# A Video Codec Incorporating Block-Based Multi-Hypothesis Motion-Compensated Prediction

*Markus Flierl, Thomas Wiegand*

**Telecommunications Laboratory**
**University of Erlangen-Nuremberg**
**Erlangen, Germany**

**[flierl,wiegand]@LNT.de**

*Bernd Girod*

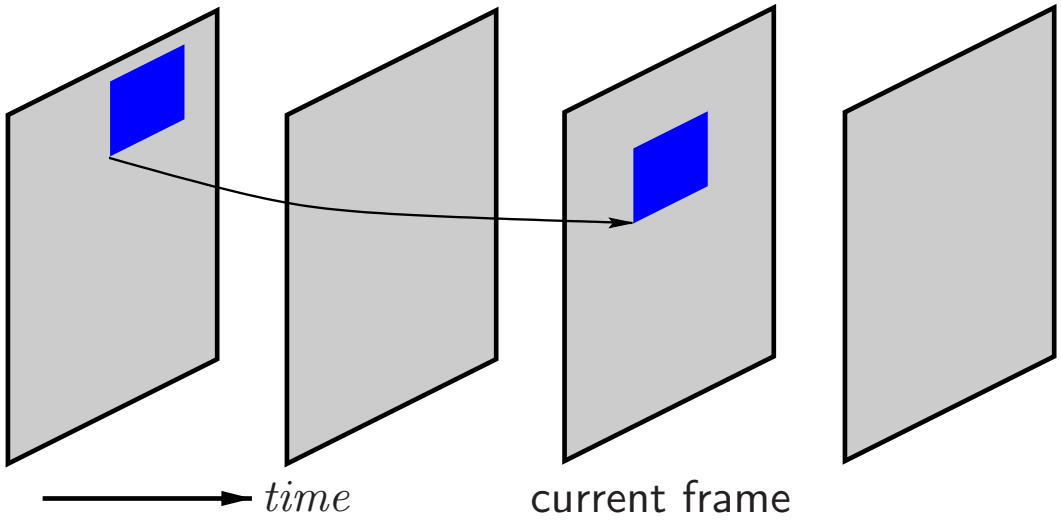
**Information Systems Laboratory**
**Stanford University**
**Stanford, CA, USA**

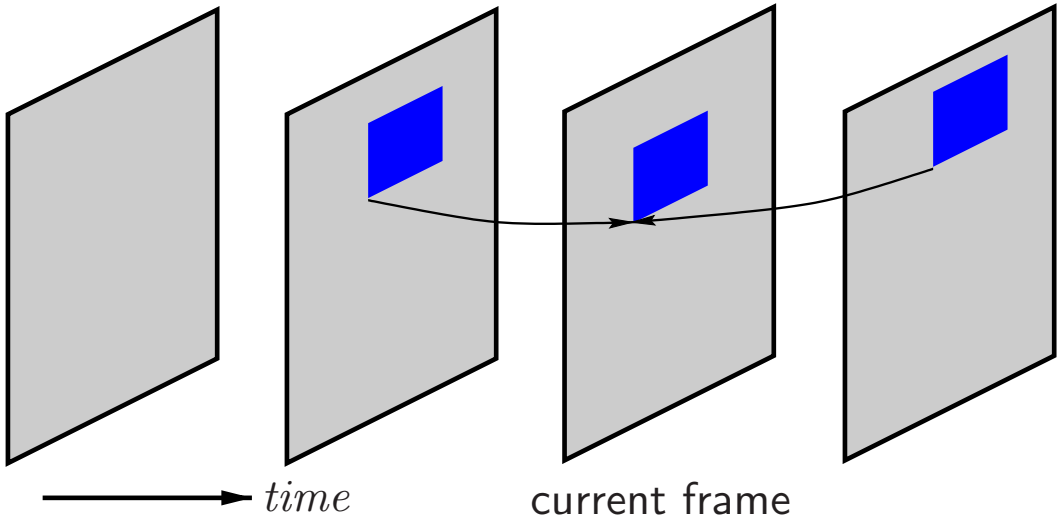
**girod@ee.stanford.edu**

# Overview

1. Multi-hypothesis motion-compensated prediction
2. Rate-constrained multi-hypothesis motion estimation
3. Integration into a hybrid video coder
4. Coding efficiency of multi-hypothesis prediction
5. Multi-hypothesis and variable block size prediction

# Motivation



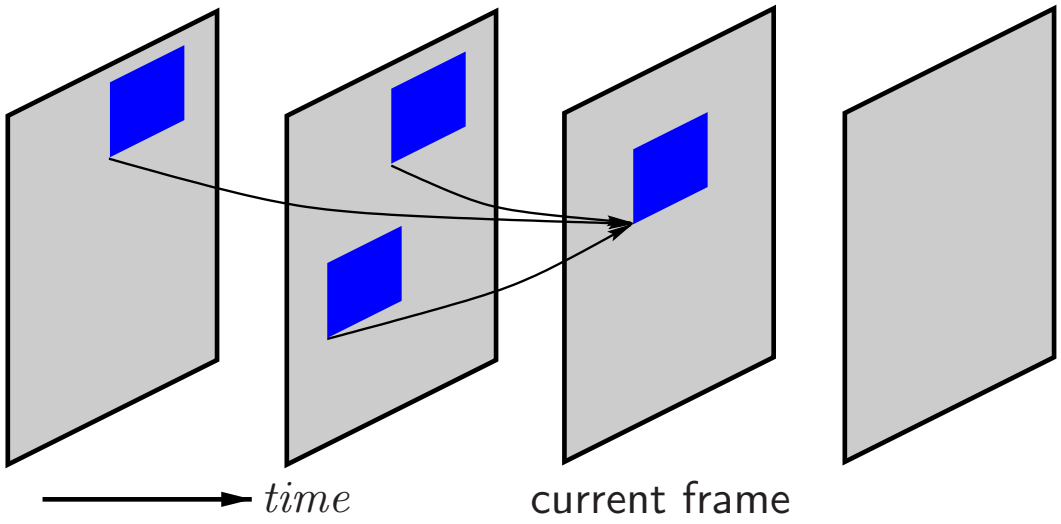




Long-term memory prediction for P-frame coding

*How about combining several prediction signals from the past?*

Bidirectional prediction for B-frame coding

# Multi-Hypothesis Motion-Compensated Prediction




Multi-hypothesis prediction for P-frame coding

$\Rightarrow$ Each prediction signal (hypothesis) is assigned a motion vector and picture reference

$\Rightarrow$ Hypotheses are linearly combined with constant scalar weights

$\Rightarrow$ Hypotheses are chosen only from previous decoded frames

# Multi-Hypothesis Motion Estimation

$\Rightarrow$ Improved prediction performance and higher bit-rate due to more than one hypothesis per block

$\Rightarrow$ A trade-off between prediction performance and bit-rate is necessary

$\hookrightarrow$ Rate-constrained multi-hypothesis motion estimation

$\Rightarrow$ The complexity of a full search algorithm for a $N$-hypothesis grows exponentially with $N$

$\hookrightarrow$ Successive improvement of $N$ optimal conditional solutions by an iterative algorithm

# Hypothesis Selection Algorithm

**0:** Assuming $N$ hypotheses $(c_1, \ldots, c_N)$, the rate-distortion cost function

$$j(c_1, \ldots, c_N) = \left\| s - \frac{1}{N} \sum_{\nu=1}^{N} c_\nu \right\|_2^2 + \lambda \sum_{\nu=1}^{N} r(c_\nu)$$

is subject to minimization for each original block $s$, given the Lagrange multiplier $\lambda$. Set $i := 0$ and guess $N$ initial hypotheses $(c_1^{(0)}, \ldots, c_N^{(0)})$.

**1:** For each hypothesis $\mu$:

**a:** Select the $\mu$-th hypothesis. All others are held constant.

**b:** Minimize the rate-distortion cost function by full search for hypothesis $c_\mu^{(i+1)}$

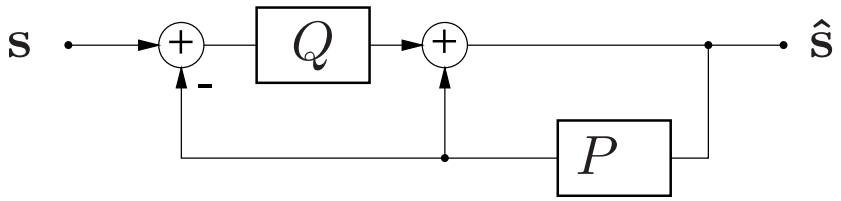
$$\min_{c_\mu^{(i+1)}} j(c_1^{(i+1)}, \ldots, c_{\mu-1}^{(i+1)}, c_\mu^{(i+1)}, c_{\mu+1}^{(i)}, \ldots, c_N^{(i)})$$

**2:** Set $i := i + 1$ and continue with step 1 until convergence.

# Multi-Hypothesis Coding Modes for H.263

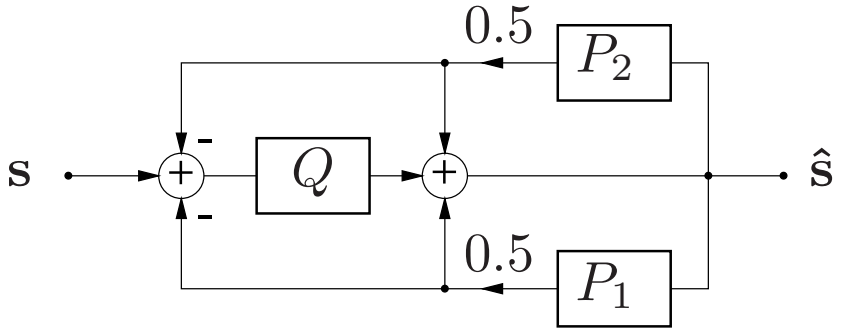
⇒ INTER-Mode

- 1 motion vector and picture reference per block
- Data for residual encoding




⇒ INTER2H-Mode

- 2 motion vectors and picture references per block
- Data for residual encoding




⇒ INTER4V2H-Mode

Multi-hypothesis block pattern indicates 1 or 2 hypotheses per $8 \times 8$ block

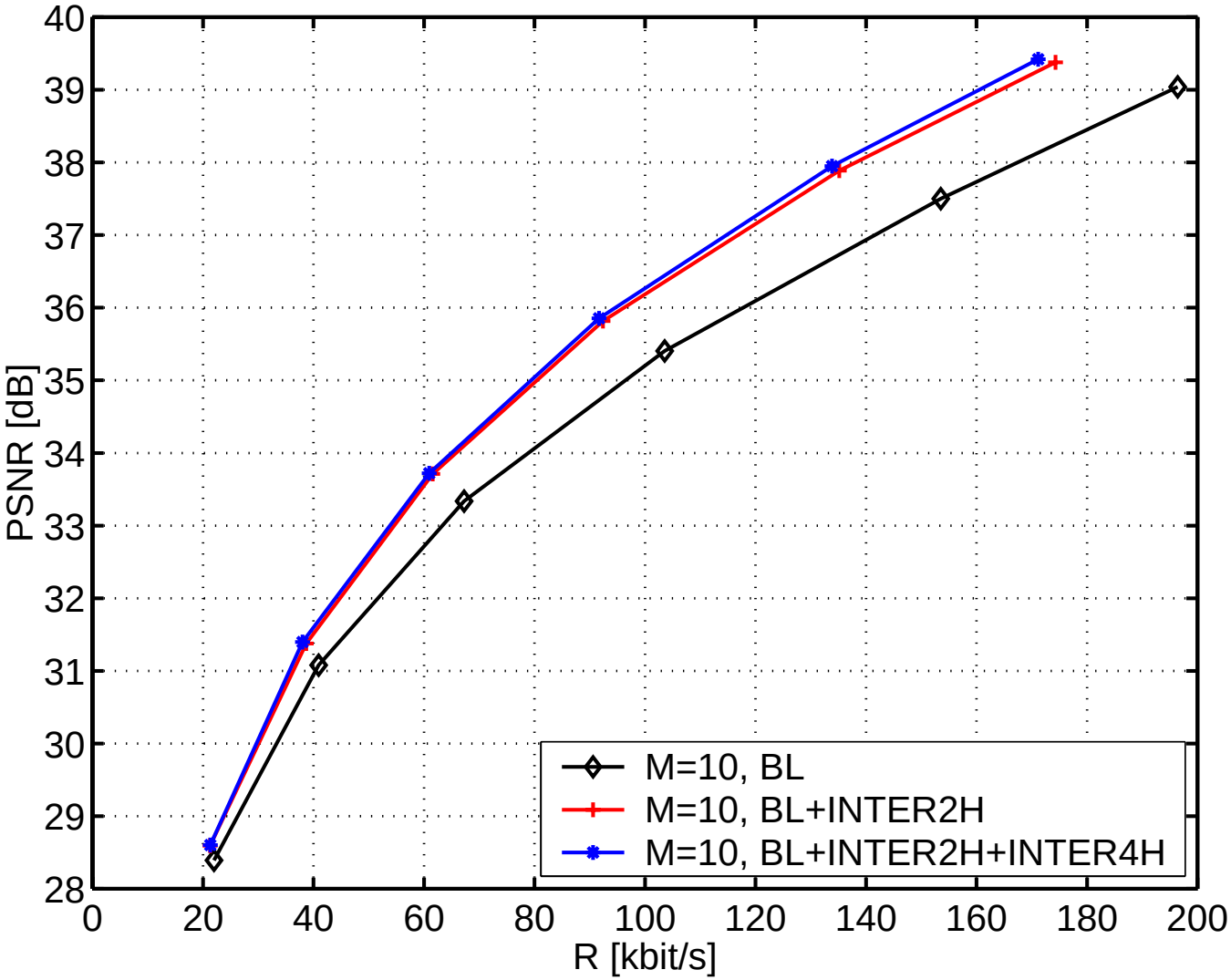
# Rate-Constrained Mode Decision

- Multi-hypothesis prediction improves the prediction signal by spending more bits for the side-information
- Encoding of the prediction error and its associated bit-rate also determine the quality of the reconstructed block

$\Rightarrow$ Rate-constrained multi-hypothesis motion estimation independent of prediction error encoding is an efficient and practical solution

$\Rightarrow$ The efficient number of hypotheses for each block has to be determined by rate-constrained mode decision

# Coding Efficiency of Multi-Hypothesis Prediction



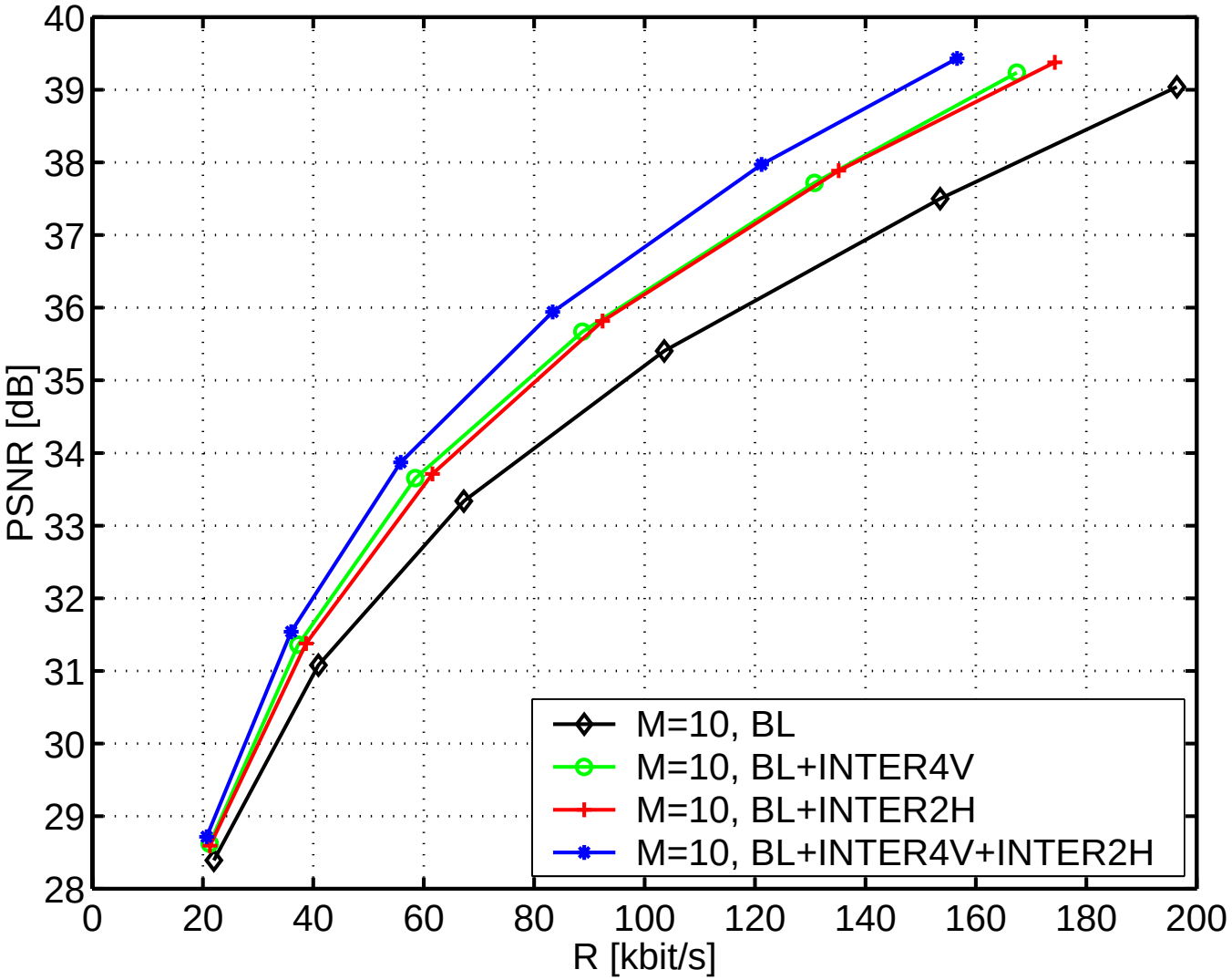

**Foreman** (QCIF, 10 fps, 10 s), $M = 10$ reference frames

⇒ Decreasing prediction performance for more than 2 hypotheses

⇒ Rate constraint prohibits a large number of hypotheses

⇒ 2 hypotheses are efficient

# Multi-Hypothesis and Variable Block Size Prediction



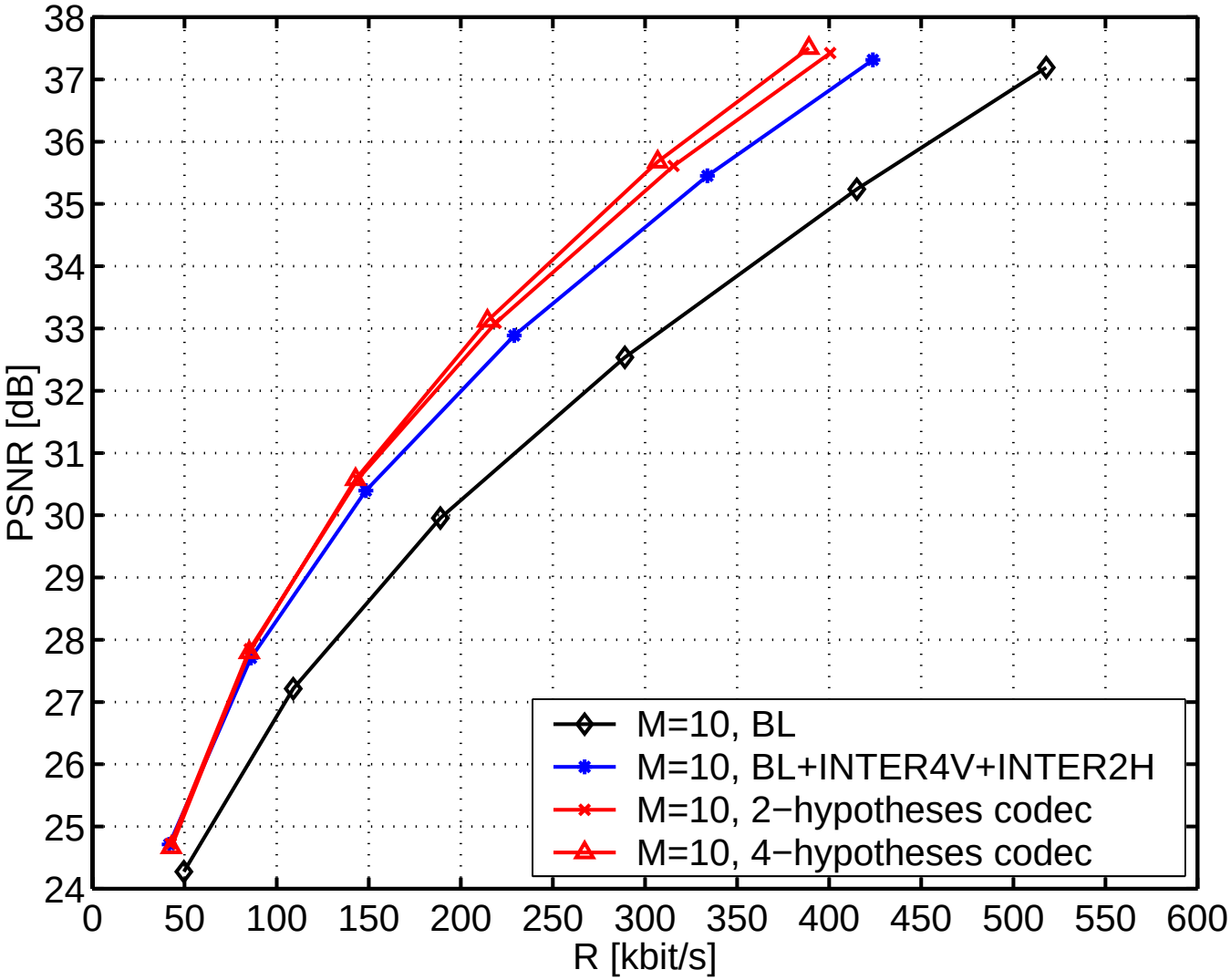

**Foreman** (QCIF, 10 fps, 10 s), $M = 10$ reference frames

$\Rightarrow$ INTER4V: 4 motion vectors and picture references for 4 $8 \times 8$ blocks

$\Rightarrow$ INTER2H: 2 motion vectors and picture references per macroblock

$\Rightarrow$ Comparable coding performance but for different scenarios

# Multi-Hypothesis Prediction for Variable Block Size



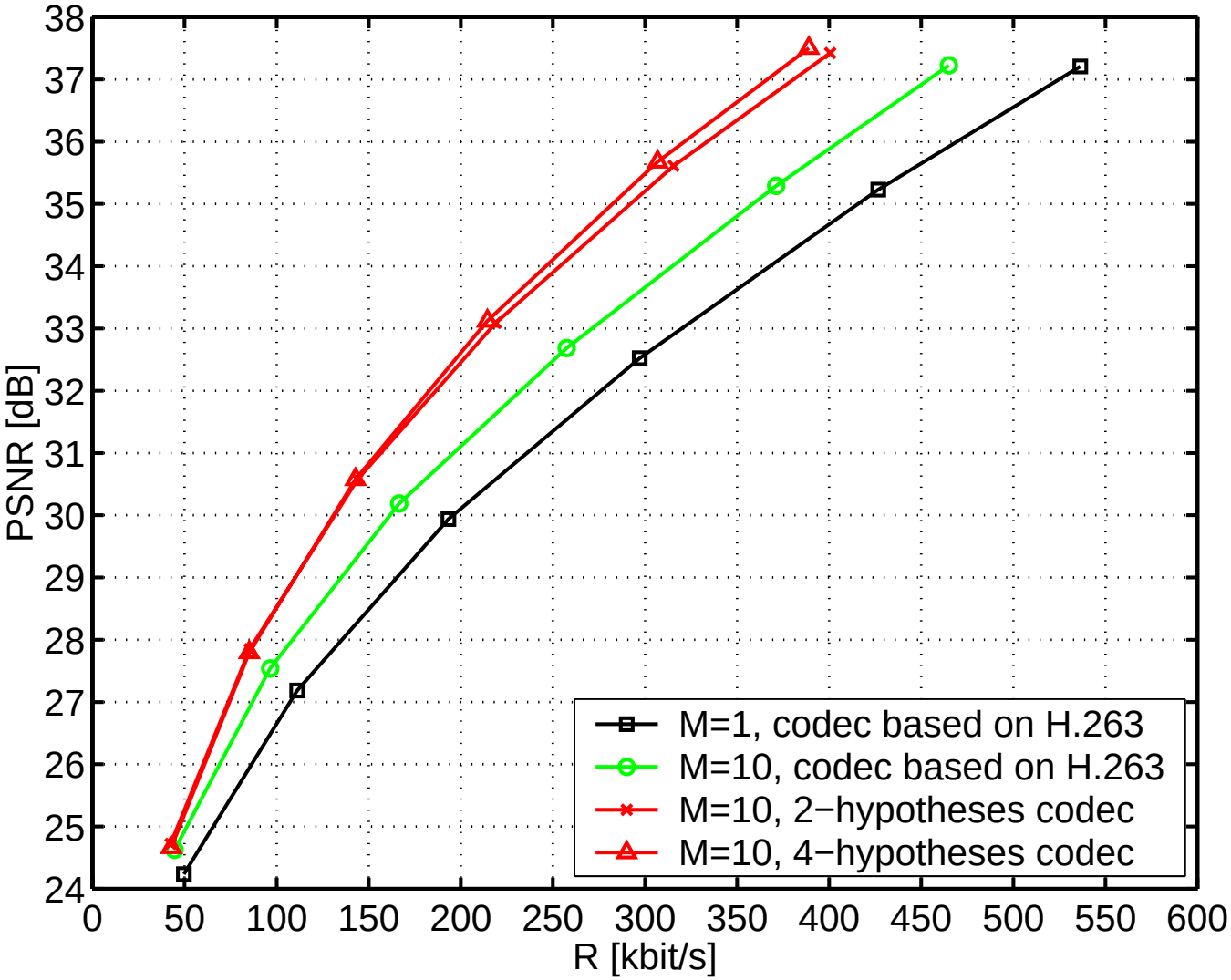

**Mobile & Calendar** (QCIF, 10 fps, 10 s), 10 ref. frames

⇒ Prediction with 2 and 4 hypotheses for $16 \times 16$ and $8 \times 8$ blocks

⇒ Additional gain for multiple $8 \times 8$ hypotheses

⇒ 2 hypotheses are efficient

# Comparison to the Reference Codec




**Mobile & Calendar** (QCIF, 10 fps, 10 s)

⇒ Gain up to 1.2 dB for long-term memory prediction with 10 reference frames

⇒ Additional gain up to 1.5 dB with 2 hypotheses

⇒ In total, up to 2.7 dB for 2-hypothesis long-term memory prediction

# Conclusions

⇒ Multi-hypotheses prediction improves efficiency of standard video compression algorithms. We observe up to 2.7 dB coding gain for the 2-hypothesis codec with 10 reference frames.

⇒ The efficient number of hypotheses for each block has to be determined by rate-constrained mode decision.

⇒ Practical video coding schemes should utilize two jointly optimized hypotheses. (Theoretical investigations on the efficient number of hypotheses are presented in the proceedings.)

⇒ Variable block size and multi-hypothesis prediction can be successfully combined.