We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

14

Condition on Word Length of Signals and Coefficients for DC Lossless Property of DWT

Masahiro Iwahashi[1] and Hitoshi Kiya[2]
[1]*Nagaoka University of Technology, Niigata,*
[2]*Tokyo Metropolitan University, Tokyo,*
Japan

1. Introduction

A discrete wavelet transform (DWT) has been widely applied to various digital signal processing techniques. It has been designed under a certain condition such as perfect reconstruction, aliasing cancellation, regularity, vanishing moment, etc. This article introduces a new condition referred to "DC lossless". It guarantees lossless reconstruction of a constant input signal (DC signal) instead of rounding of signal values and coefficient values inside a transform. The minimum word length of the values under the new condition is theoretically derived and experimentally verified.

Since JPEG 2000 algorithm based on the discrete wavelet transform (DWT) was adopted as an international standard for digital cinema video coding [1], high speed and low power implementation of a DWT has been becoming an issue of great importance [2,3]. In designing a DWT, its coefficient values and signal values are assumed to be real numbers. However, in implementation, they are rounded to rational numbers so that they are expressed with finite word length representation in binary digit. Therefore it is inevitable to have rounding errors inside a DWT processing unit.

In this article, we derive a condition on word length of coefficient values and that of signal values of a DWT such that the transform becomes lossless for a DC signal. Under this condition (DC lossless condition), it is theoretically guaranteed that an output signal contains no error in spite of rounding of coefficients and signals inside the DWT. We treat the irreversible 9-7 DWT adopted by the JPEG 2000 for lossy coding of image signals as an example.

In case of the 5-3 DWT in JPEG 2000 for lossless coding, benefiting from its lifting structure [4-6], lossless reconstruction of any signal is guaranteed even though signals and coefficients are rounded. On the contrary, it does not hold for the 9-7 DWT because of scaling for adjusting DC gain of a low pass filter in a forward transform [7]. However, we have pointed out that it became possible to be lossless for a DC signal under a certain condition on word length of coefficients and signals [8].

This DC lossless condition is a necessary condition for the regularity which has been analyzed by numerous researchers to improve coding performance of a transform. When the regularity is not satisfied, the DWT has some problems such as a checker board artifact which is observed in a reconstructed signal as unnecessary high frequency noise in flat or

smooth region of a signal [9]. It also brings about DC leakage which decreases the coding gain of a transform [10].
The regularity has been structurally guaranteed for a two channel quadrature mirror filter bank (QMF) [9] and the DCT [10] respectively. However, since these previous methods were based on the lattice structure, these are not directly applicable to the lifting structure of the 9-7 DWT. Beside these relations to the regularity, the DC lossless condition itself is also considered to be important for white balancing of a video system in which the DC signal is used as a reference input for calibration [11].
This article aims at deriving the DC lossless condition theoretically and clarifying the minimum word length of signals and coefficients. In conventional analysis, errors due to shortening of word length of signals (signal errors) were described as 'additive' to a signal [7,12]. They were treated as independent and uniformly distributed white noise. On the other hand, errors due to rounding of coefficients (coefficient errors) were described as 'multiplicative' to a signal and evaluated with the sensitivity [13-15]. It should be noted that the signal error and the coefficient error have been treated independently. Unlike those conventional approaches, we utilize mutual effect between rounding of signals and that of coefficients. Introducing a new model which unifies the coefficient error and the signal error, we define tolerance for those errors as a parameter to simultaneously control both of word length of signals and that of coefficients.
As a result of our theoretical analysis, the minimum word length of signals and that of coefficients inside the lifting 9-7 DWT are derived under the DC lossless condition. We confirm that the minimum word length derived by our analysis is shorter than that determined by a conventional approach. We also confirm that the DWT under the condition does not have the checker board for a DC signal.
This article is organized as follows. Chapter 2 defines a rounding operation and a rounding error, describes their basic properties in algebraic approach, and derives 'addition' formula and 'multiplication' formula of the rounding (modulo) operation. Application of these formulas to scaling of a signal value is introduced in chapter 3. Chapter 4 introduces the DC lossless DWT. Its usefulness is also described. Derivation process of conditions on word length of signals and coefficients is described in chapter 5. The new condition derived from the basic properties in chapter 2 is summarized in chapter 6. Other related condition derived from a conventional approach is also summarized. Theoretical results are verified and the minimum word length of the DC lossless DWT is clarified in chapter 7. This article is concluded in chapter 8.

2. Rounding operation and its basic formulas

This chapter introduces basic properties of the rounding operation focusing on 'quotient', rather than 'remainder', in modulo operation. So far, 'remainder' had been attracted numerous mathematicians' attention and various basic properties were found such as the Chinese remainder theorem in the commutative algebra (commutative ring theory). On the contrary, 'quotient' plays an important role as a 'practical' value in finite word length implementation in modern computer systems. This chapter introduces an algebraic approach of expressing 'quotient' as a practical value, and 'remainder' as a rounding error, so that it can be applied to analyzing exact behavior of rounding errors in a complex calculation procedure.

2.1 Definition of rounding operation and rounding error

In a digital calculation system, all the values of both of signals and coefficients are calculated and stored as a binary digit with finite word length. In this article, we treat a case such that a value x is expressed with a fixed point binary expression as

$$x = \sum_{p=-F}^{I-1} b_p 2^p, \quad b_p \in \{0,1\}, \quad I \geq 1, \quad F \geq 0, \quad I \in \mathbf{Z}, \quad F \in \mathbf{Z} \tag{1}$$

where b_p, $p \in \{-F, \cdots, I-1\}$, is a set of binary digit for a value x. It has I bit integer part including one sign bit and F bit fraction part. Hereinafter, F is referred to as word length of a value x. This F bit value x has a range expressed as

$$x \quad \in [-2^{I-1}, 2^{I-1} - 2^{-F}] \quad \in [-2^{I-1}, 2^{I-1}) . \tag{2}$$

For example, in case of I=1 and F=2, the maximum value is x=0.75 for $[b_0\ b_{-1}\ b_{-2}]$=[0 1 1], and the minimum value is x=-1.00 for $[b_0\ b_{-1}\ b_{-2}]$=[1 0 0].
When an F bit signal value is multiplied with a coefficient value, in a convolution of a filtering process in DWT for example, a resulting signal value has longer word length than its original value. Therefore it is rounded to F bit again. So far there are various types of rounding operations [16]. In this article, we deal with the rounding operation defined by

$$R_0[x] = \lfloor x + 2^{-1} \rfloor \quad or \quad R_0[x] = x' - (x' \bmod 1) \quad for \quad x' = x + 2^{-1} \tag{3}$$

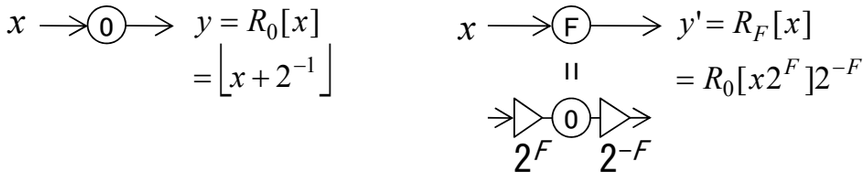
as an example. This rounding operation generates a rounding error. We denote it as

$$\Delta_0[x] = x - R_0[x] \quad or \quad \Delta_0[x] = \{ (x + 2^{-1}) \bmod 1 \} - 2^{-1} . \tag{4}$$

Expanding these expressions to an F bit case, we can define the rounding operation and the rounding error as

$$\begin{cases} R_F[x] = R_0[x2^F]2^{-F} \\ \Delta_F[x] = \Delta_0[x2^F]2^{-F} \end{cases} \tag{5}$$

for an F bit word length implementation case.
Fig.1 illustrates rounding operations expressed by these equations. The term $R_F[x]$ is a quotient, and $\Delta_F[x]$ (= x - $R_F[x]$) is related to a remainder. The former is an actual value treated in a digital system under a finite word length implementation, and the latter is a rounding error. We are now trying to develop an algebraic expression approach to exactly trace a practical value and a rounding error in a convolution processing inside a DWT.



(a) integer (b) F bit fraction

Fig. 1. Definition of the rounding operation and the rounding error. (a) An integer implementation case. (b) An F bit word length implementation case.

2.2 Basic properties of the rounding operation

Since a convolution includes additions and multiplications, we should know behavior of an addition of two values x and y. Resulting value is $R_F[x+y]$ and its rounding error is $\Delta_F[x+y]$. A multiplication result $R_F[xy]$ and its error $\Delta_F[xy]$ should be also investigated.

First of all, let's derive basic properties of the rounding operation starting with an obvious property;

$$y \in \mathbf{Z} \quad \rightarrow \quad R_0[x+y] = R_0[x] + y \quad for \quad x \in \mathbf{R}. \tag{6}$$

It represents that only a real number x can be rounded if y is an integer to calculate a rounded value of $x+y$. In this case, its rounding error becomes

$$y \in \mathbf{Z} \quad \rightarrow \quad \Delta_0[x+y] = \Delta_0[x] \qquad for \quad x \in \mathbf{R}. \tag{7}$$

It suggests that an integer y can be ignored when only the rounding error is considered in an analysis. There is another obvious property;

$$R_0[x] = 0 \quad \leftrightarrow \quad x \in [-2^{-1}, 2^{-1}). \tag{8}$$

Since the range of a rounding error is

$$\Delta_0[x] \in [-2^{-1}, 2^{-1}), \tag{9}$$

we can add two more identities;

$$\begin{cases} R_0\left[\Delta_0[x]\right] = 0, \\ \Delta_0\left[\Delta_0[x]\right] = \Delta_0[x]. \end{cases} \tag{10}$$

The equations above for F=0 can be straightforwardly extended to an $F \neq 0$ case as follows.

$$y2^F \in \mathbf{Z} \quad \rightarrow \quad \begin{cases} R_F[x+y] = R_F[x] + y \\ \Delta_F[x+y] = \Delta_F[x] \end{cases} \quad for \quad x \in \mathbf{R} \tag{11}$$

$$R_F[x] = 0 \quad \leftrightarrow \quad x \in [-2^{-1-F}, 2^{-1-F}) \tag{12}$$

$$\Delta_F[x] \in [-2^{-1-F}, 2^{-1-F}) \tag{13}$$

$$\begin{cases} R_F\left[\Delta_F[x]\right] = 0 \\ \Delta_F\left[\Delta_F[x]\right] = \Delta_F[x] \end{cases} \tag{14}$$

In addition, Eq.(12) can be extended to a more general case with an integer n as

$$R_F[x] = n2^{-F} \quad \leftrightarrow \quad x2^F \in \left[-2^{-1} + n, \quad 2^{-1} + n\right) \quad for \quad n \in \mathbf{Z}. \tag{15}$$

2.3 Basic formulas of the rounding operation

Utilizing the basic properties in Eqs.(11)-(14), we can derive an addition formula and a multiplication formula of a practical value (quotient) of the rounding operation as follows.

Addition formula

$$R_F[x+y] = R_F[x] + R_F\left[y + \Delta_F[x]\right] \quad for \quad x, y \in \mathbf{R} \tag{16}$$

Proof:

$$\begin{aligned} & R_F\left[x+y\right] \\ &= R_F\left[R_F[x] + \Delta_F[x] + y\right] && \leftarrow (4) \\ &= R_F[x] + R_F\left[\Delta_F[x] + y\right] && \leftarrow (11) \end{aligned}$$

Q.E.D.

Multiplication formula

$$R_F[xy] = R_F\left[xR_F[y]\right] + R_F\left[x\Delta_F[y] + \Delta_F\left[xR_F[y]\right]\right] \quad for \quad x, y \in \mathbf{R} \tag{17}$$

Proof:

$$\begin{aligned} & R_F[xy] \\ &= R_F\left[x\Delta_F[y] + xR_F[y]\right] && \leftarrow (4) \\ &= R_F\left[x\Delta_F[y] + \Delta_F\left[xR_F[y]\right] + R_F\left[xR_F[y]\right]\right] && \leftarrow (4) \\ &= R_F\left[x\Delta_F[y] + \Delta_F\left[xR_F[y]\right]\right] + R_F\left[xR_F[y]\right] && \leftarrow (11) \end{aligned}$$

Q.E.D.

Formulas for a rounding error (remainder) can be also derived as

$$\begin{cases} \Delta_F[x+y] = \Delta_F\left[\Delta_F[x] + \Delta_F[y]\right] \\ \Delta_F[xy] = \Delta_F\left[\Delta_F[x]R_F[y] + R_F[x]\Delta_F[y] + \Delta_F[x]\Delta_F[y]\right] \end{cases} \tag{18}$$

for real numbers x and y. These formulas have following variations;
Addition formula

$$y2^F \in \mathbf{Z} \quad \rightarrow \quad \begin{cases} R_F[x+y] = & R_F[x] + y \\ R_F[x+y] = & -\Delta_F[x] + x + y \\ \Delta_F[x+y] = & \Delta_F[x] \end{cases} \tag{19}$$

Multiplication formula

$$y2^F \in \mathbf{Z} \quad \rightarrow \quad \begin{cases} R_F[xy] = & R_F\left[\Delta_F[x]y\right] + R_F[x]y \\ R_F[xy] = & -\Delta_F\left[\Delta_F[x]y\right] + xy \\ \Delta_F[xy] = & \Delta_F\left[\Delta_F[x]y\right] \end{cases} \tag{20}$$

Especially when two kinds of word lengths are mixed in a signal processing, the following variation of the multiplication formula is conveniently applied to analyzing behavior of signals and errors in a pair of encoder and decoder [17].

$$R_{F_2}\left[xR_{F_1}[y]\right] = R_{F_2}[xy] + R_{F_2}\left[-x\Delta_{F_1}[y] + \Delta_{F_2}[xy]\right] \tag{21}$$

Proof:

$$\begin{aligned}
&R_{F_2}\left[xR_{F_1}[y]\right] \\
&= R_0\left[-x\Delta_{F_1}[y]2^{F_2} + xy2^{F_2}\right]2^{-F_2} \\
&= R_0\left[-x\Delta_{F_1}[y]2^{F_2} + \Delta_0[xy2^{F_2}] + R_0[xy2^{F_2}]\right]2^{-F_2} \\
&= R_0\left[-x\Delta_{F_1}[y]2^{F_2} + \Delta_0[xy2^{F_2}]\right]2^{-F_2} + R_0[xy2^{F_2}]\,2^{-F_2} \\
&= R_{F_2}\left[-x\Delta_{F_1}[y] + \Delta_{F_2}[xy]\right] + R_{F_2}[xy]
\end{aligned}$$

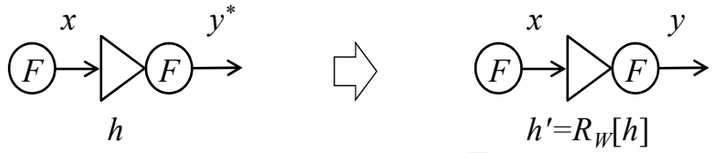
Q.E.D.

3. Application of the formulas to basic signal processing

This chapter applies the formulas to some basic signal processing cases.

3.1 Mapping invariant condition

Fig.2 illustrates a scaling of a signal value x with a coefficient value h. As illustrated in Fig.2(a), this processing maps an input value x to an output value y^* with an ideal (infinite word length) coefficient value h. Note that x has F bit word length. Output value of the multiplication is also rounded to F bit (y^* has F bit word length). In implementation, as illustrated in Fig.2(b), a coefficient value h is also rounded to W bit word length (h' has W bit word length). We aim at finding the minimum word length W of a coefficient h' such that the mapping is invariant ($y - y^* = 0$).



$h \in$ real number, $W \to \infty$ [bit] | $h' \in$ rational number, $W \to$ **min.** [bit]

(a) assumption (b) implementation

Fig. 2. Scaling of a signal value x with a coefficient value h. (a) This processing maps x to y^* with h under a given F. (b) A mapped y should be equal to y^* even though h is rounded to h'.

In case of Fig.2(b), an input value x is multiplied by a rounded value $R_W[h]$ (=h') of a given coefficient h. The result $R_W[h]x$ is rounded to $R_F[R_W[h]x]$ (=y). When it is the same as $R_F[hx]$ (=y^*), the mapping of x is invariant. It means that effect of rounding of h is nullified. This mapping invariant case is expressed as

$$E_m = 0 \quad for \quad \begin{cases} E_m = R_F\left[R_W[h]x\right] - R_F[hx] \\ R_W[h] = h - \Delta_W[h] \end{cases}. \tag{22}$$

From the basic properties, the mapping invariant condition is derived as

$$\Delta_W[h]x - \Delta_F[hx] \in \left(-2^{-1-F}, 2^{-1-F}\right] \ . \tag{23}$$

Proof:

$$\begin{aligned} & R_F\left[R_W[h]x\right] - R_F[hx] \\ &= R_F\left[hx - \Delta_W[h]x\right] - R_F[hx] \\ &= R_F\left[\Delta_F[hx] - \Delta_W[h]x\right] + R_F[hx] - R_F[hx] \\ &= R_F\left[\Delta_F[hx] - \Delta_W[h]x\right] \\ &= 0 \end{aligned}$$

$$\therefore \quad \Delta_F[hx] - \Delta_W[h]x \in \left[-2^{-1-F}, 2^{-1-F}\right)$$

Q.E.D.

The Eq.(23) also means

$$\begin{cases} \Delta_W[h] \in \left(\dfrac{\Delta_F[hx] - 2^{-1}}{x}, \quad \dfrac{\Delta_F[hx] + 2^{-1}}{x}\right], & x > 0 \\ \Delta_W[h] \in \left[\dfrac{\Delta_F[hx] + 2^{-1}}{x}, \quad \dfrac{\Delta_F[hx] - 2^{-1}}{x}\right), & x < 0 \end{cases} \tag{24}$$

which gives tolerance to the rounding error of a coefficient [8]. This is the mapping invariant condition on word length W of a coefficient under a given word length F of signals. It represents exact (not approximated) behavior of rounding errors.
Unlike the condition above, a sufficient condition can be derived by substituting the upper bound of errors and signals;

$$\left|\Delta_W[h]\right| < 2^{-1-W}, \quad \left|\Delta_F[hx]\right| < 2^{-1-F}, \quad |x| < 2^{I-1}, \tag{25}$$

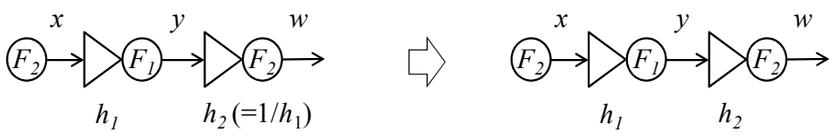
to Eq.(23). It results in the condition described as

$$W > F + I - 1 \ . \tag{26}$$

This condition is too strict and requires too long word length to guarantee the mapping invariance. In both cases of Eq.(24) and Eq.(26), the mapping invariant condition determines the minimum of word length W of a coefficient under a given word length F of signals.

3.2 Lossless condition on a scaling pair

Fig.3 illustrates a pair of two multipliers. In Fig.3(a), an input signal x has F_2 bit word length. It is scaled with a coefficient h_1, and its output value y is rounded to F_1 bit. It is re-scaled with a coefficient h_2 (=1/h_1), and its final output value w is rounded to F_2 bit. This scheme is embedded in a forward transform and a backward transform of DWT for example. It is required to regain w exactly the same as x, under a given word length set of F_1 and F_2. Note that rounding errors due to finite word length expression of coefficients h_1 and h_2 can be ignored as far as the mapping invariant condition is satisfied.



$x, y, w \in$ real number
$F_1 \rightarrow \infty$ [bit]

$x, y, w \in$ rational number
$F_1 \rightarrow$ **min.** [bit]

(a) assumption (b) implementation

Fig. 3. Scaling pair has two coefficients h_1 and h_2 (=1/h_1). (a) Output w is exactly the same as its original x. (b) This lossless property is guaranteed under a condition on F_1 and F_2.

We apply the formulas and the properties to derive the condition on F_1 and F_2. The lossless case in Fig.3(b) is described as

$$E_p = 0 \quad for \quad \begin{cases} E_p = R_{F_2}\left[h_2 R_{F_1}[h_1 x]\right] - x \\ h_1 h_2 = 1 \end{cases}. \tag{27}$$

From the basic properties, the lossless condition on a scaling pair is derived as

$$h_2 \Delta_{F_1}[h_1 x] \in \left(-2^{-1-F_2},\ 2^{-1-F_2}\right]. \tag{28}$$

Proof:

$$\begin{aligned} & R_{F_2}\left[h_2 R_{F_1}[h_1 x]\right] - x \\ &= R_0\left[-h_2 \Delta_{F_1}[h_1 x] 2^{F_2} + h_2 h_1 x 2^{F_2}\right] 2^{-F_2} - x \\ &= R_0\left[-h_2 \Delta_{F_1}[h_1 x] 2^{F_2}\right] 2^{-F_2} + x - x \\ &= R_{F_2}\left[-h_2 \Delta_{F_1}[h_1 x]\right] \\ &= 0 \end{aligned}$$

$$\therefore \quad -h_2 \Delta_{F_1}[h_1 x] \in \left[-2^{-1-F_2},\ 2^{-1-F_2}\right)$$

Q.E.D.

This condition determines the word length F_1 and F_2 of signals for an input value x. It represents exact condition such that total accumulated rounding error is nullified by the rounding just after the final multiplier with h_2. As a result, the original value x is recovered as the final output without any loss.

Unlike the exact condition above, a sufficient condition can be derived by analyzing the upper bound as follows.

$$\left| h_2 \Delta_{F_1}[h_1 x] \right| < \left| h_2 \right| 2^{-1-F_1} < 2^{-1-F_2} \tag{29}$$

As a results, when the sufficient condition given by

$$F_1 > F_2 + \log_2 \left| h_2 \right| \tag{30}$$

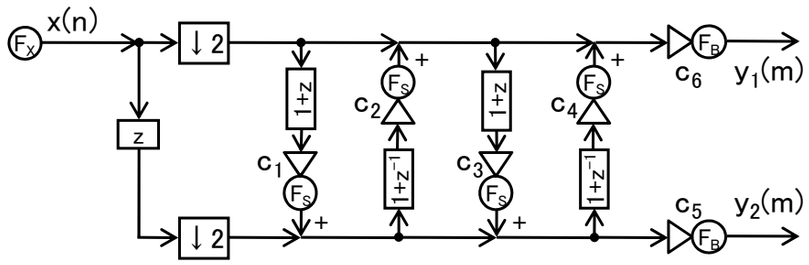
holds, the scaling pair becomes lossless. However this condition is too strict and requires too long word length of signals.

4. Application of the formulas to DWT

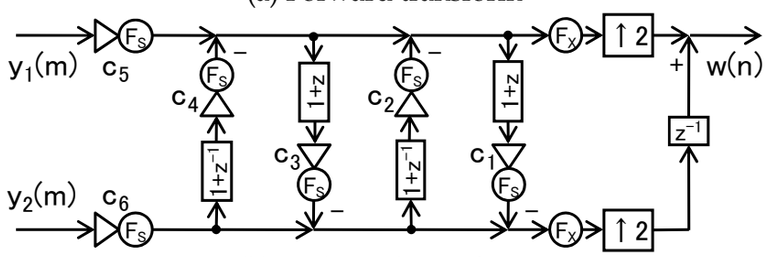
This chapter introduces the DC lossless DWT [8, 18]. Definition and its usefulness are also described. The algebraic approach based on the formulas is applied to derive conditions on word length of signals and coefficients. Derivation process is described in chapter 5.

4.1 DWT and its word length of signals and coefficients

Fig.4 illustrates the irreversible 9-7 DWT of the JPEG 2000 standard [1]. The forward transform in Fig.4(a) decomposes an input signal $x(n)$, $n \in \mathrm{N}$, $\mathrm{N}=\{n \mid 1,2, \cdots, L\}$ into band signals $y_1(m)$ and $y_2(m)$, $m \in \mathrm{M}$, $\mathrm{M}=\{m \mid 1,2, \cdots, L/2\}$. The backward transform in Fig.4(b) reconstructs the signal $w(n)$ from the band signals. In the figure, z^{-1} and ↓2 indicate the delay and the down sampler respectively.



(a) Forward transform



(b) Backward transform

Fig. 4. The irreversible 9-7 DWT of the JPEG 2000 standard.

The multiplier coefficients c_i, $i \in \mathrm{I}$, $\mathrm{I}=\{i \mid 1,2, \cdots, 6\}$ are designed under the word length long enough to be treated as real numbers. When the DWT is implemented, coefficient values are rounded to the length as short as possible to minimize total hardware complexity. Similarly, signal values are also rounded. In the figure, fraction part of each signal is shortened to F_S, F_B or F_X [bit] by a rounding operation illustrated as a circle.

Denoting the integer part as I_S [bit], total word length W_S [bit] of a signal s is defined as

$$W_S = I_S + F_S + 1 \tag{31}$$

including 1 [bit] for the sign part. Similarly, total word length W_C [bit] of a coefficient c is defined as

$$W_C = I_C + F_C + 1\,. \tag{32}$$

In Fig.4, fraction part of the input signal $x(n)$ is given as F_X [bit]. Inside the DWT, fraction part of the signals are rounded to F_S [bit] just after each of all the multiplications with c_i, $i \in \mathrm{I}$. Output signals from the forward and backward transforms are rounded to F_B [bit] and F_X [bit] respectively. Note that we do not truncate integer part of signals and that of coefficients. We are determining F_S and F_C such that the DC lossless property is satisfied.

4.2 Definition of DC lossless property and its necessity

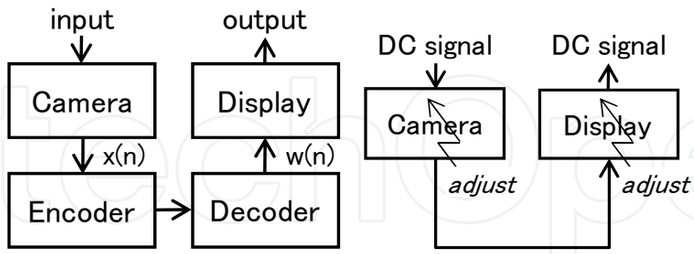
In this article, we define the DC lossless as the conjunction of the following two propositions:

$$\forall n \in \mathrm{N}, \forall m \in \mathrm{M} \left(\; x(n) = d \;\; \rightarrow \;\; y_1(m) = d \;\; \wedge \;\; y_2(m) = 0 \; \right) \tag{33}$$

$$\forall n \in \mathrm{N}, \forall m \in \mathrm{M} \left(\; y_1(m) = d \;\; \wedge \;\; y_2(m) = 0 \;\; \rightarrow \;\; w(n) = d \; \right) \tag{34}$$

for a given constant value d with F_X [bit] fraction part. When the proposition in Eq.(33) holds, the DWT has no DC leakage for the DC input signal with value d. Similarly, when the proposition in Eq.(34) is true, the reconstructed signal $w(n)$ contains no checker board artifact for the DC input signal. In the following chapters, we investigate the minimum fraction part of signals F_S [bit] which guarantees the DC lossless for given F_X and F_B [bit]. We also investigate the minimum fraction part F_{C_i} [bit] of a coefficient c_i, $i \in \mathrm{I}$ with flexibility of trading off the signal error and the coefficient error.

Fig.5(a) illustrates an example of a video system. It contains an encoder and a decoder which are composed of a forward DWT and a backward DWT. In white balancing, a camera and a display are calibrated with a constant valued input signal (DC signal) [11,19]. Therefore, it is useful for this calibration if the forward DWT and its backward do not generate any error. In this case, the camera and the display can be calibrated ignoring existence of the encoder and the decoder as illustrated in Fig.5(b). Namely, the DC lossless condition provides a low complexity DWT useful for the white balancing.



(a) video system (b) calibration

Fig. 5. The DC lossless property is useful for white balancing in a video system.

In addition, the DC lossless condition is a necessary condition for the regularity which controls smoothness of basis functions and coding performance of a transform. A DWT under the regularity does not generate the checker board artifact or the DC leakage. Harada et. al. analyzed a condition for the regularity of a two channel quadrature mirror filter bank (QMF) [9]. They confirmed that a QMF under the condition has reduced checker board artifact for an input step signal. It is expanded to a multirate system under short word length expression [20]. The regularity was structurally guaranteed for a biorthogonal linear

phase filter bank [21,22] and the DCT [10] respectively. However, since these previous methods are based on factorization of a transfer function including $(1+z^{-1})$ or $(1-z^{-1})$ in the lattice structure, these are not directly applicable to the lifting structure of the 9-7 DWT in Fig.4.
In this article, we derive the DC lossless condition theoretically in chapter 5, and determine the minimum word length of signals and that of coefficients under the condition in following chapters.

5. Derivation of condition for the DC lossless DWT

This chapter describes derivation process of the DC lossless condition.

5.1 New model for error analysis

Fig.6(a) illustrates a multiplier in the DWT circuit. An input value s has F_S [bit] fraction part and multiplied by a coefficient c'. The coefficient is originally designed as a real number c. It is rounded to a rational number c' in implementation. It produces the coefficient error:

$$\Delta c = c - c' . \tag{35}$$

Just after the multiplication, the signal is rounded to s' with F_S [bit] fraction part as

$$s' = R_{F_S}[c's] \quad = c's + e' \tag{36}$$

where e' is the signal error. From Eq.(35) and (36), the final output becomes

$$s' = cs - \Delta cs + e' . \tag{37}$$

where cs is the ideal output. This conventional model, illustrated in Fig.6(b), describes the coefficient error Δc as multiplicative to the signal s [13-15], and the signal error e' as additive [7,12]. In addition, these errors are treated independently as mutually uncorrelated noises.
Unlike these existing approaches, as illustrated in Fig.6(c), we describe the coefficient error e'' as

$$\begin{cases} s' = R_{F_S}[cs] + e'', \\ e'' = R_{F_S}\left[\Delta_{F_S}[cs] - \Delta_{F_C}[c]s\right]. \end{cases} \tag{38}$$

From Eq.(15), we utilize the fact that e'' is observed as a 'particle';

$$e'' \cdot 2^{F_S} = p \tag{39}$$

where p is an integer. Given the tolerable maximum to an integer p, word length of the coefficient c can be controlled independently of other coefficients in other sections inside the DWT. Furthermore, denoting the signal error as e' similarly to Eq.(36), the output value is described as

$$\begin{cases} s' = cs + e \\ e = e' + e'' \end{cases} \tag{40}$$

where

$$\begin{cases} e' = -\Delta_{F_S}[cs] \\ e'' = R_{F_S}\left[\Delta_{F_S}[cs] - \Delta_{F_C}[c]s\right] \end{cases} \tag{41}$$

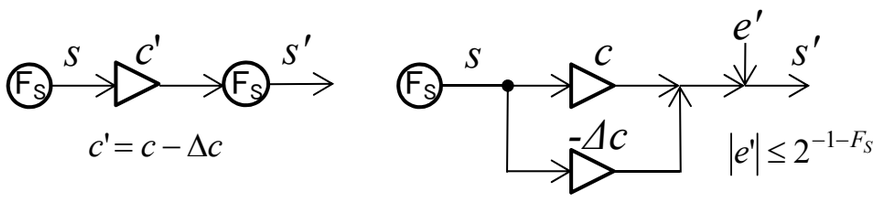
as illustrated in Fig.6(d). In this new model, both of the coefficient error e'' and the signal error e' are unified to the error e. Utilizing Eqs.(13) and (15), its absolute value is limited to

$$|e| \le (p + 2^{-1})2^{-F_S} . \tag{42}$$

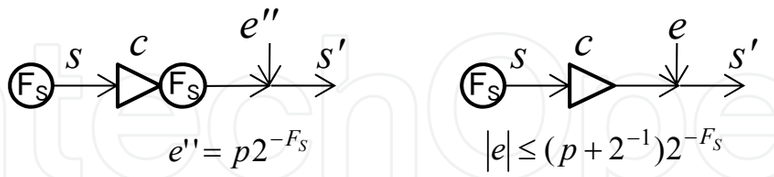
Note that the parameter p to control word length of a coefficient c is included in this equation. It is equivalent to

$$2^{-F_C + F_S + I_S - 1} \le p \tag{43}$$

where its proof is given in appendix.
Benefitting from this inequality, it becomes possible to consider mutual effect of the coefficient error and the signal error.



(a) multiplier (b) conventional model



(c) new model I (d) new model II

Fig. 6. A multiplier in the DWT and its models for error analysis.

Inside the forward DWT, the error e is propagated and added up with other errors from other multipliers. When its maximum absolute value is less than 2^{-1-F_B}, the total error is nullified by the rounding at the final output of the forward DWT. In this article, we utilize this nullification of errors at output of the DWT to derive a condition on word length such that the DC lossless defined by Eq.(33) and (34) is satisfied.

5.2 DC equivalent circuit

When the input signal is restricted to a DC signal, $x(n)$ can be described as a scalar x independent of n. The delay z^{-1} can be treated as 1 and $(1+z^{-1})$ can be replaced by 2. Therefore, instead of the circuits in Fig.4, we can use their equivalent circuits for a DC input signal in Fig.7 to derive the condition.

In Fig.7(a), a scalar x with F_X [bit] fraction part is multiplied by the rational numbers c_i, $i \in \mathrm{I}$ and rounded to F_S [bit]. Finally, the signals are rounded to F_B [bit] at its output to produce two scalars $[y_1\ y_2]$. The unified errors inside the circuit are described as

$$e_i = e_i' + e_i'' \ , \ i \in \mathrm{I} \tag{44}$$

where

$$\begin{cases} e_i' = -\Delta_{F_S}[c_i s_i] \\ e_i'' = R_{F_S}\left[\Delta_{F_S}[c_i s_i] - \Delta_{F_C}[c_i] s_i\right] \end{cases}$$

$$\begin{bmatrix} s_1 & s_3 & s_5 \\ s_2 & s_4 & s_6 \end{bmatrix} = \begin{bmatrix} 2x & 2(x+s'_2) & s_4 \\ 2(x+s'_1) & 2(s_2+s'_3) & s_3+s'_4 \end{bmatrix}$$

$$s'_i = c_i s_i + e_i = R_{F_S}[c_i s_i] \ .$$

Similarly, for the backward transform in Fig.7(b), errors are described as

$$f_i = f_i' + f_i'' \ , \ i \in \mathrm{I} \tag{45}$$

where

$$\begin{cases} f_i' = -\Delta_{F_S}[c_i t_i] \\ f_i'' = R_{F_S}\left[\Delta_{F_S}[c_i t_i] - \Delta_{F_C}[c_i] t_i\right] \end{cases}$$

$$\begin{bmatrix} t_1 & t_3 & t_5 \\ t_2 & t_4 & t_6 \end{bmatrix} = \begin{bmatrix} 2(t_3 - t'_2) & 2(t'_5 - t'_4) & y_1 \\ 2(t_4 - t'_3) & 2t'_6 & y_2 \end{bmatrix}$$

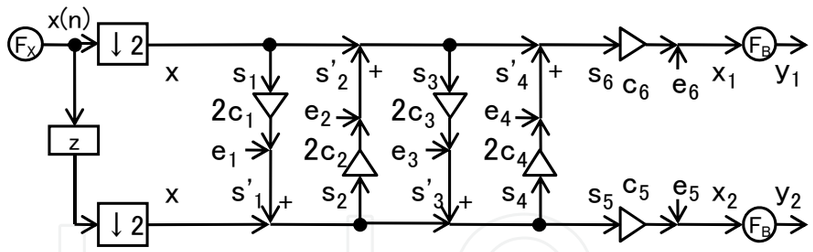
$$t'_i = c_i t_i + f_i = R_{F_S}[c_i t_i] \ .$$

Similarly to Eq.(42), these errors are described with the parameters p_i and q_i to control word length of coefficients as

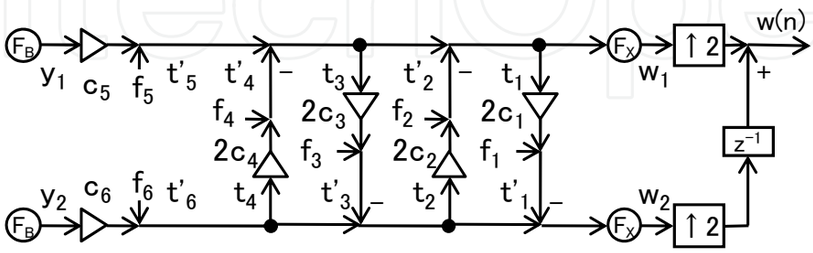
$$\left| e_i \right| \le (p_i + 2^{-1}) 2^{-F_S} \ , i \in \mathrm{I} \tag{46}$$

$$\left| f_i \right| \le (q_i + 2^{-1}) 2^{-F_S} \ , i \in \mathrm{I} \tag{47}$$

for a given word length F_s [bit] of signals.



(a) Forward transform



(b) Backward transform

Fig. 7. Equivalent circuits of the DWT for a DC input signal.

5.3 Nullification of accumulated errors

In Fig.7(a), the unified errors in Eq.(46) are propagated and accumulated inside the circuit. When the accumulated errors are nullified by the rounding at output of the forward transform, Eq.(33) is satisfied. In the figure, $\mathbf{Y}_{12}=[y_1\ y_2]^T$ is described as

$$\begin{aligned}\mathbf{Y}_{12} = R_{F_B}\Big[\ & \mathbf{I}_U e_6 + \mathbf{I}_L e_5 + \mathbf{K}\Big(\ \mathbf{I}_U e_4 + \mathbf{H}_4\Big(\ \mathbf{I}_L e_3 \\ & + \mathbf{H}_3\Big(\ \mathbf{I}_U e_2 + \mathbf{H}_2\big(\mathbf{I}_L e_1 + \mathbf{H}_1 \mathbf{I}_{UL} x\big)\Big)\Big)\Big)\Big]\end{aligned} \tag{48}$$

where

$$\mathbf{I}_U = [1 \quad 0]^T,\ \mathbf{I}_L = [0 \quad 1]^T,\ \mathbf{I}_{UL} = \mathbf{I}_U + \mathbf{I}_L$$

$$\mathbf{H}_{i\in\{1,3\}} = \begin{bmatrix} 1 & 0 \\ 2c_i & 1 \end{bmatrix},\ \mathbf{H}_{j\in\{2,4\}} = \begin{bmatrix} 1 & 2c_j \\ 0 & 1 \end{bmatrix},\ \mathbf{K} = \begin{bmatrix} c_6 & 0 \\ 0 & c_5 \end{bmatrix}.$$

It is described with the unified error matrices $\mathbf{E}_1$ and $\mathbf{E}_2$ as

$$\mathbf{Y}_{12} = R_{F_B}\left[(\mathbf{H}_{e1}\mathbf{E}_1 + \mathbf{H}_{e2}\mathbf{E}_2) + \mathbf{K}\mathbf{H}_{4321}x\right] \tag{49}$$

where

$$\begin{cases} \mathbf{H}_{e1} = \left[\mathbf{I}_U \quad \mathbf{K}\mathbf{I}_U \quad \mathbf{K}\mathbf{H}_{43}\mathbf{I}_U\right] \\ \mathbf{H}_{e2} = \left[\mathbf{I}_L \quad \mathbf{K}\mathbf{H}_4\mathbf{I}_L \quad \mathbf{K}\mathbf{H}_{432}\mathbf{I}_L\right], \\ \mathbf{H}_{43\cdots} = \mathbf{H}_4\mathbf{H}_3\cdots \end{cases}$$

and

$$\begin{cases} \mathbf{E}_1 = [e_6 \quad e_4 \quad e_2]^T \\ \mathbf{E}_2 = [e_5 \quad e_3 \quad e_1]^T. \end{cases}$$

Similarly, output values $\mathbf{W}_{12}=[w_1\ w_2]^T$ from the backward transform in Fig.7(b) are

$$\mathbf{W}_{12} = R_{F_X}\left[\ (\mathbf{H}_{e3}\mathbf{E}_3 + \mathbf{H}_{e4}\mathbf{E}_4) + (\mathbf{K}\mathbf{H}_{4321})^{-1}\mathbf{Y}_{12}\ \right] \tag{50}$$

where

$$\begin{cases} \mathbf{H}_{e3} = -[\mathbf{H}_1^{-1}\mathbf{I}_U \quad \mathbf{H}_{123}^{-1}\mathbf{I}_U \quad \mathbf{H}_{1234}^{-1}\mathbf{I}_U] \\ \mathbf{H}_{e4} = -[\mathbf{I}_L \quad \mathbf{H}_{12}^{-1}\mathbf{I}_L \quad \mathbf{H}_{1234}^{-1}\mathbf{I}_L] \\ \mathbf{H}_{12\cdots}^{-1} = \mathbf{H}_1^{-1}\mathbf{H}_2^{-1}\cdots \end{cases}$$

and

$$\begin{cases} \mathbf{E}_3 = [f_2 \quad f_4 \quad f_5]^T \\ \mathbf{E}_4 = [f_1 \quad f_3 \quad f_6]^T. \end{cases}$$

When the DWT is DC lossless, output values of the transforms are

$$\begin{bmatrix} \hat{\mathbf{Y}}_{12} \\ \hat{\mathbf{W}}_{12} \end{bmatrix} = \begin{bmatrix} \mathbf{K}\mathbf{H}_{4321}x \\ (\mathbf{K}\mathbf{H}_{4321})^{-1}\hat{\mathbf{X}}_{12} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_U \\ \mathbf{I}_{UL} \end{bmatrix} x. \tag{51}$$

Using this equation, the accumulated errors are defined as

$$\begin{bmatrix} \mathbf{E}_{y12} \\ \mathbf{E}_{w12} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_{12} \\ \mathbf{W}_{12} \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{Y}}_{12} \\ \hat{\mathbf{W}}_{12} \end{bmatrix}. \tag{52}$$

Substituting Eqs.(49), (50), (51) and using the property in Eq.(6), we have

$$\begin{bmatrix} \mathbf{E}_{y12} \\ \mathbf{E}_{w12} \end{bmatrix} = \begin{bmatrix} R_{F_B}\left[(\mathbf{H}_{e1}\mathbf{E}_1 + \mathbf{H}_{e2}\mathbf{E}_2)\right] \\ R_{F_X}\left[(\mathbf{H}_{e3}\mathbf{E}_3 + \mathbf{H}_{e4}\mathbf{E}_4)\right] \end{bmatrix}. \tag{53}$$

Applying Eq.(12), it becomes clear that when the conditions;

$$\begin{cases} \left|\mathbf{H}_{e1}\mathbf{E}_1 + \mathbf{H}_{e2}\mathbf{E}_2\right| \le \mathbf{I}_{UL}2^{-1-F_B} \\ \left|\mathbf{H}_{e3}\mathbf{E}_3 + \mathbf{H}_{e4}\mathbf{E}_4\right| \le \mathbf{I}_{UL}2^{-1-F_X} \end{cases} \tag{54}$$

are satisfied, the accumulated errors are nullified by the rounding operations at the final output of each of the forward transform and the backward transform.

6. Derived conditions on word length for the DC lossless DWT

This chapter summarizes the new condition derived from the basic properties in chapter 2, and other related condition derived from a conventional approach.

6.1 Critical condition on word length

Finally, we derive the condition on word length of coefficients and signals such that Eq.(54) is satisfied. Since the unified errors in $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ and $\mathbf{E}_4$ have the maximum in Eqs.(46) and (47) described with the parameters p_i and q_i, the DC lossless condition is also described with the parameters by substituting

$$\begin{cases} \mathbf{E}_1 = \left(\begin{bmatrix} p_6 & p_4 & p_2 \end{bmatrix}^T + \mathbf{I}_3 \cdot 2^{-1} \right) 2^{-F_S} \\ \mathbf{E}_2 = \left(\begin{bmatrix} p_5 & p_3 & p_1 \end{bmatrix}^T + \mathbf{I}_3 \cdot 2^{-1} \right) 2^{-F_S} \\ \mathbf{E}_3 = \left(\begin{bmatrix} q_2 & q_4 & q_5 \end{bmatrix}^T + \mathbf{I}_3 \cdot 2^{-1} \right) 2^{-F_S} \\ \mathbf{E}_4 = \left(\begin{bmatrix} q_1 & q_3 & q_6 \end{bmatrix}^T + \mathbf{I}_3 \cdot 2^{-1} \right) 2^{-F_S} \end{cases} \tag{55}$$

for

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$$

into Eq.(54). This is the condition we derived based on the new model described in section 5.1. We investigate the fraction part F_{Ci} [bit] of a coefficient c_i , $i \in \mathrm{I}$ as the minimum word length under the condition for a DC value x at the word length F_s [bit] of signals.

6.2 Sufficient condition on word length

As an example, in case of all the parameter in Eq.(55) are given as $p_i = q_i = p$ and $F_{Ci} = F_C$ for $\forall\, i \in \mathrm{I}$, the condition in Eq.(54) becomes

$$\begin{cases} \left(\left\| \mathbf{H}_{e1} \right\|_{L^1} + \left\| \mathbf{H}_{e2} \right\|_{L^1} \right) \cdot 2^{-F_S} (p + 2^{-1}) \;\le\; \mathbf{I}_{UL} 2^{-1-F_B} \\ \left(\left\| \mathbf{H}_{e3} \right\|_{L^1} + \left\| \mathbf{H}_{e4} \right\|_{L^1} \right) \cdot 2^{-F_S} (p + 2^{-1}) \;\le\; \mathbf{I}_{UL} 2^{-1-F_X} \end{cases} \tag{56}$$

where $\left\| \mathbf{H} \right\|_{L^1}$ denotes a column vector whose component is a sum of absolute value of all components in each row. Substituting coefficients of the 9-7 DWT [1] into Eq.(56), we have

$$\begin{cases} p \le 2^{-1+F_S-G_E} - 2^{-1} \\ G_E = 2.66 \quad [\text{bit}] \end{cases} \tag{57}$$

for $F_X = F_B = 0$. As a result, the DC lossless condition on the word length is given as

$$-\log_2 (2^{-\Delta W_C} + 2^{-\Delta W_S}) \ge G_E \tag{58}$$

where

$$\begin{bmatrix} \Delta W_C & \Delta W_S \end{bmatrix} = \begin{bmatrix} F_C - I_S & F_S \end{bmatrix}$$

and G_E is the lower bound. This means a sufficient condition for the DC lossless. Since it is too strict, the word length under this condition is redundant. Unlike this sufficient condition, our critical condition given as Eq.(54) under Eq.(55) determines the word length minimum and necessary for the DC lossless.

7. Simulation results

This chapter verifies theoretically derived conditions, and clarifies the minimum word length of the DC lossless DWT.

7.1 Word length under the sufficient condition

Utilizing the sufficient condition in section 6.2, we calculated the optimized word length under the cost function defined as $J=2^{-1}(F_C+F_S)$. The cost J is minimized for three examples. Ex.1 trades the word length between F_C and F_S, namely $F_C + F_S$ = constant. Ex.2 and Ex.3 are $F_C = F_S$ and $W_C = W_S$ respectively. Results are summarized in table 1. Table 2 summarizes word length of signals and coefficients for an 8 bit system with W_x=8 (I_x=7 and F_x=0). Ex.1 requires (F_S, F_C)=(4, 12) [bit] for signals and coefficients respectively. Ex.2 and Ex.3 require $F_S = F_C$ =11 [bit] and $W_S = W_C$ =14 [bit] respectively. The condition in Eq.(58) is plotted as a solid line in Fig.8. According to the sufficient condition, it is impossible to be DC lossless for

$$\begin{cases} F_C & \le & G_E + I_S \\ F_S & \le & G_E \end{cases} \tag{59}$$

and it is also confirmed by the figure. It guarantees the DC lossless, however the condition is too strict. Therefore the word length is redundant and there is room for further reduction.

	Ex.1 F_C+F_S= *const.*	Ex.2 $F_C = F_S$	Ex.3 $W_C = W_S$
F_C F_S	G_E+1+I_S G_E+1	$G_E+I_S^*$	$G_E+I_C^*+I_S-I_C$ $G_E+I_C^*$
W_C W_S	$G_E+2+I_S+I_C$ G_E+2+I_S	$G_E+I_S^*+1+I_C$ $G_E+I_S^*+1+I_S$	$G_E+I_C^*+I_S+1$
J	$G_E+1+I_S/2$	$G_E+I_S^*$	$G_E+I_C^*+(I_S-I_C)/2$

$I_S^*=\log_2(2^{I_S}+1)$, $I_C^*=\log_2(2^{I_C}+1)$

Table 1. Theoretically derived word length under the sufficient condition for DC lossless. F_S and F_C denote fraction part of signals and coefficients. I_S and I_C denote integer part of signals and coefficients. $W_S=I_S+F_S+1$, $W_C=I_C+F_C+1$, J is a cost function.

	Ex.1 F_C+F_S= *const.*	Ex.2 $F_C = F_S$	Ex.3 $W_C = W_S$
F_C F_S	11.66 3.66	10.67	11.25 4.25
W_C W_S	13.66 12.66	12.67 19.67	13.25
J	7.66	10.67	7.75

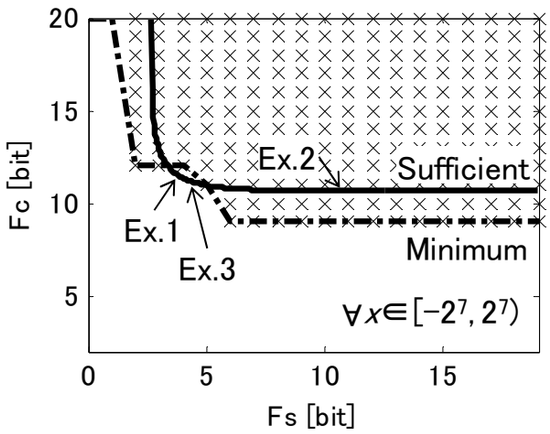
Table 2. Word length calculated with equations in table 1 for W_x=8 [bit].

7.2 Word length under the critical condition

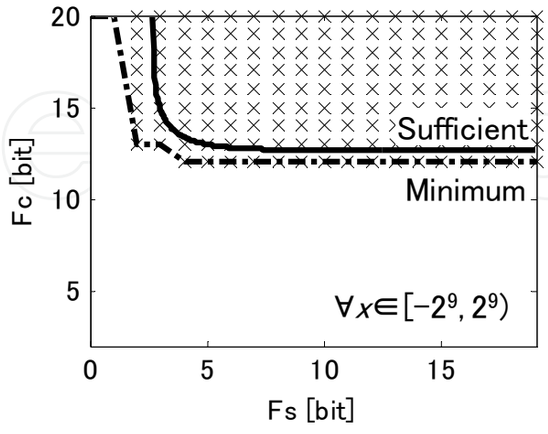
In Fig.8, a cross "×" indicates a pair (F_S, F_C) which satisfies the critical condition in section 6.1 for any 8 bit integer x with W_X=I_X=8 and F_X=0. Here in after, we denote an input DC value to a video system as

$$x_{in} = x - 2^{I_X - 1} \tag{60}$$

where x is an input value to the DC equivalent circuit in Fig.7. The minimum of F_C for each F_S is indicated as a broken line. It is clear that the word length derived by the critical condition is shorter than that determined by the sufficient condition. For example in Fig.8(a), the fraction part F_S (= F_C) is reduced from 11 [bit] to 9 [bit] for Ex.2. The word length is not shortened for Ex.1 and Ex.3. In case of Fig.8(b), F_S (= F_C) is reduced from 13 [bit] to 12 [bit] for Ex.2. (F_S, F_C) is reduced from (14, 4) to (13, 3) or (12, 4) for Ex.1. W_S (= W_C) is reduced from 16 [bit] to 15 [bit] for Ex.3. It is confirmed that the word length is shortened due to the analysis in this article.



(a) $\forall x_{in} \in [0, 2^8), \quad W_X = 8\,[\text{bit}]$

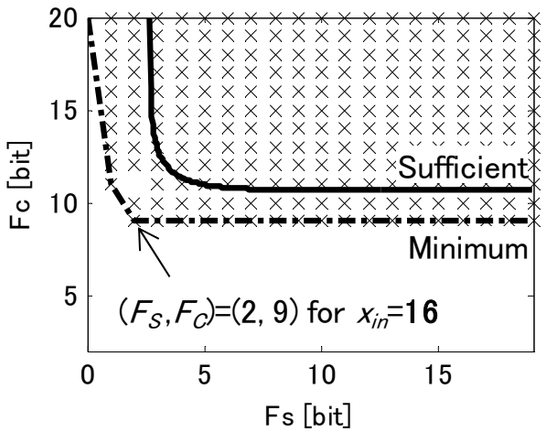


(b) $\forall x_{in} \in [0, 2^{10}), \quad W_X = 10\,[\text{bit}]$

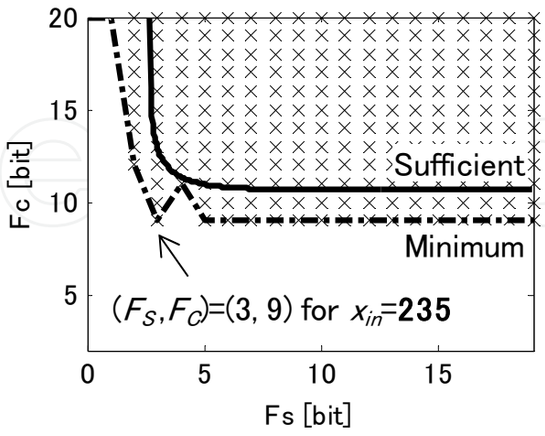
Fig. 8. Word length under the two conditions. "×" indicates (F_S, F_C) such that the DWT becomes DC lossless.

7.3 Word length for a specific value

Fig.9(a) and Fig.9(b) illustrate the word length under the conditions for the black value "16" and the white value "235" respectively. These specific values are utilized in white balancing of an 8-bit video system [19]. For example, (F_S, F_C) is reduced from (4, 12) in Fig.8(a) to (2, 9) in Fig.9(a) for Ex.1. Table 3 summarizes the minimum word length for these specific input DC values [23]. It is observed that the word length can be reduced by limiting input DC signals to a specific value. Fig.10 indicates the minimum word length F_C of coefficients for an input value x at a given word length F_S of signals. This is an example at (F_S, W_X)=(3, 8). The sufficient condition gives the same word length for any of input values. Unlike this conventional statistical analysis, our analysis gives the minimum word length shorter than that determined by the sufficient condition for each of input DC values.



(a) black value



(b) white value

Fig. 9. Word length under the two conditions for a specific value used in white balancing.

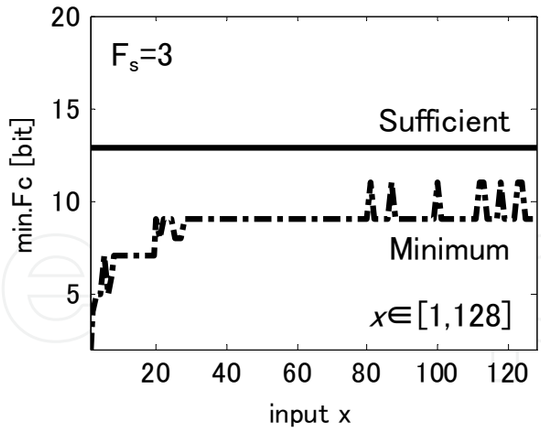


Fig. 10. The minimum word length of coefficients for each of input DC values at (F_S, W_X)=(3, 8). According to the sufficient condition, the word length is too long.

forward transform and backward transform					
input DC values		signals		coefficients	
		integer I_S [bit]	fraction F_S [bit]	integer I_C [bit]	fraction F_C [bit]
W_X= 8 bit	sufficient	8	4	2	12
	any $x_{in} \in [0,2^8)$		2		12
	x_{in}= 16 (black)		2		9
	x_{in}=235 (white)		3		9
W_X=10 bit	sufficient	10	4	2	14
	any $x_{in} \in [0,2^{10})$		2		13
	x_{in}= 64 (black)		0		8
	x_{in}=940 (white)		0		12

Table 3. The minimum word length for a specific value for white balancing of a video system.

7.4 Optimum word length assignment

Since we described tolerance for the unified errors as parameters p_i and q_i in Eq.(46) and (47), it becomes possible to simultaneously control both of word length of signals and that of coefficients. Table 4 summarizes these parameters for an input value 16 and the word length of signals at F_S=2 [bit] as an example. It indicates that [p_1 p_2 ⋯ p_6] in Eq.(46) are [0 0 0 0 0 1] for the word length of coefficients at F_C=9 [bit]. In this case, all the coefficients c_i in the

forward transform have the same length. It is worth paying attention to the fact that the parameter p_1 is the same for F_C=9, 8 and 7 [bit] for example. It means that word length of the coefficient c_1 can be reduced from 9 to 7 [bit] without any influence to the errors. Therefore, word length [F_{C1} F_{C2} … F_{C6}] of coefficients [c_1 c_2 … c_6] can be reduced from [9 9 9 9 9 9] to [7 9 7 4 6 4] according to the table.

Table 5 summarizes results of this optimum word length assignment for the forward transform. Comparing to table 3, it is observed that word length of coefficients is reduced from 9.00 [bit] to 6.17 [bit] on average for an input value x_{in}=16. Table 6 summarizes results for the backward transform. In this case, the word length is furthermore shortened. It is observed that c_6 and c_4 can be omitted since y_2 is equal to zero under the DC lossless. Fig.11 illustrates image signals reconstructed by the DWT which does not satisfy the DC lossless condition. It demonstrates the checker board artifact for reference. It is confirmed that total word length is furthermore shortened utilizing the tolerance parameters p_i and q_i introduced in this article.

F_C	p_1	p_2	p_3	p_4	p_5	p_6	y_1-x	y_2
9	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**
8	**0**	-3	**0**	**0**	**0**	0	-1	-1
7	**0**	-3	**0**	**0**	**0**	0	-1	-1
6	7	12	-8	**0**	**0**	2	6	6
5	7	-18	10	**0**	1	0	-7	-5
4	-21	-18	9	**0**	-2	**1**	-10	-12
3	35	107	7	25	9	-28	63	70

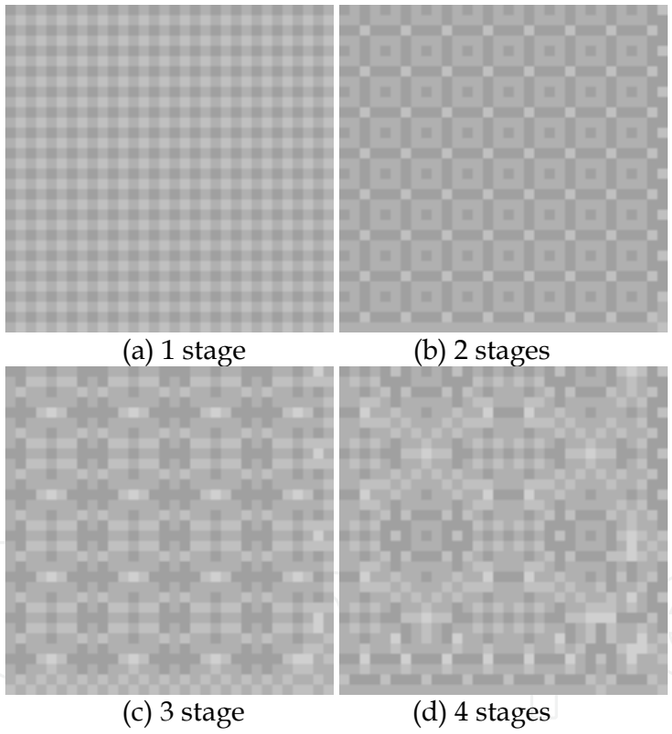
Table 4. Tolerance parameters in Eq.(46) and (47) for x_{in}=16 and F_S=2 as an example.

forward transform									
input values		signals	coefficients						
		F_S	F_{C1}	F_{C2}	F_{C3}	F_{C4}	F_{C5}	F_{C6}	*ave.*
W_X=8	x_{in}= 16 (B)	2	7	9	7	4	6	4	**6.17**
	x_{in}=235 (W)	3	9	7	9	9	1	4	**6.50**
W_X=10	x_{in}= 64 (B)	0	7	9	7	1	1	4	**4.83**
	x_{in}=940 (W)	0	7	11	-1	1	1	4	**3.83**

Table 5. The minimum word length of coefficients in the forward transform for a specific values x_{in} for a given word length of signals F_S.

backward transform									
input values		signals	coefficients						
		F_S	F_{C1}	F_{C2}	F_{C3}	F_{C4}	F_{C5}	F_{C6}	*ave.*
W_X=8	x_{in}= 16 (B)	2	9	9	7	0	8	0	**5.50**
	x_{in}=235 (W)	3	9	9	7	0	8	0	**5.50**
W_X=10	x_{in}= 64 (B)	0	7	9	7	0	8	0	**5.17**
	x_{in}=940 (W)	0	7	12	8	0	8	0	**5.83**

Table 6. The minimum word length of coefficients in the backward transform for a specific values x_{in} for a given word length of signals F_S.



(a) 1 stage (b) 2 stages

(c) 3 stage (d) 4 stages

Fig. 11. Example of reconstructed images for 128^2 pixel DC input image with x=10. Intensity is multiplied by 16.

8. Conclusions

Introducing a new model which unifies the coefficient error and the signal error, and utilizing the nullification of the accumulated errors, this article theoretically derived a condition on word length of signals and coefficients such that the 9-7 DWT of JPEG 2000 becomes lossless for a DC input signal. It was confirmed that the minimum word length

derived by the newly introduced 'critical' condition was shorter than that determined by a conventionally well known 'sufficient' condition. It was also confirmed that the DWT under the condition does not have the checker board for a DC signal. Analysis in this article contributes to build a low complexity DC lossless DWT.

9. Appendix

Proof of Eq.(43)
Eq.(39) with Eq.(41) means

$$\left|\Delta_{F_S}[cs]-\Delta_{F_C}[c]s\right| \le (p+2^{-1})2^{-F_S} \tag{A.1}$$

according to Eqs.(13) and (15). Applying the triangle inequality to the left hand side, we have

$$\left|\Delta_{F_S}[cs]-\Delta_{F_C}[c]s\right| \le \left|\Delta_{F_S}[cs]\right| + \left|\Delta_{F_C}[c]\right|\left|s\right| . \tag{A.2}$$

According to Eq.(13), each terms in the right hand side are described as

$$\begin{cases} \left|\Delta_{F_S}[cs]\right| \le 2^{-1-F_S} \\ \left|\Delta_{F_C}[c]\right|\cdot\left|s\right| \le 2^{-1-F_C}2^{I_S} \end{cases} \tag{A.3}$$

Therefore (A.2) and (A.3) under (A.1) means

$$2^{-1-F_S} + 2^{-1-F_C}2^{I_S} < (p+2^{-1})2^{-F_S}$$

and finally we have Eq.(43) as

$$2^{-F_C+F_S+I_S-1} \le p .$$

Q.E.D.

10. References

[1] ISO/IEC FCD15444-1, "JPEG2000 image coding system," March 2000.
[2] Descampe, et.al., "A flexible hardware JPEG 2000 decoder for digital cinema," IEEE Trans. circuits, systems on video technology, vol.16, issue 11, pp.1397-1410, Nov. 2006
[3] Bing-Fei Wu, Chung-Fu Lin, "Memory-efficient architecture for JPEG 2000 coprocessor with large tile image, IEEE Trans. circuits and systems II, vol.53, issue 4, pp.304-308, April 2006.
[4] H. Kiya, M. Yae, M. Iwahashi, "Linear phase two channel filter bank allowing perfect reconstruction", IEEE Proc. international symposium on circuits and systems (ISCAS), no.2, pp.951-954, May 1992.
[5] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," Technical Report 1994:7, industrial mathematics initiative, department of mathematics, university of South Carolina, 1994.
[6] M. L. Bruelers, A. W. M. van den Enden, "New Networks for Perfect Inversion and Perfect Reconstruction," IEEE Journal of selected areas in communications, vol.10, no.1, pp.130-137, Jan.1992.

[7] M. Reza, Lian Zhu, "Analysis of error in the fixed-point implementation of two-dimensional discrete wavelet transforms," IEEE Trans. circuits and systems, fundamental theory and applications, vol.52, issue 3, pp.641-655, March 2005.
[8] Kiya , M. Iwahashi, O. Watanabe, "A new class of lifting wavelet transform for guaranteeing losslessness of specific signals," IEEE international conference, acoustics, speech, and signal processing (ICASSP), pp.3273-3276, March 2008.
[9] Y. Harada, S. Muramatsu, H. Kiya, "Two channel QMF bank without checker board effect and its lattice structure," IEICE Trans. on fundamentals, vol.J80-A, no.11, pp.1857-1867, Nov. 1997.
[10] Wei Dai, T. D. Tran, , "Regularity-constrained pre- and post- filtering for block DCT-based systems," IEEE Trans. signal processing, vol.51, Issue 10, pp.2568- 2581, Oct. 2003.
[11] Hirakawa, T. W. Parks, "Chromatic adaptation and white balancing problem," IEEE Proc. international conference, image processing (ICIP), vol. III, pp.984-987, Nov. 2005.
[12] Grangetto, et.al., "Optimization and implementation of the integer wavelet transform for image coding," IEEE Trans. Image Processing, vol.11, Issue 6, pp. 596-604, June 2002.
[13] Xiao, et.al., "Coefficient sensitivity and structure optimization of multidimensional state-space digital filters," IEEE Trans. circuits, systems I, vol.45, issue 9, pp.993-998, 1998.
[14] S. Yamaki, M. Abe, M. Kawamata, "A closed form solution to L2-sensitivity minimization of second-order state-space digital filters subject to L2-scaling constraints," IEICT Trans. fundamentals of electronics, communications and computer sciences, vol.E91A, no.7, pp.1697-1705, July 2008.
[15] Y. Tonomura, S. Chokchaitam, M. Iwahashi, "Minimum hardware implementation of multipliers of the lifting wavelet transform," IEEE Proc. international conference, image processing (ICIP), pp.2499-2502, Oct. 2004.
[16] IEEE Standard 754-1985, IEEE standard for binary floating-point arithmetic.
[17] M. Iwahashi, H. Kiya, "Finite word length error analysis based on basic formula of rounding operation", the international symposium on intelligent signal processing and communication systems (ISPACS), no.86, pp.49-52, Dec. 2008.
[18] M. Iwahashi, H. Kiya, "Word length condition for DC Lossless DWT," Asia pacific signal and information processing association (APSIPA) annual summit and conference, no.TA-P2-6, pp.469–472, Oct. 2009.
[19] The society of motion picture and television engineers, "Standard for television, 1920x1080 image sample structure, digital representation and digital timing reference sequences for multiple picture rates", SMPTE 274 M-2005, Feb.2005.
[20] H. Iwai, M. Iwahashi, K. Kiya, "Methods for avoiding the checkerboard distortion caused by finite word length error in multirate system", IEICE Trans. fundamentals, vol. E93-A, no.3, pp.631-635, March, 2010.
[21] S. Oraintara, T.D. Tran, T.Q. Nquen, "A class of regular biorthogonal linear-phase filterbanks: Theory, structure, and application in image coding," IEEE Trans. signal processing, vol.51, no.12, pp.3220-3235, Dec.2003.
[22] Y. Tanaka, M. Ikehara, "First order linear phase filter banks with regularity constrains for efficient image coding," IEICE Trans. fundamentals, vol. J91-A, no.2, pp.192-201, Feb. 2008.
[23] M. Iwahashi, H. Kiya, "A lossless condition of lifting DWT for specific DC values", IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp.1458-1461, March 2010.

Discrete Wavelet Transforms - Algorithms and Applications

Edited by Prof. Hannu Olkkonen

ISBN 978-953-307-482-5
Hard cover, 296 pages
Publisher InTech
Published online 29, August, 2011
Published in print edition August, 2011

The discrete wavelet transform (DWT) algorithms have a firm position in processing of signals in several areas of research and industry. As DWT provides both octave-scale frequency and spatial timing of the analyzed signal, it is constantly used to solve and treat more and more advanced problems. The present book: Discrete Wavelet Transforms: Algorithms and Applications reviews the recent progress in discrete wavelet transform algorithms and applications. The book covers a wide range of methods (e.g. lifting, shift invariance, multi-scale analysis) for constructing DWTs. The book chapters are organized into four major parts. Part I describes the progress in hardware implementations of the DWT algorithms. Applications include multitone modulation for ADSL and equalization techniques, a scalable architecture for FPGA-implementation, lifting based algorithm for VLSI implementation, comparison between DWT and FFT based OFDM and modified SPIHT codec. Part II addresses image processing algorithms such as multiresolution approach for edge detection, low bit rate image compression, low complexity implementation of CQF wavelets and compression of multi-component images. Part III focuses watermaking DWT algorithms. Finally, Part IV describes shift invariant DWTs, DC lossless property, DWT based analysis and estimation of colored noise and an application of the wavelet Galerkin method. The chapters of the present book consist of both tutorial and highly advanced material. Therefore, the book is intended to be a reference text for graduate students and researchers to obtain state-of-the-art knowledge on specific applications.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元

Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen

IntechOpen