

Tecniche di MODULAZIONE

- Fonia e Digitale -

La modulazione è la trasformazione di un segnale in una forma adatta alla trasmissione via radio. Generalmente consiste nel traslare un segnale in banda base, quale può essere ad esempio la voce di un interlocutore, in un segnale passa banda centrato su una frequenza molto più alta. Il segnale passa banda risulta perciò modulato, mentre quello in banda base è il segnale modulante.

La modulazione può essere fatta variando l'ampiezza, la fase o la frequenza di una portante ad alta frequenza, in accordo con l'andamento del segnale da trasmettere.

La demodulazione è invece il processo inverso di estrarre il segnale in banda base da quello modulato, in modo che possa essere opportunamente utilizzato dal ricevitore.

Modulazioni analogiche

Modulazione d'Ampiezza (AM)

Con questo tipo di modulazione, l'ampiezza della portante è variata in accordo al valore istantaneo del segnale modulante. Se $s(t)$ è la portante e $m(t)$ il segnale modulante, il segnale AM può essere espresso come:

$$s_{AM}(t) = A_c [1 + m(t)] \cos(2\pi f_c t) \quad (2.1)$$

L'indice di modulazione di ampiezza è definito come il rapporto tra la massima ampiezza del segnale modulante e il valore di picco della portante. Se il segnale modulante è una sinusoidale di frequenza f_m ,

$$m(t) = \left(A_m / A_c \right) \sin(2\pi f_c t)$$

, allora l'indice di modulazione risulta:

$$k = \frac{A_m}{A_c} \quad (2.2)$$

L'indice di modulazione, spesso espresso in percentuale, non può essere superiore al 100% , se non si vuole che il messaggio venga distorto usando come demodulatore un rivelatore di inviluppo.

Lo spettro di un segnale AM si può scrivere come:

$$S_{AM}(f) = \frac{1}{2} A_c [\mathcal{D}(f - f_c) + M(f - f_c)] + \frac{1}{2} A_c [\mathcal{D}(f + f_c) + M(f + f_c)]$$

(2.3)

in cui $\mathcal{D}(f)$ è la funzione impulso unitario ed $M(f)$ è lo spettro del segnale modulante. Lo spettro consiste in un impulso alla frequenza della portante e due bande laterali che replicano lo spettro del messaggio modulante. Perciò la larghezza di banda di un segnale AM è pari a:

$$B_{AM} = 2f_m$$

(2.4) dove f_m è la massima frequenza contenuta nel segnale modulante.

La potenza di un segnale AM risulta:

$$P_{AM} = \frac{1}{2} A_c^2 [1 + 2\langle m(t) \rangle + \langle m^2(t) \rangle]$$

(2.5) in cui $\langle \bullet \rangle$

rappresenta il valore medio.

La demodulazione di un segnale AM può essere fatta in maniera coerente o incoerente. La demodulazione coerente consiste nel moltiplicare il segnale AM ricevuto $R(t) \cos(2\pi f_c t + \phi_r)$ con un'oscillazione avente la stessa frequenza:

$$v(t) = R(t) \cos(2\pi f_c t + \theta_r) \cdot A_0 \cos(2\pi f_c t + \theta_0) =$$

$$= \frac{1}{2} A_0 R(t) \cos(\theta_r - \theta_0) + \frac{1}{2} A_0 R(t) \cos(2\pi \cdot 2f_c t + \theta_r + \theta_0)$$

Modulazione di frequenza (FM)

La modulazione di frequenza consiste nel variare linearmente la frequenza istantanea della portante in accordo al segnale modulante $m(t)$:

$$s_{FM}(t) = A_c \cos\left[2\pi f_c t + \theta(t)\right] = A_c \cos\left[2\pi f_c t + 2\pi k_f \cdot \int_{-\infty}^t m(\tau) d\tau\right]$$

(2.7)

in cui k_f esprime l'entità della deviazione di frequenza (misurata in Hz/volt). L'indice di modulazione di frequenza rappresenta invece la relazione tra la massima deviazione di frequenza Δf nel segnale modulato e la banda W del segnale modulante:

$$\beta_f = \frac{\Delta f}{W}$$

(2.8)

Se il segnale modulante è $m(t) = A_m \cos(2\pi f_m t)$

, allora $\beta_f = (k_f A_m) / f_m$

e lo spettro SFM(f) contiene la portante e un numero infinito di righe attorno ad essa, poste ad una distanza pari ad un multiplo intero della frequenza modulante f_m e aventi un'ampiezza che può essere ricavata dalle funzioni di Bessel valutate in corrispondenza di β_f :

$$s_{FM}(t) = A_c \cdot \sum_{n=-\infty}^{\infty} (-1)^n J_n(\beta_f) \cos[2\pi(f_c + n f_m)t]$$

(2.9)

Se il segnale modulante non è monocromatico, lo spettro si complica notevolmente, non solo per la presenza delle righe proprie di ogni componente frequenziale, ma anche perché nascono numerose combinazioni di esse. Non è possibile ricavare analiticamente lo spettro di un segnale modulato in frequenza con un segnale a spettro continuo, però possiamo valutarne la banda occupata con la formula di Carson:

$$B = 2 \cdot (\beta_f + 1) f_m$$

(2.10)

in cui f_m è la massima frequenza presente nel segnale modulante. Anche se teoricamente la banda di un segnale FM è infinita, in pratica nella banda B scritta sopra abbiamo il 98% della potenza del segnale. Per piccoli valori dell'indice di modulazione ($b \ll 1$) abbiamo la FM a banda stretta (NBFM), la cui banda risulta $B \approx 2f_m$, perciò oltre alla portante a frequenza f_c si hanno due righe laterali a $f_c \pm f_m$; per $b \gg 1$ abbiamo invece la FM a banda larga (WBFM) che ha banda $B \approx 2b f_m = 2D f$.

Per generare un segnale FM si può usare il metodo diretto, che fa uso di un VCO controllato dal segnale modulante, oppure il metodo indiretto che fa uso della seguente approssimazione valida per segnali NBFM:

$$\begin{aligned} s_{FM}(t) &= A_c \cos[2\pi f_c t + \theta(t)] = \\ &= A_c \cos 2\pi f_c t \cdot \cos \theta(t) - A_c \sin 2\pi f_c t \cdot \sin \theta(t) \cong \\ &\cong A_c \cos 2\pi f_c t - A_c \theta(t) \sin 2\pi f_c t \end{aligned} \quad (2.11)$$

Quindi il segnale NBFM si può ottenere come somma di una portante e di un segnale di tipo AM con portante in quadratura rispetto alla prima. Un segnale WBFM si può poi ottenere da uno NBFM inviandolo a un moltiplicatore di frequenza per N , però con lo svantaggio di accrescere il rumore di fase; infatti, le variazioni aleatorie nella frequenza della portante vengono anch'esse moltiplicate per N .

I demodulatori FM, detti anche discriminatori di frequenza, possono usare varie tecniche. Una di queste consiste nell'usare un derivatore seguito da rivelatore di involuppo; infatti, dopo la derivazione si ottiene:

$$v(t) = -2\pi A_c \left[f_c + k_f m(t) \right] \sin[2\pi f_c t + \theta(t)] \quad (2.12)$$

Un altro modo è quello di sfruttare la parte laterale della curva di risposta di un filtro passa banda. Quando però è richiesta linearità in un ampio intervallo frequenziale, conviene fare uso di un convertitore frequenza-tensione che, generando un impulso di forma prefissata ad ogni ciclo del segnale d'ingresso, produce una tensione il cui valor medio è proporzionale alla frequenza istantanea. Un'altra tecnica molto nota è infine quella che fa uso di un PLL.

La modulazione di frequenza è la modulazione analogica più usata nei sistemi di comunicazione di tipo mobile, perché offre molti vantaggi rispetto alla modulazione di ampiezza.

Innanzitutto la FM ha una migliore immunità al rumore: infatti, l'informazione è contenuta nelle variazioni di frequenza anziché di ampiezza, quindi è meno soggetta ai disturbi atmosferici che tendono a causare piuttosto delle fluttuazioni nell'ampiezza. Diversamente dalla AM, la FM permette di aumentare il rapporto segnale-rumore aumentando l'indice di modulazione e quindi la banda occupata, così è possibile trovare un compromesso tra SNR e banda. Inoltre il segnale FM ha un inviluppo costante e permette quindi l'uso di efficienti amplificatori di potenza in classe C. Invece nella AM è importante avere linearità tra il segnale modulante e l'ampiezza del segnale trasmesso, perciò devono essere usati degli amplificatori in classe A o AB che non sono molto efficienti. L'uso di amplificatori di potenza che siano particolarmente efficienti è molto importante quando si ha a che fare con terminali portatili, perché da ciò dipende strettamente la durata della batteria. Infine la modulazione di frequenza ha anche un interessante proprietà di cattura del segnale più forte: se al ricevitore giungono due segnali aventi la stessa frequenza, quello che ha un livello maggiore viene demodulato mentre l'altro viene reiettato. Ciò rende i sistemi FM molto resistenti all'interferenza co-canale.

D'altra parte uno svantaggio dei sistemi FM è che richiedono un'occupazione di banda molto maggiore, proprio al fine di ottenere quei vantaggi nella riduzione degli effetti del rumore e nella proprietà di cattura. Inoltre il trasmettitore e il ricevitore per un sistema FM sono più complessi di quelli usati per la modulazione di ampiezza.

Digitalizzazione del segnale audio

I moderni sistemi di comunicazione di tipo mobile usano tecniche di modulazione di tipo digitale. Gli sviluppi tecnologici nel campo dei circuiti integrati a larghissima scala di integrazione (VLSI) e nell'elaborazione digitale dei segnali (DSP), hanno reso la modulazione digitale molto più conveniente economicamente. Inoltre essa offre molti altri vantaggi rispetto alla modulazione analogica. Ad esempio, la modulazione digitale presenta una migliore immunità al rumore, consente una più facile moltiplicazione di varie forme di informazione (voce, dati, immagini), assicura un più alto livello di sicurezza nelle comunicazioni con l'impiego di tecniche di crittografia.

Nei sistemi wireless di tipo digitale, il segnale modulante consiste in una successione temporale di simboli (impulsi), ciascuno dei quali può assumere un numero finito m di stati. Quindi ogni simbolo rappresenta n bit di informazione, con $n = \log_2 m$

. Nel caso si voglia trasmettere un segnale analogico come lo è quello vocale, occorre perciò compiere un campionamento e una successiva conversione analogico-digitale dei campioni così ottenuti. La frequenza di campionamento, in accordo con il criterio di Nyquist, deve essere almeno pari al doppio della frequenza massima contenuta nel segnale da campionare. A tale proposito si deve tener presente che, anche se il campo di udibilità va da 20Hz a 16kHz circa, in pratica si può ridurre notevolmente l'estensione dello spettro audio ai fini di una

conversazione telefonica; infatti, quello che conta è che ogni utente possa riconoscere il proprio interlocutore e capire ciò che dice. La banda assegnata ai sistemi telefonici si estende pertanto tra 300Hz e 3400Hz. Inoltre la frequenza di campionamento è stata standardizzata a 8kHz con la conversione di ogni campione su 8 bit, perciò vengono trasmessi 64kbit/sec. L'invio nel canale di trasmissione di una successione di impulsi, ognuno dei quali rappresenta un bit del segnale campionato, è, di fatto, ciò che si chiama Pulse Code Modulation (PCM), tecnica impiegata nel sistema telefonico mondiale. Il canale impiegato permette in questo caso una velocità di trasmissione maggiore di 64kbit/sec, e ciò consente di far avvenire contemporaneamente più conversazioni, usando una moltiplicazione a divisione di tempo (Time Division Multiplexing – TDM).

Nel campo delle comunicazioni cellulari, diversamente da quello delle comunicazioni telefoniche via cavo in cui c'è un'ampia disponibilità di banda, è necessario ridurre il numero di bit al secondo inviati; inoltre è importante trasmettere quanta meno potenza possibile. Facendo uso di opportune tecniche di predizione, si può arrivare a trasmettere solo 10-15kbit/sec. Tali tecniche sfruttano la correlazione che esiste tra un campione e i precedenti, per cui, anziché trasmettere tale campione, si trasmette un valore più piccolo corrispondente alla differenza tra esso ed una sua stima. Ad esempio si può calcolare una stima di tipo lineare per il campione $x(n)$:

$$\tilde{x}(n) = a_1x(n-1) + a_2x(n-2) + \dots + a_Nx(n-N) \quad (2.13)$$

e ciò che viene trasmesso è:

$$d(n) = x(n) - \tilde{x}(n) \quad (2.14)$$

Se i coefficienti a_1, a_2, \dots, a_N

sono scelti in modo opportuno, la stima $\tilde{x}(n)$

sarà abbastanza accurata e $d(n)$

sarà molto più piccolo di $x(n)$

, perciò lo si potrà trasmettere con un numero di bit inferiore. Ovviamente il ricevitore calcolerà $\tilde{x}(n)$

con lo stesso algoritmo e potrà quindi ricostruire il campione originario $x(n) = \tilde{x}(n) + d(n)$

. Una tecnica di questo tipo prende il nome di Differential PCM (DPCM), proprio perché viene trasmessa la differenza $d(n)$

. Inoltre si possono avere dei coefficienti che variano in base alle caratteristiche del segnale e allora si parla di Adaptive Differential PCM (ADPCM).

Modulazioni digitali

Per misurare le prestazioni di un particolare tipo di modulazione digitale, vengono usati due parametri: l'efficienza energetica e l'efficienza di banda. L'efficienza energetica è in grado di misurare la capacità che ha un certo tipo di modulazione digitale di trasmettere correttamente le informazioni anche con bassi livelli di potenza. Infatti, ai fini di ridurre gli effetti del rumore, è necessario aumentare sufficientemente la potenza del segnale per ottenere un'accettabile probabilità di errore. L'efficienza energetica si esprime come il rapporto tra l'energia del segnale per un singolo bit E_b e la densità spettrale di potenza del rumore N_0 , in modo che si ottenga nel ricevitore una certa probabilità di errore:

$$\eta_P = \frac{E_b}{N_0}$$

(2.15)

L'efficienza di banda descrive invece la capacità che ha un certo tipo di modulazione digitale di usare una banda modesta. In generale, se si aumenta il numero di bit al secondo inviati (bit-rate), occorre diminuire la larghezza dell'impulso associato a un simbolo da trasmettere, e ciò aumenta la banda occupata dal segnale. L'efficienza di banda è quindi definita come il rapporto tra un certo bit-rate (R) e la corrispondente banda occupata dal segnale modulato (B):

$$\eta_B = \frac{R}{B}$$

(2.16)

essa ovviamente si misura in bit/sec per Hertz, ed influenza strettamente la capacità del sistema di comunicazione.

Esiste un limite superiore all'efficienza di banda ottenibile in un determinato canale di trasmissione. Il teorema di Shannon stabilisce infatti che:

$$\eta_{Bmax} = \frac{C}{B} = \log_2 \left(1 + \frac{S}{N} \right)$$

(2.17)

dove C è la capacità del canale (in bit/sec) ed S/N il rapporto segnale-rumore in esso presente.

Nel progetto di un sistema di comunicazione digitale si deve spesso adottare un compromesso tra efficienza di banda ed efficienza energetica. Ad esempio, se si trasmettono delle

informazioni per il controllo degli errori, si aumenta la banda occupata a parità di velocità di trasmissione, ma si riduce la potenza di segnale richiesta per rimanere al di sotto di una certa probabilità di errore. Ci sono anche altri fattori che influenzano nella scelta di un particolare tipo di modulazione digitale, come la sensibilità a disturbi e interferenze nel canale o come la complessità e il costo del ricetrasmettitore.

Una scelta importante che va fatta per un sistema di comunicazione digitale è la forma degli impulsi del segnale modulante. E' chiaro che tali impulsi non potranno essere di forma rettangolare, perché occuperebbero una banda infinita, ma la limitazione della banda implica che essi saranno estesi nel tempo. Il segnale ricevuto dopo la demodulazione si presenta in questa forma:

$$g(t) = \sum_i a_i p\left(t - \frac{i}{R}\right)$$

(2.18)

in cui $p(t)$ rappresenta la forma del generico impulso, e $R=1/T$ è la velocità di trasmissione, cioè il numero di impulsi inviati in un secondo. Il ricevitore effettua il campionamento agli istanti iT in modo da ricostruire la successione a_i , ma il campione m -esimo risulta:

$$g(mT) = a_m + \sum_{i \neq m} a_i p\left(\frac{m-i}{R}\right)$$

(2.19)

per cui l'ampiezza a_m dell'impulso m -esimo viene alterata dai termini della sommatoria, dovuti agli impulsi che lo precedono e lo seguono. Questo fenomeno viene chiamato interferenza intersimbolica (ISI).

Se B è la banda in cui si vuole contenere lo spettro della successione di impulsi, Nyquist ha dimostrato che è possibile trasmetterla ad una velocità $R \leq 2B$, senza introdurre interferenza intersimbolica, purché la forma dell'impulso $p(t)$ sia scelta opportunamente. Una forma di $p(t)$ che non dà luogo ad ISI è la seguente:

$$p(t) = \frac{\text{sen}(\pi R t)}{\pi R t} = \text{sinc}(Rt) \Leftrightarrow P(f) = \frac{1}{R} \text{rect}\left(\frac{f}{R}\right)$$

(2.20)

Anche se tale impulso annulla l'ISI occupando la minima larghezza di banda possibile ($R/2$), ci sono vari motivi per cui non è conveniente usarlo. Oltre ad essere la risposta di un sistema non

causale (poiché è non nulla per $t < 0$) ed è difficile da approssimare, $p(t)$ si annulla solo ad esatti multipli di $T=1/R$ con delle oscillazioni che si smorzano come t^{-1} ; perciò un piccolo errore nel campionare esattamente negli istanti di attraversamento dello zero, causerà un errore significativo dovuto all'ISI. Uno degli impulsi più usati nelle comunicazioni di tipo mobile è allora quello a coseno rialzato:

$$p(t) = \frac{\cos(2\pi f t)}{1 - (4f/R)^2} \cdot \text{sinc}(Rt) \Leftrightarrow$$

$$\Leftrightarrow P(f) = \frac{\pi}{4fR} \cos\left(\frac{f}{2R}\right) \cdot \text{rect}\left(\frac{f}{2R}\right) \otimes \text{rect}\left(\frac{f}{R}\right)$$

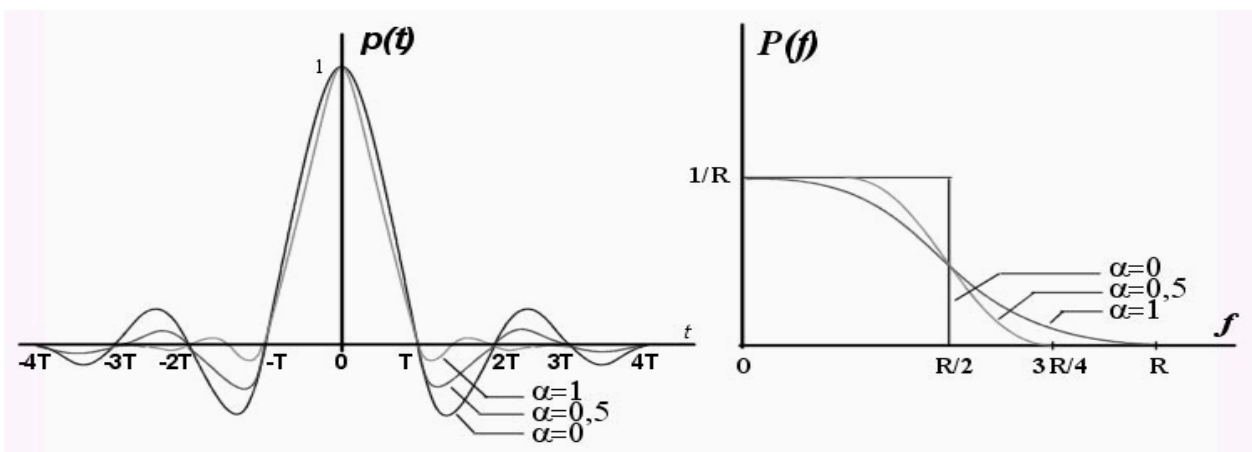
(2.21)

Il rapporto tra b e la banda minima di Nyquist $R/2$, viene chiamato roll-off:

$$\alpha = \frac{2b}{R}$$

(2.22)

Se $\alpha = 0$ si ottiene l'impulso di prima, mentre all'aumentare di α le oscillazioni laterali dell'impulso si smorzano sempre più rapidamente, ma la banda occupata aumenta. Al limite per $\alpha = 1$ si ottiene il massimo smorzamento con una banda pari a R , cioè il doppio di quella minima. Normalmente la forma dell'impulso viene realizzata tramite DSP e, poiché dovrebbe essere estesa da $-\infty$ a $+\infty$, è indispensabile adottare una sua approssimazione, generalmente



tra $-6T$ a $+6T$.

Un altro tipo di impulso molto usato è quello gaussiano:

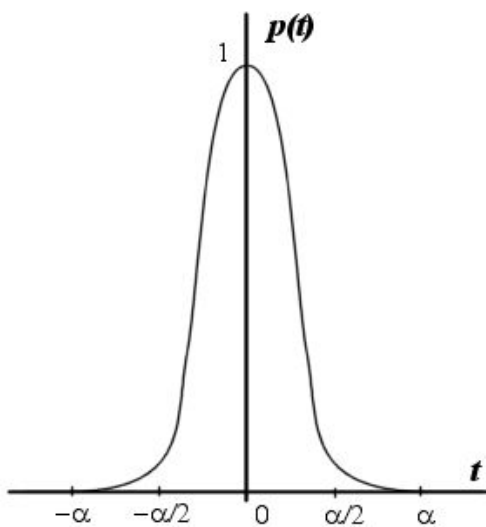
$$p(t) = \exp\left(-\frac{\pi^2}{\alpha^2} t^2\right) \Leftrightarrow P(f) = \frac{\alpha}{\sqrt{\pi}} \exp(-\alpha^2 f^2)$$

(2.23)

in cui il parametro α è legato alla banda B a -3dB del suo spettro:

$$\alpha = \frac{1}{B} \sqrt{\frac{\ln 2}{2}}$$

(2.24)



Tale impulso non si annulla in corrispondenza dei multipli del periodo di trasmissione $T=1/R$, quindi l'ISI non si annulla nemmeno teoricamente, ma ha un'importante proprietà; diversamente dall'impulso a coseno rialzato, se avvengono delle distorsioni non troppo elevate, lo spettro dell'impulso gaussiano non cambia drasticamente la propria estensione, e ciò consente di usare degli efficienti amplificatori di potenza non lineari. Per quanto riguarda il problema dell'ISI, si può constatare che i suoi effetti possono essere trascurati se il prodotto $B \cdot T$ è maggiore di 0,5.

E' stato affermato che si può limitare sensibilmente l'interferenza intersimbolica facendo in modo che gli impulsi che giungono al campionario del ricevitore abbiano una forma opportuna. Ma la forma degli impulsi ricevuti dipende ovviamente anche dalla risposta in frequenza del canale di trasmissione, nonché da quella del filtro di ricezione, il quale serve per limitare il rumore. Se $C(f)$ ed $R(f)$ sono rispettivamente le risposte del canale e del filtro di ricezione, l'impulso $g(t)$ ricevuto dipenderà da quello trasmesso $p(t)$ nel seguente modo:

$$G(f) = P(f) \cdot C(f) \cdot R(f) \Leftrightarrow g(t) = p(t) \otimes c(t) \otimes r(t)$$

(2.25)

Poiché le caratteristiche del canale non sono perfettamente note, spesso prima del campionatore si inserisce un equalizzatore, cioè un filtro la cui risposta può essere facilmente modificata in modo da compensare le caratteristiche del canale e minimizzare l'ISI. Nel caso delle comunicazioni mobili, in cui la risposta del canale è anche variabile nel tempo, si usano degli equalizzatori adattivi, i quali correggono automaticamente la loro risposta servendosi del segnale ricevuto.

Per i sistemi di comunicazione analogici è importante valutare il rapporto segnale-rumore che si riesce ad ottenere all'uscita del ricevitore, perché ciò permette di poter stimare, secondo il tipo di messaggio, se il segnale è utilizzabile soddisfacentemente. Invece per i sistemi di tipo digitale tale rapporto è meno significativo: infatti, finché il rumore non altera il segnale tanto da determinare un'interpretazione errata del campione, il ricevitore presenta alla sua uscita esattamente il campione trasmesso. Per questo è stato detto prima che i sistemi digitali sono meno sensibili al rumore. Di conseguenza, per tali sistemi risulta più interessante andare a valutare la probabilità che il ricevitore commetta un errore nella valutazione di un campione, cioè il cosiddetto Bit Error Rate (BER). Esso dipende dal tipo di modulazione impiegata ed è influenzato, oltre che dall'inevitabile rumore termico (Additive White Gaussian Noise – AWGN), da fenomeni caratteristici per un sistema di comunicazione di tipo mobile, come l'effetto Doppler o il fading del segnale ricevuto, dovuto alla propagazione di esso secondo percorsi differenti.

Modo - Phase Shift Keying - (PSK)

La tecnica di modulazione PSK fa parte delle modulazioni digitali di tipo lineare, nelle quali l'ampiezza del segnale trasmesso $s(t)$ varia linearmente con il segnale modulante $m(t)$. Anche se le modulazioni lineari hanno una buona efficienza spettrale, esse devono essere trasmesse usando degli amplificatori RF lineari, i quali sono poco efficienti in termini di potenza.

Un segnale PSK può essere scritto come:

$$s_{PSK}(t) = \sum_i A_c p(t - iT) \cos(2\pi f_c t + \varphi_i) \quad (2.26)$$

in cui la fase φ_i può assumere, in corrispondenza del simbolo i -esimo trasmesso, uno dei seguenti M valori differenti: $\varphi_i = 2\pi(k-1)/M + \vartheta$

, con $k=1,2,\dots,M$ e ϑ costante. Si parla pertanto di MPSK. Nel caso in cui $M=2$, la fase della portante può assumere solo due valori differenti e si ha la Binary PSK (BPSK):

$$v_H(t) = A_c p(t) \cos(2\pi f_c t + \vartheta)$$

$$\textcircled{R} \text{ bit 1 (2.27a)} \quad v_I(t) = A_c p(t) \cos(2\pi f_c t + \theta + \pi)$$

$$\textcircled{R} \text{ bit 0 (2.27b)}$$

$$\text{Se } m(t) = \sum_i A_c m_i p(t - iT)$$

, $m_i = \pm 1$, è il segnale modulante costituito dalla successione d'impulsi, si ottiene:

$$s_{\text{BPSK}}(t) = m(t) \cos(2\pi f_c t + \theta) \quad (2.28)$$

e quindi il segnale BPSK è equivalente ad un segnale DSB con portante soppressa. Perciò può essere generato con un modulatore bilanciato e il suo spettro coincide con quello di $m(t)$ traslato attorno alla frequenza della portante f_c .

Per quanto riguarda la demodulazione di un segnale BPSK, essa può avvenire solo in modo coerente (sincrono), cioè utilizzando un'oscillazione avente stessa frequenza e fase della portante; infatti, moltiplicando tale oscillazione con il segnale ricevuto, si ha:

$$m(t) \cos^2(2\pi f_c t + \theta) = \frac{1}{2} m(t) [1 + \cos(2\pi \cdot 2f_c t + 2\theta)] \quad (2.29)$$

ed $m(t)$ può essere estratto con un filtro passa basso. Per avere a disposizione l'oscillazione sincrona, può essere trasmesso assieme al segnale BPSK un tono pilota che viene recuperato nel ricevitore tramite un PLL, oppure può essere ricavato opportunamente, ma non in modo semplice, dallo stesso segnale BPSK.

Per evitare l'impiego di un demodulatore coerente nel ricevitore, in modo da renderlo più semplice ed economico, si può usare la Differential PSK (DPSK). Essa consiste nel trasmettere, anziché la sequenza originaria di bit $\{m_k\}$, la sequenza $\{d_k\}$ ottenuta complementando la somma modulo-2 di m_k e d_{k-1} : ciò equivale a lasciare il bit d_k inalterato dal precedente d_{k-1} se il bit originario m_k vale uno, e di invertire d_k rispetto a d_{k-1} se m_k vale zero. Questa codifica consente di demodulare il segnale moltiplicandolo per una sua replica ritardata di T ; se infatti $d_i = \pm 1$ e $m_i = \pm 1$:

$$\begin{aligned} d(t) \cos(2\pi f_c t + \theta) \cdot d(t-T) \cos[2\pi f_c (t-T) + \theta] &= \\ = m(t) \{ \cos(2\pi f_c T) + \cos[2\pi \cdot 2f_c (t-T/2) + 2\theta] \} & \\ (2.30a) \quad m(t) = d(t) \cdot d(t-T) &= \end{aligned}$$

$$= \sum_i d_i d_{i-1} p^2(t - iT) = \sum_i m_i p^2(t - iT)$$

(2.30b)

La DPSK ha quindi il vantaggio di un ricevitore meno complesso, ma ha un'efficienza energetica che è inferiore di circa 3dB rispetto alla BPSK. Ciò si spiega considerando che se viene alterato un bit d_k dal rumore, quasi sicuramente il ricevitore commetterà un errore su due bit, m_k e $m_k + 1$, perciò la probabilità di errore è circa doppia.

La Quadrature PSK (QPSK) ha un'efficienza di banda che è il doppio della BPSK, in quanto sono trasmessi due bit per ogni simbolo. Perciò la fase della portante può assumere uno di quattro valori ugualmente distanziati, e ciascuno di essi corrisponde ad una coppia di bit. Il singolo impulso modulato si può allora scrivere come:

$$v_{QPSK}(t) = A_c p(t) \cos \left[2 \pi f_c t + (i - 1) \frac{\pi}{2} + \theta \right]$$

$i = 1, 2, 3, 4$

(2.31)

Si può dimostrare che la probabilità di errore della QPSK è la stessa della BPSK, quindi l'efficienza energetica è la stessa.

Per quanto riguarda la trasmissione e la ricezione, si può sfruttare che la (2.31) è trasformabile in questa forma:

$$v_{QPSK}(t) = A_c p(t) \cos \left[(i - 1) \frac{\pi}{2} + \theta \right] \cos(2 \pi f_c t) - A_c p(t) \sin \left[(i - 1) \frac{\pi}{2} + \theta \right] \sin(2 \pi f_c t)$$

(2.32)

Ponendo $q = \pi/4$, si ottiene che il segnale QPSK è la somma di due segnali BPSK con portanti in quadratura tra loro:

$$s_{QPSK}(t) = \sum_i \frac{\sqrt{2}}{2} A_c I_i p(t - iT) \cos(2 \pi f_c t)$$

$$\begin{aligned}
 & - \sum_i \frac{\sqrt{2}}{2} A_c Q_i p(t - iT) \text{sen}(2 f_c t) = \\
 & = m_I(t) \cos(2 f_c t) - m_Q(t) \text{sen}(2 f_c t) \\
 & \left[m_I(t) \cos \omega_c t - m_Q(t) \text{sen} \omega_c t \right] \cdot \cos \omega_c t = \\
 & = \frac{m_I(t)}{2} (1 + \cos 2 \omega_c t) - \frac{m_Q(t)}{2} \text{sen} 2 \omega_c t \\
 & (2.34a) \left[m_I(t) \cos \omega_c t - m_Q(t) \text{sen} \omega_c t \right] \cdot \text{sen} \omega_c t = \\
 & = \frac{m_I(t)}{2} \text{sen} 2 \omega_c t - \frac{m_Q(t)}{2} (1 - \cos 2 \omega_c t) \\
 & \theta_k = (i - 1) \cdot \pi / 4 \\
 & \theta_k = \theta_k - \theta_{k-1} \\
 & S_{\pi/4 \text{ DPSK}}(t) = \sum_i A_c p(t - iT) \cos(2 f_c t + \theta_i) = \\
 & = \sum_i A_c I_i p(t - iT) \cos(2 f_c t) - \sum_i A_c Q_i p(t - iT) \text{sen}(2 f_c t)
 \end{aligned}$$

Modo - Frequency Shift Keying - (FSK)

Molti sistemi di comunicazione mobile usano tecniche di modulazione non lineare come la FSK, nelle quali l'ampiezza della portante è mantenuta costante, indipendentemente dal segnale modulante. Tali tecniche presentano vari vantaggi, come la possibilità di usare efficienti amplificatori in classe C, dei demodulatori più semplici, una maggiore immunità a disturbi che provocano fluttuazioni nell'ampiezza del segnale. Lo svantaggio è che occupano una banda più larga rispetto alle modulazioni lineari.

Nella Binary FSK (BFSK), la frequenza di una portante di ampiezza costante può assumere due valori:

$$v_H(t) = A_c \cos[(2 f_c + 2 \pi \cdot \Delta f) \cdot t]$$

Ⓜ bit 1 (2.36a)

$$v_L(t) = A_c \cos[(2 f_c - 2 \pi \cdot \Delta f) \cdot t]$$

Ⓜ bit 0 (2.36b)

Un modo per generare un segnale di questo tipo, potrebbe essere di commutare tra due oscillatori indipendenti, in accordo al bit da trasmettere. Però questo porterebbe ad avere una forma d'onda con delle discontinuità di fase, con problemi di allargamento dello spettro e di emissioni spurie. Perciò conviene modulare in frequenza un solo oscillatore con il segnale modulante $m(t)$ di tipo binario:

$$s_{BFSK}(t) = A_c \cos \left[2\pi f_c t + 2\pi k_f \cdot \int_{-\infty}^t m(\tau) d\tau \right]$$

$$(2.37a) \quad m(t) = \sum_i m_i p(t - iT)$$

, $m_i = \pm 1$ (2.37b)

In questo modo, anche se $m(t)$ è discontinua nelle transizioni tra un bit e il successivo, la fase della portante è continua perché è proporzionale all'integrale di $m(t)$. Supponiamo per ora che l'impulso $p(t)$ sia rettangolare. Se, ad esempio, s'impone che la fase della portante sia sempre la stessa ad ogni transizione, la continuità di essa si ottiene se l'intervallo di trasmissione T contiene un numero intero di semiperiodi dell'oscillazione trasmessa:

$$f_1 = f_c - \Delta f \Rightarrow 2\pi f_1 T = \pi k$$

$$(2.38a) \quad f_2 = f_c + \Delta f \Rightarrow 2\pi f_2 T = \pi k$$

(2.38b)

Se Δf è la massima deviazione di frequenza rispetto a quella della portante f_c , e B è la banda di $m(t)$, allora la banda occupata dal segnale BFSK è data da:

$$B_{BFSK} = 2 \cdot (\Delta f + B)$$

(2.39)

in accordo con la formula di Carson (2.10). Se si vuole ridurre tale banda, un primo passo da fare è quello di diminuire Δf :

$$f_2 - f_1|_{MSK} = 2\Delta f|_{MSK} = \frac{1}{2T} = \frac{R}{2}$$

(2.40)

che si ricava dalle (2.38) con $h=k+1$. Ciò che si ottiene è la cosiddetta Minimum Shift Keying (MSK), che è un tipo particolare di FSK a fase continua in cui la differenza di frequenza $2\Delta f$ è la minima possibile, cioè la metà del bit-rate R . La frequenza f_2 associata al bit 1 rispetto a quella f_1 associata al bit 0, presenta mezza oscillazione in più nell'intervallo T . Rispetto alla

portante, invece, f_2 presenta un quarto di oscillazione in più, mentre f_1 ne presenta un quarto in meno. Questo implica che la fase della portante aumenta di 90° ogni volta che viene trasmesso il bit 1, mentre diminuisce di 90° quando viene trasmesso il bit 0. Perciò, dalla (2.37a), scrivendo il segnale MSK nel seguente modo:

$$s_{MSK}(t) = A_c \cos[2\pi f_c t + \theta(t) + \theta_0] \quad (2.41)$$

si può capire che esso presenta una fase $\theta(t)$ che con continuità può aumentare o diminuire di 90° , in base al bit trasmesso. La demodulazione può allora essere compiuta in maniera coerente, sfruttando le componenti in fase e in quadratura della portante:

$$\begin{aligned} s_{MSK}(t) &= \frac{A_c}{2} \cos \theta(t) \cdot \cos[2\pi f_c t + \theta_0] - \frac{A_c}{2} \sin \theta(t) \cdot \sin[2\pi f_c t + \theta_0] = \\ &= I(t) \cdot \cos[2\pi f_c t + \theta_0] - Q(t) \cdot \sin[2\pi f_c t + \theta_0] \end{aligned} \quad (2.42)$$

Le componenti $I(t)$ e $Q(t)$ possono essere ricavate moltiplicando il segnale MSK per due oscillazioni in quadratura tra loro, come si è visto per altre modulazioni. Dalla conoscenza dei valori assunti da $I(t)$ e $Q(t)$ e da quelli che essi avevano per il bit precedente, si può ricavare il bit attuale, perché si può capire se la fase è aumentata oppure è diminuita di 90° .

Se si vuole ridurre ulteriormente l'estensione dello spettro, non resta altro da fare che sagomare opportunamente l'impulso $p(t)$. Nella Gaussian Minimum Shift Keying (GMSK) $p(t)$ è un impulso gaussiano. In questo modo si diminuiscono un po' le prestazioni in termini di BER, a causa della presenza dell'ISI, in cambio però di un'ottima efficienza spettrale. Infatti, eliminando le discontinuità nelle transizioni di frequenza con l'uso di impulsi gaussiani, si rendono molto meno brusche le transizioni di fase, riducendo così in modo considerevole la presenza dei lobi laterali nello spettro trasmesso. Comunque, le transizioni di fase della GMSK restano sempre di $\pm 90^\circ$ per ogni bit trasmesso, perciò essa può essere demodulata coerentemente come la semplice MSK.

La GMSK è un tipo di modulazione molto attraente per il progetto di un sistema di comunicazione wireless, sia per la ridotta occupazione di banda, sia per l'efficienza con la quale si riesce ad amplificarla, dovuta al suo inviluppo costante. Però, l'uso di impulsi gaussiani irrimediabilmente introduce l'ISI, aumentando quindi la probabilità di errore in ricezione. Poiché dalla (2.24) la forma dell'impulso gaussiano può essere completamente definita dalla sua banda B , la modulazione GMSK usata in un sistema di comunicazione viene spesso caratterizzata dal prodotto BT , cioè tra la suddetta banda e la durata T di un bit. Se si osserva lo spettro ottenuto, si nota che al diminuire di BT i lobi laterali presenti in esso decrescono in ampiezza molto

rapidamente; ad esempio, per $BT=0,5$ il massimo del primo lobo laterale si trova a -30dB rispetto a quello del lobo principale, mentre è a soli -20dB nel caso di semplice MSK ($BT=1$). Ma la riduzione di BT inevitabilmente aumenta la probabilità di errore dovuta all'ISI. In pratica conviene ridurre BT fino a quando la probabilità di errore non aumenta significativamente rispetto a quella che si avrebbe per effetto del solo canale di trasmissione.

Riferimenti:

Theodore S. Rappaport, "Wireless Communications, Principles and Practice", cap. 5

Mario Mancianti
"Lezioni di Comunicazioni Elettriche"