

# ALGORITHM AND ARCHITECTURE DESIGN FOR INTRA PREDICTION IN H.264/AVC HIGH PROFILE

*Tzu-Der Chuang, Yi-Hau Chen, Chen-Han Tsai, Yu-Jen Chen, and Liang-Gee Chen*

DSP/IC Design Lab.,  
 Graduate Institute of Electronics Engineering,  
 National Taiwan University, Taipei, Taiwan  
 Email: {peterchuang, ttchen, phenom, yjchen, lgchen}@video.ee.ntu.edu.tw

## ABSTRACT

In this paper, we propose a novel two-stage intra prediction algorithm and hardware architecture that can support H.264/AVC high profile for 1080p HD size. The proposed DCT-based open-loop intra prediction algorithm can parallel predict each sub block with nearly no quality drop and skip unnecessary reconstruction loop to meet real-time encoding constrain. With proposed reconfigurable 8-pixel parallelism processing elements, the proposed architecture can process intra prediction and reconstruction with almost 100% hardware utilization. The proposed architecture was implemented by UMC 90 nm technology with 100k gate counts at 223MHz. It is the first hardware architecture that can real-time encode 1080p HD intra frame sequence with H.264/AVC high profile.

## 1. INTRODUCTION

H.264/AVC is the newest advanced video coding standards developed by the ITU-T - ISO/IEC Joint Video Team (JVT). This new coding standard includes variable block size motion compensation, simplified integer transform and, multiple reference frame and context adaptive entropy coding. Moreover, a new intra prediction which uses neighboring pixels to predict current coding block is proposed in this standard. The intra prediction can significantly improves the coding performance in intra frame, even when motion estimation fails to find a good match, intra prediction will have a good chance to further reduce the residues.

In order to improve coding performance at high end application, a new amendment called the Fidelity Range Extensions (FRExt, Amendment I) was added to the H.264 standard, in July 2004. The FRExt project brings up a suite of some new profiles collectively called High profiles [1]. The high profiles support all features of the prior Main profile, and additionally support two main coding tools,  $8 \times 8$  integer transform and  $8 \times 8$  intra prediction. For inter prediction, macroblocks (MBs) are allowed to be encoded by  $8 \times 8$  integer transform when the block size of current sub MB mode

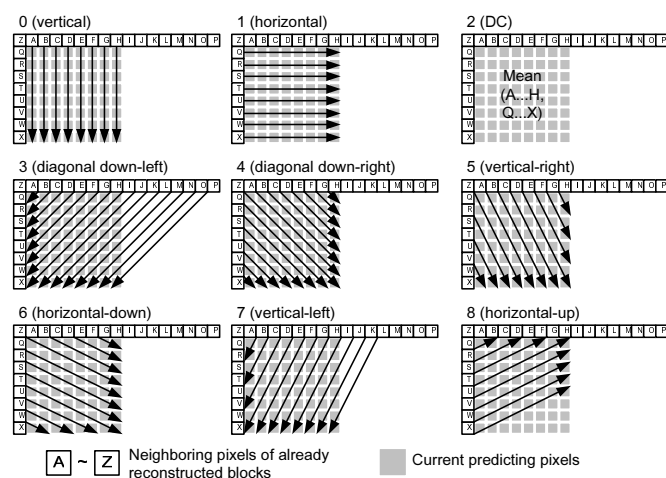
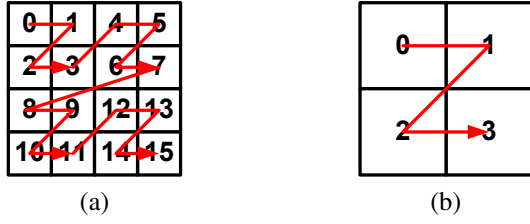


Fig. 1. Illustration of nine  $8 \times 8$  luma prediction mode

is larger than  $8 \times 8$ . For intra prediction, high profile introduces another  $8 \times 8$  spatial luma prediction by extending the concepts of Intra\_4x4 prediction as shown in Fig. 1. Similar to Intra\_4x4, Intra\_8x8 has eight different direction prediction modes and one DC prediction mode. The luma values of each sample in a given  $8 \times 8$  block are predicted from neighboring reconstructed reference pixels base on prediction modes. Besides, a distinguishing element of Intra\_8x8 prediction is a two-order binormal pre-filtering process on the boundary reference pixels before prediction step. It should be noticed that the transform of Intra\_8x8 prediction must use the new  $8 \times 8$  integer DCT transform addressed by high profile. This new intra coding tool can improve I-frame coding efficiency significantly [2].

Several researches have been published on H.264 intra hardware architecture [3, 4, 5, 6]. These previous architectures focus on main profile intra prediction which only contains Intra\_4x4 and Intra\_16x16 mode, and none of them can support Intra\_8x8 mode. In order to support Intra\_8x8 mode into hardware architecture, several problems should be solved,



**Fig. 2.** Zig-zag scan order for (a) 4x4 block and (b) 8x8 block

such as critical data dependency, insufficient operating cycles, larger memory requirement, and unbalance computation load. To solve these problems, we firstly present an efficient two-stage intra prediction hardware architecture that can support H.264 high profile at 1080p HD size 30 fps spec.

## 2. DESIGN CHALLENGES OF H.264/AVC HIGH PROFILE INTRA PREDICTION

In previous H.264/AVC designs [3, 4, 5, 6], intra prediction for baseline and main profile are well-developed for D1 and HD 720p specification. However, there are two main design challenges which lower the efficiency of above designs after Intra\_8x8 of high profile being taken into consideration.

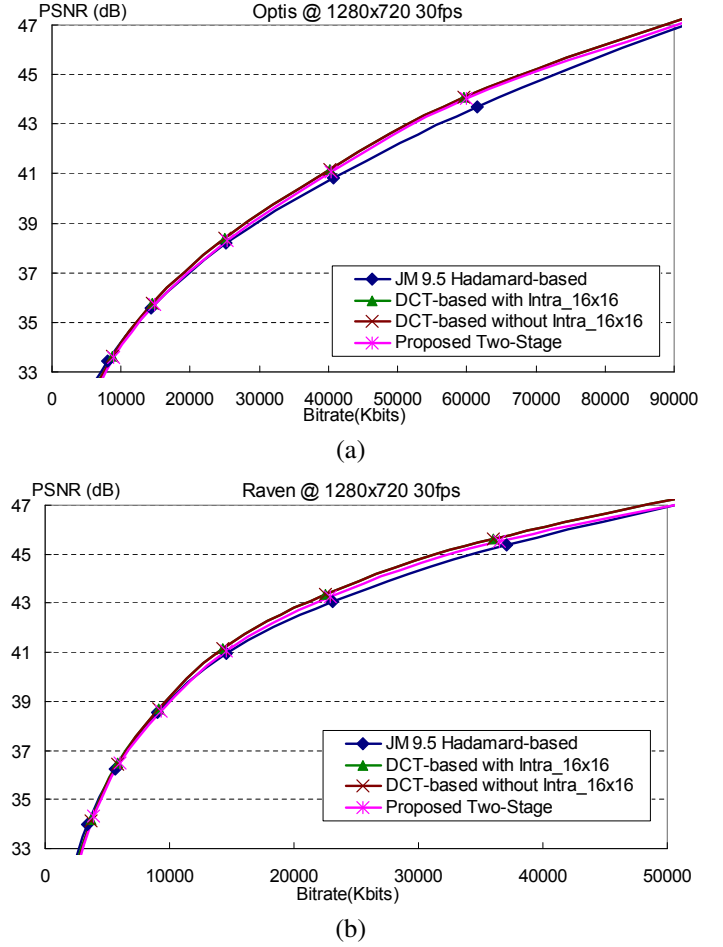
The first issues comes from the data dependency of intra prediction between each sub-block. Based on H.264/AVC standard definition for Intra\_4x4 and Intra\_8x8 modes, each sub-block should be processed by the zig-zag scan order as shown in Fig. 2, and the 13 or 25 reconstructed pixels are required for prediction as shown in Fig. 1. Since the reconstructed pixels can be only available until the neighboring blocks are predicted and reconstructed, each sub-block should be processed sequentially. In Suh's [5] and Ku's [4] design take about 1000 cycles to process one MB's intra prediction to meet HD 720p specification without Intra\_8x8. However, the above designs will require much higher operating frequency when Intra\_8x8 of high profile is taken into consideration.

The other issue is the throughput and hardware utilization. For Intra\_4x4 and Intra\_16x16 mode, four pixel parallelism is usually adopted [3, 4, 6]. But for Intra\_8x8 mode, eight pixel parallelism should be applied due to 8x8 transform. The mismatch throughput of different intra modes should be unified for intra predictor generator, transform, and reconstruction to improve the hardware utilization and processing capability.

## 3. PROPOSED HARDWARE ORIENTED INTRA PREDICTION ALGORITHM

### 3.1. DCT-based SATD Cost Function for Mode Decision

In H.264/AVC reference software, Hadamard transform is involved to calculate SATD for mode decision. It is because Hadamard transform has advantages of less computation complexity and no scaling effect. However, in H.264/AVC encod-

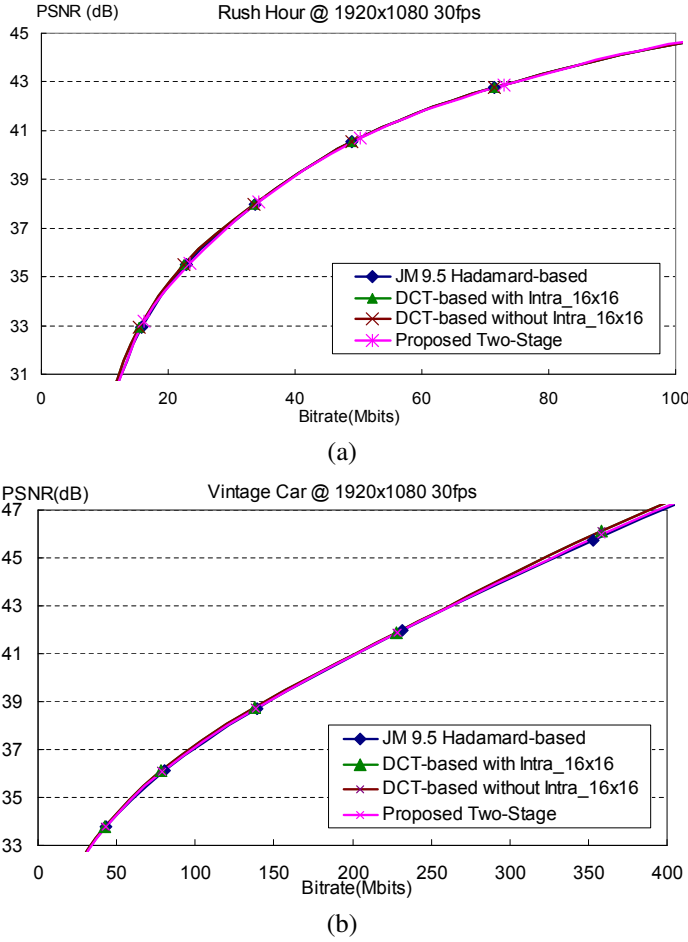


**Fig. 3.** Performance comparison of different algorithm on 720p sequence, (a) Optis and (b) Raven by 300 I-frame, CABAC encoding

ing process, the DCT transform is adopted for reconstruction loop. In order to estimate bitrate more accurately, we use 4x4 and 8x8 integer DCT as cost function for Intra\_4x4, Intra\_16x16 and Intra\_8x8 prediction mode decision to approximate the effect of transform and quantization in H.264 encoding process. As mentioned in [4], the scaling effect after DCT should be taken into account. In Intra\_4x4 mode, we choose the simplified scaling matrix in (1) for simplifying hardware architecture, where  $D$  is the 4x4 integer DCT matrix. For Intra\_8x8 mode, since all scaling factors are very similar, we do not apply any scaling matrix on it.

$$SATD(X) = DXD^T \otimes \begin{bmatrix} 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} / 2 \quad (1)$$

As shown in Fig. 3 and Fig. 4 result of "JM 9.5 Hadamard-based" and "DCT-based with Intra\_16x16", our proposed DCT-based SATD cost function for H.264/AVC high profile can improve up to 0.3dB compared to JM 9.5. The algorithm of

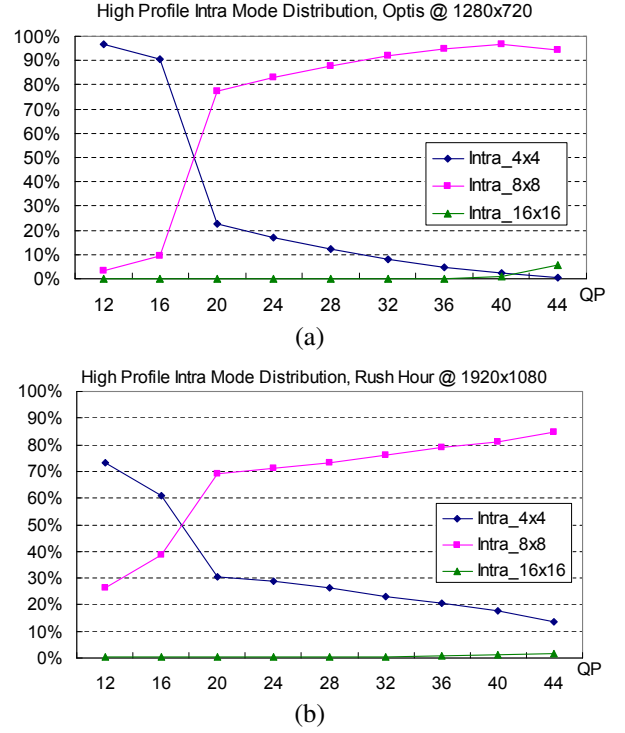


**Fig. 4.** Performance comparison of different algorithm on 1080p sequence (a) Rush Hour and (b) Vintage Car by 300 I-frame, CABAC encoding

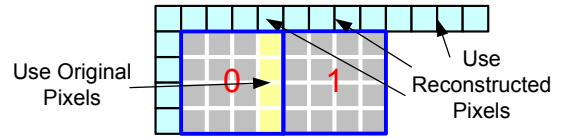
result “DCT-based without Intra\_16x16” and “Proposed Two-Stage” will be explained in Sec. 3.2 and Sec.3.3.

### 3.2. High Profile Intra Prediction Mode Distribution

In high profile, Intra\_8x8 mode provides a trade-off option between Intra\_4x4 and Intra\_16x16. The Intra\_8x8 has larger prediction size and fewer sub block headers comparing to Intra\_4x4 mode, and has more prediction modes than Intra\_16x16. Therefore we analyze the intra prediction mode distribution of high profile based on our proposed DCT-based cost estimation in Sec. 3.1. In Fig. 5, Intra\_16x16 is already replaced by Intra\_8x8 at low and medium bitrate and rarely selected at any QP. Based on Fig. 3 and Fig. 4, removing Intra\_16x16 almost has no quality drop. Hence, we remove Intra\_16x16 mode from our hardware design to simplify the hardware architecture and schedule with nearly no quality drop.



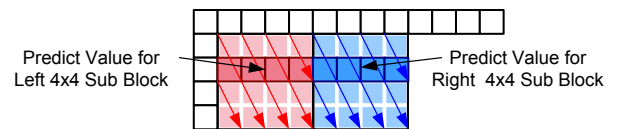
**Fig. 5.** Intra prediction mode distribution of high profile over 300 I-frames on HD sequence (a) Optis and (b) Rush Hour



**Fig. 6.** Illustration of open-loop prediction boundary pixel.

### 3.3. Hardware Oriented Open-loop Intra Prediction

In order to improve the processing parallelism limited by data dependency addressed in Sec. 2, we propose an open-loop intra prediction scheme to use original pixels instead of reconstructed pixels as boundary pixels for intra predictors. The open-loop prediction concept is based on that original pixels are close to reconstructed pixels in our target high definition application which PSNR is often greater than 35dB. For MB boundary pixels, the reconstructed pixels are still used because they are already available. Take Intra\_4x4 as an example, when block 1 in Fig.6 is processing, it uses the four



**Fig. 7.** Reconfigurable luma predictor for two 4x4 blocks.

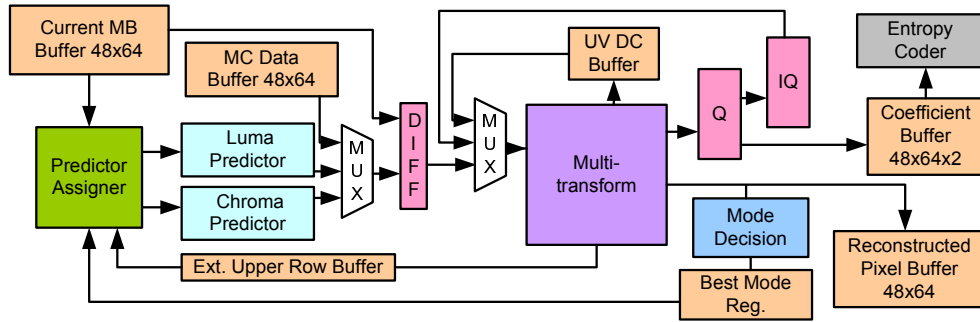


Fig. 8. Proposed high profile intra hardware architecture.

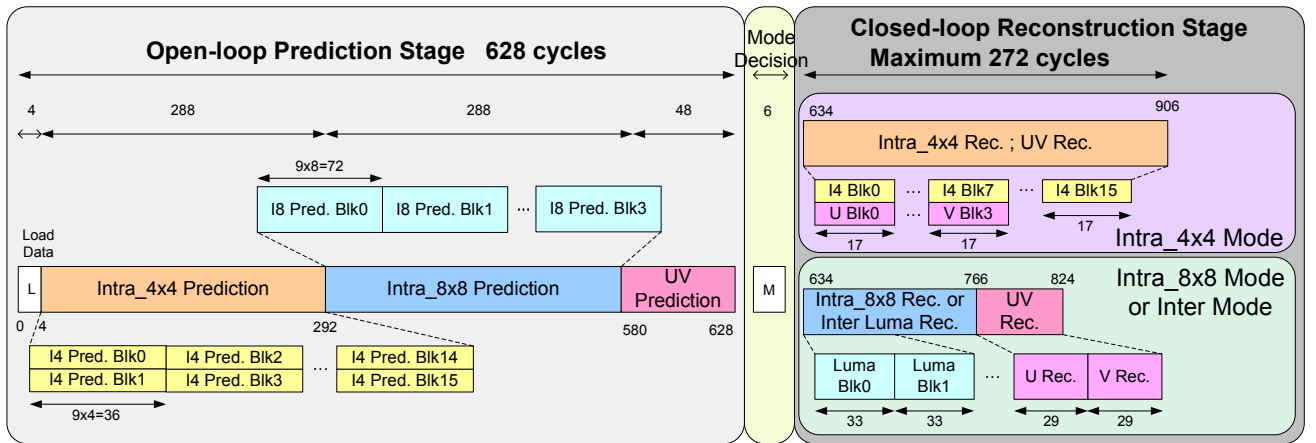


Fig. 9. Proposed open-loop prediction and closed-loop reconstruction two-stage schedule.

original pixels (yellow pixels) as its left boundary pixels and the nine reconstructed pixels from upper row MB as upper boundary pixels. As shown in Fig. 3 and Fig. 4, our proposed open-loop scheme has very slight quality degradation comparing to closed-loop DCT-based intra prediction and is still better than JM 9.5. By proposed open-loop scheme, the intra prediction of each sub block can be predicted in parallel without waiting neighboring blocks' reconstruction loop.

#### 4. PROPOSED HARDWARE ARCHITECTURE AND SCHEDULE

##### 4.1. Reconfigurable 8-pixel parallelism PE design

In order to be consistent with the throughput of 8x8 DCT in Intra\_8x8 prediction, the parallelism of our architecture is set to be 8 pixels parallel. Since the Intra\_8x8 prediction mode is similar to Intra\_4x4 prediction, we propose a reconfigurable intra luma predictor generator that can generate eight predictors for Intra\_8x8 mode, or eight predictors for two 4x4 sub blocks for Intra\_4x4 mode as shown in Fig. 7. Besides, the modified multi-transform in our previous work

is adopted[7]. This multi-transform can be configured as two 4x4 DCT/IDCT or one 8x8 DCT/IDCT transform for cost estimation and reconstruction. The proposed hardware architecture can unified throughput and improve processing capability with excellent area efficiency by using these reconfigurable 8-pixel parallel PEs as shown in Fig. 8.

##### 4.2. Proposed two stage schedule

Unlike previous prediction-reconstruction interleaved scheme [3, 4, 5], the schedule of proposed architecture can be divided into two stages, open-loop prediction and closed-loop reconstruction as shown in Fig. 9. In prediction stage, our architecture can process two 4x4 sub blocks in parallel in Intra\_4x4 mode and can process next two sub blocks immediately without reconstruction because of open-loop prediction. In this stage, only the best mode for each sub block and total MB mode cost are stored.

In reconstruction stage, only one mode is selected for reconstruction. If Intra\_4x4 mode is selected, luma 4x4 block and chroma 4x4 block will be reconstructed in parallel for higher hardware utilization as shown in Fig. 9. It is because in

**Table 1.** Gate count distribution of proposed intra prediction architecture

Module	gate count
Predictor assigner	9.8k
Luma predictor	17.3k
Chroma Predictor	1.3k
Multi-transform	27.5k
Mode decision	6.9k
Quantization	19.3k
Inverse Quant.	9.2k
FSM and buffers	9.1k
Total	100.4k

H.264 decoding process, each 4x4 luma sub block should be reconstructed in the zig-zag scan order in Fig. 2(a). Once Intra\_8x8 mode or inter mode is chosen, it will process only one 8x8 luma sub block at a time. The chroma reconstruction will be executed after four 8x8 luma blocks are done. This architecture can make 8 parallelism PEs to achieve almost 100% hardware utilization and save operating cycles from unmeaningful reconstruction. It only takes 906 cycles to process one MB.

### 4.3. Hardware Implementation Result and Comparison

The proposed hardware architecture was designed with Verilog HDL code and synthesized by UMC 90-nm technology. The total gate count is 100.4k and the gate count for each component is listed in Table 1. The comparison of previous design is in Table 2. Comparing to previous works, our design is the first hardware architecture that can support H.264 high profile intra prediction at 1080p HD spec and can process inter mode reconstruction without extra hardware cost or operating cycles.

## 5. CONCLUSION

In this paper, the hardware oriented two stage intra prediction algorithm and hardware architecture is proposed. The proposed algorithm and architecture can break intra prediction data dependency and improve hardware utilization with nearly no quality loss. Comparing to previous works, our design can support more features with lower operation frequency. The proposed architecture is the first hardware architecture that can support H.264/AVC high profile intra prediction at 1080p HD 30 fps spec.

## 6. REFERENCES

- [1] G. Sullivan, P. Topiwala, and A. Luthra, "The H.264 advanced video coding standard : Overview and introduction to the fidelity range extensions," in *SPIE Conference on Application of Digital Image Processing XXVII*, 2004.

**Table 2.** Comparison of previous works and proposed architecture

Design	[3]	[4]	[5]	Proposed
Technology	0.25um	0.18um	0.35um	0.09um
Max. operation freq.	55MHz	125MHz	108MHz	223MHz
Max. target size	720x480	1280x720	1280x720	1920x1080
Gate count	74.6k	76.8k	192.4k	100.4k
On-chip memory	14336bit	9728bit	18176bit	12288bit
Cycles/MB	<1300	<1080	<927	<906
Pixel parallelism	4	4	4	8
Decision method	DCT	DCT	Hadamard	DCT
Support Intra_8x8	No	No	No	Yes
Support 8x8 DCT	No	No	No	Yes
Support Inter Rec.	No	No	No	Yes
Freq. for 720p HD	N/A	117MHz	108MHz	98MHz
Freq. for D1	54MHz	43MHz	38.5MHz	37MHz

- [2] D. Marpe and et al., "H.264/MPEG4-AVC fidelity range extensions : Tools, profiles, performance, and application areas," in *Proc. IEEE ICIP*, Sept. 2005, vol. 1, pp. 593–596.
- [3] Y.-W. Huang and et al., "Analysis, fast algorithm, and VLSI architecture design for H.264/AVC intra frame coder," *IEEE Trans. on CSVT*, vol. 15, no. 3, pp. 378–401, Mar. 2005.
- [4] C.-W. Ku and et al., "A high-definition H.264/AVC intra-frame codec IP for digital video and still camera applications," *IEEE Trans. on CSVT*, vol. 16, no. 8, pp. 917–928, Aug. 2006.
- [5] K. Suh, S. Park, and H. Cho, "An efficient hardware architecture of intra prediction and TQ/IQIT module for H.264 encoder," *ETRI Journal*, vol. 27, 2005.
- [6] Y.-W. Huang and et al., "A 1.3tops H.264/AVC single-chip encoder for HDTV applications," in *Proc. of IEEE ISSCC*, 2005, pp. 128–588.
- [7] T.-D. Chuang, Y.-H. Chen, C.-H. Tsai, and L.-G. Chen, "Analysis and architecture design for multi-transform for H.264/AVC high profile," in *International SoC Design Conference*, 2006.