Distributed Video Coding: Status, Challenges and Outlook

Frédéric Dufaux

frederic.dufaux@ieee.org

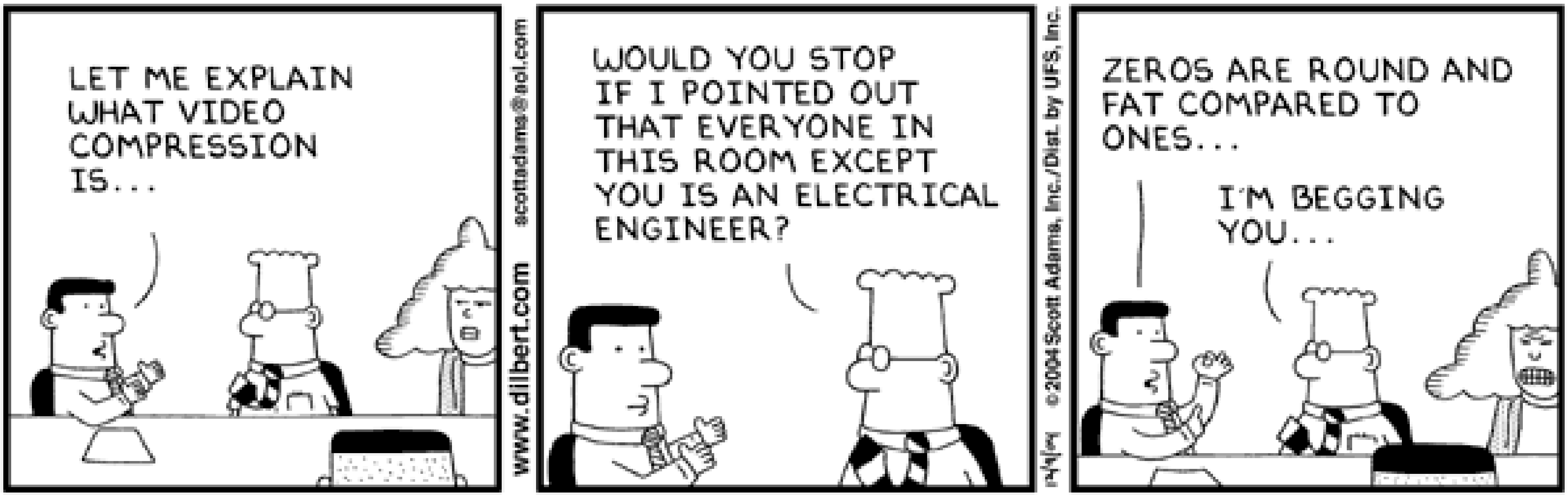
Multimedia Signal Processing Group
Swiss Federal Institute of Technology (EPFL)
Lausanne, Switzerland

Laboratoire Traitement et Communication de l'Information (LTCI) - UMR 5141 CNRS
Télécom ParisTech
Paris, France

Outline

- Video coding: context and background
- DVC: theoretical foundations
- Early DVC architectures
- Current research topics
  - Coding efficiency - the VISNET II DVC codec
  - Robust transmission
  - Multi-view video coding
- Outlook

Video Coding
Context and Background

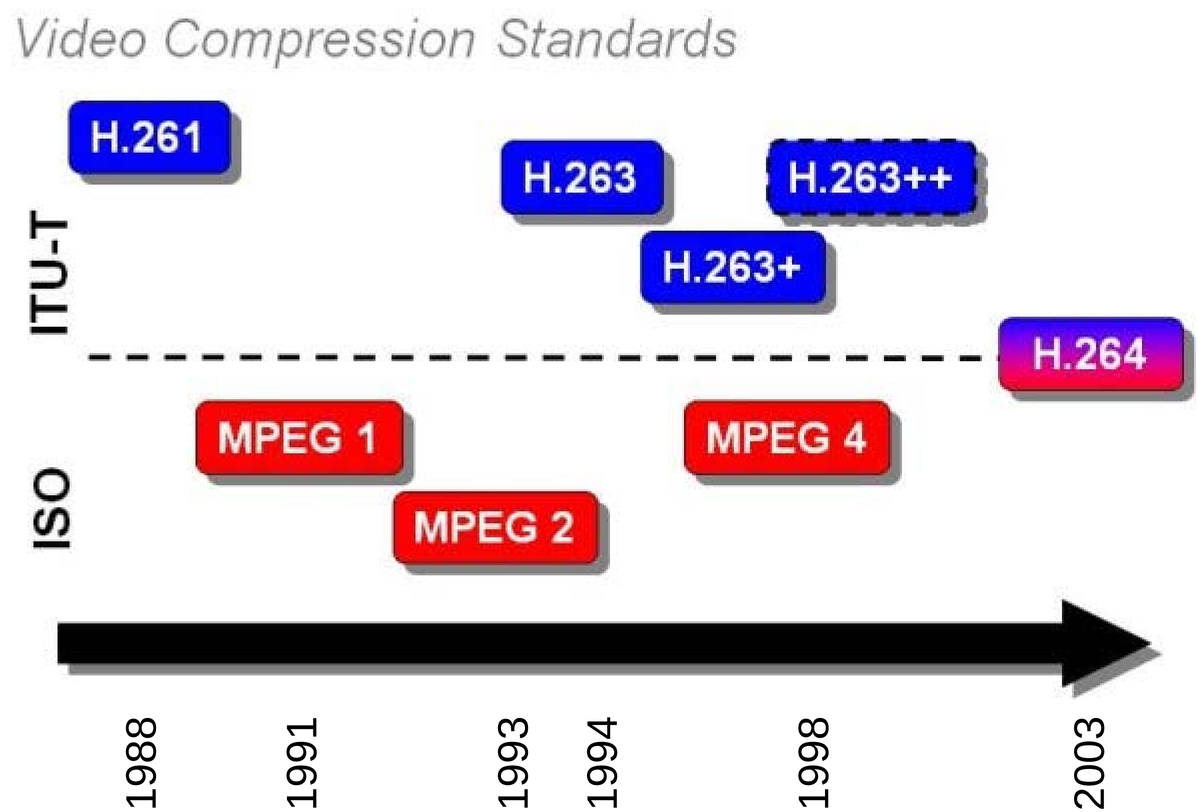

LET ME EXPLAIN WHAT VIDEO COMPRESSION IS...
WOULD YOU STOP IF I POINTED OUT THAT EVERYONE IN THIS ROOM EXCEPT YOU IS AN ELECTRICAL ENGINEER?
ZEROS ARE ROUND AND FAT COMPARED TO ONES...
I'M BEGGING YOU...
scottadams@aol.com
www.dilbert.com
© 2004 Scott Adams, Inc./Dist. by UFS, Inc.

Context

- Video coding
  - Efficient representation of video data
  - Fulfilling relevant requirements, e.g. compression, quality, error resilience, random access, interactivity
- Requirements are continuously evolving along with technological progress

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, …

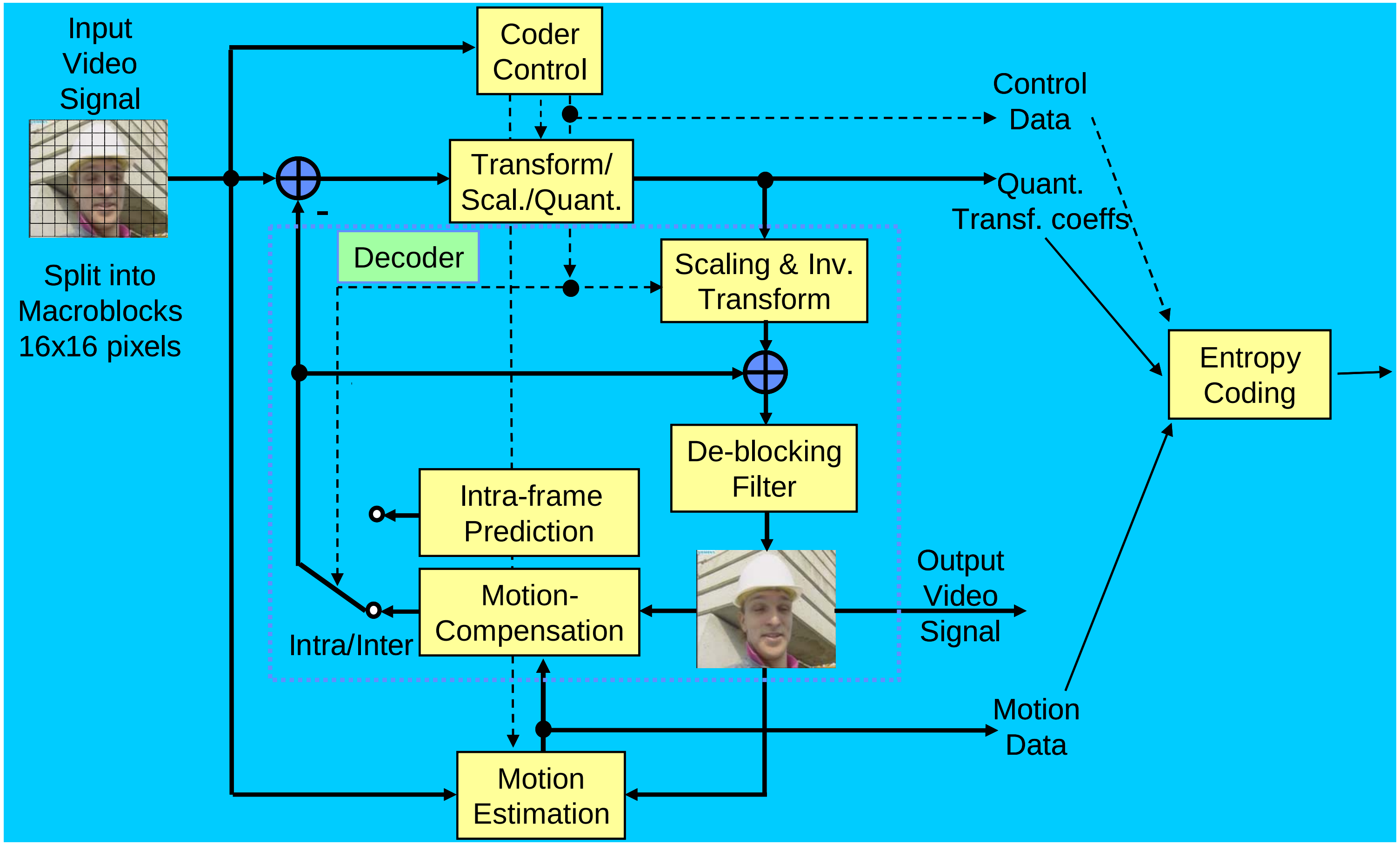


Conventional Predictive Coding

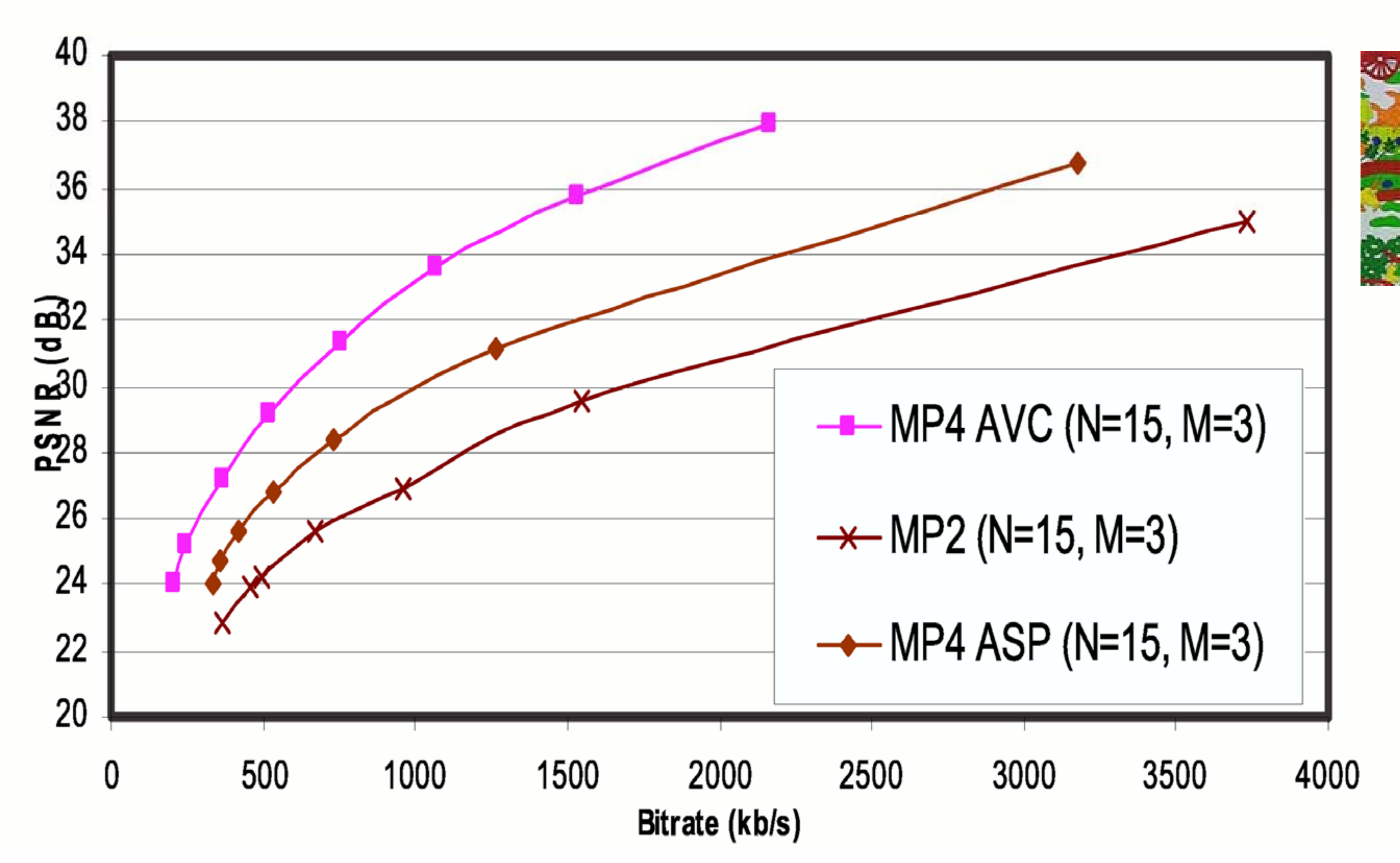
- Exploitation of the source correlation at the encoder

- 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

State-of-the-art: H.264/AVC



Performance Evolution – Mobile & Calendar



Technology Trends

- More, more, more…
  - Display resolution is increasing (Digital Cinema: 4K x 2K)
  - Ultra High Definition under development (8K x 4K)
  - HD resolutions in the mobile world (720p, then 1080p)
  - High dynamic ranges (up to 14 bits per component)
  - 4:4:4 color sampling
  - 3D, multi-view, free viewpoint

Cameras and displays seem to be ready for this 'jump forward'
However, the transmission infrastructure does not seem to be able to accommodate the associated (coded) rates !

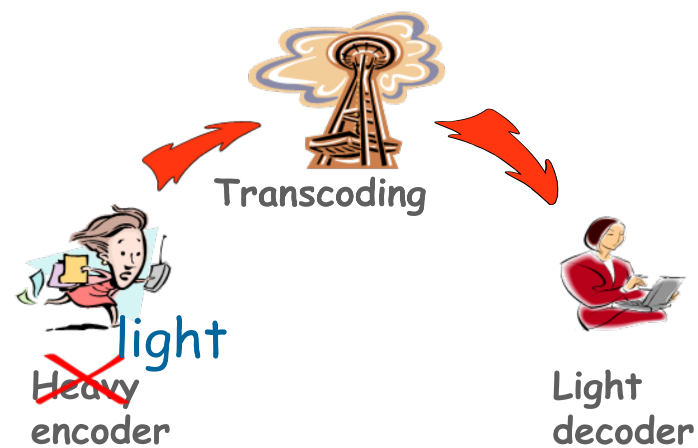
Next Generation Standard by MPEG and VCEG

- High Efficiency Video Coding (HEVC)
  - A new generation of video compression technology is needed to meet demands in bitrate
  - Compression capabilities that are clearly higher than the existing H.264/AVC High profile
  - Call for Proposals issues in January 2010, 27 proposals received, extensive subjective tests conducted by April 2010, now starting a collaborative phase

New Class of Up-Link Applications

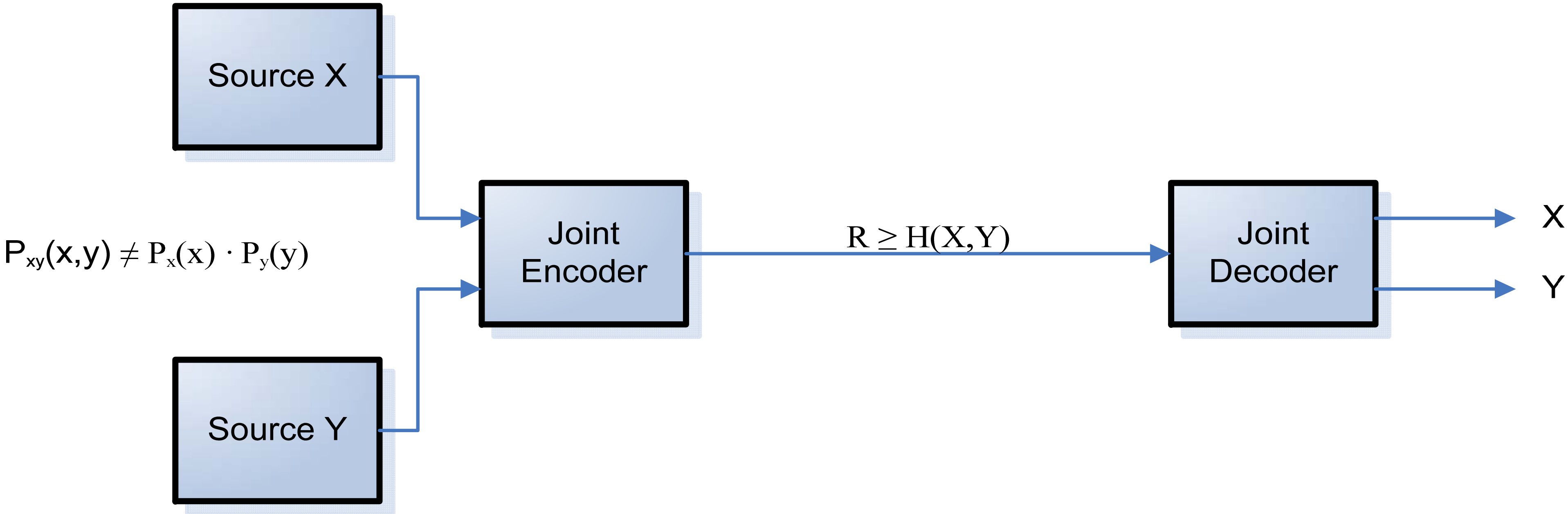
- High-resolution wireless digital video cameras
- Multimedia smartphones and PDA’s
- Low-power video sensors and surveillance cameras

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

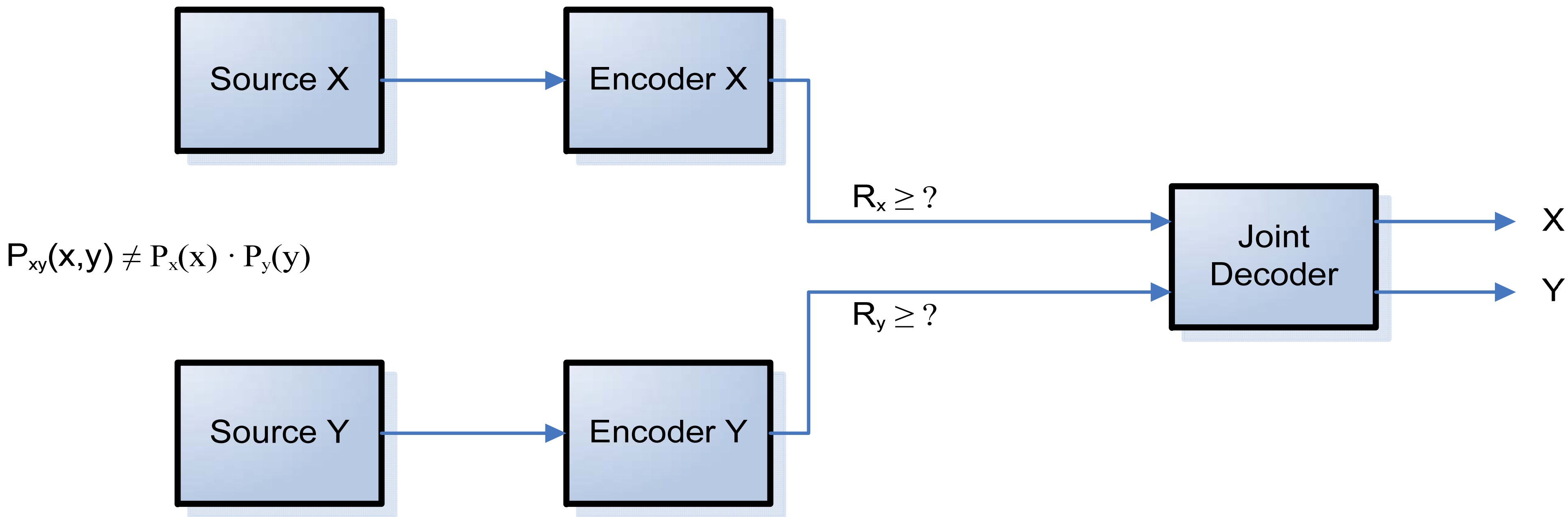


Distributed Video Coding
Theoretical Foundations

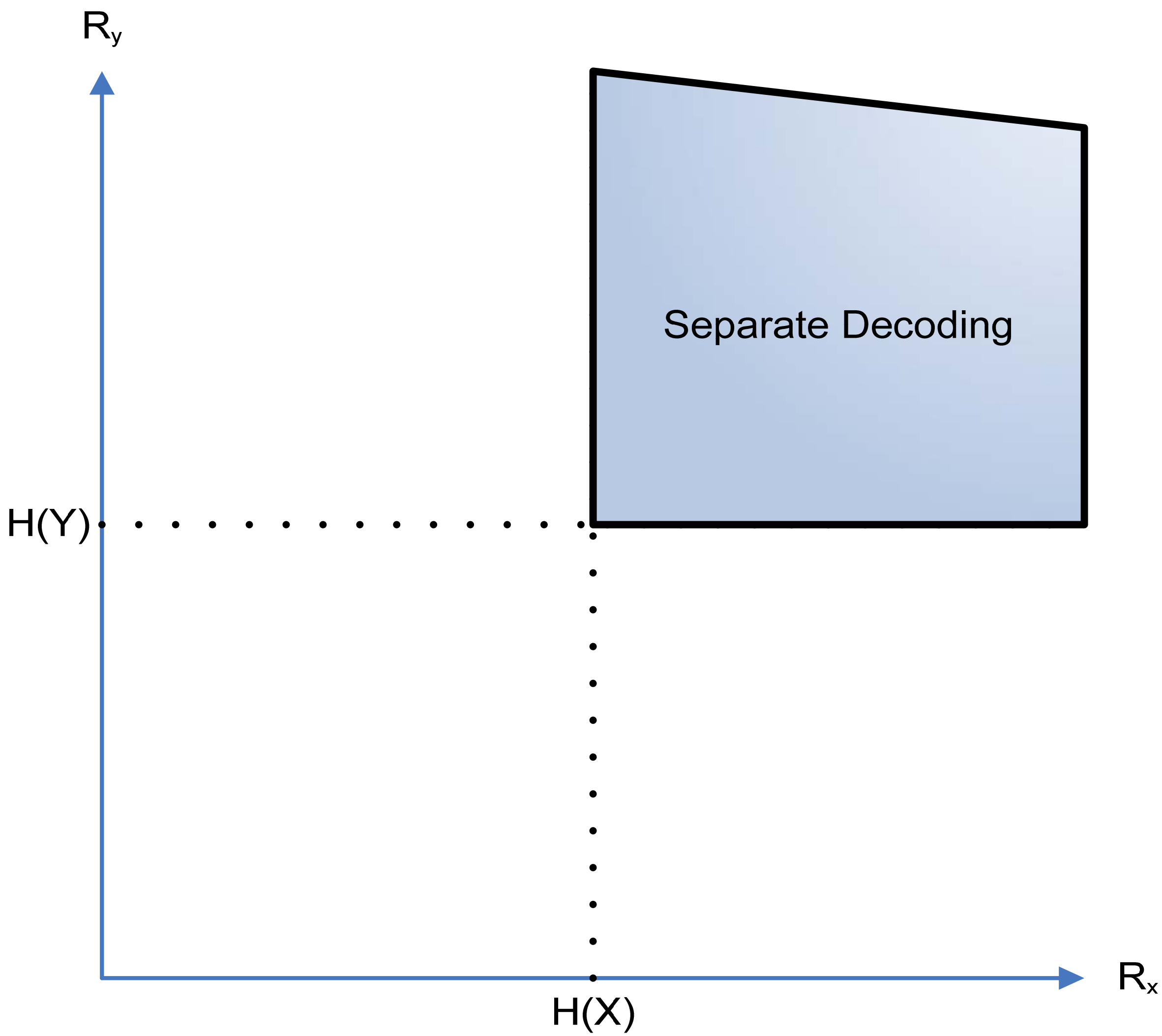
Coding of Dependent Sources



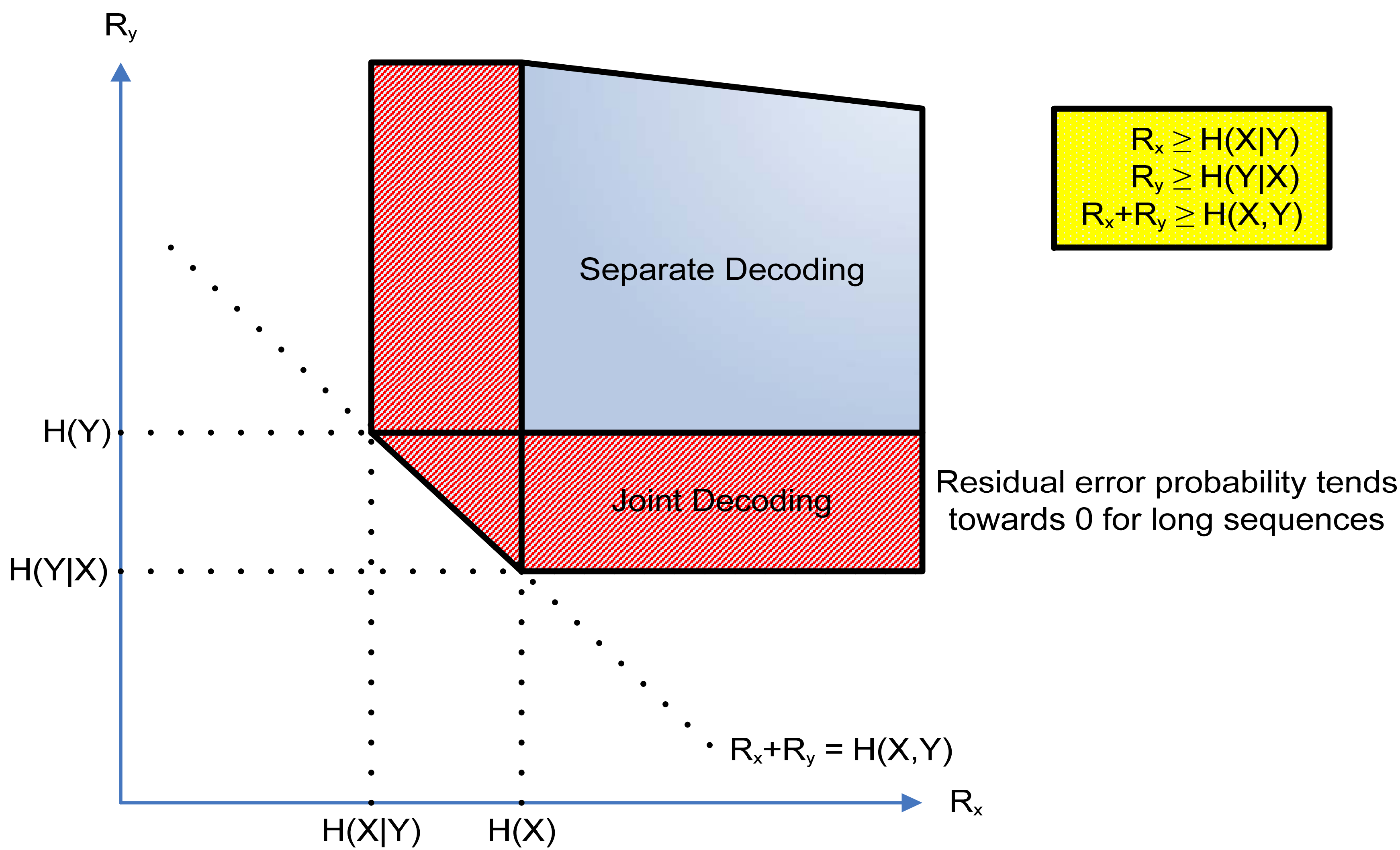
Distributed Coding of Dependent Sources



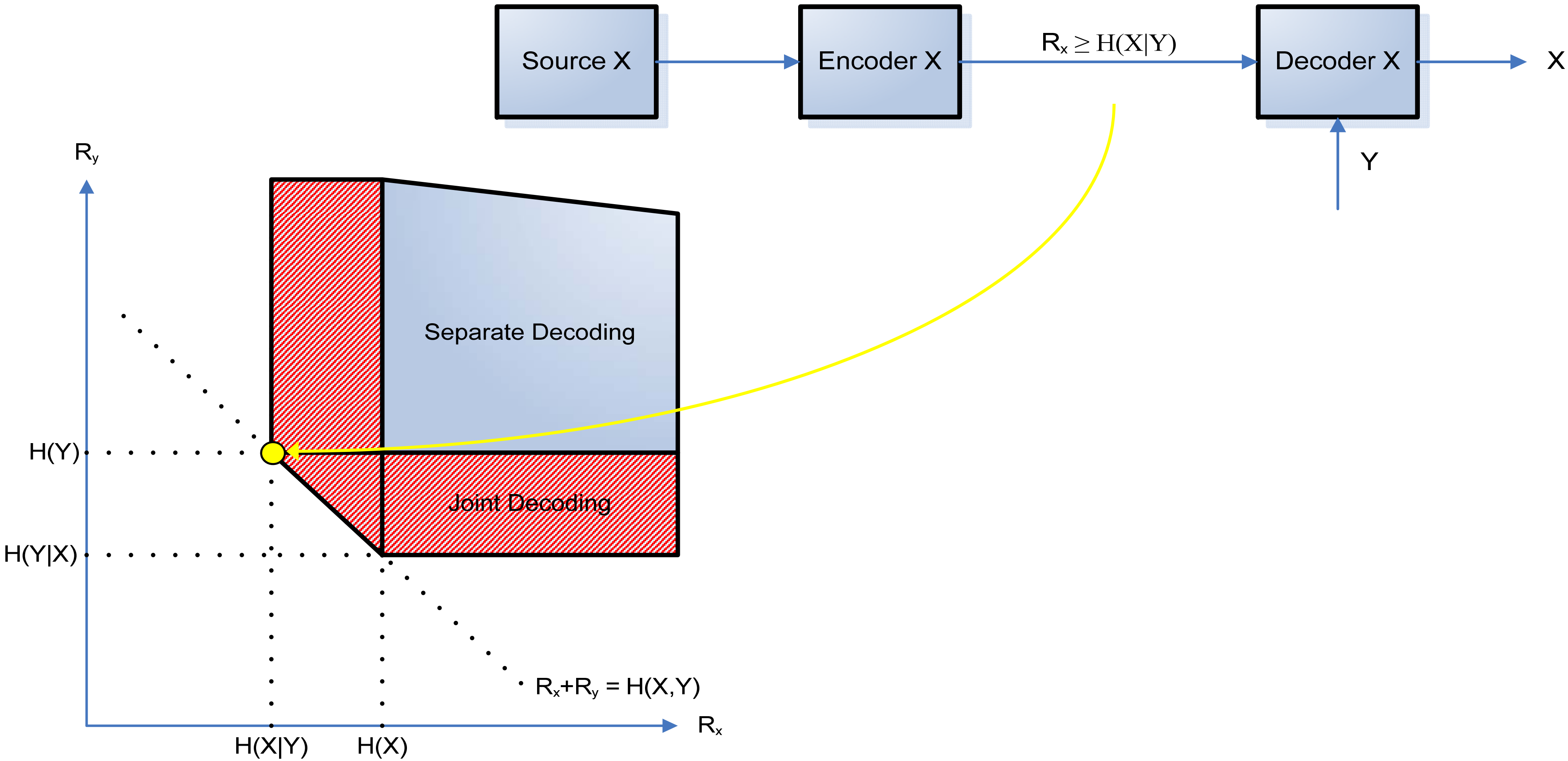
Slepian-Wolf Theorem



Slepian-Wolf Theorem

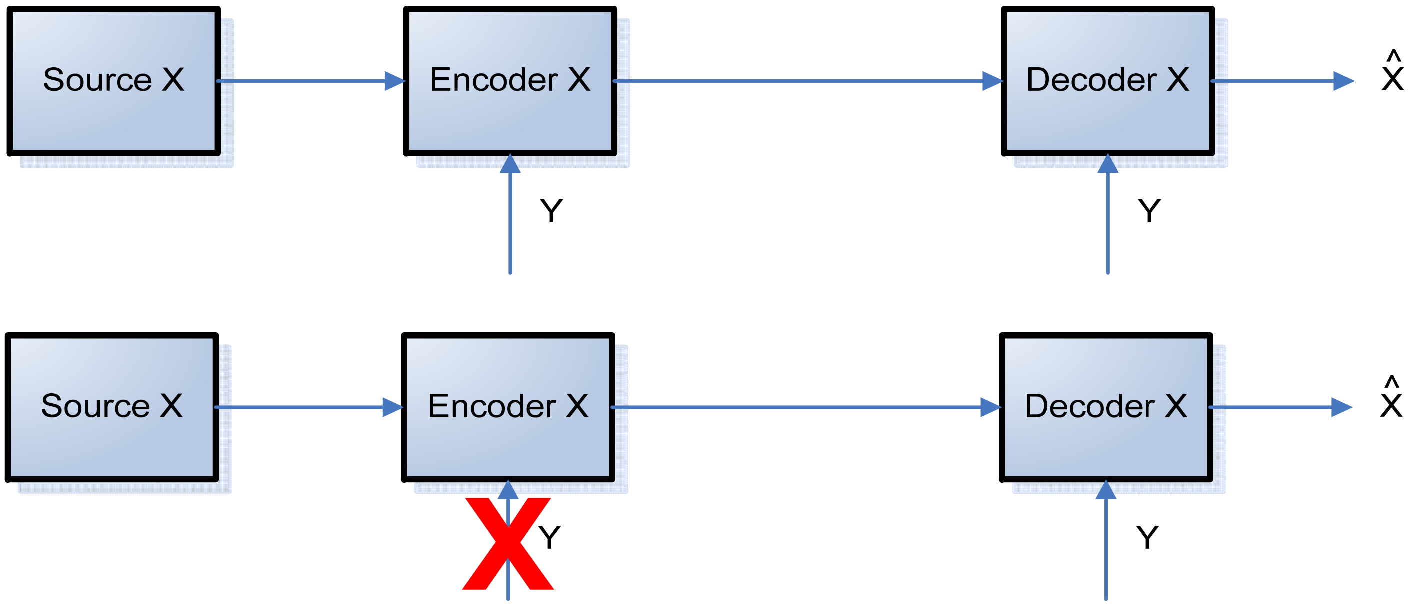


Slepian-Wolf with Decoder Side Information



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

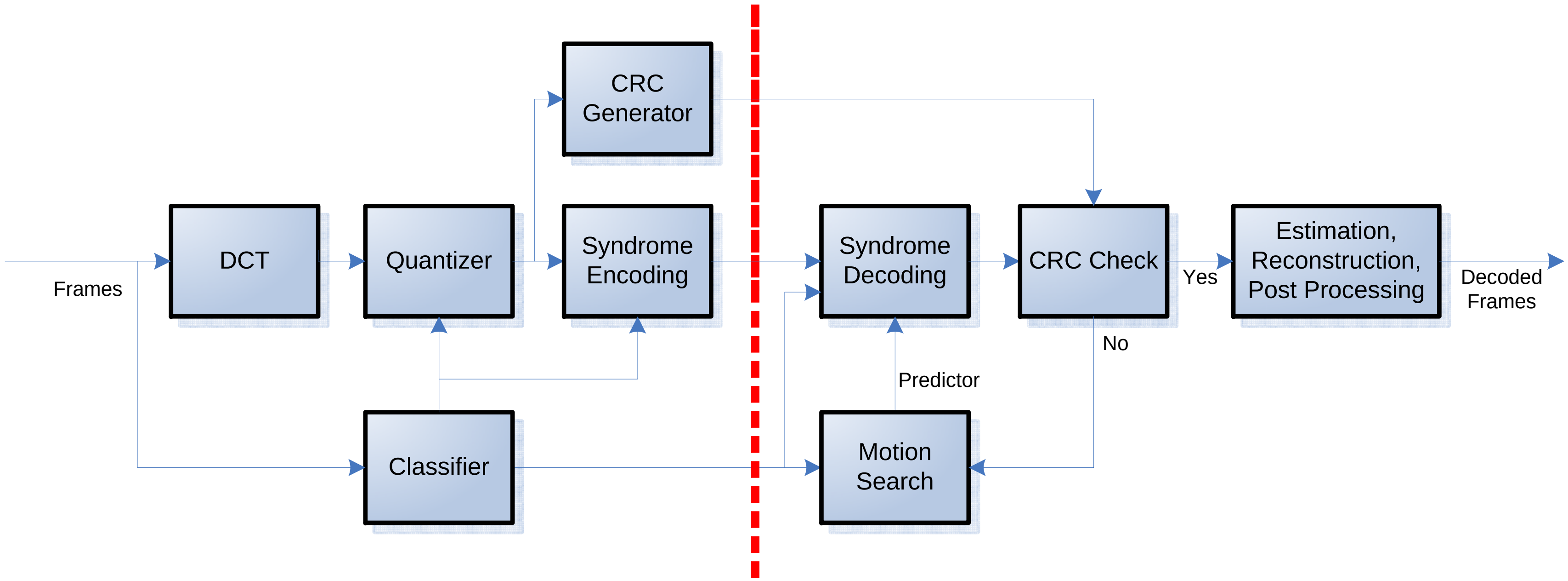
- Opportunity to re-invent video coding
  - Forget the past deterministic approach
  - Adopt a new statistical mind set

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

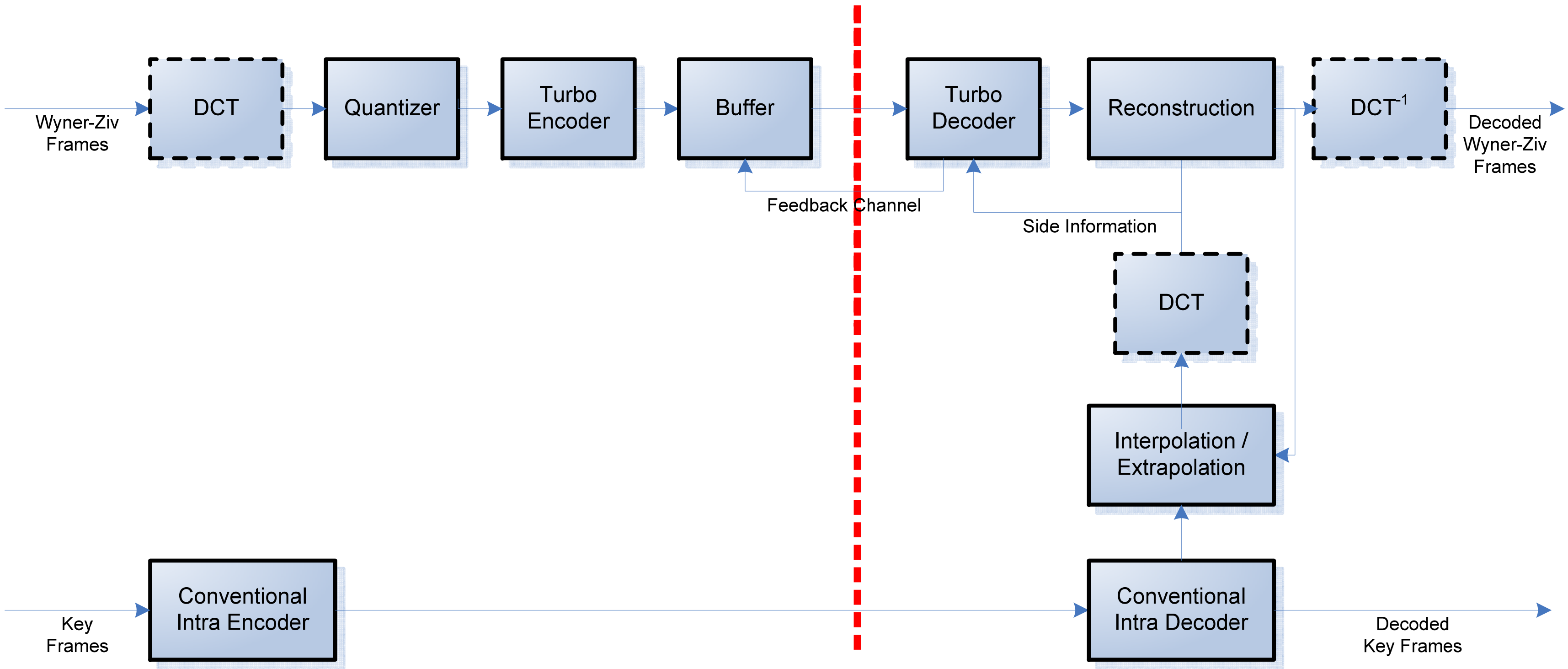
- Challenge: achieve state-of-the-art coding performance

Early DVC Architectures

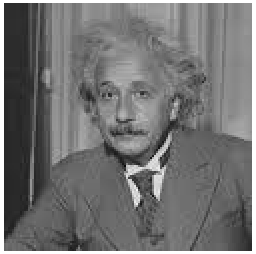
PRISM Architecture



Stanford Architecture



Current Research Topics



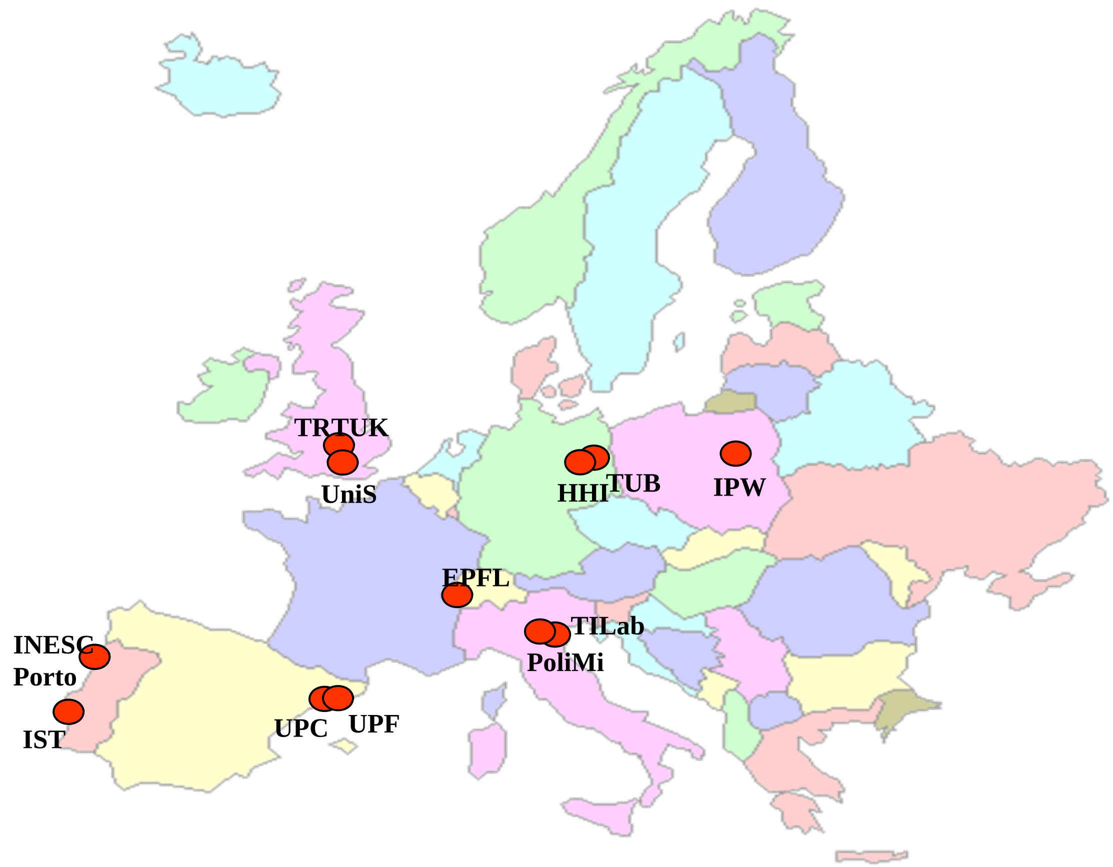
If we knew what it was we were doing,
it would not be called research, would it?
Albert Einstein

VISNET II DVC Codec

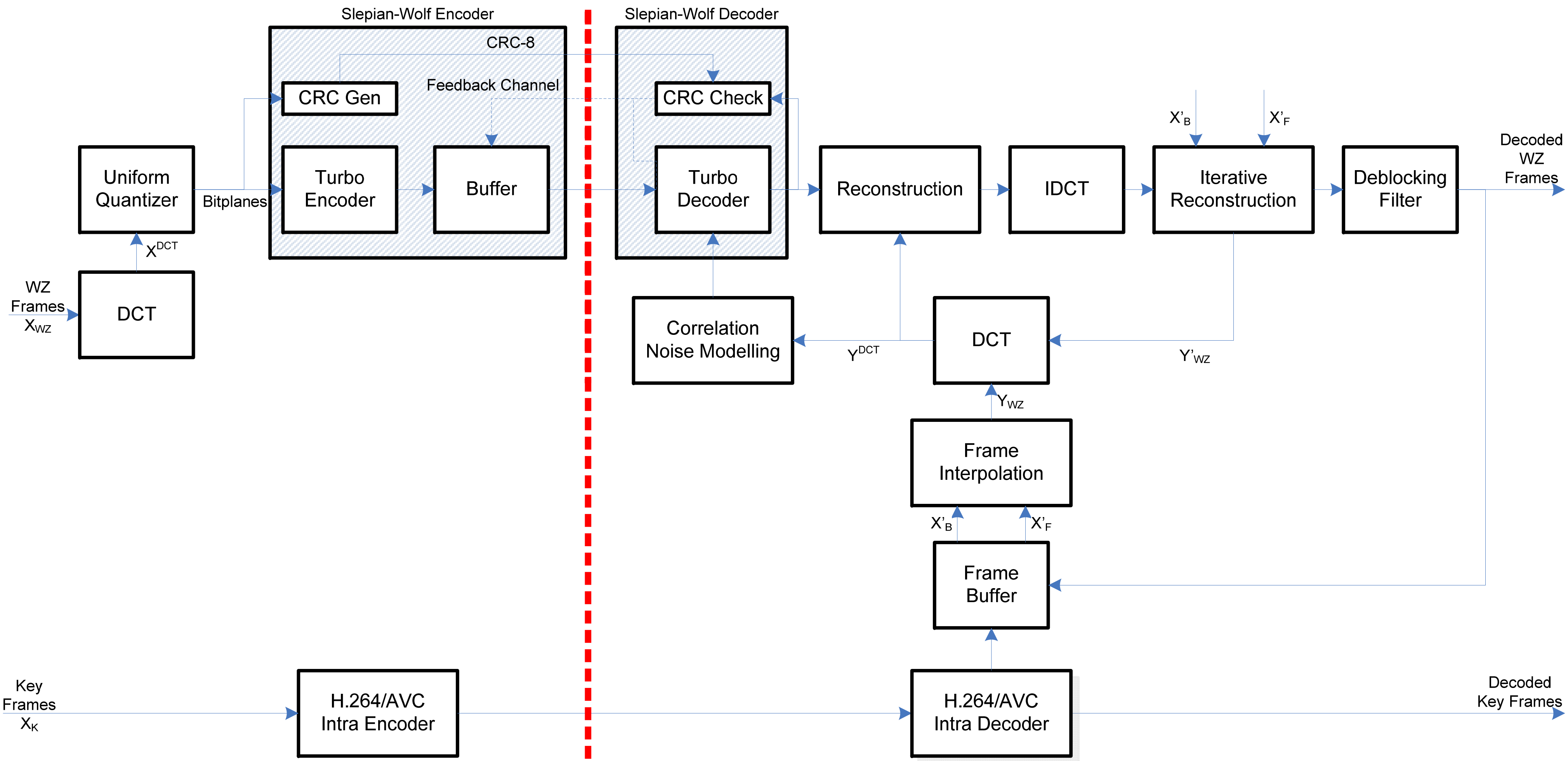
- Developed within the European Network of Excellence VISNET II
- Based on the Stanford architecture

VISNET II

Information Society Technologies



VISNET II DVC Architecture



Encoder

- Transform
  - Integer 4x4 DCT, coefficients are grouped together to form bands
- Quantization
  - Uniform quantizer: the number of levels depends on the band
  - Quantized symbols are then split into bitplanes
- Turbo Encoding
  - Bitplanes are independently encoded, starting with the most significant bitplane array
  - Parity information is transmitted upon decoder request through the feedback channel
- CRC
  - 8 bit CRC for each bitplane and sent to the decoder

Decoder

- Frame interpolation
  - Motion compensated frame interpolation, using the previous and next closer reference frames
  - Hierarchical motion estimation and spatial motion smoothing
- SI transform
  - Integer 4x4 block-based DCT is applied on the SI
- Correlation noise modeling
  - Residual statistics between WZ and SI is modeled by a Laplacian distribution, the parameter is estimated on-line at the coefficient granularity level
- Turbo decoding
  - Turbo decoding each band, bitplane by bitplane from most to least significant

Decoder

- Request stopping criterion
  - Decide whether more parity bits are needed for a bitplane
  - Decoding is considered successful if estimated error probability is smaller than 10^{-3}
- CRC checking
  - To detect residual errors when the request stopping criterion is fulfilled
- Iterative Reconstruction
  - Partially decoded frame is used to re-generate SI and iterate the reconstruction
- Deblocking filter
  - To improve both subjective and objective quality

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)

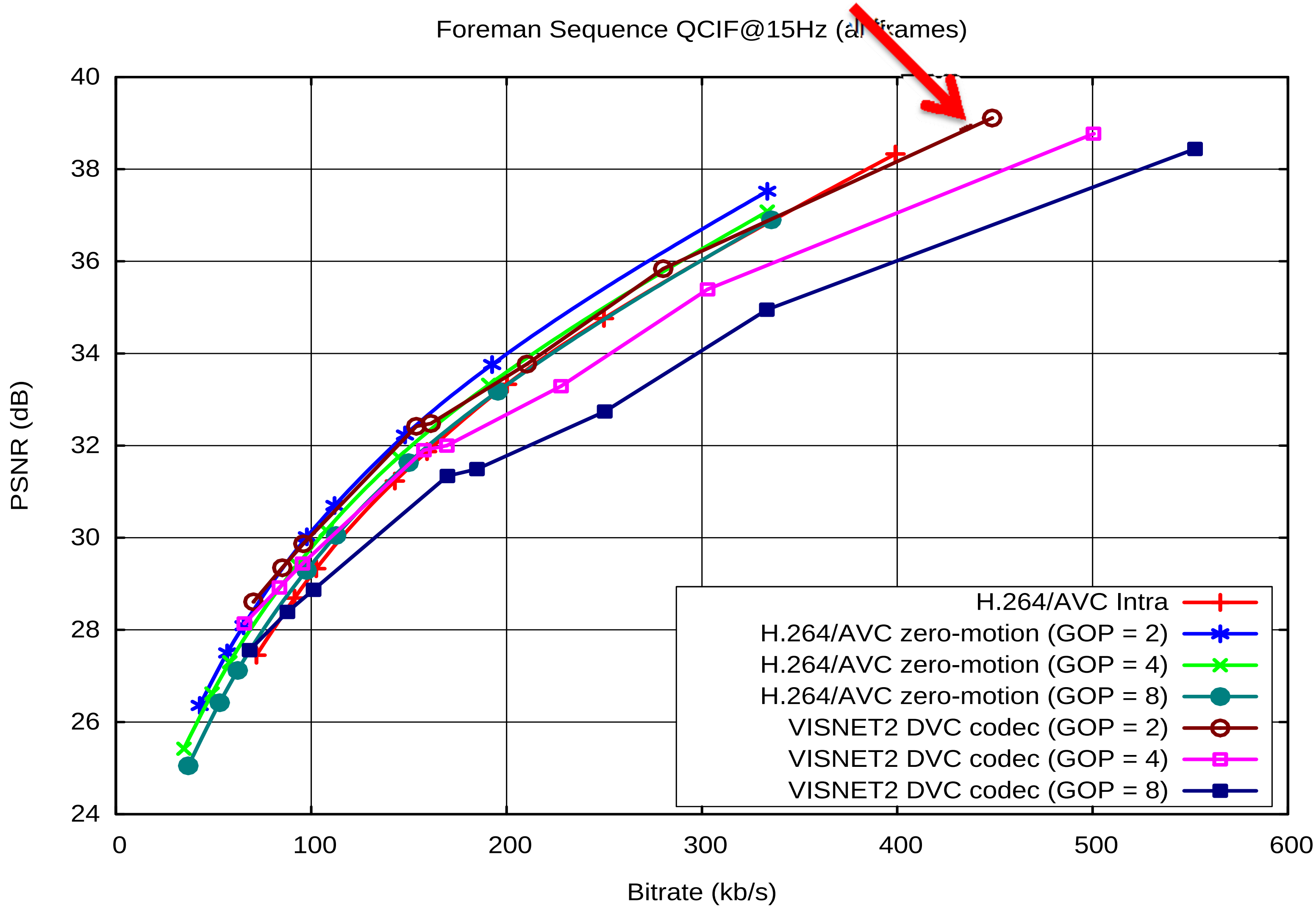
64	32	16	8
32	16	8	4
16	8	4	4
8	4	4	0

(g)

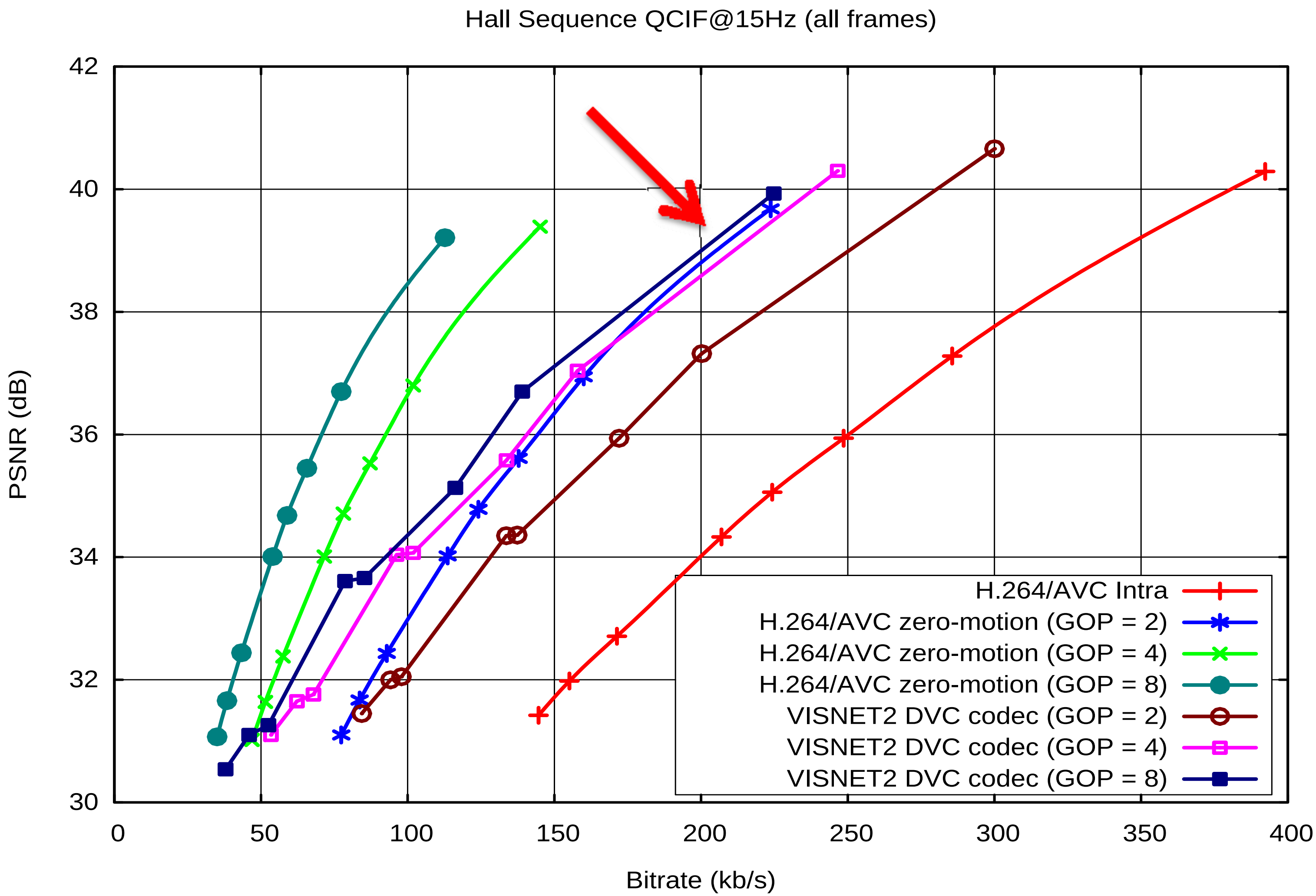
128	64	32	16
64	32	16	8
32	16	8	4
16	8	4	0

(h)

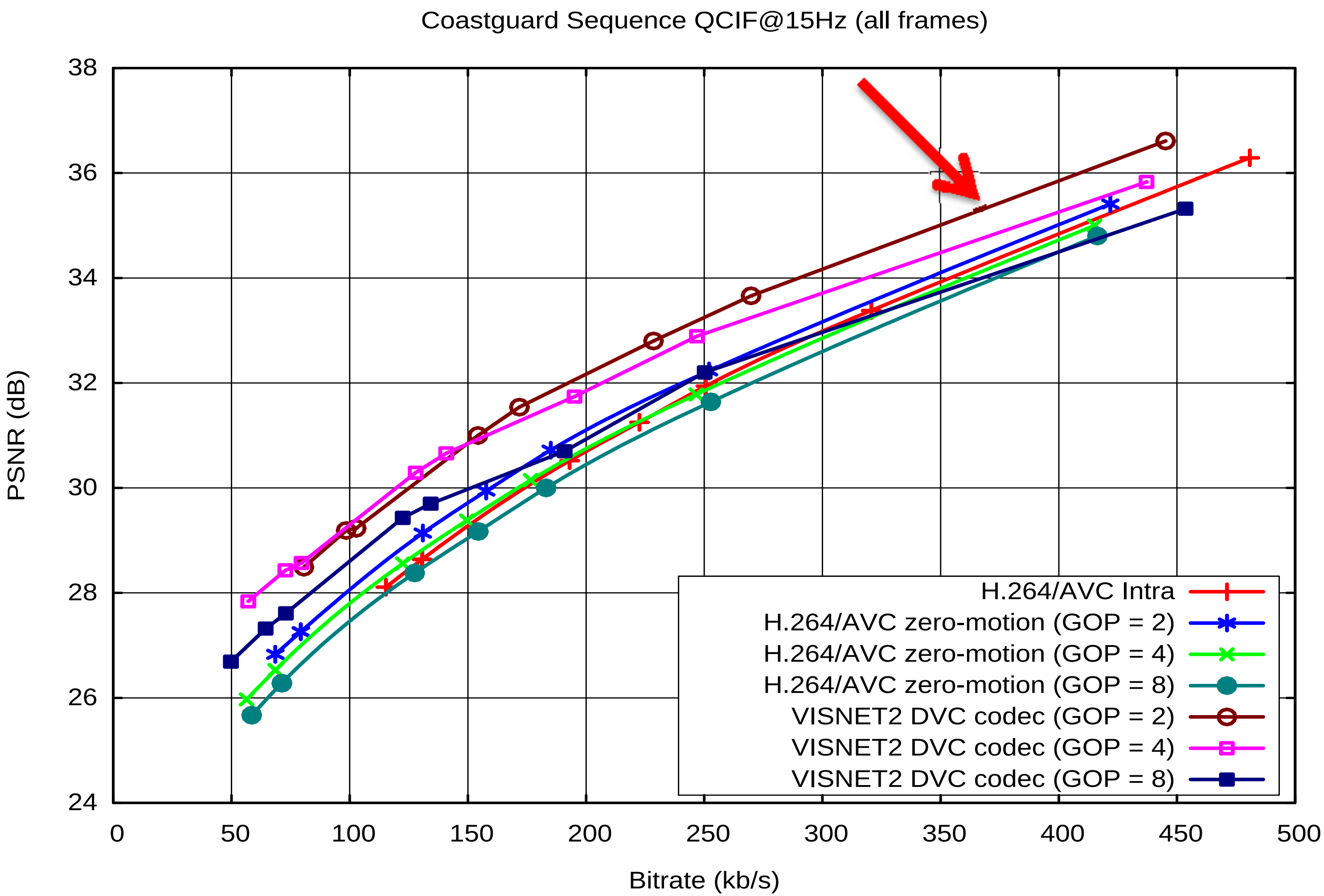
VISNET II DVC versus H.264/AVC: Foreman



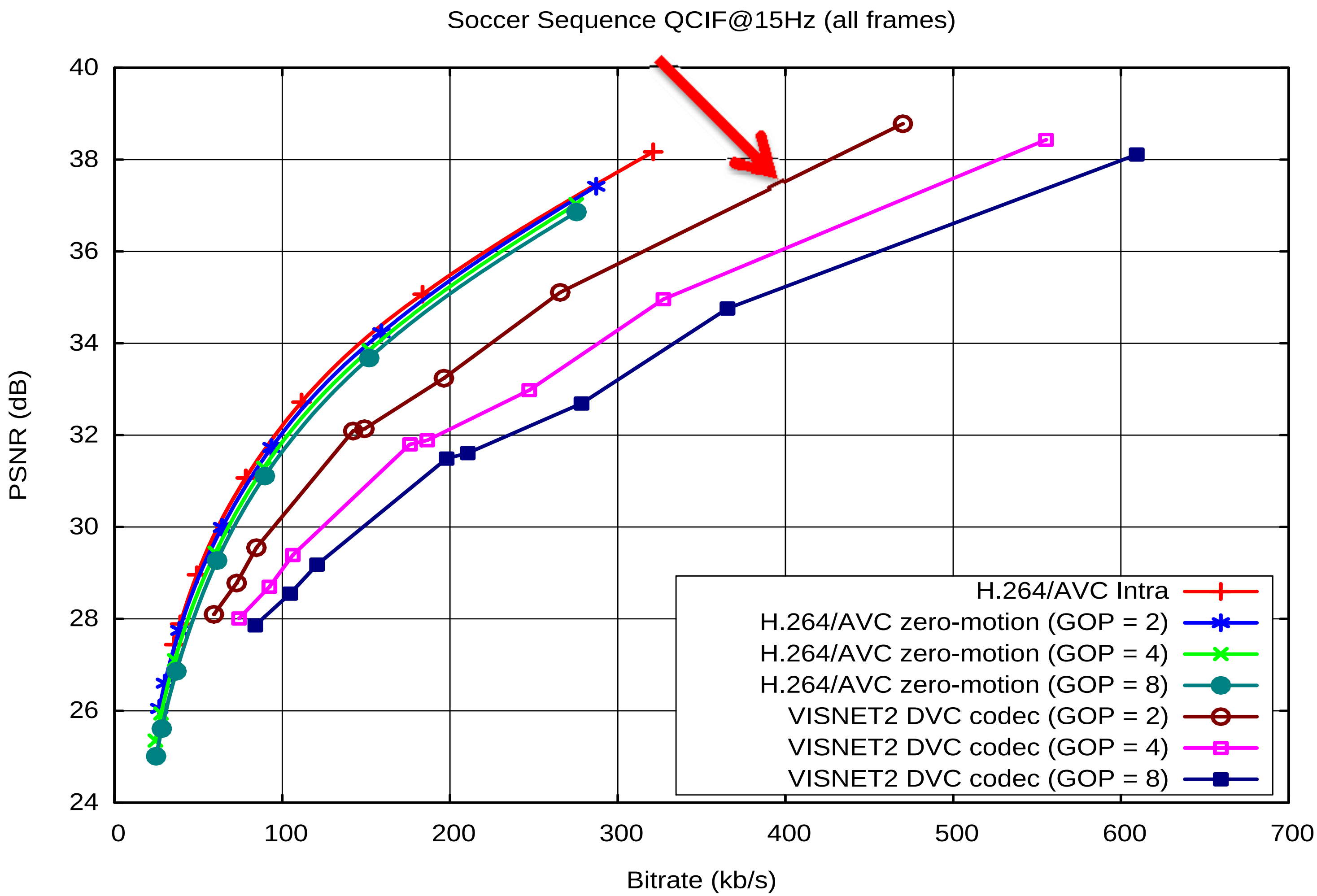
VISNET II DVC versus H.264/AVC: Hall Monitor



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

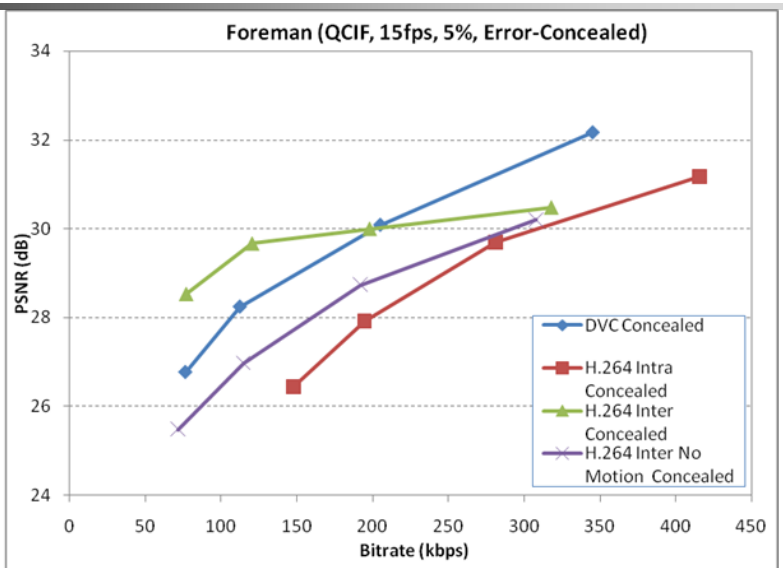
Robust Transmission

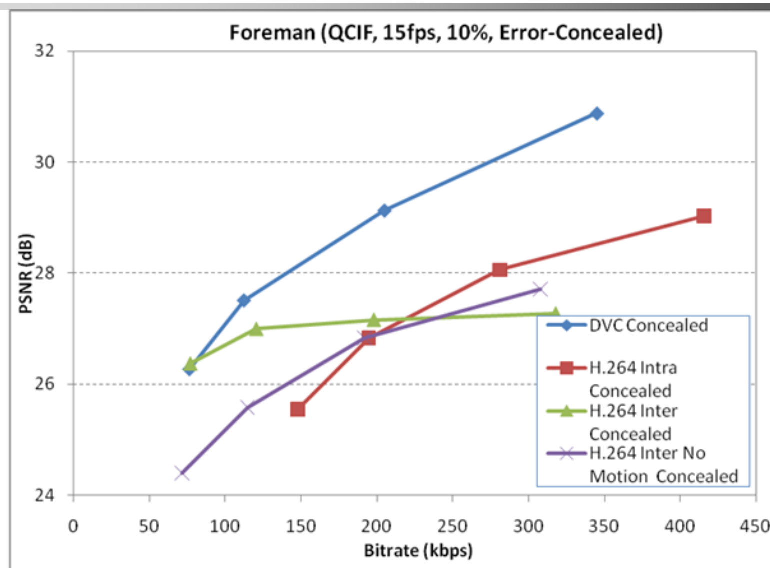
- Appealing for transmission over error-prone channels
  - Statistical framework rather than a deterministic approach
  - Absence of a prediction loop in the codec
- Decoding is successful, even in the presence of transmission errors, as long as the SI is within the noise margin of the encoded parity 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

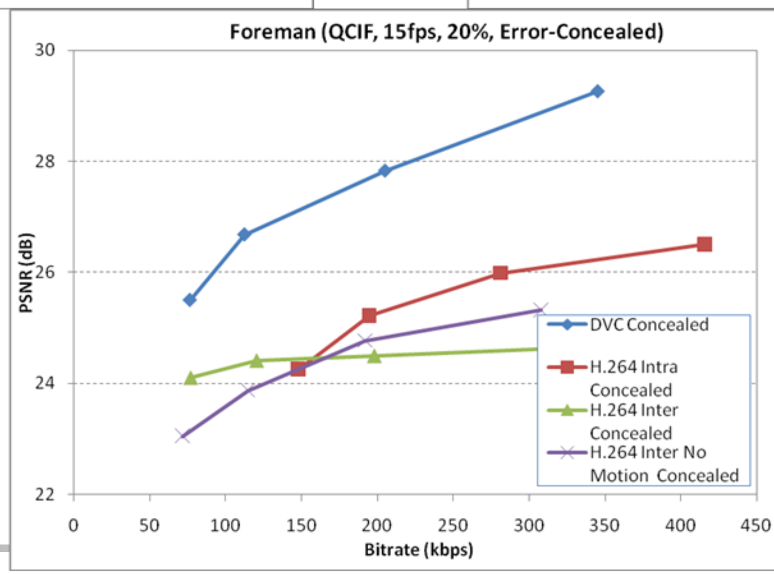
Robust Transmission

- 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

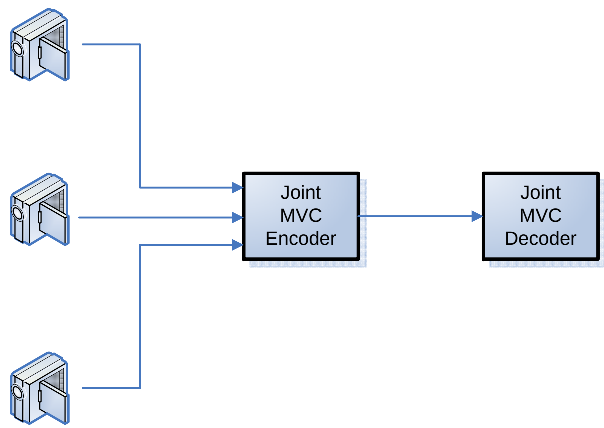
Foreman, no feedback channel



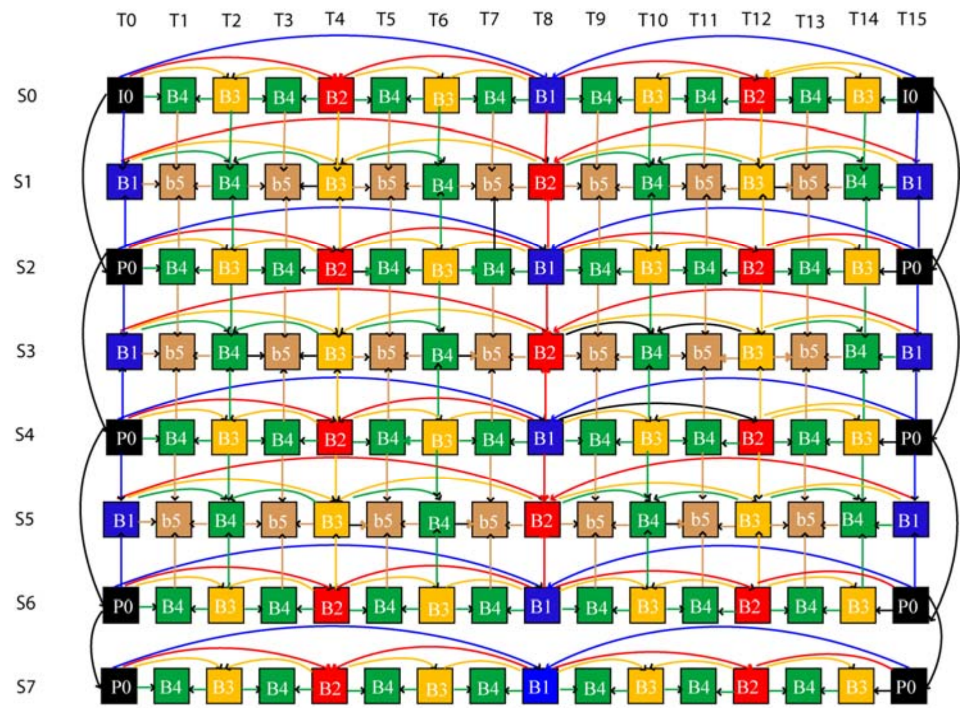




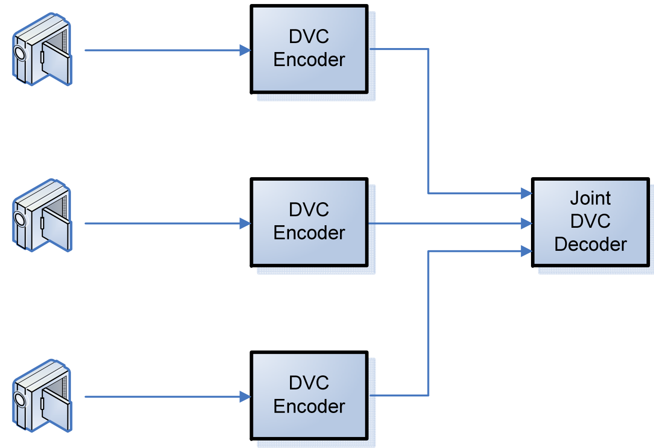
Multi-View Video Coding



- 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

Inter-View Side Information

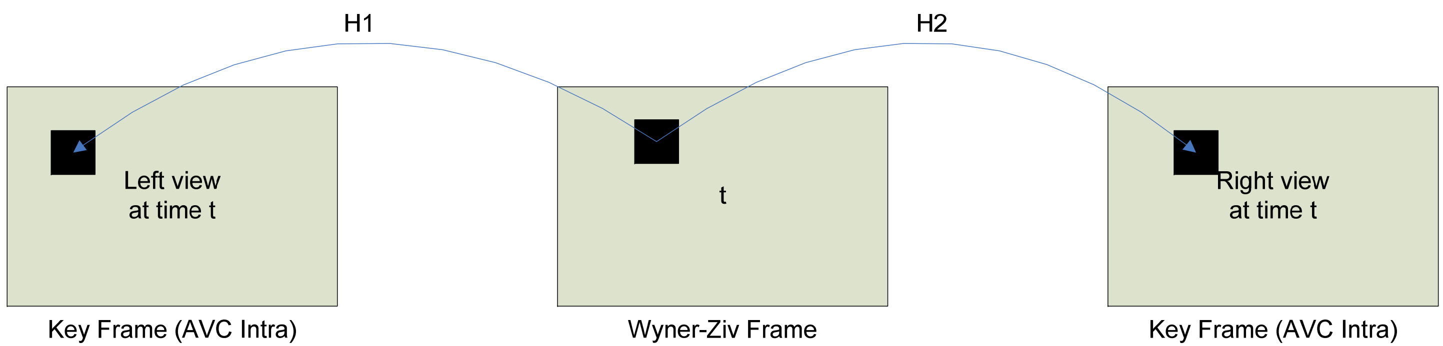
- Disparity Compensation View Prediction (DCVP)
  - Straightforward extension of MCTI
  - Disparity vectors are estimated between views
  - Interpolation at mid-point to generate SI

Inter-View Temporal Side Information

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

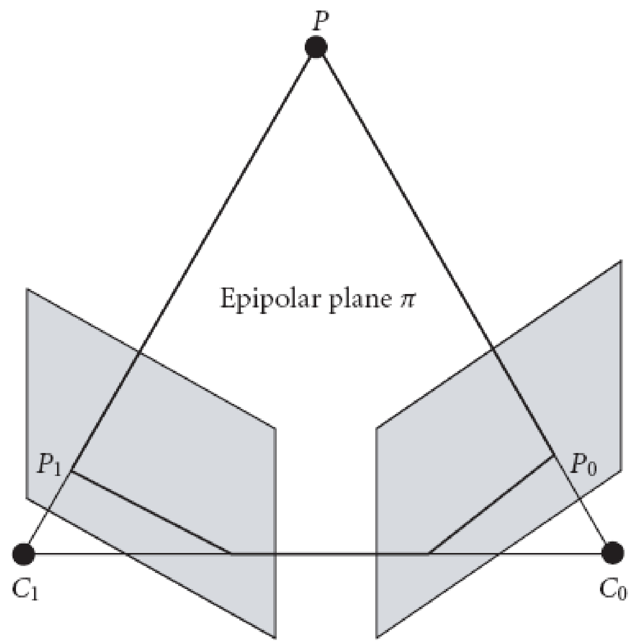
$$x_i' = \frac{a_0 + a_2 x_i + a_3 y_i}{a_6 x_i + a_7 y_i + 1}$$

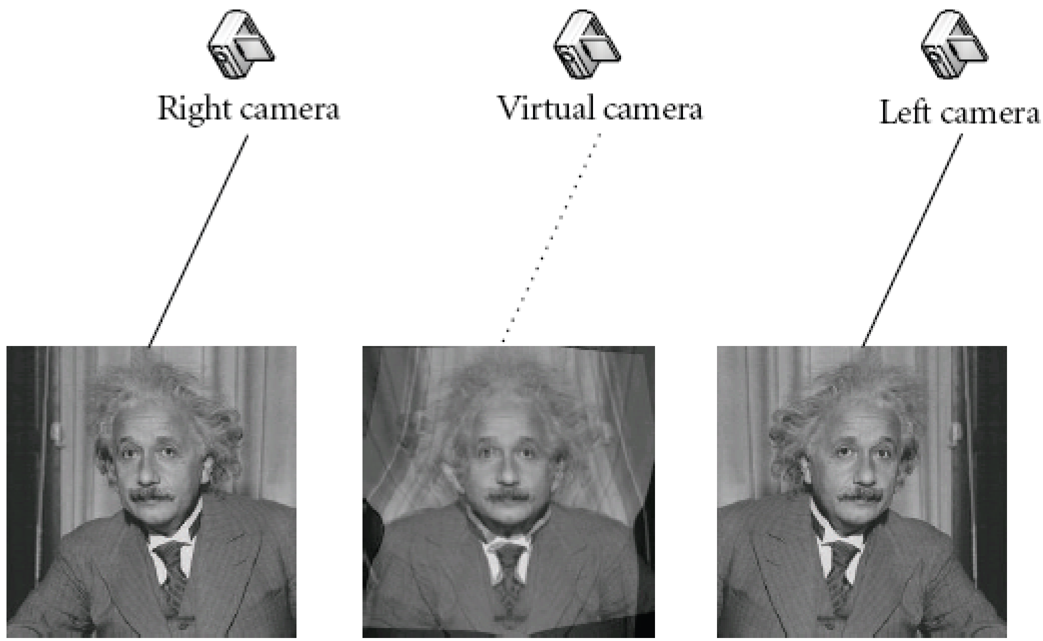
$$y_i' = \frac{a_1 + a_4 x_i + a_5 y_i}{a_6 x_i + a_7 y_i + 1}$$



Inter-View Temporal 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

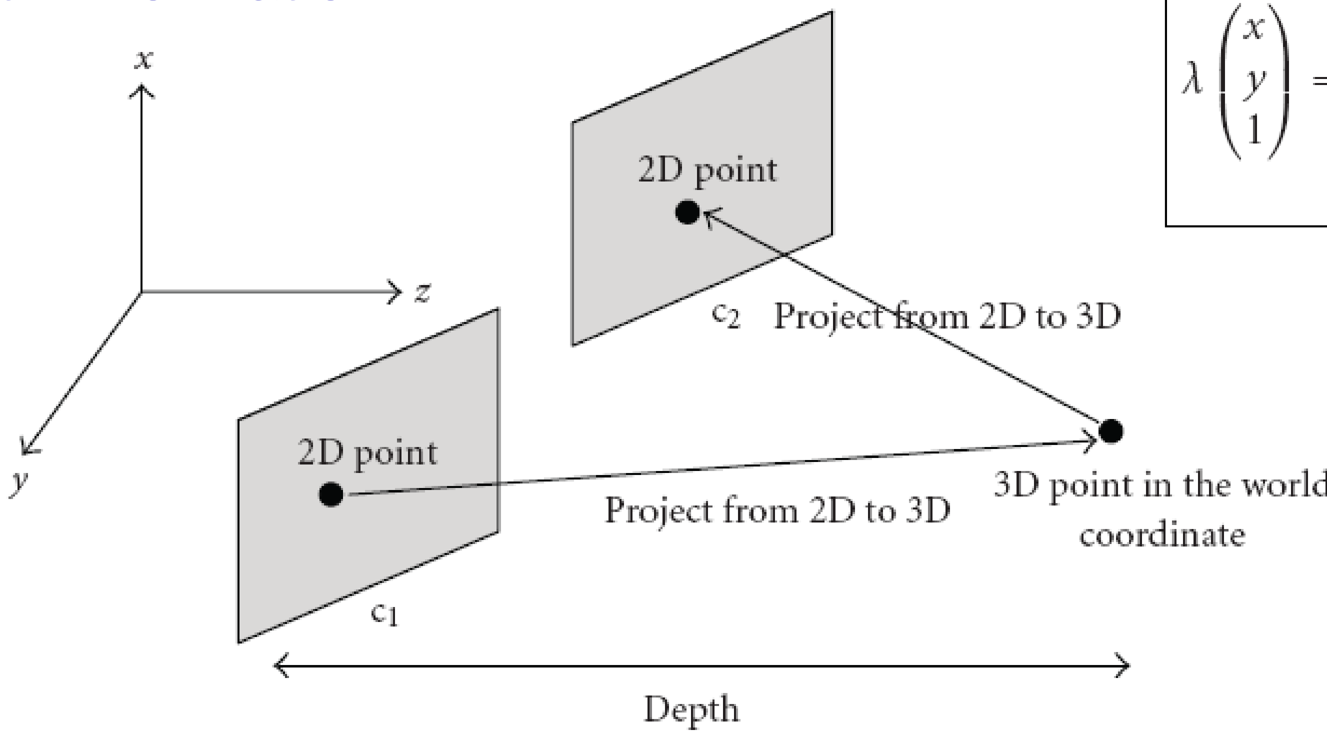




Inter-View Temporal Side Information

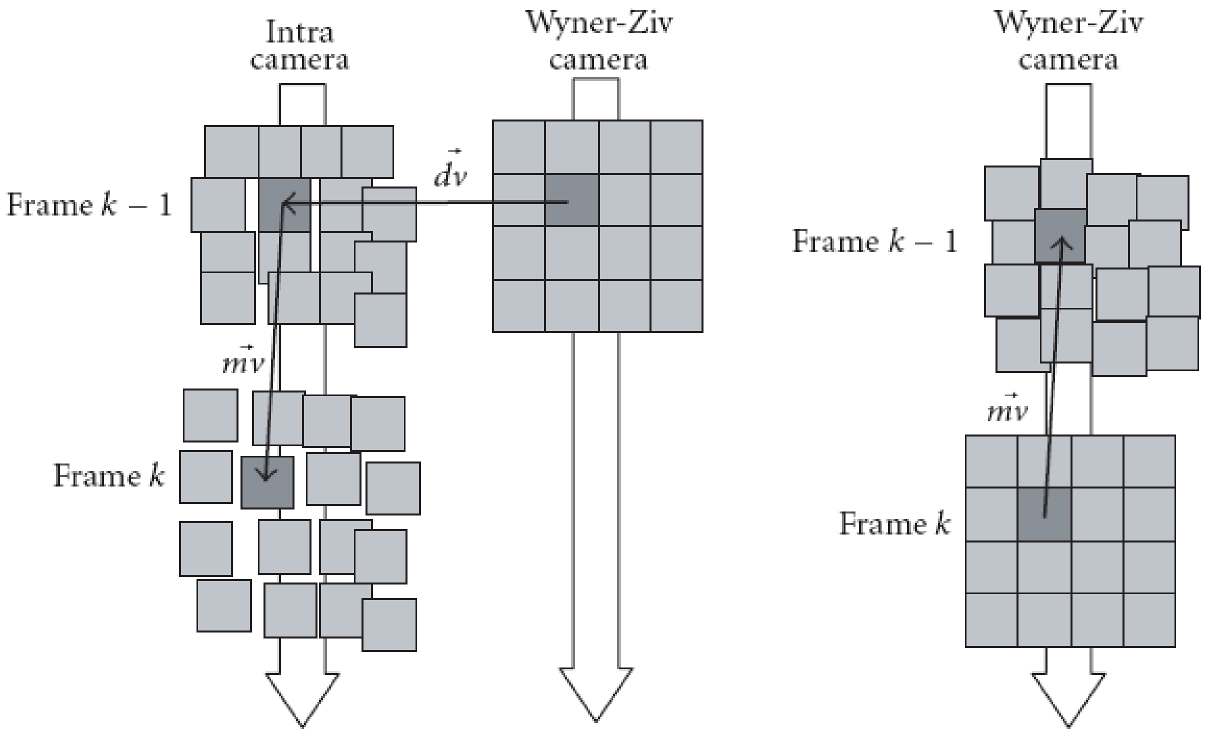
- View Synthesis Prediction (VSP)
  - Camera calibration
  - Intrinsic and extrinsic camera parameters
  - Depth information

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



Inter-View Temporal Side Information

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



Summary

- DVC is consistently better than H.264/AVC Intra, notably for video surveillance sequences
- For sequences with simple or regular motion (e.g. Coastguard), DVC may even outperform H.264/AVC No Motion
- WZ video encoding complexity is always much lower than the H.264/AVC Intra encoding complexity.
- Appealling for robust video transmission over error-prone channels
- Offer interesting architectural advantage for multi-view video coding
- But there is still a significant RD performance gap with predictive coding (full-fledge H.264/AVC)

Outlook

Disclaimer

Prediction is very difficult, especially about the future

Niels Bohr

Outlook

- Conceptual benefit
- RD performance status quo
- Complexity benefit
- Error resilience benefit
- Multi-view benefit
- Beyond video coding

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		

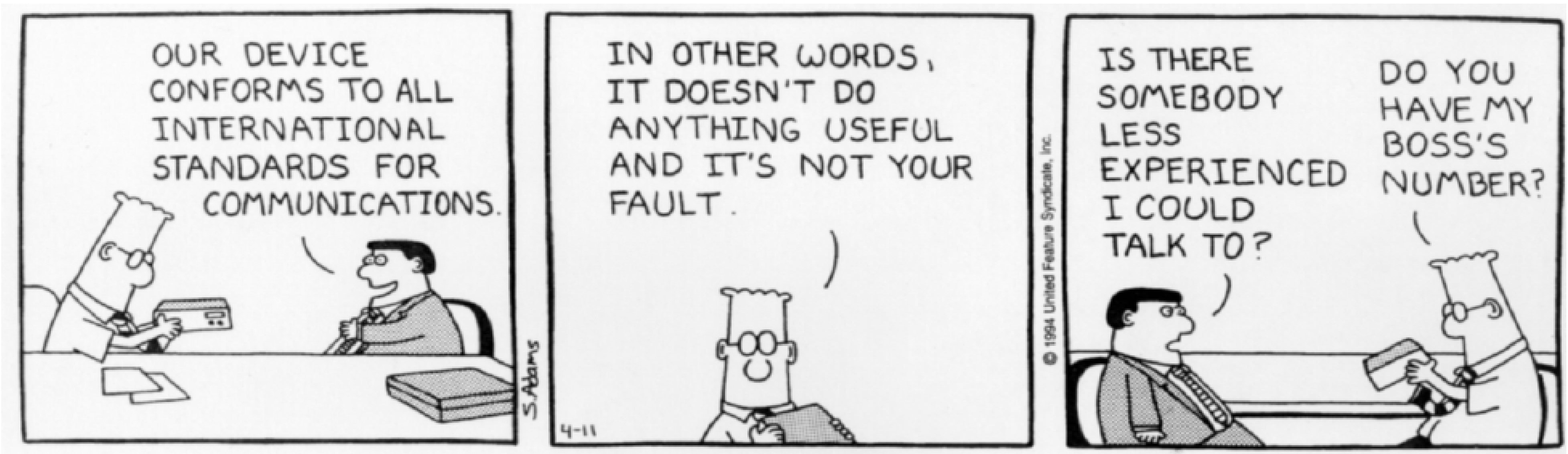
Combined Predictive and Distributed Coding ?

- Moore's law
  - Driving force of technological changes
  - More complexity may be increasingly acceptable if worthy
- Predictive video coding
  - Complex encoder and simple decoder
- Distributed video coding
  - Simple encoder and complex decoder

- Improved RD performance with a complex encoder – complex decoder design combining predictive and distributed video coding principles ?

Trends and Way Forward in Video Compression

- **Exploiting the Status Quo** - More efficient H.264/AVC non-normative coding tools
- **Smooth Approach** - Adding more efficient coding tools to the predictive (H.264/AVC) video coding architecture
  - MPEG/VCEG High Efficiency Video Coding (HEVC)
- **Less Smooth Approach** - More substantially changing the predictive (H.264/AVC) video coding architecture
  - Context Adaptive Coding or Metadata-based Coding
  - Model-Based and Inpainting-based Texture Coding
  - Advanced Transforms
- **Disruptive Approach** - Adopting a new video coding approach based on new coding principles and tools
  - Combined Predictive-Distributed Coding
  - Human Visual System
  - Compressive Sensing


OUR DEVICE CONFORMS TO ALL INTERNATIONAL STANDARDS FOR COMMUNICATIONS.
S. Adams
4-11
IN OTHER WORDS, IT DOESN'T DO ANYTHING USEFUL AND IT'S NOT YOUR FAULT.
© 1994 United Feature Syndicate, Inc.
IS THERE SOMEBODY LESS EXPERIENCED I COULD TALK TO?
DO YOU HAVE MY BOSS'S NUMBER?

Acknowledgement

- Many thanks to the individuals who contributed directly or indirectly to this presentation
  - T. Ebrahimi, M. Ouaret, S. Ye, J. Ascenso, C. Brites, A. Fernando, W. Gao, F. Pereira, S. Tubaro, A. Vetro, M. Badem, R. Martins, J. Pedro, R. Weerakkody
  - And many more…

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.
- M. Ouaret, F. Dufaux and T. Ebrahimi, “Iterative Multiview Side Information for Enhanced Reconstruction in Distributed Video Coding”, in EURASIP Journal on Image and Video Processing, (Special issue on DVC), vol. 2009, Article ID 591915, doi:10.1155/2009/591915, 2009.
- S. Ye, M. Ouaret, F. Dufaux and T. Ebrahimi, “Improved Side Information Generation for Distributed Video Coding by Exploiting Spatial and Temporal Correlations”, EURASIP Journal on Image and Video Processing, (special issue on DVC), vol. 2009, Article ID 683510, doi:10.1155/2009/683510, 2009.
- 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.

Thank you for your attention !!

Any questions ?