4.3. SISTEMI DI MODULAZIONE NUMERICI

ione

e dili so-

sito-

re di

: rin-

simo

n ge-

sa in

:sen-

doibile ezza del-

ingo-

VIHz.

una

ıpata

men-...con

1Z8

izza-

ed il

che

·R e

ilant

nale nale piezban-

osi-

una

ie di

ni fil-

sot-

gna-

L(t)

Nei sistemi numerici si tratta di trasmettere in banda passante una sequenza di simboli tratti da un alfabeto costituito da M elementi: i due simboli 1 e 0 nel caso binario, oppure M valori numerici (ampiezze o altro), oppure ancora le lettere di un testo ecc. Nella modulazione, a ciascuno dei simboli corrisponderà un particolare valore di uno dei parametri dell'onda sinusoidale portante, tipicamente l'ampiezza o la fase o la frequenza. Ma potrà anche aversi la modulazione combinata di più parametri contemporaneamente, per esempio ampiezza e fase. In molti casi si costruisce anzitutto un segnale in banda base, cioè una sequenza di impulsi di forma prefissata g(t), alla frequenza di simbolo $\frac{1}{T}$:

$$m(t) = \sum_{k} a_{k} g(t - kT)$$
 [4.1]

 a_k rappresenta l'ampiezza del k-esimo impulso, e può assumere tipicamente, nel caso binario, i valori ± 1 oppure 1 e 0, e nel caso più generale di M livelli di ampiezza i valori $\pm 1, \pm 3, ..., \pm (M-1)$. Il segnale viene poi trasferito in banda passante mediante modulazione di una portante, per esempio modulazione di ampiezza, nello stesso modo già visto per il caso analogico.

Da un punto di vista teorico, molto di quanto già detto per la trasmissione in banda base vale anche per la banda passante. Per il caso generale di trasmissione di M possibili simboli rappresentati da M segnali determinati e distinti in banda passante, si può evidentemente ancora dire che un sistema di ricezione ottimale potrà essere basato su un insieme di M filtri adattati (correlatori) alle M possibili forme d'onda di segnale. Tralasciando la teoria generale, si descrivono nel seguito i sistemi di maggior interesse per le applicazioni, calcolandone le prestazioni per il caso classico di riferimento di rumore additivo gaussiano bianco con densità (unilaterale) N_0 .

4.3.1 Modulazione di ampiezza

Per quanto già detto per il caso analogico, la modulazione (e demodulazione) in ampiezza equivale ad una pura traslazione di frequenza, e l'analisi del sistema di trasmissione può essere condotta in modo identico al caso di trasmissione in banda base. Basta considerare il canale in banda base equivalente al canale in banda passante (si veda l'Appendice B). Trasmettendo in modulazione di ampiezza un segnale del tipo [4.1] su una portante ω_0 , il ricevitore coerente sarà configurato come in fig. 4.17. La moltiplicazione per la portante coerente

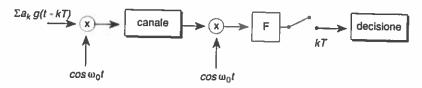


Figura 4.17 - Sistema di trasmissione numerico in modulazione d'ampiezza. [F: filtro adattato a g(t)]

 $\cos \omega_0 t$, che trasla il segnale in banda base, è seguita da un filtro adattato alla forma d'onda impulsiva g(t), e tutto ciò è chiaramente identico ad una operazione di correlazione del segnale ricevuto con il segnale $g(t)\cos \omega_0 t$. Ricordiamo a questo riguardo che, nel caso di canale con rumore bianco a densità $\frac{N_0}{2}$, il rapporto segnale-rumore all'uscita di

un filtro adattato (correlatore) è pari a $\frac{u_s^2}{\sigma_n^2} = \frac{2E_s}{N_0}$. Le prestazioni, per una modulazione a M livelli di ampiezza, sono date dalla [2.6].

Anche nel caso numerico, si può far ricorso come nel caso analogico alla modulazione a banda laterale unica o a banda laterale parzialmente soppressa, o alla modulazione d'ampiezza in quadratura. Ci soffermiamo in particolare su quest'ultima, usata in molti sistemi di trasmissione di dati numerici.

4.3.2 Modulazione d'ampiezza in quadratura (QAM)

Si è visto che due portanti isofrequenziali ma sfasate di $\frac{\pi}{2}$ sono in grado di trasportare in modulazione d'ampiezza due segnali senza interferenza mutua, in quanto perfettamente separabili in ricezione mediante demodulazione coerente. Consideriamo ad esempio due segnali numerici in banda base

$$\sum_{k} a_{k} g(t-kT); \sum_{k} b_{k} g(t-kT)$$

Questi due segnali possono naturalmente derivare dalla suddivisione in due flussi paralleli di un unico flusso di dati proveniente da una sorgente. Se i livelli di ampiezze delle due portanti, rispettivamente a_k e b_k , assumono per esempio i valori ± 1 , ± 3 , possiamo rappresentare il sistema con il diagramma vettoriale (rappresentante i segnali trasmessi sul canale con le loro ampiezze e fasi) illustrato nella fig. 4.18: i due assi rappresentano le due portanti ortogonali ($\cos \omega_0 t$; $\sin \omega_0 t$), e i punti rappresentano i segnali (vettori) trasmessi sul canale, con le loro ampiezze e fasi, risultanti dalla somma dei due canali in quadratura. Si

iı

d

e

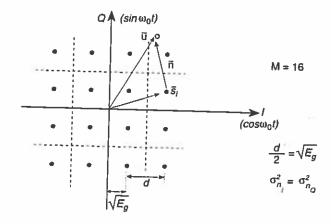


Figura 4.18 - Modulazione di ampiezza in quadratura (M=16).

le n

źί

er

li

e

e

1-

ši

tratta di una "costellazione" a 16 punti, ciascuno dei quali rappresenta un impulso sinusoidale. Se tale sistema è usato per la trasmissione di dati binari, ogni impulso porterà $\log_2 16 = 4$ bit $(a_k e b_k)$ portano ciascuno due bit). Per esempio potremo trasmettere, su un canale telefonico, 9600 bit/s modulando in ampiezza in quadratura una portante a 1800 Hz al ritmo di 9600/4 = 2400 simboli/s. La modulazione di ampiezza in quadratura, consistendo di due modulazioni di ampiezza sovrapposte ma separabili, realizza una efficienza nell'utilizzazione della banda disponibile (quantità d'informazione trasmessa al secondo per unità di banda) che è doppia rispetto alla modulazione d'ampiezza ordinaria. Nel caso di trasmissione di dati binari, possiamo analizzare il sistema come due sistemi modulati d'ampiezza completamente indipendenti (ciascuno a 4 livelli, nell'esempio di fig. 4.17); i due demodulatori di ampiezza coerenti in quadratura danno in uscita i valori stimati \hat{a}_k , \hat{b}_k e quindi i bit da essi rappresentati (fig. 4.19).

Ma evidentemente possiamo anche utilizzare in generale la costellazione di M punti per la trasmissione di M simboli qualsiasi. In questo caso a_k e b_k sono le coordinate del punto che rappresenta il simbolo trasmesso nell'intervallo kT; in ricezione i due demodulatori d'ampiezza in quadratura riproducono, in assenza di rumore, tali coordinate e quindi il simbolo trasmesso.

In presenza di rumore, le uscite dei demodulatori sono $u_t = a_k + n_t$ e $u_Q = b_k + n_Q$. Si è finora implicitamente assunto per semplicità che la forma d'onda $g(t)\cos\omega_0 t$ (ovvero $g(t)\sin\omega_0 t$) abbia energia unitaria; più in generale avremo che se E_g è l'energia di $g(t)\cos\omega_0 t$, le componenti di segnale all'uscita dei due filtri adattati (correlatori) possono essere espresse rispettivamente come $a_k \sqrt{E_g}$ e $b_k \sqrt{E_g}$ (con una normalizzazio-

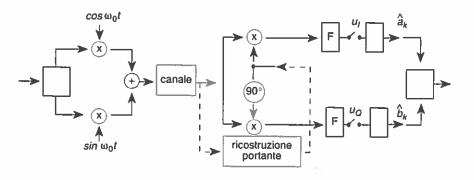


Figura 4.19 - Modulazione di ampiezza in quadratura. Schema di modulatore e demodulatore,

ne corrispondente all'uso, nella correlazione di ricezione, di un impulso di riferimento ad energia unitaria). Si osservi che con questa rappresentazione il modulo dei vettori rappresentativi dei vari segnali trasmessi $s_i(t)$ eguaglia la radice della loro energia $\sqrt{E_{st}}$. Se il rumore di canale è bianco gaussiano con densità spettrale di potenza $\frac{N_0}{2}$, le due

componenti di rumore n_l , n_Q hanno varianza pari a $\frac{N_0}{2}$, e inoltre so-

no statisticamente indipendenti (si veda l'Appendice B). Il vettore ricevuto $\overline{u} = (u_I, u_Q) = (a\sqrt{E_g} + n_I, b\sqrt{E_g} + n_Q)$ è la somma del vettore $\overline{s_i} = (a\sqrt{E_g}, b\sqrt{E_g})$ rappresentativo del segnale trasmesso e del vettore \overline{n} che rappresenta il rumore. Il punto ricevuto (u_I, u_Q) risulterà allora disperso attorno al punto trasmesso secondo una distribuzione statistica data da una campana di Gauss bidimensionale: la densità di probabilità $p[n_I, n_Q]$ del vettore rumore \overline{n} è il prodotto $p(n_I)p(n_Q)$ delle variabili gaussiane indipendenti n_I, n_Q .

Chiaramente, ricevuto un punto (u_l, u_Q) , la stima migliore del simbolo trasmesso consisterà nello scegliere tra i simboli possibili quello più vicino. Infatti questo, se i vari simboli sono a priori equiprobabili, è il simbolo per cui è massima la probabilità $p[\overline{s_i}/\overline{u}]$ di essere stato trasmesso dato il segnale ricevuto \overline{u} . Nella figura 4.18 sono anche riportate le M regioni di decisione Z_i che nel piano rappresentano il luogo dei punti più vicini a ciascuno degli M segnali possibili. Se si trasmette un dato simbolo (a_k, b_k) , in ricezione si commette errore quando il punto ricevuto (u_l, u_Q) cade esternamente alla regione di decisione relativa. Nel caso di costellazione rettangolare come in figura e per rumore gaussiano bianco, è semplice calcolare la probabilità d'errore, data l'indipendenza statistica delle componenti di rumore in quadratura; a seconda dei punti, si ha errore quando una o l'altra o entrambe

le compone
è la distana
QAM), si ca
versa tra sea
ni, che sono
quindi la pr
4 vertici, la
stanti 8 p
[1-Q][1-2
casi nell'ipo
probabilità

essendo Q

Si noti ch le/rumore è:

e quindi il se bile. Inoltre, termina le pr d, in rappor

Quanto es per $M=2^m$. scun simbolo più che la pro teressa la pro si può ricaval pensato come $\left(\frac{\log_2 M}{2}\right)$ bit analizzando s

stessi risultati Il caso di / un buon com;

che essere visto

si veda il para come due siste le componenti di rumore assumono valori in modulo maggiori di $\frac{d}{2}$ (d è la distanza tra punti vicini). Ad esempio nel caso della figura (16 QAM), si calcola la probabilità d'errore relativa ai 16 segnali, che è diversa tra segnali interni ed esterni nella costellazione. Per i punti interni, che sono 4, la probabilità di corretta decisione è $\left[1-2Q\left(\frac{d}{2\sigma_n}\right)\right]^2$ e quindi la probabilità d'errore è $1-[1-2Q]^2=4Q-4Q^2$; per i punti ai 4 vertici, la probabilità di corretta decisione è $[1-Q]^2$ mentre per i restanti 8 punti laterali la probabilità di corretta decisione è [1-Q][1-2Q]. Facendo la media delle probabilità d'errore per i vari casi nell'ipotesi di equiprobabilità a priori dei segnali, si ottiene per la probabilità di errore

$$P[E] = 3Q\left(\frac{d}{2\sigma_n}\right) - \frac{9}{4} Q^2\left(\frac{d}{2\sigma_n}\right),$$
essendo $Q\left(\frac{d}{2\sigma_n}\right) = Q\left(\sqrt{\frac{2E_g}{N_0}}\right).$

Si noti che normalmente nei sistemi pratici dove il rapporto segnale/rumore è sufficientemente alto, $Q\left(\frac{d}{2\sigma_n}\right)$ assume valori molto piccoli e quindi il secondo termine dell'espressione sopra riportata è trascurabile. Inoltre, dato l'andamento della funzione Q, il parametro che determina le prestazioni è essenzialmente la distanza minima tra i segnali d, in rapporto alla varianza del rumore.

Quanto esposto vale in generale per i vari valori di M, in particolare per $M=2^m$. Se l'informazione da trasmettere è di tipo binario, ciascun simbolo rappresenta $\log_2 M$ bit d'informazione. In questo caso, più che la probabilità di errore sul simbolo P[E] prima considerata, interessa la probabilità d'errore P[EB] relativa ai bit d'informazione, che si può ricavare dalla P[E]. Poiché in questo caso il sistema può essere pensato come due sistemi separati modulati in ampiezza a \sqrt{M} livelli $\left(\frac{\log_2 M}{2}\right)$ bit per livello, il conto della P[EB] si fa più semplicemente analizzando separatamente i due demodulatori in quadratura con gli stessi risultati già ricavati in banda base (Cap. 2).

Il caso di M=4 assume particolare rilievo, poiché si presenta come un buon compromesso tra efficienza e complessità. Il sistema può anche essere visto come modulazione di fase quaternaria (fasi $\pm \frac{\pi}{4}$, $\pm \frac{3}{4}$ π ; si veda il paragrafo successivo), oppure, nel caso di sorgenti binarie, come due sistemi binari in quadratura. Per questo sistema si ricava per

la probabilità d'errore P[E], considerando che le regioni di decisione sono i quattro quadranti, il valore

$$P[E] = 1 - P[C] = 1 - \left[1 - Q\left(\frac{d}{2\sigma_n}\right)\right]^2$$

essendo P[C] la probabilità di corretta decisione. Nel caso di sorgenti binarie (2 bit per simbolo) per il calcolo della P[EB] conviene interpretare il sistema come due sistemi binari in quadratura, e si ha subito $P[EB] = Q\left(\frac{d}{2\sigma_{\pi}}\right) = Q\left(\sqrt{\frac{2E_{g}}{N_{0}}}\right)$.

Se confrontiamo, per una sorgente che emetta $R=\frac{1}{T_b}$ bit/s, la semplice trasmissione binaria con modulazione d'ampiezza antipodale e la modulazione d'ampiezza in quadratura con M=4, a pari potenza trasmessa, si ha stessa P[EB] nei due casi. L'energia E_i è infatti la stessa poiché nel sistema QAM ciascuno dei canali in quadratura gode di una potenza di segnale che è la metà di quella disponibile per la trasmissione in modulazione d'ampiezza binaria, ma la durata di ciascun simbolo è doppia essendo pari a $T=2T_b=\frac{2}{R}$. Però la banda occupata, inversamente proporzionale alla durata di simbolo, è per il 4QAM la metà rispetto alla trasmissione binaria. Si ha cioè con il 4QAM un raddoppio dell'efficienza spettrale dovuto all'utilizzazione della quadratura, senza alcuna perdita sotto altri aspetti.

ESERCIZIO

Per una trasmissione di dati binari a R bit/s, si confronti la trasmissione con modulazione d'ampiezza binaria con la trasmissione a modulazione di ampiezza in quadratura con M = 16, dal punto di vista della protezione contro il rumore e dell'efficienza spettrale. Se effettuiamo il confronto a pari tasso d'errore, si può fare un conto approssimato semplificativo assumendo che la distanza minima tra i punti delle rispettive costellazioni sia la stessa. Ricordando che la durata di simbolo nel primo caso è pari a $T_b = 1/R$ mentre nel secondo è $4T_b$, le potenze di picco impiegate risultano rispettivamente $\frac{E_g}{T_b}$ e $\frac{18E_g}{4T_b}$. Si vede che il 16 QAM richiede una potenza di picco 4,5 volte maggiore ma una banda 4 volte inferiore. In generale quindi la scelta del tipo di modulazione più adatto dipenderà dalle condizioni di potenza e banda disponibili per la trasmissione.

ESEMPI

Si voglia trasmettere con un ponte radio un flusso di 155 Mbit/s, che rappresenta uno dei livelli fondamentali nella gerarchia standard delle velocità di trasmissione, adottata nelle reti di comunicazioni numeriche. Se la banda passante radio disponibile è di 40

NOTA
Lo scher
lizzato pr
gurazion
avere alt
essere d
rimane s
compone
vicino co
punti a re
nido d'ar
si calcola
ne vincola
radice de

ne

one

enti

ore-

bito

em-

e la

traessa

una

sioolo

:rsa-

à riopio

sen-

smisine a di vi-

le. connima

o che el se-

nente

picco

erale

dalle

lbit/s, stanınica-

di 40

MHz, si può adottare una modulazione 64 QAM (6 bit/simbolo); in tal modo la velocità di trasmissione (in baud) viene ridotta di un fattore 6, e la corrispondente banda minima richiesta è pari a 155/6 MHz. In pratica si userà per gli impulsi una forma d'onda di Nyquist con un certo arrotondamento (ad esempio α = 0.3), e ciò comporta un incremento di banda secondo un fattore (1 + α).

Se consideriamo invece il tradizionale canale telefonico della rete di telecomunicazioni analogica esistente, possiamo ritenere che esso abbia una banda passante di circa 3 KHz. Volendo trasmettere su di esso dati numerici, potremmo per esempio usare un sistema di modulazione simile a quello sopra considerato cioè un sistema 64 QAM alla velocità di 2400 baud sulla portante f₀ = 1800 Hz. Questo sistema trasporta 6·2400 = 14.400 bit/s. Nella realtà il canale telefonico analogico non è affatto ideale: il rumore e le distorsioni presenti fanno sì che tale elevata velocità di trasmissione sia raggiungibile solo con l'uso anche di equalizzazione di canale e codificazione per la correzione degli errori.

NOTA

Lo schema di modulazione illustrato in fig. 4.18 può evidentemente essere generalizzato pensando ad altri tipi di costellazioni di punti nel piano (I, Q). Oltre alla configurazione rettangolare corrispondente al QAM e che è la più semplice, si possono avere altre configurazioni più convenienti per particolari applicazioni e che possono essere descritte come modulazioni miste di ampiezza e fase. Per la demodulazione rimane sempre utilizzabile lo schema di fig. 4.19 basato su due demodulatori delle componenti in quadratura (I, Q); da queste si risale al punto della costellazione più vicino come stima del simbolo trasmesso. Per esempio con una configurazione di punti a reticolo triangolare (fig. 4.20), le regioni di decisione si presentano come un nido d'ape; sulla base di esse si definiscono le regole di decisione in ricezione e si calcola la probabilità di errore. Le limitazioni di potenza disponibile in trasmissione vincolano la costellazione ad essere interna alla circonferenza di raggio pari alla radice dell'energia di picco ammessa.

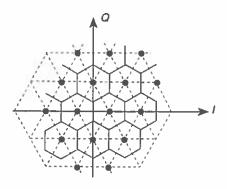


Figura 420 - Esempio di modulazione congiunta di ampiezza e di fase.

4.3.3 Demodulazione incoerente in modulazione di ampiezza numerica

Quando in ricezione non si voglia o non si possa ricostruire una portante coerente, si può in alcuni casi far uso di una demodulazione incoerente basata sulla rivelazione d'inviluppo. Se in ricezione si ha un impulso $v(t)\cos(\omega_0 t - \phi) = \cos\phi \cdot v(t)\cos\omega_0 t + \sin\phi \cdot v(t)\sin\omega_0 t$ con fase ϕ ignota, possiamo impiegare per filtrare il rumore un filtro adattato per esempio alla forma d'onda $v(t)\cos\omega_0 t$, prescindendo cioè dalla fase. Chiaramente anche l'uscita del filtro $p(t)\cos(\omega_0 t - \varphi)$ ha fase indeterminata, e si deve quindi ricorrere all'inviluppo p(t) per la decisione circa l'impulso trasmesso, inviluppo che è insensibile alla fase dell'onda portante. La ricezione è quindi basata su un filtro adattato (fase di riferimento qualsiasi) seguito da un rivelatore ad inviluppo, che viene poi campionato nell'istante di massima uscita. Questa struttura rappresenta l'ottima soluzione nel caso incoerente. La determinazione dell'inviluppo può anche essere realizzata in modo diverso, ricordando che un segnale a banda relativa stretta può essere rappresentato mediante le due componenti in quadratura u_I , u_Q e che l'inviluppo è dato da $\sqrt{u_1^2 + u_0^2}$ (si veda l'Appendice B). Il ricevitore incoerente può allora consistere in due correlatori in quadratura, come già visto nel QAM, seguiti però da quadratori per la determinazione dell'inviluppo. Un correlato-

re in fase con l'impulso in arrivo, cioè un correlatore che usasse come funzione di riferimento l'onda $\frac{1}{\sqrt{E_s}} v(t)\cos(\omega_0 t - \phi)$, darebbe in uscita il valore $\sqrt{E_s}$. Correlando invece con le funzioni (ortogonali) $\frac{1}{\sqrt{E_s}} v(t)\cos\omega_0 t$ e $\frac{1}{\sqrt{E_s}} v(t)\sin\omega_0 t$, si hanno rispettivamente i valori $\sqrt{E_s}\cos\phi$ e $\sqrt{E_s}\sin\phi$ da cui si ricava l'inviluppo.

In quanto detto, abbiamo fatto riferimento al segnale. In presenza di rumore, per quanto già detto circa la demodulazione analogica incoerente, le prestazioni, pur inferiori, non sono molto diverse dal caso coerente, a pari "distanza" tra i livelli di ampiezza e per alto rapporto segnale/rumore. Si ricordi però che la demodulazione ad inviluppo non può evidentemente essere usata quando il segnale modulante presenta ampiezze eguali in modulo ma opposte di segno (per esempio il caso antipodale $a_k = \pm 1$), in quanto il rivelatore ad inviluppo è insensibile al segno del segnale modulante.

Il caso tipico di applicazione è quello della modulazione binaria con due livelli di ampiezza 1 e 0 (modulazione "tutto o niente"). La distribuzione statistica all'uscita del rivelatore ad inviluppo non è più gaussiana come nel caso coerente: in presenza di segnale (livello 1) la somma di segnale e rumore gaussiano presenta un inviluppo con densità di probabilità data dalla legge di Rice, mentre quando si trasmette il livel-

lo 0 si ric statistica Trascura incoerent smessa, 1 trasmissirore dipe nari, si h riamente o niente" za (di pic

4.3.4 Mo Nella 1 phase shi portante

A

Se ϑ_k presentarifig. 4.21 i niente è $[0-2\pi]$. Se cide con te (QPSK) $(a_k = \pm 1, l$ Poiché vedere il se cui le corr

tra loro.



Figura 4.21 -

1erica

1 por-

ne in-

ha un

ase ϕ

to per

fase.

deter-

ne cir-

'onda

i rife-

ne poi resenl'invi-

he un

nte le

to da

allora

QAM, o. Un

usasse

be in

onali)

valori

esenza

ica in-

ıl caso

porto

lluppo le preipio il insen-

ia con distrigaus-

a somisità di I livello 0 si riceve solo rumore gaussiano e pertanto l'inviluppo è distribuito statisticamente secondo la legge di Rayleigh (si veda l'Appendice B). Trascurando le differenze di prestazioni tra demodulazione coerente e incoerente, si osservi che in questo caso, a pari potenza di picco trasmessa, la distanza tra i livelli d'ampiezza (1 e 0) è la metà rispetto alla trasmissione antipodale (1 e -1). Ricordando che la probabilità d'errore dipende dalla distanza tra i livelli che rappresentano i simboli binari, si ha che la trasmissione antipodale (con demodulazione necessariamente coerente) raggiunge le stesse prestazioni di un sistema "tutto o niente" con un livello di picco pari alla metà e quindi con una potenza (di picco) di segnale 4 volte inferiore.

4.3.4 Modulazione di fase

Nella modulazione di fase numerica (spesso indicata come PSK cioè phase shift keying), l'informazione è trasportata dalla fase di un'onda portante di frequenza e ampiezza costanti:

$$A\cos(\omega_0 t + \vartheta_k), \quad \vartheta_k = (2n+1)\frac{\pi}{M}, \quad n = 0, 1, \dots, M-1$$

Se ϑ_k può assumere M valori, ogni impulso sinusoidale può rappresentare uno tra M simboli oppure $\log_2 M$ bit d'informazione. La fig. 4.21 illustra i casi M=2, 4, 8; naturalmente la soluzione più conveniente è scegliere le fasi regolarmente distribuite nell'intervallo $[0-2\pi]$. Si osservi che il caso M=2 (indicato spesso come BPSK) coincide con una modulazione d'ampiezza binaria $(a_k=\pm 1)$, il caso M=4 (QPSK) con una modulazione d'ampiezza in quadratura $(a_k=\pm 1,b_k=\pm 1)$.

Poiché $A\cos(\omega_0 t + \vartheta_k) = A\cos\vartheta_k\cos\omega_0 t - A\sin\vartheta_k\sin\omega_0 t$, possiamo vedere il sistema come una modulazione di ampiezza in quadratura in cui le componenti in quadratura $a_k = A\cos\vartheta_k$, $b_k = A\sin\vartheta_k$ sono legate tra loro.

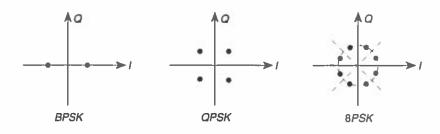


Figura 4.21 - Modulazione di fase (M=2, 4, 8).

La demodulazione ovviamente non può che essere coerente, essendo l'informazione impressa proprio sulla fase. In generale la demodulazione può essere realizzata con due demodulatori di ampiezza coerenti alle onde $\cos \omega_0 t$, $\sin \omega_0 t$ come nella modulazione di ampiezza in quadratura (nel caso M=2 ne basta uno ovviamente) (fig. 4.19). Infatti questi danno, come già visto, le componenti $a_k = A\cos \vartheta_k$, $b_k = A\sin \vartheta_k$ del vettore trasmesso (con l'aggiunta di termini spuri dovuti al rumore), e da queste si può determinare la fase ϑ_k trasmessa. Più precisamente, poiché la fase $\vartheta_k = \arctan \frac{u_Q}{u_I}$, stimata sulla base delle due componenti suddette, è affetta da errore, si determina la fase, tra le M possibili, più vicina al valore stimato. Chiaramente la probabilità di errore cresce al crescere di M, a pari valore dell'ampiezza A, poiché decresce la distanza tra i punti rappresentativi dei possibili segnali trasmessi.

Per il calcolo del tasso d'errore in presenza di rumore, si osservi anzitutto che le regioni di decisione Z_i sono gli M settori circolari di apertura $\pm \frac{\pi}{M}$ corrispondenti agli M punti. Si avrà errore in ricezione quando la deviazione angolare, rispetto al valore di fase trasmesso, prodotta dal rumore $\overline{n} = (n_i, n_Q)$ supera in modulo il valore $\frac{\pi}{M}$. Una stima semplice della probabilità di errore, tanto migliore quanto maggiore è M, si può avere ricorrendo ad un limite superiore facilmente calcolabile (per M=2 e 4 il conto esatto è immediato come già visto nella trattazione della modulazione di ampiezza).

Sia \bar{s}_i il simbolo trasmesso (fig.~4.22). Consideriamo uno dei due punti contigui, per esempio \bar{s}_{l+1} , e immaginiamo che sia, nella demodulazione, l'unica alternativa a \bar{s}_i . In questo caso di scelta binaria, si ha errore quando il vettore ricevuto \bar{u} cade nel semipiano A_{l+1} definito dalla mezzeria tra \bar{s}_i e \bar{s}_{l+1} e contenente \bar{s}_{l+1} . Lo stesso discorso può essere fatto per l'altro punto contiguo \bar{s}_{l-1} . Chiaramente la probabilità d'errore P(E), cioè la probabilità che \bar{u} cada al di fuori della regione Z_i relativa a \bar{s}_i , è inferiore alla somma delle probabilità che \bar{u} cada rispettivamente nei semipiani A_{l+1} e A_{l-1} , dato che questi sono in parte sovrapposti:

$$P(E) \le P(\overline{u} \in A_{i+1}) + P(\overline{u} \in A_{i-1})$$
 [4.2]

Queste probabilità si calcolano immediatamente considerando che la componente n_i del rumore \bar{n} lungo la congiungente \bar{s}_i con \bar{s}_{i+1} (o con \bar{s}_{i-1}) ha le stesse caratteristiche di n_i e n_Q , cioè è gaussiana con varianza pari a $\frac{N_0}{2}$ (Appendice B). Se d è la distanza di \bar{s}_i da

Figu

 \bar{S}_{i+1}

con Se I pres

e la

S dula tenz pres a div cresc per v lazio dula ampi to in

one

endo
iularenti
quaifatti
sin ϑ_k umocisa-

: due

le M di erié dei tra-

vi anari di zione

iesso,

. Una

magmente . visto

ei due demoria, si definiio può babiliegione ada rii parte

[4.2]

do che \bar{s}_{l+1} (o na con

 $i \bar{s}_i da$

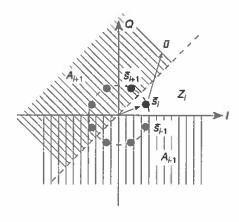


Figura 4.22 - Calcolo della probabilità di errore nei sistemi a modulazione di fase.

 \bar{s}_{i+1} (o da \bar{s}_{i-1}), si ha

$$P(\overline{u} \in A_{i+1}) = P(\overline{u} \in A_{i-1}) = P\left(n_i > \frac{d}{2}\right) = Q\left(\frac{d}{2\sigma_n}\right)$$

come si è già visto per la probabilità d'errore in trasmissione binaria. Se E_s è l'energia di segnale e quindi $\sqrt{E_s}$ è il modulo dei vettori rappresentativi dei segnali, si ha

$$d=2\sqrt{E_s}\sin\frac{\pi}{M}$$
; $\sigma_n^2=\frac{N_0}{2}$

e la [4.2] diventa

$$P(E) \le 2Q\left(\sqrt{\frac{2E_s}{N_0}}\sin\frac{\pi}{M}\right)$$

Se confrontiamo la costellazione circolare corrispondente alla modulazione di fase con la costellazione QAM, osserviamo che a pari potenza di picco impiegata, cioè a pari ampiezza massima dei vettori rappresentativi dei segnali, e a pari M, la distanza minima tra i punti tende a diventare per la modulazione di fase più piccola rispetto al QAM, al crescere di M. La modulazione di fase è pertanto normalmente usata per valori di M non superiori a 8. Un aspetto vantaggioso per la modulazione di fase in alcuni sistemi è l'ampiezza costante; ciò rende la modulazione più robusta rispetto ad altre in presenza di nonlinearità di ampiezza nei circuiti, e di incertezze in ricezione sui livelli di riferimento in ampiezza dei segnali ricevuti.

ESERCIZI

 Si confrontino le prestazioni in presenza di rumore gaussiano bianco dei sistemi di modulazione 16QAM e 16PSK, a pari potenza di picco. — Per un sistema di modulazione MPSK, si consideri il caso in cui, nella ricostruzione della portante per la demodulazione coerente, si commetta un errore di fase ϵ . Si determini il peggioramento della probabilità d'errore causato dall'errore ϵ .

NOTA

Può accadere che la portante ricostruita in ricezione per la demodulazione coerente presenti ambiguità, per esempio ambiguità di segno (180°) nel caso binario in cui si usi un quadratore per la sincronizzazione di portante (par.4.4). In tal caso la portante viene ricostruita a meno del segno e quindi può aversi in ricezione una inversione dei bit d'informazione. Si può ovviare a tale problema, oltre che con sistemi diversi di sincronizzazione, anche ricorrendo ad una precodifica differenziale dei bit in trasmissione (fig. 4.23). La sequenza di dati binari a_k viene trasformata (codifica differenziale) nella sequenza binaria $b_k = a_k \oplus b_{k-1}$ che viene poi trasmessa con ordinaria modulazione di fase binaria. La sequenza b_k ricostruita in ricezione può essere o no invertita a seconda del segno con cui è stata ricostruita la portante, ma ciò non ha conseguenze sulla ricostruzione della sequenza originaria a_k dato che la decodifica differenziale è insensibile all'inversione di b_k ($a_k = b_k \oplus b_{k-1}$). Si noti che se il bit demodulato b_k è errato, si ha un doppio errore sui bit d'informazione ricostruiti poiché risulteranno errati sia a_k che a_{k+1} . La probabilità d'errore è raddoppiata.

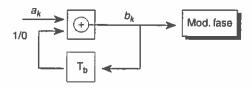


Figura 4.23 - Modulazione differenziale di fase. Precodifica dei dati.

4.3.5 Modulazione differenziale di fase. Demodulazione incoerente

Si è detto della necessità, per la demodulazione di fase coerente, della sincronizzazione per la fase di riferimento della portante. La sincronizzazione dovrà essere tanto più precisa quanto maggiore è il valore di M. Se si commette un errore di fase nella ricostruzione della portante, ciò equivale ad una rotazione di pari valore della costellazione dei segnali rispetto agli assi di riferimento, con conseguente diminuzione della distanza tra il punto trasmesso ed i limiti della regione di corretta decisione. Volendo evitare la complessità di tale sincronizzazione, si può ricorrere alla modulazione differenziale di fase (DPSK). Tale metodo consiste nel trasferire l'informazione non sulla fase assoluta ϑ_k

dell'impulso ed il to assoluto c Consideri viene rappres bolo 0 corrist una precodifi cedente para;



Figura 4.24 - Den

Il ricevitore coerente, in cu anziché la port filtrato per rid tardato del ten nente utile di s

(per semplicità ma d'onda mo uscita è quindi basta un circuit nente di rumor della demodula to è affetto da aggiuntivi, in p quindi necessar

ı cui, te, si della

one

erenrio in
so la
na insistele dei
codinessa
zione
ante,
dato
,). Si
orma-

ore è

tte
delcroclore
tandei
ione
retta
e, si
mea ϑk

dell'impulso sinusoidale ma sul cambiamento di fase $\vartheta_k - \vartheta_{k-1}$ tra un impulso ed il successivo. In tal modo non è più necessario un riferimento assoluto di fase.

Consideriamo anzitutto per semplicità il caso binario: il simbolo 1 viene rappresentato con un cambiamento di fase di 180° mentre al simbolo 0 corrisponde un cambiamento di 0°. Ciò può essere ottenuto con una precodifica differenziale dei dati (fig. 4.23), già introdotta nel precedente paragrafo.

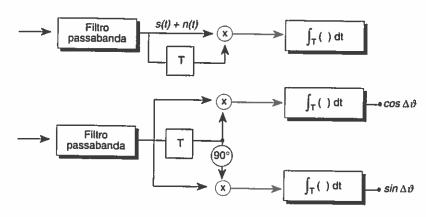


Figura 4.24 + Demodulazione differenziale di fase.

Il ricevitore può essere realizzato con un correlatore, come nel caso coerente, in cui però si utilizza come riferimento l'impulso precedente anziché la portante coerente ricostruita (fig. 4.24a). Il segnale ricevuto, filtrato per ridurre il rumore, viene moltiplicato per il segnale stesso ritardato del tempo T di simbolo; il correlatore dà in uscita una componente utile di segnale proporzionale a $\cos \Delta \vartheta (\Delta \vartheta = \vartheta_k - \vartheta_{k-1})$ poiché

$$\int_{T} \cos \omega_0 t \cdot \cos(\omega_0 t + \Delta \vartheta) dt = C \cos \Delta \vartheta$$

(per semplicità assumiamo qui che $\omega_0 T$ sia multiplo di 2π e che la forma d'onda modulante impulsiva g(t) sia rettangolare). Il segnale in uscita è quindi positivo o negativo a seconda che $\Delta\vartheta$ sia 0 o π , per cui basta un circuito a soglia per la decisione sul bit trasmesso. La componente di rumore in uscita dal correlatore è maggiore rispetto al caso della demodulazione coerente poiché il segnale (ritardato) di riferimento è affetto da rumore, e quindi si hanno in uscita termini di rumore aggiuntivi, in particolare il termine dovuto al prodotto n(t)n(t-T). È quindi necessario far precedere il correlatore da un filtro passabanda

per limitare il rumore n(t) che entra in gioco nella correlazione. Ciò non è necessario nella demodulazione coerente, che, come si è visto, è una operazione lineare di traslazione sull'asse delle frequenze: in questo caso il filtraggio può essere eseguito indifferentemente in banda passante o in banda base.

Il metodo può essere applicato al caso generale di M simboli; per M=4, ai quattro simboli, per esempio coppie di bit, sono associati cambiamenti di fase tra un impulso e il successivo pari a $\Delta\vartheta=\pm\frac{\pi}{4},\pm\frac{3\pi}{4}$. In questo caso la demodulazione può essere realizzata con due correlatori in quadratura (fig. 4.24b): le due forme d'onda di riferimento sono ricavate rispettivamente dall'impulso precedente ritardato del tempo di simbolo e dallo stesso impulso sfasato di $\frac{\pi}{2}$. Le componenti di segnale all'uscita dei due correlatori sono rispettivamente proporzionali a $\cos\Delta\vartheta$ e $\sin\Delta\vartheta$, e da esse si può ricavare il valore di $\Delta\vartheta$. Nel caso M=4 con $\Delta\vartheta=\pm\frac{\pi}{4},\pm\frac{3\pi}{4}$ basta osservare il segno delle due uscite $\cos\Delta\vartheta$ e $\sin\Delta\vartheta$ per identificare il valore di $\Delta\vartheta$ tra i 4 possibili. La decisione sul simbolo trasmesso può quindi essere presa tramite circuiti a soglia all'uscita dei due correlatori.

La struttura del demodulatore è quindi simile a quella del demodulatore in quadratura visto per la demodulazione coerente di fase e per il QAM, salvo la sostituzione, come riferimento, della portante ricostruita con l'impulso ricevuto precedente. Si può anche usare una struttura equivalente basata su un demodulatore in quadratura con un riferimento costituito da una portante con una fase β qualsiasi arbitraria. I valori della fase $\vartheta_k + \beta$ degli impulsi trasmessi sono così ricavati rispetto ad un riferimento arbitrario, ma ciò che porta l'informazione è la differenza tra le fasi di impulsi successivi. Quindi se si stima la fase di ciascun impulso rispetto al riferimento arbitrario e si calcola poi la differenza tra fasi successive, si ottiene il valore $\partial_k - \partial_{k-1}$, che è indipendente dalla fase arbitraria della portante di riferimento. Tra gli M possibili valori per $\Delta \vartheta$, si sceglierà quello più vicino al valore misurato $\hat{\vartheta}_k - \hat{\vartheta}_{k-1}$. La probabilità di errore è maggiore rispetto alla demodulazione coerente: infatti nel caso coerente essa dipende solo dall'errore presente nella fase ϑ_k del vettore ricevuto, mentre nel caso differenziale dipende da due contributi d'errore, in ϑ_k e in ϑ_{k-1} . Questo è il prezzo che si paga per l'eliminazione della necessità della sincronizzazione di portante.

4.3.6 Mod Nella i oscillatore da una fre so più sen si sinusoio

 $e f_2 = f_0 + .$

 $S_i(t)$

(FSK, free

Figura 4.25

Nel cas $\pm (M-1)$. dulazione o M frequen simbolo al mune reali; fase, e ciò cui frequen gresso. Si t controllato

{a_k} rappre
te in freque
Lo spett
calcolare cc
durata T de
La demc
di principio

e. Ciò

risto, è

n que-

banda

i; per

ti cam-

orrela-

o sono npo di

egnale

nali a M = 4 $\cos \Delta \vartheta$ sisione soglia

modu-; e per

e-rico-

strut-

n rife-

traria.

'ati ri-

ione è

la fase

poi la

indi-

gli M

surato odula-

errore

feren-

to è il

nizza-

4.3.6 Modulazione di frequenza

Nella modulazione di frequenza l'onda sinusoidale generata da un oscillatore, la cui frequenza è controllata dal segnale modulante, passa da una frequenza all'altra a seconda del simbolo da trasmettere. Nel caso più semplice di dati binari, ai simboli 1 e 0 corrisponderanno impulsi sinusoidali di durata T e di frequenza rispettivamente $f_1 = f_0 - \frac{\delta f}{2}$

e $f_2 = f_0 + \frac{\delta f}{2}$ essendo f_0 la frequenza centrale:

$$S_i(t) = A \cos \left[2\pi \left(f_0 + a_i - \frac{\delta f}{2} \right) t + \phi_i \right] \qquad a_i = \pm 1 \qquad 0 \le t \le T \quad [4.3]$$

(FSK, frequency shift keying) (fig. 4.25)

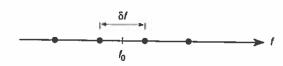


Figura 4.25 - Modulazione di frequenza numerica (FSK).

Nel caso di M simboli (MFSK), si considererà $a_i = \pm 1, \pm 3,..., \pm (M-1)$. Questa schematizzazione corrisponde a considerare la modulazione come una commutazione tra M oscillatori, generatori delle M frequenze. Ciò comporta discontinuità di fase nel passaggio da un simbolo al successivo, il che può avere aspetti negativi. È quindi più comune realizzare la modulazione di frequenza mantenendo continuità di fase, e ciò si ottiene automaticamente se si usa un unico oscillatore la cui frequenza sia controllabile in maniera continua dal segnale d'ingresso. Si tratta in sostanza di un modulatore di frequenza analogico controllato da un segnale modulante numerico in banda base m(t):

$$m(t) = \sum_{k} a_{k} g(t - kT)$$

$$s(t) = \cos[\omega_{0}t + K] m(\tau) d\tau$$
[4.4]

 $\{a_k\}$ rappresenta la sequenza di dati, e g(t) è l'impulso base modulante in frequenza.

Lo spettro occupato da questo segnale, che non è così semplice da calcolare come nella modulazione di ampiezza o di fase, dipende dalla durata T degli impulsi e dalla deviazione di frequenza δf .

La demodulazione può essere di tipo coerente o incoerente. In linea di principio la demodulazione ottimale può essere realizzata seguendo

La trasmi

il principio già discusso (Cap. 2) dell'uso di M correlatori (ovvero filtri adattati) per le M forme d'onda possibili in trasmissione. A rigore ciò è vero per il caso [4.3] in cui si ha in generale discontinuità di fase tra simboli successivi, e questi sono indipendenti tra loro. Nel caso [4.4] invece c'è dipendenza tra simboli successivi poiché la fase dell'onda ad un certo istante dipende da tutti i dati a_k precedenti. Quindi la demodulazione ottimale richiede che la decisione sul simbolo si basi sull'osservazione non solo del singolo simbolo ma dell'intera sequenza di simboli ricevuta.

Nel primo caso quindi il ricevitore può essere costituito da una batteria di M filtri adattati agli M impulsi sinusoidali, ciascuno definito da una ben precisa frequenza e fase; assumendo eguale energia e probabilità a priori per i vari segnali, il ricevitore sceglierà come segnale tramesso più probabile quello corrispondente alla massima uscita tra gli M correlatori (fig. 4.26). Si osservi che l'uso di questa configurazione ottimale non è frequente in quanto richiede in ricezione coerenza perfetta, in frequenza e fase, per tutte le M frequenze. Le prestazioni (tasso d'errore e banda occupata) dipendono dalla scelta del distanziamento δf delle frequenze. Un valore interessante è dato da $\delta f = \frac{1}{T}$; in questo caso infatti i segnali $s_i(t)$ risultano ortogonali, qualsiasi siano i valori delle fasi ϕ_i . L'ortogonalità dei segnali fa sì che, avendo travelori delle fasi ϕ_i . L'ortogonalità dei segnali fa sì che, avendo travelori delle fasi ϕ_i . L'ortogonalità dei segnali fa sì che, avendo travelori delle fasi ϕ_i . L'ortogonalità dei segnale all'uscita degli M corso il segnale $s_i(t)$, le componenti di segnale all'uscita degli M cor-

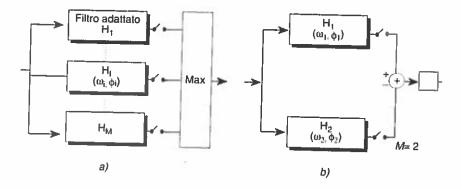


Figura 4.26 - Ricevitore coerente MFSK.

relatori sianc *i*-esimo che v cedenza); le c mico,

sono variabili sto. Se si usa stimate usand

4.3.8)

P(Ebin) rappres cola consideral 4.26b) è pari a componenti di sultano incorre della somma è | Si ha:

e quindi, rispett segnale, un din

> ES Si

gna PS

Il sistema MF5 mini di tasso d'e distinguibilità de consumo di band dine di M/T).

Si noti che si $\delta f = \frac{1}{2T}$, e quinc ϕ_i non possono el le perché si abbia

La in realiz

² Perché la correlazione $\int s_i s_j dt$, $i \neq j$, sia nulla, a rigore è necessario anche che le frequenze siano multiple di $\frac{1}{T}$. Se ciò non è, il valore della correlazione risulta trascurabile se la frequenza in gioco è molto maggiore di $\frac{1}{T}$, come di solito avviene.

o filtri e ciò è a siminvece

ne ot-

ie non vuta.

a batfinito

e pro-

ita tra razio-

erenza

azioni

anzia-

 $\frac{1}{T}$; in

iano i

trava-

ısmes-

1 cor-

2

le fre-

relatori siano tutte nulle, tranne quella corrispondente al correlatore *i*-esimo che vale $\sqrt{E_s}$ (secondo la normalizzazione già adottata in precedenza); le componenti di rumore, per rumore di canale di tipo termico.

sono variabili gaussiane *indipendenti* con varianza $\frac{N_0}{2}$, come già visto. Se si usa un demodulatore coerente, le prestazioni possono essere stimate usando l'approssimazione (che verrà meglio giustificata al par. 4.3.8)

$$P(E) \le (M-1)P(E_{bin})$$

 $P(E_{bin})$ rappresenta la probabilità d'errore del caso binario, che si calcola considerando che l'uscita di segnale del demodulatore binario (fig. 4.26b) è pari a $\pm \sqrt{E_r}$ mentre il rumore ha varianza $\sigma^2 = N_0$. Infatti le componenti di rumore all'uscita dei due filtri adattati o correlatori risultano incorrelate, essendo i segnali ortogonali, e quindi la varianza della somma è pari alla somma delle varianze delle singole componenti. Si ha:

$$P(E_{bin}) = Q\left(\sqrt{\frac{E_s}{N_0}}\right)$$

e quindi, rispetto alla trasmissione binaria antipodale a pari potenza di segnale, un dimezzamento del rapporto segnale/rumore in uscita.

ESERCIZIO

Si illustrino le prestazioni del sistema FSK binario con $\delta f = \frac{1}{T}$ utilizzando la rappresentazione geometrica (si usino i due impulsi a frequenze f_1 e f_2 come base ortogonale; la distanza d tra i punti di segnale risulta ridotta di un fattore $1/\sqrt{2}$ rispetto al sistema binario PSK a pari energia di segnale). Si veda anche il par. 4.3.7.

Il sistema MFSK ortogonale considerato ha ottime prestazioni in termini di tasso d'errore, poiché ortogonalità dei segnali significa ottima distinguibilità degli stessi in presenza di rumore. D'altra parte però il consumo di banda è notevole, crescendo linearmente con M (è dell'ordine di M/T).

Si noti che si può avere ortogonalità tra i segnali anche scegliendo $\delta f = \frac{1}{2T}$, e quindi occupando metà banda; in questo caso però le fasi ϕ_i non possono essere qualsiasi ma devono essere nella [4.3] tutte nulle perché si abbia ortogonalità.

La in realta non è così: per avere ortogonalità le fasi devono essere tutte uguali

La trasi

Demodulazione incoerente

Per i sistemi MFSK, è possibile anche una demodulazione non coerente, spesso più conveniente perché non richiede in ricezione la conoscenza delle fasi delle portanti, e presenta in molti casi prestazioni solo leggermente inferiori al caso coerente. Con riferimento allo schema del demodulatore coerente ottimale, se si pone all'uscita di ciascun filtro adattato un rivelatore ad inviluppo ci si svincola dalla necessità di conoscere esattamente le fasi ϕ_i (fig. 4.27). Valgono in questo caso considerazioni già fatte per la demodulazione incoerente in modulazione di ampiezza (si veda il par. 4.3.3). L'uscita del rivelatore corrispondente all'impulso trasmesso avrà distribuzione di Rice (Appendice B) mentre tutte le altre hanno solo la componente di rumore, se i segnali sono ortogonali, e quindi hanno distribuzione di Rayleigh. Se consideriamo per esempio il caso binario e indichiamo con u_1 , u_2 le uscite dei rivelatori corrispondenti alle due frequenze, si ha per la probabilità d'errore (supponiamo di avere trasmesso per esempio il segnale s_1 , avendosi per la probabilità d'errore $P(E/s_1) = P(E/s_2) = P(E)$,

 $P(E/s_1) = P(u_2 > u_1/s_1) = \int_0^\infty p(u_1/s_1) \left[\int_{u_1}^\infty p(u_2/s_1) du_2 \right] du_1.$ [4.5]

 $p(u_1/s_1)$ è la funzione di Rice, mentre $p(u_2/s_1)$ è la funzione di Rayleigh. In questo caso la [4.5] dà il risultato

$$P(E) = \frac{1}{2} e^{-\frac{E_s}{2N_0}}.$$

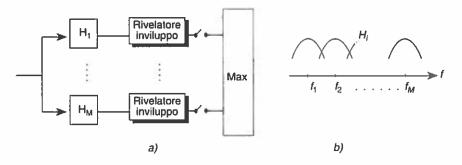


Figura 4.27 - Demodulazione MFSK incoerente.

Confrontando con la demodulazione coerente e utilizzando l'approssimazione [2.5] per la funzione Q, si verifica facilmente che la demodulazione incoerente ha prestazioni quasi eguali alla demodulazione coerente, ai livelli di rapporto segnale/rumore abbastanza alti che normalmente si incontrano.

Passan è evidente di frequenti la segnale a_k . Que rispetto all rale ottima

4.3.6.1 Mg

La moc presenta p fica. Si ha $s(t) = \cos[$

> π 2 π 2

Figura 4.28 - .

³ MSK: minit shift keying). 2 11 0

:oenosolo del ltro coonone

ono amo 'elarore

[4.5]

dosi

-

l'api deione

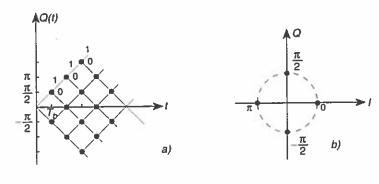
nor-

Passando a considerare la modulazione a fase continua del tipo [4.4], è evidentemente possibile demodulare con un ordinario discriminatore di frequenza come già visto nei sistemi analogici (par. 4.2.3): si ha in uscita il segnale modulante in banda base m(t) da cui si ricava l'informazione $\{a_k\}$. Questo criterio, usato quando si voglia privilegiare la semplicità rispetto alle prestazioni in termini di tasso d'errore, non è però in generale ottimale. Un esempio è dato dal sistema MSK illustrato nel seguito.

4.3.6.1 Modulazione di frequenza binaria con $\delta f = \frac{1}{2T_b}$ (MSK)³

La modulazione di frequenza binaria con $\delta f = \frac{1}{2T_b}$ e fase continua presenta proprietà particolari che ne giustificano una trattazione specifica. Si ha:

$$s(t) = \cos[\omega_0 t + \varphi(t)] = \cos\left[\omega_0 t + \frac{\pi}{2T_b} \int m(\tau) d\tau\right]; \qquad m(t) = \sum_k a_k w(t - kT_b)$$



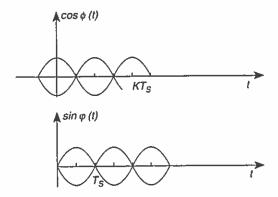


Figura 4.28 - Andamento della fase nei sistemi MSK.

³ MSK: minimum shift keying, a volte indicato anche come FFSK (fast frequency shift keying).

La to

w è l'impulso rettangolare di ampiezza unitaria e durata T_b , $a_k = \pm 1$ rappresenta l'informazione. La frequenza del generico impulso è $f_0 + a_k \frac{1}{4T_b}$, e la fase $\varphi(t) = \frac{\pi}{2T_b} \int_0^t m(\tau) d\tau$, in ogni intervallo di bit, varia

linearmente con un incremento in tale intervallo pari a $\pm \frac{\pi}{2}$ (fig. 4.28a). Il sistema raggiunge cioè, al termine di ogni intervallo di bit, una tra quattro possibili fasi $\left(0, \pm \frac{\pi}{2}, \pi\right)$ (fig. 4.28b), a seconda del bit e della fase iniziale.

Per la demodulazione, si potrebbe applicare quanto detto in generale per la demodulazione FSK, e cioè usare filtri adattati alle due frequenze, oppure anche un discriminatore seguito da un rivelatore in banda base. Ma si può far di meglio poiché il segnale MSK può essere interpretato come una modulazione di ampiezza in quadratura. Quindi si può ricorrere sia in modulazione che in demodulazione ai metodi già visti per i sistemi QAM, ottenendo così anche analoghe prestazioni. Infatti si esprima l'onda modulata in funzione delle sue componenti in quadratura rispetto alla frequenza centrale f_0 :

$$s(t) = \cos\varphi(t)\cos\omega_0 t - \sin\varphi(t)\sin\omega_0 t$$

Dall'andamento della fase $\varphi(t)$ (fig. 4.28) si vede che in ogni intervallo successivo di lunghezza $T_s = 2T_b$, compreso tra $kT_s - \frac{T_s}{2}$ e $kT_s + \frac{T_s}{2}$, si ha

$$\cos\varphi(t) = \cos\varphi(kT_s)\cos\frac{\pi t}{2T_b} = \pm\cos\frac{\pi t}{2T_b}$$
; $\varphi(kT_s) = 0$ oppure π ,

dove il segno dipende dal valore di $\varphi(kT_s)$ e quindi dalla storia dei dati precedenti. Analogamente si ha per gli intervalli successivi $(kT_s \leftrightarrow (k+1)T_s)$:

$$\sin\varphi(t) = \sin\left(kT_s + \frac{T_s}{2}\right) \sin\frac{\pi t}{2T_b} = \pm \sin\frac{\pi t}{2T_b}; \qquad \varphi\left(kT_s + \frac{T_s}{2}\right) = \pm \frac{\pi}{2}$$

In definitiva il segnale può essere espresso come

$$S(t) = \sum_{k} c_k p(t - kT_s) \cdot \cos\omega_0 t + \sum_{k} d_k p \left(t - kT_s - \frac{T_s}{2} \right) \cdot \sin\omega_0 t \quad [4.6]$$

dove l'impulso p(t) è dato da

$$p(t) = \cos\frac{\pi t}{2T_b}; -T_b \le t \le T_b$$

Figura 4.2!

 $\varphi(kT_s)$ e

Le v.

Siamc in quadra

— l'impu

— i due:

Con d portante c_k, d_k e qu dividua le d'informa coppia d_k , stema $4Q_k$ identiche :

Quanto è definita c d'energia | 'ia

ġ.

it, oit

rarein re

ıdi già n-

in

er-

ati

ivi

.6]

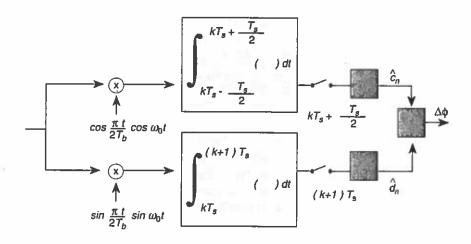


Figura 4.29 - Demodulazione d'ampiezza in quadratura per i sistemi MSK.

Le variabili binarie c_k, d_k dipendono rispettivamente dalle fasi $\varphi(kT_s)$ e $\varphi(kT_s + \frac{T_s}{2})$.

Siamo dunque in presenza di un sistema a modulazione di ampiezza in quadratura con le seguenti caratteristiche specifiche:

- l'impulso modulante in banda base p(t) ha forma sinusoidale
- i due flussi in quadratura sono sfalsati nel tempo di $\frac{T_s}{2} = T_b$.

Con due demodulatori d'ampiezza in quadratura, coerenti con la portante ω_0 e opportunamente sincronizzati, si ricuperano le sequenze c_k, d_k e quindi l'informazione trasmessa (fig. 4.29). La coppia c_k, d_k individua le fasi $\varphi(kT_s)$ e $\varphi(kT_s+T_b)$, e consente quindi di ricavare il bit d'informazione corrispondente all'intervallo di bit (kT_s, kT_s+T_b) ; la coppia d_k , c_{k+1} individua il bit successivo, e così via. Si tratta di un sistema 4QAM, e le prestazioni in termini di tasso d'errore sono quindi identiche a quelle calcolate per il QAM o il 4PSK.

Quanto alla distribuzione spettrale di potenza del segnale [4.6], essa è definita dalla forma d'onda p(t), ed è quindi proporzionale allo spettro d'energia $|P(f)|^2$ di p(t) (trasportato ovviamente sulla frequenza ω_0):

$$|P(f)|^2 = \left[\frac{\cos 2\pi f T_b}{I - (4fT_b)^2}\right]^2$$

ESERCIZIO

Si confronti lo spettro con quello del sistema 4PSK, sistema che, come si è visto, ha le stesse prestazioni del sistema MSK per quanto riguarda la probabilità d'errore.

Il sistema MSK è di efficienza analoga al 4PSK e presenta uno spettro compatto. Si possono derivare da esso altri sistemi con spettro ancora più compatto se si fa in modo che le traiettorie di fase $\varphi(t)$ (fig. 4.28) presentino andamento nel tempo ancor più graduale. Per ottenere ciò si può agire sulla forma degli impulsi modulanti. Se per esempio la sequenza di impulsi rettangolari che rappresenta i dati viene prefiltrata, prima del modulatore di frequenza, con un filtro passabasso di forma gaussiana, si ottiene un sistema, indicato come GMSK (Gaussian MSK), in cui i lobi laterali dello spettro sono a livelli estremamente ridotti. La maggior compattezza di spettro è qui ottenuta con un allargamento nel tempo degli impulsi modulanti e quindi con una conseguente interferenza intersimbolica. Di ciò occorre tener conto nella demodulazione.

4.3.7 Rappresentazione geometrica. Spazio dei segnali

Si è visto, nell'analisi della modulazione di ampiezza e di fase, come la rappresentazione geometrica o vettoriale nel piano I, Q renda più chiara e semplice la trattazione del problema della ricezione in presenza di rumore. Lo strumento della rappresentazione geometrica può essere utilmente applicato in generale. Come i segnali di un sistema QAM possono essere rappresentati come combinazioni lineari di due funzioni "ortogonali" $g(t)\cos\omega_0 t$, $g(t)\sin\omega_0 t$ e quindi come vettori nello spazio bidimensionale I, Q, così ogniqualvolta i segnali in gioco possano essere rappresentati come combinazioni lineari di N funzioni ortogonali di base, potremo rappresentarli come vettori in uno spazio a N dimensioni. Diciamo ortogonali due segnali $x_1(t)$, $x_2(t)$ se la loro correlazione temporale $\int x_1x_2dt$ è nulla. Se il generico segnale s(t) può essere espresso come

$$s(t) = a_1 x_1(t) + a_2 x_2(t) + \dots + a_N x_N(t)$$
 [4.7]

essendo $\{x_i(t)\}$ una famiglia di funzioni ortogonali, possiamo usare queste come funzioni base di riferimento cioè come assi cartesiani dello spazio rappresentativo dei segnali. Per comodità si assume che le funzioni base siano normalizzate ad energia unitaria. In questo spazio i segnali possono essere rappresentati da vettori con componenti date dai coefficienti della combinazione lineare: $\bar{s} = (a_1, a_2, ..., a_N)$. Per l'ortogonalità delle funzioni assunte come base di riferimento, si ha, come si verifica facilmente per la [4.7], che le componenti a_i del vettore \bar{s} si ricavano mediante correlazione di s(t) con la rispettiva funzione base $x_i(t)$:

$$(s(t)x_i(t)dt = a_i$$

Inoltre si ha che l'energia E_s del segnale è pari al quadrato del modulo del vettore relativo:

$$E_s = \int s^2(t)dt = a_1^2 + a_2^2 + ... + a_N^2 = |\bar{s}|^2.$$

In rice; spetto agli lative funz ponenti di risultano s

Di cons geometrica tori, region nalisi del s denza per i un'analisi la quale rir to nel para

Per il ca numerico, non si hant ne'' rappro d'errore pu le quindi t prestazioni bidimensio Se facciam \bar{s}_j , le due re niti dalla m del vettore della distar

 $P[E(i \rightarrow j)] = \frac{N_0}{2}, \text{ come}$

di M segnal il generico sun evento e li, quando cioè \overline{u} cade nione degli pie $(s_i, s_j), j$ si ha per la

on e

pet-

1CO-

.28)

.ò si

nza

del

ına,

i lo-

gior

de-

ter-

me

iara

ru-

ıtil-

onc

IZO-

ıen-

ore-

po-

cia-

rale

1.7

lue-

paoni

osenti

to-

e si

s si

(t):

no-

In ricezione si possono ricavare le coordinate del vettore ricevuto rispetto agli assi $\{x_i\}$ mediante correlazione del segnale ricevuto con le relative funzioni base; oltre alle componenti di segnale, vi sono anche componenti di rumore che, nel caso di rumore di canale gaussiano bianco, risultano statisticamente indipendenti (lasciamo la verifica al lettore).

Di conseguenza si potranno utilizzare concetti e proprietà di natura geometrica (modulo dei vettori, distanza tra i punti, ortogonalità di vettori, regioni di decisione, demodulazione a minima distanza ecc.) per l'analisi del sistema di trasmissione, esattamente come già fatto in precedenza per i sistemi QAM. La rappresentazione geometrica consente quindi un'analisi unificata dei più svariati sistemi di trasmissione numerici, per la quale rinviamo a trattazioni specifiche (un semplice esempio è riportato nel paragrafo precedente per il sistema FSK binario ortogonale).

4.3.8 Valutazione approssimata della probabilità d'errore

Per il calcolo della probabilità d'errore in un sistema di trasmissione numerico, si è visto che in qualche caso, per esempio nei sistemi QAM, non si hanno difficoltà data la particolare semplicità della "costellazione" rappresentativa dei vari segnali. Ma il calcolo della probabilità d'errore può diventare complesso per una costellazione generica. È utile quindi un metodo semplice per la valutazione approssimata delle prestazioni. Consideriamo come esempio una qualsiasi costellazione bidimensionale (si veda la fig. 4.22), ma quanto segue vale in generale. Se facciamo riferimento ad un sistema con due soli segnali (punti) si, \bar{s}_j , le due regioni di decisione sono evidentemente i due semipiani definiti dalla mezzeria tra i due punti, e si ha errore quando la componente del vettore rumore lungo la congiungente i due punti supera la metà della distanza d_y tra i punti. Ciò avviene, come noto, con probabilità $P[E(i \rightarrow j)] = Q(\frac{d_{ij}}{2\sigma_n})$; la varianza σ_n^2 della componente di rumore vale $\frac{N_0}{2}$, come la varianza delle componenti lungo gli assi I, Q. Nel caso di M segnali possibili in trasmissione, immaginiamo di avere trasmesso il generico segnale \bar{s}_i , e ricevuto il segnale \bar{u} . Avendo trasmesso \bar{s}_i si ha un evento errore E_j , nella scelta in ricezione tra gli M segnali possibili, quando uno o più tra i vettori \bar{s}_i , $j \neq i$, è più vicino a \bar{u} che non \bar{s}_i , cioè \overline{u} cade esternamente alla regione Z_i . L'evento errore E_i è cioè l'unione degli eventi errore $E_{ij} = E(i \rightarrow j)$ relativi alle M-1 possibili coppie $(s_i, s_j), j \neq i$. Poiché tali eventi E_{ij} chiaramente non sono disgiunti,

$$P(E_i) \leq \sum_{j} P(E_{ij})$$

si ha per la probabilità dell'evento errore E_i

Per la probabilità d'errore globale occorrerà poi mediare rispetto alle probabilità a priori dei segnali \bar{s}_i . Si ha così un semplice limite superiore alla probabilità d'errore, limite che si può determinare in base solo alle probabilità d'errore $P(E_{ij})$ di configurazioni binarie. Un esempio di applicazione di questo metodo di stima è già stato illustrato in precedenza per un caso particolare nei sistemi a modulazione di fase.

4.3.9 Confronto tra le prestazioni dei sistemi di modulazione numerici

I sistemi di modulazione numerici sopra presentati hanno gradi di efficienza diversi dal punto di vista dell'utilizzazione della banda di canale e della potenza di trasmissione disponibili. Se facciamo riferimento ad una sorgente che emetta dati binari al ritmo di R bit/s, l'efficienza di un sistema di modulazione nell'utilizzazione della banda può essere misurata dal rapporto R/B dove B rappresenta la banda occupata. La figura 4.30 riporta, per un dato livello di probabilità d'errore (pari a 10^{-5} nel caso illustrato in figura), l'efficienza spettrale R/B e il rapporto segnale-rumore $\frac{E_b}{N_0}$ necessario, per vari sistemi di modulazione,

in presenza di rumore bianco gaussiano di densità (unilaterale) N_0 . Si noti che l'energia disponibile per bit d'informazione $E_b = P_s/R$ non dipende dal sistema di modulazione adottato. È interessante confrontare le prestazioni ottenibili dai sistemi di modulazione pratici con i limiti posti dal teorema di Shannon sulla capacità dei canali di trasmissione. Questi limiti sono abbastanza lontani dalle prestazioni dei sistemi considerati, e l'avvicinarsi ad essi comporta l'uso di codificazioni di canale complesse. Dal teorema di Shannon sulla capacità di canale (si veda il par. 2.6) si ha che la velocità limite raggiungibile su un canale di banda B con rumore gaussiano di densità uniforme N_0 e con potenza di segnale P_s è

$$R \le B \log_2 \left(1 + \frac{P_s}{N_0 B} \right)$$

con probabilità d'errore tendente a zero.

Quindi si ha

$$\frac{E_b}{N_0} \ge \frac{2^{R/B} - 1}{R/B} \tag{4.8}$$

La [4.8] definisce nel piano di coordinate $\left(\frac{R}{B}, \frac{E_b}{N_0}\right)$ la curva limite di Shannon, e quindi la regione di piano corrispondente ai sistemi teoricamente realizzabili. Si noti che nella condizione limite di banda infinita, si ha $\frac{R}{B} \rightarrow 0$ e quindi, dalla [4.8], $\frac{E_b}{N_0} \rightarrow \ln 2 = 0,693 \ (-1.6 \ dB)$.

Figura 4.30 - Ef

Questo rapposu canale di senza errori. ferisce a prol pratici si rife d'errore cons

II quadro dell'efficienz: simboli in tras I sistemi QAN trale; ciò per la "costellazi no più "lonta spesso usati p plice, non sor MSK che ha anziché di un' de la banda o contro il rume ortogonalità t quindi di una rapporto segn in generale se ! rumore neces!

La distanz corrispondent portuni codici

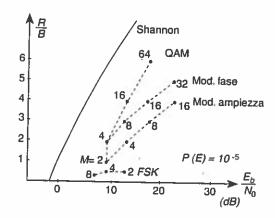


Figura 4.30 - Efficienza spettrale in funzione del rapporto segnale/rumore necessario, per vari sistemi di modulazione, e curva limite di Shannon.

Questo rappresenta il limite minimo di rapporto segnale-rumore che, su canale di banda infinita, consente teoricamente una trasmissione senza errori. Osserviamo inoltre che, mentre la curva di Shannon si riferisce a probabilità d'errore nulla, i punti rappresentativi dei sistemi pratici si riferiscono, a titolo d'esempio, ad un valore di probabilità d'errore considerato in molte applicazioni accettabile.

Il quadro risultante dalla figura mostra anzitutto come un aumento dell'efficienza spettrale comporti tipicamente l'aumento del numero di simboli in trasmissione e l'aumento del rapporto segnale-rumore richiesto. I sistemi QAM sono i più efficienti nella regione ad alta efficienza spettrale; ciò perché, a pari potenza di picco disponibile in trasmissione, la "costellazione" di simboli è migliore rispetto alle altre (i simboli sono più "lontani"). D'altra parte i sistemi a modulazione di frequenza, spesso usati perché ad inviluppo costante e demodulabili in modo semplice, non sono efficienti dal punto di vista spettrale (tranne il sistema MSK che ha caratteristiche particolari); infatti l'uso di più frequenze, anziché di un'unica portante come in QAM e PSK, inevitabilmente espande la banda occupata. Si può però ottenere in FSK un'alta protezione contro il rumore scegliendo per esempio le frequenze in modo da avere ortogonalità tra i segnali, e quindi bassa probabilità di errore. Al costo quindi di una bassa efficienza spettrale, si ottiene un abbassamento del rapporto segnale-rumore richiesto. Come indica la curva di Shannon, in generale se si vogliono ottenere valori più bassi per il rapporto segnalerumore necessario, occorre usare bande più larghe.

La distanza che separa le prestazioni dei sistemi pratici da quelle corrispondenti al limite di Shannon può essere ridotta con l'uso di opportuni codici di canale. Questi possono dare guadagni sul rapporto

segnale-rumore richiesto di alcuni dB, tipicamente fino a circa 6 dB in pratica (si veda il Cap. 6).

NOTA

Nei conti per la fig. 4.30, per la banda si è considerato il caso di riferimento ideale. Se T_s indica la durata di simbolo, si ha $T_s = T_b \log_2 M = \frac{1}{R} \log_2 M$ se codifichiamo i bit della sorgente con M simboli; la banda passante occupata è assunta pari a $B = \frac{1}{T_s}$ per i sistemi QAM e PSK, e pari a $B = \frac{M}{T_s}$ nei sistemi FSK. Per i sistemi FSK si è considerato il caso di segnali ortogonali con distanza tra le frequenze contigue pari a $\delta f = \frac{1}{T_s}$, e per la probabilità d'errore si è fatto uso della formula approssimata (par. 4.3.6).

$$P(E) \cong (M-1)Q(\sqrt{\frac{E_{s}}{N_{0}}}) = (M-1)Q(\sqrt{\frac{E_{b} \log_{2} M}{N_{0}}})$$

Per il caso FSK, si noti che, all'aumentare di M, aumenta in proporzione la banda occupata, ma aumenta anche l'energia del simbolo $E_* = E_b \log_2 M$ e quindi diminuisce la probabilità d'errore.

4.4. CENNI SULLA SINCRONIZZAZIONE

La demodulazione numerica comporta comunque la sincronizzazione di simbolo per l'identificazione in ricezione degli istanti di lettura e di decisione. Inoltre nel caso di demodulazione coerente è anche necessario ricostruire la portante in frequenza e fase. Moltissimi sono i tipi di sincronizzazione usati, a seconda delle particolari applicazioni. A volte si recupera anzitutto il sincronismo di portante, e ciò consente, nei sistemi in banda passante coerenti, di trasferire il segnale in banda base; dopo di che si può ricuperare il sincronismo di simbolo operando sul segnale in banda base. In altri casi si ricupera il sincronismo di simbolo operando sul segnale in banda passante in modo incoerente, prescindendo cioè dal ricupero della portante. In quanto segue si accenna a titolo puramente esemplificativo a possibili metodi di sincronizzazione, rinviando a testi specifici per la trattazione del tema, molto ampio e di importanza cruciale per il buon funzionamento dei demodulatori.

Sincronizzazione di portante

La ricostruzione della portante in ricezione è agevole se nello spettro del segnale ricevuto esiste una riga spettrale alla frequenza portante. L'esempio più semplice è quello della modulazione di ampiezza effettuata mediante moltiplicazione dell'onda portante per un segnale modulante a val tante dal seg traggio a bar fase. I dispos per la demos chiaramente si l'anello verrà munque all'os trale e di inse e fase; l'usciti

In molti ca sente. Da una sione dell'info tante, utile alla to segnale/rur parte ne viene ricostruire la p spetto al semp

Considerian antipodale di a 1,0 sono nella s sarà del tipo \pm Chiaramente so gnale ricevuto i ω_0 . In questo soma. Infatti quadha un segnale $\frac{1}{2}$ za $2\omega_0$ dalla quate desiderata (fix zione di fase quo occorrerà far ric potenza per elim $4\omega_0$.

Osserviamo cl biguità: con rifer



Figura 4.31 - Sincroniz

dB in

ideale. niamo i

pari a

sistemi

ze con-

appros-

banda i dimi-

zazioettura he nesono i izioni. isente, banda rrando li sime, precenna zazioampio latori.

rtante.
reffetle mo-

dulante a valor medio non nullo. In questo caso l'estrazione della portante dal segnale ricevuto può essere fatta mediante un adeguato filtraggio a banda stretta, spesso realizzato con un anello ad aggancio di fase. I dispositivi ad aggancio di fase sono stati descritti al par. 4.2.3 per la demodulazione di onde modulate in frequenza, ma possono chiaramente essere impiegati per il ricupero della portante. Allo scopo l'anello verrà progettato con una banda stretta, ma che consenta comunque all'oscillatore controllato (VCO) di agganciarsi sulla riga spettrale e di inseguire inevitabili piccole fluttuazioni spurie in frequenza e fase; l'uscita dell'oscillatore riprodurrà così l'onda portante.

In molti casi però la riga spettrale sulla frequenza portante ω_0 è assente. Da una parte, ciò è conveniente dal punto di vista della trasmissione dell'informazione poiché la potenza dedicata alla riga sulla portante, utile alla sincronizzazione, è però sprecata agli effetti del rapporto segnale/rumore utilizzabile nella demodulazione dei dati. D'altra parte ne viene anche la necessità di un sistema di sincronizzazione, per ricostruire la portante in frequenza e fase, di maggiore complessità rispetto al semplice filtraggio sopra menzionato.

Consideriamo un caso elementare, quello della modulazione binaria antipodale di ampiezza; se l'impulso modulante è rettangolare e i bit 1,0 sono nella sequenza equiprobabili e indipendenti, l'onda modulata sarà del tipo $\pm\cos\omega_0 t$ e lo spettro di potenza è continuo senza righe. Chiaramente solo operazioni non lineari potranno trasformare il segnale ricevuto in modo tale da generare la desiderata riga spettrale su ω_0 . In questo semplice caso l'uso di un quadratore risolve il problema. Infatti quadrando l'onda modulata si elimina la modulazione e si ha un segnale $\frac{1}{2} + \frac{\cos 2\omega_0 t}{2}$; viene quindi generata una riga a frequen-

za $2\omega_0$ dalla quale per divisione di frequenza si può ricavare la portante desiderata (fig. 4.31). Se invece consideriamo un sistema a modulazione di fase quaternaria (anziché binaria come nel caso precedente) occorrerà far ricorso all'elevazione del segnale modulato alla quarta potenza per eliminare la modulazione e creare una riga sulla frequenza $4\omega_0$.

Osserviamo che la portante recuperata in questo modo presenta ambiguità: con riferimento al caso binario sopra esaminato, è chiaro che

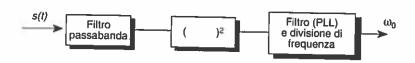


Figura 4.31 - Sincronizzazione di portante per un sistema PSK binario.

con il quadratore la portante viene ricostruita a meno del segno, cioè con una ambiguità di fase pari a π . Uno dei modi per far fronte a questa ambiguità consiste nell'uso della modulazione differenziale, in cui l'informazione sta non nella fase assoluta ma nella variazione di fase da impulso a impulso. In questo caso l'ambiguità di fase diventa irrilevante e può quindi essere accettata.

Sincronizzazione di simbolo

Quando il sincronismo di simbolo viene ricavato dopo conversione in banda base del segnale modulato, evidentemente vale quanto già detto per la trasmissione numerica in banda base (Cap. 2). Il sincronismo di simbolo può però essere ricavato anche direttamente dal segnale passabanda, indipendentemente dal ricupero della portante e dalla conversione in banda base. Considerando per esempio un segnale modulato del tipo $\sum_{k} a_{k}g(t-kT)\cos\omega_{0}t$, basterà prendere l'inviluppo o far passare il segnale attraverso un quadratore per avere, nel segnale risultante, una riga spettrale alla frequenza di simbolo $\frac{1}{T}$. Questa riga spettrale può essere utilizzata per la temporizzazione necessaria alla lettura dei simboli.

Premessa

Per tras corre fond tromagneti zione nello sione radio so onde ele canale fisic una fibra o meni atmos il solo mez sto dal me: uscita. Ad e logica trasn approssima plessa cater tiplex e cen to con una nearità), e. la trasmissi quello per c terà di adatt le trasmissi Nella pri

Nella pri data su linea menti gener sidererà essa caratteristical scrizione so

5.1. LA TR

Tipicame una coppia c