Proactive Design for Multimedia Communication Systems with Resource and Information Exchanges

Mihaela van der Schaar

Assistant professor

University of California Davis

Overview

- Challenges for wireless multimedia -> Research focus

- Scalable video coding and processing
- Cross-layer optimized wireless multimedia
- Proactive collaboration for wireless multimedia

- Research directions beyond this talk
- A new chance to reinvent multimedia compression, processing, communication & system design!

Wireless Multimedia Applications

Wireless: 802.11 WLANs,
Opportunistic SAR

- Entertainment
- Emergency services
- Surveillance
- Telemedicine
- Videoconferencing
- Remote teaching and training
- Augmented reality
- Distributed gaming

Hard delay constraints!
High bandwidth!
Loss tolerant!

IN-HOME STREAMING

STARBUCKS

WAYPORT

MEETINGS

4 Challenges

Challenges

Wireless networks provide limited QoS for *multimedia applications*

Dynamic QoS requirements

- application constraints (delay, rates) and characteristics (codec used,...)
- multimedia traffic characteristics
- usage scenarios
- user preferences

Challenges

Wireless networks provide limited QoS for *multimedia applications*

Dynamic QoS requirements

- application constraints (delay, rates) and characteristics (codec used,…)
- multimedia traffic characteristics
- usage scenarios
- user preferences

Wireless stations (WSTA) experience time-varying channel conditions

WSTA adopt different cross-layer strategies

WSTA transmission strategy influences the network dynamics

Tradeoff between fairness and efficiency

Challenges

Wireless networks provide limited QoS for *multimedia applications*

Dynamic QoS requirements

- application constraints (delay, rates) and characteristics (codec used,…)
- multimedia traffic characteristics
- usage scenarios
- user preferences

Wireless stations (WSTA) experience time-varying channel conditions

WSTA adopt different cross-layer strategies

WSTA transmission strategy influences the network dynamics

Tradeoff between fairness and efficiency

Unique constraints of *multimedia applications* change fundamental communication design principles

Existing theory

- **Information and coding theory** [Shannon and beyond]
  - "ideal" point-to-point communication setting
  - simplistic source models -> not accurate for multimedia coders
  - no delay constraints (concept of "streaming" is absent)
  - no resource management issues and policies such as fairness, etc.
  - system issues neglected – essential for *realistic* wireless multimedia communications

- **Complexity Distortion Theory** [Kolmogorov and beyond]
  - simplistic source models -> not accurate for multimedia coders
  - no consideration of the limitations, capabilities and specific features of (resource-constrained) systems and architectures

- **Optimization, Control, Microeconomic Theory**
- On-line algorithms, competitive analysis etc.

Our research aim

Contribute towards the development of a unifying theory, design and implementation of realistic **multimedia communication systems**

Our research aim

Contribute towards the development of a unifying theory, design and implementation of realistic **multimedia communication systems**

Objectives (NSF Career)

- **Traditional resource management *passively* optimizes resources**
  - Based on fixed, worst-case resource requirements
  - Do not consider the impact on other WSTAs
  - Do not consider realistic multimedia utility-cost functions

- **Proactive collaboration among competing wireless stations**
  - Influence system dynamics through resource/information exchanges
  - Users collaborate and even sacrifice short-term performances, with the incentive that overall system performance can be improved and users' temporary sacrifices will be paid back in a long term

- **Why coopetition for multimedia?**
  - Loss tolerant, delay sensitive, power sensitive

Objectives (cont.)

- Resource exchanges enabled through adapting **cross-layer** transmission strategies of participating stations
    - new cross-layer algorithms that explicitly consider multimedia
- **Rate-Distortion-Power** scalable multimedia coding and streaming
- **Formal Methods for Proactively** Designing and Optimizing Multimedia **Systems**

Our research aim

Contribute towards the development of a unifying theory, design and implementation of realistic **multimedia communication systems**

Objectives (NSF Career)

- **Traditional resource management *passively* optimizes resources**
  - Based on fixed, worst-case resource requirements
  - Do not consider the impact on other WSTAs
  - Do not consider realistic multimedia utility-cost functions

- **Proactive collaboration among competing wireless stations**
  - Influence system dynamics through resource/information exchanges
  - Users collaborate and even sacrifice short-term performances, with the incentive that overall system performance can be improved and users' temporary sacrifices will be paid back in a long term

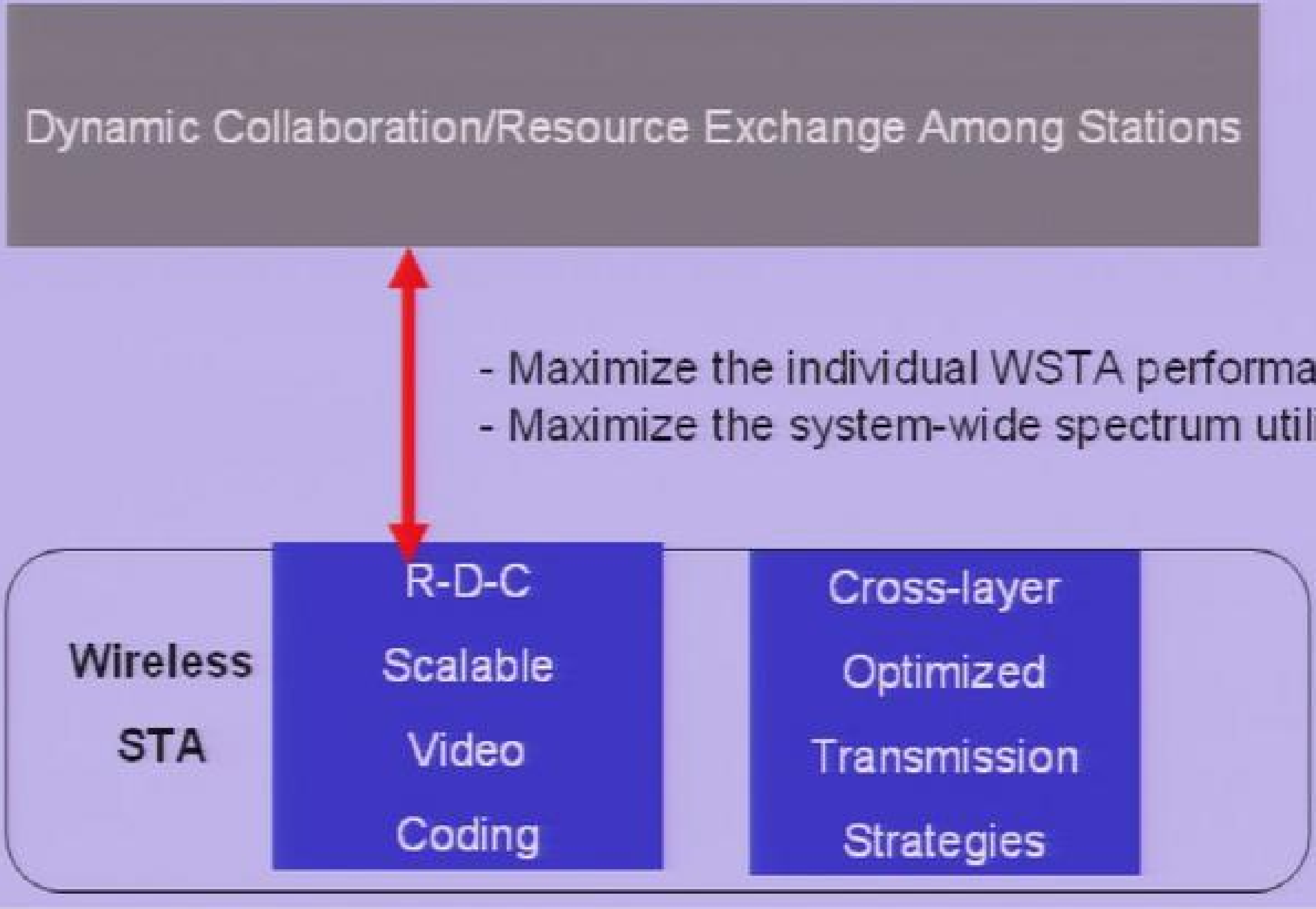
- **Why coopetition for multimedia?**
  - Loss tolerant, delay sensitive, power sensitive

Objectives (cont.)

- Resource exchanges enabled through adapting **cross-layer** transmission strategies of participating stations
  - new cross-layer algorithms that explicitly consider multimedia
- **Rate-Distortion-Power** scalable multimedia coding and streaming
- **Formal Methods for Proactively** Designing and Optimizing Multimedia **Systems**

Collaborative framework for wireless multimedia

Goal: Construct a system, where users can borrow or lend resources from the system/other users, according to their specific utility and resource awareness.



- Maximize the individual WSTA performance and
- Maximize the system-wide spectrum utilization

It all started with Scalable Video ….

Prior Scalable Video Coding standards - Not efficient for heterogeneous IP networks

- Coarse Granularity Scalability (Operate at a discrete set of bit-rates)
- Limited coding efficiency
- Overhead increases with the number of layers

What is important for multimedia communication over IP networks?

- On-the-fly & efficient adaptability to bandwidth variations, QoS levels
- Adaptation to different user & device requirements
- Complexity-scalable encoding/decoding

It all started with Scalable Video ....

Prior Scalable Video Coding standards - Not efficient for heterogeneous IP networks

- Coarse Granularity Scalability (Operate at a discrete set of bit-rates)
- Limited coding efficiency
- Overhead increases with the number of layers

What is important for multimedia communication over IP networks?

- On-the-fly & efficient adaptability to bandwidth variations, QoS levels
- Adaptation to different user & device requirements
- Complexity-scalable encoding/decoding

Our first solution - A new coding paradigm
Fine-Granularity Scalability (FGS)
[vanderSchaar - PhD thesis, '01]

FGS – embedded video coder (Successive refinement)

Goal: Achieve optimal description at each encoding stage

FGS – embedded video coder (Successive refinement)

Goal: Achieve optimal description at each encoding stage

1980 **Koshelev** proved that R-D problem is successively refinable if individual solutions of the R-D problem can be written as a Markov chain

$$\exists Q_{x_1,x_2|x} \text{ s.t.}$$

$$E\{d(X,X_i)\} \le D_i, \quad i = 1,2.$$

$$I(X;X_i) = R(D_i), \quad i = 1,2.$$

$$X \leftrightarrow X_2 \leftrightarrow X_1 \text{ is a Markov chain}$$

$$\Rightarrow \{D_1, D_2, R(D_1), R(D_2)\} \text{ is achievable}$$

1991 **Equitz** & **Cover** proved that the condition is also necessary.

FGS – embedded video coder (Successive refinement)

Goal: Achieve optimal description at each encoding stage

1980 **Koshelev** proved that R-D problem is successively refinable if individual solutions of the R-D problem can be written as a Markov chain

$$\exists Q_{x_1,x_2|x} \text{ s.t.}$$

$$E\{d(X,X_i)\} \le D_i, \quad i=1,2.$$

$$I(X;X_i)=R(D_i), \quad i=1,2.$$

$$X \leftrightarrow X_2 \leftrightarrow X_1 \text{ is a Markov chain}$$

$$\Rightarrow \{D_1,D_2,R(D_1),R(D_2)\} \text{ is achievable}$$

1991 **Equitz** & **Cover** proved that the condition is also necessary.

- Video sources are NOT successively refinable with respect to the PSNR distortion metric ☹
- Even if a source is not successively refinable, the penalty for FGS embedded coding is small ☺

Fine-Granular-Scalability (FGS) in MPEG-4

Fine-Granular-Scalability (FGS) in MPEG-4

- '98 Activity initiated by our group
  - MPEG-4 approved an FGS core-experiment
- '01 FGS became an International Standard
- Widely researched .....
  - Web search on “FGS coding” generates more than 2000 links
  - Most used scalable coder for multimedia communication research
  - Sessions dedicated to FGS at major IEEE conferences (ICIP, ICME etc.)
  - FGS opened a broad area of research (PhD theses based on FGS)
    - Optimal rate-allocation strategies (rate-shaping etc.)
    - Joint source-channel coding of FGS streams
    - Efficient streaming algorithms

Fine-Granular-Scalability (FGS) in MPEG-4

- '98 **Activity initiated by our group**

  MPEG-4 approved an FGS core-experiment

- '01 **FGS became an International Standard**
- Widely researched .....
  - Web search on "FGS coding" generates more than 2000 links
  - Most used scalable coder for multimedia communication research
  - Sessions dedicated to FGS at major IEEE conferences (ICIP, ICME etc.)
  - FGS opened a broad area of research (PhD theses based on FGS)
    - Optimal rate-allocation strategies (rate-shaping etc.)
    - Joint source-channel coding of FGS streams
    - Efficient streaming algorithms

However, FGS had coding efficiency penalty & no spatial scalability

Wavelets and motion compensation

- **Wavelet transform coding for still images (e.g. JPEG 2000)**

-> **Extension to video coding (3D wavelet video)**

- Using transforms for interframe coding goes back to '70s, '80s (e.g. Karlsson/Vetterli)
- Drawback was **lack of motion compensation**
    - Motion compensation is key to achieve high compression & visual quality, but difficult

Wavelets and motion compensation

- **Wavelet transform coding for still images (e.g. JPEG 2000)**

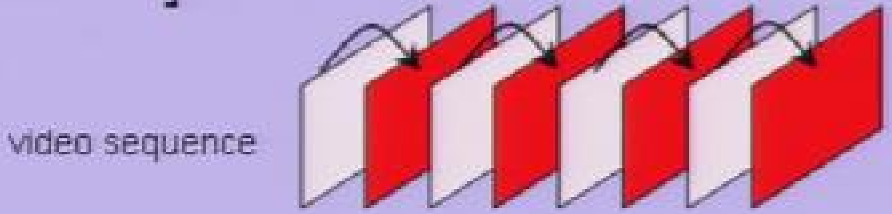
-> Extension to video coding (3D wavelet video)

- Using transforms for interframe coding goes back to '70s, '80s (e.g. Karlsson/Vetterli)
- Drawback was **lack of motion compensation**
  - Motion compensation is key to achieve high compression & visual quality, but difficult

Our contributions

- Unconstrained Motion Compensated Temporal Filtering (UMCTF)
- Fully Scalable 3-D Overcomplete Wavelet Video Coding
- 3-band temporal lifting structures
- Spatio-temporal MV scalability
- Rate-Distortion Optimized Anisotropic Motion Representation
- User-centric tradeoffs for spatio-temporal-SNR scalability
- Multiple Description Scalable Video Coding based on UMCTF

Motion compensated temporal filtering (MCTF) – background [Ohm '93]

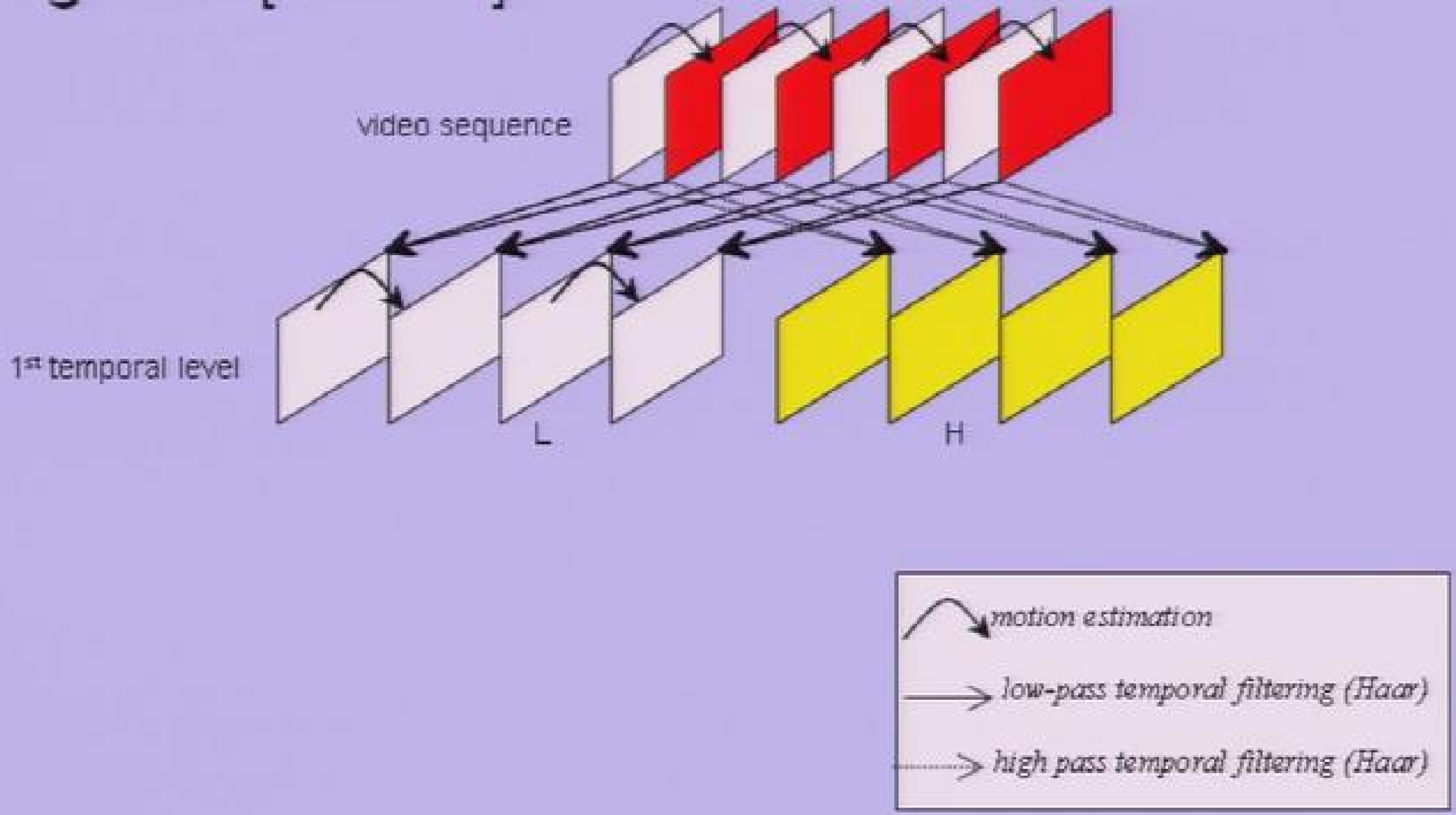


↷ *motion estimation*

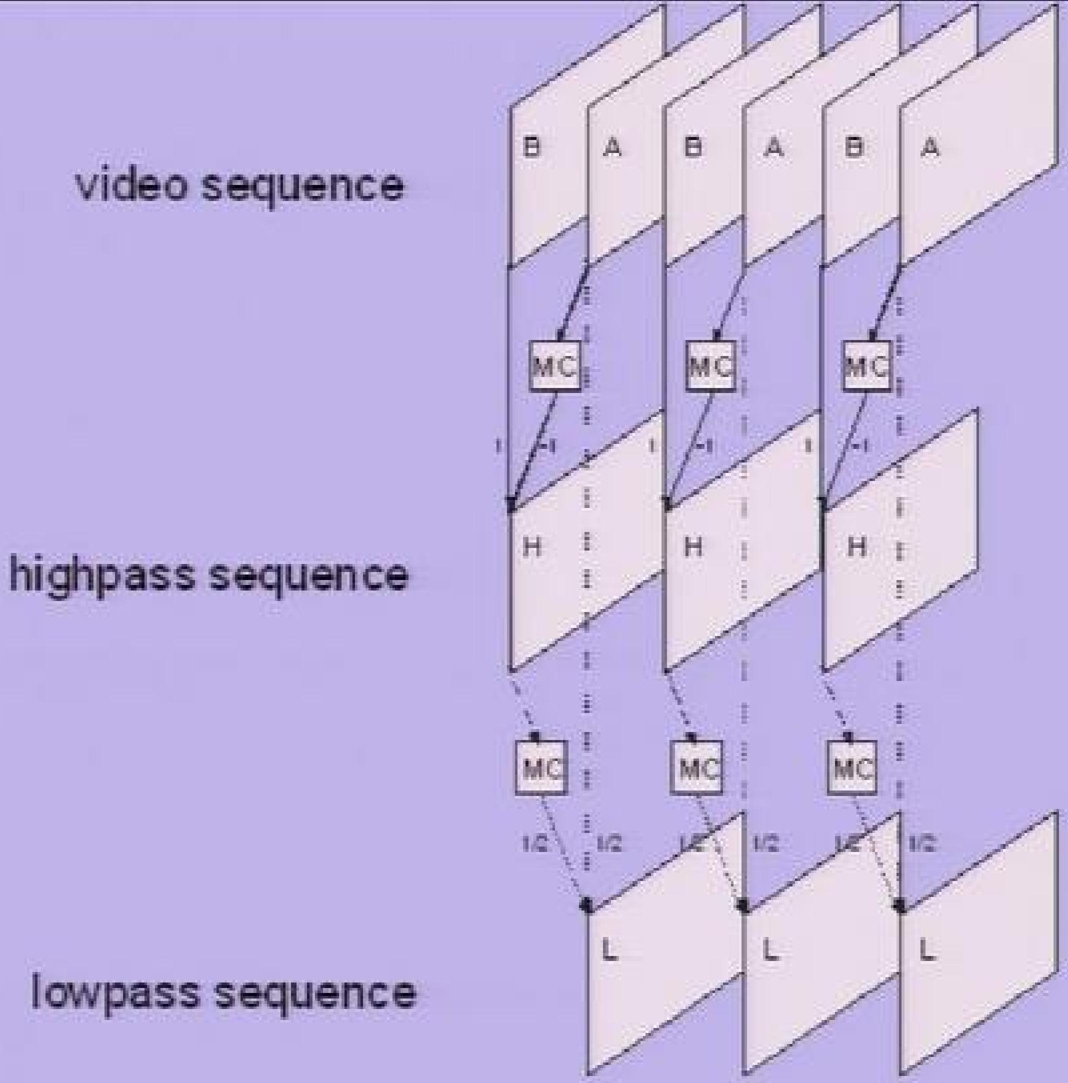
→ *low-pass temporal filtering (Haar)*

⇢ *high pass temporal filtering (Haar)*

Motion compensated temporal filtering (MCTF) – background [Ohm '93]



MCTF



video sequence

highpass sequence

lowpass sequence

$i = \{0,1\}$:

$$H_t[m+i] = \frac{1}{\sqrt{2}}\left(A_t[m+i] - \mathcal{I}_{(\cdot)}A_{t-1}[m+i-3]\right)$$

Prediction

$i = \{0,1\}$:

$$L_t[m+i-3] = \sqrt{2}A_{t-1}[m+i-3] + \mathcal{I}_{(\cdot)}H_t[m+i-1]$$

Update

Unconstrained MCTF – Adaptive temporal filtering

[vanderSchaar and Turaga '02]

Predict

$$H_t^{\lambda}[m,n] = L_t^{\lambda-1}[m,n] - \sum_{q=t-t_p^{\mathrm{init}}(\lambda)}^{t-1} \left(w_q[m,n] \cdot \alpha_q \cdot H_q^{\lambda-1}\left[m - d_m^{\mathcal{F}_{t-q}(q)}, n - d_n^{\mathcal{F}_{t-q}(q)}\right] \right)$$

$$- \sum_{q=t+1}^{t+t_p^{\mathrm{end}}(\lambda)} \left(w_q[m,n] \cdot \alpha_q \cdot H_q^{\lambda-1}\left[m - d_m^{\mathcal{B}_{q-t}(q)}, n - d_n^{\mathcal{B}_{q-t}(q)}\right] \right)$$

Unconstrained MCTF – Adaptive temporal filtering

[vanderSchaar and Turaga '02]

Predict

UMCTF weights Lifting parameters

$$H_t^{\lambda}[m,n] = L_t^{\lambda-1}[m,n] - \sum_{q=t-t_p^{\text{init}}(\lambda)}^{t-1} \left(\underline{w_q[m,n]} \cdot \underline{\alpha_q} \cdot H_q^{\lambda-1}\left[m - d_m^{\mathcal{F}_{t-q}(q)}, n - d_n^{\mathcal{F}_{t-q}(q)}\right] \right)$$

$$- \sum_{q=t+1}^{t+t_p^{\text{end}}(\lambda)} \left(\underline{w_q[m,n]} \cdot \underline{\alpha_q} \cdot H_q^{\lambda-1}\left[m - d_m^{\mathcal{B}_{q-t}(q)}, n - d_n^{\mathcal{B}_{q-t}(q)}\right] \right)$$

No. of lifting pairs

Update

Temporary Update Frame

Lifting parameters

$$Z_t\left[m - d_m^{\mathcal{B}_{q-t}(q)}, n - d_n^{\mathcal{B}_{q-t}(q)}\right] = Z_t\left[m - d_m^{\mathcal{B}_{q-t}(q)}, n - d_n^{\mathcal{B}_{q-t}(q)}\right] + w_q[m,n] \cdot \underline{\beta_q} \cdot H_q^{\lambda}[m,n]$$

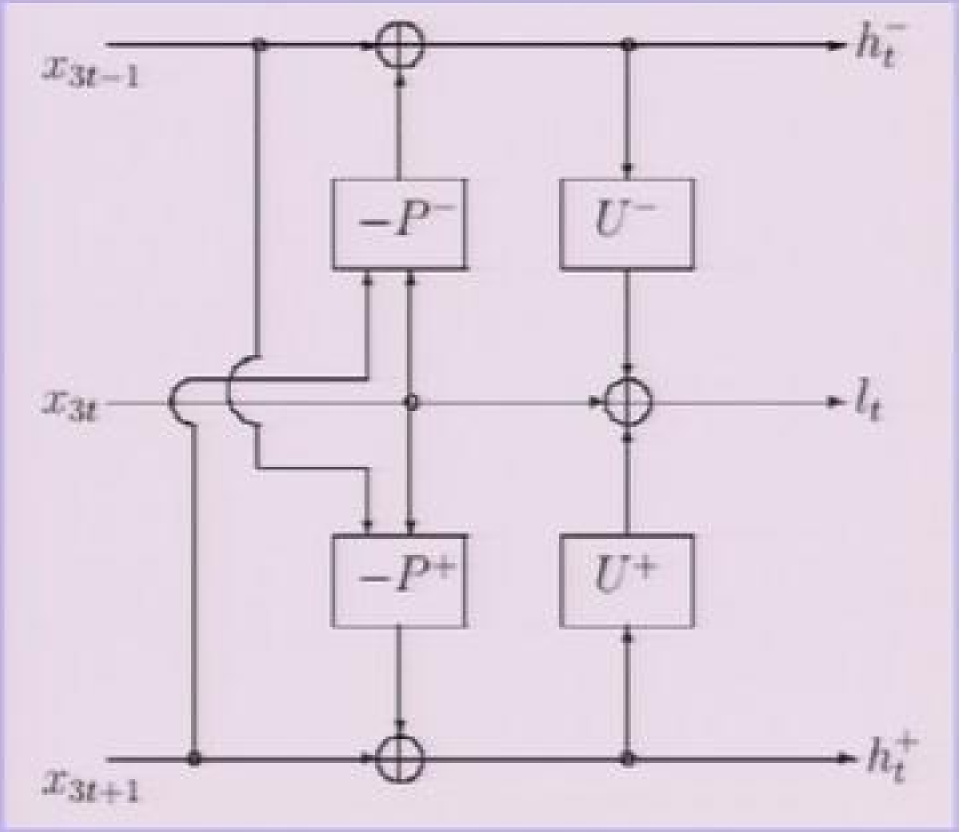
Connectivity Map

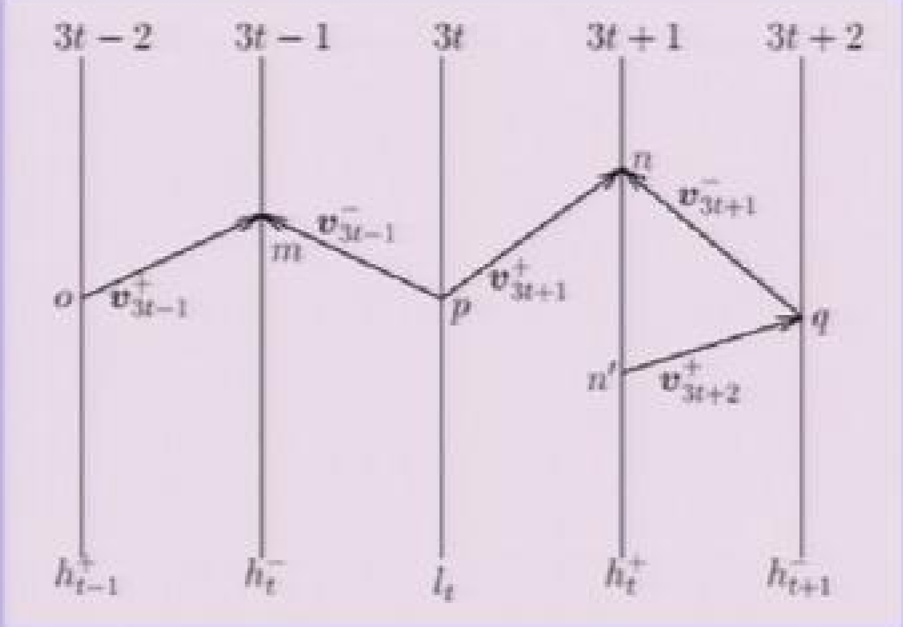
$$c_u\left[m - d_m^{\mathcal{B}_{q-t}(q)}, n - d_n^{\mathcal{B}_{q-t}(q)}\right] = c_u\left[m - d_m^{\mathcal{B}_{q-t}(q)}, n - d_n^{\mathcal{B}_{q-t}(q)}\right] + 1$$

Updated Frame

$$L_t^{\lambda}[m,n] = \left[L_t^{\lambda-1}[m,n] + \frac{1}{\max\{c_u[m,n], t^{\mathcal{B}}(\lambda)\}} Z_t[m,n] \right]$$

Example: Three-band decomposition structure with bi-directional predict operators [Tillier, Pesquet, vanderSchaar '03]

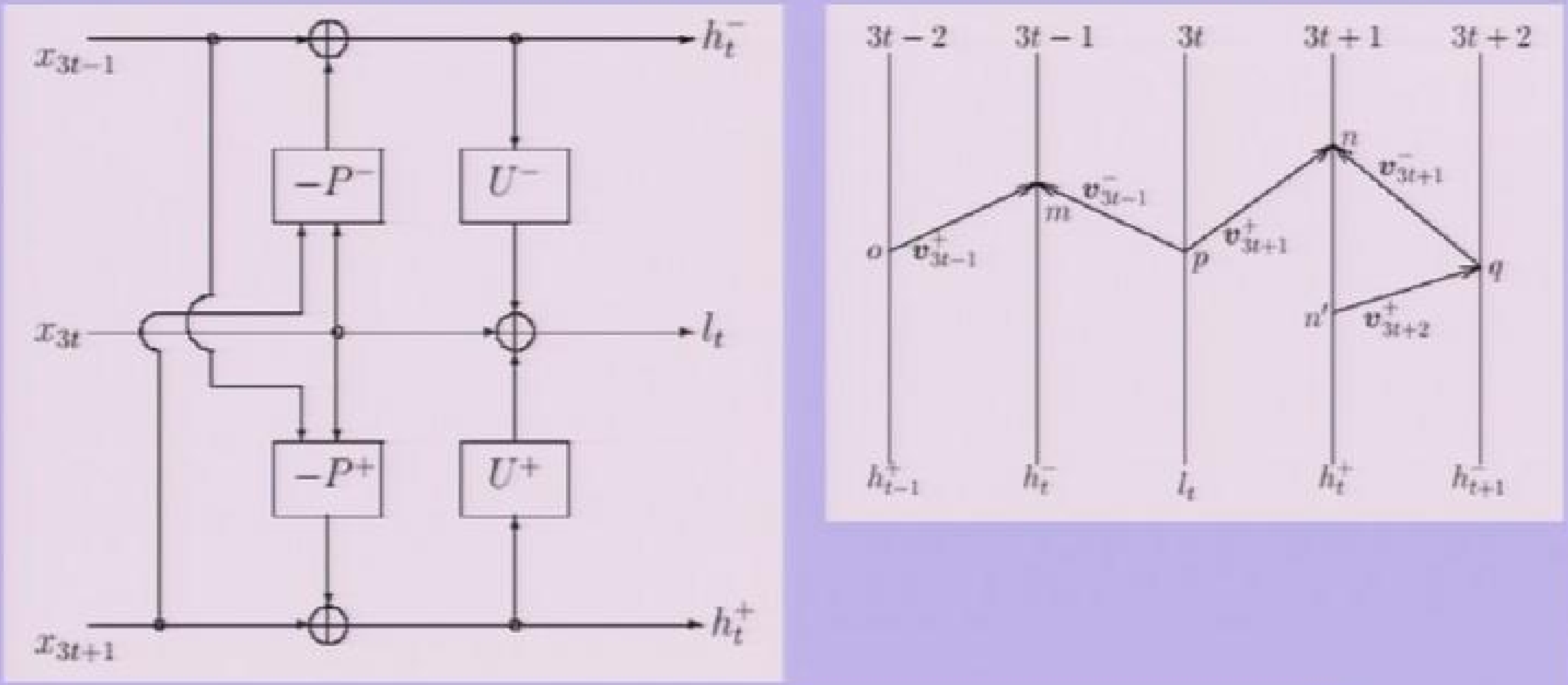




$$h_t^+(\boldsymbol{n}) = x_{3t+1}(\boldsymbol{n}) - \beta\, x_{3t+2}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^-) - (1-\beta)\, x_{3t}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^+)$$
$$h_t^-(\boldsymbol{m}) = x_{3t-1}(\boldsymbol{m}) - \beta\, x_{3t-2}(\boldsymbol{m} - \boldsymbol{v}_{3t-1}^+) - (1-\beta)\, x_{3t}(\boldsymbol{m} - \boldsymbol{v}_{3t-1}^-)$$

$$l_t(\boldsymbol{p}) = x_{3t}(\boldsymbol{p}) + \alpha h_t^+(\boldsymbol{p} + \boldsymbol{v}_{3t+1}^+) + \alpha h_t^-(\boldsymbol{p} + \boldsymbol{v}_{3t-1}^-)$$

Example: Three-band decomposition structure with bi-directional predict operators [Tillier, Pesquet, vanderSchaar '03]



$$h_t^+(\boldsymbol{n}) = x_{3t+1}(\boldsymbol{n}) - \beta\, x_{3t+2}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^-) - (1-\beta)\, x_{3t}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^+)$$

$c_q[m - u_m, n - u_n] - c_q[m - u_m, n - u_n] + 1$

Updated Frame

$$L_t^\lambda[m,n] = \left[L_t^{\lambda-1}[m,n] + \frac{1}{\max\{c_u[m,n], t^B(\lambda)\}} Z_t[m,n]\right]$$

Unconstrained MCTF – Adaptive temporal filtering

[vanderSchaar and Turaga '02]

Predict

UMCTF weights Lifting parameters

$$H_t^{\enclose{circle}{\lambda}}[m,n] = L_t^{\lambda-1}[m,n] - \sum_{q=t-t_p^{\mathrm{init}}(\lambda)}^{t-1} \left(\underline{w_q[m,n]} \cdot \underline{\alpha_q} \cdot H_q^{\lambda-1}[m - d_m^{\mathcal{F}_{t-q}(q)}, n - d_n^{\mathcal{F}_{t-q}(q)}] \right)$$

$$- \sum_{q=t+1}^{t+t_p^{\mathrm{end}}(\lambda)} \left(\underline{w_q[m,n]} \cdot \underline{\alpha_q} \cdot H_q^{\lambda-1}[m - d_m^{\mathcal{B}_{q-t}(q)}, n - d_n^{\mathcal{B}_{q-t}(q)}] \right)$$

No. of lifting pairs

Update

Temporary Update Frame

Lifting parameters

$$Z_t[m - d_m^{\mathcal{B}_{q-t}(q)}, n - d_n^{\mathcal{B}_{q-t}(q)}] = Z_t[m - d_m^{\mathcal{B}_{q-t}(q)}, n - d_n^{\mathcal{B}_{q-t}(q)}] + w_q[m,n] \cdot \underline{\beta_q} \cdot H_q^{\lambda}[m,n]$$

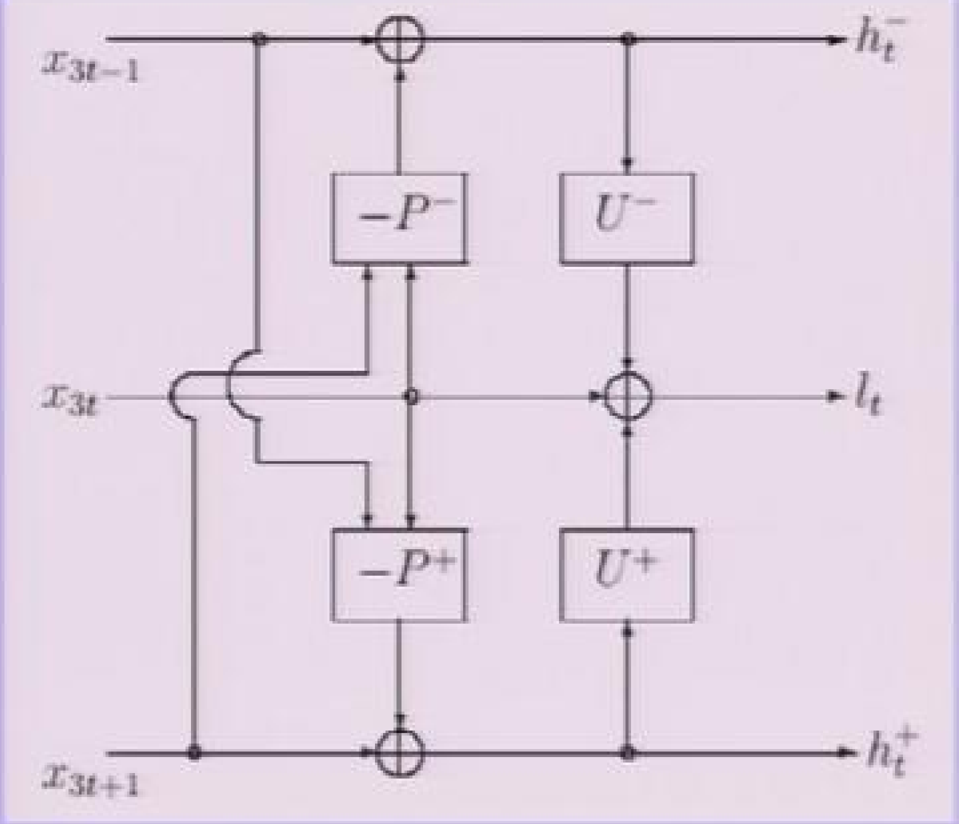
Connectivity Map

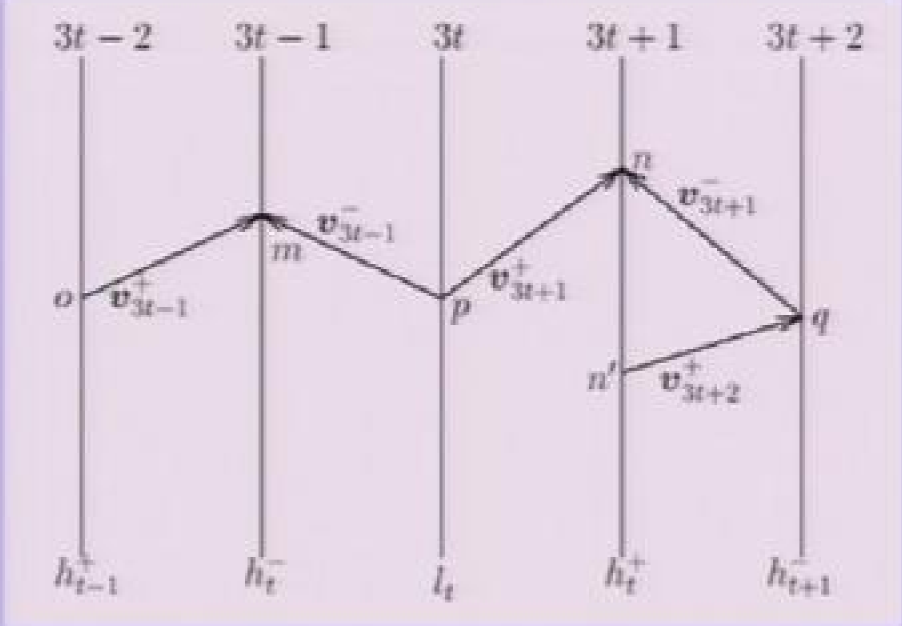
$$c_u[m - d_m^{\mathcal{B}_{q-t}(q)}, n - d_n^{\mathcal{B}_{q-t}(q)}] = c_u[m - d_m^{\mathcal{B}_{q-t}(q)}, n - d_n^{\mathcal{B}_{q-t}(q)}] + 1$$

Updated Frame

$$L_t^{\lambda}[m,n] = \left[L_t^{\lambda-1}[m,n] + \frac{1}{\max\{c_u[m,n], t^{\mathcal{B}}(\lambda)\}} Z_t[m,n] \right]$$

Example: Three-band decomposition structure with bi-directional predict operators [Tillier, Pesquet, vanderSchaar '03]

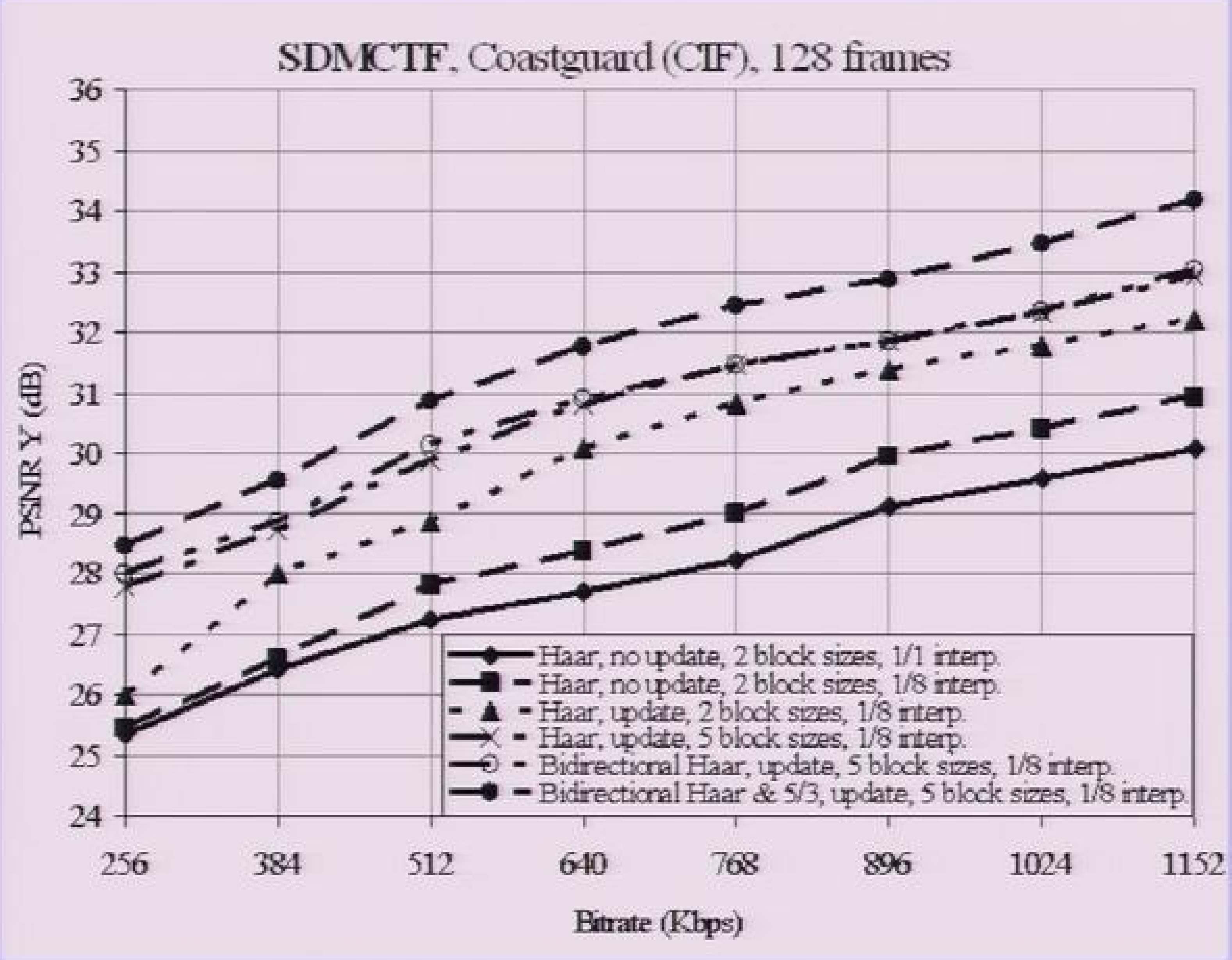




$$h_t^+(\boldsymbol{n}) = x_{3t+1}(\boldsymbol{n}) - \beta\, x_{3t+2}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^-) - (1-\beta)\, x_{3t}(\boldsymbol{n} - \boldsymbol{v}_{3t+1}^+)$$
$$h_t^-(\boldsymbol{m}) = x_{3t-1}(\boldsymbol{m}) - \beta\, x_{3t-2}(\boldsymbol{m} - \boldsymbol{v}_{3t-1}^+) - (1-\beta)\, x_{3t}(\boldsymbol{m} - \boldsymbol{v}_{3t-1}^-)$$

$$l_t(\boldsymbol{p}) = x_{3t}(\boldsymbol{p}) + \alpha h_t^+(\boldsymbol{p} + \boldsymbol{v}_{3t+1}^+) + \alpha h_t^-(\boldsymbol{p} + \boldsymbol{v}_{3t-1}^-)$$

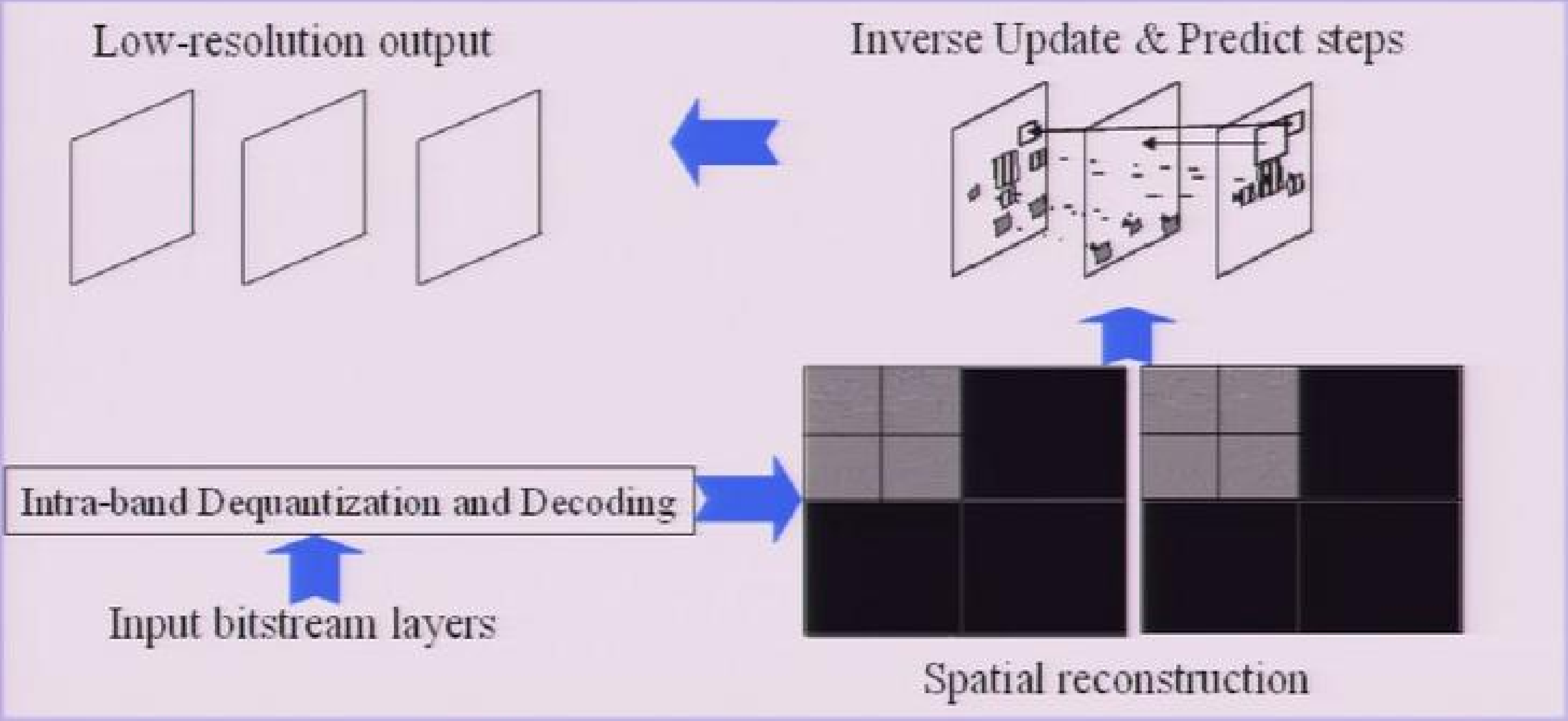
UMCTF improvements



Addressing Resolution Scalability in Video Transmission

Fundamental Problem

In the Conventional MCTF motion compensation and spatial filtering are not commutative



Low-resolution output

Inverse Update & Predict steps

Intra-band Dequantization and Decoding

Input bitstream layers

Spatial reconstruction

Wavelet Transform (WT) - before or after MC?

Wavelet Transform (WT) - before or after MC?

- Conventional: WT after MC – t+2D (SDMCTF)
  - Limited complexity ☺
  - Spatial scalability is not very efficient ☹
  - For block-based ME, Intra/Inter mode switch is not very efficient ☹
  - Discontinuities in the motion boundaries (blocking artefacts) are represented as high-frequency content in the high-frequency wavelet subbands ☹
  - ME accuracy is fixed for all spatial resolutions ☹
  - Same temporal decomposition scheme for all spatial subbands ☹

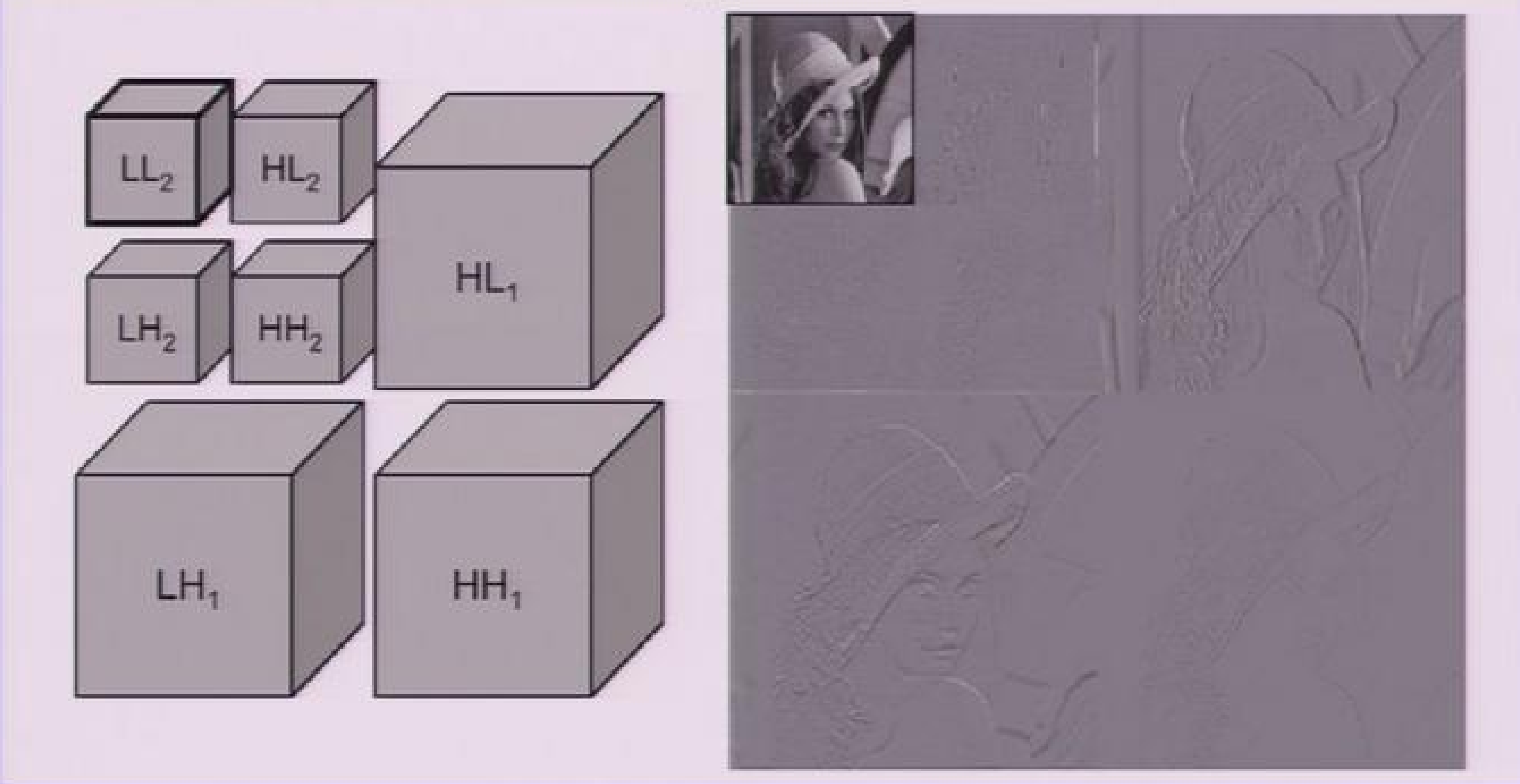
Wavelet Transform (WT) - before or after MC?

- **Conventional: WT after MC – t+2D (SDMCTF)**
  - Limited complexity ☺
  - Spatial scalability is not very efficient ☹
  - For block-based ME, Intra/Inter mode switch is not very efficient ☹
  - Discontinuities in the motion boundaries (blocking artefacts) are represented as high-frequency content in the high-frequency wavelet subbands ☹
  - ME accuracy is fixed for all spatial resolutions ☹
  - Same temporal decomposition scheme for all spatial subbands ☹
- **Our solution: WT before MC – 2D+t (IBMCTF)**
  - Multiple (separate) MC loops for wavelet bands ☺
  - No drift problem in spatial scalability ☺
  - Switching to "intra" coding mode without penalty ☺
  - Inefficiency of MC prediction in high bands ☹
    - due to shift variance of frequency-inverting alias

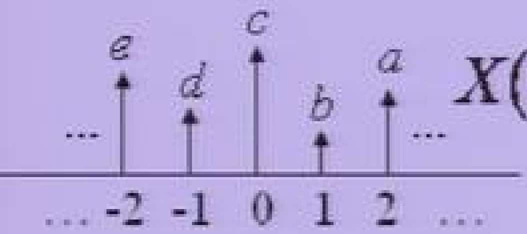
Justification for the use of Overcomplete DWT (ODWT)

- How does one perform in-band prediction and update?
- What necessitates the use of overcomplete transforms?

The "AdHoc" solution

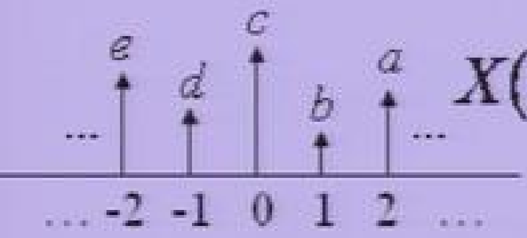


Justification for the use of ODWT



$$X(z) = \ldots + e \cdot z^{2} + d \cdot z + c + b \cdot z^{-1} + a \cdot z^{-2} + \ldots$$

Justification for the use of ODWT

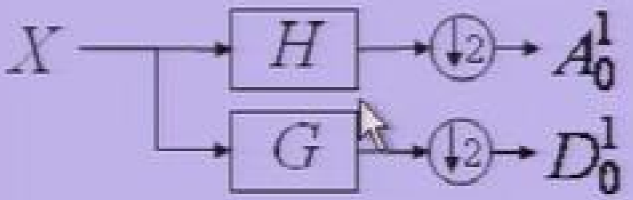


$$X(z) = \ldots + e \cdot z^{2} + d \cdot z + c + b \cdot z^{-1} + a \cdot z^{-2} + \ldots$$

Shift invariance of the DWT

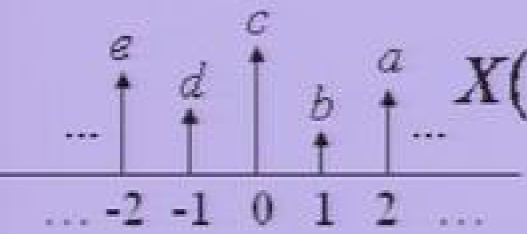
The even samples: $$X_0(z^2) = \tfrac{1}{2}\left(X(z) + X(-z)\right)$$

The odd samples: $$X_1(z^2) = \tfrac{1}{2} z^{-1}\left(X(z) - X(-z)\right)$$



$$A_0^1(z^2) = \tfrac{1}{2}\left(H(z) \cdot X(z) + H(-z) \cdot X(-z)\right)$$

Justification for the use of ODWT

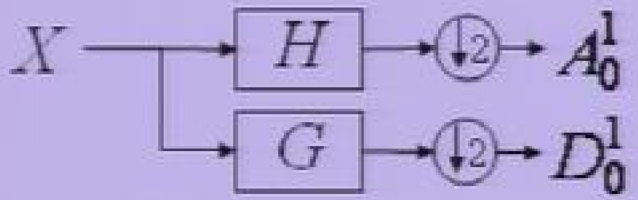


$X(z) = \ldots + e \cdot z^2 + d \cdot z + c + b \cdot z^{-1} + a \cdot z^{-2} + \ldots$

Shift invariance of the DWT

The even samples: $X_0(z^2) = \frac{1}{2}(X(z) + X(-z))$

The odd samples: $X_1(z^2) = \frac{1}{2}z^{-1}(X(z) - X(-z))$



$A_0^1(z^2) = \frac{1}{2}(H(z) \cdot X(z) + H(-z) \cdot X(-z))$

The transform of the shifted signal:

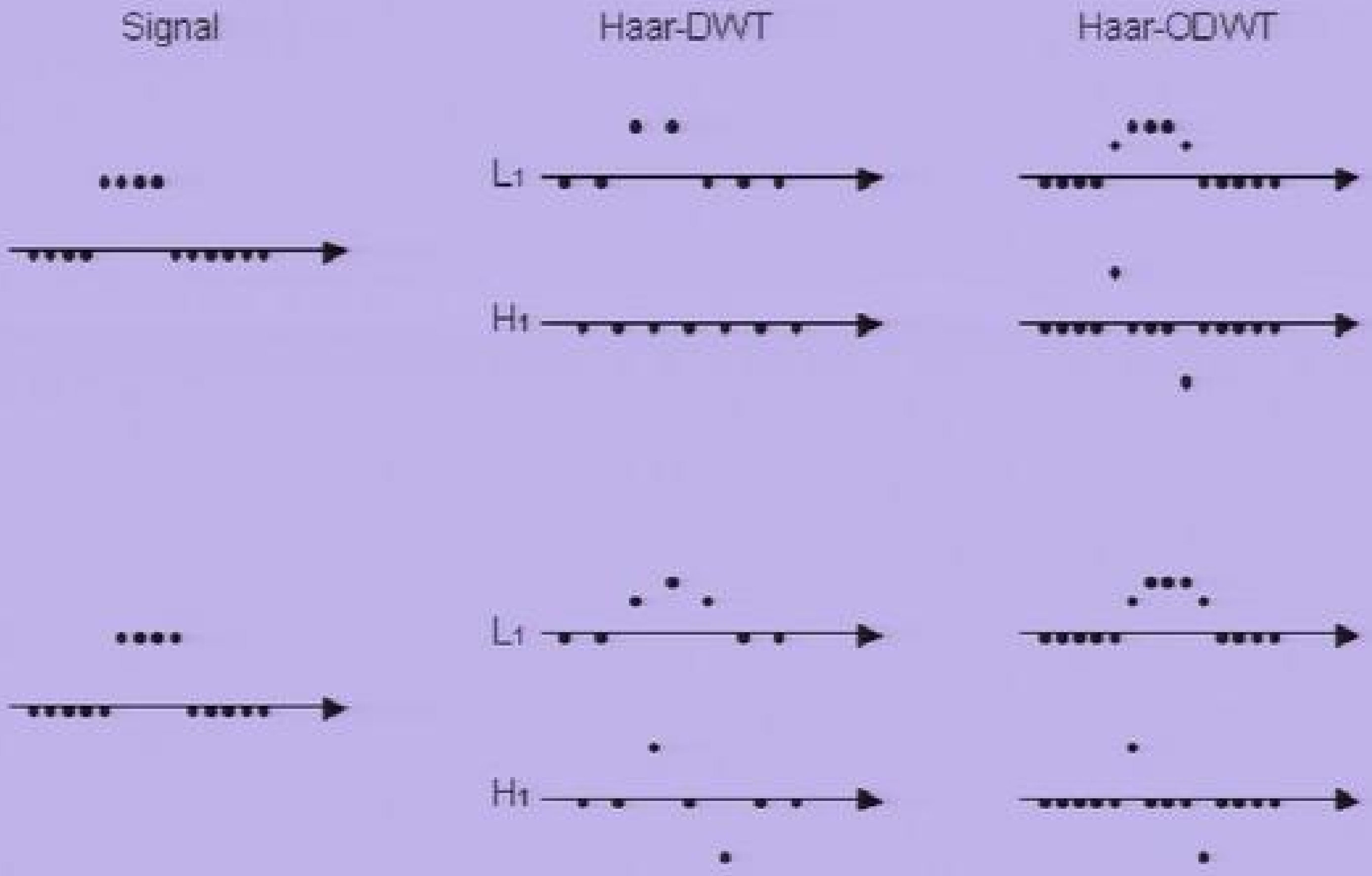
$X_s(z) = z^k X(z)$

$A_{s0}^1(z^2) = \frac{1}{2}z^k (H(z) \cdot X(z) + (-1)^k \cdot H(-z) \cdot X(-z)$

$X(z) - z^{-k}X_s(z) = 0$ BUT $A_0^1(z^2) - z^{-k}A_{s0}^1(z^2) \neq 0$

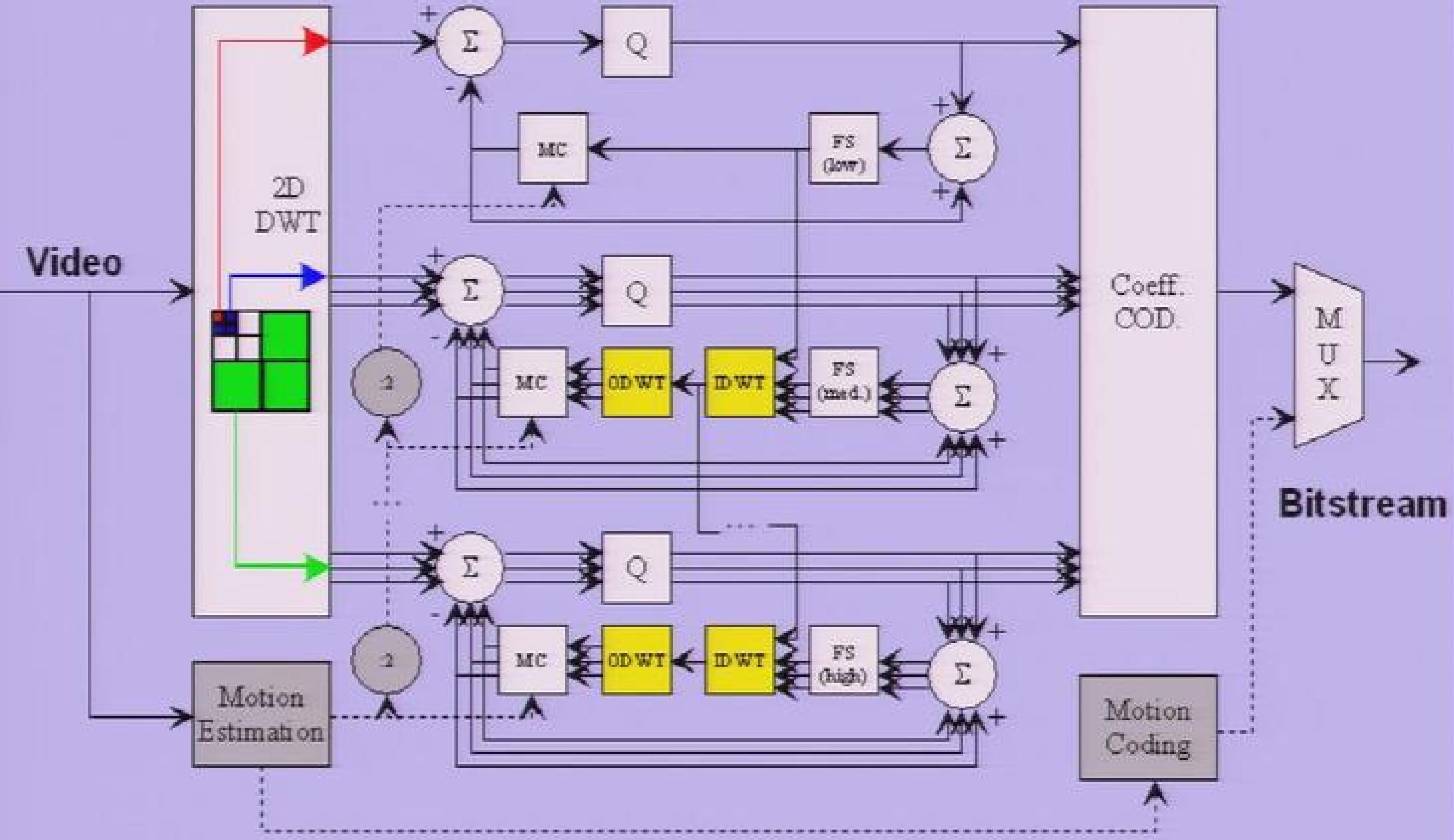
Justification for the use of ODWT

- Example for MC prediction problem in context of alias : Haar filter output of step edges



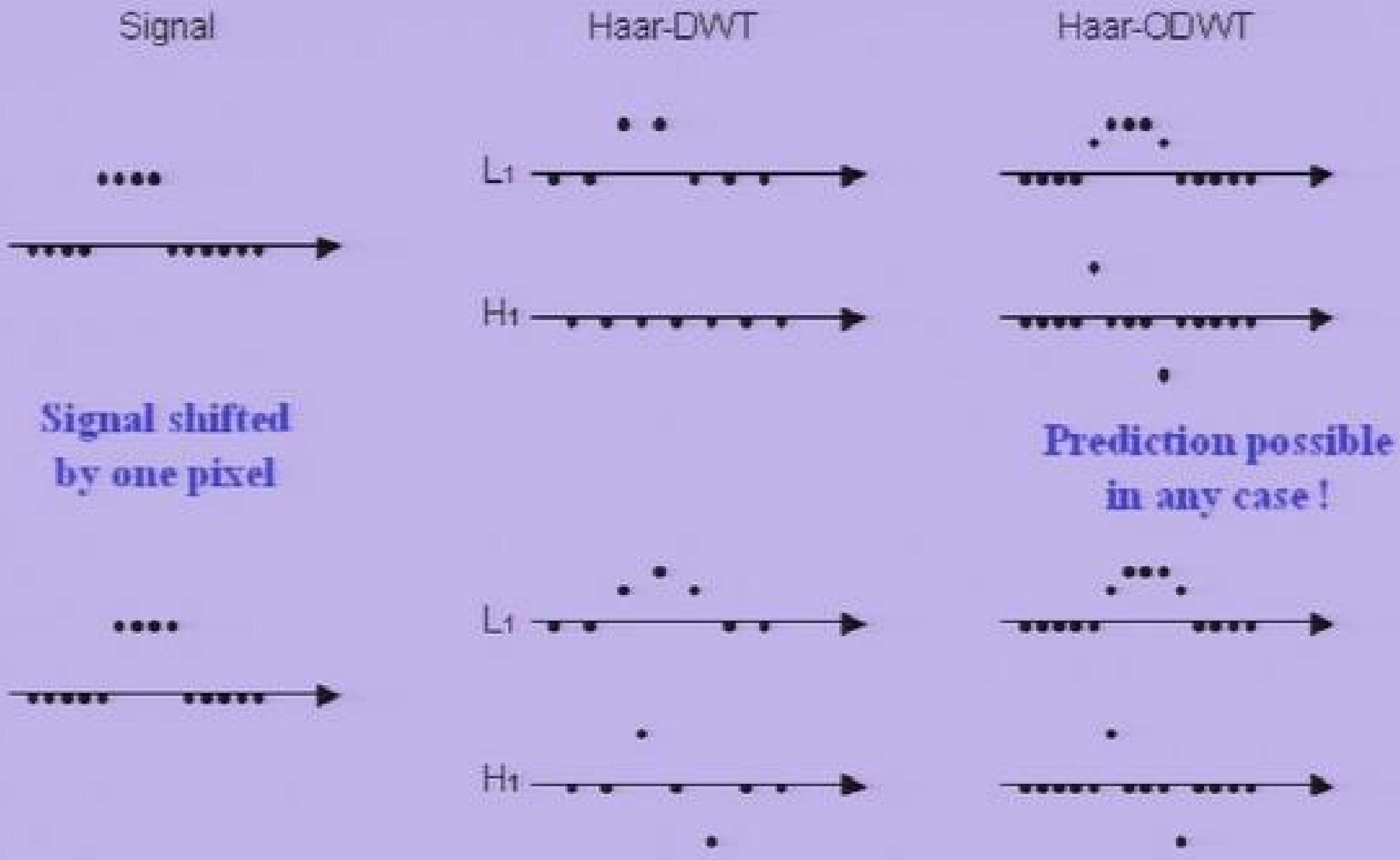
Fully Scalable 3-D Overcomplete Wavelet Video Coding

[Andreopolous, vanderSchaar '02,'03] [Ye, vanderSchaar '02]



Justification for the use of ODWT

- Example for MC prediction problem in context of alias : Haar filter output of step edges



Signal

Haar-DWT

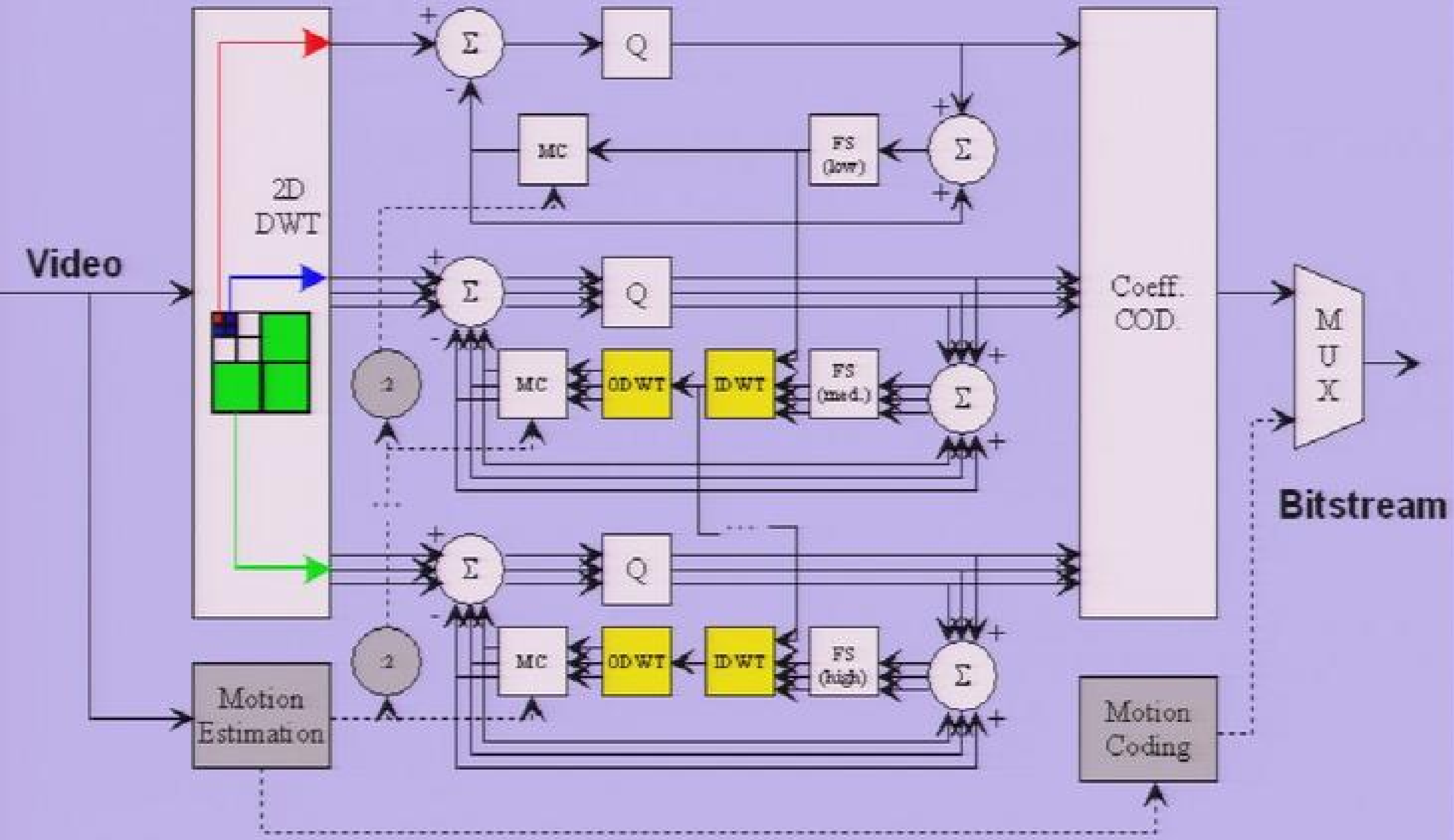
Haar-ODWT

Signal shifted by one pixel

Prediction possible in any case !

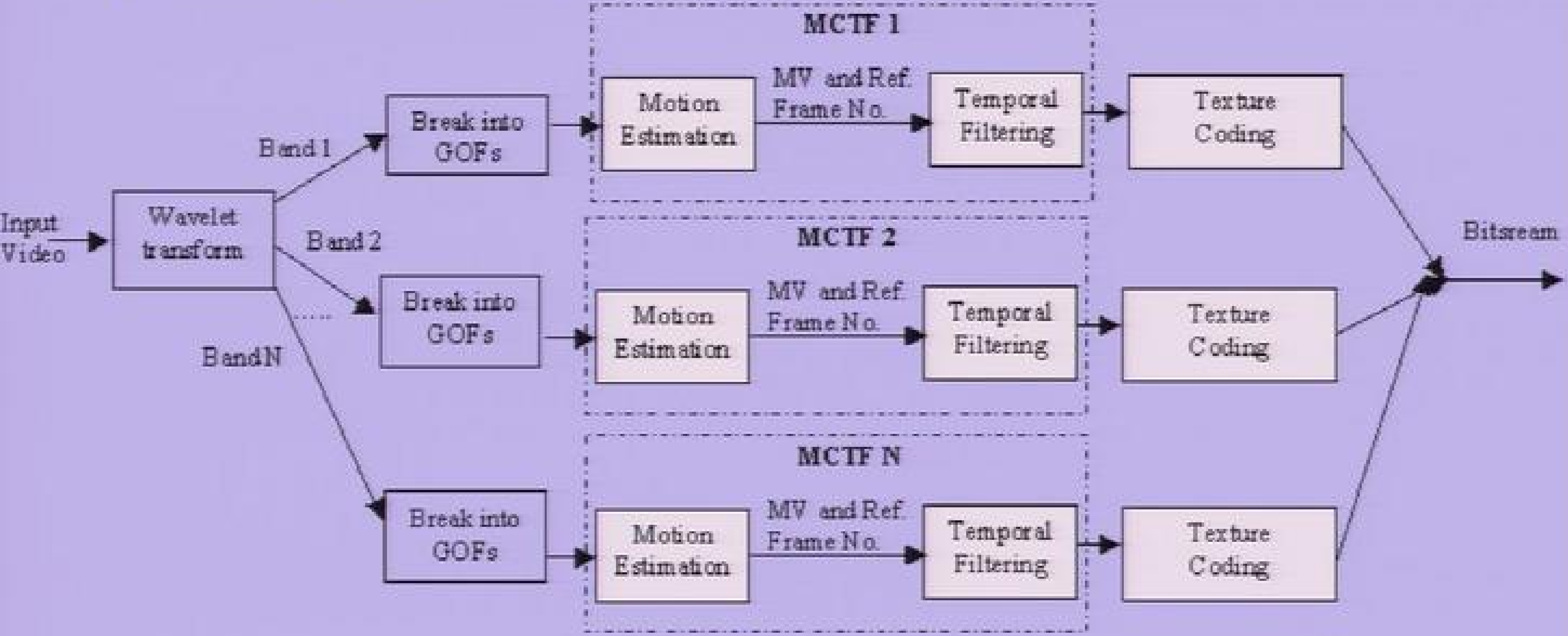
Fully Scalable 3-D Overcomplete Wavelet Video Coding

[Andreopolous, vanderSchaar '02,'03] [Ye, vanderSchaar '02]



Fully Scalable 3-D Overcomplete Wavelet Video Coding

[vanderSchaar, Andreopolous, Ye '02]



Input Video

Wavelet transform

Band 1

Band 2

......

Band N

Break into GOFs

MCTF 1

Motion Estimation

MV and Ref. Frame No.

Temporal Filtering

Texture Coding

Break into GOFs

MCTF 2

Motion Estimation

MV and Ref. Frame No.

Temporal Filtering

Texture Coding

Break into GOFs

MCTF N

Motion Estimation

MV and Ref. Frame No.

Temporal Filtering

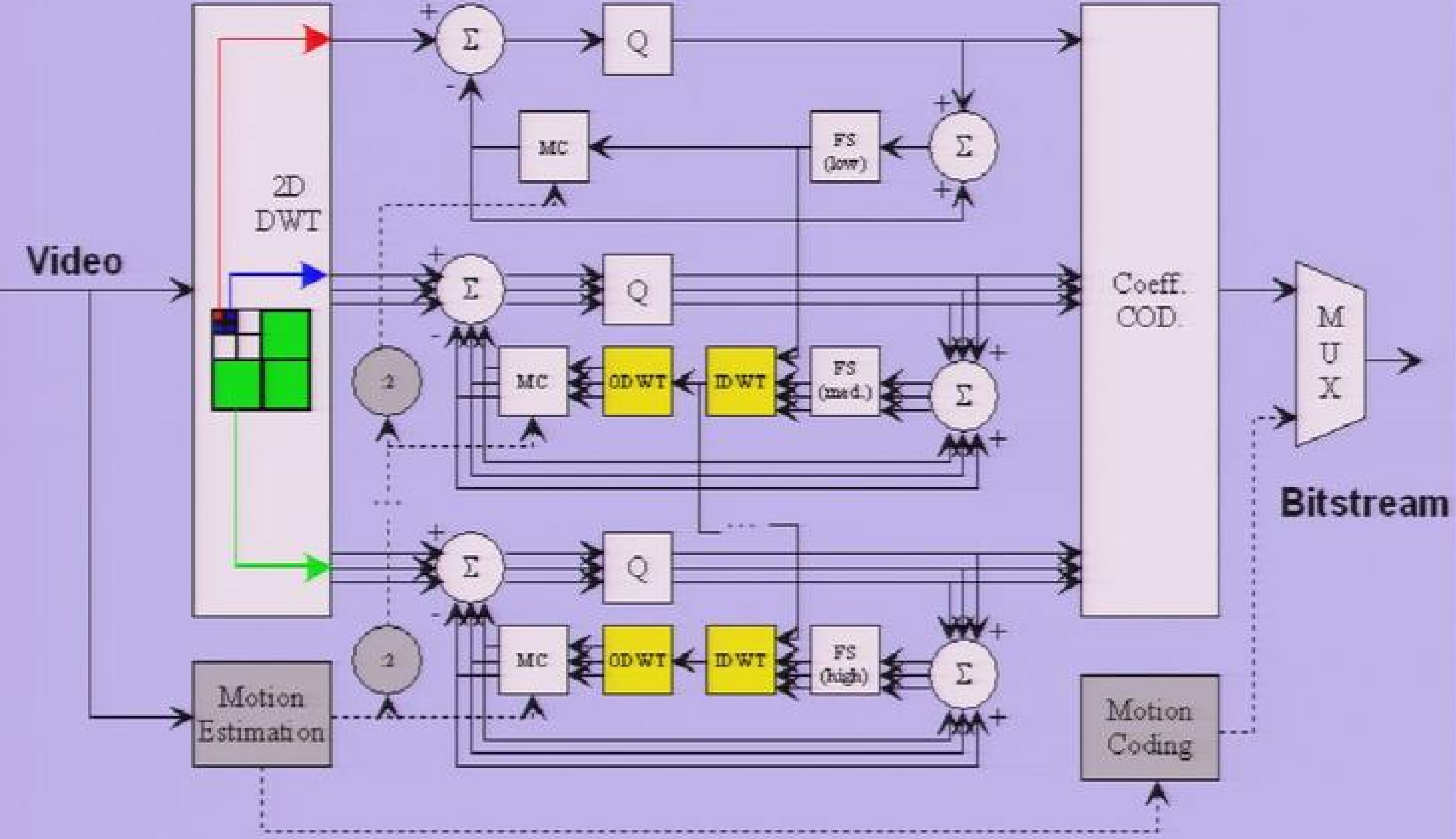
Texture Coding

Bitsream

- Different prediction structures per resolution/subband
- Different accuracy of the motion estimation is possible
- Different prediction structures per resolution/subband
- Different GOP structures
- Enables backwards compatibility with DCT standards
- Complexity adaptation per resolution

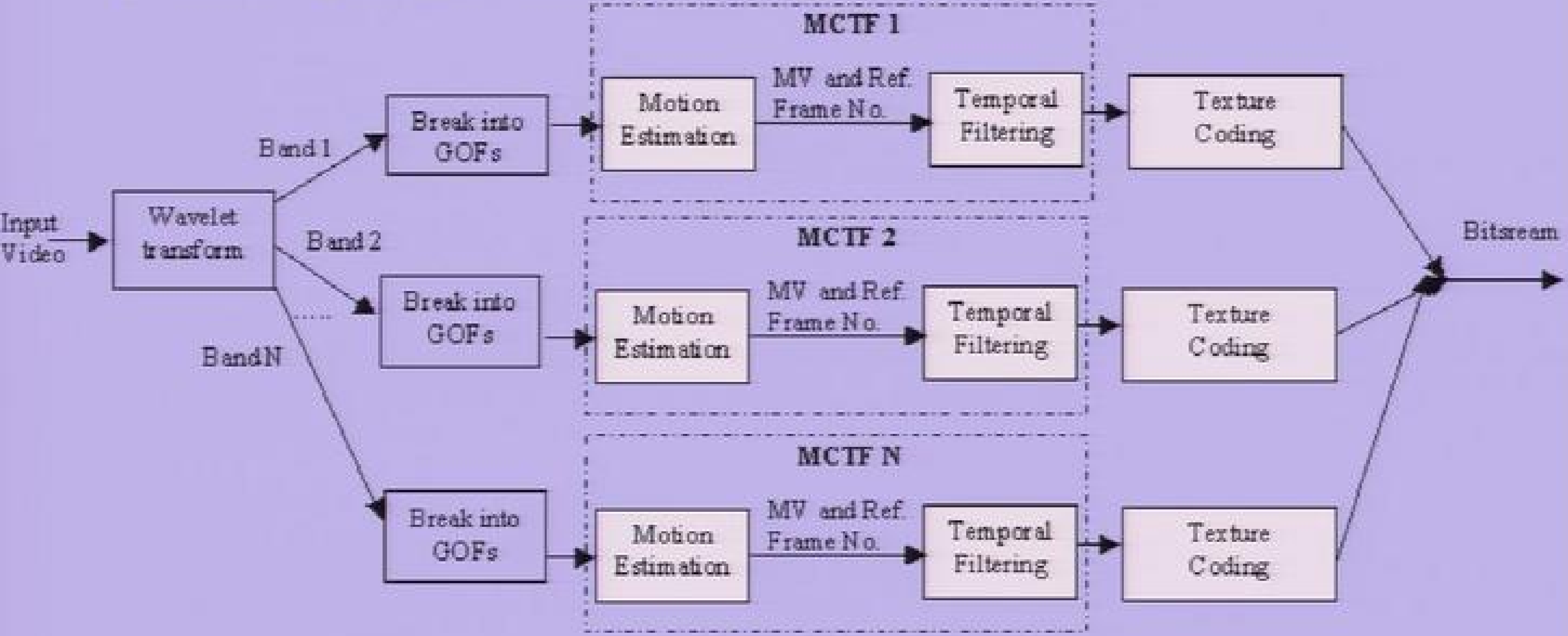
Fully Scalable 3-D Overcomplete Wavelet Video Coding

[Andreopolous, vanderSchaar '02,'03] [Ye, vanderSchaar '02]



Fully Scalable 3-D Overcomplete Wavelet Video Coding

[vanderSchaar, Andreopolous, Ye '02]



Input Video
Wavelet transform
Band 1
Band 2
Band N
MCTF 1
MCTF 2
MCTF N
Break into GOFs
Motion Estimation
MV and Ref. Frame No.
Temporal Filtering
Texture Coding
Bitsream

- Different prediction structures per resolution/subband
- Different accuracy of the motion estimation is possible
- Different prediction structures per resolution/subband
- Different GOP structures
- Enables backwards compatibility with DCT standards
- Complexity adaptation per resolution

Spatial scalability in wavelet video coders

Spatial scalability

Conventional

SDMCTF (t+2D)

Our

IBMCTF (2D+t)

Our results

ShuttleStart

Codec / Bitrates (kbps):	475	753	1335	2387
SDMCTF/Mean PSNR (dB):	41.32	42.64	43.83	44.46
IBMCTF/Mean PSNR (dB):	41.78	43.28	44.68	45.52
MPEG-4 AVC/Mean PSNR (dB):	40.38	41.81	43.25	44.72

Raven

Codec / Bitrates (kbps):	1010	1651	3041	5438
SDMCTF/Mean PSNR (dB):	38.37	39.90	41.52	42.59
IBMCTF/Mean PSNR (dB):	38.47	40.20	41.99	43.23
MPEG-4 AVC/Mean PSNR (dB):	38.33	39.86	41.47	43.11

Soccer

Codec / Bitrates (kbps):	1909	3001	5250	9246
SDMCTF/Mean PSNR (dB):	35.76	37.33	38.96	40.79
IBMCTF/Mean PSNR (dB):	35.71	37.47	39.25	41.22
MPEG-4 AVC/Mean PSNR (dB):	37.01	38.60	40.30	42.13

City

Codec / Bitrates (kbps):	1202	2148	4869	10865
SDMCTF/Mean PSNR (dB):	36.35	38.10	39.80	41.11
IBMCTF/Mean PSNR (dB):	36.61	38.41	40.24	41.74
MPEG-4 AVC/Mean PSNR (dB):	36.17	37.70	39.41	41.37

MC Wavelet video coding - current status

- Major theoretical problems seem to be resolved, but ...

- ... the present status of development is not optimum
- ...optimization for visual improvement (deblocking etc.) needed

MPEG standardization – status?

Chair MPEG scalable video coding (mid-2002 --- begin 2005)

- AVC extension based on UMCTF
- Ad-Hoc Group on Interframe Wavelet Video Coding (chair)

Our new work on video compression & processing

- Scalable video coding using oriented transforms
- User-centric video coding
- Content-aware source activation and compression for multi-camera surveillance applications (coherent source codebooks)
- Power-scalable compression algorithms

Why Operational R-D models?

- Utility-cost functions for our proactive wireless media
- Encoder optimization (R-D, but also Complexity)
- Joint source-channel coding (cross-layer)

- Two types of methodologies
  - Empirical approach - where experimental RD data is fitted to derive functional expressions [Liu '96][Zhang '97][Girod '00]
  - Analytical approach – based on traditional RD theory

  [Sakrison '68][Mallat '98][Moulin '99,'01]

- Realistic R-D models for wavelet video – missing [Wang, vanderSchaar '05]

Operational R-D models for wavelet video

Our analysis:

- **Low pass temporal frames –** similar properties as images

(based on work of Mallat, Moulin etc.)

- **High pass temporal frames –** obey Laplacian distribution & Intra-scale dependency— doubly stochastic model leading to Markov property

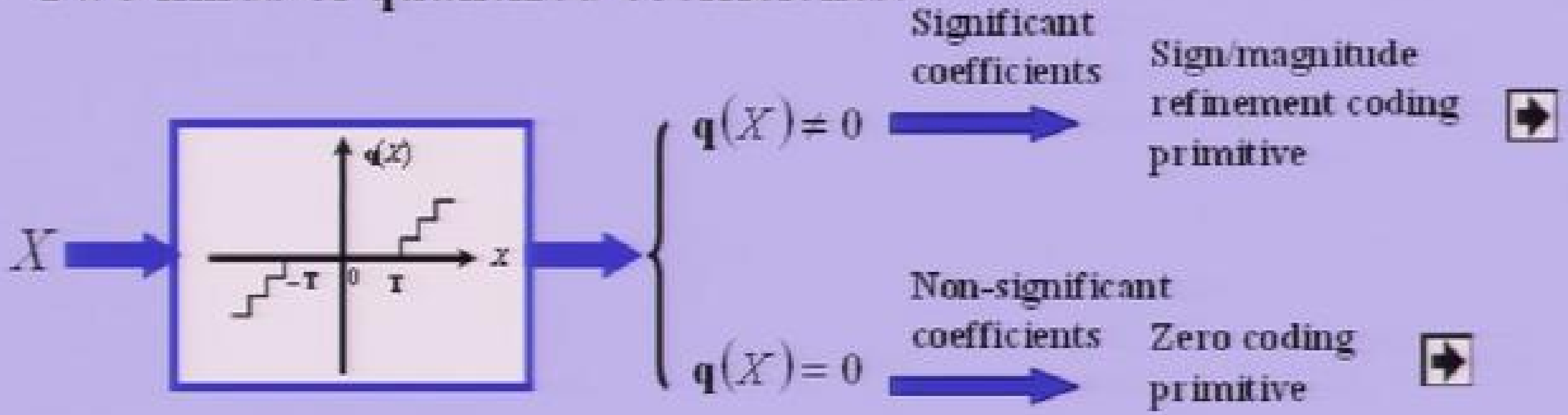
$$X \sim N(0,\theta)$$

$$\Theta \sim p(\theta) = \frac{1}{\sigma^2} e^{-\frac{1}{\sigma^2}\theta}$$

$$X \longrightarrow \Theta \longrightarrow NX$$

Bit Rate Estimation

- **Features of context adaptive coding of detail subbands:**
  - All the subbands are coded independently to achieve resolution scalability;
  - Uniform deadzone quantizer (deadzone: T, quantization step size: Δ) is used to quantize DWT coefficients X ;
  - Two kinds of quantized coefficients:

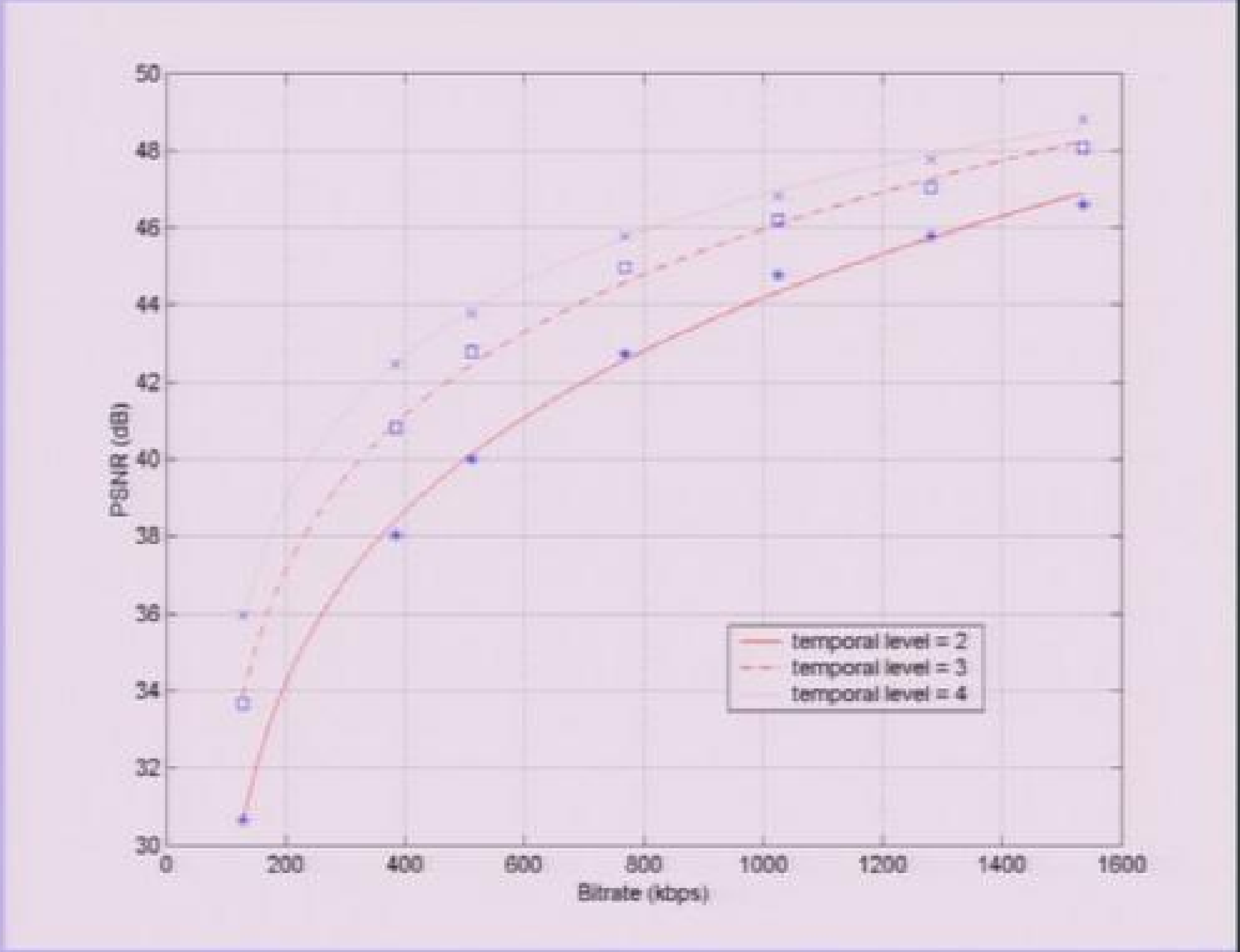


$\mathbf{q}(X) \neq 0$ Significant coefficients → Sign/magnitude refinement coding primitive

$\mathbf{q}(X) = 0$ Non-significant coefficients → Zero coding primitive

Experimental Results

(Coastguard sequence—3 spatial decomposition level)



Total coding bitrate subband (j,k):

$$\mathbf{R}_{j,k}(\upsilon) = \rho\mathbf{R}s(\upsilon) + \mathbf{Rzc}(\upsilon)$$

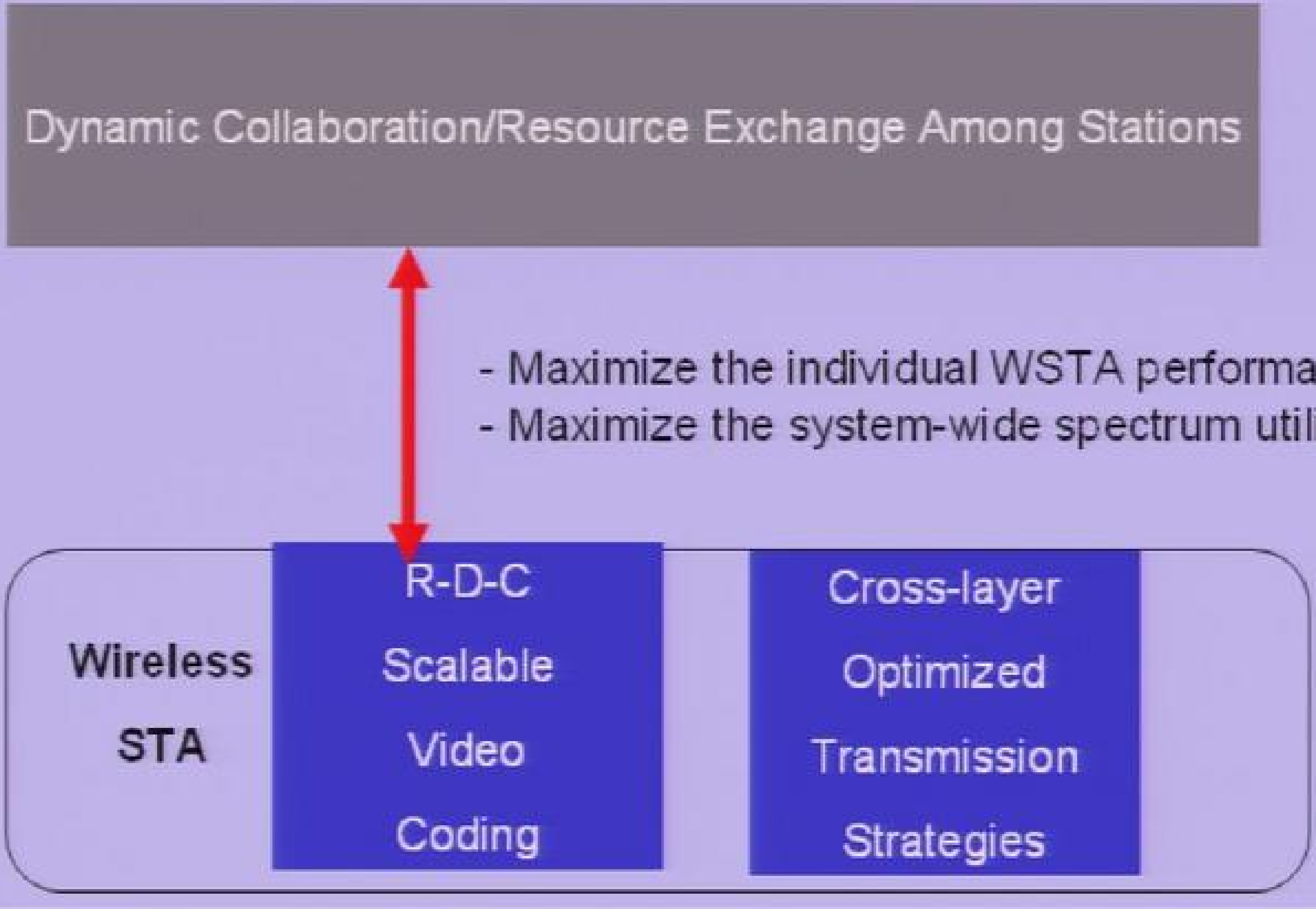
$$\upsilon = \Delta_{j,k} / \sigma_{j,k}$$

Total frame bitrate:

$$\overline{\mathbf{R}} = 4^{-J}\mathbf{R} + \sum_{j,k} 4^{-j}\mathbf{R}_{j,k}$$

Collaborative framework for wireless multimedia

Goal: Construct a system, where users can borrow or lend resources from the system/other users, according to their specific utility and resource awareness.



Dynamic Collaboration/Resource Exchange Among Stations

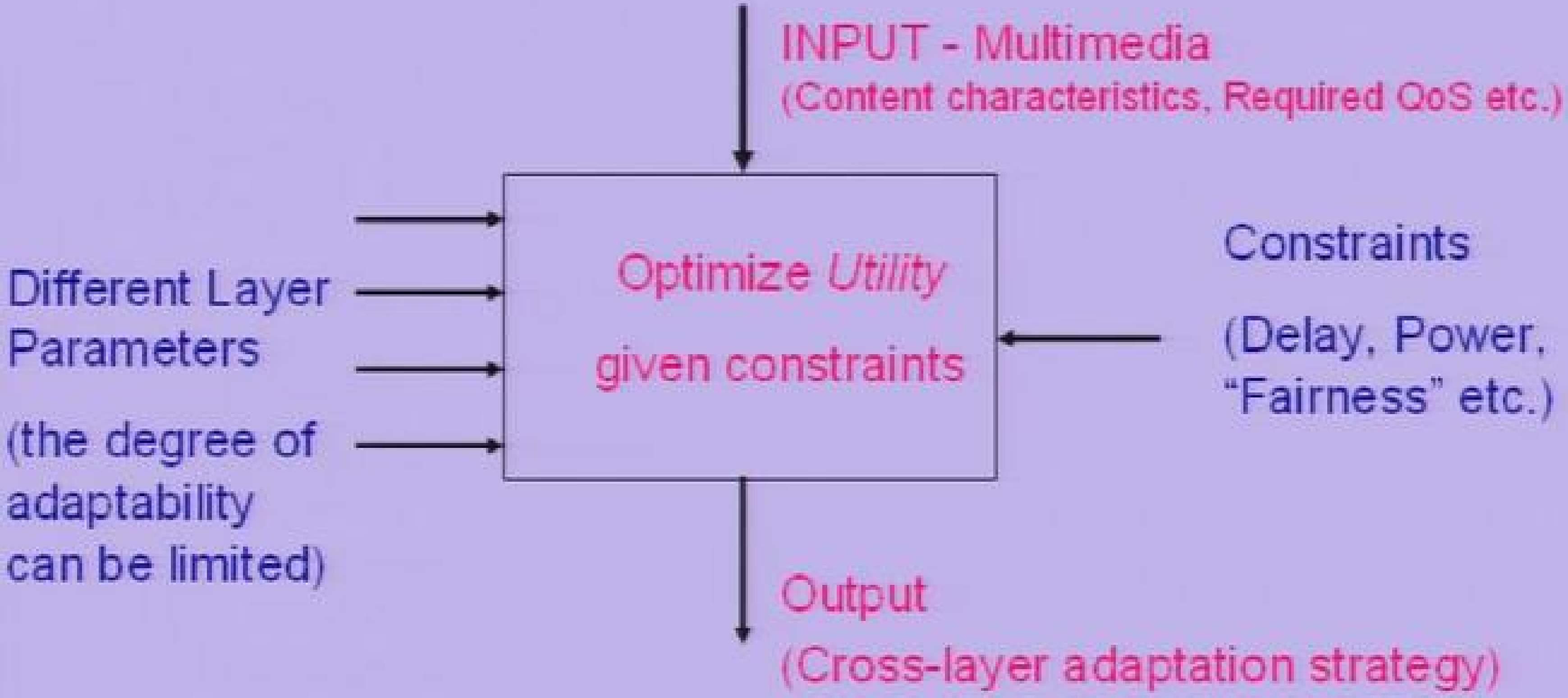
- Maximize the individual WSTA performance and
- Maximize the system-wide spectrum utilization

Wireless STA

R-D-C Scalable Video Coding

Cross-layer Optimized Transmission Strategies

Conceptual Framework (System View of Cross Layer Optimization)



INPUT - Multimedia
(Content characteristics, Required QoS etc.)

Different Layer Parameters

(the degree of adaptability can be limited)

Optimize *Utility*
given constraints

Constraints

(Delay, Power, "Fairness" etc.)

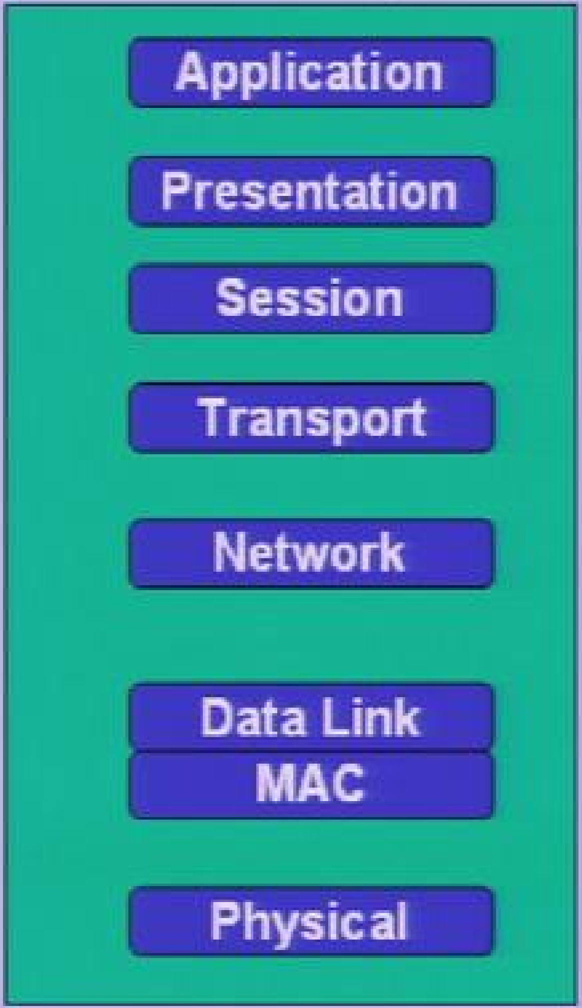
Output
(Cross-layer adaptation strategy)

- **Utility:** video quality, power, system-wide network utilization etc.

Strategies at different layers are collected into a composite strategy *S*:

$$S = \left\{ PHY_1, \ldots, PHY_{N_P}, MAC_1, \ldots, MAC_{N_M}, \ldots \right\}$$

OSI Layers



Application

Presentation

Session

Transport

Network

Data Link

MAC

Physical

- **RF**
  - Transmit power
  - Antenna direction
- **Baseband**
  - Modulation
  - Equalization
- **Link/MAC**
  - Error correction coding
  - ARQ
  - Admission Control and Scheduling
  - Packetization
- **Transport/Network**
  - TCP/UDP
  - Packetization
- **Application**
  - Compression strategies
  - Rate/Format adaptation
  - FEC/ARQ
  - Scheduling
  - Packetization

Cross-layer optimization – problem definition

Determine the optimal composite strategy

$$\boxed{S^{opt}(\mathbf{x}, mc) = \arg\max_{S} Q(S(\mathbf{x}), mc)}$$

subject to constraints

$$Delay(S(\mathbf{x}), mc) \le D_{\max} \text{ and } Power(S(\mathbf{x}), mc) \le Power_{\max}$$

given instantaneous
channel conditions $\mathbf{x} = (SNR, contention)$,
multimedia content characteristic mc,
maximum tolerable delay $D_{\max}$ and maximum power $Power_{\max}$.

Why is finding the optimal solution to this cross-layer optimization problem difficult?

Why is finding the optimal solution to this cross-layer optimization problem difficult?

- Deriving analytically **Q**, **Delay**, **Power** is often difficult and sometimes these functions are not deterministic (only worst/average values can be determined) and non-linear;
- some of the strategies *PHYi*, *MACi*, *Transi*, *Appi* depend on other strategies deployed at the same or other layers.;
- the wireless channel conditions may change continuously;

Why is finding the optimal solution to this cross-layer optimization problem difficult?

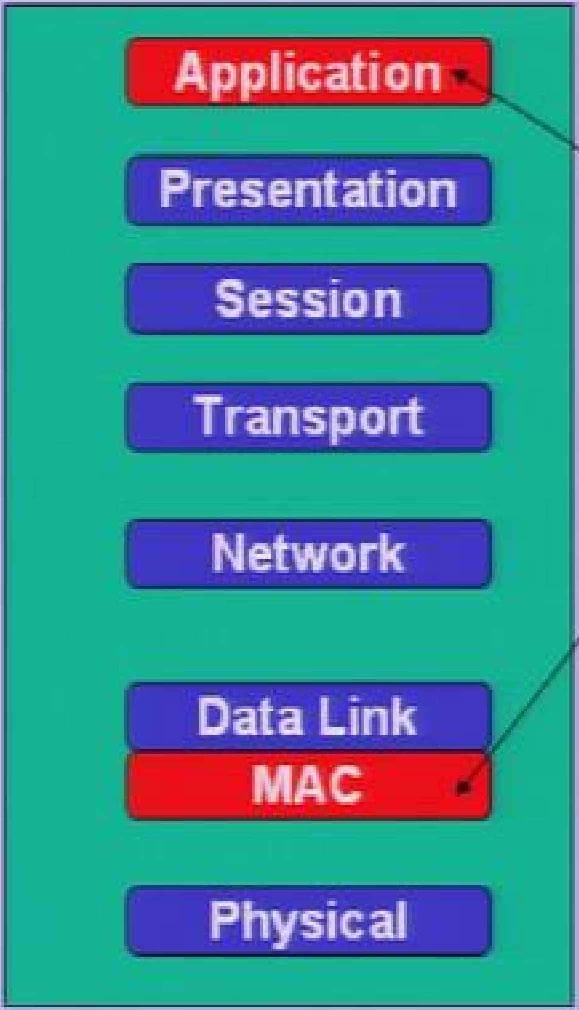
- Deriving analytically **Q**, **Delay**, **Power** is often difficult and sometimes these functions are not deterministic (only worst/average values can be determined) and non-linear;
- some of the strategies *PHYi, MACi, Transi, Appi* depend on other strategies deployed at the same or other layers.;
- the wireless channel conditions may change continuously;
- the multimedia traffic characteristics vary dynamically;
- different power and implementation constraints;.
- interaction among stations

Goal: formal procedures need to be established for optimal initializations, grouping of transmission strategies at different stages, and ordering etc

Classification of cross-layer solutions [vanderSchaar, Shankar '05]

Cross-layer video optimization

[Li, vanderSchaar '03][[vanderSchaar, Choi, Krishnamachari '03][Shankar, vanderSchaar '04][Krishnaswamy, vanderSchaar '04][vanderSchaar, Shankar '05]



Examples

- **MAC** – retransmission limit adaptation, packetization
- **Application** - packetization, rate adaptation and prioritized scheduling strategies
- **Cross-layer** – MAC+ Application layers

Does cross-layer optimization help?

- Example: MAC research (e.g. Choi et al, Goldsmith et al) has shown the importance of adapting the packet-size *L*

Does cross-layer optimization help?

- Example: MAC research (e.g. Choi et al, Goldsmith et al) has shown the importance of adapting the packet-size L

$$P_e^m(L) = 1 - (1 - p_b^m)^L$$

$$Throughput = \frac{L}{L + L_m^{header}} * (1 - P_e^m(L))$$

Optimal packet-size determined by MAC:

$$L^* = \frac{-L_m^{header} + \sqrt{\left(L_m^{header}\right)^2 - \frac{L_m^{header}}{2\log(1 - p_b^m)}}}{2}$$

Does cross-layer optimization help?

- Example: MAC research (e.g. Choi et al, Goldsmith et al) has shown the importance of adapting the packet-size *L*

$$P_e^m(L) = 1-(1-p_b^m)^L$$

Apply this solution to wireless video

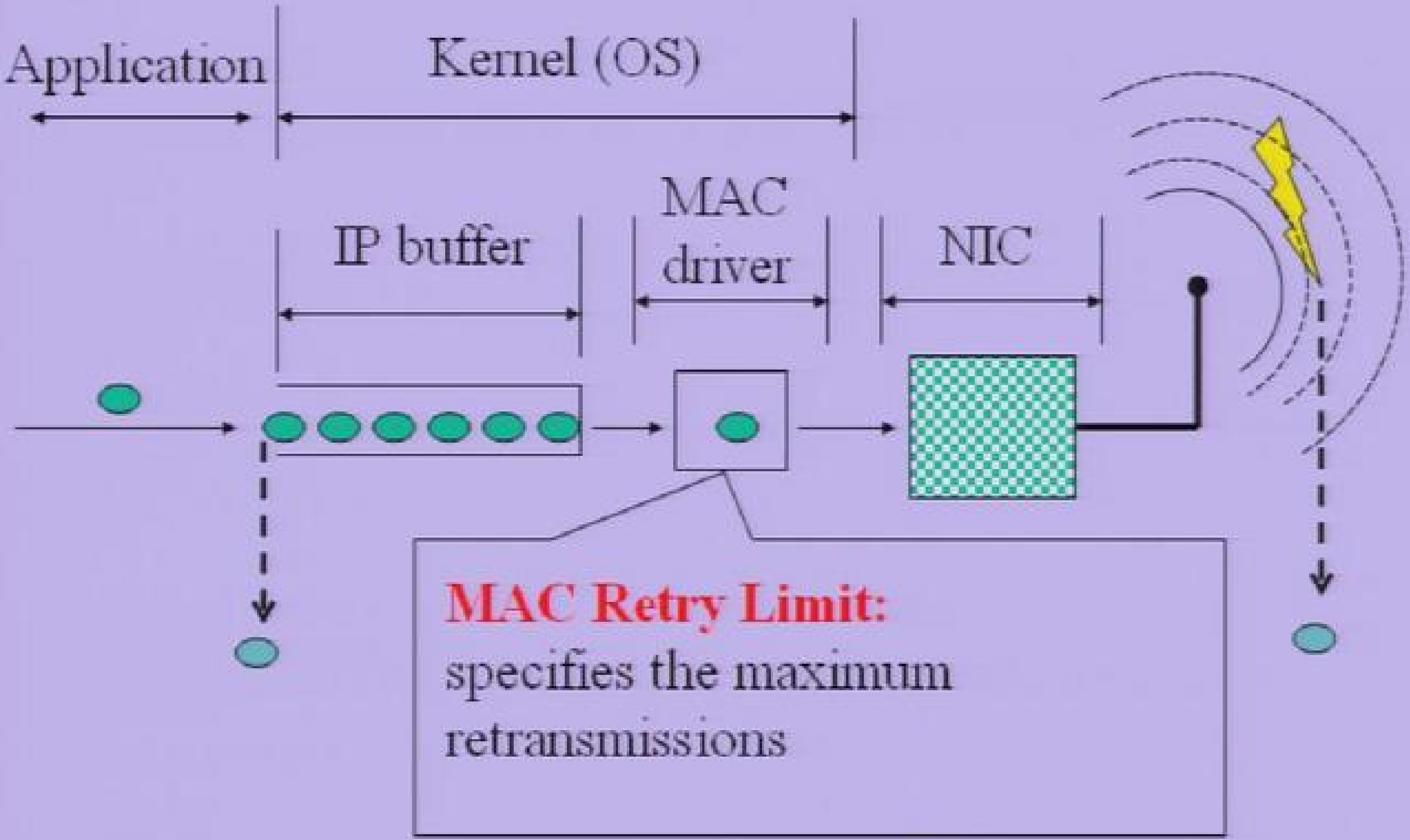
$$Throughput = \frac{L}{L+L_m^{header}} * (1-P_e^m(L))$$

Optimal packet-size determined by MAC:

$$L^* = \frac{-L_m^{header} + \sqrt{\left(L_m^{header}\right)^2 - \frac{L_m^{header}}{2\log(1-p_b^m)}}}{2}$$

p_b^m	PSNR for *L*= 500 bytes	PSNR for *L*=1000 bytes	PSNR for L^* determined by MAC
0.000006	32.86	30.65	27.90
0.000010	30.93	28.10	31.20
0.000030	28.76	25.43	26.86
0.000050	24.01	23.09	25.12

Implementation constraints affect cross-layer optimization



Application

Kernel (OS)

IP buffer

MAC driver

NIC

MAC Retry Limit:
specifies the maximum retransmissions

Fluid Model (similar results using M/G/1 model)

Video characteristics

- Constant arrival rate of multimedia packets λ

Channel characteristics

- Packet loss probability (at the PHY) without retransmissions P
- Service rate of the link C

- Link packet erasure rate (after *T* retransmissions) $p_L(T,P) = P^{T+1}$
- Mean number of transmissions $s(T,P) = \dfrac{1-P^{T+1}}{1-P}$
- Effective utilization factor of the link $\rho(P) = \lambda / C(1-P)$
- Buffer overflow rate $p_B(T,P) = \dfrac{\lambda s(T,P) - C}{\lambda s(T,P)} = 1 - \dfrac{1}{\rho(P)} \dfrac{1}{1-P^{T+1}}$

Fluid Model

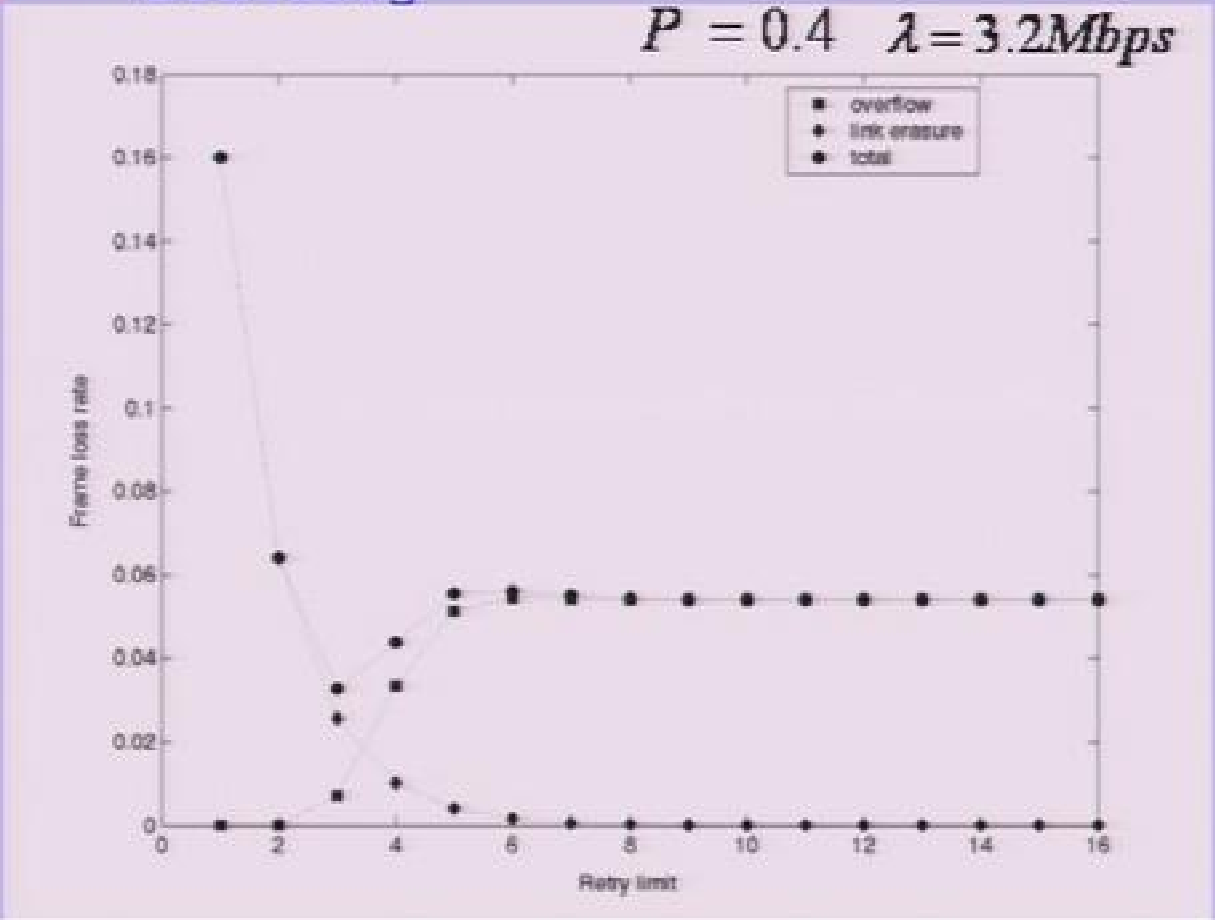
- The overall loss rate

$$p_T(T,P) = p_B(T,P) + p_L(T,P) = 1 - \frac{1}{\rho(P)}\frac{1}{1-P^{T+1}} + P^{T+1}$$

$P - fixed$

Monotonically Increasing

Monotonically Decreasing



$P = 0.4 \quad \lambda = 3.2 Mbps$

Fluid Model

- The overall loss rate

$$p_T(T,P) = p_B(T,P) + p_L(T,P) = 1 - \frac{1}{\rho(P)}\frac{1}{1-P^{T+1}} + P^{T+1}$$

$P - fixed$

Monotonically Increasing

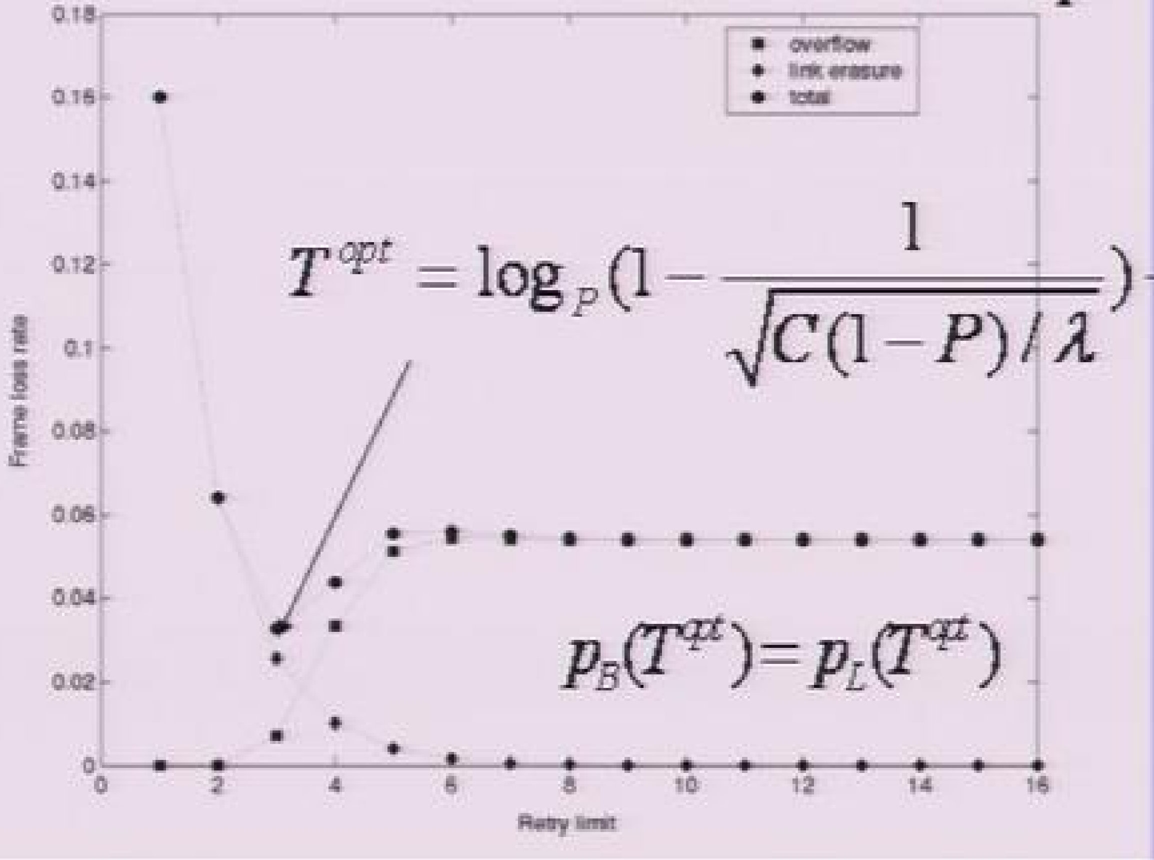
Monotonically Decreasing

Optimal retry limit

- multimedia traffic characteristics, channel conditions, Buffer sizes

⇒ **Requires Real-time Adaptation** [Li, vanderSchaar '03])

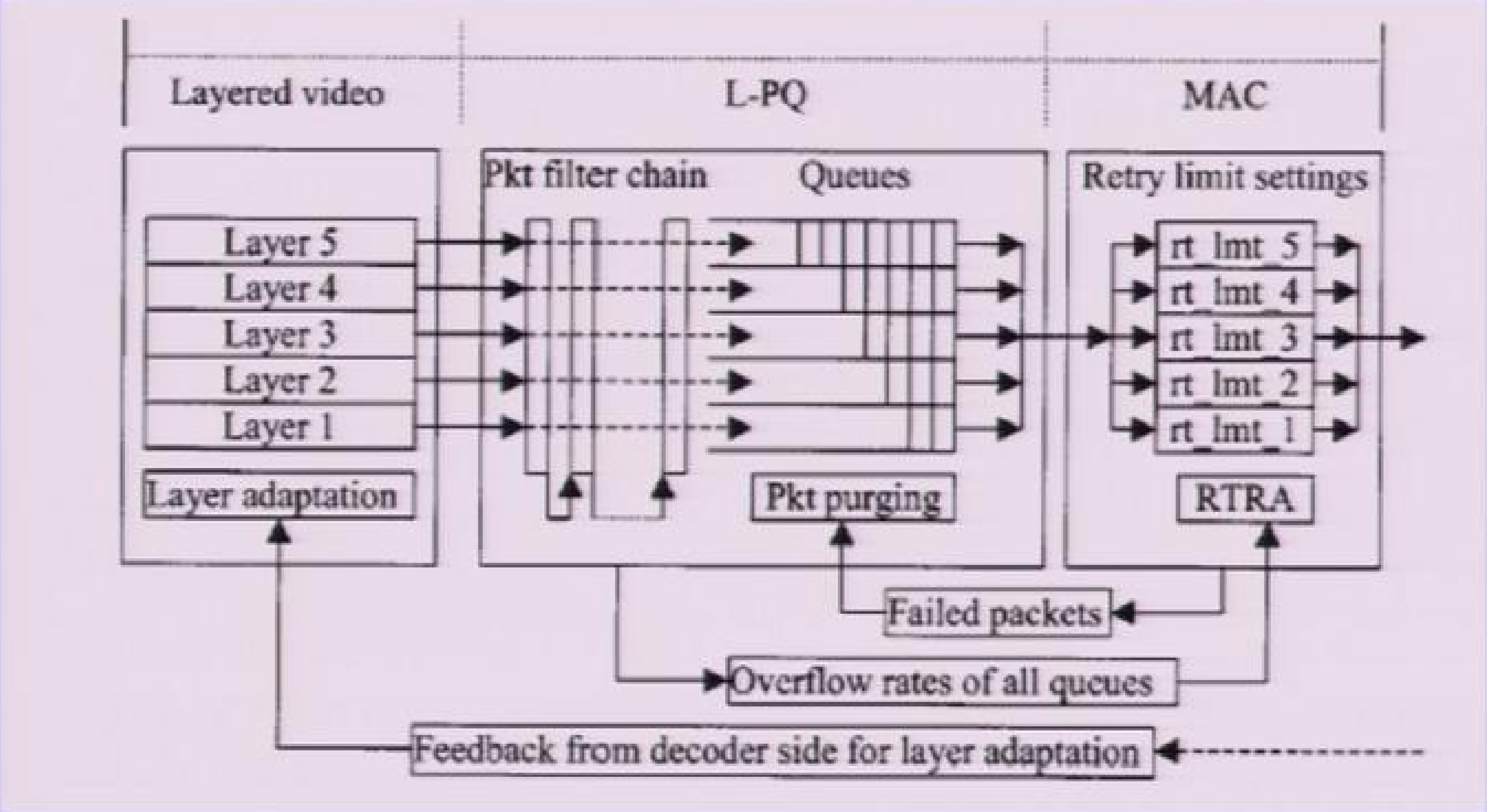
$P = 0.4 \quad \lambda = 3.2Mbps$



$$T^{opt} = \log_P\left(1 - \frac{1}{\sqrt{C(1-P)/\lambda}}\right) - 1$$

$$p_B(T^{opt}) = p_L(T^{opt})$$

How to perform cross-layer video optimization?



$$S^{opt}(\mathbf{x}, mc) = \arg\min_{S} D(S(\mathbf{x}), mc)$$

Application layer: Prioritization, Scheduling, Packet size
MAC: Retransmission

Joint Application-MAC cross-layer optimization

- Expected associated distortion

$$\overline{D}_{p,s} = P(succ) \times D_{p,s}^{Quant,R} + P(fail) D_{p,s}^{loss}$$

- Expected number of transmissions for any packet

$$\overline{T} = \sum_{t=1}^{T_{\max}+1} t p_L^{t-1}(1-p_L) + P(fail)(T_{\max}+1)$$

- Expected additional transmission rate (overhead)

$$\overline{R}_{p,s} = (\overline{T}-1) L_{p,s} + L^{Header}$$

- Cross-layer optimization problem

$$\left(T_{\max,s}^{opt}, L_s^{opt}\right) = \arg\min_{(T_{\max}, L)} \left[\sum_{p=1}^{P_s} \left(\overline{D}_{p,s} + \lambda \overline{R}_{p,s} \right) \right]$$

Joint Application-MAC cross-layer optimization

Retransmission limits for different priority packets $\mathbf{T} = \begin{bmatrix} T_1 & \dots & T_N \end{bmatrix}$

Average number of link retransmissions $\mathbf{s} = \begin{bmatrix} s_1(T_1, P) & \dots & s_N(T_N, P) \end{bmatrix}$

Departure rates from queues to the link (APP layer R-D scheduling)

$$\mathbf{\Lambda} = \begin{bmatrix} \Lambda_1 & \dots & \Lambda_N \end{bmatrix}$$

System-wide average packet retransmissions

$$\overline{s}(\mathbf{T}, P) = \frac{\mathbf{\Lambda} \cdot \mathbf{s}(\mathbf{T}, P)}{\mathbf{\Lambda} \cdot \mathbf{1}} \qquad \mathbf{1} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^T$$

Overflow rate of the multiqueue system $p_B(\mathbf{T}, P) = \dfrac{\lambda \overline{s}(\mathbf{T}, P) - C}{\lambda \overline{s}(\mathbf{T}, P)}$

Link erasure rate $p_L(\mathbf{T}, P) = P^{T+1}$

Joint Application-MAC cross-layer optimization

Retransmission limits for different priority packets $\mathbf{T}=\left[T_1 \quad \dots \quad T_N\right]$

Average number of link retransmissions $\mathbf{s}=\left[s_1(T_1,P) \quad \dots \quad s_N(T_N,P)\right]$

Departure rates from queues to the link (APP layer R-D scheduling)

$$\mathbf{\Lambda}=\left[\Lambda_1 \quad \dots \quad \Lambda_N\right]$$

System-wide average packet retransmissions

$$\overline{s}(\mathbf{T},P)=\frac{\mathbf{\Lambda}\cdot\mathbf{s}(\mathbf{T},P)}{\mathbf{\Lambda}\cdot\mathbf{1}} \qquad \mathbf{1}=\left[1 \quad 1 \quad \dots \quad 1\right]^T$$

Overflow rate of the multiqueue system $p_B(\mathbf{T},P)=\dfrac{\lambda\overline{s}(\mathbf{T},P)-C}{\lambda\overline{s}(\mathbf{T},P)}$

Link erasure rate $p_L(\mathbf{T},P)=P^{T+1}$

MAC shadow retry limit -> retransmission limit vector **Tsrl**

Actual retransmission limit vector **Tre** (with unequal elements)

Iterative algorithm [Li and vanderSchaar '03] for computing **Tre**

Subjective video quality experiment

- We evaluate the impact of these strategies on the perceived video quality by performing a visual experiment according to CCIR Recommendation 500-4
- selected five scales are:
    - very annoying (1),
    - annoying (2),
    - slightly annoying (3),
    - perceptible but not annoying (4),
    - imperceptible (5).

Deployed strategies	**Visual Score**
No optimization at MAC & App.	1.4
MAC layer optimization (RTRO)	1.9
Application layer optimization	3.8
Joint Application-MAC cross-layer optimization	4.6

Cross-layer solutions – new solution needed

Cross-layer solutions – new solution needed

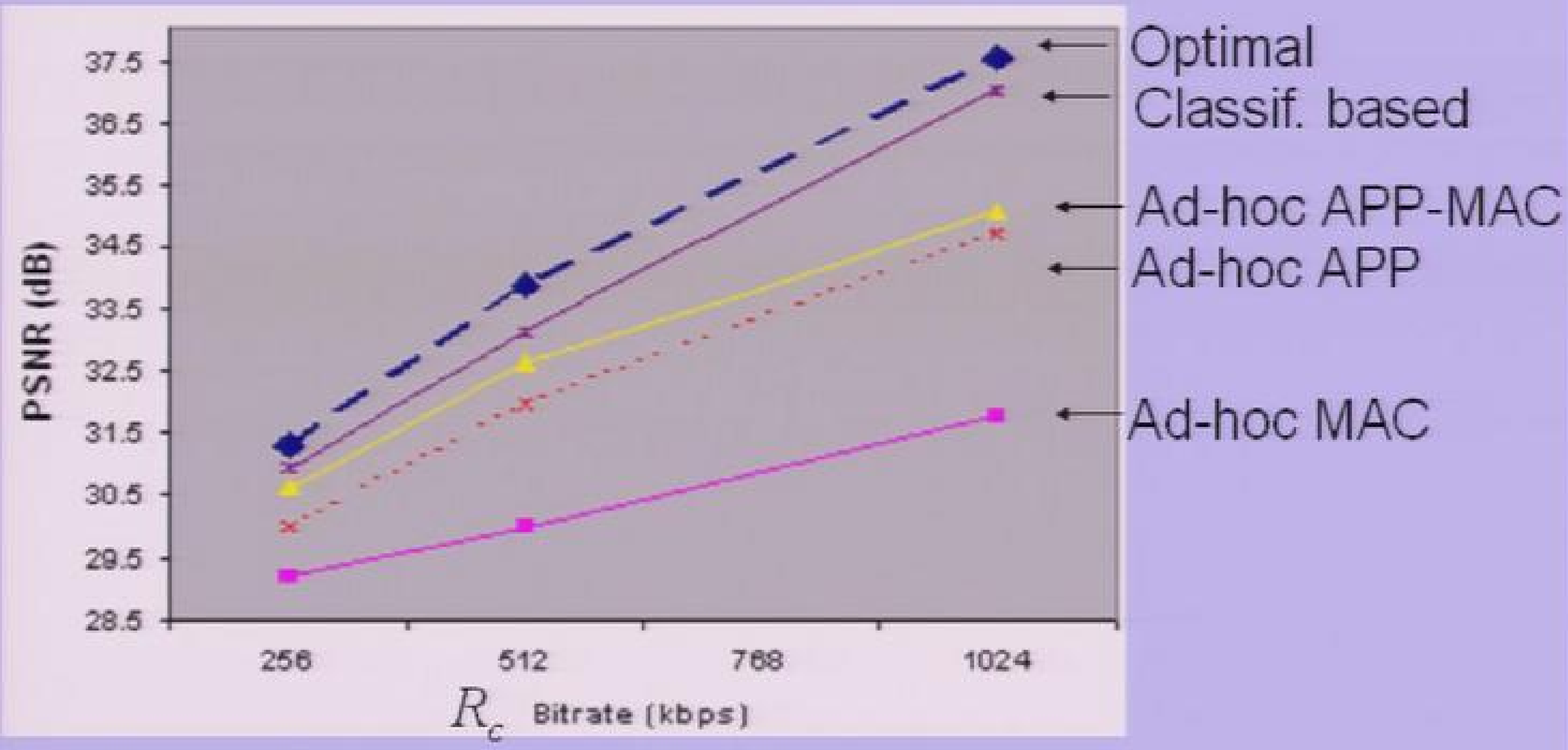
- Integrated optimization approach – very complex -> unsuitable for real-time multimedia

- Current solutions: ad-hoc heuristics

- Our new approach [Wong, vanderSchaar, Turaga '05]
  - Determine OFFLINE optimal cross-layer solution for classes of content, channel conditions, protocol implementation
  - Use ON-LINE classification techniques to choose the optimized solution
    - Video and Channel features -> Strategy choices
  - This de-facto solution can be used as is or further improved (i.e. serve only as initialization) -> Learn on the fly new, improved solutions
  - Another advantage: user subjective metrics (not PSNR) can be used

Cross-layer results using classification

$$P(fail) = 0.1$$



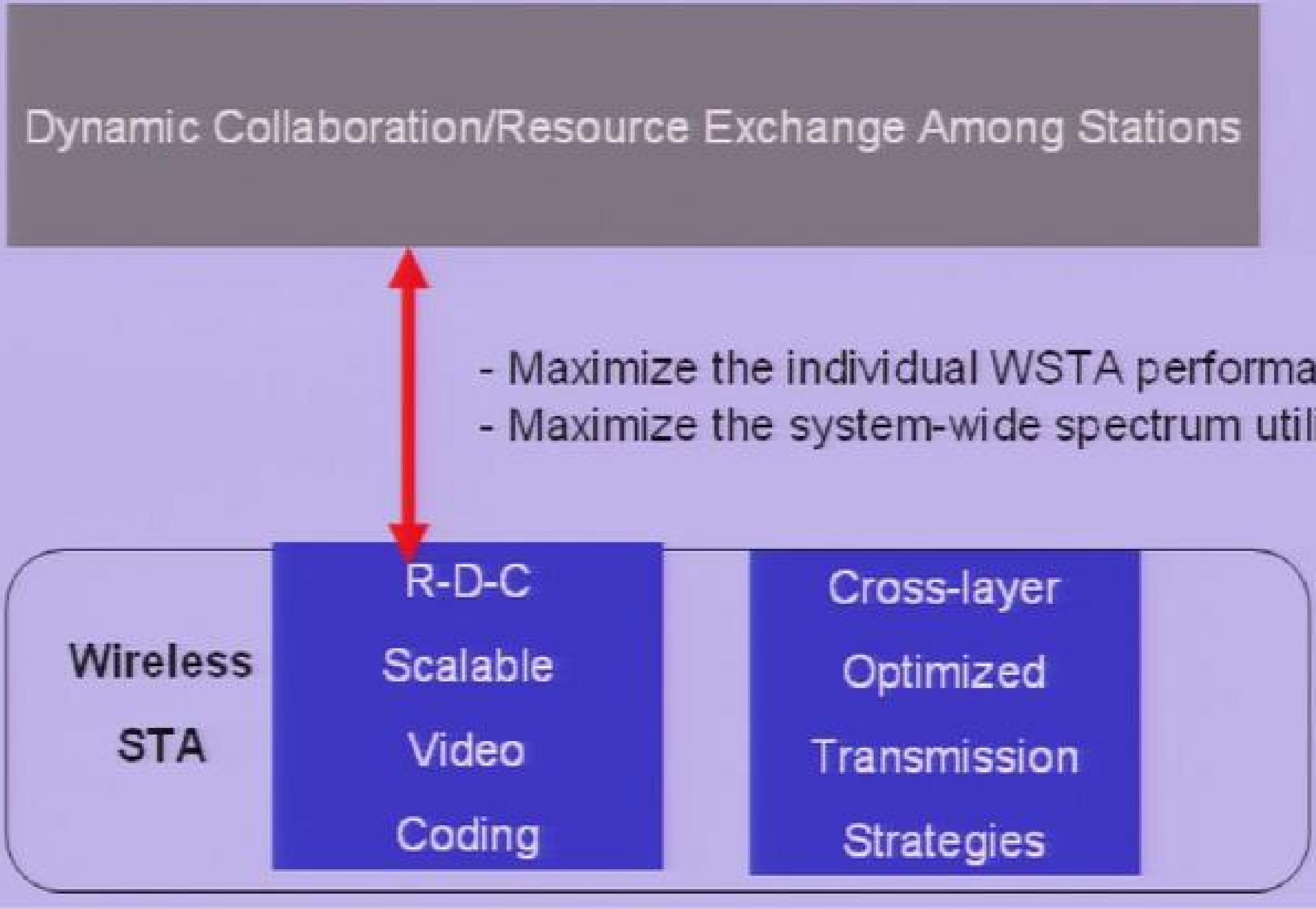
Future research

- Wireless and Internet multimedia communication with Resource and Information Exchanges
- Multimedia compression and communication over OSAR
- Content-Aware Multi-camera systems

- Formal Methods for Designing and Optimizing Multimedia algorithms on resource-constrained (embedded) systems

Collaborative framework for wireless multimedia

Goal: Construct a system, where users can borrow or lend resources from the system/other users, according to their specific utility and resource awareness.



Dynamic Collaboration/Resource Exchange Among Stations

- Maximize the individual WSTA performance and
- Maximize the system-wide spectrum utilization

Wireless STA

R-D-C Scalable Video Coding

Cross-layer Optimized Transmission Strategies

New proactive framework for wireless multimedia

- **Fairness based on contention resolution protocols** [vanderschaar, Shankar '05]
  - Workload- Generalized Processor Scheduler [Gallager]

$$\frac{W_i(t_1,t_2)}{W_j(t_1,t_2)} \ge \frac{\phi_i}{\phi_j}, j = 1, 2, \ldots, \boxed{n} \longleftarrow \text{Number of WSTAs}$$

GPS advantages (guaranteed throughput, independent service) cannot be preserved if WSTAs use different cross-layer optimization strategies

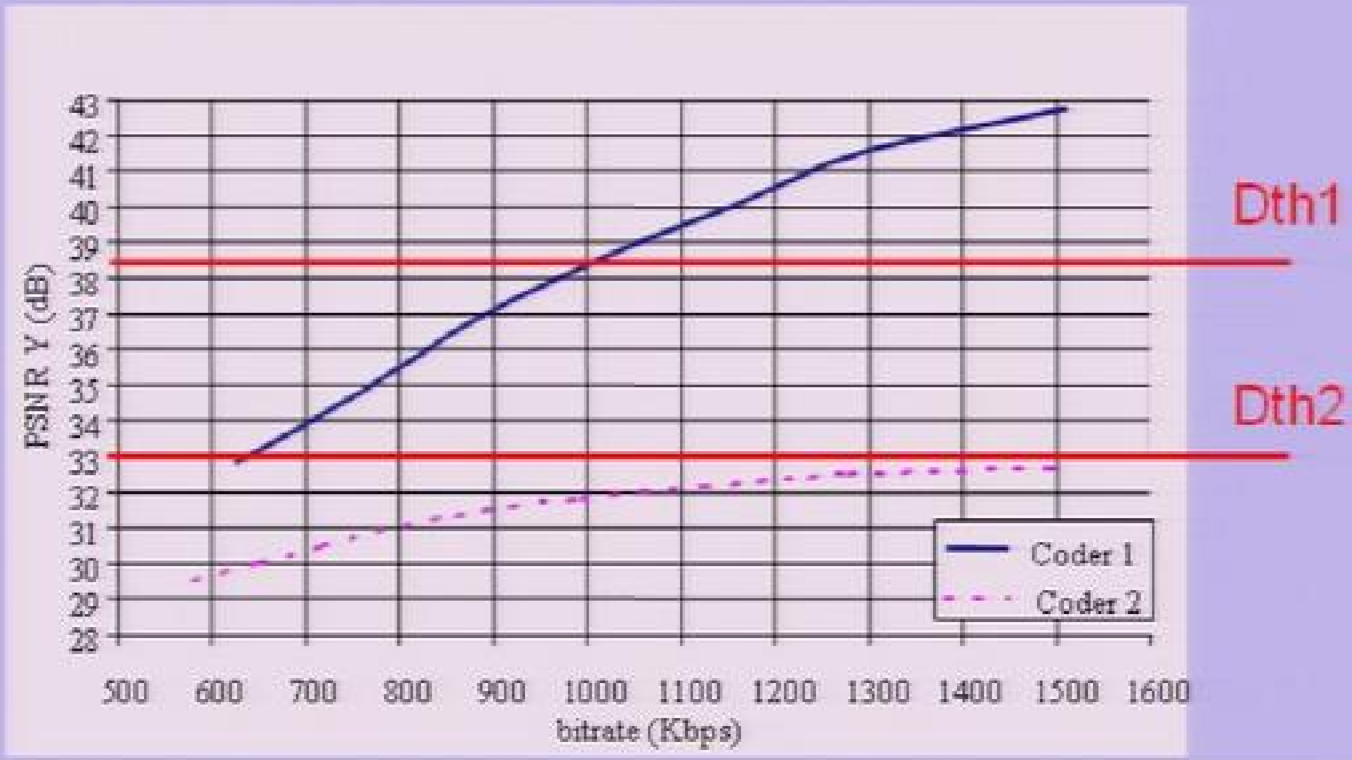
  - Air-Time Fairness
  - For multimedia: Delay or Distortion Fairness

Distortion Fair Scheduling

$$\frac{D_i(t_1,t_2)}{\phi_i} \geq \frac{D_j(t_1,t_2)}{\phi_j}$$

How to Provide Distortion Fair transmission time?

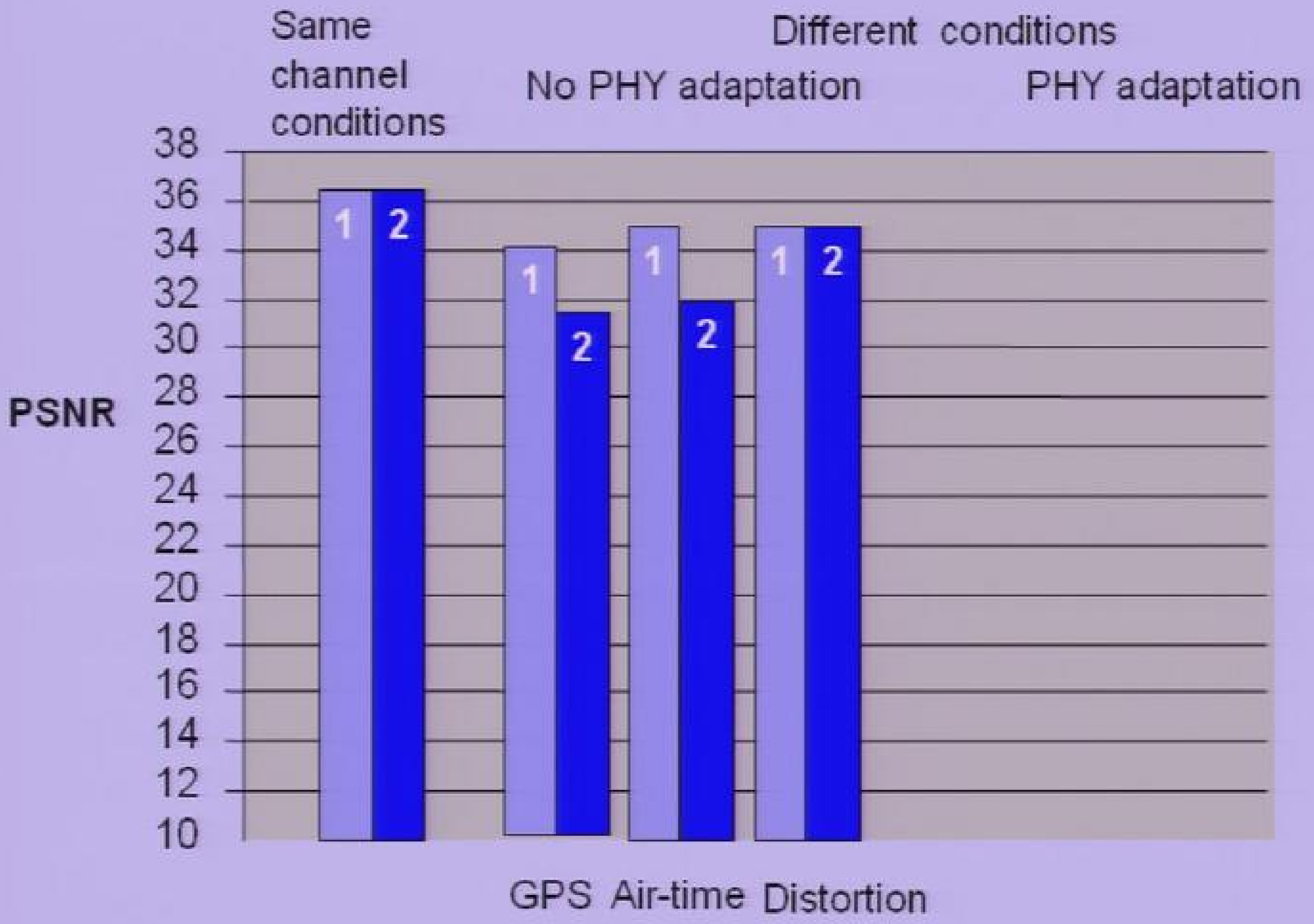
- Rate-Distortion models needed
- Use equal distortion, or different quality "levels"



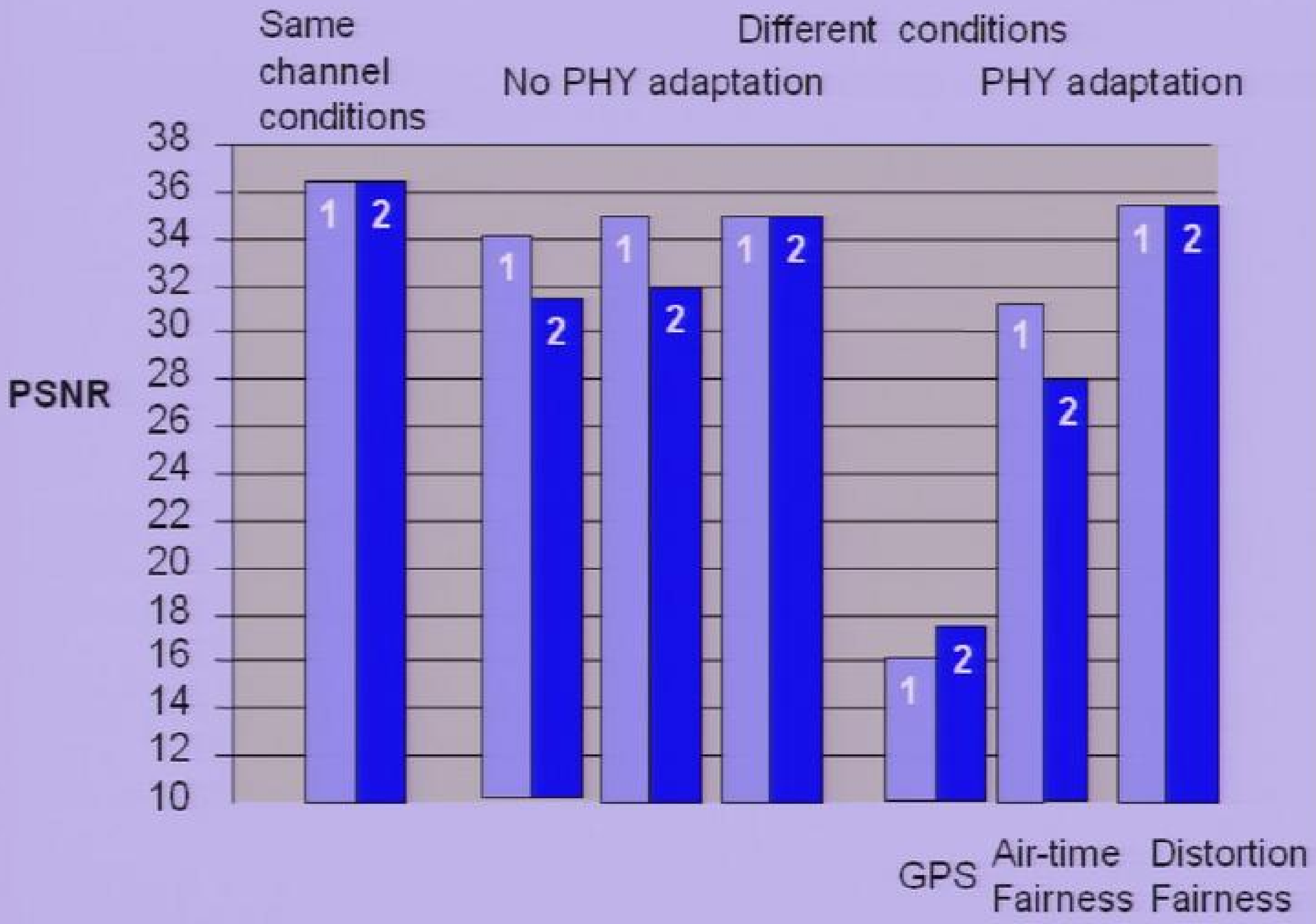
WSTAs interaction – Fairness

Football (SD – resolution, 30 Hz), 2 Mbps, 400ms Delay

WSTAs interaction – Fairness



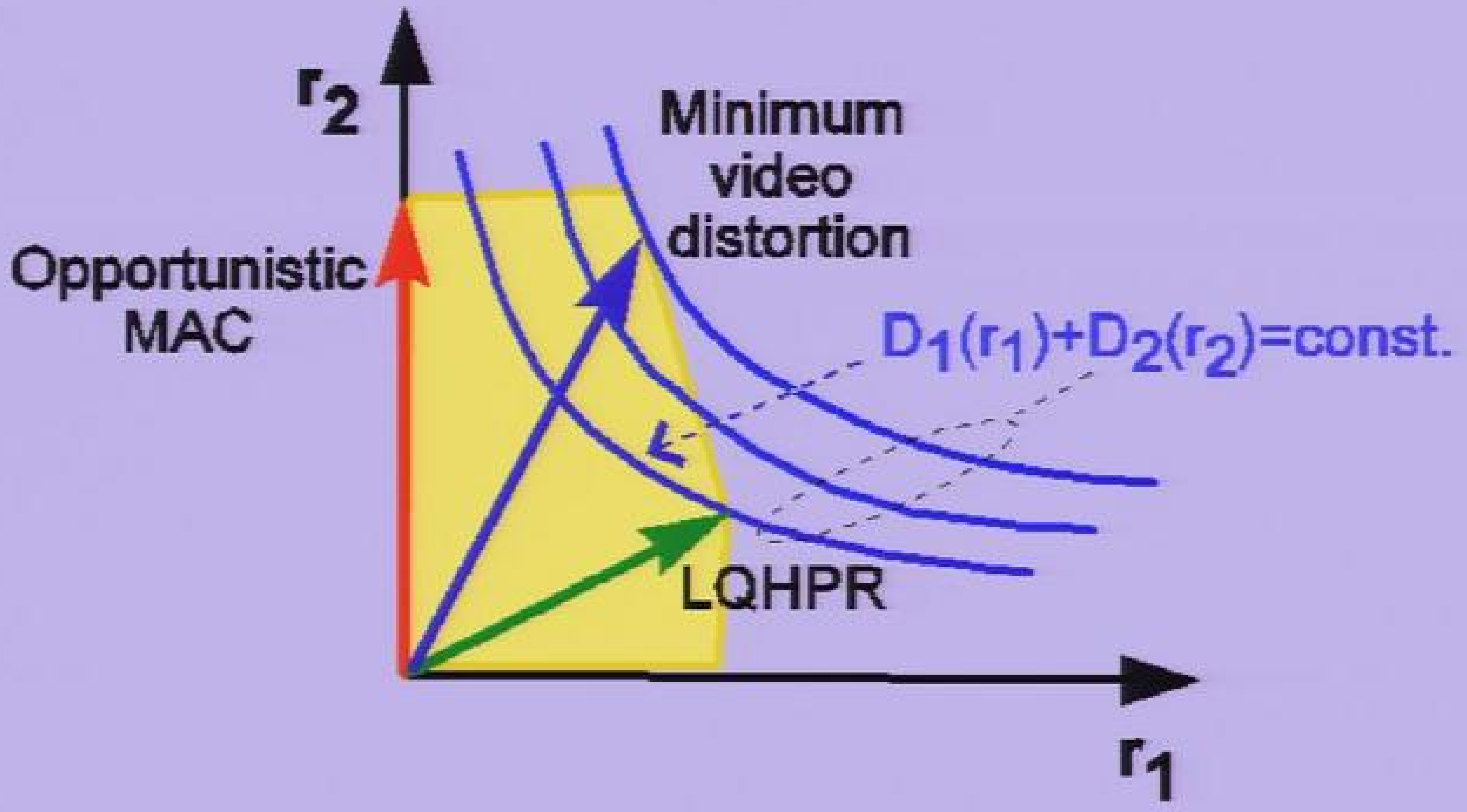
WSTAs interaction – Fairness



Do conventional information theory results for wireless communications hold for multimedia?

[Scaglione, vanderSchaar '05]

Opportunistic MAC or Longest Queue Highest Rate?



New proactive framework for wireless multimedia

- Fair resource management, but passive resource allocation

New proactive framework for wireless multimedia

- **Fair resource management, but passive resource allocation**
- **Proactive resource management based on coopetition among WSTAs** [NSF Career] **–**

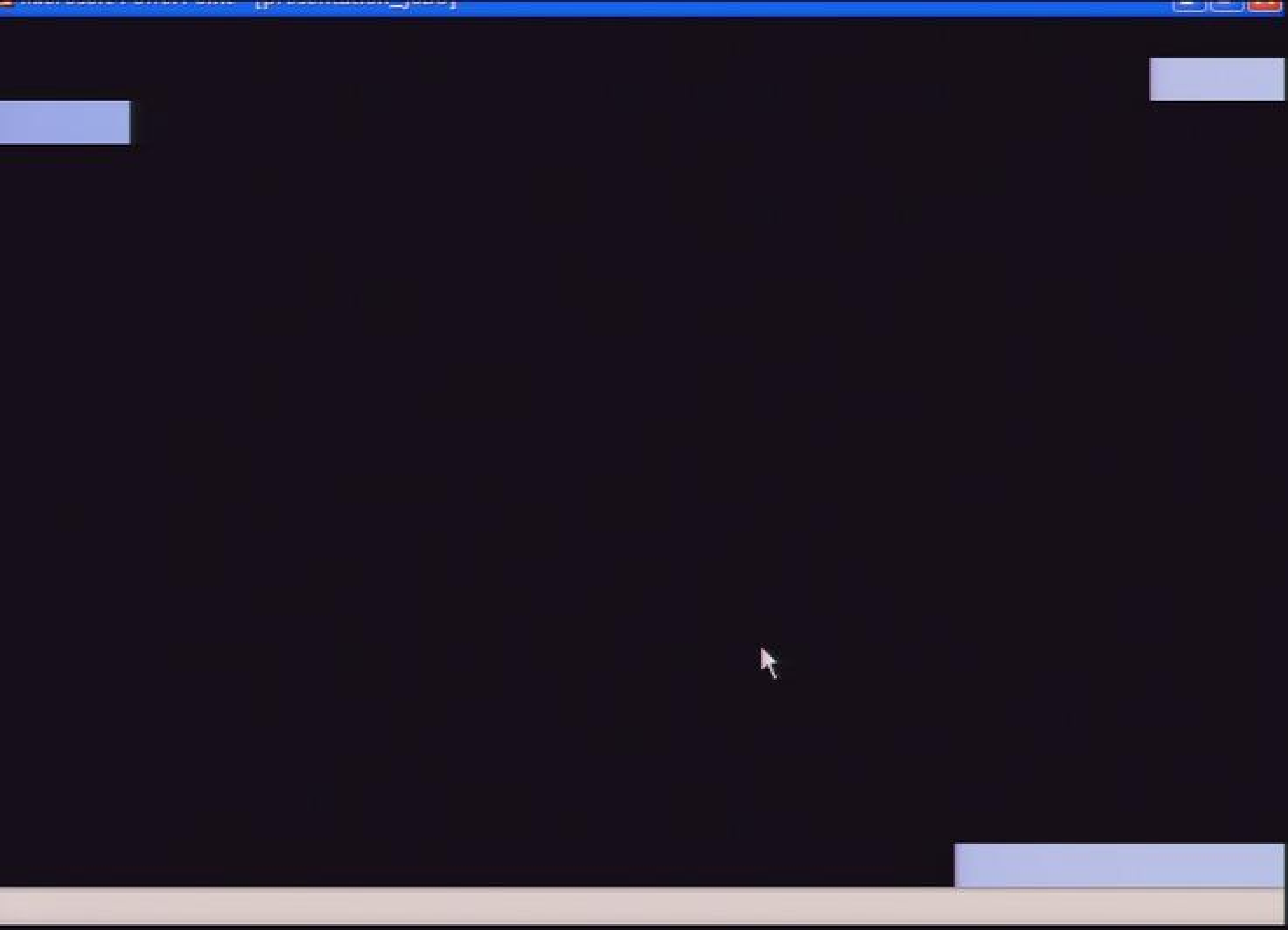
  borrows ideas from **on-line algorithms, game theory**
  - Wireless multimedia - game played between the competing WSTAs with no, partial or full information and different utility-cost functions
  - Significant improvements in quality and system resource utilization possible [Larcher, vanderschaar '04][sood, vanderschaar '05]

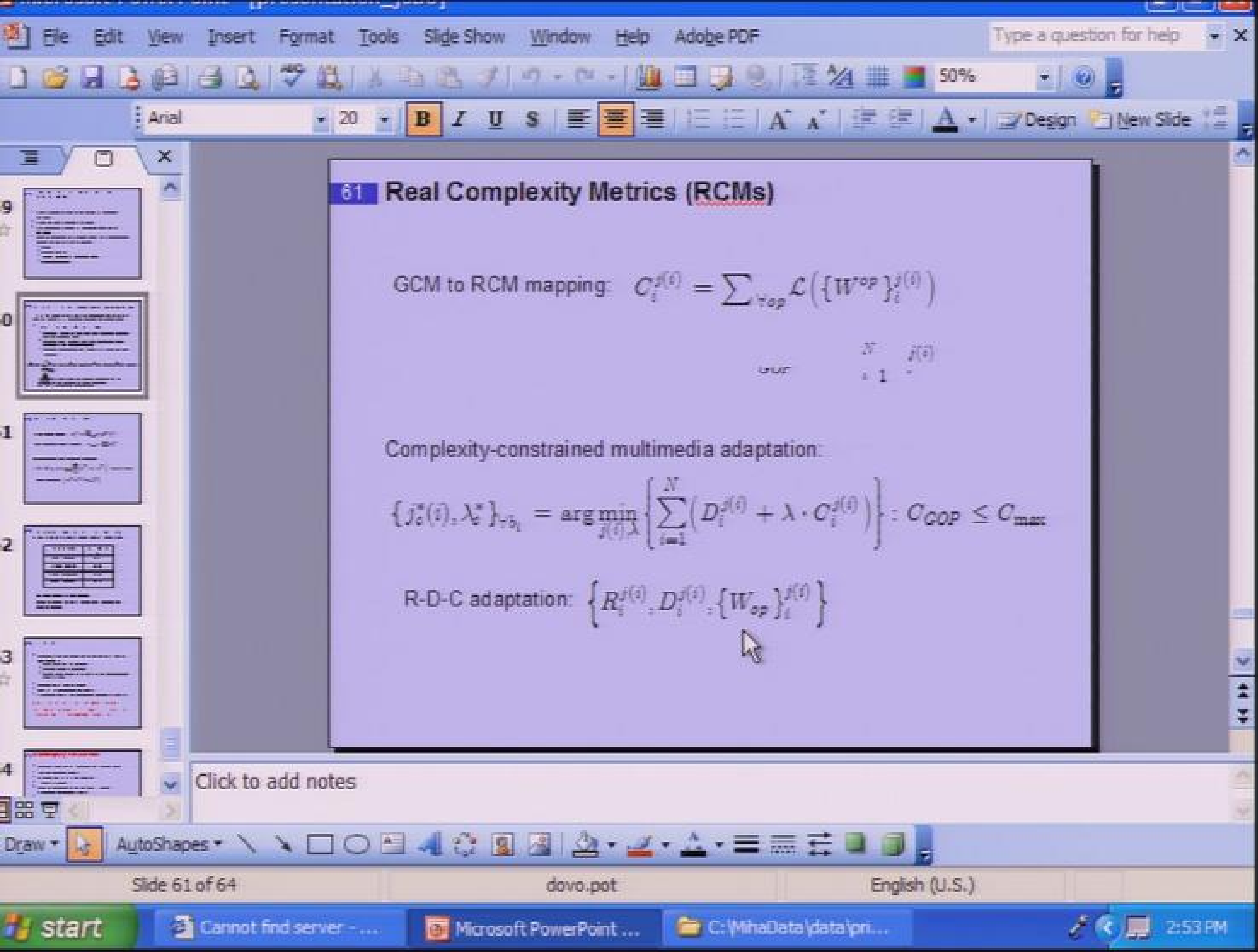
Formal Methods for Designing and Optimizing Multimedia Systems

Limitation of existing approaches

- Systems are designed based on worst-case scenarios for multimedia
- System layer currently *does not cooperate* with the multimedia applications to achieve optimal R-D-C tradeoffs
- Currently only ad-hoc solutions for R-D-C optimization
- Coarse levels of multimedia complexity (profiles)

Formal methods for multimedia on resource-constrained devices





61

Real Complexity Metrics (RCMs)

GCM to RCM mapping: $C_i^{j(i)} = \sum_{\forall op} \mathcal{L}\left(\{W^{op}\}_i^{j(i)}\right)$

GOP $\quad {}^{N}_{i\;1} \quad {}^{j(i)}$

Complexity-constrained multimedia adaptation:

$$\{j_c^*(i), \lambda_c^*\}_{\forall b_i} = \arg\min_{j(i),\lambda} \left\{ \sum_{i=1}^{N} \left(D_i^{j(i)} + \lambda \cdot C_i^{j(i)} \right) \right\} : C_{GOP} \le C_{\max}$$

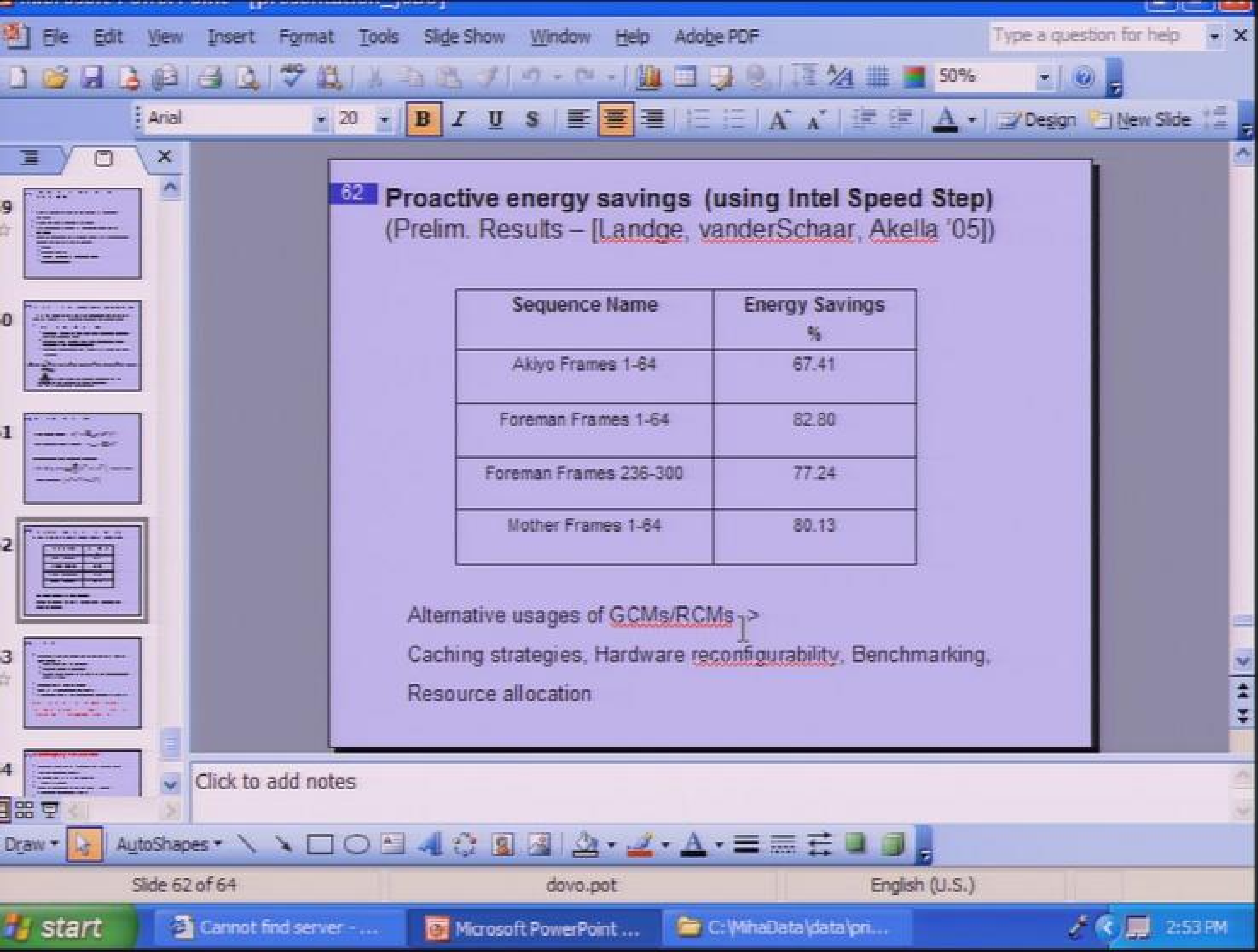
R-D-C adaptation: $\left\{ R_i^{j(i)}, D_i^{j(i)}, \{W_{op}\}_i^{j(i)} \right\}$

Click to add notes

Slide 61 of 64

dovo.pot

English (U.S.)



Proactive energy savings (using Intel Speed Step)

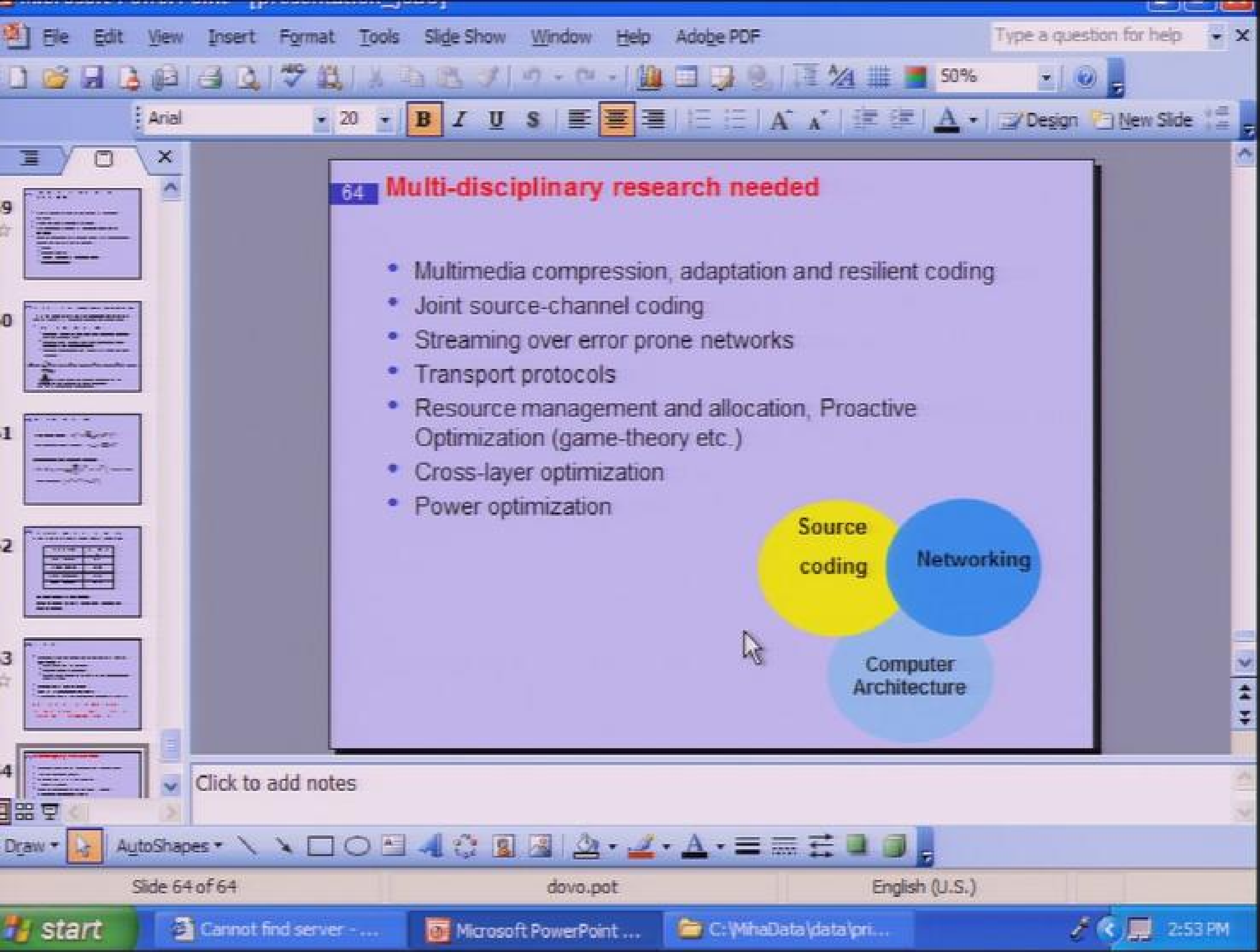
(Prelim. Results – [Landge, vanderSchaar, Akella '05])

Sequence Name	Energy Savings %
Akiyo Frames 1-64	67.41
Foreman Frames 1-64	82.80
Foreman Frames 236-300	77.24
Mother Frames 1-64	80.13

Alternative usages of GCMs/RCMs ->

Caching strategies, Hardware reconfigurability, Benchmarking,

Resource allocation



64

Multi-disciplinary research needed

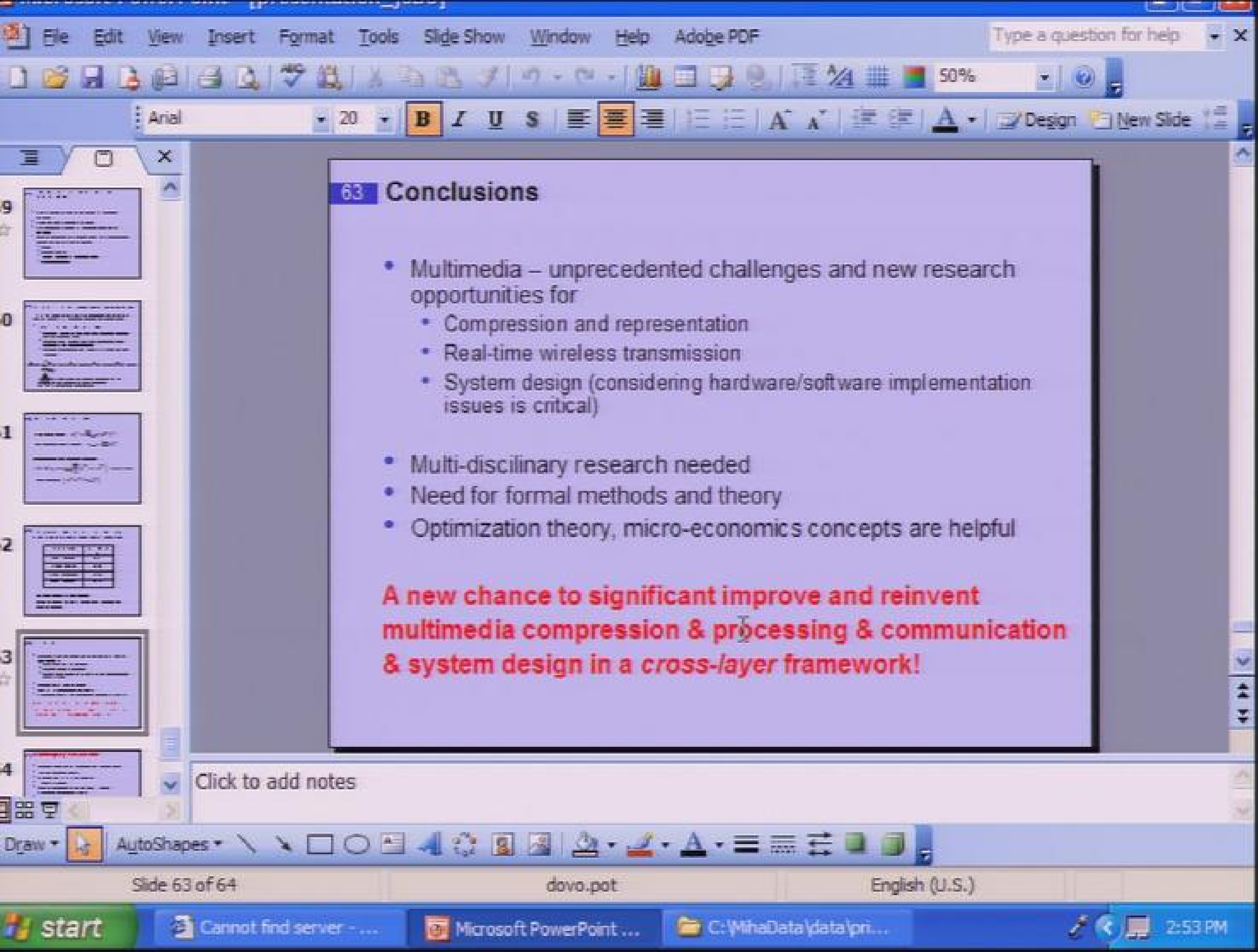
- Multimedia compression, adaptation and resilient coding
- Joint source-channel coding
- Streaming over error prone networks
- Transport protocols
- Resource management and allocation, Proactive Optimization (game-theory etc.)
- Cross-layer optimization
- Power optimization

Source coding

Networking

Computer Architecture

Click to add notes



63

Conclusions

- Multimedia – unprecedented challenges and new research opportunities for
  - Compression and representation
  - Real-time wireless transmission
  - System design (considering hardware/software implementation issues is critical)
- Multi-discilinary research needed
- Need for formal methods and theory
- Optimization theory, micro-economics concepts are helpful

A new chance to significant improve and reinvent multimedia compression & processing & communication & system design in a *cross-layer* framework!

Conclusions

Conclusions

- Multimedia – unprecedented challenges and new research opportunities for
    - Compression and representation
    - Real-time wireless transmission
    - System design (considering hardware/software implementation issues is critical)

Conclusions

- Multimedia – unprecedented challenges and new research opportunities for
  - Compression and representation
  - Real-time wireless transmission
  - System design (considering hardware/software implementation issues is critical)

- Multi-discilinary research needed
- Need for formal methods and theory
- Optimization theory, micro-economics concepts are helpful

A new chance to significant improve and reinvent multimedia compression & processing & communication & system design in a *cross-layer* framework!

Preparing to stand by...