Abstract

HEVC has introduced tools for parallel processing first time ever in a video compression standard. It has been found to be well received and appreciated by video coding experts. In this paper, we carry out the analysis of these tools to understand (in practicality) the effectiveness of achieving the purpose for which they are targeted. This paper will also help readers understand the intricacies involved in realizing such tools on various hardware platforms.

Introduction

High Efficiency Video Coding (HEVC/H265) is the most contemporary video compression standard, jointly developed by ITU-T VCEG and ISO/IEC MPEG standard bodies. HEVC, successor of H.264/AVC, is targeted to achieve half the bit-rate compared to its predecessor for the same video quality. Along with primary goal of achieving higher compression and accommodating support for higher resolutions like 8K, HEVC also embraced parallel processor architectures by introducing tools for parallel encoding/decoding within a frame.

H.264/AVC has slice feature, supporting parallel processing. Slices were introduced mainly for fast resynchronization in case of bit-stream error or packet loss during transmission. Though it can be used for parallel processing, single parallel coding option does not fit for all architectures & will not serve every application need. Employing slices for parallel processing has the following limitations,

- Decreases coding efficiency due to CABAC prediction dependency breakages at slice boundaries
- Causes load balancing issues in both encoder & decoder as size and complexity of slices may vary from picture to picture.
- Introduces parsing overheads in the decoder because of the need for identification of each slice entry point

In order to overcome these limitations, HEVC has included two more coding tools in addition to slices, namely TILES and Wavefront Parallel Processing (WPP), specifically designed to enable high level parallel processing.

In this white paper, we provide an analysis on the need for parallel coding tools and the forces that caused to introduce such tools HEVC. We also describe HEVC parallel tools in detail and discuss on realization approaches for these tools on various architectures. We cover the challenges, issues
and compromises involved in realizing them and the necessary care that needs to be taken into consideration while designing an efficient parallelization strategy for encoding or decoding. Along with describing compromises in video quality, this paper provides details on the applications that can take advantages of such parallel tools.

**Driving Forces for Parallel Coding Tools:**

Computational complexity of HEVC encoding/decoding is multi-fold compared to its predecessor H.264/AVC because of increased number of coding combinations. Due to increased complexity, real-time realization of HEVC codec on single core CPUs is not practical. Capability of single-core has reached the maximum over the previous years, paving way for multi-core architectures. For video resolution beyond UHD such as 4Kx2K and 8Kx4K, low-latency and real-time processing are beyond reality with single-core. Recent days, demands for software solutions are high compared to demands for hardware counterparts due to various factors such as cost, flexibility, feasibility and other parameters. In order to overcome above mentioned issues, along with other reasons, and to make best use of multi-core architectures, HEVC introduced parallel coding tools known as TILES and WPP.

![Fig.1 Rapid Growth in Higher resolution support by Video capture & Display Devices](image1)

![Fig.2 Ubiquitous Video World](image2)
HEVC Parallel Tools

The parallel tools introduced in HEVC are Wavefront Parallel Processing (WPP) and Tiles. Both tools allow subdivision of each picture into multiple partitions that can be processed in parallel. Each partition contains integer number of Coding Tree Blocks (CTB) that may or may not have dependencies on CTBs of other partitions. When WPP or Tiles are enabled, bit-stream contains entry point offsets that indicate the start position of each picture partition which is necessary for each core to immediately access the partition.

1. TILES:

Tiles are aimed at achieving parallelism without impacting much on the video quality. Enabling tiles allows picture to split into rectangular regions (sub-set of picture), which aid in independent encoding or decoding of the current picture. Picture can be split uniformly or non-uniformly as tiles. However, the number of tiles in vertical and horizontal direction in a picture is constrained by level signaled in bit-stream. Entry point for each tile is sent in the slice header. HEVC allows coding schemes such as, single slice - multi tile, multi slice - single tile, single slice – single tile and multi slice – multi tile. A Maximum of 20 tile columns and 22 tile rows are allowed as per the HEVC specification and HEVC also limits the minimum width of tile to 256 & height to 64 pixel lines. Though tiles help in achieving load balancing, enabling loop filter across tiles might pose a threat. Resetting entropy coding at the start of each tile and breaking of dependencies such as intra prediction, motion vector information and SAO may introduce visual artefacts across the tile boundaries and deteriorate the coding efficiency achieved by encoder. Below diagram demonstrates division of picture into tiles.
2. Wavefront Parallel Processing (WPP):

Wavefront Parallel Processing (WPP) is another major boost for achieving parallelism. WPP achieves higher compression and less visual artifacts than TILES, since WPP doesn't break the coding dependencies as TILES do. WPP splits the slices into CTU rows and encoding or decoding of each row is dependent on the previous CTU row. WPP introduces parallelism in entropy coding in which each row can be encoded or decoded once two CTU's in the
previous row are processed, and each row is processed in the raster scan order. CABAC is terminated at each row to enable parallelism. HEVC, equipped with WPP, exploits the multi-core architecture to the best extent compared to the utilization of multicores by H.264\AVC. Since CABAC context is inferred from the previous row, higher synchronization among cores plays a vital role. Inter-core communication and synchronization are the major drawbacks in WPP compared to tiles. But recent advances in architectures, which make inter-core communication & synchronization ease, are suitable for WPP.

**Implementation Challenges**

Irrespective of all the advantages that were brought into the table by HEVC parallel tools, practical implementation of these tools on real time HEVC encoder/decoder solutions come with a great set of design challenges. Few of them are very specific to encoder while others are common for both encoder and decoder implementations. Constraints brought in by these parallel tools vary across different implementations based on requirement, design and architecture of interest. In the following section, we discuss various design and development issues that arise while leveraging HEVC parallel tools on an encoder/decoder solution.

**A. Coding Efficiency:**

Coding Efficiency is the ability to minimize bit rates necessary for representation of video content to reach a given level of video quality—or, as alternatively formulated, to maximize the video quality achievable within a available bit rate.[1] WPP and Tiles degrade coding efficiency as they both break the sequential nature of video coding.
Basically there are two main reasons for reduced coding efficiency caused by these parallel tools.

i) Due to unavailability of neighbor data at tile boundaries. In case of Tiles, as picture is divided into independently decodable units, neighbor data will not be available for the blocks at tile boundaries even if they belong to same slice. This can greatly affect the efficiency of intra prediction as reference pixels will not be available. In inter coding, formation of merge list would be handicapped due to unavailability of neighbors - thereby increasing bits for Motion Vector coding.

ii) Due to discontinuity in context for Context Adaptive entropy coding. HEVC adapts Context Adaptive Binary Arithmetic Coding (CABAC), which uses probability model that depends greatly on previously coded content. At the beginning of every Tile, context has to be reset as it does not depend on previously coded data of the picture. In case of WPP, context for every row gets initialized from previous row data, which has no knowledge of rest of the picture coded.

Experimental result shows that there is approximately 2% loss in BD-Bitrate due to introduction of WPP or Tiles. And it is also observed that, at very low bit-rates, tile boundaries can be visible in reconstructed pictures.

B. Memory:

Generally, parallel tools are used when multiple homogeneous or heterogeneous processing resources are available. When parallel tools are introduced, common data used by different instances, especially reference data used for motion compensation, has to be replicated at every core, which increases memory footprint and bandwidth.

C. Scalability:

Scalability is a requirement for every multi-core solution but is also an issue while using these parallel tools. In multi-core architectures with huge number of available cores, it is not advisable to divide picture into so many tiles as it may adversely affect the video quality. WPP poses a different problem at low resolutions when the number of available cores are more than the number of CTU rows. These reasons are enough to make the solution unfriendly in situations where the numbers of available cores change dynamically.

D. Core Utilization & Load Balancing:

Core utilization indicates efficiency of the multi-core design. When using WPP, every row should wait for 2 CTUs encoding/decoding of previous row for CABAC context initialization. It is observed that, at pipeline build-up stage, this delay propagates as number of CTU rows executing simultaneously on different cores grows, thereby
leading to increased core idleness. Imagine a situation with low resolution- it is possible that when last core starts processing a row, first core would have finished processing the first row and will be idle if there are no rows to be processed. Also, at pipeline ramp-down stage, different rows might take different time for processing there by introducing core idleness in some of the cores. Similar problem of load balancing arises when tiles are used as it is possible that different tiles need different amount of time for processing respective data.

E. Slice Header Preparation Delay:

When using any of the parallel tools in HEVC, it is needed to encode offset in bit stream for each independent unit, which brings in a new constraint for preparation of slice header. Due to this, slice header preparation/encoding has to wait until all CTUs of the slice are encoded. This can be a problem for emulation prevention especially in low delay cases where partial frame data needs to be passed to application for reducing delay. Other option is to create a new slice for every independent unit, which will lead to drop in coding efficiency.

F. Reconstructed Image Stitching:

In Multi-core systems with distributed memory, reconstructed image stitching becomes a major bottleneck for improving the execution speed. Different parts of the picture, which needs to be stitched together for referencing and displaying, are processed by various cores. This can also increase core idleness as cores will generally be idle when reconstructed image is moved into a common memory.

G. Rate Control:

Rate control is responsible for maintaining desired bit-rate for encoded data. In encoder solutions, it is important to control the bits consumed within the picture at low bit rates and to maintain constant bit rate in low delay applications. This requires feedback from entropy coder about already consumed bits. In case of WPP, every core independently processes each row where it is impossible to get feedback from entropy coder. Similarly, different tiles are unaware of the status of other tiles leading to lack of information for implementing Rate Control in Sub-Frame level.

H. MTU size constraint

For applications having a hard constraint on MTU size, it is pretty difficult and challenging to match the MTU size when these two parallel tools are used.

In case of WPP, there is a complication where the slices starting in the middle of a row has to end in the same row. Similarly in multi-tiles case, the constraint of slice CTUs being limited to a single TILE makes it difficult for meeting the MTU requirement.
Applicable Areas

These parallel tools usage is almost inevitable for real-time applications that use HD and/or beyond HD resolutions. So, UHD applications gets the advantage of achieving real-time performance using these tools. And, applications that demand loss less coding can take advantage of these tools due to the higher complexity of entropy module.

Conclusion

As current trend is moving towards usage of multi-core architecture, there is greater need for parallel processing tools such as tiles and WPP. This paper can be used to understand necessity of Parallel tools, challenges involved in designing encoder/decoder solutions that use these tools and compromises that should be made for implementation. This can be a major boon in multi-core solutions for achieving real time performance if implemented with care.

References

PathPartner Involvement in Multi-Core HEVC Development

PathPartner has extensive experience in the field of video encoders & decoders. Design, development of video codecs such as MPEG-2, MPEG-4, H264, HEVC, AVS, VP6/7, RV and VC-1 etc. on various platforms became essential, as multiple architectures & standards are available & used by end users/devices. Porting of video codecs involves greater challenges on multi-core platforms. Multi-core platforms are of various kinds like homogeneous (symmetric/asymmetric) general purpose multi-cores, hardware accelerator multi-cores or heterogeneous architectures. The key challenges for both design & development of video codecs vary with each architecture/platform. PathPartner has significant experience in design & development of almost all video codecs ranging from proprietary to public standards and on architectures ranging from simple to very complex multi-core platforms. PathPartner provides licensable codec IPs along with expertise services for both porting & optimization of all video, audio & speech codecs. PathPartner is currently active in development of HEVC video codec (both encoder & decoder) on multi-core platforms such as Cortex-A15, CPU+GPU and x86. PathPartner has readily available HEVC decoder multi-core solution on Cortex-A15, which runs real-time for 1080p60 and is scalable upto 4k. PathPartner's PC based HEVC encoder solution (developed in modular way for easy porting) has achieved video quality that is comparable with HM encoder. Also, PathPartner has licensable FPGA solution for HEVC decoder 1080p30. Recently we completed the implementation of HEVC encode & decode real-time solution for UHD on TI’s c66x DSP based multi-chip platform.

About PathPartner

PathPartner based out of California, USA and Bangalore India is a leading provider of Consulting, Services and Solutions for digital media centric devices market. Our services range from Product Engineering/R&D and System Integration to Middleware, Applications and System solutions

With an expert management team which has rich experience in Technology, Engineering & Business practices, PathPartner has Sales & marketing presence in USA, Europe, Korea and Taiwan. The company specializes in addressing challenges faced by leading OEMs, Silicon and OS providers in their product development

http://www.pathpartnertech.com/
sales@pathpartnertech.com