# Transactions Letters

## Jointly Optimized Mode Decisions in Redundant Video Streaming

J. Christian Schmidt and Kenneth Rose, *Fellow, IEEE*

***Abstract*—This letter investigates source-channel coding for error-resilient video streaming using redundant encoding. We estimate the end-to-end distortion per redundantly encoded macroblock (MB) via extension of the recursive optimal per-pixel estimate to encompass redundant transmissions. Redundant encoding is formulated as joint optimization of the MB parameters in the primary and redundant transmissions. We present three encoding strategies with different gain-complexity tradeoffs. The proposed methods are general in nature, and could be implemented on top of any (hybrid) video codec. Simulation results employing H.264's redundant slice mechanism show significant performance gains over conventional error-resilient encoding methods and naive redundant encoding schemes.**

***Index Terms*—Error resilience, H.264, redundant slices, source-channel coding, video streaming.**

### I. Introduction

Video streaming over packet networks remains a fundamental challenge due to the best-effort nature of the underlying network and lack of end-to-end quality of service (QoS). This motivates ongoing research into error resilience mechanisms to mitigate the impact of packet loss.

A traditional video streaming system consists of several key modules which can be modified to improve error robustness. The encoder may adjust its macroblock (MB) coding mode decisions, e.g., using intra refresh to stop error propagation. Error propagation can be reduced by appropriate selection of motion parameters such as motion vector (MV) and reference frame. Packetization techniques can improve error resilience by dispersing spatially adjacent MBs into different packets [e.g., H.264's flexible macroblock ordering (FMO)], or scattering data temporally to guard against burst losses. Data partitioning can be used in tandem with unequal error protection. At the transport level, channel coding tools such as forward error correction (FEC) or automatic re-transmission requests (ARQ) can protect data packets. Finally, an error-resilient decoder may perform suitable error concealment.

Manuscript received July 15, 2009; revised May 20, 2010; accepted August 15, 2010. Date of publication October 14, 2010; date of current version April 1, 2011. This work was supported in part by the University of California MICRO Program, Applied Signal Technology, Inc., Cisco Systems, Inc., Dolby Laboratories, Inc., Qualcomm, Inc., and Sony-Ericsson, Inc. This paper was recommended by Associate Editor S. Li.

The authors are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 USA (e-mail: schmidt@ece.ucsb.edu; rose@ece.ucsb.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2010.2087453

Rate-distortion optimization for live video streaming depends critically on accurate end-to-end distortion estimation. We resort to the "recursive optimal per-pixel estimate" (ROPE) [1], [2], which has been successfully applied to MB coding mode and quantization parameter (QP) selection [1], [2], error-resilient motion estimation/compensation [3] and reference picture selection [4], multiple description (video) coding (MDC) [5], and joint mode and QoS selection [6]. An error-resilient encoder must balance the conflicting objectives of minimizing error propagation versus efficient compression. Rather than allocate the entire bit budget for source coding, some rate may be designated for channel protection, trading some source coding fidelity for reduced effective packet loss rate (PLR). Channel coding mechanisms are subject to practical drawbacks. ARQ trades feedback effectiveness versus latency and scalability. FEC rate allocation is usually performed at the packet level, *after* encoding. Due to uneven packet sizes and padding, it is difficult to estimate the effective coding rate and loss probability at encode time. A Trellis-based algorithm has been proposed in [6], albeit at the cost of delay and complexity. In practice, channel coding mechanisms lower the effective PLR experienced by the source coder, mitigating but not eliminating the traditional issue of error propagation.

Recently, redundant encoding has been proposed [9]–[18], e.g., as enabled by the redundant slice mechanism in H.264 [7]. Wu and Boyce [9] investigated streaming of precompressed video. After offline compression with several redundant representations of variable quality, the delivery policy (which redundant slices to transmit) is optimized during transmission. Baccichet *et al.* [10] proposed a redundant encoding scheme for live video streaming. The primary picture is coded independently from the redundant slices, which may cover the entire picture or just a region of interest (ROI). Redundant MBs use the same coding parameters as the primary MBs, but a larger QP selected on a GOP-by-GOP basis via a concealment distortion model. In the MDC scheme in [11], each frame is split into two slice groups. Two balanced descriptions are generated by combining the primary slices from one group with the redundant slices from the other.

In [12], a custom FMO pattern distributes MB losses and enables improved concealment. To lower the risk of simultaneous primary and redundant slice loss due to channel fading, a method for dynamic redundant slice allocation is presented. Another approach transmits the edge information of the encoded frame via redundant slices, guiding image-inpainting

1051-8215/$26.00 © 2010 IEEE

concealment via structural hints [13]. [14] proposes redundant encoding of the MVs without residual.

Rane *et al.* [15] combined redundant slices and Reed-Solomon coding to generate a Wyner-Ziv bitstream. The parity information is generated over the redundant slices, which use the same MB parameters as the primary bitstream, but a higher QP. The decoder regenerates redundant data for received primary slices, and uses it and the received parity information to recover the redundant information for *lost* primary data.

Finally, two methods for picture-adaptive redundant encoding are proposed in [16]. Hierarchical encoding uses a fixed structure that divides each GOP into smaller units, and allocates redundant pictures between them. Adaptive redundant picture allocation identifies which pictures need protection most, based on heuristics for distortion propagation. We have previously investigated redundant encoding in [17] and [18]; this letter builds on our previous contributions in this area.

## II. Redundant Encoding

### A. Motivation

Several issues affect redundant encoding performance. In terms of redundancy allocation, adaptability is coarse. Some methods redundantly encode every frame [10], [11]. Others protect only some frames [16] or a ROI [10], [12], which only works well in scenarios with a clear ROI. Redundancy allocation is based on heuristic distortion models. To adjust the redundant data rate, the QP is usually increased by an offset over the primary QP [11], [16]. [14] only retransmits the MVs, but no coded residual. Since not all MBs contribute equally toward end-to-end distortion, a locally adaptive approach should allocate rate toward important areas by adjusting the redundant QP, based on accurate distortion estimation. This is a tradeoff of limiting error propagation versus introducing drift when the secondary data is reconstructed.

Finally, there is the issue of coding mode selection for the primary and redundant MB encodings, and overall RD optimization (RDO). Both encodings contribute to the rate and end-to-end distortion costs. Hence, their coding parameters and rates should be decided jointly. Existing work simplifies mode decision, e.g., encoding the primary and redundant picture independently (JM reference software [8] or as proposed in [16]), or using the same MB modes for both encodings [10], [11], [15]. Joint MB-level decisions naturally provide the local adaptivity outlined above, and enable the encoder to optimize the error resilience versus drift tradeoff. Previously, error-resilient live video streaming has been posed as optimization of the MB coding mode and QP [1]. In this letter, we investigate redundant encoding as a mode decision problem over the joint space of the primary and redundant MBs.

### B. End-to-End Distortion Estimation

Let $f_n^i$ denote the original value of pixel $i$ in frame $n$, and let $\hat{f}_n^i$ denote its *encoder* reconstruction. The reconstructed value at the *decoder* is $\tilde{f}_n^i$. The expected distortion for this pixel is

$$\begin{aligned} d_n^i &= E\{(f_n^i - \tilde{f}_n^i)^2\} \\ &= (f_n^i)^2 - 2f_n^i E\{\tilde{f}_n^i\} + E\{(\tilde{f}_n^i)^2\}. \end{aligned} \tag{1}$$

The computation of $d_n^i$ requires the first and second moments of each random variable in the sequence $\tilde{f}_n^i$, $(\tilde{f}_n^i)^2$. These can be computed recursively, using the previous frame's moments $\tilde{f}_{n-1}^i$, $(\tilde{f}_{n-1}^i)^2$. For redundant encoding with a primary and a secondary description (in separate packets), there are four possible channel outcomes (denoted by the binary variables $b_1^i, b_2^i$), leading to three different decoding results:

1) primary data received and reconstructed (the secondary data is irrelevant in this case, i.e., $b_1^i = 0, b_2^i = 0, 1$);
2) primary data lost, but secondary data received and reconstructed ($b_1^i = 1, b_2^i = 0$), and finally;
3) both transmissions lost and the affected region needs to be concealed ($b_1^i = b_2^i = 1$).

Assuming iid packet loss for simplicity,[1] these outcomes have probabilities $1-p$, $p(1-p)$, and $p^2$. The first and second moments of $\tilde{f}_n^i$, $(\tilde{f}_n^i)^2$ can then be calculated as

$$\begin{aligned} E\{\tilde{f}_n^i\} = \quad & (1-p)E\{\tilde{f}_n^i | b_1^i = 0\} \\ & + p(1-p)E\{\tilde{f}_n^i | b_1^i = 1, b_2^i = 0\} \\ & + p^2 E\{\tilde{f}_n^i | b_1^i = b_2^i = 1\} \end{aligned} \tag{2}$$

$$\begin{aligned} E\{(\tilde{f}_n^i)^2\} = \quad & (1-p)E\{(\tilde{f}_n^i)^2 | b_1^i = 0\} \\ & + p(1-p)E\{(\tilde{f}_n^i)^2 | b_1^i = 1, b_2^i = 0\} \\ & + p^2 E\{(\tilde{f}_n^i)^2 | b_1^i = b_2^i = 1\}. \end{aligned} \tag{3}$$

Let $\tilde{f}_{n,1}^i, (\tilde{f}_{n,1}^i)^2$ and $\tilde{f}_{n,2}^i, (\tilde{f}_{n,2}^i)^2$ denote the *successful* reconstruction and its squared value of the primary and secondary coded data, respectively. With

$$E\{\tilde{f}_n^i | b_1^i = 0\} = E\{\tilde{f}_{n,1}^i\} \tag{4}$$

$$E\{(\tilde{f}_n^i)^2 | b_1^i = 0\} = E\{(\tilde{f}_{n,1}^i)^2\} \tag{5}$$

$$E\{\tilde{f}_n^i | b_1^i = 1, b_2^i = 0\} = E\{\tilde{f}_{n,2}^i\} \tag{6}$$

$$E\{(\tilde{f}_n^i)^2 | b_1^i = 1, b_2^i = 0\} = E\{(\tilde{f}_{n,2}^i)^2\} \tag{7}$$

and using simple frame copy concealment ($\tilde{f}_n^i = \tilde{f}_{n-1}^i$)

$$E\{\tilde{f}_n^i | b_1^i = b_2^i = 1\} = E\{\tilde{f}_{n-1}^i\} \tag{8}$$

$$E\{(\tilde{f}_n^i)^2 | b_1^i = b_2^i = 1\} = E\{(\tilde{f}_{n-1}^i)^2\} \tag{9}$$

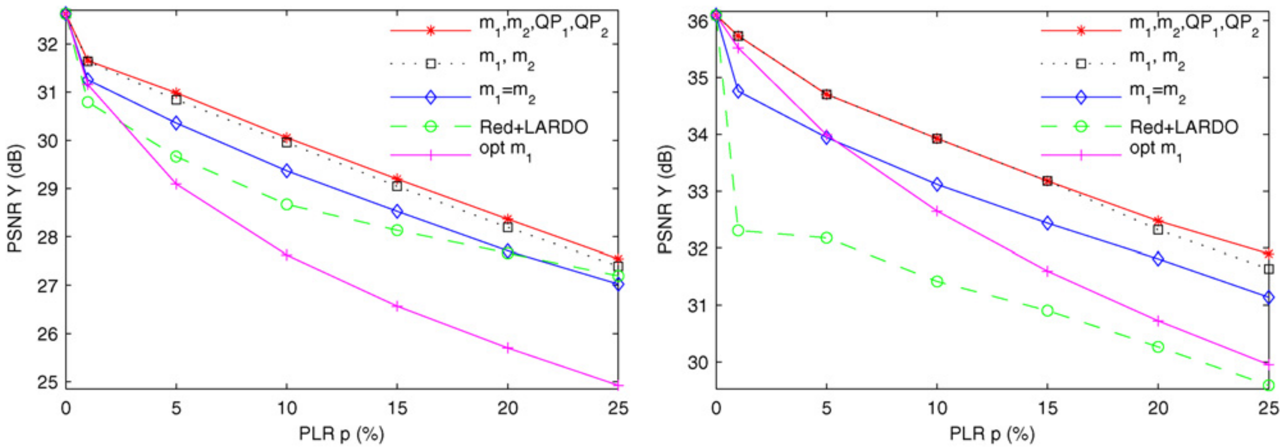
we rewrite (2), (3) as

$$\begin{aligned} E\{\tilde{f}_n^i\} = \quad & (1-p)E\{\tilde{f}_{n,1}^i\} \\ & + p(1-p)E\{\tilde{f}_{n,2}^i\} + p^2 E\{\tilde{f}_{n-1}^i\} \end{aligned} \tag{10}$$

$$\begin{aligned} E\{(\tilde{f}_n^i)^2\} = \quad & (1-p)E\{(\tilde{f}_{n,1}^i)^2\} \\ & + p(1-p)E\{(\tilde{f}_{n,2}^i)^2\} + p^2 E\{(\tilde{f}_{n-1}^i)^2\}. \end{aligned} \tag{11}$$

We need to compute the moments for successful reconstruction of the primary and secondary data. The recursion step is identical for both descriptions. Hence, $r = 1, 2$ denotes the description index in the equations below. For intra-coding with

[1] Note that ROPE is extensible to other channel models.




(a) (b)

Fig. 1. Delivery performance, PSNR Y versus PLR $p$ (CIF, 15 frames/s). (a) *Foreman*, 300 kb/s, $p = 0\ldots25\%$. (b) *News*, 200 kb/s, $p = 0\ldots25\%$.

constrained prediction, the decoder reconstruction is directly given by the transmitted value

$$E\{\tilde{f}^i_{n,r}\}(I) = \hat{f}^i_{n,r} \tag{12}$$

$$E\{(\tilde{f}^i_{n,r})^2\}(I) = (\hat{f}^i_{n,r})^2. \tag{13}$$

In inter-coding, pixel $i$ is predicted from pixel $j = i + mv$ in the previous frame $n-1$, i.e., the encoder prediction is $\hat{f}^j_{n-1}$.[2] The prediction error $e^i_{n,r}$ is quantized to the value $\hat{e}^i_{n,r}$, which is transmitted together with the motion vector $mv$. Even if the current packet is received correctly, the decoder uses the *decoder* reconstruction of pixel $j$ in the previous frame, $\tilde{f}^j_{n-1}$, for prediction, which is potentially different from the value used by the encoder $\tilde{f}^i_{n,r} = \tilde{f}^j_{n-1} + \hat{e}^i_{n,r}$. The first and second moments of $\tilde{f}^i_n$ for an inter-coded pixel are

$$E\{\tilde{f}^i_{n,r}\}(P) = E\{\tilde{f}^j_{n-1}\} + \hat{e}^i_{n,r} \tag{14}$$

$$\begin{aligned} E\{(\tilde{f}^i_{n,r})^2\}(P) &= E\left\{(\tilde{f}^j_{n-1} + \hat{e}^i_{n,r})^2\right\} \\ &= E\{(\tilde{f}^j_{n-1})^2\} + 2\hat{e}^i_{n,r}E\{\tilde{f}^j_{n-1}\} + (\hat{e}^i_{n,r})^2. \end{aligned} \tag{15}$$

The expected end-to-end distortion (per MB) is

$$E\{D_{MB}\} = \sum_{i \in MB} \left((f^i_n)^2 - 2f^i_n E\{\tilde{f}^i_n\} + E\{(\tilde{f}^i_n)^2\}\right). \tag{16}$$

*C. RD Optimization*

Given (10)–(16), error-resilient encoding can be posed as a well-known RD optimization problem

$$\min_k J^{(k)}_{MB} = \min_k \left(E\{D^{(k)}_{MB}\} + \lambda R^{(k)}_{MB}\right) \tag{17}$$

where $k$ indexes the encoding modes, $\lambda$ is the Lagrange factor, and $R^{(k)}_{MB} = R^{(k)}_1 + R^{(k)}_2$ is the rate for mode $k$.

For non-redundant encoding using ROPE [1], mode decision optimizes the main MB parameters: $\{mode, QP\}$. Other parameters have been included in the optimization, e.g., MVs and reference frame(s) [3], [4], but their impact is additive and not central to the problem of optimal redundant encoding considered here. For redundant encoding with a primary and secondary transmission, the decision space encompasses the parameters of *both* coded MBs: $\{mode_1, mode_2, QP_1, QP_2\}$.

Equations (10)–(16) illustrate that the primary MB has a larger weight toward end-to-end distortion than the redundant MB. The redundant MB should be coded at a lower rate $R_2 < R_1$, i.e., $QP_2 > QP_1$. A smaller primary $QP_1$ allows for superior decoded video quality, but leaves a smaller rate budget for the secondary encoding. In case of primary loss, the secondary reconstruction will thus be of poorer quality, and introduce more drift. It is not clear how to best distribute the rate $R_{MB}$ between the two encodings, i.e., what choices are optimal for $QP_1$ and $QP_2$. A larger QP results in lower rate from a smaller coded residual, while the rate for the side information is (mostly) unchanged. Therefore, the primary MB $mode_1$ may not be the optimal choice for $mode_2$. Recall that both MB encodings impact end-to-end performance. Optimal performance can only be obtained by jointly considering all parameter combinations for the MB pair.

1) Compute the rates $R^{(k)}_1$ and values $\tilde{f}^{i,(k)}_{n,1}$, $(\tilde{f}^{i,(k)}_{n,1})^2$ for successful reconstruction of all available combinations of $\{mode_1, QP_1\}^{(k)}$ for the *primary* MB encoding (indexed by $k$):
   a) P-SKIP;
   b) inter modes with different values $QP^{(k)}_1$;
   c) intra modes with different $QP^{(k)}_1$.
2) Repeat step 1 for the *redundant* MB transmission: determine rates $R^{(l)}_2$ and values $\tilde{f}^{i,(l)}_{n,2}$, $(\tilde{f}^{i,(l)}_{n,2})^2$ for successful reconstruction of all available combinations of $\{mode_2, QP_2\}^{(l)}$ for the secondary MB (indexed by $l$):
   a) P-SKIP (due to different secondary MB decisions, the secondary P-SKIP may be different from the primary);
   b) inter modes with different $QP^{(l)}_2$ (recycle primary MVs, but requantize residual with $QP^{(l)}_2$);
   c) intra modes with different $QP^{(l)}_2$.
3) For each combination $(k, l)$ of $\{mode_1, QP_1\}^{(k)}$ and $\{mode_2, QP_2\}^{(l)}$, determine the combined rate $R^{(k,l)}_{MB}$ and end-to-end distortion $E\{D_{MB}\}^{(k,l)}$.

[2]ROPE has been successfully extended to subpixel motion, whose estimation involves cross-correlation terms due to interpolation (e.g., [2]). This issue is orthogonal to the estimation of redundantly coded data considered here.

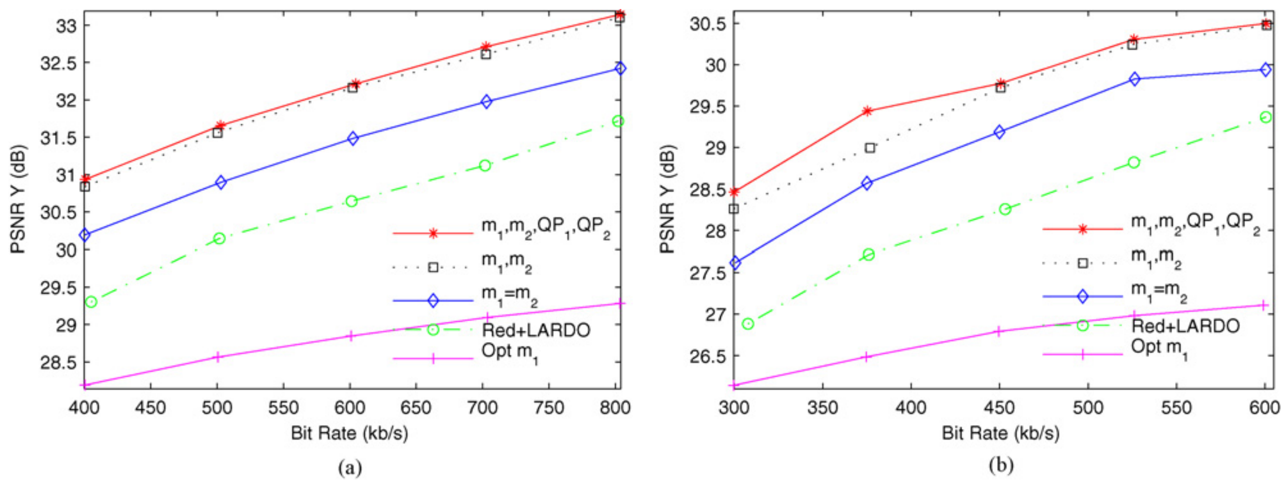



Fig. 2. Delivery performance, PSNR Y versus rate (CIF, 15 frames/s). (a) *Foreman*, 400–800 kb/s, $p = 10\%$. (b) *Table Tennis*, 300–600 kb/s, $p = 20\%$.

4) For $\lambda$ given by the rate control, pick the parameter combination $(k, l)$ that achieves the best Lagrangian

$$\min_{(k,l)} J^{(k,l)} = \min_{(k,l)} \left( E\{D_{MB}\}^{(k,l)} + \lambda R_{MB}^{(k,l)} \right). \tag{18}$$

We refer to this method as full-joint optimization. The number of parameter combinations considered is large; reasonable complexity can be achieved by pruning the decision space.

### D. Reduced-Complexity Approaches

In operational RDO, there is a close relationship between the Lagrangian factor $\lambda$ and the QP, e.g., H.264 uses [19]

$$\lambda = 0.85 * 2^{(QP-12)/3}. \tag{19}$$

Thus, we separate the MB QP consideration, and focus on the optimal combination of primary and secondary coding mode $\{mode_1, mode_2\}$. $QP_1$ is selected by the rate control, and we use a fixed offset $dQP$ for $QP_2 = QP_1 + dQP$. We will evaluate the impact of the choice of $dQP$ later.

Combined mode selection significantly reduces encoding complexity, but it still considers $N^2$ combinations for $N$ available modes. A simplified encoding algorithm could indeed use the same mode for the primary and secondary MB as in [10], [11], and [15]: $\{mode_1 = mode_2\}$. Recall that the key contribution of this letter is accurate end-to-end distortion estimation of the MB pair. Optimization over a limited parameter space may not realize maximum gains, but better distortion estimation should still improve performance over other approaches.

## III. Experimental Results

### A. Simulation Details

We implemented the proposed algorithms for redundant encoding on top of the JM 13.2 reference software [8]. For comparison, we implemented a redundant encoding algorithm similar to the current state of the art [10], [11]. Every frame is encoded redundantly, and the redundant transmission uses the same MB modes and motion information as the primary picture, but a higher secondary QP. Mode decision for the primary MB is based on loss-aware RD optimization (LARDO) [20] with $K = 100$ simulated decoders.[3] Unlike in [10] and [11], we select a fixed $dQP$ offset for the secondary transmission (constant throughout the entire sequence); the displayed results use the best $dQP$ for the respective sequence.

In the results, full-joint optimization, combined primary and secondary mode selection and our low complexity encoding algorithm are labeled as "$m_1, m_2, QP_1, QP_2$", "$m_1, m_2$", "$m_1 = m_2$", respectively. The redundant encoding scheme with LARDO is labeled "*Red+LARDO*". We also provide results of conventional non-redundant MB mode selection using ROPE ("*Opt* $m_1$"). All sequences were encoded at 15 frames/s, CIF resolution, each packet contains one slice and is $\leq 512$ bytes. The bitstreams were then simulated at different PLRs, and results averaged over 500 loss patterns at each PLR.

### B. Performance Versus PLR p

The first experiment compares the algorithms across a PLR range of $p = 1, \ldots 25\%$. For the *Foreman* sequence [300 kb/s, 150 frames, Fig. 2(a)], full-joint optimization outperforms all other methods: 0.9–1.1 dB over basic redundant encoding with LARDO at low to medium PLR, and 0.4 dB at high PLR. Optimal non-redundant encoding is close at $p = 1\%$, but then falls behind the redundant encoding schemes. Combined mode selection (without QP consideration) only results in a slight PSNR drop of 0.2 dB relative to full-joint optimization. Low-complexity encoding trails by another 0.4–0.6 dB. Redundant encoding uses 6%–16% ($p = 1, \ldots, 25\%$) of the total rate.

For the sequence news [200 kb/s, 150 frames, Fig. 1(b)], full-joint optimization and combined mode selection perform similarly at low and medium PLR, with a 0.2–0.3 dB gap at high PLR. Low-complexity redundant encoding trails by 1 dB at low PLR due to poor adaptivity; the gap narrows to 0.5 dB at high PLR. The rate for the secondary description varies between 5% and 18% of the total. Note the bad performance of the general redundant encoding scheme. This

[3]LARDO with $K = 100$ has significantly higher complexity than ROPE.

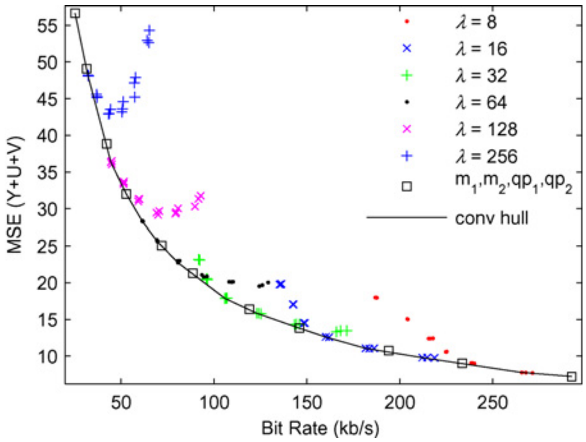



Fig. 3. Operational RD points and lower convex hull: MSE (per pixel) versus effective bit rate (*Akiyo*, $p$ = 10%, $\lambda$ = 8, 16, ... 256).

is partially caused by poor adaptivity (redundancy allocation): our naive implementation redundantly encodes every frame, when only some frames need full protection, most are fine protecting an ROI, and some may need none. However, part of the performance degradation is due to suboptimal distortion estimation.

### C. Performance Versus Bit Rate

We also compared performance across a range of target bit rates, with the PLR fixed. For *Foreman* [400–800 kb/s, PLR $p = 10\%$, Fig. 2(a)], the proposed algorithms outperform current state-of-the-art redundant coding. Full-joint optimization consistently gains 1.4–1.6 dB over LARDO-based redundant coding. Equivalently, it can achieve a 30–40% bit rate reduction. Joint-mode optimization still enables most of the gains: performance only drops by 0.2 dB, or an equivalent rate increase of about 4–6%. Performance under our low complexity algorithm degrades by 0.6 dB, or incurs a rate increase of 17–25%. Similar results are shown in Fig. 2(b) (*Table Tennis*, 150 frames, 300–600 kb/s, $p$ = 20%).

### D. Complexity Versus Performance Tradeoff, Impact of $dQP$

Our last experiment evaluates pure operational RD performance. With rate control disabled, we varied $\lambda = 8, \ldots, 256$, and generated one encoding for each $\lambda$ using full-joint optimization. Additional encodings were performed for each triplet $\{\lambda, QP_1, dQP\}$ ($QP_1 = 24, 26, \ldots, 40$, $dQP = 6, 8, 10$), using combined mode selection. Fig. 3 plots all RD points for the *Akiyo* sequence ($p$ = 10%, MSE per pixel versus effective bit rate), and the lower convex hull.

Full-joint optimization achieves points on the lower convex hull for all $\lambda$ (intermediate values added). For the other points, each set corresponds to the RD performance of combined mode selection under one particular value of $\lambda$. For each value of $\lambda$, there are several points on, or close to, the lower convex hull. We conclude that combined mode selection is able to achieve performance close to full-joint optimization, and that the difference is primarily due to rate control issues and coarser-grain adaptation. Further, note that RD points with the same ($\lambda$, $QP$) pair form clusters. The choice of the secondary QP offset $dQP$ has minor impact, but appears non-critical.

## IV. Conclusion and Future Work

We presented a new error-resilient encoding framework that enables accurate end-to-end distortion estimation for redundant encoding, jointly considering the contributions from the primary and secondary transmissions. Based on this framework, we proposed several algorithms to select the primary and secondary MB parameters (modes and QPs). Results show significant RD performance gains over current redundant coding schemes and conventional ROPE-based optimal MB coding mode selection. The proposed scheme can be further enhanced beyond the results presented here: The algorithm could employ principles of multiple descriptions (e.g., prediction from different reference pictures) and enable improved reconstruction when both descriptions are received. FMO [7] could enable improved concealment within our scheme: the gains from optimal redundant encoding and improved concealment should be additive. Finally, note that while redundant encoding is inefficient compared to true scalable representation, it is more error-resilient than SVC, where the loss of the base layer renders the bit stream useless.

## References


[1] R. Zhang, S. L. Regunathan, and K. Rose, “Video coding with optimal inter/intra-mode switching for packet loss resilience,” *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.

[2] H. Yang and K. Rose, “Advances in recursive per-pixel end-to-end distortion estimation for robust video coding in H.264/AVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 7, pp. 845–856, Jul. 2007.

[3] S. Wan and E. Izquierdo, “Rate-distortion optimized motion-compensated prediction for packet loss resilient video coding,” *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1327–1338, May 2007.

[4] A. Leontaris and P. C. Cosman, “Video compression for lossy packet networks with mode switching and a dual-frame buffer,” *IEEE Trans. Image Process.*, vol. 13, no. 7, pp. 885–897, Jul. 2004.

[5] B. A. Heng, J. G. Apostolopoulos, and J. S. Lim, “End-to-end rate-distortion optimized MD mode selection for multiple description video coding,” *EURASIP J. Appl. Signal Process.*, vol. 2006, no. 32592, p. 12, 2006.

[6] E. Masala, H. Yang, K. Rose, and J. C. De Martin, “Rate-distortion optimized slicing, packetization and coding for error-resilient video transmission,” in *Proc. DCC*, 2004, pp. 182–191.

[7] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Rec. H.264, ISO/IEC 14496-10 AVC, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Mar. 2005.

[8] *H.264/AVC Software Coordination* [Online]. Available: http://iphome.hhi.de/suehring/tml/

[9] Z. Wu and J. M. Boyce, “Adaptive error resilient video coding based on redundant slices of H.264/AVC,” in *Proc. IEEE ICME*, Jul. 2007, pp. 2138–2141.

[10] P. Baccichet, S. Rane, A. Chimienti, and B. Girod, “Robust low-delay video transmission using H.264/AVC redundant slices and flexible macroblock ordering,” in *Proc. IEEE ICIP*, vol. 4. Sep.–Oct. 2007, pp. 93–96.

[11] T. Tillo, M. Grangetto, and G. Olmo, “Redundant slice optimal allocation for H.264 multiple description coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 59–70, Jan. 2008.

[12] B. Katz, S. Greenberg, N. Yarkoni, N. Blaunstien, and R. Giladi, “New error-resilient scheme based on FMO and dynamic redundant slices allocation for wireless video transmission,” *IEEE Trans. Broadcasting*, vol. 53, no. 1, pp. 308–319, Mar. 2007.

[13] L. Yang and H. Xiong, “Redundant multiscale structure coding for error resilient video completion,” in *Proc. IEEE ISCAS*, May 2008, pp. 3578–3581.

[14] M. B. Dissanayake, C. T. E. Hewage, S. T. Worrall, W. A. C. Fernando, and A. M. Kondoz, "Redundant motion vectors for improved error resilience in H.264/AVC coded video," in *Proc. IEEE ICME*, Apr.–Jun. 2008, pp. 25–28.

[15] S. Rane, P. Baccichet, and B. Girod, "Systematic lossy error protection of video signals," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 10, pp. 1347–1360, Oct. 2008.

[16] C. Zhu, Y. Wang, M. M. Hannuksela, and H. Li, "Error resilient video coding using redundant pictures," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 1, pp. 3–14, Jan. 2009.

[17] J. C. Schmidt and K. Rose, "Macroblock-based retransmission for error-resilient video streaming," in *Proc. IEEE ICIP*, vol. 5. Oct. 2008, pp. 2308–2311.

[18] J. C. Schmidt and K. Rose, "Jointly optimized mode decisions in redundant video streaming," in *Proc. IEEE ICASSP*, Apr. 2009, pp. 797–800.

[19] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, Jul. 2003.

[20] T. Stockhammer, T. Wiegand, and S. Wenger, "Optimized transmission of H.26L/JVT coded video over packet-lossy networks," in *Proc. IEEE ICIP*, vol. 2. 2002, pp. 173–176.