Appunti di Comunicazioni Elettriche

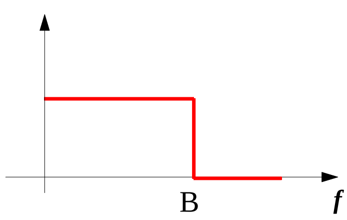
Capitolo 7

Trasmissione numerica - Parte III

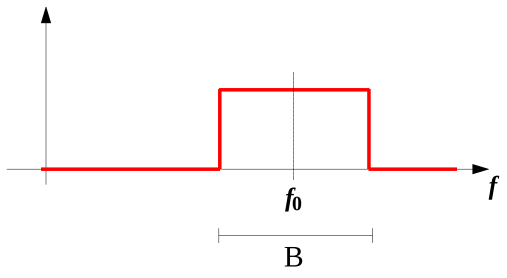
Modulazione di segnali numerici

INTRODUZIONE

Nei discorsi che abbiamo fatto fino ad ora, abbiamo sempre considerato un *sistema di trasmissione numerico* <u>*in banda base*</u>, nel quale cioè il mezzo trasmissivo utilizzato è di tipo passa-basso (come ad esempio il cavo coassiale) e quindi consente la trasmissione dei segnali direttamente nella propria banda (detta appunto *banda base*).

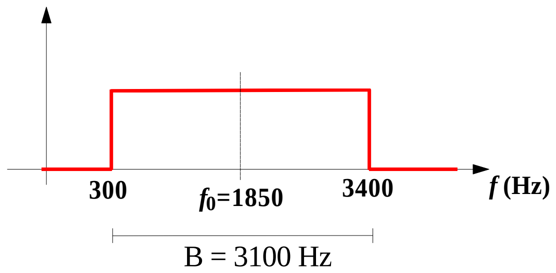


Vediamo adesso cosa succede se il mezzo trasmissivo è passa-banda (come l'atmosfera terrestre usata nei ponti radio terrestri e nei collegamenti via satellite), ossia lascia passare frequenze comprese in un certo intervallo B, simmetrico rispetto ad una *frequenza centrale* f_0:



In questo caso, dobbiamo fare in modo che il segnale da trasmettere abbia tutte le proprie componenti spettrali comprese nell'intervallo $\left[f_0 - \frac{B}{2}, f_0 + \frac{B}{2}\right]$. Per ottenere questo, è necessario effettuare una **modulazione**.

Un tipico esempio di mezzo passa-banda è il **doppino telefonico**, che è progettato in modo da far passare solo le componenti spettrali del *segnale telefonico*, ossia componenti spettrali notoriamente comprese tra 300 Hz e 3400 Hz:

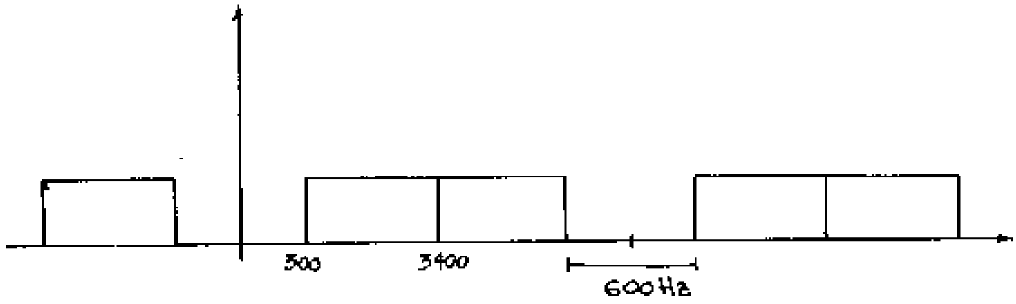


Supponiamo, per semplicità, che il canale sia perfettamente equalizzato (per cui il segnale in uscita ha attenuazione costante in banda), abbia caratteristiche costanti nel tempo e abbia caratteristica di fase rettilinea con la frequenza (il che significa che il *ritardo di gruppo* del canale è rigorosamente costante). Queste ipotesi individuano il cosiddetto **canale telefonico ideale**.

Supponiamo allora di voler effettuare una trasmissione numerica su questo canale, il che significa che intendiamo trasmettere due distinte forme d'onda, corrispondenti ai simboli logici 0 ed 1, con l'obbiettivo che tali forme d'onda, in ricezione, consentano di distinguere, con una accettabile probabilità di errore, i simboli stessi 0 ed 1. In linea di massima, potremmo anche evitare di effettuare una modulazione: dovremmo scegliere le forme d'onda in modo tale che il loro spettro sia compreso nella banda di 3100 Hz centrata sulla frequenza di 1850 Hz. Qui subentra però un problema: sappiamo, infatti, che, all'uscita del filtro di ricezione, abbiamo bisogno di forme ad intersimbolo nullo ed abbiamo visto che godono di questa caratteristica tutte le forme d'onda il cui spettro, ripetuto a passo 1/T (dove 1/T è, in generale, il **baud rate**, ossia la frequenza di trasmissione di forme d'onda in linea[1]), fornisce uno spettro costante. Nel nostro caso, a causa del

[1] Ricordiamo ancora una volta la differenza tra frequenza di cifra e baud rate: la frequenza di cifra quantifica la velocità di trasmissione dei bit, in quanto rappresenta il numero di bit inviati nell'unità di tempo, mentre il baud rate è il numero di forme d'onda inviate nell'unità di tempo. Queste due quantità coincidono per un sistema binario, mentre sono diverse per un sistema multilivello: ad esempio, in un sistema a M=4 livelli, le forme d'onda inviate nell'unità di tempo sono 1/T, per cui il baud rate è 1/T, ma a ciascuna forma d'onda sono associati 2 bit, per cui la frequenza di cifra è 2/T bit/sec, ossia il doppio del baud rate.

gap di 300 Hz che separa lo spettro delle forme d’onda trasmesse dalla frequenza nulla[2], lo spettro periodicizzato non potrà mai fornire una costante, ma darà quanto illustrato nella figura seguente::



Come evidenziato dalla figura, si crea un gap di 600 Hz, tra una replica e l’altra, che può essere riempito solo con la modulazione.

TIPI DI MODULAZIONE NUMERICA

Abbiamo visto che le tipologie di modulazione analogica si dividono nelle seguenti categorie:

- si parla di “**modulazione AM**” (AM=Amplitude Modulation) quando viene fatta variare l’ampiezza della portante c(t) in modo proporzionale al segnale da trasmettere s(t);
- si parla di “**modulazione PM**” (PM=Phase Modulation) quando viene fatta variare la fase di c(t) in modo proporzionale a s(t);
- si parla di “**modulazione FM**” (FM= Frequency Modulation) quando viene fatta variare la frequenza di c(t) in modo proporzionale a s(t).

Le stesse possibilità, ma nomi diversi, si hanno per la modulazione numerica:

- si parla di “**modulazione ASK**” (ASK=Amplitude Shift Keying) quando viene fatta variare l’ampiezza di c(t) in modo proporzionale a s(t);
- si parla di “**modulazione PSK**” (PSK=Phase Shift Keying) quando viene fatta variare la fase di c(t) in modo proporzionale a s(t);
- si parla di “**modulazione FSK**” (FSK=Frequency Shift Keying) quando viene fatta variare la frequenza di c(t) in modo proporzionale a s(t).

Il nostro scopo è descrivere questi tre tipi di modulazione e decidere, se possibile, quale sia il più conveniente in rapporto alle specifiche di progetto.

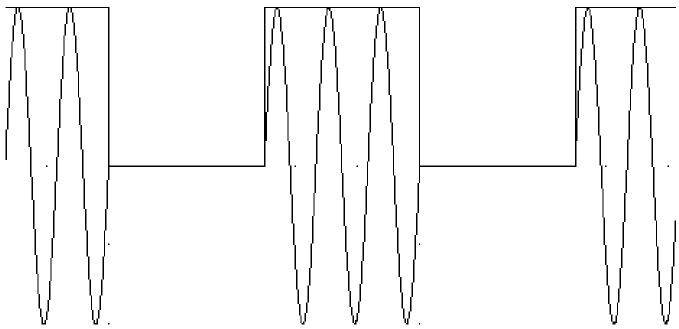
[2] Ricordiamo una proprietà importante: se un filtro lineare riceve in ingresso un segnale con un certo numero di componenti spettrali, l’uscita sarà composta o dalle stesse componenti spettrali in ingresso o solo da alcune di esse, in quanto il filtro, data appunto la sua linearità, non può produrre in uscita delle frequenze diverse da quelle ricevute in ingresso. Di conseguenza, se lo spettro del segnale in ingresso presenta componenti spettrali comprese da 300 Hz e 3400 Hz, nella migliore delle ipotesi anche il segnale di uscita conterrà le stesse componenti; in ogni caso, non ci potranno essere frequenze inferiori a 300 Hz e superiori a 3400 Hz. Questo è il motivo per cui, nel discorso circa la periodicizzazione dello spettro, si considera in uscita dal filtro uno spettro avente la stessa banda di quello in ingresso al filtro stesso.

Segnaliamo inoltre che esiste un particolare tipo di modulazione numerica, che prende il nome di **P.A.M.** (*Pulse Amplitude Modulation*): si tratta sempre di una modulazione numerica di ampiezza, come la ASK citata poco fa, ma con la differenza che il segnale modulato è ancora in banda base.

Modulazione di ampiezza: ASK

SCELTA DEL TIPO DI MODULAZIONE DI AMPIEZZA

Così come nel caso analogico, la modulazione di ampiezza di una portante c(t), ad opera di un segnale numerico s(t), consiste nel far variare l’ampiezza di c(t) in modo proporzionale a s(t). Al contrario, però, del caso analogico, *nel caso numerico il segnale modulante s(t) è la sequenza di rettangoli di opportuna ampiezza.* Quindi, con la ASK noi andiamo a modulare una portante sinusoidale mediante dei rettangoli. Ad esempio, se adottiamo una codifica di tipo *ortogonale*, per cui trasmettiamo un rettangolo di una certa ampiezza per l’1 e niente per lo 0, il segnale modulato, da inviare sul canale di trasmissione, è del tipo seguente:

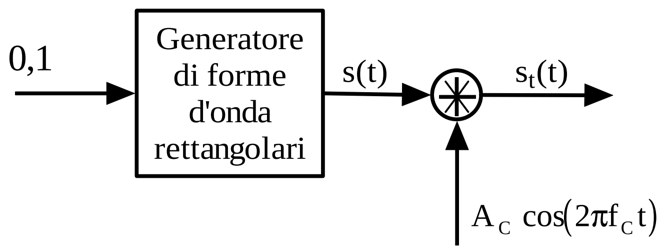


Nella figura è tracciato sia il segnale modulato (costituito da “pezzi” di sinusoide) sia il segnale modulante, cioè la sequenza di rettangoli: la sequenza binaria considerata in questo caso è 10101.

Il modo più semplice che può venire in mente per ottenere un simile segnale modulato è quello di usare un oscillatore che oscilli alla frequenza della portante: si accende l’oscillatore (cioè si accende la sua alimentazione[3]) quando bisogna trasmettere un 1, mentre lo si lascia spento quando bisogna trasmettere uno 0. Tuttavia, è intuitivo capire che questo modo di procedere è sconsigliabile: infatti, un oscillatore è un sistema reazionato costituito anche da elementi dotati di memoria (condensatori e induttori); questi generano dei poli che a loro volta generano dei transitori all’accensione ed allo spegnimento, il che “sporcherebbe” troppo il segnale da trasmettere.

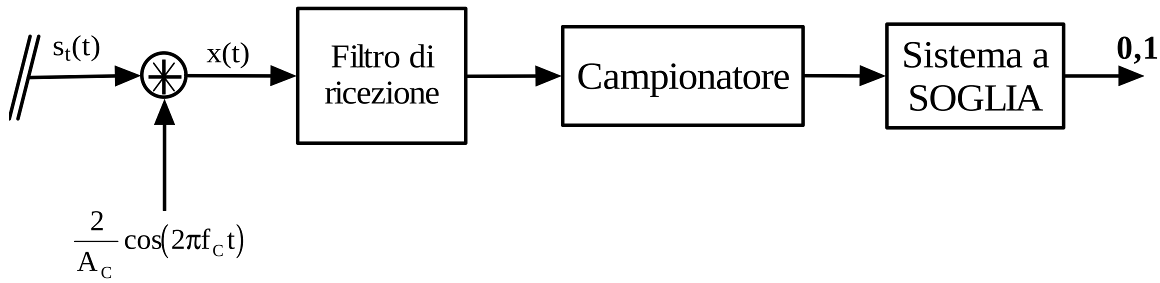
Molto più efficace è invece il metodo seguente: si considera un oscillatore che oscilla a frequenza stabile e poi si usa un moltiplicatore (o **mixer**). Lo schema è dunque quello classico visto anche nella modulazione analogica:

[3] In generale, gli oscillatori sono circuiti progettati in modo che, accendendo semplicemente la tensione di alimentazione, si generi un segnale sinusoidale, alla frequenza prestabilita, che dopo un certo tempo si stabilizza, mantenendo praticamente costante la propria ampiezza. Ci sono, d’altra parte, alcuni oscillatori (come il **ring oscillator**) nei quali l’alimentazione è sempre accesa, ma il funzionamento (cioè la generazione dell’oscillazione) è abilitato da un particolare segnale di **enable**: se l’enable vale 1, allora l’oscillatore oscilla, altrimenti è inerte. L’uso di una linea enable rappresenta il modo migliore per controllare l’oscillazione di un oscillatore.

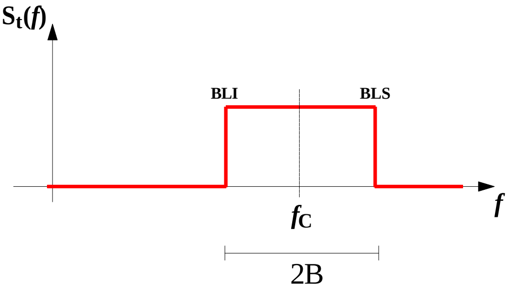


Nel caso in cui sia stata adottata una **codifica ortogonale** per il segnale s(t), è evidente che il prodotto di s(t) per l'oscillazione locale darà zero quando s(t)=0 (0 logico) e darà la portante amplificata quando s(t)=V (1 logico). Si ottiene così la cosiddetta *modulazione del tipo tutto o niente* (**OOK** - *On Off Keying*): a livello concreto, per realizzare il modulatore in questa situazione non serve più il moltiplicatore, ma basta usare banalmente un **interruttore**, che mette a massa l'uscita del circuito quando bisogna trasmettere uno 0 e lascia passare invece il segnale dell'oscillatore quando bisogna trasmettere un 1.

In ricezione, il modo di effettuare la demodulazione è ovviamente identico al caso analogico, per cui si usa, a valle del canale trasmissivo, il classico **demodulatore coerente**:



Questo sistema presenta però uno svantaggio: supponiamo che il segnale modulante s(t) occupi una banda B e corrisponda ad una frequenza di cifra f_S; se trasmettiamo in banda base e scegliamo un roll off δ=0 (per le forme d'onda ad intersimbolo nullo in uscita dal filtro di ricezione), sappiamo che otteniamo la massima velocità di trasmissione compatibilmente con la banda assegnata: occupiamo infatti una banda B e trasmettiamo a frequenza di cifra f_S=2B (cioè 2B bit al secondo: per esempio, per un segnale modulante di banda B=1kHz, trasmettiamo 2kbit/sec). Se invece trasmettiamo con modulazione di ampiezza, sappiamo che il segnale modulato $s_t(t)$ ha uno spettro di banda 2B centrato sulla frequenza f_C della portante:



La frequenza di cifra, però, rimane 2B, per cui continuiamo a trasmettere 2B bit al secondo, ma occupando una banda doppia rispetto a prima. Quindi, *rispetto alla trasmissione in banda base, la modulazione ASK comporta uno spreco di banda, quantificabile con un fattore 2.*

Per ridurre questo gap, si può allora pensare di ricorrere alla **modulazione di ampiezza in banda laterale unica** (che in analogico avevamo indicato con **SSB**, ossia *Single Side Band*): sappiamo, infatti, che, una volta effettuata la modulazione così come indicato prima, il contenuto informativo della banda laterale superiore (BLS) e di quella laterale inferiore (BLI) è lo stesso, per cui possiamo pensare di trasmettere una sola di tali bande: così facendo, occupiamo una banda nuovamente pari a B e quindi riguadagniamo il fattore 2 che prima invece perdevamo.

Tuttavia, non sempre è applicabile questo tipo di modulazione: sappiamo, infatti, che, a causa della non idealità dei filtri a nostra disposizione, per poter filtrare completamente una delle due bande e per poter ricostruire l'altra in ricezione, è necessario che lo spettro del segnale modulante sia nullo alle basse frequenze, cioè non si deve estendere fino alla continua. Il fatto è che i segnali di tipo numerico sono segnali passa-basso, che cioè includono la continua, per cui questo esclude la possibilità di usare la SSB per la trasmissione numerica.

Una via di mezzo può essere allora la **modulazione con banda laterale parzialmente soppressa** (che in analogico avevamo indicato con **VSB**, ossia *Vestigial Side Band*): anziché filtrare completamente una delle due bande laterali del segnale in uscita dal moltiplicatore, se ne considera una minima parte, progettando poi un filtro opportuno, in ricezione, che usi questa "parte" per la ricostruzione del segnale s(t) di partenza. In questo modo, è tollerabile anche la presenza di una continua.

Anche la modulazione VSB presenta però i suoi problemi, legati agli errori di fase tra la portante modulata e l'oscillazione locale, isofrequenziale con la portante, generata in ricezione. Cominciamo col ricordare che, quando entrambe le bande del segnale modulato vengono trasmesse (quindi ci riferiamo ad una modulazione di ampiezza del tipo DSB-SC), un eventuale sfasamento tra portante modulata e oscillazione locale provoca semplicemente una attenuazione del segnale ricostruito[4]. Questa attenuazione provoca un peggioramento del rapporto segnale/rumore, che però è recuperabile con un aumento di pochi dB della potenza in trasmissione, in base a quanto sappiamo a proposito del legame tra la potenza trasmessa e la probabilità di errore[5]. Al contrario, trasmettendo parzialmente una delle due bande, lo sbilanciamento tra di esse può essere dannoso: un eventuale sfasamento tra portante modulata e oscillazione locale provoca non solo una perdita di qualche dB della potenza in uscita, ma anche una forma d'onda che cambia tanto di più quanto maggiore è lo sfasamento stesso. Questo impedirebbe di controllare l'intersimbolo all'uscita del filtro in ricezione, il che sappiamo essere inaccettabile.

Quindi, riassumendo, *la tecnica VSB rende troppo schiavi dell'errore di fase tra portante modulata e oscillazione locale, per cui si usano comunque metodi di trasmissione in cui entrambe le bande laterali vengono trasmesse.*

Rimane a questo punto il problema della perdita di un fattore 2 nel rapporto tra velocità di trasmissione dei bit e banda occupata sul mezzo trasmissivo:

$$\text{in banda base} \longrightarrow \frac{\text{velocità}}{\text{banda}} = \frac{2B}{B} = 2$$

$$\text{con modulazione di ampiezza} \longrightarrow \frac{\text{velocità}}{\text{banda}} = \frac{2B}{2B} = 1$$

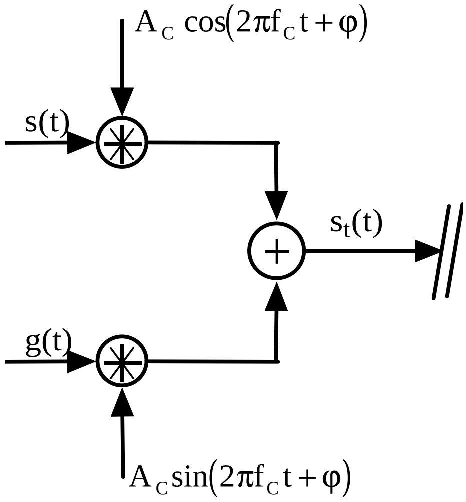
[4] Il motivo fondamentale di questo lo si intuisce facendo riferimento al metodo dei vettori rotanti: quando sono presenti entrambe le bande laterali, ossia entrambi i vettori, a sinistra e a destra di quello della portante (che coincide con l'asse verticale), comunque la composizione delle due bande consente di mantenere il riferimento di fase, in quanto la composizione dei vettori dà comunque un ulteriore vettore A in fase con la portante; l'eventuale sfasamento tra portante e oscillazione locale fa si, semplicemente, che non venga preso il vettore A così com'è, ma una sua proiezione, causando appunto semplicemente una attenuazione.

[5] Ci riferiamo al fatto che, aumentando di N dB la potenza di segnale trasmessa, si ottiene una riduzione della probabilità di errore di N ordini di grandezza: questo, lo ricordiamo, discende dalla relazione $p(\varepsilon)=Q(\gamma)$.

(i valori appena riportati si riferiscono evidentemente al caso di roll off δ=0, che però generalmente non è quello realmente attuato).

Si ricorre allora nuovamente al metodo di **modulazione con due portanti in quadratura**, che non viene praticamente mai usato in analogico (salvo che per i *segnali di crominanza* nella trasmissione del *segnale televisivo*), ma che invece, come vedremo adesso, si rivela molto utile in digitale.

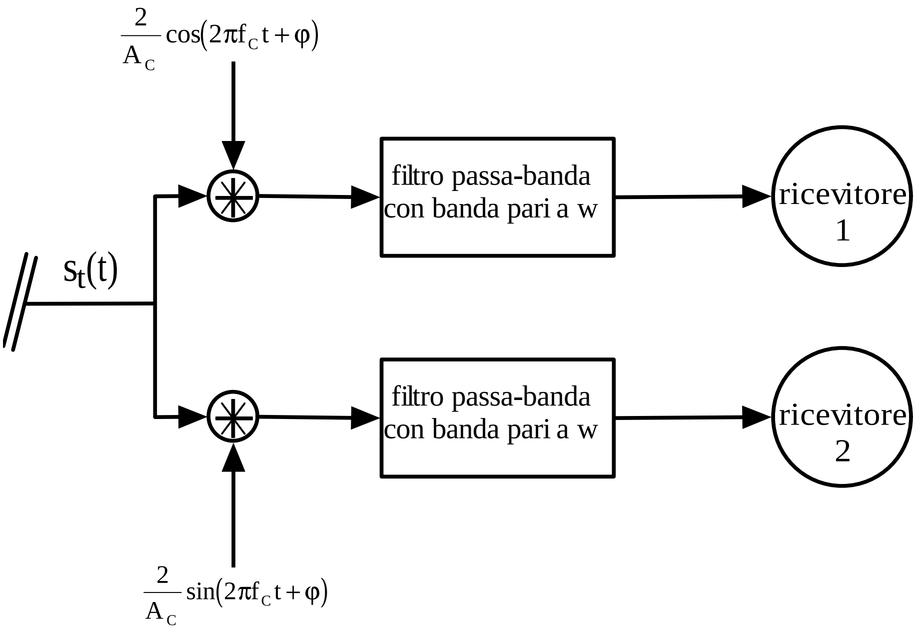
Lo schema che abbiamo adottato, nella trasmissione analogica, per l'apparato modulatore era il seguente:



Lo scopo è di trasmettere, mediante lo stesso mezzo trasmissivo, due distinti segnali s(t) e g(t): ciascuno modula d'ampiezza, mediante la tecnica DSB-SC, una portante a frequenza f_C, con la differenza che uno modula una portante Coseno e l'altro una portante Seno. I due segnali modulati vengono successivamente sommati per ottenere l'unico segnale da trasmettere attraverso il canale:

$$s_t(t) = s(t)A_C \cos(2\pi f_C t + \varphi) + g(t)A_C \sin(2\pi f_C t + \varphi)$$

Nell'ipotesi che il canale sia ideale, per cui non introduce distorsioni su $s_t(t)$, $s_t(t)$ è proprio il segnale che arriva in ingresso all'*apparato di demodulazione*, che possiamo schematizzare nel modo seguente:



Anche l'apparato demodulatore consta di due rami, in ciascuno dei quali viene effettuata una demodulazione del segnale modulato $s_t(t)$ secondo sempre la tecnica DSB-SC; in particolare, tale demodulazione consiste nella moltiplicazione di $s_t(t)$ per il segnale $\frac{2}{A_C}\cos(2\pi f_C t+\varphi)$ nel caso di s(t) e per il segnale $\frac{2}{A_C}\sin(2\pi f_C t+\varphi)$ nel caso di g(t), e nel successivo filtraggio.

Senza ripetere i dettagli analitici già visti nel caso analogico, le uscite dei due moltiplicatori sono le seguenti:

$$x_s(t) = s(t) + s(t)\cos(4\pi f_C t + 2\varphi) + g(t)\sin(4\pi f_C t + 2)$$
$$x_g(t) = s(t)\sin(2\pi f_C t + \varphi) + g(t) - g(t)\cos(4\pi f_C t + 2\varphi)$$

Filtrando questi due segnali sulla banda base B (si suppone che entrambi i segnali occupino la stessa banda B, in modo che il segnale trasmesso $s_t(t)$ occupi banda complessiva 2B), le uscite sono proprio s(t) e g(t).

E' intuitivo dunque comprendere l'utilità di questo metodo di trasmissione: se a ciascuna forma d'onda s(t) e g(t) associamo un bit, il metodo è tale che, occupando ancora una banda 2B sul mezzo trasmissivo, vengano trasmessi 2 bit per volta, ossia una velocità di trasmissione doppia della frequenza di cifra: quindi

$$\text{modulazione con 2 portanti in quadratura} \longrightarrow \frac{\text{velocità}}{\text{banda}} = \frac{4B}{2B} = 2$$

Abbiamo così recuperato il fattore 2 che con la DSB-SC avevamo invece perso rispetto alla trasmissione in banda base.

A questo punto, bisogna chiedersi se questo sistema di modulazione sia meno sensibile del precedente (VSB) nei confronti degli errori di fase tra portanti modulate e oscillazioni locali generate in ricezione. Questo discorso è stato già fatto, anche a livello analitico, con riferimento alla modulazione analogica, e si era trovato che la sensibilità, su ciascun ramo, a tali sfasamenti era sicuramente maggiore rispetto alla normale tecnica DSB-SC: per esempio, si era trovato che il segnale in uscita dal filtro posto sul ramo inferiore era

$$g_u(t) = s(t)\sin(\varphi_1 - \varphi) + g(t)\cos(\varphi_1 - \varphi) \neq g(t)$$

In base a questa espressione, *lo sfasamento j_1 tra portanti modulate e portanti in quadratura non porta solo ad una attenuazione, ma anche ad una interferenza con l'altro segnale.* Nel caso analogico, questo effetto è molto fastidioso, mentre lo è sicuramente di meno nel caso numerico: qui, infatti, il problema non è tanto quello di ricavare il segnale così come è stato trasmesso, bensì semplicemente quello di discriminare i livelli del segnale in modo da optare per 0 o per 1. Di conseguenza, finche l'interferenza tra i due segnali consente comunque una buona capacità di discriminare, si tratta di un fenomeno sicuramente tollerabile.

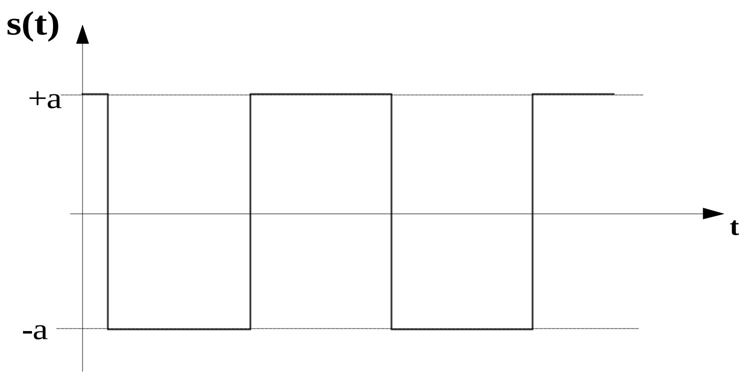
Quindi, *al contrario di quanto abbiamo concluso nel caso analogico, nel caso numerico il metodo di modulazione con due portanti in quadratura è decisamente utile.*

ANALISI DELLA MODULAZIONE ASK

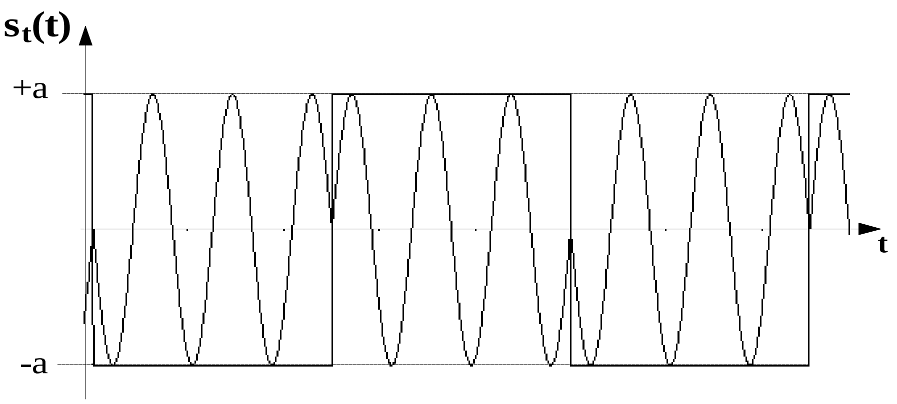
Con l'uso delle due portanti in quadratura abbiamo dunque trovato il modo di usare, sostanzialmente, la tecnica di modulazione con doppia banda laterale (e portante soppressa) senza perdere niente in quanto a rapporto tra velocità di trasmissione e banda occupata sul mezzo trasmissivo. E' allora opportuno fare qualche osservazione su questo tipo di modulazione quando il segnale modulante è di tipo numerico.

Il caso più semplice di ASK è quello descritto nel paragrafo precedente, in cui una portante sinusoidale viene modulata da un segnale modulante avente forma d'onda rettangolare. Tuttavia, l'ASK è valida anche nel caso in cui la forma d'onda modulante è di altro tipo. Si può, però, individuare una differenza fondamentale tra i due casi.

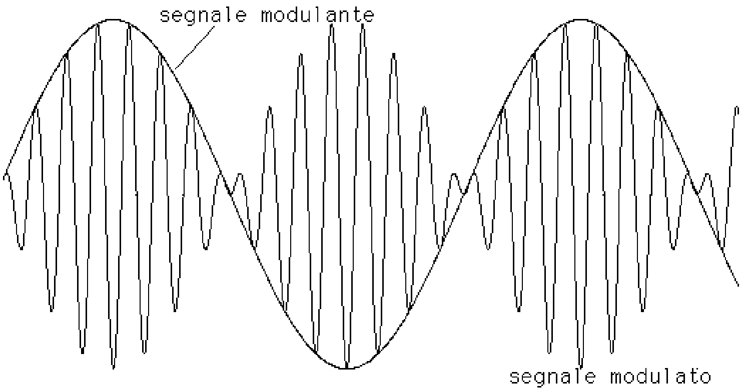
Supponiamo di aver scelto una codifica di linea di tipo *antipodale*, per cui associamo rettangoli di ampiezza +a al simbolo 1 e rettangoli di ampiezza -a al simbolo 0. Se, per esempio, volessimo trasmettere la sequenza logica 10101, il segnale modulante avrebbe la seguente forma d'onda:



Questo segnale va a modulare (cioè moltiplica) la portante sinusoidale a frequenza prefissata f_0, dando origine ad una forma d'onda del tipo seguente:



Si nota che *il segnale modulato è ancora una sinusoide, a frequenza f_0, di ampiezza costante, che però cambia la propria fase (di 180°) ogni volta che il segnale modulante cambia segno.* Se, invece, usiamo una forma d'onda modulante qualsiasi, si ha ancora una variazione di fase in corrispondenza dei cambiamenti di segno di tale forma d'onda modulante, ma si ha anche una variazione della ampiezza. Se, ad esempio, il segnale modulante è una sinusoide, allora sappiamo bene che il segnale modulato è del tipo seguente:



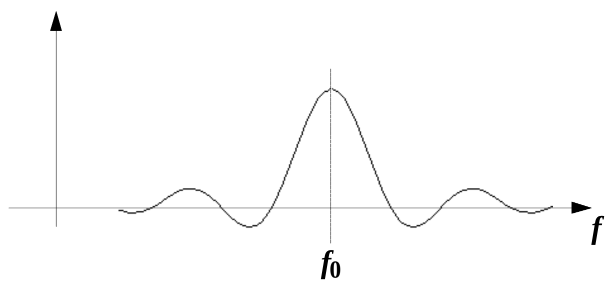
(si è utilizzata una sinusoide a frequenza evidentemente inferiore a quella della portante)

Quindi, nel primo caso (segnale modulante di forma rettangolare) si ottiene una *forma d'onda ad inviluppo costante* (l'inviluppo è una retta orizzontale), mentre nel secondo caso no.

Questa differenza è importante in quanto una forma d'onda ad inviluppo costante consente l'utilizzo, in trasmissione, di un amplificatore spinto fino in saturazione (che aumenta notevolmente l'efficienza rispetto ad un tradizionale stadio di uscita[6]). Infatti, supponiamo di inviare la forma d'onda ad inviluppo costante in un amplificatore e supponiamo che l'ampiezza dell'ingresso sia tale da mandare in saturazione l'amplificatore stesso: allora, l'amplificatore invece di presentare in uscita una sinusoide analoga a quella di ingresso (a parte l'eventuale amplificazione/attenuazione), fornisce una sinusoide squadrata (al limite anche un'onda quadra), la quale risente sicuramente almeno del segnale del segnale in ingresso. Ciò significa che la forma d'onda in uscita, a meno del ritardo dovuto al funzionamento dei dispositivi, sarà sincrona con la sinusoide in ingresso, ovvero coincideranno gli istanti di attraversamento dello zero delle due forme d'onda.

La differenza sostanziale tra il segnale in uscita dall'amplificatore in saturazione e quello in ingresso (ossia la portante moltiplicata) per un'onda rettangolare è nello spettro.

Cominciamo dalla sinusoide modulata dall'onda rettangolare: dato che un prodotto nel dominio del tempo equivale ad una convoluzione in frequenza, dobbiamo effettuare la convoluzione tra lo spettro della portante, che è un impulso a frequenza f_0, e lo spettro del rettangolo, che è del tipo $\sin(f)/f$; questa convoluzione sposta semplicemente lo spettro del rettangolo sulla frequenza f_0, in modo da ottenere il seguente spettro:

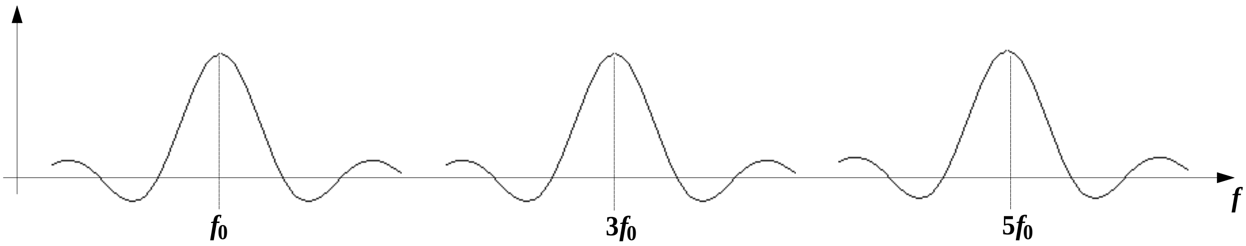


E' chiaro che le code del $\sin(f)/f$ si estendono fino all'infinito, ma si possono fare almeno due osservazioni: in primo luogo, il decadimento a 0 delle code fa si che, comunque, ad una certa distanza dalla frequenza centrale, l'ampiezza sia praticamente trascurabile; in secondo luogo, dato che, in ricezione, l'effetto del filtro (che deve tirar fuori forme d'onda a intersimbolo nullo) è anche quello di operare un

[6] Ricordiamo che uno stadio di uscita è un circuito che assorbe una certa quantità di potenza dall'alimentazione e cede una parte di essa al carico, dissipando invece la parte rimanente nei dispositivi che costituiscono il circuito stesso: gli stati di uscita migliori hanno una *efficienza di conversione* (rapporto tra la potenza trasferita al carico e la potenza assorbita all'alimentazione) il cui valore massimo (prettamente teorico) è circa del 75%. Al contrario, un amplificatore spinto in saturazione cede al carico praticamente tutta la potenza assorbita dall'alimentazione, per cui ha una efficienza di circa il 100%.

filtraggio delle componenti spettrali, si può comunque pensare di filtrare il segnale anche prima della trasmissione, in modo da eliminare le code e lasciare invariato solo il lobo principale del $\sin(f)/f$. In aggiunta a questo, ricordiamo che la "forma" di quel $\sin(f)/f$ dipende dalla durata dell'impulso rettangolare: quanto più tale impulso è lungo, tanto più il lobo principale del $\sin(f)/f$ si stringe (e si alza) attorno a f_0 e tanto più velocemente si attenuano le altre code.

Passiamo adesso al segnale in uscita dall'amplificatore in saturazione: il fatto che la sinusoide venga squadrata significa, nell'ipotesi di ottenere approssimativamente un'onda quadra con duty-cicle[7] del 50%, che dobbiamo moltiplicare il rettangolo modulante non più solo con l'armonica a frequenza f_0, ma anche con armoniche a frequenze che sono multipli dispari di f_0 ($3f_0$, $5f_0$,...). Questo comporta che lo spettro $\sin(f)/f$ del rettangolo venga replicato più volte e, in particolare, venga posizionato sia su f_0 sia anche su multipli dispari di f_0:



Anche in questo secondo caso si può ottenere in ricezione solo lo spettro del segnale di interesse: basta usare un filtro passa-banda centrato attorno alla frequenza f_0 e di banda passante sufficiente: il risultato è proprio lo spettro $\sin(f)/f$ centrato su f_0, ossia appunto la sinusoide a frequenza f_0 modulata da un'onda rettangolare.

Tra l'altro, è evidente che non potremo mai trasmettere, nella realtà, un segnale avente lo spettro indicato nella figura, in quanto la banda passante a nostra disposizione sul mezzo trasmissivo non sarà mai infinita, ma terrà conto del fatto che il mezzo deve ospitare più canali di trasmissione. Di conseguenza, dovremo aver cura, prima di trasmettere, di filtrare solo la replica spettrale di nostro interesse, che è chiaramente quella centrata su f_0.

In conclusione, *utilizzando una modulazione ASK, con rettangoli modulanti di tipo antipodale, possiamo dunque usare un amplificatore in saturazione per le generazione del segnale da trasmettere in linea, a patto poi di filtrare opportunamente il segnale da inviare.*

Il discorso appena fatto vale anche se, al posto di una oscillazione puramente sinusoidale, usiamo come portante direttamente un'onda quadra: tuttavia, questo è vero solo se il segnale modulante è di tipo rettangolare, in quanto, in caso contrario, mentre il segnale modulato avrà ancora uno spettro periodico, il segnale modulante non sarà più riconoscibile.

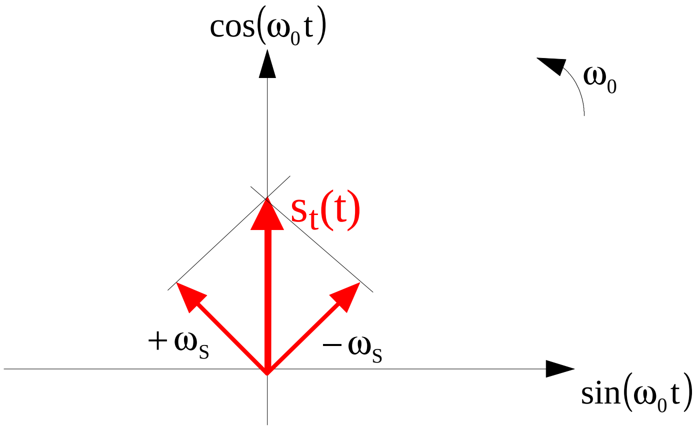
TRASMISSIONE QAM E MODULAZIONE DI FASE PSK

Passiamo adesso al metodo di trasmissione con due portanti in quadratura, nota con la sigla **QAM**, che sta appunto per *Quadrature Amplitude Modulation*. Il nostro scopo è quello di schematizzare in modo più diretto questo tipo di trasmissione.

A tal fine, ci serviamo nuovamente della rappresentazione mediante vettori rotanti: ricordiamo, perciò, che, dato un generico segnale sinusoidale, di pulsazione ω_S, che modula in DSB-SC una

[7] Ricordiamo che un'onda quadra con *duty-cycle* del 50% è tale per cui, in ogni singolo periodo T, il livello alto ed il livello basso sono mantenuti per un tempo uguale, che quindi coincide con T/2. In poche parole, in ogni periodo, il segnale deve essere per metà periodo al livello alto e per metà periodo a livello basso. Squadrando una sinusoide, si ottiene approssimativamente questo risultato, ottenendo un periodo dell'onda quadra uguale a quello della sinusoide di partenza.

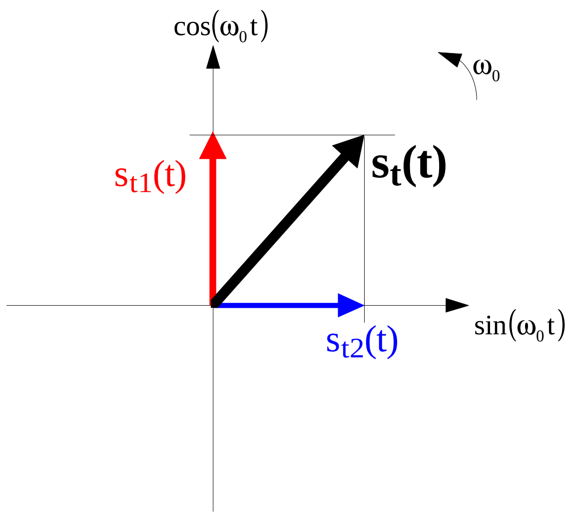
portante sinusoidale, di pulsazione ω_0, possiamo rappresentare il segnale modulato $s_t(t)$, in un piano che ruota sincronicamente con la portante, mediante un vettore che giace sull’asse della portante stessa e la cui ampiezza è funzione del tempo:



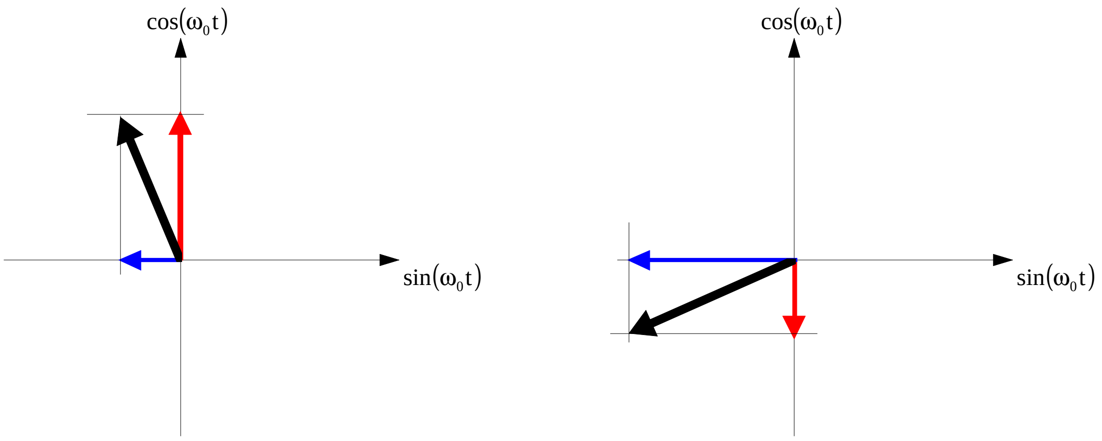
L’asse verticale è quello corrispondente alla portante, per cui su tale asse si trova il segnale modulato: esso è il risultato della composizione dei due vettori rotanti, di ampiezza metà, che ruotano con velocità $+\omega_S$ (cioè ω_S in senso antiorario) e $-\omega_S$ (cioè ω_S in senso orario). Questi due vettori, rappresentativi di due sinusoidi a frequenza f_0+f_S e f_0-f_S, sono, rispettivamente, la banda laterale superiore e la banda laterale inferiore del segnale modulato.

E’ chiaro che, per un generico segnale modulante s(t), la schematizzazione delle due bande laterali con dei vettori non ha più senso, ma è comunque esplicativa di quello che vogliamo dire.

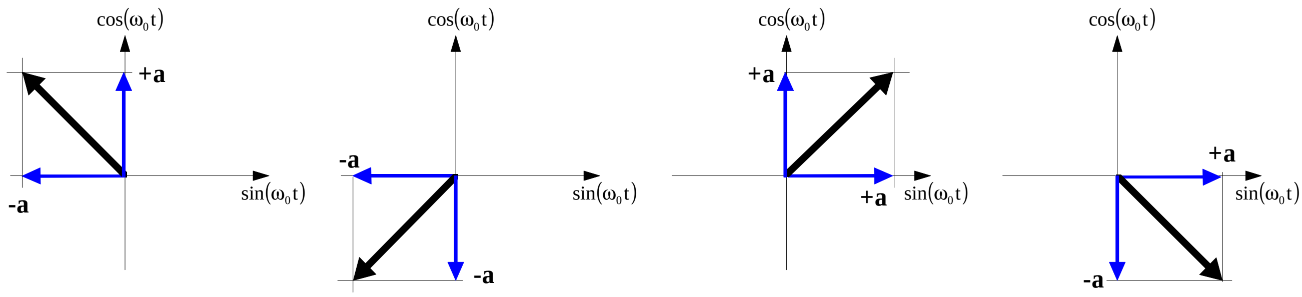
Nel caso della trasmissione con due portanti in quadratura, quello che noi facciamo è usare due segnali, $s_1(t)$ e $s_2(t)$, che modulano due portanti, isofrequenziali e in quadratura, con la tecnica DSB-SC; di conseguenza, nel piano rotante, avremo due vettori, rappresentativi dei due segnali modulati $s_{t1}(t)$ e $s_{t2}(t)$, giacenti uno sull’asse orizzontale e uno sull’asse verticale:



Per la trasmissione, i due segnali modulati vengono sommati, in modo da trasmettere la risultante $s_t(t)$. L’ampiezza e la fase di questa risultante, in ogni istante, dipendono dalla ampiezza, nello stesso istante, dei due segnali modulati, la quale ampiezza dipende a sua volta dall’ampiezza e dal segno, sempre nello stesso istante, dei due segnali modulanti $s_1(t)$ e $s_2(t)$. La figura precedente, per esempio, mostra quello che succede in un generico istante in cui $s_1(t)$ e $s_2(t)$ hanno entrambi ampiezza positiva, ma ci sono, teoricamente, infinite altre situazioni, come quelle indicate qui di seguito:



In realtà, le possibili situazioni diventano in numero finito qualora i segnali modulanti siano di tipo numerico, ossia rettangoli di due sole possibili ampiezze (ad esempio +a e -a). In questo caso, infatti, le possibilità sono solo 4:



Abbiamo dunque 4 diversi vettori rotanti, di uguale ampiezza (pari ad $a\sqrt{2}$), che <u>differiscono tra loro per la fase</u>:

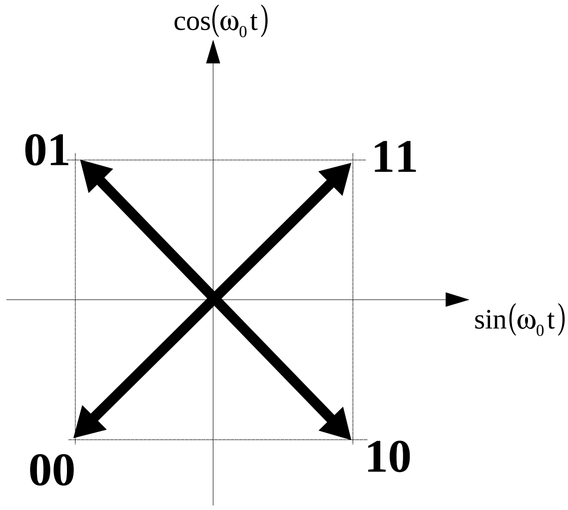
(+a,+a)→ +45°
(+a,-a)→ -45°
(-a,+a)→ +135°
(-a,-a)→ -135°

In ricezione, quindi, l'informazione è portata dalla fase del segnale ricevuto, in base alla tabella appena riportata. Questo mostra come, <u>nel caso della trasmissione numerica, con rettangoli modulanti di tipo antipodale, non sia ben delineato il confine tra modulazione di ampiezza e modulazione di fase</u>.

Naturalmente, il discorso appena fatto vale solo se il segnale modulante è di tipo rettangolare, per cui la conclusione cui siamo pervenuti è che, <u>*nel caso di trasmissione numerica, l'unica forma d'onda che è sensato utilizzare in trasmissione è una forma d'onda rettangolare.*</u>

<u>La quantità di informazione che tale forma d'onda contiene è sostanzialmente contenuta nel segno, dato che l'ampiezza è costante</u>. Effettuando una modulazione in DSB-SC con un segnale rettangolare antipodale, all'uscita del modulatore si ottiene una sinusoide con inviluppo costante, il cui segno cambia, in funzione del bit da trasmettere, da periodo di cifra a periodo di cifra. D'altra parte, cambiare il segno ad una sinusoide equivale a cambiare la sua fase di π: di conseguenza, si giunge sostanzialmente ad una **modulazione di fase**, che in numerico si indica con la già citata sigla <u>**PSK**</u>.

Tornando adesso alle 4 possibilità indicate prime, appare evidente che possiamo associare, a ciascuna possibilità (quindi sostanzialmente a ciascuna fase del segnale modulato) una combinazione di 2 bit:



Scegliendo, dunque, l'ampiezza dei due segnali modulanti, possiamo trasmettere, sullo stesso canale, cioè occupando sempre la stessa banda (che sarà 2B, se B è la banda del singolo segnale modulante), due simboli binari. Questo sistema consiste perciò in una **modulazione di fase a 4 livelli** (brevemente **4-PSK**).

Quindi, in definitiva, *la modulazione QAM con segnale modulante rettangolare antipodale è interpretabile come una modulazione di fase a 4 livelli.*

E' chiaro che i livelli possono essere anche più di 4, in base alle possibili ampiezze dei due segnali modulanti: per ottenere, ad esempio, 8 livelli, dovremo considerare 3 diverse ampiezze per i segnali modulanti, in modo appunto da poter ottenere $2^3=8$ 8 diverse combinazioni, ossia 8 fasi diverse per il segnale modulato inviato sul canale. In questo modo, possiamo trasmettere, per ogni periodo di cifra, 3 simboli binari, per cui, conservando ancora la stessa banda, trasmettiamo a velocità tripla rispetto a quella che otterremmo trasmettendo in banda base oppure usando un semplice sistema DSB-SC:

$$\text{in banda base} \longrightarrow \frac{\text{velocità}}{\text{banda}} = \frac{1\left(\frac{\text{bit}}{\text{periodo}}\right)}{B} = \frac{1\left(\frac{\text{bit}}{1/2B(\text{sec})}\right)}{B} = \frac{2B}{B} = 2$$

$$\text{con la DSB-SC} \longrightarrow \frac{\text{velocità}}{\text{banda}} = \frac{2B}{2B} = 1$$

$$\text{con la 4-PSK} \longrightarrow \frac{\text{velocità}}{\text{banda}} = \frac{4B}{2B} = 2$$

$$\text{con la 8-PSK} \longrightarrow \frac{\text{velocità}}{\text{banda}} = \frac{6B}{2B} = 3$$

In generale, si deduce quindi la seguente relazione generale: $\boxed{b = \log_2 M}$

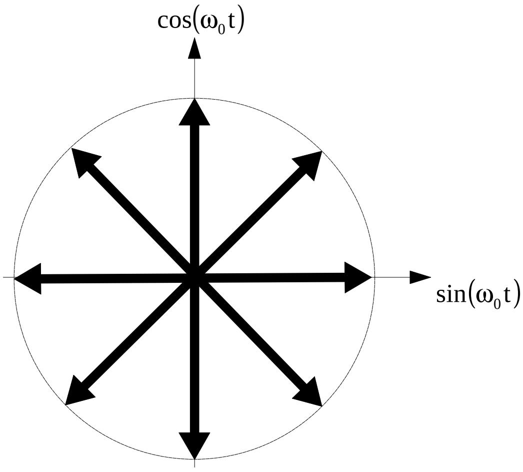
Per ottenere M livelli (di fase), abbiamo bisogno di b ampiezze diverse per i rettangoli modulanti, ossia, in altre parole, *ogni forma d'onda inviata sul canale convoglia una informazione pari a* $\log_2 M$. Inoltre, con b ampiezze diverse, possiamo ottenere un rapporto velocità/banda pari a $\boxed{\frac{2bB}{2B} = b}$.

Probabilità di errore in un sistema M-PSK

In base a quanto detto poco fa, quanto maggiore è il numero b di livelli (per ora consideriamo solo potenze di 2, ma vedremo che non si tratta di un requisito indispensabile), tanto maggiore è la velocità di trasmissione, a parità di banda occupata sul canale. Tuttavia, si pone un importante problematica: in ricezione, il nostro interesse è quello di individuare, con una certa probabilità di errore, la combinazione di bit che è stata trasmessa; di conseguenza, per un sistema M-PSK, quello che conta, ai fini della capacità discriminativa del ricevitore, è la distanza tra i vertici dei vettori rappresentativi delle diverse configurazioni.

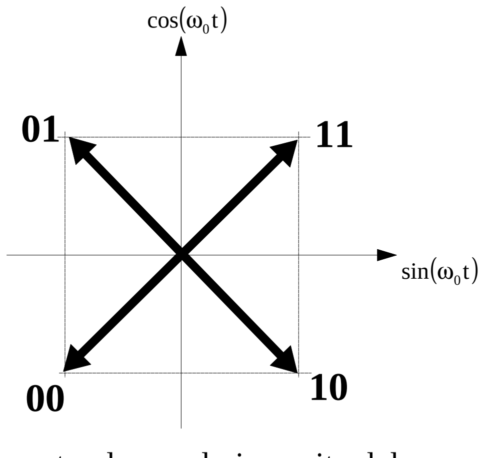
Nel caso della 4-PSK, questa distanza è ovviamente quella massima ottenibile, in quanto corrisponde al massimo sfasamento (pari a 180°) dei vettori, per cui si ottiene la minore probabilità di errore possibile.

Se passiamo alla 8-PSK, la distanza tra gli 8 vettori diminuisce e quindi aumenta la probabilità di errore:



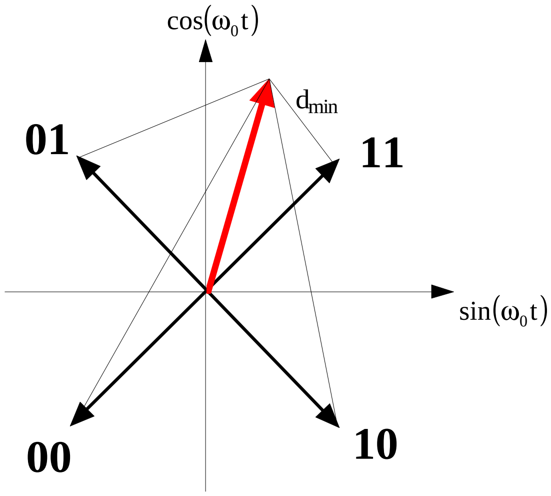
Infittendo ulteriormente il numero di livelli, è evidente che la probabilità di errore, a parità di ampiezza dei vettori, aumenta. Vediamo allora con maggiore dettaglio quello che succede.

Partiamo dal caso semplice di M=4 livelli:



Se non ci fosse rumore sovrapposto al segnale in uscita dal mezzo trasmissivo, il segnale ricevuto può corrispondere solo a uno di quei 4 vettori rotanti. Al contrario, in presenza di rumore, quello che arriva è un vettore di ampiezza e fase casuali[8], come ad esempio quello della figura seguente:

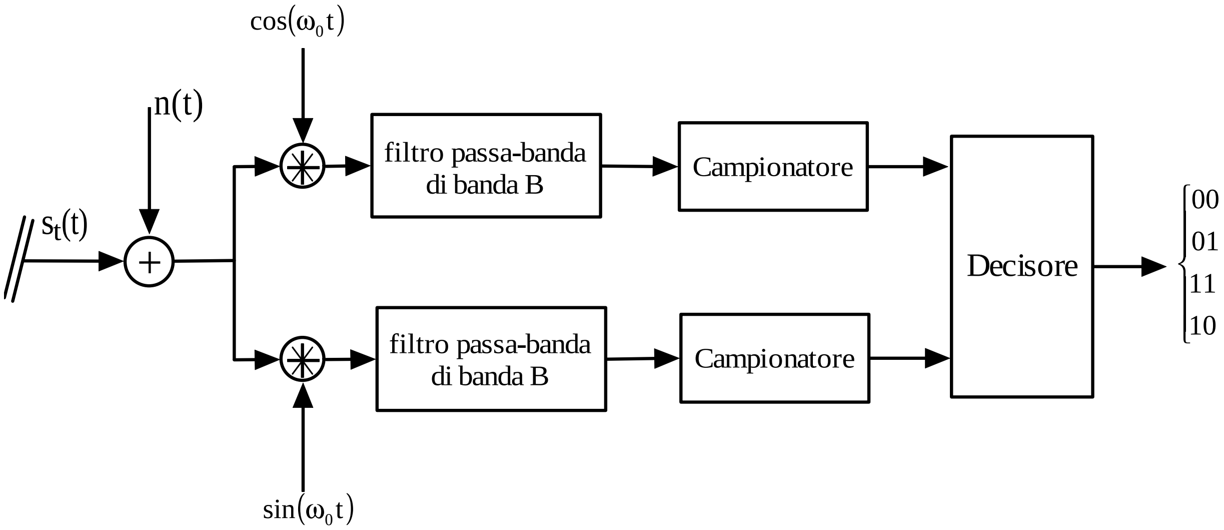
[8] Ricordiamo che il rumore sovrapposto al segnale è un rumore passa-banda, rappresentabile perciò come somma di un termine in fase ed uno in quadratura, ciascuno di ampiezza casuale. Di conseguenza, la composizione del vettore rappresentativo del segnale con tali due componenti dà origine ad un vettore di ampiezza e fase casuali.



L'operazione di decisione del ricevitore corrisponde, in generale, ad una ripartizione in sottospazi: nel caso di M=4, i sottospazi coincidono con i quadranti del piano rotante, per cui il compito del ricevitore è semplicemente quello di verificare in quale quadrante si trova il vettore rotante che è stato ricevuto. Questo criterio, come detto, corrisponde ad un **criterio di distanza minima** da una delle 4 possibili configurazioni: i punti del primo quadrante, per esempio, sono a distanza minima dalla configurazione 11, per cui questa sarà la configurazione per cui opterà il decisore per tutti i vettori che cadono nel suddetto quadrante.

Se volessimo fare un calcolo della probabilità di errore, dovremmo procedere in modo del tutto analogo a quello seguito per la trasmissione in banda base: una volta individuata la statistica del rumore sovrapposto al segnale, è immediato ricavare la statistica del segnale complessivo in uscita dal campionatore e, sulla base di questa statistica e della posizione scelta per le soglie (il sistema ha più di 2 livelli, per cui le soglie sono più di 1), si possono calcolare le probabilità di errore sulle singole combinazioni e quindi la probabilità media di errore $p(\varepsilon)$.

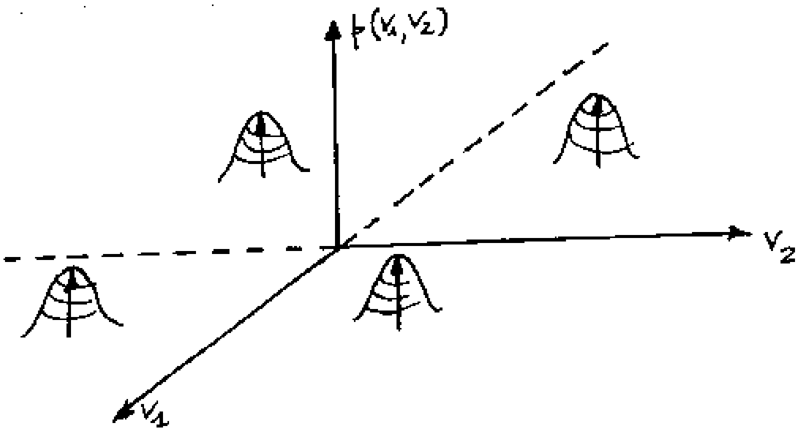
La difficoltà, essenzialmente di tipo matematico, in questo caso è proprio nello studio del rumore. A tal fine, riprendiamo lo schema generale dell'apparato demodulatore per la QAM (che abbiamo visto essere equivalente alla 4-PSK):



Il rumore n(t) è il classico rumore bianco con distribuzione gaussiana delle ampiezze; esso si distribuisce su entrambi i rami dell'apparato, per cui, all'uscita dei due campionatori, abbiamo due rumori passa-banda, chiaramente uguali tra loro, ancora con distribuzione gaussiana. Il decisore opera considerando la coppia v_1, v_2 di tensioni in uscita dai due campionatori (nell'istante di campionamento) e stabilendo a quale configurazione di bit questa coppia corrisponde. Si deduce, quindi, che si ottengono 4 densità di probabilità, che sono ovviamente probabilità condizionate alla trasmissione di 00, 01, 11 ed 10. Se non ci fosse rumore, si avrebbero 4 impulsi, per cui il sistema non sbaglierebbe mai; data, invece, la presenza del rumore, le 4 densità di probabilità condizionate sono gaussiane bidimensionali a simmetria circolare tutte uguali tra loro.

Qui subentra, dunque, la difficoltà matematica cui si accennava prima: lo spazio da considerare è adesso bidimensionale, dato che il decisore ha a che fare con *coppie di valori di tensione*.

Volendo dare una rappresentazione grafica di quello che stiamo dicendo, possiamo usare uno spazio tridimensionale, in cui i due assi orizzontali corrispondono ai valori delle tensioni v_1 e v_2, mentre l'asse verticale indica il valore delle 4 probabilità condizionate, indicate con $p(v_1,v_2)$:



Data la perfetta simmetria e uguaglianza delle gaussiane e data, come al solito, l'equiprobabilità delle combinazioni[9], il modo più sensato di procedere è quello di porre le soglie a metà tra una gaussiana e l'altra. Fissate le soglie, le singole probabilità di errore, da cui poi risalire alla probabilità di errore $p(\varepsilon)$ complessiva, corrispondono ai volumi occupati dalle code delle 4 gaussiane.

Nel caso di una 8-PSK, il procedimento è assolutamente identico, con la differenza che le configurazioni possibili sono 8, per cui anche le gaussiane diventano 8. A parità di ampiezza dei vettori rispetto alla 4-PSK ed a parità di rumore (che è sempre lo stesso a prescindere dal numero di livelli e, in generale, a prescindere dal segnale ricevuto), è ovvio che la probabilità di errore aumenta: infatti, diminuisce la distanza tra una gaussiana[10] e l'altra, per cui aumentano i volumi delle code e quindi aumenta $p(\varepsilon)$.

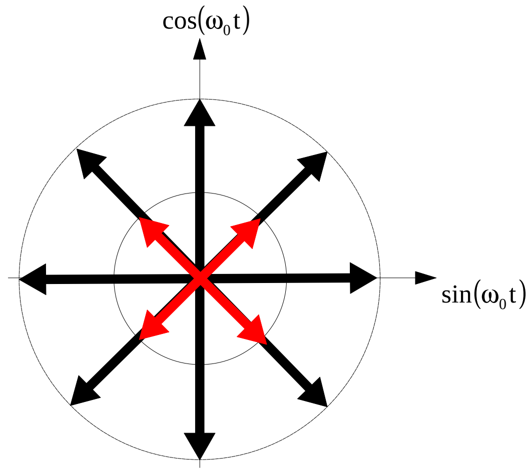
Passando ad una 16-PSK, la cosa peggiora ulteriormente e così via.

[9] Così come, per un sistema binario (cioè a due soli livelli), è sempre lecito assumere che i bit siano equiprobabili, in un sistema multilivello è lecito assumere che le combinazioni siano equiprobabili. A questo proposito, però, è bene fare una osservazione, che è stata già fatta in precedenza: in generale, non è assolutamente detto che i bit (o le configurazioni) siano equiprobabili, ma, altrettanto in generale, si fa spesso in modo che lo diventino, usando le già citate procedure di scrambling. Nonostante questo, però, ci sono anche casi in cui la non equiprobabilità è talmente evidente che viene sfruttata: si pensi al caso della trasmissione dei FAX, dove, associando 1 al nero e 0 al bianco, lo 0 è estremamente più probabile dell' 1, visto che un FAX è un foglio bianco con un certo numero di punti neri corrispondenti alle frasi.

[10] Ricordiamo che anche le gaussiane circolari bidimensionali, come quelle monodimensionali, si estendono teoricamente fino all'infinito, ma in realtà si può assumere che siano nulle a sufficiente distanza dal centro. Questo significa che, mentre in linea del tutto teorica la probabilità di errore non è mai nulla, nella pratica può essere comunque resa trascurabile: basta fare in modo che le soglie (sarebbe meglio parlare di **linee di confine**) taglino le gaussiane dove essere hanno un livello sufficientemente basso.

TRASMISSIONE M-QAM

Facciamo adesso il seguente ragionamento. *In genere, in un sistema di trasmissione numerico si cerca di trasmettere il più velocemente possibile, compatibilmente con la banda a disposizione e ovviamente con i vincoli sulla probabilità di errore.* Abbiamo visto che, se la banda a disposizione è limitata, l'unico modo di aumentare la velocità di trasmissione è quello di usare un sistema multilivello, come il sistema M-PSK visto prima (M sono i livelli). D'altra parte, abbiamo anche visto che, essendo il rumore sempre lo stesso, la probabilità di errore aumenta all'aumentare del numero M di livelli, in quanto le gaussiane sono sempre le stesse, ma sono maggiori come numero e quindi sono più vicine. Allora, *se si vuole aumentare il numero di livelli, mantenendo però costante la probabilità di errore, l'unico modo è quello di mantenere costante la distanza tra le gaussiane, il che significa aumenta l'ampiezza dei vettori:*



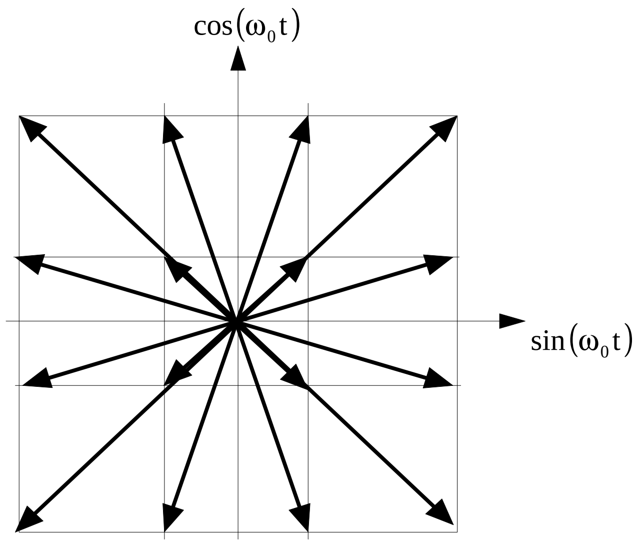
Nella figura sono indicati un sistema 4-PSK ed un sistema 8-PSK: la maggiore ampiezza dei vettori di quest'ultimo sistema fa si che la distanza tra gli estremi dei vettori, cioè la distanza tra le gaussiane, sia la stessa che si ottiene con il 4PSK, usando una ampiezza minore. La probabilità di errore è la stessa nei due casi.

Il fatto è che aumentare l'ampiezza dei vettori significa necessariamente aumentare la potenza in trasmissione e non è detto che si possa effettuare questo aumento. Questo è il motivo per cui, quando la complessità del sistema multilivello sale, non sia più conveniente usare il PSK: solitamente, non si va oltre il 16-PSK, in quanto, per un numero superiore di livelli, l'efficienza del sistema in termini di potenza non è più accettabile.

D'altra parte, però, la trasmissione QAM consente comunque di ottenere qualche altro risultato importante. Infatti, con riferimento sempre alla figura precedente, si osserva che, nel caso della 8-PSK, gran parte dello spazio all'interno del cerchio tratteggiato rimane praticamente inutilizzata. Si può allora sfruttare questo spazio per aumentare il numero di punti, considerando sempre che, ai fini della probabilità di errore finale, ciò che conta è solo la distanza tra un punto e quelli immediatamente adiacenti.

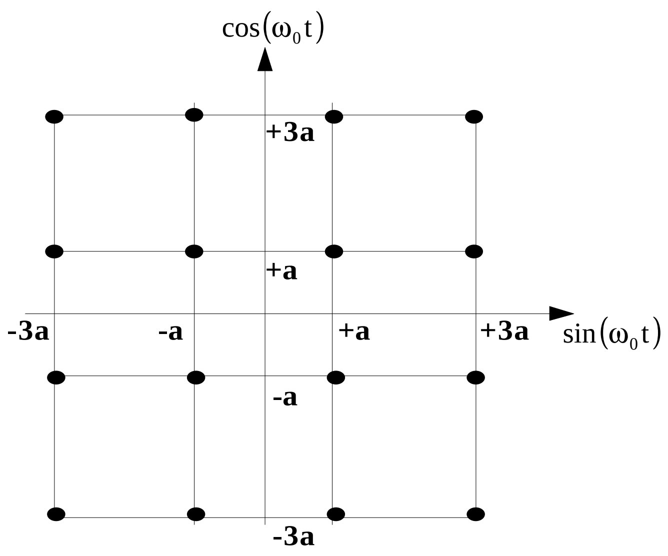
Se vogliamo sfruttare questo ulteriore spazio, non possiamo più basarci solo sulla fase dei vettori, ma dobbiamo necessariamente considerarne anche l'ampiezza: usando, infatti, vettori di ampiezza e fase diversa, siamo teoricamente in grado di coprire ogni punto dello spazio interno al cerchio. Ecco, quindi, che si passa dalla M-PSK alla **M-QAM**, ossia un *sistema multilivello basato sull'ampiezza oltre che sulla fase, del segnale ricevuto.*.

Chiaramente, dobbiamo fare in modo da coprire punti che siano tutti equidistanti tra loro. Per ottenere questo, anziché un cerchio dobbiamo considerare un quadrato:



Nella figura è stato considerato un sistema costituito da 16 livelli equidistanti tra loro. Il modo con cui realizzare praticamente questo sistema è abbastanza semplice: *basta scegliere in modo opportuno le ampiezze dei due segnali modulanti, in modo da ottenere un sistema multilivello sia sulla fase, come nel caso M-PSK, sia anche sull'ampiezza.*

Per esempio, per ottenere il sistema **16-QAM**, dobbiamo considerare 4 livelli per ogni portante, in modo da ottenere le 4^2=16 possibili configurazioni indicate in figura. Per ottenere l'equidistanza tra i vari punti (cioè tra i vertici dei vari vettori), basta considerare, sia per una portante sia per l'altra, 4 livelli di valore $-3a, -a, +a, +3a$:

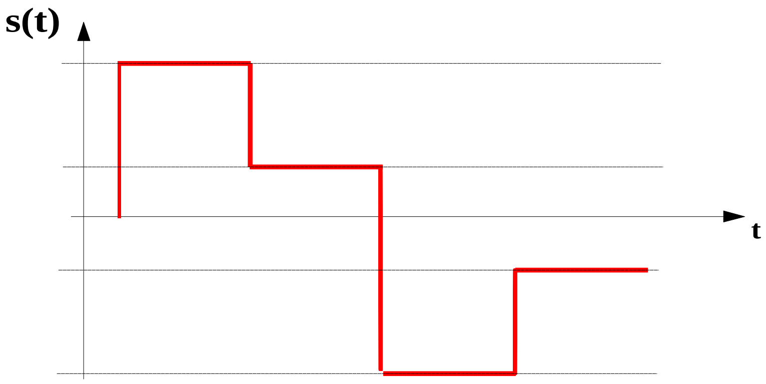


I punti del piano, corrispondenti alle varie configurazioni, costituiscono la cosiddetta **costellazione**. Quella appena esaminata è una costellazione da 16, ma ci sono ovviamente costellazioni comprendenti un numero maggiore di punti (sempre potenze di 2 per il momento).

Per ottenere un sistema 32-QAM, avremo bisogno di 5 livelli per portante, che saranno ad esempio $-4a, -2a, 0, +2a, +4a$: otterremo una costellazione da 32 punti.

E' importante osservare che, *nel passaggio dalla M-PSK alla M-QAM, la forma d'onda trasmessa non ha più inviluppo costante, proprio perché le informazioni sono adesso anche nelle ampiezze.*

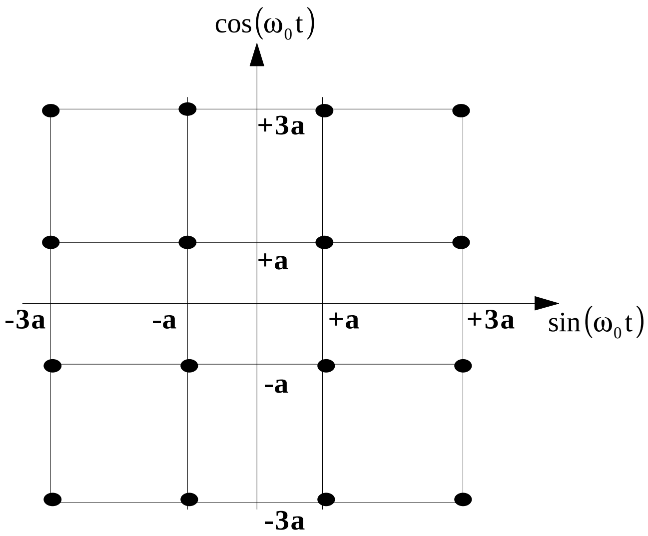
Per esempio, in un sistema 16-QAM, con 4 possibili ampiezze, la forma d'onda modulante può essere del tipo seguente:



<u>Usando questo segnale per modulare la portante sinusoidale, è chiaro che il segnale modulato non ha più inviluppo costante</u>. Questo spiega il concetto precedentemente esposto, secondo cui il sistema M-QAM è un multilivello sia sulla fase (come l' M-PSK) sia sull'ampiezza.

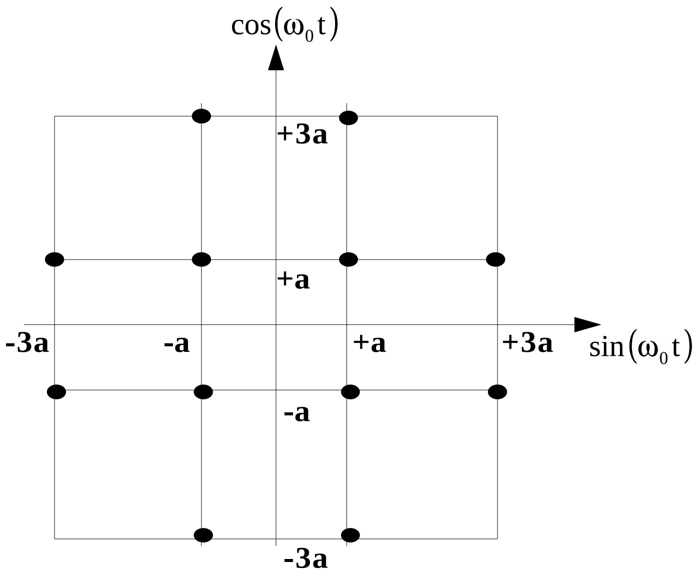
Di conseguenza, in trasmissione non è più possibile usare un amplificatore in saturazione, in quanto la saturazione si potrebbe ottenere solo per le 2 ampiezze maggiori e non per le altre due, producendo così un segnale fortemente distorto. *Questo è dunque il prezzo che si paga, nel sistema M-QAM, ottenendo in cambio una maggiore efficienza nell'uso della potenza trasmessa.*

Un'altra importante considerazione da fare riguarda la potenza di picco da trasmettere in un sistema M-QAM. Facciamo ancora una volta riferimento ad un sistema **16-QAM** (4 livelli per portante):



Considerando che la potenza di picco di un segnale è il suo valore massimo (positivo o negativo) elevato al quadrato e considerando che, in un generico sistema QAM, dobbiamo trasmettere due segnali, ciascuno con una propria potenza di picco, è ovvio che *la potenza di picco del segnale complessivo $s_t(t)$ da trasmettere sul canale è determinata dalle configurazioni relative ai vertici del quadrato:* tali configurazioni, infatti, corrispondono ai valori massimi dei due segnali $s_{t1}(t)$ e $s_{t2}(t)$ da sommare e trasmettere.

Generalmente, le limitazioni, in trasmissione, sono proprio per la potenza di picco, per cui può spesso essere necessario limitare tale potenza. Nel caso della 16-QAM riportata in figura, l'unico modo di ridurre la potenza di picco è quello di non trasmettere le 4 configurazioni estreme. Così facendo, si ottiene *ancora un sistema QAM multilivello, dove però il numero di configurazioni non è più una potenza di 2:* infatti, delle 16 configurazioni possibili per un 16-QAM, ne eliminiamo 4 (cioè imponiamo che non vengano mai trasmesse), portandoci così ad un sistema <u>**12-QAM**</u>:



Non è detto che le configurazioni estreme vadano necessariamente eliminate. In generale, il principio che si può seguire può essere quello di *riempire, nel modo migliore, l'area avente per confine nuovamente un cerchio.* Per esempio, nel sistema ormai più diffuso, che è il **256-QAM**, invece di trasmettere tutte le possibili configurazioni (sarebbero 16 in orizzontale e 16 in verticale), le configurazioni vicino ai vertici vengono spostate vicino ai punti medi dei lati, in modo appunto da ottenere approssimativamente un cerchio e, ovviamente, da rispettare sempre un passo regolare tra i punti del reticolo.

In effetti, si potrebbe pensare anche ad altre tecniche, per esempio notando che la distanza tra due punti estremi della diagonale è maggiore della distanza tra due punti estremi di un lato del quadrato: si può allora pensare di imporre che ogni punto sia equidistante da quelli che lo circondano. Tuttavia, le complicazioni tecnologiche sono in questi casi tali da sconsigliare questo tipo di soluzioni.

Complementi sul sistema PSK

Esaminiamo adesso brevemente il sistema PSK in base alla sua naturale definizione, cioè quella di un sistema nel quale si fa variare la fase della portante in modo proporzionale al segnale modulante:

$$\begin{cases} s(t) \\ c(t) = \cos(2\pi f_0 t) \end{cases} \longrightarrow \cos(2\pi f_0 t + ks(t))$$

E' evidente che, essendo il segnale modulante s(t) di tipo numerico, ossia un rettangolo di sole due ampiezze, anche la fase del segnale modulato potrà avere solo due valori: per esempio, supponendo una codifica antipodale con ampiezze +A e -A, la fase del segnale modulato potrà essere solo +kA oppure -kA.

Dato che varia solo la fase della sinusoide, mentre ampiezza e frequenza rimangono inalterate, è evidente che anche il sistema PSK è uno schema di modulazione ad inviluppo costante. Di conseguenza, per le considerazioni già fatte in precedenza, possiamo anche qui squadrare il segnale e poi filtrare attraverso un filtro passa-banda centrato sulla frequenza della portante e di banda opportuna.

MODULAZIONE DI FREQUENZA FSK

Dopo la modulazione di ampiezza e la modulazione di fase, vediamo adesso come è possibile modulare di frequenza una portante sinusoidale mediante un segnale di tipo numerico, ossia una sequenza di impulsi rettangolari di ampiezza e durata prefissate. Si parla di **modulazione FSK**.

Il concetto è sempre quello di far variare, in ciascun istante, la frequenza della portante in modo proporzionale al valore assunto in quello stesso istante dal segnale da trasmettere s(t). Consideriamo perciò la generica portante sinusoidale

$$c(t) = A_C \cos\left(2\pi f_0 t + \varphi(t)\right)$$

Indichiamo con $\theta(t) = 2\pi f_0 t + \varphi(t)$ l'argomento del coseno.

Si definisce "**frequenza istantanea**" la quantità

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

Sostituendo l'espressione di $\theta(t)$, otteniamo che

$$f_i(t) = f_0 + \frac{1}{2\pi}\frac{d\varphi(t)}{dt}$$

<u>La caratteristica della modulazione di frequenza è quella di far variare, istante per istante, questa grandezza in modo proporzionale al valore assunto, in ciascun istante, dal segnale s(t)</u>. Vediamo allora quale legame intercorre tra questa grandezza e s(t) stesso: si pone $\boxed{f_i(t) = f_0 + ks(t)}$, per cui, uguagliando con la precedente espressione (nel riquadro), di ottiene

$$f_0 + ks(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt}$$

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

$$\theta(t) = 2\pi f_0 t + 2\pi k \int_{-\infty}^{t} s(\tau)d\tau$$

Questo è dunque l'argomento del segnale modulato.

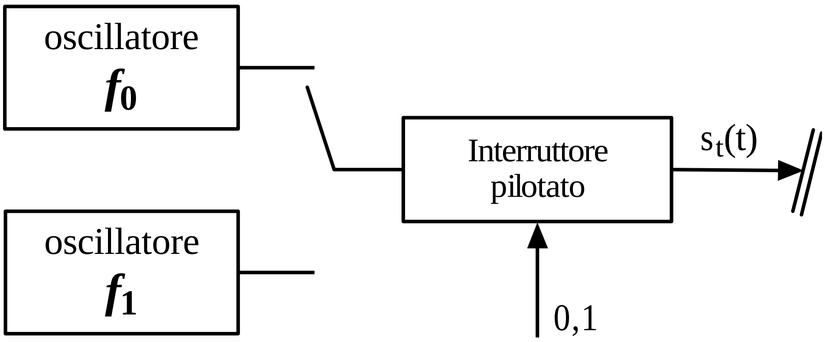
Evidentemente, ad ogni valore di s(t) corrisponderà un preciso valore di $f_i(t)$ e quindi anche un preciso valore dell'argomento $\theta(t)$ del coseno della portante. <u>A questo punto, quindi, subentra la natura del segnale modulante</u>: essendo il segnale di tipo numerico, esso corrisponderà, come detto, ad una sequenza di rettangoli, di durata prefissata T, con solo 2 possibili ampiezze: nel caso di codifica ortogonale, per esempio, si ha

$$\text{trasmissione di uno } 0 \longrightarrow s(t) = 0 \longrightarrow f_i(t) = f_0$$

$$\text{trasmissione di un } 1 \longrightarrow s(t) = a \longrightarrow f_i(t) = f_0 + ka = f_1$$

Il risultato è dunque semplicemente quello di <u>*trasmettere una sinusoide a frequenza f_0 quando il simbolo è 0 e una sinusoide a frequenza f_1 quando il simbolo è 1*[11].</u>

Ci sono due modi diversi di ottenere questo tipo di segnale. Un modo abbastanza semplice è quello di usare due distinti oscillatori, che oscillano a frequenze f_0 e f_1, e di inviare sul canale l'uscita dell'uno o dell'altro a seconda del valore del segnale modulante s(t):



Il segnale modulante s(t) pilota quindi un commutatore che collega il canale a uno dei due oscillatori, scelto appunto in base al valore di s(t) in ogni istante.

E' evidente che, per evitare brusche discontinuità nel segnale da trasmettere, è necessario effettuare le commutazioni negli istanti di zero di entrambe le sinusoidi, le cui frequenze, quindi, vanno scelte in modo opportuno, in accordo anche al valore del periodo di cifra T: bisogna fare in modo che, sia all'inizio sia alla fine di ogni periodo T, entrambe le sinusoidi valgano 0.

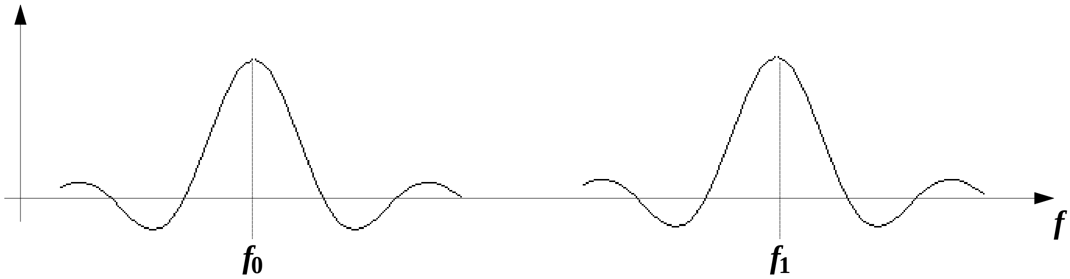
Il problema è che, <u>normalmente, non si riesce a commutare esattamente in corrispondenza dello zero delle due sinusoidi, il che provoca un salto di tensione nella transizione da una oscillazione all'altra</u>. Queste ripetute discontinuità, pur non portando alcuna informazione, allargano lo spettro del segnale da trasmettere, che quindi necessita di più banda sul mezzo trasmissivo.

Un modo alternativo di procedere è allora quello di usare un oscillatore controllato in tensione (**VCO** - *Voltage Controlled Oscillator*), pilotato dalla forma rettangolare in ingresso. Con questo dispositivo, non ci sono discontinuità per il seguente motivo: la frequenza di oscillazione viene variata cambiando il valore della capacità di un condensatore inserito nell'anello di reazione dell'oscillatore stesso; allora, anche variando istantaneamente il valore della capacità, non potrà mai variare istantaneamente il valore della carica da essa immagazzinata. Di conseguenza, la tensione generata dall'oscillatore non presenterà salti bruschi, ma, d'altra parte, presenterà comunque dei transistori in corrispondenza di ogni transizione.

Questi due modi di realizzare il modulatore FSK portano dunque allo stesso formato di trasmissione in linea, ma producono segnali che possono avere spettri anche sensibilmente diversi. Non solo, ma tali segnali, oltre ad avere una diversa occupazione di banda, pur convogliando la stessa identica informazione, richiedono tecniche di demodulazione diverse. Vediamo allora quali sono queste tecniche.

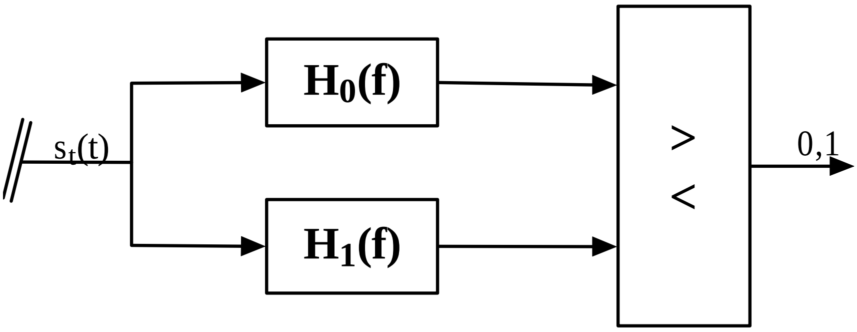
Intanto, è immediato comprendere come sia fatto lo spettro del segnale trasmesso: abbiamo infatti visto in precedenza che, data una generica sinusoide, a frequenza f_0, modulata da un segnale rettangolare, il segnale modulante ha spettro a forma di $\sin(f)/f$ centrato sulla frequenza f_0. Nel nostro caso, le possibili frequenze sono due, per cui avremo o l'uno o l'altro spettro tra quelli indicati nella figura seguente:

[11] La codifica di linea usata per una modulazione FSK non potrà mai essere antipodale, in quanto sappiamo bene che frequenze positive e negative sono la stessa cosa. Potremo quindi usare solo una codifica ortogonale.

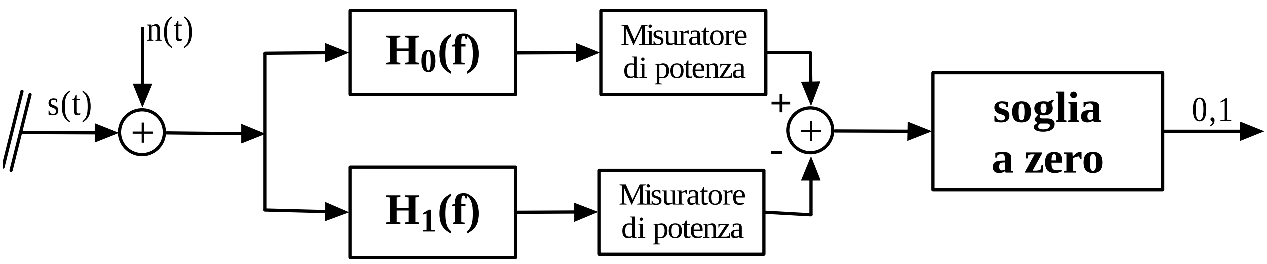


Allora, il ricevitore ottimo è costituito da due filtri, la cui risposta all'impulso somigli alle forme d'onda per le quali sono usati: il filtro relativo al simbolo 0 sarà un filtro passa-banda centrato alla frequenza f_0, mentre quello relativo al simbolo 1 sarà un filtro passa-banda centrato attorno a f_1.

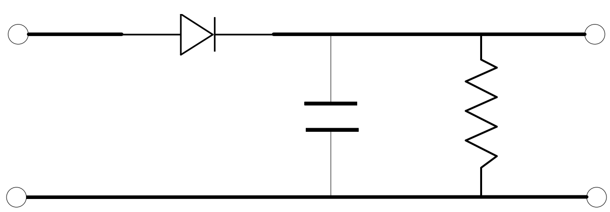
Supponiamo, per semplicità, che i due filtri siano adattati e lasciamo perdere la questione dell'intersimbolo nullo all'uscita dei due filtri stessi. Quando i due filtri ricevono il segnale trasmesso, l'uscita maggiore sarà ovviamente quella del filtro corrispondente al simbolo ricevuto (che, a meno di errori, coincide con quello trasmesso). Basta allora fare un confronto tra le due uscite, selezionando quella maggiore e optando quindi per il simbolo corrispondente:



Come effettuare il confronto tra le due uscite? Se i due filtri sono adattati, le uscite, per ciascun simbolo trasmesso, sono qualcosa di proporzionale alla funzione di autocorrelazione di una sinusoide (a frequenza f_0 o f_1) di durata T. Il simbolo per cui decidere sarà allora quello corrispondente al filtro che in uscita produce un segnale di potenza di picco maggiore. Si può allora adottare un semplice schema del tipo seguente:



Il misuratore di potenza può essere semplicemente un demodulatore ad inviluppo, in quanto ciò che interessa è il valore massimo del segnale in uscita dal filtro, in quanto tale valore determina la potenza di picco. Il suo schema circuitale sarà quindi del tipo seguente:



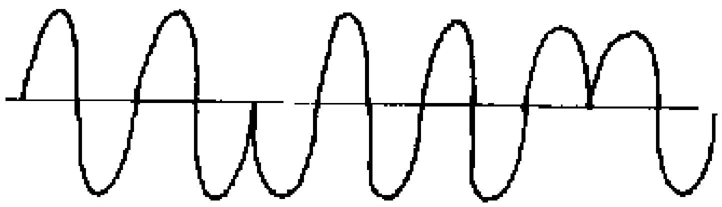
DEMODULAZIONE PSK

Abbiamo prima osservato che, per come è realizzata, la modulazione FSK richiede necessariamente una codifica di linea di tipo ortogonale. Allora, in base a quanto abbiamo in precedenza visto a proposito della codifica antipodale e di quella ortogonale, sappiamo che la codifica antipodale consente un risparmio di 3dB di potenza media[12]. Da questo punto di vista, quindi, *dato che la modulazione FSK non consente una codifica antipodale, mentre la PSK si, è evidente che quest'ultima è più efficiente, in quanto richiede 3dB in meno di potenza media per ottenere la stessa probabilità di errore.*

A fronte di questo vantaggio, la PSK ha, però, un problema in più: infatti, avendo visto che la PSK è una forma particolare di modulazione di ampiezza, è ovvio che la demodulazione deve essere *coerente*, nel senso che è necessario generare, in ricezione, una oscillazione sincrona, in frequenza ed in fase, con la portante modulata.

L'unico modo di procedere, in modo del tutto analogo a quanto visto a proposito del problema di estrazione del timing, è quello di ricavare l'informazione relativa alla fase direttamente dal segnale ricevuto.

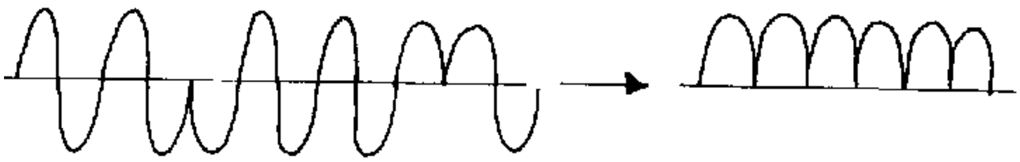
Consideriamo dunque un *segnale PSK*, ossia fondamentalmente una sinusoide, a frequenza prefissata, che presenta dei salti di fase di π, a distanza T, ogni volta che un simbolo è diverso dal precedente:



Se f_0 è la frequenza della portante, abbiamo visto che lo spettro del segnale modulato è a forma di $\sin(f)/f$ centrato sulla frequenza f_0. Allora, sembrerebbe sufficiente mandare questo segnale in ingresso ad un filtro passa-banda molto stretto centrato sulla frequenza f_0: ci si aspetterebbe, in uscita, un segnale approssimativamente sinusoidale a frequenza f_0. Al contrario, in uscita dal filtro non si ottiene niente, proprio a causa dei ripetuti salti di fase del segnale. Possiamo infatti vedere la cosa in modo intuitivo: con probabilità 0.5 noi "diciamo" al filtro di dare in uscita una sinusoide con fase 0 e con probabilità 0.5 gli diciamo di dare una uscita una sinusoide con fase 180°; dato, però, che il filtro ha una banda molto stretta, e quindi una risposta all'impulso molto ampia, esso media i comandi in ingresso su un intervallo temporale molto lungo, per cui alla fine fornisce una uscita praticamente nulla.

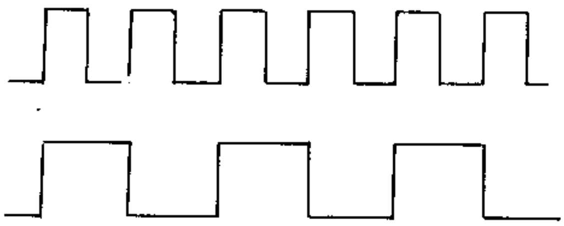
A questo punto, se vogliamo in uscita una sinusoide che si possa sfruttare, dobbiamo necessariamente ricorrere ad una operazione non lineare. Dato, allora, che i problemi sono i salti di fase di π, che nel caso della sinusoide equivalgono semplicemente ad un cambiamento di segno, ci basta raddrizzare. Se raddrizziamo a doppia semionda, otteniamo una sequenza perfettamente periodica di semionde positive:

[12] Ricordiamo, infatti, che, se il picco del segnale antipodale è metà di quello del segnale ortogonale, la potenza media è pari ad 1/2 di quella nel caso ortogonale, corrispondente appunto ad un risparmio di 3dB. Se, invece, i picchi sono uguali, allora è la codifica ortogonale a far risparmiare i 3dB di potenza media.



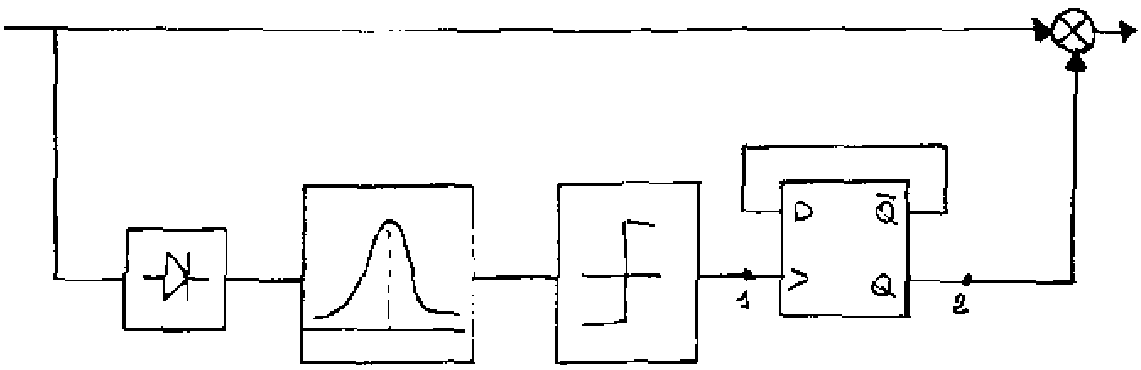
Questo segnale è periodico e quindi ha spettro a righe; inoltre, esso contiene notoriamente una componente a frequenza $2f_0$, che quindi possiamo tranquillamente estrarre mediante un successivo filtraggio passa-banda molto stretto.

Quindi, con un raddrizzamento a doppia semionda e un successivo filtraggio passa-banda molto stretto, otteniamo una sinusoide a frequenza $2f_0$. D'altra parte, a noi serve una sinusoide sincrona con la portante, ossia a frequenza f_0, per cui dobbiamo effettuare una divisione di frequenza. Questa operazione può essere effettuata utilizzando un **flip-flop di tipo D**: infatti, sappiamo che, con le opportune connessioni, usando come segnale di clock un'onda quadra di periodo τ, l'uscita Q del flip-flop è a sua volta un'onda quadra di periodo 2τ.



Questo equivale ad una divisione per 2 della frequenza del clock, per cui possiamo procedere nel modo seguente: data la sinusoide a frequenza $2f_0$ in uscita dal filtro, la squadriamo, usando un semplice *comparatore*, in modo da ottenere un'onda quadra con la stessa frequenza (e quindi con periodo $\tau=1/2f_0$); usando questo segnale come clock del flip-flop, otteniamo in uscita un'onda quadra di periodo 2τ, ossia di frequenza f_0. Da quest'onda quadra potremmo successivamente isolare la componente armonica a frequenza f_0, ma non è necessario, in quanto sappiamo che, usando direttamente questo segnale come oscillazione locale, la demodulazione coerente funziona ugualmente: infatti, oltre ad estrarre lo spettro del segnale modulante, il demodulatore coerente (cioè fondamentalmente il moltiplicatore), produce repliche traslate di questo stesso spettro, che vengono però eliminate dal successivo filtraggio passa-basso con banda B pari a quella del segnale modulante.

Lo schema definitivo, per la generazione dell'oscillazione locale, è il seguente:



Così facendo, abbiamo ottenuto un segnale, sia pure non sinusoidale, alla frequenza f_0 desiderata, ma non siamo certi di quale sia la fase, che può essere 0 oppure π [13]. Il fattore che determina l'una o l'altra fase è lo stato del circuito all'inizio di tutta procedura.

Quindi, il sistema appena descritto fornisce una oscillazione a frequenza f_0 , ma tale oscillazione può essere coerente o alla fase 0 oppure alla fase π. La conseguenza è che dalla demodulazione si ottiene o la sequenza corretta di 1 e 0 oppure una sequenza complementata, dove cioè gli 1 sono diventati 0 e viceversa.

Questo problema va ovviamente risolto[14]. La soluzione adottata è la seguente: o all'inizio della trasmissione oppure di tanto in tanto, si trasmettono delle *configurazione canoniche* (dette **preamboli**), note sia al trasmettitore sia al ricevitore; quando quest'ultimo riceve e riconosce una configurazione canonica, verifica se essa ha il segno giusto oppure no: se il segno non è giusto, il ricevitore cambia segno all'oscillazione locale, riottenendo così il sincronismo esatto.

Codifica differenziale DPSK

L'uso di preamboli, descritto poco fa, per la demodulazione coerente nella trasmissione PSK funziona sempre, tranne in un caso: può infatti capitare che, a seguito di una irregolarità di funzionamento (ad esempio dovuta ad un disturbo impulsivo), si verifichi una commutazione spuria (cioè non desiderata) dell'oscillazione locale. Se questo accade, i bit saranno necessariamente tutti sbagliati da quel momento in poi.

Dobbiamo allora trovare il modo di mantenere lo schema di trasmissione descritto prima, svincolandoci però dall'indeterminazione di fase. Si può ottenere questo obbiettivo mediante la cosiddetta **codifica differenziale** (***DPSK*** - *Differential PSK*).

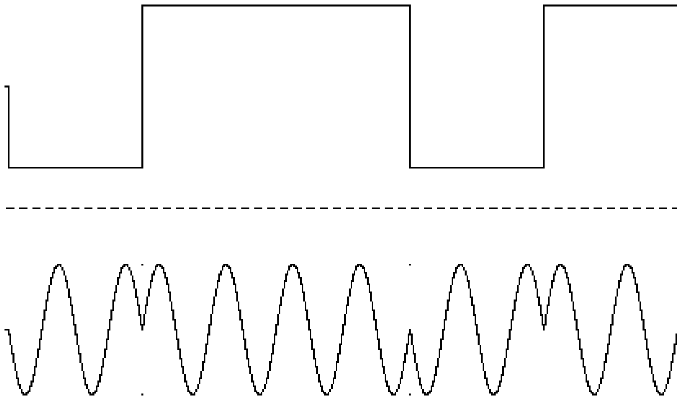
Dato che è possibile misurare, in maniera affidabile, la presenza di salti di fase nel segnale ricevuto, *si può pensare di legare l'informazione simbolica (cioè 0 o 1) non più al valore assoluto della fase, bensì alle singole transizioni di fase:* ad esempio, se c'è un salto di fase, gli si associa il simbolo 1, mentre, se non c'è un salto di fase, gli si associa il simbolo 0.

Per capire il concetto, riferiamoci ad un esempio concreto. Supponiamo che la sequenza da trasmettere sia 01101.

Nel primo caso (PSK), si moltiplica la portante sinusoidale $\cos(2\pi f_0 t)$ per un rettangolo positivo in corrispondenza della trasmissione di un 1 e per un rettangolo negativo per la trasmissione di 0. Allora, per trasmettere la sequenza 01101 si ottiene quanto segue:

[13] Sussiste un principio generale: volendo usare quel meccanismo per ottenere una divisione di frequenza, la divisione effettivamente si ottiene, ma essa si paga con una indeterminazione di fase pari a 2π/d, dove d è il numero per cui bisogna dividere. Nel nostro, caso vogliamo dimezzare la frequenza, ossia dividere per d=2, per cui l'indeterminazione è di π. Non sappiamo se la fase dell'onda quadra ottenuta ha fase 0 o fase π.

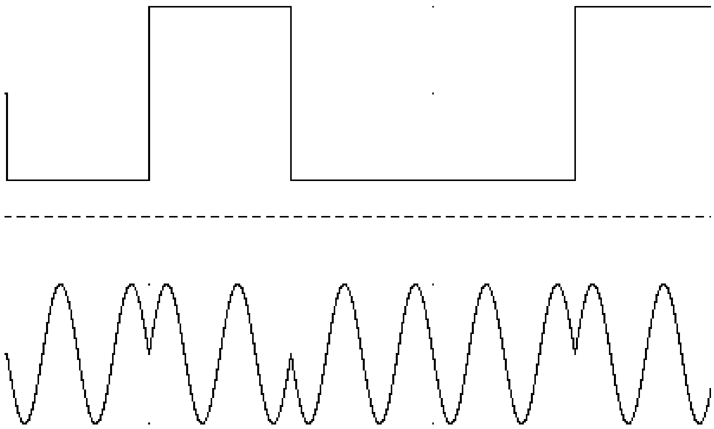
[14] Osserviamo che tutto il procedimento appena descritto si potrebbe comunque evitare se, oltre al segnale modulato, venisse inviato, dal trasmettitore sul mezzo trasmissivo, una sinusoide aggiuntiva sovrapposta al segnale, ma ovviamente sistemata al di fuori dello spettro del segnale: in ricezione, il ricevitore si deve agganciare a questa sinusoide, in modo da generare l'oscillazione locale necessaria alla demodulazione.



La forma d'onda in alto è la sequenza di rettangoli inviati al moltiplicatore, mentre la forma d'onda in basso è la portante $\cos(2\pi f_0 t)$ modulata dai suddetti rettangoli: in pratica, l'effetto dei rettangoli è quello di lasciare invariata la fase di $\cos(2\pi f_0 t)$ quando c'è da trasmettere 1 e di invertire la fase (cioè il segno) di $\cos(2\pi f_0 t)$ quando c'è da trasmettere uno 0. Così facendo, quindi, si associa l'informazione simbolica, per ogni periodo T, alla fase della sinusoide in quel periodo.

Al contrario, nella codifica DPSK, l'informazione viene ricavata dalle variazioni di fase: *in corrispondenza della trasmissione di 1 si fa cambiare la fase della sinusoide rispetto alla fase del periodo precedente; in corrispondenza della trasmissione di 0, invece, la fase resta immutata rispetto a quella del periodo precedente.*

Nel caso della sequenza 01101, supponendo che la fase corrispondente al primo bit (0) della sequenza sia la stessa della figura precedente[15], abbiamo quanto segue:

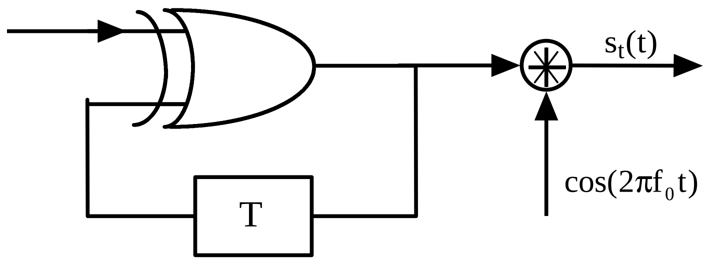


I primi due periodi di forma d'onda sono gli stessi, dopo di che subentra la differenza dovuta proprio al diverso principio.

Per ottenere questo secondo tipo di codifica (**DPSK**, informazione associata alle variazioni di fase), non si tratta più di scegliere un rettangolo positivo per l' 1 e uno negativo per lo 0: *se si vuole trasmettere un 1, bisogna far variare la fase rispetto al simbolo precedente e quindi il rettangolo dovrà essere di segno opposto al precedente; se si vuole trasmettere uno 0, invece, la fase deve rimanere invariata rispetto al simbolo precedente e quindi il rettangolo dovrà essere dello stesso segno del precedente.*

[15] E' necessario fare questa ipotesi preliminare, per due motivi: in primo luogo, in base al principio su cui si basa la codifica DPSK, la forma d'onda da associare ad un simbolo dipende da quella associata al simbolo precedente, per cui, non essendo nota questa, dobbiamo fare una ipotesi di partenza; in secondo luogo, si è scelta la stessa fase usata nella figura precedente al fine di semplificare il confronto.

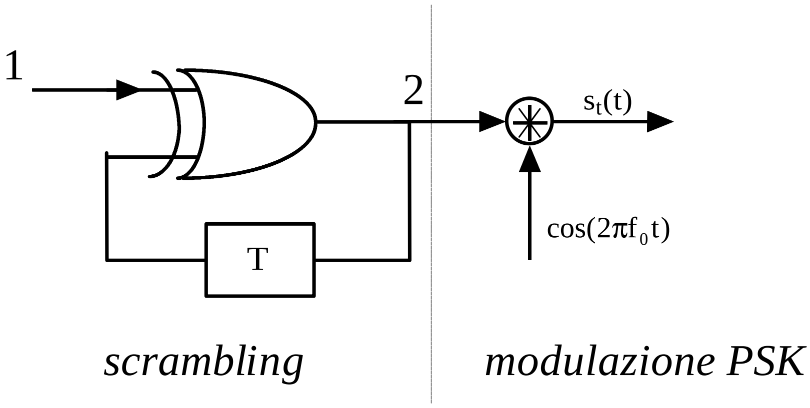
Da un punto di vista pratico, l'operazione da eseguire è descritta nella figura seguente:



Il moltiplicatore riceve in ingresso la portante sinusoidale $\cos(2\pi f_0 t)$ che deve moltiplicare per una forma d'onda rettangolare (sempre di durata T); se il bit da inviare è uguale a quello trasmesso del periodo precedente, allora la forma d'onda usata per il bit precedente deve essere usata anche per il bit in esame; se, invece, il bit in esame è diverso dal precedente, allora il rettangolo usato nel periodo precedente va adesso complementato.

Questa operazione è una cosiddetta *negazione condizionata*, realizzabile mediante una porta *XOR*, come nella figura.

In effetti, possiamo vedere la modulazione DPSK come una sequenza di due operazioni successive: data la sequenza di bit da trasmettere (punto 1 nella figura seguente), usiamo un circuito che modifichi tale sequenza in modo opportuno; successivamente, la sequenza modificata (punto 2 nella figura seguente), costituisce una nuova sequenza di bit da modulare secondo la tecnica PSK:



Abbiamo quindi effettuato uno **scrambling** della sequenza da trasmettere, dopo di che viene effettuata una normalissima modulazione PSK. Lo scrambling è tale da modificare la sequenza originale nel modo seguente: se il bit considerato è uguale al precedente, per cui non ci deve essere alcuna variazione di fase, lo scrambling fornisce uno 0, mentre invece, se è diverso dal precedente, per cui ci deve essere una variazione di fase di π, lo scrambling fornisce un 1.

Per esempio, consideriamo quello che abbiamo fatto nell'esempio precedente, dove la sequenza informativa da trasmettere era

$$\underset{\uparrow}{0} \quad 1 \quad 1 \quad 0 \quad 1$$

Supponiamo di aver appena trasmesso il primo bit della sequenza (lo 0 indicato dalla freccia) mediante un impulso negativo (corrispondente ad uno 0 per la PSK); il bit successivo è un 1, che deve indicare una variazione di fase sulla portante; ci basta usare un rettangolo di segno opposto al precedente, cioè un rettangolo positivo (corrispondente ad un 1 per la PSK). Il simbolo successivo è un altro 1, per cui dobbiamo nuovamente cambiare la fase della portante, usando un impulso di segno opposto, cioè un rettangolo negativo (corrispondente ad uno 0 per la PSK). Arriva poi uno 0, che deve lasciare invariata la fase della portante rispetto al periodo precedente: dobbiamo allora usare un altro rettangolo negativo. Infine c'è un 1, per cui dobbiamo cambiare nuovamente la fase della portante, usando un impulso positivo.

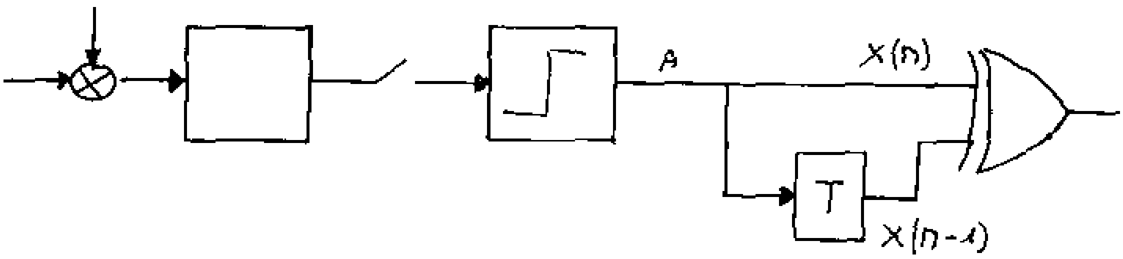
In conclusione, la sequenza di bit prodotta dallo scrambling è la seguente:

0 1 0 0 1

Questa sequenza fissa le forme d'onda rettangolari da usare per moltiplicare la portante sinusoidale.

In ricezione, dovremo evidentemente fare l'operazione duale di quella appena descritta. Dovremo sempre verificare, sul segnale ricevuto, la permanenza o la variazione di fase tra due periodi successivi e scegliere di conseguenza la fase dell'oscillazione locale. Fatta questa scelta, avremo ottenuto una demodulazione coerente (in frequenza, ma non in più in fase, che non ci interessa in quanto abbiamo accertato di non poterla controllare), dopo la quale potremo porre il filtro di ricezione, il campionatore ed il decisore.

Lo schema è il seguente:

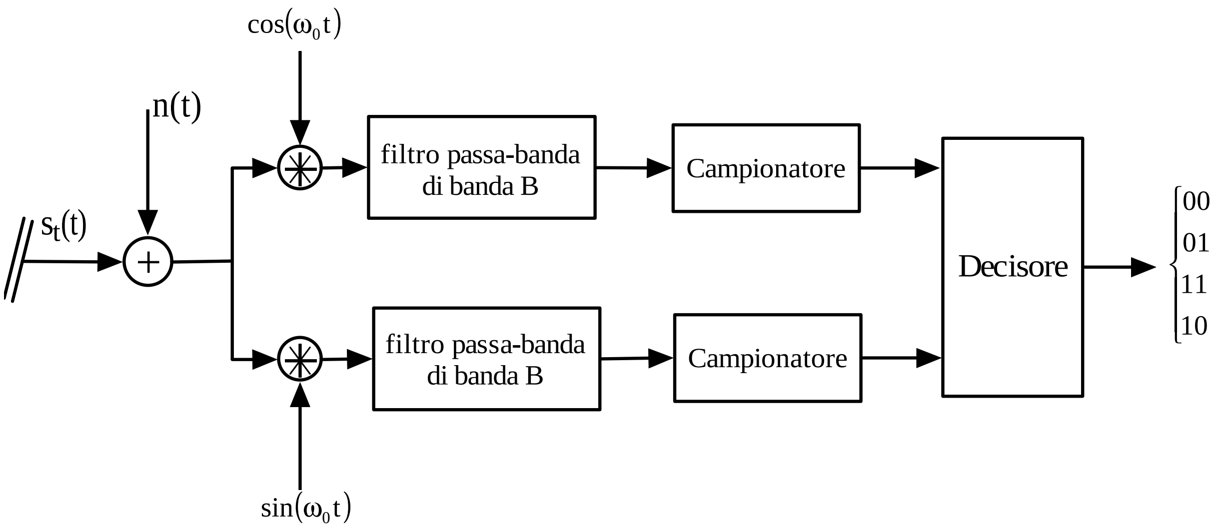


Ovviamente, all'uscita del decisore (cioè nel punto A indicato nella figura), non otteniamo la sequenza di bit di informazione che abbiamo trasmesso (cioè i bit nel punto 1 della figura precedente), bensì la sequenza di bit dopo lo *scrambling*, ossia i bit in ingresso al modulatore PSK (cioè nel punto 2 della figura precedente). Ciò significa che dobbiamo adesso fare una operazione che consenta di passare dalla sequenza nel punto 2 alla sequenza nel punto 1, cioè la sequenza originaria. Basta allora confrontare l'ultimo bit con il precedente: se sono uguali, allora il bit di informazione è 0, mentre in caso contrario si tratta di un 1.

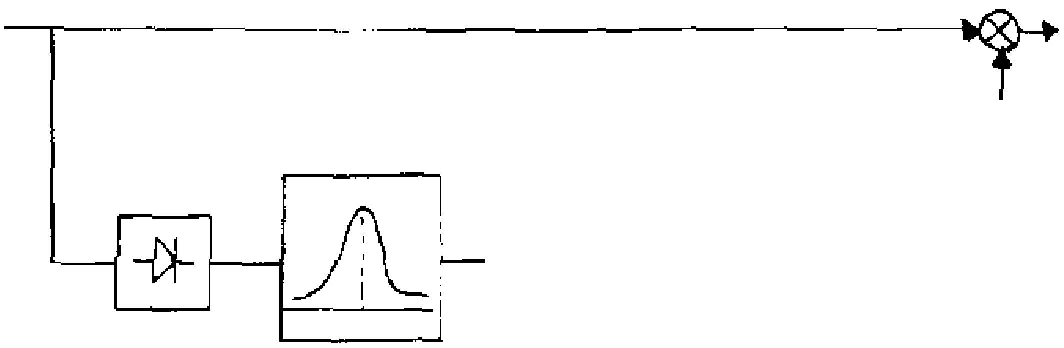
Con questo procedimento, ci siamo sganciati dall'indeterminazione di fase legata alla divisione di frequenza (da $2f_0$ a f_0), ma abbiamo ottenuto anche uno svantaggio: se il decisore commette un errore su un bit, questo errore risulta quanto meno duplicato, in quanto risulterà sicuramente sbagliata anche la decisione successiva.

Demodulazione nel PSK multilivello e M-DPSK

Tutto il discorso appena fatto può essere generalizzato anche al sistema **M-PSK**, nel quale cioè si usano due portanti in quadratura, ciascuna delle quali è modulata in ampiezza (e quindi in fase) da un segnale numerico a $b = \log_2 M$ ampiezze. Anche in questo caso, infatti, si pone il problema della demodulazione coerente in ricezione:



Dobbiamo questa volta generare, in ricezione, sia un riferimento sincrono col Coseno sia un riferimento sincrono col Seno. In questo caso, però, le fasi possibili del segnale $s_t(t)$ ricevuto dal mezzo trasmissivo non sono più 2, ma in numero M: per esempio, in un sistema M-PSK, abbiamo 4 possibili fasi (45°,135°,225°,315°). Se prendiamo il segnale $s_t(t)$, lo raddrizziamo e lo filtriamo, otteniamo ancora una sinusoide a frequenza $2f_0$, ma essa sarà ancora modulata PSK con fase 0°/180°:



Abbiamo dunque bisogno di raddrizzare ulteriormente per giungere ad una definizione non ambigua della fase. Ovviamente, dopo il secondo raddrizzamento, la componente spettrale che ci interessa è a $4f_0$, per cui abbiamo bisogno ancora una volta di una divisione di frequenza. Il problema è che, dovendo dividere per 4, abbiamo bisogno di 2 stadi flip-flop in cascata, per cui l'ambiguità sulla fase è adesso di $2\pi/d=2\pi/4=\pi/2$. Ci ritroviamo, perciò, nella stessa condizione del PSK a due sole fasi, in quanto possiamo nuovamente confondere una combinazione binaria con qualsiasi altra.

Dobbiamo allora ricorrere nuovamente ad una codifica differenziale (per cui parleremo di **M-DPSK**), ossia dobbiamo associare l'informazione simbolica alle variazioni di fase, tra un periodo e l'altro, della portante modulata e non più alla fase assoluta, in ciascun periodo, della stessa portante. Dato che in un sistema 4-PSK trasmettiamo coppie di bit in ogni periodo T, dovremo associare ciascuna coppia ad una variazione di fase: le possibili variazioni di fase sono adesso 4 (0°,90°,180°,270°) e a ciascuna variazione corrisponderà una coppia di bit.

Anche in questo caso, dunque, abbiamo la possibilità di sganciarci dalla indeterminazione nella identificazione della fase, fermo restando, però, che questo si paga con una possibile moltiplicazione degli errori.

OSSERVAZIONE: ULTERIORI VANTAGGI DEL SISTEMA NUMERICO MULTILIVELLO

Abbiamo visto in precedenza quale sia un grosso vantaggio di un sistema numerico multilivello e cioè la possibilità di aumentare la velocità di trasmissione pur lasciando invariata la banda: ad esempio, avendo a disposizione una banda B passa-basso per trasmettere, possiamo trasmettere alla massima velocità di 2B bit al secondo, mentre invece, avendo a disposizione una banda 2B passa-banda, possiamo trasmettere, per esempio, a velocità di 6B bit al secondo con una modulazione 8-QAM (equivalente ad una 8-PSK).

Un altro motivo che ci potrebbe indurre ad usare un sistema multilivello è il seguente: supponiamo di voler trasmettere ad una prefissata frequenza di cifra f_S, il che significa un numero prefissato di bit al secondo; se usiamo una codifica binaria (cioè 2 soli livelli), dobbiamo trasmettere tante forme d’onda, per unità di tempo, quanti sono i bit, per cui risulta in pratica fissata la *durata equivalente* della forma d’onda trasmessa. Allora, se vogliamo ridurre la banda occupata, non possiamo far altro che aumentare la durata equivalente e questo lo otteniamo ancora una volta con un sistema multilivello: ad esempio, passando da un sistema a 2 soli livelli ad uno a 4 livelli, conservando la stessa velocità di trasmissione, possiamo raddoppiare la durata della singola forma d’onda.

Il fatto di aumentare la durata temporale, e quindi di stringere la banda, si ripercuote anche sul filtro di ricezione, che avrà una banda minore, ossia integrerà il segnale in ingresso per un intervallo di tempo più lungo. Questa operazione di integrazione ha un effetto diverso sul segnale utile e sul rumore ad esso sovrapposto: infatti, quanto più lungo è il tempo di integrazione, tanto minore è il contributo del rumore sul segnale, dato che quest’ultimo mantiene la propria coerenza durante tutto il periodo di cifra, mentre il rumore no, data la sua casualità.

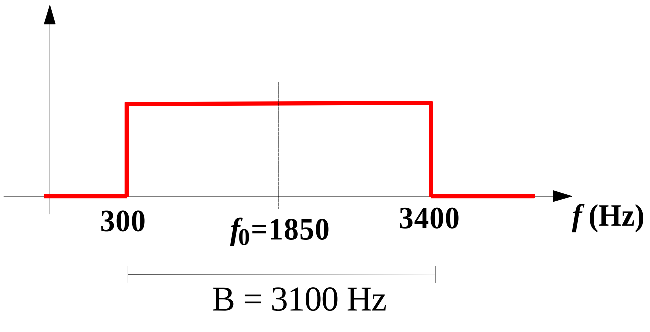
Quindi, riassumendo, il sistema multilivello può essere dimensionato secondo due criteri:

- *se la banda è fissata, lo si dimensiona per arrivare ad una velocità di trasmissione maggiore;*
- *se, invece, la velocità di trasmissione è fissata, allora lo si dimensiona per stringere la banda occupata e quindi aumentare il tempo di integrazione in ricezione.*

ESEMPIO: TRASMISSIONE NUMERICA SU CANALE TELEFONICO IDEALIZZATO

Consideriamo un **canale telefonico idealizzato**, ossia un mezzo trasmissivo passa-banda con attenuazione costante nella banda compresa tra 300 Hz e 3400 Hz:

canale telefonico idealizzato

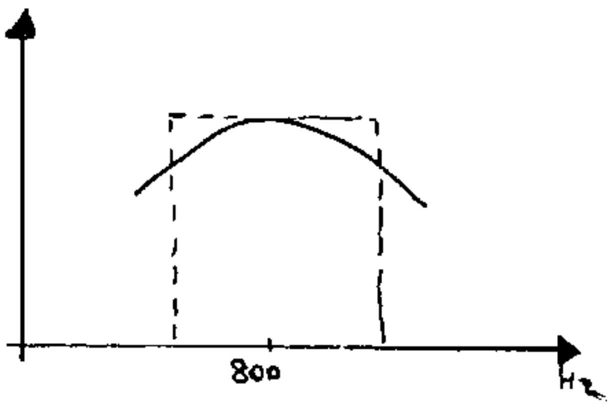


Vogliamo effettuare una trasmissione numerica, su questo mezzo trasmissivo, del segnale telefonico e vogliamo dimensionare il sistema in modo da ottenere una probabilità di errore $p(\varepsilon)=10^{-9}$ (il che significa sbagliare 1 bit ogni miliardo).

Nel risolvere questo esercizio, consideriamo la potenza media come parametro qualificante della potenza di segnale. A rigor di logica, questa non è una scelta corretta, in quanto è noto che la potenza media emessa da un parlatore non è una grandezza deterministica, visto che dipende dalle condizioni fisiche e di umore del parlatore stesso. D'altra parte, però, è anche noto che, nella trasmissione telefonica, non viene mai trasmessa una sola conversazione per volta, ma un certo numero di conversazioni in contemporanea (usando tecniche a divisione di frequenza o, specialmente nella trasmissione numerica, a divisione di tempo): di conseguenza, se il canale telefonico che noi consideriamo è uno dei tanti canali che finiscono in un sistema di trasmissione multiplo, allora la potenza che l'amplificatore deve essere in grado di amplificare può essere valutata considerando appunto la potenza media, effettuando cioè una media delle potenze emesse da tutti i possibili parlatori. Considerando che ci sono sicuramente parlatori forti e parlatori deboli, il principio è congruente con la realtà.

A questo punto facciamo un'altra ipotesi: supponiamo che il dimensionamento del trasmettitore sia tale che il canale utilizzato garantisca, in uscita, un rapporto segnale-rumore di **30 dBp**, inteso come rapporto tra la potenza media di segnale (in base alle considerazioni di poco fa) e la potenza media di rumore pesato.

Cominciamo allora a chiederci perché si parla di *dB pesati* (**dBp**). Il motivo è che l'udito dell'uomo non presenta una sensibilità costante per tutte le frequenze, ma, al contrario, presenta la massima sensibilità in corrispondenza della frequenza di circa 800 Hz (cioè approssimativamente della frequenza centrale del *campo udibile*, che va notoriamente da 20 Hz a 16 kHz) e poi la sensibilità decresce andando verso i due estremi. Un andamento schematico di come varia tale sensibilità con la frequenza è riportato nella figura seguente:



In base a queste considerazioni, *bisogna supporre che, in ingresso al mezzo trasmissivo, ci sia un filtro che tenga conto della sensibilità dell'orecchio umano, così come facevamo, nel caso del segnale televisivo, per tener conto della sensibilità dell'occhio alle variazioni di colore.* Dobbiamo cioè osservare che, *nella misura della potenza di rumore, è inutile pesare allo stesso modo tutte le componenti spettrali di rumore, mentre è opportuno pesare poco (praticamente trascurare) le componenti in bassa e in alta frequenza, alle quali l'orecchio umano è praticamente "sordo".*

Mentre nel caso del segnale televisivo parlavano di *filtro videometrico*, adesso parliamo di **filtro psofometrico**; in particolare, ci interessa definire la **banda equivalente di rumore** del filtro psofometrico, notoriamente definita come la banda di un filtro con funzione di trasferimento rettangolare che racchiude la stessa area sottesa dalla funzione di trasferimento effettiva. Si trova che tale banda equivalente di rumore, nel caso del filtro psofometrico, vale $\mathbf{B_N=1.7kHz}$.

Questo dato ci consente di eliminare la *pesatura psofometrica del rumore*, passando dalla definizione del rapporto segnale-rumore pesato (cioè potenza di segnale rapportata alla potenza di rumore pesata nella banda B_N) al rapporto tra la potenza media di segnale P_R e la potenza di rumore pesato questa volta nella banda B effettiva del canale usato per la trasmissione:

$$\frac{P_R}{h_n B_N} = 30\mathrm{dBp} \xrightarrow{\text{eliminando la pesatura psofometrica}} \frac{P_R}{h_n B} = \_\_\,\mathrm{dB}$$

Ovviamente, la banda B da considerare è quella del canale telefonico, che abbiamo detto essere ampia 3100 Hz. Facendo i calcoli, abbiamo quanto segue:

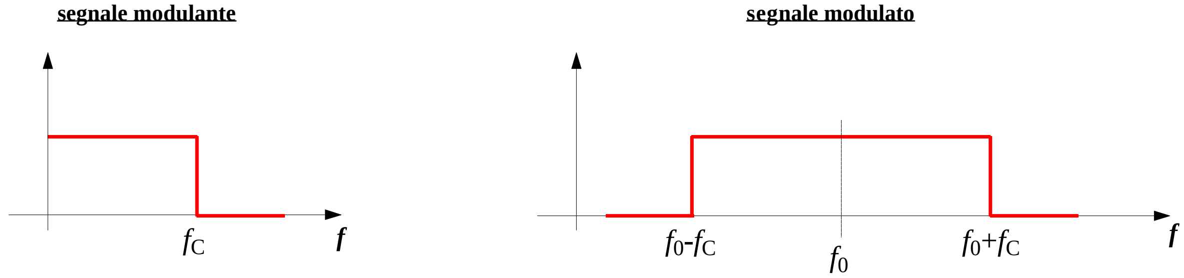
$$\frac{P_R}{h_n B_N} = 10^{\frac{30}{10}} = 10^3 \longrightarrow \frac{P_R}{h_n} = 10^3 B_N = 1.7 \cdot 10^6 \longrightarrow \frac{P_R}{h_n} = 1.7 \cdot 10^6 = x \cdot B = x \cdot \left(3.1 \cdot 10^3\right)$$

da cui ricaviamo quindi che $x = \dfrac{1.7 \cdot 10^6}{3.1 \cdot 10^3} = 548.4 \Leftrightarrow 27.4\mathrm{dB}$.

Arrotondando a 5 la cifra decimale, possiamo dunque concludere che il rapporto S/N garantito all'uscita del mezzo trasmissivo è

$$\left.\frac{S}{N}\right|_{OUT} = \frac{P_R}{h_n B} = 27.5\mathrm{dB}$$

Abbiamo adesso tutte le specifiche necessarie riguardo l'uscita del mezzo trasmissivo. Dobbiamo adesso scegliere il tipo di modulazione da usare: la modulazione si rende chiaramente necessaria in quanto il canale è passa-banda, mentre il segnale numerico che intendiamo trasmettere è un segnale passa-basso. Scegliamo allora una modulazione **PSK a 2 livelli** (equivalente ad una modulazione di ampiezza se le forme d'onda modulanti sono dei rettangoli antipodali). Questo significa, come sappiamo, che, se f_C è la banda del segnale che vogliamo trasmettere, dopo la modulazione otterremo un segnale di banda $2f_C$, con le due bande laterali disposte simmetricamente rispetto alla frequenza f_0 della portante:



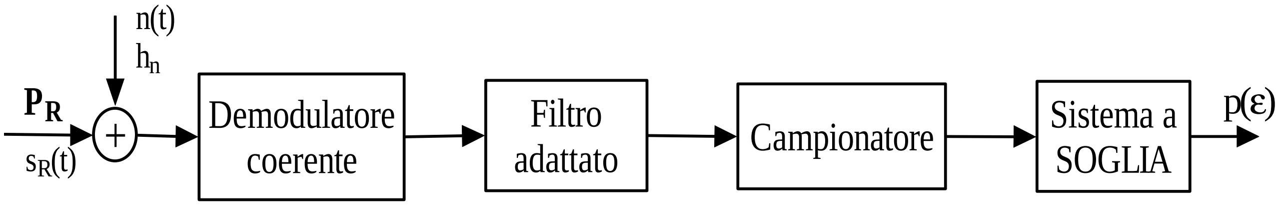
E' allora ovvio che, se vogliamo sfruttare a pieno la banda che abbiamo a disposizione sul mezzo trasmissivo, dobbiamo porre f_0 a metà di tale banda, per cui sarà

$$f_0 = \frac{300 + 3400}{2} = 1850 \;\; \mathrm{Hz}$$

Con questa scelta, abbiamo a disposizione, per il segnale modulante, una banda pari esattamente a metà di quella del mezzo trasmissivo, ossia

$$f_C = \frac{B}{2} = \frac{3100}{2} = 1550 \;\; \mathrm{Hz}$$

Il segnale modulante dovrà dunque avere questa banda. In effetti, quindi, *possiamo adesso procedere assimilando il sistema con modulazione PSK come un sistema di trasmissione in banda base, dove però la banda vale adesso B_B=1550 Hz.* Di conseguenza, possiamo procedere, nei conti, così come abbiamo fatto in un esercizio precedente, salvo però a considerare, in questo caso, un elemento in più in ricezione, vale a dire il *demodulatore coerente* posto immediatamente a valle del mezzo trasmissivo:



Per prima cosa, dobbiamo passare dalla p(ε) a valle del decisore al rapporto segnale-rumore a valle del campionatore: ipotizzando, come sempre fatto nei discorsi precedenti, che il rumore sovrapposto sia quello termico e decidendo subito di porre la soglia a metà, sappiamo che la formula cui fare riferimento è

$$p(\varepsilon) = P(\varepsilon\,|\,1) = P(\varepsilon\,|\,0) = Q(\gamma) = \int_{\gamma}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

Il passaggio avviene dunque fondamentalmente da p(ε) alla quantità γ, che ricordiamo essere, in generale, pari a $\dfrac{V_1 - V_0}{2\sigma_n}$.

Fissata p(ε), dobbiamo conoscere γ e quindi dobbiamo conoscere come la funzione Q(γ) varia con γ stesso. Ricordiamo, allora, che, volendo ridurre p(ε)=Q(γ) di un fattore 10, è necessario aumentare γ di 1 dB. Noi dobbiamo trovare quale valore di γ serve per ottenere Q(γ)=p(ε)=10^{-9}. Dalle apposite tabelle, si trova che, per ottenere p(ε)=10^{-6}, è necessario prendere γ=13.5 dB. Allora, in base a quanto appena detto, per ottenere p(ε)=10^{-9}, dobbiamo prendere **γ=16.5 dB**, ricordando che $(\gamma)_{dB} = 20\log_{10}\gamma$.

Noto γ (che in pratica rappresenta il rapporto segnale/rumore, in termini di tensione, a valle del campionatore), possiamo associare questo parametro al rapporto segnale/rumore $\left.\dfrac{S}{N}\right|_U$ a valle del filtro di ricezione.

Qui, però, dobbiamo operare una ulteriore scelta, scegliendo tra una codifica ortogonale e una codifica antipodale:

$$\text{antipodale} \longrightarrow \left.\frac{S}{N}\right|_U = \frac{P_{S,p}}{P_{N,m}} = \gamma^2$$

$$\text{ortogonale} \longrightarrow \left.\frac{S}{N}\right|_U = \frac{P_{S,p}}{P_{N,m}} = 4\gamma^2$$

dove $P_{S,p}$ rappresenta la potenza di picco del segnale all'uscita del filtro di ricezione e $P_{N,m}$ rappresenta la potenza media del rumore all'uscita dello stesso filtro.

Optiamo, ad esempio, per una *codifica antipodale*, per cui la definizione da considerare è la prima delle due appena elencate. Per passare all'ingresso del filtro di ricezione, dobbiamo scegliere il tipo di filtro, ossia la funzione di trasferimento. A questo scopo, consideriamo allora che siamo in un contesto in cui si presenta una limitazione di potenza in termini di potenza media, per cui il filtro ottimale è il *filtro adattato*: per questo filtro, sappiamo che il rapporto S/N in uscita è numericamente pari al rapporto S/N in ingresso, a patto di definire quest'ultimo come rapporto tra la potenza media di segnale e la potenza di rumore che cade nella banda convenzionale $f_S/2$:

$$\left.\frac{S}{N}\right|_U = \frac{P_R}{h_n \frac{f_S}{2}}$$

Possiamo dunque scrivere, con riferimento a quanto trovato prima, che

$$\left.\frac{S}{N}\right|_U = \left.\frac{S}{N}\right|_{IN} = \frac{P_R}{h_n \frac{f_S}{2}} = \gamma^2 = 16.5\text{dB}$$

Prima di andare avanti, ci conviene fissare il valore della *frequenza di cifra* f_S (cioè del numero di simboli trasmessi nell'unità di tempo). A tal fine, dobbiamo considerare che, ancora una volta, facciamo un *progetto alla Nyquist*, per cui vogliamo forme d'onda ad intersimbolo nullo in uscita dal filtro adattato. Di conseguenza, la frequenza cifra è determinata, essendo fissata la banda B_B del sistema, dal valore del *roll off* δ:

$$B_B = \frac{f_S}{2}(\delta + 1)$$

Allo scopo di ottenere la massima velocità di trasmissione, ci poniamo nel caso limite di δ=0, da cui consegue che $f_S = 2B_B = 2 \cdot 1550 = 3100$ bit/sec.

Possiamo dunque scrivere, con riferimento all'espressione di $\left.\frac{S}{N}\right|_{IN}$, che

$$\left.\frac{S}{N}\right|_{IN} = \frac{P_R}{h_n \frac{3100}{2}} = \frac{P_R}{h_n \cdot 1550} = 16.5\text{dB}$$

A questo punto, dobbiamo passare dall'ingresso del filtro adattato all'ingresso del *demodulatore coerente*, in modo da ottenere il rapporto S/N all'uscita del mezzo trasmissivo.

Il discorso da fare è assolutamente identico a quello seguito nel caso analogico:

- per quanto riguarda la potenza media di segnale P_S all'ingresso del demodulatore, sappiamo che essa è metà della potenza media P_R all'uscita del demodulatore stesso: infatti, la demodulazione produce un segnale che non ha più le oscillazioni sinusoidali, per cui, dato che il fattore di picco di una sinusoide è 1/2, il rapporto tra la potenza media della portante modulata e la potenza media del segnale modulante è 1/2;
- stesso discorso per la potenza media di rumore: se h_n è la densità spettrale di rumore all'uscita del demodulatore, quella in ingresso è $h'_n = h_n / 2$.

Quindi, all'ingresso del demodulatore, ossia all'uscita del mezzo trasmissivo, abbiamo

$$\left.\frac{S}{N}\right|_{OUT} = \frac{P_S}{h'_n \cdot 3100} = \frac{\frac{P_R}{2}}{\frac{h_n}{2} \cdot 3100} = \frac{1}{2}\frac{P_R}{h_n \cdot 1550} = \frac{1}{2}\left.\frac{S}{N}\right|_{IN}$$

Abbiamo dunque trovato che il rapporto S/N in ingresso al demodulatore è metà di quella in uscita dal demodulatore stesso: dato che in uscita avevamo trovato prima che $\left.\frac{S}{N}\right|_{IN} = 16.5\text{dB}$, deduciamo che in ingresso deve risultare

$$\left.\frac{S}{N}\right|_{OUT} = 16.5\text{dB} - 3\text{dB} = 13.5\ \text{dB}$$

La conclusione cui siamo pervenuti è dunque quella per cui ci servono 13.5 dB di rapporto S/N in uscita dal mezzo trasmissivo per ottenere una probabilità di errore 10^{-9}. C'è però da fare una osservazione: tutto questo discorso vale se il segnale modulante inviato in ingresso al modulatore coerente è rettangolare, per cui la modulazione PSK è equivalente alla modulazione ASK; al contrario, se le forme d'onda modulanti sono di altro tipo, allora la modulazione è rigorosamente una ASK; la differenza tra i due casi è importante: infatti, se si rinuncia al filtro adattato perché si vogliono trasmettere forme d'onda rettangolari, bisogna considerare che, tra l'ingresso e l'uscita del filtro di ricezione c'è una perdita, di circa 0.5dB, dovuta al *fattore di forma*. Sotto questa ipotesi, quindi, se vogliamo ottenere i 16.5 dB all'uscita del filtro di ricezione, dovremo considerare non più 16.5 dB anche in ingresso, ma 17 dB. Di conseguenza, il rapporto S/N in uscita dal mezzo trasmissivo non dovrà essere più di 13.5 dB, ma di **14 dB**.

Ci accorgiamo, a questo punto, che, a fronte della nostra richiesta di 14dB, il mezzo trasmissivo fornisce un rapporto segnale/rumore di 27.5dB (in base ai calcoli fatti all'inizio). Abbiamo perciò a disposizione un sovrappiù di potenza di 13.5dB, che ci conviene impiegare in qualche modo: dato che la banda a disposizione è fissata, possiamo pensare di aumentare la velocità di trasmissione mettendo insieme un *sistema multilivello*.

Un modo per realizzare un sistema multilivello è quello di usare un semplice sistema QAM, che utilizza due portanti in quadratura, ciascuna modulata da un segnale numerico: se, per ogni portante, utilizziamo un *normale* segnale modulante numerico a 2 livelli, otteniamo un sistema ad M=4 livelli, nel quale la velocità di trasmissione è raddoppiata: mentre prima trasmettevamo 3100 bit/sec, adesso trasmettiamo 6200 bit/sec, a parità di banda occupata.

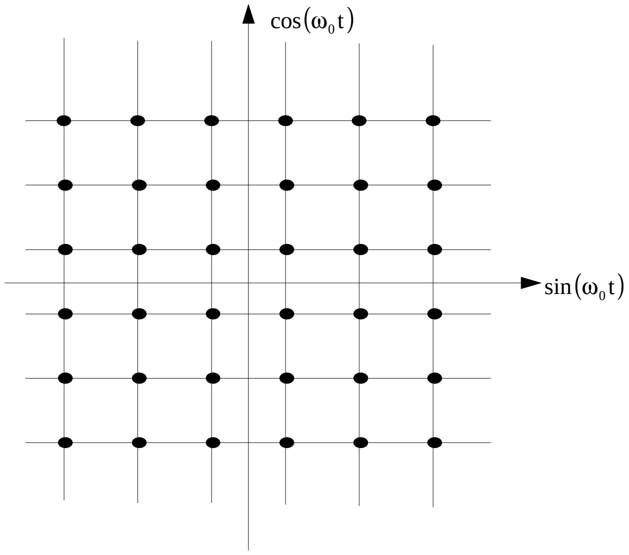
Vediamo allora se i 13,5dB di potenza a disposizione sono sufficienti per implementare un sistema a 4 livelli. E' immediato fare questa verifica, in quanto basta ricordare che un sistema QAM necessita solo di 3dB in più di potenza in trasmissione rispetto ad un sistema a 2 soli livelli: infatti, dobbiamo trasmettere due segnali numerici, con le stesse caratteristiche, sulle due portanti ortogonali tra loro, per cui, dato che la potenza del segnale somma di due segnali ortogonali è pari alla somma delle potenze, dobbiamo raddoppiare la potenza media rispetto ad un solo canale, ossia aumentarla di 3dB.

Quindi, usando un semplice sistema QAM, spendiamo 3dB in più rispetto a prima, per cui abbiamo ancora a disposizione $14(\text{dB}) - 3(\text{dB}) = 11(\text{dB})$ di potenza. Possiamo allora pensare di utilizzare questi dB di potenza disponibile realizzando un sistema M-QAM (QAM multilivello, con M livelli per portante), ossia un sistema nel quale i due segnali modulanti hanno più di 2 ampiezze. Infatti, ci ricordiamo della formula, introdotta proprio nella parte sui sistemi numerici multilivello, secondo cui l'incremento di potenza media, conseguente all'uso di un sistema a M livelli, rispetto ad un sistema a 2 soli livelli è dato da $\frac{M^2-1}{3}$. Ci basta quindi imporre che questo incremento sia minore o tutt'al più uguale agli 11dB di potenza a nostra disposizione:

$$\frac{M^2-1}{3} \leq 10^{\frac{11}{10}} = 10^{1.1} \longrightarrow M^2 \leq 3 \cdot 10^{1.1} + 1 = 38.77 \longrightarrow M \leq 6.22$$

Se vogliamo che M sia una potenza di 2, è ovvio che i livelli che possiamo utilizzare sono M=4. Questi sono però i livelli per ciascuna portante; avendo 2 portanti in quadratura, abbiamo ottenuto un sistema **16-QAM**, nel quale, quindi, la velocità di trasmissione è quadruplicata rispetto al sistema iniziale: possiamo cioè trasmettere con una frequenza di cifra $f_S = 4 \cdot 3100 = 12400$ bit/sec, ossia un valore quadruplo rispetto al baud rate (che ricordiamo essere pari alla frequenza di cifra solo nel caso di sistema 2 soli livelli).

Tornando adesso al dimensionamento, possiamo andare anche oltre le prestazioni ottenute poco fa: infatti, dato che deve essere M≤6.22, possiamo anche pensare di considerare M=6 livelli per ciascuna portante: con 6 livelli per portante, otteniamo un sistema 36-QAM, la cui *costellazione di punti*, nel *piano dei segnali*, è fatta nel modo seguente:



Se vogliamo comunque ricondurci ad un numero di simboli pari ad una potenza di 2, possiamo pensare di lasciare inutilizzate le 4 configurazioni estreme (ottenendo tra l'altro una riduzione della potenza di picco): così facendo, quindi, otteniamo un sistema **32-QAM**, con il quale quintuplichiamo la velocità di trasmissione rispetto al sistema iniziale, arrivando quindi ad una frequenza di cifra $f_S = 5 \cdot 3100 = 135400$ bit/sec.

Questo è dunque il massimo che possiamo ottenere per la banda che ci è stata assegnata.

Possiamo adesso osservare un'altra cosa: tutto il progetto appena svolto si è basato sull'ipotesi di considerare un roll off δ=0 per le forme d'onda (ad intersimbolo nullo) in uscita dal filtro di ricezione. Sappiamo, però, che, nella pratica, δ=0 è una scelta quasi mai realizzabile; al contrario, si considera un δ non nullo, il che, ovviamente, a parità di banda a disposizione, riduce la velocità di trasmissione rispetto al limite teorico ottenibile con δ=0.

Un modo per fissare il valore di δ è proprio quello di partire da una assegnata frequenza di cifra e da una assegnata banda e di calcolare quindi il δ corrispondente. Per esempio, supponiamo che, per le nostre esigenze, sia sufficiente raggiungere una frequenza di cifra $f_S = 3 \cdot 3100 = 9600$ bit/sec [16]. Se usassimo un sistema 16-QAM, con 4 livelli per portante, otterremmo una velocità di 12400 bit/sec

[16] Questo valore della frequenza di cifra potrebbe anche derivare da limitazioni sul **modem** che usiamo per inviare i segnali sul canale: per esempio, fino a qualche anno fa, i modem raggiungevano una velocità massima di 9600 bit/sec, per cui erano questi dispositivi ad imporre la frequenza di cifra.

adottando δ=0, mentre otterremmo una velocità chiaramente minore se usassimo un δ>0; andiamo allora a scegliere il valore di δ necessario, ad ottenere, in un sistema 16-QAM, una velocità di 9600 bit/sec:

$$B_B = \frac{f_B}{2}(\delta+1) \longrightarrow \delta = \frac{2B_B}{f_B} - 1 = \frac{12400}{9600} - 1 = 0.29$$

Nel fare questo calcolo, abbiamo considerato $2B_B$=12400 bit/sec in quanto stiamo considerando un sistema 16-QAM, nel quale la massima frequenza di cifra, pari a 2 volte la banda occupata[17], deve essere 12400 bit/sec.

Quindi, in conclusione, per ottenere, mediante un sistema 16-QAM, una frequenza di cifra $f_S = 9600$ bit/sec, dobbiamo utilizzare un roll off δ=0.29.

ESEMPIO: TRASMISSIONE DEL SEGNALE DEL TELEVIDEO

Un tipico esempio di segnale di tipo numerico è quello contenente le informazioni del Televideo. Questo segnale viene trasmesso "*insieme*" al segnale televisivo, per cui cerchiamo in primo luogo di capire cosa significa "*insieme*".

Richiamiamo intanto alcune nozioni fondamentali circa il segnale televisivo. Intanto, gli **standard europei** circa il segnale televisivo prevedono le seguenti specifiche di base:

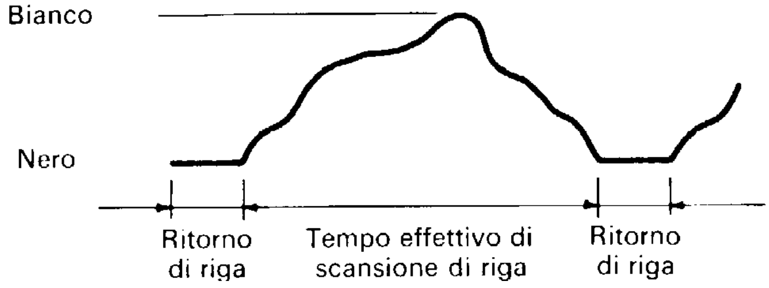
- in primo luogo, si adotta una *frequenza di quadro* di 25 Hz, il che significa che vengono trasmessi 25 quadri al secondo (in realtà, in base al sistema della scansione interlacciata, sappiamo che vengono trasmessi 50 semiquadri al secondo);
- in secondo luogo, la risoluzione verticale è fissata in 625 righe per ogni quadro.

E' chiaro quindi che vengono trasmesse $625 \cdot 25 = 15625$ righe/sec. Questo valore consente di ricavare il cosiddetto *periodo di riga*, ossia il tempo dedicato alla scansione di ciascuna riga:

$$T_{riga} = \frac{1}{15625} = 64 \frac{\mu \text{sec}}{\text{riga}}$$

Di questi 64μsec, non tutti possono essere utilizzati per contenere i segnali di luminanza (per la TV in bianco e nero) ed eventualmente di crominanza (per la TV a colori): è necessario, infatti, inserire, tra una riga e l'altra, i *segnali di sincronismo*, che consentano al pennello elettronico del televisore di individuare sia la fine di una riga sia la fine di un quadro. Per inserire questi segnali di sincronismo, si riservano 14μsec dei 64μsec teoricamente destinati a ciascuna riga: in tal modo, il tempo a disposizione per le informazioni di luminanza e crominanza è di 50 μsec (*tempo effettivo di scansione di riga*).

[17] E' importante osservare, in questo discorso, che la banda da inserire nella formula per il calcolo del δ non è tutta la banda a disposizione sul canale, ma la sua metà (cioè la cosiddetta *banda base equivalente*), dato che si usa un sistema di modulazione. Questo, ovviamente, se si usa una modulazione del tipo DSB-SC. Se la modulazione usata fosse del tipo SSB (anche se nella pratica non si utilizza mai), allora la banda da inserire nella formula è, in effetti, tutta la banda a radiofrequenza disponibile sul canale.



Negli altri 14µsec, il segnale viene mantenuto al livello del nero (black forced), al fine di recuperare la continua[18], per 9÷10 µsec, mentre il tempo rimanente (circa 5µsec) è destinato all'*impulso di sincronismo*:

La zona di nero a cavallo degli impulsi di sincronismo non serve solo a recuperare la continua, ma anche a spegnere il pennello elettronico quando ha finito di scandire una riga e deve riposizionarsi all'inizio della riga successiva. Lo stesso vale, ovviamente, per gli *impulsi di quadro*, ossia quando è necessario scandire un nuovo quadro.

Veniamo adesso al Televideo. Non tutte le 625 righe di ogni quadro vengono effettivamente visualizzate sullo schermo: le righe non visualizzate, in numero tutt'altro che trascurabile, sono dette **righe non attive** e si è pensato di utilizzarle per trasmettere informazioni di altro tipo: inizialmente, si inserirono in queste righe solo dei *segnali campione*, ovvero dei segnali di forma ben precisa che consentono, se visualizzati su un oscilloscopio posto all'uscita del televisore, di stabilire se tutta la catena di trasmissione, dal trasmettitore al ricevitore, è accettabile o meno.

Oltre ai segnali campione, si è poi pensato di inserire altre informazioni e, in particolare, si è optato per l'uso di forme d'onda che consentono di trasmettere un segnale numerico: si tratta, appunto, del **Televideo**.

E' chiaro che bisogna evitare che il televisore si accorga che si è inserito qualcosa nelle righe non attive: basta allora che le forme d'onda utilizzate non interferiscano con i segnali di sincronismo. Quindi, *possiamo inserire qualsiasi segnale, in ciascuna riga non attiva, a patto che esso sia compreso in un intervallo di tempo di 50 msec e abbia una altezza compresa tra il livello del nero ed il livello del bianco.* A questi requisiti soddisfano dunque le forme d'onda che trasportano le informazioni del Televideo.

[18] Ricordiamo il motivo per cui è necessario recuperare la continua in ricezione: il segnale televisivo no ha statistica stazionaria, per cui, a seconda della scena, il livello nero del segnale cambia. Dato che non si può effettuare un collegamento facendo passare anche la continua (per via del mezzo trasmissivo che può non essere passa-basso), è necessario interporre un condensatore prima della trasmissione: questo significa perdere il valor medio a breve periodo, per cui il segnale televisivo continua a salire o a scendere. Se questo segnale viene mandato a comandare il pennello elettronico del televisore, ci si trova una serie di strisciate che rendono il segnale inservibile. E' quindi necessario riportare il riferimento al suo esatto valore e questa informazione viene appunto inserita nell'intervallo a cavallo del sincronismo: una volta riconosciuto l'impulso di sincronismo, il ricevitore legge la tensione appena dopo il fronte di salita del sincronismo stesso e quindi sottrae tale tensione al segnale.

Fatte queste premesse, entriamo nel merito dell'esercizio, che sarà svolto con criteri analoghi a quelli seguiti nell'esercizio precedente.

Per prima cosa, supponiamo che il mezzo trasmissivo utilizzato garantisca, in uscita, un rapporto segnale/rumore di **40 dBp**, inteso come rapporto tra la tensione bianco-nero (v_{bn}) e la tensione efficace di rumore (v_n). La prima operazione da compiere è passare dai dBp ai dB, eliminando la *pesatura videometrica del rumore*: sappiamo che dobbiamo considerare 8dB in meno se il rumore sovrapposto è bianco oppure 16dB in meno negli altri casi. Supponendo che il rumore sia bianco, deduciamo che il mezzo trasmissivo garantisce, in uscita, un rapporto segnale/rumore di **32 dB**.

In base a considerazioni fatte in precedenza, il dimensionamento del sistema si deve basare su due specifiche di fondo: la probabilità di errore p(ε) all'uscita del decisore e la frequenza di cifra f_S. Per quanto riguarda la prima, fissiamo arbitrariamente il valore $p(\varepsilon)=10^{-6}$ (cioè 1 bit sbagliato ogni milione). Per quanto riguarda, invece, la frequenza di cifra, possiamo fare il seguente ragionamento: in primo luogo, sappiamo che il segnale televisivo occupa una banda di circa 5 MHz; se trasmettessimo in banda base, avremmo dunque bisogno di una banda passante di 5MHz; se, invece, il mezzo è passa-banda, per cui dobbiamo usare la modulazione, abbiamo bisogno di una banda maggiore. Supponiamo allora che il mezzo trasmissivo sia passa-banda e ci metta a disposizione 10 MHz di banda: dato che la modulazione di ampiezza (ASK) richiederebbe proprio 10 MHz, scegliamo questo tipo di trasmissione, considerando una portante a frequenza pari esattamente a metà della banda a nostra disposizione. Con questo ragionamento, ***il nostro sistema di trasmissione è equivalente ad un sistema di trasmissione in banda base con banda B=5 MHz.*** Fissata dunque la banda, la frequenza di cifra dipende dal valore del roll off δ secondo la relazione $B=\frac{f_S}{2}(\delta+1)$. Se volessimo la massima velocità di trasmissione, dovremmo prendere δ=0. Al contrario, scegliamo l'estremo opposto, ossia **δ=1**. In tal modo, la frequenza di cifra viene a coincidere numericamente con la banda occupata, per cui trasmettiamo a $f_S = 5\ \text{Mbit/sec}$.

In realtà, però, c'è da fare una precisazione: ***è vero che, nel momento in cui inviamo i bit in trasmissione, ne inviamo 5Mbit al secondo, ma è altrettanto vero che non trasmettiamo bit con continuità:*** delle 625 righe di ogni quadro, infatti, solo alcune sono destinate ad accogliere le informazioni del Televideo, per cui noi trasmettiamo ad "intermittenza". Andiamo allora a calcolare quanto tempo, per ogni quadro, è destinato alla trasmissione numerica: supponiamo, per esempio, di usare 2 sole righe attive per la trasmissione del Televideo; dato che in 1 secondo sono trasmessi 25 quadri, noi trasmettiamo 50 righe di Televideo per ogni secondo; dato che ognuna riga mette a disposizione 50μsec per la trasmissione delle informazioni, deduciamo che il tempo da noi impiegato per la trasmissione numerica è

$$50(\text{righe})\cdot 50*10^{-6}\left(\frac{\text{sec}}{\text{righe}}\right)=2.5(\text{m sec})$$

Quindi, in ogni secondo noi abbiamo a disposizione 2.5 msec per la trasmissione numerica: quindi, la frazione di tempo a nostra disposizione è $\chi=\frac{2.5(\text{m sec})}{1(\text{sec})}=2.5*10^{-3}$.

La conseguenza di questo ragionamento è la seguente: se avessimo la possibilità di trasmettere i bit con continuità, noi trasmetteremmo a $f_S = 5\ \text{Mbit/sec}$; dato che, invece, pur inviando 5 Mbit/sec, trasmettiamo per una frazione χ=0.0025 di tempo, allora la **velocità media di ricezione**, cioè la frequenza di cifra effettiva, è $f_{S,\text{reale}}=\chi\cdot f_S = 7.5\ \text{kbit/sec}$.

Fatta questa precisazione, possiamo comunque fare i nostri conti supponendo di poter usare il mezzo trasmissivo con continuità, ossia usando una frequenza di cifra $f_S = 5\ \text{Mbit/sec}$.

Partiamo a questo punto dalla probabilità di errore p(ε) a valle del decisore. Sappiamo che, per passare da valle a monte del decisore, ci basta considerare la relazione Q(γ)=p(ε), dove $\gamma = \frac{V_1 - V_0}{2\sigma_n}$, avendo indicato con V_1 e V_0 i livelli che si misurerebbero, in uscita dal campionatore, in assenza di rumore (caratterizzato da una deviazione standard σ_n, uguale sia all’uscita sia all’ingresso del campionatore).

Avendo preso p(ε)=10^{-6}, sappiamo che si ottiene $(\gamma)_{dB} = 20\log_{10}\gamma = 13.5 \text{ dB}$.

Abbiamo dunque valutato il rapporto S/N a valle del campionatore. Possiamo adesso passare al rapporto segnale/rumore $\left.\frac{S}{N}\right|_U$ a valle del filtro di ricezione, definito come rapporto tra la potenza di picco $P_{S,p}$ del segnale e la potenza media $P_{N,m}$ di rumore: optando ancora una volta per una codifica antipodale, abbiamo che γ^2 è numericamente uguale ad $\left.\frac{S}{N}\right|_U$, per cui

$$\left.\frac{S}{N}\right|_U = \frac{P_{S,p}}{P_{N,m}} = \gamma^2$$

Per passare all’ingresso del filtro di ricezione, dobbiamo scegliere il tipo di filtro, ossia la funzione di trasferimento. Scegliamo ancora una volta il <u>*filtro adattato*</u>: sotto questa ipotesi, sappiamo che il rapporto S/N all’uscita del filtro è numericamente pari al rapporto S/N in ingresso, a patto di definire quest’ultimo come rapporto tra la potenza media di segnale e la potenza di rumore che cade nella *banda convenzionale* $f_S/2$:

$$\left.\frac{S}{N}\right|_U = \left.\frac{S}{N}\right|_{IN} = \frac{P_R}{h_n \frac{f_S}{2}}$$

dove P_R è la potenza media (di segnale) ricevuta. Questa potenza media, dato che le forme d’onda trasmesse sono pezzi di sinusoide modulata da rettangoli di tipo antipodale, coincide con la potenza di picco, visto che i rettangoli antipodali (cioè positivi o negativi, ma con la stessa ampiezza) hanno un *fattore di picco* unitario. Di conseguenza, se indichiamo con a l’ampiezza dei rettangoli, la potenza di picco sarà a^2, per cui possiamo scrivere che

$$\left.\frac{S}{N}\right|_U = \left.\frac{S}{N}\right|_{IN} = \frac{a^2}{h_n \frac{f_S}{2}}$$

In base a quanto detto prima, questo rapporto è numericamente pari al rapporto S/N in uscita dal filtro, cioè al γ^2, per cui

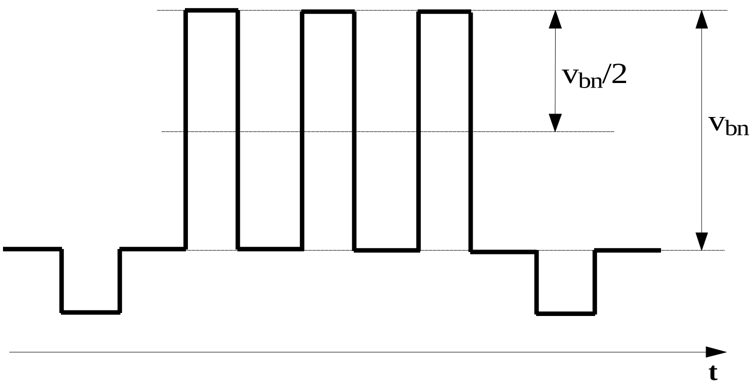
$$\left.\frac{S}{N}\right|_U = \left.\frac{S}{N}\right|_{IN} = \frac{a^2}{h_n \frac{f_S}{2}} = \gamma^2 = 13.5\text{dB}$$

In realtà, dobbiamo fare, anche in questo caso, una osservazione a proposito del <u>*fattore di forma*</u>: abbiamo prima scelto di ottenere, in uscita dal filtro, forme d’onda con δ=1; avendo fissato la funzione di trasferimento del filtro, le forme d’onda in ingresso al filtro risultano automaticamente fissate; il problema, però, è che queste forme d’onda avranno comunque delle code, per cui mal si adattano ad essere costrette in un rettangolo di durata non superiore a 50μsec. Le forme d'onda che è sensato utilizzare sono necessariamente delle forme d'onda rettangolari[19]. Con questa scelta, quindi, passando da forme d'onda (approssimativamente) rettangolari all'ingresso del filtro a forme d'onda con δ=1 all'uscita del filtro, quest'ultimo non potrà più essere adattato, per cui dobbiamo tener conto del disadattamento considerando 0.5dB di perdita nel rapporto segnale/rumore. In conclusione, quindi, all'ingresso del filtro richiediamo un rapporto segnale/rumore

$$\left.\frac{S}{N}\right|_{IN} = 13.5\text{dB} + 0.5\text{dB} = 14 \ \text{dB}$$

A questo punto, dovremmo passare a monte del demodulatore, in modo da confrontarci con i 32 dB garantiti in uscita dal mezzo trasmissivo, ma evitiamo questo passaggio, supponendo che i suddetti 32 dB siano già relativi all'uscita del demodulatore.

Con questa ipotesi, dovremmo confrontare quindi i 14 dB necessari per avere $p(\varepsilon)=10^{-6}$ con i 32 dB garantiti dal mezzo trasmissivo. Sembrerebbe, dunque, come nell'esercizio precedente, che ci sia un sovrappiù di potenza non utilizzato. In realtà, bisogna fare una importante precisazione: i 14 dB che noi abbiamo ricavato si riferiscono a trasmissione di forme d'onda di tipo antipodale, ossia rettangoli di ampiezza positiva o negativa seconda che sia trasmesso 1 o 0. Tuttavia, <u>per come è strutturato il segnale televisivo (nel quale vogliamo inserire le informazioni numeriche), non è possibile scendere al di sotto del livello del nero</u>. Di conseguenza, se vogliamo usare delle forme d'onda antipodali, esse non potranno mai avere valor medio coincidente con il livello del nero. Indicata con v_{bn} l'escursione bianco-nero del segnale televisivo[20], i rettangoli dovranno avere valor medio $v_{bn}/2$ e ampiezza (positiva o negativa) pari anch'essa a $v_{bn}/2$:



Quindi, il rapporto S/N in ingresso al filtro sarà

$$14\text{dB} = \left.\frac{S}{N}\right|_{IN} = \frac{\left(\frac{v_{bn}}{2}\right)^2}{h_n \frac{f_S}{2}} = \frac{v_{bn}^2}{4h_n \frac{f_S}{2}}$$

[19] E' ovvio che le forme d'onda non potranno mai essere perfettamente rettangolari, in quanto, se così fosse, occuperebbero una banda infinita; al contrario, è necessario restringere lo spettro alla banda a disposizione, per cui non si tratterà proprio di rettangoli, ma dei classici *rettangoli sottoposti ad un filtraggio passa-basso*.

[20] che è pari al 70% dell'escursione picco-picco del segnale

D'altra parte, abbiamo definito, all'inizio, il rapporto S/N di 32 dB come rapporto tra potenza picco-picco (intesa come escursione bianco-nero al quadrato v_{bn}^2) e potenza media di rumore nella banda B=5MHz: quindi

$$\frac{S}{N} = \frac{v_{bn}^2}{h_n B} = 32\text{dB}$$

Vediamo allora di legare questo rapporto S/N a quello all'ingresso del filtro:

$$\left.\frac{S}{N}\right|_{IN} = \frac{v_{bn}^2}{4h_n \frac{f_S}{2}} = \frac{v_{bn}^2}{2h_n f_S}$$

Dato che abbiamo scelto δ=1, la banda B occupata e la frequenza di cifra coincidono numericamente, per cui

$$\left.\frac{S}{N}\right|_{IN} = \frac{v_{bn}^2}{2h_n f_S} = \frac{v_{bn}^2}{2h_n B} = \frac{1}{2}\frac{S}{N}$$

Abbiamo dunque trovato che, in unità naturali, il rapporto S/N all'ingresso del filtro è metà del rapporto S/N che abbiamo inizialmente definito all'uscita del mezzo trasmissivo: per effettuare un confronto, dobbiamo dunque aumentare $\left.\frac{S}{N}\right|_{IN}$ di 3dB, ottenendo 17dB.

La conclusione è dunque che, per ottenere una p(ε) di 10^{-6}, ci basta ottenere 17 dB di rapporto S/N all'uscita del mezzo trasmissivo, quando invece quest'ultimo ce ne fornisce 32dB. Abbiamo perciò nuovamente un sovrappiù di potenza, che in questo caso è di 15dB. Pur considerando che i 17dB sono un limite minimo teorico, che comunque noi aumenteremo di qualche unità in modo da abbassare la probabilità di errore di qualche ordine di grandezza, ci sono dunque dB di potenza inutilizzati. Possiamo pensare di impiegare questi dB per realizzare un sistema multilivello? In questo caso non è possibile, in quanto, in mancanza di una equalizzazione accurata del canale, bisogna tener conto dei cammini multipli (**multipath**) dovuti alle stratificazioni dell'atmosfera e alle riflessioni immancabili.

ESEMPIO: TRASMISSIONE NUMERICA DEL SEGNALE TV SU PONTE RADIO

Supponiamo nuovamente di voler trasmettere il segnale televisivo a colori in formato numerico. Come mezzo di trasmissione scegliamo questa volta un **ponte radio** e, in particolare, supponiamo che la distanza tra trasmettitore e ricevitore sia L=100 km. In base ai noti problemi di visibilità tra le antenne, sappiamo che una distanza di 100 km non si può coprire con una sola tratta, mentre sicuramente ne bastano 2, lunghe ciascuna 50 km.

Le specifiche da rispettare sono come al solito 3:

- in primo luogo, abbiamo visto, in un esercizio precedente, che la frequenza di cifra necessaria alla trasmissione numerica del segnale TV a colori, secondo lo standard **CCIR 4:2:2**, è $f_S = 162\ \text{Mbit/sec}$;
- in secondo luogo, sempre in accordo allo stesso esercizio, abbiamo stabilito che la probabilità di errore accettabile è p(ε)=10^{-7} (derivante dal tollerare 1 solo bit sbagliato per ogni quadro);

- infine, supponiamo di avere a disposizione, per la trasmissione, una banda complessiva B_{RF}=40MHz, che sarà ovviamente centrata su una frequenza centrale f_0 dell'ordine dei GHz.

L'obbiettivo è di dimensionare il sistema, determinando sostanzialmente la potenza da trasmettere, per rispettare queste specifiche.

La prima considerazione da fare riguarda proprio la banda a nostra disposizione sul mezzo trasmissivo: non essendo un sistema passa-basso, abbiamo la necessità di effettuare una modulazione, che dovremo scegliere essenzialmente tra PSK, QAM e FSK (non consideriamo la ASK in quanto sappiamo che, usando forme d'onda modulanti di tipo rettangolare, essa è del tutto equivalente alla PSK). *Se la banda a disposizione non è una risorsa che possiamo scegliere a piacimento, il parametro discriminante è sicuramente la potenza, nel senso che conviene scegliere la modulazione più efficiente in termini di potenza da trasmettere:* la modulazione FSK è sicuramente la meno efficiente da questo punto di vista, in quanto ....(?)..... ; tra la PSK e la QAM, invece, ci sono pro e contro: la PSK è sicuramente più efficiente per quanto riguarda la trasmissione, nel senso che, generando forme d'onda modulate ad inviluppo costante, consente l'uso di amplificatori in saturazione, aventi efficienza estremamente alta; d'altra parte, però, se dovessimo poi scegliere un sistema multilivello, la QAM è più vantaggiosa rispetto alla PSK in quanto consente un maggiore sfruttamento della potenza ai fini della probabilità di errore: ricordiamo infatti che, mentre la PSK prevede l'uso di costellazioni di punti disposti su un cerchio, la QAM consente la disposizione dei punti ai bordi e all'interno di un quadrato.

Ad ogni modo, per il momento scegliamo la modulazione PSK. Questo significa, se usiamo forme d'onda modulanti di tipo rettangolare, che effettuiamo una modulazione di ampiezza della portante sinusoidale (e una demodulazione coerente in ricezione), per cui lo spettro del segnale modulato occupa banda doppia rispetto alla banda B del segnale modulante. Ciò significa, che la banda del segnale modulante dovrà essere metà di quella a disposizione sul mezzo trasmissivo:

$$B = \frac{B_{RF}}{2} = 20\ \text{MHz}$$

Dobbiamo quindi ragionare con *un sistema di trasmissione equivalente, in banda base, di banda pari a 20 MHz.*

In questi 20 MHz noi dobbiamo trasmettere ad una velocità di 162 Mbit/sec: usando un sistema binario, anche nell'ipotesi di scegliere roll off δ=0 in uscita dal filtro di ricezione, avremmo bisogno di una banda numericamente pari alla metà di 162 MHz, ossia 81 MHz, mentre invece abbiamo a disposizione solo 20 MHz. Non possiamo far altro che usare un sistema multilivello.

Supponiamo, allora, di prendere δ=0 (il che, nella realtà, non è fattibile): con 20 MHz di banda a disposizione, il baud rate, con un normale sistema PSK binario, è 40 Mbit/sec; se usiamo un sistema a 4-PSK, otteniamo invece 80 Mbit/sec; se usiamo allora un sistema 16-PSK, arriviamo a 160 Mbit/sec. Per trasmettere a 162 Mbit/sec, l'unica possibilità è allora la seguente: dobbiamo ricorrere ad un sistema a 32 livelli, con il quale potremmo trasmettere a 200 Mbit/sec, nel quale però non consideriamo δ=0, ma un opportuno valore di δ maggiore di 0. Abbiamo quanto segue:

$$B = \frac{f_B}{2}(\delta+1) \longrightarrow \delta = \frac{2B}{f_B} - 1 = \frac{f_{B,max}}{f_{B,desiderato}} - 1 = \frac{200\cdot 10^6}{162\cdot 10^6} - 1 = 0.23$$

Quindi, *per trasmettere a 162 Mbit/sec, in una banda base (equivalente) di 20 MHz, dobbiamo adottare d=0.23 all'uscita del filtro di ricezione.*

A questo punto possiamo fare i normali conti per il dimensionamento di un sistema numerico, partendo dalla probabilità di errore p(ε) a valle del decisore.

Per passare da valle a monte del decisore, ci basta considerare la relazione Q(γ)=p(ε), dove $\gamma = \frac{V_1 - V_0}{2\sigma_n}$, avendo indicato con V_1 e V_0 i livelli che si misurerebbero, in uscita dal campionatore, in assenza di rumore (caratterizzato da una deviazione standard σ_n, uguale sia all'uscita sia all'ingresso del campionatore).

Se fosse p(ε)=10^{-6}, sappiamo che si otterrebbe $(\gamma)_{dB} = 20\log_{10}\gamma = 13.5 \text{ dB}$. Dato che, invece, vogliamo p(ε)=10^{-7}, dobbiamo considerare $(\gamma)_{dB} = 14.5 \text{ dB}$.

Possiamo adesso passare al rapporto segnale/rumore $\left.\frac{S}{N}\right|_U$ a valle del filtro di ricezione, definito come rapporto tra la potenza di picco $P_{S,p}$ del segnale e la potenza media $P_{N,m}$ di rumore: optando ancora una volta per una codifica antipodale, abbiamo che γ^2 è numericamente uguale ad $\left.\frac{S}{N}\right|_U$, per cui

$$\left.\frac{S}{N}\right|_U = \frac{P_{S,p}}{P_{N,m}} = \gamma^2$$

Per passare all'ingresso del filtro di ricezione, dobbiamo scegliere il tipo di filtro. Se scegliamo il *filtro adattato*, sappiamo che il rapporto S/N all'uscita del filtro è numericamente pari al rapporto S/N in ingresso, a patto di definire quest'ultimo come rapporto tra la potenza media di segnale e la potenza di rumore che cade nella *banda convenzionale* $f_S/2$:

$$\left.\frac{S}{N}\right|_U = \left.\frac{S}{N}\right|_{IN} = \frac{P_R}{h_n \frac{f_S}{2}}$$

dove P_R è la potenza media (di segnale) ricevuta. Questa potenza media, dato che le forme d'onda trasmesse sono pezzi di sinusoide modulata da rettangoli di tipo antipodale, coincide con la potenza di picco. Di conseguenza, se indichiamo con a l'ampiezza dei rettangoli, la potenza di picco sarà a^2, per cui possiamo scrivere che

$$\left.\frac{S}{N}\right|_U = \left.\frac{S}{N}\right|_{IN} = \frac{a^2}{h_n \frac{f_S}{2}}$$

In base a quanto detto prima, questo rapporto è numericamente pari al rapporto S/N in uscita dal filtro, cioè al γ^2; se però consideriamo il disadattamento del filtro, adottando un fattore di forma di 0.5dB, abbiamo che

$$\left.\frac{S}{N}\right|_U = \left.\frac{S}{N}\right|_{IN} = \frac{a^2}{h_n \frac{f_S}{2}} = \gamma^2 + 0.5\text{dB} = 15\text{dB}$$

A questo punto, dobbiamo passare a monte del demodulatore coerente, valutando come cambiano la potenza di segnale e quella di rumore: ci ricordiamo, allora, da quanto visto a suo tempo nel caso della demodulazione analogica, che la potenza media di segnale raddoppia dall'ingresso all'uscita, mentre rimane invariata la potenza media di rumore: infatti, mentre raddoppia la densità spettrale di potenza del rumore (se h_n è quella in uscita, in ingresso è $h_n/2$), allo stesso tempo si dimezza la banda in cui pesare il rumore, che passa da f_S all'ingresso a $f_S/2$ in uscita. Indicato allora con $\left.\frac{S}{N}\right|_{MEZZO}$ il rapporto S/N a monte del demodulatore (e quindi all'uscita del mezzo trasmissivo), abbiamo che

$$\left.\frac{S}{N}\right|_{MEZZO} = \frac{\frac{a^2}{2}}{\frac{h_n}{2} f_S} = \frac{1}{2}\frac{a^2}{h_n \frac{f_S}{2}} = \frac{1}{2}\left.\frac{S}{N}\right|_{IN} = \frac{1}{2}\frac{a^2}{h_n \frac{f_S}{2}} \xrightarrow{\text{in unità logaritmiche}} \left.\frac{S}{N}\right|_{MEZZO} = \left.\frac{S}{N}\right|_{IN} - 3dB = 12dB$$

A questo punto, avendo stabilito quanto deve valere il rapporto S/N in uscita dal mezzo, ci basta valutare h_n per calcolare la potenza a^2 necessaria in ricezione, ossia quindi il valore di a.

Nota la potenza minima necessaria in ricezione, per passare alla potenza minima da trasmettere su ogni tratta in trasmissione dobbiamo considerare l'attenuazione introdotta dal mezzo trasmissivo. A questo proposito, sappiamo che, nel caso del ponte radio, le attenuazioni sono due: *attenuazione in spazio libero*, dovuta alla divergenza sferica delle onde, e *attenuazione equivalente complessiva*, dovuta essenzialmente ai cammini multipli.

Se supponiamo di usare, tra una tratta e l'altra, apparecchiature di pura amplificazione (che cioè si occupano solo di amplificare in potenza il segnale ricevuto e ritrasmetterlo), allora la potenza da trasmettere su ogni tratta sarà semplicemente

$$P_{T,min} = P_{R,min} + \alpha_{eq} + \alpha_{SL}$$

Facciamo osservare che quest'ultima relazione può essere interpretata nel modo seguente: mentre la quantità $P_{R,SL} = P_{R,min} + \alpha_{eq}$ rappresenta la *minima potenza in ricezione in propagazione in spazio libero* (cioè in assenza di attenuazione in spazio libero), aggiungendo l'attenuazione α_{SL} otteniamo la potenza minima in trasmissione:

$$P_{T,min} = P_{R,SL} + \alpha_{SL}$$

Dobbiamo dunque valutare α_{SL} e $\alpha_{eq} = \alpha_S + 10\log_{10} N$. Per quanto riguarda α_S, sappiamo che può essere valutata come il reciproco della probabilità di fuori servizio del sistema: prendendo, per quest'ultima, il solito valore $P_{F.S.T} = 10^{-3}$, si ottiene α_S=30dB, per cui

$$P_{T,min} = P_{R,min} + 30[dB] + 10\log_{10} 2 + \alpha_{SL}$$

Per quanto riguarda, invece, l'attenuazione di spazio libero, sappiamo che è una quantità deterministica (al contrario di α_{eq} che abbiamo appena visto essere una quantità statistica) in quanto dipende dal guadagno G_T dell'antenna trasmittente, dall'area efficace A_{eff} dell'antenna ricevente e dalla lunghezza L del collegamento. Supponiamo ad esempio che risulti α_{SL}=100dB (su ogni tratta), per cui

$$P_{T,min} = P_{R,min} + 33[dB] + 100[dB]$$

La conclusione è che dobbiamo trasmettere 133 dB in più rispetto alla minima potenza richiesta in ricezione[21].

[21] E' chiaro che questa è la condizione minima teorica di funzionamento, ma nella realtà dovremo comunque garantirci un buon margine di sicurezza: ad esempio, anziché trasmettere 133dB in più rispetto a $P_{R,min}$, ne trasmetteremo almeno 143dB in più.

Queste considerazioni valgono dunque nel caso in cui decidiamo semplicemente di amplificare il segnale tra una tratta e l'altra. Vediamo invece adesso se e come cambiano le cose se decidiamo di rigenerare il segnale tra una tratta e l'altra.

Se estraiamo la sequenza di bit tra una tratta e l'altra, ci riconduciamo, in pratica, ad un sistema formato da due sistemi numerici in cascata, ciascuno dei quali relativo ad un collegamento di lunghezza 50 km. Vediamo allora di riproporre considerazioni già fatte in precedenza.

Se ogni tratta diventa sostanzialmente un sistema di trasmissione numerico a sé stante, su ciascuna tratta si potranno verificare degli errori, ma abbiamo già osservato in precedenza che *è praticamente impossibile che su più tratte si sbagli lo stesso bit*. Questa considerazione consente di trascurare la probabilità dell'evento "bit sbagliato sia sulla prima sia sulla seconda tratta"[22], il che significa che il numero N_{tot} di bit sbagliati alla fine è la somma del numero di bit sbagliato sulla prima e poi sulla seconda tratta (gli errori sulle varie tratte si vanno sommando gli uni agli altri):

$$N_{tot} = N_1 + N_2$$

Se la probabilità di errore sul generico bit è abbastanza piccola e se gli errori di tratta in tratta sono statisticamente indipendenti tra loro[23], si può affermare che *la probabilità di errore complessiva è pari alla somma delle probabilità di errore sulle singole tratte:*

$$P_{tot}(\varepsilon) = \sum_{i=1}^{N} p_i(\varepsilon)$$

(ovviamente, nel nostro caso è N=2)

Questa formula è l'analogo della formula in base alla quale, nei sistemi analogici multitratta su cavo, il rapporto rumore/segnale complessivo è pari alla somma dei singoli rapporti segnale/rumore.

Ovviamente, la probabilità di errore complessiva $P_{tot}(\varepsilon)$ coincide con la $p(\varepsilon)$ considerata nei sistemi a singola tratta, cioè con la $p(\varepsilon)=10^{-7}$ che noi richiediamo all'uscita del sistema complessivo:

$$\sum_{i=1}^{N} p_i(\varepsilon) = 10^{-7}$$

Se stessimo considerando una trasmissione via cavo (come fatto in un esercizio precedente), ci basterebbe richiedere che la probabilità di errore $p_i(\varepsilon)$ di ogni tratta sia N volte più piccola della probabilità di errore totale $P_{tot}(\varepsilon)$. Questo perché si può assumere che tutte le tratte abbiano un uguale comportamento dal punto di vista del rumore. Al contrario, *nel caso di un ponte radio, in cui ogni tratta si comporta diversamente dalle altre a causa dei cammini multipli, non ha senso richiedere che risulti* $p_i(\varepsilon) = P_{tot}(\varepsilon)/N$.

E' molto più sensato, invece, fare un discorso di questo tipo: date le N tratte da cui è composto il sistema, è realistico pensare che, in ciascun intervallo di tempo di osservazione, ci possa essere 1 sola tratta particolarmente svantaggiata rispetto alle altre, ossia affetta da una attenuazione di gran lunga superiore rispetto alle altre. Questa assunzione di partenza consente di stabilire quanto deve valere $P_{tot}(\varepsilon)$. Infatti, supponiamo per esempio che il sistema abbia N=4 tratte, la prima delle quali abbia una attenuazione di x dB; supponiamo inoltre che le altre tratte abbiano i seguenti valori di attenuazione:

---

[22] Se uno stesso bit venisse sbagliato nella prima tratta e poi anche nella seconda, alla fine arriverebbe corretto al ricevitore. Se la probabilità che questo accada non fosse trascurabile, i discorsi da fare cambierebbero decisamente rispetto a quelli che facciamo noi.

[23] il che equivale a dire che devono essere indipendenti i disturbi che, in ciascuna tratta, agiscono sul segnale provocando gli errori

$$
\begin{array}{ll}
1° \text{ tratta: } x \text{ dB} \longrightarrow & p_1(\varepsilon) \\
2° \text{ tratta: } x+5 \text{ dB} \longrightarrow & p_2(\varepsilon) = p_1(\varepsilon)\cdot 10^5 \\
3° \text{ tratta: } x+10 \text{ dB} \longrightarrow & p_3(\varepsilon) = p_1(\varepsilon)\cdot 10^{10} \\
4° \text{ tratta: } x+2 \text{ dB} \longrightarrow & p_4(\varepsilon) = p_1(\varepsilon)\cdot 10^2
\end{array}
$$

In base a quanto abbiamo visto in precedenza a proposito del legame tra degradazione del rapporto S/N in uscita e degradazione della probabilità di errore[24], è evidente che, tra le 4 tratte, la più svantaggiata è la 3°, che sperimenta una attenuazione di 10 dB e quindi una probabilità di errore di 10 ordini di grandezza superiore a quella della tratta meno svantaggiata, che in questo caso è la prima. Non solo, ma confrontando proprio $p_1(\varepsilon)$ e $p_3(\varepsilon)$, si vede che, per ogni bit sbagliato dalla prima tratta, la terza tratta ne sbaglia 10^{10}, cioè un numero estremamente più grande. Anche rispetto alle altre due tratte, il numero di bit sbagliato dalla terza tratta è di gran lunga maggiore: ad esempio per ogni bit sbagliato dalla 2° tratta, la terza tratta ne sbaglia 10^5, ossia 1 a 100 mila. E' evidente, allora, che *la probabilità di errore complessiva viene praticamente a coincidere con la probabilità di errore della tratta più svantaggiata.*

In effetti, quindi, il discorso, fatto in precedenza, secondo il quale si considerava una tratta in affievolimento profondo e le altre quasi in attenuazione di spazio libero, nel caso delle apparecchiature rigenerative si traduce nel discorso secondo il quale *la maggior parte dei bit sbagliati alla fine del collegamento provengono da una sola tratta e precisamente da quella che si trova con una attenuazione (dovuta ai cammini multipli) maggiore di tutte le altre.*

Ovviamente, dato che la tratta più svantaggiata non è sempre la stessa, la conclusione del discorso è che bisogna dimensionare il sistema in modo tale che su ciascuna tratta sia garantita la stessa probabilità di errore, pari alla probabilità di errore $p(\varepsilon)$ complessiva:

$$p(\varepsilon) = P_{tot}(\varepsilon) = p_i(\varepsilon)$$

Quindi, nel caso di apparecchiature rigenerative, la probabilità di errore resta la stessa sia che si consideri una sola tratta sia che si considerino più tratte.

Con questo modello abbiamo dunque trasferito il concetto di *fuori servizio dell'intero sistema* al concetto di **fuori servizio della singola tratta**: se la generica tratta sperimenta una probabilità di errore inferiore a quella massima tollerabile, allora il sistema è sopra soglia, altrimenti è fuori servizio. Questo significa che la $P_{F.S.i}$ (probabilità di fuori servizio della generica tratta) rappresenta la probabilità dell'evento "probabilità d'errore della generica tratta maggiore della probabilità di errore richiesta a tutto il sistema:

$$P_{F.S.i} = P\left(p_i(\varepsilon) > p(\varepsilon)\right)$$

E' chiaro che l'evento "fuori servizio della singola tratta" si può definire per tutte le tratte. Non solo, ma è chiaro che il sistema va fuori servizio se almeno una delle tratte è fuori servizio: in termini probabilistici, questo significa l'evento "fuori servizio del sistema" è dato dall'unione degli eventi "fuori servizio dell' i-sima tratta":

$$\boxed{P_{F.S.T} = \sum_{i=1}^{N} P_{F.S.i}}$$

[24] Per ogni diminuzione di 1 dB del rapporto S/N, la probabilità di errore aumenta di 1 ordine di grandezza

dove chiaramente $P_{F.S.T}$ è la probabilità di fuori servizio dell'intero sistema, cioè il dato che ci viene fornito per eseguire il progetto.

A questo punto, si può fare una ulteriore ipotesi: *è ragionevole assumere che la statistica degli affievolimenti profondi sia la stessa per tutte le tratte*, il che significa che anche le $P_{F.S.i}$ sono tutte uguali e quindi che

$$\boxed{P_{F.S.T} = N \cdot P_{F.S.i}}$$

La conclusione cui siamo giunti è dunque duplice: *mentre la probabilità di errore su ogni singola tratta è uguale a quella di tutto il sistema, la probabilità di fuori servizio di ogni singola tratta è N volte più piccola della probabilità di fuori servizio di tutto il sistema.* Vediamo a quali calcoli ci portano queste considerazioni.

Intanto, se $p_i(\varepsilon)=p(\varepsilon)$, è evidente che i calcoli sulla potenza minima da ricevere su ogni tratta sono identici a quelli fatti nel caso della singola tratta, per cui diciamo immediatamente che in uscita da ogni "pezzo" di mezzo trasmissivo, cioè in ingresso ad ogni demodulatore, abbiamo bisogno di un rapporto S/N di **12 dB**, cui corrisponde un certo valore $P_{R,min}$ di potenza da ricevere (valore che dipende dall'entità del rumore sovrapposto).

La potenza minima in trasmissione, necessaria a ricevere $P_{R,min}$, si calcola in modo analogo a quanto fatto prima, ma con una sostanziale differenza: intanto, dobbiamo sempre ragionare su una sola tratta, per la quale dobbiamo considerare una attenuazione equivalente e una attenuazione in spazio libero; l'attenuazione equivalente, per una singola tratta, coincide con l'attenuazione supplementare α_S, che quindi è il reciproco della probabilità di fuori servizio $P_{F.S.i}$; questa però non è pari a $P_{F.S.T}$, ma è N volte più piccola: nel nostro caso, N=2, per cui $P_{F.S.i} = 0.5*10^{-3}$ e quindi $\alpha_S = 33\,\text{dB}$ (essendosi dimezzata la probabilità di fuori servizio, l'attenuazione è raddoppiata, ossia è aumentata di 3dB).

Possiamo dunque scrivere che

$$P_{T,min} = P_{R,min} + \alpha_S + \alpha_{SL,i} = P_{R,min} + 33[\text{dB}] + \alpha_{SL,i}$$

dove l'attenuazione in spazio libero $\alpha_{SL,i}$ della generica tratta è ancora di 100 dB.

In conclusione, *ai fini del calcolo della potenza da trasmettere, nulla cambia effettuando una pura amplificazione oppure una rigenerazione intermedia del segnale: numericamente si ottiene lo stesso risultato.*

Questa fondamentale conclusione dipende sostanzialmente da due ipotesi di fondo:

- in primo luogo, il sistema complessivo è costituito dalla cascata di più sottosistemi e quindi le prestazioni del sistema dipendono da uno solo dei sottosistemi;
- in secondo luogo, la statistica delle attenuazioni supplementari deve essere tale che ad una probabilità di fuori servizio pari ad 1/N corrisponde la possibilità di far fronte ad una attenuazione N volte maggiore.

Autore: **SANDRO PETRIZZELLI**
e-mail: sandry@iol.it
sito personale: http://users.iol.it/sandry
succursale: http://digilander.iol.it/sandry1