From Fourier Transform to Wavelet Transform

Basic Concepts

OCTOBER 27, 1999 WHITE PAPER *Ali M. Reza, Spire Lab, UWM*

Abstract

This white paper briefly describes formulations of the Fourier transform and the wavelet transform applied to signals and systems. Characteristics of discrete Fourier transform (DFT) and its time-dependent version, referred to as discrete short-time Fourier transform (DSTFT), are explained. DSTFT is compared with the wavelet transform in terms of resolution. Wavelet transform is presented as a tool for multiresolution presentation of the signal. In this case, formulation of continuous wavelet transform (CWT), wavelet frame transform (WFT), and discrete wavelet transform (DWT) is presented.

Introduction

The Fourier transform has long been a principle analytical tool in such diverse fields as linear systems, optics, probability theory, quantum physics, antennas, and signal analysis. This mathematical tool originally was used for analysis of continuous signals and systems [1]. Application of digital technology in signals and systems resulted in modifications and development of the Fourier transform for discrete signals and systems. Computational aspects of the Fourier transform were further developed to speedup the computation that is demanded for real applications. Further advances in digital hardware technology, along with high-speed computational algorithm for the Fourier transform, resulted in extensive application of this mathematical tool.

One of the areas that significantly benefited from this advancement is digital signal processing. With present day technology, it is possible to calculate the Fourier transform of a real-time discrete signal; e.g., speech signal or digital videos, process the result in transform domain, and carry out the inverse transform all in real-time. Computational solutions for variety of hardware implementation of the Fourier transform are available. Most of these solutions are in public domain and are free.

The Fourier transform, with its wide range of applications, like many other mathematical tools, has its limitations. For example, this transformation cannot be applied to non-stationary signals. These signals, e.g. speech and image, have different characteristics at different time or space. Although the modified version of the Fourier transform, referred to as short-time (or time-variable) Fourier transform can resolve some of the problems associated with non-stationary signals, but does not address all issues of concern. The short-time Fourier transform is extensively used in speech signal processing but rarely, if ever, used in image processing.

The wavelet transform, which was developed independently on different fronts, is gradually substituting the Fourier transform in some essential signal processing applications. Multiresolution signal processing, used in computer vision; subband coding, developed for speech and image compression; and wavelet series expansions developed in applied mathematics, have been recognized as different views of a single theory. Wavelet transform applies to both continuous and discrete signals. This transformation provides a general technique that is applicable to many tasks in signal processing[2].

The wavelet transform is successfully applied to non-stationary signals for analysis and processing and provides an alternative to the short-time Fourier transform (STFT). In contrast to STFT, which uses a single analysis window, the wavelet transform uses short windows at high frequencies and long windows at low frequencies. This flexibility is introduced in the spirit of so-called “*constant Q*’ or constant relative bandwidth frequency analysis. For some applications it is desirable to obtain the wavelet transform as signal decomposition onto a set of basis functions, referred to as wavelets. These basis functions are obtained from a single prototype wavelet by dilations and contractions (scaling) as well as shifts. Recent surge in application of wavelet transform in various areas of signal processing resulted from the effectiveness of this mathematical tool for analysis and synthesis of signals.

In the following, first different formulations of the Fourier transform are presented. Details of discrete Fourier transform and time-dependent (short-time) Fourier transform are further discussed. Then the wavelet transform is presented and its properties and characteristics for discrete signals are detailed.

Fourier Transform

The essence of the Fourier transform of a waveform is to decompose or separate the waveform into a sum of sinusoids of different frequencies. In other words, the Fourier transform identifies or distinguishes the different frequency sinusoids, and their respective amplitudes, which combine to form an arbitrary waveform. The Fourier transform is then a frequency domain representation of a function. This transform contains exactly the same information as that of the original function; they differ only in the manner of presentation of the information[2]. Fourier analysis allows one to examine a function from another point of view, the transform domain.

Mathematically, this relationship is stated by a pair of equations denoting the forward and inverse transformation. In the case of continuos function, the transform pair, known as Fourier Transform (FT) is given by

$$\begin{aligned} &\text{Forward FT:} \quad X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2p\,f\,t} dt \\ &\text{Inverse FT:} \quad x(t) = \int_{-\infty}^{\infty} X(f) e^{+j2p\,f\,t} df \end{aligned} \tag{1}$$

In above relations, $j = \sqrt{-1}$, $x(t)$ is the continuous function in time and $X(f)$ is its corresponding Fourier transform, which is a continuous function in frequency. This formula is mainly applied to functions with bounded energy. In other words, $x(t)$ should be an energy signal satisfying the following bound.

$$\int_{-\infty}^{\infty} |x(t)|^2 dt < \infty \tag{2}$$

This Fourier transform is mainly used for theoretical analysis and design of continuous signals and systems. For example, when designing an analog filter, the filter frequency response is obtained by applying the Fourier transform to the impulse response of the filter. Also when analyzing an energy signal, the signal spectrum is obtained by using the Fourier transform. This transformation is also used in many areas of applied mathematics like solving differential equations.

In the case of continuos periodic functions, the function does not have a finite energy. If $x(t)$ is periodic with a period of T and fundamental frequency of $f_o = 1/T$, $x(t)$ satisfies $x(t) = x(t+T)$ for all t's, and if it has a finite power, the periodic function can then be expressed as a linear combination of harmonically related sinusoidal functions. The pair of equations, which defines the Fourier series (FS) of a periodic function, is stated by

$$\text{Forward FS:} \quad c_k = \frac{1}{T}\int_{-T/2}^{T/2} x(t)e^{-j2p\,k\,f_o}dt$$
$$\text{Inverse FS:} \quad x(t) = \sum_{k=-\infty}^{\infty} c_k e^{+j2p\,k\,f_o} \tag{3}$$

In above relations, c_k's are Fourier series coefficients of $x(t)$. The condition of having finite power for the periodic function $x(t)$ is stated by the following bound.

$$\frac{1}{T}\int_{-T/2}^{T/2} \left|x(t)\right|^2 dt < \infty \tag{4}$$

This transform converts a continuous periodic function to a sequence of complex numbers. In general, this sequence is infinite. However, in most practical cases, only finite number of c_k's have significant values. The Fourier series expansion of periodic functions or functions with finite support, finite duration, allows analysis and design of signals and systems under some very special conditions. This transformation is also used in many areas of applied mathematics like solving partial differential equations.

Advances in computers and digital technology resulted in design of discrete signals and systems and modifications in the Fourier transform. The Fourier transform that is applied to discrete sequences and referred to as discrete time Fourier transform (DTFT) is defined by the following pair of equations.

$$\text{Forward DTFT:} \quad X\left(e^{j2p\,f}\right) = \sum_{n=-\infty}^{\infty} x[n]e^{-j2p\,n\,f}$$
$$\text{Inverse DTFT:} \quad x[n] = \int_{-p}^{p} X\left(e^{j2p\,f}\right)e^{+j2p\,n\,f}df \tag{5}$$

In above relations, $x[n]$ is the discrete function and $X\left(e^{j2p\,f}\right)$ is its corresponding Fourier transform. The transform function is continuous and periodic in the frequency domain, with the period of $2p$. In this formulation, the frequency variable, f, is normalized by the sampling frequency f_s. In other words, if f_a is the actual frequency in Hz, $f = f_a / f_s$ is the normalized frequency used in (5). This transformation is commonly used for analysis of discrete signals and systems.

Calculation of DTFT by computer can only be carried out for finite sequences and for discrete samples of $X\left(e^{j2p\,f}\right)$ in frequency domain. These requirements and constraints result in another formulation of the Fourier transform that is defined for periodic discrete functions. Let $x[n]$ be a periodic sequence with a period of N; i.e., $x[n] = x[n+N]$ for all n's, the pair of the Fourier transform relations, referred to as discrete Fourier transform (DFT), for $x[n]$, is defined by

$$\text{Forward DFT:} \quad X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2p\,nk/N} \qquad \text{for } k = 0,1,2,\cdots,N-1$$
$$\text{Inverse DFT:} \quad x[n] = \frac{1}{N}\sum_{n=0}^{N-1} X[k]e^{+j2p\,nk/N} \qquad \text{for } n = 0,1,2,\cdots,N-1 \tag{6}$$

In above relations, both discrete functions $x[n]$ and its DFT, $X[k]$, are periodic with the same period N. For graphical purposes and better visual representation of the signal spectrum, the forward DFT can be calculated for $M >> N$ points in frequency domain as follows.

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2p\,n\,k/M} \quad \text{for } k = 0,1,2,\cdots,M-1 \tag{7}$$

This formulation is obtained from the forward DTFT in (5) by using $\Delta f = 1/M$ for the sampling of the normalized frequency.

Although different formulations of the Fourier transform have real application in analyzing signals and systems, but only the last one, DFT relations shown in (6), is practically used in real world computations. Some of the applications of DFT in signal processing are spectrum estimation, feature extraction, and frequency domain filtering. Due to advances in fast computation algorithms for DFT, known as Fast Fourier Transform (FFT)[3], and high-speed hardware implementation, this approach is used for real-time digital signal processing (DSP). It is therefore necessary to address performance and limitation issues of DFT for various applications.

Let $x[n]$ for $n = 0,1,2,\cdots,N-1$, be the sequence of real numbers obtained from sampling an analog temporal signal with sampling period of T seconds. The actual duration of this signal is therefore equal to $T_o = NT$ seconds. When calculating the DFT of this sequence, the resultant sequence, $X[k]$, is in general a complex sequence in frequency domain. The actual distance between frequencies associated to the two consecutive samples of $X[k]$ is $1/NT$ Hertz (Hz). Due to the symmetry properties of $X[k]$ and sampling constraints, center of $X[k]$ sequence corresponds to the maximum frequency of the signal. This frequency is $f_{\max} = (N/2)\cdot(1/NT) = 1/2T$ Hz, which is determined by the sampling period T. Resolution of DFT is fixed at $\Delta f = 1/NT = 1/T_o$ Hz and is depended on the duration of the original analog signal. Increasing number of samples by reducing the sampling period does not change the overall resolution.

One main assumption in using DFT for calculation of the spectrum of a discrete signal is that the observed signal is stationary during the observation time T_o. In other words, the spectrum of the signal is assumed to remain the same during the observation time. For most practical signals, this assumption is not valid. For example, in speech signals, spectrum of the signal may vary significantly from one point to another. This depends on the contents of the speech and the sampling period. In this case and other similar cases, the Fourier transform is modified such that a two-dimensional time-frequency representation of the signal is obtained. The modified Fourier transform referred to as short-time or time-dependent Fourier transform, depends on a window function. For the discrete signals, this transformation, referred to as discrete short-time Fourier transform (DSTFT)[4][5] is obtained by using a window function, $g[\ell]$, where

$$\begin{aligned} g[\ell] &\neq 0 \quad \text{for} \quad 0 \leq \ell \leq L-1 \\ g[\ell] &= 0 \quad \text{for} \quad \ell < 0 \;\text{ or }\; \ell \geq L \end{aligned} \tag{8}$$

The pair of equations, which define the DSTFT of a discrete sequence, is stated by

$$\begin{aligned}\text{Forward DSTFT:}\quad & X[n,k]=\sum_{\ell=0}^{M-1}x[n+\ell]g[\ell]e^{-j2p\,\ell k/N}\\ \text{Inverse DSTFT:}\quad & x[n+\ell]=\frac{1}{Ng[\ell]}\sum_{k=0}^{N-1}X[n,k]e^{-j2p\,\ell k/N}\end{aligned}\tag{9}$$

The index k in (9) is similar to the frequency index in DFT given in (6). The resultant forward Fourier transform in this case provides estimates of the instantaneous frequency spectrum of the signal at any desired time. The window $g[\ell]$ has a stationary origin, and as n changes, the signal slides past the window so that, at each value of n, a different portion of the signal is viewed. The operation detailed in (9) can be carried out by using linear filtering. For example, the $k-\text{th}$ component of the forward transform, $X[n,k]$, can be obtained by filtering $x[n]$ with an FIR filter whose impulse response is

$$h_k[n]=g[-n]\cdot e^{j2p\,kn/N}\tag{10}$$

The main purpose of the window in the time-dependent Fourier transform is to limit the extent of the transformed sequence so that the spectral characteristics are reasonably stationary over the duration of the window function. The more rapidly the signal characteristics change, the shorter the window should be. Resolution in frequency depends on the duration of the window function. In the discrete case and for the uniform window, the actual frequency resolution, in terms of the sampling period T, equals to $\Delta f = 1/LT$ which is the inverse of the actual size of the window. For other shaped windows; e.g. the raised cosine function, the resolution is obtained from $\Delta f \cong a/MT$ in which, $0 < a \le 1$ depends on the shape of the window.

In general the resolution of the DSTFT can be related to the bandwidth of the window sequence. If we use RMS (Root-Mean Square) as a measure of bandwidth, the resultant function for resolution is

$$\Delta f = \left(\sum_{k=0}^{N-1}k^2\left|G[k]\right|^2 \Big/ \sum_{k=0}^{N-1}\left|G[k]\right|^2\right)^{1/2}\tag{11}$$

In (11), $G[k]$ is obtained by calculating DFT of the window sequence as follows.

$$G[k]=\sum\nolimits_n g[n]e^{-j2pnk/N};\ \text{for } k=0,1,2,\cdots,N-1\tag{12}$$

In this approach, two sinusoids will be discriminated only if they are more than Δf apart. Similarly, the spread in time is given by Δt as

$$\Delta t = \left(\sum\nolimits_n n^2\left|g[n]\right|^2 \Big/ \sum\nolimits_n\left|g[n]\right|^2\right)^{1/2}\tag{13}$$

This parameter indicates resolution in time. In other words, two pulses in time can be discriminated only if they are more than Δt apart. As the window becomes shorter, frequency resolution decreases. On the other hand, as the window length decreases, the ability to resolve changes with time increases. Consequently, the choice of window length becomes a trade-off between frequency resolution and time resolution. Resolution in time and frequency cannot be arbitrarily small, because their product is lower bounded.

$$\text{Time-Bandwidth Product} = \Delta t \Delta f \geq \frac{1}{4p} \qquad (14)$$

This is referred to as the uncertainty principle, or Heisenberg inequality. When the window function, $g[\ell]$, is Gaussian the time-bandwidth product satisfy the equality in (14). The general meaning of (14) is that one can only trade time resolution for frequency resolution, or vice versa. We should also point out that, for a fixed window, the time resolution as well as the frequency resolution for all time and frequency instants become fixed. Figure 1 shows the resolution grids on the time-frequency plane.

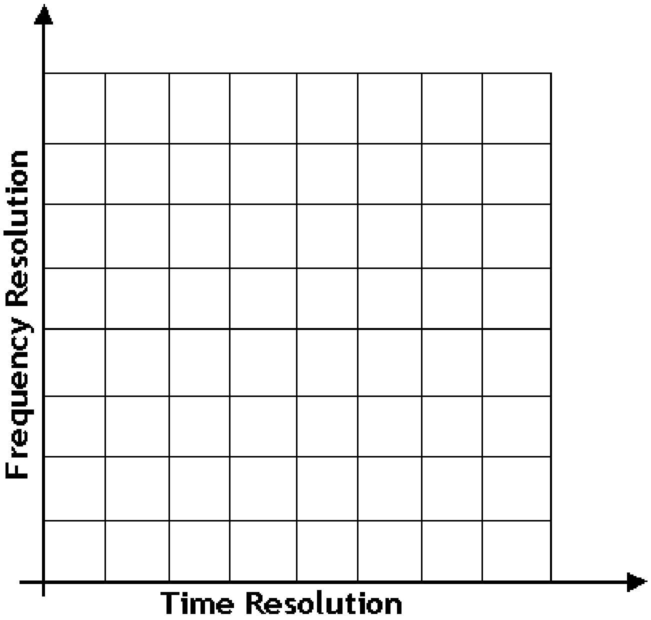

Figure 1 – Time-frequency resolution of STFT

In DSTFT, there is a trade off between the desired resolution in the frequency domain, which is inversely proportional to the actual length of the window in time, and the assumption of short-time stationary. Based on this trade off, the window function is determined. In general, for DSTFT, after deciding about the window function, the frequency and time resolutions are fixed for all frequencies and all times respectively. This approach does not allow any variation in resolutions in terms of time or frequency.

Calculation of DSTFT via filtering approach requires N complex FIR filters to get N components of $X[n,k]$, for $k = 0,1,2,\cdots,N-1$. As an example, the real parts of three of these complex FIR filters are shown in Figure 2. These functions indicate how different frequencies appear inside a fixed window. The number of samples at all frequencies is the same. This number is determined by the size of the window sequence.

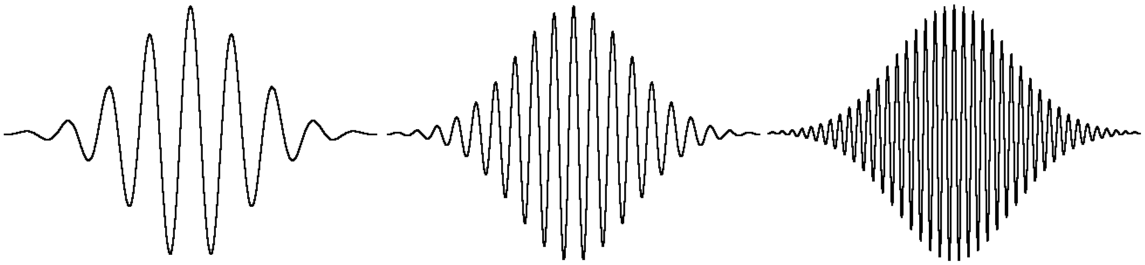

Figure 2 – Real parts of three different filters associated to STFT.

Wavelet Transform

Wavelet transform can be defined for different class of functions. The intention in this transformation is to address some of the shortcomings of the STFT. Instead of fixing the time and the frequency resolutions Δt and Δf , one can let both resolutions vary in time-frequency plane in order to obtain a multiresolution analysis. This variation can be carried out without violating the Heisenberg inequality in (14). In this case, the time resolution must increase as frequency increases and the frequency resolution must increase as frequency decreases. This can be obtained by fixing the ratio of Δf over f to be equal to a constant c :

$$\frac{\Delta f}{f} = c \tag{15}$$

In terms of the filter bank terminology, the analysis filter bank consists of band-pass filters with constant relative bandwidth (so-called "*constant-Q*" analysis). The way that the time-frequency plane is resolved in this approach is as shown in Figure 3. In this case, the frequency responses of the analysis filters in the filter bank are regularly spaced in a logarithmic scale. These filters are naturally distributed into octaves.

With this approach, the time resolution becomes arbitrarily good at high frequencies, while the frequency resolution becomes arbitrarily good at low frequencies. Two very close short bursts can eventually be separated if one goes to higher analysis frequencies in order to increase time resolution. The wavelet analysis, as explained, works best if the signal is composed of high frequency components of short duration plus low frequency components of long duration. The concept of changing resolution at different frequencies can be obtained by introducing what is referred to as "wavelet packets"[2]. Depending on the signal, arbitrary time-frequency resolutions, within the uncertainty bound in (14), can be chosen. As an example, three of these functions for three different frequencies are shown in Figure 4.

The continuous wavelet transform (CWT) is based on the above ideas. In this case, all impulse responses of the analysis filters in the filter bank are defined as scaled (i.e., stretched or compressed) versions of the same prototype $y\,(t)$. For example, for a scale factor of g , the filter impulse response becomes

$$y_g(t) = \left(1 \big/ \sqrt{|g|}\right) y\,(t/g\,) \tag{16}$$

The constant $1 \big/ \sqrt{|g|}$ is used for normalization purpose. Based on this formulation, the CWT is defined as follows.

$$W(t\,,g\,) = \left(1 \big/ \sqrt{|g|}\right) \int x(t) \bar{y} \left[\left(t - t\,\right)/g\, \right] dt \tag{17}$$

In (17), the function $\bar{y}(\cdot)$ is the complex conjugate of the function $y(\cdot)$.

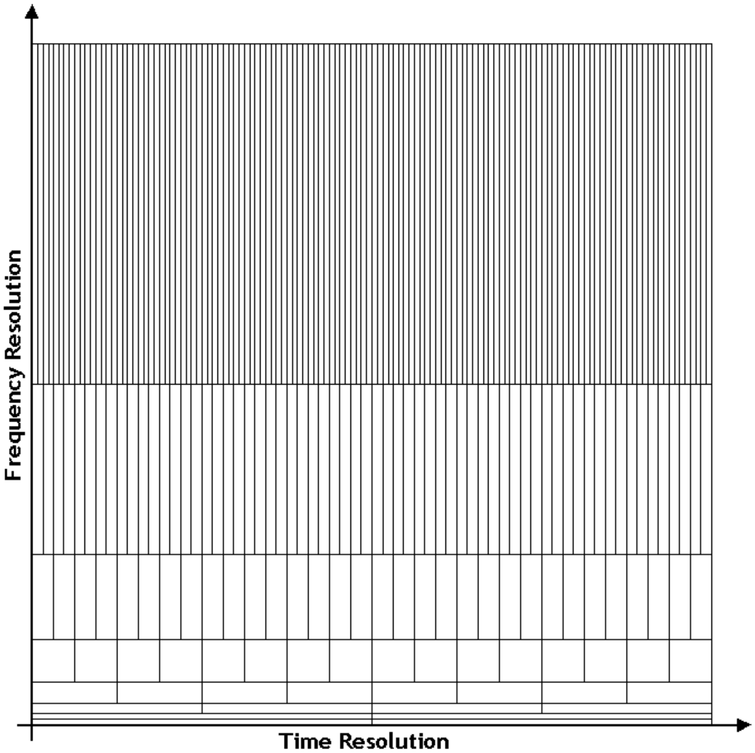

Figure 3 – Time-frequency resolution for constant Q filter bank

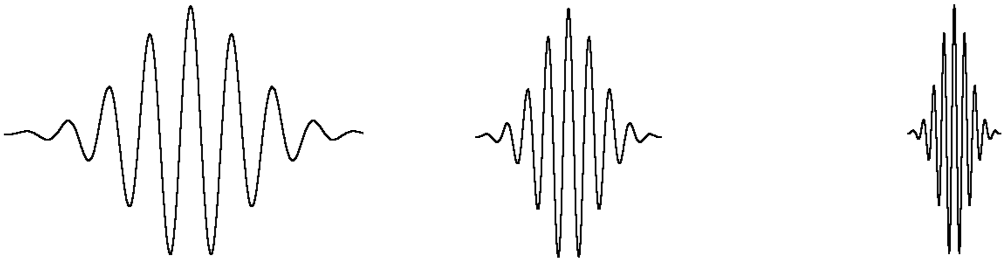

Figure 4 - Three different wavelets with different frequencies and time duration's.

The same prototype, $y(t)$, called the basic wavelet, is used for all the filter impulse responses and no specific scale is privileged, i.e., the wavelet analysis is self-similar at all scales. By choosing a modulated window for the basic filter, as follows, the connection between the STFT and wavelet transform becomes clearer.

$$y(t) = j(t)e^{-j2p f_o t} \tag{18}$$

In this case, the frequency response of the analysis filter, with the general scaling of g, will have its center frequency at $f = f_o / g$. In general, $y(t)$ can be any band pass function. Particularly, we can choose real-valued functions and avoid complex-valued transforms. It is also evident that, the local frequency is associated with the scaling factor g. Consequently, the local frequency, whose definition depends on the basic wavelet, is no longer linked to frequency modulation but is now related to scaling. This is the reason why the terminology "scale" is often preferred to "frequency" for CWT.

Another way to look at the wavelet function in (17) is to consider it as expansion in terms of basis function. Any general signal can be represented as decomposition into wavelets, i.e., the original function is synthesized by adding elementary building blocks, of constant shape but different size and amplitude. In this approach, one can design a set of basis functions by choosing a proper basic wavelet $y(t)$ (mother wavelet) and use a delayed and scaled version of that. By only assuming real functions, the basis function or wavelet at scale g ($g > 0$) and time delay t is obtained from the following equation.

$$y_{g,t}(t) = \frac{1}{\sqrt{g}} y\left(\frac{t-t}{g}\right) \qquad (19)$$

Based on function expansion theory, the forward and inverse continuous wavelet transform (CWT) is written as follows.

$$\begin{aligned} &\text{Forward CWT:} \quad W(t,g) = \int x(t) y_{g,t}(t)dt \\ &\text{Inverse CWT:} \quad x(t) = \frac{c}{g^2}\iint_{g>o} W(t,g) y_{g,t}(t)\, dg\, dt \end{aligned} \qquad (20)$$

In (20), constant c depends on $y(t)$. The reconstruction or inverse transformation is satisfied whenever $y(t)$ is of finite energy and band pass (oscillates in time like a short wave). For sufficiently regular $y(t)$, the reconstruction condition is

$$\int y(t)dt = 0 \qquad (21)$$

Under this condition, the continuously labeled basis functions (wavelets) $y_{g,t}(t)$ behave in the wavelet analysis and synthesis just like an orthonormal basis. By appropriately discretizing the time-scale parameters, t, g, and choosing the right mother wavelet, $y(t)$, it is possible to obtain a true orthonormal basis. The natural way is to discretize the scaling variable g in a logarithmic manner ($g = g_o^{-j}$) and to use Nyquist sampling rule, based on the spectrum of function $x(t)$, to discretize t at any given scale ($t = kg_o^{-j}T$). The resultant wavelet functions are then as follows.

$$y_{j,k}(t) = g_o^{j/2} y(g_o^j t - kT)\ ; \quad \text{where } j \text{ and } k \text{ are integers} \qquad (22)$$

If g_o is close enough to one and if T is small enough, then the wavelet functions are over-complete and signal reconstruction takes place within non-restrictive conditions on $y(t)$. On the other hand, if the sampling is sparse, e.g., the computation is done octave by octave ($g_0 = 2$), a true orthonormal basis will be obtained only for very special choices of $y(t)$.

Based on the assumption that wavelet functions are orthonormal:

$$\langle y_{j,k}(t), y_{\ell,m}(t)\rangle \equiv \int y_{j,k}(t) y_{\ell,m}(t)dt = \begin{cases} 1 & \text{for j} = \ell \text{ and } k = m \\ 0 & \text{for j} \neq \ell \text{ or } k \neq m \end{cases} \qquad (23)$$

the wavelet transform, referenced as wavelet frame transform (WFT) in this case, is obtained as follows.

$$\begin{array}{ll} \text{Forward WFT:} & W[k,j]=\int x(t)y_{j,k}(t)dt \\ \text{Inverse WFT:} & x(t)=\sum_{j}\sum_{k}W[k,j]y_{j,k}(t) \end{array} \tag{24}$$

It is shown that under certain conditions, it is possible to use two different sets of wavelet functions for wavelet frame transform (WFT). In this case, one set is used for analysis or expansion of the function and the second one is used for synthesis or reconstruction of the function. The two sets of wavelet functions should satisfy the biorthogonality condition. Let $y_{j,k}(t)$'s be the set of analysis wavelet functions and $\tilde{y}_{j,k}(t)$'s be the set of synthesis wavelet functions. The WFT transform pair in this case is as follows.

$$\begin{array}{ll} \text{Forward WFT:} & W[k,j]=\int x(t)y_{j,k}(t)dt \\ \text{Inverse WFT:} & x(t)=\sum_{j}\sum_{k}W[k,j]\tilde{y}_{j,k}(t) \end{array} \tag{25}$$

In this case, the biorthogonality condition is specified as follows.

$$\left\langle y_{j,k}(t),\tilde{y}_{\ell,m}(t)\right\rangle \equiv \int y_{j,k}(t)\tilde{y}_{\ell,m}(t)dt=\begin{cases}1 & \text{for j}=\ell \text{ and } k=m \\ 0 & \text{for j}\neq\ell \text{ or } k\neq m\end{cases} \tag{26}$$

For discrete time cases, (22) is generally used with $g_0 = 2$, the computation is done octave by octave. In this case, the basis for a wavelet expansion system is generated from simple *scaling* and *translation*. The generating wavelet or mother wavelet, represented by $y(t)$, results in the following two-dimensional parameterization of $y_{j,k}(t)$'s .

$$y_{j,k}(t)=2^{j/2}y(2^{j}t-k)\ ; \quad \text{for } j,k=1,2,\cdots \tag{27}$$

The $2^{j/2}$ factor in (27) normalizes each wavelet to maintain a constant norm independent of scale j. In this case, the discretizing period in t is normalized to one and is assumed that it is the same as the sampling period of the discrete signal ($t = k2^{-j}$). All useful wavelet systems satisfy the *multiresolution* conditions. In this case, the lower resolution coefficients can be calculated from the higher resolution coefficients by a tree-structured algorithm called *filter-bank* [6]. In wavelet transform literatures, this approach is referred to as discrete wavelet transform (DWT).

In practice, the signal under analysis is a discrete signal, which is assumed to be the highest available resolution for that signal. By using the discrete wavelet transform (DWT), which is exactly the same as the multiresolution analysis in the context of filter banks, all desired lower resolution components can be calculated from the original sampled signal. It is shown that, under the condition of orthogonality, or biorthogonality, the original sampled signal can be obtained from the set of lower resolution coefficients.

The multiresolution idea is better understood by using a function represented by $j(t)$ and referred to as scaling function. A two-dimensional family of functions is generated, similar to (27), from the basic scaling function by

$$j_{j,k}(t) = 2^{j/2} j\,(2^j t - k)\,; \quad \text{for } j,k \in 1,2,\cdots \qquad (28)$$

Any continuous function, $x(t)$, can be represented, at a given resolution or scale j_o, by a sequence of coefficients given in $x_{j_o}(t) = \sum_k x_{j_o}[k] j_{j_o,k}(t)$ expansion. In other words, the sequence $x_{j_o}[k]$ is the set of samples of the continuous function $x(t)$ at resolution j_o. Higher values of j correspond to higher resolution[7].

Discrete signals are assumed samples of continuous signals at known scales or resolutions. In this case, it is not possible to obtain information about higher resolution components of that signal. It is however, desired to use the given samples to obtain the lower resolution representation of the same signal. This can be achieved by imposing some properties on the scaling functions. The main required property is the nesting of the spanned spaces by the scaling functions. In other words, for any integer j, the functional space spanned by $\{j_{j,k}(t);\ \text{for } k \in 1,2,\cdots\}$ should be a subspace of the functional space spanned by $\{j_{j+1,k}(t);\ \text{for } k \in 1,2,\cdots\}$ [8][9].

The nesting of the space spanned by $j\,(2^j t - k)$ is achieved by requiring that $j\,(t)$ be represented by the space spanned by $j\,(2t)$. In this case, the lower resolution function, $j\,(t)$, can be expressed by a weighted sum of shifted version of the same scaling function at the next higher resolution, $j\,(2t)$, as follows.

$$j\,(t) = \sum_k h(k)\sqrt{2}\,j\,(2t - k) \qquad (29)$$

The set of coefficients $h(k)$'s is called the scaling function coefficients (or the scaling filter or the scaling vector) and $\sqrt{2}$ maintains the norm of the scaling function with scale of two. This recursive equation is fundamental to the theory of the scaling function and is referred to by different names in literatures. It is called the refinement equation, the multiresolution analysis equation, or the dilation equation. Design of a wavelet system is the choosing of the set of coefficients represented by $h(k)$'s, in (29).

Important features of a function or signal can be described better by defining a slightly different set of functions that span the difference between the spaces spanned by various scales of the scaling function. It is shown that these functions are the same wavelet functions discussed earlier. Since it is assumed that these wavelets reside in the space spanned by the next narrower scaling function, they can be represented by a weighted sum of shifted version of the scaling function $j\,(2t)$ as follows.

$$y\,(t) = \sum_k h'(k)\sqrt{2}\,j\,(2t - k), \quad k \in 1,2,\cdots \qquad (30)$$

The set of coefficients $h'(k)$'s is called the wavelet function coefficients (or the wavelet filter). Scaling function coefficients, $h(k)$, and wavelet coefficients, $h'(k)$, can be related to each other by some imposing conditions. For example, by forcing the orthogonality condition on the integer translates of the wavelets, and based on the requirement that wavelets should span the "difference" spaces, it is shown[10] that

$$h'(k) = \pm(-1)^k h(1-k) \tag{31}$$

Alternatively, by requiring a mirror symmetry between the two filters frequency responses, in half-band filter banks[11], the relation between these two filters become

$$h'(k) = \pm(-1)^k h(k) \tag{32}$$

The function generated by (30) gives the prototype or mother wavelet $y(t)$ for a class of expansion functions of the form shown in (27).

It is shown[10] that any continuous function can be represented by the following expansion, defined in terms of a given scaling function and its wavelet derivatives:

$$x(t) = \sum_{k=-\infty}^{\infty} c_{j_o}(k) j_{j_o,k}(t) + \sum_{j=j_o}^{\infty} \sum_{k=-\infty}^{\infty} d_j(k) y_{j,k}(t) \tag{33}$$

In this expansion, the first summation gives a function that is a low resolution or coarse approximation of $x(t)$ at scale j_o. For each increasing j in the second summation, a higher or finer resolution function is added, which adds increasing details. The choice of j_o sets the coarsest scale whose space is spanned by $j_{j_o,k}(t)$. The rest of the function is spanned by the wavelets providing the high-resolution details of the function. The set of coefficients in the wavelet expansion represented by (33) is called the *discrete wavelet transform* (DWT) of the function $x(t)$.

These wavelet coefficients, under certain conditions, can completely describe the original function, and in a way similar to Fourier series coefficients, can be used for analysis, description, approximation, and filtering. If the scaling function is well behaved, then at a high scale, samples of the signal are very close to the scaling coefficients.

In order to work directly with the wavelet transform coefficients, we should present the relationship between the expansion coefficients at a given scale in terms of those at one scale higher. This relationship is especially practical by noting the fact that the original signal is usually unknown and only a sampled version of the signal at a given resolution is available. As mentioned before, for well- behaved scaling or wavelet functions, the samples of a discrete signal can approximate the highest achievable scaling coefficients.

It is shown[10] that the scaling and wavelet coefficients at scale j are related to the scaling coefficients at scale $j+1$ by the following two relations.

$$c_j(k) = \sum_m h(m-2k) c_{j+1}(m) \tag{34}$$

$$d_j(k) = \sum_m h'(m-2k) c_{j+1}(m) \tag{35}$$

Equations (34) and (35) state that scaling coefficients at higher scale, along with the wavelet and scaling filters, $h(k)$ and $h'(k)$ respectively, can be used to calculate the wavelet and scaling coefficients or *discrete wavelet transform* coefficients, at lower scales.

In practice, a discrete signal, at its original resolution is assumed the corresponding scaling coefficients. For a given wavelet system, with known wavelet filters $h(k)$ and $h'(k)$, it is possible to use (34) and (35), in a recursive fashion, to calculate the discrete wavelet transform coefficients at all desired lower scales. In most engineering applications, the wavelet systems are chosen such that the two wavelet filters have finite number of non-zero coefficients. In signal processing terminology, these filters are referred to as *finite impulse response* (FIR) filters. Under this assumption, and by using ideas from multirate signal processing literature [6], it is possible to calculate the two summations in (34) and (35) by using two FIR filters. Outputs of these filters are calculated for only even indices and the filters are used with their indices being negated. These differences can be incorporated into the filtering operation by decimation of the output of the filter[6] (in which every other sample is removed from the signal) and by reversing the order of the filter coefficients. In other words, $c_{j-1}(k)$ is obtained by first filtering $c_j(k)$ with a FIR filter whose impulse response is $h_o(k) = h(-k)$ and then decimating the filter output by only keeping every other samples of the output signal. Similarly $d_{j-1}(k)$ is obtained by first filtering $c_j(k)$ with a FIR filter whose impulse response is $h_1(k) = h'(-k)$ and then decimating the result by only keeping every other samples of the output signal[11]. It turns out that always the scaling filter, h_o, and the wavelet filter, h_1, are respectively lowpass and highpass filters.

Mathematically, by assuming that $c_J(k) = x(k)$, the DWT coefficients are obtained by using the following set of equations:

$$\begin{cases} c_{j-1}(k) = \sum_m h(m-2k)c_j(m) \\ d_{j-1}(k) = \sum_m h'(m-2k)c_j(m) \end{cases} \; ; \; j = J, J-1, \cdots, j_o + 1 \tag{36}$$

These calculations are continued until $c_{j_o}(k)$ and $d_{j_o}(k)$ are calculated. Then the collection of these coefficients, namely $\left\{d_{J-1}(k), d_{J-2}(k), \cdots, d_{j_o+1}(k), d_{j_o}(k), c_{j_o}(k)\right\}$, is called DWT of the original signal $x(k)$. Closer look at the number of coefficients obtained for DWT reveals that this number equals to the original number of points in the discrete signal. This is due to the decimation by two that operates in each stage of the process.

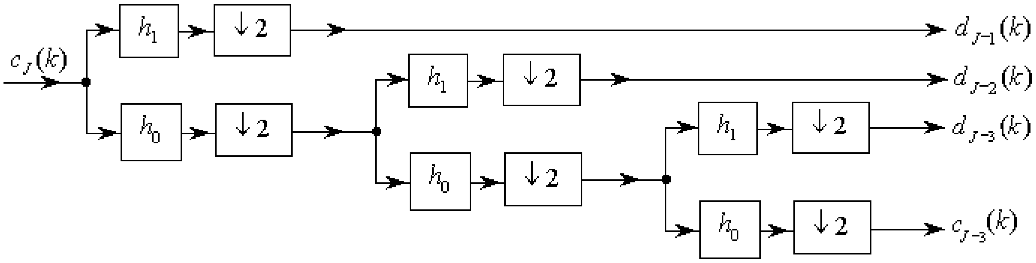

Figure 5 - Three-stage wavelet decomposition, DWT analysis, tree.

Since the pair of filters used in the calculation of DWT is complementing lowpass and highpass filters; the final result provides sequences that are coming from different frequency bands of the original signal. Discrete wavelet transform divides the original signal bandwidth in a logarithmic fashion. The first DWT sequence, $d_{J-1}(k)$'s, are the signal components from the upper half of the signal bandwidth. The lower half of the bandwidth is divided between the rest of DWT sequences. Similarly, the second DWT sequence,

$d_{J-2}(k)$'s, are the signal components from the upper half of the remaining bandwidth of the signal. In other words, this sequence is the one-forth of the total bandwidth and is half of the lower half of the original bandwidth. The remaining bandwidth is similarly divided between the rest of DWT sequences. This idea is better explained in Figure 6. These filters are referred to as analysis filters in filter bank as well as wavelet literatures.

With reference to (33), in which the original continuous signal is written in its wavelet expansion form, mathematical relations for synthesis filters can be derived. It has been shown[10] that the higher resolution scaling coefficients are related to the lower resolution scaling and wavelet coefficients by the following relationship.

$$c_{j+1}(k) = \sum_m c_j(m)\, h(k-2m) + \sum_m d_j(k) h'(k-2m) \qquad (37)$$

This equation indicates how the DWT sequences at resolution j_o can be used, in an iterative fashion, to reconstruct the scaling coefficients at the highest achievable resolution, J .

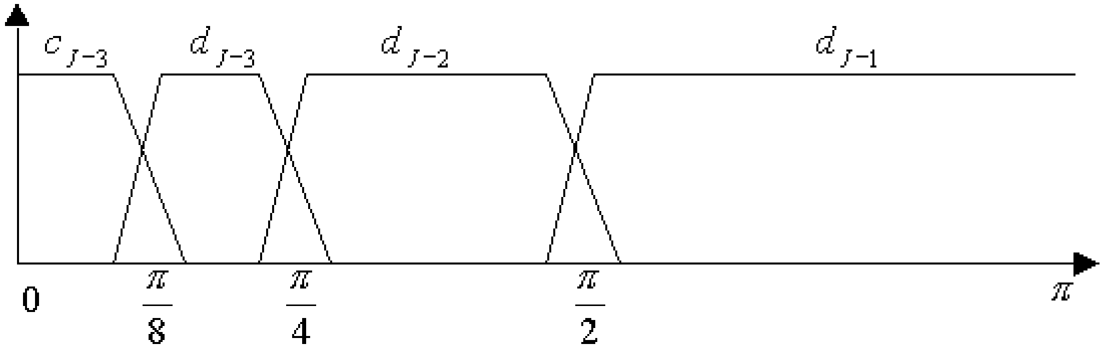

Figure 6 - Logarithmic division of the frequency band between different DWT sequences.

In this case, the way that the two filters are utilized is different from the analysis stage. Here, for any given k , the DWT coefficients are multiplied by every other coefficient of these filters. In other words, the DWT coefficients can be extended by inserting one zero between every two consecutive samples and then normally filter them by using $g_o(k) = h(k)$ and $g_1(k) = h'(k)$ filters. The insertion of zeros between DWT coefficients is represented by $\boxed{\uparrow 2}$, which is the complementary of the decimation operation, shown in Figure 5 by $\boxed{\downarrow 2}$. In this case, the filter coefficients do not need to be reversed in their orders. The three-stage synthesis operation is better explained in Figure 7.

The analysis filter bank efficiently calculates the DWT using banks of digital filters and down-samplers. Similarly, the synthesis filter bank efficiently calculates the inverse DWT by reconstructing the original discrete signal by using up-samplers and banks of digital filters. Properties of the wavelet functions are best developed and understood through the scaling function. If the scaling function has compact support, then the wavelet is composed of a finite sum of scaling functions given in (30). In practical situations, where the wavelet transform is being used as a computational tool in signal processing, the expansion is made finite by using a finite support scaling function.

Researchers in related engineering and applied mathematics areas have developed many different wavelet transform systems each with specific properties. The difference between these wavelet transforms is mainly in their scaling functions and the way that they are developed. As stated before, there are two major classes of wavelet transform systems. One class consists of orthogonal wavelets and the other one consists of biorthogonal wavelets. Other wavelet transform systems, not included in the two main categories, have generally limited applications.

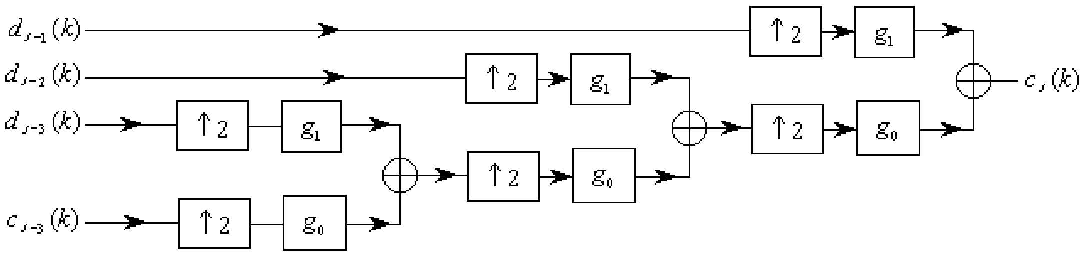

Figure 7 - Three-stage wavelet reconstruction, DWT synthesis tree.

In the orthogonal wavelet systems, knowledge of scaling filter is sufficient for design of the analysis and synthesis filters[11]. For a given even size FIR scaling filter, $h(k)$, we have

$$\begin{cases} g_0(k) = h(k) \\ h_0(k) = g_0(K-1-k) \\ g_1(k) = -(-1)^k h_0(k) \\ h_1(k) = (-1)^k g_0(k) \end{cases} \tag{38}$$

Requiring orthogonality uses up a large number of the degrees of freedom, results in complicated design equations and prevents linear phase analysis and synthesis filter banks. By allowing nonorthogonal basis and dual basis, in biorthogonal wavelet systems, we can attain greater flexibility in achieving other goals.

In this case, by relaxing the orthogonality constraint, a pair of scaling functions and their corresponding scaling filters are introduced[11][12]. If $h(k)$ and $\tilde{h}(k)$ represent the pair of scaling filters, then, they should satisfy

$$\sum_k \tilde{h}(k)h(k-2\ell) = d(\ell) \tag{39}$$

The multiresolution definition for biorthogonal wavelet systems is given as follows:

$$j\,(t) = \sum_k h(k)\sqrt{2}j\,(2t-k), \qquad k \in 1,2,\cdots \tag{40}$$

$$\tilde{j}\,(t) = \sum_k \tilde{h}(k)\sqrt{2}\tilde{j}\,(2t-k), \qquad k \in 1,2,\cdots \tag{41}$$

The biorthogonal wavelets, $y\,(t)$ and $\tilde{y}\,(t)$, are defined by

$$y\,(t) = \sum_k (-1)^k \tilde{h}(1-k)\sqrt{2}j\,(2t-k) = \sum_k g_1(k)\sqrt{2}j\,(2t-k), \qquad k \in 1,2,\cdots \tag{42}$$

$$\tilde{y}\,(t) = \sum_k (-1)^k h(1-k)\sqrt{2}\tilde{j}\,(2t-k) = \sum_k h_1(k)\sqrt{2}\tilde{j}\,(2t-k), \qquad k \in 1,2,\cdots \tag{43}$$

Under some technical conditions, we can expand functions using the wavelets, $y(t)$, and reconstruct them using their dual functions, $\tilde{y}(t)$.

There are many different approaches in design of biorthogonal wavelets. In almost all of them, one desired requirement is to make the resultant wavelet FIR filters to have linear phases. In order to have the perfect reconstruction property[11], in this case, the following conditions should be satisfied.

$$\begin{cases} g_0(k) = h(k) \\ h_0(k) = \tilde{h}(k) \\ g_1(k) = -(-1)^k h_0(k) \\ h_1(k) = (-1)^k g_0(k) \end{cases} \tag{44}$$

In summary, for both cases, with reference to (38) and (44), the forward and inverse wavelet transform for discrete function $x[k]$ is obtained through the following relations:

$$\text{Forward DWT:} \quad \begin{cases} c_J(k) = x(k) \\ \begin{bmatrix} c_{j-1}(k) = \sum\limits_m h_o(m-2k) c_j(m) \\ d_{j-1}(k) = \sum\limits_m h_1(m-2k) c_j(m) \\ \text{for } j = J, J-1, \cdots, j_o + 1 \end{bmatrix} \\ \left\{ d_{J-1}(k), d_{J-2}(k), \cdots, d_{j_o+1}(k), d_{j_o}(k), c_{j_o}(k) \right\} \end{cases} \tag{45}$$

$$\text{Inverse DWT:} \quad \begin{cases} c_{j+1}(k) = \sum\limits_m c_j(m) g_o(k-2m) + \sum\limits_m d_j(k) g_1(k-2m) \\ \text{for } j = j_o, j_o + 1, \cdots, J-1 \\ x(k) = c_J(k) \end{cases}$$

If both biorthogonal filters are restricted to have linear phases, then the lowpass filters, $h_0(k)$ and $g_0(k)$, are symmetric. The highpass filters, $h_1(k)$ and $g_1(k)$, are symmetric when both filters are odd and they are anti-symmetric when both filters are even.

Conclusion

In this white paper, different formulations of the Fourier transform and the wavelet transform are presented. Some of the properties of these transforms are briefly discussed. It is pointed out that the standard Fourier transform is mainly applicable to stationary signals. For non-stationary signals, the time-dependent short-time Fourier transform is presented. Limitation of this formulation with regard to fixed temporal/spatial and frequency resolution is discussed.

The wavelet transform, as a multiresolution technique is presented and compared with Fourier transform in terms of its variable resolution property. Different formulation of the wavelet transform is also discussed. At the end, relevant references to the corresponding literatures are given for interested readers.

Bibliography

[1] Oppenheim, A. V. and Schafer, R. W., *Discrete Signal Processing*. Second Edition, Prentice Hall Inc., Upper Saddle River, New Jersey, 1999.

[2] Riol, O., and Vetterli, M., “Wavelets and signal processing,” IEEE Signal Processing Magazine. Vol. 8, No. 4, pp. 14-38, October 1991.

[3] Brigham, E. O., *The Fast Fourier Transform*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1974.

[4] Allen, J. B. and Rabiner, L. R., “A unified approach to short-time Fourier analysis and synthesis,” *Proceedings of IEEE.* Vol. 65, No. 11, pp. 1558-1564, 1977.

[5] Portnoff, M. R., “Time-frequency representation of digital signals and systems based on short-time Fourier analysis,” *IEEE Transactions on Acoustic, Speech, and Signal Processing*. Vol. 28, pp. 55-69, February 1980.

[6] Vaidyanathan, P. P., *Multirate Systems and Filter banks.* Prentice Hall Inc., Englewood Cliffs, New Jersey, 1993.

[7] Vaidyanathan, P. P. and Djokovic, I., “Wavelet transforms,” in *The Circuits and Filters Handbook*. Edited by W. K. Chen, Chapter 6, pp. 134-219, CRC Press and IEEE Press, Boca Raton, Florida, 1995.

[8] Daubechies, I., *Ten Lectures on Wavelets*. SIAM, Philadelphia, Pennsylvania, 1992.

[9] Mallat, S. G., “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Transactions on Pattern Recognition and Machine Intelligence*. Vol. 11, No. 7, pp. 674-693, July 1989.

[10] Burrus, C. S.; Gopinath, R. A.; and Guo, H., Introduction to Wavelets and Wavelet Transforms: A Primer. Prentice Hall Inc., Upper Saddle River, New Jersey, 1998.

[11] Strang, G. and Nguyen, T., Wavelets and Filter Banks. Wellesley-Cambridge Press, Wellesley, Massachusetts, 1996.

[12] Cohen, A.; Daubechies, I.; and Feauveau, J. C., “Biorthogonal bases of compactly supported wavelets,” *Communications on Pure and Applied Mathematics*. Vol. 45, pp. 485-560, 1992.