ELEC9344: Speech and Audio Processing

Time-Frequency Analysis

Chapter 8

Time-Frequency Analysis

- Spectral analysis is an important technique that is utilised widely in many different applications, from control systems to speech processing.

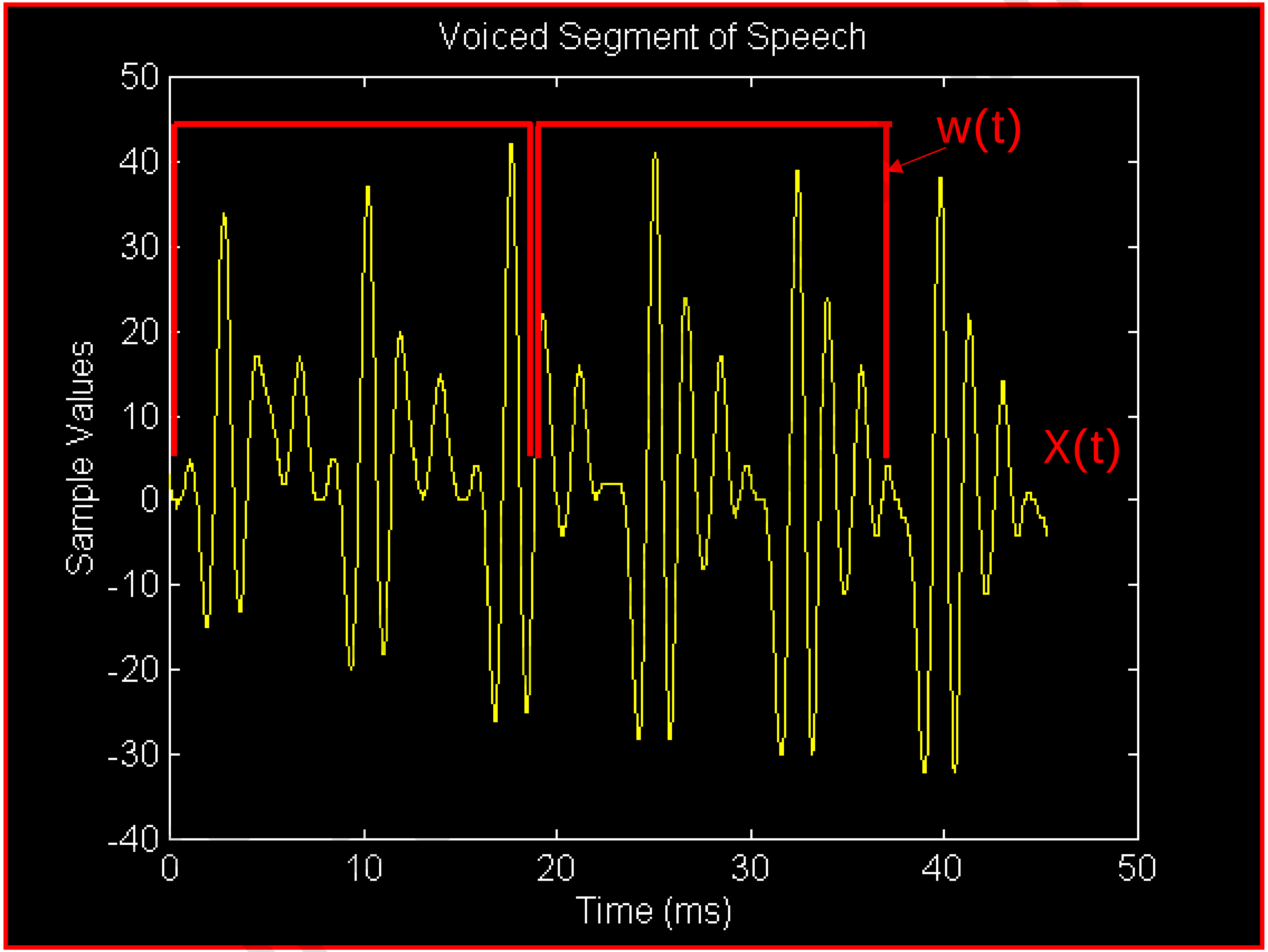
- Traditionally the Short Time Fourier Transform (STFT) has been used to extract the relevant frequency information from the signal requiring analysis in the particular application.

Short-Time Fourier Transform (STFT)

- Conventional STFT-based spectral analysis techniques build a time-frequency representation of a signal by taking the Fourier Transform (FT) of a windowed section of the signal.

- The window is then moved along the signal in time producing a succession of estimates of the spectral components of the signal. This works well for signals composed of stationary components (e.g. sine waves) and for slowly varying signals.

- However, if a signal contains rapidly changing transient events, such as those found in speech, a problem can arise in the use of the STFT.



Voiced Segment of Speech
w(t)
X(t)
Sample Values
Time (ms)
50
40
30
20
10
0
-10
-20
-30
-40
0
10
20
30
40
50

STFT

- To achieve effective temporal localisation of such high frequency events, a short analysis window is required.
- It is a property of the STFT that the frequency resolution is inversely proportional to the window length. This implies that improved time resolution can be achieved only at the expense of poorer frequency resolution.
- Thus, choosing a window function short enough to localise the high frequency transients will reduce the STFT's ability to distinguish between two adjacent frequency components.
- Conversely, choosing a longer window function will give better frequency resolution, but poorer time resolution.
- However, many applications, for example speech processing, require both good frequency resolution and good time resolution, hence a trade-off must be achieved between time and frequency resolution when using the STFT.

STFT

If the window function is defined as w(t) and the signal as s(t), the STFT can be expressed as:

$$STFT = (\omega, b) = \int_{-\infty}^{\infty} x(t)\, w(t-b)\, e^{-j\omega t}\, dt$$

where 'b' is the time shift of the window.

The above equation provides a time-frequency representation of the windowed signal s(t).w(t-b). The function w(t-b) is a shifted version of a standard window function with the window now centred at '*b*' seconds.

For example, a Morlet window is given by the following expression:

$$w(t) = e^{-\frac{1}{2}t^2} \Rightarrow w(t-b) = e^{-\frac{1}{2}(t-b)^2}$$

STFT

- The major problem intrinsic to the use of window function for the STFT namely the fixed time-frequenncy resolution at all points in the time-frequency plane.

- It is the property of the STFT that the frequency resolution is inversely proportional to the window length, which implies that improved time resolution can be achieved only at the expense of poorer frequency resolution

- Thus, choosing a window function short enough to localise the high frequency transients will reduce the STFT's ability to distinguish between two adjacent frequency components. Conversely, choosing a longer window function will give better frequency resolution, but poorer temporal information

Discrete -Time equivalent of STFT (using the Morlet Window)

$$STFT(\theta,b) = \sum_{m=-N/2}^{m=N/2} w(m-b)s(m)e^{-j\theta m}$$

$$STFT(k,b) = \sum_{m=-N/2}^{m=N/2} e^{\frac{-1}{2}(m-b)^2} s(m)e^{-j\frac{2\pi k}{N}m}$$

where N is the window length fixed for STFT and k =0, 1, 2,...N. The shift parameter ‘b’ is the discrete time at which the analysis window is centred. s(m) is the sampled signal.

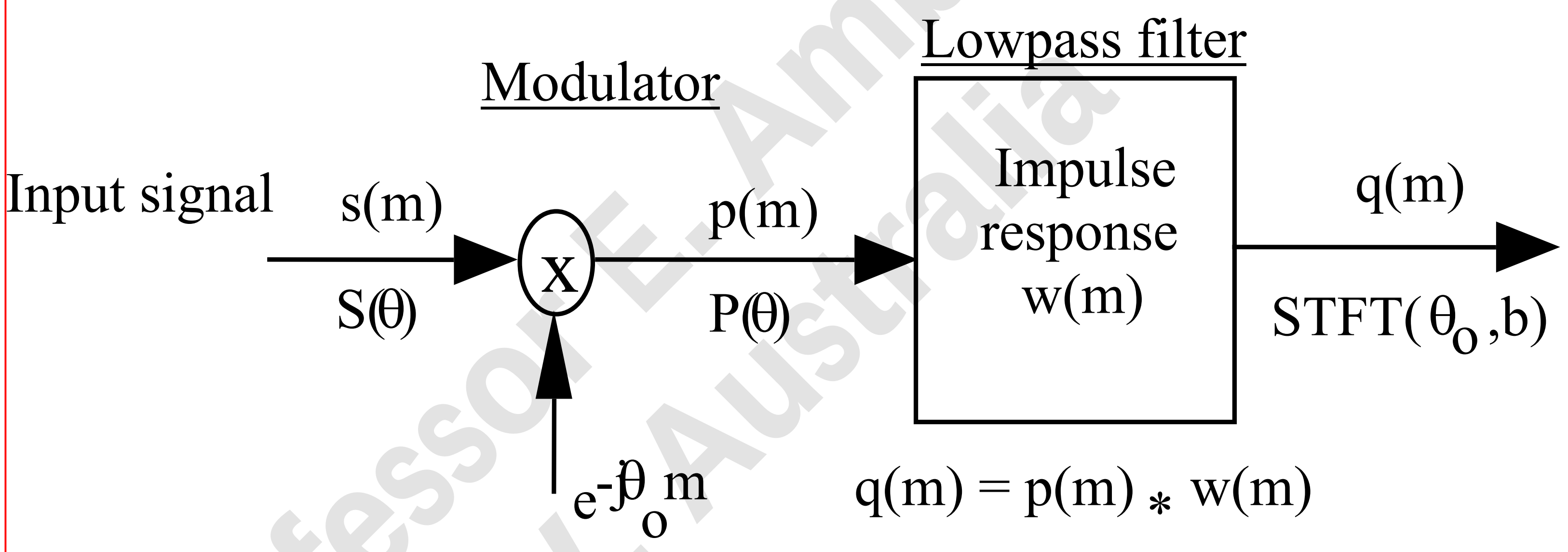
Frequency Resolution of the STFT

$$STFT(\theta,b) = \sum_{m=-N/2}^{m=N/2} w(m-b)\, s(m)e^{-j\theta m}$$

- It is obvious that for each value θ, STFT(θ,n) is the convolution of the sequence s(m).$e^{-j\theta m}$ with the sequence ω(m).
- Therefore for a particular digital frequency θ_0, the STFT(θ_0,n) can be viewed as the convolution of the window impulse response with a frequency shifted version of the signal.

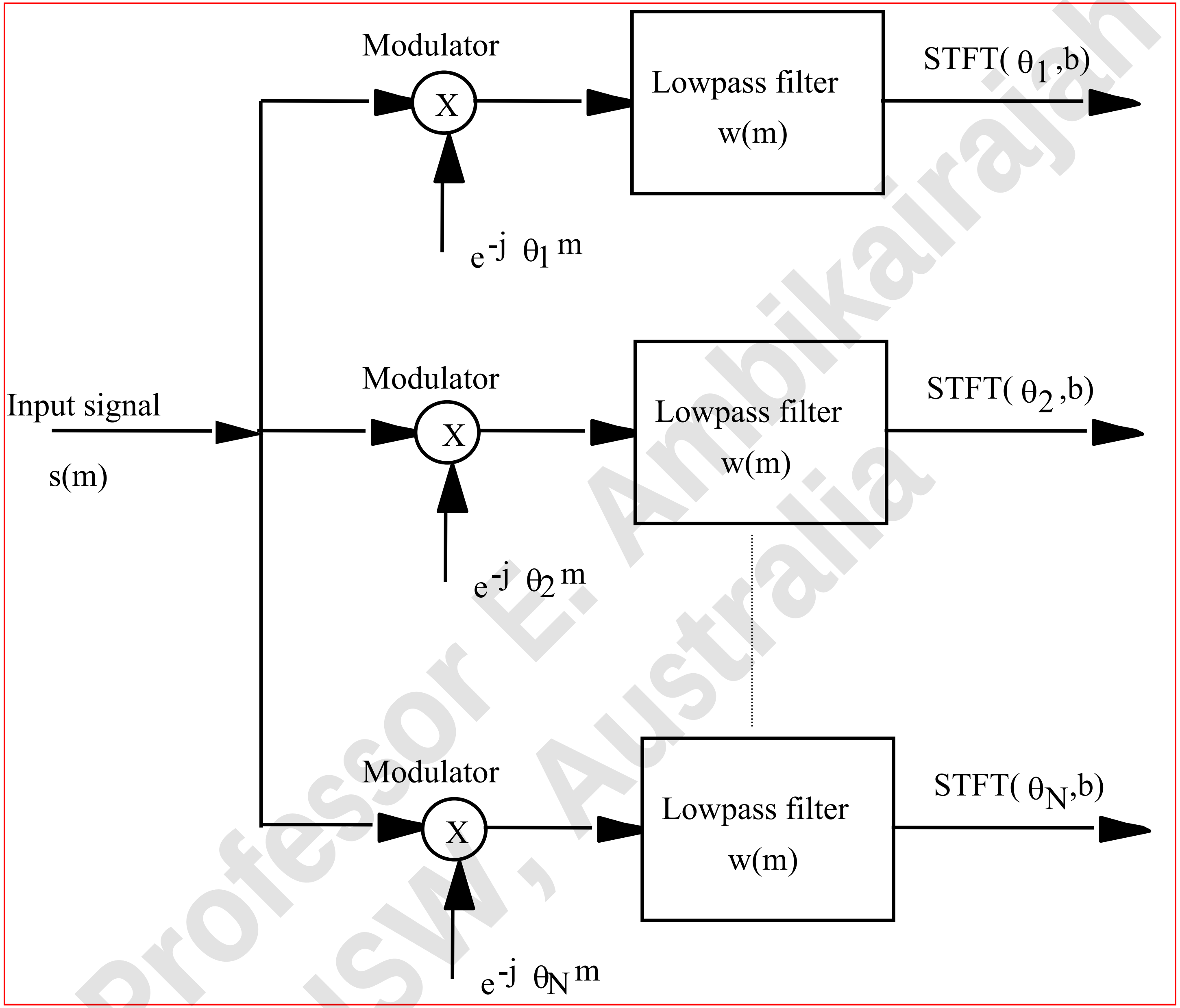
STFT as Modulation and Filtering

$$STFT(\theta,b)=\sum_{m=-N/2}^{m=N/2} w(m-b)\,s(m)e^{-j\theta m}$$



w(m) plays the **role** of the impulse response of a linear shift invariant system. The frequency shifting is caused by the multiplication of the exponential sinusoid .

- By using a bank of these structures and by varying θ_o in the range $-\pi$ to π one can evaluate STFT(θ,b) at various digital frequencies.
- The entire discrete STFT could thus be realised by a bank of modulators and lowpass filters.
- The bandwidth of each of the lowpass filters is identical because they all have the same impulse response.
- The frequency resolution of the STFT will therefore be the same at all frequencies and equal to the lowpass filter bandwidth.
- The STFT will not discriminate between different frequencies which lie within this passband of the lowpass filter.



STFT realised by a bank of modulators and lowpass filters

Time-Frequency Resolution of the STFT

- In time-frequency analysis of a non-stationary signal, it is desirable to maximise both the time resolution (Δt) and the frequency resolution(Δf). This implies minimising the product of Δt. Δf.

- The frequency Resolution is associated with the window function bandwidth and that two sinusoids will be discriminated only if they are more than Δf Hz apart.

- The time resolution is associated with the window function length and that two pulses in time will be discriminated only if they are more than Δt sec apart.

Time-Frequency Resolution of the STFT

- Because both Δt and Δf are controlled by the same window length, it is not possible to decrease one without increasing the other. In fact, the time-bandwidth product $\Delta t . \Delta f$ is lower bounded (Rioul and Vetterli, 1991).

- Time Bandwidth Product = $\Delta t \Delta f \geq (1/4\pi)$

- Once a window has been chosen for the STFT, then both the time resolution and the frequency resolution are fixed. Thus, the window can be chosen to give good frequency resolution or good time resolution, but not both.

Time-Frequency Resolution of the STFT…..

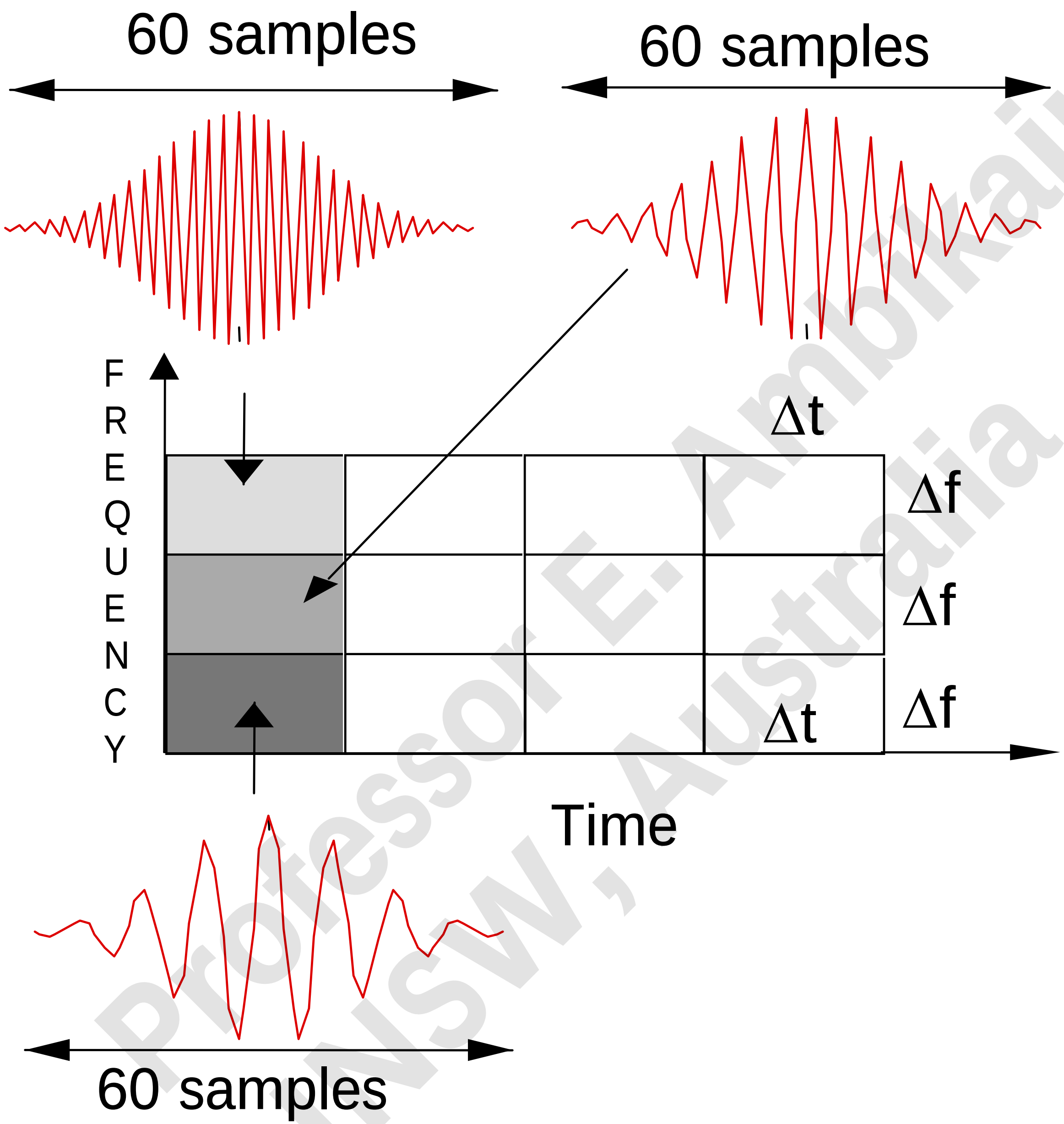
- It is the property of the STFT that frequency resolution is inversely proportional to window length $\Delta f \ \alpha$ (1/window length)
- Choosing a window function short enough to localise transient events will reduce the STFT's ability to resolve closely spaced low-frequency components
- Conversely, choosing a longer window function will give better frequency resolution, but poorer time resolution

- However, many applications, including speech processing require both good frequency resolution and good time resolution, hence a trade off must be achieved between time and frequency resolution when using the STFT

- Note that the STFT performs a constant Bandwidth analysis which implies a variable Q analysis (same bandwidth for both high and low frequencies)

Analysis Freq.	Constant BW	Q factor
200 Hz	100Hz	2
1000 Hz	100 Hz	10
3500 Hz	100 Hz	35

STFT



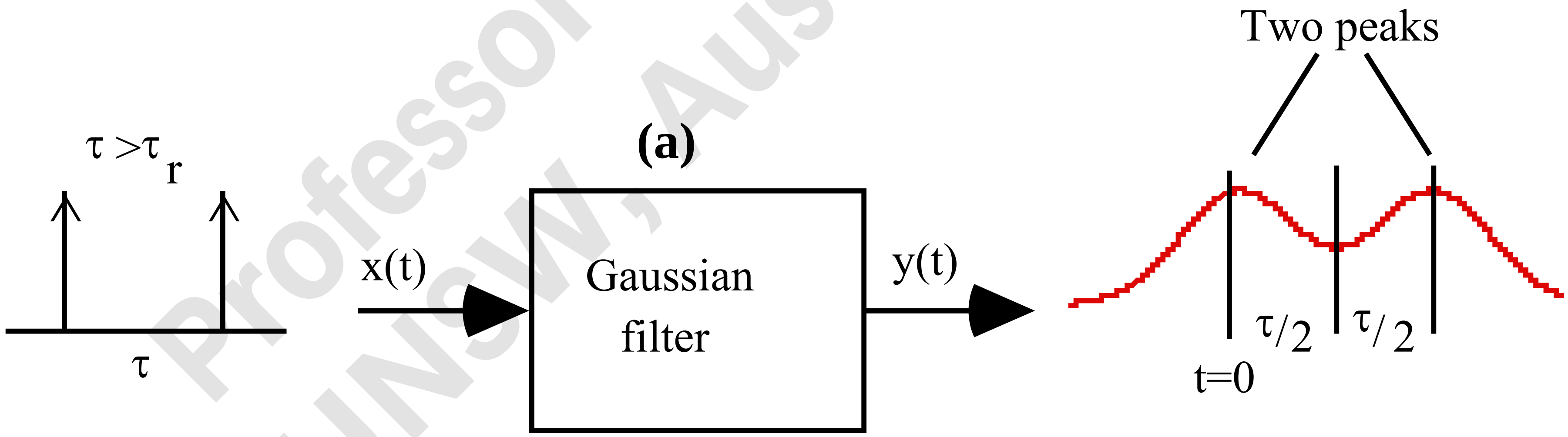
The tiling of the time-frequency plane for the STFT where the time frequency resolution is fixed for all analysis frequencies by the choice of the window function

Time resolution Analysis of a Gaussian Window/Filter

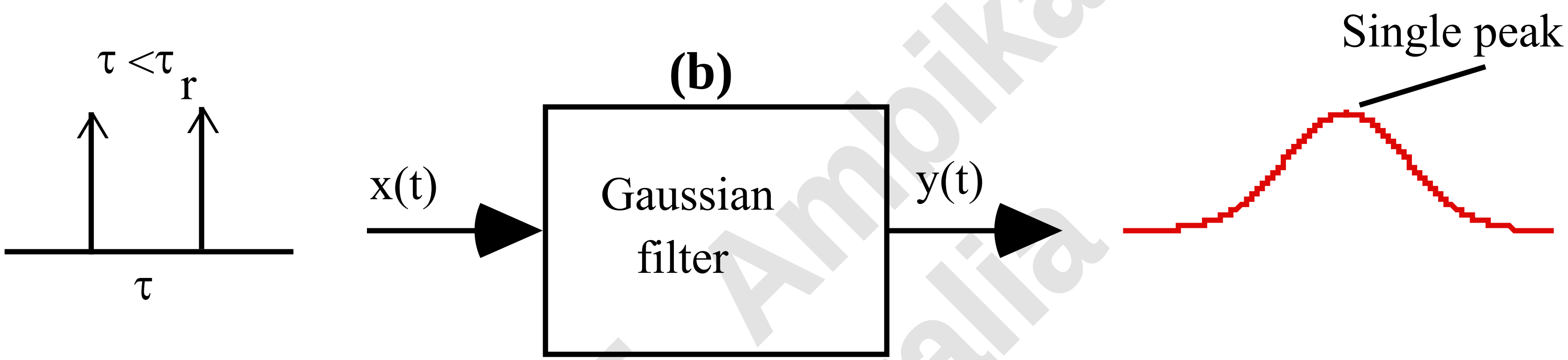
The impulse response of the Gaussian filter is given by:

$$h(t) = e^{-\alpha t^2}$$

If we define the time resolution as τ_r and the two impulses are separated by at least τ_r when passed through the Gaussian filter, the output of the filter will still lead to separate peaks in the response (see figure (a)).



Time resolution Analysis of a Gaussian Window/Filter......



However, if the pulses are not separated by τ_r then the output of the filter shows only one peak (see figure (b)).

In order to have two impulses separated by t secs, the response y(t) must be:

$$y(t) = h(t) + h(t-\tau)$$

$$y(t) = e^{-\alpha t^2} + e^{-\alpha (t-\tau)^2}$$

From figure(a) it is clear that $y'' > 0$ *when* $t = \frac{\tau}{2}$

$$\left.\frac{\partial^2 y(t)}{\partial t^2}\right|_{t=\frac{\tau}{2}} = -2\alpha e^{\frac{-\alpha\tau^2}{4}}[2-\alpha\tau^2]$$

$$= 0 \qquad \text{when } t = \frac{\tau}{2} \text{ (saddle point)}$$

$$[2-\alpha\tau^2] = 0$$

$$\text{Time resolution } \Delta t \text{ (sec)} = \tau_r = \sqrt{\frac{2}{\alpha}}$$

Frequency Resolution of the Gaussian Filter

The FT of the Gaussian filter is

$$H(\omega) = \int_{-\infty}^{\infty} e^{-\alpha t^2} e^{-j\omega t} dt$$

$$H(\omega) = \sqrt{\frac{\pi}{\alpha}}\, e^{-\frac{\omega^2}{4\alpha}}$$

In order to obtain the 3dB bandwidth,

let $\omega = \omega_B$, *Hence*

$$\frac{1}{\sqrt{2}} = e^{-\frac{\omega_B^2}{4\alpha}}$$

The Bandwidth (Δ f in radians) is :

$$\Delta f = 2\left[\sqrt{4\alpha\,(\ln \sqrt{2}\,)}\right]$$

Time-Bandwidth Product for a Gaussian filter

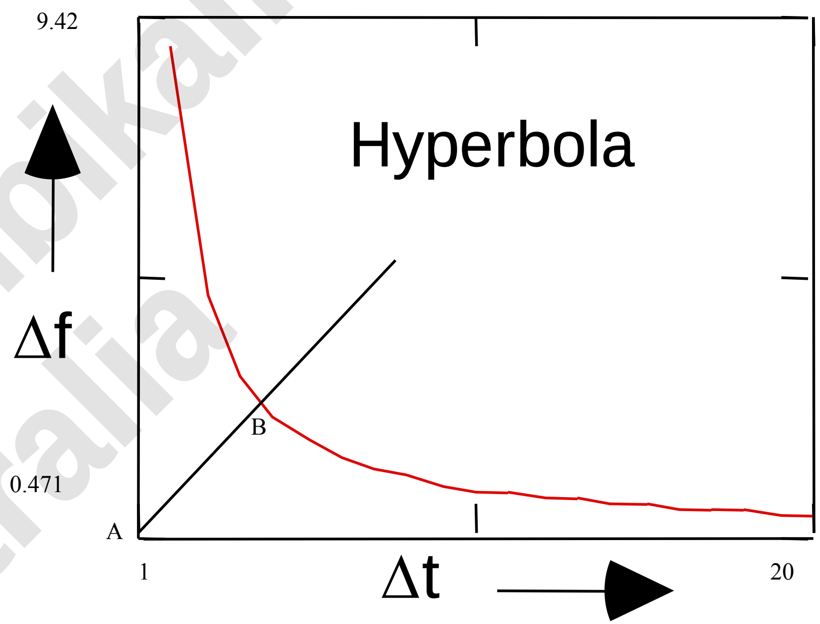
$$let \quad \alpha = 1/2$$

$$\Delta t \, \Delta f = 16 \left[\sqrt{(\ln \sqrt{2}\,)} \right]$$

The following graph shows Δf versus Δt of the above equation.

Similarly, the Time-Bandwidth product for other window functions can be calculated and plotted.

From the graph it can be proved that the distance AB is minimum for the Gaussian window function as compared to any other window functions.



Rectangular (Δt.Δf)=0.88

Triangular =0.64

Hamming =0.65

Gaussian = 0.53

Introduction to Wavelet Transform (WT)

- Wavelet Transform is a relatively new technique for signal analysis, and have recently found applications in many areas including Speech and Audio processing, Image processing, and Biomedical signal analysis.

Basis Functions

- In spectral analysis of a signal, a set of basis functions is used to identify frequency components that are present in the signal, for example, in the STFT, the basis functions are sinusoids of various frequencies.

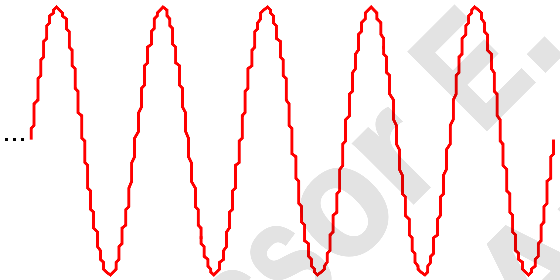
- The Wavelet Transform (WT) has been proposed as a potentially superior alternative to the STFT. In the WT, wavelets are used as the basis functions.

Wavelets

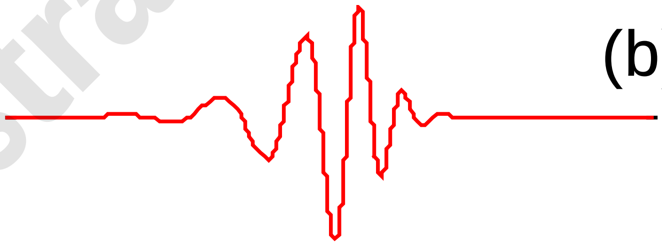
- A wavelet is an oscillatory waveform of finite duration that has an average value of zero, in contrast to sinusoids, which have infinite duration.

- Also, where sinusoids are smooth and regular, wavelets tend to be asymmetric and irregular.

(a)



Sinusoid



(b)

Wavelet

For example, the above Figure (a) shows a sinusoidal waveform, while Figure (b) displays a typical wavelet.

Short-Time Fourier Transform (STFT)

- Conventional STFT-based spectral analysis techniques build a time-frequency representation of a signal by taking the Fourier Transform of a windowed section of the signal.

- The window is then moved along the signal in time producing a succession of estimates of the spectral components of the signal.

Short-Time Fourier Transform ……….

- This works well for signals composed of stationary components and for slowly varying signals.

- However, if a signal contains both slowly-varying components and rapidly changing transient events, such as speech signals, a problem can arise in the use of the STFT.

Short-Time Fourier Transform ....

- To achieve effective temporal localisation of transient events, a short analysis window is required.

- It is a property of the STFT that frequency resolution is inversely proportional to window length, which implies that improved time resolution can be achieved only at the expense of poorer frequency resolution.

Short-Time Fourier Transform ....

- Thus, choosing a window function short enough to localise the high frequency transients will reduce the STFT's ability to resolve closely-spaced low-frequency components.
- Conversely, choosing a longer window function will give better frequency resolution, but poorer time resolution.
- However, many applications, for example speech processing, require both good frequency resolution and good time resolution, hence a trade-off must be achieved between time and frequency resolution when using the STFT.

Wavelet Transform

- In contrast to the STFT, which uses a single analysis window, the Wavelet Transform (WT) uses shorter windows at higher frequencies and longer windows at lower frequencies.

- This provides good frequency resolution but poorer time resolution at lower frequencies and good time resolution but poor frequency resolution at higher frequencies.

Wavelet Transform ……

- The loss in time resolution at lower frequencies does not present major difficulties in many applications e.g. speech processing, as lower frequency components in speech are relatively constant.

- The loss in frequency resolution at higher frequencies is also of minor consequence in speech processing, as high frequency components are composed primarily of transient events whose separation and identification require good time resolution.

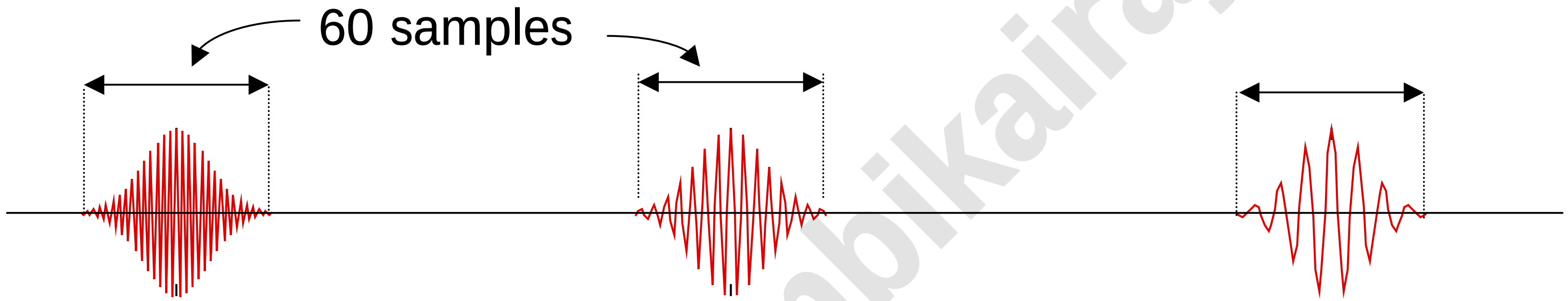
Wavelet Transform……

- The basis functions used in the WT are derived from a single primary (‘mother’) wavelet by time compression and dilation, whereas the basis functions used in the STFT are derived by varying the frequency of a sinusoid.

- A further difference between the STFT and the WT is that the basis functions in the STFT are of infinite duration, whereas the wavelets used in the WT are of finite duration (and hence finite energy); this is referred to as ‘compact support’.

Basis Function for STFT:

$$w(t)\, e^{-j\omega_0 t} = e^{-\alpha t^2}\, e^{-j\omega_0 t}$$



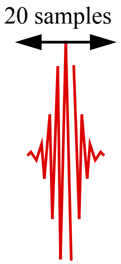
The basis function used to analyse high frequency components in the case of the STFT (window size = 60 samples, cycles in window = 28)

The basis function used to analyse medium frequency components in the case of the STFT (window size = 60 samples, cycles in window = 14)

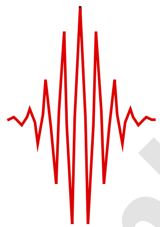
The basis function used to analysis low frequency components in the case of the STFT (window size = 60 samples, cycles in window = 7)

Basis Function for WT:

$$w\left(\frac{t}{a}\right) e^{-\frac{j\omega_0 t}{a}} = e^{-\alpha\left(\frac{t}{a}\right)^2} e^{-\frac{j\omega_0 t}{a}}$$



a=0.6



a=1.0

a=1.4

A high frequency primary wavelet. (Short window = 20 samples, cycles in window = 7)
Note: A Gaussian window is used here.

The primary wavelet above, stretched in time to produce a medium frequency wavelet. (medium window = 40 samples, cycles in window = 7)

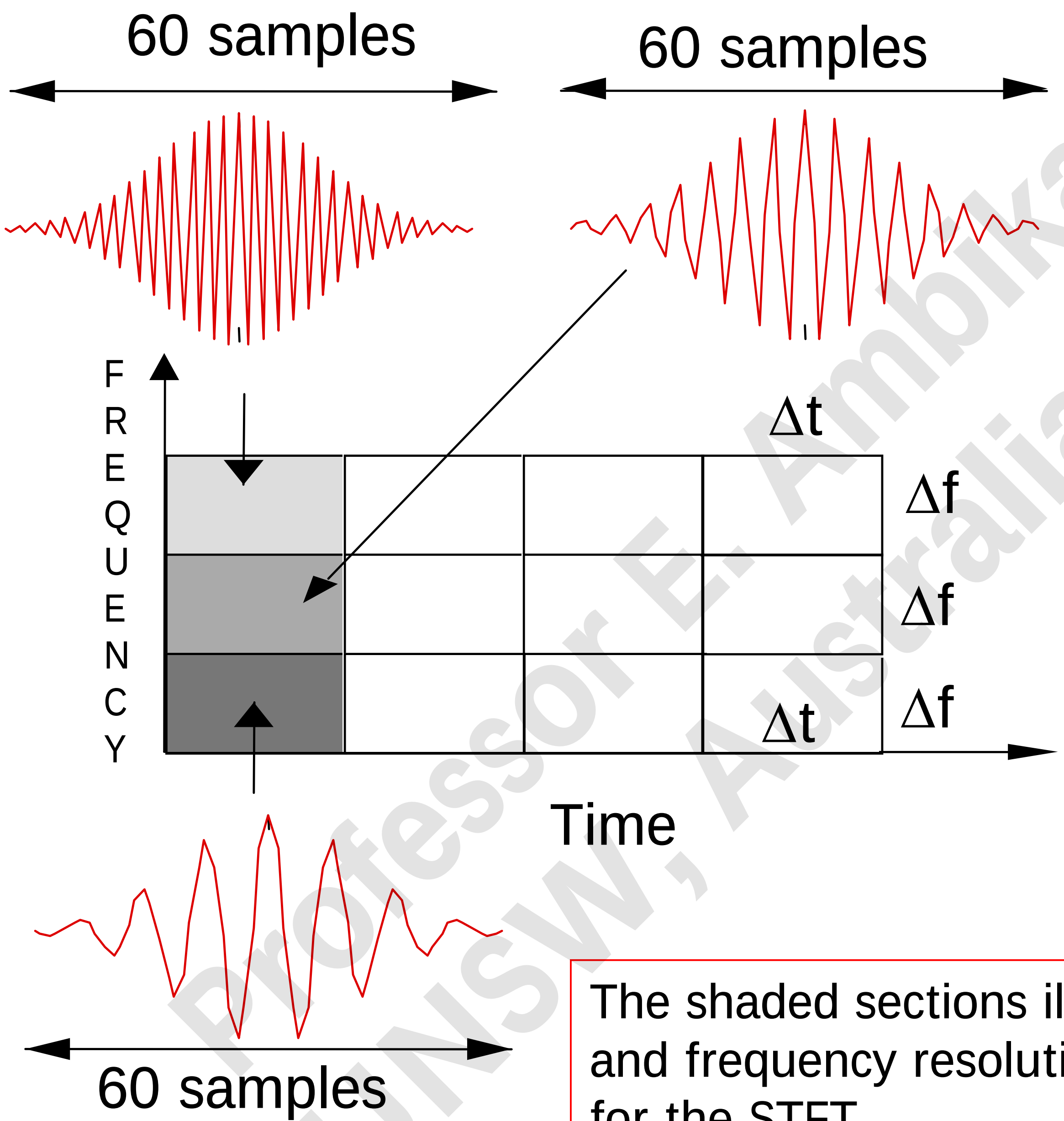
The primary wavelet above further stretched in time to produce a low frequency wavelet. (long window = 60 samples, cycles in window =7)

WT performs Constant Q analysis

- The fact that the wavelets used all have the same number of cycles means that the WT provides a constant-Q frequency analysis.
- The primary wavelet can be thought of as a bandpass filter, and the constant-Q property of the other bandpass filters (daughter wavelets) follows because they are all simply compressed or dilated versions of the primary wavelet.

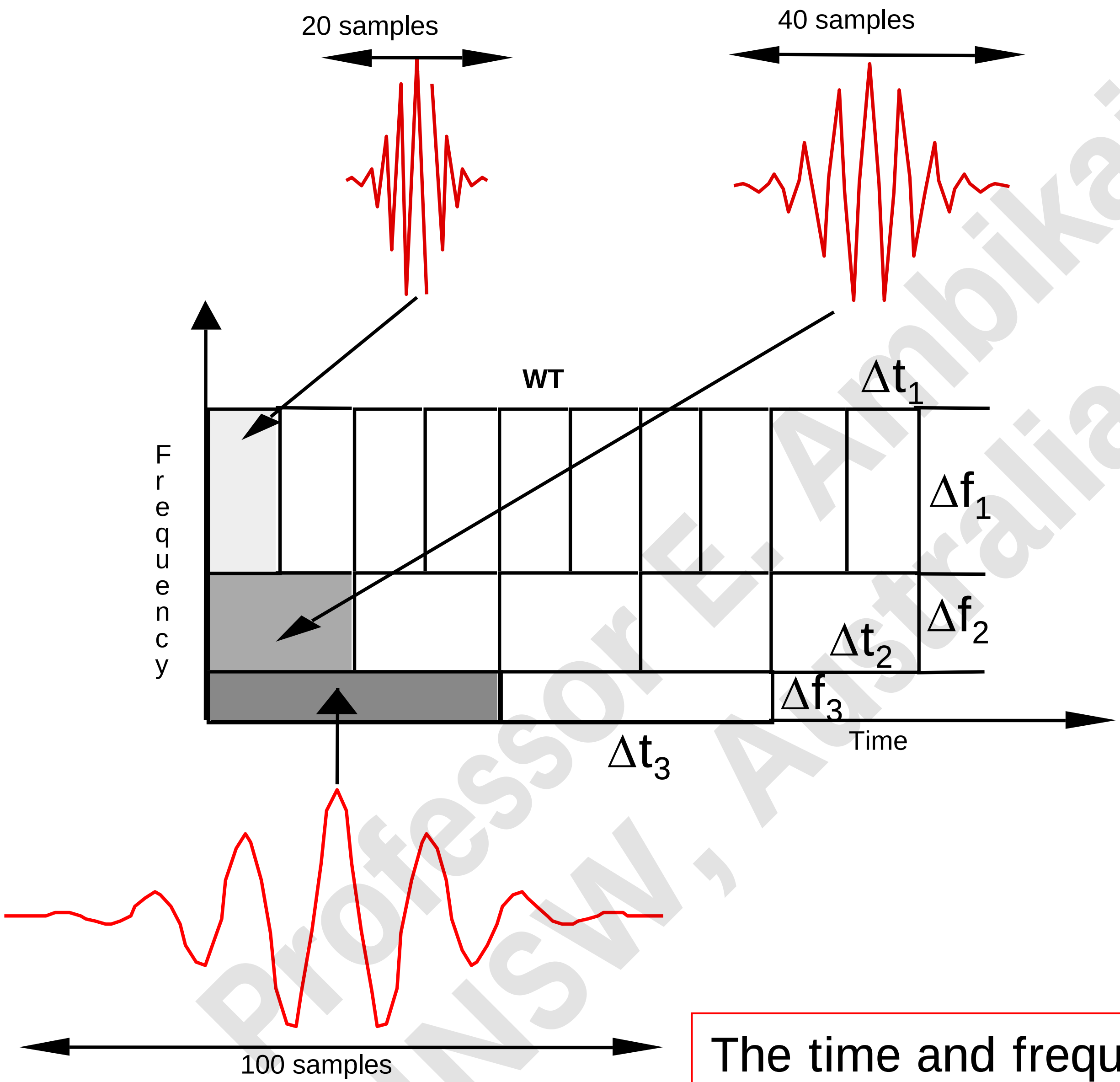
STFT Performs Constant Bandwidth Analysis

- In STFT, the width of the window is constant for all frequencies and, as a result, the basis function uses many cycles (high-Q) to analyse high frequency components as opposed to a smaller number of cycles (low-Q) for analysing low frequency components.



The tiling of the time-frequency plane for the STFT where the time frequency resolution is fixed for all analysis frequencies by the choice of the window function.

The shaded sections illustrate how the time and frequency resolution are both constant for the STFT.



Note that in both cases (STFT and WT), the *product* of time resolution and frequency resolution is constant.

Figure shows the tiling of the time-frequency plane for the WT, where high frequencies are analysed with good time resolution and low frequencies are analysed with good frequency resolution.

The time and frequency resolution are traded off against each other in the WT.

Morlet Wavelet

- Further insight into the bandpass filter-like nature of the wavelet transform can be obtained by considering the Morlet wavelet, which is given by the following equation:

$$g(t) = w(t)\cos(\omega_0 t) = e^{\frac{-t^2}{2}}\cos(\omega_0 t)$$

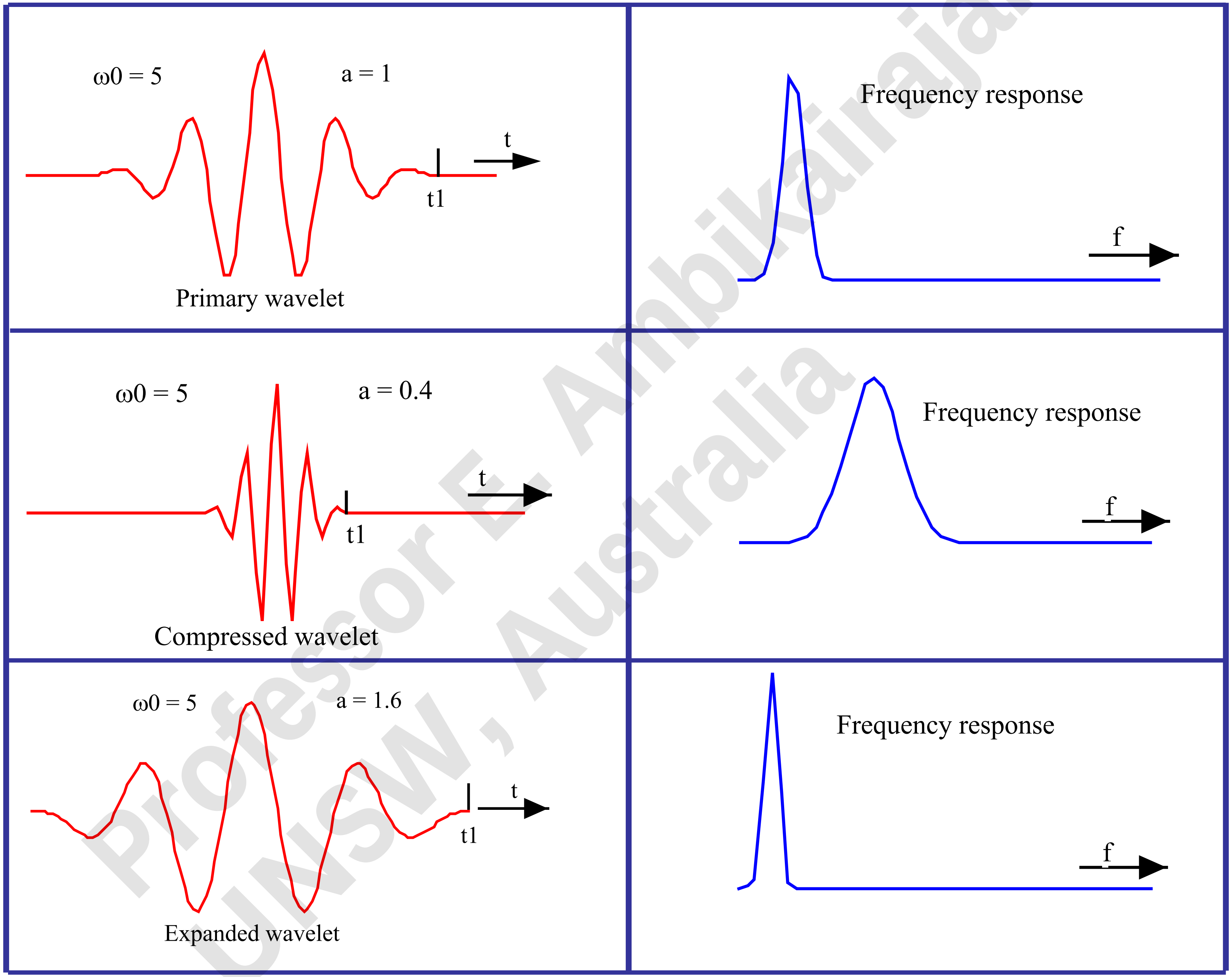
- where ω_0 is the frequency of the mother wavelet. The wavelet g(t/a) is derived by expansion if a > 1, or by compression if a < 1. The above equation can now be modified to explicitly include this dependence on the factor '*a*' (usually referred to as the 'scale factor').

Morlet wavelet

$$\mathrm{g}\left(\frac{\mathrm{t}}{\mathrm{a}}\right) = w\left(\frac{t}{a}\right)\cos\left(\omega_0 \frac{t}{a}\right) = e^{\frac{-1}{2}\left(\frac{t}{a}\right)^2} \cos\left(\omega_0 \frac{t}{a}\right)$$

The time function and corresponding frequency response for the wavelet described by the above equation is plotted for various values of ‘a’.

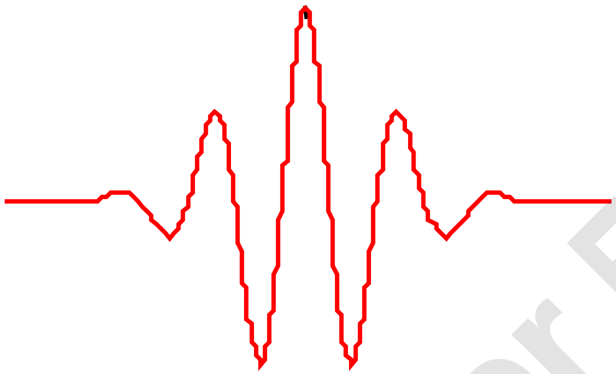
Note: The bandwidth is different in each case.


ω0 = 5
a = 1
t
t1
Primary wavelet
Frequency response
f
ω0 = 5
a = 0.4
t
t1
Compressed wavelet
Frequency response
f
ω0 = 5
a = 1.6
t
t1
Expanded wavelet
Frequency response
f

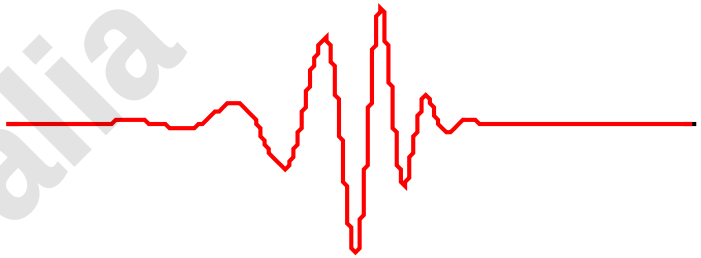
Mathematical Conditions for wavelet basis function

- One important point to remember is that the wavelet transform does not have a single set of basis functions (i.e. a single mother wavelet) like the Fourier transform, which utilises just sine and cosine functions.
- Instead, the wavelet transform can choose from an infinite set of possible mother wavelets.
- A number of mathematical conditions must be satisfied for a function to be admissible as a wavelet basis function.
- These conditions essentially require the wavelet function to be absolutely integrable and square integrable, to be composed of only positive frequency components and to have zero DC component.

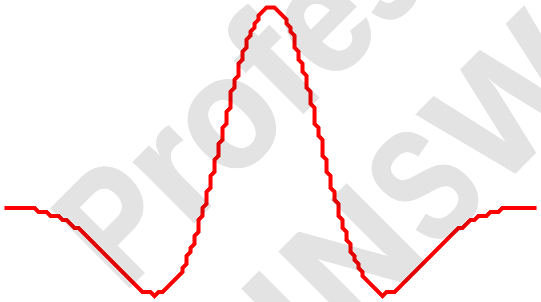
There are different types of wavelet families whose characteristics vary according to several criteria. All wavelets decay quickly to zero, i.e. they have finite energy and compact support (zero valued outside a certain time interval). Examples of one-dimensional wavelets are given below:



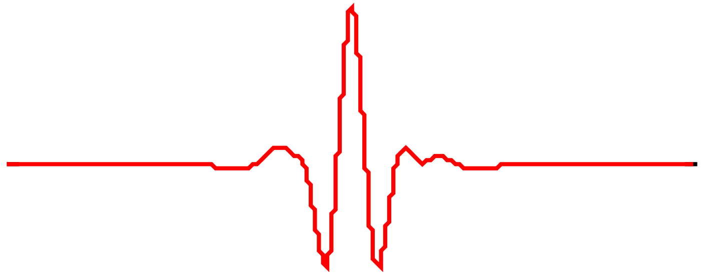
Morlet/Gaussian



Daubechies (db8)



Mexican hat



Symlet (sym8)

Discrete STFT vs Discrete WT

The discrete-time STFT is evaluated according to the following equation:

$$STFT(k,b) = \sum_{m=-N/2}^{m=N/2} w(m-b)\, s(m)\, e^{-j\frac{2\pi k}{N}m}$$

where s(m) is the signal being analysed, T is the sampling period, N is the window width, which is fixed for the STFT, and k=0,1,....N-1. The parameter b specifies the number of samples by which the analysis window is shifted along the signal under analysis, while w(m-b) is the window function, shifted by b samples.

Discrete WT

$$g\left(\frac{t}{a}\right) = e^{\frac{-1}{2}\left(\frac{t}{a}\right)^2} e^{-\frac{j\omega_0 t}{a}}$$

The discrete WT (assuming a Morlet wavelet) is calculated as follows:

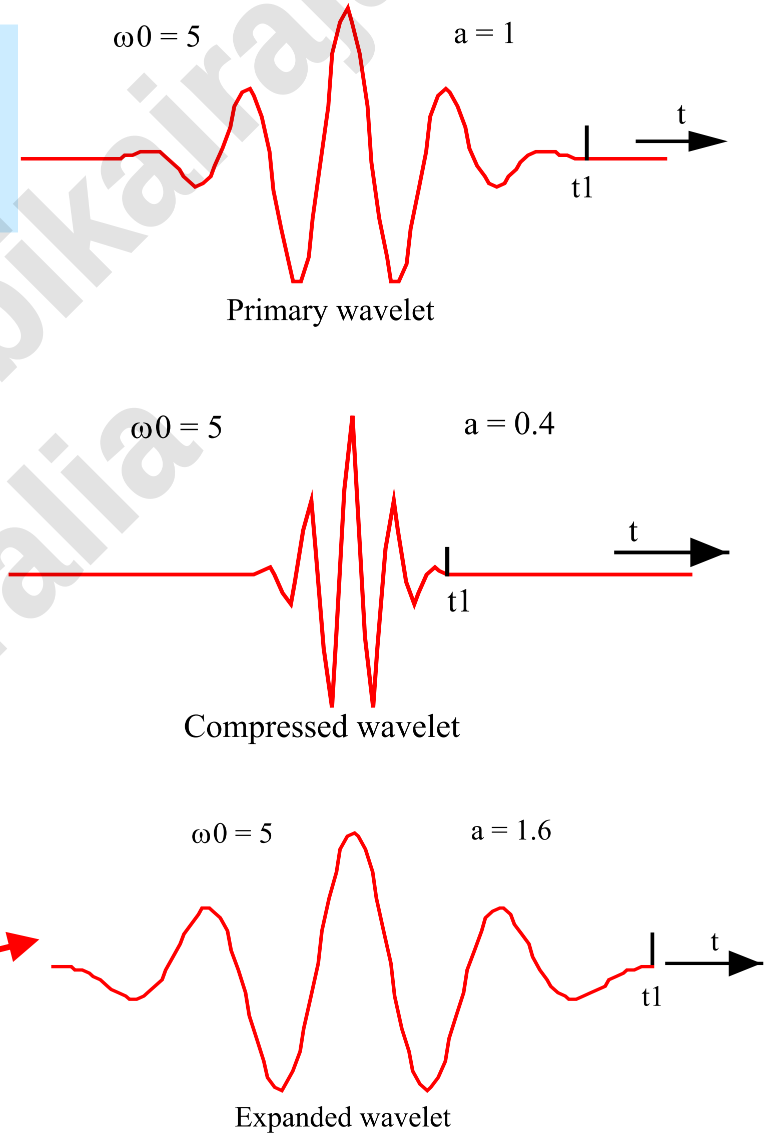
$$DWT(a,b) = \frac{1}{\sqrt{a}} \sum_{m=-N(a)}^{m=N(a)} e^{-\frac{1}{2}\left(\frac{[m-b]T}{a}\right)^2} s(m) e^{\left(\frac{-jw_0 mT}{a}\right)}$$

where *a* is the positive valued scale parameter, and ωo is the frequency of the prototype wavelet. The constant 1/sqrt(a) is used for energy renormalisation. N(a) is a positive constant associated with the wavelet length such that wavelet is non-zero in the range -N(a) to +N(a).

$$DWT(a,b) = \frac{1}{\sqrt{a}} \sum_{m=-N(a)}^{m=N(a)} e^{-\frac{1}{2}\left(\frac{[m-b]T}{a}\right)^2} s(m) e^{\left(\frac{-jw_0 mT}{a}\right)}$$

It is clear that expanding a prototype wavelet with frequency f_0 will result in a longer wavelet with a lower associated frequency. In fact, the associated wavelet frequency is inversely proportional to the scale parameter 'a' as indicated by

$$f = \frac{f_0}{a}; \quad \omega = \frac{\omega_0}{a}$$



$$DWT(a,b) = \frac{1}{\sqrt{a}} \sum_{m=-N(a)}^{m=N(a)} e^{-\frac{1}{2}\left(\frac{[m-b]T}{a}\right)^2} s(m) e^{\left(\frac{-jw_o mT}{a}\right)}$$

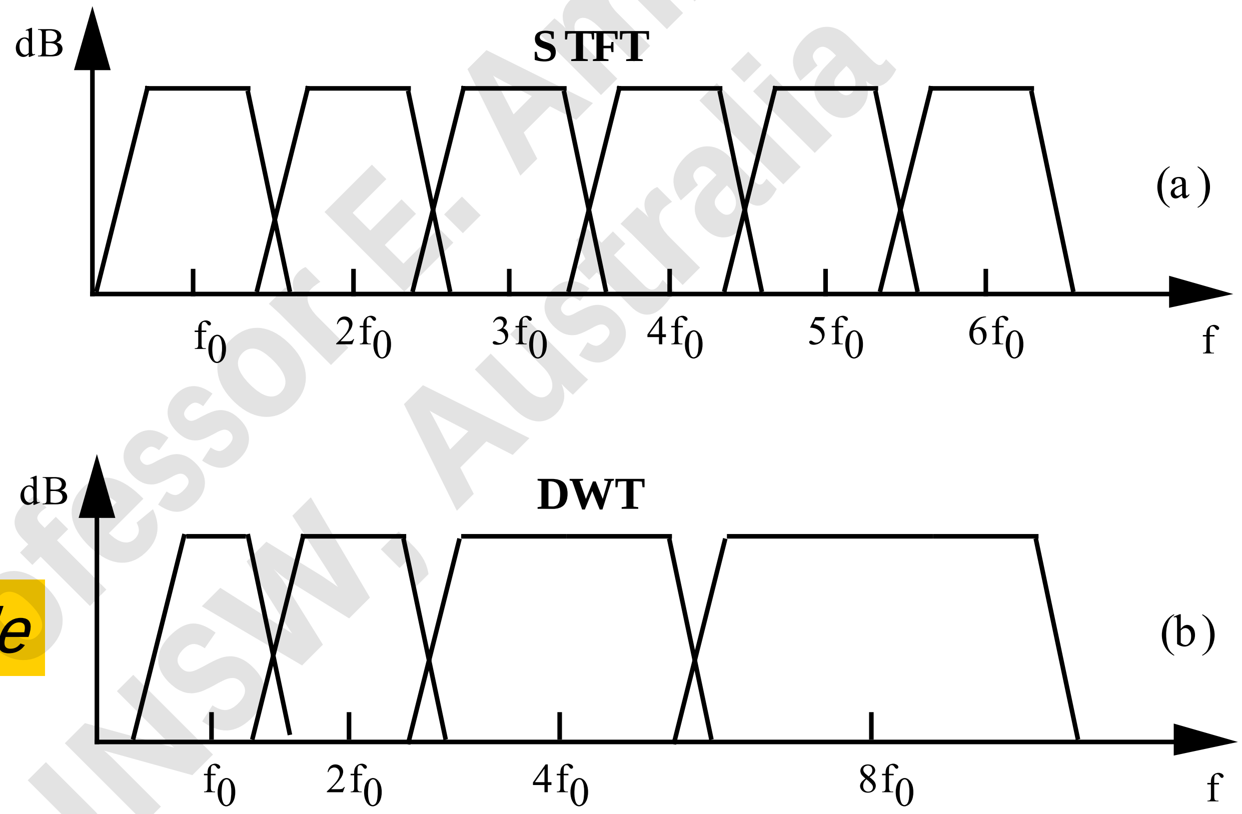
In order to analyse the signal, the scale parameter a (see above equation) is often given an initial large value (corresponding to the lowest-frequency prototype wavelet) and is then decreased in regular increments to analyse the signal in more detail at higher frequencies.

Since spectral properties are frequently displayed on a logarithmic frequency scale, it is customary to use:

$$a = \frac{1}{2^j}$$

With this definition, integral increments in j result in octave increments in a (j = 0, 1, 2, 3, 4…). This is called the *dyadic scale* (a is modified by powers of two).

Figure below shows that the frequency analysis performed by the STFT results in a spectrum with increments equally spaced along the frequency axis, and with equal bandwidth, while the DWT performs frequency analysis with variable bandwidth (constant-Q) and logarithmically spaced increments along the frequency axis.



dyadic scale

Example

- Figure shows (next slide) the frequency analysis carried out on a speech signal using STFT and DWT.
- The speech signal is the word ‘eight’ sampled at 8 kHz.
- The top window shows the time-frequency analysis carried out using the DWT while the middle window shows the STFT spectrogram obtained.
- Note that both analyses show good frequency localisation in the low frequency formant region.
- However, at high frequencies the time localisation is much better for the DWT than for the STFT.
- The window size used for the analysis of the STFT was 300 samples. In this example, a Morlet window was used, and the frequency of the mother wavelet was 3.9 Hz.

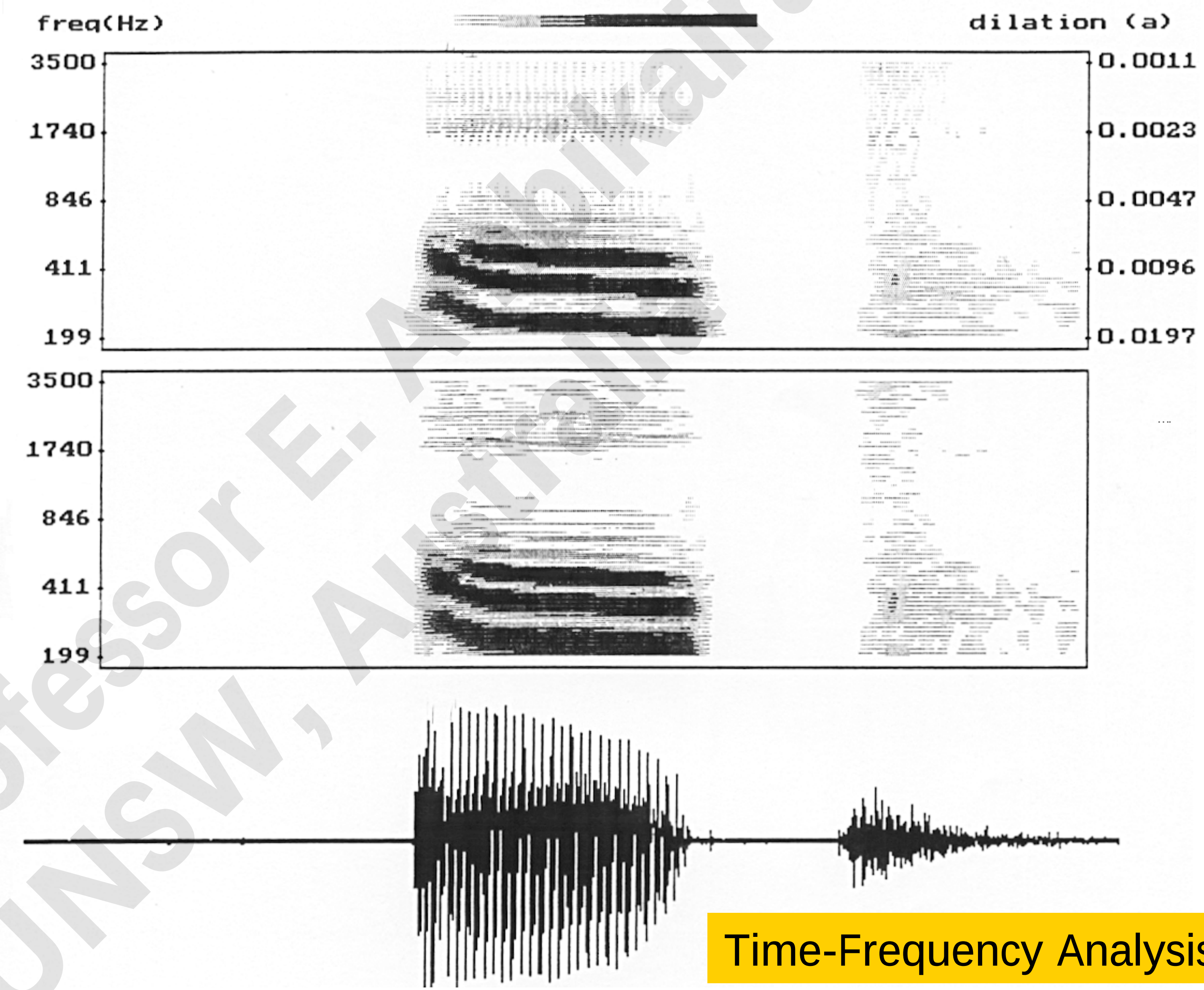
Comparison of DWT (top window) and STFT (middle window) analysis of the utterance ‘eight’ (bottom window).

WT

STFT

utterance

‘eight’



Time-Frequency Analysis

Alternative Implementation of the Discrete Wavelet Transform

- The DWT given by the following Equation

$$\mathrm{DWT}(a,b)=\frac{1}{\sqrt{a}}\sum_{m=-N(a)}^{m=N(a)} e^{-\frac{1}{2}\left(\frac{[m-b]T}{a}\right)^2} s(m) e^{\left(\frac{-jw_o mT}{a}\right)}$$

  is not the most practical way of evaluating the wavelet transform, as strictly speaking, an infinite number of wavelet functions is required to represent a signal.
- However, it is possible to obtain an equivalent formulation to the DWT which uses a finite number of basis functions.
- This form of the DWT uses a *pair* of wavelet functions for calculation.

- In this formulation, a signal s(n) can be decomposed in terms of translates and dilates of a primary wavelet w(n):

$$s(n) = \sum_{k} c(k) g_k(n) + \sum_{j=1}^{L} \sum_{k} d(j,k) w_{jk}(n)$$

- where n= 0, 1, 2, ...; j (scale factor) = 1, 2, 3, ...; k (translation factor) = 0, 1, 2, 3, ...; g(n) is a so-called scaling function, and the primary wavelet w(n) is normally obtained from the scaling function.

- The primary wavelet is normally a bandpass filter with a centre frequency of ω_0 and hence

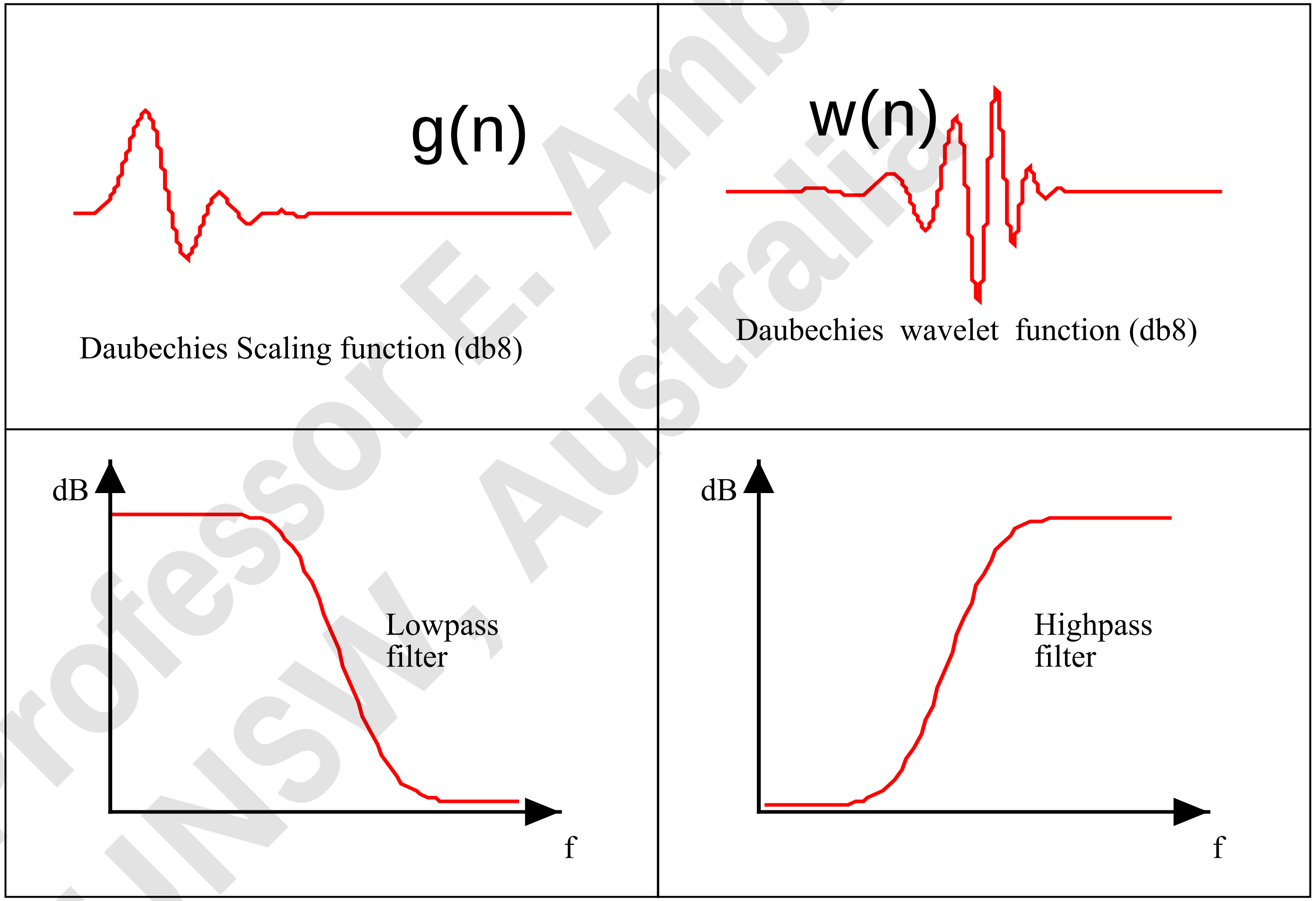
$$\sum_{n=0}^{N-1} w(n) = 0$$

- Similarly, g(n) is a normalised lowpass filter and hence

$$\sum_{n=0}^{N-1} g(n) = 1$$

- The coefficients c(k) (see previous slide) represent the approximation of the original signal s(n), while the coefficients d(j,k) represent the details of the original signal s(n) at different scale factors.

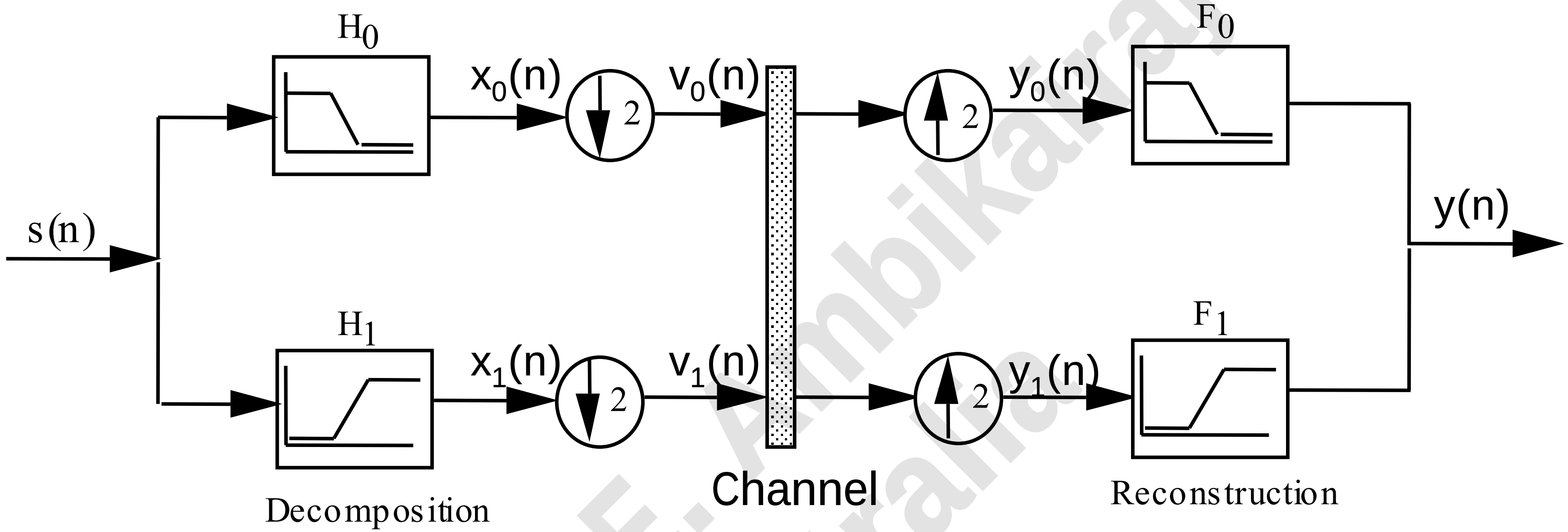
- Figure (below) gives an example of the Daubechies scaling function g(n) and the corresponding wavelet function w(n). The frequency response of each function is also shown.



Wavelets and Filter Banks

- Two methods developed independently in the late seventies and early eighties which lead naturally to fast implementation methods for the discrete wavelet transform are subband coding and multiresolution signal analysis.
- An important development was that of Quadrature Mirror Filters (QMF) which allows a signal to be split using non-ideal filters into two down-sampled subband signals, and subsequently reconstructed without aliasing distortion
- The down sampling of the signal components during decomposition introduces aliasing distortion.
- However, it turns out that by carefully choosing filters for the decomposition and reconstruction, it is possible to "cancel out" the effects of aliasing.

Two-Channel Quadrature-Mirror Filter Bank



For perfect reconstruction, the decomposition and reconstruction filters are related by the following equations:

$H_1(z) = H_0(-z)$ {demonstrates the mirror image property of the filters}

$F_0(z) = H_1(-z)$ { satisfies alias cancellation condition}

$F_1(z) = -H_0(-z)$ { satisfies alias cancellation condition}

Mirror Image Filters

- Let $h_0(n)$ be some FIR lowpass filter with real coefficients. The mirror filter is defined as

$$h_1(n)=(-1)^n h_0(n)$$

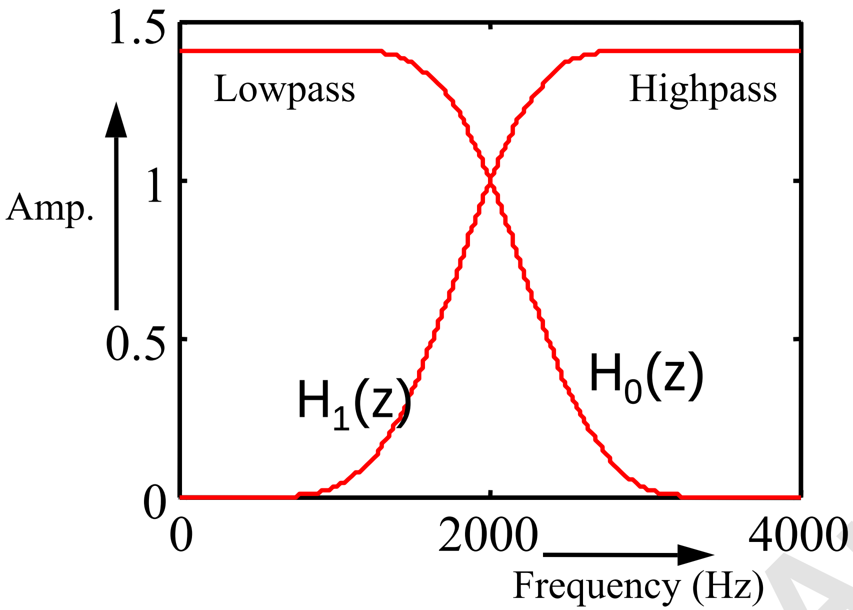
Therefore $\quad H_1(z) = H_0(-z)$

$$H_1(\theta) = H_1(z)\Big|_{z=e^{j\theta}}$$

$$H_0(-z)\Big|_{z=e^{j\theta}} = H_0(-e^{j\theta}) = H_0(e^{j(\theta-\pi)}) = H_0(\theta-\pi)$$

$$H_1(\theta) = H_0(\theta-\pi) \Rightarrow |H_1(\theta)| = |H_0(\theta-\pi)|$$

This demostrates the mirror image property of H_0 and H_1 about $\theta = \pi/2$. Hence justifying the name quadrature mirror filters (QMF)



QMF Filters

•Perfect reconstruction filterbanks then followed, with orthogonal and biorthogonal solutions.

• The orthogonal filters provide an orthogonal transform, i.e. the same signal s(n) is projected onto new perpendicular axis. The orthogonal transform can be used to separate out the noise and decorrelate the signal.

•The biorthogonal transform provides new axis which are not necessarily perpendicular, but no information is lost.

Example

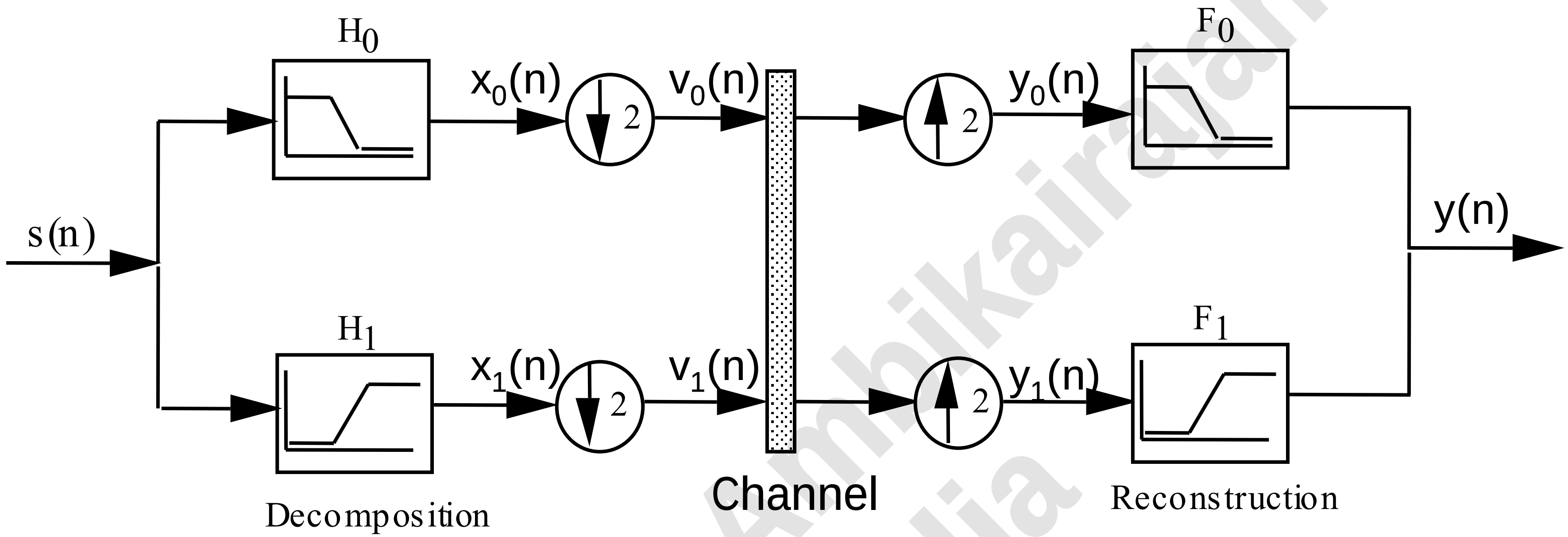
- Consider a two-channel QMF bank with the analysis filter given by $H_0(z) = 1+z^{-1}$
- The second analysis filter is therefore

$$H_1(z) = H_0(-z) = 1-z^{-1}$$

and the corresponding synthesis filters for an alias-free realisation

$$F_0(z) = H_0(z) = 1+z^{-1} \quad (= H_1(-z))$$

$$F_1(z) = -H_0(-z) = -[1-z^{-1}] = 1+ z^{-1}$$

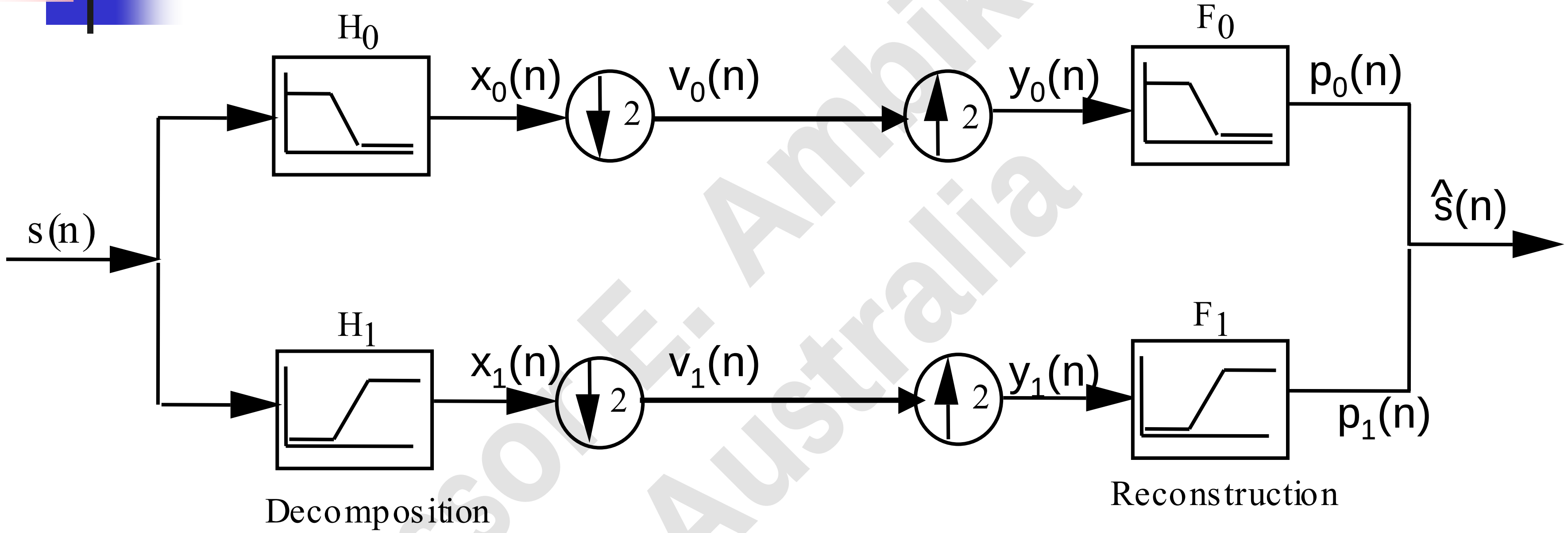


If the down-sampling and the up sampling factors are equal to the number of bands of the filter bank, then the output y(n) can be made to retain some or all of the characteristics of the input x(n) by properly choosing the filters in the structure.

In this case, the filter bank is said to be critically sampled filter bank

In speech applications, the speech signal s(n) is first split into a number of subband signals by means of analysis filterbank. The subband signals are then processed and finally combined by a synthesis filterbank resulting in an output signl y(n).

Analysis of the two-channel QMF Bank



In subband coding applications, $v_0(n)$ and $v_1(n)$ are quantised, encoded and transmitted to the receiver.

We assume ideal operation here, with no coding and transmission errors, so we focus on the analysis and synthesis filters

The Fast Wavelet Transform

- The concept of multiresolution analysis of a signal was introduced in the 1980s in the form of a pyramidal decomposition, which decomposes a signal into a coarse approximation and additional detail components. **This idea is central to wavelet analysis.**

- Mallat (1989) established the link between filter banks and wavelets and proposed an algorithm for implementing wavelet expansions as a set of filter banks.

- Daubechies (1988) first reconstructed wavelets using filter coefficients, and indeed most useful wavelets are derived iteratively from particular filters rather than the other way around.

The Fast Wavelet Transform

- Mallat's algorithm for performing wavelet decomposition and reconstruction using a tree structured filter bank is illustrated in the next slide.

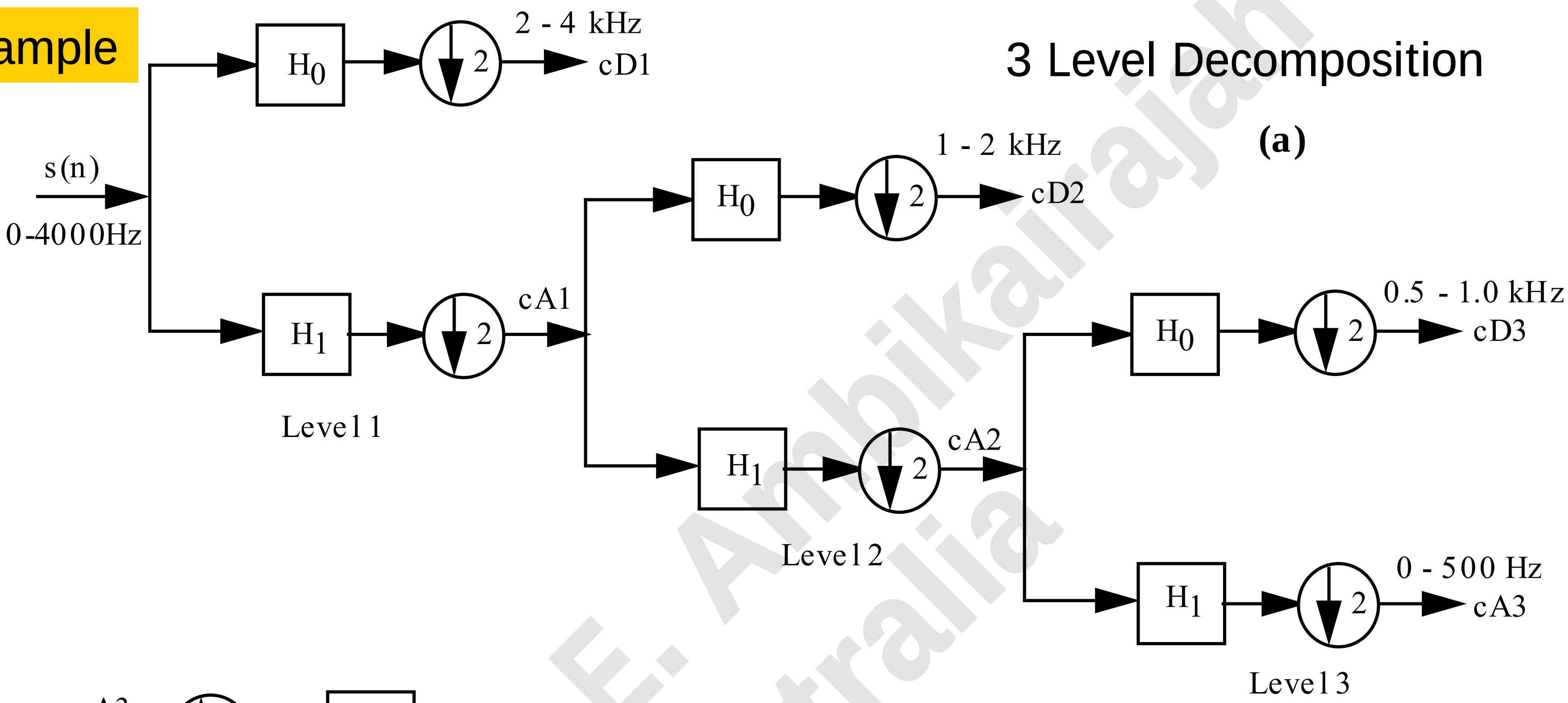
- . This algorithm is usually referred to as the Fast Wavelet Transform (FWT).

- The original signal s(n) is filtered by a lowpass and highpass filter pair followed by decimation by 2 to form the approximation (referred to as cA1) and detail (cD1) signals at Level 1.

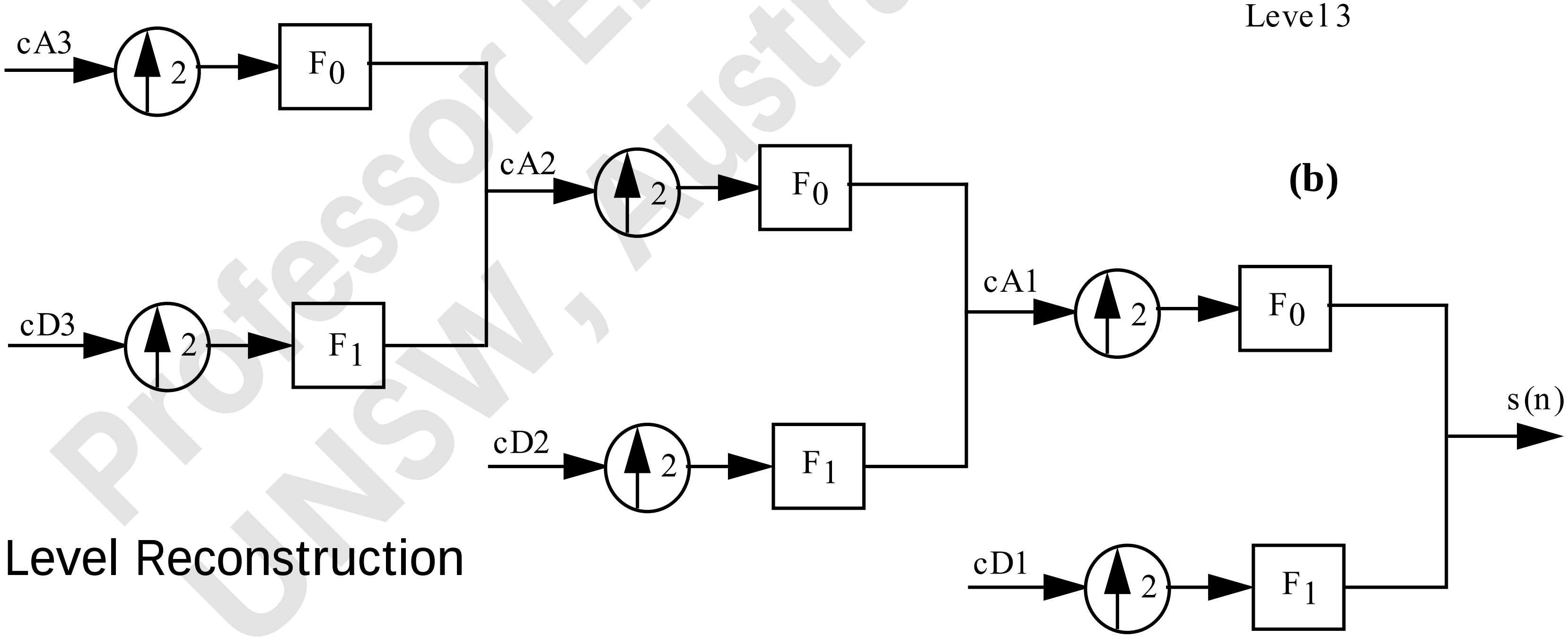
- The decomposition process can be iterated to further levels as illustratedbefore, with successive approximation components being further decomposed so that one signal is broken down into many lower resolution components.

Example

3 Level Decomposition



(a)



(b)

3 Level Reconstruction

- Theoretically this could continue indefinitely but in practice the procedure is limited by the number of samples in the original signal s(n).

- Where this length is a power of 2, for example 2^j, then j levels of decomposition are possible, resulting in an approximation cAj and a detail cDj of length 1 sample.

- The approximations are the low frequency components of the signal, the details are the high frequency components (see previous slide). The reconstruction process consists of interpolation and filtering as shown in the diagram

- . For the example shown, which consists of 3 levels of decomposition, the original signal may be reconstructed from signals cA3, cD3, cD2 and cD1. These signals are called **wavelet coefficients**.

- In practice, using the Fast Wavelet Transform for signal processing applications involves three aspects:
  - breaking up the signal to obtain the wavelet coefficients (FWT),
  - modifying the wavelet coefficients to carry out whatever processing is required (e.g. noise reduction), and
  - reassembling the signal from the modified wavelet coefficients ('Inverse FWT').

Note: It is possible to develop a multiband analysis/synthesis filterbank by iterating a two-channel QMF bank.

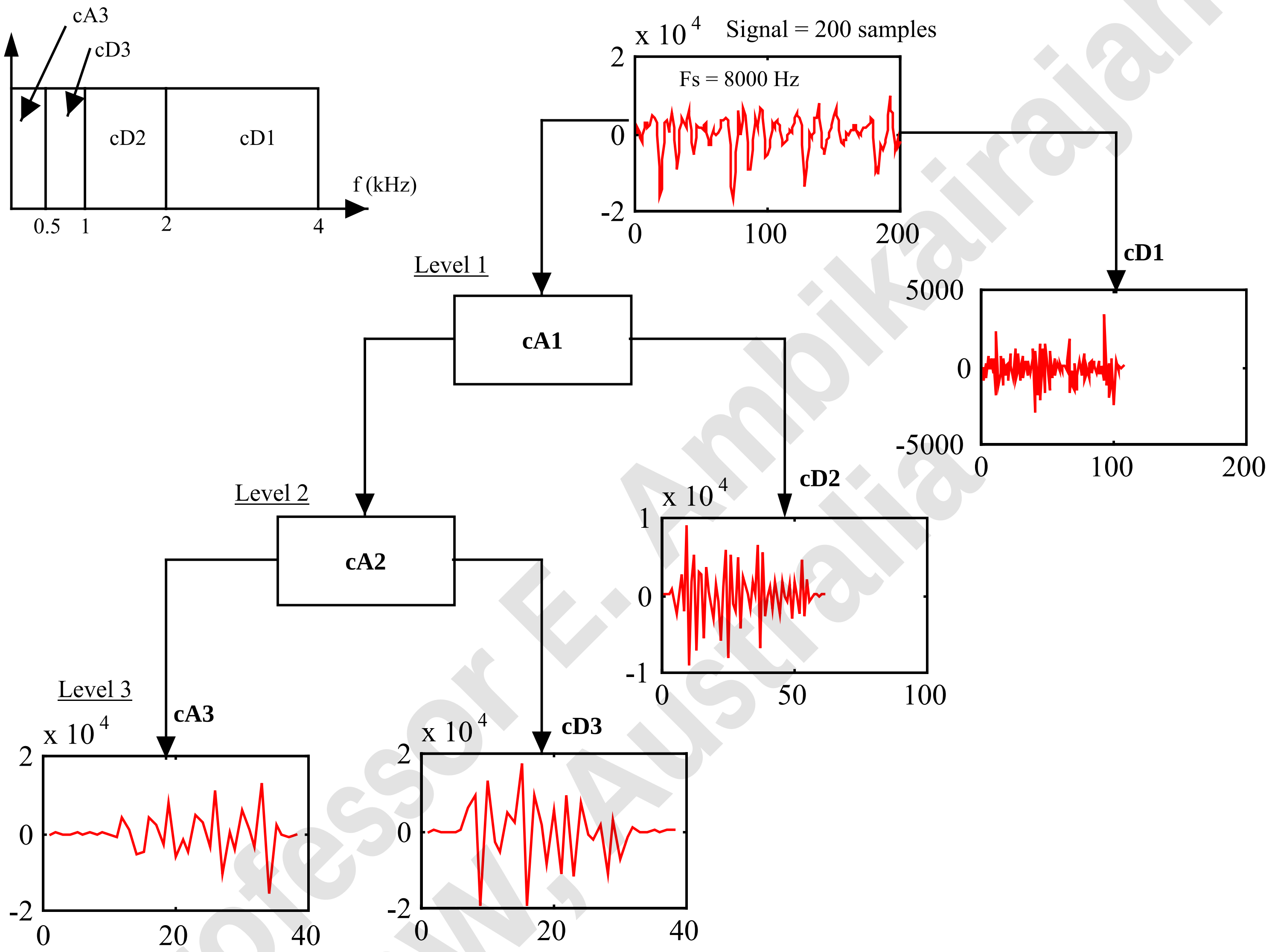
Example of a Multi-Level Decomposition

- The diagram in the next slide shows a 3-level decomposition using Daubechies wavelet filters (db8) and implemented using Matlab.

- In this example, the input signal consists of a frame of 200 speech samples.

- After the first level of decomposition (filtering and down sampling) the detail coefficients cD1 consist mainly of high frequency components, and the length of cD1 is actually more than half the length of the original signal, i.e. cD1 is 107 samples long as opposed to 100.

- This is a consequence of the filtering process, which is implemented by convolving the signal with a filter impulse response. The convolution smears the signal, introducing several extra samples into the output vector. Similarly cD2 = 61 as opposed to 50 and cD3 is 38 as opposed to 25.



3-Level Decomposition. The speech signal s(n) = 200 samples long. The wavelet coefficients cD1 = 107 samples; cD2 = 61; cD3 =38 and cA3 = 38 samples long.

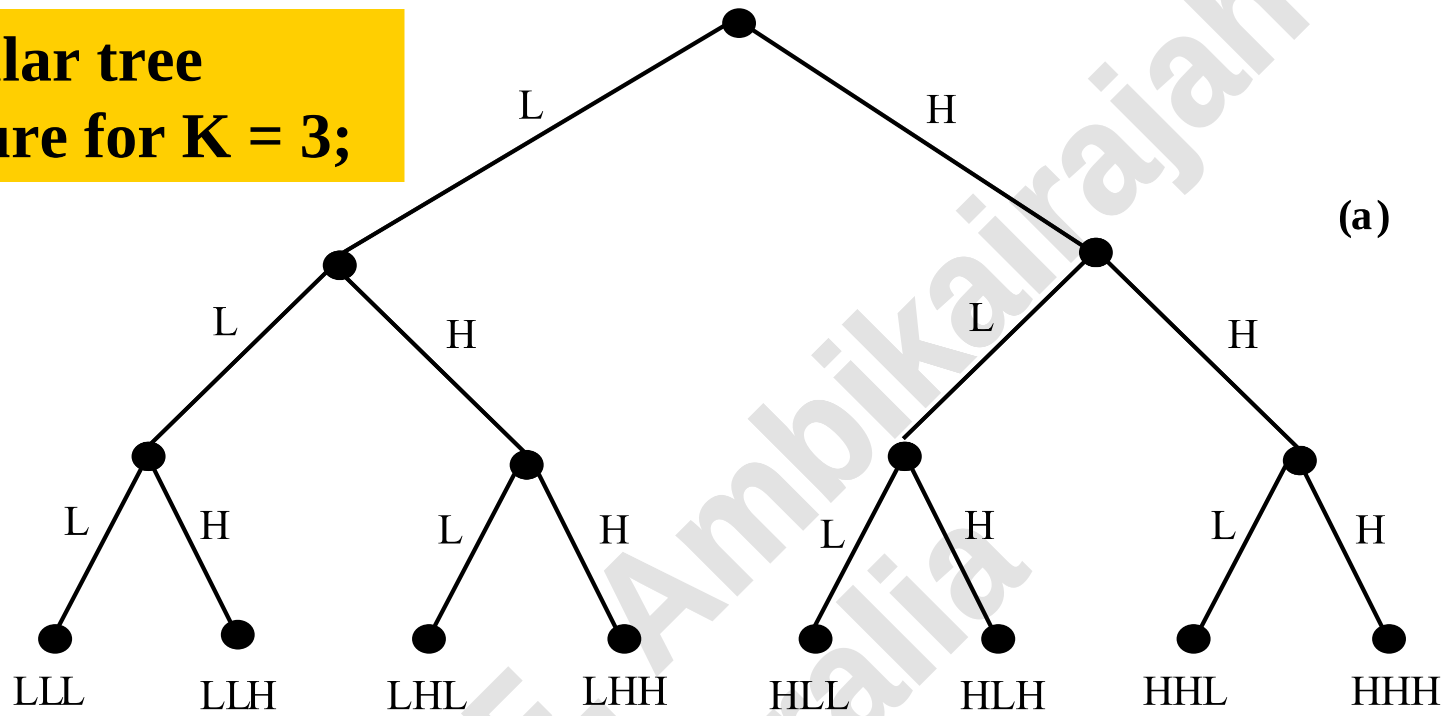
Regular Binary Subband tree

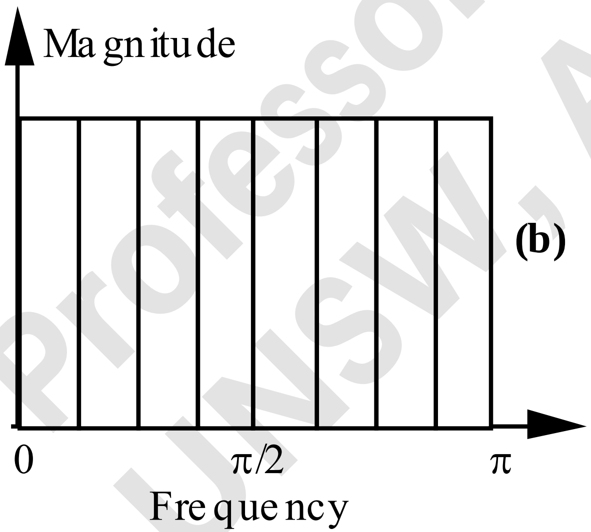
- The two-channel filterbank divides the input spectrum into two equal subbands, yielding the low (L) and the high (H) subbands.
- This two-band QMF split can again be applied to the L and H half-bands to generate the quarter bands.
- When this procedure is repeated K times, 2^K equal bandwidth subbands are obtained.
- This approach provided the maximum possible frequency resolution of $\pi/2^K$ within K levels.
- This spectral analysis structure is called a K-level regular binary tree.

Regular Binary Subband tree….

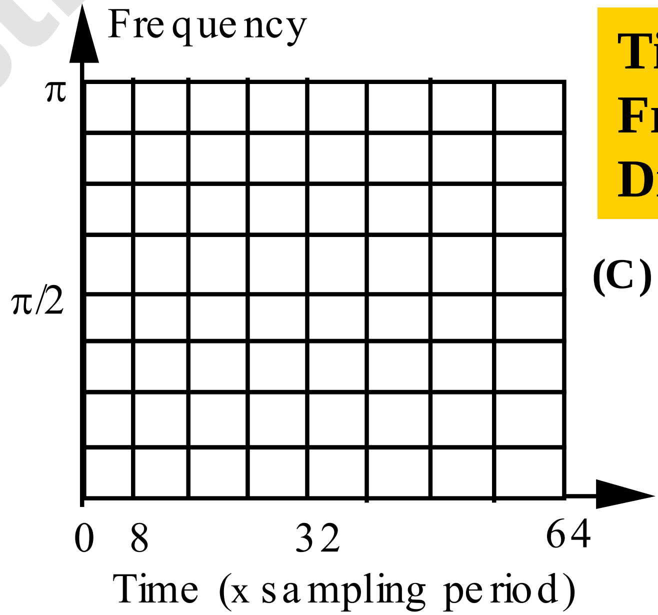
- For K = 3 the regular binary tree structure that employs QMF and the corresponding frequency band split are shown in Figure (a) and (b) (next slide)
- Figure (c) shows the frequency diagram for the 8-band regular binary subband decomposition.
- We assume that the filter outputs are computed for each input block of 8 samples, which will result in one new subband sample generated for each maximally decimated subband and for each input block.

A regular tree structure for K = 3;

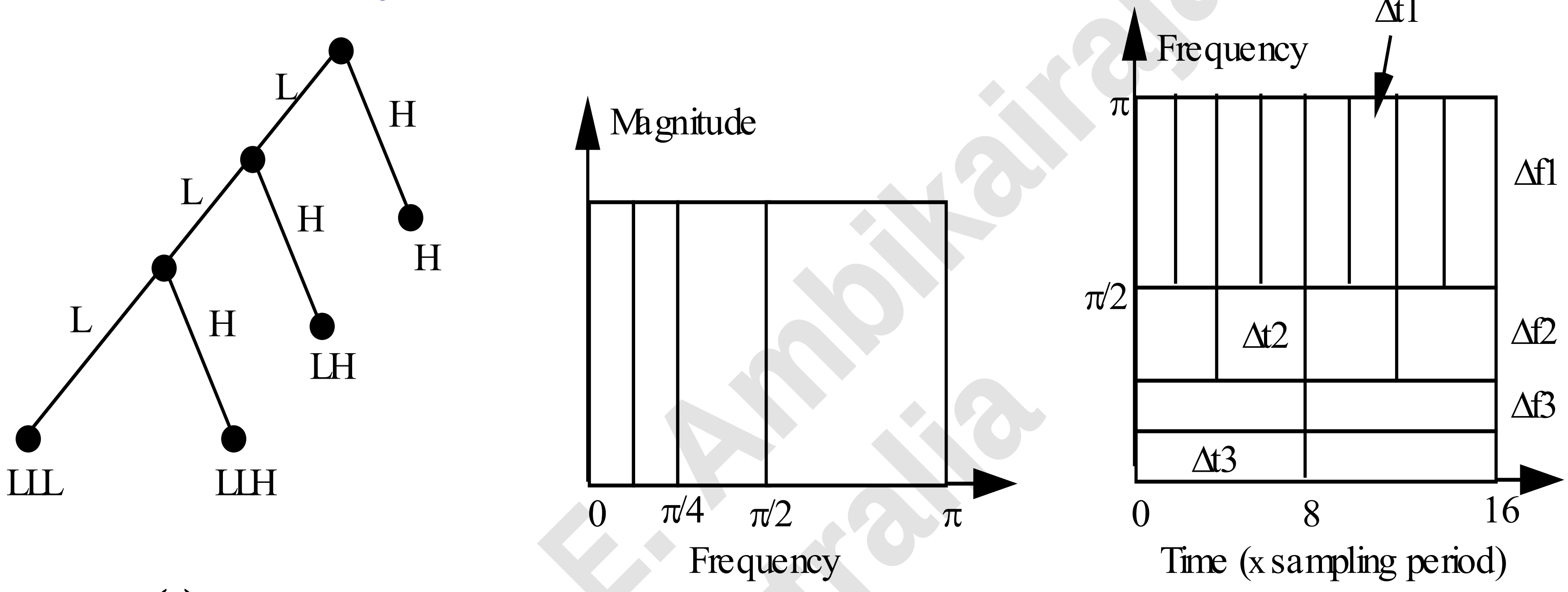




Time-Frequency Diagram



Dyadic Subband Tree



(a) (b) (c)

In many applications all of the subbands of the regular tree may not be required and as a result some of the fine frequency resolution subbands can be combined to yield larger bandwidth frequency bands. This implies the irregular termination of the tree branches.

One of the irregular tree structures is called **dyadic**, or octave band tree.

It splits only the lower half of the spectrum into two equal bands at any level of the tree. Therefore, the higher half-band component of the signal at any level of the tree is not decomposed any further.

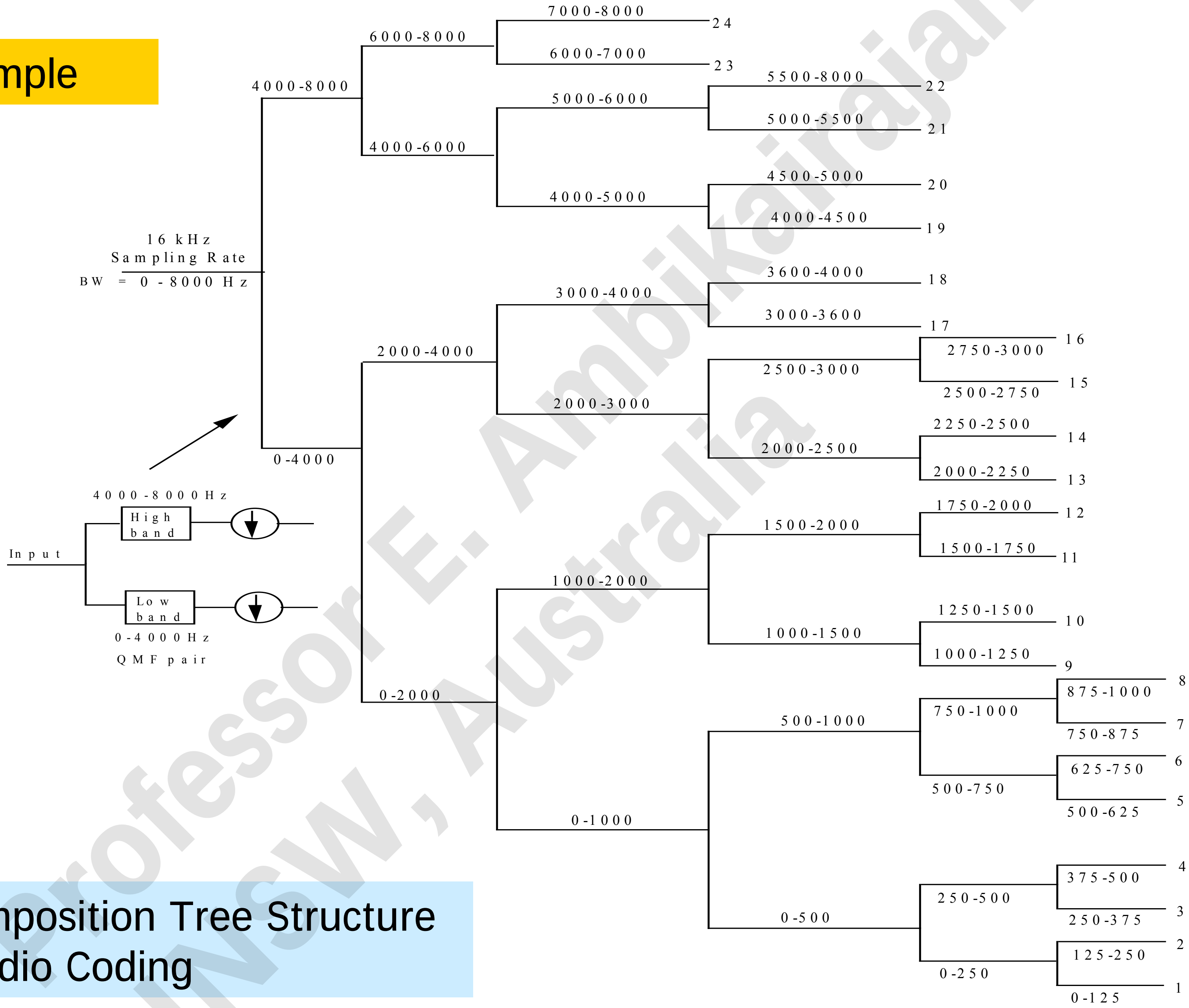
Wavelet Packet Decomposition

- The dyadic tree structured decomposition shown before can be generalised in several ways.
- First, instead of splitting a signal into two bands at a time, we can split into several bands.
- Second, signals which in the dyadic tree decomposition are not split any further can themselves be further decomposed into subbands. In this way we can obtain a very general tree structure, and by modifying the synthesis filterbank correspondingly, we can retain the perfect reconstruction property. This is called the **wavelet packet expansion.** .

Wavelet Packet Decomposition

- An example of a 24-band WP representation is shown in the next slide where the sampling rate is 16 kHz.
- This filterbank structure is identified because it has sufficient resolution for direct implementation of the psychoacoustic model.
- Also the subband bandwidths and centre frequencies closely approximate the critical bands

Example



Decomposition Tree Structure for Audio Coding

Spectral Reordering

- The subband numbering of the diagram (see previous slide) does not take into account the switching of the highpass and lowpass spectra as the output of each highpass branch in the decomposition tree is decimated.

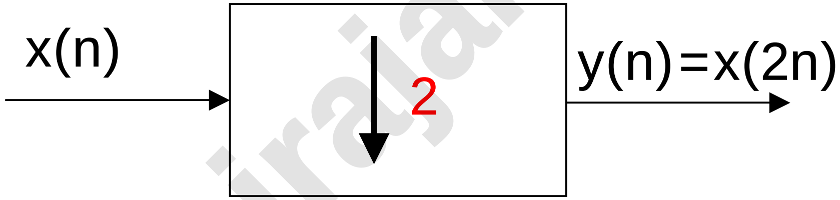
- Appropriate numbers for reordering the spectra can be illustrated, for example, using a 3 level Wavelet Packet decomposition tree as shown in Table 1.

Band No: ->	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**		**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
				L									**H**				
Level 1	1	2	3	4	5	6	7	8		16	15	14	13	12	11	10	9
			L			**H**						**L**			**H**		
Level 2	1	2	3	4	8	7	6	5		16	15	14	13	9	10	11	12
	L		**H**		**L**		**H**			**L**		**H**		**L**		**H**	
Level 3	1	2	4	3	8	7	5	6		16	15	13	14	9	10	12	11

Table 1: Appropriate numbers for reordering the sub-band spectra
L - Lowpass sub-band; H - Highpass subband

Appendix A

Decimation by 2



Aliasing term

$$Y(z) = \frac{1}{2}\left[X\left(z^{\frac{1}{2}}\right) + X\left(-z^{\frac{1}{2}}\right)\right]$$

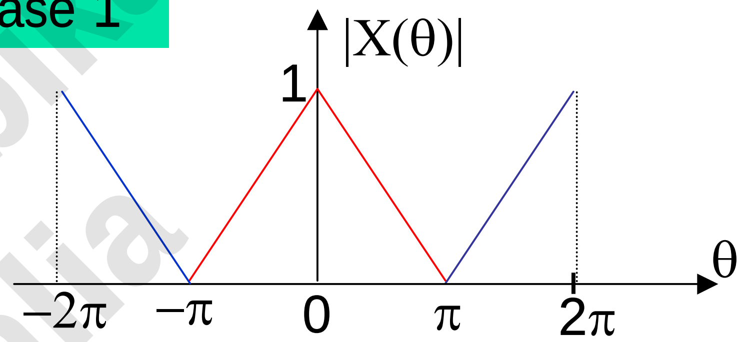
$$Y(\theta) = \frac{1}{2}\left[X\left(e^{\frac{j\theta}{2}}\right) + X\left(-e^{\frac{j\theta}{2}}\right)\right]$$

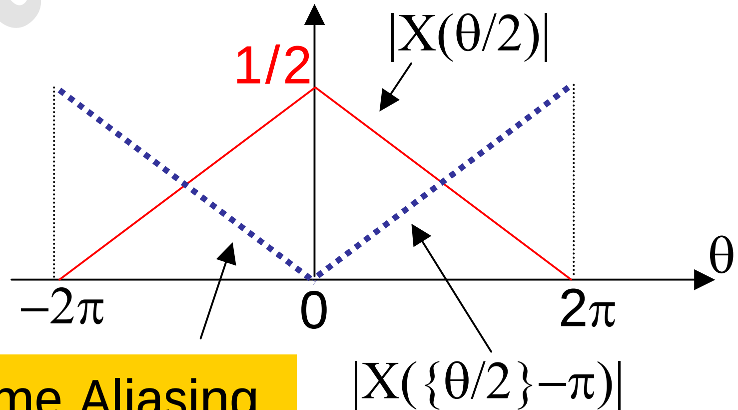
$$= \frac{1}{2}\left[X\left(e^{\frac{j\theta}{2}}\right) + X\left(e^{j\left(\frac{\theta}{2} - \pi\right)}\right)\right]$$

$$= \frac{1}{2}\left[X\left(\frac{\theta}{2}\right) + X\left(\frac{\theta}{2} - \pi\right)\right]$$

Aliasing term

Case 1



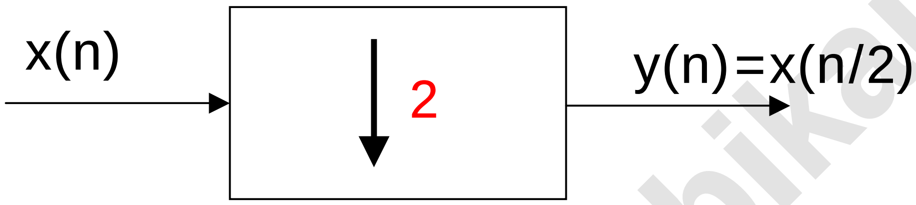


Extreme Aliasing

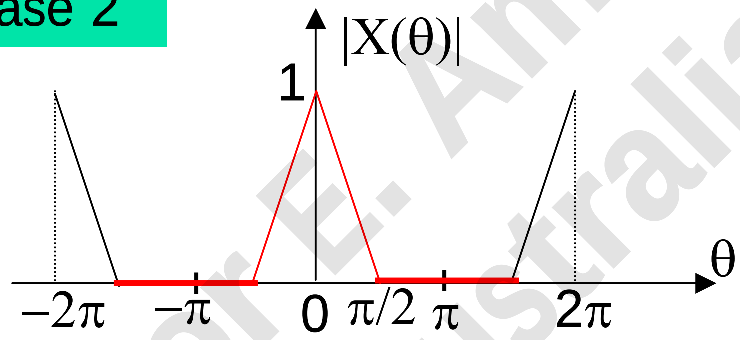
Stretch X(θ) by a factor 2 to obtain X(θ/2)

The spectrum is stretched by down sampling

Decimation by 2

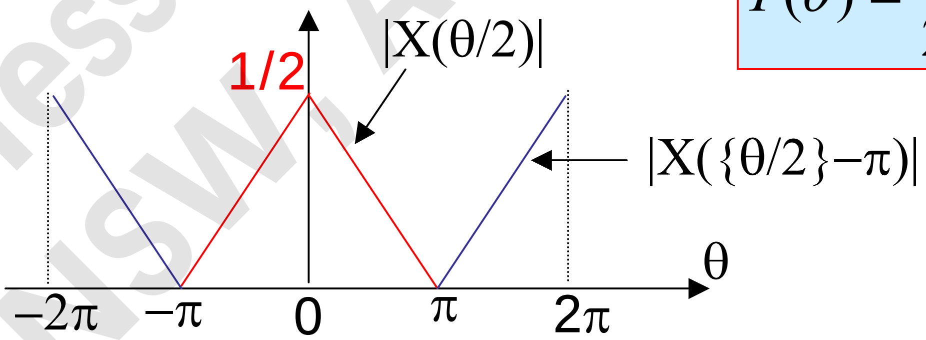


Case 2

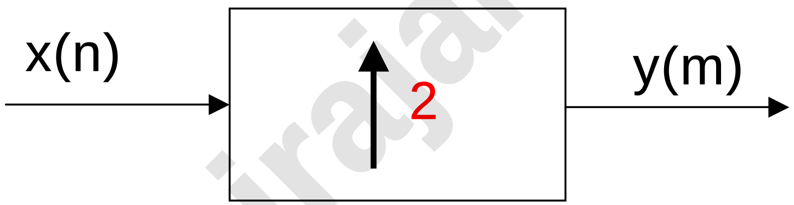


$$Y(\theta) = \frac{1}{2}[X(\frac{\theta}{2}) + X(\frac{\theta}{2} - \pi)]$$

No Aliasing

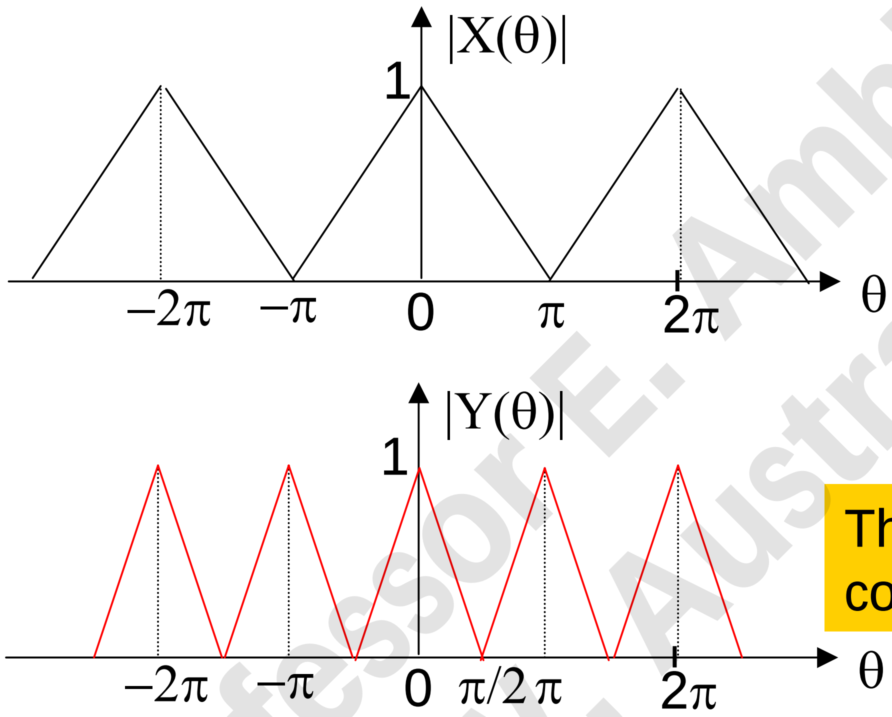


Interpolation by 2



$$Y(z) = [X(z^2)]$$

$$Y(\theta) = [X(e^{j2\theta})]$$
$$= [X(2\theta)]$$



The upsampled spectrum has compressed images of X(θ)

The spectrum is compresseed by upsampling