

A High Performance and Low Power Hardware Architecture for H.264 Transform Coding

Jubli Kashyap, Virendra Kumar Yadav

Abstract— In the search for ever better and faster video compression standards H.264 was created. H.264 promises to be an excellent video format for use with a large range of applications and need for hardware acceleration of its very computationally intensive parts. To address this need, this paper proposes architecture for the discrete transform (DCT) and quantization blocks from H.264. The first set of architectures for the DCT and quantization were optimized for power, which resulted in transform and quantizer blocks that use 10.5623 mW Power. All of the designs were synthesized for Cadence BuildGate Synthesis CMOS technology, as well as the combined DCT and Quantization blocks went through comprehensive place and route flow.

Index Terms—CMOS Technology, DCT, H.264, JVT, ITU-T, SoC, Quantization, YUV System, Zero Shift.

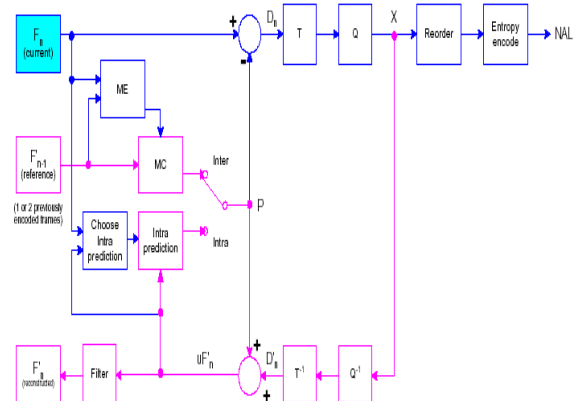
1 INTRODUCTION

Due to the remarkable progress in the development of products and services offering full-motion digital video, digital video coding currently has a significant economic impact on the computer, telecommunications, and imaging industry. This raises the need for an industry standard for compressed video representation with extremely increased coding efficiency and enhanced robustness to network environments. Since the early phases of the technology, international video coding standards have been the engines behind the commercial success of digital video compression. ITU-T H.264/MPEG-4 (Part 10) Advanced Video Coding (commonly referred as H.264/AVC) is the newest entry in the series of international video coding standards. It was developed by the Joint Video Team (JVT), which was formed to represent the cooperation between the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) [3]-[5]. Compared to the currently existing standards, H.264 has many new features that makes it the most powerful and state-of-the-art standard. Network friendliness and good video quality at high and low bit rates are two important features that distinguish H.264 from other standards. The usual 8x8 DCT is the basic transformation in H.264. This eliminates any mismatch issues between the encoder and the decoder.

The demand of multimedia communications on mobile and portable applications is growing nowadays. To realize multimedia communications, implementing a video compression standard is essential in any multimedia processing system-on-a-chip (SoC). There have been reports on the very large-scale integration (VLSI) implementation of MPEG-4 video recently. The emerging efficient H.264 or MPEG-4 Part 10 standard can greatly reduce the bandwidth and storage requirements for multimedia data. The VLSI implementation of H.264 is a challenge since an

H.264 baseline decoder is approximately three times more complex than an H.263 baseline decode. Implementational flexibility is an important factor of concern for SoC designs. Since the traditional hardwired design is less flexible, the processor-based implementation is a preferred choice. VLSI implementation can be categorized into three types, hardwired, digital-signal-processor-based, and hybrid. To achieve higher performance with flexibility, the hybrid architecture has been proposed.

2 DESIGN REQUIREMENTS OF THE H.264 TRANSFORM



Block Diagram of H.264 Encoder

2.1 Discrete Cosine Transform

The DCT is conceptually similar to the DFT, except:

The DCT does a better job of concentrating energy into lower order coefficients than does the DFT for image data. The DCT is purely real, the DFT is complex (magnitude and phase). The DCT is purely real, the DFT is complex (magnitude and phase).

A DCT operation on a block of pixels produces coefficients that are similar to the frequency domain coefficients produced by a DFT operation. An N-point DCT has the same frequency resolution as and is closely related to a 2N-point DFT. The N frequencies of a 2N point DFT correspond to N points on the upper half of the unit circle in the complex frequency plane.

Assuming a periodic input, the magnitude of the DFT coefficients is spatially invariant (phase of the input does not matter). This is not true for the DCT. For most images, after transformation the majority of signal energy is carried by just a few of the low order DCT coefficients. These coefficients can be more finely quantized than the higher order coefficients. Many higher order coefficients may be quantized to 0 (this allows for very efficient run-level coding).

$$F(u, v) = \frac{2}{N} C(u)C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \left[\frac{(2x+1)u\pi}{2N} \right] \cos \left[\frac{(2y+1)v\pi}{2N} \right]$$

$$f(i, j) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v) F(u, v) \cos \left[\frac{(2x+1)u\pi}{2N} \right] \cos \left[\frac{(2y+1)v\pi}{2N} \right]$$

Which can be written in matrix form, where the rows of [T] are the DCT basis vectors, as:

$$[f]_{N \times N} = [T^T]_{N \times N} [F]_{N \times N} [T]_{N \times N}$$

$$[F]_{N \times N} = [T]_{N \times N} [f]_{N \times N} [T^T]_{N \times N}$$

In the formulas, F(u,v) is the two-dimensional NxN DCT. u, v, x, y = 0,1,2,...N-1 where x, y are spatial coordinates in the sample domain and u, v are frequency coordinates in the transform domain.

C(u), C(v) = 1/(square root (2)) for u, v = 0.

C(u), C(v) = 1 otherwise.

MPEG specifies the spatial samples to be represented in 9 bits and the coefficients to be represented in 12 bits. The dynamic range of the coefficients is specified as [-2048:+2047].

- Jubli Kashyap is currently working as Assistant Professor (EC Deptt.) in NIEC, Guru Gobind Singh Indraprastha University, Delhi, India, Mobile No.- 09213356485. E-mail: jubli_k@rediffmail.com
- Virendra Kumar Yadav is currently working as Technical Services professional, IBM, India Mobile No.- 09212428516, E-mail: virendra7n@rediffmail.com.

$$[f]_{N \times N} = [T^T]_{N \times N} [F]_{N \times N} [T]_{N \times N}$$

$$[F]_{N \times N} = [T]_{N \times N} [f]_{N \times N} [T^T]_{N \times N}$$

2.3 Color Transformation

The human visual systems are most sensitive to changes in luminance and less to changes in chrominance.

RGB must be converted to the other color systems.

YUV System

$$Y = 0.299R + 0.587G + 0.114B ;$$

$$U = B - Y ;$$

$$V = R - Y ;$$

Zero Shift

After transformation Y, U and V are in the range of [0,255] Zero shift changes this range to [-128,127]

DCT The strength of transform coding in achieving data compression is that the image energy of the most natural scenes is mainly concentrated in the low frequency region and hence into a few transform coefficients.

2.4 Quantization

Quantization allows us to reduce the accuracy with which the DCT coefficients are represented when converting the DCT to an integer representation.

It tends to make many coefficients zero, specially those for high spatial frequencies.

Two standard tables are available for quantization.

2.5 Zigzag Scan

The zigzag pattern used in the JPEG algorithms orders the basis functions from low to high spatial frequencies .

It facilitate entropy coding by encountering the most likely non-zero coefficient first.

2.6 Picture Format

The picture is divided into number of Macro Blocks (MB) In 4:4:4 sampling each MB contains four Y-blocks, four R-Y blocks and four B-Y blocks.

In 4:2:2 sampling each MB contains four Y-blocks, two R-Y blocks and two B-Y blocks.

* Color difference signal is sub-sampled in Horizontal direction

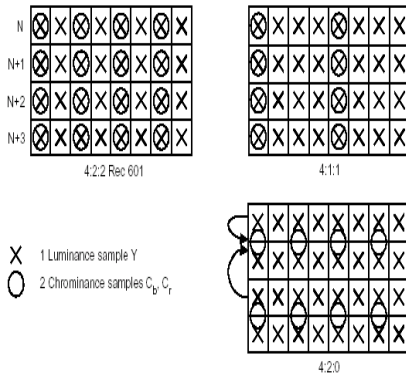
* The bit rate = 720X576X25X8 + 360X576X25X (8+8) = 166Mb/s

In 4:2:0 sampling each MB contains four Y-blocks, one R-Y block and one B-Y block

* Color difference signal is sub-samples in both horizontal and vertical direction

* Bit rate= $720 \times 576 \times 25 \times 8 + 360 \times 288 \times 25 \times (8+8) = 124 \text{ Mb/s}$.

Motion vectors are associated with MB's(Microblocks)



[2] Joint Video Team (JVT) of ITU-T VCEG and ISO/IEC MPEG, Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification, ITU-T Rec. H.264 and ISO/IEC 14496-10 AVC, May 2003.

[3] ITU-T Rec. H.264 / ISO/IEC 11496-10, "Advanced Video Coding", Final Committee Draft, Document JVTE022, September 2002

[4] INTRODUCTION TO DATA COMPRESSION 3RD EDITION BY KHALID SAYOOD.

[5] H.264 AND MPEG-4 VIDEO COMPRESSION BY IAIN E G Richardson. Published by John Wiley & Sons, September 2003.

[6] Multimedia Communications by Fred Halsal, Pearson Educaton,2004.

3 IMPLEMENTATION RESULTS

The proposed architecture is implemented in Verilog HDL. The implementation is verified with RTL simulations using Cadence ncLaunch. The Verilog RTL is then synthesized to a Cadence BuildGate Synthesis tool. The resulting netlist is placed and routed to the Cadence SoC Encounter. The first set of architectures for the DCT and quantization were optimized for power, which resulted in transform and quantizer blocks that use 10.5623 mW Power.

4 REFERENCES

[1] I. Richardson, H.264 and MPEG-4 Video Compression, Wiley, 2003