When applied recursively, these formulae define the *fast wavelet transform*; the relations (35) and (36) define the forward transform, while (37) defines the inverse transform.

Now, from the fact that $H(0) = G(\pi) = 1$ and $G(0) = H(\pi) = 0$, we see that $H(\omega)$ acts like a low pass filter for the interval $[0, \pi/2]$ and $G(\omega)$ similarly behaves like a band pass filter for the interval $[\pi/2, \pi]$. Equation (8) (respectively (12)) then implies that the major part of the energy of the functions in $V_0$ (respectively $W_0$) is concentrated in the intervals $[0, \pi]$ (respectively $[\pi, 2\pi]$). The basic behavior of the dual functions is the same. In an approximate sense, this means that the wavelet expansion splits the frequency space into dyadic blocks $[2^j\pi, 2^{j+1}\pi]$ with $j \in \mathbf{Z}$ [103, 104].

In signal processing this idea is known as subband filtering, or, more specifically, as quadrature mirror filtering. Quadrature mirror filters were studied before wavelet theory. The decomposition step consists of applying a low-pass ($\widetilde{H}$) and a band-pass ($\widetilde{G}$) filter followed by downsampling ($\downarrow 2$) (i.e. retaining only the even index samples), see Figure 1. The reconstruction consists of upsampling ($\uparrow 2$) (i.e. putting a zero between every two samples) followed by filtering and addition. One can show that the conditions (27) correspond to the exact reconstruction of a subband filtering scheme. More details about this can be found in [115, 132, 133, 134].

An interesting problem now is: given a function $f$, determine, with a certain accuracy and in a computationally favorable way, the coefficients $\lambda_{n,l}$ of a function in the space $V_n$, which are needed to start the fast wavelet transform. A trivial solution could be

$$\lambda_{n,l} = f(l/2^n).$$

Other sampling procedures, such as (quasi-)interpolation and quadrature formulae were proposed in [1, 2, 85, 120, 128, 138].

An implementation of a fast wavelet transform in pseudo code is given in the appendix.

**10. Examples of wavelets.** Now that we have discussed the essentials of wavelet multiresolution analysis, we take a look at some important properties of wavelets.

*Orthogonality.* Orthogonality is convenient to have in many situations, e.g. it directly links the $\mathbf{L}^2$ norm of a function to the norm of its wavelet coefficients by

$$\|f\| = \sqrt{\sum_{j,l} \gamma_{j,l}^2}.$$

In the biorthogonal case these two quantities are only equivalent. Another advantage of orthogonal wavelets is that the fast wavelet transform is a unitary transformation (i.e. its adjoint is its inverse). Consequently, its condition number is equal to 1, which is the optimal case. (Recall that the condition number of a linear transformation $A$ is defined as $\|A\|.\|A^{-1}\|$). This is of importance in numerical calculations. It means that an error present in the initial data will not grow under the transformation, and that stable numerical computations are possible.

If the multiresolution analysis is orthogonal (remember that this includes semiorthogonal wavelets), the projection operators onto the different subspaces yield optimal approximations in the $\mathbf{L}^2$ sense.

*Compact support.* If the scaling function and wavelet are compactly supported, the filters $H$ and $G$ are finite impulse response filters, so that the summations in the fast wavelet transform are finite. This obviously is of use in implementations. If they are not compactly supported, a fast decay is desirable so that the filters can be approximated reasonably by finite impulse response filters.

*Rational coefficients.* For computer implementations it is of use if the filter coefficients $h_k$ and $g_k$ are rationals or, even better, dyadic rationals. Multiplication by a power of two on a computer corresponds to shifting bits, which is a very fast operation.

*Symmetry.* If the scaling function and wavelet are symmetric, then the filters have generalized linear phase. The absence of this property can lead to phase distortion. This is important in signal processing applications.

*Smoothness.* The smoothness of wavelets plays an important role in compression applications. Compression is usually achieved by setting small coefficients $\gamma_{j,l}$ to zero, and thus leaving out a component $\gamma_{j,l}\,\psi_{j,l}(x)$ from the original function. If the original function represents an image and the wavelet is not smooth, the error can easily be detected visually. Note that the smoothness of the primary functions is more important to this aspect than that of the dual. Also, a higher degree of smoothness corresponds to better frequency localization of the filters. Finally, smooth basis functions are desired in numerical analysis applications where derivatives are involved.

*Number of vanishing moments of the dual wavelet.* We saw earlier that this is important in singularity detection and characterization of smoothness spaces. Also, it determines the convergence rate of wavelet approximations of smooth functions. Finally, the number of vanishing moments of the dual wavelet is connected to the smoothness of the wavelet (and vice versa).

*Analytic expressions.* As previously noted, an analytic expression for a scaling function or wavelet does not always exists but in some cases it is available and nice to have. In harmonic analysis, analytic expressions of the Fourier transform are particularly useful.

*Interpolation.* If the scaling function satisfies

$$\varphi(k) = \delta_k \quad \text{for} \quad k \in \mathbf{Z},$$

then it is trivial to find the function of $V_j$ that interpolates data sampled on a grid with spacing $2^{-j}$, since the coefficients are equal to the samples.

As could be expected, it is not possible to construct wavelets that have all these properties and there is a trade-off between them. We now take a look at several compromises.

*Examples of orthogonal wavelets.*

(i) Two simple examples of orthogonal scaling functions are the box function $\chi_{[0,1]}(x)$ and the Shannon sampling function $\operatorname{sinc}(\pi x)$. The orthogonality conditions are easy to verify, either in the time or frequency space. The corresponding wavelet for the box function is the *Haar wavelet*

$$\psi_{Haar}(x) = \chi_{[0,1/2]}(x) - \chi_{[1/2,1]}(x),$$

and the Shannon wavelet is

$$\psi_{Shannon}(x) = \frac{\sin(2\pi x) - \sin(\pi x)}{\pi x}.$$

These two, however, are not very useful in practice, since the first has very low regularity and the second has very slow decay.

(ii) A more interesting example is the *Meyer wavelet* and scaling function [106]. These functions belong to $C^\infty$ and have faster than polynomial decay. Their Fourier transform is compactly supported. The scaling function and wavelet are symmetric around 0 and 1/2, respectively, and the wavelet has an infinite number of vanishing moments.

(iii) The *Battle-Lemarié wavelets* are constructed by orthogonalizing B-spline functions using (20) and have exponential decay [12, 95]. The wavelet with $N$ vanishing moments is a piecewise polynomial of degree $N-1$ that belongs to $C^{N-2}$.

(iv) Probably the most frequently used orthogonal wavelets are the original Daubechies wavelets [47, 49]. They are a family of orthogonal wavelets indexed by $N \in \mathbb{N}$, where $N$ is the number of vanishing wavelet moments. They are supported on an interval of length $2N-1$. A disadvantage is that, except for the Haar wavelet (which has $N=1$), they cannot be symmetric or antisymmetric. Their regularity increases linearly with $N$ and is approximately equal to $0.2075N$ for large $N$. In [137] a different family with regularity asymptotically equal to $0.3N$ was presented. In [50] three variations of the original family, all with orthogonal and compactly supported functions, are constructed:

1. The previous construction does not lead to a unique solution if $N$ and the support length are fixed. One family is constructed by choosing, for each $N$, the solution with closest to linear phase (or closest to symmetry). In fact, the original family corresponds to choosing the extremal phase.

2. Another family has more regularity, at the price of a slightly larger support length $(2N+1)$.

3. In a third family, the scaling function also has vanishing moments ($\mathcal{M}_p = 0$ for $0 < p < N$). This is of use in numerical analysis applications where inner products of arbitrary functions with scaling functions have to be calculated very fast [17]. Their construction was asked by Ronald Coifman and Ingrid Daubechies therefore named them *coiflets*. They are supported on an interval with length $3N-1$.

*Examples of biorthogonal wavelets.*

(i) Biorthogonal wavelets were constructed by Albert Cohen, Ingrid Daubechies and Jean-Christophe Feauveau in [31]. Here $\Delta(\omega)$ is chosen equal to $e^{-i\omega}$, and thus

$$G(\omega) = -e^{-i\omega}\,\overline{\widetilde{H}(\omega+\pi)} \quad \text{and} \quad \widetilde{G}(\omega) = -e^{-i\omega}\,\overline{H(\omega+\pi)}.$$

In one of the families constructed in [31], the scaling functions are the cardinal B-splines and the wavelets too are spline functions. All functions including the dual ones have compact support and linear phase. Moreover, all filter coefficients are dyadic rationals. A disadvantage is that for small filter lengths, the dual functions have very low regularity.

(ii) Semiorthogonal spline wavelets were constructed by Charles Chui and Jianzhong Wang in [23, 24, 25]. The scaling functions are cardinal B-splines of order $m$ and the wavelet functions are splines with support $[0, 2m-1]$. All primary and dual functions still have generalized linear phase and all coefficients used in the fast

TABLE 1
*A quick comparison of wavelet families.*

| wavelet family | compact support | | analytic expression | | symmetry | orthogonality | | compact support $\widehat{\psi}$ |
|---|---|---|---|---|---|---|---|---|
| | primary | dual | primary | dual | | semi | full | |
| a | x | x | o | o | o | x | x | o |
| b | x | x | x | o | x | o | o | o |
| c | x | o | x | x | x | x | o | o |
| d | o | o | o | o | x | x | x | x |
| e | o | o | x | x | x | x | x | o |

a: Daubechies' orthogonal wavelets
b: biorthogonal spline wavelets
c: semiorthogonal spline wavelets
d: Meyer wavelet
e: orthogonal spline wavelets

wavelet transform are rationals. A powerful feature here is that analytic expressions for the wavelet, scaling function, and dual functions are available. A disadvantage is that the dual functions do not have compact support, but have exponential decay instead. The same wavelets, but in a different setting, were also derived by Akram Aldroubi, Murray Eden and Michael Unser in [129, 131]. They also showed that for $N$ going to infinity, the spline wavelets converge to Gabor functions [130].

(iii) Other semiorthogonal wavelets can be found in [89, 109, 110, 113]. A characterization of all semiorthogonal wavelets is given in [1, 2].

The properties of some of the orthogonal, biorthogonal and semiorthogonal wavelet families are summarized in Table 1.

*Examples of interpolating scaling functions.*

(i) The Shannon sampling function

$$\varphi_{Shannon} = \frac{\sin(\pi x)}{\pi x},$$

is an interpolating scaling function. It is band limited, but it has very slow decay.

(ii) An interpolating scaling function, whose translates also generate $V_0$, can be found by letting

$$\widehat{\varphi}_{interpol}(\omega) = \frac{\widehat{\varphi}(\omega)}{\sum_l \varphi(l) e^{-i\omega l}},$$

provided that the denominator does not vanish [1, 2, 129, 138]. Even if $\varphi$ is compactly supported, $\varphi_{interpol}$ is in general not compactly supported. The cardinal spline interpolants of even order are constructed this way [118].

(iii) An interpolating scaling function can also be constructed from a pair of biorthogonal scaling functions as

$$\varphi_{interpol}(x) = \int_{-\infty}^{+\infty} \varphi(y + x)\, \overline{\widetilde{\varphi}(y)}\, dy.$$

The interpolation property immediately follows from the biorthogonality condition. In the case of an orthogonal scaling function this is just its autocorrelation function. The interpolating function and its translates do not generate the same space as $\varphi$ and its translates. This construction, starting from the Daubechies orthogonal or biorthogonal wavelets, yields a family of interpolating functions which had been studied earlier by Gilles Deslauriers and Serge Dubuc in [56, 57]. These functions are smooth and compactly supported. More information can also be found in [61, 117]. A natural choice for the wavelet here is $\psi(x) = \varphi(2x-1)$ and this is a typical example of a wavelet with a non-vanishing integral. The dual scaling function is a Dirac impulse and the dual wavelet is a linear combination of Dirac impulses (and has several vanishing moments). We still have a fast wavelet transform with finite impulse response filters.

(iv) Also wavelets can be interpolating. In [2] wavelets that are both symmetrical and interpolating were constructed.

**11. Wavelets on an interval.** So far we have been discussing wavelet theory on the real line (and its higher dimensional analogs). For many applications, the functions involved are only defined on a compact set, such as an interval or a square, and to apply wavelets then requires some modifications.

**11.1. Simple solutions.** To be specific, let us discuss the case of the unit interval $[0,1]$. Given a function $f$ on $[0,1]$, the most obvious approach is to set $f(x) = 0$ outside $[0,1]$, and then use wavelet theory on the line. However, for a general function $f$ this "padding with 0s" introduces discontinuities at the endpoints 0 and 1; consider for example the simple function $f(x) = 1$, $x \in [0,1]$. Now, as we have said earlier, wavelets are effective for detecting singularities, so artificial ones are likely to introduce significant errors.

Another approach, which is often better, is to consider the function to be periodic with period 1, $f(x+1) = f(x)$. Expressed in another way, we assume that the function is defined on the torus and identify the torus with $[0,1]$. Wavelet theory on the torus parallels that on the line. In fact, note that if $f$ has period 1, then the wavelet coefficients on a given scale satisfy $\langle f, \psi_{j,k} \rangle = \langle f, \psi_{j,k+2^j} \rangle$, $k \in \mathbb{Z}$, $j \geq 0$. This simple observation readily allows us to rewrite wavelet expansions on the line as analogous ones on the torus, with wavelets defined on $[0,1]$. A periodic multiresolution analysis on the interval $[0,1]$ can be constructed by periodizing the basis functions as follows,

$$\varphi^*_{j,l}(x) = \chi_{[0,1]}(x) \sum_m \varphi_{j,l}(x+m) \quad \text{for} \quad 0 \leqslant l < 2^j \quad \text{and} \quad j \geqslant 0. \tag{38}$$

If the support of $\varphi_{j,l}(x)$, is a subset of $[0,1]$, then $\varphi^*_{j,l}(x) = \varphi_{j,l}(x)$. Otherwise $\varphi_{j,l}(x)$ is chopped into pieces of length 1, which are shifted onto $[0,1]$ and added up, yielding $\varphi^*_{j,l}(x)$. Similar definitions hold for $\psi^*_{j,l}$, $\widetilde{\varphi}^*_{j,l}$ and $\widetilde{\psi}^*_{j,l}$. The algorithm in the appendix describes the periodic fast wavelet transform. This "wrap around" procedure is satisfactory in many situations (and certainly takes care of functions like $f(x) = 1$, $x \in [0,1]$). However, unless the behavior of the function $f$ at 0 matches that at 1, the periodic version of $f$ has singularities there. A simple function like $f(x) = x$, $x \in [0,1]$, gives a good illustration of this.

A third method, which works if the basis functions are symmetric, is to use reflection across the edges. This preserves continuity, but introduces discontinuities in the first derivative. This solution is sometimes satisfactory in image processing applications.

**11.2. Meyer's boundary wavelets.** What really is needed, are wavelets intrinsically defined on $[0,1]$. We sketch a construction of orthogonal wavelets on $[0,1]$, recently presented by Yves Meyer [107]. We start from an orthogonal Daubechies scaling function with $2N$ non-zero coefficients:

$$\varphi(x) = 2\sum_{k=0}^{2N-1} h_k\,\varphi(2x-k). \tag{39}$$

It is easy to see that $\operatorname{clos}\{x : \varphi(x) \neq 0\} = [0, 2N-1]$, and, as a consequence,

$$B_{j,k} = \operatorname{clos}\{x : \varphi_{j,k}(x) \neq 0\} = [2^{-j}k, 2^{-j}(k+2N-1)]. \tag{40}$$

This implies that for sufficiently small scales $2^{-j}$, $j \geq j_0$, a function $\varphi_{j,k}$ can only intersect at most one of the endpoints 0 or 1. Let us restate this in a different way. Define the set of indices

$$S_j = \{k : B_{j,k} \cap (0,1) \neq \emptyset\}.$$

We define three subsets of this set containing the indices of the basis functions at the left boundary, in the interior, and at the right boundary:

$$\begin{aligned} S_j^{(1)} &= \{k : 0 \in B_{j,k}^\circ\} \\ S_j^{(2)} &= \{k : B_{j,k}^\circ \subset (0,1)\} \\ S_j^{(3)} &= \{k : 1 \in B_{j,k}^\circ\}. \end{aligned}$$

Here $E^\circ$ denotes the interior of the set $E$. For sufficiently large $j$ the sets $S_j^{(1)}$ and $S_j^{(3)}$ are disjoint and

$$S_j = S_j^{(1)} \cup S_j^{(2)} \cup S_j^{(3)}.$$

It is easy to write down what these sets are more explicitly:

$$\begin{aligned} S_j^{(1)} &= \{k : -2N+2 \leqslant k \leqslant -1\} \\ S_j^{(2)} &= \{k : 0 \leqslant k \leqslant 2^j - 2N + 1\} \\ S_j^{(3)} &= \{k : 2^j - 2N + 2 \leqslant k \leqslant 2^j - 1\}. \end{aligned}$$

Note, in particular, that the sets $S_j^{(1)}$ and $S_j^{(3)}$ contain the indices of $2N-2$ functions, independently of $j$. We now let $V_j^{[0,1]}$ denote the restriction of functions in $V_j$:

$$V_j^{[0,1]} = \{f : f(x) = g(x),\ x \in [0,1], \text{ for some function } g \in V_j\}.$$

Clearly, since the $V_j$ form an increasing sequence of spaces,

$$V_j^{[0,1]} \subset V_{j+1}^{[0,1]},$$

and $V_j^{[0,1]}$, $j \geq j_0$, form a multiresolution analysis of $L^2([0,1])$. It is also obvious that the functions in $\{\varphi(x-l)|_{[0,1]} : l \in S_j\}$ span $V_j^{[0,1]}$. Here $g(x)\,|_{[0,1]}$ denotes the restriction of $g(x)$ to $[0,1]$. Not quite as obvious is the fact that the functions in this collection are linearly independent, and hence form a basis for $V_j^{[0,1]}$. In order to obtain an orthonormal basis, we may argue as follows. As long as the function $\varphi_{j,k}$ lives entirely inside $[0,1]$, restricting it to $[0,1]$ has no effect. In particular, the functions $\varphi_{j,k}$, $k \in S_j^{(2)}$ are still pairwise orthogonal. A key observation now is that for $k \in S_j^{(1)}$, $l \in S_j^{(2)} \cup S_j^{(3)}$,

$$\text{(41)} \qquad \langle \varphi_{j,k}, \varphi_{j,l} \rangle_{[0,1]} = \int_0^1 \varphi_{j,k}(x)\, \varphi_{j,l}(x)\, dx = \int_{-\infty}^{+\infty} \varphi_{j,k}(x)\, \varphi_{j,l}(x)\, dx = 0,$$

and similarly when $k \in S_j^{(3)}$, $l \in S_j^{(2)} \cup S_j^{(1)}$. We see that the three collections $\{\varphi(x-l)|_{[0,1]} : l \in S_j^{(1)}\}$, $\{\varphi(x-l)|_{[0,1]} : l \in S_j^{(2)}\}$, and $\{\varphi(x-l)|_{[0,1]} : l \in S_j^{(3)}\}$ are mutually orthogonal. So, since the functions in $\{\varphi(x-l)|_{[0,1]} : l \in S_j^{(2)}\}$ already form an orthonormal set, there only remains to separately orthogonalize the functions in $\{\varphi(x-l)|_{[0,1]} : l \in S_j^{(1)}\}$ and in $\{\varphi(x-l)|_{[0,1]} : l \in S_j^{(3)}\}$. This is easily accomplished with a Gram-Schmidt procedure.

Now, if we let $W_j^{[0,1]}$ denote the restriction of functions in $W_j$ to $[0,1]$, then we have that

$$\text{(42)} \qquad V_{j+1}^{[0,1]} = V_j^{[0,1]} + W_j^{[0,1]}.$$

So, the basis elements in $V_j^{[0,1]}$ together with the restriction of the wavelets $\psi_{j,k}$ to $[0,1]$ span $V_{j+1}^{[0,1]}$. However, there are $2^j + 2N - 2$ wavelets that intersect $[0,1]$, and, since $\dim V_{j+1}^{[0,1]} - \dim V_j^{[0,1]} = 2^j$ we have too many functions. The restrictions of the wavelets in $W_j$ that live entirely inside $[0,1]$ are still mutually orthogonal and, by an observation similar to (41), they are also orthogonal to $V_j^{[0,1]}$. There are $2N-2$ wavelets whose support intersects an endpoint. However, we only need $N-1$ basis functions at each endpoint. One can now use (30) to write out the dependencies, and construct $N-1$ basis functions at each endpoint. After that we just apply a Gram-Schmidt procedure again, and we have an orthonormal basis for $W_j^{[0,1]}$.

This elegant construction of Yves Meyer has a couple of disadvantages. Among the functions $\varphi_{j,k}$ that intersect $[0,1]$ there are some that are almost zero there. Hence, the set $\{\varphi_{j,k}\}_{k\in S_j}$ is almost linearly dependent, and, as a consequence, the condition number of the matrix, corresponding to the change of basis from $\{\varphi_{j,k}\}_{k\in S_j}$ to the orthonormal one, becomes quite large. Furthermore, we have $\dim V_j^{[0,1]} \neq \dim W_j^{[0,1]}$, which means that there is an inherent imbalance between the spaces $V_j^{[0,1]}$ and $W_j^{[0,1]}$, which is not present in the case of the whole real line.

**11.3. Dyadic boundary wavelets.** As we noted earlier (33) all polynomials of degree $\leqslant N-1$ can be written as linear combinations of the $\varphi_{j,l}$ for $l \in \mathbf{Z}$. Hence, the restriction of such polynomials to $[0,1]$ are in $V_j^{[0,1]}$. Since this fact is directly linked to many of the approximation properties of wavelets, any construction of a multiresolution analysis on $[0,1]$ should preserve this. The construction in [5, 32, 33] uses this as a starting point and is slightly different from the one by Yves Meyer. Let us briefly describe this construction as well. Again we start with an orthogonal Daubechies scaling function $\varphi$ with $2N$ non-zero coefficients, and assume that we have picked the scale fine enough so that the endpoints are independent as before. By (33) and, since the $\{\varphi_{j,k}\}$ is an orthonormal basis for $V_j$, each monomial $x^\alpha$, $\alpha \leqslant N-1$, has the representation $x^\alpha = \sum_k \langle x^\alpha, \varphi_{j,k} \rangle\, \varphi_{j,k}(x)$. The restriction to $[0,1]$ can then be written

$$x^\alpha|_{[0,1]} = \left( \sum_{k=-2N+2}^{0} + \sum_{k=1}^{2^j-2N} + \sum_{k=2^j-2N+1}^{2^j-1} \right) \langle x^\alpha, \varphi_{j,k} \rangle \, \varphi_{j,k}(x)|_{[0,1]}.$$

If we let

$$x^\alpha_{j,L} = 2^{j(\alpha+1/2)} \sum_{k=-2N+2}^{0} \langle x^\alpha, \varphi_{j,k} \rangle \, \varphi_{j,k}(x)|_{[0,1]}$$

and, similarly,

$$x^\alpha_{j,R} = 2^{j(\alpha+1/2)} \sum_{k=2^j-2N+1}^{2^j-1} \langle x^\alpha, \varphi_{j,k} \rangle \, \varphi_{j,k}(x)|_{[0,1]},$$

then

$$2^{j/2}(2^j x)^\alpha|_{[0,1]} = x^\alpha_{j,L} + 2^{j(\alpha+1/2)} \sum_{k=1}^{2^j-2N} \langle x^\alpha, \varphi_{j,k} \rangle \, \varphi_{j,k}(x)|_{[0,1]} + x^\alpha_{j,R}.$$

We let the spaces $\tilde{V}_j$, $j \geq j_0$, that form a multiresolution analysis of $L^2([0,1])$, be the linear span of the functions $\{x^\alpha_{j,L}\}_{\alpha \leqslant N-1}$, $\{x^\alpha_{j,R}\}_{\alpha \leqslant N-1}$, and $\{\varphi_{j,k}|_{[0,1]}\}_{k=1}^{2^j-2N}$:

$$\tilde{V}_j = \overline{\{x^\alpha_{j,L}\}_{\alpha \leqslant N-1}} \cup \overline{\{\varphi_{j,k}\}_{k=1}^{2^j-2N}} \cup \overline{\{x^\alpha_{j,R}\}_{\alpha \leqslant N-1}}$$

Finding an orthonormal basis for $\tilde{V}_j$ is easy; in fact, the collections $\{x^\alpha_{j,L}\}_{\alpha \leqslant N-1}$, $\{\varphi_{j,k}\}_{k=1}^{2^j-2N}$, and $\{x^\alpha_{j,R}\}_{\alpha \leqslant N-1}$ are mutually orthogonal, and all of the functions in these are linearly independent. We thus only have to orthogonalize the functions $x^\alpha_{j,L}$ and $x^\alpha_{j,R}$ to get our orthonormal basis. Note that, by construction, $\dim \tilde{V}_j = 2^j$ and all polynomials of degree $\leq N-1$ are in $\tilde{V}_j$. It is also easy to see that

$$\tilde{V}_j \subset \tilde{V}_{j+1}.$$

To get to the corresponding wavelets we let $\tilde{W}_j$ be the orthogonal complement of $\tilde{V}_j$ in $\tilde{V}_{j+1}$. The wavelets $\psi_{j,k}$ with $1 \leqslant k \leqslant 2^j - 2N$ are all in $\tilde{V}_{j+1}$ and live entirely inside $[0,1]$. The remaining $2N$ functions required for an orthonormal basis of $\tilde{W}_j$, can be found, for example by using (30) again.

This last construction carries over to more general situations. For example, we can also use biorthogonal wavelets and much more general closed sets than [0, 1] [5, 33, 87].

There are also other constructions of wavelets on [0, 1]. In fact, for historical perspective it is interesting to notice that Franklin's original construction [70] was given for [0, 1]. Another interesting one, in the case of semiorthogonal spline wavelets, has been given by Charles Chui and Ewald Quak [19]; we refer to the original paper for details.

**12. Wavelet packets.** A simple, but most powerful extension of wavelets and multiresolution analysis are wavelet packets [37, 38]. In this section it will be useful to switch to the following notation:

$$m_e(\omega) = H^e(\omega)\, G^{1-e}(\omega) \quad \text{for} \quad e = 0, 1.$$

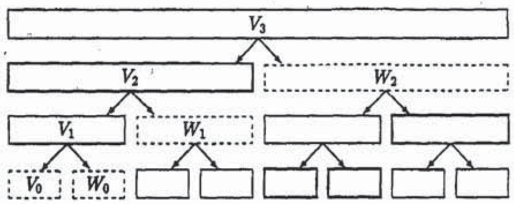


FIG. 4. *Wavelet packets scheme.*

The fundamental observation is the following fact, called the *splitting trick* [22, 30, 106]:
*Suppose that the set of functions* $\{f(x-k) \mid k \in \mathbb{Z}\}$ *is a Riesz basis for its closed linear span* $S$. *Then the functions*

$$f_k^0 = \frac{1}{\sqrt{2}} f^0(x/2-k) \quad \textit{and} \quad f_k^1 = \frac{1}{\sqrt{2}} f^1(x/2-k) \quad \textit{for} \quad k \in \mathbb{Z},$$

*also constitute a Riesz basis for* $S$, *where*

$$\widehat{f^e}(\omega) = m_e(\omega/2)\,\widehat{f}(\omega/2).$$

We see that the classical multiresolution analysis is obtained by splitting $V_j$ with this trick into $V_{j-1}$ and $W_{j-1}$ and then doing the same for $V_{j-1}$ recursively. The wavelet packets are the basis functions that we obtain if we also use the splitting trick on the $W_j$ spaces. So starting from a space $V_j$, we obtain, after applying the splitting trick $L$ times, the basis functions

$$\psi^L_{e_1,\ldots,e_L;j,k}(x) = 2^{(j-L)/2}\psi^L_{e_1,\ldots,e_L}(2^{j-L}x-k),$$

with

$$\widehat{\psi}^L_{e_1,\ldots,e_L}(\omega) = \prod_{i=1}^{L} m_{e_i}(2^{-i}\omega)\,\widehat{\varphi}(2^{-L}\omega).$$

So, after $L$ splittings, we have $2^L$ basis functions and their translates over integer multiples of $2^{L-j}$ as a basis of $V_j$. The connection between the wavelet packets and the wavelet and scaling functions is

$$\varphi = \psi^L_{0,\ldots,0} \quad \text{and} \quad \psi = \psi^L_{1,0,\ldots,0}.$$

However, we do not necessarily have to split each subspace at every stage. In Figure 4 we give a schematical representation of a space and its subspaces after using the splitting on 3 levels. The top rectangle represents the space $V_3$ and each other rectangle corresponds to a certain subspace of $V_3$ generated by wavelet packets. The slanted lines between the rectangles indicate the splitting, the left referring to the filter $m_0$ and the right to $m_1$. The dashed rectangles then correspond to the wavelet multiresolution analysis $V_3 = V_0 \oplus W_0 \oplus W_1 \oplus W_2$. The bold rectangles correspond to a possible wavelet packet splitting and a basis with functions

$$\left\{\psi_0^1(4x-k),\, \psi_{1,1}^2(2x-k),\, \psi_{0,0,1}^3(x-k),\, \psi_{1,0,1}^3(x-k) \mid k \in \mathbf{Z}\right\}.$$

For the dual functions, a similar procedure has to be followed.

In the Fourier domain, the splitting trick corresponds to dividing the frequency interval essentially represented by the original space into two parts. So the wavelet packets allow more flexibility in adapting the basis to the frequency contents of a signal.

It is easy to develop a fast wavelet packet transform. It just involves applying the same low and band pass filters also to the coefficient of functions of $W_j$ again in an iterative manner. This means that, starting from $M$ samples, we construct a full binary tree with $(M \log_2 M)$ entries. The power of this construction lies in the fact that we have much more freedom in deciding which basis functions we will use to represent the given function. We can choose to use the set of $M$ coefficients of the tree to represent the function that is optimal with respect to a certain criterion. This procedure is called ***best basis selection***, and one can design fast algorithms that make use of the tree structure. The particular criterion is determined by the application, and which basis functions that will end up in the basis depends on the data.

For applications in image processing, entropy-based criteria were proposed in [40]. The best basis selection in that case has a numerical complexity of $\mathcal{O}(M)$. Applications in signal processing can be found in [36, 139].

This wavelet packets construction can also be combined with wavelets on an interval and wavelets in higher dimensions [55].

**13. Multidimensional wavelets.** Up till now we have focused on functions of one variable and the one-dimensional situation. However, there are also wavelets in higher dimensions. A simple way to obtain these is to use tensor products. To fix ideas, let us consider the case of the plane. Let

$$\Phi(x,y) = \varphi(x)\,\varphi(y) = \varphi \otimes \varphi(x,y),$$

and define

$$V_0 = \{f : f(x,y) = \sum_{k_1,k_2} \lambda_{k_1,k_2}\, \Phi(x-k_1, y-k_2), \lambda \in l^2(\mathbf{Z}^2)\}.$$

Of course, if $\{\varphi(x-l) \mid l \in \mathbf{Z}\}$ is an orthonormal set, then $\{\Phi(x-k_1, y-k_2)\}$ form an orthonormal basis for $V_0$. By dyadic scaling we obtain a multiresolution analysis of $L^2(\mathbf{R}^2)$. The complement $W_0$ of $V_0$ in $V_1$ is similarly generated by the translates of the three functions

$$\Psi^{(1)} = \varphi \otimes \psi, \quad \Psi^{(2)} = \psi \otimes \varphi, \quad \text{and} \quad \Psi^{(3)} = \psi \otimes \psi. \tag{43}$$

There is another, perhaps even more straightforward, wavelet decomposition in higher dimensions. By carrying out a one-dimensional wavelet decomposition for each variable separately, we obtain

$$f(x,y) = \sum_{i,l} \sum_{j,k} \langle f, \psi_{i,l} \otimes \psi_{j,k} \rangle\, \psi_{i,l} \otimes \psi_{j,k}(x,y). \tag{44}$$

Note that the functions $\psi_{i,l} \otimes \psi_{j,k}$ involve two scales, $2^{-i}$ and $2^{-j}$, and each of these functions are (essentially) supported on a rectangle. The decomposition (44) is therefore called the *rectangular wavelet decomposition* of $f$ while the functions in (43) are the basis functions of the *square wavelet decomposition*. For both decompositions, the corresponding fast wavelet transform consists of applying the one-dimensional fast wavelet transform to the rows and columns of a matrix.

These simple constructions are insufficient in many cases. What we need sometimes are wavelets intrinsically constructed for higher dimensions. One of the interesting problems here is how to split a space into complementary subspaces. In the univariate case we split into two spaces, each with essentially the same "size." If we use the square tensor product basis in $d$ dimensions, we split into $2^d$ subspaces, $2^d - 1$ of which are spanned by wavelets. There are several constructions of nonseparable wavelets that use this kind of splitting. One of the problems here is, given the scaling function, is there an easy way, cf. (19), to find the wavelets? This was studied in [54, 113, 121]. Another idea is to still try to split into just two subspaces. This involves the use of different lattices [99]. In the bivariate case, Ingrid Daubechies and Albert Cohen constructed smooth, compactly supported, biorthogonal wavelets, using ideas from the univariate construction [29].

By now, there is a lot of material about multivariate wavelets. However, we shall leave this topic for now and just mention some other possibilities such as hexagonal lattices, and Clifford valued wavelets [6, 9, 34].

## 14. Applications.

**14.1. Data compression.** One of the most common applications of wavelet theory is data compression. There are two basic kinds of compression schemes: lossless and lossy. In the case of lossless compression one is interested in reconstructing the data exactly, without any loss of information. We consider here lossy compression. This means we are ready to accept an error, as long as the quality after compression is acceptable. With lossy compression schemes we potentially can achieve much higher compression ratios than with lossless compression.

To be specific, let us assume that we are given a digitized image. The compression ratio is defined as the number of bits the initial image takes to store on the computer divided by the number of bits required to store the compressed image. The interest in compression in general has grown as the amount of information we pass around has increased. This is easy to understand when we consider the fact that to store a moderately large image, say a $512 \times 512$ pixels, 24 bit color image, takes about 0.75 MBytes. This is only for still images; in the case of video, the situation becomes even worse. Then, we need this kind of storage for each frame, and we have something like 30 frames per second. There are several reasons other than just the storage requirement for the interest in compression techniques. However, instead of going into this, let us now look at the connection with wavelet theory.

First, let us define, somewhat mathematically, what we mean by an image. Let us for simplicity discuss an $L \times L$ grayscale image with 256 grayscales (i.e. 8 bit). This can be considered to be a piecewise constant function $f$ defined on a square

$$f(x,y) = p_{ij} \in \mathbf{N}, \quad \text{for} \quad i \leqslant x < i+1 \quad \text{and} \quad j \leqslant y < j+1 \quad \text{and} \quad 0 \leqslant i,j < L,$$

where $0 \leqslant p_{ij} \leqslant 255$. Now, one of the standard procedures for lossy compression is through transform coding, see Figure 5. The most common transform used in this context is the "Discrete Cosine Transform", which uses a Fourier transform of the

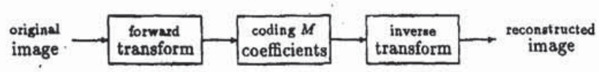


FIG. 5. *Image transform coding.*

image $f$. However, we are more interested in the case when the transform is the fast wavelet transform.

There are in fact several ways to use the wavelet transform for compression purposes [101, 102]. One way is to consider compression to be an approximation problem [58, 59]. More specifically, let us fix an orthogonal wavelet $\psi$. Given an integer $M \geqslant 1$, we try to find the "best" approximation of $f$ by using a representation

$$f_M(x) = \sum_{kl} b_{jk}\,\psi_{jk}(x) \quad \text{with } M \text{ non-zero coefficients } b_{jk}. \tag{45}$$

The basic reason why this potentially might be useful is that each wavelet picks up information about the image $f$ essentially at a given location and at a given scale. Where the image has more interesting features, we can spend more coefficients, and where the image is nice and smooth we can use fewer and still get good quality of approximation. In other words, the wavelet transform allows us to focus on the most relevant parts of $f$. Now, to give this mathematical meaning we need to agree on an error measure. Ideally, for image compression we should use a norm that corresponds as closely as possible to the human eye [58]. However, let us make it simple and discuss the case of $L^2$.

So we are interested in finding an optimal approximation minimizing the error $\|f - f_M\|_{L^2}$. Because of the orthogonality of the wavelets this equals

$$\left(\sum_{jk} |\langle f, \psi_{jk}\rangle - b_{jk}|^2\right)^{1/2}. \tag{46}$$

A moment's thought, reveals that the best way to pick $M$ non-zero coefficients $b_{jk}$, making the error as small as possible, is by simply picking the $M$ coefficients with largest absolute value, and setting $b_{j,k} = \langle f, \psi_{jk}\rangle$ for these numbers. This then yields the optimal approximation $f_M^{opt}$.

Another fundamental question is which images can be approximated well by using the procedure just sketched. Let us take this to mean that the error satisfies

$$\|f - f_M^{opt}\|_{L^2} = \mathcal{O}(M^{-\beta}), \tag{47}$$

for some $\beta > 0$. The larger $\beta$, the faster the error decays as $M$ increases and the fewer coefficients are generally needed to obtain an approximation within a given error. The exponent $\beta$ can be found easily, in fact it can be shown that

$$\left(\sum_{M\geq 1} (M^{\beta}\|f - f_M^{opt}\|_{L^2})^p \frac{1}{M}\right)^{1/p} \approx \left(\sum_{jk} |\langle f, \psi_{jk}\rangle|^p\right)^{1/p} \tag{48}$$

with $1/p = 1/2 + \beta$. The maximal $\beta$ for which (47) is valid can be estimated by finding the smallest $p$ for which the right-hand side of (48) is finite. The expression on the right is one of many equivalent norms on the Besov space $\dot{B}_p^{2\beta,p}$ (Besov spaces are smoothness spaces generalizing the Lipschitz continuous functions). The $\beta$ in the left-hand side of (48) is actually not exactly the same as in (47). However, for practical purposes, the difference is of no consequence.

**14.2. Operator analysis.** As mentioned earlier, interest in wavelets historically grew from the fact that they are effective tools for studying problems in partial differential equations and operator theory. More specifically, they are useful for understanding properties of so-called *Calderón-Zygmund operators*.

Let us first make a general observation about the representation of a linear operator $T$ and wavelets. Suppose that $f$ has the representation

$$f(x) = \sum_{jk} \langle f, \psi_{jk} \rangle \psi_{jk}(x).$$

Then,

$$Tf(x) = \sum_{jk} \langle f, \psi_{jk} \rangle T\psi_{jk}(x),$$

and, using the wavelet representation of the function $T\psi_{jk}(x)$, this equals

$$\sum_{jk} \langle f, \psi_{jk} \rangle \sum_{il} \langle T\psi_{jk}, \psi_{il} \rangle \psi_{il}(x) = \sum_{il} \left( \sum_{jk} \langle T\psi_{jk}, \psi_{il} \rangle \, \langle f, \psi_{jk} \rangle \right) \psi_{il}(x).$$

In other words, the action of the operator $T$ on the function $f$ is directly translated into the action of the infinite matrix $A_T = \{ \langle T\psi_{jk}, \psi_{il} \rangle \}_{il,jk}$ on the sequence $\{ \langle f, \psi_{jk} \rangle \}_{jk}$. This representation of T as the matrix $A_T$ is often referred to as the "standard representation" of $T$ [17]. There is also a "nonstandard representation". For virtually all linear operators there is a function (or, more generally, a distribution) $K$ such that

$$Tf(x) = \int K(x,y) f(y)\, dy.$$

The nonstandard representation of $T$ is now simply the (two-dimensional) wavelet coefficients of the kernel $K$, using the square decomposition $\{ \langle K, \Psi^{(j)}_{k_1,k_2} \rangle \}$ (again, we have more than one wavelet function in two dimensions), while the standard representation corresponds to the rectangular decomposition.

Let us then briefly discuss the connection with Calderón-Zygmund operators. Consider a typical example. Let $H$ be the *Hilbert transform*,

$$Hf(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(s)}{x-s}\, ds.$$

The basic idea now is that the wavelets $\psi_{jk}$ are approximate eigenfunctions for this, as well as for many other related (Calderón-Zygmund) operators. We note that if $\psi_{jk}$ were exact eigenfunctions, then we would have $H\psi_{jk}(x) = \lambda_{jk}\psi_{jk}(x)$, for some number $\lambda_{jk}$ and the standard representation would be a diagonal "matrix":

$$A_H = \{ \langle H\psi_{il}, \psi_{jk} \rangle \} = \{ \lambda_{il} \langle \psi_{il}, \psi_{jk} \rangle \} = \{ \lambda_{il}\, \delta_{il,jk} \}.$$

This is unfortunately not the case. However, it turns out that $A_T$ is in fact an almost diagonal operator, in the appropriate, technical sense, with the off diagonal elements quickly becoming small. To get some idea why this is the case, note that for large $|x|$, we have, at least heuristically,

$$H\psi_{jk}(x) \approx \frac{1}{\pi x}\int \psi_{jk}(y)\,dy.$$

A priori, the decay of the right-hand side would thus be $\mathcal{O}(1/x)$, which of course is far from the rapid decay of a wavelet $\psi_{jk}$ (remember that some wavelets are even zero outside a finite set). Recall, however, that $\psi_{jk}$ has at least one vanishing moment so the decay is in fact much faster than just $\mathcal{O}(1/x)$, and the shape of $H\psi_{jk}(x)$ resembles that of $\psi_{jk}(x)$. By expanding the kernel as a Taylor series,

$$\frac{1}{x-s} = \frac{1}{x}\left(1+\frac{s}{x}+\frac{s^2}{x^2}\cdots\right),$$

we see that the more vanishing moments $\psi$ has, the faster the decay of $H\psi_{j,k}$.

So, for a large class of operators, the matrix representation, either the standard or the nonstandard, has a rather precise structure with many small elements. In this representation, we then expect to be able to compress the operator by simply omitting small elements. In fact, note that this is essentially the same situation, as in the case of image compression, the "image" now being the kernel $K(x, y)$. Hence, if we could do basic operations, such as inversion and multiplication, with compressed matrices, rather than with the discretized versions of $T$, then we may significantly speed up the numerical treatment. This program of using the wavelet representations for the efficient numerical treatment of operators was initiated in [17]. We also refer to [4, 3] for related material and many more details.

In a different direction, because of the close similarities between the scaling function and finite elements, it seems natural to try wavelets where traditionally finite element methods are used, e.g. for solving boundary value problems [84]. There are interesting results showing that this might be fruitful; for example, it has been shown [17, 46, 111, 140] that for many problems the condition number of the $N \times N$ stiffness matrix remains bounded as the dimension $N$ goes to infinity. This is in contrast with the situation for regular finite elements where the condition number in general tends to infinity.

One of the first problems we have to address when discussing boundary problems on domains, is how to take care of the boundary values and the fact that the problem is defined on a finite set rather than on the entire Euclidean plane. This is similar to the problem we discussed with wavelets on an interval, and, indeed, the techniques discussed there can be often used to handle these two problems [5, 8].

Wavelets have also been used in the solution of evolution equations [11, 76, 93, 98]. A typical test problem here is *Burgers' equation*:

$$\frac{\partial u}{\partial t} + u\,\frac{\partial u}{\partial x} = \nu\,\frac{\partial^2 u}{\partial x^2}.$$

The time discretization is obtained here using standard schemes such as Crank-Nicholson or Adams-Moulton. Wavelets are used in the space discretization. Adaptivity can be used both in time and space [11].

One of the nice features of wavelets and finite elements is that they allow us to treat a large class of operators or partial differential equations in a unified way, allowing for example general PDE solvers to be designed. In specific instances, though, it is possible to find particular wavelets, adapted to the operator or problem at hand [10, 44, 45, 88]. In [16], Gregory Beylkin develops fast wavelet-based algorithms for the solution of differential equations.

Note: Applications in statistics such as the smoothing of data were investigated by David Donoho and Iain Johnstone in [62, 63, 64, 65].

**Acknowledgement.** We would like to thank Gilbert Strang and the referees for useful comments and suggestions, and Maurits Malfait and Nathan Hall for helping us with the presentation.

## REFERENCES


[1] A. Aldroubi and M. Unser, *Families of wavelet transforms in connection with Shannon's sampling theory and the Gabor transform*, in [21], pp. 509–528.

[2] ———, *Families of multiresolution and wavelet spaces with optimal properties*, Numer. Funct. Anal. and Optimiz., (1993), pp. 417–446.

[3] B. Alpert, *A class of bases in $L^2$ for the sparse representation of integral operators*, SIAM J. Math. Anal., 24 (1993), pp. 246–262.

[4] B. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin, *Wavelet-like bases for the fast solution of second-kind integral equations*, SIAM J. Sci. Comp., 14 (1993), pp. 159–184.

[5] L. Andersson, N. Hall, B. Jawerth, and G. Peters, *Wavelets on closed subsets of the real line*, in [119], pp. 1–61.

[6] L. Andersson, B. Jawerth, and M. Mitrea, *The Cauchy singular integral operator and Clifford wavelets*, in [13], pp. 525–546.

[7] E. W. Aslaksen and J. R. Klauder, *Unitary representations of the affine group*, J. Math. Phys., 9 (1968), pp. 2267–2275.

[8] P. Auscher, *Wavelets with boundary conditions on the interval*, in [21], pp. 217–236.

[9] ———, *Ondelettes fractales et Applications*, PhD thesis, Université de Paris-Dauphine, 1989.

[10] P. Auscher and P. Tchamitchian, *Conjecture de Kato sur les ouverts de* **R**, Rev. Mat. Iberoamericana, 8 (1992), pp. 149–199.

[11] E. Bacry, S. Mallat, and G. Papanicolaou, *A wavelet based space-time adaptive numerical method for partial differential equations*, RAIRO Mathematical Modelling and Numerical Analysis, 26.

[12] G. Battle, *A block spin construction of ondelettes*, Comm. Math. Phys., 110 (1987), pp. 601–615.

[13] J. Benedetto and M. Frazier, eds., *Wavelets: Mathematics and Applications*, CRC Press, Boca Raton, 1993.

[14] J. J. Benedetto, *Irregular sampling and frames*, in [21], pp. 445–507.

[15] M. A. Berger, *Random affine iterated function systems: Curve generation and wavelets*, SIAM Rev., 31 (1989), pp. 614–627.

[16] G. Beylkin, *On wavelet-based algorithms for solving differential equations*, in [13], pp. 449–466.

[17] G. Beylkin, R. Coifman, and V. Rokhlin, *Fast wavelet transforms and numerical algorithms I*, Comm. Pure Appl. Math., 44 (1991), pp. 141–183.

[18] A. Cavaretta, W. Dahmen, and C. Micchelli, *Subdivision for Computer Aided Geometric Design*, Memoirs Amer. Math. Soc., 93 (1991).

[19] C. Chui and E. Quak, *Wavelets on a bounded interval*, in Numerical Methods of Approximation Theory, D. Braess and L. L. Schumaker, eds., Birkhäuser Verlag, Basel, 1992, pp. 1–24.

[20] C. K. Chui, *An Introduction to Wavelets*, Academic Press, San Diego, 1992.

[21] ———, ed., *Wavelets: A Tutorial in Theory and Applications*, Academic Press, San Diego, 1992.

[22] C. K. Chui and C. Li, *Non-orthogonal wavelet packets*, SIAM J. Math. Anal., 24 (1993), pp. 712–738.

[23] C. K. Chui and J. Z. Wang, *A cardinal spline approach to wavelets*, Proc. Amer. Math. Soc., 113 (1991), pp. 785–793.

[24] ———, *A general framework of compactly supported splines and wavelets*, J. Approx. Theory, 71 (1992), pp. 263–304.

[25] ——, *On compactly supported spline wavelets and a duality principle*, Trans. Amer. Math. Soc., 330 (1992), pp. 903–915.
[26] Z. Ciesielski, *Constructive function theory and spline systems*, Studia Math., 52 (1973), pp. 277–302.
[27] A. Cohen, *Biorthogonal wavelets*, in [21], pp. 123–152.
[28] ——, *Ondelettes, analyses multiresolutions et filtres miroirs en quadrature*, Ann. Inst. H. Poincaré, Anal. Non Linéaire, 7 (1990), pp. 439–459.
[29] A. Cohen and I. Daubechies, *Non-separable bidimensional wavelet bases*. Preprint AT&T Bell Laboratories, New Jersey, 1991.
[30] ——, *On the instability of arbitrary biorthogonal wavelet packets*, SIAM J. Math. Anal., 24 (1993), pp. 1340–1354.
[31] A. Cohen, I. Daubechies, and J. Feauveau, *Bi-orthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 45 (1992), pp. 485–560.
[32] A. Cohen, I. Daubechies, B. Jawerth, and P. Vial, *Multiresolution analysis, wavelets and fast algorithms on an interval*, C. R. Acad. Sci. Paris Sér. I Math., I (1993), pp. 417–421.
[33] A. Cohen, I. Daubechies, and P. Vial, *Multiresolution analysis, wavelets and fast algorithms on an interval*, Appl. Comput. Harmon. Anal., 1 (1993), pp. 54–81.
[34] A. Cohen and J.-M. Schlenker, *Compactly supported bidimensional wavelet bases with hexagonal symmetry*, Constr. Approx., 9 (1993), pp. 209–236.
[35] R. R. Coifman, *A real variable characterization of $H^p$*, Studia Math, 51 (1974).
[36] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, *Signal processing and compression with wave packets*, in Proceedings of the International Conference on Wavelets, Marseille, 1989, Y. Meyer, ed., Masson, Paris.
[37] R. R. Coifman, Y. Meyer, and V. Wickerhauser, *Size properties of wavelet packets*, in [116], pp. 453–470.
[38] ——, *Wavelet analysis and signal processing*, in [116], pp. 453–470.
[39] R. R. Coifman and R. Roohberg, *Representation theorems for holomorphic and harmonic functions in $L^p$*, Astérisque, 77 (1980), pp. 11–66.
[40] R. R. Coifman and M. L. Wickerhauser, *Entropy based algorithms for best basis selection*, IEEE Trans. Inform. Theory, 38 (1992), pp. 713–718.
[41] D. Colella and C. Heil, *Characterizations of scaling functions: I. Continuous solutions*, SIAM J. Math. Anal., To appear.
[42] ——, *The characterization of continuous four-coefficient scaling functions and wavelets*, IEEE Trans. Inform. Theory, 38 (1992).
[43] J. M. Combes, A. Grossmann, and P. Tchamitchian, eds., *Wavelets: Time-Frequency Methods and Phase Space*, Inverse problems and theoretical imaging, Springer-Verlag, 1989.
[44] S. Dahlke and I. Weinreich, *Wavelet bases adapted to pseudo-differential operators*, Appl. Comput. Harmon. Anal., To appear.
[45] ——, *Wavelet-Galerkin-methods: An adapted biorthogonal wavelet basis*, Constr. Approx., 9 (1993), pp. 237–262.
[46] W. Dahmen and A. Kunoth, *Multilevel preconditioning*, Numer. Math., 63 (1992), pp. 315–344.
[47] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 41 (1988), pp. 909–996.
[48] ——, *The wavelet transform, time-frequency localization and signal analysis*, IEEE Trans. Inform. Theory, 36 (1990), pp. 961–1005.
[49] ——, *Ten Lectures on Wavelets*, no. 61 in CBMS-NSF Series in Applied Mathematics, SIAM, Philadelphia, 1992.
[50] ——, *Orthonormal bases of compactly supported wavelets II. Variations on a theme*, SIAM J. Math. Anal., 24 (1993), pp. 499–519.
[51] I. Daubechies, A. Grossmann, and Y. Meyer, *Painless nonorthogonal expansions*, J. Math. Phys., 27 (1986), pp. 1271–1283.
[52] I. Daubechies and J. C. Lagarias, *Two-scale difference equations I. Existence and global regularity of solutions*, SIAM J. Math. Anal., 22 (1991), pp. 1388–1410.
[53] ——, *Two-scale difference equations II. Local regularity, infinite products of matrices and fractals*, SIAM J. Math. Anal., 23 (1992), pp. 1031–1079.
[54] C. de Boor, R. A. DeVore, and A. Ron, *On the construction of multivariate (pre)wavelets*, Constr. Approx., 9 (1993), pp. 123–166.
[55] B. Deng and B. Jawerth, *Biorthogonal wavelet packets on closed intervals*. Preprint.
[56] G. Deslauriers and S. Dubuc, *Interpolation dyadique*, in Fractals, Dimensions non entières

et applications, Masson, Paris, Paris, 1987, pp. 44–55.

[57] ——, *Symmetric iterative interpolation processes*, Constr. Approx., 5 (1989), pp. 49–68.

[58] R. A. DeVore, B. Jawerth, and B. J. Lucier, *Image compression through wavelet transform coding*, IEEE Trans. Inform. Theory, 38 (1992), pp. 719–746.

[59] ——, *Surface compression*, Comput. Aided Geom. Des., 9 (1992), pp. 219–239.

[60] R. A. DeVore and B. J. Lucier, *Wavelets*, in Acta Numerica 1, Cambridge University Press, 1991, pp. 1–56.

[61] D. L. Donoho, *Interpolating wavelet transforms*. Preprint Department of Statistics, Stanford University, 1992.

[62] ——, *Unconditional bases are optimal bases for data compression and for statistical estimation*, Appl. Comput. Harmon. Anal., 1 (1993), pp. 100–115.

[63] D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness by wavelet shrinkage*. Preprint Department of Statistics, Stanford University, 1992.

[64] ——, *Ideal spatial adaptation via wavelet shrinkage*. Preprint Department of Statistics, Stanford University, 1992.

[65] ——, *New minimax theorems, thresholding, and adaptation*. Preprint Department of Statistics, Stanford University, 1992.

[66] T. Eirola, *Sobolev characterization of solutions of dilation equations*, SIAM J. Math. Anal., 23 (1992), pp. 1015–1030.

[67] H. G. Feichtinger and K. Gröchenig, *Theory and practise of irregular sampling*, in [13], pp. 305–363.

[68] ——, *Irregular sampling theorems and series expansions of band-limited functions*, SIAM J. Math. Anal., 23 (1992), pp. 530–556.

[69] G. Fix and G. Strang, *Fourier analysis of the finite element method in Ritz-Galerkin theory*, Stud. Appl. Math, 48 (1969), pp. 265–273.

[70] P. Franklin, *A set of continuous orthogonal functions*, Math. Ann, 100 (1928), pp. 522–529.

[71] M. Frazier and B. Jawerth, *Decomposition of Besov spaces*, Indiana Univ. Math. J., 34 (1985), pp. 777–799.

[72] ——, *The $\varphi$-transform and applications to distribution spaces*, in Function Spaces and Applications, M. C. et al., ed., no. 1302 in Lecture Notes in Math., 1988, pp. 223–246.

[73] ——, *A discrete transform and decompositions of distribution spaces*, J. Func. Anal, 93 (1990), pp. 34–170.

[74] M. Frazier, B. Jawerth, and G. Weiss, *Littlewood-Paley theory and the study of function spaces*, no. 79 in Regional Conference Series in Mathematics, American Mathematical Society, Providence, 1991.

[75] J. Geronimo, D. Hardin, and P. R. Massopust, *Fractal functions and wavelet expansions based on several scaling functions*, J. Approx. Theory, To appear.

[76] R. Glowinski, W. M. Lawton, M. Ravrchol, and E. Tenenbaum, *Wavelet solution of linear and nonlinear elliptic parabolic and hyperbolic problems in one space dimension*, in Proceedings of the 9th International Conference on Numerical Methods in Applied Sciences and Engineering, SIAM, Philadelphia, 1990.

[77] A. Grossmann, R. Kronland-Martinet, and J. Morlet, *Reading and understanding continuous wavelet transforms*, in [43], pp. 2–20.

[78] A. Grossmann and J. Morlet, *Decompostion of Hardy functions into square integrable wavelets of constant shape*, SIAM J. Math. Anal., 15 (1984), pp. 723–736.

[79] ——, *Decomposition of functions into wavelets of constant shape, and related transforms*, in Mathematics and Physics, Lectures on Recent Results, L. Streit, ed., World Scientific Publishing, Signapore, 1985.

[80] A. Grossmann, J. Morlet, and T. Paul, *Transforms associated to square integrable group representations I. General results*, J. Math. Phys., 26 (1985), pp. 2473–2479.

[81] A. Haar, *Zur Theorie der orthogonalen Funktionen-Systeme*, Math. Ann., 69 (1910), pp. 331–371.

[82] D. P. Hardin, B. Kessler, and P. R. Massopust, *Multiresolution analyses based on fractal functions*, J. Approx. Theory, 71 (1992), pp. 104–120.

[83] C. E. Heil and D. F. Walnut, *Continuous and discrete wavelet transforms*, SIAM Rev., 31 (1989), pp. 628–666.

[84] S. Jaffard, *Wavelet methods for fast resolution of elliptic problems*, SIAM J. Numer. Anal., 29 (1992), pp. 965–986.

[85] A. J. E. M. Janssen, *The Zak transform: a signal transform for sampled time-continuous signals*, Philips J. Res., 43 (1988), pp. 23–69.

[86] B. Jawerth, *On Besov spaces*, Tech. Report 1, Lund, 1977.

[87] B. Jawerth and G. Peters, *Wavelets on non smooth sets of $\mathbb{R}^n$*. Preprint, 1993.

[88] B. Jawerth and W. Sweldens, *Wavelet multiresolution analyses adapted for the fast solution of boundary value ordinary differential equations*, in Sixth Copper Mountain Conference on Multigrid Methods, N. D. Melson, T. A. Manteuffel, and S. F. McCormick, eds., NASA Conference Publication 3224, 1993, pp. 259–273.

[89] R.-Q. Jia and C. A. Micchelli, *Using the refinement equations for the construction of pre-wavelets II: Powers of two*, in Curves and Surfaces, P. J. Laurent, A. L. Méhauté, and L. L. Schumaker, eds., Academic Press, New York, 1991.

[90] S. Kelly, M. Kon, and L. Raphael, *Pointwise convergence of wavelet expansions*, Bull. Amer. Math. Soc. (N.S.), 29 (1994).

[91] J. R. Klauder and B.-S. Skagerstam, *Coherent States*, World Scientific, Signapore, 1985.

[92] T. H. Koornwinder, ed., *Wavelets: an elementary treatment of theory and applications*, no. 1 in Series in Approximations and Decompositions, World Scientific, Signapore, 1993.

[93] A. Latto and E. Tenenbaum, *Les ondelettes à support compact et la solution numérique de l'équation de burgers*, C. R. Acad. Sci. Paris Sér. I Math., 311 (1990), pp. 903–909.

[94] W. M. Lawton, *Necessary and sufficient conditions for constructing orthonormal wavelets bases*, J. Math. Phys., 32 (1991), pp. 57–61.

[95] P.-G. Lemarié, *Ondelettes a localisation exponentielle*, J. Math. Pures Appl., 67 (1988), pp. 227–236.

[96] ———, ed., *Les Ondelettes en 1989*, no. 1438 in Lecture Notes in Math., Springer-Verlag, 1990.

[97] P.-G. Lemarié and Y. Meyer, *Ondelettes et bases hilbertiennes*, Rev. Mat. Iberoamericana, 2 (1986), pp. 1–18.

[98] Y. Maday, V. Perrier, and J.-C. Ravel, *Adaptivité dynamique sur bases d'ondelettes pour l'approximation d'équations aux dérivées partielles*, C. R. Acad. Sci. Paris Sér. I Math., I (1991), pp. 405–410.

[99] W. Madych, *Some elementary properties of multiresolution analysis of* $L^2(\mathbf{R})$, in [21], pp. 259–294.

[100] S. Mallat and W. L. Hwang, *Singularity detection and processing with wavelets*, IEEE Trans. Inform. Theory, (1992), pp. 617–643.

[101] S. Mallat and S. Zhong, *Wavelet transform maxima and multiscale edges*, in [116], pp. 67–104.

[102] ———, *Characterization of signals from multiscale edges*, IEEE Trans. Patt. Anal. Mach. Intell., 14 (1992), pp. 710–732.

[103] S. G. Mallat, *Multifrequency channel decompositions of images and wavelet models*, IEEE Trans. Acoust. Speech Signal Process., 37 (1989), pp. 2091–2110.

[104] ———, *Multiresolution approximations and wavelet orthonormal bases of* $L^2(\mathbf{R})$, Trans. Amer. Math. Soc., 315 (1989), pp. 69–87.

[105] ———, *A theory for multiresolution signal decomposition: The wavelet representation*, IEEE Trans. Patt. Anal. Mach. Intell., 11 (1989), pp. 674–693.

[106] Y. Meyer, *Ondelettes et Opérateurs, I: Ondelettes, II: Opérateurs de Calderón-Zygmund, III:* (with R. Coifman), *Opérateurs multilinéaires*, Hermann, Paris, 1990. English translation of first volume is published by Cambridge University Press.

[107] ———, *Ondelettes sur l'intervalle*, Rev. Mat. Iberoamericana, 7 (1992), pp. 115–133.

[108] ———, *Wavelets: Algorithms and Applications*, SIAM, Philadelphia, 1993.

[109] C. A. Micchelli, *Using the refinement equations for the construction of pre-wavelets*, Numerical Algorithms, 1 (1991), pp. 75–116.

[110] C. A. Micchelli, C. Rabut, and F. I. Utretas, *Using the refinement equations for the construction of pre-wavelets III: Elliptic splines*, Numerical Algorithms, 1 (1991), pp. 331–352.

[111] P. Oswald, *On discrete norm estimates related to multilevel preconditioners in the finite element method*, in Constructive Theory of Functions, Proc. Int. Conf. Varna 1991, Bulg. Acad. Sci., Sofia, 1992, pp. 203–214.

[112] J. Peetre, *New Thoughts on Besov Spaces*, Duke Univ. Math. Series, Durham, NC, 1976.

[113] S. D. Riemenschneider and Z. Shen, *Wavelets and pre-wavelets in low dimensions*, J. Approx. Theory, 71 (1992), pp. 18–38.

[114] O. Rioul, *Simple regularity criteria for subdivision schemes*, SIAM J. Math. Anal., 23 (1992), pp. 1544–1576.

[115] O. Rioul and M. Vetterli, *Wavelets and signal processing*, IEEE Signal Proc. Mag., (1991), pp. 14–38.

[116] M. B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, and L. Raphael, eds., *Wavelets and their Applications*, Jones and Bartlett, 1992.

[117] N. Saito and G. Beylkin, *Multiresolution representations using the autocorrelation func-*

*tions of compactly supported wavelets*. Preprint University of Colorado at Boulder.

[118] I. J. SCHOENBERG, *Cardinal interpolation and spline functions*, J. Approx. Theory, 2 (1969), pp. 167–627.

[119] L. L. SCHUMAKER AND G. WEBB, eds., *Recent Advances in Wavelet Analysis*, Academic Press, 1993.

[120] M. J. SHENSA, *Wedding the à trous and Mallat algorithms*, IEEE Trans. Signal Process., 40 (1992), pp. 2464–2482.

[121] J. STÖCKLER, *Multivariate wavelets*, in [21], pp. 325–356.

[122] G. STRANG, *Wavelets and dilation equations: A brief introduction*, SIAM Rev., 31 (1989), pp. 614–627.

[123] ———, *Wavelet transforms versus Fourier transforms*, Bull. Amer. Math. Soc. (N.S.), 28 (1993), pp. 288–305.

[124] G. STRANG AND G. FIX, *A Fourier analysis of the finite element variational method*, in Constructive aspects of Functional Analysis, Edizione Cremonese, Rome, 1973.

[125] R. S. STRICHARTZ, *How to make wavelets*, Amer. Math. Monthly, 100 (1993), pp. 539–556.

[126] J. O. STRÖMBERG, *A modified Franklin system and higher order spline systems on $\mathbb{R}^n$ as unconditional bases for Hardy spaces*, in Conference on Harmonic Analysis in Honor of Antoni Zygmund, B. et al., ed., vol. II, Univ. of Chicago Press, 1981, pp. 475–494.

[127] W. SWELDENS AND R. PIESSENS, *Asymptotic error expansions of wavelet approximations of smooth functions II*, Tech. Report 1993:4, Industrial Mathematics Initiative, Department of Mathematics, University of South Carolina.

[128] ———, *Quadrature formulae and asymptotic error expansions for wavelet approximations of smooth functions*, SIAM J. Numer. Anal., To appear.

[129] M. UNSER AND A. ALDROUBI, *Polynomial splines and wavelets — a signal processing perspective*, in [21], pp. 543–601.

[130] M. UNSER, A. ALDROUBI, AND M. EDEN, *On the asymptotic convergence of b-spline wavelets to Gabor functions*, IEEE Trans. Inform. Theory, 38 (1992), pp. 864–872.

[131] ———, *A family of polynomial spline wavelet transforms*, Signal Process., 30 (1993), pp. 141–162.

[132] P. P. VAIDYANATHAN, *Theory and design of M-channel maximally decimated quadrature mirror filters with arbitrary M, having perfect reconstruction property*, IEEE Trans. Acoust. Speech Signal Process., 36 (1987), pp. 476–492.

[133] M. VETTERLI, *Filter banks allowing perfect reconstruction*, Signal Process., 10 (1986), pp. 219–244.

[134] M. VETTERLI AND C. HERLEY, *Wavelets and filter banks: theory and design*, IEEE Trans. Acoust. Speech Signal Process., 40 (1992), pp. 2207–??

[135] L. F. VILLEMOES, *Wavelet analysis of refinable functions*, in Curves and Surfaces II, P. J. Laurent, A. L. Méhauté, and L. L. Schumaker, eds., A. K. Peters, Boston.

[136] ———, *Energy moments in time and frequency for two-scale difference equation solutions and wavelets*, SIAM J. Math. Anal., 23 (1992), pp. 1119–1543.

[137] H. VOLKMER, *On the regularity of wavelets*, IEEE Trans. Inform. Theory, 38 (1992), pp. 872–876.

[138] G. G. WALTER, *A sampling theorem for wavelet subspaces*, IEEE Trans. Inform. Theory, 38 (1992), pp. 881–884.

[139] M. V. WICKERHAUSER, *Acoustic signal compression with wavelet packets*, in [21], pp. 679–700.

[140] J.-C. XU AND W.-C. SHANN, *Galerkin-wavelet methods for two-point boundary value problems*, Numer. Math., 63 (1992), pp. 123–142.

**Appendix: The periodic fast wavelet transform algorithm.** We will give here a pseudo code implementation of the periodic fast wavelet transform. We assume that $len\_hp$ coefficients $h_k$ are non-zero, starting with the one with index $k = min\_hp$. Similar assumptions hold for the $g_k$, $\tilde{h}_k$, and $\tilde{g}_k$ with lengths $len\_gp$, $len\_hd$ and $len\_gd$, and starting indices $min\_gp$, $min\_hd$ and $min\_gd$ respectively. These coefficients are stored in 4 vectors such that

$$hp[k] = a\, h_{k+min\_hp}, \quad gp[k] = a\, g_{k+min\_gp},$$

$$hd[k] = b\, \tilde{h}_{k+min\_hd}, \quad \text{and} \quad gd[k] = b\, \tilde{g}_{k+min\_gd},$$

where $a\,b = 2$. We start with $2^n$ coefficients $\lambda_{n,l}$ of a function of $V_n$ and can thus apply $n$ steps of the algorithm. These coefficients are initially stored in a vector $v$. The computed wavelet coefficients are stored in a vector $w$ such that

$$w = [\ \lambda_{0,0}\quad \gamma_{0,0}\quad \gamma_{1,0}\quad \gamma_{1,1}\quad \gamma_{2,0}\quad \cdots\quad \gamma_{2,3}\quad \cdots\quad \gamma_{n-1,0}\quad \cdots\quad \gamma_{n-1,2^{n-1}-1}\ ].$$

The algorithms are written in such a way to reduce operations in the inner loops. They are however not highly optimized not to affect readability too much. The index notation $a\,(b)\,c$ stands for $a$, $a+b$, $\ldots$, $c$ and the operator floor($a$) rounds $a$ to the nearest integer towards minus infinity.

```
for j ← n − 1 (−1) 0
    w[0 (1) 2^(j+1) − 1] ← 0
    for l ← 0 (1) 2^j − 1
        i ← (2 * l + min_hd) mod 2^(j+1)
        for k ← 0 (1) len_hd
            w[l]  ←  w[l] + hd[k] * v[i]
              i   ←  (i + 1) mod 2^(j+1)
        end for
        i ← (2 * l + min_gd) mod 2^(j+1)
        ls ← l + 2^j
        for k ← 0 (1) len_gd
            w[ls]  ←  w[ls] + gd[k] * v[i]
               i   ←  (i + 1) mod 2^(j+1)
        end for
    end for
    v ← w[0 (1) 2^j − 1]
end for
```

```
for j ← 1 (1) n
    v[0 (1) 2^j − 1] ← 0
    for k ← 0 (1) 2^j − 1
        i ← (floor((k − min_hp)/2)) mod 2^(j−1)
        lb ← (k − min_hp) mod 2
        for l ← lb (2) len_hp
            v[k]  ←  v[k] + hp[l] * w[i]
               i  ←  (i − 1) mod 2^(j−1)
        end for
        i ← (floor((k − min_gp)/2)) mod 2^(j−1)
        lb ← (k − min_gp) mod 2
        for l ← lb (2) len_gp
            v[k]  ←  v[k] + gp[l] * w[i + 2^(j+1)]
               i  ←  (i − 1) mod 2^(j−1)
        end for
    end for
    w[0 (1) 2^j − 1] = v
end for
```

# The Lifting Scheme:
# A New Philosophy in Biorthogonal Wavelet Constructions


Wim Sweldens

Katholieke Universiteit Leuven Belgium, Department of Computer Science

Wim.Sweldens@cs.kuleuven.ac.be


## ABSTRACT


In this paper we present the basic idea behind the lifting scheme, a new construction of biorthogonal wavelets which does not use the Fourier transform. In contrast with earlier papers we introduce lifting purely from a wavelet transform point of view and only consider the wavelet basis functions in a later stage. We show how lifting leads to a faster, fully in-place implementation of the wavelet transform. Moreover, it can be used in the construction of second generation wavelets, wavelets that are not necessarily translates and dilates of one function. A typical example of the latter are wavelets on the sphere.



**Keywords:** wavelet, biorthogonal, in-place calculation, lifting


## 1 Introduction

At the present day it has become virtually impossible to give the definition of a "wavelet". The research field is growing so fast and novel contributions are made at such a rate that even if one manages to give a definition today, it might be obsolete tomorrow. One, very vague, way of thinking about wavelets could be:

"*Wavelet* are **building blocks** *that can* **quickly decorrelate** *data.*"

This sentence at least incorporates three of the main features of wavelets. First of all, they are *building blocks* for general data sets or functions. Mathematically we say that they form a basis or, more general a frame. This means that each element of a general class can be written in a stable way as a linear combination of the wavelets. If we denote the wavelets by $\psi_i$ and the coefficients by $\gamma_i$, we can write a general function $f$ as

$$f = \sum_i \gamma_i \psi_i.$$

Secondly, wavelets have the power to *decorrelate*. This means that the representation of the data in terms of the wavelet coefficients $\gamma_i$ is somehow more "compact" then the original representation. In information-theoretic jargon, we say that the entropy in the wavelet representation is smaller then in the original representation. In approximation-theoretic jargon, we want to get an accurate approximation of $f$ by only using a small fraction of the wavelet coefficients.

The way to get this decorrelation power is to construct wavelets which already in some way resemble the data we want to represent. More specifically, we would like the wavelets to have the same correlation structure as the data. For example, most signals we encounter in daily life have both correlation in space and frequency. Samples which are spatically close are much more correlated then ones that are far apart, and frequencies often occur in bands. To analyze and represent such signals we need wavelets that are local in space and frequency. Typically this is achieved by building wavelets which have compact support (localization in space), which are smooth (decay towards high frequencies), and which have vanishing moments (decay towards low frequencies).

Finally, we want to *quickly* find the wavelet representation of the data. More precisely, we want to switch between the original representation of the data and its wavelet representation in a time proportional to the size of the data. The fast decorrelation power of wavelets is the key to applications such as data compression, fast data transmission, noise cancellation, signal recovering, and fast numerical algorithms.

The purpose of this paper is to introduce the *lifting scheme*, a new tool in the construction of biorthogonal wavelets. The main difference with classic constructions such as[1-3] is that it does not employ the Fourier transform. Until recently, the Fourier transform has been instrumental in wavelet constructions. The underlying reason is that wavelets are traditionally defined as translates and dilates of one function, and translation and dilation become algebraic operations after Fourier transform. The wavelet construction then relies on certain polynomial factorizations. We refer to wavelets which are translates and dilates of one function as *first generation wavelets*. In the case of first generation wavelets, the lifting scheme will never come up with wavelets which somehow could not be found by the techniques developed by Cohen, Daubechies, and Feauveau in.[2] Nevertheless, using lifting to (re)construct these wavelets has the following advantages:

1. It allows a faster implementation of the wavelet transform. Traditionally, the fast wavelet transform is calculated with a two-band subband transform scheme. In each step the signal is split into a high pass and low pass band and then subsampled. Recursion occurs on the low pass band. The lifting scheme makes optimal use of similarities between the high and low pass filters to speed up the calculation. In some cases the number of operations can be reduced by a factor of two.
2. The lifting scheme allows a fully in-place calculation of the wavelet transform. In other words, no auxiliary memory is needed and the original signal (image) can be replaced with its wavelet transform.
3. In the classical case, it is not immediately clear that the inverse wavelet transform actually is the inverse of the forward transform. Only with the Fourier transform one can convince oneself of the perfect reconstruction property. With the lifting scheme, the inverse wavelet transform can immediately be found by undoing the operations of the forward transform. In practise, this comes down to simply reversing the order of the operations and changing each $+$ into a $-$ and vice versa.
4. The lifting scheme is a very natural way to introduce wavelets in a classroom. Indeed, since it does not rely on the Fourier transform, the properties of the wavelets and the wavelet transform do not appear as

somehow "magical" to students who do not have a strong background in Fourier analysis.

Since lifting does not rely on the Fourier transform, it can be used to construction wavelets in settings where translation and dilation, and thus the Fourier transform, cannot be used. We refer to such wavelet as *second generation wavelets*. Typical examples are:

1. *Wavelets on bounded domains*: The construction of wavelets on domains in a Euclidean space is needed in applications such as image segmentation and the numerical solution of partial differential equations. A special case is the construction of wavelets on an interval, which is needed to transform finite length signals without introducing artifacts at the boundaries.
2. *Wavelets on curves and surfaces*: To analyze data that live on curves or surfaces or to solve equations on curves or surfaces, one needs wavelets intrinsically defined on these manifolds, independent of parametrization.
3. *Weighted wavelets*: Diagonalization of differential operators and weighted approximation require a basis adapted to weighted measures. Wavelets biorthogonal with respect to a weighted inner product are needed.
4. *Wavelets and irregular sampling*: Many real life problems require basis functions and transforms adapted to irregularly sampled data.

It is obvious that wavelets adapted to these setting cannot be formed by translation and dilation. The Fourier transform can thus no longer be used as a construction tool. The lifting scheme provides an alternative.

There are two ways to introduce lifting. The first one is concerned with the basis functions, i.e. the scaling functions, dual scaling functions, wavelets, and dual wavelets, and how lifting affects them. This approach was taken in the original papers.[9,10] In this paper, however, we follow a different approach, namely we first discuss how lifting affects the wavelet transform. We have found this to be a much more natural way to introduce lifting. In a later section, we will briefly mention what happens to the basis functions. Of course, theoretically both approaches are equivalent. In fact one can be seen as adjoint to the other.

## 2 The basic idea behind lifting

A canonical case of lifting consists of three stages, which we refer to as: *split*, *predict*, and *update*. We here describe the basic idea behind each and later work out a concrete example. Assume we start with an abstract data set, which we refer to as $\lambda_0$. We know this data set has some correlation structure and we would like to exploit it to obtain a more compact representation.

In the first stage we split the data into two smaller subsets $\lambda_{-1}$ and $\gamma_{-1}$. (We use negative indices here because the convention is that the smaller the data set, the smaller the index.) We refer to $\gamma_{-1}$ as the wavelet subset. We do not impose any no restriction on how the data should be split, nor on the relative size of each of the subsets. The only thing we need is some procedure to join $\lambda_{-1}$ and $\gamma_{-1}$ back into the original data set $\lambda_0$. The easiest possibility for the split is a simply brutal cut of the data set into two disjoint parts. This choice we refer to as the *Lazy wavelet*. Think for example of cutting an image into two parts with a pair of scissors.

As we said before, we would like to get a more compact representation of $\lambda_0$. Consider the case were $\gamma_{-1}$ does not contain any information (e.g. that part of the image is entirely black). Then we would immediately have a more compact representation since we can simply replace $\lambda_0$ with the smaller set $\lambda_{-1}$. Indeed the extra part needed to reassemble $\lambda_0$ does not contain any information.

Obviously this situation hardly ever occurs in practise. Therefore, in a second stage, we try to use the $\lambda_{-1}$ subset to predict the $\gamma_{-1}$ subset based on the correlation present in the original data. If we can find a prediction operator $\mathcal{P}$, independent of the data, so that

$$\gamma_{-1} = \mathcal{P}(\lambda_{-1}),$$

then again we can replace the original data set with $\lambda_{-1}$, since now we can predict the part missing to reassemble $\lambda_0$. The construction of a prediction operator is typically based on some model of the data which reflects its correlation structure. Obviusly the prediction operator $\mathcal{P}$ cannot be dependent on the data, otherwise we would hide information in $\mathcal{P}$.

Again, in practise it might not be possibly to exactly predict $\gamma_{-1}$ based on $\lambda_{-1}$. However, $\mathcal{P}(\lambda_{-1})$ is likely to be close to $\gamma_{-1}$. Thus we might want to replace $\gamma_{-1}$ with the difference between itself and its predicted value $\mathcal{P}(\lambda_{-1})$. If the prediction is reasonable, this difference will contain much less information then the original $\gamma_{-1}$ set. We denote this abstract difference operator with a $-$ sign and thus get

$$\gamma_{-1} := \gamma_{-1} - \mathcal{P}(\lambda_{-1}).$$

The wavelet subset now encodes how much the data deviates from the model on which $\mathcal{P}$ was built.

We now have some more insight in how to split the original data set. Indeed, in order to get the maximal data reduction from prediction, we need the subsets $\lambda_{-1}$ and $\gamma_{-1}$ to be maximally correlated. Cutting an image into a left and right part might not be the best idea since pixels on the far left and the far right are hardly correlated. Predicting the right half of the image based on the left is thus a though job. A better idea is to interlace the two sets. We will come back to this later.

At this moment we can replace the original data with the smaller set $\lambda_{-1}$ and the wavelet set $\gamma_{-1}$. With a good prediction, the two subsets $\{\lambda_{-1}, \gamma_{-1}\}$ yield a more compact representation then the orginal set $\lambda_0$. We can now iterate this scheme. We split $\lambda_{-1}$ into two subsets $\lambda_{-2}$ and $\gamma_{-2}$ and then replace $\gamma_{-2}$ with the difference between $\gamma_{-2}$ and $\mathcal{P}(\lambda_{-2})$. After $n$ steps we have replaced the original data with the wavelet representation $\{\lambda_{-n}, \gamma_{-n}, \cdots, \gamma_{-1}\}$. Given that the wavelet sets encode the difference with some predicted value based on a correlation model, this is likely to give a more compact representation.

This scheme sounds promising, but in some cases we are not completely satisfied. The reason is that we often want some global properties of the original data set to be maintained in the smaller versions $\lambda_{-j}$. For example, in the case of an image, we would like the smaller images $\lambda_{-j}$ to have the same overall brightness, i.e. the same

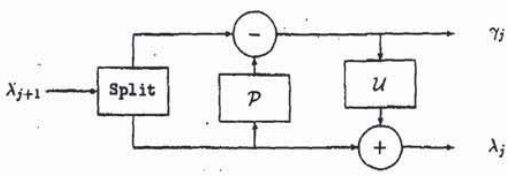


Figure 1: *The lifting scheme: split, predict, and update.*

average pixel value. If the splitting stage is simply subsampling and we iterate the scheme till $\lambda_{-n}$ is only 1 pixel, that pixel will be an arbitrary pixel from the original image. We would rather have the last value to be the average of all the pixel values in the original image. In fact, we are facing a worst case example of a problem known as aliasing.

We can solve part of this problem by introducing a third stage. The idea is to find a better $\lambda_{-1}$ so that a certain scalar quantity $Q()$, like e.g. the mean, is preserved, or

$$Q(\lambda_{-1}) = Q(\lambda_0).$$

We could do so by finding a new operator to extract $\lambda_{-1}$ directly from $\lambda_0$, but we decide not to for two reasons. First, this would create a scheme which is very hard to invert. Secondly, we would like to reuse the work already done maximally. Therefore we propose to use the already computed wavelet set $\gamma_{-1}$ to update $\lambda_{-1}$ so that the latter preserves $Q()$. In other words, we construct an operator $\mathcal{U}$ and update $\lambda_{-1}$ as

$$\lambda_{-1} := \lambda_{-1} + \mathcal{U}(\gamma_{-1}).$$

The three stages of lifting are depicted in a block diagram in Figure 1. Again we can now iterate the scheme. This leads to the following wavelet transform algorithm (with a C-like syntax):

$$\texttt{For j = -1 downto -n:} \quad \begin{cases} \{\lambda_j, \gamma_j\} & := \ \texttt{Split}(\lambda_{j+1}) \\ \gamma_j & -= \ \mathcal{P}(\lambda_j) \\ \lambda_j & += \ \mathcal{U}(\gamma_j). \end{cases}$$

We can now illustrate one of the nice properties of lifting: once we have the forward transform, we can immediately derive the inverse. The only thing to do is to reverse the operations and toggle $+$ and $-$. This leads to the following algorithm for the inverse wavelet transform:

$$\texttt{For j = -n to -1:} \quad \begin{cases} \lambda_j & -= \ \mathcal{U}(\gamma_j) \\ \gamma_j & += \ \mathcal{P}(\lambda_j) \\ \lambda_{j+1} & := \ \texttt{Join}(\lambda_j, \gamma_j). \end{cases}$$

## 3 A simple example

In this section we consider a simple example to illustrate the ideas of the previous section. Suppose we sample a signal $f(t)$ with sampling distance $\Delta t = 1$. We denote the original samples by $\lambda_0 = \{\lambda_{0,k} = f(k) \mid k \in \mathbf{Z}\}$.

We first need to define the split stage. As mentioned in the previous section this implies splitting the data into two parts which are maximally correlated. As the correlation in most signals is local, i.e. neighboring samples are much more correlated then ones that are far apart, we simply subsample the data into even and odd indexed samples. This is an example of a Lazy wavelet transform. We obtain two sequences $\lambda_{-1}$ and $\gamma_{-1}$ with coefficients

$$\lambda_{-1,k} := \lambda_{0,2k} \quad \text{and} \quad \gamma_{-1,k} := \lambda_{0,2k+1} \quad \text{for} \quad k \in \mathbf{Z}. \tag{1}$$

Next we need to find a operator to predict $\gamma_{-1}$ based on $\lambda_{-1}$. Again assuming maximal correlation amongst neighboring samples, we simply suggest to predict an odd sample $\lambda_{0,2k+1}$ as the average of its two (even) neighbors: $\lambda_{-1,k}$ and $\lambda_{-1,k+1}$. The difference with this prediction then becomes

$$\gamma_{-1,k} := \gamma_{-1,k} - 1/2\,(\lambda_{-1,k} + \lambda_{-1,k+1}). \tag{2}$$

The model used to build $\mathcal{P}$ is a function piecewise linear over intervals of length 2. If the original signal complies with the model, all wavelet coefficients in $\gamma_{-1}$ are zero. In other words, the wavelet coefficients measure to which extend the original signal ***fails to be linear***. Their expected value is small. In terms of frequency content, the wavelet coefficients capture high frequencies present in the original signal.

However, the frequency localization of the signals $\lambda_{-1}$ and $\gamma_{-1}$ is far from ideal. It would be nice if the $\lambda_{-1}$ signal somehow captures the low frequencies, and the $\gamma_{-1}$ the high frequencies. Right now the $\lambda_{-1}$ signal is simply subsampled and its frequency content thus stretches out over the whole band of the original signal. Again, we have the worst case example of aliasing. In the update stage, we can reduce the amount of aliasing by at least assuring that the DC component ends up entirely in the $\lambda_{-1}$ part. In other words, we would like the average of the signal to be maintained in $\lambda_{-1}$, or

$$\sum_k \lambda_{-1,k} = 1/2 \sum_k \lambda_{0,k}.$$

This is precisely the scalar quantity $Q()$ of the previous section which we would like to preserve. We therefore update the $\lambda_{-1,k}$ with the help of the wavelet coefficients $\gamma_{-1,k}$. Again we use the neighboring wavelet coefficients and thus propose an update $\mathcal{U}$ of the form:

$$\lambda_{-1,k} := \lambda_{-1,k} + A\,(\gamma_{-1,k-1} + \gamma_{-1,k}).$$

To find $A$ we calculate the average:

$$\sum_k \lambda_{-1,k} = \sum_k \lambda_{0,2k} + 2\,A \sum_k \gamma_{-1,k} = (1 - 2\,A) \sum \lambda_{0,2k} + 2\,A \sum \lambda_{0,2k+1}.$$

From this we see that the correct choice to maintain the average is $A = 1/4$. One step in the wavelet transform is shown in the scheme in Figure 2. By iterating this scheme we get a complete wavelet transform. The inverse transform can be derived immediately as shown in the previous section. Note that at no point we used the Fourier transform.

$$
\begin{tikzcd}
\cdots & \lambda_{j,2k} \arrow[dl, "-1/2"'] \arrow[d, dashed] \arrow[dr, "-1/2"] & \lambda_{j,2k+1} \arrow[d] & \lambda_{j,2k+2} \arrow[dl, "-1/2"'] \arrow[d, dashed] \arrow[dr, "-1/2"] & \cdots \\
\cdots \arrow[dr, "1/4"] & \lambda_{j,2k} \arrow[d] & \gamma_{j-1,k} \arrow[dl, "1/4"'] \arrow[d, dashed] \arrow[dr, "1/4"] & \lambda_{j,2k+2} \arrow[d] & \cdots \arrow[dl, "1/4"'] \\
\cdots & \lambda_{j-1,k} & \gamma_{j-1,k} & \lambda_{j-1,k+1} & \cdots
\end{tikzcd}
$$

Figure 2: *The lifting scheme: Split, calculate the wavelet coefficients $\gamma_{j-1,m}$ as the failure to be linear, and use them to update the $\lambda_{j-1,k}$.*

The wavelet transform presented here in fact is the $(N = 2, \widetilde{N} = 2)$ biorthogonal wavelet transform of Cohen-Daubechies-Feauveau.[2] This simple example already shows how the lifting scheme can speed up the implementation of the wavelet transform. Classicly the $\{\lambda_{-1,k}\}$ coefficients are found as the convolution of the $\{\lambda_{0,k}\}$ coefficients with the filter $\widetilde{h} = \{-1/8, 1/4, 3/4, 1/4, -1/8\}$. This step would take 6 operations per coefficient while lifting only needs 3.

This is only one simple instance of lifting. A whole family of biorthogonal wavelets can be constructed by varying the three stages of lifting:

1. *Split:* Other choices but the Lazy wavelet are possible as an initial split. A typical alternative is the Haar wavelet transform.
2. *Predict:* In wavelet terminology the prediction step establishes the number of vanishing moments ($N$) of the dual wavelet. In other words, if the original signal is a polynomial of degree less than $N$, all wavelet coefficients will be zero. In our example $N = 2$, but higher order schemes are easily obtained by involving more neighbors.
3. *Update:* Again in wavelet terminology, the update step establishes the number of vanishing moments ($\widetilde{N}$) of the primal wavelet. In other words the transform preserves the first $\widetilde{N}$ moments of the $\lambda_j$ sequences. The example above had $\widetilde{N} = 2$ (one extra because of symmetry). Again higher order ones can be constructed by involving more neighbors. In some cases, namely when the split stage already creates a wavelet with a vanishing moment (such as the Haar) the update stage can be omitted.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\lambda_{-3,0}$ | $\gamma_{-3,0}$ | $\gamma_{-2,0}$ | $\gamma_{-2,1}$ | $\gamma_{-1,0}$ | $\gamma_{-1,1}$ | $\gamma_{-1,2}$ | $\gamma_{-1,3}$ |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\lambda_{-3,0}$ | $\gamma_{-1,0}$ | $\gamma_{-2,0}$ | $\gamma_{-1,1}$ | $\gamma_{-3,0}$ | $\gamma_{-1,2}$ | $\gamma_{-2,1}$ | $\gamma_{-1,3}$ |

Figure 3: *Mallat (above) and Lifting (below) organization of wavelet coefficients for* $n = 3$.

## 4 In-place calculation

In this section we introduce another feature of lifting: in-place calculation. This means we can replace the original data set with its wavelet transform without having to allocate extra memory. As is well known, the FFT has a similar feature provided one starts with bit-reversing the original samples. The basic idea behind in-place calculation of the wavelet transform using lifting is quite similar.

Assume the original signal has length $2^n$ and the original samples are stored in a vector so that sample $\lambda_{0,k}$ sits in memory location $k$ $(0 \leq k < 2^n)$. In the traditional organization of the wavelet coefficients, as proposed by Mallat in,[6,5] a wavelet coefficient $\gamma_{j,k}$ is stored in location $2^{n+j} + k$, see Figure 3 for an example with $n = 3$. To obtain in-place calculation with lifting, we propose a different ordering of the wavelet coefficients. The idea is to store the coefficient $\gamma_{j,k}$ in location $2^{-j-1} + 2^{-j}\,k$, see Figure 3. Essentially all the coefficients depicted in one column of Figure 2 are stored in the same location. The Lazy wavelet transform is then immediate. Since all other operations can be done with $+=$ or $-=$ operations, we have a fully in-place calculation. Figure 4 shows the in-place organization of the wavelet coefficients of the classic Lena image.

It is also possible to first rearrange the original samples to end up with the Mallat organization of the coefficients after an in-place calculation. The sample $\lambda_{0,k}$ then has to be stored in location $m(k)$ which can be found as follows. Consider the $n$-bit binary representation of $k$: $\mathbf{k}_0\mathbf{k}_1\mathbf{k}_2\cdots\mathbf{k}_n$. Next isolate the trailing zeros (if any) as $\mathbf{k}_0\mathbf{k}_1\mathbf{k}_2\cdots\mathbf{k}_j100\ldots0$. Now the binary expansion of $m(k)$ is $00\cdots01\mathbf{k}_0\mathbf{k}_1\mathbf{k}_2\cdots\mathbf{k}_j$. In other words, it can be seen a partial bit reversal.

## 5 The basis functions

So far, we only explained how lifting affects the wavelet transform. In this section we briefly describe what happens to the basis functions. The basic idea to find a basis function given the transform is very simple. If you want to construct the basis function associated with a coefficient $\lambda_{j,k}$ of $\gamma_{j,k}$, simply put that coefficient to one, all other coefficients to zero, and perform an inverse wavelet transform starting from level $j$. This is known as the *cascade algorithm*.[4] We refer to the basis functions associated with the $\lambda_{j,k}$ (respectively $\gamma_{j,k}$) as scaling functions (respectively wavelets) and denote them with $\varphi_{j,k}$ (respectively $\psi_{j,k}$). If you are interested in the discrete basis, do an inverse transform till level 0, while if you are interested in the continuous basis, do an inverse transform add infinitum.

Figure 4: *In-place organization of Lena wavelet coefficients.*

This way it is easy to see that the Lazy wavelet simple corresponds to a Dirac sequence or a Dirac function. In case of the simple example mentioned earlier, the scaling functions are given by

$$\varphi_{j,k}(x) = \Lambda(2^j x - k),$$

where $\Lambda$ is the classical "Hat" function: $\Lambda(x) = \max\{0, 1-|x|\}$. Essentially all scaling functions are translates and dilates of one particular function. The same is true for the wavelet. By doing one step of the inverse transform we see that

$$\psi_{j,k}(x) = \psi(2^j x - k),$$

where $\psi(x)$ is given by

$$\psi(x) = \Lambda(2x-1) - 1/4\,\Lambda(x) - 1/4\,\Lambda(x+1). \tag{3}$$

In case we would not have an update stage, the wavelet would simply be $\psi(x) = \Lambda(2x-1)$.

This gives us another insight into how lifting constructs wavelets. A new wavelet $\psi(x)$ is found as an old wavelet $\Lambda(2x-1)$ (the one without updating) combined with scaling functions on the *same* level, $\Lambda(x)$ and $\Lambda(x+1)$. This opposed to the classical case were a wavelet is constructed as a linear combination of Hat functions on the next *finer* level, namely $\Lambda(2x-k)$ with $k \in \mathbf{Z}$. The coefficients in (3) are chosen so that the wavelet has two vanishing moments. This is another way to get to the 1/4 coefficients encountered earlier. The prediction stage actually corresponds to a similar operation on the dual wavelet, which is sometimes referred to as dual lifting. We thus start from an almost trivial case, the Lazy wavelet, and gradually build a new wavelet with improved properties, by adding in new basis functions. This is the inspiration behind the name "lifting scheme."

Let us now write the original signal as

$$f(x) = \sum_k \lambda_{0,k}\, \varphi_{0,k}\,.$$

After wavelet transform, the same signal can be written as

$$f(x) = \sum_k \lambda_{-n,k}\, \varphi_{-n,k} + \sum_{j=-n}^{-1} \sum_k \gamma_{j,k}\, \psi_{j,k}.$$

This is precisely a representation with "building blocks that decorrelate" which we were after in the introduction. Because of the correlation structure, many of the wavelet coefficients will be small. We can thus obtain an accurate approximation with only a small number of coefficients by simply omitting the wavelet coefficients below a certain threshold.

## 6 Second generation wavelets

The original motivation for the lifting was to construct wavelets in settings were no Fourier transform is available. The theory of lifting for second generation wavelets is given in.[10] Here we just introduce the basic idea.

The key point in each setting is to define the initial split operation. The easiest choice again is the Lazy wavelet transform. The prediction and updating stage are then quite similar to the ones described above. The main difference lies in the fact that the filter coefficients used in the prediction and update operator might vary depending on location. Indeed, if one wants to account for local irregularities, one cannot use the same filters everywhere. This is precisely why the Fourier transform can no longer be used. As a result the wavelets are not translates and dilates of one function. However, they still enjoy all the nice properties of first generation wavelets, such as fast transform and decorrelation power. They still are compactly supported, smooth, and have vanishing moments.

Let us first consider the case of wavelets on an interval. We essentially want to transform a signal of arbitrary finite length without the use of ad hoc solutions such as zero padding, periodization, or reflection around the edges. The Lazy wavelet transform can still be subsampling even and odd samples. In our simple example on the real line, the predicted value for an odd sample was based on its neighboring even samples left and right. In case of a finite length signal, the same idea can be used as long as the sample is sufficiently far away from the boundary. At the left boundary, if not enough even samples on the left are available to predict an odd sample, one simply replaces the missing samples on the left by extra samples on the right. In other words, we look for more correlated data where it is available. For example, suppose we use cubic interpolation to predict an odd sample based upon 4 neighboring even samples. Away from the boundary we use 2 samples on the left and 2 on the right. Close to the left boundary we might have to use 1 on the left and 3 on the right, or even none on the left and 4 on the right. This automatically leads to filters adapted for boundary constructions. It assures that all wavelets, including the ones at the boundary have the same number of vanishing moments. The wavelet coefficients for the Lena image in Figure 4 were calculated this way. Examples of boundary wavelets can be found in.[11]

In the case of irregular samples the same philosophy can be used. The fact that the sample locations are not on a regular grid poses no problems for the local polynomial prediction or update. The filters now change everywhere to account for the different locations of the samples. Here one has several choices for the Lazy wavelet. One idea is to still use even and odd subsampling. This way the imbalances in the sampling distances are maintained throughout the hierarchy. The alternative is to subsample in such a manner that the ratio of the largest versus smallest sampling distance approaches one. Which one is better depends on the situation at hand. For practical examples we again refer to.[11]

Another typical example of second generation wavelets are wavelets defined on curves, surfaces, or general manifolds. More particularly, the lifting scheme was used to construct wavelets on a sphere in.[7] These spherical wavelets are used for the efficient representation of data that naturally lives on a sphere. Examples are topographic data (earth elevation), bidirectional reflection functions, astrophysical data, and environment maps. These wavelets were used for spherical image processing in.[8]

## 7 Future developments

In this paper we gave a short introduction to lifting, a new method to construct wavelets. For more details and applications we need to refer to the original papers. Here we just would like to mention a few developments which are currently under investigation.

- Wavelets on general surfaces: The construction of the spherical wavelets does not fundamentally rely on the special properties of the sphere. It only uses the fact that one can recursively cut the sphere into spherical triangles. Therefore the construction can be generalized to more general surfaces or manifolds.
- Wavelet packets: The lifting scheme can also be used in the construction of wavelet packets. It is not hard to come up with a Lazy wavelet packet, i.e. a transform which also recursively splits the $\gamma_{-j}$ sets. On these splittings again the prediction and update operators can be used.
- M-band wavelets: Again it is not very difficult to invent a Lazy M-band wavelet. Now we need to find several prediction and update operators.
- Wavelet frames: The lifting scheme can be used to construct overcomplete representations or frames. The key again lies in the correct definition of the Lazy wavelet. The two sets $\gamma_{-1}$ and $\lambda_{-1}$ coming from the split can have some amount of overlapping information or redundancy. Prediction and updating then would lead to wavelet frames.

## Acknowledgment

The author is Senior Research Assistant of the National Fund of Scientific Research Belgium (NFWO). Part of this work was done at the University of South Carolina with support from NSF EPSCoR Grant EHR 9108772 and DARPA Grant AFOSR F49620-93-1-0083. The author would also like to acknowledge Interval Research in Palo Alto where he was visiting when writing this paper. Also special thanks to Peter Schröder for many most inspiring and stimulating discussions. The code to generate the in-place coefficients of the Lena image was written by two South Carolina graduate students, Gabriel Fernández and Senthil Periaswamy, who the author had the pleasure of working with in CS564 and CS798. Finally, it was Geoff Davis who pointed out that the in-place organization of the wavelet coefficients actually corresponds to a partial bit reversal of the indices.

## 8 REFERENCES


1. C. K. Chui. *An Introduction to Wavelets.* Academic Press, San Diego, CA, 1992.
2. A. Cohen, I. Daubechies, and J. Feauveau. Bi-orthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45:485–560, 1992.
3. I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41:909–996, 1988.
4. I. Daubechies. *Ten Lectures on Wavelets.* CBMS-NSF Regional Conf. Series in Appl. Math., Vol. 61. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.
5. S. G. Mallat. Multifrequency channel decompositions of images and wavelet models. *IEEE Trans. Acoust. Speech Signal Process.*, 37(12):2091–2110, 1989.
6. S. G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L_2(\mathbf{R})$. *Trans. Amer. Math. Soc.*, 315(1):69–87, 1989.
7. P. Schröder and W. Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. *Computer Graphics, (SIGGRAPH '95 Proceedings)*, 1995.
8. P. Schröder and W. Sweldens. Spherical wavelets: Texture processing. In P. Hanrahan and W. Purgathofer, editors, *Rendering Techniques '95.* Springer Verlag, Wien, New York, August 1995.
9. 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.
10. W. Sweldens. The lifting scheme: A construction of second generation wavelets. Technical Report 1995:6, Industrial Mathematics Initiative, Department of Mathematics, University of South Carolina, 1995.
11. W. Sweldens and P. Schröder. Building your own wavelets at home. Technical Report 1995:5, Industrial Mathematics Initiative, Department of Mathematics, University of South Carolina, 1995.


**Notes:**

The lifting scheme has connections with many other developments. Space does not allow us to mention them here. They are all pointed out in the original papers, which leads to a bibliography of over a 100 items!

The technical reports from South Carolina are available as Postscript files through anonymous ftp to `ftp.math.sc.edu`, in the directories `/pub/imi_94` and `/pub/imi_95`.

# CREW: Compression with Reversible Embedded Wavelets


Ahmad Zandi James D. Allen Edward L. Schwartz Martin Boliek

RICOH California Research Center
2882 Sand Hill Road, Suite 115, Menlo Park, CA 94025
e-mail: zandi@crc.ricoh.com



## ABSTRACT

Compression with Reversible Embedded Wavelets (CREW) is a unified lossless and lossy continuous-tone still image compression system. It is wavelet-based using a "reversible" approximation of one of the best wavelet filters. Reversible wavelets are linear filters with non-linear rounding which implement exact-reconstruction systems with minimal precision integer arithmetic. Wavelet coefficients are encoded in a bit-significance embedded order, allowing lossy compression by simply truncating the compressed data. For coding of coefficients, CREW uses a method similar to Shapiro's zerotree, and a completely novel method called Horizon. Horizon coding is a context based coding that takes advantage of the spatial and spectral information available in the wavelet domain. CREW provides state of the art lossless compression of medical images (greater than 8 bits deep), and lossy and lossless compression of 8-bit deep images with a single system. CREW has reasonable software and hardware implementations.


## 1 Introduction

Since CREW uses a "reversible" approximation of one of the best known wavelet filters, its performance is equal to or better than other existing methods in both lossy and lossless modes. It encodes the wavelet coefficients in a bit-significance embedded order similar to Shapiro [Sha93], and a completely novel method called *Horizon*. Horizon coding is a context-based coding that takes advantage of the spatial and spectral information available in the wavelet domain. While Zerotree is rightfully considered to be one of the best encoding methods of the wavelet coefficients, it is not efficient when it reaches the lesser significant bits of the coefficients. This usually is tolerable in a lossy system, but for lossless compression the encoding of the less significant bits are of prime importance. Horizon coding, which is particularly useful for the lesser significant bits is general and powerful enough to be used alone.

The original motivation for this system is the compression of medical images, although the same feature set could be very useful for other applications such as pre-press images, satellite images, document processing, world wide web, and other communication systems. At this time, for various reasons, medical image compression is considered to belong to the lossless realm. However, there are definitely future possibilities for the use of lossy compression. Perhaps the image is kept in a lossless compressed form prior to the diagnosis and archived for permanent records using lossy compression. For such a scheme, it is very desirable to have a single system

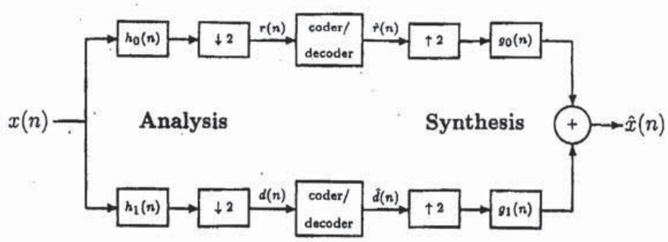


Figure 2.1: Block Diagram of a Wavelet Analysis/Synthesis System

that can perform both lossy and lossless compression. Moreover, with an *embedded* compression system, i.e., compressed data is in a visually important order, lossy compression can be performed by a simple truncation of the compressed bit stream.

In section 2 the basics of wavelet decomposition are explained, and reversible wavelets are defined. In section 3, two reversible wavelet transforms are described in detail. The first one, the S-transform is used to make the definitions easy to comprehend, while the second, the RTS-transform, is the suggested transform. Section 4 describes the embedded entropy coding, including a brief description of Shapiro's Zerotree, and Horizon context model. In section 5, the implementation of CREW in software and hardware is discussed. Section 6 contains some experimental results, comparing the performance of CREW with other system.

## 2 Wavelet Decomposition of Digital Signals

A wavelet transform[1] is defined by a pair of FIR analysis filters $h_0(n), h_1(n)$, and synthesis filters $g_0(n), g_1(n)$. The filters $h_0$ and $g_0$ are the low-pass and $h_1$ and $g_1$ are the high-pass. For an input signal, $x(n)$, the filters $h_0$ and $h_1$ are applied and the results are decimated by 2 (critically subsampled) to generate the transform signals $r(n)$ and $d(n)$, so-called the *reference* and the *detail* signals (analysis part of Figure 2.1). In the synthesis part the transformed signals are upsampled by 2 (a zero is inserted after every term) and then passed through the synthesis filters. Coefficients of the reference signal $r(n)$ are processed through the low-pass synthesis $g_0$ and the coefficients of the detail signal $d(n)$ through the high-pass synthesis filter $g_1$ (synthesis part of Figure 2.1). In this paper we are only interested in *quadrature mirror filters*[2], i.e., the synthesis filters are defined in terms of the analysis filters as follows:

$$\begin{cases} g_0(n) &= (-1)^n h_1(n) \\ g_1(n) &= -(-1)^n h_0(n). \end{cases}$$

The coder/decoder blocks contain all the processing in the transformed domain, e.g., quantization, coding etc. The filters can be recursively applied to the reference and detail signal. Of special interest are the *pyramidal* systems, in which the filters are

[1] For basic wavelet transformation we adopt the terminology and notations of [VBL94].

[2] For details and extensive references on QMF systems cf. [SA90].

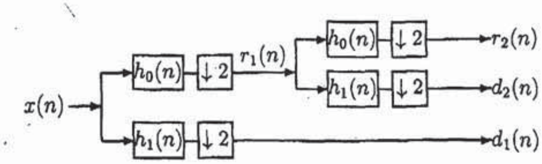


Figure 2.2: Block Diagram of a two-level Pyramidal Transform

recursively applied only to the reference signal. Figure 2.2 shows the block diagram of a two-level (the filters applied twice) pyramidal system.

**Definition 1: Exact Reconstruction Systems:** The system in Figure 2.1 is called exact reconstruction if the signals, $x(n)$ and $\hat{x}(n)$ are identical up to a multiplicative constant and a delay term [LGT88].

**Definition 2: Efficient Reversible Systems:** A reversible system is an implementation of an exact-reconstruction system, in integer arithmetic, so that a signal with integer coefficients can be losslessly recovered. An efficient reversible system is a reversible system with transform matrix[3] of determinant $\approx \pm 1$.

The construction of reversible transforms is by no means difficult. Given enough precision, any exact reconstruction transform can be made reversible. The challenge is to construct *"efficient"* reversible transforms. Intuitively, efficient means that a reasonable and practical entropy coder can efficiently encode the coefficients. While efficiency criteria might be specific to a particular encoder, in general, systems with determinant $= \pm 1$ are efficient. Like any transform which is used for coding, good energy compaction is also a primary factor. The goal is to design a single system which performs well in both lossy and lossless modes. The concepts will be made more clear in the next section with two examples.

## 3 Two Examples of Reversible Transforms

### 3.1 Exact Reconstruction

**Example 1: Hadamard Transform:** In normalized form it has the following filter coefficients:

$$\begin{cases} h_0 &= \frac{1}{\sqrt{2}}(1,1) \\ h_1 &= \frac{1}{\sqrt{2}}(1,-1), \end{cases} \tag{3.1}$$

which is clearly an exact-reconstruction transform.

**Example 2: TS-transform (Two-Six-transform):** For the origin and qualities of the TS-transform cf. Remark 1 below. TS-transform is defined by the following filter coefficients:

$$\begin{cases} h_0 &= \frac{1}{\sqrt{2}}(1,1) \\ h_1 &= \frac{1}{8\sqrt{2}}(-1,-1,8,-8,1,1). \end{cases} \tag{3.2}$$

This is also an exact-reconstruction transform.

[3]For the definition of the transform matrix cf. [SA90].

### 3.2 Basic Reversible Versions

The reversible transforms, in general, are non-linear. Hence they will be defined as expressions. However, the linear system approximation which is useful for evaluation will also be given. Recall that for the input signal $x(n)$, $r(n)$ and $d(n)$ are the reference and the detail signal respectively. The reversible S-transform will be used as a simple example before explaining the reversible TS-transform.

**Example 3: S-transform:** An efficient reversible version of the Hadamard transform, Example 3, known as the S-transform [SAAJ91, SP93] is defined as follows:

$$\left\{ \begin{array}{lcl} r(0) & = & \lfloor \frac{x(0)+x(1)}{2} \rfloor \\ d(0) & = & x(0) - x(1). \end{array} \right. \tag{3.3}$$

Notice that this is an approximation to the linear transform (with determinant $= -1$.) The floor function in the definition of $r(0)$ is the source of non-linearity.

$$\left\{ \begin{array}{lcl} r(0) & = & \frac{x(0)+x(1)}{2} \\ d(0) & = & x(0) - x(1). \end{array} \right. \tag{3.4}$$

A constructive proof for the reversibility of the S-transform is the inverse transform:

$$\left\{ \begin{array}{lcl} x(0) & = & r(0) + \lfloor \frac{d(0)+1}{2} \rfloor \\ x(1) & = & r(0) - \lfloor \frac{d(0)}{2} \rfloor. \end{array} \right. \tag{3.5}$$

The idea behind the reversibility of the S-transform is the observation of two facts. One, knowledge of the sum and the difference of two integers are sufficient to recover the numbers. Two, the sum and the difference have the same parity, i.e., they share the same least significant bit. Hence the integer division by 2 (or a shift right by 1) in Eq. 3.3, eliminates a redundant least significant bit. The S-transform, where redundant information can be detected and easily eliminated, is an example of efficient reversible transform.

**Example 4: RTS-Transform (Reversible TS-transform):** An efficient reversible version of the TS-transform, which we call RTS-transform is defined as follows:

$$\left\{ \begin{array}{lcl} r(0) & = & \lfloor \frac{x(0)+x(1)}{2} \rfloor \\ r(1) & = & \lfloor \frac{x(2)+x(3)}{2} \rfloor \\ r(2) & = & \lfloor \frac{x(4)+x(5)}{2} \rfloor \\ \vdots & & \\ d(0) & = & \lfloor \frac{-\lfloor \frac{x(0)+x(1)}{2} \rfloor + 4(x(2)-x(3)) + \lfloor \frac{x(4)+x(5)}{2} \rfloor}{4} \rfloor \\ \vdots & & \end{array} \right. \tag{3.6}$$

Notice that this is an approximation to the following linear version (with determinant $= -1$) of the TS-transform:

$$\begin{cases} r(0) &= \frac{x(0)+x(1)}{2} \\ r(1) &= \frac{x(2)+x(3)}{2} \\ r(2) &= \frac{x(4)+x(5)}{2} \\ \vdots & \\ d(0) &= \frac{-x(0)-x(1)+8(x(2)-x(3))+x(4)+x(5)}{8} \\ \vdots & \end{cases} \tag{3.7}$$

The proof that the RTS-transform is reversible is quite simple. We show how to recover $x(2)$ and $x(3)$ from $r(0), r(1), r(2)$ and $d(0)$, the recovery of other samples are similar. Notice the expression for $d(0)$ in Eq. 3.6, can be written as,

$$d(0) = \lfloor \frac{-r(0) + 4(x(2) - x(3)) + r(2)}{4} \rfloor.$$

From this it follows that:

$$x(2) - x(3) = d(0) - \lfloor (-r(0) + r(2))/4 \rfloor.$$

Hence $x(2) - x(3)$ is completely known. This, combined with $r(1) = \lfloor (x(2) + x(3))/2 \rfloor$ in Eq. 3.6, and the use of inverse S-transform Eq. 3.5, leads to the recovery of $x(2)$ and $x(3)$.

**Remark 1:** Le Gall and Tabatabai in [LGT88] use a design procedure based on the factorization of a product filter into two linear phase low-pass components. These correspond to the low-pass analysis and synthesis filters. By using the quadrature mirror properties the high-pass filters are derived. In their most important example, which is by now classical, the following product filter is factored:

$$P(Z) = 1/16(1 + Z^{-1})^3(-1 + 3Z^{-1} + 3Z^{-2} - Z^{-3}).$$

Two factorizations are given in [LGT88],

$$\begin{cases} P(Z) &= [1/4(1 + Z^{-1})^3] \times [1/4(-1 + 3Z^{-1} + 3Z^{-2} - Z^{-3})] \\ P(Z) &= [1/2(1 + Z^{-1})^2] \times [1/8(1 + Z^{-1})(-1 + 3Z^{-1} + 3Z^{-2} - Z^{-3})]. \end{cases} \tag{3.8}$$

A third factorization,

$$P(Z) = [1/2(1 + Z^{-1})] \times [1/8(1 + Z^{-1})^2(-1 + 3Z^{-1} + 3Z^{-2} - Z^{-3})],$$

which was not mentioned in that paper, is, in fact, the rational version of the TS-transform (Eq. 3.7.) Speck in [Spe93] considers and analyses this third factorization. In [VBL94] the normalized version Eq. 3.2, is evaluated together with several thousand other wavelet transforms, and is rated as one of the overall best.

**Remark 2:** In both the S-transform and the RTS-transform, the reference signal $r(n)$ has the same range of values as the input signal $x(n)$, e.g., if the range of $x(n)$ is from 0 to 255 the same is true about $r(n)$. This property is especially important in a pyramidal system where the reference signal is successively decomposed.

**Remark 3:** Both normalized Hadamard transform and the normalized TS-transform have especially simple implementations in two dimensions, the domain of digital images. Recall that in the separable two dimensional wavelet transform the image is decomposed into four blocks, the so-called LL, LH, HL, and HH [Sha93]. Each L corresponds to an application of the low-pass filter $h_0$, and each H corresponds to an application of the high-pass filter $h_1$. If we denote $h_0$, and $h_1$ to be the low-pass and the high-pass filters of the normalized TS-transform, Eq. 3.2, and similarly $h_0^r$, and $h_1^r$ for the rational version, Eq. 3.7, then

$$\begin{cases} h_0^r &= \frac{\sqrt{2}}{2} h_0 \\ h_1^r &= \sqrt{2} h_1. \end{cases}$$

Therefore the LL, and HH components of the rational TS-transform, Eq. 3.7 are 1/2 and 2 times the corresponding components of the normalized TS-transform. Moreover the components LH, HL are identical. This leads to an efficient implementation of the TS-transform through the rational TS-transform. More to the point for this article is the fact that if the rational TS-transform, Eq. 3.7, is replaced by the the reversible TS-transform (RTS-transform), Eq. 3.6, a very good approximation of the normalized TS-transform is realized, which in addition is reversible. Notice that the scale factors of 1/2 and 2 are especially easy to implement.

## 4 Embedded Entropy Coding

In most transform-based compression systems the coefficients are entropy coded. Of special importance for us are the so-called "embedded" coders. Briefly, an embedded coding is a system in which the coded bit stream is ordered by visual significance or, more accurately, ordered with respect to some error metric cf. also [Sha93]. The embedded order used in this paper is bit-significance in the transform domain, the same as used in [Sha93].

The Zerotree [Sha93] is an efficient embedded coding method of the wavelet coefficients, which takes advantage of the inherent similarity of different bands in the transform domain.

Horizon embedded coding, introduced below, is a spatial-spectral context model which uses the same embedding order, i.e., bit-significance, as the Zerotree. Non-embedded context models have been proposed to encode signed integers which take advantage of spatial correlations of coefficients [Lan91]. In Horizon coding the high correlation of neighboring pixels, in addition to the similarities of different bands, are utilized by context dependent entropy coding. The Horizon context dependent coding is especially attractive for encoding the low order bits, which must be encoded in a lossless system. Moreover the coding can start with the Zerotree, or some other spectral context model, and change to Horizon after any number of bit-planes.

### 4.1 Zerotree

The most important part of the Zerotree embedded coding of the wavelet coefficients is a prediction method which, according to Shapiro, is based on "the basic hypothesis - if a coefficient at a coarse scale is insignificant with respect to a threshold then all of its descendants are also insignificant." The descendants are defined with respect to a tree structure defined on the wavelet coefficients, which takes advantage of the similarity of the bands at different resolutions [Sha93]. The other part of the Zerotree embedded coding is a bit-significance embedding method to encode signed integers. In the so-called *dominant pass* the integers in sign-magnitude form, are encoded one bit at a time from the MSB to LSB, by the prediction method of Zerotree, until the first "on" bit is detected. The sign is encoded at this time, which is the logical embedding order of the sign bit. The remaining of the bits are encoded without the Zerotree prediction, in the so-called *subordinate pass.*

### 4.2 Horizon: A Spatial-Spectral Context Model

Horizon context model addresses the bit-significance embedded encoding of the wavelet coefficients by a binary entropy coder. There are three basic contexts designed for embedded encoding of signed integers. Hence these can be described independently of any wavelet system. Recall that in the bit-significance embedded coding of signed integers, the sign bit is encoded with the first "on" bit (starting from the MSB). Therefore, prior to the occurrence of the sign bit, the set of events consists of 0,1,-1. After the sign bit is encoded the set of events is 0,1. The first context is used to encode the event of zero bit vs. non-zero bit when the sign bit has not yet been encoded. The second context is used to encode the sign bit if the previous event was non-zero. Context three is used to encode the zero bit vs. the one bit if the sign bit is already encoded. In the terminology of Shapiro the first two contexts are used during the dominant pass and the third context during the subordinate pass. Notice also that the second context is used at most once for every integer.

The basic contexts of the Horizon model can be extended for the coding of the wavelet coefficients. Briefly, the fact that the wavelet transform is localized both in space and frequency makes it possible to use regional contexts in the same band as well as between bands contexts. The regional contexts in the same band are similar to the JBIG or the lossless JPEG systems. The between band spectral contexts can use modeling such as the tree structure used in Zerotree, or use the so-called similar coefficients from the other bands, provided that the system stays causal.

## 5 Computation in Software and Hardware

CREW is suitable for implementation in both software and hardware. It can be performed in two passes where the first pass generates all the transform coefficients and the second pass embeds and encodes them, or it can be performed in a unique one pass mode with memory management.

The filters chosen in CREW are easy to implement for both encoding and decoding. The implementation of the forward transform is immediate from Eq. 3.6. Figure 5.1 shows a hardware implementation of the inverse transform (the more interesting case). As in the forward case, four additions/subtractions are required. A total of four multiplications/divisions are hardwired shifts in hardware and three shift instructions

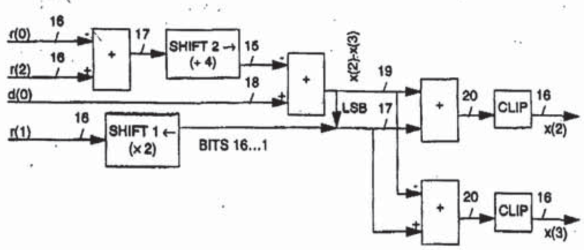


Figure 5.1: An Implementation of the Inverse RTS-Transform

in software. After $2 \times r(1)$ is computed (bits 16...1), its least significant bit (LSB) is taken from the computed value of x(2)-x(3). This operation has zero gate cost in hardware and is two logic operations in software.

For lossless decompression, the clip operation in Figure 5.1 simply shifts its input right by one (divides by two) and may drop (or otherwise ignore) the three most significant bits. In the lossy case, where quantization can cause the reconstructed value to be out of range, the three most significant bits must be checked and out of range results must be changed to the minimum or maximum allowable value.

For images where a full frame can fit in memory allowing two pass implementation, memory/data flow management is not a difficult issue. Even for $1024 \times 1024$ 16 bit medical images (2 Mbytes in size), requiring a full frame buffer is probably reasonable. However larger images (for example A4, 400 DPI 4-color images are about 50 Mbytes), performing the wavelet transform with a limited amount of line buffer memory is desirable. A one pass method reduces the memory required by about a factor of 100 compared to using a full frame buffer for this example.

Because only the high pass filter is overlapped, the largest filter support region is defined by a cascade of low pass filters followed by a high pass filter. For a four level decomposition, the largest support region is $(2^3 \times 6) \times (2^3 \times 6) = 48 \times 48$ pixels, as shown in Figure 5.2. Note that for computational efficiency, redundant calculations due to overlap are done only once. Thus, only $16 \times 16$ new pixels are used in calculations for each region.

At the moment CREW uses the binary adaptive arithmetic coder known as Q-coder.

## 6 Experimental Results

Two sets of images are used for experimentation, a class of $512 \times 512$ USC gray scale 8-bit deep images, and a class of medical images of different modalities. Medical images "cr", "dsa", "xray" are $1024 \times 1024$, and are 10 bits deep. Images "ct" and "mri" are $512 \times 512$ and are 12 bits deep. Tables 1 and 2 are the lossless results, for the medical and the USC images respectively. The results related to the USC images (Table 2) are compared with JPEG lossless (with QM-coder), and bit-plane JBIG of the gray coded image [AT94]. The results on medical imaging (Table 1) are compared with DPCM which uses three neighbor pixels for prediction and Huffman

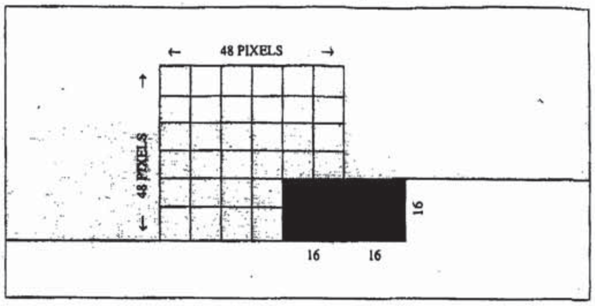

Figure 5.2: Image and Coefficients in line Buffer; 4-level pyramid

codes the prediction errors. Table 3 is the results of lossy compression of the USC images. The MSE results are compared with JPEG lossy with arithmetic coding at the same compression ratio. JPEG lossy with arithmetic coding was chosen because of its superior rate/distortion over baseline JPEG. In each case JPEG is used roughly at low, and high compression ratio. The truncation of CREW's compressed bit stream was used to decompress at exactly the same ratio. As can be observed state of the art lossy and lossless performance is achieved for all cases.

## 7 Conclusion

CREW is essentially an exact reconstruction transform scheme with good energy compaction. Moreover, it is specially designed so that the transform coefficients have small redundancy and are easy to entropy code. Since CREW has good compression efficiency at any level of quantization it is well suited for embedded coding. Embedded coding of the coefficients makes the quantization level a function of the length of the coded stream. Hence, quantization is performed with truncation. Without truncation the image is recovered losslessly.

Horizon context dependent entropy coding, together with the Shapiro's Zerotree is used to achieve such a system. In addition to the lossless and lossy state-of-the-art compression performance with a single system, efficient software and hardware implementations are also possible. Other features of CREW are multi-resolution and progressive capabilities.

**TABLE 1**
**Compression ratio of lossless compression of Medical images.**

| compression method | cr | ct | dsa | mri | xray |
|---|---|---|---|---|---|
| CREW | 2.43 | 5.26 | 2.89 | 3.23 | 2.58 |
| DPCM | 2.34 | 3.95 | 2.64 | 2.86 | 2.41 |
| JBIG | 2.25 | 4.92 | 2.72 | 2.68 | 2.46 |

**TABLE 2**
**Compression ratio of lossless compression of USC images.**

| compression method | couple | crowd | lax | lena | man | woman1 | woman2 | average |
|---|---|---|---|---|---|---|---|---|
| CREW | 1.63 | 1.88 | 1.34 | 1.84 | 1.69 | 1.66 | 2.37 | 1.73 |
| JPEG | 1.54 | 1.87 | 1.31 | 1.72 | 1.64 | 1.58 | 2.28 | 1.66 |
| JBIG | 1.53 | 1.75 | 1.31 | 1.69 | 1.59 | 1.58 | 2.10 | 1.62 |

**TABLE 3**
**Mean Square Error (MSE) of lossy compression of USC images.**

| compression method | | couple | crowd | lax | lena | man | woman1 | woman2 |
|---|---|---|---|---|---|---|---|---|
| low ratio | CREW | 23.69 | 17.29 | 54.72 | 14.69 | 22.20 | 22.41 | 6.59 |
| | JPEG | 29.62 | 20.10 | 87.80 | 17.08 | 29.98 | 33.08 | 6.55 |
| high ratio | CREW | 42.17 | 30.65 | 99.70 | 21.15 | 40.08 | 38.30 | 9.73 |
| | JPEG | 49.86 | 36.10 | 137.00 | 27.74 | 48.84 | 52.29 | 11.32 |

**References**


[AT94] R. Arps and T. Truong. Comparison of international standards for lossless still image compression. *Proceedings of the IEEE*, 82(6):889–899, June 1994.

[Lan91] G. Langdon. Sunset: a hardware algorithm for lossless compression of gray scale images. In *Medical Imaging V: Image Capture, Formatting, and Display*, volume 1444, pages 272–282. SPIE, March 1991.

[LGT88] D. Le Gall and A. Tabatabai. Sub-band coding of digital images using symmetric short kernel filters and arithmetic coding techniques. In *International Conference on Acoustics, Speach and Signal Processing*, pages 761–765, New York, 1988. IEEE.

[SA90] E. Simoncelli and E. Adelson. Subband Transforms. In J. Woods, editor, *Subband Coding*, chapter 4. Kluwer Academic Publishers, Norwell, Massachusetts, 1990.

[SAAJ91] I. Shah, O. Akiwumi-Assani, and B. Johnson. A chip set for lossless image compression. *IEEE Journal Solid-State Circuits*, 26(3):237–244, 1991.

[Sha93] J. Shapiro. An embedded hierarchical image coder using zerotrees of wavelet coefficients. *Proc. IEEE Data Compression Conference*, pages 214–223, 1993.

[SP93] A. Said and W. Pearlman. Reversible image compression via multiresolution representation and predictive coding. In *Visual Communications and Image Processing*, volume 2094, pages 664–674. SPIE, November 1993.

[Spe93] D. Speck. Low-complexity subband coding for image compression. Technical Report UCSC-CRL-94-03, Computer Research Laboratory, University of California at Santa Cruz, May 1993.

[VBL94] J. Villasenor, B. Belzer, and J. Liao. Filter evaluation and selection in wavelet image compression. In *Data Compression Conference*, Snowbird, Utah, 1994. IEEE.

08/633,896

UNITED STATES DEPARTMENT OF COMMERCE
Patent and Trademark Office
Address: COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

| SERIAL NUMBER | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. |
|---|---|---|---|
| 08/633,896 | 04/17/96 | MEANY J | 8190-43 |

26M2/0724

GUY R GOSNELL
BELL SELTZER PARK AND GIBSON
POST OFFICE DRAWER 34009
CHARLOTTE NC 28234

| EXAMINER | |
|---|---|
| TADAYON, B | |
| ART UNIT | PAPER NUMBER |
| 2616 | 5 |

DATE MAILED: 07/24/97

This is a communication from the examiner in charge of your application.
COMMISSIONER OF PATENTS AND TRADEMARKS

☒ This application has been examined ☐ Responsive to communication filed on ________ ☐ This action is made final.

A shortened statutory period for response to this action is set to expire 3 month(s), ~~________~~ days from the date of this letter.
Failure to respond within the period for response will cause the application to become abandoned. 35 U.S.C. 133

**Part I THE FOLLOWING ATTACHMENT(S) ARE PART OF THIS ACTION:**

1. ☒ Notice of References Cited by Examiner, PTO-892.
2. ☒ Notice of Draftsman's Patent Drawing Review, PTO-948.
3. ☒ Notice of Art Cited by Applicant, PTO-1449.
4. ☐ Notice of Informal Patent Application, PTO-152.
5. ☐ Information on How to Effect Drawing Changes, PTO-1474.
6. ☐ ________

**Part II SUMMARY OF ACTION**

1. ☒ Claims 1-30 are pending in the application.

   Of the above, claims ________ are withdrawn from consideration.

2. ☐ Claims ________ have been cancelled.
3. ☐ Claims ________ are allowed.
4. ☒ Claims 1-3, 5-14, 16-29 are rejected.
5. ☒ Claims 4, 15, 30 are objected to.
6. ☐ Claims ________ are subject to restriction or election requirement.
7. ☐ This application has been filed with informal drawings under 37 C.F.R. 1.85 which are acceptable for examination purposes.
8. ☐ Formal drawings are required in response to this Office action.
9. ☐ The corrected or substitute drawings have been received on ________. Under 37 C.F.R. 1.84 these drawings are ☐ acceptable; ☐ not acceptable (see explanation or Notice of Draftsman's Patent Drawing Review, PTO-948).
10. ☐ The proposed additional or substitute sheet(s) of drawings, filed on ________, has (have) been ☐ approved by the examiner; ☐ disapproved by the examiner (see explanation).
11. ☐ The proposed drawing correction, filed ________, has been ☐ approved; ☐ disapproved (see explanation).
12. ☐ Acknowledgement is made of the claim for priority under 35 U.S.C. 119. The certified copy has ☐ been received ☐ not been received ☐ been filed in parent application, serial no. ________; filed on ________.
13. ☐ Since this application apppears to be in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under Ex parte Quayle, 1935 C.D. 11; 453 O.G. 213.
14. ☐ Other

EXAMINER'S ACTION

PTOL-326 (Rev. 2/93)

## Part III DETAILED ACTION

### Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. § 103 which forms the basis for all obviousness rejections set forth in this Office action:

> A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.
>
> Subject matter developed by another person, which qualifies as prior art only under subsection (f) or (g) of section 102 of this title, shall not preclude patentability under this section where the subject matter and the claimed invention were, at the time the invention was made, owned by the same person or subject to an obligation of assignment to the same person.

2. Claims 1-3, 5-14, 16-29 are rejected under 35 U.S.C. § 103 as being unpatentable over Seshadri et al.

As to claims 1, 7, 12, 22, 28, Seshadri et al. Teaches an error resilient method of encoding data (column 2 lines 28-56); generating a plurality of code words representative of respective portions of data (column 2 lines 28-56), wherein each code word comprises a first portion and an associated second portion (column 2 lines 28-56), wherein said code word generating step comprises: generating the first portion of each code word, the first portion including information representative of a predetermined characteristic of the associated second portion (column 2 lines 28-56, figure 1, abstract); and generating the second portion of each code word, the second portion including information representative of the respective portion of the data (column 2 lines 28-56, figure 1, abstract); and providing error protection to at least one of the first portions of the plurality of the code words while maintaining any error protection provided to the respective second portion associated with the at least one first portion at a lower level than the error protection provided to the respective first portion (column 2 lines 28-56).

As to claims 2-3, 8-9, 13-14, 23-24, 29, Seshadri et al. Does not specifically teach entropy coding; second portions having predetermined number of characters, and first portions which includes information representative of the predetermined number of characters.

As to claims 5-6, 10-11, 16-17, 25-26, 28, Seshadri et al. Does not specifically teach storing the at least one first portion, and storing the respective second portion; transmitting first portion, and transmitting respective second portion via a second data link.

As to claims 12, 18-19, 20-21, 22, 27, Seshadri et al. Does not specifically teach compressing data; transforming data based on the a predetermined transformation function; quantizing and encoding data; wavelet transform; biorthogonal wavelet transform; transformed coefficients, and coefficients below a threshold; establishing a clipping threshold such that the ratio is at least as great as a predetermined clipping ratio.

However, entropy coding; having predetermined number of characters; storing portions; transmitting portions; compressing data; transforming data based on the a predetermined transformation function; quantizing and encoding data; wavelet transform; biorthogonal wavelet transform; transformed coefficients, and coefficients below a threshold; establishing an adaptive threshold; are all well-known and routinely used in the art. (Official Notice)

It would have been obvious to a person of ordinary skill in the art at the time the invention was made to incorporate these features into the image processing and data encoding. Because entropy coding is used to compress the data; having predetermined number of characters is used in run-length encoding; storing portions is used to gather the data; transmitting portions is used to send the data to the receiver end; compressing/ transforming data/ quantizing/ encoding data/ wavelet transform/ transformed coefficients are used to encode/ compress/ reduce the size of data; an adaptive threshold is used to make the system more flexible and adjustable.

3. Claims 4, 15, 30 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

4. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Nelson et al., Nadan, and Baggen et al. Teach parity/ code word, data enable sequences, and interleaved phases.

5. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dr. Bijan Tadayon whose telephone number is (703) 308-7595. The fax number is (703) 308-9051 or (703) 308-9052.

B.T.
Dr. Bijan Tadayon
July 16, 1997

Leo Boudreau
LEO BOUDREAU
SUPERVISORY PATENT EXAMINER
GROUP 2600

**BEST COPY**

OLD TOP AND BOTTOM EDGES, SN--n—APART AND DISCARD CARBON

| FORM PTO-892 (REV. 2-92) | U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE | SERIAL NO. | GROUP ART UNIT | ATTACHMENT TO PAPER NUMBER |
|---|---|---|---|---|
| NOTICE OF REFERENCES CITED | | 08/633,896 | 2616 | 5 |
| | | APPLICANT(S) MEANY ET AL. | | |

**U.S. PATENT DOCUMENTS**

| * | | DOCUMENT NO. | DATE | NAME | CLASS | SUB-CLASS | FILING DATE IF APPROPRIATE |
|---|---|---|---|---|---|---|---|
| | A | 5289501 | FEB94 | SESHADRI ET AL | 375 | 17 | NOV91 |
| | B | 5168493 | DEC92 | NELSON ET AL. | 370 | 84 | NOV91 |
| | C | 5321750 | JUN94 | NADAN | 380 | 20 | MAY92 |
| | D | 5471486 | NOV95 | BAGGEN ET AL | 371 | 37.1 | FEB95 |
| | E | | | | | | |
| | F | | | | | | |
| | G | | | | | | |
| | H | | | | | | |
| | I | | | | | | |
| | J | | | | | | |
| | K | | | | | | |

**FOREIGN PATENT DOCUMENTS**

| * | | DOCUMENT NO. | DATE | COUNTRY | NAME | CLASS | SUB-CLASS | PERTINENT SHTS. DWG | PP. SPEC. |
|---|---|---|---|---|---|---|---|---|---|
| | L | | | | | | | | |
| | M | | | | | | | | |
| | N | | | | | | | | |
| | O | | | | | | | | |
| | P | | | | | | | | |
| | Q | | | | | | | | |

**OTHER REFERENCES (Including Author, Title, Date, Pertinent Pages, Etc.)**

| | |
|---|---|
| R | |
| S | |
| T | |
| U | |

| EXAMINER | DATE |
|---|---|
| DR. B. Tadayon | 16JUL97 |

* A copy of this reference is not being furnished with this office action.
(See Manual of Patent Examining Procedure, section 707.05 (a).)

Form PTO 948 (Rev. 10-94) U.S. DEPARTMENT OF COMMERCE - Patent and Trademark Office Application No 633896

# NOTICE OF DRAFTSPERSON'S PATENT DRAWING REVIEW

PTO Draftpersons review all originally filed drawings regardless of whether they are designated as formal or informal. Additionally, patent Examiners will review the drawings for compliance with the regulations. Direct telephone inquiries concerning this review to the Drawing Review Branch, 703-305-8404.

The drawings filed (insert date) 4/17/96, are
A. ✓ not objected to by the Draftsperson under 37 CFR 1.84 or 1.152.
B.____ objected to by the Draftsperson under 37 CFR 1.84 or 1.152 as indicated below. The Examiner will require submission of new, corrected drawings when necessary. Corrected drawings must be submitted according to the instructions on the back of this Notice.

1. DRAWINGS. 37 CFR 1.84(a): Acceptable categories of drawings: Black ink. Color.
   - __ Not black solid lines. Fig(s)____
   - __ Color drawings are not acceptable until petition is granted. Fig(s)____
2. PHOTOGRAPHS. 37 CFR 1.84(b)
   - __ Photographs are not acceptable until petition is granted. Fig(s)____
   - __ Photographs not properly mounted (must use brystol board or photographic double-weight paper). Fig(s)____
   - __ Poor quality (half-tone). Fig(s)____
3. GRAPHIC FORMS. 37 CFR 1.84 (d)
   - __ Chemical or mathematical formula not labeled as separate figure. Fig(s)____
   - __ Group of waveforms not presented as a single figure, using common vertical axis with time extending along horizontal axis. Fig(s)____
   - __ Individuals waveform not identified with a separate letter designation adjacent to the vertical axis. Fig(s)____
4. TYPE OF PAPER. 37 CFR 1.84(e)
   - __ Paper not flexible, strong, white, smooth, nonshiny, and durable. Sheet(s)____
   - __ Erasures, alterations, overwritings, interlineations, cracks, creases, and folds copy machine marks not accepted. Fig(s)____
   - __ Mylar, velum paper is not acceptable (too thin). Fig(s)____
5. SIZE OF PAPER. 37 CFR 1.84(f): Acceptable sizes:
   - 21.6 cm. by 35.6 cm. (8 1/2 by 14 inches)
   - 21.6 cm. by 33.1 cm. (8 1/2 by 13 inches)
   - 21.6 cm. by 27.9 cm. (8 1/2 by 11 inches)
   - 21.0 cm. by 29.7 cm. (DIN size A4)
   - __ All drawing sheets not the same size. Sheet(s)____
   - __ Drawing sheet not an acceptable size. Sheet(s)____
6. MARGINS. 37 CFR 1.84(g): Acceptable margins:

Paper size

| | 21.6 cm. X 35.6 cm. (8 1/2 X 14 inches) | 21.6 cm X 33.1 cm. (8 1/2 X 13 inches) | 21.6 cm. X 27.9 cm. (8 1/2 X 11 inches) | 21.0 cm. X 29.7 cm. (DIN Size A4) |
|---|---|---|---|---|
| T | 5.1 cm. (2") | 2.5 cm. (1") | 2.5 cm. (1") | 2.5cm. |
| L | .64 cm. (1/4") | .64 cm. (1/4") | .64 cm. (1/4") | 2.5 cm. |
| R | .64 cm. (1/4") | .64 cm. (1/4") | .64 cm. (1/4") | 1.5 cm. |
| B | .64 cm. (1/4") | .64 cm. (1/4") | .64 cm. (1/4") | 1.0 cm. |

Margins do not conform to chart above.
Sheet(s)____
__Top (T) __ Left (L) __ Right (R) __ Bottom (B)

7. VIEWS. 37 CFR 1.84(h)
   REMINDER: Specification may require revision to correspond to drawing changes.
   - __ All views not grouped together. Fig(s)____
   - __ Views connected by projection lines or lead lines. Fig(s)____
   
   Partial views. 37 CFR 1.84(h) 2
   - __ View and enlarged view not labled separatly or properly. Fig(s)____
   
   Sectional views. 37 CFR 1.84 (h) 3
   - __ Hatching not indicated for sectional portions of an object. Fig(s)____
   - __ Cross section not drawn same as view with parts in cross section with regularly spaced parallel oblique strokes. Fig(s)____
8. ARRANGEMENT OF VIEWS. 37 CFR 1.84(i)
   - __ Words do not appear on a horizontal, left-to-right fashion when page is either upright or turned so that the top becomes the right side, except for graphs. Fig(s)____
9. SCALE. 37 CFR 1.84(k)
   - __ Scale not large enough to show mechanism with crowding when drawing is reduced in size to two-thirds in reproduction. Fig(s)____
   - __ Indication such as "actual size" or scale 1/2" not permitted. Fig(s)____
10. CHARACTER OF LINES, NUMBERS, & LETTERS. 37 CFR 1.84(l)
    - __ Lines, numbers & letters not uniformly thick and well defined, clean, durable, and black (except for color drawings). Fig(s)____
11. SHADING. 37 CFR 1.84(m)
    - __ Solid black shading areas not permitted. Fig(s)____
    - __ Shade lines, pale, rough and blurred. Fig(s)____
12. NUMBERS, LETTERS, & REFERENCE CHARACTERS. 37 CFR 1.84(p)
    - __ Numbers and reference characters not plain and legible. 37 CFR 1.84(p)(l) Fig(s)____
    - __ Numbers and reference characters not oriented in same direction as the view. 37 CFR 1.84(p)(l) Fig(s)____
    - __ English alphabet not used. 37 CFR 1.84(p)(2) Fig(s)____
    - __ Numbers, letters, and reference characters do not measure at least .32 cm. (1/8 inch) in height. 37 CFR(p)(3) Fig(s)____
13. LEAD LINES. 37 CFR 1.84(q)
    - __ Lead lines cross each other. Fig(s)____
    - __ Lead lines missing. Fig(s)____
14. NUMBERING OF SHEETS OF DRAWINGS. 37 CFR 1.84(t)
    - __ Sheets not numbered consecutively, and in Arabic numerals, beginning with number 1. Sheet(s)____
15. NUMBER OF VIEWS. 37 CFR 1.84(u)
    - __ Views not numbered consecutively, and in Arabic numerals, beginning with number 1. Fig(s)____
    - __ View numbers not preceded by the abbreviation Fig. Fig(s)____
16. CORRECTIONS. 37 CFR 1.84(w)
    - __ Corrections not made from prior PTO-948. Fig(s)____
17. DESIGN DRAWING. 37 CFR 1.152
    - __ Surface shading shown not appropriate. Fig(s)____
    - __ Solid black shading not used for color contrast. Fig(s)____

COMMENTS:

ATTACHMENT TO PAPER NO. REVIEWER [illegible signature] DATE 8/12/96

OIF JAN 26 1998 PATENT & TRADEMARK OFFICE JC10

Attorney's Docket No. 8190-43 PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re: Meany and Martens
Serial No.: 08/633,896 Group Art Unit: 2616
Filed: April 17, 1996 Examiner: B. Tadayon
For: ERROR RESILIENT METHOD AND APPARATUS FOR ENTROPY CODING

January 23, 1998

Assistant Commissioner for Patents
Washington, DC 20231

## AMENDMENT

Sir:

In response to the Official Action dated July 24, 1997, please amend the above-identified application as follows:

IN THE SPECIFICATION:

Page 1, line 8, between "methods" and "apparatus", please insert --and--.

Page 29, line 35, please delete "5A", and insert --5B-- therefore.

Page 30, line 6, please delete "5A", and insert --5B-- therefore.

IN THE CLAIMS:

Please amend independent Claim 1, 7, 12, 22 and 28 as follows:

1. (Amended) An error resilient method of encoding data comprising the steps of:

generating a plurality of code words representative of respective portions of the data, wherein each code word comprises a first portion and an associated second portion, and wherein said code word generating step comprises the steps of:

a1

49

In re: Meany and Martens
Serial No. 08/633,896
Filed: April 17, 1996
Page 2

generating the first portion of each code word, [the first portion] wherein said first portion generating step comprises the step of including information within the first portion that is representative of a predetermined characteristic of the associated second portion; and

generating the second portion of each code word, [the second portion] wherein said second portion generating step comprises the step of including information within the second portion that is representative of the respective portion of the data; and

providing error protection to at least one of the first portions of the plurality of code words while maintaining any error protection provided to the respective second portion associated with the at least one first portion at a lower level than the error protection provided to the respective first portion.

7. (Amended) A data encoding apparatus comprising:

code word generating means for generating a plurality of code words representative of respective portions of the data, wherein each code word comprises a first portion and an associated second portion, and wherein said code word generating means comprises:

first generating means for generating the first portion of each code word, [the first portion] said first generating means comprising means for including information within the first portion that is representative of a predetermined characteristic of the associated second portion; and

second generating means for generating the second portion of each code word, [the second portion] said second generating means comprising means for including information within the second portion that is representative of the respective portion of the data; and

error protection means for providing error protection to at least one of the first portions of the plurality of code words while maintaining any error protection provided to the respective second portion associated with the at least one first portion at a lower level than the error protection provided to the respective first portion.

12. (Amended) An error resilient method of compressing data comprising the steps of:

transforming the data based upon a predetermined transformation function;

quantizing the transformed data such that the quantized data has fewer unique coefficients than the transformed data; and

encoding the quantized data, said encoding step comprising the steps of:

generating a plurality of code words, representative of respective portions of the data, which have respective first and second portions, wherein [the first portion includes] said code word generating step comprises the steps of including information within the first portion that is representative of a predetermined characteristic of the associated second portion, and [wherein the second portion includes] including information within the second portion that is representative of a respective portion of the data; and

providing error protection to at least one of the first portions of the plurality of code words while maintaining any error protection provided to the respective second portion associated with the at least one first portion at a lower level than the error protection provided to the respective first portion.

22. (Amended) An error resilient data compression apparatus comprising:

a data transformer for transforming the data based upon a predetermined transformation function;

a data quantizer for quantizing the transformed data such that the quantized data has fewer unique coefficients than the transformed data; and

a data encoder for encoding the quantized data, said data encoder comprising:

code word generating means for generating a plurality of code words, representative of respective portions of the data, which have respective first and second portions, wherein said code word generating means comprises means for including [the first portion includes] information within the first portion that is representative of a predetermined characteristic of the associated second portion, and means for including [wherein the second portion includes] information within the second portion that is representative of a respective portion of the data; and

error protection means for providing error protection to at least one of the first portions of the plurality of code words while maintaining any error protection provided to the respective second portion associated with the at least one first portion at a lower level than the error protection provided to the respective first portion.

28. (Amended) A computer readable memory for storing error resilient encoded data, the computer readable memory comprising:

a storage medium for storing the error resilient encoded data, said storage medium being partitioned into a first error protected data block and a second data block, wherein any error protection provided by said second data block is at a lower level than the error protection provided by said first data block; and

a plurality of code words, representative of respective portions of the original data, which have respective first and second portions, wherein the first portion of each code word includes information representative of a predetermined characteristic of the associated second portion, and wherein the associated second portion of each code word includes information representative of a respective portion of the original data,

wherein at least one of the first portions of the plurality of code words is stored in the first data block of said storage medium such that the at least one first portion is error protected, and wherein the respective second portion associated with the at least one first portion is stored in the second data block of said storage medium such that any error protection provided to the respective second portion associated with the at least one first portion is at a lower level than the error protection provided to the respective first portion.

REMARKS

Applicants would like to thank the Examiner for the thorough review of the present application and for the indication that Claims 4, 15 and 30 define patentable subject matter and would be allowable if rewritten in independent form. Each independent claim, namely, Claims 1, 7, 12, 22 and 28, has been amended to more clearly define the invention, as explained more fully below. The specification has also been amended to correct several obvious informalities that were noted during our review of the specification in the course of preparing the present Amendment. As discussed in detail below, the amended set of claims includes recitations which further patentably distinguish the claimed invention over the cited reference.

The Invention

The claimed invention provides an error resilient method and apparatus for entropy coding data which includes code word generating means for generating a plurality of code words representative of respective items in the data set. Each code word has two portions which we shall hereafter refer to as "fields", namely, a first or prefix field which is susceptible to bit errors, and an associated second or suffix field which is resilient to bit errors. According to the claimed invention, the code words are generated such that a bit error in the prefix field of a code word could result in a potential loss of code word synchronization, while a bit error in the suffix field of a code word shall only affect that particular code word. More specifically, the code words are generated such that a bit error in the suffix field of a code word will not result in a loss of code word synchronization, but the resulting misdecoded value shall, instead, fall within a predetermined range about the correct value. Thus, according to the claimed invention, the error resilient method and apparatus for entropy coding data shall be suitable for use with unequal error protection such that the prefix fields are encoded with a higher level of error protection and the suffix fields are encoded with a lower level of error protection, if any at all.

As claimed, the code word generating means includes prefix generating means and suffix generating means for generating the prefix and suffix fields of each code word, respectively. In particular, the prefix field includes information representative of a predetermined characteristic of the associated suffix field. As defined by dependent Claims 3, 9, 14, 24 and 29, each prefix field preferably includes information representative of the predetermined number of characters, such as bits, which form the associated suffix field of the code word. The prefix field may also include information representative of other characteristics of the associated suffix field, such as the contiguous or consecutive range of coefficient values which the associated suffix field may represent. In addition, each suffix field includes information representative of respective portions of the original data. Consequently, even though the suffix fields are not error protected or are only provided with a relatively low level of error protection, the method and apparatus of the claimed invention can correctly determine the length of the suffix field of a code word even if there should be of one or more bit errors within the said suffix field, provided that the associated prefix field is decoded correctly, i.e., without the occurrence of a bit error. Accordingly, in order to provide a high probability that the prefix field is decoded correctly, the method and apparatus of the claimed invention encodes the prefix field with a higher relative level of error protection.

According to one advantageous embodiment set forth in Claims 12-27 in which the data has been quantized, the quantized coefficients can be characterized using a "histogram" which is a discrete distribution consisting of a number of individual "bins", each of which represent the frequency or probability of occurrence of a quantized coefficient value. In other words, each bin is associated with a particular quantization interval which has as its frequency a count of the number of occurrences of coefficients whose values fall within the associated quantization interval.

According to this embodiment of the error resilient method and apparatus for encoding data of the claimed invention, the prefix field of each code word includes information representative of the number of bits K which form the associated suffix field of the code word. Furthermore, the prefix field can also include information representative of the specific histogram bin within which quantized coefficient value resides. The suffix field will, in turn, identify one particular quantized coefficient value within the respective histogram bin. In aggregate, the prefix and suffix field of each code word shall together include information representative of a specific coefficient valve residing within a specific bin of the quantized coefficient histogram.

In other words, the prefix field of this exemplary embodiment includes the information representative of a set of quantized coefficient values while the suffix field includes the information representative of a specific coefficient value among the set designated by the prefix field. Thus, if the prefix field of a code word is decoded correctly, i.e., without the occurrence of a bit error, the length of the associated suffix field and the range of coefficient values which may be represented by the associated suffix field will be known. As a result, the effects of one or more bit errors on the suffix field will be isolated to a specific code word, thereby limiting such errors to a misdecoded coefficient value which is constrained to that range of values determined by the prefix field, i.e., the range of valves within the respective histogram bin. Accordingly, the error resilient method and apparatus for encoding data according to the claimed invention effectively reduces, if not prevents, catastrophic errors in an efficient manner.

The Amended Claims Are Patentable

Claims 1-3, 5-14 and 16-29 were rejected under 35 U.S.C. §103 as being unpatentable over U.S. Patent No. 5,289,501 to Seshadri et al. Each of the independent claims, namely, Claims 1, 7, 12, 22 and 28, have been amended to further patentably distinguish the claimed invention over the cited reference, as explained in detail below.

The Seshadri '501 patent describes a technique for transmitting information in digital form over fading channels. In order to provide error protection for the transmitted information, the technique described by the Seshadri '501 patent accepts a stream of data that has been subdivided into different classes that merit different levels of error protection. For example, the class of data meriting the highest level of error protection may be the most important data and/or the data most susceptible to error, while the class of data meriting the lowest level of error protection may be the least important data and/or the data least susceptible to errors. Once the different classes have been separately scrambled, each class is redundancy coded using a different, respective channel code. Thus, the technique described by the Seshadri '501 patent provides unequal error protection to the different classes of data. Following encoding of the data, the encoded data is modulated prior to being transmitted over free space communication channels to remote digital cellular mobile radio cell sites.

Like the Seshadri '501 patent, the data encoding method and apparatus of the claimed invention utilizes unequal error protection to provide different levels of error protection. With respect to the claimed invention, for example, the first portions of the code words have a higher level of error protection than the