A discrete-time multiresolution theory

Olivier Rioul

A Discrete-Time Multiresolution Theory

Olivier Rioul

Abstract—**Multiresolution analysis and synthesis for discrete-time signals is described in this paper. Concepts of scale and resolution are first reviewed in discrete time. The resulting framework allows one to treat the discrete wavelet transform, octave-band perfect reconstruction filter banks, and pyramid transforms from a unified standpoint. This approach is very close to previous work on multiresolution decomposition of functions of a continuous variable, and the connection between these two approaches is made. We show that they share many mathematical properties such as biorthogonality, orthonormality, and regularity. However, the discrete-time formalism is well suited to practical tasks in digital signal processing and does not require the use of functional spaces as an intermediate step.**

I. Introduction

WAVELET transforms [21], octave-band filter banks [23], [24], and pyramid transforms [3] have been used for different purposes in various fields of signal processing; they are now recognized as different views of a common theory.[1] This paper establishes the links between these techniques for discrete-time signals, focusing on basis decomposition in time rather than frequency decomposition. This topic is close to the one studied in a recent work by Shensa [22] that clarifies the relationship of various discrete-time and continuous-time wavelet transforms using a similar formalism. While [22] focuses on implementation issues of the "continuous wavelet transform" (to be defined below) for signal analysis purposes, we consider only a discrete-time version of the wavelet transform that is more appropriate for coding applications. We assume one-dimensional signals; extension to more dimensions may not be straightforward [25], but the ideas are the same as in one dimension.

There are various types of wavelet transforms, for which the connection to filter banks is more or less obvious. The continuous wavelet transform (CWT) is the first type of wavelet transform defined in the literature [12]. It maps a one-dimensional analog signal $x(t)$ to a set of wavelet coefficients which vary continuously over time b and scale a:

$$\mathrm{CWT}\ (a, b) = a^{-1/2} \int x(t)\, \psi^* \left(\frac{t-b}{a} \right) dt. \quad (1)$$

The author is with the Centre National d'Etudes des Télécommunications, CNET/PAB/RPE/ETP, 92131 Issy-Les-Moulineaux, France.

[1]The reader is referred to [21] and [24] for tutorial descriptions of multiresolution theory and multirate filter banks, and to [8], [15]–[17] for a mathematical study of wavelets.

The "wavelet" functions $\psi((t-b)/a)$ are used to bandpass filter the signal. This can be seen as a kind of time-varying spectral analysis in which scale a plays the role of a local frequency: As a increases, wavelets are stretched and analyze low frequencies, while for small a, contracted wavelets analyze high frequencies. Therefore, the time-frequency extents of the wavelets vary according to a "constant-Q" scheme [21], with various resolutions in time or frequency (multiresolution analysis), whereas techniques based on short-time Fourier transforms analyze the signal at constant resolution [21].

In applications such as signal coding or compression, time-scale parameters (b, a) are sampled so that the signal is represented by wavelet coefficients in an economical manner. A typical choice is $a = 2^j$, $j \in \mathbf{Z}$ (octave by octave computation [21]); in addition, wavelets are shifted in proportion to their temporal extent, i.e., $b = k2^j$. Wavelet "basic functions" become

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \quad (2)$$

and wavelet coefficients are inner products of the signal with the $\psi_{j,k}(t)$'s,

$$w_k^j = \int x(t) \psi_{j,k}^*(t)\, dt. \quad (3)$$

Note that this "sampled" wavelet transform still applies to continuous-time (analog) signals. It has been extensively studied by Meyer [17] and Mallat [15], [16] who built a complete mathematical theory based on "multiresolution spaces" of functions. In fact, some of the ideas behind the theory were already found in image coding and computer vision [1], [3]. Based on this theory, several authors [4], [8], [9], [17] constructed special wavelet prototypes $\psi(t)$ for which perfect reconstruction of the signal $x(t)$ is possible; the synthesis takes the form of a "wavelet series,"

$$x(t) = \sum_{j \in \mathbf{Z}} \sum_{k \in \mathbf{Z}} w_k^j\, \tilde{\psi}_{j,k}(t) \quad (4)$$

where the synthesis wavelets $\tilde{\psi}_{j,k}(t)$ are defined similarly as (2) with wavelet prototype $\tilde{\psi}(t)$. As a special case, orthonormal wavelets are obtained if one has $\tilde{\psi}(t) = \psi(t)$.

In order that the coefficients (3) be computed efficiently, Mallat [15] derived a iterative algorithm that was transposed to compute the reconstruction part. Mallat's direct and inverse algorithms turn out to be exactly an analysis and synthesis octave-band filter bank [23], [24], which is used on discrete-time signals. Filter bank theory is therefore closely related to wavelets. The former has

been previously developed for applications such as subband coding of speech [7] or images [25], and has greatly influenced the latter. In fact, known FIR filter design techniques [23] were independently used by Daubechies [8], [9] to construct finite-length orthonormal wavelets; her construction is based on the classical MAXFLAT digital filters [13] which have been recently related to Lagrangian interpolators [22]. Recently, FIR filter design methods were developed [4], [26] to construct linear phase, "biorthogonal" wavelets for which $\tilde{\psi}(t) \neq \psi(t)$.

Therefore, the connection between wavelet series and filter banks has been extensively studied. However, the "discrete-time side" of wavelet series was mainly considered as a technical step either for the derivation of fast algorithms which rely on the filter bank structure (for further developments see, e.g., [20]) or for the construction of continuous-time wavelet bases *via* filter design [4], [8], [9], [26]. On the other hand, filter banks were studied from the viewpoint of subband decomposition, which masks the time decomposition properties. In this paper, we focus on octave-band perfect reconstruction filter banks seen as multiresolution decomposition of discrete-time signals, performed by a discrete wavelet transform (DWT),[2] to be defined in Section VII.

With the notable exception of Shensa [22], very little can be found on this subject in the literature, in which it is generally admitted that analog wavelets are the interesting objects that underly discrete-time filter banks. Here, we adopt a dual point of view, and consider discrete wavelet transforms as the interesting objects. A discrete-time presentation of multiresolution theory may allow new developments; it is certainly more appropriate for practical tasks in digital signal processing than the continuous-time one. As an example, various image compression schemes were derived based on the wavelet model for analog signals [2], [27], but they are essentially discrete in nature.

This paper focuses on discrete-time signals and develops a discrete-time multiresolution theory where "basis functions" on which the signal is decomposed are discrete-time wavelets. It is self-contained: the reader should not need to have previous knowledge on wavelets, although some rudimentary knowledge on vector spaces is assumed. Note that since the overlap between discrete-time and continuous-time approaches is important, the developments and ideas of this paper are known to wavelet experts, yet they do not appear in the literature as they are presented here.

II. Notations

Throughout this paper we use the following notations. Most of them are classical; in particular, the operator (or matrix) notations appear in other works (see, e.g., [8]) and have been reused by Shensa in [22].

x or $\{x_n\}$ denotes the original signal, which is a discrete-time, complex-valued sequence. Its nth sample will be noted x_n. δ denotes the "pulse" signal, whose samples are $\delta_0 = 1$ and $\delta_n = 0$ if $n \neq 0$.

$\boldsymbol{G}$ and $\boldsymbol{H}$ denote the low-pass and high-pass filtering operator, respectively. The corresponding impulse responses are g_n and h_n. Therefore, $\boldsymbol{G}x$ is a low-pass filtered signal, i.e., the result of the discrete convolution $(\boldsymbol{G}x)_n = \Sigma_k x_k g_{n-k}$. When needed, the sample index of the input is shown, as in $(\boldsymbol{G}x_k)$. $\boldsymbol{I}$ is the identity operator, $\boldsymbol{I}x = x$, of impulse response δ.

$\uparrow x$ and $\downarrow x$ denote the upsampled and downsampled signals, respectively. They are defined by $(\uparrow x)_{2n} = x_n$, $(\uparrow x)_{2n+1} = 0$, and $(\downarrow x)_n = x_{2n}$. Upscaled and downscaled signals, to be defined in Section III, are denoted by $\boldsymbol{G} \uparrow x$ and $\downarrow \boldsymbol{G}' x$, respectively.

The inner product of x with y is $\langle x, y \rangle = \langle x_n, y_n \rangle = \Sigma_n x_n y_n^*$. In particular, $\|x\|^2 = \langle x, x \rangle$ is the squared norm of x.

III. Scaling in Discrete Time

This section defines scale for discrete-time signals. This definition is inspired from the commonly used scale of road maps: given a "real" object, the scale of a representation of this object is the ratio of the length unit of the representation to the corresponding real length. Here the "real object" is an original discrete-time signal $\{x_n\}$, which is at scale 1 by definition, and scale refers to discrete time: a scaled version of $\{x_n\}$ is either 1) "upscaled": a discrete-time signal similar to $\{x_n\}$, but sampled at a higher rate; or 2) "downscaled": a discrete-time signal similar to $\{x_n\}$, but sampled at a lower rate.

Since scale is a relative notion, we focus on the description of scale-changing operators that map a signal into a scaled version of it. Throughout the paper we restrict to changes of scale by an integer power of two. We therefore study two basic scaling operators: 1) "upscaling" operator (by a factor 2): a discrete-time equivalent to the dilation $x(t) \rightarrow x(t/2)$; and 2) "downscaling" operator (by a factor 2): a discrete-time equivalent to the contraction $x(t) \rightarrow x(2t)$.

Obviously, the scale notion is related to multirate systems [6]. For example, a downscaled version of x_n could be its down*sampled* version, $(\downarrow x)_n = x_{2n}$. However, upsampling $(\uparrow x)_n = x_{n/2}$ if n is even, 0 otherwise, is not a good candidate for upscaling, since it inserts a zero between every other sample of x_n, hence x and $\uparrow x$ do not have similar evolutions in time. This can be corrected by further interpolating the samples, as in the following example, illustrated in Fig. 1:

$$\text{Upscaled version of } x_n = \begin{cases} x_{n/2} & \text{if } n \text{ is even} \\ (x_{(n-1)/2} + x_{(n+1)/2})/2 & \text{otherwise.} \end{cases} \quad (5)$$

[2]The acronym DWT is sometimes used to denote wavelet series decomposition of continuous-time signals (3), (4). Here DWT refers to a type of wavelet transform which decomposes discrete-time signals, as in [22]. Although we shall see that the DWT reduces to an octave-band filter bank, we use the wavelet terminology to emphasize its multiresolution (time) decomposition properties, as opposed to the well-known subband (frequency) decomposition properties.

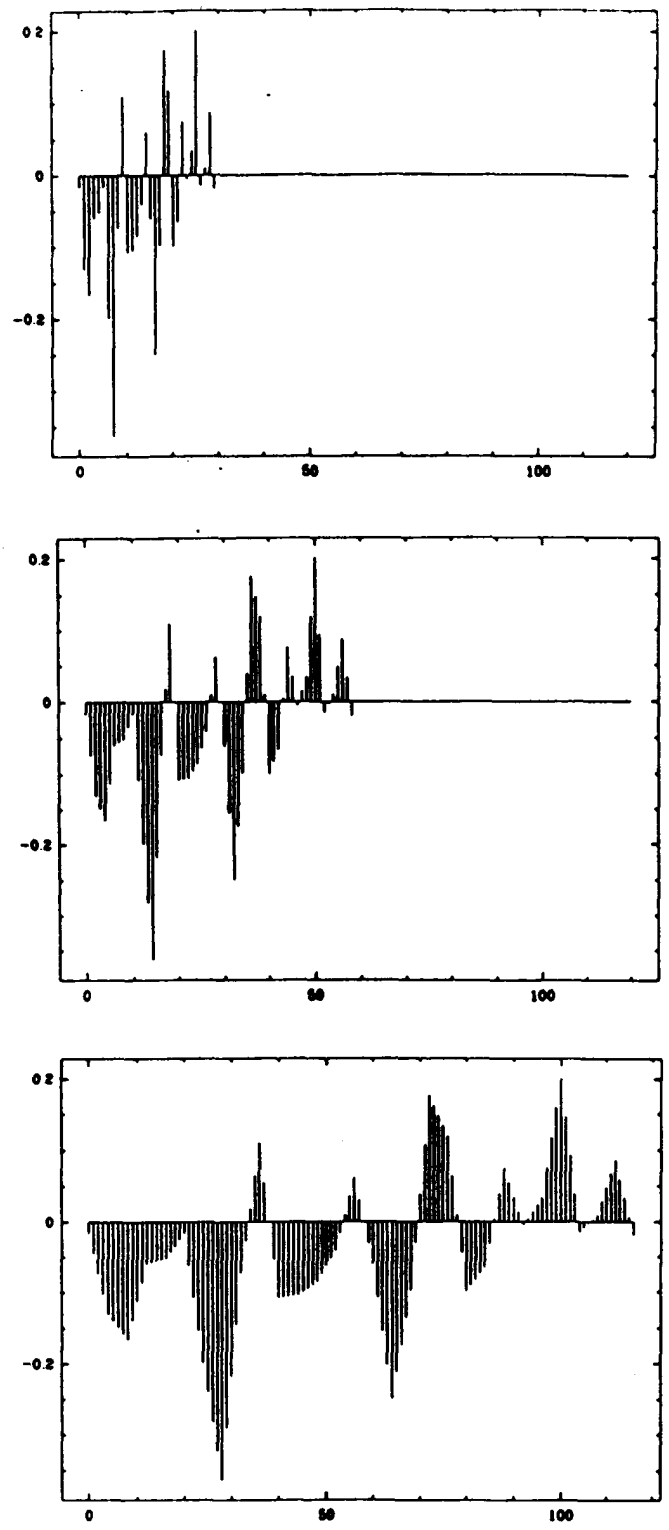


Fig. 1. Two successive upscalings (by a factor 2) of an original discrete-time signal (top of figure). Upscaling is performed by first-order interpolation (5). Upscaled signals are stretched in time, but no resolution is added (see Section V).

In order to determine general expressions for up- and downscaling operators, we need some basic assumptions:

1) Scaling operators are linear.

2) When scaling a signal, time shifts are scaled accordingly. This is

- If y_n is the upscaled version of x_n, then y_{n-2k} is the upscaled version of x_{n-k}.
- If y_n is the downscaled version of x_n, then y_{n-k} is the downscaled version of x_{n-2k}.

3) Scaled versions have similar time evolutions (shape preservation). The third point is difficult to be properly expressed. It is not considered until Section X, where it is connected to the "regularity" property.

The first two assumptions result in the following characterizations, proven in Appendix A:

- The upscaled version of x is of the form $\boldsymbol{G} \uparrow x$, where $\boldsymbol{G} \uparrow$ denotes upsampling followed by filtering with some impulse response $\{g_n\}$:

$$(\boldsymbol{G} \uparrow x)_n = \sum_k x_k g_{n-2k}. \tag{6}$$

- The downscaled version of x is of the form $\downarrow \boldsymbol{G}' x$, where $\downarrow \boldsymbol{G}'$ denotes filtering with some impulse response $\{g'_n\}$, followed by downsampling:

$$(\downarrow \boldsymbol{G}' x)_n = \sum_k x_k g'_{2n-k}. \tag{7}$$

Note that the impulse responses $\{g_n\}$ and $\{g'_n\}$ are not necessarily equal. In the sequel, we assume that they are impulse responses of low-pass filters; intuitively this is required by assumption 3)—Section X gives a theoretical justification of it. As an example, (5) corresponds to a 3-tap upscaling filter $g_{\pm 1} = 0.5$, $g_0 = 1$.

Of course, these up and down "scaling" operators are merely the usual interpolation and decimation operators present in a two-band filter bank [6], [23], [24]. (The corresponding flow graphs are shown in Fig. 2.) However, we keep using the "scale" terminology in the following because it is more appropriate for multiresolution decomposition of discrete-time signals. Note that the operator notation used here is very easily connected to flow-graph implementation; for example, $\boldsymbol{G} \uparrow\downarrow \boldsymbol{G}' x$ means that the input x successively encounters a filter of impulse response g', a downsampler, an upsampler, and finally a filter of impulse response g.

Now, using only these two operators (6), (7), one can compute scaled versions of the original signal $\{x_n\}$ at all dyadic scales $s = 2^{-i}$, where $i \in \mathbf{Z}$. The rule is that the upscaling operator (6) doubles the scale, while the downscaling operator (7) halves it. Since the "scaling filters" $\{g_n\}$ and $\{g'_n\}$ remain fixed at all scales, all scale transitions are performed the same way, and no particular scale is privileged. However, a scaled version of a given signal $\{x_n\}$ at some scale s is not unique. As an example, signals of the form $(\boldsymbol{G} \uparrow)^N (\downarrow \boldsymbol{G}')^M (\boldsymbol{G} \uparrow)^M (\downarrow \boldsymbol{G}')^N x$ are all "scaled" versions of x at scale 1, for all nonnegative values of N and M. Therefore, to characterize scaled versions of x, we need another parameter, namely, resolution, which is defined in Section V.

IV. Hermitian Transposition and Inner Products

Before introducing the concept of resolution, it is convenient to say a few words about Hermitian transposition of operators and inner products. In matrix notation, the

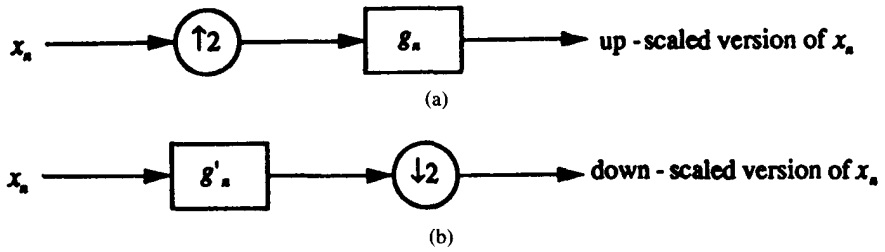


Fig. 2. Flow graphs of up- and downscaling operators. (a) Upscaling (6) is upsampling followed by filtering. (b) Downscaling (7) is filtering followed by downsampling. The filters of impulse response $\{g_n\}$ and $\{g'_n\}$ are called scaling filters.

upscaling operator is

$$G \uparrow x = \begin{pmatrix} \vdots & & & & & \\ & g_4 & g_2 & g_0 & g_{-2} & g_{-4} & \\ \cdots & g_5 & g_3 & g_1 & g_{-1} & g_{-3} & \cdots \\ & g_6 & g_4 & g_2 & g_0 & g_{-2} & \\ & & & \vdots & & & \end{pmatrix} \cdot \begin{pmatrix} \vdots \\ x_{-1} \\ x_0 \\ x_1 \\ \vdots \end{pmatrix}. \tag{8}$$

Its Hermitian transpose is, by definition, the operator obtained from (8) by transposition and complex conjugation, i.e.,

$$\begin{pmatrix} & & \vdots & & \\ & g_1^* & g_2^* & g_3^* & g_4^* & g_5^* & \\ \cdots & g_{-1}^* & g_0^* & g_1^* & g_2^* & g_3^* & \cdots \\ & g_{-3}^* & g_{-2}^* & g_{-1}^* & g_0^* & g_1^* & \\ & & & \vdots & & & \end{pmatrix}. \tag{9}$$

This is exactly the matrix form of a downscaling operator (7), with scaling filter $\{g^*_{-n}\}$ in place of $\{g_n\}$. Following the filter bank terminology [24], $\{g^*_{-n}\}$ is called the "paraconjugate" sequence of $\{g_n\}$ (the connection with filter banks will be explained in Section VIII). Paraconjugation will be denoted in this paper with a tilde symbol:

$$\tilde{g}_n = g^*_{-n}. \tag{10}$$

We have thus seen that the Hermitian transpose of the upscaling operator $G \uparrow$ is the paraconjugate downscaling operator $\downarrow \tilde{G}$. Similarly, the Hermitian transpose of the downscaling operator $\downarrow G'$ is the paraconjugate upscaling operator $\tilde{G}' \uparrow$.

Hermitian transposition is useful for several reasons. First, it has a flow-graph interpretation: it is well known [5] that transposing a linear operator amounts to transposing its flow graph: the Hermitian transposed flow graph is thus obtained by reversing the directions of all arrows—summing nodes become branching nodes and *vice versa*—and conjugating the multipliers. From the discussion above, it follows that the two flow graphs of Fig. 2 are each other's Hermitian transpose if and only if the scaling filters $\{g_n\}$ and $\{g'_n\}$ are each other's paraconjugate (10). In any case, the flow-graph computational structures of up- and downscaling operators are always tranpose of each other. This has useful implications for deriving algorithms [20]; once an algorithm has been derived to compute downscaling, the transposed algorithm computes upscaling (and vice versa), at the same computational cost if both scaling filters have the same length. This easily carries over to direct and inverse wavelet transforms (see Section VIII and [20]).

Another characterization of Hermitian transposition uses inner products and is useful in the sequel. The inner product of two discrete-time signals $\{x_n\}$ and $\{y_n\}$ can be written

$$\langle x, y \rangle = (\cdots x_{-1} \quad x_0 \quad x_1 \cdots) \begin{pmatrix} \vdots \\ y^*_{-1} \\ y^*_0 \\ y^*_1 \\ \vdots \end{pmatrix}. \tag{11}$$

This definition requires that signals have finite energy, i.e., $\|x\| < \infty$, which is assumed throughout this paper. The inner product $\langle x, y \rangle$ measures the "similarity" between x and y and permits to interpret signals as geometric vectors: for example, the signals x and y are "orthogonal" if $\langle x, y \rangle = 0$. They are "orthonormal" if they moreover have a unit norm, i.e., $\|x\| = \|y\| = 1$. Using (11), it is easy to show that the Hermitian transpose $O^\dagger$ of some operator O can be alternatively defined as satisfying

$$\langle x, Oy \rangle = \langle O^\dagger x, y \rangle \tag{12}$$

or, equivalently,

$$\langle x, O^\dagger y \rangle = \langle Ox, y \rangle \tag{13}$$

for any signals x and y of finite energy. These equations do not introduce any new concept: only notations are new. They are useful because of their conciseness: any operator on the left-hand side of an inner product can be brought to the right-hand side after Hermitian transposition, and vice versa. For scaling operators we can thus write

$$\langle x, \boldsymbol{G} \uparrow y \rangle = \langle \downarrow \tilde{\boldsymbol{G}} x, y \rangle \tag{14}$$

$$\langle x, \downarrow \boldsymbol{G}' y \rangle = \langle \tilde{\boldsymbol{G}}' \uparrow x, y \rangle \tag{15}$$

which, respectively, stand for the apparently more complicated formulas

$$\sum_n x_n \left(\sum_k g_{n-2k} y_k \right)^* = \sum_n \left(\sum_k \tilde{g}_{2n-k} x_k \right) y_n^*$$

and

$$\sum_n x_n \left(\sum_k g'_{2n-k} y_k \right)^* = \sum_n \left(\sum_k \tilde{g}'_{n-2k} x_k \right) y_n^*.$$

To simplify notations, proofs in the Appendix will often use (14) and (15).

V. Discrete-Time Resolution and Biorthogonality

In this section we define the resolution notion for characterizing different versions of an original signal $\{x_n\}$ at the same scale. Intuitively, the more information is present in a scaled version of $\{x_n\}$, the higher the resolution. More precisely, we define the resolution parameter r as follows: A scaled version of $\{x_n\}$, obtained by action of up- and/or downscaling operators on $\{x_n\}$, is at resolution $r = 2^{-j}$ ($j \geq 0$) if it is characterized by one sample every other 2^j sampling period of $\{x_n\}$.

For example, x itself is at resolution 1. Its downscaled version $\downarrow \boldsymbol{G}' x$ (7), which lives at scale $1/2$, is also at resolution $1/2$, because the downscaling operator rejects half of the samples. Note that in general, the resolution of a signal cannot exceed its scale, otherwise it would be characterized by more samples than are actually present in the signal. Therefore, we always have

$$r(y) \leq s(y) \tag{16}$$

for any signal y. Since we start from an original signal $\{x_n\}$ at resolution 1, all resolutions considered in this paper are negative powers of two: $r = 2^{-j}$, $j \geq 0$.

How is resolution affected by up and downscaling operators? We have seen that downscaling, when applied to the original signal $\{x_n\}$, halves both scale and resolution. In contrast, since an upscaled signal is computed directly from the signal coefficients, resolution is not increased by up-scaling, which simply "magnifies" a signal without adding new details. Also, despite the low-pass filtering present in upscaling (6), information will not necessarily be lost when upscaling a signal. This is clear in the example (5). We therefore assume in this paper that the upscaling operator $\boldsymbol{G} \uparrow$ is one to one:

$$x \neq y \quad \text{implies} \quad \boldsymbol{G} \uparrow x \neq \boldsymbol{G} \uparrow y. \tag{17}$$

In fact, this can be proven under weak conditions as follows. Let $z = x - y$. If $\boldsymbol{G} \uparrow z = 0$, then from (6), the outputs of convolutions $g_{2n} * z_n$ and $g_{2n+1} * z_n$ vanish. Seen in the frequency domain, this implies $z_n = 0$ when the frequency response $G(e^{j\omega})$ does not vanish for $\omega = \omega_0$ and $\omega_0 + \pi$, a condition which is always satisfied in practical systems, e.g., when the scaling filter is FIR half-band low pass.

To summarize, there are two important rules concerning changes of scale and resolution:

1) When replacing the original signal x by its downscaled version $\downarrow \boldsymbol{G}' x$, scale and resolution are halved. In particular,

$$s(\downarrow \boldsymbol{G}' x) = s(x)/2 \quad \text{and} \quad r(\downarrow \boldsymbol{G}' x) = r(x)/2. \tag{18}$$

2) Upscaling any signal y via $\boldsymbol{G} \uparrow$ doubles its scale but leaves resolution unchanged, i.e.,

$$s(\boldsymbol{G} \uparrow y) = 2s(y) \quad \text{and} \quad r(\boldsymbol{G} \uparrow y) = r(y). \tag{19}$$

Another operator allows one to halve resolution while leaving scale unchanged; it plays a central role in the following and is defined by

$$\boldsymbol{A}x = \boldsymbol{G} \uparrow \downarrow \boldsymbol{G}' x \tag{20}$$

that is, downscaling followed by upscaling. From rules (18) and (19), this signal is at scale 1 and resolution $1/2$ when x is the original signal. Operator $\boldsymbol{A}$ is therefore called "approximation operator at half the resolution." Note that $\boldsymbol{A}$ does not reduce to a filter because it is not shift invariant. Its flow graph is depicted in Fig. 3.

Using the two scaling operators (6), (7), one can compute an infinite set of different versions of an original signal $\{x_n\}$ which have different scales and resolutions. Now, under some special conditions on these two operators it is possible that a scaled version at given scale s and resolution r is unique. In this case, only two intuitive notions (scale and resolution) are necessary to characterize all versions of x that might be present in multiresolution systems.

To achieve this, several equivalent conditions must be fulfilled. The up and downscaling operators must satisfy the property

$$\downarrow \boldsymbol{G}' \boldsymbol{G} \uparrow = \boldsymbol{I} \tag{21}$$

i.e., upscaling followed by downscaling leaves the signal unchanged. Under this condition, a scaled version of $\{x_n\}$ at given scale $s = 2^{-i}$ and resolution $r = 2^{-j}$ is unique; it is given by the formula $(\boldsymbol{G} \uparrow)^{j-i} (\downarrow \boldsymbol{G}')^j x$, proven in Appendix B. Note that (21) implies that the downscaling operator $\downarrow \boldsymbol{G}'$ halves both scale and resolution only if it is applied to signals such as x itself (see (18)), for which $s = r$. However, for "overscaled" signals, such as $\boldsymbol{A}x$, for which $s > r$, the operator $\downarrow \boldsymbol{G}'$ only halves scale and leaves resolution unchanged.

Another condition is that the approximation operator at half the resolution (20) is a projector,

$$\boldsymbol{A}^2 = \boldsymbol{A} \tag{22}$$

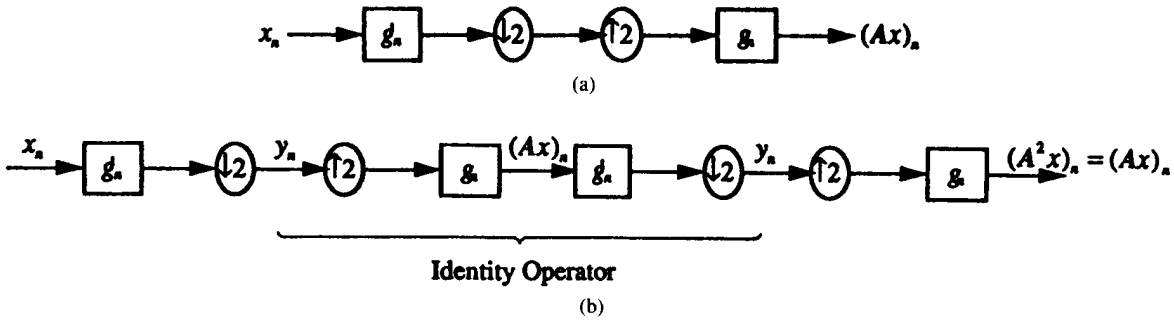


Fig. 3. A flow-graph illustration of the fact that the approximation Ax of an original signal x at half the resolution is a projection. (a) Downscaling, followed by upscaling approximate x at half its resolution. (b) Reapproximating Ax by A leaves Ax unchanged. This is equivalent to (21), i.e., to the condition that upscaling, followed by downscaling, is the identity operator.

i.e., reapproximating the approximation Ax leaves it unchanged. In fact, (21) implies that $A^2 = G \uparrow\downarrow G'G \uparrow\downarrow G'$ reduces to $G \uparrow\downarrow G' = A$ by simplification of the middle term $\downarrow G'G \uparrow = I$. This illustrated using flow graphs in Fig. 3(b).

Appendix B proves that conditions (21) and (22) are in fact equivalent, and that another equivalent condition is the "biorthogonality" property that deserves attention. The two families of shifted scaling impulse responses $\{g_{n-2k}\}$ and $\{\tilde{g}'_{n-2k}\}$, indexed by k, satisfy

$$\langle g_{n-2k}, \tilde{g}'_{n-2l}\rangle = \sum_n g_{n-2k} g'_{2l-n} = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

(recall that $\tilde{g}'$ is the paraconjugate (10) of g'). Therefore, biorthogonality is necessary and sufficient for scaled versions of the original signal $\{x_n\}$ to be uniquely determined by their scale and resolution parameters. This should always hold in "coherent" multiresolution systems—for which multiresolution approximations are unique. Biorthogonality will therefore be assumed in the sequel.

Orthonormality is a special case of biorthogonality for which one further imposes that scaling filters are paraconjugate of each other,

$$g_n = \tilde{g}'_n \tag{24}$$

so that the family of shifted signals $\{g_{n-2k}\}$ form an orthonormal set,

$$\langle g_{n-2k}, g_{n-2l}\rangle = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

From (24) and the discussion of Section IV it follows that orthogonality is a shorthand for the combination of two properties: biorthogonality, plus the condition that up- and downscaling operators are Hermitian transpose of each other.

VI. Multiresolution Residue Signals and Pyramid Transforms

This section uses the definitions and properties of scale and resolution discussed above to give a general definition of discrete-time multiresolution signal decomposition, and reviews the pyramid transform [1], [3] within this framework. We assume that scale and resolution characterize scaled signals as discussed in the preceding section (conditions (21), (22), or (23)). Intuitively, in a multiresolution analysis, the original signal $\{x_n\}$ is decomposed into several multiresolution components associated to different resolutions, while during synthesis, the signal is reconstructed from its multiresolution components.

From Section V, a signal at some resolution r contains all the information necessary to obtain versions at lower resolutions $r' \le r$: For example, the version of original signal x at scale 2^{-i} and resolution 2^{-j}, namely. $(G \uparrow)^{j-i}(\downarrow G')^j x$, is brought to resolution 2^{-j-1} by applying the operator $(G \uparrow\downarrow G')^{j-i+1}$. Therefore, to avoid redundancy of information in a multiresolution analysis, the signal is decomposed into residue signals that catch "details" from one resolution to the next finer one. These residue signals are defined by difference as follows. The "residue" signal of $\{x_n\}$ at scale s and resolution r is the signal which, added to the scaled version of $\{x_n\}$ at same scale and resolution, increases resolution from r to $2r$. That is

$$\text{scaled signal } (s, r) + \text{residue signal } (s, r) = \text{scaled signal } (s, 2r). \tag{26}$$

We have assumed

$$s \ge 2r \tag{27}$$

so that the right-hand side of (26) is well defined according to (16).

Now, we can define a general multiresolution decomposition of an original signal $\{x_n\}$ as a collection of residue signals at successive resolutions $1/2$, $1/4$, $1/8$, $\cdots$. These residue signals are computed during multiresolution analysis. During multiresolution synthesis, the signal $\{x_n\}$ is reconstructed starting from a low resolution scaled version of $\{x_n\}$, by applying (26) iteratively to increase resolution until the highest resolution 1 is reached, which gives the original signal back. Note that with this definition, a multiresolution decomposition differs from another only by the scales of multiresolution components.

The pyramid transform is a direct application of these ideas. It was first introduced by Burt and Adelson [3] for

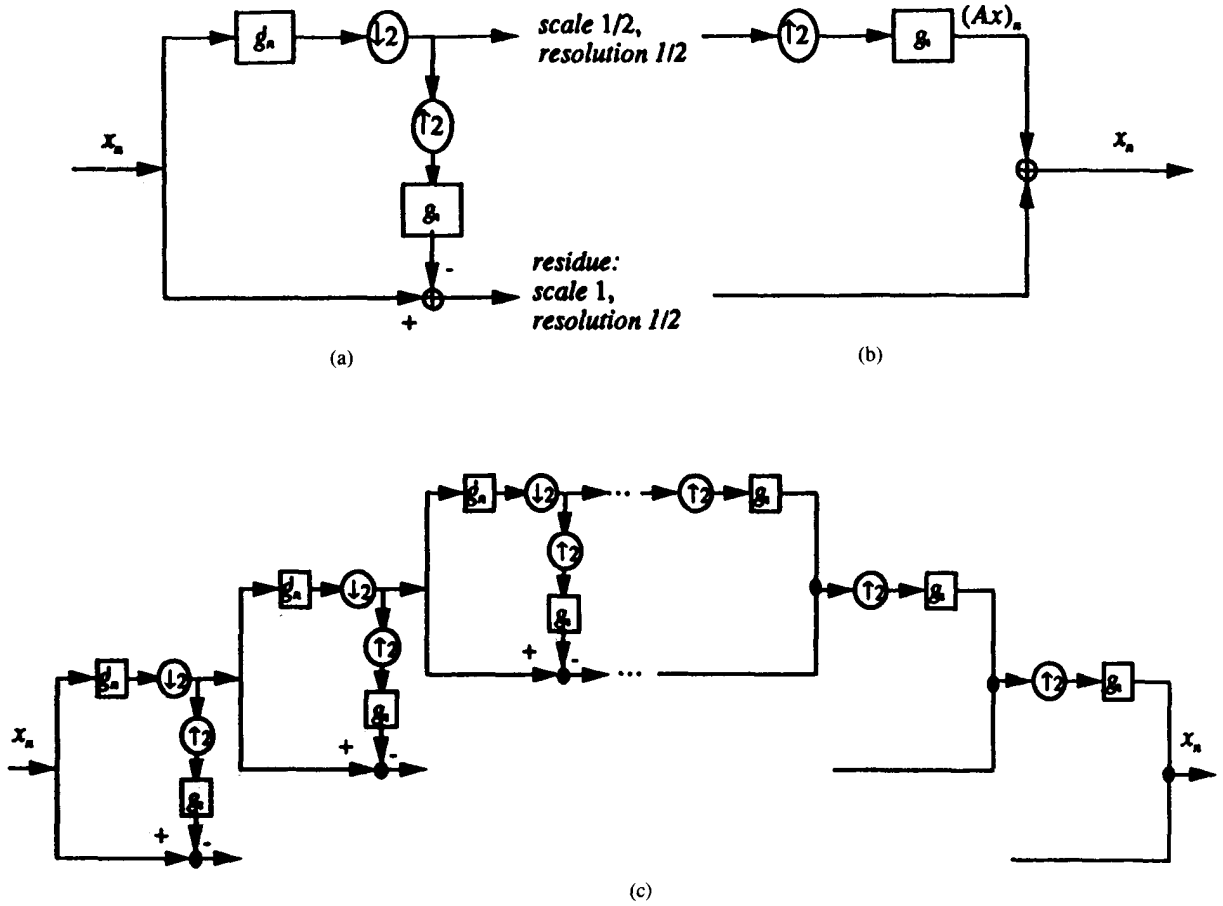


Fig. 4. Flow-graph implementation of the pyramid transform. (a) One step of decomposition; the original signal x is decomposed into a version of x at half its scale, and a residue signal at the same scale and half the resolution, the latter being obtained by difference. (b) The reconstruction part uses (26). One simply adds back what was subtracted before. (c) The elementary cell of (a) and (b) is iterated J times to provide a full pyramidal decomposition on J "octaves" (here $J = 3$).

image coding purposes; here we describe pyramid decompositions for one-dimensional signals in the framework of this paper. A pyramid transform on J "octaves" decomposes the signal $\{x_n\}$ into the collection of residue signals at scale $2^{-(j-1)}$ and resolution 2^{-j}, where $j = 1, \cdots, J$, plus a low resolution version of $\{x_n\}$, namely the scaled signal at scale and resolution 2^{-J}. This description is sufficient to fully describe a pyramid transform.

It is easy to connect this to the original description of Burt and Adelson [3], by deriving a computational tree-structure for its implementation: Start with one step of decomposition, i.e., let $J = 1$. From the above definition, the original signal x is decomposed into two components: its residue signal at scale 1 and resolution $1/2$, which, from (20), (26) is

$$x - Ax = x - G \uparrow \downarrow G'x \tag{28}$$

plus its scaled version at scale and resolution $1/2$, i.e., $\downarrow G'x$. The corresponding flow graph is depicted in Fig. 4(a). To reconstruct x, $\downarrow G'x$ is brought back to scale 1 and (26) is applied, i.e.,

$$G \uparrow (\downarrow G'x) + (x - G \uparrow \downarrow G'x) = x. \tag{29}$$

In other words, one simply adds what has been previously subtracted. The corresponding flow graph is depicted in Fig. 4(b). This decomposition readily extends to a full computation of a pyramid transform on J octaves. Simply note that from the definition of general multiresolution decomposition, scale and resolution parameters of multiresolution residue signals are halved at each stage (octave) of the decomposition. Now, since any residue signal is a difference of scaled versions of x, the rules (18), (19) apply for residue signals as well. Therefore, scale and resolution parameters of a multiresolution residue signal are halved when replacing x by $\downarrow G'x$. This amounts to iterating the basic computational structure (one-step decomposition) of Fig. 4 on $(\downarrow G')^j x$ at each level $j = 0, \cdots, J - 1$. This gives the flow graph of Fig. 4(c).

In Burt and Adelson's terminology [3], the above multiresolution residue signals form a Laplacian pyramid, while the set of versions of $\{x_n\}$ at scales $2^{-i}(i = 0, \cdots, J)$ is called a Gaussian pyramid. The terminology "pyramid" comes from the fact that muiltiresolution components are computed at successive scales, from scale 1 ("base" of pyramid) to scale $2^{-(J-1)}$ ("top" of pyramid). "Gaussian" and "Laplacian" were named after the type of scaling filters used in [3].

Recall that one always has perfect reconstruction since what has been subtracted is added back during synthesis (see (29)). Therefore, there is no constraint at all on scaling filters $\{g_n\}$ and $\{g'_n\}$. Even the basic biorthogonality constraint (23), which ensures uniqueness of scaled versions of x at a given scale, is not necessary for the scheme to work. There is a price to pay, however: the multiresolution residue signals correspond to scales $s = 2^{-i}(i = 0, \cdots, J-1)$ that are always twice their resolution $r = 2^{-i-1} = s/2$. Therefore the transform is overcomplete: starting from an original signal sampled at rate $1/T$, the multiresolution components of a pyramid transform are sampled at rate $1/T(1 + 1/2 + 1/4 + \cdots + 2^{-J}) \approx 2/T$. This means that there are about twice as many transform coefficients as original signal samples. (In two dimensions this factor becomes $4/3$ [1], [8].)

In contrast with the pyramid transform, the discrete wavelet transform, presented below, is not overcomplete (there are as many wavelet coefficients as signal samples) but requires design constraints on scaling filters.

VII. The Discrete Wavelet Transform and Perfect Reconstruction Filter Banks

We have seen in the preceding section that a potential drawback of the pyramid transform is its overcompleteness, due to the fact that the residue signals involved are "overscaled", i.e., they satisfy $s = 2r$. In a "discrete wavelet transform (DWT)," each residue signal is "critically sampled," i.e., its scale and resolution parameters are equal, $r = s$. To describe the DWT we therefore need to extend the basic definition of residue signals (26) to this case, which violates restriction (27).

We are led to consider two operators, $\boldsymbol{H}\uparrow$ and $\downarrow \boldsymbol{H}'$, defined similarly as $\boldsymbol{G}\uparrow$ and $\downarrow \boldsymbol{G}'$, with impulse responses $\{h_n\}$ and $\{h'_n\}$, respectively. Define $\downarrow \boldsymbol{H}'x$ as the residue signal of x at scale $1/2$ and resolution $1/2$, which is brought back to scale 1 by applying $\boldsymbol{H}\uparrow$; this gives $\boldsymbol{H}\uparrow\downarrow \boldsymbol{H}'x$ as the residue signal at scale 1 and resolution $1/2$ defined by (28), i.e.,

$$x - \boldsymbol{G}\uparrow\downarrow \boldsymbol{G}'x = \boldsymbol{H}\uparrow\downarrow \boldsymbol{H}'x. \tag{30}$$

This condition is simply a rewriting, in operator notation, of perfect reconstruction (without delay) of a two-band filter bank [23] depicted in Fig. 5(a), with low-pass filters $\{g_n\}$, $\{g'_n\}$ and high-pass filters $\{h_n\}$, $\{h'_n\}$. This definition of residue signals at scale and resolution $r = s = 1/2$ is immediately extended to other values of $r = s$ by application of the rules (18), (19) as in the preceding section. This gives $\downarrow \boldsymbol{H}'(\downarrow \boldsymbol{G}')^{j-1}x$ as the residue signal at scale and resolution $r = s = 2^{-j}(j > 0)$.

Using this extended definition of multiresolution residue signals, a DWT on J "octaves" is simply defined as the transformation of a signal $\{x_n\}$ into the following "wavelet coefficients" $\{w_n^j\}$, $j = 1, \cdots, J$, which are precisely the residue signals $\downarrow \boldsymbol{H}'(\downarrow \boldsymbol{G}')^{j-1}x$ at scale and resolution $r = s = 2^{-j}$, and a low resolution signal, $\{v_n^J\}$, the scaled version of x at scale and resolution 2^{-J}. To reconstruct $\{x_n\}$, residue signals are first upscaled by means of $\boldsymbol{H}\uparrow$ (as in (30)), then definition (26) is applied iteratively to increase resolution until resolution 1 is reached, i.e., until the original signal $\{x_n\}$ is recovered. This is performed by the "inverse DWT" (IDWT).

The DWT and the IDWT are easily recognized to be an octave-band filter bank [23]: To halve both resolution and scale of residue signals, x is replaced by $\downarrow \boldsymbol{G}'x$ according to rule (18), i.e., the two-band filter bank of Fig. 5(a) is applied to the scaled version $(\downarrow \boldsymbol{G}')^j x$ at each level j. This gives the flow graph of Fig. 5(b), where the analysis filter bank computes the DWT, whereas the synthesis filter bank computes the IDWT. Note that this filter bank is critically sampled, i.e., there are as many computed wavelet coefficients as signal samples, as opposed to the pyramid transform, which is overcomplete by 50% (Section VI). However, low-pass ($\{g_n\}$, $\{g'_n\}$) and high-pass ($\{h_n\}$, $\{h'_n\}$) filters are constrained to satisfy the perfect reconstruction property (30).

Although the computational structures of the DWT and octave-band perfect reconstruction filter bank are identical, the DWT provides an alternative formalism, which describes it as a multiresolution decomposition in time, rather than as a subband frequency decomposition. This formalism can be developed by defining "basis functions" onto which the signal is decomposed: The basic analysis (synthesis) "scaling sequence" is $\tilde{g}'_n$ (g_n, respectively), while the basic analysis (synthesis) "wavelet" is the corresponding high pass impulse response $\tilde{h}'_n$ (h_n, respectively). Other scaling sequences and wavelets are obtained from the "basic" ones by successive upscalings:

- Analysis scaling sequences:

$$\tilde{g}'^j = (\tilde{\boldsymbol{G}}'\uparrow)^{j-1}\tilde{g}'. \tag{31}$$

- Analysis wavelets:

$$\tilde{h}'^j = (\tilde{\boldsymbol{G}}'\uparrow)^{j-1}\tilde{h}'. \tag{32}$$

- Synthesis scaling sequences:

$$g^j = (\boldsymbol{G}\uparrow)^{j-1}g. \tag{33}$$

- Synthesis wavelets:

$$h^j = (\boldsymbol{G}\uparrow)^{j-1}h. \tag{34}$$

Appendix C shows that the wavelet coefficients w_k^j of the original signal $\{x_n\}$ at octave j ($j = 1, \cdots, J$), i.e., the residue signal at scale and resolution 2^{-j}, are inner products of $\{x_n\}$ with the corresponding analysis wavelets:

$$w_k^j = \langle x_n, \tilde{h}'^j_{n-2^j k}\rangle, \qquad j = 1, \cdots, J. \tag{35}$$

Similarly the low resolution component is

$$v_k^J = \langle x_n, \tilde{g}'^J_{n-2^J k}\rangle. \tag{36}$$

An alternative definition of the DWT is thus (35), (36). The inverse DWT reconstructs the signal as a linear combination of shifted synthesis wavelets weighted by the corresponding wavelet coefficients, plus a very low res-

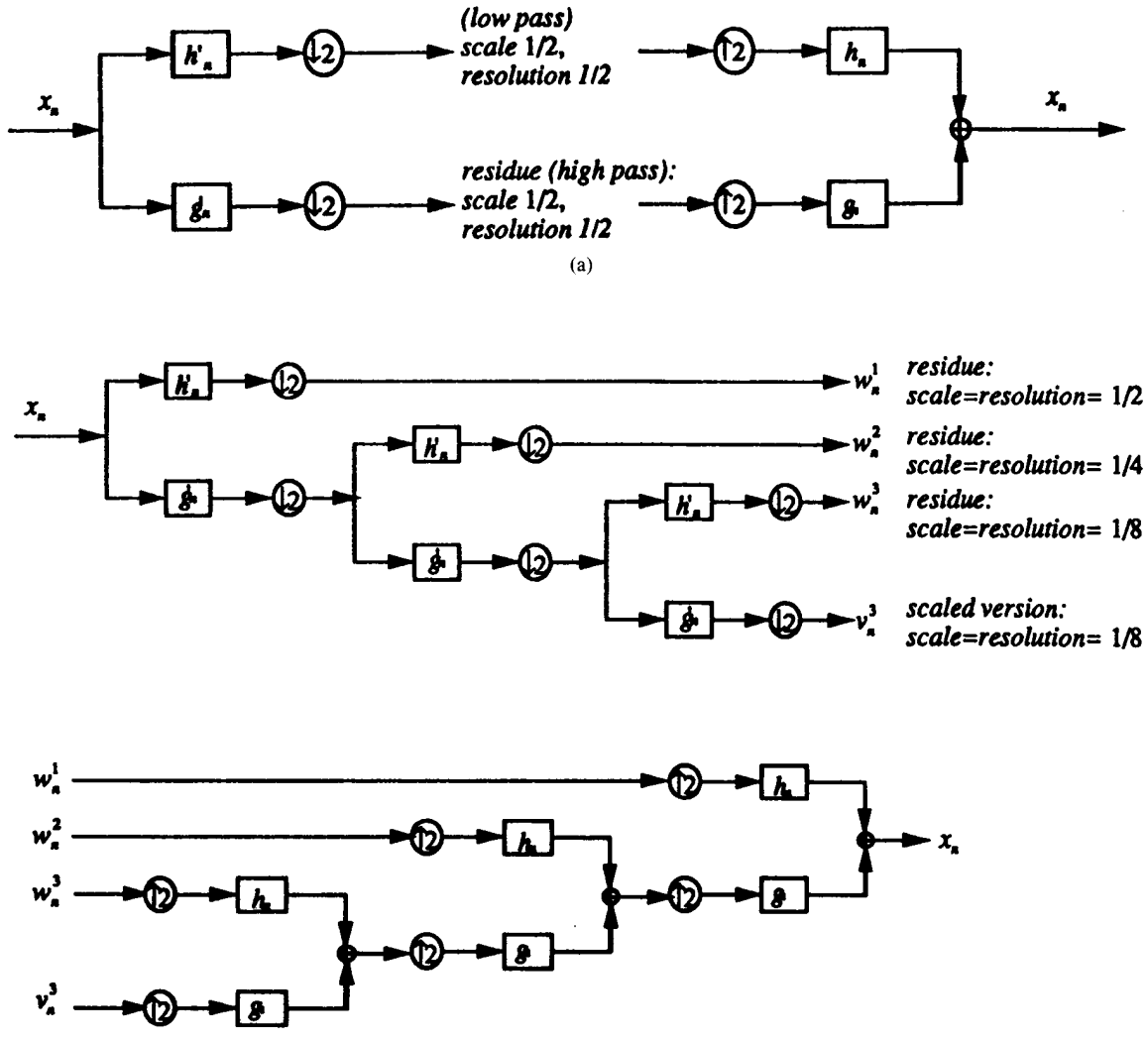


Fig. 5. Flow-graph implementation of the discrete wavelet transform (DWT). (a) One step of decomposition/reconstruction. The original signal x is decomposed into a version of x at half its scale and resolution, and a residue signal also at half its scale and resolution. This is simply a two-band perfect reconstruction filter bank; the decomposition (analysis) part uses the low-pass "scaling" filter $\{g'_n\}$ and the high-pass "wavelet" filter $\{h'_n\}$. The reconstruction (synthesis) part uses the low-pass "scaling" filter $\{g_n\}$ and the high-pass "wavelet" filter $\{h_n\}$. (b) The elementary cell of (a) is iterated J times to provide a full DWT on J "octaves" (here $J = 3$).

olution approximation of x, obtained similarly:

$$x_n = \sum_{j=1}^{J} \sum_k w_k^j h_{n-2^j k}^j + \sum_k v_k^J g_{n-2^J k}^J. \tag{37}$$

Equation (37), also proven in Appendix C, is simply a rewriting of the previous definition of the inverse DWT. As a result, wavelets and scaling sequences are the discrete-time "basis functions" present in a multiresolution decomposition: for example, wavelet $h_{n-2^j k}^j$, shifted in proportion to its scale,[3] contributes for the representation of x at resolution 2^{-j}, around location $k2^{-j}$ in the signal.

[3]In this paper, scale is defined from the viewpoint of the signal. Therefore in this presentation of the DWT, scale and resolution parameters of DWT components are equal. Scale can also be defined from the viewpoint of the wavelets (32), (34) that are "analyzing" the signal [12], [15], [16]. Small resolutions would then correspond to large scales for wavelets, and scale and resolution parameters would be reciprocal of each other for wavelet coefficients.

VIII. Biorthogonal and Orthogonal Wavelets

The biorthogonality property of $\{g_n\}$ and $\{g'_n\}$ (23) derived in Section V ensures that a scaled (or residue) version of an original signal $\{x_n\}$ at a given scale and resolution is unique. In a DWT, we moreover have the constraint of perfect reconstruction (30) of the two-band filter bank of Fig. 5(a). Then, Appendix D shows that biorthogonality carries over to "wavelets" $\{h_n\}$ and $\{h'_n\}$. In fact, restricting to FIR filters for convenience, the four filters g, g', h, h' are simply related (Appendix D):

$$\begin{cases} h_n = (-1)^n g'_n \\ h'_n = (-1)^{n+1} g_n. \end{cases} \tag{38}$$

(When the filters are moreover assumed to be causal, this set of equations is slightly affected and induces a delay on

the output of the filter bank.) To design a DWT, it is therefore sufficient to find two low-pass filters satisfying biorthogonality (23) and to define high-pass filters by (38). In addition, the two families of analysis and synthesis wavelets are mutually biorthogonal not only across time shifts at a given scale, but also across scales:

$$\langle h^j_{n-2jk}, \tilde{h}'^j_{n-2jl} \rangle = \begin{cases} 1 & \text{if } k = l \text{ and } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{39}$$

A proof is given in Appendix D. In fact Appendix D shows that biorthogonality is already implied by perfect reconstruction, and therefore always occurs in a DWT.

Biorthogonality of wavelets (39) is perhaps easier to understand if we note that, in (37), the term contributing to resolution 2^{-j} is

$$\sum_k \langle x_n, \tilde{h}'^j_{n-2jk} \rangle h^j_{n-2jk}. \tag{40}$$

Now (39) implies that (40) is a projection of x onto the subspace of signals spanned by the wavelets $\{h^j_{n-2jk}\}$, since if x belongs to this subspace, then using (39) one finds that (40) reduces to x. In other words, owing to biorthogonality, the DWT decomposes a signal into projections of this signal onto subspaces corresponding to different resolutions 2^{-j} ($j = 1, \cdots, J$). This point of view is the discrete counterpart of the original theory of Mallat [15], [16] and Meyer [17].

The Orthonormal Case

We have seen in Section VI that the orthonormal case is a special case of biorthogonality for which one further imposes (24), i.e., $g_n = \tilde{g}'_n$. From (38) this implies $h_n = \tilde{h}'_n$; analysis and synthesis wavelets are equal. From the discussion of Section IV, it follows that the analysis and synthesis filter banks of Fig. 5(b), which, respectively, perform the DWT and IDWT, are Hermitian transpose of each other. In other words the DWT is an orthogonal transform: it is composed of lossless or paraunitary two-band filter banks [24] for which analysis and synthesis filters are paraconjugate of each other. The orthogonal discrete wavelet transform was first connected to the pyramid transform, in a similar manner as in Section VII, by Adelson *et al.* [1].

From (39) and the relation $h_n = \tilde{h}'_n$, the wavelets form an orthonormal basis:

$$\langle h^j_{n-2jk}, h^i_{n-2il} \rangle = \begin{cases} 1 & \text{if } k = l \text{ and } i = j \\ 0 & \text{otherwise} \end{cases} \tag{41}$$

hence (40) is now an orthogonal projection of x. This means that among all signals y belonging to the subspace spanned by the $\{h^j_{n-2jk}\}$, (40) is the one that minimizes the quadratic distance $\|x - y\|^2$. Therefore, orthonormality can also be thought of as the condition under which multiresolution components are most similar to the original signal [15].

Orthonormality is often considered as an essential property for coding applications (see, e.g., [15]), where the coding strategy is based on an LMS error criterion. It also simplifies the design: scaling and wavelet sequences are simply related since (38) reduces to $h_n = (-1)^n g^*_{L-1-n}$ for causal filters. However, it is well known [8], [9], [23], [26] that orthonormal filters cannot be of linear phase, except for a trivial choice; this has motivated the search for biorthogonal, linear phase wavelets [4], [26].

IX. Comparison with Wavelet Series

There is a remarkable parallelism between the DWT, presented in the previous section, and its continuous-time counterpart, which has been developed for functions of a continuous variable by Meyer [17], Mallat [15], [16], Daubechies [8] and other authors. We refer to the latter as the "wavelet series decomposition" (see (3) and (4)). This section compares the two models within the framework of this paper. The relationship of discrete and continuous wavelet transforms was also clarified by Shensa [22].

The analog model uses a continuous version of the inner product, namely,

$$\langle x(t), y(t) \rangle = \int x(t)\, y^*(t)\, dt \tag{42}$$

and applies to analog signals $x(t)$ of finite energy $\|x(t)\|^2 = \langle x(t), x(t) \rangle$. Also, upscaling is the simple dilation $x(t) \rightarrow x(t/2)/\sqrt{2}$. (In the presentation of the DWT above, the constant $\sqrt{2}$ has been integrated into the discrete scaling sequences.) As in the DWT, one defines analysis and synthesis basic scaling functions, denoted by $\hat{\varphi}(t)$ and $\varphi(t)$, respectively, and analysis and synthesis basic wavelets $\hat{\psi}(t)$ and $\psi(t)$. The whole set of scaling functions and wavelets is defined as in (31)–(34) by successive upscalings. For example, the synthesis wavelets are $\psi^j(t) = 2^{-j/2}\psi(2^{-j}t)$, and the other basis functions involved, namely, $\varphi^j(t)$, $\hat{\psi}^j(t)$, and $\hat{\varphi}^j(t)$ are defined similarly. Then, the wavelet series coefficients are (compare with (35), (36))

$$\begin{aligned} W^j_k &= \langle x(t), \hat{\psi}^j(t - 2j^k) \rangle \\ V^J_k &= \langle x(t), \hat{\varphi}^J(t - 2^J k) \rangle \end{aligned} \tag{43}$$

and the reconstruction part uses a wavelet series (compare with (37))

$$x(t) = \sum_{j=1}^{J} \sum_k W^j_k \psi^j(t - 2^j k) + \sum_k V^J_k \varphi^J(t - 2^J k). \tag{44}$$

Analog wavelets are biorthogonal (compare with (39))

$$\langle \psi^j(t - 2^j k), \hat{\psi}^i(t - 2^i l) \rangle = \begin{cases} 1 & \text{if } k = l \text{ and } i = j \\ 0 & \text{otherwise} \end{cases} \tag{45}$$

and, of course, orthonormality occurs when analysis and synthesis wavelets are equal. Other properties, such as

phase linearity [4], [26] and finite or "compact" support [4], [8], [26], are also expressed similarly as in the discrete-time case.

Therefore, the sole ability of the wavelet transform to do multiresolution signal decomposition, using orthonormal or biorthogonal bases, should not be critical in deciding whether to choose the discrete-time model or the continuous-time one, because both models share the same properties. In fact, the parallelism that exists between the discrete-time and the continuous-time formalisms is so strong that a DWT can always be deduced from a biorthogonal analog wavelet basis [4]. Specifically, discrete and continuous-time "basis functions" are related by [4]

$$\begin{aligned}\hat{\varphi}(t) &= \sqrt{2}\sum_n \tilde{g}'_n \hat{\varphi}(2t-n)\\ \hat{\psi}(t) &= \sqrt{2}\sum_n \tilde{h}'_n \hat{\varphi}(2t-n)\\ \varphi(t) &= \sqrt{2}\sum_n g_n \varphi(2t-n)\\ \psi(t) &= \sqrt{2}\sum_n h_n \varphi(2t-n). \end{aligned} \tag{46}$$

Furthermore, additional properties satisfied by continuous-time wavelets, such as orthonormality, symmetry, and finite support are automatically fulfilled by the associated discrete-time wavelets [4], [8]. Thus, a continuous wavelet series decomposition scheme induces an associated DWT, the latter being implemented as an octave-band filter bank. Moreover, fed by the discrete input $\{V_k^0\}$, this filter bank provides the continuous-time wavelet series coefficients $\{W_k^j\}$ $(j = 1, \cdots, J)$ and $\{V_k^J\}$ [4]. The algorithm was first derived by Mallat [15] in the context of orthonormal wavelets and was generalized for other types of continuous wavelets transforms by Shensa [22].

However, any arbitrary DWT cannot always be deduced from a wavelet series decomposition scheme, because this would imply that the obtained discrete wavelets satisfy constraints other than biorthogonality. For example, they would satisfy the relations [4] $\Sigma_n h_n = \Sigma_n \tilde{h}'_n = 0$, which are not always met in perfect reconstruction filter banks (see [23]).

This has motivated several researchers to determine the minimal conditions under which continuous-time wavelet bases can be deduced from discrete-time ones. Necessary and sufficient conditions were recently derived, which turned out to be quite technical [4], [14]. It was previously shown [8], however, that a sufficient condition for the equivalence between the continuous-time case and the discrete-time case is the regularity property, which is by itself interesting; it is discussed in the next section.

X. Filter Regularity

Regularity is well understood for continuous-time wavelets; they are regular if they are at least continuous, possibly with several continuous derivatives. Evidently this cannot be expressed directly on discrete-time signals, and the regularity property seems to be an exception to the rule that the discrete-time and continuous-time models share the same properties. Nevertheless, the aim of this section is to show that regularity can be defined for discrete-time wavelets, and that the two definitions of regularity are in fact equivalent.

Regularity, introduced by wavelet theory as a smoothness condition on continuous-time wavelet basis functions, was soon recognized as a new design constraint for perfect reconstruction filter banks which is used to construct regular continuous-time wavelets [4], [8], [26]. Therefore, we want to characterize this constraint as a smoothness condition on the discrete-time "basis functions" defined in Section VII. These "basis sequences" are obtained by successive up-scalings: the aim is to find the conditions on the upscaling operator $\boldsymbol{G}\uparrow$ (6) under which signals $(\boldsymbol{G}\uparrow)^j x$ vary smoothly, even for large j. We then say that the underlying scaling filter $\{g_n\}$ is "regular."

In a biorthogonal DWT, for example, regularity of the scaling filter $\{g_n\}$ implies regularity for all synthesis discrete wavelets and scaling sequences, since they are defined by successive up-scalings associated to $\{g_n\}$ (33), (34). On the other hand, regularity of analysis discrete wavelets and scaling sequences (31), (32) holds if the scaling filter $\{\tilde{g}'_n\}$ is regular. Note that regularity can also be studied for pyramid transforms because their equivalent basis functions are also of the form $(\boldsymbol{G}\uparrow)^j x$, and in fact regularity was first observed by Burt and Adelson [3] in the framework of pyramidal decomposition.

Regularity is believed to be useful in multiresolution decomposition schemes for several reasons [18], [21]. In a DWT, for example, any error occurring in a wavelet coefficient w_k^j results, from (37), in a perturbation in the synthesized signal which is equal to the corresponding synthesis wavelet sequence $\{h_{n-2jk}^j\}$. It is therefore natural to require that this perturbation be smooth, rather than discontinuous, or even fractal, as in the example shown in Fig. 6(a). This may be useful in, e.g., image coding applications where a "fractal" perturbation is likely to "strike the eye" much more than a smooth one, for the same SNR level [2]. On the other hand, requiring that the signal be analyzed by smooth basis functions $\{\tilde{h}'^j_n\}$ ensures that no "artificial" discontinuity (i.e., not due to the signal itself) appears in the wavelet coefficients $w_k^j = \langle x_n, \tilde{h}'^j_{n-2jk}\rangle$. This should also be of interest for compression purposes [2], [27].

A. Definition of Regularity

We now define regularity for discrete-time signals $(\boldsymbol{G}\uparrow)^j x$, i.e., for responses to the operator $(\boldsymbol{G}\uparrow)^j$, restricting to the impulse responses $g_n^j = ((\boldsymbol{G}\uparrow)^j \delta)_n$ for convenience. The following definition of regularity is inspired from Hölder regularity of continuous-time functions [11], [19]: For $0 < \alpha \leq 1$, g_n^j is said to be regular of order $r = \alpha$ if it satisfies

$$|g_{n+1}^j - g_n^j| \leq c(2^{-j})^\alpha \tag{47}$$

where c is the constant independent of j and n. In other

words, the slopes

$$\Delta g_n^j = (g_{n+1}^j - g_n^j)/2^{-j} \tag{48}$$

of the curve g_n^j plotted against $n2^{-j}$, are constrained to increase less than $2^{j(1-\alpha)}$ as j increases: for $\alpha < 1$ they may indefinitely increase (Fig. 6(b) shows an example which satisfies (47) for $\alpha = 0.550 \cdots$ [8], [9], [11]); for $\alpha = 1$ they are always bounded (see Fig. 6(c)). Of course, the regularity condition (47) is stronger as α increases, thereby imposing smoother time evolutions of the sequence $\{g_n^j\}$.

In order to extend this definition to higher regularity orders, we impose (47) on the "discrete derivatives" of g_n^j. The first "discrete derivative" of g_n^j is the sequence of slopes (48). The Nth discrete derivative $\Delta^N g_n^j$ is defined by applying N times the difference operator Δ. Now, g_n^j is said to be regular of order $r = N + \alpha$ $(0 < \alpha \leq 1)$ if its Nth discrete derivative $\Delta^N g_n^j$ is regular of order α, i.e., if

$$|\Delta^N g_{n+1}^j - \Delta^N g_n^j| \leq c(2^{-j})^\alpha. \tag{49}$$

The example of Fig. 6(c) is regular of order $r > 2.102$ [19]. Again this definition imposes a stronger condition as r increases; here condition (49) is imposed on "slopes of slopes" and therefore requires very smooth evolutions of g_n^j.

B. Connection with Wavelet Series

It can be shown [19] that regular discrete wavelet and scaling sequences converge toward continuous-time functions as $j \to \infty$. More precisely, the discrete curves $\{g_n^j\}$, $\{\tilde{g}_n^{\prime j}\}$, $\{h_n^j\}$ and $\{\tilde{h}_n^{\prime j}\}$, when plotted against $n2^{-j}$, uniformly converge to $\varphi(t)$, $\hat{\varphi}(t)$, $\psi(t)$, and $\hat{\psi}(t)$, respectively. In addition, these limit functions have the same Hölder regularity order r as their discrete counterparts [19], Hölder regularity $r = N + \alpha$ $(0 < \alpha \leq 1)$ for the continuous-time case being defined similarly as

$$|\varphi^{(N)}(t + h) - \varphi^{(N)}(t)| \leq c|h|^\alpha \tag{50}$$

where $\varphi^{(N)}(t)$ is the Nth derivative of $\varphi(t)$ (compare with (49)). Note that a Hölder regularity order $r > N$ implies N continuous derivatives.

Now, the continuous-time limits $\varphi(t)$, $\hat{\varphi}(t)$, $\psi(t)$, and $\hat{\psi}(t)$ can be used to define a wavelet series decomposition as in Section IX. Compactly supported continuous-time wavelets have been designed by this method [4], [8], [9]. Therefore, owing to regularity, the identification between discrete-time and continuous-time wavelet schemes is complete: it is only a matter of opinion to state that continuous-time wavelets underlie discrete-time ones (from (46)) or *vice versa* (from the limit process). Likewise, the regularity property can be seen either on continuous-time functions (50) or on discrete-time sequences (49).

C. A First Necessary Condition for Regularity

Regularity of the scaling filter $\{g_n\}$ also implies shape preservation by the upscaling operator $G \uparrow$ (this is Assumption 3) of Section III). In other words, upscaling "stretches" the sequence $\{g_n^j\}$ but does not affect its shape: we have $g_n^j \approx g_{2n}^{j+1} \approx g_{2n+1}^{j+1}$, the approximations being sharper as j increases. Now since $g^{j+1} = G \uparrow g^j$, using (6), we have $g_{2n}^{j+1} = \Sigma_k g_{2k} g_{n-k}^j \approx (\Sigma_k g_{2k}) g_n^j$, hence $\Sigma_k g_{2k} = 1$. Similarly it can be shown that $\Sigma_k g_{2k+1} = 1$, starting from g_{2n+1}^{j+1} ([19] contains a more rigorous proof). These conditions are necessary conditions for shape preservation and regularity. They can be rewritten as $\Sigma_n g_n = 2$ and $\Sigma_n (-1)^n g_n = 0$, i.e.,

$$G(e^{j\omega}) = \begin{cases} 2 & \text{if } \omega = 0 \\ 0 & \text{if } \omega = \pi \end{cases} \tag{51}$$

where $G(e^{j\omega})$ is the transfer function of the scaling filter of impulse response $\{g_n\}$. This justifies that scaling filters are preferably low pass. In fact, they are often chosen to be half-band low-pass filters, a property that is justified by stronger regularity conditions (see below). Note that the first condition in (51) is simply a renormalization: In an orthogonal DWT, for example, they are normalized such that [8] $\Sigma_n g_n = \Sigma_n \tilde{g}_n' = \sqrt{2}$ so that orthogonality (25) holds. Hence, the order of magnitude of $\{g_n^j\}$ decreases as $2^{-j/2}$ as j increases, and one has to renormalize $\{g_n\}$ according to (51) so that the order of magnitude of $\{g_n^j\}$ is preserved for different j's.

Fig. 6 shows several examples of iterated wavelet sequences $\{h_n^j\}$ corresponding to different choices of $\{g_n\}$. In Fig. 6(a), the curve $\{h_n^j\}$ rapidly diverges as j increases. This example does not even fulfill (51), therefore it is not continuous. The example shown in Fig. 6(b) is the first Daubechies wavelet of length 4 [8]. It satisfies (51) but is clearly not very regular; this shows that (51) is not sufficient to obtain a high regularity order. Fig. 6(c) is a very regular example (see below). The example of Fig. 6(d) is particularly interesting: it corresponds to one of Smith and Barnwell filters derived in [23]. Mathematically speaking, this example is not regular because (51) is not satisfied. However, in this case, the scaling filter has a 40-dB attenuation in the stopband, and the value of $G(e^{j\omega})$ at $\omega = \pi$ is thus very small (about 10^{-2}). As a result, (51) is "almost" satisfied and the iterated sequences $\{h_n^j\}$ "pretend" to be regular for small j's. For large j's the curves eventually diverge, with strong oscillations near the wavelet modes (see Fig. 6(d)). Although rejected by the mathematical definition above, such wavelets may be "regular enough" in applications where the number of octaves in the multiresolution decomposition is not too large.

D. Necessary Conditions for Regularity: "Flat Filters" or "Vanishing Moments"

A regularity order $r > N$ requires more than (51). In fact it can be shown [4], [19] that the scaling transfer function $G(e^{j\omega})$ is necessarily of the form

$$G(e^{j\omega}) = \left(\frac{1 + e^{-j\omega}}{2}\right)^{N+1} F(e^{j\omega}). \tag{52}$$

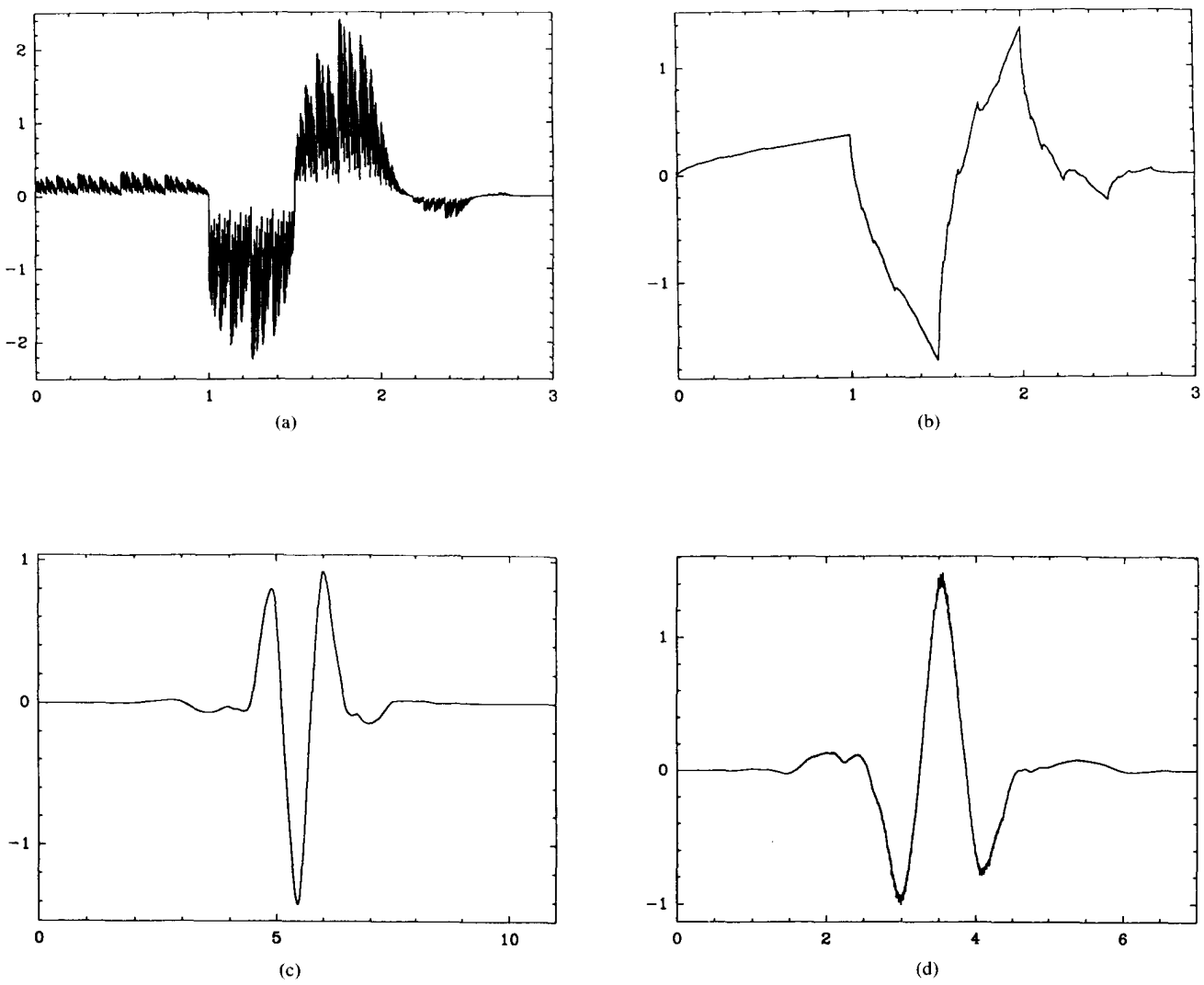


Fig. 6. Iterated wavelet sequences $\{h_n^j\}$, plotted against $n2^{-j}$, for four different choices of the scaling filter $\{g_n\}$ ($j = 11$). (a) $g_0 = g_1 = 0.7$, $g_2 = -g_3 = 0.1$, $g_n = 0$ otherwise. This example satisfies orthonormality (25), and can therefore be used in a perfect reconstruction paraunitary filter bank (see Section VIII). However, the obtained curve is a highly irregular, fractal function. In fact, this curve rapidly diverges as j increases indefinitely. (b) Daubechies wavelet of length 4: $g_0 = (1 + \sqrt{3})/4\sqrt{2}$, $g_1 = (3 + \sqrt{3})/4\sqrt{2}$, $g_2 = (3 - \sqrt{3})/4\sqrt{2}$, $= (1 - \sqrt{3})/4\sqrt{2}$ (after [8]). The regularity order is $r = 0.55 \cdots$. The obtained curve rapidly converges, as j increases indefinitely, to a continuous, but not differentiable, function. (c) Daubechies "closest to linear phase" wavelet of length 12 (after [19]). The regularity order, as determined in [19] is at least 2.102. (d) This example corresponds to the 8-tap scaling filter designed by Smith and Barnwell [23]. Mathematically speaking, the limit process diverges: due to small, but rapid oscillations, the curves slowly diverges as j increases indefinitely. However, the obtained curves look "reasonably regular" for small j's, due to the fact that the scaling filter's transfer function is strongly attenuated in the stop band (see Section X).

Condition (51) follows for $N = 0$. Here (52) implies that up to N derivatives of the transfer function $G(e^{j\omega})$ vanish at $\omega = \pi$, hence the spectrum of a regular scaling filter is "flat" at half the sampling frequency. It is also flat at $\omega = 0$ (with the same order of vanishing derivatives), which tends to imply that the scaling filter should be half-band low pass.

In a DWT, using (38) one finds

$$\sum_n n^i \tilde{h}'_n = 0, \qquad i = 0, \cdots, N \tag{53}$$

and similarly for $\{h_n\}$ if the analysis scaling filter $\{\tilde{g}'_n\}$ is "flat." Equation (53) means that the wavelet sequence $\{\tilde{h}'_n\}$ has $N + 1$ vanishing moments [4], [9], and this can be written similarly for the continuous-time wavelet $\hat{\psi}(t)$ [4], [16], [17]. For example, the wavelet shown in Fig. 6(c) has six vanishing moments. This property may be interesting by itself for some applications [9].

E. Estimating Regularity

The condition (52) or (53) can be easily integrated in a filter design procedure [4], [26]. However, it is only a necessary condition for regularity: there exist nonregular

examples for which (52) holds [19]. It is therefore important to *a posteriori* determine the regularity order of a computed scaling filter $\{g_n\}$, and several regularity estimates have been derived for the purpose [8], [11]. Daubechies' estimate [8] is based on the determination of maxima of spectra. Unfortunately, it requires many computations to determine a good estimate and is not optimal in general [19]. Daubechies and Lagarias [10], [11] have derived a method based on matrix algebra which yields optimal regularity estimates in some instances. The author derived a method in [19], which, unlike earlier estimates, is based on the discrete-time approach described in this section. It is easily implementable, applicable in all cases, and gives optimal estimates for Hölder regularity.

XI. Conclusion

This paper has developed an alternative view of multiresolution theory which focuses on discrete-time signals. It is based on precise notions of scale and resolution in discrete-time (Sections III and V). We have described both the pyramid transform (Section VI) and the discrete wavelet transform (Sections VII and VIII) using these notions. Pyramid transforms uses "overscaled" multiresolution components, while scale and resolution parameters of DWT coefficients are equal.

Biorthogonality is derived as an essential condition for scaled versions of an original signal to be characterized by scale and resolution parameters (Section V). In a DWT, it is also implied by the perfect reconstruction property (Section VIII). Orthonormality is a special case of biorthogonality, in which DWT and inverse DWT flow graphs are self transposed.

The discrete-time multiresolution theory derived here shares the same properties as the continuous-time multiresolution theory of Mallat [15], [16] and Meyer [17]. These properties include basis expansion, scale and resolution notions, orthogonality or biorthogonality, and regularity. Therefore, it is only a matter of taste to decide whether analog wavelets underly discrete-time ones or vice versa. The DWT provides a coherent alternative, which has advantages and drawbacks. For example, a change of scale is evidently not so easily expressed for discrete sequences that it is for continuous-time signals. The discrete approach, however, avoids technical proofs or makes them easier and readily provides numerical algorithms. Of course, discrete-time multiresolution theory also gives a new way of looking at filter banks; it describes them as a temporal multiresolution decomposition rather than as a subband frequency decomposition.

New criteria in filter design are also brought by wavelets. In particular, there is a, presumably important, notion of regularity. In Section X we have briefly discussed the regularity property in the framework of discrete-time wavelets. Although regular filters have been used in practical systems [2], [15] it is still not clear whether regularity is to play an important role in applications such as image coding.

Appendix A
General Expressions for Scaling Operators

Proof of (6): Let $\{g_n\}$ be the impulse response of the upscaling operator, that is, the upscaled version of the pulse signal δ_n. By assumption 2), g_{n-2k} is the response to δ_{n-k}. Since the input signal can be written

$$x_n = \sum_k x_k \delta_{n-k}$$

its upscaled version is, using linearity 1),

$$\sum_k x_k g_{n-2k}$$

which is (6). □

Proof of (7): Let $\{g_n^0\}$ be the impulse response of the downscaling operator to $\{\delta_n\}$, and $\{g_{n-1}^1\}$ be the impulse response to $\{\delta_{n-1}\}$. The input signal can be written

$$x_n = \sum_k x_{2k} \delta_{n-2k} + x_{2k+1} \delta_{n-1-2k}.$$

Its downscaled version is therefore

$$\sum_k x_{2k} g_{n-k}^0 + x_{2k+1} g_{n-1-k}^1.$$

Now let $\{g'_n\}$ be defined by $g'_{2n} = g_n^0$ and $g'_{2n+1} = g_n^1$. The downscaled version becomes

$$\sum_k x_{2k} g'_{2n-2k} + x_{2k+1} g'_{2n-2k-1}$$

which reduces to (7). □

Appendix B
Derivation of Biorthogonality

From any possible expression of upscaled versions of $\{x_n\}$, simplify using (21) each time this is possible to obtain a unique expression for a version of $\{x_n\}$ at a given scale 2^{-i} and resolution $2^{-j}(j \geq i)$, namely, $(\boldsymbol{G} \uparrow)^{j-i}(\downarrow \boldsymbol{G}')^j x$. Condition (21) is therefore sufficient for uniqueness of a version of x at a given scale and resolution. In addition, since $\boldsymbol{A}$ approximates at half the resolution, it should leave $\boldsymbol{A}x$ itself unchanged. That is, both $\boldsymbol{A}x$ and $\boldsymbol{A}^2x$ are at scale 1 and resolution $1/2$, so they are equal and (22) is a necessary condition for uniqueness. We have seen in Section V that (21) implies (22). Therefore, if we show the converse implication, then both conditions (21) and (22) will be equivalent to uniqueness.

Proof of (22) ⇒ (21): Equation (22) can be written, using (20), $\boldsymbol{G} \uparrow y = 0$, where $y = \downarrow \boldsymbol{G}'(\boldsymbol{G} \uparrow \downarrow \boldsymbol{G}' - \boldsymbol{I})x$. Now since upscaling is one to one (17), this yields $y = 0$, i.e.,

$$(\downarrow \boldsymbol{G}'\boldsymbol{G} \uparrow - \boldsymbol{I}) \downarrow \boldsymbol{G}'x = 0$$

for all signals x. We now prove that any signal y with finite energy can be written $\downarrow \boldsymbol{G}'x$ for some x. We can always write $y = y' + y''$, where y' belongs to the range of the downscaling operator $\downarrow \boldsymbol{G}'$ (hence it can be put in the form $\downarrow \boldsymbol{G}'x$ for some x), and where y'' is orthogonal to any signal of this range:

$$\langle y'', \downarrow \boldsymbol{G}'z \rangle = 0 \qquad \text{for all signals } z.$$

Using the transposition property (15) one obtains

$$\langle \tilde{\boldsymbol{G}}' \uparrow y'', z \rangle = 0 \qquad \text{for all } z$$

hence $\tilde{\boldsymbol{G}}' \uparrow y'' = 0$, which implies $y'' = 0$ similarly as for $\boldsymbol{G} \uparrow$ (17). Therefore any signal y can be written $y = y' = \downarrow \boldsymbol{G}' x$ for some x. Now from the equation $(\downarrow \boldsymbol{G}' \boldsymbol{G} \uparrow - \boldsymbol{I}) \downarrow \boldsymbol{G}' x = 0$, (21) immediately follows. □

This shows that uniqueness is equivalent to (21) and to (22). We now show that biorthogonality is also equivalent to (21).

Proof: Using the transposition property (15) and the relation

$$\langle \delta_{n-k}, \delta_{n-l} \rangle = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{otherwise} \end{cases}$$

we have

$$\begin{aligned} \langle g_{n-2k}, \tilde{g}'_{n-2l} \rangle &= \langle (\boldsymbol{G} \uparrow \delta)_{n-2k}, (\tilde{\boldsymbol{G}}' \uparrow \delta)_{n-2l} \rangle \\ &= \langle (\boldsymbol{G} \uparrow \delta_{n-k}), (\tilde{\boldsymbol{G}}' \uparrow \delta_{n-l}) \rangle \\ &= \langle (\downarrow \boldsymbol{G}' \boldsymbol{G} \uparrow \delta_{n-k}), \delta_{n-l} \rangle \\ &= \langle \delta_{n-k}, \delta_{n-l} \rangle \\ &= \begin{cases} 1 & \text{if } k = l \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

which shows that (23) and (21) are equivalent.

Appendix C
Discrete Wavelet Bases

Proof of (35) and (36): The residue signal at resolution and scale 2^{-j} can be written

$$\begin{aligned} w_k^j &= (\downarrow \boldsymbol{H}' (\downarrow \boldsymbol{G}')^{j-1} x)_k \\ &= \langle (\downarrow \boldsymbol{H}' (\downarrow \boldsymbol{G}')^{j-1} x_n, \delta_{n-k} \rangle. \end{aligned}$$

Using the transposition property (15) we have

$$\begin{aligned} w_k^j &= \langle x_n, (\tilde{\boldsymbol{G}}' \uparrow)^{j-1} \tilde{\boldsymbol{H}}' \uparrow \delta_{n-k} \rangle \\ &= \langle x_n, ((\tilde{\boldsymbol{G}}' \uparrow)^{j-1} \tilde{h}')_{n-2jk} \rangle \end{aligned}$$

which reduces to (35) by definition of $\tilde{h}'^j$ (32). One proves similarly that $v^J = (\downarrow \boldsymbol{G}')^J x$ reduces to (36).

Proof of (37): To reconstruct the signal, the wavelet coefficients w^j, and v^J are brought back to scale 1 and (26) is applied. This gives

$$x = \sum_{j=1}^{J} (\boldsymbol{G} \uparrow)^{j-1} \boldsymbol{H} \uparrow w^j + (\boldsymbol{G} \uparrow)^J v^J.$$

Using the formula $w_n^j = \Sigma_k w_k^j \delta_{n-k}$ (and similarly for v^J), linearity and (33), (34), this equation is easily seen to reduce to (37). □

Appendix D
Perfect Reconstruction and Biorthogonality

Proof of (38) and Biorthogonality (39), Assuming Perfect Reconstruction (30): It is well known [23], [24] that perfect reconstruction of the filter bank of Fig. 5(a) can be written as two conditions on z transforms of filters, namely,

$$\begin{cases} G(z)G'(z) + H(z)H'(z) = 2 & \text{(D1)} \\ G(z)G'(-z) + H(z)H'(-z) = 0. & \text{(D2)} \end{cases}$$

Assuming (noncausal) FIR filters, z transforms are polynomials in z and z^{-1}. Now, consider four polynomial factors of low-pass filters, defined by $G(z) = G_H(z)G_{H'}(z)$ and $G'(z) = G'_H(z)G'_{H'}(z)$, which satisfy, from (D2), $H(z) = G_H(z)G'_H(-z)$ and $H'(z) = -G_{H'}(-z)G'_{H'}(z)$. Consequently, a common factor of (D1) is $G_H(z)G'_{H'}(z)$; both terms divides a constant polynomial, and can therefore be chosen equal to one, $G_H(z) = G'_{H'}(z) = 1$. We therefore end up with $H(z) = G'(-z)$ and $H'(z) = -G(-z)$, which is (38), and perfect reconstruction reduces to

$$G(z)G'(z) - G(-z)G'(-z) = 2$$

which is (23).

We have seen in Section V that (23) is equivalent to $\downarrow \boldsymbol{G}' \boldsymbol{G} \uparrow = \boldsymbol{I}$. From the relations (38) one similarly proves that $\downarrow \boldsymbol{H}' \boldsymbol{H} \uparrow = \boldsymbol{I}$. Now, using definitions (32) and (34), and the transposition property (15), the left-hand side of (39) can be written

$$\begin{aligned} &\langle (\boldsymbol{G} \uparrow)^{j-1} \boldsymbol{H} \uparrow \delta_{n-k}, (\tilde{\boldsymbol{G}}' \uparrow)^{j-1} \tilde{\boldsymbol{H}}' \uparrow \delta_{n-l} \rangle \\ &\quad = \langle \downarrow \boldsymbol{H}' (\downarrow \boldsymbol{G}')^{j-1} (\boldsymbol{G} \uparrow)^{j-1} \boldsymbol{H} \uparrow \delta_{n-k}, \delta_{n-l} \rangle \end{aligned}$$

which reduces to the right-hand side of (39). □

Acknowledgment

The author would like to express his appreciation to I. Daubechies of Rutgers University, A. Grossmann of the University of Marseilles-Luminy, and Y. Meyer of Paris-Dauphine University, for introducing him to mathematical aspects of wavelets when it was very new. He is also grateful to M. Vetterli and C. Herley of Columbia University, New York, for valuable discussions on filter banks, pyramids, and wavelets. Finally, he thanks P. Duhamel of CNET, Paris, for carefully reading the manuscript, and the anonymous reviewers for their valuable feedback.

References

[1] E. H. Adelson, E. Simoncelli, and R. Hingorani, "Orthogonal pyramid transforms for image coding," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 845, pp. 50-58, Oct. 1987.

[2] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using vector quantization in the wavelet transform domain," in *Proc. 1990 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Albuquerque, NM, Apr. 3-6, 1990, pp. 2297-2300.

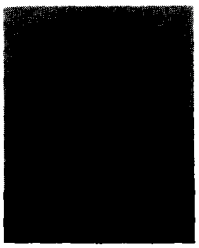
[3] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, no. 4, Apr. 1983.

[4] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, to be published.

[5] R. E. Crochiere and A. V. Oppenheim, "Analysis of linear digital networks," *Proc. IEEE*, vol. 63, pp. 581-595, Apr. 1975.

[6] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[7] R. E. Crochiere, S. A. Weber, and J. L. Flanagan, "Digital coding of speech in subbands," *Bell Syst. Tech. J.*, vol. 55, pp. 1069–1085, Oct. 1976.
[8] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, no. 7, pp. 909–996, 1988.
[9] I. Daubechies, "Orthonormal bases of compactly supported wavelets, II. Variations on a theme," *SIAM J. Math. Anal.*, to be published.
[10] I. Daubechies and J. C. Lagarias, "Two-scale difference equations, I. Existence and global regularity of solutions," *SIAM J. Math. Anal.*, vol. 22, no. 5, pp. 1338–1410, Sept. 1991.
[11] I. Daubechies and J. C. Lagarias, "Two-scale difference equations, II. Local regularity, infinite products of matrices and fractals," *SIAM J. Math. Anal.*, to be published.
[12] P. Goupillaud, A. Grossmann, and J. Morlet, "Cycle-octave and related transforms in seismic signal analysis," *Geoexploration*, vol. 23, pp. 85–102, 1984/1985.
[13] O. Herrmann, "On the approximation problem in nonrecursive digital filter design," *IEEE Trans. Circuit Theory*, vol. CT-18, no. 3, pp. 411–413, May 1971.
[14] W. M. Lawton, "Necessary and sufficient conditions for constructing orthonormal wavelet bases," Aware, Rep. AD900402, Cambridge, MA, 1990.
[15] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, no. 7, pp. 674–693, July 1989.
[16] S. Mallat, "Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$," *Trans. Amer. Math. Soc.*, vol. 315, no. 1, pp. 69–87, Sept. 1989.
[17] Y. Meyer, *Ondelettes et Opérateurs*, Tome I. Paris: Herrmann, 1990.
[18] O. Rioul, "Simple, optimal regularity estimates for wavelets," in *Proc. 6th Eur. Signal Processing Conf. (EUSIPCO'92)*, Brussels, Belgium, August 24–27, 1992.
[19] O. Rioul, "Simple regularity criteria for subdivision schemes," *SIAM J. Math. Anal.*, vol. 23, no. 6, pp. 1544–1576, Nov. 1992.
[20] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continuous wavelet transforms," *IEEE Trans. Inform. Theory* (Special Issue on Wavelet Transforms and Multiresolution Signal Analysis), vol. 38, no. 2, pp. 569–586, Mar. 1992.
[21] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Mag.*, vol. 8, no. 4, pp. 14–38, 1991.
[22] M. J. Shensa, "The discrete wavelet transform: Wedding the á trous and Mallat algorithms," *IEEE Trans. Signal Processing*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992.
[23] M. J. T. Smith and T. P. Barnwell III, "Exact reconstruction techniques for tree structured subband coders," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 3, pp. 434–441, June 1986.
[24] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: A Tutorial," *Proc. IEEE*, vol. 78, no. 1, pp. 56–93, July 1990.
[25] M. Vetterli, "Multidimensional subband coding: Some theory and algorithms," *Signal Processing*, vol. 6, no. 2, pp. 97–112, Feb. 1984.
[26] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE Trans. Signal Processing*, vol. 40, no. 9, pp. 2207–2232, Sept. 1992.
[27] W. R. Zettler, J. Huffman, and D. C. P. Linden, "Application of compactly supported wavelets to image compression," *Proc. SPIE Int. Soc. Opt. Eng.*, vol. 1244, pp. 150–160, Feb. 1990.



Olivier Rioul was born in Strasbourg, France, in 1964. He received the Dipl. Ing. degree from the Ecole Polytechnique, Palaiseau, France, in 1987, the Dipl. Ing. Telecom. degree from the Ecole Nationale Supérieure des Télécommunications (ENST), Paris, in 1989, and the Doctorat d'Ingénieur des Télécommunications (Ph.D.) degree from the ENST, in 1993.

He worked for AT&T Bell Laboratories in 1988. In 1989, he joined the Centre National d'Études des Télécommunications (CNET-France Télécom) in Issy-Les-Moulineaux, France, where he is currently a Research Engineer in the Centre de Recherche en Physique de l'Environnement (Unité mixte CNRS/CNET). His research interests include wavelets, multirate signal processing, computational complexity, image coding, and signal analysis.