

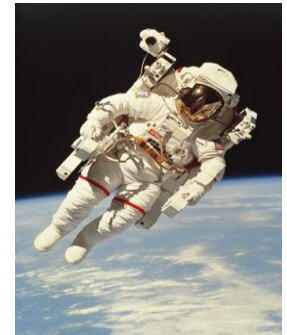
Compressão de Imagens em Movimento

Padrão H.264

PTC2547

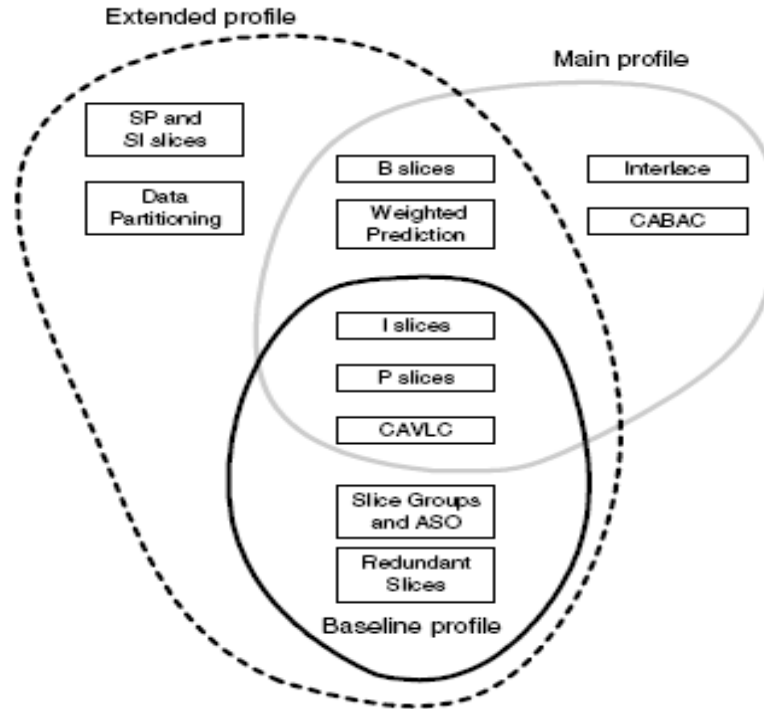
Princípios de Televisão Digital

Guido Stolfi – 09/2016



- Resumo; Níveis e Perfis
- Ferramentas de Codificação Específicas
- Comparações entre MPEG-2 e H.264
- Codificador HEVC H.265

- Finalizado em Março de 2003 e aprovado pelo ITU-T em Maio de 2003
- Padrão Internacional 14496-10 ou MPEG-4 part 10 AVC (Advanced Video Coder) do ISO/IEC
- Complexidade computacional: ~5x MPEG-2
- Ganho de até 50% na taxa de bits em relação ao MPEG-2



Perfil	Recursos Adicionais
Baseline	Sem imagens tipo B, 8 bits, 4:2:0
Extended	Com imagens B, Fatias SI e SP, c/ entrelaçamento
Main	Com CABAC, sem fatias SI e SP, sem ordem arbitrária de macroblocos
High	Com transformada 8 x 8 e 4 x 4
High 10	Amostras de 8, 9 ou 10 bits
High 4:2:2	Formato 4 : 2 : 2
High 4:4:4	Formato 4 : 4 : 4

- Codificação de imagens (fatias) tipo “I” e tipo “P”
- Compactação de dados por CAVLC (Context Adaptive Variable Length Coding)
- Baixa latência
- Aplicações: Videofone, videoconferência, telefonia celular, etc.

- Codificação de imagens (fatias) tipo “I”, “P” e “B”, com predição ponderada
- Fatias SI e SP para permitir chaveamento entre fluxos de dados diferentes
- Particionamento de Dados permite maior robustez a erros de transmissão
- Aplicações: “streaming”, edição de vídeo digital

- Codificação de imagens (fatias) tipo “I”, “P” e “B”, com predição ponderada
- Suporte para vídeo com entrelaçamento
- Compactação com CABAC (Context Adaptive Binary Arithmetic Coding)
- Preditor para blocos tipo “I”
- Sem fatias SI e SP, sem particionamento de dados
- Aplicações: radiodifusão, produção e edição de vídeo

- Transformada 8 x 8 e 4 x 4
- Digitalização das amostras com 9 a 12 bits
- Amostragem de Crominância com maior resolução
- Aplicações: produção de vídeo, cinema digital, aplicações profissionais

Nível	Formato de Vídeo (macroblocos por quadro)	Taxa de Bits (baseline ~ high)	Taxa de Bits (high 4:4:4)
1	Até 396	~ 1 Mb/s	~ 3 Mb/s
2	Até 1620	2 ~ 5 Mb/s	8 ~ 16 Mb/s
3	Até 5120	10 ~ 25 Mb/s	40 ~ 80 Mb/s
4	Até 8704	20 ~ 60 Mb/s	80 ~ 200 Mb/s
5	Até 36864	100 ~ 300 Mb/s	0,5 ~ 0,9 Gb/s

Recursos	Baseline	Extended	Main	High	High 10	High 4:2:2	High 4:4:4
Fatias I e P	S	S	S	S	S	S	S
Fatias B		S	S	S	S	S	S
Fatias SI e SP		S					
Imagens de Referencia Múltiplas	S	S	S	S	S	S	S
Filtro anti-blocagem	S	S	S	S	S	S	S
Codificação CAVLC	S	S	S	S	S	S	S
Codificação CABAC				S	S	S	S
Ordenação Flexível de Macroblocos (FMO)	S	S					
Orientação Arbitrária de Fatias (ASO)	S	S					
Transformada Adaptativa 8x8 vs. 4x4				S	S	S	S

Recursos	Baseline	Extended	Main	High	High 10	High 4:2:2	High 4:4:4
Fatias Redundantes (RS)	S	S					
Particionamento de Dados		S					
Video Entrelaçado (PicAFF, MBAFF)		S	S	S	S	S	S
Formato 4:2:0 de Crominância	S	S	S	S	S	S	S
Vídeo Monocromático (4:0:0)				S	S	S	S
Formato 4:2:2						S	S
Formato 4:4:4							S
Amostras de 8 bits	S	S	S	S	S	S	S
Amostras de 9 e 10 bits					S	S	S
Amostras de 11 e 12 bits							S

Nível	Macroblos por segundo (max)	Macroblocos por Imagem (max)	Taxa de bits p/ Baseline, Ext. e Main (max)	Taxa de bits para perfil High (max)	Taxa de bits para High 10 (max)	Taxa de bits para High 4:2:2 e 4:2:0 (max)
1	1485	99	64 kb/s	80 kb/s	192 kb/s	256 kb/s
1 b	1485	99	128 kb/s	160 kb/s	384 kb/s	512 kb/s
1.1	3000	396	192 kb/s	240 kb/s	576 kb/s	768 kb/s
1.2	6000	396	384 kb/s	480 kb/s	1152 kb/s	1536 kb/s
1.3	11880	396	768 kb/s	960 kb/s	2304 kb/s	3072 kb/s
2	11880	396	2 Mb/s	2,5 Mb/s	6 Mb/s	8 Mb/s
2.1	19800	792	4 Mb/s	5 Mb/s	12 Mb/s	16 Mb/s
2.2	20250	1620	4 Mb/s	5 Mb/s	12 Mb/s	16 Mb/s

Nível	Macroblos por segundo (max)	Macroblocos por Imagem (max)	Taxa de bits p/ Baseline, Ext. e Main (max)	Taxa de bits para perfil High (max)	Taxa de bits para High 10 (max)	Taxa de bits para High 4:2:2 e 4:2:0 (max)
3	40500	1620	10	12,5 Mb/s	30 Mb/s	40 Mb/s
3.1	108000	3600	14	17,5 Mb/s	42 Mb/s	56 Mb/s
3.2	216000	5120	20	25 Mb/s	60 Mb/s	80 Mb/s
4	245760	8192	20	25 Mb/s	60 Mb/s	80 Mb/s
4.1	245760	8192	50 Mb/s	62,5 Mb/s	150 Mb/s	200 Mb/s
4.2	522240	8704	50 Mb/s	62,5 Mb/s	150 Mb/s	200 Mb/s
5	589824	22080	135 Mb/s	168,75 Mb/s	405 Mb/s	540 Mb/s
5.1	983040	36864	240 Mb/s	300Mb/s	720 Mb/s	960 Mb/s

Nível				1	1.1	1.2	1.3
Macroblocos/quadro (max)				99	396	396	396
Macroblocos/segundo (max)				1485	3.000	6.000	11.800
Formato	Largura	Altura	MB's	Quadros por segundo (max)			
SQCIF	128	96	48	30,9	62,5	125,0	172
QCIF	176	144	99	15,0	30,3	60,6	120
QVGA	320	240	300	-	10,0	20,0	39,6
525 SIF	352	240	330	-	9,1	18,2	36,0
CIF	352	288	396	-	7,6	15,2	30,0
525 HHR	352	480	660	-	-	-	-

Nível				2	2.1	2.2
Macroblocos/quadro (max)				396	792	1.620
Macroblocos/segundo (max)				11.880	19.800	20.250
Formato	Largura	Altura	MB's	Quadros por segundo (max)		
SQCIF	128	96	48	172,0	172,0	172,0
QCIF	176	144	99	120,0	172,0	172,0
QVGA	320	240	300	39,6	66,0	67,5
525 SIF	352	240	330	36,0	60,0	61,4
CIF	352	288	396	30,0	50,0	51,1
525 HHR	352	480	660	-	30,0	30,7
625 HHR	352	576	792	-	25,0	25,6
VGA	640	480	1200	-	-	16,9
524 4SIF	704	480	1320	-	-	15,3
525 SD (D-1)	720	480	1350	-	-	15,0
4CIF	704	576	1584	-	-	12,8
625 SD	720	576	1620	-	-	12,5

Nível				3	3.1	3.2
Macroblocos/quadro (max)				1.620	3.600	5.120
Macroblocos/segundo (max)				40.500	108.000	21.6000
Formato	Largura	Altura	MB's	Quadros por segundo (max)		
QVGA	320	240	300	135,0	172,0	172,0
525 SIF	352	240	330	122,7	172,0	172,0
CIF	352	288	396	102,3	172,0	172,0
525 HHR	352	480	660	61,4	163,6	172,0
625 HHR	352	576	792	51,1	136,4	172,0
VGA	640	480	1200	33,8	90,0	172,0
524 4SIF	704	480	1320	30,7	81,8	163,6
525 SD (D-1)	720	480	1350	30,0	80,0	160,0
4CIF	704	576	1584	25,6	68,2	136,4
625 SD	720	576	1620	25,0	66,7	133,3
SVGA	800	600	1900	-	56,8	113,7
XGA (XVGA)	1024	768	3072	-	35,2	70,3
720p HD	1280	720	3600	-	30,0	60,0
4VGA	1280	960	4800	-	-	45,0
SXGA	1280	1024	5120	-	-	42,2

Nível				4	4.1	4.2
Macroblocos/quadro (max)				8.192	8.192	8.192
Macroblocos/segundo (max)				245.760	245.760	589.824
Formato	Largura	Altura	MB's	Quadros por segundo (max)		
525 SD (D-1)	720	480	1350	172,0	172,0	172,0
4CIF	704	576	1584	155,2	155,2	172,0
625 SD	720	576	1620	151,7	151,7	172,0
SVGA	800	600	1900	129,3	129,3	172,0
XGA (XVGA)	1024	768	3072	80,0	80,0	160,0
720p HD	1280	720	3600	68,3	68,3	136,5
4VGA	1280	960	4800	51,2	51,2	102,4
SXGA	1280	1024	5120	48,0	48,0	96,0
525 16SIF	1408	960	5280	46,5	46,5	93,1
16CIF	1408	1152	6336	38,8	38,8	77,6
4SVGA (UXGA)	1600	1200	7500	32,8	32,8	65,5
1080 HD	1920	1088	8160	30,1	30,1	60,2
2K x 1K	2048	1024	8192	30,0	30,0	60,0
4XGA	2048	1536	12288	-	-	-

Nível				5	5.1
Macroblocos/quadro (max)				21.696	36.864
Macroblocos/segundo (max)				589.824	983.040
Formato	Largura	Altura	MB's	Quadros por segundo (max)	
XGA (XVGA)	1024	768	3072	172,0	172,0
720p HD	1280	720	3600	163,8	172,0
4VGA	1280	960	4800	122,9	172,0
SXGA	1280	1024	5120	115,2	172,0
525 16SIF	1408	960	5280	111,7	172,0
16CIF	1408	1152	6336	93,1	155,2
4SVGA (UXGA)	1600	1200	7500	78,6	131,1
1080 HD	1920	1088	8160	72,3	120,5
2K x 1K	2048	1024	8192	72,0	120,0
4XGA	2048	1536	12288	48,0	80,0
16VGA	2560	1920	19200	30,7	51,2
3616 x 1536 (2,53 : 1)	3616	1536	21696	27,2	45,3
3676 x 1536 (2,39 : 1)	3680	1536	22080	26,7	44,5
4K x 2K	4096	2048	32768	-	30,0
4096 x 2304 (16 : 9)	4096	2304	36864	-	26,7

Ferramentas de Codificação do H.264



Figure 3.12 Residual (no motion compensation)



Figure 3.13 Residual (16×16 block size)

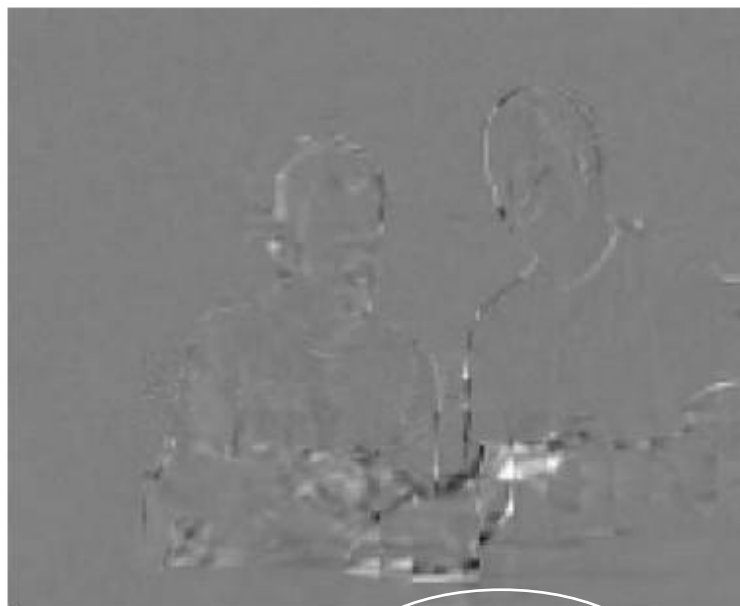


Figure 3.14 Residual (8×8 block size)

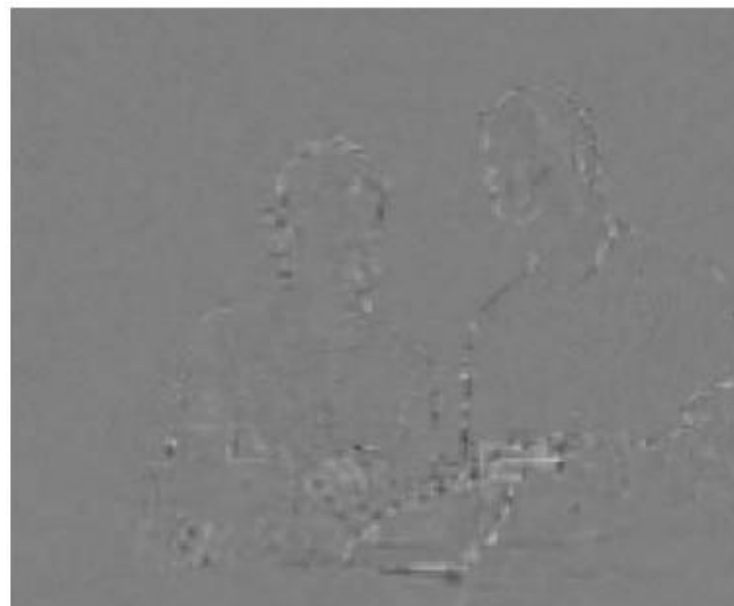


Figure 3.15 Residual (4×4 block size)

- Quanto menor o tamanho do bloco utilizado para predição de movimento, menor o erro resultante.
- Isso implica em maior carga computacional no mecanismo de busca, além de maior quantidade de bits transmitidos (maior quantidade de vetores de movimento).
- O padrão H.264 usa então blocos de tamanho variável, que podem ser escolhidos de acordo com as características de homogeneidade do movimento detectado.

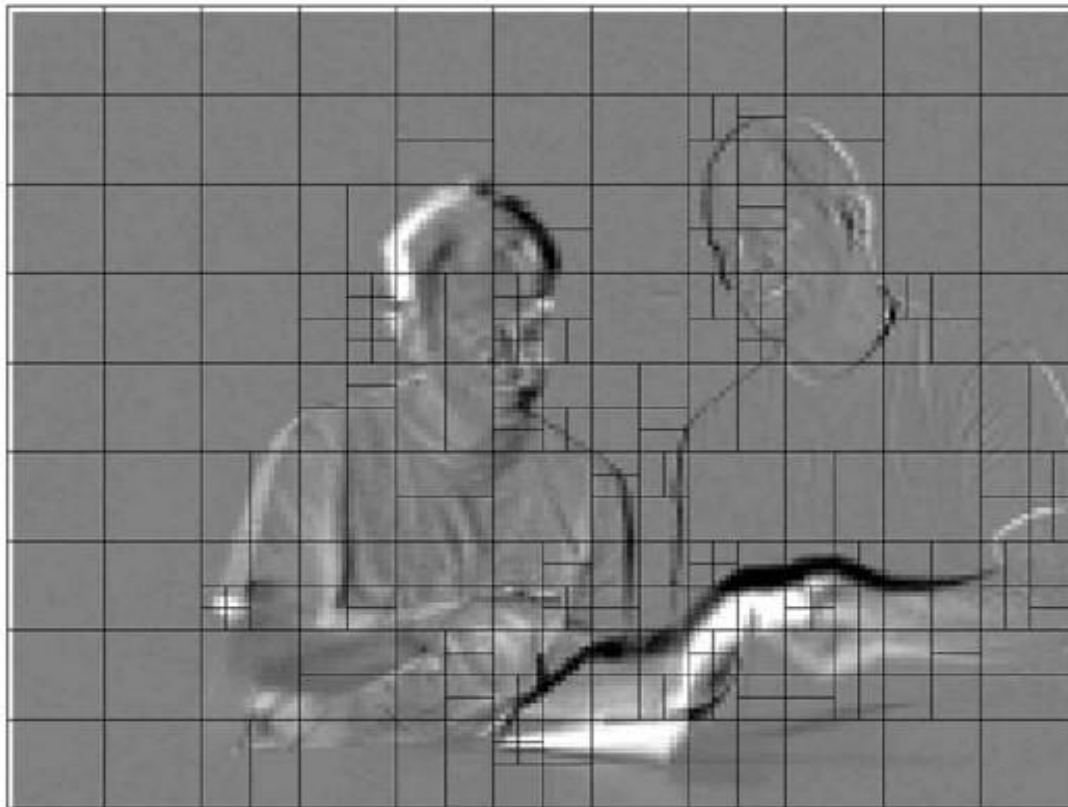
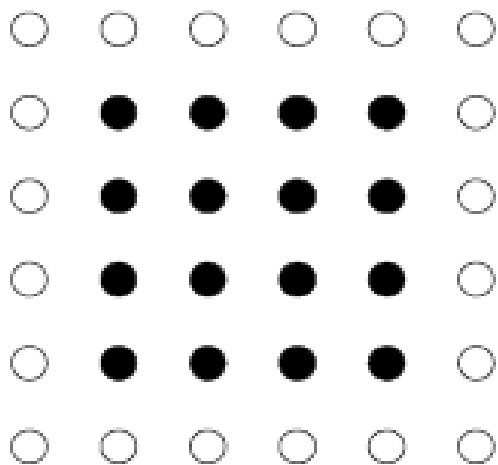
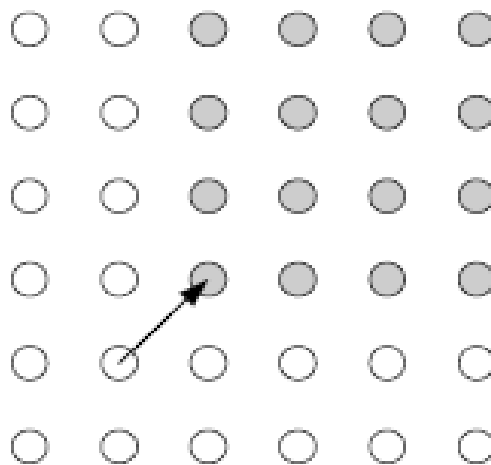


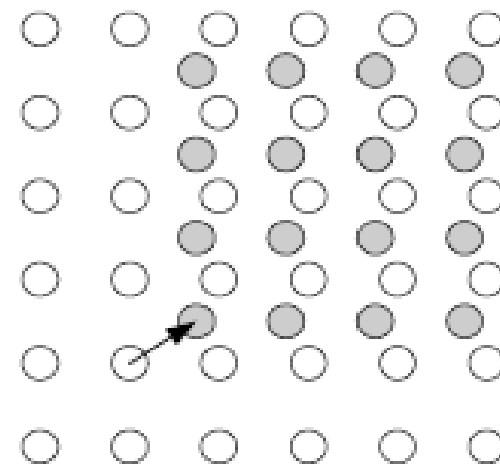
Figure 6.12 Residual (without MC) showing choice of block sizes



Macrobloco
a ser compensado



Vetor de Movimento
com precisão de
1 pixel



Vetor de Movimento
com precisão
fracionária

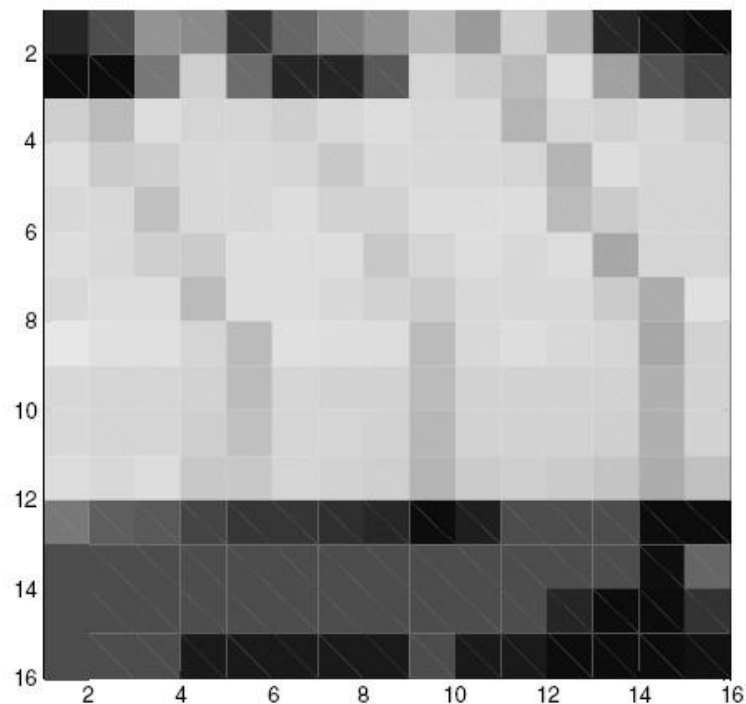


Figure 3.16 Close-up of reference region

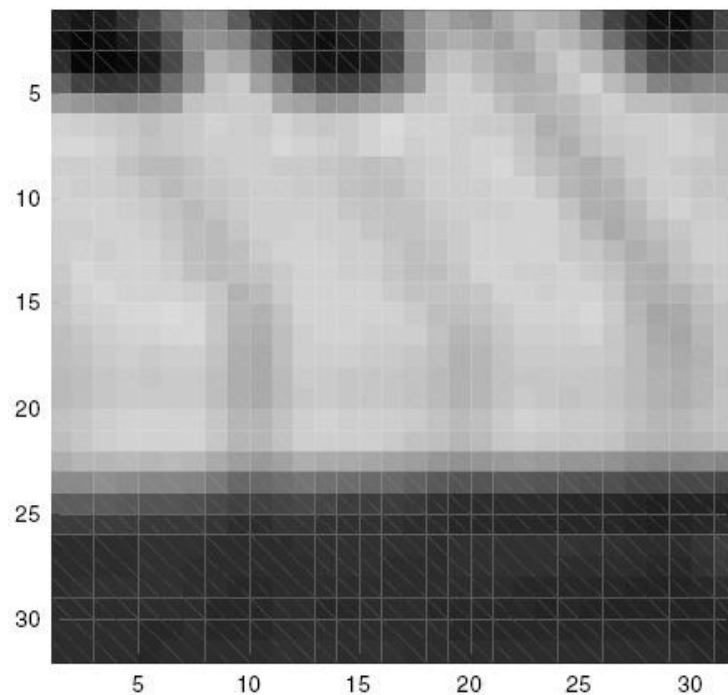


Figure 3.17 Reference region interpolated to half-pixel positions

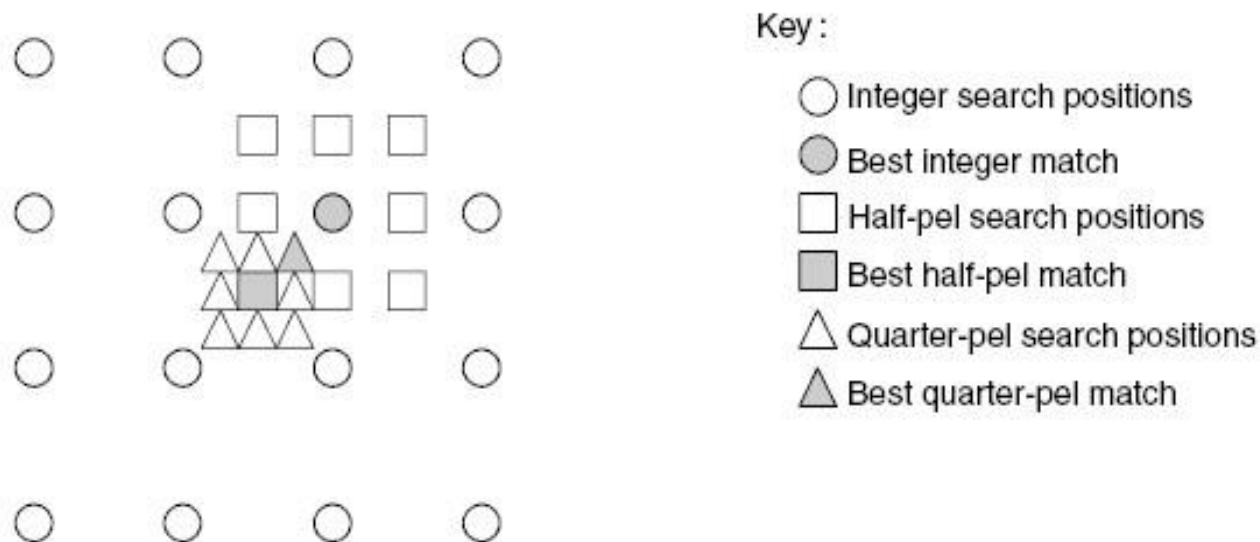
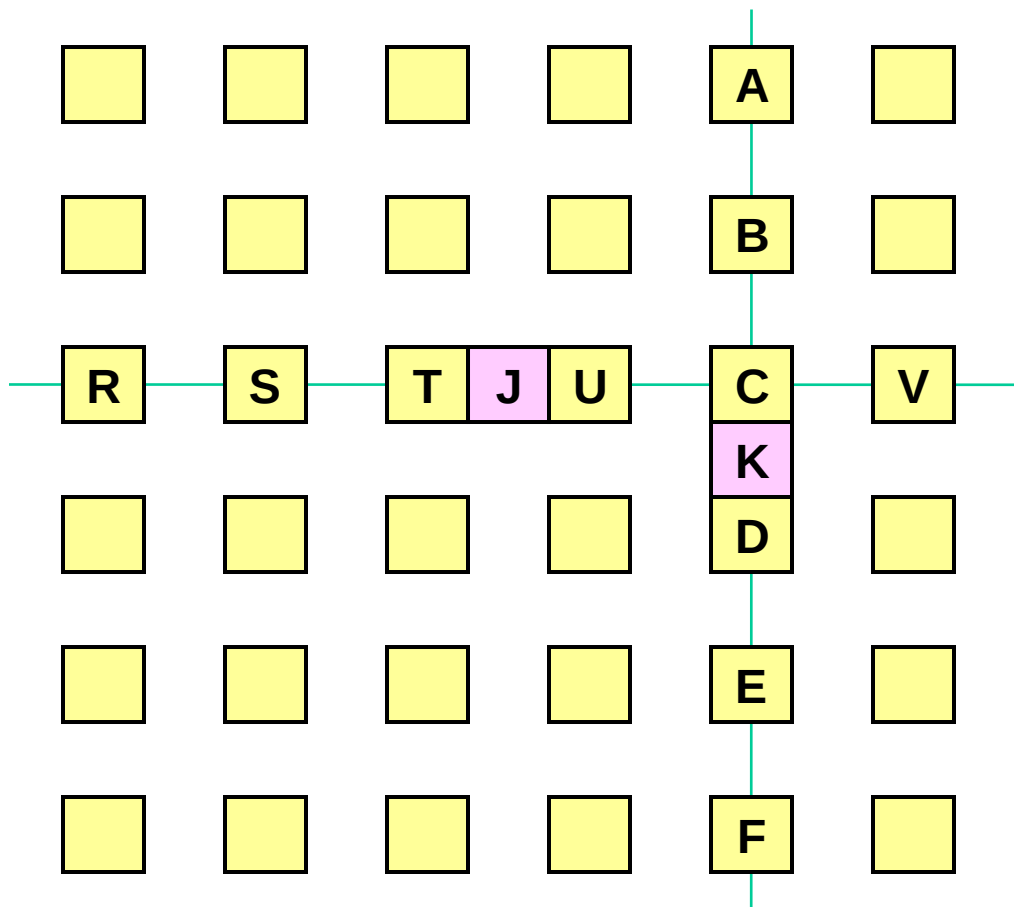


Figure 3.18 Integer, half-pixel and quarter-pixel motion estimation

Table 3.1 SAE of residual frame after motion compensation (16×16 block size)

Sequence	No motion compensation	Integer-pel	Half-pel	Quarter-pel
'Violin', QCIF	171945	153475	128320	113744
'Grasses', QCIF	248316	245784	228952	215585
'Carphone', QCIF	102418	73952	56492	47780

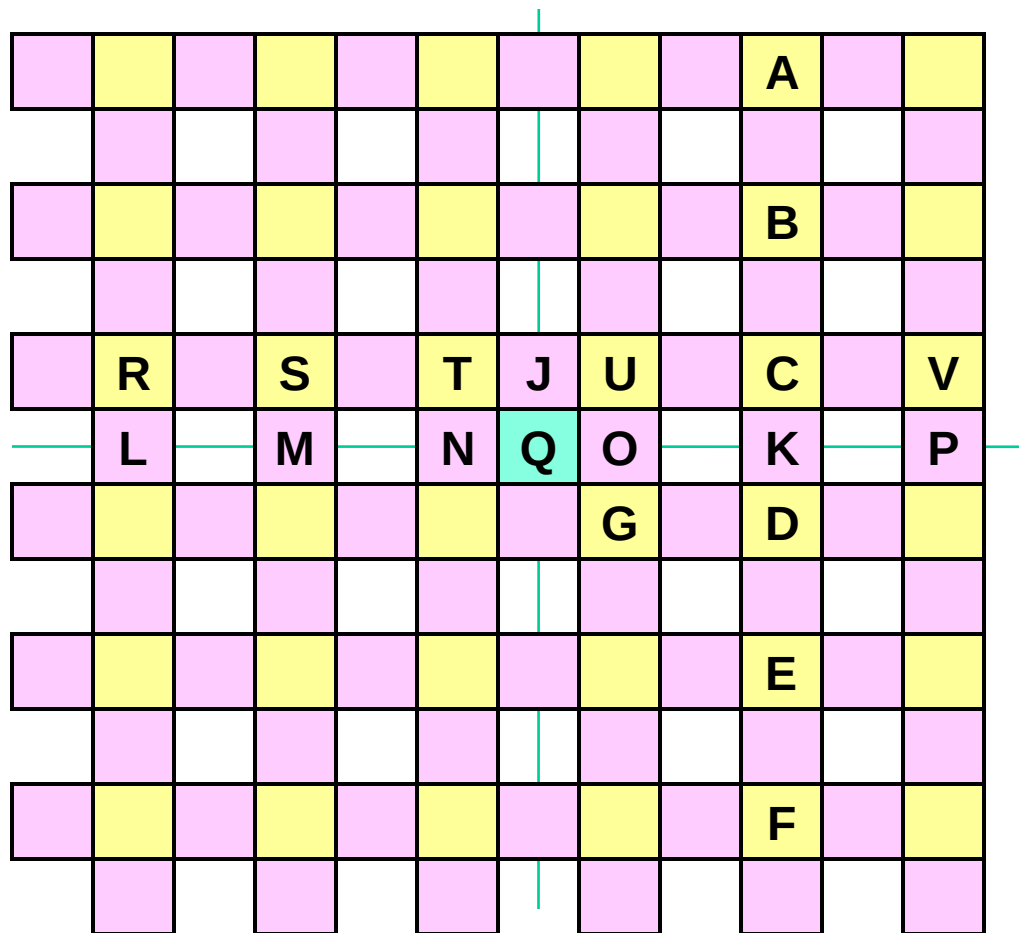
SAE = Sum of Absolute Error
(soma tória do m3dulo dos erros de reconstru33o)



1ª. Etapa (1/2 pixel):
Interpolação de Pixels
adjacentes a pontos de
amostragem, usando
filtro FIR de 6 coeficientes
na horizontal e vertical

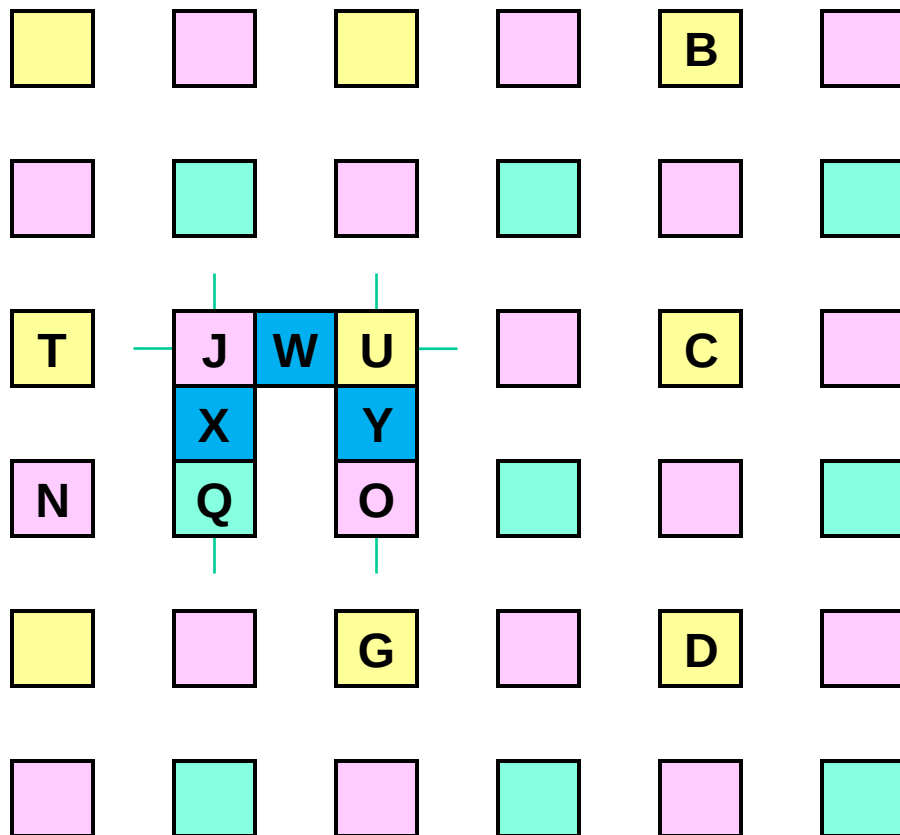
$$J = (R - 5S + 20T + 20U - 5C + V)/32$$

$$K = (A - 5B + 20C + 20D - 5E + F)/32$$



2ª. Etapa (1/2 pixel):
Interpolação dos Pixels
adjacentes aos pontos já
interpolados, usando o
mesmo filtro FIR

$$Q = (L - 5M + 20N + 20O - 5K + P) / 32$$

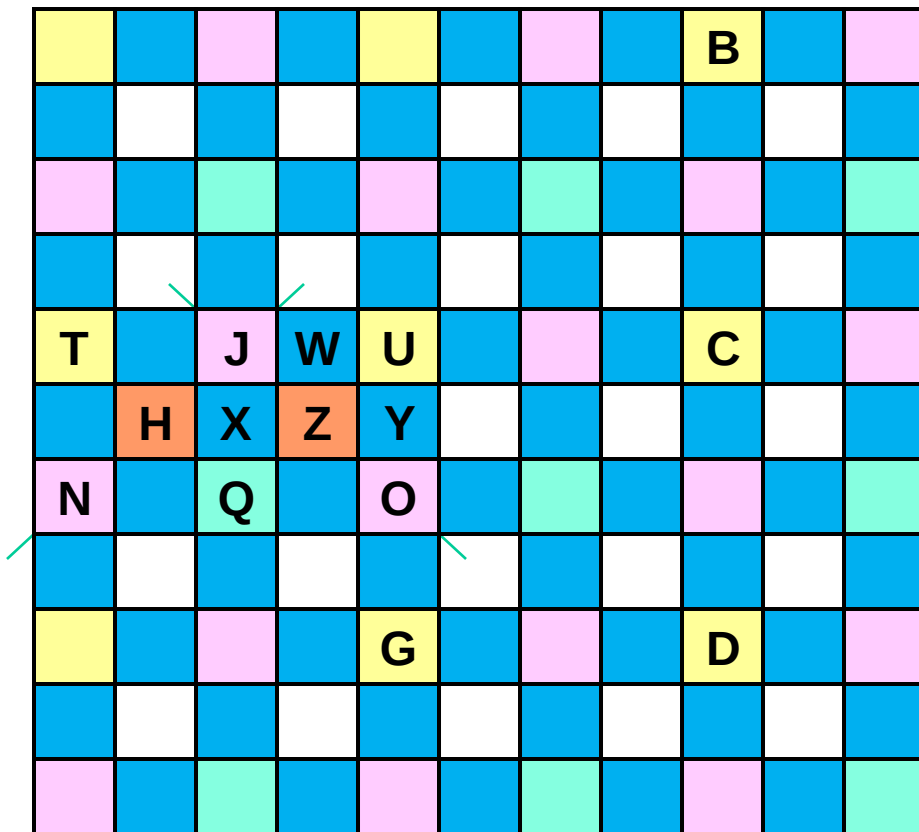


3ª. Etapa (1/4 pixel):
Interpolação dos Pixels
adjacentes aos pontos
calculados, usando
interpolação linear (média)

$$W = (J + U) / 2$$

$$X = (J + Q) / 2$$

$$Y = (U + O) / 2$$



4ª. Etapa (1/4 pixel):
Interpolação dos Pixels
faltantes, usando
interpolação linear em
diagonal a partir de pontos
na grade de ½ pixel

$$H = (J + N) / 2$$

$$Z = (J + O) / 2$$

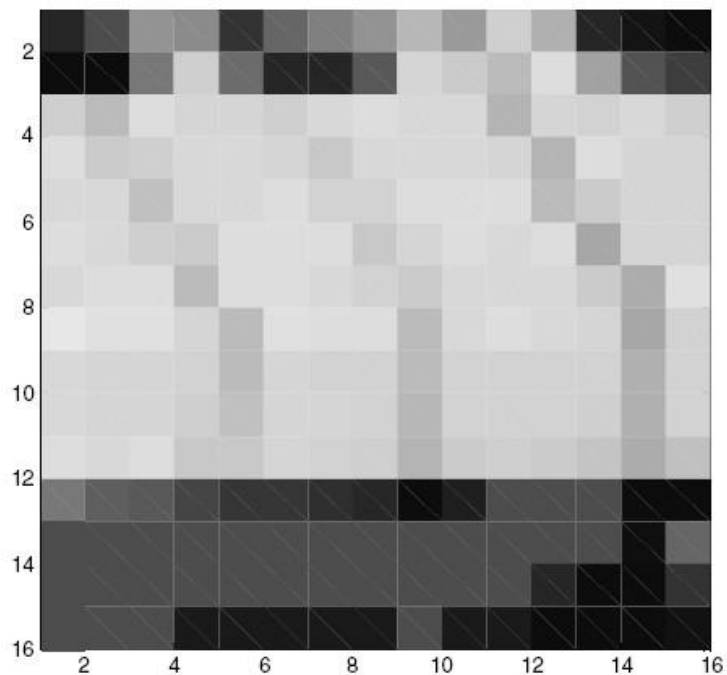


Figure 3.16 Close-up of reference region

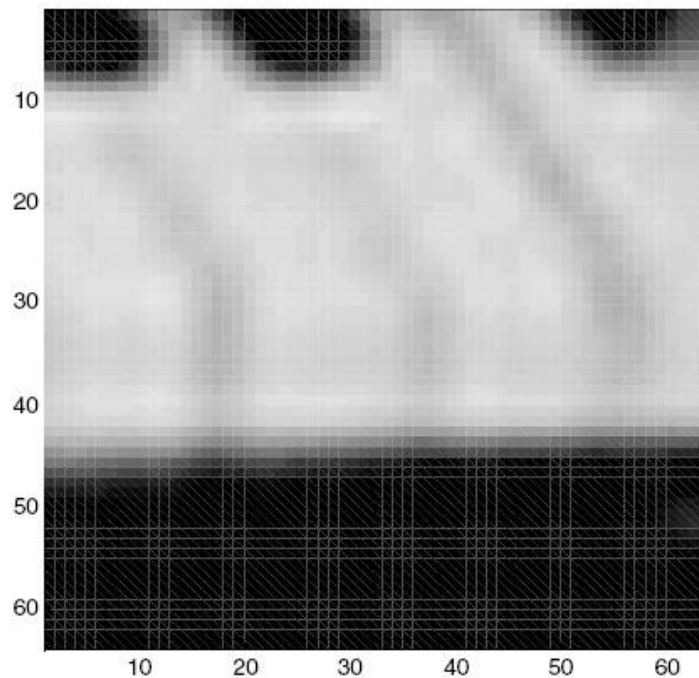
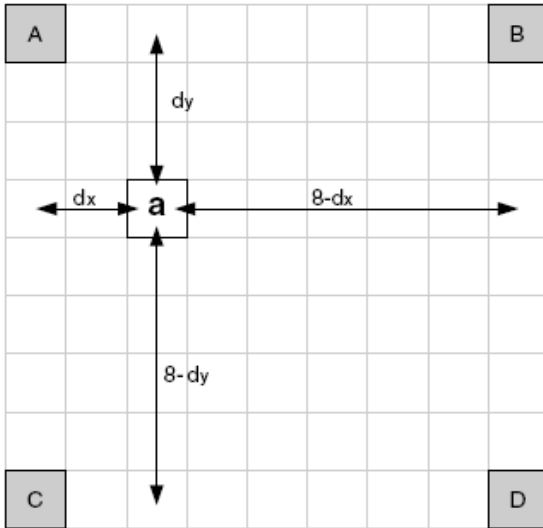


Figure 6.16 Luma region interpolated to quarter-pel positions



Já que a luminância é interpolada em $\frac{1}{4}$ de pixel, a crominância deve ser interpolada em $\frac{1}{8}$

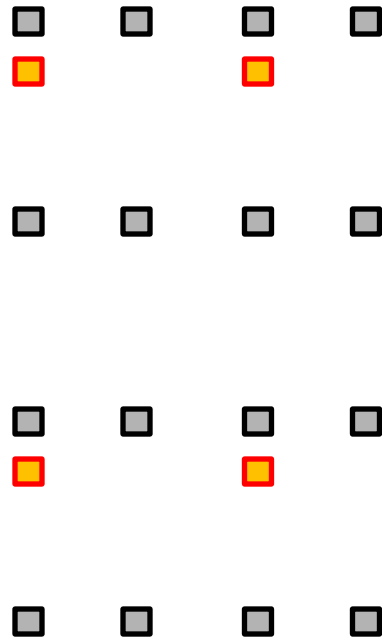
Interpolação bi-linear entre 4 amostras de crominância

$$a = \frac{1}{64} \left[(8-dx)(8-dy)A + dx(8-dy)B + (8-dx)dy \cdot C + dx \cdot dy \cdot D \right]$$

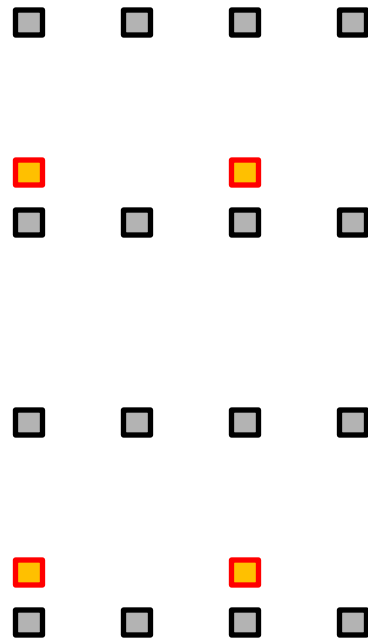


Gray Luminância

Yellow Crominância



Campo Top



Campo Bottom

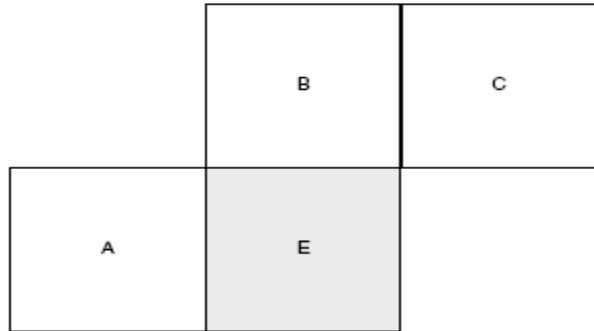
■ Luminância

■ Crominância

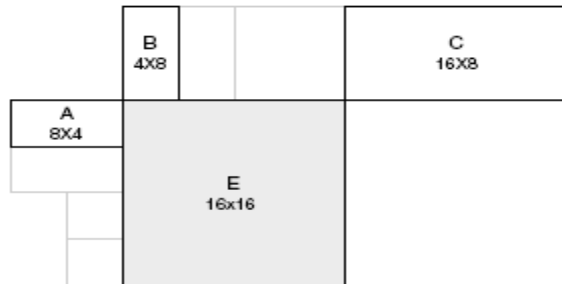
- Para cada macrobloco decodificado no receptor, o vetor de movimento é obtido pela soma de um vetor gerado por um preditor (MVp) com um erro de predição (MVd), que é transmitido junto com o macrobloco

$$MV = MVp + MVd$$

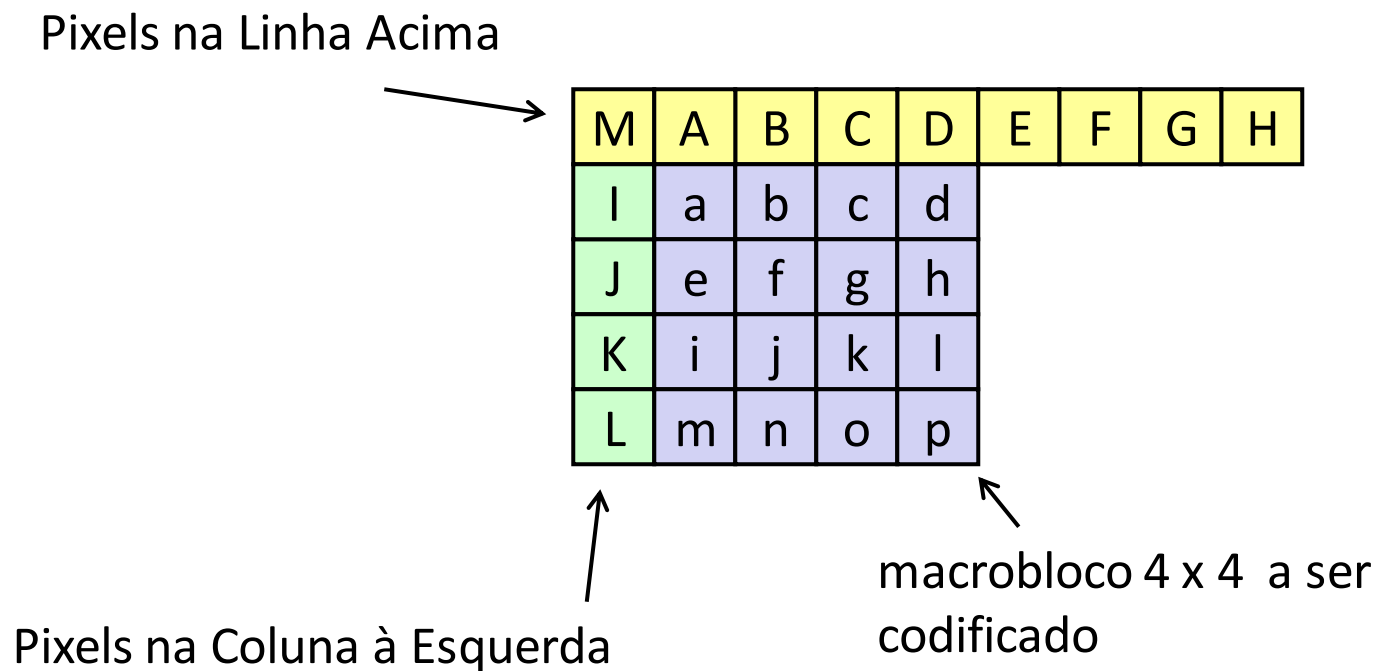
- A geração do vetor previsto (MVp) depende das disponibilidades e dimensões dos macroblocos superiores e à esquerda do atual.

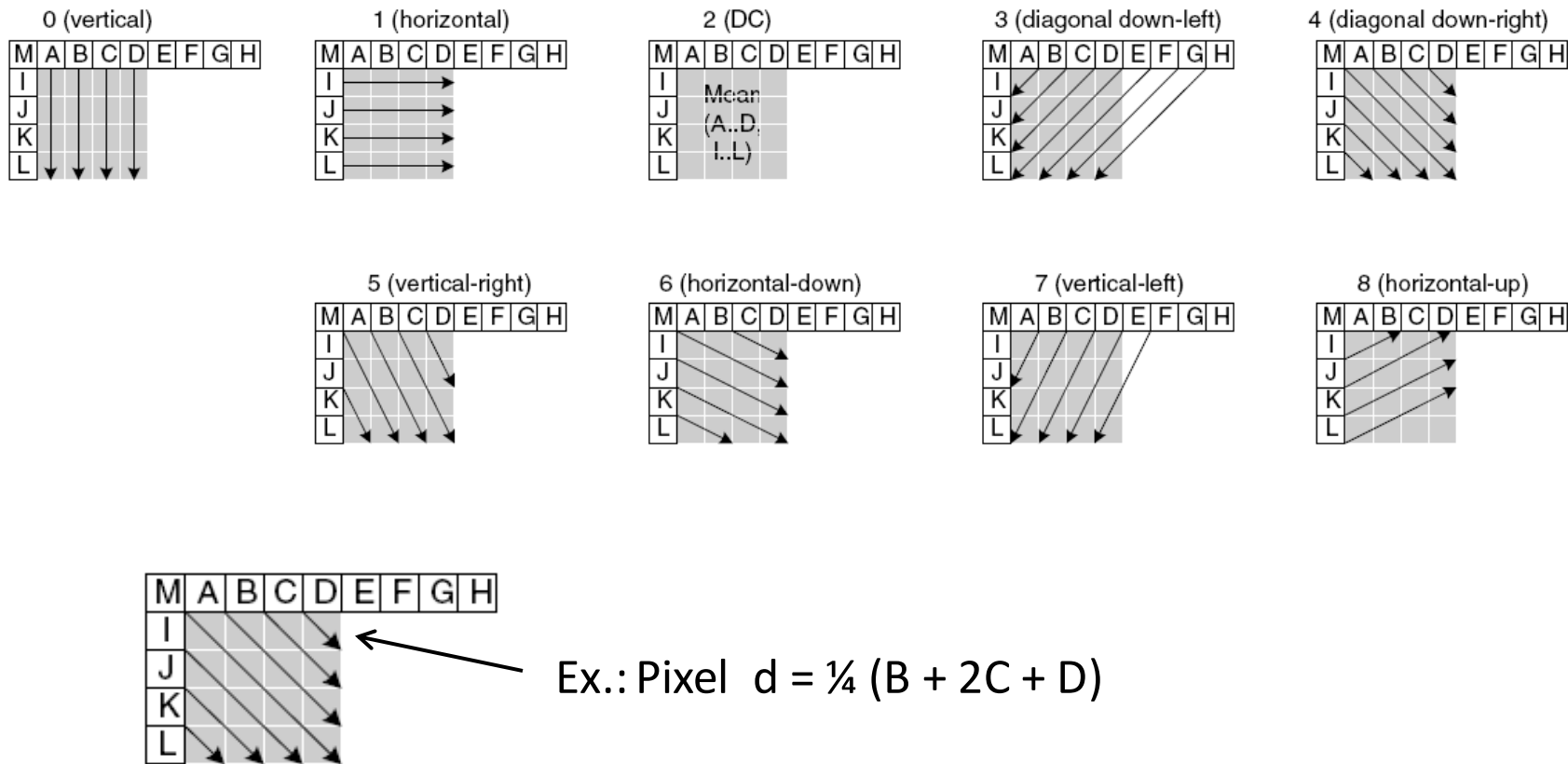


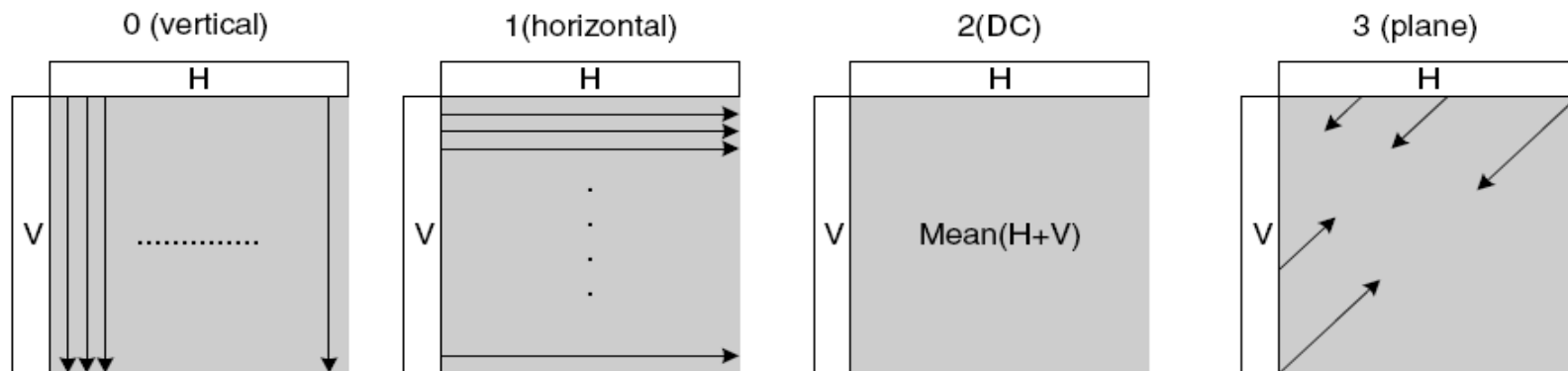
$$MV(E) = \text{Mediana } (MV(B), MV(C), MV(A))$$

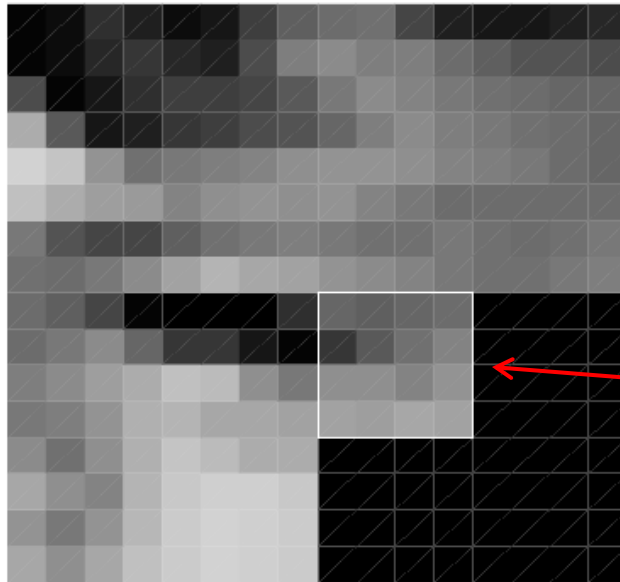


- Preditor em imagens tipo I:
 - Preditor espacial linear, extrapola pixels de um bloco a partir dos pixels vizinhos (à esquerda e acima)
 - Blocos de Luminância 4 x 4 admitem 9 tipos de predição linear
 - Blocos de Luminância 16 x 16 admitem 4 tipos
 - Blocos de Crominância admitem 4 tipos de preditor







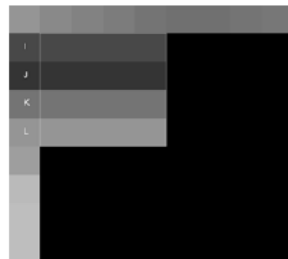


macrobloco 4 x 4 a ser
codificado

Exemplo de Predição Intra 4 x 4



0 (vertical), SAE = 317



1 (horizontal), SAE = 401



2 (DC), SAE = 317



3 (diag down/left), SAE = 350



4 (diag down/right), SAE = 466



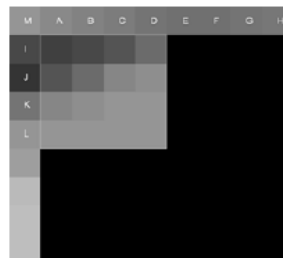
5 (vertical/right), SAE = 419



6 (horizontal/down), SAE = 530



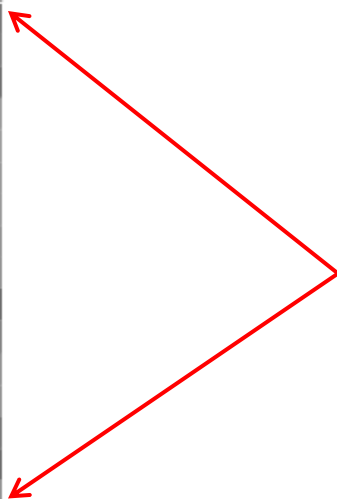
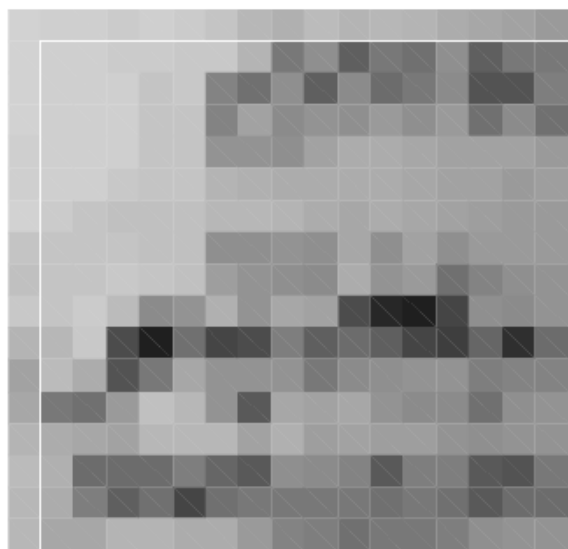
7 (vertical/left), SAE = 351



8 (horizontal/up), SAE = 203

menor erro de
predição



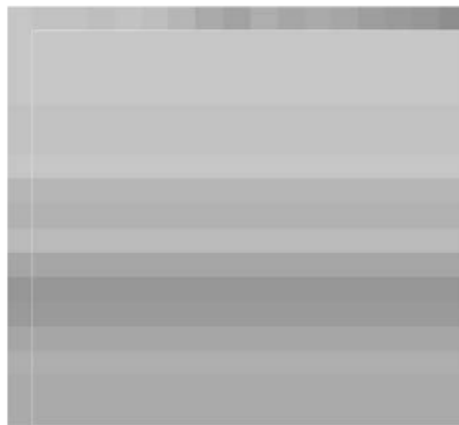


macrobloco 16x16 a ser
codificado

Exemplo de Predição Intra 16 x 16



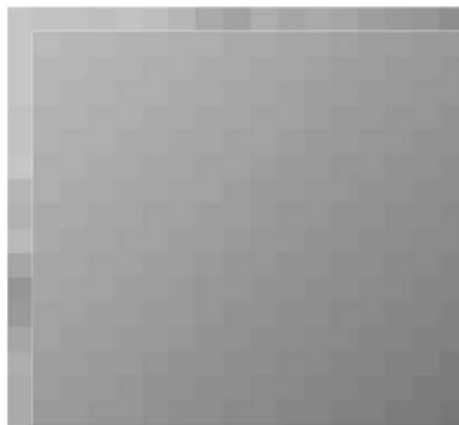
0 (vertical), SAE = 3985



1 (horizontal), SAE = 5097



2 (DC), SAE = 4991

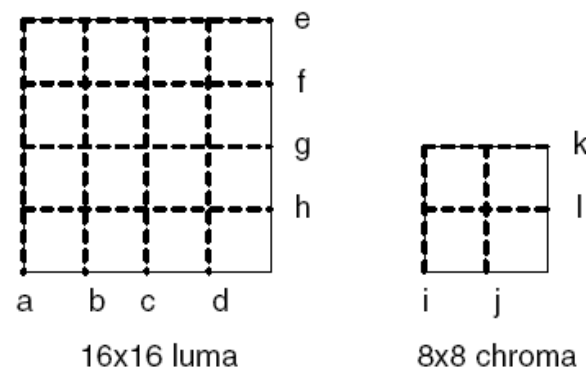


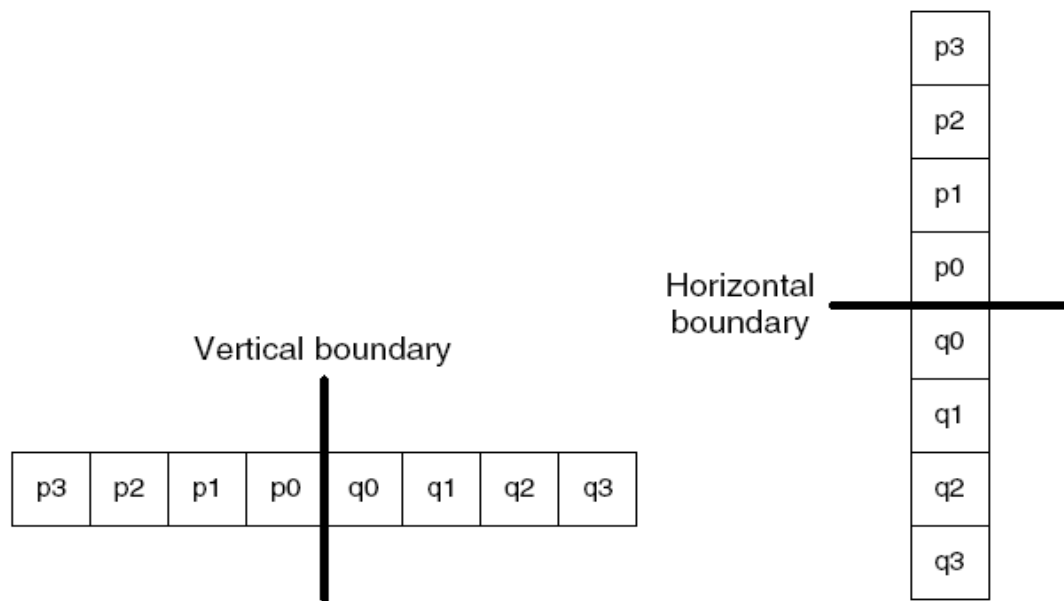
3 (plane), SAE = 2539

menor erro de
predição



- No codificador, é aplicado um filtro espacial para reduzir descontinuidades entre os blocos 4 x 4, depois da reconstrução (transformada) e antes da imagem ser armazenada como possível referência
- No receptor, é aplicado antes da exibição e da armazenagem
- O filtro é ligado ou desligado sob comando do codificador





Exemplos:

Filtro tipo 1: $p_0' = f(p_1, p_0, q_0)$

Filtro tipo 2: $p_2' = f_1(p_3, p_2, p_1, p_0, q_0)$

$p_1' = f_2(p_2, p_1, p_0, q_0)$

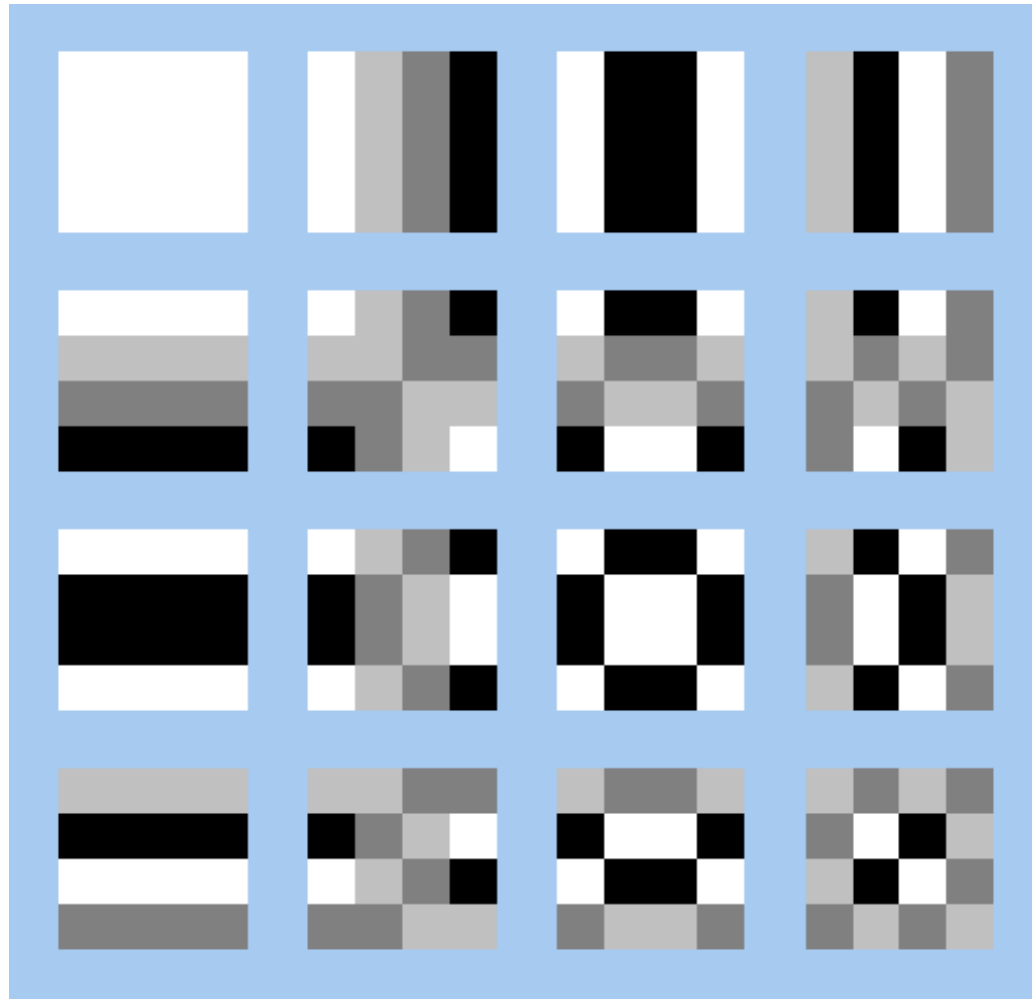
$p_0' = f_3(p_2, p_1, p_0, q_0, q_1)$

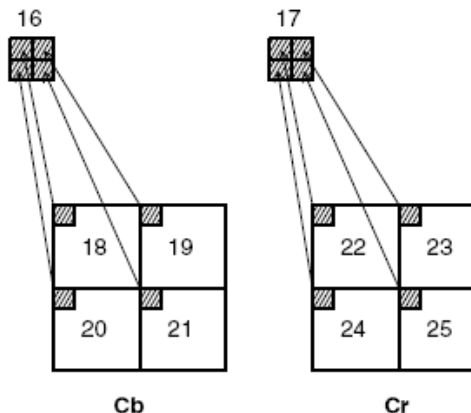
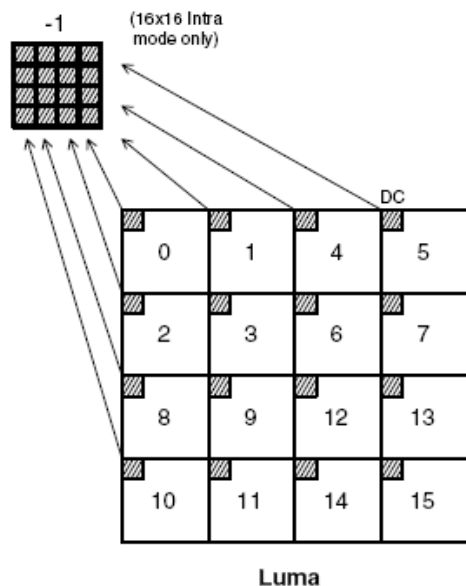


- Blocos 4 x 4 de Luminância: Transformada Inteira, similar à DCT

$$Y = C \cdot X \cdot C^T \otimes E$$

$$C = \begin{bmatrix} +1 & +1 & +1 & +1 \\ +2 & +1 & -1 & -2 \\ +1 & -1 & -1 & +1 \\ +1 & -2 & +2 & -1 \end{bmatrix} \quad E = \begin{bmatrix} \frac{1}{4} & \frac{1}{\sqrt{40}} & \frac{1}{4} & \frac{1}{\sqrt{40}} \\ \frac{1}{\sqrt{40}} & \frac{1}{10} & \frac{1}{\sqrt{40}} & \frac{1}{10} \\ \frac{1}{4} & \frac{1}{\sqrt{40}} & \frac{1}{4} & \frac{1}{\sqrt{40}} \\ \frac{1}{\sqrt{40}} & \frac{1}{10} & \frac{1}{\sqrt{40}} & \frac{1}{10} \end{bmatrix}$$

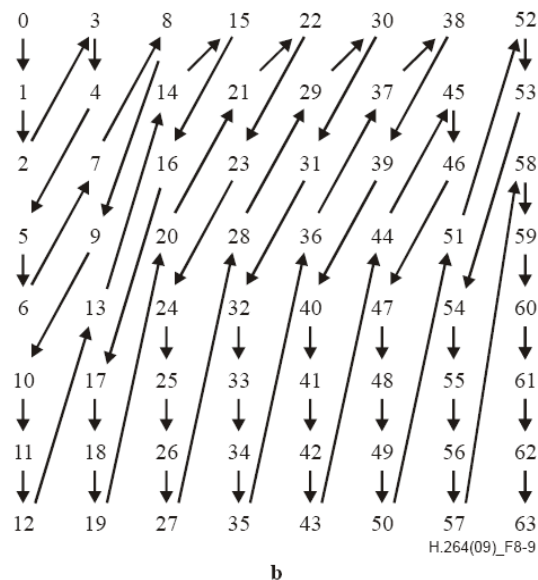
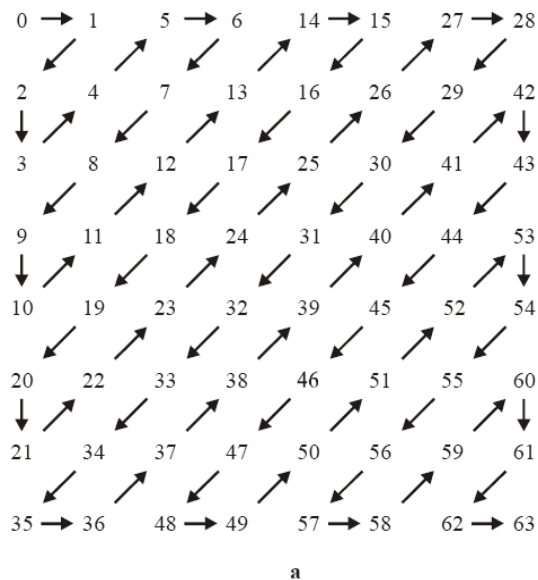
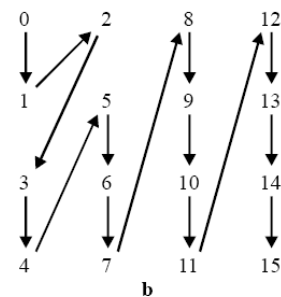
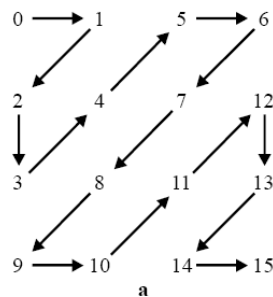




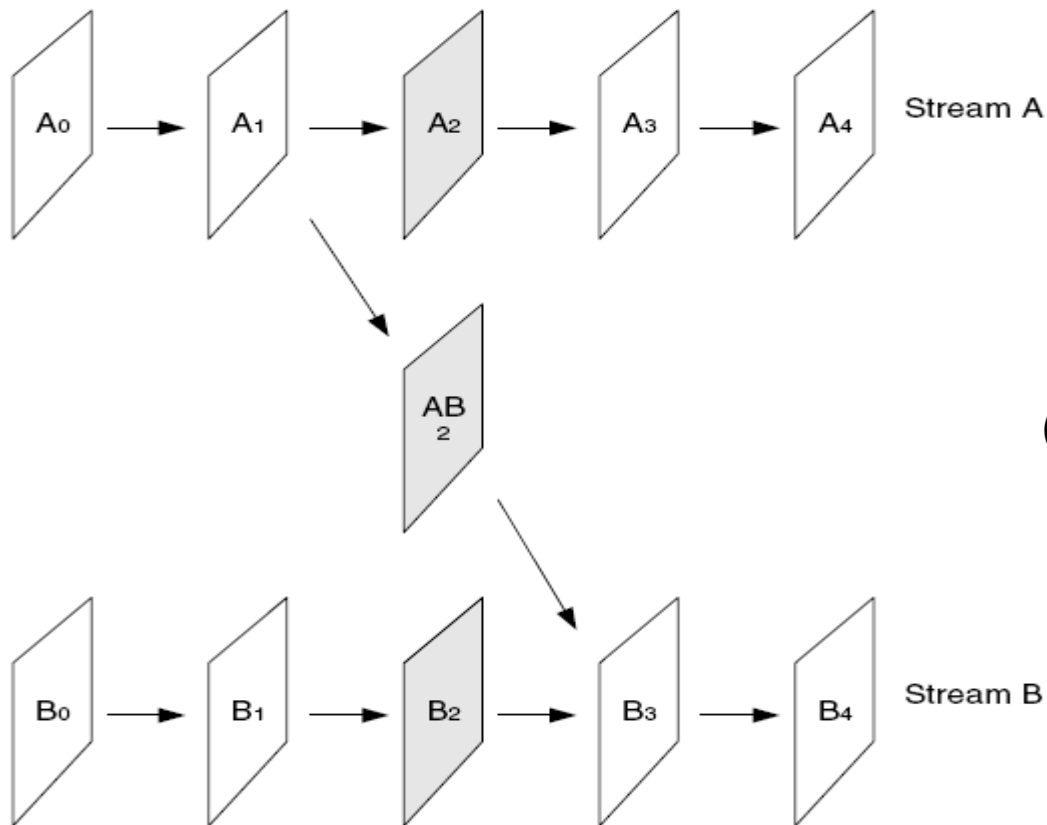
$$\mathbf{W}_{QD} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{W}_D \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

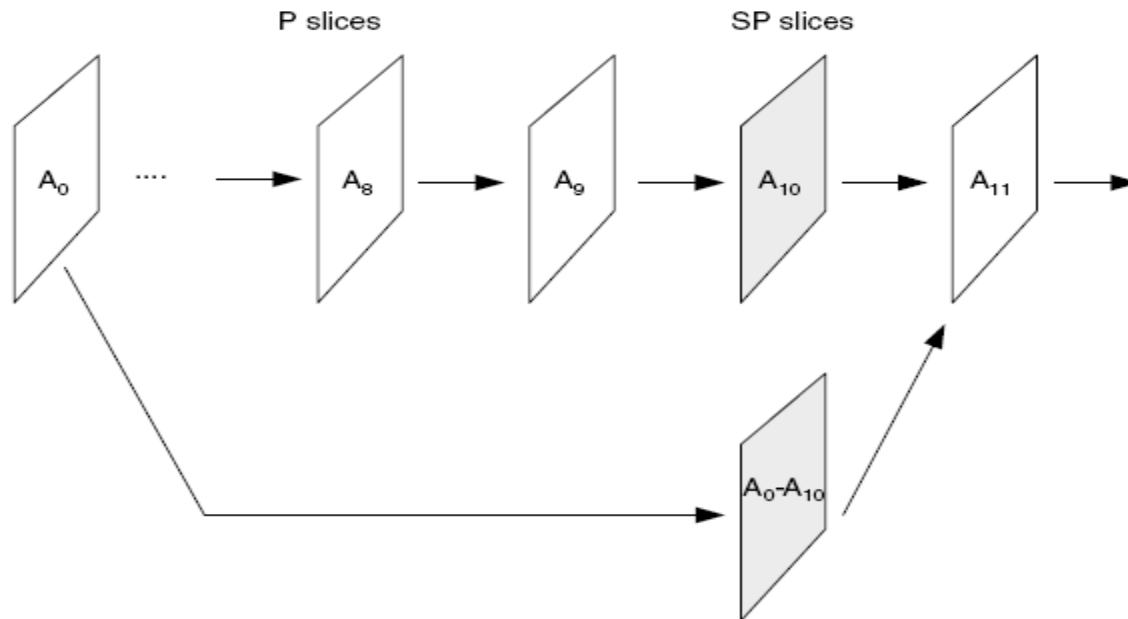
$$\mathbf{Y}_D = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{W}_D \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \right) / 2$$

Blocos 4 x 4



Blocos 8 x 8





(Perfil “Extended”)

- Codificação Entrópica:
 - CAVLC (*Context Adaptive Variable Length Coder*): Tabelas VLC adaptam-se às características da imagem de acordo com os símbolos já transmitidos
 - CABAC (*Context Adaptive Binary Arithmetic Coder*): Baseia-se em modelos para as probabilidades condicionais de ocorrência dos símbolos

- Os dados de uma Fatia são subdivididos em 3 partições:
- A = Cabeçalhos da Fatia e dos Macroblocos
- B = Coeficientes codificados para os macroblocos tipo “intra” da fatia
- C = Coeficientes para os macroblocos tipo “inter” (P, B)
- Cada partição é transmitida em unidades separadas
- A partição A pode ter prioridade maior

Comparações entre MPEG e H.264



MPEG



H.264 (sem filtro anti-blocagem)

Distorções de Reconstrução



Figure 6.33 Reconstructed, $QP = 36$ (no filter)



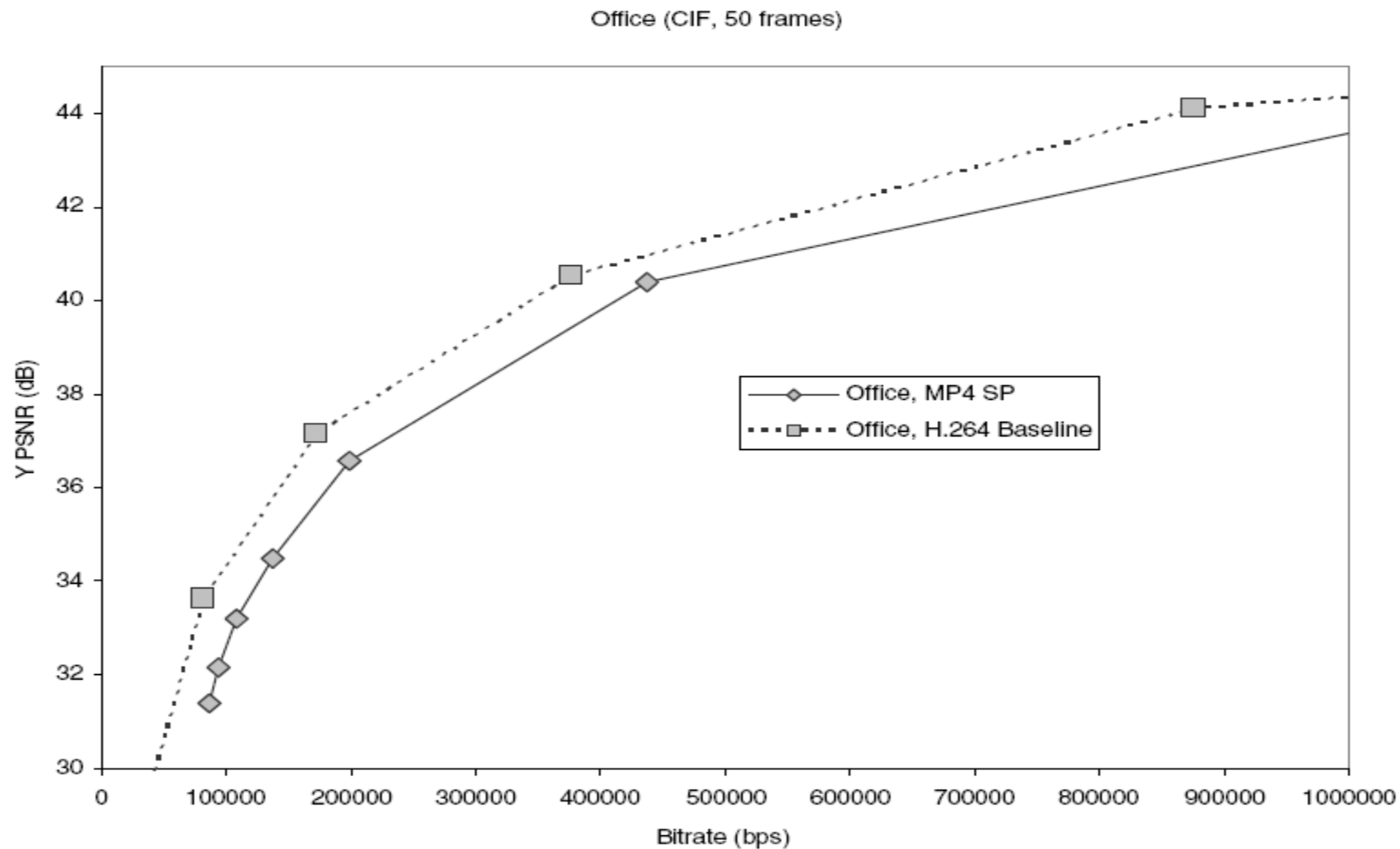
Figure 6.34 Reconstructed, $QP = 36$ (with filter)

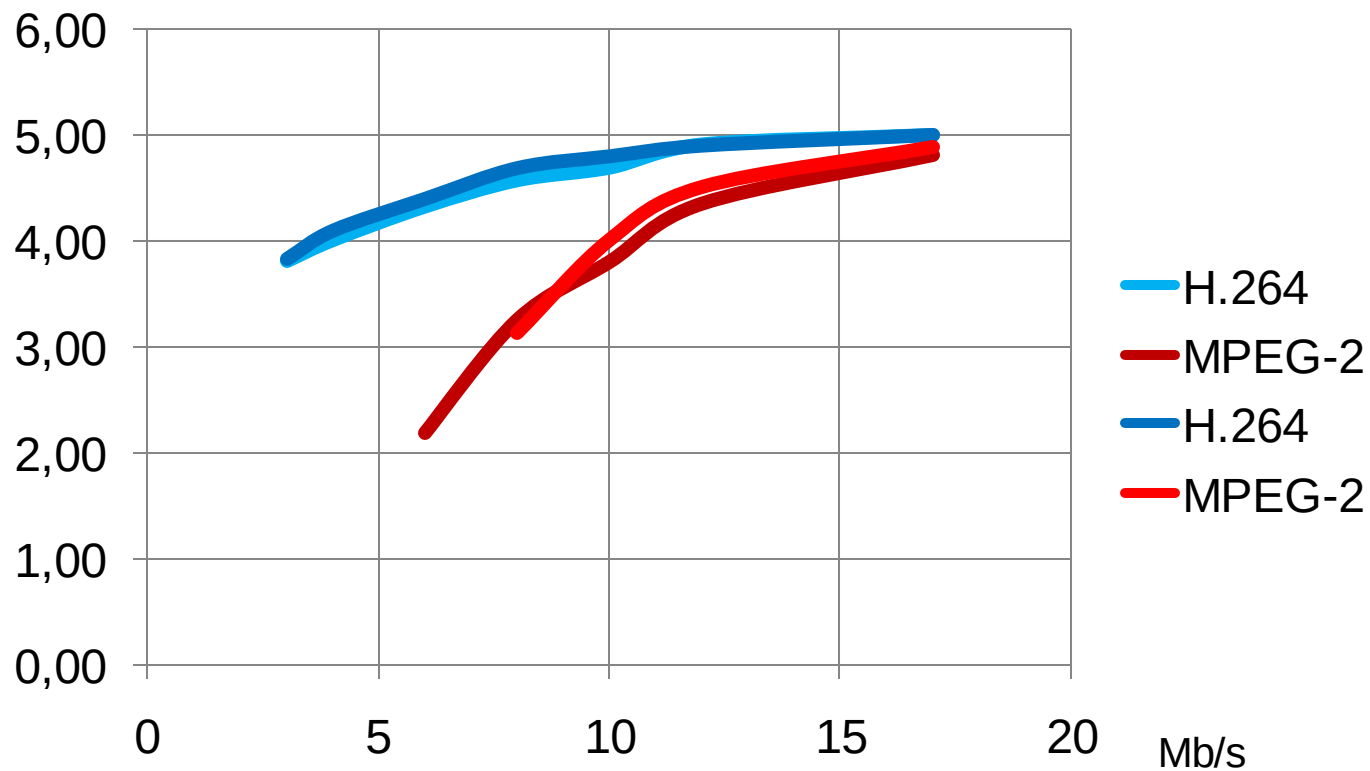












Avaliação subjetiva de sequencias de vídeo HDTV

Codificador HEVC / H.265

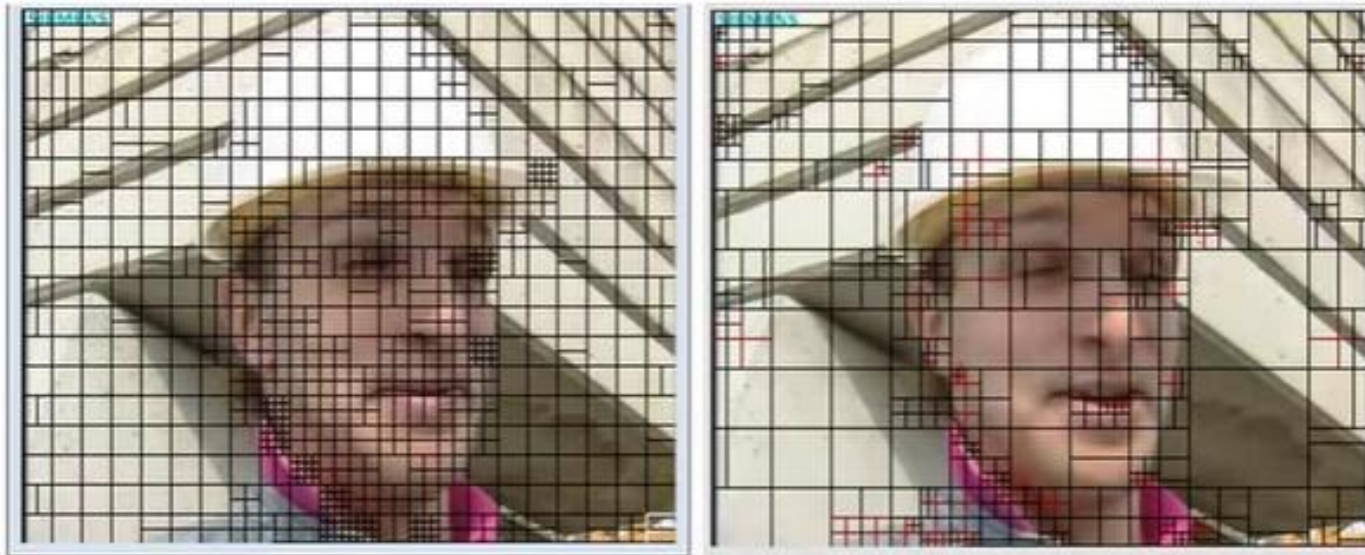
- ISO/IEC MPEG + ITU-T VCEG

(Moving Pictures Experts Group e Video Coding Experts Group)

- ISO/IEC: 23008-2 MPEG-H Parte 2
- ITU-T: H.265
- Finalizado em 25/01/2013
- Ratificado em 13/04/2013
- Metas: redução para 50% da taxa de bits em relação a H.264; suporte a UHDV (8k)

- Recursos:
 - “Coding Tree Unit”: 64 x 64, 32 x 32, 16 x 16, 8 x 8, 4 x 4
 - Vários espaços de Cromaticidade: NTSC, PAL, Rec.601, Rec.709, Rec.2020, RGB, etc.
 - Focado em vídeo sem entrelaçamento
 - Transformadas inteiras 4x4, 8x8, 16x16 e 32x32
 - Preditores Intra com mais direções de interpolação
 - Codificação CABAC apenas

- Partições 16 x 16 vs. 64 x 64



- Comparação com H.264:

