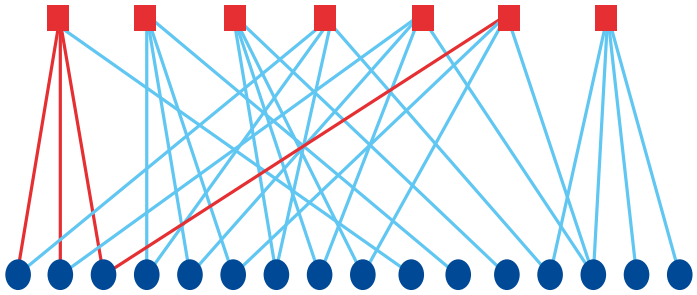
Che cosa è, come funziona

Rivelazione, Correzione e Mascheramento degli Errori

Parte II



Marzio **Barbero,**
Natasha **Shpuza**

1. Introduzione

La prima parte di questo articolo è stata pubblicata nel numero di aprile 2008 [1], a cui è seguita, nel numero successivo, una ricostruzione storica [2] degli eventi che hanno portato, a partire dalla formulazione della teoria di Shannon, ai recenti sviluppi delle tecniche per la protezione dell'informazione. Tali tecniche sono fondamentali per le prestazioni dei sistemi di telecomunicazione e memorizzazione dei dati sviluppati e standardizzati negli ultimi anni.

Tali articoli evidenziano il ruolo dei codici convoluzionali e codici RS (Reed Solomon) sia nelle comunicazioni spaziali, sia nei sistemi di diffusione televisiva di prima generazione (DVB-S e DVB-T). Il § 2 che segue è dedicata ad una illustrazione della codifica convoluzionale, avendo già trattato i codici RS in [1] § 7.

Si è visto in [2] § 7 che la riscoperta dei codici LDPC, proposti nel lontano 1962 da Robert Gallager, è alla base delle prestazioni dei sistemi utilizzati per le missioni spaziali del futuro e dei sistemi di diffusione televisiva di seconda generazione (DVB-S2, DVB-T2 e DVB-C2). Le prestazioni di questi sistemi raggiungono il limite teorico di Shannon ([2] § 2). A questi codici è dedicata il § 4 dell'articolo.

Sommario

Questa è l'ultima parte di un trittico di articoli, di cui i primi due sono stati pubblicati nei numeri di aprile e agosto del 2008. Lo scopo è quello di fornire informazioni di base utili alla comprensione delle tecniche utilizzate per proteggere i dati dagli errori introdotti dai canali di trasmissione o dai sistemi di memorizzazione e registrazione. Tali tecniche sono alla base dei sistemi di diffusione televisiva, in particolare gli standard DVB. In questo articolo conclusivo sono considerati i codici di tipo convoluzionale, utilizzati nella prima generazione DVB, e i codici LDPC, adottati nella seconda generazione. Questi codici trovano ora applicazione anche per migliorare le prestazioni dei servizi di downloading e streaming basati su protocollo IP e per incrementarne le prestazioni in termini di capacità dei sistemi di memorizzazione (hard-disk).

2. Codifica convoluzionale

La codifica convoluzionale e la codifica a blocchi costituiscono le due forme principali di FEC. Essi differiscono fra loro poiché i codici convoluzionali non spezzano il flusso di dati da codificare in blocchi di lunghezza fissa, bensì la ridondanza è aggiunta in modo continuo al flusso codificato.

Si è visto in [2] che la codifica convoluzionale ha trovato impiego sin dalle prime missioni spaziali, grazie alla semplicità costruttiva del codificatore, realizzabile grazie a pochi flip-flop e alcune porte logiche.

In figura 1 è rappresentato, come esempio, il "miglior" codice con *rate* R=1/2 e L=3.

Il *code rate* R e la *constraint length* L sono i due parametri principali che caratterizzano il codice. Nella codifica convoluzionale si opera sul flusso binario seriale: R è il rapporto k/n dove n è il numero di bit in uscita dal codificatore in corrispondenza di k bit di informazione in ingresso. I bit in uscita sono generati in funzione dei k bit in ingresso e dei precedenti L-1 blocchi di k bit, per cui si ha memoria del flusso di dati già codificato. Un ulteriore parametro che caratterizza il codice è la distanza libera d_{free} (*free distance*), la distanza di Hamming minima fra differenti sequenze codificate.

Acronimi e sigle	
3GPP	3rd Generation Partnership Project
BCH	Bose, Chaudhuri, Hocquenghem (codice)
DTMB	Digital Terrestrial Multimedia Broadcast
DVB	Digital Video Broadcasting, (www.dvb.org)
-RCS	Return Channel via Satellite
FEC	Forward Error Correction
FLUTE	File Delivery over Unidirectional Transport
G.hn	home network
IRA	Irregular Repeat Accumulate
LDPC	Low-Density Parity-Check
LT	Luby Transform (codice)
MBMS	Multimedia Broadcast/Multicast Services
MPEG	Motion Picture Expert Group
RAPTOR	RAPid TORnado
RS	Reed Solomon (codice)
SISO	Soft-In-Soft-Out
SoC	System on Chip
UMTS	Universal Mobile Telecommunications System

Fig. 1 - Il codificatore nello schema è composto da due componenti elementari: flip-flop e sommatori binari (XOR). E' caratterizzato da un *code rate* R=1/2, quindi il numero di bit in uscita è il doppio di quelli entranti. Il flusso in ingresso è applicato al registro a scorrimento che dispone di uscite intermedie (prese) in corrispondenza di ciascun stadio; ad ogni bit in arrivo all'ingresso, il flusso avanza di un colpo di clock e l'uscita è ottenuta prelevando alternativamente i bit ottenuti alle due uscite 1 e 2; il bit-rate in uscita è quindi doppio rispetto a quello in ingresso. La *constraint length* L corrisponde al numero di prese e la presenza o meno delle connessioni alle uscite dei flip-flop corrispondono ai polinomi generatori del codice G_1=101 (cioè manca la presa intermedia) e G_2=111 (l'uscita 2 dipende dai tre bit presenti nello shift). I polinomi generatori sono generalmente espressi in ottale e quindi questo codice è noto come L=3 (5,7), d_{free}=5.

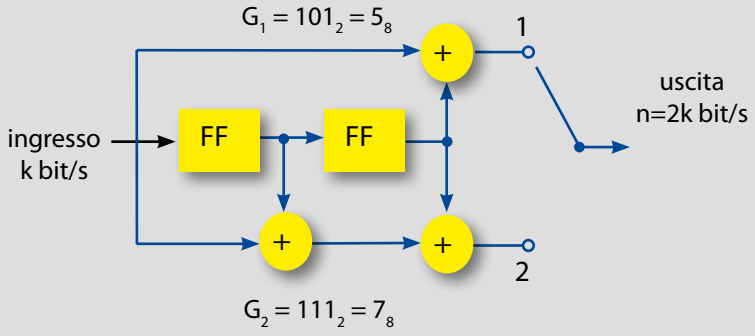
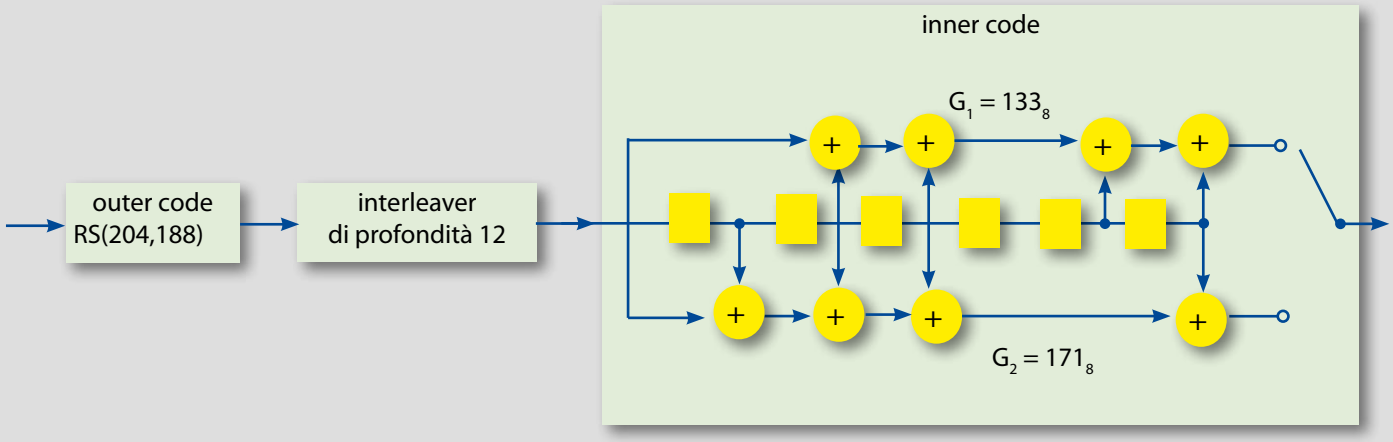


Fig. 2 - Nel sistema DVB-S lo schema del trasmettitore prevede un codice esterno accorciato RS (204,188), seguito da un interleaver con profondità 12 e dal codice interno, convoluzionale con code rate R=1/2, *constraint length* L=7, polinomi generatori $G_1=171_8$ e $G_2=133_8$, d_{free}=10.



La struttura di un codificatore di questo tipo corrisponde a quella di una macchina a stati finiti. Il numero di stati cresce esponenzialmente con il crescere di L: per valori piccoli di L nella decodifica è utilizzato l'algoritmo di Viterbi ([2] § 4), che presenta il vantaggio di richiedere un tempo fisso per la decodifica. In pratica si confronta una sequenza piuttosto lunga di bit ricevuti con tutte le possibili sequenze e si sceglie quella più prossima (criterio di massima verisimiglianza), ricavando da essa k bit ogni n bit ricevuti.

La decodifica introduce un ritardo, proporzionale alla sequenza esaminata, e una decisione errata può influenzare anche le successive decisioni: si può avere propagazione degli errori ed in tal caso gli errori si presentano a burst.

Per limitare gli effetti dei burst di errori si può ricorrere alla tecnica di due codici posti in cascata, un codice esterno seguito da un codice interno, così come si è visto parlando dei codici prodotto. In questo caso però si utilizza il termine codici concatenati (*concatenated codes*) per indicare questa tecnica.

Lo schema adottato per le comunicazioni spaziali a partire dal 1977 era basato su il codice convoluzionale R=1/2 e L=7 come codice interno, da simboli costituiti da byte protetti da un codice esterno RS(255,233,33), in grado di correggere fino a 16 byte errati.

Anche nel caso del sistema di diffusione da satellite di prima generazione, il DVB-S, è stato adottato uno schema analogo (figura 2), basato su un codice convoluzionale R=1/2 e L=7. In questo caso si adotta un RS accorciato RS(204,188) derivato dal RS(255,239) e in grado di correggere fino a 8 byte errati o 16 byte nel caso di *erasure* ([1], § 7.3).

Per ridurre gli effetti dei burst di errori è utilizzata la tecnica dell'interleaving ([1], § 6). Lo schema di interleaving (usato in trasmissione) e de-interleaving (utilizzato in ricezione) convoluzionale è quello di figura 3.

La codifica convoluzionale permette la ricezione in caso di canali particolarmente rumorosi, ma richiede un'elevata ridondanza. Con R=1/2, metà della capacità del canale è utilizzata per la ridondanza.

Si può ridurre tale ridondanza, per sfruttare maggiormente la capacità per trasmettere i dati utili, se le caratteristiche del canale lo consentono e se si può accettare una riduzione di d_{free} e la conseguente minore robustezza.

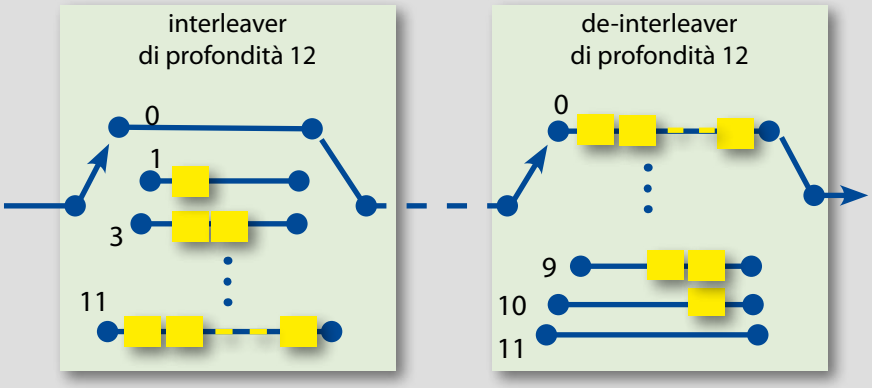


Fig. 3 - Nel sistema DVB-S l'interleaver consiste in 12 rami, ciclicamente interconnessi al flusso di byte in ingresso: ciascun ramo j è costituito da uno shift register (FIFO) di Mj celle. Il valore di M è pari a 17, rapporto fra il numero di byte costituenti il codice (204) e la profondità di interleaving (12). Al ramo 0 non corrisponde alcun registro a scorrimento, e quindi il ritardo introdotto è nullo, al ramo j=11 corrisponde un registro di 17*11=187 byte. Il de-interleaver ha una struttura analoga, ma l'indice j=0 corrisponde al ritardo massimo (187 byte) e quello 11 al ritardo nullo.

A tale scopo si utilizza la tecnica di *puncturing*, cioè si "perfora" il flusso di dati avviati al modulatore, ovvero si eliminano alcuni dei bit in base ad una opportuna matrice (tabella 1). In fase di decodifica è noto il valore di code rate *R* utilizzato dal codificatore: tanto più è prossimo a 1, tanto più bassa è la ridondanza e meno robusto il codice. Il valore di *R* può essere variato nel tempo in funzione delle caratteristiche dal canale; si consideri ad esempio il caso di una sonda spaziale che si allontana dalla terra: al diminuire dalla potenza ricevuta, si accetta una riduzione della velocità dei dati ricevuti, pur di continuare l'acquisizione delle preziose informazioni raccolte.

3. I Turbo Codici

Nei primi anni '90 l'uso dei codici a blocco di tipo RS e di quelli convoluzionali distavano ancora da quelle teoricamente raggiungibili (limite di Shannon) di più di 3dB, ovvero gli schemi di codifica praticamente realizzabili allora richiedevano che i sistemi di comunicazione dovessero utilizzare, a parità di prestazioni, un'energia almeno doppia della minima teorica.

Nel 1993 due ingegneri elettronici francesi, Claude Berrou e Alain Glavieux, proposero uno schema di codifica che consentiva un miglioramento tale da approssimarsi al limite di circa 0,7 dB.

Questi codici vennero denominati Turbo Codici perché nella decodifica si utilizza un percorso di retroazione, analogamente a quanto avviene in campo automobilistico con i motori turbo.

Tab. 1 - Matrici di perforazione usate per le telecomunicazioni satellitari e specificate per lo standard DVB-S. Se, ad esempio, si vuole utilizzare il codificatore di figura 1 con un *code rate R*=2/3, si invieranno solo i bit in posizione pari in uscita dal ramo superiore e tutti i bit in uscita dal ramo inferiore.

R	1/2	2/3	3/4	5/6	7/8
puncturing matrix	1 1	10 11	101 110	10101 11010	1000101 1111010
d_{free}	10	6	5	4	3

Un turbo codice è formato dalla concatenazione parallela di due codici separati da un interleaver.

In figura 4, a titolo di esempio, è riportato lo schema del codificatore turbo utilizzato per il sistema di telefonia mobile di terza generazione UMTS.

In generale la scelta dei codificatori e dell'interleaver è libera, ma la maggior parte delle realizzazioni si basa sui criteri adottati per quello in figura: i due codificatori sono identici; il codice è sistematico, ovvero i bit in ingresso sono anche presenti all'uscita; l'interleaver legge i bit in ordine pseudo-casuale.

La scelta dell'interleaver è fondamentale per il progetto dello schema di un turbo codice. L'interleaver pseudo casuale o pseudo-random è definito da un generatore di numeri pseudo casuali o da una look-up table.

Ha due scopi fondamentali. Essendo posto all'ingresso del secondo encoder, la sua uscita ha caratteristiche statistiche completamente diverse dall'uscita del primo encoder, in particolare per quanto riguarda il peso, cioè il numero di 1 presenti. L'uso dell'interleaving pseudocasuale all'ingresso del secondo encoder rende i due flussi ottenuti completamente scorrelati, anche in uscita dai due decoder corrispondenti, e ciò è particolarmente vantaggioso in fase di decodifica.

Fondamentale per ottenere prestazioni ottimali è l'impiego di un metodo *soft -decision* per la decodifica. In un decoder *SISO* la decisione non è basata su una soglia *(hard-decision)* per decidere se il simbolo ricevuto è 0 oppure 1, ma il decoder elabora un valore reale (*soft*) ottenuto dal demodulatore e fornisce in uscita per ciascun bit una stima della probabilità che il bit trasmesso sia un 1.

Nel decoder turbo, le uscite dei due decoder forniscono stime degli stessi bit, ma i bit sono trasmessi in sequenze differenti e ciò permette di trarre un significativo guadagno dalla comparazione delle due informazioni, dopo una appropriata riorganizzazione dei dati stimati.

Un ulteriore guadagno è ottenuto reiterando le stime più volte, usando alternativamente i valori stimati dai due decoder, fino a quando viene deciso in modo definitivo (*hard*) se al bit ricevuto è assegnato il valore 0 oppure 1.

I turbo codici sono utilizzati, oltre che nel sistema UMTS, dallo standard DVB-RCS.

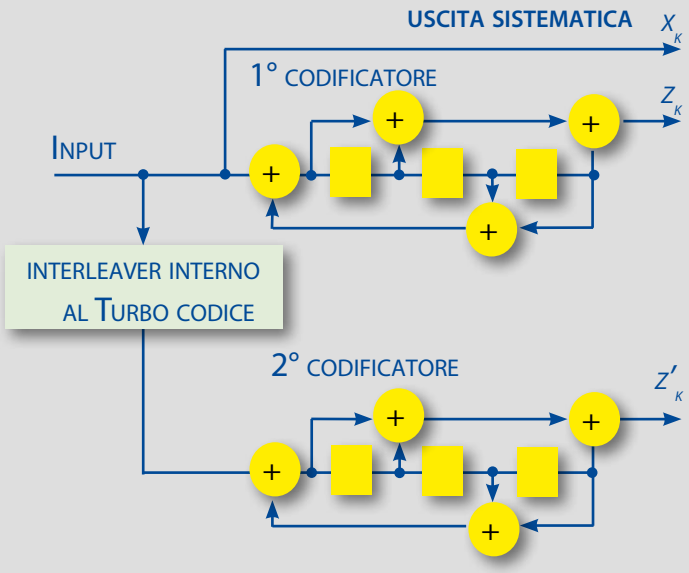


Fig. 4 - Codificatore utilizzato per i sistemi UMTS: segue i criteri di progetto indicati nel documento originale di Berrou e Glavieux del 1993 e utilizza una coppia di semplici codificatori convoluzionali identici. Per ogni bit in ingresso x_k vengono generati un bit x_k (il codice è sistematico) e due bit di parità z_k e z'_k: il code rate *R* è 1/3.

4. I codici LDPC

Con la scoperta dei Turbo Codici si avviò una rivalutazione degli schemi di codifica proposti nel passato, caratterizzati da una ridotta complessità, sfruttatabile per realizzare schemi di decodifica iterativi. Nell'ambito di tale analisi degli schemi precedentemente trascurati fu ripreso il lavoro iniziato con la tesi di dottorato di Robert Gallager nel 1962, sui codici LDPC.

In quanto codici lineari a blocco, i codici LDPC possono essere rappresentati mediante matrici (si veda come esempio il codice di Hamming [1] § 4): la matrice per il calcolo dei bit di parità H e la matrice generatrice G. Per migliorare l'efficienza del codice, la matrice H deve essere costruita in modo che la distanza minima sia la più grande possibile: ciò implica che la matrice sia "sparsa", ovvero gli 1 siano in numero ridotto rispetto agli 0, da qui la denominazione di codice LDPC, cioè a bassa densità dei bit di parità. La matrice H è caratterizzata da n colonne e da n-k righe, dove k sono i bit di parità.

Se il numero di 1 presenti sulle colonne è costante e se il numero di 1 presenti sulle righe è costante il codice LDPC si dice regolare, altrimenti è detto irregolare. In figura 5 è un esempio molto semplice di codice LDPC con una matrice H (20,15).

I codici LDPC hanno il vantaggio di offrire prestazioni prossime a quelle teoriche (limite di Shannon), di essere adatti a differenti tipi di canale e di richiedere tempi di decodifica che crescono linearmente con le dimensioni del blocco. Inoltre sono realizzabili schemi che permettono un elevato grado di parallelismo sia in fase di codifica che di decodifica.

Sono stati proposti diversi algoritmi per costruire matrici H, anche utilizzando schemi di generazione pseudo casuali. Non è un problema complesso quello di generare codici LDPC con buone prestazioni e quelli con prestazioni migliori sono di tipo irregolare. La difficoltà di progettazione consiste nel mantenere bassa la complessità del codificatore e del decodificatore.

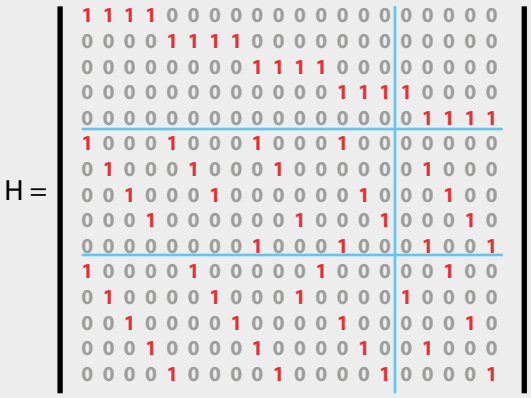


Fig. 5 - L'esempio proposto da Gallager nel 1963. E' basato su una matrice H (*parity check matrix*) costituita da 20 colonne e 15 righe caratterizzata da avere 4 elementi a 1 per ciascuna riga e tre elementi a 1 per ciascuna colonna.

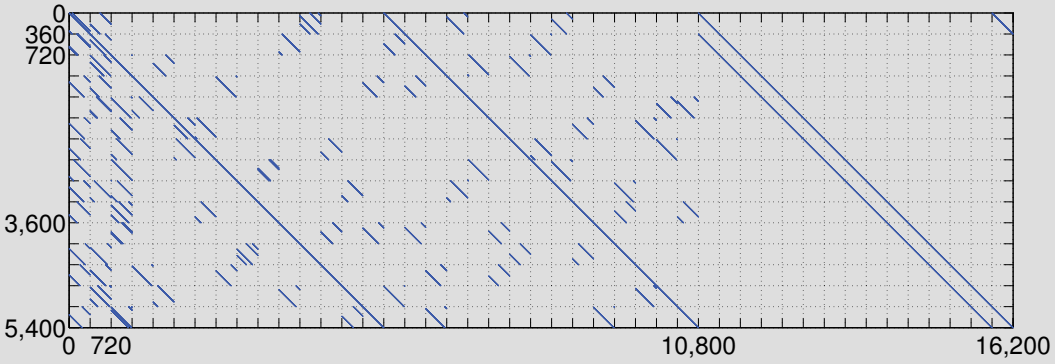


Fig. 6 - In pratica la matrice H deve essere di grandi dimensioni ed è importante definire la sua struttura in modo da minimizzare la complessità del codificatore e del decodificatore, anche in termini di memoria per rappresentarla. Nell'esempio la disposizione degli 1, rappresentati dalle linee diagonali, è tale da consente un risparmio di memoria e la possibilità di individuare una struttura in sottomatrici. Questa matrice è costituita da 16200 colonne e 5400 righe ed è adottata nei sistemi DVB di seconda generazione.

Infatti le prestazioni dei codici migliorano al crescere delle dimensioni del blocco, come già era stato evidenziato da Elias nel 1955. I codici LDPC possono avere prestazioni migliori dei Turbo Codici quando la lunghezza del blocco è elevata, dell'ordine di alcune decine di migliaia di bit (figura 6)

Una rappresentazione alternativa del codice è quella grafica, introdotta da Tanner (figura 7). Se la matrice H è scelta senza alcuna restrizione, le connessioni, rappresentabili con il grafico di Tanner, appaiono a distribuite a caso e l'accesso alle n connessioni implica n cicli di clock. E' quindi vantaggioso adottare strutture di H tali da consentire una parallelizzazione parziale, in modo che un certo numero di nodi venga processato in parallelo.

Così come per i Turbo codici, anche per quelli LDPC il guadagno in termini di prestazioni è ottenuto grazie alla decodifica iterativa. Il numero di iterazioni è elevato (almeno 30) e il numero di calcoli, seppur semplici, per ciascuna iterazione è proporzionale alle dimensioni della matrice, pertanto la complessità totale della decodifica è superiore a quella richiesta per i Turbo Codici. Al crescere del numero di iterazioni, cresce la quantità di memoria necessaria e si incrementa la latenza (il ritardo nella decodifica).

Anche nel caso dei codici LDPC le realizzazioni pratiche utilizzano la decodifica *soft*, che consente una più rapida convergenza dell'algoritmo di decodifica.

I codici LDPC sono stati adottati per la prima volta in uno standard dal gruppo DVB-S2 [3]. Lo schema scelto per il sistema di diffusione via satellite di seconda generazione DVB-S2, è stato successivamente adottato per il DVB-T2 per il terrestre [4] e DVB-C2 per la distribuzione via cavo

La struttura prescelta è basata sulla concatenazione di due codici: un BCH come codice esterno e un LDPC come codice interno.

Il codice LDPC utilizzato dal DVB è denominato *Extendet IRA code*, ha il vantaggio di una ridotta (lineare) complessità del codificatore. La lunghezza di parola può essere n=64800 (per le trame normali) o n=16200 (per le trame corte). Per comprendere

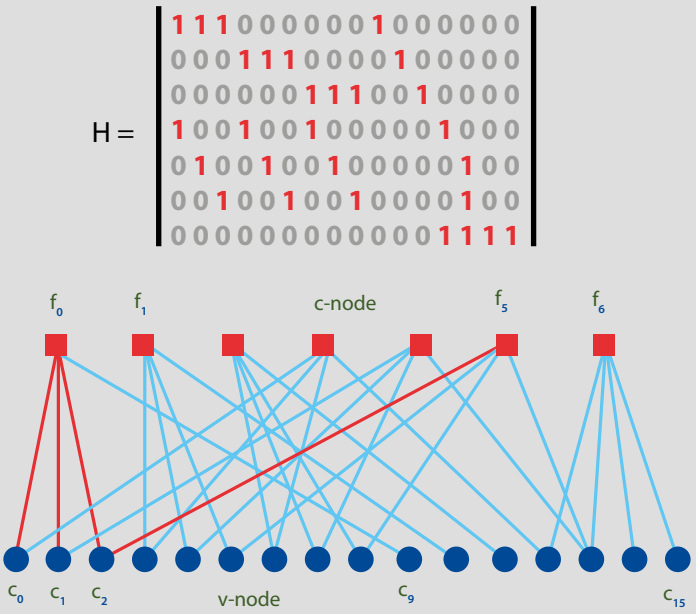


Fig. 7 - Nella rappresentazione proposta da Tanner il codice è descritto da un grafico bipartito.
Consiste di due tipi di nodi: i v-node e i c-node.
I v-node (*variable-node*) sono di solito indicati con la lettera *c* e sono in numero pari al numero di bit che costituiscono la parola di codice, ovvero *n*, quante le colonne della matrice H. I c-node (*check node*) sono indicati con la lettera *f* e sono tanti quante le righe di H, ovvero *n-k*. Il c-node f_i è connesso con il v-node c_j solo se l'elemento h_{ij} della matrice H è pari a 1.
L'algoritmo di decodifica prevede che ciascun nodo *f* riceva i contributi dai nodi *c* ad esso collegati e determini i valori di parità ritenuti corretti. Il calcolo può essere *hard*, semplice XOR dei contributi, oppure di tipo *soft*. Nel passo successivo tali valori sono inviati ai nodi *c*, e per ciascuno di essi è determinato, in base al valore precedente e ai contributi dei nodi *f* connessi, il nuovo valore di *c*.
Il processo è iterativo, ha termine quando non vi sono più variazioni nei valori di *c*, e quindi la parola di codice calcolata è ritenuta corretta, oppure quando si raggiunge il numero massimo previsto di iterazioni.
Una scelta opportuna della disposizione degli 1 nella matrice H, e quindi dei collegamenti fra i nodi del grafico, consente di effettuare i calcoli in parallelo, diminuendo il tempo di latenza dovuto alla decodifica.
Nell'esempio riportato, la matrice H in alto e il grafico di Tanner in basso corrispondono allo stesso codice, molto semplice. Il nodo f_0 è calcolato in base ai contributi c_0, c_1, c_2 e c_9 e, nel passo successivo, contribuisce, insieme a f_5, a determinare il nuovo valore di c_2...

la complessità di decodifica, si consideri che, con la lunghezza di 64800 bit, ad ogni iterazione è necessario accedere e calcolare circa 300000 dati e il numero di iterazioni per garantire le prestazioni è pari a 30.

Per ridurre la complessità e la latenza nelle fasi di codifica e decodifica, i codici LDPC adottati dai sistemi DVB di seconda generazione sono caratterizzati da "matrici di parità" H sparse, disposte in modo da consentire l'individuazioni di sottomatrici, su cui operare in parallelo (figura 6).

L'uso dei codici LDPC permette un'elevata flessibilità nella scelta del *code rate,* infatti *R* può assumere per il DVB/S2 i valori 1/4, 1/3, 2/5, 1/2, 3/5, 2/3, 3/4, 4/5, 5/6, 8/9, e 9/10; mentre nel caso del DVB/T2 i valori possibili sono 1/2, 3/5, 2/3, 3/4, 4/5, 5/6.

Altri standard che utilizzano questi codici sono le nuove versioni di WiMax mobile e WiFi. Anche il sistema di televisione terrestre DTMB della Repubblica Popolare Cinese adotta uno schema basato su BCH e LDPC, con una lunghezza di 7488 bit.

5. Mascheramento

Nella prima parte [1] si è visto che l'adozione dei codici serve a rivelare la presenza degli errori e, normalmente, a correggerli riducendo la probabilità di errore residua a valori tali da rendere non percepibili o accettabili gli effetti dovuti agli errori non corretti.

In [1] § 5 si è accennato agli effetti del superamento della capacità correttiva del codice: in ricezione si passa da una condizione ottimale, in cui tutti gli errori sono corretti, ad una condizione in cui, anziché recuperare integralmente l'informazione, vengono introdotti ulteriori errori e si perviene rapidamente ad una condizione di non funzionamento, di non-servizio.

Nel caso in cui il servizio lo consenta, si richiede la ritrasmissione dell'insieme di dati non ricevuti correttamente: ciò è possibile se è disponibile un canale di ritorno e se i dati possono essere ritrasmessi, ad esempio nel caso in cui si tratti di file di dati prodotti da un server. Oppure si attende una nuova ricezione degli stessi dati, nel caso di ritrasmisisone ciclica, come avviene per alcuni servizi quali il Televideo, esempio riportato in [1].

Nel caso della diffusione di informazioni audio e video, in broadcasting o streaming, si può invece sfruttare la ridondanza residua, ancora presente nell'informazione ricevuta anche quando siano state utilizzate efficienti tecniche di compressione e codifica.

Le strategie di correzione e di interleaving consentono non solo di ridurre, grazie ad esempio alla concatenazione di codici, il numero di errori residui, cioè quelli che superano le capacità correttive del codice, ma anche di fornire una valutazione dell'attendibilità dei dati, in particolare se si adottano tecniche di decisione *soft*, al fine di segnalarla allo stadio di decodifica successivo.

Al momento in cui si estraggono dai dati associati al flusso binario le informazioni relative ai pixel d'immagine o ai campioni audio, è quindi nota l'affidabilità di tali informazioni e, se è elevata la probabilità che siano erronee, è possibile ricorrere alla tecnica del mascheramento (*concealment*).

Ad esempio, nel caso del segnale video, è possibile utilizzare la correlazione con i pixel contigui o fra blocchi e macroblocchi (nel caso di codifica MPEG) contigui nel tempo (appartenenti ai quadri precedenti o successivi) e ricostruire l'informazione mancante con pixel, blocchi o macroblocchi stimati a partire da quelli presenti in memoria e ritenuti corretti. Nel caso di perdita di gran parte dell'informazione (ad esempio in presenza di fading o rumore impulsivo) il mascheramento avviene congelando l'intera immagine presente in memoria, fino a quando la decodifica riprende correttamente.

Le tecniche di mascheramento di solito danno origine a difetti percepibili dall'utente, ma meno fastidiosi rispetto all'interruzione completa del servizio o alla visualizzazione di porzioni di immagine completamente scorrelate.

6. Cancellazioni (erasure)

Una sempre maggior mole di traffico dati si attua sulla rete Internet. Il protocollo internet (IP) prevede l'organizzazione dei dati in pacchetti dotati di una intestazione (*header*) che racchiude l'indirizzo la sorgente e la destinazione del pacchetto e spesso anche un numero che ne indica la posizione assoluta o relativa all'interno della sequenza che compone il flusso (*stream*) di dati. I pacchetti vengono instradati, seguendo i percorsi ritenuti più opportuni, attraverso la rete fino a raggiungere la destinazione. A volte, per le ragioni più varie (ad esempio *overflow* dei *buffer* presenti nei *router* intermedi), alcuni dei pacchetti non raggiungono la destinazione oppure i pacchetti ricevuti non vengono considerati validi, perché sono rivelati errori non correggibili.

In questi casi si potrebbe procedere con la richiesta di ritrasmissione dei pacchetti mancanti, cioè cancellati (*erasure*), ma di fatto tale approccio può risultare non praticabile a causa della distanza fra il ricevitore e la sorgente (e quindi del tempo di latenza), della complessità o tipologia della rete (canali *wireless* o via satellite) o dal tipo di sorgente (un server che deve servire contemporaneamente più utenze e quindi non è in grado di gestire le richieste di ritrasmissione).

I *fountain code* sono codici che offrono una soluzione: la sorgente invia l'informazione in modo ridondante, tale da consentire la sua ricostruzione anche da parte del ricevitore soggetto ad un numero di erasure non superiore a quello massimo prevedibile (*worst case*). Il nome *fountain* è rappresentativo del concetto della fontana (la sorgente) che riempie il bicchiere (il ricevitore) anche nel caso in cui parte dell'acqua (i pacchetti dati) prodotti dalla fontana va persa.

Il *fountain code* produce, a partire da *k* simboli, un flusso teoricamente illimitato di simboli; il decodificatore è in grado di recuperare i *k* simboli a partire da un insieme di *n* simboli ricevuti; il codice ha buone prestazioni se *n* è prossimo a *k*, e se il tempo di decodifica è direttamente proporzionale a *k*. Il simbolo è, genericamente, un vettore di bit, ad esempio un pacchetto dati.

I codici LT sono stata la prima classe di *fountain code* utilizzabile in pratica. I *Raptor code* [5] sono una delle classi di *fountain code* caratterizzati da un tempo di codifica e decodifica lineare rispetto a *k*. Questi codici hanno due stadi di codifica: un *pre-code* o *outer-code* e un *inner-code*. Il *pre-code* può, a sua volta, essere la concatenazione di due codici, ad esempio un codice di Hamming e un LDPC. Il codice interno è un codice LT.

Alcuni dei più recenti standard quali il MBMS nell'ambito del 3GPP per la telefonia mobile di terza generazione e DVB-H [6] e DVB-SH per il datacast verso i dispositivi mobili adottano il protocollo FLUTE che prevede, opzionalmente, l'uso dei codici *Raptor*.

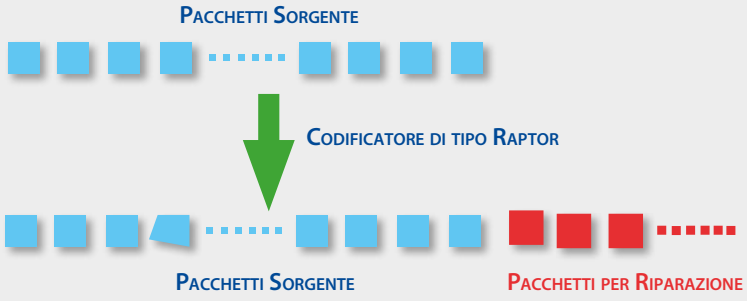


Fig. 8 - I dati che costituiscono i file da trasmettere sono generalmente protetti da FEC, a livello fisico. Il file è organizzato in un numero fisso di *source symbol*; a livello applicazione il codificatore *Raptor* genera, oltre ai *source symbol,* un numero di *repair symbol* variabile in funzione del numero di *erasure* previsti nelle condizioni peggiori, cioè di pacchetti che si ritiene probabile vadano persi. Se è disponibile un canale di ritorno, si possono anche prevedere servizi in cui vengano generati, su richiesta, ulteriori *repair symbol*, destinati agli utenti che non sono stati in grado di ricostruire l'intera informazione. Lo scopo è quello di adattare la banda utilizzata, minimizzando quella necessaria a contrastare le *erasure* e riducendo al minimo la latenza, almeno per gli utenti che godono di migliori condizioni di ricezione.

I codici *Raptor* previsti in FLUTE *sono* di tipo sistematico, cioè i simboli del messaggio originale sono compresi fra quelli ricevuti. Si prevede la mappatura dei file in simboli denominati simboli sorgente (*source symbol*) e la generazione di simboli aggiuntivi utilizzati per la riparazione (*repair symbol*) una appropriata strategia basata sull'adozione di codici per la correzione degli errori consente di minimizzare sia l'effetto della perdita di pacchetti sia l'occupazione di banda (figura 8).

7. Oltre il limite?

Nel mese di luglio è stato diffuso un comunicato stampa su un nuovo dispositivo SoC basato su LDPC per gestire la memorizzazione dei dati su hard-disk. In particolare esso trova applicazione per i dischi da 2,5 pollici, il segmento a crescita più veloce nel mercato degli hard-disk, portando la capacità a 320 GB e quindi consentendo di continuare l'andamento nella crescita di capacità di memorizzazione, che raddoppia ogni 18 mesi.

Rappresenta l'ultimo progresso, in ordine di tempo, che possiamo attribuire al contributo essenziale delle tecniche di protezione degli errori.

Gli articoli che compongono questo trittico illustrano una storia, quella che ha portato alle prestazioni finali dei sistemi attuali. Le tappe principali di questa storia sono riassunte in figura 9.

Abbiamo visto in [2] che è una storia che ha un inizio ben definito, temporalmente: coincide con la la pubblicazione dell'articolo di Shannon nel 1948. Fin dall'inizio la storia introduce due dei principali protagonisti e indica il possibile lieto fine. I protagonisti sono lo stesso C. E. Shannon, e R.W. Hamming, il cui codice è citato in tale articolo come esempio di codice "efficiente". E il lieto fine è costituito dal raggiungimento del limite di Shannon.

Negli anni '50, il computer, appena nato, è il primo campo di applicazione. Hamming inventa il codice proprio per rendere possibile il funzionamento del computer, che altrimenti, a causa degli errori, si blocca in continuazione. Il codice RS (Reed -Solomon) è del 1960; Reed raggiunge la notorietà per aver realizzato un computer compatto, delle dimensioni di una scrivania.

Negli anni '60 e '70, l'attenzione si focalizza sulle comunicazioni spaziali; il Jet Propulsion Laboratory e la NASA diventano il punto di aggregazione degli esperti che collaborano per mettere a punto i sistemi che consentano l'acquisizione delle informazioni raccolte dalle missioni spaziali. Protagonisti sono A.J. Viterbi e I.S. Reed. Il codice Reed-Muller, l'algoritmo di Viterbi per la decodifica dei codici convoluzionali e infine i codici concatenati RS e convoluzionali con decodifica di Viterbi sono le tappe che permettono l'esplorazione dei pianeti del sistema solare a partire dal 1969 fino alla fine del secondo millennio.

Tornando al campo della memorizzazione: gli schemi basati su codici prodotto RS trovano un ampia diffusione sia per la registrazione su supporto magnetico (disco o nastro) che su disco ottico. Sia il CD (1982) che il DVD (1996) lo adottano.

E lo schema che consente le esplorazioni spaziali è anche alla base dei sistemi di diffusione televisiva digitale. Lo schema RS-interleaving-codice convoluzionale-decodifica di Viterbi è adottato per la prima trasmissione digitale sperimentale via satellite delle immagini in alta definizione durante i campionati mondiali di Italia '90 e successivamente dal DVB-S (1994), il sistema di diffusione televisiva via satellite, normalizzato dal gruppo presieduto da Mario Cominetti, del Centro Ricerche Rai.

Il codice RS all'inizio degli anni '90 è il protagonista assoluto nei tre campi (memorizzazione, comunicazioni spaziali, diffusione televisiva digitale) e sembra che i 3 dB che separano dal lieto fine, il raggiungimento del limite di Shannon, siano un ostacolo insormontabile, quando i Turbo codici, "inventati" nel 1993, rendono il limite a portata di mano.

Sono trascorsi pochi anni, necessari perché la novità venisse recepita da alcuni standard di telecomunicazioni (DVB-RCS, HSPA per la telefonia e WiMax), quando il gruppo che si occupa del successore del sistema di diffusione televisiva via satellite, il DVB-

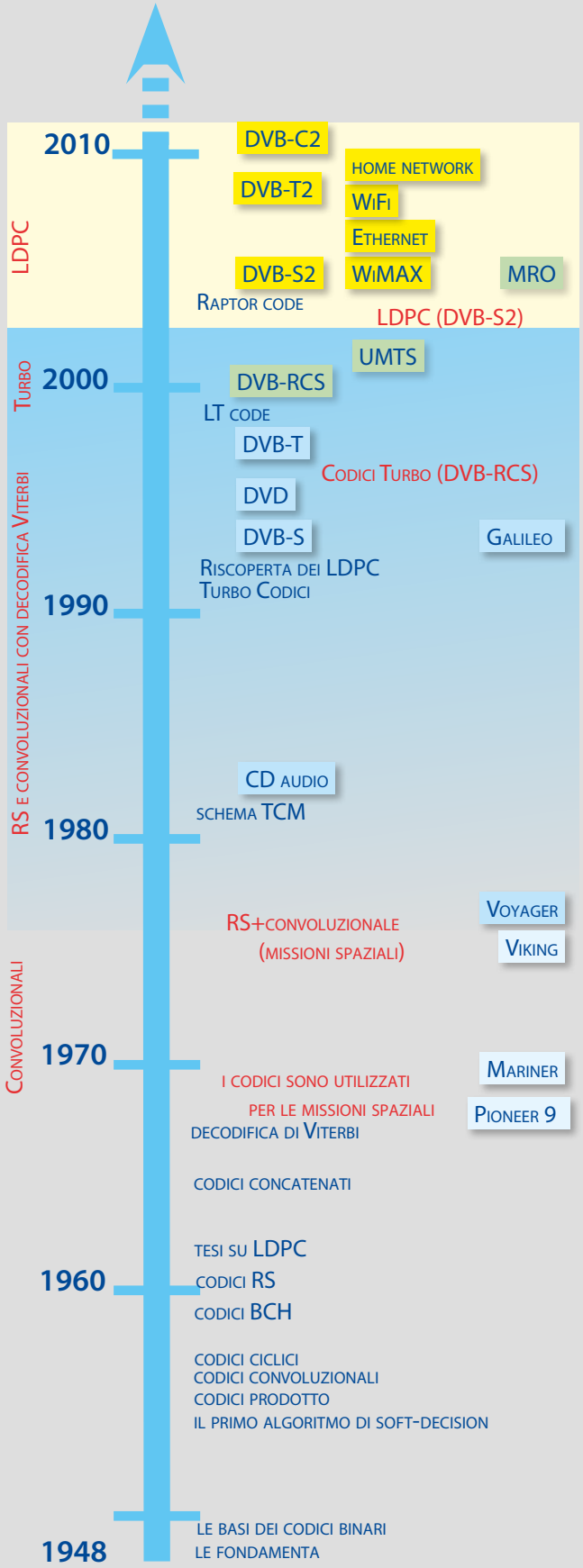


Fig. 9 - L'articolo di C. E. Shannon nel 1948 costituisce le fondamenta della teoria matematica dei codici e indica relazione fra capacità e banda di un canale soggetto a rumore gaussiano.

Gli anni '50 sono densi di importanti intuizioni: viene descritto il primo algoritmo di soft-decision (Wagner, 1954), sono inventati i codici prodotto e viene evidenziato che, per il raggiungimento del limite di Shannon, occorre utilizzare codici con elevata lunghezza di blocco (P. Elias, 1955), sono inventati i codici convoluzionali (P. Elias, 1955), i codici ciclici (E. Prange, 1957), e con la fine del decennio i codici BCH (A. Hocquenghem, 1959, e R.C. Bose e D.K. Ray-Chaudhuri, 1960).

Degli anni '60 sono le basi teoriche degli schemi di codifica oggi più diffusi: i codici RS (I. S. Reed e G. Solomon, 1960) e LDPC (R. Gallager, 1962), la concatenazione dei codici (G.D. Forney, 1965). Inizia l'uso dei codici per le missioni spaziali (Reed-Muller, 1969). I protagonisti in quegli anni collaborano al Jet Propulsion Laboratory della NASA, fra gli altri, Viterbi e Reed. L'algoritmo di Viterbi è del 1967. Nel 1977, con le missioni Voyager, è introdotto lo schema basato sul codice RS concatenato con codice convoluzionale e decodifica di Viterbi. Inizia il lungo periodo di predominanza degli schemi basati su RS e codice convoluzionale.

Un ulteriore significativo passo di avvicinamento al limite è ottenuto applicando la codifica convoluzionale ai simboli della modulazione: è lo schema Trellis Coded Modulation (Ungerboeck, 1982).

Le tecniche di integrazione su larga scala rendono finalmente possibile l'adozione degli schemi di decodifica anche su blocchi di elevata lunghezza, guadagnando quindi in efficienza. Questi schemi possono essere adottati negli standard destinati a prodotti per il grande pubblico. Il primo è il CD audio (1982), che adotta uno schema basato sul prodotto di codici RS. Analogo schema, ma con migliori prestazioni, caratterizza il DVD (1996).

Agli inizi degli anni '90 il limite non è ancora stato raggiunto, ma i risultati degli ultimi 15 anni di sviluppo e realizzazioni sembrano stabili. Uno schema efficiente e collaudato è quello adottato per le missioni spaziali. E' un'ottima ragione per utilizzarlo anche per il nascente standard di diffusione da satellite, DVB-S (1994).

L'evento inatteso, che riduce la distanza che ancora separa dal limite è l'invenzione dei Turbo codici (C. Berrou, A. Glavieux, P. Thitimasjshima, 1993). I turbo codici vengono adottati dagli standard DVB-RCS e UMTS (ETSI TS 125 212, 2001). In campo spaziale è utilizzato dal MRO (Mars Reconnaisance Orbiter, 2005).

Nel 1994 sono riscoperti gli LDPC.

Sono "la soluzione finale", vengono adottati per le nuove generazioni di standard per le comunicazioni: DVB-S2 (2005), WiMAX (IEEE-80216e, 2005), 10GBase-T Ethernet (802.3an, 2006), WiFi (IEEE 820.11n, 2007), DVB-T2 (2008), G.hn (ITU G.9960, 2009), DVB-C2(2010).

S2, presieduto da Alberto Morello del Centro Ricerche Rai, mette in competizione codici Turbo e LDPC.

Vincono gli LDPC, e in breve tempo il successo degli schemi basati su LDPC si estende alle nuove generazioni di standard: dopo il DVB-S2 (2005), è adottato per gli altri standard televisivi DB-T2 per la diffusione terrestre e DVB-C e per la distribuzione via cavo, e dagli standard per i collegamenti a microonde WiMAX, per le reti wireless WiFi, Ethernet 10GBase-T, per la rete domestica G.hn con distribuzione su linee elettriche, telefoniche e coassiali fino a 1 Gbit/s (ITU G.9960, 2009).

Il loro uso si afferma anche a livello di protezione dei servizi per downloading e streaming (protocollo FLUTE).

E infine, come abbiamo visto all'inizio, arrivano anche a sostituire i sistemi basati su RS anche per i sistemi di memorizzazione.

Il comunicato stampa citato proclama l'attualità della nuova architettura basata su LDPC, in sostituzione su quella basata su codici RS, inventata quasi 50 anni fa. E' vero: l'articolo di Reed-Solomon è esattamente di 50 anni fa, del 1960, ma quello di Gallager, che descrive i codici LDPC, è del 1962, quarantotto anni fa.

Ai risultati di oggi hanno contribuito tutti i protagonisti della nostra storia. Il limite di Shannon è raggiunto da codici con elevata lunghezza blocco, come preconizzava Elias nel 1955, massimizzando la distanza di Hamming fra le parole di codice, utilizzando la soft-decision, adottando schemi che consentano di effettuare la decodifica in parallelo. Ma soprattutto il limite è raggiunto dagli schemi attuali perché oggi la densità di integrazione e la velocità di calcolo consentono di integrare memoria necessaria e algoritmo in una piccola superficie di silicio: in definitiva di realizzare un SoC.

La nostra storia è dunque una storia a lieto fine, il limite indicato nel 1948 è raggiunto, dopo circa mezzo secolo. Eppure, come è tipico nell'ambito scientifico, l'obiettivo non è il raggiungimento del limite, ma è il suo superamento, lo scoprire ciò che ci attende, oltre l'orizzonte.

Quale sarà la continuazione di questa storia, il progresso dovuto ai codici per la protezione contro gli errori, nel campo della memorizzazione, delle esplorazioni spaziali e delle comunicazioni sulla terra?

Bibliografia

1. *M. Barbero, N. Shpuza:* "Rivelazione, Correzione e Mascheramento degli Errori - Parte I", Elettronica e Telecomunicazioni, Aprile 2008.

2. *M. Barbero, N. Shpuza*: "Prossimi al limite di Shannon, 60 anni dopo", Elettronica e Telecomunicazioni, Agosto 2008.

3. *V. Mignone, A. Morello:* "Il sistema DVB-S2 di seconda generazione per la trasmissione via satellite e Unicast", Elettronica e Telecomunicazioni, Dicembre 2003.

4. *V. Mignone, A. Morello, G. Russo, P. Talone:* "DVB-T2 - la nuova piattaforma per la televisione digitale terrestre", Elettronica e Telecomunicazioni, Dicembre 2008.

5. *A. Shokrollahi*: "Raptor Codes",IEEE Trans. on Information Theory, vol. 52, No.6, June 2006

6. A. *Bertella, P. Casagranda, D. Milanesio e M. Tabone*: "Il sistema DVB-H per la TV Mobile", Elettronica e Telecomunicazioni, Dicembre 2005.

Gli articoli legati alla evoluzione degli algoritmi di protezione dagli errori, a partire da quello di Shannon, sono riportati nella bibliografia di [2].

Le specifiche dei vari sistemi DVB possono essere reperite in:
www.dvb.org/technology/standards/index.xml