Hardware analysis of transform and quantization

(JCTVC-G132)

Madhukar Budagavi, Vivienne Sze, Mangesh Sadafale

Texas Instruments

Joint Collaborative Team on Video Coding (JCT-VC)
of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11
7th Meeting: Geneva, CH, 21-30 November, 2011

Overview

- Transforms based on matrix multiplication is useful feature to have for HEVC core transform
  - DSPs, ARM, SIMD based devices equipped with 8-way MACs are already available
  - Hardwired SIMD engines are also popular in hardware codecs that supports multiple standards

Overview - Hardware

- Transform and quantization complexity needs to be evaluated by considering a **complete solution** that takes into consideration forward and inverse transform and quantization and de-quantization
- Forward transform, Inverse transform, Quantization, De-quantization typically implemented in a single hardware block **to reduce number of pipeline stages and to reuse commonality**
- Hardware block also needs to support **transforms & quants from multiple video coding standards**
- Two architectures presented for HM4 to demonstrate sharing between forward and inverse transform and quantization
- Architecture 1: Unifies forward+inverse
  - 5-10% more area than forward only, 17-21% more area than inverse only
- Architecture 2: Hardwired SIMD based, good for multi-standard support
  - Inverse+Forward+Quant+Dequant

Need for complete solution for transform and quantization

- Mobile phones, tablet computers etc. need to support both video capture and playback on same device
- Video conferencing needs encoder/decoder running simultaneously
- For video encoder applications such as video surveillance, encoder would require support both forward and inverse transform
- Even set-top-boxes where decoder complexity is a primary concern, we would still need to build in a light-weight encoder to support PVR/DVR functionality
- Therefore, it is important that the transform and quantization design we select for HEVC is friendly to a complete solution that includes:
  - Forward transform, Inverse transform
  - Quantization, De-quantization

Two architectures to illustrate hardware sharing in HM4 design

- First architecture makes use of a useful symmetry property between forward and inverse HM4 transforms to reduce overall area of hardware block that needs to implement both forward and inverse transform
- Second architecture is a special-purpose SIMD architecture that can implement both transform and quantization

HM4 Inverse 4-point transform

Inverse transform

$$\begin{bmatrix} Y0 \\ Y1 \\ Y2 \\ Y3 \end{bmatrix} = \begin{bmatrix} C32 & C16 & C32 & C48 \\ C32 & C48 & -C32 & -C16 \\ C32 & -C48 & -C32 & C16 \\ C32 & -C16 & C32 & -C48 \end{bmatrix} \begin{bmatrix} X0 \\ X1 \\ X2 \\ X3 \end{bmatrix} \Rightarrow$$

Even $\begin{bmatrix} Z0 \\ Z1 \end{bmatrix} = \begin{bmatrix} C32 & C32 \\ C32 & -C32 \end{bmatrix} \begin{bmatrix} X0 \\ X2 \end{bmatrix}$

Odd $\begin{bmatrix} Z2 \\ Z3 \end{bmatrix} = \begin{bmatrix} -C48 & C16 \\ -C16 & -C48 \end{bmatrix} \begin{bmatrix} X1 \\ X3 \end{bmatrix}$

$$\begin{bmatrix} Y0 \\ Y1 \\ Y2 \\ Y3 \end{bmatrix} = \begin{bmatrix} Z0 - Z3 \\ Z1 - Z2 \\ Z1 + Z2 \\ Z0 + Z3 \end{bmatrix}$$

HM4 Forward 4-point transform

Forward

$$\begin{bmatrix} P0 \\ P1 \\ P2 \\ P3 \end{bmatrix} = \begin{bmatrix} C32 & C32 & C32 & C32 \\ C16 & C48 & -C48 & -C16 \\ C32 & -C32 & -C32 & C32 \\ C48 & -C16 & C16 & -C48 \end{bmatrix} \begin{bmatrix} M0 \\ M1 \\ M2 \\ M3 \end{bmatrix} \Rightarrow$$

$$\begin{bmatrix} K0 \\ K1 \\ K2 \\ K3 \end{bmatrix} = \begin{bmatrix} M0+M3 \\ M1+M2 \\ -M1+M2 \\ -M0+M3 \end{bmatrix}$$

$$\begin{bmatrix} P0 \\ P2 \end{bmatrix} = \begin{bmatrix} C32 & C32 \\ C32 & -C32 \end{bmatrix} \begin{bmatrix} K0 \\ K1 \end{bmatrix}$$ Even

$$\begin{bmatrix} P1 \\ P3 \end{bmatrix} = \begin{bmatrix} -C48 & -C16 \\ C16 & -C48 \end{bmatrix} \begin{bmatrix} K2 \\ K3 \end{bmatrix}$$ Odd

Commonality between HM4 forward and inverse transform for 4-points

$$\begin{bmatrix} K0 \\ K1 \\ K2 \\ K3 \end{bmatrix} = \begin{bmatrix} M0+M3 \\ M1+M2 \\ -M1+M2 \\ -M0+M3 \end{bmatrix}$$

Even part matrix is identical

Even $\begin{bmatrix} Z0 \\ Z1 \end{bmatrix} = \begin{bmatrix} \text{C32} & \text{C32} \\ \text{C32} & \text{-C32} \end{bmatrix} \begin{bmatrix} X0 \\ X2 \end{bmatrix}$ ⟷ $\begin{bmatrix} P0 \\ P2 \end{bmatrix} = \begin{bmatrix} \text{C32} & \text{C32} \\ \text{C32} & \text{-C32} \end{bmatrix} \begin{bmatrix} K0 \\ K1 \end{bmatrix}$ Even

Odd $\begin{bmatrix} Z2 \\ Z3 \end{bmatrix} = \begin{bmatrix} \text{-C48} & \text{C16} \\ \text{-C16} & \text{-C48} \end{bmatrix} \begin{bmatrix} X1 \\ X3 \end{bmatrix}$ ⟷ $\begin{bmatrix} P3 \\ P1 \end{bmatrix} = \begin{bmatrix} \text{-C48} & \text{C16} \\ \text{-C16} & \text{C48} \end{bmatrix} \begin{bmatrix} K3 \\ K2 \end{bmatrix}$ Odd

$$\begin{bmatrix} Y0 \\ Y1 \\ Y2 \\ Y3 \end{bmatrix} = \begin{bmatrix} Z0-Z3 \\ Z1-Z2 \\ Z1+Z2 \\ Z0+Z3 \end{bmatrix}$$

Odd part matrix is identical when forward inputs and outputs are flipped

HM4 Inverse 8-point transform

Inverse transform

$$\begin{bmatrix} Y0 \\ Y1 \\ Y2 \\ Y3 \\ Y4 \\ Y5 \\ Y6 \\ Y7 \end{bmatrix} = \begin{bmatrix} C32 & C8 & C16 & C24 & C32 & C40 & C48 & C56 \\ C32 & C24 & C48 & -C56 & -C32 & -C8 & -C16 & -C40 \\ C32 & C40 & -C48 & -C8 & -C32 & C56 & C16 & C24 \\ C32 & C56 & -C16 & -C40 & C32 & C24 & -C48 & -C8 \\ C32 & -C56 & -C16 & C40 & C32 & -C24 & -C48 & C8 \\ C32 & -C40 & -C48 & C8 & -C32 & -C56 & C16 & -C24 \\ C32 & -C24 & C48 & C56 & -C32 & C8 & -C16 & C40 \\ C32 & -C8 & C16 & -C24 & C32 & -C40 & C48 & -C56 \end{bmatrix} \begin{bmatrix} X0 \\ X1 \\ X2 \\ X3 \\ X4 \\ X5 \\ X6 \\ X7 \end{bmatrix}$$

Even
$$\begin{bmatrix} Z0 \\ Z1 \\ Z2 \\ Z3 \end{bmatrix} = \begin{bmatrix} C32 & C16 & C32 & C48 \\ C32 & C48 & -C32 & -C16 \\ C32 & -C48 & -C32 & C16 \\ C32 & -C16 & C32 & -C48 \end{bmatrix} \begin{bmatrix} X0 \\ X2 \\ X4 \\ X6 \end{bmatrix}$$

Odd
$$\begin{bmatrix} Z4 \\ Z5 \\ Z6 \\ Z7 \end{bmatrix} = \begin{bmatrix} -C56 & C40 & -C24 & C8 \\ -C40 & C8 & -C56 & -C24 \\ -C24 & C56 & C8 & C40 \\ -C8 & -C24 & -C40 & -C56 \end{bmatrix} \begin{bmatrix} X1 \\ X3 \\ X5 \\ X7 \end{bmatrix}$$

$$\begin{bmatrix} Y0 \\ Y1 \\ Y2 \\ Y3 \\ Y4 \\ Y5 \\ Y6 \\ Y7 \end{bmatrix} = \begin{bmatrix} Z0-Z7 \\ Z1-Z6 \\ Z2-Z5 \\ Z3-Z4 \\ Z3+Z4 \\ Z2+Z5 \\ Z1+Z6 \\ Z0+Z7 \end{bmatrix}$$

HM4 Forward 8-point transform

Forward

$$\begin{bmatrix} P0 \\ P1 \\ P2 \\ P3 \\ P4 \\ P5 \\ P6 \\ P7 \end{bmatrix} = \begin{bmatrix} C32 & C32 & C32 & C32 & C32 & C32 & C32 & C32 \\ C8 & C24 & C40 & C56 & -C56 & -C40 & -C24 & -C8 \\ C16 & C48 & -C48 & -C16 & -C16 & -C48 & C48 & C16 \\ C24 & -C56 & -C8 & -C40 & C40 & C8 & C56 & -C24 \\ C32 & -C32 & -C32 & C32 & C32 & -C32 & -C32 & C32 \\ C40 & -C8 & C56 & C24 & -C24 & -C56 & C8 & -C40 \\ C48 & -C16 & C16 & -C48 & -C48 & C16 & -C16 & C48 \\ C56 & -C40 & C24 & -C8 & C8 & -C24 & C40 & -C56 \end{bmatrix} \begin{bmatrix} M0 \\ M1 \\ M2 \\ M3 \\ M4 \\ M5 \\ M6 \\ M7 \end{bmatrix}$$

$$\begin{bmatrix} K0 \\ K1 \\ K2 \\ K3 \\ K4 \\ K5 \\ K6 \\ K7 \end{bmatrix} = \begin{bmatrix} M0+M7 \\ M1+M6 \\ M2+M5 \\ M3+M4 \\ -M3+M4 \\ -M2+M5 \\ -M1+M6 \\ -M0+N7 \end{bmatrix}$$

$$\begin{bmatrix} P0 \\ P2 \\ P4 \\ P6 \end{bmatrix} = \begin{bmatrix} C32 & C32 & C32 & C32 \\ C16 & C48 & -C48 & -C16 \\ C32 & -C32 & -C32 & C32 \\ C48 & -C16 & C16 & -C48 \end{bmatrix} \begin{bmatrix} K0 \\ K1 \\ K2 \\ K3 \end{bmatrix}$$

Even

$$\begin{bmatrix} P1 \\ P3 \\ P5 \\ P7 \end{bmatrix} = \begin{bmatrix} -C56 & -C40 & -C24 & -C8 \\ C40 & C8 & C56 & -C24 \\ -C24 & -C56 & C8 & -C40 \\ C8 & -C24 & C40 & -C56 \end{bmatrix} \begin{bmatrix} K4 \\ K5 \\ K6 \\ K7 \end{bmatrix}$$

Odd

Commonality between HM4 forward and inverse transform for 8-points

Even

$$\begin{bmatrix} Z0 \\ Z1 \\ Z2 \\ Z3 \end{bmatrix} = \begin{bmatrix} C32 & C16 & C32 & C48 \\ C32 & C48 & -C32 & -C16 \\ C32 & -C48 & -C32 & C16 \\ C32 & -C16 & C32 & -C48 \end{bmatrix} \begin{bmatrix} X0 \\ X2 \\ X4 \\ X6 \end{bmatrix}$$

Odd

$$\begin{bmatrix} Z4 \\ Z5 \\ Z6 \\ Z7 \end{bmatrix} = \begin{bmatrix} -C56 & C40 & -C24 & C8 \\ -C40 & C8 & -C56 & -C24 \\ -C24 & C56 & C8 & C40 \\ -C8 & -C24 & -C40 & -C56 \end{bmatrix} \begin{bmatrix} X1 \\ X3 \\ X5 \\ X7 \end{bmatrix}$$

$$\begin{bmatrix} Y0 \\ Y1 \\ Y2 \\ Y3 \\ Y4 \\ Y5 \\ Y6 \\ Y7 \end{bmatrix} = \begin{bmatrix} Z0-Z7 \\ Z1-Z6 \\ Z2-Z5 \\ Z3-Z4 \\ Z3+Z4 \\ Z2+Z5 \\ Z1+Z6 \\ Z0+Z7 \end{bmatrix}$$

Even part matrix same as 4-pt transform

Odd part matrix differs only in sign of circled entries

$$\begin{bmatrix} K0 \\ K1 \\ K2 \\ K3 \\ K4 \\ K5 \\ K6 \\ K7 \end{bmatrix} = \begin{bmatrix} M0+M7 \\ M1+M6 \\ M2+M5 \\ M3+M4 \\ -M3+M4 \\ -M2+M5 \\ -M1+M6 \\ -M0+N7 \end{bmatrix}$$

Even

$$\begin{bmatrix} P0 \\ P2 \\ P4 \\ P6 \end{bmatrix} = \begin{bmatrix} C32 & C32 & C32 & C32 \\ C16 & C48 & -C48 & -C16 \\ C32 & -C32 & -C32 & C32 \\ C48 & -C16 & C16 & -C48 \end{bmatrix} \begin{bmatrix} K0 \\ K1 \\ K2 \\ K3 \end{bmatrix}$$

Odd

$$\begin{bmatrix} P1 \\ P3 \\ P5 \\ P7 \end{bmatrix} = \begin{bmatrix} -C56 & -C40 & -C24 & -C8 \\ C40 & C8 & C56 & -C24 \\ -C24 & -C56 & C8 & -C40 \\ C8 & -C24 & C40 & -C56 \end{bmatrix} \begin{bmatrix} K4 \\ K5 \\ K6 \\ K7 \end{bmatrix}$$

Area savings of HM4 joint fwd+inv design

Frequency (MHz)	32-pt HM4 inv only (Kgates)	32-pt HM4 fwd only (Kgates)	32-pt HM4 inv+fwd only (Kgates)	Percent additional area of inv+fwd over forward only	Percent additional area of inv+fwd over inverse only
150	116	126	139	10%	**20%**
200	122	139	146	5%	**20%**
250	130	148	156	5%	**20%**
300	140	157	169	8%	**21%**
350	150	170	181	6%	**21%**
400	164	182	192	5%	**17%**
450	175	195	206	6%	**18%**
500	189	216	226	5%	**20%**

Architecture 2: Special purpose SIMD

- Further hardware sharing between transform and quantization is possible using a hardwired SIMD architecture
  - JCTVC-F446 presented results for unified architecture based on 32-way SIMD for HM4
- Hardwired SIMD also useful in supporting multiple codecs using common hardware
- Important for design to be friendly for SIMD implementation e.g. implementable using matrix multiplication
- Comparison between partial butterfly HM4 SIMD and partial butterfly F251 SIMD
  - HM4 – 8-bit coefficients
  - F251 – 14-bit coefficients
  - F251 SIMD expected to have larger area when compared to HM4 SIMD

Conclusions

- Transforms based on matrix multiplication is useful feature to have for HEVC core transform
  - DSPs, ARM, SIMD based devices equipped with 8-way or 16-way multiplier are already available
  - Hardwired SIMD engine is also popular in hardware codecs that supports multiple standards
- Transform and quantization complexity needs to be evaluated by considering a **complete solution** that takes into consideration forward and inverse transform and quantization and de-quantization
- Two architectures presented for HM4 to demonstrate sharing between forward and inverse transform and quantization
  - Architecture 1: Unifies forward+inverse
    - 5-10% more area than forward only, 17-21% more area than inverse only
  - Architecture 2: Hardwired SIMD based, good for multi-standard support
    - Inverse+Forward+Quant+Dequant