

Optimal Quantized Lifting Coefficients for the 9/7 Wavelet

S. Barua[†], K. A. Kotteri^{*}, A. E. Bell^{*} and J. E. Carletta[†]

^{*}Virginia Tech, E-mails: kkotteri@vt.edu, abell@vt.edu

[†]University of Akron, E-mails: sb22@uakron.edu, carlett@uakron.edu

Abstract— The lifting structure has been shown to be computationally efficient for implementing filter banks. The hardware implementation of a filter bank requires that the lifting coefficients be quantized. The quantization method determines compression performance, hardware size, hardware speed and energy. We investigate the implementation of two lifting coefficient sets, rational and irrational, for the biorthogonal 9/7 wavelet. Six different approaches are used to find optimal quantized lifting coefficients from these sets. We find that the best hardware and PSNR performance is obtained using the rational coefficient set quantized with gain compensation and lumped scaling.

I. INTRODUCTION

Real-time image transmission from mobile wireless sensors requires low-power, low-cost, high-speed implementation in hardware. This paper investigates the optimal hardware implementation of the discrete wavelet transform (DWT) that is at the heart of the JPEG2000 image compression standard. A lifting approach is used—a method that offers computational advantages over the traditional convolution approach. We focus on the biorthogonal 9/7 wavelet filters of the JPEG2000 lossy coder; we design and evaluate new fixed-point approximations of the lifting coefficients for implementation on a field programmable gate array (FPGA) [1].

We employ several hardware and compression performance metrics. Hardware performance metrics include size or cost, throughput, and the energy required to process an image. Before implementation, hardware cost is estimated in terms of T , the total number of non-zero terms used when writing all the lifting coefficients in canonical-signed-digit (CSD) format [2]. After implementation, hardware cost is measured directly in terms of the *number of logic elements* used. Compression performance is evaluated by comparing the compressed and original images using *peak signal-to-noise ratio* (PSNR).

This paper is organized as follows. Section II provides background on the lifting structure and coefficients for the biorthogonal 9/7 wavelet filters. Section III describes our quantization methods and compares the hardware and compression performance of our new quantized coefficient designs. Our conclusions are presented in Section IV.

II. BACKGROUND

Fig. 1 shows the structure of one stage of a two-channel biorthogonal filter bank. For the 9/7 DWT, filters $H(z)$ and $G(z)$ are symmetric FIR filters with nine and seven taps, respectively. Traditionally, the filters are implemented using

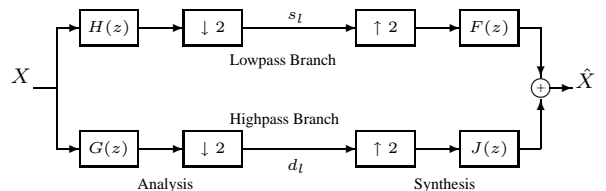


Fig. 1. Two-channel biorthogonal scalar wavelet filter bank.

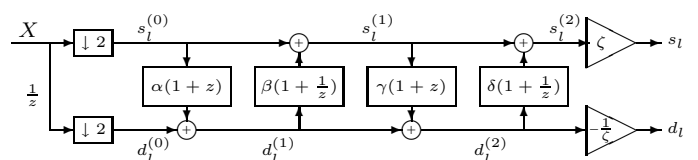


Fig. 2. Lifting structure of the biorthogonal 9/7 DWT.

convolution. This implementation is non-polyphase, and suffers from inefficient hardware utility and low throughput.

Lifting is a more efficient polyphase alternative to the traditional convolution implementation. A lifting structure requires fewer computations and has twice the throughput. The lowpass (LPF) and highpass (HPF) filters in Fig. 1 are replaced with a sequence of shorter lift and update filters. As shown in Fig. 2, a lifting implementation of the 9/7 DWT uses four two-tap symmetric filters, followed by a scaling factor. The unquantized values of lifting coefficients α , β , γ , δ and ζ are obtained by factoring the 9/7 polyphase matrix; the values are listed in Table I under the column heading “Irrational” [3]. The synthesis lifting structure inverts the scaling factor and reverses the order of the lift and update filters.

The irrational lifting coefficients in Table I have infinitely long binary representations; for a fast hardware implementation these coefficients must be *quantized*, or approximated in fixed-point. In general, the more CSD terms used to represent the lifting coefficients (the higher the value of T), the closer the quantized lifting coefficients are to the unquantized coefficients—and consequently, the closer the PSNR values achieved by the fixed-point hardware are to the PSNR values obtained with unquantized coefficients. The lifting structure is inherently *orthogonal*; when synthesis immediately follows analysis (i.e. no compression), analysis is *exactly* inverted regardless of the quantization used for α , β , γ , δ and ζ . Thus lifting has a significant advantage over convolution where perfect reconstruction is not guaranteed after the coefficients are quantized.

The unquantized 9/7 analysis LPF and HPF have four zeros

This material is based upon work supported by the National Science Foundation under Grants 9876025, 0218672, and 0217894.

TABLE I
LIFTING COEFFICIENTS FOR THE BIORTHOGONAL 9/7 WAVELET FILTERS.

| | Irrational | Rational |
|----------|------------------|---------------|
| α | -1.58613434... | -3/2 |
| β | -0.0529801185... | -1/16 |
| γ | 0.882911076... | 4/5 |
| δ | 0.443506852... | 15/16 |
| ζ | 1.14960439... | $4\sqrt{2}/5$ |

at $z = -1$ and 1 respectively. These zeros determine the smoothness of the analysis and synthesis scaling functions and prevent DC leakage in the analysis HPF. A set of rational lifting coefficients was obtained by moving two zeros of the analysis LPF away from $z = -1$ [4], [5]. These coefficients are listed in Table I under ‘‘Rational’’. The rational lifting coefficients generate PSNR values almost exactly equal to the irrational coefficient PSNR values [5]. The rational coefficients are also more easily quantized than the irrational; all except γ and ζ have finite binary representations and can be represented exactly—with no approximation—with a small T .

III. NEW QUANTIZATION METHODS AND RESULTS

A. Quantization Objectives

In the biorthogonal 9/7 filter bank shown in Fig. 1, the synthesis section exactly inverts the analysis section, so that the reconstructed image, \hat{X} equals the original image, X (to within an integer shift l) when the following PR conditions are met:

$$F(z)H(z) + J(z)G(z) = 2z^{-l}(\text{no-distortion}) \quad (1)$$

$$F(z)H(-z) + J(z)G(-z) = 0. \quad (\text{no-aliasing}) \quad (2)$$

The no-aliasing condition is satisfied by design in the usual way. This reduces the problem from the design of four filters to two; in lifting we are concerned with the design of the two analysis filters $H(z)$ and $G(z)$.

Quantization of the five lifting coefficients changes the magnitude response of the analysis filters. Previous research has shown that compression performance requires that the magnitude response of the quantized analysis filters ($H'(z)$ and $G'(z)$) closely approximate the magnitude response of the unquantized 9/7 filters ($H(z)$ and $G(z)$) [2]. Quantized values of α , β , γ and δ determine the shape of the magnitude responses; whereas, quantized values of ζ and $1/\zeta$, determine the DC gain and Nyquist gain of $H'(z)$ and $G'(z)$ respectively.

In the next section we develop quantization techniques for both the irrational and rational lifting coefficients that attempt to: (1) perfectly reconstruct the DC component ($|H'(0)||G'(\pi)| = 2$); (2) prevent DC leakage and obviate the checkerboarding artifact at low bit rates ($|G'(0)| = 0$); and, (3) minimize the mean-squared-error (MSE) of the magnitude responses of the quantized LPF and HPF ($|H'(\omega)|$, $|G'(\omega)|$).

B. Quantization of irrational coefficients

Lifting coefficient quantization can be posed as a problem of allocating a fixed number T of CSD terms to the six

coefficients (α , β , γ , δ , ζ and $1/\zeta$) while retaining acceptable subjective quality for the reconstructed image. Here, ‘‘acceptable’’ subjective quality is defined by the absence of the checkerboarding artifact. (For all approaches, the T used is shown in Table II, and is the smallest number of terms that did not result in checkerboarding.) We developed the following three approaches for quantizing the irrational lifting coefficients.

1) *Mostly Uniform Allocation (MUA)*: Here all six lifting coefficients are allotted the same base number of CSD terms and then the largest magnitude coefficients are allotted extra terms. Once the number of terms for a given coefficient is decided, the unquantized coefficient is quantized to the closest CSD coefficient with that number of terms. We used three terms as the base, allotting three extra terms for α and ζ , for a total of $T = 21$ terms. Table II lists these coefficients under ‘‘irrational: MUA’’.

2) *Exhaustively Searched Allocation (ESA)*: All possible allocations of a fixed number T of CSD terms to the six lifting coefficients are examined. For each allocation, α , β , γ and δ are quantized to the closest CSD coefficient with the allotted number of terms; together, they determine the shape of the magnitude response of the quantized filters. Next, a gain factor ζ_{comp} is computed such that the DC gain of the analysis LPF equals $\sqrt{2}$; this gain compensates for the change in DC gain due to quantization. Quantized versions of the gain factors are then obtained by approximating ζ_{comp} and $1/\zeta_{comp}$ using the terms allotted to them. The following cost function, CF_{ESA} , is used to evaluate the quantized coefficient sets:

$$CF_{ESA} = MSE_h + MSE_g + dev_dc, \quad (3)$$

where

$$MSE_h = \frac{1}{\pi} \int_0^\pi (|H(\omega)| - |H'(\omega)|)^2 d\omega,$$

$$MSE_g = \frac{1}{\pi} \int_0^\pi (|G(\omega)| - |G'(\omega)|)^2 d\omega,$$

$$dev_dc = |2 - |H'(0)||G'(\pi)||.$$

CF_{ESA} measures the deviation of $H'(z)$ from $H(z)$, the deviation of $G'(z)$ from $G(z)$, and the degree to which an image’s DC component is reconstructed. The smallest number of CSD terms for which the reconstructed images did not exhibit checkerboarding was $T = 19$. The coefficient set judged to be best resulted in the smallest CF_{ESA} using 19 terms. Table II lists the resulting coefficients under ‘‘irrational:ESA’’.

3) *Simulated Annealing (SA)*: Simulated annealing searches for the best quantized coefficient set given a *domain* for each coefficient [6][7]. The domain denoted $\mathcal{D}(B, T_{max})$ is the set of all real numbers between B and $-B$ whose CSD representation is up to B bits wide and has up to T_{max} non-zero terms. SA has the advantage that constraints can be placed on the bit widths of the lifting coefficients; however, there is no direct control over the total T for the set.

Coefficients are quantized in two passes. In the first pass, SA iteratively searches for the best four-coefficient set $\{\alpha, \beta, \gamma, \delta\}$. For this pass, scaling factors ζ and $1/\zeta$ are set to 1, and the

TABLE II
QUANTIZED LIFTING COEFFICIENTS FOR BIORTHOGONAL 9/7 WAVELET FILTERS.

| | Quantized irrational coefficients | | | Quantized rational coefficients | | |
|----------------|-----------------------------------|-----------------|------------|---------------------------------|----------------|-------------------|
| | MUA T=21 | ESA T=19 | SA T=19 | MUA T=20 | MUA-LS T=19 | MUA-LSGC T=21 |
| α | -1.5859375 | -1.59375 | -1.5546875 | -1.5 | -1.5 | -1.5 |
| β | -0.052734375 | -0.0546875 | -0.0546875 | -0.0625 | -0.0625 | -0.0625 |
| γ | 0.8828125 | 0.8828125 | 0.85546875 | 0.7998046875 | 0.7998046875 | 0.7998046875 |
| δ | 0.44140625 | 0.4453125 | 0.4453125 | 0.46875 | 0.46875 | 0.46875 |
| ζ | 1.1484375 | 1.140625 | 1.1328125 | 1.13134765625 | 0.7998046875 | 0.7998046875 |
| $1/\zeta$ | -0.87109375 | -0.876708984375 | -0.8828125 | -0.883789063 | -1.25 | -1.25030517578125 |
| Lumped scaling | N/A | N/A | N/A | N/A | $\sqrt{2}$ | $\sqrt{2}$ |

TABLE III
PSNR AND HARDWARE PERFORMANCE OF THE SIX QUANTIZED LIFTING COEFFICIENT DESIGNS.

| | | Unquantized | Quantized irrational coefficients | | | Quantized rational coefficients | | |
|------------------------|-------|-------------|-----------------------------------|------------------|------------|---------------------------------|----------------|------------------|
| | | | MUA T=21 | ESA T=19 | SA T=19 | MUA T=20 | MUA-LS T=19 | MUA-LSGC T=21 |
| Aerial2 | 1:1 | 58.75 | 56.13 | 58.76 | 58.67 | 58.35 | 57.41 | 58.80 |
| | 8:1 | 32.76 | 32.75 | 32.76 | 32.75 | 32.76 | 32.76 | 32.76 |
| | 32:1 | 27.99 | 27.98 | 27.99 | 27.97 | 27.98 | 27.98 | 27.98 |
| | 100:1 | 24.79 | 24.78 | 24.78 | 24.75 | 24.78 | 24.78 | 24.77 |
| Bike | 1:1 | 58.7 | 54.53 | 58.70 | 58.52 | 58.03 | 56.58 | 58.75 |
| | 8:1 | 32.83 | 32.81 | 32.82 | 32.8 | 32.82 | 32.82 | 32.82 |
| | 32:1 | 24.81 | 24.79 | 24.80 | 24.78 | 24.81 | 24.81 | 24.81 |
| | 100:1 | 21.24 | 21.22 | 21.21 | 21.1 | 21.21 | 21.2 | 21.20 |
| Woman | 1:1 | 58.68 | 55.48 | 58.68 | 58.55 | 58.23 | 57.15 | 58.74 |
| | 8:1 | 33.68 | 33.68 | 33.69 | 33.7 | 33.69 | 33.69 | 33.69 |
| | 32:1 | 27.29 | 27.28 | 27.28 | 27.22 | 27.28 | 27.28 | 27.28 |
| | 100:1 | 24.51 | 24.49 | 24.48 | 24.44 | 24.48 | 24.48 | 24.48 |
| size (#logic elements) | | | 2705 | 2284 | 2441 | 1981 | 1958 | 2123 |
| f_{max} (MHz) | | | 60.54 | 72.59 | 68.96 | 79.92 | 77.08 | 70.66 |
| Latency(clocks) | | | 53 | 51 | 51 | 47 | 49 | 49 |
| Energy (mjoules) | | | 8.69 | 8.06 | 8.3 | 7.56 | 7.52 | 7.81 |
| approx format, s_l | | | (47, -38) | (41, -32) | (45, -36) | (40, -31) | (39, -30) | (39, -30) |
| details format, d_l | | | (40, -31) | (40, -31) | (38, -29) | (34, -25) | (27, -17) | (39, -29) |

quality of a candidate set is measured by the cost function, CF_{SA} . The cost function uses the MSE between the quantized and unquantized magnitude responses. MSE in the stopband is weighted more heavily than MSE in the passband to ensure that $G'(z)$ has a magnitude response close to 0 at $z = 1$.

$$CF_{SA} = 0.75(MSE_{stopband}) + 0.25(MSE_{passband}),$$

where

$$MSE_{stopband} = \frac{2}{\pi} \int_{\pi/2}^{\pi} (|H(\omega)| - |H'(\omega)|)^2 d\omega + \frac{2}{\pi} \int_0^{\pi/2} (|G(\omega)| - |G'(\omega)|)^2 d\omega,$$

$$MSE_{passband} = \frac{2}{\pi} \int_0^{\pi/2} (|H(\omega)| - |H'(\omega)|)^2 d\omega + \frac{2}{\pi} \int_{\pi/2}^{\pi} (|G(\omega)| - |G'(\omega)|)^2 d\omega$$

In the second pass, SA uses the best quantized $\{\alpha, \beta, \gamma, \delta\}$ from the first pass and iteratively searches for the best quantized scaling factors ζ and $1/\zeta$. For this pass, the cost function CF_{ESA} from equation (3) is used.

Optimal quantized lifting coefficients were obtained using $\mathcal{D}(9, 4)$ for $\{\alpha, \beta, \gamma, \delta\}$ and $\mathcal{D}(8, 4)$ for the scaling factors. Using smaller bit widths or fewer non-zero digits resulted in checkerboarding. The optimal lifting coefficient set used a total of $T = 19$ CSD terms; Table II lists these coefficients under “irrational:SA”.

C. Quantization of rational coefficients

Of the rational lifting coefficients depicted in Table I, α , β and δ can be represented exactly in CSD with a small number of terms, but γ , ζ and $1/\zeta$ have infinitely long CSD representations. Hence the fixed-point quantization of γ , ζ and $1/\zeta$ require approximations. We developed the following three approaches for quantizing these three coefficients.

1) *Mostly Uniform Allocation (MUA)*: We again use the mostly uniform allocation scheme described above; the coefficients required $T = 20$ and are listed in Table II under “rational:MUA”.

2) *With Lumped Scaling (MUA-LS)*: The rational gain factors ζ and $1/\zeta$ include factors of $\sqrt{2}$ and $1/\sqrt{2}$, respectively. We can avoid implementing these factors in hardware by lumping them together. For example, one 2-D DWT stage

includes both row and column filtering and two factors are required. We defer the $\sqrt{2}$ factors until two can be combined into one factor of 2, 1/2 or 1. Multiplication and division by 2 are easily implemented as bit shifts.

With lumped scaling, γ and ζ are both equal to 0.8, which has an infinitely long binary (and CSD) representation. Therefore, γ and ζ must be approximated by quantization; the other coefficients, including $1/\zeta = 1.25$, are represented *exactly*. γ and ζ are quantized to γ' and ζ' by allocating the remaining T s to each coefficient. A total of $T = 19$ terms are used for the set and the coefficients are labeled “rational:MUA-LS” in Table II.

3) *With Lumped Scaling and Gain Compensation (MUA-LSGC)*: This method uses lumped scaling to avoid implementation of the $\sqrt{2}$ factors, but modifies $1/\zeta$ from its original value of 1.25 in order to compensate for the change in DC gain caused by quantization of γ . $1/\zeta$ is quantized using the terms allotted to it. A total of $T = 21$ terms are used and the coefficients are labeled “rational:MUA-LSGC” in Table II.

D. Results

Each of the quantized lifting coefficient sets in Table II was used to compute a five-level, non-expansive, symmetric extension DWT of three different grayscale images. The eight-bit grayscale images are part of the standardized digitized image set provided by ITU [8]. Perfect reconstruction (i.e. a 1:1 compression ratio) and three compression ratios (8:1, 32:1 and 100:1) were examined.

1) *Compression Performance*: Table III compares the PSNR performance of the six quantized coefficient designs. Among the quantized irrational designs, the ESA method performed best; among the quantized rational designs, all three designs performed equally well under compression, but at 1:1 the gain compensation technique employed in MUA-LSGC outperforms MUA-LS and MUA. Among all the quantized coefficient designs, the rational coefficients quantized with lumped scaling and gain compensation (MUA-LSGC) yield the best PSNR performance. The superior performance of ESA and MUA-LSGC can be understood in light of a previously identified quantized filter property: no-distortion MSE [2]. No-distortion MSE was $5.8e-11$ for ESA and $5.8e-19$ for MUA-LSGC; the other methods all resulted in MSEs on the order of 10^{-7} or 10^{-8} . All six designs resulted in similar smoothness of the analysis and synthesis scaling functions; moreover, the coding gain of all six designs (9.69-9.71) was almost identical to the unquantized coding gain (9.71).

2) *Hardware Performance*: Digital hardware was implemented on an Altera Apex EP20K1000EFC672-1X field programmable gate array for each of the six sets of coefficients. A multiplierless architecture is used, whereby multiplications are replaced by shifts and additions, organized in a fully pipelined tree of carry save adders with a ripple carry adder at the base of the tree. The Quartus II v2.1 software package is used for synthesis, placement and routing of structural VHDL descriptions of the architectures. The fixed-point formats of all signals are chosen so as to keep all bits of information, with

no truncation or round-off; hence, the formats depend on the coefficients used.

Table III depicts the hardware performance of the six quantized coefficient designs. The maximum possible operating frequency f_{max} is obtained via timing analysis, and indicates system throughput. The energy shown is that required to compute a five-level, two-dimensional DWT of a 512×512 block of random pixels. The table also shows the fixed-point formats of the approximation and detail coefficients produced; the notation is (n, f) : n is the total number of bits, including the sign, and 2^f is the weight of the least significant bit.

The quantized rational designs are smaller, faster, and require less energy than the quantized irrational designs. The primary reason is that the rational coefficients have narrower bit widths. (For example, the first coefficient, -1.5, requires only three bits.) In the lifting structure, intermediate signals grow in bit width after each lifting stage, with the amount of growth determined by the corresponding coefficient. In the rational designs, the bit width grows more slowly, with bit widths near the input of the system being particularly small. These smaller bit widths imply narrower adders, and smaller hardware. Smaller hardware translates into lower energy consumption. Throughput is a consequence of the width of the largest adder in the system, so lower bit widths at the output make for faster, higher performance systems.

IV. CONCLUSION

Optimal, quantized lifting coefficients for the biorthogonal 9/7 wavelet filters were derived using our new method: we refer to them as MUA-LSGC. The best design began with the unquantized, rational coefficients proposed previously [4], [5]. The MUA-LSGC set offers the best image compression performance (in terms of PSNR) with a small, fast hardware implementation. The energy required by MUA-LSGC to compute the DWT of an image is lower than any of the quantized coefficient designs that began with the irrational lifting coefficients.

REFERENCES

- [1] A. Skodras, C. Christopoulos, and T. Ebrahimi, “The JPEG2000 still image compression standard,” *IEEE Signal Processing Mag.*, vol. 18, no. 5, pp. 36–58, September 2001.
- [2] K. A. Kotteri, A. E. Bell, and J. E. Carletta, “Design of multiplierless, high-performance, wavelet filter banks with image compression applications,” *IEEE Trans. Circuits Syst. II*, to appear in Dec 2003.
- [3] I. Daubechies and W. Sweldens, “Factoring wavelet transforms into lifting steps,” *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247–269, 1998.
- [4] D. Tay, “A class of lifting based integer wavelet transform,” in *Proc. IEEE Int’l Conference on Image Processing*, vol. 1, 2001, pp. 602–605.
- [5] Z. Guangjun, C. Lizhi, and C. Huowang, “A simple 9/7-tap wavelet filter based on lifting scheme,” in *Proc. IEEE Int’l Conference on Image Processing*, vol. 2, 2001, pp. 249–252.
- [6] N. Benvenuto, M. Marchesi, and A. Uncini, “Applications of simulated annealing for the design of special digital filters,” *IEEE Trans. Signal Processing*, vol. 40, no. 2, pp. 323–332, February 1992.
- [7] A. Corona, M. Marchesi, C. Martini, and S. Ridella, “Minimizing multimodal functions of continuous variables with the simulated annealing algorithm,” *ACM Trans. Math. Software*, vol. 13, no. 3, pp. 262–280, September 1987.
- [8] *T.24: Standardized digitized image set*, ITU Std., June 1998.