A 249Mpixel/s HEVC video-decoder chip for Quad Full HD applications


As Published: http://dx.doi.org/10.1109/ISSCC.2013.6487682

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

Version: Author’s final manuscript

Citable link: http://hdl.handle.net/1721.1/93239

Terms of Use: Creative Commons Attribution-Noncommercial-Share Alike

Detailed Terms: http://creativecommons.org/licenses/by-nc-sa/4.0/
The latest video coding standard High Efficiency Video Coding (HEVC) [1] provides 50% improvement in coding efficiency compared to H.264/AVC, to meet the rising demand for video streaming, better video quality and higher resolutions. The coding gain is achieved using more complex tools such as larger and variable-size coding units (CU) in a hierarchical structure, larger transforms and longer interpolation filters. This paper presents the first integrated circuit which supports Quad Full HD (QFHD, 3840x2160) video decoding for HEVC draft standard. It addresses new design challenges for HEVC (‘H.265’) with three primary contributions: 1) a system pipelining scheme which adapts to the variable-size largest coding unit (LCU) and provides a two-stage sub-pipeline for memory optimization; 2) unified processing engines to address hierarchical coding structure and many prediction and transform block sizes in area-efficient ways; 3) a motion compensation (MC) cache which reduces DRAM bandwidth for the large LCU and meets the high throughput requirement due to long filters.

The HEVC standard uses 8-bit integer transform for 4 to 32-pt Inverse DCT (IDCT) and 4-pt Inverse DST (IDST). The largest TU is 32x32 with seven smaller TUs – three square (4x4, 8x8, 16x16) and four non-square (4x16, 8x32, 16x4, 32x8). Compared to H.264/AVC, this is an 8x increase in 1-D transform logic and a 16x increase in transpose memory for the largest TU. To reduce logic area, only four row or column pixels in a TU are transformed every cycle by the variable-size partial 1-D transform block shown in Figure 9.5.3. We extend the typical recursive decomposition of a 2Npt IDCT into N-pt IDCT and NxN matrix multiplication to our partial architecture with even-odd index sorting. The NxN matrix has only N unique elements differing only in sign. The 4x4 matrix, for example, has only four elements – 89, 75, 50, 18. We exploit this property to use Multiple Constant Multiplication (MCM) [6] instead of N full multipliers. This reduces the variable-size partial transform area from 96K to 71K gates. Two MCM planes are needed for partial 4x4, 8x8 and 16x16 matrix multiplications to meet the required throughput. The transpose memory would need 125K gate register array. Instead, it is implemented using four single-port SRAMs for 4 pixel/cycle read or write that is fully pipelined with the transform computation. Some pixels are stored in a register-based row cache to avoid a stall when switching from column to row transform.

### Acknowledgements:
This work was funded by Texas Instruments. The authors thank the TSMC University Shuttle Program for chip fabrication.

### References:
Figure 9.5.1: System block diagram and pipelining scheme in variable-size pipeline block (VPB).

Figure 9.5.2: Unified prediction engine and memory-efficient two-stage sub-pipeline in PPB and sub-PPB.

Figure 9.5.3: Unified 2-D inverse transform engine with variable-size partial transform for all square/non-square TUs with 4 to 32-pt IDCT and 4-pt IDST.

Figure 9.5.4: Four-parallel high throughput MC cache and twisted 2-D DRAM mapping.

Figure 9.5.5: Chip specifications and measurement results.

Figure 9.5.6: Comparison with state-of-the-art video decoders.