A HIGH-PERFORMANCE ARCHITECTURE OF JPEG2000 ENCODER

Damian Modrzyk, and Michał Staworko

Integrated Circuits Development Department, Evatronix SA, Gliwice, Poland
Institute of Telecommunications, Warsaw University of Technology, Warsaw, Poland
e-mail: damian.modrzyk@evatronix.com, m.staworko@elka.tele.pw.edu.pl

ABSTRACT
This article presents hardware architecture of JPEG2000 encoder core, oriented for HD video broadcast and surveillance applications. Thanks to developed efficient 2-D DWT engine that is capable of computing four coefficients per clock cycle, and adopted two EBCOT TIER-1 modules, with smart switching of the channels, the maximum compression speed of 180 Msamples/s at 100 MHz, in lossy mode is achieved. The architecture is implemented in VHDL and synthesized for FPGA devices, and ASIC 0.13 μm CMOS technology. Performance simulations, conducted on a set of natural images and video sequences, have revealed that the encoder is capable of processing 1080p 4:4:4 signal with a speed of 30 frames per second. Additionally, an excellent quality of reconstructed images has been observed, with respect to the reference, software encoder.

1. INTRODUCTION
In the age of acquisition and processing of high resolution images and video, efficient systems of signal compression-decompression play a fundamental role. This issue is extremely important in the case of 3D video signals, where simultaneous processing of more than one frame is assumed. Present multimedia applications require also resolution and quality scalability, flexible rate control, region of interest coding, increased colour depth, error resilience and other demanding features. JPEG2000 still-image compression standard [1][2], which outperforms significantly JPEG standard, meets all these demands in a single codestream. Thanks to many sophisticated coding tools, such as discrete wavelet transformation (DWT) and two-tiered embedded block coding with optimized truncation (EBCOT) [3], it allows to obtain excellent compression performance and is commonly used in digital cinema (DCI) and HDTV applications.

By efficiency of a compression system we mean its extensive processing power. Nowadays, the requested throughput of video encoders is so enormous that use of their software implementations is out of the question in real-time applications. The only solution for a high speed, low latency realisations of complex image encoding algorithms are specialized hardware implementations. Obviously, the superior throughput of hardware solutions is always balanced by FPGA or ASIC area resources and power consumption. In the literature, several architectures of high-performance JPEG2000 encoder have been proposed [4-11]. All these implementations meet different design challenges that the researchers try to overcome. Firstly, the mismatch between word-based DWT dataflow and sequential, computational intensive bit-plane processing of EBCOT algorithm requires significant amount of on-chip memory resources or high external memory bandwidth, to buffer data, passed between these two blocks. Secondly, an entropy coding algorithm in JPEG2000 requires sophisticated control mechanisms and many branching conditions. In hardware it results in complex, combinatorial logic and considerable length of the critical paths. The design maximum frequency is also deteriorated due to long arithmetic paths in the DWT processor. Finally, an enormous complexity of JPEG2000 compression algorithms, which enforces use of pipeline computations, reflects in significant area of the encoder chip.

In this work we propose a high performance, hardware architecture of JPEG2000 encoder targeted for HD image resolutions. High throughput of the presented design is achieved by two main strategies. Firstly, very efficient 2-D DWT engine that computes four coefficients per one clock cycle is applied in the encoder. Secondly, parallel processing of subbands is obtained through exploiting two channels of EBCOT TIER-1 module, with a smart switching technique. What is more, single TIER-1 module incorporates several techniques of fast and efficient entropy coding, elaborated for the base JPEG2000 design [12-14]. The encoder utilizes external SDRAM memory to store intermediate, compression results, what minimizes the internal memory resources and enables realisation of the pipelining at a frame level. All these issues, together with improved external memory access scheduling, result in a compression performance that meets requirements of real-time HDTV applications. The rest of this paper is organized as follows. Section 2 contains a brief description of JPEG2000 coding tools. Proposed architecture of the JPEG2000 encoder together with performance improvement strategies are gathered in Section 4. Finally, conclusions and future work aspects are summarized in Section 5.

2. JPEG2000 OVERVIEW
The block diagram of JPEG2000 encoder is presented in Figure 1. Input image components are firstly pre-processed by conducting DC level shifting and multiple component transform on the samples. The image plane is partitioned into rectangular, non-overlapping tiles, which are compressed independently. Each tile is decomposed by discrete wavelet transformation into subbands with a certain number of decomposition levels. DWT coefficients are then uniformly quantized, to be further partitioned into code-blocks. The entropy coder processes the bit-planes of coefficients in the code-block, to form the output JPEG2000 codestream [2].

![Block diagram of JPEG2000 encoder](image_url)

Figure 1 – Block diagram of JPEG2000 encoder.
Rate controller, and rate-distortion optimization blocks are not directly standardised in the norm, however they are given informative and pose an important part of the encoder.

2.1 Multiple Component Transform

JPEG2000 standard enables compression of multiple-component images (e.g. RGB, YCbCr) of arbitrary bit-depths. To provide better compression efficiency the standard defines two types of de-correlating transformation derived on the first three input image components. These are reversible component transform (RCT) Eq.1, used in lossless compression mode and irreversible component transform (ICT) Eq.2, utilized in lossy mode. Prior to computation of the transformations, DC level shifting on input samples is performed, in the case of their unsigned representation.

\[
Y = 0.299R + 0.587G + 0.114B
\]
\[
C_b = -0.16875R - 0.33126G + 0.5B
\]
\[
C_r = 0.5R - 0.41869G - 0.08131B
\]  
\[
C_y = R - G
\]  
\[
C_c = B - G
\]  

\(Y\) denotes here luminance component, whereas \(C_p, C_y\) are chrominance components.

2.2 Discrete Wavelet Transform

Input image, after MCT transform is partitioned into rectangular, non-overlapping blocks – so-called tiles, which in turn are processed independently by discrete wavelet transform algorithm. DWT applied in JPEG2000 is a two-dimensional (2-D), multi-level filtering method that consists of two 1-D filtering operations performed in vertical and horizontal directions respectively [2]. Each 1-D wavelet transform decomposes array of samples into low-pass set – downsampled, low-resolution approximation of the original signal, and high-pass set – downsampled residual of the original signal. As a result of DWT decomposition, the tile is divided into four subbands – LL, HL, LH, HH, which contain transform coefficients with different horizontal and vertical spatial frequency characteristics. The LL subband can be further, recursively decomposed in a dyadic fashion. As an example three-level DWT decomposition, which results in 10 subbands is presented in Figure 2.

![Figure 2 – Three-level DWT decomposition of image tile.](image)

JPEG2000 defines both irreversible and reversible discrete wavelet transformation. The first one is implemented as 9/7-tap Daubechies filter [2], whereas the second is implemented by means of 5/3-tap filter. The filters can be implemented in a convolution-based or a lifting-based fashion, with symmetric extensions of the samples at the signal boundaries. In hardware realisations, the most popular is a lifting approach, which has many advantages over traditional FIR filtering [15]. It is also recommended in the JPEG2000 standard. Irreversible DWT in JPEG2000 consists of four lifting steps and two scaling steps as it is presented in Eq.3.

\[
\begin{align*}
Y(2n-1) &= X_{in}(2n-1) + \alpha(X_{in}(2n) + X_{in}(2n+2)) \\
Y(2n) &= X_{in}(2n) + \beta(Y(2n-1) + Y(2n+1)) \\
Y(2n+1) &= Y(2n) + \gamma(Y(2n-1) + Y(2n+2)) \\
Y(2n+2) &= Y(2n+1) + \delta(Y(2n-1) + Y(2n+1))
\end{align*}
\]  
\[
\begin{align*}
Y(2n) &= X_{in}(2n) + \frac{Y(2n-1)+Y(2n+2)}{2} \\
Y(2n+1) &= Y(2n) + \frac{Y(2n-1)+Y(2n+2)}{4}
\end{align*}
\]  

where \(\alpha, \beta, \gamma, \delta\) denote lifting coefficients, \(K\) is a scaling factor and \(X_{in}(2n), X_{in}(2n+1)\) represent even and odd samples of input, respectively. Reversible DWT consists of only two lifting steps. High-pass and low-pass coefficients are calculated using following equations:

\[
\begin{align*}
Y(2n-1) &= X_{in}(2n-1) + \alpha(X_{in}(2n) + X_{in}(2n+2)) \\
Y(2n) &= X_{in}(2n) + \beta(Y(2n-1) + Y(2n+1)) \\
Y(2n+1) &= Y(2n) + \gamma(Y(2n-1) + Y(2n+2)) \\
Y(2n+2) &= Y(2n+1) + \delta(Y(2n-1) + Y(2n+1))
\end{align*}
\]  
\[
\begin{align*}
Y(2n) &= X_{in}(2n) + \frac{Y(2n-1)+Y(2n+2)}{2} \\
Y(2n+1) &= Y(2n) + \frac{Y(2n-1)+Y(2n+2)}{4}
\end{align*}
\]  

2.3 Quantization

After discrete wavelet transform, the dynamic range of coefficients is decreased by scalar quantization process. The only exception is lossless compression mode, where the quantization step is equal 1. In other case, subband DWT coefficients – \(a_b(u,v), v\), are quantized according to Eq.5:

\[
q_U(u,v) = \text{sign}(a_b(u,v)) \left\lfloor \frac{a_b(u,v)}{\Delta_b} \right\rfloor \Delta_b = 2^{k_b - \epsilon_b} \left(1 + \frac{H_b}{2^m}\right)
\]

\(q_U(u,v)\) represents quantization step of a particular subband and \(R_b\) denotes dynamic range of the coefficients, which is dependent of image component bit depth and type of the subband (LL, HL, LH or HH). \(k_b, \epsilon_b, H_b\) are the quantization step’s exponent and mantissa respectively. Quantized coefficients are passed to entropy coding algorithm in a sign-magnitude representation.

2.4 Embedded Block Coding

The embedded block coding in JPEG2000 is divided into two parts – TIER-1 and TIER-2. During TIER-1 coding, each subband, containing quantized DWT coefficients is partitioned into rectangular code-blocks, whose dimensions must be power of two. In Figure 3, there is presented an example of subband decomposition into sixteen code-blocks of size 64x64, and further layers of the entropy coding algorithm – bit-planes and stripes. The bit-planes are processed from the most significant one (containing non-zero element) to the least significant one. Each bit of the analyzed bit-plane is encoded by one of the three coding processes called: significant propagation, magnitude refinement and clean-up [1][2]. The coding passes utilize a special scan order to process the stripes. As a result of bit-plane coding, a pair of context and binary decision values is generated for each bit position, and given at the input of MQ-encoder, adopted in the JPEG2000 standard. The context data generated by bit-plane coder is used to derive the estimated probability value, from the predefined look-up table with 47 entries. Code-bytes obtained from arithmetic encoder are distributed across one or more layers within the codestream [2]. Each layer successively contributes a certain level of improvement into reconstructed image. Code data representing a specific tile, layer, component and resolution form packets.

TIER-2 algorithm holds the process of final codestream generation. It is done by formation of several headers (main header, tile
headers etc.), with special markers and marker segments, and incorporating code-bytes collected in the packets. These can be put into codestream in five different orders, what is defined in the standard as a progression order [1][2].

### 3. PROPOSED ENCODER ARCHITECTURE

Figure 4 presents block diagram of the proposed, high-performance JPEG2000 encoder. It consists of interface controller module, multiple component transform (MCT) module, discrete wavelet transform engine (DSP), DWT coefficients de-multiplexer (CHANNEL DMUX), two EBCOT TIER-1 modules, EBCOT TIER-2 stream formatter and memory management units: TIER-1 MMU, TIER-2 MMU, external memory scheduler. Configuration of the encoder is performed via interface controller block. It communicates with a master device using three interfaces. Interrupts interface informs about compression state within the encoder. Configuration interface enables setting of the compression parameters such as decomposition levels, code-block size, tile size, filter mode and etc. Codestream interface is a JP2 stream output. Despite communication with external host, the interface controller provides synchronization and enables processing in the encoder sub-modules (MCT, DSP, EBCOT). Intermediate results of the encoder are buffered in the external SDRAM memory, to be further processed by DSP block.

#### 3.1 MCT Module

Input image samples are transferred directly to multiple component transform module through video input interface, synchronized by separate, 200 MHz clock signal – clk_mct. The module performs DC level shifting and then applies reversible component transform or irreversible component transform according to Eq. 1 and Eq. 2. The encoder architecture of the module is multi-level pipelined in order to decrease critical path. Obtained YCbCr coefficients are buffered in the external memory, to be further processed by DSP block.

#### 3.2 DSP Module

DSP module is an implementation of discrete wavelet transformation and quantization algorithms. The module performs multi-level, two-dimensional reversible or irreversible transform, with symmetric boundary extension. Presented encoder provides up to five DWT levels in lossy compression and up to seven levels in lossless mode. The DSP block consists of four sub-modules as presented in Figure 5. The sub-modules take at the input a compression parameters set, asserted by interface controller module.

![Figure 5 – Internal structure of DSP block.](image)

The first, DSP_IN block is a tile formatter. It reads blocks of 136x136 image samples (128x128 with four samples of the overlap on each boundary), from the external memory utilizing memory scheduler. Then, it sends to DWT processor four samples simultaneously: even sample from even line – le_so_data, odd sample from even line – le_so_data, even sample from odd line – lo_so_data, and odd sample from odd line – lo_so_data. In the case of DWT levels greater than 1, blocks of LL subband samples are read from the external memory to provide further decompositions.

DSP_DWT is an efficient implementation of the 2-D discrete wavelet transform. The two-dimensional filtering is partitioned into two vertical 1-D (performed first) and two horizontal 1-D, lifting-based transform engines. Such a structure, depicted in Figure 6, after a latency of four image lines, allows to obtain four DWT coefficients (from LL, HL, LH and HH subband respectively) in one clock-cycle. Each 1-D DWT processor has eight-level pipeline optimized to minimize rounding error between the stages. The proposed DWT module enables 5/3-tap and 9/7-tap filtering in the same hardware architecture. Choice of the transform type is programmable. When reversible transform is enabled, the result coefficients are taken from the fourth stage of the pipeline. Before sending DWT coefficients to the output, the data are scaled according to steps 5-6 of Eq.3. DSP_DWT block can process incoming samples with a maximum throughput of 800 Msamples/s, when synchronized by 200 MHz clock pulse. In the presented encoder the performance is equal 400 Msamples/s, assuming 100 MHz pixel clock. DSP_QUANT sub-module is a straightforward implementation of the scalar quantization applied in ISO/IEC 15444-1 standard. The block performs simultaneous quantization of four input coefficients, derived for all DWT subbands. The quantization process is three-level pipelined, with the stages that correspond to Eq.5 formula.
Maximum performance of single EBCOT TIER-1 channel is approximately 50 Msamples/s at 100 MHz, in lossless compression mode. To increase the encoder performance we propose to incorporate two such TIER-1 channels. This solution is beneficial because processing of $HL$, $LH$ and $HH$ subbands is significantly shorter, than compression of $LL$ subband. It can be explained by residual, low-amplitude content in the high-pass code-blocks, being in contrast to low-pass, high-amplitude coefficients observed in the $LL$ code-blocks. Conducted experiments revealed that during $LL$ subband processing, by first TIER-1 channel, the second block is capable of compressing both $HL$ and $LH$ subbands. Then, only $HH$ subband is left for processing, what can be done by either channel 0 or channel 1, depending of the end of coding order. Compression of $HH$ subband is extremely rapid, comparing to other subbands.

3.4 EBCOT TIER-2 Module

EBCOT TIER-2 module implements final stage of the compression process. It performs sequence of tasks to form the final JPEG2000 codestream. When TIER-1 processing of all code-blocks ends, the TIER-2 module initiates second part of the rate control algorithm. It relies on accurate selection of truncation points rates that meet target bit-rate constraint. Next, the module starts producing JP2 stream, consisting of several headers and block-coded data assembled in contiguous segments – packets. The module supports two (component-resolution and resolution-component) progression orders from five defined in JPEG2000 standard. The packets are coded with a use of bit-stuffing technique and utilizing hierarchical tag trees to effectively create reduced-resolution levels of data arrays [14]. Typical, operating frequency of EBCOT TIER-2 module is 100 MHz, however it can be synchronized by maximum 150 MHz clock pulse.

4. IMPLEMENTATION RESULTS, PERFORMANCE

The JPEG2000 encoder design was implemented in VHDL hardware description language and synthesised for FPGA devices and ASIC technology. The encoder was configured to compress images of resolution up to 2048x2048 and pixel components dynamic range of 8-bit per sample. Maximum allowed tile size is equal 2048x2048. Such a configuration requires 64 MB of the external SDRAM memory. The core was extensively verified using our hardware-software verification platform. Its compression efficiency has been evaluated, by utilizing a repository of HD test images of 1280x720 and 1920x1080 resolutions, encoded with different parameter sets. Implementation results for Xilinx Virtex-5, Altera Stratix III FPGAs and TSMC 0.13 $\mu$m CMOS process are presented in Table 1. For the needs of comparison with competitive JPEG2000 architectures, the encoder chip specifications, under TSMC 0.18 $\mu$m CMOS technology are shown in Table 2. The best
evaluation of the proposed encoder vs. other existing works, can be found in performance index value, defined as design throughput per unit area, at 1 MHz. The PI obtained for our core is slightly smaller than indexes of Sanyo [6] and Chang [7] solutions, however none of the competing designs is capable of compressing the signal with a throughput of 180 Msamples/s. This work was also compared to three commercial JPEG2000 encoder products, aimed for HD video broadcast. Results are presented in Table 3 and concern 1080p signal at 4:2:2 sub-sampling. It can be seen that our solution provides the highest frame rate. Additionally proposed encoder core, in contrast to other commercial solutions, supports 1080p30 fps in 4:4:4 signal format. Remarkably is also the fact that at higher bit-rates this core introduces smaller distortions into compressed image than the reference, software encoder. It can be seen in form of R&D curve in Figure 8, obtained for one of the 1920x1080 test images. This situation is particularly visible when number of DWT levels is greater than two and can be explained by rounding error elimination strategies, applied in DWT module.

Table 1 – JPEG2000 encoder implementation results.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gate count</td>
<td>179 K</td>
<td>N/A</td>
<td>N/A</td>
<td>180 K</td>
</tr>
<tr>
<td>Core area</td>
<td>19 mm²</td>
<td>13 mm²</td>
<td>20.1 mm²</td>
<td>&lt; 20 mm²</td>
</tr>
<tr>
<td>Frequency</td>
<td>100 MHz</td>
<td>54 MHz</td>
<td>42 MHz</td>
<td>100 MHz</td>
</tr>
<tr>
<td>Power</td>
<td>2 W</td>
<td>0.9 W</td>
<td>384 mW</td>
<td>450 mW</td>
</tr>
<tr>
<td>Msamples/s</td>
<td>180</td>
<td>70</td>
<td>124</td>
<td>66</td>
</tr>
<tr>
<td>Perf. index</td>
<td>0.095</td>
<td>0.1</td>
<td>0.148</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 2 – JPEG2000 chip comparison in 0.18 µm CMOS process.

Table 3 – Commercial JPEG2000 encoder cores comparison.

5. CONCLUSIONS

In the paper we presented a novel, hardware architecture of JPEG2000 encoder, targeted for HD video applications. To meet high-performance requirements of the compression system, an efficient 2-D DWT engine is proposed and integrated in the encoder architecture, together with two parallel channels of EBCOT TIER-1 module. As a result of applied hardware optimizations the maximum throughput of 180 Msamples/s has been achieved for 100 MHz clock and 0.333 bpp compression rate. The proposed encoder core, at competitive area resources, provides superior frame rate and excellent compression quality for HD and Full HD video material. Future works will be focused on constraining power consumption of the design.

REFERENCES


Figure 8 – The proposed encoder codestream quality vs. reference JJ2000 5.1 Java application. Applied encoder settings: two DWT levels, code-block size – 64x64, tile size – 1024x1024.