Rate-Constrained Multi-Hypothesis Motion-Compensated Prediction for Video Coding

Markus Flierl

Telecommunications Laboratory
University of Erlangen-Nuremberg
Erlangen, Germany

flierl@LNT.de

Thomas Wiegand

Image Processing Department
Heinrich-Hertz-Institute
Berlin, Germany

wiegand@hhi.de

Bernd Girod

Information Systems Laboratory
Stanford University
Stanford, CA, USA

girod@ee.stanford.edu

Overview

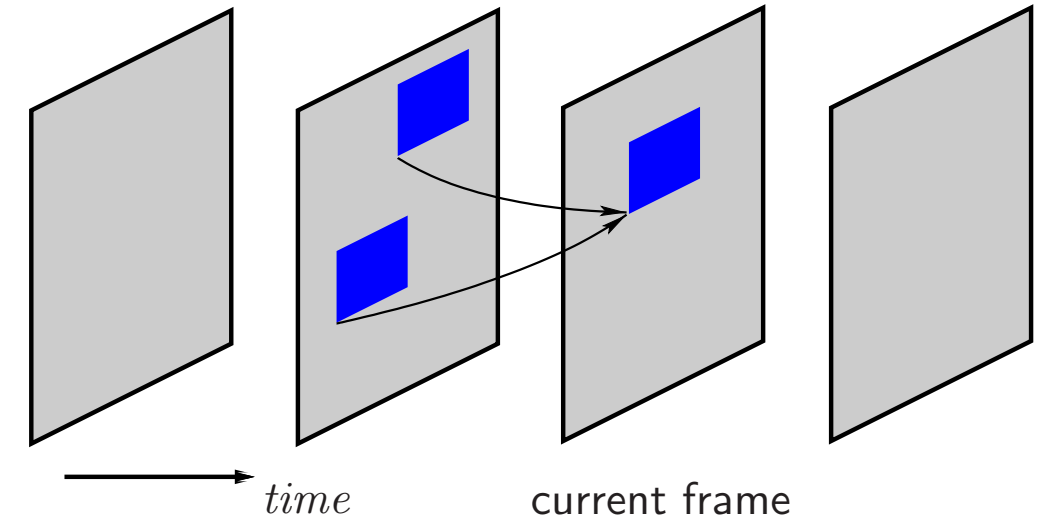
1. Multi-hypothesis motion-compensated prediction with multiple reference frames

2. Rate-constrained multi-hypothesis motion estimation

3. Integration into a hybrid video coder

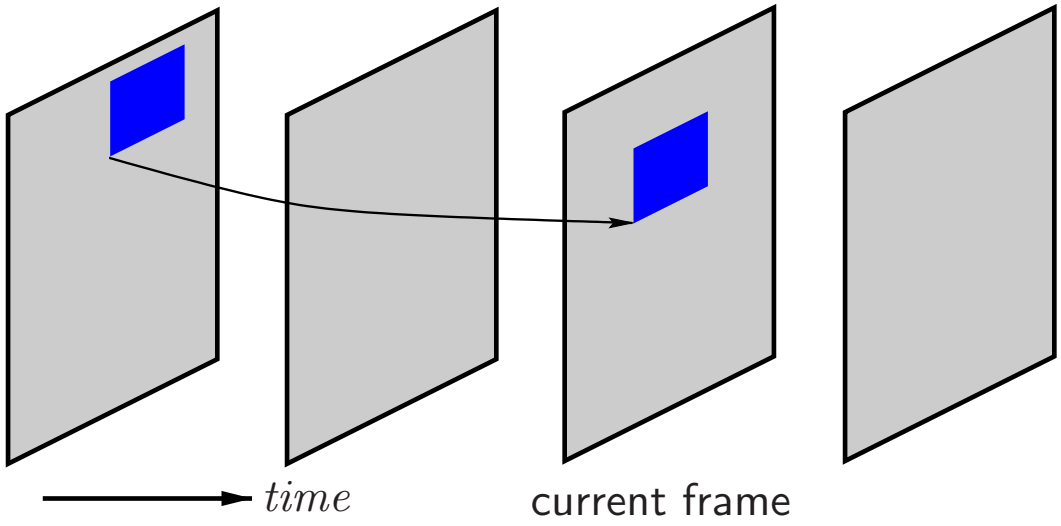
4. Experimental results for multiple hypotheses on multiple reference frames

Motivation



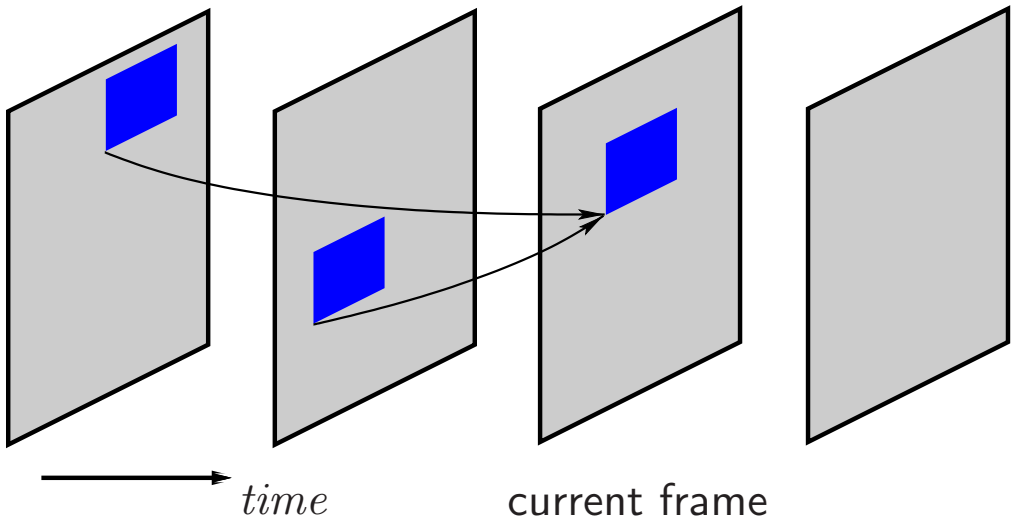
Multi-hypothesis prediction for P-Frame coding with one reference frame

What do we gain when we combine both concepts?



Long-term memory prediction for P-frame coding (requires picture reference)

Multi-Hypothesis Motion-Compensated Prediction



Multi-hypothesis prediction for P-frame coding with two hypotheses and multiple reference frames

$\Rightarrow$ 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 (motion vectors and picture references) is necessary

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

$\Rightarrow$ A full search algorithm for a 2-hypothesis is not practical.

$\hookrightarrow$ Successive improvement of 2 suboptimal conditional solutions by an iterative algorithm

Hypothesis Selection Algorithm

0: Assuming 2 hypotheses (c_1, c_2), the rate-distortion cost function

$$j(c_1, c_2) = \left\| s - \frac{1}{2}c_1 - \frac{1}{2}c_2 \right\|_2^2 + \lambda[r(c_1) + r(c_2)]$$

is subject to minimization for each original block s, given the Lagrange multiplier λ. Set $i := 0$ and guess 2 initial hypotheses $(c_1^{(0)}, c_2^{(0)})$.

1: Minimize the rate-distortion cost function by full search for

a: hypothesis $c_1^{(i+1)}$ while fixing hypothesis $c_2^{(i)}$

$$\min_{c_1^{(i+1)}} j(c_1^{(i+1)}, c_2^{(i)})$$

b: and hypothesis $c_2^{(i+1)}$ while fixing the complementary hypothesis.

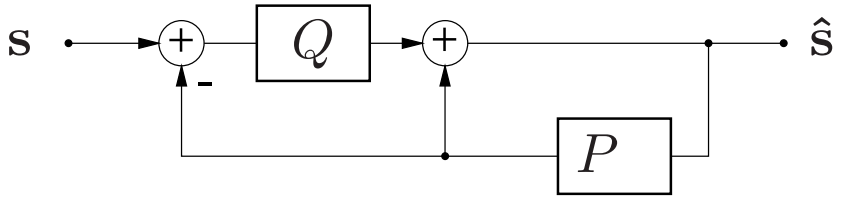
$$\min_{c_2^{(i+1)}} j(c_1^{(i+1)}, c_2^{(i+1)})$$

2: Set $i := i + 1$ and continue with step 1 as long as the cost function decreases.

Multi-Hypothesis Modes for H.263

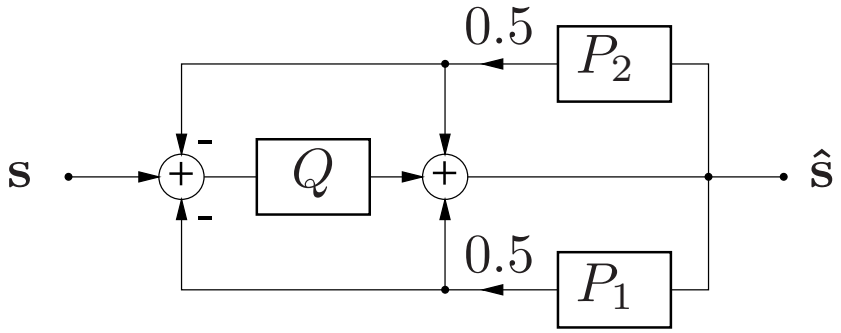
⇒ INTER-Mode

- 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 8×8 block

Rate-Constrained Decisions

- 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 determines the quality of the reconstructed block

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

$\Rightarrow$ A rate-constrained decision which also incorporates the encoding of the prediction error determines whether 1 or 2 hypotheses are used.

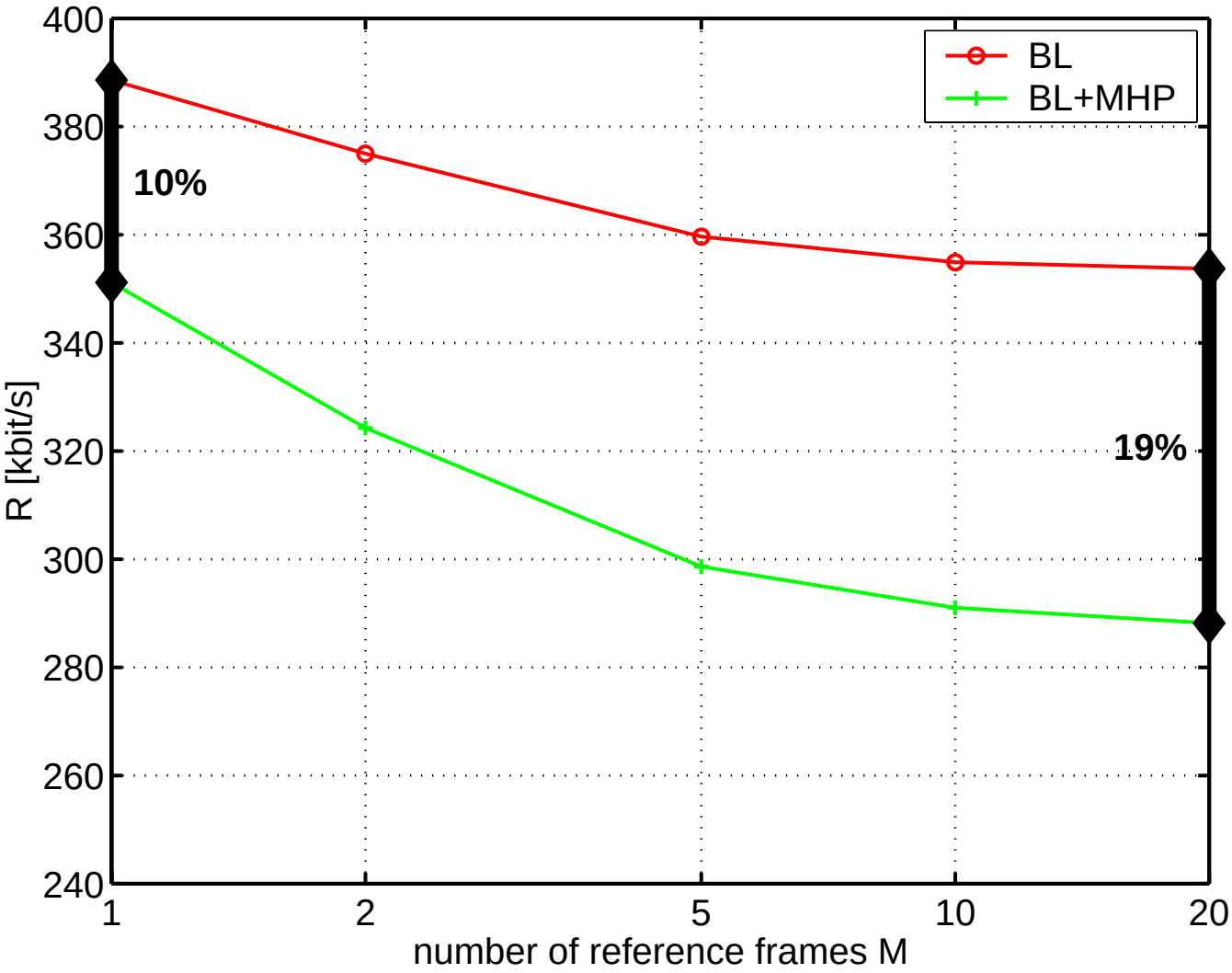
Previous Results

$\Rightarrow$ Practical video coding schemes should utilize two jointly optimized hypotheses. (VCIP 2000)

$\Rightarrow$ Theoretical investigations on the efficient number of hypotheses support this finding. (VCIP 2000)

$\Rightarrow$ Variable block size and multi-hypothesis prediction can be successfully combined. (VCIP 2000)

Multiple Hypotheses on Multiple Reference Frames



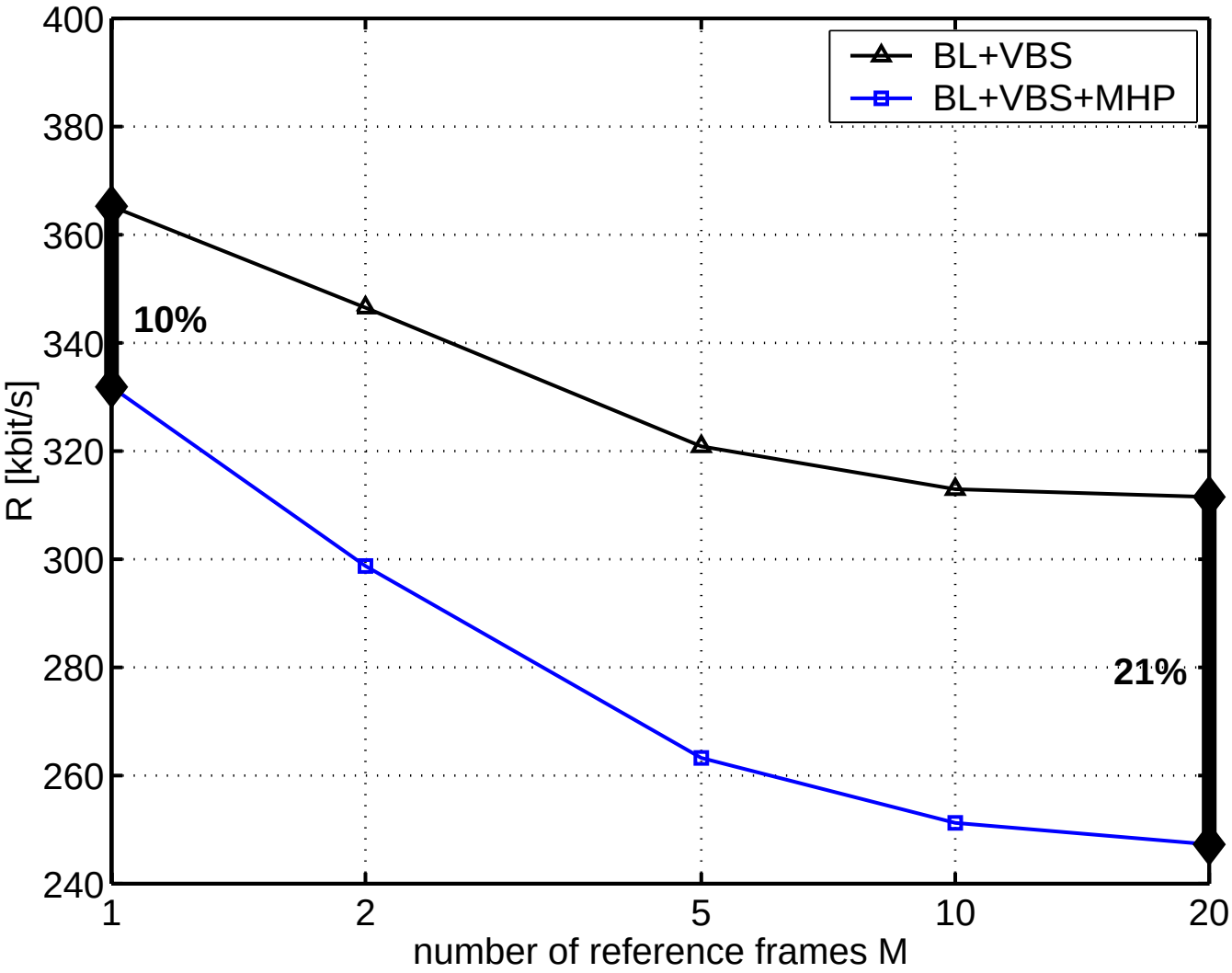
- Fixed block size prediction with and without MH prediction

⇒ MH prediction with 1 reference frame saves 10% of bit-rate

⇒ MH prediction with 20 reference frames saves 19% of bit-rate

Mobile & Calendar (QCIF, 10 fps, 10 s) at 34 dB PSNR

Multiple Hypotheses on Multiple Reference Frames



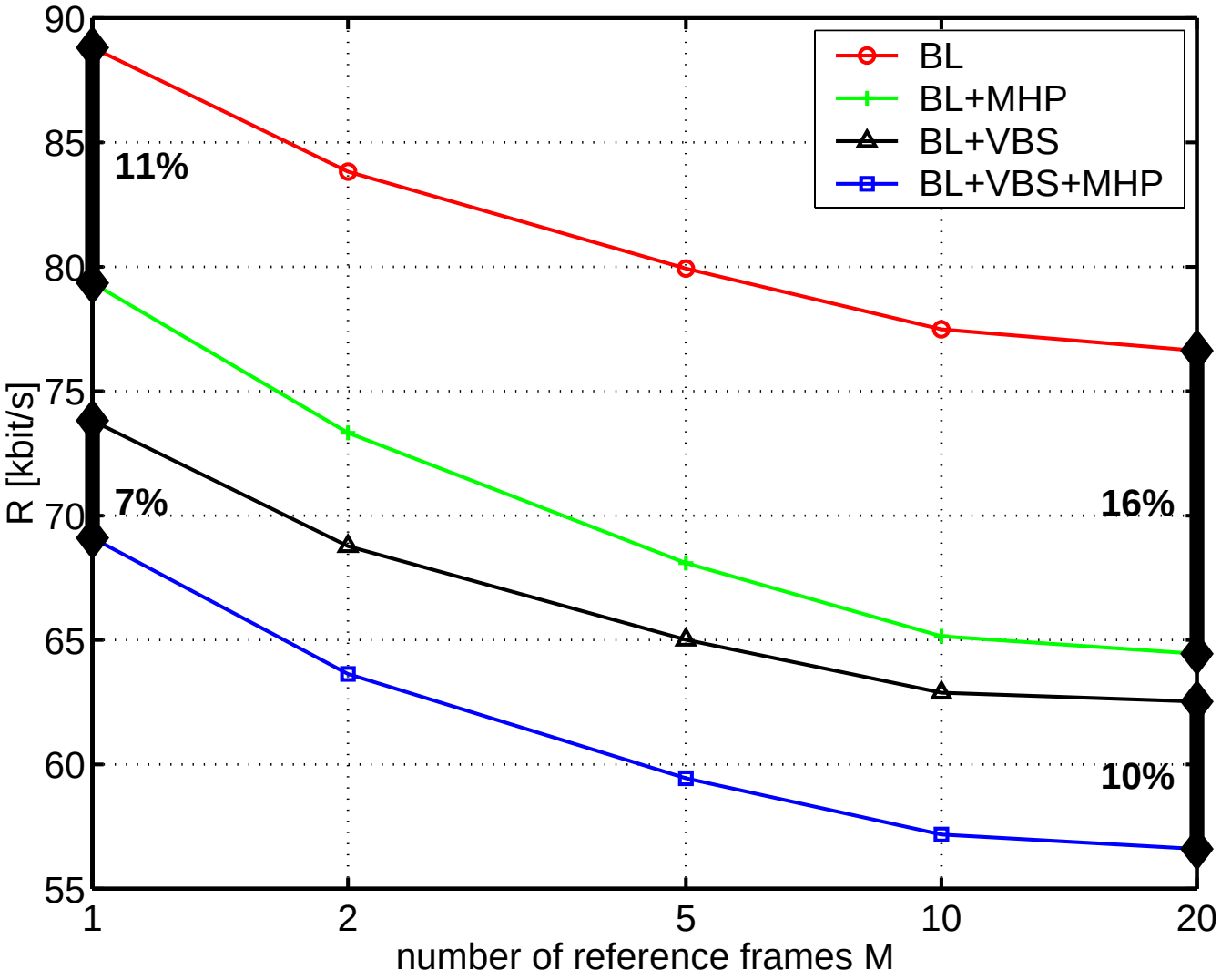
Mobile & Calendar (QCIF, 10 fps, 10 s) at 34 dB PSNR

- Variable block size prediction with and without MH prediction

⇒ MH prediction with 1 reference frame saves 10% of bit-rate

⇒ MH prediction with 20 reference frames saves 21% of bit-rate

Multiple Hypotheses on Multiple Reference Frames



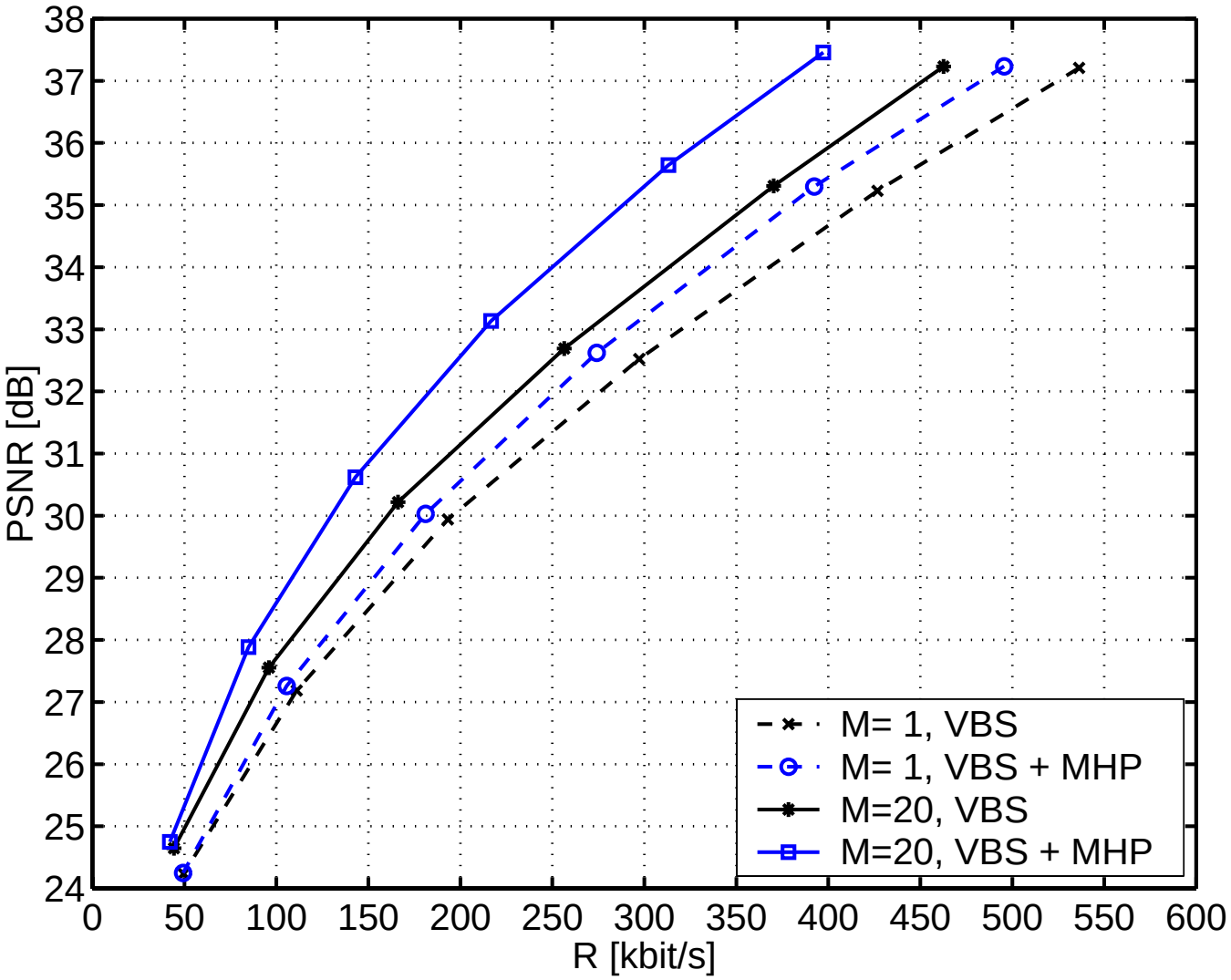
Foreman (QCIF, 10 fps, 10 s) at 34 dB PSNR

- Fixed and variable block size prediction with and without MH prediction

⇒ MH prediction with 1 reference frame saves 11% for FBS and 7% for VBS prediction

⇒ MH prediction with 20 reference frames saves 16% for FBS and 10% for VBS prediction

Multiple Hypotheses on Multiple Reference Frames



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

- VBS + MH prediction with 1 and 20 reference frames.

⇒ MH prediction with 1 reference frame gains up to 0.8 dB

⇒ MH prediction with 20 reference frame gains up to 1.6 dB

⇒ Efficiency of MH prediction improves for a larger number of reference frames

Example: Mobile & Calendar

$M = 1$ without MHP
Sequence at 200 kbit/s and 30 dB

$M = 20$ with MHP
Sequence at 200 kbit/s and 33 dB

Conclusions

$\Rightarrow$ Long-term memory enhances the efficiency of multi-hypothesis motion-compensated prediction for video coding.

$\Rightarrow$ This observation is independent from the block size used for prediction.

$\Rightarrow$ The bit-rate savings saturate for 20 reference frames.