画像情報特論 (3)

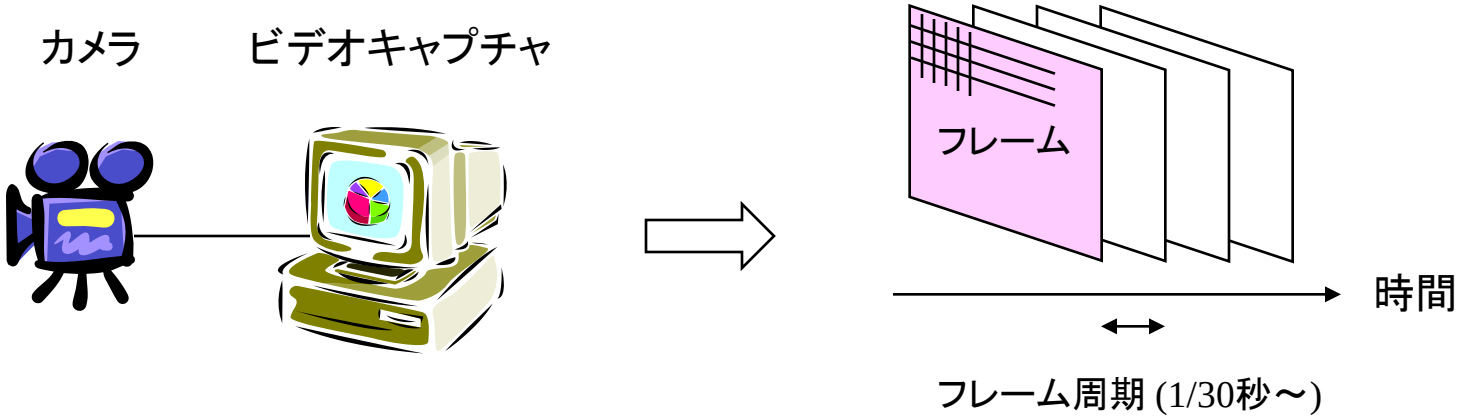
- ビデオ圧縮: H.264/AVC

情報理工学専攻　甲藤二郎
E-Mail: katto@waseda.jp

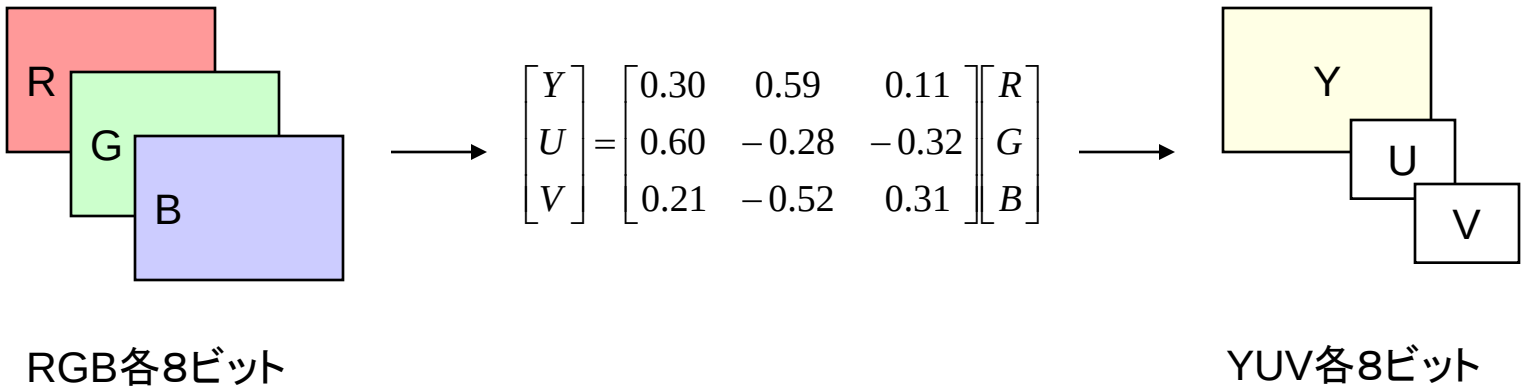
ビデオ圧縮の基礎

ディジタルビデオ (1)

- 時間方向・空間方向のサンプリング

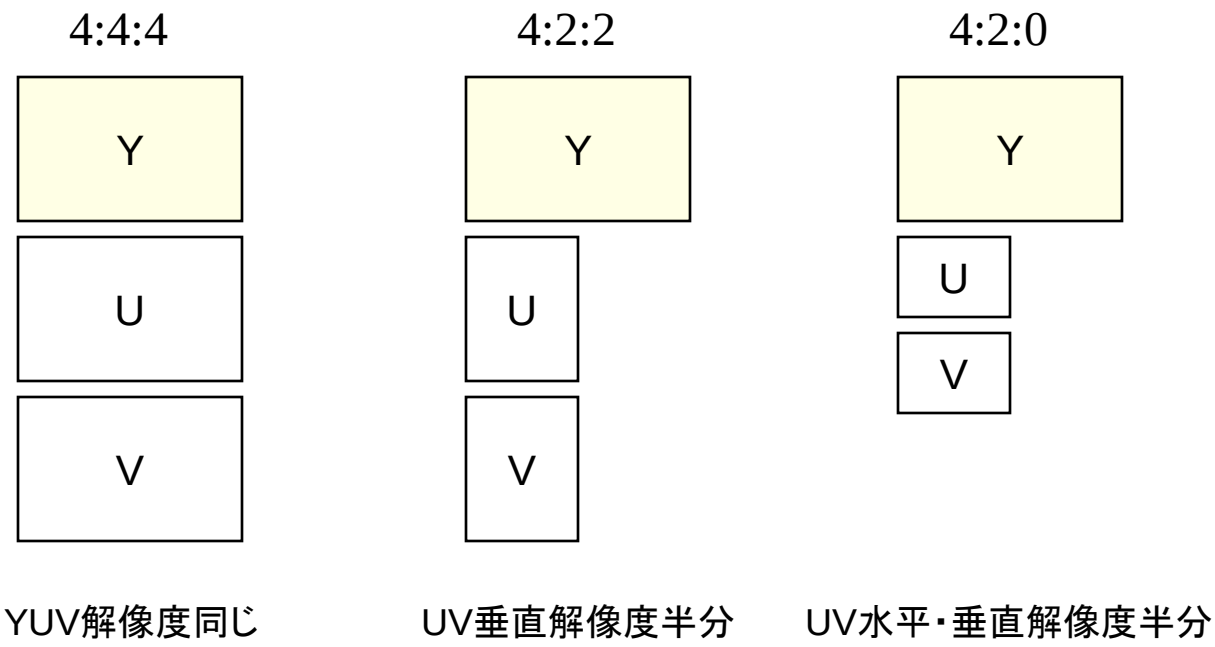


- RGB / YUV 変換



ディジタルビデオ (2)

- CCIR 601 フォーマット

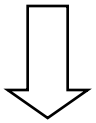


- 通常のビデオ圧縮: 4:2:0 フォーマット
- 高画質ビデオ圧縮: 4:2:2 フォーマット

ディジタルビデオ (3)

- 莫大な情報量 (RGB各8ビット無圧縮の場合)

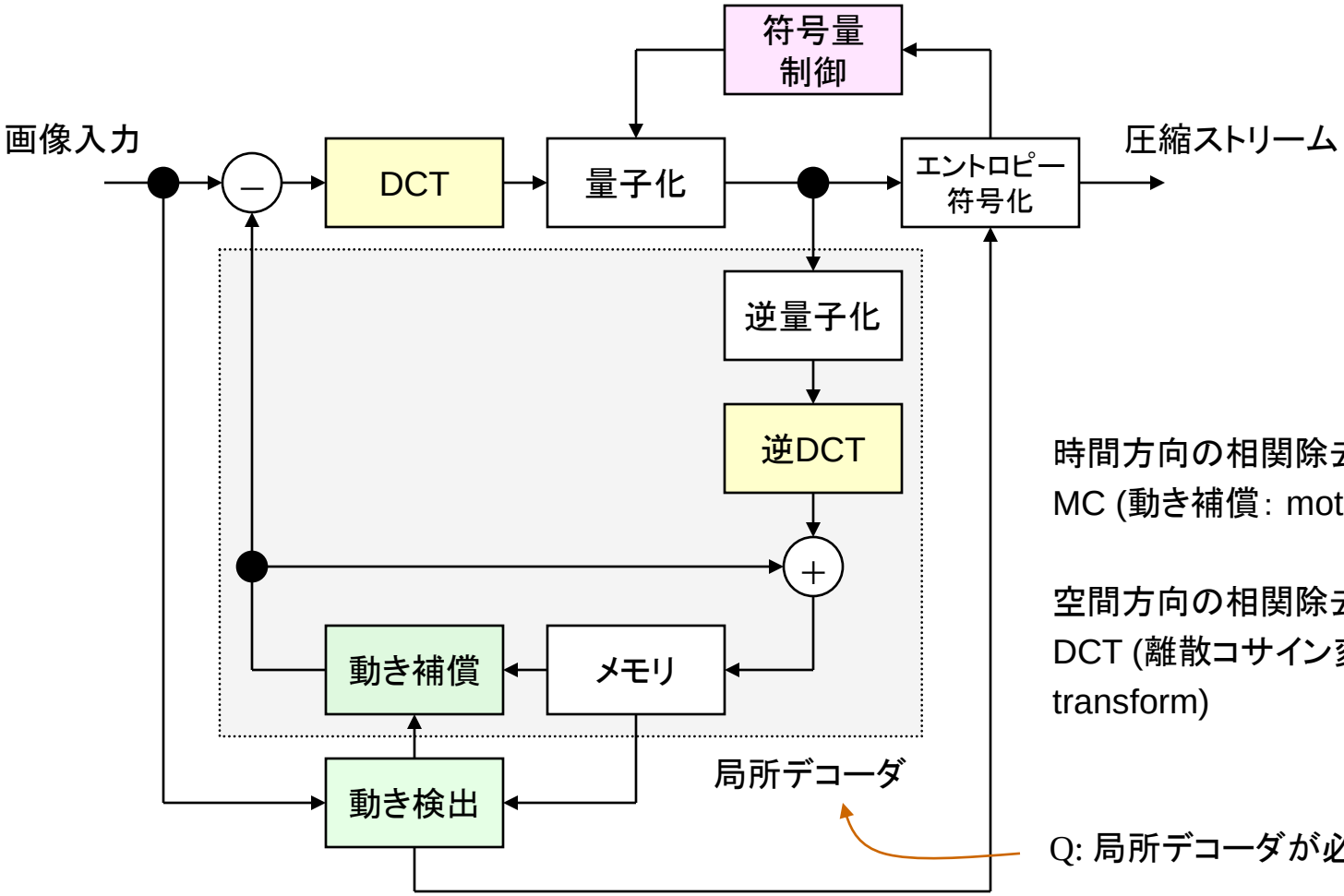
用途	解像度	データ量
TV会議	352x240	21Mbit/s
TV	720x480	83Mbit/s
HDTV	1920x1080	498Mbit/s



データ圧縮の必要性

ビデオ圧縮の仕組み

- MC+DCT ハイブリッド予測符号化 (30年間変わらない方式)



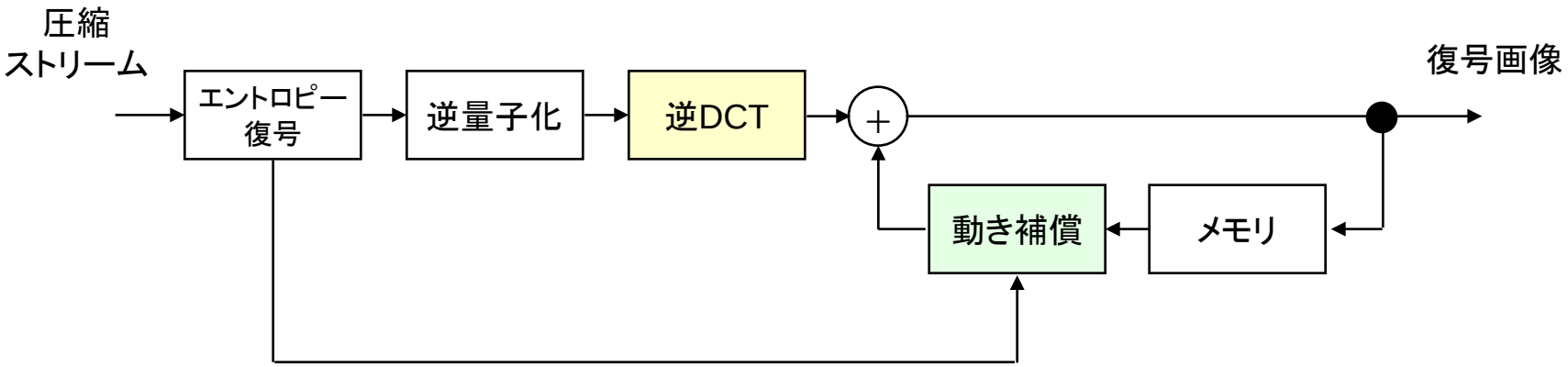
時間方向の相関除去:
MC (動き補償: motion compensation)

空間方向の相関除去:
DCT (離散コサイン変換: discrete cosine transform)

Q: 局所デコーダが必要な理由を説明せよ

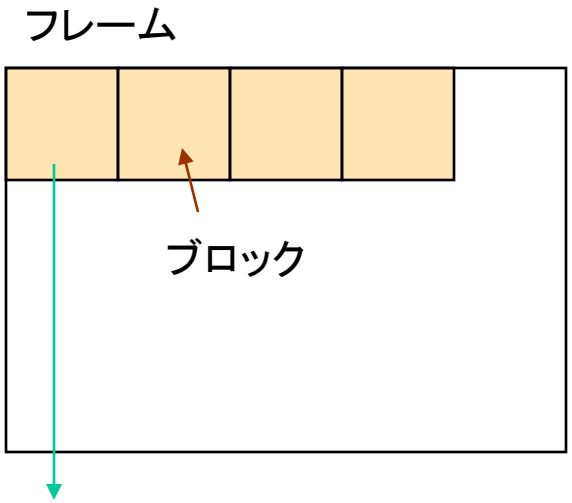
ビデオ復号の仕組み

- エンコーダのローカルデコードに同じ



フレーム内符号化

- DCT



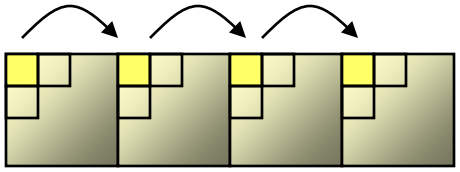
DCT (離散コサイン変換)

画像信号の性質:
隣接画素間の相関が非常に高い
(相関係数: 0.9 ～)

直交変換

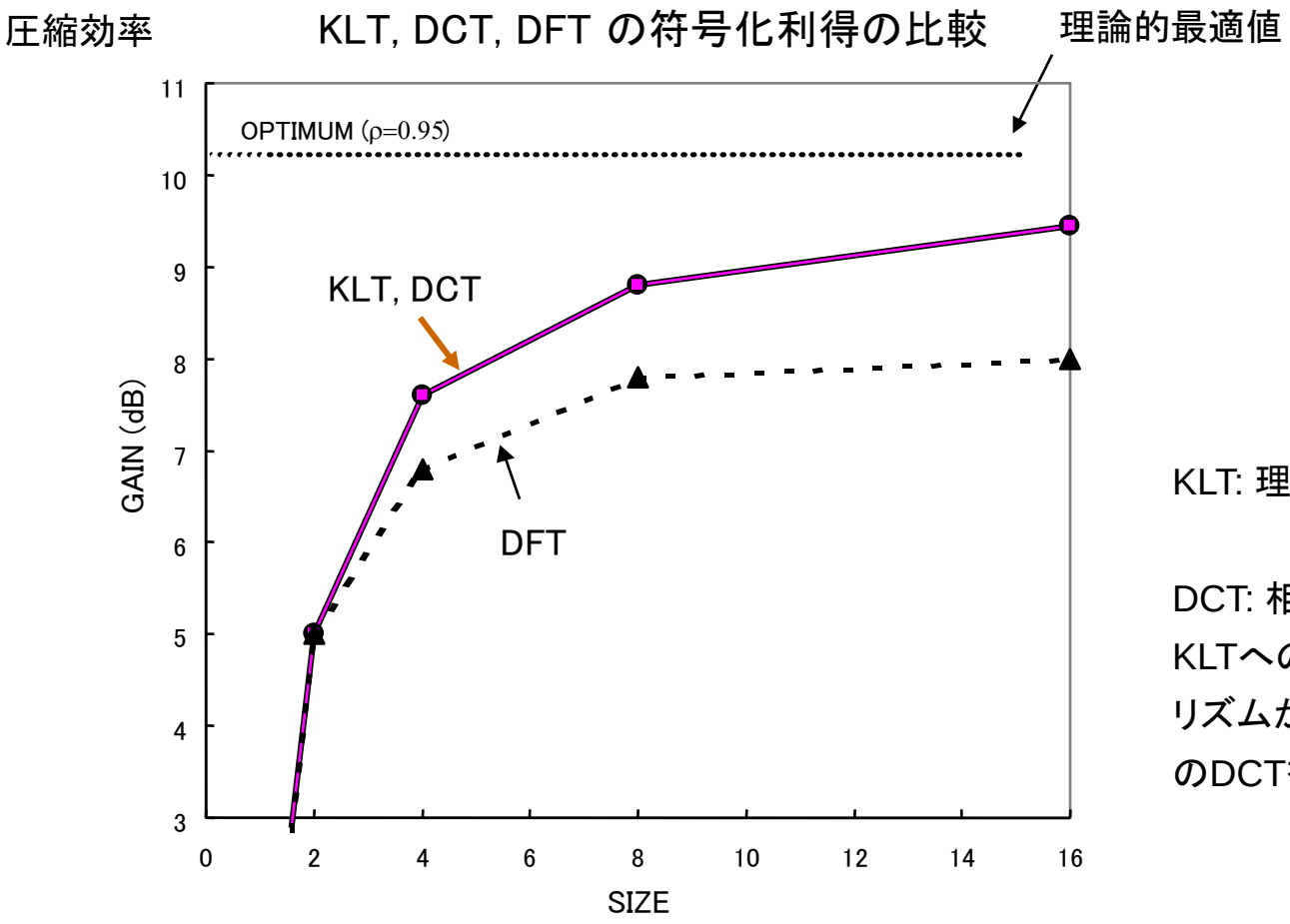
特定の変換係数にエネルギーが集中

隣接ブロック間でさらに予測 (特に直流成分)



直交変換 (1)

- DCTが使われる理由

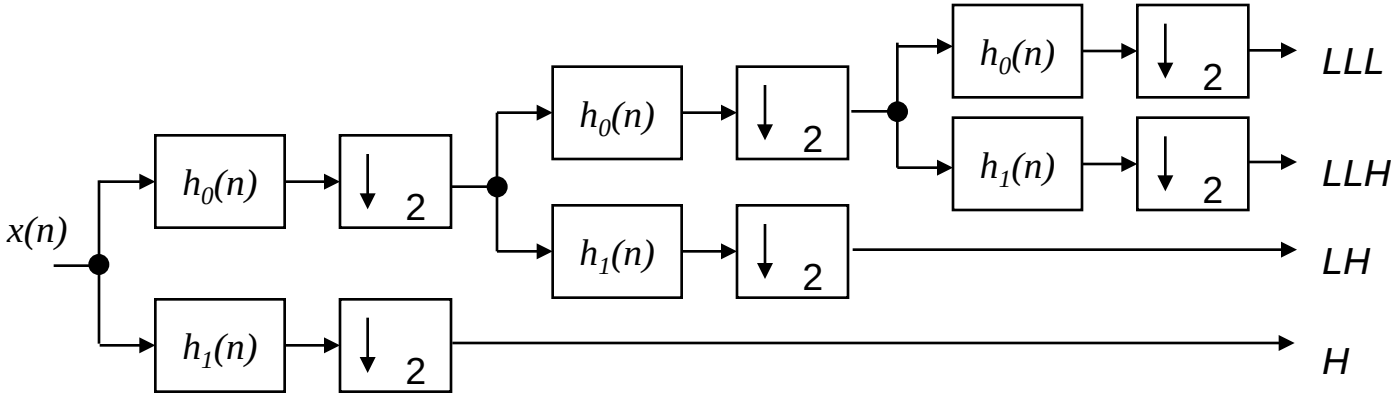


KLT: 理論的に最適な直交変換。

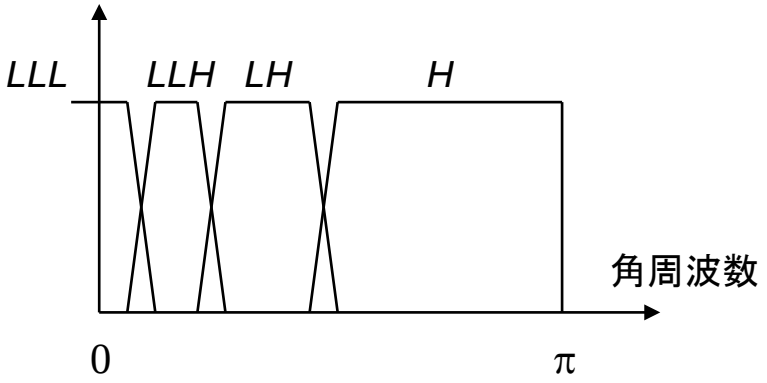
DCT: 相関の高い入力に対するKLTへの漸近性、及び高速アルゴリズムが存在。通常は 8x8 サイズのDCTを使用。

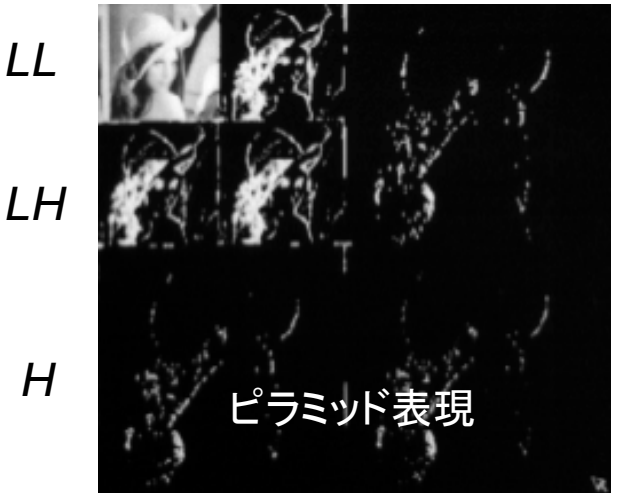
直交変換 (2)

- Wavelet 変換 (対抗)



2分割フィルタバンクのツリー接続



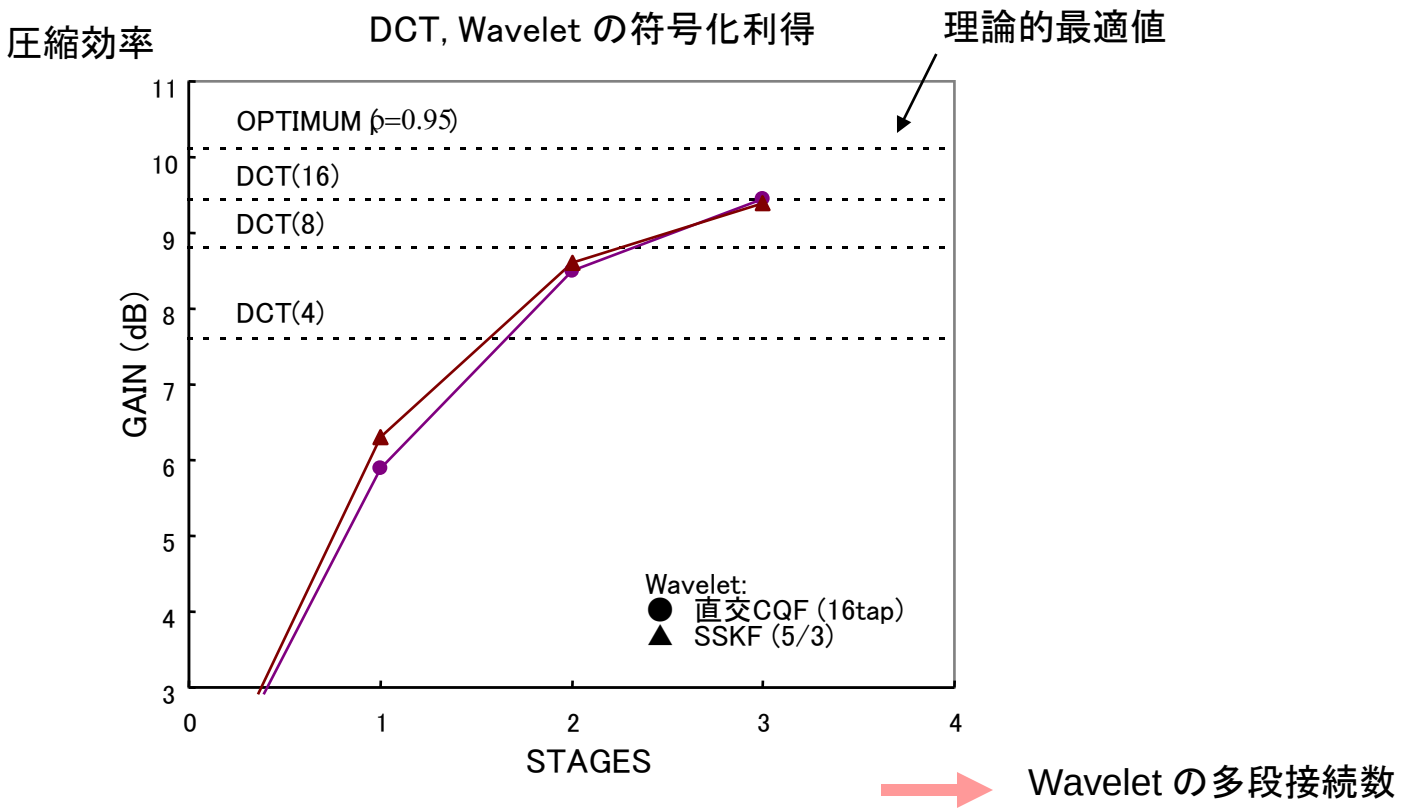


長所: ブロックひずみが少ない

短所: ブロック動き補償と相性が悪い

直交変換 (3)

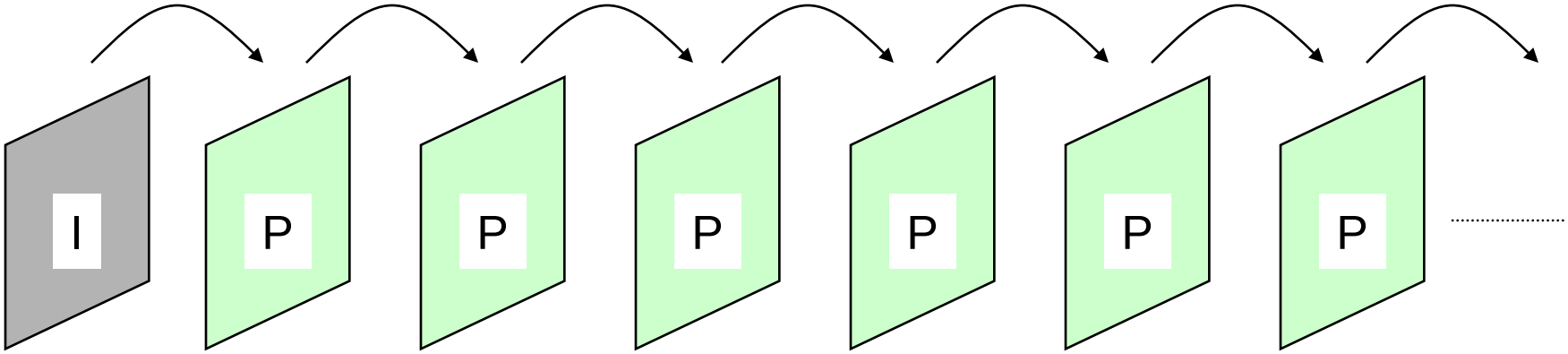
- DCT と Wavelet の比較



- DCT: 動画 (ビデオ) 圧縮
- Wavelet: 静止画圧縮 (JPEG-2000)

フレーム間符号化 (1)

- IP 予測



ビデオ信号の性質:

隣接フレーム間の相関が非常に高い

(相関係数: 0.9 〜)

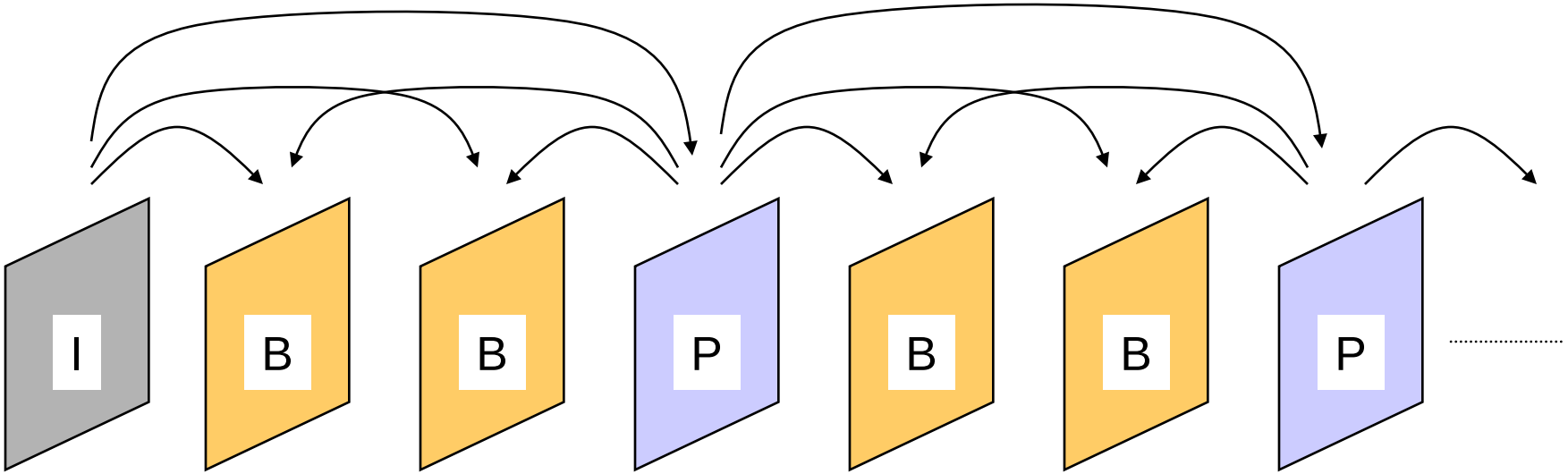
フレーム間の予測誤差がほとんどゼロ

- I: I ピクチャ (フレーム内符号化)
- P: P ピクチャ (フレーム間符号化)

さらに動き検出・動き補償予測

フレーム間符号化 (2)

- IPB 予測



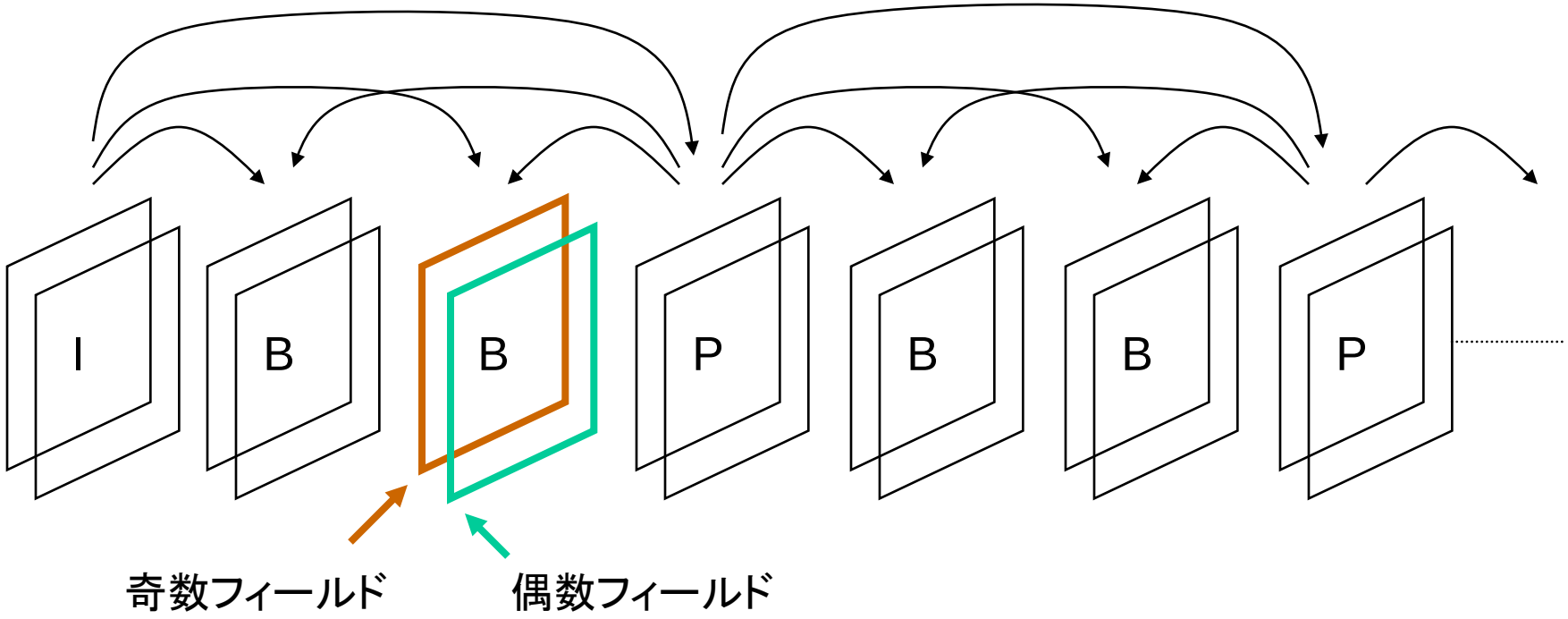
片方向で予測を行うより、両方向で
予測を行うほうが予測効率が高い
(ただし、フレーム間の距離に依存)

- I: I ピクチャ (フレーム内符号化)
- P: P ピクチャ (片方向予測)
- B: B ピクチャ (両方向予測)

→ 予測効率の改善

フレーム間符号化 (3)

- フレーム・フィールド適応予測



ディジタルTV放送に対応 (MPEG-2)

- 動き補償: フィールド予測、フレーム予測、デュアルプライム予測
- DCT: フレームDCT、フィールドDCT

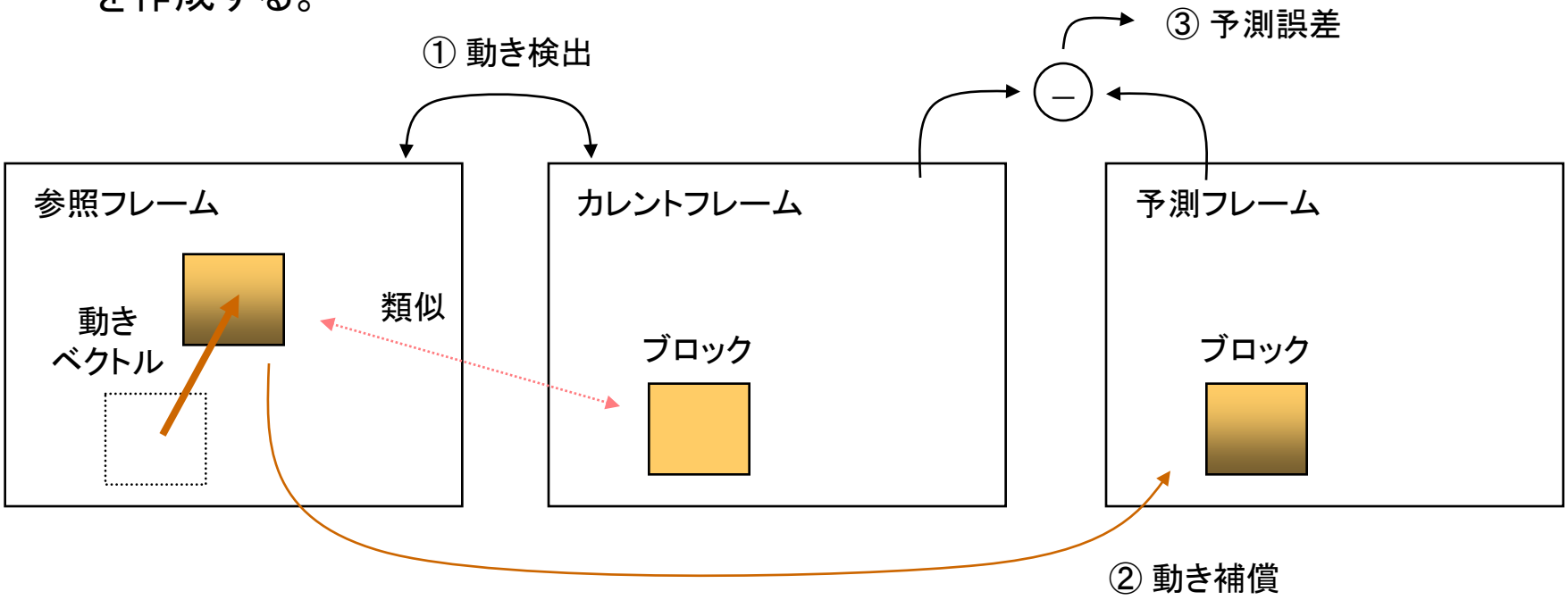
動き検出と動き補償 (1)

・動き検出 (ブロックマッチング):

過去の画像 (参照フレーム) から、現在の画像 (カレントフレーム) に最も類似しているブロックを探索し、動きベクトルを求める。

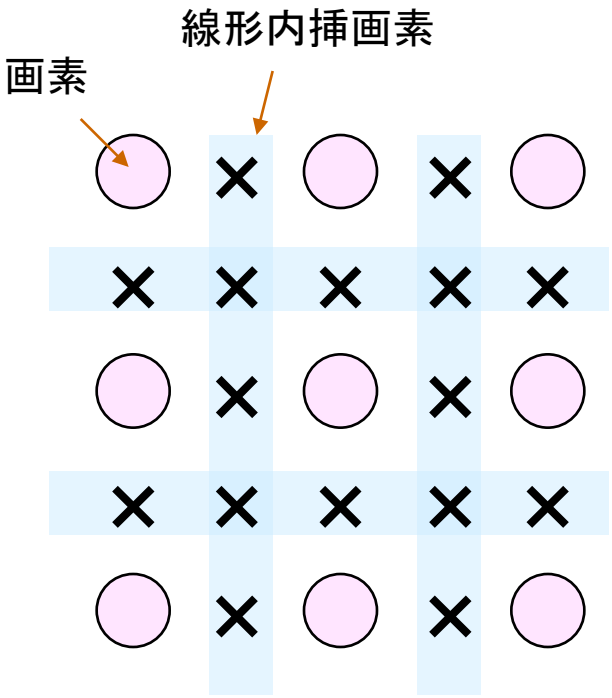
・動き補償:

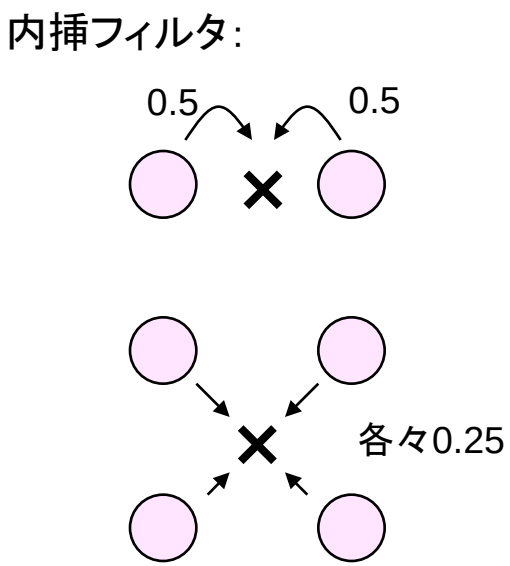
動き検出で求めた動きベクトルから、カレントフレームの予測画像 (予測フレーム) を作成する。



動き検出と動き補償 (2)

- 半画素精度動き補償:
  線形内挿を行い、0.5 画素精度の動きベクトルを算出し、予測画像を作成。

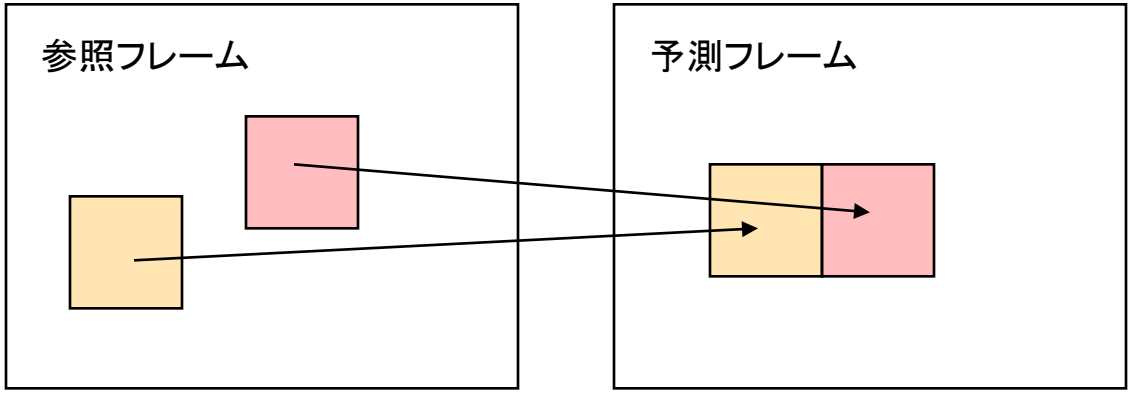




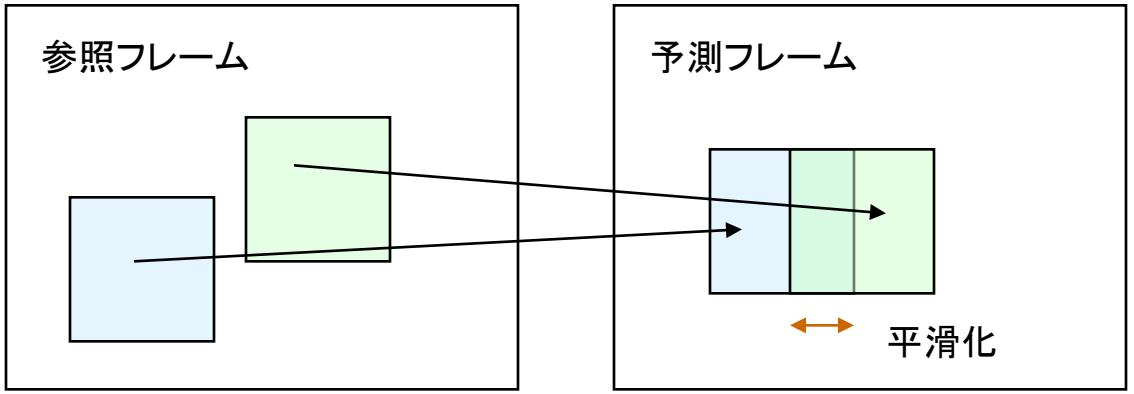
動き検出と動き補償 (3)

- オーバーラップ動き補償:
  隣接ブロックの動きベクトルも利用し、ブロックの平滑化加算によって予測画像を作成。

通常のブロックマッチング



オーバーラップ動き補償



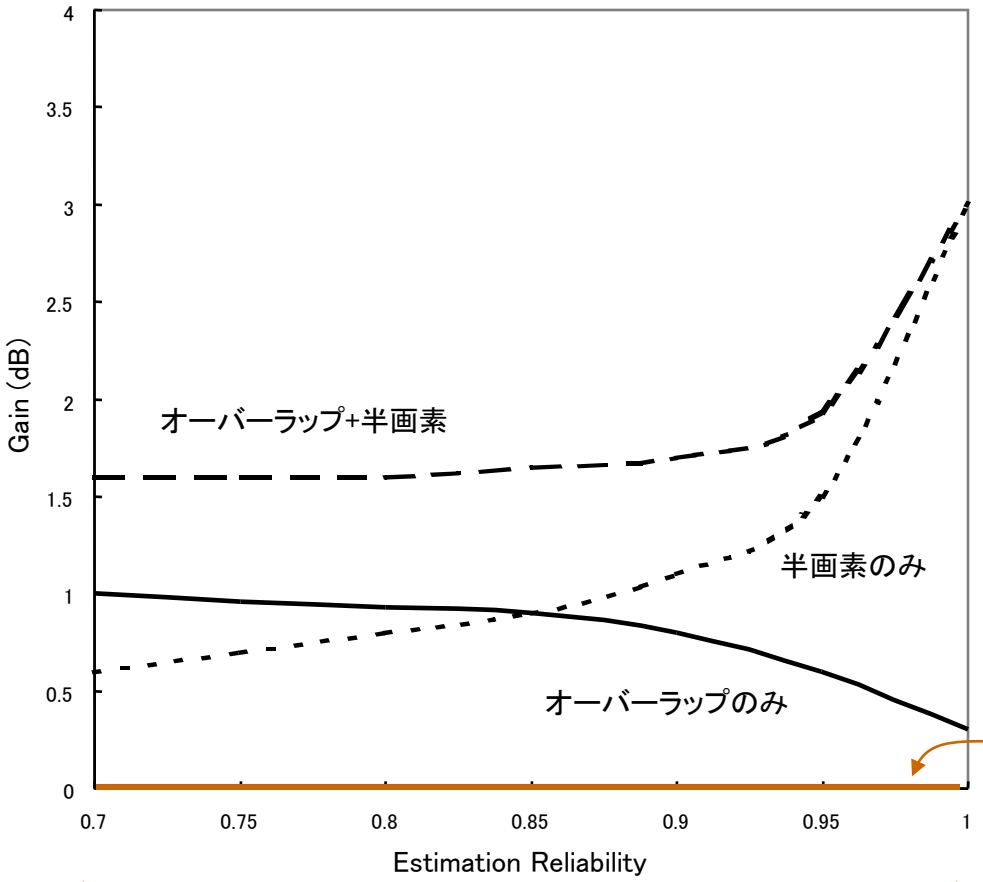
平滑化：台形ウィンドウ、
コサインウィンドウなど。

動き検出と動き補償 (4)

- 特性比較

整数画素精度・ブロック動き補償に対する
半画素精度・オーバーラップ動き補償の予測利得

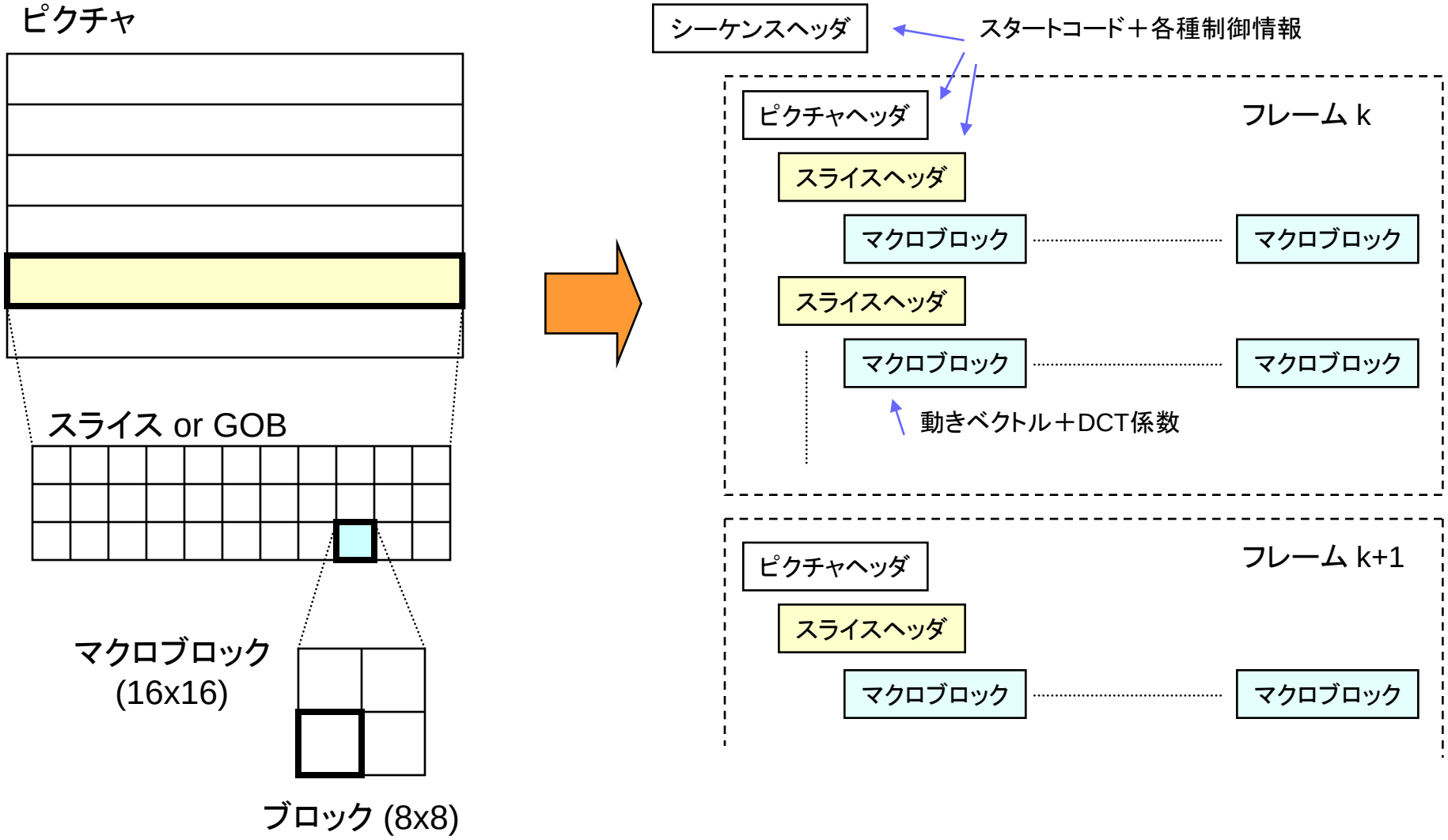
予測誤差の
低減効果



動き予測しにくい
画像の場合

動き予測しやすい
画像の場合

ビットストリーム構造


ピクチャ
スライス or GOB
マクロブロック
(16x16)
ブロック (8x8)
シーケンスヘッダ
スタートコード＋各種制御情報
ピクチャヘッダ
フレーム k
スライスヘッダ
マクロブロック
マクロブロック
スライスヘッダ
マクロブロック
マクロブロック
動きベクトル＋DCT係数
ピクチャヘッダ
フレーム k+1
スライスヘッダ
マクロブロック
マクロブロック

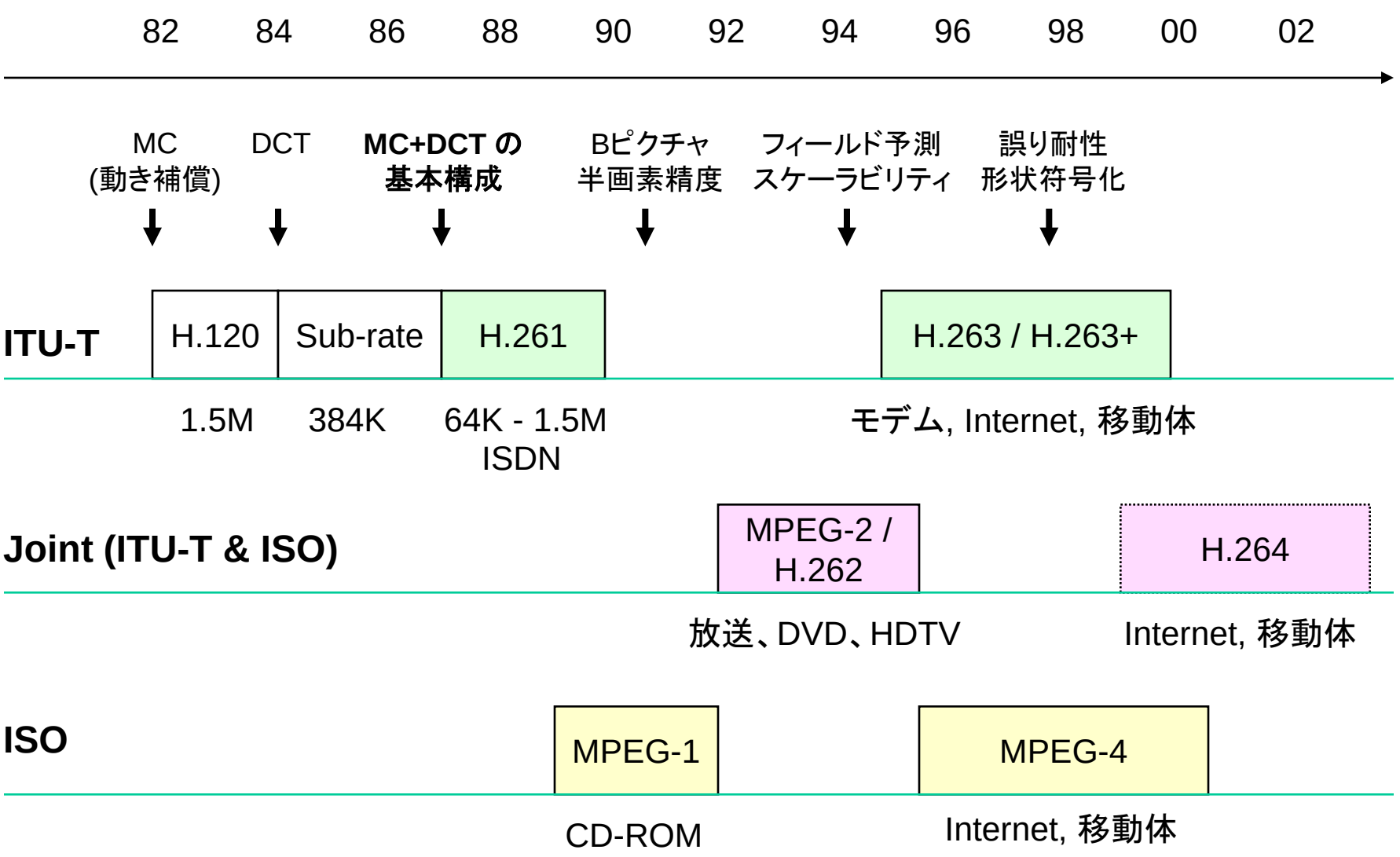
国際標準方式

国際標準方式 (1)

団体	名称	時期	符号化レート	当初の用途
ITU-T	H.261	1990年	64kb/s～2Mb/s	ISDN用テレビ電話
	H.263	1996年	数十kb/s～	アナログ回線用テレビ電話
	H.263+	1998年	数十kb/s～	インターネット、移動体
	H.264	2003年	数十kb/s～	インターネット、移動体
ISO	MPEG-1	1992年	～1.5Mb/s	CD-ROM
	MPEG-2*	1995年	数Mb/s～数十Mb/s	ディジタル放送
	MPEG-4	1999年	数十kb/s～	インターネット、移動体

* MPEG-2/H.262、H.264/AVC (MPEG-4 Part 10) はISOとITU-Tのジョイント規格

国際標準方式 (2)


82
84
86
88
90
92
94
96
98
00
02
MC
(動き補償)
DCT
MC+DCT の
基本構成
Bピクチャ
半画素精度
フィールド予測
スケーラビリティ
誤り耐性
形状符号化
ITU-T
H.120
Sub-rate
H.261
H.263 / H.263+
1.5M
384K
64K - 1.5M
ISDN
モデム, Internet, 移動体
Joint (ITU-T & ISO)
MPEG-2 /
H.262
H.264
放送、DVD、HDTV
Internet, 移動体
ISO
MPEG-1
MPEG-4
CD-ROM
Internet, 移動体

国際標準方式 (3)

- 代表的な機能の比較

名称	MC+DCT	1/2画素	IPB予測	フィールド	形状符号化	再同期	スケーラビリティ
H.261	○	-	-	-	-	-	-
H.263	○	○	△	-	-	-	-
MPEG-1	○	○	○	-	-	○	-
MPEG-2	○	○	○	○	-	○	○
H.263+	○	○	△	-	△	○	○
MPEG-4	○	○	○	○	○	○	FGS
H.264	○	○	○	-	△	○	SVC

インターネット放送で有効
+ 符号量制御 (後述)

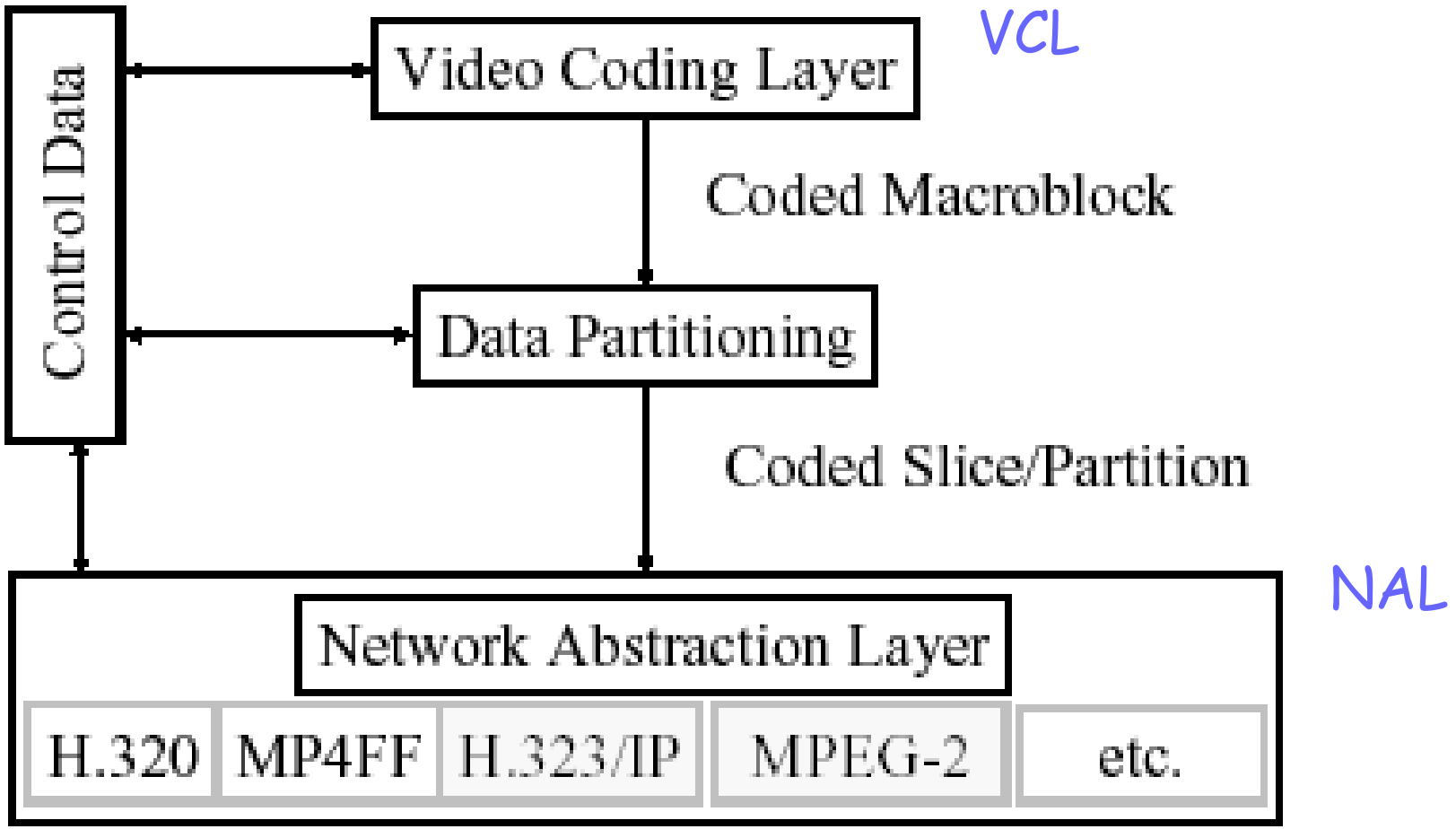
ITU-T H.264/AVC (MPEG-4 Part 10)

H.264/AVC の応用

- **Broadcast** over cable, satellite, cable modem, DSL, terrestrial
- Interactive or serial storage on optical and magnetic devices (e.g. **DVD**)
- **Conversational** services over IDSN, Ethernet, LAN, DSL, wireless and mobile networks
- Video-on-demand (**VoD**) or multimedia streaming services over ISDN, cable modem, DSL, LAN, wireless and mobile networks
- Multimedia messaging services (**MMS**) over ISDN, DSL, Ethernet, LAN, wireless and mobile networks

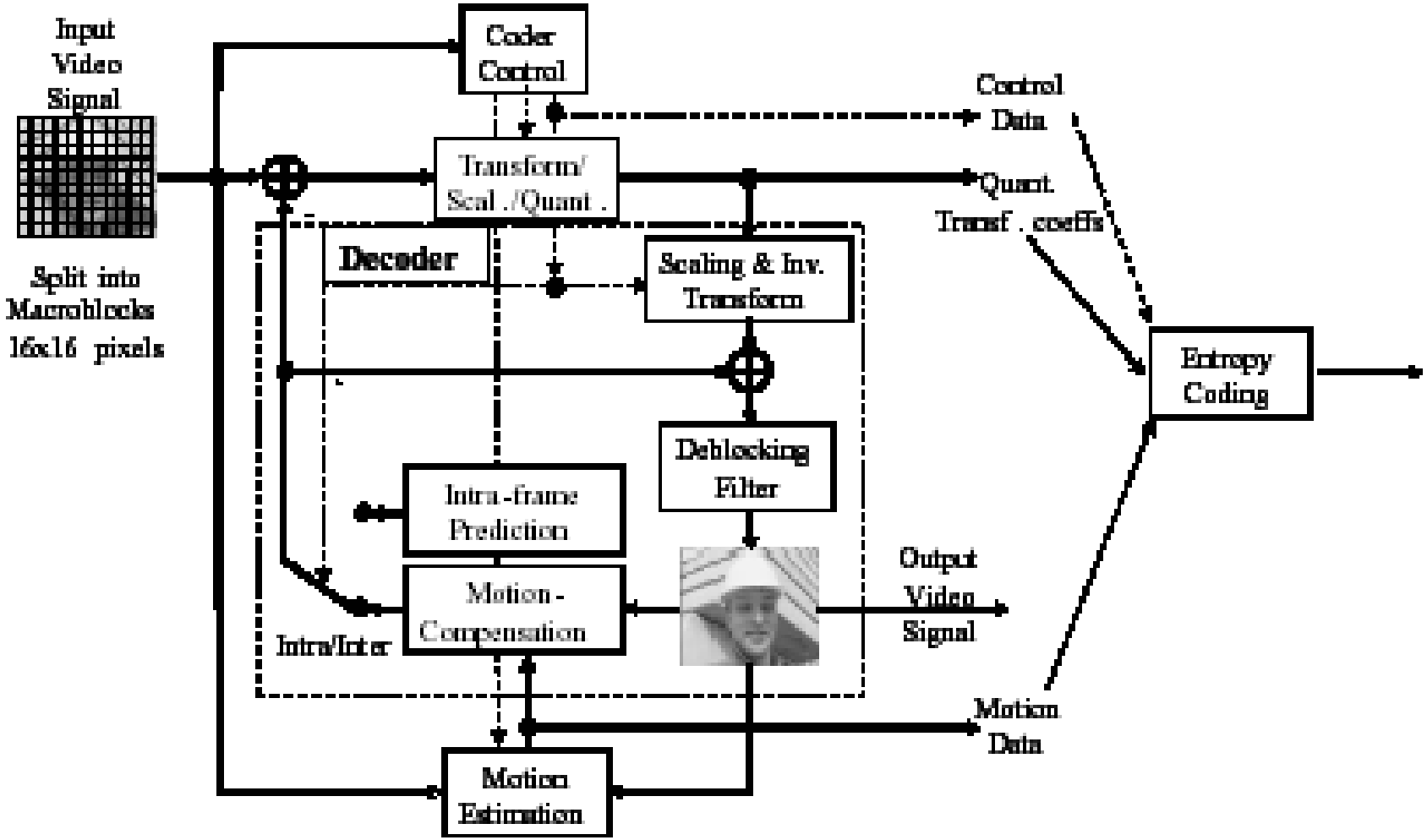
T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

H.264/AVC の基本構成 (1)



T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

H.264/AVC の基本構成 (2)



T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

Improved Prediction

- Variable block-size motion compensation (4x4 to 16x16)
- Quarter-sample-accuracy motion compensation (1/4pel)
- Picture boundary extrapolation for motion compensation
- Multiple reference picture motion compensation
- Flexibility in picture display ordering
- Flexibility in picture referencing
- Prediction with weights and offsets (for fading)
- Motion inference in skipped area
- Directional spatial prediction for intra coding
- In-the-loop de-blocking filtering

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

Improved Coding Efficiency

- Small block-size transform (4x4)
- Hierarchical block transform (4x4, 8x8, 16x16)
- Short word-length transform (16-bit arithmetic)
- Exact-match inverse transform (no calculation error)
- Arithmetic entropy coding (CABAC)
- Context-adaptive entropy coding (CAVLC & CABAC)

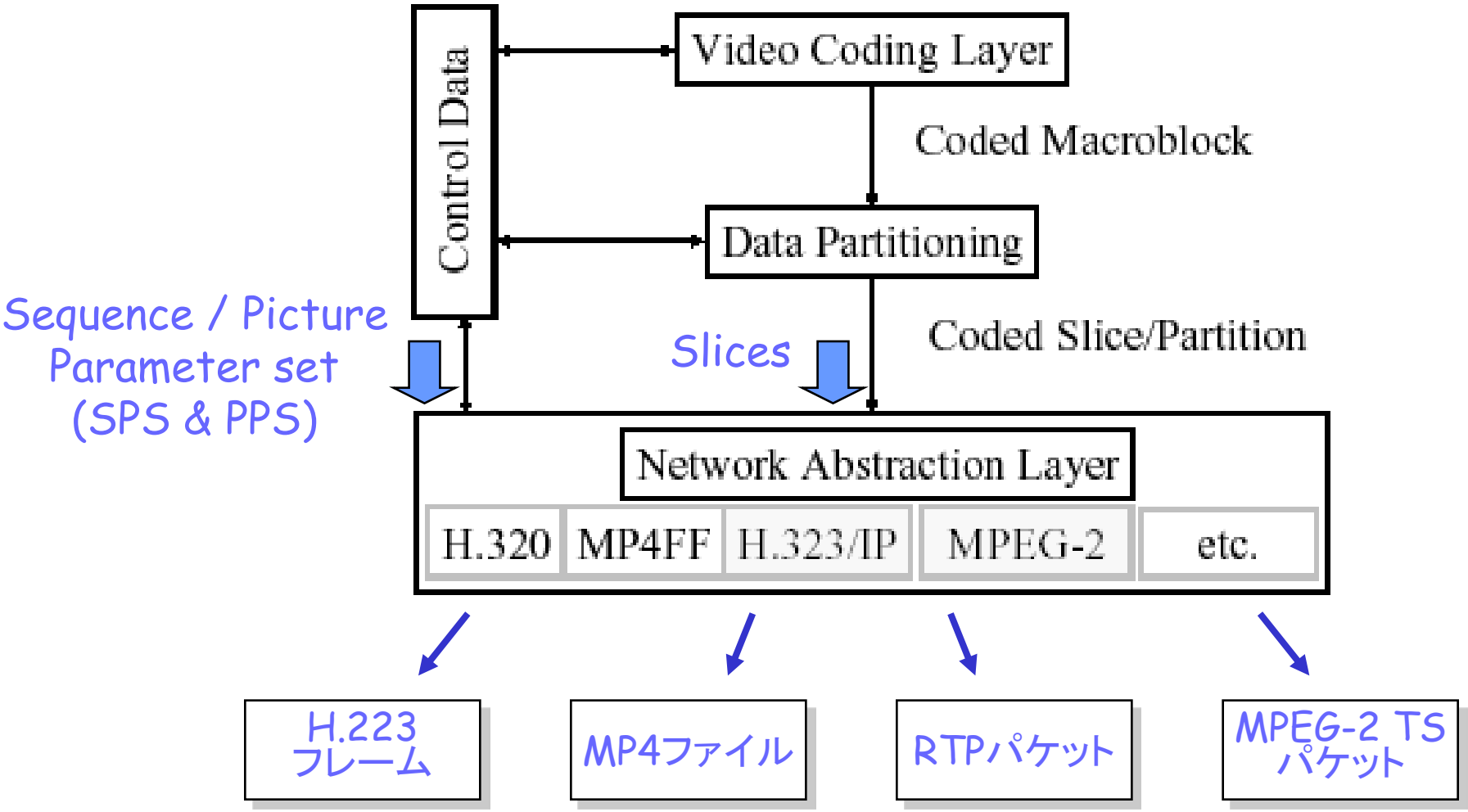
T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

Robustness & Flexibility

- Flexible "parameter set" structure
- NAL unit syntax structure (network abstraction)
- Flexible slice size (like MPEG-1)
- Flexible macro-block ordering by slice groups
- Arbitrary slice ordering
- Redundant picture representation (after data loss)
- Data partitioning (for error resiliency)
- SP/SI picture types (for decoder synchronization and switching)

T.Wiegand et al: "Overview of the H.264/AVC Video Coding Standard", IEEE CSVT, Jul.2003.

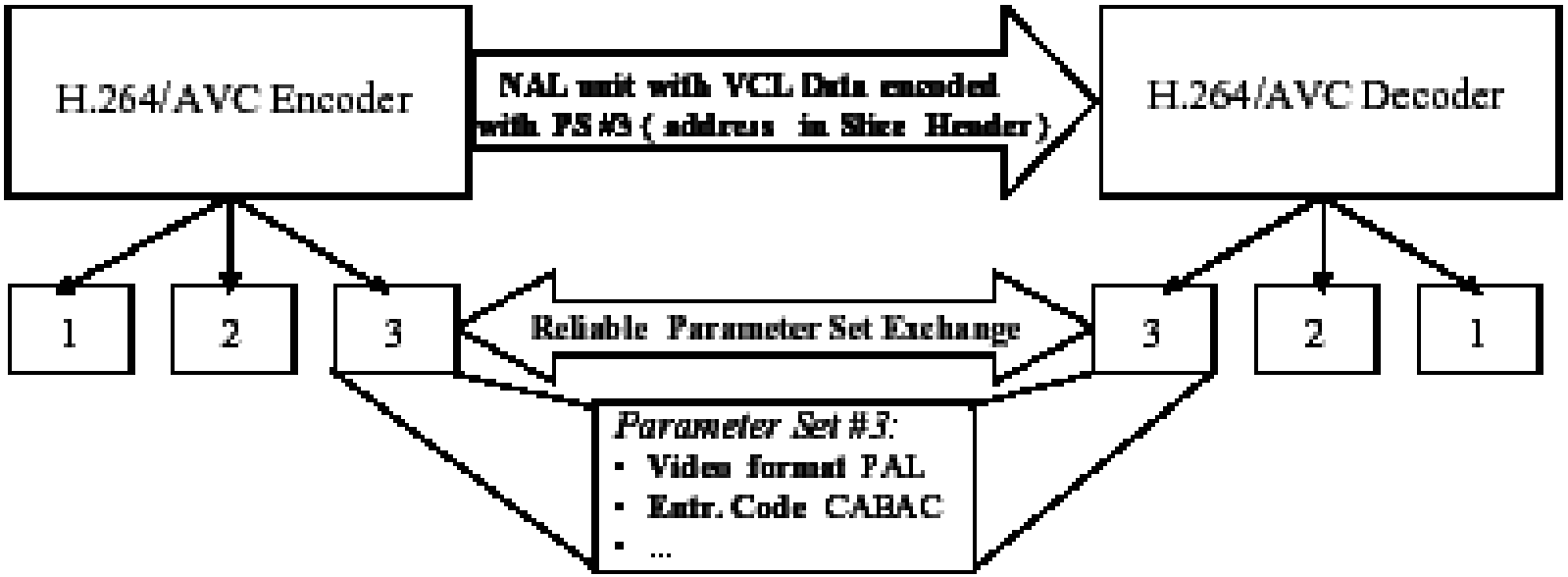
1. NAL: Network Abstraction



T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

1. NAL: Network Abstraction

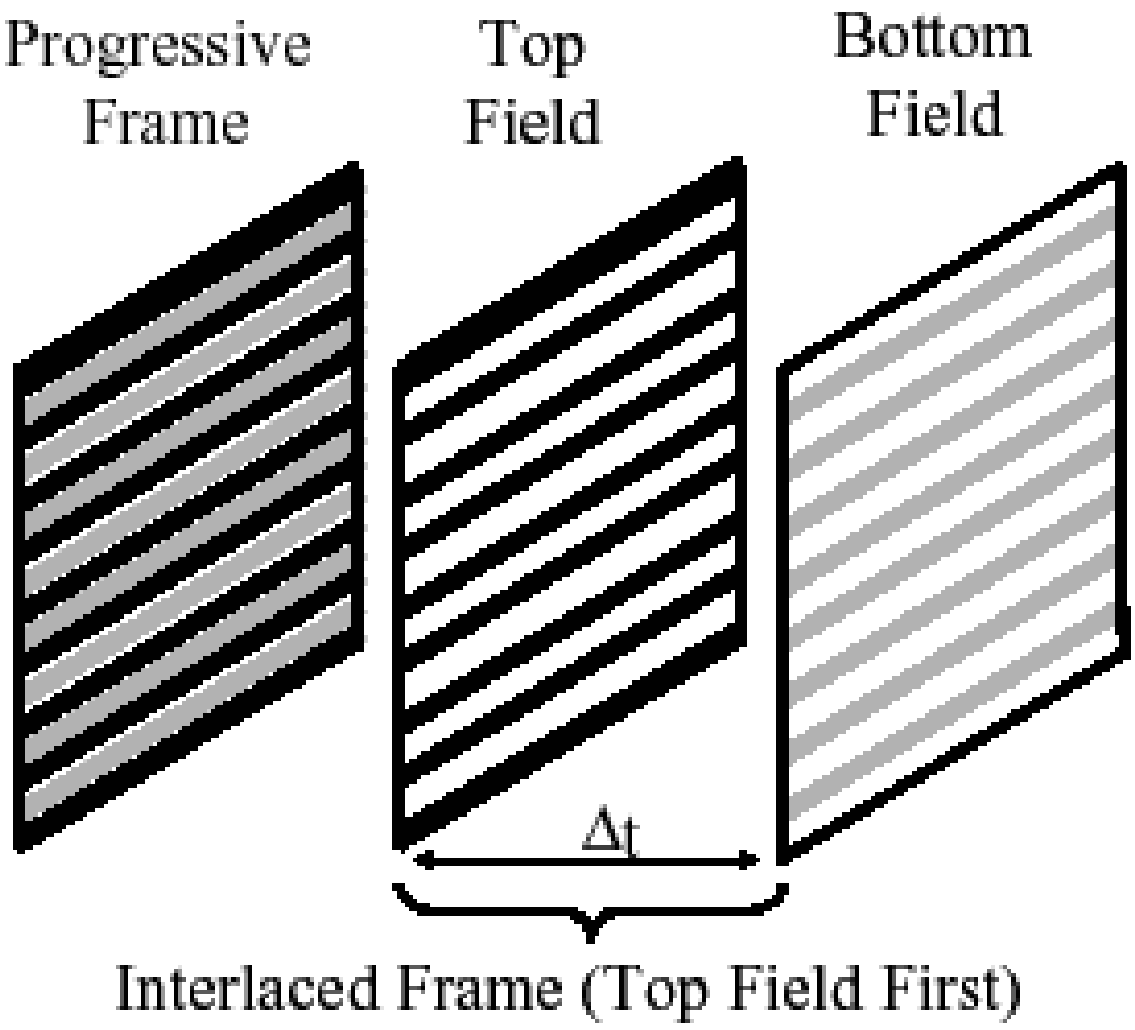
In-band signaling (e.g. by RTP)



Out-band signaling (e.g. by TCP)

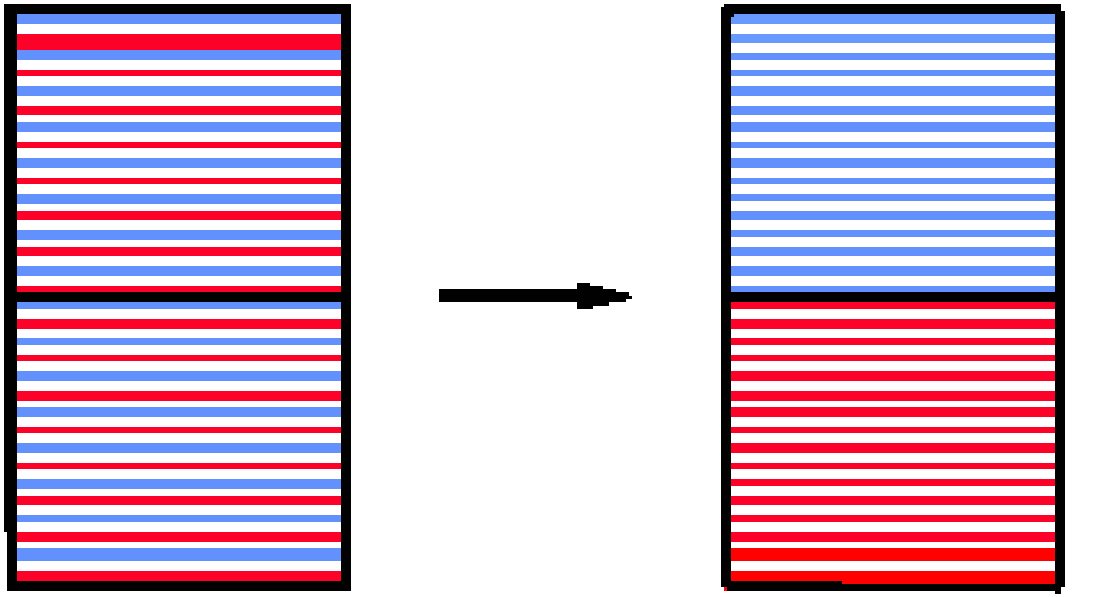
T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

2. Frame & Field



T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

2. Frame & Field



A Pair of Macroblocks
in Frame Mode

Top/Bottom Macroblocks
in Field Mode

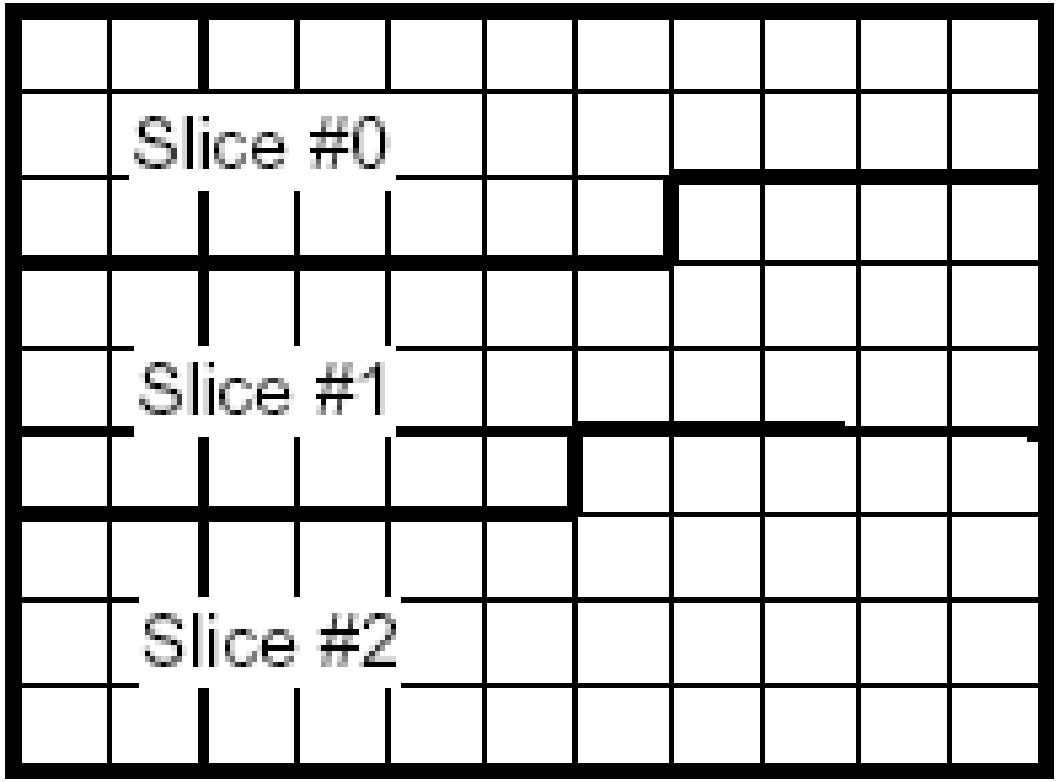
PAFF: picture-adaptive frame/field coding (odd/even fields)

MBAFF: macroblock-adaptive frame/field coding (top/bottom MBs)

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

3. Slice

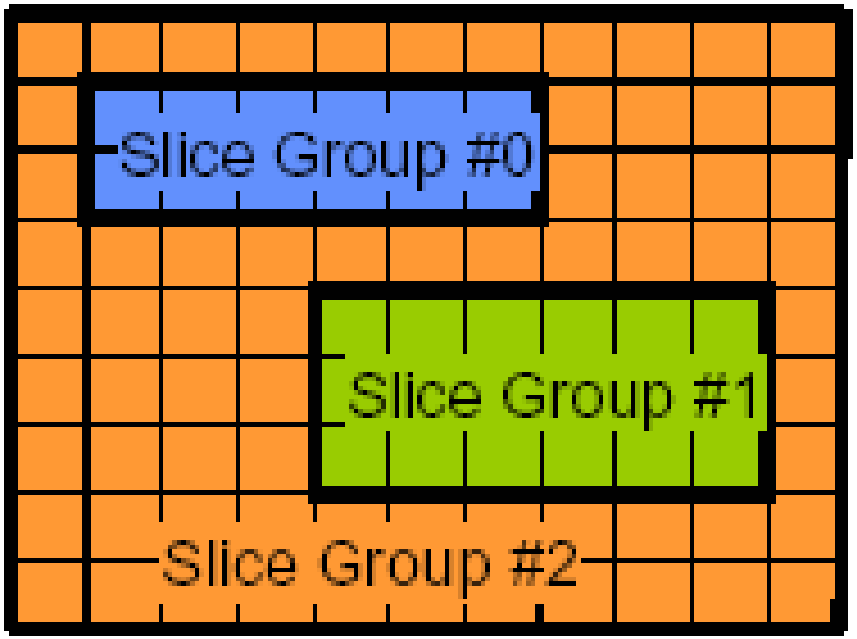
H.264/AVC の符号化の基本単位:

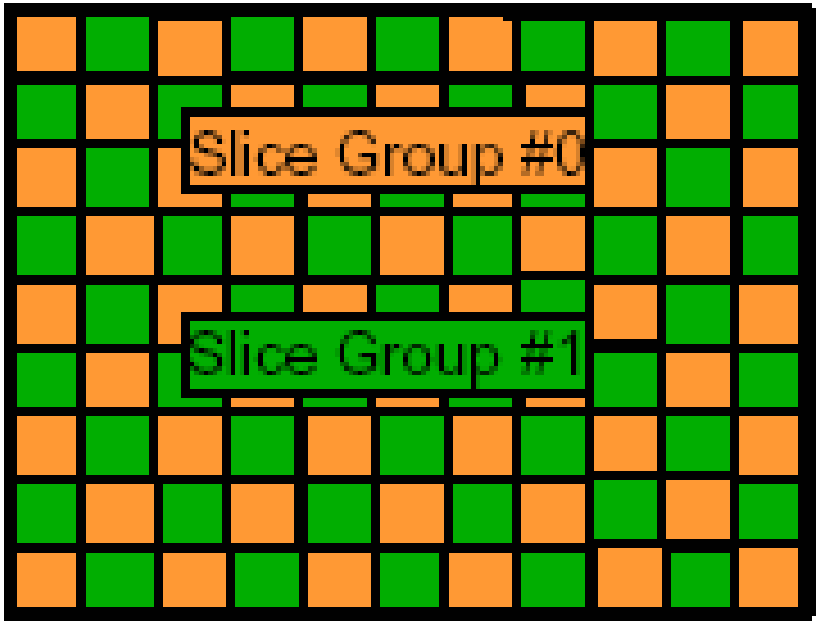


Not using FMO (flexible macroblock ordering)

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

3. Slice





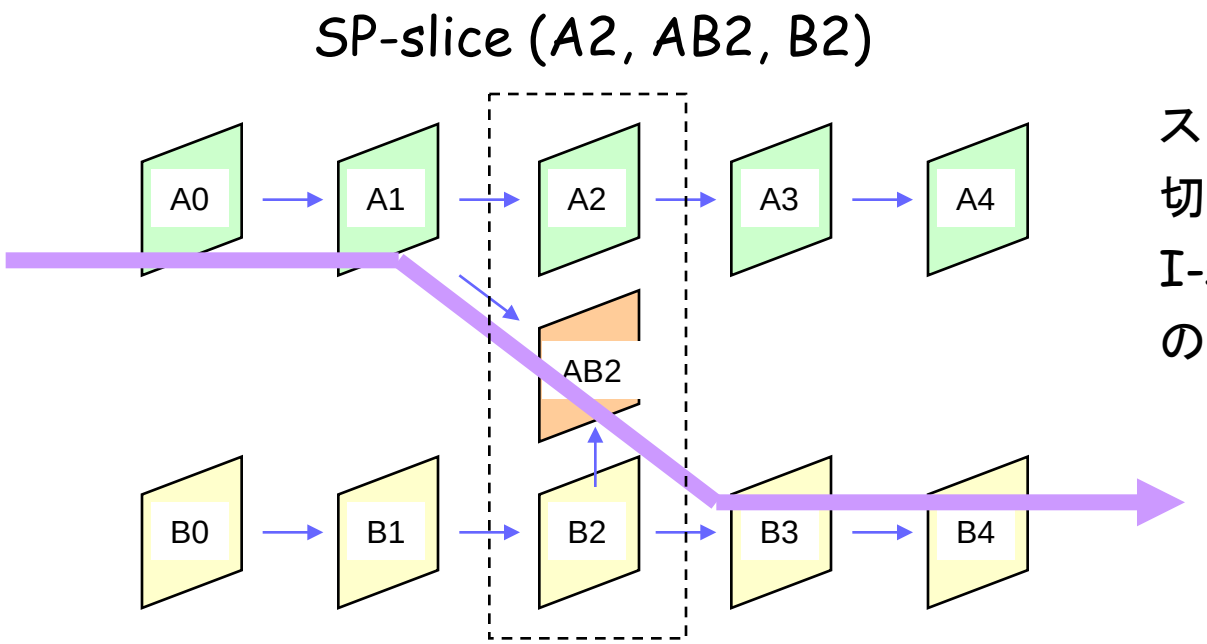
Using FMO (flexible macroblock ordering)

= Slice groups ~ ROI (Region of Interest)

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

3. Slice

- I-slice: I macroblock only
- P-slice: I,P macroblocks
- B-slice: I,P,B macroblocks
- SP/SI-slice: (for bit-stream switching)



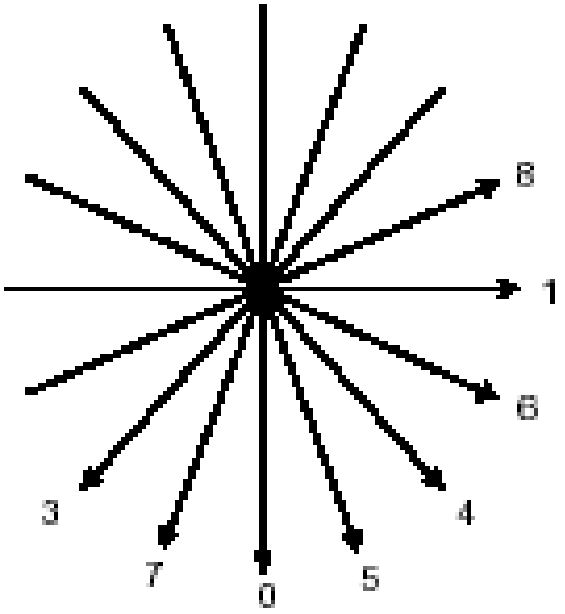
ストリーム A から B への切り替え、早送り等：
I-slice よりも SP-slice の方が効率がよい

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

4. Intra-Frame Prediction

- **Intra_4x4**

Prediction directions



encoded samples

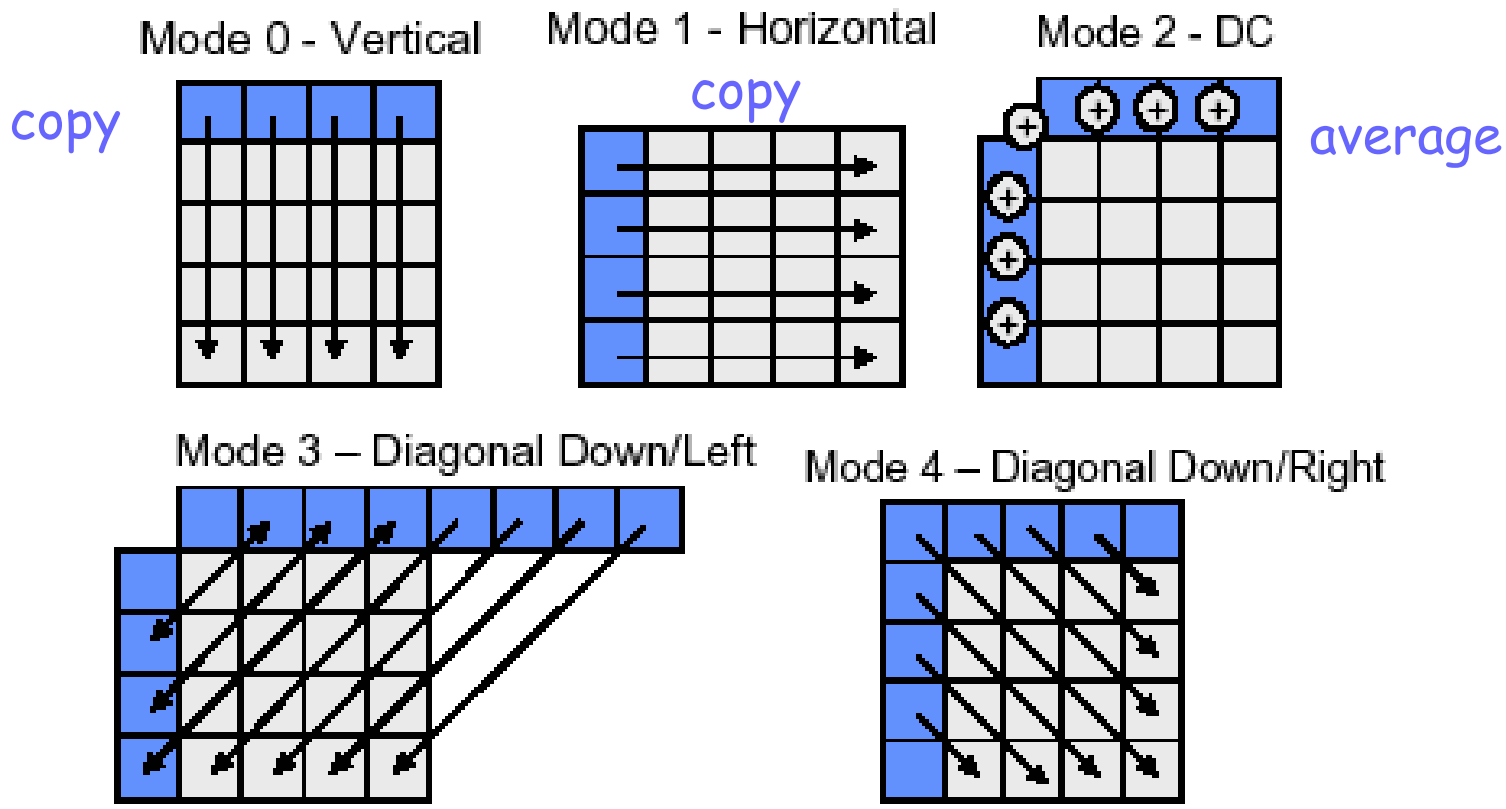
	Q	A	B	C	D	E	F	G	H
I		a	b	c	d				
J		e	f	g	h				
K		i	j	k	l				
L		m	n	o	p				

4x4 block to be predicted

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

4. Intra-Frame Prediction

- **Intra_4x4: nine modes**



Modes 3 ~ 8: diagonal-down-left, diagonal-down-right, vertical-right, horizontal-down, vertical-left, and horizontal-up

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

4. Intra-Frame Prediction

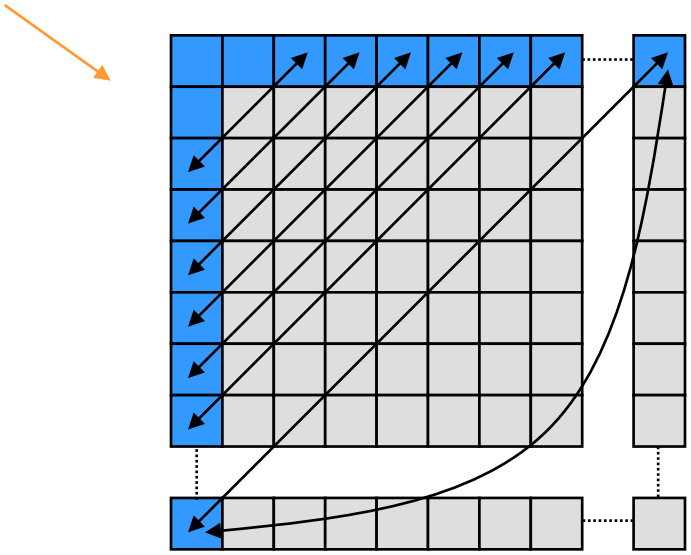
- **Intra_16x16: four modes**

  Mode 0: vertical prediction

  Mode 1: horizontal prediction

  Mode 2: DC prediction

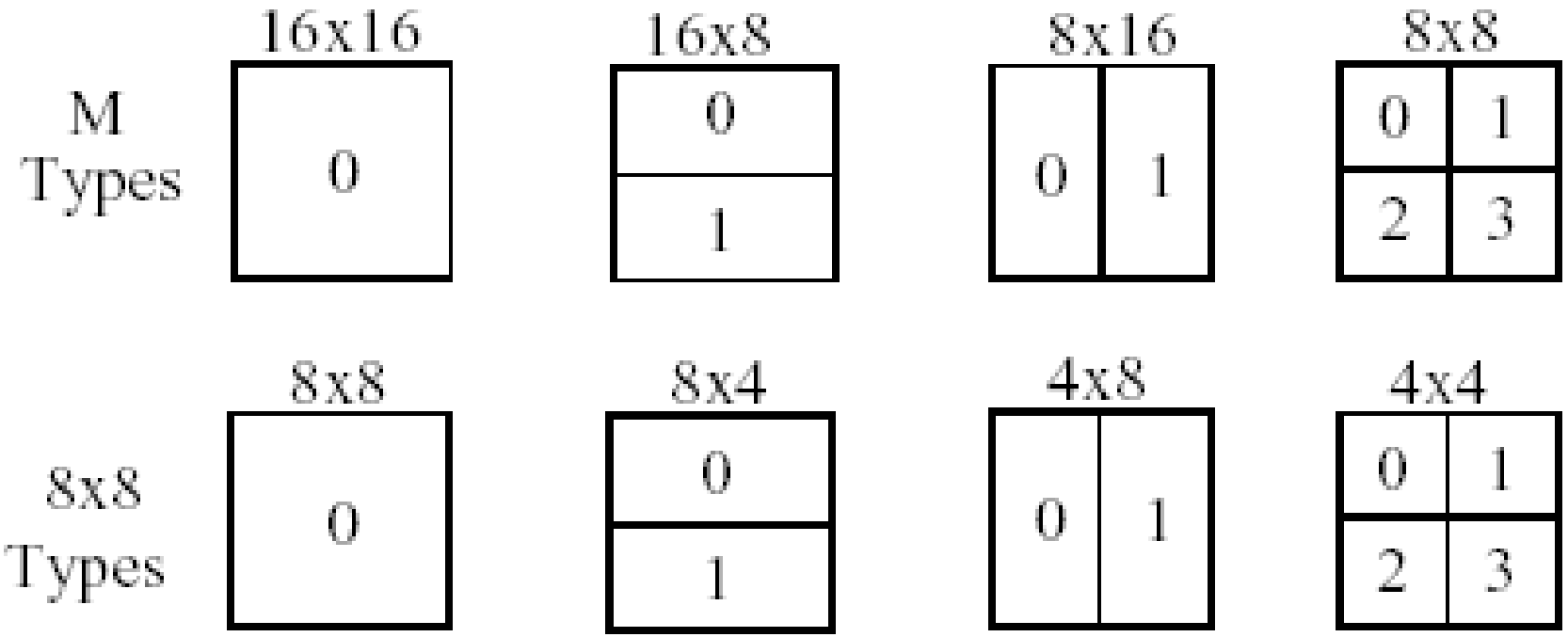
  Mode 3: plane prediction



T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

5. Inter-Frame Prediction

- Macroblock partitioning (for variable block-size MC)

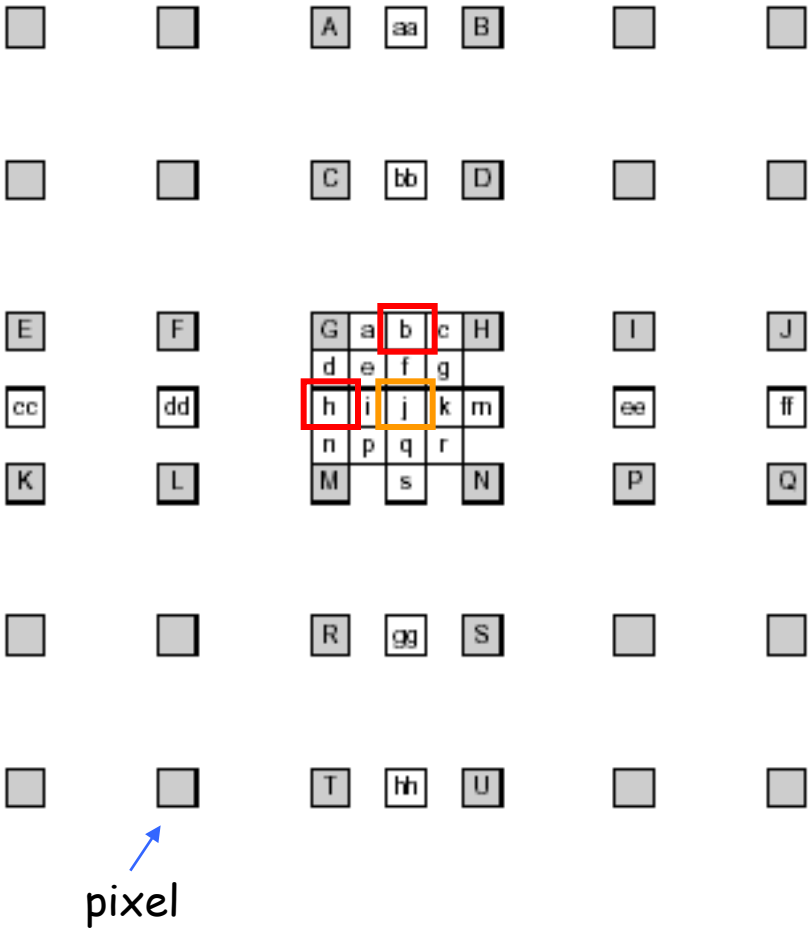


Maximally 16 motion vectors (4x4 case)

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

5. Inter-Frame Prediction

- Quarter-pel accuracy motion compensation



1/2-pel accuracy: 6-tap FIR

$$b_1 = (E - 5F + 20G + 20H - 5I + J)$$
$$b = (b_1 + 16) >> 5$$
$$h_1 = (A - 5C + 20G + 20M - 5R + T)$$
$$h = (h_1 + 16) >> 5$$
$$j_1 = (cc - 5dd + 20h_1 + 20m_1 - 5ee + ff)$$
$$j = (j_1 + 512) >> 10$$

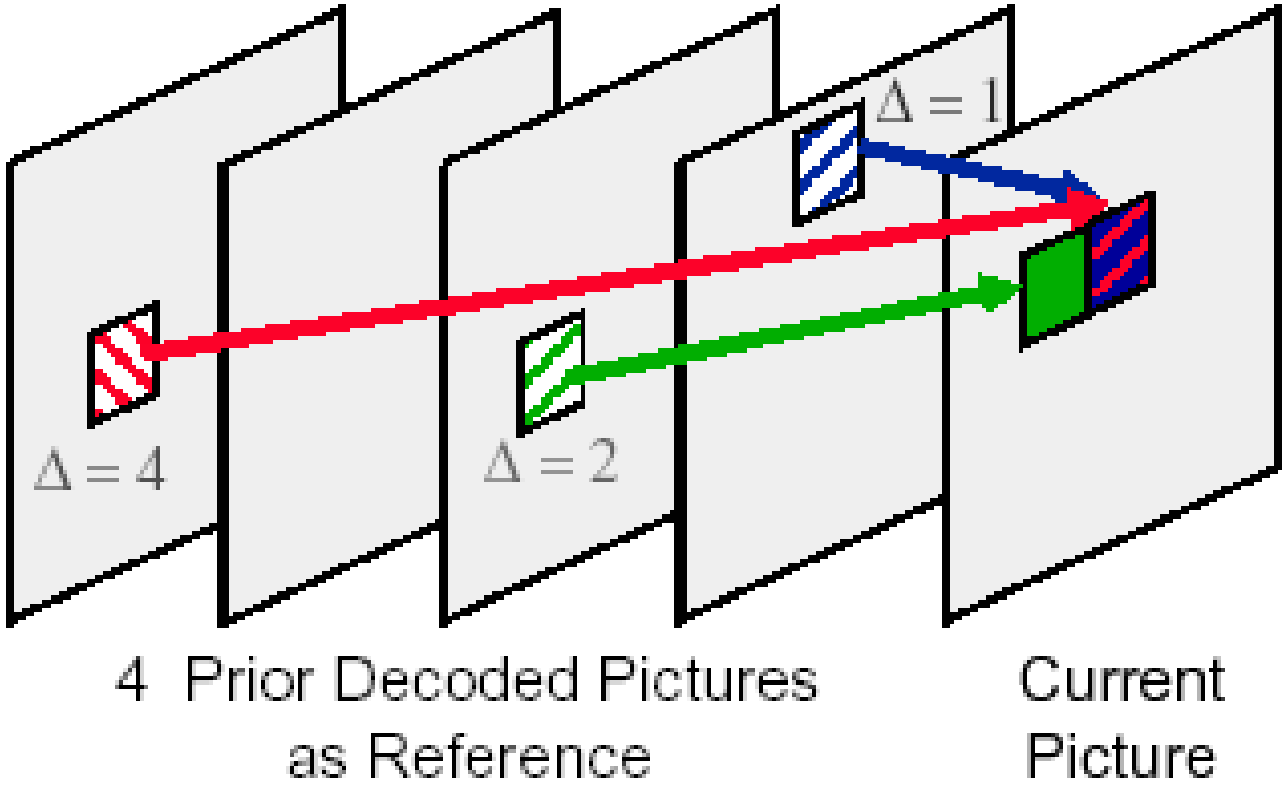
1/4-pel accuracy: 2-tap FIR

$$a = (G + b + 1) >> 1$$
$$e = (b + h + 1) >> 1$$

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

5. Inter-Frame Prediction

- Multi-frame (multi-reference) motion compensation

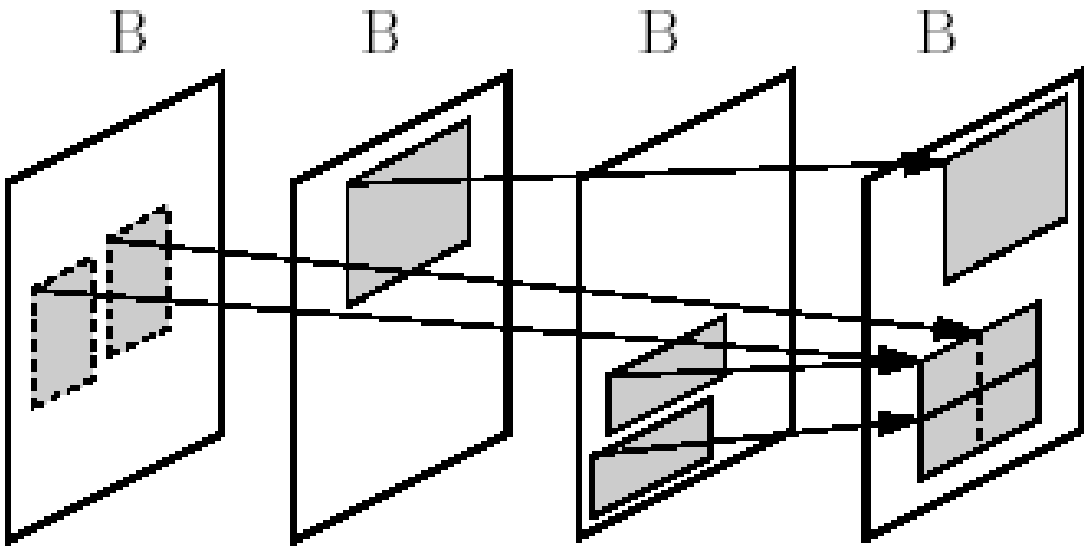


Applied to P-slices and B-slices

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

5. Inter-Frame Prediction

- **B-slices: *generalized* B-pictures**



- P-Skip mode
- B-Skip mode

+ no coefficients

- List 0 prediction
- List 1 prediction
- "Bi-predictive" prediction: using two reference pictures
- Direct prediction: MV prediction and no overhead transmission

M.Flierl et al: "Generalized B-Pictures and the Draft H.264/AVC Video Coding Standard", IEEE CSVT, Jul.2003.

6. Transform

- **4x4 integer transform**

$$H_{4x4} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

16-bit integer arithmetic

$G_{4x4} = 5.38\,\text{dB for } \rho = 0.90$

$$H_{DCT(4x4)} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ a & b & -b & -a \\ 1 & -1 & -1 & 1 \\ b & -a & a & -b \end{bmatrix}$$

$a = \sqrt{2}\cos\frac{\pi}{8} \approx 1.306,\ b = \sqrt{2}\cos\frac{3\pi}{8} \approx 0.541$ $\quad G_{DCT(4x4)} = 5.39\,\text{dB for } \rho = 0.90$

H.S.Malvar et al: “Low-Complexity Transform an Quantization in H.264/AVC”, IEEE CSVT, Jul.2003.

7. Entropy Coding

- **exp-Golomb code: for syntax elements except quantized transform coefficients**

code	syntax elements
1	0
0 1 0	1
0 1 1	2
00 1 00	3
00 1 01	4
00 1 10	5
00 1 11	6
000 1 000	7
000 1 001	8
⋮	⋮

Exp-Golomb code:

$$\underbrace{00\cdots0}_{prefix}\quad\underbrace{0}_{separator}\quad\underbrace{xx\cdots x}_{suffix}$$

$0,1,2,\cdots$

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

7. Entropy Coding

- **CAVLC: context-adaptive variable length coding (for quantized transform coefficients)**

7 6 -2 0 -1 0 0 1 0 0 0 0 0 0 0 0 (16 coefficients)

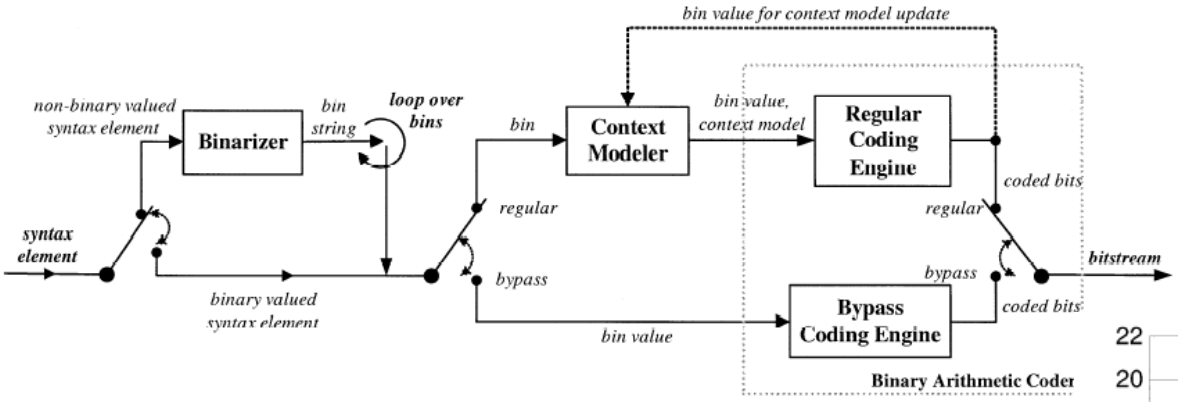
- N = 5: number of non-zero quantized coefficients
- T1s = 2: number of '±1' at the end of scan (trailing 1s)
- sign flag = '-' '+' : sign of T1s
- coefficient values in reverse order → '-2' '6' '7'
- total zeros = 3: number of zeros before the last non-zero
- run before = 2, 1: how total zeros are distributed in reverse order

Context-adaptive: 過去の符号化結果に応じて適応的に VLC table を切替える

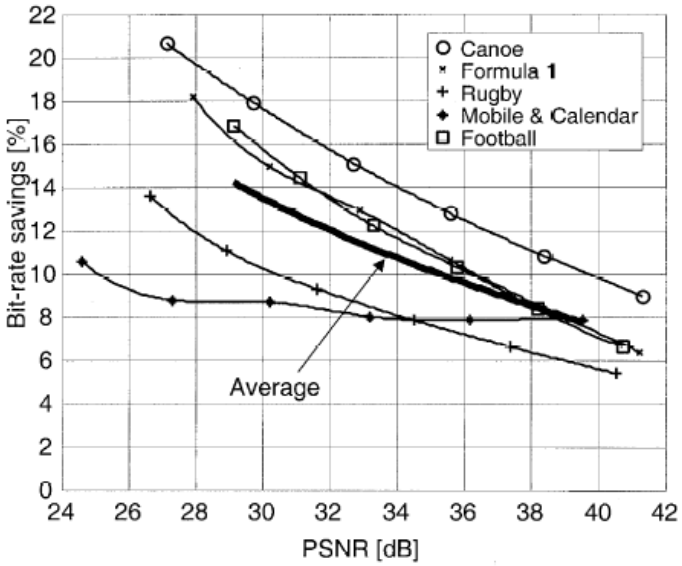
T.Wiegand et al: "Overview of the H.264/AVC Video Coding Standard", IEEE CSVT, Jul.2003.

7. Entropy Coding

- **CABAC: context-adaptive binary arithmetic coding (for quantized transform coefficients)**



vs CAVLCの改善効果

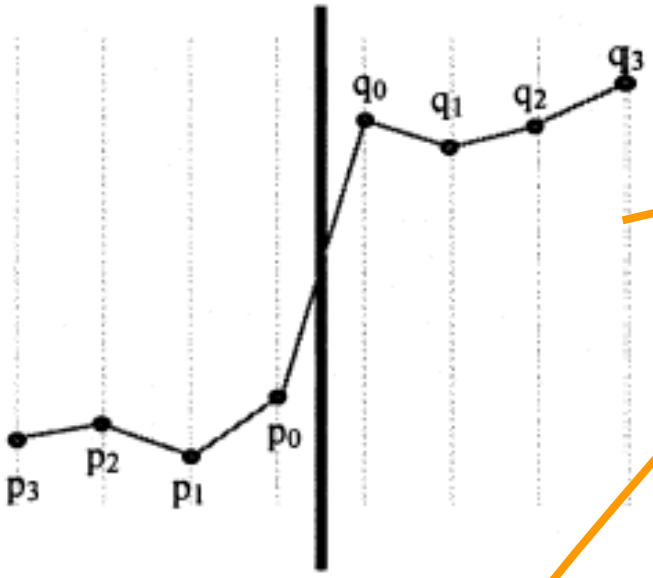


- Binarization: 多値・二値変換
- Context modeling: 確率モデル更新・適応
- Binary arithmetic coding

D.Marpe et al: “Context-Based Adaptive Binary Arithmetic Coding …”, IEEE CSVT, Jul.2003.

8. In-Loop Deblocking Filter

block boundary



「ブロックひずみ」と「エッジ」をどのように区別するか？

$$\begin{cases} |p_0 - q_0| < \alpha(QP) \\ |p_1 - p_0| < \beta(QP) \\ |q_1 - q_0| < \beta(QP) \end{cases}$$

ならばフィルタリングを実行する。

QP: Quantization Parameter (0~51)

e.g. $\alpha(QP) = 0.8 \cdot \left(2^{QP/6} - 1\right)$

$\beta(QP) = 0.5 \cdot QP - 7$

$$p_0' = (p_2 + 2p_1 + 2p_0 + 2q_0 + q_1)/8$$

$$p_1' = (p_2 + p_1 + p_0 + q_0)/4$$

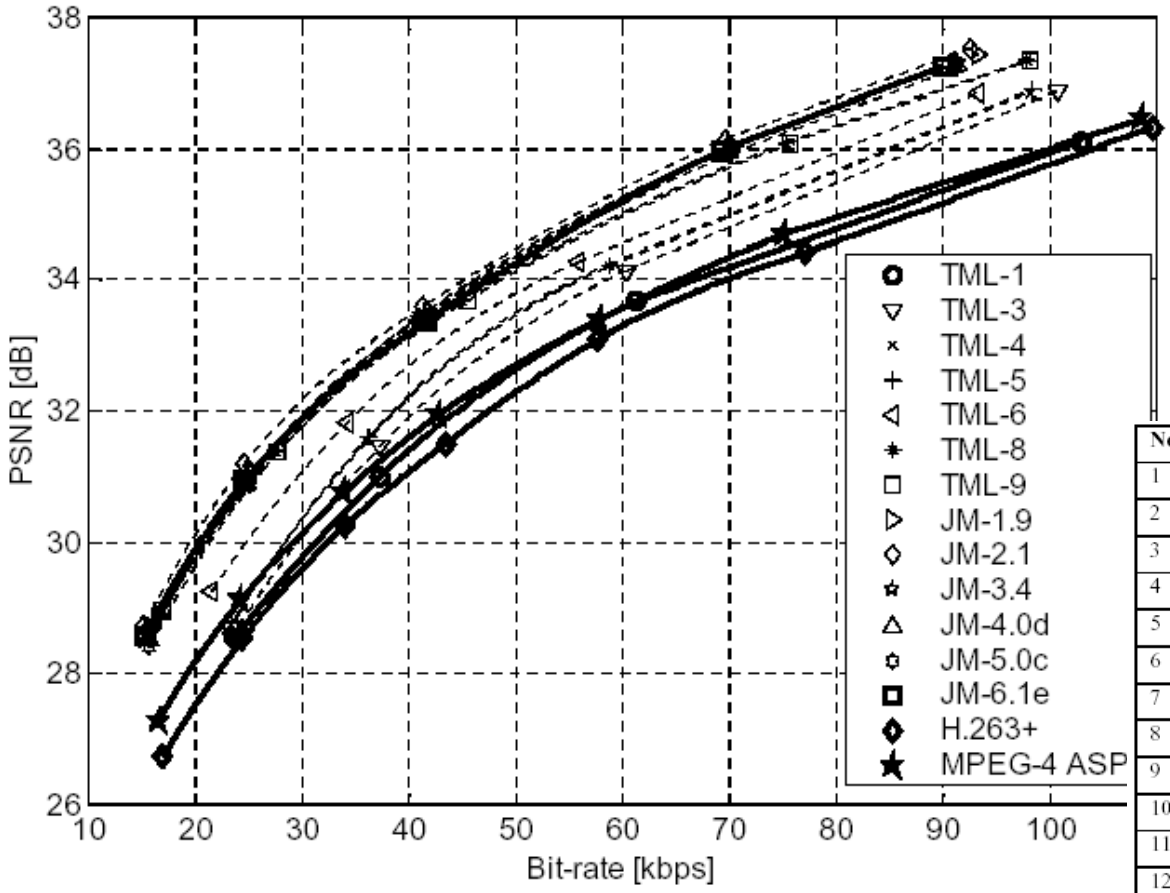
$$p_2' = (2p_3 + 3p_2 + p_1 + p_0 + q_0)/8$$

conventional

H.264/AVC

P.List et al: “Adaptive Deblocking Filter”, IEEE CSVT, Jul.2003.

H.264/AVC Evolution

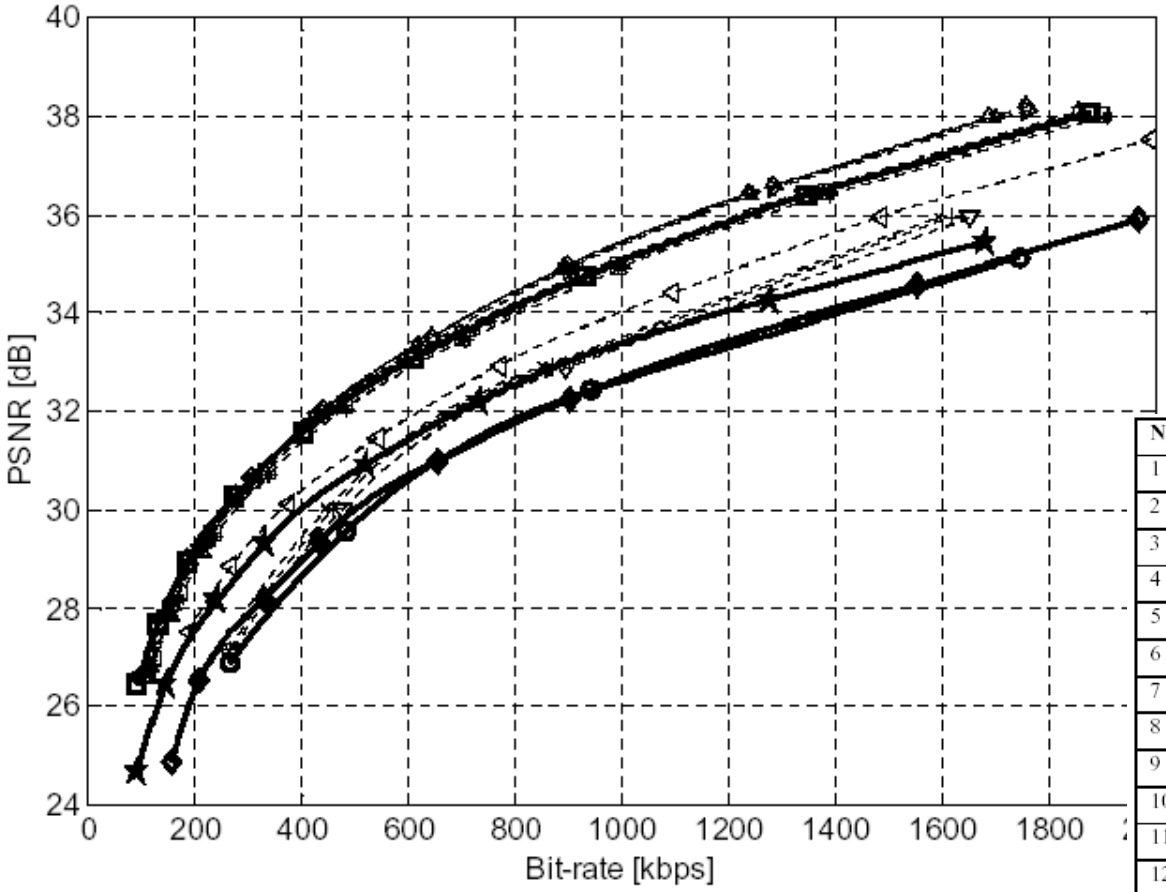


Foreman

No	TML/JM	Date	Location
1	TML-1	Aug. 1999	Berlin, Germany
2	TML-2*	Oct. 1999	Red Bank, NJ, USA
3	TML-3	Feb. 2000	Geneva, Switzerland
4	TML-4	May, 2000	Osaka, Japan
5	TML-5	Aug. 2000	Portland, OR, USA
6	TML-6	Jan. 2001	Eibsee, Germany
7	TML-7*	Apr. 2001	Austin, TX, USA
8	TML-8	May 2001	Porto Seguro, Brazil
9	TML-9	Sep. 2001	Santa Barbara, CA, USA
10	JM-1	Dec. 2001	Pattaya, Thailand
11	JM-2	Feb. 2002	Geneva, Switzerland
12	JM-3	May 2002	Fairfax, VA, USA
13	JM-4	July 2002	Klagenfurt, Austria
14	JM-5	Oct. 2002	Geneva, Switzerland
15	JM-6	Dec. 2002	Awaji, Japan
16	Final	Mar. 2003	Pattaya, Thailand

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

H.264/AVC Evolution



Tempete

No	TML/JM	Date	Location
1	TML-1	Aug. 1999	Berlin, Germany
2	TML-2*	Oct. 1999	Red Bank, NJ, USA
3	TML-3	Feb. 2000	Geneva, Switzerland
4	TML-4	May, 2000	Osaka, Japan
5	TML-5	Aug. 2000	Portland, OR, USA
6	TML-6	Jan. 2001	Eibsee, Germany
7	TML-7*	Apr. 2001	Austin, TX, USA
8	TML-8	May 2001	Porto Seguro, Brazil
9	TML-9	Sep. 2001	Santa Barbara, CA, USA
10	JM-1	Dec. 2001	Pattaya, Thailand
11	JM-2	Feb. 2002	Geneva, Switzerland
12	JM-3	May 2002	Fairfax, VA, USA
13	JM-4	July 2002	Klagenfurt, Austria
14	JM-5	Oct. 2002	Geneva, Switzerland
15	JM-6	Dec. 2002	Awaji, Japan
16	Final	Mar. 2003	Pattaya, Thailand

T.Wiegand et al: “Overview of the H.264/AVC Video Coding Standard”, IEEE CSVT, Jul.2003.

圧縮効率の改善効果 (非公式)

手法	効果
CABAC	10～15%
可変ブロック動き補償	～5%
Integer 変換	～5%
複数参照ピクチャ	～5%
R-D最適化 (次回)	10～15%
総計	30%以上