TRASMISSIONE NUMERICA IN BANDA BASE

Trasmissione numerica in banda base

Per trasmettere una sequenza di cifre binarie su un canale di trasmissione passa basso (in banda base) la tecnica più comune consiste nell'inviare una successione di forme d'onda di tipo passa basso $g(t-kT)$, repliche traslate di una forma d'onda prefissata $g(t)$, ciascuna con ampiezza positiva o negativa in accordo con il flusso dei dati da trasmettere: livello negativo per l'uno e positivo per lo zero, o viceversa. La tecnica è indicata con la sigla PAM (*Pulse Amplitude Modulation*). $f_s=1/T$ è la frequenza di simbolo (o frequenza di cifra), ed è uguale al *bit rate* nel caso di trasmissione binaria (2-PAM).

Il segnale viene trasmesso su un canale per il quale assumiamo il modello AWGN (Additive White Gaussian Noise) per cui il segnale al ricevitore sara' $\Sigma\, a_k\, g(t-kT) + n(t)$, dove $n(t)$ è il rumore che si suppone gaussiano. Si vuol ricostruire la sequenza dei dati a_k con la minima probabilità di errore. Si esamineranno nell'ordine i seguenti aspetti:

- come effettuare la decisione ottima, nel caso di trasmissione di un solo impulso
- da cosa dipende e come si calcola la probabilità di errore
- come si può trasmettere una sequenza di dati senza interferenza reciproca
- quali forme d'onda $g(t)$ si possono usare, e quale banda occupano
- come inviare più di un bit per ciascuna forma d'onda (trasmissione multilivello)
- vantaggi e svantaggi della trasmissione multilivello

Riconoscimento di una forma d'onda nota immersa in rumore (1)

Supponiamo di trasmettere un solo bit a_0. Rcevuta la forma d'onda $r(t)=a_0\, g(t) + n(t)$, dove $a_0 = \pm 1$ e $n(t)$ è rumore gaussiano bianco con densità spettrale di potenza (bilatera) $N_0/2$, si vuol decidere sul segno di a_0 con la minima probabilità di errore. A tale scopo si considera il segno della variabile (casuale)

$$y = \int_{-\infty}^{\infty} r(t)\, w(t)\, dt$$

media pesata dei valori della forma d'onda ricevuta. Il motivo della pesatura è evidente se si pensa agli istanti (o intervalli) di tempo in cui il segnale è nullo: con $w(t)$ non nullo non si preleverebbe segnale utile ma solo rumore! In tali istanti sarà quindi $w(t)=0$, e l'effettivo intervallo di integrazione sarà limitato alla durata di $g(t)$. Non è difficile intuire che $w(t)$ sarà piccolo dove il segnale utile ha piccola ampiezza, e che $w(t)$ avrà valori negativi dove $g(t)$ è minore di zero, per dare contributo positivo al segnale utile (il contributo di rumore non dipende dal segno di $w(t)$, come si vedrà tra breve).

Queste considerazioni rendono verosimile che la pesatura ottima $w(t)$ sia proporzionale a $g(t)$. La costante di proporzionalità non ha alcuna importanza: la decisione sarà infatti basata solo sul segno di y.

Riconoscimento di una forma d'onda nota immersa in rumore (2)

Possiamo separare y in due contributi

Il segnale utile alla decisione è $$s_0 = a_0 \int_{-\infty}^{\infty} g(t)\, w(t)\, dt$$

mentre il disturbo prodotto dal rumore è $$n_0 = \int_{-\infty}^{\infty} n(t)\, w(t)\, dt$$

n_0 sarà una variabile casuale, essendo $n(t)$ un processo. In particolare si può vedere come la lettura in $t=0$ del processo $n(t)$ attraverso un filtro con risposta all'impulso $w(-t)$ infatti:

$$n(t) * w(-t)\Big|_{t=0} = \int_{-\infty}^{\infty} n(\tau)\, w(\tau)\, d\tau$$

sappiamo allora che n_0 é una variabile casuale gaussiana con valor medio nullo e varianza (vedi "processi attraverso sistemi LTI"):

$$\sigma_n^2 = \frac{N_0}{2} \int_{-\infty}^{\infty} w^2(t)\, dt$$

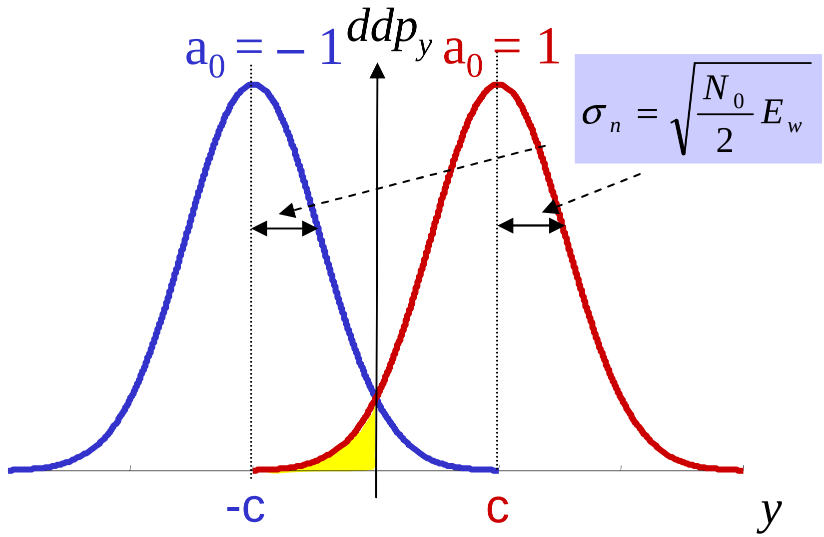
Riconoscimento di una forma d'onda nota immersa in rumore (3)

Poniamo: $c = \int_{-\infty}^{\infty} g(t) w(t)\, dt$ e $\int_{-\infty}^{\infty} w^2(t)\, dt = E_w \quad \Rightarrow \quad \sigma_n^2 = \frac{N_0}{2} E_w$

le ddp di y sotto le due ipotesi $a_0 = 1$ e $a_0 = -1$ sono rappresentate in figura.
Il decisore che minimizza la probabilità d'errore sceglie per a_0 l'ipotesi che massimizza la $P(a_0 | y)$ (MAP, massima probabilità a posteriori). Se i valori di a_0 sono equiprobabili a priori, tale scelta coincide con quella del decisore a massima verosimiglianza (ML) che sceglie per a_0 l'ipotesi che massimizza la $p(y|a_0)$ (verosimiglianza di a_0), cioè il segno di y.

Se $a_0 = 1$ il decisore commette un errore quando $y<0$, e cioè se il disturbo è minore di –*c*; se $a_0 = -1$ si ha errore quando il disturbo è maggiore di *c*. In entrambi i casi si ha la stessa probabilità d'errore (area in giallo):

$$P_e = Q\left(\sqrt{\frac{c^2}{\sigma_n^2}}\right)$$



quindi $w(t)$ va scelto in modo da rendere massimo il rapporto tra la varianza del segnale utile c^2 e la varianza del disturbo.

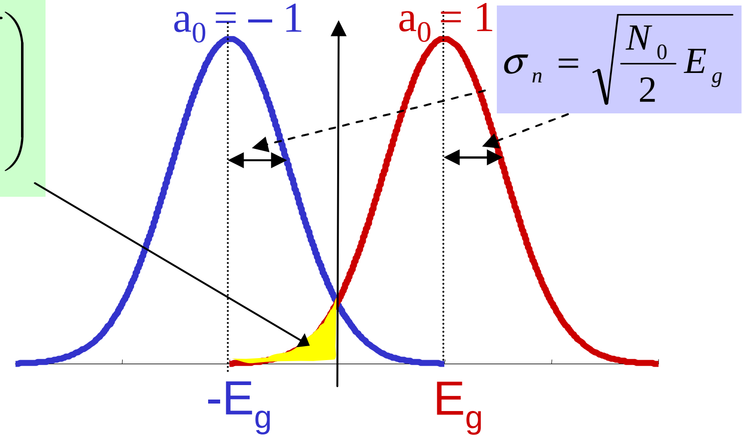
Riconoscimento di una forma d'onda nota immersa in rumore (4)

Possiamo scrivere tale rapporto, come:

$$\frac{c^2}{\sigma_n^2} = \frac{\left| \int_{-\infty}^{\infty} w(t)\, g(t)\, dt \right|^2}{\frac{N_0}{2} \int_{-\infty}^{\infty} w^2(t)\, dt} \le \frac{\int_{-\infty}^{\infty} |w(t)|^2\, dt \int_{-\infty}^{\infty} |g(t)|^2\, dt}{\frac{N_0}{2} \int_{-\infty}^{\infty} w^2(t)\, dt} = \frac{2E_g}{N_0}$$

Il limite massimo si ottiene poi facilmente: basta scegliere $w(t)$ proporzionale (o più semplicemente uguale) a $g(t)$! Così l'ampiezza del segnale utile, la varianza del disturbo e la probabilità d'errore diventano rispettivamente:

$$s_0 = a_0\, E_g \qquad \sigma_n^2 = \frac{N_0}{2} E_g \qquad P_e = Q\left(\sqrt{\frac{2E_g}{N_0}}\right)$$



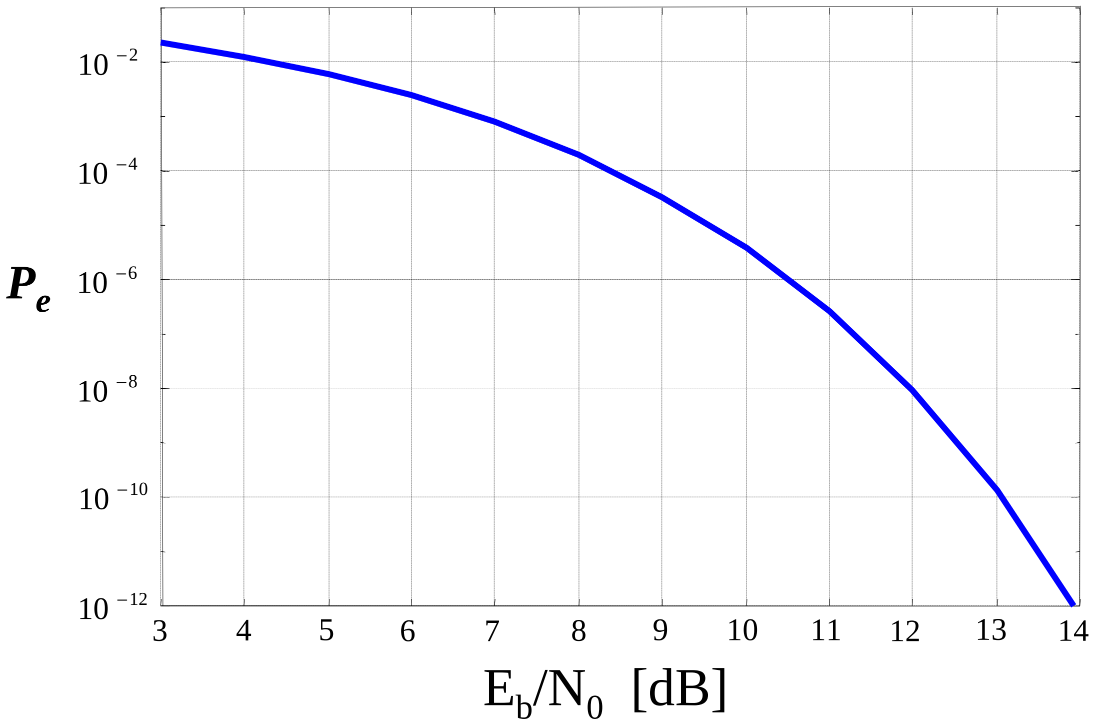
N.B. E' importante osservare che tali risultati non dipendono dall'effettiva forma d'onda elementare $g(t)$, ma solo dalla sua energia. La probabilità d'errore dipende solo dall'energia di $g(t)$ e dalla densità spettrale del rumore $N_0/2$.

Probabilità di errore e rapporto segnale-rumore

La probabilità di errore viene solitamente data in funzione di E_b/N_0 dove E_b è l'energia spesa per ciascun bit di informazione: nel caso in esame (2-PAM) $E_b = E_g$. La probabilità di errore è spesso indicata con la sigla BER (*Bit Error Rate*).

E_b/N_0 viene espresso in unità logaritmiche (dB). Si può notare (vedi figura) che la probabilità di errore decresce rapidamente all'aumentare dell'energia del segnale.

E_b/N_0 caratterizza la qualità del collegamento. Più spesso si parla di Rapporto-Potenza di Segnale-Potenza di Rumore (SNR). Tale quantità se misurata a valle del filtro di ricezione si definisce come:



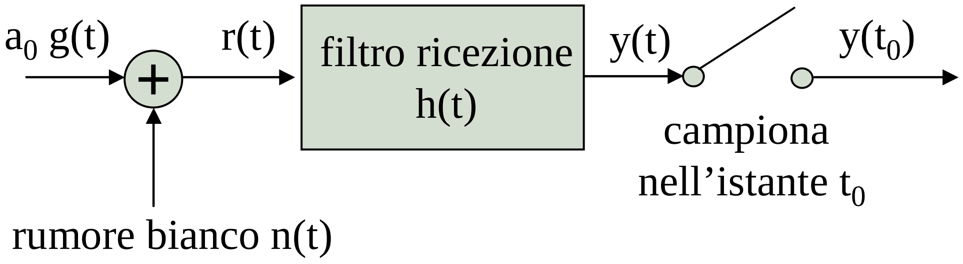
$$SNR = \frac{E\left[s_0^2\right]}{E\left[n_0^2\right]}$$

E nel nostro caso, come abbiamo visto, risulta:

$$SNR = \frac{E_g^2}{E_g N_0 / 2} = \frac{2E_b}{N_0}$$

Filtro adattato (1)

Soprattutto nel passato, la correlazione tra il segnale ricevuto *r(t)* e la forma d'onda elementare *g(t)* veniva calcolata mediante il "filtro adattato".



L'uscita del filtro all'istante t_0 è data da $y(t_0) = \int_{-\infty}^{\infty} r(\tau)\, h(t_0 - \tau)\, d\tau$ e coincide con l'uscita desiderata $\int_{-\infty}^{\infty} r(\tau)\, g(\tau)\, d\tau$ se $h(t_0 - \tau) = g(\tau)$ cioè se $h(t) = g(t_0 - t)$.

La risposta all'impulso del filtro adattato è quindi data dalla forma d'onda elementare *g(t)* ribaltata rispetto all'asse dei tempi e ritardata del tempo t_0; il ritardo t_0 deve essere sufficiente a rendere causale, e quindi realizzabile, il filtro.

Trasmissione di una sequenza di forme d'onda

Si trasmette la successione di forme d'onda $\Sigma\, a_k\, g(t-kT)$ e si riceve la forma d'onda $r(t)=\Sigma\, a_k\, g(t-kT) + n(t)$. In ricezione si calcolano, come nel caso di un solo simbolo, le correlazioni y_k di $r(t)$ con $g(t-kT)$. Si consideri ad esempio la correlazione corrispondente a $k=0$ (per determinare il bit a_0). Si ottiene

$$y_0 = \int_{-\infty}^{\infty} r(t)g(t)dt = \int_{-\infty}^{\infty} \sum_{-\infty}^{\infty} a_k g(t-kT)g(t)dt + \int_{-\infty}^{\infty} n(t)g(t)dt =$$

$$= a_0 E_g + \sum_{k\neq 0} a_k \int_{-\infty}^{\infty} g(t-kT)g(t)dt + \int_{-\infty}^{\infty} n(t)g(t)dt$$

cioè, oltre al segnale desiderato e all'inevitabile rumore, la possibile interferenza di tutti gli altri simboli a_k ($k \neq 0$). Per eliminare questa interferenza intersimbolica (ISI) basta progettare la forma d'onda elementare $g(t)$ in modo che risulti

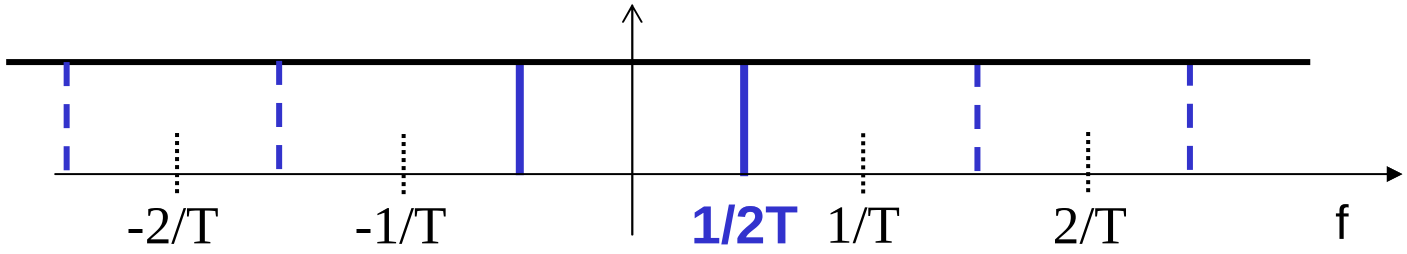
$$\int_{-\infty}^{\infty} g(t-kT)g(t)dt = 0$$

per ogni $k \neq 0$. Si può ovviamente ottenere questo risultato con forme d'onda $g(t)$ di durata T (e quindi con repliche separate temporalmente). Tuttavia ridurre la durata di $g(t)$ significa aumentarne la banda. Le forme d'onda elementari che si usano in pratica si sovrappongono nel tempo (hanno durata molto maggiore di T) e occupano una banda molto minore. Questa soluzione è stata suggerita da *Nyquist*.

Trasmissione senza interferenza intersimbolica (1)

La condizione $\int_{-\infty}^{\infty} g(t-kT)\, g(t)\, dt = 0$ per ogni $k \neq 0$ equivale a $f(kT)=0$ per $k \neq 0$, dove $f(t)=g(t)*g(-t)$. Si devono quindi trovare forme d'onda *f(t)* che abbiano zeri equispaziati (forme d'onda di *Nyquist*).

Le forme d'onda di *Nyquist*, campionate con passo *T*, hanno un solo campione diverso da zero. Se ne deduce che la ripetizione periodica, con passo in frequenza $f_c=1/T$, della trasformata *F(f)* è una costante. Se ne deduce anche che la forma d'onda di *Nyquist* con banda minima ha trasformata rettangolare, con banda *1/2T*.

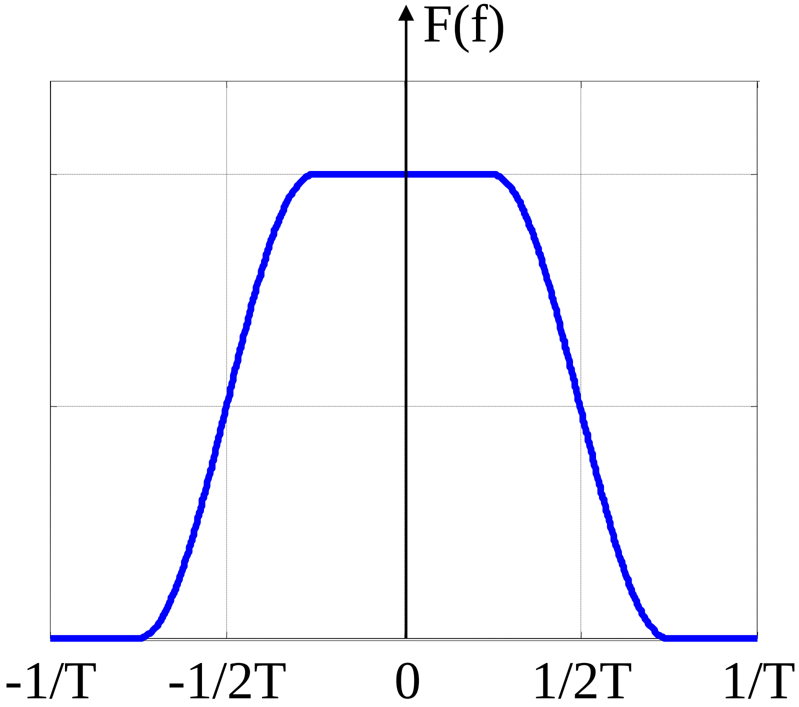


f(t) e *g(t)* sono seni cardinali (di durata infinita!): $f(t) = g(t) = \dfrac{\sin \pi t/T}{\pi t/T}$

La durata molto grande e le code che si esauriscono molto lentamente rendono questa forma d'onda di interesse solo teorico. In pratica è inevitabile usare una banda un po' maggiore del minimo teorico.

Trasmissione senza interferenza intersimbolica (2)

In figura è mostrata una soluzione con banda maggiore della minima teorica. La ripetizione periodica, con passo $1/T$, di *F(f)* dà una costante. Si noti che il modulo di *G(f)* è dato dalla radice quadrata di *F(f)*. La fase di *G(f)* è arbitraria (infatti *F(f)=G(f)G*(f)* risulta comunque reale), quindi solitamente si sceglie la soluzione con fase nulla.

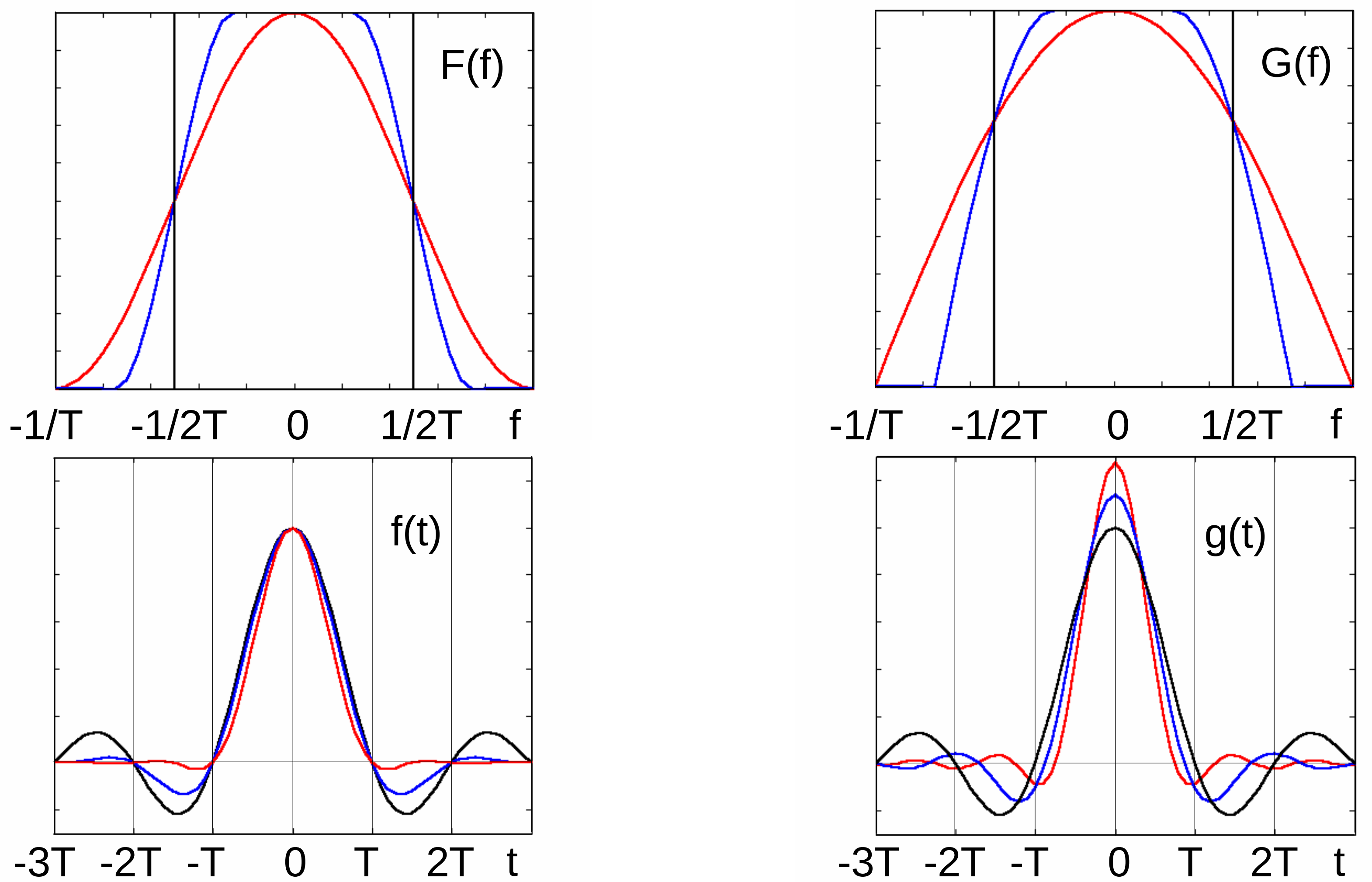


L'incremento di banda occupata relativo a quella minima (1/2T) si chiama *roll-off* e si indica con α. Tale valore é compreso tra 0 e 1 per cui la banda risulta compresa tra 1/2T e 1/T (un valore tipico di roll-off é $\alpha=0.4$). Inoltre, in questo caso (2-PAM) tempo di simbolo (T) e tempo di bit (T_b) coincidono, e sono pari all'inverso del bit rate (R_b):

$$B = \frac{1}{2T}(1+\alpha) = \frac{R_b}{2}(1+\alpha)$$

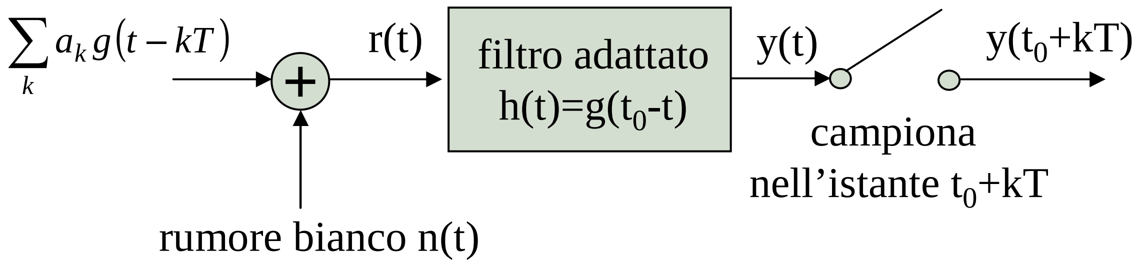
Trasmissione senza interferenza intersimbolica (3)

Esempi di forme d'onda di Nyquist e a “radice di Nyquist” (N.B.: radice nelle frequenze, non nel tempo!). Si noti che in generale la forma d'onda trasmessa *g(t)* non ha valore nullo in *t=kT* per $k \neq 0$ (non ne ha nessun motivo!)



Filtro adattato (2)

Abbiamo detto che quando si trasmette una sequenza di dati binari a_k, il ricevitore deve ripetere l'operazione di correlazione del segnale ricevuto per ciascun tempo di simbolo per poter prendere un decisione su ciascun simbolo a_k. E' facile verificare che se si campiona l'uscita del filtro adattato agli istanti $t_0 + kT$ si ottengono le correlazioni tra il segnale ricevuto $r(t)$ e le repliche traslate $g(t\text{-}kT)$ della forma d'onda elementare:



$$y(t_0+kT)=\int_{-\infty}^{\infty} r(\tau)\,g(t_0-t+\tau)\,d\tau\Big|_{t=t_0+kT} = \int_{-\infty}^{\infty} r(\tau)\,g(\tau-kT)\,d\tau$$

Con <u>un solo filtro</u> si ottengono quindi <u>tutte le correlazioni</u> che servono quando si trasmette una <u>sequenza</u> di cifre binarie.

Trasmissione multilivello: 4-PAM (1)

Supponiamo che al ritmo di bit R_b il sistema di trasmissione appena visto richieda una banda minima superiore a quella disponibile. Che fare?

Si possono prevedere per l'ampiezza a_k più valori, per esempio quattro valori (-3,-1,1,3) (4-PAM) invece di due (-1,1); è possibile distinguere tra quattro alternative: ciò corrisponde a trasmettere due bit d'informazione per ciascun simbolo trasmesso. Ai quattro livelli possono infatti essere associate coppie di bit: 00, 01, 10 e 11. Con tale modifica, i segnali trasmesso e ricevuto rimangono:

$$s(t) = \sum_{k=0}^{(N-1)/2} a_k\, g(t-kT), \qquad r(t) = s(t) + n(t)$$

Il legame tra banda minima e ritmo di simbolo rimane invariato, ma cambia il rapporto tra ritmo di simbolo ($1/T$) e ritmo di bit (R_b), visto che per ogni simbolo trasmesso, si inviano due bit:

$$R_b = \frac{2}{T}$$

$$\boxed{B =} \frac{1}{2T}(1+\alpha) = \boxed{\frac{R_b}{4}(1+\alpha)}$$

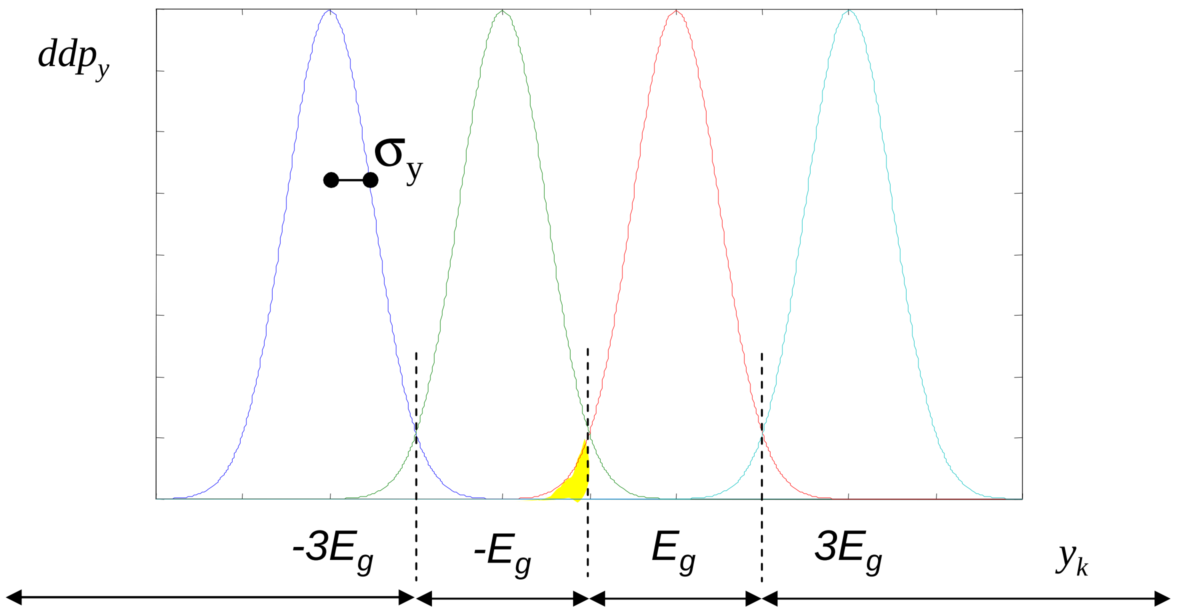
Pertanto la banda occupata si dimezza rispetto al caso 2-PAM, a pari R_b.

Trasmissione multilivello: 4-PAM (2)

In ricezione si dovrà confrontare la correlazione *y* di *r(t)* e *g(t-kT)* con opportune soglie per distinguere tra le quattro possibili ampiezze e scegliere l'ipotesi di a_k per cui la ddp del ricevuto y_k é massima.

Nota: nel caso multilivello occorre fare attenzione al guadagno del ricevitore, per posizionare esattamente le soglie. In figura si é mantenuta la scelta $w(t)=g(t)$ per cui in assenza di rumore i valori possibili di y_k sono $-3E_g$ $-E_g$,E_g ,$3E_g$.

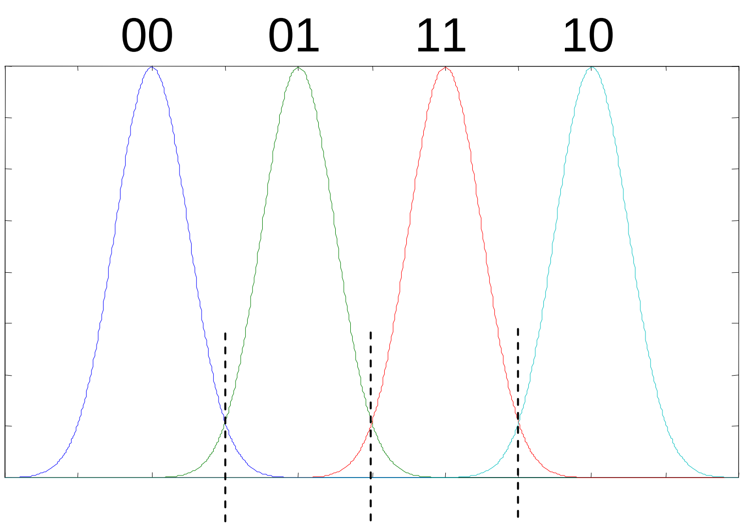
Per esempio



Trasmissione multilivello 4-PAM (3)

A pari energia E_g la probabilità di errore è praticamente uguale a quella del caso binario: infatti rimangono uguali la varianza di n_0 e la distanza tra il livello trasmesso e la soglia (E_g), per cui la probabilità rappresentata dall'area gialla vale $Q\left(\sqrt{2E_g/N_0}\right)$. Per i due livelli interni la probabilità d'errore é doppia, perché si può sbagliare in entrambe le direzioni. Inoltre si deve osservare che:

- l'energia media trasmessa, supponendo equiprobabili i quattro livelli, è data da $(9E_g+E_g+E_g+9E_g)/4=5E_g$
- l'energia spesa per ciascun bit d'informazione è la metà dell'energia media $5E_g$; infatti con un simbolo si trasmettono due bit d'informazione.



Oss: poiché si sbaglia pressoché sempre a favore di un livello adiacente, si associano le coppie di bit ai quattro livelli secondo la codifica di *Gray*: due livelli adiacenti differiscono per un solo bit. In tal modo in caso di errore, si sbaglia un solo bit della coppia:

$$P_b = \frac{1}{2} P_e$$

Trasmissione multilivello 4-PAM (4)

Riassumendo, la trasmissione a quattro livelli (4-PAM) richiede un ricevitore come quello del caso binario, ma con un decisore a tre soglie. La probabilità d'errore dipende dal livello che trasmettiamo (doppia per i due livelli interni), per cui quella media sarà:

$$P_e = (3/2)Q\left(\sqrt{2E_g/N_0}\right)$$

L'energia media per simbolo è $E_s=5E_g$ ed ogni simbolo porta due bit ($E_s=2E_b$). Con mapping di Gray:

$$\boxed{P_b = \frac{3}{4}Q\left(\sqrt{\frac{4E_b}{5N_0}}\right)}$$

Per ottenere le stesse prestazioni del 2-PAM occorre una maggiore energia: bisogna aumentare il rapporto E_b/N_0 di un fattore 5/2 = 4 dB (trascurando il fattore 3/4).

Si ha però una riduzione della banda necessaria a pari bit rate:

$$\boxed{B} = \frac{1}{2T}(1+\alpha) = \boxed{\frac{R_b}{4}(1+\alpha)}$$

Trasmissione multilivello M-PAM (1)

Il passaggio al 4-PAM consente un risparmio di banda al costo di un aumento di energia per bit. Se necessario, il numero dei livelli si può aumentare ulteriormente: p.e. con 8-PAM, cioé 8 livelli (-7,-5,-3,-1,1,3,5,7) si trasmettono tre bit per simbolo e la banda viene ridotta a un terzo rispetto al caso binario. Ad esempio per trasmettere 5 Mb/s con due livelli in pratica occorrono almeno 3 MHz (banda minima $R_b/2$=2.5 MHz). Con 4 livelli occorrono 1.5 MHz e con 8 livelli basta 1 MHz.

L'energia media però diventa $E_s=(49E_g+25E_g+9E_g+E_g+E_g+9E_g+25E_g+49E_g)/8=21E_g$ e questa corrisponde a $3E_b$. Quindi per ottenere la stessa P_e del 2-PAM occorre aumentare il rapporto E_b/N_0 di un fattore 21/3 = 8.5 dB.

Generalizzando possiamo dire che un sistema M-PAM occupa una banda:

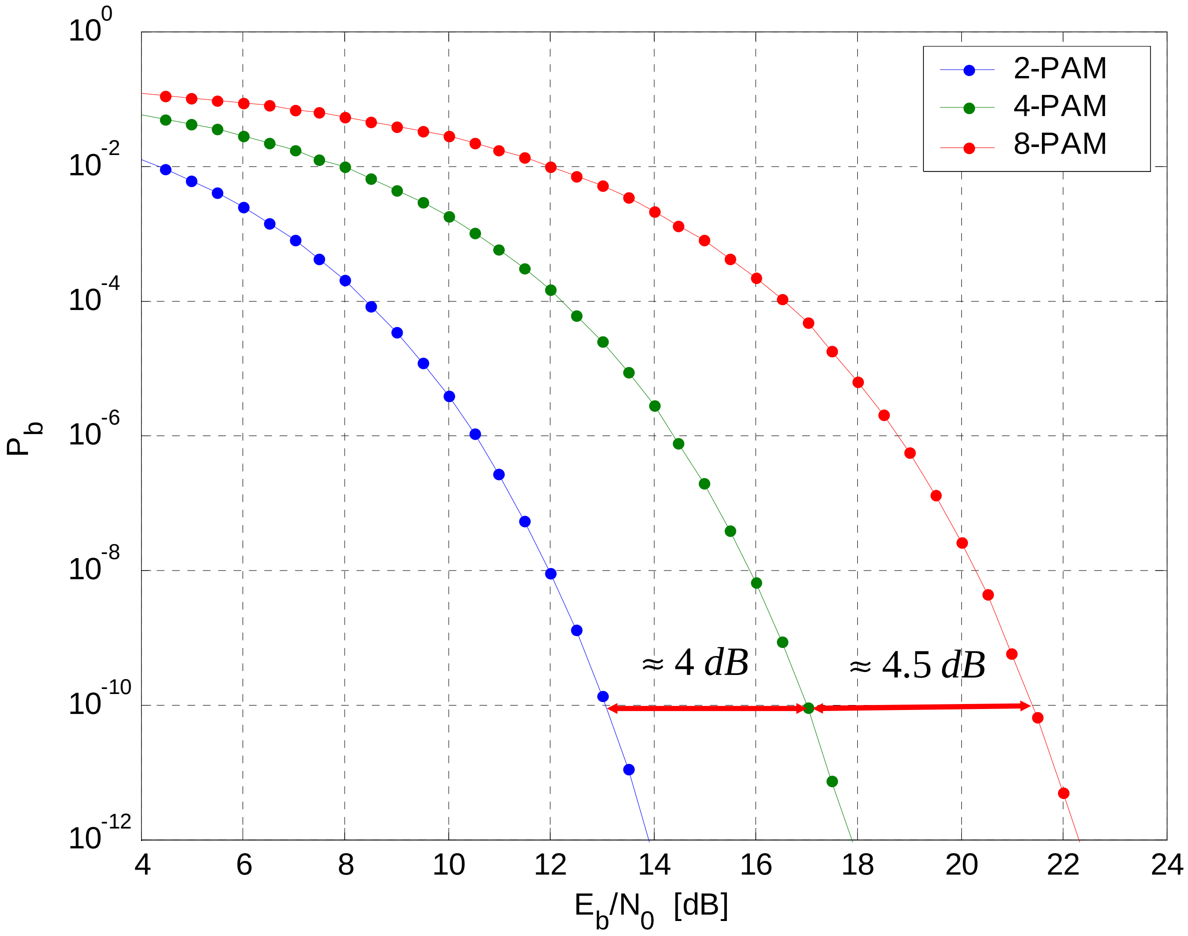
$$\boxed{B = \frac{R_b}{2}\frac{1}{\log_2 M}(1+\alpha)}$$

E per quanto riguarda le prestazioni

$$\left.\begin{aligned} E_s &= \frac{M^2-1}{3}E_g \\ E_s &= E_b \log_2 M \end{aligned}\right\} \Rightarrow \quad P_b = 2\frac{M-1}{M\log_2 M}Q\left(\sqrt{\frac{3\log_2 M}{M^2-1}\frac{2E_b}{N_0}}\right)$$

In pratica però molto raramente si superano 16 o 32 livelli.

Trasmissione multilivello M-PAM (2)



Esempio:

Si vuole trasmettere 1 Mb/s su un canale passa-basso AWGN con banda $B_c = 235\ kHz$.

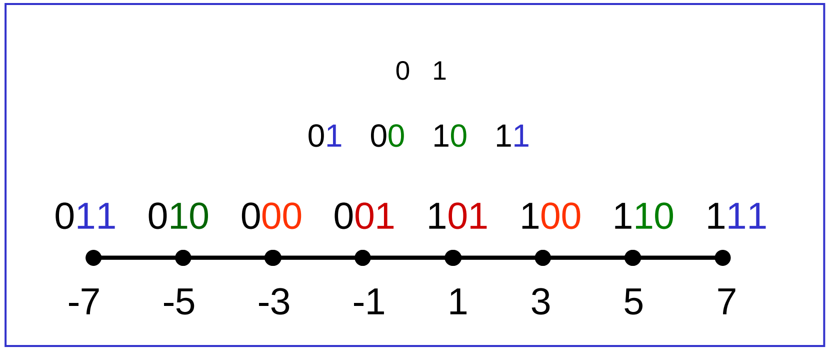
2-PAM $\qquad B_{\min} = \dfrac{10^6}{2} \dfrac{1}{\log_2 2} = 500\ kHz > B_c \qquad$ Banda disponibile non sufficiente!

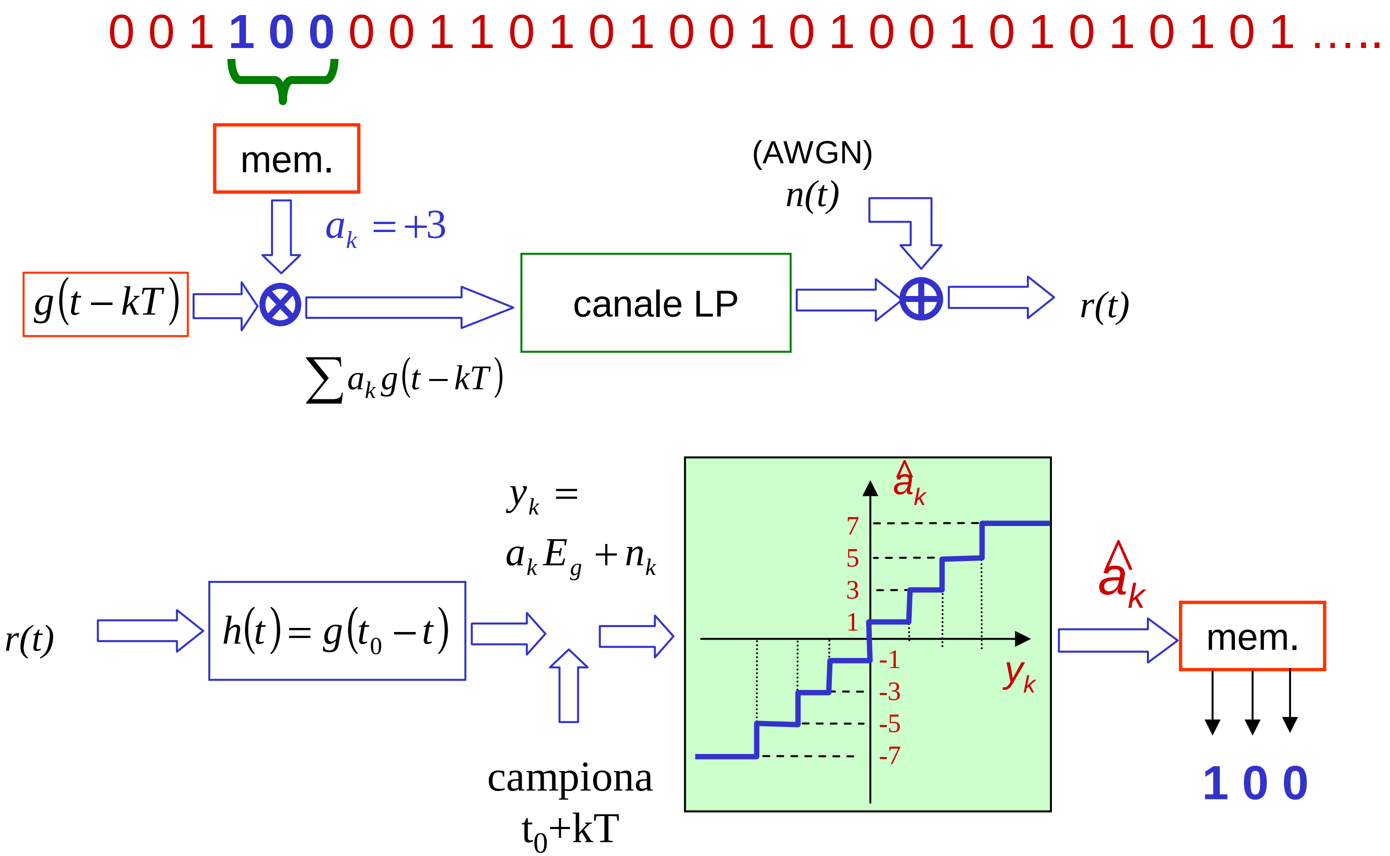
4-PAM $\qquad B_{\min} = \dfrac{10^6}{2} \dfrac{1}{\log_2 4} = 250\ kHz > B_c \qquad$ Banda disponibile non sufficiente!

8-PAM $\qquad B_{\min} = \dfrac{10^6}{2} \dfrac{1}{\log_2 8} = 167\ kHz < B_c \qquad$ Ok! Anche con roll-off 40%

Il ritmo di simbolo sarà: $\dfrac{1}{T} = \dfrac{R_b}{3}, \quad T = 3\ \mu s$

La costellazione 8-PAM con mapping di Gray:





Che E_b/N_o ci vuole per avere $P_b=10^{-8}$?

Con un 2-PAM sarebbero 12 dB (vedi curva prestazioni precedente), con un 8-PAM occorrono 12+8.5=20.5 dB.