Appunti di Comunicazioni Elettriche
Capitolo 3
Modulazione angolare

Introduzione

Abbiamo visto in precedenza che la *modulazione di ampiezza* di una portante sinusoidale $c(t) = A_C \cos(2\pi f_C t + \varphi(t))$, ad opera di un generico segnale (analogico) modulante s(t), consiste fondamentalmente nel trasmettere tale portante facendone però variare, in ciascun istante, l'ampiezza in modo proporzionale al valore assunto dal segnale s(t) in quell'istante: il **segnale modulato** risulta perciò essere del tipo

$$\underbrace{s_t(t) = A_C s(t) \cos(2\pi f_C t + \varphi(t))}_{\text{modulazione AM}}$$

La **modulazione di fase** (**PM**, *Phase Modulation*) si effettua sempre trasmettendo la portante, ma facendone variare, istante per istante ed in modo lineare, la fase. Il segnale modulato di fase è dunque nella forma seguente:

$$\boxed{\underbrace{s_t(t) = A_C \cos(2\pi f_C t + K_P s(t))}_{\text{modulazione PM}}}$$

Infine, per quanto riguarda la **modulazione di frequenza** (**FM**, *Frequency Modulation*), il concetto è ovviamente quello di far variare, in ciascun istante, la frequenza della portante in modo proporzionale al valore assunto in quell'istante dal segnale da trasmettere s(t). Tuttavia, in questo caso l'espressione del segnale modulato è un po' più complicata da ottenere, in quanto non è SOLO la frequenza che subisce variazioni. Vediamo allora i relativi dettagli matematici.

Consideriamo intanto il generico segnale portante sinusoidale:

$$c(t) = A_C \cos(2\pi f_C t + \varphi(t))$$

Indichiamo con $\theta(t) = 2\pi f_C t + \varphi(t)$ l'argomento del coseno.

Si definisce **pulsazione istantanea** di c(t) la variazione temporale di θ(t), ossia la sua derivata:

$$\omega_i(t) = \frac{d\theta(t)}{dt}$$

Dividendo per 2π, si ottiene chiaramente la **frequenza istantanea**:

$$\boxed{f_i(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt}}$$

La modulazione di frequenza consiste allora nel far variare, istante per istante, la frequenza istantanea della portante in modo proporzionale al valore assunto, in ciascun istante, dal segnale s(t). Si fa cioè in modo che la frequenza istantanea risulti legata al segnale modulante dalla seguente relazione:

$$f_i(t) = f_C + k_F s(t)$$

dove f_C è la frequenza della portante e k_F una costante da scegliere in modo opportuno.

In pratica, quindi, l'effetto del segnale modulante s(t) è quello di far variare la frequenza istantanea della portante rispetto al valore f_C, costante nel tempo, che aveva prima della modulazione. A questo proposito, si definisce **deviazione di frequenza** (**$\Delta f(t)$**)la variazione della frequenza istantanea della portante rispetto al valore f_C:

$$\Delta f(t) = f_i(t) - f_C = k_F s(t)$$

E' evidente che questa deviazione di frequenza dipende da s(t), per cui è funzione del tempo; essa raggiungerà perciò un valore massimo negativo quando s(t) raggiunge il suo massimo negativo ed un valore massimo positivo quando s(t) è al suo valore massimo positivo. La differenza tra questi valori prende il nome di **deviazione di frequenza picco-picco**:

$$\Delta f_{PP} = (\Delta f)_{\substack{\max \\ posit}} - (\Delta f)_{\substack{\max \\ neg}} = k_F \left[(s(t))_{\substack{\max \\ posit}} - (s(t))_{\substack{\max \\ neg}} \right]$$

Un caso particolare si ottiene quando il segnale s(t) ha escursione simmetrica rispetto al proprio valore medio: ad esempio, se il segnale modulante è $s(t) = \cos(\omega_S t)$, allora il valore medio è zero e i valori estremi sono +1 e -1. In casi come questo, la quantità Δf_{PP} è pari al doppio della cosiddetta **deviazione di frequenza di picco** (simbolo: **Δf_P**), intesa come la massima escursione di Δf in positivo o in negativo:

$$\Delta f_{PP} = 2(\Delta f)_{\max} = 2\Delta f_P = 2k_F (s(t))_{\max}$$

Nel caso particolare in cui s(t) è sinusoidale con valor medio nullo e ampiezza unitaria, allora $(s(t))_{\max} = 1$, per cui $\Delta f_{PP} = 2\Delta f_P = 2k_F$.

Se, invece, il segnale modulante è $s(t) = A_S \cos(\omega_S t)$, cioè ancora a valor medio nullo ma con ampiezza A_S non più unitaria, allora risulta evidentemente $\Delta f_{PP} = 2\Delta f_P = 2k_F A_S$. Di questo risultato ci serviremo in seguito.

Torniamo adesso ai passaggi di prima e, in particolare, all'espressione del segnale modulato FM. Abbiamo visto che ad ogni valore di s(t) corrisponde un preciso valore di $f_i(t)$ e quindi anche un preciso valore dell'argomento $\theta(t)$ del coseno della portante. Vediamo allora quale relazione intercorre tra s(t) e $\theta(t)$: in base a come abbiamo definito $f_i(t)$ e in base all'ultima relazione scritta, è evidente che

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt} = f_C + k_F s(t)$$

Questa è una equazione differenziale nella incognita $\theta(t)$: risolvendo si ottiene evidentemente

$$\theta(t) = 2\pi f_C t + 2\pi k_F \int_{-\infty}^{t} s(\tau) d\tau$$

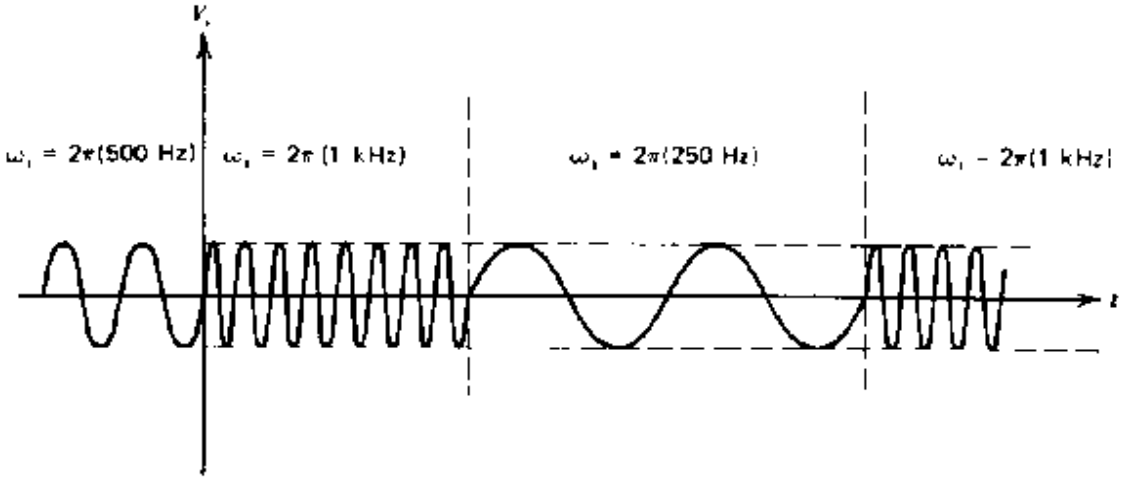
Questo è dunque l'argomento del segnale modulato, il quale quindi risulta avere la seguente espressione:

Autore: Sandro Petrizzelli

$$\underbrace{s_t(t) = A_C \cos\left(2\pi f_C t + 2\pi k_F \int_{-\infty}^{t} s(\tau)d\tau\right)}_{\text{modulazione FM}}$$

La cosa importante da sottolineare è la seguente: l'argomento del coseno in quest'ultima relazione dipende evidentemente da come è fatto il segnale s(t); è possibile che questo segnale abbia una struttura tale che, una volta effettuata l'integrazione, vengano modificate sia la fase sia la frequenza della portante. Questo per dire che *la modulazione di frequenza NON comporta necessariamente solo la variazione della frequenza della portante, ma può comportare anche la variazione della fase*. La cosa sarà più chiara dagli esempi seguenti.

Ad ogni modo, riportiamo, nella figura seguente, l'andamento di un tipico segnale sinusoidale modulato in frequenza:

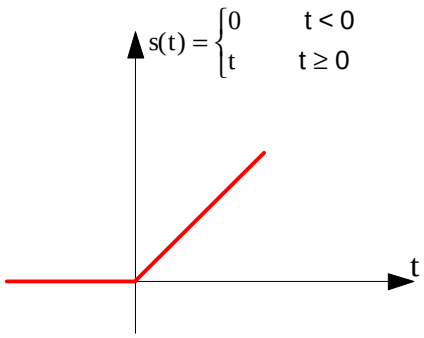


Si tratta di una sinusoide alla quale viene fatta variare, ad intervalli di tempo successivi, la frequenza: da un valore iniziale di 500 Hz, si passa ad 1kHz, poi a 250Hz ed infine ancora ad 1 kHz.

Questa figura evidenzia come la modulazione di frequenza sia, fondamentalmente, una trasformazione non lineare che, pur mantenendo costante l'inviluppo della portante, produce un segnale modulato il cui spettro ha una banda in generale maggiore del valore corrispondente alla modulazione di ampiezza. Come vedremo, a questa espansione di banda corrisponde un miglioramento della protezione contro i disturbi presenti sul canale trasmissivo.

Esempio: modulazione FM e PM della rampa

Supponiamo che il segnale analogico da trasmettere sia la rampa:



Usando la classica portante cosinusoidale, le corrispondenti espressioni generali dei segnali modulati, rispettivamente, in frequenza e in fase sono le seguenti:

$$\underbrace{s_t(t) = A_C \cos\left(2\pi f_C t + K_P s(t)\right)}_{\text{modulazione PM}}$$

$$\underbrace{s_t(t) = A_C \cos\left(2\pi f_C t + 2\pi k_F \int_{-\infty}^{t} s(\tau)d\tau\right)}_{\text{modulazione FM}}$$

E' evidente che, per t<0, in entrambi i casi il segnale modulato corrisponde alla portante in quanto s(t) vale zero. Vediamo perciò cosa succede per t≥0.

Per la modulazione di fase abbiamo quanto segue:

$$s_t(t) = A_C \cos\left(2\pi f_C t + K_P t\right) = A_C \cos\left((2\pi f_C + K_P)t\right) = A_C \cos\left(2\pi\left(f_C + \frac{K_P}{2\pi}\right)t\right)$$

Ciò che si ottiene non è dunque una variazione istantanea della fase rispetto a quella della portante, che è sempre nulla, bensì una variazione istantanea della frequenza, che risulta essere $f_C + \frac{K_P}{2\pi}$.

La cosa interessante è che, <u>in questo caso, la variazione della frequenza è costante nel tempo</u>, nel senso che la frequenza assume sempre lo stesso valore in ciascun istante: questo fatto è legato alle caratteristiche del segnale modulante a rampa .

Vediamo invece cosa succede modulando in frequenza:

$$s_t(t) = A_C \cos\left(2\pi f_C t + 2\pi k_F \int_0^t \tau d\tau\right) = A_C \cos\left(2\pi f_C t + \pi k_F t^2\right) = A_C \cos\left(2\pi\left(f_C + k_F \frac{t}{2}\right)t\right)$$

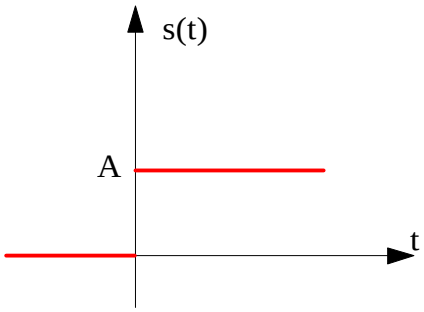
Anche in questo caso, non ci sono variazioni di fase, mentre varia la frequenza della portante, che risulta essere $f_C + k_F \frac{t}{2}$.

Autore: Sandro Petrizzelli

La differenza con la modulazione di fase è però nel fatto che la frequenza varia istante per istante: essa aumenta linearmente all’aumentare del tempo.

Esempio: modulazione FM e PM del gradino

Supponiamo adesso che il segnale analogico da trasmettere sia un gradino di altezza A:



Partiamo sempre dalle espressioni generali dei segnali modulati, rispettivamente, in frequenza e in fase:

$$\underbrace{s_t(t) = A_C \cos\left(2\pi f_C t + K_P s(t)\right)}_{\text{modulazione PM}}$$

$$\underbrace{s_t(t) = A_C \cos\left(2\pi f_C t + 2\pi k_F \int_{-\infty}^{t} s(\tau)d\tau\right)}_{\text{modulazione FM}}$$

Anche in questo caso, per $t<0$, il segnale modulato coincide con la portante. Vediamo invece per $t\geq 0$.

Modulando in fase, otteniamo

$$s_t(t) = A_C \cos\left(2\pi f_C t + K_P A\right)$$

per cui, questa volta, la frequenza rimane quella della portante, mentre c’è uno sfasamento rispetto alla portante pari a $K_P A$. Tale sfasamento è costante nel tempo.

Modulando invece in frequenza, abbiamo quanto segue:

$$s_t(t) = A_C \cos\left(2\pi f_C t + 2\pi k_F \int_0^t A d\tau\right) = A_C \cos\left(2\pi f_C t + 2\pi A k_F t\right) = A_C \cos\left(2\pi\left(f_C + A k_F\right)t\right)$$

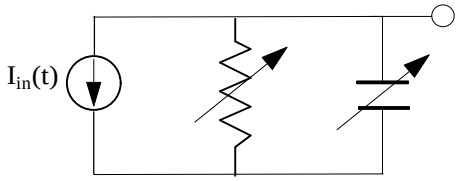
Si osserva che non ci sono variazioni di fase, mentre cambia la frequenza, che diventa $f_C + Ak_F$. Anche in questo caso, la nuova frequenza rimane costante nel tempo.

Si osserva, in particolare, che la frequenza risulta aumentata, così come accadeva anche nel caso della rampa: questo è un risultato generale, nel senso che *la frequenza della portante, nella modulazione FM, aumenta quando s(t) è positivo mentre diminuisce quando s(t) è negativo.*

Semplice circuito per la modulazione di fase

Anche se ne parleremo ampiamente più avanti, possiamo facilmente renderci conto di come si possa ottenere una modulazione di fase di una portante sinusoidale.

Consideriamo infatti il circuito seguente:



Il generatore di corrente fornisce una corrente sinusoidale $I_{in}(t) = I_C \cos(\omega_C t)$ che alimenta il parallelo tra una resistenza ed un condensatore. In termini sistemistici, abbiamo dunque il segnale sinusoidale che entra in ingresso ad un sistema lineare tempo-invariante, rappresentabile perciò con una funzione di trasferimento H(ω): il modulo e la fase di tale funzione di trasferimento dipendono dai parametri R e C degli elementi resistivi e reattivi che costituiscono il sistema.

L'uscita del sistema, a regime, è un segnale ancora sinusoidale e isofrequenziale con l'ingresso, data la linearità, il cui modulo e la cui fase dipendono dal modulo e dalla fase di H(ω), oltre che dal modulo e dalla fase dell'ingresso: se y(t) è l'uscita (in questo caso la tensione $V_R(t)=RI_R(t)$), essa ha dunque espressione

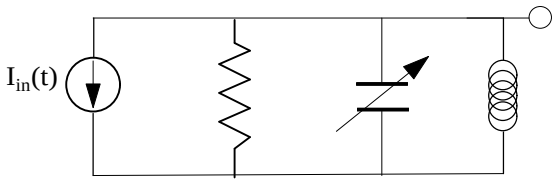
$$y(t) = Y(\omega_C)\sin(\omega_C t + \varphi(\omega_C))$$

dove

$$\varphi(\omega_C) = \langle H(\omega_C)$$
$$Y(\omega_C) = I_{in}\left|H(\omega_C)\right|$$

Facendo dunque variare R o C o entrambi in base al segnale modulante s(t), otteniamo un segnale modulato la cui fase φ(ω) varia proporzionalmente ad s(t). Gli elementi R e C dovranno dunque essere una resistenza ed una capacità variabili entrambi elettronicamente (in modo appunto che le variazioni siano attuate dal segnale modulante) ed è possibile ottenere questo usando dei diodi la cui polarizzazione avvenga appunto tramite s(t).

Se la frequenza ω_C della portante è particolarmente elevata, è preferibile usare, al posto di quel circuito, un circuito risonante RLC parallelo, in cui l'unico parametro variabile sia la capacità:



Di questi circuito parleremo ampiamente in seguito.

Autore: Sandro Petrizzelli

Modulazione FM

Caso particolare: singolo tono modulante

Ci concentriamo, in questi paragrafi, sulla modulazione di frequenza.

Studiamo un caso particolare, in cui il segnale analogico s(t) modulante è una sinusoide (il cosiddetto **singolo tono modulante**):

$$s(t) = A_S \cos(2\pi f_S t)$$

Vediamo qual è il risultato della modulazione di frequenza di tale segnale: abbiamo detto che l'espressione generale del segnale modulato FM è

$$s_t(t) = A_C \cos\left(2\pi f_C t + 2\pi k_F \int_{-\infty}^{t} s(\tau)d\tau\right)$$

per cui dobbiamo andare a sostituire l'espressione di s(t) e fare i calcoli.

Al fine di semplificarci tali calcoli (ed in particolare il calcolo dell'integrale), supponiamo che s(t) sia nullo prima di t=0 e sia $s(t) = A_S \cos(2\pi f_S t)$ dopo t=0. In tal modo, il segnale modulato risulta essere il seguente:

$$s_t(t) = A_C \cos\left(2\pi f_C t + 2\pi k_F \int_{-\infty}^{t} s(\tau)d\tau\right) = A_C \cos\left(2\pi f_C t + 2\pi k_F \int_{0}^{t} A_S \cos(2\pi f_S \tau)d\tau\right) =$$

$$= A_C \cos\left(2\pi f_C t + 2\pi k_F A_S \left(\frac{1}{2\pi f_S}\sin(2\pi f_S t)\right)\right) = A_C \cos\left(2\pi f_C t + \frac{k_F A_S}{f_S}\sin(2\pi f_S t)\right)$$

A questo punto, ci ricordiamo di una definizione già data in precedenza: in presenza di segnale modulante $s(t) = A_S \cos(2\pi f_S t)$, abbiamo detto che la quantità $k_F A_S$ rappresenta la **deviazione di frequenza di picco** della portante: indicandola con Δf_P, otteniamo

$$s_t(t) = A_C \cos\left(2\pi f_C t + \frac{\Delta f_P}{f_S}\sin(2\pi f_S t)\right)$$

Si chiama inoltre **indice di modulazione** la quantità

$$m = \frac{\Delta f_P}{f_S}$$

ossia il rapporto tra la deviazione di frequenza di picco e la frequenza del tono modulante. Con questa seconda posizione, l'espressione del segnale modulato diventa

$$\boxed{s_t(t) = A_C \cos(2\pi f_C t + m\sin(2\pi f_S t))}$$

Questa è l'espressione più generale possibile per il segnale modulato FM nel caso di singolo tono modulante.

Spettro e occupazione di banda del segnale modulato

Una caratteristica importante del segnale modulato FM è, come sappiamo, l'occupazione di banda, ossia l'intervallo di frequenze da esso coperto: infatti, è anche in base a questa caratteristica che è necessario progettare i mezzi trasmissivi ed i vari filtri usati negli apparati di modulazione/demodulazione.

Al fine di valutare l'occupazione di banda del segnale $s_t(t)$, ci mettiamo inizialmente in una condizione particolare: supponiamo infatti di adottare un basso indice di modulazione, il che vale a dire **m<<1**. Vediamo cosa succede.

Intanto, a prescindere da quella condizione particolare, possiamo usare le formule di duplicazione del coseno per scrivere che il segnale modulato è

$$s_t(t) = A_C \cos(2\pi f_C t + m\sin(2\pi f_S t)) = A_C \cos(2\pi f_C t)\cos(m\sin(2\pi f_S t)) - A_C \sin(2\pi f_C t)\sin(m\sin(2\pi f_S t))$$

A questo punto, il fatto che **m<<1** ci consente di fare le seguenti due approssimazioni:

$$\cos(m\sin(2\pi f_S t)) \cong 1$$
$$\sin(m\sin(2\pi f_S t)) \cong m\sin(2\pi f_S t)$$

Sulla base di queste approssimazioni, il segnale modulato diventa il seguente:

$$s_t(t) = A_C \cos(2\pi f_C t) - mA_C \sin(2\pi f_C t)\sin(2\pi f_S t)$$

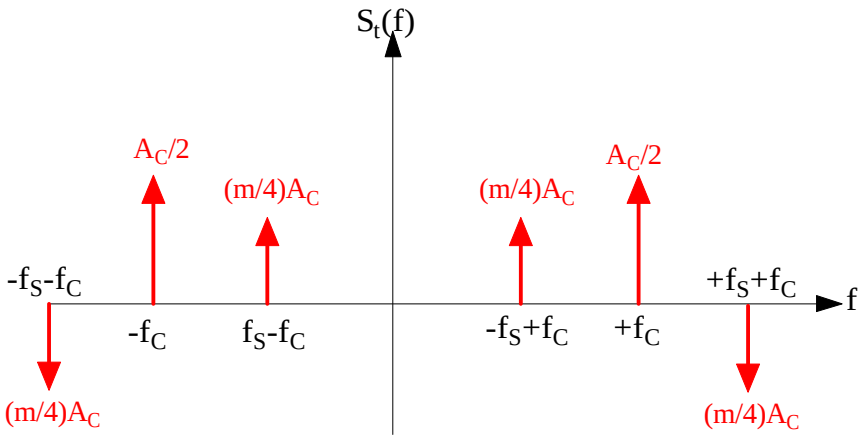
Sfruttando le opportune formule trigonometriche, possiamo inoltre sviluppare il secondo termine a secondo membro, ottenendo quanto segue:

$$s_t(t) = A_C \cos(2\pi f_C t) - \frac{mA_C}{2}\cos(2\pi(f_C + f_S)t) + \frac{mA_C}{2}\cos(2\pi(f_C - f_S)t)$$

Passiamo adesso al dominio della frequenza: lo spettro di questo segnale è il seguente:

$$S_t(f) = A_C\left[\frac{1}{2}\delta(f - f_C) + \frac{1}{2}\delta(f + f_C)\right] - \frac{mA_C}{2}\left[\frac{1}{2}\delta(f - f_C - f_S) + \frac{1}{2}\delta(f + f_C + f_S)\right] + \frac{mA_C}{2}\left[\frac{1}{2}\delta(f - f_C + f_S) + \frac{1}{2}\delta(f + f_C - f_S)\right]$$

Abbiamo dunque 6 impulsi, posizionati in posizioni (ovviamente) simmetriche rispetto all'origine delle frequenze:



Si deduce dunque che i valori non nulli dello spettro si hanno nei due intervalli $[-f_C-f_S,-f_C+f_S]$ e $[-f_S+f_C,f_S+f_C]$, mentre al di fuori di tali intervalli si hanno solo valori nulli. La banda occupata dal segnale è quindi di ampiezza $2f_S$.

Questo fatto è ovviamente strettamente legato all'ipotesi di partenza di un indice di modulazione m molto piccolo. Si può già intuire che, per m generico, la banda del segnale modulato aumenta: il motivo sta nel modo in cui varia, al variare di m, il coseno presente nella espressione

$$s_t(t) = A_C \cos\left(2\pi f_C t + m\sin\left(2\pi f_S t\right)\right)$$

Vediamo ad ogni modo che cosa accade dal punto di vista matematico.
Ricordando, dalla teoria dei numeri complessi, che vale la relazione

$$e^{j\vartheta} = \cos\vartheta + j\sin\vartheta$$

possiamo riscrivere il segnale modulato nel modo seguente:

$$s_t(t) = A_C \,\mathrm{Re}\left\{e^{j2\pi f_C t} e^{jm\sin(2\pi f_S t)}\right\}$$

A questo punto, è evidente che il segnale $e^{jm\sin(2\pi f_S t)}$ è un segnale periodico di periodo $1/f_S$, in quanto è tale il segnale presente all'esponente e la funzione esponenziale è una funzione lineare. Sappiamo allora che un qualsiasi segnale periodico è esprimibile mediante uno sviluppo in serie di Fourier: se poniamo

$$\tilde{s}(t) = e^{jm\sin(2\pi f_S t)}$$

il suo sviluppo in serie sarà

$$\tilde{s}(t) = \sum_{k=-\infty}^{+\infty} c_K e^{j2\pi k f_S t}$$

Per determinare questo sviluppo, dobbiamo evidentemente calcolarne i coefficienti: usando la normale definizione, abbiamo che

$$c_K = \frac{1}{T}\int_{-T/2}^{+T/2} \tilde{s}(t) e^{-j2\pi k f_S t} dt = \frac{1}{T}\int_{-T/2}^{+T/2} e^{jm\sin(2\pi f_S t)} e^{-j2\pi k f_S t} dt = \frac{1}{T}\int_{-T/2}^{+T/2} e^{jm\sin(2\pi f_S t) - j2\pi k f_S t} dt = f_S \int_{-1/2f_S}^{+1/2f_S} e^{jm\sin(2\pi f_S t) - j2\pi k f_S t} dt$$

dove $T=1/f_S$. Facendo inoltre il cambio di variabile $x=2\pi f_S t$, abbiamo che

$$c_K = f_S \int_{-\pi}^{+\pi} e^{jm\sin x - jkx} \frac{1}{2\pi f_S} dx = \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{jm\sin x - jkx} dx$$

A questo punto, la risoluzione di quell'integrale è possibile ma è tutt'altro che agevole. Si verifica comunque che il risultato è un numero reale per ogni k: poniamo allora

$$c_K = \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{jm\sin x - jkx} dx = J_{K,m}$$

per cui il nostro sviluppo in serie risulta essere

$$\tilde{s}(t) = \sum_{k=-\infty}^{+\infty} J_{K,m} e^{j2\pi k f_S t}$$

Tornando all'espressione del segnale modulato $s_t(t)$, esso diventa

$$s_t(t) = A_C \, \mathrm{Re}\left\{ e^{j2\pi f_C t} \sum_{k=-\infty}^{+\infty} J_{K,m} e^{j2\pi k f_S t} \right\}$$

Il termine esponenziale $e^{j2\pi f_C t}$ non dipende chiaramente da k, per cui lo portiamo dentro la sommatoria; inoltre, la parte reale è un operatore lineare, per cui possiamo portare fuori dalle parentesi graffe i termini costanti: quindi, il segnale modulato assume l'espressione

$$s_t(t) = A_C \sum_{k=-\infty}^{+\infty} J_{K,m} \, \mathrm{Re}\left\{ e^{j2\pi f_C t} e^{j2\pi k f_S t} \right\}$$

dove abbiamo portato fuori anche i termini $J_{K,m}$ in quanto abbiamo detto che risultano essere reali.

Applicando infine la relazione $e^{j\vartheta} = \cos\vartheta + j\sin\vartheta$, possiamo concludere che

$$s_t(t) = A_C \sum_{k=-\infty}^{+\infty} J_{K,m} \cos\left(2\pi\left(f_C + k f_S\right)t\right)$$

Questa è dunque l'espressione più generale possibile, nel dominio del tempo, in presenza di un singolo tono modulante FM e di un indice di modulazione generico.

Dato che a noi interessa l'occupazione di banda di questo segnale, effettuiamo la trasformata di Fourier per passare al dominio della frequenza: è immediato verificare che lo spettro di $s_t(t)$ risulta essere

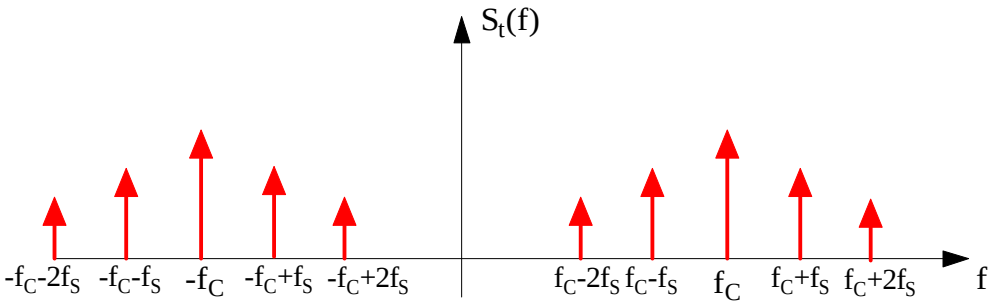
$$\boxed{S_t(f) = A_C \sum_{k=-\infty}^{+\infty} \frac{J_{K,m}}{2} \left[\delta\left(f - \left(f_C + k f_S\right)\right) + \delta\left(f + \left(f_C + k f_S\right)\right)\right]}$$

Autore: Sandro Petrizzelli

Abbiamo dunque una successione di infiniti impulsi: questo significa che il segnale modulato presenta teoricamente banda infinita.

C'è però da osservare un fatto: nel valutare, con appositi metodi numerici, i valori dei coefficienti $J_{K,m}$, si trova che essi assumono valori (reali) sempre più piccoli al crescere del valore assoluto di k. Questo significa che gli impulsi di cui è composto $S_t(f)$ non hanno tutti la stessa area: al contrario, fissato l'indice di modulazione m, essi presentano il massimo valore in corrispondenza di k=0, cioè in corrispondenza di $\pm f_C$, mentre poi presentano valori via via decrescenti all'aumentare di k in valore assoluto.

Possiamo cioè schematizzare la cosa nel modo seguente:

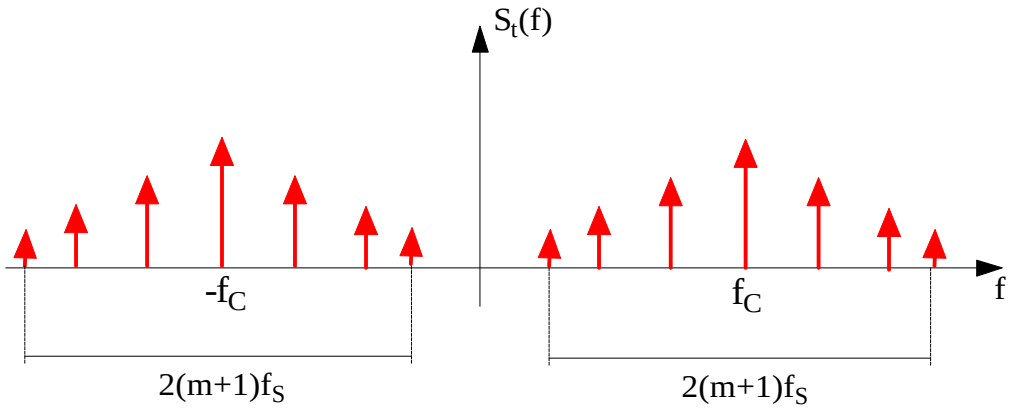


Questo fatto è molto importante in quanto consente di fare il seguente discorso: se l'area degli impulsi va via via decrescendo, possiamo ritenere che esisterà un valore di |k| oltre il quale l'area degli impulsi sarà talmente piccola da poterla ritenere nulla (cioè praticamente da poter ritenere nullo il contenuto energetico ad essa associato). Detto in altri termini, esisterà un valore di |k| oltre il quale noi potremo trascurare il contributo energetico (o di potenza) del segnale modulato. Questo valore, per un indice di modulazione abbastanza minore di 1, è approssimativamente |k|=2.

Possiamo dunque ritenere che il contenuto energetico (o di potenza) del segnale sia concentrato in un certo intorno della frequenza della portante: in particolare, lo standard internazionale considera la cosiddetta **approssimazione di Carson**, secondo la quale l'ampiezza dell'intorno, centrato in f_C, entro il quale è concentrata l'energia del segnale ha ampiezza

$$B_{RF} = 2(m+1)f_S$$

dove ricordiamo ancora che m è l'indice di modulazione e f_S la frequenza del tono modulante che stiamo considerando.



La quantità $\mathbf{B_{RF}}$ (dove il pedice "RF" sta per *RadioFrequency*, ossia radiofrequenza) prende il nome di **banda di Carson**.

Ricordando che l'indice di modulazione è stato definito come $m = \frac{\Delta f_P}{f_S}$, la banda di Carson assume l'espressione

$$B_{RF} = 2\left(\frac{\Delta f_P}{f_S} + 1\right) f_S = 2\left(\Delta f_P + f_S\right)$$

In base a questa espressione, la banda occupata dalla portante modulata di frequenza da un singolo tono è il doppio della somma della frequenza f_S del tono modulante e della deviazione di frequenza di picco Δf_P.

Questa formula si può in realtà generalizzare per **segnali modulati qualsiasi**, affermando che la banda di Carson è

$$B_{RF} = 2\left(\Delta f_P + f_{max}\right)$$

dove f_{max} è la massima frequenza del segnale modulante, mentre Δf_P è sempre la **deviazione di frequenza di picco**.

Questa formula è però relativa al caso in cui la deviazione di frequenza è simmetrica rispetto al valor medio f_C, ossia al caso in cui la frequenza istantanea della portante modulata varia, in positivo e in negativo, della stessa quantità massima Δf_P rispetto al valore f_C. Questo è, per esempio, quello che accade nel caso del singolo tono modulante esaminato poco fa, dove la legge di modulazione è sinusoidale e quindi la frequenza istantanea oscilla tra f_C+2f_S e f_C-2f_S.

Nel caso in cui, invece, la deviazione di frequenza non sia simmetrica rispetto al valor medio f_C, allora la formula si modifica nel modo seguente:

$$\boxed{B_{RF} = \Delta f_{PP} + 2f_{max}}$$

dove f_{max} è ancora la massima frequenza del segnale modulante, mentre Δf_{PP} la **deviazione di frequenza picco-picco**, ossia la differenza tra la massima e la minima frequenza istantanea della portante modulata.

E' ovvio che, se $\Delta f_{PP}=2\Delta f_P$, ricadiamo nella formula precedente, per cui possiamo assumere quest'ultima come espressione generale della banda di Carson.

Ad ogni modo, utilizzando l'approssimazione di Carson, concludiamo che *il segnale modulato FM ha una occupazione di banda limitata, ma comunque maggiore rispetto a quella del segnale modulante di partenza*.

Il fatto che la banda del segnale modulato possa essere considerata come limitata significa, evidentemente, che possiamo tranquillamente usare lo stesso mezzo trasmissivo per trasmettere più di un segnale modulato in FM. Naturalmente, dobbiamo fare in modo che gli spettri dei vari segnali non si sovrappongano: è una cosa facile da fare in quanto, nota la frequenza di ciascun segnale, è sufficiente spaziare in modo opportuno la frequenza della sua portante rispetto alle altre.

Esempio numerico: trasmissione FM del segnale radiofonico

Facciamo subito un esempio concreto a proposito della banda di Carson.

Supponiamo di voler trasmettere, su un mezzo trasmissivo passa-banda, il **segnale radiofonico**: tale segnale è in tutto e per tutto assimilabile al segnale musicale, per cui possiamo assumere che si tratti di un segnale passa-basso di banda f_{max} **= 15 kHz**. Essendo il mezzo trasmissivo di tipo passa-banda, la modulazione si rende necessaria per allocare il segnale nella banda passante del mezzo stesso.

Nella **trasmissione del segnale radiofonico in FM**, ad ogni canale viene riservata una banda passante di **180 kHz**. Ciò significa che la banda del segnale modulato, che sappiamo essere data da $B_{RF} = \Delta f_{PP} + 2f_{max}$ (*banda di Carson*), dovrà essere di 180kHz: quindi abbiamo che

$$B_{RF} = \Delta f_{PP} + 2f_{max} = 180\text{kHz} \longrightarrow \Delta f_{PP} = 180\text{kHz} - 2f_{max} = 180\text{kHz} - 2 \cdot 15\text{kHz} = 150\text{kHz}$$

Abbiamo cioè trovato che la massima deviazione di frequenza Δf_{pp} picco-picco a nostra disposizione è di 150 kHz.

Se supponiamo che il segnale musicale abbia valor medio nullo, possiamo ritenere che la deviazione di frequenza sia simmetrica rispetto al valor medio (pari alla frequenza f_C della portante), il che significa che la deviazione di frequenza di picco vale $\Delta f_P = \Delta f_{PP} / 2 = 75\text{kHz}$.

Adesso modifichiamo leggermente le specifiche del problema, supponendo che la trasmissione sia in forma numerica: ciò significa che, avendo a disposizione il segnale analogico da trasmettere, dobbiamo prima campionarlo, poi quantizzarlo ed infine usarlo per modulare di frequenza la portante. I calcoli sono abbastanza semplici:

- in primo luogo, se il segnale analogico ha banda 15 kHz, per il **campionamento** dobbiamo scegliere, in base al *teorema del campionamento*, una frequenza di campionamento pari almeno a 2*15kHz=30 kHz; nella pratica, se si vuole ad esempio trasmettere il segnale con **qualità CD**, la frequenza di campionamento è fissata a **44.1 kHz**; ciò significa che l'uscita del campionatore genera 44100 campioni al secondo;
- il secondo passo è la **quantizzazione** e dobbiamo perciò scegliere il numero di bit da associare a ciascun campione: non vanno bene 8 bit per campione, in quanto questo è quello che si fa per la trasmissione in forma numerica del segnale telefonico, il quale ha notoriamente una qualità abbastanza scadente; gli standard della *qualità CD* indicano invece un quantizzazione di **16 bit** per campione. Da ciò consegue che il numero di bit in uscita dal quantizzatore nell'unità di tempo è

$$N = 44100\left(\frac{\text{campioni}}{\text{secondo}}\right) \cdot 16\left(\frac{\text{bit}}{\text{campione}}\right) = 705600\left(\frac{\text{bit}}{\text{secondo}}\right)$$

Come si vedrà in seguito nel capitolo sulla trasmissione numerica, il segnale che consente la trasmissione di 705600 bit/sec è un segnale passa-basso che, nell'ipotesi di poter raggiungere i limiti ideali di funzionamento, occupa una banda di N/2 Hz, ossia **352.8 kHz**. Questo è dunque il valore che dovremmo sostituire a f_{max} nel calcolo di B_{RF}:

$$180\ \text{kHz} = B_{RF} = \Delta f_{PP} + 2f_{max} = \Delta f_{PP} + 2 \cdot 352.8\ \text{kHz}$$

E' evidente, allora, che si otterrebbe una deviazione di frequenza Δf_{PP} negativa, il che non ha senso. Questo significa che, con la banda assegnata, non possiamo effettuare una trasmissione numerica del segnale musicale in modulazione di frequenza. In realtà, vedremo che non è proprio così, in quanto le tecniche di modulazione con segnali numerici consentono di ovviare a questo ostacolo, raggiungendo velocità di trasmissione (cioè bit/sec) molto elevate con occupazione di banda relativamente piccola.

Rumore nella tecnica di modulazione FM

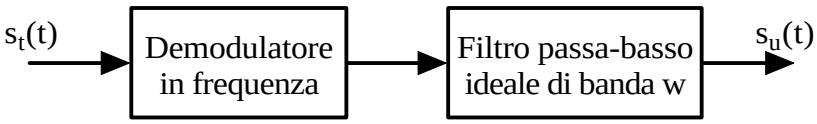
Presenza del rumore

Il dispositivo classico usato, nella demodulazione FM, per la determinazione in ricezione della frequenza istantanea dell'onda modulata è il cosiddetto **discriminatore**. Si tratta di un circuito che compie fondamentalmente due operazioni: in primo luogo, trasforma l'onda modulata in frequenza (ad inviluppo costante) in un'onda il cui inviluppo è proporzionale alla frequenza istantanea; successivamente, preleva tale inviluppo mediante un classico *demodulatore d'inviluppo*.

Il nucleo del demodulatore è dunque costituito da un circuito la cui funzione di trasferimento varia in modulo, nell'intorno della frequenza portante, in modo lineare con la frequenza. Per esempio, si può ricorrere ad un circuito risonante ad una frequenza vicina a quella della portante, usandolo lungo un tratto il più possibile lineare della sua caratteristica.

Normalmente, il discriminatore è preceduto da un **limitatore**. Infatti, l'inviluppo del segnale all'ingresso del demodulatore non sarà in generale costante, ma a causa della presenza del rumore presenterà delle fluttuazioni, ossia sostanzialmente una modulazione spuria di ampiezza: questa deve essere rimossa perché non vada ad aggiungersi alla modulazione prodotta dal segnale utile e ciò si realizza appunto con un circuito di limitazione delle ampiezze.

Senza scendere, comunque, nei dettagli realizzativi, consideriamo un **demodulatore FM**, supposto ideale (quindi non rumoroso):

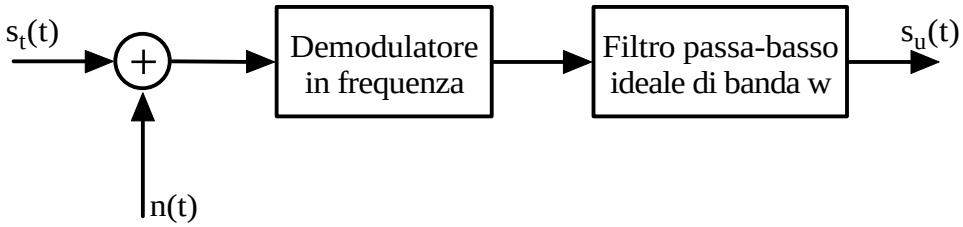


Il segnale modulato, ossia il segnale trasmesso sul canale (supposto anch'esso ideale) e che arriva in ingresso al demodulatore, ha espressione analitica

$$s_t(t) = A_C \cos\left(2\pi f_C t + 2\pi k_F \int_{-\infty}^{t} s(\tau)d\tau\right)$$

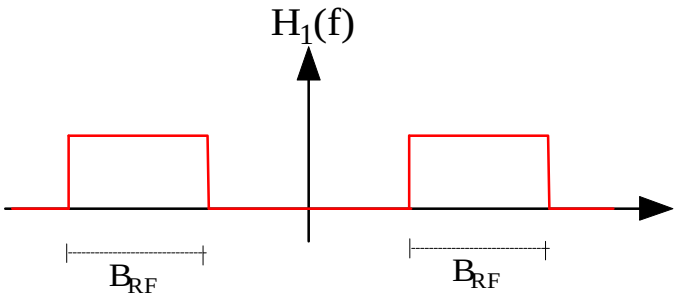
dove ovviamente s(t) è il segnale analogico oggetto della trasmissione.

Al fine di tener conto della presenza di rumore introdotto dagli apparati di modulazione, poniamo a monte del demodulatore un sommatore che aggiunge ad $s_t(t)$ il solito rumore n(t) gaussiano bianco:



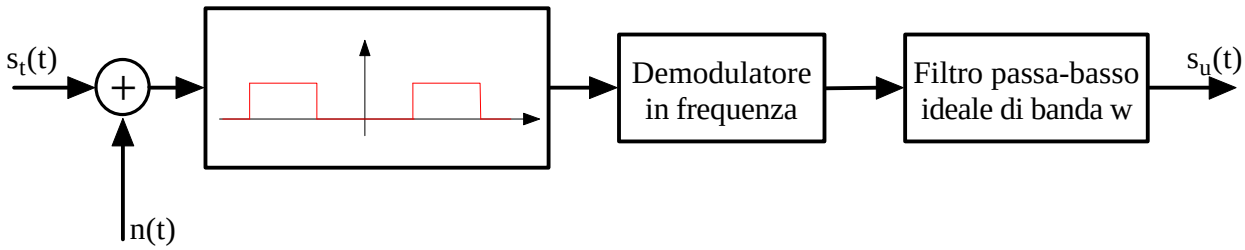
Ancora una volta, in accordo a quanto fatto nei casi precedenti di modulazione, possiamo subito porre un filtro passa-banda subito dopo il sommatore: questo filtro è fatto in modo da lasciar passare inalterato il segnale utile $s_t(t)$ e da filtrare invece le componenti di rumore eterne all'intervallo di frequenza in cui il segnale utile è definito.

Il filtro avrà perciò una funzione di trasferimento del tipo seguente:



Chiaramente, B_{RF} è la cosiddetta **banda di radiofrequenza** che, nella approssimazione di Carson, vale $B_{RF}=\Delta f_{PP}+2w$, dove w è la banda del segnale modulante s(t) (passa-basso) mentre Δf_{PP} è la **deviazione di frequenza picco-picco**.

Lo schema di demodulazione diventa dunque il seguente:



All'uscita dal filtro passa-banda abbiamo il segnale $x(t)=s_t(t)+n_F(t)$. Sostituendo l'espressione del segnale modulato, abbiamo

$$x(t)=A_C\cos\left(2\pi f_C t+2\pi k_F\int_{-\infty}^{t}s(\tau)d\tau\right)+n_F(t)$$

Per quanto riguarda il *rumore filtrato*, essendo di tipo passa-banda sappiamo che è esprimibile come somma di due componenti di rumore in *quadratura* tra loro:

$$n_F(t) = n_I(t)\cos(2\pi f_C t) + n_q(t)\sin(2\pi f_C t)$$

Se allora poniamo

$$\text{ampiezza}: \; r(t) = \sqrt{n_I^2(t) + n_q^2(t)}$$

$$\text{fase}: \qquad \varphi(t) = \text{arctg}\left(\frac{n_I(t)}{n_q(t)}\right)$$

possiamo riscrivere il rumore filtrato nella forma

$$n_F(t) = r(t)\cos(2\pi f_C t + \varphi(t))$$

per cui il segnale in uscita dal filtro passa-banda assume l'espressione

$$x(t) = A_C \cos\left(2\pi f_C t + 2\pi k_F \int_{-\infty}^{t} s(\tau)d\tau\right) + r(t)\cos(2\pi f_C t + \varphi(t))$$

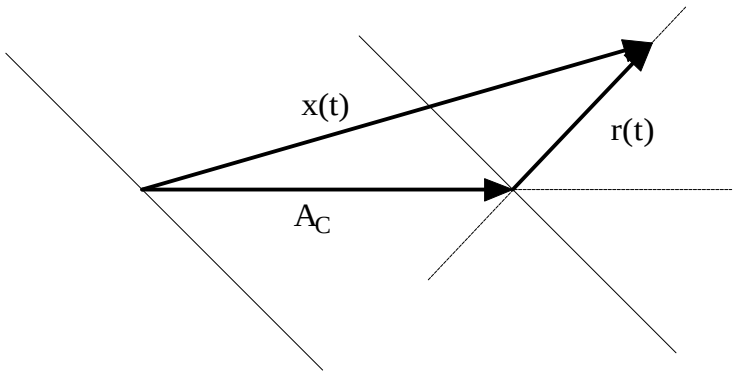
Se inoltre poniamo

$$\Phi(t) = 2\pi k_F \int_{-\infty}^{t} s(\tau)d\tau$$

possiamo ancora una volta riscrivere x(t) nella forma

$$x(t) = A_C \cos(2\pi f_C t + \Phi(t)) + r(t)\cos(2\pi f_C t + \varphi(t))$$

Questa espressione è utile in quanto ci consente di rappresentare le due componenti di x(t) in termini di *vettori rotanti*:



piano rotante

L'angolo che il vettore x(t) forma con l'asse orizzontale (e quindi col vettore A_C) è q(t), mentre F(t) è l'angolo che il vettore x(t) forma con l'asse inclinato. Il piano considerato è un piano rotante, che quindi ruota con pulsazione $2pf_C$ coincidente con quella della portante (che quindi appare ferma)

La risultante x(t) ha una ampiezza che evidentemente non dipende dal termine di segnale $\Phi(t)$ che a noi interessa. Valutiamo allora la fase di x(t), ossia l'angolo formato tra l'asse orizzontale del piano rotante e x(t) appunto: esso risulta essere

$$\theta(t) = \Phi(t) + \text{arctg}\frac{r(t)\sin\left(\varphi(t) - \Phi(t)\right)}{A_C + r(t)\cos\left(\varphi(t) - \Phi(t)\right)}$$

Si tratta in pratica dell'angolo del segnale x(t) che va in ingresso al demodulatore vero e proprio.

Il demodulatore, che deve tirar fuori s(t), non deve far altro che eseguire la derivata di q(t), in modo da ottenere la pulsazione istantanea, e successivamente dividere per 2p, in modo da ottenere la frequenza istantanea: essendo un po' complessa la derivata dell'espressione appena trovata per $\theta(t)$, facciamo qualche ipotesi semplificativa:

- in primo luogo, supponiamo che la potenza di rumore sia molto minore della potenza del segnale utile: questa ipotesi corrisponde a dire che, al denominatore dell'argomento della "arctg", possiamo trascurare il secondo termine rispetto ad A_C, per cui

$$\theta(t) \cong \Phi(t) + \text{arctg}\frac{r(t)\sin\left(\varphi(t) - \Phi(t)\right)}{A_C}$$

- in secondo luogo, supponiamo di poter trascurare il termine $\Phi(t)$ rispetto al termine $\varphi(t)$ nell'argomento del "seno", per cui l'espressione di $\theta(t)$ diventa

$$\theta(t) \cong \Phi(t) + \text{arctg}\frac{r(t)\sin\left(\varphi(t)\right)}{A_C} = \Phi(t) + \text{arctg}\frac{n_q(t)}{A_C} \cong \Phi(t) + \frac{n_q(t)}{A_C}$$

Andando allora a calcolare la derivata, divisa per 2π, si trova quanto segue:

$$x_1(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt} = \frac{1}{2\pi}\frac{d}{dt}\left[\Phi(t) + \frac{n_q(t)}{A_C}\right] = \frac{1}{2\pi}\frac{d\Phi(t)}{dt} + \frac{1}{2\pi A_C}\frac{dn_q(t)}{dt}$$

Il segnale $x_1(t)$ è dunque ciò che viene fuori da quello che nello schema abbiamo indicato come **demodulatore di frequenza**. Ricordando che

$$\Phi(t) = 2\pi k_F \int_{-\infty}^{t} s(\tau)d\tau$$

è evidente che tale segnale ha anche quest'altra espressione:

$$x_1(t) = k_F s(t) + \frac{1}{2\pi A_C}\frac{dn_q(t)}{dt}$$

Abbiamo dunque ancora una volta la somma del segnale utile $k_F s(t)$ e di una componente di rumore.

L'ultimo passo è il filtraggio di questo segnale: l'effetto è un segnale $s_u(t)$ corrispondente ad $x_1(t)$ privato delle componenti in frequenza esterne all'intervallo $[-w,+w]$.

Rapporto segnale-rumore

Possiamo allora andarci a calcolare il rapporto segnale rumore.
Per quanto riguarda la potenza del segnale, essa vale

$$S = E\left[\left(k_F s(t)\right)^2\right] = k_F^2 P_S$$

Per quanto riguarda la potenza del rumore, abbiamo che

$$N = E\left[\left(\frac{1}{2\pi A_C}\frac{dn_q(t)}{dt}\right)^2\right] = \frac{1}{A_C^2}E\left[\left(\frac{1}{2\pi}\frac{dn_q(t)}{dt}\right)^2\right]$$

Per calcolare questa potenza, dobbiamo fare qualche osservazione preliminare.

Intanto, possiamo considerare il segnale $\frac{1}{2\pi}\frac{dn_q(t)}{dt}$ come l'uscita di un particolare "derivatore" al quale arriva in ingresso il segnale $n_q(t)$:

La funzione di trasferimento del derivatore è

$$H_{deriv}(f) = \frac{1}{2\pi} j2\pi f = jf$$

dove $j2\pi f$ è la funzione di trasferimento del *derivatore ideale*.

Sulla base di ciò, possiamo determinare lo spettro di potenza del segnale $\frac{1}{2\pi}\frac{dn_q(t)}{dt}$ sapendo che lo spettro di potenza del segnale $n_q(t)$ è $S_X(f) = \frac{kT}{2}$ (spettro bilatero): abbiamo infatti che

$$S_Y(f) = S_X(f)\left|H_{deriv}(f)\right|^2 = \frac{kT}{2}f^2 = \frac{kT}{2}f^2$$

L'espressione $\boxed{S_Y(f) = \frac{kT}{2}f^2}$ è di importanza fondamentale per capire i pregi della modulazione/demodulazione di frequenza: essa infatti mostra che *un demodulatore di frequenza ha l'effetto per cui il rumore in uscita non è più bianco (come nel caso della demodulazione di ampiezza), ma ha una densità spettrale di potenza che cresce proporzionalmente con il quadrato della frequenza.*

Autore: Sandro Petrizzelli

In altre parole, un demodulatore FM, ricevendo in ingresso un rumore con contenuto energetico costante su tutte le frequenze, effettua una *sagomatura del rumore* in modo tale che il suo contenuto energetico sia maggiore alle alte frequenze. Questo può essere un vantaggio quando il segnale a cui tale rumore è sovrapposto ha un contenuto informativo concentrato sulle basse frequenze, ossia laddove il rumore è meno rilevante: tipico sarà il caso del **segnale televisivo**, che sarà esaminato in seguito.

Noto dunque lo spettro di potenza $S_Y(f)$ del segnale $\frac{1}{2\pi}\frac{dn_q(t)}{dt}$, possiamo calcolarci la sua potenza, ricordando che essa è pari all'area sottesa, tra -w e +w (visto che stiamo considerando le trasformate bilatere) di $S_Y(f)$: riprendendo le formule ricavate prima, abbiamo dunque che la potenza di rumore in uscita dal demodulatore vale

$$N=\frac{1}{A_C^2}E\left[\left(\frac{1}{2\pi}\frac{dn_q(t)}{dt}\right)^2\right]=\frac{1}{A_C^2}\int_{-w}^{+w}S_Y(f)df=\frac{1}{A_C^2}\int_{-w}^{+w}\frac{kT}{2}f^2df=\frac{1}{A_C^2}\frac{kT}{2}\int_{-w}^{+w}f^2df=\frac{1}{A_C^2}\frac{kT}{2}\left[\frac{f^3}{3}\right]_{-w}^{+w}=$$

$$=\frac{kTw^3}{3A_C^2}$$

Possiamo dunque concludere che il rapporto S/N in uscita dal demodulatore è

$$\left(\frac{S}{N}\right)_{OUT}=\frac{3A_C^2k_F^2}{kTw^3}P_S$$

Questa espressione può anche essere scritta in una forma più significativa:

$$\left(\frac{S}{N}\right)_{OUT}=3\frac{1}{\frac{kT}{2}w}\left(\frac{k_F}{w}\right)^2\frac{A_C^2}{2}P_S$$

In questa formula, kT/2 rappresenta la densità spettrale di potenza *bilatera* del rumore termico, che indichiamo normalmente con h_n; k_F rappresenta la deviazione di frequenza picco-picco, che possiamo indicare con Δf_{PP}; inoltre, indicata con P_R la potenza media del segnale modulato FM, ossia la potenza media ricevuta dal canale, si ricava facilmente che essa vale

$$P_R=E\left[s_t^2(t)\right]=E\left[A_C^2\cos^2\left(2\pi f_Ct+2\pi k_F\int_{-\infty}^{t}s(\tau)d\tau\right)\right]=A_C^2E\left[\cos^2\left(2\pi f_Ct+2\pi k_F\int_{-\infty}^{t}s(\tau)d\tau\right)\right]=\frac{A_C^2}{2}$$

dove abbiamo tenuto conto che il segnale modulato è un segnale sinusoidale, di ampiezza costante (pari ad A_C) e di valor medio nullo, per cui il suo quadrato è a sua volta una sinusoide di ampiezza A_C^2 e valor medio 1/2.

Di conseguenza, possiamo riscrivere l'espressione del rapporto S/N in uscita da un demodulatore FM come

$$\boxed{\left(\frac{S}{N}\right)_{OUT}=3\left(\frac{\Delta f_{PP}}{w}\right)^2\frac{P_R}{h_nw}}$$

Rapporto S/N in ingresso ed in uscita al demodulatore

Possiamo adesso confrontare il rapporto S/N tra l'uscita e l'ingresso del demodulatore FM:

$$\left(\frac{S}{N}\right)_{OUT} = 3\left(\frac{\Delta f_{PP}}{w}\right)^2 \frac{P_R}{h_n w}$$

$$\left(\frac{S}{N}\right)_{IN} = \frac{P_R}{P_{rumore}} = \frac{P_R}{h_n B_{RF}} = \frac{P_R}{h_n (\Delta f_{PP} + 2w)}$$

(dove abbiamo considerato la *densità spettrale bilatera* h_n del rumore in ingresso).

Facciamo il rapporto al fine di effettuare un confronto:

$$\frac{\left(\frac{S}{N}\right)_{IN}}{\left(\frac{S}{N}\right)_{OUT}} = \frac{\frac{P_R}{h_n (\Delta f_{PP} + 2w)}}{3\left(\frac{\Delta f_{PP}}{w}\right)^2 \frac{P_R}{h_n w}} = \frac{1}{3}\frac{1}{(\Delta f_{PP} + 2w)}\frac{w^3}{\Delta f_{PP}^2}$$

Questa formula non è per la verità molto significativa. Al contrario, possiamo ricavarne un'altra più importante.

Conservando la definizione sul rapporto S/N in uscita del demodulatore, cambiamo la definizione di rapporto S/N a monte, scegliendo questa volta di misurare la potenza di rumore non più nella *banda di Carson*, ma nella banda w del segnale modulante (passa-basso):

$$\left(\frac{S}{N}\right)_{OUT} = 3\left(\frac{\Delta f_{PP}}{w}\right)^2 \frac{P_R}{h_n w}$$

$$\left(\frac{S}{N}\right)_{IN} = \frac{P_R}{P_{rumore}} = \frac{P_R}{h_n w}$$

Facendo il rapporto tra le due quantità, otteniamo allora

$$\frac{\left(\frac{S}{N}\right)_{IN}}{\left(\frac{S}{N}\right)_{OUT}} = \frac{\frac{P_R}{h_n w}}{3\left(\frac{\Delta f_{PP}}{w}\right)^2 \frac{P_R}{h_n w}} = \frac{1}{3\left(\frac{\Delta f_{PP}}{w}\right)^2} = \frac{1}{3}\left(\frac{w}{\Delta f_{PP}}\right)^2$$

Quello che si nota, da questa espressione, è che il rapporto S/N subisce un aumento o una diminuzione, a parità di banda del segnale modulante s(t), a seconda del valore di Δf_{PP}: evidentemente, se risulta $\frac{1}{3}\left(\frac{w}{\Delta f_{PP}}\right)^2 < 1$, il rapporto S/N subisce un miglioramento in uscita rispetto al valore che aveva in ingresso. Di conseguenza, si dovrà procedere a scegliere Δf_{PP} in modo tale che

$$\Delta f_{PP}^2 > \frac{w^2}{3}$$

Quanto maggiore è la deviazione di frequenza (ossia l'indice di modulazione), tanto maggiore è l'aumento del valore del rapporto S/N, ossia tanto migliori sono le prestazioni del demodulatore. Questo è in accordo anche all'espressione trovata prima per F, che diminuisce all'aumentare di Δf_{PP}.

Queste considerazioni mostrano dunque come *la modulazione angolare, e in particolare quella di frequenza, sia una modulazione efficiente, nel senso che essa può consentire, a parità di rumore h_n in ricezione ed a parità di rapporto S/N in uscita, di operare con minore potenza rispetto alla trasmissione diretta (cioè senza modulazione) o alla modulazione d'ampiezza ([1])*. Ovviamente, lo svantaggio è quello di richiedere un allargamento dello spettro da trasmettere, ossia una maggiore occupazione di banda. Ci si può esprimere, quindi, dicendo che *la maggiore efficienza della modulazione FM, rispetto alla modulazione AM oppure alla trasmissione in banda base, si ottiene al prezzo di una maggiore occupazione di banda*.

Il vantaggio della modulazione angolare, rispetto alla modulazione d'ampiezza, è dunque la maggiore protezione contro il rumore, pagata appunto con più banda occupata. Un ulteriore vantaggio, in molte applicazioni, è dato anche dal fatto che il segnale modulato ha inviluppo costante, il che rende la modulazione robusta nei confronti di eventuali non linearità circuitali.

La condizione di soglia

In base alle considerazioni conclusive del paragrafo precedente, si potrebbe pensare che sia conveniente elevare il più possibile il valore di Δf_{PP}, ossia l'indice di modulazione. Questo non è assolutamente vero, in quanto bisogna tener conto che una grandezza strettamente legata al valore di Δf_{PP} è proprio la banda di radiofrequenza B_{RF} occupata dal segnale modulato: aumentando Δf_{PP}, si ottiene un aumento di B_{RF}.

Questo aumento di B_{RF} presenta due controindicazioni fondamentali:

- la prima è nella maggiore occupazione di frequenza sul mezzo trasmissivo: se ci viene assegnata una certa banda B_{RF} sul mezzo, potremo scegliere un valore massimo di Δf_{PP}, imposto appunto dal valore fisso di B_{RF} (lo si vedrà meglio nell'esempio che seguirà);
- la seconda, ancora più importante, è che, all'aumentare di B_{RF}, non facciamo altro che far passare una maggiore quantità di rumore nel demodulatore, il che può invalidare l'ipotesi, fatta in precedenza, per cui la potenza del rumore stesso è molto minore di quella del segnale.

Concentriamoci su quest'ultimo aspetto: l'espressione

[1] Ricordiamo, infatti, che, in un sistema di trasmissione con modulazione di ampiezza DSB-SC, il rapporto S/N all'uscita del demodulatore rimane invariato rispetto all'ingresso, per cui il sistema è del tutto equivalente ad un sistema di trasmissione diretta in banda base.

$$\left(\frac{S}{N}\right)_{OUT} = 3\left(\frac{\Delta f_{PP}}{w}\right)^2 \frac{P_R}{h_n w}$$

per il rapporto S/N in uscita dal demodulatore di frequenza vale, come si è detto, fin quando il rumore ingurgitato dal demodulatore è piccolo: infatti, in questo caso, nella formula

$$\theta(t) = \Phi(t) + \text{arctg}\frac{r(t)\sin(\varphi(t) - \Phi(t))}{A_C + r(t)\cos(\varphi(t) - \Phi(t))}$$

è possibile trascurare il termine di rumore $r(t)\cos(\varphi(t) - \Phi(t))$ a denominatore e il termine di segnale $\Phi(t)$ a numeratore.

Finché queste approssimazioni sono valide, allora l'espressione di $\left(\frac{S}{N}\right)_{OUT}$ è quella ricavata e quindi essa indica che possiamo garantirci le stesse prestazioni riducendo la potenza in ricezione (e quindi quella in trasmissione) e aumentando Δf_{PP}.

Tuttavia, man mano che diminuiamo P_R e aumentiamo Δf_{PP}, il rumore diventa sempre più confrontabile con il segnale, fino a quando si raggiunge una condizione di soglia, per una particolare coppia di valori di P_R e Δf_{PP}, in corrispondenza della quale il demodulatore subisce un rapido deterioramento di prestazioni.

Esiste una relazione matematica che, quando verificata, garantisce la correttezza dell'ipotesi per cui la potenza del rumore è molto minore di quella del segnale: si tratta della cosiddetta **condizione di soglia**, secondo la quale deve risultare

$$\left.\frac{P_R}{h_n B_{RF}}\right|_{dB} > 10(dB)$$

dove P_R è la **potenza ricevuta**, ossia la potenza che arriva in ingresso all'apparato di demodulazione (ossia, anche, la potenza in uscita dal canale: tale potenza, nell'ipotesi di canale ideale, coincide con la **potenza trasmessa**, ossia la potenza del segnale modulato in uscita dal modulatore), mentre $h_n B_{RF}$ è la potenza di rumore ingurgitata dal demodulatore e calcolata in questo caso tramite la *densità spettrale monolatera* $\mathbf{h_n = kT}$ di rumore ([2]).

Quella condizione impone in pratica un limite minimo alla differenza tra la potenza del segnale ricevuto e la potenza di rumore ad esso sovrapposto. Finché questo limite minimo è garantito, il demodulatore garantisce buone prestazioni; nel momento in cui la condizione viene meno, le prestazioni calano notevolmente.

Riuscire quindi a migliorare le prestazioni contro il rumore ricorrendo ad un aumento dell'indice di modulazione trova un limite ineludibile in questi fenomeni di soglia.

Questo fatto evidenzia una ulteriore differenza tra un sistema di modulazione FM ed un sistema di modulazione d'ampiezza: fin quando viene soddisfatta la condizione di soglia, il sistema con modulazione FM è sicuramente quello che garantisce, a parità di potenza trasmessa, le migliori prestazioni possibili; nel

[2] Il motivo per cui consideriamo una densità spettrale di rumore monolatera (solo frequenze positive), mentre nei conti precedenti abbiamo considerato quella bilatera, è che la condizione di soglia viene da un discorso prettamente fisico sul funzionamento del demodulatore e quindi non ha senso considerare le frequenze negative, che nella realtà non esistono ma solo un artificio matematico introdotto dall'analisi di Fourier.

momento in cui, invece, la condizione di soglia viene a mancare, per esempio a causa di una eccessiva riduzione della potenza trasmessa P_T, il demodulatore FM è praticamente non adottabile, mentre il demodulatore AM dà prestazioni tanto peggiori quanto minore è P_T.

In altre parole, *mentre le prestazioni di un demodulatore AM decrescono linearmente al diminuire di P_T, quelle di un demodulatore FM hanno un andamento diverso, in quanto sono ottime al di sopra della soglia, mentre sono pessime al di sotto della soglia.*

La demodulazione e l'effetto del rumore

Abbiamo visto che il rumore all’uscita del demodulatore FM ha uno spettro di potenza di tipo *parabolico*, il che significa che sono le componenti a frequenza più elevata che contribuiscono maggiormente al rumore totale d'uscita. Si può allora pensare di ridurre tali componenti di rumore mediante un opportuno filtraggio (detto **deenfasi**) all'uscita del discriminatore.

Tuttavia, così facendo però si riducono anche le componenti di segnale alle frequenze alte, distorcendo il segnale stesso. Si può però rimediare a tale distorsione con un prefiltraggio del segnale modulante in trasmissione (**enfasi**), che sia ovviamente del tutto complementare al filtraggio di deenfasi in ricezione (deve cioè risultare che la cascata dei due filtri di enfasi e deenfasi non distorca il segnale): si può ottenere questo facendo in modo che il filtro in trasmissione abbia funzione di trasferimento inversa rispetto a quella del filtro di deenfasi.

Osservazione

Per concludere, facciamo una osservazione a proposito del concetto di *potenza* del segnale utile.

E’ infatti importante richiamare l’attenzione sulle differenze esistenti tra il caso in cui l’amplificatore finale in trasmissione è limitato in potenza media e quello in cui esso è limitato in potenza di picco: *se la limitazione è sulla potenza di picco, come spesso accade, il vantaggio della modulazione angolare sugli altri tipi di trasmissione è particolarmente consistente, in quanto eventuali non linearità dell’amplificatore in trasmissione non hanno alcun effetto sul segnale trasmesso.* Ciò significa, come ritorneremo a dire in seguito, che è possibile usare amplificatori in trasmissione spinti fino alla saturazione, in quanto, così facendo, si ottiene il massimo del rendimento, ossia si riesce a trasferire al carico (sarebbe l’antenna trasmittente) quasi tutta la potenza prelevata dalle alimentazioni.

Questa considerazione spiega perché, molto spesso, sia conveniente usare la modulazione angolare anche con deviazione efficace minore di 1.

Esempio numerico

In un sistema di trasmissione FM, il canale di trasmissione ha una attenuazione di 90dB in una banda passante di 1 MHz centrata intorno alla frequenza della portante. Nelle ipotesi che

- il segnale modulante abbia una banda di 100 kHz;
- il fattore di rumore delle apparecchiature di ricezione sia di 10dB;
- la potenza in uscita dal modulatore sia di 10mW,

calcolare la deviazione di frequenza utilizzabile, compatibile con la banda a disposizione, affinché il sistema lavori sopra soglia.

Risoluzione

Ricordando che la banda occupata dal segnale modulato FM vale $B_{RF} = \Delta f_{PP} + 2f_{max} = 2(\Delta f_P + f_{max})$, deduciamo che la deviazione di frequenza di picco vale

$$\Delta f_P = \frac{B_{RF}}{2} - f_{max}$$

Il valore massimo di Δf_P si ottiene imponendo che la banda a radiofrequenza occupata sia esattamente pari a quella messa a disposizione sul canale, ossia 1 MHz:

$$(\Delta f_P)_{max} = \frac{(B_{RF})_{max}}{2} - f_{max} = \frac{1\ \text{MHz}}{2} - 100\ \text{kHz} = 400\ \text{kHz}$$

La condizione affinché il sistema considerato lavori sopra-soglia è rappresentata, in unità naturali, da

$$\frac{P_R}{h_n B_{RF}} \geq 10$$

dove P_R è la potenza di segnale ricevuta dal demodulatore, h_n la densità spettrale monolatera di potenza del rumore sovrapposto al segnale in ingresso al demodulatore e B_{RF} la banda (centrata sulla portante) occupata dal segnale modulato (il termine $h_n B_{RF}$ rappresenta infatti la potenza media del rumore sovrapposto al segnale utile).

E' immediato calcolarsi la potenza media ricevuta, che è pari a quella trasmessa (10 mW) diminuita della attenuazione (90 dB) introdotta dal mezzo trasmissivo:

$$P_R[\text{dB}] = P_T[\text{dB}] - \alpha[\text{dB}] = (10\text{mW})_{\text{dBm}} - 90[\text{dB}] = -80[\text{dBm}] \longrightarrow P_R = 10^{-11}\,\text{W}$$

La densità spettrale (monolatera) di rumore è valutabile come $\mathbf{h_n = FkT_0}$, dove possiamo ritenere che T_0 sia la temperatura ambiente (293°K) e dove abbiamo introdotto il fattore di rumore (=10dB per ipotesi) per tenere conto della rumorosità intrinseca del demodulatore ([3]): sotto questa ipotesi, abbiamo che

[3] Ricordiamo che, se il demodulatore non fosse rumoroso, il termine di rumore in ingresso sarebbe semplicemente quello che è accompagnato al segnale e che quindi si modella con la classica $h_n=kT_0$. Volendo invece considerare la rumorosità del demodulatore, si considera un termine aggiuntivo di rumore, che è quantificabile, per definizione, o tramite il **fattore**

$$(h_n)_{dBm/Hz} = 10\log_{10}\frac{FkT_0}{1mW} = 10\log_{10}\frac{kT_0}{1mW} + 10\log_{10} F == -174[dBm] + 10[dB] =$$
$$= -164[dBm/Hz] \longrightarrow h_n = 4*10^{-20} \text{ W/Hz}$$

Possiamo dunque riscrivere la condizione di soglia, in unità naturali, come

$$10 \le \frac{P_R}{h_n B_{RF}} = \frac{25*10^7}{B_{RF}}$$

da cui quindi otteniamo che

$$B_{RF} \le 25MHz$$

La condizione di soglia è dunque sempre soddisfatta, visto che il valore massimo di B_{RF} è 1 MHz per le specifiche imposte dalla traccia. Possiamo dunque usare il massimo valore della deviazione di frequenza di picco, ossia 400 kHz.

Autore: **Sandro Petrizzelli**
e-mail: sandry@iol.it
sito personale: http://users.iol.it/sandry

di rumore (che va a moltiplicare kT_0) oppure tramite la **temperatura equivalente di rumore** (nel qual caso si considera in ingresso kT_0+kT_{eq})