18-796

Multimedia Communications:
Coding, Systems, and Networking

*Prof. Tsuhan Chen*
tsuhan@ece.cmu.edu

# H.263, H.263v2, and H.26L

## Very Low Bit Rate Video Coding

- Started around Nov 1993 by ITU-T SG 15
  - PSTN and mobile network: about 10~24 kbits/s
- Near-term
  - H.263 and H.263 Version 2 (“H.263+”)
  - Enhancements of H.261 (p×64)
- Long-term
  - H.26L
  - Fundamentally different from H.261



## H.263

- H.261 combined with MPEG-like features optimized at very low bit rates
- Compared to H.261
  - More picture formats, different GOB structures
  - Half-pel motion compensation, no loop filtering
  - 3-D VLC table
  - Four negotiable options
- Performance
  - About 3~4 dB better PSNR than H.261 at <64 kbits/s
  - 30% bit rate saving compared with MPEG-1

# Picture Formats

- More allowable formats than H.261

| | Sub-CIF | QCIF | CIF | 4CIF | 16CIF |
|---|---|---|---|---|---|
| Pels/line | 128 | 176 | 352 | 704 | 1408 |
| Lines | 96 | 144 | 288 | 576 | 1152 |

- Chrominance Sampling: 4:2:0 (same as H.261)



# GOB Structures

| GOB 0 |
|---|
| GOB 1 |
| GOB 2 |
| GOB 3 |
| GOB 4 |
| GOB 5 |
| GOB 6 |
| GOB 7 |
| GOB 8 |
| GOB 9 |
| GOB 10 |
| GOB 11 |
| GOB 12 |
| GOB 13 |
| GOB 14 |
| GOB 15 |
| GOB 16 |
| GOB 17 |

CIF

| GOB 0 |
|---|
| GOB 1 |
| GOB 2 |
| GOB 3 |
| GOB 4 |
| GOB 5 |
| GOB 6 |
| GOB 7 |
| GOB 8 |

QCIF

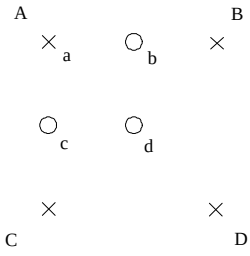
| GOB 0 |
|---|
| GOB 1 |
| GOB 2 |
| GOB 3 |
| GOB 4 |
| GOB 5 |

sub-QCIF

# Half-Pel Prediction

- Resolution of MVs is half-pel
- MV range: [–16,15.5]

- Bilinear interpolation to fill in pels



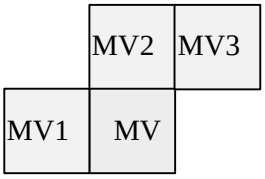

× Integer pixel position

○ Half pixel position

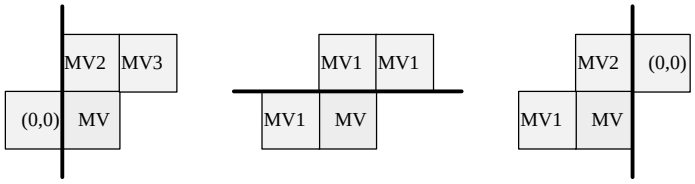
a = A
b = (A + B + 1) / 2
c = (A + C + 1) / 2
d = (A + B + C + D + 2) / 4



# Predictive Coding of Motion Vectors




MV: Current motion vector
MV1, MV2, MV3: predictors
prediction = median(MV1,MV2,MV3)




Picture boundary or GOB boundary

# 3D VLC Table

- A symbol is (last,run,level)
  - Last =1: indicates the last nonzero coefficient
  - No need for EOB

| LAST | RUN | LEVEL | CODE |
|---|---|---|---|
| 0 | 0 | 1 | 10s |
| 0 | 0 | 2 | 1111 s |
| 0 | 0 | 3 | 0101 01s |
| 0 | 0 | 4 | 0010 111s |
| 0 | 0 | 5 | 0001 1111 s |
| 0 | 0 | 6 | 0001 0010 1s |
| 0 | 0 | 7 | 0001 0010 0s |
| 0 | 0 | 8 | 0000 1000 01s |
| 0 | 0 | 9 | 0000 1000 00s |
| 0 | 0 | 10 | 0000 0000 111s |
| 0 | 0 | 11 | 0000 0000 110s |
| 0 | 0 | 12 | 0000 0100 000s |
| … | … | … | … |

| LAST | RUN | LEVEL | CODE |
|---|---|---|---|
| … | … | … | … |
| 1 | 0 | 1 | 0111 s |
| 1 | 0 | 2 | 0000 1100 1s |
| 1 | 0 | 3 | 0000 0000 101s |
| 1 | 1 | 1 | 0011 11s |
| 1 | 1 | 2 | 0000 0000 100s |
| 1 | 2 | 1 | 0011 10s |
| 1 | 3 | 1 | 0011 01s |
| 1 | 4 | 1 | 0011 00s |
| 1 | 5 | 1 | 0010 011s |
| 1 | 6 | 1 | 0010 010s |
| 1 | 7 | 1 | 0010 001s |
| 1 | 8 | 1 | 0010 000s |



# Negotiable Options in H.263

- Unrestricted Motion Vector Mode
- Advanced Prediction Mode
- PB-Frame Mode
- Syntax-based Arithmetic Coding Mode

- Usage
  - The decoder signals the encoder which of the options it has the capability to decode. If the encoder supports some of these options, it may enable them

## Unrestricted Motion Vector (UMV) Mode

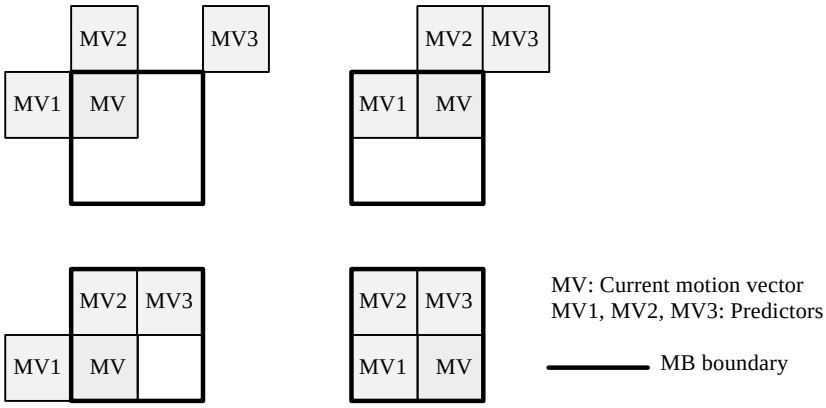
- Motion vectors may point outside the picture
  - Edge pels are repeated
  - Significant gain for movement along the edge of the pictures, especially for smaller picture formats
- Extension of the motion vector range
  - [–31.5, 31.5] instead of [–16, 15.5]
  - Especially useful for 4CIF, 16CIF
- Good for camera movement and background motion



## Advanced Prediction Mode (AP)

- Four motion vectors for each MB (one for each 8×8 block)
  - Use more bits, but give better prediction. Encoder decides which MBs to apply
- Overlapped block motion compensation (OBMC)
  - Less blocking artifacts
- Motion vectors may point outside as in UMV

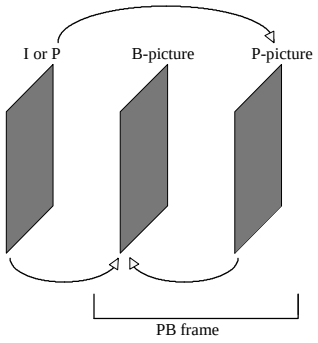
# MV Prediction for AP






# PB-Frame (PB) Mode

- A PB-frame consists of two pictures
  - P-picture: Predicted from the last decoded picture
  - B-picture: Predicted from both the last decoded picture and the P-picture currently being coded
- Coding Process
  - MVs of B derived from MVs of P
  - One MB in P followed by one MB in B



## Syntax-Based Arithmetic Coding (SAC) Mode

- Arithmetic coding is used instead of VLC
- Cumulative frequencies are provided
- About 5% less bits at the same SNR
  - Inter frames: 3~4%
  - Intra blocks and frames: ~10%



## H.263 Version 2 (H.263v2)

- 1996~1998, also known as H.263+
- Enhancements of H.263
  - as opposed to H.26L
- Example features
  - Extended source formats
  - 16 negotiable options
  - Supplemental enhancement information

## Custom Source Formats

- Higher picture clock frequency (PCF)
- Custom picture formats
- Custom pixel aspect ratios (PAR)

| Pixel Aspect Ratio | Pixel Width : Pixel Height |
|---|---|
| Square | 1:1 |
| CIF | 12:11 |
| 525-type for 4:3 picture | 10:11 |
| CIF for 16:9 picture | 16:11 |
| 525-type for 16:9 picture | 40:33 |
| Extended PAR | m:n, m and n are relatively prime |

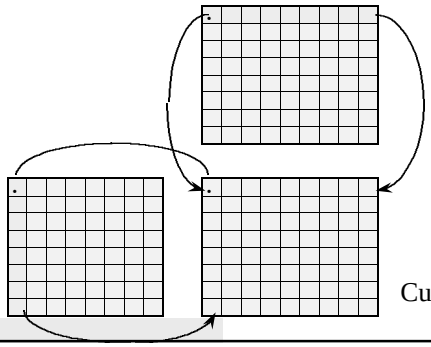


## New Negotiable Options

- Coding efficiency
  - Advanced Intra Coding Mode
  - Alternate Inter VLC Mode
    - Use intra table for inter DCT
  - Deblocking Filter Mode
    - Depending on quantization step size
  - Modified Quantization Mode
    - More flexible changes of quantization step sizes
    - Finer quantization for chrominance
    - Extended DCT range
  - Improved PB-Frame Mode
    - Forward, backward, or bi-directional

# Advanced Intra Coding

- A separate VLC table for intra DCT
- Modified inverse quantization
- Spatial prediction of DCT coefficients
  - DC only, vertical DC & AC, horizontal DC & AC






# New Negotiable Options (cont.)

- Error robustness
  - Slice Structure Mode
    - Sequential or arbitrary
    - Rectangular or not
  - Reference Picture Selection Mode
    - Multiple reference pictures are stored
    - Avoid using pictures that contain errors
  - Independent Segment Decoding Mode
    - Prevent error propagation

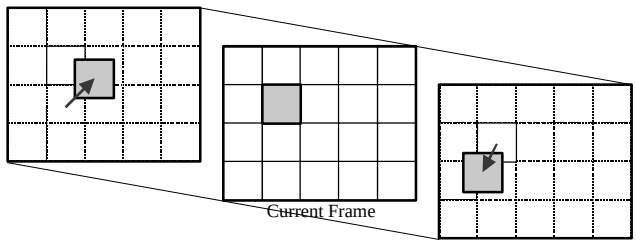
# New Negotiable Options (cont.)

- Scalability
  - Decode partial information from partial bitstream
  - To fit various bandwidth requirements
  - To fit terminals with different capabilities
- Scalability Mode
  - Temporal Scalability : Bi-directional prediction
  - Spatial Scalability
  - SNR Scalability



# Bi-Directional Prediction

- Prediction from the previous frame, or the prediction from the future frame, or an average of both is used as the final prediction.

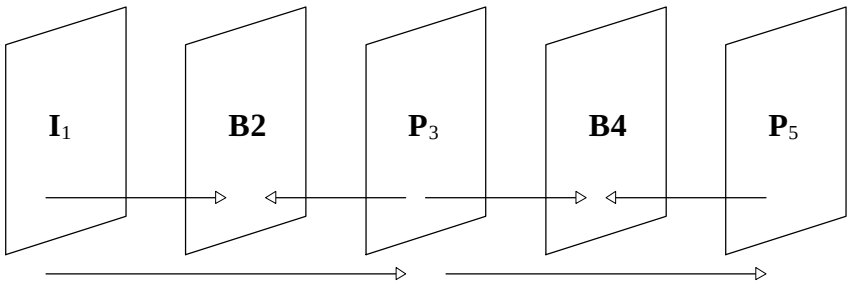


# Bi-Directional Prediction (cont.)

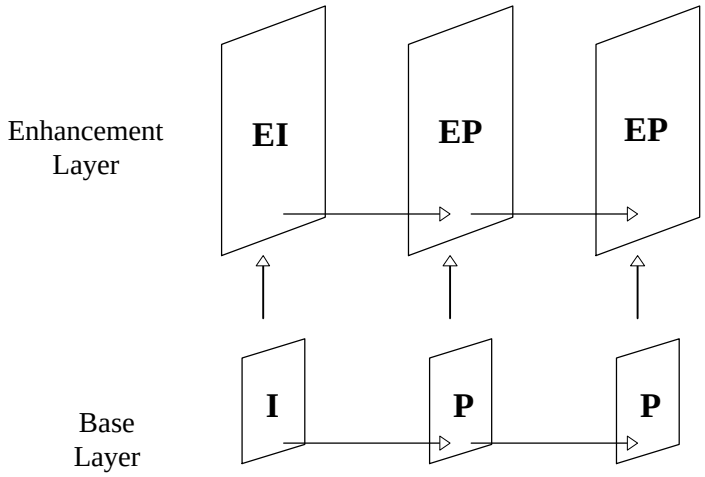
- Advantages
  - Higher coding efficiency
  - No un-covered background problem
  - Increased frame rate with few extra bits
- Disadvantages
  - Two frames stores are needed at the decoder
  - More delay



# Temporal Scalability

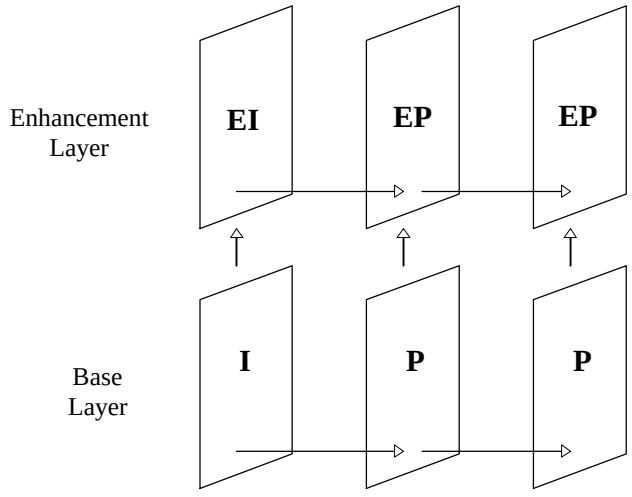


# Spatial Scalability

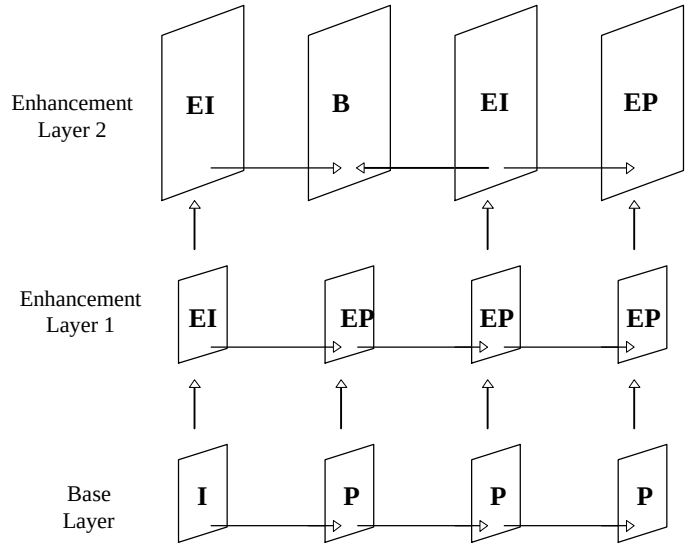





# SNR Scalability



# Multilayer Scalability



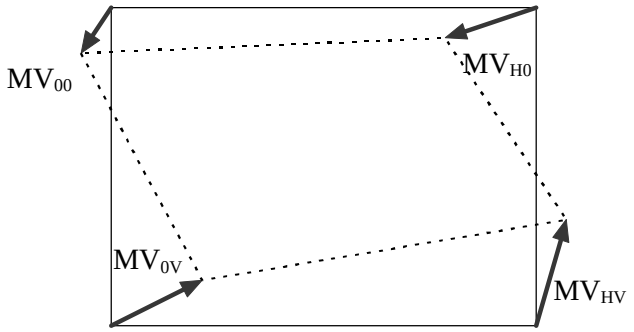



# New Negotiable Options (cont.)

- Others
  - Reduced-Resolution Update Mode
    - Update a picture at a lower spatial resolution
    - To retain details in stationary background
  - Reference Picture Resampling Mode

# Reference Picture Resampling

- For source format changes
- Global motion compensation
- Special-effect warping






# Modifications to UMV Mode

- Single value VLC
  - Easy implementation
- Reversible VLC table
  - Better error resilience
- Larger motion vector range
  - Depending on the picture size
  - Up to [–256, 255.5]

# Supplemental Enhancement Information

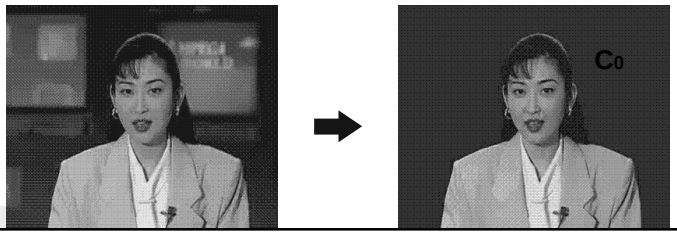
- Can be discarded by the decoder
- Enhanced features
  - Picture freeze and release
    - Complete or partial picture
  - Tagging information
    - Snapshot
    - Video segment start/end
    - Progressive refinement start/end
  - **Chroma key**: to represent transparent pixels



# Description

- A video sequence $f(x,n)$ with regions $R_i$
- For each region $R_i$, replace the non-object area with a special color $C_0$

$$g(\mathbf{x},n)=\begin{cases} f(\mathbf{x},n) & \text{if } \mathbf{x}\in\Re_i \\ C_0 & \text{if } \mathbf{x}\in\overline{\Re_i} \end{cases}$$



## Description (cont.)

- The sequence $g(x,n)$ is then coded and transmitted
- The color $C_0$ and a threshold $T$ are sent as side information to the decoder
- Chroma-keying
  - Based on $C_0$ and $T$ , the decoder detects transparency and recovers the region boundary

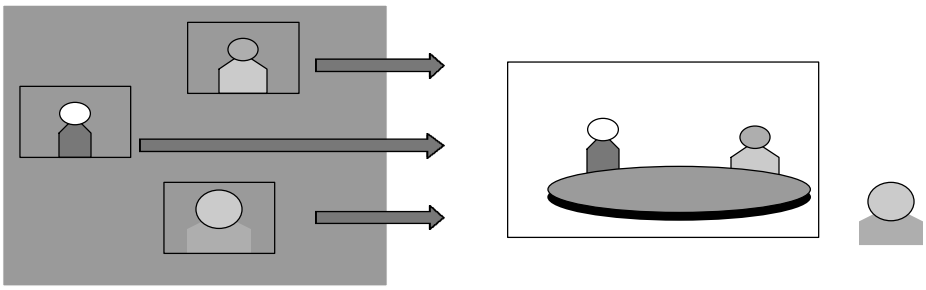


## Advantages of Chroma Keying

- Low complexity
- Efficiency
  - Minimal side information
  - Implicit shape coding
  - Solves “uncovered background”
- Minimal syntax change

# Example Application

- Virtual video conference

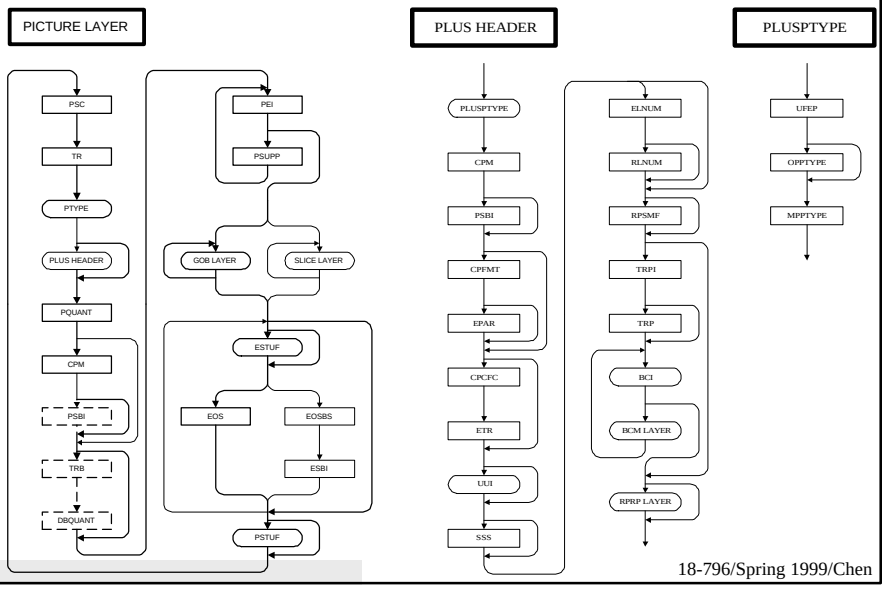




# Test Model Near-Term (TMN)

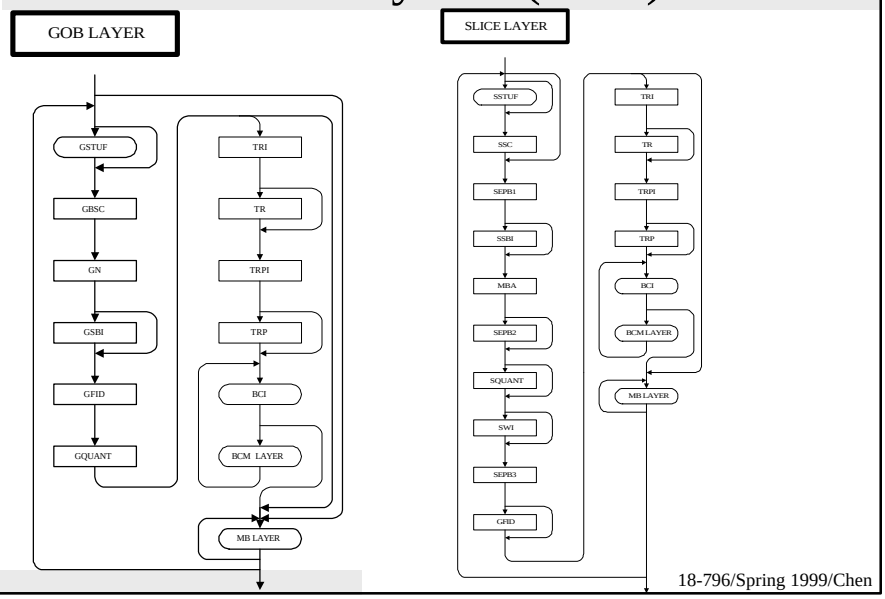
- Encoder specifications
- Rate control
  - Depending on buffer fullness
    - Skip pictures
    - Increase quantization step sizes

# Bitstream Syntax

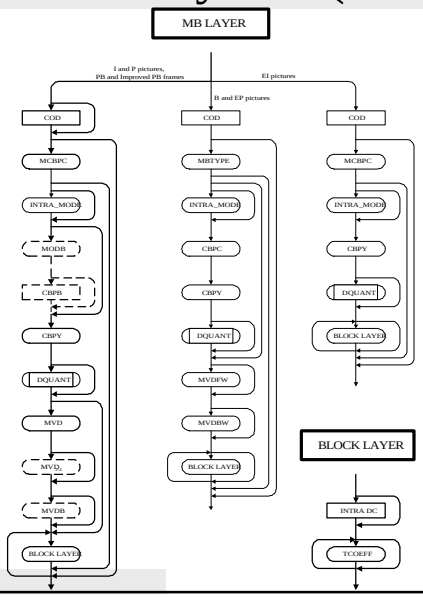





# Bitstream Syntax (cont.)



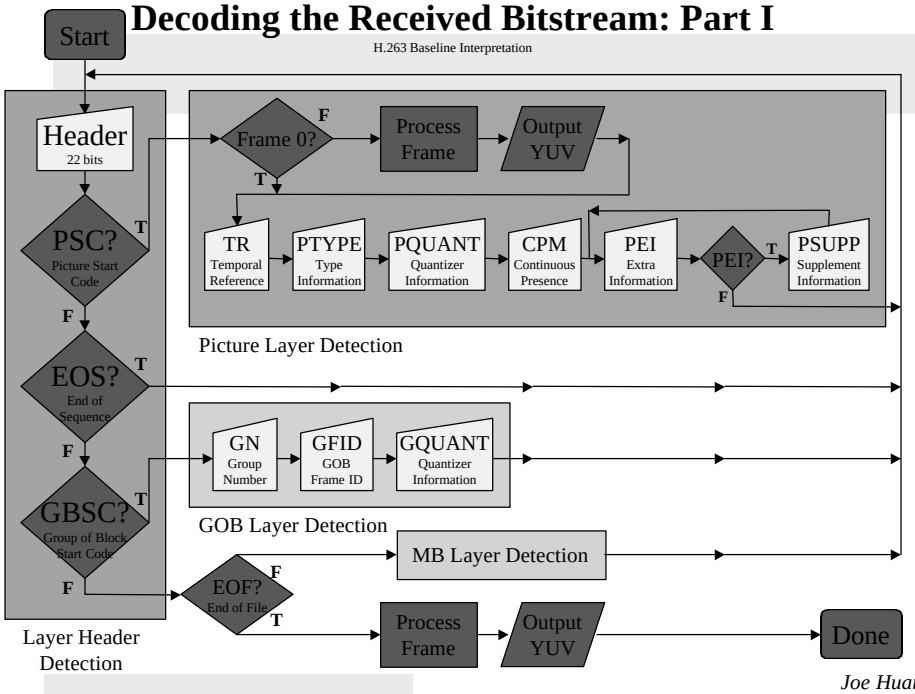
# Bitstream Syntax (cont.)






# Decoding the Received Bitstream: Part I

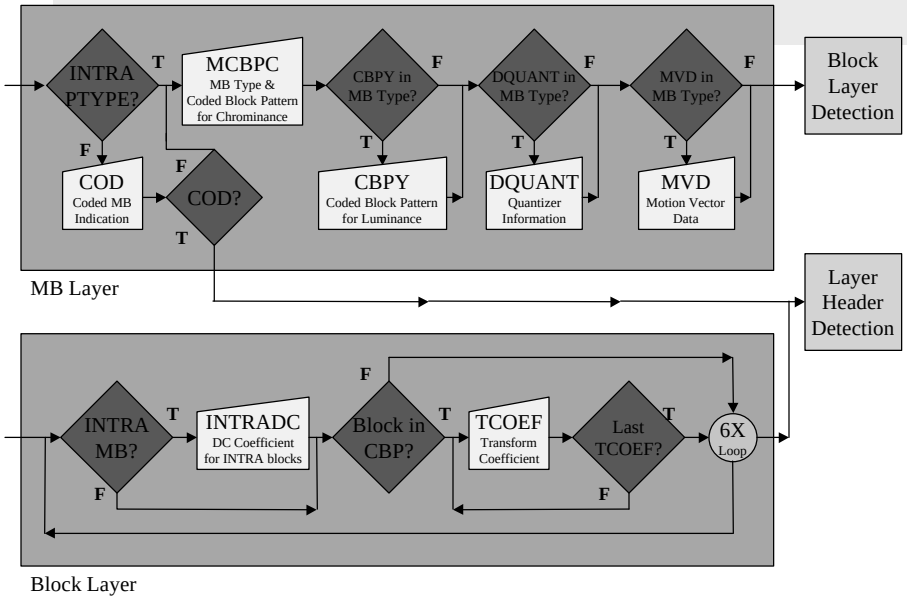
H.263 Baseline Interpretation




*Joe Huang*

## Decoding the Received Bitstream: Part II

H.263 Baseline Interpretation




*Joe Huang*

## H.26L

- Better quality and more functionalities
- Call for Proposals, Jan 1998
  - Very low bit rates, real-time, low end-to-end delay
  - Low complexity permitting software implementations
  - Enhanced error robustness, including mobile networks
  - Adaptable rate control mechanisms
  - Variety of source materials

18-796/Spring 1999/Chen

## Applications of H.26L

- Real-time conversational services
- Internet video applications
- Sign language and lip-reading communication
- Video storage and retrieval services (e.g. VOD)
- Video store and forward services (e.g. video mail)
- Multi-point communication over heterogeneous networks



## Tentative Time Schedule

| | |
|---|---|
| Jan 1998 | Call for proposals |
| Nov 1998 | Evaluation of the proposals |
| Jan 1999 | 1st Test Model of H.26L (TML1) |
| Nov 1999 | Final Major Feature Adoptions |
| Aug 2001 | Determination |
| May 2002 | Decision |

# H.263++

- Further extensions of H.263. Targeted for decision in Nov'2000
- Four KTAs
  - Data partitioning with reversible VLCs
  - 4×4 motion and DCT
  - Adaptive quantization
  - Long-term/Background frame store use



# References


- ITU-T Q.15/16, Gary Sullivan, ed., "Draft Text of Recommendation H.263 Version 2 ("H.263+") for Decision," Sept. 1997
- ftp://standard.pictel.com/video-site/
- Joan L. Mitchell et al., Sec. 19.3, *MPEG Video: Compression Standard*, Chapman & Hall, New York, NY
- Barry G. Haskell, Atul Puri, Arun N. Netravali, Sec 17.1, *Digital Video : An Introduction to MPEG-2*, Chapman & Hall, New York, NY

# Results of CMU H.263 Baseline

*Khalid Goudeaux, Joseph Huang, Kehua Jiang,*
*Ta-Chien Lin, Deepak Turaga*

## CMU H.263 Codec Result (1) - Akiyo_qcif

300 Frames Video

9,123,840 bits/s

L: Quant 4

R: Original

30 Hz: 99,552 bits/s

5 Hz: 36,807 bits/s

## CMU H.263 Codec Result (2) - Akiyo_qcif

L: Quant 8

R: Quant 4

30 Hz: 40,404 bits/s

5 Hz: 16,538 bits/s

L: Quant 16

R: Quant 4

30 Hz: 16,968 bits/s

5 Hz: 7,025 bits/s



## CMU H.263 Codec Result (3) - Akiyo_qcif

L: Quant 24

R: Quant 4

30 Hz: 11,188 bits/s

5 Hz: 4,495 bits/s

L: Quant 31

R: Quant 4

30 Hz: 9,506 bits/s

5 Hz: 3,576 bits/s

## CMU H.263 Codec Result (1) - news_qcif


300 Frames Video

9,123,840 bits/s

L: Quant 4

R: Original

30 Hz: 232,802 bits/s

5 Hz: 82,658 bits/s



## CMU H.263 Codec Result (2) - news_qcif


L: Quant 8

R: Quant 4

30 Hz: 104,955 bits/s

5 Hz: 40,359 bits/s

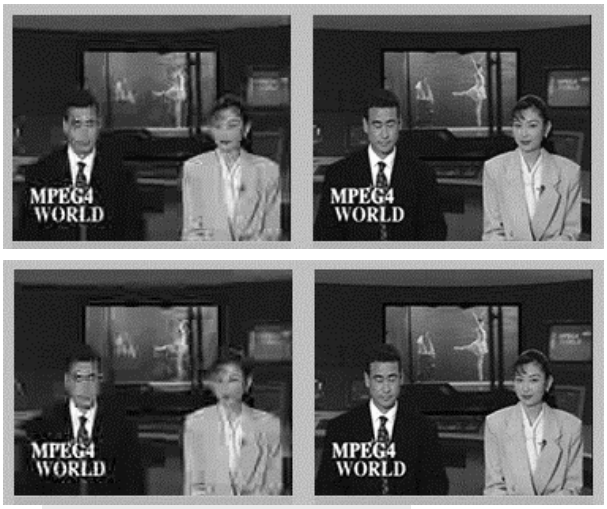

L: Quant 16

R: Quant 4

30 Hz: 44,680 bits/s

5 Hz: 18,333 bits/s

## CMU H.263 Codec Result (3) - news_qcif




L: Quant 24

R: Quant 4

30 Hz: 27,892 bits/s

5 Hz: 11,616 bits/s

L: Quant 31

R: Quant 4

30 Hz: 21,529 bits/s

5 Hz: 8,898 bits/s



## CMU H.263 Codec Result (1) - trevor_qcif

150 Frames Video

9,123,840 bits/s

L: Quant 4

R: Original

30 Hz: 332,698 bits/s

5 Hz: 118,248 bits/s

## CMU H.263 Codec Result (2) - trevor_qcif

L: Quant 8

R: Quant 4

30 Hz: 114,525 bits/s

5 Hz: 57,312 bits/s

L: Quant 16

R: Quant 4

30 Hz: 62,213 bits/s

5 Hz: 26,732 bits/s



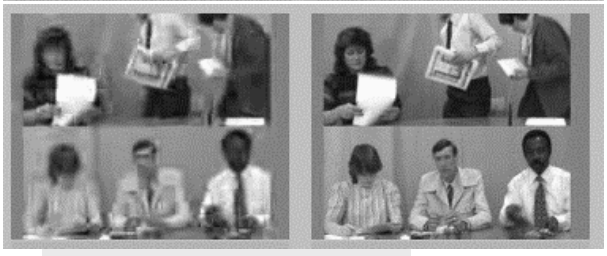
## CMU H.263 Codec Result (3) - trevor_qcif

L: Quant 24

R: Quant 4

30 Hz: 41,098 bits/s

5 Hz: 17,735 bits/s



L: Quant 31

R: Quant 4

30 Hz: 33,848 bits/s

5 Hz: 14,106 bits/s