

3

Formati e Conversioni

In un mondo ideale ci sarebbe un solo formato di audio digitale (ci sarebbero anche tante altre cose ma non è questa la sede per parlarne). Purtroppo, come sempre quando la tecnologia si applica al mercato, le scelte tecniche vengono influenzate dai fattori più disparati e i risultati sono spesso sconfortanti.

Teoricamente è possibile digitalizzare un segnale analogico con qualunque combinazione di frequenza di campionamento F_c (che si misura in **Hz**) e di profondità di quantizzazione p (che si misura in **bit**). Il parametro F_c determina la banda passante utile (che è un po' meno di $F_c/2$) e p determina la dinamica (tradizionalmente stimata in **6dB** per ogni bit utilizzato).

Vediamo ora una lista dei principali formati lineari multibit.

44100Hz/16bit. È il formato del CD Audio, dà una banda passante di circa **20KHz** con circa **96dB** di dinamica. Forse non tutti sanno perché è stato scelto proprio **44100**. Le prime apparecchiature di registrazione digitale erano basate su registratori video; per ovvi motivi di semplicità di progetto era bene avere un numero intero di campioni per ogni linea di quadro.

Il formato **NTSC** usa **30 frame** di **525** linee al secondo (di cui **490** utilizzabili); con **3** campioni per linea si ottiene una F_c di **$30 \times 490 \times 3 = 44100$** .

D'altra parte il formato **PAL/SECAM** usa **25 frame** di **625** linee al secondo (di cui **588** utilizzabili); con **3** campioni per linea si ottiene una F_c di **$25 \times 588 \times 3 = 44100$** , semplice no?

Purtroppo **$44100/2 = 22050$** e la frequenza di Nyquist è troppo vicina al limite della banda audio per premettere un'agevole conversione **D/A** e **A/D**, con tutti i problemi che ne derivano.

48000Hz/16bit. Quando è stato progettato un sistema di registrazione dedicato all'audio (il **DAT**) la frequenza di campionamento massima è stata elevata a **48KHz**. In questo caso la frequenza di Nyquist sale a **24KHz** rendendo molto meno problematici i filtri pre- e post-digitalizzazione. Storicamente fu una sorpresa per i recensori scoprire la drammatica differenza di qualità tra i due formati.

Con il miglioramento della tecnologia elettronica sono divenuti disponibili convertitori **A/D** e **D/A** con **18, 20** e **24** bit. Questo ha fatto ha creato un certo numero di formati intermedi, utilizzati specialmente nelle apparecchiature da studio che sfruttavano la maggiore risoluzione disponibile per ampliare la gamma dinamica almeno prima del riversamento su **CD**. Contemporaneamente è sorto il problema di spremere il meglio da quanto disponibile in studio al momento della creazione del master da stampare.

Dopo molti anni è stato infine possibile definire il formato fisico del successore del **CD**. Il **DVD** (*Digital Versatile Disc*) è un contenitore grezzo che nelle sue varie versioni può contenere da **5** a **18 GigaByte** di dati. Il fatto che sia "**versatile**" non è stato proprio un bene visto che sono oggi disponibili troppi formati **DVD** alternativi e la maggior parte dei consumatori se ne sta alla finestra aspettando che si decidano.

96000Hz/24bit tra le varie possibilità di audio su **DVD** questa è per la più diffusa, questo è anche il formato più ricco disponibile nelle apparecchiature commerciali a basso costo (perfino nelle moderne schede audio per PC). la frequenza di Nyquist è **48KHz** e questo risolve pressoché tutti i problemi legati al filtraggio. I soliti spiritosi hanno pensato di interpretare i vantaggi di questa banda passante allargata e sono cominciate a spuntare casse dotate di super-tweeter ad ultrasuoni (poco diffuse nel mercato occidentale). I **24** bit di profondità di quantizzazione forniscono una gamma dinamica teorica superiore al rumore termico delle normali apparecchiature. Alcuni recensori hanno criticato pesantemente questa scelta giudicandola uno spreco. Bisogna notare che a fronte di un leggero incremento della occupazione in bit si riesce ad eliminare praticamente tutti i problemi legati alla quantizzazione: il fatto che **144dB** non siano raggiungibili significa solamente che i problemi veri saranno legati alla qualità delle elettroniche è in particolare dei convertitori, lasciando molto spazio al miglioramento.

88200Hz/24bit è un formato esotico, usato da pochi puristi intelligenti per le registrazioni destinate al riversamento su **CD**. Peccato, perché sarebbe stato probabilmente lo standard ideale per il formato audio definitivo.

192000Hz/24bit Un'altra possibilità di audio su **DVD**, decisamente uno spreco, forse una scusa per mettere poco materiale sul **DVD** mantenendo alto il prezzo.

La presenza di più di un formato implica la necessità di passare dall'uno all'altro. Il resto della lezione è dedicato al problema della conversione di frequenza, la conversione di profondità di quantizzazione è altrettanto importante e interessante e merita un puntata tutta sua.

Conversioni di un fattore intero

Il caso più semplice se presenta quando si deve passare da una frequenza f_0 ad un suo multiplo o sottomultiplo. Esempi tipici sono

$$48000 \leftrightarrow 96000 \leftrightarrow 192000$$

oppure

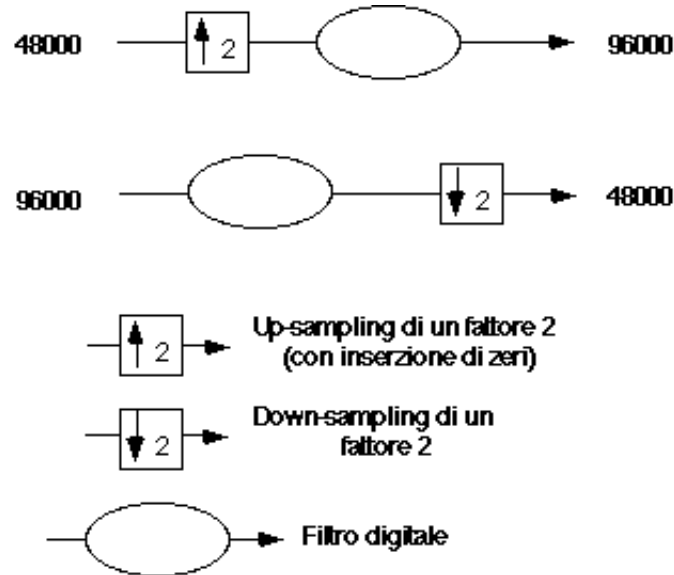
$$44110 \leftrightarrow 88200 \leftrightarrow 176400.$$

Studiamo separatamente il caso dell'*up-sampling* e del *down-sampling*.

Up-sampling intero

La moltiplicazione della frequenza di campionamento di un fattore **2**, **4** o anche **8** è un passo tipico nella conversione **D/A** multibit di qualità come abbiamo già visto. Il primo passo per moltiplicare la frequenza per **n** consiste nell'inserire **n-1** zeri tra un campione e l'altro. Si ottiene un segnale campionato alla frequenza più alta che conserva esattamente la stessa informazione (e lo stesso spettro) del segnale di partenza. Bisogna quindi eliminare le immagini della banda "buona" e questo si ottiene con un filtraggio digitale passa-basso. Nella pratica non si inseriscono davvero gli zeri ma si costruisce in modo opportuno il filtro digitale. Ricordo che altre tecniche come l'interpolazione polinomiale o l'interpolazione con *spline* sono solo un nome esotico per qualcosa che in questo contesto è comunque sempre solo e soltanto un filtraggio digitale.

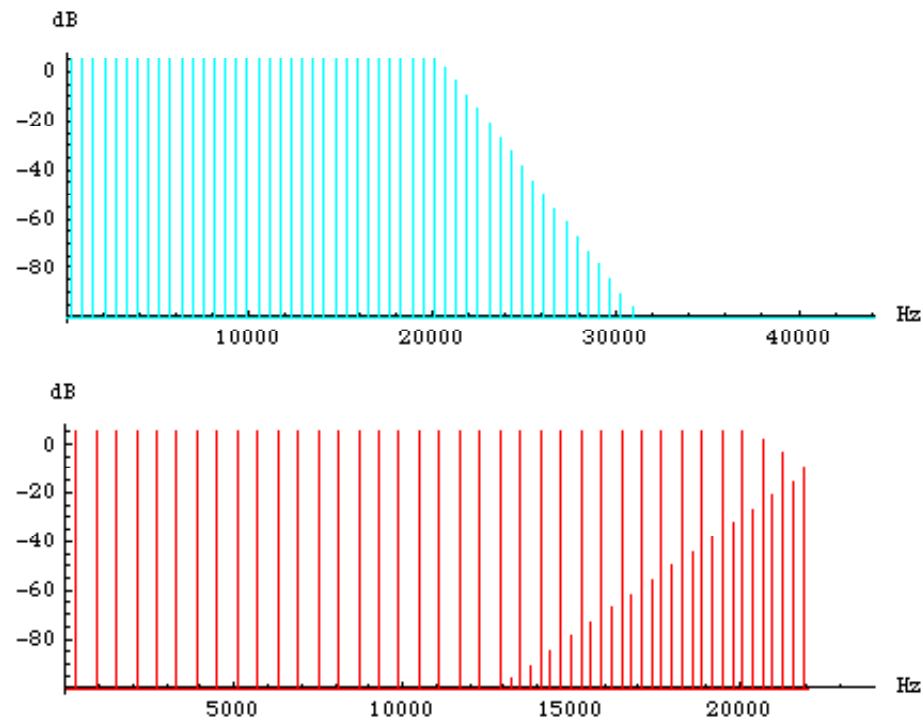
La qualità della conversione è legata alla bontà del filtro e non vi sono ostacoli ad ottenere risultati allo stato dell'arte come i migliori lettori CD possono testimoniare.



Conversione con *up-sampling* 1:2 e *down-sampling* 2:1 .

Down-sampling *intero*

La divisione della frequenza di campionamento di un fattore n è un'operazione altrettanto semplice. Se si eliminassero semplicemente $n-1$ campioni consecutivi ogni n si otterrebbe un segnale campionato alla frequenza più bassa che ha come spettro quello originale tagliato e ribattuto. Occorre quindi fare precedere il sottocampionamento da un filtraggio passa basso, anche in questo caso tenendo conto dei campioni che verranno buttati dopo il filtraggio il progetto si semplifica drasticamente.



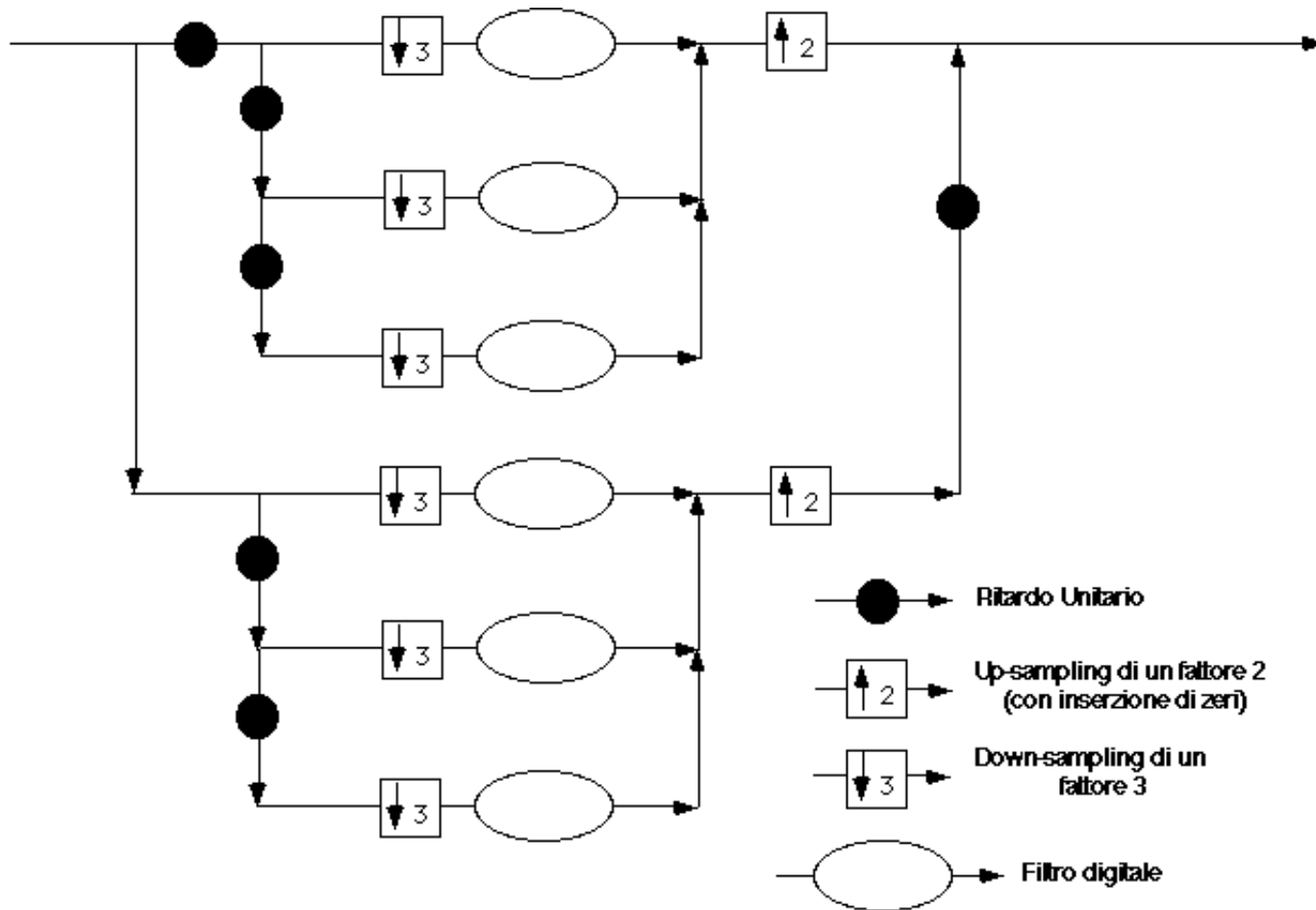
Lo spettro di un segnale con componenti oltre i 30Khz, campionato a 88200 (in celeste), lo spettro risultante dopo un sottocampionamento 88200 -> 44100 senza filtraggio preventivo. segnale (in rosso).

L'uso tipico di questa tecnica si ha quando si vuole preparare un master a **44100Hz** partendo da un originale a **88200Hz**. Si noti che se si deve anche operare una riduzione della profondità di quantizzazione questa deve essere effettuata successivamente: infatti il *Noise Shaping* deve essere **l'ultima** operazione prima della stampa del CD.

Conversioni di un fattore razionale

Se le frequenze da convertire stanno in un rapporto fisso non intero (per esempio **48KHz** e **32KHz** stanno come **3** a **2**) si può effettuare la conversione effettuando prima un *up-sampling* (da **48KHz** a **96KHz**) e poi un *down-sampling* (da **96KHz** a **32KHz**). Fare il contrario sarebbe drammatico oltre che idiota: alla fine avremmo solo **8KHz** di banda passante invece dei **16KHz** possibili. Poiché il sovracampionamento termina con un filtraggio passa basso e il sottocampionamento inizia con un filtraggio passa basso uno dei due filtri può essere eliminato utilizzando un filtro che operi a **96KHz** con frequenza di taglio sotto i **16KHz**.

Questo procedimento non presenta particolari inconvenienti fino a che il rapporto è formato da numeri piccoli, ma è possibile semplificarne l'implementazione con una tecnica alternativa (detta *polyphase filtering*). Lavorando di carta e penna nel dominio della trasformata **Z** (qualcosa di ancora più brutto e cattivo della trasformata di Fourier) si riesce a riscrivere il processo di cui sopra in un altro equivalente in cui invece di un filtraggio alla frequenza più elevata si devono fare molti filtri alla frequenza più bassa, ognuno calcolato con i coefficienti opportuni. Nel caso della conversione da **48000** e **32000** abbiamo **6** filtri che operano direttamente a **16KHz** (e quindi molto più semplici da realizzare). Gli input dei filtri sono i campioni originali presi uno ogni tre, opportunamente sfalsati (ogni sottosequenza di campioni entra in due filtri). I risultati dei filtri vengono opportunamente ricombinati. Se qualcuno vuole approfondire i dettagli e seguire la dimostrazione può consultare il capitolo sulle conversioni nel testo dello Zolzer.

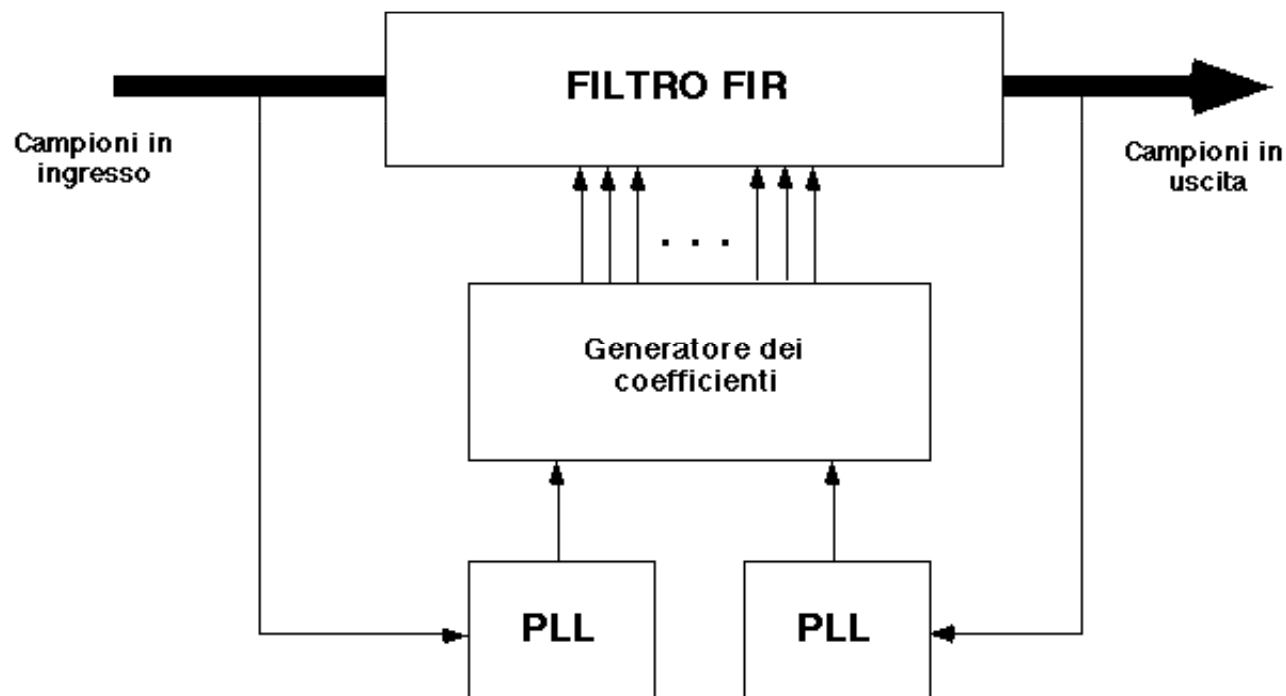


Conversione con rapporto $3/2$ con filtri *polyphase* (figura adattata dal testo di Zolzer p.225). Si noti che le operazioni iniziali di ritardo e sottocampionamento si risolvono in un instradamento dei campioni; l'operazione finale equivale a prendere alternativamente un campione da una linea e uno dall'altra.

Conversioni asincrone

Il caso più sfavorevole si presenta quando le due frequenze non stanno un rapporto semplice oppure il rapporto di frequenza deve poter variare durante la conversione. Il modello da cui si parte per questa conversione è quello di un passaggio dal dominio digitale a quello analogico seguito da un nuovo campionamento. Questo “disastro” viene compiuto davvero tutte le volte che prendendo il segnale dall’uscita analogica di un apparecchio digitale si registra di nuovo in digitale. N.B.: questo è anche quello che fanno i pirati per fare le copie da vendere o da mettere in rete 5 minuti dopo l’uscita sul mercato dei cosiddetti CD “protetti”.

Quando la conversione deve avvenire all’interno di una apparecchiatura si cerca invece di eliminare il passaggio all’analogico rendendolo virtuale. In pratica ad ogni istante i campioni dell’ingresso vengono utilizzati per pilotare un filtro interpolatore a coefficienti variabili (determinati in base alle specifiche o ad un’analisi dei clock in ingresso e in uscita). Si trovano facilmente degli integratini che fanno tutto in modo molto economico (come il **Crystal 8420**) ed è molto semplice imbottire apparecchiature professionali di tutte le conversioni che servono.



Struttura di un convertitore asincrono (figura ripresa dal testo di Watkinson p.148).

Nella pratica audio la conversione **96000** -> **44100** è usata tutte le volte che del materiale da studio registrato col nuovo standard **DVD** deve venire riversato anche su **CD**. La cosa è ragionevole: sarebbe poco realistico sperare che i professionisti, mentre realizzano una registrazione su **48** canali, disponessero di due sistemi di registrazione uno a **96000** per il **DVD** uno a **88200** per il **CD**. Però la cosa che fa veramente arrabbiare è che questo ripiego, adottato evidentemente per risparmiare, talvolta venga sbandierato come un pregio di quel **CD**. I puristi, che registrano direttamente in stereo usano la frequenza di **88200** e la conversione per il *master* avviene certamente senza danni. Quindi se vedete **CD** con scritto sopra **“REGISTRATO a 24/96”**, compratelo se vi piace, ma tenete presente che quella scritta evidenzia un possibile difetto, certamente non un pregio.

Talvolta in alcune apparecchiature consumer si usa anche la conversione opposta (**44100 -> 96000** oppure **44100 -> 192000**). Perché si scelga questa strada invece del tradizionale e robusto oversampling non mi è evidente (anche se a distanza e senza conoscere i criteri di progetto si possono solo fare ipotesi basate su dati parziali).

Ipotesi 1. Il convertitore asincrono viene inserito per ridurre il *jitter*. Alcuni autori infatti suggeriscono di sfruttare il rapporto variabile tra il clock di ingresso e quello di uscita per ottenere un clock in uscita assolutamente variabile a partire da uno in ingresso liberamente fluttuante. Per esempio nelle specifiche del convertitore asincrono **Crystal 8420** si parla di una riduzione fino a **50dB** del *jitter* a **30Khz**. Resta il dubbio che il rimedio (la conversione asincrona) non sia meno pericoloso del male (il *jitter*).

Ipotesi 2. Il convertitore di frequenza viene introdotto per utilizzare lo stesso stadio di conversione **D/A** sia per la decodifica **CD** che quella **DVD** (questa scelta può ridurre i costi di progetto anche per apparecchiature solo **CD** permettendo di utilizzare parti progettate per il **DVD**).

Bisogna considerare anche che la complicazione della conversione asincrona (magari implementata con integrati commerciali) non permette quell'ottimizzazione dei filtri digitali ai fini dell'ascolto, che invece è possibile con le conversioni a frequenza fissa.

Trasporto del segnale digitale

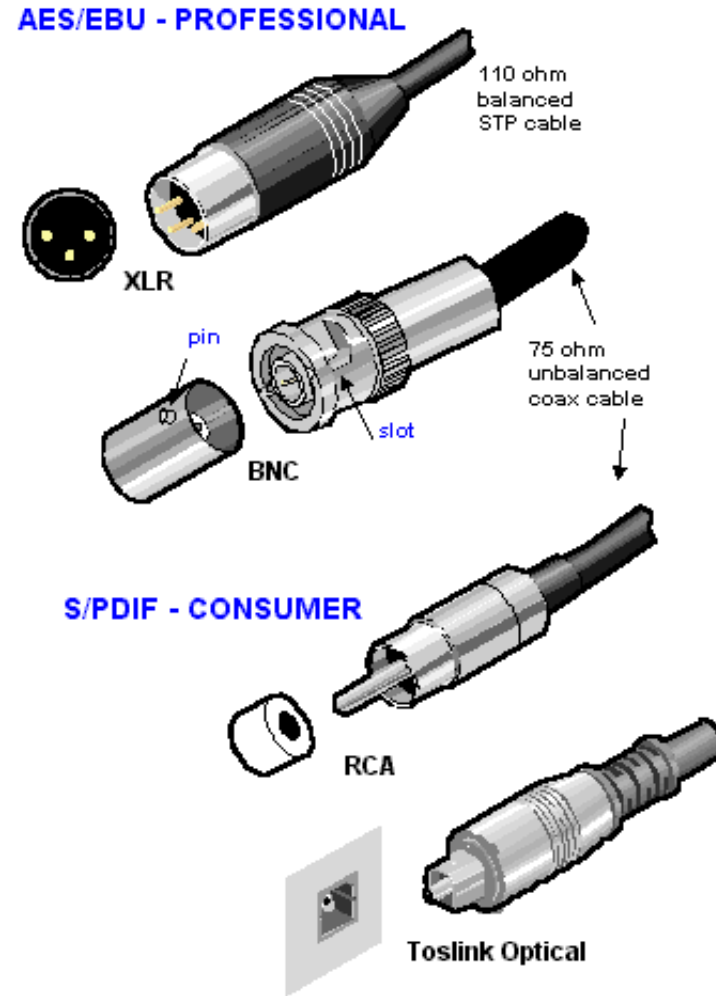
Formati seriali SP/DIF e AES-EBU

I protocolli tipici di trasmissione del segnale digitale sono i seguenti:

- **SP/DIF (Sony Philips Digital Interface)** Il segnale viene codificato in modo sequenziale e viaggia o su cavo coassiale video (da 75Ω) o su fibra ottica.
- **AES/EBU** usato in campo professionale è simile al SP/DIF tipo ma il segnale viaggia su una linea bilanciata (il cavo audio bilanciato usato in ambiente professionale).

Il bit dei campioni audio vengono impaccati con una codifica detta **biphase-mark** che permette di avere un cavo insensibile alla polarità e di trasmettere il clock insieme al segnale. In entrambi i casi il clock è sepolto nel segnale e va ricostruito al momento della ricezione. In questo caso la ricostruzione (che di solito avviene con un PLL) può essere difettosa dando origine a perdita di dati.

Una costosa alternativa, inevitabile quando si devono interfacciare molte diverse apparecchiature digitali (tipicamente negli studi) è far viaggiare un clock di riferimento su linee separate (**word clock**). In questo caso non vi sono problemi di ricostruzione del clock.

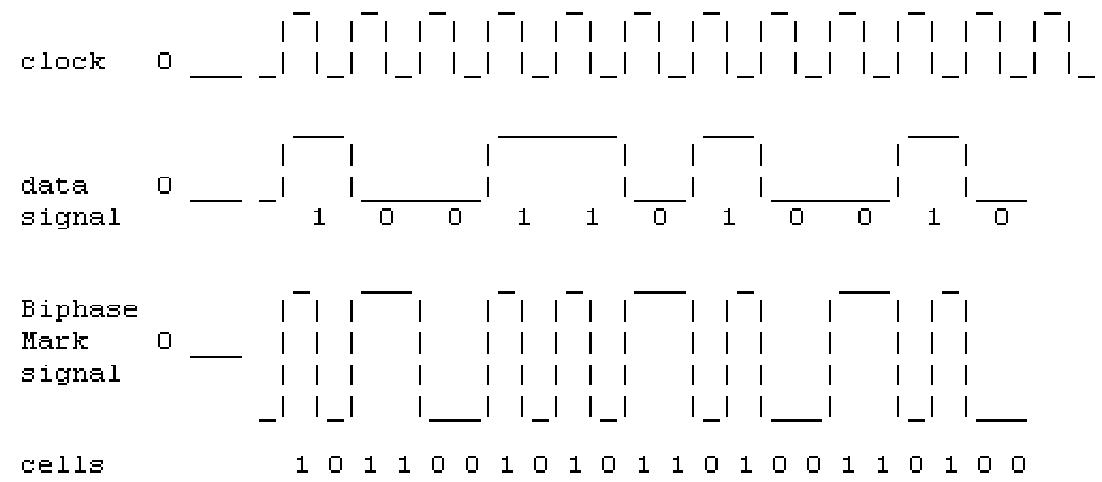


Le frequenze in gioco sono

2.8224 Mbit/s (44.1 kHz sampling rate, CD, DAT)

3.072 Mbit/s (48 kHz sampling rate, DAT)

e siamo nella banda dei segnali video (il cavo **SP/DIF** è un normale cavo video da 75Ω)



Formati seriali **USB** e **FireWire**

Il bus **USB 1.0** e **1.1**. era stato introdotto per connettere periferiche a bassa velocità (mouse, tastiere, floppy-disk). Con un limite teorico intorno ai **10 Mbit/sec** può essere usato anche per Hard-Disc e Penne e lettori mp3.

Attenzione, se i dati musicali viaggiano sotto forma di file (.wav o .mp3) la loro integrità è garantita, ma la trasmissione è off-line.

Attenzione, se i dati musicali viaggiano in streaming diretto (come quando si collega un convertitore o una scheda audio, si ha la possibilità di suonarli in tempo reale ma la loro integrità non è garantita (si possono perdere campioni).

I bus **USB 2.0** e **FireWire** con la loro capacità di **480** e **400 Mbit/sec** (**800** per il **FireWire** ad alte prestazioni) permettono di connettere schede multicanale con formati estesi e con una ragionevole certezza di non perdere dati.