# Recent, Current and Future Developments in Video Coding

*Jens-Rainer Ohm*

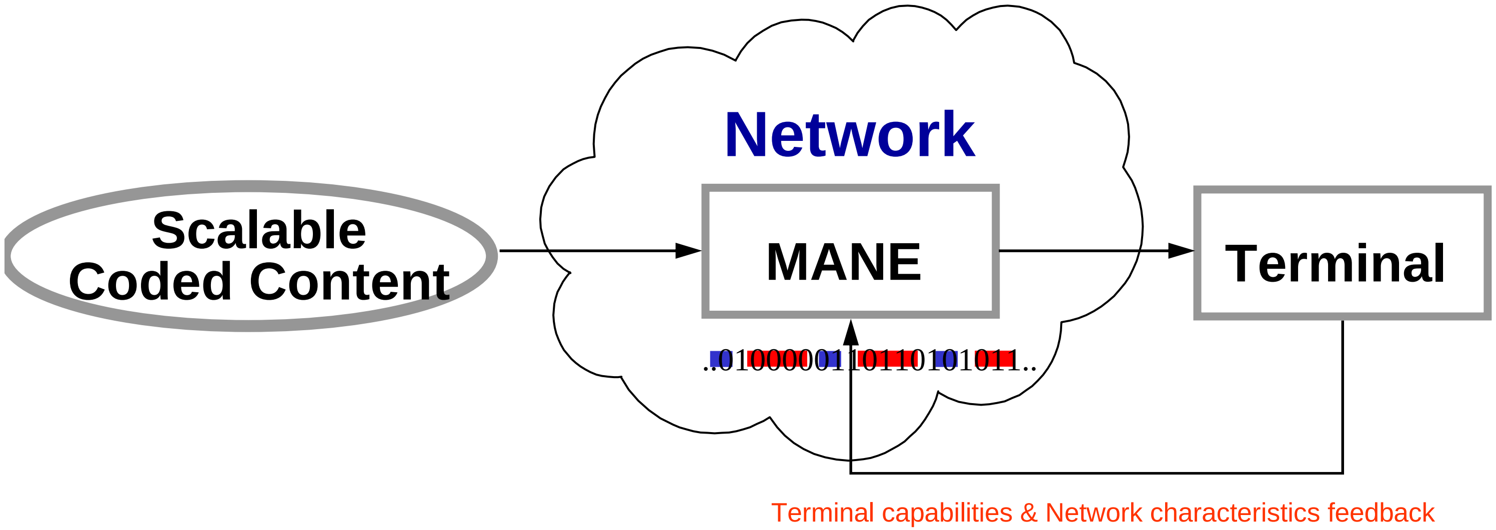
Institut für Nachrichtentechnik
RWTH Aachen University

# Outline

- Recent and current activities in MPEG Video and JVT
  - Scalable Video Coding
  - Multiview Video Coding and 3D Video
  - MPEG-C Reconfigurable Video Coding
- The Future: Increasing Compression Performance?
  - Will Video data rates and formats continue increasing?
  - Examples of tools for improved compression

# Scalable Video Coding

- Universal Media Access:
  code once and then customize the stream to access content
  - **“Anytime”**
  - from **“Anywhere”** (i.e. using any access network - wireless, internet etc.)
  - and by **“Anyone”** (i.e. with any terminal complexity)
- Compatibility of different formats/resolutions

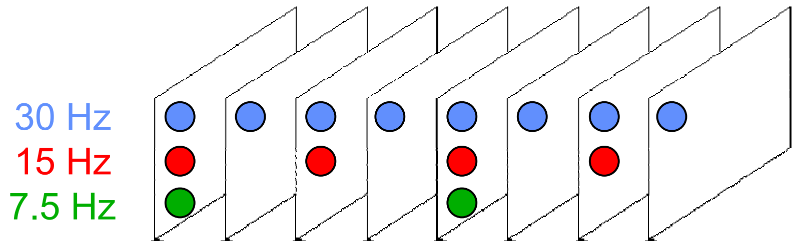





J.-R. Ohm · Institut für Nachrichtentechnik · RWTH Aachen University

# Scalability of Video - Modalities

- **Temporal: change of frame rate**




- **Spatial: change of frame size**

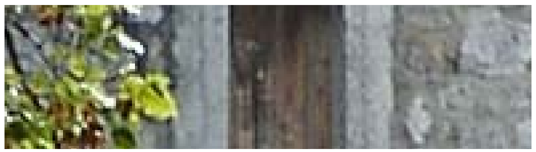

QCIF

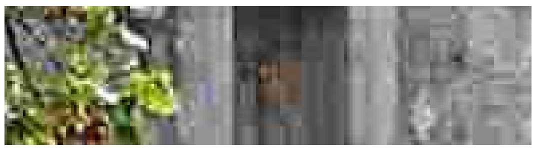

CIF


TV


HDTV

- **Fidelity: change of quality (a.k.a. SNR)**


High rate


Low rate

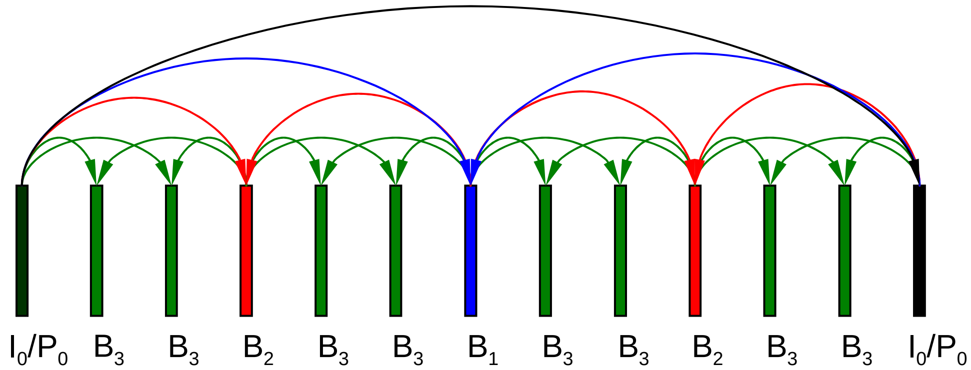
## Example: Format Enhancements in Mobile TV

- Spatial and temporal format extension
- Introduction of QVGA services

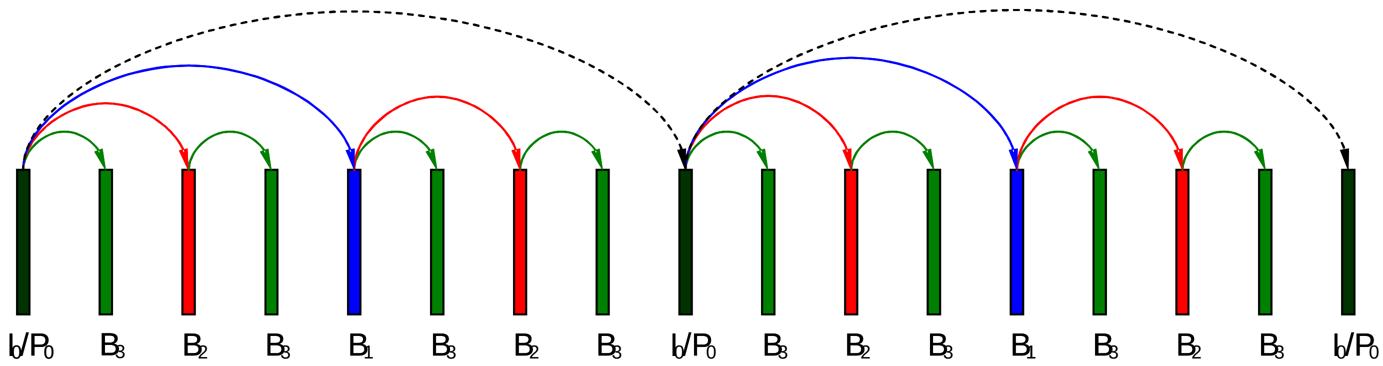
- Backwards-compatible introduction of VGA services

## Hierarchical prediction structures

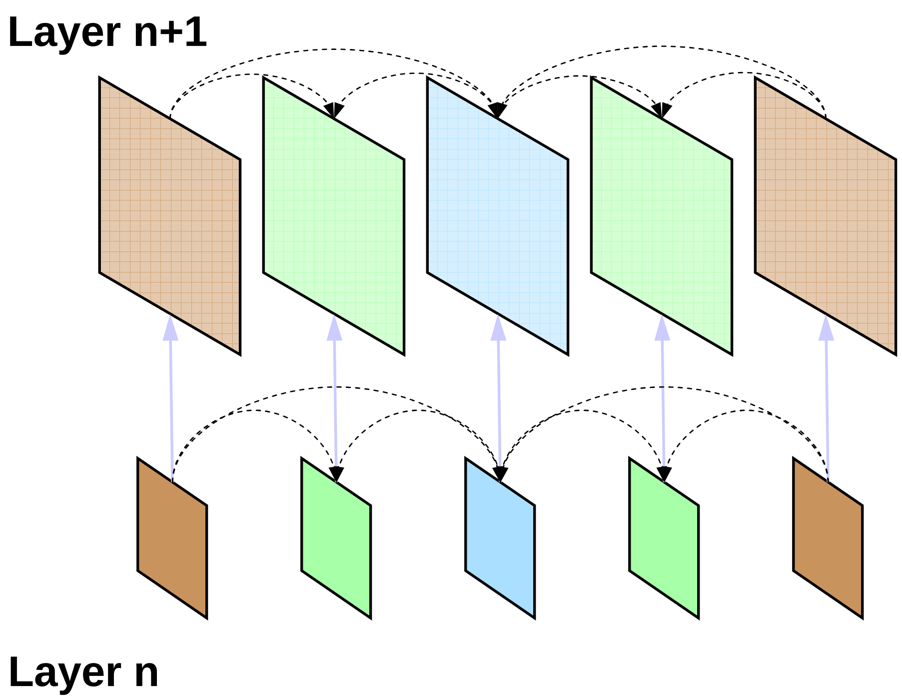
- Non-dyadic temporal scalability




- Low-delay prediction structure (structural delay is 0)



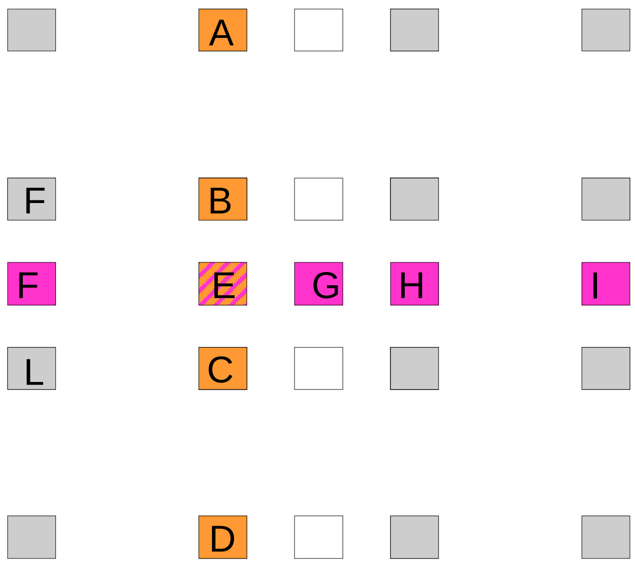
# Layered Coding for Spatial Scalability



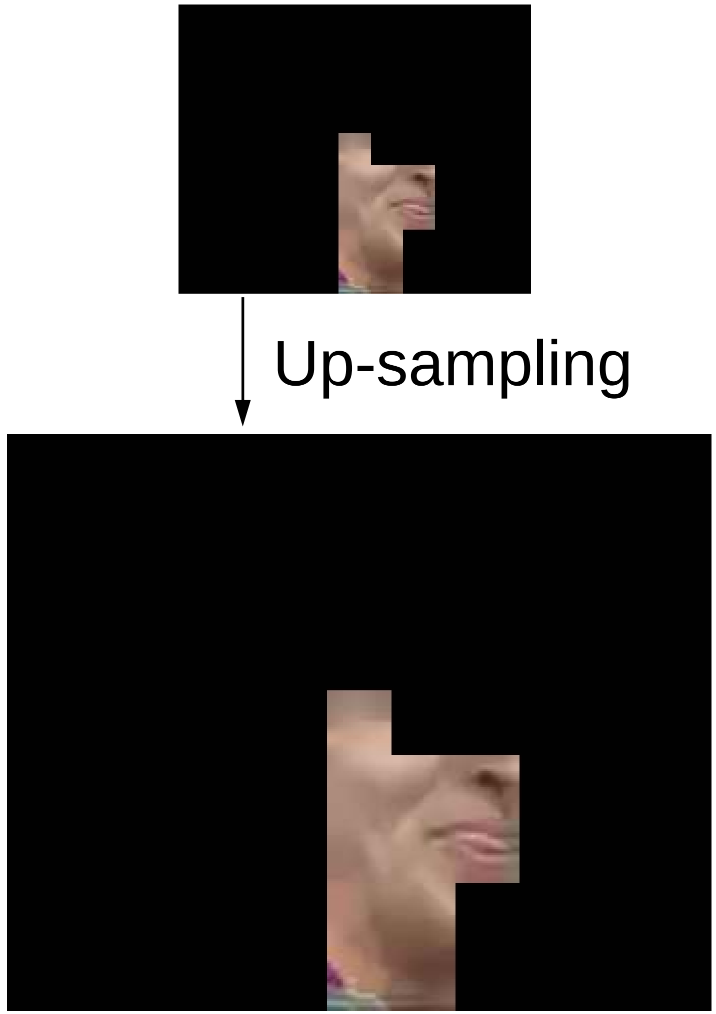

- Layered coding
  - Oversampled pyramid for each resolution:
    e.g. QCIF, CIF, 4CIF, 16CIF
  - MC prediction structures of all layers are aligned
- Inter-layer prediction: Switchable prediction (with upsampling)
  - Prediction of intra macroblocks
    Prediction of partitioning and motion information
  - Prediction of residual data

# Intra Prediction

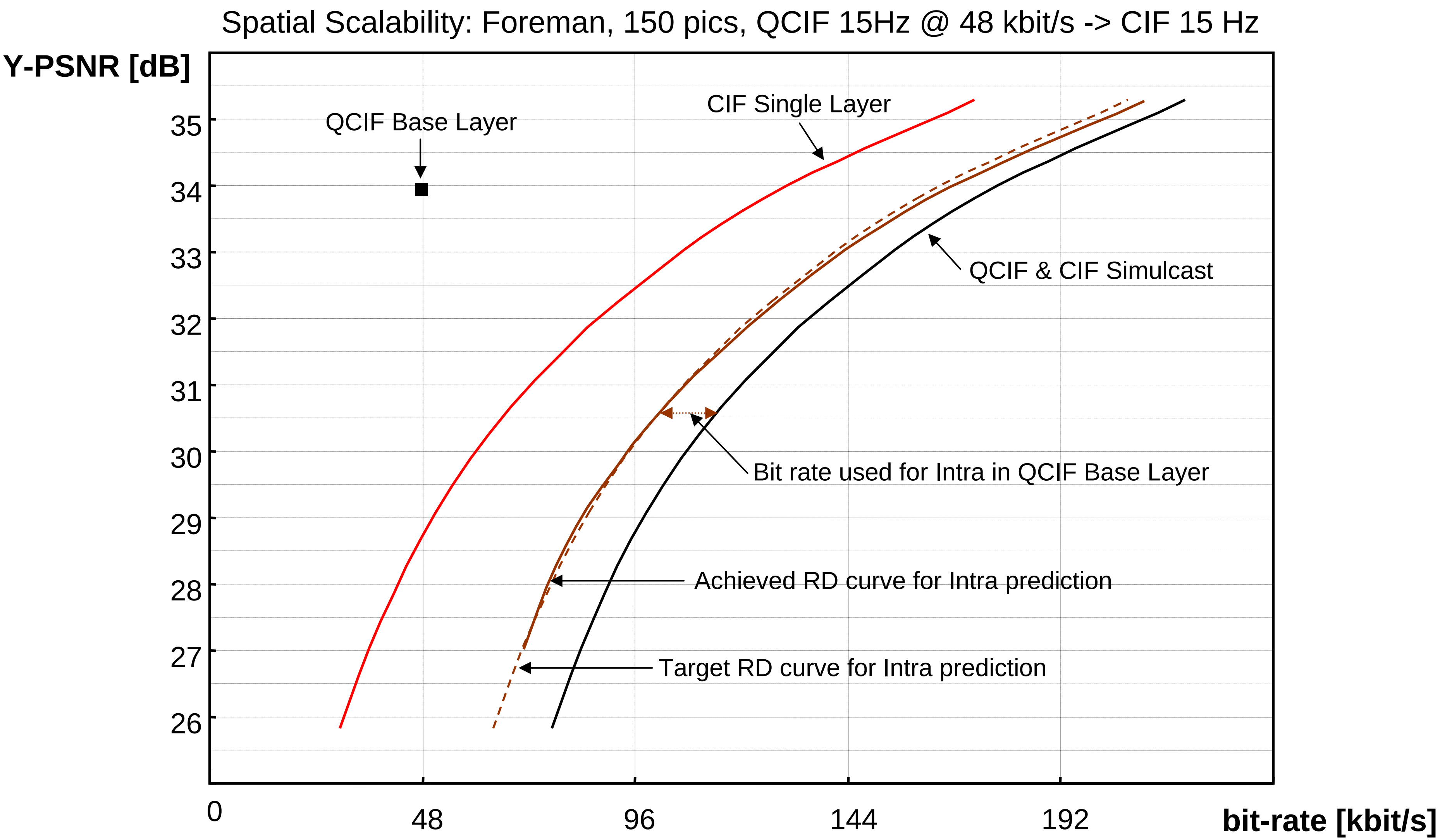
## 4-tap filter: [-3, 19, 19, -3]




full sample positions
half sample positions

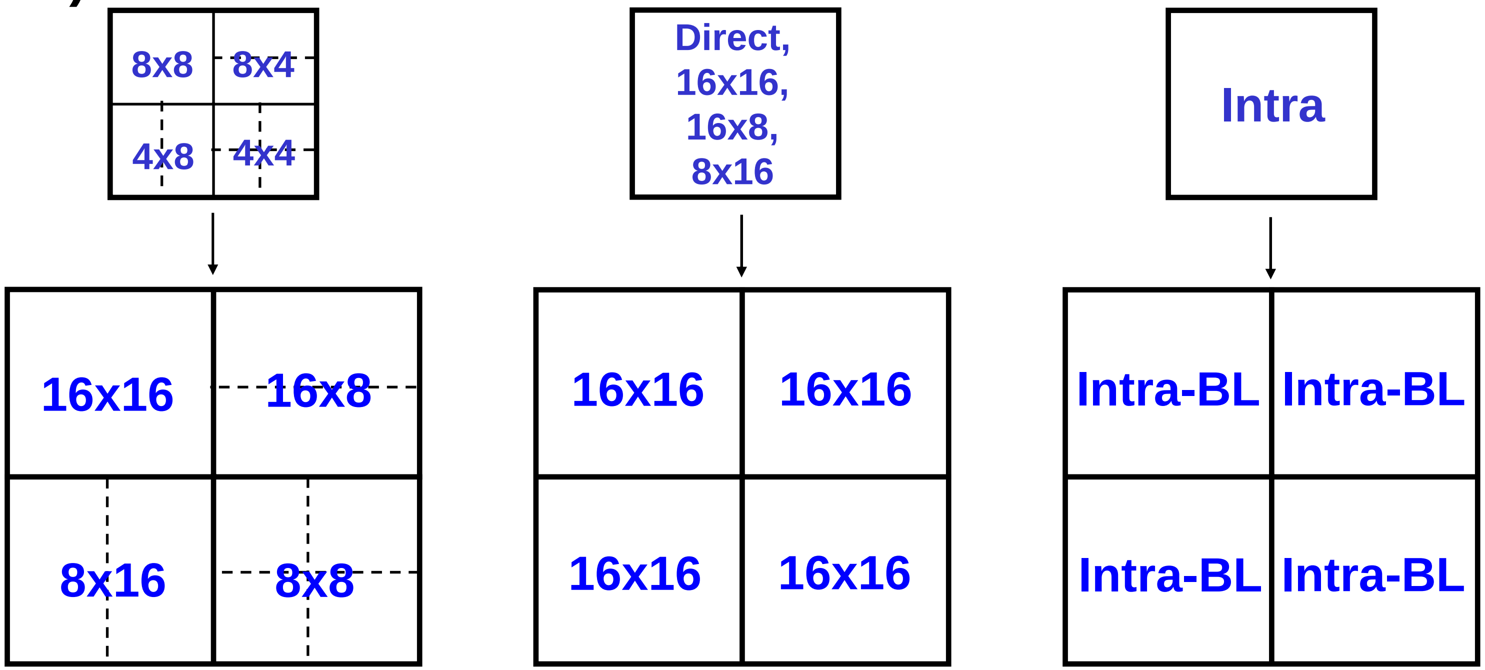


# Spatial Prediction of Data



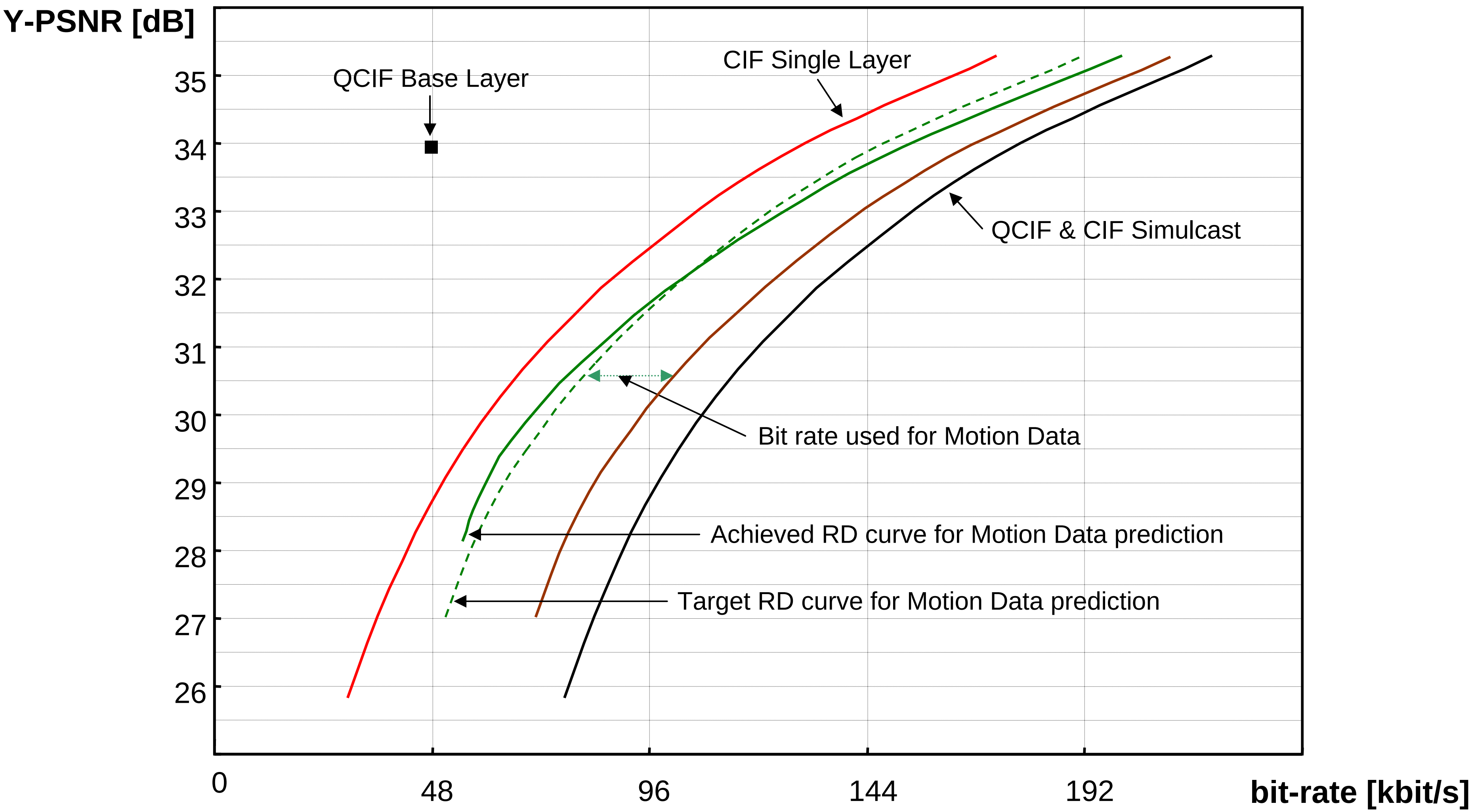
## Spatial Prediction of Motion Data

- Upsample macroblock partitioning as switchable partitioning predictor
- Multiply motion vectors by 2 and use them as switchable predictors
  (keep list 0, list 1, bi-predictive and reference indices information)



# Spatial Prediction of Data

Spatial Scalability: Foreman, 150 pics, QCIF 15Hz @ 48 kbit/s -> CIF 15 Hz

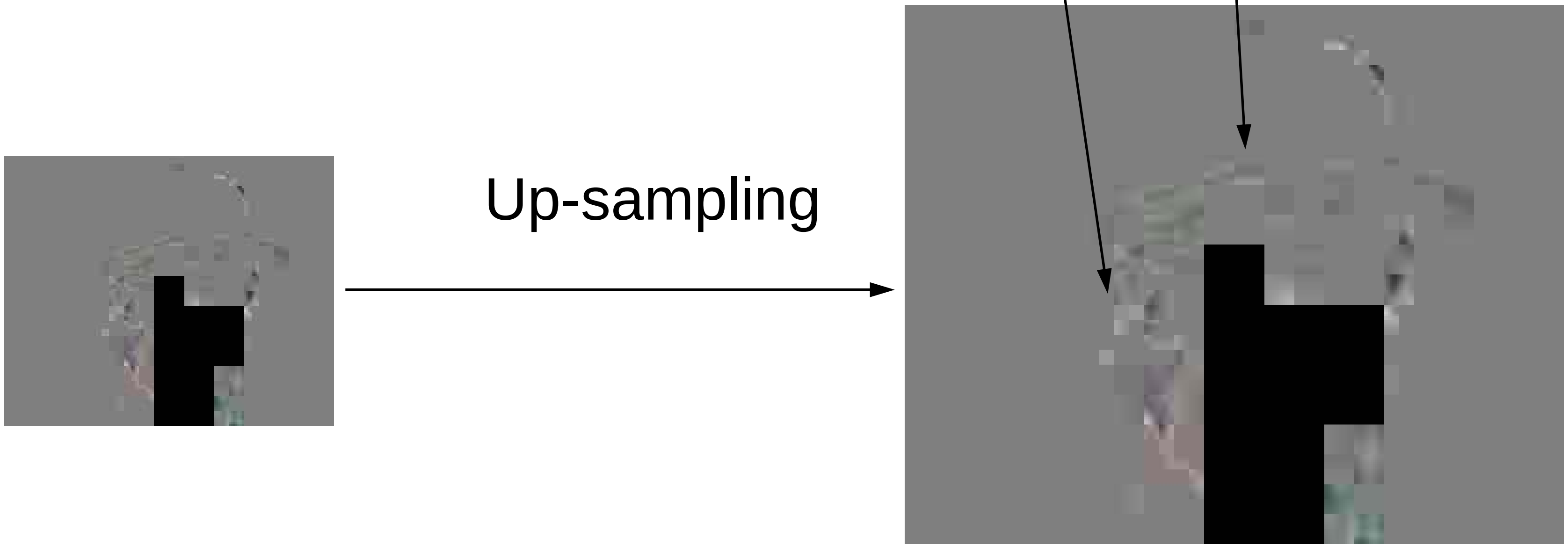




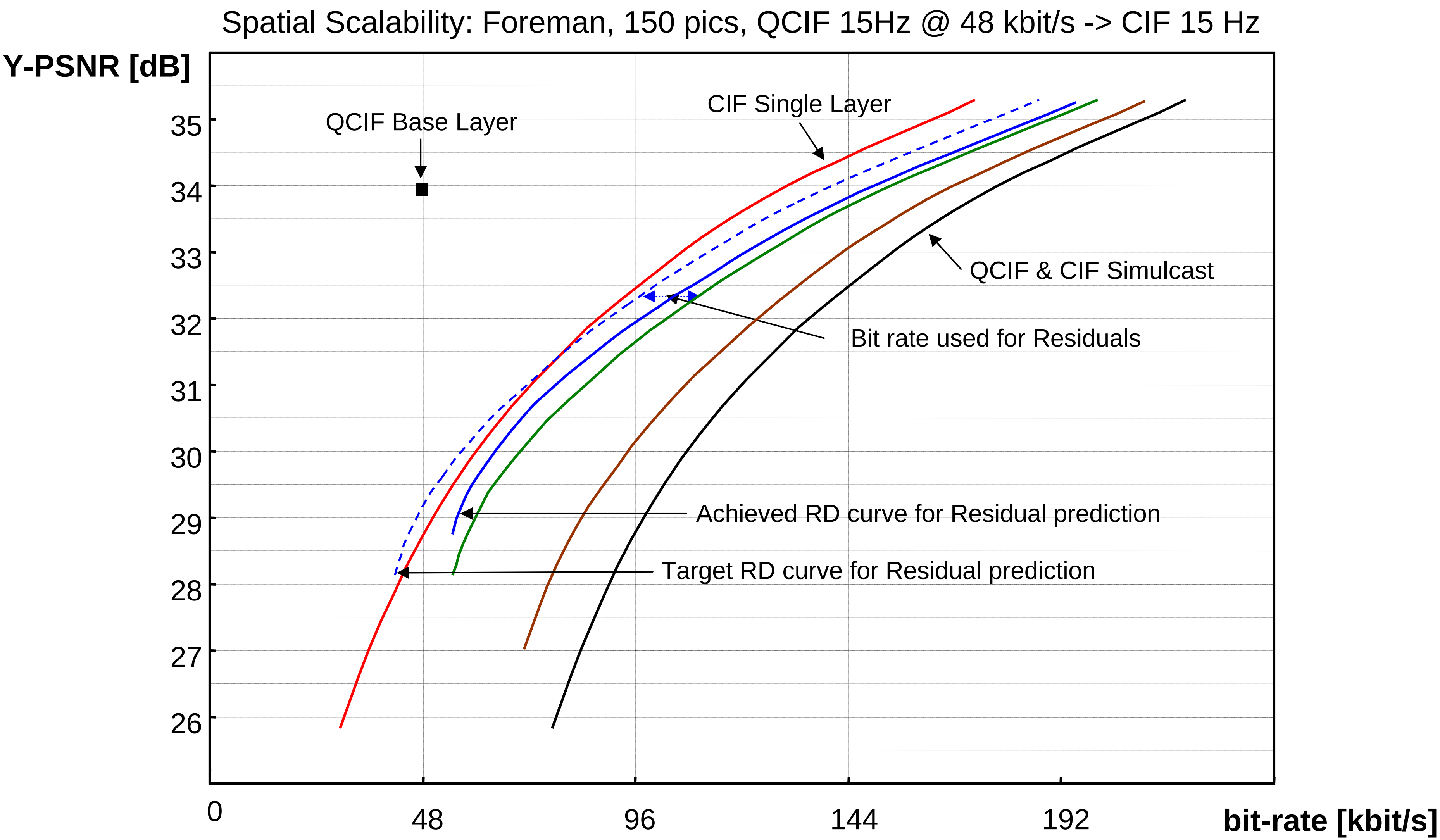

J.-R. Ohm Institut für Nachrichtentechnik RWTH Aachen University

## Block-wise bi-linear up-sampling filter

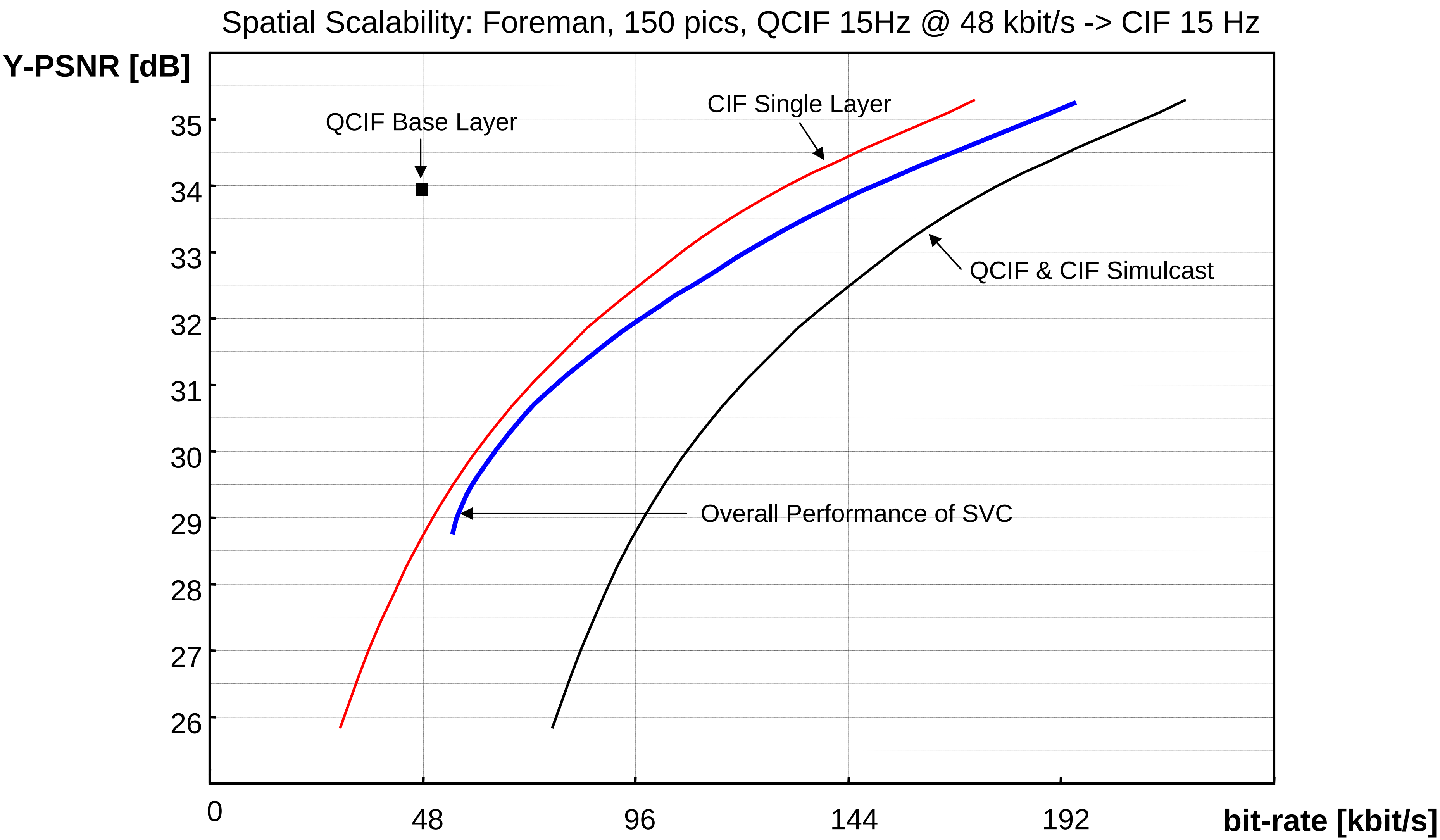
- Block-wise: block boundaries of 4x4 or 8x8 blocks
- bi-linear: small 4x4 block size



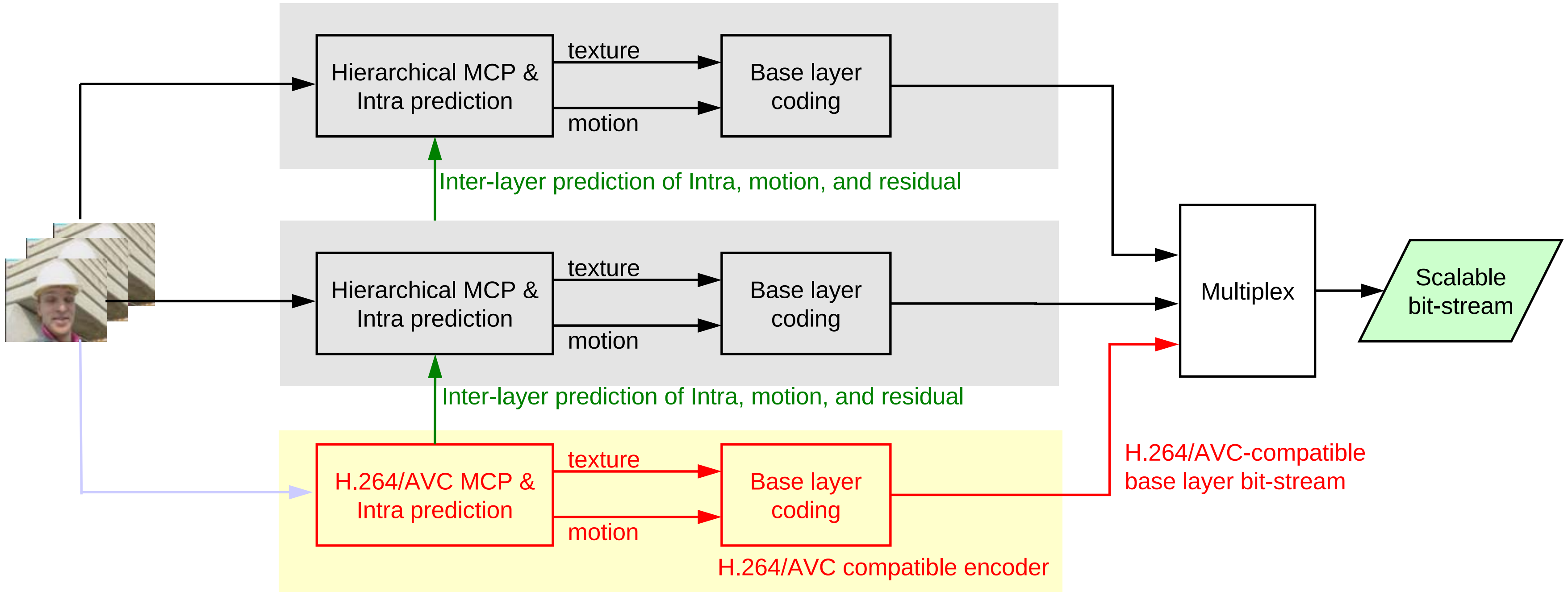
# Spatial Prediction of Data



# Spatial Prediction of Data



# Coarse-grain SNR scalability (CGS)




- Extreme case of (extended) spatial scalability
  - resolution ratio is equal to 1, no cropping
  - Requantization of residual
  - no upsampling (motion, texture) required
  - singe-loop decoding !

## Single-loop Decoding

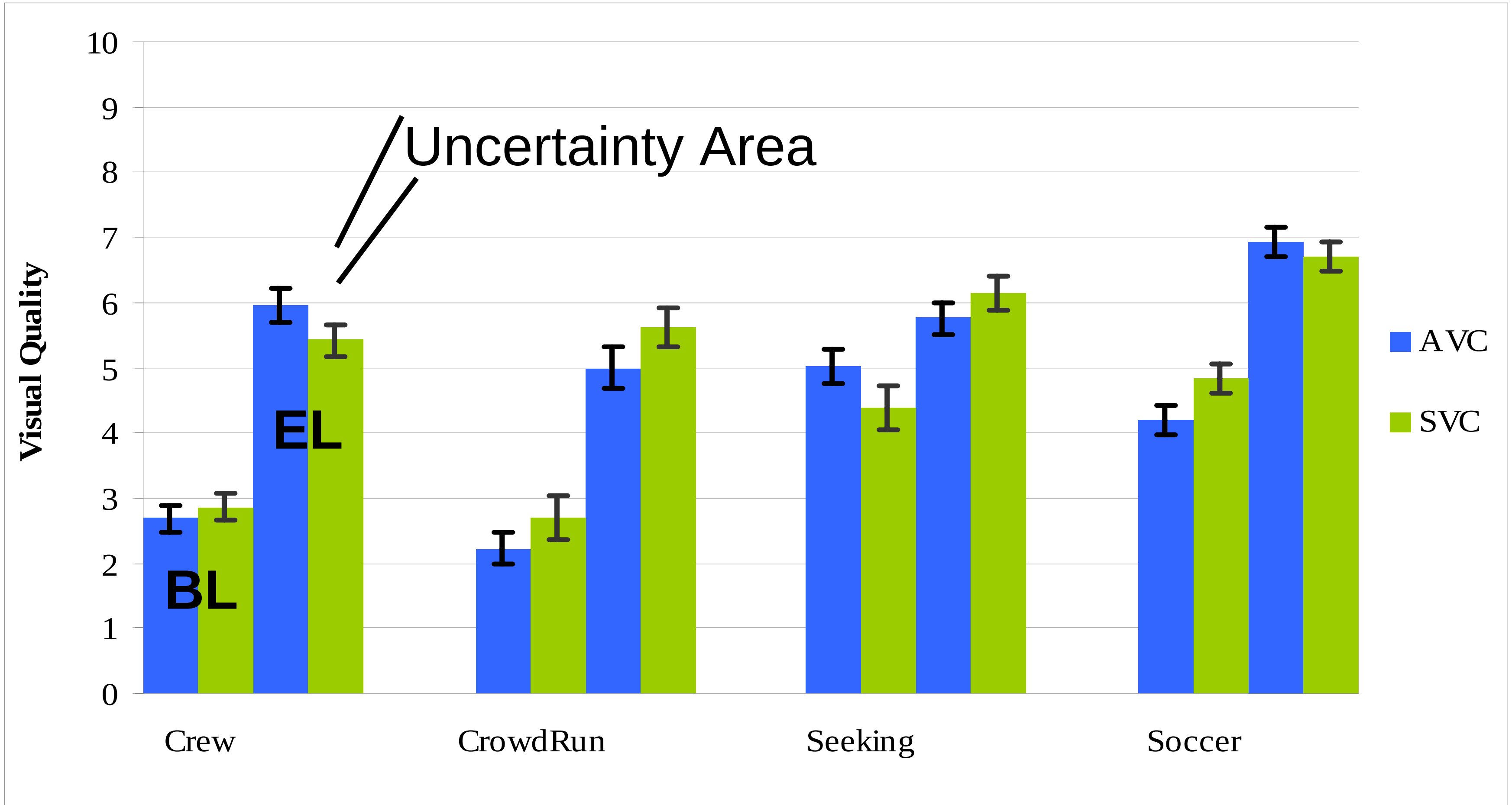
- Unlike MPEG-2, H.263, MPEG-4 scalability, the current SVC uses
  - inter-layer intra prediction is restricted to base layer macroblock that are coded in intra mode
  - single motion compensation loop (including deblocking) is sufficient at decoder side
  - only pictures of highest layer are stored in the decoded picture buffer
- Impact
  - additionally required complexity for supporting spatial and SNR scalability is small
  - Minor impact on compression performance (0 – 0.5 dB)

## Subjective Testing

- Subjective Performance Evaluation of SVC
  - Several test cases for spatial, quality and combined scalability
  - Three application scenarios: TV broadcasting, Conversational, Movie Production
  - Three profiles tested: Scalable Baseline, Scalable High, Scalable High Intra
  - SVC compared to AVC either at 10% additional bit rate or at the same bit rate
  - Visual quality evaluated in subjective tests

## Results of Subjective Testing

- Baseline Broadcasting SNR Scalability QVGA - QVGA



# Multiview and 3D Video Coding

# Multi-View Coding

- Goal: Joint compression of an entire set of video views (captured by multiple cameras

- Possible applications:
  - Free-viewpoint systems (e.g. interactive navigation)
  - 3D TV, multiview displays
- Challenges:
  - How much compression performance can be gained by utilizing inter-view redundancies
  - There may be more than one usage scenario (e.g. different types of displays), which would require different types of optimization

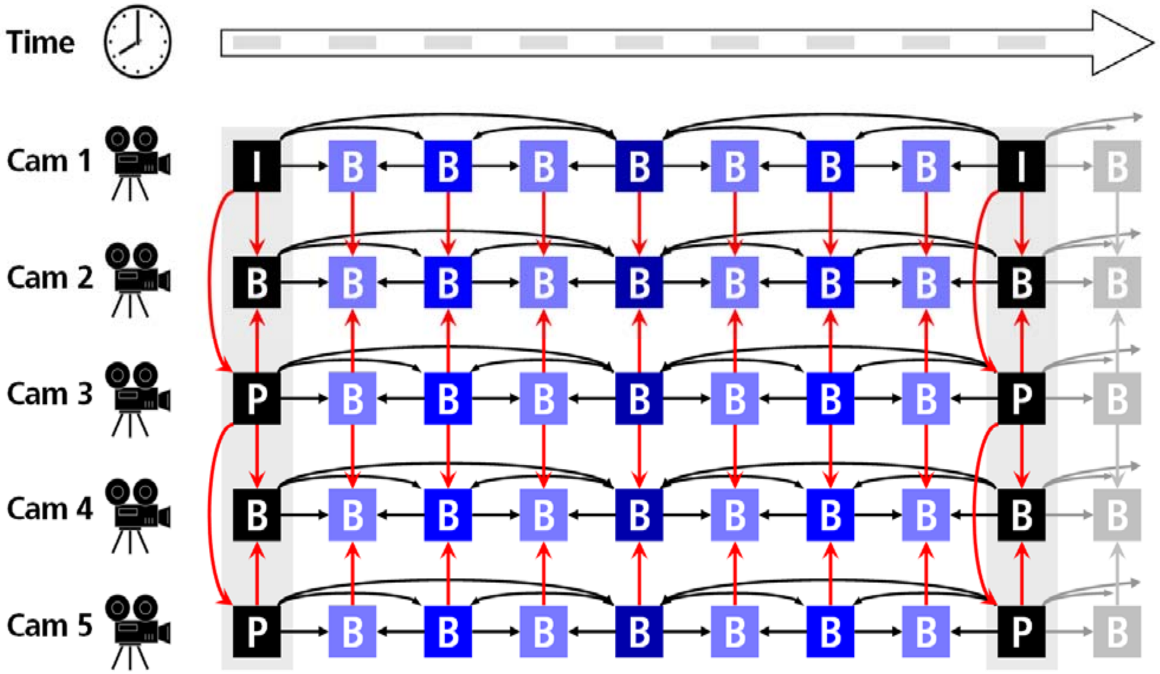
# MVC Applications

- 3D displays have largely improved recently, and are targeting consumer markets
  - High-end displays require many views
  - Autostereoscopic displays for several users and with view correction are coming
  - Low cost stereoscopic e.g. for mobiles
  - *In these cases, all available views need to be decoded*

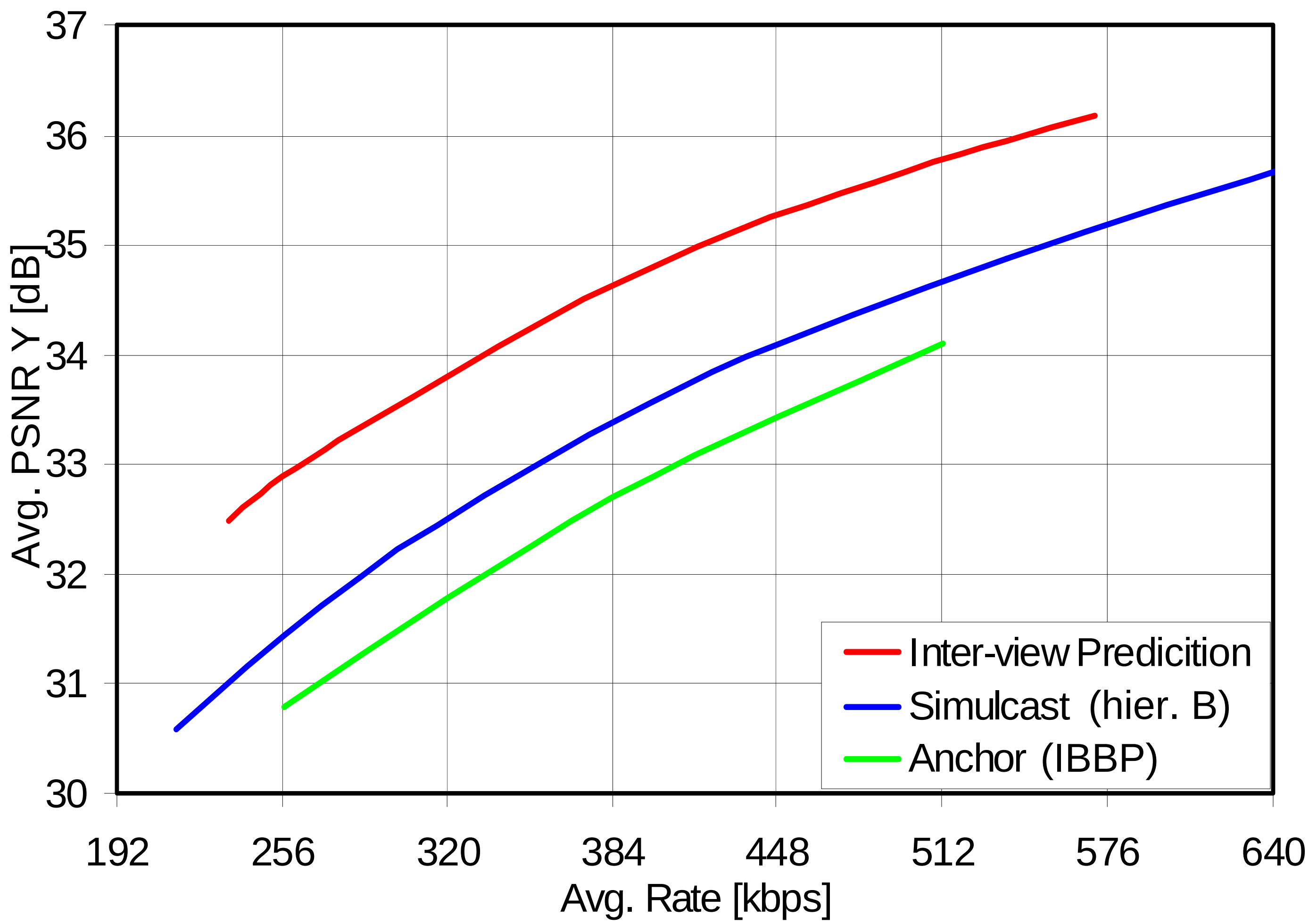
- Adaptation of view direction
  - Support head motion parallax viewing
  - Integral imaging
  - *In these cases, only some views are selectively decoded (all must be available for choice)*

## MVC Reference Model

- Fully compatible to MPEG4-AVC at slice layer and below
- Can be seen as reorganization of input images into a single stream prior to encoding
- Uses hierarchical B-pictures combined in temporal and inter-view dimension



# Objective Results Ballroom



## Overall Observations and Results

- Without any changes at slice layer and below, roughly 20% bit rate reduction can be achieved allowing inter-view prediction
  - Sophisticated temporal+inter-view prediction structures can be implemented based on flexible reference picture indexing of AVC
  - No distinction between motion vectors and disparity vectors in this case
- Efficient methods for buffer and view access management have been defined, including possibilities for low delay and random access

## Overall Observations and Results

- Using additional tools (changing AVC at macroblock level), only 12% additional bitrate saving have been achieved so far, in particular using
  - Illumination compensation

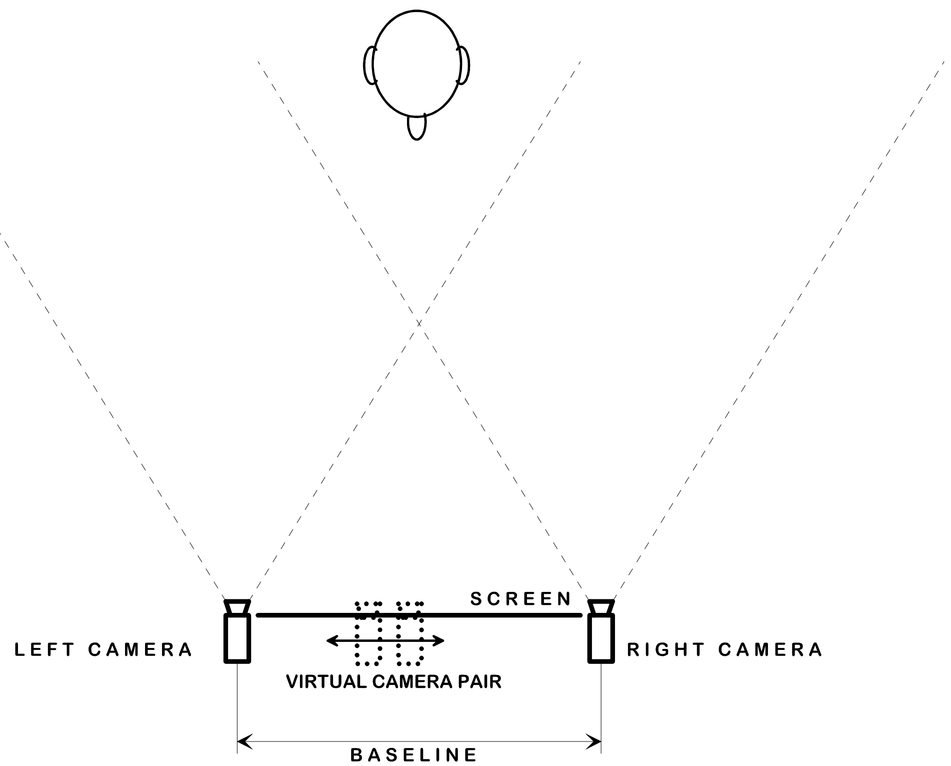

  - Combined motion/disparity vector coding (e.g. disparity-based motion skip)
- It appears that these achievements are not (yet) mature for standardization, first phase MVC standard (ready July 2008) will not include them

# 3D Video

- MVC (in JVT) targets joint compression of a set of multiple views, typically dense camera setup
- 3D Video is a complementary technology to enable generating a continuum of views from a much more sparse view set
  - Most probably requiring depth map representation/ compression and interpolation/rendering method
  - Most probably MVC as video compression for the sparse views
- Focus of the work: Upcoming 3D (M-view) displays in consumer applications
  - Goal: Much less data rate than direct compression of all M views by MVC
  - Good subjective quality, but certainly not measurable by PSNR
- Still unclear whether high-quality view synthesis (interpolation) needs to be normative

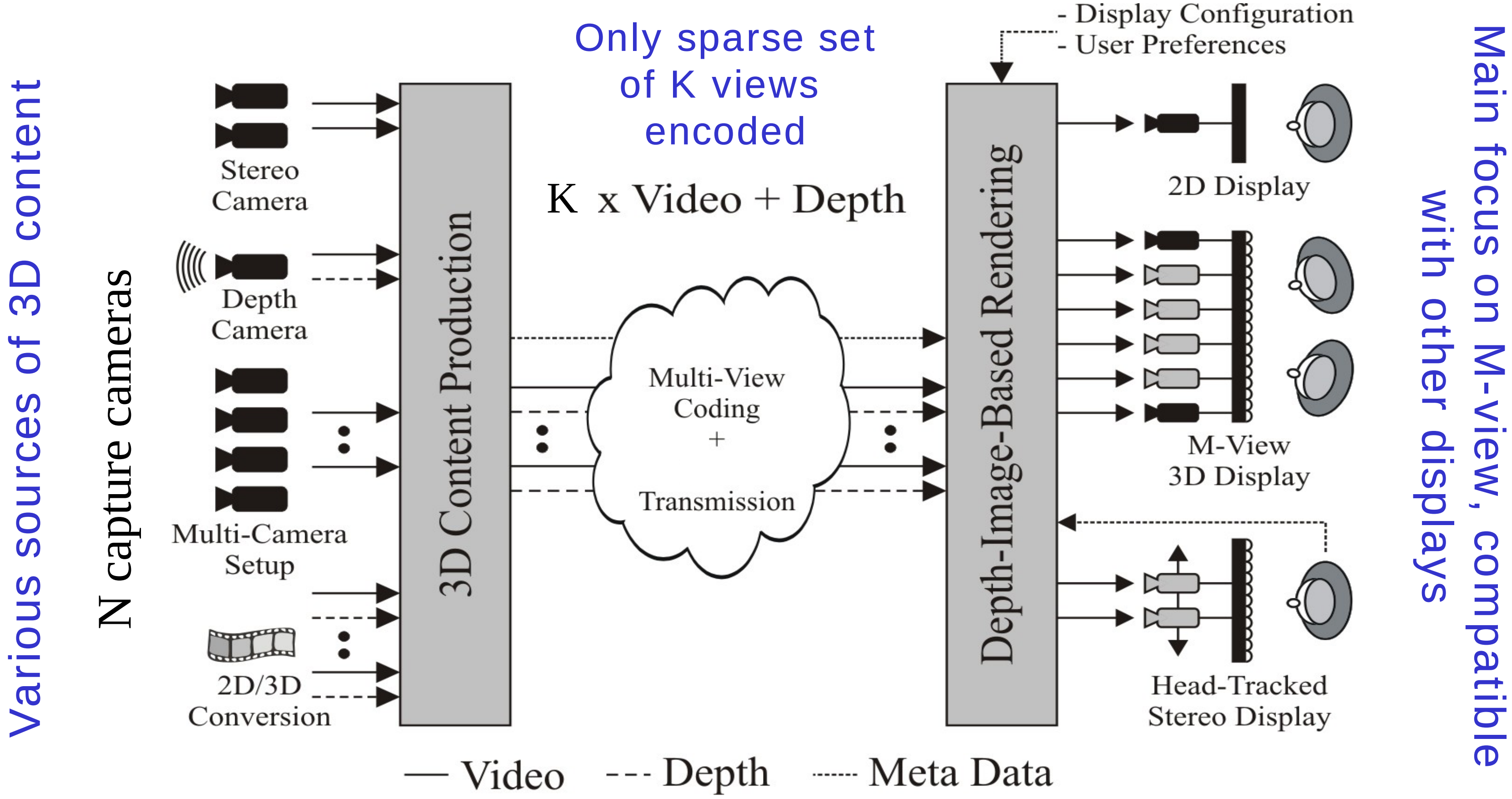
# 3D Video

- Basic principle: Advanced stereoscopic and multiview displays need support to present various views




- Maximum angle between leftmost and rightmost position expected to be around 20 degrees for the upcoming generations of displays

- High-level view of 3DV technology chain

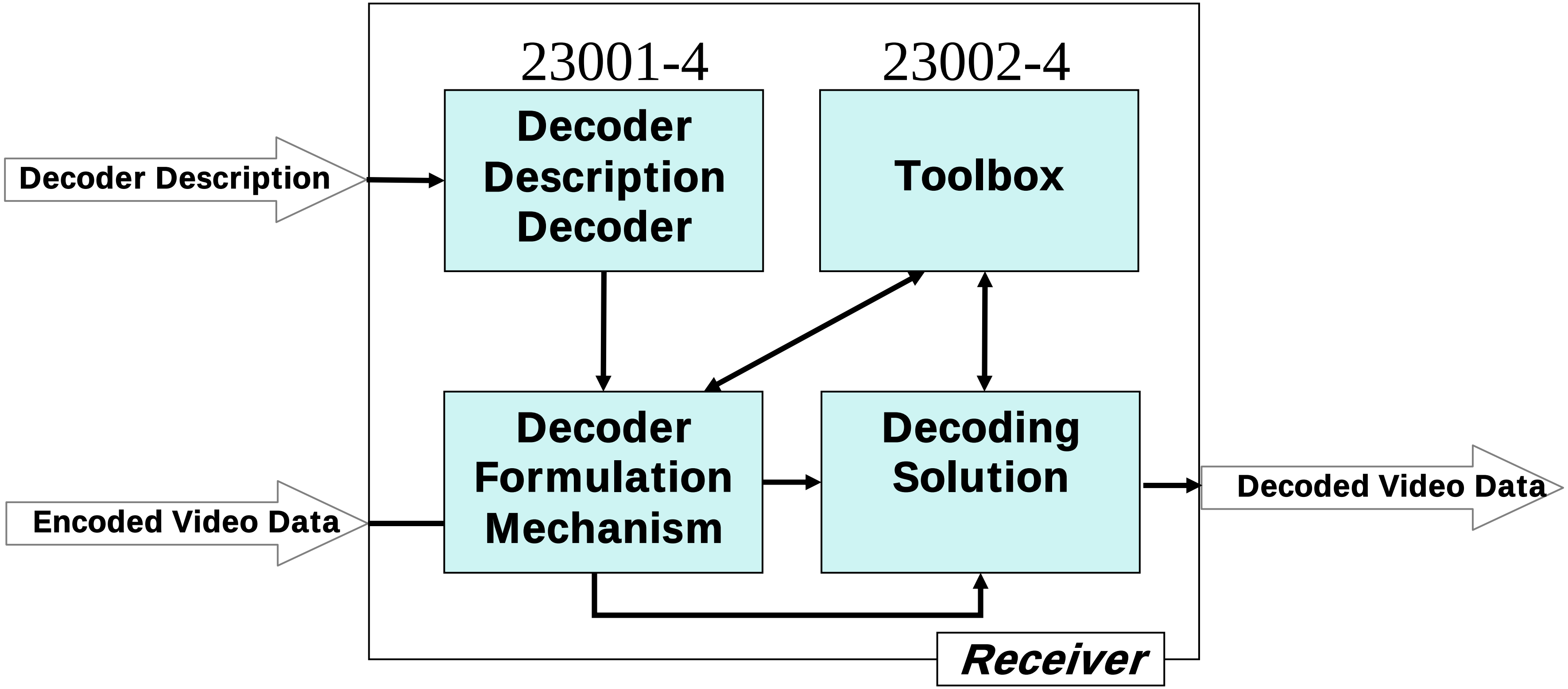


# Reconfigurable Video Coding

## Reconfigurable Video Coding

- Approach: Break down monolithic standards into more elementary building blocks
- Possible applications:
  - Simplified, more flexible standards development
  - Tailoring of existing standards for specific applications (small modifications simple)
  - Devices could "learn" updates of standards
- Current status: Both parts expected for October 2008
  - 21001-4 Codec Configuration Description (useful for non-video and non-MPEG devices)
  - 21002-4 Video Tool Library (including only video tools defined by MPEG)

# Reconfigurable Video Coding



## Status of RVC Standardization

- Codec Configuration Description (23001-4)
  - Based on subset of formal description language CAL (RVC-CAL) – data-flow oriented, timing and synchronization fully supported
  - Bitstream syntax expressed using MPEG-21 BSDL schema (RVC-BSDL), which enables efficient representation by XML-based description
- Straightforward to use this throughout the entire design flow of development of video devices
  - Tools for automatic generation of C code and VHDL are available
  - Beyond the "conventional" C simulation, formal description provides rigid testing of timing behaviour, data flow etc.

# Status of RVC Standardization

- Video Tool Library (23002-4)
  - Currently building blocks from existing standards defined as "functional units" (FUs) are implemented
  - Pseudocode formulation in style of old MPEG standards & CAL based reference software implementation
  - First version will support MPEG-4 Simple Profile, MPEG-2 Main Profile and AVC Baseline Profile
  - Future versions for more MPEG profiles and new tools
  - For more advanced entropy decoders (CABAC, CAVLC) dedicated FUs need to be defined

# Future Video Coding –
# Further Improving Compression?

# Increasing Compression Performance

- Until now, we have seen a continuous tendency with new standards appearing mostly for the purpose of increased compression performance
  - Additional functionality seems "nice to have“, but not necessarily mandatory
  - For consumer applications, mainly the "low-to-medium" rate points where motion compensation is effective (not too much distortion, sufficiently far from lossless)
- Higher compression will certainly be urged by demand for higher resolution
  - Beyond HD & stereo/multiview for the home
  - MPEG-4 AVC was mainly developed for CIF/QCIF at medium/low rates, increasing usage for (low-rate) HD

## Increasing Compression Performance

- Today's video codecs (such as AVC) still working by same principle as developed around 1990
  - Separating texture and motion coding
  - Have however become much more intelligent in encoder decisions (mode switching, rate/distortion based decisions in AVC)
  - Paradigm of normative bitstream & decoder (and non-normative encoder) is traditionally used in video
- Towards lower rates, high percentage of entire bit allocation goes into mode and motion info
  - Dramatic break-down of quality below certain rate points (sequence dependent)
- Saving data rate in motion/mode and texture (intra in particular) seems to be key to improve compression

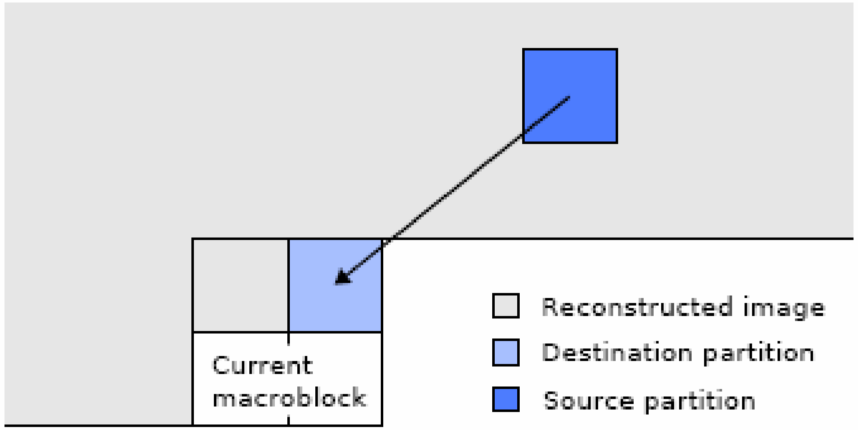
## Increasing Compression Performance

- Possible improvements could be possible by:
  - Texture synthesis (spatially, temporally)
    - Including structured and noisy textures
  - Improved intra prediction & coding
  - Improved motion models, including receiver-side motion update
  - Adaptive transforms or other basis functions
  - Improved motion compensation by adaptive filters
  - Closer integration of motion compensation loop and texture coding?
- Most of these require higher complexity, more "intelligence" also at the decoder side
- Another approach to go could be to decrease complexity while achieving same good results as AVC

- Improved intra prediction by texture modeling

- (Two identical objects in an image need to be redundantly coded.)
- Redundancy of recurring texture features is not fully captured.

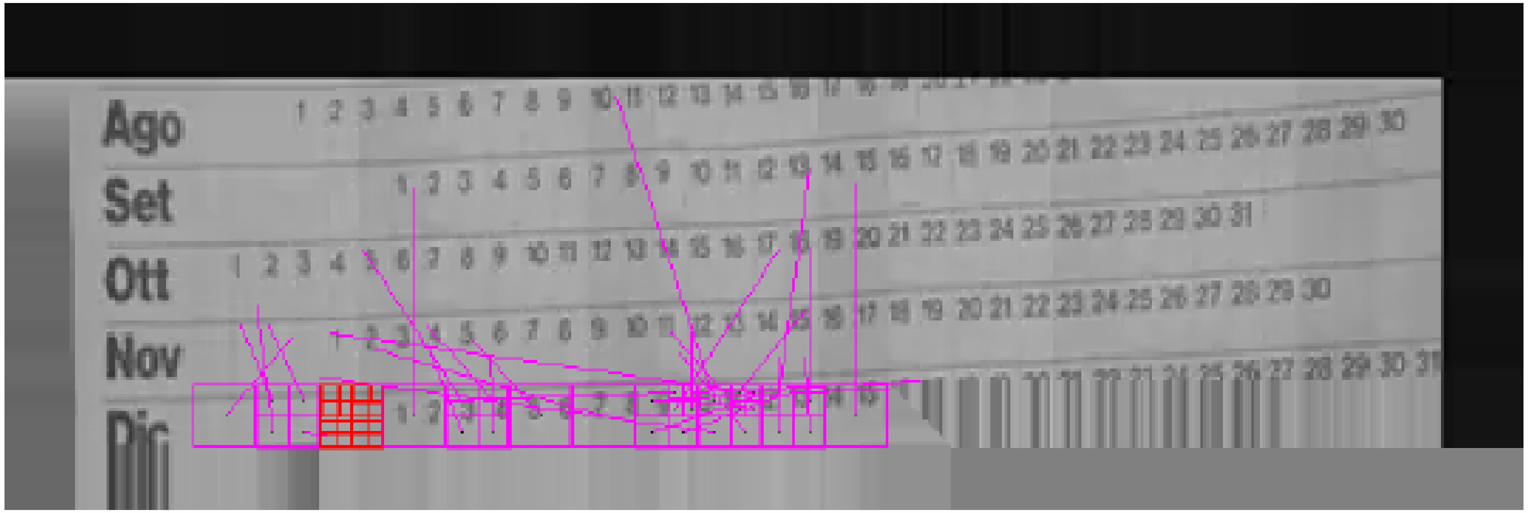
- Displacement intra prediction




- "motion compensation" within one frame [Kondo et al., 2004]
- extrapolation of partly unreconstructed partitions
- quarter-pel refinement
- simple, fast

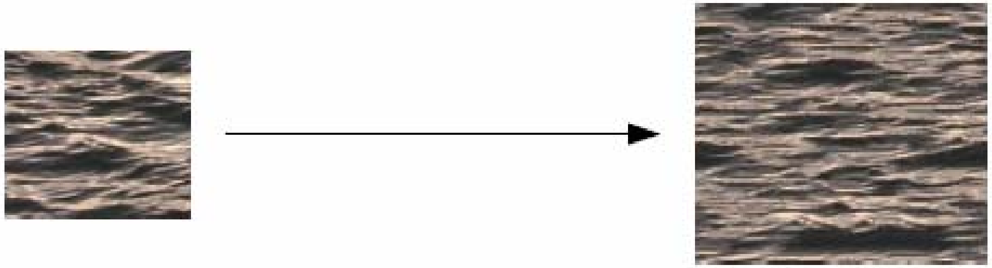
## Examples from RWTH Work

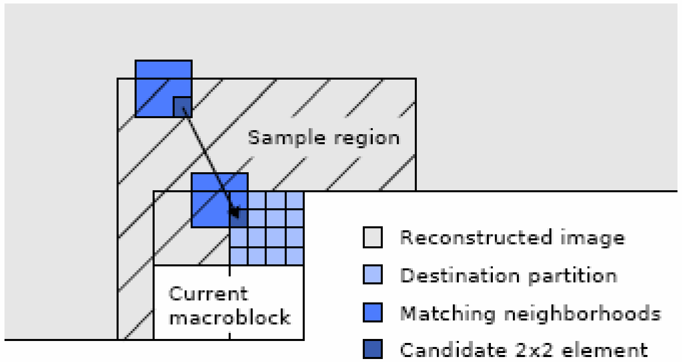
- Displacement intra prediction – example of prediction image for "spincalendar" sequence



# Examples from RWTH Work

- Markovian texture prediction by "template matching"

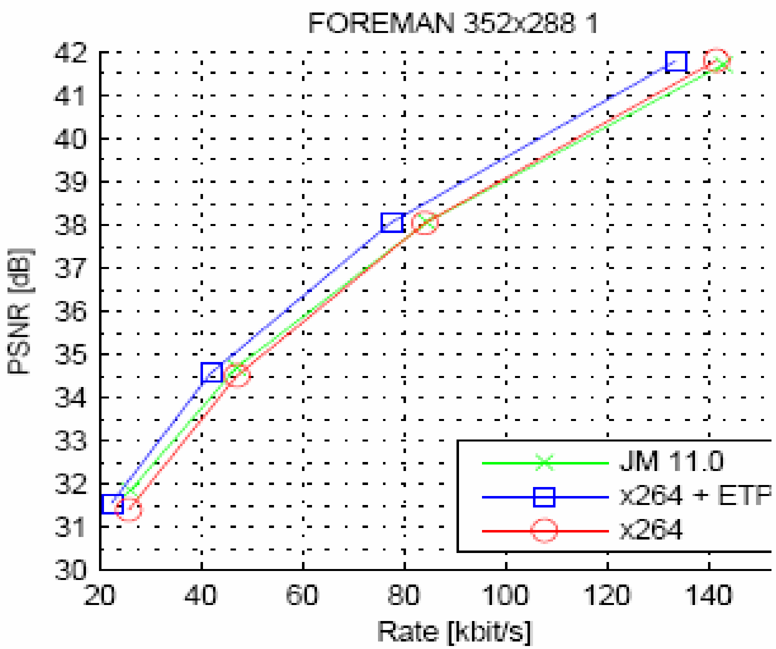


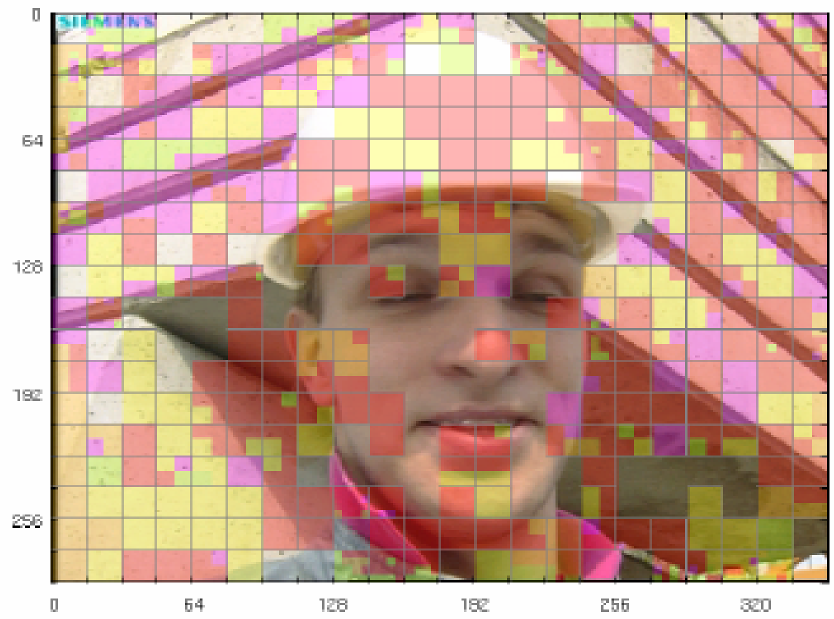



- both encoder and decoder perform search: no parameters
- weighted averaging of candidate elements
- good quality predictors, but complex

## Examples from RWTH Work

- New intra texture prediction - results



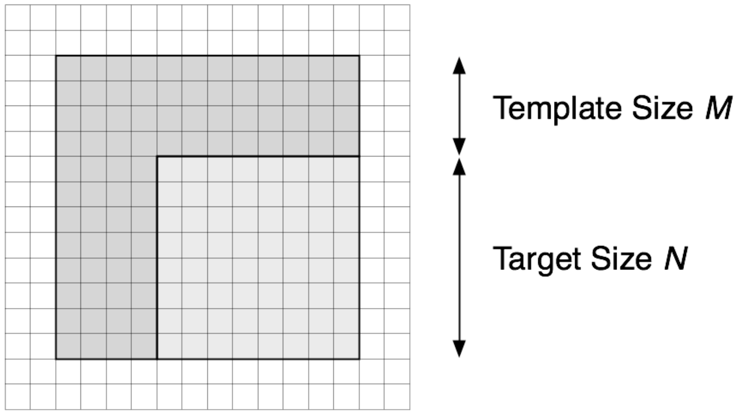
- ca. 5% average rate savings
- new macroblock type: DIP MTP simplified H.264/AVC intra
- uncolored: conventional H.264/AVC intra

## Examples from RWTH Work

- New intra texture prediction - results

| Sequence | Res. | ΔPSNR [dB] | ΔRate [%] |
|---|---|---|---|
| Bus | CIF | 0.347 | –4.217 |
| City | CIF | 0.292 | –4.358 |
| Concrete | CIF | 0.204 | –1.575 |
| Crew | CIF | 0.217 | –3.717 |
| Football | CIF | 0.202 | –2.835 |
| Foreman | CIF | 0.620 | –9.946 |
| Harbour | CIF | 0.141 | –1.707 |
| Mobile | CIF | 0.342 | –3.122 |
| Shuttlestart | CIF | 0.231 | –4.382 |
| Soccer | CIF | 0.246 | –4.488 |
| Spincalendar | CIF | 1.786 | –22.606 |
| Table | CIF | 0.542 | –8.107 |
| Average | | 0.431 | –5.922 |
| City | 4CIF | 0.405 | –6.096 |
| Spincalendar | 4CIF | 2.814 | –33.931 |

- Extension of template matching to inter prediction
  - Basic principle: Motion estimation at decoder, comparing neighboring template of current block against reference frame



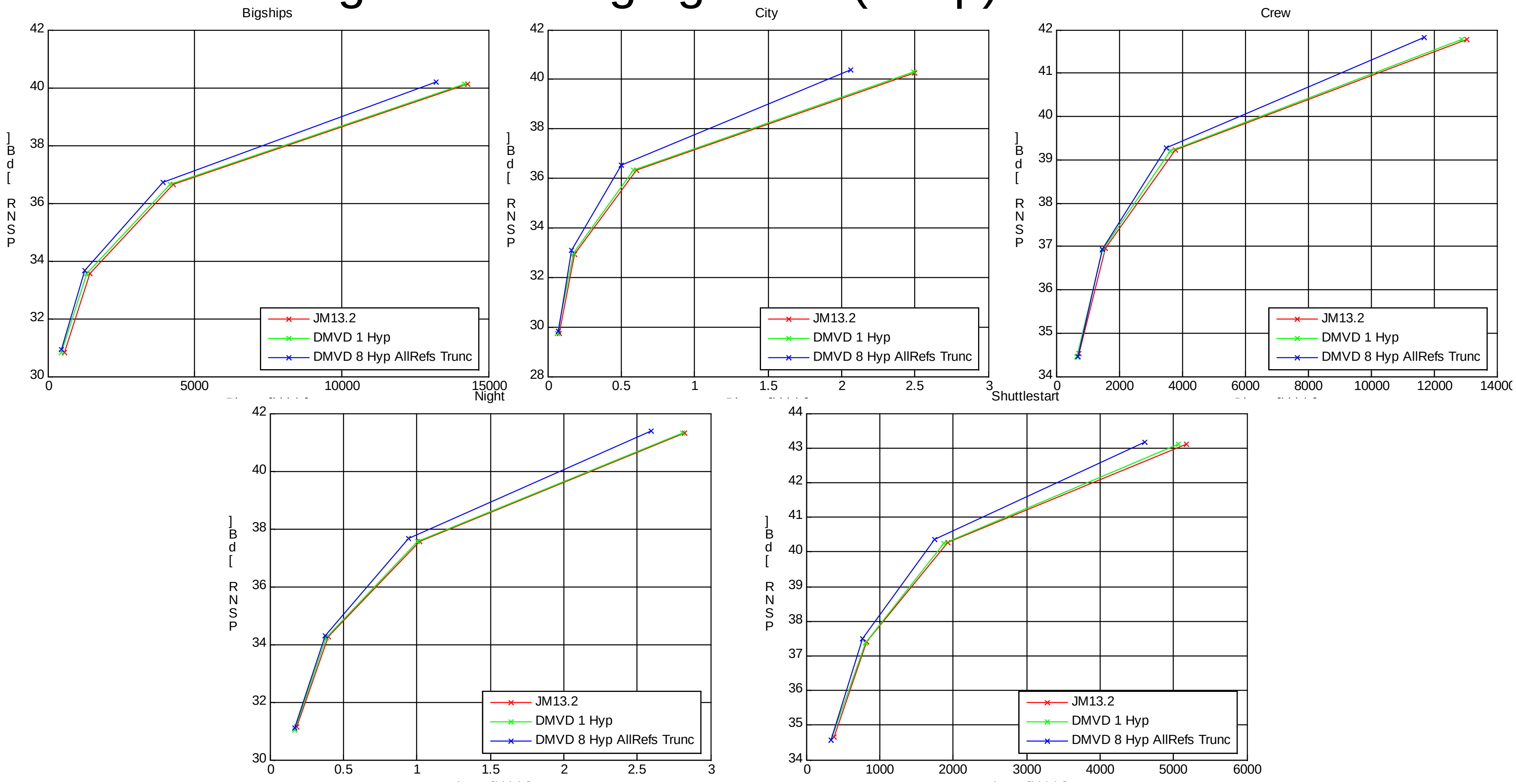

  - Template can be derived from the decoded signal or from the prediction signal
  - Small search range around predictor (±1..4 pixels)

## Examples from RWTH Work

- Advantages of decoder-side motion vector derivation using template prediction
  - If selected (signaled by an additional mode definition) it is not necessary to transmit motion vector
  - If combined with multi-frame prediction, it is not even necessary to signal the reference frame index
  - Weighted averaging of several (best, second best etc.) matches becomes appropriate (unlike conventional coding, where a motion vectors would need to be encoded for each

# Examples from RWTH Work

- Combination of template inter prediction, multi-frame and weighted averaging – HD (720p) results



- Template inter prediction – overall results in bit rate reduction

| Sequence | Resolution | DMVD 1 Hyp | DMVD 2 Hyp | DMVD 4 Hyp | DMVD 8 Hyp |
|---|---|---|---|---|---|
| foreman | CIF 30Hz | -5.2209 | -6.5985 | -7.7978 | -6.7912 |
| mobile | CIF 30Hz | -1.5080 | -3.2016 | -5.4924 | -5.4051 |
| paris | CIF 15Hz | -1.0622 | -1.6025 | -2.0272 | -1.2819 |
| tempete | CIF 30Hz | -1.6167 | -2.7669 | -4.4590 | -4.8984 |
| bigships | 720p 60Hz | -5.9903 | -8.1042 | -14.1659 | -13.4904 |
| city | 720p 60Hz | -5.4663 | -9.0854 | -17.1830 | -19.1836 |
| crew | 720p 60Hz | -4.1780 | -5.4089 | -7.3916 | -7.6865 |
| night | 720p 60Hz | -1.8803 | -4.6359 | -7.3306 | -7.7954 |
| shuttlestart | 720p 60Hz | -2.8586 | -5.0851 | -9.4766 | -10.9685 |
| **average** | **CIF** | **-2.3520** | **-3.5424** | **-4.9441** | **-4.5941** |
| | **720p** | **-4.0747** | **-6.4639** | **-11.1096** | **-11.8249** |
| | **Overall** | **-3.3090** | **-5.1655** | **-8.3693** | **-8.6112** |

# Summary

- Continuing efforts made to resolve previously open problems in video coding – example SVC
  - Not more than 10% rate increase for scalability functionality has been achieved, otherwise it would not be acceptable

- Increasing compression stays most important factor of development – looking for increasing resolution and multiview
  - Desire for higher pixel rates grows faster than affordable network bandwidth (both for wireless and wired)
  - Bit rate must not scale up with resolution

- Current paradigms in video standardization may need revisits
  - Just some examples: non-SNR quality criteria, role of RVC, roles of encoder/bitstream/decoder/postprocessing