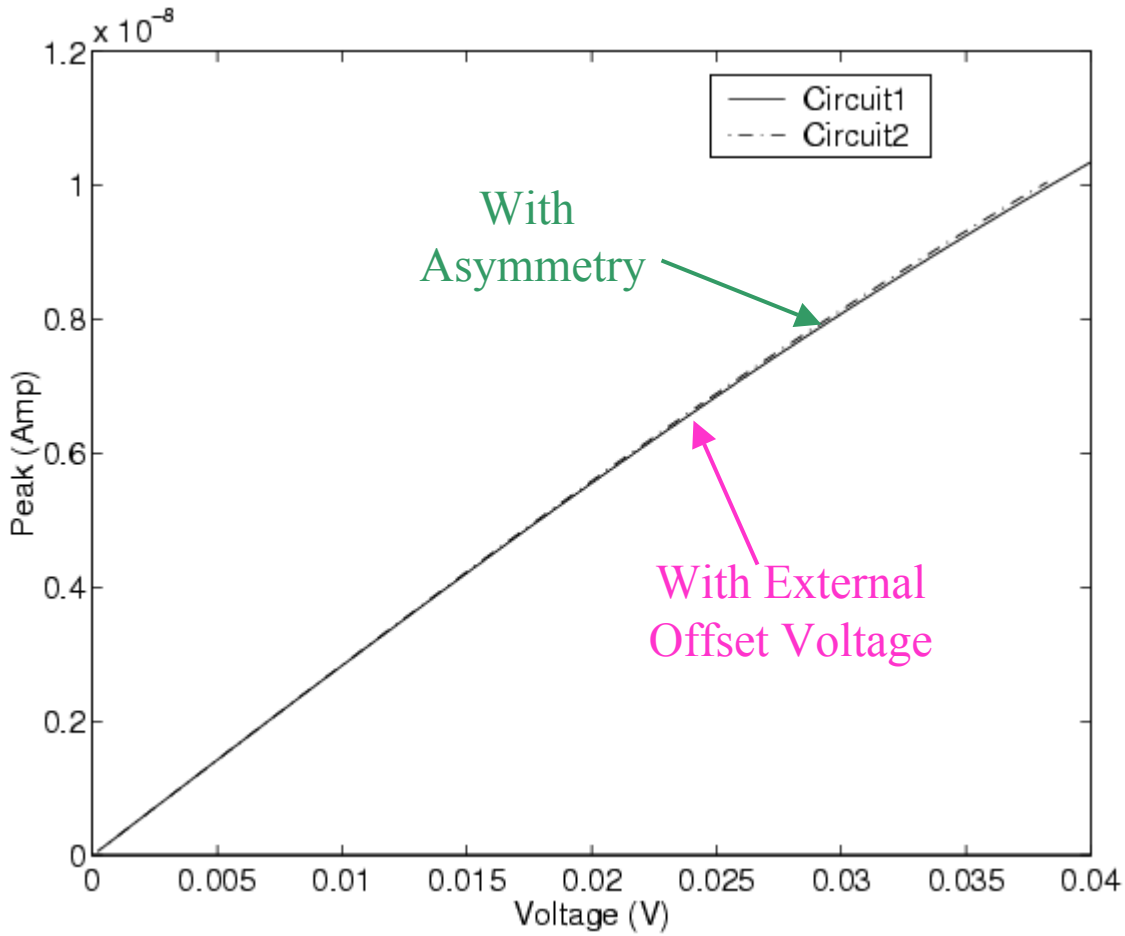
Error Curves



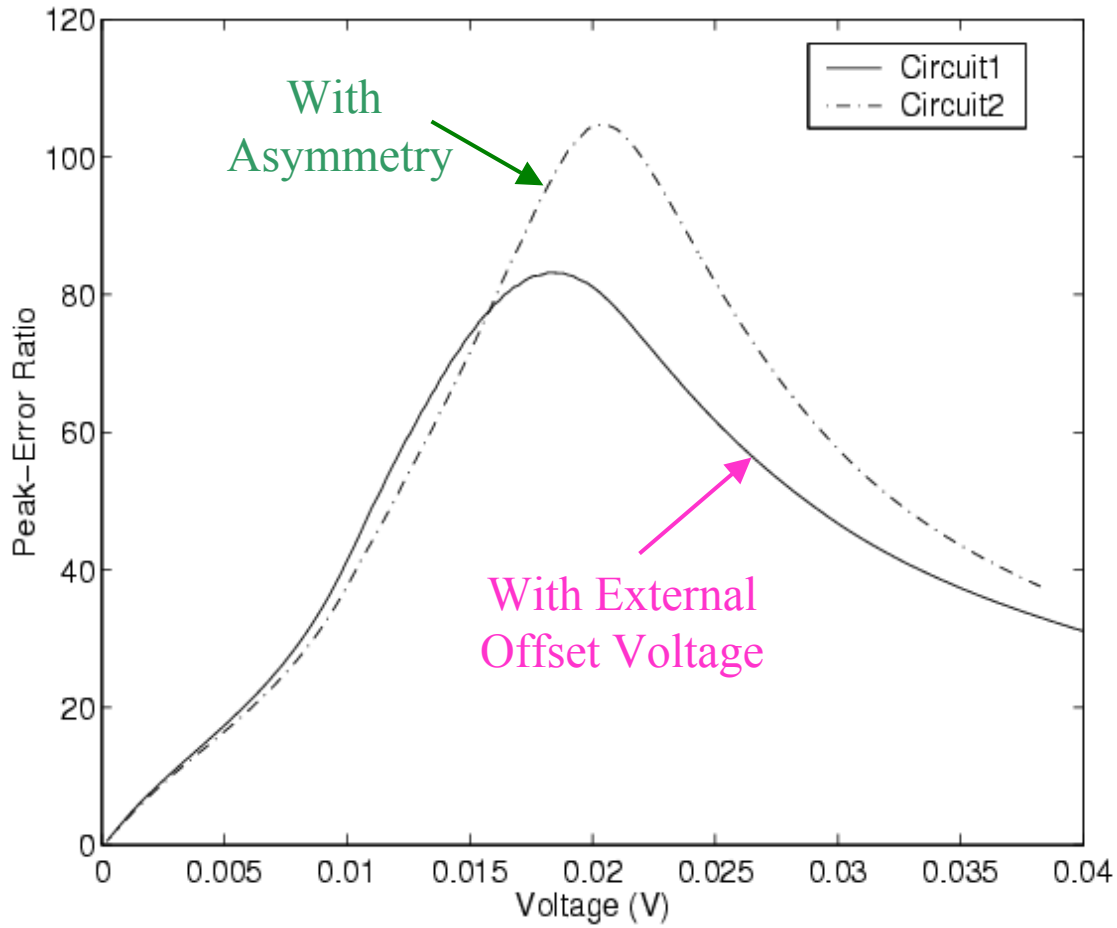
RMS Error Vs Offset Voltage



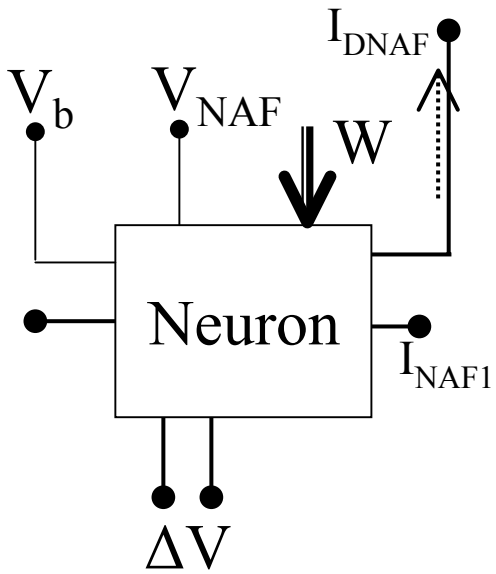
Peak DNAF Vs Offset Voltage

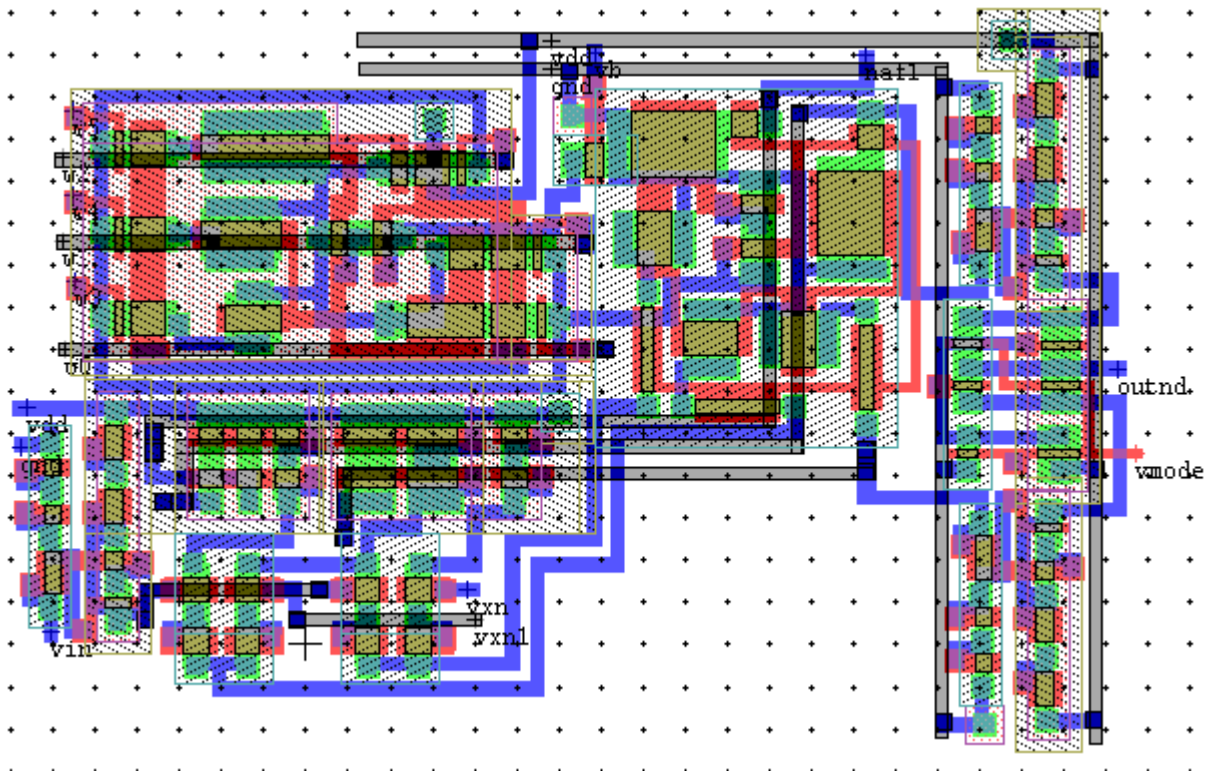


Peak-Error-Ratio Vs Offset Voltage

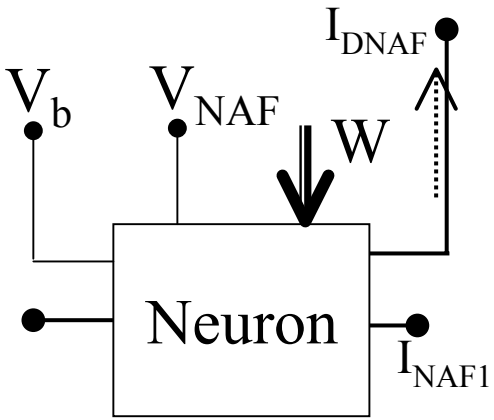


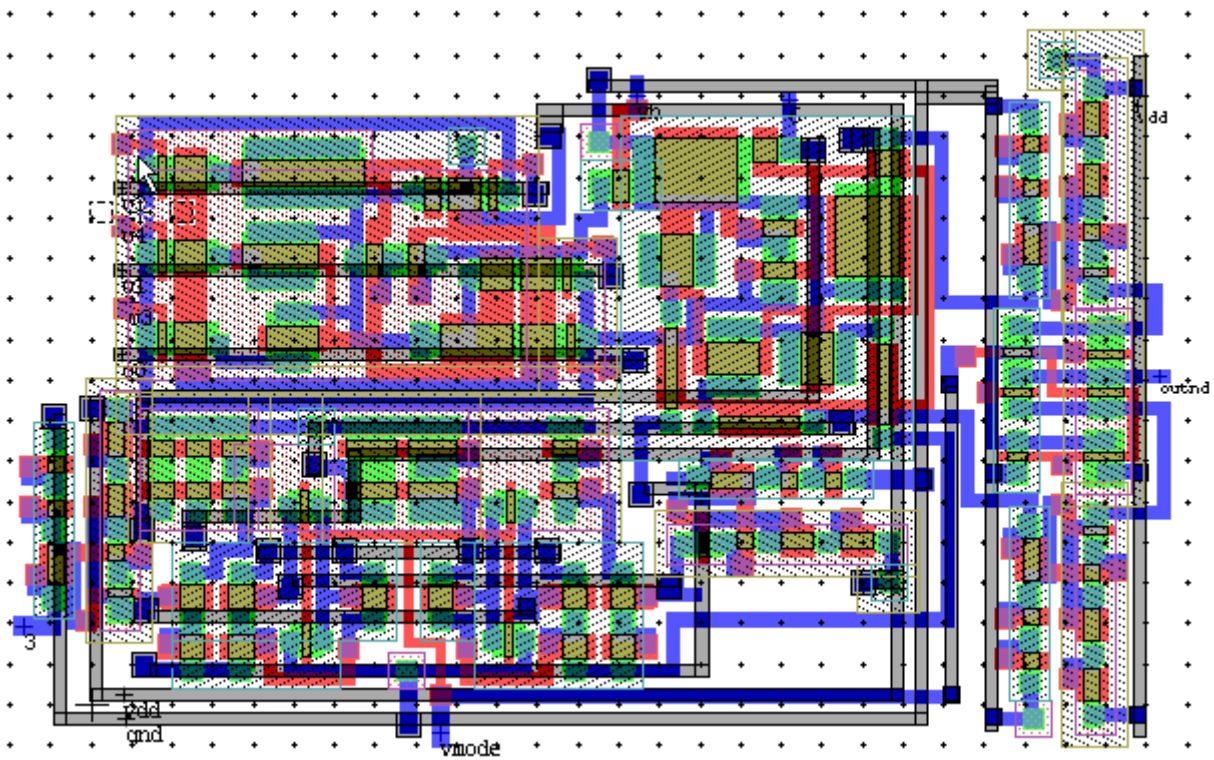
Neuron (with External Offset Voltage)





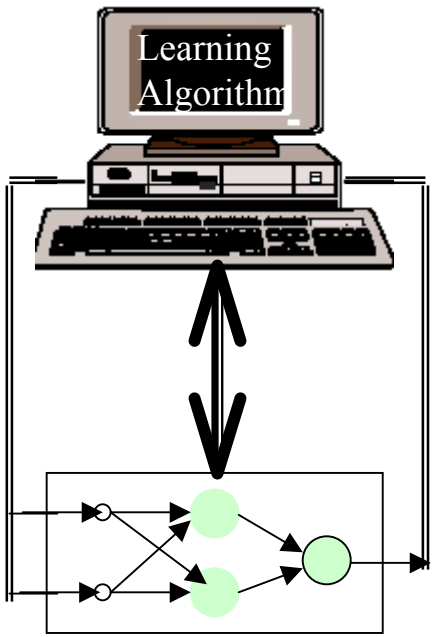
Neuron (with Asymmetry)





Low Power Feedforward Neural Network

Training



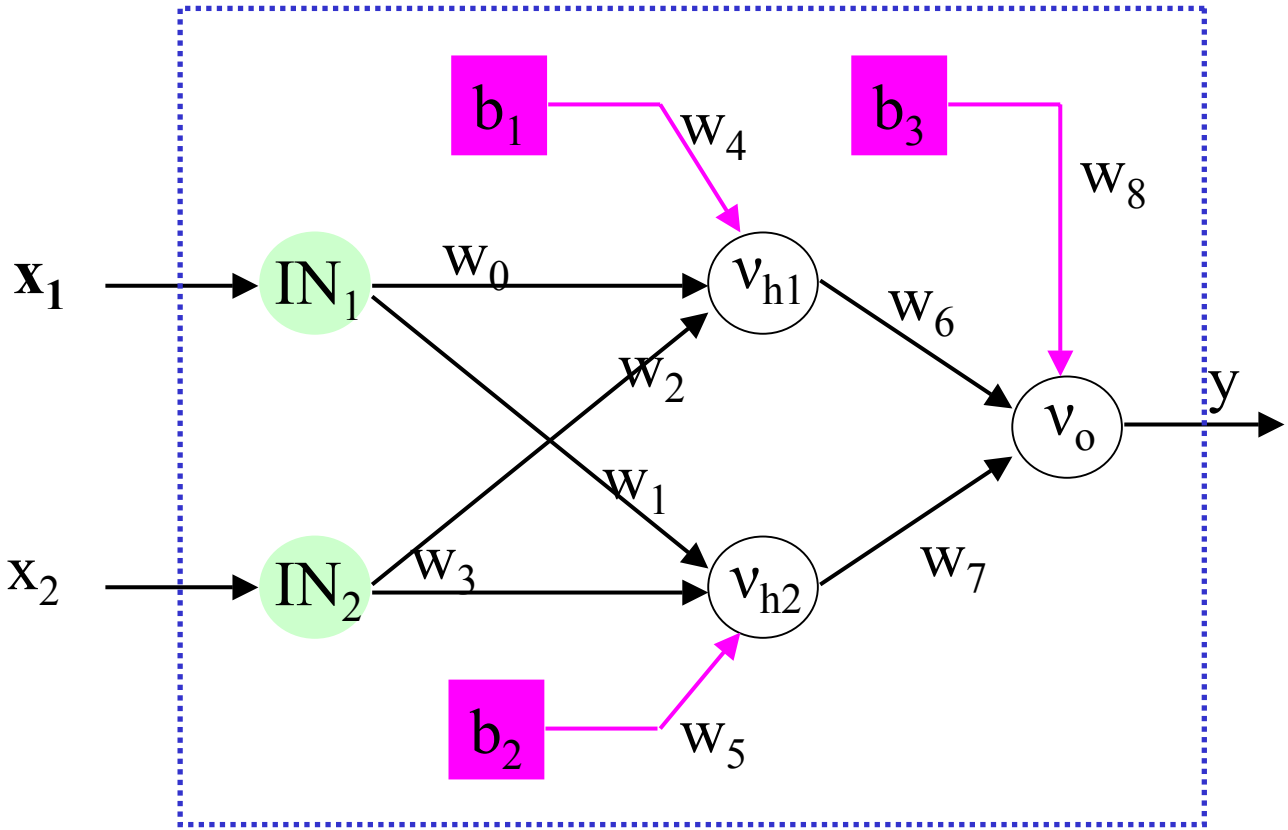
NN System

- **Chip-in-the-Loop Training**

- Learning Algorithm on the host PC
- Weight update from PC

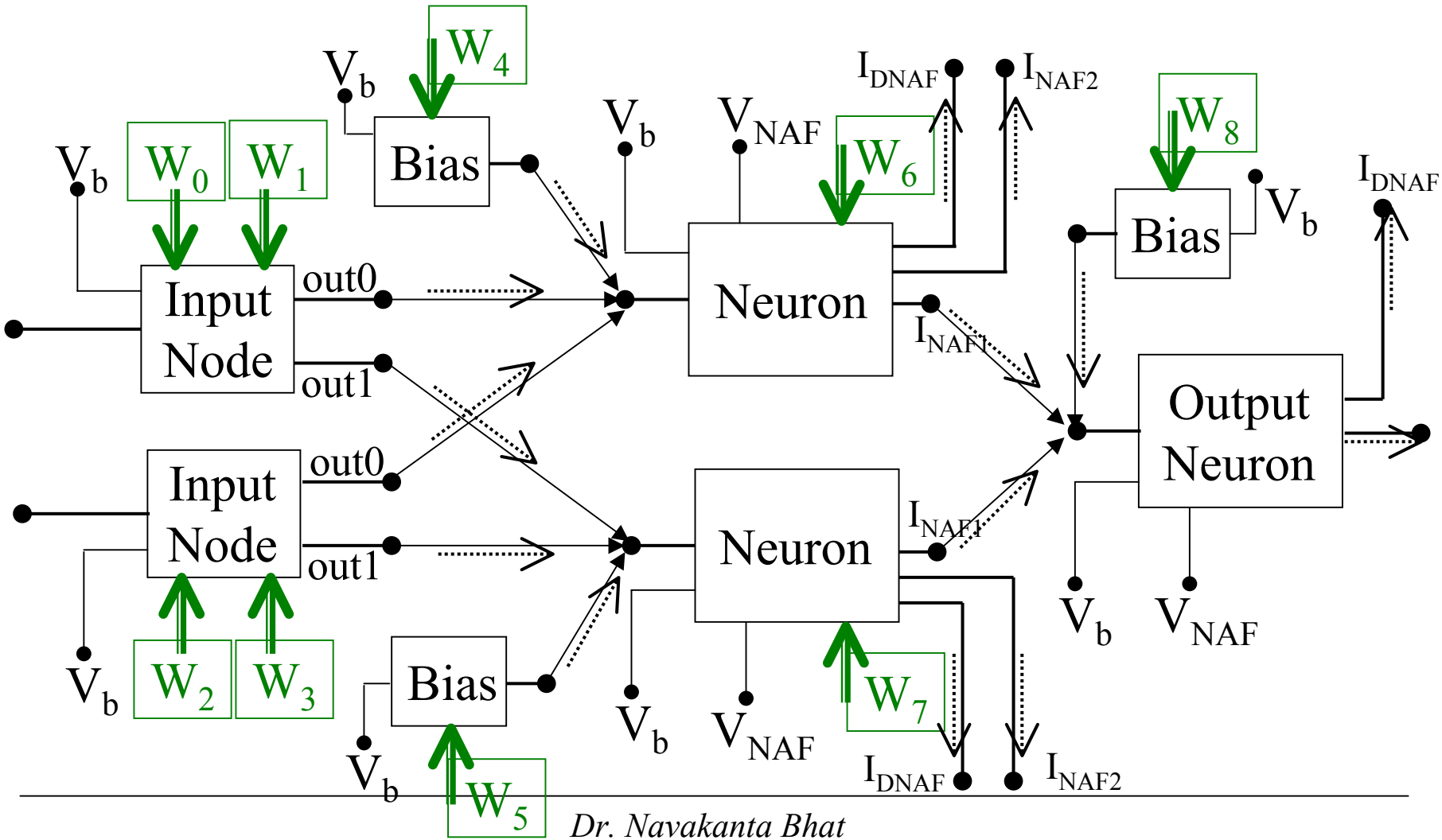
- Can accommodate for offsets & non-idealities
- Programmable Neural Network
- Flexibility in Learning Algorithm
- No learning overhead
- Serial weight update - slow learning
- PC is required for training

Feedforward Neural Network

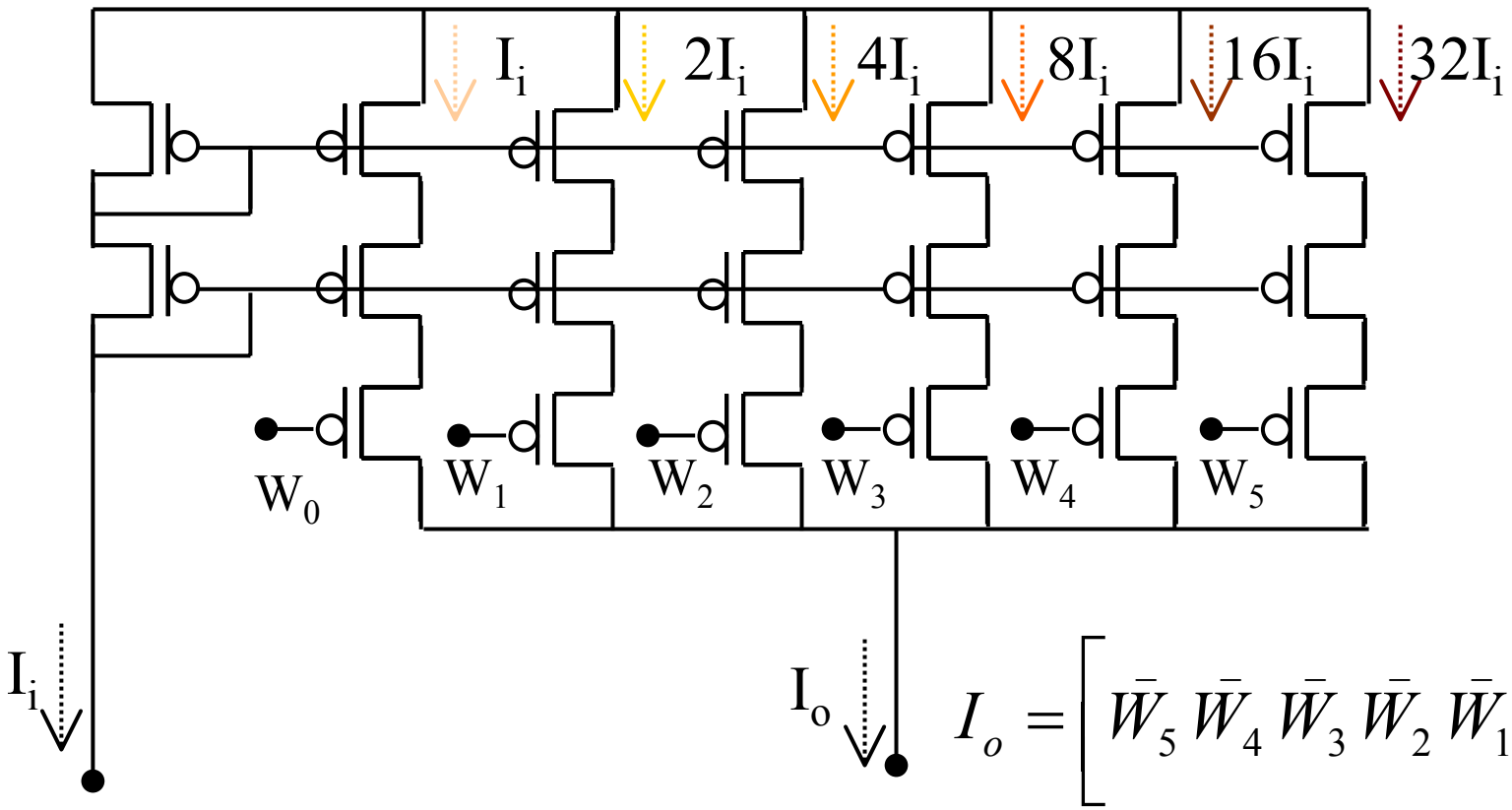


Input Layer | Hidden Layer | Output Layer

Macro-Architecture

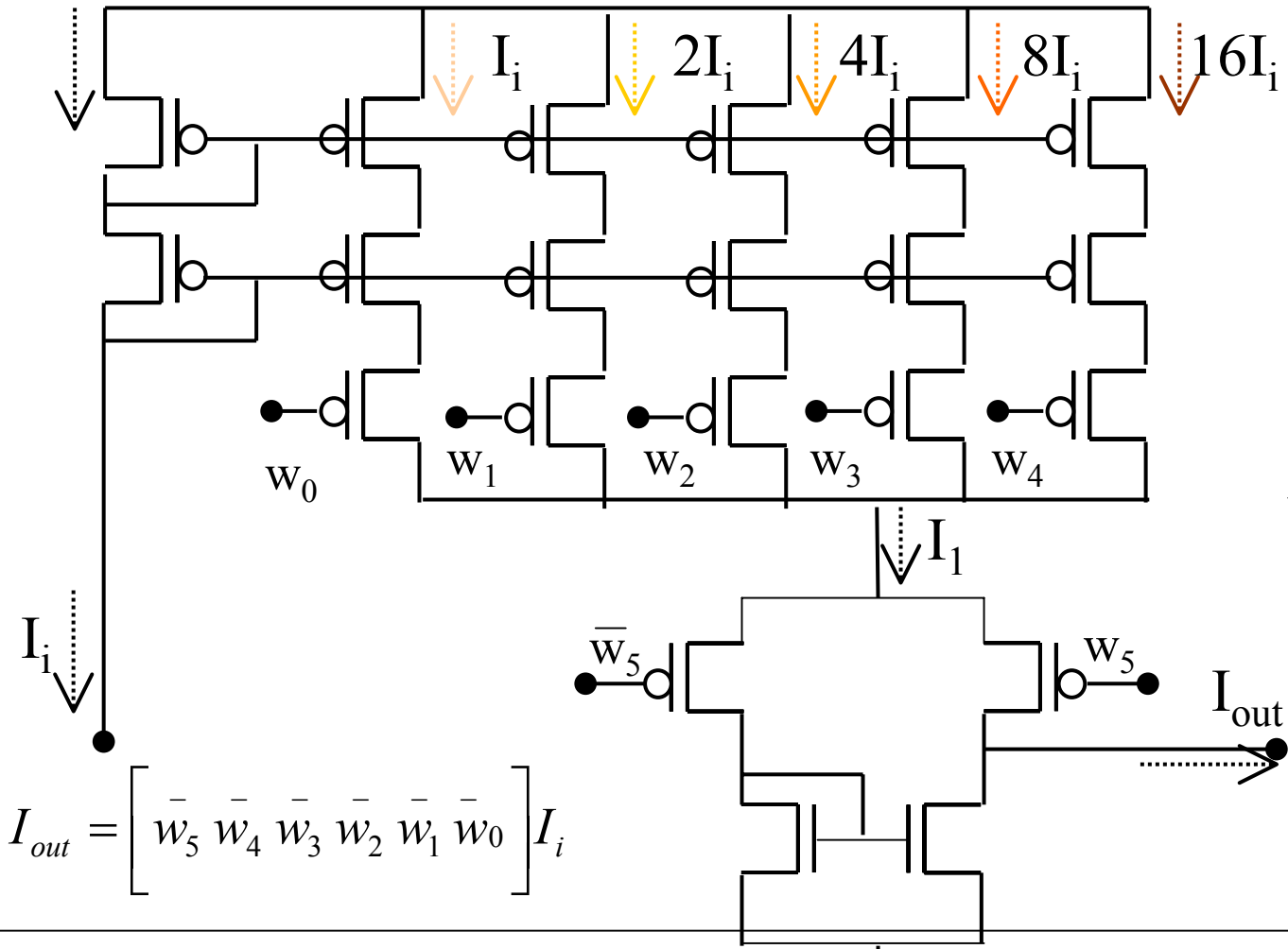


Synaptic Weights: MDAC



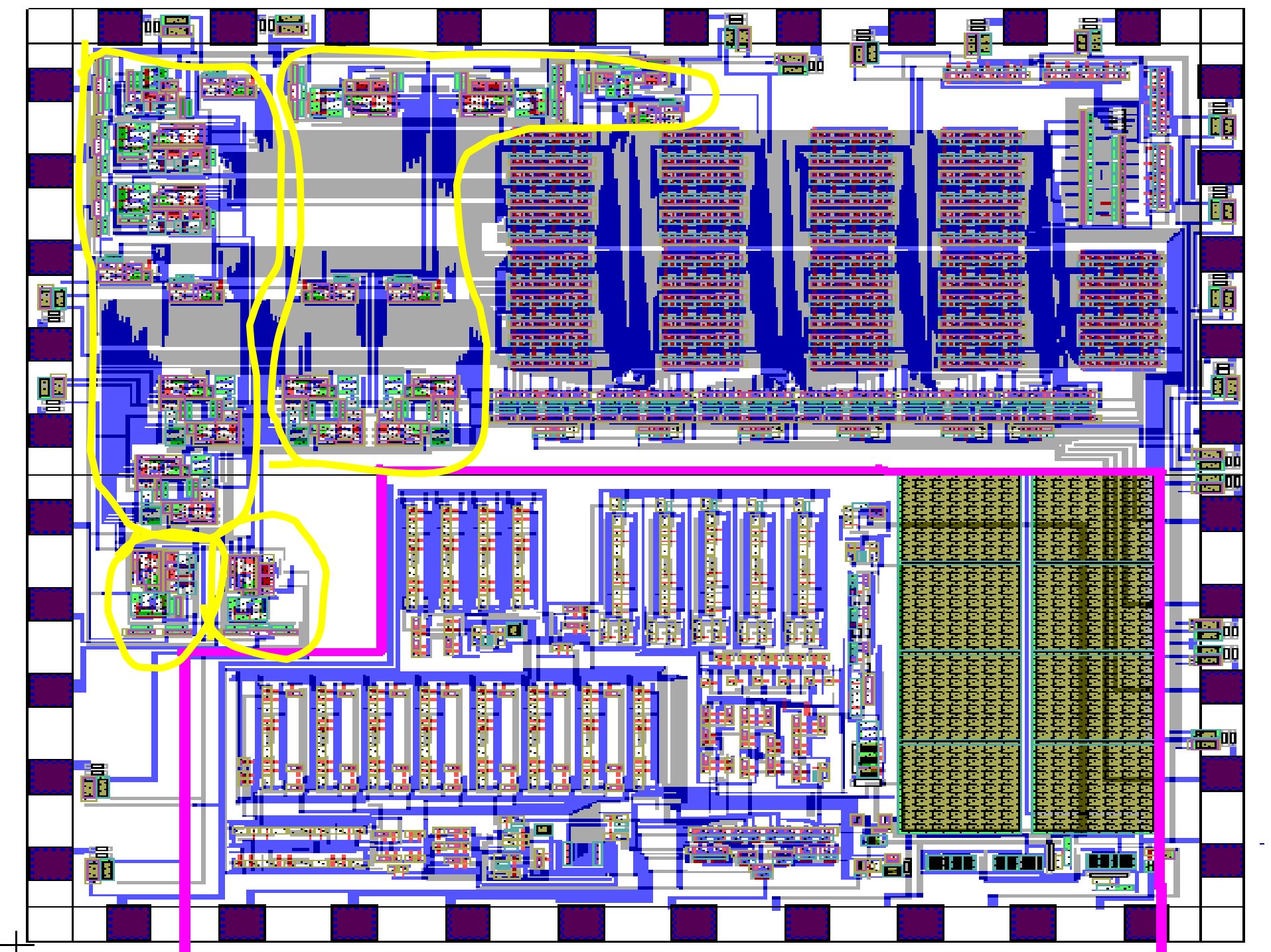
$$I_o = \left[\bar{W}_5 \bar{W}_4 \bar{W}_3 \bar{W}_2 \bar{W}_1 \bar{W}_0 \right] I_i$$

Synaptic Weights: MDAC

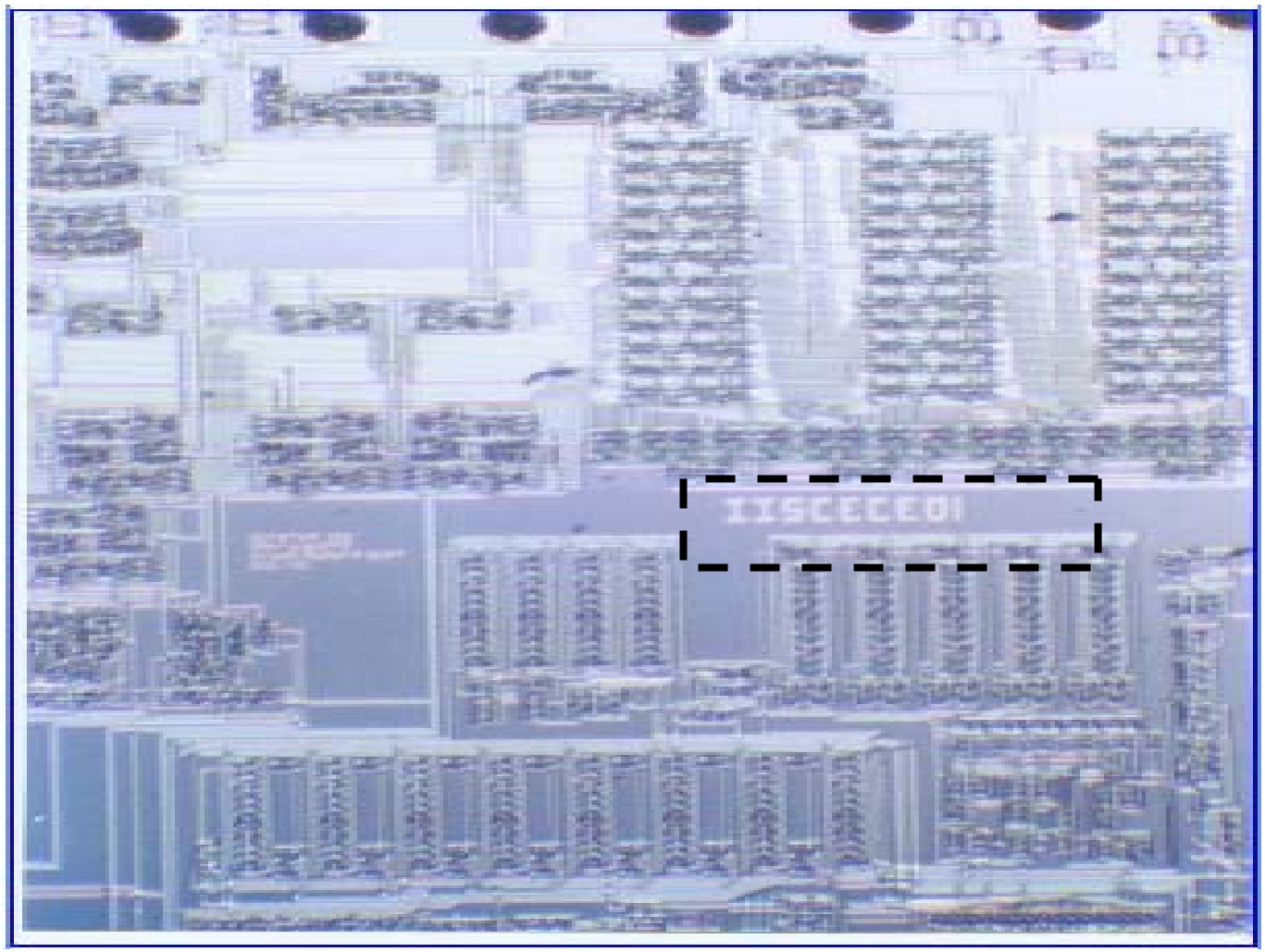


W5 is used as sign bit

$$I_{out} = \begin{bmatrix} \bar{w}_5 & \bar{w}_4 & \bar{w}_3 & \bar{w}_2 & \bar{w}_1 & \bar{w}_0 \end{bmatrix} I_i$$



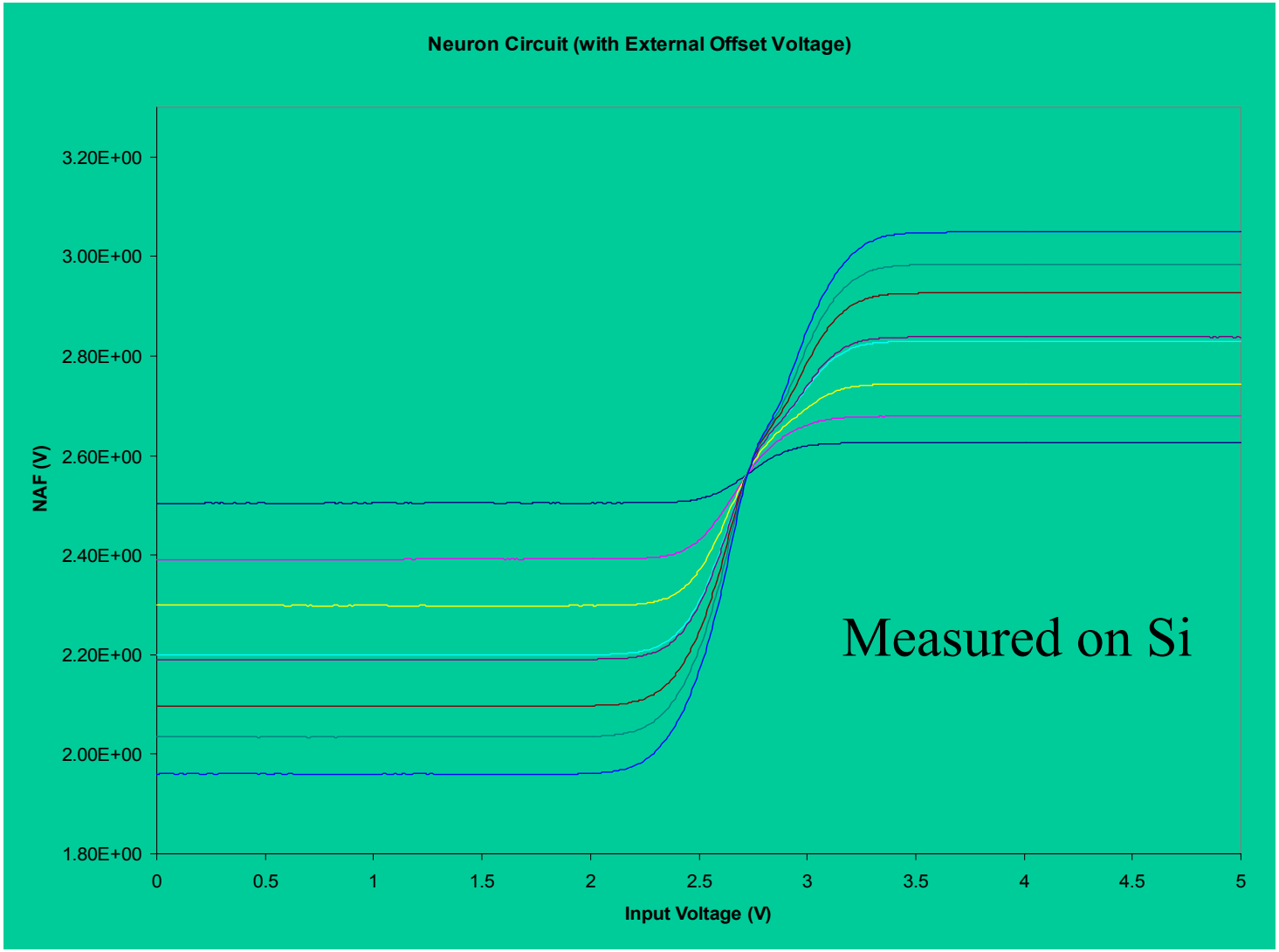
The Chip Micrograph



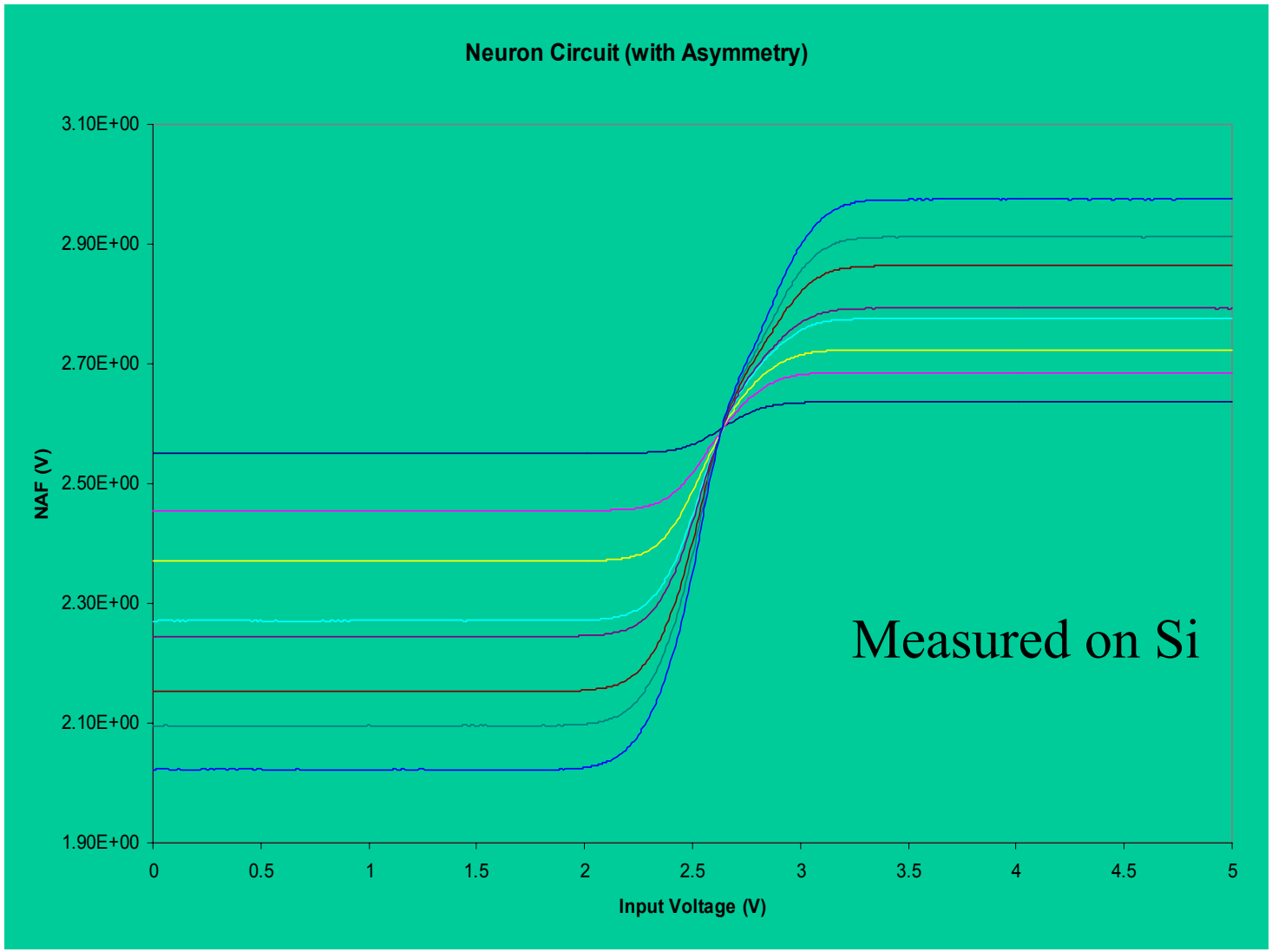
The design was prototyped through MOSIS on AMIS 1.5µm technology

Dr. Navakanta Bhat

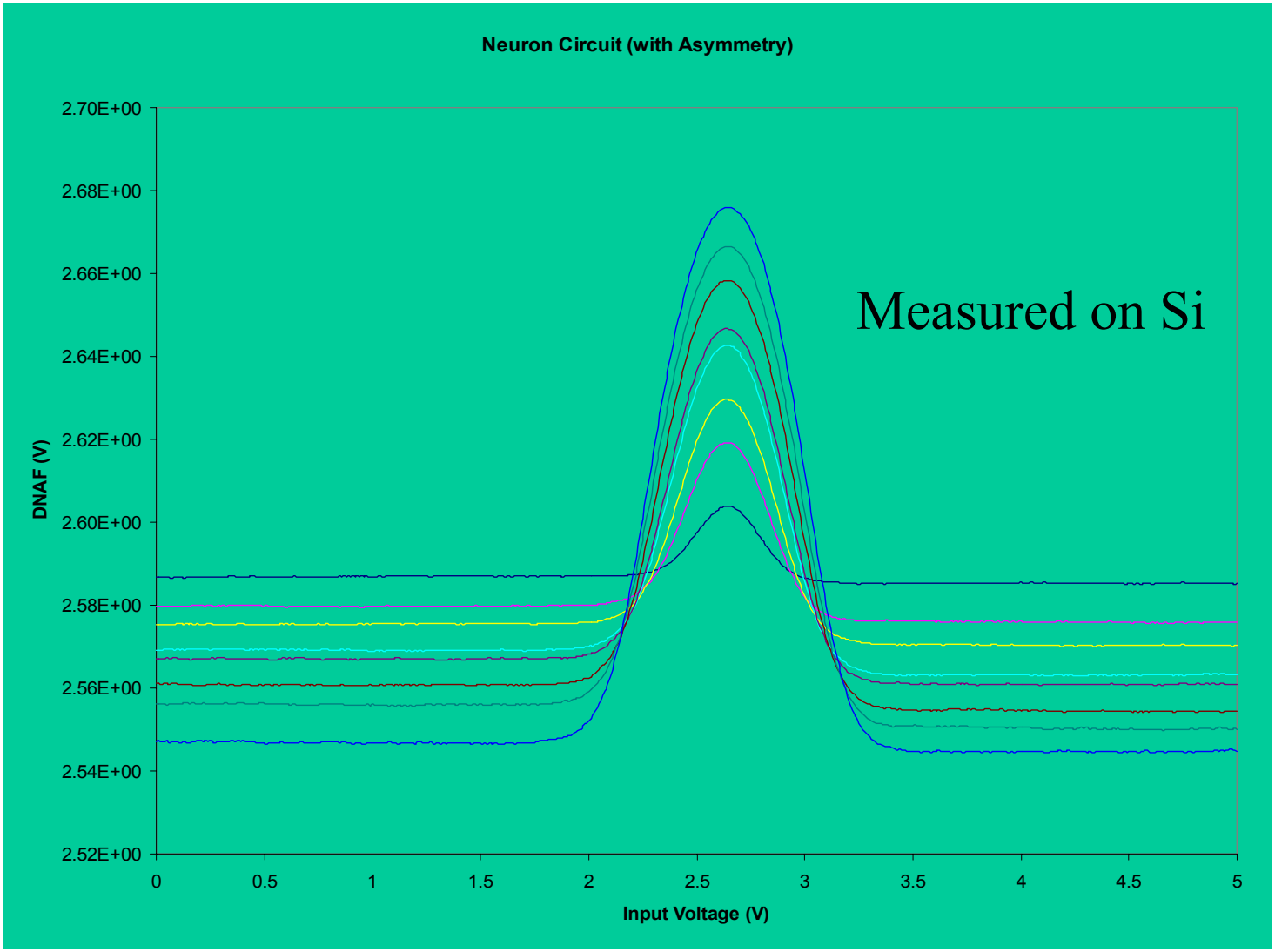
Neuron (with External Offset Voltage)



Neuron (with Asymmetry)



DNAF for Neuron with Asymmetry

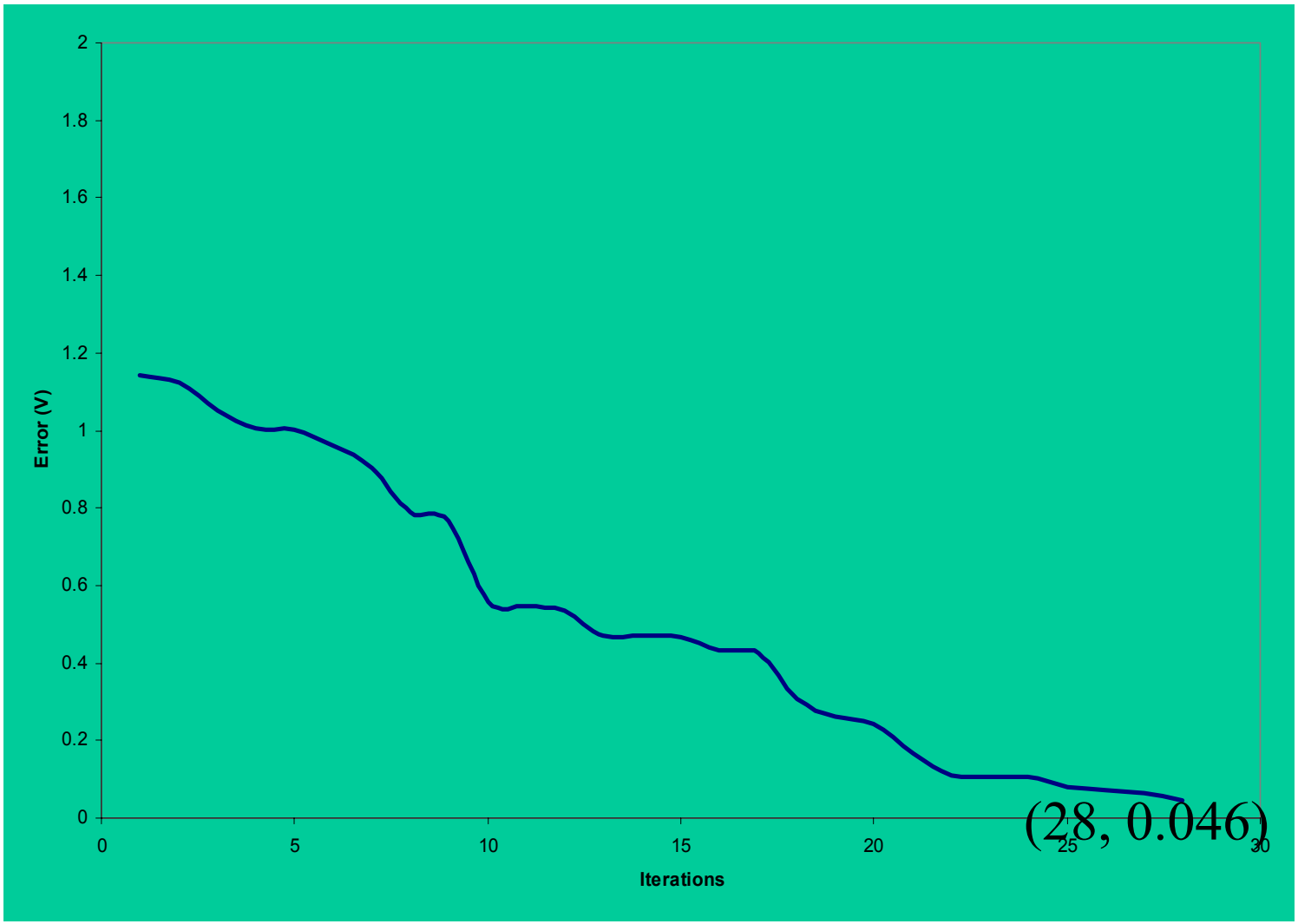


Training

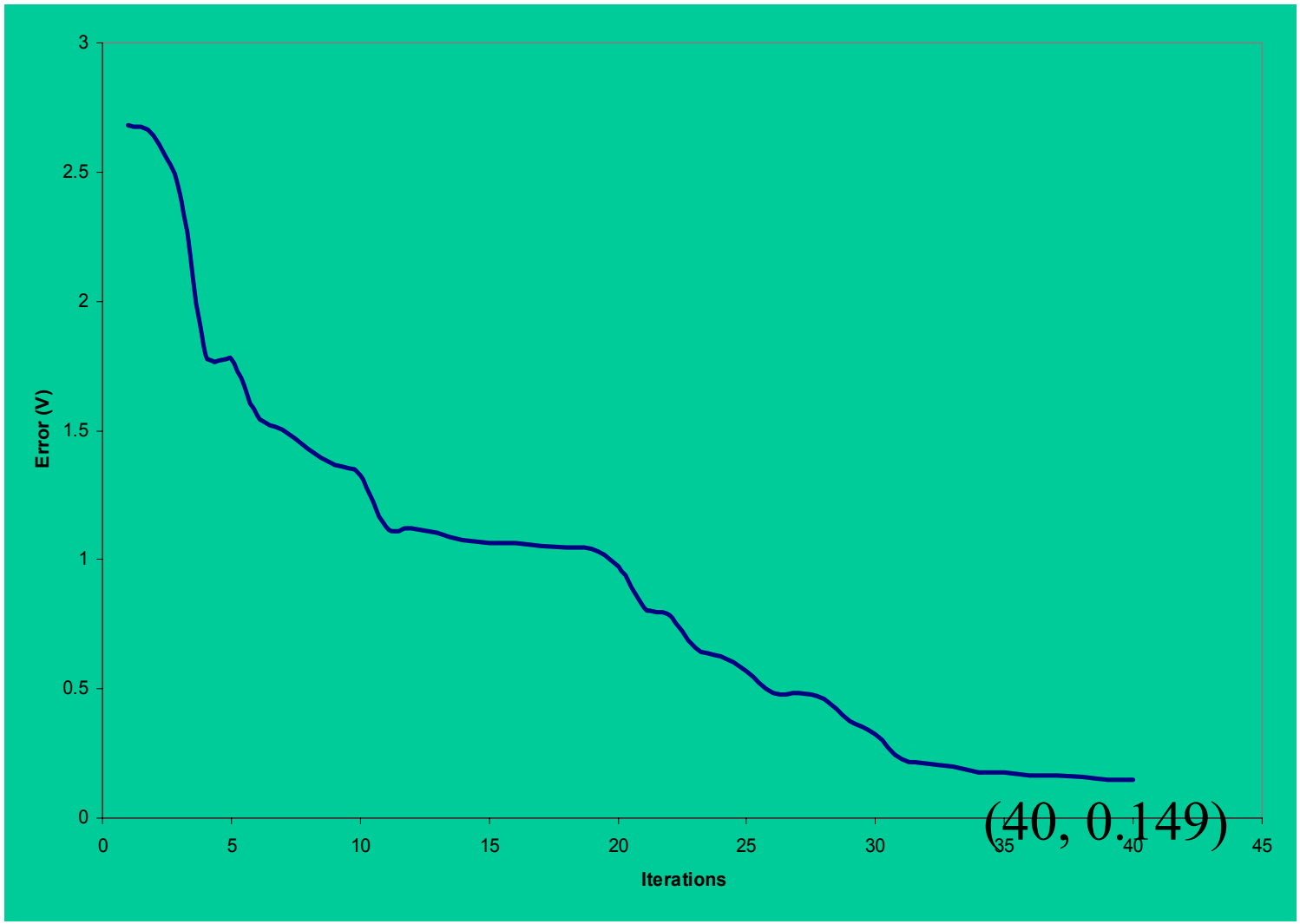
- Parallel Perturbative Algorithm
- Logic OR & Logic AND Functions
- I/P patterns: (0,0), (0,5), (5,0) and (5,5)
- O/P: Logic 0 - 1.9V; Logic 1 - 2.9V

$$Error = \sum_{i=1}^{4} |d_i - y_i|$$

Training with OR Function



Training with AND Function



Floating Gate Transistor as Analog Memory

FGMOS

The floating gate MOSFET can store charge on the FG

This forms a useful block in Neural Networks in order to update the weights in analog fashion

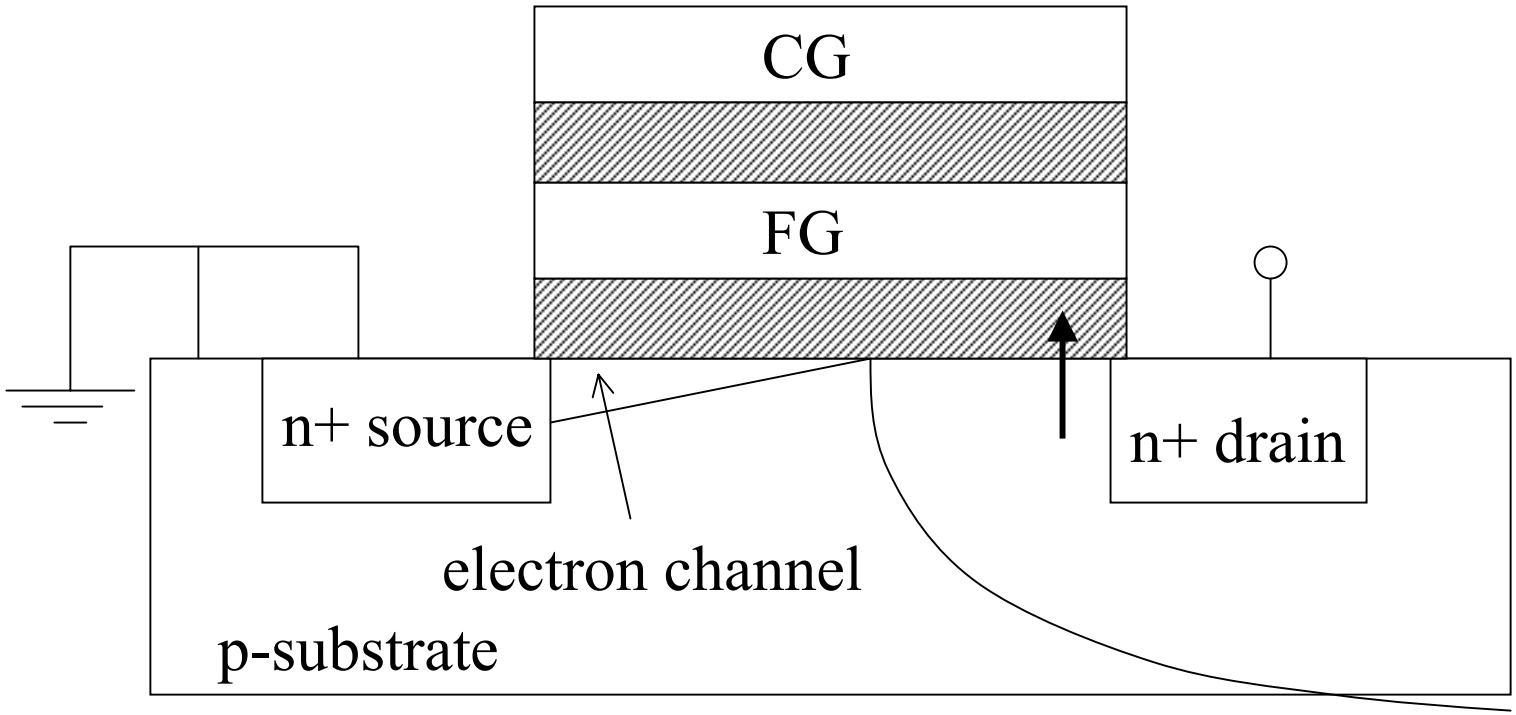
The technology has to support the double poly floating gate device

The floating gate transistor can correct itself for the of process variations

Charge storage on FGNMOS

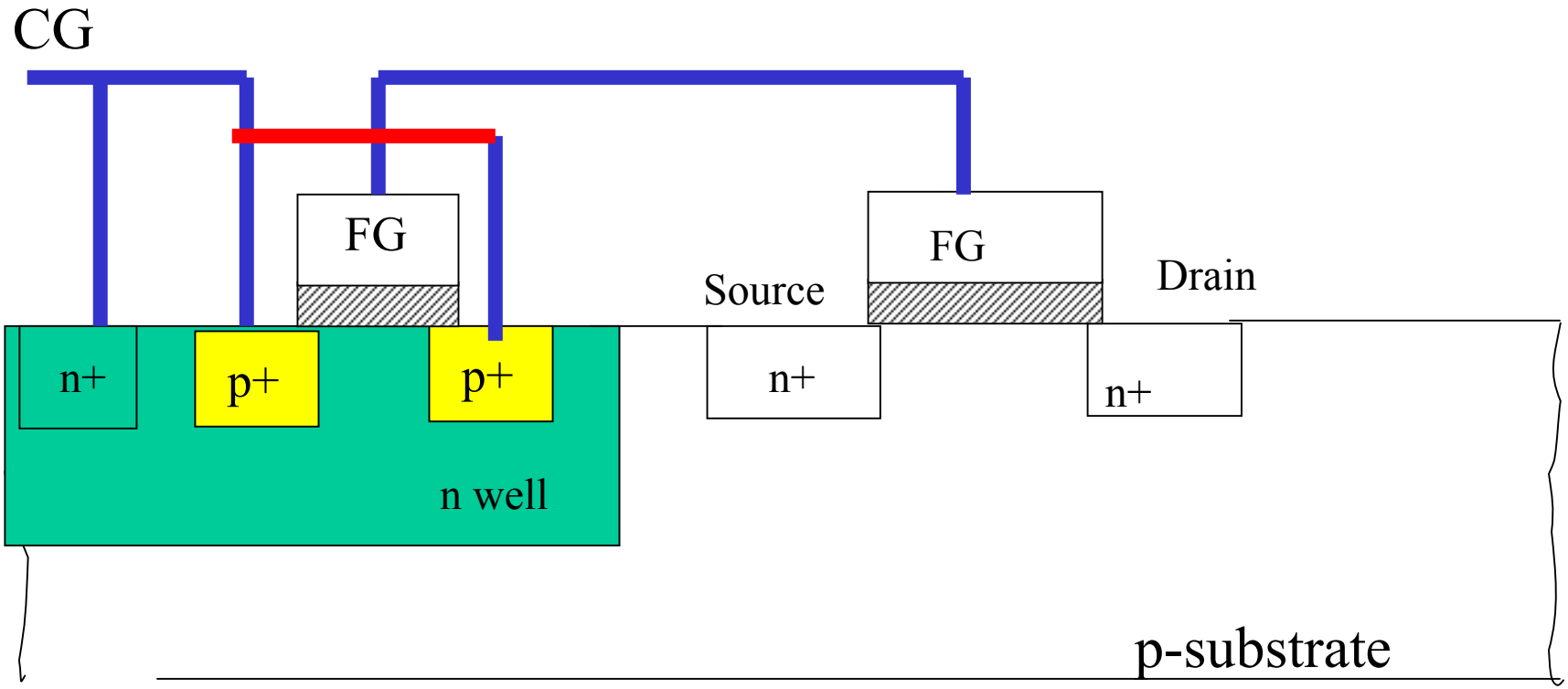
The electron charge is stored into the FG by hot carrier Injection (program), while the charge is taken out (erase) By tunneling through oxide

Program and erase are done only under high voltages



FGMOS with single poly?

There has been attempts to realize FGMOS in the conventional digital technology



This device has been reasonably successful and precision Adaptive analog circuits have been built using this