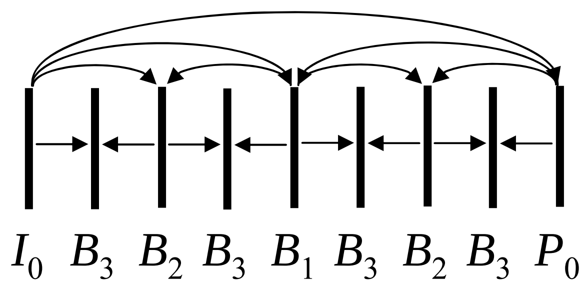
A Convex Optimization Approach for Determining Rate-Distortion Optimized H.264/AVC and SVC Transform Coefficients

***Martin Winken**, Heiko Schwarz, Thomas Wiegand*

Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Introduction

- Usage of motion-compensated prediction causes manifold inter-picture dependencies, e.g.:



- Inter-picture dependencies:
  - Pictures B_n can only depend on pictures B_m with $m < n$
  - Typically considered using a cascading of quantization parameter (QP) values.
- **Idea:** Consider dependencies also in selection of transform coefficients

General approach

- Write motion-compensated prediction signal as matrix product:

motion-compensated prediction signal

motion matrix

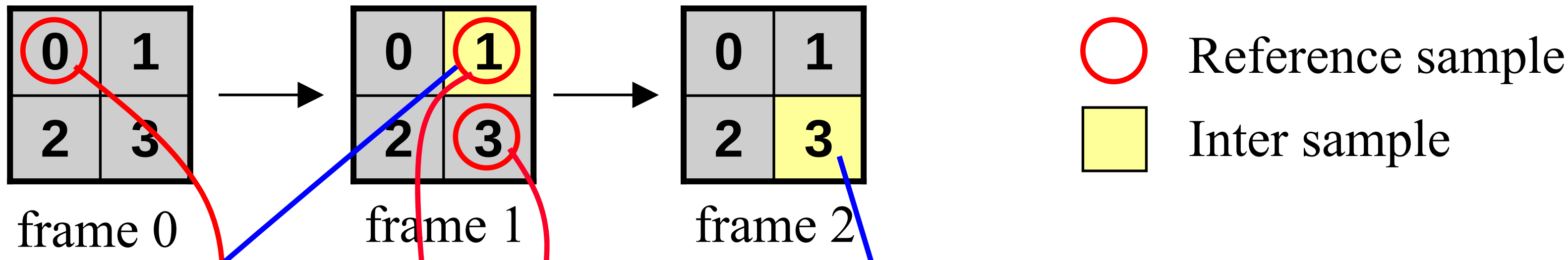
reconstructed signal

$$\hat{\mathbf{s}}_{\mathrm{MC}} = \mathbf{M}\,\mathbf{s}$$

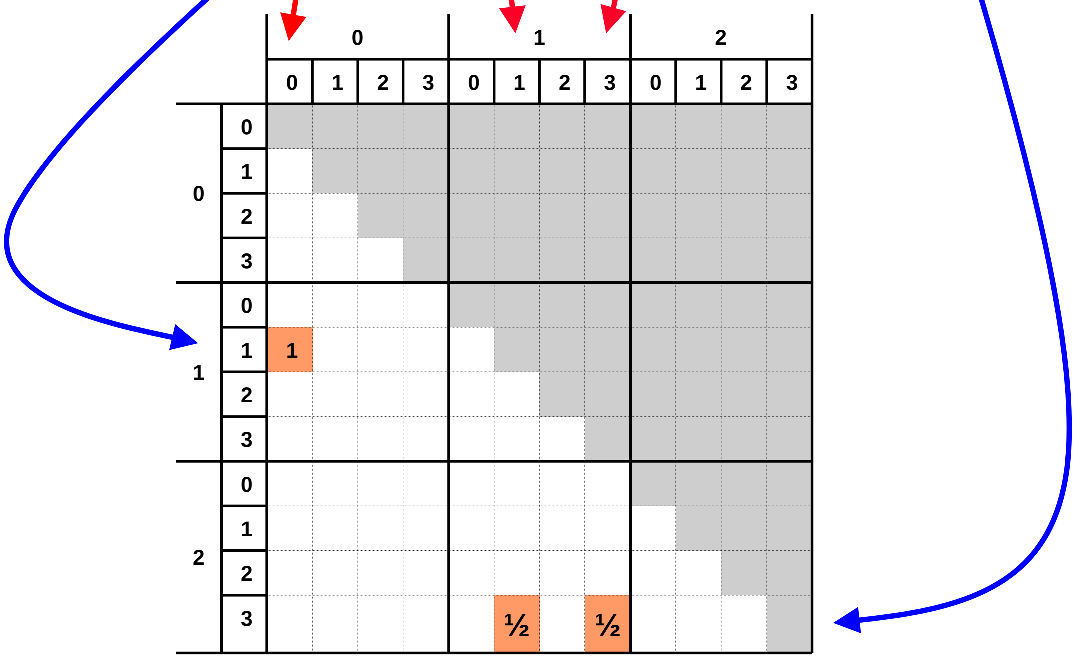
- **Note:** The vectors contain **all** the samples of the sequence!

Example: How to construct matrix M?

- Simple example using 3 pictures of size 2×2:



- Construction of motion matrix **M**:



$$\hat{\mathbf{s}}_{MC} = \mathbf{M}\,\mathbf{s}$$

non-zero entry

always zero entry

Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Write decoding process as matrix operation

- **General:**

reconstructed signal, prediction signal, residual signal

$$\mathbf{s} = \hat{\mathbf{s}} + \mathbf{r}$$

- **Rewrite prediction signal:**

static predictor

$$\mathbf{s} = (\mathbf{Ms} + \mathbf{p}) + \mathbf{r}$$

- **Rewrite residual signal:**

transform coefficients

$$\mathbf{s} = (\mathbf{Ms} + \mathbf{p}) + \mathbf{Tc}$$

scaling / inv. transform matrix

Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Inaccuracies of the linear signal model

- **Linear signal model for the decoder**

motion data

transform coefficients

$$\mathbf{s} = \mathbf{M}\mathbf{s} + \mathbf{T}\mathbf{c} + \mathbf{p}$$

reconstructed sample values

static predictor

scaling & inverse transform

- **Inaccuracies when using H.264/AVC due to**
  - rounding / clipping
  - deblocking filter

Rate-constrained optimization

- **General Lagrangian optimization approach:**

$$\min \left\{ \mathrm{D}(\mathbf{c}) + \lambda \cdot \mathrm{R}(\mathbf{c}) \right\}$$

- **Use mean-squared error distortion measure:**

$$\mathrm{D}(\mathbf{c}) = \left\| \mathbf{s} - \mathbf{v} \right\|_2^2 = (\mathbf{s} - \mathbf{v})^T (\mathbf{s} - \mathbf{v})$$

original signal

- **Use piece-wise linear rate model:**

$$\mathrm{R}(\mathbf{c}) = \left\| \mathbf{c} \right\|_1 = \sum_i \mathrm{abs}(\mathbf{c}_i)$$

Quadratic Program Formulation

- Select transform coefficient values by solving for $\mathbf{c}_{\mathbf{opt}}$:

original video signal

minimize $$(\mathbf{s}-\mathbf{v})^T(\mathbf{s}-\mathbf{v})+\lambda\sum_i \mathrm{abs}(\mathbf{c}_i)$$

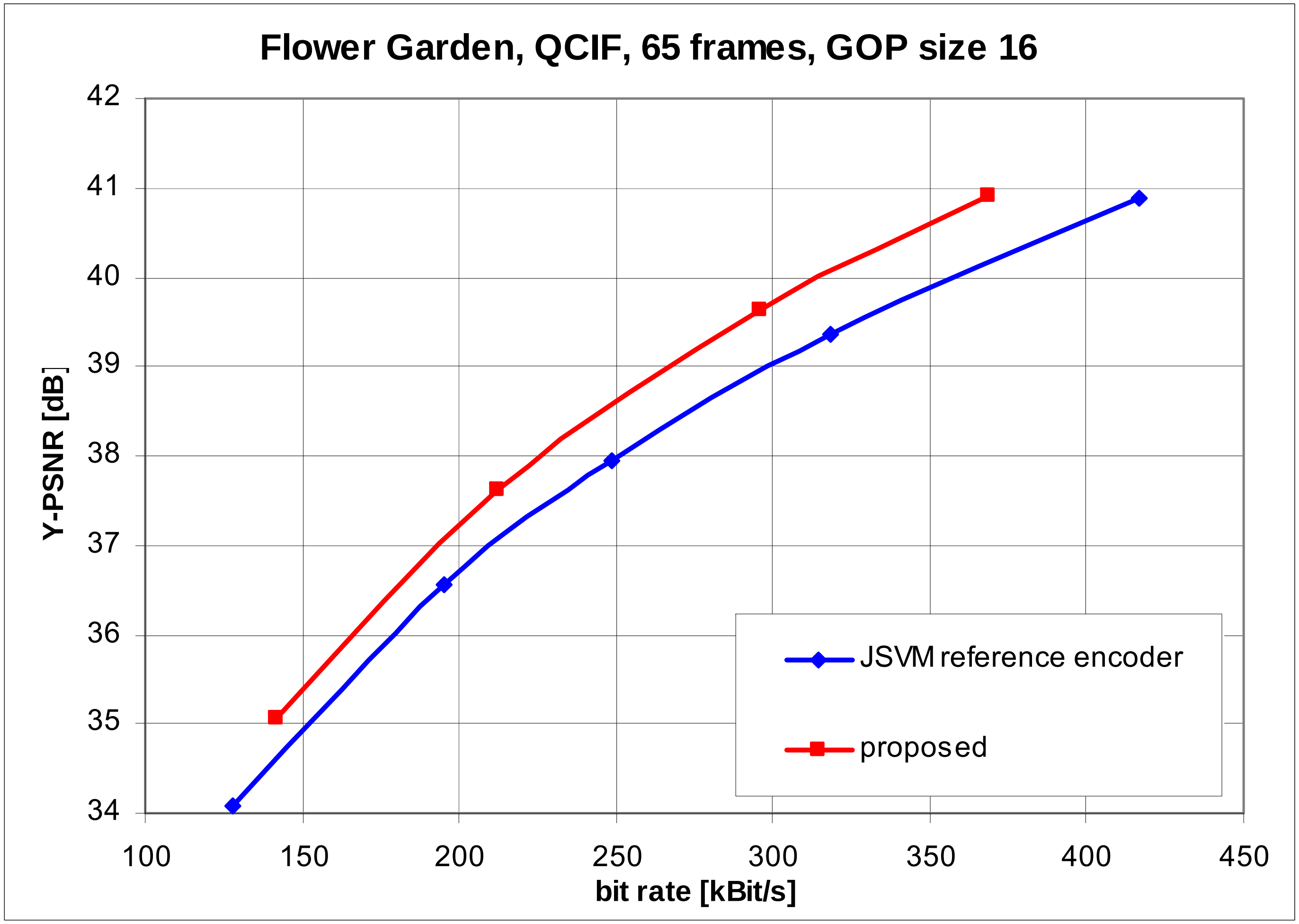
subject to $$\mathbf{s}=\mathbf{M}\mathbf{s}+\mathbf{T}\mathbf{c}+\mathbf{p}$$

- Problems of this kind are called "quadratic program"
- Can be solved by standard numerical optimization algorithms!
- Matrices **M** and **T** are very large, but highly sparse

Side note: There's no "real" quantization!

- In our approach, there is no "classical" quantization
- Instead, transform coefficient values are obtained by solving a numerical optimization problem (quadratic program)

Experimental Results



Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

M. Winken, H. Schwarz, T. Wiegand – A Convex Optimization Approach for Determining Rate-Distortion Optimized H.264/AVC and SVC Transform Coefficients

Extension for SVC Spatial Scalability

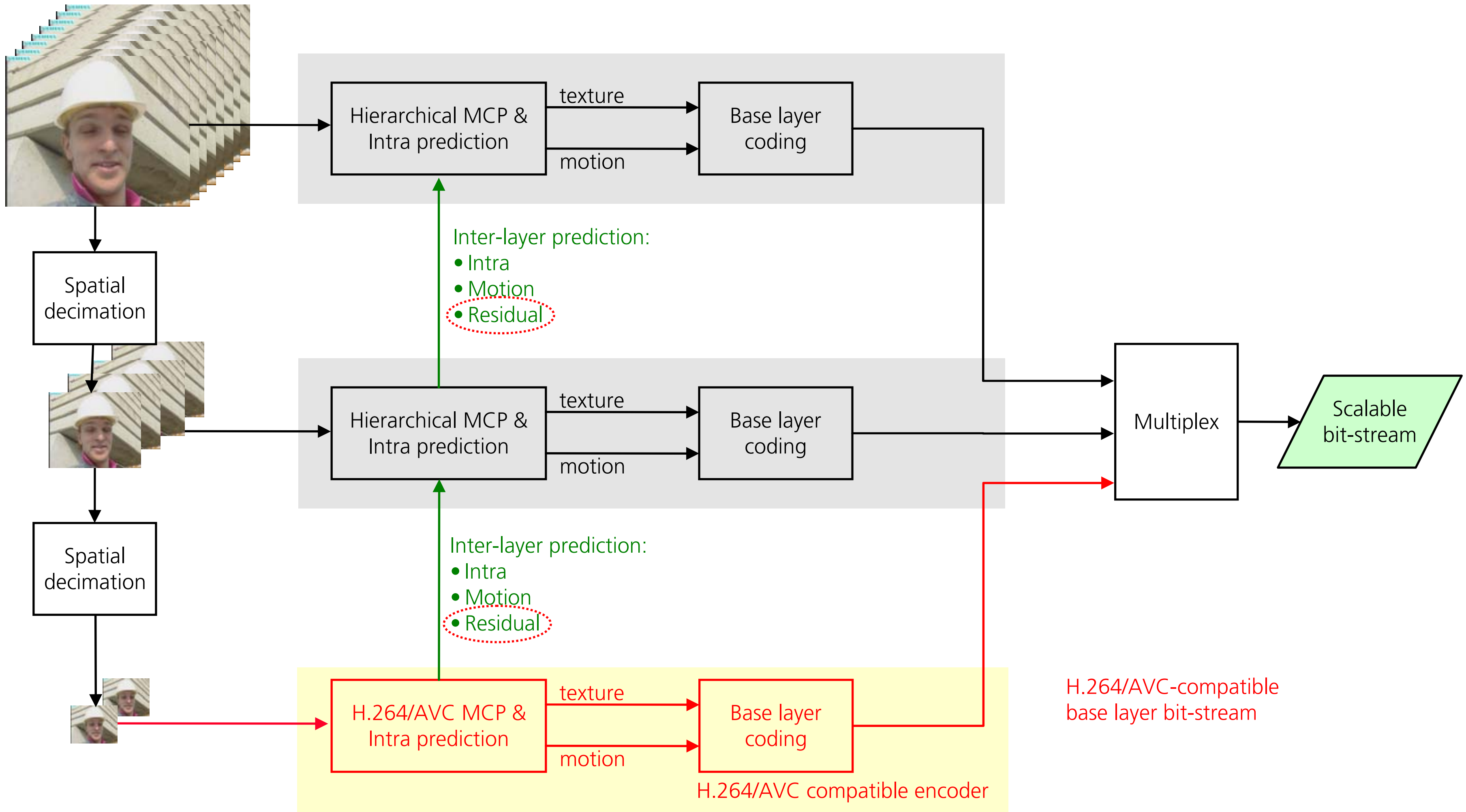
Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

SVC Extension of H.264/AVC

- **The recently finalized Scalable Video Coding (SVC) extension of H.264/AVC provides:**
  - Temporal scalability
  - Quality scalability
  - Spatial scalability

- **Idea: Include the additional dependencies introduced by spatial scalability in the optimization**

SVC Spatial Scalability



Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Linear signal model for spatial scalability

- **Base layer model is unchanged**

$$\mathbf{s}_0 = \mathbf{M}_0\mathbf{s}_0 + \mathbf{T}_0\mathbf{c}_0 + \mathbf{p}_0$$

- **Enhancement layer has two additional terms**

$$\mathbf{s}_1 = \mathbf{M}_1\mathbf{s}_1 + \mathbf{T}_1\mathbf{c}_1 + \mathbf{p}_1 + \mathbf{B}\mathbf{s}_0 + \mathbf{R}\mathbf{T}_0\mathbf{c}_0$$

Inter-layer
Intra prediction

Inter-layer
residual prediction

Optimization Problem for Spatial Scalability

- **Minimized a weighted sum of base and enhancement layer R-D costs:**

$$\begin{bmatrix} c_{0,opt} \\ c_{1,opt} \end{bmatrix} = \underset{[c_0 \quad c_1]^T}{\arg\min} \begin{Bmatrix} (1-w)\cdot\alpha\cdot J_0(c_0) + \\ w\cdot J_1(c_0, c_1) \end{Bmatrix}$$

α : normalization factor

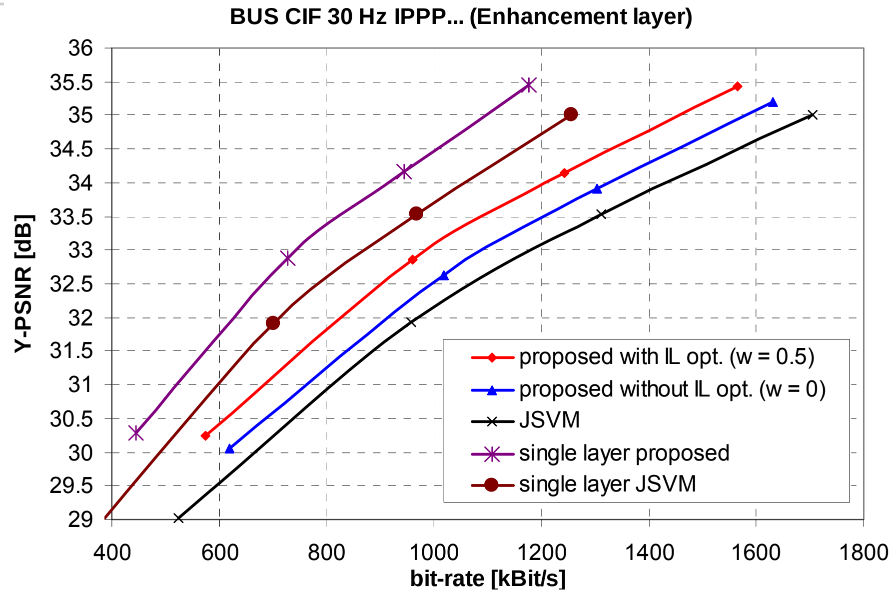
w : weighting factor

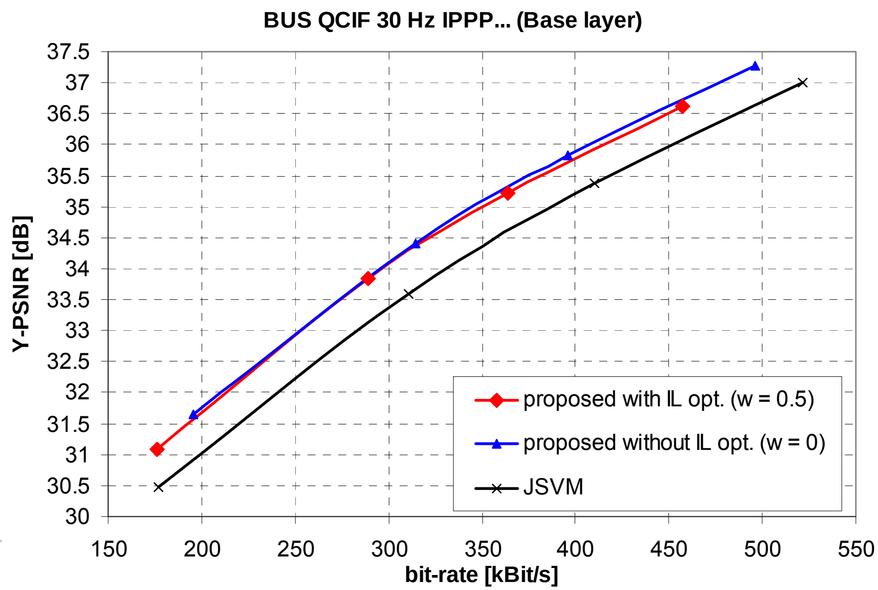
with base and enhancement layer R-D costs

$$J_0(c_0) = D_0(c_0) + \lambda_0 \cdot R_0(c_0)$$

$$J_1(c_0, c_1) = D_1(c_0, c_1) + \lambda_1 \cdot (R_0(c_0) + R_1(c_1))$$

Experimental Results





Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Conclusion

- New approach for selecting transform coefficient values has been presented
- Basic idea: Consider inter-picture dependencies due to motion-compensated prediction and inter-layer prediction
- Gains are up to 0.8 dB PSNR