



# Università degli Studi di Bergamo

---

**DIPARTIMENTO DI INGEGNERIA E SCIENZE APPLICATE**



# RETI INTERNET MULTIMEDIALI

---

**Compressive Video:  
MPEG**

*Il documento è adattato da materiale cortesemente messo a disposizione dal Prof. Stefano Paris e dal Prof. Vittorio Trecordi*

# Introduzione

---

- Il gruppo di esperti MPEG (Motion Picture Experts Group) si occupa della definizione dei seguenti standard
  - MPEG-1, MPEG-2, MPEG-4: sistemi e tecniche di codifica audio/video
  - MPEG-7: sistemi di ricerca di contenuti audio/video in database multimediali e descrittori base (forme, colori, etc.)
  - MPEG-21: sistemi per lo scambio di contenuti multimediali attraverso diverse tecnologie di rete/dispositivi e gestione della protezione dei diritti

# Introduzione

---

- MPEG definisce un sistema di codifica asimmetrico
  - Il codificatore richiede molte più risorse computazionali del decodificatore
  - Il decodificatore è ben specificato dallo standard, mentre i dettagli implementativi del codificatore no
  - Sviluppato inizialmente per applicazioni non real-time (es. video storage o video streaming)
- Gli standard definiti in MPEG sono organizzati in tre parti principali
  - Sistema: multiplazione di video, audio, e altri dati per lo storage e la trasmissione (definizione del *bitstream*)
  - Codifica video
  - Codifica audio

# MPEG

---

## Codifica Video

# Introduzione

- I diversi standard di *codifica video* definiti da MPEG si rivolgono a diverse tipologie di applicazioni
  - Esigenze di risorse (banda, complessità, ecc.)
  - Qualità del segnale video riprodotto

Standard	MPEG-1	MPEG-2
Risoluzione (Y)	352 x 240	720 x 480
Frame rate	30 fps	30 fps
Bit rate	1.5 Mbps	15 Mbps
Applicazioni	Video Storage, Broadcast (VCR quality)	Digital Broadcast TV, HDTV

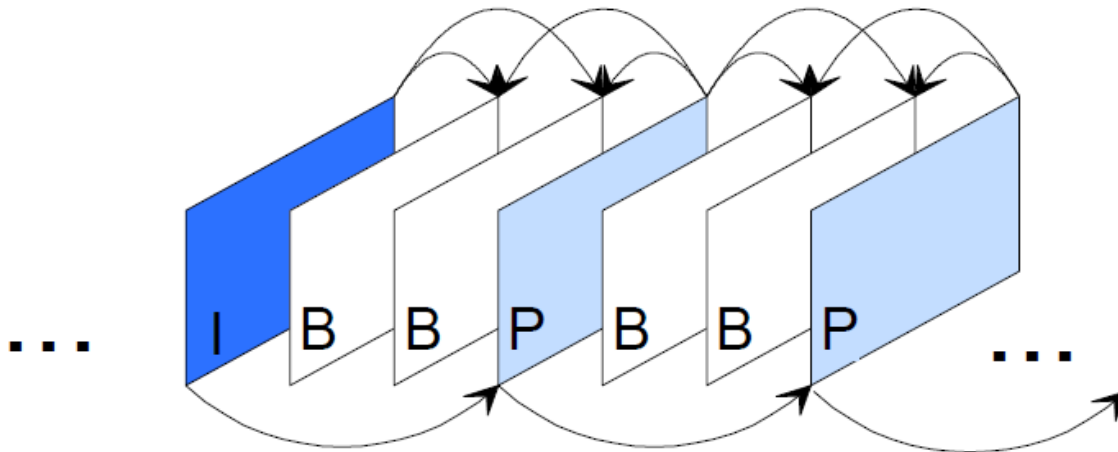
# MPEG - Group of Pictures

---

- Il flusso video viene codificato come una *sequenza di frame* (o *picture*)
- Esistono tre tipologie di frame
  - I-frame: Intra-Coded Pictures
  - P-frame: Predictive-Coded Pictures (forward prediction)
  - B-frame: Bi-Directional Predictive Coded Pictures (bi-directional prediction)
- L'insieme dei frame viene suddiviso in sottoinsiemi detti Group of Pictures (GOP)
  - Ogni GOP è composta da una I-frame, e diverse P-frame e B-frame

# MPEG - Group of Pictures

- Le B-frame sono computazionalmente onerose ma sono molto vantaggiose in termini di compressione
- Le I-frame sono utili nel caso di cambi di scena



Type	Size	Compress
I	18K	7:1
P	6K	20:1
B	2.5K	50:1
Avg	4.8K	27:1

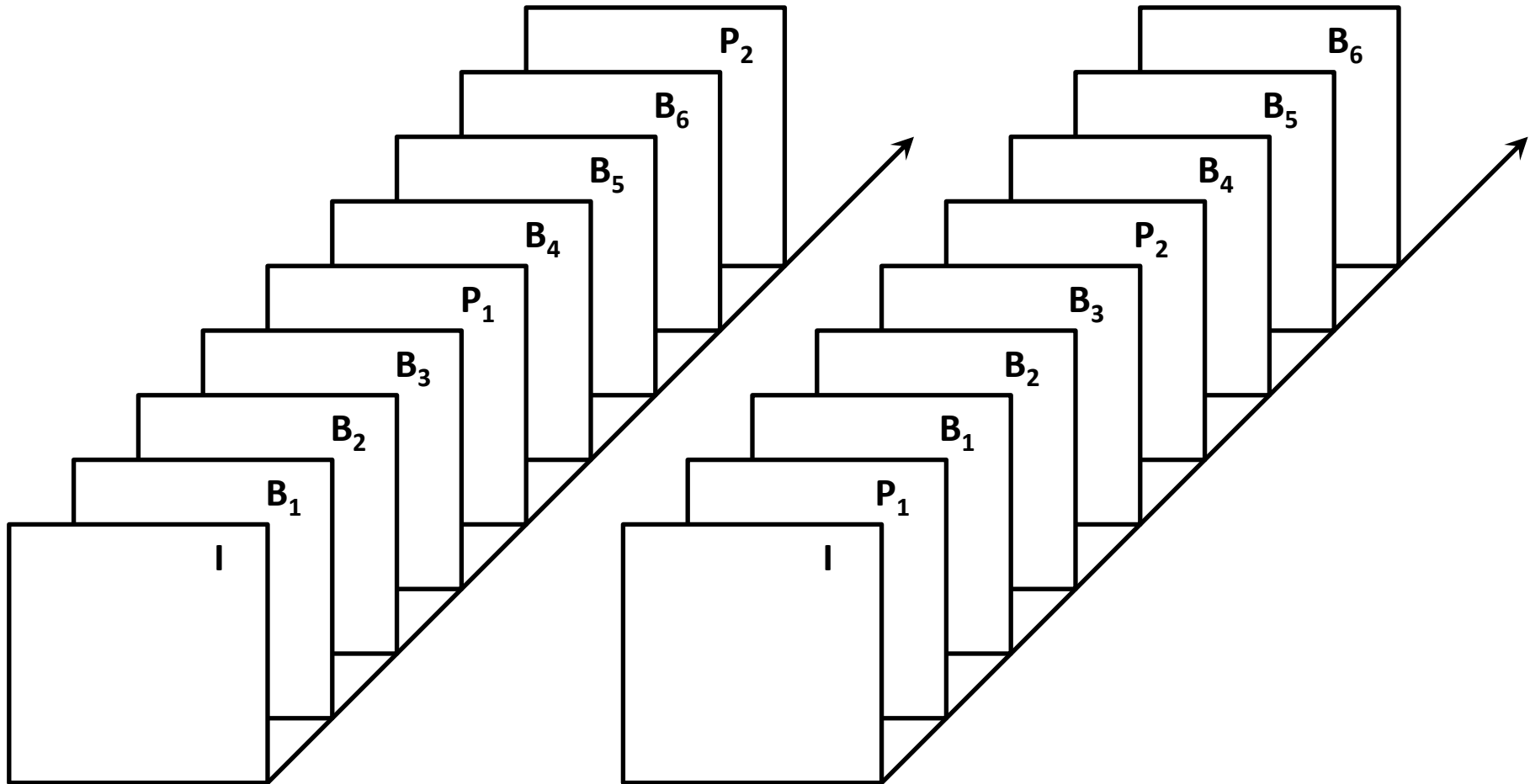
# MPEG - Group of Pictures

---

- A causa delle tecniche di predizione utilizzate, al ricevitore la sequenza di frame decodificati risulterà diversa rispetto a quella di visualizzazione
  - Un P-frame dipende da un I-frame precedente
  - Un B-frame dipende sia da un frame precedente sia da un frame successivo (siano essi I-frame o P-frame)
- E' necessario riordinare la sequenza di frame prima di visualizzarli



# MPEG - Group of Pictures



Ordine di visualizzazione

Ordine di trasmissione/ricezione

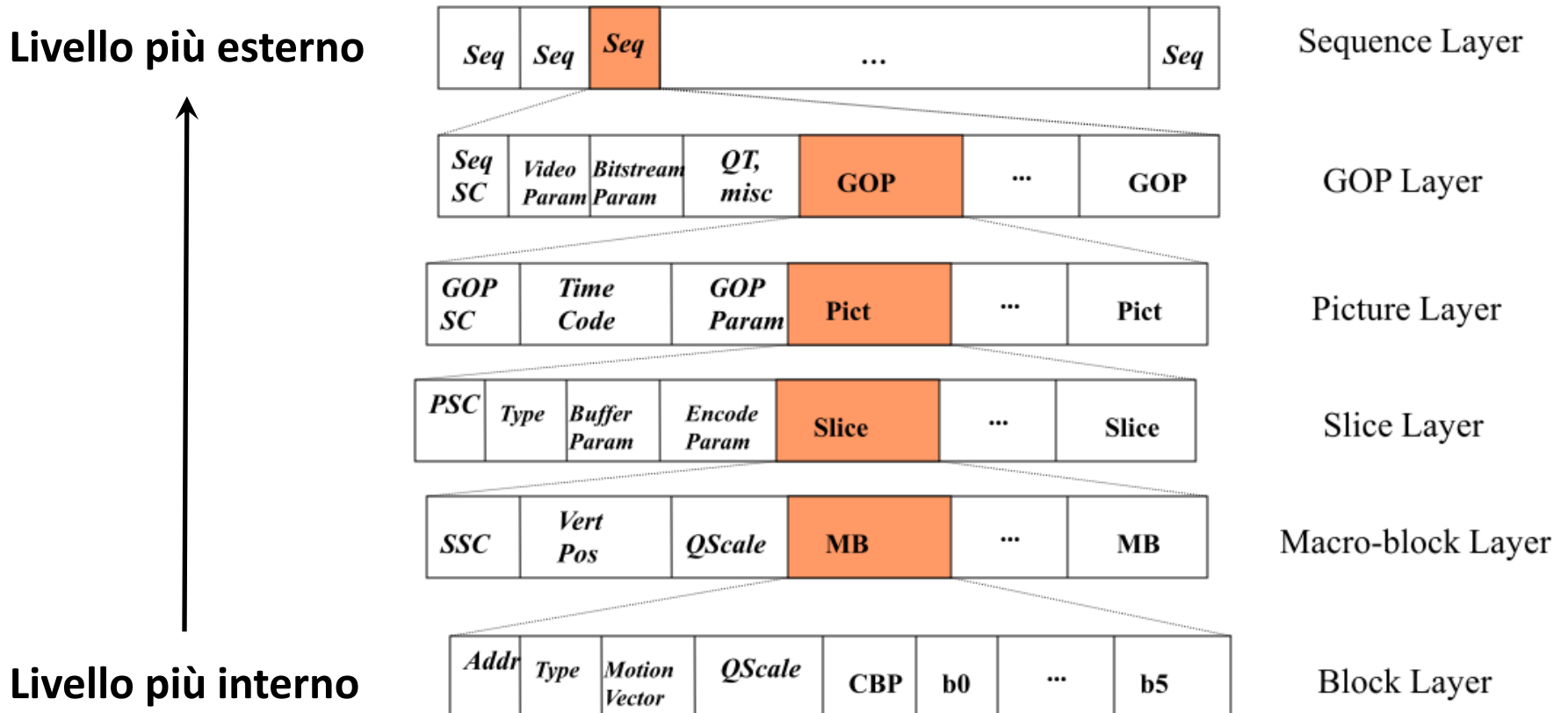
# MPEG – Formato Video

---

- Il flusso video è codificato in una struttura formata da diversi livelli gerarchici
  1. Blocco 8x8 (sia per luminanza Y che per crominanza U e V)
  2. Macro-blocco (sotto-campionamento 4:2:0)
    - 4 blocchi luminanza Y (una matrice 16x16)
    - 2 blocchi crominanza U e V (uno U e uno V)
  3. Slice: sequenza contigua di macro-blocchi (differenti slice sono codificate ognuna separatamente dall'altra)
  4. Picture (I, P o B)
  5. Group of Picture (GOP)
  6. Sequenza

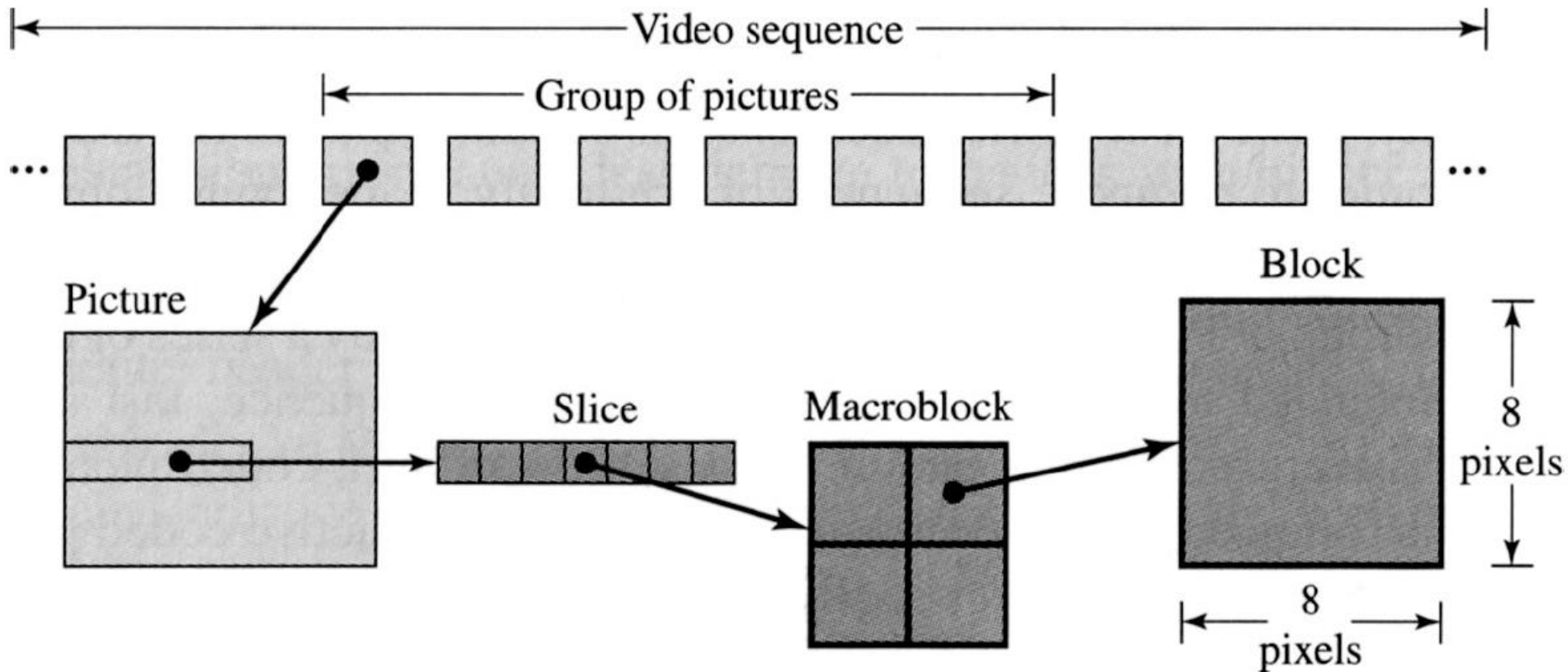
# MPEG – Formato Video

## ■ MPEG bitstream: struttura gerarchica



# MPEG – Formato Video

## ■ MPEG bitstream: struttura gerarchica



# Codifica Intra-frame (I-frame)

---

- Un frame viene codificato utilizzando solo le informazioni contenute nel frame stesso
  - Utilizzato per i *I-frame*
- Per ognuno dei blocchi 8x8 di luminanza e crominanza si effettua
  1. *DCT*: coefficienti trasformati con DCT
  2. *Quantizzazione*: output della DCT quantizzato
  3. *Entropy coding*: coefficienti quantizzati vengono codificati utilizzando uno schema lossless
- Il processo è analogo a quello eseguito da JPEG sequenziale

# Codifica Intra-frame (I-frame)

## ■ Codifica dei coefficienti DC di un I-frame

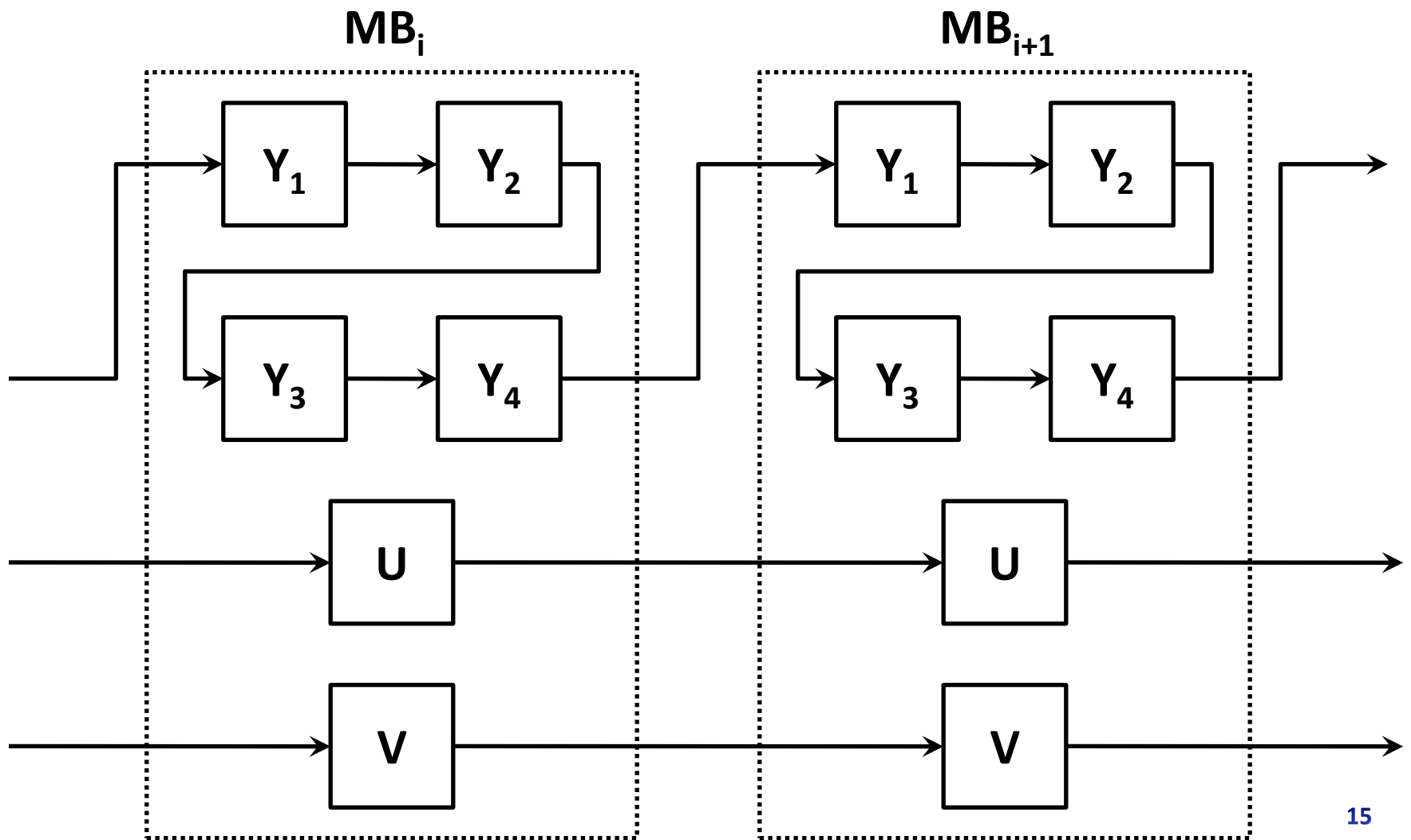
- La ridondanza dei coefficienti DC dei blocchi 8x8 adiacenti è ridotta utilizzando una codifica DPCM (Differential Pulse Coded Modulation)

$$\Delta DC_k^i = DC_k^i - P_k^i$$

$P_k^i$  → Predittore del valore DC del blocco  $i$ -esimo del macro-blocco  $k$ -esimo

- Il segnale risultante dalla differenza è codificato utilizzando una compressione lossless, definita dallo standard
  - Il predittore è definito utilizzando l'informazione relativa al coefficiente DC del blocco precedente
- ## ■ L'ordine di predizione dei coefficienti DC dipende dall'ordine di codifica dei blocchi 8x8

# Codifica Intra-frame (I-frame)



# Codifica Intra-frame (I-frame)

---

- Codifica dei coefficienti AC di un I-frame
  - Molti dei coefficienti AC ad alta frequenza sono nulli
  - E' più efficiente codificare la posizione e il valore dei (pochi) coefficienti non nulli
- La sequenza (secondo uno scan a zig-zag) di coefficienti AC quantizzati viene codificata utilizzando una sequenza di simboli *<Run:Level>*
  - *Run*: indica il numero di elementi nulli che precedono un elemento non nullo
  - *Level*: indica il valore dell'elemento non nullo (successivo alla sequenza di elementi nulli)
- Ogni simbolo *<Run:Level>* viene codificato con un codice lossless standardizzato



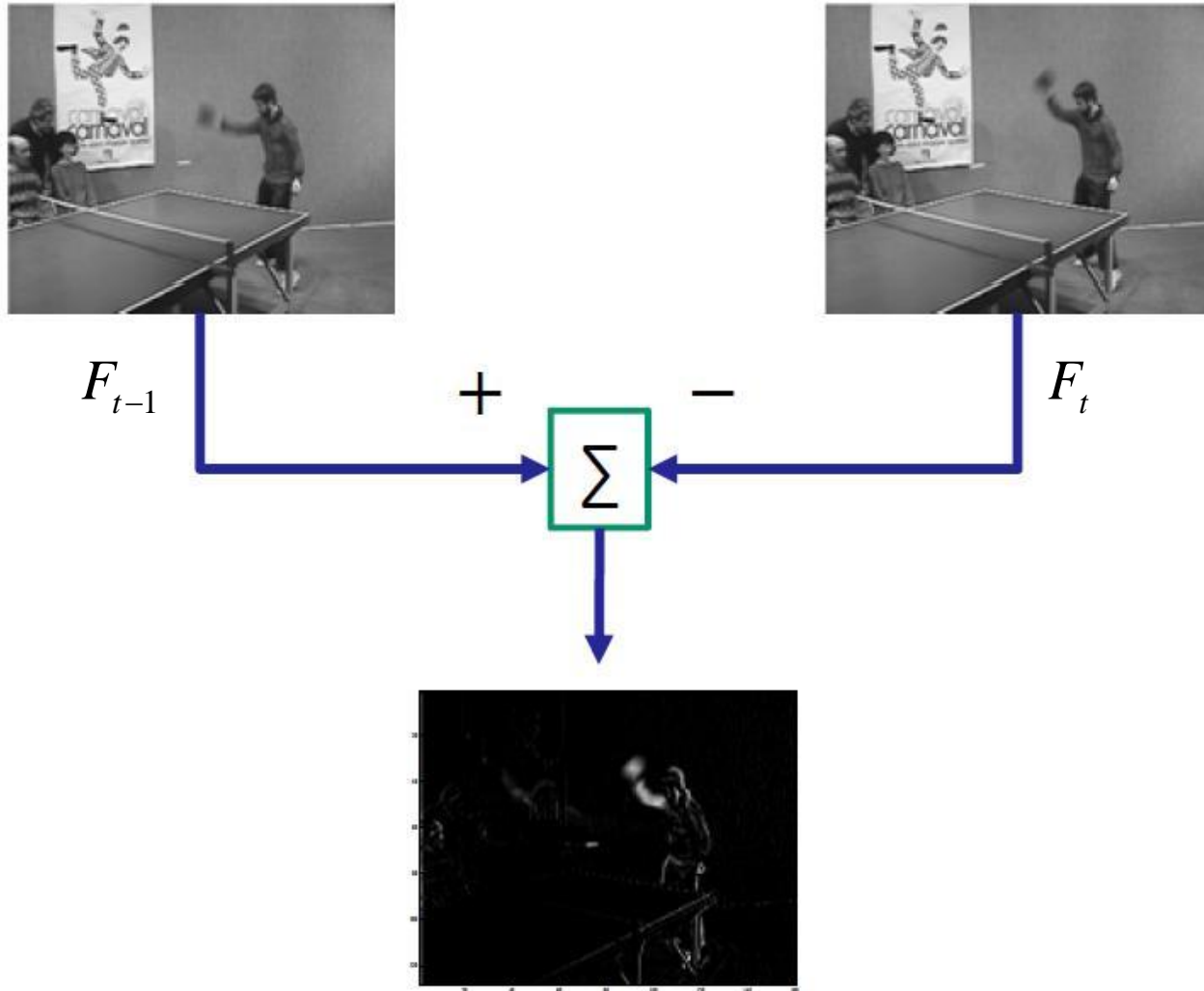


# Codifica Inter-frame (P-frame)

---

- In generale due frame consecutivi sono molto simili tranne per alcuni dettagli dovuti
  - Movimento (della videocamera e degli oggetti)
  - Variazione della luce (movimento delle ombre)
- Se si codifica il frame risultante dalle differenze tra il frame corrente e il frame precedente si aumenta la compressione
  - Poca variazione nel frame risultante (pochi coefficienti DCT)

# Codifica Inter-frame (P-frame)

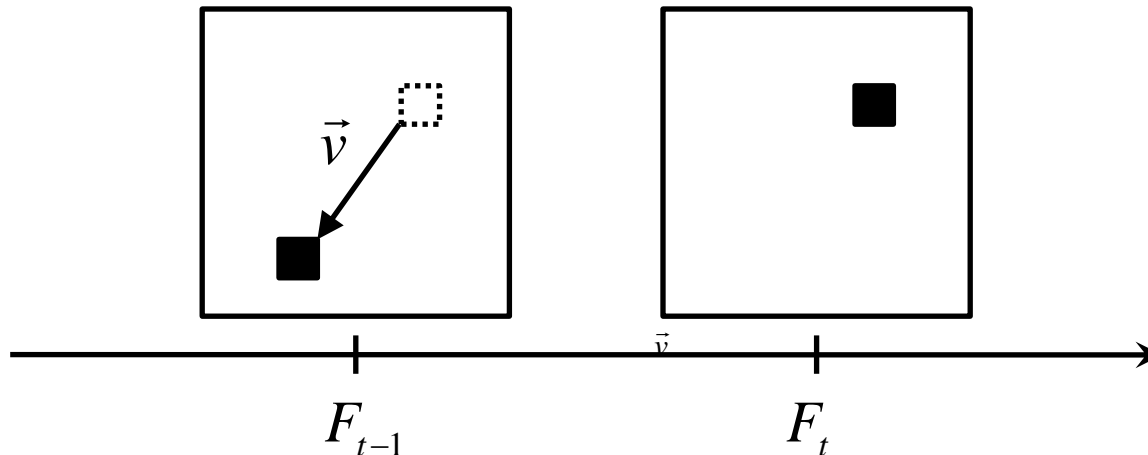


# Codifica Inter-frame (P-frame)

---

- Per migliorare ulteriormente la codifica si usa la tecnica di *block-based motion compensated prediction error*
- E' necessario eseguire la stima del movimento dei macro-blocchi tra due frame consecutivi (tecnica di *motion estimation*)
- Prima di eseguire la differenza si genera un nuovo frame utilizzando il frame di riferimento e l'informazione ottenuta dall'algoritmo di motion estimation (vettore di movimento)
  - Tecnica di *motion compensation*
  - Errore di predizione: differenza tra il frame corrente e la versione ottenuta dalla predizione (motion compensated frame)

# Codifica Inter-frame (P-frame)



**Codifica**

$$F_t^M = mc(\tilde{F}_{t-1}, \vec{v})$$

$$D_t = F_t - F_t^M$$

$$P_t = enc(D_t) \left. \vphantom{enc(D_t)} \right\} \text{Invio}$$

$$\vec{v}' = enc(\vec{v}) \left. \vphantom{enc(\vec{v})} \right\} \text{Invio}$$

**Decodifica**

$$\vec{v} = dec(\vec{v}') \left. \vphantom{dec(\vec{v}')} \right\} \text{Ricezione}$$

$$D_t = dec(P_t) \left. \vphantom{dec(P_t)} \right\} \text{Ricezione}$$

$$F_t^M = mc(\tilde{F}_{t-1}, \vec{v})$$

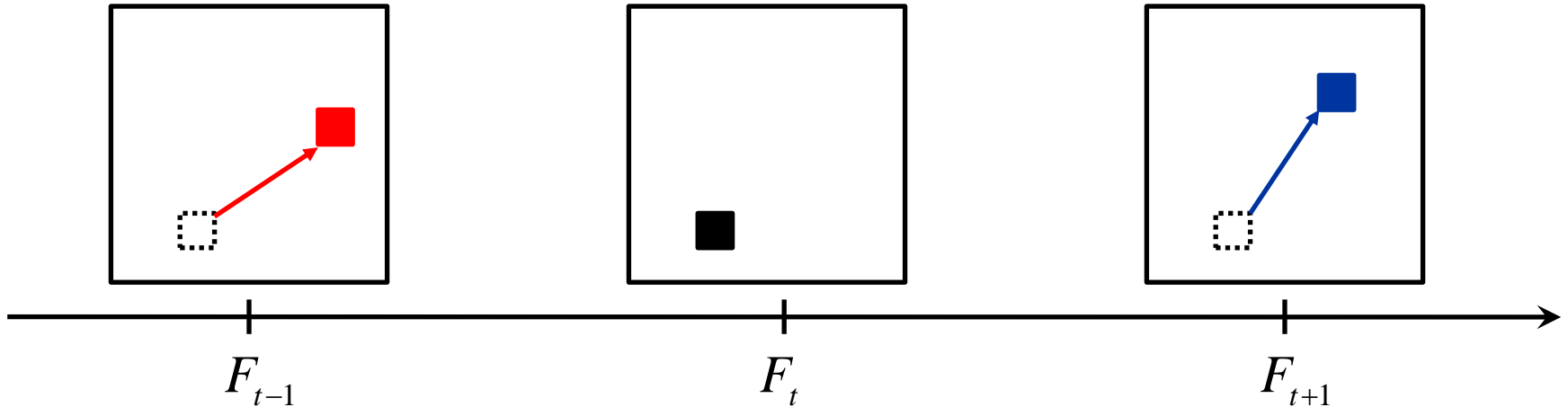
$$F_t = F_t^M + D_t$$

# Codifica Inter-frame (B-frame)

---

- La predizione bi-direzionale utilizza due frame di riferimento
- Per ogni macro-blocco del frame predetto vengono generati due vettori di movimento
  - Un vettore relativo a un frame nel passato
  - Un vettore relativo a un frame nel futuro
- La predizione bidirezionale permette di ottenere un miglior rapporto di compressione
- I B-frame non vengono utilizzati come frame di riferimento per la predizione di altri frame

# Codifica Inter-frame (B-frame)



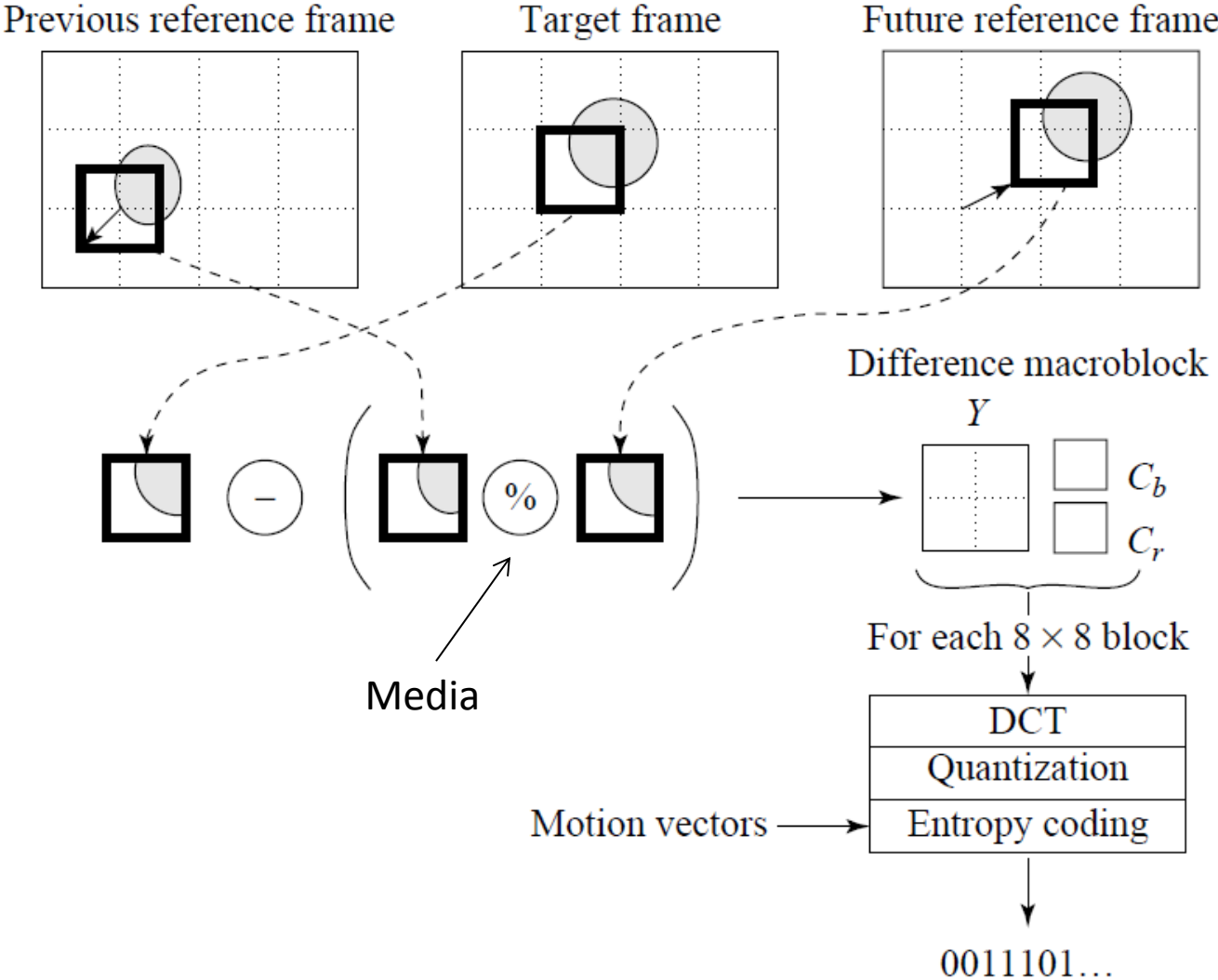
$$\vec{v}_t = \alpha_1 v_{t-1} + \alpha_2 v_{t+1}$$

$\alpha_1 = 0.5 \wedge \alpha_2 = 0.5 \longrightarrow$  **Bi-directional Prediction**

$\alpha_1 = 1 \wedge \alpha_2 = 0 \longrightarrow$  **Forward Prediction**

$\alpha_1 = 0 \wedge \alpha_2 = 1 \longrightarrow$  **Backward Prediction**

# Codifica Inter-frame (B-frame)





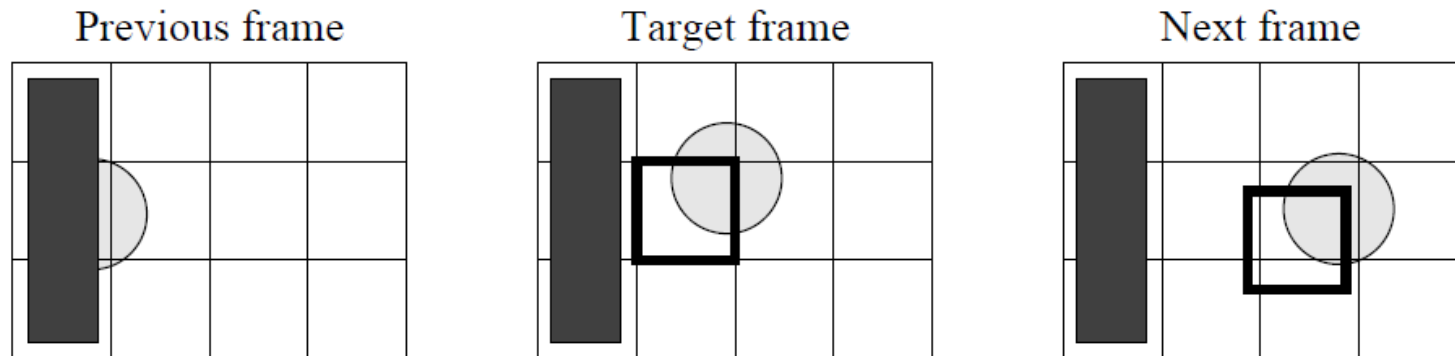
# Codifica Inter-frame (B-frame)

---

- Non tutti i macro-blocchi dei B-frame derivano da predizione bi-direzionale
- I macro-blocchi che compongono i B-frame sono codificati utilizzando predizione
  - Bidirezionale
  - Backward
  - Forward
  - Nessuna (Intra-coded)
- Viene utilizzata la tipologia di predizione che minimizza l'errore di predizione della componente di luminanza

# Codifica Inter-frame (B-frame)

- In una situazione come quella in figura, ad esempio, verrà utilizzata solamente *backward prediction*
  - L'algoritmo di motion compensation non trova un buon match nella frame precedente ma lo trova nella frame successiva



# Codifica Inter-frame

---

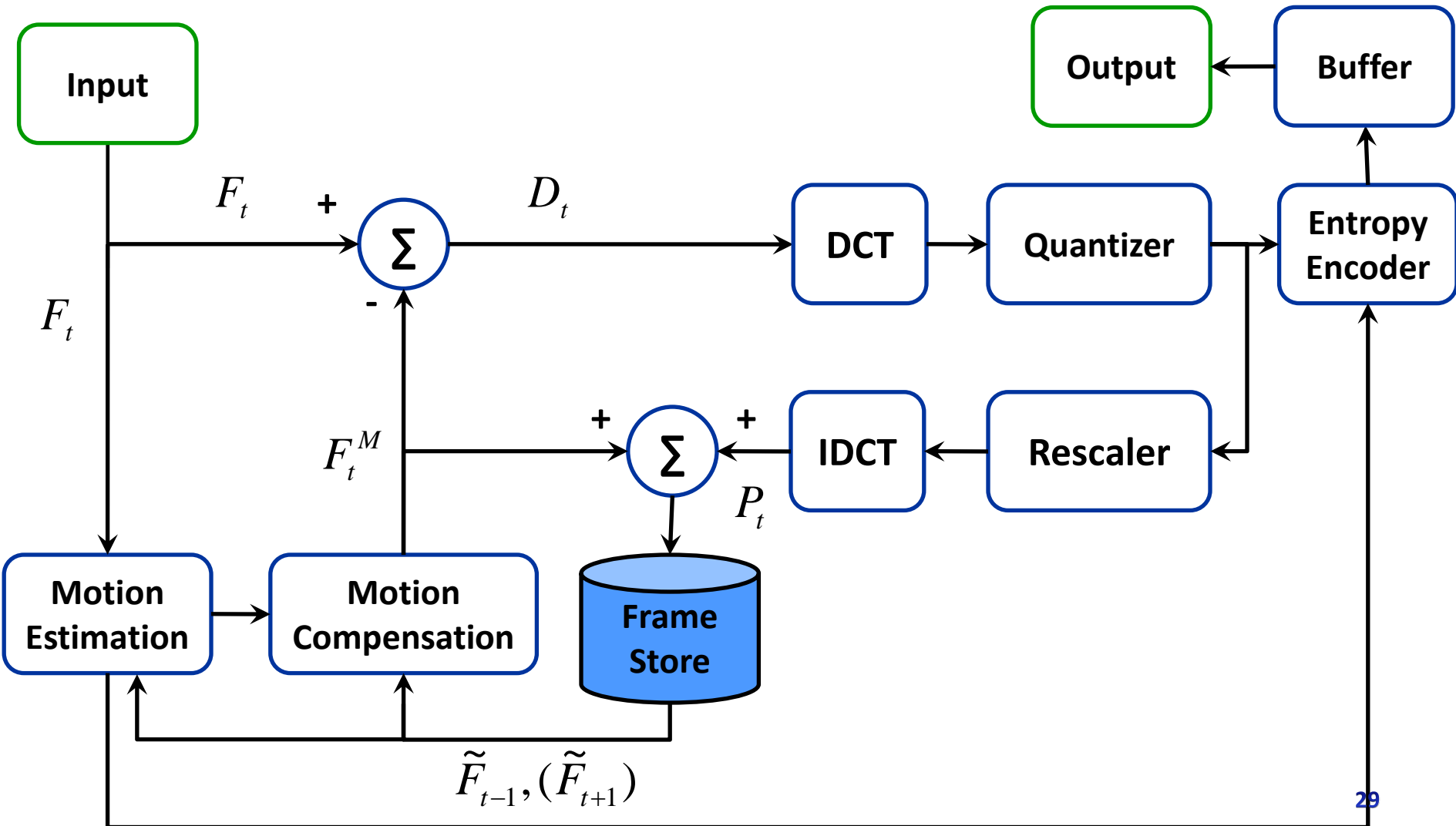
- L'informazione necessaria alla ricostruzione del macro-blocco generato dalla codifica Inter-frame (P o B) è formata da
  - Vettore di movimento
  - Errore di predizione
- La trasformazione DCT è applicata all'errore di predizione, prima della quantizzazione e della codifica lossless

# MPEG - Codifica

---

- Il codificatore MPEG è composto dai seguenti componenti
  - Controller: controllo e sincronizzazione del flusso video
  - Modulo DCT
  - Quantizzatore
  - Entropy encoder
  - Reconstruction Module (Rescaler + Modulo IDCT)
  - Motion Estimator (predizione del movimento)
  - Motion Compensator (compensazione del movimento)
- Il Reconstruction Module e il Motion Compensation sono necessari per generare i frame di riferimento con cui eseguire la predizione del movimento

# MPEG - Codificatore

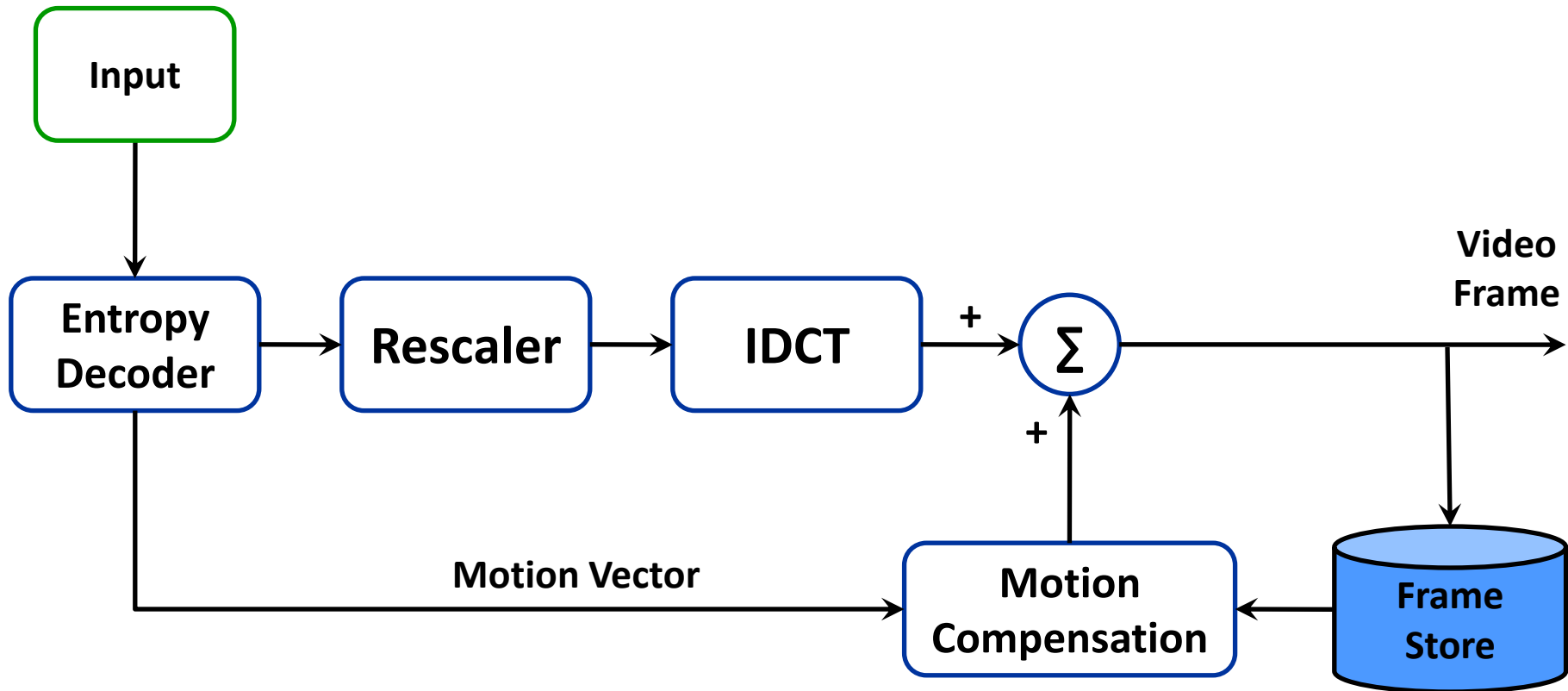


# MPEG - Decodifica

---

- Il decodificatore MPEG è più semplice del codificatore (codifica asimmetrica)
  - La ricostruzione è guidata dalle informazioni inviate dal codificatore (coefficienti della DCT e motion vector)
- E' composto essenzialmente da tre componenti
  - Entropy decoder
  - Reconstruction Module
  - Motion Compensation

# MPEG - Decodificatore



# MPEG

---

## IDCT Mismatch



# IDCT Mismatch

---

- L'efficienza di compressione dipende generalmente dalla codifica inter-frame
- Nella codifica inter-frame l'informazione relativa a un frame predetto (P o B) si somma all'informazione decodificata da precedenti frame
- Minime variazioni prodotte dalle IDCT del codificatore e del decodificatore si propagano fino a un I-frame (inizio del GOP successivo)
  - Se non controllate diventano percettibili come distorsione del segnale
- La differenza prodotta dalle IDCT del codificatore e del decodificatore è nota come *IDCT mismatch error*

# IDCT Mismatch

---

- L'errore di IDCT mismatch non è da confondere con l'errore dovuto alla quantizzazione dei coefficienti della DCT
  - Entrambe le IDCT operano sugli stessi coefficienti quantizzati
- Lo standard non definisce l'algoritmo per il calcolo della IDCT
  - Indica solo delle tolleranze espresse come errori
- La precisione finita con cui può essere rappresentata l'output della IDCT contribuisce all'errore di mismatch

# IDCT Mismatch

---

- Lo standard definisce una procedura per la misura dell'accuratezza della IDCT
- Un set di valori di riferimento (reference data set) composto da 10000 blocchi 8x8 viene fornito per il calcolo dell'accuratezza
- La propria implementazione della IDCT deve essere eseguita sullo stesso data set
- Due insiemi di valori
  - Reference pixel  $\rightarrow f_k(y, x)$
  - Reconstructed pixel  $\rightarrow \hat{f}_k(y, x)$

# IDCT Mismatch

---

- L'errore di misura tra i dati di riferimento e quelli calcolati dalla propria IDCT deve rispettare certi requisiti

$$e_k(y, x) = \hat{f}_k(y, x) - f_k(y, x)$$

- Requisiti:

Peak error ←  $ppe_k(y, x) = \max \{ |e_k(y, x)| \} \leq 1 \quad \forall k = 1, 2, \dots, 10000$

Peak mean square error ←  $pmse(y, x) = \frac{\sum_{k=1}^{10000} e_k^2(y, x)}{10000} \leq 0.06$

# IDCT Mismatch

## ■ Requisites:

Overall mean square error ← 
$$omse = \frac{\sum_{x=0}^7 \sum_{y=0}^7 \sum_{k=1}^{10000} e_k^2(y, x)}{640000} \leq 0.02$$

Peak mean error ← 
$$pme(y, x) = \frac{\sum_{k=1}^{10000} e_k(y, x)}{10000} \leq 0.015$$

Overall mean error ← 
$$ome = \frac{\sum_{x=0}^7 \sum_{y=0}^7 \sum_{k=1}^{10000} e_k(y, x)}{640000} \leq 0.0015$$

# IDCT Mismatch

---

- Per limitare la propagazione degli errori di IDCT Mismatch, deve essere eseguita una codifica intra-frame (I-Frame) ogni 132 P-frame
  - Processo noto come *force updating*
- Nella pratica il *force updating* è raro, poiché gli I-frame appaiono con una frequenza di 1 ogni 10-20 frames
  - Semplifica le procedure di fast forward e fast reverse playback, semi-random access

# MPEG

---

## Motion Estimation

# Motion Estimation

---

- A differenza dell'algoritmo di Motion Compensation, lo standard non definisce un algoritmo di Motion Estimation
  - Motion Estimation: stima del vettore di movimento
  - Motion Compensation: utilizzo del vettore di movimento per predire la trasformazione di un frame
- L'algoritmo di Motion Estimation è il componente più importante nella codifica MPEG
  - Dalla sua implementazione dipendono il rapporto di compressione e la qualità della codifica



# Motion Estimation

---

- L'algoritmo di motion estimation determina un vettore di movimento per ogni macro-blocco in cui è suddiviso il frame
- Assunzione: i vettori di movimento sono costanti in un macro-blocco
  - Un unico vettore di movimento è assegnato al macro-blocco composto dai 4 blocchi 8x8 di luminanza e i 2 blocchi 8x8 di crominanza
- La ridondanza informativa tra vettori di movimento appartenenti a macro-blocchi adiacenti è ridotta codificandone la differenza (codifica predittiva)

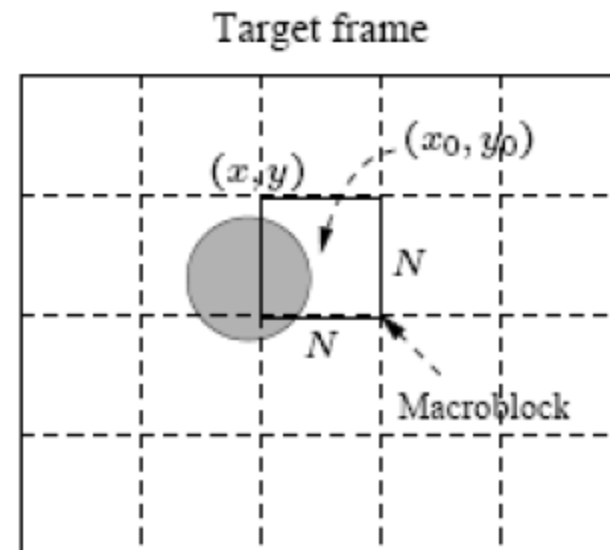
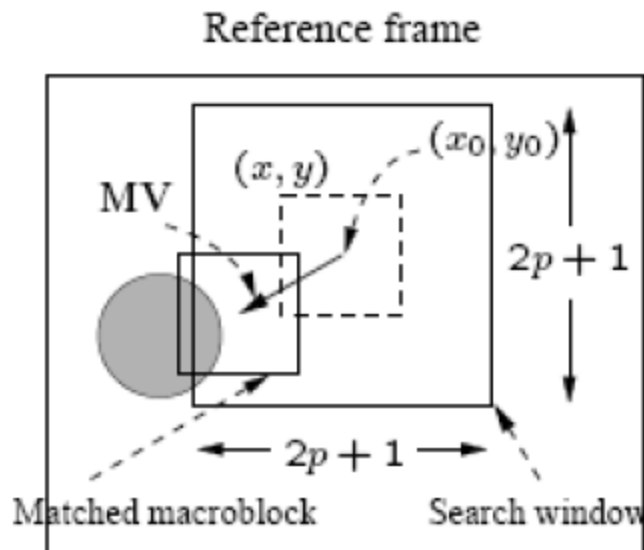
# Motion Estimation

---

- Algoritmo di motion estimation
  - Suddividi il frame corrente in macro-blocchi
  - Stima un vettore di movimento per ogni macro-blocco utilizzando un frame di riferimento
  - All'interno di una finestra di ricerca del frame di riferimento (area  $W^2$ ), cerca il macro-blocco che minimizza una misura dell'errore (*best match*)
  - Genera il vettore di movimento, definito come offset in termini di coordinate tra il blocco da codificare nel frame corrente e il macro-blocco best match nel frame di riferimento
- Per minimizzare la complessità computazionale si utilizzano delle strategie di ricerca efficienti

# Motion Estimation

- La ricerca si effettua nelle vicinanze (spostamento orizzontale e verticale  $[-p, +p]$  pixel, considerando il centro del macro-blocco)
- La dimensione della finestra di ricerca è quindi  $W^2 = (2p + 1)^2$



# Criterio di Matching

- Esistono diverse misure dell'errore che si vuole minimizzare
  - Mean Absolute Error (MAE)
  - Sum of Absolute Error (SAE)
  - Mean Squared Error (MSE)

$$MAE(x, y) = \frac{1}{N^2} \sum_{i=0}^N \sum_{j=0}^N \left| F_t(i, j) - \tilde{F}_{t-1}(i+x, j+y) \right|$$

$$SAE(x, y) = \sum_{i=0}^N \sum_{j=0}^N \left| F_t(i, j) - \tilde{F}_{t-1}(i+x, j+y) \right|$$

NB:  $N = 16$

$$MSE(x, y) = \frac{1}{N^2} \sum_{i=0}^N \sum_{j=0}^N \left( F_t(i, j) - \tilde{F}_{t-1}(i+x, j+y) \right)^2$$

# Criterio di Matching

- Generalmente si utilizza  $MAE(x, y)$  come misura dell'errore
  - $x$  e  $y$  sono riferiti allo spostamento orizzontale e verticale tra il macro-blocco da codificare e il macro-blocco attualmente considerato del frame di riferimento
  - $i$  e  $j$  indicano gli indici dei pixel corrispondenti tra i due macro-blocchi

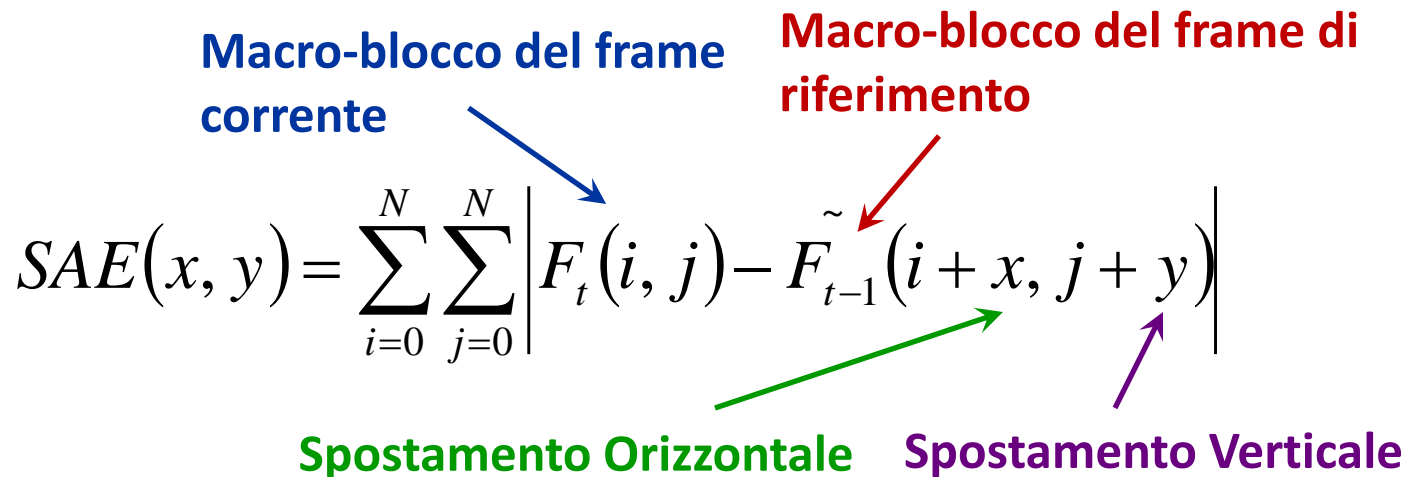
Macro-blocco del frame corrente

Macro-blocco del frame di riferimento

$$SAE(x, y) = \sum_{i=0}^N \sum_{j=0}^N \left| F_t(i, j) - \tilde{F}_{t-1}(i+x, j+y) \right|$$

Spostamento Orizzontale

Spostamento Verticale



# Strategie di Ricerca

---

- Strategie di ricerca del vettore che minimizza l'errore
  - Full Search
  - 2-D Logarithmic Search
  - Three-Step Search
  - Hierarchical Search

# Strategia Full Search

## ■ Strategia Full Search

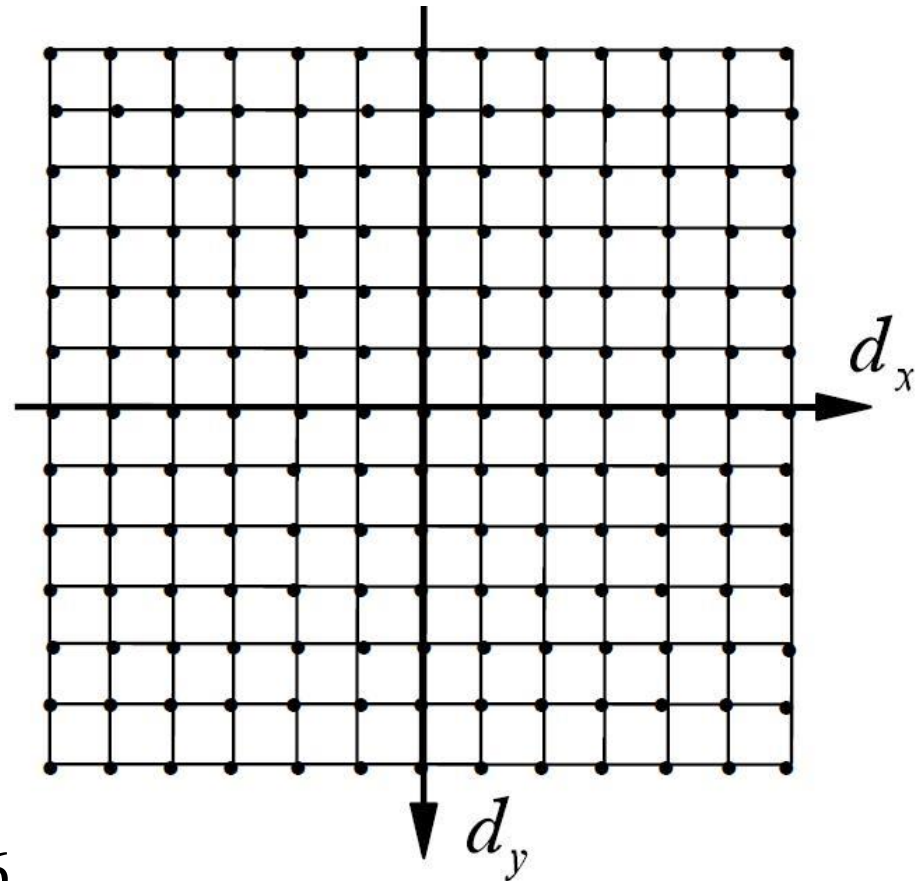
- Prova tutte le combinazioni ottenute variando  $x$  e  $y$  all'interno dell'area di ricerca  $W^2$
- Seleziona la combinazione che minimizza la misura dell'errore scelta
- Trova sempre la soluzione ottima all'interno della finestra

## ■ Complessità

- Assunzione usando MAE: 3 operazioni per ogni confronto (sottrazione, valore assoluto, addizione)



$$W^2 \cdot N^2 \cdot 3 \rightarrow O(W^2 \cdot N^2) \quad N = 16$$



# Strategia 2-D Logarithmic Search

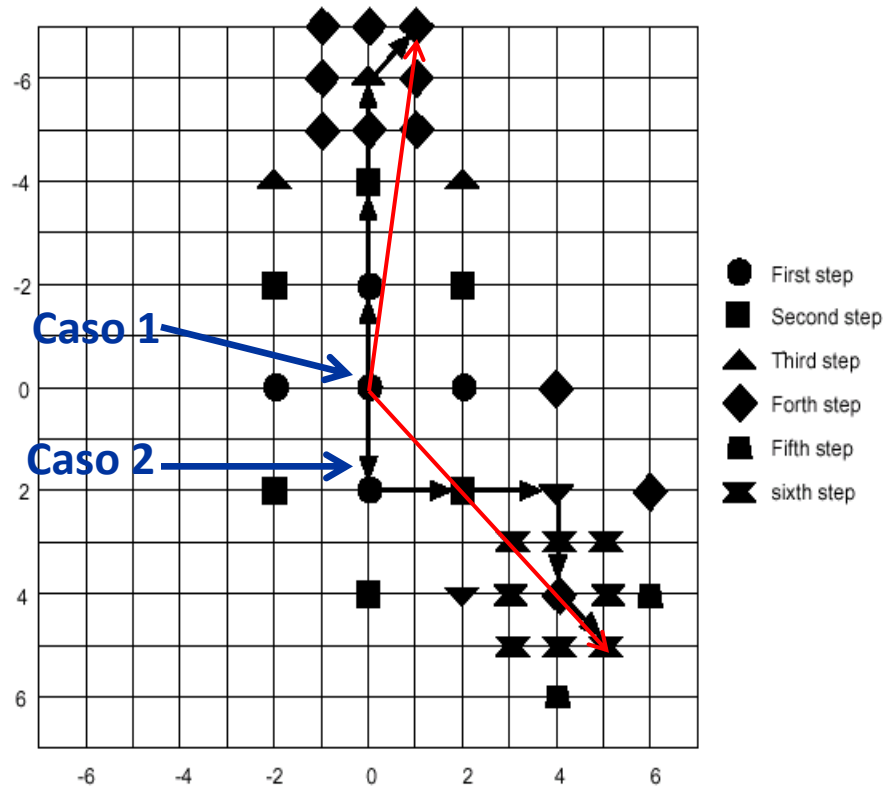
---

- Strategia 2-D Logarithmic Search
  - Trova una soluzione sub-ottima, ma richiede meno risorse computazionali della Full Search
  - Versione 2-D della ricerca binaria (o dicotomica)
  - Il confronto della misura dell'errore (solitamente MAE) è effettuata solo tra i macro-blocchi di una area che contiene 5 punti vicini
  - L'area è spostata iterativamente, centrandola nel punto che minimizza la misura dell'errore



# Strategia 2-D Logarithmic Search

- La misura MAE è calcolata solo per i 5 punti all'interno dell'area quadrata disposta in diagonale (con offset di 2 pixel)
- L'area viene spostata ad ogni iterazione e centrata nel punto che minimizza la misura MAE
- Il lato dell'area viene dimezzato quando
  1. Il punto selezionato si trova sul bordo dell'area di ricerca
  2. Il punto selezionato si trova nel mezzo dell'area corrente
- Quando l'area è dimezzata, la ricerca viene effettuata su tutti gli 8 punti adiacenti al centro



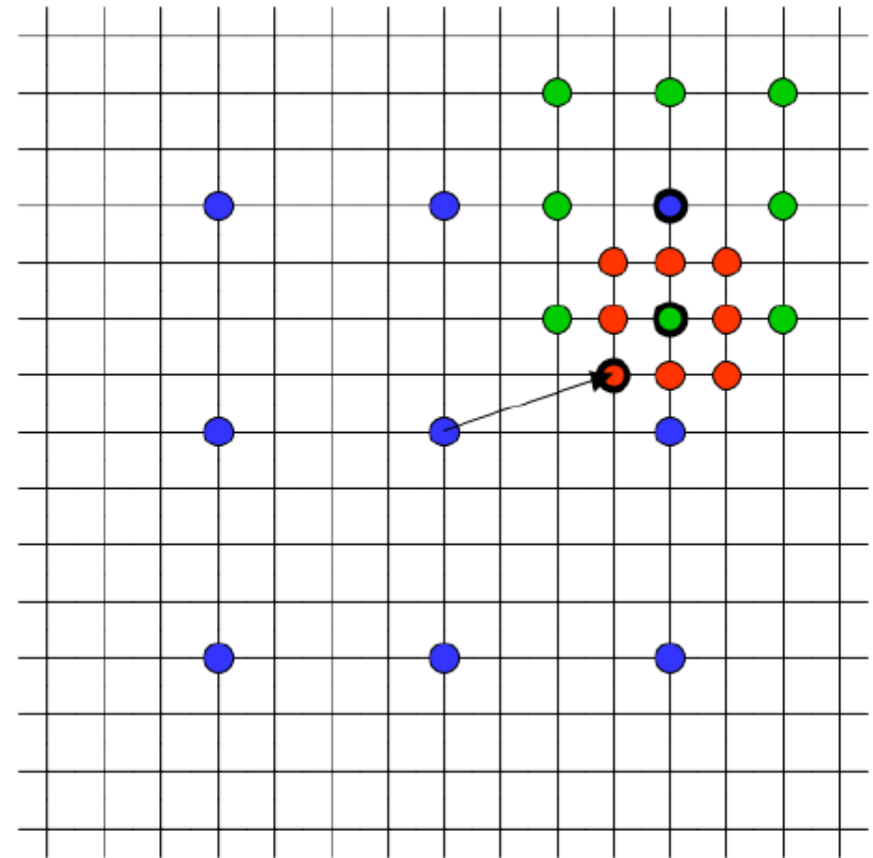
# Strategia Three-Step Search

---

- Strategia Three-Step Search
  - Richiede meno risorse computazionali della Full Search e della 2-D Logarithmic Search
  - Il numero di passi dell'algoritmo, a differenza della 2-D Logarithmic Search, è noto a priori ed è pari a 3
  - In genere, porta a peggiori rapporti di compressione rispetto a Full Search e 2-D Logarithmic Search

# Strategia Three-Step Search

- Tre Step
  - Step 1: -4,+4 pixel
  - Step 2: -2,+2 pixel
  - Step 3: -1,+1 pixel
- Si trova il miglior match ad ogni step e ci si sposta centrando il nuovo quadrato di indagine in tale punto, dimezzando la dimensione del quadrato
- Allo Step 3 si seleziona il punto tra i 9 rimasti che offre il miglior match e si è trovato il vettore di movimento

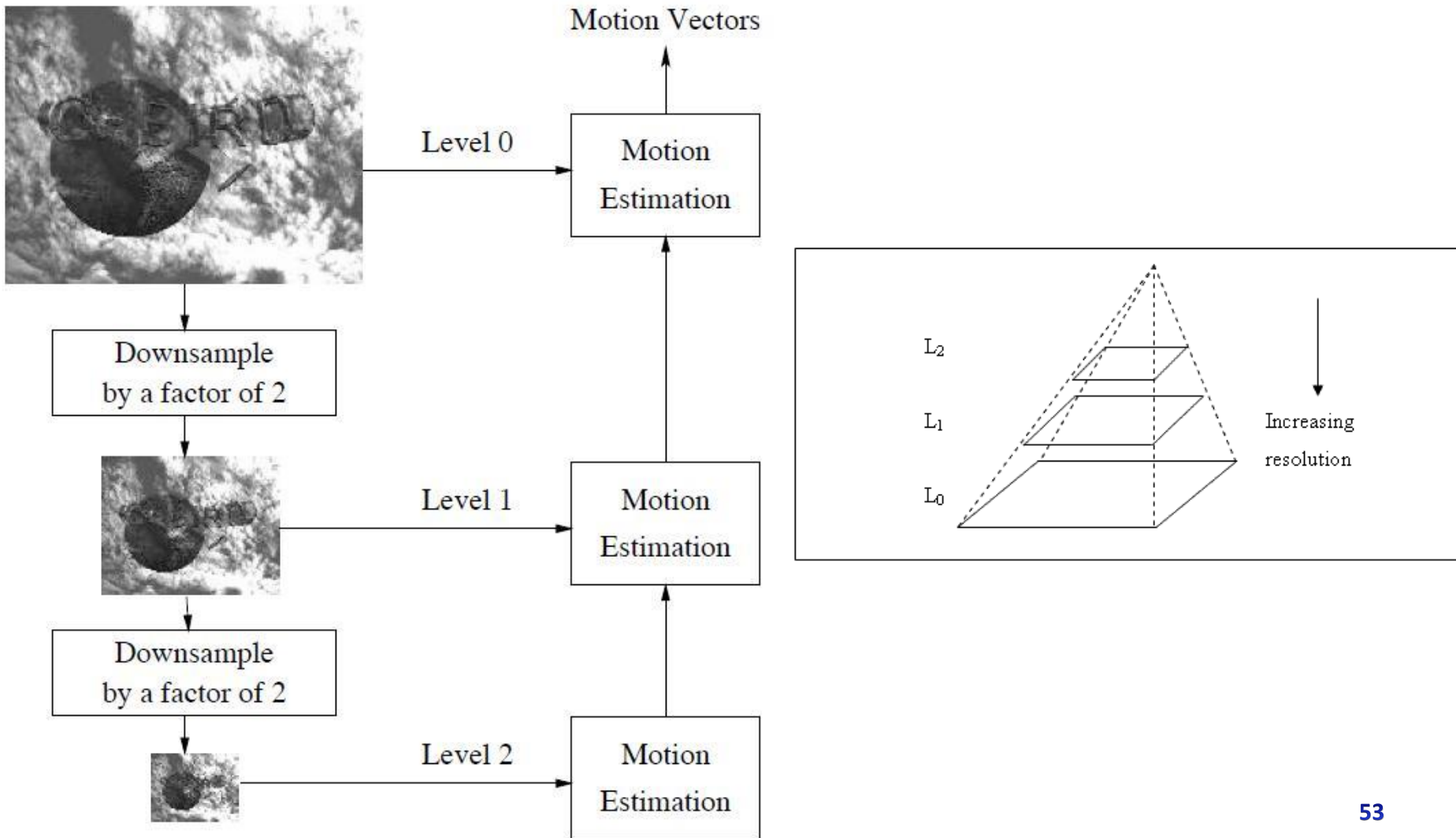


# Strategia Hierarchical Search

---

- Strategia Hierarchical Search
  - La procedura di ricerca è eseguita utilizzando diverse versioni dei frame a risoluzioni diverse (ottenute mediante sotto-campionamento)
  - Trova il vettore di movimento che minimizza l'errore nel frame a più bassa risoluzione
  - Si migliora la stima modificando il vettore iterativamente incrementando la risoluzione dei frame
- La Hierarchical Search offre flessibilità nel trade-off tra complessità ed efficienza di compressione

# Strategia Hierarchical Search



# **MPEG-2**

---

**Algoritmo e differenze rispetto ad MPEG-1**

# Introduzione

---

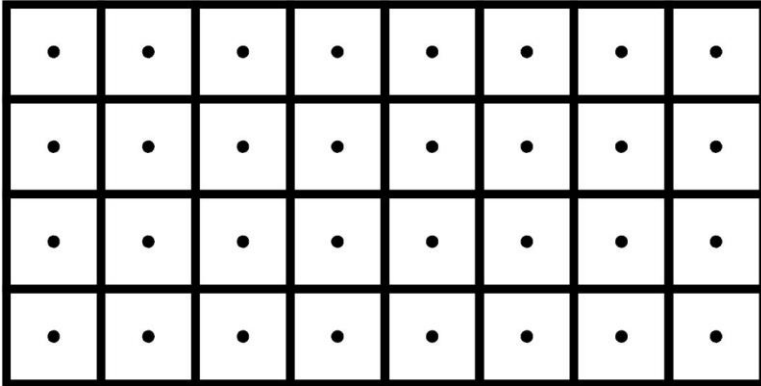
- MPEG-2 è stato sviluppato per supportare applicazioni con data rate maggiori di quelle che utilizzano la codifica MPEG-1
- MPEG-2 condivide molti degli aspetti della codifica di MPEG-1
- MPEG-2 definisce un meccanismo di codifica per segnali video interlacciati, mentre MPEG-1 prevede solo video a scansione progressiva
- MPEG-2 definisce tre formati di sotto-campionamento delle componenti di cromaticità  $C_b$  e  $C_r$ 
  - 4:2:0 → Componenti  $C_b$  e  $C_r$  sottocampionati nella direzione orizzontale e verticale (come MPEG-1)
  - 4:2:2 → Componenti  $C_b$  e  $C_r$  sottocampionati nella sola direzione orizzontale
  - 4:4:4 → Nessun sottocampionamento

# Introduzione

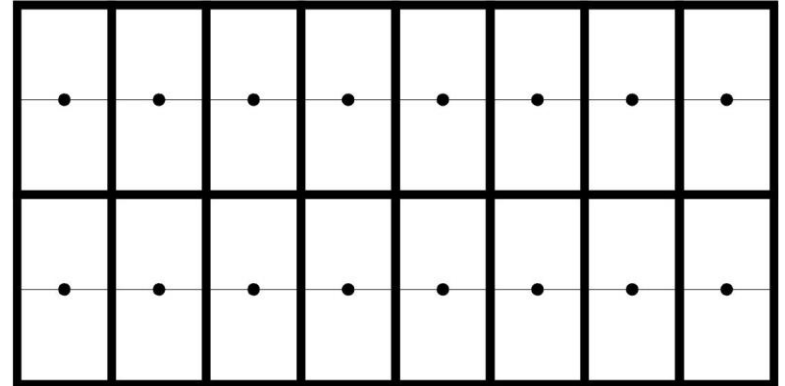
## Sotto-Campionamento

---

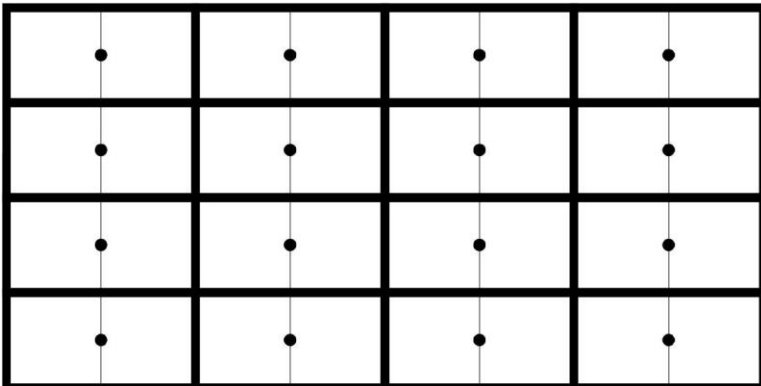
4:4:4



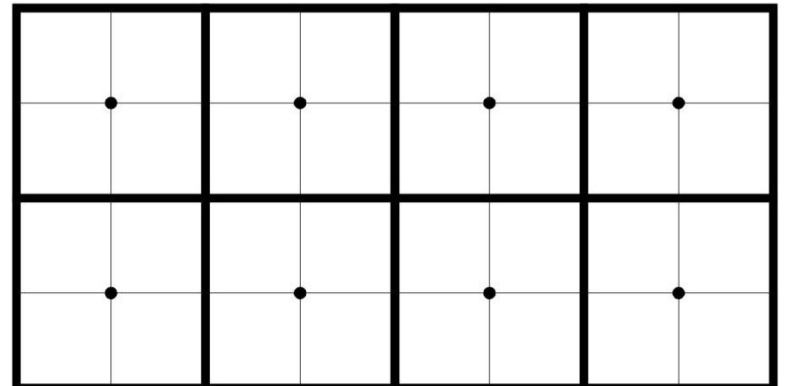
4:4:0



4:2:2

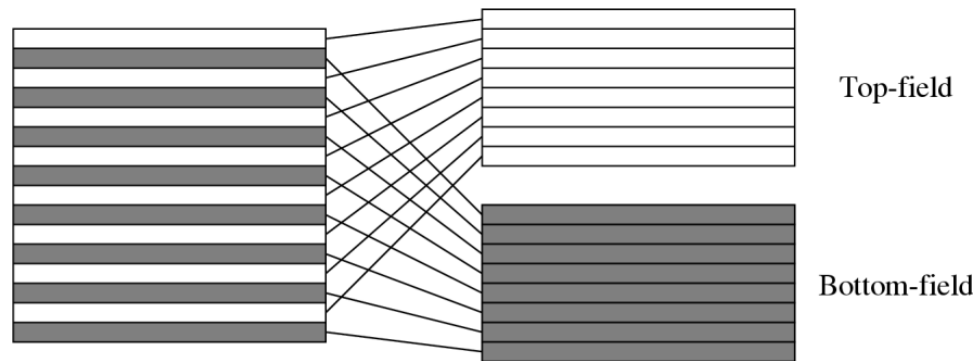


4:2:0

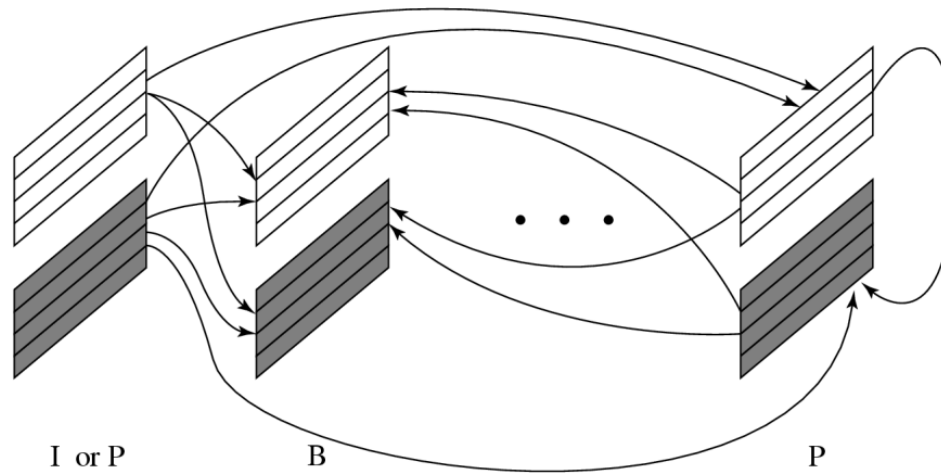




# Interlacciamento e Field Prediction



(a)



(b)



# Formato Video

---

- MPEG-2 specifica diversi formati video, definiti dalla combinazione di due parametri
  - **Profilo:** definisce un sottoinsieme della sintassi di MPEG-2
    - Riguarda gli algoritmi di compressione e risoluzione del colore
    - E' rivolto ad una applicazione specifica
    - Descrive diverse modalità di codifica applicabili
    - Definisce la complessità massima del decodificatore
  - **Livello:** introduce vincoli su alcuni parametri specificati dai profili
    - Risoluzione massima dell'immagine
    - Massimo bitrate
- MPEG-2 definisce 5 diversi profili
  - Simple Profile (SP)
  - Main Profile (MP)
  - SNR Scalable Profile (SNR)
  - Spatially Scalable (Spt)
  - High Profile (HP)

# Formato Video

## ■ Profili definiti da MPEG-2

Profile	Algorithms
HIGH	Supports all functionality of the SPATIAL Scalable Profile plus 3 layers with SNR and Spatial Scalable coding modes 4:2:2 YUV picture format
SPATIAL Scalable	Supports all functionality of the SNR Scalable Profile plus Spatial Scalable coding modes (2 layers) 4:0:0 YUV
SNR Scalable	Supports all functionality of the MAIN Profile plus SNR Scalable coding (2 layers) 4:2:0 YUV
MAIN (core)	Nonscalable coding supporting all functionality of the SIMPLE profile plus B-picture prediction modes
SIMPLE	Nonscalable coding supporting coding progressive and interlaced video random access I, P-picture prediction modes 4:2:0 YUV

# Formato Video

- MPEG-2 definisce 4 diversi livelli
  - Low Level (LL)
  - Main Level (ML)
  - High Level 1440 (HL-1440)
  - High Level (HL)
- Esempio nel caso di Simple Profile @ Main Level

Level	Parametro	Limite
Main (SP@ML)	Campioni/linea	720
	Linee/frame	576
	Frame/secondo	30
	Bitrate	15 Mbps

# Formato Video

LEVEL/ PROFILE	SIMPLE	MAIN	SNR	SPATIAL	HIGH
HIGH		1920x1152 80 Mb/s			1920x1152 100 Mb/s
HIGH 1400		1440x1152 60 Mb/s		1440x1152 60 Mb/s	1440x1152 80 Mb/s
MAIN	720x576 15 Mb/s	720x576 15 Mb/s	720x576 15 Mb/s		720/576 20 Mb/2
LOW		352x288 4 Mb/s	352x288 4 Mb/s		

# MPEG-2

---

## System Layer

# Introduzione

---

- Oltre a definire la codifica video e audio, MPEG specifica
  - La struttura dei pacchetti utilizzati per il trasporto del flusso multimediale
  - I meccanismi di sincronizzazione
  - Un protocollo di trasporto che garantisce l'integrità della trasmissione del flusso multimediale
- Il livello di sistema di MPEG definisce come integrare e sincronizzare diversi flussi dati elementari (Elementary Streams, ES)
- Elementary Stream: singolo flusso multimediale, solitamente output di un codificatore
  - Audio
  - Video
  - Dati e controllo (es. sottotitoli)

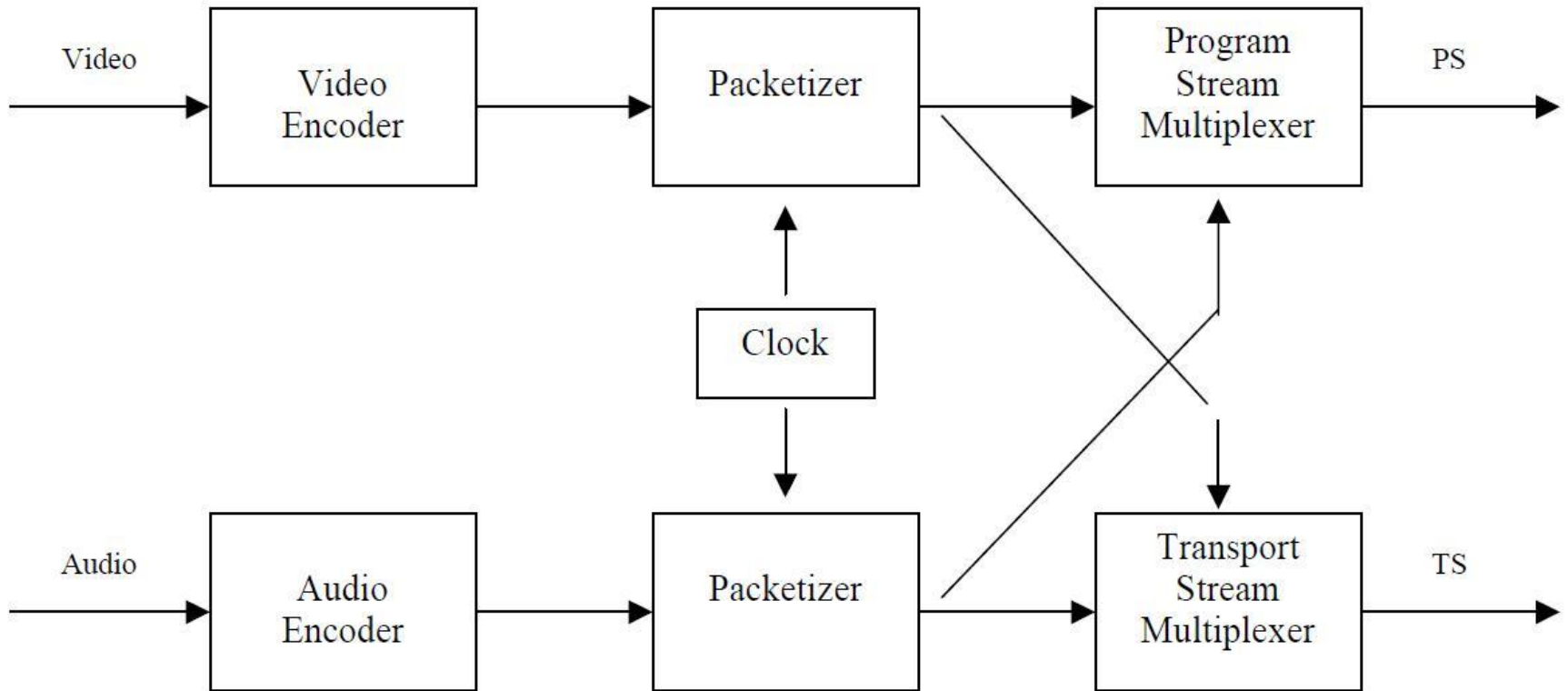


# Introduzione

---

- I flussi ES vengono dapprima strutturati in pacchetti per formare il Packetized Elementary Stream (PES)
- I diversi PES contengono un timestamp per la sincronizzazione dei diversi flussi
- I diversi PES vengono successivamente multiplati per formare un unico flusso di trasporto
- MPEG definisce due alternative modalità per la generazione del flusso di trasporto (multiplazione)
  - Program Stream (PS)
  - Transport Stream (TS)

# Introduzione



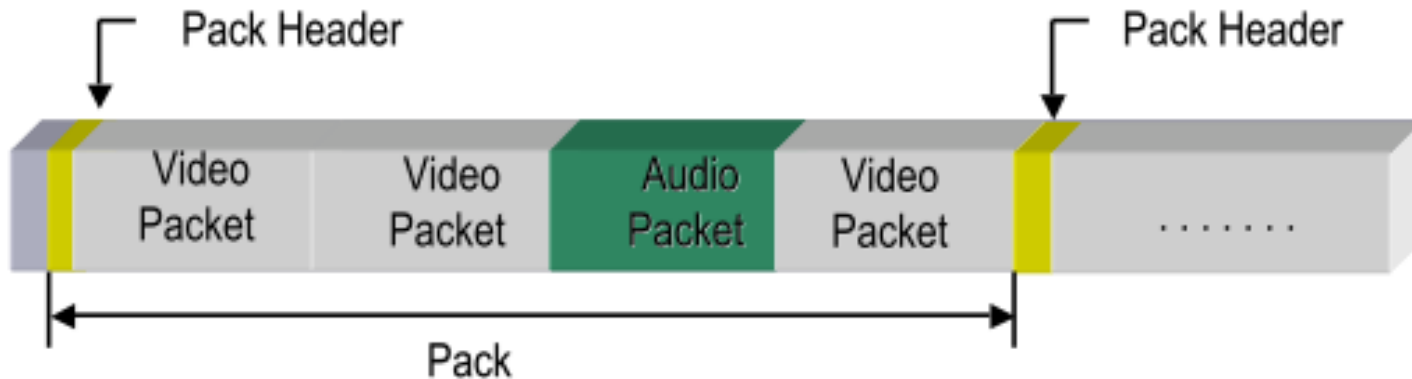
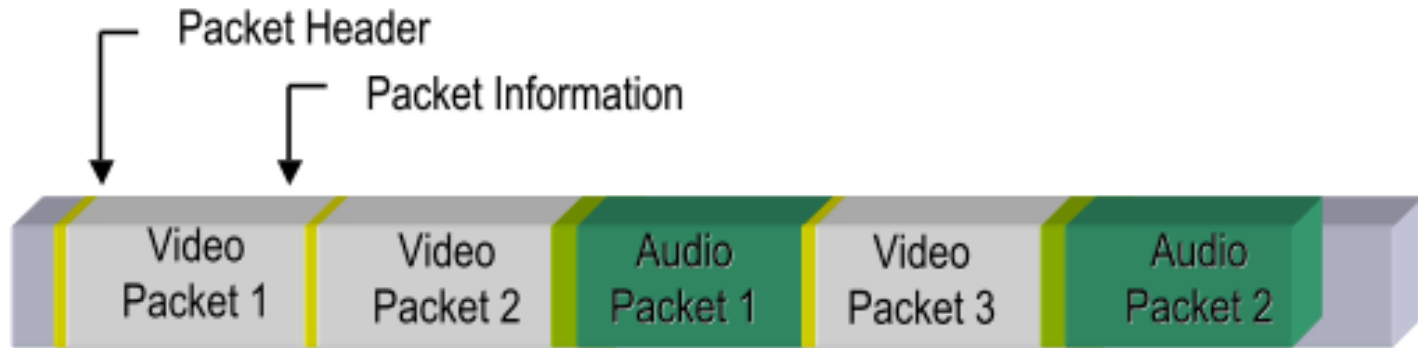
# Program Stream

---

- MPEG PS è stato progettato per la trasmissione/memorizzazione di contenuti multimediali attraverso canali di comunicazione ad alta affidabilità, ossia con errori di trasmissione trascurabili (es. CD-ROM)
- MPEG PS è composto da flussi audio/video pacchettizzati (PES) multiplati assieme e trasmessi/memorizzati attraverso pacchetti di dimensione variabile (*pack*)
- I flussi audio e video sono sincronizzati e possono essere decodificati e riprodotti in modo sincrono

# Formato Program Stream

---



# Formato Program Stream

---

- Un insieme di PES audio/video/data è multiplato in un pack
- Ad ogni pack viene aggiunto un header, che trasporta informazioni relative a
  - Datarate di picco
  - Numero di stream multiplati
  - Informazioni temporali addizionali
- Il Program Stream è in grado di moltiplicare solo un insieme di PES audio/video/data relativi a uno stesso contenuto multimediale

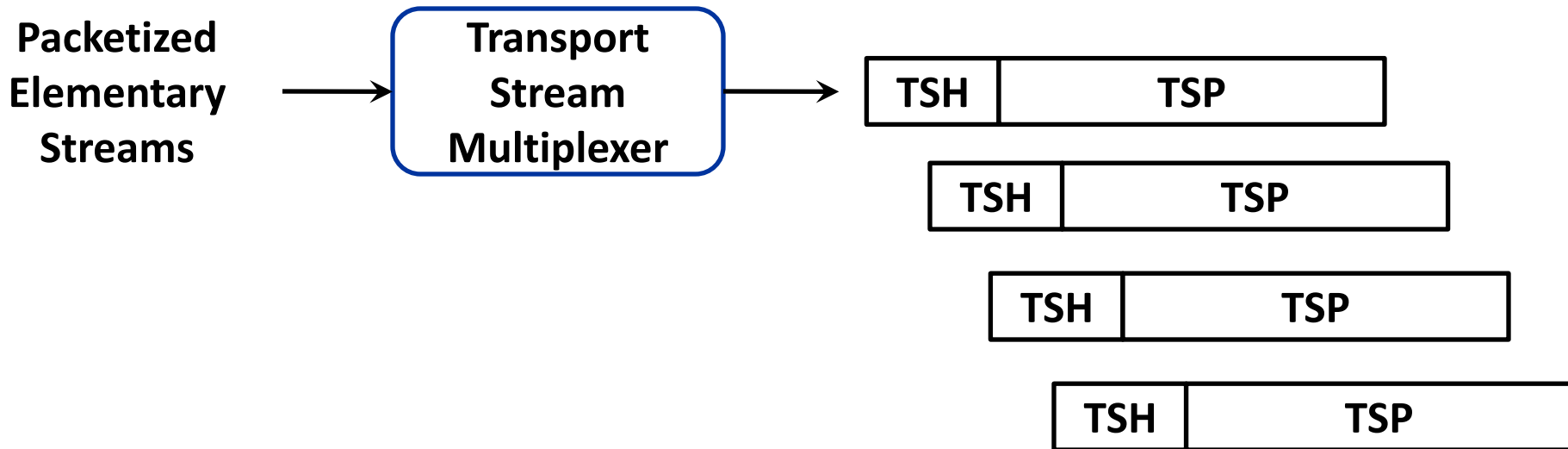
# Transport Stream

---

- MPEG TS è pensato per la trasmissione di contenuti multimediali attraverso reti poco affidabili (errori di trasmissione non trascurabili)
- MPEG TS è composto da più flussi audio e/o video (PES) non necessariamente sincronizzati, multiplati assieme e trasmessi attraverso pacchetti di dimensione fissa (188 byte)
- Il flusso definito dalla modalità TS è composto da
  - Un Header lungo 4 byte
  - Un Payload lungo 184 byte

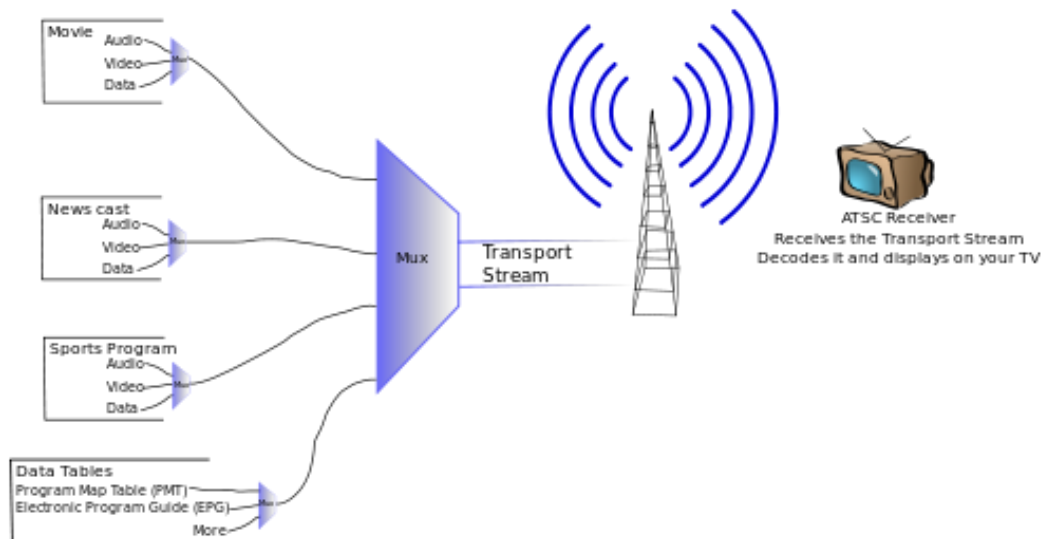
# Transport Stream

---



# Transport Stream

- Transport Stream è ampiamente diffuso nel broadcasting terrestre/satellitare
  - MPEG TS permette la multiplazione di più flussi audio/video PES relativi a differenti contenuti multimediali





# MPEG-4

---

## Sommario

# Introduzione

---

- MPEG-4 è uno degli standard più recenti di compressione di contenuti multimediali
  - E' nato come estensione di MPEG-1 e MPEG-2
- MPEG-4 estende gli standard che lo precedono adottando una codifica basata su oggetti (*object-based coding*)
  - Garantisce maggiore efficienza (alti rapporti di compressione)
  - Il flusso video può essere codificato manipolando sia gli oggetti visivi che la scena
- MPEG-4 è stato definito per un ampio spettro di applicazioni con bitrate compresi tra 5 kbit/s e 10 Mbit/s

# Codifica VOP-based

---

- Il formato del flusso video codificato (bitstream) definito da MPEG-4 è molto diverso da quello definito dai suoi predecessori
- MPEG-4 definisce una struttura gerarchica di rappresentazione della scena visualizzata orientata agli oggetti
  - Video-Object Sequence (VS): scena completa
  - Video Object (VO): oggetto specifico della scena
  - Video Object Layer (VOL): esistono multipli VOL per un VO in caso di codifica scalabile
  - Video Object Plane (VOP): snapshot di un VO in uno specifico istante
  - Group of VOPs (GOV): gruppo di VOP considerati insieme

# Codifica VOP-based

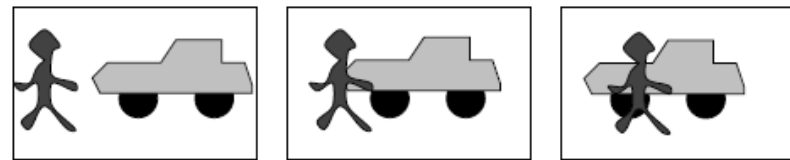
---

- In MPEG-1 e MPEG-2 la codifica è invece eseguita considerando tutto il frame (*codifica frame-based o block-based*)
  - E' possibile che un vettore di movimento per un macro-blocco non si riferisca a del reale movimento
- Nel caso di codifica VOP-based ogni VOP presenta una forma arbitraria (e non fissa come un macro-blocco)
  - Permette di ottenere un unico vettore di movimento che è consistente col reale movimento dell'oggetto

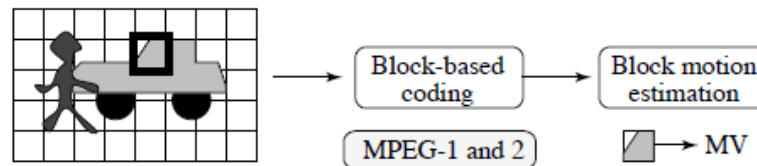
# Codifica VOP-based

Complessità

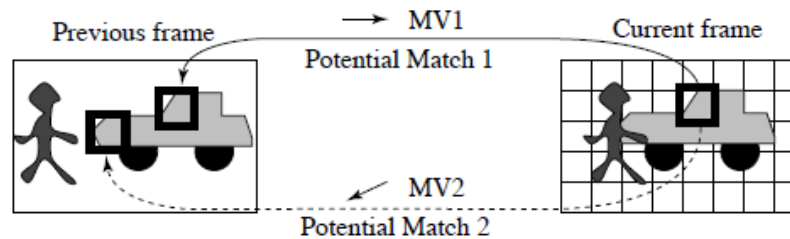
Rapporto di Compressione



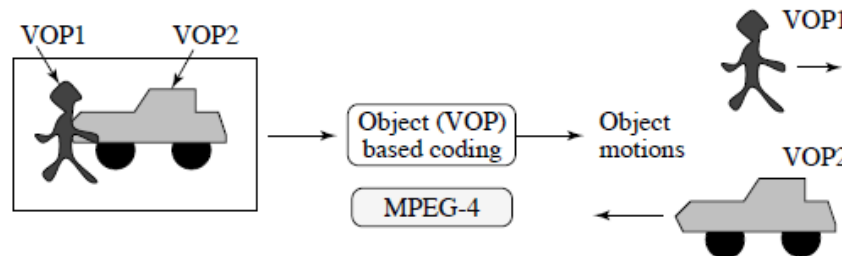
(a)



(b)



(c)



(d)

# Codifica VOP-based

---

- La codifica VOP-based definita da MPEG-4 sfrutta anche la tecnica di motion compensation
  - I-VOP: codifica Intra-frame applicata a un VOP
  - P-VOP: codifica Inter-frame con forward prediction applicata a un VOP
  - B-VOP: codifica Inter-frame con bi-directional prediction applicata a un VOP
- La codifica VOP-based risulta più complessa della frame-based poiché un VOP può avere una forma arbitraria
  - Si deve codificare anche l'informazione che identifica la forma oltre al vettore di movimento e i coefficienti della DCT

# Approfondimenti

---

## ■ Libri

- Chad Fogg; Didier J. LeGall; Joan L. Mitchell; William B. Pennebaker, “MPEG Video Compression Standard”, Springer 1996
- Iain Richardson, “H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia”, Wiley 2003

## ■ Articoli

- D. LeGall, “MPEG: a Video Compression Standard for Multimedia Applications”. Communications of the ACM, vol. 34, no. 4, pp. 46-58, April 1991
- IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on the H.264/JVC Video Coding Standard, July 2003