Elettronica Analogica II

Alberto Tibaldi

28 agosto 2011

Indice

Capitolo 1

Stabilità degli amplificatori operazionali

Quando si studia o progetta qualcosa che riguardi gli amplificatori operazionali (e non solo), bisogna avere la garanzia che il circuito sia stabile, ossia che, **quando l'ingresso è nullo**, l'uscita converga a un valore finito. Se l'uscita *non sta ferma*, il circuito sarà dunque instabile.

La prima cosa da fare, dunque, sarà **annullare gli ingressi**.

Dalla teoria dei controlli, si sa che un sistema è stabile se i suoi poli stanno sul semipiano sinistro. Per studiare ciò, dunque, si introduce l'idea di **amplificazione di anello**: un sistema con reazione è infatti un sistema ad anello chiuso.

Si ha un segnale di errore ε che pilota l'uscita attraverso l'amplificazione; un esempio di ciò è per esempio il servosterzo: la controreazione consiste nel misurare la posizione delle ruote mediante un trasduttore, riportando indietro l'uscita, calcolandone l'errore ε, e utilizzandolo per pilotare le ruote in modo da annullarlo.

Elettronicamente, dato un generatore v_e di ingresso, si preleva la sua v_u, la si porta indietro e la si sottrae, e ciò fa sì da avere $v_u = v_e$, a meno di un certo errore ε. Si noti che ε è più piccolo tanto più grande è l'amplificazione del dispositivo.

A questo punto, da elettronici, potremmo chiederci: non bastava un filo per fare ciò? Beh, in realtà no: v_e è infatti non ideale, dunque si avrebbero dei partitori che rovinano tutto. Ciò che ci serve, di fatto, è qualcosa che *faccia la fatica* di portare la tensione giusta al carico: questa *fatica*, di fatto, la fa l'operazionale.

Se si vuole amplificare 10, si deve fare in modo che $\frac{1}{10}$ della tensione di uscita torni indietro, ossia che essa sia imposta uguale alla v_e.

Ci si pone a questo punto un problema fondamentale, già citato: quello della **stabilità**. Perchè ha senso chiedersi se un sistema è stabile? Beh, semplicemente, perchè se il sistema non fosse stabile, allora vorrebbe dire che una oscillazione, che *gira* per la retroazione, continuerebbe ad autosostenersi, diventando alla fine più importante dell'ingresso stesso, facendo a un certo punto divergere l'uscita (fino al punto di saturazione).

Per studiare la stabilità si analizza il guadagno d'anello: si taglia l'anello in un punto (a patto di prestare attenzione su quale punto si utilizza, nella fattispecie si eviti di tagliare per esempio dei partitori), dunque si introduce un segnale e si vede quanto ne torna indietro. Fare ciò con l'operazionale è abbastanza semplice, dal momento che distinguere i blocchi A e β non è molto complicato. Si apre dunque l'anello, si introduce un v_p, e si vede quanto segnale torna indietro all'altro capo del filo rotto: esso sarà v_r.

Dal momento che stiamo studiando una controreazione, ossia una reazione negativa, a frequenze sufficientemente basse avrò un segnale in opposizione di fase rispetto a quello introdotto; si definisce dunque il **guadagno di anello** come:

$$T \triangleq -\frac{v_r}{v_p}$$

Cosa sarebbe questa grandezza? Beh, tolto il segnale di ingresso del sistema, si apre l'anello in un punto, ottenendo due capi; a un capo si introduce un segnale v_p, all'altro si vedrà *tornare indietro* un altro segnale, v_r. Il rapporto tra i due indica quanto del segnale introdotto nell'anello viene mandato indietro: questa è la *reazione*. A seconda di quanto segnale torna indietro, il guadagno di anello sarà più o meno alto. Più alto è il guadagno di anello, più è il segnale che viene riportato indietro: il guadagno di anello è una funzione che quantifica quanto di un segnale nell'anello torna indietro (a scapito dell'uscita per esempio).

Cosa capita dunque? Beh, al crescere della frequenza di v_p vengono a intervenire gli elementi reattivi del circuito, che siano in A o che siano in β, dunque la fase inizia a ruotare, e il modulo a calare. A un certo punto la fase potrebbe esser talmente ruotata da essere cambiata di 180°, i quali, aggiunti ai 180° di partenza (dovuti al fatto che la reazione cambia di segno il segnale di ritorno da quello di partenza, essendo negativa), fanno arrivare una rotazione di 360°: il segnale v_r sarà in fase a v_p. Ciò è veramente pessimo: in questo caso, infatti, la reazione diventa positiva. Esistono sostanzialmente tre possibilità:

- $|T| = 1$: in questo caso ingegneristicamente non si rientra mai, dal momento che il modulo dovrebbe valere esattamente 1. Se questo caso fosse possibile, e chiudessi l'anello, avrei una tensione di uscita che rimane quella che è: si ha un oscillatore che non si autoinnesca, che non ha bisogno di innesco.

- $|T| < 1$: in questo caso la v_r che torna è più piccola di v_p, dunque quando chiudo l'anello e tolgo v_p, a ogni ciclo v_r si riduce di ampiezza, fino ad annichilirsi. Questa è una condizione di stabilità, per il circuito.

- $|T| > 1$: se per ogni ciclo $v_r > v_p$, capita che, al contrario di prima, il sistema tende ad avere l'uscita che diverge, anche senza la v_p: ad anello chiuso basta il rumore nel circuito per autoinnescare una reazione che prima o poi renderà il sistema instabile.

Questo studio, questo criterio, basato sull'analisi del guadagno di anello, è chiamato **criterio di Bode**.

Si noti che, se si vuole progettare un amplificatore, ed esso è instabile per una qualche frequenza anche non interessante nel range di amplificazione, esso diventa comunque inutilizzabile: se l'uscita satura per una qualsiasi frequenza, l'amplificatore va fuori linearità e non è più un amplificatore. Vi è un altro modo per *vedere* questo criterio, per utilizzarlo: valutato T e la frequenza in cui $|T|$ vale 1 (o 0 dB per gli amici), la fase è o meno pari a 180^circ ? Beh, se siamo sotto i 180 gradi siamo a posto, e l'amplificatore è stabile, altrimenti no.

In realtà, la stabilità non ci basta: vorremmo anche un certo margine, in modo che la stabilità non sia determinata al pelo: le tolleranze di fabbricazione o le derive infatti potrebbero rendere comunque il sistema inutilizzabile: usare anche solo un banale gradino di ingresso potrebbe essere infatti problematico, dal momento che il gradino contiene tutte le frequenze, e dunque l'amplificatore potrebbe avere una risposta *isterica* anche al solo gradino, a una delle frequenze di eccitazione.

Applichiamo questo discorso a un amplificatore operazionale controreazionato:

Se si taglia l'anello (in un punto tale da non distruggere partitori), si può vedere che:

$$v_r = -A_d\beta$$

Per β, si può fare un ragionamento particolare: siccome non conosco R_i e R_u, per *mettermi al sicuro*, posso ipotizzare che l'amplificatore sia **nel peggior caso possibile**, e stimare un $|T|$ più grande del dovuto. In questa ipotesi, posso porre $R_u = 0$ (in modo che il partitore vada a favore del β) e $R_i \to \infty$ (in modo che si elimini un rapporto di partizione con l'ingresso);

siamo fortunati, perchè questo è sostanzialmente il caso dell'amplificatore operazionale ideale, molto semplice da studiare. Si può dire, dato ciò, che:

$$\beta \sim \frac{R_1}{R_1 + R_2}$$

Si noti che poi le R_i possono essere Z_i, e dunque β potrebbe di suo introdurre delle rotazioni di fase, ma per ora ignoriamo cose del genere.

Come si comporta A_d in un amplificatore operazionale? Beh, esso avrà un andamento di questo genere:

Si hanno sostanzialmente 3 poli: f_{p1}, f_{p2}, f_{p3}. Il primo polo è dovuto al primo stadio ed è a bassa frequenza, il secondo al secondo stadio. Si può ragionevolmente supporre che i poli siano abbastanza lontani tra loro. β è minore o uguale a 1, dunque sarà minore o uguale a 0 dB; T si calcola come il prodotto $A\beta$, dunque sarà semplicemente A_d, traslato *in basso* di β (che, se non si hanno reazioni reattive, è costante). Ciascun polo introdurrà un comportamento di questo genere:

$$\frac{1}{1 + j\frac{\omega}{\omega_p}}$$

Ossia, se $\omega = \omega_p$, ho $\frac{1}{1+j}$, la cui fase è $-45°$. Se poi $\omega \gg \omega_p$, avrò circa $\frac{1}{j\frac{\omega}{\omega_p}}$, che porterà una rotazione di fase in ritardo di 90°.

AL fine di migliorare le cose, dal momento che, oltre il secondo polo, abbiamo $|T| > 1$, e dunque il sistema è instabile, posso o modificare β, per esempio in modo da abbassare il guadagno, o abbassare l'A_d dell'operazionale.

Noi, opereremo sull'operazionale: modificando lo schema interno si può fare in modo che, con un certo margine, si abbia $|T| > 1$ fino a quando la fase ha qualcosa meno di 180° di rotazione; si può dire che il margine ideale sia di circa 45°. Quando T taglia dunque l'asse 0 dB, si deve aver avuto 135° di rotazione di fase, dunque il secondo polo deve essere posizionato più o meno esattamente sull'asse 0 dB.

Come si può fare ciò? Beh, un'idea potrebbe essere: se prendo il primo polo e lo porto indietro, ottengo che vi sarà più attenuazione, dunque scegliendo la posizione corretta potrei riuscire a mettere il secondo polo esattamente sull'asse 0 dB. Ragionando: se il secondo polo è sull'asse 0 dB, ovviamente prima di esso agirà solo il primo polo. Per raggiungere T_0, dunque, il primo polo dovrà essere molto più indietro di quanto non fosse prima. Facendo così abbiamo introdotto un polo a 2 o 3 Hz, e dunque, se da un lato abbiamo migliorato la stabilità del sistema, dall'altro abbiamo perso amplificazione.

Ora, come si fa ciò in pratica? Beh, in sostanza, abbiamo detto che il primo polo è dovuto al primo amplificatore, ossia allo stadio differenziale. Volendo dunque ridurre la posizione del polo, dobbiamo introdurre una capacità C tale da mettere in parallelo a C_0 un polo; questa capacità C dovrebbe però essere molto grossa, per abbassare di molto la posizione del polo, dunque essa sarà impossibile da integrare. Si ha però una possibilità: invece di mettere una capacità in parallelo a C_0, la si può mettere a cavallo dei due stadi, in modo da sfruttare a nostro favore l'effetto Miller, e l'amplificazione della capacità.

Il parallelo R_1C_1 è il responsabile del primo polo *naturale* dell'operazionale, mentre R_2C_2 del secondo; gli altri poli derivano da vari parametri dello schema, quali C_B. Aggiungere C non aggiunge di fatto poli, dal momento che si trova in una maglia di condensatori, dunque degenere, ma fa altre cose:

$$\omega_{p1} \sim -\frac{1}{g_{m2}R1R_2C}$$

Ciò è abbastanza ovvio: $g_{m2}R_2$ è infatti l'amplificazione di tensione dovuta al secondo stadio, dunque questo è l'effetto Miller, che moltiplica C, e la fa ritornare in ingresso più grossa.

Per il secondo polo:

$$\omega_{p2} \sim \frac{-g_{m2}C}{C(C_1 + C_2) + C_1C_2} \sim \frac{-g_{m2}}{C_1 + C_2}$$

Ciò è invece molto inaspettato, ma altrettanto bello: se valutassi questo rapporto, vedrei che il secondo polo è stato *buttato* a frequenze molto elevate, nell'ordine dei GHz! Questo fenomeno è detto **pole splitting**: il primo polo va giù di frequenza, il secondo va molto su. Ciò che questa C inoltre fa è aggiungere uno zero: questa capacità infatti fa sì che si introduce una certa corrente direttamente dal primo al secondo stadio; se esiste una frequenza per cui la corrente nella capacità è uguale a $g_{m2}v_{u1}$, dunque, si ha che tutta la corrente nella capacità è quella del pilotato, dunque sull'uscita non va altro, e $v_u = 0$ (ottenendo uno zero di trasmissione). Si ha che:

$$v_{u1}sC = g_{m2}v_{u1}$$

Dunque

$$s = \frac{g_{m2}}{C}$$

Questo zero si trova al semipiano di destra, dunque la rete è a fase non minima.

A questo punto, due possibilità: se abbiamo operazionali a bipolari, essendo la transconduttanza alta questo zero va a frequenze alte; altrimenti, con i MOSFET non è così, e si deve introdurre qualche artificio che spiegheremo in seguito.

Come si fa, in pratica, a **compensare l'amplificatore**? Beh, un'osservazione: a causa del pole splitting, il secondo polo è ai GHz, dunque quello che era il *terzo polo naturale* diventa il *secondo polo* dopo la compensazione. Partendo dunque dalla pulsazione di quest'ultimo, risalendo con una pendenza di - 20 dB/dec (andando verso sinistra), si continua fino a incrociare T_0 (ossia T a $j\omega = 0$). ω_{p3} è data dal datasheet mediante il parametro f_T, dal momento che per chi produce l'integrato è facile tenerla d'occhio). Si ha che:

$$T_0 = A_0\beta = g_{m1}R_1g_{m2}R_2\beta$$

Si ha che:

$$\omega_{p1} = \frac{\omega_T}{|T_0|}$$

Dal momento che 20 dB/dec significa **proporzionalità inversa**, dunque si può dire che:

$$f_aG_a = f_bG_b$$

Dove f sono frequenze, G guadagni. Da prima, dunque:

$$\omega_{p1} = \frac{\omega_T}{g_{m1}R_1g_{m2}R_2\beta} = \frac{1}{Cg_{m2}R_1R_2}$$

Dunque:

$$C = \frac{g_{m1}\beta}{\omega_T}$$

La capacità da mettere a cavallo del primo e del secondo stadio va dimensionata mediante questa formula, al fine di ottenere la stabilità, e questa dipende da β. Si noti, tuttavia, che essa **non dipende dall'amplificazione dell'amplificatore operazionale**.

La situazione più **estrema** per β è il caso del **voltage follower**, ossia $\beta = 1$: in questo caso C deve essere la più grande possibile, dal momento che di per sè β è il più alto che si potrebbe avere, ossia dal momento che la reazione è la più grande che si possa avere, dunque il caso più critico. Per vederlo in un altro modo, in questo caso $A_d = T$, dunque si ha sostanzialmente che non si ha un β che *di suo* abbassi un po' il diagramma di Bode di T

rispetto a quello di A_d, ottenendo il caso pessimo. Per $\beta = 1$, si ottiene il cosiddetto *full-compensated op-amp*: amplificatori operazionali compensati per ogni tipo di reazione.

Ora, ricordiamo che:

$$\omega_z \sim \frac{g_{m2}}{C} \qquad \omega_T \sim \frac{g_{m1}}{C}$$

Se $\omega_z \gg \omega_T$, siamo contenti, dal momento che, essendo uno zero sul semipiano destro, esso introduce una rotazione di fase identica a quella di un polo sul semipiano sinistro; se però la pulsazione dello zero è molto maggiore a quella del secondo polo, ricordando che la pulsazione del secondo polo è la massima pulsazione per cui si ha un guadagno di anello superiore o uguale a 1, si avrà sì una rotazione di fase, ma per frequenze per le quali il guadagno di anello è molto inferiore a 1, dunque possiamo stare tranquilli. Se però ω_z è nella banda utile, essa introduce un grosso ritardo di fase, riducendo il margine di stabilità. Si considerino ora nel dettaglio due casi:

- Nel caso dei sistemi a BJT, si ha che $g_{m2} \gg g_{m1}$: la transconduttanza di un transistore bipolare è infatti linearmente dipendente (cresce) con la corrente che vi è dentro; il secondo stadio avrà una corrente molto più grossa del primo (I_0 il primo, KI_0 il secondo), dunque possiamo stare tranquilli.

- Nel caso dei sistemi MOS, non possiamo stare così tranquilli, dal momento che noi vorremmo guadagno molto elevato; si può vedere che, aumentando I_0, riduciamo r_0, e facciamo aumentare di poco g_m, dunque non si può fare in modo che $g_{m2} > g_{m1}$: ridurremmo troppo il guadagno. Nel caso dei MOSFET dunque si introducono artifici circuitali: al posto della sola capacità di compensazione si introduce in serie ad essa anche una piccola resistenza R: capita che a sinistra si ha v_{u1}, a destra $v_u = 0$ (per ipotesi, dal momento che stiamo calcolando la frequenza dello zero di trasmissione), dunque:

  $$\frac{v_{u1} - 0}{R + \frac{1}{sC}} = g_{m2} v_{u1} \Longrightarrow \frac{sC}{1 + sRC} = g_{m2}$$

  Dunque:

  $$sC = g_{m2}(1 + sRC)$$

  Dunque

$$s_z = \frac{g_{m2}}{C(1 - R g_{m2})}$$

Se $R = \frac{1}{g_{m2}}$, la pulsazione dello zero va a frequenze idealmente infinite, realmente elevatissime. R poi viene riportata all'ingresso divisa per l'amplificazione (per effetto Miller), non influenzando ciò che capita.

1.1 Banda passante di un amplificatore

Spesso nei libri si parla di banda passante di un amplificatore, indicando quella banda per cui l'amplificazione non si è ancora ridotta di 3 dB. 3 dB possono essere pochi, tanti: dipende sostanzialmente da cosa si deve fare, dunque come indicatore non è il massimo.

Per poter introdurre il discorso che si intende fare, si introduce una certa simbologia:

- A_∞ è il guadagno (o, più generalmente, la funzione di trasferimento) che noi idealmente vorremmo ottenere per il sistema intero (nei nostri casi, per l'amplificatore operazionale reazionato), supponendo che l'operazionale e tutti i componenti siano ideali. Si tratta di *ciò che vorremmo ottenere*.
- A_d è l'amplificazione dell'amplificatore operazionale non retroazionato.
- β è il blocco di retroazione.
- T è il guadagno di anello.
- A_f è *ciò che si ottiene in realtà*: dato un amplificatore operazionale retroazionato, al fine di ottenere un determinato sistema, questo è ciò che si ottiene in realtà. Esso vale:

$$A_f = A_\infty \frac{T}{1+T} + A_0 \frac{1}{1+T} \sim A_\infty \frac{T}{1+T}$$

A_0 è l'amplificazione che avrei tra ingresso e uscita se spegnessi l'operazionale, ossia parametri parassiti e resistenze di ingresso e uscita; questi si trascurano quasi sempre.

In frequenza capita che, se siamo nelle frequenze per cui $|T| \gg 1$:

$$|A_f| \sim |A_\infty|$$

Ossia, si può dire che *siamo in banda passante* se $|T| \gg 1$.

Se $|T| \ll 1$, ho che:

$$|A_f| \sim |A_\infty| \, |T|$$

In questo caso, siamo fuori dalla banda passante.

L'ultimo caso, è quello per cui $|T| \sim 1$: in generale si può dire che:

$$T = |T| \, \mathrm{e}^{j\angle T}$$

Definisco $\varphi = \angle T$. Si sa inoltre che:

$$A_f = A_\infty \frac{1}{1 + \frac{1}{T}}$$

Applicando il teorema di Carnot (noto anche come *teorema del coseno*), vedo che:

$$|A_f| = \frac{|A_\infty|}{\sqrt{1 + \frac{1}{|T|^2} + \frac{2}{|T|}\cos(\varphi)}}$$

Se $\psi \sim 45°$, ottengo un sovraguadagno di circa 3 dB; si dice che questo valore per ψ sia ottimo, dal momento che da un lato non introduce troppa inerzia nel sistema (non è troppo *lento*), ma dall'altro la risposta del sistema non è troppo isterica (nel senso che se il margine è troppo basso, come già detto si hanno diverse sovraelongazioni che rallentano la convergenza dell'uscita).

1.2 Slew rate

Parlare di banda passante ha senso solo quando si fa l'ipotesi di linearità, ossia nell'ipotesi in cui il segnale sia così piccolo da mantenere il circuito in linearità. A limitare le prestazioni di un sistema non vi sono solo le questioni concernenti la banda passante, dal momento che sono coinvolti anche fenomeni di non linearità: lo slew rate è, di fatto, uno di questi fenomeni.

Per **slew rate** si intende la massima velocità di variabilità possibile sull'**uscita**:

$$S_R \triangleq \left.\frac{\mathrm{d}v_u}{\mathrm{d}t}\right|_{max}$$

Questa è, in altre parole, la massima pendenza, nel tempo, che si può avere sull'uscita.

Si noti che questo fenomeno **non va confuso** con il limite di banda passante: data per esempio una funzione di trasferimento con un polo dominante, se metto in ingresso un generatore di gradino ideale (a pendenza dunque infinita), a causa del polo nel tempo il circuito risponde con un esponenziale, del tipo:

$$v_u = V_{UM}\left[1 - \mathrm{e}^{-\frac{t}{\tau}}\right]$$

Dove $\tau = RC$.

Da qua:

$$\frac{\mathrm{d}v_u}{\mathrm{d}t} = \frac{V_{UM}}{\tau}\mathrm{e}^{-\frac{t}{\tau}}$$

Ciò ci dice che si ha una $\frac{\mathrm{d}v_u}{\mathrm{d}t}$ limitata, dove il massimo è per $t = 0$; dalla prima formula, poi, si conosce l'andamento del segnale.

Ciò che tutte queste formule non mostrano è che (a meno di effetti di saturazione che comunque non ci riguardano), si hanno vincoli sull'ampiezza del segnale di ingresso: idealmente, per ciò che sappiamo dalla banda passante, al crescere dell'ingresso cresce anche l'uscita, e l'andamento è sempre lo stesso: ciò, in altre parole, ci dice che $\frac{\mathrm{d}v_u}{\mathrm{d}t}$ può anche valere ∞, dal momento che la banda passante limita il contenuto armonico in assoluto, non in dipendenza dell'ampiezza del segnale.

Si consideri a questo punto un segnale sinusoidale:

$$v_u(t) = V_{UM}\sin(\omega t)$$

Il limite di slew rate ci dice che *più di una certa pendenza non posso avere*: a parità di frequenza, dunque, la sola ampiezza diventa determinante.

Facciamo due conti: perchè non vi sia distorsione da slew rate:

$$\frac{\mathrm{d}v_u}{\mathrm{d}t} < S_R$$

Ma

$$\frac{\mathrm{d}v_u}{\mathrm{d}t} = \omega V_{UM}\cos(\omega t) < S_R$$

Dunque, considerando che il massimo del coseno è 1:

$$\omega V_{UM} < S_R$$

Dunque

$$V_{UM} \leq \frac{S_R}{\omega}$$

Ossia, in questo caso, si ha un andamento di questo tipo:

Questo andamento esponenziale ci suggerisce qual è il massimo valore che può essere introdotto senza distorsione, a una certa frequenza.

Da cosa dipende, **in pratica**, lo slew rate? Beh, ricordando lo schema interno dell'amplificatore operazionale, si può dire che sulla capacità C ci sarà:

$$v_C = v_u - v_{u1} \sim v_u$$

Dal momento che $v_u \gg v_{u1}$.

Volendo una certa $\frac{\mathrm{d}v_u}{\mathrm{d}t}$, dunque, dovrò avere una $\frac{\mathrm{d}v_C}{\mathrm{d}t}$ molto rapida: **è la capacità colei che limita lo slew rate**: per cambiare tensione ai propri capi, essa deve far scorrere una certa corrente, dunque:

$$\frac{\mathrm{d}v_C}{\mathrm{d}t} = \frac{I_C}{C}$$

Per migliorare lo slew rate devo mandare tanta corrente nella capacità, in modo da variare rapidamente la carica, dunque la tensione. Il problema è che la massima corrente che si possa gestire è quella per cui si ha un gradino di tensione differenziale in ingresso, tale da interdire Q_2, e facendo condurre tutto nello specchio, ottenendo, su C, di fatto, I_0:

$$S_R \leq \frac{I_0}{C}$$

Ma sappiamo che:

$$C = \frac{g_{m1}}{\omega_T}$$

Dunque:

$$S_R = \frac{I_0 \omega_T}{g_{m1}}$$

ω_T dipende dalla tecnologia, e $\frac{I_0}{g_{m1}}$ è un rapporto di grandezze tra loro legate. Nel BJT esse dipendono da V_T, cosa intoccabile, mentre nel MOS si han meno problemi, dal momento che g_m è piccola, e si può giocare su altri parametri.

Per il BJT sono stati studiati sistemi per migliorare questo rapporto, basati sull'introduzione di reazioni all'interno del circuito; ciò si fa, dunque, degenerando il primo stadio, controreazionandolo, cosa poco piacevole, dal momento che migliora sì lo slew rate, ma riducendo notevolmente il guadagno.

Capitolo 2

Il rumore nei circuiti integrati

2.1 Rumore nel dominio della frequenza

Anzichè studiare il rumore nel dominio del tempo, è possibile farne una caratterizzazione nel dominio della frequenza: l'andamento sarà presumibilmente continuo, dunque si avrà una qualche distribuzione della potenza per unità di frequenza. In altre parole, per ogni frequenza si avrà una certa potenza, che sarà espressa da noi in termini di $\frac{\mathrm{V}^2}{\mathrm{Hz}}$ o come radice (si parla di radice di potenza).

Si noti che si ha un forte collegamento con i valori efficaci:

$$V_{n(rms)} = \int_0^\infty V_n^2(f)df$$

la radice di questa espressione sarà dunque il valore efficace: V_{rms} può dunque essere pensato come la somma, su tutti i possibili contributi di frequenza, della densità spettrale di potenza.

Presentiamo a questo punto le distribuzioni di rumore in frequenza più comuni:

- Il rumore bianco (costante in frequenza): tipico per esempio del rumore termico, ossia del rumore dovuto all'agitazione degli elettroni.

- Rumore flicker (o rumore $\frac{1}{f}$, o *rumore rosa*):

$$V_n^2(f) = \frac{K^2}{f}$$

  Questo è comune nei dispositivi attivi allo stato solido, a causa di **trappole** presenti nei dispositivi che trattengono i portatori, rilasciandoli non uniformemente rispetto alla frequenza del segnale. Ciò conta molto per frequenze basse, ma tende a ridursi al crescere della frequenza. Si noti che la potenza decade con la frequenza, non con il quadrato della frequenza, dunque la pendenza di questo rumore, rappresentato su un diagramma di Bode, sarà:

$$10 \log_{10} \frac{1}{f} = -10 \log_{10}(f)$$

ossia, - 10 dB/dec.

2.1.1 Rumore in ingresso a un due-porte

Cosa capita se ho del rumore in ingresso a un due porte?

Data $A(f)$ la funzione di trasferimento tra ingresso e uscita del 2-porte, e il $V_{ni}^2(f)$ in ingresso, come si calcola il $V_{nu}^2(f)$?

Si noti che quello in ingresso non è un vero generatore di tensione, bensì di potenza (per quanto i simboli utilizzati spesso coincidano): il simbolo sembra fuoriviante, anche se può tornare utile per ricondursi alla fisica.

Si può dire che:

$$V_{nu}^2(f) = V_{ni}^2(f) \, |A(f)|^2$$

per le radici di potenza:

$$V_{nu}(f) = V_{ni}(f) \, |A(f)|$$

Questo, perchè si ha a che fare con i valori efficaci, che son dunque trattati mediante il modulo.

Esiste un metodo molto approssimato per la stima del rumore a partire dal diagramma di Bode della caratteristica di rumore sull'uscita: il **metodo della tangente rosa**. Si costruisce su di un diagramma di Bode la retta a pendenza - 10 dB/dec tangente al grafico (che dunque lo interseca in un unico punto); questo sarà il punto per cui si avran i massimi contributi di rumore, e intorno al quale converrà fare l'integrazione.

2.1.2 Modellistica del rumore

Resistori

Si consideri un generico resistore R: per la fisica, dato che $T > 0$ K, in serie ad esso vi sarà una potenza di rumore equivalente dall'espressione:

$$V_n^2(f) = 4k_BTR$$

Questo è rappresentabile mediante un equivalente Thevenin o Norton:

Se si integrasse questa espressione tra 0 e ∞, essendo un rumore bianco, sembrerebbe che la potenza prodotta dal resistore sia infinita; il problema è che in un resistore reale non si ha solo una resistenza, ma anche parametri parassiti, ad esempio una capacità. Un modello più completo potrebbe essere dunque:

$$N_R = 4k_BTR\frac{\pi}{2}\frac{1}{2\pi RC} = \frac{k_BT}{C}$$

L'integrazione è stata effettuata usando un trucco non riportato, basato sul calcolo della banda equivalente di un filtro passa basso a un singolo polo; ciò può essere usato per integrare ogni situazione in cui si abbia una pendenza pari a -20 dB/dec.

La dipendenza della potenza di rumore è solo da C, dove C è la capacità parassita.

Modello equivalente di rumore

Sull'ingresso di un circuito ci sarà un certo rumore termico, dipendente dalla capacità del resistore in ingresso; al fine di ridurre il rumore, però, si può anche introdurre una capacità, in modo da spostare il polo e ridurre *manualmente* il rumore.

Questo rumore però (termico) non è l'unico rumore esistente: una generica rete (per esempio un amplificatore) ha al proprio interno altre fonti di rumore; ciò che si fa è rappresentare il rumore interno mediante due generatori equivalenti di ingresso, per poi considerare la rete priva di rumore. i *generatori* si possono determinare così:

1. Se cortocircuito al riferimento l'ingresso, non si ha contributo della corrente in ingresso sull'uscita; prendo dunque l'uscita, e la *riporto all'ingresso*: questo sarà il contributo di tensione.

2. Se *apro* l'ingresso, il generatore di tensione non potrà pilotare l'uscita, ma il generatore di corrente sì, dunque ciò che si riporta all'ingresso sarà il solo contributo del generatore di corrente.

Ciò che si puà fare poi è, al posto di due generatori, usarne uno solo, equivalente: $V_{neq}^2(f)$, che varrà, per ogni carico:

$$V_{neq}^2 = V_{ni}^2(f) + I_{ni}^2(f)R_g^2 + V_{nR}^2(f)$$

Considerando il caso più generale in cui si ha anche un Z_g e un Z_{in}, si può dire che:

$$V_{in}^2(f) = V_{neq}^2(f) \left| \frac{Z_{in}}{Z_{in} + Z_g} \right|^2$$

2.1.3 Elementi attivi

Diodi

Si ha a che fare molto spesso con giunzioni, dunque con diodi; cosa capita in questi casi? Beh, dove ci sono giunzioni nei semiconduttori, vi sono sia contributi di rumore bianco, sia di rumore rosa, ma non solo: si ha a che fare con il cosiddetto *rumore shot*: anche il **shot** è un rumore bianco, ma esso dipende dal fatto che, quando si ha corrente, la conduzione avviene *per pacchetti*, quantizzata.

Si ha un'espressione del tipo:

$$V_D^2(f) = 2qI_D$$

Dove q è la carica dell'elettrone.

Un inciso molto importante: r_d, ossia la resistenza differenziale (la resistenza di segnale) del diodo, non è un resistore, ma una grandezza imposta per modellare determinate fenomenologie, dunque **non introduce rumore termico**.

$$r_d = \frac{V_T}{I_D}$$

Un diodo può essere usato come generatore di rumore noto, a meno che non si usi uno Zener: esso infatti, se usato in regione **avalanche**, contiene un altro tipo di rumore (che nella trattazione non verrà introdotto), molto fastidioso.

Transistori bipolari a giunzione

Usando la modellistica prima proposta, è possibile calcolare i due generatori equivalenti per il BJT.

Si ha a che fare con due giunzioni, dunque con rumore shot, e con qualche resistenza interna, che genererà del rumore; un esempio classico di queste resistenze è la $r_{bb'}$, ossia la resistenza della base, che è però un resistore reale, nato a causa del fatto che la base è una parte di semiconduttore poco drogata. A causa delle trappole, inoltre, si avrà a che fare con del rumore flicker:

sperimentalmente si è infatti visto che una componente molto importante di rumore flicker è sulla giunzione base-emettitore. Questo, ovviamente, si somma alle già presenti componenti shot, presenti in ogni giunzione: sia in quella base-emettitore, sia in quella base-collettore. Si ha, alla fine, che:

$$I_{ni}^2 = 2qI_B + \frac{K}{f} + \frac{2qI_C}{\beta^2(f)}$$

Di questi tre termini, il primo è il rumore shot nella giunzione base-emettitore, il secondo il rumore flicker in base, il terzo il rumore shot sulla giunzione base-collettore. L'unico (di solito) importante è il primo: il flicker diventa importante solo a basse frequenze, mentre il shot base-collettore solo quando $\beta(f)$ diventa molto piccolo, ed essendo al denominatore fa esplodere il numeratore).

Per quanto riguarda il generatore di tensione, i contributi derivano dalla resistenza di base $r_{bb'}$ (termico), e dal rumore shot del collettore, riportato in ingresso mediante la g_m (al quadrato, dal momento che si parla di potenze!):

$$V_{ni}^2(f) = 4k_BTr_{bb'} + \frac{2qI_C}{g_m^2}$$

Ora, un po' di manipolazione algebrica: sappiamo infatti che

$$g_m = \frac{I_C}{V_T} = \frac{I_Cq}{k_BT}$$

Dunque:

$$\frac{2qI_C}{g_m}\frac{k_BT}{qI_C} = 2\frac{k_BT}{g_m}$$

Dunque:

$$V_{ni}^2(f) = 4k_BT\left[r_{bb'} + \frac{1}{2g_m}\right]$$

I rumori shot e flicker **non sono presenti** dal momento che, assumendo $r_{bb'} \ll r_\pi$, cosa ragionevole per I_C relativamente basse, dal momento che:

$$r_\pi = \frac{\beta V_T}{I_C}$$

ma supponendo $r_{bb'} \sim 0$, quando si chiude la base in corto circuito, in sostanza si annichilisce, sotto il punto di vista delle tensioni (e delle cadute di tensioni su $r_{bb'}$), questi contributi. **Date queste ipotesi**, si può considerare, come già detto in precedenza per il discorso generale, che $V_{in,eq}^2(f)$ è la somma della V_{ni}^2, delle eventuali tensioni dovute a effetti del generatore (termici), e della I_{ni}^2, moltiplicata **per la sola** R_B^2 (al quadrato dal momento che è una potenza): di fatto $V_{in,eq}^2$ è un generatore equivalente Thevenin del **solo ingresso**, dunque non si considera, per modellare il solo ingresso (avendo noi **portato fuori** dal circuito i generatori V_{ni}^2 e I_{ni}^2), alcun contributo dell'impedenza di ingresso.

Transistori MOS a effetto di campo

Nel caso dei MOSFET la modellistica è semplificata, dal momento che i generatori di corrente sul gate possono essere considerati ininfluenti (si ha un minimo di contributo, ma non molto importante rispetto ad altri).

Sperimentalmente, si può osservare la presenza di un rumore flicker presente in prossimità dell'interfaccia ossido-semiconduttore: il transistore MOS conduce infatti in un canale, situato vicino alla superficie di contatto ossido-semiconduttore, dove vi sono trappole che catturano e rilasciano, secondo un processo casuale rosa, i portatori. Sperimentalmente si è verificato che questo processo è dipendente dalla corrente drain-source, secondo una certa funzione, del tipo:

$$N_{flicker}(f) = K\frac{I_D^{\alpha}}{f}$$

Dove I_D è la corrente di polarizzazione, K una costante dipendente dal dispositivo, α una costante tra 0,5 e 2.

Il rumore flicker nei MOSFET è stato largamente studiato, e ciò che è stato scoperto è che esso presenta una forte dipendenza dalla geometria del dispositivo, più che direttamente dalla corrente di polarizzazione I_D: a seconda della geometria (inversamente proporzionalmente rispetto all'ampiezza dell'area utile) infatti il numero di stati-trappola presenti varia, dunque più spesso si può trovare questa espressione del rumore flicker **già riportato all'ingresso**, dunque pronto per essere sommato ai contributi del generatore equivalente:

$$N_{in,flicker}(f) = \frac{K_f}{WLC_{ox}f}$$

Dove K_f tipicamente vale:

$$K_f = 3 \times 10^{-24}\mathrm{V^2F}$$

Il canale, inoltre (senza voler approfondire il discorso dell'analisi a canale graduale), è dotato di una certa resistenza, che **si dimostra** (ma qua non si fanno i calcoli in dettaglio, calcoli che richiederebbero l'integrazione sul canale) valere:

$$R_{canale} = \frac{2}{3}\frac{1}{g_m}$$

Questa resistenza introduce dunque un **rumore termico**. In definitiva, il generatore verrà usualmente definito come:

$$V_{ni}^2(f) = 4k_B T \frac{2}{3}\frac{1}{g_m} + \frac{K_f}{WLC_{ox}f}$$

Volendo essere pristini, si ha anche un contributo della corrente di gate di leakage, che scorre granulosamente, provocando un rumore shot:

$$I_{shot}^2(f) = 2qI_G$$

Ciò potrebbe essere sommato, in ingresso, a un contributo capacitivo donato da C_{gs}, ma ciò non verrà mai tenuto in conto nella trattazione.

Si ricordi che, sotto il punto di vista del rumore, i transistori a canale p sono migliori; un modo per ridurre il rumore (flicker) è quello di avere transistori grossi, e capacità elevata (dunque, ossido sottile).

C_{gs} non è stata, di fatto, tenuta in conto nel modello da noi introdotto: se si considerasse il MOSFET ad alte frequenze, sarebbe necessario tenere conto anche dei fenomeni di leakage e della corrente di drain riportata al gate, cosa che non si farà nella trattazione.

Cenni all'amplificatore operazionale

Si deve ricordare che si ha un certo $V_{ni}^2(f)$, su un morsetto (non conoscendo il segno del rumore, si può mettere, a propria discrezione, su un morsetto piuttosto che su un altro), e, se il primo stadio è a bipolari, si avrà anche un contributo di corrente. I generatori modellano ciò che sta dentro all'operazionale, dunque, per quanto non se ne terrà mai conto, si ricordi che potrebbero esistere delle correlazioni.

Capitolo 3

Filtri su circuiti monolitici

Volendo realizzare filtri su circuiti integrati, una prima difficoltà sta nell'integrare induttori: per avere poli complessi coniugati dovrò dunque usare, in qualche modo, dei circuiti attivi. Un secondo problema sono poi le resistenze: per ottenere grossi valori di resistenza dovrei mettere insieme, in un integrato, molte *resistenze quadre*, occupando inutilmente molta area, e ottenendo effetti sgradevoli, quali resistori non lineari.

Un modo per realizzare resistenze, come si vedrà meglio in seguito, è basata sui filtri tempo-discreti, a condensatori commutati: se infatti non siamo in grado di realizzare con precisione dei resistori, siamo in grado di realizzare con precisione (anche se a valori bassi) dei condensatori; il brutto è che, per commutare i condensatori, si deve però utilizzare un segnale a frequenze elevate, dunque questi filtri vanno benissimo, ma su segnali a bassa frequenza (come si vedrà meglio in seguito), come in banda audio, ma, andando su con la frequenza, la velocità degli switch dovrebbe essere troppo elevata.

Si vorrebbe inoltre avere la possibilità di *accordare* queste grandezze: le tolleranze di un integrato sono infatti pessime, dunque poter effettuare delle modifiche a posteriori sarebbe ottimo.

L'obiettivo sarà quello di realizzare resistori a resistenza elettricamente variabile, e una tecnica per far ciò è basata sugli OTA (Operational Transconductance Amplifiers), ossia gli **amplificatori di transconduttanza**.

Una nota: se vogliamo fare (come si tende sempre più a voler fare) dei sistemi on-chip (mettendo nello stesso integrato sia la parte analogica sia quella digitale), si hanno dei problemi: se da un lato i circuiti digitali sono molto poco suscettibili al rumore (per quanto sopra una certa soglia si ottengano risultati veramente pessimi, rispetto ai circuiti analogici, dal momento che quando si incomincia a leggere male tra uni e zeri si rischia di finire in stati imprevedibili), essi ne generano tanto, a causa dei fronti molto ripidi. Questi disturbi vanno poi ad accoppiarsi con parametri parassiti poichè, essendo tutto nello stesso chip, il campo elettromagnetico non deve percorrere molta strada: si hanno grossi problemi di autocompatibilità elettromagnetica.

Dato un circuito analogico che deve andare a parlare con un qualche carico *ascoltatore*, dato il potenziale di riferimento 0 V, esso può essere condiviso da più circuiti, dal momento che è un riferimento; il fatto è che potrebbero esservi resistenze parassite che *sporcano* questo riferimento, ottenendo in circuiti diversi riferimenti diversi. Se poi si hanno piste sovrapposte, si possono avere anche altri effetti, come accoppiamenti di tipo capacitivo, attraverso i vari substrati.

Fatti questi discorsi, cosa si può fare? Beh, separare i ritorni dei 0 V non si può fare, dal momento che si ha a che fare, in un integrato, con troppi circuiti a bordo del chip; la soluzione migliore è dunque quella di utilizzare una soluzione **bilanciata**, in modo che, anche se si ha un disturbo sul 0 V, esso sarà di certo di modo comune, modo che potrà essere eliminato dalla struttura bilanciata.

Inoltre, se si hanno delle capacità verso il substrato, se tutto è abbastanza simmetrico, i disturbi saranno di modo comune. Ciò ci dice che conviene dunque, quando possibile, usare soluzioni di tipo bilanciato.

Come si può passare dalla semplice idea a una soluzione circuitale? Consideriamo per esempio un integratore:

Questo circuito può essere ottenuto, in forma bilanciata, studiando i punti di 0 V, *riflettendo* il circuito rispetto a questi punti, costruendo un **circuito speculare**. Ciò significa che sarebbe però necessario raddoppiare il numero dei componenti, anche se ciò spesso non è necessario: è possibile utilizzare un singolo operazionale con due stadi di uscita: esso è simile agli operazionali già noti, con due stadi finali. Il costo aggiuntivo dunque, in termini di area, è irrisorio.

3.1 Filtri $g_m - C$

Abbiamo detto che fare i condensatori è facile, ma le resistenze (o almeno quelle tempo-continue) no. Si potrebbero usare amplificatori operazionali, ma essi hanno un grosso problema: una banda passante abbastanza limitata.

Come già accennato, la soluzione al problema è basata sull'uso di un nuovo tipo di dispositivo: gli amplificatori di transconduttanza. Questo è il simbolo funzionale:

dove, idealmente:

$$I_u = g_m(V_1 - V_2)$$

Idealmente, ho delle impedenze molto alte sia in ingresso, sia in uscita (in ingresso ho dei gate di MOSFET, in uscita le r_d dei medesimi). Essi si possono realizzare mediante uno stadio differenziale di ingresso, e un amplificatore di corrente in uscita:

Si avrà che:

$$I_u = I_2 - I_1' = g_m(V_1 - V_2)$$

Un oggetto di questo genere ha un unico stadio, un polo dominante, più il polo e lo zero dello specchio, che si *compensano*, introducendo un dispositivo sempre stabile. Non serve dunque (a meno di casi particolari) utilizzare una compensazione, e si ha una banda passante dunque molto più elevata rispetto a quella di un amplificatore operazionale.

In realtà questi stadi sono un po' più complicati di come mostrato: I_u ha infatti un andamento lineare in uscita solo se la differenza di tensioni $(v_1 - v_2)$ è piccola (decine di mV); servono dunque artifici per migliorare la dinamica di modo differenziale. Si noti però un fatto: il guadagno in corrente dipende da g_m, che, all'incirca, vale:

$$g_m = \sqrt{2 I_D \mu C_{ox} \frac{W}{L}}$$

ma

$$I_D = \frac{I_0}{2}$$

Variando dunque I_0 è possibile variare il guadagno, ottenendo di fatto dispositivi accordabili.

In realtà, sarebbe necessario tenere in conto anche i parametri parassiti, cosa che in realtà come per l'operazionale non si fa quasi mai:

In ingresso si hanno le capacità del sistema MOS, in uscita una conduttanza G_0 data dalle r_d in parallelo, quindi una C_u anch'essa in parallelo.

3.2 Circuiti basati sull'amplificatore di transconduttanza

Resistore con un capo al riferimento

Come si fa a realizzare un filtro senza resistori? Beh, si sa che una resistenza ha un comportamento di questo genere:

Data una tensione V_e di ingresso riferita a 0 V, si ha:

$$I_e = \frac{V_e}{R}$$

Si può fare dunque qualcosa del genere:

Dato v_e nell'ingresso (non si specifica ancora se + o -), esso è la Δv (essendo l'altro connesso a 0 V); se dunque lo mettiamo sul morsetto invertente, la corrente entrerà nel dispositivo anzichè uscire da esso; se non abbiamo carichi, dunque, la corrente potrà solo essere prelevata dall'ingresso, e ciò comporta avere esattamente la caratteristica prevista:

$$I_u = v_e g_m$$

Ma

$$I_e = I_u = v_e g_m$$

Dunque

$$R_{eq} = \frac{1}{g_m}$$

Ossia, abbiamo trovato un modo per realizzare una resistenza senza ricorrere a resistori.

Volendo provare a usare il modello non ideale, si ottiene:

$$I_e = V_e \left[sC_i + sC_u + g_m + G_0 \right]$$

Dunque:

$$R_{eq} = \frac{1}{G_0 + g_m + s(C_i + C_u)} \sim \frac{1}{g_m}$$

Si avrebbe a che fare in realtà con un polo, che però è situato a frequenze altissime, dunque non ci preoccupa.

Resistore flottante

Se il resistore non è riferito al potenziale di riferimento, ossia è flottante, si ha qualcosa di questo genere:

$$I = \frac{V_1 - V_2}{R}$$

Dove si **suppone** che $V_1 > V_2$.

Come si può fare? Beh, l'obiettivo è quello di *sentire* la differenza di tensione $V_1 - V_2$, dunque realizzare una corrente a essa proporzionale. Volendo inoltre che la corrente venga *risucchiata* dal primo operazionale, e *espulsa* dal secondo, servirà qualcosa di questo genere:

Supponendo $g_{m1} = g_{m2} = g_m$, si ha:

$$I = g_m(V_1 - V_2)$$

Essendo infatti i punti 1 e 2 alla stessa tensione, non entrando corrente nei morsetti, si ottiene proprio ciò.

Sommatore di correnti

Un altro esempio di circuiti può essere il sommatore di correnti:

Integratore

Per fare un integratore, data una tensione di ingresso, si vuole avere una corrente costante da poter mandare a caricare una capacità. Mediante l'amplificatore operazionale, ciò che si faceva era sostanzialmente *creare una massa virtuale*, in modo da avere un punto che non cambiasse la tensione, e dunque neanche la corrente da integrare, cosa che invece sarebbe capitata mettendo semplicemente una resistenza e una capacità: al caricarsi della capacità, essa varia la propria tensione, variando però anche la tensione del resistore, dunque la corrente, ottenendo di fatto l'andamento esponenziale ben noto; con l'operazionale il collegamento al potenziale di riferimento visto dalla resistenza è *inflessibile*, dunque anche la corrente sul resistore costante; non entrandone nei morsetti, essa andrà a caricare la capacità.

Ciò che si fa con gli OTA è semplicemente ciò:

Nota: questo è un integratore non invertente, cosa che, con i circuiti basati gli amplificatori operazionali, non si poteva realizzare.

Induttore con un capo al riferimento

A questo punto si vorrebbe realizzare un circuito che si comporti come un'induttanza, ossia per cui:

$$I_L = \frac{V_e}{sL}$$

Ciò risulta utile ogni volta che si deve a che fare con l'introduzione di poli complessi coniugati in una rete elettronica. Sembra una cosa complicata, ma gli elementi per realizzare ciò ci sono già tutti: prima si è visto infatti che l'integratore ha un andamento del tipo:

$$V_C = \frac{g_m}{sC} V_e$$

Se riuscissimo ad assorbire in ingresso una corrente proporzionale alla tensione di uscita, si potrebbe ottenere la funzione di trasferimento desiderata. Ciò però è banale: basta utilizzare un altro OTA, in questa maniera:

$$I = \frac{g_{m1} g_{m2}}{sC} V_e$$

Dunque:

$$L_{eq} = \frac{C}{g_{m1} g_{m2}}$$

Induttore flottante

Volendo realizzare un'induttanza flottante, tra due nodi:

$$I_L = \frac{V_1 - V_2}{sL}$$

La base sarà ancora un integratore, ma ora si deve fare in modo (come già fatto per il resistore flottante) di produrre una corrente dipendente dalla differenza di tensioni:

$$I_L = \frac{g_{m1} g_{m2}}{sC} (V_1 - V_2)$$

Da qua

$$L_{eq} = \frac{C}{g_{m1} g_{m2}}$$

Realizzazione di buffer

Un problema degli OTA è il fatto che, avendo un'elevata impedenza di uscita, possono essere facilmente *condizionati dal carico*: la funzione di trasferimento del circuito sarà fortemente funzione del carico applicato. Una soluzione per questo problema è l'usare come carico un altro OTA, data la sua elevatissima impedenza di ingresso; si avrà dunque qualcosa del tipo:

In uscita da ogni OTA si avrà una certa Y_0. L'amplificazione globale sarà, supponendo che tutti gli OTA siano identici tra loro:

$$V_u = \left(\frac{g_m}{Y_0} \right)^{n-1} \left(\frac{g_m}{Y_L + Y_0} \right) (V_{in} - V_u)$$

Dunque:

$$V_u\left(1+\left(\frac{g_m}{Y_0}\right)^{n-1}\frac{g_m}{Y_L+Y_0}\right)=\left(\frac{g_m}{Y_0}\right)^{n-1}\left(\frac{g_m}{Y_L+Y_0}\right)V_{in}$$

Da qui:

$$\frac{V_u}{V_{in}}=\frac{\left(\frac{g_m}{Y_0}\right)^{n-1}\left(\frac{g_m}{Y_L+Y_0}\right)}{1+\left(\frac{g_m}{Y_0}\right)^{n-1}\left(\frac{g_m}{Y_L+Y_0}\right)}$$

Dove n è il numero degli OTA usati. Ciò è spesso spiegato in molti libri, ma si può anche usare un singolo operazionale come voltage follower: quando si mettono infatti diversi stadi basati sugli OTA in cascata (2 o 3), si deve compensare, dal momento che si introducono nella rete diversi poli, dunque per compensare si riduce l'amplificazione, e di solito non ne vale la pena.

Se $g_m \gg Y_p$, cosa che avviene con appunto due o tre OTA, si ha che:

$$\frac{V_u}{V_{in}}\sim 1$$

3.3 Filtri MOSFET-C

Una tecnica alternativa per la realizzazione di filtri è quella basata sulla tecnica MOSFET-C: essa è basata sull'uso di transistori MOS a effetto di campo, in zona resistiva. Per quanto questa struttura abbia diversi problemi, quali il fatto di dover usare degli amplificatori operazionali (dunque poca banda), essendo semplice ricondurre le celle *classiche* (come la Sallen-Key) a questa tecnologia, spesso i progettisti la prediligono.

In zona resistiva, per $V_{DS} < V_{GS} - V_{th}$, si ha che:

$$I_D=\mu C_{ox}\frac{W}{L}\left[(V_{GS}-V_{th})V_{DS}-\frac{V_{DS}^2}{2}\right]$$

La I_D è controllabile, ma dipende anche dalla V_{DS}; se derivo:

$$\frac{\partial I_D}{\partial V_{DS}}=\mu C_{ox}\frac{W}{L}\left[(V_{GS}-V_{th})-V_{DS}\right]$$

Questa conduttanza non è costante, ma dipende da V_{DS}, dunque non è lineare.

Per linearizzare, si usa una struttura bilanciata, con v_S identica sui due source, v_C uguale sui due gate, v_e e $-v_e$ sugli ingressi (situati nel drain). Facciamo due conti:

$$i_{D1} = \mu C_{ox} \frac{W}{L} \left[(v_C - v_S - V_{th})(v_e - v_S) - \frac{(v_e - v_S)^2}{2} \right]$$

$$i_{D1} = \mu C_{ox} \frac{W}{L} \left[(v_C - v_S - V_{th})(-v_e - v_S) - \frac{(-v_e - v_S)^2}{2} \right]$$

Calcoliamo la differenza delle due correnti drain-source:

$$i_{D1} - i_{D2} = \mu C_{ox} \frac{W}{L} \left[2v_e(v_C - v_S - V_{th}) + 2v_e v_S \right] =$$

$$= \mu C_{ox} \frac{W}{L} 2v_e \left(v_C - V_{th} \right)$$

Ciò elimina dunque la dipendenza da v_S; derivando a questo punto rispetto alla tensione di ingresso v_e, e calcolando il reciproco, si ottiene:

$$R_{eq} = \frac{1}{2\mu C_{OX} \frac{W}{L}(v_C - V_{th})}$$

Come si può realizzare questa cosa? Beh, serve un qualche dispositivo in grado di mantenere uguali le v_S con la massima precisione possibile: questo è un amplificatore operazionale!

3.4 Circuiti per l'accordo di filtri integrati monolitici tempo continui

Prima di introdurre questo argomento, si vuole ricordare il problema delle tolleranze di fabbricazione: le g_m si possono infatti controllare con tolleranze del 20%, i condensatori con tolleranze del 10%, dunque le costanti di tempo, pari a $\frac{C}{g_m}$, saranno dell'ordine di 30% di tolleranza.

Dal momento che però i rapporti tra grandezze omogenee sullo stesso integrato si riconducono a rapporti tra aree, i rapporti possono essere progettati con tolleranze anche dell'1%, e la stessa cosa vale per capacità e transistori. Le tecniche di progetto che si vogliono introdurre partiranno da questa idea.

Su un integrato posso fare qualcosa di questo tipo: se il filtro è costituito da amplificatori di transconduttanza, posso introdurre sistemi per la loro *gestione*, o meglio per la gestione, con elevata precisione, della loro g_m, in modo da regolare g_m mediante una tensione di controllo, magari in maniera automatica. Un'idea può essere la seguente:

Si consideri per ipotesi che, se v_C aumenta, aumenta anche la g_m dell'OTA.

L'operazionale impone che la v_i valga 0, dunque che sul nodo B vi sia una massa virtuale; sulla resistenza R cadrà una tensione pari a $+V_p$, con il + dalla parte di B. Si ha dunque che V_R coincide con V_p, ma dunque che:

$$V_R = V_p$$

$$V_p = g_m V_p R$$

Da qua:

$$g_m = \frac{1}{R}$$

Dunque, a seconda di R, e della sua tolleranza, si ottiene una g_m altrettanto precisa; questa v_C in uscita inoltre può essere utilizzata per pilotare altri amplificatori di transconduttanza presenti nello stesso integrato, poichè *uguali* a quelli che già abbiamo.

Questa è una reazione negativa: se per caso g_m fosse più grande del dovuto, la tensione sul punto B si abbasserebbe, ma dunque il segnale, invertito e portato a v_C, andrebbe ad abbassare la g_m, *controllandola* con questo meccanismo di controreazione. La capacità C limita la banda passante, dunque il rumore: questo circuito, dopo un transitorio, funziona quasi solo in continua, dunque limitare la sua banda può solo fargli bene. Questo controllo si può realizzare alla fabbricazione dell'integrato, magari utilizzando un potenziometro.

Questo metodo è molto bello, ma ha un grosso problema: non è automatico! Si dovrebbe fare un trimming per ogni oggetto prodotto, e questo inoltre con l'invecchiamento, con la temperatura, con qualsiasi cosa, non varrebbe più. Esistono però metodi automatici di trimming: si dovrebbe misurare la frequenza di risonanza e il Q del filtro, dunque correggere automaticamente in modo da avere sempre i parametri ottimali.

Ciò che si fa all'interno dell'integrato è implementare in un punto il filtro, e in un altro punto un dispositivo che riproduca alcune delle grandezze fondamentali del filtro, con elementi identici: questo sarà detto **master**. Il filtro vero e proprio è detto **slave**, dal momento che esso è controllato da ciò che fanno il controllo e il master. In questo modo, lo slave fa solo da filtro, e non dobbiamo mandare in ingresso nessun riferimento: introdurre diversi segnali a diverse frequenze potrebbe provocare strani effetti, direttamente sul filtro, per questo si cerca di disaccoppiare i circuiti in cui si introduce da un lato riferimenti e dall'altro i segnali veri e propri. A volte può essere utile agire sul filtro anche per modificare il Q, ma raramente: essendo il Q un rapporto di grandezze omogenee, esso sarà realizzabile con elevata precisione; si dovrebbe mettere un circuito in grado di agire sull'ampiezza del segnale.

Alcune osservazioni:

- la frequenza di V_r non deve essere troppo prossima a quella di risonanza dello slave, se no si rischia di *sporcare* il segnale utile; d'altra parte non deve manco essere troppo diversa, altrimenti le risposte di master e slave sono troppo diverse;

- se si varia la frequenza di V_r, varierà il segnale di controllo, dunque la risposta del master, e la frequenza di risonanza del filtro varierà di conseguenza; questo significa che il filtro è accordabile dall'esterno: **filtro tracking**.

Il blocco master+controllo può essere realizzato mediante un PLL:

Il master è un VCO, e l'uscita ha una frequenza dipendente dalla tensione di controllo. Sostanzialmente, il controllo agisce in modo che, man mano che ci si avvicina alla frequenza giusta, si tende ad agganciarla. Questa tensione di uscita viene poi mandata allo slave: essa sarà significativa dal momento che, nel master, si avrà un componente analogo a quelli dello slave, un *componente critico*.

Si propone a questo punto un esempio di schema che si può utilizzare per la realizzazione di un VCO:

Vediamo come funziona: variando v_C al solito si varia g_m, dunque I, dunque la frequenza; a circuito spento, si suppone che il comparatore (a due soglie, entrambe positive) sia spento; questo significa che V_u è nulla, e che dunque il condensatore è scarico. Quando attacco tutto, tutta la corrente va a caricare la capacità, dal momento che T_3 è interdetto, dunque non passa corrente, dunque neanche in T_1 e T_2, e tutta la corrente deve per forza andare nel diodo, dunque nella capacità, facendo aumentare V_u linearmente nel tempo. Quando V_u supera entrambe le soglie, scatta il comparatore, e dunque si ha che T_3 satura, diventando come un corto circuito; sul punto A si ha una tensione pari a $V_{sat3} + V_{BE1}$, il diodo dunque si interdice, tutta la corrente va nel ramo con i transistori, e trova il lato debole dello specchio, dunque T_2 rispecchia la corrente, va a scaricare il condensatore, con in modulo la stessa corrente con cui era stato caricato, fino a che non si sorpassano entrambe le soglie. A questo punto si ricomincia come prima, e così via.

V_u' è un'onda quadra alla stessa frequenza dell'onda triangolare: essa sarà l'uscita del VCO. Si ha che:

$$T_1 = (V_{S2} - V_{S1})\frac{C}{I} = (V_{S2} - V_{S1})\frac{C}{g_m V_T}$$

Controllando v_C controllo g_m, dunque la frequenza.

Capitolo 4

Filtri a condensatori commutati

A questo punto, al posto dei resistori, vogliamo usare un processo tempo-discreto, in modo da realizzare qualcosa di equivalente e semplicemente utilizzabile.

Per un generico resistore, come ben noto:

$$I = \frac{V_2 - V_1}{R}$$

Utilizziamo qualcosa di questo genere:

Si deve assolutamente evitare che Φ_1 e Φ_2 siano entrambi chiusi: ciò potrebbe comportare un corto circuito tra V_1 e V_2.

Ora: se apro Φ_2, chiudo Φ_1, ho V_1 ai capi di C, dopo un certo transitorio di carica per la capacità; la carica sulla capacità Q sarà pari a:

$$Q_1 = V_1 C$$

A questo punto, se apro Φ_1 e chiudo Φ_2, supponendo $V_1 > V_2$, la capacità dovrà cedere una carica pari a ΔQ:

$$\Delta Q = (V_1 - V_2)C$$

Dunque, alla fine di questo processo, avrò una carica pari a:

$$Q_2 = V_2 C$$

Allora, se apro e chiudo, ogni volta vi sarà uno scambio di carica pari a ΔQ, e dunque questa si aggiungerà e toglierà ogni volta; questo significa che, a ogni ciclo, la capacità andrà a scambiare dei *pacchetti* di carica.

La corrente che passa, I, sarà:

$$I = \frac{\Delta Q}{T_C}$$

Dove T_C è il periodo del clock che intercorre per il ciclo.

Si ricordi che questa è una corrente fortemente *granulosa*, quantizzata; per vederla *fluida* è necessario che i segnali che vengono utilizzati in questi circuiti abbiano frequenze ben inferiori rispetto a quelle usate per pilotare il circuito: la velocità con cui si variano V_1 e V_2 deve essere sufficientemente più bassa di quella di switching (vedremo che deve essere rispettato di sicuro Nyquist, ma si deve avere molto di più).

Si noti che, anche se non sembra, questo sistema dissipa potenza: ciò deriva dal fatto che gli interruttori hanno delle resistenze in serie, quando si chiudono; capita dunque che:

Dalla V_1, quando carico il condensatore, ho un andamento esponenziale della corrente: all'inizio si ha $\frac{V}{R_S}$ (tutto cade sulla R_S, essendo il condensatore vuoto infatti esso è come un corto), poi:

$$i_C(t) = \frac{V}{R_S}\mathrm{e}^{-\frac{t}{\tau}}$$

L'energia sarà dunque pari a:

$$E = \int_0^\infty V\frac{V}{R_S}\mathrm{e}^{-\frac{t}{RC}}dt = -\frac{V^2}{R_S}R_SC\mathrm{e}^{-\frac{t}{RC}}\Big|_0^\infty = \frac{V^2}{R_S}R_SC = V^2C$$

Sapendo che l'energia immagazzinata in una capacità è:

$$E_C = \frac{1}{2}CV^2$$

Ho che in R_S si dissipa metà della potenza, mentre l'altra metà si accumula in C; quando poi svuoto la capacità, lo faccio su un'altra resistenza serie, dove si svuoterà l'altra metà di CV^2; avendo che:

$$P = \frac{E}{T} = \frac{E}{T_C} = V^2\frac{C}{E} = \frac{V^2}{R_{eq}}$$

Ho che la potenza di fatto si dissipa su questa resistenza equivalente, esattamente come in una qualsiasi resistenza.

Un esempio di circuito che può generare i due segnali in controfase e non sovrapposti potrebbe essere la seguente implementazione di Flip-Flop:

Si noti che, se usassimo un flip-flop NAND; non si potrebbe ottenere lo stesso risultato.

Si avrebbe infatti un andamento, nel tempo, di questo tipo:

Il periodo del clock deve essere abbastanza lungo rispetto alla costante di tempo $\tau = R_SC$, in modo da permettere alla capacità di caricarsi; supponendo che $R_S = 1$ kΩ, e che $C = 10$ pF, si avrebbe

$$\tau = R_S C = 10\,\text{ns}$$

Per avere un errore dell'un per mille è necessario aspettare circa 7 τ, dunque circa 70 ns per la scarica: la frequenza di clock dovrà per forza essere inferiore a circa 7 MHz.

Un circuito spesso utilizzato per realizzare un generatore di clock a frequenza fissa potrebbe essere il seguente:

Il condensatore agisce come un *corpo rigido* per i cambi di tensione: in questo modo si riesce a mandare la tensione da una parte all'altra, invertendo lo stato.

Si ha un andamento nel tempo di questo tipo:

L'ingresso di ciascuno di questi derivatori ha una forma di questo tipo:

Si usano dei diodi di clamp per far in modo da tagliare la tensione in ingresso ai transistori, per evitare che vengano distrutti; i diodi dunque fanno in modo che gli ingressi non vadano sopra la tensione di alimentazione. Introdurre la resistenza R' fuori dall'inverter fa sì che essa si *prenda la tensione in più*, dissipandola ed evitando effetti strani. Se R' è grande rispetto a R, allora la costante di tempo su cui si calcola l'oscillazione non verrà perturbata molto.

4.1 Circuiti a capacità commutate

Amplificatore invertente

Si supponga ora di voler realizzare un amplificatore invertente, basandosi sul principio delle capacità commutate. Lo schema di partenza è il seguente:

Un'idea potrebbe esser quella di implementare direttamente le resistenze, come finora visto, mediante i processi prima visti:

Questo schema non funziona, sostanzialmente per il seguente motivo: in questo schema, l'amplificatore operazionale non ha mai l'anello chiuso, dunque non è controreazionato! Un'idea migliore è la seguente:

Quando Φ_1 è chiuso, il condensatore C_1 è collegato a V_e, dunque dopo un transitorio si accumula una carica $Q = C_1 V_e$; nel mentre, C_2 tende a scaricarsi sull'uscita. Quando Φ_2 è chiuso, e gli altri sono spenti, si ha che C_1 è collegato a C_2, dunque la carica presente in C_1 verrà trasferita in C_2, ottenendo:

$$V_u = \frac{C_1}{C_2} V_e$$

Questo dunque è il guadagno dell'amplificatore.

4.1.1 Integratori

Per fare filtri un modo è quello di realizzare dei circuiti integratori:

Si ricordi che questo circuito funziona sostanzialmente perchè, se l'amplificatore operazionale è in linearità, v_d è circa nulla, dunque il punto A si trova a massa virtuale; scorre dunque una corrente pari a

$$I = \frac{v_e}{R}$$

questa corrente viene dunque deviata interamente nella capacità C, dove viene integrata.

La funzione di trasferimento è:

$$V_u = \frac{V_e}{R}\frac{-1}{sC}$$

Dunque:

$$\frac{V_u}{V_e} = -\frac{1}{sRC} = -\frac{1}{j\frac{f}{f_0}}$$

Dove

$$f_0 = \frac{1}{2\pi RC}$$

L'idea per la realizzazione, mediante condensatori commutati, di questo circuito, è quella di sostituire banalmente alla resistenza R l'ormai solito circuito:

Al posto di R introduciamo dunque R_{eq}, ottenendo:

$$\frac{V_u}{V_e} = -\frac{f_C C_1}{j\omega C_2}$$

Dove f_C è la frequenza del clock impiegato per pilotare il circuito a condensatori commutati. Volendo si può banalmente calcolare f_0, come:

$$f_0 = \frac{C_1 f_C}{2\pi C_2}$$

Cosa ci dice ciò? Beh, volendo un certo rapporto tra la frequenza di clock e la frequenza f_0, si dovrà avere:

$$\frac{f_0}{f_C} = \frac{C_1}{2\pi C_2}$$

Dunque, la frequenza che determina le prestazioni dell'integratore dipende dal rapporto di due capacità (grandezza che si può realizzare con grande precisione), e dalla frequenza degli interruttori (ma la sintesi di frequenza non è difficile da realizzare); se vario f_C, vario dunque la frequenza dell'integratore: si ha un meccanismo di trimmering, basato sulla variazione di frequenza di clock f_C.

Data la presenza del 2π, volendo che $f_C \sim 100 f_0$, devo impostare le capacità in modo che:

$$\frac{C_1}{2\pi C_2} = \frac{1}{100} \longrightarrow C_1 \sim 16 C_2$$

Purtroppo, a differenza di come potrebbe sembrare, non sono tutte rose e fiori: il rapporto tra C_1 e C_2 sarebbe ben definibile, se non fosse per i parametri parassiti presenti nel circuito. Volendo esplicitare gli interruttori mediante i MOSFET con cui vengono realizzati, si ha qualcosa di questo tipo:

Questi transistori presentano diversi parametri parassiti: il primo è la capacità tra gate e canale; sul gate si ha un'onda quadra di notevole ampiezza, dunque si hanno, date le controfasi, due onde quadre in opposizione di fase alla frequenza del clock, onde che iniettano carica nella capacità di integrazione, variando imprevedibilmente il comportamento dell'uscita. Si consideri anche solo dunque il blocco Φ_2 (onda quadra sul secondo MOSFET), C_p relativa alla capacità gate-canale, e C_2: questo circuito, di fatto, è un amplificatore invertente, con amplificazione

$$A_v = -\frac{C_p}{C_2}$$

Se C_2 è grossa, questo effetto si può ridurre; se inoltre f_C è molto più grande di f_0, ossia della frequenza che vogliamo trattare, gli effetti parassiti saranno già stati tagliati dagli altri filtri.

Il problema è che questa non è l'unica capacità: vi sono ancora le varie capacità tra MOSFET e substrato: quella sull'ingresso, essendo in parallelo a un generatore di segnale di tensione, non ci da grossi problemi, mentre invece le altre *sporcano* C_1, dal momento che sono in parallelo a essa! Essendo poi $C_1 \ll C_2$, può capitare che C_1 sia dell'ordine dei pF, e dunque che i parametri parassiti siano effettivamente confrontabili con la capacità C_1.

Sono state dunque studiate delle soluzioni atte a risolvere il problema delle capacità parassite, come strutture di questo tipo:

Si usano quattro interruttori al posto di due: non è un problema, dal momento che integrare due MOSFET in più ha un costo assolutamente irrisorio in termini di area.

Come funziona questo circuito? Vediamo!

- Quando Φ_2 è chiuso e Φ_1 è aperto, C_1 si scarica completamente, e dunque $V_{C1} = 0$.

- Quando si scambia, dunque Φ_2 aperto e Φ_1 chiuso, C_1 vede v_e a sinistra, e massa virtuale a destra, dunque assumerà una carica $\Delta Q = v_e C_1$; questa carica è stata assunta dal momento che è passata una corrente che è andata anche in C_2: la corrente che ha caricato C_1 dunque passa anche su C_2, e ha versato anche in C_2 una ΔQ, carica che però nell'istante seguente viene mantenuta: C_2, di fatto, somma per ogni ciclo la carica, effettuando l'integrazione. Si ha che la corrente che carica C_1 e C_2 è pari a:

$$I = \frac{\Delta Q}{T_C}$$

  Dunque:

$$R_{eq} = \frac{T_C}{C_1} = \frac{1}{C_1 f_C}$$

Prima di continuare lo studio di questo circuito, si supponga di invertire la temporizzazione: ora, si ha che C_1 e C_2 si caricano assieme, e quando C_1 si scarica C_2 mantiene; si può usare una configurazione differente:

Come funziona?

- Quando si chiudono i Φ_1 e si aprono i Φ_2, C_1 si carica a v_e, mentre C_2 non si carica nè scarica, ma mantiene ciò che già aveva dentro: non si ha transito di corrente. Su C_1 si finisce per avere, al solito, una carica pari a $C_1 v_e$.

- Quando si inverte la configurazione, dunque si chiudono i Φ_2 e aprono i Φ_1, accade che tutta la carica presente in C_1 si va a svuotare in C_2, con una nota: dal momento che C_1 è polarizzato in segno opposto, vi sarà una corrente, in C_2, fluente da destra verso sinistra, e non viceversa come prima; data la convenzione di utilizzatore, dunque, la tensione su C_2 sarà dello stesso verso di v_u, ottenendo, di fatto, un integratore non invertente.

A questo punto, abbiamo visto come realizzare, a condensatori commutati, un integratore non invertente; perchè questo circuito però è meglio di prima? Beh, scopriamolo! Sostituendo agli interruttori ideali i MOSFET, si ottiene:

Le capacità tra gate e canale sono sempre fastidiose come prima, dunque sotto questo punto di vista nessun miglioramento; il miglioramento vi è sotto il punto di vista delle capacità tra drain e substrato:

- C_{p1} non dà al solito fastidio, poichè è in parallelo all'ingresso;
- C_{p2}, se l'interruttore è chiuso va in parallelo a C_{p1}, non dando problemi, mentre se l'interruttore è aperto l'altro è chiuso, dunque è in parallelo a un corto circuito (idealizzato, dal momento che il transistore ha comunque resistenze parassite) verso massa, e non da contributo sull'uscita.
- C_{p3} è sempre con tensione 0 al proprio capo: se il MOSFET 4 conduce allora è tra massa e massa virtuale, altrimenti è in parallelo a un corto circuito.
- C_{p4} è sempre connessa tra massa e massa virtuale.

Questa struttura dunque sente meno l'influenza delle capacità parassite, C_1 non è in alcun modo *sporcata*, e ora il rapporto è realizzabile con grande precisione.

4.1.2 Analisi formale dell'integratore

Si consideri l'ultimo schema analizzato:

A questo punto, prima si proporrà uno studio delle forme d'onda nel dominio del tempo, per poi passare al dominio della trasformata $\mathcal{Z}$, e dunque in frequenza

Nel dominio del tempo, si ha un andamento di questo tipo:

v_e deve variare molto lentamente rispetto al clock; prima che Φ_1 vada a 1, C_1 è scarico; quando si chiude, la capacità deve caricarsi a v_e; quando si apre Φ_1 e si chiude Φ_2, la capacità C_1 si scarica a 0 e la carica che ha accumulato va nella C_2, facendo variare la v_u che (essendo un integratore), avrà già avuto un valore prima, accumulato. Si ha dunque che, al generico istante discreto nT_C:

$$v_u(nT_C) = v_u[(n-1)T_C] + v_e[(n-1)T_C]\frac{C_1}{C_2}$$

Spieghiamo: la tensione di uscita all'n-esimo istante è uguale a quella presente nell'istante precedente, più la v_e dell'istante precedente (dal momento che solo all'n-esimo istante si può scaricare la carica), con un'amplificazione pari a $\frac{C_1}{C_2}$. Si passi dunque alla $\mathcal{Z}$:

$$V_u = V_u z^{-1} + V_e z^{-1}\frac{C_1}{C_2}$$

Dunque:

$$V_u \left(1 - z^{-1}\right) = \frac{C_1}{C_2} V_e z^{-1}$$

Dunque:

$$\frac{V_u}{V_e} = \frac{C_1}{C_2}\frac{z^{-1}}{1 - z^{-1}}$$

Ma noi sappiamo che:

$$\frac{C_1}{C_2} = 2\pi\frac{f_0}{f_C}$$

Dunque:

$$\frac{V_u}{V_e} = 2\pi\frac{f_0}{f_C}\frac{z^{-1}}{1 - z^{-1}}$$

Ora, passando al dominio della frequenza:

$$\frac{V_u}{V_e} = H(f) = 2\pi\frac{f_0}{f_C}\frac{\mathrm{e}^{-j2\pi\frac{f}{f_C}}}{1 - \mathrm{e}^{-j2\pi\frac{f}{f_C}}}$$

Noi vorremmo che

$$H(f) = \frac{1}{j\frac{f}{f_0}}$$

Ma questo sembra lontano dall'ideale. Non è del tutto vero: introduciamo alcune manipolazioni algebriche, come moltiplicare e dividere per $j\frac{f}{f_0}$, e per $\mathrm{e}^{j\pi\frac{f}{f_C}}$:

$$H(f) = \frac{j\frac{f}{f_0}}{j\frac{f}{f_0}} 2\pi\frac{f_0}{f_C}\frac{\mathrm{e}^{-j\pi\frac{f}{f_C}}}{\mathrm{e}^{j\pi\frac{f}{f_C}} - \mathrm{e}^{-j\pi\frac{f}{f_C}}} =$$

$$= \frac{1}{j\frac{f}{f_0}}\frac{\frac{\pi f}{f_C}}{\sin\left(\pi\frac{f}{f_C}\right)}\mathrm{e}^{-j\pi\frac{f}{f_C}}$$

Si ha a che fare con due tipi di errore, rispetto al risultato atteso:

- un errore sul modulo, dipendente da $\frac{f}{f_C}$: questo ci dice che più la frequenza di clock f_C è alta, più il termine tende a 1. Questo fatto completa il già annunciato discorso riguardo alla lentezza dei segnali rispetto alla frequenza di clock: per questo specifico caso, abbiamo quantificato l'errore.

  Idealmente, al crescere di f, si dovrebbe aver un andamento di questo genere; in pratica, non si arriva mai ad avere qualcosa di prossimo a $f = f_C$, dunque l'andamento non si ha mai. Supponendo che

  $$f = \frac{f_C}{10}$$

  L'errore sarebbe:

  $$\frac{\frac{\pi}{10}}{sin\frac{\pi}{10}} \sim 1$$

- per quanto riguarda il secondo errore, esso è sulla fase: la fase è molto più suscettibile del modulo rispetto alle frequenze alte, infatti, nell'esempio di $f_C = 10f$, si avrebbero circa 18° di errore, dunque un errore molto elevato.

  Si fa in modo dunque da avere una frequenza di clock molto più elevata, almeno 10 volte quella del segnale, ma si consiglia molto caldamente di avere almeno 100 volte di differenza.

Aliasing

I problemi non sono finiti: oltre a tutto ciò che si è detto, si ha ancora il problema dell'aliasing. La presenza dell'aliasing comporta che l'operazione di campionamento faccia replicare la banda base:

Se si ha un disturbo, anche esso verrà replicato: se esso è prossimo alla frequenza di campionamento, ce lo ritroveremo pure in banda base. Il fenomeno del campionamento fa sì che i disturbi vengano ribaltati, e inseriti nel segnale utile.

Un dettaglio: il campionamento, in realtà, non viene fatto nel circuito visto, ma a priori: all'ingresso del filtro! Ciò che trattiamo con quel circuito deve essere già stato campionato, e la banda che si usa dal campionatore in poi è solo la banda del circuito.

Questo, per dire che, volendo evitare il fenomeno dell'aliasing, è indispensabile fare in modo che all'ingresso non vi siano disturbi alla frequenza del clock; l'unico modo per garantire ciò è mettere un filtro passa basso tempo-continuo davanti a tutto, per attenuare i disturbi alla frequenza del clock, in modo tale da renderli poco importanti. Davanti a un filtro tempo discreto devo dunque metterne uno tempo continuo, ma molto *blando*, facile da progettare: un RC, magari con componenti esterni all'integrato. Allo stesso modo un filtro analogo a questo andrà messo sull'uscita, in modo da *tagliare i gradini* dovuti al processo di campionamento, in maniera da **ricostruire** la continuità del segnale.

Integratore differenziale

Proponiamo a questo punto un altro circuito:

Al posto dei tradizionali quattro interruttori usiamo solo due deviatori (comodità di disegno, ma la realizzazione è del tutto analoga). Quando gli interruttori sono nella posizione 1, si ha che C_1 si carica con una carica pari a:

$$C_1(V_1 - V_2)$$

Nella posizione 2 poi, C_1 si scarica, ottenendo:

$$\Delta Q = (V_{e1} - V_{e2})C_1$$

Questa è la ΔQ che si da per ogni T_C; se si divide dunque per T_C, si ottiene la corrente, dunque se si moltiplica per l'impedenza della capacità di integrazione, si ottiene la tensione di uscita:

$$V_u = (V_1 - V_2)\frac{C_1}{sC_2T_C}$$

Questo è un integratore che integra la differenza dei due ingressi.

Filtro biquadratico a condensatori commutati

Volendo realizzare a condensatori commutati la solita cella biquadratica a variabili di stato, si può fare qualcosa di questo tipo:

Prima di tutto, identifichiamo i vari ruoli: C_2 e C_3 sono capacità di integrazione, mentre le altre sono **capacità serbatoio**, che conterranno provvisoriamente la carica da accumulare nelle altre. Il primo operazionale deve avere in sè i contributi dell'ingresso, del secondo operazionale e dell'uscita di sè stesso. Come fare dunque? Beh, in sostanza, un'idea è quella di introdurre, ovunque serva una resistenza, la solita struttura a deviatori (che abbiam capito essere migliore dell'altra); si possono dunque sostanzialmente identificare due fasi: una fase di carica, e una fase di integrazione.

- Nella fase 1 si collegano le varie capacità C_1, C_1', C_1'' ai contributi da sommare, caricandole provvisoriamente delle varie tensioni. C_1 sarà collegata tra ingresso e v_{u2}, C_1' tra v_{u1} e massa virtuale, C_1'' tra v_{u1} e 0 V, in modo da portare carica al secondo operazionale al colpo successivo.

- Nella fase 2 si ha che C_1 viene collegata a C_2 e scollegata dall'ingresso e da v_{u2}, poi C_1' viene collegata a C_2, e C_1'' viene collegata a C_3, in modo da svuotare tutto.

4.2 Pompe di carica

Si vuole a questo punto introdurre una tecnica per realizzare dei trasformatori DC-DC a condensatori commutati. Essi servono in sostanza per cambiare i livelli di tensione, in valore, in segno, o entrambe le cose. Nei circuiti integrati infatti sono presenti poche fonti di alimentazione, ma possono servire molti livelli di tensione, senza che essi vengano per forza introdotti dall'esterno (sia più alti, sia più bassi); mediante questi trasformatori è possibile realizzare ciò.

L'idea di base di queste tecniche è l'uso dei **flying capacitors**: si tratta di **capacità volanti**, ossia capacità che prendono una tensione, e poi la portano verso il carico, mediante un sistema di interruttori.

L'idea migliore per comprendere queste tecniche, è quella di vedere alcuni esempi pratici di applicazione.

4.2.1 Invertitore di segno

Si vuole realizzare un circuito in grado di ottenere:

$$V_u = -V_e$$

Come si può fare? Una realizzazione circuitale è la seguente:

C_f è il flying capacitor; ai capi del carico si introduce una capacità grossa (che verrà introdotta in uscita a ogni circuito), dal momento che questo è un sistema campionato, dunque, dal momento che la corrente arriva a *pacchetti grossi*, è necessario un serbatoio, un *volano*, in modo da poter vedere, sull'uscita, la media dei pacchetti.

Quando gli interruttori sono in posizione 1, la C_f si carica con V_{in}; nella posizione 2, quindi, in uscita si avrà la stessa tensione, invertita di segno: $-V_{in}$.

Cosa capita in pratica? Beh, si ha che:

$$i_L = \frac{v_u}{R_L}$$

Dove R_L è il carico. Si suppone che la capacità C_T sia grossa abbastanza da poter mantenere costante la tensione di uscita: i pacchetti di carica, rispetto a tutta la carica presente in C_T, hanno un contributo irrisorio sull'uscita. Ricordando che:

$$I = \frac{\Delta Q}{T_C}$$

Si ha:

$$\Delta Q = \frac{v_u}{R_L} T_C$$

Ossia, per ogni tempo T_C si ha un ΔQ uguale alla corrente per questo tempo. Al termine della fase 2, C_f sarà carica secondo una tensione V_u, dunque in sostanza:

$$\Delta Q = (V_{in} - V_u) C_f$$

Questo è il ΔQ ottenuto passando da V_{in} a V_u.
Ora: ponendo uguali le due ΔQ, avrò che:

$$\frac{V_u}{R_L} T_C = (V_{in} - V_u) C_f$$

Da qua

$$V_u \left(\frac{T_C}{C_f} - C_f \right) = -V_{in} C_f$$

Dunque:

$$V_u = -V_{in} \frac{C_f}{\frac{T_C}{R_L} + C_f} = -V_{in} \frac{1}{1 + \frac{T_C}{C_f R_L}}$$

Dunque, tutta questa roba equivale ad avere un partitore di tensione della V_{in}, cambiata di segno, tra una R_{eq}, modellante le imperfezioni nel rendimento del circuito a condensatori commutati, e la R_L di carico: questa espressione, infatti, si può scrivere come:

$$V_u = -V_{in} \frac{R_K}{R_L + \frac{T_C}{C_f}} = -V_{in} \frac{R_L}{R_L + R_{eq}}$$

4.2.2 Dimezzatore di tensione

Volendo realizzare una tensione pari a metà di una certa tensione:

$$V_u = \frac{1}{2}V_e$$

Si può realizzare qualcosa del genere:

Ciò che si fa è mettere alternativamente in parallelo all'uscita uno dei due C_a: la tensione in uscita sarà pari a $\frac{V_e}{2}$ grazie al partitore capacitivo, e poi, quando C_f è in posizione 1, si ha un parallelo che farà perdere un minimo di corrente; v_u sarà più bassa di $\frac{v_e}{2}$, dunque C_f tende a caricarsi. Allo stesso modo il colpo dopo si ha un C_f più carico, che si scarica, dal momento che la tensione in uscita sarà un poco più alta di metà di v_e. Volendo fare il conto:

$$i_u = \frac{V_u}{R_L}$$

Da qua

$$\Delta Q = \frac{V_u}{R_L}T_C$$

A regime il ΔQ deve essere preso da C_f:

$$\Delta Q = \left(\frac{V_e}{2} - V_u\right)C_f$$

Dovendo essere uguali i ΔQ, si ha:

$$\left(\frac{V_e}{2} - V_u\right)C - f = \frac{V_u}{R_L}T_C$$

Dunque

$$V_u\left(\frac{T_C}{R_L} + C_f\right) = \frac{V_e}{2}C_f$$

Quindi:

$$\frac{V_u}{V_e} = \frac{1}{2}\frac{C_f}{C_f + \frac{T_C}{R_L}}$$

La situazione è analoga alla precedente: si ha il solito rapporto di ripartizione tra la R_L e la R_{eq}.

4.2.3 Duplicatore di tensione

A capacità commutate si può realizzare un circuito in grado di duplicare la tensione rispetto a quella di ingresso, ossia:

$$V_u = 2V_e$$

Come si può realizzare? Beh, sostanzialmente, si farà qualcosa di questo genere:

L'idea sostanzialmente è: in una fase faccio caricare la capacità volante C_f a V_e, e la fase dopo configuro gli switch in modo da metterla **in serie** al generatore di tensione, in modo da far vedere, al carico, una tensione pari a $2V_e$.

Come esiste il modo per duplicare la tensione, esiste naturalmente anche il modo per triplicare la tensione: sostanzialmente, è sufficiente ripetere (in maniera intelligente) lo schema di prima, facendo in modo che in una fase ci siano, anzichè un singolo condensatore, due condensatori, che si caricano a V_e; nella fase 2 questi verranno messi entrambi in serie al generatore di ingresso, sommando due volte la tensione ai propri capi, ottenendo:

$$V_u = 3V_e$$

4.2.4 Alcune non-idealità delle pompe di carica

Abbiamo visto, per tutti i casi, che il sistema si comporta come un circuito di questo genere:

Dunque, rispetto alla condizione *ideale*, si ha questo partitore, su questa resistenza equivalente R_{eq}, che ci da notevolmente fastidio. Per tutti i casi, si è visto che R_{eq} è:

$$R_{eq} = \frac{T_C}{C_f}$$

Questa formula ci suggerirebbe che dobbiamo avere la f_C (frequenza di clock) più alta possibile, e C_f molto grande. Purtroppo, queste due cose non vanno molto d'accordo: se C_f è grande e T_C piccolo, C_f ha poco tempo per caricarsi alla tensione V_e, dunque non riuscirà a caricarsi a sufficienza e la pompa di carica non farà il proprio mestiere. Si ricordi, infatti, che C_f si carica secondo una costante di tempo, dal momento che gli interruttori hanno una resistenza serie R_{on}:

$$\tau = R_{on}C_f$$

A noi serve che:

$$\tau \ll \frac{T_C}{2}$$

Chiediamo che sia minore di $\frac{T_C}{2}$ dal momento che il ciclo di ricarica è pari a sola metà del clock: per metà clock si ha una fase (ricarica) per l'altra l'altra fase (scarica). L'errore da chiedersi chiaramente non sarà, come nei casi precedenti, dell'un per mille, però non si deve sottovalutare la cosa:

$$R_{on}C_f \ll T_C \longrightarrow R_{on} \ll \frac{T_C}{C_f} = R_{eq}$$

Dunque, si deve pretendere che:

$$R_{on} \ll R_{eq}$$

Con la tecnologia, si può dire quale sia la R_{on} dei MOS, e ciò ci permette di dimensionare una resistenza equivalente idonea.

Si ricordi che C_T, ossia il **tank capacitor**, deve essere tanto grande da non far calare la I_u anche quando esso non è collegato; volendo dimensionarlo, si può usare come parametro un Δv_u accettabile, ossia una massima differenza di tensione accettabile prima di aver una scarica esagerata.

$$\Delta v_u = \frac{I_u}{C_T}\frac{T_C}{2}$$

Ossia, si chiede che C_T sia dimensionata in funzione di T_C, della corrente di uscita, e della caduta di tensione accettabile.

4.2.5 Regolatori di tensione

Abbiamo finora visto come realizzare dei trasformatori DC-DC, ma non dei regolatori. La differenza sta sostanzialmente nel fatto che i regolatori sono in grado di permettere la **regolazione**, ossia in grado di ottenere tensioni precise in uscita. Ad esempio, volendo avere in uscita 2,2 V, per quanto ne sappiamo ora, non si può fare.

L'idea potrebbe essere: abbiamo detto che abbiamo un fattore che ci infastidisce, R_{eq}, dato dalla non idealità del circuito. Ciò che si potrebbe fare è introdurre un controllo su R_{eq}, in modo da farla aumentare o ridurre a seconda dell'uscita che si vuole ottenere. Ciò che posso fare dunque è, mediante uno dei circuiti già visti (duplicatore, triplicatore, o parenti) avere una tensione più alta di quella che ci serve, poi confrontare la tensione in uscita dalla pompa di carica con quella che vogliamo ottenere, ottenuta mediante un diodo Zener o mediante un circuito band-gap. Perchè non usare direttamente uno di questi circuiti? Beh, sostanzialmente, per questioni energetiche: una pompa di carica è in grado di fornire una potenza maggiore rispetto a circuiti come quelli prima citati, per quanto, per soluzioni fuori da circuiti integrati, ci siano idee ancora migliori per il pilotaggio di carichi di vario tipo.

Una volta fornita questa tensione, più alta del dovuto, la si confronta con un comparatore di soglia con isteresi, che controlla il clock: per un certo segnale in uscita dal comparatore il clock sarà bloccato, bloccando tutte le funzioni della pompa di carica, facendo scaricare il C_T e facendo abbassare la tensione, mentre per l'altro segnale tutto il circuito funzionerà regolarmente, facendo aumentare la tensione di uscita, fino a quando non si supereranno le soglie di nuovo, continuando con questo ciclo. L'ampiezza delle soglie sostanzialmente coincide con il ripple della tensione di uscita.

Ciò di fatto coincide con l'avere una R_{eq} variabile: regolando R_{eq} regolo la tensione di uscita.

Un'idea per la realizzazione del sistema può esser la seguente:

Capitolo 5

La distorsione

In questa trattazione si parlerà soprattutto di distorsione dovuta ad effetti di non linearità; per poter però dare un'idea di ciò che capita in ambito di distorsione lineare, se ne daranno alcuni cenni, in modo da poter effettuare un confronto tra le due cose.

5.1 Cenni sulla distorsione lineare

Si supponga di aver a che fare con un sistema LTI (Lineare Tempo-Invariante): esso può essere caratterizzato mediante una funzione di trasferimento $H(f)$, per cui vale il principio di sovrapposizione degli effetti.

Dato un ingresso $X(f)$ e un'uscita $Y(f)$, la funzione di trasferimento costituisce il seguente legame:

$$Y(f) = X(f)H(f)$$

Questa è una proprietà dei sistemi LTI.

Ora, se si ragiona sul modulo di $X(f)$, esso può essere sviluppato in serie di Fourier: esso avrà una fondamentale e una serie di armoniche, infinite. Il modulo dell'uscita avrà questo comportamento:

$$|Y(f_0)| = |X(f_0)|\,|H(f_0)|$$

Questo, per ciascuna armonica presente.

Il contenuto armonico dell'uscita è esattamente uguale a quello dell'ingresso: il *numero* delle armoniche è lo stesso; le ampiezze delle armoniche sono però modificate, a causa del fenomeno della **distorsione lineare**: essa agisce sul modulo delle armoniche, sull'ampiezza, senza però variare il numero delle armoniche presenti. Ciò capita ad esempio nei filtri.

5.2 Distorsione non lineare

La distorsione non lineare è sostanzialmente dovuta al fatto che i componenti con cui si ha a che fare (nella fattispecie quelli elettronici, come i transistori) sono non lineari: essi hanno caratteristiche non lineari.

Nella parte precedente della trattazione non abbiamo mai considerato questi effetti di non linearità dal momento che ci siamo messi in condizione di **piccolo segnale**:

Il punto Q rappresenta il **punto di funzionamento a riposo**, ossia **in assenza di segnale**. Se il segnale è sufficientemente piccolo, è possibile approssimare il suo andamento d'uscita con una retta tangente in Q, e usare questo per calcolare la tensione di uscita.

Se il segnale fosse più grande, si avrebbe, dall'usare la retta all'usare la caratteristica effettiva, un errore molto importante: questa è una distorsione. Vediamo quali sono le caratteristiche dei circuiti non lineari.

Per i circuiti non lineari, **non** vale il principio di sovrapposizione degli effetti: dato un sistema non lineare a due ingressi:

$$f(x, y) \neq f(x, 0) + f(0, y)$$

Inoltre:

$$f(Ax) \neq Af(x)$$

Si ha di peggio: il concetto di funzione di trasferimento, per come lo intendiamo, **non esiste**.

Il fenomeno più interessante, per questa trattazione, è il fatto che si **generano delle frequenze**: se in ingresso ho un certo numero di armoniche, in uscita ne ho un numero diverso; se in ingresso dunque introduco un segnale monocromatico a frequenza f_0, in uscita potrò avere un offset sulla continua (non è certo), e tutte o parte delle armoniche a frequenza multipla della fondamentale: $2f_0$, $3f_0$, e così via.

Si supponga di avere un sistema non lineare, con ingresso E e uscita U:

$$U = f(E)$$

Dove f è una funzione non lineare tra ingresso e uscita. Si potrebbe avere qualcosa di questo tipo:

Dato dunque un segnale di ingresso E monocromatico del tipo:

$$E = E_q + \hat{E}\cos(\omega t)$$

Dove $\hat{E}$ è sufficientemente grande da provocare fenomeni di non linearità, si avrà ciò: sviluppando f in serie di Taylor attorno a Q, si ottiene:

$$f(E) \sim f(E_q) + (E - E_q) \left.\frac{\mathrm{d}U}{\mathrm{d}E}\right|_{E=E_q} + \frac{(E-E_q)^2}{2!} \left.\frac{\mathrm{d}^2U}{\mathrm{d}E^2}\right|_{E=E_q} + ...$$

Ciascuna derivata si definisce come a_i, dove i è l'ordine di derivazione, ottenendo un'espressione del tipo:

$$U - U_q = a_1(E - E_q) + a_2(E - E_q)^2 + a_3(E - E_q)^3 + ...$$

Ma si osservi che:

$$E - E_q = \hat{E}\cos(\omega t)$$

Dunque:

$$U - U_q = \hat{E}a_1\cos(\omega t) + a_2\hat{E}^2\cos^2(\omega t) + ...$$

Si ricordano a questo punto alcune delle fondamentali formule goniometriche, utili per i calcoli di questo tipo:

$$\cos^2(\omega t) = \frac{1}{2} + \frac{1}{2}\cos(2\omega t)$$

$$\cos^3(\omega t) = \frac{3}{4}\cos(\omega t) + \frac{1}{4}\cos(3\omega t)$$

$$\cos^4(\omega t) = \frac{3}{8}\cos(2\omega t) + \frac{1}{8}\cos(4\omega t)$$

Utilizzandole, si può sostituire, e ottenere:

$$U - U_q = a_1\hat{E}\cos(\omega t) + a_2\hat{E}^2\left(\frac{1}{2} + \frac{1}{2}\cos(2\omega t)\right) + a_3\hat{E}^3\left(\frac{3}{4}\cos(\omega t) + \frac{1}{4}\cos(3\omega t)\right) +$$

$$+ a_4\hat{E}^4\left(\frac{3}{8} + \frac{1}{2}\cos(\omega t) + \frac{1}{8}\cos(4\omega t)\right)$$

Ordiniamo dunque i termini secondo la loro pulsazione:

$$U - U_q = \left[\frac{1}{2}a_2\hat{E}^2 + \frac{3}{8}a_4\hat{E}^4 + ...\right] + \cos(\omega t)\left[a_1\hat{E} + a_3\hat{E}^3\frac{3}{4} + ...\right] +$$

$$+\cos(2\omega t)\left[\frac{1}{2}a_2\hat{E}^2+\frac{1}{2}a_4\hat{E}^4\right]+\cos(3\omega t)\left[a_3\hat{E}^3\frac{1}{4}+...\right]+\cos(4\omega t)\left[a_4\hat{E}^4\frac{1}{8}+...\right]$$

Cosa significa ciò? Beh, data una singola sinusoide, in uscita abbiamo trovato una continua e un certo numero di armoniche. Il termine continuo, in altre parole, rappresenta uno scostamento del punto di lavoro.

Si raggruppa ciascuna parentesi in variabili A_i:

$$U - U_0 = A_0 + A_1\cos(\omega t) + A_2\cos(2\omega t) + ...$$

Alcune considerazioni:

- i termini del tipo $\hat{E}^{pari}$ generano le ampiezze delle armoniche **pari**; per contro, quelli dispari generano armoniche dispari;
- Se la funzione non lineare è dispari, ci saranno solo armoniche dispari; in alternativa, se la funzione è pari, avremo solo multipli pari; questo, per le proprietà dello sviluppo di Taylor.

Si consideri a questo punto una piccola variante: un sistema **debolmente lineare**. Si tratta di un sistema non lineare, ma **non troppo**! Ciò si può dire se tutte le potenze più alte sono trascurabili rispetto alle altre. In queste condizioni, per esempio:

$$A_0 = \frac{1}{2}a_2\hat{E}^2$$

$$A_1 = a_1\hat{E}$$

$$A_2 = \frac{1}{2}a_2\hat{E}^2$$

$$A_3 = \frac{1}{4}a_3\hat{E}^3$$

In queste condizioni si possono definire i coefficienti di distorsione armonica: essi sono una misura di **quanto il sistema è non lineare**!

$$D_2 = \frac{A_2}{A_1} = \frac{\frac{1}{2}a_2\hat{E}^2}{a_1\hat{E}} = \frac{1}{2}\frac{a_2}{a_1}\hat{E}$$

$$D_3 = \frac{A_3}{A_1} = \frac{1}{4}\frac{a_3}{a_4}\hat{E}^2$$

$$THD = \sqrt{D_2^2 + D_3^2 + ...}$$

Il THD è sostanzialmente un parametro in grado di quantificare la qualità (la non linearità) del circuito.

5.2.1 Distorsione di intermodulazione

Le armoniche secondarie costituiscono di sicuro un problema, ma fino a un certo punto: esse, infatti, potrebbero essere filtrate, se nel circuito è presente un filtro di qualche tipo!

Esiste in realtà un altro tipo di distorsione, molto più fastidiosa, in cui non è possibile effettuare filtraggi di alcun tipo.

Si consideri, in ingresso a un circuito, una somma di due segnali monocromatici, del tipo:

$$E - E_q = \hat{E}_a \cos(\omega_a t) + \hat{E}_b \cos(\omega_b t)$$

In uscita, si avrà qualcosa del tipo:

$$U - U_q = a_1(E - E_q) + a_2(E - E_q)^2 + ...$$

Ora, il contenuto delle parentesi deve essere sostituito con la somma dei due seni; per dare un'idea:

$$(E - E_q)^2 = (\hat{E}_a \cos^2(\omega_a t) + 2\hat{E}_a \hat{E}_b \cos(\omega_a t)\cos(\omega_b t) + \hat{E}_b^2 \cos^2(\omega_b t))$$

e questo solo con un termine del secondo ordine!

Se il sistema fosse lineare, dovrei ottenere solo ω_a e ω_b. Ciò che ottengo ora, in pratica, è un insieme di contributi:

$$\omega_a - \omega_b \quad 2\omega_a - \omega_b \quad 2\omega_b - \omega_a \quad \omega_a + \omega_b \quad 2\omega_a + \omega_b$$

Le più fastidiose di fatto sono quelle vicino, o prima di f_a e f_b: esse sono, a tutti gli effetti, impossibili da filtrare, o perchè troppo vicine, o perchè addirittura all'interno della banda passante.

Questi contributi di intermodulazione si possono identificare con il proprio ordine di prodotto: sommando i coefficienti che moltiplicano le singole ω (esempio, $2\omega_a - \omega_b$ ha ordine 3), si può quantificare il contributo, e il fastidio che provocano (quelle di ordine 3 sono particolarmente fastidiose).

Se $|T| \sim 1$, posso trascurare e dire:

$$|A_f| \sim \frac{|A_\infty|}{\sqrt{2 + 2\cos(\varphi)}}$$

Sapendo che il margine di fase ψ si definisce come:

$$\psi = \pi - \varphi$$

usando la goniometria, si può dire che:

$$\cos(\varphi) = \cos(\pi - \psi) = -cos(\psi)$$

Da qua:

$$|A_f| = \frac{|A_\infty|}{\sqrt{2 - 2\cos(\psi)}} = \frac{|A_\infty|}{2\sin\left(\frac{\psi}{2}\right)}$$