Chapter 1

Introduction

Visual information have been an integral part of human lives since the inventions of the photographic film, image capturing technologies, video recording devices and television systems in the early 20th century. The emergence beyond HD formats (eg., 4K, 8K) and its growing popularity have increased the diversity of services. It is evident that the importance of video applications in our daily life has increased significantly. Entertainment, internet, broadcasting are only a few of the primary consumers in which huge demand of video contents can be found in the form of TV broadcasting, digital cinema, internet streaming, 3D-video, etc. In this modern era, a major source of entertainment is online TV like YouTube, Amazon Prime, Netflix, etc. Almost 86% of the internet audience is watching online video [1] and more than 500 hours of video is uploaded to YouTube in every minute [2]. The massive boost up of consumer video has been observed in mobile devices which are going to dominate the demand of video in the future. Indeed, the average person on the street has become an amateur video processor. The demand for high quality video in portable devices such as smart-phones, laptops, tablets, cameras, etc. is increasing with time.

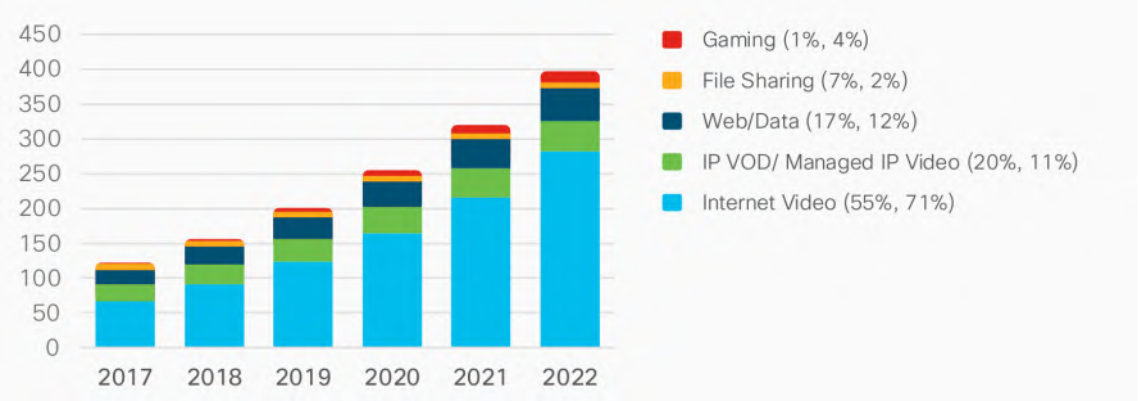

FIGURE 1.1: IP traffic consumed by different applications [1]

Video applications are becoming popular in surveillance, as safety and security are becoming the primary concern. By the year 2022, video surveillance traffic on the internet will jump to seven times as compared to that of 2017 [1]. Video applications are becoming imperative for surveillance of public places, national security, threat detection and object tracking. Emergency services like ambulance, fire control, police are also very much dependent on the video applications.

Video conferencing, advertising, vision-based control system are playing important roles in the field of business and automation industry. Not only that, video applications are extending their area towards the field of health care, automobile, education, etc. Artificial intelligence, machine learning applications, internet of things, the augmented reality are adding a new dimension in those video applications. The momentum achieved by video technology is so high that it is almost impossible to predict its future.

A major portion of internet traffic is consumed by visual data in the current time, which is an indication of popularity of the video applications. The graph in Fig. 1.1 represents the trends of internet video in the coming years as per the survey carried out by Cisco. According to the annual Cisco Global Cloud Index (2017-2022), the

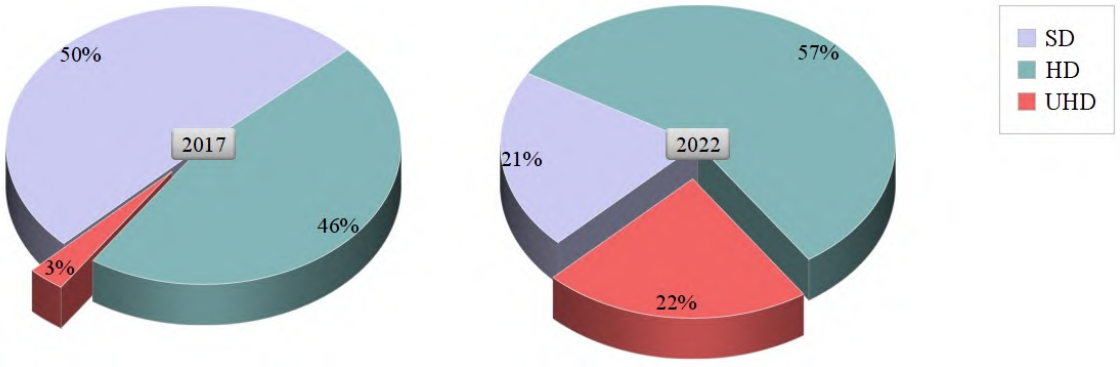

FIGURE 1.2: Percentage of global video traffic consumed by different video resolutions [1]

wireless and mobile data traffic will be 71% of total IP traffic by 2022, and 82% of this total IP traffic will be due to video applications.

Additionally, introduction of Ultra-High Definition (UHD) has made it more pronounced. Users are demanding increased video quality, increased quantities of video content, more extensive access, and better reliability. Fig. 1.2 shows the trends of UHD video in terms global video traffic percentage. By the year 2022, 62% of the flat-panel TV will support UHD streaming in which bit-rate required is more than double the HD video bit-rate and nine times more than Standard-Definition (SD) video bit-rate. This is creating a major tension between the available capacity per user in the network and the bit-rates required to transmit video content at the desired quality. Hence, video coding is emerging as a popular research topic in recent time.

The progress in video technology was started more than hundred years ago when the first ever motion picture called 'The Horse in Motions' was made in 1878 [3]. The first video with color was produced in 1905 and the videotape in 1951 [3]. The revolution in the video applications was possible only after the introduction

of digital video. Analog video, used in the early era, was difficult to process and store. Easiness of handling digital video is far better than analog video as storing and processing are simpler. Therefore, the discussion in this thesis is restricted to digital video only. Digital video is a sequence of digital images which are known as frames. The number of frames displayed per unit time is known as the frame rate and it is measured in frames per second (FPS). Every frame is a collection of pixels which are represented by a fixed number of bits. The quality of video depends proportionally on the number of bits used to represent a pixel and FPS.

A raw, uncompressed digital video requires large data to be stored or transmitted. For example, an uncompressed high definition video (i.e., 1980×1080 pixels per frame) with 24 FPS of 1.5 hours duration requires 806 GB to store and 1.2 Gbps bit-rate to transmit. The 4G mobile (4×4 LTE) supports maximum bandwidth of 326 Mbps [4]. Clearly, raw video data is expensive, and data handling for digital video is a major challenge. Hence, video codec is an essential tool for every video application to effectively utilize the communication bandwidth and storage area.

Video codec refers to an encoder as well as a decoder, as shown in Fig. 1.3. The encoder converts the raw video sequences to bit-stream after applying various compression algorithms. Those compression algorithms remove redundant information and reduce the total number of bits to be stored or transmitted for efficient communication. Generally, video signals deal with four types of redundancies.

FIGURE 1.3: Structure of general encoding and decoding system (CODEC)

- **Temporal redundancy**: It is the similarity between successive video frames and inter-prediction block is used to remove this redundancy it.
- **Spatial redundancy**: It is the amount of similarity present in pixel values of the same video frame and intra-prediction block is used to remove it.
- **Perceptual redundancy**: Human eyes can not sense abrupt changes in the video sequence. This limitation of human eyes is referred to as perceptual redundancy. Transform and quantization block is used to remove perceptual redundancy.
- **Statistical redundancy**: It is the amount of similarity present in the data-stream. Entropy coding block is used to remove it.

The decoder performs an exactly inverse process to that of the encoder and restores the video signal. Video coding may be either lossy and lossless. In lossy coding, redundant data are pruned forever and cannot be recovered during decoding. In lossless coding, the decoded video is exactly the same as that of the original, and no information is lost.

Codecs are based on the mathematical principles of information theory. The primary motive of any video codec is to encode a video using minimum number of bits with high perception quality. However, a practical codec always tries to maintain a balance among the following trades-off depending on the field of applications.

- **Quality** vs **Bit-rate**: The number of bits used per second to encode a video sequence is known as bit-rate. A high-quality video requires high bit-rate. On the other hand, the restriction on bit-rate is achieved by compromising video quality. For a particular bit-rate, the quality of video depends on the efficiency

of the video coding standard. In this regard, HEVC/H.265 achieves the best result to date.

- **Complexity** vs **Efficiency**: In general, the efficiency of the video codec always achieved at the expense of algorithmic and architectural complexities. Codecs with more features have more efficiency as well as high complexity.
- **Latency** vs **Quality**: Algorithm complexity increases with efficiency. Hence, a more efficient algorithm incurs more implementation cost as well as more delay.

The primary goal of any codecs is to achieve more compression efficiency as compared to its predecessor without any sacrifice in the perceptual quality. To achieve that, each video standard introduces more complex algorithm which increases the hardware complexity in proportion. Different varieties of video codec have been invented in last few decades. Among those, High Efficiency Video Coding (HEVC) is the latest video coding standard proposed by the Joint Collaborative Team for Video Coding (JCT-VC). Efficiency of the HEVC is the best till date. But, its architectural complexity is also large.

1.1 Brief History of Video Codec Standards

A standard for video coding is essential for inter-operability. The need for video coding standard was realized about three decades ago. Two major organizations, International Telecommunication Union (ITU) and the International Organization for Standardization (ISO) are involved from the beginning of video codec standardization. The evolution of video coding standards with time is presented in Fig. 1.4. In the year 1984, ITU introduced the first video compression standard, namely

H.120 [5]. However, it was neither efficient nor good enough for practical uses. The first compression standard which had some practical uses was H.261 [6] introduced by Video Coding Expert Group (VCEG) of ITU Telecommunication standardization sector (ITU-T) in 1988. This was the first block-based hybrid compression algorithm using a combination of transformation, temporal Differential Pulse Code Modulation (DPCM) and motion compensation.

ISO also started working on video standard in parallel to ITU and Moving Picture Expert Group (MPEG) was founded in 1988 [7]. MPEG delivered its first ever video coding standard in 1992 followed by MPEG-2 in 1994. Target of MPEG-2 was to support broadcasting of digital video. ITU recommended it as H.262. In the year 1995, ITU-T proposed the H.263 standard [8] to support video conferencing, internet, mobile telephony, etc. H.263 has different versions and there are many annexes.

The MPEG-4 standard [9] was developed by introducing some new features as well as adopting many features from related standards. This standard supports different

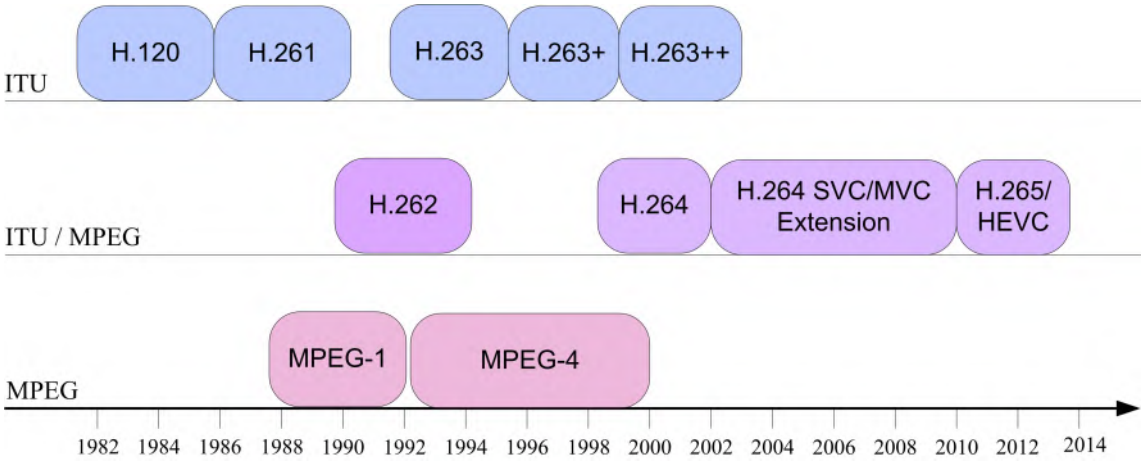

FIGURE 1.4: Evolution of video coding standards [4]

profiles which provide various tool sets for the use in different applications.

H.264 or Advanced Video Coding (AVC) [10] is perhaps the most popular and widely used video coding format. It was developed under the partnership between ITU-T and ISO/IEC known as Joint Video Coding Team (JVT) to provide good video quality at lower bit-rate than its predecessors MPEG-2, MPEG-4 and H.263 standards. It has been adopted in many video coding applications such as set-top boxes, Blu-ray, HD-DVD which support HD resolutions as well as in TV broadcasting standards.

HEVC or H.265 [11] is the latest video coding standard and is expected to be adopted in future multimedia products. It is capable of reducing the bit-rate by 50% to its predecessor for same perceptual quality [11]. This directly implies that better video quality can be delivered over the same bit-rate or bit-rate can be halved to attain the same video quality. The first version of HEVC was published in April 2013 by JCT-VC. The JCT-VC committee then proceeded to develop extensions of HEVC, namely Format Range Extensions (RExt), Scalable HEVC (SHVC) and Screen Content Coding (SCC). Meanwhile, the multiview (MV-HEVC) and 3D (3D-HEVC) video coding extensions of HEVC were developed by another committee, namely the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V).

1.2 Motivation and Problem Description

Use of better quality with less distortion is the recent trend in visual technology. As a result, video resolution and frame rate are increasing quadratically. However, bandwidth and the storage space are limited and are the main constraints to cope-up with growing demands of high-quality video. Video coding standards are introduced for bandwidth savings and so as to reduce storage space with minimum distortion.

TABLE 1.1: Performance comparison of different video codec for interactive applications [13]

Algorithm	Bit-rate reduction compared to			
	H.262	H.263 HLP	MPEG-4 ASP	H.264/MPEG-4
H.263 HLP	27.8%	-	-	-
MPEG-4 ASP	37.4%	13.2%	-	-
H.264/MPEG-4	67.0%	54.1%	46.8%	-
HEVC MP	80.1%	72.3%	67.9%	40.3%

Till date HEVC achieves the best results among the existing video coding standards [11].

HEVC/H.265 is the latest video coding standard jointly developed by the ITU and ISO. A reference software, called the HEVC test model (HM) [12], is being developed along with the draft of the standard. It supports three encoder configurations known as All-Intra (AI), Low-Delay (LD), and Random-Access (RA) configuration. It is expected that HEVC will achieve on an average 50% bit-rate saving in comparison to H.264/AVC for the same video quality [11]. A comparative study of efficiency measured among different video codecs is detailed in [13] and results are presented in Table 1.1.

There are challenges to use HEVC in video applications. Among many more, the main challenge to use HEVC in video applications is its complexity. A large number of additional features are used to achieve high coding efficiency in HEVC. These features simultaneously increase the complexity of HEVC encoder as well as decoder. It has been observed that HEVC encoder is several times more complex than H.264/AVC encoder [14], whereas complexity of the decoder is almost double than that of the H.264/AVC decoder [15].

The high hardware complexity is a major concern for the most efficient video standard. Complexity optimization of HEVC algorithm is imperative, especially for smart handheld battery operated portable devices. Smart portable devices have brought a revolution in the field of information technology and the demand for video applications in these devices is experiencing phenomenal growth. According to the annual Cisco Global Cloud Index (2017-2022) [1] the wireless and mobile data traffic will be 71% of total IP traffic by 2022. Therefore, the video codec with high efficiency and lesser complexity is must for such portable devices. However, the hardware design for HEVC with optimized area and power is a challenging task.

Video codec in portable devices need to satisfy the following requirements in addition to bit-rate reduction:

- Low complexity: The area and the hardware complexity should be minimum.
- Low delay: Processing time should be minimum to support real-time operations.
- Low power: Minimum power consumption is expected to maximize battery lifetime.
- High throughput: Must have maximum pixel processing ability to support high resolution video.

There is a very good scope of hardware optimization in HEVC standard; both in algorithmic and architectural level. In algorithm level, computational complexity can be significantly reduced by adopting suitable approximation schemes. A study has been carried out on HEVC profile to show the time spent on different functions while using HM software [14] and a portion of the results are presented in Table 1.2. It

TABLE 1.2: Time distribution among different classes in HM 8.0 for sequence BasketballDrive_1980 × 1080 with QP=27 [14]

Different Classes	Encoder configuration	
	All-Intra	Random Access
TEncSearch	11.8	7.4
TComTrQuant	24.4	10.7
TComRdCost	9.8	38.8
TComInterpolationFilter	0.0	19.8
TComYUV	0.1	1.7
partialButterfly	8.7	4.0
TComDataCU	5.8	2.7
TEncSbac	8.4	3.5
TEncEntropy	1.2	0.6
TEncBinCABAC	2.2	0.9
TComPrediction	10.0	1.1
TComPattern	6.6	0.4
memcpy/memset	11.0	7.1

has been observed that a significant amount of time is spent in transform and quantization function (TComTrQuant class in HM) where Rate-Distortion Optimized (RDO) quantization takes place. For example, about a quarter of the total encoding time is spent in the TComTrQuant class in the AI configuration as shown in Table 1.2. TComTrQuant class uses higher order Discrete Cosine Transform (DCT), Inverse DCT (IDCT). Therefore, the efficient design of DCT and IDCT architectures is essential for encoding and decoding a video sequence. Furthermore, it is possible to reduce area and power requirements with an effective approximation of DCT and IDCT architectures.

1.3 Related Works

The efficiency of the video, as well as the image coding algorithm, greatly depends on the transform coding process. As a result, relentless efforts can be observed in the literature to discover an effective transform coding algorithm. Different methods of transform coding such as the Hadamard transform [16], Haar transform [17], Fourier transform [18] and several discrete trigonometric transforms [19], [20] have been explored [21–24]. Among those, it can be observed that a significant amount of research is dedicated to DCT of type II and III [25].

DCT was first ever introduced by Ahmed et al. [26] in 1974. After that, type II and III DCT have become the most popular as forward and inverse transform in the field of image and video coding. This is because they closely approximate the Karhunen-Loeve Transform (KLT) [27] which can exploit the correlation in stationary Markov-I signals [20]. At the same time, DCT is more feasible for hardware implementation as it doesn't depend on the characteristics of the input signal. Due to this efficient DCT algorithm design has been a major scientific effort in the circuits, systems, and signal processing community.

In general, two dimensional (2D) DCT is used in video as well as image compression. Hardware architecture of the DCT does not have regular structure as that of the Fast Fourier Transform (FFT). Hence, several architectures are proposed to implement it efficiently. The 2D DCT can be implemented in hardware using either direct [28–30] or indirect methods [31–34]. The direct method is characterized by high performance but complicated implementation structure. It demands high hardware cost, power consumption and greater design efforts. As a consequence, a very few architectures are proposed in the literature using direct method approach.

In the indirect method, row-column decomposition approach is used to compute 2D transforms. Two 1D transform cores are used with a transpose memory to compute row and column transform in full parallel architecture. In folded architecture, row and column transforms are computed sequentially by using the same 1D transform circuit with some transposition logic [35]. The indirect method of 2D DCT computation is worthwhile over direct computation in terms of hardware efficiency. As a consequence, hardware cost and power consumption associated with this type of architecture are less as compared to that of the architectures by direct implementation method.

Hardware complexity and compression efficiency are proportional to the order of DCT. The lower order DCTs are vastly applied for image compression. Several methods of 8-point DCT computation have been proposed, such as the Lee DCT factorization [36], Arai DCT scheme [37], Chen's factorization method [38], and the Loeffler DCT algorithm [39]. Among these, the technique introduced by Chen has been frequently considered in the development of architectures for the computation of one dimensional (1D) transforms.

Higher order DCTs are becoming popular in the most recent video coding standard like AVS [40], HEVC to handle high and ultra high definition video. Larger size DCT provides better energy compaction and thereby reduces the bit-rate. However, this coding efficiency is achieved at the cost of increased complexity and latency. This poses several challenges to codec and system designers, especially when the real-time operation is demanded.

One of the reasons behind the high complexity in DCT architecture is the use of large number of multipliers which incur high hardware cost. From this perspective, CORDIC based approach [41] is widely used for the multiplier less low-power DCT architecture design. Since the CORDIC algorithm leads to a very regular

structure suitable for VLSI implementation, it has been widely used to implement DCT[42–46]. Other designs use Distributed Arithmetic (DA) [33, 34] to avoid the multiplication computations. This approach uses look-up tables and accumulators instead of multipliers to compute inner products and has been widely used in many DSP applications such as DFT, DCT, convolution, and digital filters. Other than CORDIC and DA, another effective method of complexity reduction is the approximation technique which actually compromises with the accuracy of the DCT output.

The approximation method provides an alternative solution to reduce computational complexity of the DCT [20]. However, a meaningful estimation of DCT is necessary with reduction in computational complexity. Several research papers are published on the approximated DCT architecture which demonstrate the complexity and accuracy trade-off according to the field of applications. These papers use different techniques to approximate a DCT matrix. Some of the prominent methods of approximations are as follows:

- Some of the researchers [47, 48] use functions such as signum, ceil and floor, to approximate the coefficients of the exact DCT matrix.
- Scaling and rounding-off is used in large number of articles [49–52].
- Pruning technique is used in [53].
- Brute force computation over a reduced search space is used in [54].
- Derivation of higher order DCT based on approximated lower order DCT is proposed in [55].
- Single variable matrix parameterization of existing approximations is shown in [56].

- Multiplier-free DCT approximation for radio-frequency multi-beam digital aperture-array space imaging is presented in [57].

In general, the selected coefficients in the very low complexity approximated DCT matrices are defined in the set $\{0, \pm\frac{1}{2}, \pm 1\}$. For example, the coefficients in the signed DCT (SDCT) matrix proposed in [47] are obtained by replacing the original coefficient by their sign. The Bouguezel–Ahmad–Swamy (BAS) series of algorithms [55, 56, 58–61] proposed several DCT matrices by replacing coefficients with 0, 1/2, 1. Similarly, Cintra and Bayer [54, 62] have proposed two transforms derived from 0 and ±1 as coefficients of the transform matrix. Thus, such transformations possess null multiplicative complexity because the required arithmetic operations can be implemented exclusively by means of binary additions and bit-shifting operations. Effectively, DCT approximations have been considered for applications in real-time video transmission and processing [63, 64], satellite communication systems [20], portable computing applications [20], radio-frequency smart antenna array [57], and wireless image sensor networks [65].

HEVC uses variable block size higher order DCT for energy compaction and its size varies from 4×4 to 32×32. High -order and -precision DCT reduces the bit rate up to 10% [66]. However, higher order DCT requires a large number of high precision multipliers to reduce encoder-decoder mismatch and drifting error. This increases the computational complexity as well as the hardware cost significantly. Several works have been proposed earlier to derive DCT like transform which can reduce computational complexity and drifting errors simultaneously. Among them, the most popular approximation is Integer Cosine Transform (ICT) [67] in which a large number scales real-valued DCT coefficients and then those are rounded to the nearest integer. ICT has almost similar coding performance as that of the real-valued DCT and its hardware implementation requires less number of resources.

Such type of integer DCT is proposed and adopted in HEVC core transform [66]. Its hardware implementation presented in [68] uses Multiple Constant Multiplications (MCM) [69] which obviates the need for multipliers. The complexity of this architecture is further reduced by exploiting the symmetry and anti-symmetry properties of the DCT matrix using Chen's algorithm [38]. The HEVC encoder requires Walsh Hadamard Transform (WHT) to determine the Sum of Absolute Transformed Differences (SATD) in the prediction unit. Architecture proposed in [70] exploits the relationship between WHT and DCT. Such type of architecture uses precomputed WHT results from the prediction unit to calculate DCT. However, it requires a large number of rotation units which increase the hardware cost and delay. Masera et al. [71] proposed an approximate method which dynamically skips some rotations depending on the characteristics of the input signal.

In HEVC, DCT and entropy coding are recursively performed in each quadtree depth level during RDO process. It is observed that RDO process [72] consumes a major portion of the power and hardware resources. It means, optimizing the RDO process can significantly reduce the power consumption and complexity of the encoder. A way of optimizing RDO process is by approximating forward- and inverse- DCT. Many approximation techniques have already been proposed in [73, 74] which are very popular for image compression. However, most of them are not devised for video compression as it demands more accuracy than image.

Few more architectures [75–85] are proposed for HEVC. These architectures can be distinguished in to three categories. Some of them are DCT designs, whereas others are either IDCT or unified designs. The unified architectures shares same hardware for forward and inverse transforms to reduce area overhead. The common thing about these all architectures is they all try to reduce power and area overhead. Different innovative design methods has been introduced to reduce power and area.

Few of the designs have applied approximation techniques to reduce the computational complexity. However, those sacrifice accuracy and hence, quality of the video degrades.

1.4 Thesis Objectives

This thesis focuses on the design and implementation of higher order transform coding used in HEVC and its assessment. Transform coding is essential compression tool which is associated with the image and video coding algorithm from the beginning. The objective of the transform coding is to convert the information from spatial domain to frequency domain. In frequency domain, most of the high-frequency components are discarded because the human visual system is less sensitive to high-frequency components. In this context, Discrete Cosine Transform is the most preferred transform to convert the information from spatial domain to frequency domain. Almost all the video codecs use DCT and IDCT for data compression and decompression, respectively. However, HEVC uses larger size of DCT and IDCT as compared to the earlier codecs. This feature with many more increase the efficiency and the complexity of the HEVC standard. Hence, the objectives of this thesis are as follows:

- To study and understand the operating principle of different video coding standards, how HEVC differs from these standards and find out which is the most critical task performed in it.
- In HEVC, transform coding plays vital role in achieving high compression efficiency. So, to study recent advances in the transform coding operation performed in HEVC.

- As stated earlier, complexity of the transform coding is high in HEVC standard. Therefore, it is desirable to propose a transform coding architecture which maintains low hardware complexity and high coding efficiency.

- The purpose of HEVC is to support ultra high resolution video. Therefore, it is desired that the transform coding architectures must have high throughput and low latency in addition to the hardware complexity reduction. So, to proposed an architecture to cater this requirement.

- All the proposed architectures must comply the requirements mentioned in the HEVC draft. Therefore, it is necessary to assess the performance of the developed coding solution and check its utility in the video coding context.

1.5 Contribution and Scope of the Work

The aim of this thesis is to investigate the efficient transform for HEVC, its assessment and architectural implementation. So, this thesis proposes few efficient DCT architectures which reduce the hardware complexity with negligible coding loss. An approximation scheme for DCT as well as IDCT algorithm is also proposed in this thesis which significantly reduces hardware complexity of HEVC core transform architecture. An effective utilization method of the complex higher order DCT architecture has also been proposed to increase the throughput and to reduce the latency of HEVC transform architecture. This work focusses only on the transform coding stage. All the other stages were retained as specified in the HEVC standard.

The accuracy and the video encoding performance of all the proposed as well as existing methods are measured using HEVC reference software model [12] of version 16.15 (i.e., HM-16.15). It was the latest version when we started the work. All

the hardware architectures are described using Verilog hardware description language. Synopsys Design Compiler and Xilinx Vivado Design Suite version 2016.2 are used for hardware implementations. Additionally, Isim simulator along with the Mathworks MATLAB software are used for simulation and verification purposes.

1.6 Thesis Outline

This thesis is organized into six chapters. In this chapter, we have introduced the video codec and its purpose, a brief historical overview of video codec, literature review, the motivation of the work performed, contribution and the scope of the work.

Chapter 2 focuses on the definition of the terms and parameters associated with video coding technology. To create a strong background of the reader, a quick overview of the video coding process has described. Brief description of the HEVC standard has also been provided in this chapter.

In Chapter 3, transform coding used in HEVC is discussed in detail and its performance assessment in terms of coding efficiency and hardware complexity has been studied. The integer transform has been introduced and the advantages as well as disadvantages are detailed. An alternative to integer transform has been proposed in this chapter to reduce complexity of the HEVC core. The coding performance of the proposed transform unit has been measured with the help of HEVC reference software. The proposed design is implemented on ASIC as well as FPGA platforms to assess the hardware cost and operating speed, and results are compared with the existing designs.

In Chapter 4, the approximate designs of the HEVC transform architectures are discussed and assessed in terms of coding performance and hardware complexity. A new method of WHT architecture is proposed. At the same time, a new approximation scheme is proposed and its performance is analyzed. It has been observed that the new approximation scheme significantly reduces the hardware complexity with little sacrifice in coding efficiency.

In Chapter 5, a new method to increase the throughput of the transform coding architecture by hardware sharing technique is proposed. The proposed hardware sharing method utilizes all the hardware resources efficiently, hence the complexity of the architecture remains almost the same. However, it increases throughput and reduces latency, which is essential for UHD and other higher resolution videos.

Finally, conclusions are drawn in Chapter 6. The major findings and the main contributions of this work are briefed in this chapter. A discussion on the future work in the related areas is also presented in this chapter.