

# Progetto di circuiti integrati analogici

Alberto Tibaldi

15 agosto 2011

# Indice

<b>1</b>	<b>Introduzione alle tecnologie al silicio</b>	<b>3</b>
1.1	Passi di progetto . . . . .	4
1.2	Introduzione ai processi CMOS . . . . .	6
1.2.1	Latch-up e sottomassa . . . . .	8
1.3	Richiami sul transistore MOS . . . . .	10
1.3.1	Modelli del transistore MOS . . . . .	15
1.3.2	Proprietà dinamiche del transistore MOS . . . . .	16
1.4	Componenti passivi . . . . .	21
1.4.1	Resistori integrati . . . . .	21
1.4.2	Condensatori integrati . . . . .	26
1.5	Matching di componenti integrati . . . . .	27
<b>2</b>	<b>Stadio di amplificazione elementare MOS</b>	<b>34</b>
2.1	Analisi dello stadio elementare - polarizzazione . . . . .	34
2.2	Prestazioni dinamiche . . . . .	39
2.3	Analisi di rumore . . . . .	41
2.4	Distorsione . . . . .	43
2.4.1	Pre-post-distorsione . . . . .	43
2.4.2	Controreazione . . . . .	45
<b>3</b>	<b>Stadio di amplificazione cascode</b>	<b>48</b>
3.1	Dallo stadio a gate comune allo stadio cascode . . . . .	48
3.2	Caratteristiche statiche dello stadio cascode . . . . .	50
3.3	Risposta in frequenza . . . . .	53
3.4	Analisi del rumore . . . . .	54
3.5	Folded cascode . . . . .	55
<b>4</b>	<b>Circuiti di polarizzazione</b>	<b>58</b>
4.1	Specchi di corrente . . . . .	58
4.2	Specchio di corrente con elevata dinamica di uscita. . . . .	60
4.3	Riferimenti di corrente integrati . . . . .	63

4.3.1	Soluzione banale . . . . .	63
4.3.2	Soluzione retroazionata 1 . . . . .	65
4.3.3	Soluzione retroazionata <i>quasi definitiva</i> . . . . .	66
4.3.4	Circuito finale . . . . .	67
4.4	Riferimenti di tensione integrati . . . . .	68
4.4.1	Introduzione . . . . .	68
4.4.2	Band-gap di Kuijk . . . . .	71
4.4.3	Band-gap di Widlar . . . . .	72
4.4.4	Implementazione a MOS . . . . .	73
4.4.5	Implementazione per basse tensioni di alimentazione . . . . .	73
<b>5</b>	<b>Amplificatori ad alto guadagno</b>	<b>75</b>
5.1	Amplificatore differenziale . . . . .	75
5.1.1	Dinamica di uscita del modo differenziale . . . . .	79
5.1.2	Dinamica di ingresso di modo comune . . . . .	81
5.2	Stadio differenziale telescopico (cascode) . . . . .	82
5.3	Stadio completamente differenziale come cascata di stadi differenziali elementari . . . . .	83
5.3.1	Feedback di modo comune - reprise . . . . .	87
5.4	Stadi folded cascode - rail-to-rail . . . . .	88
5.4.1	Dinamica di uscita di modo differenziale . . . . .	88
5.4.2	Dinamica di ingresso di modo comune . . . . .	89
5.5	Amplificatori single-ended - parametri . . . . .	93
5.5.1	Amplificazione di modo comune . . . . .	94
5.5.2	PSRR: Power Supply Rejection Ratio . . . . .	97
5.5.3	Offset . . . . .	100
5.5.4	Principali soluzioni per eliminare l'offset - circuiti tempo-discreti . . . . .	106
5.6	Risposta in frequenza - compensazione . . . . .	109

# Capitolo 1

## Introduzione alle tecnologie al silicio

Il taglio che si intende attribuire a questa trattazione è parzialmente tecnologico, parzialmente circuitale, incentrato sui circuiti di tipo integrato: da diversi decenni, infatti, i circuiti analogici si realizzano come parte di un sistema integrato. Ci occuperemo dunque, in questa trattazione, soprattutto della parte analogica, dunque di segnali fundamentalmente in banda base.

Su questa fetta di silicio, che può avere dimensioni variabili dai 2 ai 5 millimetri, si ha grosso modo a che fare, per il 70-80%, con elettronica digitale. La rimanente parte del sistema è dedicata un po' a una parte a RF, e un po' all'analogica. I rimanenti elementi sul silicio sono sostanzialmente *pad*: si tratta di piazzole, più circuiti che permettono di pilotare ciò che è stato integrato sul silicio. Questo è un tipico esempio di SoC : System on Chip.

Oltre che di SoC si parla di SiP: System in Package; si tratta di sistemi in cui si ha a che fare con più die, collegati tra loro mediante dei bond che permettono la connessione elettrica dei vari die.

Fondamentale è parlare di *area*: questi circuiti sono infatti realizzati mediante processi CMOS (comunemente si utilizzano processi a 90 nanometri, anche se sono in sviluppo processi a 27 e 45 nanometri). Il processo tecnologico da utilizzare è sostanzialmente deciso da chi occupa la maggior parte di area del silicio: dal momento che i circuiti digitali sono molto estesi, sono sostanzialmente i progettisti digitali a stabilire cosa utilizzare, e questo è il CMOS; non a caso, infatti, il grosso del mercato è regolato dai CMOS, dal momento che massimizzano le prestazioni della parte digitale del circuito (per quanto per l'analogica non siano la scelta migliore).

Perchè integriamo? Beh, sostanzialmente, più si integra, più si riducono i costi: se si ha infatti la possibilità di lanciare sul mercato un sistema integrato, più si riesce a caricarvi sopra componenti e più esso sarà performante,

ma d'altra parte più esso sarà *piccolo*: integrare, riduce i costi. Un sistema integrato viene fabbricato per volumi di produzione per 10000 pezzi all'anno; l'acquisto può costare qualche dollaro (pezzo venduto in volume); il 40% del costo del dispositivo è dovuto all'incapsulamento, e il costo è legato all'area occupata, ma **non** a ciò che si mette dentro: i due elementi che determinano il costo sono l'**area** e il **numero di maschere** necessarie per la fabbricazione. Un processo da 350 nanometri, in uso 10-15 anni fa, è meno costoso di uno da 90 nm.

## 1.1 Passi di progetto

In fase di progetto, si ha sostanzialmente a che fare con due tipi di errore:

- di progetto (colpa sostanzialmente dei simulatori, o di un cattivo progettista);
- di layout: di fatto ci sono problemi che, anche con un buon progetto, non si possono vedere prima di aver ottenuto il layout.

Si studi a questo punto il diagramma di flusso per il progetto di un dispositivo, di un sistema integrato analogico; esso verrà proposto per punti.

1. Determinazione delle specifiche funzionali: esse devono essere implementate dal sistema integrato; si parla di specifiche di tipo naturalmente elettrico, ma anche meccanico, termico, di invecchiamento.
2. A questo punto, si parla della scelta del processo da utilizzare; vi sono, dunque, due possibilità:
  - il processo può essere scelto dal progettista analogico (cosa che capita molto raramente);
  - il processo è scelto dal progettista digitale (ciò che capita di solito); in questo caso, si ha un secondo sotto-step, in cui si definisce il **design kit**, ossia il kit dei componenti che possono essere utilizzati sul progetto.

Può capitare che la scelta del processo sia influenzata dalle specifiche richieste: se per esempio ci servisse una dinamica di 5 volt, anche se il processo non lo consente, è necessario effettuare delle modifiche, in modo da ottenere delle uscite che permettano valori di tensione più elevati.

3. Scelto il processo, si passa alla fase di **progetto**: si tratta sostanzialmente della scelta della topologia circuitale, dato l'insieme di specifiche e il processo. Progettare significa, sostanzialmente, dimensionare i componenti; si consideri questo semplice esempio:  
Una volta scelta la topologia (primo passo), è necessario dimensionare  $W$  e  $l$ , e scegliere il tipo di resistenza da integrare.
4. Progetto del layout: si progetta il piazzamento di ogni singolo componente; questa è un'operazione che, nei circuiti analogici, va effettuata **manualmente**; si effettua dunque, alla fine, un controllo di conformità tra progetto e layout, mediante tool automatici.
5. Tape-out : si rilasciano le maschere per la fabbricazione dell'integrato; questo è un punto molto importante, dal momento che, a 90 nanometri, un set di maschere costa 300-400 mila euro, dunque è necessario essere particolarmente sicuri di ciò che si è fatto prima di questo passo, dal momento che rifare le maschere è gravissimo.
6. Una volta uscite le maschere, si passa alla fabbricazione (tra il passo precedente e questo ci sono dall'1 ai 3 mesi, necessari per avere il silicio).
7. Test del processo: sulla stessa fetta dell'integrato, si producono delle zone dove inserire i transistori del design kit; ciò che si vuole sostanzialmente fare è caratterizzare i componenti del processo, per vedere se tutto è andato a buon fine. Si verifica sostanzialmente dunque che i componenti siano realizzati secondo i parametri desiderati.
8. Test funzionale: si verifica il comportamento, a livello **funzionale**, del circuito.
9. Si realizza il package del sistema.
10. Test meccanici e chimici: una volta ottenuto il package e verificate le proprietà elettriche e funzionali, si verificano quelle meccaniche e chimiche, mediante attacchi di tipo chimico o prove con vibrazioni.
11. Si effettuano le prove di invecchiamento, ovviamente utilizzando processi di invecchiamento rapido, indotto da cicli termici e chimici.
12. Si passa, infine, alla fase di produzione.

La durata di questo processo richiede da 1 a 2 anni circa. Si ricordi che, se il processo va male, si perdono sostanzialmente i mesi dell'attesa del

silicio, e di questi servono mediamente due cicli. Ciò che si deve cercar di non sbagliare assolutamente è il progetto: sbagliare il progetto, come già detto, è gravissimo, a causa degli elevatissimi costi da affrontare. Per questo motivo, i CAD si sono notevolmente sviluppati: per cercare di controllare il progetto e il progettista.

## 1.2 Introduzione ai processi CMOS

La quasi totalità dei processi CMOS è realizzata utilizzando un **isolamento a giunzione**. Cosa significa? Perché si parla di ciò? Cerchiamo di presentare il problema: vogliamo realizzare per esempio un diodo su di un substrato di silicio. Focalizziamoci meglio sul problema: consideriamo un substrato di silicio di tipo  $p$ , spessore del substrato che può variare dai 150 ai 500 micron; sopra di esso, si costruisce uno strato epitassiale, ancora tipo  $p$  (si noti che potrebbe anche essere tipo  $n$ ), su cui costruiamo i circuiti.

Sulla superficie della fetta vengono diffuse o impiantate regioni con drogaggio differente. Volendo realizzare un diodo, per esempio, si può diffondere o impiantare, sullo strato epitassiale (abbiamo detto tipo  $p$ ), una sacca tipo  $n$ , di 3 o 4 micron (per esempio), all'interno della quale si produce un'altra sacca, ancora più piccola, drogata  $p^+$ , ossia fortemente drogata tipo  $p$ . Fatto ciò, si ricopre la superficie di ossido, si mettono i contatti metallici, e questo è, di fatto, il circuito.

Tutto finito? Beh, non abbiamo ancora motivato l'isolamento: ogni volta che piazziamo un diodo, ne viene fuori *gratuitamente* un altro: ogni volta che si impianta una sacca tipo  $n$  su uno strato tipo  $p$ , si forma di fatto un altro diodo; questo è un effetto assolutamente indesiderato, dal momento che introduce elementi parassiti. Volendo rappresentare questi effetti in termini circuitali, si ottiene qualcosa di questo genere:

I diodi sono back-to-back, dunque ci verrebbe da dire che non causano problemi; purtroppo non è così: se non sono isolati, potrebbe intervenire l'effetto transistor, e fare entrare in conduzione il sistema, ottenendo enormi perdite.

Ciò che di solito si fa, è il seguente stratagemma:

si realizza un contatto di superficie,  $p^+$ , che si collega al potenziale più basso (nel caso lo strato epitassiale sia di tipo  $p$ ; se lo strato epitassiale è tipo  $n$ , si introduce un contatto di superficie tipo  $n^+$  e lo si collega al potenziale più alto), in modo da polarizzare inversamente tutte le giunzioni parassite. Questo, è il metodo di isolamento a giunzione.

Questi processi con isolamento a giunzione hanno dei limiti: se si considerano queste giunzioni parassite, esse vengono polarizzate in modo da risultare circuiti aperti; in realtà non capita ciò, dal momento che noi sappiamo che:

$$I_D = I_S \left( e^{\frac{V_D}{nV_T}} - 1 \right)$$

Con l'isolamento, si vuole  $V_D < V_\gamma$ , e dunque si ha che  $I_D \sim -I_S$ ; finchè di diodo parassita ce n'è uno solo, pazienza; il brutto è che ogni componente introdotto sul circuito integrato, di fatto, introduce giunzioni, dunque diodi parassiti, che incrementano il leakage. Cerchiamo di quantificare questo leakage: in generale si può scrivere che:

$$I_S = \frac{qA_j n_i^2 D_n}{N_A L_n} = \frac{qA_j}{L_n N_A} n_i^2(T) \frac{k_B T}{q} \mu_n(T)$$

Ma noi sappiamo, dalla teoria di Dispositivi Elettronici, che:

$$n_i^2(T) = DT^3 e^{-\frac{V_{bi}}{V_T}}$$

Dove  $V_{bi}$  è la tensione di built-in della giunzione. In risultante:

$$I_S = ET^{4-n} e^{-\frac{V_{bi}}{V_T}}$$

Questo significa che la corrente di saturazione inversa,  $I_S$ , è fortemente dipendente dalla temperatura  $T$  (dal momento che è presente esplicitamente, nonchè nell'esponenziale).  $n$  è un coefficiente circa uguale a 1,5:

al crescere di  $t$ , l'argomento dell'esponenziale si riduce, ma la corrente dunque sale.

Si consideri un esempio numerico: data  $I_S = 1$  pA a 27 gradi celsius, a 120 gradi diventa pari a 100 pA, e a 200 gradi a 92 microampere: alzando la temperatura, i circuiti di superficie non fanno più il proprio mestiere, poichè il leakage aumenta notevolmente.

Nei circuiti che studieremo, si avrà che:

$$\overline{I_D} = 1 \div 100 \mu A$$

Dall'alto, il layout può avere questo aspetto:

Se tagliamo nella sezione A-A, possiamo vedere qualcosa del genere:

Si hanno due giunzioni: a sinistra tra il source e il substrato, dunque, nel circuito, una giunzione è collegata tra 0 V e 0 V, ma l'altra è collegata tra drain e 0 V. Abbiamo quantificato prima quanta corrente può passare in uno di questi *diodi* parassiti, e abbiamo detto quanta corrente può servire per definire il punto di lavoro del transistor; la verità è che, a 200 gradi, la

corrente sul diodo è molto simile a quella sul drain del transistor MOS, dunque, di fatto, il transistor non funziona! Ciò che modifica il comportamento del transistor è, di fatto, la temperatura! Per questo motivo risulta fondamentale mettere dei case, dei dissipatori: togliendo temperatura, si riducono le correnti di leakage, e così si migliora il comportamento del sistema.

Esistono sul mercato processi anche detti SoI: dato un substrato di silicio, vi si fa crescere sopra dell'ossido, che separa i vari componenti; elettricamente, sembra una soluzione ottima, ma da un lato è molto costosa, e dall'altro l'ossido, oltre a essere un ottimo isolante elettrico, è anche un ottimo isolante termico, e noi abbiamo bisogno di conduttori termici, dal momento che abbiamo bisogno di togliere il calore, non di mantenerlo lì!

Come mai il substrato è di 500 micron? Beh, sostanzialmente, le dimensioni sono scelte al fine di prevenire stress di tipo meccanico: di solito infatti il substrato viene incollato a uno slug metallico, e a causa di variazioni termiche si ha il rischio di generare microfratture, le quali distruggerebbero il dispositivo.

### 1.2.1 Latch-up e sottomassa

Il problema delle correnti parassite, può causare altre sotto-problematiche; si consideri per esempio un inverter CMOS:

Per realizzare l'isolamento servirebbe una sacca  $n$  con dentro i terminali del pMOS; per isolare la sacca  $n$  al solito si fa un contatto  $n^+$  e lo si collega al potenziale  $V_{DD}$  (in questo caso, dunque, il più alto); con un semplice inverter abbiamo molte, moltissime giunzioni; ciò che capita generalmente ogni volta che si piazza un MOS, è piazzare un BJT parassita! Ognuna di queste sacche con terminali all'interno, infatti, riesce a realizzare di questi circuiti. Si considerino le resistenze del substrato e della sacca su cui si impianta il pMOS, e l'effetto transistor; si avrà qualcosa del genere:

Se proviamo a rappresentare il circuito risultante dall'analisi di questi parametri parassiti, si trova qualcosa di questo genere:

Se per qualche motivo l'alimentazione va sotto l'uscita, il transistor parassita 1 ha una corrente che cade sulla resistenza, accendendo il transistor 2, provocando una reazione a catena che rovina il comportamento del circuito. Questo fenomeno è detto latch-up. Il substrato viene drogato fortemente (nel nostro caso dunque  $p^+$ ), dal momento che se ne vuole ridurre la resistenza, in modo da ridurre il fenomeno del latch-up; per questo motivo, generalmente, il substrato, nei processi comunemente usati, è molto drogato.

Il problema appena presentato fa parte di un problema ben più vasto:

Il substrato è collegato al potenziale più basso, al fine di isolare le varie parti del circuito; supponendo di avere, nel nostro circuito integrato, 100000

sacche di tipo  $n$ , per integrare diversi dispositivi, se una sola di queste sacche viene collegata a un potenziale più basso del substrato (per un qualsiasi motivo), viene acceso un transistor con un emettitore e 99999 collettori, che condurrà, introducendo un leakage enorme; questo è il **problema del sottomassa**. Questo transistor parassita avrà dei parametri molto bassi, per esempio possiamo supporre che  $\beta$  valga  $10^{-4}$ , ma se si fa passare qualche milliamperere nel circuito esteso, nei collettori parassiti passa qualche microampere, ottenendo di fatto leakage dello stesso ordine di grandezza dei segnali; ciò può provocare malfunzionamento o addirittura distruzione del circuito.

La presenza di questi diodi parassiti introduce leakage, ma non solo: questi diodi sono sostanzialmente legati a effetti di tipo capacitivo: finché tutti questi elementi parassiti sono collegati in modo da non influenzare il comportamento del segnale non abbiamo problemi, ma la verità è che potrebbero benissimo anche introdurre limitazioni di banda; ciascuna giunzione, infatti, introduce una capacità del tipo:

$$C_j = \frac{C_{j0}}{\sqrt[m]{1 + \frac{V_j}{\phi_0}}}$$

Ciò che risulta evidente è che  $C_j$  è una capacità non lineare: essa dipende infatti dalla tensione applicata ai suoi capi. Ciò che possiamo fare è linearizzare questa capacità: dato un certo punto di lavoro, possiamo calcolare la retta tangente mediante l'operazione di derivazione, per poi utilizzare la capacità risultante quando si sta calcolando il modello di piccolo segnale.

Il nostro obiettivo è avere una dinamica molto elevata in uscita, dunque durante il segnale si potrebbero sentire gli effetti di non linearità della capacità; per risolvere ciò è possibile o essere sicuri del comportamento del dispositivo mediante simulazioni, oppure calcolare una capacità efficace, equivalente, definibile per esempio nel seguente modo:

$$C_{j,eff} = \frac{Q(V_2) - Q(V_1)}{V_{j2} - V_{j1}}$$

Di solito, in pratica, ciò che si fa è considerare, come capacità efficace:

$$C_{j,eff} = \frac{C_{j0}}{2}$$

$C_{j0}$  è uno dei parametri noti nel modello del transistor (si tratta della massima capacità possibile, per unità di area).

La capacità finale dipende sostanzialmente dalla lunghezza del transistor ( $L$ ), dall'ampiezza delle regioni di diffusione (in cui si introduce il drogante), e dalla larghezza  $W$  del dispositivo:

$$C_{DB} = A_D \frac{C_{j0}}{2} = L_{CD} W \frac{C_{j0}}{2}$$

Dove  $L_{CD}$  è la larghezza delle regioni di diffusione del transistor (la larghezza di drain).

La presenza di queste capacità parassite può essere importante sulle prestazioni in frequenza del circuito; al fine di ridurle, è possibile cambiare il processo, utilizzando un isolamento a ossido (il già citato SoI):

Se al posto dell'isolamento a giunzione si utilizza un isolamento a ossido, si introduce sostanzialmente un ossido spesso, in modo da ridurre le capacità tra le varie sacche e il bulk; i tranches (distanza tra una sacca e l'altra, riempita di ossido) è sostanzialmente decisa dal processo, mentre si sceglie, per lo spessore dell'ossido rispetto allo strato sotto (al substrato), qualche micron.

Si consideri per esempio la realizzazione di un inverter con isolamento a ossido:

Di solito l' $n$ -well si fa profondo tanto quanto le pozze  $p$  (e viceversa), in maniera da ridurre le capacità interne al well; le capacità verso il bulk, come già detto, vengono invece ridotte mediante lo spessore dell'ossido.

### 1.3 Richiami sul transistor MOS

Prima di introdurre dei veri e propri concetti sulle topologie da utilizzare per il progetto dei circuiti integrati, è necessario effettuare alcuni richiami riguardo al transistor MOS. Si farà fondamentalmente riferimento a processi con isolamento a giunzione, ottimizzati per i sistemi digitali (essendo quelli che più frequentemente capita di utilizzare).

Avremo a che fare fondamentalmente con due tipi di realizzazione per i transistori:

In un caso, il body risulta essere vincolato al potenziale più basso del circuito, ottenendo, di fatto, solo tre terminali liberi. Nel secondo caso, in cui si ha a che fare con un **doppio isolamento**, si hanno cinque terminali, di cui uno occupato e quattro liberi: in questo modo, dunque, si ha un terminale in più libero rispetto al caso precedente (dal momento che il body non va vincolato al potenziale di riferimento):

Analizziamo più nel dettaglio il caso di transistor a doppio isolamento:

esso, di fatto, presenta diverse giunzioni parassite. Si ha a che fare sostanzialmente con un diodo (parassita) tra body e drain,  $D_3$ , uno tra body e source,  $D_4$ , uno tra body e ISO ( $D_1$ ), uno tra ISO e substrato ( $D_2$ ); ISO è un terminale di isolamento, da qua il nome (ed è il quinto terminale, che

noi non avremo accessibile). Il body si trova a una tensione  $V_B$ , la quale deve per forza essere minore di  $V_D$  e di  $V_S$ ; se collego a  $V_{DD}$ , dunque alla tensione di polarizzazione, il terminale ISO, in un colpo solo polarizzo inversamente sia  $D_1$  sia  $D_2$ . Se poi collego il corpo del transistor (il body) al source, ottengo un totale isolamento: il source del transistor avrà tensione bassa, dal momento che è collegato al body (il quale è a tensione bassa);  $V_D$  è senza dubbio positiva rispetto a  $V_B$ , dunque tutti i diodi, in questo modo, sono polarizzati inversamente (e l'isolamento a giunzione, di fatto, realizzato). Questo è ciò che capita quando si ha il classico simbolo del MOSFET: in realtà, volendo sviluppare il modello in un passo successivo, si dovrebbe vedere l'ultimo disegno presentato.

Ora, date le due tecnologie possibili, si consideri il seguente circuito:

Cosa cambia in questi circuiti? Beh, il comportamento è senza dubbio analogo, ma una differenza in effetti c'è: la tensione di soglia:

- nel caso del circuito realizzato con il transistor direttamente costruito su strato epitassiale (con dunque il body collegato direttamente al potenziale di riferimento), si ha che:

$$V_{th} = V_{th0} + \gamma \left( \sqrt{V_{SB} + |2\phi_F|} - \sqrt{2\psi_G} \right)$$

ossia, si deve tenere conto dell'effetto body.

- nel caso del circuito realizzato con il transistor a doppio isolamento, si ha:

$$V_{th} = V_{th0}$$

La grandezza che dunque sarà fondamentalmente influenzata, nel circuito, sarà la tensione sul source! Infatti:

$$V_{S1} = V_{in} - V_{GS1}$$

$$V_{S2} = V_{in} - V_{GS2}$$

Studiamo a questo punto qualche altro parametro: si sa che:

$$V_{GS} = V_{th} + V_{OD}$$

Dove  $V_{OD}$  è detta **tensione di overdrive**, ed è calcolabile come:

$$V_{OD} = \sqrt{\frac{2I_D}{\mu_n C_{OX} \frac{W}{L}}}$$

La corrente di drain diminuisce fino a quando non va a 0; la tensione per cui capita ciò è la tensione di overdrive: quando la tensione  $V_x$  diventa tale da spegnere il transistor, non si ha più corrente.

Per un transistor in stato di saturazione, si sa che:

$$i_D = \beta_n \frac{W}{L} V_{OD}^2 (1 + \lambda V_{DS})$$

Questa espressione è valida, a patto che:

$$\begin{cases} V_{GS} > V_{TH} \\ V_{DS} > V_{OD} \end{cases}$$

Se invece il transistor è in zona triodo, si ha che:

$$\begin{cases} V_{GS} > V_{TH} \\ V_{DS} < V_{OD} \end{cases}$$

In questo caso, l'equazione che modella il comportamento del transistor è:

$$i_D = \beta_n \frac{W}{L} (2v_{OD}V_{DS} - V_{DS}^2)$$

Se poi si è in zona di triodo profondo, si ha che il termine quadratico diventa annullabile, dunque in sostanza il comportamento del dispositivo diviene analogo a quello di una resistenza.

### **Esempio teorico-pratico: campionatore elementare**

Si vuole a questo punto proporre un esempio atto a chiarire e fissare i concetti appena proposti. Volendo realizzare un campionatore ideale, si può utilizzare uno schema di questo tipo:

Vogliamo a questo punto tracciare la caratteristica tensione-corrente, cercando di comprendere in quale regione di funzionamento si trova il transistor, in questo circuito; si supponga di fissare  $V_G$  alla tensione di alimentazione, di avere una certa  $V_{in}$  (il cui valore verrà specificato più avanti), e una  $V_1$ , che rappresenta la tensione tra drain e source.  $V_{in}$  è sostanzialmente il segnale da campionare, il cui valore verrà immagazzinato nella capacità  $C$ .

A questo punto, sappiamo che:

$$0 < V_{in} < V_{DD}$$

Ora, ci sono sostanzialmente due casi:

- se  $V_1$  è positiva, il transistoro ha il drain a destra; se ipotizzo  $V_{in} = 0$ , cadranno 3.3 volt su  $V_{GS}$ , dunque  $V_1$  decide la corrente che deve passare nel transistoro;
- se  $V_1$  è negativa, ciò che fa passare corrente nel canale è di nuovo  $V_1$ , dunque source e drain scambiano i propri ruoli rispetto al caso precedente.  $V_{GS}$  decide lo stato di conduzione del transistoro. Essa è data da:

$$V_{GS} = V_G - V_S$$

ma

$$V_S = V_{in} - V_1$$

dunque

$$V_{GS} = V_G - (V_{in} - V_1) > V_{th}$$

Si noti che  $V_{th}$  è una soglia che dipende dalla  $V_{SB}$ , dunque dalla carica sulla capacità  $C$ . In che stato siamo dunque? Beh, per scoprirlo, basta verificare se la tensione tra drain e source (dunque  $V_1$ ) è o meno maggiore della tensione di overdrive:

$$V_{DS} > V_{OD} \implies V_1 > V_G - V_{in} + V_1 - V_{th}$$

Dunque:

$$V_G - V_{in} - V_{th} < 0$$

Quasi sempre questa diseguaglianza non è verificata e dunque, per  $V_1 < 0$ , siamo in zona triodo; altrimenti, di saturazione.

Si ha una curva del tipo:

La parte *a destra* è relativa al primo punto, e viceversa; si vede che a sinistra si ha un andamento prevalentemente *triode*, mentre a destra qualcosa di più simile alla saturazione, dove si ha come parametro per  $V_{GS}$  la tensione  $V_{in}$  (che in effetti la modula).

## Transistori per la realizzazione di amplificatori - weak inversion

Quando si vogliono utilizzare i transistori per realizzare circuiti di amplificazione, un parametro molto importante è la transconduttanza: essa, in sostanza, regola il guadagno del circuito. Supponiamo a questo punto di polarizzare in saturazione il transistor, e utilizzare da qui il modello di piccolo segnale:

Supponendo a questo punto che:

$$\begin{cases} V_{GS} > V_{TH} \\ V_{DS} > V_{OD} \end{cases}$$

Possiamo dire che:

$$i_D = \beta_n \frac{W}{L} (v_{GS} - V_{th})^2$$

Dove:

$$v_{GS} = V_{GS} + v_{gs}$$

A questo punto, sviluppiamo il quadrato, al fine di dividere la parte continua e la parte variabile:

$$i_D = \beta_n \frac{W}{L} [(V_{GS} - V_{th})^2 + 2(V_{GS} - V_{th})v_{gs} + v_{gs}^2]$$

ricordando dunque che:

$$g_m = \left. \frac{\partial i_D}{\partial V_{GS}} \right|_{V_{DS}=\text{costante}}$$
$$i_D \sim I_D + 2\beta_n \frac{W}{L} V_{OD} v_{gs}$$

Il termine che moltiplica  $v_{gs}$  è detto  $g_m$ , ossia è la transconduttanza del circuito. Si può dire anche che:

$$g_m = \frac{2I_D}{V_{OD}}$$

Se riportassimo il valore della transconduttanza al variare di  $V_{DS}$ , otterremmo qualcosa di questo genere:

Il nostro obiettivo, al fine di avere un amplificatore con un guadagno molto elevato, è quello di far crescere la transconduttanza del circuito. Tra i vari parametri in grado di comandare la transconduttanza, come si può vedere, vi è anche il rapporto di aspetto, con però una particolare indicazione:

aumentare il rapporto di aspetto certamente aumenta la transconduttanza del transistor, ma allo stesso modo aumenta anche i parametri parassiti, riducendo la banda.

C'è un particolare accorgimento che si può fare, al fine di realizzare ottimi amplificatori: si può polarizzare il transistor MOS in uno stato di **weak inversion**, ossia **debole inversione**: se tengono il drain a una tensione sufficientemente elevata, si può dire che il comportamento del transistor abbia un andamento di questo genere:

$$i_D = \frac{W}{L} I_{D0} e^{\frac{V_{GS}}{nV_T}}$$

$I_{D0}$  è un parametro del modello del circuito. Volendo derivare la transconduttanza, a questo punto, si ha:

$$g_m = \frac{W}{L} I_{D0} \frac{1}{nV_T} e^{\frac{V_{GS}}{nV_T}} = \frac{I_D}{nV_T}$$

Cosa ci dice ciò? Questa è molto simile alla transconduttanza di un transistor bipolare. Si ha, a questo punto, un andamento del genere:

Il primo punto, a transconduttanza, è quello nella regione di weak inversion; si scende dunque, quando si entra in regione di saturazione.

Una nota: parlare di weak inversion implica dire che nel canale, a differenza della condizione di strong inversion, si ha una bassa densità di carica; se tuttavia si costruisce un transistor molto grosso, in modo da avere una carica totale confrontabile con quella di un MOS equivalente operante in stato di strong inversion, si può ottenere al contempo una buona corrente e una transconduttanza elevata. Questo fatto ha un drawback: di fatto le prestazioni in bassa frequenza sono ottime, ma la banda, a causa delle elevate dimensioni del transistor, è molto ridotta.

### 1.3.1 Modelli del transistor MOS

Verranno a questo punto presentati i modelli che utilizzeremo, per il transistor nMOS.

#### Modello di ampio segnale

Il modello di ampio segnale può essere il seguente:

questo rappresenta, come si può vedere, un modello semplificato di ampio segnale. Esso è utile al fine di discutere le dinamiche di ingresso e di uscita del transistor. Una nota: in questo modello, l'effetto body è tenuto in conto,

ma nascosto nella tensione di soglia, ossia in  $V_{th}$ , la quale è funzione della tensione tra source e bulk:

$$i_D = \beta_n \frac{W}{L} (V_{GS} - V_{th}(V_{SB}))^2$$

### Modello di piccolo segnale

Il modello di piccolo segnale potrebbe essere il seguente:

Si noti che in questo ambito è stato esplicitato un generatore di transconduttanza, che quantifica l'effetto body. La transconduttanza di body è pari a:

$$g_{mb} = \left. \frac{\partial i_D}{\partial V_{SB}} \right|_{V_{GS}=\text{costante}, V_{DS}=\text{costante}} = \frac{\partial i_D}{\partial V_{th}} \frac{\partial V_{th}}{\partial V_{SB}} = -g_m \frac{\gamma}{2\sqrt{V_{SB} + 2\phi_F}}$$

Ciò ci può far pensare a un fatto: al fine di realizzare un circuito di amplificazione, vi sono due possibilità: quella classica, ossia introdurre il segnale dal gate, e una nuova, ultimamente molto documentata in letteratura: quella di sfruttare l'effetto body per introdurre il segnale, e farlo amplificare.

### 1.3.2 Proprietà dinamiche del transistor MOS

Si cercherà a questo punto di identificare le capacità parassite presenti in un MOS, evidenziando i problemi ad esse collegate. Si considerino le varie sezioni del transistor MOS.

In generale si può identificare, per ogni regione di sovrapposizione, una lunghezza  $L_D$ , indicante il fatto che si ha un overlap dell'ossido con una delle regioni; essa introduce una capacità parassita tra il gate e il source, e una tra il gate e il drain; supponendo che  $L_D$  sia uguale sia per source sia per drain ( $C_3$  e  $C_4$ ), le due capacità saranno uguali: queste capacità sono sostanzialmente legate a dei condensatori a facce piane parallele, dunque a capacità fisse, costanti nel tempo. Si ha poi a che fare con altre due capacità: una per la regione di svuotamento ( $C_2$ ), e una tra gate e canale ( $C_1$ ).

Un'osservazione: **di solito**, nei processi,  $L < W$ .

Si faccia riferimento alla sezione AA: il modello che si può ricavare per il transistor è il seguente:

Questo ci fa capire che il modello da utilizzare, in termini di capacità parassite, dipende dallo stato di polarizzazione che utilizziamo: a seconda della regione di funzionamento che utilizziamo, vi saranno alcune capacità, invece che altre. Questo fatto non è positivo, come vedremo tra breve.

Prima di tutto, è possibile quantificare le varie capacità:

$$C_3 = C_4 = C_{ox}WL_D$$

Queste capacità sono fisse, dal momento che dipendono dalla tecnologia presente: dal fatto che si introducono veri e propri condensatori parassiti.

Restano da discutere ancora due capacità, ma per esse bisognerà aggiungere qualcosa: sarà necessario fare ipotesi sullo stato di conduzione.

- Se il MOS è in stato OFF, il canale di fatto non esiste, dunque si ha che  $C_1 = 0$ : non essendoci canale, non si ha la capacità. Ciò che si ha è una capacità che collega direttamente gate a bulk:

$$C_{GB} = C_{ox}WL$$

In questo ambito,  $C_1$  diventa una capacità tra gate e substrato: dove dovrebbe esserci il canale si ha, di fatto, il substrato, dal momento che non si ha inversione e dunque formazione di cariche; si possono dunque identificare due sostanziali contributi:

$$C_{GB} = C_1 + C_{ext}$$

Per ora, le capacità tra gate e drain e gate e source sono dovute alle sole  $C_3$  e  $C_4$ , dunque sono molto piccole (dal momento che  $L_D$  si suppone essere piccolo, e queste capacità sostanzialmente dipendono dalla geometria del sistema).

- Se il MOS è in stato di saturazione, si formerà di sicuro il canale, che però sarà strozzato; si ha la regione di inversione, dunque la  $C_1$  nel senso che intendevamo all'inizio (capacità collegata al canale), ma si ha anche una capacità tra gate e drain la capacità risultante tra gate e drain è comunque sostanzialmente dominata dalla  $C_4$ , dunque, in sostanza, niente di nuovo; poi, si ha una capacità tra gate e source, e in particolare, si ha che:

$$C_{GS} \sim C_3 + \frac{2}{3}C_{ox}WL$$

- Quando si passa alla regione triodo, le capacità  $C_{GS}$  e  $C_{GD}$  diventano sostanzialmente uguali; in particolare, si ha che:

$$C_{GS} \sim C_{GD} \sim \frac{1}{2}C_{ox}WL$$

Volendo, si può graficare l'andamento delle capacità, per poter trarre alcune conclusioni:

Si vede chiaramente che, se il dispositivo è OFF, la capacità prevalente è quella tra gate e bulk, mentre le altre sono sostanzialmente inesistenti; man mano che si va verso la zona di saturazione, la capacità tra gate e source aumenta, mentre quella tra gate e bulk riduce il proprio valore; infine, in regione triodo, la capacità tra gate e bulk resta trascurabile, mentre quelle tra gate e drain e gate e source sono circa uguali.

Abbiamo dunque visto che, effettivamente, al variare dello stato del dispositivo variano le capacità interessate; si consideri, per comprendere le problematiche legate a questo fatto, il circuito del campionario ideale:

Con il transistor spento, il grosso della capacità è contenuta nella  $C_{GB}$ ; data la definizione del problema, ciò accade per  $t < t_0$ ; quando il transistor è acceso, dunque per  $t > t_0$ , la situazione cambia completamente, dal momento che entrano violentemente in gioco le capacità  $C_{GS}$  e  $C_{GD}$ , mentre la  $C_{GB}$  diventa pressochè nulla. Questo fatto però è assolutamente negativo: il fatto che capitino qualcosa del genere ci dice che si hanno degli scambi di carica da qualche parte: quando infatti la capacità tra gate e bulk si annichilisce, si ha che la carica da qualche parte deve pur andare (essendosi accumulata), e tra le varie parti andrà anche nel condensatore  $C$  che mantiene in memoria le informazioni da campionare: questo è molto negativo, dal momento che si altera la memoria del circuito, rovinando l'informazione. Questo fenomeno è detto **charge injection**.

In generale, se il transistor è pilotato da un ampio segnale, ossia da un segnale che fa cambiare la regione di funzionamento del dispositivo, la matrice delle capacità cambierà, dunque il circuito sarà tempo-variabile: non solo i parametri del modello varieranno, ma varierà il modello stesso: si avrà, di fatto, una topologia differente del circuito.

Quando si è per esempio sotto l'ipotesi di funzionamento di piccolo segnale, si vincola sostanzialmente la regione di funzionamento: si dice che il punto di lavoro del MOS non varia nel tempo, dunque che la topologia del circuito equivalente è costante nel tempo, e anche i suoi parametri.

Nel caso del MOS, solitamente, si fa riferimento alle capacità  $C_{gs}$ ,  $C_{gd}$ , ignorando le altre; ci sarebbero anche  $C_{js}$  e  $C_{jd}$  (si noti che l'uso dei pedici minuscoli fa riferimento a un modello di segnale), dunque si otterrebbe qualcosa del genere:

Questo modello di segnale ha diversi parametri; uno dei più importanti, per quanto raramente citati nei corsi di elettronica, è il guadagno di corrente. Esso può essere definito come:

$$\frac{i_d}{i_g}$$

ossia, come rapporto della corrente sul drain e di quella sul gate, tenendo conto dei parametri parassiti.

Come mai siamo interessati a questo parametro? Si osservi il problema, applicato a un problema reale: si supponga di avere un microprocessore, in uscita a esso un condensatore, e un driver in mezzo:

si deve cercar di scaricare e caricare, ciclicamente, questo condensatore. Come fare? Beh, sostanzialmente, ogni volta che si parla di driver, ci si riconduce a degli inverter:

Come funziona ciò? Beh, per far passare l'uscita da basso a alto si deve spegnere l'nMOS e accendere il pMOS; viceversa per il ciclo di scarica (nel quale la corrente girerà in verso opposto).

La velocità di carica/scarica dipende sostanzialmente da  $\tau = RC$ : si deve cercare di diminuire la resistenza, utilizzando il rapporto di aspetto. Problema: per diminuire la resistenza bisognerebbe aumentare il rapporto di aspetto, dunque le dimensioni del transistor; tanto più il transistor però è grosso, tanto più aumenteranno i parametri parassiti.

Ciò che si vuole realmente massimizzare è il rapporto tra la corrente erogata e quella introdotta: tanto maggiore sarà il guadagno in corrente, tanto minore sarà il numero di inverter da mettere in cascata al fine di pilotare la capacità, il carico. Man mano che si dovrà gestire sempre più corrente si dovranno aumentare le dimensioni del transistor, dunque ottenendo di fatto blocchi sempre più grossi nella cascata.

I processi tecnologici vengono solitamente caratterizzati rispetto alle prestazioni del transistor, quantificate mediante i guadagni in corrente: tanto più elevato è il guadagno in corrente, tanto migliori saranno i nostri transistori.

Si faccia riferimento a un transistor MOS che si pilota in corrente, chiuso in corto sul drain:

Volendo realizzarlo in laboratorio bisogna utilizzare il precedente schematico: polarizzare e disaccoppiare la base (il disaccoppiamento si può fare con una resistenza  $R_G$ , con una capacità o con un qualche bias-T; il segnale è un segnale di corrente, introdotto in gate; in uscita si ha un misuratore della variazione di corrente dovuta alla  $i_{in}$ .

Il circuito equivalente sarà il seguente:

Si può vedere chiaramente che:

$$i_d = g_m v_{gs} - i_x$$

Ma si può dire che:

$$i_x = sC_{gd}v_{gs}$$

dunque:

$$i_d = (g_m - sC_{gd})v_{gs}$$

Poi, si può vedere che la  $i_{in}$  si va a ripartire tra due capacità: quella  $C_{gs}$ , e quella  $C_{gd}$ ; su entrambe, dato il corto da drain a riferimento, vi sarà la stessa tensione; dunque:

$$i_{in} = s(C_{gs} + C_{gd})v_{gs}$$

da qua:

$$v_{gs} = \frac{i_{in}}{s(C_{gd} + C_{gs})}$$

Si può finalmente scrivere che:

$$i_d = i_{in} \frac{g_m - sC_{gd}}{s(C_{gs} + C_{gd})}$$

Dunque:

$$\frac{i_d}{i_{in}} = \frac{g_m - sC_{gd}}{s(C_{gs} + C_{gd})}$$

Questa espressione presenta un polo nell'origine, e uno zero nel semipiano di destra:

La pulsazione dello zero è:

$$\omega_z = \frac{g_m}{C_{gd}}$$

Cerchiamo di semplificare questa equazione, in modo da trovare qualcosa di semplice e che permetta di valutare la bontà del processo tecnologico. Per  $\omega \ll \omega_z$ , l'espressione è approssimabile a:

$$\frac{i_d}{i_{in}} \sim \frac{g_m}{s(C_{gs} + C_{gd})} \sim \frac{\mu_n C_{ox} \frac{W}{L} v_{OD}}{s \frac{2}{3} C_{ox} W L}$$

questo, trascurando  $C_{gd}$ . Semplificando:

$$= \frac{\mu_n v_{OD}}{\frac{2}{3} L^2 s}$$

Per tirare su il guadagno di corrente, dunque, questa espressione ci dice che bisogna realizzare canali **corti**.

Ora: il modello di transistori MOS a canale corto è certamente diverso da quelli di transistori a canale lungo; con il canale corto, in realtà, la dipendenza non è più dal reciproco di  $L^2$ , ma solo dal reciproco di  $L$ , ma comunque ciò che è detto è sempre vero: al diminuire della lunghezza di canale, il guadagno in corrente aumenta.

Un parametro molto importante, per la classificazione delle prestazioni del processo, è la  $f_T$ , ossia la frequenza di transizione allo 0 del guadagno in corrente: si noti che questa non è assolutamente una caratteristica in grado di quantificare il comportamento dinamico del nostro dispositivo, ma solo la qualità del processo: ai fini del progetto non è un dato interessante. Si può vedere che:

$$f_T = \frac{\omega_z}{2\pi} = \frac{\mu_n v_{OD}}{\frac{4}{3}\pi L^2}$$

## 1.4 Componenti passivi

### 1.4.1 Resistori integrati

Un'operazione possibile e anzi spesso necessaria è quella di integrare resistori, utilizzando alcuni passi del processo CMOS. Generalmente, un resistore ha un aspetto di questo tipo:

La lunghezza  $L$  generalmente predomina su  $w$  e  $t$ ; si sa, dalla legge di Ohm, che:

$$R_{AB} = \rho \frac{L}{S}$$

dove

$$S = wt$$

in altro modo, si vede:

$$R_{AB} = \frac{\rho L}{t w} = R_S \frac{L}{w}$$

Dove  $R_S$ , nella trattazione, sarà sempre chiamata **resistenza superficiale**.

Fondamentale dunque non è tanto il valore dei singoli  $L$  e  $w$ , quanto del loro rapporto: una resistenza infatti sarà data da diversi quadri, messi vicini, in modo da formare la geometria appena proposta. Visto che l'area è un

parametro molto *costoso*, questa resistenza sarà da realizzare in uno spazio ridotto: massimizzare il rapporto, ma non le dimensioni! In realtà, purtroppo, non è possibile fare resistenze infinitamente piccole: in ogni manuale di processo, infatti, si ha a che fare con qualcosa di questo genere:

Si ha un andamento di questo tipo, per la resistenza superficiale;  $w$  non può essere ridotto più di tanto, e anzi di solito si utilizza un certo valore in poi: quello per cui la resistenza superficiale, per quanto bassa, tende a divenire costante: si hanno infatti grosse tolleranze sulle aree realizzabili, di conseguenza tocca scegliere questo caso di worst-case.

Verranno ora proposte le principali tecnologie da utilizzare.

### ***n*-well**

Un primo modo per realizzare un resistore integrato è mediante l'uso di un *n*-well, ossia di una sacca drogata *n*:

Si hanno processi che possono mettere a disposizione diverse  $R_S$ : da 100 ohm al quadro a un/dieci kilohm al quadro; ancora una volta si ha a che fare con accoppiamenti parassiti verso il substrato.

Un'alternativa è quella di introdurre anche uno strato solo  $n^+$ :

In questo caso, dal momento che lo strato  $n^+$  superficiale è molto drogato, la resistenza diminuisce:

$$R_S \sim 10 \div 200 \Omega / \square$$

I due processi appena visti hanno un difetto: questi resistori come già detto sono accoppiati capacitivamente al substrato; il fatto che ci sia questa diffusione, può contenere un effetto di questo tipo:

Se applichiamo un potenziale, otteniamo qualcosa di molto negativo: dato  $A$  potenziale più alto di  $B$ , la regione svuotata sarà più larga verso  $B$  che verso  $A$ : a seconda della tensione applicata la dimensione della sezione della resistenza varierà!

Questo significa che le resistenze così realizzate hanno un valore che dipende dalla tensione applicata ai suoi capi: con esse, è problematico anche solo stabilire un rapporto di partizione, dal momento che le resistenze dipendono dalla tensione.

Sono da preferire i resistori del secondo tipo, a drogaggio più elevato, dal momento che la presenza di più drogante permette di attenuare l'effetto.

### **Resistore isolato**

Un'altra soluzione potrebbe essere la seguente:

Anche questo processo è molto simile a un processo già visto in precedenza: si collega il terminale di isolamento a quello con il potenziale più alto, in modo da rendere tendenzialmente costante tutta la regione di svuotamento. Questo processo è lo stesso del nMOS a doppio isolamento.

### **Resistore in polisilicio**

Un'alternativa, è la seguente:

L'epitassia permette di isolare meglio rispetto ai casi precedenti; inoltre, è possibile realizzare valori abbastanza variabili.

### **Resistenze metal**

Ancora un metodo:

Si introduce qualcosa di questo genere: uno strato di metallo (Metal 1), molto sottile, dunque uno di ossido, uno di metallo (Metal 2) un po' più spesso del predecessore, e così via: Metal 1 è sostanzialmente largo qualche decimo di micron, mentre le più spesse sono qualche micron (usualmente, questo processo è realizzato in allumina).

Queste piste, queste resistenze, sono di solito utilizzate per interconnessioni: le più sottili possono tollerare poca corrente, di conseguenza si utilizzano solo per le connessioni corte, locali, lunghe non più di qualche millimetro; volendo utilizzarle in altre applicazioni, esse presentano parametri fortemente parassiti (date le loro dimensioni). Andando su Metal 2 o su Metal 3 si riduce la  $R_S$ , dunque si ha la possibilità di realizzare interconnessioni a livelli più alti, per collegare i vari blocchi tra loro.

### **Tolleranze nei resistori**

I componenti passivi sono affetti da tolleranze di fabbricazione: volendo classificare i resistori, essi vengono definiti con delle tolleranze di fabbricazione, di vario tipo. Analizzeremo, ora, le principali tolleranze di fabbricazione.

Prima di tutto, una tolleranza è sul valore nominale del resistore:

$$\frac{\Delta R}{R} = \frac{R_1 - R_n}{R_n}$$

Dove  $R_1$  è il valore effettivamente realizzato con il processo, e  $R_n$  il valore nominale.

Altro parametro è la variazione con la temperatura:

$$R_1 = R_n(1 + k_T \Delta T)$$

Altro ancora, la tensione:

$$k_V = \frac{R(V_1) - R(V_2)}{R(V_1)}$$

o, ancora, normalizzando rispetto alla differenza di tensione:

$$k_V = \frac{R(V_1) - R(V_2)}{V_1 - V_2}$$

Tra i più importanti, infine, il parametro di tolleranza sul matching:

$$M_T = \frac{R_1 - R'_1}{\frac{R_1 + R'_1}{2}}$$

Cosa significa ciò? Se, sullo stesso substrato, realizzo due resistori identici tra loro, si avranno delle tolleranze di fabbricazione; questo è il parametro di tolleranza sul matching.

Si ha, per i vari parametri, la seguente tabella, che li classifica:

Le più importanti colonne sono quelle che riguardano la tolleranza sul valore nominale, e sul valore di matching.

Si può osservare che le tolleranze di fabbricazione sono davvero molto elevate; volendo per esempio realizzare il solito stadio a source comune, si può avere qualcosa di questo tipo:

Supponendo di aver polarizzato lo stadio in modo da avere  $I_{D0} = 10$  mA, si ha:

$$\bar{V}_0 = V_{DD} - RI_{D0}$$

Se si tiene conto delle tolleranze, si hanno variazioni enormi rispetto al punto del lavoro: si arriva tranquillamente a variazioni del 30%.

Se i progetti fossero realizzati sui valori assoluti, l'integrazione su silicio non avrebbe senso; questo per fortuna non è così, dal momento che la maggior parte dei circuiti è fatta utilizzando coppie di componenti: i componenti come si suol dire **matched**. Questo è un concetto importantissimo, che bisogna tenere bene a mente: il progetto di un circuito integrato deve essere basato su rapporti di grandezze: concettualmente, fare valori singolarmente precisi è impossibile, ma ciò che è possibile fare, all'interno di un integrato, è realizzare componenti *sbagliati alla stessa maniera*.

Per quanto riguarda i coefficienti di tensione, anche essi sono molto importanti; un esempio classico nei quali questa tolleranza si fa sentire, è nel problema dei *partitori di tensione*: essi sono molto utilizzati, per esempio per un convertitore ADC, per stabilire le tensioni di ingresso: se supponiamo

che la corrente nelle resistenze sia trascurabile, o di avere delle resistenze di  $n$ -well, cosa si avrà? Beh, vediamo:

Volendo avere delle resistenze elevate, usare gli  $n$ -well sarà fondamentale. Si ha però un problema: lo spessore fisico delle resistenze è  $t_0$ , ma quello reale utilizzabile è dipendente dalla regione di svuotamento: da  $t_0$  sarà necessario togliere  $x_d$ , ottenendo:

$$t_i = t_0 - x_d$$

dove  $t_i$  è l'effettivo spessore della resistenza. Dato che si parla di  $n$ -well, ciascuna resistenza ha il substrato che deve essere collegato obbligatoriamente al potenziale di riferimento. Ciò è problematico, perchè, se da un lato tutti i substrati sono collegati al riferimento, dall'altro la tensione ai capi del resistore cambia! Alcune saranno collegate a tensioni prossime a 2 V, altre a tensione prossime a 1 V! Questo significa che la tensione di polarizzazione del diodo varia, e dunque anche la resistenza opposta dal dispositivo! Si ha che:

$$\frac{R_{128}}{R_1} = \frac{t_1}{t_{128}} = \frac{t_0 - x_{d,1}}{t_0 - x_{d,128}}$$

dove, dalla teoria dei dispositivi, si sa che:

$$x_d = \left[ \frac{2\epsilon_0\epsilon_r}{N_D N_A} (V_R + \phi_0) \right]^{\frac{1}{2}}$$

Dati, per esempio,  $t_0 = 2\mu m$ ,  $N_D = 10^{16}$ ,  $\phi_0 = 0,7$  V, si può calcolare che:

$$\frac{R_1}{R_{128}} = 0,6$$

Il rapporto di partizione varia violentemente: man mano che il  $\Delta V$  aumenta, il rapporto di partizione cambia.

Cosa capiterebbe, con resistori a doppio isolamento?

Sostanzialmente, potremmo collegare il corpo a qualcosa di diverso dal GND, dunque per esempio alla **testa** del resistore: in questo modo tutte le resistenze sono *sbagliate*, ma allo stesso modo.

Purtroppo, questo resistore ha uno svantaggio: il miglioramento ottenuto è pagato in area! Infatti, ora, si ha qualcosa di questo genere:

La sacca  $p$  ora è più drogata di quella dell'epitassia, dunque le resistenze saranno minori a parità di superficie occupata, e per ottenere gli stessi valori di resistenza sarà necessaria più area; si ha inoltre ancora da introdurre l'isolamento, che aumenta ancora l'area necessaria.

## 1.4.2 Condensatori integrati

Il modo più semplice per avere un condensatore integrato è il seguente:

Utilizzare un condensatore MOS: si collegano il source e il drain con un corto circuito, dunque i due terminali sono questi.

Quantificare la capacità non è banale: sotto all'ossido si ha il canale, e a seconda che esso ci sia o non ci sia, la capacità assunta dal dispositivo è diversa. Volendo usare il MOS come condensatore, dunque, bisogna sapere che la capacità tra i punti  $A$  e  $B$  dipende dalla tensione tra i due punti stessi:

Si tratta di una capacità non lineare: essa si comporterà veramente come tale solo al di sopra di una certa soglia,  $V_{th}$  (che non è per forza quella che abbiamo sempre inteso come tensione di soglia del MOS).

Ciò porta a un'altra osservazione: i terminali  $A$  e  $B$  non sono intercambiabili: non è possibile scambiare i due terminali, dal momento che, dipendendo la capacità dalla tensione ai suoi capi, il segno della tensione è fondamentale.

In un caso, se infatti il gate  $A$  è al potenziale più alto, la capacità sarà polarizzata oltre la  $V_{th}$ ; nell'altro modo, il canale non sarà formato, e la capacità sarà differente.

Non è tutto: oltre a  $C_{AB}$  nominale, si hanno delle capacità *gratis*, aggiuntive, parassite: si introducono delle  $C_j$  che, a seconda di dove e come il condensatore è posizionato, si comportano e appaiono in modo diverso, cambiando il funzionamento del circuito.

Il problema del canale formato si può risolvere sfruttando l'ossido di gate, pre-formando il canale:

In questo caso il canale è preformato, dunque sempre presente, e il valore della capacità tendenzialmente costante. Resta comunque il problema della capacità parassita verso il substrato, e non solo: oltre alla problematica appena proposta, ve ne è un'altra (in entrambi i casi): il condensatore stesso presenta resistenze parassite! Si può infatti vedere che esso ha un comportamento di questo tipo:

In realtà, alla fine, in molti casi, il modello utilizzato è molto semplificato rispetto a questo:

In questo modo risulta possibile calcolare la costante di tempo associata a questa capacità, tenendo conto dell'effetto resistivo:

$$\tau = C_{AB} \frac{R_{ch}}{4}$$

Spesso, assieme al valore della capacità, tra i parametri si assegna infatti  $\tau$ , basandosi su un modello come questi.

## Condensatori metal-metal

Un'alternativa alla tecnologia prima proposta sono i condensatori metal-metal:

Si hanno due armature, Metal1-Metal2 (in alternativa, in polisilicio), di questo genere.

Anche in questo caso esistono capacità parassite: prima, esse insistevano tra il nodo B e massa, ma ora anche il nodo A ha piccole capacità parassite, che derivano dalla metallizzazione del polisilicio sul gate del transistor (per quanto piccole). Anche questi condensatori dunque sono affetti da parametri parassiti, che dipendono da dove il condensatore è stato piazzato.

### Note finali

Una nota: come già accennato (e come si rivedrà meglio tra breve), una buona idea potrebbe essere quella di utilizzare componenti matchati tra loro; il matching che si ottiene tra condensatori della stessa specie è di gran lunga migliore rispetto a quello di resistori matchati: per questo motivo si cerca di preferire l'uso di condensatori matchati rispetto ai resistori; anche se, infatti:

$$\frac{\Delta C}{C} \sim 30\%$$

si ha che

$$M_T \sim 0,05\%$$

Altro parametro molto importante è la capacità specifica: nel caso del MOS:

$$C_{ox} \sim 2 \div 3 fF/\mu m^2$$

nel caso di un condensatore Metal1-Metal2:

$$C_{ox} \sim 0,01 \div 0,02 fF/\mu m^2$$

Questo significa che, in sostanza, conviene di molto utilizzare i condensatori MOS (essendo la capacità specifica molto maggiore).

## 1.5 Matching di componenti integrati

Volendo realizzare due componenti A e B integrati uguali tra loro (che si tratti di resistori, condensatori o transistor), essi devono avere, al fine di essere *matchati*, non solo la stessa forma e la stessa area occupata, ma anche essere

circondati da componenti uguali: è necessario introdurre, nel problema, una totale simmetria.

Volendo introdurre un componente  $C$ , esso rovinerebbe la simmetria, e renderebbe vano ciò che si intende fare; piuttosto, ciò che si può fare è cercare di mantenere la simmetria, includendo altri componenti, elettricamente ridondanti, ma in grado di far ottenere la simmetria (per quanto non verranno mai utilizzati).

Parlando di condensatori (ma in realtà non solo), si sa che la capacità è sicuramente collegata all'area contenuta tra le facce:

$$C_{AB} = \epsilon_0 \epsilon_r \frac{WL}{t}$$

Si noti che si ha anche una capacità di perimetro,  $C_p$ ; ciò comporta il fatto che una generica capacità, oltre alla capacità di area  $C_A$ , avrà questa seconda capacità:

$$C_{AB} = C_A + C_p$$

La capacità perimetrale dipende da alcuni aspetti; uno, può essere l'errore di allineamento nelle maschere; nel caso dei condensatori a facce piane parallele, queste potrebbero essere capacità di bordo, di *fringe*:

per questo motivo si cerca di realizzare la capacità con l'armatura inferiore più spessa della superiore: si riducono gli effetti di bordo.

Gli errori di fabbricazione, di costruzione delle capacità integrate, riguardano sostanzialmente errori nella realizzazione di  $W$  e  $L$ :

Nella realtà può capitare che si costruisca qualcosa in cui l'ossido abbia un andamento di questo genere; avere una deformazione di questo tipo dell'ossido potrebbe cambiare il valore della capacità, anche di molto, rispetto a quanto atteso. Chiamando  $\Delta L$  la lunghezza di errore massima, si può avere:

$$W_1 = W_{1d} - 2\Delta L$$

$$L_1 = L_{1d} - 2\Delta L$$

Il valore effettivo di  $C_1$  sarà:

$$C_1 = C_{ox}(L_{1d} - 2\Delta L)(W_{1d} - 2\Delta L)$$

approssimando al primo ordine l'espressione, si ottiene:

$$\sim C_{ox}W_{1d}L_{1d} - 2\Delta L(W_{1d} + L_{1d})C_{ox}$$

Dove il primo termine rappresenta il valore nominale delle capacità (dipendendo da  $C_{ox}$  e dalle dimensioni desiderate), mentre il secondo è un termine di errore, dipendente dalla tolleranza  $\Delta L$  ottenuta. Volendo calcolare l'errore, dunque:

$$\frac{\Delta C_1}{C_1} = \frac{2\Delta L(W_{1d} + L_{1d})C_{ox}}{C_{ox}W_{1d}L_{1d}} = \frac{2\Delta L(W_{1d} + L_{1d})}{W_{1d}L_{1d}}$$

ora, si noti che al numeratore si ha il doppio della somma di lunghezza e larghezza desiderate: esso, sostanzialmente, è il perimetro della capacità  $C_1$ ; al denominatore si ha il prodotto di lunghezza e larghezza: l'area di un rettangolo!

$$= \Delta L \frac{P_1}{A_1}$$

Torniamo, dopo questa parentesi, al discorso del matching: si supponga di dover realizzare in un integrato una coppia di capacità  $C_1$  e  $C_2$ ; nella parentesi precedente, abbiamo quantificato l'errore, dovuto a problemi geometrici, sul valore della capacità; il rapporto delle due capacità sarà (tenendo conto degli errori):

$$\frac{C_1}{C_2} = \frac{C_{1d} \left(1 + \frac{\Delta C_1}{C_1}\right)}{C_{2d} \left(1 + \frac{\Delta C_2}{C_2}\right)}$$

usando la formula appena ricavata:

$$= \frac{C_{1d} \left(1 + \Delta L \frac{P_1}{A_1}\right)}{C_{2d} \left(1 + \Delta L \frac{P_2}{A_2}\right)}$$

Al fine di avere un rapporto il più possibile simile a quello desiderato:

$$\frac{C_1}{C_2} = \frac{C_{1d}}{C_{2d}}$$

I due condensatori devono avere lo stesso rapporto perimetro/area; se infatti:

$$\frac{P_1}{A_1} = \frac{P_2}{A_2}$$

ottengo proprio ciò che ho desiderato.

Questo è vero, se la capacità deve essere uguale a un'altra; cosa posso fare, se però voglio realizzare un circuito con condensatori diversi? Si supponga

di aver a che fare con condensatori con area diversa, uno  $A_1$ , uno  $A_2$ , relativi a  $C_1$  e  $C_2$  differenti. Definisco:

$$\alpha \triangleq \frac{A_2}{A_1}$$

Dunque, volendo ottenere un valore di  $\alpha$  molto preciso, avrò che:

$$\frac{P_1}{A_1} = \frac{4}{l_1}$$

$$\frac{P_2}{A_2} = \frac{4}{\sqrt{\alpha} l_1}$$

Il criterio che si utilizza per ottenere lo stesso rapporto perimetro-area è quello del **matching unitario**: per poter realizzare  $\alpha$  con elevata precisione, è necessario realizzare il condensatore  $C_2$  come replica di tanti condensatori  $C_1$ , posti in parallelo tra loro:

in questa maniera,  $C_2$  è il parallelo di tanti  $C_1$ , e si ha che:

$$\frac{P_1}{A_1} = \frac{P_2}{A_2}$$

La stessa tecnica può essere applicata sui transistori MOS: volendo per esempio fare uno specchio di corrente:

si può richiedere che:

$$\frac{I_2}{I_1} = K > 1$$

allora, sappiamo, dalla teoria, che:

$$K = \frac{\frac{W_2}{L_2}}{\frac{W_1}{L_1}}$$

dove, però, dalla teoria precedentemente vista, sappiamo che:

$$\frac{W}{L} = \frac{W_d}{L_d} \frac{1 - \frac{2\Delta l}{W_d}}{1 - \frac{2\Delta l}{L_d}}$$

sviluppando secondo Taylor al primo ordine:

$$\sim \frac{W_d}{L_d} \left( 1 - 2\frac{\Delta l}{W_d} + 2\frac{\Delta l}{L_d} \right)$$

Sostituendo in  $K$ :

$$K \sim \frac{\frac{W_{d2}}{L_{d2}} \left(1 - 2\frac{\Delta l}{W_{d2}} + 2\frac{\Delta l}{L_{d2}}\right)}{\frac{W_{d1}}{L_{d1}} \left(1 - 2\frac{\Delta l}{W_{d1}} + 2\frac{\Delta l}{L_{d1}}\right)}$$

di solito, nei processi,  $W \gg L$ ; assunta questa ipotesi, si può semplificare l'espressione, come:

$$K \sim \frac{\frac{W_{d2}}{L_{d2}} \left(1 + 2\frac{\Delta l}{L_{d2}}\right)}{\frac{W_{d1}}{L_{d1}} \left(1 + 2\frac{\Delta l}{L_{d1}}\right)}$$

Per ottenere un buon matching tra i due transistori, si deve richiedere che:

$$L_{d1} = L_{d2}$$

ossia che la lunghezza sia la stessa.

Questa è un'analisi dell'errore dovuto alle imprecisioni nella realizzazione delle dimensioni fisiche.

### Errori sui gradienti

I problemi che peggiorano il matching sono molto svariati; un altro errore, tra i componenti integrati, riguarda i gradienti presenti nel sistema. Si supponga, per esempio, di avere qualcosa del tipo:

$$C_{ox} = C'_{ox} + mx$$

Questo, per esempio, è dovuto a gradienti causati dalla crescita dell'ossido, che modificano il valore di  $C_{ox}$  al variare della posizione spaziale considerata.

Si immagini, ora, di voler realizzare due capacità (potrebbero essere resistori, transistori, qualsiasi cosa!):  $C_A$ , più grossa, e  $C_B$ . Si ha:

$$\frac{C_A}{C_B} = \frac{C_{A1} + C_{A2}}{C_B} = \frac{2C'_{ox} + m(x_1 + x_2)}{C'_{ox} + mx_3}$$

Volendo a questo punto imporre che, effettivamente,  $C_A = 2C_B$ , si deve imporre:

$$2 = \frac{2C'_{ox} + m(x_1 + x_2)}{C'_{ox} + mx_3}$$

questo, è vero esclusivamente se:

$$x_3 = \frac{x_1 + x_2}{2}$$

Ossia, se il componente  $C_B$  è in mezzo ai due componenti usati per  $C_A$ .

Questo, è vero nel caso si sia fatta un'ipotesi di gradiente costante, ossia variazione lineare della funzione di capacità specifica. Il matching unitario può essere utilizzato per migliorare il problema: usando tutti i componenti a partire da un singolo componente, mettendone tanti in parallelo, è sufficiente fare così: si mette un componente in mezzo a tutti gli altri, e poi si creano i collegamenti in parallelo, minimizzando i problemi:

### Cenni sui problemi nell'impiantazione dei MOS

Un problema, come già spiegato, è legato a problemi nel realizzare la geometria con precisione; un secondo problema, altrettanto importante, riguarda il drogaggio. Volendo realizzare un MOS, si deve fare qualcosa del genere:

L'impiantazione ionica è il processo utilizzato per la realizzazione delle sacche  $n$ . Al fine di non introdurre crepe nel materiale, ciò che di solito si fa è utilizzare un angolo di incidenza per l'impiantazione, di 7 o 8 gradi: a causa di questo, tuttavia, si ha una regione di ombra, sotto il punto di vista del drogaggio: ci sarà una zona che avrà un'intensità di drogante minore rispetto alle altre. Ciò dunque sostanzialmente introduce un problema: le due regioni di source e drain saranno diverse e dunque, di fatto, non più intercambiabili!

Ciò che bisogna fare è dunque etichettare chi è il source e chi è il drain: questi transistori sono sostanzialmente diversi

Infatti, se un terminale è un po' più drogato di un altro, si ha per esempio una resistenza parassita più grande da una parte che dall'altra: in questo modo, se con un circuito i problemi sono minimi, con l'altro diventano assolutamente importanti, e dai quali non si può prescindere.

Al fine di risolvere questo problema si ha una tecnica di layout detta di **baricentro comune**: il layout corretto, secondo questa tecnica, è quello realizzato nella seguente maniera: dovendo piazzare più transistori MOS, si chiede che tutte le diagonali colleganti i vari terminali tra loro (tra dispositivi opposti), passi per lo stesso punto:

L'applicazione di questa tecnica riesce a limitare l'effetto ombra.

Il drogaggio influenza le tensioni di soglia, nei transistori MOS; le soglie dei transistori potrebbero essere diverse tra loro, proprio a causa dei diversi livelli di drogaggio. La differenza delle tensioni di soglia è inversamente proporzionale alla radice quadrata delle dimensioni del transistor moltiplicate tra loro:

$$\Delta V_{th} = \frac{K_{\Delta V_{th}}}{\sqrt{WL}}$$

Tanto più i transistori sono grandi, tanto minore sarà il mismatch sulle tensioni di soglia del transistor. Questo per esempio è molto importante quando si deve fare una coppia differenziale!

## Capitolo 2

# Stadio di amplificazione elementare MOS

Prima di parlare del progetto vero e proprio, alcune considerazioni riguardo alle tecniche e ai risultati: progettare significa identificare un insieme di parametri, e la topologia del circuito, al fine di soddisfare delle specifiche. Le specifiche possono essere le più svariate: sul guadagno, sulla banda, sullo slew rate, sulle dinamiche, sul rumore. Abbiamo dei vincoli, e dei costi, e non solo: le varie specifiche possono essere collegate tra loro, dunque soddisfarne una può implicare il fatto che altre non siano soddisfabili: a volte, non è possibile soddisfare tutte le specifiche assieme.

### 2.1 Analisi dello stadio elementare - polarizzazione

Si prenda in considerazione il seguente stadio:

Si ha un generatore in continua per polarizzare il gate, dunque un generatore di segnale. La caratteristica di ingresso è un qualcosa di questo genere:

Uniamo i due disegni: in generale, per questo stadio, si ha che:

$$V_0 = V_{DD} - Ri_D$$

Si supponga che il transistor sia polarizzato in stato di saturazione; si avrà che:

$$i_D = \beta_n \frac{W}{L} (V_{IN} - V_{th})^2$$

Abbiamo dei limiti: quando  $V_0 = v_{OD}$ , si ha un passaggio alla regione triodo; questa non ci piace, dal momento che la regione triodo non è una regione in cui valga il legame lineare tra ingresso e uscita (è una regione a carattere circa resistivo). In regione di saturazione, si può dire che il guadagno dello stadio (analizzando il circuito) valga:

$$\frac{v_0}{v_{in}} = -g_m R = -2\beta_n \frac{W}{L} v_{OD} R$$

Al crescere della tensione di overdrive, come si vede, cresce il guadagno: la transconduttanza (considerando  $R$  fissa), ha un andamento del tipo:

Solitamente, al fine di avere una transconduttanza piuttosto elevata, si polarizza il transistor in stato di saturazione.

Il guadagno è una specifica fondamentale: uno degli obiettivi per chi progetta amplificatori è proprio quello di avere un guadagno molto alto: un sistema di questo tipo è molto utile se controreazionato, dunque se chiuso su di un anello; avendo un guadagno elevato, gli effetti della reazione saranno ottimi.

Si ha a che fare con due parametri liberi: la tensione di polarizzazione di gate,  $V_{GS}$ , e la resistenza di drain,  $R$ : entrambi possono contribuire al miglioramento di due specifiche sostanziali, ossia il guadagno e la dinamica.

Si supponga di avere una resistenza  $R$  fissa, e di aver scelto una certa  $V_{GS}$  tale per cui si ha un certo punto di lavoro; volendo agire su  $V_{GS}$ , per tirare su il guadagno potremmo volerci portare vicino alla zona triodo! Ciò non è positivo, dal momento che l'ascissa del punto di lavoro, ossia la tensione del punto di lavoro, deve essere al centro della possibile variazione, e non vicino a uno dei bound: se  $V_0 \sim v_{OD}$ , la dinamica è molto limitata, dal momento che la tensione potrà variare poco rispetto al punto dove sta. Nella migliore delle ipotesi, scegliendo in modo opportuno il punto di lavoro, si deve avere che:

$$v_{OD} < V_0 < V_{AL}$$

Dunque, il punto ideale, è:

$$V_{DS0} = \frac{V_{AL} + v_{OD}}{2}$$

In questo modo, tenendo conto del fatto che la dinamica deve essere simmetrica, essa è massimizzata.

L'altro parametro su cui si può lavorare è la resistenza  $R$ : si supponga di aver fissato una certa tensione  $V_{GS0}$ , che non verrà cambiata per ora: modificando  $R$  si modifica sostanzialmente la pendenza della retta di carico:

essa, infatti, è pari a  $-\frac{1}{R}$ : se aumento  $R$ , riduco la retta di carico. La retta di carico va a intercettare la curva che ci interessa, che sarà relativa al parametro  $V_{GS0}$ : di nuovo, più abbasso la resistenza più il punto di lavoro si avvicina alla regione triodo, aumentando il guadagno ma peggiorando la dinamica.

Variando contemporaneamente i due parametri, è possibile massimizzare la dinamica e ottenere un buon guadagno: giocando assieme con  $R$  e  $V_{GS}$ , si riesce a mantenere  $V_{DS0}$  nel punto centrale della dinamica, diminuendo la tensione di ingresso: si può, con questo trucco, far aumentare il guadagno.

Proviamo a utilizzare alcune formule: si sa, dall'analisi del modello di piccolo segnale, che il guadagno di tensione è pari a:

$$A_v = -2\beta_n \frac{W}{L} v_{OD}$$

ora, esprimiamo la tensione di overdrive come funzione della corrente:

$$v_{OD} = \sqrt{\frac{i_{DQ}}{\beta_n \frac{W}{L}}}$$

Dunque:

$$\begin{aligned} A_v &= -2\beta_n \frac{W}{L} \sqrt{\frac{i_{DQ}}{\beta_n \frac{W}{L}}} R = \\ &= -2\sqrt{\beta_n \frac{W}{L}} i_{DQ} \frac{V_R}{i_{DQ}} \end{aligned}$$

questo, dopo aver espresso  $R$  come:

$$R = \frac{V_R}{i_{DQ}}$$

dove  $V_R$  è la tensione continua sulla resistenza  $R$ . Dunque:

$$= -2\sqrt{\beta_n \frac{W}{L}} \frac{V_R}{\sqrt{i_{DQ}}}$$

A questo punto, facendo crescere la resistenza di carico, aumenta il guadagno! Una soluzione potrebbe usare un resistore integrato  $R$  molto grosso, ma non va bene: occuperebbe troppa area!

L'idea di base potrebbe essere la seguente:

Si supponga di fissare una  $V_{GS0}$  di ingresso: serve un carico in grado di presentare una resistenza molto alta, con però un offset in grado di far

passare una corrente di polarizzazione: si tratta, in sostanza, di un qualcosa con caratteristica quasi piatta.

Cosa meglio di un generatore di corrente?! Esso permette di avere un'e-scursione notevole della corrente di uscita, in quanto piatto, e comunque tenere un buon guadagno.

Qual è il guadagno di questo stadio? Beh, qualcosa del genere:

Si avrà:

$$\frac{v_0}{v_{in}} = -g_m r_0$$

Dove  $r_0$  è la resistenza dovuta alla modulazione della lunghezza del canale del transistor di transconduttanza:

$$r_0 \sim \frac{1}{\lambda I_{DQ}} = K_d \frac{L}{I_{DQ}}$$

Per far crescere  $r_0$ , il transistor deve essere **lungo**.

Il generatore di corrente in realtà si può realizzare semplicemente mediante l'elettronica nota: esso è detto **carico attivo**: questo generatore di corrente, di fatto, sarà costituito da uno specchio di corrente:

Dall'intersezione delle caratteristiche di  $M_2$  (che sarà il vero e proprio carico attivo del sistema) e  $M_1$  (l'amplificatore, o meglio il transconduttore), si ottiene il punto di lavoro. Essendo il carico attivo, la caratteristica di  $M_2$  sarà la curva di carico dell'amplificatore.

Calcoliamo il guadagno per questo circuito, utilizzando il modello di piccolo segnale:

Si vede chiaramente che tra gate e source di  $M_2$  c'è solo una continua atta a mantenere la polarizzazione, dunque il generatore di transconduttanza di  $M_2$  sarà di sicuro spento. Si avrà dunque:

$$A_v = \frac{v_0}{v_{in}} = -g_m (r_{01} \oplus r_{02})$$

Questo, tanto per incominciare con le brutte notizie, di sicuro sarà già più basso dell'altro. Volendo a questo punto sostituire la transconduttanza con la funzione precedentemente introdotta, si ha:

$$A_v = -2 \sqrt{\beta_n \frac{W}{L}} \Big|_1 i_{DQ} (r_{01} \oplus r_{02})$$

Ma:

$$r_{01} \oplus r_{02} = \left( \frac{1}{r_{01}} + \frac{1}{r_{02}} \right) = \left( \frac{1}{K_d L_1} + \frac{1}{K_d L_2} \right)^{-1} = \frac{K_d}{I_{DQ}} \left( \frac{1}{L_1} + \frac{1}{L_2} \right)^{-1} =$$

$$= \frac{K_d}{I_{DQ}} L_1 \oplus L_2$$

Dunque, il guadagno può essere riscritto, supponendo per semplicità  $L_1 = L_2 = L$ :

$$\begin{aligned} A_v &= -K_d \frac{L}{\sqrt{I_{DQ}}} \sqrt{\beta_n \frac{W}{L}} \Big|_1 = \\ &= -\sqrt{\beta_n} K_d \frac{\sqrt{WL}}{\sqrt{I_{DQ}}} \end{aligned}$$

A parole, per far crescere il guadagno, devo ridurre  $I_{DQ}$ , e aumentare l'area del transistor usato da transconduttore. Si ha un andamento di questo genere:

1 mA in microelettronica è una corrente enorme, 100 nA molto piccola: solitamente la corrente va dai 10 ai 100  $\mu\text{A}$ . Il guadagno piatto coincide, sostanzialmente, con una condizione di weak inversion.

### Note conclusive della sezione

Alcune note: prima di tutto, diciamo due parole sulla dinamica di uscita. Fissati il punto di lavoro per  $M_1$  e per  $M_2$ , la dinamica di uscita è stata automaticamente definita: la resistenza di uscita infatti ora è definita da  $r_{01} \oplus r_{02}$ , e ciò che sta tra il limite della regione triodo è la dinamica. Si può dire che:

$$v_{OD1} < V_0 < V_{AL} - v_{OD2}$$

Questa è la dinamica di uscita; dovrò dunque fissare il punto di lavoro come punto centrale:

$$\bar{V}_0 = \frac{V_{AL} - v_{OD2} + v_{OD1}}{2} \sim \frac{V_{AL}}{2}$$

questo, se le due tensioni di overdrive sono simili tra loro.

Il guadagno di questo circuito può andare grosso modo da 50 a 200: questo è abbastanza irrisorio, rispetto ai guadagni che vorremmo. La soluzione a questo punto può essere o realizzare, in qualche modo, un unico stadio con un guadagno enorme, o mettere tanti di questi stadi in cascata.

Un'ulteriore nota: supponendo di mettere una resistenza  $R$  di valore piccolo (un 10 kilohm) come carico, lo stadio di amplificazione non funzionerebbe: senza mettere una qualche capacità di disaccoppiamento, il punto di lavoro verrebbe completamente distrutto, e il sistema lavorerebbe in condizioni che non potremmo prevedere (senza rifare i conti).

## 2.2 Prestazioni dinamiche

A questo punto, si consideri il circuito equivalente di piccolo segnale dell'amplificatore, tenendo in conto le capacità parassite; l'obiettivo, ora, è quello di studiare la risposta in frequenza, al variare dei parametri di progetto.

Il modello in questione è qualcosa del genere:

Si sta considerando un equivalente Norton per l'ingresso.

Considerandole note, verranno ora solo richiamate le principali espressioni per l'amplificatore:

$$A_v = -g_{m1}R_{out} \frac{1 - \frac{s}{s_z}}{\left(1 + \frac{s}{s_{p1}}\right) \left(1 + \frac{s}{s_{p2}}\right)}$$

Dove:

$$s_z = \frac{g_{m1}}{C_{gd}}$$

$$s_{p1} = -\frac{1}{R_S(1 + g_{m1}R_{out}C_{gd}) + R_S C_{in} + R_{out}(C_{out} + C_{in})}$$

$$s_{p2} = -\frac{R_S[C_{in} + (1 + g_m R_{out})C_{gd}] + R_{out}(C_{gd} + C_{out})}{R_S R_{out}(C_{in}C_{gd} + C_{in}C_{out} + C_{gd}C_{out})}$$

Il guadagno ha un comportamento di questo tipo:

A questo punto, queste formule possono essere studiate al variare dei vari parametri; quello che possiamo fare noi, a mano, è utilizzare delle condizioni particolari, al fine di maneggiare espressioni semplici:

- $C_{out}$  molto maggiore delle altre capacità nel circuito;
- $C_{in}$  molto maggiore delle altre capacità nel circuito;
- $C_{ext}$  molto maggiore delle altre capacità nel circuito, dove  $C_{ext}$  è una capacità che si collega a ponte tra ingresso e uscita.

Questi sono i tre casi fondamentali che andremo ad analizzare, per conoscere il comportamento.

### Primo caso

Nel primo caso, facendo i limiti per  $C_{out} \rightarrow \infty$ , si ottiene:

$$s_{p1} \sim \frac{-1}{R_{out}C_{out}}$$
$$s_{p2} \sim -\frac{1}{R_S(C_{in} + C_{gd})}$$

mentre  $s_z$  ha la solita espressione. A questo punto, al fine di effettuare un'analisi sotto il punto di vista dei componenti:

$$s_{p1} = -\frac{2I_{DQ}}{C_{out}K_dL_1}$$
$$s_{p2} = -\frac{1}{R_S\left(\frac{2}{3}C_{ox}WL + C_{ox}LDW_1\right)}$$

quest'ultima, considerando che al posto di  $C_{in}$  ci sia esclusivamente una capacità tra gate e source, calcolata come al solito. Per quanto concerne  $s_z$ :

$$s_z = \frac{\sqrt{4\beta_n i_{DQ}}}{C_{ox}L_D\sqrt{W_1L_1}}$$

I parametri ai quali siamo fondamentalmente interessati sono  $W_1$  e  $i_{DQ}$ : si può vedere che, al crescere di  $W_1$ , lo zero si sposta verso sinistra, il secondo polo verso destra (si avvicinano all'origine), mentre il primo polo rimane sostanzialmente inalterato; per quanto riguarda  $i_{DQ}$ , invece, tutti sono influenzati: lo zero aumenta la propria frequenza, come anche il primo polo, mentre il secondo polo resta sostanzialmente costante

### Secondo caso

Il secondo caso, nel quale si carica l'ingresso con una capacità elevata rispetto alle altre, verrà trattato più rapidamente; se si ha che:

$$R_{out}(C_{gd} + C_{out}) \ll R_S(C_{in} + C_M)$$

e poi

$$C_{in} \gg C_M$$

si ha sostanzialmente ciò:

In sostanza, il secondo polo e lo zero abbassano in modulo la propria frequenza, mentre il primo polo rimane pressochè costante, all'aumentare

della dimensione  $W_1$ ; all'aumentare di  $i_D$ , invece il primo polo continua a rimanere costante, ma il secondo polo e lo zero tendono ad aumentare il modulo della propria frequenza.

### Terzo caso

Si consideri a questo punto il terzo caso, ossia l'aggiunta di una  $C_{ext}$  esterna, in parallelo a quella che chiamiamo  $C_{gd}$ ; ciò che sostanzialmente capita è che si ha un qualcosa del genere:

$$s_{p1} \sim -\frac{1}{g_{m1}R_S R_{out}C_{ext}}$$

$$s_{p2} = -\frac{g_m C_{ext}}{C_{in}C_{out} + C_{ext}(C_{in} + C_{out})}$$

$$s_z = \frac{g_m}{C_{ext} + C_{gd}}$$

In questo caso, all'aumentare del valore della capacità introdotta esternamente, si ha che lo zero tende ad abbassare la propria frequenza, mentre per i due poli si ha il cosiddetto fenomeno del **pole splitting**: il secondo polo aumenta in modulo la propria frequenza, e il primo polo la riduce in modulo.

Possono essere utili anche le variazioni al variare di  $W_1$  e  $i_D$ .

## 2.3 Analisi di rumore

Verrà a questo punto proposta un'analisi riguardante il rumore. Si consideri il seguente circuito:

Per ciascun MOSFET distinguiamo la sorgente di rumore, mediante i generatori di rumore; si farà dunque l'analisi, supponendo che i generatori siano nulli, e si va a calcolare la potenza di rumore portata sull'uscita.

L'idea è sostanzialmente la seguente: dato il circuito, riportare tutti i rumori all'ingresso, considerando una rete noise-less, senza rumore; da qua, dato un unico contributo di rumore all'ingresso di tutto il circuito, si potrà calcolare la potenza di uscita.

L'idea di solito è quella di partire da questo modello equivalente di rumore: dove

$$i_{n,eq}^2 = i_{n,thermal}^2 + g_m^2 v_{n,flicker}^2$$

questo, grazie al fatto che si possono considerare le sorgenti scorrelate. Si può dimostrare che:

$$i_{n,thermal}^2 \triangleq i_{n,t}^2 = 4k_B T \gamma g_m$$

dove  $\gamma$  è un parametro di processo, legato alla lunghezza di canale, che di solito vale  $\frac{2}{3}$ ; per il secondo contributo:

$$v_{n,flicker}^2 \triangleq v_{n,f} = \frac{K_f}{C_{ox} W L} \frac{1}{f}$$

Come si comporterà dunque il rumore? Beh, sotto una certa  $f_C$ , detta **frequenza d'angolo**, il termine inversamente proporzionale alla frequenza tende a divergere, e tutto il rumore è dovuto al rumore flicker; da lì in poi, di sicuro, il rumore termico sarà il prevalente (essendo esso un rumore bianco); normalmente,  $f_C = (0.5 \div 2)$  MHz.

Ora, è possibile calcolare la corrente complessiva di rumore, utilizzando la sovrapposizione degli effetti delle potenze, come la somma delle due potenze:

$$i_n^2 = i_{n,t}^2 + i_{n,f}^2 = 4k_B T \gamma g_m + \frac{K_f g_m^2}{C_{ox} W L} \frac{1}{f}$$

Passando dunque alla potenza di ingresso, in termini di equivalente di tensione:

$$v_{n,eq,in}^2 = 4k_B T \gamma \left[ 1 + \frac{g_{m,2}}{g_{m,1}} \right] + v_{n,f}^2 \left[ 1 + \frac{g_{m,2}^2}{g_{m,1}^2} \right]$$

Da questa espressione, finalmente, è possibile passare ai parametri di processo e sostituire tutto. Per semplicità, quello che si fa di solito è distinguere le casistiche, e considerare solo un contributo di rumore, dunque fare tutto in funzione della frequenza. Se dunque  $f < f_C$ , si ha solo il rumore flicker, e dunque:

$$v_{n,eq,in}^2 \sim \frac{K_f}{C_{ox} W_1 L_1} \left[ 1 + \frac{\mu_p}{\mu_n} \right] \frac{1}{f}$$

Dove  $\mu_n$  e  $\mu_p$  sono le mobilità dei portatori.

Al fine di ridurre il rumore, come si può vedere, l'area del dispositivo deve essere elevata. Se imponiamo dunque il fatto che  $L_1 = L_2 = L$ , possiamo giocare anche solo più sui  $W_i$ . Problema: se aumentiamo l'area  $WL$ , aumentiamo i parametri parassiti, nella fattispecie le capacità, fatto che riduce la banda del sistema; d'altra parte, ricordiamo, si può aumentare anche il guadagno.

Volendo rimanere in uno stato in cui si hanno basse frequenze, bisogna aumentare  $g_{m1}$ , e ridurre  $g_{m2}$ ; per cambiare il secondo, è necessario lavorare sul rapporto di aspetto, dunque sulla dinamica di uscita del sistema; questo

significa ridurre, a parità di corrente, il limite superiore della dinamica di uscita. Si ricordi dunque che il transistoro di carico attivo,  $M_2$ , contribuisce alla definizione di  $R_{out}$  e di  $C_{out}$ .

## 2.4 Distorsione

La distorsione può essere sostanzialmente ricondotta a due casi: da un lato, si può parlare di distorsione con gli stadi in saturazione, e qualche armonica che si viene a formare; altro caso è la discussione della distorsione nel senso che i transistori sono fuori dallo stato di saturazione, di linearità.

Proponiamo alcuni risultati:

si ha che

$$V_O = V_{DD} - R_D i_D$$

dove

$$i_D = \beta_n (V_{IN} + v_{in} \cos(\omega t) - V_{TH})^2$$

In virtù di questa relazione, per la quale si avrebbe sostanzialmente ipotizzato la presenza dello stato di saturazione, si hanno componenti armoniche di secondo ordine, e una continua.

Ciò che verrà fatto in questa sezione è proporre due tecniche per la riduzione della distorsione.

### 2.4.1 Pre-post-distorsione

In un sistema non lineare, si ha sostanzialmente che un generico blocco  $f$  potrebbe introdurre distorsione, ossia potrebbe avere una caratteristica non lineare; un'idea potrebbe dunque essere quella di introdurre, prima o dopo esso, un altro blocco, tale per cui si abbia una distorsione opposta a quella introdotta dall'altro blocco:

Questa è la cosiddetta tecnica di pre-post-distorsione: un blocco introduce una certa distorsione, e un altro una distorsione identica, reciproca, più un certo guadagno  $K$  atto a introdurre il guadagno del circuito.

Come si può fare in pratica? Beh, un'applicazione potrebbe essere la seguente:

Si potrebbe semplicemente introdurre nel circuito un altro transistoro, collegato *a diodo*; in questa maniera, si avrebbe:

$$V_O = V_{DD} - V_{GS2}$$

dove

$$V_{GS2} = V_{TH2} + V_{OD} = V_{TH2} + \sqrt{\frac{i_{D2}}{\beta_n \frac{W}{L}|_2}}$$

Ma, come sappiamo,  $i_{D2}$  è in relazione con la tensione di ingresso:

$$V_O = V_{DD} - \left[ V_{TH2} + \sqrt{\frac{i_{D2}}{\beta_n \frac{W}{L}|_2} \beta_n \frac{W}{L}|_1} (V_{GS1} - V_{TH1})^2 \right]$$

Dunque:

$$V_O = V_{DD} - \left[ V_{TH2} + \sqrt{\frac{W}{L}|_1 \frac{W}{L}|_2} (V_{GS1} - V_{TH1}) \right]$$

dove

$$V_{GS1} = V_{IN} + v_{in}$$

da qua:

$$\begin{aligned} V_O &= V_{DD} - \left[ V_{TH2} + \sqrt{\frac{W}{L}|_1 \frac{W}{L}|_2} (V_{IN} + v_{in} - V_{TH1}) \right] = \\ &= V_{DD} - \left( V_{TH2} + \sqrt{\frac{W}{L}|_1 \frac{W}{L}|_2} (V_{IN} - V_{TH1}) \right) - \sqrt{\frac{W}{L}|_1 \frac{W}{L}|_2} v_{in} \end{aligned}$$

Dunque, il guadagno di bassa frequenza sarà:

$$A_v = \frac{v_o}{v_{in}} = \sqrt{\frac{W}{L}|_1 \frac{W}{L}|_2}$$

Questo, dunque, sarà vero a prescindere dalla polarizzazione! Ora, se  $L_1 = L_2 = L$ , avremo che:

$$A_v = \sqrt{\frac{W_1}{W_2}}$$

L'amplificatore, in layout, sarà fatto così:

Si noti che **non** siamo passati, nelle nostre considerazioni, dal modello di piccolo segnale: non abbiamo introdotto alcuna approssimazione nella determinazione della nostra teoria!

Volendo ora analizzare il modello di piccolo segnale di questo schema, si avrà qualcosa di questo genere:

L'impedenza vista dall'uscita del primo stadio è  $g_m^{-1}$ ! Si avrà che:

$$\frac{v_o}{v_{in}} = -\frac{g_{m1}}{g_{m2}}$$

Cosa ci dice ciò? Beh, volendo un guadagno elevato, dovremo ridurre la  $g_{m2}$ , e aumentare la  $g_{m1}$ . Questo schema purtroppo comporta degli svantaggi: la dinamica di uscita del sistema infatti è piuttosto ridotta. Questo fatto è implicabile al transistor collegato a diodo: aumentando la  $g_m$  di un transistor, infatti, ciò che si fa è aumentare

$$g_m = \frac{2i_{DQ}}{v_{OD}}$$

se si aumenta la tensione di overdrive, si va (guardando il transistor 2) ad abbassare il limite superiore della dinamica.

Per aumentare  $g_m$ , dovremmo aumentare le dimensioni del transistor; per fare questo, dunque, si aumenta la capacità parassita!

In queste capacità parassite passerà parte della corrente, e in questo modo, andando su in frequenza, le correnti che passeranno nelle capacità parassite diventeranno significative, e avremo che

$$i_{D1} \neq i_{D2}$$

Questa tecnica di pre-post-distorsione, dunque, funziona solo a patto di mantenere i transistori **piccoli**: questo implica il fatto che, con questo circuito, non sarà contemporaneamente possibile avere un guadagno elevato e una reiezione efficace della distorsione.

Il transistor può essere sia tipo  $p$ , sia tipo  $n$ ; l'importante, è che esso sia collegato a diodo.

## 2.4.2 Controreazione

La seconda tecnica per la reiezione della distorsione presente nel circuito è basata sulla controreazione. Di tutte le possibili reazioni introducibili nel circuito, la soluzione più semplice è quella di introdurre una retroazione sul source, mediante una  $R_S$ :

Questa  $R_S$  legge la corrente di drain (la quale è uguale alla corrente di source) e introduce una tensione sulla maglia di ingresso, controllandolo. Questo circuito viene detto **stadio a source comune degenerato di source**.

Studiamo, prima di tutto, l'amplificazione di piccolo segnale: si ha che

$$i_d = g_m v_{gs} = g_m (v_{in} - R_S i_d)$$

e

$$v_{gs} = v_{in} - R_S i_d$$

dunque:

$$i_d (1 + g_m R_S) = g_m v_{in}$$

e

$$\frac{i_d}{v_{in}} = \frac{g_m}{1 + g_m R_S} \triangleq G_m$$

Questa è una transconduttanza. Questa, potrebbe portarci a vedere questo dispositivo come un dispositivo equivalente, con transconduttanza equivalente  $G_m$ . Anche per questo caso, è possibile considerare solo dei valori notevoli per  $G_m$ :

- se  $g_m R_S \ll 1$ ,  $G_m \sim g_m$
- in caso contrario,

$$G_m \sim \frac{1}{R_S}$$

A questo punto, consideriamo le caratteristiche del dispositivo per  $R_S = 0$  e  $R_S \neq 0$ ; per  $R_S = 0$ , ossia se il dispositivo non è reazionato, si ha:

se  $R_S \neq 0$ , si ha il secondo grafico. Ora: se si ha che  $g_m R_S \gg 1$ , la linearità è stata molto aumentata: sia  $i_D$ , sia  $G_m$ , presenteranno tratti molto regolari; per  $i_D$  infatti si avrà, dopo un primo tratto quadratico, la prevalenza di un tratto lineare, dato dal fatto che  $R_S$  farà prevalere il proprio peso su tutti gli altri; allo stesso modo,  $G_m$  tenderà a diventare costante, a un valore prossimo a  $1/R_S$  (come già anticipato). Dato dunque  $g_m R_S \gg 1$ , si potrà dire che:

$$A_v \sim -\frac{R_D}{R_S}$$

Dal momento che però il nostro desiderio è quello di integrare su silicio questo stadio, potrò fare qualcosa di questo genere:

In questo caso, si avrà che:

$$A_v = -\frac{g_{m3}}{g_{m2}}$$

e questo dipende sostanzialmente dal rapporto dei  $W$  dei due transistori di carico!

Il vantaggio principale della degenerazione di source è l'introduzione di un'elevata impedenza di uscita; si può infatti vedere, utilizzando un generatore di test, che:

L'impedenza di uscita, grazie alla controreazione, è divenuta elevatissima:

$$r_{out} = \frac{v_T}{i_T} = r_0 + R_S r_0 (g_m + g_{mb}) + R_S \sim g_m R_S r_0$$

# Capitolo 3

## Stadio di amplificazione cascode

### 3.1 Dallo stadio a gate comune allo stadio cascode

Precedentemente, è stato introdotto, come stadio di amplificazione elementare, lo stadio a source comune; è possibile realizzare altri stadi elementari, quale per esempio lo stadio a gate comune:

Si polarizza il gate a un potenziale costante, e si spinge dal source una certa corrente, in questa maniera; sempre al source si collega un generatore di segnale, che sarà l'ingresso del sistema (supponendo ovviamente che il transistor sia in stato di saturazione). Si consideri ora il modello di segnale:

La tensione di uscita dello stadio sarà:

$$v_o = -R_D i_d = -R_D g_m v_{gs}$$

Valutiamo a questo punto la tensione di pilotaggio  $v_{gs}$ ; per fare ciò, è necessario tenere presente l'impedenza vista dal source del transistor (che sarà, come al solito,  $g_m^{-1}$ ; dunque:

$$v_{gs} = -\frac{g_m^{-1}}{g_m^{-1} + R_S} v_s$$

finalmente:

$$\frac{v_o}{v_s} = \frac{g_m R_D}{1 + g_m R_S}$$

se poi si ha che  $g_m R_S \gg 1$ ,

$$A_v = \frac{R_D}{R_S}$$

e se invece  $g_m R_S \ll 1$

$$A_v = g_m R_D$$

Questo stadio ha un comportamento molto simile a quello dello stadio a source comune, senza però essere invertente. Esso ha proprietà molto utili, quale la presenza di un'impedenza di uscita molto elevata: volendo utilizzare un generatore di prova, si può vedere che:

$$v_{gs} = -R_S i_T$$

dunque

$$v_T = -v_{gs} + (i_T - g_m v_{gs}) r_0$$

sostituendo la prima nella seconda e dividendo direttamente per  $i_T$ :

$$\frac{v_T}{i_T} = R_S + r_0(1 + g_m R_S)$$

Dunque, si ha la stessa impedenza di uscita che si aveva nello stadio a degenerazione di source: esso però può essere utilizzato per far crescere l'impedenza di uscita, e dunque il guadagno.

Si può realizzare una piccola variante, ossia utilizzare questo stadio, pilotato però in corrente:

Se siamo in regione attiva, le correnti di source e drain sono identiche:  $i_S = i_D$ ; in questo caso, però, imponiamo il fatto che sul source si può avere anche un segnale di corrente, ossia che:

$$i_S = I_S + i_s(t)$$

In questo modo, si ha:

$$R_{in} = \frac{v_D}{i_D} = -R_D$$

Questa è la transresistenza dello stadio: se si pilota lo stadio in corrente, esso sostanzialmente si comporta come una resistenza: uscita in tensione, ingresso in corrente. Se piloto in corrente questo circuito, in ingresso posso utilizzare un equivalente Norton invece di uno thevenin: il generatore di corrente non sarà ideale, ma avrà una  $r_0$  parallelo; facendo i conti, si può vedere la stessa resistenza di uscita di prima. Al fine di realizzare il generatore

di corrente, è possibile utilizzare un secondo transistor (che per noi sarà  $M_1$ , il quale fungerà da transconduttore: in ingresso metteremo una certa tensione  $v_{in}$ , ed esso in uscita fornirà una corrente, che sarà quella da usare per pilotare  $M_2$ , il transresistore.

Si ha:

$$v_o = -R_D g_m v_{gs}$$

dunque:

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

Questo guadagno è esattamente uguale a quanto visto precedentemente: esso è sostanzialmente uguale al guadagno di uno stadio a source comune! Tutto ciò che abbiamo fatto è inutile! Usato in questo modo, questo circuito non serve a niente, dal momento che è uguale a un source comune; nella pratica, questo stadio ha un nome preciso: stadio **cascode**, e, utilizzato in una maniera particolare, diventa molto potente.

## 3.2 Caratteristiche statiche dello stadio cascode

Ci poniamo a questo punto l'obiettivo di calcolare il guadagno intrinseco, ossia il guadagno con un carico ottimo: un generatore ideale di corrente. Si avrà:

$$\frac{v_o}{v_{in}} = -g_m R_{out}$$

questo, lo sapevamo; quanto vale però  $R_{out}$ ? Beh, vediamo:  $v_{gs1} = 0$ , dunque:

$$R_{out} \sim g_{m2} r_{o1} r_{o2}$$

dunque, il guadagno sarà:

$$A_v \sim -g_{m1} g_{m2} r_{o1} r_{o2}$$

Molto interessante: nel caso (per esempio) i due transistori fossero uguali, sostanzialmente si otterrebbe un guadagno che è il quadrato di quello di uno stadio a source comune: molto più grosso! A noi dunque starà caricare il cascode con un generatore di corrente a impedenza elevata!

Come fare? Beh, sicuramente, potremmo usare un carico attivo, come questo:

Funziona? Beh, si ha qualcosa del tipo:

$$A_v = -g_{m1} [(g_{m2}r_{02}r_{01}) \oplus r_{03}]$$

Ma, se  $r_{03}$  è dell'ordine di grandezza delle altre, l'operazione di somma armonica riduce drasticamente il valore della resistenza! Facendo una cosa del genere, si ricade nei limiti dello stadio a source comune!

Come si può risolvere il problema in maniera efficace? Sostanzialmente, caricando lo stadio cascode con un altro stadio cascode:

In questa maniera, si carica il cascode con qualcosa con la stessa impedenza (almeno, in termini di ordini di grandezza).

Dato questo stadio, valutiamone la dinamica: consideriamo metà stadio per volta (tanto la seconda metà gli è assolutamente simmetrica):

Siamo per ora interessati al solo studio del limite inferiore della dinamica (quello superiore verrà ricavato per simmetria): al fine di determinarlo, bisogna ricordare qual è la condizione che ci vincola maggiormente: il fatto che entrambi i transistori devono essere in saturazione; definito il nodo  $x$ , si ha che:

$$V_x = V_G - V_{GS2}$$

Si vede che  $V_x$  coincide sostanzialmente con  $v_{DS1}$ ; si deve dunque avere che:

$$V_x > V_{OD1}$$

Questa è un'ipotesi che permette di garantire la saturazione. Fatta questa ipotesi, si ragiona su  $M_2$ : si deve avere una  $v_{DS2}$  tale per cui:

$$v_{DS2} > V_{OD2}$$

Cerchiamo di determinare questo vincolo:  $v_{DS}$  è la differenza tra la tensione al drain e quella al source del transistor;  $v_{D2}$  è la tensione di uscita (essendo il drain del transistor 2 collegato all'uscita), e  $v_{S2}$  è sostanzialmente  $V_G - V_{GS2}$ ; si avrà dunque:

$$v_{D2} - v_{S2} = v_{D2} - (V_G - V_{GS2})$$

dunque:

$$v_0 - V_G > -V_{TH2}$$

ossia:

$$v_0 > V_G - V_{TH2}$$

Una novità: il limite inferiore della dinamica è sostanzialmente questo, ed è dettato da  $V_G$ , parametro di progetto. Siccome noi vogliamo una dinamica massima, dovremo tirare giù  $V_G$ ; abbiamo dei vincoli anche sotto questo punto di vista, dal momento che il guadagno deve essere tenuto alto: il guadagno infatti dipende da  $r_{01}$ , che è:

$r_{01}$  è l'inverso della pendenza della curva rappresentata: per farla crescere, è necessario di sicuro allungare il canale, ma non solo: dobbiamo tenere il punto di lavoro  $Q$  abbastanza avanti, altrimenti la pendenza cala, la resistenza diminuisce, e il guadagno diminuirà di conseguenza. Al fine di soddisfare entrambe le specifiche, ciò che si fa è fissare  $M_1$  al bordo della regione triodo, scegliendo dunque  $V_G$  in modo che:

$$V_x = V_{OD1}$$

In questo caso, il limite inferiore della dinamica di uscita sarà:

$$v_{o,min} = v_{OD1} + v_{OD2}$$

Questa scelta permette di soddisfare entrambe le specifiche.

Finora, avevamo ottenuto solo vantaggi dallo stadio a source comune al cascode: si ha una dinamica inferiore: questo è il prezzo che abbiamo pagato per aumentare il guadagno.

Volendo introdurre il carico designato, il secondo cascode, si può dire che il limite superiore della dinamica di uscita sia così calcolabile:

se  $V_x = V_{OD1}$ , si ha che  $V_y = V_{DD} - V_{OD3}$ ; facendo un ragionamento duale, si trova:

$$V_0 < V_{DD} - (v_{OD3} + v_{OD4})$$

La caratteristica sarà un qualcosa di questo genere:

Se si considera la caratteristica statica del circuito, per  $V_{IN} = 0$ , si ha sostanzialmente che  $M_1$  è spento, ossia che si comporta come un circuito aperto; nella pratica, esso è un circuito aperto fino a un certo punto: ci sono delle correnti di leakage che comunque fanno passare un po' di corrente per  $M_2$ , il quale non sarà di sicuro acceso pienamente, ma comunque in uno stato di debole inversione, dunque nella regione quadratica: la  $V_{GS2}$  è di poco inferiore alla tensione di soglia. Per quanto riguarda il cascode di carico, esso vorrebbe sostanzialmente forzare la corrente, ma sotto è tutto spento (o quasi):  $M_4$  dunque è in regione triodo, e ha su di sé una tensione costante;

$M_3$  e  $M_4$  sono in regione triodo, dunque hanno una caduta di tensione molto bassa, e quello che sostanzialmente si ha è che  $V_0 = V_{DD}$ . Questo, fino a quando la  $V_{IN}$  non arriva alla soglia del transistor 1: da quel punto, tutti entrano in conduzione, e in saturazione, fino a quando, a forza di tirare su la tensione, per  $V_{IN} = V_{OD1} + V_{OD2}$ , si ha che in triodo ci vanno i transistori di sotto!

La tensione  $V_x$  si adagia a  $V_{G1} - V_{TH2}$  proprio a causa del fatto che si hanno queste perdite di leakage: esse portano il transistor in debole inversione, e quindi un po' gli permettono di condurre.

### 3.3 Risposta in frequenza

Torniamo a questo punto alla struttura di base (quella con un generico carico resistivo), al fine di trovare informazioni riguardo alla banda del sistema:

In questo circuito è possibile identificare sostanzialmente tre nodi:  $A$ ,  $B$ ,  $X$ . Si può vedere che il nodo  $X$  è caricato da  $C_{gd1}$  e da  $C_{gs2}$ .

Ciò che ci serve, al fine di fare il progetto, sono delle espressioni approssimate, che ci permettano di capire quali sono le prestazioni in frequenza del circuito, e queste sostanzialmente per i tre nodi.

Questo stadio può essere visto come un common source che pilota un common gate: il common source, ossia lo stadio in basso, è già stato studiato; sappiamo inoltre che questo stadio a common source vede in sostanza solo  $g_{m2}^{-1}$ , ossia l'impedenza vista dentro al source di  $M_2$ . Tutto si riconduce allo studio di questo circuito equivalente di piccolo segnale:

Si ha:

$$s_{pA} = -\frac{1}{R_S (C_{gs} + (1 + g_{m1}g_{m2}^{-1}) C_{gd1})}$$

questo, utilizzando il teorema di Miller; essa è approssimabile a:

$$s_{pA} \sim -\frac{1}{R_S g_{m1} g_{m2}^{-1} C_{gd1}}$$

Essendo  $g_{m2}^{-1}$  piccola, questo polo si troverà plausibilmente a una frequenza elevata: questo risultato comporta una sorpresa molto piacevole, ossia il fatto che lo stadio cascode non è influenzato dall'effetto Miller. Per gli altri elementi:

$$s_{pX} = -\frac{g_{m2}}{C_{gs2}}$$

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

Infine, il terzo polo, relativo al nodo B, ossia il polo dominante del sistema, sarà:

$$s_{pB} = -\frac{1}{R_{out}C_0}$$

La risposta in frequenza di questo stadio sarà:

### 3.4 Analisi del rumore

Effettuiamo a questo punto una analisi del rumore, come al solito sul circuito di base; questa si può fare sostituendo i modelli ideali dei transistori con quelli che tengono conto del rumore, e dunque, tenendo conto della scorrelazione delle sorgenti di rumore, applicando il principio di sovrapposizione delle potenze. Si otterrà:

Abbiamo due transistori (supponiamo che la resistenza, che modella il carico del circuito, sia ideale: più avanti la sostituiremo col carico dello stadio). Due transistori, due contributi:

$$v_{n0}^2 = v_{n01}^2 + v_{n02}^2$$

A questo punto, calcoliamo i contributi della tensione di uscita: per quanto riguarda  $v_{n01}^2$ , basta spegnere il transistor 2, che sarà dunque solo sostituito dalla sua resistenza  $r_{02}$ ; si avrà dunque che  $i_1 = i_{D1}$ , e dunque:

$$v_{n01}^2 = R_D^2 i_{n1}^2$$

Per quanto riguarda il transistor  $M_2$ , dobbiamo spegnere  $M_1$  e far accendere  $M_2$ , si avrà qualcosa di duale: si può osservare che la corrente  $i_1$  e quella sul transistor spento,  $i_3$ , sono uguali; a questo punto, si può vedere che:

$$i_1 = i_3 = i_2 + g_{m2}v_{gs2}$$

dove, però

$$v_{gs2} = -i_3 r_{01}$$

dunque:

$$i_1 (1 + g_{m2}r_{01}) = i_2$$

dunque

$$i_1 = \frac{i_2}{1 + g_{m2}r_{01}}$$

allora

$$v_{n02}^2 = \left( \frac{R_D}{1 + g_{m2}r_{01}} \right)^2 i_{n2}^2$$

Cosa ci dice in pratica questa formula? Beh, se  $g_{m2}r_{01}$  è grosso, il rumore dovuto al secondo transistor diventa trascurabile:  $M_2$  contribuisce in modo marginale al rumore presente sull'uscita. Se dunque si studia lo stadio completo, si ha qualcosa di questo genere:

I rumori dovuti a  $M_2$  e  $M_4$  sono trascurabili rispetto agli altri; si avrà dunque che:

$$v_{n0}^2 \sim R_D^2 (i_{n1}^2 + i_{n3}^2)$$

Questa, poi, può essere portata sull'ingresso, volendo completare l'analisi, semplicemente dividendo per il guadagno  $A_v$  (noto da ciò che è stato fatto in precedenza):

$$v_{n,in}^2 = \frac{v_{n0}^2}{A_v^2}$$

### 3.5 Folded cascode

Oltre al classico cascode (anche detto **telescopico**), esistono gli stadi cascode di tipo *ripiegato*: *folded cascode*. Sostanzialmente, essi sono basati sulla seguente idea: si usa un transistor  $M_1$  a source comune, dunque

A differenza del cascode classico, un transistor è  $n$ , uno è  $p$ : nella fattispecie, è lo stadio a gate comune a essere un  $p$ -MOS. Esiste anche una versione duale di questo stadio, ossia quella con transconduttore di tipo  $p$ :

Di fatto, della corrente  $I_B$ , ossia della corrente fornita dal generatore di corrente di polarizzazione, un po' di essa viene presa da  $M_1$ , la restante parte da  $M_2$ ; l'equazione al nodo  $A$  sarà:

$$i_{D2} = I_B - i_{D1}$$

Facendo riferimento allo schema con transconduttore  $n$ , la tensione di uscita  $V_{out}$  sarà dunque sostanzialmente proporzionale alla corrente che passa sul carico del transconduttore; si può vedere che, se il transistor 1 è sotto soglia, l'uscita è massima: tutta la corrente va verso il carico. Da quando

si va sopra soglia, e man mano che si alza  $V_{in}$ , si avrà che la corrente si abbasserà, fino a una certa  $V_{in}^*$ , tensione per cui sarà il carico ad andare sotto soglia. Questo sostanzialmente significa che man mano che  $M_1$  si accende, esso ruberà sempre più corrente, e a un certo punto tutta la corrente  $I_B$  verrà portata via. Si può vedere che:

$$V_{in}^* = V_{TH1} + \sqrt{\frac{I_B}{\beta_n \frac{W_1}{L_1}}}$$

Questo, con l'overdrive che corrisponde alla corrente  $I_B$ ! Ossia, se in sostanza tutta la  $I_B$  va a passare dentro a  $M_1$ .

Possiamo fissare il punto di lavoro  $Q$  al centro: qui, entrambi i transistori sono polarizzati in regione di saturazione; modulando dunque la tensione di ingresso, si modula la corrente di drain, ma dunque  $i_{D2}$ , che è la corrente che passa nel carico; questo, fino a quando la relazione è verificata. Si ha che

$$v_o = R_L i_{d2} = -R_L i_{d2} = -g_{m1} R_L v_{in}$$

Ancora una volta, il guadagno è quello del source comune, ma ora l'impedenza vista dal drain sarà molto elevata: essa sarà circa pari a  $g_{m2} r_{01}$ , moltiplicata ancora per la resistenza che carica il nodo  $A$ ; essa è

$$r_A = r_{03} \oplus r_{01}$$

Per quanto riguarda il folded cascode:

$$R_{out} \sim g_{m2} r_{02} (r_{03} \oplus r_{01})$$

mentre nel cascode standard si ha:

$$R_{out} \sim g_{m2} r_{02} r_{01}$$

abbiamo dunque leggermente peggiorato l'impedenza di uscita, e il guadagno sarà conseguentemente un poco più basso di prima. Cosa si può dire invece riguardo la dinamica di uscita? Beh, si deve far sì che il nodo  $x$  abbia circa, come tensione:

$$V_x = v_{OD1}$$

In questo stadio, il limite inferiore della dinamica è sostanzialmente dipendente da  $R_L$ ; il limite superiore, invece, sarà tale per cui

$$V_A = V_{DD} - v_{OD3}$$

Questo, perchè in questo nodo la tensione di uscita può arrivare a:

$$v_0 < V_{DD} - (v_{OD3} + v_{OD2})$$

La tensione di uscita, nel cascode standard, doveva essere maggiore di  $v_{OD1} + v_{OD2}$ , quindi la differenza è poca; il limite superiore della dinamica invece non tiene conto del transistor di ingresso, ossia del transconduttore, dunque spesso saremo più interessati a questo tipo di soluzione.

# Capitolo 4

## Circuiti di polarizzazione

Nei sistemi integrati si deve sempre avere a che fare con almeno un circuito per la generazione di tensioni e correnti di riferimento: questi sono infatti fondamentali per le prestazioni del sistema.

Dato un circuito integrato, esiste almeno un circuito che, a partire dalla tensione di alimentazione  $V_{AL}$ , genera un insieme di tensioni e correnti di riferimento, che possono essere erogate o assorbite.

### 4.1 Specchi di corrente

Se disponiamo di un riferimento di corrente, potremo per esempio replicare una certa corrente quante volte vogliamo, mediante uno specchio di corrente. Cosa fa, in pratica, uno specchio di corrente? Beh, data una certa corrente di riferimento  $I_{REF}$ , generata mediante un circuito che ancora non conosciamo (e che verrà introdotto in questo capitolo), la si converte a una  $V_{GS1}$ , mediante una qualche funzione non lineare  $f$  realizzata dal transistor costituente il lato debole dello specchio; dunque, da  $V_{GS1}$ , il lato forte realizzerà una funzione del tipo  $Kf^{-1}$ , e produrrà in uscita una  $I_0$ .

Prima di tutto, discutiamo gli specchi di corrente dunque: essi sono fondamentali, come parte del circuito di polarizzazione: essi sono il mattone fondamentale che permette la distribuzione delle correnti di riferimento.

Ciascun MOS è caratterizzato da un certo rapporto di aspetto ( $W/L$ ); se siamo in saturazione, abbiamo inoltre che

$$v_{DS2} > v_{OD2}$$

Dunque, si può scrivere, in queste condizioni, che:

$$I_{REF} = \mu_n \frac{C_{OX}}{2} \frac{W_1}{L_1} (v_{GS} - V_{TH1})^2 (1 + \lambda_1 V_{DS1})$$

$$I_0 = \mu_n \frac{C_{OX}}{2} \frac{W_2}{L_2} (v_{GS} - V_{TH2})^2 (1 + \lambda_1 V_{DS2})$$

Se calcoliamo il rapporto di specchiaggio, otteniamo:

$$\frac{I_0}{I_{REF}} = \frac{\mu_n \frac{C_{OX}}{2} \frac{W_1}{L_1} (v_{GS} - V_{TH1})^2 (1 + \lambda_1 V_{DS1})}{\mu_n \frac{C_{OX}}{2} \frac{W_2}{L_2} (v_{GS} - V_{TH2})^2 (1 + \lambda_1 V_{DS2})}$$

Ora, se chiediamo che  $\lambda_1 = \lambda_2 = \lambda$ , e che  $V_{TH1} = V_{TH2}$ , si ha che:

$$\frac{I_0}{I_{REF}} = \frac{\frac{W_1}{L_1} (1 + \lambda V_{DS1})}{\frac{W_2}{L_2} (1 + \lambda_1 V_{DS2})}$$

Se le due  $V_{DS}$  sono diverse, in generale, si avrà un errore sulla corrente distribuita! Infatti, essa, non sarà completamente controllabile con il solo rapporto di aspetto, o almeno non fino a quando si utilizza questo schema. Come si può fare? Beh, un circuito alternativo potrebbe essere il seguente:

Se si fa in modo che  $V_x = V_y$  con  $V_B$ , si riesce a ottenere un rapporto di specchiaggio  $K$  ideale: in questo modo infatti si è imposto che la parte specchiante del circuito sia quella costituita dai soli  $M_1$  e  $M_2$ , mentre la restante parte è solo atta a imporre l'uguaglianza delle tensioni tra drain e source, in modo da ridurre l'errore. Questo specchio è migliore del precedente, ma presenta un grosso problema: abbiamo pagato la soluzione del nostro problema, riducendo la dinamica di specchiaggio. Da qui, può saltare fuori quest'altro circuito:

In questo ambito, si ha che  $M_1$  e  $M_2$  fanno **lo specchio**;  $M_3$  controlla la tensione di gate di  $M_2$ , e  $M_4$  genera la tensione di gate che dovrà polarizzare opportunamente  $M_3$ . Si ha che:

$$V_x = V_{GS1} = V_{TH1} + v_{OD1}$$

dunque

$$V_y = V_{GS1} + V_{GS4} - V_{GS3} = V_{TH1} + v_{OD1} + V_{TH4} + v_{OD4} - V_{TH3} - v_{OD3}$$

Se le tensioni di soglia e le tensioni di overdrive dei transistori 3 e 4 sono uguali, si ha che:

$$V_x = V_y$$

Inoltre

$$v_{OD5} = \sqrt{\frac{I_0}{\beta_n \frac{W_3}{L_3}}} = \sqrt{\frac{I_{REF}}{\beta_n \frac{W_4}{L_4}}}$$

dunque

$$\frac{I_0}{I_{REF}} = \frac{W_3}{W_4}$$

Se siamo dunque in grado di garantire un buon matching tra  $M_3$  e  $M_4$ , al fine di imporre  $V_x = V_y$ , otteniamo un buon risultato; si tenga comunque presente che il matching più importante sarà comunque quello tra  $W_1$  e  $W_2$ : quello relativo allo specchiaggio.

Ora: si supponga di utilizzare, come carico di questo specchio, un altro transistor: qual è il limite inferiore della dinamica di uscita? Beh, si ha che:

$$V_{0,min} = V_y + v_{OD3}$$

ma

$$V_y = V_{GS1} + v_{OD3}$$

dunque, se si ha che:

$$\frac{W_2}{W_1} = \frac{W_3}{W_4} = K$$

allora:

$$V_{0,min} = V_{TH1} + v_{OD1} + v_{OD3}$$

e questa, grosso modo, potrebbe valere un volt.

In un processo che produce dispositivi con tensioni da 0 a 3,3 V, sopra perdiamo la solita  $v_{OD}$  (circa qualche centinaio di millivolt), ma sotto circa un volt: la dinamica è stata troppo ridotta!

## 4.2 Specchio di corrente con elevata dinamica di uscita.

Sostanzialmente, l'idea dietro questo tipo di circuiti è: dati i soliti due transistori configurati a specchio, se ne introducono altri due, pilotati dalla stessa tensione  $V_B$ : essa dovrà essere tale da avere  $V_x = V_y$ :

Come si può ottenere questa  $V_B$ ? Beh, con un circuito retroazionato! Fintanto che i transistori sono percorsi da corrente, si ha che  $V_x = V_y$ ; prima

si aveva che  $M_1$  è un diodo, dunque  $V_{DS1}$  doveva per forza essere pari a  $V_{GS1}$ ; ora  $V_{DS1}$  può essere anche più piccola della tensione tra gate e source, dunque si può, di fatto, aumentare la dinamica! Si tengano ben presenti dunque i nostri due obiettivi:

- avere  $V_x = V_y$ , in modo da avere un rapporto di partizione molto preciso;
- avere la massima dinamica di uscita, configurando una  $V_B$  tale per cui si abbia  $V_{0,min} = v_{OD2} + v_{OD4}$ .

Se si sceglie dunque la  $V_B$  in modo tale da avere  $v_{DS2} = v_{OD2}$ , si avrà massimizzato la dinamica di uscita. Questo significa che, sostanzialmente, i transistori verranno polarizzati **al bordo della regione triodo**.  $V_B$  ovviamente avrà dei limiti: si deve infatti avere che:

$$V_{DS1} \geq v_{OD1}$$

$$v_{DS3} \geq v_{OD3}$$

però, come si può vedere studiando la topologia del circuito,

$$v_{DS1} = V_B - v_{GS3}$$

$$v_{DS3} = v_{GS1} - (V_B - v_{GS3})$$

sostituendo, si può vedere che:

$$V_B \geq v_{OD1} + v_{OD3} + V_{TH3}$$

$$V_B \leq V_{TH1} - v_{OD1} + V_{TH3}$$

Al che:

$$v_{OD3} + v_{OD1} + V_{TH3} < V_{TH1} + V_{TH3} + v_{OD1}$$

dunque

$$v_{OD3} < V_{TH1}$$

## Generazione di $V_B$

A questo punto, si supponga di avere a disposizione una certa  $I_{REF}$ , generata con un circuito di precisione (che verrà introdotto in seguito); come si può generare una tensione  $V_B$  in modo tale da polarizzare i transistori nella maniera corretta, a partire da  $I_{REF}$ ? Beh, si consideri il seguente circuito:

Il problema è avere una  $V_B$  in qualche modo proporzionale alla  $I_{REF}$ ; si deve avere dunque che:

$$V_B = v_{OD1} + V_{GS3}$$

Data una corrente, per poter ottenere una certa tensione, si possono per esempio utilizzare dei transistori collegati a diodo, e modificare il rapporto di aspetto  $W/L$  per ottenere la tensione che si vuole, fissata la corrente. Si noti che  $V_B$  però non è realmente costante, dal momento che deriva dalle variazioni dei parametri di processo, o della temperatura: essa deve essere una tensione non tanto progettata, quanto adattata dall'esterno. Al fine di realizzare un matching, sarà necessario utilizzare dei transistori, con una configurazione particolare. Ciò che cercheremo di fare, è ottenere con un blocco la sola  $V_{GS3}$ , con uno la sola  $v_{OD1}$ . Per  $V_{GS3}$ , non è molto difficile: basta mettere un transistor uguale a  $R_3$ , in configurazione diodo. Per  $v_{OD1}$ , si può fare qualcosa di un po' più complicato:

Infatti, si ha che:

$$v_{DS6} = V_{GS6} - V_{GS7}$$

nel caso i transistori sono matchati, questa differenza si riconduce alla sola differenza delle tensioni di overdrive:

$$= v_{OD6} - v_{OD7}$$

A questo punto, dal momento che  $v_{DS6} < v_{OD6}$ , questo  $M_6$  deve essere per forza di cose in zona triodo.  $M_6$  va configurato in modo che  $v_{DS6} = v_{OD1}$ ; scrivo dunque le equazioni, per il modello in regione triodo, tenendo conto che in questi transistori si ha una corrente pari a  $I_{REF}$ :

$$I_{REF} = \beta_n \frac{W_6}{L_6} [2(v_{GS6} - V_{TH})v_{DS6} - v_{DS6}^2]$$

imponendo il fatto che  $v_{DS6} = v_{OD1}$ , si ottiene:

$$\frac{I_{REF}}{\beta_n \frac{W_6}{L_6}} = v_{OV1}^2 + 2v_{OD7}v_{OD1}$$

a questo punto, è sufficiente sostituire  $v_{OD1}$  e  $v_{OD7}$  come funzione di  $I_{REF}$ , e il gioco è fatto:

$$\frac{I_{REF}}{\beta_n \frac{W_6}{L_6}} = \frac{I_{REF}}{\beta_n \frac{W_1}{L_1}} + 2 \frac{I_{REF}}{\beta_n \sqrt{\frac{W_1}{L_1}} \sqrt{\frac{W_7}{L_7}}}$$

Questo significa che le variazioni di  $I_{REF}$  si ripercuotono su tutto il resto del sistema; semplificando un po' di fattori, si ottiene che:

$$\frac{1}{\frac{W_6}{L_6}} = \frac{1}{\frac{W_1}{L_1}} + \frac{2}{\sqrt{\frac{W_1}{L_1}} \sqrt{\frac{W_7}{L_7}}}$$

questa è la relazione di progetto da utilizzare. Se poi si fa in modo che:

$$\frac{W_1}{L_1} = \frac{W_7}{L_7}$$

si ottiene:

$$\frac{W_6}{L_6} = \frac{1}{3} \frac{W_1}{L_1}$$

### 4.3 Riferimenti di corrente integrati

A questo punto, il problema che intendiamo affrontare è come ottenere una  $I_{REF}$ , ossia una corrente di riferimento il più possibile precisa, in un circuito.

Cosa si intende per *precisa*? Beh, sostanzialmente, la nostra richiesta è quella di non avere variazioni di questa corrente al variare della tensione di alimentazione del circuito. Dato un generico circuito, si vuole che:

$$\frac{\Delta I_{REF}}{\Delta V_{DD}} \sim 0$$

$$\frac{\Delta I_{REF}}{\delta T} \sim 0$$

Dove  $T$  è la temperatura.

Verranno a questo punto presentate tre soluzioni per questo problema.

#### 4.3.1 Soluzione banale

La prima soluzione, tanto semplice quanto inefficace, potrebbe essere la seguente:

Come generatore di corrente, utilizzare una resistenza. Valutiamo la variazione di  $I_{OUT}$  al variare di  $V_{DD}$ :

$$\begin{aligned}\frac{\Delta I_{OUT}}{\Delta V_{DD}} &= \frac{g_{m1}^{-1}}{R + g_{m1}^{-1}} g_{m2} = \\ &= \frac{g_{m2}}{1 + Rg_{m1}}\end{aligned}$$

se poi si ha che  $g_{m1}R \gg 1$ , si ha che:

$$\frac{\Delta I_{OUT}}{\Delta V_{DD}} = \frac{g_{m2}}{g_{m1}} \frac{1}{R}$$

Questa espressione è proporzionale al rapporto  $W_2/W_1$ ; al fine di quantificare, possiamo valutare la sensibilità:

$$\begin{aligned}S_{V_{DD}}^{I_{OUT}} &= \frac{\frac{\Delta I_{OUT}}{I_{OUT}}}{\frac{\Delta V_{DD}}{V_{DD}}} = \\ &= \frac{V_{DD}}{I_{OUT}} \frac{\Delta I_{OUT}}{\Delta V_{DD}}\end{aligned}$$

sostituendo ciò che è stato precedentemente visto, e notando che si ha che  $I_{OUT} = KI_{IN}$ , dove  $K$  è il rapporto di aspetto:

$$\sim \frac{V_{DD}}{KI_{IN}} \frac{g_{m2}}{Rg_{m1}}$$

ma

$$\frac{g_{m2}}{g_{m1}} \sim K$$

dunque

$$= \frac{V_{DD}}{I_{in}} \frac{1}{R}$$

ricordando a questo punto (osservando il circuito) che

$$I_{IN} = \frac{V_{DD} - V_{GS1}}{R}$$

si può trovare finalmente che:

$$S_{V_{DD}}^{I_{OUT}} = \frac{V_{DD}}{V_{DD} - V_{GS1}}$$

Se dunque si ha (come normalmente è) che  $V_{DD} \gg V_{GS1}$ , sostanzialmente questa sensibilità è circa unitaria; in altre parole, variazioni della tensione di alimentazione si ripercuotono sulla variazione di corrente con un fattore quasi unitario. Inoltre, si ha dipendenza dalla variazione della temperatura, secondo resistenza  $R$ : il 30 % di variazione rispetto a  $R$  provoca una variazione di  $I_{OUT}$  del 30 %.

### 4.3.2 Soluzione retroazionata 1

L'unico modo che al giorno d'oggi si ha per avere circuiti generanti una corrente di riferimento il più possibile indipendente dalla tensione di polarizzazione è quello di utilizzare la retroazione.

Sostanzialmente, una prima idea è la seguente:

Come funziona questo circuito? Beh, si supponga di eccitare il circuito con una tensione di polarizzazione che, al momento dell'accensione, cresce come una rampa al variare del tempo  $t$ ; se  $V_{DD} = 0$ , tutto è spento; man mano che si fa salire la tensione, i due transistori, inizialmente spenti, cambiano il proprio stato nel seguente modo:  $M_1$  è spento, e si può accendere solo dopo che su  $M_2$  passa una certa corrente; infatti, la tensione ai capi di gate e source di  $M_1$  dipende sostanzialmente da ciò che cade sulla resistenza  $R_2$ , e ciò dipende a sua volta da quanto passa sul transistor  $M_2$ . Dal momento che, per valori bassi di  $V_{DD}$  (all'inizio della rampa),  $M_1$  è completamente spento, su  $R_1$  non si ha corrente, ma manco caduta di tensione, e dunque l'intera  $V_{DD}$  è tendenzialmente sul gate di  $M_2$ ;  $M_2$  pian piano grazie a questo meccanismo si accende, e così  $V_{GS1}$  inizia ad aumentare, con un andamento dapprima quadratico. La tensione tra gate e source di  $M_1$  continua ad aumentare quadraticamente fino a quando non si supera, all'istante  $t_2$ , la tensione di soglia per il transistor 1, e in questa maniera si *accende* la reazione e così la tensione di  $M_2$  non seguirà più  $V_{DD}$ , ma si stabilizzerà a  $V_{G2}$ . Cosa capita se si ha una variazione di tensione? Sostanzialmente, se  $V_{DD}$  aumenta, aumenta la corrente su  $R_1$ , ma dunque anche la caduta di tensione su di essa; la corrente di uscita, in virtù della retroazione, rimane sostanzialmente costante: il nodo A, grazie alla retroazione, rimane sostanzialmente allo stesso potenziale. Questo, ovviamente, se il guadagno di anello è sufficientemente elevato.

Proviamo a introdurre un po' di dati: la corrente di uscita,  $I_{OUT}$ , è pari a:

$$I_{OUT} = \frac{V_{GS1}}{R_2} = \frac{1}{R_2} [V_{TH1} + v_{OD1}] = \frac{1}{R_2} \left[ V_{TH1} + \sqrt{\frac{I_{IN}}{\beta_n \frac{W_1}{L_1}}} \right]$$

si può dimostrare che:

$$S_{V_{DD}}^{I_{OUT}} = \frac{v_{OD1}}{2V_{GS1}} S_{V_{DD}}^{I_{IN}}$$

Cosa significa ciò? Beh, consideriamo un'espressione di caso peggiore: se  $S_{V_{DD}}^{I_{IN}} = 1$ , si ha generalmente che  $v_{OD1} \sim 100$  mV, e  $v_{GS1} \sim 0,8$  V; si ha qualcosa del tipo:

$$\frac{0,1}{1,6} \times 1$$

questo è ben minore di 1/10.

Questo circuito è decisamente superiore al precedente, ma presenta ancora un problema: la dipendenza diretta da parte della temperatura.

### 4.3.3 Soluzione retroazionata *quasi definitiva*

Una soluzione candidata a essere la definitiva, o quasi, potrebbe essere la seguente: dato uno specchio di corrente, si legge la corrente di uscita con un altro specchio, ottenendo ciò.

Fintanto che i transistori sono in regione di saturazione, si deve capire quanto vale questa corrente. Un'idea, potrebbe essere quella di ripetere l'analisi precedentemente proposta:

Finchè  $V_{DD} = 0$ , tutto è spento, e niente da dire. Ciò che capita è che, man mano che si va su di tensione, in teoria non si accende: questo circuito funzionerebbe, se ci fosse già una corrente al suo interno! Infatti, esso sostanzialmente funzionerebbe se vi fosse già una  $I_{IN}$  sul transistore  $M_1$  a diodo, e se vi fosse già una tensione che pilota lo specchio superiore. Servirebbero almeno delle correnti di leakage, in modo da far passare della corrente.

Questo circuito ha sostanzialmente un problema, che ora cercheremo di spiegare. Si considerino i due stadi di specchiaggio, considerati singolarmente con le proprie caratteristiche:

Teoricamente, considerando tutto ideale (e la realtà si discosta solo di poco), le due caratteristiche sono sovrapposte; dal momento che sappiamo che i punti in comune tra le due caratteristiche sostanzialmente costituiscono il punto di lavoro del circuito, questo circuito presenta infiniti punti di lavoro, come infinite sono le intersezioni tra due rette sovrapposte!

Si ha però una buona notizia: questo circuito, decisamente problematico e strano, non presenta dipendenza da parte della tensione di alimentazione.

### 4.3.4 Circuito finale

Sfruttando l'idea precedente, si modifica il circuito nel seguente modo:

Dove  $K > 1$ .

In questo circuito, si compone uno specchio tradizionale con uno specchio di tipo Widlar; l'uscita del Widlar viene specchiata dallo specchio superiore all'ingresso del Widlar. Come funziona questo circuito? Beh, sostanzialmente, esso è composto dall'incrocio di due caratteristiche: una è la caratteristica osservata nel precedente circuito, l'altra è quella del Widlar. Per comprendere il problema dobbiamo semplicemente chiederci, dunque, quale sia la transcaratteristica dello specchio di Widlar. Come funziona questo specchio? Beh, sostanzialmente, per  $I_{REF}$  molto bassa, si ha che la caduta di tensione sulla resistenza  $R_S$  che degenera  $M_2$  è molto piccola, dunque si ha una crescita, con una pendenza maggiore di 1 (dovuta al fatto che  $K > 1$ ) della  $I_{OUT}$  (la quale viene specchiata dal Widlar con un rapporto di specchiaggio pari a  $K$ ). Quando  $I_{OUT}$  diventa grande abbastanza da rendere la caduta di tensione su  $R_S$  simile alla  $V_{GS2}$ , la caratteristica tende a saturare.

Esistono due intersezioni delle caratteristiche, dunque due punti di lavoro: un punto A, corrispondente a una soluzione con tutte le tensioni e correnti nulle, e un punto B, che sarà quello sul quale vorremo progettare.

Studiamo a questo punto il sistema retroazionato, aprendo l'anello e introducendo le stesse condizioni di carico prima presenti:

Sappiamo, dalla teoria dei controlli, che il guadagno di anello  $T$  è pari a:

$$T = -\frac{i'_T}{i_T}$$

Ora, consideriamo il caso del punto di lavoro A: se si ha una variazione della corrente di ingresso  $i_T$ , a causa dello specchio Widlar si avrà in uscita una corrente  $K$  volte più grande;  $K > 1$ , dunque l'anello, nell'intorno di A, è instabile: il sistema tende a *fuggire* da A: esso è un punto di equilibrio, ma dal quale, se si esce, il sistema tende ad allontanarsi.

Consideriamo a questo punto di trovarci in un intorno di B: a fronte di un aumento di  $i_T$ , l'uscita non cresce allo stesso modo, ma ha una variazione minore, dal momento che si è in prossimità della regione di saturazione; si guardi il modello di segnale:

si ha qualcosa del genere:

$$i_{out} = g_{m2}v_{gs2}$$

$$v_{gs2} = v_{gs1} - R_S i_{out}$$

dunque

$$i_{out}(1 + g_{m2}R_S) = g_{m2}v_{gs1} = g_{m2}\frac{i_T}{g_{m1}}$$

dunque:

$$i_{out} = i'_T = \frac{g_{m2}}{g_{m1}}i_T \frac{1}{1 + R_S g_{m2}}$$

Dunque, se si ha che  $R_S g_{m2} \gg 1$ , si è sostanzialmente intorno ad A, e  $|T| \gg 1$  (il sistema è instabile); se però  $R_S g_{m2} \gg 1$ , si ha che:

$$|T| \sim \frac{1}{R_S g_{m1}}$$

e questo è minore di 1: B è un punto di lavoro **stabile** per il nostro sistema.

A questo punto, le brutte notizie: gli unici terminali disponibili per il circuito sono  $V_{DD}$  e il GND:

Dunque, è fondamentale che si possa avere la garanzia che questo circuito **parta**, ossia si tolga dal punto di lavoro banale e vada a lavorare in B.

## 4.4 Riferimenti di tensione integrati

### 4.4.1 Introduzione

A questo punto, si vuole ricavare qualcosa di duale rispetto a prima: data una tensione di alimentazione  $V_{DD}$  (che potrebbe essere variabile nel tempo per le più svariate ragioni), come si può fare per ottenere una tensione indipendente dalle variazioni di tensione di alimentazione, e magari dalla temperatura?

In questa introduzione verranno proposti sostanzialmente alcuni circuiti che di per sè non funzionano, ma che costituiscono i mattoni fondamentali per la costruzione di qualcosa di seriamente funzionante.

La prima idea che potrebbe venirci in mente è la seguente:

Ciò non ha molto senso: supponendo che le resistenze siano tecnologicamente ben realizzate, questo circuito garantisce un rapporto di partizione costante; data una variazione di tensione, la tensione di uscita varierebbe in maniera proporzionale: questo circuito, come riferimento di tensione, non ha senso.

Un'idea decisamente migliore potrebbe essere la seguente: si sa, dalle equazioni del transistor bipolare, che:

$$I_0 = I_S e^{\frac{V_{BE}}{\eta V_T}}$$

La tensione di riferimento, dato il transistor bipolare connesso a *diodo*, potrebbe essere semplicemente la tensione tra base ed emettitore del medesimo. Volendo invertire l'espressione, però, si ottiene qualcosa di sgradito:

$$V_{BE} = \eta V_T \ln \left( \frac{I_0}{I_S} \right)$$

Se si dispone di un generatore di corrente, e questa corrente viene fatta passare nel transistor a diodo, essa può essere il riferimento di tensione; purtroppo, però, questa tensione di riferimento così ottenuta dipende dalla temperatura  $T$ , sia per quanto riguarda  $V_T$ , sia per quanto riguarda  $I_S$ . Si sa, dalla teoria dei Dispositivi Elettronici, che:

$$\frac{\partial V_{BE}}{\partial T} \sim -2, 2mV/^\circ C$$

Questo significa che più si scalda, più  $V_{BE}$  diminuisce. Dal momento che a noi interessa utilizzare un eventuale riferimento in temperature comprese tra - 20 gradi e 150 gradi, si può avere qualcosa del genere:

tendenzialmente, l'andamento è piuttosto lineare.

Ora, si consideri un altro circuito:

In questo caso, considerando i due transistori esattamente uguali tra loro (o almeno, idealmente), si ha:

$$V_{BE1} = \eta V_T \ln \left( \frac{I_0}{I_{S1}} \right)$$

$$V_{BE2} = \eta V_T \ln \left( m \frac{I_0}{I_{S2}} \right)$$

Si calcoli, dunque, la differenza delle due:

$$V_{BE2} - V_{BE1} = \eta V_T \ln \left( m \frac{I_0}{I_{S2}} \frac{I_{S1}}{I_0} \right)$$

Abbiamo calcolato in questo senso la differenza poichè stiamo considerando  $m > 1$ . Semplificando (considerando i transistori tutti uguali tra loro):

$$V_{BE2} - V_{BE1} = \eta V_T \ln m$$

Dunque, la differenza delle due  $V_{BE}$ , maggiore di 0, ha una deriva di tensione positiva al variare della temperatura: il termine variabile è sostanzialmente  $V_T$ , ed esso aumenta con la temperatura.

Schema molto simile a questo è il duale: dati due generatori di corrente identici tra loro, e due transistori tali per cui

$$T_2 = mT_1$$

ossia, tali per cui  $T_2$  è costituito da  $m$  transistori uguali a  $T_1$  messi in parallelo tra loro, si ha questo schema:

In questo caso avremo:

$$V_{BE1} - V_{BE2} = \eta V_T \ln \left( \frac{I_0}{I_{S1}} \frac{I_{S2}}{I_0} \right)$$

In questo caso le due correnti di saturazione dei transistori non saranno uguali; dal momento che però si ha a che fare con un  $T_2$  costituito da  $m$   $T_1$  identici tra loro, possiamo supporre senza sbagliar molto che

$$I_{S2} = mI_{S1}$$

Dunque, si possono introdurre le dovute semplificazioni, e ottenere qualcosa di molto simile (a meno di un segno) a prima:

$$V_{BE1} - V_{BE2} = \eta V_T \ln m$$

Come mai abbiamo introdotto tutti questi circuiti? C'è un motivo: buona parte dei riferimenti di tensione per circuiti reali, attualmente utilizzati, ha un qualcosa di questo tipo:

Dove la tensione di riferimento ottenuta in uscita è un qualcosa del tipo:

$$V_0 = \alpha(K_1T - K_2T) + V_0'$$

cosa significa? In uscita vogliamo una combinazione lineare delle due tensioni, prodotte da due circuiti separati; questi due circuiti saranno uno un circuito con un andamento dalla temperatura crescente, l'altro decrescente; combinando in maniera appropriata  $K_1$  e  $K_2$ , si potrà fare in modo da avere le due derivate termiche una opposta all'altra, ed eliminare tra loro i suddetti contributi.

Questo principio di *compensazione* sarà applicato a tutti i circuiti reali, che ora studieremo.

## 4.4.2 Band-gap di Kuijk

Studiamo a questo punto il primo circuito reale per la produzione di un riferimento di tensione:

Questo è un circuito con 2 retroazioni; proponiamone un'analisi.

Se il guadagno di anello della reazione negativa è sufficientemente elevato, essendo  $A_d$  molto elevato, si ha che

$$v_d \sim 0$$

dunque, dall'equazione alla maglia:

$$v_{BE1} - v_{BE2} = R_2 I_2$$

ossia, la differenza delle tensioni tra base ed emettitore è uguale alla caduta di tensione sulla resistenza  $R_2$ . Dunque:

$$\eta V_T \ln \left( \frac{I_1}{I_{S1}} \frac{I_{S2}}{I_2} \right) = R_2 I_2$$

Le due resistenze  $R_1$  sono ovviamente uguali tra loro, e sottoposte alla stessa caduta di tensione (grazie all'operazionale), dunque  $I_1 = I_2$ ; inoltre, ipotizzando che, come precedentemente fatto

$$T_2 = m T_1$$

si ha:

$$\eta V_T \ln m = R_2 I_2$$

Ignoriamo per ora le derivate dovute al resistore  $R_2$ :  $I_2$  cresce con la temperatura  $T$ , per il primo membro; si ha che:

$$V_0 = V_{BE1} + R_1 I_1$$

dunque,  $V_{BE1}$  ha una variazione secondo  $-2,2 \text{ mV}/^\circ\text{C}$ . Calcoliamo le derivate rispetto a  $T$ :

$$\frac{\partial V_0}{\partial T} = \frac{\partial V_{BE}}{\partial T} + \frac{R_1}{R_2} \frac{k_B}{q} \ln m$$

vogliamo che questa sia nulla; dunque:

$$\frac{\partial V_{BE}}{\partial T} = -\frac{R_1}{R_2} \frac{k_B}{q} \ln m$$

Si hanno due parametri di progetto: il rapporto tra le resistenze, e  $m$ . Si può vedere che, a  $27^\circ$ , si ha qualcosa del genere:

La tensione di regolazione di questo sistema, per cui si ha una tensione di uscita indipendente, presenta un massimo in prossimità dei 27 gradi, ed è pari a 1.25 V : la tensione di band gap!

### 4.4.3 Band-gap di Widlar

Un altro modo per ottenere un risultato simile al precedente è il seguente:

in questo caso, si ha a che fare con  $T_1 = T_2$ , ossia con transistori uguali tra loro; il transistor 3 ha solo il compito di dare la tensione al circuito, mentre il 4 funge sostanzialmente da riferimento *compensante*. Si ha, data la presenza di  $R_1 \neq R_2$ , che:

$$V_{BE1} - V_{BE2} = R_3 I_2$$

dunque

$$V_T \ln \left( \frac{I_1}{I_{S1}} \frac{I_{S2}}{I_2} \right) = R_3 I_2$$

Se il circuito è ben progettato, si ha che  $V_{BE1} = V_{BE2}$ ; dunque:

$$I_1 = \frac{V_0 - V_{BE1}}{R_1}$$

$$I_2 = \frac{V_0 - V_{BE1}}{R_2}$$

dunque:

$$\frac{I_1}{I_2} = \frac{R_2}{R_1}$$

Da ciò, ricaviamo  $V_0$ :

$$V_0 = V_{BE4} + R_2 I_2 = V_{BE4} + R_2 \frac{V_T \ln \left( \frac{R_2}{R_1} \right)}{R_3}$$

Questa espressione è piuttosto simile a prima.

#### 4.4.4 Implementazione a MOS

Finora si è parlato di circuiti basati su transistori bipolari, ma in pratica, nei nostri processi, vorremmo utilizzare transistori MOS. Ciò che si potrebbe fare è semplicemente utilizzare lo stesso processo dei MOS ai quali siamo abituati, e sfruttare il fatto che si hanno degli elementi parassiti:

Dato un pMOS è possibile ricavare dei transistori PNP come questo, transistori che possono essere ottenuti *gratuitamente*, con il processo, in quanto parassiti. I MOS poi possono essere o spenti (e dunque ignorati) o usati per la polarizzazione del circuito. Un circuito basato su una di queste idee potrebbe essere il seguente:

Questo circuito sfrutta proprio questo tipo di idee; dal momento che  $M_1$  e  $M_2$  possono essere progettati come si vuole, si sceglie di imporre la stessa  $V_{DS}$ , in modo che possa scorrere la stessa corrente sui due transistori PNP ottenuti; si ha:

$$I_0 = KI_1$$

dove

$$I_1 = \frac{k_B T \ln m}{q R_2}$$

e  $K$  è il rapporto di specchiaggio dello specchio a MOS. Questo non è *bilanciato* come gli altri, dunque questo circuito, per come è stato genericamente caricato, ha semplicemente una corrente proporzionale alla temperatura: un **termometro**. Di fatto, tutti gli schemi di termometri sono sostanzialmente basati su celle di questo tipo.

#### 4.4.5 Implementazione per basse tensioni di alimentazione

I processi più moderni, come quello a 90 nm, hanno in uscita tensioni anche minori di 1,2 V, o comunque della tensione di band gap; tutti i circuiti finora visti, di fatto, sono inutilizzabili per tensioni di alimentazione inferiore ai 1,8 V: volendo tenere conto anche di questo fatto, si può utilizzare il seguente circuito, ideato nell'università di Pavia:

La cella base è uguale alla precedente, con in più un carico resistivo in uscita, e le due resistenze  $R_0$ . Analizziamolo:  $I$  è data da due contributi,  $I_1$  e  $I_2$ ; si ha:

$$I_1 = \frac{V_{BE1}}{R_0}$$

$I_2$  deriva dall'equazione alla maglia:

$$R_2 I_2 = V_{BE1} - V_{BE2}$$

da qua

$$I_2 = \frac{1}{R_2} \frac{k_B T}{q} \ln m$$

sommando i contributi:

$$I = \frac{V_{BE1}}{R_0} + \frac{1}{R_2} \frac{k_B T}{q} \ln m$$

Ciò è in grado di generare tensioni di riferimento di qualche centinaio di mV.

# Capitolo 5

## Amplificatori ad alto guadagno

### 5.1 Amplificatore differenziale

Finora sono stati analizzati amplificatori single-ended, ossia a ingresso e uscita riferiti a massa. In realtà essi non sono gli unici amplificatori esistenti.

Volendo portare un segnale a un ricevitore, le interconnessioni potrebbero essere lunghe, o vicine a un segnale digitale, dunque a un segnale con un'elevata variazione di tensione al variare del tempo; con segnali di questo genere, i parassiti potrebbero attivarsi, provocando variazioni nell'informazione contenuta nel segnale (a causa del crosstalk).

Non è tutto: se si ha un circuito con impulsi di corrente che commuta, come per esempio una pompa di carica, la corrente assorbita potrebbe avere un comportamento di questo tipo:

Un generico circuito integrato potrebbe avere un comportamento di questo tipo: ogni circuito che assorbe a impulsi funziona sostanzialmente grazie alla presenza di una capacità di bypass  $C_b$ ; in un circuito di questo tipo, si può fare l'equazione alla maglia, e vedere che:

$$V'_{DD} = V_{DD} - 2L_p \frac{di_0}{dt}$$

La tensione di alimentazione dunque avrà un andamento di questo tipo:

Questi *rimbalzi di alimentazione* potrebbero essere anche picchi di 200 o 300 mV, questi potrebbero cambiare la tensione  $V_0$  di uscita del dispositivo, e dunque variare il comportamento del circuito.

Di solito, sia questo problema sia il problema del crosstalk possono essere ridotti nella stessa maniera: mediante l'uso di strutture di tipo differenziale.

In questo caso, si definisce un'uscita come:

$$V_0 = V_{01} - V_{02}$$

Se i transistori sono polarizzati in regione attiva, si può dire che:

$$v_{01} = -g_{m1}R_{D1}\frac{v_1}{2}$$

$$v_{02} = -g_{m2}R_{D2}\frac{v_2}{2}$$

Se:

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

e

$$R_{D1} = R_{D2} = R_D$$

si ha:

$$\frac{v_0}{v_1} = -g_m R_D$$

Il guadagno, formalmente, è simile a quello di uno stadio a source comune; in pratica, questo rigetta il disturbo dovuto alla fluttuazione dell'alimentazione, dal momento che entrambi i transistori avranno la stessa fluttuazione, e la simmetria della struttura permette di eliminarlo.

Se gli ingresso del circuito fossero disturbati dallo stesso segnale comune, la differenza delle correnti al drain **non** sarebbe disturbata da ciò.

Questa struttura è molto valida, ma ha un problema: essa **raddoppia** l'area di silicio occupata.

Nel caso i due lati del circuito differiscono per solo una piccola tensione differenziale, l'unica asimmetria starebbe proprio in essi: nei generatori differenziali; tutto il resto deve essere simmetrico.

Per funzionare da amplificatore, ovviamente questo stadio deve avere i transistori in saturazione, il segnale nella dinamica, e il punto di lavoro del circuito ben definito. Per questo circuito, il punto di lavoro è sostanzialmente definito dal fatto che le tensioni di ingresso abbian un certo valore, e così quella di alimentazione, che va per le resistenze. Se si modificano le componenti di modo comune introdotte in alimentazione per questo circuito, si va a modificare anche il punto di lavoro. Per garantire sia lo stato di saturazione, sia per definire in maniera univoca il punto di lavoro, si preferisce a questa soluzione la seguente:

Si polarizza **in corrente** questo circuito:  $I_{SS}$  permette di **svincolare** il punto di lavoro dalla tensione di ingresso di modo comune. Questo perchè in questo modo il punto di lavoro non dipende più dalla tensione di ingresso: la corrente di bias è imposta dal generatore di corrente, e dunque non dipende

più dalla tensione sul punto A (a meno di non-idealità del generatore). Le variazioni di modo comune si ripercuotono dunque sul solo punto A, e il problema a questo punto diviene il problema di svincolare la corrente erogata dal generatore dalla tensione ai propri capi.

Si consideri il solo punto di lavoro: per  $v_d = 0$  e  $v_C > 0$ , in modo da avere  $M_1$ ,  $M_2$  e  $M_3$  in saturazione, se  $M_1 = M_2$  e  $R_{D1} = R_{D2} = R_D$ , la tensione di uscita di bias è:

$$V_{01} = V_{02} = V_{DD} - R_D \frac{I_{SS}}{2}$$

questo, dal momento che  $V_{GS1} = V_{GS2} = V_{GS}$ . Volendo considerare la tensione di uscita di bias di modo differenziale, se il sistema è perfettamente simmetrico, si ha:

$$V_0 = V_{01} - V_{02} = 0$$

Finchè tutto è simmetrico, la tensione di uscita **non dipende dal modo comune**.

Per quanto riguarda il modo differenziale:

$$v_{01} = V_{DD} - R_D i_{D1}$$

Ma  $i_{D1}$ , ossia la corrente totale sul drain, dipende dalla  $v_{GS}$ , dunque, data l'equazione alla maglia di ingresso:

$$v_{CM} + \frac{v_d}{2} - v_{GS1} + v_{GS2} + \frac{v_d}{2} - v_{CM} = 0$$

dunque

$$v_{GS1} - v_{GS2} = v_d$$

eliminando le tensioni di soglia, tendenzialmente tutte uguali (a patto che i transistori siano matchati):

$$v_{OD1} - v_{OD2} = v_d$$

Però, noi sappiamo che:

$$v_{OD} = \sqrt{\frac{i_D}{\beta_n \frac{W}{L}}}$$

riscrivendo le espressioni, si può dunque trovare:

$$\sqrt{i_{D1}} - \sqrt{i_{D2}} = \sqrt{\beta_n \frac{W}{L}} v_d$$

Ora, se ricordo che:

$$I_{SS} = i_{D1} + i_{D2}$$

ottengo

$$i_{D1,2} = \frac{I_{SS}}{2} \pm \left[ \beta_n \frac{W}{L} v_d \sqrt{\frac{4I_{SS}}{\mu_n C_{OX} \frac{W}{L}} - v_d^2} \right]$$

Si ricordi che tutto ciò è verificato, a patto che i transistori siano in saturazione. Volendo calcolare la corrente di uscita, si ha qualcosa del tipo:

$$i_0 = i_{D1} - i_{D2} = \beta_n \frac{W}{L} v_d \sqrt{\frac{2I_{SS}}{\mu_n C_{OX} \frac{W}{L}} - v_d^2}$$

se  $v_d \sim 0$ , ci si può ricondurre a una relazione corrente/tensione differenziale di tipo lineare; infatti, volendo linearizzare questa espressione, si può vedere che:

$$G_m = \frac{\partial i_0}{\partial v_d} = \frac{2I_{SS} - \mu_n C_{ox} \frac{W}{L} v_d^2}{\left[ \frac{4I_{SS}}{\mu_n C_{ox} \frac{W}{L}} - v_d^2 \right]^{1/2}}$$

L'andamento di questa transcaratteristica, al variare di  $v_d$ , presenta un massimo per  $v_d = 0$ :

Questa è equivalente in sostanza alla transconduttanza di un singolo transistorore, percorso da una corrente pari a  $\frac{I_{SS}}{2}$ . Questo risultato sostanzialmente afferma che se si applicano due tensioni uguali nei due ingressi, e una tensione differenziale in continua nulla, si avrà la massima transconduttanza; se si applica una tensione differenziale in continua, la transconduttanza sarà un poco più bassa; si ha tutto l'interesse a portare la tensione differenziale in continua a 0, al fine di alzare il guadagno.

Si sa che:

$$G_m(0) = g_m = \sqrt{2\mu_n C_{ox} \frac{W}{L} \frac{I_{SS}}{2}}$$

questo, dalle normali formule del MOS; come detto, la corrente è metà della  $I_{SS}$ . Il guadagno di questo stadio sarà:

$$v_0 = G_m(0)R_D = g_m R_D$$

A questo punto, possiamo passare al modello di piccolo segnale: si ha che

$$v_{gs1} - v_{gs2} = v_d$$

dunque:

$$g_{m1}v_{gs1} + g_{m2}v_{gs2} = 0$$

da qui

$$v_{gs1} = -v_{gs2}$$

risultato:

$$v_{gs1} = \frac{v_d}{2}$$

$$v_{gs2} = -\frac{v_d}{2}$$

Il nodo A si trova a massa virtuale: esso non è pilotato dalla tensione di modo differenziale, dal momento che essa non può cambiare il potenziale ai suoi capi. Osservando le uscite:

$$v_{01} = -g_m R_D \frac{v_d}{2}$$

$$v_{02} = g_m R_D \frac{v_d}{2}$$

$$A_{d1} = \frac{v_{01}}{v_d} = -g_m \frac{R_D}{2}$$

$$A_{d2} = g_m \frac{R_D}{2}$$

dunque:

$$A_d = \frac{v_0}{v_d} = \frac{v_{01} - v_{02}}{v_d} = -g_m R_D$$

### 5.1.1 Dinamica di uscita del modo differenziale

Si prenda in considerazione il seguente circuito, considerando prevalentemente il contributo di modo differenziale:

Questo circuito è simmetrico; a fronte di un aumento del segnale di tensione si avrà in uscita un'inversione (dato il verso della tensione di uscita

scelta), e un'amplificazione di una certa entità. L'obiettivo della sottosezione è quello di evidenziare quali sono i limiti della dinamica di uscita, ossia quali sono i massimi valori raggiungibili dalla tensione di uscita.

Per quanto riguarda il limite superiore, ciò che polarizza questi transistori è sostanzialmente una coppia di resistenze; le resistenze si comportano alla stessa maniera qualsiasi sia la loro tensione; la massima tensione che potrà esservi sui nodi di uscita sarà  $V_{DD}$ , ossia la tensione di alimentazione.

Per quanto riguarda il limite inferiore, si dovrà garantire che i transistori  $M_1$  e  $M_2$  siano in stato di saturazione; si deve dunque avere che:

$$v_{DS1} \geq v_{OD1}$$

stessa cosa per il transistoro 2, con gli stessi conti. Si avrà che:

$$v_{D1} = v_{O1}$$

e che

$$v_{S1} = v_{CM} + \frac{v_d}{2} - v_{GS1}$$

dunque:

$$v_{O1} \geq v_{OD1} + v_{CM} + \frac{v_d}{2} - V_{TH1} - v_{OD1}$$

dunque:

$$v_{O1} > v_{CM} - v_{TH}$$

si ha, definitivamente, che:

$$v_{CM} - V_{TH} < v_{O1} < V_{DD}$$

Volendo poi massimizzare la dinamica, si deve fare in modo che il punto di lavoro sia al centro della dinamica:

$$\bar{V}_{O1} = V_{DD} - R_D \frac{I_{SS}}{2}$$

dunque, si deve porre ciò uguale a:

$$\frac{V_{DD} + v_{CM} - V_{TH}}{2}$$

Brutte notizie: tutte queste grandezze hanno una dipendenza dalla tensione di ingresso di modo comune.

## 5.1.2 Dinamica di ingresso di modo comune

Si vuole a questo punto fare un ragionamento simile al precedente, per la dinamica di ingresso di modo comune. Si supponga per semplicità (non dovrebbe cambiare molto) che non vi sia del segnale differenziale in ingresso. Vogliamo calcolare i limiti della dinamica di ingresso di modo comune.

Se tutti i transistori sono in saturazione, si ha che:

$$v_{01} = v_{02} = V_{DD} - R_D \frac{I_{SS}}{2}$$

si ha inoltre che:

$$V_A = v_{CM} - v_{GS1}$$

Se riduciamo la tensione di ingresso di modo comune,  $V_A$  va giù con essa; al diminuire di questa i transistori  $M_1$  e  $M_2$  non hanno problemi (dal momento che la loro tensione tra drain e source aumenta, portandoli distanti dalla regione di funzionamento triodo), ma  $M_3$  sì; si avrà, come vincolo:

$$v_{DS3} = v_{CM} - v_{GS1} > v_{OD3}$$

Dunque:

$$v_{CM} > v_{GS1} + v_{OD3} = V_{TH} + v_{OD1} + v_{OD3}$$

Per quanto riguarda il limite superiore, invece, i nodi 1 e 2 sono a tensione costante fino a quando non si va in regione triodo; ora, dunque, i transistori critici saranno quelli del transconduttore differenziale:

$$v_{DS1} = V_{DD} - R_D \frac{I_{SS}}{2} - (v_{CM} - V_{GS1}) \geq v_{OD1}$$

dunque:

$$V_{DD} - R_D \frac{I_{SS}}{2} + V_{TH} > v_{CM}$$

Si ha qualcosa di questo tipo:

Purtroppo la dinamica di ingresso di modo comune e quella di uscita di modo differenziale si interferiscono: ci sarà solo una regione *comune* di funzionamento per cui l'amplificatore farà il proprio mestiere. L'obiettivo di noi progettisti sarà, oltre ad ottenere determinate prestazioni, quelli di massimizzare la sovrapposizione delle aree, evitando a tutti i costi la disgiunzione degli intervalli.

Il circuito ora progettato è un primo esempio di amplificatore completamente differenziale:

Come si vede il modello è leggermente diverso da quello dell'amplificatore operazionale; uno stadio di questo tipo (realizzato come abbiamo fatto) presenta sostanzialmente il problema dell'interferenza della dinamica di ingresso di modo comune sulla dinamica di uscita di modo differenziale.

L'obiettivo che ora ci poniamo è quello di ottenere uno stadio amplificatore differenziale con le seguenti caratteristiche:

- guadagno differenziale  $A_d$  molto elevato;
- massima dinamica di uscita di modo differenziale;
- massima dinamica di ingresso di modo comune;
- reiezione del modo comune.

## 5.2 Stadio differenziale telescopico (cascode)

Il primo obiettivo che potremmo porci è quello di aumentare il guadagno; una prima idea potrebbe essere quella di utilizzare qualcosa di simile rispetto a ciò che abbiamo usato in precedenza per gli stadi single-ended.

Si parta dal seguente stadio:

Questo stadio è semplicemente identico allo stadio precedente, utilizzando però un carico attivo al posto di un carico resistivo. Si imposta in modo tale che tutti gli specchi siano 1:1, tranne quello relativo a  $M_7$  e  $M_0$ ; a riposo,  $M_3$  e  $M_4$  devono rilasciare  $I_{SS}/2$ . Il guadagno vale:

$$A_d = -g_m r_{01} \oplus r_{03}$$

Uno stadio di questo genere ha 40 / 50 dB di guadagno; per quanto riguarda la dinamica di uscita di modo differenziale, invece:

$$v_{CM} - V_{TH} < v_{01} < V_{DD} - v_{OD3}$$

Per quanto riguarda invece la dinamica di modo comune:

$$V_{GS1} + v_{OD0} < v_{CM} < \bar{V}_{01} + V_{TH}$$

Le due dinamiche (ingresso di modo comune e uscita di modo differenziale) sono peggiorate: abbiamo ottenuto un guadagno tendenzialmente più elevato, perdendo però in dinamica.

Sotto il punto di vista del guadagno, si può fare ancora di meglio: se invece di un semplice carico attivo si utilizza un cascode, caricato con un altro cascode, si ha qualcosa del genere:

Uno stadio del genere può raggiungere guadagni dell'ordine dei 100 dB. Il guadagno elevatissimo di questo stadio, purtroppo, si paga con la dinamica: questo è il cosiddetto **amplificatore differenziale telescopico**.

Per quanto riguarda la dinamica di uscita, si ha qualcosa del tipo:

$$V_{01} + V_{B1} - V_{GS1} < v_{01} < V_{B2} + V_{SG5} - v_{OD5}$$

Da qua:

$$V_{B1} - V_{TH} < v_{01} < V_{B2} + V_{TH}$$

si ha una dipendenza **indiretta** dal modo comune. Per quanto riguarda il modo comune, invece:

$$V_{GS1} + v_{OD0} < v_{CM} < V_{B1} - v_{OD3}$$

infatti, si deve avere:

$$V_{DS1} > v_{01}$$

dunque

$$V_{B1} - V_{GS3} - (v_{CM} - V_{GS1}) > v_{OD1}$$

da qui

$$V_{B1} - V_{TH} - v_{OD3} - V_{CM} + V_{TH} + v_{OD1} > v_{OD1}$$

dunque

$$v_{CM} < V_{B1} - v_{OD3}$$

come prima scritto.

La dinamica di questo stadio è molto piccola: poche centinaia di mV! A meno che ovviamente la tensione di ingresso di modo comune non sia appartenente al range di valori appena definito: a meno che non sia dunque sostanzialmente fissata! Ciò è abbastanza irragionevole.

### 5.3 Stadio completamente differenziale come cascata di stadi differenziali elementari

Al fine di far crescere il guadagno la prima soluzione studiata è stata quella basata sullo stadio telescopico, ossia su di uno stadio cascode differenziale. Un'idea alternativa è quella di mettere in cascata più stadi differenziali, ottenendo qualcosa di questo genere:

In questo schematico, si può vedere un primo specchio differenziale, e poi un secondo, dove i transistori  $M_5$  e  $M_6$  rappresentano i **transconduttori** del secondo stadio. Questo stadio, utilizzando la simbologia prima introdotta, può essere semplicemente rappresentato in questa maniera:

Il guadagno complessivo di modo differenziale, per questi due stadi, sarà:

$$\frac{v_0}{v_d} = \frac{v_0}{v'_0} \frac{v'_0}{v_d}$$

volendo esplicitarlo:

$$\frac{v_0}{v_d} = (g_{m6}r_{06} \oplus r_{08})(g_{m1}r_{01} \oplus r_{03})$$

Per quanto riguarda le dinamiche, si ha qualcosa di questo genere: per il secondo stadio, si ha

Questo vale per i singoli transistori  $M_6$  e  $M_7$ . Si ha però una nota: sappiamo che la tensione di uscita è data da:

$$v_0 = v_{01} - v_{02}$$

e che quando si ha il minimo valore di uscita di  $v_{01}$  si ha il massimo di  $v_{02}$ , e viceversa; dovremo dunque dire che:

$$v_0 > v_{OD7} - (V_{DD} - v_{OD6}) = -V_{DD} + 2v_{OD}$$

dove in realtà si dovrebbero sommare  $v_{OD6}$  e  $v_{OD7}$ . Si può ragionare allo stesso modo e trovare che:

$$v_0 < V_{DD} - v_{OD5} - v_{OD8}$$

dunque:

$$-V_{DD} + 2v_{OD} < v_0 < V_{DD} - 2v_{OD}$$

Abbiamo ottenuto due ottimi risultati: da un lato, abbiamo aumentato il guadagno dello stadio (essendo esso dato dal prodotto dei guadagni di due stadi differenziali), ma dall'altro abbiamo eliminato la dipendenza della dinamica di uscita di modo differenziale dalla tensione di ingresso di modo comune: nelle formule precedentemente presentate infatti la dipendenza dalla tensione di modo comune è sparita.

Per quanto riguarda invece la tensione di ingresso di modo comune, per il limite inferiore si può dire come al solito che:

$$v_{CM} > V_{GS1} + v_{OD0}$$

Per quanto riguarda invece il limite superiore, si dovrà tenere conto della tensione di uscita media ai nodi 1 e 2, ossia della tensione di uscita a riposo per il primo stadio differenziale. Supponiamo che

$$\bar{v}'_{01} = \bar{v}'_{02}$$

In questo caso, si ha che:

$$V_{DS1} = V_{DS2} = \bar{v}'_{01} - (V_{GS1} - v_{CM}) > v_{OD1}$$

dunque, si può semplicemente dire che

$$v_{CM} < \bar{v}'_{01} + V_{TH1}$$

Abbiamo un problema: la tensione di uscita  $\bar{v}'_{01}$  non è nota, dal momento che essa discende dall'intersezione di due curve quasi piatte.

Con piccolissime variazioni di  $V_{GS}$  o di  $i_D$  si hanno enormi variazioni di  $V_{DS}$ , di  $v_0$ , dal momento che l'intersezione tra queste curve è molto *mobile*: si può andare in triodo per variazioni minime.

A questo punto, facciamo alcune considerazioni: si supponga di indurre una variazione della tensione di uscita (di una delle due uscite) di 3 volt (sostanzialmente rail-to-rail, dall'alimentazione inferiore a quella superiore); se il secondo stadio guadagnasse per esempio 100, la variazione che si avrà sulle uscite del primo stadio sarà di 30 mV: molto bassa:

$$\Delta V_{SG5} = \frac{3V}{100} = 30mV$$

Se questo amplificatore si collega in feedback, per esempio in questo modo, supponendo che  $A_{d1}A_{d2} = 80dB$  (per esempio):

Si ha, supponendo che i transistori siano tutti in regione di saturazione, una variazione di 30 mV nell'intorno del punto di lavoro.

In questo sistema la tensione di uscita del primo stadio non è condizionata dalla tensione di modo comune: di fatto, questo accade dal momento che si ha un secondo stadio, che amplificherà le variazioni di tensione in uscita dal primo stadio, provocando le variazioni sull'uscita.

Qual è il problema? Dire che il transistoro sia in saturazione non è banale, per il discorso prima fatto: non siamo in grado di conoscere il valore medio della tensione di uscita del primo stadio, dunque non possiamo neanche dire con certezza di avere i transistori del primo stadio in saturazione. Resta dunque il problema di fissare le tensioni di polarizzazione. Questo problema si presenta per qualsiasi stadio differenziale: le retroazioni sono efficaci per il solo modo differenziale, ma non sono in grado di stabilizzare, di controllare la tensione di modo comune.

Si consideri ora uno stadio elementare:

Controllare anche in questo stadio più semplice le uscite non è possibile.

Ciò che si deve fare è utilizzare un circuito di polarizzazione per il solo modo comune, al fine di fissare, di controllare la **sola** componente di modo comune. Negli stadi differenziali, un buon obiettivo sarebbe anche quello di disaccoppiare le retroazioni per il modo comune e quelle per il modo differenziale: fare in modo che la reazione per ciascun modo sia insensibile rispetto all'altro modo. Per quanto riguarda il modo comune, l'obiettivo è fissare un valore di  $V_0$  tale per cui:

$$V_0 = \frac{V_{01} + V_{02}}{2}$$

L'obiettivo è quello di controllare le correnti dei transistori in modo da fissare le  $V_{01}$  e  $V_{02}$  al valor medio, cercando di non interferire con il modo differenziale.

Una soluzione può essere la seguente:

Si può prendere  $V_0$  come la tensione media tra le due, semplicemente utilizzando una rete composta da due resistenze uguali: esse prendono la tensione, e in questo modo, prendendo un nodo in mezzo alle due, si **media-no** le due tensioni; mediante un amplificatore differenziale (un operazionale per esempio) si pone questa tensione  $V_0$  uguale a una certa tensione di riferimento  $V_{REF}$ , ottenuta per esempio mediante uno dei circuiti di riferimento precedentemente presentati. Questo operazionale piloterà un transistor, che potrà essere collegato al nodo A, in modo da regolare la corrente in questo punto, in modo da regolare le tensioni di uscita al valore desiderato.

A seconda di cosa si sceglie come amplificatore di confronto con  $V_{REF}$ , si ottengono risultati più o meno precisi: se l'obiettivo è una tensione molto precisa, allora ben verrà l'operazionale; se si ridurrà il guadagno (introducendo un normale amplificatore differenziale) si ridurrà anche la precisione con cui si otterrà la tensione media; ciò tendenzialmente non è molto problematico.

Si consideri dunque il seguente schema di funzionamento:

In questo modo, si hanno  $\bar{v}'_{01} = \bar{v}'_{02}$ ; volendole fissare a circa 1.65 volt (la metà tra 0 e 3.3):

$$V_{SG5} = V_{SG6} = V_{TH} + v_{OD} + \sqrt{\frac{2I_{SS}}{2\mu_n C_{ox} \frac{W}{L}}} = \sqrt{\frac{I_{SS}}{\mu_n C_{ox} \frac{W}{L}}}$$

dunque:

$$\bar{v}'_{01} = \bar{v}'_{02} = V_{DD} - \bar{V}_{SG5}$$

Ora che abbiamo ottenuto la tensione media, qual è la massima tensione di ingresso di modo comune accettabile per questo stadio? Beh, semplicemente:

$$v_{CM} < (V_{DD} - V_{SG5}) + V_{THn}$$

Mettendo dei numeri, si può vedere meglio che:

$$v_{CM} < 3.3V - (V_{THp} - V_{THn}) - v_{OD5}$$

Ciò è prossimo alla  $V_{DD}$ !

Utilizzare stadi in cascata può essere interpretabile nella seguente maniera: il primo stadio traduce il segnale differenziale in un segnale single-ended; il secondo si comporta come un source comune, il quale non ha problemi di modo comune, essendo l'uscita da esso indipendente.

### 5.3.1 Feedback di modo comune - reprise

Di fatto il circuito finora presentato ha qualche controindicazione: se si introducono queste resistenze  $R$  sulle uscite, le uscite verranno sostanzialmente caricate da una resistenza pari a  $2R$ ; questa cosa si va a ripercuotere sul guadagno; infatti:

$$A_d = g_{m1}(r_{01} \oplus r_{03} \oplus (2R))$$

$A_d$  viene drasticamente ridotto dal parallelo di queste grandezze. Come risolvere questo problema? Beh, sostanzialmente, utilizzando circuiti più complicati ma efficaci.

Una prima soluzione potrebbe essere la seguente:

Il nostro problema è che le resistenze abbassano il guadagno; per ottenere un risultato uguale al precedente, senza ridurre il guadagno, si possono utilizzare degli stadi source follower come buffer, polarizzati da altri due transistori.

Un'altra soluzione potrebbe essere la seguente:

In questo caso, si ha che, quando si ha una variazione delle tensioni di uscita, i due stadi differenziali vengono sbilanciati rispetto alla tensione di riferimento. La corrente del transistorore  $M_x$  è la quantità regolata (la somma delle correnti dei due differenziali). Ad anello chiuso, si tende, dalla situazione di sbilanciamento, ad andare verso la  $V_{ref}$ .

Purtroppo, i circuiti di controllo del modo comune in realtà non sono indipendenti da quelli di modo differenziale: essi limitano in qualche modo la dinamica del circuito. Se uno qualsiasi dei transistori, anche nell'ambito

del circuito di reazione di modo comune, può capitare che si vada fuori dinamica: se il circuito di retroazione non funziona, si perde il controllo del modo comune, e dunque si rischia di mandare fuori linearità l'intero circuito.

## 5.4 Stadi folded cascode - rail-to-rail

A questo punto si può introdurre un'ulteriore soluzione al problema della realizzazione di stadi completamente differenziali: gli stadi folded-cascode. L'idea sostanzialmente è la seguente:

I transistori 9 e 10 devono fornire sia alla coppia differenziale, sia allo stadio di guadagno in tensione, che sarà costituito da uno stadio cascode.  $V_{B1}$  e  $V_{B2}$  saranno tensioni di polarizzazione ottenute a partire da un'ideale rete di polarizzazione.

Il sostanziale salto è stato passare dallo stadio telescopico a questo, e questo per tentare di ottenere un guadagno elevato, ma anche una dinamica elevata.

Si ricordi lo stadio folded cascode:

Sostanzialmente quanto più si apre il transistor  $M_A$ , tanto più vi è corrente su esso. La corrente in esso dipende sostanzialmente dalla  $v_{GS}$ ; il guadagno sarà:

$$A_V = G_{mA} (r_{0c} g_{mc} (r_{0B} \oplus r_{0A}))$$

Per questa struttura, si può fare un ragionamento molto simile, e ottenere:

$$A_{d,0} = g_{m1} [(r_{07} g_{m7} (r_{09} \oplus r_{01})) \oplus (g_{m5} r_{05} r_{03})]$$

### 5.4.1 Dinamica di uscita di modo differenziale

Si vuole a questo punto valutare la dinamica di uscita per il modo differenziale. Come si potrà vedere, la tensione di uscita dipende sostanzialmente dalle tensioni di polarizzazione  $V_{B1}$  e  $V_{B2}$ . Al fine di garantire il funzionamento del circuito, si deve garantire la linearità dei transistori; si consideri per esempio il transistor 5:

$$v_{DS5} > v_{OD5}$$

ma

$$v_{DS5} = v_{01} - (V_{B2} - v_{GS5}) > v_{OD5}$$

ordinando e svolgendo  $v_{GS5} = v_{TH5} + v_{OD5}$ , si ottiene

$$v_{01} > V_{B2} - V_{TH5}$$

Come si vede, ciò è legato a una delle tensioni di polarizzazione: se si varia in modo opportuno la tensione di polarizzazione, allora è possibile aumentare la dinamica di modo differenziale di uscita. Ciò che si può fare nella fattispecie è scegliere  $V_{B2}$  in modo da portare il transistor a funzionare al bordo della regione triodo; se dunque

$$V_{B2} = V_{THn} + v_{OD5} + V_{OD3}$$

allora

$$v_{01} > v_{OD3} + v_{OD5}$$

Le stesse considerazioni possono essere fatte anche per il limite superiore della dinamica:

$$v_{SD7} > v_{OD7}$$

dunque

$$v_S = V_{B1} + v_{SG7}$$

$$v_D = v_{01}$$

da qua

$$V_{B1} + v_{GS7} - v_{01} > v_{OD7}$$

da qua

$$v_{01} < v_{B1} + V_{TH7}$$

se polarizzo dunque in regione triodo:

$$v_{01} < V_{DD} - (v_{OD9} + v_{OD7})$$

## 5.4.2 Dinamica di ingresso di modo comune

Cosa si può dire, per quanto riguarda questo stadio, sulla dinamica di ingresso di modo comune? Beh, come sempre, si può vedere che:

$$v_{GS1} + v_{OD0} < v_{CM}$$

Per quanto riguarda il limite superiore della dinamica, invece, si avrà dipendenza da  $V_x$ ; si dovrà verificare il fatto che

$$v_{DS1} > v_{OD1}$$

dunque

$$V_{DD} - v_{OD9} - (v_{CM} - v_{GS1}) > v_{OD1}$$

da qua

$$V_{DD} - v_{OD9} + V_{TH1} > v_{CM}$$

dunque:

$$v_{GS1} + v_{OD0} < v_{CM} < V_{DD} - v_{OD9} + V_{THn}$$

Ciò è molto interessante: questa tensione, di fatto, è **superiore** alla tensione di alimentazione; in altre parole, è possibile introdurre nel circuito una tensione di modo comune di poco superiore alla tensione di alimentazione, senza comunque mandarlo fuori linearità. Questo è il grande vantaggio degli amplificatori folded cascode.

Questo amplificatore è fully-differential, dunque si avrà bisogno di un circuito di reazione per il modo comune. In questi casi è possibile o utilizzare i circuiti precedentemente presentati, o qualcosa di più economico e semplice, per quanto anche impreciso. Si consideri la seguente idea:

Si degenera di source uno stadio, con una resistenza  $R_x$  variabile (che, come si vede, può essere realizzata mediante un transistoro in regione triodo); comandando la resistenza con la tensione di uscita, facendo in modo che:

$$R_x = \frac{K}{V_0}$$

si può far in modo da abbassare  $R_x$  al crescere della  $V_0$ , riducendo la tensione di uscita.

Si applichi questa idea al nostro stadio:

A seconda di come variano  $v_{01}$  e  $v_{02}$ , le resistenze variano, e così si realizza la reazione, a patto che i transistori siano in regione triodo. La resistenza equivalente sarà:

$$R_{eq} = R_{DSA} \oplus R_{DSB}$$

dove

$$R_{DS} = \frac{1}{\beta_n \frac{W}{L} (v_{GS} - V_{TH})}$$

dunque, si può scrivere:

$$R_{eq} = \frac{1}{\beta_n \frac{W}{L} \Big|_A (v_{01} - V_{THn}) + \beta_n \frac{W}{L} \Big|_B (v_{02} - V_{THn})}$$

A questo punto, si assuma:

$$\frac{W}{L} \Big|_A = \frac{W}{L} \Big|_B = \frac{W}{L}$$

si ha:

$$R_{eq} = \frac{1}{\beta_n \frac{W}{L} (v_{01} + v_{02} - 2V_{THn})}$$

Ora: come è possibile, da queste formule, dimensionare i transistori? Beh, si può scrivere che:

$$v_{GS12} - v_{GS3} = R_{eq} I_{eq}$$

questo, per l'equazione alla maglia. Qui si fa intervenire una prima approssimazione: si considera il fatto che  $I_{eq}$ , la somma delle due correnti, valga  $2I_B$ ; si avrà dunque:

$$V_{THn} + v_{OD12} - (V_{THn} + v_{OD3}) = 2R_{eq} I_B$$

dunque, passando alle formule esplicite delle tensioni di overdrive:

$$\sqrt{\frac{I_B}{\beta_n \frac{W}{L} \Big|_{12}}} - \sqrt{\frac{I_B}{\beta_n \frac{W}{L} \Big|_3}} = 2R_{eq} I_B$$

A questo punto, al fine di alleggerire la notazione, si definisca un rapporto dei rapporti di aspetto,  $K$ :

$$K \triangleq \frac{\frac{W}{L} \Big|_{12}}{\frac{W}{L} \Big|_3}$$

si ottiene:

$$\sqrt{\frac{1}{\beta_n \frac{W}{L} \Big|_3}} \left( 1 - \frac{1}{\sqrt{K}} \right) = 2R_{eq} \sqrt{I_B}$$

Combinando le due equazioni, fissando  $K$ , si può ricavare il rapporto di specchiaggio del transistor  $A$ .

Si può vedere che questa equazione avrà dipendenza dai parametri di progetto; ciò è negativo, dal momento che essi hanno molteplici dipendenze, tuttavia la semplicità di questo circuito, e le basse richieste in termini di area, lo rendono comunque molto appetibile.

Esistono altri modi di utilizzare questa tecnica:

1. In questo caso si riduce la dinamica di modo differenziale; in questo caso, utilizzando questa configurazione, si va a ridurre il limite superiore; nel caso precedente, quello analizzato prima di questa osservazione, il limite inferiore. Un modo per migliorare ciò sarebbe ridurre la resistenza, ma ciò significa aumentare le dimensioni dei transistori, dunque il carico capacitivo proposto.
2. In questa configurazione si riduce il limite inferiore della dinamica di ingresso di modo comune.

Si consideri, a questo punto, questa struttura:

In questo caso si utilizza una coppia differenziale di tipo  $p$ , che inietta corrente nei nodi  $z$  e  $w$ . Si provi, per questa struttura, a ricavare la dinamica di ingresso di modo comune. Si vede che il limite superiore della dinamica è dettato dal generatore di corrente, ossia da  $M_{0p}$ ; sarà:

$$v_{CM} < V_{DD} - (v_{OD0p} + v_{SG1p})$$

per quanto riguarda il limite superiore, si ha qualcosa del tipo:

$$v_{CM} + v_{SG1p} - v_{OD3} > v_{OD1}$$

dunque, ordinando e sviluppando

$$v_{CM} > v_{OD3} - V_{THp}$$

molto interessante: si ha qualcosa al di sotto della tensione di riferimento, della tensione negativa di alimentazione. Questo significa che in questa struttura, duale alla precedente, si hanno le dinamiche di ingresso di modo comune duali.

L'idea che dunque intendiamo utilizzare è la seguente:

Questo schema sostanzialmente mette insieme le due topologie, con la speranza di trarne il meglio. In effetti, è proprio così: se si lavora al centro della dinamica, entrambe le coppie differenziali saranno accese, ossia lavoreranno, amplificheranno, e si avrà un certo guadagno; man mano che ci si

allontana da questo punto ideale, ossia man mano che si aumenterà o diminuirà la tensione di modo comune, si avrà che una delle due coppie sarà *più accesa* dell'altra, ma dunque solo una contribuirà al guadagno globale, che però sarà comunque ragionevole.

Il fondamentale problema di questa topologia è stato appena spiegato: se una delle due coppie è spenta, il guadagno varierà: il guadagno è funzione della tensione di ingresso di modo comune. Si ha, se entrambe le coppie differenziali sono accese:

$$A_{d0} = g_{mn}R_{out} + g_{mp}R_{out} = (g_{mn} + g_{mp})R_{out}$$

Questo, però, se entrambi i differenziali sono accesi. Più generalmente, si avrà una transconduttanza equivalente  $G_m(v_{CM})$  di questo tipo:

Se poi il sistema è retroazionato, si può sperare che questa  $G_m$ , che sostanzialmente andrà a lavorare sul guadagno di anello, sia sufficientemente grande da rendere la funzione di trasferimento del sistema come quella desiderata: il  $A_f$  dovrà essere idealmente uguale al  $A_\infty$ .

## 5.5 Amplificatori single-ended - parametri

Si vuole a questo punto studiare come realizzare, a partire da uno stadio completamente differenziale, uno stadio single-ended. Si riparta da qui:

Esistono almeno due soluzioni che permettano di ottenere, a partire da questo stadio, uno stadio con uscita riferita a massa o a  $V_{DD}$ . Le si osservi:

Sotto il punto di vista dei guadagni, queste due soluzioni sono molto simili; ciò che rende una soluzione preferibile all'altra è invece la dinamica: la soluzione a destra ha migliori performance sotto il punto di vista della dinamica di modo comune. Ciò che si può osservare infatti è che la soluzione a destra, per quanto richieda più area, presenta due specchi di corrente, ciascuno costituente il carico di ciascun transistor dello specchio; ciò permette sostanzialmente, dal momento che si carica mediante il lato debole, di fissare la tensione di drain dei transistori della coppia; ciò non è fatto nel circuito a sinistra, dove non si fissa la tensione.

A partire dagli stadi più elaborati, utilizzati precedentemente (con secondo stadio di amplificazione ad alto guadagno, cascode), si può fare qualcosa del genere:

Si può utilizzare uno *specchio migliorato*, mediante questo semplice collegamento: ciò permette di migliorare lo stadio sotto il punto di vista della stabilità della tensione di uscita, e di renderlo single ended.

### 5.5.1 Amplificazione di modo comune

La reiezione della tensione di modo comune è effettuata dallo stadio di ingresso; volendo dunque discutere sull'amplificazione di modo comune, è sufficiente riferirsi alla sola coppia differenziale in ingresso al sistema:

Ciò che di solito si fa per effettuare l'analisi è passare al modello di piccolo segnale:

si ha:

$$v_{o1} = -R_{D1}i_{d1}$$

$$v_{o2} = -R_{D2}i_{d2}$$

e

$$v_o = v_{o1} - v_{o2}$$

A questo punto, come siamo soliti fare, possiamo scrivere l'equazione al nodo  $A$ , e vedere che:

$$g_{m1}v_{gs1} + g_{m2}v_{gs2} = \frac{v_{cm} - v_{gs1}}{R_{SS}}$$

Per ispezione, si può vedere che:

$$v_{gs1} = v_{gs2}$$

dunque:

$$v_{gs1} = \frac{v_{cm}}{1 + R_{SS}(g_{m1} + g_{m2})}$$

Ora: possiamo calcolare la tensione di uscita del transistor 1 come:

$$v_{o1} = -R_{D1}g_{m1} \frac{v_{cm}}{1 + R_{SS}(g_{m1} + g_{m2})}$$

$$v_{o2} = -R_{D2}g_{m2} \frac{v_{cm}}{1 + R_{SS}(g_{m1} + g_{m2})}$$

dunque, è possibile calcolare l'amplificazione di modo comune come:

$$A_{cm} = \frac{v_o}{v_{cm}} = \frac{R_{D2}g_{m2} - R_{D1}g_{m1}}{1 + R_{SS}(g_{m1} + g_{m2})}$$

Questa è l'amplificazione di modo comune.

Si hanno due possibilità, per ridurre questa amplificazione: avere  $R_{SS}$  molto elevata (dunque un generatore di corrente sostanzialmente ideale), oppure avere un circuito molto simmetrico, ossia avere

$$R_{D2}g_{m2} = R_{D1}g_{m1}$$

in questo caso, indipendentemente dal generatore, è possibile reiettare il modo comune. Per fare ciò si dovrebbe avere un ottimo matching dei nostri dispositivi.

In realtà, se si ricorda la discussione iniziale, il matching è importante, ma non è tutto: l'amplificatore differenziale è basato sull'amplificazione degli sbilanciamenti, dunque delle dissimmetrie, introdotte da un segnale di modo differenziale: intrinsecamente, quando si eccita un amplificatore differenziale, si genera una dissimmetria: ciò rende, di fatto, la reiezione del modo comune impossibile (in termini di simmetrizzazione totale del circuito).

Un caso particolare potrebbe essere quello di avere resistenze perfettamente matchate; in questo caso, si avrebbe:

$$R_{D1} = R_{D2} = R_D$$

In tal caso, il guadagno di modo comune si riduce a:

$$\frac{v_o}{v_{cm}} = \frac{R_D(g_{m1} - g_{m2})}{1 + R_{SS}(g_{m1} + g_{m2})} = \frac{\Delta g_m R_D}{1 + 2R_{SS}\bar{g}_m}$$

Volendo dunque quantificare le prestazioni del circuito in termini di reiezione del modo comune, si definisce di solito un parametro, detto CMRR:

$$CMRR = \left| \frac{A_d}{A_{CM}} \right|$$

in questo caso, date le approssimazioni precedentemente introdotte, si ottiene:

$$CMRR = \frac{(1 + 2\bar{g}_m R_{SS})\bar{g}_m}{\Delta g_m}$$

Questa è l'espressione del common mode rejection ratio.

Si consideri a questo punto il seguente circuito, rappresentante una coppia differenziale eccitata da un solo segnale di modo comune:

Si consideri a questo punto il modello di segnale di questo circuito: in esso stiamo considerando le resistenze tra drain e source solo per i transistori  $M_3$  e  $M_4$ , ma non per i transistori  $M_1$  e  $M_2$ . Come mai? Prima di tutto, alcune ipotesi introduttive: stiamo ipotizzando che i transistori  $M_1$  e  $M_2$ ,

come anche  $M_3$  e  $M_4$ , siano perfettamente matchati. Ciò che si può vedere che è che i transistori  $M_1$  e  $M_2$  hanno anche la stessa  $v_{gs}$ ; si avrà dunque che:

$$i_{d1} = g_{m1}v_{gs1}$$

$$i_{d2} = g_{m2}v_{gs2}$$

$$g_{m1}v_{gs1} + g_{m2}v_{gs2} = \frac{v_{cm} - v_{gs1}}{R_{SS}}$$

e

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

dunque, si può ricavare che:

$$2g_m v_{gs} R_{SS} = v_{cm}$$

Questo circuito, a tutti gli effetti, è assolutamente equivalente a questo, ridisegnato:

Ciò ci permette di giustificare con facilità il fatto che non sono state aggiunte quelle resistenze: in pratica ci sono, però se ne può anche non tenere conto, dal momento che la degenerazione di source aumenta la resistenza vista dal drain, rendendo trascurabili quelle tra drain e source. Si ha che

$$v_o = r_{o4}i_o$$

dove

$$\begin{aligned} i_o = g_{m4}v_{sg4} - i_{d2} &= g_{m4}(i_{d1}r_{o3} \oplus g_{m3}^{-1}) - i_{d1} = \left( g_{m1} \frac{1}{\frac{1}{r_{o3}} + g_{m3}} - 1 \right) i_{d1} = \\ &= \frac{g_{m4} - (g_{m3} + r_{o3}^{-1})}{r_{o3}^{-1} + g_{m3}} \frac{g_{m1}v_{cm}}{1 + 2R_{SS}g_{m1}} \end{aligned}$$

per il matching, si può dire che  $g_{m3} = g_{m4}$ ; alla fine di tutto, rimane:

$$\frac{v_o}{v_{cm}} = \frac{1}{(1 + r_{o3}g_{m3})(1 + 2R_{SS}g_{m1})} r_{o4}g_{m1}$$

a questo punto, si consideri, come ipotesi, il fatto che  $r_{o3}g_{m3} \gg 1$ ; si ha:

$$\frac{v_o}{v_{cm}} = \frac{g_{m1}}{g_{m3}(1 + 2R_{SS}g_{m1})}$$

Si può dunque ricavare che:

$$A_{cm} = \frac{(g_{m1} - g_{m2})r_{o3} - \frac{g_{m1}}{g_{m3}}}{1 + R_{SS}(g_{m1} + g_{m2})}$$

dunque:

$$CMRR = (r_{o4} \oplus r_{o2})(g_{m3}(1 + 2R_{SS}g_{m1}))$$

Passando ai parametri tecnologici, si può scrivere:

$$CMRR = \frac{2}{v_{OD3}} \frac{2}{v_{OD1}} \frac{1}{\lambda_0} \frac{1}{\lambda_2 + \lambda_4}$$

Il nostro obiettivo, in sede di progetto, è rendere molto grande questo parametro; per fare ciò, purtroppo, dobbiamo andare a scapito dell'area: è necessario ridurre i parametri  $\lambda$ , e per fare ciò è necessario allungare i transistori; allo stesso modo sarebbe necessario ridurre la tensione di overdrive, e per questo motivo è necessario aumentare il rapporto di specchiaggio. Alla fine della fiera, ciò che si deve fare è aumentare l'area del dispositivo, cosa critica per noi.

### 5.5.2 PSRR: Power Supply Rejection Ratio

L'amplificazione di modo comune è legata alla reiezione di disturbi aggiunti alle alimentazioni; il fatto che possano essere presenti dei disturbi nelle alimentazioni è un qualcosa di cui non ci siamo ancora occupati, quantomeno in questo capitolo: se ne è già discusso quando si parlava di riferimenti di corrente o di tensione.

Quando si parla del progetto degli amplificatori, bisogna purtroppo tenere in conto il fatto che le tensioni di alimentazione sono affette da disturbi:

Una cosa del genere potrebbe capitare per esempio in sistemi di questo tipo:

Se in un sistema si hanno assieme blocchi digitali e blocchi analogici, alimentati mediante la stessa tensione di alimentazione, può capitare una cosa piuttosto spiacevole: solitamente i circuiti digitali assorbono corrente in maniera impulsiva, ossia mediante impulsi; le tensioni di alimentazione, alle quali saranno collegati anche i circuiti di tipo analogico, avranno dunque un valore di questo tipo:

$$v_{AL} = V_{DD} - L_{DD} \frac{d}{dt} i_{CORE} - (L_{SS} \frac{d}{dt} i_{CORE} + V_{SS})$$

Ciò è particolarmente importante nei sistemi a bassa tensione. Il fatto che la durata di questi impulsi di corrente è prossima al tempo di commutazione delle porte, il quale è dell'ordine dei nanosecondi, rende particolarmente importanti questi effetti.

Si noti che disaccoppiare le tensioni di alimentazione per la parte digitale e per quella analogica non è sufficiente: in realtà, essendo tutti i circuiti nello stesso chip di silicio, si avrà ancora un accoppiamento parassita attraverso il substrato, detto *crosstalk del substrato*.

Ora: spegniamo le sorgenti all'ingresso e definiamo, dati:

$$\begin{cases} v_d = 0 \\ v_{cm} = 0 \\ v_{ss} = 0 \end{cases}$$

$$A^+ = \frac{v_o}{v_{dd}}$$

$$A^- = \frac{v_o}{v_{ss}}$$

Si definisce a questo punto il PSRR come:

$$PSRR^+ = \left| \frac{A_d}{A^+} \right|$$

$$PSRR^- = \left| \frac{A_d}{A^-} \right|$$

Definiti questi parametri, si può dire che il disturbo sull'alimentazione  $v_{dd}$  porta sull'uscita delle variazioni; questo fatto può essere modellato in una maniera più intelligente, in modo da poter tener conto di nozioni già note e semplificarci i calcoli: ciò a tutti gli effetti coincide con il considerare un amplificatore alimentato mediante tensioni di alimentazione non corrotte, con però dei disturbi aggiuntivi riportati sull'ingresso:

Il fatto che si conosce l'entità del disturbo permette di valutare la sua entità rispetto al segnale di ingresso, ossia al segnale di modo differenziale che ecciterà il sistema. Questo calcolo, a tutti gli effetti, può ricordare ciò che si fa per la stima del rumore.

Si consideri per esempio una struttura single ended, come la seguente:

Questo circuito andrebbe analizzato mediante il modello di piccolo segnale nella sua interezza. Si noti che i due stadi di amplificazione sono polarizzati dalla stessa tensione: questo fatto potrebbe sembrare una cattiva scelta, ma vedremo che non è così.

Si consideri prima di tutto il transistoro a diodo, unito con il generatore di corrente (il transistoro): si ha, considerando un modello di segnale, qualcosa di questo genere:

questo circuito ci mostra che, di fatto,  $v_{SG8}$  è modulata, e così con lei saranno modulate anche le correnti sui lati forti: il lato debole dello specchio trasduce la corrente che vi si impone in una tensione, che dovrà pilotare gli altri transistori; il fatto che questa tensione è variabile con la tensione di alimentazione, dunque, andrà a ripercuotersi su  $v_{gs8}$  nella seguente maniera:

$$\frac{v_{gs8}}{v_{dd}} = \frac{1}{1 + r_{o9}g_{m8}}$$

Se le tensioni sono modulate, lo saranno dunque anche le correnti. Si avrà dunque, tenendo conto di questo effetto, un equivalente di piccolo segnale di questo tipo:

Di fatto si avrà una ripercussione sull'uscita di ciò che è stato prima detto, ma per due motivazioni: si hanno di fatto due percorsi, uno *indiretto*, e uno *diretto* verso l'uscita.

- Per quanto riguarda il percorso indiretto, sostanzialmente il fatto che si ha una variazione di  $v_{dd}$  comporta una variazione di modo comune sull'uscita: la situazione descritta nello schema è di fatto analoga a quella studiata in precedenza, ma ora con coppie differenziali di tipo  $p$ : anzichè introdurre un segnale di modo comune sotto forma di variazione di  $v_{gs}$ , si introduce una variazione della corrente in ingresso a una coppia differenziale. Si avrà:

$$\left. \frac{v_o}{v_{dd}} \right|_{diff} = \frac{v_o}{v_o'} \frac{v_o'}{v_{dd}}$$

due termini: il secondo è l'amplificazione del secondo stadio, il secondo è l'amplificazione di modo comune del primo stadio: due quantità note! Si avrà:

$$= g_{m6}(r_{o7} \oplus r_{o6}) \frac{g_{m1}}{g_{m3}} \frac{-v_{cm}}{1 + 2g_m r_{o5}}$$

- Per quanto riguarda il contributo, si vede che, per quanto riguarda il secondo stadio, si ha semplicemente qualcosa del tipo:

$$\left. \frac{v_o}{v_{dd}} \right|_{II} = \frac{r_{o6}}{r_{o6} + r_{o7}}$$

Il guadagno risultante sarà:

$$A^+ = \left[ \frac{1}{1 + \frac{r_{o7}}{r_{o6}} - \frac{1}{2} \frac{g_{m6}(r_{o7} \oplus r_{o6})}{g_{m3}r_{o5}}} \right]$$

Imponendo l'eguaglianza dei due termini, si può avere:

$$A^+ = 0$$

dunque, si può ottenere un PSRR che tende a infinito. Ordinando le espressioni e sostituendo tutto, si ha:

$$\frac{1}{PSRR} = \frac{1}{g_{m1}(r_{o2} \oplus r_{o4})g_{m6}r_{o7}} - \frac{1}{2g_{m1}(r_{o2} \oplus r_{o4})g_{m3}r_{o5}}$$

se si chiede che

$$g_{m6}r_{o7} = 2g_{m3}r_{o5}$$

si ottiene un PSRR infinito. Per questo motivo la scelta iniziale di polarizzare i due stadi con lo stesso circuito di polarizzazione risulta essere saggia: polarizzando con la stessa corrente, con la stessa tensione, si riesce di fatto, progettando in maniera idonea tutto, a ottenere una compensazione degli errori introdotti dall'alimentazione, dunque una loro reiezione.

### 5.5.3 Offset

Si consideri un generico amplificatore operazionale, configurato in questo modo:

Se si applica una tensione differenziale nulla, l'uscita teoricamente dovrebbe essere a metà della dinamica, dunque a un certo valore nominale  $v_{0N}$ . Ci si aspetterebbe di aver a che fare con una caratteristica di questo tipo:

La verità è che le caratteristiche sono diverse da questa: volendo produrre un certo numero di amplificatori a partire dallo stesso progetto e sullo stesso chip, quelle che si potrebbero misurare sono tutte caratteristiche molto diverse tra loro. Ciò è negativo: un minimo mismatch potrebbe di fatto far saturare l'uscita, anche se la tensione differenziale è idealmente nulla.

Data la curva nominale, si può definire il concetto di offset di uscita, come:

$$v_{0,off} \triangleq |v_{01} - v_{0N}|$$

Ciò che di solito si fa, come d'altra parte si fa per il rumore, è riportare ciò all'ingresso, in modo da confrontare con il segnale di ingresso e verificare che il disturbo non sia simile al segnale di ingresso:

$$v_{off} = \frac{v_{0,off}}{A_d}$$

L'approccio che normalmente si utilizza per la modellizzazione è il seguente:

Si dice che un amplificatore reale è modellabile mediante un amplificatore non affetto da offset, eccitato da un generatore equivalente, in grado di introdurre una variazione sull'uscita uguale a quella che si ha in presenza di offset. Questo è un altro modo (più circuitale, più visivo) di definire l'offset: uno sbilanciamento dell'ingresso tale da avere l'uscita al valore nominale.

Dalla trattazione si potrebbe pensare che esiste un solo tipo di offset: l'idea potrebbe essere basata sul fatto che la presenza di mismatching dovuta alle tolleranze sul matching dei componenti provochi asimmetrie. La verità è che ciò è vero, ed è sicuramente importante, ma non è l'unico contributo: questo è un termine *casuale*, ma non è l'unico, dal momento che si ha anche un termine *sistematico*, che si può prevedere.

Generalmente, un sistema di amplificazione non è costituito da un singolo stadio; quello che si ha è un sistema costituito da diversi stadi di amplificazione, ciascuno dei quali introdurrà un contributo di offset.

La tensione di uscita  $v_{off}$  tiene conto di tutti gli offset, amplificati dai vari stadi di amplificazione. Ciascuna tensione di offset è costituita sia dal contributo casuale, sia dal contributo sistematico, contributo che dunque noi possiamo ridurre progettando in maniera opportuna ciascun blocco di amplificazione.

Si ha:

$$v_{0,off} = v_{0,off1}A_2A_3 + v_{0,off,2}A_3 + v_{0,off3}$$

Volendo riportare all'ingresso questa espressione, ossia trovare la tensione di offset in ingresso, si ottiene:

$$v_{off} = \frac{v_{0,off}}{A_1A_2A_3} \frac{v_{0,off1}}{A_1} + \frac{v_{0,off2}}{A_1A_2} + \frac{v_{0,off3}}{A_1A_2A_3}$$

Cosa ci dice questa espressione? Si può vedere che la tensione di offset relativa al primo stadio di amplificazione sarà diviso per il solo guadagno dell'amplificatore 1; gli altri, dai prodotti dei guadagni antecedenti. Il primo stadio, dunque, sarà, almeno sotto il punto di vista dell'offset, lo stadio più critico: progettare il primo stadio sarà assolutamente fondamentale sotto il punto di vista dell'offset.

Noi, da progettisti, siamo sostanzialmente in grado di controllare il solo offset sistematico: l'offset casuale è poco controllabile, dal momento che è dovuto alle tolleranze. Verrebbe spontaneo pensare che, una volta progettato in maniera molto accurata il primo stadio sotto il punto di vista degli offset sistematici, gli altri siano più semplici; ciò è tendenzialmente vero, ma d'altra parte fino a un certo punto: bisogna sempre e comunque fare in modo che gli offset introdotti dai suddetti stadi sia sufficientemente basso, dal momento che, se è troppo significativo, anche se diviso per molto guadagni darà comunque fastidio.

L'offset è una somma di errori sistematici e casuali. Considerando per esempio il seguente circuito:

Si può scrivere che:

$$\frac{W}{L}\Big|_1 = \frac{W}{L}\Big|_2$$

$$\frac{W}{L}\Big|_3 = \frac{W}{L}\Big|_4$$

Per questo circuito dunque si deve discutere di entrambe le fonti di offset.

### Offset sistematico

Ipotizzando di analizzare il solo modo comune, è possibile vedere che la struttura è perfettamente simmetrica se

$$v_{GS6} = v_{GS3}$$

Questo dunque implica  $i_{D1} = i_{D2}$ .

Ora, ricordando che

$$v_{GS6} = V_{THn} + v_{OD6}$$

si deve avere:

$$V_{TH6} + v_{OD6} = V_{TH3} + v_{OD3} \implies v_{OD6} = v_{OD3}$$

A questo punto, dunque, è possibile riscrivere secondo le formule precedentemente analizzate questa espressione in modo da trovare:

$$\sqrt{\frac{i_{D3}}{\beta_n \frac{W}{L}\Big|_3}} = \sqrt{\frac{i_{D6}}{\beta_n \frac{W}{L}\Big|_6}}$$

semplificando i vari termini comuni, quindi, è possibile trovare una relazione tra correnti e dimensioni dei transistori:

$$\frac{i_{D6}}{i_{D3}} = \frac{\frac{W}{L}|_6}{\frac{W}{L}|_3}$$

Dallo schema tuttavia si può osservare che  $i_{D7}$  e  $i_{D6}$  sono coincidenti, dal momento che i due transistori si trovano sullo stesso ramo (non consideriamo correnti sull'uscita); si può inoltre applicare lo stesso ragionamento su 7 e 5 dal momento che essi (dallo schema, sempre), condividono la stessa tensione di overdrive. Con gli stessi ragionamenti di prima dunque è possibile scrivere:

$$\frac{i_{D7}}{i_{D5}} = \frac{\frac{W}{L}|_7}{\frac{W}{L}|_5}$$

Osservando lo schema,  $i_{D5}$  è la corrente che va ad alimentare lo stadio differenziale; questo vuol dire che, in questo caso, essa si ripartisce in due contributi esatti nei transistori 1 e 2, che poi hanno come generatore lo specchio 3-4. Dunque:

$$i_{D3} = \frac{i_{D5}}{2}$$

e quindi

$$\frac{i_{D7}}{i_{D3}} = 2 \frac{\frac{W}{L}|_7}{\frac{W}{L}|_5}$$

Questa dunque può essere riportata al transistor 6, essendo esso legato al 7 (le due correnti di drain come già detto sono coincidenti). Quindi:

$$\frac{i_{D6}}{i_{D5}} = 2 \frac{\frac{W}{L}|_7}{\frac{W}{L}|_5} = 2 \frac{\frac{W}{L}|_6}{\frac{W}{L}|_5}$$

Questa ultima relazione è quella che, se rispettata, permette di eliminare il contributo sistematico all'offset.

### Contributo del mismatch dei transistori

Una volta analizzato con il modo comune il contributo sistematico dell'offset e visto come annullarlo, si faccia riferimento, per quanto riguarda il contributo del mismatch dei transistori, al seguente circuito:

Si vuole discutere l'offset casuale per questo circuito. Questo è causato da mismatch dei transistori della coppia differenziale, e dunque dal fatto che

$$\frac{W}{L}|_1 \neq \frac{W}{L}|_2$$

e

$$V_{TH1} \neq V_{TH2}$$

Si immagini a questo punto di applicare all'ingresso una tensione tale da avere una  $i_o$  del tipo

$$i_o = i_{D1} - i_{D2}$$

questa deve essere portata a 0 dalla tensione  $v_d$  introdotta all'ingresso. Dunque, si ha:

$$\begin{aligned} v_{off} &= V_{GS1} - V_{GS2} = V_{TH1} + v_{OD1} - (V_{TH2} + v_{OD2}) = \\ &= (V_{TH1} - V_{TH2}) + (v_{OD1} - v_{OD2}) \end{aligned}$$

Dunque, se si ha, per le componenti DC,

$$I_{D1} = I_{D2} = I_D$$

definendo  $\Delta V_{TH} \triangleq V_{TH1} - V_{TH2}$ , è possibile ottenere

$$v_{off} = \Delta V_{TH} + \left[ \sqrt{\frac{I_D}{\beta_n \frac{W}{L}|_1}} - \sqrt{\frac{I_D}{\beta_n \frac{W}{L}|_2}} \right]$$

A questo punto, al fine di riscrivere le espressioni in un modo diverso, si definiscono

$$\begin{aligned} \overline{\frac{W}{L}}_{1,2} &\triangleq \frac{\frac{W}{L}|_1 + \frac{W}{L}|_2}{2} \\ \Delta \frac{W}{L}_{1,2} &= \frac{W}{L}|_1 - \frac{W}{L}|_2 \end{aligned}$$

da qua, si ha

$$\begin{aligned} \frac{W}{L}|_1 &= \overline{\frac{W}{L}}_{1,2} + \frac{1}{2} \Delta \frac{W}{L}_{1,2} \\ \frac{W}{L}|_2 &= \overline{\frac{W}{L}}_{1,2} - \frac{1}{2} \Delta \frac{W}{L}_{1,2} \end{aligned}$$

Sostituendo nell'espressione di  $v_{off}$  si può ottenere

$$v_{off} = \Delta V_{TH} + \frac{I_D}{\beta_n \frac{W}{L}|_{1,2}} \left[ \frac{1}{\sqrt{1 + \frac{\Delta \frac{W}{L}|_{1,2}}{\frac{W}{L}|_{1,2}}}} - \frac{1}{\sqrt{1 - \frac{\Delta \frac{W}{L}|_{1,2}}{\frac{W}{L}|_{1,2}}}} \right]$$

Sviluppando questa espressione mediante serie di Taylor è possibile vedere che:

$$v_{off,1,2} = \Delta V_{TH} + \frac{\bar{v}_{OD,1,2}}{2} \left( \frac{\Delta \frac{W}{L}|_{1,2}}{\frac{W}{L}|_{1,2}} \right)$$

Per quanto riguarda gli offset casuali dunque si ha a che fare con due contributi di offset:

- il primo, legato al mismatch delle tensioni di soglia;
- il secondo, legato al mismatch dei rapporti di aspetto.

Si può vedere che

$$\Delta V_{TH1,2} \sim \pm 3 \frac{K_{VTH}}{\sqrt{WL}}$$

dove la  $K$  è una costante di processo; nel caso del processo 350 nm in uso come esempio, essa vale

$$K_{VTH} = 15 \text{ mV mm}$$

Questo significa che se i transistori vengono fatti corti e larghi, è possibile ottenere un offset inferiore al millivolt.

Il secondo termine invece è legato al solo mismatch geometrico, e questo si può ridurre facendo crescere il  $\frac{W}{L}$ . Si noti che anche i transistori di polarizzazione contribuiscono alla tensione di offset.

Se  $v_d = 0$  si ha un  $\Delta I_{D1,2} \neq 0$ ; questo si può scrivere anche come

$$\Delta I_{D1,2} = \bar{g}_{m1,2} v_{off1,2}$$

Questa viene anche specchiata, e la coppia di specchiaggio introduce un ulteriore offset, quindi un'ulteriore  $\Delta I_{D3,4}$ ; questo significa che, sulla corrente di uscita, si ha un offset risultante pari a

$$\Delta i_o = \Delta I_{1,2} + \Delta I_{3,4}$$

Si immagini a questo punto di mettere un  $V_{BIAS}$ , e solo delle tensioni di modo comune,  $v_{CM}$ .  $V_{BIAS}$ , a questo punto, mantiene fisso il punto di lavoro. Su  $V_B$  ciò che arriva è  $i_0 = i_{D4} - i_{D2}$ ; l'errore totale dunque dipende dalla somma delle due  $\Delta I_D$  appena scritte.

Al fine di vedere la cosa in un modo diverso, è possibile prendere questo errore e riportarlo all'ingresso, ottenendo la tensione di offset all'ingresso:

$$\begin{aligned} v_{off} &= v_{off1,2} + \frac{\Delta I_{D3,4}}{\bar{g}_{m1,2}} = \\ &= v_{off1,2} + \frac{v_{off3,4} \bar{g}_{m3,4}}{\bar{g}_{m1,2}} = \\ &= \Delta V_{TH1,2} + \frac{\bar{v}_{OD1,2}}{2} \left( \frac{\Delta \frac{W}{L} |_{1,2}}{\frac{W}{L} |_{1,2}} \right) + \frac{\bar{v}_{OD1,2}}{\bar{v}_{OD3,4}} \left[ \Delta V_{TH3,4} + \frac{\bar{v}_{OD3,4}}{2} \left( \frac{\Delta \frac{W}{L} |_{3,4}}{\frac{W}{L} |_{3,4}} \right) \right] \end{aligned}$$

Da questa formula è possibile osservare che si possono usare dei transistori 3,4 piccoli e 1,2 grossi, in maniera da giocare sulle  $v_{OD}$  e cercare in questo modo di ridurre la tensione di offset.

Ovviamente, oltre al problema dell'offset si dovrebbe anche tenere conto di quello delle derive, ossia del fatto che si possono avere variazioni nel tempo degli offset.

### 5.5.4 Principali soluzioni per eliminare l'offset - circuiti tempo-discreti

Un'idea per eliminare il contributo dell'offset è quello di utilizzare dei circuiti tempo-discreti.

#### Recupero offset all'uscita

Una prima osservazione: è buona cosa evitare l'uso di transistori grossi, dal momento che essi occupano molta area e comunque sono più lenti di quelli di dimensioni ridotte.

Un primo circuito potrebbe essere il seguente:

Questo è un circuito a tempo discreto, ed opera in due fasi che saranno ora descritte.

- $\phi_1$ : memorizzazione dell'offset. In questa fase, supponendo che il disturbo sia modellato mediante la tensione di offset esterna, si avrà una  $v_{o,off}$  composta da due contributi:

$$v_{o,off} = v_{C1} + v_{C2}$$

- $\phi_2$ : durante questa seconda fase, gli switch  $S_3, S_4, S_5, S_6$  sono spenti,  $S_1$  e  $S_2$  accesi. Questo significa che:

$$v_o = -v_{C2} + v'_o - v_{C1}$$

semplicemente facendo l'equazione alle maglie. Ma

$$v'_o = (v_{in} + v_{off})A_d$$

quindi

$$A_d v_{off} = v_{o,off} = v_{C1} + v_{C2}$$

dunque

$$v_o = A_d v_{in}$$

Questa soluzione funziona ma solamente se l'amplificatore ha un guadagno non troppo elevato (inferiore a 1000). Nella fase  $\phi_1$  infatti, quando si memorizza, se l'amplificatore ha amplificazione troppo elevata, l'offset legato ai mismatch della coppia differenziale potrebbe portare l'amplificatore fuori dalla dinamica di uscita.

Un altro difetto di questa soluzione è il fatto che si può solo usare in circuiti a capacità commutate, dal momento che viene facilmente occupato in un sistema come un integratore o simili. Tuttavia essa non si presta bene ad essere implementata in circuiti a tempo continui (inoltre il transconduttore è pure accoppiato capacitivamente al resto del circuito).

### Recupero dell'offset all'ingresso

Al fine di eliminare il problema principale della precedente topologia, è possibile ricorrere a una topologia diversa. L'idea è quella di fare il recupero dell'offset all'ingresso, invece che all'uscita, utilizzando il seguente schema:

Anche in questo caso si ha a che fare con due fasi. Durante la fase  $\phi_1$ , si ha  $S_3, S_4, S_5, S_6$  accesi, mentre  $S_1, S_2$  spenti. Il risultato è quello di "staccare la sorgente e reazionare l'amplificatore", in questo modo:

Il risultato in questo caso è, di nuovo,

$$v_{out} = A_d v_d$$

(questo è vero solo in questa fase).  
ma anche

$$-v_{out} = v_d v_{off}$$

dunque, eguagliando le due cose, è possibile dire che

$$\frac{v_{out}}{A_d} + v_{off} = -v_{out}$$

quindi

$$-v_{out} \left( 1 + \frac{1}{A_d} \right) = v_{off}$$

al contrario di prima, se  $A_d \gg 1$ ,  $v_{out} \sim -v_{off}$ . Infine, si ha

$$v_{off} = v_{C1} + v_{C2}$$

e in questo modo quindi la tensione di offset è stata immagazzinata nei due condensatori.

Per quanto riguarda la seconda fase,  $\phi_2$ , si ha il seguente comportamento: qui, si ha

$$v_{C2} + v_{IN} + v_{C1} - v_{off} = v_d$$

quindi, da prima

$$v_{C1} + v_{C2} = v_{off}$$

e quindi

$$v_d = v_{IN}$$

e questo significa che l'offset è stato eliminato dall'ingresso.

### **Circuito alternativo 1**

Senza un circuito di recupero dell'offset, un circuito come quello che segue avrebbe un offset dell'ordine delle decine/centinaia di mV.

Un altro esempio di circuito per il recupero dell'offset potrebbe essere il seguente:

In questo caso, si richiede che i circuiti  $M_1$  e  $M_2$  siano piccoli. L'idea che si usa è quella di utilizzare un circuito digitale, con un contatore binario la cui uscita va a un convertitore D/A, il quale va a pilotare le correnti.

In una prima fase si collegano gli ingressi ad una tensione di alimentazione, in modo da polarizzare i gate dei transistori. L'ingresso sarà differenziale. Quindi  $S_1$  e  $S_2$  sono accesi,  $S_3$  e  $S_4$  sono spenti. Quindi, si può considerare la parte digitale. Il convertitore digitale analogico va a pilotare le correnti sui rami, a ogni commutazione. Quando l'uscita commuta si aggiunge o toglie un  $\Delta I$  in modo tale da eliminare l'offset.

### Circuito alternativo 2

Un altro schema alternativo (che si può trovare sul Razavi) è il seguente:

In questo circuito si hanno due amplificatori di transconduttanza e uno di transresistenza, dove il secondo amplificatore di transconduttanza è un feedback che recupera l'offset, con  $C_1$  e  $C_2$ . Ancora una volta, in una fase si stacca l'ingresso, e si caricano i condensatori; l'offset al transconduttore 2 è l'offset d'uscita, quindi nella seconda fase si effettua l'eliminazione.. Anche in questo caso, il circuito funziona a fasi, e ci sono due fasi.

È possibile dimostrare che durante la fase  $\phi_1$ , in cui  $S_1, S_2$  sono accesi,  $S_3, S_4$  spenti.

$$v_{out} = v_{off1} \frac{G_{m1}}{G_{m2}} + v_{off2}$$

## 5.6 Risposta in frequenza - compensazione

Gli amplificatori che si progettano e utilizzano devono avere un guadagno molto elevato, ma solamente per frequenze basse.

Come si vede dal diagramma di Bode, si deve avere

$$v_o = A_d v_d$$

ma questo solamente per basse frequenze. Il primo polo,  $f_{p1}$ , deve infatti essere posizionato per frequenze basse, di solito non più di 100 Hz.

L'argomento è quello della compensazione; volendo mettere questo amplificatore  $A_d$  in un anello, si finisce per avere, come noto dalla teoria della retroazione,

$$A_f = \frac{1}{\beta} \frac{T}{1 + T}$$

Questa  $A_f$  è l'amplificazione del sistema retroazionato. Supponendo che  $\beta$  sia costante (il blocco di retroazione),

$$T = A_d(f)\beta$$

In generale,  $T$  ha un andamento in frequenza di questo genere:

Questo è l'andamento di  $A_f(f)$ .

Si ricordi che non si ha solo una variazione del modulo di  $T$ , ma anche una variazione di fase, e anche essa è significativa. Infatti, un segnale non viene solo attenuato, ma anche sfasato, dal momento che esso subisce un certo ritardo  $\tau$ . Al crescere della frequenza dunque non solo si ha attenuazione, ma anche questo ritardo.

Ciò che si deve fare nei sistemi retroazionati è vedere qual è la differenza tra ciò che c'è all'ingresso e all'uscita. La reazione deve essere negativa, e non è detto che il ritardo non porti ad avere addirittura una reazione positiva, se troppo elevato. La condizione di instabilità è

$$|T| = 1, \quad \angle T = 180^\circ$$

Quello che si può fare è utilizzare il criterio di Bode al fine di studiare la stabilità:

Se all'intersezione del diagramma dell'ampiezza con il livello 0 dB si ha  $\angle T < 180^\circ$ , il sistema è stabile. L'obiettivo è garantire la stabilità, richiedendo un certo PM (Phase Margin): esso è definito come:

$$\text{PM} = 180^\circ + \angle T(f_T)$$

Solitamente, un progetto è realizzato in modo da avere  $\text{PM} > 70^\circ$ . L'analisi di stabilità, solitamente, si fa per il caso peggiore, ossia  $\beta = 1$ ; questa è la condizione di "voltage follower", dal momento che in questo caso si riporta all'ingresso l'intera uscita.

Ciò che si può fare in pratica è identificare la frequenza di transizione dell'asse 0 dB,  $f_T$ , e quindi si effettua l'analisi della stabilità su  $T$ . Se  $\beta$  è costante in frequenza, si ha una traslazione di questa curva con  $\beta$  (e il caso critico è  $\beta = 1$ ).

Quando si esegue la compensazione in frequenza ci si può riferire al margine di fase o al margine di guadagno; il margine di fase è stato definito poco sopra, quello di guadagno, cercando la frequenza per la quale la rotazione di fase è di  $180^\circ$ , e si vede quanto vale il guadagno; se il guadagno è maggiore di 1, il sistema è instabile.

Al fine di valutare il guadagno di anello si deve aprire l'anello, eccitare l'anello con un generatore di test, e quindi verificare cosa torna indietro. Ovviamente, la polarizzazione del circuito deve essere mantenuta.

Dal momento in cui si apre l'anello, il taglio non deve cambiare le condizioni di carico e di polarizzazione del sistema; in altre parole, il fatto di lavorare ad anello aperto non deve cambiare le condizioni di lavoro del sistema; l'apertura dell'anello serve a noi a fare una misura, non a cambiare il punto di lavoro del sistema.

Alternative sono metodi basati sul tenere l'anello chiuso per la bassa frequenza, mostrando impedenze molto elevate (induttanze molto elevate) a certe frequenze (elevate), mantenendo però la DC intatta.

Considerando ora un sistema del tipo

Il valore del margine di fase modifica la risposta al gradino del sistema. Nel caso di un margine di fase molto basso, come  $5^\circ$ , il sistema è poco stabile, dunque al gradino il sistema risponde con delle sovraelongazioni molto elevate, la cui ampiezza decade esponenzialmente, e la velocità di decadimento dalla stabilità del sistema. Al contrario, con un margine di fase di  $60^\circ$ , si ha qualcosa di opposto, ossia una salita lenta ma senza sovraelongazioni.

Studiando dunque la risposta al gradino, è possibile capire quale sia il margine di fase.

Si noti una cosa molto importante: l'instabilità prescinde dall'ampiezza dell'ingresso; questo, per dire che non è necessario mettere un'eccitazione elevata per rendere instabile un sistema; basta del rumore (che è di fatto sempre presente). Se un circuito è instabile, è instabile sempre, indipendentemente dall'ingresso.

Anche in frequenza si può osservare la presenza di un margine di fase di una certa entità; sul diagramma di Bode, si vede infatti la presenza di un overshoot in prossimità della  $f_T$ . Tanto più il margine di fase è piccolo, tanto più questo overshoot è elevato. Il motivo per cui il margine deve essere maggiore di  $60^\circ$ , è il fatto che l'amplificazione, in altri casi, non è piatta.

Si consideri per esempio il seguente amplificatore:

In questo circuito si ha a che fare con due punti ad alta impedenza,  $x$  e  $y$ ; sicuramente, la funzione di trasferimento presenta due poli. Il diagramma di Bode dunque sarà qualcosa del tipo:

La funzione di trasferimento in frequenza presenta due a sinistra. Ci possono poi essere zeri a destra causati dalle capacità tra gate e drain, per esempio per quanto riguarda  $M_6$ .

Si hanno i due poli  $p_1$  e  $p_2$ , poi  $p_m$  e  $z_m$ , e lo zero a destra  $z$  a frequenze molto elevate, causato appunto dalle "capacità a ponte".

A seconda del guadagno, il diagramma di Bode può modificarsi, ma il senso è che un sistema di questo genere può tranquillamente essere instabile. Per questo motivo, è necessario introdurre per questo stadio una compensazione (a meno che non lo si voglia utilizzare come comparatore, quindi fuori da un anello analogico).

Il modo più semplice per rendere stabile questo sistema è quello di spostare il primo polo a frequenze più basse, in modo tale da ottenere qualcosa di questo genere:

Spostando infatti indietro in frequenza il primo polo, ciò che di fatto si fa è far guadagnare molto l'amplificatore in una banda più stretta, ma d'altro canto far sì che quando il diagramma di Bode taglia l'asse 0 dB, lo faccia con una pendenza di 20 dB/dec. Quindi,

$$f_{p1} = \frac{1}{2\pi r_{0x}(C_x + C_c)}$$

in questo modo, spostando quanto si deve, si fa in modo che si possano avere almeno 60° di margine di fase.

L'idea per fare ciò sarebbe introdurre una capacità  $C_C$  tra  $r_{0x}$  (nodo  $x$ ) e massa; il problema è che la capacità necessaria per fare ciò è superiore al nF, e quindi impossibile da integrare. Questa è l'idea di principio della compensazione a polo dominante, ma di fatto è impossibile da applicare.

L'idea corretta dunque è quella di applicare una compensazione di tipo Miller, basata su un'idea diversa. Si tratta di una tecnica che permette di modificare la posizione dei poli sfruttando l'effetto Miller, aggiungendo un condensatore di compensazione per esempio tra l'uscita del primo stadio e l'ingresso del secondo stadio. Si ha il "pole splitting": dal momento che si aggiunge ciò, il polo  $p_1$  si porta a frequenze più basse (come desiderato), ma  $p_2$  si sposta a frequenze più alte (cosa assolutamente desiderabile).  $p_m$  e  $z_m$  restano dove sono, ma lo "zero a destra" si sposta a frequenze più basse.

Questa ultima osservazione è fastidiosa, dal momento che se lo zero fosse a frequenze elevate, non ci darebbe problemi; esso, infatti, non cambierebbe i margini di fase. Essendo esso uno zero a destra ed essendo spostato dal  $C_C$  di Miller a frequenze più basse, esso potrebbe trovarsi nella regione utile e quindi influenzare negativamente il margine di fase; per questo motivo, è necessario lavorare su ciò con maggiore attenzione. Si consideri il seguente modello circuitale:

Si ha:

$$A_{d0} = g_{m1}r_{0x}g_{m6}r_{0y}$$

quindi, si ha che

$$\omega_{p1} = \frac{1}{g_{m6}r_{0x}r_{0y}C_C}$$

$$\omega_{p2} = \frac{g_{m6}}{C_y}$$

$$\omega_z = \frac{g_{m6}}{C_C}$$

Lo zero a destra si ha quando tutta la corrente su  $C_C$  va a finire su  $g_{m6}v_{gs6}$ ; quindi, alla  $s_z$ , si ha:

$$s_z \implies i_C = g_{m6}v_{gs6}$$

quindi

$$i_C = s_z C_C v_C = s_z C_C v_{gs6}$$

quindi

$$s_z = + \frac{g_{m6}}{C_C}$$

Un'idea per spostare a frequenze più elevate questo zero può essere quella di aggiungere una resistenza  $R_C$  in serie a  $C_C$ . In questo caso, infatti,

$$\frac{v_{gs6}}{R_C + \frac{1}{s_z C_C}} = g_{m6}v_{gs6}$$

quindi, in questo caso,

$$\frac{s_z C_C}{1 + s_z R_C C_C} = g_{m6}$$

da cui

$$s_z(1 - g_{m6}R_C)C_C = g_{m6}$$

$$s_z = \frac{g_{m6}}{(1 - g_{m6}R_C)C_C}$$

Imponendo  $g_{m6}R_C = 1$ , il risultato è lo spostamento dello zero a frequenza idealmente infinita. Un'alternativa invece è quella di fare  $g_{m6}R_C \gg 1$ ; in questo caso, il termine negativo al denominatore prevale sul termine positivo, dunque lo zero si può addirittura spostare nel semipiano di sinistra, in modo che il suo contributo alla fase diventi positivo. Addirittura, quello che si potrebbe provare a fare è spostare questo zero fino a farlo coincidere con il secondo polo, in modo da ottenere un amplificatore operativo a banda molto più larga.

Il secondo polo è quello che limita a banda dell'amplificatore operativo retroazionato. A questo punto, per avere un margine di fase di  $60^\circ$ , si può dimostrare che

$$\frac{\omega'_{p2}}{\omega_T} = 2, 2$$

Tanto più  $\omega'_{p2}$  è spostato a frequenze elevate, tanto meglio è. Se poi addirittura si riuscisse a fare sovrapporre zero e polo, sarebbe il massimo. Ad ogni modo, semplificando, se  $g_{m6}R_C \gg 1$ ,

$$s_z \rightarrow -\frac{1}{R_C C_C}$$

Si noti che  $g_m$  e  $R_C$  non sono nè matchati nè matchabili, dal momento che sono diversi e non direttamente correlati. Questo significa che avere esattamente la sovrapposizione sarebbe molto difficile, a causa delle tolleranze.

Proseguendo, come si può realizzare in pratica ciò? L'idea è sostanzialmente questa:

Al fine di realizzare la resistenza, si può utilizzare il transistor  $M_8$ . Esso deve essere in regione triodo ma, dal momento che non si ha corrente continua tra drain e source (a causa della capacità), di sicuro la condizione sarà soddisfatta.

A questo punto:

$$V_{GS6} + V_{GS8} = V_{GS9} + V_{GS10}$$

quindi, se  $V_{GS6} = V_{GS9}$ , si ha di sicuro  $V_{GS8} = V_{GS10}$ ; dunque

$$v_{OD6} = v_{OD9}$$

dalla relazione dalle tensioni di overdrive è possibile passare a quella su correnti continue e rapporti

$$\sqrt{\frac{I_{B1}}{\beta_n \frac{W}{L}|_6}} = \sqrt{\frac{I_{B2}}{\beta_n \frac{W}{L}|_9}}$$

questo implica

$$\frac{I_{B1}}{I_{B2}} = \frac{\frac{W}{L}|_6}{\frac{W}{L}|_9}$$

questa è la condizione che garantisce che i due overdrive siano uguali e dunque che  $v_{OD6} = v_{OD9}$ .

Dunque, la stessa cosa si deve fare per gli altri due transistori.

Si ricorda che l'obiettivo finale è ottenere

$$g_{m6}r_{DS6} = 1$$

quindi

$$\mu_n C_{OX} \frac{W}{L} \Big|_6 v_{OD6} \frac{1}{\mu_n C_{OX} \frac{W}{L} \Big|_8} = 1$$

quindi

$$\frac{v_{OD8}}{v_{OD6}} = \frac{\frac{W}{L} \Big|_6}{\frac{W}{L} \Big|_8}$$

imponendo tuttavia come fatto prima che

$$v_{OD6} = v_{OD9}$$

si ha

$$\frac{v_{OD8}}{v_{OD6}} = \frac{v_{OD10}}{v_{OD9}} = \frac{\sqrt{\frac{I_{B1}}{\beta_n \frac{W}{L} \Big|_{10}}}}{\sqrt{\frac{I_{B1}}{\beta_n \frac{W}{L} \Big|_9}}}$$

Quindi, al fine di riprendere questa condizione e trasformarla in una condizione sui soli rapporti di aspetto, al fine di spostare lo zero in una regione stabile

$$\frac{\frac{W}{L} \Big|_6}{\frac{W}{L} \Big|_8} \sqrt{\frac{\frac{W}{L} \Big|_{10}}{\frac{W}{L} \Big|_9}} = 1$$

Imponendo questa condizione, lo zero è in regione stabile.