

Energy-Efficient Multi-Codec Bitrate-Ladder Estimation for Adaptive Video Streaming

Vignesh V Menon^{*†}, Reza Farahani[†], Prajit T Rajendran[‡],
Samira Afzal[†], Klaus Schoeffmann[†], Christian Timmerer[†]

^{*} Video Communication and Applications Department, Fraunhofer HHI, Berlin, Germany

[†] Alpen-Adria-Universität, Klagenfurt, Austria

[‡] Universite Paris-Saclay, CEA, List, F-91120, Palaiseau, France

Abstract—With the emergence of multiple modern video codecs, streaming service providers are forced to encode, store, and transmit bitrate ladders of multiple codecs separately, consequently suffering from additional energy costs for encoding, storage, and transmission. To tackle this issue, we introduce an online energy-efficient Multi-Codec Bitrate ladder Estimation scheme (MCBE) for adaptive video streaming applications. In MCBE, quality representations within the bitrate ladder of new-generation codecs (e.g., *High Efficiency Video Coding* (HEVC), *Alliance for Open Media Video 1* (AV1)) that lie below the predicted rate-distortion curve of the *Advanced Video Coding* (AVC) codec are removed. Moreover, perceptual redundancy between representations of the bitrate ladders of the considered codecs is also minimized based on a *Just Noticeable Difference* (JND) threshold. Therefore, random forest-based models predict the VMAF score of bitrate ladder representations of each codec. In a live streaming session where all clients support the decoding of AVC, HEVC, and AV1, MCBE achieves impressive results, reducing cumulative encoding energy by 56.45%, storage energy usage by 94.99%, and transmission energy usage by 77.61% (considering a JND of six VMAF points). These energy reductions are in comparison to a baseline bitrate ladder encoding based on current industry practice.

Index Terms—HTTP Adaptive Streaming; Multi-Codec Streaming; Per-Title Encoding; Energy-Aware Streaming; Just Noticeable Difference.

I. INTRODUCTION

The emergence of novel video formats and standards has facilitated content delivery across various platforms and devices. *HTTP Adaptive Streaming* (HAS) delivery systems, such as those based on the *MPEG Dynamic Adaptive Streaming over HTTP* (DASH) [1] standard or *Apple HTTP Live Streaming* (HLS) [2], have emerged as the dominant technologies utilized by service providers to deliver live video content [3], [4]. In such systems, each codec requires its own set of representations, i.e., bitrate ladders [5], [6]. For example, *Advanced Video Coding* (AVC) [7] and *High Efficiency Video Coding* (HEVC) [8] have distinct bitrate ladders. Initially, streaming services used AVC for wider device compatibility [9]. However, as newer devices with HEVC and *Alliance for Open Media Video 1* (AV1) [10] support becomes prevalent, HEVC and AV1-encoded bitrate ladder representations are introduced. Recent years have developed new formats such as *Versatile Video Coding* (VVC) [11], *Essential Video Coding* (EVC) [12], and *Low Complexity Enhancement Video Coding* (LCEVC) [13]. Over time, streaming systems have evolved

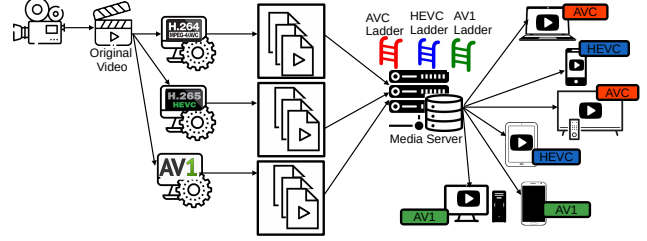


Fig. 1: An example of a multi-codec streaming system.

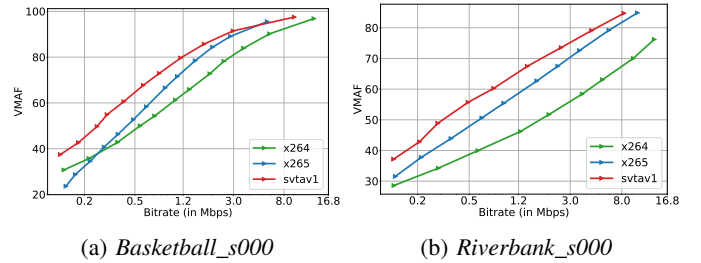


Fig. 2: Rate-distortion (RD) curves of representative sequences of VCD dataset [15], encoded with JPTS bitrate ladder [16] for x264 [17], x265 [18], and svtav1 [19] encoders.

to accommodate multiple codecs, with older devices relying solely on AVC, some newer devices using HEVC streams, and certain devices supporting both AVC and HEVC, including seamlessly switching between them [9] (cf. Fig. 1). Handling such multi-codec deployments requires generating ABR bitrate ladders of each codec separately, considering the range of codecs to be supported by the receiving device population based on their decoding capabilities [14].

The estimation of a multi-codec bitrate ladder, as proposed in this paper, is based on the fact that, in some cases, the compression efficiency of AVC is better than new-generation video codecs, i.e., at low bitrates [20], [21]. Furthermore, the compression efficiency of codecs saturates at very high target bitrates, as they become similar to lossless coding. The regions where each codec performs better than others depend on the *complexity of the video content* [22]. An example is shown in Fig. 2 where the cross-over bitrate between the quality of x264 [17] and x265 [18] is at approximately $b_1 = 0.3$ Mbps for *Basketball_s000*, while the cross-over bitrate between the quality of x265 and svtav1 is at approximately $b_2 = 5.0$ Mbps. This means that, at bitrates lower than b_1 , x264 outperforms

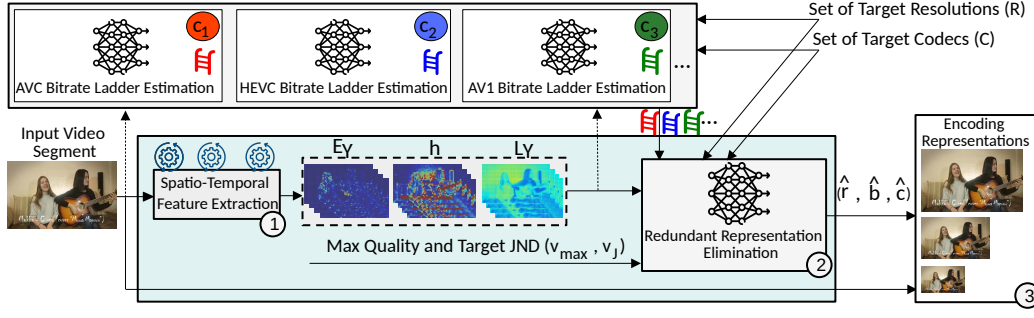


Fig. 3: Online encoding using MCBE envisioned in this paper for adaptive video streaming.

x265, while at bitrates higher than b_2 , x265 outperforms svtav1. On the other hand, for *Riverbank_s000*, svtav1 remains superior throughout the bitrate range. This is because old-generation codecs may excel in scenarios where the content does not leverage the advanced coding tools and techniques introduced by the new-generation codecs. x265 encoding of *Basketball_s000* at bitrates lower than b_1 can be eliminated, as clients can be served with x264 representations (due to comprehensive support of AVC).

Encoding video content into multiple representations in various bitrate-resolution pairs for each codec results in substantial computational workload and energy consumption [23]. Additionally, the storage and transmission of these representations further contribute to the overall energy consumption [24]. When unnecessary high-bitrate representations (of new-generation codecs) are eliminated, the energy consumption of the streaming system is significantly reduced [25]. This is because the energy consumption of AVC is significantly lower than that of new-generation video codecs [26], [27]. As video streaming continues to grow in popularity and usage, finding energy-efficient solutions to optimize the multi-codec bitrate ladder becomes crucial to mitigate the environmental impact and reduce operational costs for streaming service providers [28].

In this light, this paper proposes an online Multi-Codec Bitrate Ladder Estimation (MCBE) scheme for adaptive video streaming applications. A lightweight algorithm is proposed to *eliminate redundant representations of the bitrate ladders of new-generation video codecs*, based on their predicted perceptual quality. Therefore, random forest models [29] are trained to estimate the VMAF score of each representation based on low-complexity spatio-temporal features of the input video segment. Note that other quality metrics can be envisioned, which are subject to our future work. When AVC performs better (in terms of perceptual quality) than or is identical as HEVC and/or another new-generation codec (e.g., AV1) in a bitrate range, the corresponding new-generation codec representations are eliminated from the bitrate ladder. This is because clients can be served with AVC representations with better RD performance. Moreover, the bitrate ladder representations with a perceptual quality difference within a given *Just Noticeable Difference* (JND) [30] threshold are eliminated. Finally, it is worth noting that MCBE can be used

in conjunction with state-of-the-art bitrate-ladder prediction schemes [16], [31]–[33]).

II. MCBE ARCHITECTURE

Adaptive video streaming systems often use bitrate-ladder prediction methods to enhance the Quality of Experience (QoE) for users [32], [34], [35]. The architecture of the proposed MCBE scheme is shown in Fig. 3. MCBE receives input bitrate ladders for each codec, e.g., c_1 , c_2 , and c_3 for AVC, HEVC, and AV1 codecs, respectively. Other codecs may be envisioned as part of future work but are supported by the current architecture. It extracts DCT-energy-based features and eliminates redundant representations based on the predicted quality metric (i.e., VMAF in this paper) of each representation. MCBE comprises three phases (cf. Fig. 3):

- ① Spatio-temporal feature extraction (Section II-A)
- ② Redundant representation elimination (Section II-B)
- ③ Encoding of the segments using the selected bitrate-resolution pairs of each codec

A. Spatio-Temporal Feature Extraction

MCBE uses the following DCT-energy-based features [36], extracted using open-source VCA v2.0 video complexity analyzer [22] for every segment:

- 1) Average luma texture energy (E_Y)
- 2) Average gradient of the luma texture energy (h)
- 3) Average luminescence (L_Y)

B. Redundant Representation Elimination

In this paper, the VMAF score $v_{r_t, b_t, c}$ of the t^{th} representation of the codec c is modeled as a function of the video content complexity features and the target representation (i.e., resolution r_t and bitrate b_t) [37], [38], as shown in the following equation:

$$v_{r_t, b_t, c} = f_V(E_Y, h, L_Y, r_t, b_t, c) \quad (1)$$

Random forest models [29] which are hyperparameter-tuned with the parameters $min_samples_leaf=1$, $min_samples_split=2$, $n_estimators=100$, and $max_depth=14$ are trained for each codec $c \in \mathcal{C}$ and resolution $r \in \mathcal{R}$ to predict VMAF. Input to the model for each codec-resolution are $[E_Y, h, L_Y, b]$. The pseudo-code of the *redundant representation elimination* method is shown in Algorithm 1. This algorithm consists of two primary steps as follows:

Algorithm 1: Redundant representation elimination.

Inputs:

M : number of supported codecs
 \mathcal{C} : set of all codecs c_1, c_2, \dots, c_M in order of priority
 N_c : number of representations for codec c
 $(\hat{r}_t, \hat{b}_t, c)$ pairs $\forall c \in \mathcal{C}, t \in N_c$
 v_J : target JND

Output: Q : Set of selected representations

Step 1:

```
for each  $c \in \mathcal{C}$  do
   $t = 2$ 
  while  $t \leq N_c$  do
    if  $\hat{v}_{c, \hat{r}_t, \hat{b}_t} > v_{\max}$  or  $\hat{v}_{c, \hat{r}_t, \hat{b}_t} - \hat{v}_{c, \hat{r}_{t-1}, \hat{b}_{t-1}} < v_J$ 
    then
      Eliminate  $(\hat{r}_t, \hat{b}_t, c)$  from the ladder
       $\hat{N}_c = N_c - 1$ 
     $t = t + 1$ 
```

Step 2:

```
 $Q = \{(\hat{r}_t, \hat{b}_t, c_1)\}, t \in \hat{N}_{c_1}$ 
for each  $c \in \{c_2, \dots, c_M\}$  do
  for each  $t \in \hat{N}_c$  do
     $(\tilde{r}_i, \tilde{b}_i, c_1) \leftarrow \arg \min_i | \hat{b}_{i, c_1} - \hat{b}_t | \text{ s.t. } \hat{b}_t \geq \hat{b}_{i, c_1}$ 
     $(\tilde{r}_j, \tilde{b}_j, c_1) \leftarrow \arg \min_j | \hat{b}_{j, c_1} - \hat{b}_t | \text{ s.t. } \hat{b}_t \leq \hat{b}_{i, c_1}$ 
    RD curve  $L$  between  $(\tilde{r}_i, \tilde{b}_i, c_1)$  and  $(\tilde{r}_j, \tilde{b}_j, c_1)$ :
    
$$v = \frac{\hat{v}_{c_1, \tilde{r}_j, \tilde{b}_j} - \hat{v}_{c_1, \tilde{r}_i, \tilde{b}_i}}{\tilde{b}_j - \tilde{b}_i} \cdot (\hat{b}_t - \tilde{b}_i) + \hat{v}_{c_1, \tilde{r}_i, \tilde{b}_i}$$

    if  $(\hat{v}_{c, \hat{r}_t, \hat{b}_t} \text{ is above } L)$  then
      Add  $(\hat{r}_t, \hat{b}_t, c)$  to  $Q$ .
```

Step 1: In practice, it is often observed that the VMAF scores of different representations are highly similar, leading to perceptual redundancy in the bitrate ladder. Consequently, this redundancy implies a wastage of energy during the encoding, storage, and transmission of data, without any improvement in QoE. To minimize this perceptual redundancy, MCBE leverages the concept of the JND threshold, which represents the minimum threshold at which the human eye can perceive differences in quality [39]–[41]. A fixed JND threshold denoted as v_J is input from the streaming service provider. If the VMAF difference between two representations is lower than v_J , the higher bitrate representation among them is eliminated. Furthermore, when the predicted VMAF is greater than the maximum VMAF above which the representation is deemed perceptually lossless (v_{\max}), the corresponding representation is eliminated from the bitrate ladder [42]. This way, MCBE lowers the overall energy requirement for encoding.

Step 2: RD points (based on the bitrates predicted by the bitrate ladder estimators (cf. Fig. 3) and the corresponding predicted VMAF scores) of each representation of new-generation codecs are geometrically compared to the predicted RD curve of the previous generation codec. The representation is eliminated if the point is below the RD curve of the previous generation codec.

In the final phase, the encoding process is performed exclusively for the selected bitrate-resolution combinations (\hat{b}, \hat{r}) of each codec (\hat{c}) for every video segment.

III. EXPERIMENTAL RESULTS

A. Test Setup

In this paper, 400 sequences (80% of the sequences) from the Video Complexity Dataset [15] are used as the training dataset, and the remaining (i.e., 20%) are used as the test dataset. The sequences are encoded at 30fps with the fastest encoding preset supported by the considered encoders on a dual-processor server with Intel Xeon Gold 5218R (80 cores, frequency at 2.10 GHz) with $\mathcal{C} = \{\text{x264 v1.1, x265 v3.5, svtav1 v1.6}\}$. VCA and the encoders specified in \mathcal{C} are run using eight CPU threads with x86 SIMD optimization [44]. The resolutions specified in the Apple HLS authoring specifications [43] are considered in the evaluation, i.e., $\mathcal{R} = \{360\text{p, } 432\text{p, } 540\text{p, } 720\text{p, } 1080\text{p, } 1440\text{p, } 2160\text{p}\}$. In all experiments, the average target JND function (v_J) is considered as two [45], four, and six [46] based on current industry practices. Accordingly, v_{\max} is set as 98, 96, and 94, respectively. This paper uses the following state-of-the-art encoding bitrate ladder prediction schemes in conjunction with MCBE:

- 1) Default HLS bitrate ladder [43] for each codec/encoder.
- 2) OPTE [33], where optimized resolutions for the set of bitrates in the HLS bitrate ladder are predicted for each encoder.
- 3) JTSPS [16], where optimized bitrate-resolution pairs are predicted for JND-aware efficient encoding for each encoder.

Note that separate bitrate ladders are generated for each encoder in the state-of-the-art encoding schemes.

B. Latency and Accuracy Analysis

Spatio-temporal features (cf. Section II-A) are extracted at a rate of 370fps. The overall inference time (including the feature extraction time, VMAF prediction time, and inference time) for a 4s video segment of 2160p resolution is 0.37s. Hence, the additional latency introduced by MCBE is negligible. The average mean absolute error (MAE) of VMAF prediction for all resolutions is observed to be 2.42, which is acceptable for live-streaming applications.

C. Storage Consumption Analysis

Fig. 4 shows the rate-distortion (RD) curves of selected video sequences (segments) encoded using JTSPS bitrate ladder prediction method for x264, x265, and svtav1. It is observed that there are bitrate regions where the new-generation codecs (i.e., HEVC and AV1) have lower RD performance compared to AVC. MCBE eliminates the representations of new-generation codecs when their predicted VMAF is lower than the RD curve of the AVC encoding. In Fig. 4, dot marks indicate the eliminated representations. Furthermore, it is also observed that MCBE removed the perceptual redundancy between multiple codec representations based on the JND threshold of six VMAF points. Table I shows the storage reduction (ΔS) using MCBE in conjunction with the HLS bitrate ladder, OPTE, and JTSPS. As v_J increases, more representations are eliminated, which reduces the storage needed. HLS

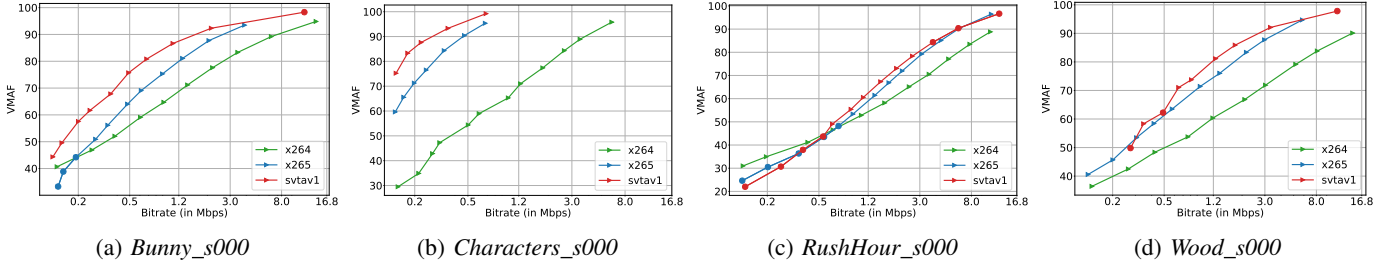


Fig. 4: RD curves of representative segments (a) *Bunny_s000* ($E_Y=22.40$, $h=4.70$, $L_Y=129.21$), (b) *Characters_s000* ($E_Y=45.42$, $h=36.88$, $L_Y=134.56$), (c) *RushHour_s000* ($E_Y=47.75$, $h=19.70$, $L_Y=101.66$), and (d) *Wood_s000* ($E_Y=124.72$, $h=47.03$, $L_Y=119.57$) of VCD dataset [15] encoded using MCBE (x264, x265, svtav1). Here, JTPS [16] is considered as the bitrate ladder prediction method, and $v_J = 6$. Representations marked using dots indicate the eliminated representations.

TABLE I: Average performance results using MCBE compared to HLS, OPTE, and JTPS bitrate ladders prediction methods for various target encoder combinations.

| MCBE configuration | | HLS ladder [43] | | | OPTE [33] | | | JTPS [16] | | |
|----------------------|-------|------------------|------------|------------------|------------------|------------|------------------|------------------|------------|------------------|
| Target encoders | v_J | ΔE_{enc} | ΔS | ΔE_{sto} | ΔE_{enc} | ΔS | ΔE_{sto} | ΔE_{enc} | ΔS | ΔE_{sto} |
| (x264, x265) | 2 | -34.05% | -43.70% | -68.30% | -36.03% | -46.12% | -70.97% | -15.82% | -10.20% | -19.36% |
| | 4 | -47.72% | -60.48% | -84.38% | -49.96% | -63.00% | -86.31% | -16.07% | -9.53% | -18.15% |
| | 6 | -58.09% | -69.91% | -90.94% | -59.75% | -72.50% | -92.44% | -16.18% | -11.27% | -21.26% |
| (x264, svtav1) | 2 | -34.50% | -43.31% | -67.87% | -36.87% | -45.73% | -70.55% | -12.76% | -8.79% | -16.81% |
| | 4 | -48.19% | -59.95% | -83.96% | -51.17% | -62.82% | -86.18% | -12.82% | -8.27% | -15.86% |
| | 6 | -58.61% | -69.18% | -90.50% | -61.06% | -72.60% | -92.49% | -12.90% | -9.38% | -17.88% |
| (x264, x265, svtav1) | 2 | -20.42% | -53.20% | -78.10% | -18.55% | -53.63% | -78.50% | -22.56% | -14.57% | -27.01% |
| | 4 | -41.67% | -69.15% | -90.49% | -39.68% | -69.49% | -90.69% | -23.74% | -17.81% | -32.45% |
| | 6 | -56.45% | -77.61% | -94.99% | -54.34% | -78.32% | -95.30% | -27.80% | -22.62% | -40.12% |

bitrate ladder and OPTE representations have high perceptual redundancy compared to JTPS, as JTPS representations are predicted with a perceptual gap of one JND [16]. Hence, storage reduction is significantly high with HLS bitrate ladder and OPTE, compared to JTPS.

D. Energy Consumption Analysis

This section evaluates energy consumption using MCBE in terms of (i) encoding (ΔE_{enc}), and (ii) storage (ΔE_{sto}). The *CodeCarbon* tool [47] is used to calculate the encoding energy. The storage energy is modeled inspired by [48] as $E_{sto} = S_d \cdot P_b \cdot T_s$, where S_d is the video data size (in bits), P_b is power consumption per bit (in W/bit), and T_s is the time taken for data to be stored (in hours).

Table I illustrates the average energy reduction achieved in encoding and storage using MCBE compared to the alternative schemes. Negative values in the table indicate the extent of the reduction in energy consumption. Compared to the state-of-the-art, the results show significant encoding energy reduction ΔE_{enc} for MCBE. For instance, in a streaming session with devices supporting AVC, HEVC, and AV1 decoding, and considering a JND of six VMAF points, MCBE achieves energy reductions of up to 56.45%, 54.34%, and 27.80% when compared to HLS bitrate ladder encoding, OPTE, and JTPS, respectively. This substantial reduction in encoding energy is primarily because MCBE eliminates the need to encode segments for all representations in all x264, x265, and svtav1 bitrate ladders as explained previously in Section II-B. Instead, MCBE selects the representation with the lowest energy requirement

(i.e., AVC representation) for encoding when representations in different codecs have the same VMAF value. Consequently, MCBE predominantly includes all x264 representations and only the higher VMAF representations from the other codecs, leading to a reduced amount of data to encode ΔE_{enc} and store ΔE_{sto} , thus consuming less storage energy. For example, compared to the HLS ladder, MCBE reduces the data to store by 77.61%, resulting in a remarkable 94.99% less energy consumed for storage during a streaming session with devices supporting AVC, HEVC and AV1 decoding and a JND of six VMAF points.

IV. CONCLUSIONS

This paper proposed MCBE, an online energy-efficient multi-codec JND-aware bitrate ladder estimation scheme for adaptive streaming applications. MCBE includes an algorithm to determine an optimized multi-codec encoding bitrate ladder, where redundant representations of new-generation video codecs are eliminated. Furthermore, perceptual redundancy within the representations of each codec is minimized by eliminating representations based on the JND threshold. MCBE can be used in conjunction with any bitrate ladder estimation scheme. MCBE on average, yields encoding, storage, and transmission energy savings of 56.45%, 77.61%, and 94.99%, respectively, compared to the state-of-the-art HLS bitrate ladder encoding, for a streaming session with devices supporting AVC, HEVC, and AV1 decoding, considering a JND of six VMAF points.

REFERENCES

- [1] I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming Over the Internet," *IEEE MultiMedia*, vol. 18, no. 4, pp. 62–67, April 2011.
- [2] R. Pantos, Ed. and W. May, "HTTP Live Streaming," *RFC 8216*, <https://www.rfc-editor.org/info/rfc8216>, August 2017.
- [3] R. Farahani et al., "Hybrid P2P-CDN Architecture for Live Video Streaming: An Online Learning Approach," in *2022 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2022, pp. 1911–1917.
- [4] R. Farahani et al., "ARARAT: A Collaborative Edge-Assisted Framework for HTTP Adaptive Video Streaming," *IEEE Transactions on Network and Service Management (TNSM)*, pp. 625–643, 2022.
- [5] A. Bentalb et al., "A Survey on Bitrate Adaptation Schemes for Streaming Media Over HTTP," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 562–585, 2019.
- [6] C. Timmerer, M. Smole, and C. Mueller, "Efficient Multi-Codec Support for OTT Services: HEVC/H. 265 and/or AV1?" 2018. [Online]. Available: <http://www.itec.aau.at/bib/files/TimmererC010218.pdf>
- [7] T. Wiegand et al., "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.
- [8] G. J. Sullivan et al., "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [9] Y. A. Reznik, "Toward Efficient Multicore Streaming," *SMPTE Motion Imaging Journal*, vol. 132, no. 4, pp. 16–25, 2023.
- [10] J. Han et al., "A Technical Overview of AV1," *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1435–1462, 2021.
- [11] B. Bross et al., "Overview of the Versatile Video Coding (VVC) Standard and its Applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736–3764, 2021.
- [12] J. Samuelsson et al., "MPEG-5 EVC," in *SMPTE 2019*, 2019, pp. 1–11.
- [13] S. Battista et al., "Overview of the Low Complexity Enhancement Video Coding (LCEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7983–7995, 2022.
- [14] Y. A. Reznik et al., "Optimal multi-codec adaptive bitrate streaming," in *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2019, pp. 348–353.
- [15] H. Amirpour et al., "VCD: Video Complexity Dataset," in *Proceedings of the 13th ACM Multimedia Systems Conference*. New York, NY, USA: Association for Computing Machinery, 2022, p. 234–239.
- [16] V. V. Menon et al., "JND-aware Two-pass Per-title Encoding Scheme for Adaptive Live Streaming," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2023.
- [17] VideoLAN, "x264." [Online]. Available: <https://www.videolan.org/developers/x264.html>
- [18] VideoLAN, "x265." [Online]. Available: <https://www.videolan.org/developers/x265.html>
- [19] Alliance of Open Media (AOM), "SVT-AV1." [Online]. Available: <https://gitlab.com/AOMediaCodec/SVT-AV1>
- [20] A. Zbrovskiy, C. Feldmann, and C. Timmerer, "A Practical Evaluation of Video Codecs for Large-Scale HTTP Adaptive Streaming Services," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 998–1002.
- [21] A. V. Katsenou et al., "A Subjective Comparison of AV1 and HEVC for Adaptive Video Streaming," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 4145–4149.
- [22] V. V. Menon et al., "Green Video Complexity Analysis for Efficient Encoding in Adaptive Video Streaming," in *First International ACM Green Multimedia Systems Workshop (GMSys '23)*, 2023.
- [23] M. G. Koziri et al., "Efficient cloud provisioning for video transcoding: Review, open challenges and future opportunities," *IEEE Internet Computing*, vol. 22, no. 5, pp. 46–55, 2018.
- [24] J. Baliga et al., "Green cloud computing: Balancing energy in processing, storage, and transport," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 149–167, 2010.
- [25] D. Lorenzi, "QoE- and Energy-Aware Content Consumption For HTTP Adaptive Streaming," in *Proceedings of the 14th Conference on ACM Multimedia Systems*, 2023, p. 348–352.
- [26] T. Laude et al., "A Comparison of JEM and AV1 with HEVC: Coding Tools, Coding Efficiency and Complexity," in *2018 Picture Coding Symposium (PCS)*, 2018, pp. 36–40.
- [27] I. Bender et al., "Compression Efficiency and Computational Cost Comparison between AV1 and HEVC Encoders," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.
- [28] V. V. Menon, "Video Coding Enhancements for HTTP Adaptive Streaming," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, p. 6905–6909.
- [29] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, 2001.
- [30] D. Yuan et al., "Visual JND: A Perceptual Measurement in Video Coding," *IEEE Access*, vol. 7, pp. 29 014–29 022, 2019.
- [31] S. Zvezdakov, D. Kondranin, and D. Vatolin, "Machine-Learning-Based Method for Content-Adaptive Video Encoding," in *2021 Picture Coding Symposium (PCS)*, 2021, pp. 1–5.
- [32] A. V. Katsenou, J. Sole, and D. R. Bull, "Content-agnostic Bitrate Ladder Prediction for Adaptive Video Streaming," in *2019 Picture Coding Symposium (PCS)*, 2019, pp. 1–5.
- [33] V. V. Menon et al., "OPTE: Online Per-Title Encoding for Live Video Streaming," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 1865–1869.
- [34] J. De Cock et al., "Complexity-based consistent-quality encoding in the cloud," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 1484–1488.
- [35] A. Zbrovskiy et al., "FAUST: Fast Per-Scene Encoding Using Entropy-Based Scene Detection and Machine Learning," in *2021 30th Conference of Open Innovations Association FRUCT*, 2021, pp. 292–302.
- [36] N. B. Harikrishnan et al., "Comparative evaluation of image compression techniques," in *2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)*, 2017, pp. 1–4.
- [37] V. V. Menon et al., "Transcoding Quality Prediction for Adaptive Video Streaming," in *Proceedings of the 2nd Mile-High Video Conference*, 2023, p. 103–109.
- [38] V. V. Menon et al., "Video Quality Assessment with Texture Information Fusion for Streaming Applications," 2023.
- [39] J. Y. Lin et al., "Experimental design and analysis of JND test on coded image/video," in *Applications of Digital Image Processing XXXVIII*, vol. 9599. SPIE, 2015, pp. 324–334.
- [40] H. Wang et al., "MCL-JCV: a JND-based H. 264/AVC video quality assessment dataset," in *2016 IEEE International Conference on image processing (ICIP)*. IEEE, 2016, pp. 1509–1513.
- [41] H. Wang et al., "VideoSet: A large-scale compressed video quality dataset based on JND measurement," *Journal of Visual Communication and Image Representation*, vol. 46, pp. 292–302, 2017.
- [42] V. V. Menon et al., "Content-adaptive Variable Framerate Encoding Scheme for Green Live Streaming," 2023.
- [43] Apple Inc., "HLS Authoring Specification for Apple Devices." [Online]. Available: https://developer.apple.com/documentation/http_live_streaming/hls_authoring_specification_for_apple_devices
- [44] P. K. Tiwari et al., "Accelerating x265 with intel® advanced vector extensions 512," *White Paper on the Intel Developers Page*, 2018. [Online]. Available: <https://www.intel.com/content/dam/develop/external/us/en/documents/mcw-intel-x265-avx512.pdf>
- [45] A. Kah et al., "Fundamental relationships between subjective quality, user acceptance, and the VMAF metric for a quality-based bit-rate ladder design for over-the-top video streaming services," in *Applications of Digital Image Processing XLIV*, vol. 11842, International Society for Optics and Photonics. SPIE, 2021, p. 118420Z.
- [46] J. Ozer, "Finding the Just Noticeable Difference with Netflix VMAF," Sep. 2017. [Online]. Available: <https://streaminglearningcenter.com/codecs/finding-the-just-noticeable-difference-with-netflix-vmef.html>
- [47] BCG-GAMMA and MILA, "CodeCarbon." [Online]. Available: <https://codecarbon.io/>
- [48] A. Bianco, R. Mashayekhi, and M. Meo, "Energy consumption for data distribution in content delivery networks," in *2016 IEEE International Conference on Communications (ICC)*. IEEE, 2016, pp. 1–6.