Power-Rate-Distortion Analysis for Wireless Video Communication under Energy Constraints

Zhihai He, *Member, IEEE,* Yongfang Liang, *Student Member, IEEE,* Lulin Chen, *Senior Member, IEEE*
Ishfaq Ahmad, *Senior Member, IEEE* and Dapeng Wu *Member, IEEE*

Abstract— **Mobile devices performing video coding and streaming over wireless and pervasive communication networks are limited in energy supply. To prolong the operational lifetime of these devices, an embedded video encoding system should be able to adjust its computational complexity and energy consumption as demanded by the situation and its environment. To analyze, control, and optimize the rate-distortion (R-D) behavior of the wireless video communication system under the energy constraint, we develop a power-rate-distortion (P-R-D) analysis framework, which extends the traditional R-D analysis by including another dimension, the power consumption. Specifically, in this paper, we analyze the encoding mechanism of typical video coding systems, and develop a parametric video encoding architecture which is fully scalable in computational complexity. Using dynamic voltage scaling (DVS), an energy consumption management technology recently developed in CMOS circuits design, the complexity scalability can be translated into the energy consumption scalability of the video encoder. We investigate the R-D behavior of the complexity control parameters and establish an analytic P-R-D model. Both theoretically and experimentally, we show that, using this P-R-D model, the video coding system is able to automatically adjust its complexity control parameters to match the available energy supply of the mobile device while maximizing the picture quality. The P-R-D model provides a theoretical guideline for system design and performance optimization in mobile video communication under energy constraints.**

Index Terms— **Energy consumption, rate-distortion analysis, wireless video, complexity scalability.**

I. INTRODUCTION

VIDEO encoding and streaming over wireless communication networks is envisioned for a wide range of applications, such as battlefield intelligence, surveillance, reconnaissance, security monitoring, emergency response, disaster rescue, environmental tracking, tele-medicine, and multimedia systems in consumer electronics [4]. In wireless video communication, video capture, compression and network streaming operate on the mobile devices with limited energy. A primary factor in determining the utility or operational lifetime of the mobile communication device is how efficiently it manages its

energy consumption. The problem becomes even more critical with the power-demanding video encoding functionality integrated into the mobile computing platform [1].

A. The Research Problem

Video encoding and data transmission are the two dominant power-consuming operations in wireless video communication, especially over wireless LAN, where the typical transmission distance ranges from 50m to 100m. Experimental studies show that for relative small picture sizes, such as QCIF (176×144) videos, video encoding consumes about $\frac{2}{3}$ of the total power for video communication over Wireless LAN [1], [19]. For pictures of higher resolutions, it is expected that the fraction of power consumption by video encoding will become even higher. From the power consumption perspective, the effect of video encoding is two-fold. First, efficient video compression significantly reduces the amount of the video data to be transmitted, which in turn saves a significant amount of energy in data transmission. Second, more efficient video compression often requires higher computational complexity and larger power consumption in computing. These two conflicting effects imply that in practical system design there is always a tradeoff among the bandwidth R, power consumption P, and video quality D. Here, the video quality is often measured by the mean square error (MSE) between the encoded picture and original one, also known as the source coding distortion. To find the best trade-off solution, we need to develop an analytic framework to model the power-rate-distortion (P-R-D) behavior of the video encoding system. To achieve flexible management of power consumption, we also need to develop a video encoding architecture which is fully scalable in power consumption.

B. Related Work

Many algorithms have been reported in the literature to reduce the encoding computational complexity. A statistical modeling approach is proposed in [21] to predict the zero DCT coefficients after quantization. Based on the prediction, the DCT computation for those zeros coefficients can be saved. Fast and low-power motion estimation algorithms have been developed to reduce the computational complexity of the motion estimation module [3], [14]. Since there is no motion estimation for INTRA macroblocks (MB's), the INTRA ratio parameter, which is the fraction of INTRA MB's in the video frame, can be used to control the motion estimation complexity in the video encoder [19]. A parametric scheme

for scalable motion estimation and DCT has been proposed in [8]. Hardware implementation technologies have also been developed to improve the video coding speed [14], [25].

To our best knowledge, there has been no analytic framework for modeling the P-R-D behavior of the video encoding system. Rate-distortion (R-D) analysis has been one of the major research focus in information theory and communication for the past few decades, from the early Shannon's source coding theorem for asymptotic R-D analysis of generic information data [6], to recent R-D modeling of modern video encoding systems [10], [12], [13], [22]. For video encoding on the mobile devices and streaming over the wireless network, it is needed to consider another dimension, the power consumption, to establish a theoretical basis for R-D analysis under energy constraints. In energy-aware video encoding, the coding distortion is not only a function of the encoding bit rate as in the traditional R-D analysis, but also a function of the power consumption P. In other words,

$$D = D(R, P), \tag{1}$$

which describes the P-R-D behavior of the video encoding system. The P-R-D model provides a theoretical basis, as well as a practical guideline, for system design and performance optimization in wireless communication. Using the P-R-D model, we can perform energy consumption control on each mobile device. At the system level, for example in a wireless sensor network, we can perform across-node energy optimization and network lifetime maximization.

C. The Proposed Research

In this work, we develop an analytic framework to model, control and optimize the P-R-D behavior of typical video encoding systems. This is accomplished by two major steps. First, we develop a video encoding architecture which is fully scalable in power consumption. Specifically, we introduce several control parameters into the video encoder to control the power consumption of the major encoding modules. Second, we analyze the R-D behaviors of these control parameters. The integration of the R-D models for the control parameters results in a comprehensive P-R-D model for the video coding system. Based on the P-R-D model, we develop a quality optimization scheme to determine the best configuration of complexity control parameters according to the power supply level of the mobile device to maximize the video presentation quality.

D. Paper Organization

The rest of the paper is organized as follows. In Section II, we analyze the encoding complexity of a typical video encoder, and investigate the fundamental approach to design a complexity-scalable video encoding system. To translate the complexity scalability into energy scalability, we present the dynamic voltage scaling (DVS), a recently developed power management technology in CMOS circuits design. In Section III, we present a complexity scalable motion estimation (ME) scheme and study the R-D behavior of the ME complexity control parameter. In Section IV, we present a complexity

scalable scheme which is able to collectively control the power consumption of the remaining modules in the video encoder. The R-D behavior of the complexity control parameter is also analyzed. An integrated P-R-D model is presented in Section V. The quality optimization and complexity control parameters configuration are also discussed in Section V. The power-scalable video encoding scheme is summarized in Section VI. Section VII presents the experimental results. Some concluding remarks are given in Section VIII.

II. ENCODER COMPLEXITY ANALYSIS AND POWER CONSUMPTION

In this section, we analyze the computational complexity of the major encoding modules in a typical video encoding system. Based on the complexity profile, we outline a complexity scalable architecture for video encoding. We then consider the DVS CMOS design technology and discuss its application in energy scalable encoding system design.

A. Encoder Complexity Profile

Typical video encoders, including all the standard video encoding systems, such as MPEG-2 [16], H.263 [17], and MPEG-4 [23], employ a hybrid motion compensated DCT encoding scheme. Specifically, as shown in Fig. 1, they have the following major encoding modules: motion estimation (ME) and compensation (COMP), DCT, quantization (QUANT), entropy encoding (ENC) of the quantized DCT coefficients, inverse quantization (DQUANT), inverse DCT (IDCT), picture reconstruction (RECON), and interpolation (INTERP) [23]. For the ease of exposition, the DCT, IDCT, QUANT, DQUANT and RECON modules are collectively referred to as PRECODING. In this way, the video encoder has only three major modules: ME, PRECODING, and ENC. The PRECODING can be considered as the data representation module.
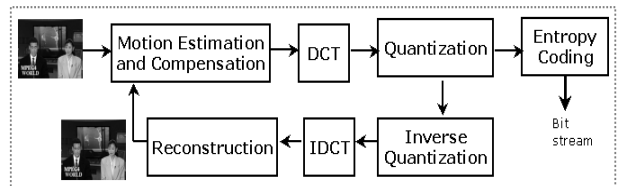

Fig. 1.　Block diagram of a typical video encoder. For intra MB or frames, motion estimation and compensation are not needed.

To analyze the run-time complexity of the major encoding modules, we run the MPEG-4 video encoder on an 866 MHz Pentium III PC and profile its computational complexity, measured as the average processor cycles. The test video sequences are "Akiyo", "News", and "Carphone" in QCIF format encoded at 15 fps and 64 kbps. In Table I, we list the percentage CPU occupancy for the major encoding modules. (We have also evaluated the encoder CPU occupancy with other video sequences and different frame rate and bit rate settings. Only a slight difference from the results in Table I has been observed.) It can be seen that ME is the most computation-intensive module, consuming about one-third of

Component	Akiyo	News	Carphone
ME	30.4%	32.6%	33.1%
COMP	9.1%	8.4%	8.7%
DCT	10.5%	9.2%	9.2%
QUANT	4.9%	4.6%	5.1%
ENC	4.7%	5.4%	4.5%
DQUANT	1.9%	1.5%	2.0%
IDCT	2.3%	2.9%	2.6%
RECON	7.5%	6.9%	7.2%
INTERP	14.3%	12.8%	13.2%
RC	7.4%	7.9%	7.6%
Other	6.5%	7.3%	6.7%

the processor cycles. The PRECODING modules collectively consume about 50% of the total processor cycles. The ENC module, which is basically a bit splicing engine, uses a relative small amount of the total CPU time, especially at low coding bit rates. In addition, its computational complexity mainly depends on the coding bit rate.

B. Complexity Scalable Encoder Design

As discussed in Section I-C, to design a video encoder which is fully scalable in power consumption, we need to introduce several encoder parameters to control the computational complexity of the major encoding modules. Specifically, in this work, the complexity control parameter for the ME module is the number of SAD (sum of absolute difference) computations per frame, denoted by λ_{ME}. This is based on the observation that the ME process is simply a sequence of SAD computations to find the MB position of the minimum SAD. Therefore, the computational complexity of ME, denoted by $\mathcal{C}_{ME}$, is simply given by

$$\mathcal{C}_{ME} = \lambda_{ME} \cdot \mathcal{C}_{SAD}, \tag{2}$$

where $\mathcal{C}_{SAD}$ represents the complexity of one SAD computation between the current MB and its reference MB. Here, the computational complexity is measured by the number of processor cycles used by the operation. A detailed description of the parametric ME design, optimal resource allocation of the SAD computations, and R-D analysis of the λ_{ME} complexity parameter will be presented in Section III. By analyzing the encoding architecture of the video encoding system, we find that it is possible to control the computational complexity of all the PRECODING modules using one single parameter λ_{PRE}, which is the number of non-zero MB's in the video frame. Here, "non-zero" means the MB has non-zero DCT coefficients after quantization. Let $\mathcal{C}_{NZMB}$ and $\mathcal{C}_{PRE}$ be the PRECODING computational complexity of one non-zero MB (NZMB) and the whole video frame, respectively. From Section IV, we will see that,

$$\mathcal{C}_{PRE} = \lambda_{PRE} \cdot \mathcal{C}_{NZMB}. \tag{3}$$

A detailed description of the parametric PRECODING design, dynamic rate control, and R-D analysis of the complexity

control parameter λ_{PRE} will be presented in Section IV. The ENC module, as a variable length coding (VLC) engine, mainly consists of VLC table look-up and bit splicing of the codewords. The computational complexity of the ENC module, denoted by $\mathcal{C}_{ENC}$, is approximately proportional to R. Therefore, we have

$$\mathcal{C}_{ENC} = S \cdot R \cdot \mathcal{C}_{BIT}, \tag{4}$$

where $\mathcal{C}_{BIT}$ is the per bit ENC complexity, and S is the size of the picture. Here, S is needed because R represents the coding bit rate in the unit of bits per pixel. The computational complexity of the video encoder $\mathcal{C}$, measured by the number of processor cycles per second, is given by

$$\begin{aligned} &\mathcal{C}(R; \lambda_{ME}, \lambda_{PRE}, \lambda_F) \\ &= \lambda_F \cdot (\lambda_{ME}\mathcal{C}_{SAD} + \lambda_{PRE}\mathcal{C}_{NZMB} + S \cdot R \cdot \mathcal{C}_{BIT}), \end{aligned}$$

where λ_F is the encoding frame rate. This model presents a complexity-scalable architecture for video encoding, whose computational complexity is mainly controlled by the parameter set $\{\lambda_{ME}, \lambda_{PRE}, \lambda_F\}$. It can be seen that, in the proposed complexity scalable video coding design, we try to find the "atom operations" that have fixed computational complexity, and decompose the overall video encoding into these atom operations. Specifically in this work the atom operations are the MB SAD computation, the PRECODING of one MB, and the per-bit ENC operation.

C. Dynamic Voltage Scaling and Encoder Energy Consumption

In the previous section, we have outlined a parametric video encoding architecture which is fully scalable in computational complexity. To translate the complexity scalability into energy scalability, we need to consider the energy-scaling technologies in hardware design. To dynamically control the energy consumption of the microprocessor on the portable device, a CMOS circuits design technology, named *dynamic voltage scaling (DVS)*, has been recently developed [18], [20]. In CMOS circuits, the power consumption P is given by

$$P = V^2 \cdot f_{CLK} \cdot C_{EFF}, \tag{5}$$

where V, f_{CLK}, and C_{EFF} are the supply voltage, clock frequency, and effective switched capacitance of the circuits [7]. Since the energy is power times time, and the time to finish an operation is inversely proportional to the clock frequency. Therefore, the energy per operation E_{op} is proportional to V^2 ($E_{op} \propto V^2$). This implies that lowering the supply voltage will reduce the energy consumption of the system in a quadratic fashion. However, lowering the supply voltage also decreases the maximum achievable clock speed. More specifically, it has been observed that f_{CLK} is approximately linearly proportional to V [7]. Therefore, we have

$$P \propto f_{CLK}^3, \quad \text{and} \quad E_{op} \propto f_{CLK}^2. \tag{6}$$

It can be seen that the CPU can reduce its energy consumption substantially by running more slowly. For example, according to (6), it can run at half speed and thereby use only $\frac{1}{4}$ of the energy for the same number of operations. This is the key idea

behind the DVS technology. Variable chip makers, including AMD [2] and Intel [15], have recently announced and sold processors with this energy-scaling feature. In conventional system design with fixed supply voltage and clock frequency, clock cycles, and hence energy, are wasted when the CPU workload is light and the processor becomes idle. Reducing the supply voltage in conjunction with the clock frequency eliminates the idle cycles and saves the energy significantly. It should be noted that in practice the energy saving is less than the amount suggested by the model in (6). In this work, we just use this model to translate the computational complexity into the energy consumption of the hardware. Certainly, if available, more accurate DVS energy consumption model can be used to improve the energy management performance.

The DVS technology provides an enabling hardware technology for our energy-scalable video encoding system design. Using the parametric complexity scalability scheme outlined in Section II-B, we can flexibly control the number of processor cycles per second $\mathcal{C}$ of the video encoder by choosing appropriate complexity control parameters $\{\lambda_{ME}, \lambda_{PRE}, \lambda_F\}$. With DVS, we can adjust the supply voltage V such that the corresponding clock frequency f_{CLK} matches $\mathcal{C}$. According to Eqs. (5) and (6),

$$P = C_{EFF} \cdot [\mathcal{C}(R; \lambda_{ME}, \lambda_{PRE}, \lambda_F)]^3. \qquad (7)$$

In other words, for a given power supply level of the mobile device, we can determine the encoding complexity by

$$\mathcal{C}(R; \lambda_{ME}, \lambda_{PRE}, \lambda_F) = \Phi(P), \quad \Phi(P) = (\frac{P}{C_{EFF}})^{\frac{1}{3}}. \quad (8)$$

It should be noted that if a different DVS model is used, the expression of $\Phi(.)$ should be changed accordingly. This power consumption model describes a parametric energy-scalable video encoding architecture whose energy consumption is controlled by the parameter set $\{\lambda_{ME}, \lambda_{PRE}, \lambda_F\}$. In the following sections, we will describe each energy-scalability parameter in detail and model its R-D behavior. The R-D models, along with the DVS power consumption model in (8), will be integrated together to establish a comprehensive P-R-D analysis framework.

III. COMPLEXITY-SCALABLE MOTION ESTIMATION AND R-D ANALYSIS

In this section, we analyze the computational complexity of the ME module, and propose a complexity scalability scheme to control the computational complexity of the ME module using the parameter λ_{ME}, which is the number of SAD computations per frame. We present an adaptive method to allocate the SAD computations among the MB's to optimize the picture quality. The R-D behavior of the complexity control parameter λ_{ME} is also analyzed.

A. Complexity Scalable Motion Estimation Design

In block-based video coding, the objective of motion estimation is to find the best match in the reference frame for every MB in the current frame. The search for the SAD-optimal motion vector problem can be formulated as

$$(u_0, v_0) = \arg \min SAD(u, v) \qquad (9)$$

where $SAD(u, v)$ represents the sum of absolute difference (SAD) between the current MB and the reference MB at a relative position of (u, v). We can see that the ME process is simply a sequence of SAD computations to find the motion vector which has the minimum SAD. Note that the computational complexity of each MB SAD is a constant. Therefore, the overall computational complexity of the ME module is linearly proportional to the number of SAD computations λ_{ME}, as in (2). In the proposed energy scalable framework, λ_{ME} is determined by system-level power management and quality optimization. At the frame-level, the λ_{ME} SAD computations are allocated among the MB's in the video frame to optimize the picture quality.

B. Dynamic Allocation of the SAD Computations

It is well known that the moving objects in the video scene contribute most to the overall visual quality. This suggests that in motion estimation under energy constraints, we need to allocate the available λ_{ME} SAD computations among the MB's according to their motion characteristics to optimize the overall picture quality. Let (mv_x, mv_y) be the motion vector of the MB. The block motion activity (BMA) factor of the MB, denoted by ma is defined as

$$ma = |mv_x| + |mv_y|. \qquad (10)$$

At the frame level, we introduce a *motion history matrix* (MHM), denoted by $\mathcal{M} = [m_{ij}]_{MR \times MC}$, where MR and MC are the numbers of MB's per row and per column, respectively. Initially, we set $m_{ij} = 1$. After a frame is coded, each entry is updated as follows:

$$m_{ij} = \begin{cases} m_{ij} + 1, & \text{if } ma = 0; \\ 0, & \text{else.} \end{cases} \qquad (11)$$

Here, ma is the BMA factor of the (i, j)-th MB in the coded frame. The larger the value of m_{ij}, it is of higher probability that this MB is a static block, and less SAD computations can be allocated to this MB. Fig. 2 shows the MHM for the "Sean" sequence. Note that each entry of the MHM is linearly scaled and represented by the gray level of a MB, ranging from 0 to 255. We can see that the MHM captures not only the motion history but also the locations of the object motion. Most importantly, this MHM approach has very low computation overhead and is very cost-effective in practice.

Using the MHM, we can allocate the λ_{ME} SAD computations among the MB's. The number of SAD computations allocated to the (i, j)-th MB, denoted by $nsad_{ij}$, is determined by

$$nsad_{ij} = \frac{1}{N-1} \left[1 - \frac{m_{ij}}{\displaystyle\sum_{(k,l) \geq (i,j)} m_{kl}} \right] \cdot Nsad, \qquad (12)$$

where N is the number of MB's left so far that need to perform the motion estimation, and $Nsad$ is the available number of SAD computations. Here, $N - 1$ is a normalization factor, because

$$\sum_{(i,j)} \left[1 - \frac{m_{ij}}{\displaystyle\sum_{(k,l) \geq (i,j)} m_{kl}} \right] = N - 1. \qquad (13)$$

Initially, $Nsad$ is set to be λ_{ME}. Suppose the motion search range is SR. If $nsad_{ij} \geq (2 \cdot SR + 1)^2$, it means the computational power is enough to perform a full search for this block. Otherwise, the diamond motion search algorithm in [24] is used to find the motion vector, whose complexity, indicated by the number of search layers, is controlled by $nsad_{ij}$.
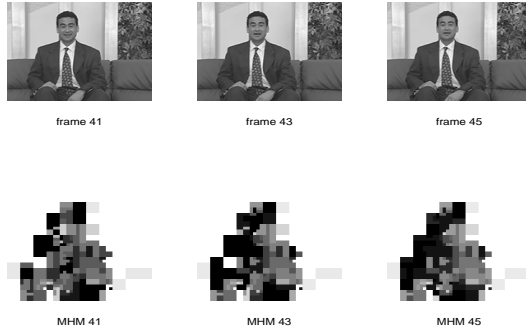

Fig. 2. MHMs of the "Sean" sequence.

C. Modeling the R-D Behavior of λ_{ME}

To analyze the R-D behavior of the complexity control parameter λ_{ME}, we need to investigate the relation between λ_{ME} and the frame SAD S_f, which is the average SAD per pixel in the motion compensated difference frame. To this end, we collect the frame SAD statistics for different λ_{ME} from several test video sequences. Fig. 3 plots the frame SAD S_f as a function of λ_{ME} for two QCIF video sequences: "Akiyo" and "Foreman". The simulation results suggest the following relation between λ_{ME} and S_f:

$$S_f(\lambda_{ME}) = \beta_0 + \beta_1 \cdot e^{-\beta_2 x}, \quad x = \frac{\lambda_{ME}}{\lambda_{ME}^{max}}, \qquad (14)$$

where the model parameters β_0, β_1, and β_2 are estimated by the statistics of previous frames; and λ_{ME}^{max} is the maximum value of λ_{ME}. Besides the SAD, another operation called SSD (sum of square difference), which is the square difference between the current MB and its reference, is often used in motion estimation. In hardware design, the SSD is more advantageous than the SAD because the subtraction and multiplication operations can be completed by a single instruction [9]. In motion estimation, SAD and SSD have similar performance because SSD linearly increases with the SAD. Therefore, the proposed complexity control is also applicable to the SSD-based ME. Simulation with SSD yields similar result as shown in Fig. 3, and the complexity model in (14) also applies to SSD. In this case, the frame SSD S_f becomes the variance of the difference frame. From Section V we will see that the final P-R-D model needs the variance information for R-D analysis. Therefore, hereafter, we assume SSD is used for ME.

IV. COMPLEXITY-SCALABLE PRECODING AND R-D ANALYSIS

In this section, we present a parametric complexity scalability scheme to collectively control the computational complexity of the PRECODING modules, namely, the DCT, QUANT,
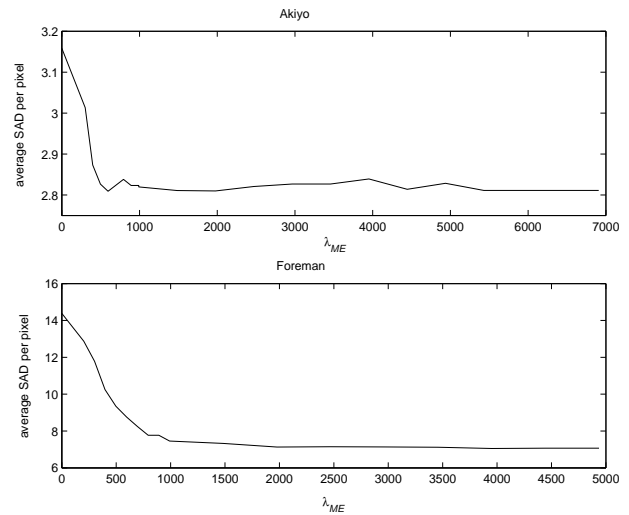

Fig. 3. Frame SAD as a function of λ_{ME}.

DQUANT, IDCT, and RECON modules. We then analyze the R-D behavior of the PRECODING complexity control parameter.

A. Complexity-Scalable PRECODING Design

In typical video encoding as illustrated in Fig. 1, DCT is applied to the difference MB after motion estimation and compensation, or the original MB if its coding mode is INTRA. After the DCT coefficients are quantized, DQUANT, IDCT, and RECON are performed to reconstruct the MB for motion prediction of the next frame. In transform coding of videos, especially at low coding bit rates, the DCT coefficients in the MB might become all zeros after quantization. We refer to this MB as an all-zero MB (AZMB). Otherwise, it is called a non-zero MB (NZMB). In international standards for video encoding, such as MPEG-2, H.263, and MPEG-4, "non-zeros" also means the CBP (coded block pattern) value of the MB is non-zero. If we can predict an MB to be AZMB, all the above PRECODING operations can be skipped, because the output of DQUANT and IDCT of an AZMB is still an AZMB, and the reconstructed MB is exactly the reference MB used in motion estimation and compensation. Therefore, the encoder can simply copy over the reference MB to reconstruct the current MB. This is a unique property of the AZMB, which can be used to reduce the computational complexity of the video encoder [11].

In this work, the unique property of the AZMB is used to design a complexity scalability scheme for the PRECODING modules. Let $\{x_{nk}|0 \leq n,k \leq 7\}$ be the coefficients in the different MB after motion estimation. For INTRA MB's, $\{x_{nk}\}$ are the original pixels in the video frame. Let $\{y_{ij}|0 \leq i,j \leq 7\}$ be the DCT coefficients. According to the definition of DCT, we have

$$y_{ij} = \frac{1}{4} C_i C_j \sum_{n=0}^{7} \sum_{k=0}^{7} x_{nk} \cos(i\pi \frac{2n+1}{16}) \cos(j\pi \frac{2k+1}{16}),$$

where

$$C_i = \left\{ \begin{array}{ll} \frac{1}{\sqrt{2}} & \text{if } \ i = 0, \\ 1 & \text{else} \end{array} \right. \qquad C_j = \left\{ \begin{array}{ll} \frac{1}{\sqrt{2}} & \text{if } \ j = 0, \\ 1 & \text{else.} \end{array} \right.$$

We can see that

$$|y_{ij}|^2 \leq \sum_{n=0}^{7} \sum_{k=0}^{7} |x_{nk}|^2. \qquad (15)$$

Note that the right-hand side is the SSD of the difference MB, which is already computed during the motion estimation. This suggests us that the SSD could be an efficient and low-cost measure to predict the AZMB. After motion estimation and compensation, let $\{SSD_i | 1 \leq i \leq M\}$ be the SSD values of the M MB's in the video frame sorted in an ascending order. In the proposed complexity scalability scheme for PRECODING, we force the first $M - \lambda_{PRE}$ MB's to be AZMB's, and treat the remaining λ_{PRE} MB's as NZMB's to which the PRECODING operations are applied. Let $\mathcal{C}_{NZMB}$ be the number of processor cycles needed by the PRECODING operations to finish one NZMB. The value of $\mathcal{C}_{NZMB}$ can be obtained either by theoretical cycle estimation of the PRECODING modules, or from simulation statistics. In practice, the value of $\mathcal{C}_{NZMB}$ may vary slightly from MB to MB. Note that the power management and energy consumption control operate on a level much higher than the MB. For example, in real-world applications, it is sufficient to adjust the system power control parameters for every 5 seconds, which have 150 frames (if coded 30 frame per second) and thousands of MB's. At this level, in its average sense, it is quite reasonable to assume $\mathcal{C}_{NZMB}$ is a constant. The overall complexity of the PRECODING modules, denoted by $\mathcal{C}_{PRE}$ is then given by

$$\mathcal{C}_{PRE} = \lambda_{PRE} \cdot \mathcal{C}_{NZMB}. \qquad (16)$$

We refer to this type of complexity scalability scheme as λ_{PRE} - scalability.

B. Dynamic Rate Control

In the proposed PRECODING complexity scalability scheme, the first $M - \lambda_{PRE}$ MB's are encoded as AZMB's to scale down the computational complexity of the PRECODING modules. Since the DCT coefficients in the AZMB's are all zeros, which do not need any encoding bits. All the available bit budget, denoted by R_T, will be allocated to the NZMB's. In this work, we adopt the linear rate control (LRC) algorithm developed in our previous work [13] to perform dynamic bit allocation and rate control. The LRC algorithm is based on a linear rate model. Specifically, we have found that in typical video encoding, including the standard MPEG-2, H.263, MPEG-4, and JVT coding, the coding bit rate R is a linear function of ρ, the fraction of zeros among the quantized transform coefficients. In other words,

$$R = \theta \cdot (1 - \rho), \qquad (17)$$

where θ is a constant. For a detailed treatment of the linear rate model and the LRC algorithm, see [13]. One unique feature of the LRC algorithm is that it always divides the picture into two groups: coded and uncoded MBs, and balances the bit budget between these two groups using the linear rate model. Such type of rate control mechanism allows a dynamic bit relocation from the AZMB's to the NZMB's, as well as a near-optimal bits allocation among the NZMB's.

As far as the subjective video quality is concerned, the proposed scalability and dynamic rate control scheme also performs reasonably well. As mentioned in Section III-B, the moving objects in the scene contribute most to the video presentation quality, and have unique significance in subjective video quality evaluation. In motion estimation and compensation, these regions of the picture often correspond to blocks with relatively large SSD values. In the proposed complexity scalability and dynamic rate control scheme, the saved AZMB bits are added to these blocks, resulting in an improved visual quality within these regions. Fig. 4 shows the 150-th frame of "Foreman" encoded at 192 kbps and 15 fps, and the 80-th frame of "Carphone" encoded at 64 kbps and 15 fps with 100% and 20% PRECODING complexity. It can be seen that low complexity PRECODING still maintains a perceptually acceptable picture quality. It should also be noted that the blocks with SSD below the threshold often correspond to picture regions with smooth spatial or temporal variation. The slightly degraded quality in these regions can be easily restored by post-processing techniques, such as deblocking, deringing, or temporal smoothing, at the receiver side.

Fig. 4. Coded video quality comparison for Frame 150 of "Foreman" and Frame 80 of "Carphone" when (A) 100% blocks are encoded; (B) 20% blocks are encoded.

C. R-D Behavior of The Complexity Control Parameter λ_{PRE}

The dynamic rate control is a near-optimal bit allocation process. Based on the mathematical formulation for optimal bit allocation, we analyze the R-D behavior of the complexity control parameter λ_{PRE}. Let $\{\sigma_i^2 | 1 \leq i \leq M\}$ be the variance of the MB's in the video frame sorted in an ascending order. Let R be the target coding bit rate in bits per pixel (bpp). According to the classic R-D distortion formula [6], the distortion of the i-th MB is given by

$$D_i(R_i) = \sigma_i^2 \cdot 2^{-2\gamma R_i}, \qquad (18)$$

where R_i is the bit rate of the i-th MB, and γ is a model constant. The optimal bit allocation can be then formulated as

$$D = \min_{\{R_i\}} \frac{1}{M} \sum_{i=1}^{M} \sigma_i^2 \cdot 2^{-2\gamma R_i}, \qquad (19)$$

$$s.t. \qquad \frac{1}{M} \sum_{i=1}^{M} R_i = R. \qquad (20)$$

The minimum distortion obtained by the optimal bit allocation is

$$D = (\prod_{i=1}^{M} \sigma_i^2)^{\frac{1}{M}} \cdot 2^{-2\gamma R}. \qquad (21)$$

In our complexity scalability scheme, the first $M - \lambda_{PRE}$ MB's are encoded as AZMB's, while the remaining λ_{PRE} MB's are encoded as NZMB's. In this case, the bit rate of each AZMB is zero, and its coding distortion, denoted by D_i^z, is exactly the variance of the difference MB, i.e.,

$$D_i^z = \sigma_i^2 \cdot 2^{-2\gamma \cdot 0} = \sigma_i^2, \quad 1 \le i \le M - L,$$

where $L = \lambda_{PRE}$ is introduced to simplify the notation. Since all the coding bits are allocated among the NZMB's, according to (21), the coding distorting of each NZMB, denoted by D_i^{nz}, is given by

$$D_i^{nz} = (\prod_{i=M-L+1}^{M} \sigma_i^2)^{\frac{1}{L}} \cdot 2^{-2\gamma \frac{M \cdot R}{L}}, M - L + 1 \le i \le M.$$

The overall distortion D of the video frame, which is average distortion of the AZMB's and NZMB's, is given by

$$\begin{aligned} D &= D(L) = \frac{1}{M}[\sum_{i=1}^{M-L} D_i^z + \sum_{i=M-L+1}^{M} D_i^{nz}] \\ &= \frac{1}{M}[\sum_{i=1}^{M-L} \sigma_i^2 + L(\prod_{i=M-L+1}^{M} \sigma_i^2)^{\frac{1}{L}} 2^{-2\gamma \frac{MR}{L}}]. \end{aligned} \qquad (22)$$

To derive the expression for $D(L)$, we consider the continuous-time version of (22). Note that $\{\sigma_i^2 | 1 \le i \le M\}$ is an increasing series. Fig. 5 shows $\{\sigma_i^2\}$ for the 100-th frame of the "Foreman". Experiments on other video frames and other video sequences yield similar results. This suggests us that it is reasonable to model $\{\sigma_i^2\}$ with the following linear function

$$\mathcal{G}(t) = A \cdot t, \quad t \in [0, 1], \qquad (23)$$

such that

$$\sigma_i^2 = \mathcal{G}(\frac{i}{M}), \quad 1 \le i \le M. \qquad (24)$$

Here A is a positive constant. It should be noted that at the right end of the curve, the linear approximation is not accurate. However, since the R-D modeling is a statistical procedure to model the behavior of the whole frame, which has a large number of MB's, the approximation error within this small region won't affect much the performance of the whole model. Our simulation results which will be presented later confirm this observation. Similarly, we define $y = \frac{L}{M}$, and consider $D(y)$ as the continuous-time version of $\{D(L)\}$, i.e.,
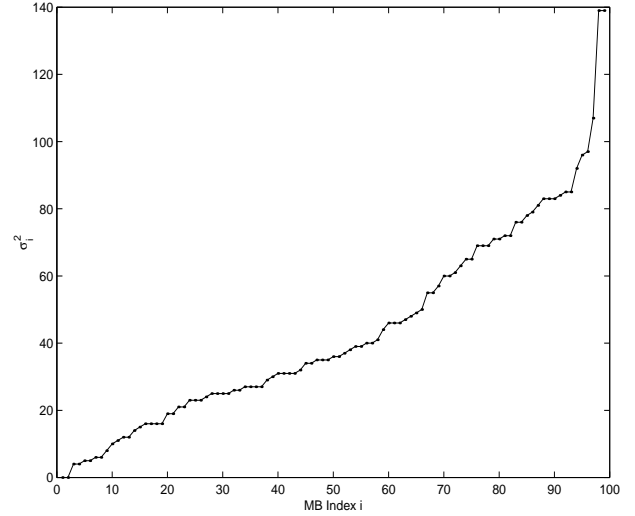
$$D(y) = D(\frac{L}{M}).$$

Fig. 5. The MB variances sorted in an ascending order for the 100-th frame of "Foreman".

Note that the first term on the right-hand side of (22) can be written as

$$\begin{aligned} \frac{1}{M} \sum_{i=1}^{M-L} \sigma_i^2 &= \int_0^{1-y} \mathcal{G}(t)dt = \int_0^{1-y} A \cdot t \, dt \\ &= \frac{A}{2}(1 - y)^2, \end{aligned} \qquad (25)$$

where $y = \frac{L}{M}$ represents the fraction of NZMB's in the video frame. Let

$$Z = (\prod_{i=M-L+1}^{M} \sigma_i^2)^{\frac{1}{L}}.$$

We have,

$$\begin{aligned} \ln(Z) &= \frac{M}{L} \cdot \frac{1}{M} \sum_{i=M-L+1}^{M} \ln \sigma_i^2 = \frac{1}{y} \int_{1-y}^{1} \ln(At) \, dt \\ &= \ln A - \frac{1}{y}[y + (1 - y) \ln(1 - y)]. \end{aligned} \qquad (26)$$

Therefore,

$$D(y) = A[\frac{1}{2}(1 - y)^2 + y e^{\frac{-1}{y}[y + (1-y) \ln(1-y)]} \cdot 2^{-2\gamma \frac{R}{y}}]. \qquad (27)$$

This model describes the complexity-rate-distortion (C-R-D) behavior of the PRECODING modules. To test the accuracy of the C-R-D model, we implement the PRECODING complexity scalability in the MPEG-4 encoder and generate the $D(y)$ curves for a set of coding bit rates R, ranging from 0.01 bpp to 1.0 bpp. Fig. 6 shows the actual $D(y)$ curves for the 100-th frame of "Foreman" and those estimated with (27). It can be seen that the estimation is very accurate.

D. Parameters Estimation and Model Simplification

The C-R-D model for the PRECODING modules given by (27) has one parameter A. Note that

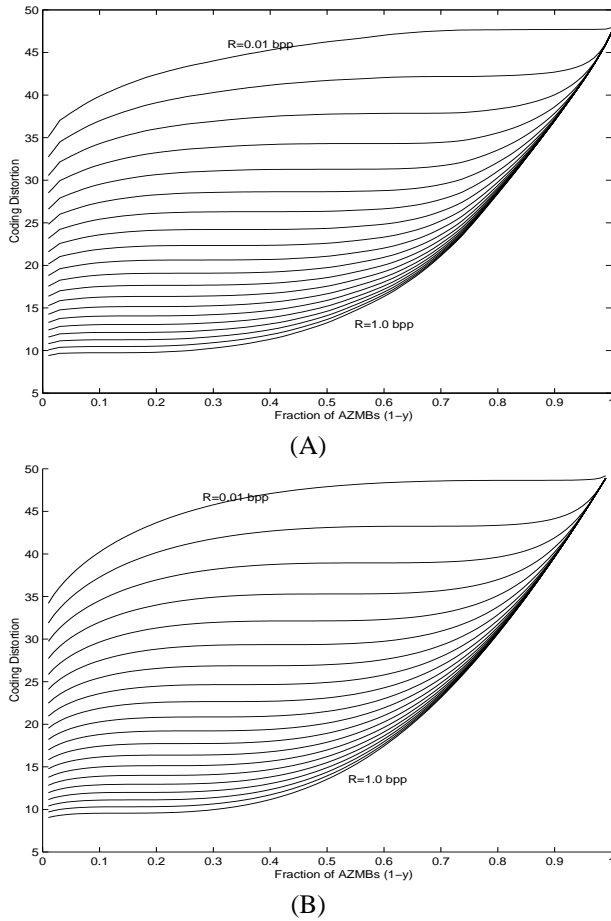$$\frac{1}{M} \sum_{i=1}^{M} \sigma_i^2 = \int_0^1 \mathcal{G}(t)dt = \frac{A}{2}. \qquad (28)$$

(A)

(B)

Fig. 6. Plot of $D(y)$ for different bit rates ($0.01bpp \leq R \leq 1.0bpp$): (A) actual results; (B) estimated by the C-R-D model in (27).

Therefore, A can be estimated by

$$A = \frac{2}{M} \sum_{i=1}^{M} \sigma_i^2 = \frac{2}{M} \sum_{i=1}^{M} SSD_i. \qquad (29)$$

The C-R-D model in (27) will be used for energy consumption control and picture quality optimization. Since the model is highly nonlinear, it is not suitable for mathematical optimization. Therefore, we need to simplify the formulation of the C-R-D model, specifically the exponential term in (27). Taylor expansion yields the following linear approximation,

$$e^{\frac{-1}{y}[y+(1-y)\ln(1-y)]} \simeq (\frac{1}{e} + \frac{1}{e^3}) + (1 - \frac{1}{e} - \frac{1}{e^3})(1-y). \quad (30)$$

Fig. 7 shows the nonlinear exponential function (solid line) and its linear approximation (dashed line). It can be seen that approximation error is relatively small. With the linear approximation, the PRECODING C-R-D model becomes,

$$D(y) = A[\frac{1}{2}(1-y)^2 + y(1+a_0 y) \cdot 2^{-2\gamma \frac{R}{y}}], \qquad (31)$$

where
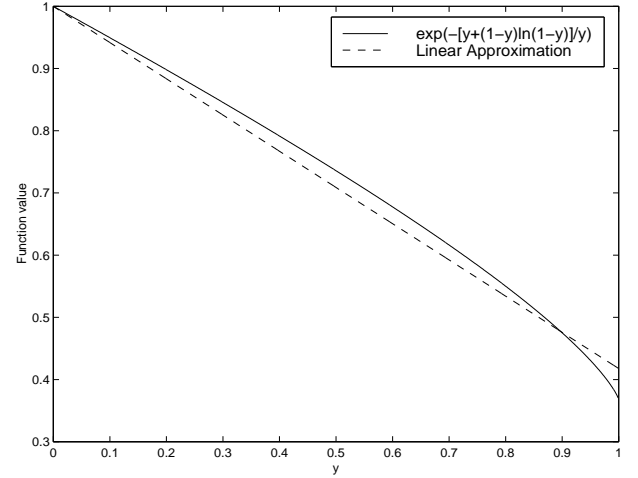
$$a_0 = \frac{1}{e} + \frac{1}{e^3} - 1. \qquad (32)$$

Fig. 7. Linear approximation of $e^{\frac{-1}{y}[y+(1-y)\ln(1-y)]}$

.

V. INTEGRATED POWER-RATE-DISTORTION MODEL

A. Considering the Frame Rate

In Section III, we have derived the complexity-scalability model for the ME module. For a complexity target of λ_{ME} SSD computations, the average MB variance is given by

$$\frac{1}{M} \sum_{i=1}^{M} \sigma_i^2 = \beta_0 + \beta_1 \cdot e^{-\beta_2 x}, \quad x = \frac{\lambda_{ME}}{\lambda_{ME}^{max}}. \qquad (33)$$

According to (29) and the PRECODING C-R-D model in (31), we have

$$
\begin{aligned}
D &= D(R; x, y) \\
&= 2(\beta_0 + \beta_1 \cdot e^{-\beta_2 x})[\frac{1}{2}(1-y)^2 \\
&\quad + y(1 + a_0 y) \cdot 2^{-2\gamma \frac{R}{y}}],
\end{aligned}
\qquad (34)
$$

where x and $y = \frac{\lambda_{PRE}}{M}$ are the normalized complexity control parameters. Both x and y range from 0 to 1, with 0 and 1 representing the lowest and highest computational complexity, respectively. It should be noted that the distortion in (34) only measures the quality for a single frame. The research in video quality evaluation suggests that the video presentation quality should be measured not only by the spatial quality of a single frame, but also by the temporal quality in motion smoothness [5]. Therefore, the encoding frame rate λ_F plays a very important role in quality evaluation. It is also a key parameter in energy consumption control. For example, at lower frame rates, more energy can be allocated to each frame to improve the spatial quality. However, in this case, the temporal video quality degrades. Although many results have been published in subjective video quality evaluation [5], most of them focus on experimental studies. For quality optimization of video coding, we need an analytic, mathematically tractable model to describe the video presentation quality. The experimental results in [5] suggest that the video presentation quality D_v should consist of two parts: the spatial quality of a single picture $D_{spatial}$ and the temporal motion quality $D_{temporal}$. $D_{spatial}$ is given by (34). $D_{temporal}$ depends on the encoding

frame rate. In typical video decoding and display, if a video frame is skipped, the previous decoded picture stays on the screen until the next frame is decoded. In other words, the decoder reconstruction of the skipped frame is the copy of its previous decoded frame. From the video encoder point of view, the ME complexity x, the PRECODING complexity y, and the bit rate R of the skipped video frame are all zeros. Therefore, from (34), we can see that its coding distortion is given by

$$D_{temporal} = D(R; x, y)|_{R=0, x=0, y=0} = \beta_0 + \beta_1, \quad (35)$$

which is the MSE between the skipped frame and its previous reconstruction. Let ω_s and ω_t be the perceptual weight on the spatial quality and temporal quality, respectively. The experimental results in [5] suggests that ω_s and ω_t should be a function of the frame rate. For example, if the video encoder encodes only one frame per minute, although each picture has very high quality, the viewer will complain about the bad video streaming service because he has missed a lot of important motion information and the spatial information in between. In this work, we choose the perceptual weight as follows,

$$\omega_t = (1 - z)^2, \quad \omega_s = 1 - \omega_t, \quad (36)$$

where $z = \frac{\lambda_F}{f_{max}}$, and f_{max} is the maximum frame rate with a default value of 30 fps. Therefore, the video presentation quality is defined as

$$\begin{aligned} D_v &= \omega_s \cdot D_{spatial} + \omega_t \cdot D_{temporal} \\ &= (1 - z)^2 (\beta_0 + \beta_1) + 2(2z - z^2)(\beta_0 + \beta_1 e^{-\beta_2 x}) \\ &\quad \cdot [\frac{1}{2}(1 - y)^2 + y(1 + a_0 y) \cdot 2^{-2\gamma \frac{R}{y}}]. \end{aligned} \quad (37)$$

B. R-D Optimized Power Consumption Control

From (7), we can derive the relationship between the power consumption and the complexity control parameters,

$$\Phi(P) = z(C_1 x + C_2 y + C_3 R), \quad (38)$$

where C_1, C_2, and C_3 are constants. For a given power supply level P and a given rate R, we need to find the best configuration of the complexity parameters for the ME and PRECODING modules to maximize the picture quality. Mathematically, this can be formulated as in (39). The minimization parameters (x, y, z) can be obtained using binary search of the minimum point. Note that the battery often has an operational lifetime of several hours, several days, or even several weeks. Therefore, there is no need to adjust the power control parameters too often, say every second, because the power supply condition doesn't change that quickly. Suppose the adjustment period is 5 seconds. This means we only need solve the the R-D optimized power control problem in (39) once per 5 second. Therefore, the overhead of the power control is relatively small. In our future work, we shall investigate the possibility of further simplification of the model and its solution as well.

VI. R-D Optimized Power-Scalable Video Encoding

Using the P-R-D model and the optimal configuration of the power control parameters, the video encoder is able to achieve the R-D optimized power consumption scalability. The R-D optimized power-scalable video encoder system operates as follows:

- **Step 1, Determining the model parameters:** In (39), the ME model parameters β_0, β_1, β_2 are estimated from the statistics of previous frames using linear regression. a_0 is a constant determined by (32). The model parameter γ is also determined from the R-D statistics of the previous frames. At beginning stage, for example the first second of video encoding, no power control is applied, because the system has sufficient power supply.
- **Step 2, Optimization:** Find the optimal complexity control parameters $\{x, y, z\}$ using (39). This step is executed only if the power control is triggered according to the adjustment frequency, for example, once per 5 seconds.
- **Step 3, Frame rate and ME complexity control:** Set the encoding frame rate to be $\lambda_F = z \cdot f_{max}$. The available SSD computations for ME is given by $\lambda_{ME} = x \cdot \lambda_{ME}^{max}$. Using the MHM-based allocation scheme presented in Section III-B to allocate the SSD computation among the MB's. Using the fast and efficient diamond ME scheme to find the motion vector and the minimum SSD for each MB. The number of diamond search layers is controlled by the allocated SSD computations.
- **Step 4, PRECODING complexity control:** Find the $(1 - y) \cdot M$ MB's with the smallest SSD values and force them to be AZMB's. The PRECODING operation is applied to the remaining NZMB's. Dynamic rate control is used to reallocate the bits from the AZMB's to the NZMB's.

It can be seen that the complexity of the major encoding modules is controlled by the parameter set $\{x, y, z\}$ to match the power supply level of the mobile device. At the same time, these parameters are configured according to the P-R-D model such that the overall video quality is optimized.

VII. Experimental Results

To evaluate the performance of the P-R-D model and the power-scaling video encoding system, we implement the proposed P-R-D model and power scalability scheme in the public domain H.263+ encoder. Similar performance is expected for other coding systems, such as MPEG-2 and MPEG-4. In our simulations, the maximum search points for each MB λ_{ME}^{max} is 50, and the maximum frame rate $f_{max} = 30$ fps. To test the accuracy of the P-R-D model, we run the video encoder over the "Foreman" QCIF sequence at 128 kbps and 15 fps for different complexity control parameters (x, y) and measure the corresponding distortion. Fig. 8 shows the actual distortion function $D(x, y)$. The estimation given by the P-R-D model is shown in Fig. 9. We can see that model estimation is quite accurate. Simulations over other test videos yield similar results. For a given bit rate R and device power supply level, using (39) the encoder can find the best configuration of complexity control parameters to maximize

$$\min_{\{x,y,z\}} D_v(R; x, y, z) = (1-z)^2(\beta_0 + \beta_1) + 2(2z - z^2)(\beta_0 + \beta_1 e^{-\beta_2 x})[\frac{1}{2}(1-y)^2 + y(1 + a_0 y) \cdot 2^{-2\gamma \frac{R}{y}}],$$

$$s.t. \quad \Phi(P) = z(C_1 x + C_2 y + C_3 R). \tag{39}$$

the video quality. Figs. 10 to 12 show the picture distortion, and the optimal control parameters $\{x, y, z\}$ as functions of the percentage of power consumption for different coding bit rates R. Some interesting observations can be made: (1) As the encoder scales down its power consumption, as a percentage of its maximum power consumption level, the video quality degrades. The video encoding automatically changes from high quality motion video coding (when the energy supply is plenty) to still image coding (when the device is running out of energy). (2) At lower bit rates, the ME wins over the PRECODING in power allocation, because the ME is computation-hungry but the PRECODING is bit-rate-hungry; hence, as shown in Fig. 10, the complexity for the ME is high but the complexity for the PRECODING is low. Fig. 13 shows the "Carphone" QCIF video coded at 64 kbps and 15 fps for different power consumption levels. We can see the picture quality degradation is very graceful. Fig. 14 shows the achievable minimum distortion D as a function of R and the power P. To view the P-R-D model in more detail, we plot the D-P curves for different bit rates, ranging from 0.01 bpp to 1.0 bpp in Fig. 15. Fig. 16 shows the D-R curves at different power consumption levels. We can see that when the power supply level is low, the $D(R)$ function is almost flat, which means the video processing and encoding efficiency is very low; hence, in this case, more bandwidth does not improve the video presentation quality. We can see that the P-R-D model has direct applications in energy management, resource allocation, and QoS provisioning in wireless video communication, especially over wireless video sensor networks.
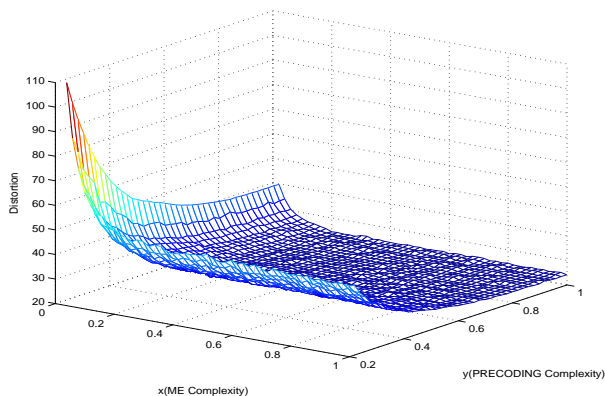

Fig. 9. The complexity-distortion surface $D(x, y)$ estimated by the P-R-D model.
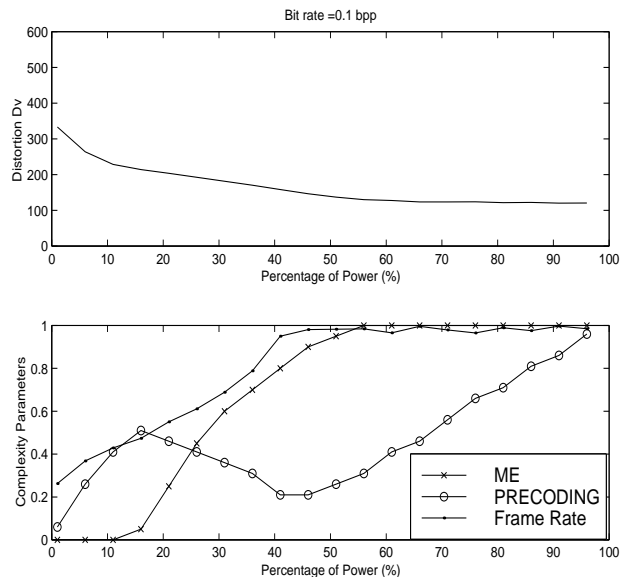
Fig. 10. R-D optimized power control for the "Football" CIF video at $R = 0.1bpp$, about 150 kbps at 15 fps.

Fig. 8. Actual complexity-distortion surface $D(x, y)$

VIII. CONCLUDING REMARKS

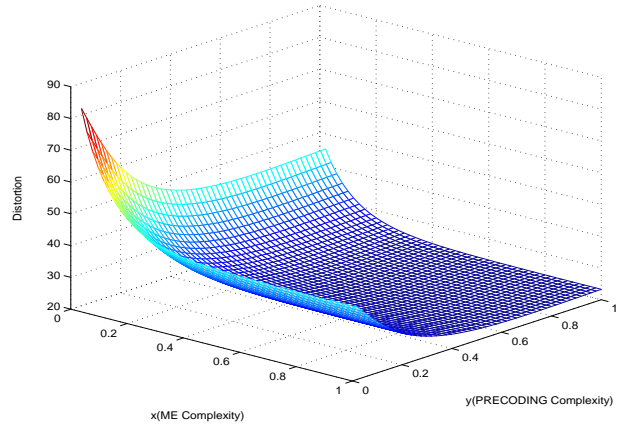There are two major contributions in this work. First, based on the complexity analysis of typical video encoding systems and the DVS CMOS design technology, we have developed a parametric video encoding architecture which is fully scalable in power consumption. Second, we have successfully extended the traditional R-D analysis by considering another dimension, the power consumption, and established the P-R-D analysis framework for mobile video encoding and communication under energy constraints. Using the P-R-D model, given a power supply level and a bit rate, the power-scalable video encoder is able to find the best configuration of complexity control parameters to maximize the video quality. The P-R-D analysis establishes a theoretical basis and provides a practical guideline in system design and performance optimization for
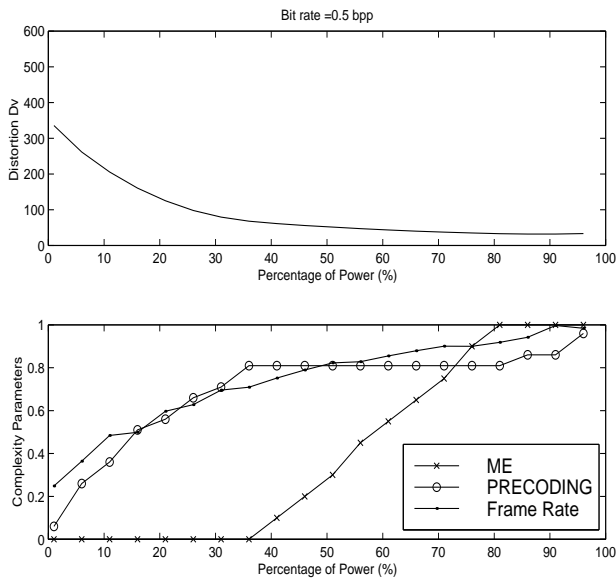
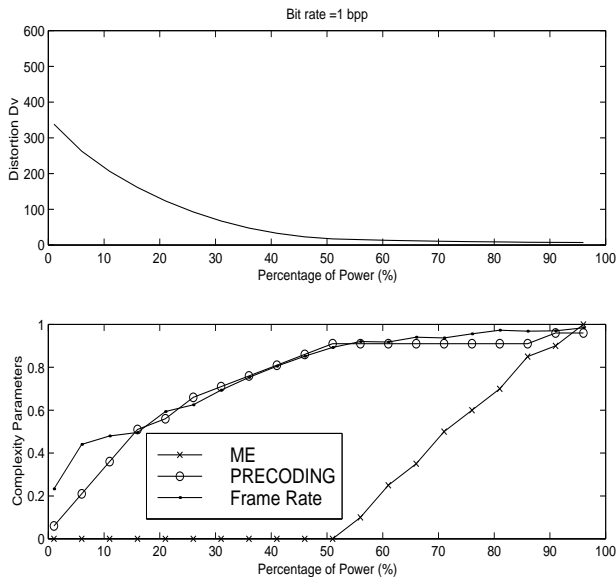Fig. 11. R-D optimized power control for the "Football" CIF video at $R = 0.5bpp$.

Fig. 13. The encoded "Carphone" QCIF sequence at 64 kbps and 15 fps for different power supply level.

Fig. 12. R-D optimized power control for the "Football" CIF video at $R = 1.0bpp$.

Fig. 14. The P-R-D Model.

wireless video communication under energy constraints. In our future work, we will use the P-R-D model developed in this paper for joint resource allocation and control for video encoding and wireless transmission.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] P. Agrawal , J.-C. Chen, S. Kishore, P. Ramanathan, and K. Sivalingam, "Battery power sensitive video processing in wireless networks," *Proceedings IEEE PIMRC'98*, Boston, September 1998.

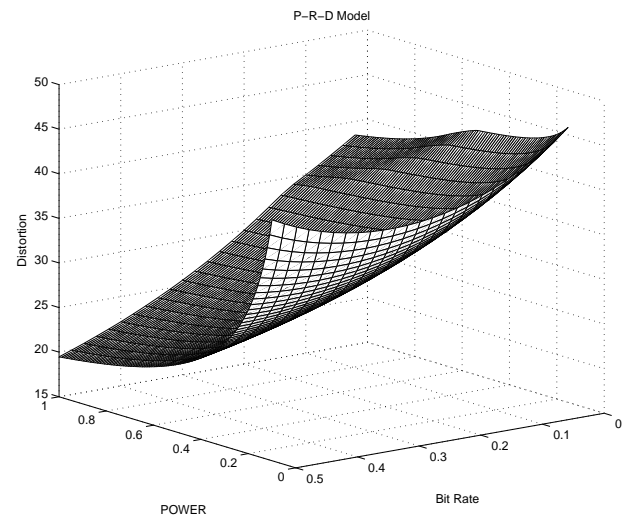[2] AMD Inc, "AMD PowerNow!TM Technology Platform Design Guide for Embedded Processors," http://www.amd.com/epd/processors.

[3] S. M. Akramullah, I. Ahmad, and M. L. Liou, "Optimization of H.263 video encoding using a single processor computer: performance tradeoffs and benchmarking," *IEEE Trans. On Circuits and System for Video Technology*, vol. 11, pp. 901 - 915, August 2001.

[4] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communication Magazine*, pp. 102 - 114, August 2002.

[5] R. T. Apteker, J. A. Fisher, V. S. Kisimov, and H. Neishlos "Video acceptability and frame rate" *IEEE Multimedia*, vol 2, No. 3, pp. 32-40, 1995.

[6] T. Berger, *Rate Distortion Theory,* Prentice Hall, Englewood Cliffs, NJ, 1984.

[7] T. Burd and R. Broderson, "Processor Design for Portable Systems," *Journal of VLSI Signal Processing,* vol. 13, no. 2, pp. 203–222, August 1996.

[8] W. P. Burleson, P. Jain, and S. Venkatraman, "Dynamically Parameterized Architecture for Power-Aware Video Coding: Motion Estimation and DCT," *Proceedings of the Second USF International Workshop on Digital and Computational Video*, 2001.

[9] L. Chen, Z. He, S. Sethuraman, and C. W. Chen, "MPEG-4 encoder implementation on MAP-CA DSP," *Proceedings of International Conference on Consumer Electronics*, Los Angeles, CA, June 2002.

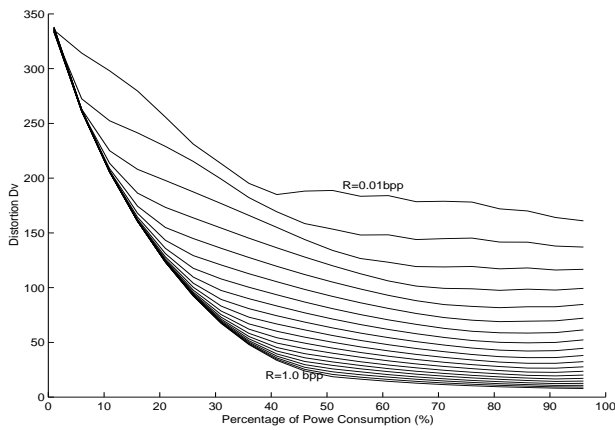[10] T. Chiang, Y.-Q. Zhang, "A new rate control scheme using quadratic

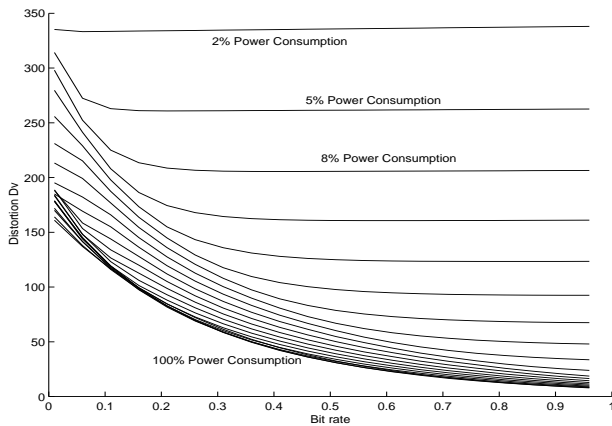Fig. 15.   The D-P curves for different bit rates.

Fig. 16.   The D-R curves for different power consumption levels.

rate distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.7, pp. 246 – 250, February 1997.

[11] B. Erol, F. Kossentini, and H. Alnuweiri, "Efficient coding and mapping algorithms for software-only real-time video coding at low bit rates," *IEEE Trans. On Circuits and System for Video Technology*, vol. 10, pp. 843 - 856, Sep. 2000.

[12] Z. He and S. K. Mitra, "A Unified Rate-Distortion Analysis Framework for Transform Coding," *IEEE Transactions on Circuits and System on Video Technology,* vol. 11, pp. 1221 -1236, December 2001.

[13] Z. He and S. K. Mitra, "A linear source model and a unified rate control algorithm for DCT video coding," *IEEE Transactions on Circuits and System on Video Technology*, vol. 12, pp. 970 - 982, November 2002.

[14] Z.-L. He, C.-Y. Tsui, K.-K. Chan, and M. Lion, "Low-power VLSI design for motion estimation using adaptive pixel truncation," *IEEE Trans. On Circuits and System for Video Technology*, vol. 10, August 2000.

[15] Intel Inc, "Intel XScale Technology," http: // www.intel.com / de-sign/intelxscale.

[16] "MPEG-2 Video Test Model 5," *ISO/IEC JTC1/SC29/WG11 MPEG93/457*, April 1993.

[17] ITU-T, "Video coding for low bit rate communications," *ITU-T Recom-mendation H.263*, version 1, version 2, January 1998.

[18] J. Lorch and A. Smith, "Improving dynamic voltage scaling algorithms with PACE," *Proceedings of the ACM SIGMETRICS 2001 Conference,* June 2001.

[19] X. Lu, Y. Wang, and E. Erkip, "Power efficient H.263 video transmission over wireless channels," *Proceedings of 2002 International Conference on Image Processing*, Rochester, New York, September 2002.

[20] R. Min, T. Furrer, and A. Chandrakasan, "Dynamic Voltage Scaling Techniques for Distributed Microsensor Networks", *IEEE Computer Society Workshop on VLSI (WVLSI '00)*, April 2000, pp. 43-46.

[21] I. M. Pao and M. T. Sun, "Statistical Computation of Discrete Cosine Transform in Video Encoders," *Journal of Visual Communication and Image Representation*, vol. 9, no. 2, pp.163-170, June 1998.

[22] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, pp. 172 – 185, February 1999.

[23] T. Sikora, "The MPEG-4 video standard verification model," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 7, pp. 19–31, February 1997.

[24] A. M. Tourapis, O. C. Au, and M. L. Liou, "Predictive Motion Vector Field Adaptive Search Technique (PMVFAST) - Enhancing Block Based Motion Estimation," *proceedings of Visual Communications and Image Processing 2001 (VCIP-2001)*, San Jose, CA, January 2001.

[25] J. Villasenor, C. Jones, and B. Schoner, "Video Communications using Rapidly Reconfigurable Hardware," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, pp. 565-567, Dec. 1995.

PLACE PHOTO HERE

Zhihai He received the B.S. degree from Beijing Normal University, Beijing, China, and the M.S. degree from Institute of Computational Mathematics, Chinese Academy of Sciences, Beijing, China, in 1994 and 1997 respectively, both in mathematics, and the Ph.D. degree from University of California, Santa Barbara, CA, in 2001, in electrical engineer-ing. In 2001, he joined Sarnoff Corporation, Prince-ton, NJ, as a Member of Technical Staff. In 2003, he joined the Department of Electrical and Computer Engineering, University of Missouri, Columbia, as an assistant professor. He received the 2002 IEEE Transactions on Circuits and Systems for Video Technology Best Paper Award, and the SPIE VCIP Young Investigator Award in 2004. His current research interests include image/video processing and compression, network transmission, wireless communication, computer vision analysis, sensor network, and embedded system design. He is a member of the Visual Signal Processing and Communication Technical Committee of the IEEE Circuits and Systems Society, and serves as Tech-nical Program Committee member or session chair of several international conferences.

PLACE PHOTO HERE

Yongfang Liang Yongfang Liang received the B.S. and M.S. degrees from Zhongshan(Sun Yat-sen) University, Guangzhou, China, in 1998 and 2001 respectively, all in electrical engineering. From 2001 to 2002, he was with the HuaWei Technologies Corporation, Shenzhen, as Researcher and Software Engineer in the Multimedia Department. Currently he is pursuing the Ph.D. degree in computer science and engineering, at the Universiy of Texas at Ar-lington. His current research interests include video compression, wireless multimedia communication, and image processing.

PLACE PHOTO HERE

Lulin Chen Lulin Chen received his BS degree in Electrical Engineering from the University of Science and Technology of China in 1982, his MSEE from East China Normal University in 1985, and his PhD degree in Electrical-Optical Engineering from Shanghai Institute of Technical Physics (SITP), Chinese Academy of Sciences (CAS), in 1992. From 1992 to 1995 he was an Assistant/Associate Researcher at SITP, CAS. He was a Doctoral Researcher at the Institute of Industrial Sciences at University of Tokyo, Japan (1993-1994) and a Postdoctoral Research Fellow in the Department of Electrical Engineering at University of Rochester (1995-1997). Since August 2001, he has been a Technology Leader at Sarnoff Corporation. Prior to his current position, he was with some industrial organizations including Xerox and PictureTel. He received Awards of Daheng, Science and Technology Advances, and Natural Science from CAS in 1992, 1993 and 1995, respectively, and invented many technical approaches in prototype systems and product lines. His current research interests include image/video processing and networking, parallel computation, codec technology, embedded media system, and system on chip.

PLACE PHOTO HERE

Dapeng Wu (S'98–M'04) received B.E. in Electrical Engineering from Huazhong University of Science and Technology, Wuhan, China, in 1990, M.E. in Electrical Engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 1997, and Ph.D. in Electrical and Computer Engineering from Carnegie Mellon University, Pittsburgh, PA, in 2003. From July 1997 to December 1999, he conducted graduate research at Polytechnic University, Brooklyn, New York. During the summers of 1998, 1999 and 2000, he conducted research at Fujitsu Laboratories of America, Sunnyvale, California, on architectures and traffic management algorithms in the Internet and wireless networks for multimedia applications.

Since August 2003, he has been with Electrical and Computer Engineering Department at University of Florida, Gainesville, FL, as an Assistant Professor. His research interests are in the areas of networking, communications, multimedia, signal processing, and information and network security. Currently he is an associate editor for the IEEE Transactions on Vehicular Technology.

Dr. Wu received the IEEE Circuits and Systems for Video Technology (CSVT) Transactions Best Paper Award for Year 2001.

PLACE PHOTO HERE

Ishfaq Ahmad Dr. Ahmad received a BSc degree in Electrical Engineering from the University of Engineering and Technology , Lahore , Pakistan , in 1985, and an MS degree in Computer Engineering and a PhD degree in Computer Science from Syracuse University, New York, U.S.A., in 1987 and 1992, respectively. His recent research focus has been on developing parallel programming tools, scheduling and mapping algorithms for scalable architectures, heterogeneous computing systems, distributed multimedia systems, video compression techniques, and web management. His research work in these areas is published in over 125 technical papers in refereed journals and conferences, with best paper awards at Supercomputing 90 (New York), Supercomputing '91 (Albuquerque), and 2001 International Conference on Parallel Processing (Spain). He is currently a full professor of computer science and engineering in the CSE Department of the University of Texas at Arlington.

Prioir to joining UT Arlington, he was an associate professor in the Computer Science Department at HKUST in Hong Kong. At HKUST, he was also the director of the Multimedia Technology Research Center , an officially recognized researh center that he conceived and built from scratch. The center was funded by various agencies of the Government of the Hong Kong Special Administrative Region as well as local and international industries. With more than 40 personnel including faculty members, postdoctoral fellows, full-time staff, and graduate students, the center engaged in numerous research and development projects with academia and industry from Hong Kong, China, and the U.S. Particular areas of focus in the center are video (and related audio) compression technologies, videotelephone and conferencing systeme. The center commercialized several of its technologies to its industrial partners world wide.

He has participated in the organization of several international conferences and is an associate editor of Cluster Computing, Journal of Parallel and Distributed Computing, IEEE Transactions on Circuits and Systems for Video Technology, IEEE Concurrency, and IEEE Distributed Systems Online.