
Chapter 4

Wavelet Transform and Denoising

4.1 Why wavelet

Fourier transform based spectral analysis is the dominant analytical tool for frequency domain analysis. However, Fourier transform cannot provide any information of the spectrum changes with respect to time. Fourier transform assumes the signal is stationary, but PD signal is always non-stationary. To overcome this deficiency, a modified method-short time Fourier transform allows to represent the signal in both time and frequency domain through time windowing function [29]. The window length determines a constant time and frequency resolution. Thus, a shorter time windowing is used in order to capture the transient behavior of a signal; we sacrifice the frequency resolution. The nature of the real PD signals is nonperiodic and transient as shown in Fig 18, Fig 19; such signals cannot easily be analyzed by conventional transforms. So, an alternative mathematical tool- wavelet transform must be selected to extract the relevant

time-amplitude information from a signal. In the meantime, we can improve the signal to noise ratio based on prior knowledge of the signal characteristics.

In this work, we stated only some keys equations and concepts of wavelet transform, more rigorous mathematical treatment of this subject can be found in [30-35]. A continuous-time wavelet transform of $f(t)$ is defined as:

$$CWT_{\psi} f(a, b) = W_f(b, a) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (19)$$

Here $a, b \in R$, $a \neq 0$ and they are dilating and translating coefficients, respectively. The asterisk denotes a complex conjugate. This multiplication of $|a|^{-\frac{1}{2}}$ is for energy normalization purposes so that the transformed signal will have the same energy at every scale. The analysis function $\psi(t)$, the so-called mother wavelet, is scaled by a , so a wavelet analysis is often called a time-scale analysis rather than a time-frequency analysis. The wavelet transform decomposes the signal into different scales with different levels of resolution by dilating a single prototype function, the mother wavelet. Furthermore, a mother wavelet has to satisfy that it has a zero net area, which suggest that the transformation kernel of the wavelet transform is a compactly support function (localized in time), thereby offering the potential to capture the PD spikes which normally occur in a short period of time [36].

4.2 Discrete Wavelet transform and Multiresolution Analysis

One drawback of the CWT is that the representation of the signal is often redundant, since a and b are continuous over R (the real number). The original signal can be completely reconstructed by a sample version of $W_f(b, a)$. Typically, we sample $W_f(b, a)$ in dyadic grid, i.e.,

$$a = 2^{-m} \quad \text{and} \quad b = n 2^{-m} \quad (20)$$

$m, n \in Z$, and Z is the set of positive integers. Substituting (19) into (18), we have

$$DWT_{\psi} f(m, n) = \int_{-\infty}^{\infty} f(t) \psi_{m,n}^*(t) dt$$

where $\psi_{m,n}(t) = 2^{-m} \psi(2^m t - n)$ is the dilated and translated version of the mother wavelet $\psi(t)$.

The family of dilated mother wavelets of selected a and b constitute an orthonormal basis of $L^2(R)$. In addition, we sample $W_f(b, a)$ in dyadic grid, this wavelet transform is also called dyadic-orthonormal wavelet transform. Due to the orthonormal properties, there is no information redundancy in the discrete wavelet transform. In addition, with this choice of a and b , there exists the multiresolution analysis (MRA) algorithm, which decompose a signal into scales with different time and frequency resolution. MRA is designed to give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies.

The fundamental concept involved in MRA is to find the average features and the details of the signal via scalar products with scaling signals and wavelets. In the PD signals we have seen, sharp spikes are observed when PD occurs. The spikes are typically of high frequency and we are able to discriminate the PD spikes with other noises through the decomposition of MRA into different levels.

The differences between different mother wavelet functions (e.g. Haar, Daubechies, Coiflets, Symlet, Biorthogonal and etc.) consist in how these scaling signals and the wavelets are defined. The choice of wavelet determines the final waveform shape; likewise, for Fourier transform, the decomposed waveforms are always sinusoid. To have a unique reconstructed signal from wavelet transform, we need to select the orthogonal wavelets to perform the transforms.

The wavelet decomposition results in levels of approximated and detailed coefficients. The algorithm of wavelet signal decomposition is illustrated in Fig 22. Reconstruction of the signal from the wavelet transform and post processing, the algorithm is shown in Fig 23. This multi-resolution analysis enables us to analyze the signal in different frequency bands; therefore, we could observe any transient in time domain as well as in frequency domain.

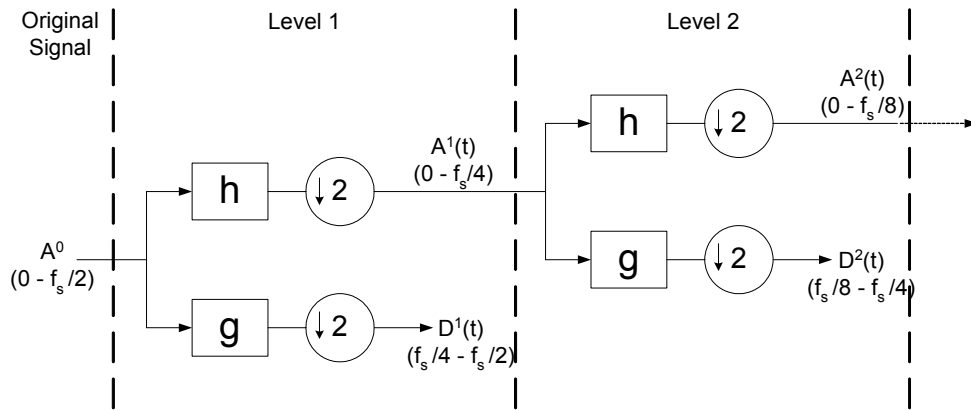


Fig 22. Multi-resolution wavelet decomposition. h = low-pass decomposition filter; g = high-pass decomposition filter; $\downarrow 2$ = down-sampling operation. $A^1(t)$, $A^2(t)$ are the approximated coefficient of the original signal at levels 1, 2 etc. $D^1(t)$, $D^2(t)$ are the detailed coefficient at levels 1,2.

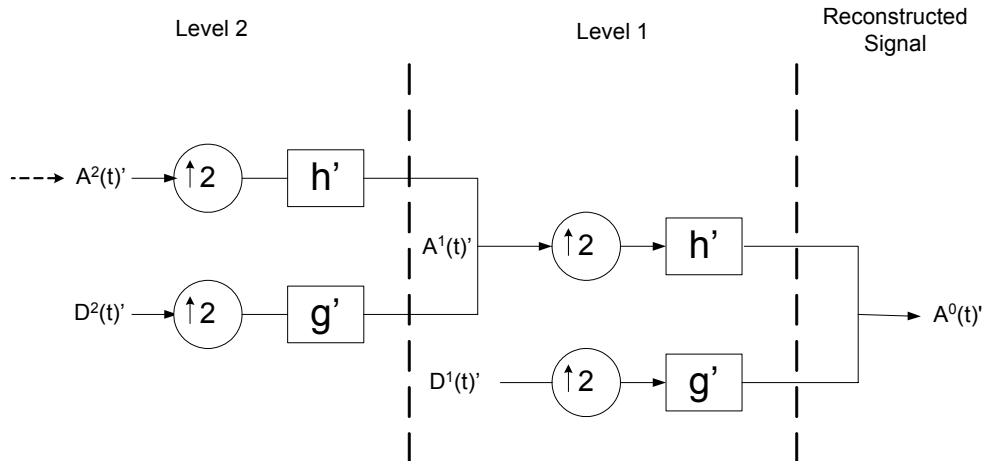


Fig 23. Multi-resolution wavelet reconstruction. h' = low pass reconstruction filter; g' = high-pass reconstruction filter; $\uparrow 2$ = up-sampling operation. $A^1(t)'$, $A^2(t)'$ are the processed or non-processed approximated coefficient of the original signal at levels 1, 2 etc. $D^1(t)'$, $D^2(t)'$ are the processed or non-processed detailed coefficient at levels 1,2.

The relation between the low-pass and high-pass filter and the scalar function $\psi(t)$ and the wavelet $\phi(t)$ can be states as following:

$$\phi(t) = \sum_k h[k]\phi[2t - k]$$

$$\psi(t) = \sum_k g[k]\phi[2t - k]$$

The relation between the low-pass filter and high-pass filter is not independent to each other, they are related by:

$$g[L - 1 - n] = (-1)^n \cdot h[n]$$

where $g[n]$ is the high-pass, $h[n]$ is the low-pass filter, L is the filter length (total number of points). Filters satisfying this condition are commonly used in signal processing, and they are known as the Quadrature Mirror Filters (QMF). The two filtering and down-sampling operation can be expressed by:

$$A^i[k] = \sum_n A^{i-1}(t) \cdot h[2k - n]$$

$$D^i[k] = \sum_n A^{i-1}(t) \cdot g[2k - n]$$

The reconstruction in this case is very easy since the halfband filters form the orthonormal bases. The above procedure is followed in reverse order for the reconstruction. The signals at every level are upsampled by two, passed through the synthesis filters $g^s[n]$, and $h^s[n]$ (highpass and lowpass, respectively), and then added. The interesting point here is that the analysis and synthesis filters are identical to each other, except for a time reversal. Therefore, the reconstruction formula becomes (for each layer)

$$A^i[k] = \sum_{k=-\infty}^{\infty} (D^{i+1}[k] \cdot g[-n + 2k] + A^{i+1}[k] \cdot h[-n + 2k])$$

4.3 Wavelet-based Denoising

The general wavelet denosing procedure is as follows [29, 37, 38]:

- Apply wavelet transform to the noisy signal to produce the noisy wavelet coefficients to the level which we can properly distinguish the PD occurrence.
- Select appropriate threshold limit at each level and threshold method (hard or soft thresholding) to best remove the noises.
- Inverse wavelet transform of the thresholded wavelet coefficients to obtain a denoised signal.

4.3.1 Wavelet selection

To best characterize the PD spikes in a noisy signal, such as Fig 18, and Fig 20, we should select our “mother wavelet” carefully to better approximate and capture the transient spikes of the original signal. “Mother wavelet” will not only determine how well we estimate the original signal in terms of the shape of the PD spikes, but also, it will affect the frequency spectrum of the denoised signal. The choice of mother wavelet can be based on eyeball inspection of the PD spikes [29], or it can be selected based on correlation γ (21) between the signal of interest and the wavelet-denoised signal [38], or based on the cumulative energy (22) over some interval where PD spikes occur.

$$\gamma = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 (Y - \bar{Y})^2}} \quad (21)$$

where \bar{X} and \bar{Y} are the mean value of set X and Y , respectively.

$$E = \sum X^2 \quad (22)$$

where E is the energy and X is the signal vector.

We choose to select the mother wavelet based on the last two methods: correlation between two signals and cumulative energy over some interval of PD spike occurrence. We found that the two methods give us a very similar outcome.

4.3.2 Threshold limits

Many methods for setting the threshold have been proposed. The most time-consuming way is to set the threshold limit on a case-by-case basis. The limit is selected such that satisfactory noise removal is achieved. For a Gaussian noise [39]; if we apply orthogonal wavelet transform to the noise signal, the transformed signal will preserve the Gaussian nature of the noise, which the histogram of the noise will be a symmetrical bell-shaped curve about its mean value. From theory, four times the standard deviation would cover 99.99% of the noise. Therefore, we could set the threshold be 4.5 times of the standard deviation of the wavelet-transformed signal to remove the Gaussian noise in the signal.

We have found that for the fiber optic signals, we could simply apply the standard deviation methods, since the signal is mostly white noises (see Fig 18 and **Appendix B**), however for the PZT signals, we should set the threshold case-by-case to best denoise the signals.

Two rules are generally used for thresholding the wavelet coefficients (soft/hard thresholding). Hard thresholding sets zeros for all wavelet coefficients whose absolute value is less than the specified threshold limit. It has shown that hard thresholding provides an improved signal to noise ratio [29]. In this study, we adopt the hard thresholding method.

4.3.3 Level of Decomposition

From the previous section, we have known that the wavelet transform is constituted by different levels. The maximum level to apply the wavelet transform depends on how many data points contain in a data set, since there is a down-sampling by 2 operation from one level to the next one. In our experience, one factor that affects the number of

level we can reach to achieve the satisfactory noise removal results is the signal-to-noise ratio (SNR) in the original signal. Generally, the measured signals from the PZT sensors have higher SNR than that of the measured signals from fiber optic sensors. So to process the PZT data, we need more level of wavelet transform (e.g. 12) to remove most of its noise. For the fiber optic sensor data, we could only go up to 4 or 5 level otherwise we would remove much of the PD signal, therefore the PD spikes wouldn't be captured.