A Hierarchical Control Scheme for Energy Quota Distribution in Hybrid Distributed Video Coding

Muhammad Usman Karim Khan, Muhammad Shafique, Jörg Henkel
Karlsruhe Institute of Technology (KIT), Chair for Embedded Systems, Karlsruhe, Germany
{muhammad.khan, muhammad.shafique, henkel}@kit.edu

ABSTRACT

Hybrid Distributed Video Coding (HDVC) has emerged as an attractive solution for energy and resource-constrained video processing systems in different domains like distributed video sensors, mobile devices with dynamically varying energy levels, or hybrid paradigms where mobile devices are communicating with high-end servers. In this paper, we present a hierarchical control scheme for determining and distributing the energy quota in HDVC encoder and decoder under scenarios of dynamically varying energy levels and user constraints. The energy quotas are computed and controlled online at various hierarchical levels (like group of frames, frames, and macroblocks), while jointly accounting for the computation and transmission energy. For energy reduction, the processing of selective regions is intelligently distributed at both encoder and decoder sides, considering the texture and motion properties of different video regions and available spatial/temporal correlations. Experimental results demonstrate that our scheme provides on average 20%-25% reduced energy consumption compared to state-of-the-art HDVC schemes. The key to high energy efficiency is to leverage the video content properties.

Categories and Subject Descriptors: C.3 [Special-Purpose and Application-Based Systems]: Real-time and embedded systems; I.4.2 [Compression (Coding)]: Approximate Methods

General Terms: Algorithms, Design, Management, Performance

Keywords: Video Coding, Distributed Video Coding (DVC), Hybrid Distributed Video Coding (HDVC), Motion Estimation, Low-Power, Adaptivity, Energy Budgeting

1. INTRODUCTION & RELATED WORK

The high compression efficiency of the H.264/AVC [2] advanced video coding standard has enabled a wide range of applications like video sensors under low-bandwidth constraints, surveillance with smart cameras, entertainment, mobile multimedia, etc. To achieve high compression, H.264 employs variable block-sized motion estimation, deblocking filter, and context-adaptive entropy coding [3]. This results in a significantly increased computational complexity and energy consumption (~10x relative to MPEG-4 ASP [4][11]). H.264/AVC follows the predictive video coding (PVC) model where the prediction (motion estimation, motion compensation, etc.) is performed at the encoder side. In this scenario, computational requirements of the H.264 decoder are 10x-20x lesser compared to that of the H.264 encoder [1][4]. However, due to high compression efficiency, the transmission power for PVC is quite low. Therefore, PVC is typically well-suited for scenarios where encoding devices have high computational power available, while decoders are resource/power-constrained devices, like video streaming for mobile devices, etc.

Recently, Distributed Video Coding (DVC) has been emerged as an attractive solution for scenarios where the encoding devices are resource-constrained requiring low-complexity encoding (like in-field wireless video sensor nodes, small autonomous flying robots, mobile devices with low processing capabilities, etc.), while the decoding devices have high computational power and can execute high-complexity decoding (like servers, high-end transmission head) [13][6][15][27][28][23][12]. DVC paradigm provides means to shift/offload the computational workload from encoder to decoder. A DVC encoder typically consumes only 7% of the total power consumed by a PVC H.264 encoder (see Fig. 1a) [16][17]. In DVC paradigm, instead of the encoder, the decoder performs the Motion Estimation (ME) with interpolation, extrapolation, and upsampling, while exploiting the inter-frame correlation in order to generate an estimate of the input video sequence. At the DVC decoder side, Slepian-Wolf decoder and ME with interpolation contribute towards 90% of the total decoding complexity [15] (see DVC details in Section 1.1). To improve the estimation quality, the DVC decoder demands auxiliary information from the encoder (i.e. parity bits generated by turbo coding), which results in a much higher transmission power compared to the PVC case (see Fig. 1b). More are the parity bits, higher is the video quality at the decoder side and higher is the transmission energy at the encoder-side [29][30].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CODES+ISSS’12, October 7–12, 2012, Tampere, Finland.
Copyright 2012 ACM 978-1-4503-1426-8/12/09...$15.00.

Fig. 1 (a) Relative computation power comparison for different coding systems; (b) Comparing the computation and transmission power for an ASIC-based video sensor [16]; wireless transmitter of [18]; H.264 implementation of [17]

Fig. 1b shows that for the “Hall” test video sequence, the transmission power of a DVC-based system is about 4x more than that of the PVC-based H.264 coding system for a 65 nm ASIC-based video sensor node [13]. Note, although in this example of “Hall” sequence, the power of three codecs is almost similar, the power consumption of PVC and DVC can significantly vary depending upon the video sequence properties1 (see case study in Section 2) [11][16]. The major drawback of DVC is lower video quality compared to that provided by PVC [25]. In short, DVC is beneficial in scenarios, where the encoder-side devices are computation-constrained and have sufficient transmission power, while the decoder-side devices have sufficient computational power available.

1 For surveillance type sequences with low motion, DVC power consumption is lower than PVC; while for entertainment quality sequences with hectic motion, PVC power consumption is lower than DVC.
Issues with PVC and DVC: Both PVC and DVC individually become power-/energy-wise inefficient in scenarios where both encoder- and decoder-side devices have sufficient computational and/or transmission power to deliver the required quality of service (frame rate, etc.). Examples of such scenarios are: (1) collaborative distributed video sensor networks for smart energy-aware surveillance; (2) mobile devices – with dynamically varying battery energy levels – communicating with each other or with high-end power-constrained laptops/servers; (4) heterogeneous communicating devices from different vendors with distinct energy consumption properties, etc.[13][18][31]; see Fig. 2.

Hybrid Distributed Video Coding (HDVC) as an Emerging Coding Paradigm: To cope with the energy-related issues for video coding in such dynamic scenarios, Hybrid Distributed Video Coding (HDVC) has emerged as an attractive solution [24][25][10]. HDVC aims at combining the positive aspects of both PVC and DVC, i.e. providing high encoded video quality close to PVC and low computational power close to DVC. In HDVC, the decoder-side ME complexity is relaxed at the cost of partial ME at the encoder side. The partial ME at the encoder side results in better reconstruction of frames at the decoder side, that corresponds to a high video quality and low energy consumption at the decoder side. However, it incurs high energy consumption at the encoder side due to ME processing. Note, ME is the most energy consuming functional block and may consume up to 65% of the total encoding energy [5][11]. Better ME at the HDVC encoder side also reduces the number of parity bits that results in low transmission energy [8]. However, this depends upon the amount of ME to be performed at the encoder side (see Section 2 for detailed analysis). Complex motion sequences may require more ME at the HDVC encoder side to meet the quality of service requirements, which may not be feasible due to the unavailability of sufficient computational and battery energy resources at the HDVC encoder side.

Challenging Problem: In order to realize HDVC in resource- and energy-constrained devices subjected to dynamically varying scenarios, an adaptive energy quota distribution scheme is required. Such a scheme needs to compute and control the energy quota for determining the ME effort at both HDVC encoder and decoder sides, such that overall energy (considering computation and transmission) is minimized, while satisfying the user constraints.

Prominent state-of-the-art techniques in low-power HDVC (like [9][21][22]) determine the number of Macroblocks (MB: 16x16 pixel blocks) to be processed at the encoder side in a raster scan order, use global motion models, send a low quality reference or process it at the encoder side based upon MB’s motion intensity. Our experiments in Section 2 illustrate that objects in 16x16 pixel blocks) to be processed at the encoder side in a raster scan order or based on mere motion intensity may lead to inefficient energy distribution in HDVC. As a result, these state-of-the-art techniques may quickly exhaust the available energy quotas for the ME of non-important MBs. In general, state-of-the-art techniques in HDVC and DVC may not be energy efficient as they do not exploit (i) video content properties (like texture, motion, etc.); (ii) the relationship between video content properties and the ME effort and parity bits produced; and (iii) correlation of MBs. Therefore, the energy quota distribution and control scheme need to perform intelligent partitioning of ME in HDVC by exploiting the video content properties while minimizing the overall energy.

Our Novel Concept and Contributions in a Nutshell: To address the above-discussed challenges, we propose a novel hierarchical control scheme for HDVC that performs video content-aware adaptive energy quota distribution and control at multiple hierarchical levels (i.e. group of frames, frame, and MB) for both HDVC encoder and decoder. Our scheme reduces the overall energy consumption by jointly considering for the computation and transmission energy under scenarios of dynamically varying energy levels and user constraints. Our scheme accounts for video content properties (i.e. texture and motion) for selecting MBs from the regions of interest (ROIs) to perform selective ME, while efficiently utilizing the allocated energy quotas. Fig. 3 shows an overview of our HDVC codec with novel contributions in colored filled-blocks.

![Fig. 3 Our HDVC system with hierarchical control scheme for run-time adaptive energy quota distribution and control](image-url)

Our scheme employs:

1) A group-of-frame-level energy control algorithm (Section 3.1) that computes the target energy quota for a complete group of frames depending upon the battery level and user-defined coding duration constraints. It then controls the target quota of the subsequent group-of-frames in a feedback loop.

2) A frame-level control energy algorithm (Section 3.2) that adaptively distributes the group-level energy quota among different frames depending upon their temporal distances. It thereby, accounts for the changing temporal correlation among different frames. The energy quota among different frames is controlled to avoid violation of group-level quota.

3) An ROI-identification and extrapolation algorithm (Section 3.3) that detects the ROIs by evaluating the motion properties of different MBs w.r.t. their spatial and temporal neighboring MBs. A ranking mechanism is used to select important ROI-MBs for HDVC encoder-side ME depending upon the frame-level energy quota.

4) An MB-level energy control algorithm (Section 3.4) that distributes the frame level energy quota among the selected ROI-MBs starting from the highest rank value. It controls the MB-level energy quota by back-propagating the weighted error between consumed and target MB energy.
side, the allocated energy quota is used for determining the ME/SI configuration for the MBs that are not selected for the encoder-side ME.

**Paper Organization**: Section 1.1 presents an overview of PVC, DVC, and HDVC. Section 2 presents a motivational analysis of HDVC for energy and correlation. Our novel hierarchical control scheme for energy quota distribution is presented in Section 3. Section 4 presents the results followed by conclusion in Section 5.

1.1 DVC/HDVC: Terminology and Basics

The DVC concept originated from the works of Slepian-Wolf [14] and Wyner-Ziv [6]. Prominent research projects are DISCOVER [19][13][20]. Due to resource constraints, the DVC encoder only performs spatial predictions, i.e. so-called Intra-frame encoding of Key-Frames (denoted as I-frames in this paper). Other video frames between two I-frames are called Non-Key-Frames or Wyner-Ziv frames (denoted as W-frames in this paper). For W-frames, encoder only sends the auxiliary information, i.e. parity bits generated using turbo coding. The decoder reconstructs the W-frames using decoder-side ME and interpolation and uses parity bits to correct the estimates. The basic principle of decoder-side reconstruction is to exploit the correlation between two I-frames. The decoder-side estimate of W-frames is called Side Information (SI). This estimate is improved by requesting more parity bits from the encoder side. See more details on SI generation mechanisms in [8][32]. Note that a better estimate of SI at the decoder implies better reconstructed video quality and lesser transmission of parity bits from the encoder, thus reducing the transmission energy.

HDVC extends the concepts of DVC and PVC by adaptively distributing the ME at both encoder and decoder sides [24]. It is beneficial in scenarios, where both encoding and decoding devices are resource/power-constrained. The group of W-frames and preceding I-frame is called the Group of W-frames (GOW); see Fig. 4. The number of frames in a GOW is denoted as size of GOW (SGOW=m+1); 'm' is the number of W-frames and 'I' is the I-frame.

Before proceeding to the details of our novel scheme, we present an experimental case study to accentuate the research challenges and issues related to the energy quota distribution problem in HDVC. This analysis provides the foundation for this work.

2. MOTIVATIONAL CASE STUDY: HDVC ENERGY AND CORRELATION ANALYSIS

We have performed an experimental case study for the scenario shown in Fig. 2, where two energy-constrained mobile devices are communicating with each other over wireless network using an HDVC video codec; see experimental setup in Section 4. This case study demonstrates the impact of diverse video properties on the distribution of computation and transmission energy. Additionally, we have implemented the approach of [9] to illustrate the limitations of state-of-the-art that allocates the energy quota for ME of the MBs in the raster scan order, i.e. ignoring the video content properties.

Fig. 5 shows varying distribution of computation and transmission energy for different test video sequences with diverse texture and motion properties. Note that for a given number of MBs for which ME has to be processed at the HDVC encoder side, there is an increase in the ME energy due to extensive ME search for fast moving MBs. Moreover, since for the remaining MBs decoder may not get accurate matches, it results in increased transmission energy of encoder due to more parity bits. Such an increase in the computation and transmission energy can be seen in the “Coastguard” sequence (Fig. 5a), which contains river water with ripples and moving boats that are hard to predict. Due to low texture/motion, “Mother and Daughter” and “Hall” sequences have reduced transmission and computation energy compared to the other two sequences with same Peak Signal to Noise Ratio (PSNR)=36 dB; see Fig. 5a.

**Hint-1**: More energy quota should be allocated for encoder side ME in case of video scenes with high texture and motion. Higher energy savings for computation and transmission at the encoder-side can be obtained for the low-texture and low-motion video scenes, as the decoder – with a high probability – generates good quality reconstructed video frames. The key is to leverage the video content properties during the energy quota distribution in order to balance the ME computation energy and the parity bit transmission energy at the encoder side, such that the video quality achieved at the decoder side is high.

Even within a video scene, due to their diverse texture and motion properties, it is important to study that which MBs have the highest impact on the computation and transmission energy. For this, we have performed experiments with different number of MBs processed for encoder-side ME. More encoder-side ME leads to high ME computation energy, but reduced transmission energy (see Fig. 5b). However, for a given number of MBs to be processed for ME, the energy consumption highly depends upon which MBs are selected for encoder-side ME. State-of-the-art techniques in HDVC (like [9]) select MBs in the raster select for ME processing. Since MBs of ROI typically do not lie on the raster scan order (see “Foreman” and “football players” in see Fig. 6), such techniques may lead to high transmission energy in case of few MBs processed for encoder-side ME; see Fig. 5b. Moreover, ME of background MBs (low-texture/motion, static MBs; see Fig. 6) may not provide effective reduction in the parity bits, thus leading to higher transmission energy (see Fig. 5b). Since such MBs can be easily regenerated at the decoder side, there is no need to waste energy for encoder-side ME of such MBs. Furthermore, ME of many complex MBs of background regions may quickly exhaust the available energy quotas for the ME. Therefore, it is beneficial to
spend encoder-side energy in the ME of the complex MBs of the ROI, because in this case, the decoder will have difficulty in estimating such MBs (as our results illustrate in Section 4).

**Hint-2:** Computation and transmission energy in HDVC highly depends upon the MB types selected for the encoder-side ME. Higher overall energy savings and better video quality can be achieved if the MBs of objects in ROI are selected for encoder-side ME processing. Such knowledge needs to be incorporated during the energy quota distribution. The key challenges are low-overhead ROI identification, ROI-based MB selection for different video scenes, and adaptive energy quota distribution considering this knowledge.

Fig. 7 demonstrates that there is extensive spatial and temporal correlation between the MBs of the same frame and neighboring frames, respectively. Typically MBs of objects with low variance have high spatial correlation [36][37][38][39]. Similarly slow-moving MBs have high temporal correlation. Fig. 7 shows motion vector drift (motion vector difference of spatial neighbors) and motion vector difference between the current and collocated MBs (i.e. temporal neighbors). This correlation can be used to efficiently predict the important MBs for which encoder performs the ME to reduce the prediction error at the decoder-side. However, the highly-correlated MBs are left for the decoder, as decoder has a high probability to find a good approximation for such spatially and temporally correlated MBs, thus resulting in less number of parity bits.

**Hint-3:** The match of the MBs with high spatial and temporal correlation can be accurately predicted at the decoder-side without excessive decoder-side ME. MBs with high spatial and temporal correlation need not to be processed for encoder-side ME, thus using the encoder-side energy quota for more important MBs. The key challenge is to account for motion vector drift for ROI identification and ranking ROI-MBs.

**Summarizing the motivational analysis:** The key research challenges for reducing the energy consumption of HDVC are:

a) Adaptive energy quota distribution and control at various levels under dynamically varying scenarios;

b) Leveraging video content properties to enable efficient control on energy distribution and to achieve high energy reduction

c) Low-overhead Region of Interest (ROI) identification and ROI-based MB selection depending upon MB properties,

d) Motion-based ranking of MBs for efficient MB selection;

### 3. Hierarchical Control Scheme for Energy Quota Distribution

Fig. 8 illustrates the detailed operational flow of our hierarchical control scheme that adaptively distributes and controls the energy quota of HDVC encoder and decoder at various hierarchical levels, i.e. GOW, frame, and Macroblock (MB) levels. The inputs to our scheme are (i) offline energy analysis of HDVC ME at both encoder and decoder sides; (ii) user constraints (like required coding duration); and (iii) HDVC configurations (like resolution, frame rate, $S_{GOW}$). The output is the list of selected MBs with their allocated energy quota. The MB-level energy quota is used to obtain an appropriate ME configuration for the selected MBs. Afterwards, the consumed energy during encoding and decoding are monitored and fed back to the scheme along with the motion statistics (like motion vectors). Our scheme operates in four major steps:

**Step-1)** **GOW-Level Energy Quota Distribution and Control (Section 3.1):** First, user defined constraints and available energy from the battery level are used to determine the potential coding duration. If the potential coding duration is less than the user-defined constraint, the frame rate and GOW size are re-adjusted to fulfill the user constraints in the available battery level. This coding duration along with the offline ME energy analysis is used to derive the initial target energy quota for each GOW. Since the consumed energy of a GOW may differ from the allocated energy quota, the target of the upcoming GOW is recomputed using a feedback control mechanism. Our scheme employs a simple PID controller to control the target of the upcoming GOWs. In case of a deviation (i.e. a significant change between two I-frames, for instance, due to a scene cut), the initial target for a GOW is reset for the controller. The GOW-level energy quota is forwarded to the frame-level energy quota distributor.

**Step-2)** **Frame-Level Energy Quota Distribution and Control (Section 3.2):** At the frame level, the energy quota is distributed

---

**Fig. 7 Spatial and temporal correlation in “Foreman” sequence**

**Fig. 8 Detailed operational flow of our hierarchical control scheme for energy quota distribution**
among the I-frame and ‘m’ W-frames. The energy quota of the I-frame is computed by performing a history-based prediction using the energy consumption of ‘k’ previous I-frames. Afterwards, the energy quota of ‘m’ W-frames is obtained. This quota is distributed among ‘m’ W-frames considering the temporal distance of a given W-frame from the preceding I-frame and relative variance difference of two I-frames at the borders of the current GOW. The temporal distance based energy quota distribution accommodates the changing temporal correlation inside a given GOW due to the longer temporal distance. At the decoder side, the energy quota distribution primarily focuses on the bi-directional ME between two I-frames, which are used for motion interpolation for different W-frames.

Step-3) ROI Identification and Extrapolation (Section 3.3): In order to intelligently distribute the frame-level energy quota among different MBs of a frame, our scheme employs a ROI-based MB selection and ROI-driven energy quota distribution. The ROI is identified for I-frames considering the motion drift of a given MB w.r.t. its neighboring MBs. Since MBs in the object/ROI have high spatial and temporal correlation, our scheme selects the ROI-MB with high rank value, which is quantified as the motion vector drift. It is based on the analysis of Section 2, that decoder can generate the estimates for W-frames for highly-correlated MBs. Typically, MBs at the object boundaries are selected by our scheme. An ROI map containing the ROI-MBs sorted w.r.t. their rank values in a descending order (starting with the highest rank value). This facilitates generating a high-quality SI at the decoder side that leads to reduced energy at the encoder side. Note, our scheme does not waste energy for the homogeneous and slow-moving MBs, because decoder can estimate these MBs with a high probability. Instead the energy is spent on complex MBs with high texture and fast motion, such that the SI generation at the decoder side can be improved that leads to less parity bits and consequently low transmission energy. In order to reduce the ROI overhead, the I-frame ROI-MBs are extrapolated for W-frames. The extrapolation is considered by fitting the projected ROI-MB location to the nearest MB boundary.

Step-4) MB-Level Energy Quota Distribution and Control (Section 3.4): First the maximum number of selected MBs is computed based on the available frame-level energy quota and best case ME (i.e. ME configuration with the least energy). Since in a resource-constrained scenario, sufficient processing power may not be available to meet the throughput constraints, the number of selected MBs is readjusted. For the number of selected MBs, ROI-MBs from the ROI map are extracted starting from the highest rank value. The energy quota of each MB is computed and controlled in a feedback manner. Depending upon the allocated energy quota, an appropriate ME configuration is selected and ME is performed. The consumed energy is monitored. This consumed energy may differ from the allocated energy quota. Therefore, the target of the subsequent MB is adjusted by back-propagating the error between the consumed and target MB energy in a weighted fashion. The energy quota of each MB is computed and controlled in a weighted fashion. The total consumed energy is fed back for controlling the frame-level and GOW-level energy quota distribution.

In the following, we present detailed algorithms of different operational steps of our scheme.

### 3.1 GOW-Level Energy Quota Distribution & Control

Fig. 9 illustrates the pseudo-code for computing the GOW-level target energy quota for both encoder and decoder sides. Each GOW has ‘m’-SGOW–1 W-frames and ‘I’ I-frame (see Section 1.1). When the HDVC is initialized the encoder and decoder side target GOW energy quotas (E_TGOW(Enc,Dec)) are computed considering the available battery levels and user constraints; lines 2-13. First the potential encoding or decoding durations are computed considering the available battery levels (line 4; Eq. 1) and minimum of two is selected to determine the overall potential coding duration; line 5. Note, E_{Enc} and E_{Dec} are the average energy consumption of I-frames and W-frames that are obtained using an offline analysis for various test video sequences [33].

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>GOW_EnergyControl ( ): Input: Initialization Signal Init; Deviation Signal Dev; User defined duration Dc, Battery levels BL_{Dec}, BL_{Enc}, Frame rate FR; Size of GOW SGOW; Consumed GOW energy E_{GOW(Enc,Dec)}.</td>
</tr>
<tr>
<td>2.</td>
<td>if(Init OR Dev) then</td>
</tr>
<tr>
<td>3.</td>
<td>m ← SGOW - 1; SGOW' ← SGOW; FR' ← FR; sumEi ← 0;</td>
</tr>
<tr>
<td>4.</td>
<td>∀i(Enc,Dec) D; getDur(BL_i, m, FR, E_{avg}, E_{Wavg});</td>
</tr>
<tr>
<td>5.</td>
<td>D ← min(D_{Enc}, D_{Dec});</td>
</tr>
<tr>
<td>6.</td>
<td>D_{req} ← max(D, DU);</td>
</tr>
<tr>
<td>7.</td>
<td>if(D_{DU} &gt; D) then</td>
</tr>
<tr>
<td>8.</td>
<td>m ← SGOW' - 1;</td>
</tr>
<tr>
<td>9.</td>
<td>∀i(Enc,Dec) ETGI ← (BL_i × (m + 1))/D_{req} × FR);</td>
</tr>
<tr>
<td>10.</td>
<td>D_{req} ← max(D, DU);</td>
</tr>
<tr>
<td>11.</td>
<td>if(D_{DU} &gt; D) then</td>
</tr>
<tr>
<td>12.</td>
<td>EncDec GOWS ← GOWS;</td>
</tr>
<tr>
<td>13.</td>
<td>GOWGOWs ← ETGI;</td>
</tr>
<tr>
<td>14.</td>
<td>else</td>
</tr>
<tr>
<td>15.</td>
<td>GOWs ← GOWs;</td>
</tr>
<tr>
<td>16.</td>
<td>GOWs ← GOWs;</td>
</tr>
<tr>
<td>17.</td>
<td>sumEi ← sumEi + E_{avg};</td>
</tr>
<tr>
<td>18.</td>
<td>GOWGOWs ← ETGI + ((K_p × E_i) + (K_v × sumE)) + (K_D × ΔE));</td>
</tr>
<tr>
<td>19.</td>
<td>}</td>
</tr>
<tr>
<td>20.</td>
<td>return (E_{GOW(Enc,Dec)} × FR' × m, S_{GOW});</td>
</tr>
</tbody>
</table>

Fig. 9 Pseudo-code for computing and controlling the target energy quota at the GOW level for both encoder and decoder

If the potential coding duration is less than the user-defined coding duration constraint, the required duration is set (line 6) and the frame rate (FR) and GOW size (SGOW) are readjusted to accommodate the required coding duration in the available battery levels; lines 7-10. Afterwards, the target GOW energy quotas are computed; lines 11-12. Since the consumed energy of a GOW may differ from the target GOW energy quota, our scheme adapts the target GOW energy quotas in a feedback control mechanism; lines 14-19. The error between the consumed energy and target energy quota is computed after for each GOW. Our scheme employs a simple PID controller to control the target energy quota of the next GOW. K_p, K_v, and K_D are the proportional, integral, and derivative gains. K_p reduces the error, K_v eradicates the steady error effects, and K_D ameliorates the stability. These gains are computed using the Ziegler-Nichols Method [26] using the settings given in Eq. 2 (K_p = 0.8; T_c = 2).

\[
D = \frac{(BL_i × (m + 1))/(E_{avg} + m × E_{wavg}) × FR)}{(1}
\]

\[
K_p = 0.6 × K_p; K_v = K_v/(0.5 × T_c); K_D = K_D × (0.125 × T_c); (2)
\]

Line 2 shows that in case of a deviation (Dev; line 2), the target GOW quota is reset and forwarded to the controller. The (Dev) is denoted as a significant change between two I-frames, such that the decoder cannot create a high quality estimate of the W-frames. It is denoted by a scene cut or significant texture difference (\( \xi_{fr} - \xi_{fr+1} \)) between two consecutive I-frames \( IF_i \) and \( IF_{i+1} \). The deviation signal is computed using Eq. 3, where \( \xi_{fr} \) denotes the texture threshold.

\[
Dev = \begin{cases} 1; & \text{if} \{ \text{sceneCut} \} \ OR \ (\xi_{fr} - \xi_{fr+1}) \ > \ \xi_{th} \} \\ 0; & \text{Otherwise} \end{cases} 
\]
3.2 Frame-Level Energy Quota Distribution and Control

The GOW-level energy quota is distributed among the I-frames (E_{IF}) and the W-frames (E_{WTF}) depending upon S_{GOW}.

Encoder-Side Energy Quota Distribution: The energy quota is distributed to the I-frame for intra encoding and ROI identification (E_{IF}=E_{IF}+E_{ROI}), and to the W-frames for selective ME, transmission, channel coding, and ROI extrapolation: \( E_{WTF}=E_{WTF}(d)+E_{CC}+E_{ROI} \).

E_{IF} Prediction: The energy quota for the I-frame in the \( i^{th} \) GOW is allocated based on the consumed energy of the previous ‘\( k \)’ I-frames (E_{ICF}) using Eq. 4. In order to avoid noise in the energy consumption median operator is used for ‘\( k-1 \)’ I-frames in the history, while the immediately previous I-frame’s energy is used to compensate for the sudden error effects.

\[
E_{IF} = \text{pred} \left( E_{ICF(i-1)} \right) + \frac{E_{ICF(i-2)} + \cdots + E_{ICF(i-k)}}{2}
\] (4)

E_{WTF} Prediction: The total target energy quota for ‘\( m \)’ W-frames (\( E_{WTF}=E_{TGOW}-E_{IF} \)) is computed and distributed among individual W-frames considering their temporal distance (see Eq. 5).

\[
E_{WTF} = \frac{1 - \nu}{m - (j-1)} + \sum_{k=1}^{\nu} E_{WTF} - \frac{E_{WTF} - E_{CFWk}}{m-1}
\] (5)

such that, \( \nu = (1 - \min\left(\nu_{IF}, \nu_{IF-1}\right) / \max\left(\nu_{IF}, \nu_{IF-1}\right) \)\)

The amount of temporal correlation for a given W-frame w.r.t. its preceding I-frame decreases depending upon its temporal distance. It is due to the fact that objects in farther video frames have more temporal distance that may not be captured by ME under a given search range and energy quota constraints. Therefore, extensive search is required for farther W-frames, and less motion search may be sufficient for the closer W-frames. To facilitate this, our scheme adaptively distributes the \( E_{WTF} \) energy quota among W-frames depending upon their temporal distance from the preceding I-frame and variance difference (\( \nu \)) between the bordering I-frames; see Eq. 5. The variance difference \( \nu \) of two consecutive I-frames hints towards the quality of the temporal correlation. Depending upon the variance difference, a part of \( E_{WTF} \) (first addend without \( \nu \) in Eq. 5) is equally distributed in order to keep the fairness of allocation in order to compensate for the energy consumption other than ME computations, like \( E_{IF}, E_{CC} \) and \( E_{ROI} \). However, the other part (second addend with \( \nu \) in Eq. 5) is distributed based on the temporal distance to compensate for the temporal correlation changes.

After the encoding of each W-frame, the energy quota of the next W-frame is controlled depending upon the consumed energy of the previous W-frames (E_{CFW}); see Eq. 5. For instance, for \( m=4 \), the energy quota of WF_1, WF_2, WF_3, and WF_4 can be computed as follows:

\[
E_{WTF1} = \left(1 - \nu\right) / 4 + \nu / 10 \times (E_{WTF1})
\]

\[
E_{WTF2} = \left(1 - \nu\right) / 3 + \nu / 6 \times (E_{WTF1} - E_{CFW1})
\]

\[
E_{WTF3} = \left(1 - \nu\right) / 2 + \nu / 3 \times (E_{WTF1} - E_{CFW1} + E_{CFW2})
\]

\[
E_{WTF4} = (E_{WTF1} - (E_{CFW1} + E_{CFW2} + E_{CFW3}))
\]

Decoder-Side Energy Quota Distribution: At the decoder side, the energy quota is mainly distributed among the key I-frames in order to perform the bi-directional ME. For W-frames an average quota based on the history is reserved considering the energy consumption for SI generation/interpolation and Slepian-Wolf decoder. The size of history is ‘\( l \)’ previous GOWs each having ‘\( m \)’ W-frames; see Eq. 6. The remaining quota of the \( i^{th} \) GOW is allocated to the decoding of I-frames \( E_{ICF} \) and bi-directional ME \( E_{ME} \); Eq. 7.

\[
E_{WTFi} = m \times \left\{ \begin{array}{ll} A & \text{if } y \in \{1, \ldots, L\} \\ E_{CFW(i-1)}, & \text{otherwise} \end{array} \right\}
\] (6)

\[
E_{WTFi} = E_{TGOW} - E_{IF} - E_{ME} = \left( E_{IF} - E_{CD} \right) / 2
\] (7)

3.3 ROI Identification and Extrapolation

As previously discussed, HDVC encoder-side ME of selective MBs in ROI may lead to significant energy reduction due to improved SI and parity bits reduction. Typically, the decoder cannot perform a good estimation for the MBs with high motion and/or variance, especially at the object boundaries. Therefore, it is of key importance in HDVC to send energy at the encoder side to perform ME of the ROI-MBs. Note, if the boundary of the moving region is extracted, the HDVC decoder can exploit the correlation of MBs belonging to the same object to predict the other MBs of the moving objects/regions. Therefore, the key focus of our scheme is to accurately extract the ROI-MBs at the object boundaries (see Fig. 10). Our scheme exhibits an integrated motion-based algorithm for ROI identification and ROI-MB selection; see Fig. 11. For a low-overhead scheme, the ROI map is generated only at key frames and extrapolated for all the W-frames see Fig. 12. The ROI map is regenerated when decoder requests the ROI re-computation due to significant degradation in the SI-quality as a result of accumulation of prediction error; Fig. 12. For each frame, the extrapolation is done considering the ROI in the immediately preceding frame (see later in this section).

![Fig. 10 ROI identification at an I-frame using MVD and extrapolation for 2 W-frames of foreman sequence](image)

<table>
<thead>
<tr>
<th>MVD</th>
<th>Accumulator</th>
<th>List</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>(16,16), (16,32), (48,208)</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>(32, 48), (96, 128)</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>(96,96), (352, 64), (208, 80), (128, 154)</td>
</tr>
</tbody>
</table>

Table I: An example hash table used as an ROI-Map

The ROI-MBs are identified based on the Motion Vector Drift (MVD; Eq. 8) of given MBs w.r.t. their neighboring MBs; see lines 2-5. Note, this MVD is also denoted as “rank” value to perform ROI-based MB selection at the encoder side. Higher is the MVD of a MB, higher the rank it gets, i.e. more energy needs to be spent for this MB to perform extensive ME in order to obtain a good match. Note, MVD already denotes an inverse function of correlation in the neighborhood. Therefore, selecting highest rank MBs as ROI-MBs complies with the motivational study of Section 2, where we discussed that the HDVC decoder can regenerate estimate the highly-correlated MBs (i.e. low-ranked) without requesting a large number of parity bits from the HDVC encoder.

\[
\overline{mvd} = \sum_{i=1}^{L} \sum_{j=0}^{l-1} \left| \overline{mv(x,y)} - \overline{mv(x+i,y+j)} \right| \times SAD(x,y)
\] (8)

(\(x,y\)) are the location coordinates of a given MB and \( SAD(x,y) \) is the best matching cost in terms of Sum of Absolute Difference (SAD) of the current MB w.r.t. the best matching MB in the previous
frame. MVD value ensures that only the MBs at the high MV gradient and with high motion cost are included in the ROI. This means that for a moving object, only the MBs at the object boundaries are selected as ROI-MBs. The MVD acts as an index to a hash table, which stores the accumulated number of MBs having the same MVD and their coordinates; see Table I. This list is referred as the ROI map which is *rank-wise sorted* in a descending order, i.e., starting from the ROI-MB with the highest rank value. An example ROI identification procedure is shown in Fig. 10 for the “Foreman” test video sequence, where only the boundaries of the moving regions are detected.

1. **GenerateROI-MAP**(*Input*: Previous reconstructed I-frame \(I_F\), Motion Vector and SAD arrays from ME \(mv, SAD\), Current Frame \(CF\); *Output*: rank-wise sorted ROI map \(ROI\))

2. \(\forall MB \in CF\) \{ 
3. \(mv(d, y) \leftarrow ComputeMVD(mv, SAD)\); 
4. \(IncrementHashTable(mv(d, y))\); 
5. \(AddtoList(mv(d, y), (x, y))\); 
6. \(counter \leftarrow 0\); 
7. \(\forall i \in \text{Last index of HashTable} \rightarrow 0\) \{ 
8. \(counter \leftarrow counter + \text{HashTable}(i)\); 
9. \(\text{if} (counter \geq T_h) \{\lambda \leftarrow i; \text{break}\}\); 
10. \} 
11. ROI \(\leftarrow \text{getMBList}(\lambda, \text{last});\) 
12. \(\text{return}(ROI);\)

**Fig. 11** Pseudo-code for extracting the ROI of a frame

**Extrapolation of the ROI Map for W-frames:** First, the reflected location vector (\(\text{reflectLoc}\)) is computed by subtracting the motion vector of an ROI-MB to obtain a potential extrapolation of its motion. Afterwards, a fitting to the MB boundary is performed by comparing the ROI-MB with extrapolated motion to the neighboring MB boundaries, as shown in Fig. 13. The vector \(r\) is computed to locate the nearest MB boundary using Eq. 9. *Minimum* denotes a function that returns the vector corresponding to the minimum magnitude out of \(r_1, r_2, r_3, r_4\). The \(r\) vector is added in the reflected location to determine the extrapolated ROI-MB.

\[
\text{reflectLoc} = \text{ROIMB}_{(F_{t+1})} - m_{\text{VMB}(F_t)} \\
\bar{r} = \text{Minimum}(r_1, r_2, r_3, r_4) \quad (9)
\]

\[
\text{ROIMB}_{(F_{t+1})} = \text{reflectLoc} + \bar{r}
\]

**Fig. 12** ROI identification and extrapolation within a sequence

**3.4 MB-Level Energy Quota Distribution and Control**

Fig. 14 shows the ROI-driven MB selection and energy distribution. The input is the frame-level energy quota of the W-frame and the ROI map sorted w.r.t. the rank values of the MBs in a descending order. First, the maximum number of selected MBs \(N_{\text{Emax}}\) that need to be considered for encoder-side ME is computed using the frame-level energy quota and the energy required for transmission, channel coding, and ROI extrapolation; line 2. Note, the transmission energy is a function of the transmission distance ‘\(d\)’ [34]. Since in several cases, the underlying platform may not be able to provide the required throughput due to limited computational resources and/or limited processing power. Therefore, the maximum number of MBs \(N'\) that can be processed by the underlying platform are computed in line 3. The maximum number of selected MBs \(N'\) is determined as the minimum of \(N_{\text{Emax}}\) and \(N_{\text{Fmax}}\); line 4. Afterwards, \(N'\) MBs are extracted from the ROI map starting from the highest rank value; lines 5-7. If \(N\) is greater than the number of MBs in the ROI, additional MBs are extracted from the non-ROI map; line 8-9.

1. **ROI-DrivenMBEnergyDistribution()**: \(*Input*: Hardware throughput constraint \(T_{\text{HW}}\); rank-wise sorted map of ROI \(ROI\); List of MBs outside the ROI \(\text{ROI} nonROI\); Energy quota of the current frame \(E_{WF}\); Energy for transmission for a given transmission distance ‘\(d\)’ \(E_T(d)\); Energy of channel coding \(E_{CC}\); Energy of extrapolating ROI \(E_{ROIex}\); \(*Output*: Consumed Energy of W-frame, \(E_{WF}\)**

2. \(N_{\text{Emax}} \leftarrow \frac{(E_{WF} - E_{\text{Tx}(d)} - E_{CC} - E_{ROIex})}{E_{WF}} \), \(\text{getAvgMBEnergy}()\)

3. \(N_{\text{Fmax}} \leftarrow \frac{T_{\text{HW}}}{FR} \); 
4. \(N \leftarrow \min(N_{\text{Emax}}, N_{\text{Fmax}})\); 
5. \(N_{ROI} \leftarrow \text{getNumMBs}(ROI)\); 
6. \(\text{if}(N = N_{ROI}) \{ \text{ListMBs} \leftarrow ROI;\) 
7. \(\text{if}(N < N_{ROI}) \{ \text{ListMBs} \leftarrow \text{getMBs}(ROI, N);\) 
8. \(\text{if}(N > N_{ROI}) \{ \text{ListMBs} \leftarrow \text{getMBs}(ROI, N) + \text{getMBs}(\text{ROI nonROI});\) 
9. \(\Delta E \leftarrow (E_{WF} - E_{\text{Res} - E_{\text{Tx}(d)} - E_{CC} - E_{ROI})/N;\) 
10. \(\text{ECWF} \leftarrow 0; \text{ETM} \leftarrow \Delta E;\) 
11. \(\forall \text{mbListMBs} \{ // \text{loop over all MBs in the list of selected MBs}\) 
12. \(\text{getAvgMBEnergy}()\) \(\text{getMEconfig}()\); 
13. \(\text{confME} \leftarrow \text{getMEconfig}()\); 
14. \(\text{EMB} \leftarrow \text{performME}()\); 
15. \(\text{ECWF} \leftarrow \text{ECWF} + \text{EMB};\) 
16. \(\Delta E \leftarrow \text{ET} - \text{EMB};\) 
17. \(\text{ETM} \leftarrow \text{ET} + \alpha \times \Delta E;\) 
18. \} 
19. \(\text{return}(\text{ECWF});\)

**Fig. 14** Pseudo-code for ROI-driven MB selection and MB-level energy distribution and control

The target energy quota for one MB is computed depending upon the number of selected MBs; line 10. Afterwards, all the MBs in the selected MB list are processed for ME starting from the highest rank MBs first; lines 12-18. An appropriate ME configuration is selected considering various ME configurations that provide energy vs. quality tradeoffs; line 13. Multiple ME configurations can be formulated depending upon the search window size, search pattern types, termination rules, initial search point prediction, etc. Note, ME configuration is not the contribution of this paper. In this paper, we deployed the ME configuration technique of [11] since we consider it crucial to deploy a state-of-the-art ME configuration technique to evaluate the quality of our work in a fair way.

After the ME is completed for the MB, the consumed energy \(E_{\text{MB}}\) is monitored; line 14. The consumed energy can be different
from the target energy or the average energy of the selected ME configuration. Therefore, we adapt the energy quota of the MB and the average energy value of the ME configuration in order to have a better ME configuration selection for the subsequent MBs. Depending upon the error between the target and the consumed energies (see line 16), the target of the next MB is readjusted by back-propagating the energy error; line 17. The weighting factor $\alpha$ controls the strength of back-propagation. The total consumed energy of the W-frame is returned to control the frame-level energy quota (line 15, 19); see Section 3.2.

Note, the decoder-side MB-level energy distribution and ranking are very similar to that at the encoder side, except that the energy quota for the I-frame includes the energy for forward and backward ME.

4. RESULTS AND EVALUATION

4.1 Experimental Setup

In order to evaluate the energy efficiency of our scheme, we have developed a complete HDVC system with 5 different ME configuration classes at encoder and 2 ME configuration classes at decoder. The configuration of ME classes are borrowed from the work of [11]. The Intra-frame encoder for I-frame coding is based on H.264/AVC standard. For W-frame coding and estimation, the algorithms for side information, parity-bits generation, and the quality matrix Q4 are based on [35]. For transmission energy estimation, the model of [34] is employed. For transmission energy estimation, the distance of 100 meters is assumed. Based on this model, the energy for one bit transmission is given as $\alpha = 1 \mu J$/bit. The computation energy results are obtained from the synthesis results using a 90 nm technology.

We have compared our ROI-based energy quota distribution and control scheme with Raster-scan based scheme of [9]. For fairness of comparison, we have employed the same I-frame coding algorithm, same ME algorithm, and same SI generation unit. The coding is performed under the similar target PSNR constraints. The energy results illustrate the benefit of our energy distribution and control scheme and ROI-based decisions. Note, the overhead of our scheme is included in the energy results. Detailed discussion on the overhead results can be found in Section 4.4.

4.2 Comparison with State-of-the-Art

Experiments in Fig. 15 illustrate the energy consumption comparison of our scheme with state-of-the-art for different video sequences. Fig. 15 shows that our scheme provides on average 20% energy reduction (maximum 25%) compared to state-of-the-art. Higher savings are obtained for “Coastguard”, “Foreman” and “Hall” sequences, as they exhibit more motion. Note, the savings for “Mother & Daughter” and “News” sequences are relatively less due to limited motion content in the sequences. In such scenarios, adaptive ME already terminates earlier as it quickly determines the best match. This result also demonstrates that in case of high-motion and high-texture, our scheme exhibits a higher potential for energy savings compared to state-of-the-art.

Fig. 16 shows detailed energy consumption comparison between our scheme and state-of-the-art at MB-level for “Foreman” sequence. It is noteworthy in Fig. 16 that our scheme has allocated more energy quota to the Foreman face as it is detected as ROI. However, due to the raster scan nature, state-of-the-art only captures the moving hat of the “Foreman” object, while the significant energy quota is wasted on the background wall MBs (marked with circles), which could have been easily generated by the HDVC decoder. As a result, our scheme also provides PSNR improvement even under same quality constraints, because the decoder has a much better match of the “Foreman” face in the case of our scheme compared to that of the state-of-the-art. This implies that to achieve the quality comparable to that of our scheme, state-of-the-art will have to spend even more energy. However, due to the video dynamics, only a controlled rate based encoding is possible and a controlled quality based encoding is typically not possible.

Fig. 17 depicts the quality (PSNR) comparison on frame-level between the two schemes for the cases of same transmission energy. For the frames at labels A and B, the deviation between the two key-frames is high, thus resulting in a higher variance between the two frames. Our scheme allocates more energy to the frame farther away from the key-frame A and thus provides better PSNR compared to state-of-the-art. The raster-scan based scheme fails to allocate energy quota to important blocks, thus the results in PSNR loss. Our scheme achieves 2 dB higher PSNR in these cases. For low variance key-frames (i.e. frames at labels C and D in Fig. 17), the energy distribution among the non-key frames is nearly constant and therefore, the PSNR difference between C and D is less between the two schemes. Note, both schemes achieve the same quality at the I-frames, as we provide the same I-frame coding algorithms to both schemes and the difference only lies in the energy quota distribution and control for W-frames. Therefore, the PSNR for I-frames is same for both schemes, while PSNR difference only occurs for W-frames in this case. This is to ensure a high
4.3 Detailed Experimental Evaluation

Fig. 18 and Fig. 19 illustrate the detailed observations for quality and energy consumption when using our scheme. Fig. 18 shows that our scheme tends to increase the relative quality if the value of m increases, where m is the number of W-frames between two I-frames. This observation is shown for “Foreman” sequence encoded with QP=18 and QP=36 for the I-frames. The later W-frames in the GOW are usually of lower quality compared to the earlier W-frames because of the increased temporal de-correlation with the I-frames. Since our approach allocates these later W-frames a higher energy quota, a better estimate (or match) of the current MB in I-frames can be found. This leads to a quality increase even for these frames, which will not be possible when allocating energy without consideration of temporal distance, i.e. equally allocating energy to all W-frames.

Fig. 19 shows the energy quota adaptation at the (a) frame-level; and (b) GOW-level (see Section 3.1 and 3.2 for algorithms). Fig. 19 shows that our scheme keeps the energy quota distribution close to the target energy; thus leading to a reduced controller error. An interesting observation in Fig. 19b is that the error between the target and consumed energy at the GOW-level is almost insignificant. This is due to the hierarchical control of the target energy quota which minimizes the risk of target energy quota violations. Fig. 19a shows that over the period, the controller error (i.e. difference between the target and consumed energy) is reducing. This is due to the adaptation of the controller output. Note that in Fig. 19a, results for only W-frames are shown, because the variance in the energy consumption for I-frames is insignificant and are adapted using a different equation; Eq. 4.

![Graph showing energy consumption comparison](image1)

Fig. 18 Difference in PSNR for the proposed ROI and raster-scan approaches at (a) m=1 and (b) m=4.

![Graph showing energy consumption comparison](image2)

Fig. 19 (a) Proposed and consumed energy per frame (b) proposed and consumed energy per GOW.

4.4 Overhead

Table II shows the performance, area and energy overhead of the proposed scheme after synthesis and place&route results for a 90nm technology. The implementation includes one divider and three multipliers, which are shared among the different calculation steps. For the GOW-, Frame- and MB-Level calculations 4cycles/23cycles (depending on init/dev in Fig. 9), 2+2+m+4cycles and 6+3*NROI cycles are required, respectively (see parameters in the algorithm shown in Fig. 14). Table II shows that the energy overhead is insignificant compared to the savings of our scheme.

Moreover, the ROI extrapolation requires only 2 additions and a search for minimum magnitude vector within 4 vectors. This is a negligible overhead compared to the encoding process and our energy savings.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Latency [Cycles]</th>
<th>Area [GE]</th>
<th>Energy [nWs]</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOW-Level</td>
<td>5/23</td>
<td>11,103</td>
<td>18.4</td>
</tr>
<tr>
<td>Frame-Level</td>
<td>2+12</td>
<td>8,327</td>
<td>11.1</td>
</tr>
<tr>
<td>MB-Level</td>
<td>6+360</td>
<td>14,238</td>
<td>292.2</td>
</tr>
<tr>
<td>Total</td>
<td>1501</td>
<td>62,638</td>
<td>1211.2</td>
</tr>
</tbody>
</table>

Table II: Performance, area and energy overhead of our scheme (m=4, k=5, NROI=120)

5. CONCLUSION

We have presented a hierarchical control scheme for adaptive energy quota distribution and control for HDVC. Our scheme controls the distribution of energy quota at various levels namely: (i) GOW-level; (ii) Frame-level; and (iii) MB-level. To achieve high energy distribution efficiency, our scheme exploits the video content properties and region of interest. An intelligent MB selection for ME leads to reduced transmission energy. Overall, our scheme provides on average 20% energy reduction compared to state-of-the-art energy management scheme for HDVC. Consideration of ROI and video content properties in the HDVC paradigm facilitates a high-quality SI generation at the decoder-side in addition to efficiently distributing energy at both encoder and decoder sides. Due to its high energy efficiency and adaptivity to provide higher quality for ROIs, our scheme enables HDVC in resource- and energy-constrained devices subjected to dynamically varying scenarios, involving complex motion and longer GOW lengths.

6. ACKNOWLEDGMENT

We would like to thank Mr. Florian Kriebel, Karlsruhe Institute of Technology (KIT), for his contribution to the hardware implementation.
7. REFERENCES