Distributed Video Coding (DVC)

Frederic Dufaux
LTCI - UMR 5141 - CNRS
TELECOM ParisTech
frederic.dufaux@telecom-paristech.fr

Montpellier, April 30, 2013

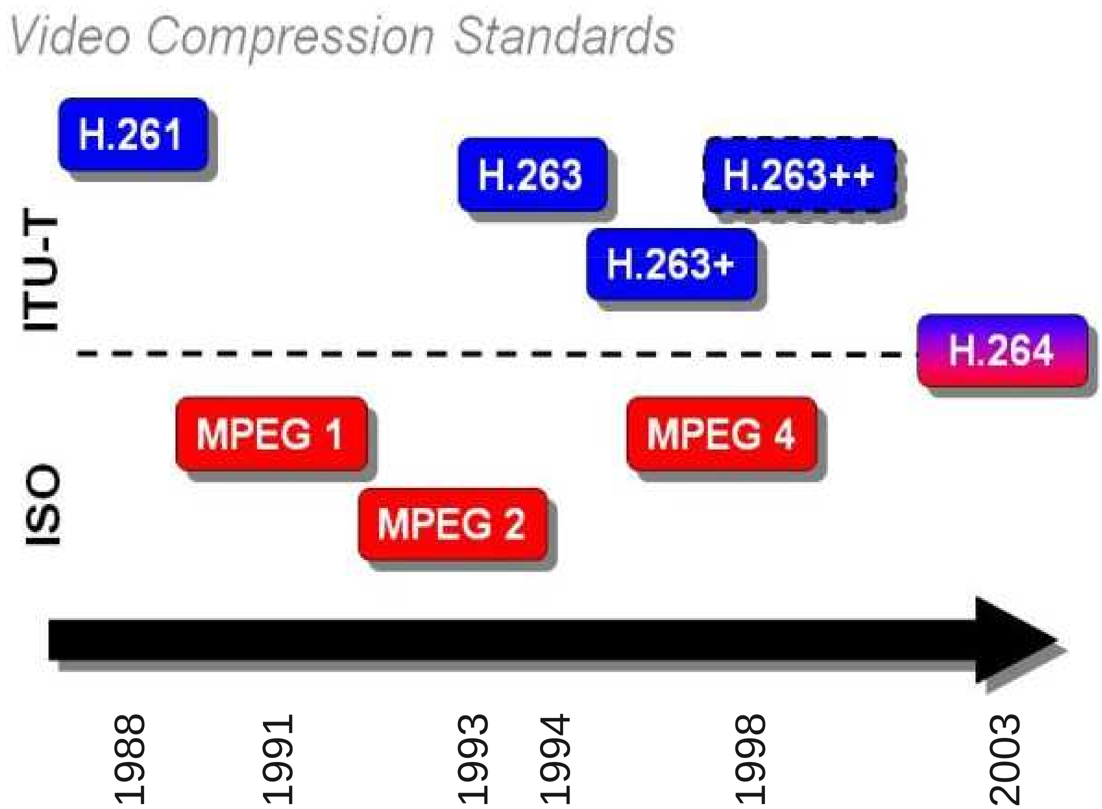
Outline

- Context and background
- Theoretical foundations
- Distributed Video Coding (DVC)
- Multiview video coding
- Conclusions

Context and Background

Video Coding Standards

- **Compression efficiency**
  - Typically 50% gain every 5 years
  - Adding more efficient coding tools / modes to the familiar predictive video coding architecture
  - Functionalities such as scalability, error resilience, interactivity, low complexity, random access, …



Conventional Predictive Coding

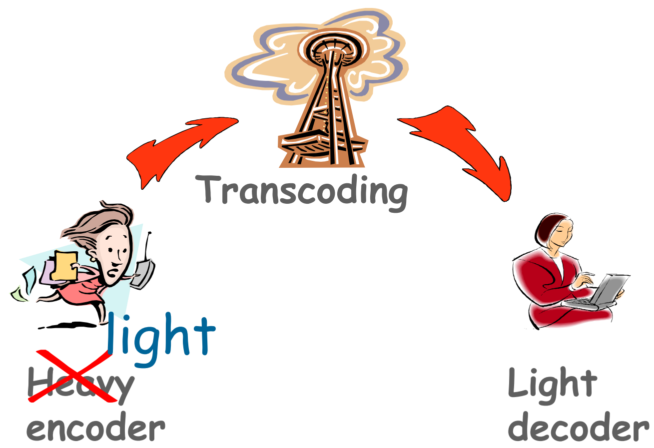
- Exploitation of the source correlation at the encoder

- High coding efficiency
- Rigid partition of complexity
  - High complexity encoder
  - Low complexity decoder
  - More appropriate for a broadcast model (downlink)
- Fragile in the presence of packet/frame losses
  - Drift due to prediction loop in encoder

New Class of Up-Link Applications

- High-resolution wireless digital video cameras
- Multimedia smartphones and PDA's
- Low-power video sensors and surveillance cameras

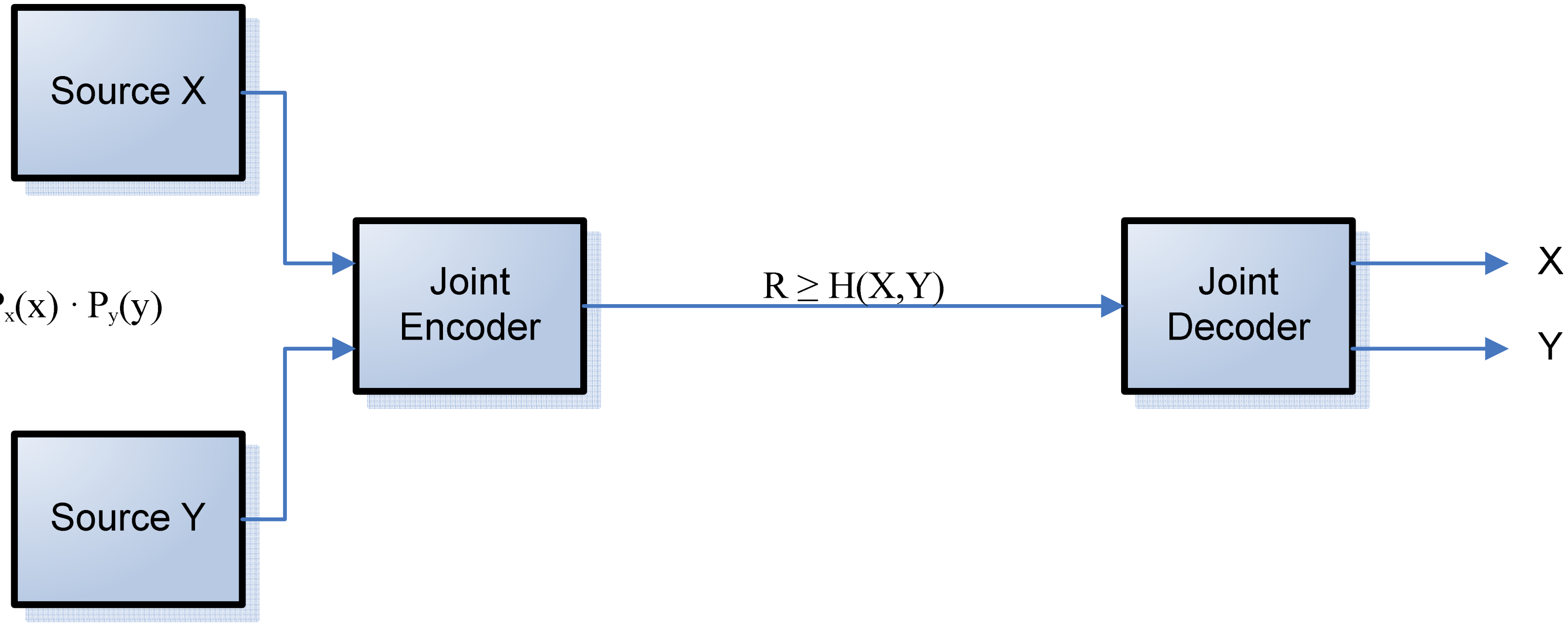
- Challenges
  - High coding efficiency
  - Flexible partition of complexity
    - *Low complexity encoder*
    - *High complexity decoder*
  - Robustness to packet/frame losses
  - Low latency



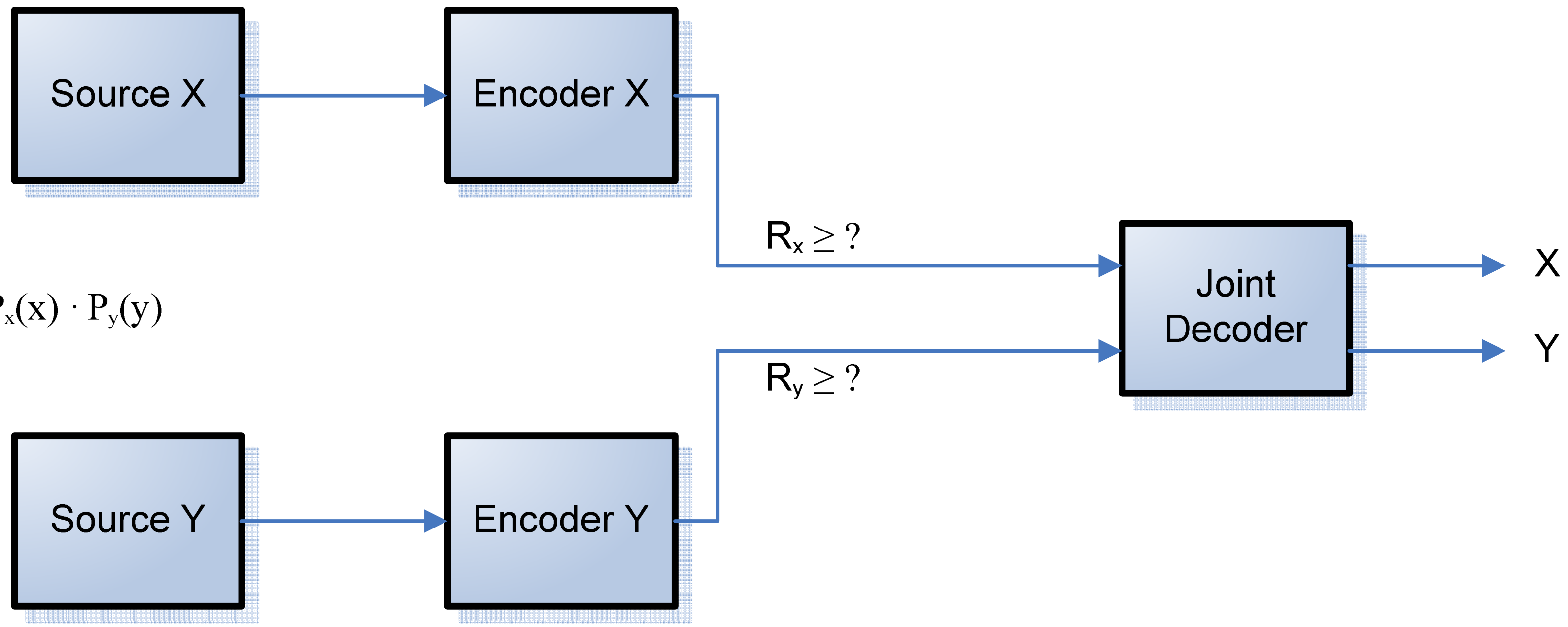
Theoretical Foundations

Coding of Dependent Sources

$P_{xy}(x,y) \neq P_x(x) \cdot P_y(y)$

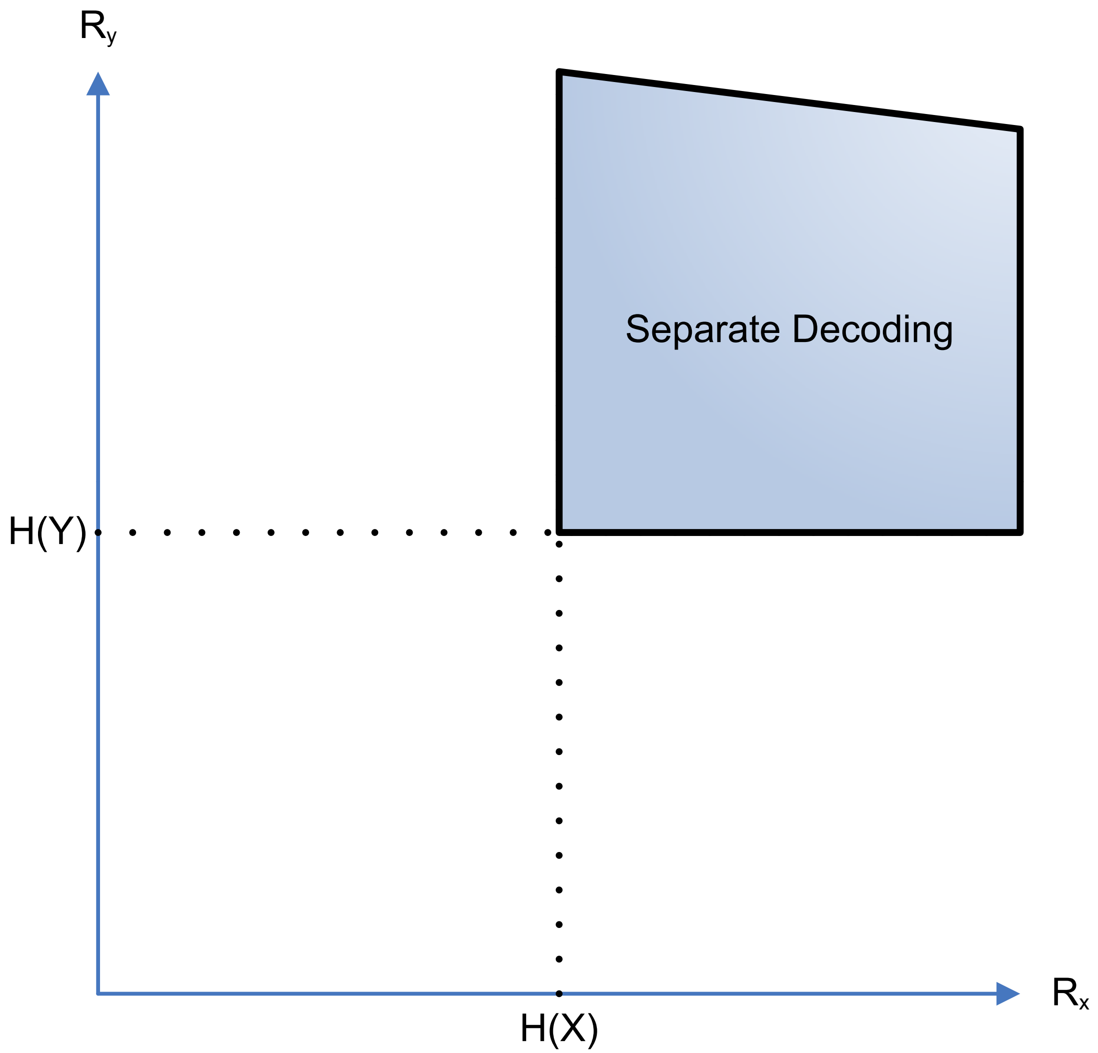


Distributed Coding of Dependent Sources

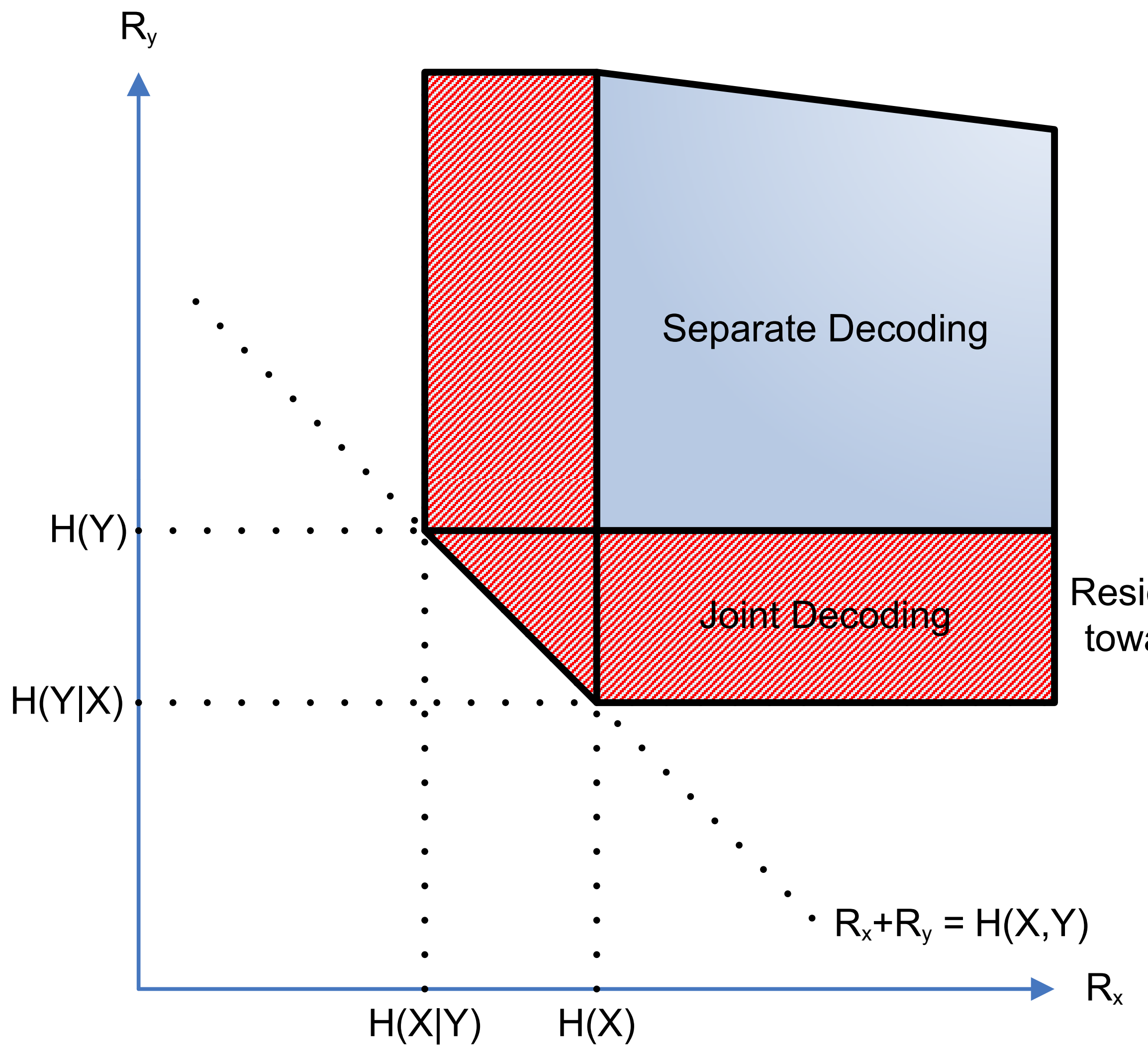


$P_{xy}(x,y) \neq P_x(x) \cdot P_y(y)$

Slepian-Wolf Theorem



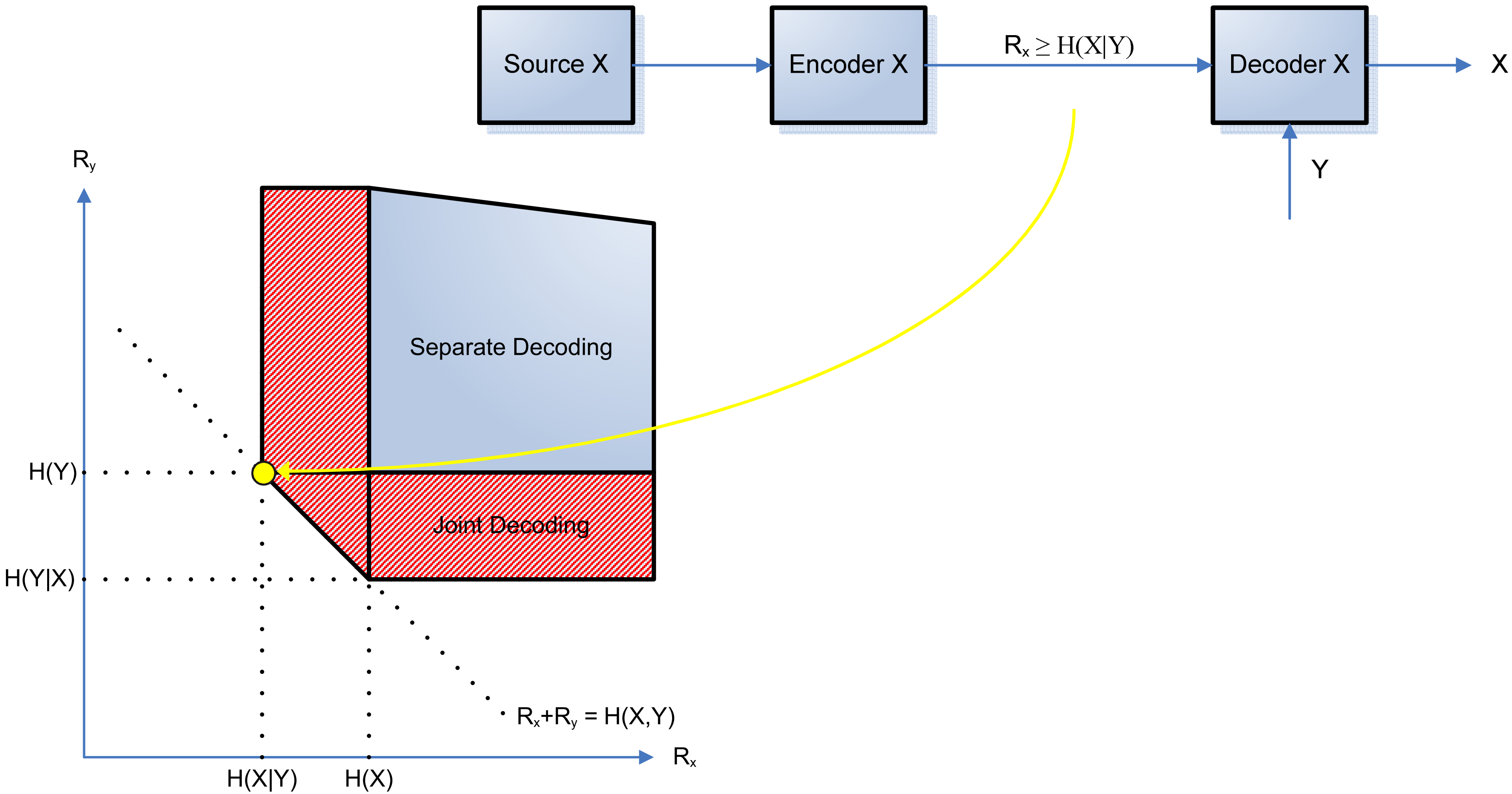
Slepian-Wolf Theorem



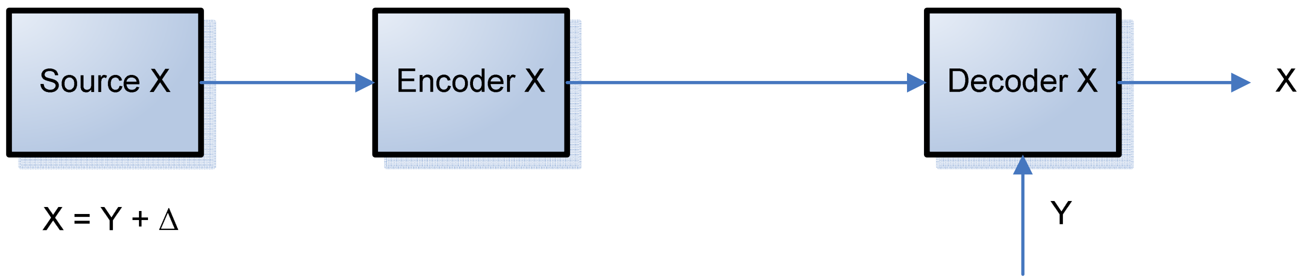
$R_x \geq H(X|Y)$
$R_y \geq H(Y|X)$
$R_x+R_y \geq H(X,Y)$

Residual error probability tends towards 0 for long sequences

Slepian-Wolf with Decoder Side Information



Slepian-Wolf with Decoder Side Information

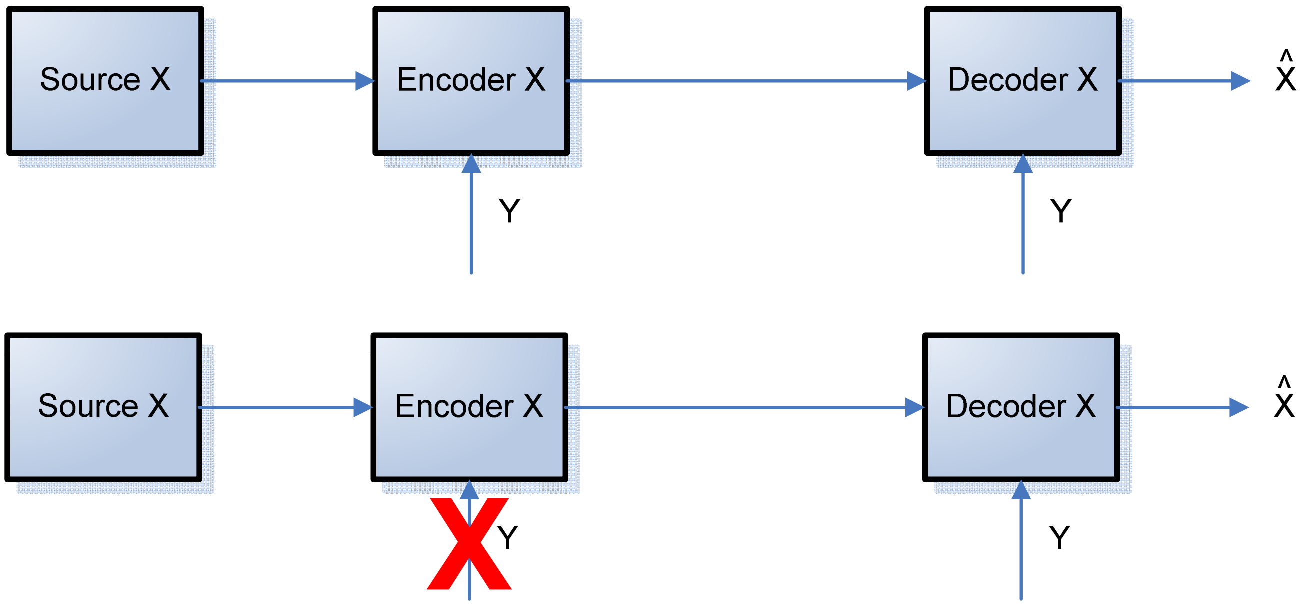


- Y is a guess of X
  - Better guess results in better coding efficiency

- Y is a noisy version of X with channel errors Δ
  - Encoder generates parity bits to protect against channel errors
  - Decoder performs error-correcting decoding

Wyner-Ziv Theorem

- Extension to lossy coding



- No rate-distortion performance loss
  - Gaussian statistics and MSE distortion
  - Later on: only innovation X-Y needs to be Gaussian

Opportunities

- Opportunity to re-invent video coding
  - Forget the past deterministic approach
  - Adopt a new statistical mind set

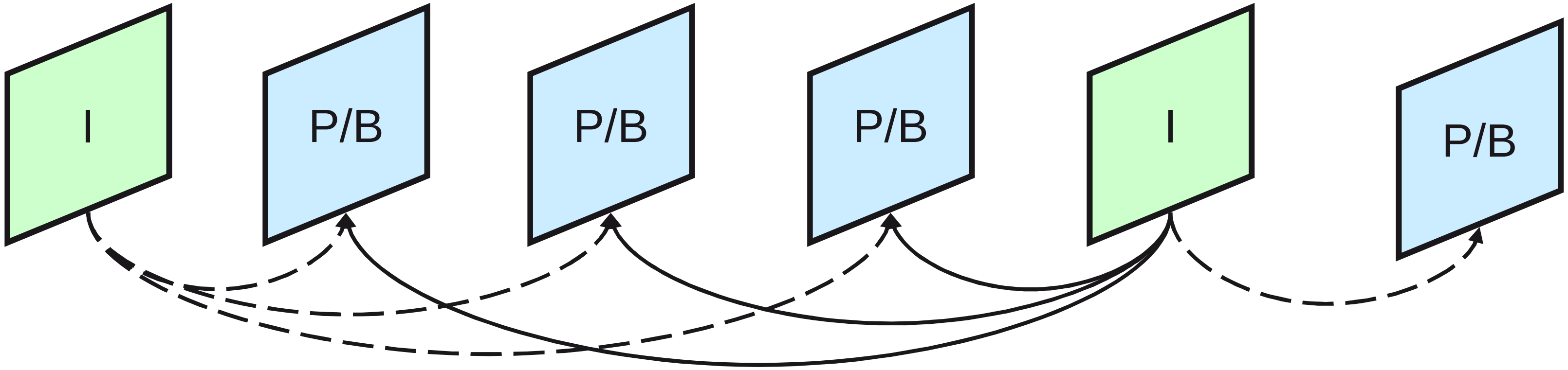
- Flexible complexity partition
- Intrinsic joint source-channel coding robust to errors
- Codec independent scalability
- Multiview coding exploiting correlation between views

- Challenge: achieve state-of-the-art coding performance

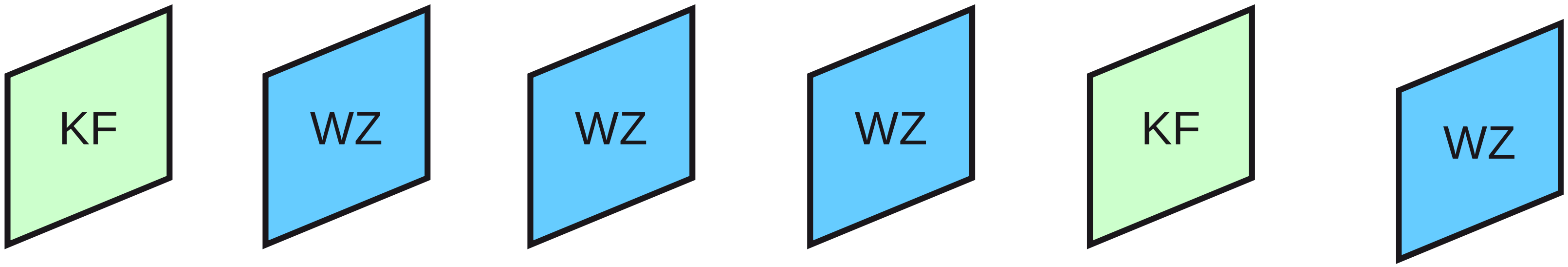
Distributed Video Coding (DVC)

Application of DVC to low complexity mono-view video

- Hybrid video coding



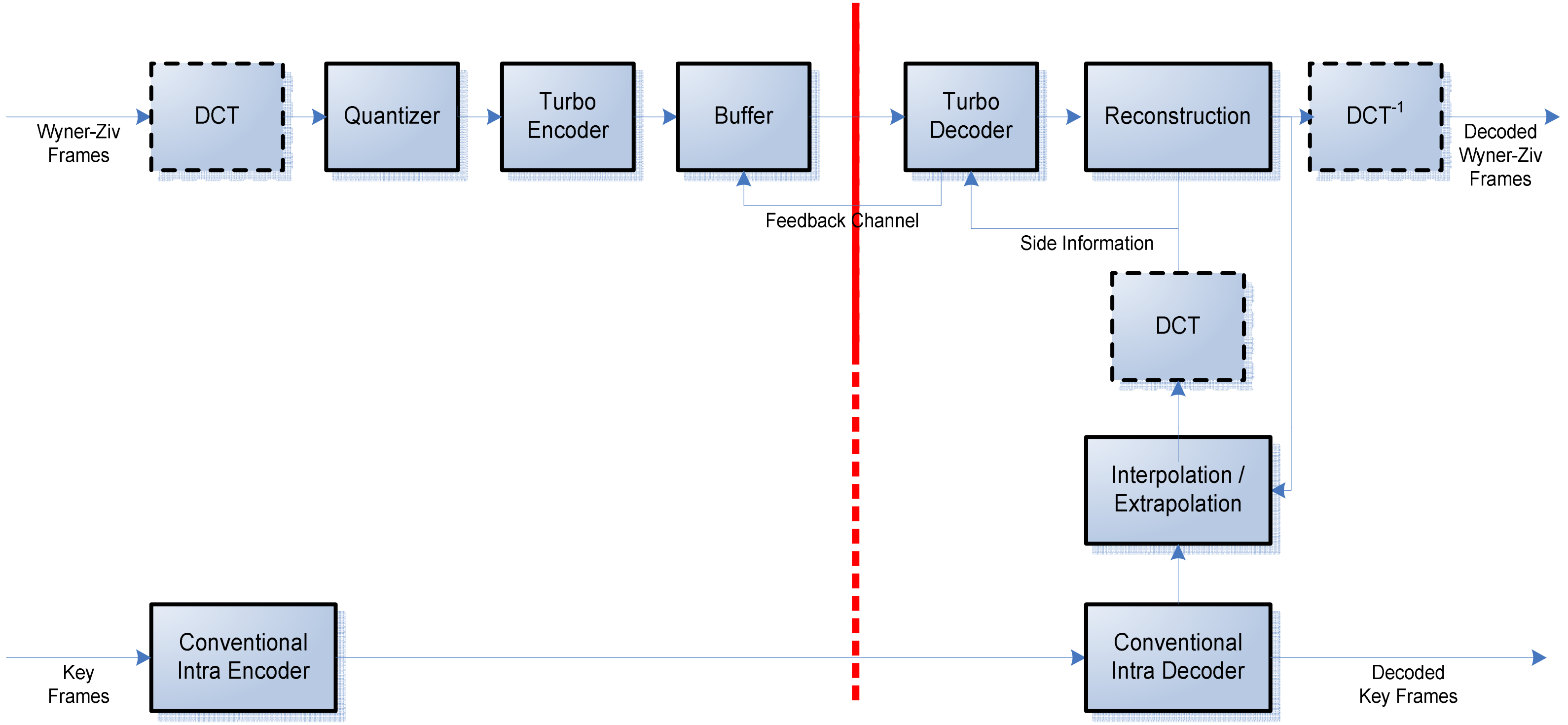
- Distributed video coding



Application of DVC to low complexity mono-view video

- **Key frames are coded as Intra frames**
- **For WZ frame only parity bit are coded**
  - Pixel domain coding
  - Transform domain coding
  - No prediction! (KF are not supposed to be known)
- **Side information is needed to reconstruct WZFs**
  - SI amounts to an estimation of the current WZF, based on information available at the decoder
- **Orders of magnitudes simpler than INTRA (10 times) and INTER (100 times) coding**

Pixel-domain and Transform-domain Architectures



Pixel-domain and Transform-domain Architectures

- Sequences divided into Group of Pictures (GOP)
  - First frame of GOP is Intra coded (key frame)
  - Remaining frames encoded using distributed coding (WZ frames)
- Pixel-domain and transform-domain
- Quantized values split into bitplanes which are Turbo encoded
- Decoder
  - Motion compensated interpolation/extrapolation to generate SI
  - Parity bits of WZ frames requested via feedback channel
  - SI and parity bits using in the turbo decoder to reconstruct bitplanes

Image interpolation in DVC

- **Problem:**
  - Given images I_{k-1} and I_{k+1}, find the best estimation of image I_k
- **Typical Side Information generation problem**
- **Current solutions use block-matching motion estimation and compensation**
- **Looking for backward and forward motion vector fields**

Image interpolation in DVC: the DISCOVER algorithm

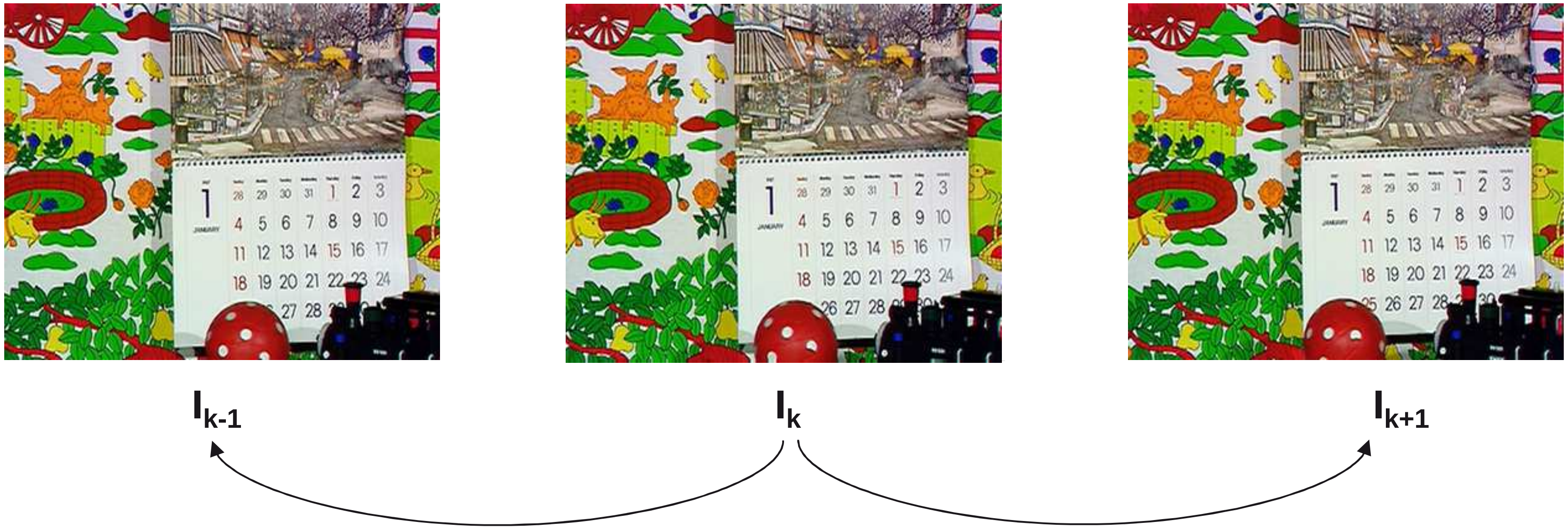
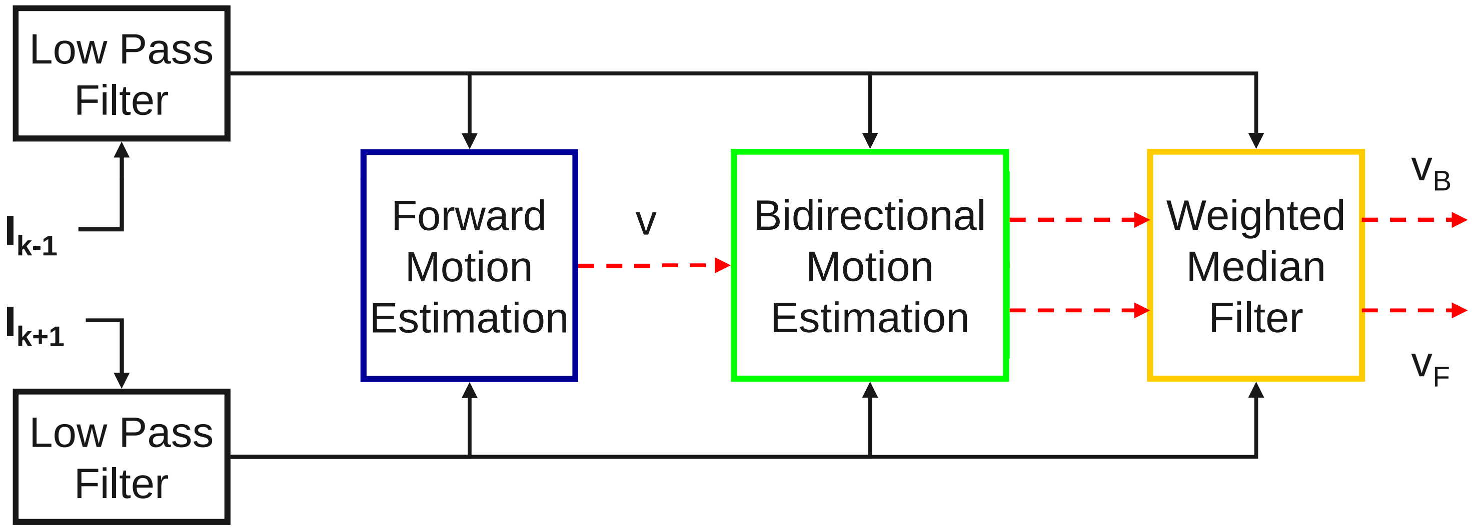
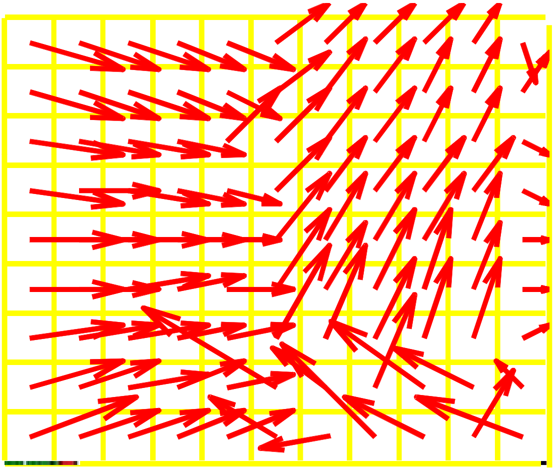


Image interpolation in DVC: the DISCOVER algorithm



The DISCOVER algorithm: Forward ME

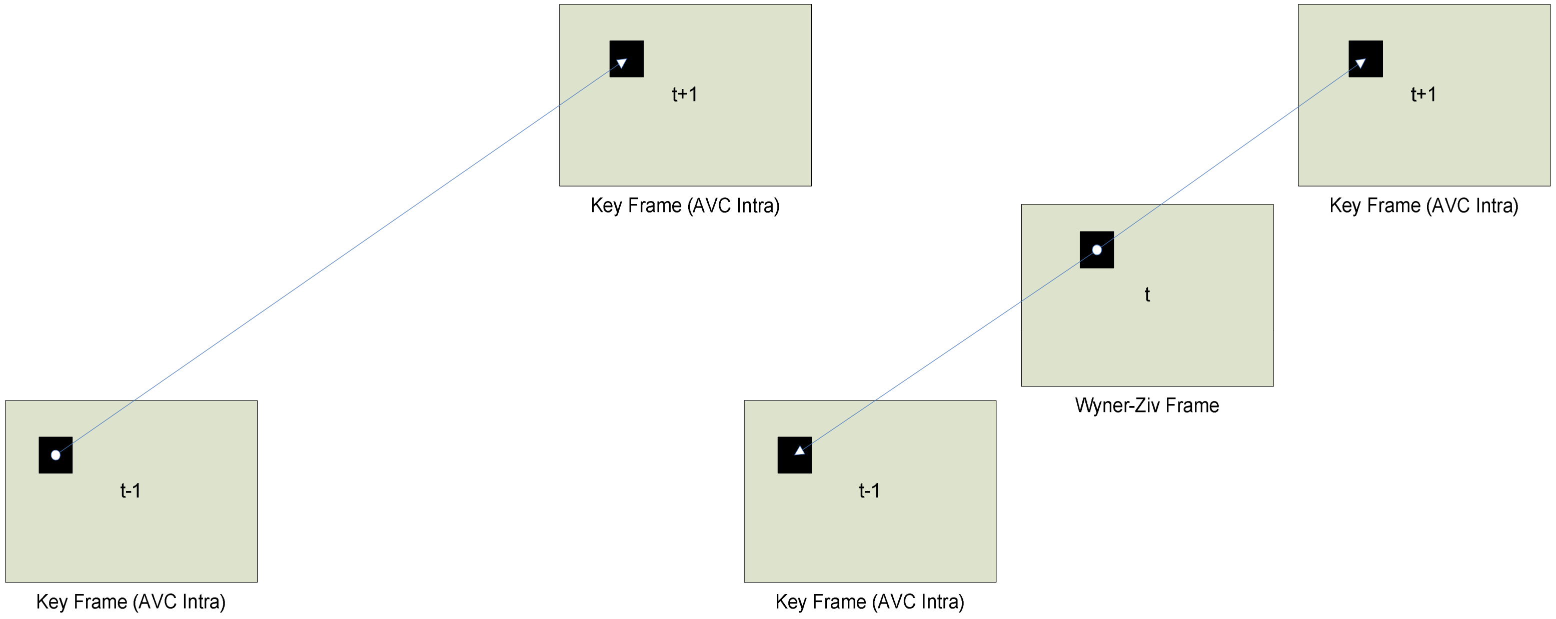
$\mathbf{I}_{k-1}$



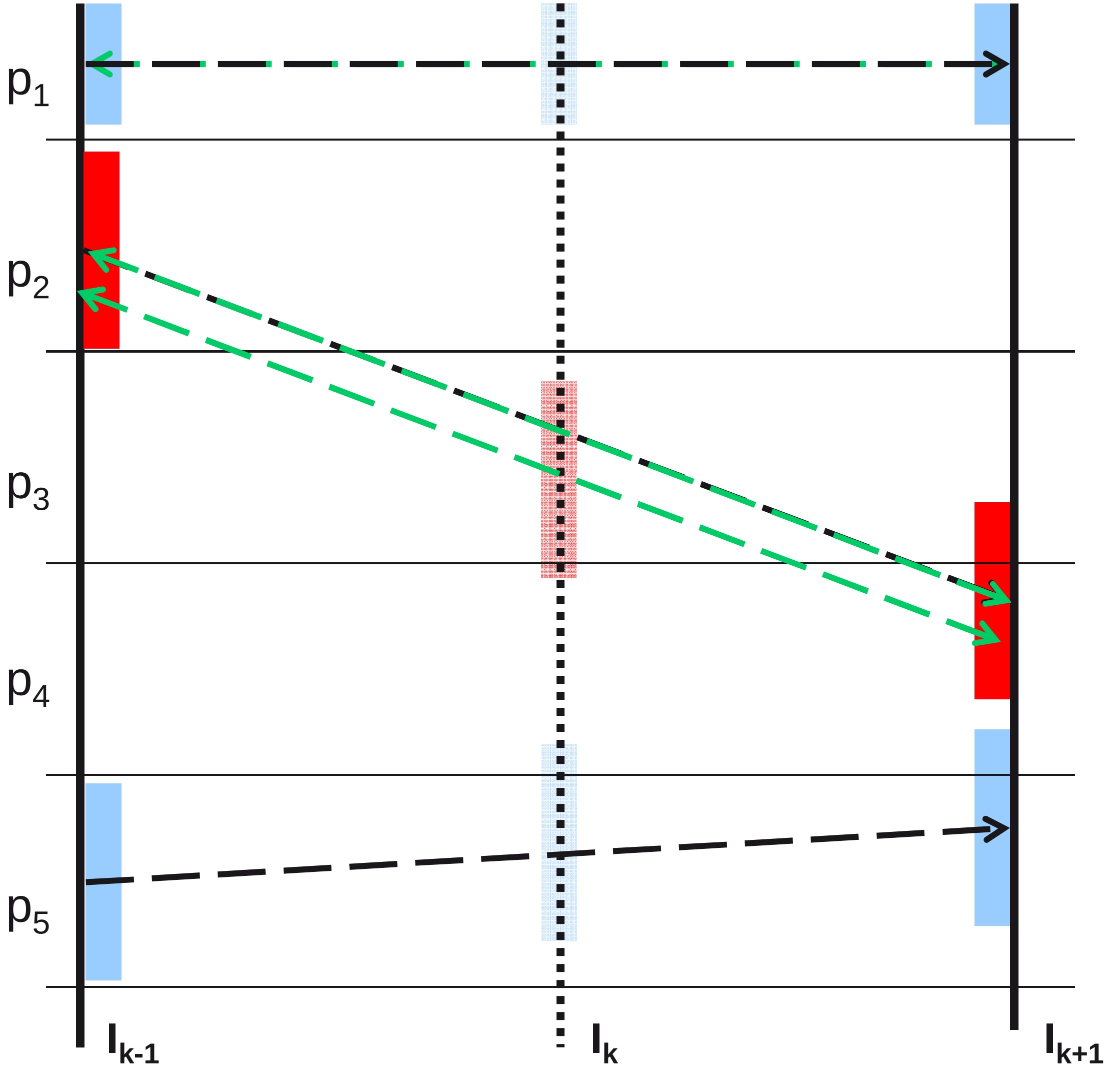
$\mathbf{I}_{k+1}$

$$d(\mathbf{v}) = d\left(B_{k-1}^{(\mathbf{p})}, B_{k+1}^{(\mathbf{p}+\mathbf{v})}\right) \qquad \mathbf{v}^* = \arg\min_{\mathbf{v}} d(\mathbf{v})$$

The DISCOVER algorithm: Split of monodirectional vectors



The DISCOVER algorithm: Split of monodirectional vectors

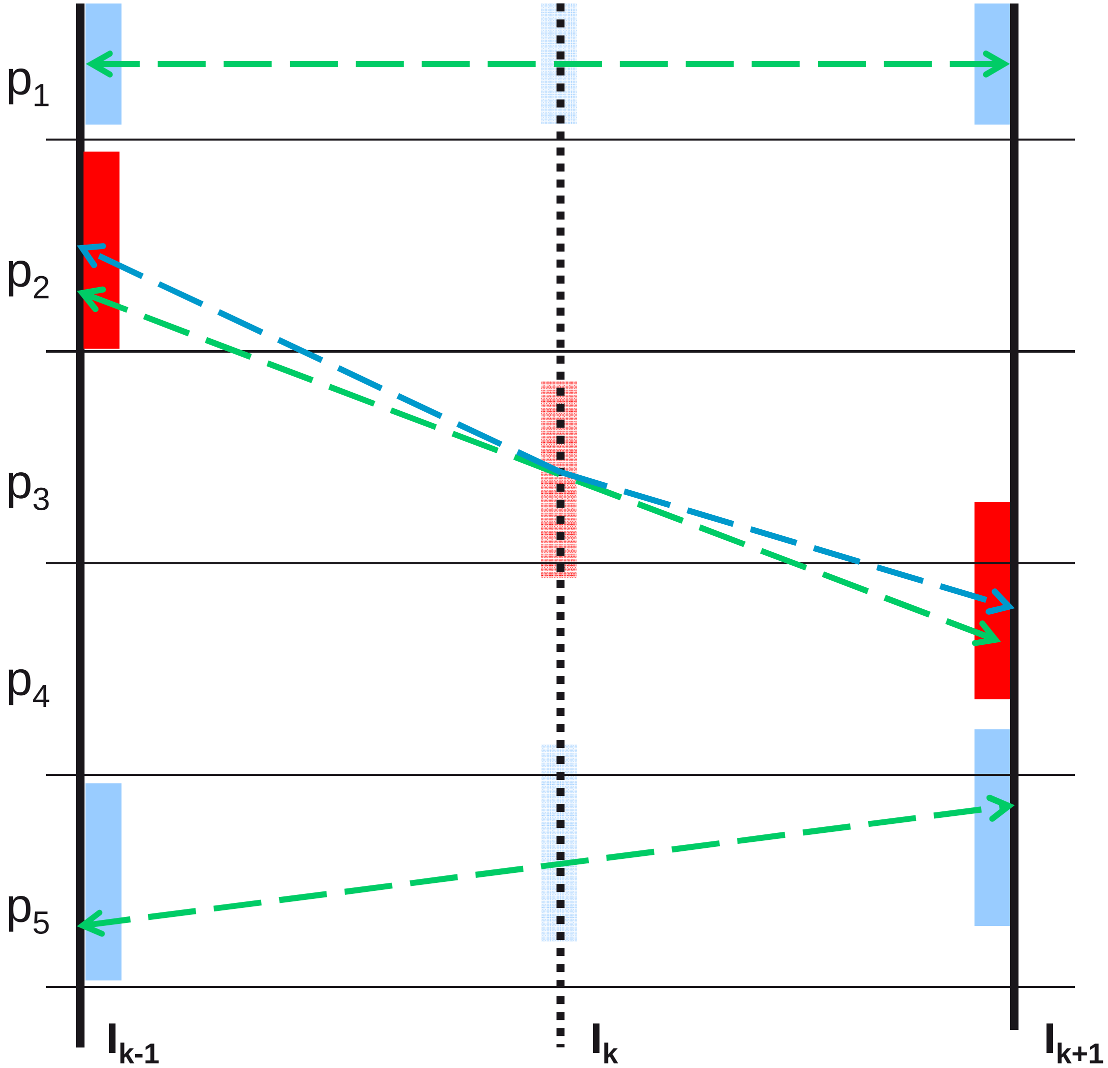


Block centered on pel p_4

$v_B(p_1) = -v(p_1)/2$
$v_F(p_1) = v(p_1)/2$

$v_B(p_2+v(p_2)/2) = -v(p_2)/2$
$v_F(p_2+v(p_2)/2) = v(p_2)/2$

The DISCOVER algorithm: Refinement of bidirectional vectors



The DISCOVER algorithm: Refinement and Median Filtering

- **Split MVs are further refined with a block matching in a small window near their value**
- **Median filtering is performed to enforce regular MVs**
- **The two motion-compensated images are added to produce the Side Information**

The DISCOVER algorithm: Sample interpolated image

PSNR:

26.4 dB

Test Conditions




- Spatial resolution: QCIF.
- Temporal resolution: 15 Hz (i.e. 7.5 Hz for the WZ frames with GOP=2).
- GOP size: 2, 4 and 8.

16	8	0	0
8	0	0	0
0	0	0	0
0	0	0	0

(a)

32	8	0	0
8	0	0	0
0	0	0	0
0	0	0	0

(b)

32	8	4	0
8	4	0	0
4	0	0	0
0	0	0	0

(c)

32	16	8	4
16	8	4	0
8	4	0	0
4	0	0	0

(d)

32	16	8	4
16	8	4	4
8	4	4	0
4	4	0	0

(e)

64	16	8	8
16	8	8	4
8	8	4	4
8	4	4	0

(f)

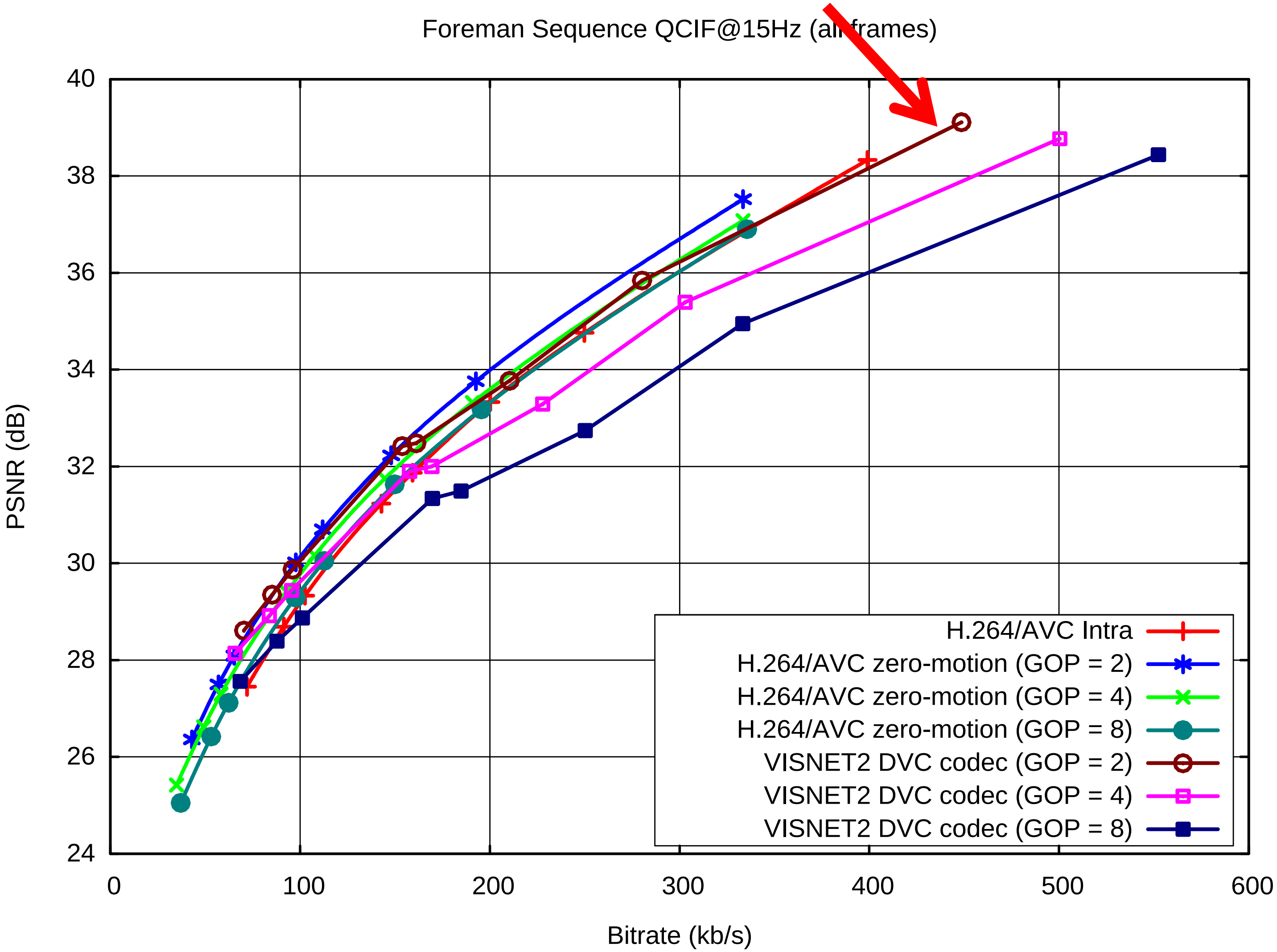
64	32	16	8
32	16	8	4
16	8	4	4
8	4	4	0

(g)

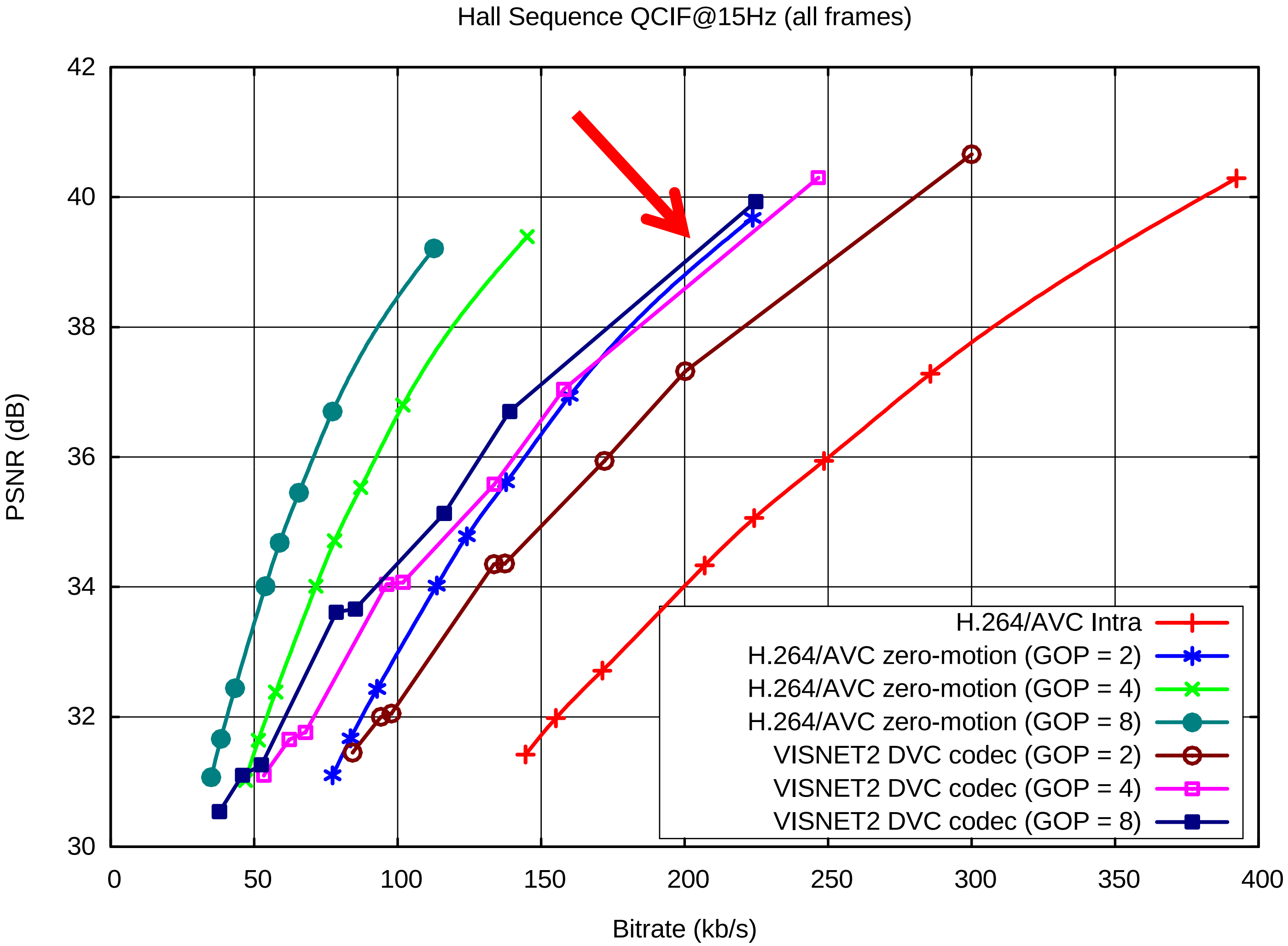
128	64	32	16
64	32	16	8
32	16	8	4
16	8	4	0

(h)

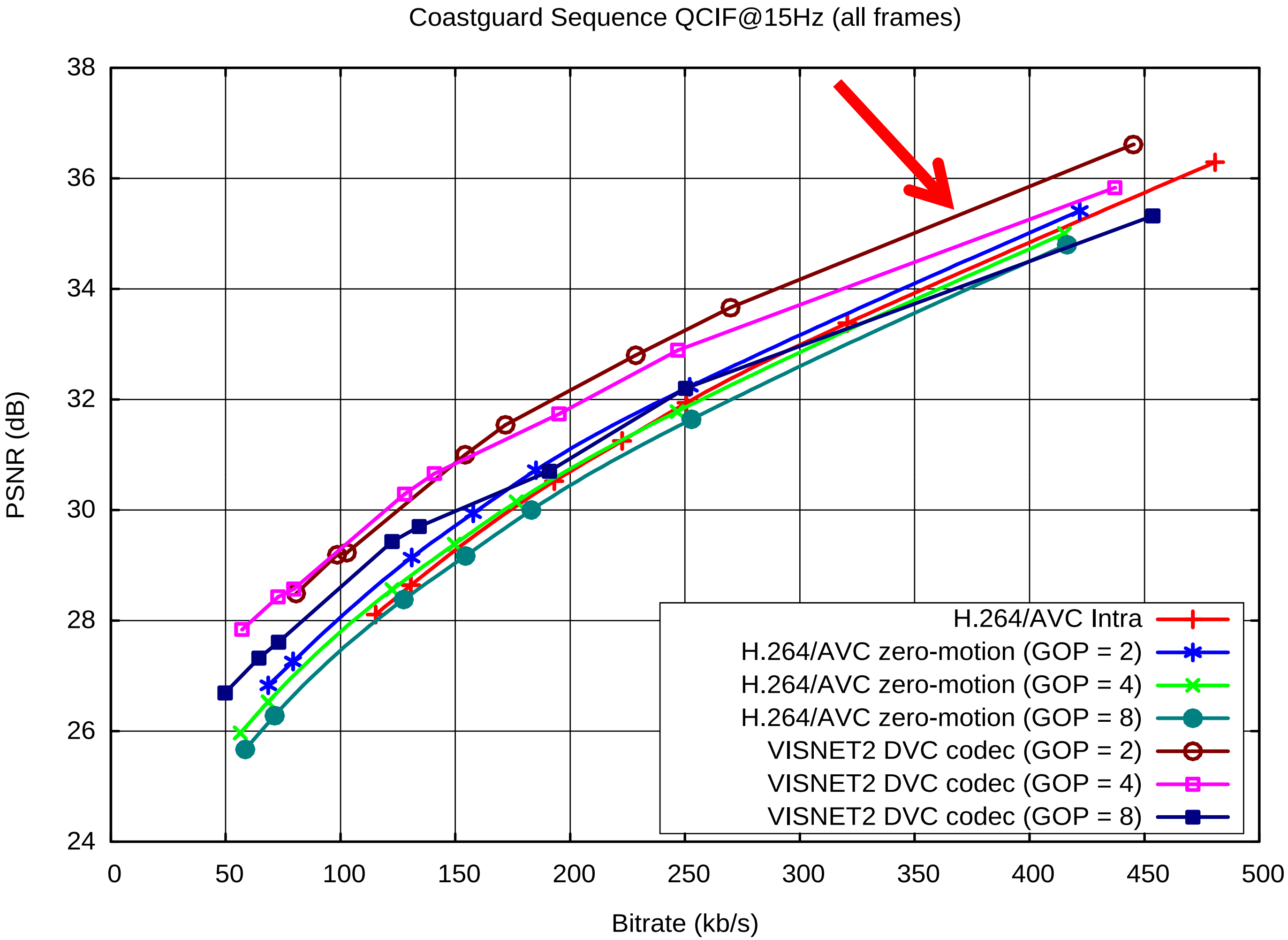
VISNET II DVC versus H.264/AVC: Foreman



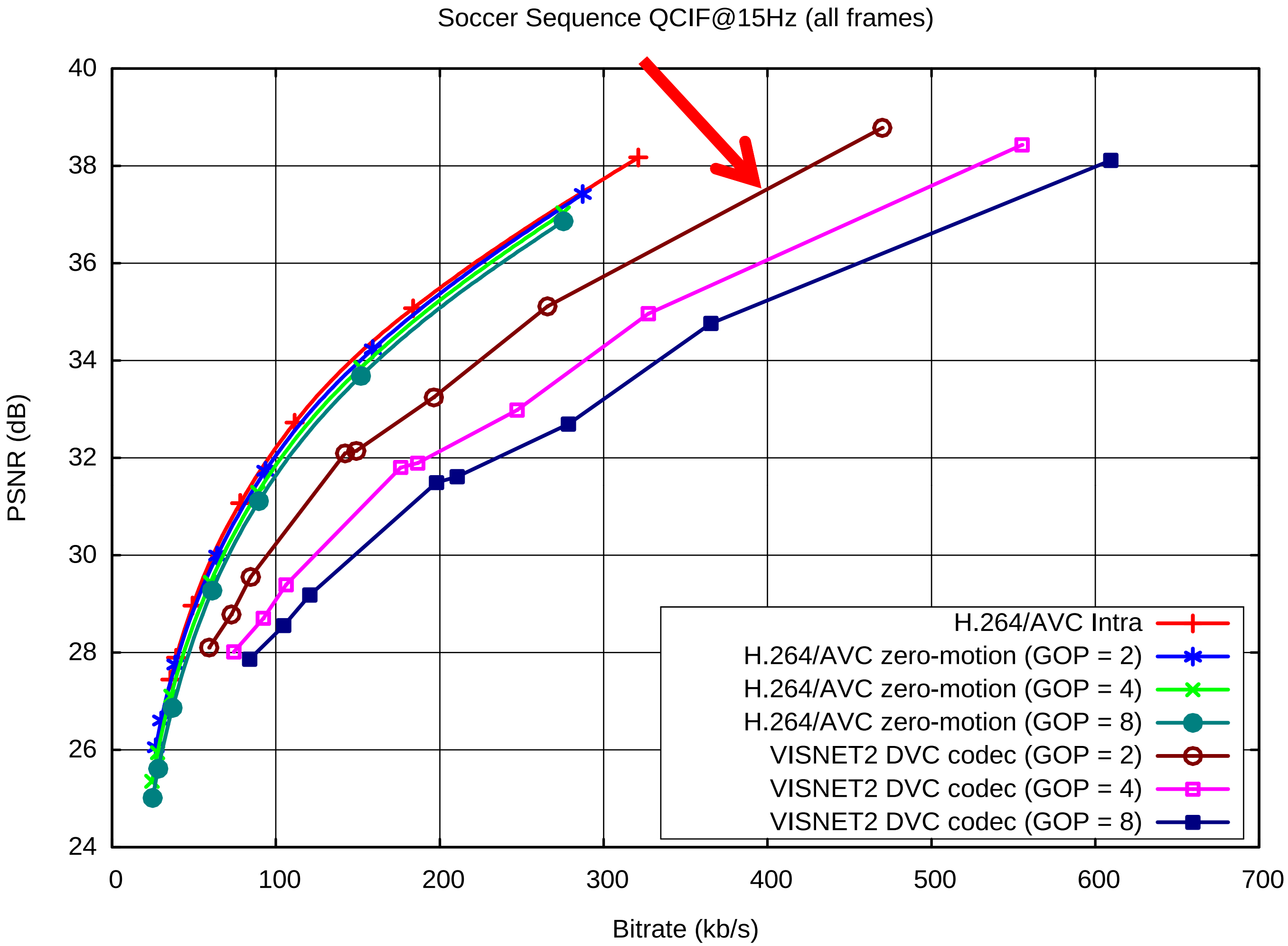
VISNET II DVC versus H.264/AVC: Hall Monitor



VISNET II DVC versus H.264/AVC: Coastguard



VISNET II DVC versus H.264/AVC: Soccer



Complexity

- **WZ frame encoding complexity is approximately 1/6 of the H.264/AVC Intra or H.264/AVC No Motion encoding complexity**
- **However, DVC decoding complexity is much higher (some orders of magnitude) than H.264/AVC Intra or H.264/AVC No Motion decoding complexity**
- **DVC decoding complexity is strongly dependent on the quality of SI**
- **Substantial on-going work on fast and parallel implementations of channel decoding algorithms**

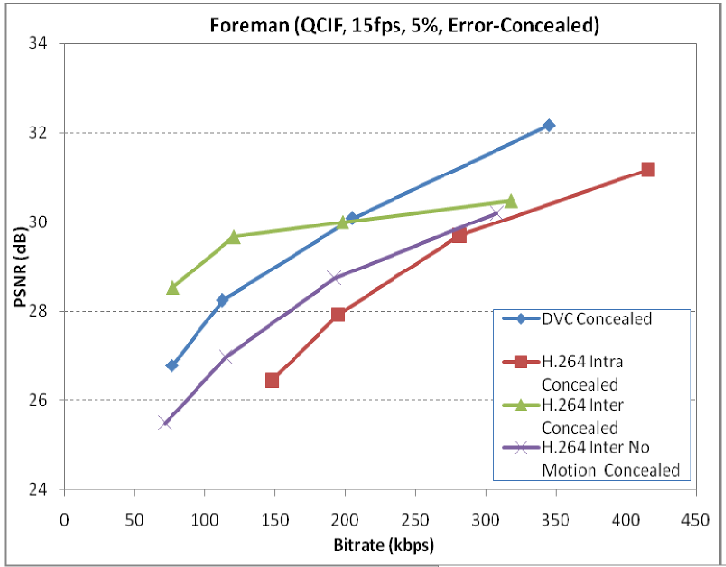
Robust Transmission

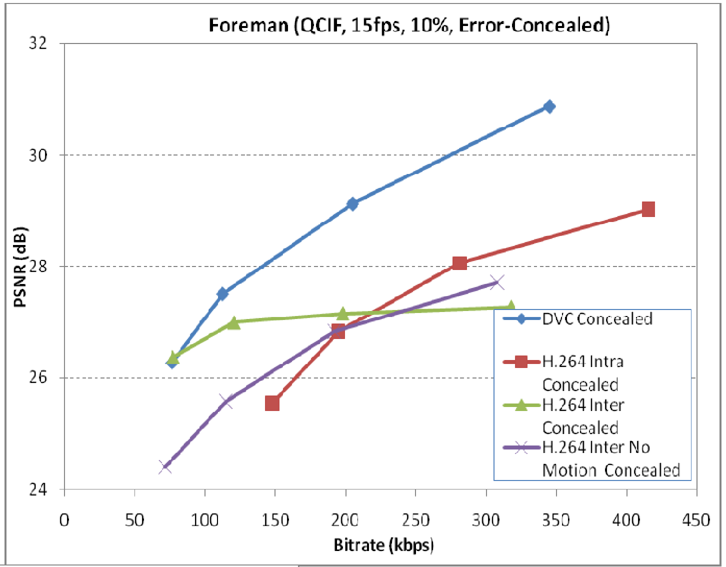
- **Appealing for transmission over error-prone channels**
  - Statistical framework rather than a deterministic approach
  - Absence of a prediction loop in the codec
- **Decoding is successful, even in the presence of transmission errors, as long as the SI is within the noise margin of the encoded parity bits**
- **Scalable schemes robust to packet losses both in the base and enhancement layers**
- **Increase the robustness of standard encoded video by adding redundant information encoded according to distribute coding principles**

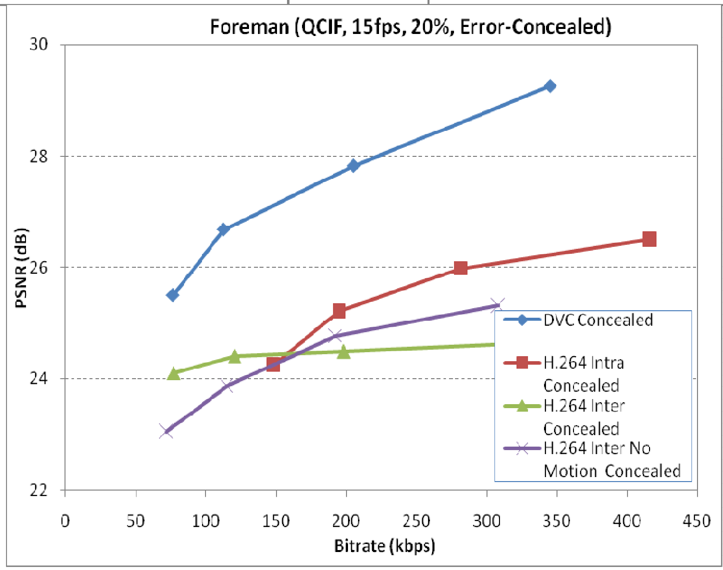
Robust Transmission

- **DVC**
  - WZ frames: hybrid spatial and temporal error concealment
  - Key frames: JM error concealment
- **H.264/AVC**
  - JM 11.0
  - Flexible Macroblock Ordering (FMO)
  - JM error concealment
- **With/without feedback channel**
  - Automatic Repeat reQuest (ARQ)
- **Packet Loss Rate**
  - 5%, 10%, 20%, error patterns from VCEG

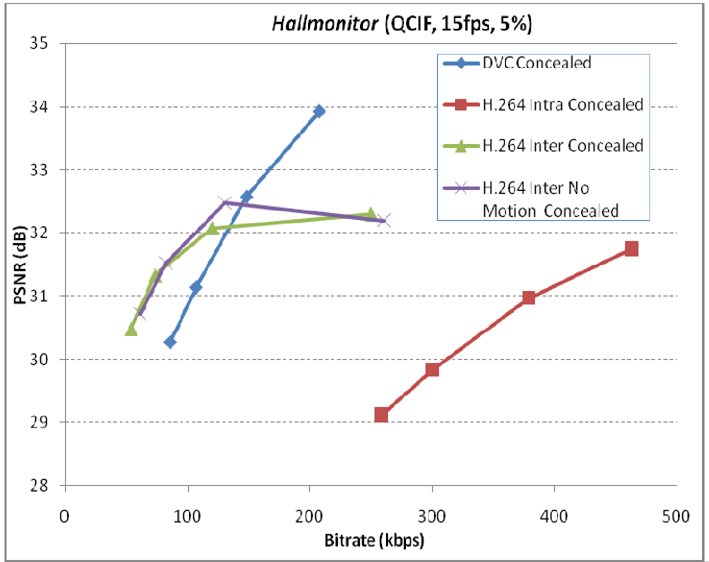
Foreman, no feedback channel

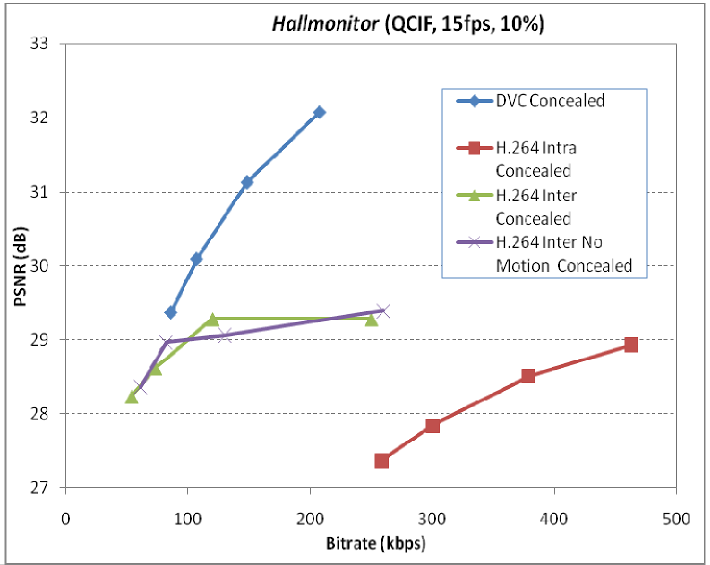


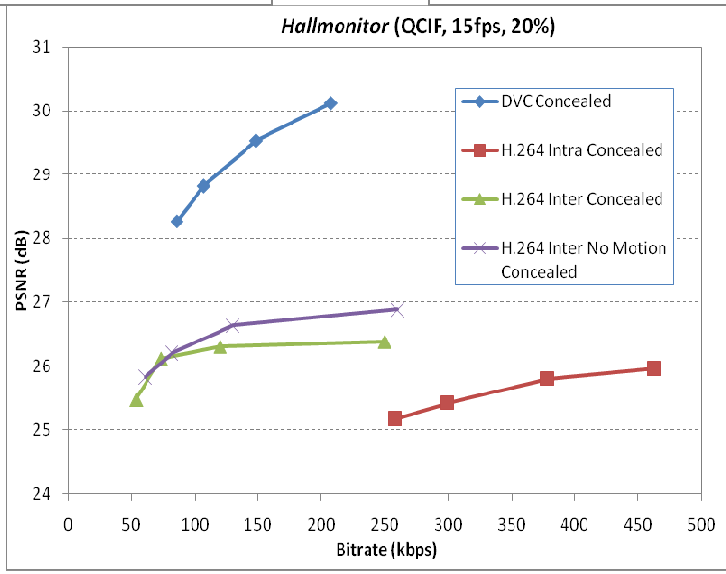




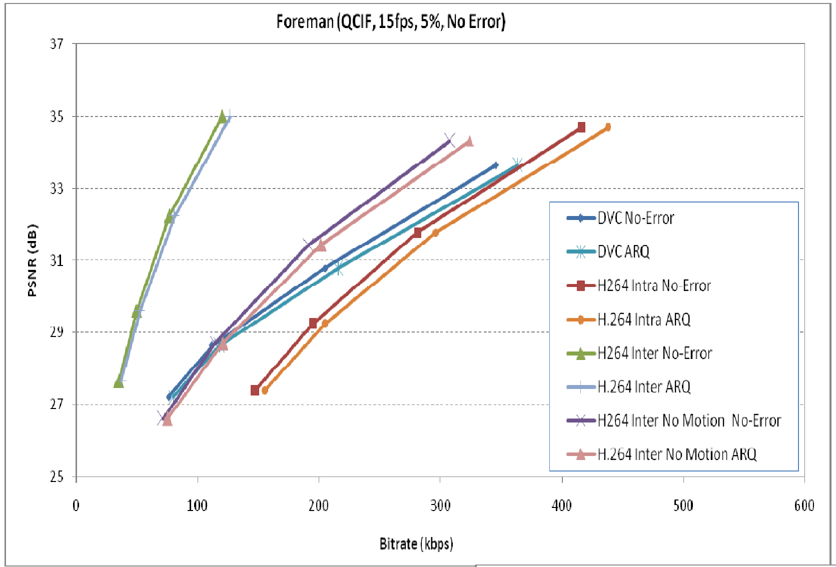
Hall Monitor, no feedback channel

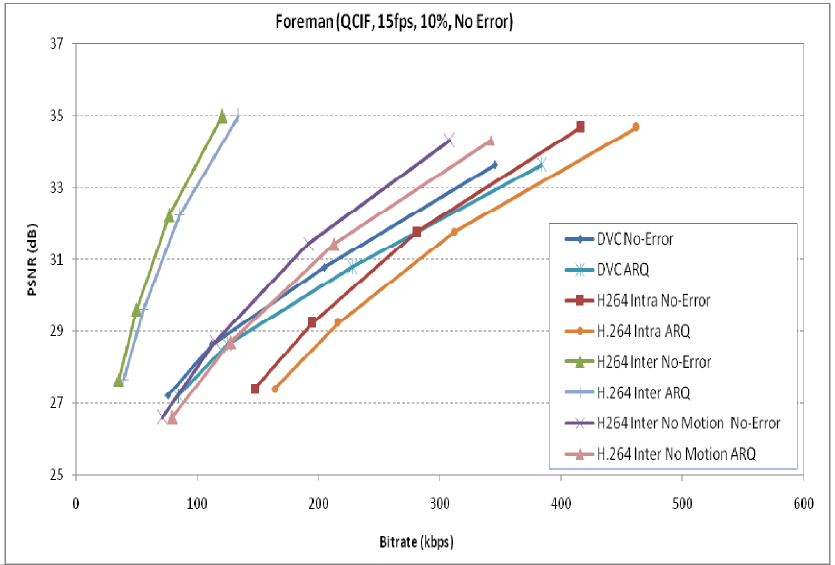


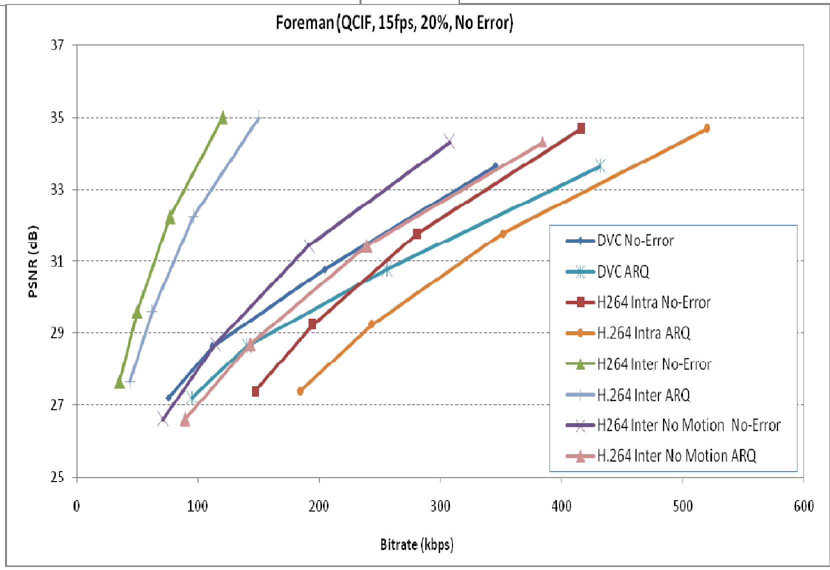




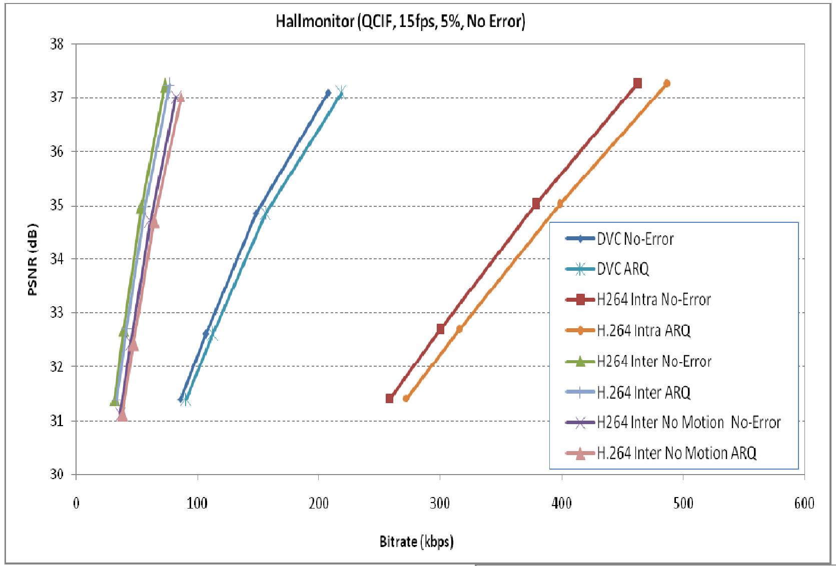
Foreman, feedback channel

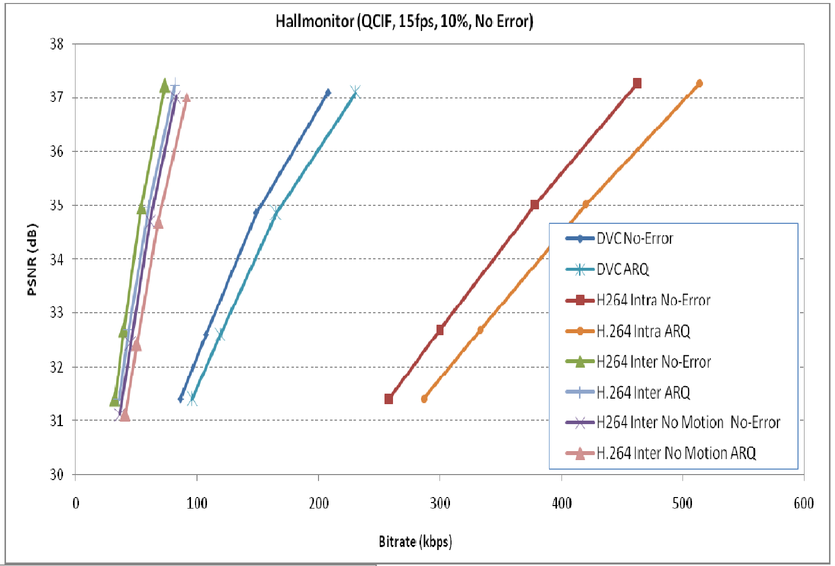


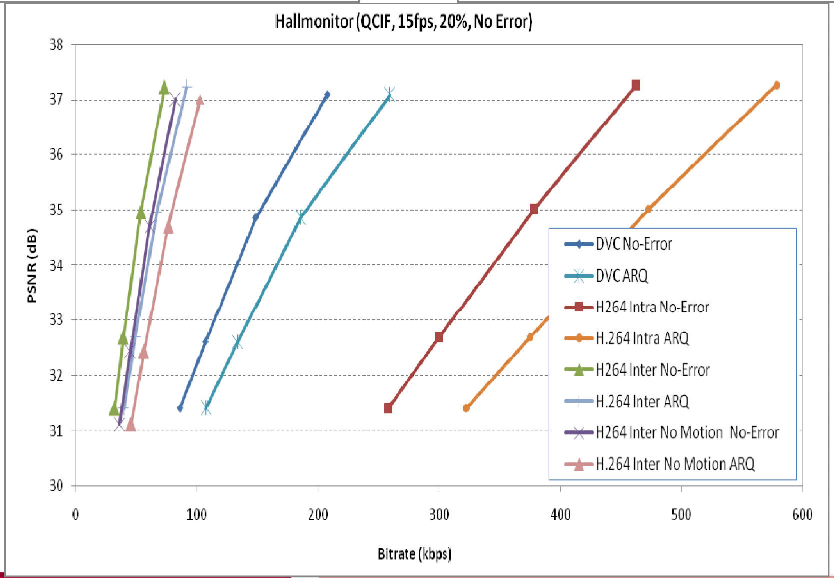




Hall Monitor, feedback channel







Multiview Video Coding

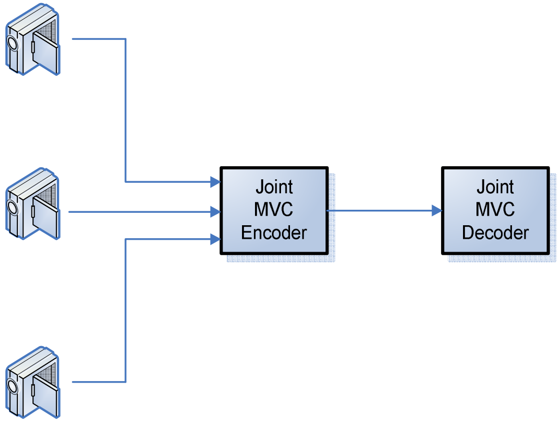
Multiview video coding

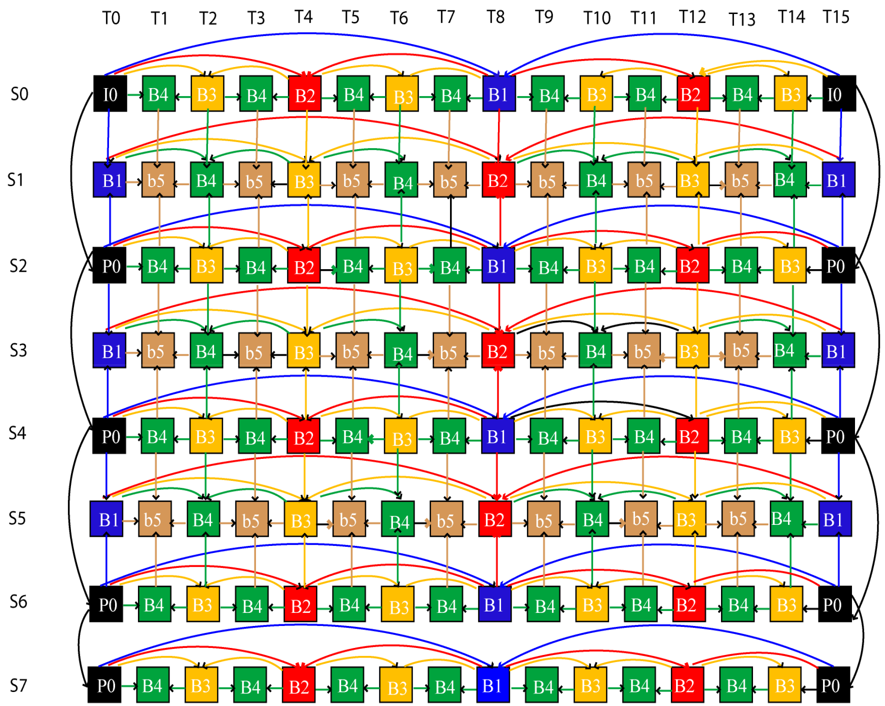
- **Emerging problem**
- **Camera arrays, stereoscopic video**

- **Inter-view correlation and disparity estimation**
- **Temporal correlation and motion estimation**

- **Huge complexity → DVC techniques**
- **Conceptually close to the monoview case**
  - Key frames and Wyner-Ziv frames

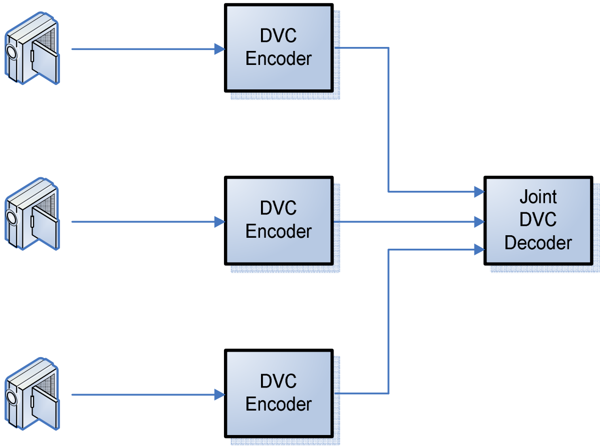
Multi-View Video Coding





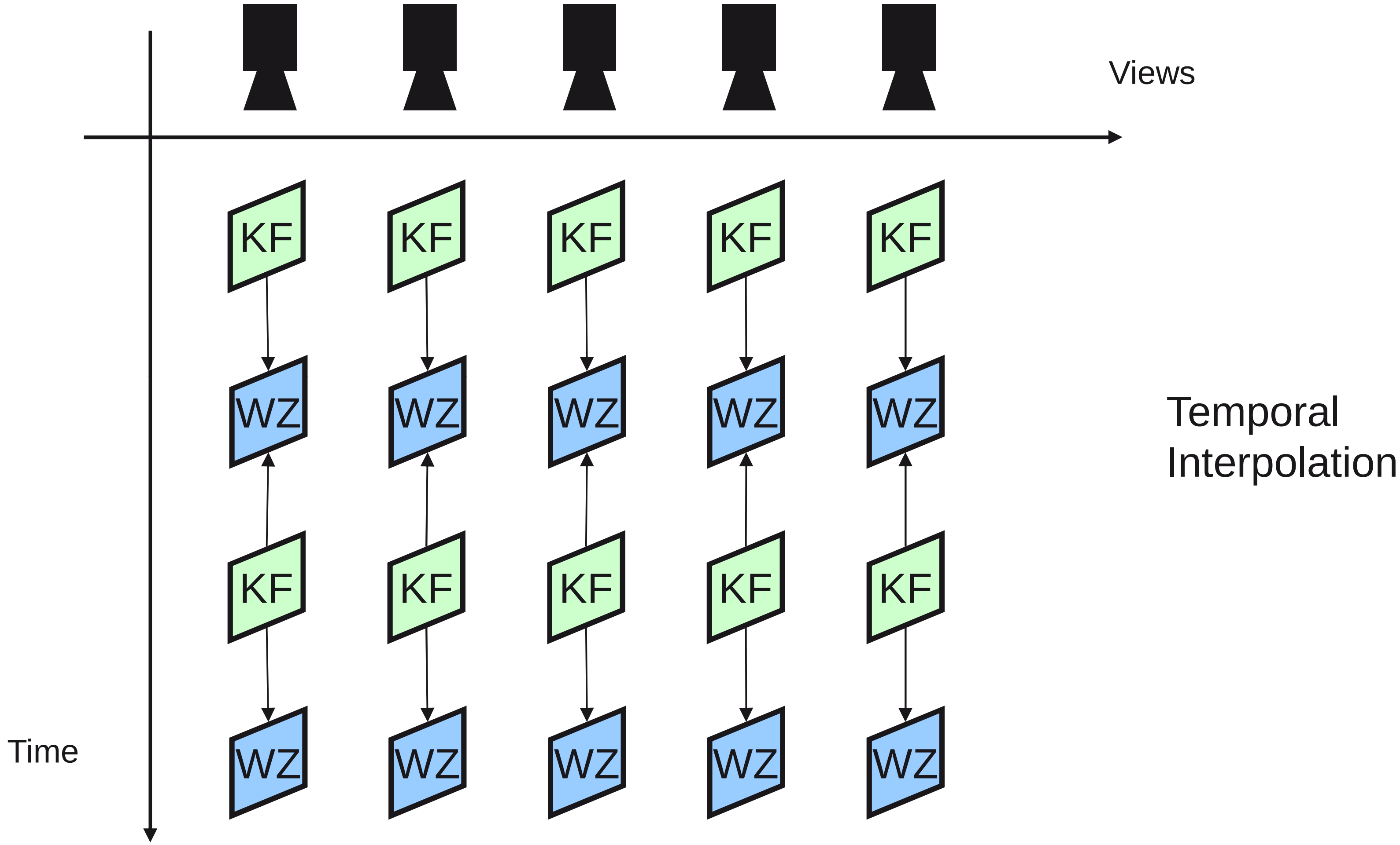
- MVC
  - Extension of AVC
  - Block-based predictive coding along time and across views
  - Very complex encoder
  - Cameras have to communicate

Multi-View Distributed Video Coding

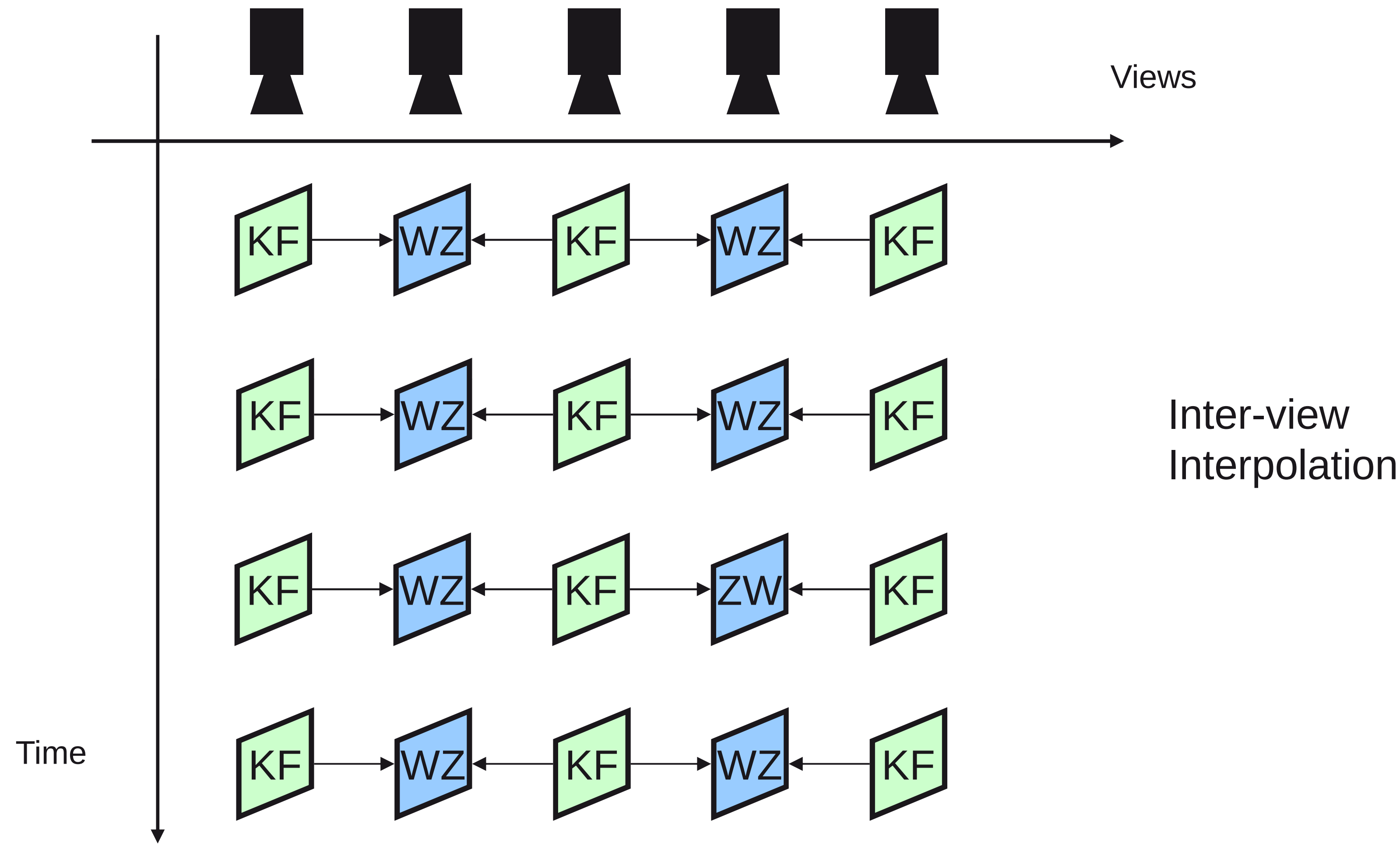


- DVC
  - Low complexity / lower power consumption encoder
  - Exploit inter-view correlation without communication between cameras

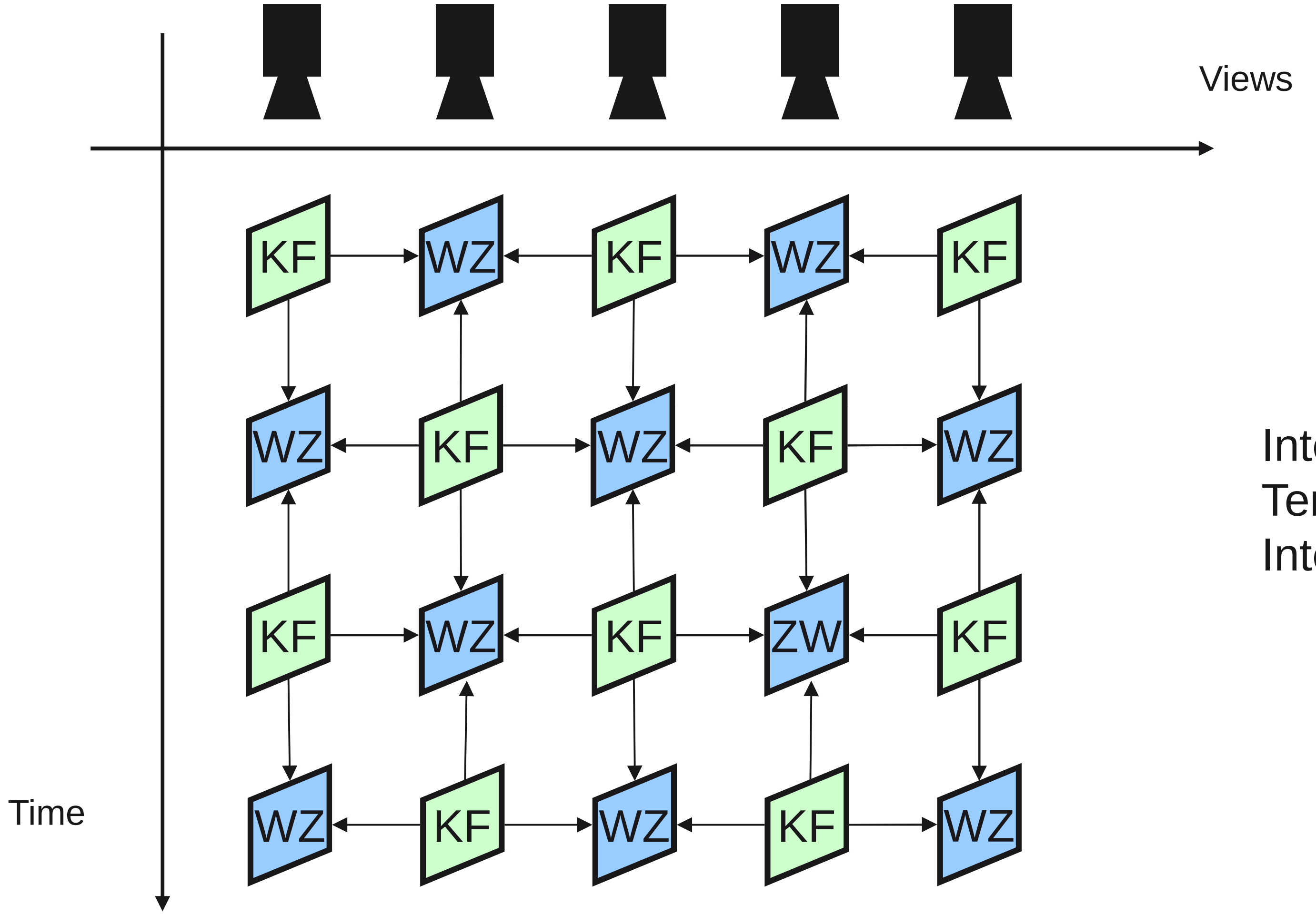
Multiview video coding: possible schemes



Multiview video coding: possible schemes

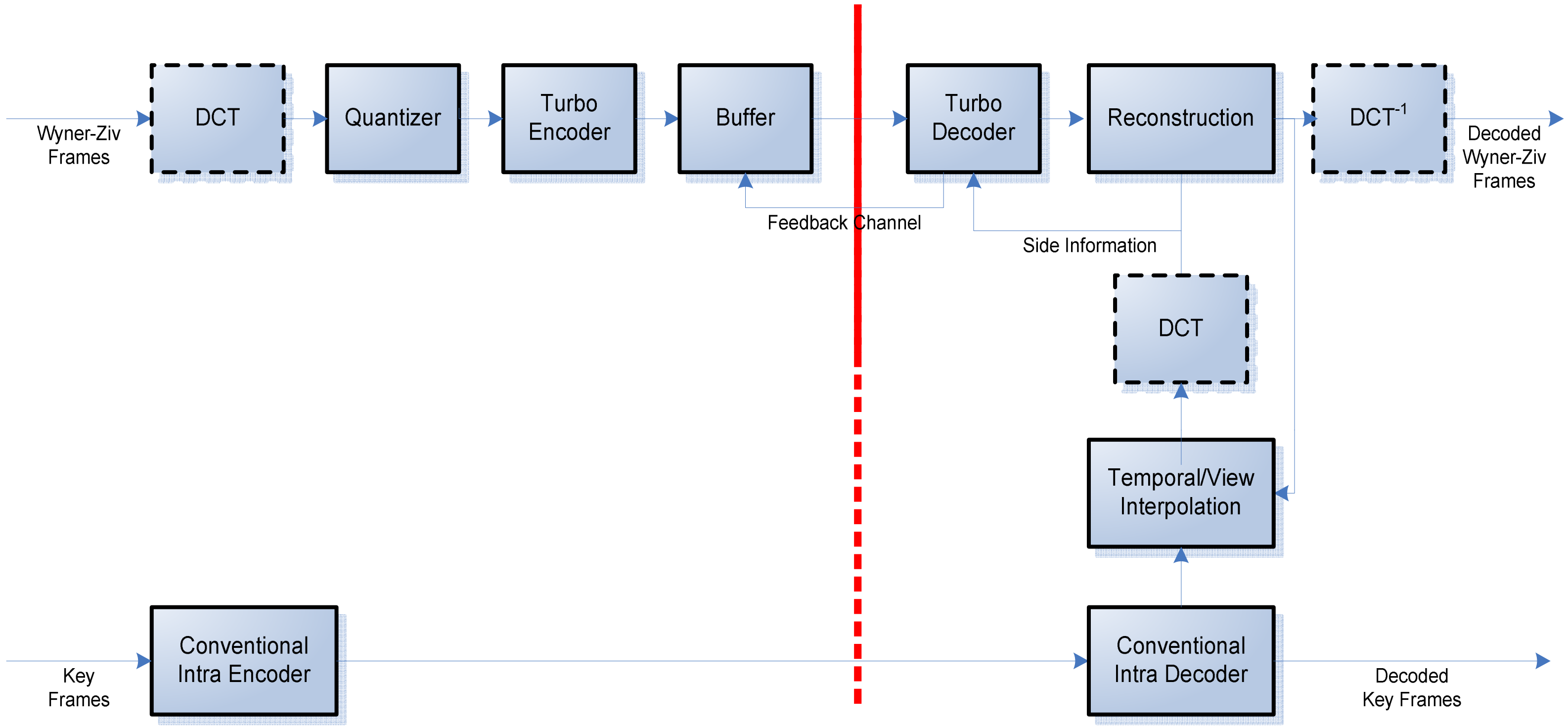


Multiview video coding: possible schemes



Interview and Temporal Interpolation

Multiview DVC



Inter-View Side Information

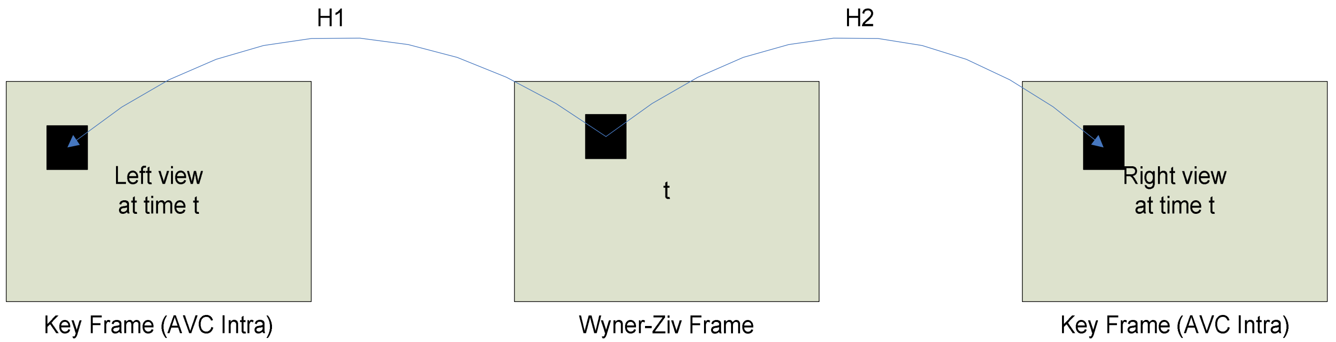
- Disparity Compensation View Prediction (DCVP)
  - Straightforward extension of MCTI
  - Disparity vectors are estimated between views
  - Interpolation at mid-point to generate SI

Inter-View Side Information

- Homography
  - Homography relating the central view to side views
  - Assumption that the scene is planar
  - Parameters have to be computed once

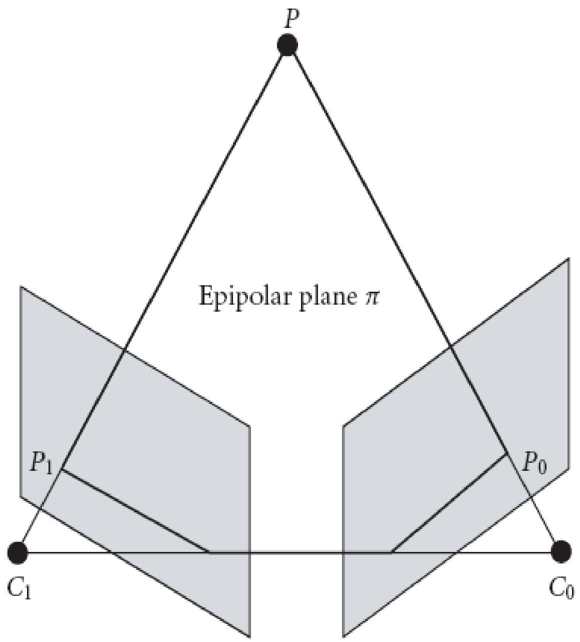
$$x_i' = \frac{a_0 + a_2 x_i + a_3 y_i}{a_6 x_i + a_7 y_i + 1}$$

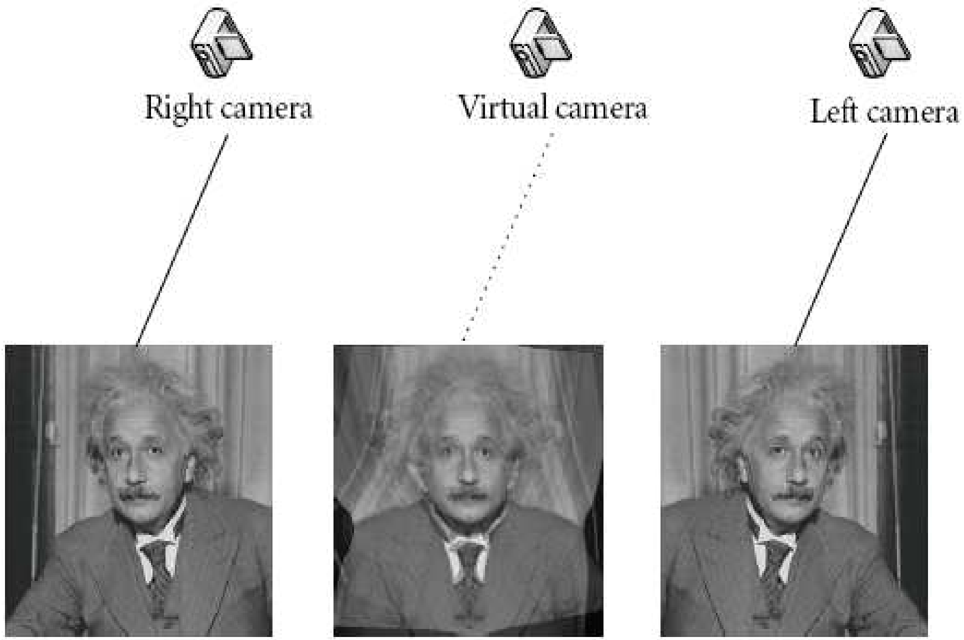
$$y_i' = \frac{a_1 + a_4 x_i + a_5 y_i}{a_6 x_i + a_7 y_i + 1}$$



Inter-View Side Information

- View Morphing (VM)
  - Fundamental matrix: map a point in one camera and its epipolar line in the other camera
  - Requires at least seven point correspondences

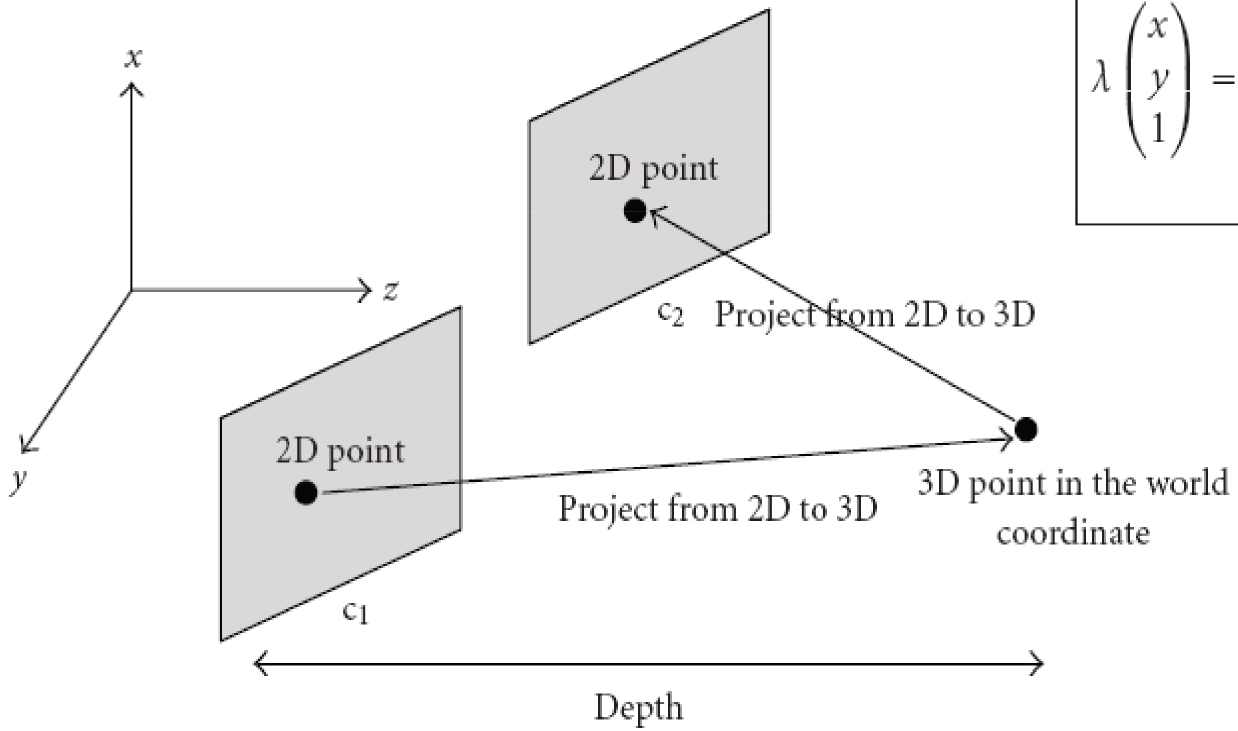




Inter-View Side Information

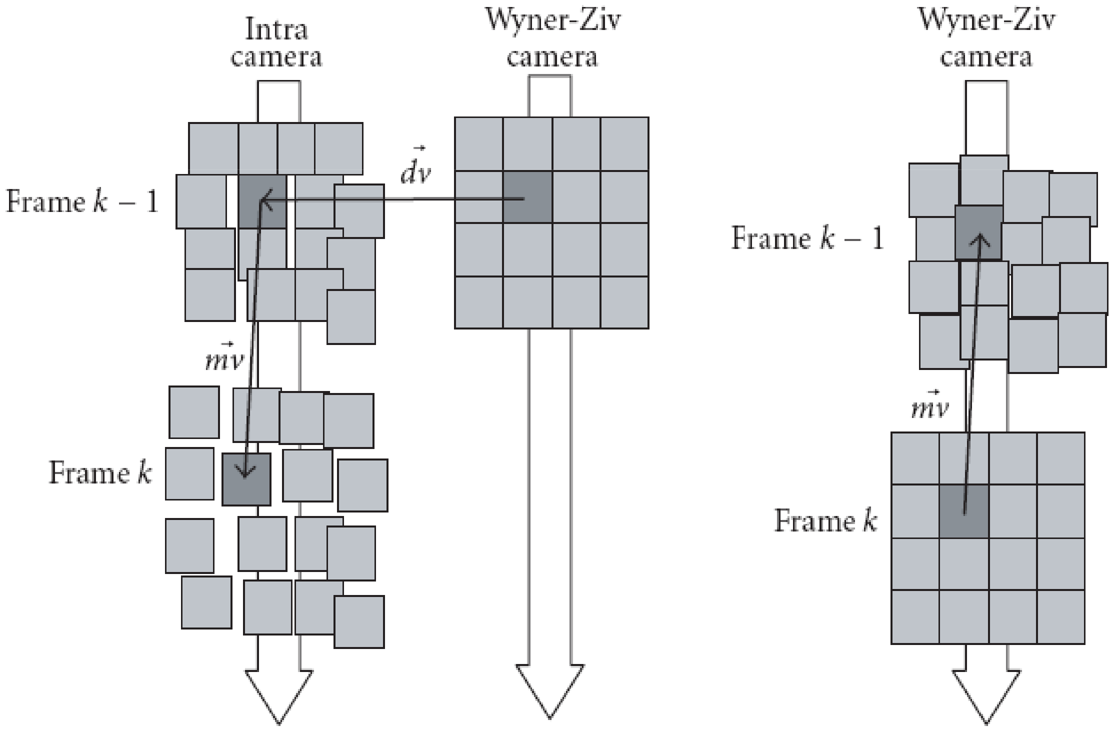
- View Synthesis Prediction (VSP)
  - Camera calibration
  - Intrinsic and extrinsic camera parameters
  - Depth information

$$\lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = A \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_{3D} \\ Y_{3D} \\ Z_{3D} \\ 1 \end{pmatrix}$$



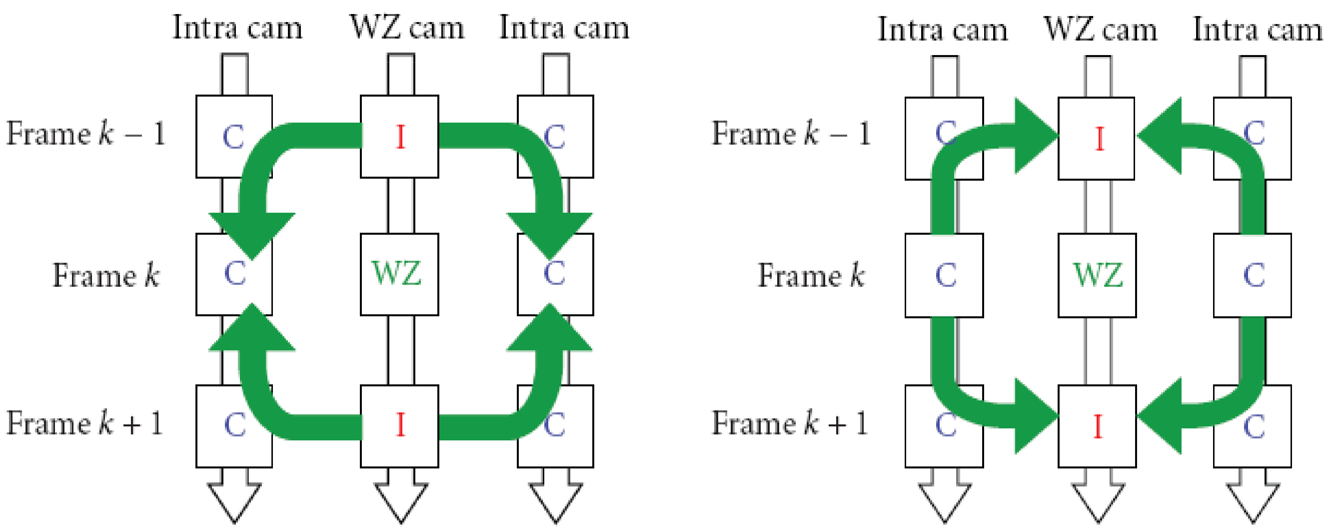
Inter-View Side Information

- Multi-View Motion Estimation (MVME)
  - Compute motion vectors in a side view
  - Apply them to current view (WZ frame) using disparity vectors

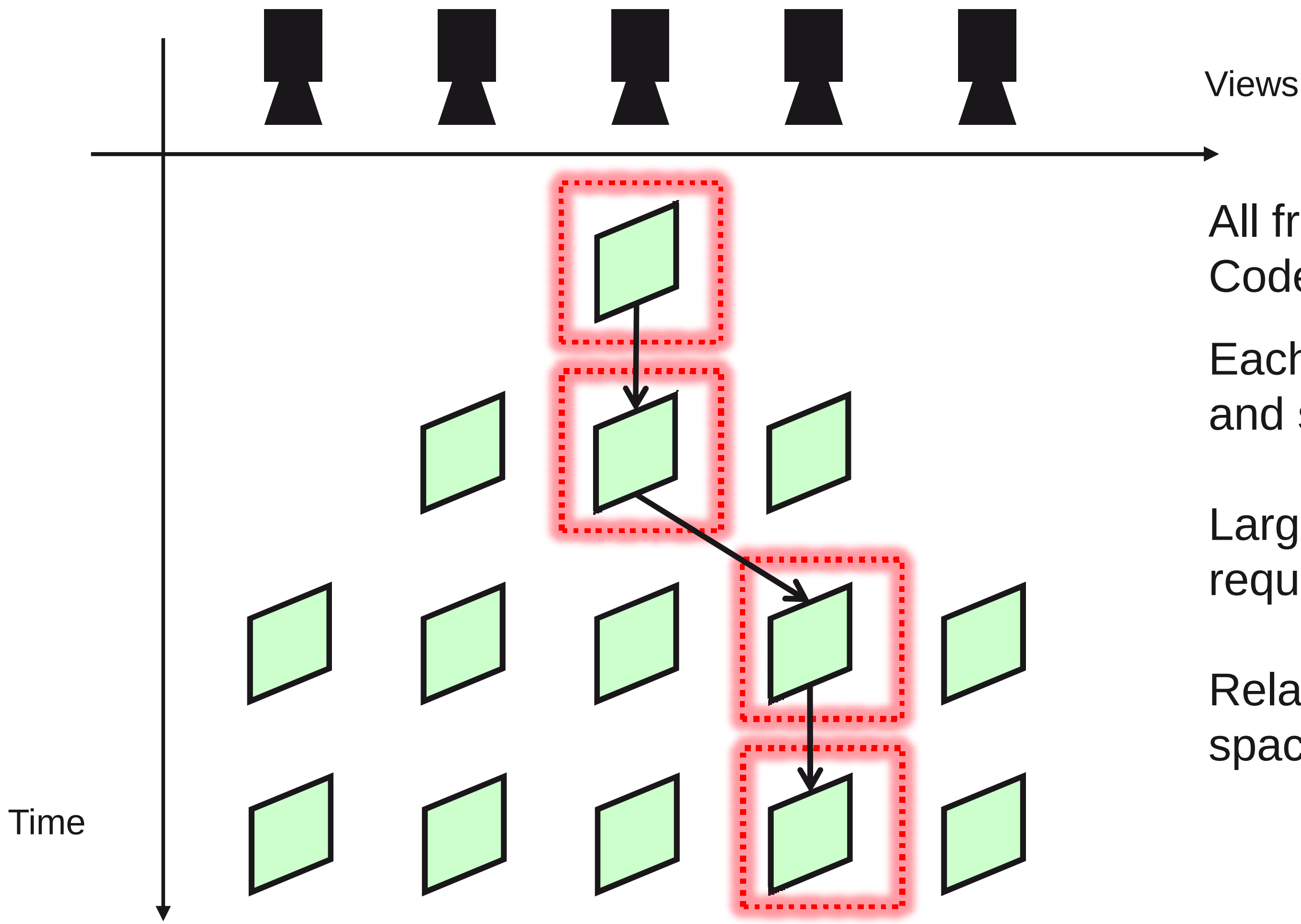


Inter-View Temporal Side Information

- Multi-View Motion Estimation (MVME)
  - 8 different possible paths
  - Weighted average using reliability measure (MSE or SAD of matching error)



Application to IMVS: Interactive Multiview Video Streaming



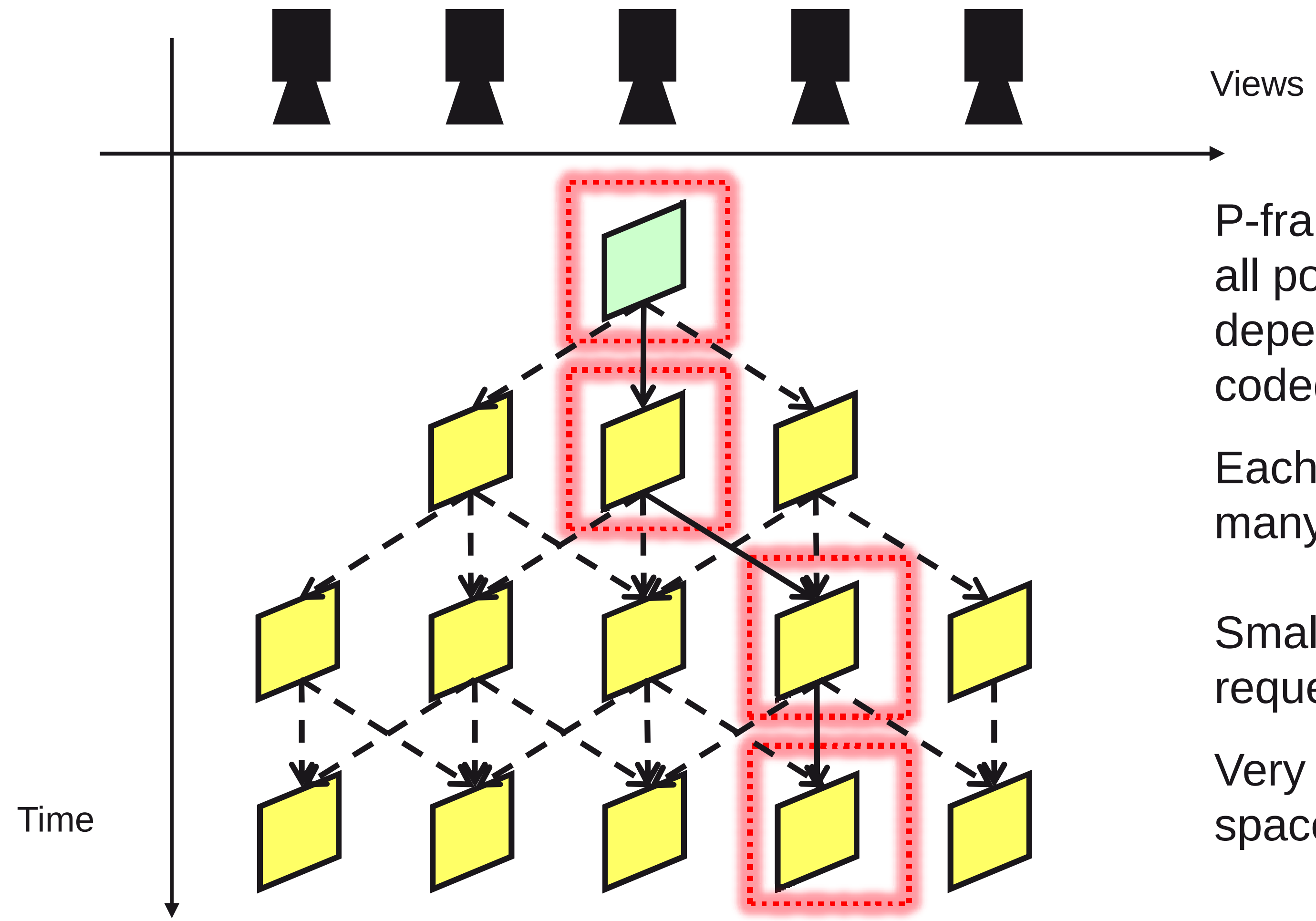
All frames are Intra Coded

Each image is coded and stored only once

Large bandwidth requested

Relatively low server space requested

Application to IMVS: Interactive Multiview Video Streaming



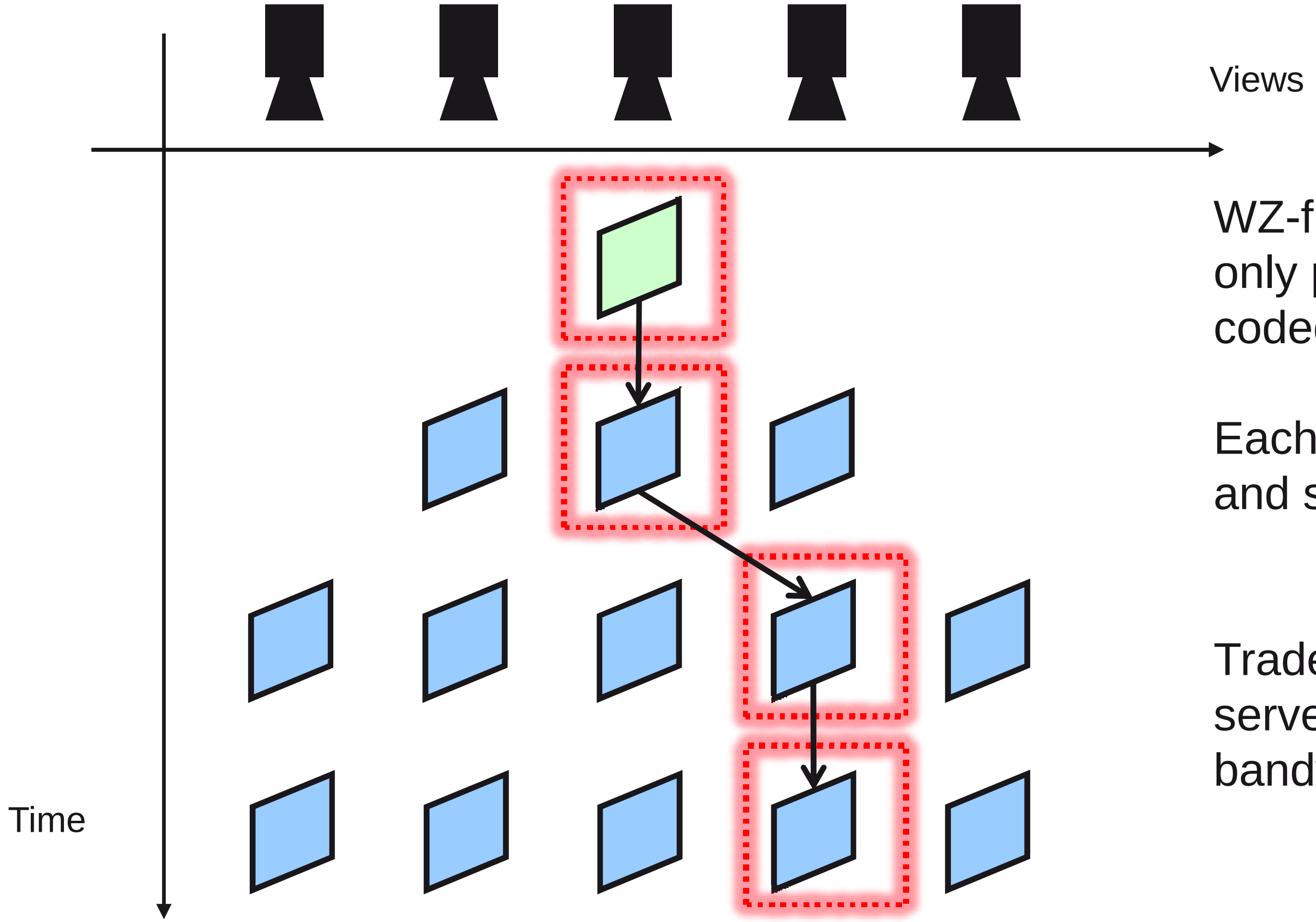
P-frames are used: all possible frame dependencies are coded

Each image is coded many times

Smallest bandwidth requested

Very large server space requested

Application to IMVS: Interactive Multiview Video Streaming



WZ-frames are used: only parity bits are coded

Each image is coded and stored only once

Trade-off between server space and bandwidth

Conclusions

Most Promising Applications

Application	Flexible allocation of codec complexity	Improved error resilience	Codec independent scalability	Exploitation of multi-view correlation
Wireless video cameras	X	X		
Wireless low-power surveillance	X	X	X	X
Mobile document scanner	X	X		
Video conferencing with mobile devices	X	X		
Mobile video mail	X			
Disposable video cameras	X			
Visual sensor networks	X	X	X	X
Networked camcorders	X	X		X
Distributed video streaming	X	X	X	
Multiview video entertainment	X			X
Wireless capsule endoscopy	X	X		

Conclusions

- **DVC allows very low-complexity video coding**
  - In theory without loss in RD performance
  - In practice some loss seems unavoidable
- **DVC allows graceful degradation in unreliable environment**
  - Joint source/channel coding naturally applies to the channel coding used in DVC
- **DVC enables MVC with low computational power**
  - Distributed exploitation of inter-view correlation

Further reading

- J. Slepian and J. Wolf, “Noiseless Coding of Correlated Information Sources”, IEEE Trans. on Information Theory, vol. 19, no. 4, pp. 471-480, July 1973.
- A. Wyner and J. Ziv, “The Rate-Distortion Function for Source Coding with Side Information at the Decoder”, IEEE Trans. on Information Theory, vol. 22, no. 1, pp. 1-10, January 1976.

- R. Puri, A. Majumdar, and K. Ramchandran, “PRISM: A Video Coding Paradigm with Motion Estimation at the Decoder”, IEEE Transactions on Image Processing, vol. 16, no. 10, pp. 2436-2448, October 2007.
- B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, “Distributed Video Coding”, Proceedings of the IEEE, vol. 93, no. 1, pp. 71-83, January 2005.
- C. Guillemot, F. Pereira, L. Torres, T. Ebrahimi, R. Leonardi and J. Ostermann, “Distributed Monoview and Multiview Video Coding”, IEEE Signal Processing Magazine, vol. 24, no. 5, pp. 67-76, September 2007.
- P.L Dragotti and M. Gastpar, Distributed Source Coding: Theory, Algorithms and Applications, Academic Press, February 2009.
- F. Dufaux, W. Gao, S. Tubaro, A. Vetro, “Distributed Video Coding: Trends and Perspectives”, EURASIP Journal on Image and Video Processing, (review article, special issue on DVC), vol. 2009, Article ID 508167, doi:10.1155/2009/508167, 2009.
- J. Ascenso, C. Brites, F. Dufaux, A. Fernando, T. Ebrahimi, F. Pereira and S. Tubaro, “The VISNET II DVC Codec: Architecture, Tools and Performance”, in Proc. 18th European Signal Processing Conference (EUSIPCO 2010), Aalborg, Denmark, August 2010.
- F. Pereira, L. Torres, C. Guillemot, T. Ebrahimi, R. Leonardi and S. Klomp, “Distributed Video Coding: Selecting the most promising application scenarios”, Signal Processing: Image Communication, Vol. 23, no. 5, pp. 339-352, June 2008.