An introduction to wavelet transforms: a tutorial approach

C M Leavey, M N James, J Summerscales and R Sutton

Paper submitted 8 November 2002

Non-destructive testing (NDT) and condition monitoring techniques are among the most rapidly developing engineering disciplines. In the past decade methods have been sought to enable real-time or on-line analysis of degradation in structures(1-3)*. Methods developed in the 1980s based on modal analysis*(4) *have resurfaced, partly due to new developments in signal processing. Many of the on-line NDT techniques currently under research utilise wavelet transforms*(5-9) *to decompose spatial signals into the time/frequency domain. The aim of this paper therefore, is to provide a friendly tutorial to the two most commonly used forms of the wavelet transform. It is hoped that this article will provide interested readers with sufficient knowledge of the subject to allow productive usage of the various software packages offering wavelet technologies. Being aware of the building knowledge provided both by literature and the internet, a list of useful references and world-wide-web links has been provided at the end of the paper.*

1. Introduction

The wavelet transform is a relatively new tool to be taken up by the engineering community. Particular strengths lie in its noise and data reduction abilities, which have been exploited in a number of practical applications(10 and 11). These, and many other aspects of the wavelet transform have led to growing interest in the technique by engineers involved in non-destructive testing (NDT). NDT applications are varied, ranging from de-noising(12) to the measurement of dispersion curves for multimode Lamb waves(13 and 14). These rely on the ability of the wavelet transform to decompose a signal into spatially distributed frequency components that can then be selectively filtered according to the requirements of the application.

Lamb wave propagation in plates is generally found to be a depth-related phenomenon, the depth affecting the dispersion properties of the waves. Neithammer and Jacobs(13) have demonstrated that the use of wavelets for crack detection in plates produces superior results to the more commonly used short-time-Fourier-transform (STFT). A similar technique has been applied to the analysis of dynamic strain data(15), where the decomposition of signals taken from beams allowed identification of phase velocities and hence improved accuracy in the determination of crack lengths and location. Wavelets are particularly good in the analysis of noisy data through decomposition of the frequency components contained within a signal. Many applications exist where the wavelet transform has been used to remove unwanted noise from a signal allowing for improved damage identification. Sasi *et al*(16) applied the wavelet transform to analysis of eddy-current data taken from stainless steel cladding tubes. In this instance a discrete version of the wavelet transform was used to improve the signal-to-noise ratio.

Common applications of the discrete versions of the wavelet transform are in data reduction and feature extraction. Due to the physical significance of the wavelet transform in the analysis of oscillating signals, feature extraction and data reduction can often be considered as being the same thing. A wavelet packet approach was adopted by Yang(17) to improve classification of damage in structures. This work found that wavelet packets were particularly good at identifying characteristics relating to damage. For a number of damage scenarios, wavelets provided very high detection rates.

Wavelets are simply mathematical functions exhibiting some kind of oscillatory behaviour. Unlike sinusoidal functions, where the oscillation dominates the entire signal, wavelets show only localised oscillation. Wavelet analysis involves breaking a signal into different frequency components by comparing the signal to a number of differently sized wavelet functions. Wavelets are closely allied to Fourier analysis, but have the advantage of being able to overcome some of their commonly associated constraints. To understand how this is achieved, a review of the issues follows.

Signal processing is the term given to the process of extracting, analysing and modifying information contained within signals. Signal processing, or digital signal processing (DSP), is an extremely important engineering technology spanning many disciplines including image processing, condition monitoring, instrumentation and control, telecommunications and biomedicine(18). Signals generally carry time-varying information such as acceleration, force, temperature and strain. Simple analysis of such data has enabled engineers and scientists to determine derived properties relating to the physical process/system under consideration (for example derivation of Young's modulus from stress/satran data).

Mr Christopher Leavey, BEng(Hons), MPhil, GMIMarEST, MInstP is a research student in the Department of Mechanical and Marine Engineering at the University of Plymouth and a member of the Marine and Industrial Dynamic Analysis Research Group. His main research interests lie in smart structures particularly autonomous condition/health monitoring and life prediction of carbon fibre reinforced plastics.

Professor M Neil James, BSc(Eng), PhD, Hon DTech is currently Professor of Mechanical Engineering and Dean of the Faculty of Technology at the University of Plymouth. He is a Chartered Engineer and a Fellow of the Institute of Materials, Minerals & Mining.

Dr John Summerscales, IEng, FIMMM, FInstNDT, FIAQP is Reader in Composites Engineering in the Department of Mechanical and Marine Engineering at the University of Plymouth. He is a member of the Advanced Composites Manufacturing Centre with research interests in composites manufacturing, process control, total quality management, non-destructive testing, microscopical analysis, natural fibre reinforcements, smart materials and intelligent structures. Contact address: ACMC/DMME, University of Plymouth, Plymouth PL4 8AA, UK. Tel/Fax: +44 1752 23 2650; e-mail: jsummerscales@plymouth.ac.uk

Professor Robert Sutton, BEng(Tech), MEng, PhD, CEng, MIMechE, FIEE is Professor of Control Systems Engineering in the Department of Mechanical and Marine Engineering at the University of Plymouth and a member of the Marine and Industrial Dynamic Analysis Research Group. His main research interests lie in the application of advanced control engineering and artificial intelligence techniques to control problems. Ongoing research is concerned with the use of fuzzy logic, neurofuzzy algorithms, artificial neural networks and adaptive search algorithms in the design of novel control systems for industrial and marine plant.

Contact address: Department of Mechanical and Marine Engineering, University of Plymouth, Drake Circus, Plymouth, Devon PL4 8AA. E-mail: christopher.leavey@students.plymouth.ac.uk

In the analysis of vibration data, time-related changes to a physical variable may not be sufficient to characterise fully the dynamic behaviour of a system. Often the most valuable data in such areas relates to how a system reacts to changes in frequency. In the determination of this *frequency response* data two options are available. Either the system's response to direct sinusoidal excitation can be analysed or, more commonly, the time response can be transformed via mathematical methods to the frequency domain. Fourier showed how simple periodic functions could be represented by an infinite series of sinusoidal functions. His work eventually led to the development of the Fourier transform named after him. The Fourier transform is an extension of the simpler Fourier series, which allows non-periodic signals (such as those of interest in DSP) to be broken down into a finite number of sinusoidal functions of differing amplitude and frequency. The Fourier transform reveals both the frequency components present in the time signal and the magnitude of the oscillations (in other words the frequency response). For a review of the key concepts refer to the literature[18-23]. Despite the continued dominance of the Fourier transform, the technique has a number of well-established constraints. The Fourier transform is defined over infinite time. This means that the effect of transforming a time signal, having finite duration, to the frequency domain is to convolve[1] (*) the Fourier transform of the signal with the Fourier transform of the window function, Figure 1. The frequency spectra of the window function can often prove detrimental to the time signal's frequency representation, reducing the power and introducing additional components in an effect known as leakage. Further frequency components can be introduced due to edge effects where the window function causes discontinuities in the time signal.

An additional limitation on the use of the Fourier transform is the lack of time information in the resulting transform. Fourier techniques were designed for use with stationary signals. A signal is stationary if its statistical properties, for example average and variance, do not change with time. However, most practical signals contain time-varying frequency data and so cannot be classified in this manner. An attempt to solve this time resolution problem has been made through the development of the STFT. STFT uses small time-shifted windows to approximate time/frequency information, giving bands of frequencies over time increments. The accuracy of the STFT in terms of time and frequency resolution is highly dependent on the size of the window selected for the transform. Large windows offer good frequency resolution but poor time resolution while short windows offer good time resolution but poor frequency resolution.

The resolution problem associated with the STFT has been found to be a result of Heisenberg's uncertainty principle[18], which states that time and frequency cannot be resolved simultaneously, *in other words it is impossible to achieve accuracy in both quantities at the same time*. One possible solution to this problem is to adopt a flexible windowing strategy and selectively transform the signal according to the time/frequency requirements. This is the method used in wavelet analysis, a technique formally presented in the 1980s[24], and fast gaining popularity within DSP and in many industrial applications.

Undertaking an analysis using wavelets is a relatively straightforward process. Complications are introduced through the differing approaches adopted and the conditions required to satisfy each implementation. The two most commonly used forms of the wavelet transforms are the continuous wavelet transform (CWT) and the discrete wavelet transform (DWT). Section 2 reviews the CWT providing a simple illustration of how the transformation process is achieved. Having gained a fundamental knowledge of the CWT, the DWT is then explained in section 3. Examples are used throughout this and an attempt is made to relate the transform of the CWT to that of the DWT. To conclude these two sections a note on the selection of appropriate transforms (*CWT or DWT*) and their corresponding wavelet functions is given in section 4. A review of the relevant mathematics associated with the CWT and DWT, including further recommended reading, is given in Appendices I and II.

[1] Convolution is a special kind of 'product' of functions. Convolution has the property that the Fourier (or Laplace) transform of *f(x) *g(x)* is the Fourier transform of *f(x)* multiplied by the Fourier transform of *g(x)*. Reference[22] provides further details of convolution theory.

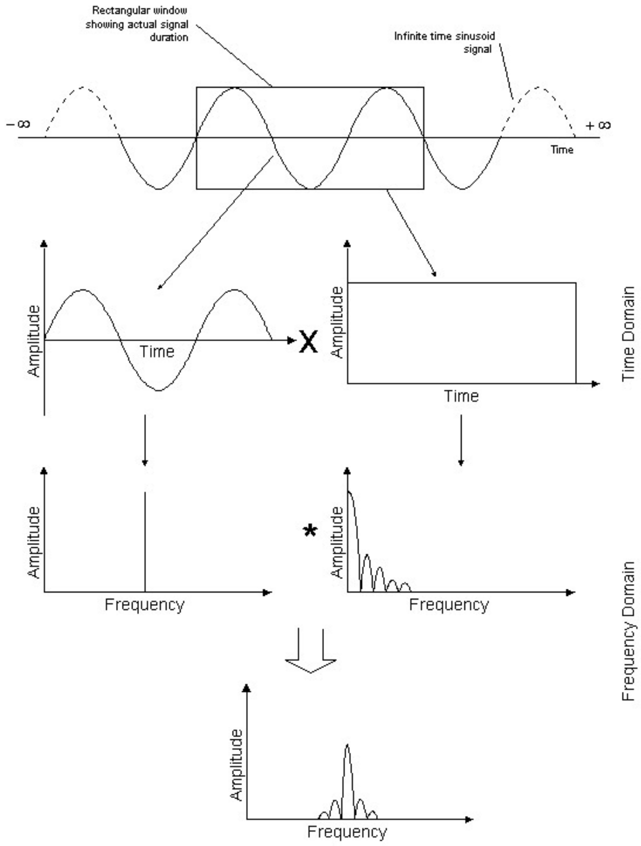

Figure 1. Fourier transform windowing and leakage

2. Overview of the continuous wavelet transform

Of the wavelet transforms available, the CWT is possibly the simplest to visualise, providing a convenient introduction to the subject as a whole. The CWT is similar to the Fourier transform where an arbitrary function of time can be represented by an infinite summation of sinusoidal functions and their multiplicative coefficients. In wavelet analysis sinusoidal functions are replaced with wavelet functions. The essence of the calculation is then to determine the coefficients necessary to accurately portray the time function. (See Appendix I).

The wavelet is a mathematical function usually of time. The terms *wavelet* and *wavelet function* are used interchangeably herein. A wavelet is a small wave: it must be oscillatory and have a limited duration. The use of functions localised in time removes the windowing requirements commonly found in Fourier techniques. However, the term windowing or tiling is still common in wavelet analysis. The leakage effects suffered by the Fourier transform (see section 1 above) caused through windowing are not present in the wavelet transform and therefore signal information is retained. Wavelets must meet a number of mathematical conditions which limit the types of functions that may be used in the analysis. The discrete transform must satisfy a greater number of conditions than the CWT. The reasons for this are given in the following two sections. Three examples of wavelet functions are shown in Figure 2.

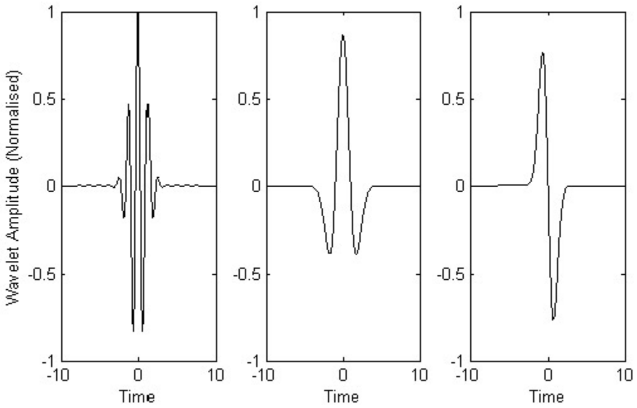

Figure 2. Examples of wavelets – (a) Morlet; (b) Mexican hat; (c) Gaussian

The definition of a wavelet requires an understanding of the process of scaling and translation. To *scale* a wavelet means to stretch or dilate: this ensures that the energy contained within the scaled wavelets is the same as the original '*mother*' wavelet. As the wavelet is stretched in the horizontal x-axis direction it is *squashed* in the vertical y-axis direction. Translation moves the wavelet, usually in the positive direction, along the x-axis. In most cases the x-axis will represent time, while the y-axis is amplitude. In practice, the horizontal axis is not always 'time' but *will always* contain time-varying information. The transform process is achieved through continued scaling and translation of the mother wavelet along the length of a signal.

Before a transform can be performed, a wavelet function must be selected. For the current discussion a generic function will be assumed; for information on wavelet selection see section 4. The term *scale* is often used in relation to the frequency of the wavelet function. Scale is inversely proportional to frequency, so low scale relates to the most tightly packed (high frequency) wavelets. A consequence of this is that typical wavelet decompositions are a function of time and scale as opposed to time and frequency. Wavelet transforms generally begin using low scale (high frequency) wavelet functions progressing to high scales (low frequency) where the wavelet is at its most dilated. The analysing wavelet is set at the beginning of the signal, $t = 0$. The product of the wavelet and signal are then integrated over all time. The results are then normalised giving the wavelet coefficient for that scale and translation. The wavelet is then moved by some small increment in the positive time direction and a new coefficient calculated, this process being continued until the end of the signal is reached. At this point the scale is increased, the wavelet being stretched and dilated, and the analysing signal returned to the starting point, *ie* $t = 0$, allowing calculation of the wavelet coefficients at the next scale. The transformation is complete once the signal has been analysed for all scales, Figure 3.

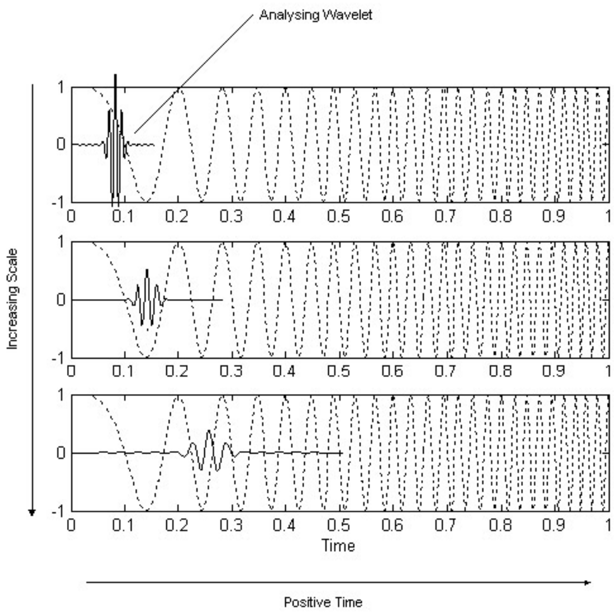

Figure 3. Wavelet analysis overview

The wavelet transform performs a comparison of wavelet to signal. A high degree of similarity exists between the two functions when the coefficients at that translation and scale are large. If the two are dissimilar the coefficients are small. In this way the transform process gives an indication of the frequency content of the signal. As an approximate relationship exists between scale and frequency, a large coefficient at a particular scale implies the presence of a particular frequency. The ability of the wavelet transform to give accurate time and frequency information is constrained by the uncertainty principle in the same way as Fourier transforms. Stated simply, this means that it is not possible to achieve both good time and good frequency resolution simultaneously. This is clearly demonstrated with the Fourier transform where there exists no time information at all due to the infinite window implicit in the technique. The STFT achieves time resolution through the reduction of this window: as the window is reduced, time resolution improves at the expense of frequency resolution.

Analysis of differing scales in the wavelet transform provides flexible windowing. The resulting transform will have varying time/frequency resolution as demonstrated in Figure 4. It is clear from the wavelet transform process that the lower scales (higher frequencies) occupy small windows where the wavelet is highly compacted. As the window size is reduced, time resolution improves at the expense of frequency resolution (as in STFT). Figure 4 illustrates this by the tall thin tiles located at the top of the diagram. The higher scale (lower frequency) components are stretched and occupy a larger window, thus having poorer time resolution and better frequency resolution (again this is similar to STFT with large windows). The flexible tiling scheme adopted by the wavelet transform provides a trade-off between frequency and time resolution. This usually proves advantageous in the analysis of most signals. Real signals tend to exhibit high frequency components for short durations and low frequency components for long durations.

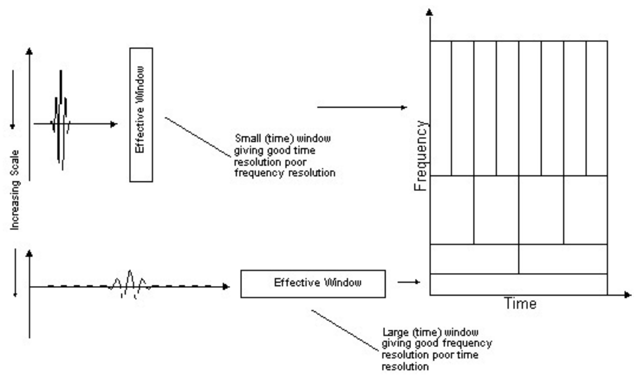

Figure 4. Wavelet analysis flexible windowing scheme

Having briefly examined the CWT transform process the obvious question that remains is *what does a signal look like in the time-frequency/time-scale domain*? Consider a simple sinusoidal signal of finite duration having a frequency of 50 *Hz*: the frequency domain representation is a single vertical line (neglecting leakage), see Figure 5. At any point within the duration of the signal the frequency will be 50 *Hz*, hence in this instance the time-frequency graph of this function can be constructed by projecting the single component of Figure 5 throughout the signal's time duration, Figure 6. Time-frequency and time-scale plots can be displayed in

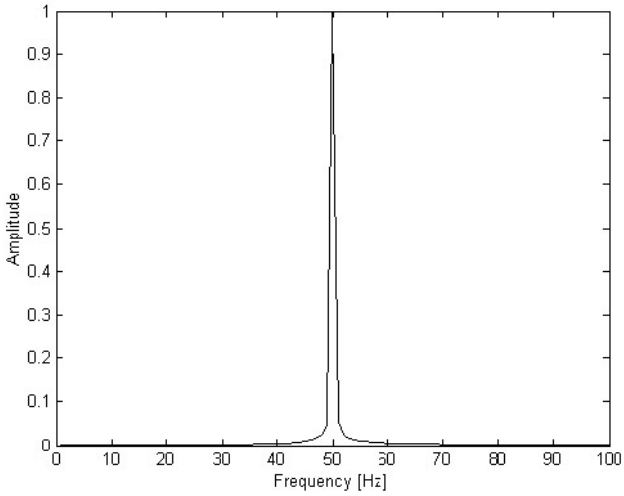

Figure 5. Fourier transform of a 50 Hz sinusoidal function

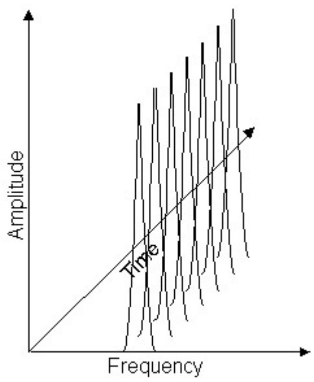

Figure 6. Discrete frequency components with common frequency

either three-dimensions (*time, frequency and amplitude*) or two-dimensions (*time and frequency*). The above example is plotted in two-dimensions, Figure 7. In this instance the axes are time and frequency due to the use of a STFT. For a wavelet decomposition the axes will be time and scale. In Figure 7, high amplitude components are shown in white and zero amplitude in black, grey shading providing intermediate levels. The visualization of a sinusoidal signal with multiple harmonics (sinusoids with frequencies of integer factors greater than the fundamental frequency) can be represented in the same manner as the above example. Each horizontal band relates directly to the frequency and duration of the individual harmonics. Time-scale plots using

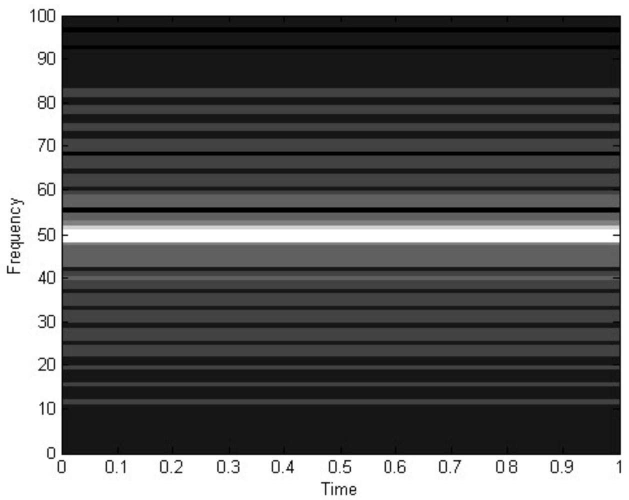

Figure 7. Time-frequency representation of a 50 Hz sinusoid

wavelet analysis are generally quite similar to those produced by using time-frequency methods; there are, however, a number of notable differences as demonstrated in the following example. In this instance the signal to be analysed consists of a low frequency sine wave (50 *Hz*) of duration 1 second, followed by a high frequency sine wave (150 *Hz*). For the STFT and the continuous wavelet transforms the time-frequency/time-scale plots are shown in Figure 8(a) and 8(b). The two figures are inverted as scale is inversely proportional to frequency. The effect of the resolution trade-off discussed previously is also quite clear from the wavelet transform of Figure 8(b). The low frequency components derived from dilated wavelets offer greater frequency resolution *(poorer time resolution)* than the high frequency components. The first harmonic covers a frequency range of approximately 20 *Hz* while the second covers about 70 *Hz*; this scale to frequency was approximated using the following relationship:

$$s = \frac{F_c \cdot \Delta}{f}$$

where f is the frequency, s the scale, Δ is the sample period and F_c the centre frequency defined as an approximate measure of the oscillatory nature of the basis function at its centre. Looking at the STFT of Figure 8(a) there appears to be an anomaly in the diagram, a vertical line separating the two dominant harmonics. This anomaly is a direct result of the time-frequency resolution problem associated with fixed windowing (selection of smaller time windows can go some way to reduce this problem but at the cost of frequency information). Clearly, the solution to the resolution problem offered by the wavelet transform produces additional complexities in the interpretation of the coefficients. The overall effect, however, is beneficial.

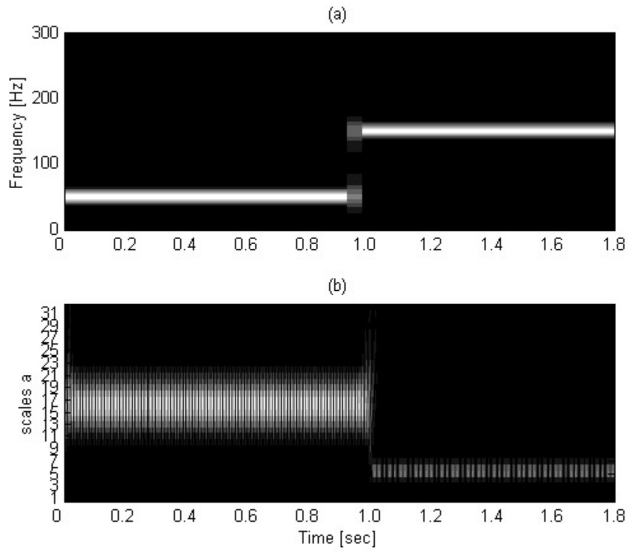

Figure 8. Comparison of STFT and CWT – (a) STFT; (b) Continuous wavelet transform using Morlet basis

3. Overview of the discrete wavelet transform

The above review considered the continuous wavelet transform. This provided a demonstration of some of the key concepts involved in wavelet analysis. The CWT is a computationally demanding algorithm so, as with Fourier techniques, the discrete wavelet transform (DWT) has been developed. The DWT differs from the CWT in that the method of computation utilises subband coding (an alternative technique for computation of the DWT using pyramidal coding can be found in[19]). As there is a great deal of redundancy in the data contained within the CWT, the DWT utilises sampling of both scale and time data, thereby producing a substantially faster algorithm. Scale and time are sampled in powers of two (2^1, 2^2… *etc*); in most texts this is commonly termed

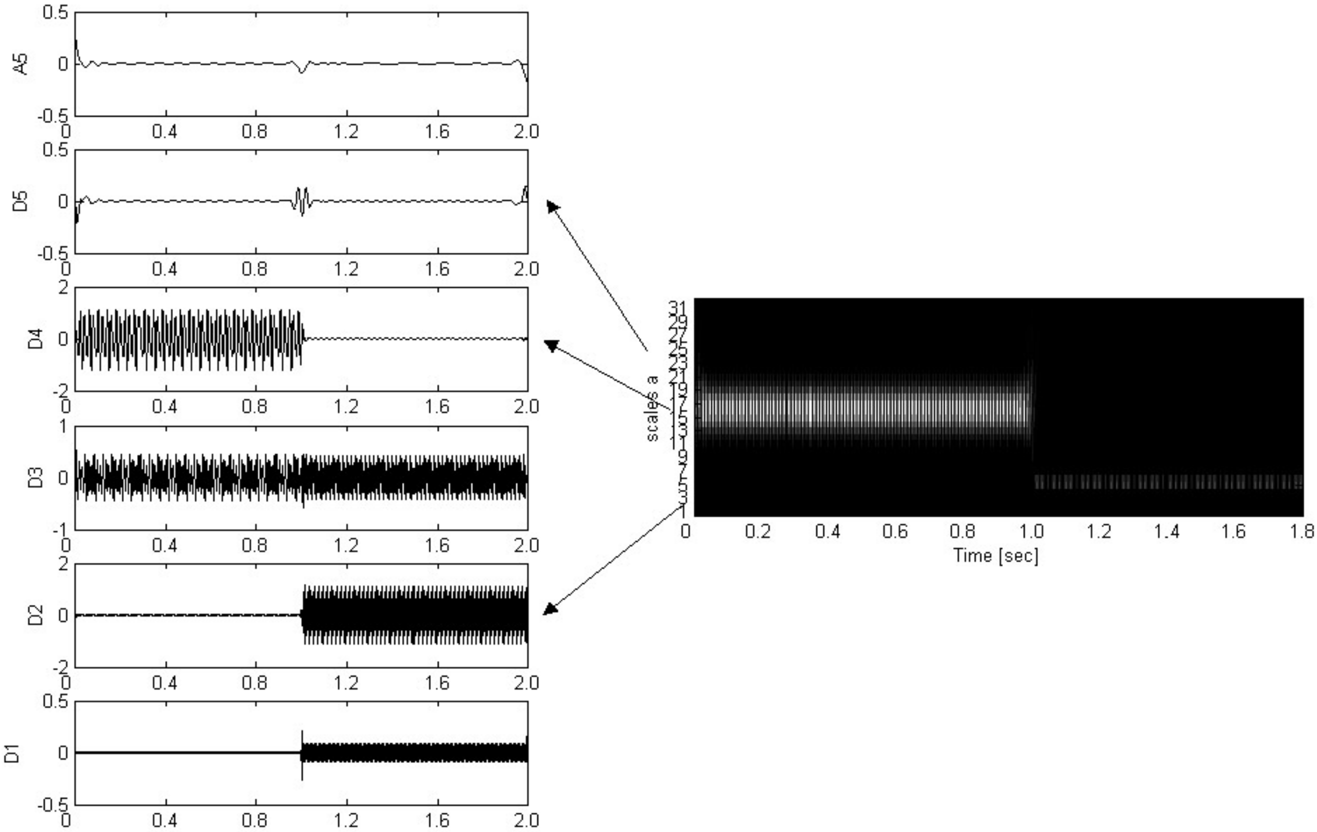

Figure 9. CWT and DWT using Daubechies db7 basis

dyadic sampling. DWT data is generally displayed as a series of plots where each plot represents a specific scale. Using the notion of dyadic sampling these will be scales 2, 4, 8, etc. The data above (Figure 8) reanalysed using a DWT algorithm is shown in Figures 9 and 10, where Figure 10 shows the left-hand-side of Figure 9. The analysing wavelet in this instance was developed by Daubechies[25], known as the 'db7', see Figure 11. (See Appendix II).

Figure 9 demonstrates how the DWT diagrams relate to the CWT diagram (Figure 8). Examining the two diagrams simultaneously it can be seen that the high amplitude parts of the signals relate to the light areas in the CWT. The essential information displayed in the CWT can be fully represented in the six plots of the DWT. The Y-axis notation is 'A' for an approximation and 'D' for a detail; D1 is low scale with increasing numbers for higher scales. The approximation usually shows the underlying trend of the data: the low frequency components. In this instance it is more or less a straight line. The remaining detail levels show how the signal changes between time samples. The DWT contains the same essential information as the CWT (see the lower four graphs of Figure 10, particularly 'D2' and 'D4' which relate to scales of 4 and 16 respectively). As a point of interest, the 'D3' level found in Figure 10 is not present in the CWT of Figure 8 (scale) due to the adoption of a different wavelet function.

Because the DWT uses dyadic sampling it is easy to relate the levels 'D1' to D5' of Figure 10 to the scales shown in Figure 8. The CWT of Figure 8 is analysed over 32 scales, hence the DWT must be analysed over 5 levels: 2^1, $2^2 \ldots 2^5$ giving the samples scales *2, 4…32*. These five levels represent the detail diagrams displayed in

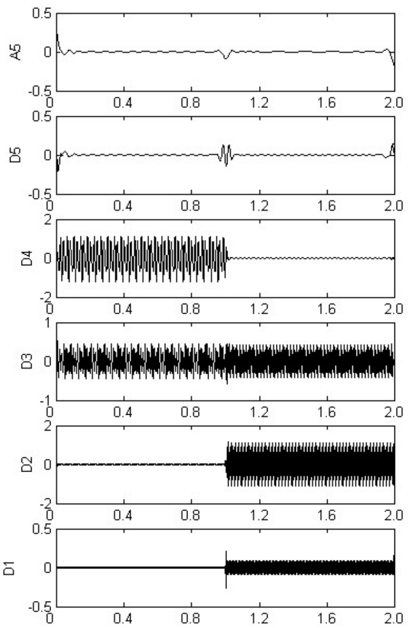

Figure 10. Example of DWT using Daubechies db7 basis

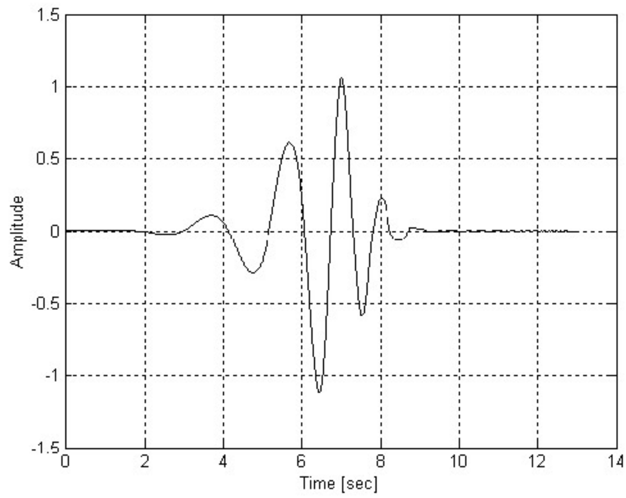

Figure 11. Daubechies db7 wavelet function

Figure 10 where level *D1* corresponds to scale 2 *(2^1)*, *D2* to scale 4 *(2^2)*, *D3* to scale 8 *(2^3)*, etc. The DWT differs from the CWT in that it makes use of a scaling function in addition to the wavelet function previously discussed; this is the reason for the additional diagram in Figure 10. In DWT analysis, the wavelet function is used to determine the detail coefficients while the scaling function determines the approximation coefficients. The nature of the approximation coefficients depends on the level of decomposition that is undertaken. Each level of decomposition removes more high frequency data from the approximation coefficients. In the case of Figure 10 only a small amount of low frequency data has been retained, leaving the general trend of the signal.

In the frequency domain the wavelet function has band-pass properties as shown in Figure 12. Scaling (stretching) of the wavelet function acts to compress the frequency band-pass window, due to a reduction in the wavelet's sampling time, and reduces the band-pass frequencies. Time compression of the wavelet by a factor of 2 will stretch the frequency spectrum of the wavelet by a factor of 2 and also shift all frequency components up by a factor of 2; this is shown diagrammatically in Figure 13 where the wavelet function is represented by the Greek letter ψ. From Figure 13 it is clear that the wavelet transform can be calculated through the dyadic sampling and filtering of the wavelet function. The problem with this is that in order to cover the entire frequency spectra from zero to the upper bounds, an infinite number of scaled wavelet functions would be required. The solution is to produce a function having low-pass properties using the scaling function mentioned previously, see Figure 14. The DWT is therefore calculated by passing the signal through a filter-bank consisting of a high- and low-pass filter: the wavelet function and scaling function respectively. On each pass through the filter bank the scaling function captures the low frequency data from the previous approximation. The effect of dyadic sampling of the scaling function is to progressively reduce the low frequency components contained in each subsequent approximation. The filtering process as a whole is shown in Figures 15 and 16, which demonstrate the actions of both the wavelet and scaling function in the process of decomposition.

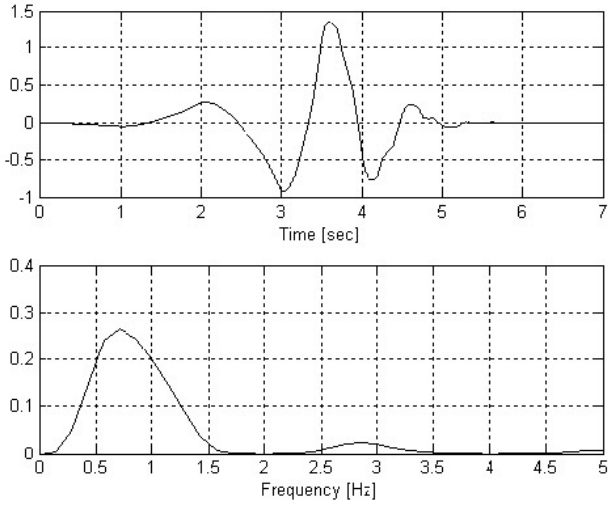

Figure 12. Daubechies db4 wavelet basis and its Fourier transform showing the band-pass nature of the wavelet function

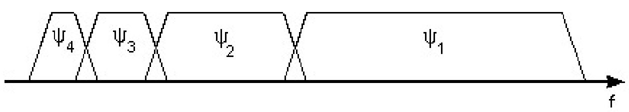

Figure 13. Wavelet spectra resulting from dilation of the mother wavelet (after Ref 25)

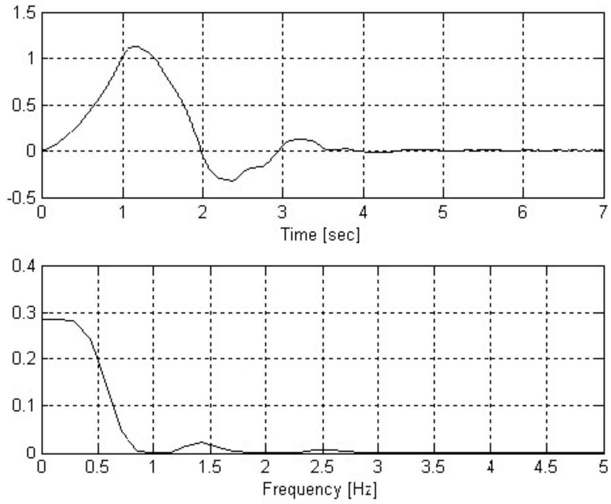

Figure 14. Daubechies db4 scaling function and its Fourier transform showing the low-pass nature of the function

4. A note on transform and wavelet selection

The foundations of wavelet analysis lie in linear algebra and as such some of the terminology is common to both. The wavelet functions are commonly referred to as bases (*basis in singular*). A basis or basis set essentially describes a coordinate system which is linearly independent, allowing a signal or a function to be written as a linear combination of the basis set. A simple example of a commonly used basis set are the unit vectors i, j and k used in three-dimensional Euclidean space. Fourier analysis uses the same principles of linear independence in the construction of an infinite summation of sinusoids, themselves being basis functions. For clarity, the idea of linear independence with reference to the Fourier series means that none of the bases can be constructed from a summation of any of the other bases. Choice of a wavelet can often appear quite arbitrary[27], but should be guided by the following:

Orthogonal or Nonorthogonal – The orthogonality of a wavelet has a significant impact on the type of decomposition that can be performed on a signal. Orthogonality implies that vectors are at right angles to one another. The general definition of the scalar product of two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ is $\boldsymbol{A.B} = \boldsymbol{AB}cos\phi$, where ϕ is the angle between the vectors. If the vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal (in other words at right angles to one another) this product is zero as the angle ϕ will be 90°. Strictly speaking, the use of the CWT does not require an orthogonal basis, although most are at least partially orthogonal. With reference to Figure 8, 9 and 10 it is easy to see that the CWT contains a great deal of information as compared to the DWT, which requires an orthogonal basis (*note that despite the additional information in the CWT representation all the essential detail is fully captured in the DWT*). A lot of redundant information exists in the CWT and subsequently the transform and its inverse are computationally demanding. Through the use of orthogonal bases (DWT) it is possible to reduce the information required for decomposition and reconstruction of the signal and hence provide opportunities for data reduction. This is the approach adopted in image compression and has been successfully used in [28] and [29] to replace the more commonly used cosine transform. Calculation of the CWT can use either orthogonal or nonorthogonal bases. Despite its computational requirement, this algorithm can give a good overview of a signal's features where smooth, continuous variations in wavelet amplitude are expected[27].

Complex or Real – As with Fourier techniques, wavelet bases can be found in real and complex forms. Complex functions provide the opportunity of determining both magnitude and phase information

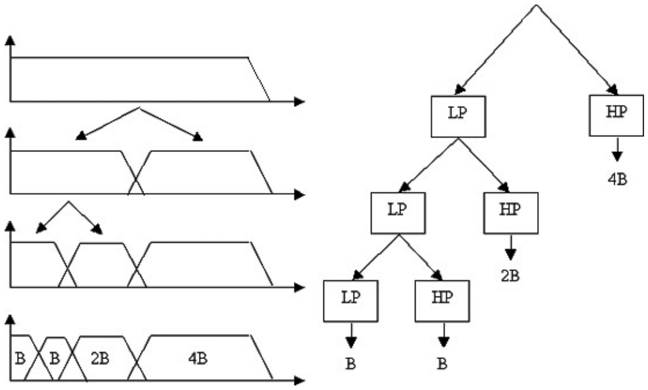

Figure 15. Decomposition of a signal through application of the DWT

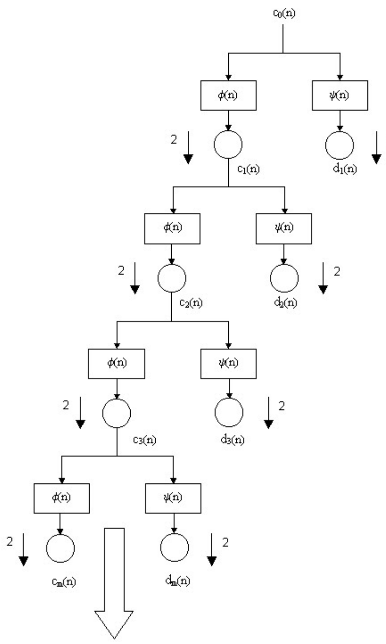

Figure 16. Decomposition process using high(right) and low(left) pass filters

from a signal; this has been suggested to be useful in the analysis of oscillating signals[27]. More recently, complex wavelet functions have found application in signal processing where the difficulty of applying the DWT to consistently process time-shifted signals (*fundamentally the same signal captured at a different point in time*) has prevented the use of pattern recognition techniques[29].

Width – Remembering the windowing problems of the STFT, the width of the wavelet basis has a similar impact on the time-frequency resolution. For a wide basis function, time resolution will be sacrificed at lower scales whereas a short basis function will sacrifice frequency resolution at high scales.

Shape – The shape of the wavelet is one of the most important considerations in the selection of a basis. Generally, the shape of the function should show similar characteristics to the signal being analysed. In some cases a basis can be selected on the function it is to perform, for example the Haar function (see Figure 17) is particularly good at detecting discontinuities in a signal due its own discontinuous nature.

5. Discussion

The wavelet transform (CWT and DWT) is essentially a time/frequency analysis tool which goes some way to solving the resolution problems of Heisenberg's uncertainty principle. There are clearly some notable differences in the two representations highlighted by this review. The CWT offers maximum detail

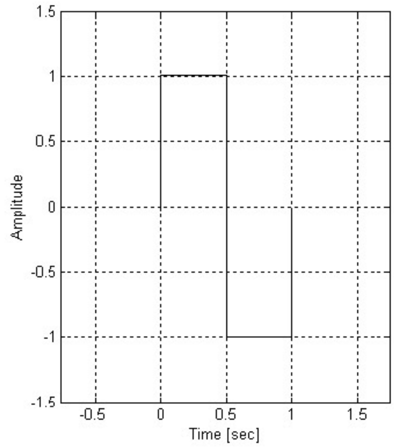

Figure 17. Haar wavelet basis

and time/frequency data but is difficult to use practically due to its computational demands. Where real-time analysis is required, the CWT is unsuitable despite being the more stable of the two algorithms in terms of shift invariance.

The DWT has been suggested for use in data reduction applications and it is clear that in real-time analysis this is the more logical of the two algorithms to use. The shift invariance problem of the algorithm is an area of current research and solutions have been suggested through the use of complex bases[30]. A further difference is clearly evident in the type of basis that can be used with each of the algorithms. The CWT makes use of a single wavelet basis, whereas the DWT requires both wavelet and scaling functions; the DWT also requires that both functions are orthonormal. A further extension to the DWT algorithm is available with wavelet packet analysis. This enables each of the detail coefficients to be filtered in the same way as the approximation coefficients, hence providing flexibility in signal representation with respect to the design objectives.

There are a number of time/frequency algorithms available, complementing the wavelet transforms. The STFT is possibly one of the most well-known time/frequency techniques. STFT can be extremely useful in the analysis of certain signals providing an appropriate balance can be made between time and frequency resolution. The wavelet transform and STFT are both linear transforms obeying superposition; an extremely useful nonlinear algorithm exists in the Wigner-Ville transform[31] which gives instantaneous energy information between time samples. Cohen[32] presents a good review of time/frequency distributions including a discussion of the relative merits of each.

6. Conclusion

An overview of the continuous and discrete wavelet transforms has been given in the above text including where applicable examples of their use in the analysis of signals. A number of references have been given for additional study, in particular detailing the supporting mathematics. This paper attempts to develop and explain some of the key terms attached to wavelet analysis providing a good introductory level text for those new to the subject. (See Appendix I and II).

Acknowledgement

The principal author (CML) is grateful to the EPSRC for support under the Doctorial Training Accounts scheme.

World Wide Web Resources

1. Valens C. 1999, *A Really Friendly Guide to Wavelets*: http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html
2. Institute of Electrical and Electronics Engineers 1995, *An*

Introduction to Wavelets: http://www.amara.com/IEEEwave/IEEEwavelet.html
3. Wavelet Digest: http://www.wavelet.org/wavelet/index.html
4. Polikar R. 1999, The Wavelet Tutorial: http://engineering.rowan.edu/~polikar/WAVELETS/WTtutorial.html
5. Lewalle J. 1995, *Tutorial on Continuous Wavelet Analysis of Experimental Data*: http://www.ecs.syr.edu/faculty/lewalle/tutor/tutor.html
6. Altmann J. 1996, *Surfing the Wavelets*: http://www.wavelet.org/wavelet/tutorial/wavelet.htm
7. Wavelet Forum: http://www.ondelette.com/servlets/com.ondelette.servlet.webforum.WebForumServlet?_FORUMFILE_=forumWavelet.properties
8. Mathsoft's Wavelet Resource Page: http://www.mathsoft.com/wavelets.html

NB: All web links were known to be active on August 2002.

References

1. P Hajela and F J Soeiro, 'Recent Developments in Damage Detection Based on System Identification Methods', Structural Optimization, Vol 2, pp 1-10, 1990.
2. D Armon, Y Ben-Haim and S Brun, 'Crack Detection in Beams by Rank-Ordering of Eigenfrequency Shifts', Mechanical Systems and Signal Processing, Vol 8, No. 1, pp 81-91, 1994.
3. K D Hjelmstad and S Shin, 'Crack Identification in a Cantilever Beam from Modal Response', Journal of Sound and Vibration, Vol 198, No 5, pp 527-545, 1996.
4. C-P Fritzen and K Bohle, 'Model-Based Damage Identification from Changes of Modal Data – A Comparison of Different Methods', Key Engineering Materials, Vols 204-205, pp 65-74, 2001.
5. S Y Huang, G Z Qi and J C S Yang, 'Orthonormal Wavelets for System Identification' 14th Modal Analysis Conference, Vol 1, pp 696-701, 1996.
6. Q Wank and X M Deng, 'Damage Identification with Spatial Wavelets' International Journal of Solids and Structures, Vol 36, pp 3443-3468, 1999.
7. S T Quek, Q Wang, L Zhang and K K Ang, 'Sensitivity Analysis of Crack Detection in Beams by Wavelet Technique', International Journal of Mechnaical Sciences, Vol 43, pp 2899-2910, 2001.
8. Y-F Chang, 'Wavelet Deconvolution Before Scanning in Ultrasonic Non-Destructive Testing', Insight, Vol 44, No 11, pp 694-699, 2002.
9. Q Wang, and X Deng, 'Damage Detection with Spatial Wavelets', International Journal of Solids and Structures, Vol 36, pp 3443-3468, 1999.
10. B Mazieres, 'A new approach for fingerprint image compression,' Tech. Rep. LA-UR-98-4298, Los Alamos Nat'l. Lab, Jan. 1998.
11. C J Lian, K F Chen, H H Chen, L G Chen, 'Lifting based discrete wavelet transform architecture for JPEG2000', 2001 IEEE International Symposium on Circuits and Systems (ISCAS'2001), Sydney, Australia, May 2001.
12. J N Watson, P S Addison and A Sibbald, 'Temporal filtering of NDT data using wavelet transforms', Retrieved February 18, 2003 from the World Wide Web: http://www.ce.utexas.edu/em2000/papers/JNWatson.pdf
13. M Niethammer and J Jacobs, 'Time-frequency representation of Lamb waves using the reassigned spectrogram', Retrieved March 23, 2002 from the World Wide Web: http://www.ce.gatech.edu/~ljacobs/Marc_note_JASA.pdf
14. M A Hamstad, 'An illustrated overview of the use and value of a wavelet transformation to acoustic emission technology', Retrieved February 12, 2003 from the World Wide Web: http://www.vallen.de/zdownload/pdf/hamstad.pdf
15. S T Quek, Q Wang, L Zhang and K H Ong, 'Practical issues in the detection of damage in beams using wavelets', Smart Materials and Structures, Vol 10, pp 1009 – 1017, 2001.
16. B Sasi, B P C Rao, S Thirunavukkarasu, T Jayakumar and P Kalyanasundaram, 'Wavelet transform based method for eddy current testing of cladding tubes', Retrieved January 4, 2003 from the World Wide Web: http://www.nde2002.org/papers/017P.pdf
17. X Yang, 'Reducing measuring points of automatic damage detection system based on wavelet packet analysis under noisy conditions', Retrieved September 9, 2002 from the World Wide Web: http://www.wbmt.tudelft.nl/pto/research/conferences/Proceedings/Sci2001/Paperspdf/P001221.PDF
18. E C Ifeachor and B W Jervis, 'Digital Signal Processing: A Practical Approach', 2nd Edition, Prentice-Hall, 2002, Harlow, England, ISBN:0-201-59619-9.
19. R B Randall, 'Frequency Analysis', 3rd Edition, Brüel and Kjær, 1987, Denmark, ISBN: 87-87355-07-8.
20. D E Newland, 'An Introduction to Random Vibrations, Spectral and Wavelet Analysis', 3rd Edition, 1993, Longman Scientific and Technical, Bath, England, ISBN: 0-582-21584-6.
21. J A Cadzow and H F Van Landingham, 'Signals Systems and Transforms', Prentice-Hall, 1985, Englewood Cliffs, New Jersey, ISBN: 0-13-809542-6.
22. S V Vaseghi, 'Advanced Digital Signal Processing and Noise Reduction', 2nd Edition, Wiley, 2000, Chichester, New York, ISBN: 0-471-62692-9.
23. P V O'Neil 'Advanced Engineering Mathematics', 4th Edition, Brooks/Cole Publishing, 1995, California, USA, ISBN: 0-534-94320-0.
24. P Goupillard, A Grossman and J Morlet, 'Cycle-Octave and Related Transforms in Seismic Signal Analysis', Geoexploration, Vol 23, pp 85-102,1984.
25. I Daubechies, 'Ten Lectures on Wavelets', CBMS-NSF regional conference series in applied mathematics 61, 1992, Philadelphia: SIAM.
26. C Valens, 'A Really Friendly Guide to Wavelets', 1999, Retrieved June 24, 2002 from the World Wide Web: http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html
27. C Torrence and P C Gilbert, 'A Practical Guide to Wavelet Analysis', Bulletin of the American Metrological Society, Vol 79, No 1, pp 61-78, 1998.
28. C M Brislawn, 'Fingerprints go Digital', Notices of the AMS, Vol 42, No 11, pp 1278-1283, 1995.
29. D Lemire, 'White Paper on Image Compression using the Fast Wavelet Transform', Concrete Industries Research Association, 1999.
30. N Kingsbury, 'Complex Wavelets for Shift Invariant Analysis and Filtering of Signals', Applied and Computational Harmonic Analysis, Vol 10, pp 234-253, 2001.
31. J C Goswami and A K Chan, 'Fundamentals of Wavelets Theory, Algorithms and Applications', John Wiley and Sons, 1999, Inc., New York, USA, ISBN: 0-471-19748-3.
32. L Cohen, 'Time-Frequency Distributions – A Review', Proceedings of the IEEE, Vol 77, No 7, pp 941-981, 1989.
33. A D Poularikas, 'The Transforms and Applications Handbook', 2nd Edition, IEEE Press, 1999, Florida, USA, ISBN:0-8493-8595-4.
34. C S Burrus, A G Ramesh and H Guo, 'Introduction to Wavelets and Wavelet Transforms: A Primer', Prentice-Hall, 1998.
35. Y M Hawwar and A M Reza 'Nonlinear Filtering in the Wavelet Transform Domain', IEEE International Conference on Image Processing, Vol 3, pp 266-269, 2000.
36. D B Percival and A T Walden 'Wavelet methods for Time Series Analysis', Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge Press, 2000, New York, USA, ISBN: 0-521-64068-7.

Appendix I – Continuous Wavelet Transform

The bases used in the CWT and DWT, whilst sharing some similarities, demonstrate a number of differences as will be shown below. The CWT's forward transform is defined by the inner product of the signal with the wavelet basis. The inner product, also known as the scalar or dot product, maps the projection of the basis function onto the signal. Equation [1] defines a measure of similarity between the signal and the basis function:

$$W(s,\tau) = \langle f(t), \psi(s,\tau) \rangle = \int f(t)\psi^*(s,\tau)dt \quad [1]$$

where *f(t)* is some function of time (*the signal*), $W(s, \tau)$ are the wavelet coefficients, $\psi(s, \tau)$ is the wavelet basis, s and τ are the scale and translation (time) parameters respectively and * represents complex conjugation. As described in section 2 the wavelet transform is calculated through scaling and translation of the wavelet basis. To achieve scaling of the wavelet function it is divided by the scale parameter s. However, using this alone is insufficient to produce appropriately proportioned energy in the wavelet coefficients. The scaling is therefore normalised producing:

$$\psi(s,\bullet) = \frac{1}{\sqrt{s}}\psi \quad [2]$$

Equation [2] restricts the type of basis that can be used for this type of analysis to those which are square integrable. In other words:

$$\psi \in L^2(\Re^n) \quad [3]$$

and

$$\int \psi^2(x)dx < \infty \quad [4]$$

In fact due to the energy normalisation implied in Equation [2], Equation [4] can be restated as:

$$\int \psi^2(x)dx = 1 \quad [5]$$

The translation parameter shown in Equation [1] can now be introduced to Equation [2]; to ensure that each successive translation occurs at the unit time as opposed to unit scale the scale parameter is also included giving the scaled and translated wavelet:

$$\psi(s,\tau) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \quad [6]$$

Hence [1] can be expressed as,

$$W(s,\tau) = \langle f(t), \psi(s,\tau) \rangle = \frac{1}{\sqrt{s}}\int f(t)\psi^*\left(\frac{t-\tau}{s}\right)dt \quad [7]$$

In order to reconstruct the function *f(t)* from its wavelet coefficients the inverse CWT is defined:

$$f(t) = \frac{1}{C_\psi}\iint W(s,\tau)\frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right)d\tau\frac{ds}{s^2} \quad [8]$$

where:

$$C_\psi = \int \frac{|\psi(\omega)|}{|\omega|}d\omega \quad [9]$$

Note that in equation [9] above that $\psi(\omega)$ is the frequency domain representation of the wavelet basis. Equation [8] leads to one of the most important conditions of the CWT wavelet basis; that of admissibility:

$$\int \frac{|\psi(\omega)|}{|\omega|}d\omega < \infty$$

It is clear from Equations [8] and [9] that in order to recover a signal from its wavelet coefficients C_ψ must be convergent (*admissibility*); this also implies that the Fourier transform, $\psi(\omega)$ evaluated at zero frequency must be zero:

$$\|\psi(\omega)\|_{\omega=0} = 0 \quad [10]$$

Using knowledge of Fourier techniques it is clear that the frequency domain representation of the wavelet is a band pass function localised in frequency. Transformation of a band pass function from the frequency domain to the time domain will always produce a signal with an oscillatory nature. It can therefore be stated from Equations [9] and [10] that the wavelet basis must be oscillatory in order that a signal can be reconstructed using the inverse function defined in Equation [7]. A further condition ensuring the wave-like behaviour of the wavelet function is that it should have a zero mean, in other words:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \quad [11]$$

In order that the CWT avoids the windowing problems inherent in Fourier techniques it must be localised in time. This way information is not lost through transformation due to leakage. Combining this with the results of Equations [9], [10] and [11] leads to the definition of a wavelet namely: a function localised in time with zero mean and having band pass characteristics.

A review of the relevant mathematics relating to the CWT has been given above including the wavelet requirements needed to satisfy decomposition and reconstruction of a signal. The reader wishing to find out more about the CWT is referred to key papers (25, 27, 32), while additional detail of the CWT can be found in books(30, 33, 34).

Appendix II – Discrete Wavelet Transform

The CWT, while being extremely useful for performing exploratory analysis, is a very slow algorithm. From a practical stance there is a need to produce a fast decomposition tool akin to the fast Fourier transform (FFT) which then provides real-time application opportunities. The main problem that exists in the discretization of the CWT is the reconstruction of the signal from the wavelet bases. The DWT forces some additional requirements on the type of basis that can be used in the transform process. The CWT is highly redundant; this redundancy occurs because the CWT analyses a signal at all scales and translations also requiring an infinite number of analysing wavelets. Appropriate sampling of scale and translation and a reduction in the number of wavelets used for decomposition produces a substantially faster algorithm, the DWT. The first step in the generation of the DWT is the discretization of the wavelet bases. Recalling Equation [6], a wavelet is expressed by:

$$\psi(s,\tau) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right)$$

To create a discrete version of Equation [6], both scale and translation are sampled giving:

$$\psi(j,k) = \frac{1}{\sqrt{s_0{}^j}}\psi\left(\frac{t-k\tau_0 s_0{}^j}{s_0{}^j}\right) \quad [12]$$

where j and k are integers, s_0 is a fixed dilation step and τ_0 is a translation factor; t in this instance gives the time steps over which the wavelet is defined. It is clear from Equation [12] that τ_0 is a function of the dilation parameter s_0. Typically s_0 is selected as 2 (dyadic) and τ_0 as 1; in this way both scale and translation are dyadic.

Using dyadic sampling of scale and time means that only special types of bases can be used to reconstruct the signal. To this end Daubechies(25) presents an equivalent admissibility condition for the discrete version of the wavelet transform *vis*:

$$A\|f\|^2 \le \sum \left|\left\langle f, \psi_{j,k} \right\rangle\right|^2 \le B\|f\|^2 \quad \text{....................[13]}$$

Equation [13] defines a frame with bounds A and B. These can be calculated from s_0 and τ_0 and describe the accuracy of the reconstruction. Equation [13] produces a trade-off between accuracy of the reconstruction and the constraints on the basis function: as A and B come closer the reconstruction becomes more accurate and further constrains the selection of the basis. For the discrete case ($A = B$) the wavelet coefficients behave like an orthonormal basis. In this case, for reconstruction to take place the wavelet bases *must* also be orthonormal.

The practical implementation of the DWT makes use of the subband components (section 2). Decomposition is achieved through filtering of the signal using high pass and low pass filters. This filtering technique is known as multiresolution analysis (MRA) because the signal is broken down into discrete frequency bands of varying resolution (recall the time/frequency resolution issue from section 2). The DWT introduces the scaling function providing the low pass filter part of the filter bank. The scaling function $\phi(t)$, is not quite as strictly defined as the wavelet function as it does not need to satisfy admissibility nor need it be oscillatory. However, it must satisfy the condition of orthonormality. Hence, for the scaling function and the wavelet function the following applies:

$$\left\langle \phi_{j,k}(t), \phi_{l,m}(t) \right\rangle = \int \phi_{j,k}(t), \phi^*_{l,m}(t)dt = \begin{cases} 1 \; for \; j = l \; and \; k = m \\ 0 \; for \; j \ne l \; and \; k \ne m \end{cases} \quad \text{..[14]}$$

$$\left\langle \psi_{j,k}(t), \psi_{l,m}(t) \right\rangle = \int \psi_{j,k}(t), \psi^*_{l,m}(t)dt = \begin{cases} 1 \; for \; j = l \; and \; k = m \\ 0 \; for \; j \ne l \; and \; k \ne m \end{cases} \quad \text{..[15]}$$

The wavelet and scaling function are sampled and translated in a dyadic manner as shown in Equation [12]. For implementation these can now be expressed as:

$$\phi_{j,k}(t) = \frac{1}{\sqrt{{s_0}^j}} \cdot \phi\left(2^{-j}t - k\right) \quad \text{......................[16]}$$

$$\psi_{j,k}(t) = \frac{1}{\sqrt{{s_0}^j}} \cdot \psi\left(2^{-j}t - k\right) \quad \text{.....................[17]}$$

Let us define the wavelet coefficients produced through low and high-pass filtering as c_n and d_n, where n represents the level of decomposition. The signal is passed through the scaling and wavelet filters giving the first set of decomposition coefficients c_1 and d_1, these are low and high frequency respectively. The second iteration filters c_1 into c_2 and d_2, the third c_2 into c_3 and d_3 etc. until the appropriate level of decomposition is achieved *ie* to level n. Using equations [16] and [17] and taking the inner product the wavelet coefficients are defined as:

$$c_j = \frac{1}{\sqrt{{s_0}^2}} \sum_n \left\langle \phi_{j,k}, \phi_{j-1,n} \right\rangle c_{j-1} \quad \text{.....................[18]}$$

$$d_j = \frac{1}{\sqrt{{s_0}^2}} \sum_n \left\langle \psi_{j,k}, \psi_{j-1,n} \right\rangle d_{j-1} \quad \text{......................[19]}$$

Using equations [18] and [19] the reconstruction of the signal can be written:

$$f(t) = \sum \frac{1}{\sqrt{{s_0}^2}} c_M \phi\left(2^{-M}t - k\right) + \sum_{i=1}^{M} \sum \frac{1}{\sqrt{{s_0}^2}} d_i \phi\left(2^{-i}t - k\right) \quad \text{..[20]}$$

Equation [20] is the discrete wavelet series decomposition and enables practical implementation of the reconstruction formula. Equations [18] to [20] are the essence of decomposition and reconstruction and result in a transform that is more efficient than the FFT.

A brief overview of the mathematics required for the DWT has been given. For readers wishing to study this area further, reference can be made to the following[25, 27, 33, 35 and 36] which provide detailed studies and proofs.

Nomenclature

Notation

c_i	Discrete wavelet coefficients for ith level decomposition.
d_i	Discrete scaling coefficients of ith level decomposition.
Δ	Sampling period.
F_c	Centre frequency.
$f(t)$	Continuous valued time function representing a signal.
$\phi(\bullet)$	Scaling function.
j	Level (index) for scale in discrete form.
k	Index for time in discrete form.
L^2	Vector space of all square integrable functions.
$\Re^n$	Space of real-valued n dimensional vectors.
s	Scale.
$\psi(\bullet)$	Wavelet function – both continuous and discrete forms.
t	Actual time – continuous.
τ	Constant for time-shifting of wavelet functions.
$W(\bullet)$	Wavelet coefficients.
ω	Frequency.
x	Dummy function variable.

Operations

$\langle \bullet, \bullet \rangle$	Inner product
$\bullet$	Used to represent an arbitrary functional variable.
$*$	Convolution
$\lvert n \rvert$	Modulus or absolute value of n.
$\| \; \|$	Norm or length of a vector.
$\in$	Contained in.