BH001/2 COURSE: SYSTEM SOLUTION

MULTIMEDIA SYSTEMS AND VIDEO CODING TECHNIQUES

INSTRUCTOR: MINH N. TRAN
SOFTWARE ENGINEERING DIVISION
R-CAR SOFTWARE SOLUTION 2 – MIDDLEWARE 2

AGENDA

- ❑History of video coding
- ❑Overview: color space
- ❑Overview: video coding

- ❑Overview: MPEG-1 video codec
- ❑Overview: MPEG-2 video codec
- ❑Overview: H.264/AVC video codec
- ❑Overview: Multimedia SoC and software layers

HISTORY OF VIDEO CODING

MOVING PICTURE CODING

- **Moving picture**, also called **film** or **movie** or **motion picture**, is *a series of still-images* which, when shown on a screen, creates the illusion of moving images due to the **phi phenomenon**.

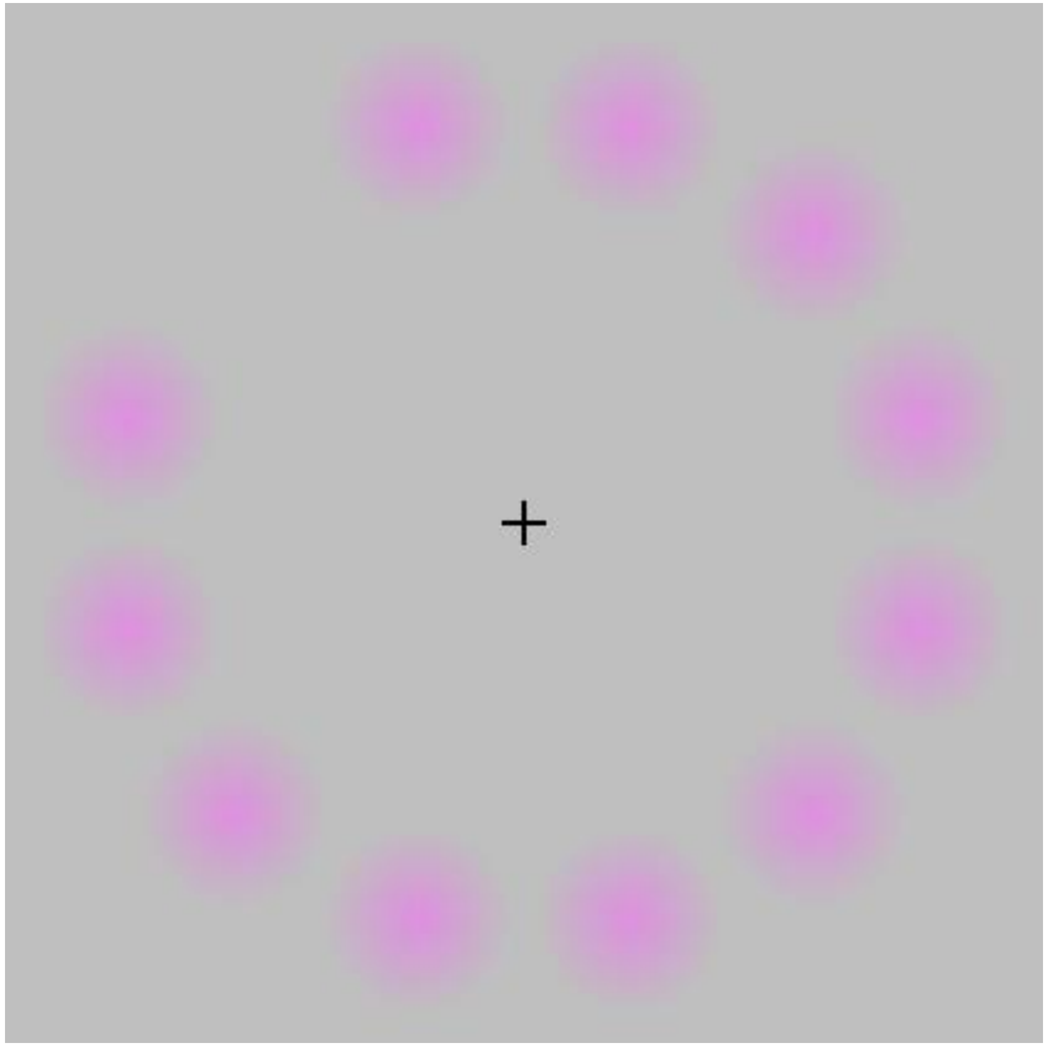
- **Coding** is a term from **coding theory** which may refer to either *data compression* or *error-correcting codecs* or *cryptography* or *line coding*.

- Within the scope in this training material, **moving picture coding** refers to **data compression** techniques that can be applied on **a series of still-images**.

- **CODEC** is a short term of “en**co**de” and “**dec**ode”. “Encode” is similar to “compress” and “decode” is similar to “decompress”. Encrypt and decrypt are sound familiar here but actually off-topic.

MOVING PICTURE CODING: PHI PHENOMENON

- **Phi phenomenon**: a series of still-images is displayed one-by-one in a short period to represent a circle movement.
- A still-image represents a *nick of time* or a *brief moment* in the time series.
- There is *nothing moving*, but human eyes and brain captured the picture information and then be *under an illusion that the circle is moving*.



(An animation image to represent Phi Phenomenon)

BIG IDEAS FOR EVERY SPACE

MOVING PICTURE CODING ➔ STANDARDIZATION

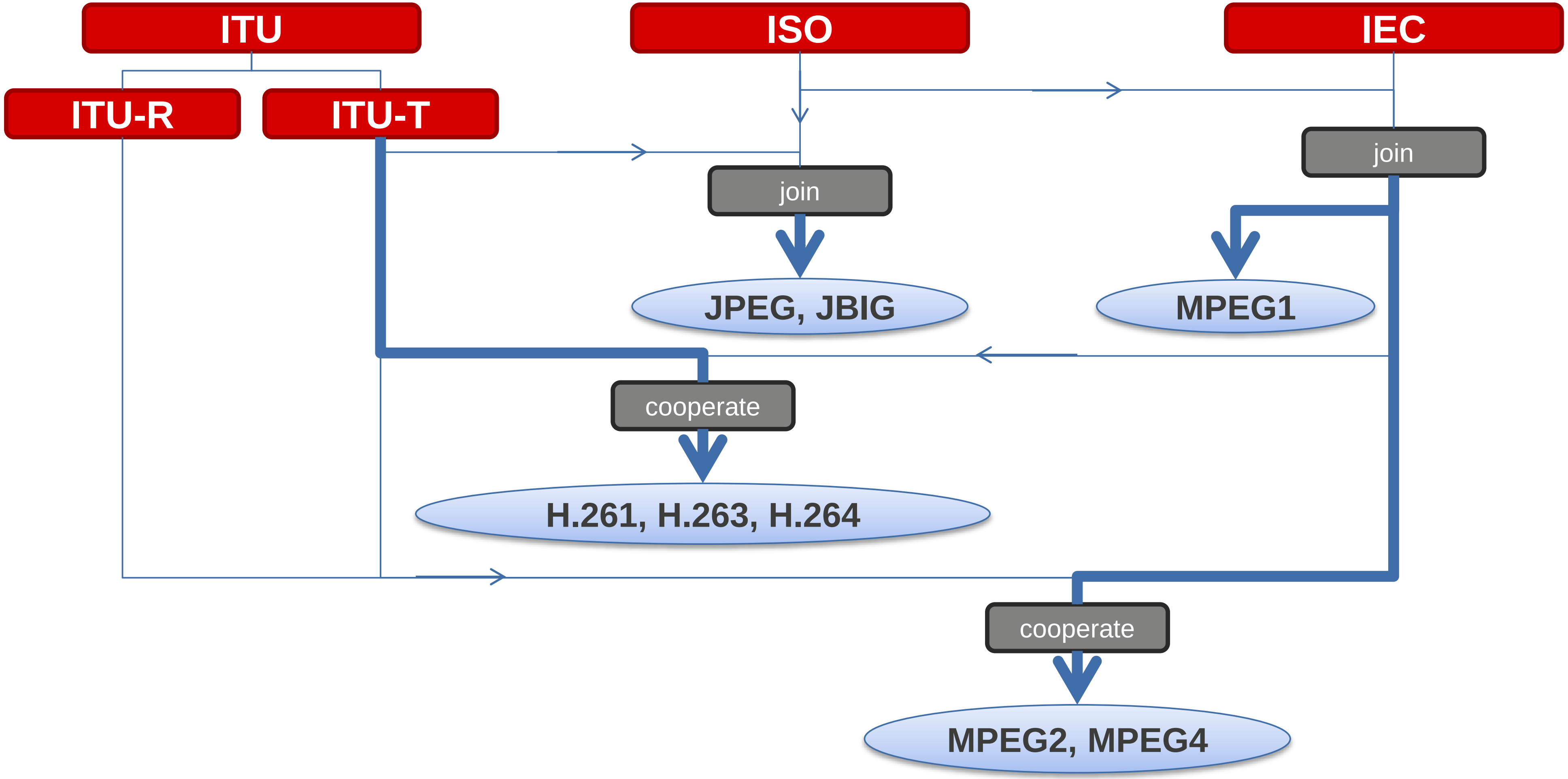
- There are thousands different technical methods to conduct "**coding**" in moving picture area. ➔ Struggling to choose, huh?

- When moving picture area become popular, it requires a **standardization** in coding techniques. The standard does not necessarily represent the best technical solution, but rather attempt to achieve a compromise between various requirements.

- However, there are different regions and countries where *a standard cannot be applied for some reasons*. Still a lot of moving picture coding standards are available in the world, but only few famous ones are being widely supported.

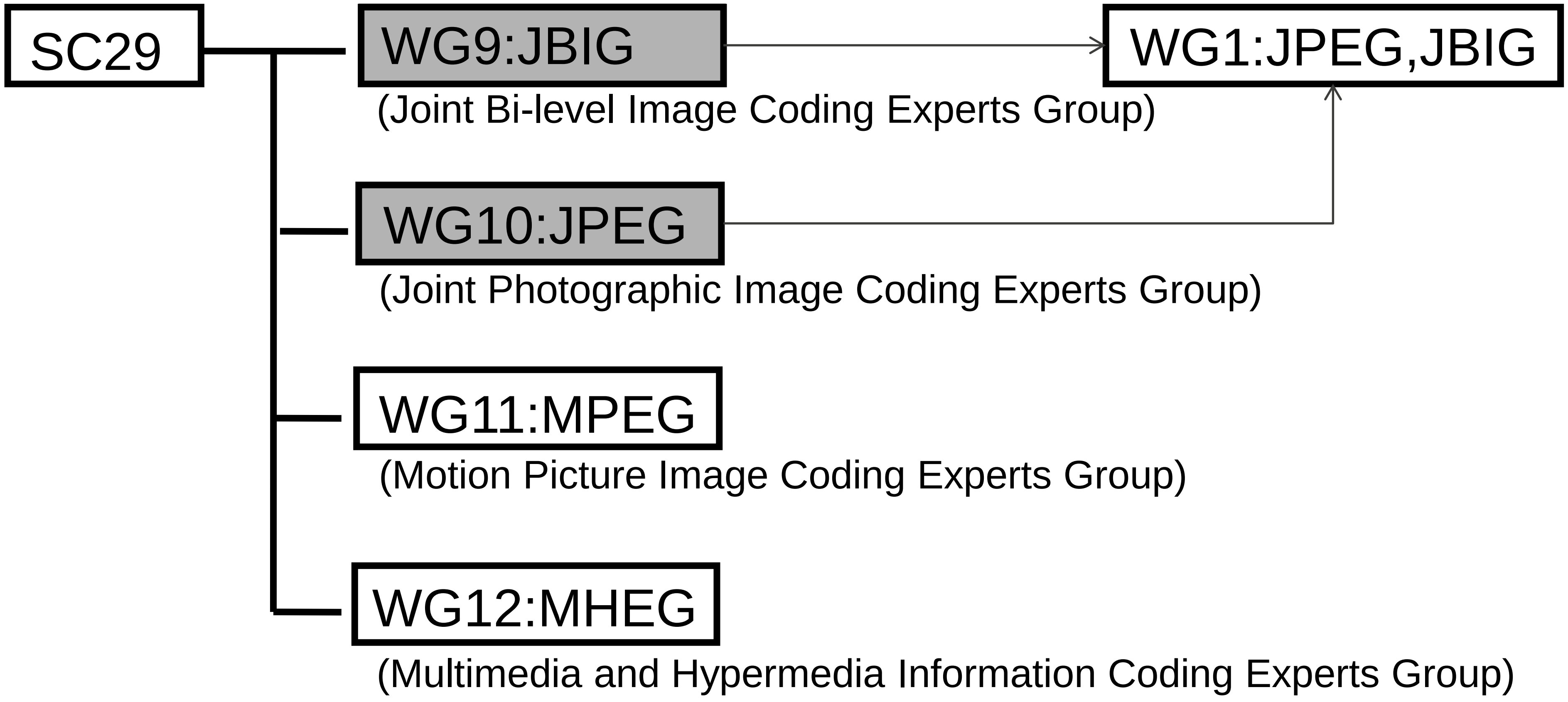
MOVING PICTURE CODING → STANDARDIZATION

- The standards are established among various organizations

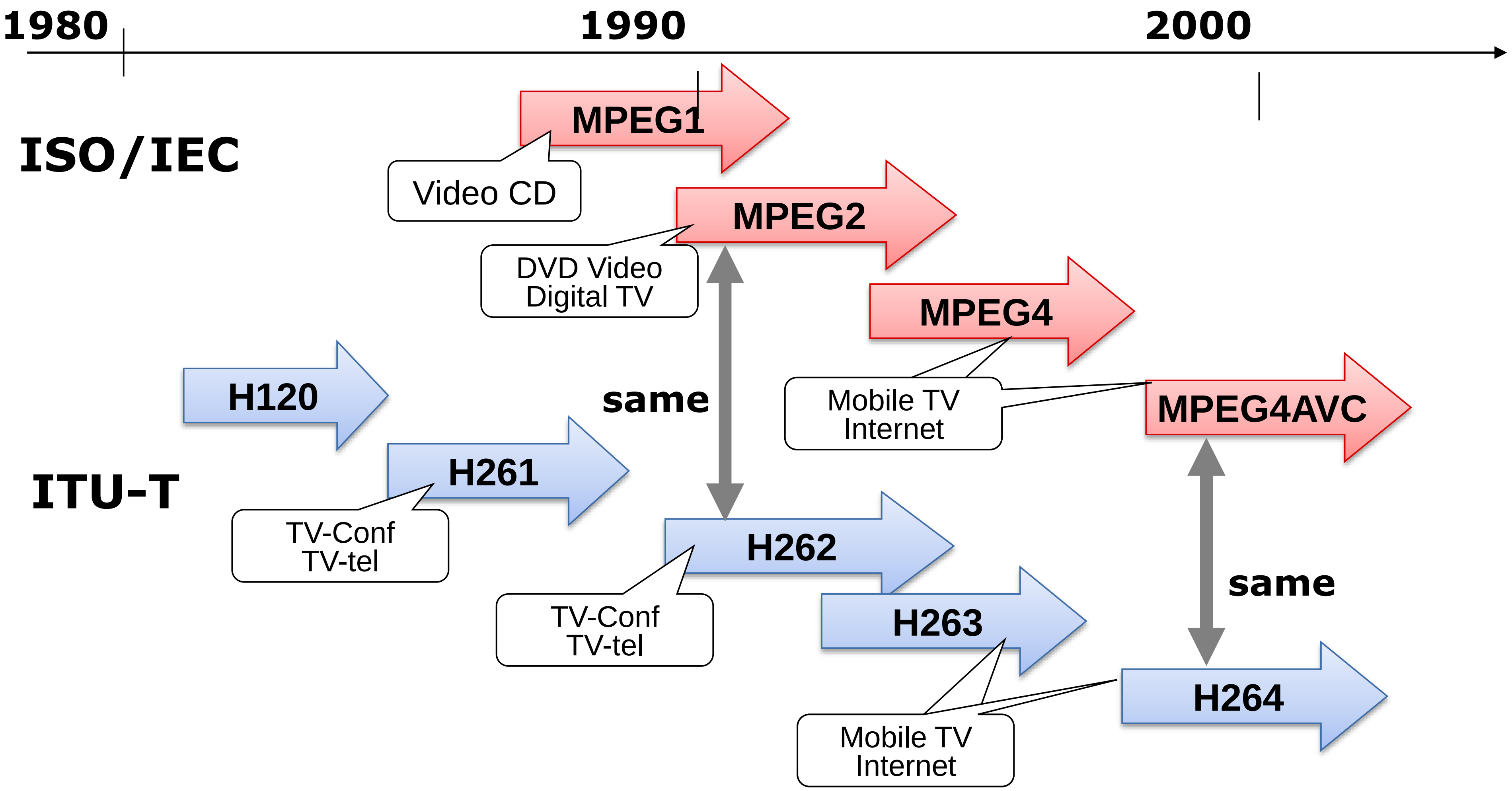


MOVING PICTURE CODING → STANDARDIZATION

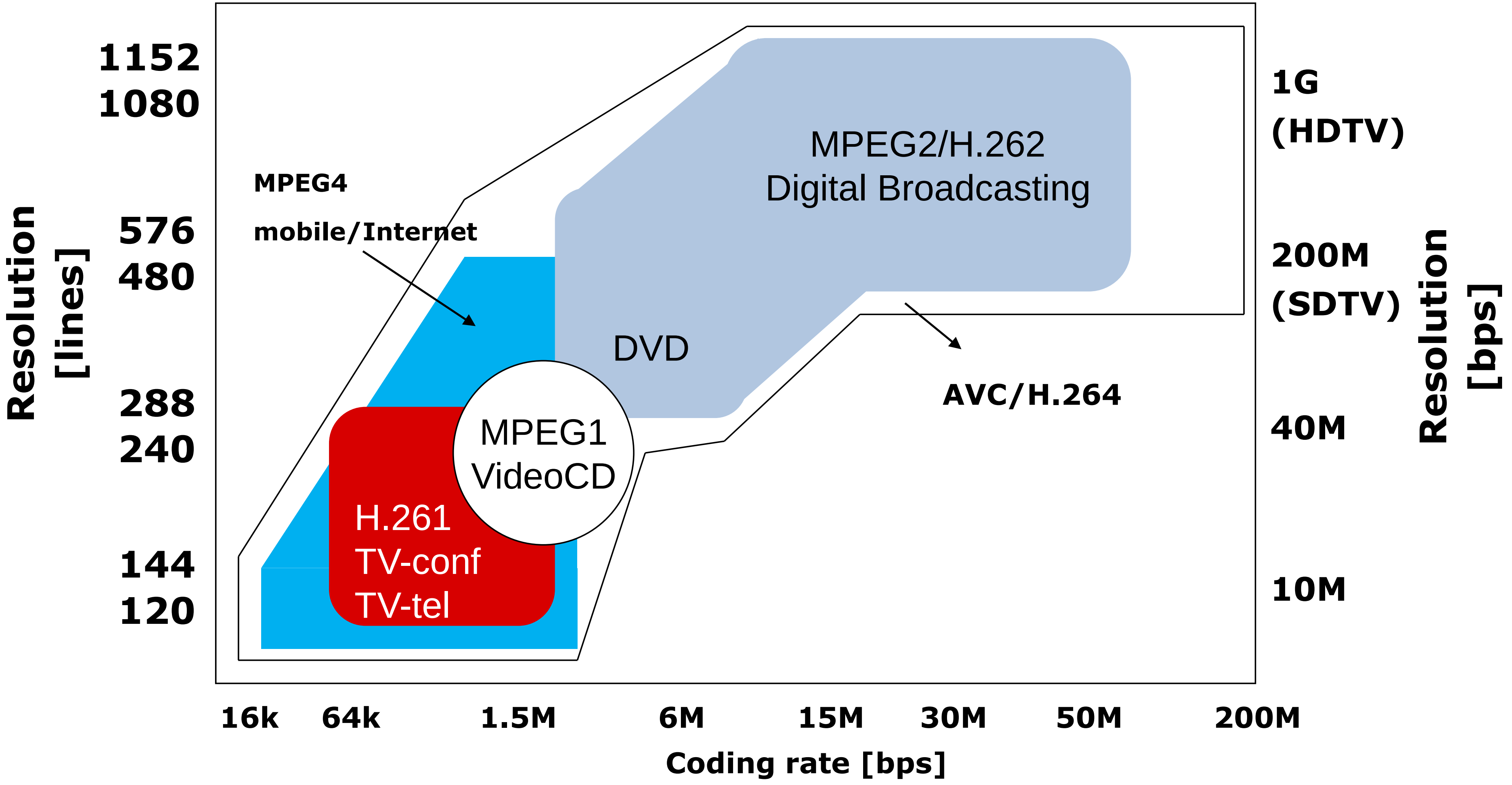
ISO/IEC JTC1



MOVING PICTURE CODING ➔ STANDARDIZATION

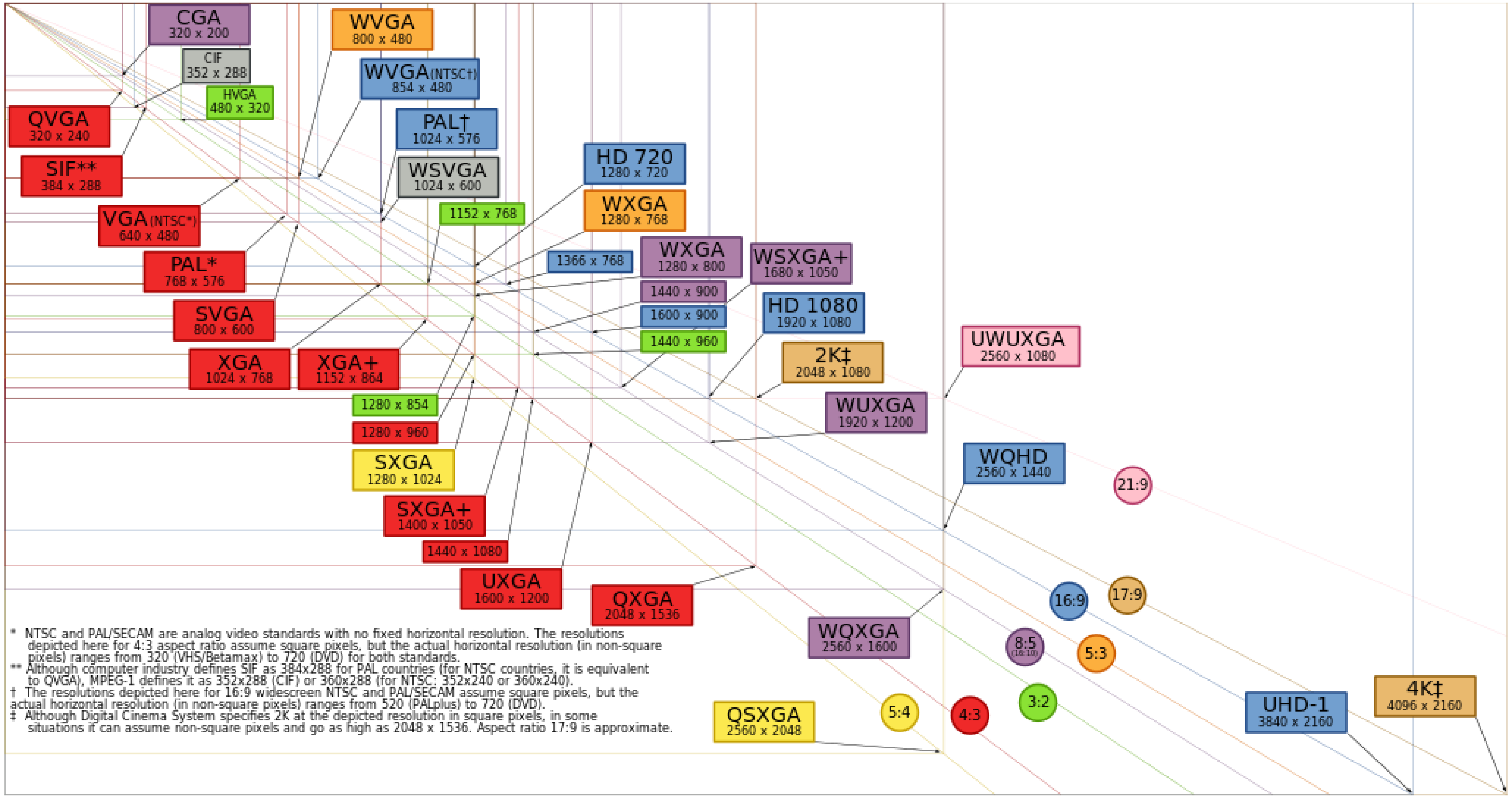


MOVING PICTURE CODING ➔ STANDARDIZATION



BASIC TERMS: FRAME RESOLUTION

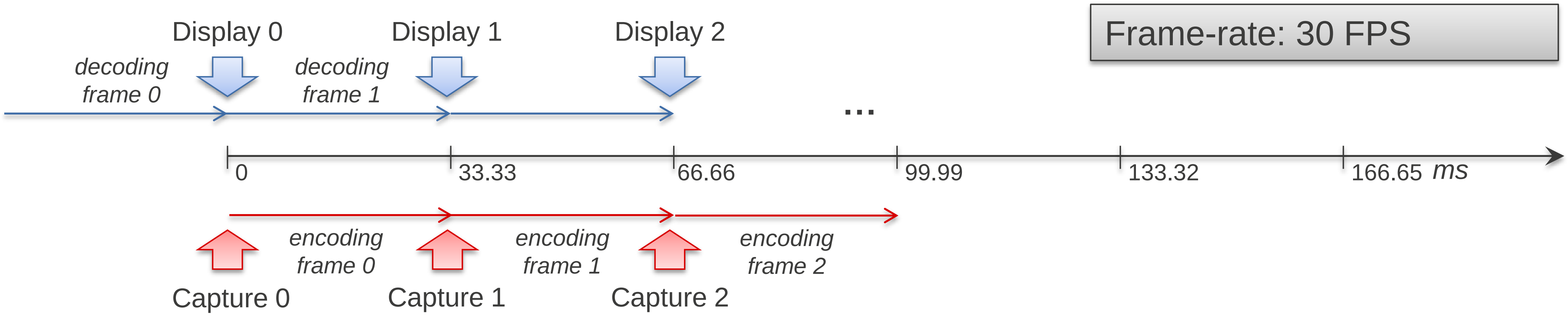
- Number of pixels count in digital image that is defined by two integer numbers: pixel columns (**width**) and pixel rows (**height**).



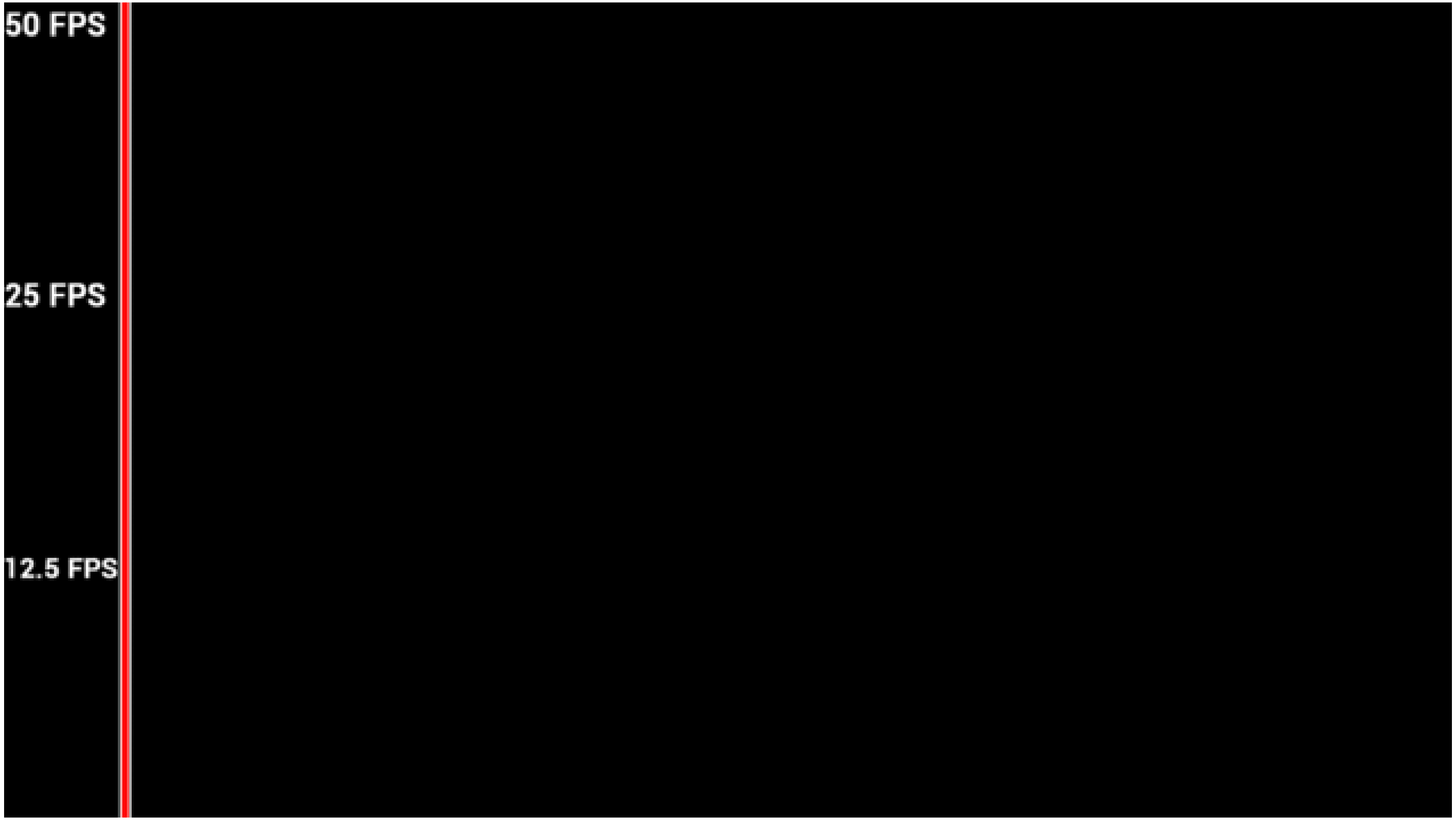
BASIC TERMS: FRAME RATE (FPS)

- The frequency at which an image device produces unique consecutive images called frames. The rate is expressed as either **frames per second** (FPS) or Hertz (Hz). In video coding, frame-rate usually refers to FPS which means the number of frames that system can build per second (24 FPS, 25 FPS, 30 FPS, etc.)

- For smoothly video processing, frames to be captured/displayed should be ready before/after a specific timestamp.



BASIC TERMS: FRAME RATE (FPS)



(An animation image to represent different frame rate visualizations)

BASIC TERMS: BITRATE (BPS)

- A number of bits that are processed per unit of time. Bitrate is either variable bitrate (VBR) or constant bitrate (CBR).

- In video coding, bitrate refers to the **number of bits used per unit of playback time** to represent a continuous video playback after source coding.
  - It also means the **required data transfer speed** to avoid interrupt of playback, for example: Youtube 720p (2.5 Mbps), DVD (9.8 Mbps), Blu-ray 1080p (40 Mbps)

- Bitrate decides the **quality of frame to be displayed**. If the same source is encoded into different bitrate streams, the higher bitrate stream will provide better quality of decoded frames than the lower bitrate stream.

BASIC TERMS: BITRATE (BPS)

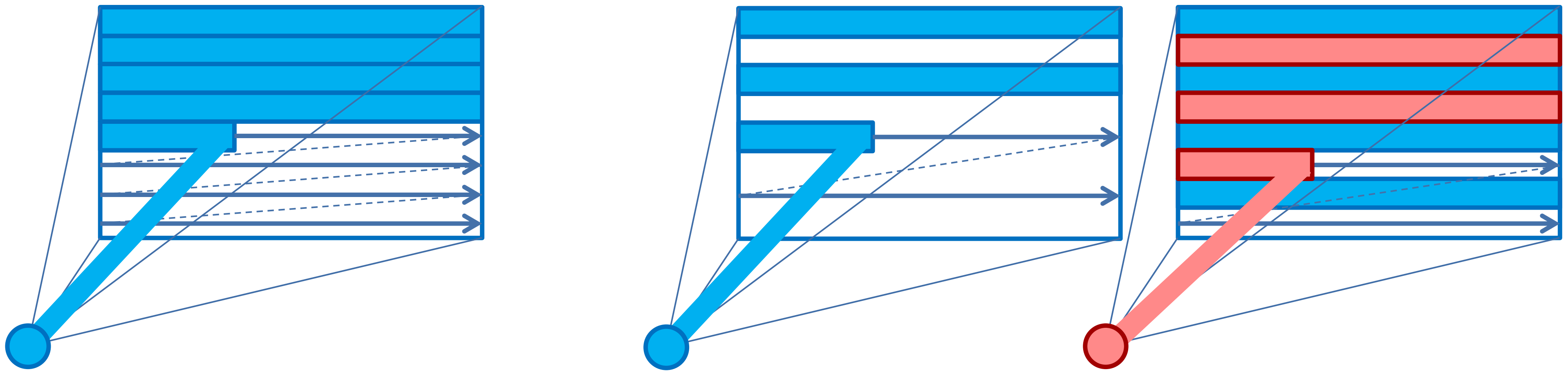
High bitrate

Low bitrate

BASIC TERMS: VIDEO SCANNING

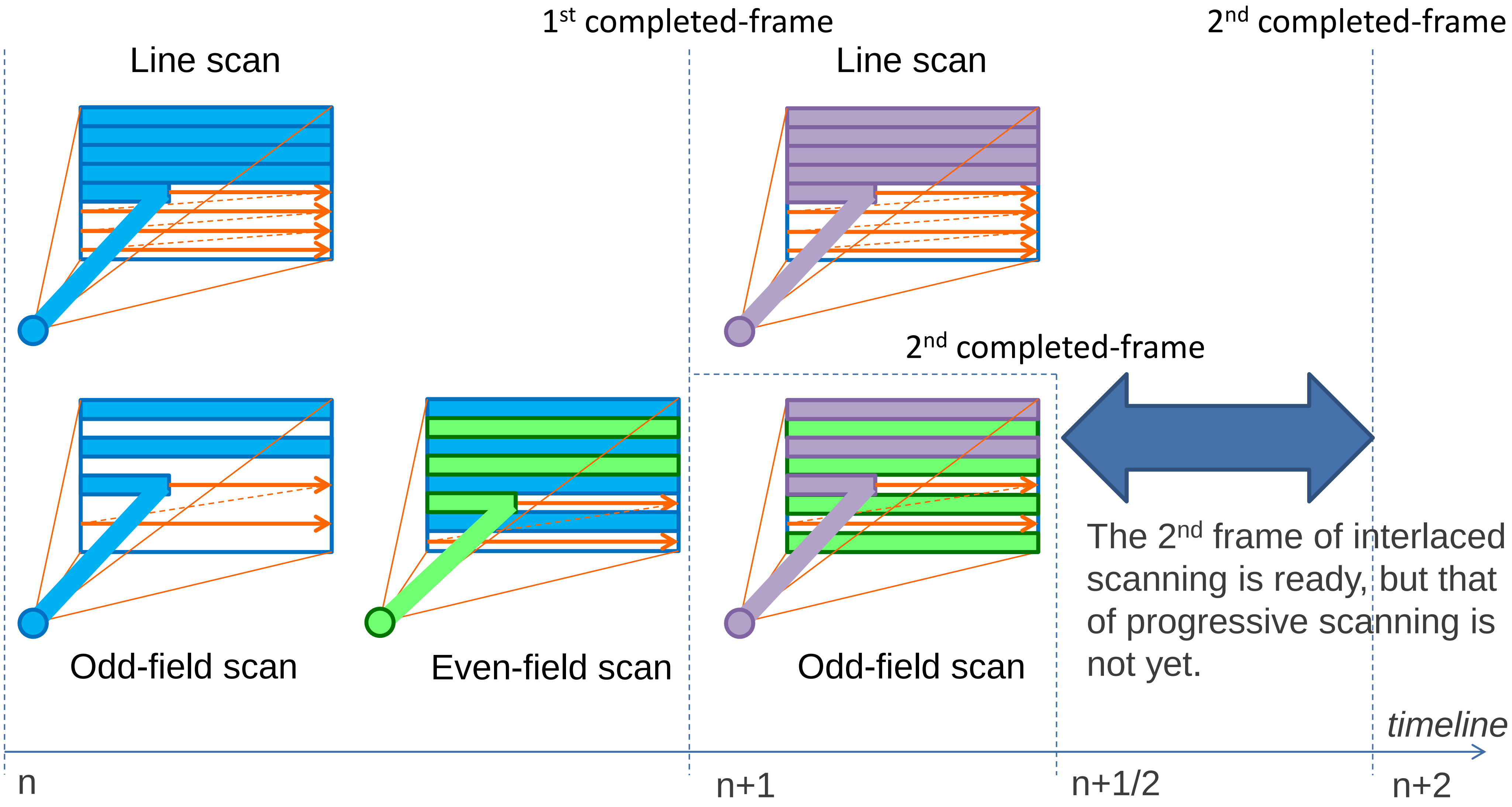
- As the video information is actually 2 dimensional, video signal scans the 2-dimensional information
- There 2 types of scanning: **Progressive Scan** and **Interlaced Scan**

- Progressive (non-interlaced) scan
  - It scans 2 dimensional area progressively
- Interlaced scan
  - it scans 2 dimensional area alternatively



Line scan　　Odd-field scan　　Even-field scan

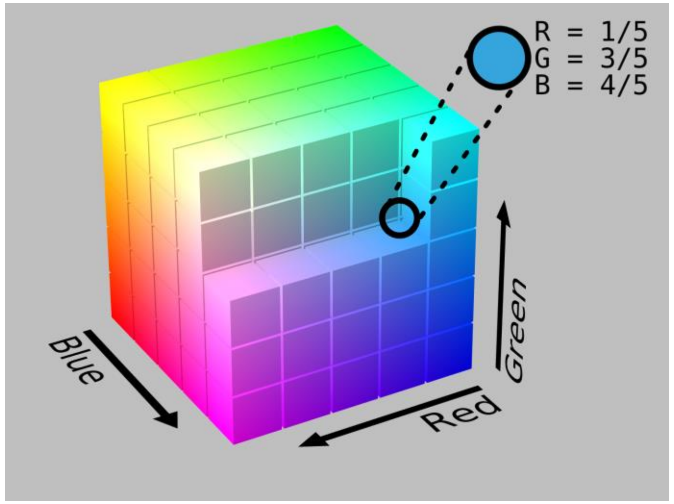
BASIC TERMS: VIDEO SCANNING



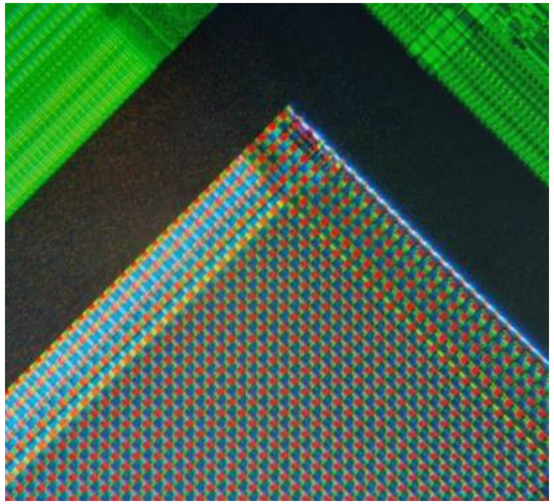
OVERVIEW: COLOR SPACE

COLOR SPACE: RGB

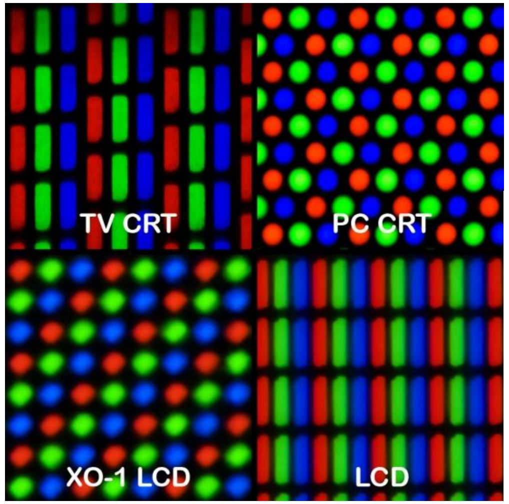
- Color is a chromaticity combination of the **Red**, **Green**, and **Blue**.
- Almost input sources (camera, etc.) provide RGB data naturally.
- Almost output sources (display, etc.) receive RGB data naturally.
- A (digital) pixel represents any color combined from the elements **Red**, **Green**, and **Blue**. If one element is 8-bit depth, then one pixel is 24-bit depth.
- **Compression on RGB data pixels** is **less effective** in general.



Chromaticity combination
(Source from Wikipedia)



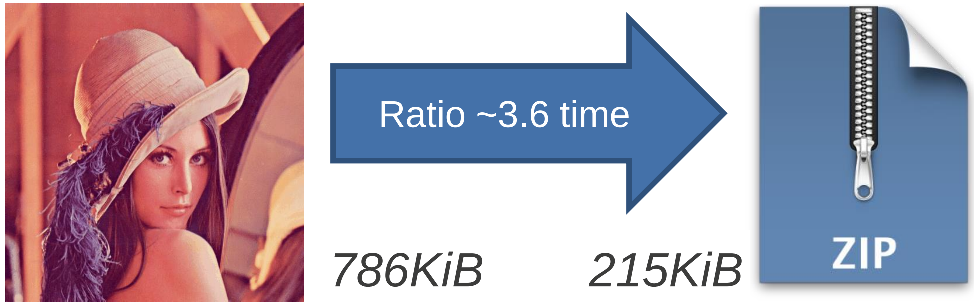
Pixels on camera devices
(Source from Wikipedia)



Pixels on display devices
(Source from Wikipedia)

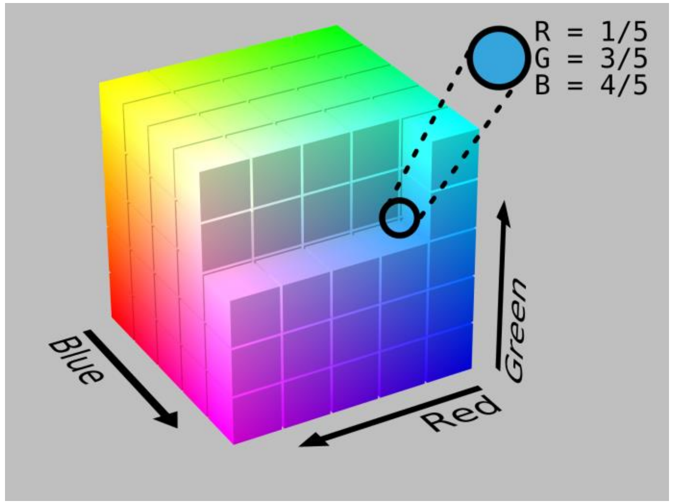
COLOR SPACE: RGB

- Color is a chr
- Almost input s ally.
- Almost output rally.
- A (digital) pixe nents Red, Green, and Blue. If one e epth.

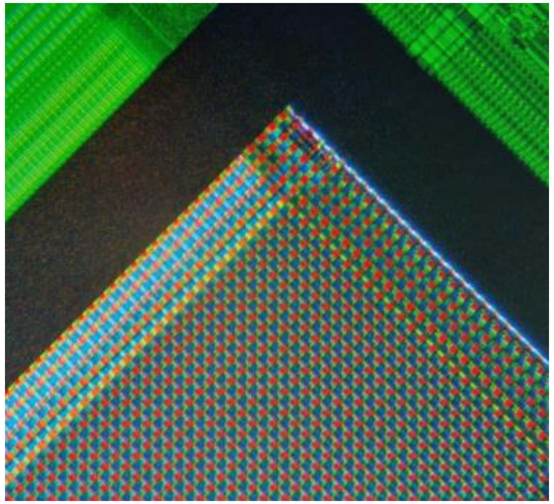


Original (Lenna) Compressed

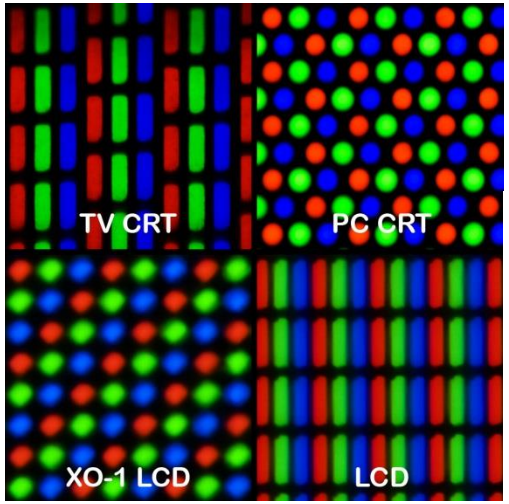
- **Compression on RGB data pixels** is **less effective** in general.



Chromaticity combination (Source from Wikipedia)



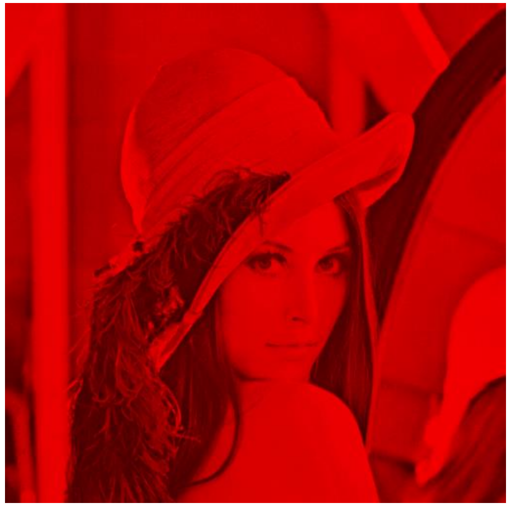
Pixels on camera devices (Source from Wikipedia)

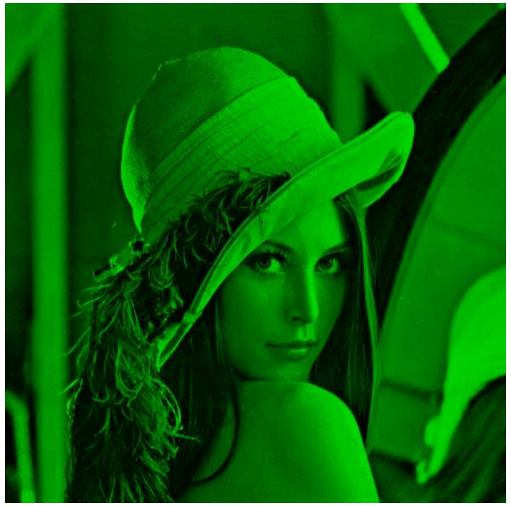


Pixels on display devices (Source from Wikipedia)

COLOR SPACE: RGB


Original (Lenna)


Red


Green


Blue

Each color channel contains both **luma** that helps to detect picture information and **chroma** that helps to detect color information by human-eyes.


Luma on grayscale

COLOR SPACE: RGB

Original (Lenna)

Red

Green

Blue

Each color channel contains both **luma** that helps to detect picture information and **chroma** that helps to detect color information by human-eyes.

Luma on grayscale

Scientists proved that human-eyes are more sensitive on **luma** differences than **chroma** differences. <u>Then it’s possible to reduce the **chroma** and retain the **luma**.</u>

That sounds great!

COLOR SPACE: RGB

Original (Lenna)

Red

Green

Blue

Each color channel contains both **luma** that helps to detect picture information and **chroma** that helps to detect color information by human-eyes.

Luma on grayscale

Scientists proved that human-eyes are more sensitive on **luma** differences than **chroma** differences. Then it's possible to reduce the **chroma** and retain the **luma**.

However, in RGB color space, reducing **chroma** will also lead to reducing **luma**. Hence it's not effective to do so with RGB.

Reducing the chroma and retaining the luma…

REDUCING THE CHROMA AND RETAINING THE LUMA…

Reducing the chroma retaining the luma…

Reducing the chroma and retaining the luma…

Reducing the chroma and retaining the luma…

…

WE NEED TO WORK WITH ANOTHER COLOR SPACE…

Reducing the chroma retaining the luma…

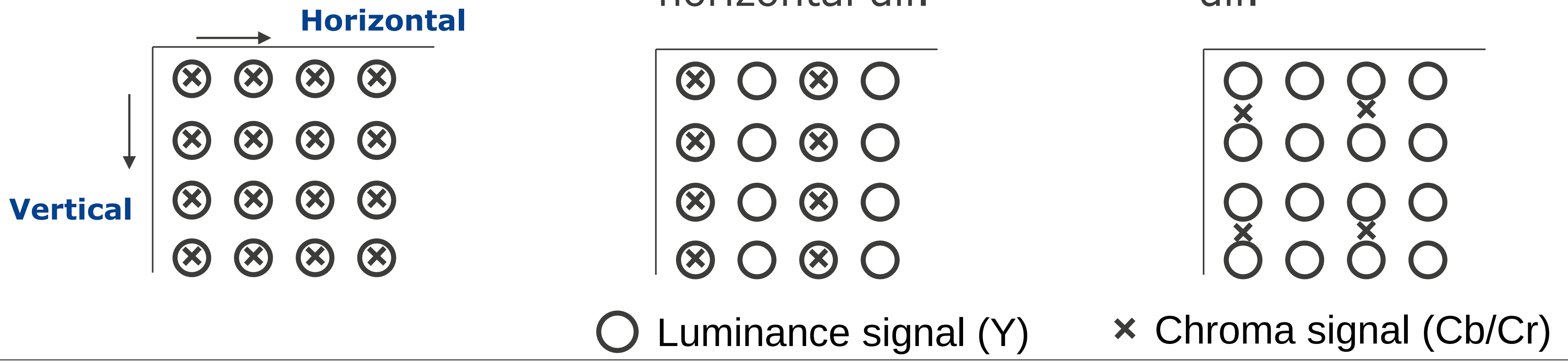
COLOR SPACE: YUV

- YCbCr color space separates the luminance from the color information. YUV is another name of this “color space” format.
- YCbCr can compress data by representing less resolution for chroma information than for luma information, taking advantage of the human visual system's less sensitive for color differences than for luminance.
- Y is the luma component and Cb and Cr are the blue-difference and red-difference chroma components.
- (We also have Cg which is the green-difference. But Cg component can be reconstructed from Y and Cb and Cr components or vice versa so that YCbCr is widely used instead of YCbCg or YCrCg or else.)

COLOR SPACE: YUV

- On the color information, there are 3 types of widely used YCbCr format which are 4:4:4, 4:2:2, and 4:2:0. The number means the sampling frequency compared to color burst signal.

- (4:4:4) has same data density as RGB, while (4:2:2) has 2/3, and (4:2:0) has 1/2 of data density as RGB, respectively
  - **4 : 4 : 4**
    - Each pixel has Y and Cb, Cr data
  - **4 : 2 : 2**
    - Chroma have only half sample in horizontal dir.
  - **4 : 2 : 0**
    - Chroma have only half samples in both dir.



COLOR SPACE: YUV


Original (Lenna)

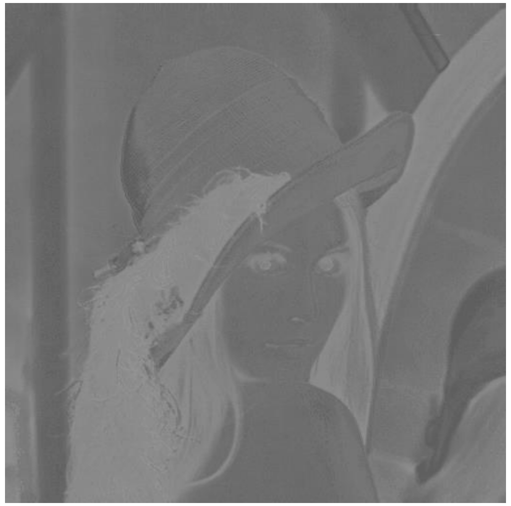

Red


Green


Blue

Both Blue-difference and Red-difference channels contain **chroma** data and are **not as detail as** Red, Green, Blue or **luma** channels.


Luma on grayscale


Blue-difference


Red-difference

COLOR SPACE: YUV

Original (Lenna)

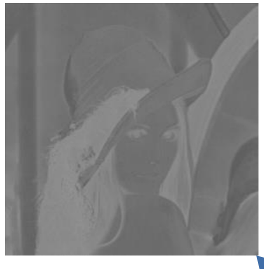
Red

Green

Blue

Both Blue-difference and Red-difference channels contain **chroma** data and are **not as detail as** Red, Green, Blue or **luma** channels.

Luma on grayscale



Blue-difference (1/4)

Red-difference (1/4)

Reducing **chroma** does not affect much on human-eyes.

COLOR SPACE: YUV

- Color conversion is defined by ITU-R BT 601 recommendation.

- Range of Y Cb Cr

Y (luminance)	:[16..235]
Cb (chrominance)	:[16..240] (center is 128)
Cr (chrominance)	:[16..240] (center is 128)

- Range of RGB (RGB are compensated by Gamma)

R (red)	:[0..255]
G (green)	:[0..255]
B (blue)	:[0..255]

- Conversion equations:

$$Y = 0.257 \times R + 0.504 \times G + 0.098 \times B + 16$$
$$Cb = -0.148 \times R - 0.291 \times G + 0.439 \times B + 128$$
$$Cr = 0.439 \times R - 0.368 \times G - 0.071 \times B + 128$$
$$R = 1.164 \times (Y-16) + 1.596 \times (Cr-128)$$
$$G = 1.164 \times (Y-16) - 0.813 \times (Cr-128) - 0.392 \times (Cb-128)$$
$$B = 1.164 \times (Y-16) + 2.017 \times (Cb-128)$$

COLOR SPACE: YUV (EXAMPLE)

- Calculate one TV picture data volume with following conditions:
  - Picture size : 720 [pixels/line] x 480 [lines/frame]
  - Color format : (4:4:4)
  - Data depth: 8 bits for each Y/Cb/Cr

    720[pixels/line]*480[lines/frame]*8[bits/pixel]*3 = 720*480*24[bits/frame]

- Calculate data rate of TV signal with conditions above and following:
  - Frame rate : 30 [Frames/sec]

    720*480*24[bits/frame]*30[frames/sec] = 720*480*24*30[bps]

- Calculate TV picture data volume with conditions above but Color format is (4:2:0)

    720[pixels/line]*480[lines/frame]*8[bits/pixel]*3*1/2 = 720*480*12[bits/frame]

- Calculate data rate of TV signal with conditions above (4:2:0) and following:
  - Frame rate : 30 [Frames/sec]

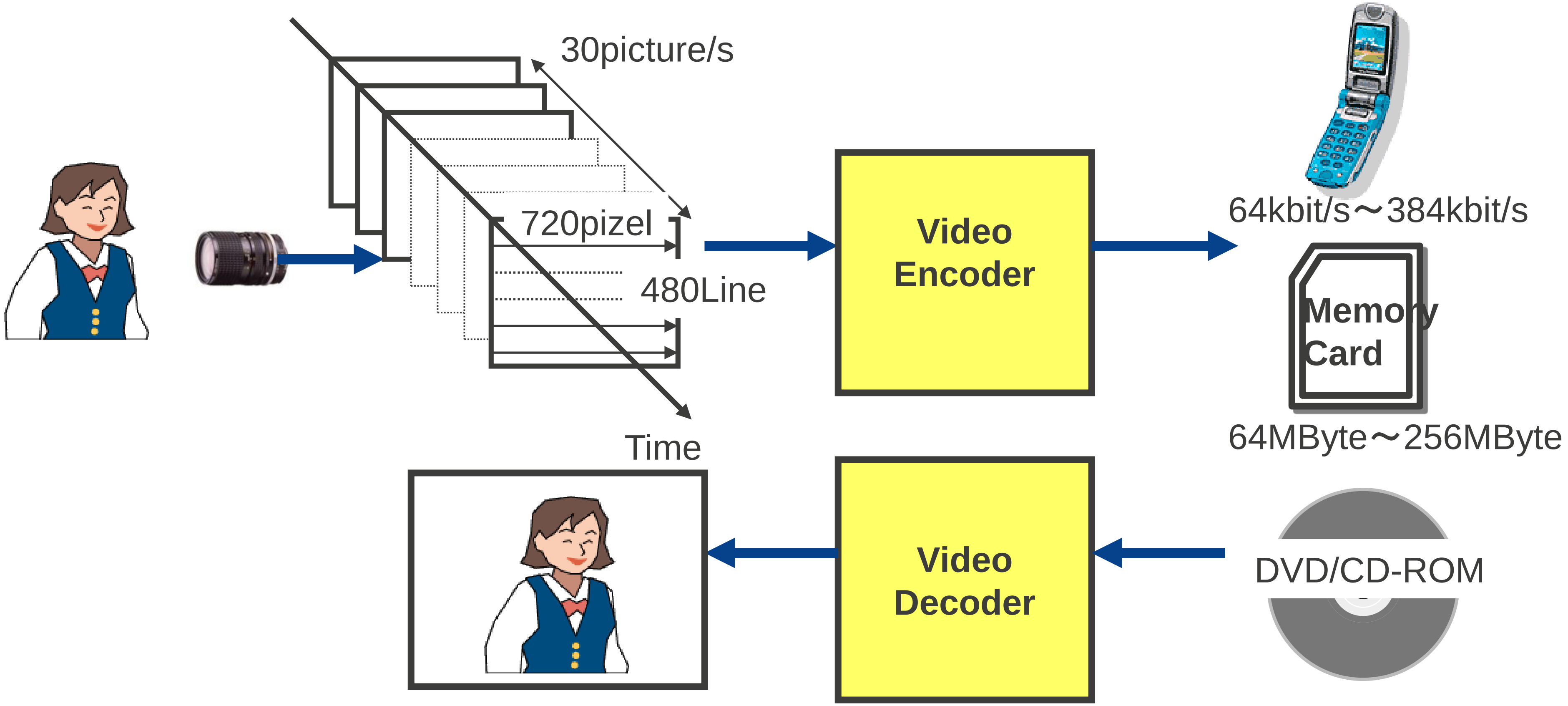
    720*480*12[bits/frame]*30[frame/sec] = 720*480*12*30[bps]

OVERVIEW: VIDEO CODING

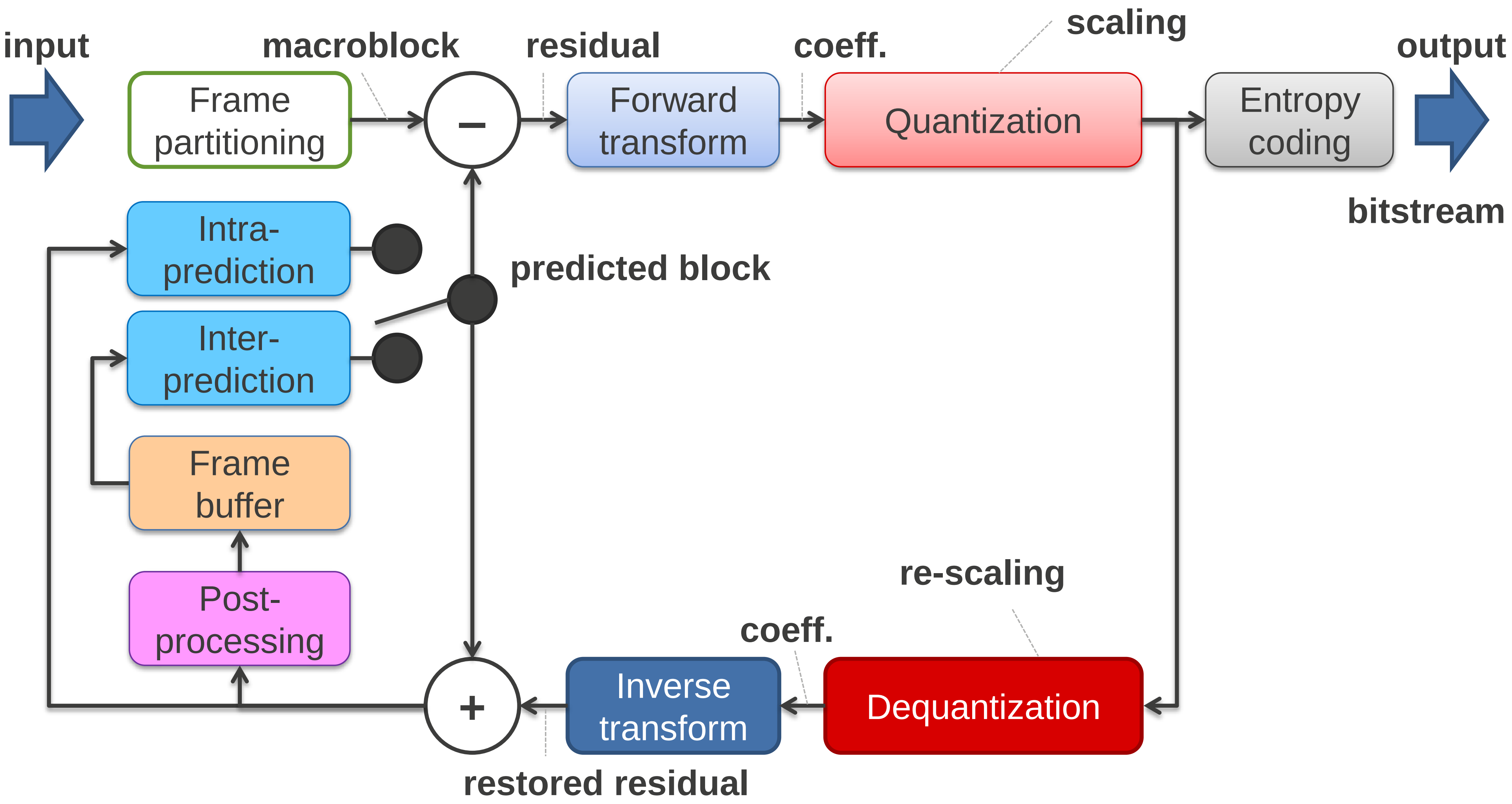
BASIC VIDEO CODING TECHNOLOGY

❑ Importance of Video coding

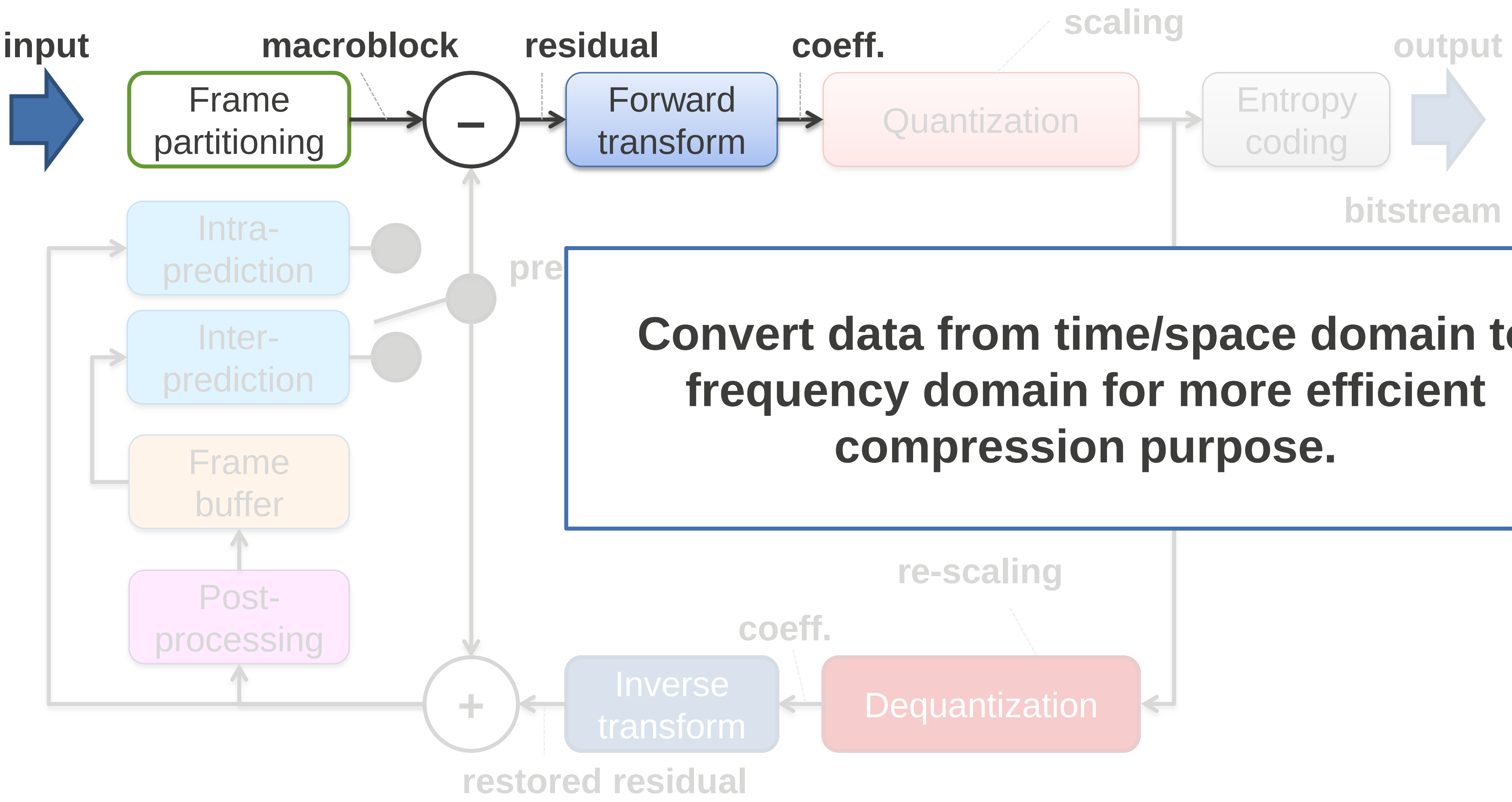
- Original data rate : (720pixel)x(480Line)x(24bit/RGB)x(30picture/s) = **249Mbit/s**
- Traffic data rate : Digital Terrestrial Broadcasting (**192 k bit/s**), DVD (**3～8Mbit/s**)



HYBRID BLOCK-BASED ENCODING FLOW

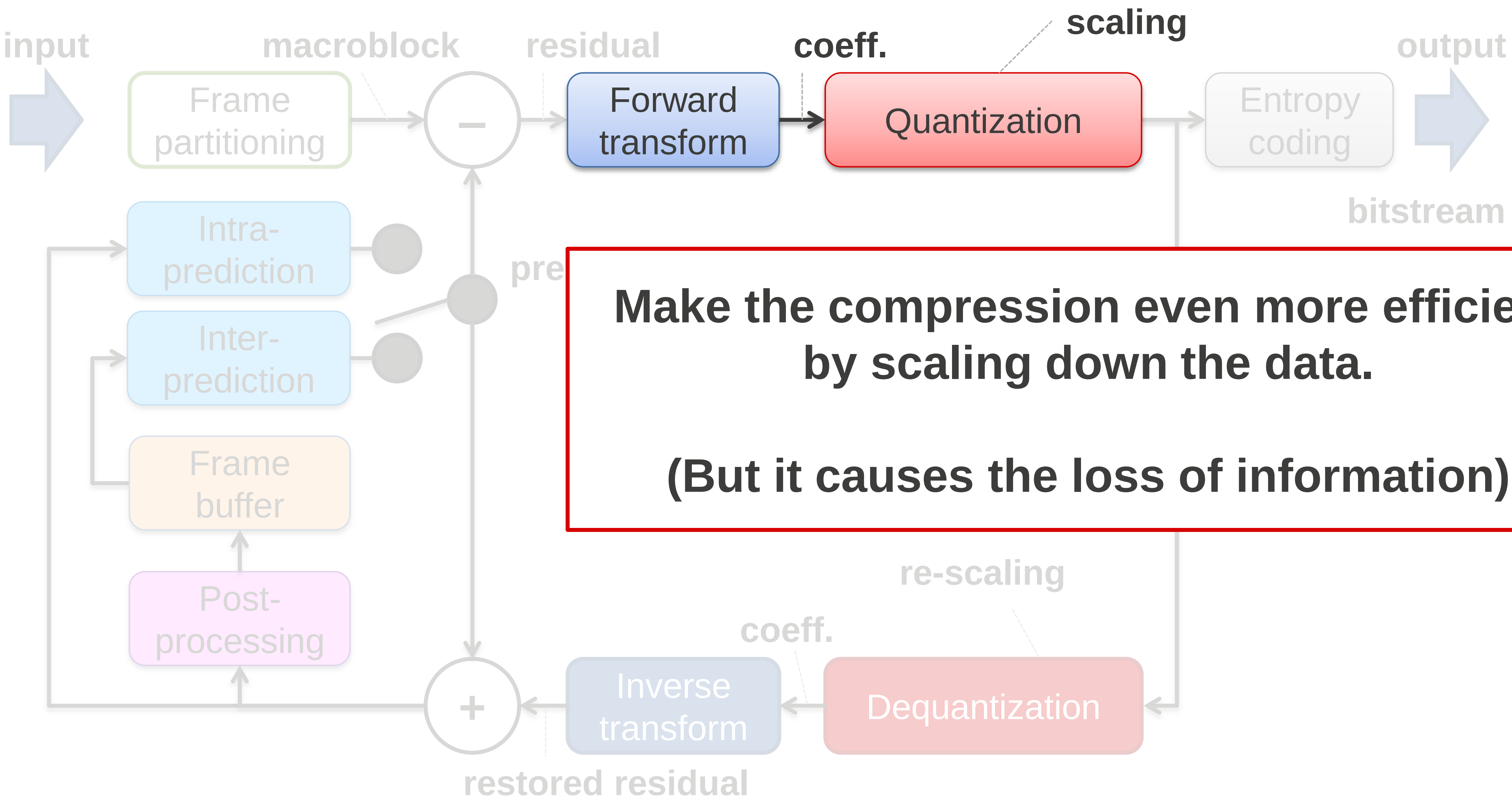


HYBRID BLOCK-BASED ENCODING FLOW



Convert data from time/space domain to frequency domain for more efficient compression purpose.

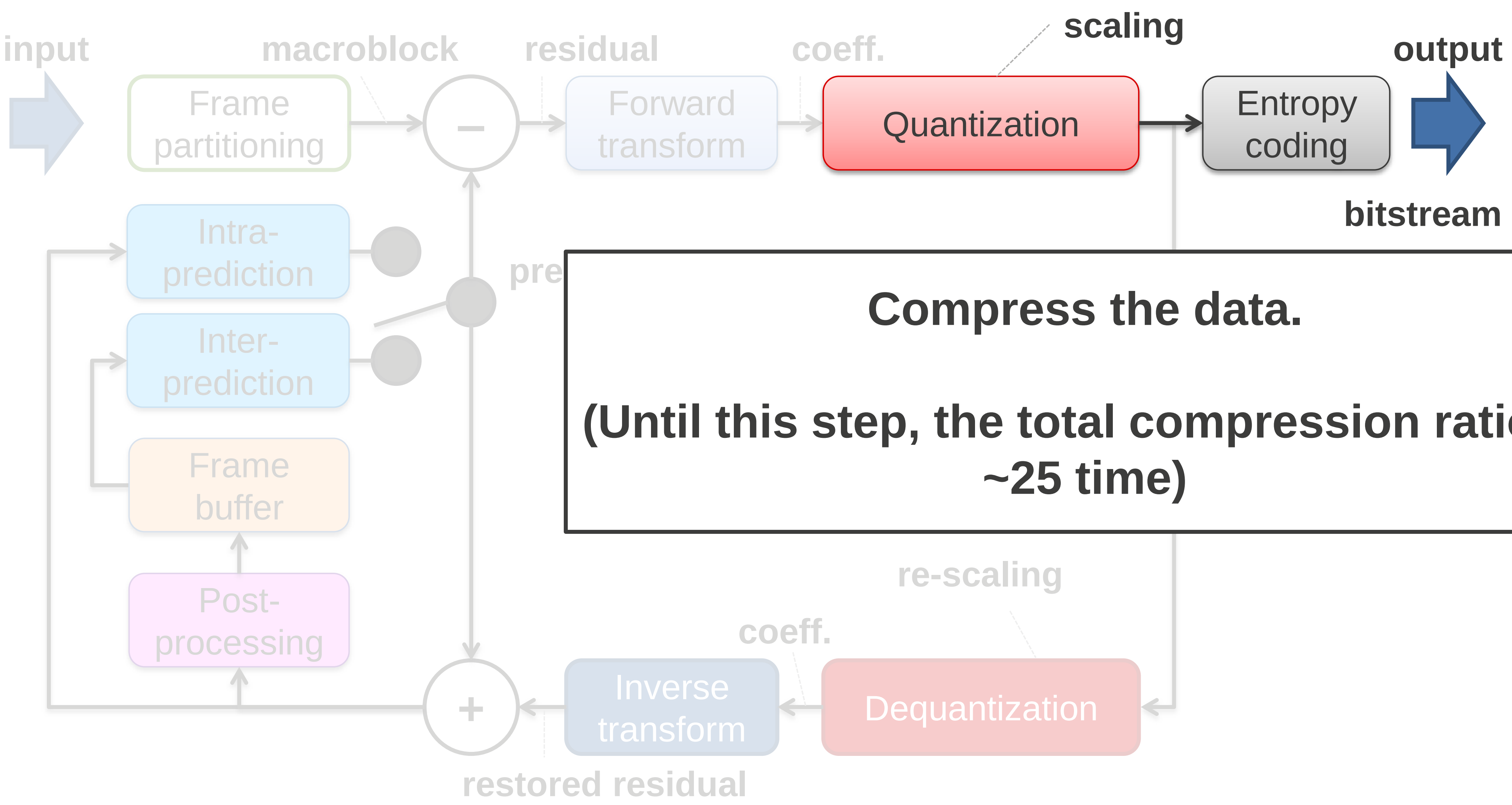
HYBRID BLOCK-BASED ENCODING FLOW



Make the compression even more efficient by scaling down the data.

(But it causes the loss of information)

HYBRID BLOCK-BASED ENCODING FLOW

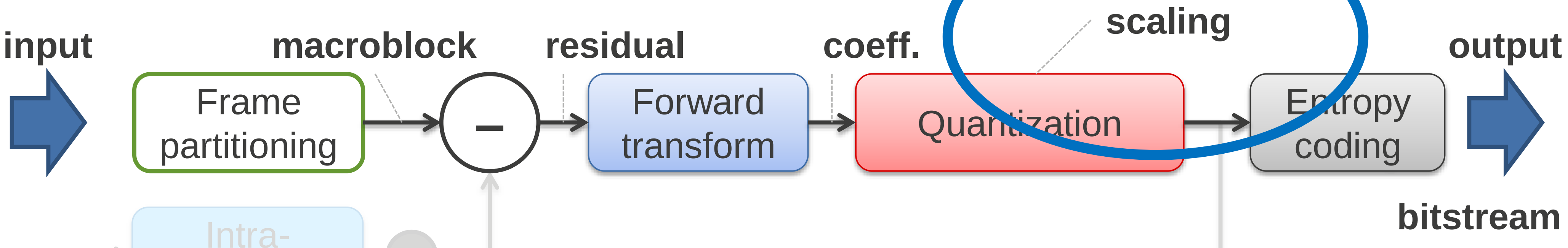


Compress the data.

(Until this step, the total compression ratio is ~25 time)

HYBRID BLOCK-BASED ENCODING FLOW

It's better than normal compression because of **this**.



Compress the data.

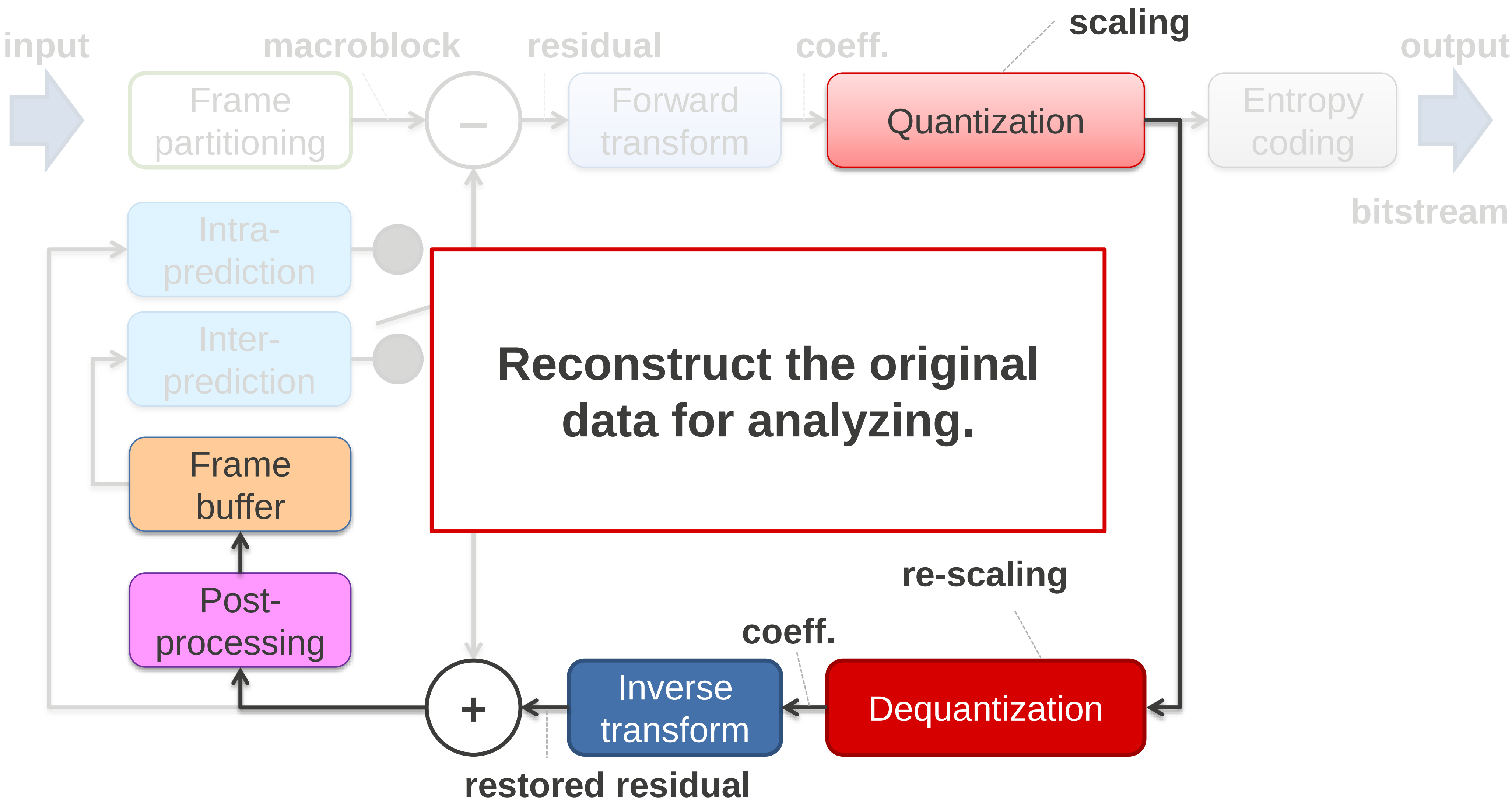
(Until this step, the total compression ratio is ~25 time)

Ratio ~3.6 time

786KiB *215KiB*

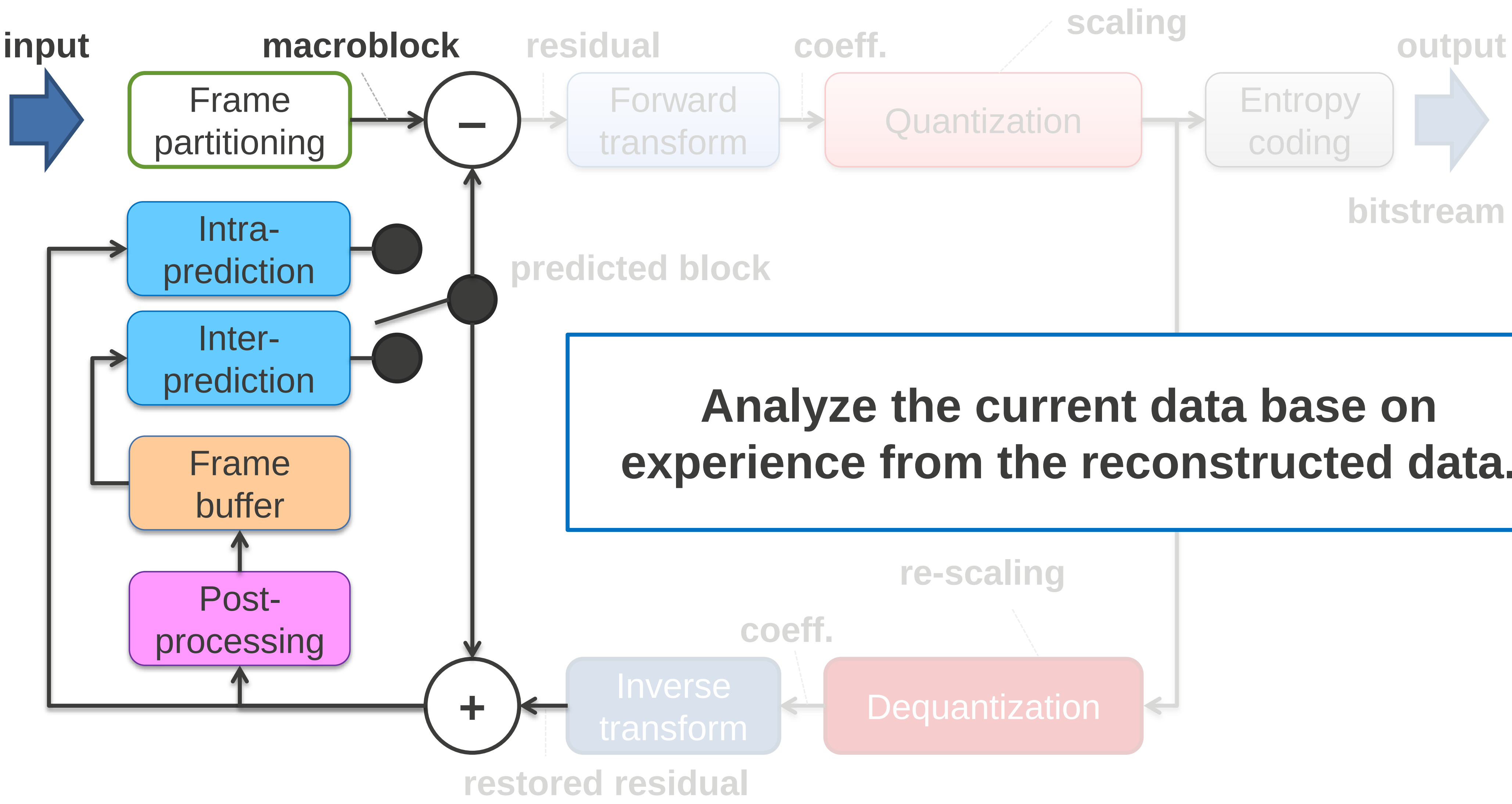
Original (Lenna) Compressed

HYBRID BLOCK-BASED ENCODING FLOW



Reconstruct the original data for analyzing.

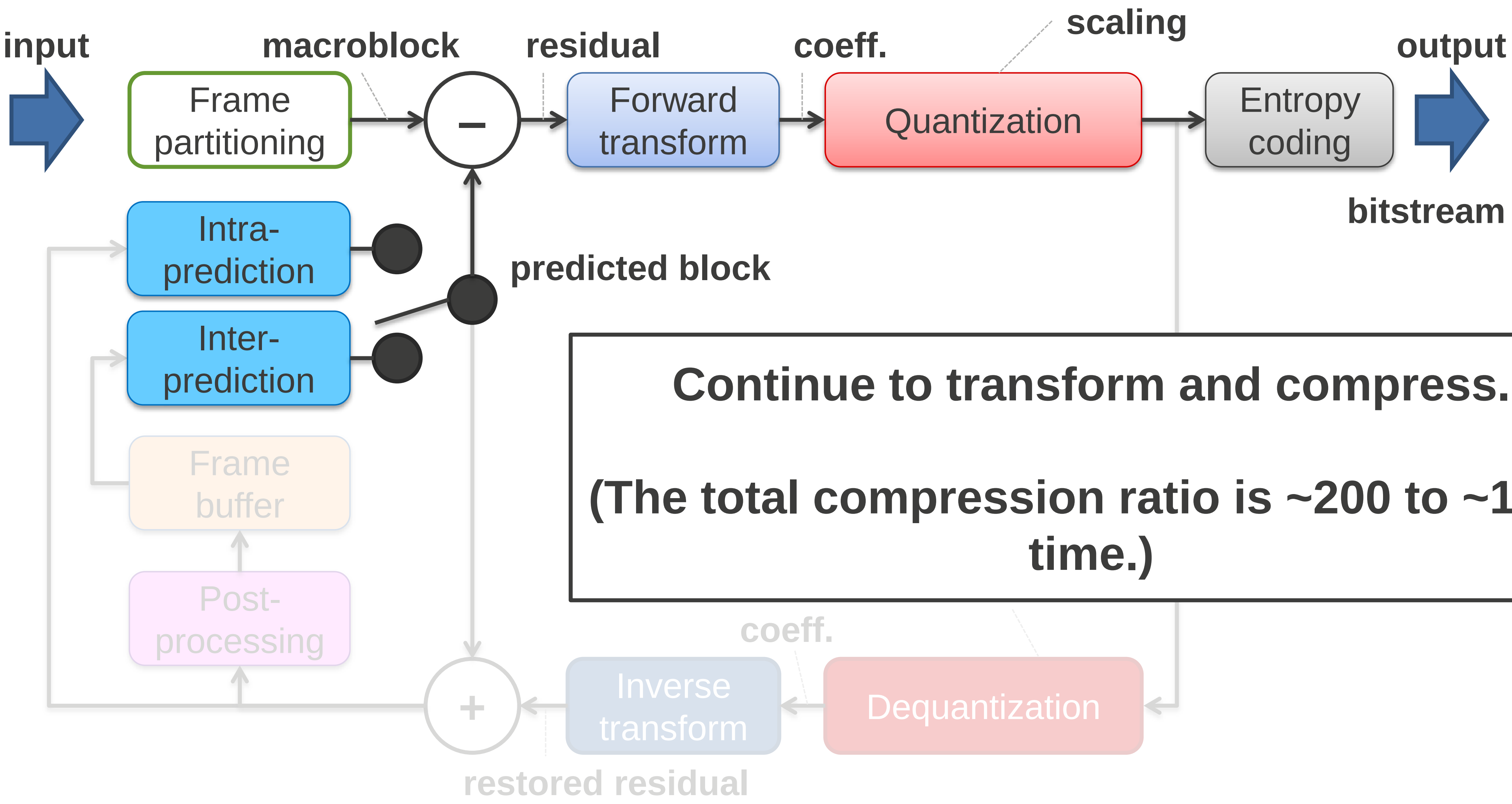
HYBRID BLOCK-BASED ENCODING FLOW



Analyze the current data base on experience from the reconstructed data.

HYBRID BLOCK-BASED ENCODING FLOW

From the 2nd time encoding.

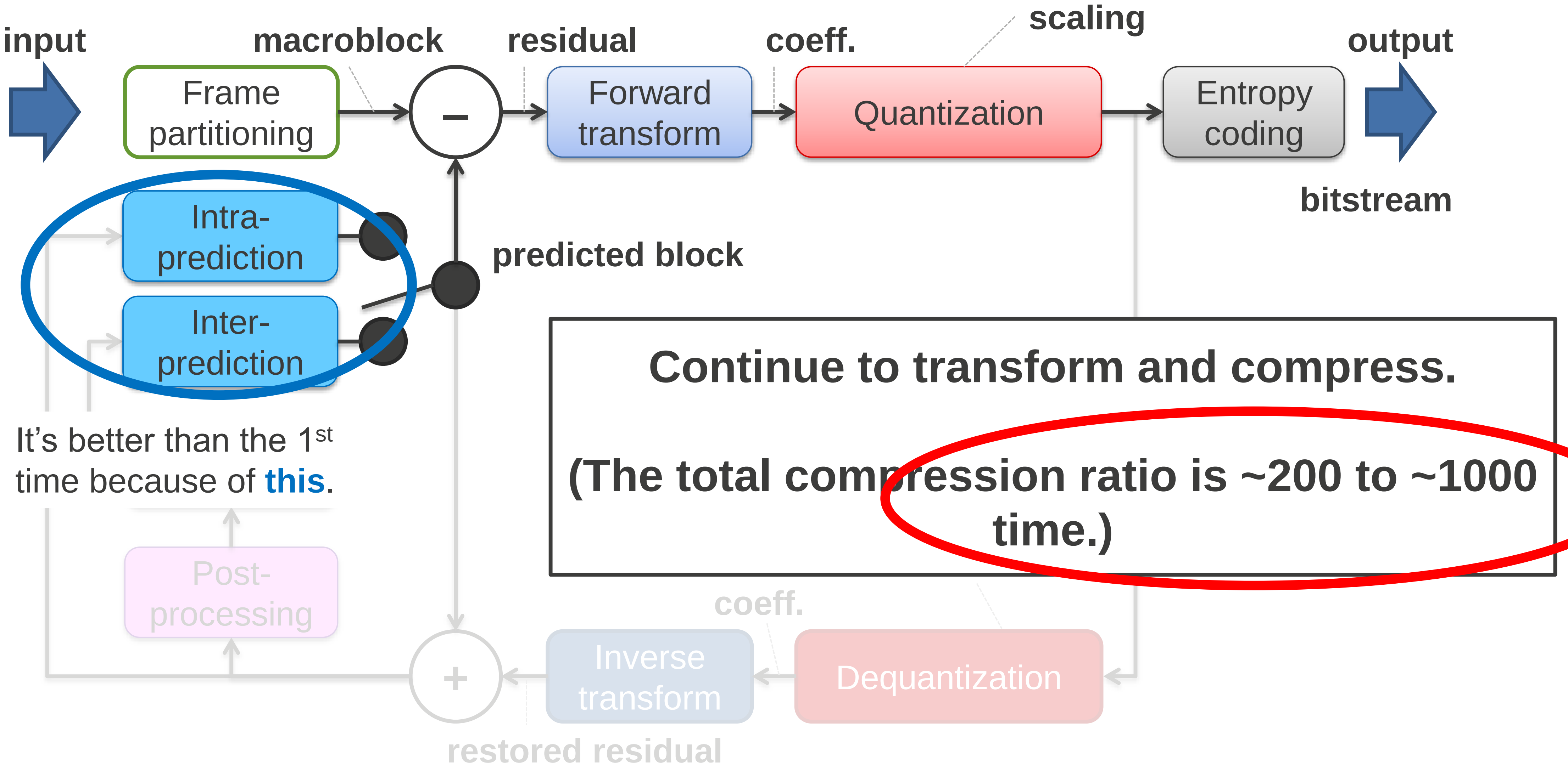


Continue to transform and compress.

(The total compression ratio is ~200 to ~1000 time.)

HYBRID BLOCK-BASED ENCODING FLOW

From the 2nd time encoding.

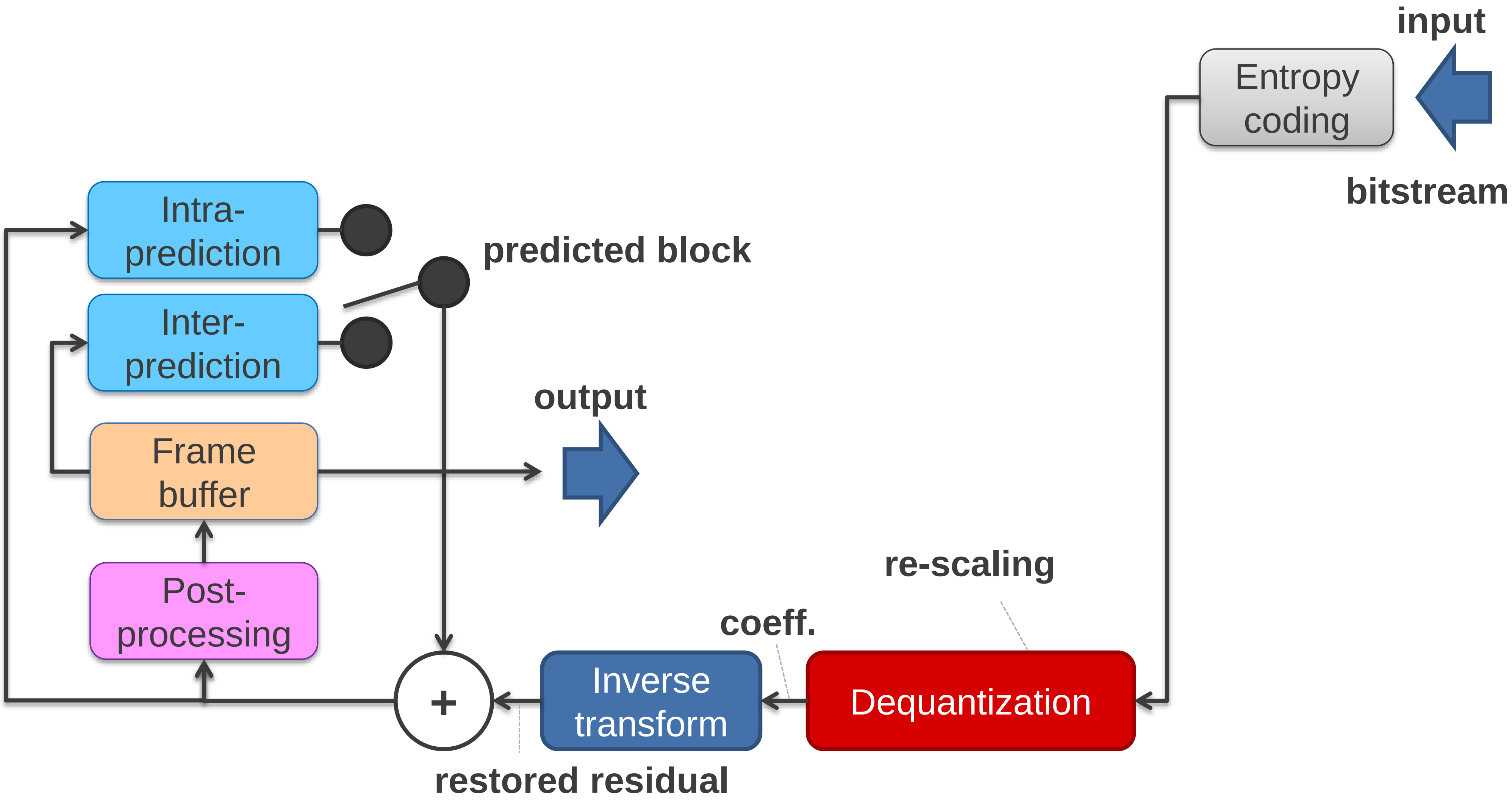


It's better than the 1st time because of **this**.

Continue to transform and compress.

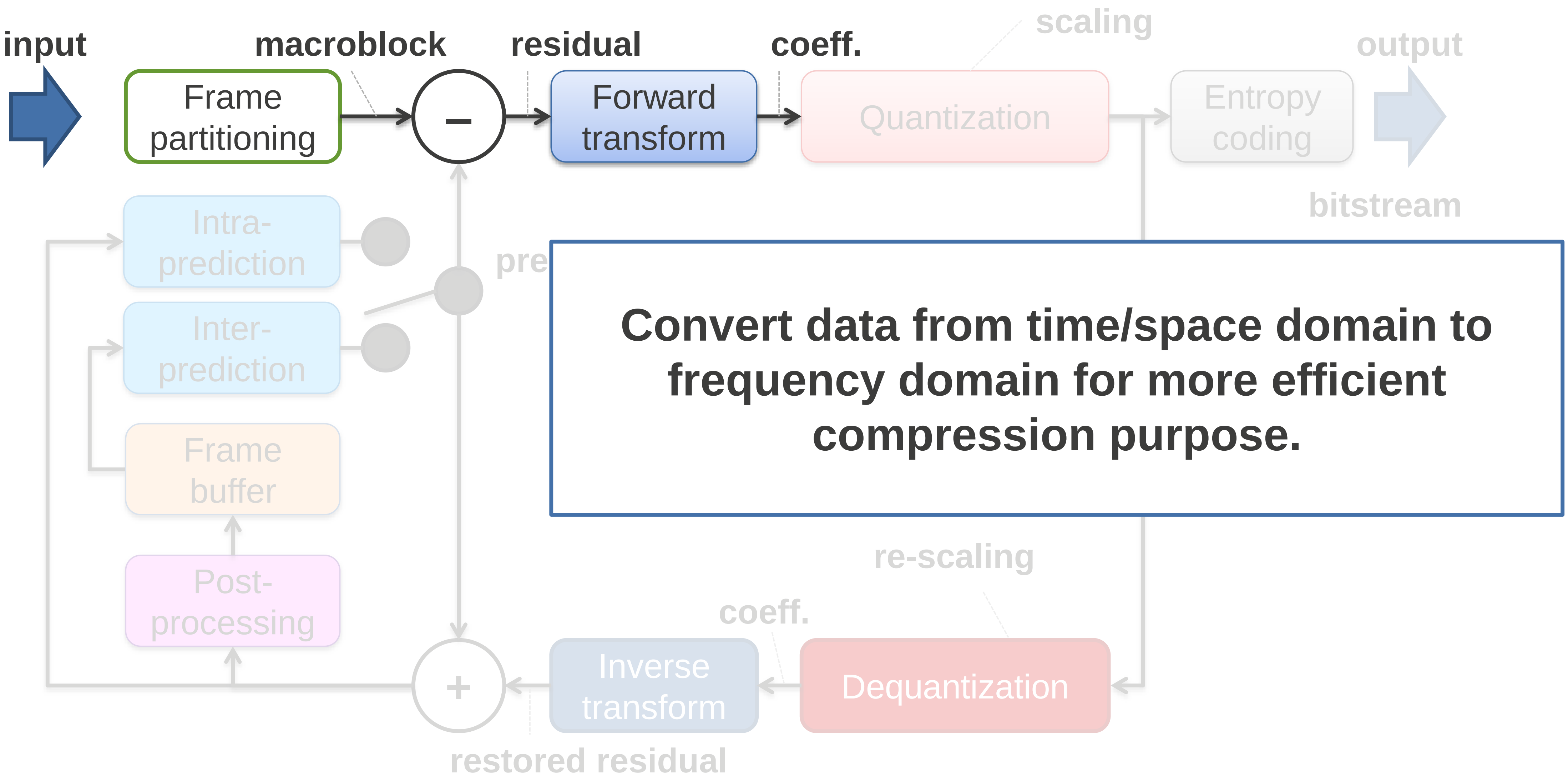
(The total compression ratio is ~200 to ~1000 time.)

HYBRID BLOCK-BASED DECODING FLOW



IN DETAIL…

HYBRID BLOCK-BASED ENCODING FLOW



Convert data from time/space domain to frequency domain for more efficient compression purpose.

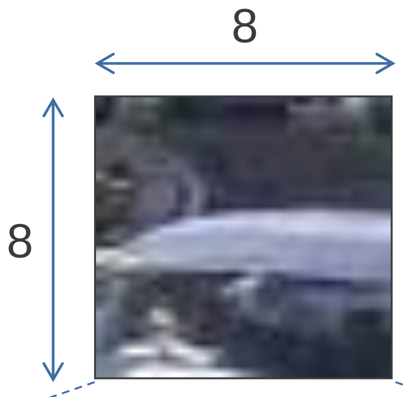
TRANSFORM

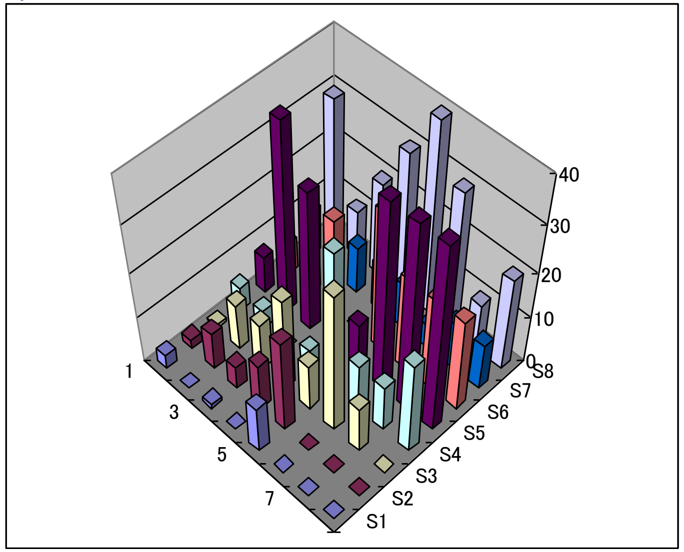
- Convert image or motion-compensated residual data into another domain, the transform domain.

- Data in the transform domain is de-correlated (to **reduce the autocorrelation of a signal with itself**) and compact.

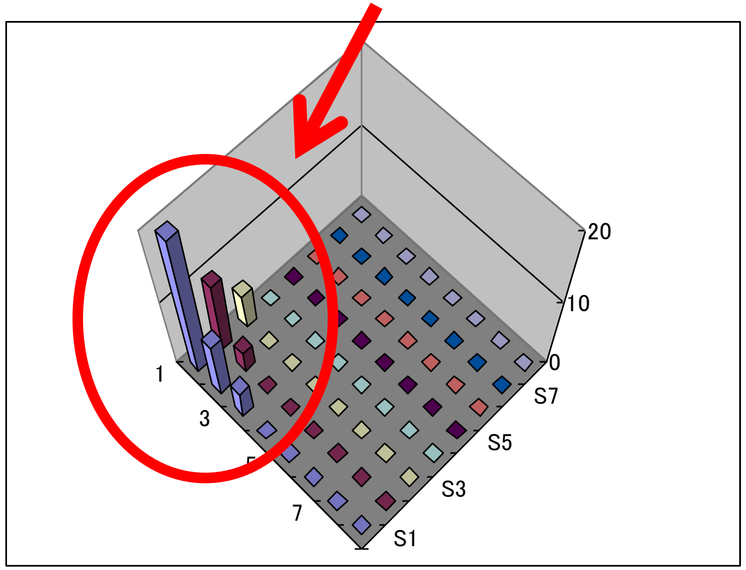
- Popular transform: **DCT (Discrete Cosine Transform)**

TRANSFORM





Data that we're interesting on for compression.

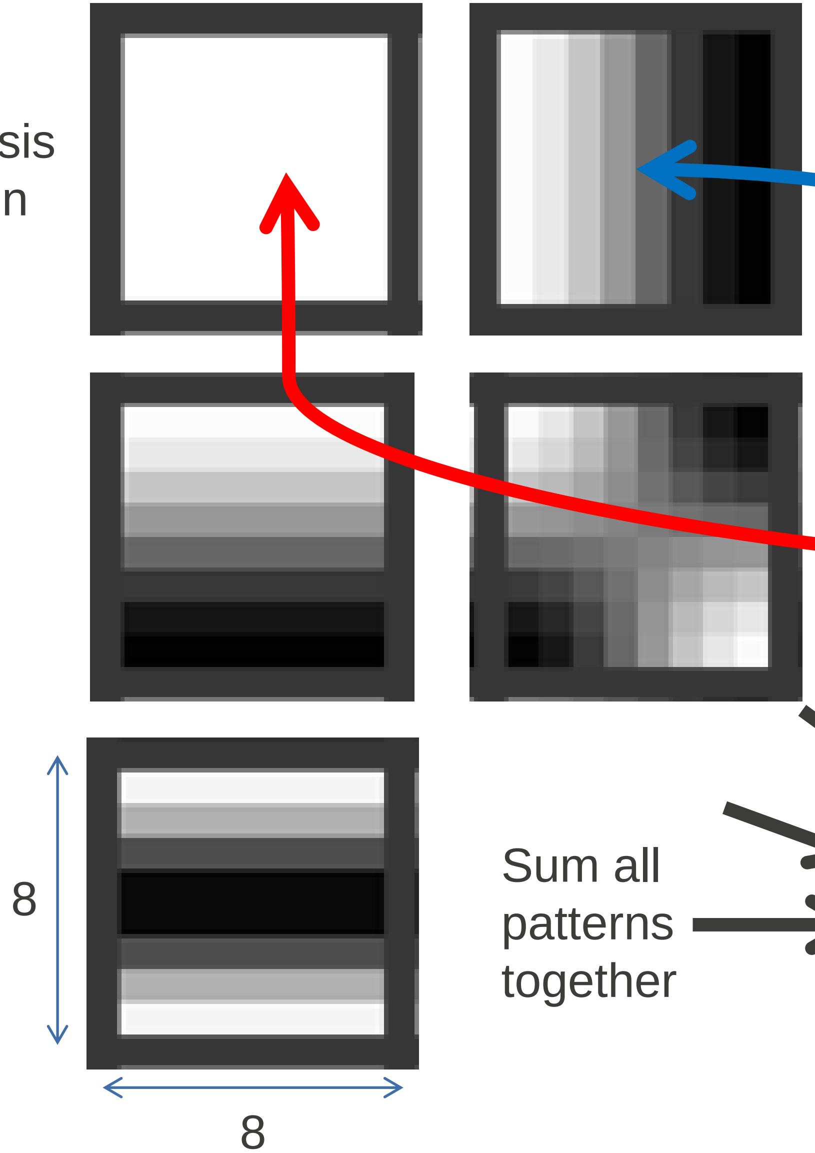


Input of DCT process is a block of 8x8 sampling data.

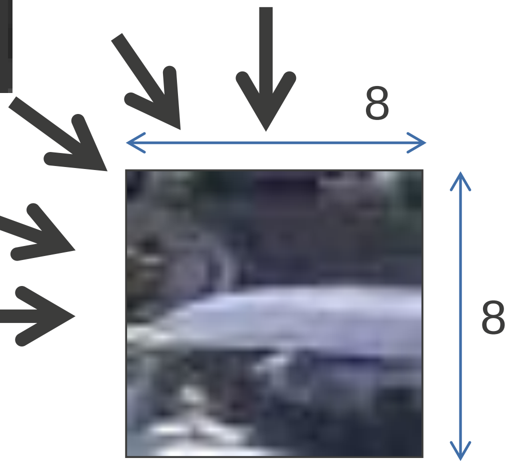
Output of DCT process is a block of 8x8 coefficients which is used for DCT basis functions.

TRANSFORM

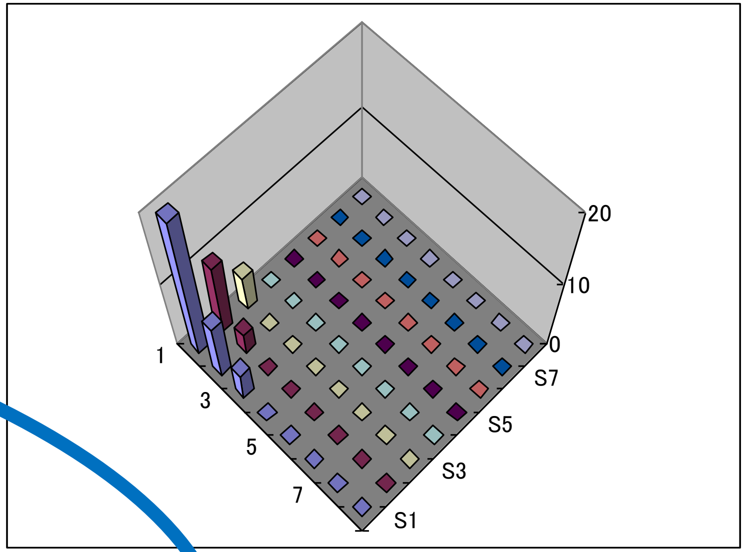
DCT basis function



Sum all patterns together



512 bits



multiply

17	9	5	0	0	0	0	0
8	3	0	0	0	0	0	0
2	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

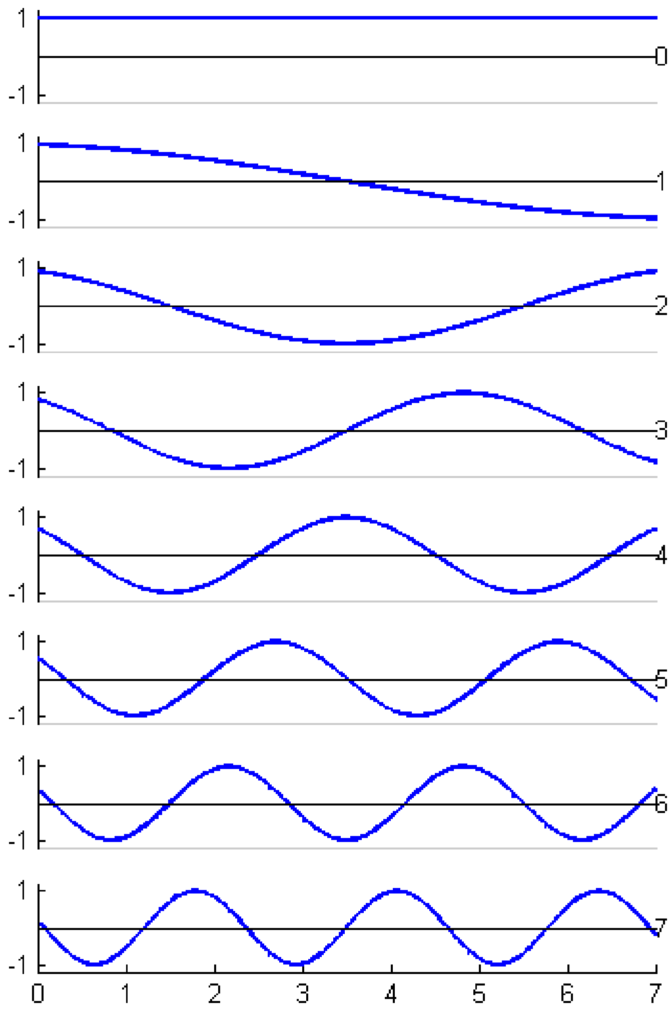
48 bits

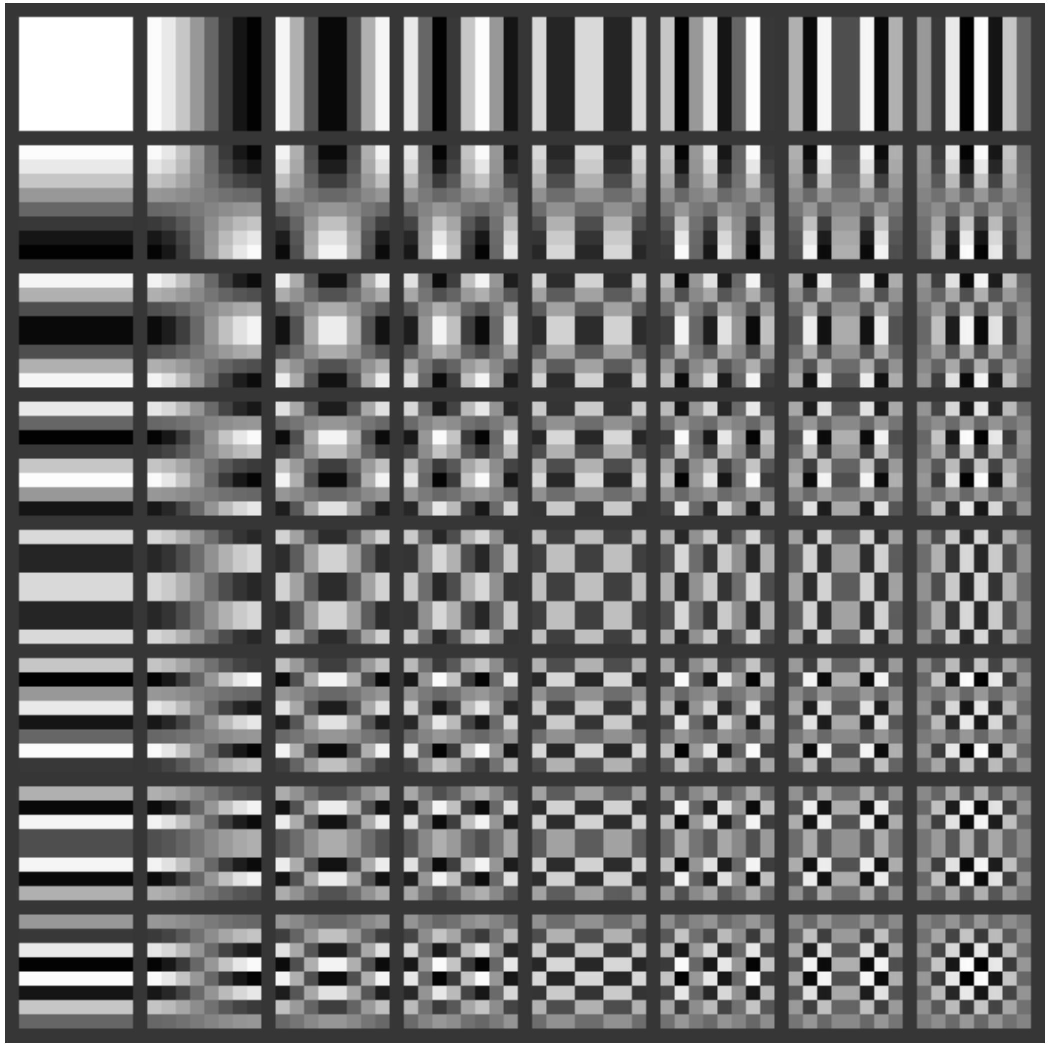
RENESAS

TRANSFORM

DCT 8x8 Basis Patterns

8 points

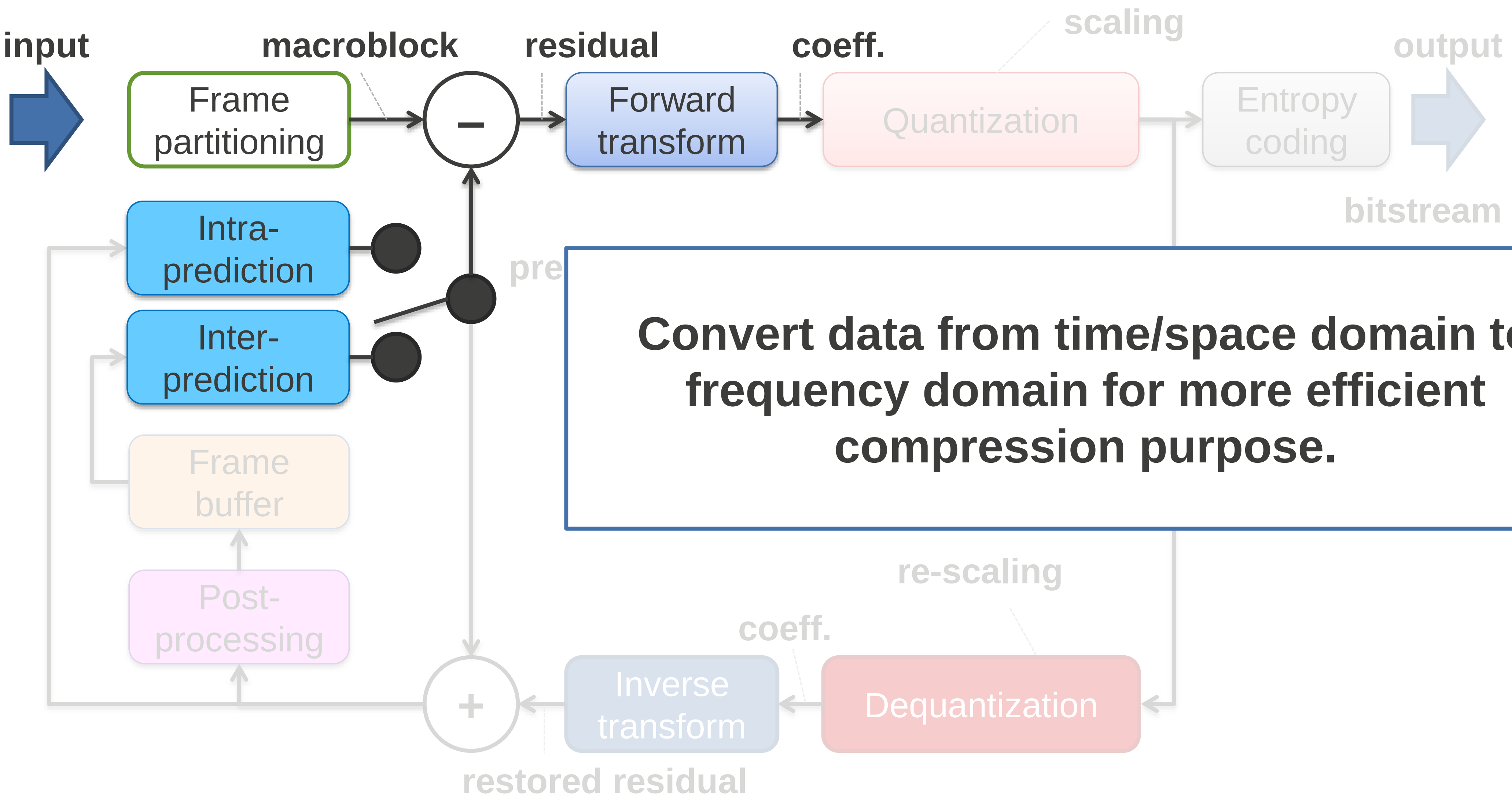




8 blocks

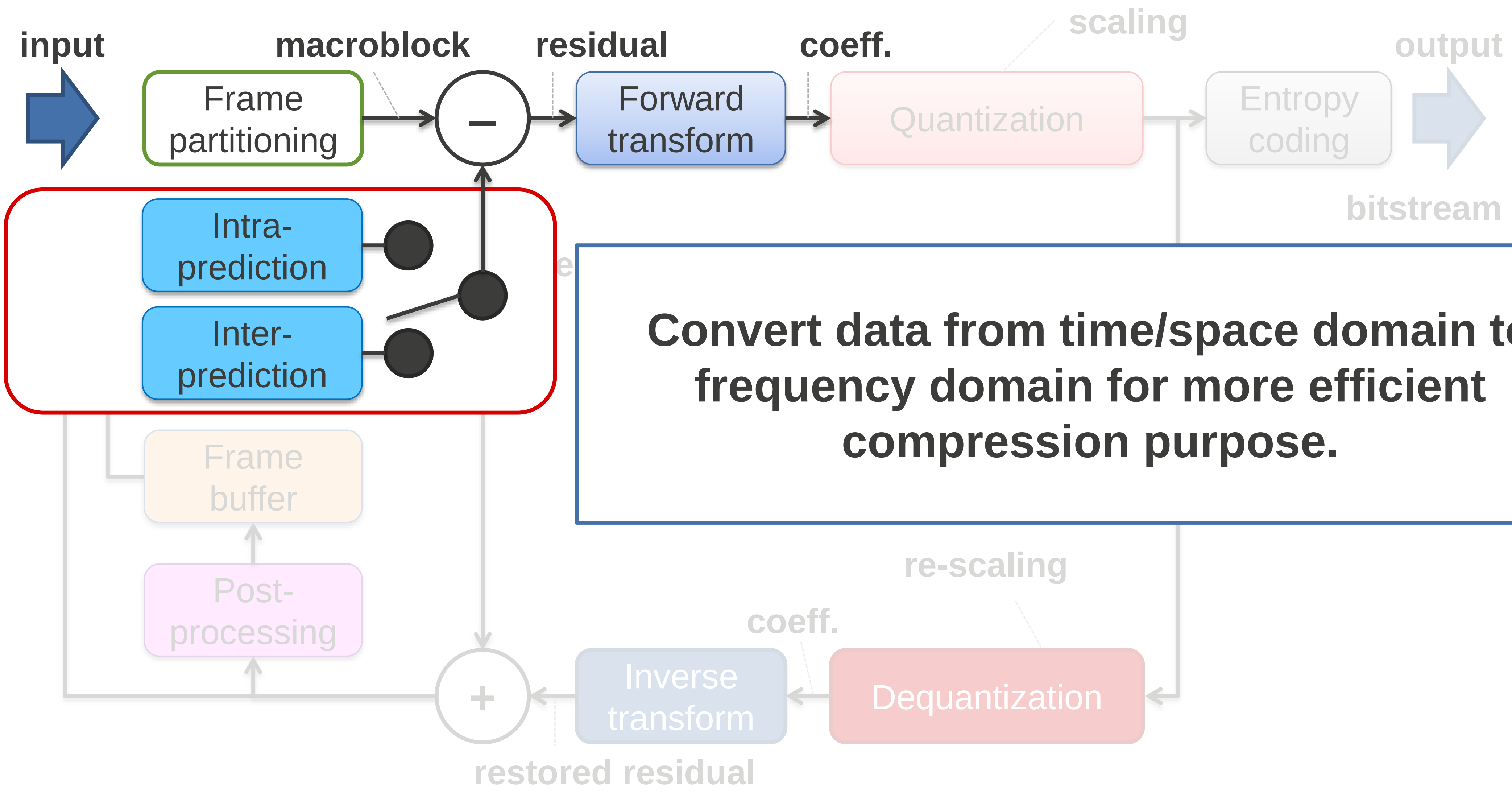
AND TRANSFORMING THE ANALYZED DATA…

HYBRID BLOCK-BASED ENCODING FLOW



Convert data from time/space domain to frequency domain for more efficient compression purpose.

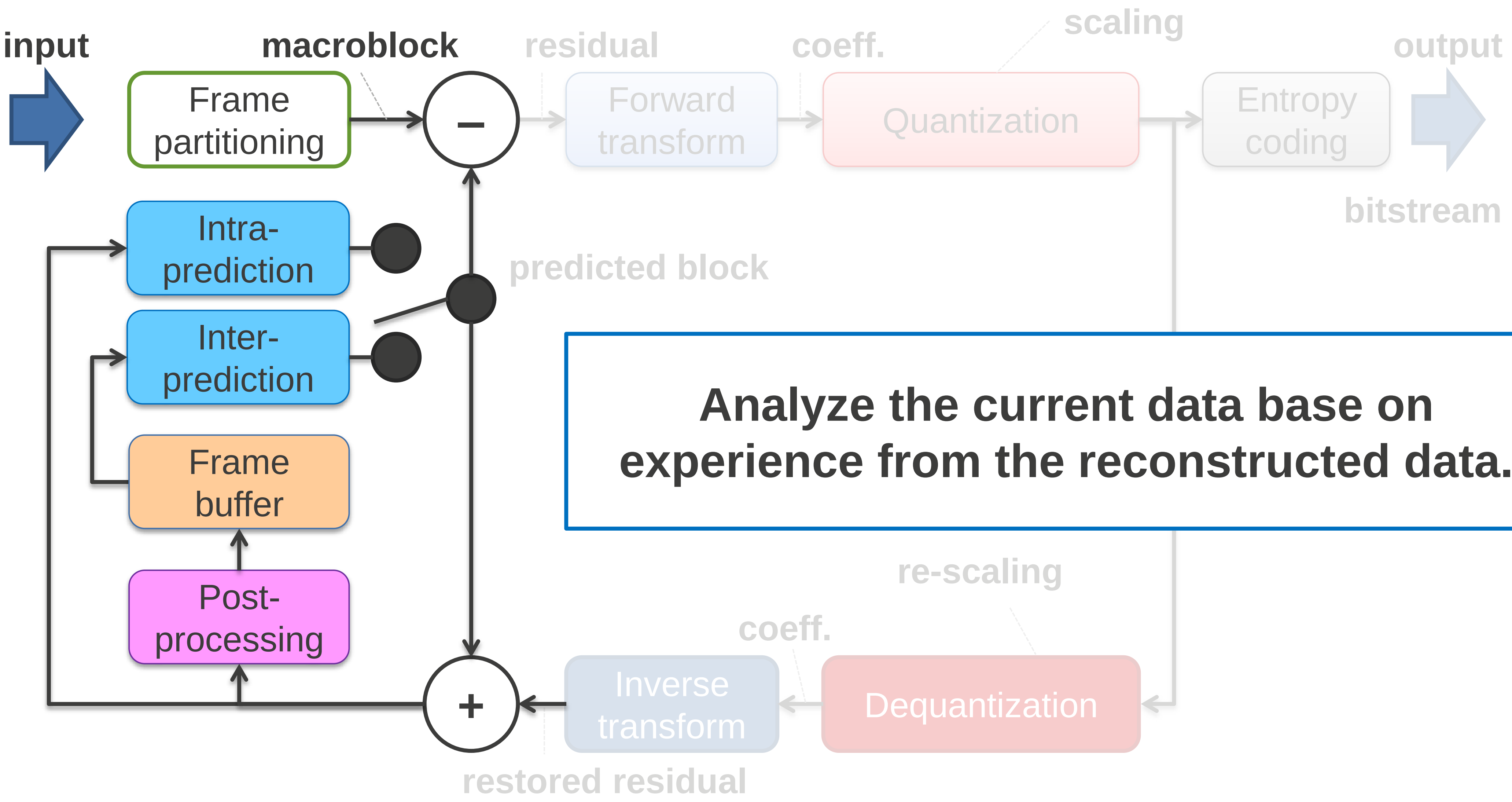
HYBRID BLOCK-BASED ENCODING FLOW



Convert data from time/space domain to frequency domain for more efficient compression purpose.

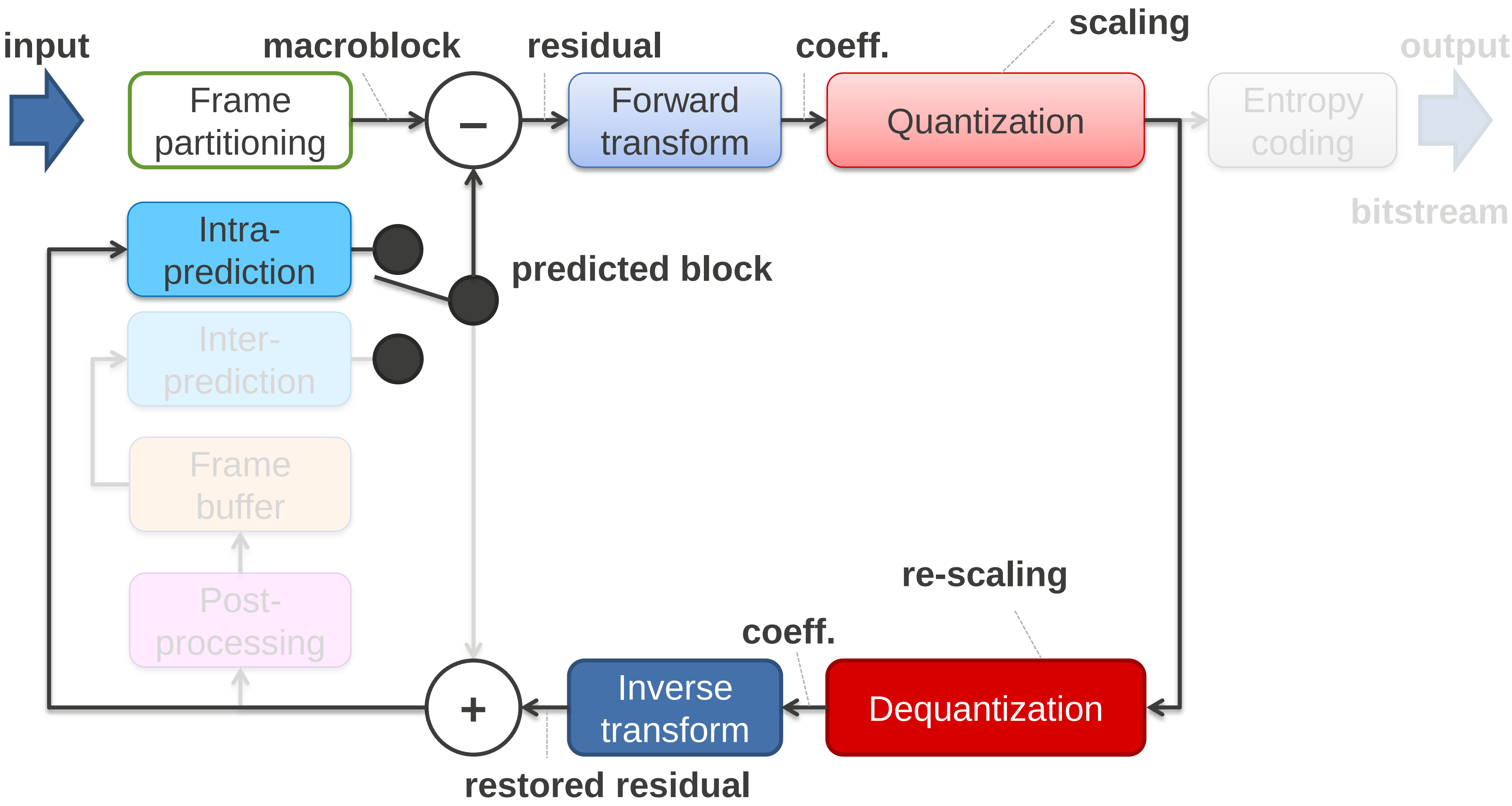
HOW TO ANALYZE THE DATA BASE ON PREVIOUS DATA…

HYBRID BLOCK-BASED ENCODING FLOW



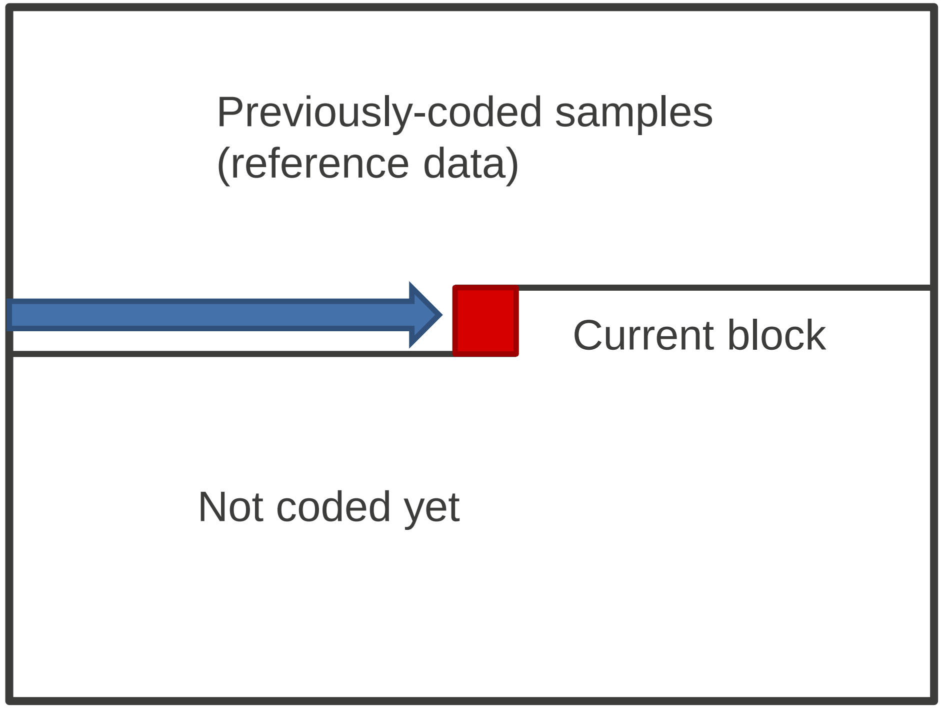
Analyze the current data base on experience from the reconstructed data.

HYBRID BLOCK-BASED ENCODING FLOW

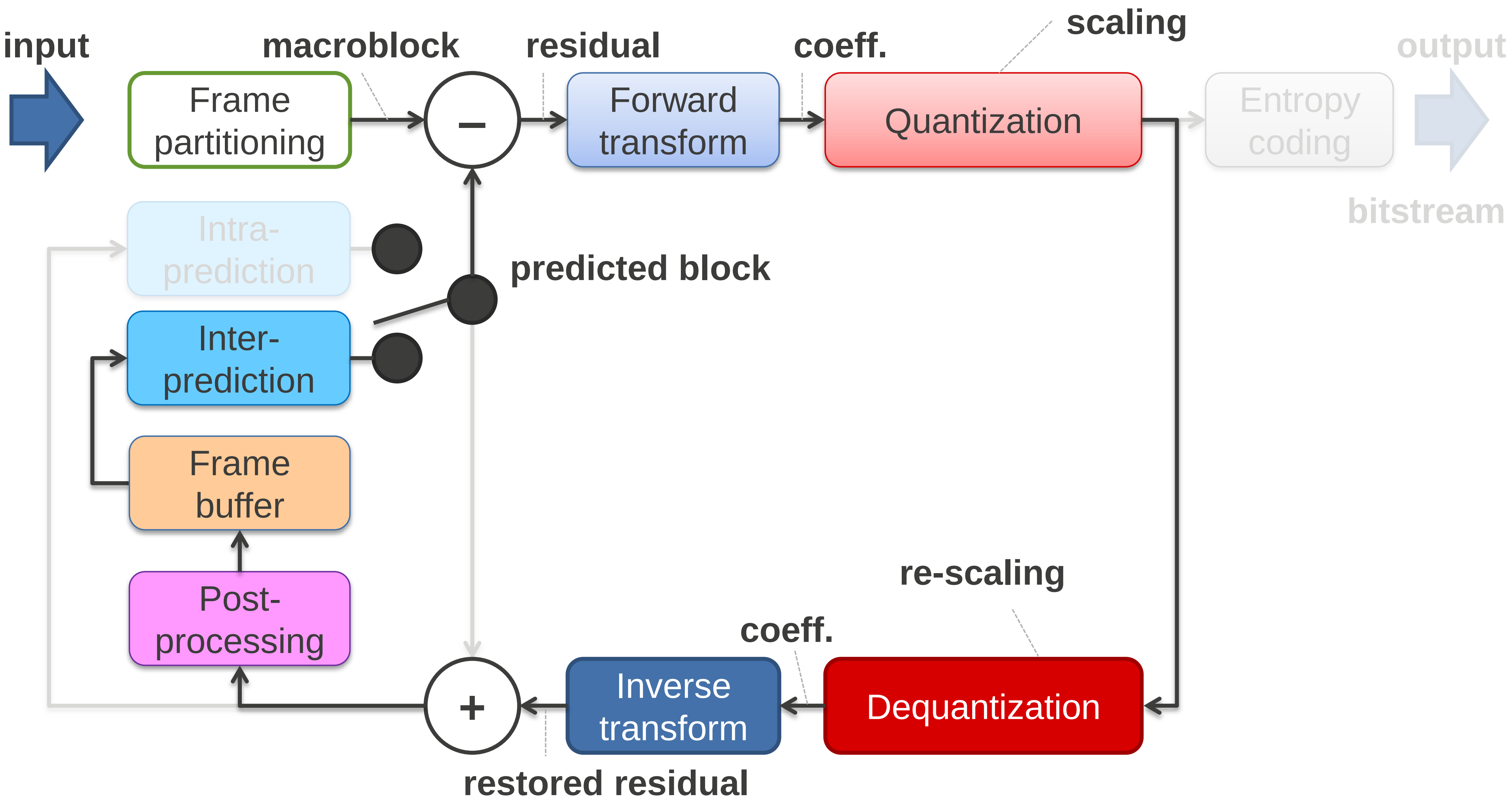


PREDICTION

- It's possible to have some redundant data within a frame ➔ use **intra prediction**.
- **Intra prediction**: The current block is predicted from previously-coded samples in the same frame.

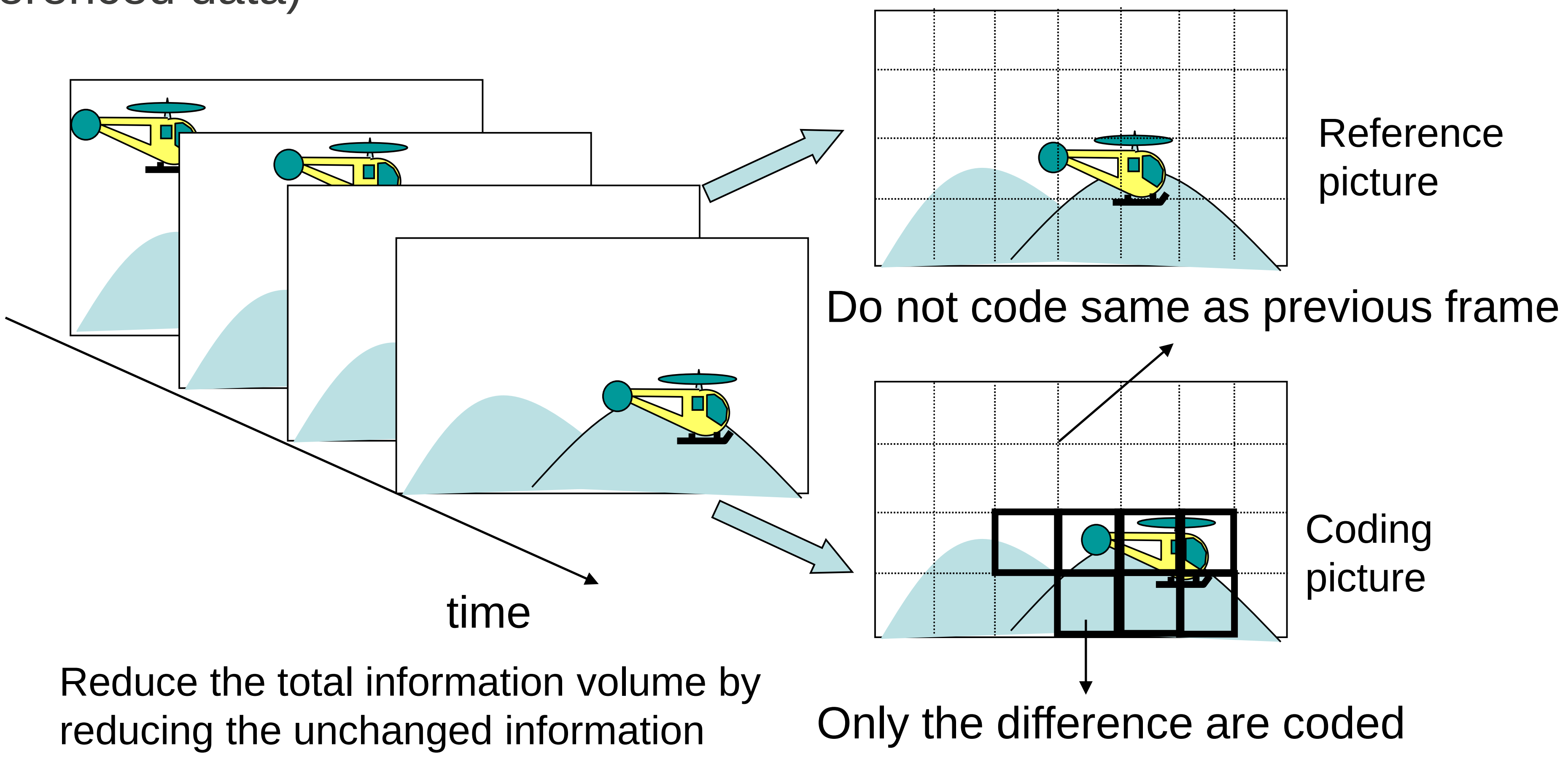


HYBRID BLOCK-BASED ENCODING FLOW



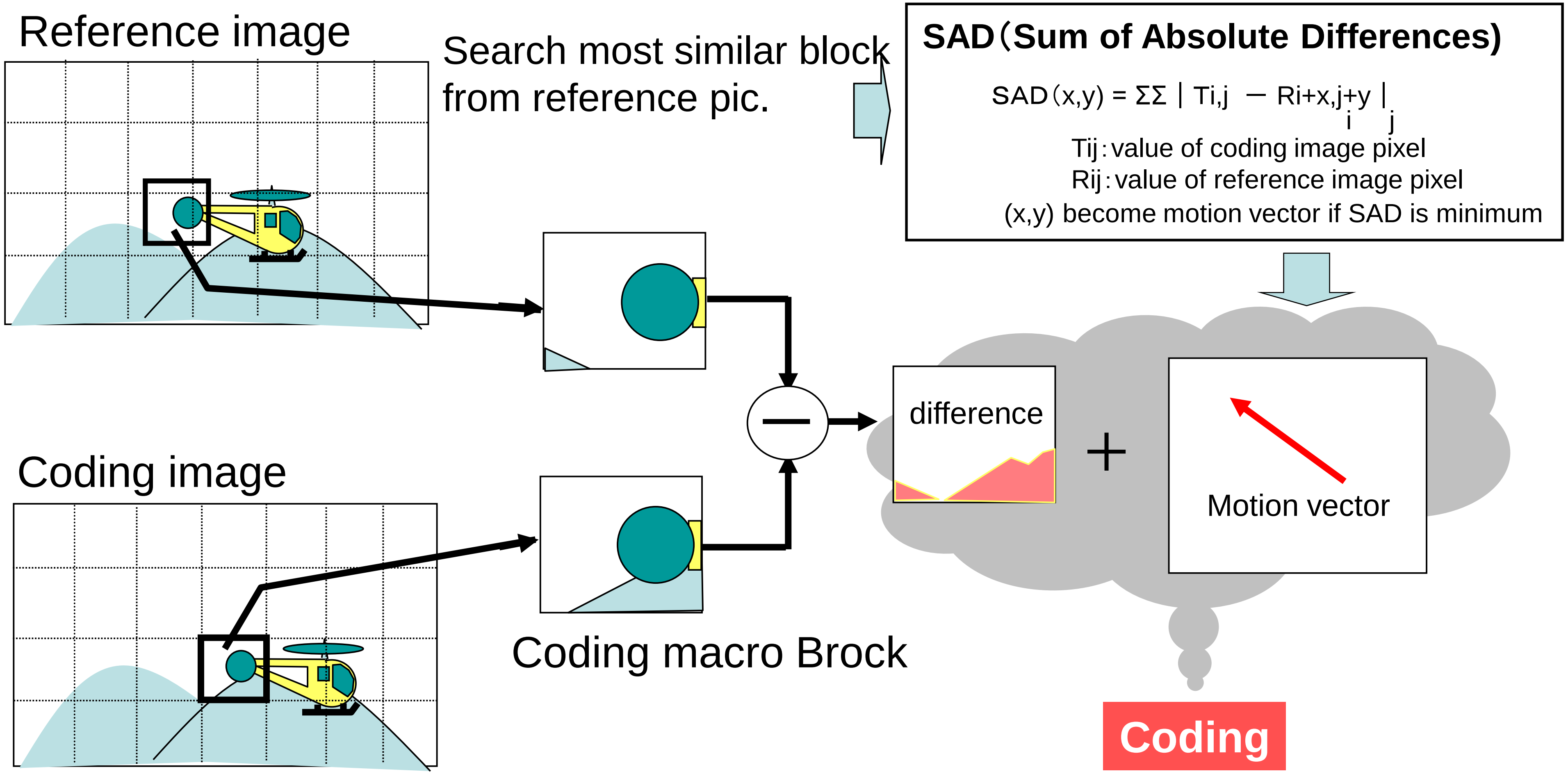
PREDICTION

- A sequence of frames have the same background ➔ use **inter prediction**.
- **Inter prediction**: reduce information using inter-frame correlation (only coding the differenced data)



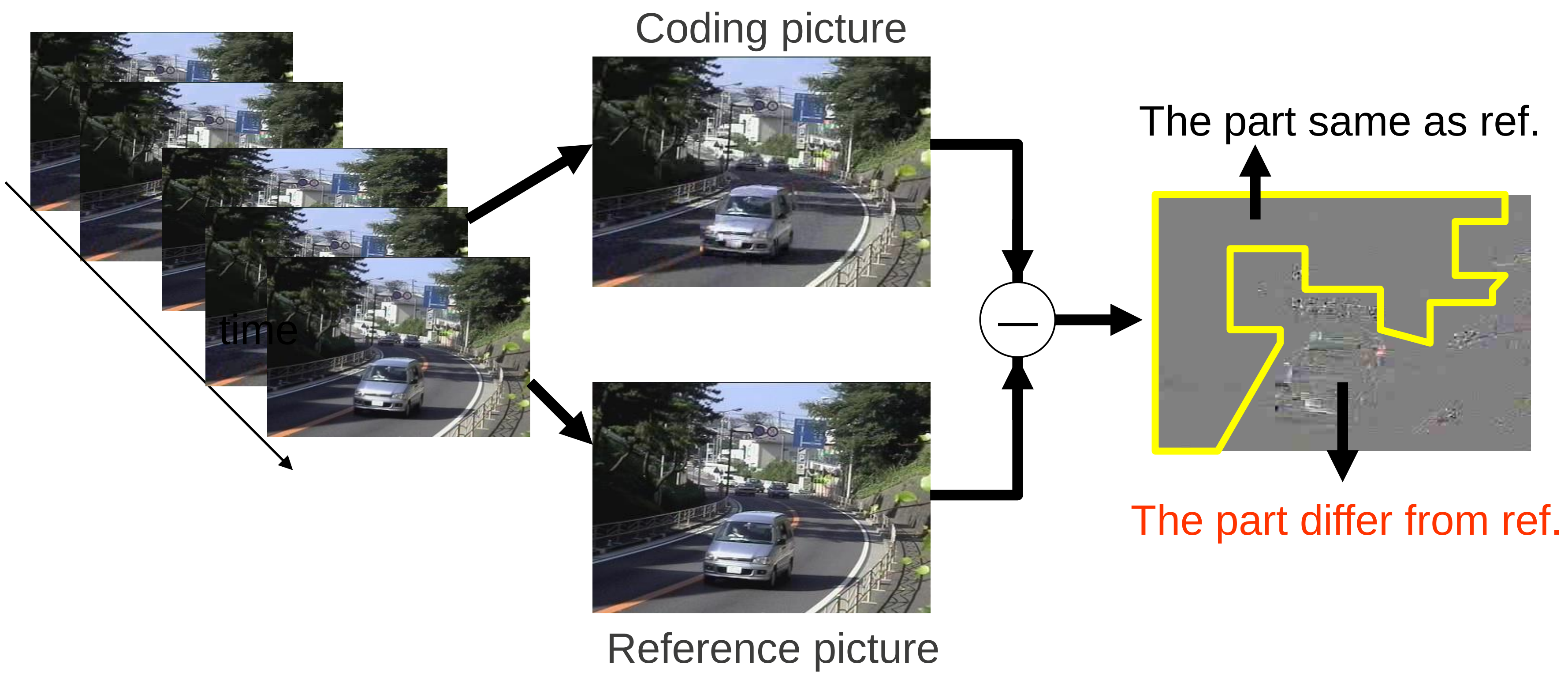
PREDICTION

- **Motion Vector** : To minimize differential code
- **Motion Estimation** : Process to search most similar block



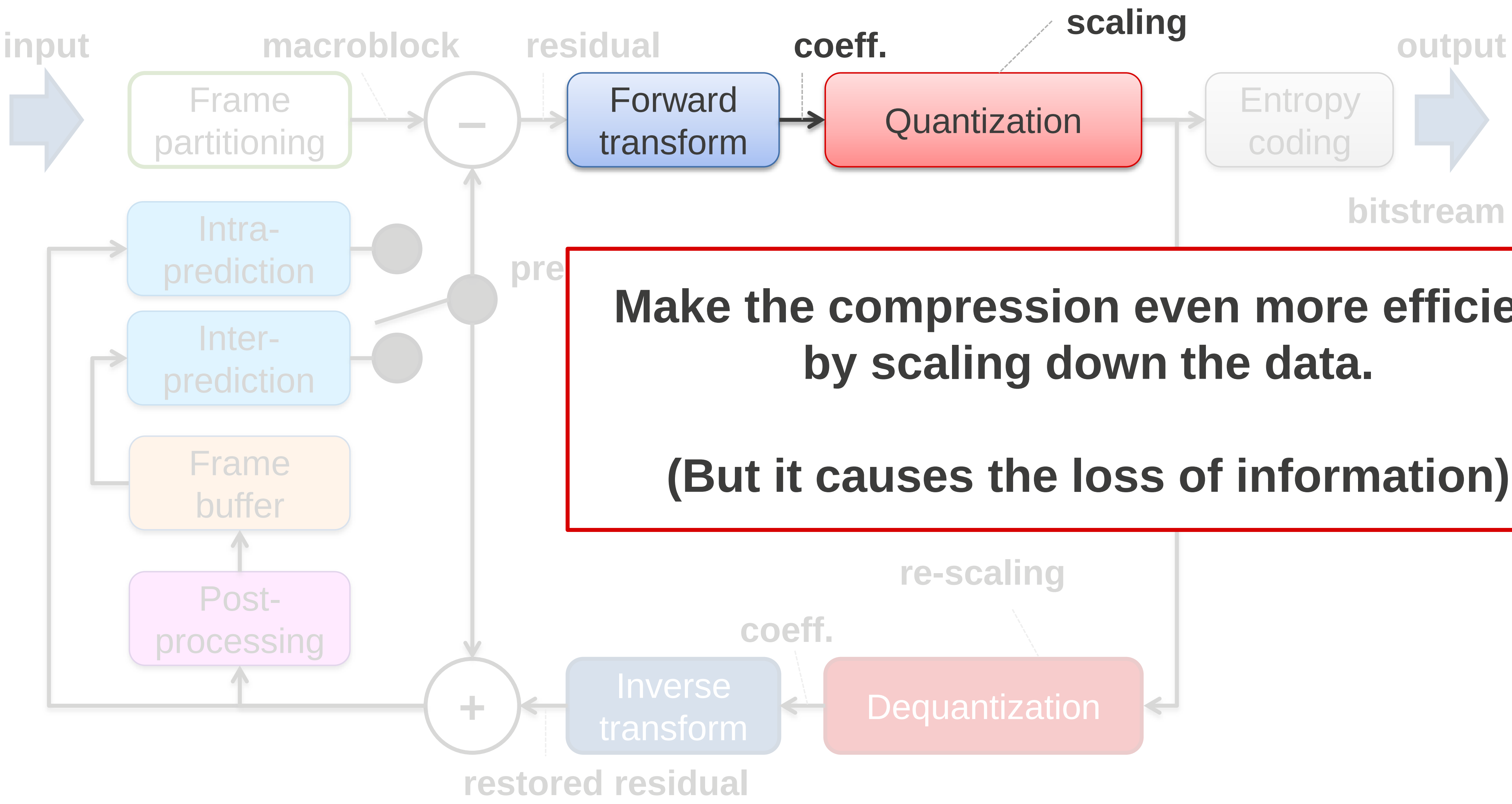
PREDICTION

- It's not important if either previous frame or next frame becomes reference frame. The differences between reference frame and coded frame are still the same.
  → **Bi-directional reference is also possible.**



WORKING THE OUTPUT OF TRANSFORMING PROCESS…

HYBRID BLOCK-BASED ENCODING FLOW



Make the compression even more efficient by scaling down the data.

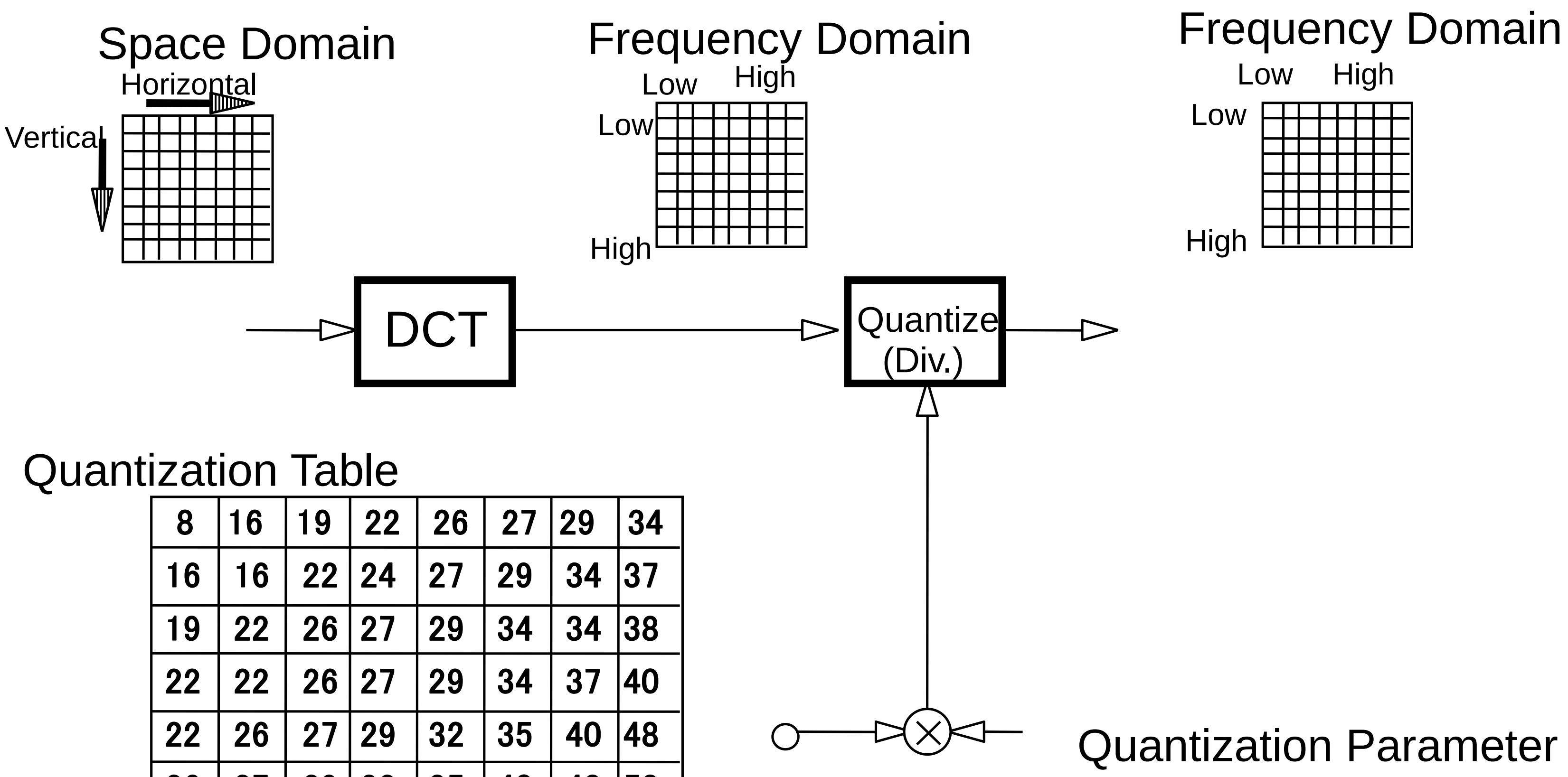
(But it causes the loss of information)

QUANTIZATION

- Quantizer maps a signal with a range of values X to a quantized signal with a **reduced range of values Y**.
- It should be possible to **represent the quantized signal with fewer bits** than the original since the range of possible values is smaller.
- Quantization means "*scale value of data to be something more simple*".
- **Scaling value will naturally cause a lost of data information**.

QUANTIZATION

DCT/Quantization

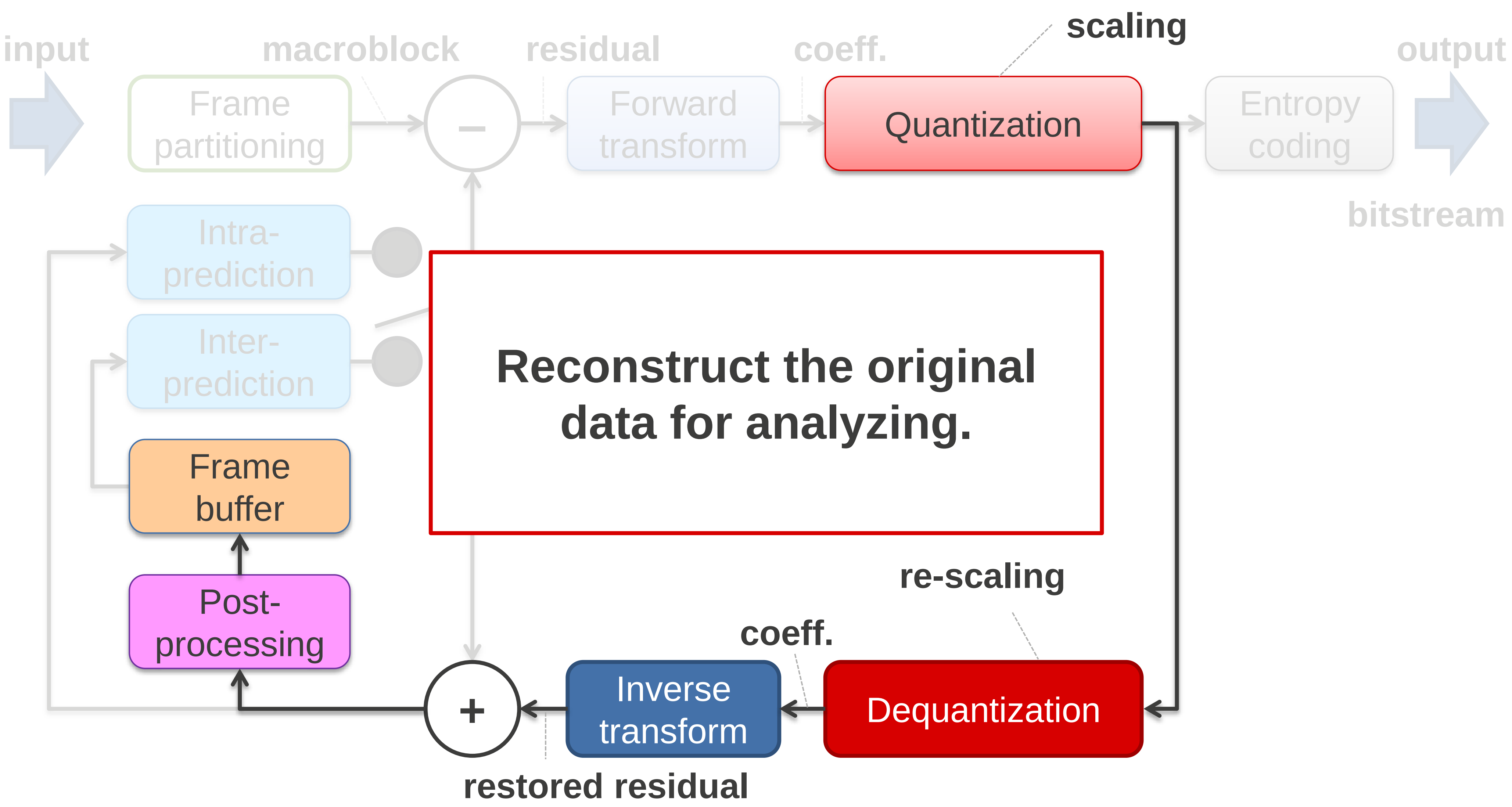


Quantization Table

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

RECONSTRUCTING DATA FOR LATER ANALYZING…

HYBRID BLOCK-BASED ENCODING FLOW



Reconstruct the original data for analyzing.

I-QUANTIZATION / I-DCT

- The reconstructed image may have a **small changes in color information and object recognition**. However, **human-visualization cannot recognize** difference in term of pixel values between original data and reconstructed data.

DCT / Quantization / I-quantization / I-DCT

Space Domain

Horizontal →

Vertical ↓

85	62	111	141	134	136	126	114
96	85	129	136	118	107	78	58
159	141	127	112	90	109	122	143
227	189	129	100	109	120	113	109
68	57	65	112	127	120	126	125
117	103	122	124	117	114	110	103
61	85	113	114	111	109	87	86
101	124	133	109	81	89	111	113

Input signal

⇨

Space Domain

87	63	98	149	145	132	131	115
96	86	118	139	113	94	82	56
170	144	130	113	93	114	140	132
224	179	135	104	90	111	127	114
66	59	72	103	123	132	129	116
115	114	118	119	109	100	104	113
53	90	123	128	118	105	90	81
106	132	129	102	94	105	106	98

Output signal

I-QUANTIZATION / I-DCT

139	144	149	153	155	155	155	155
144	151	153	156	159	156	156	156
150	155	160	163	158	156	156	156
159	161	162	160	160	159	159	159
159	160	161	162	162	155	155	155
161	161	161	161	160	157	157	157
162	162	161	163	162	157	157	157
162	162	161	161	163	158	158	158

(a) source image samples

235.6	-1.0	-12.1	-5.2	2.1	-1.7	-2.7	1.3
-22.6	-17.5	-6.2	-3.2	-2.9	-0.1	0.4	-1.2
-10.9	-9.3	-1.6	1.5	0.2	-0.9	-0.6	-0.1
-7.1	-1.9	0.2	1.5	0.9	-0.1	0.0	0.3
-0.6	-0.8	1.5	1.6	-0.1	-0.7	0.6	1.3
1.8	-0.2	1.6	-0.3	-0.8	1.5	1.0	-1.0
-1.3	-0.4	-0.3	-1.5	-0.5	1.7	1.1	-0.8
-2.6	1.6	-3.8	-1.8	1.9	1.2	-0.6	-0.4

(b) forward DCT coefficients

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

(c) quantization table

15	0	-1	0	0	0	0	0
-2	-1	0	0	0	0	0	0
-1	-1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

(d) normalized quantized coefficients

240	0	-10	0	0	0	0	0
-24	-12	0	0	0	0	0	0
-14	-13	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

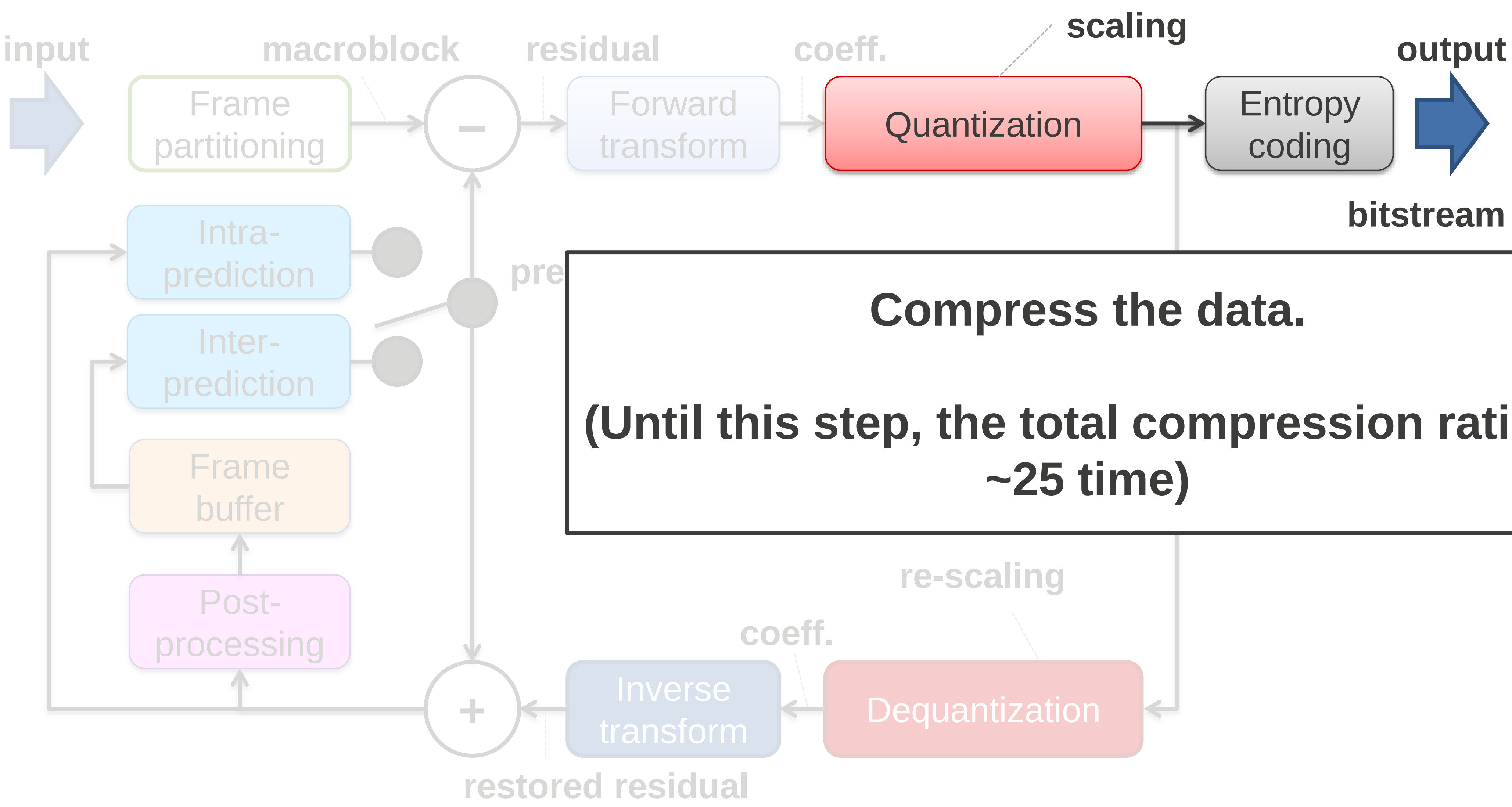
(e) denormalized quantized coefficients

144	146	149	152	154	156	156	156
148	150	152	154	156	156	156	156
155	156	157	158	158	157	156	155
160	161	161	162	161	159	157	155
163	163	164	163	162	160	158	156
163	164	164	164	162	160	158	157
160	161	162	162	162	161	159	158
158	159	161	161	162	161	159	158

(f) reconstructed image samples

COMPRESSING DATA…

HYBRID BLOCK-BASED ENCODING FLOW

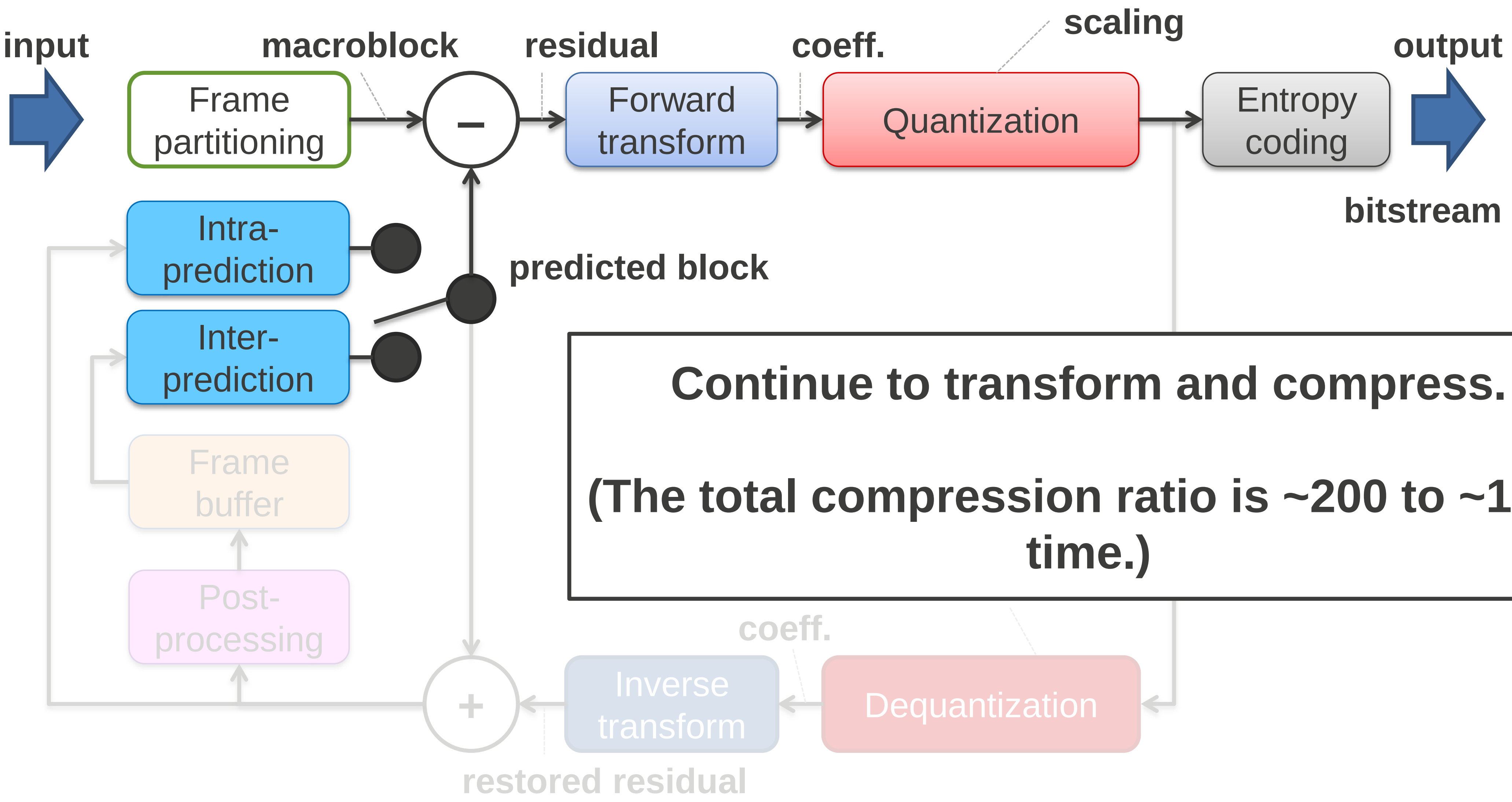


Compress the data.

(Until this step, the total compression ratio is ~25 time)

HYBRID BLOCK-BASED ENCODING FLOW

From the 2nd time encoding.



Continue to transform and compress.

(The total compression ratio is ~200 to ~1000 time.)

ENTROPY CODING

- **Entropy coding**: To assign code according to occurrence probability.
- It converts a series of symbols (quantized transform coefficients, run-level, motion vectors, picture headers, etc.) representing elements of the video sequence into a compressed bitstream suitable for transmission or storage.

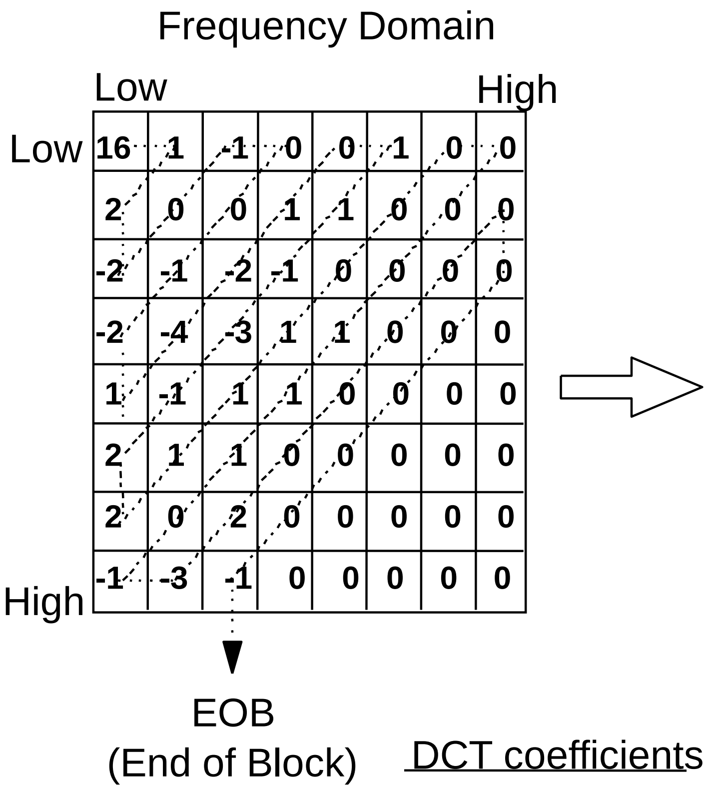
Value	Probability	Normal code	VLC
1	1/2	001	1
2	1/4	010	01
3	1/8	011	001
4	1/8	100	000

Coding value	1,2,1,1,2,3,1,4
In normal code	001,010,001,001,010,011,001,100（24bit)
In VLC (Variable Length Coding)	1,01,1,1,01,001,1,000(14bit)

ENTROPY CODING / ZIGZAG SCAN

- ZigZag Scan is fixed between encoder and decoder, but can be customized depending to the codec specification itself.

Zigzag Scan/VLC

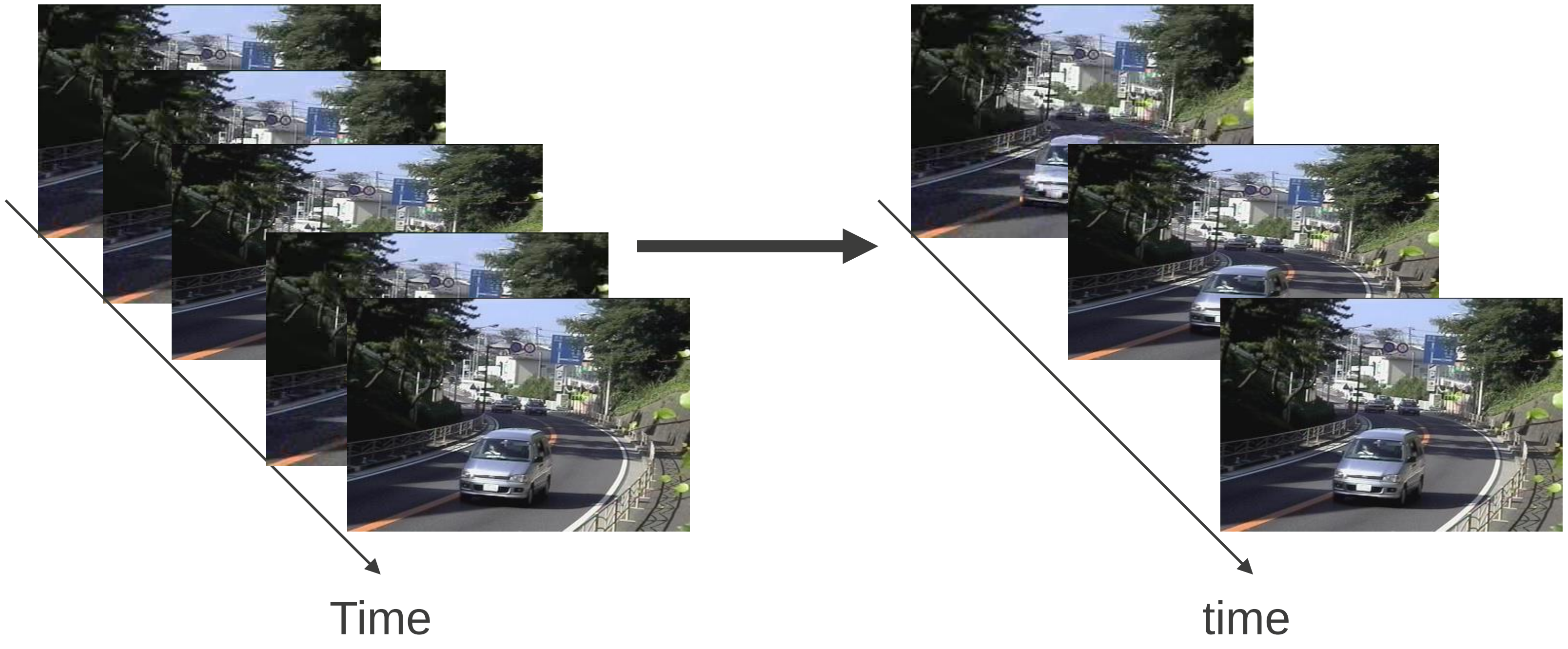


DCT Coeff.	Zero Run	Value	Bit length
DC coeff.		16	15
AC coeff.	0	1	3
	0	2	5
	0	-2	5
	1	-1	4
	2	-1	5
	0	-2	5
	0	1	3
	0	-4	8
	0	-2	5
	0	1	3
	1	1	4
	0	1	3
	0	-1	3
	0	-3	6
	0	-1	3
	0	2	5
	0	2	5
	0	1	3
	0	1	3
	0	1	3
	6	1	3
	0	1	3
	0	1	3
	1	-1	4
	0	-3	6
	0	2	5
	10	-1	9
EOB			2
Total			134

THAT'S ALMOST ABOUT VIDEO CODING, BUT…

OTHER BASIC VIDEO CODING TECHNOLOGY (1)

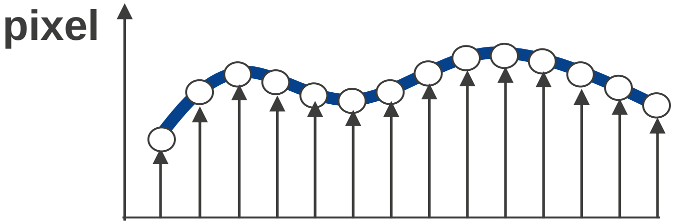
- **Dropping frame**: To decimate pictures by time.
- **Dropping frame** = **resampling data** ➔ **cause a big lost of information**
  - In encoding viewpoint, input-captured may provide more frames than the processing capability of encoder.
  - In decoding viewpoint, output-displayed may slow down the processing of decoder due to its limitation in FPS.



OTHER BASIC VIDEO CODING TECHNOLOGY (1)

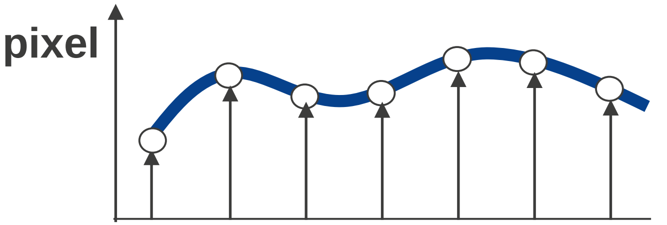
- ❑ **Sub-sampling**: To decimate pixels number in a picture
  (change resolution of picture)
- ❑ **Sub-sampling** = **resizing image to be smaller one**
- ❑ Less information ➔ less data size ➔ better compression ➔ poor quality display

Picture



After cut off high frequency using filter,
Decimate some pixels in a picture.

Picture



BASIC VIDEO CODING TECHNOLOGY: SUMMARY

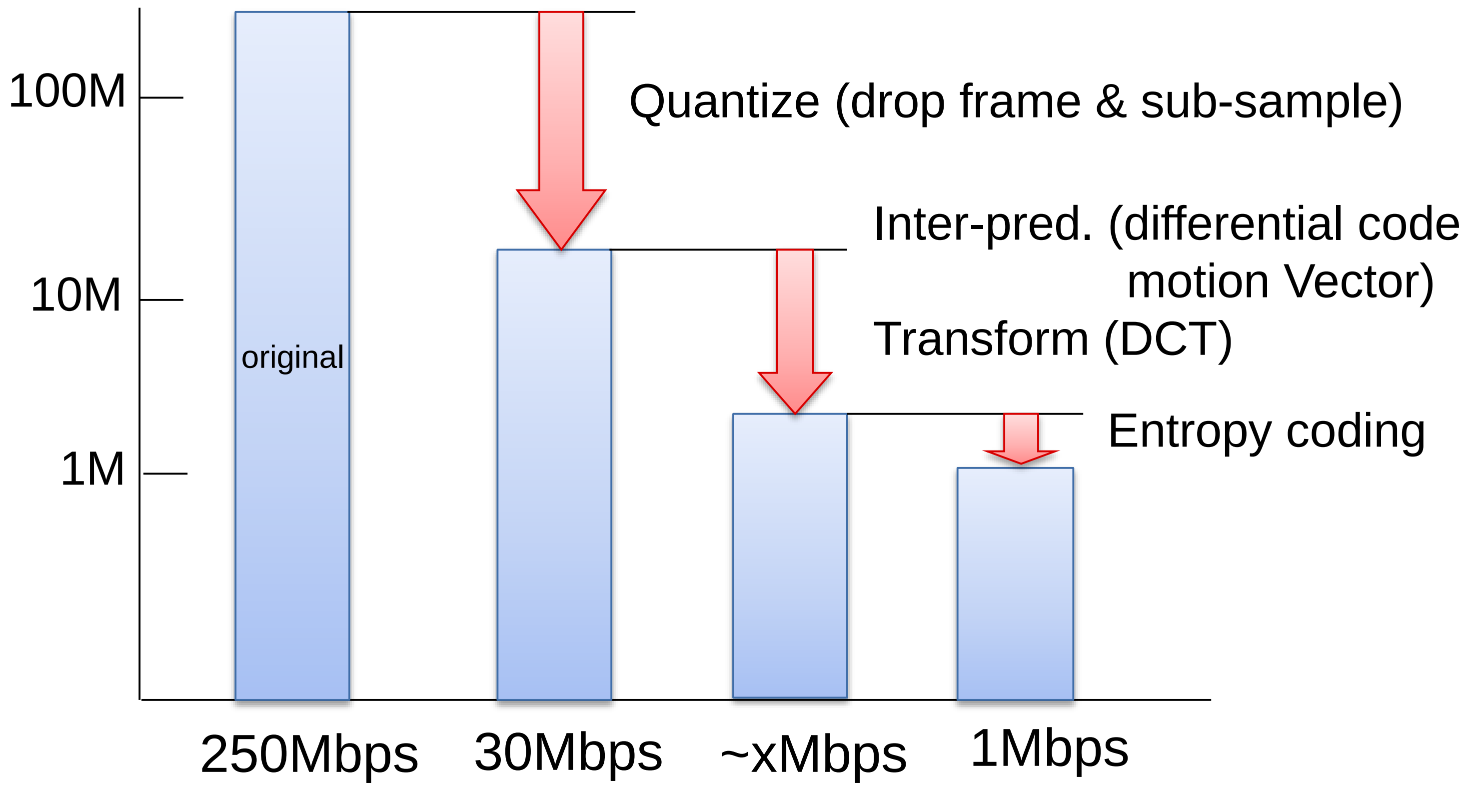
Group	Method		Effectiveness	Deterioration
Cut of information (irreversible)	Drop frame		◎	large
	Sub-sampling		◎	large
Reduce of information (reversible/ irreversible)	In spatial	**Intra prediction (DC prediction)**	○	medium
	In temporal	**Inter prediction (difference motion vector)**	○	medium
	Time-Frequency Transform	**DCT**	○	medium
Use statistic feature (reversible)	Allocate suitable code	**Entropy coding Zigzag scan**	△	small (non)

BASIC VIDEO CODING TECHNOLOGY: SUMMARY

Group	Method		Effectiveness	Deterioration
Cut of information (irreversible)	Drop frame		◎	large
	Sub-sampling		◎	large
Reduce of information (reversible/ irreversible)	In spatial	**Intra prediction (DC prediction)**	○	medium
	In temporal	**Inter prediction (difference motion vector)**	○	medium
	Time-Frequency Transform	**DCT**	○	medium
Use statistic feature (reversible)	Allocate suitable code	**Entropy coding Zigzag scan**	△	small (non)

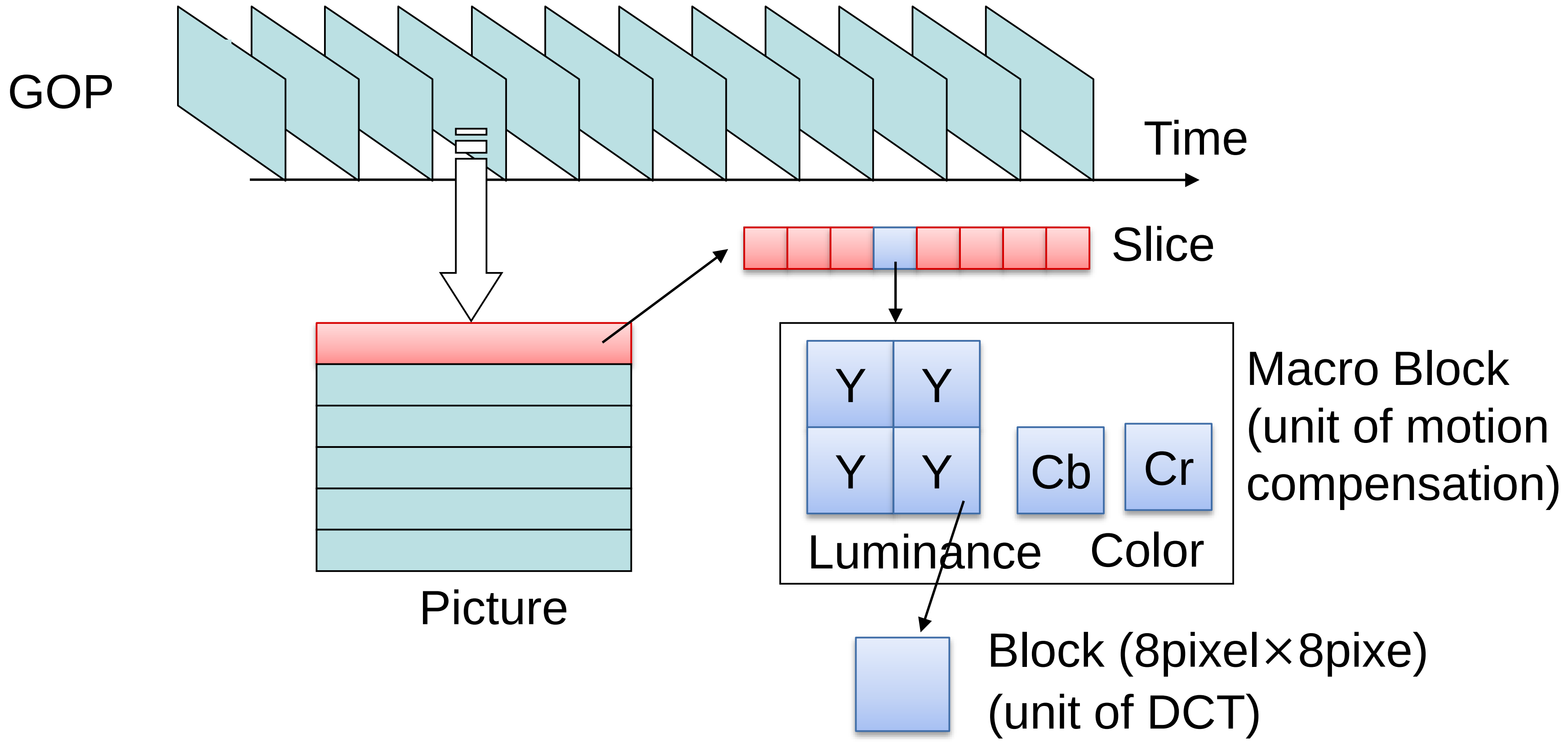
New CODEC is invented to improve those groups.

OTHER BASIC VIDEO CODING TECHNOLOGY (1)



OTHER BASIC VIDEO CODING TECHNOLOGY (1)

Block	: 8[pixel]x8[line]pixel
Macro block	: 16[pixel]x16[line]pixel, consists of several blocks
Slice	: consists of several macro blocks
Picture or VOP	: (Video object plane) one video image
GOP or VOL	: Group Of Picture



OVERVIEW: MPEG-1 VIDEO CODEC

MPEG-1 VIDEO

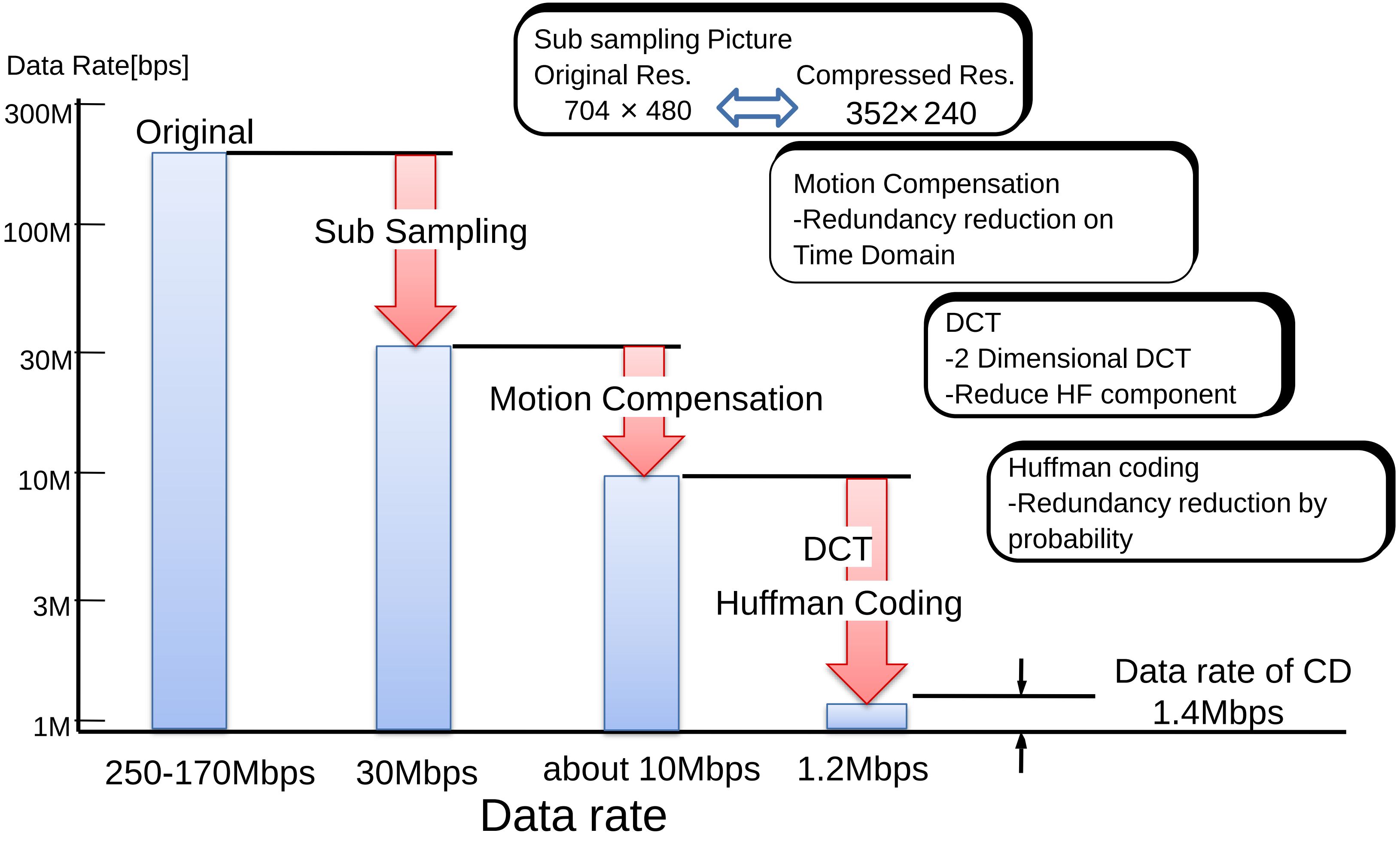
- MPEG1 video compression standard is developed by ISO/MPEG to realize video encode/decode with data rate about 1.5Mbps using compact disk (CD).

- Coding algorithm of MPEG1 Video consists of 3 major technologies:
  - **Motion compensation (MC):** Redundancy reduction in Time domain
  - **Conversion coding (DCT)**: Redundancy reduction in Frequency domain
  - **Entropy coding**: Redundancy reduction by Probability/Statistics such as Huffman coding, Run length coding

Items	Specifications
Source format	SIF(Source input format); 352x240x30 / 352x288x25
Video format	Progressive (Frame picture)
Color format	(4:2:0)
Bitrate	About 1.5 Mbps
Target media	CD (Compact Disc) / DAT
Target quality	VCR (Video Cassette Recorder)

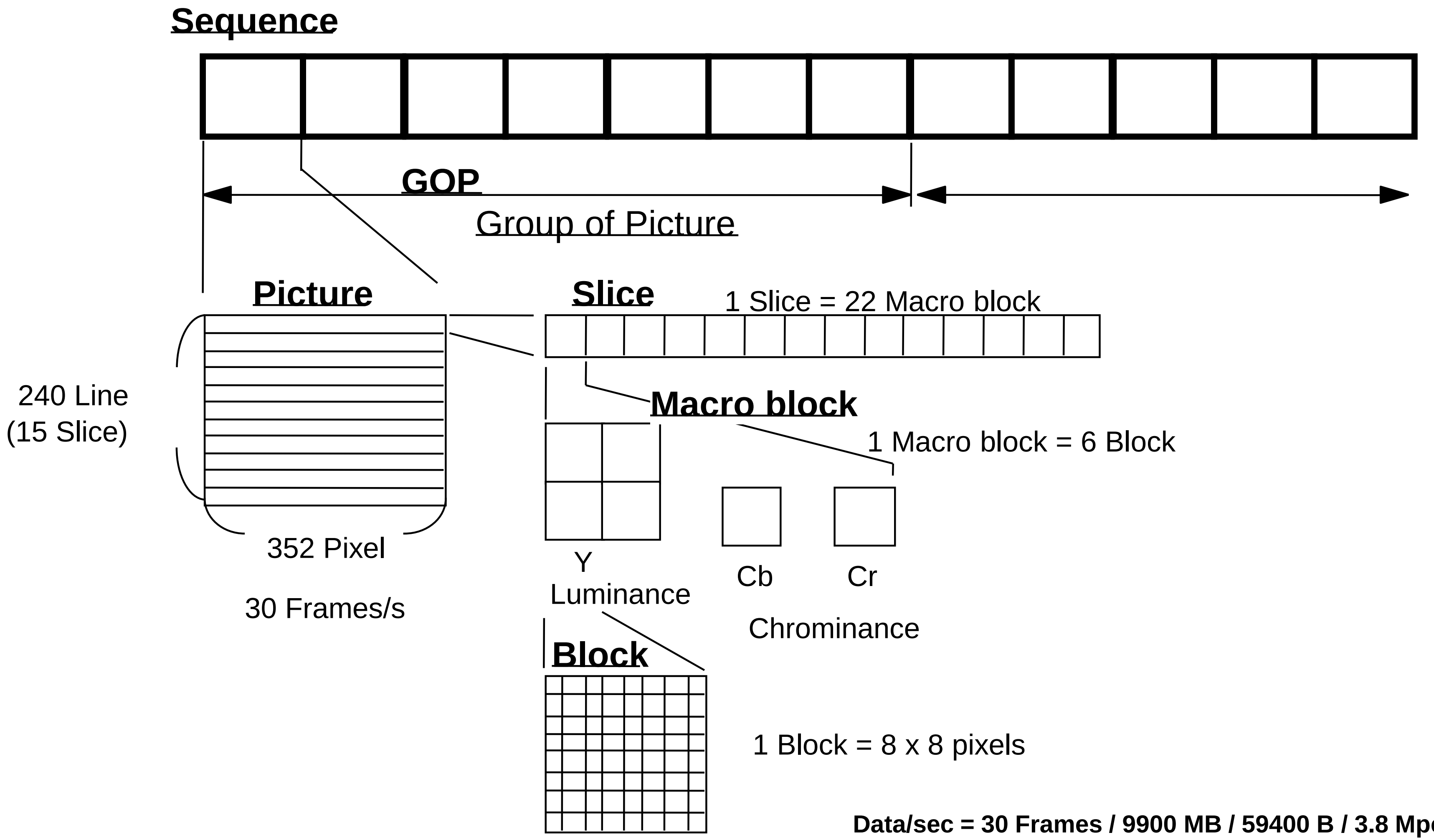
MPEG-1 VIDEO

- The migrate path to reduce data rate is shown below:



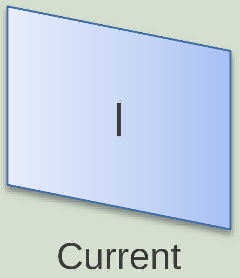
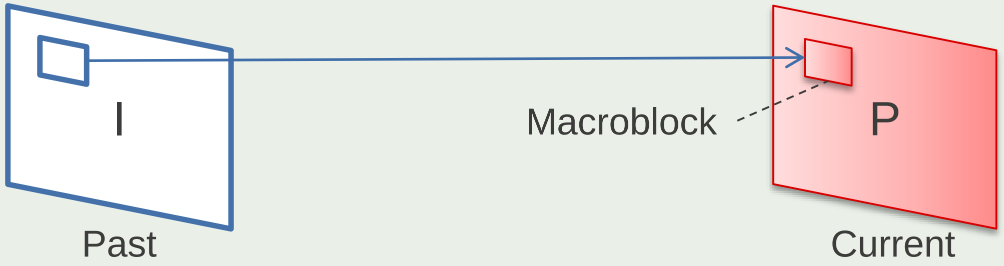
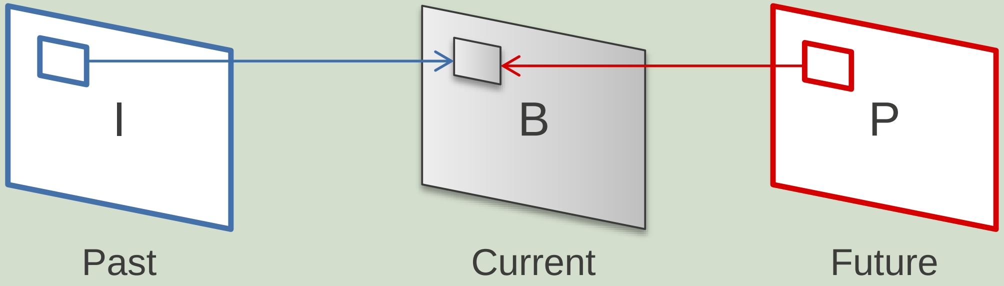
MPEG-1 VIDEO

- The MPEG1 video has following layers:

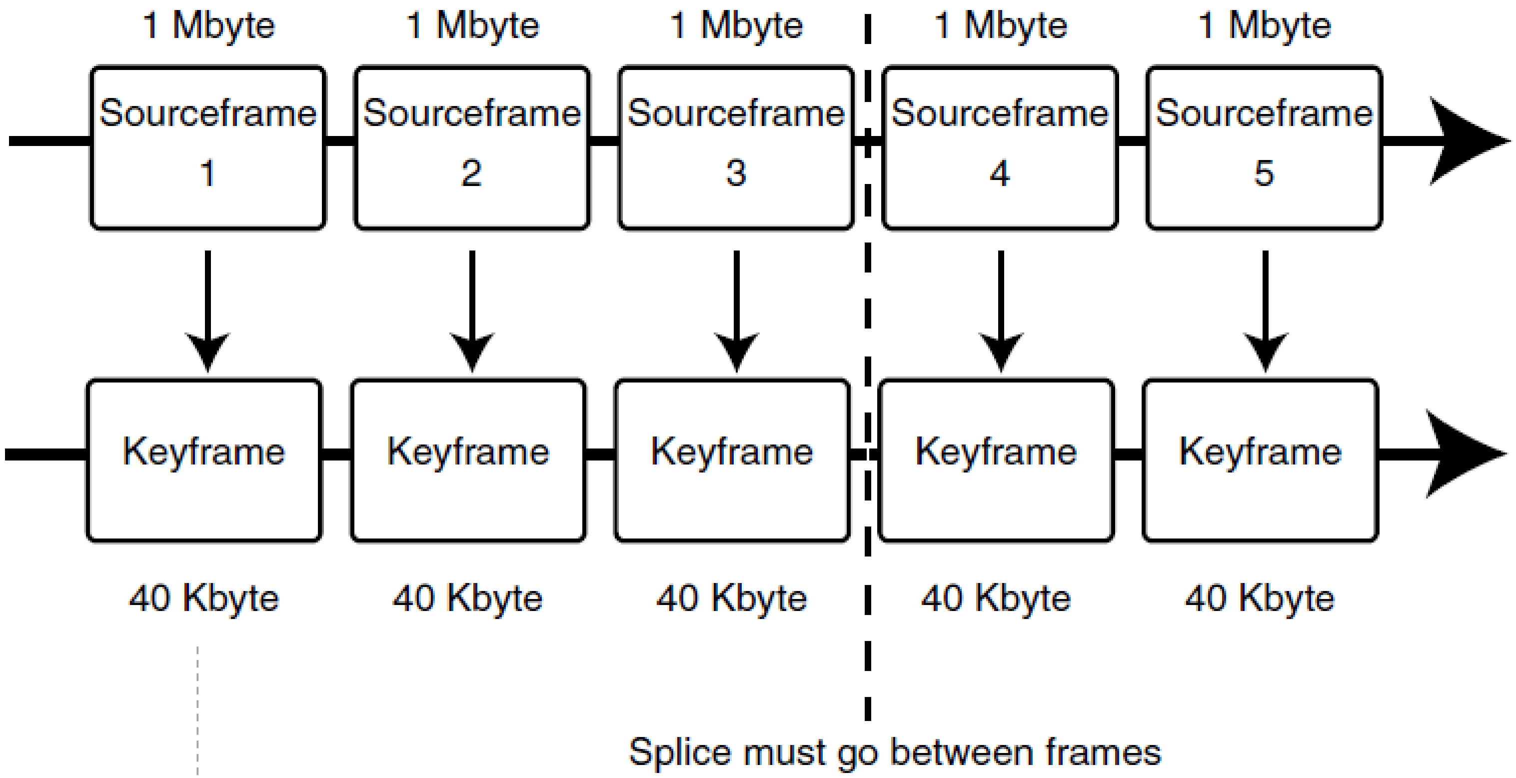


Data/sec = 30 Frames / 9900 MB / 59400 B / 3.8 Mpel

MPEG-1 VIDEO

Picture types	Prediction types	Reference relation
I-Picture/ Keyframe	Intra-Frame Prediction	
P-Picture	Inter-Frame Prediction	
B-Picture	Inter-Frame Bi-direction Prediction	

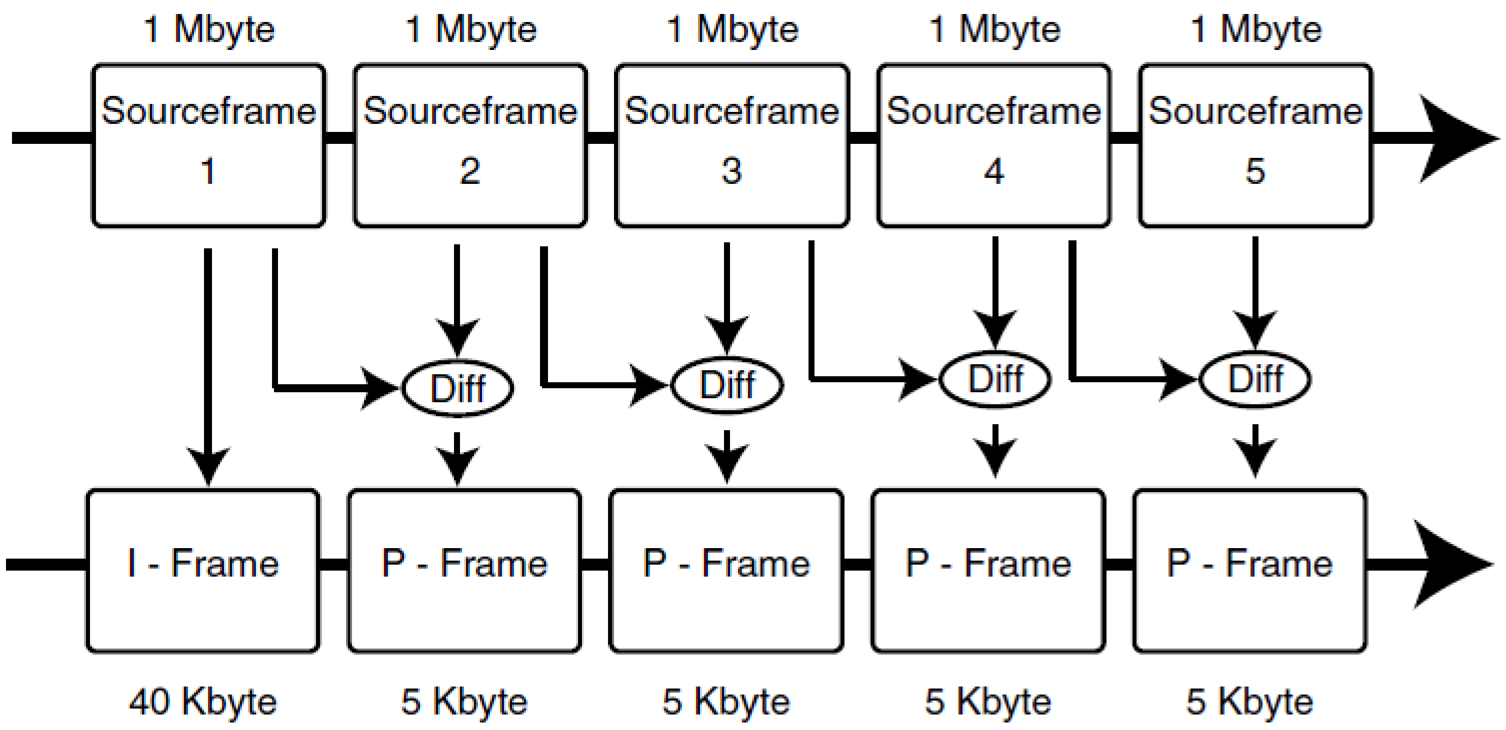
MPEG-1 VIDEO: I VS P VS B



Efficiency

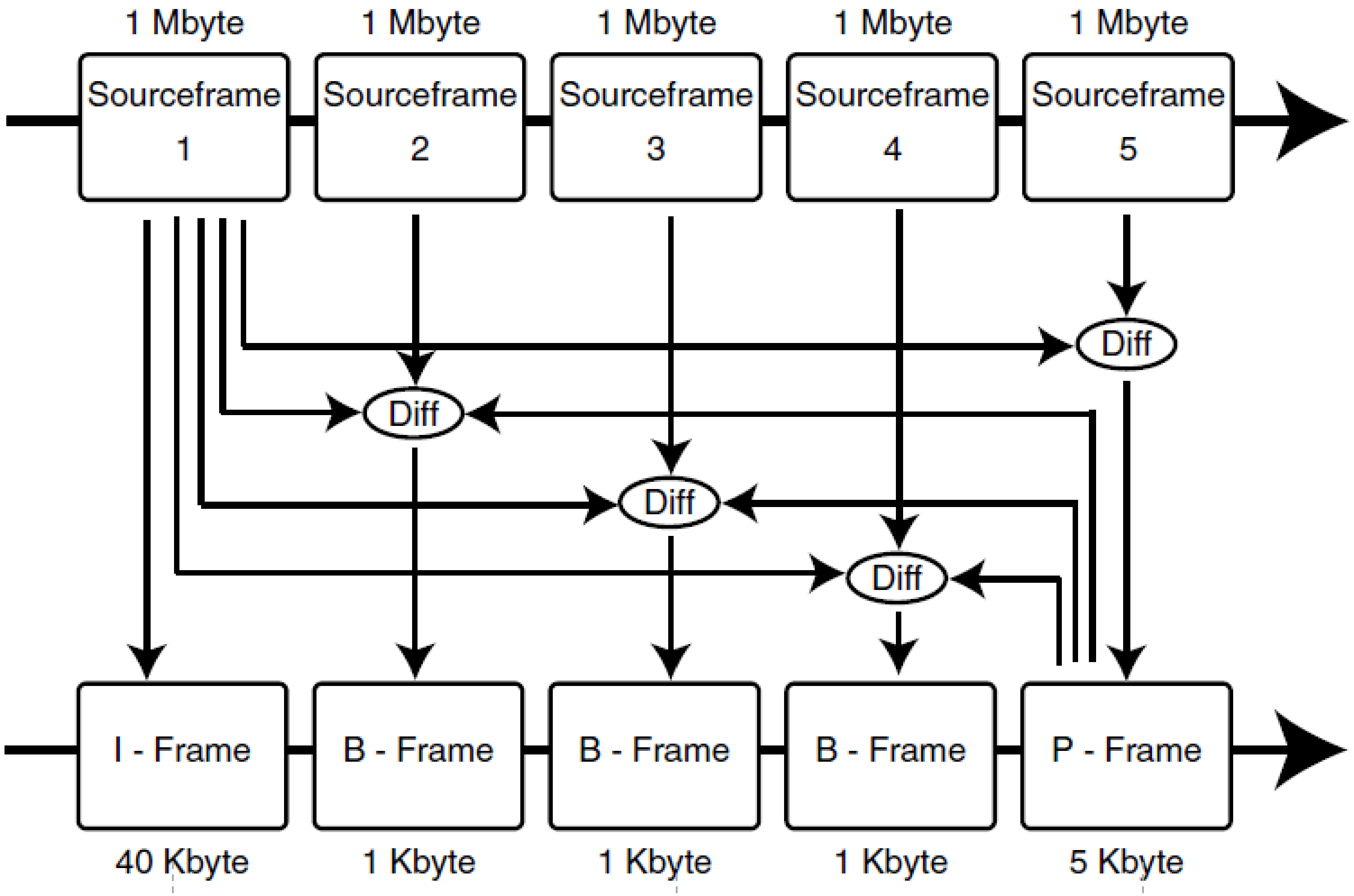
25

MPEG-1 VIDEO: I VS P VS B



Efficiency 25 200

MPEG-1 VIDEO: I VS P VS B



Efficiency

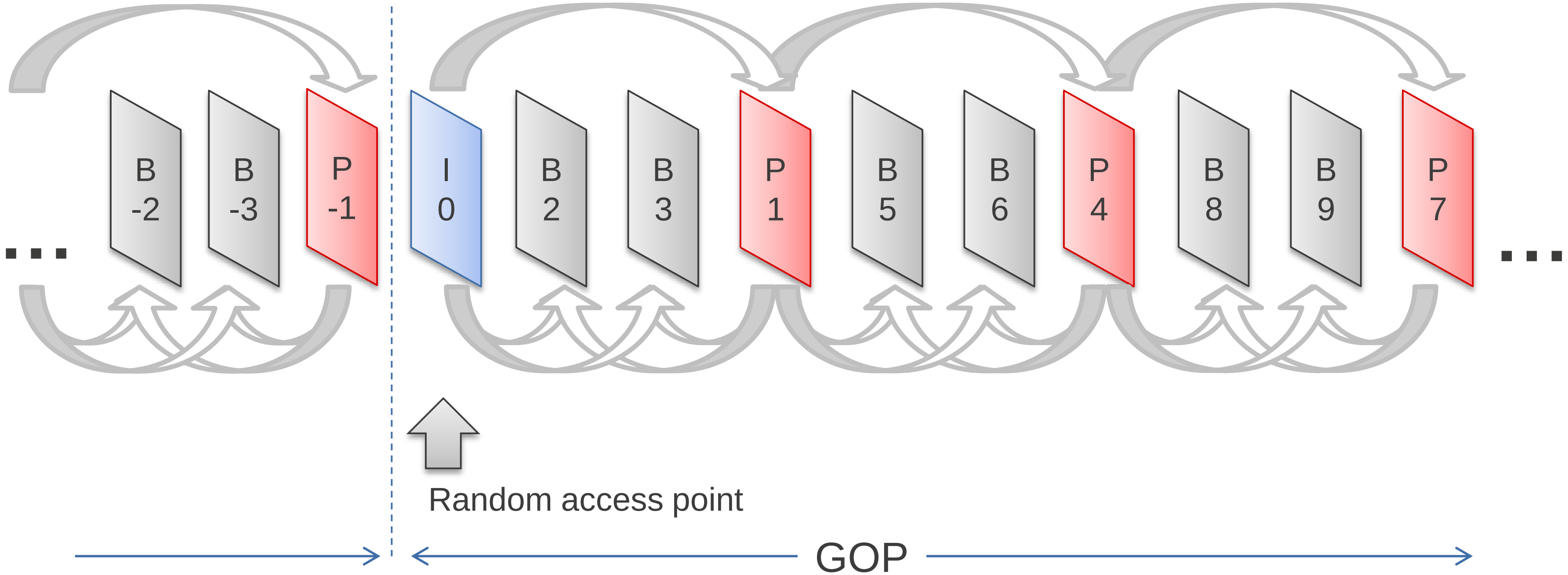
25

1000

200

MPEG-1 VIDEO

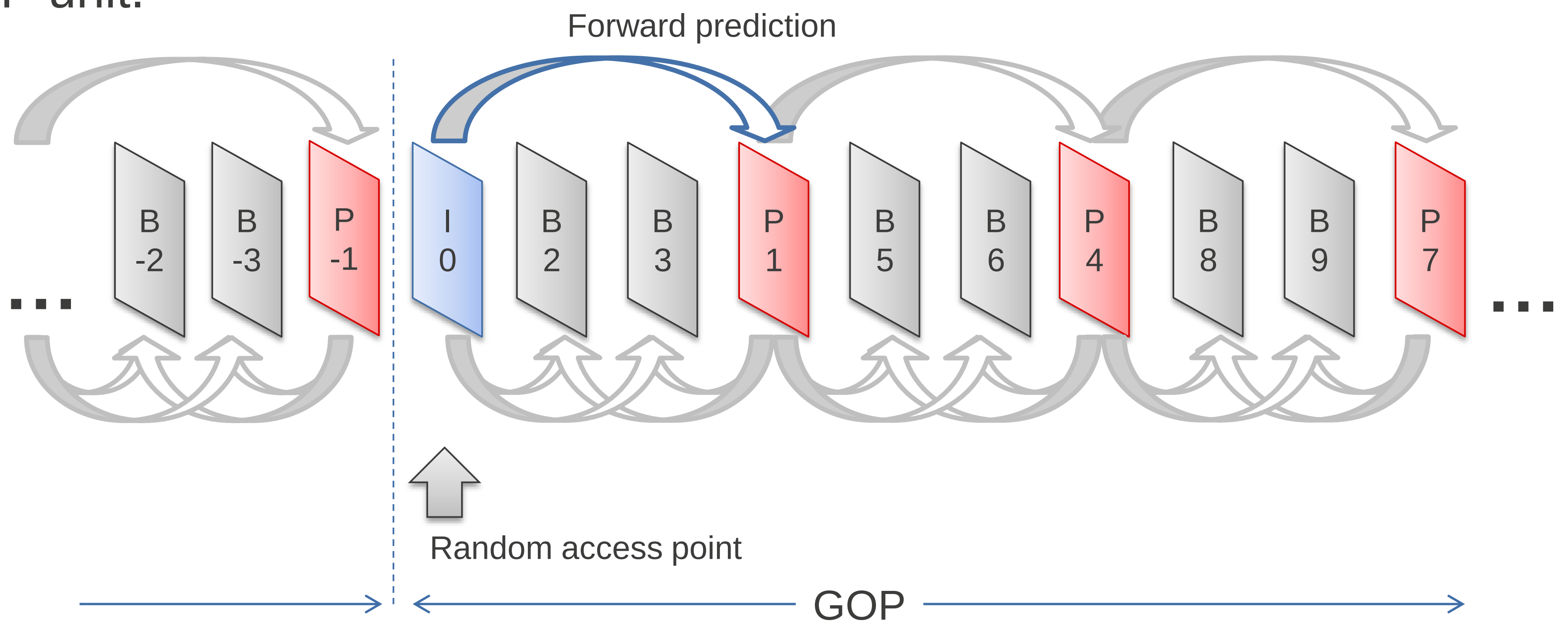
- Prediction structure and GOP (Group of Picture): Prediction structure is based on a GOP unit.



Adaptive selection of I/P/B prediction for each Macro Block

MPEG-1 VIDEO

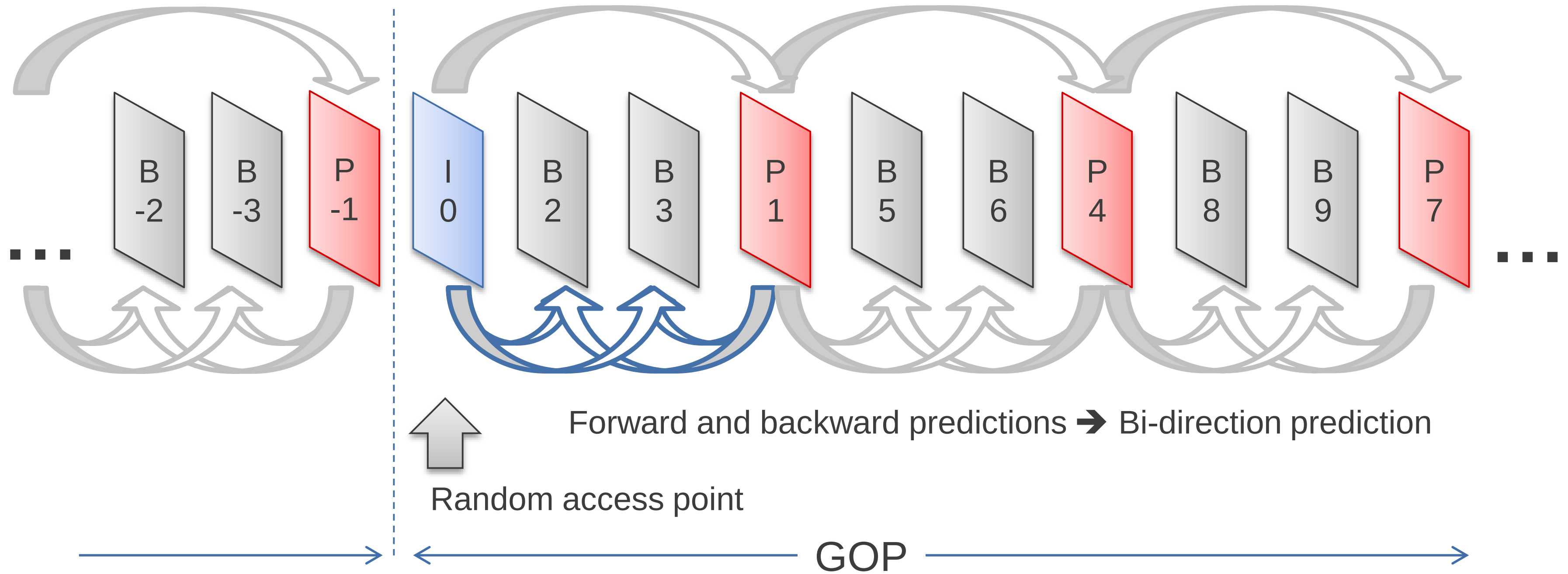
- Prediction structure and GOP (Group of Picture): Prediction structure is based on a GOP unit.



Adaptive selection of I/P/B prediction for each Macro Block

MPEG-1 VIDEO

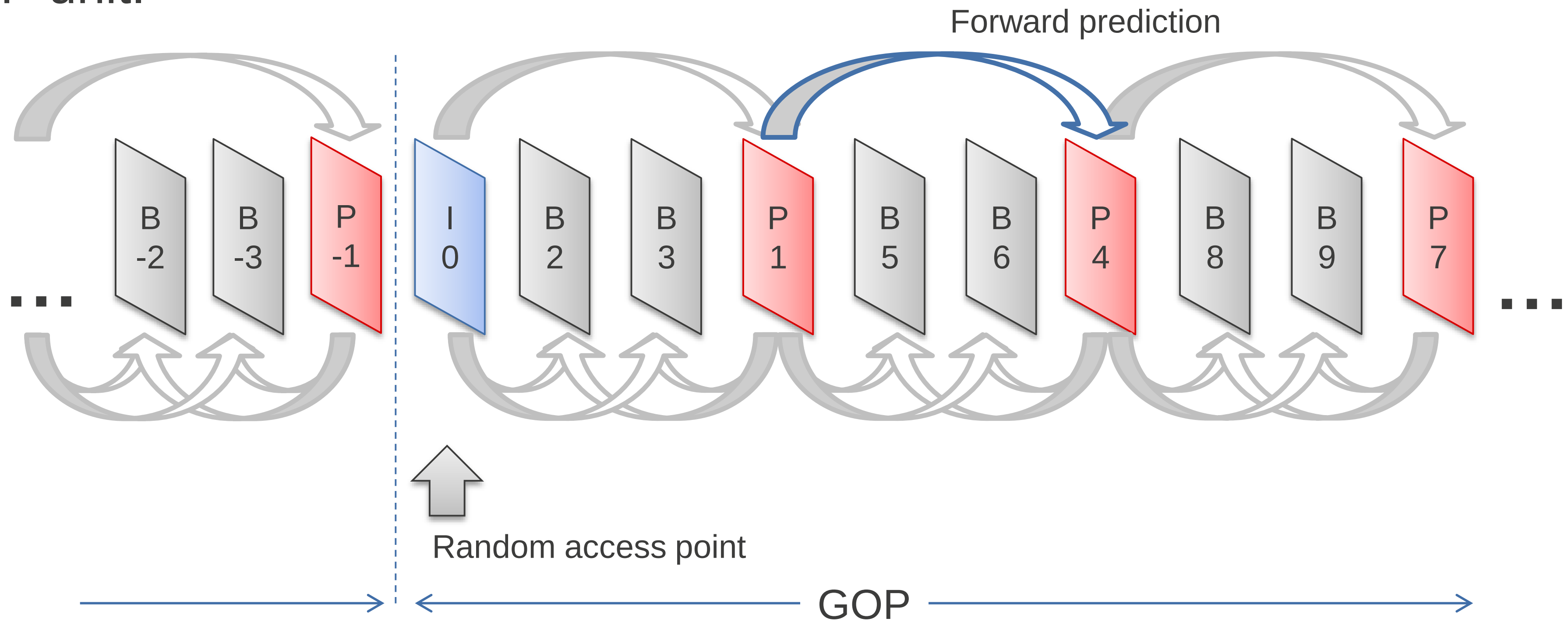
- Prediction structure and GOP (Group of Picture): Prediction structure is based on a GOP unit.



Adaptive selection of I/P/B prediction for each Macro Block

MPEG-1 VIDEO

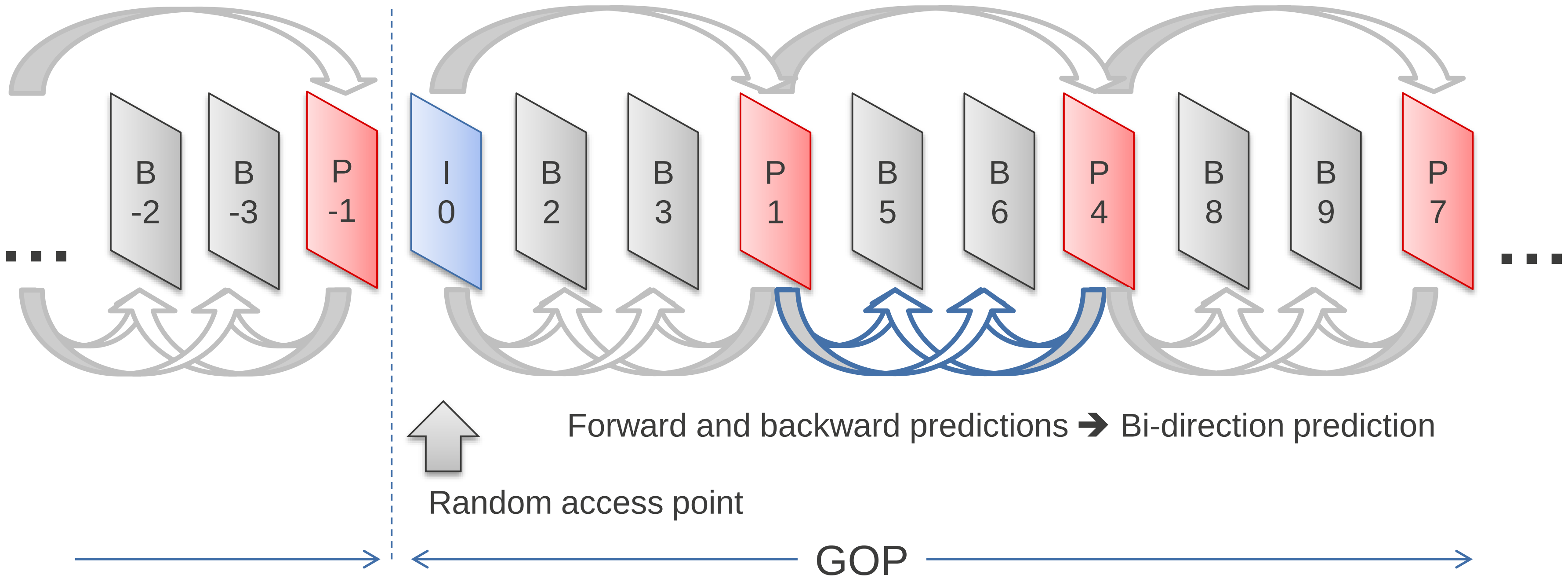
- Prediction structure and GOP (Group of Picture): Prediction structure is based on a GOP unit.



Adaptive selection of I/P/B prediction for each Macro Block

MPEG-1 VIDEO

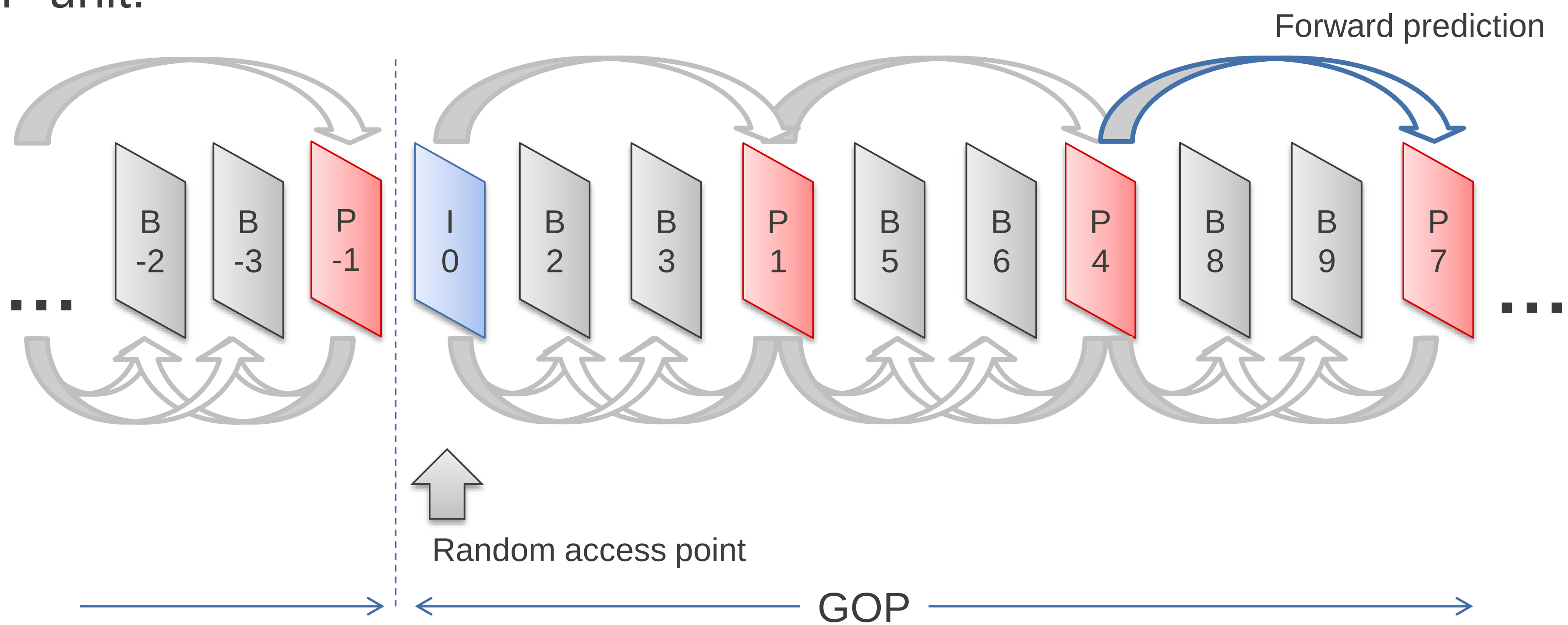
- Prediction structure and GOP (Group of Picture): Prediction structure is based on a GOP unit.



Adaptive selection of I/P/B prediction for each Macro Block

MPEG-1 VIDEO

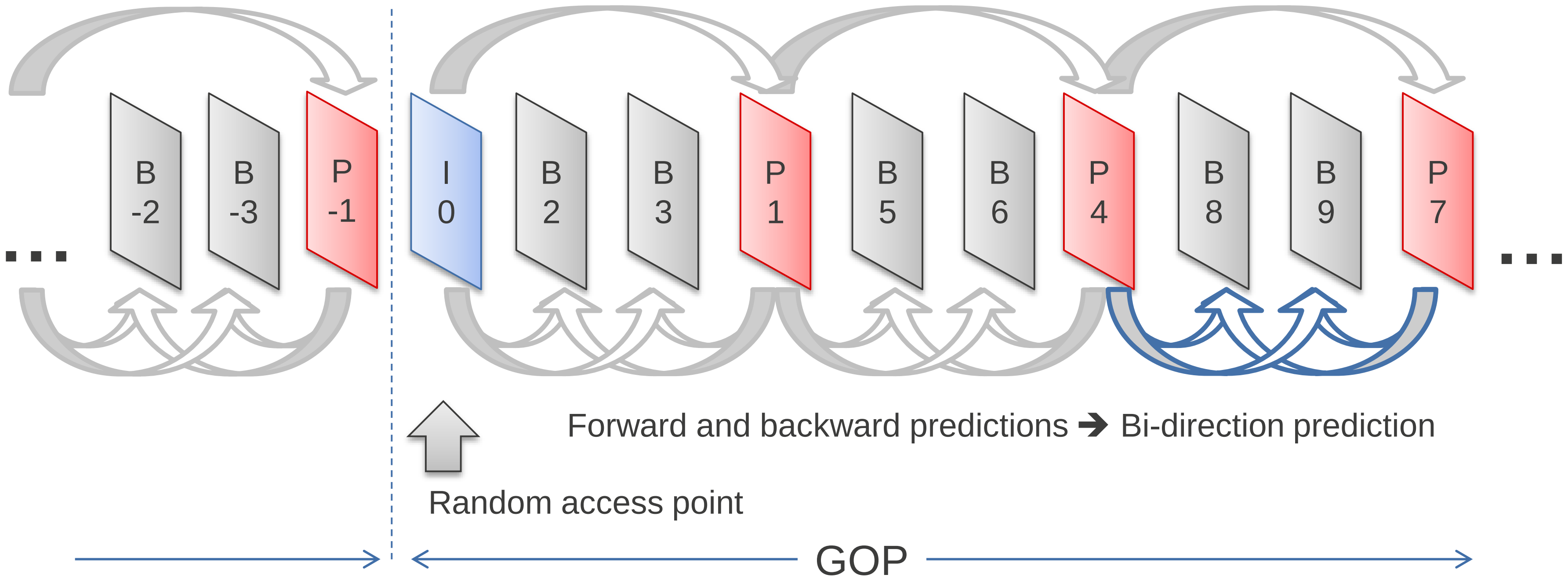
- Prediction structure and GOP (Group of Picture): Prediction structure is based on a GOP unit.



Adaptive selection of I/P/B prediction for each Macro Block

MPEG-1 VIDEO

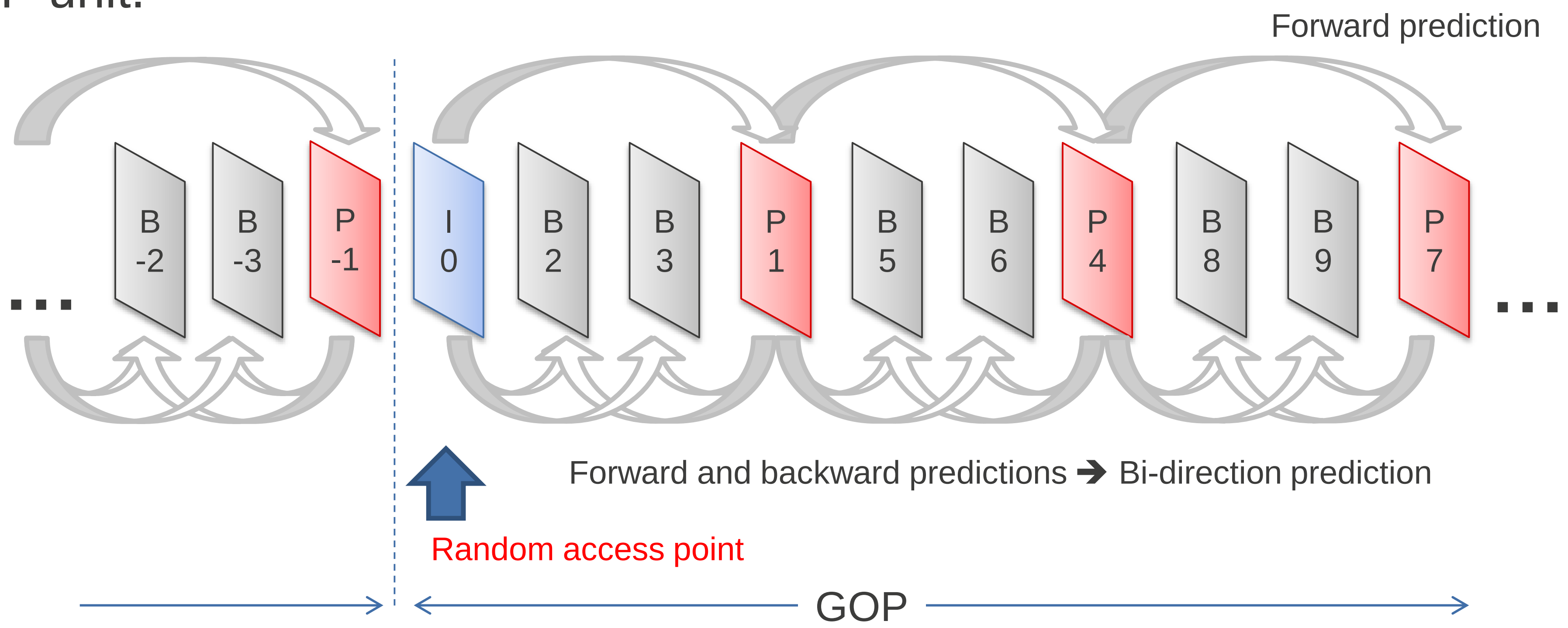
- Prediction structure and GOP (Group of Picture): Prediction structure is based on a GOP unit.



Adaptive selection of I/P/B prediction for each Macro Block

MPEG-1 VIDEO

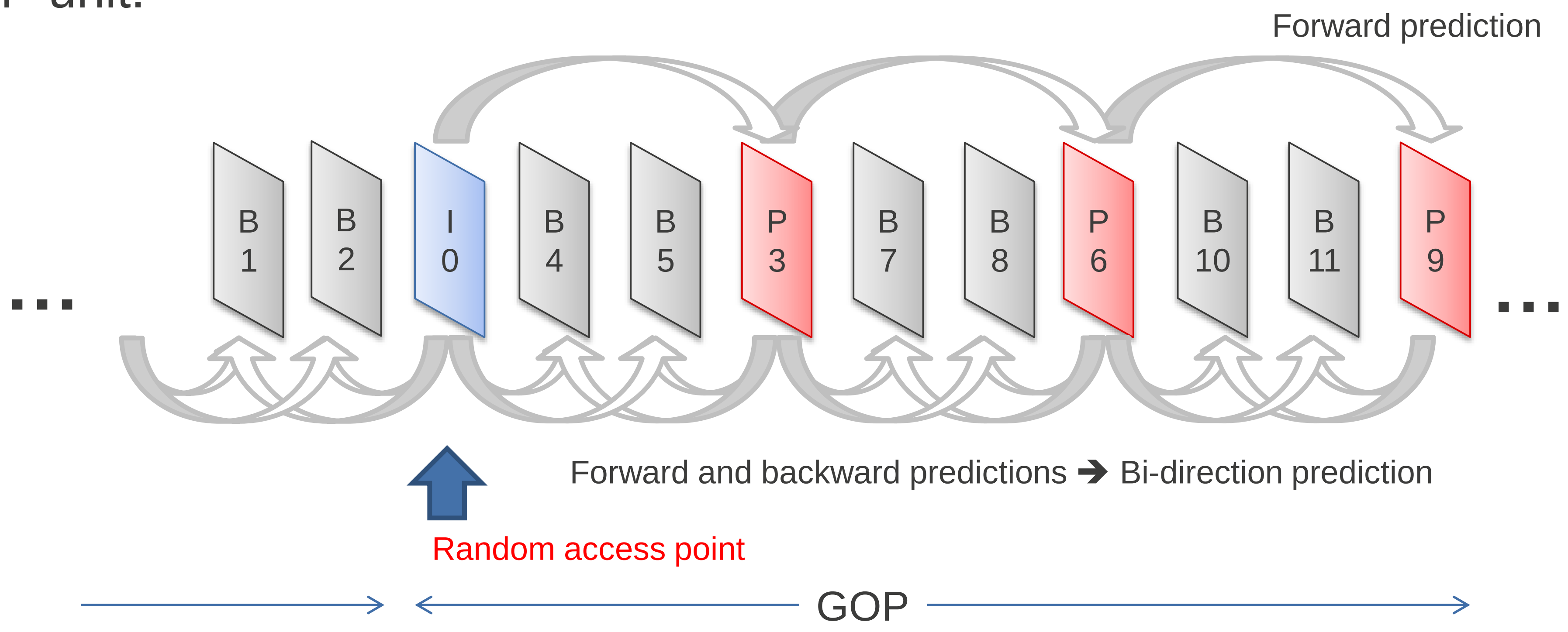
- Prediction structure and GOP (Group of Picture): Prediction structure is based on a GOP unit.



Adaptive selection of I/P/B prediction for each Macro Block

MPEG-1 VIDEO

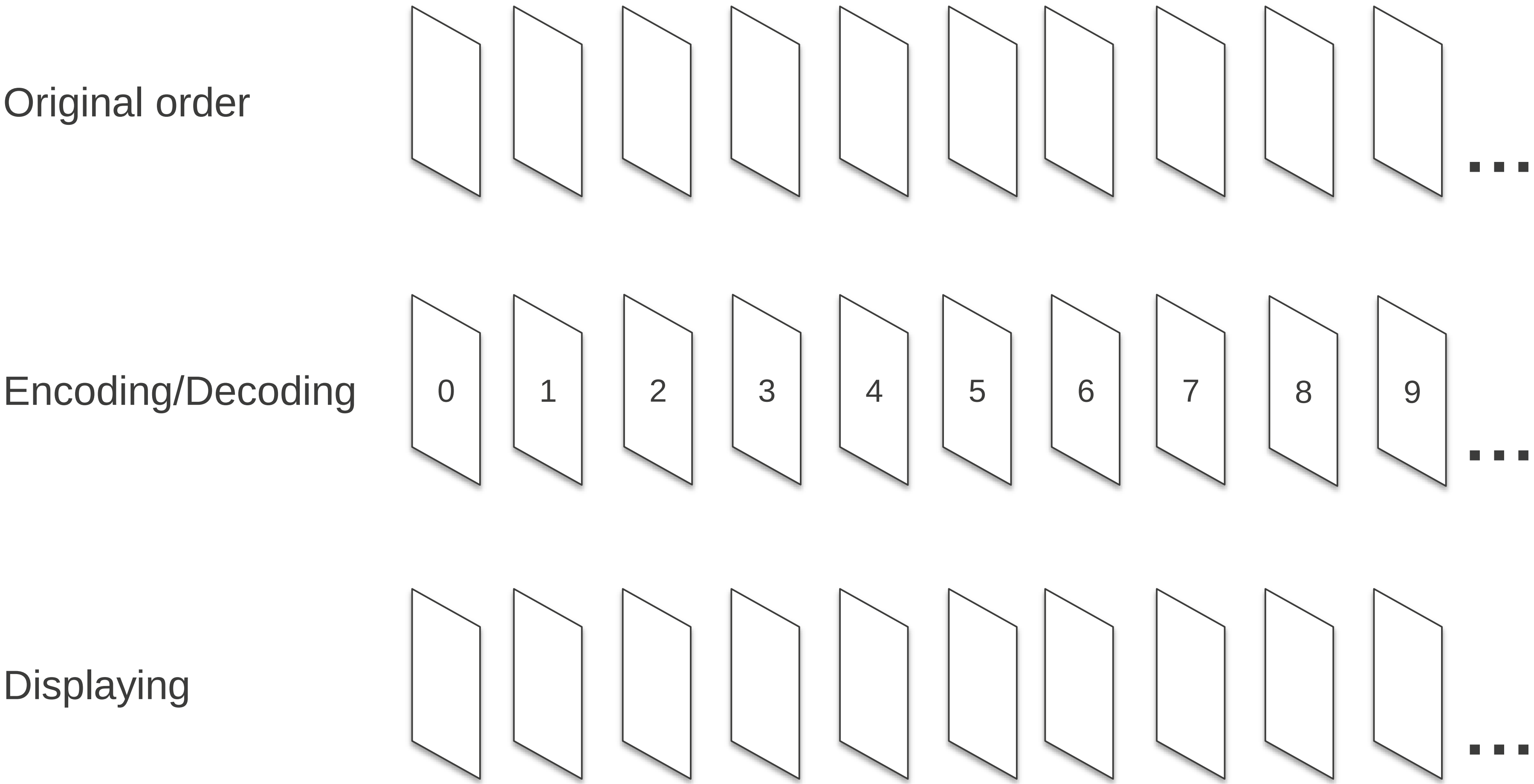
- Prediction structure and GOP (Group of Picture): Prediction structure is based on a GOP unit.



Adaptive selection of I/P/B prediction for each Macro Block

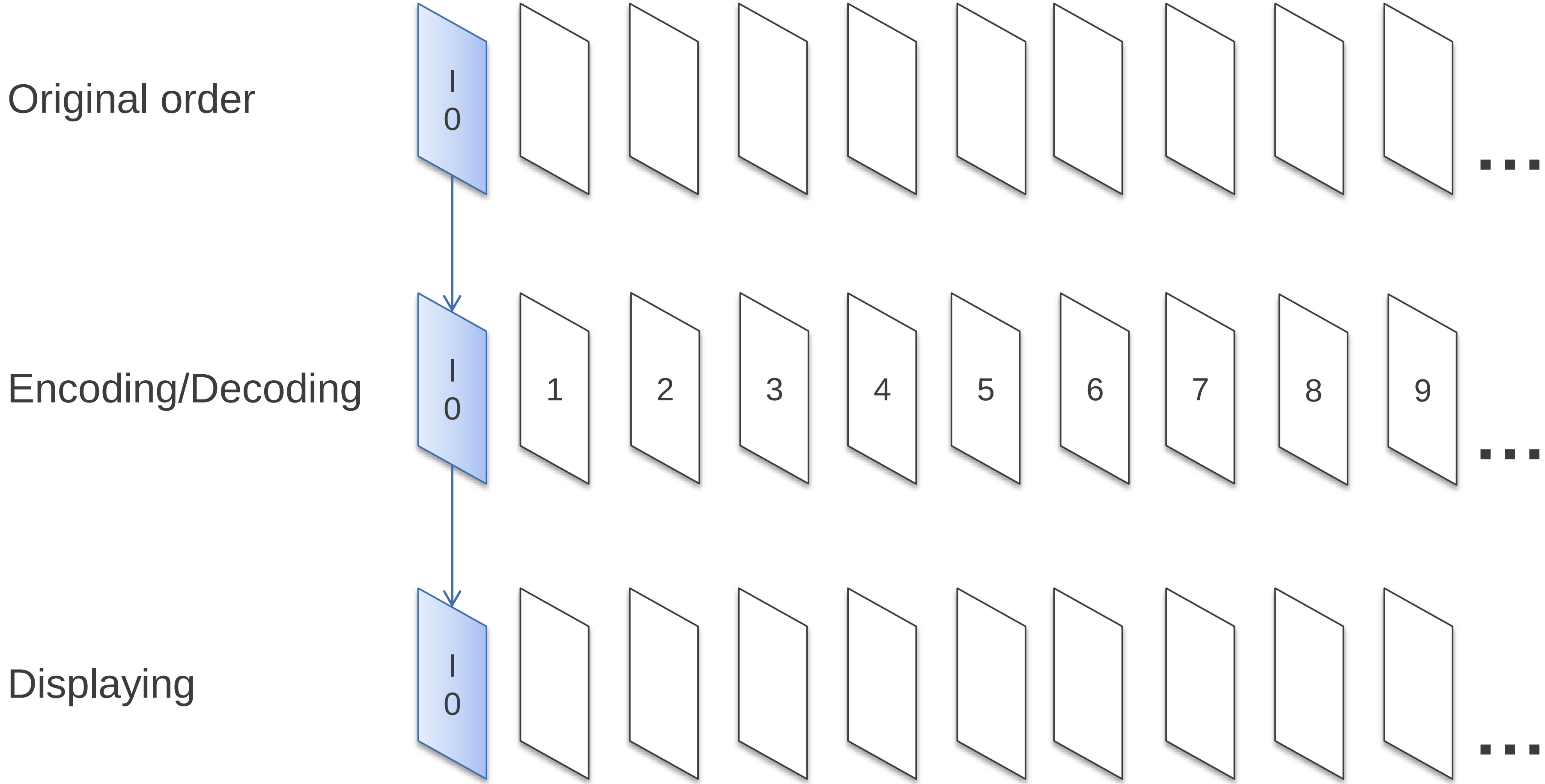
MPEG-1 VIDEO

- Picture re-ordering – The difference between decoding / encoding order and displaying order.



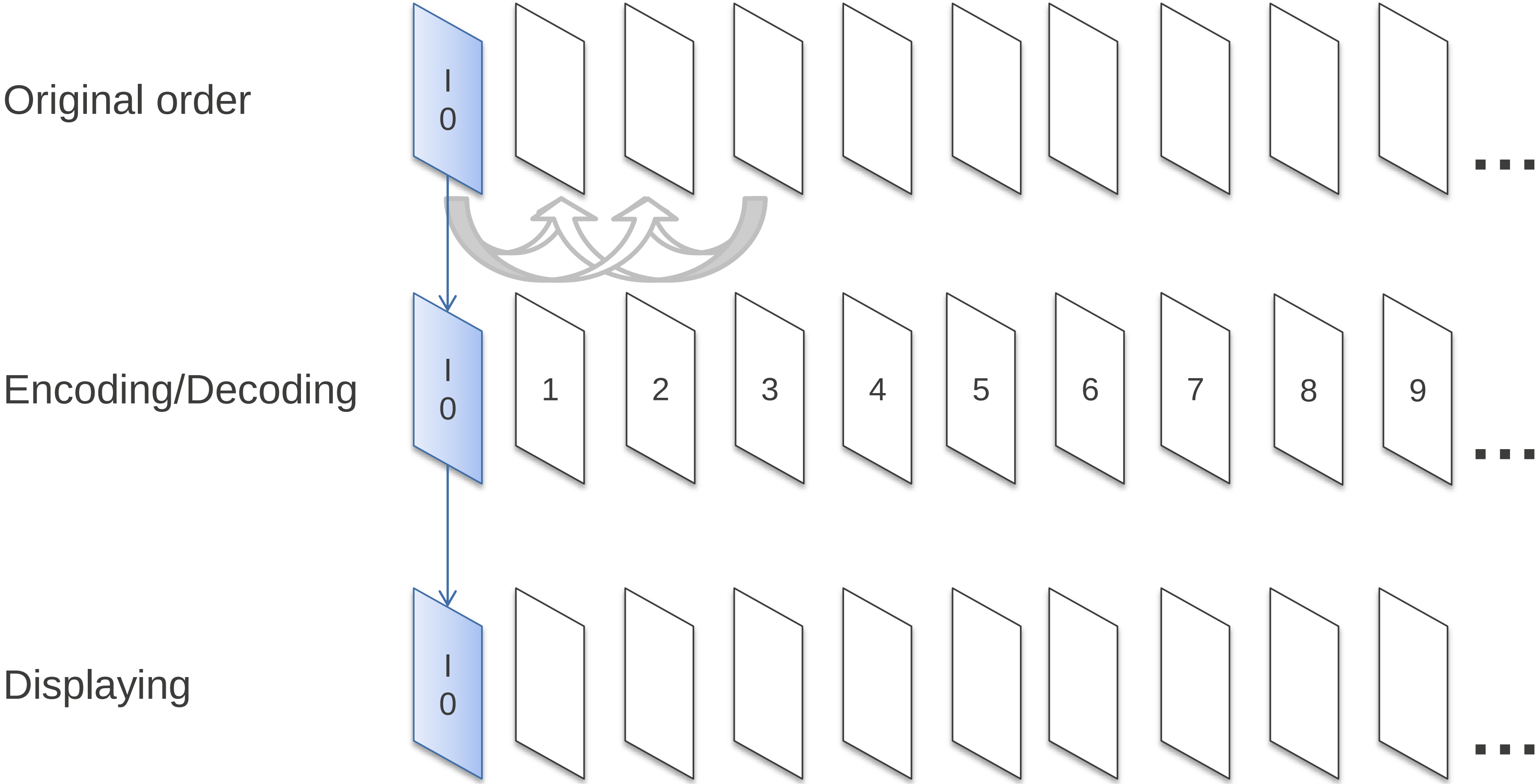
MPEG-1 VIDEO

- Picture re-ordering – The difference between decoding / encoding order and displaying order.



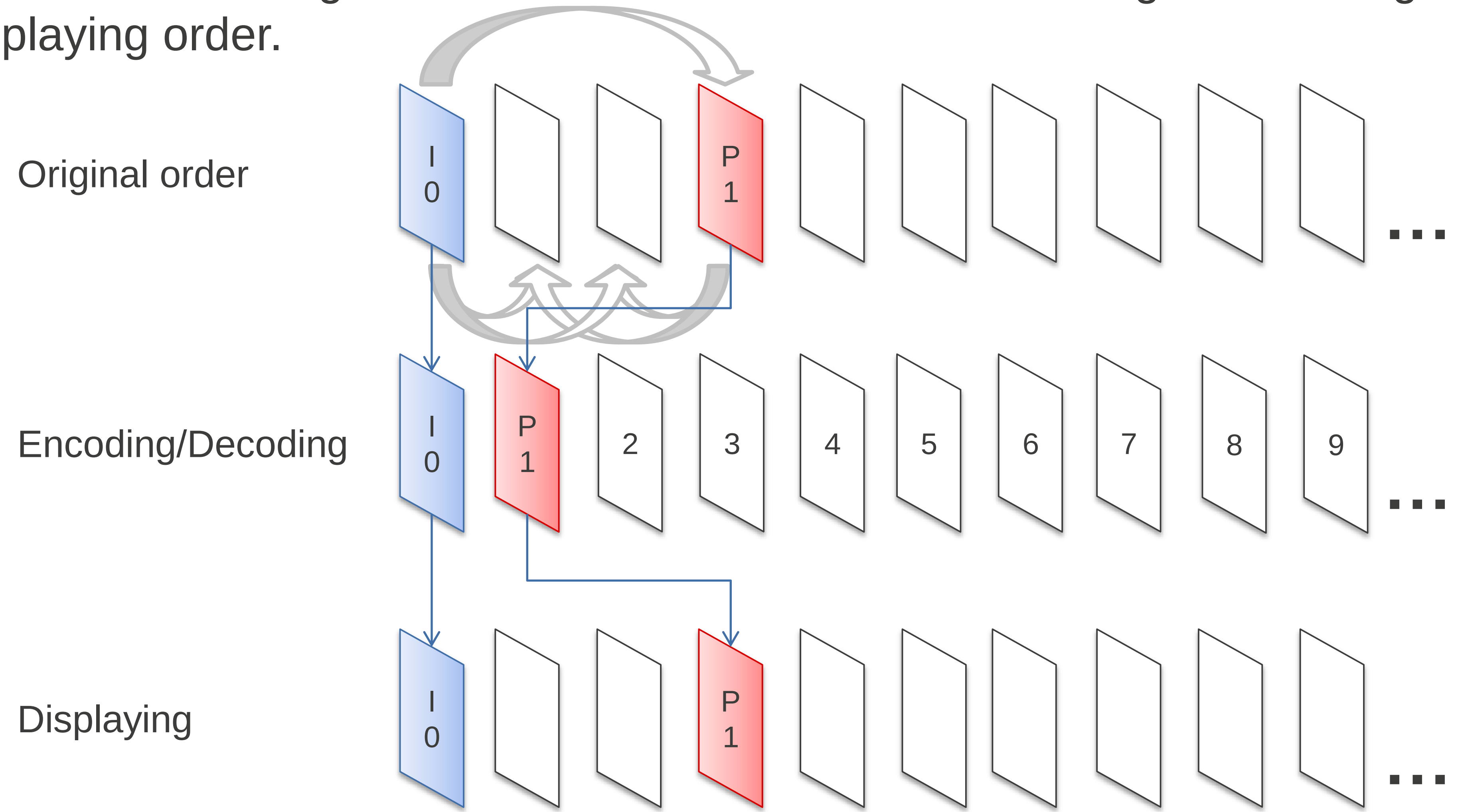
MPEG-1 VIDEO

- Picture re-ordering – The difference between decoding / encoding order and displaying order.



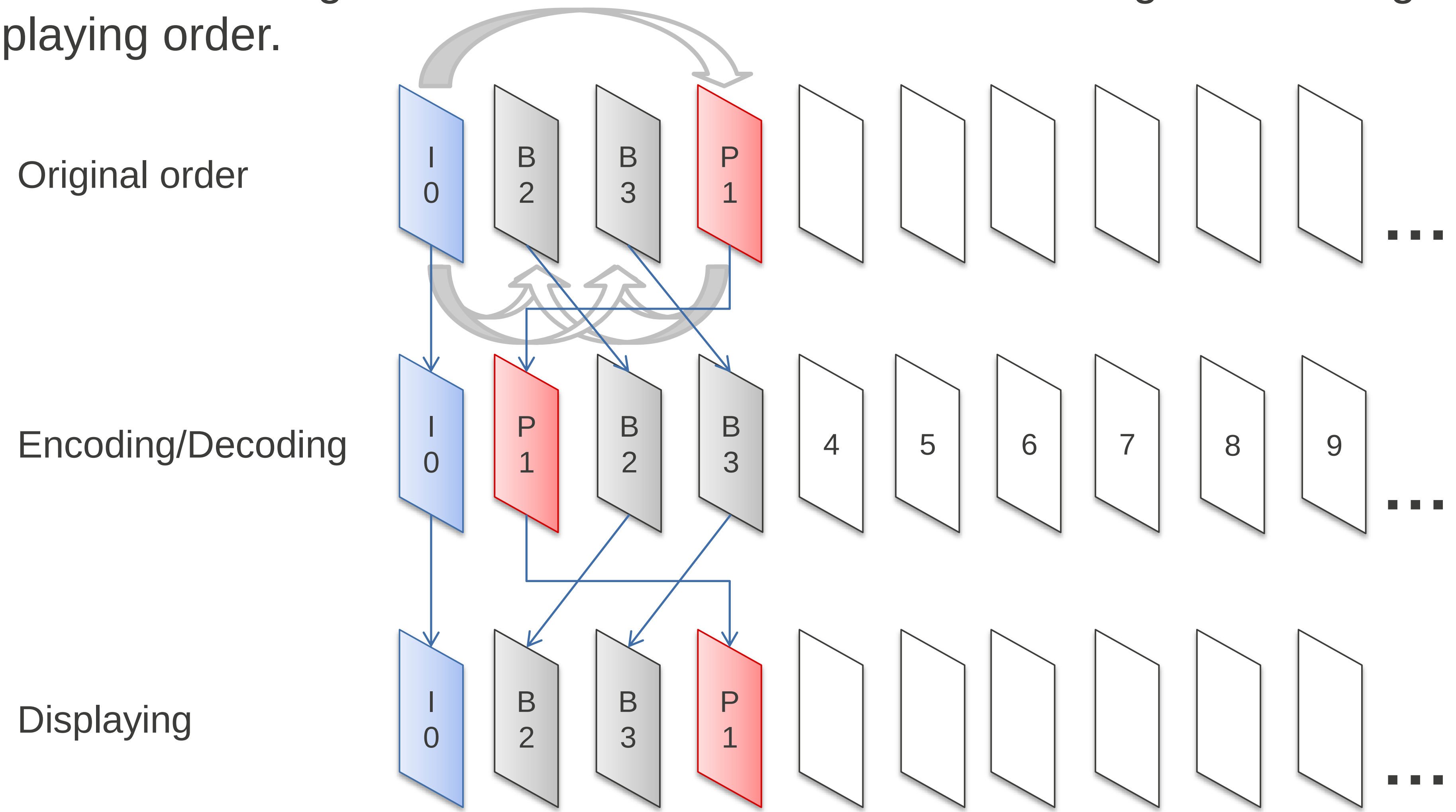
MPEG-1 VIDEO

- Picture re-ordering – The difference between decoding / encoding order and displaying order.



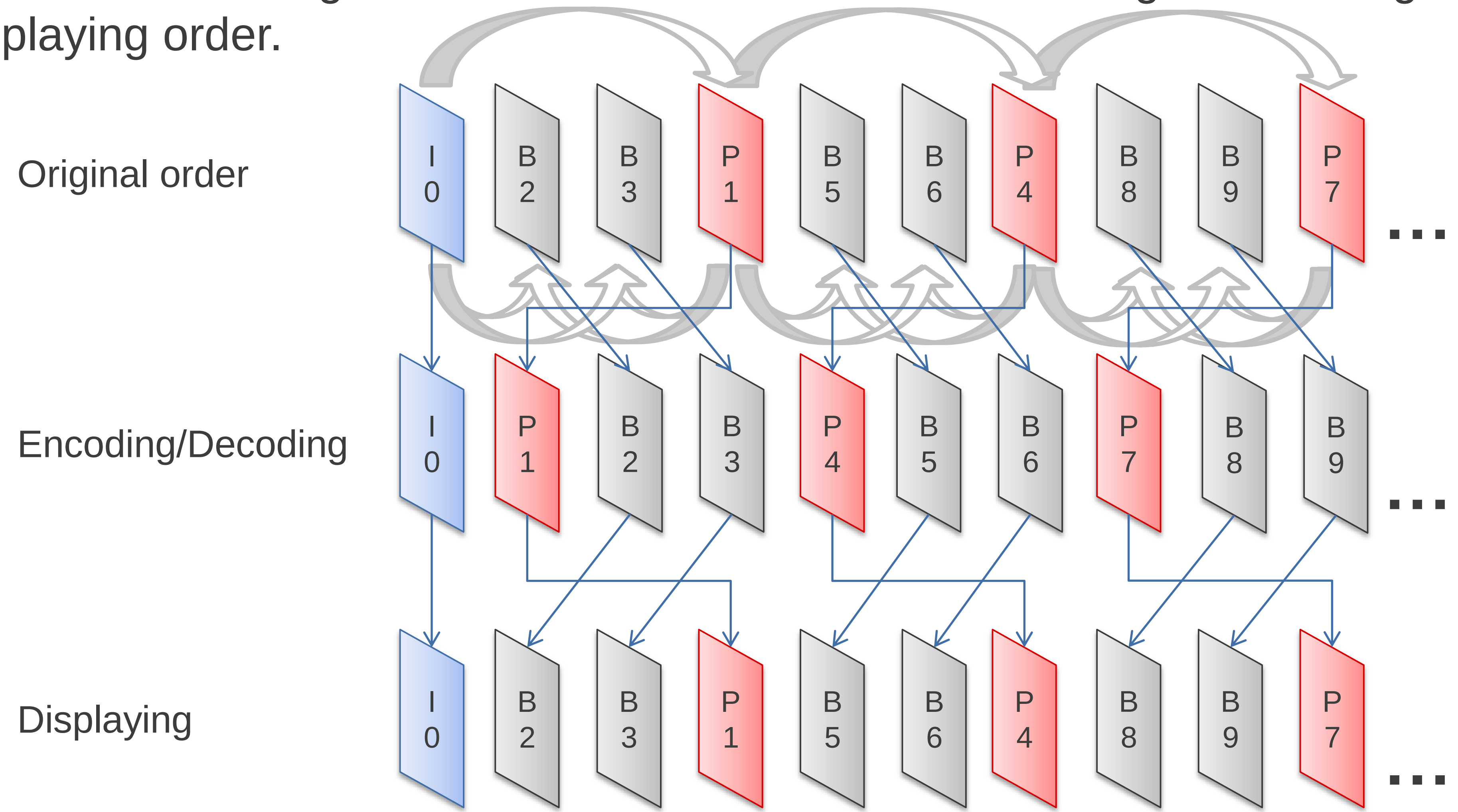
MPEG-1 VIDEO

- Picture re-ordering – The difference between decoding / encoding order and displaying order.



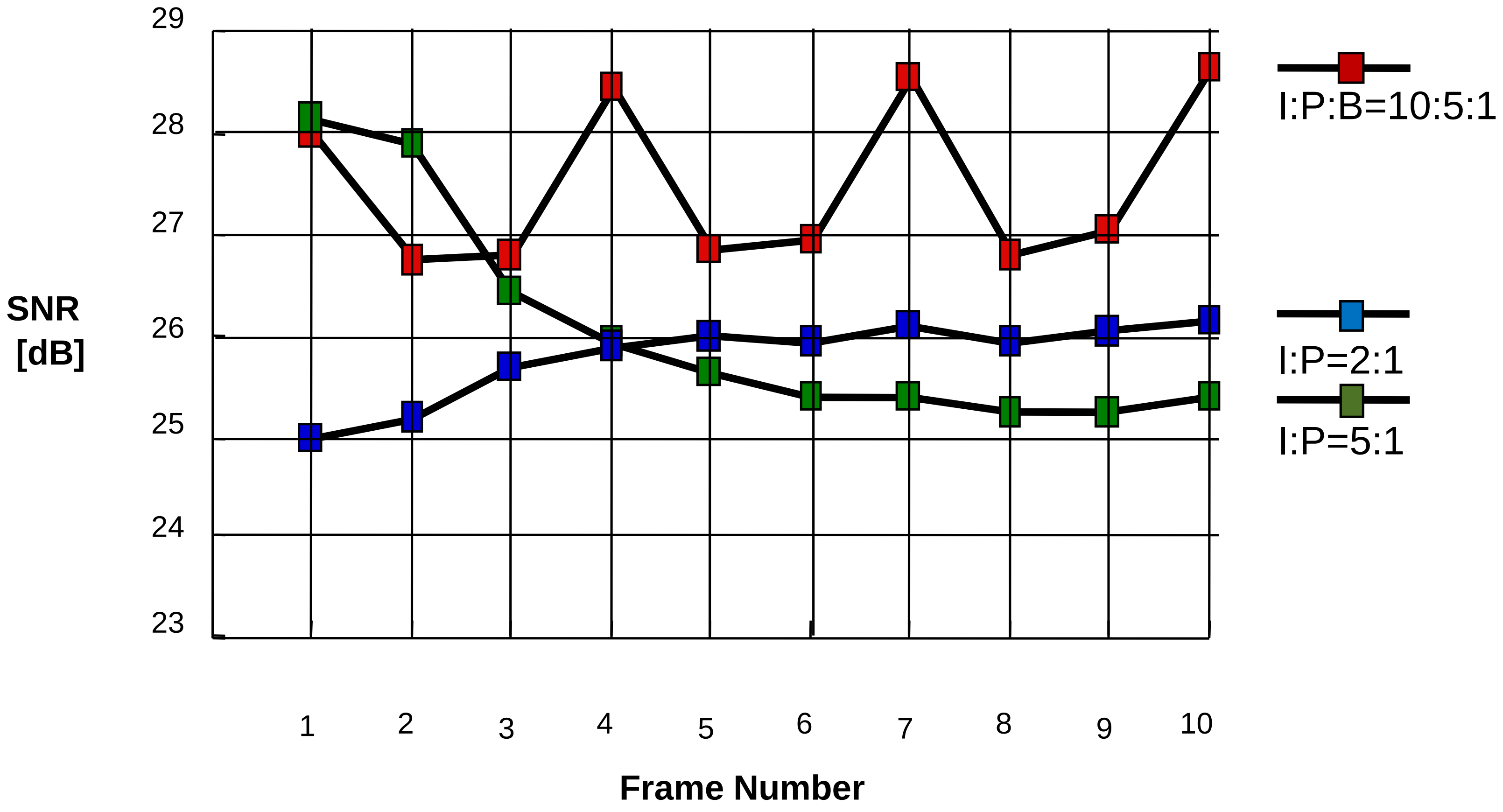
MPEG-1 VIDEO

- Picture re-ordering – The difference between decoding / encoding order and displaying order.



MPEG-1 VIDEO

- Efficiency of using B-Picture
  - Assign many data to I and P to make good picture quality for I/P
  - B picture will be improved by adjacent I/P picture



OVERVIEW: MPEG-2 VIDEO CODEC

MPEG-2 VIDEO

- The target application of MPEG2 Video standard is Digital TV or Digital Optical Disc system.
- MPEG2 Video also supports HDTV & SDTV. (MPEG3 is supposed to be HDTV coding standard, but it is skipped.)

Items	Specifications
Source format	SDTV / HDTV; SIF is also supported
Video format	Progressive / Interlace (Frame/Field picture)
Color format	(4:2:0); (4:2:2) is optional
Bitrate	Max.15 (SDTV) – 80 (HDTV) Mbps
Target media	Digital TV / Optical Disc
Target quality	Studio quality

MPEG-2 VIDEO

❑ Comparison between MPEG1 and MPEG2

	MPEG1	MPEG2
System	• Package Media only • Program Stream only • Unique system clock	• Package Media and broadcasting • Program Stream and Transport Stream • Multiple Service (channel)
Video	• Frame operation (non interlace)	• Frame/field adaptive operation (Interlace/non interlace) • Profile/Level support • Scalability support
Audio	• Stereo (2 channel) • Musicam	• Multi channel/Surround • Half Sampling Rate • NBC coding(AC-3)

MPEG-2 VIDEO

❑ Profile and Level

High Level 1920x1152 @ 60fps		x	MP@HL 80MBit/s	x	x	HP@HL 100MBit/s
High-1440 Level 1440x1152 @ 60fps		x	MP@HL-1440 60MBit/s US ATSC	x	Spatially@HL-1440 60MBit/s	HP@HL-1440 80MBit/s EU dTTb
Main Level 720x576 @ 30fps		SP@ML 15MBit/s CATV	MP@ML 15MBit/s Generic	SNR@ML 15MBit/s	x	HP@ML 20MBit/s
Low Level 352x288 @ 30fps		x	MP@LL 4MBit/s	SNR@LL 4MBit/s	x	x
		Simple Profile	Main Profile	SNR Scalable Profile	Spatially Scalable Profile	High Profile
Function	B-picture		○	○	○	○
	SNR scalability			○	○	○
	Spatial scalability				○	○
	Temporal scalability					
	4:2:2					○
	Data partitioning					

OVERVIEW: H.264/AVC VIDEO CODEC

H.264/AVC VIDEO

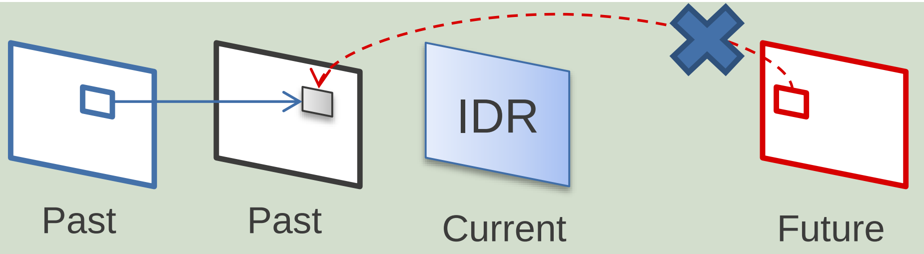
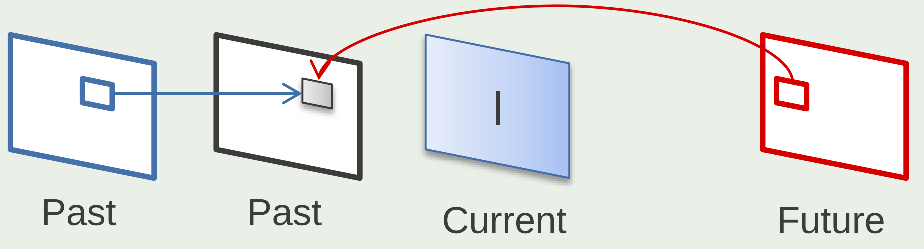
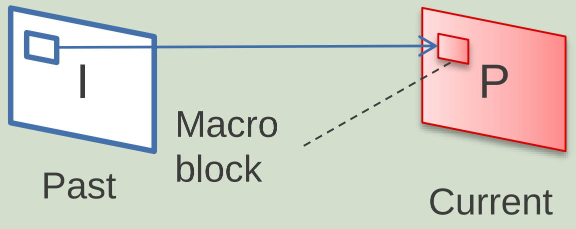
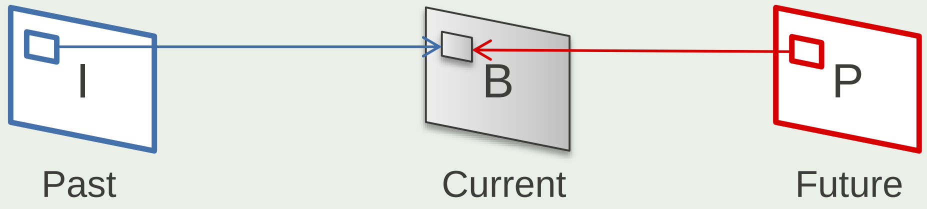
- Improvement of image quality (**70% improve from MPEG-4**)
  - Some reports 2~3 dB improve in PSNR (in the case all tools are used)
- Network　Friendliness
- Points of improvement
  - Intra-prediction
    - increase multi-mode prediction
    - de-blocking filter
    - integer transform by 4x4
  - Inter-prediction
    - variable block-size motion compensation
    - multi-reference
  - Entropy coding
    - CABAC (Context-based Adaptive Binary Arithmetic Coding), but not supported in　Baseline profile
- Separate Coding layer from network layer
  - VCL: Video Coding Layer
  - NAL: Network Adaptation Layer

H.264/AVC VIDEO

- Major differences of video coding methods are listed below

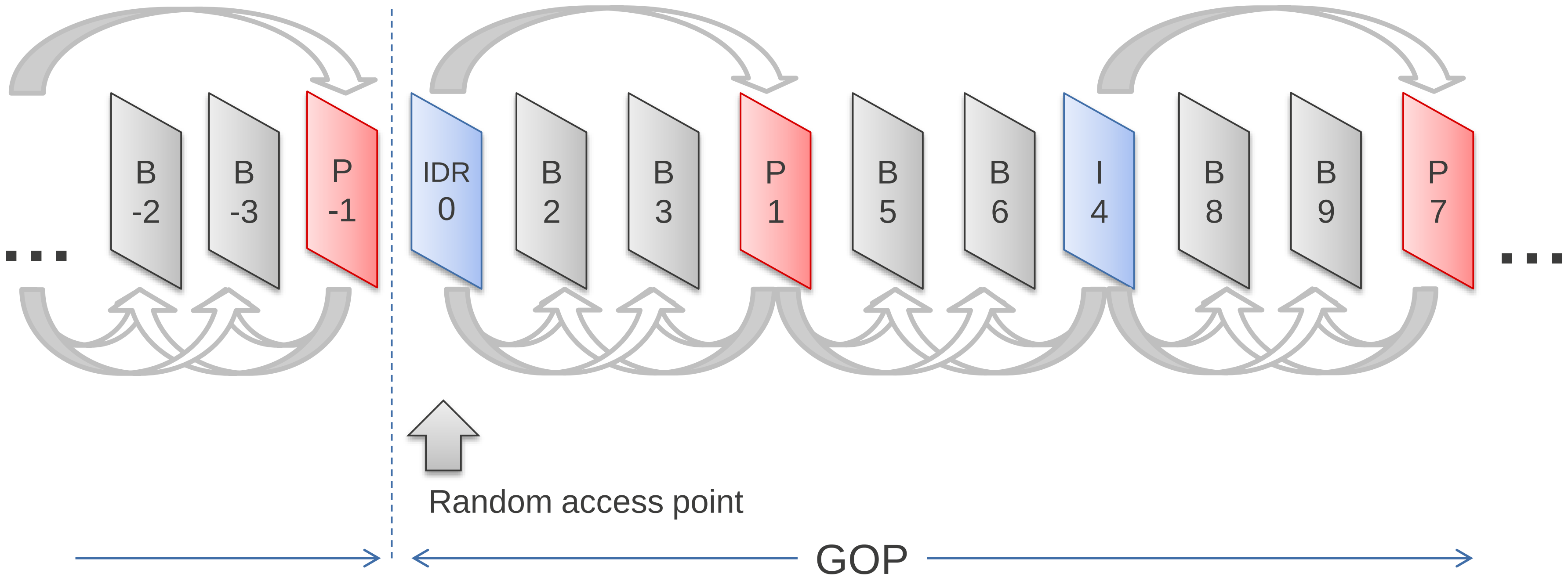
Specifications	MPEG1	MPEG2	H.264
Source format	Frame	Frame/Field	Frame/Field
Frame rate	25~30fps	24~60fps	Variable
Bit rate	Up to 1.5Mbps	Up to 60Mbps	16kbps~240Mbps
Picture type	I/P/B	I/P/B	I/P/B (slice)
Intra prediction	DC coef. prediction	DC coef. prediction	16x16, 4x4 intra prediction
Inter prediction	16x16	16x16	16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4
Prediction picture accuracy	1/2 cell	1/2 cell	1/4 cell
Transfer coding	8X8 DCT	8x8 DCT	4x4 DCT Integer
Loop filter	No	No	Yes

H.264/AVC VIDEO

Picture types	Prediction types	Reference relation
IDR-Picture/ Keyframe	Intra-Frame Prediction (with Long-term prediction reset)	
I-Picture/ Keyframe	Intra-Frame Prediction	
P-Picture	Inter-Frame Prediction	
B-Picture	Inter-Frame Bi-direction Prediction	

H.264/AVC VIDEO

- Prediction structure and GOP (Group of Picture): Prediction structure is based on a GOP unit.



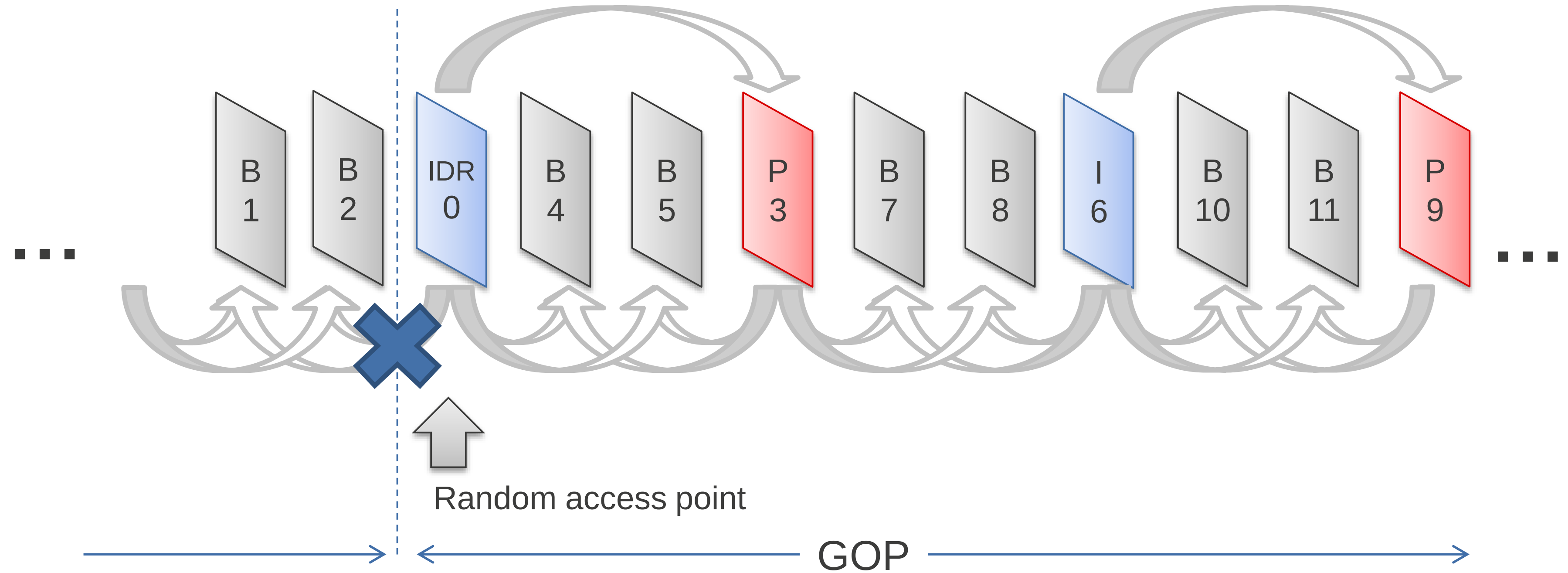
Adaptive selection of I/P/B prediction for each Macro Block

IDR (Instantaneous Decoding Refresh)

➔ Special I-picture that restricts reference between frames before and after it.

H.264/AVC VIDEO

- Prediction structure and GOP (Group of Picture): Prediction structure is based on a GOP unit.



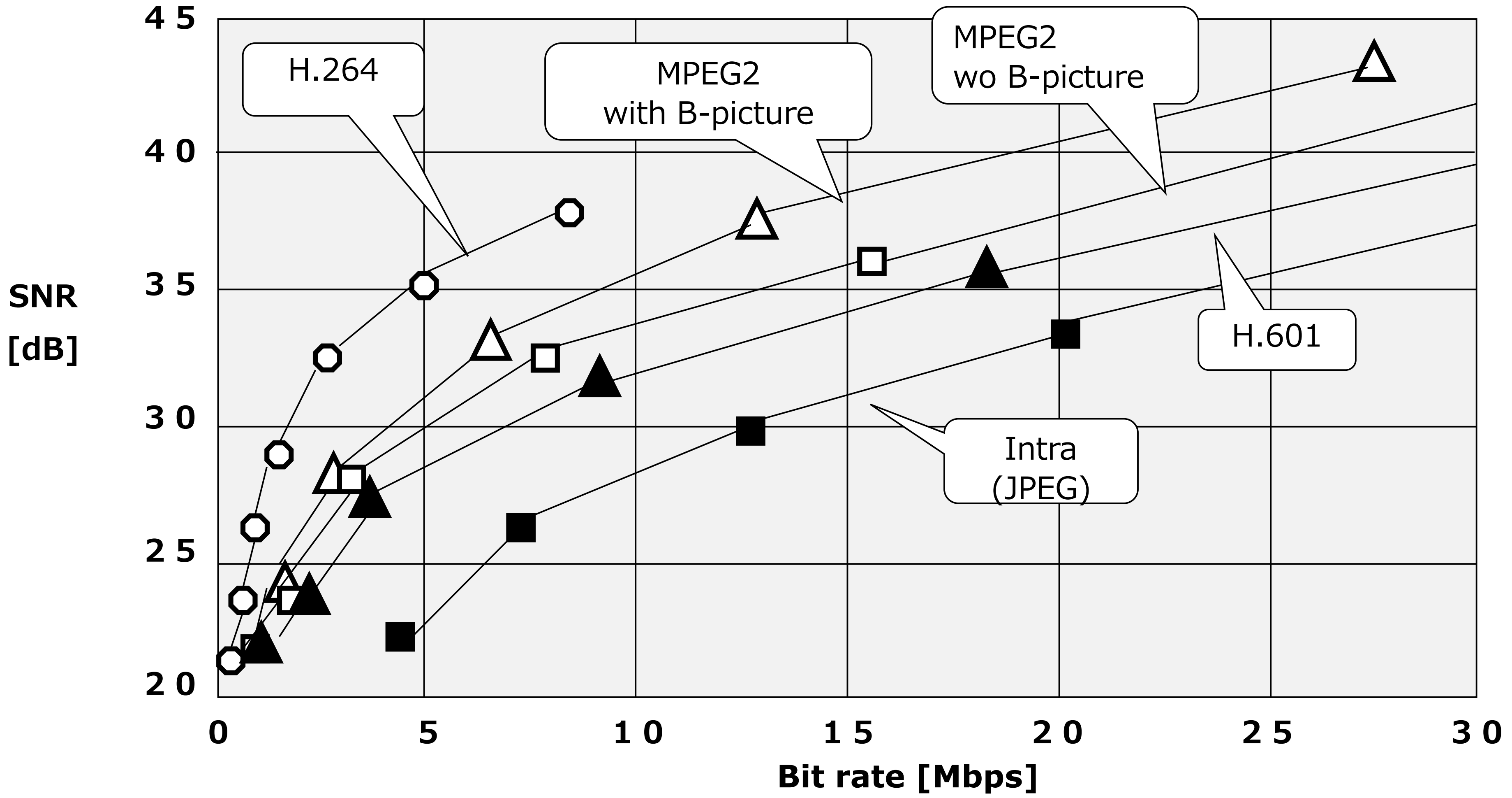
Adaptive selection of I/P/B prediction for each Macro Block

IDR (Instantaneous Decoding Refresh)

➔ Special I-picture that restricts reference between frames before and after it.

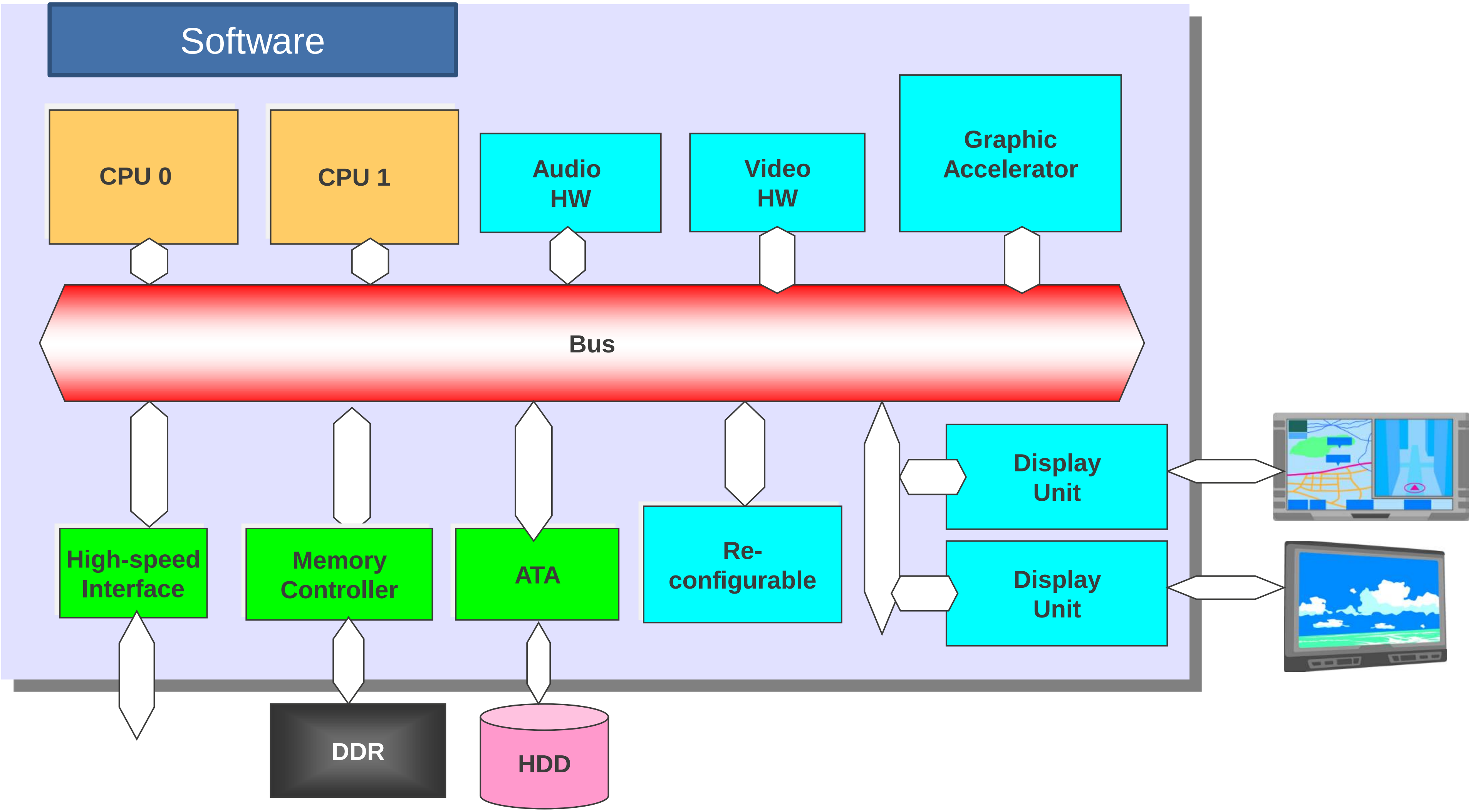
H.264/AVC VIDEO

- Performance of various coding methods are compared in figure below.



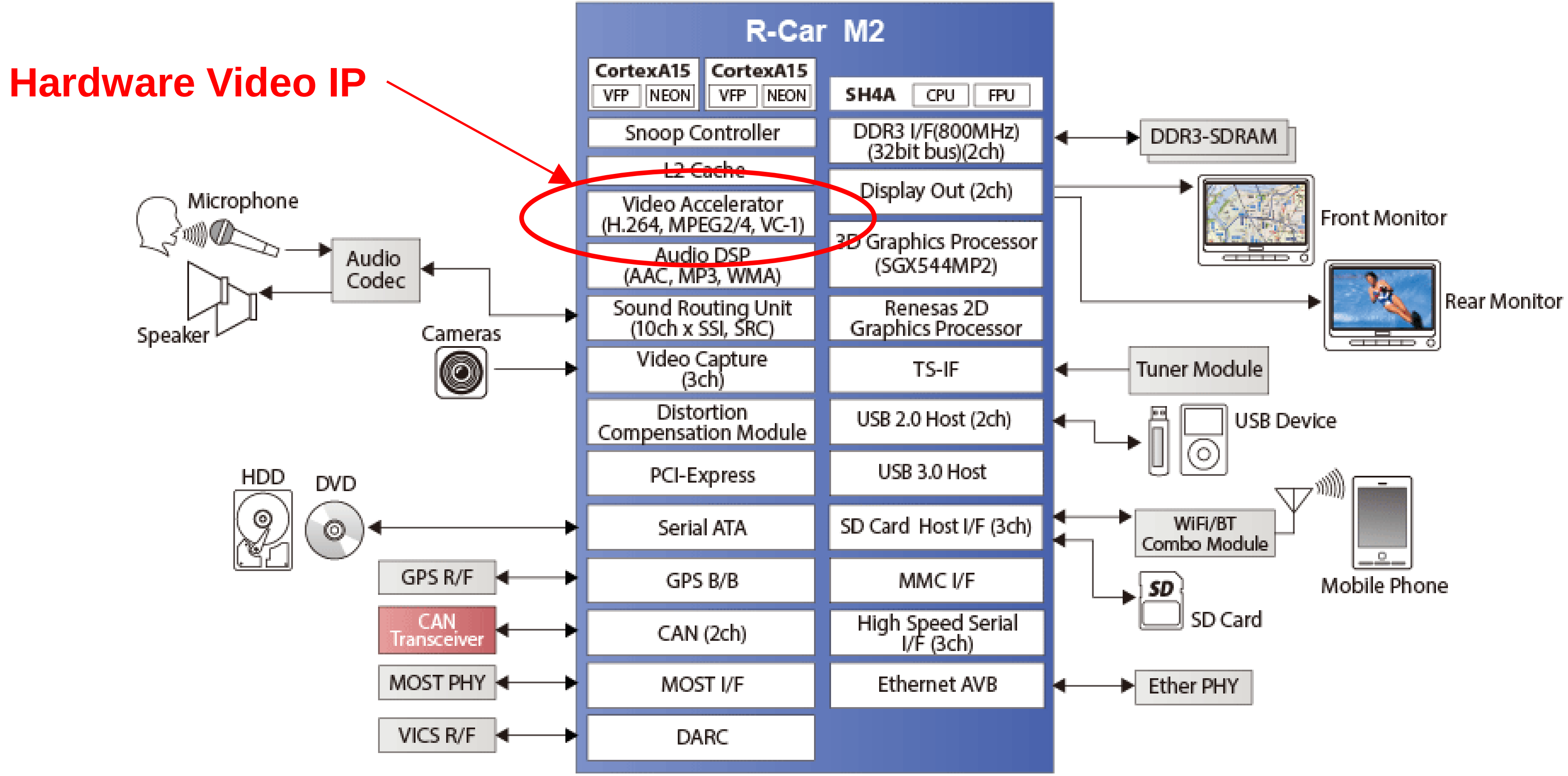
OVERVIEW: MULTIMEDIA SOC AND SOFTWARE LAYERS

SYSTEM-ON-CHIP ARCHITECTURE

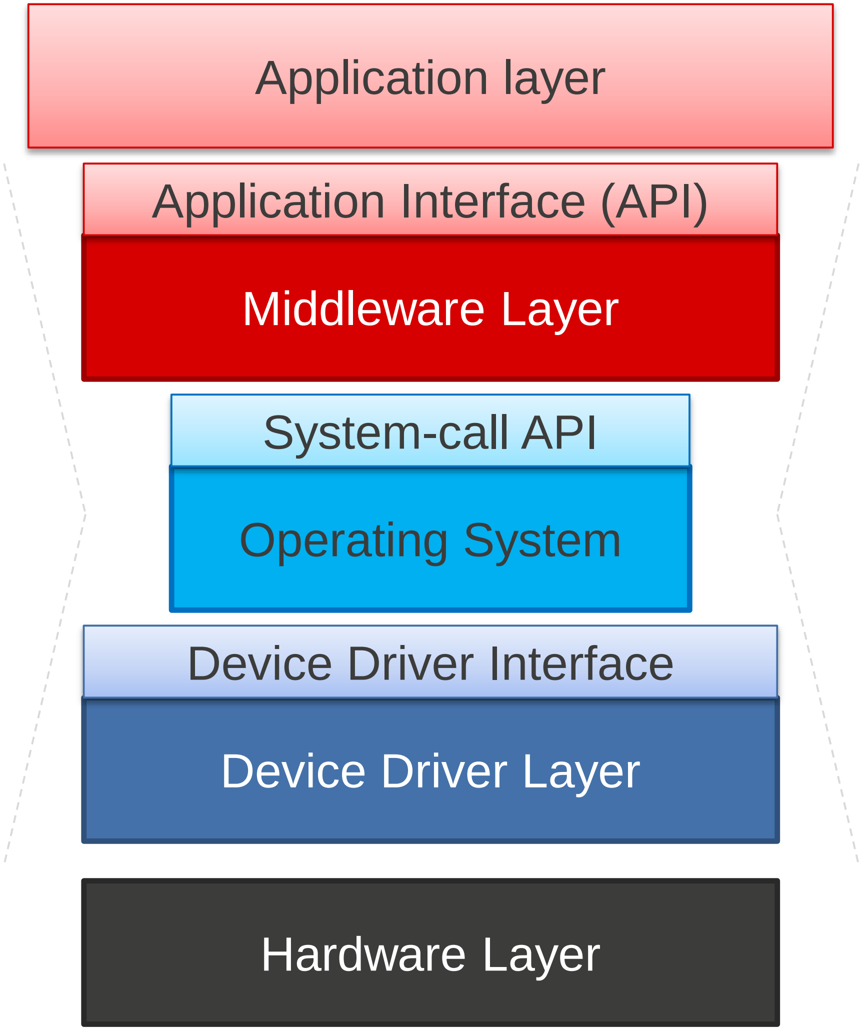


SYSTEM-ON-CHIP ARCHITECTURE

- Renesas R-Car M2



SYSTEM-ON-CHIP ARCHITECTURE



Navigation, video/audio playback, web browser, parent control, etc.

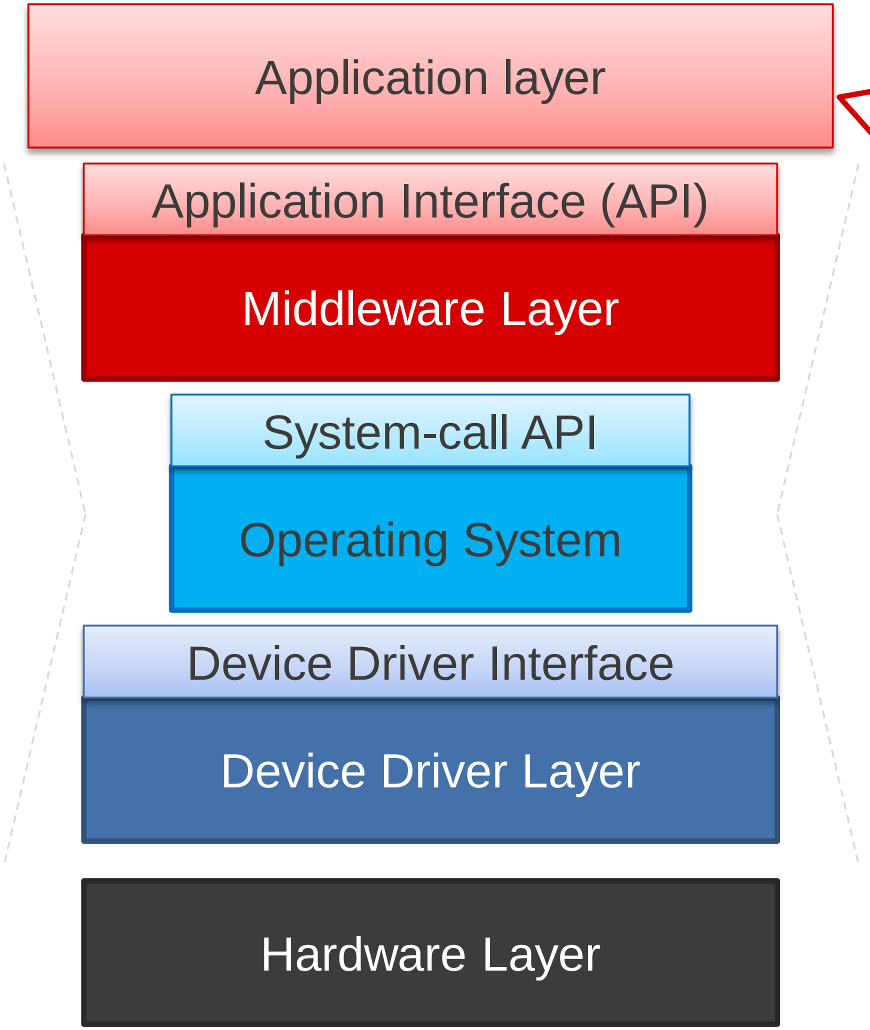
OpenMAX, Gstreamer, Stagefright, Digital Television Middleware, etc.

Windows, Linux, Android, QNX, INTEGRITY, etc.

Power management, thermal control, video/audio coding control, etc.

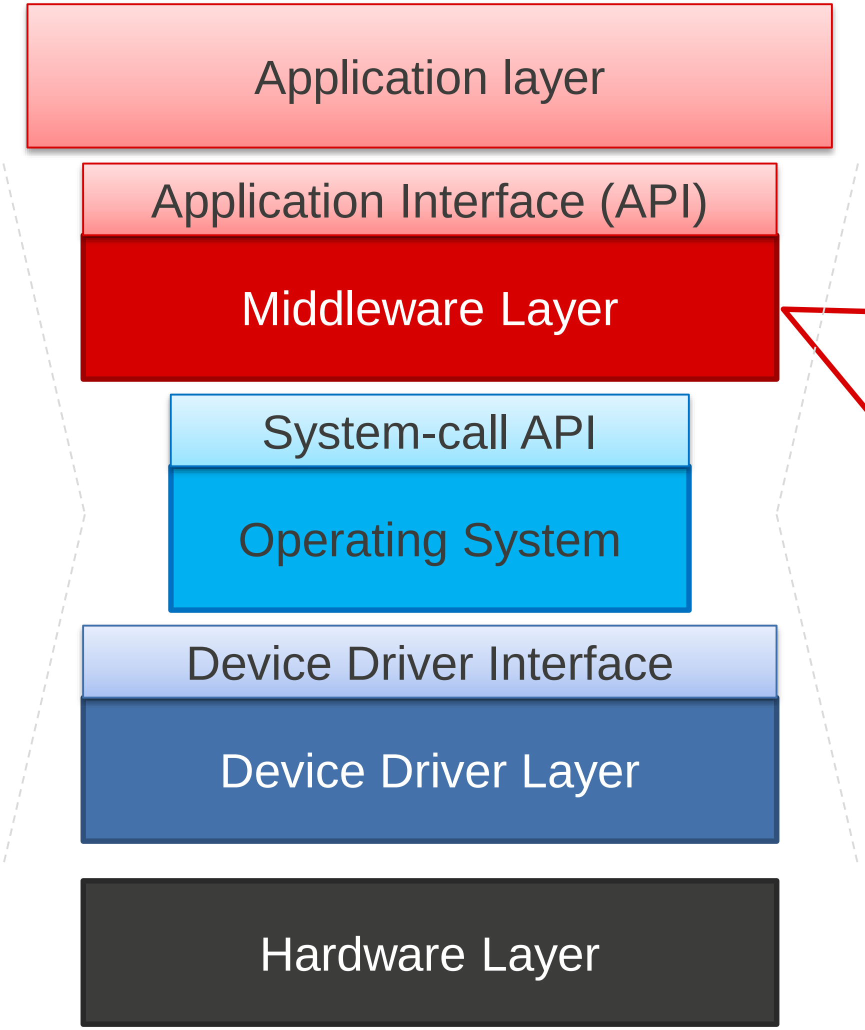
Video/Audio decoder IP, camera sensor, thermal sensor, USB port, etc.

SYSTEM-ON-CHIP ARCHITECTURE



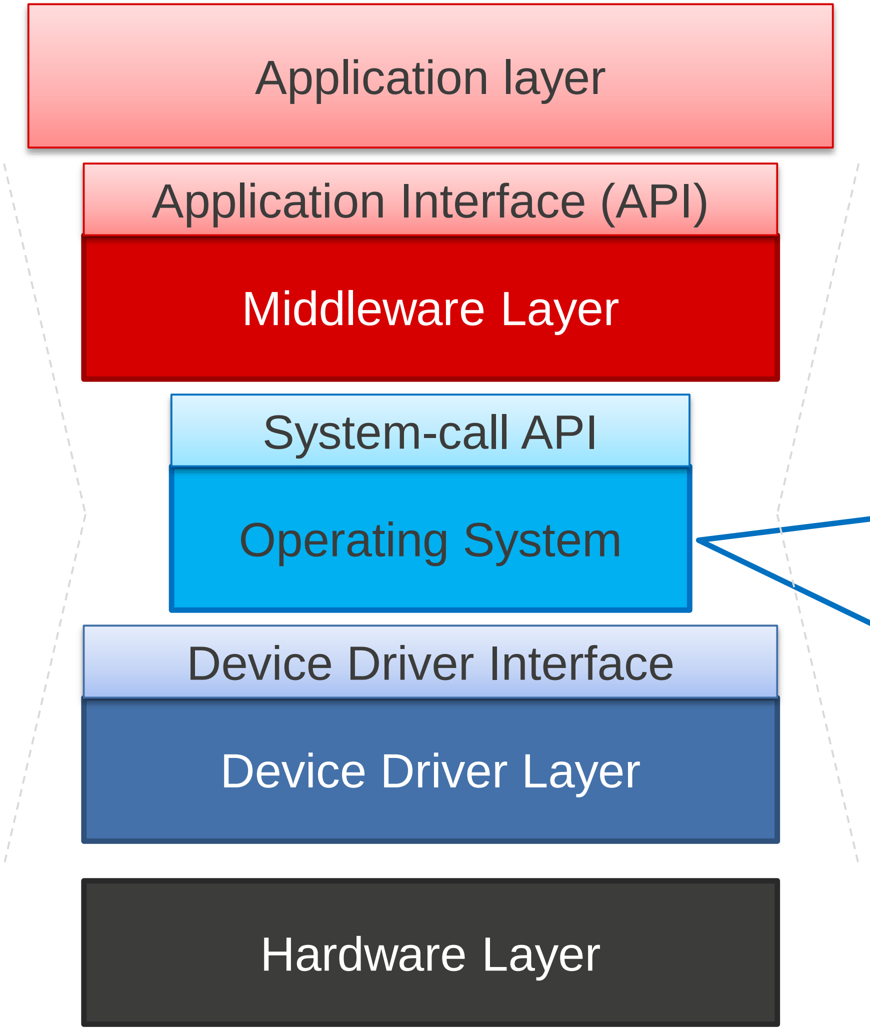
- ❑ Product-specific software.
- ❑ Final specification is defined here.
- ❑ Customized for the actual product.

SYSTEM-ON-CHIP ARCHITECTURE



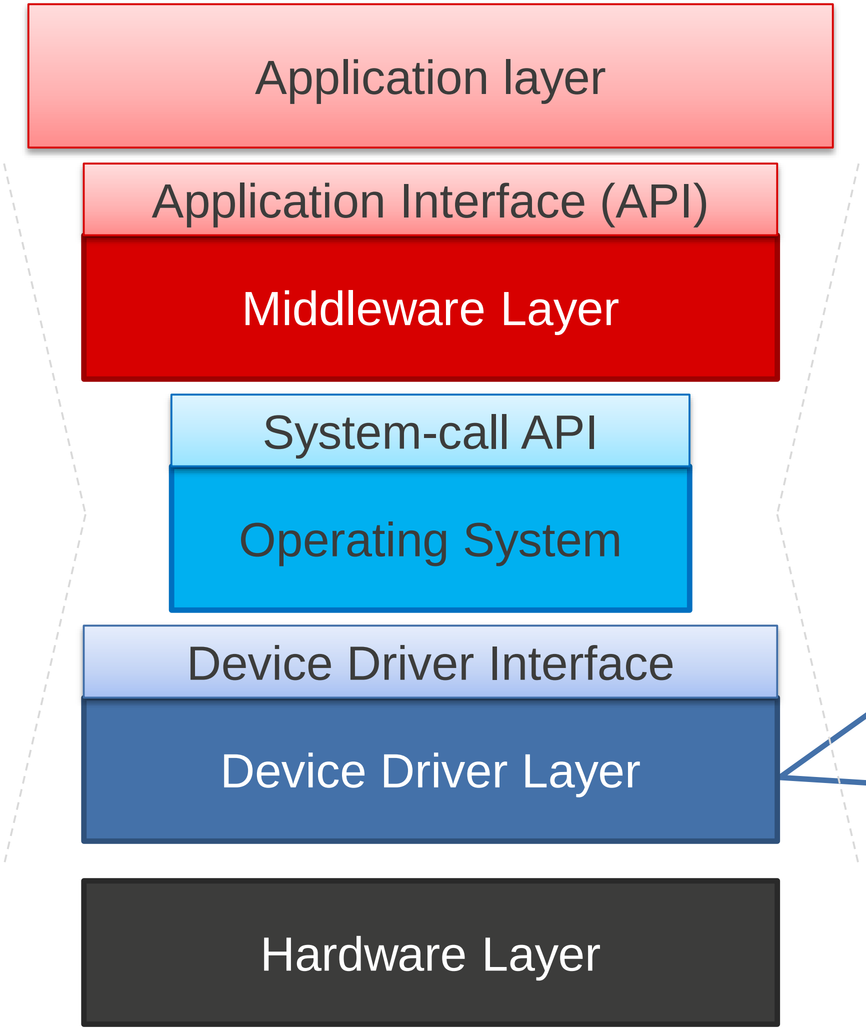
- ❑ Function-specific software.
- ❑ Common functions are realized here.
- ❑ Some functions can be replaced by hardware implementation.
- ❑ Not customized for the actual product because its mission is to generalize processing of various lower layer and hardware layer for the portability and maintainability of upper layer.

SYSTEM-ON-CHIP ARCHITECTURE



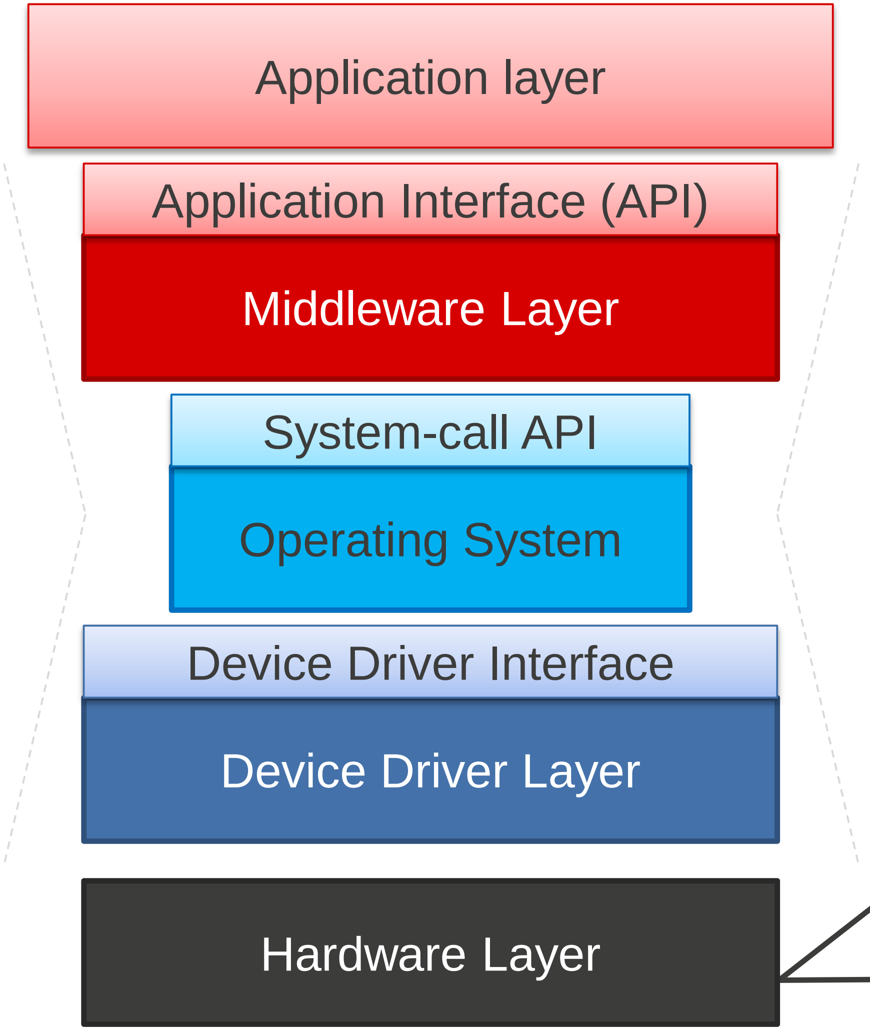
- ❑ Operating system offers general tools and software circumstance.
- ❑ However, it almost does nothing in the business logic of product that is realized both by other software layers and by hardware layer.
- ❑ It only provides service and helper for other layers to implement the business logic of product.

SYSTEM-ON-CHIP ARCHITECTURE



- ❑ Hardware-specific software.
- ❑ Hardware-dependent specification is assimilated here.
- ❑ It controls the logic of hardware, but need helper from OS layer to communicate with hardware layer.

SYSTEM-ON-CHIP ARCHITECTURE

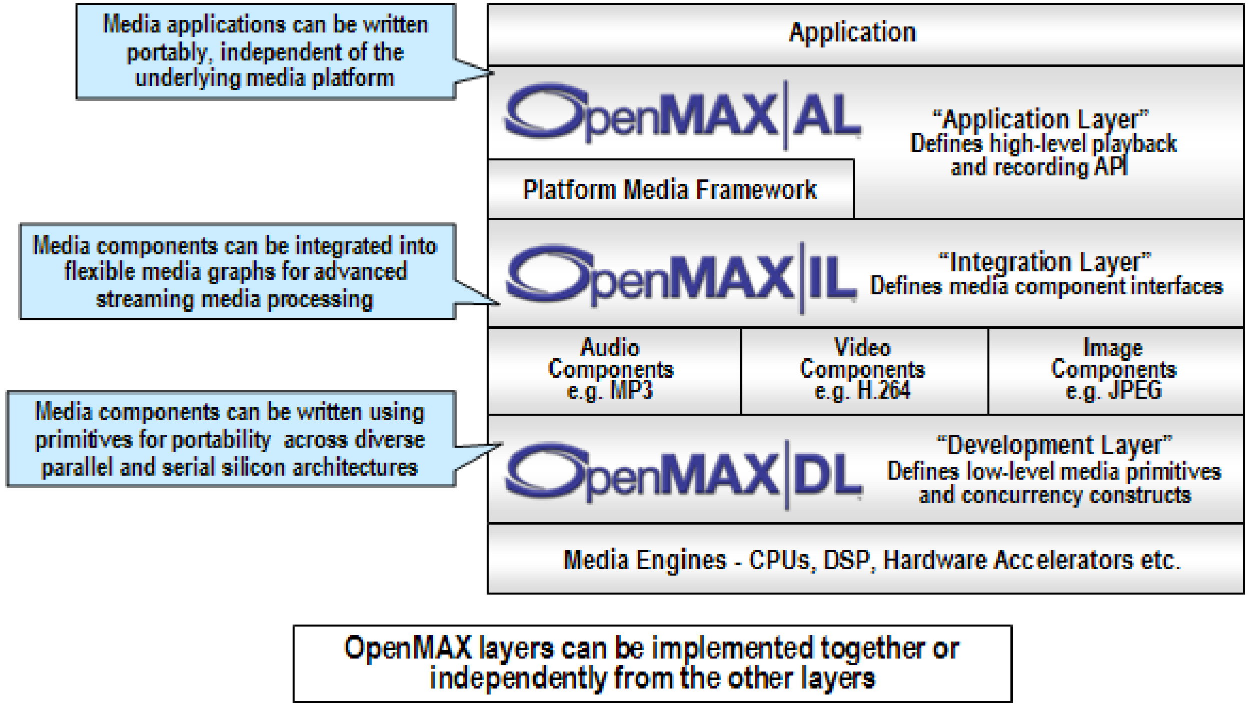


- ❑ Product-specific hardware.
- ❑ With the same logic implementation, software is for optimal cost and hardware is for optimal performance.

OPENMAX

- OpenMAX is a royalty-free, cross-platform set of C-language programming interfaces that provides abstractions for routines especially useful for audio, video, and still images.

- OpenMAX IL serves as a low-level interface for audio, video, and imaging codecs used in embedded devices.



- **It gives applications and media frameworks the ability to interface with multimedia codecs and supporting components in a unified manner.**


One Renesas