Συμπίεση Δεδομένων: Συμπίεση Ψηφιακού Βίντεο

Αλέξανδρος Ελευθεριάδης
Αναπ. Καθηγητής & Marie Curie Chair
Τμήμα Πληροφορικής και Τηλεπικοινωνιών
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

eleft@di.uoa.gr, (210) 727-5210

[Διάλεξη 4η]

Outline

- H.264/AVC Video Coding Standard
  - History and Brief Introduction
  - Adoption and Products
- Scalable Video Coding (SVC)
  - Temporal scalability
  - SNR scalability
  - Spatial scalability
- Scalable video transmission
  - Receiver-driven layered multicast
  - Video conferencing demo

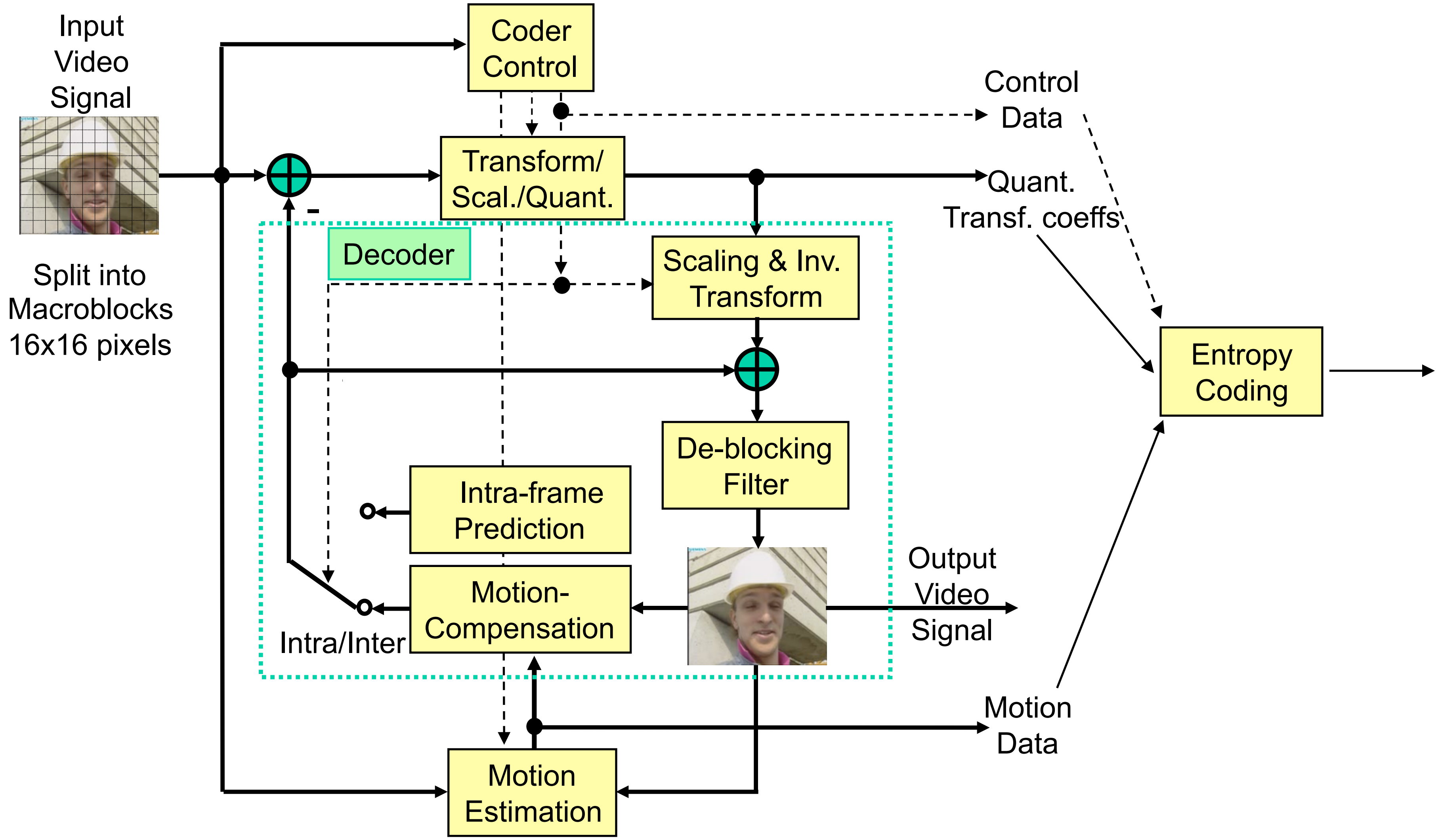
H.264 AVC (Advanced Video Coding)

Σημ.: Μερικές διαφάνειες και διαγράμματα είναι από τον Thomas Wiegand, HHI

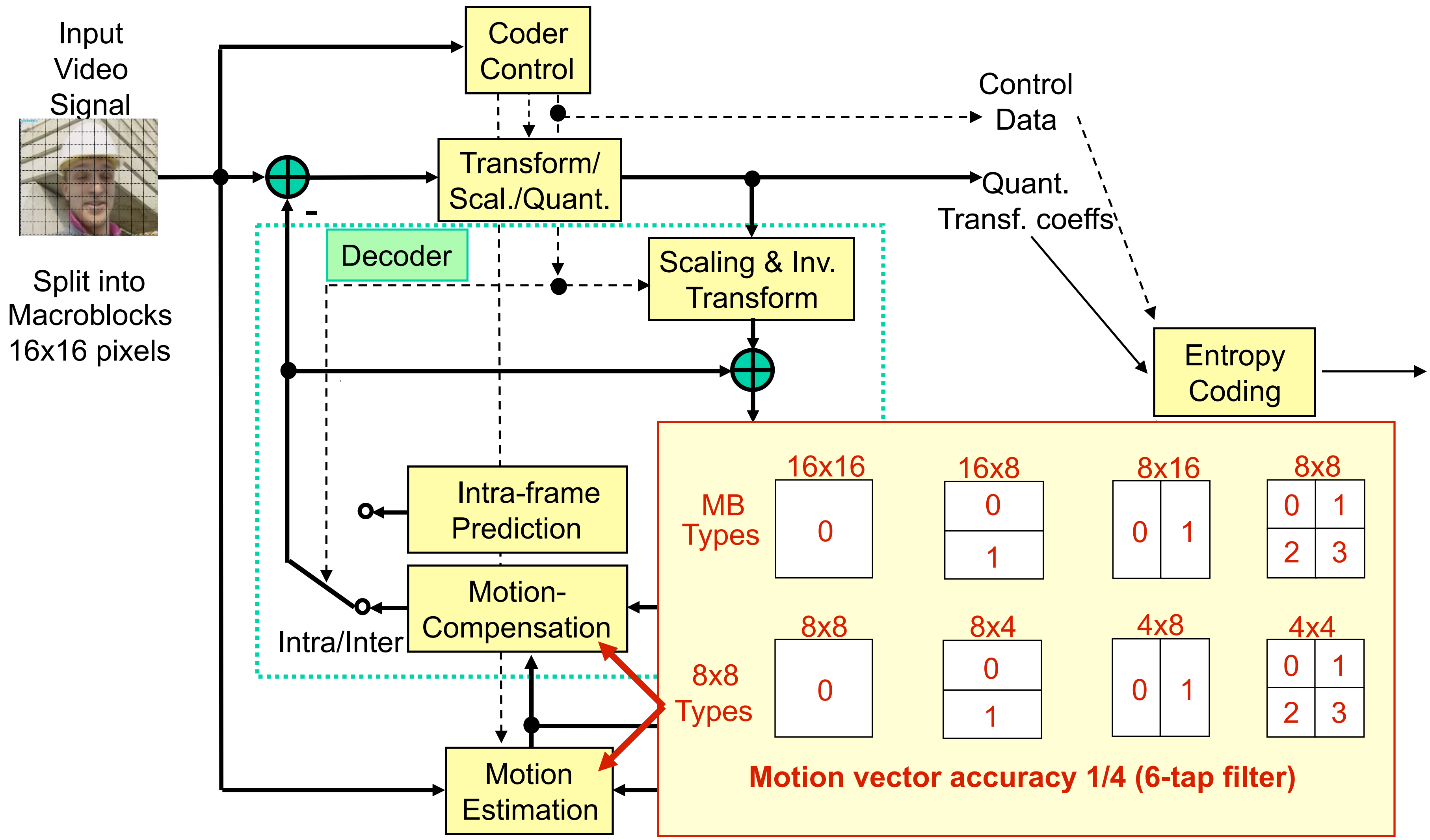
Standardization of H.264/AVC

- **1993-1997**: **VCEG** planning phase
  H.26P ⇨ H.263, H.26L ⇨ H.264/AVC
- **August 1999**: **1st test model (TML-1)**
- **December 2001**: Formation of **Joint Video Team (JVT)** between **VCEG** and **MPEG** ⇨ joint project **H.264/AVC** (similar to H.262/MPEG2 Video)
- **JVT Chairs**: Gary Sullivan (Microsoft), Ajay Luthra (Motorola), and Thomas Wiegand (HHI)
- **ITU-T | ISO/IEC Approval**: **Spring 2003**
- **ITU-T | ISO/IEC Approval of Fidelity Range Extensions: Autumn 2004**
- **ITU-T | ISO/IEC Approval of Scalability Extension**: **July 2007**
- **Initiation of Multi-view Video Extension**: **July 2006**

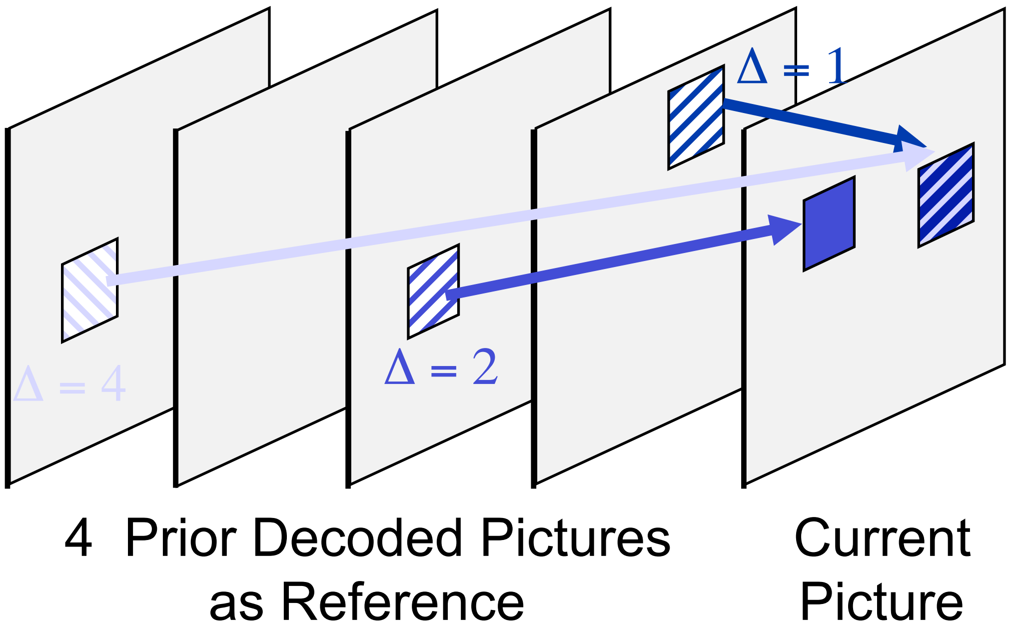
Hybrid Video Coding Structure



Motion Compensation Accuracy



Multiple Reference Pictures and Generalized B Pictures

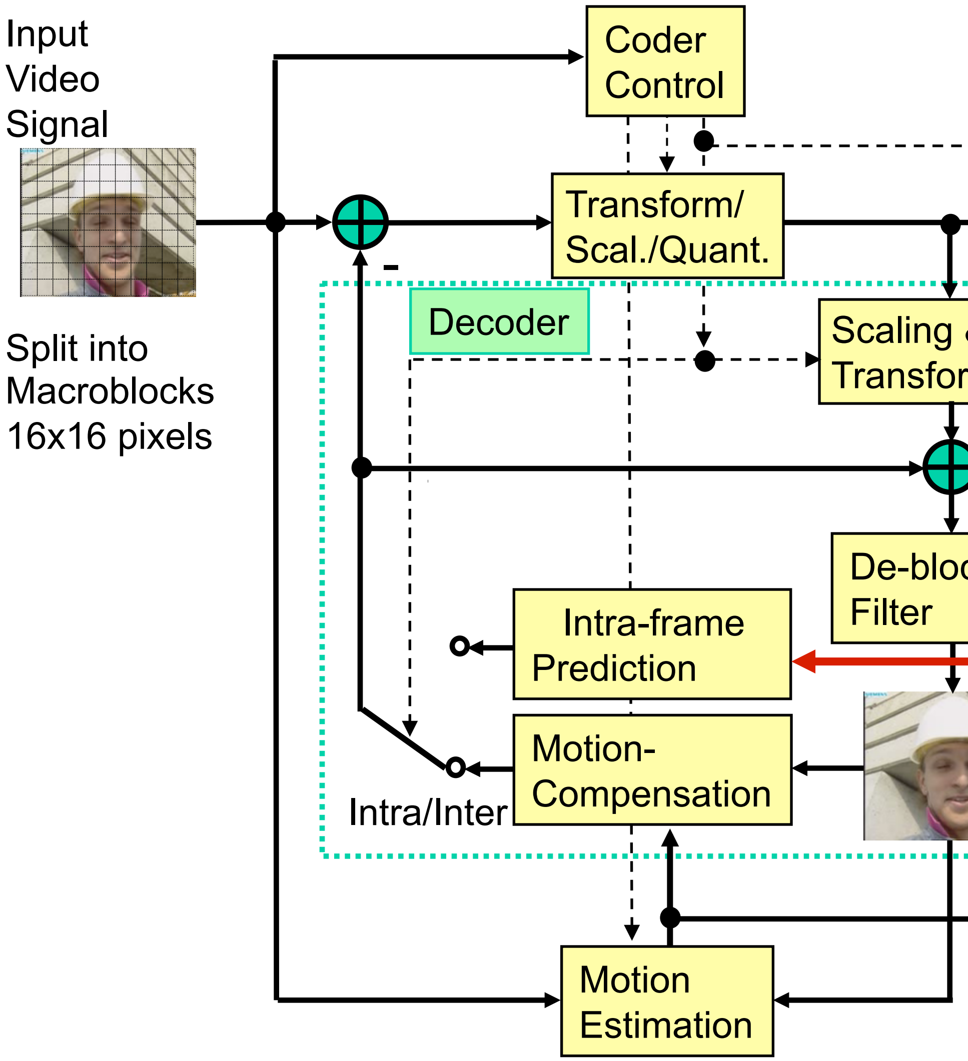


1. Extend motion vector by reference picture index Δ
2. Provide reference pictures at decoder side
3. In case of bi-predictive coding (B pictures): decode 2 sets of motion parameters

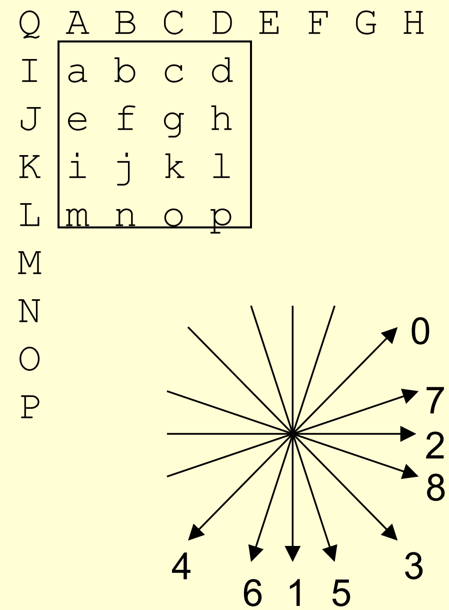
Flexible buffering of reference pictures generalizes B pictures:

- **B pictures can be reference pictures (decoupling of concepts)**
- **Reference picture for a B picture can be any prior decoded picture**

Intra Prediction

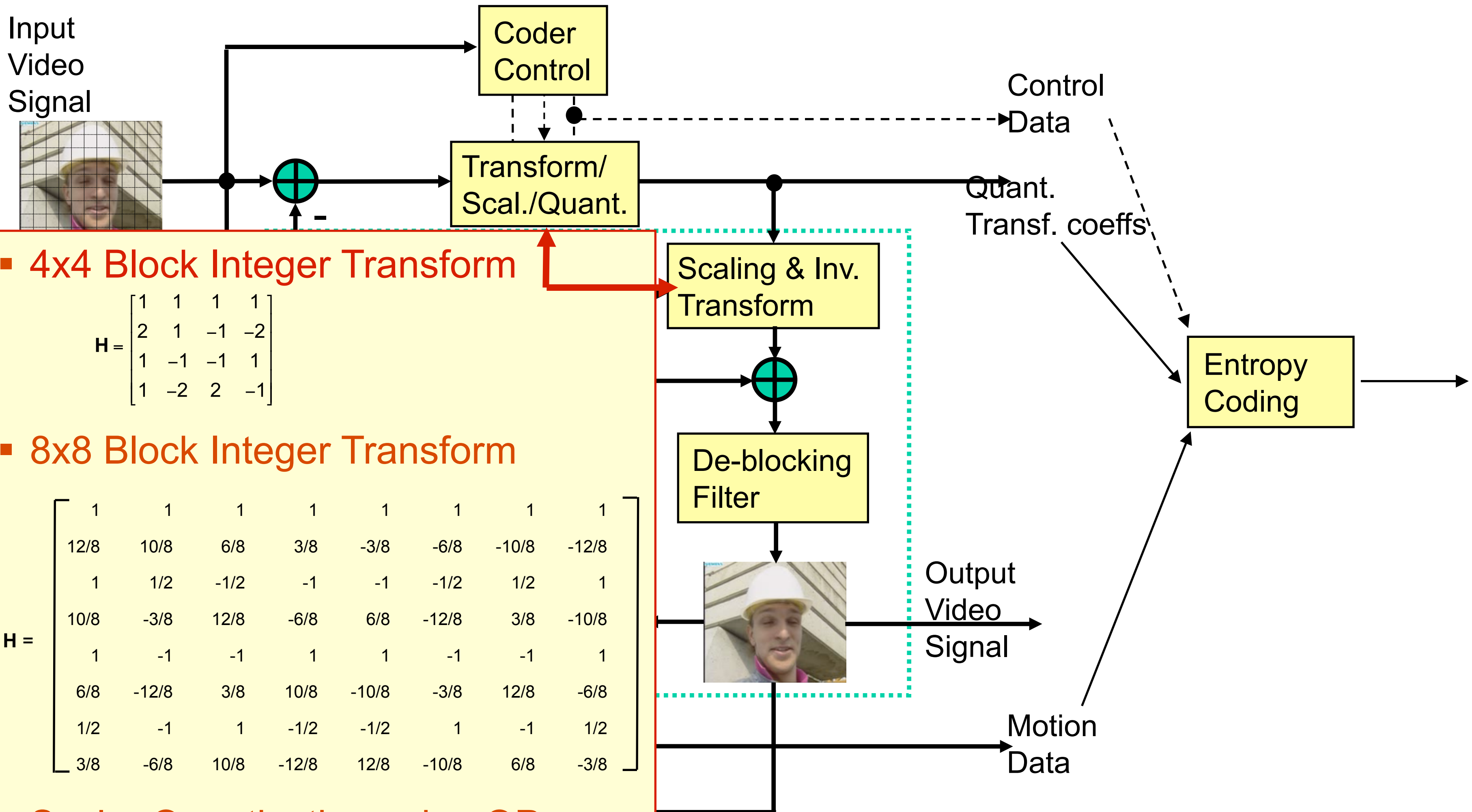


- Directional spatial prediction (9 types for luma, 1 chroma)



- e.g., Mode 3:
  diagonal down/right prediction
  a, f, k, p are predicted by
  (A + 2Q + I + 2) >> 2

Transform Coding



- 4x4 Block Integer Transform

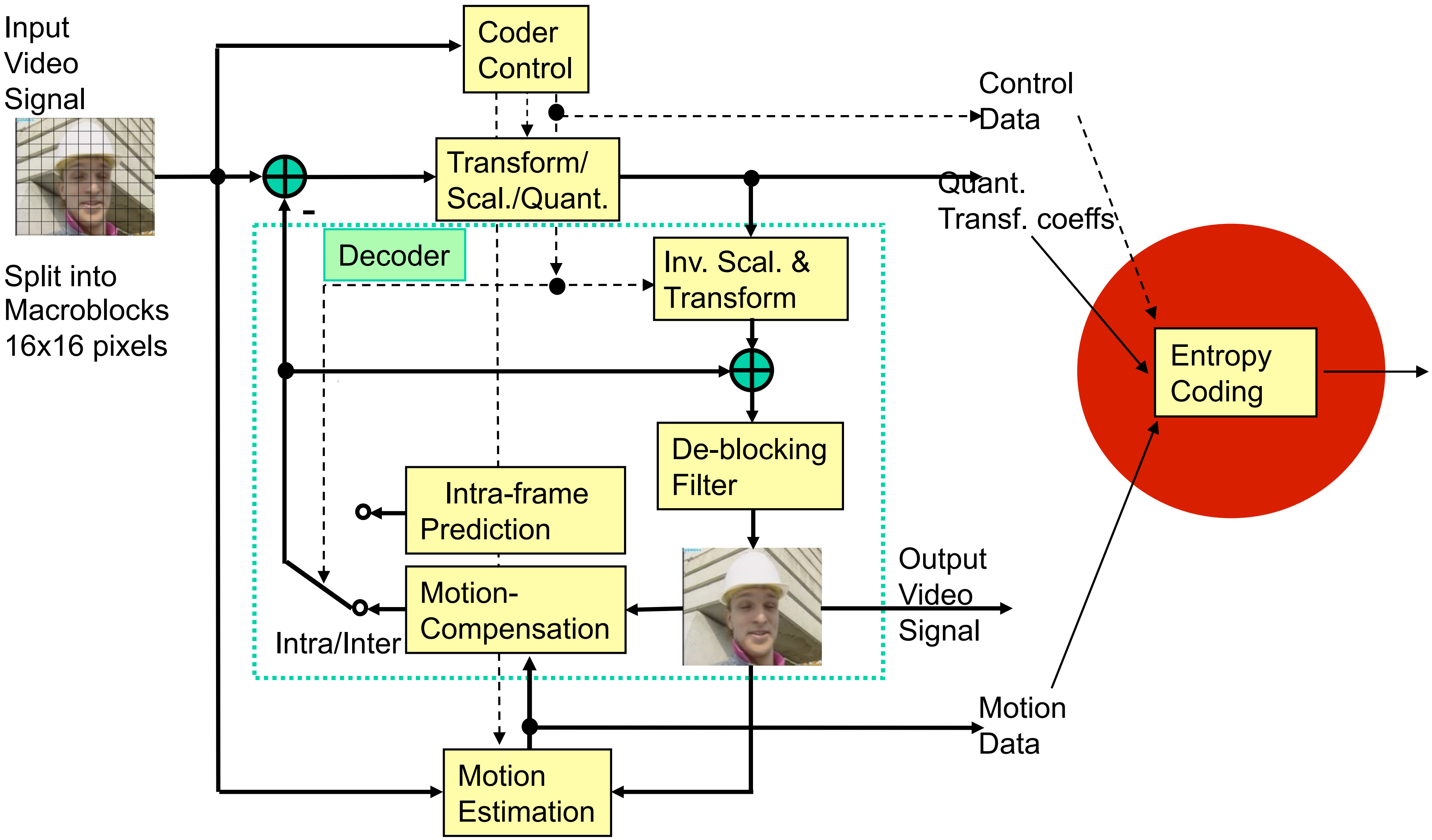
$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

- 8x8 Block Integer Transform

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 12/8 & 10/8 & 6/8 & 3/8 & -3/8 & -6/8 & -10/8 & -12/8 \\ 1 & 1/2 & -1/2 & -1 & -1 & -1/2 & 1/2 & 1 \\ 10/8 & -3/8 & 12/8 & -6/8 & 6/8 & -12/8 & 3/8 & -10/8 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 6/8 & -12/8 & 3/8 & 10/8 & -10/8 & -3/8 & 12/8 & -6/8 \\ 1/2 & -1 & 1 & -1/2 & -1/2 & 1 & -1 & 1/2 \\ 3/8 & -6/8 & 10/8 & -12/8 & 12/8 & -10/8 & 6/8 & -3/8 \end{bmatrix}$$

- Scalar Quantization using QP to Adjust Step Size / Bit Rate

Entropy Coding



Variable Length Coding

- Exp-Golomb code is used universally for all symbols except for transform coefficients (Section 9.1)

Bit string form	Range of codeNum
1	0
0 1 x_0	1-2
0 0 1 x_1 x_0	3-6
0 0 0 1 x_2 x_1 x_0	7-14
0 0 0 0 1 x_3 x_2 x_1 x_0	15-30
0 0 0 0 0 1 x_4 x_3 x_2 x_1 x_0	31-62
…	…

```
leadingZeroBits = -1;
for( b = 0; !b; leadingZeroBits++ )
    b = read_bits( 1 )
```

The variable codeNum is then assigned as follows:

$$\text{codeNum} = 2^{\text{leadingZeroBits}} - 1 + \text{read\_bits(leadingZeroBits)}$$

Context-Adaptive VLC (CAVLC)

- Context adaptive VLCs for coding of transform coefficients
  - Contexts are built dependent on transform coefficients
  - No end-of-block, but number of coefficients is decoded
  - Coefficients are scanned backwards

Context-based Adaptive Binary Arithmetic Codes (CABAC)

- Usage of adaptive probability models for most symbols
- Exploiting symbol correlations by using contexts
- Restriction to binary arithmetic coding
  - Simple and fast adaptation mechanism
  - Fast binary arithmetic codec based on table look-ups and shifts only
- Average bit-rate saving over CAVLC 10-15% for broadcast video

NAL Unit Format and Types

NAL unit header	NAL unit payload

NAL unit header: **1 byte consisting of**

- **forbidden_bit (1 bit): may be used to signal that a NAL unit is corrupt (useful e.g. for decoders capable to handle bit errors)**
- **nal_storage_idc (2 bit): signals relative importance, and if the picture is stored in the reference picture buffer**
- **nal_unit_type (5 bit): signals 1 of 10 different NAL unit types**
  - Coded slice (regular VCL data),
  - Coded data partition A, B, C (DPA, DPB, DPC),
  - Instantaneous decoder refresh (IDR),
  - Supplemental enhancement information (SEI),
  - Sequence and picture parameter set (SPS, PPS),
  - Picture delimiter (PD) and filler data (FD).

NAL unit payload**: an emulation prevented sequence of bytes.**

Application Standards Adoption Status

- 3GPP (recommended in rel 6)
- 3GPP2 (optional for streaming service)
- ARIB (Japan mobile segment broadcast)
- ATSC (preliminary adoption for robust-mode back-up channel)
- Blu-ray Disc Association (mandatory for Video BD-ROM players)
- DLNA (optional in first version)
- DMB (Korea - mandatory)
- DVB (specified in TS 102 005 and one of two in TS 101 154)
- DVD Forum (mandatory for HD DVD players)
- IETF AVT (RTP payload spec approved as RFC 3984)
- ISMA (mandatory specified in near-final rel 2.0)
- SCTE (under consideration)
- US DoD MISB (US government preferred codec up to 1080p)
- (And of course MPEG and the ITU-T)

Product Examples

- Video conferencing (Polycom, Tandberg, Sony)
- HDTV satellite (DirecTV, BSkyB, Echostar, …)
- Terrestrial TV (France, Estonia, Norway, Brazil, …)
- IPTV (AT&T, France Telecom, British Telecom, Deutsche Telekom, KPN, Belgacom, …)
- Mobile TV (Korea, Japan, Italy, USA, Qatar, Malaysia, …)
- Mobile phones (Nokia, Samsung, SonyEricsson, Apple, …)
- Mobile video players (Sony PSP, Apple iPod, …)
- Blu-ray disc and HD DVD (Sony, Samsung, Toshiba, LG, …)
- AVC HD Camcorders (Panasonic, Sony, Hitachi, …)
- Internet video streaming (Apple Quicktime, Adobe Flash-YouTube, …)

- *Video application area where H.264/AVC is not present: Digital Cinema*

Coder Control – Rate-Distortion Optimization (RDO)

- **Constrained problem**:

$$\min_{\boldsymbol{p}} D(\boldsymbol{p}) \quad \text{s.t.} \quad R(\boldsymbol{p}) \leq R_T$$

D Distortion
R Rate
R_T Target rate
p Parameter Vector

- **Unconstrained Lagrangian formulation**:

$$\boldsymbol{p}' = \arg\min_{\boldsymbol{p}} \left\{ D(\boldsymbol{p}) + \lambda \cdot R(\boldsymbol{p}) \right\}$$

with λ controlling the rate-distortion trade-off

- **Rate-Constrained Mode Decision**

$$D_2(M|QP)+\lambda_M R(M|QP)$$

M Evaluated macroblock mode out of a set of possible modes
Q Value of quantizer for transform coefficients
l_M Lagrange parameter for mode decision
D_2 Sum of squared differences (luminance & chrominance)
R Number of bits associated with header, motion, transform coefficients

- **Rate-Constrained Motion Estimation**

$$D_1(\boldsymbol{m},\Delta)+\lambda_D R(\boldsymbol{m},\Delta)$$

m Motion vector containing spatial displacement and picture reference parameter D
l_D Lagrange parameter for motion estimation
D_1 Sum of absolute differences (luminance)
R Number of bits associated with motion information

Αναφορές για RDO

- IEEE Signal Processing Magazine, Vol. 15, Nr. 6, 1998 – αφιέρωμα σε τεχνικές RDO για εικόνες/βίντεο

Ιδιαίτερα:

- G. Sullivan, T. Wiegand, “Rate-distortion optimization for video compression” (pp. 74-90).

Επιτεύγματα στην Κωδικοποίηση Βίντεο