International Conference on Image Processing 2005

Coding Efficiency of Video Sensor Networks

Markus Flierl

Stanford University
Max Planck Center for Visual Computing and Communication

Motivation

- 3-dimensional scene that evolves in time
- Observed by multiple video cameras located at different positions
- Each camera signal is coded locally
- The cameras are connected directly to the network
- One remote decoder is able to reconstruct arbitrary views

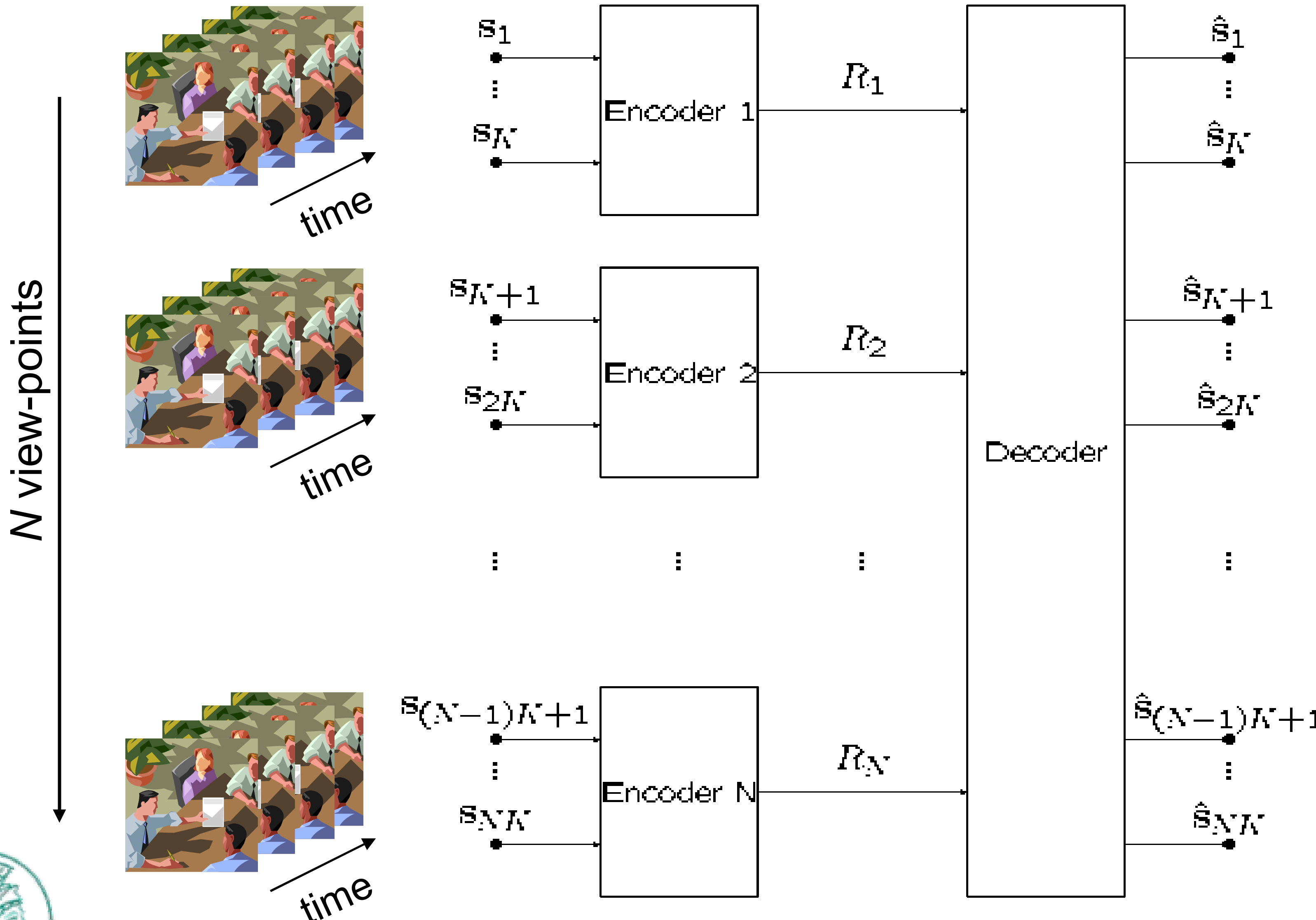
Outline

- Communication Scenario
- Coding of One Video Signal with Side Information
- Efficiency Study
  - *Model for Transform-Coded Video Signals*
  - *Conditional Karhunen-Loeve Transform*
  - *Relative Conditional Eigendensities*
- Collaborative Coding of Multiple Video Signals
- Performance with Gaussian Assumption

Communication Scenario

- Multiple video cameras that are connected to a network and encode highly correlated signals
- Joint decoder recovers arbitrary views
- Scenarios:
  - *All encoders communicate with each other and compress the signals jointly*
  - *Encoders do not communicate with each other but rely solely on joint decoding*
  - *Combination of above scenarios*
- At high rates, all scenarios may achieve the same rate distortion bound

Communication Scenario



We are Interested in …

the efficiency of

- *collaborative motion-compensated transform encoding/decoding*

when compared to

- *non-collaborative motion-compensated transform coding*

depending on

- *the number of cameras*
- *the size of the MCT GOP at each sensor*
- *the correlation among the view-point signals*

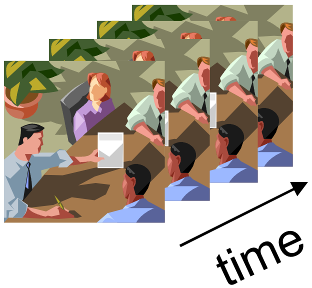
Coding One Video Signal with Side Information

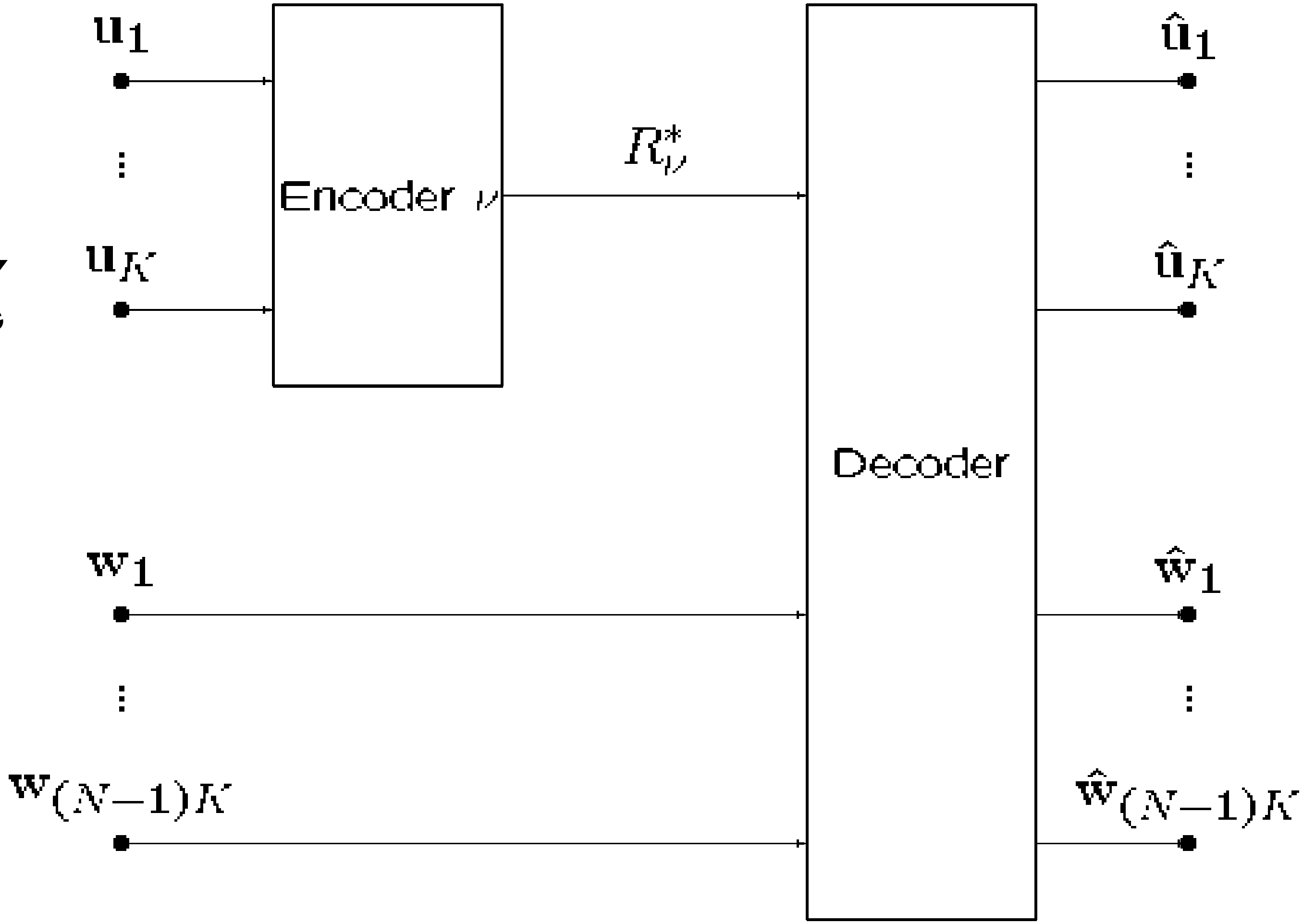
- Choose set **s** of *K* input pictures to be encoded
- Set **w** of *(N-1)K* side information pictures
- At high rates:
  - ***Reconstructed side information at the decoder approaches the original side information, i.e.,***

    $$\hat{\mathbf{w}} \longrightarrow \mathbf{w}$$

  - ***Wyner-Ziv coding scheme***
  - ***Rate distortion function of chosen encoder is bounded by the conditional rate distortion function and the bound is achieved for Gaussian Signals [Wyner & Ziv, 1976]***

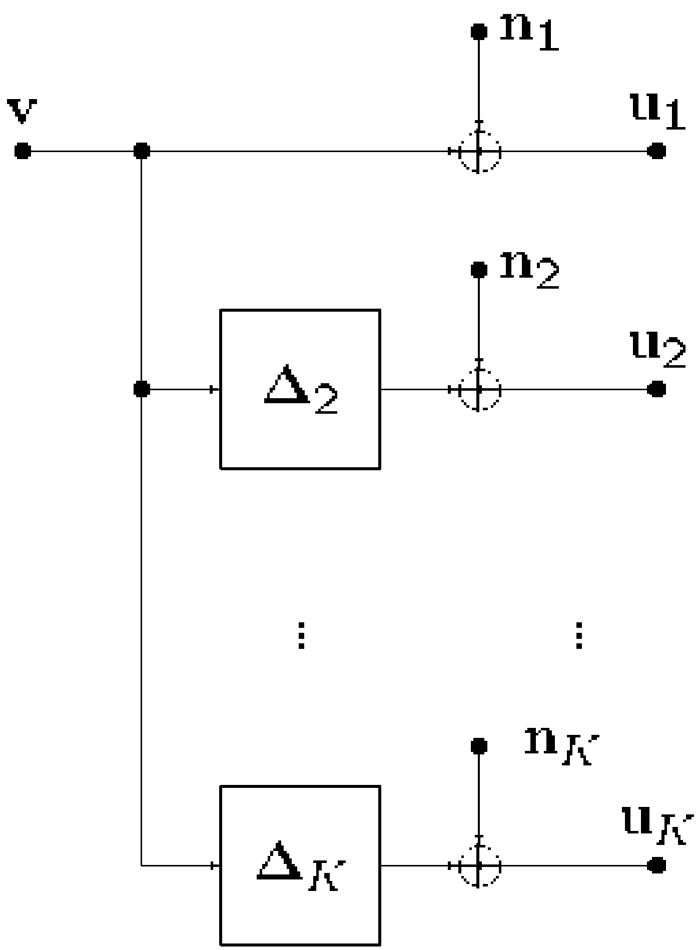
Coding One Video Signal with Side Information





(N-1)K
pictures

Signal Model for Subband Coding of Video



Model for coding with motion-compensated lifted wavelets [Flierl & Girod, 2003]

$\mathbf{v}$ model picture

Δ_k k-th displacement error

$\mathbf{n}_k$ k-th model error of motion compensation

$\mathbf{u}_k$ k-th motion-compensated signal

Signal Model for Subband Coding of Video

- Basic idea:
  - *Reversible true motion trajectories*
  - *Reversible estimated motion trajectories*
  - *Identical accuracy of motion compensation*
- Power spectral densities of *K* pictures:

$$\frac{\Phi_{\mathbf{uu}}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} = \begin{pmatrix} 1+\alpha & P & \cdots & P \\ P & 1+\alpha & \cdots & P \\ \vdots & \vdots & \ddots & \vdots \\ P & P & \cdots & 1+\alpha \end{pmatrix}$$

$\alpha(\omega, \sigma_{\mathbf{n}}^2)$ normalized PSD of motion model error

$P(\omega, \sigma_{\mathbf{\Delta}}^2)$ characteristic function of displacement error

Coding One Video Signal with Side Information

- Very accurate disparity compensation
- Consider model error **z** of disparity compensation
- Side information is a noisy version of the video signal to be encoded:

$$\underline{\mathbf{w}}_\mu = \underline{\mathbf{u}} + \underline{\mathbf{z}}_\mu$$

- The set of model error images **z** is statistically independent of the set of input pictures **u**.
- Matrix of conditional power spectral densities:

$$\Phi_{\mathbf{u}|\mathbf{w}} = \Phi_{\mathbf{uu}} \left[(N-1)\Phi_{\mathbf{uu}} + \Phi_{\mathbf{zz}} \right]^{-1} \Phi_{\mathbf{zz}}$$

Conditional Karhunen-Loeve Transform

- Conditional KLT of $\Phi_{\mathbf{u}|\mathbf{w}}(\omega)$ for *K* motion-compensated pictures **u** given *(N-1)K* side information pictures **w**:
  - *First eigenvector adds all components and scales with* $1/\sqrt{K}$
  - *For the remaining eigenvectors, any orthonormal basis can be used that is orthogonal to the first eigenvector*
- Independent of side information, i.e., side information is not required at the encoder
- Motion-compensated Haar wavelet meets these requirements

Relative Conditional Eigendensities

- Compare the conditional eigendensities for collaborative coding $\Lambda_k^*(\omega)$ to the corresponding for non-collaborative coding $\Lambda_k(\omega)$.

$$\frac{\Lambda_1^*(\omega)}{\Lambda_1(\omega)} = \frac{\gamma}{[N-1]Q+\gamma} \cdot \frac{Q + \frac{\gamma K P}{[N-1][Q+KP]+\gamma}}{Q+KP},$$

$$\frac{\Lambda_k^*(\omega)}{\Lambda_k(\omega)} = \frac{\gamma}{[N-1]Q+\gamma}, \qquad k = 2, 3, \ldots, K$$

$\gamma(\omega, \sigma_{\mathbf{z}}^2)$
normalized PSD of
disparity model error

$Q(\omega, \sigma_{\mathbf{n}}^2, \sigma_{\boldsymbol{\Delta}}^2)$
normalized PSD
$Q = 1 + \alpha - P$

Collaborative Coding of Multiple Video Signals

- Power of each camera signal is the same
- Signal of the current sensor always serves as a reference view-point for disparity compensation
- The model error of disparity compensation has the same variance independent of the current sensor/reference view-point

→ Each sensor shows the same rate distortion performance

Performance Bounds via Conditional Eigendensities

- Rate difference to non-collaborative MCT coding for each picture *k* of the ν-th sensor:

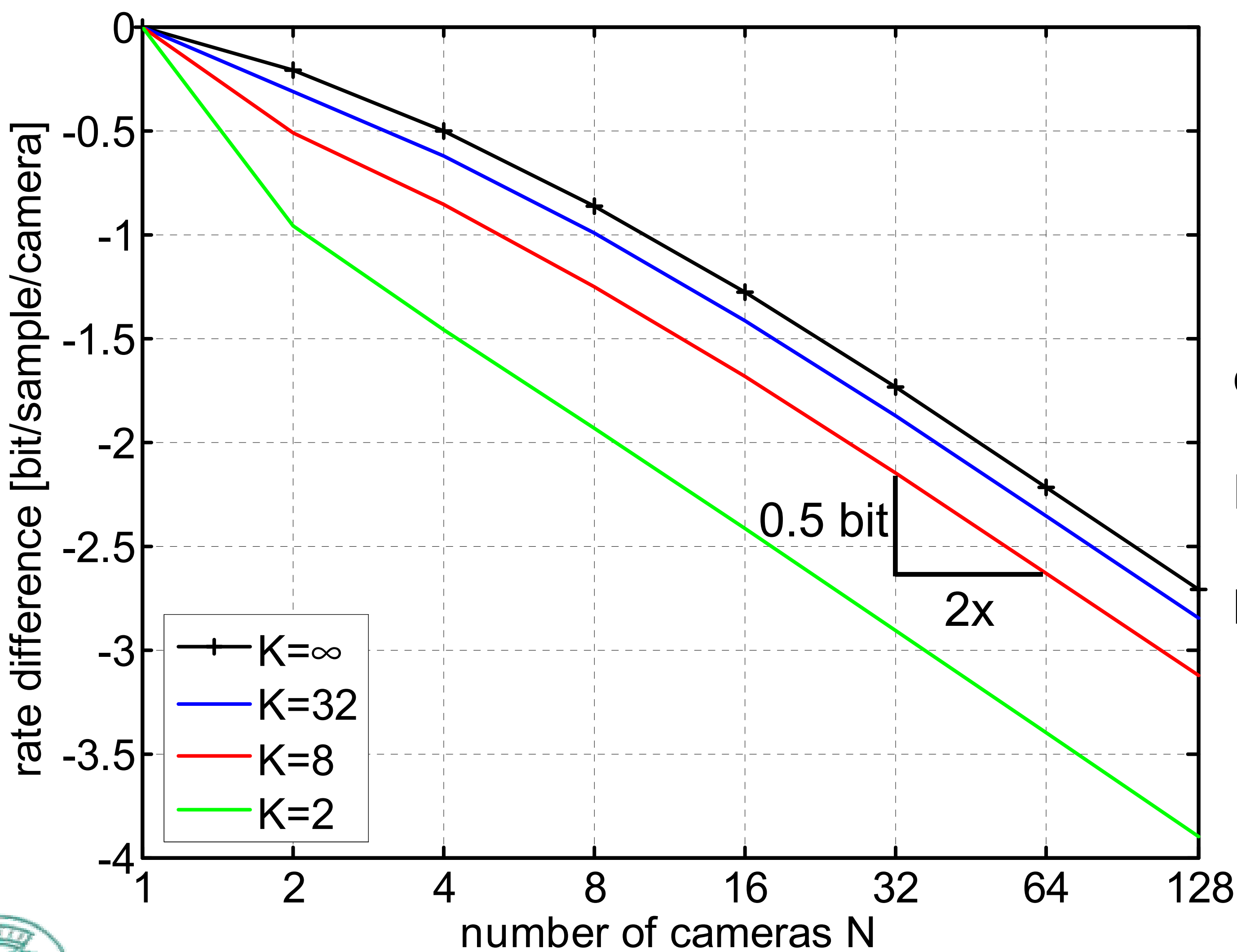
$$\Delta R^*_{\nu,k} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log_2 \left(\frac{\Lambda^*_k(\omega)}{\Lambda_k(\omega)} \right) d\omega$$

  - *Measures maximum bit-rate reduction*
  - *Compares to optimum non-collaborative motion-compensated transform coding*
  - *For the same mean squared reconstruction error*
  - *For Gaussian signals*

- Average rate difference for each camera:

$$\Delta R^* = \frac{1}{NK} \sum_{\nu=1}^{N} \sum_{k=1}^{K} \Delta R^*_{\nu,k}$$

Rate Difference with Gaussian Assumption at High Rates

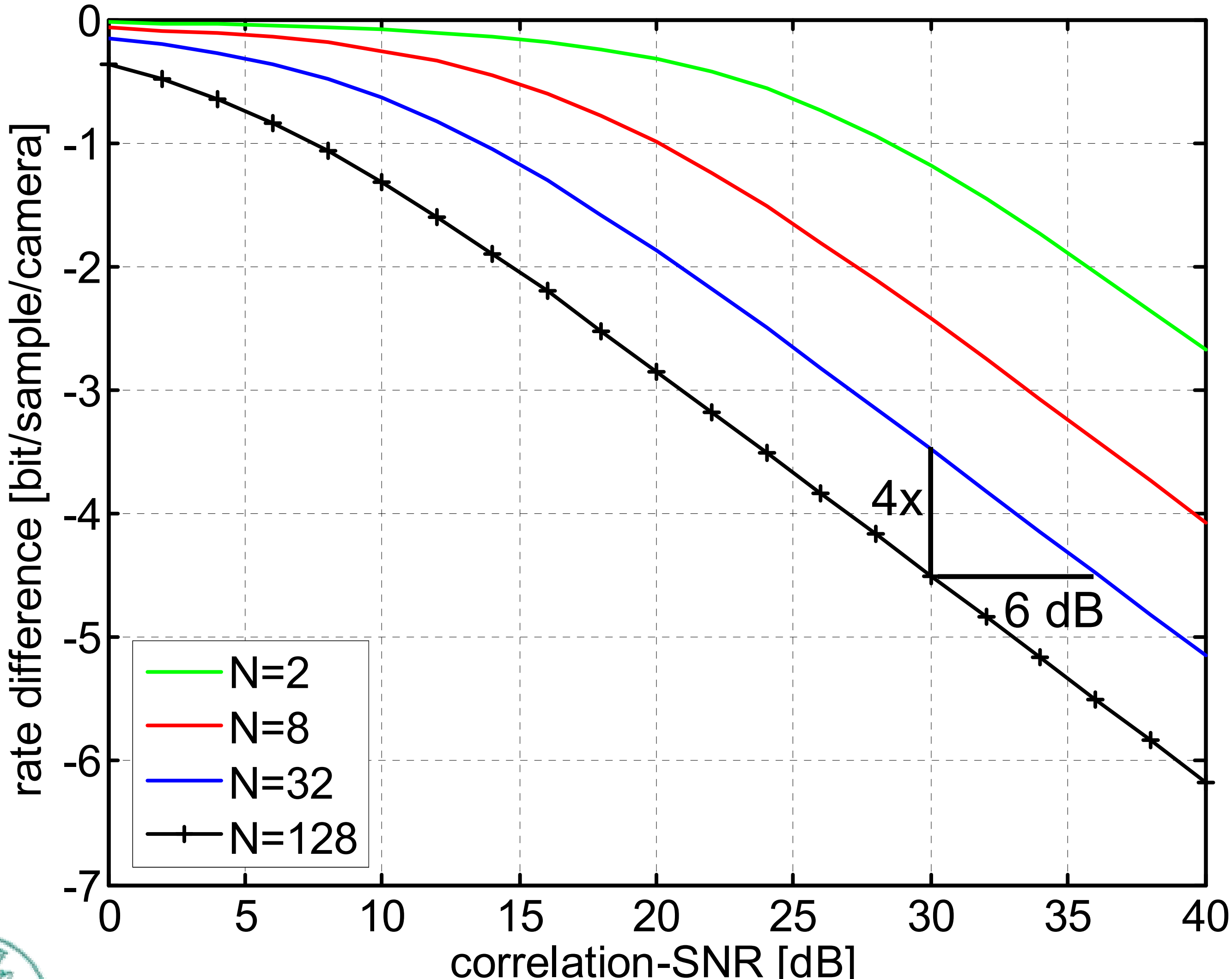


c-SNR = 20 dB

RNL = -30 dB

half-pel accuracy

Rate Difference with Gaussian Assumption at High Rates



K = 32

RNL = -30 dB

half-pel accuracy

Conclusions

- For *N* video sensors, we have compared the efficiency of collaborative motion-compensated coding to non-collaborative motion-compensated coding in terms of rate difference at high rate.
- For a large number of cameras,
  - *doubling the number decreases the rate difference at most by 0.5 bit per sample per camera.*
  - *quadrupling the number compensates the correlation-SNR at least by 6 dB.*