University of Victoria

CSC 461 Project Report

AV1 Encoding

Shaun Nguyen
V00945109

Tirth Patel
V00950586

December 11, 2023

Table of Contents

Introduction

In the landscape of digital video, the demand for high-quality, bandwidth-efficient encoding has fueled the development of advanced video codecs. Among these, AV1 (AOMedia Video 1) stands out as an open-source and royalty-free codec, designed to address the evolving needs of video content delivery.

AV1 represents a collaborative effort by the Alliance for Open Media (AOMedia), featuring major industry players like Netflix, Google, and others. The codec is engineered to deliver exceptional compression efficiency without compromising on visual quality, making it a compelling solution for a wide array of applications, from online streaming to video conferencing.

AV1 Characteristics

- **Royalty-Free and Open Source:** AV1 is an open-source and royalty-free video codec, which means it can be freely used without incurring licensing fees. This accessibility fosters widespread adoption and innovation.

- **Exceptional Compression Efficiency:** AV1 achieves superior compression efficiency compared to its predecessors. It can deliver high-quality video at lower bitrates, resulting in reduced storage requirements and bandwidth usage.

- **Versatility Across Resolutions and Bitrates:** AV1 is designed to handle a wide range of resolutions, from low-resolution videos to ultra-high-definition content. It adapts well to varying bitrate requirements, making it suitable for diverse applications, including streaming, video conferencing, and broadcasting.

- **Optimized for Web and Streaming:** AV1 is particularly well-suited for web-based applications and online streaming services. Its efficient compression enables faster

loading times for web content and smoother streaming experiences, even under challenging network conditions.

Figure 1: AV1 Characteristics

- **State-of-the-Art Video Quality:** AV1 excels in preserving video quality, even at low bitrates. It incorporates advanced coding techniques and tools to minimize artifacts and enhance visual fidelity, making it an ideal choice for high-quality video delivery.

- **Wide Industry Support:** AV1 has garnered support from a broad spectrum of industry players, including major technology companies, content creators, and streaming platforms. This widespread backing enhances its potential for becoming a standard in video encoding.

- **Adaptive Bitrate Streaming (ABR) Compatibility:** AV1 supports Adaptive Bitrate Streaming, allowing content providers to dynamically adjust video quality based on the viewer's network conditions. This ensures a seamless viewing experience without interruptions.

- **Hardware Acceleration and Decoding Support:** As AV1 gains traction, hardware support for encoding and decoding is becoming more prevalent. This facilitates efficient real-time processing and playback on a variety of devices, including smartphones, smart TVs, and streaming media players.

- **Constant Evolution and Improvement:** The AV1 standard is subject to continuous development and enhancement. Ongoing efforts by the AOMedia consortium contribute to refining the codec, addressing limitations, and ensuring its competitiveness in the ever-evolving landscape of video compression.

- **Forward-Looking Technology:** AV1 represents a forward-looking approach to video encoding, incorporating the latest advancements in compression technology. Its design considers emerging trends such as 8K resolution, high dynamic range (HDR), and immersive multimedia experiences.

Technical Overview

I. High-level syntax

The AV1 bitstream is organized into open bitstream units (OBUs), which are then used in the AV1 decoding process. Each OBU consists of a variable-length sequence of bytes and includes a header and payload. The header identifies the OBU type and specifies the payload size. Common OBU types are as follows:

- **Sequence Header:** Contains information for the entire sequence, including the sequence profile and activation of specific coding tools.
- **Temporal Delimiter:** Indicates the frame's presentation timestamp and defines a temporal unit for frames sharing the same timestamp, beneficial in scalable coding.
- **Frame Header:** Configures coding details for a specific frame, specifying inter or intraframe type, reference frames, and probability model updates.

- **Tile Group:** Holds tile data for a frame, with each tile independently decodable and contributing to the reconstructed frame.

- **Frame:** Combines the frame header and tile data, reducing overhead compared to separate frame header and tile group OBUs.

- **Metadata:** Carries additional information like high dynamic range, scalability, and timecode.

- **Tile List:** Similar to a tile group OBU, it contains tile data, but each tile has an extra header indicating its reference frame index and position in the current frame. This enables selective tile decoding, useful for applications like light field technology.

II. Reference frame system

Reference Frames

The AV1 codec can store up to eight frames in its decoded frame buffer, with seven frames available as reference frames when encoding a new frame. These reference frames are assigned unique indexes from 1 to 7. Typically, frames with indexes 1–4 are used for preceding frames in the display order, and indexes 5–7 are for subsequent frames. Compound interprediction can combine two references, leading to either unidirectional or bidirectional compound predictions. Unidirectional compound predictions are limited to four specific pairs, while bidirectional compound predictions support all 12 combinations. This limitation reduces the total number of possible compound reference frame combinations. It is based on the idea that balanced reference frames on both sides of the current frame tend to yield better predictions. After encoding a frame, the encoder can choose to replace a reference frame in the decoded frame buffer and explicitly signal this in the bitstream. Additionally, there is an option to skip updating the decoded frame buffer, which is particularly useful in high-motion videos where some frames have less relevance to neighboring frames.

Alternate Reference Frame

The alternate reference frame (ARF) in AV1 is a frame that's encoded and stored in the decoded frame buffer but may not be displayed. Its main role is to act as a reference frame for

subsequent frames in the processing pipeline. When sending a frame for display, the AV1 codec has the option to either code a new frame or use an existing frame in the decoded frame buffer, known as "show existing frame." If an ARF is eventually displayed, it can effectively be used to code a future frame in a pyramid coding structure.

Moreover, the encoder can generate a frame to reduce prediction errors among multiple displayed frames. This can involve applying temporal filtering to consecutive original frames, creating an ARF with reduced noise and retaining common information. To optimize the overall rate-distortion performance, the encoder typically uses a lower quantization step size to encode this common information (the ARF). However, a drawback is that it adds an extra frame for decoders to process, potentially affecting hardware capacity. To balance compression performance and decoder throughput, each level definition establishes a maximum decode rate, limiting the number of permissible synthesized ARF frames, considering both displayable frames and ARFs not used as "show existing frames."

Frame Scaling

The AV1 codec offers the option to reduce the resolution of a source frame for compression and then restore it to the original resolution. This feature comes in handy when a few frames are too complex to compress effectively within the target streaming bandwidth. The downscaling factor is restricted to a range of 8/16–15/16. The reconstructed frame is initially upscaled linearly to its original size and subsequently undergoes a loop restoration filter as part of the postprocessing. Both the linear upscaling filter and loop restoration filter operations are defined as standard procedures, which we'll delve into further in Section VII-D. To keep hardware implementation cost-effective without requiring additional line buffers beyond what's needed for regular frame decoding, the rescaling is limited to the horizontal direction. The upscaled and filtered version of the decoded frame becomes available as a reference frame for coding subsequent frames.

III. Superblock and Tile

Super block

The superblock size in AV1 can be either 128×128 luma samples or 64×64 luma samples, and this choice is indicated in the sequence header. Superblocks can be divided into smaller coding blocks, each with its unique prediction and transform modes. The coding of a superblock depends only on its adjacent superblocks to the left and above.

Tile

In AV1, a tile is a rectangular collection of superblocks that is limited in spatial referencing to within the tile boundaries. This allows independent coding of titles within a frame, aiding efficient multithreading for both encoding and decoding. Tiles can range from a minimum size of one superblock to a maximum width of 4096 luma samples, with the maximum tile size being 4096×2304 luma samples. Up to 512 tiles are permitted in a frame.
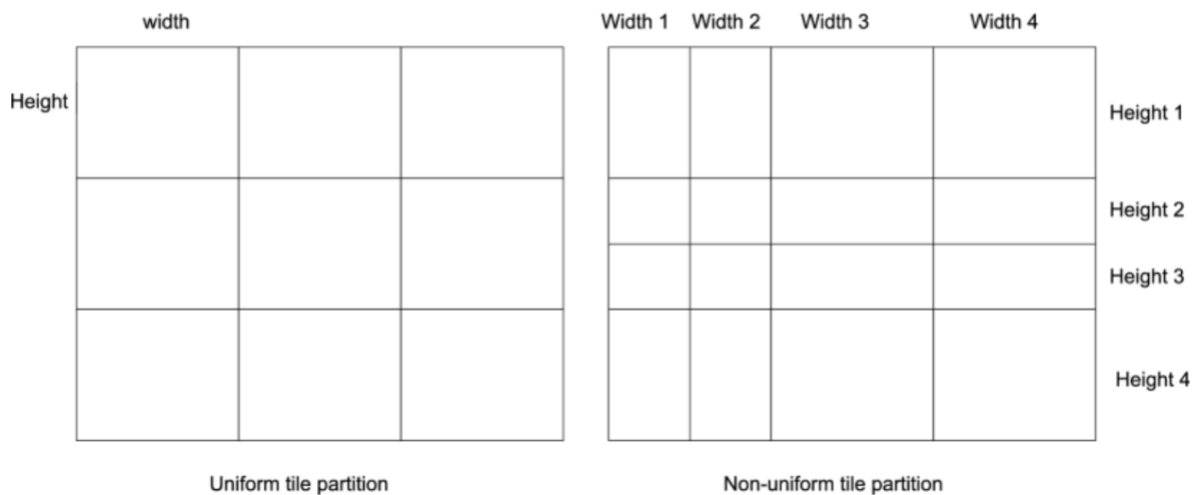

Figure 2: Uniform and non-uniform title size

AV1 offers two ways to specify tile sizes. The uniform tile size option assumes all tiles within a frame are of the same size, except those on the frame's bottom or right boundary. The number of tiles vertically and horizontally is identified in the bitstream, and tile dimensions are derived based on the frame size. The second option, nonuniform tile size, uses a lattice arrangement of tiles with nonuniform spacing both vertically and horizontally. Tile dimensions must be

specified in units of superblocks in the bitstream. This option accommodates variations in computational complexity across superblocks within a frame due to differences in video signal statistics. It enables the use of smaller tile sizes for regions with higher computational demands, helping to distribute the workload among threads. This is particularly valuable when ample computing resources and minimal frame coding latency are required. An example illustrating the two tile options is provided in Figure 2.

The choice between uniform and nonuniform tile sizes, as well as the specific tile sizes, is determined for each individual frame. It's important to mention that postprocessing filters are used to smooth out any potential coding artifacts, like blocking artifacts, along the boundaries between tiles.

IV. Coding Block Operations

Coding Block Partitioning

AV1 retains the recursive block partitioning design from VP9, allowing superblocks to be divided into smaller block sizes for coding. To minimize overhead in predicting highly correlated video signals, especially in 4k videos, AV1 supports a maximum coding block size of 128×128 luma samples. Each block level has ten possible partition options (Fig 3), offering greater coding flexibility. However, the minimum coding block size is expanded to 4×4 luma samples, enhancing prediction quality for complex videos. These extensions bring coding flexibility but may pose challenges for hardware decoders. Special constraints based on block size are introduced to address such issues.
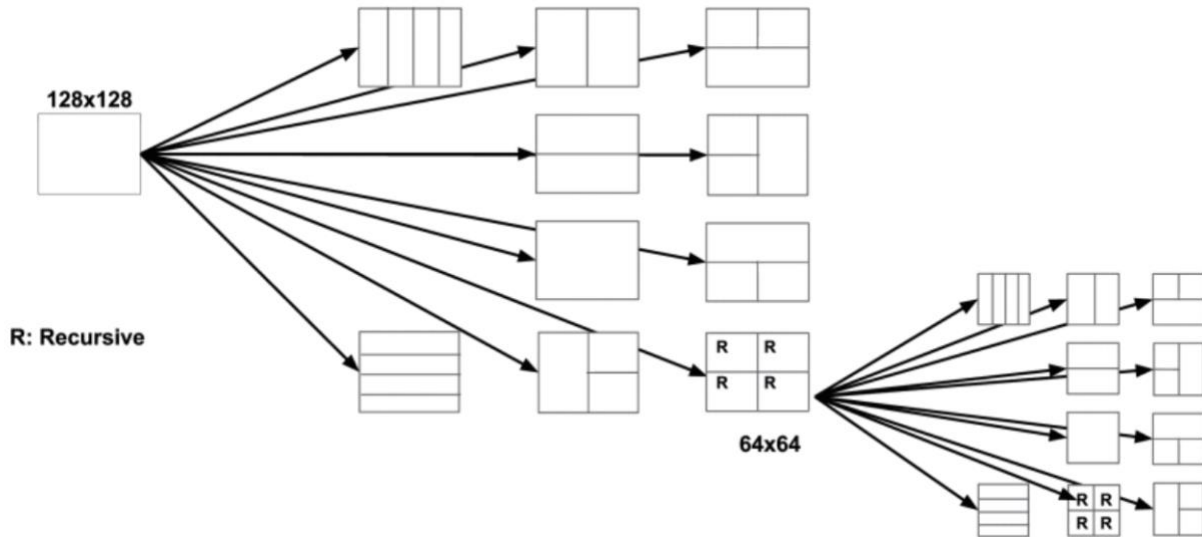
Figure 3: Recursive block partition

Intraframe Prediction

In intramode coding blocks, the prediction modes for the luma component and both chroma components are separately indicated in the bitstream. The luma prediction mode is encoded using a probability model determined by the prediction context from the neighboring coding blocks. The coding of the chroma prediction mode depends on the luma prediction mode. In intraprediction, transform blocks are used, and it relies on previously decoded boundary pixels as a reference.

Interframe Prediction

AV1 offers a comprehensive set of tools for leveraging temporal correlations in video signals. These include adaptive filtering in translational motion compensation, affine motion compensation, and highly flexible compound prediction modes.

Dynamic Motion Vector Referencing Scheme

Motion vector coding significantly affects the overall bit rate in video codecs. To reduce this, modern codecs often use predictive coding for motion vectors and then encode the differences using entropy coding. The accuracy of this prediction plays a crucial role in coding efficiency. AV1 uses a dynamic motion vector referencing approach that gathers candidate motion vectors

from nearby spatial and temporal neighbors and ranks them to optimize entropy coding efficiency.

Transform Coding

AV1 inherits the transform coding scheme in VP9 and extends its flexibility in terms of both the transform block sizes and the kernels. Transform coding is applied to the prediction residual to remove the potential spatial correlations. Transform blocks within a coding block share the same transform size. Four square transform sizes are supported 4×4, 8×8, 16×16, and 32×32.

A set of separable 2-D transform types, constructed by combinations of 1-D discrete cosine transform (DCT) and asymmetric discrete sine transform (ADST) kernels, is selected based on the prediction mode.
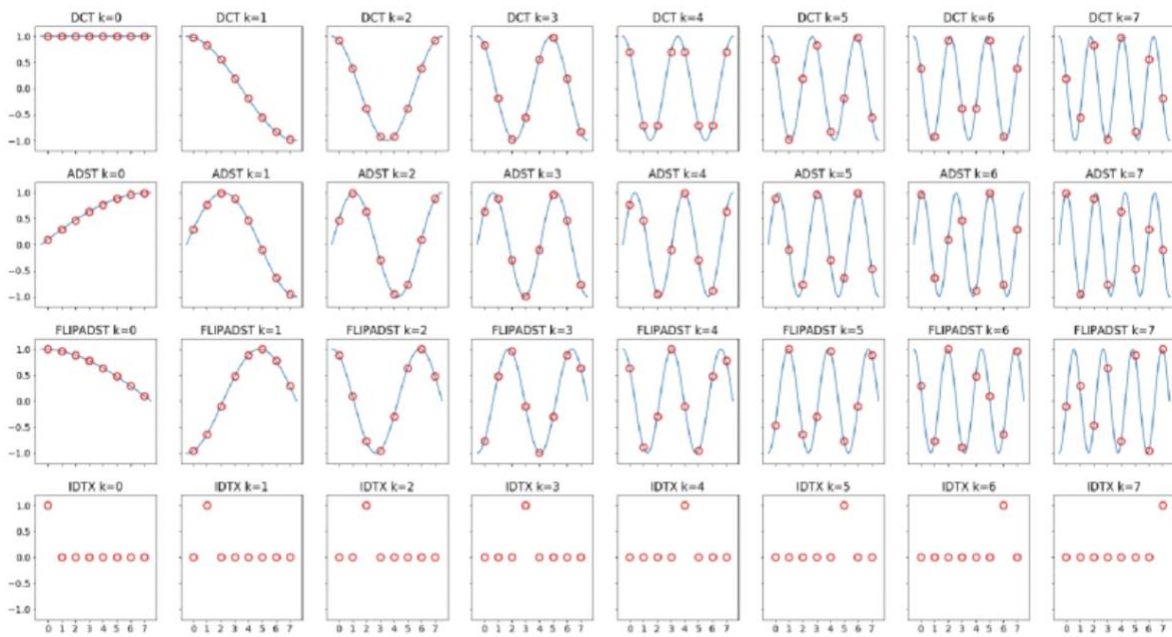
Transform kernels of DCT, ADST, FLIPADST, and IDTX for dimension N=8

Figure 4: Transform kernels

V. Entropy Coding System

VP nine uses a more complicated version of the binary be coder to compress all its elements, AV1 further expands upon this by creating an adaptive multi symbol arithmetic encoder. Each

syntax element in AV1 is a member of a specific alphabet of N elements, and a context consists of a set of N probabilities together with a count to facilitate fast early adaptation. The probabilities are stored as 15-bit cumulative distribution functions (CDFs). This allow more frequent syntax elements have higher probability and are stored using less symbols.

AV1 at Netflix

Netflix consistently advances video compression and analysis due to the high frequency of their content playback—every saved bit translates to real cost savings. An exemplary illustration of this commitment is Netflix's development of the Video Multimethod Assessment Fusion (VMAF) metric. Created in collaboration with two universities, VMAF serves as a key tool for evaluating video quality through computer analysis and has remained integral to Netflix's work.

VMAF is utilized to validate results from Netflix's Per-Shot Encoding, a technique adjusting encoding parameters for each shot in a film, deviating from fixed parameters for the entire film. This innovative approach has been discussed in previous talks, including insights into Netflix's per-title encoding.

Figure 5: Netflix AV1

Netflix's significant contribution to the video codec landscape is AV1. As a founding member of the Alliance for Open Media (AoM), Netflix recognized the need for a better codec. By contributing to an open standard, Netflix not only addressed its own requirements but also fostered a collaborative community, attracting contributions to both the codec's development and widespread adoption.

Recent developments include streaming AV1 to Android clients, taking advantage of AV1's native support for 10-bit video—a distinguishing feature from other codecs like AVC and HEVC. There are visual enhancements in scenes such as skies and water achieved through the use of 10-bit.
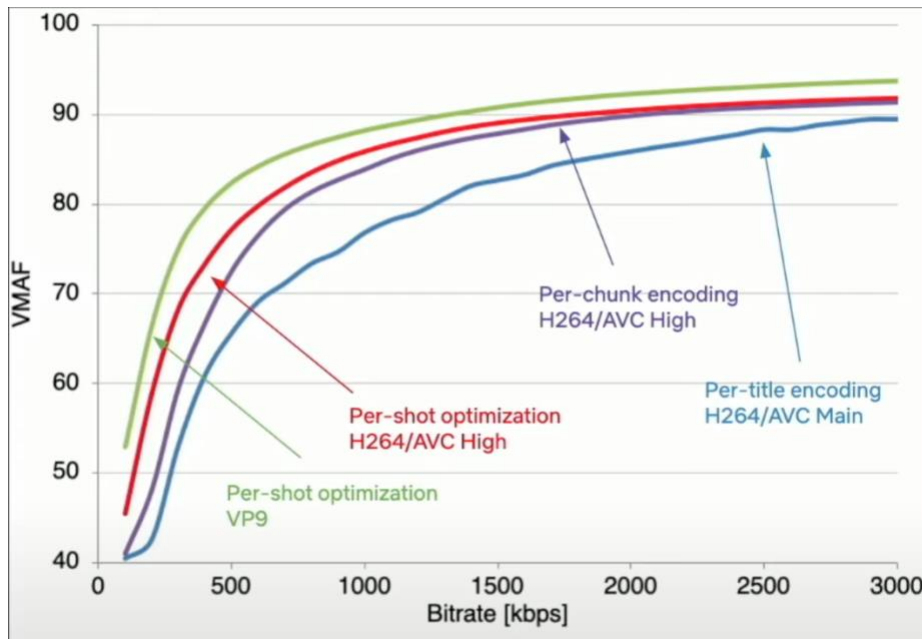

Figure 6: VMAF (quality metric) vs Bitrate

Another standout feature of AV1 is Film Grain synthesis, designed to enhance encoding efficiency by removing random film grain during encoding and reapplying similar noise to maintain the original aesthetic. This technique can result in up to a 30% reduction in bitrate. Netflix has collaborated with Intel on the SVT-AV1 software video encoder. Leveraging Intel's

SVT technology optimized for Xeon chips, SVT-AV1 aims to increase adoption by delivering a data center-ready, optimized encoder.

SVT-AV1: open-source AV1 encoder and decoder

The Scalable Video Technology for AV1 (SVT-AV1) is an open-source AV1 codec implementation developed collaboratively by Intel and Netflix. SVT-AV1 aims to provide high performance and scalability in AV1 video encoding. A collaboration between Intel and Netflix since August 2018, this project has seen active contributions from other partner companies and the open-source community since its open-sourcing. It is available on GitHub under a BSD + patent license.

Architectural Features:

Intel aimed to create a high-performance and scalable AV1 encoder with SVT-AV1. The encoder employs parallelization at various stages of the encoding process, adapting to the number of available cores, thus reducing encoding time while maintaining compression efficiency. Key architectural features include multi-dimensional parallelism, multi-stage partitioning decisions, block-based mode decisions, RD-optimized classification, and open-loop hierarchical motion estimation.

One of the key architectural features of SVT-AV1 is its multi-stage partitioning decision mechanism. This feature is crucial in the encoding process as it determines how to divide the video frames into smaller partitions for more efficient compression. SVT-AV1 utilizes multi-dimensional parallelism in this partitioning decision, optimizing the use of available computing resources, especially multiple CPU cores. This allows SVT-AV1 to adapt to the number of available cores, including newer servers with a significant core count.
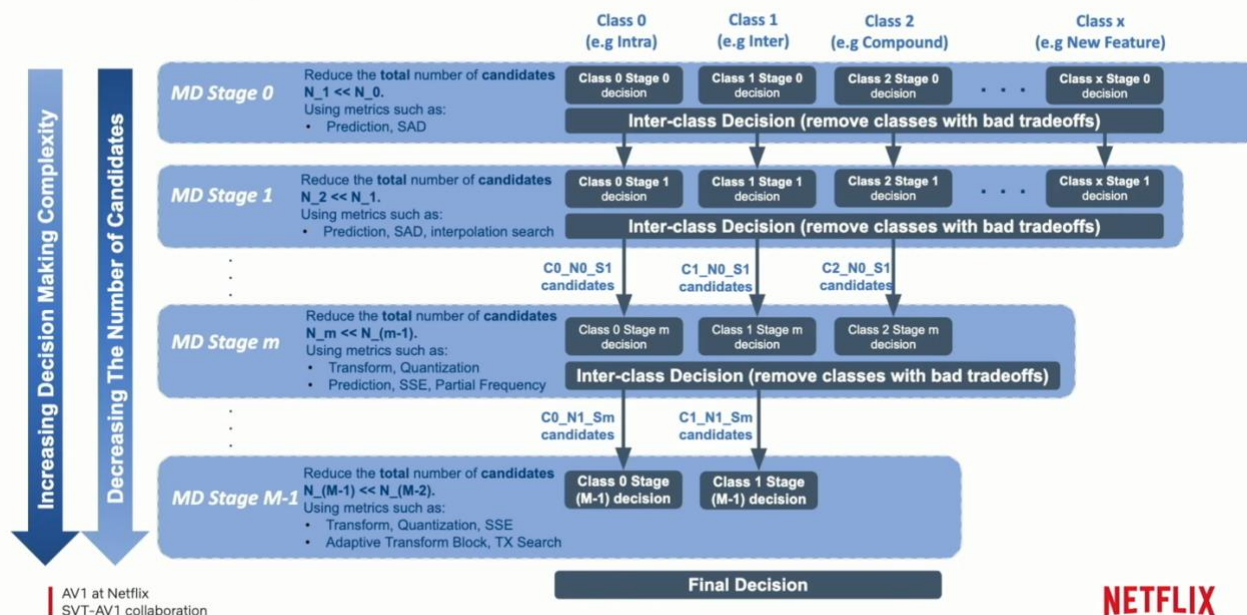
Figure 7: Multi-stage multi-class mode decision

Additionally, SVT-AV1 employs a multi-stage class mode decision approach. Mode decision is a fundamental aspect of video encoding where the encoder determines the best coding mode (such as inter or intra prediction) for each block in a frame. The multi-stage class mode decision in SVT-AV1 involves a classification process that helps the codec make informed decisions about the coding mode for different blocks. This can significantly impact compression efficiency and overall video quality.

These multi-stage decisions in SVT-AV1 contribute to achieving a balance between compression efficiency and performance. By optimizing the partitioning and mode decision processes through multiple stages, SVT-AV1 can efficiently utilize computing resources, resulting in competitive compression efficiency and reduced encoding time.

Compression Efficiency and Performance:

In terms of compression efficiency, SVT-AV1 rivals libaom at slow speed settings. The encoder's performance improvements over time can be observed on the Video Codec Tracker website.

SVT-AV1's 2-pass encoding mode, utilizing 4-thread mode, competes well with libaom. The encoder shows a 16.5% decrease in encoding time compared to libaom, while maintaining slightly superior compression efficiency, as indicated by various quality metrics like PSNR, VMAF, and MS-SSIM.
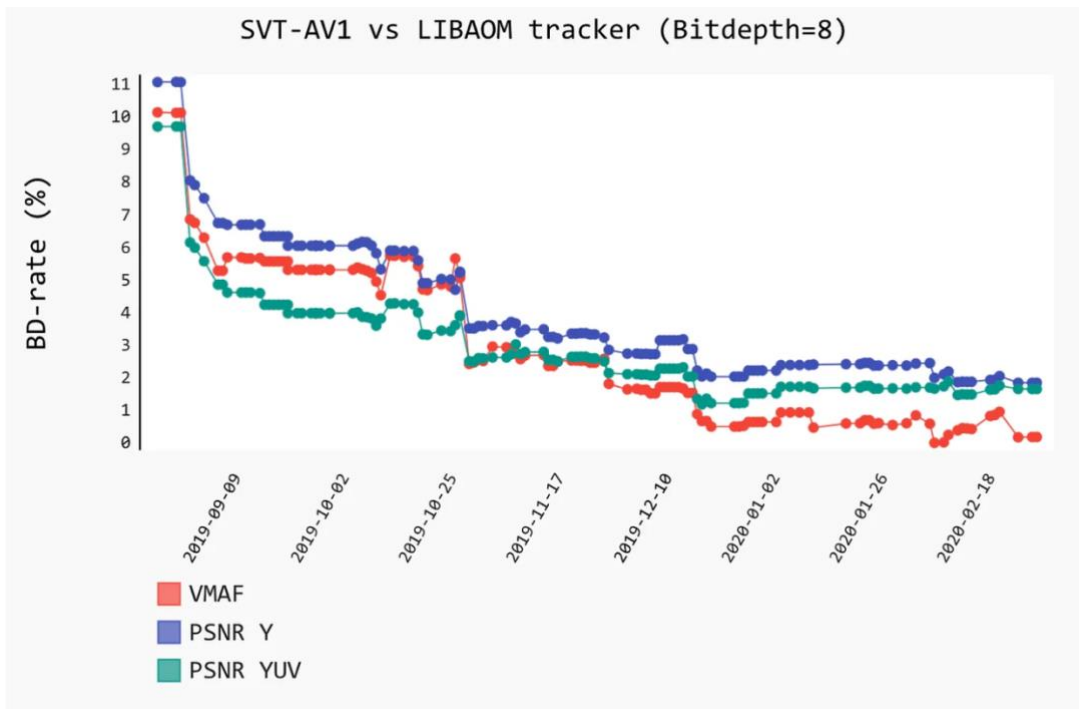

Figure 8: SVT-AV1 vs LIBOM

Decoder Performance:

In decoding, SVT-AV1 demonstrates slightly faster performance than libaom in the 1-thread mode, with even larger improvements in the 4-thread mode. Particularly noteworthy are the significant speed gains observed when decoding bitstreams with multiple tiles using the 4-thread mode. The decoder's performance is deemed satisfactory for a research decoder, prioritizing experimentation over production-level optimization.

Testing Framework:

To ensure codec conformance, SVT-AV1 has been extensively covered with unit tests and end-to-end tests, employing the Google Test framework. These tests, automatically triggered for

each pull request through GitHub actions, support sharding and parallel execution to expedite the review process.
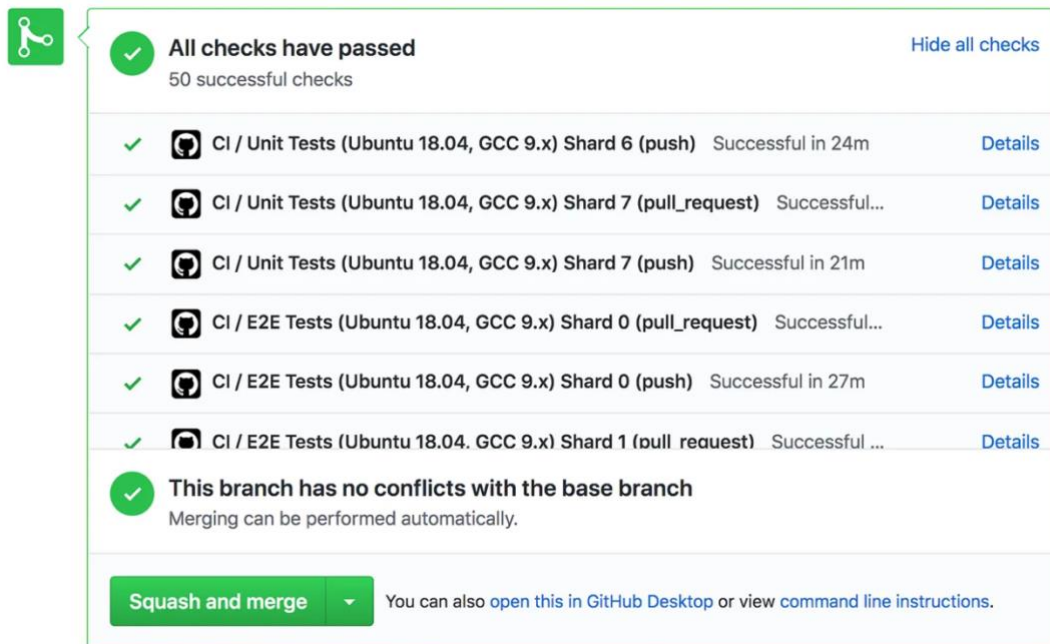

Figure 9: SVT-AV1 Testing

Overall, SVT-AV1 represents a collaborative effort between Intel and Netflix, delivering a high-performance AV1 codec with notable compression efficiency improvements and satisfactory decoder performance for research purposes. The project has embraced openness and community contributions, fostering ongoing enhancements to the AV1 codec implementation.

Current Development

Current state of AV1 playback support (2023), covering which browsers, mobile devices, smart TVs, consoles and streaming sticks are compatible with AV1 codec so far. On Sep 12, 2023, Apple announced the A17 Pro chip in their new iPhone 15 Pro would include a dedicated AV1 decoder. This is a big line in the sand for Apple and for the wider industry and will hopefully prove to be the day that revitalized interest and momentum for AV1 adoption across the industry.
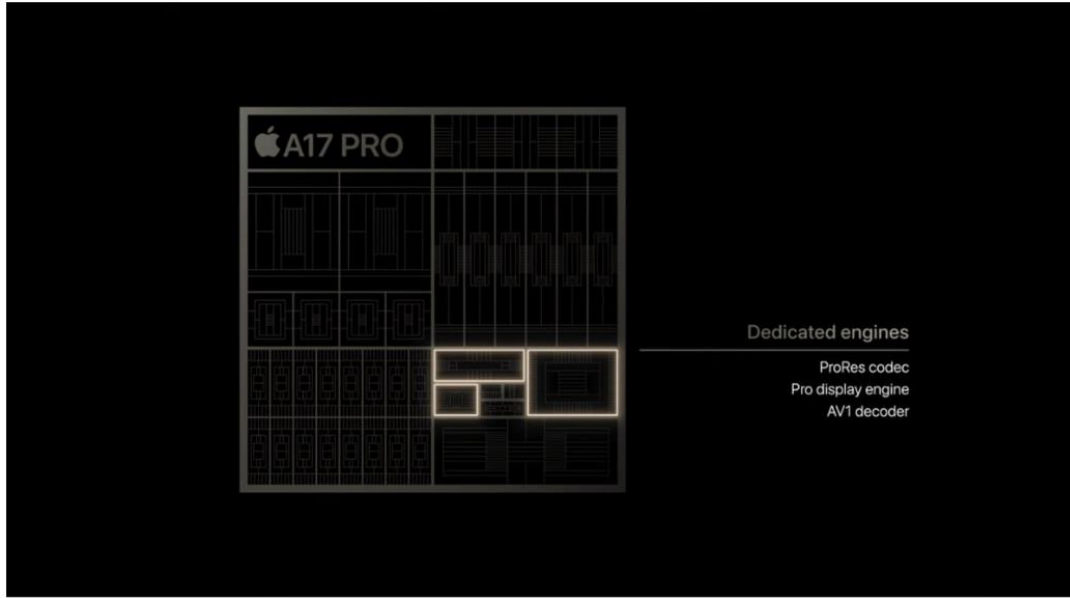
Figure 10: Apple A17 Pro Chip

In the beginning of 2023, as anticipation grew for Apple's official endorsement of AV1, Meta took proactive steps by independently incorporating AV1 support for their Reels videos on Facebook and Instagram, extending to iOS devices. Meta expressed confidence in AV1 as the most promising codec for their video products in the foreseeable future. The accompanying image illustrates the substantial enhancement in visual quality achieved with AV1 in comparison to VP9 and H.264, all while maintaining a consistent bitrate.

Screenshot comparing video codec quality for low bandwidth streams. Source: Meta Engineering Blog

Figure 11: H.264 vs VP9 vs AV1

Generally speaking, the Chrome browser and Android ecosystem handle AV1 well across phones, tablets, smart TVs and set-top boxes/streaming sticks. Unfortunately, the same cannot be said for Safari and iOS where support had been lacking until the iPhone 15 Pro announcement.

Chrome	Edge	Firefox	Safari	Android	Windows	iOS/macOS
✅	✅	✅	❌	✅	✅	❌

Figure 12: AV1 Software Support

Individual Contributions

Shaun Nguyen
- Project proposal
- Creating website
- Demo presentation
- Report: Technical Overview, Current Development

Tirth Patel
- Project Proposal
- Creating and maintaining websites
- Project Updates: Oct 23, Nov 6, Nov 20
- Project Demo
- Project Report: Topic Intro and background, Netflix implementation of Av1
- Formatting Final Report

References

[1] R. Trafford-Jones, "Video: AV1 at Netflix," *The Broadcast Knowledge*, Feb. 24, 2020. https://thebroadcastknowledge.com/2020/02/24/video-av1-at-netflix/#video (accessed Dec. 12, 2023).

[2] N. T. Blog, "SVT-AV1: an open-source AV1 encoder and decoder," *Medium*, Mar. 13, 2020. https://netflixtechblog.com/svt-av1-an-open-source-av1-encoder-and-decoder-ad295d9b5ca2 (accessed Dec. 12, 2023).

[3] Y. Chen *et al.*, "An Overview of Coding Tools in AV1: the First Video Codec from the Alliance for Open Media," *APSIPA Transactions on Signal and Information Processing*, vol. 9, 2020, doi: https://doi.org/10.1017/ATSIP.2020.2.

[4] "Netflix Research," *research.netflix.com*. https://research.netflix.com/publication/AV1%20at%20Netflix (accessed Dec. 12, 2023).

[5] J. Han et al., "A Technical Overview of AV1," in Proceedings of the IEEE, vol. 109, no. 9, pp. 1435-1462, Sept. 2021, doi: 10.1109/JPROC.2021.3058584. https://ieeexplore.ieee.org/document/9363937

[6] "AV1 Playback Support: Apple adds iPhone AV1 decoder [2023 Update]," *bitmovin.com*, Sep. 12, 2023. https://bitmovin.com/av1-playback-support

[7] "How Meta brought AV1 to Reels," Engineering at Meta, Feb. 21, 2023. https://engineering.fb.com/2023/02/21/video-engineering/av1-codec-facebook-instagram-reels/