INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392 - I

SMR.379/31

COURSE ON BASIC TELECOMMUNICATIONS SCIENCE

9 January - 3 February 1989

An Introduction to the DISCRETE FOURIER TRANSFORM (DFT) and the FAST FOURIER TRANSFORM (FFT)

J.G. LUCAS

Dept. of Electrical Engineering, University of Sydney, Australia

CONTENTS.

Summary

Discrete transforms [e.g., the Discrete Fourier transform (DFT) and the Fast Fourier Transfrom (FFT)] are applied to discrete data points (which have typically been obtained by sampling a continuous waveform) and give a discrete result. In other words an input of sample values at equal discrete intervals of time give rise to numbers out at equal intervals of frequency.

To fully understand the relevance and usefulness of Discrete Transforms it is useful to start by considering the Fourier Series approach and then develop its relationship to the Fourier integral. The sampling of a continuous waveform to produce discrete data is then discussed and from there it is a simple matter to relate the Fourier integral to the Discrete Fourier Transform (DFT) and the Fast Fourier Transform (FFT).

In the final section the simple extension of these techniques to 2 dimensional fields of data is discussed.

1. FOURIER SERIES

It is well known that the majority of practical waveforms can be expressed in Fourier Series form as a constant plus an infinite series sum of sines and cosines. Sines and cosines are not the only functions which can be used for the summation (e.g., Walsh, Bessel and Legendre Function are also used) but sinusoids have the widest application and are probably the easiest to understand!

This is of course the Fourier series.

Specifically the Fourier series technique is applied to <u>periodic</u> functions which are defined by $f(t) = f(t \pm KT)$

where $K = 1, 2, 3 \ldots\ldots$

i.e., the periodic function has period T.

The Fourier series approach says that a periodic function can be represented by an infinite sum of sinusoidal terms plus a DC term as:

$$f(t) = {}^{a_0}/_2 + \sum_{n=1}^{\infty} a_n \cos(n\omega_0 t) + \sum_{n=1}^{\infty} b_n \sin(n\omega_0 t) = f(t + KT) \text{ -----(1)}$$

where K is any integer and ω_0 is related to the period by $\omega_0 = \frac{2\pi}{T}$.

The priviso that this representation is correct is that the function is well behaved (satisfies the Dirichlet conditions).

i.e., (i) that the integral of the function $f(t)$ i.e., $\int_{t'}^{t'+T} |f(t)|dt$ is finite over the finite interval T.

(ii) that there are only a finite number of finite discontinuities in the function.

All of these sinusoidal terms fit exactly into the basic (or fundamental) period T. When n=1 we have terms $\sin\omega_0 t$ and $\cos\omega_0 t$ which are the fundamental (or first harmonic) terms.

The coefficients from eqn. (1) are obtained by multiplying through by cosine and sine terms e.g., to find the coefficient a_n multiply (1) through by $\cos(m\omega_0 t)$:

$$\cos(m\omega_0 t)\ f(t) = \frac{a_0}{2}\cos(m\omega_0 t) + \left|\sum_{n=1}^{\infty} a_n \cos(n\omega_0 t)\right| \cos(n\omega_0 t)$$

$$+ \left|\sum_{n=1}^{\infty} b_n \sin n\omega_0 t\right| \cos m\omega_0 t \text{ ------------(2)}$$

Integrate both sides of this equation over the basic period T

and using the fact that

$$\int_t^{t+T} \cos(n\omega_0 t)\cos(m\omega_0 t)dt = 0 \quad : m \neq n$$
$$= 1 \quad : m = n$$

) these are called the "orthogonality" conditions.

$$\& \int_t^{t+T} \sin(n\omega_0 t)\cos(m\omega_0 t)dt = 0$$

to obtain:

$$a_n = \frac{2}{T}\int_{t'}^{t'+T} f(t)\cos(n\omega_0 t)dt \text{ ---------------------------(3)}$$

b_n can be similarly found by multiplying through by the sine term $\{\sin(m\omega_0 t)\}$ and again integrating.

Then:

$$b_n = \frac{2}{T}\int_{t'}^{t'+T} f(t)\sin(n\omega_0 t)dt \text{ ---------------------------(4)}$$

(Setting n=0 the DC term is simply obtained:)

$$\frac{a_0}{2} = \frac{1}{T}\int_{t'}^{t'+T} f(t)dt \text{ ---------------------------(5)}$$

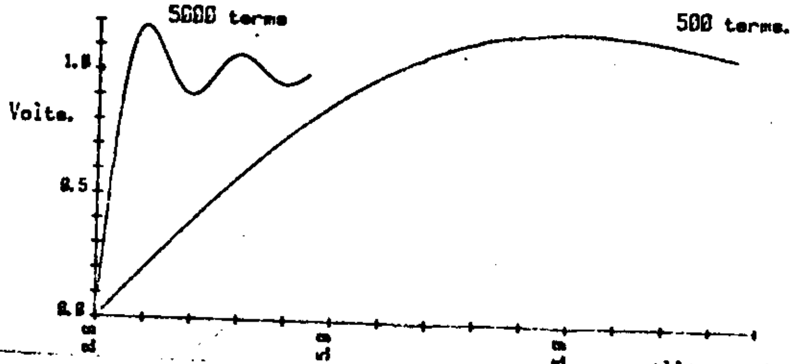
The result of the Fourier series summation is not exact when there are discontinuities involved. At a discontinuity the limit of the Fourier Series is equal to the mean value of the function at the discontinuity i.e., $\frac{y(t-) + y(t+)}{2}$. The effect when any finite number of terms is taken is an overshoot characteristic which is known as the Gibbs phenomenon.

An example best illustrates this effect:

Consider the Fourier Series analysis of a 50 Hz square wave.

The result is simply given by $f(t) = \sum_{n=1}^{N} \frac{1}{2n-1} \sin(2\pi(2n-1)ft)$ where f = 50Hz. Plotting this result near the origin for two values of N:

FOURIER SERIES ANALYSIS OF A 50 Hz SQUARE WAVE.



Increasing the number of terms brings the overshoot closer to the discontinuity - the peak value of the overshoot does not change as shown in the example of a square wave discontinuity.

Complex Form of Fourier Series

We can express $\cos x = \frac{e^{jx} + e^{-jx}}{2}$

& $\sin x = \frac{e^{jx} - e^{-jx}}{2j}$

So the Fourier series can be written as

$$f(t) = \frac{a_0}{2} + \sum_{1}^{\infty} a_n \left(\frac{e^{jn\omega_0 t} + e^{-jn\omega_0 t}}{2}\right) + \sum_{1}^{\infty} b_n (\qquad\qquad)$$

or

$$f(t) = \frac{a_0}{2} + \sum_{1}^{\infty} \tfrac{1}{2}(a_n - jb_n)\, e^{jn\omega_0 t} + \sum_{1}^{\infty} \tfrac{1}{2}(a_n + jb_n)\, e^{-jn\omega_0 t}$$

(this second term can be written as
($\sum_{-1}^{-\infty} \tfrac{1}{2}(a_n - jb_n)\, e^{jn\omega_0 t}$
(since $a_{-n} = a_n$ & $b_{-n} = -b_n$

$$\therefore \quad f(t) = \sum_{-\infty}^{\infty} \tfrac{1}{2}(a_n - jb_n)\, e^{jn\omega_0 t}$$

$$\text{or} \quad f(t) = \sum_{-\infty}^{\infty} C_n\, e^{jn\omega_0 t} \qquad \text{------------(6)}$$

where

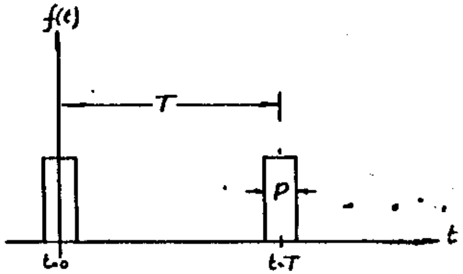
$$C_n = \tfrac{1}{2}(a_n - jb_n) = \frac{1}{T}\int_{t'}^{t'+T} f(t)\, e^{-jn\omega_0 t}\, dt \qquad \text{------------(7)}$$

When n=0 the DC term is clearly $a_o/2$.

II. THE FOURIER TRANSFORM

1. The Relationship between Fourier Series and the Fourier Transform

It will be shown that the Fourier Transform is a limiting form of the Fourier Series. This could equally have been done the other way round by showing that the Fourier Series is a limiting form of the Fourier Transform. In this note the Fourier Series was taken as the simplest starting point so it is logical to show the Fourier Transform as a limit of that series form. The result will be developed by considering a very simple example which is the repetitive pulse waveform shown.



The period is T and the pulse has length P, (all in seconds) where $P < T$.

The pulse only contributes to the integral for P/T of the total period. So the Fourier Series coefficients are (from (7)):

$$C_n = \frac{1}{T}\int_{-P/2}^{+P/2} 1.e^{-jn\omega_0 t}\,dt$$

For n = 0

$$C_o = \frac{P}{T}$$

for n = 1, 2, ...

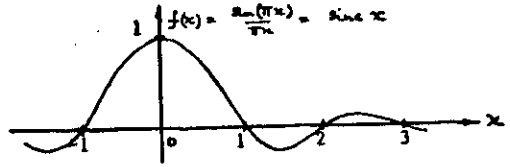
$$C_n = \frac{1}{-jn\omega_0 T}\left|e^{-jn\omega_0 t}\right|_{-P/2}^{P/2}$$

$$= \frac{1}{-jn\omega_0 T}\{\cos(n\omega_0 P/2) - j\sin(n\omega_0 P/2) - \cos(n\omega_0 P/2) - j\sin(n\omega_0 P/2)\}$$

$$\therefore \quad C_n = \frac{P}{T}\,\frac{\sin(n\pi P/T)}{n\pi P/T} \qquad \text{------------------------------(8)}$$

This has the form $\frac{\sin(\pi x)}{\pi x}$ which is often called sinc x.

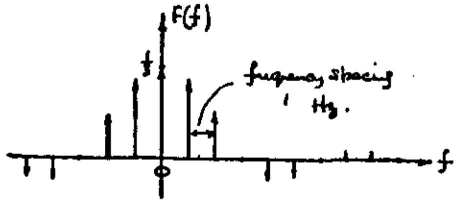
The sinc x function has the following infinite form:



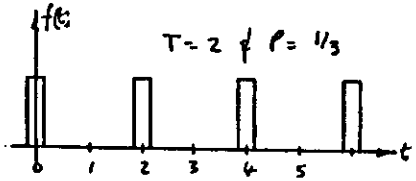
To see what the frequency domain result looks like consider the following numerical example:

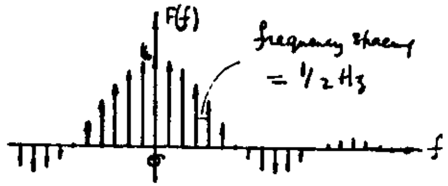
Let T = 1 second and P = 1/3 second.

Then plotting the dc term, the fundamental and harmonics as a series of impulses with appropriate amplitudes gives the following time/frequency graphs:

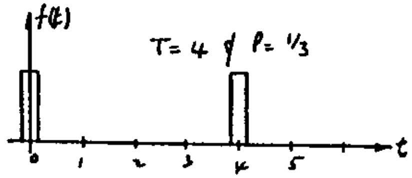


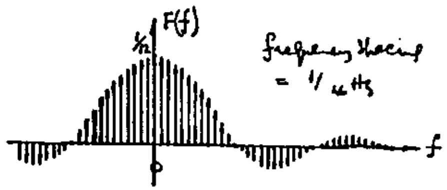
Removing every second pulse:





Removing every second pulse:





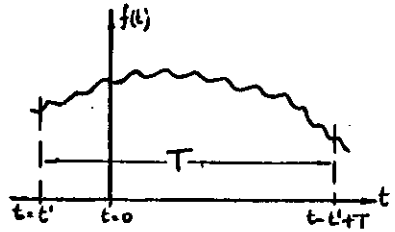
Now if the period T between pulses is taken bigger and bigger it can be seen that the spectrum is made up of more and more lines contained under the $\frac{\sin x}{x}$ envelope. The relative shape of the envelope does not change. In fact the absolute magnitude of the envelope reduces as the period increases since the signal energy is obviously decreasing.

However absolute spectrum magnitude is often not as important as the relative magnitude and it can be normalised to unity each time. Clearly as the limit of an infinite period is approached the lines move closer and closer together and finally when there is only one pulse remaining they become continuous. When they do so this is the Fourier Transform. It is immediately clear that the Fourier Transform applies to waveforms without a period i.e., to non-periodic or aperiodic waveforms.

These spectrum plots show the occurrence of both positive and negative frequencies in perfect symmetry about the zero frequency axis. The "negative" frequencies do not exist in practice but appear because the sinusoid is represented mathematically as the sum of two exponentials ($e^{j\omega t}$ and $e^{-j\omega t}$).

Developing the Fourier Transform from the Fourier Series.

Consider a general time function f(t) which is not assumed to be particularly periodic. In fact it can be forced to be periodic by considering any segment T long to be repetitive. Then let T get bigger and bigger i.e., move towards an 'aperiodic' signal. Before T goes to infinity the Fourier Series form can certainly be applied and it is clear that a segment T seconds long is being considered.



i.e.,

$$f(t) = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega_0 t} \qquad \text{---(9)}$$

where

$$C_n = \frac{1}{T}\int_{t'}^{t'+T} f(t)e^{-jn\omega_0 t}\,dt$$ where referring to the diagram $t' < t < t'+T$ ---(10)

So in fact the signal will be built up of component frequencies $n \times \frac{2\pi}{T}$, T being the basic period.

Since we are taking T to be very large indeed the integral may be made symmetrical about the origin as that will not affect the result. Then obtain:

$$C_n = \frac{1}{T}\int_{-T/2}^{+T/2} f(t)\, e^{-jn\omega_0 t}\,dt \qquad \text{---(11)}$$

Substituting C_n in the Fourier Series form of f(t) shown in equation (2) gives:

$$f(t) = \sum_{n=-\infty}^{\infty} \left(\frac{1}{T}\int_{-T/2}^{+T/2} f(t)\, e^{-jn\omega_0 t}\,dt\right) e^{jn\omega_0 t} \qquad \text{---(12)}$$

The ω_0 here is $\frac{2\pi}{T}$ (i.e., the fundamental).
As T gets bigger and bigger then $\frac{1}{T}$ becomes very small indeed and since it has the dimensions of frequency it can be called: $\frac{1}{T} = \Delta f$

$$\therefore \quad f(t) = \sum_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(t)\, e^{-j2\pi ft}\,dt\right] e^{j2\pi ft} df \qquad \text{---(14)}$$

now in equation 14 let:

$$F(f) = \int_{-\infty}^{\infty} f(t)\, e^{-j2\pi ft}\,dt$$

and then:

$$f(t) = \int_{-\infty}^{\infty} F(f)\, e^{j2\pi ft}\,df \qquad \text{---(15)}$$

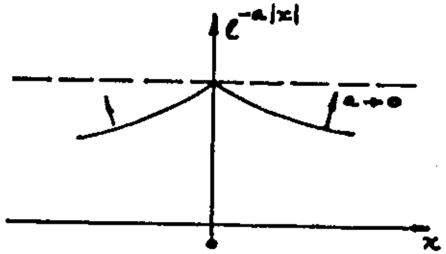
Notice the notation which is used, capital letters for functions in the Transform Domain and small case letters for functions in the original domain. This is the simplest symmetrical form of the Fourier Transform and will be used exclusively here.

In the mathematical development the 2π could have been reorganised in a number of ways. However the simple symmetrical form presented here is preferred since in the other arrangements premultipliers of $\frac{1}{2\pi}$ or $\frac{1}{\sqrt{2\pi}}$ have to be (painfully) accounted for.

Ref.

The Fourier Transform of a Periodic Waveform

Strictly speaking the Fourier Transform applies to aperiodic signals and considerable effort is now required to obtain the Fourier transforms of useful signals like the impulse, the step or to a periodic signal like a sinusoid. None of these mathematically satisfies the existence conditions of the transform and to get around this problem it is often necessary to premultiply the function to be transformed by $e^{-a|x|}$ as in the Figure. This always ensures a finite waveform and the required transform is then obtained by setting a equal to zero in the result.



The sinusoid logically has a line spectrum and has an impulsive form (see Appendix 2) in the frequency domain.

With periodic waveforms an approach which generally allows the development of the transform is to consider the Fourier series form of the periodic function:

$$f(t) = \sum_{n=-\infty}^{\infty} C_n e^{+j2\pi n f_0 t} \qquad \text{here } f_0 = \frac{1}{T} \text{ ---------------(16)}$$

where the C_n are the complex constants.

Then the Fourier Transform (FT) can be simply taken as:

$$FT(f(t)) = FT\left(\sum_{n=-\infty}^{\infty} C_n e^{j2\pi n f_0 t}\right)$$

$$FT(f(t)) = \sum_{n=-\infty}^{\infty} C_n FT(e^{j2\pi n f_0 t}) \qquad \text{-----------------------------(17)}$$

Now the Fourier Transform of a single exponential is simply a (shifted) unit impulse (Appendix 2) so can be written as:

$$FT(f(t)) = \sum_{n=-\infty}^{\infty} C_n \mu_0(f - nf_0) \qquad \text{Where } \mu_0(x) \text{ is a unit impulse at } x = 0. \text{ ------------(18)}$$

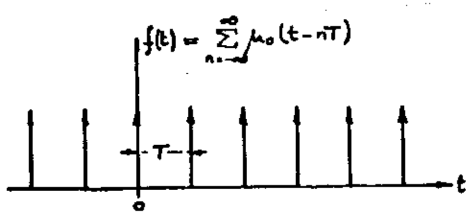
This result is a series of impulses of strength the Fourier Series Coefficients. So the transform of a PERIODIC function is always a series of impulses. This is a very important and quite general result.

3. Sampling

To process analogue signal information in a digital computer it is necessary to convert the analogue information into digital or number form. For example, using an A-D convertor. This convertor samples the analogue signal at equal intervals of time and it can be shown that, provided the spacing between samples satisfies certain constraints, then the original signal can be exactly reconstructed from these sample values. This is a remarkable result.

It is necessary first to develop a mathematical expression for sampling.

The product of a function by a unit impulse exactly samples the junction at the unit impulse instant. Repeated sampling of an analogue signal at equal intervals if time can be expressed as the product of the analogue signal by an infinite series of unit impulses of the form:



Clearly this is a periodic function of period T and can be expressed as:

$$f(t) = \sum_{n=-\infty}^{\infty} \mu_0(t-nT) \qquad \text{------(19)}$$

If f(t) is expressed as a Fourier Series

i.e.,

$$f(t) = \sum_{n=-\infty}^{\infty} C_n e^{j2\pi n f_0 t}$$

where $f_0 = \frac{1}{T}$ ------(20)

To find C_n the integral is taken over the basic interval T (i.e., from -T/2 to +T/2).

i.e.,

$$C_n = \frac{1}{T}\int_{-T/2}^{T/2} f(t) e^{-jn\omega_0 t}$$

$$= \frac{1}{T}\int_{-T/2}^{T/2} \mu_0(t) e^{-jn\omega_0 t}\, dt$$

) due to the impulse
) this integral only
) exists when
) t = 0.

$$= \frac{1}{T} \quad \text{(which is a constant)} \qquad \text{------(21)}$$

so

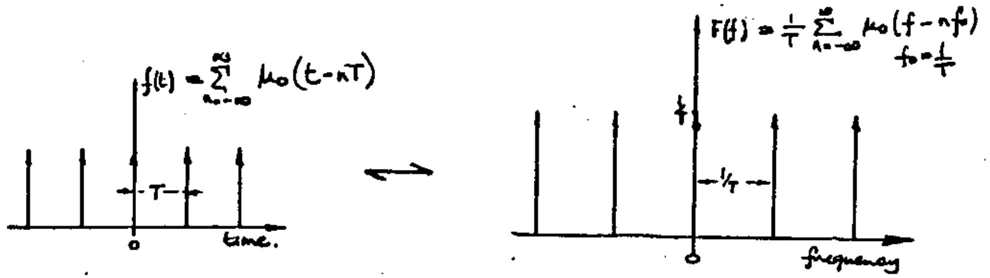
$$f(t) = \frac{1}{T}\sum_{n=-\infty}^{\infty} e^{j2\pi f_0 t}$$

Again this is the sum of simple exponentials. ------(22)

So that the Fourier Transform can be directly written as (see Appendix 2)

$$F(f) = \frac{1}{T}\sum_{n=-\infty}^{\infty} \mu_0(f - nf_0) \qquad f_0 = \frac{1}{T} \qquad \text{------(23)}$$

So we have a Fourier Transform pair:



Plotting this result it can be seen that the function is effectively its own Fourier Transform. This particular function is very widely used since it describes the process of sampling and is given the special symbol Ш(x) (pronounced 'shah') with impulses occurring at integral values of x. With the time samples which occur at interval T (on page 15) the sampling function is expressed as Ш(t/T).

The transformed frequency samples which are shown above which occur at a frequency interval of 1/T are expressed as Ш(tT).

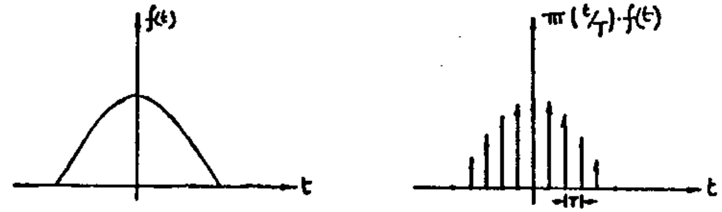
Thus we can write the relationship between the sampling function and its transform in compact notation as:

$$Ш(t/T) \leftrightarrow \tfrac{1}{T}Ш(Tf) \qquad \text{--(24)}$$

A function of time f(t) sampled at interval T is then compactly expressed as:

$$Ш\left(\frac{t}{T}\right) f(x) = \sum_{n=-\infty}^{\infty} f(nT)\mu_0(t-nT) \qquad \text{--------------------------------(25)}$$

A pictorial example

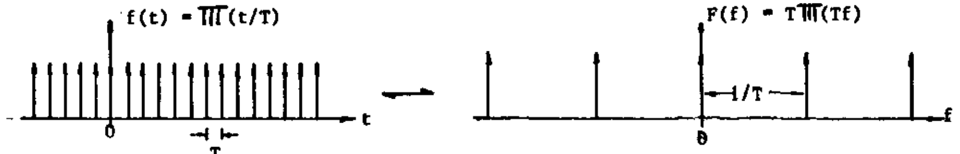


It is important to note that in any practical sampling situation there can only be a FINITE number of data points. In other words the samples are GATED. In the transform domain the effect of this finite gate has to be carefully taken into account. The transform of the gate is developed in Appendix 1.

3.1 Sampling Rate Requirements

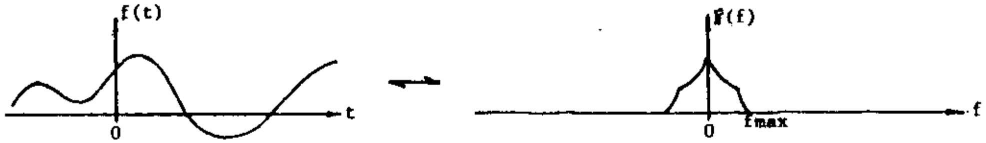
In order to work out the minimum sampling rate which is required so that the original analogue signal can be completely reconstructed in the computer this sampling process must be carefully considered in both the time and frequency scale consider first the sampling function:

Time Domain | Frequency Domain



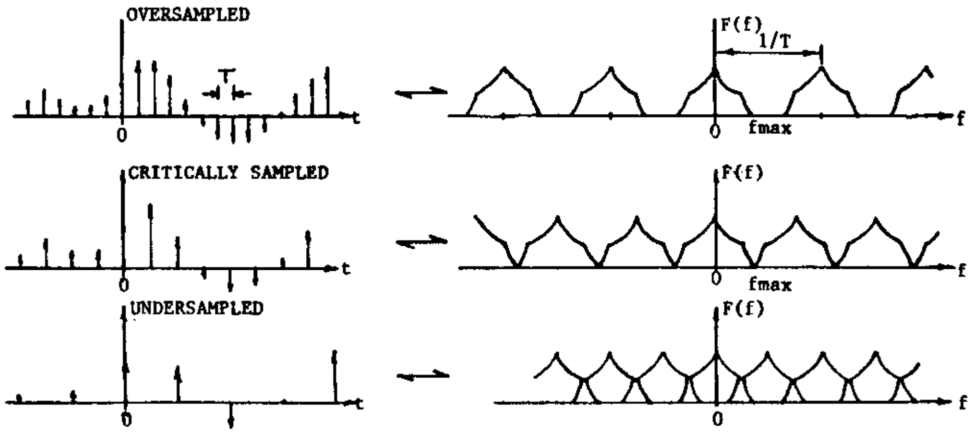
Now the continuous waveform to be sampled is f(t):

and this has the transform F(f): (which is bandlimited *1)[1]
*see footnote



In the time domain the sampled function is Ш(t/T) ×f(t)

Now multiplication in the time domain is equivalent to convolution[2] in the frequency domain which in this case means replication.



[1]The frequency spectrum is illustrated here as a real even function for simplicity. In practice there would be an odd imaginary component also.
[2]The concept of convolution is discussed in Appendix 2.

Parts C, D and E of this diagram allow the deduction of the very important sampling theorm which states that an analogue signal can be completely reconstructed from its sampled values provided:

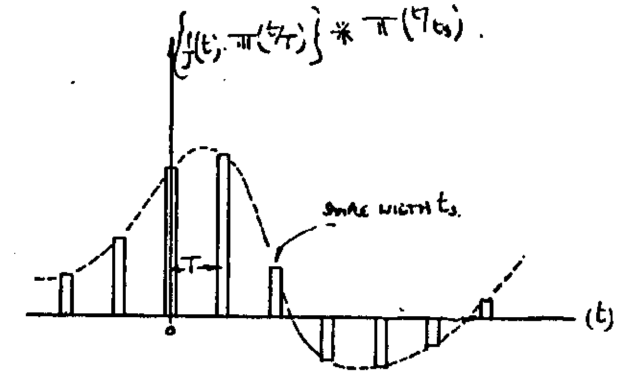
(i) that the signal being sampled is band limited (which simply means that it contains no frequencies above f_{max})

and

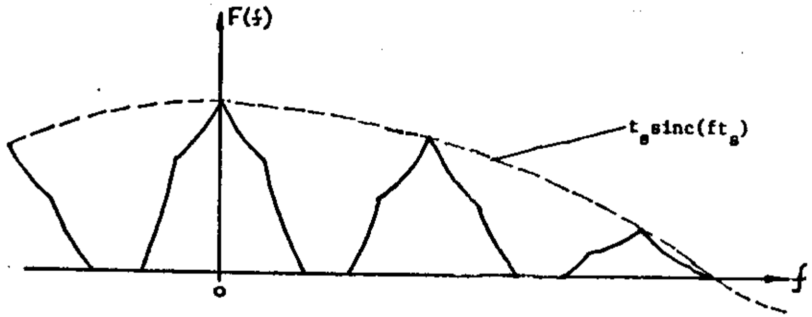
(ii) that the signal is sampled at a rate which is at least twice the highest frequency component contained in this signal (i.e., $f_{sample} \geq 2 f_{max}$).

5. PRACTICAL SAMPLING PROBLEMS

It is worth mentioning some of the problems which are associated with sampling. The first and most obvious is that the samples are not perfect impulses of zero time duration but are finite in length.



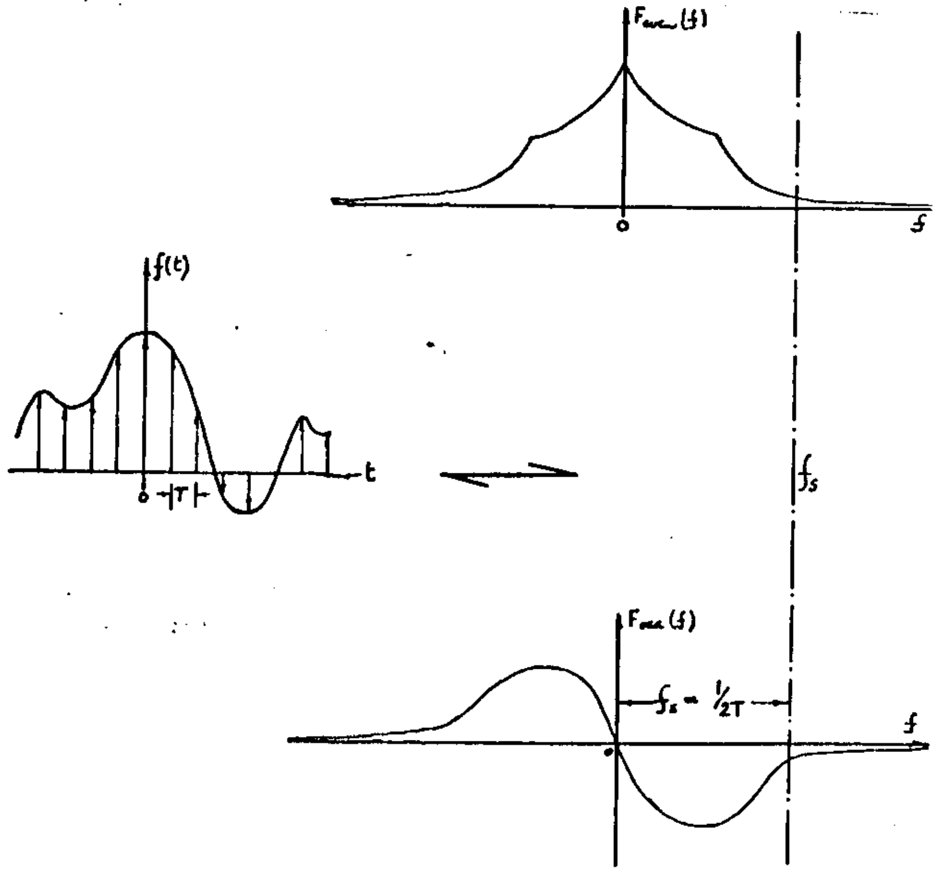
This has the effect in the frequency domain of:



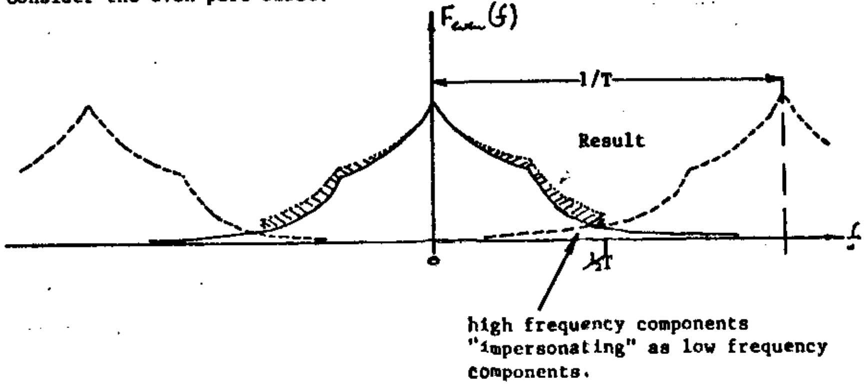
It is clear that the sample width can be greater than 10% of the sampling period T without severe distortions to the useful central spectrum.

The second effect is rather more subtle. What happens when the sampling frequency or sampling rate is not sufficient? In

practice a signal is not contained within a definite band of frequencies but rather tends to "tail off" in the frequency domain. These tails approximate an exponential decay which can continue a long way up the frequency range. A decision on where the "cut-off" occurs is then made on a cost basis and sampling is carried out accordingly. In such a case what results? The effects are most easily visualised by separating the transform into even and odd parts.

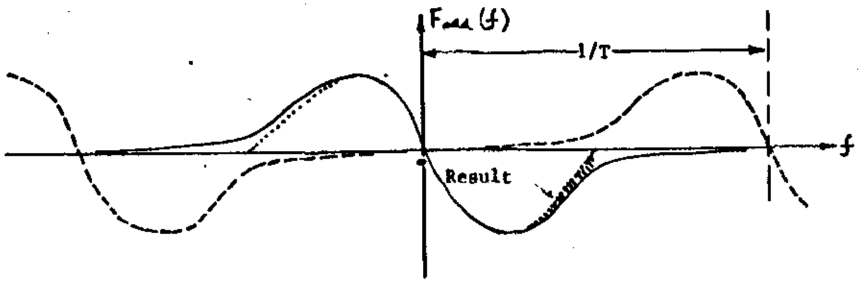


Consider the even part first:



This even part of the spectrum is clearly reinforced by undersampling. This effect is known as "aliasing".

Consider the odd part.



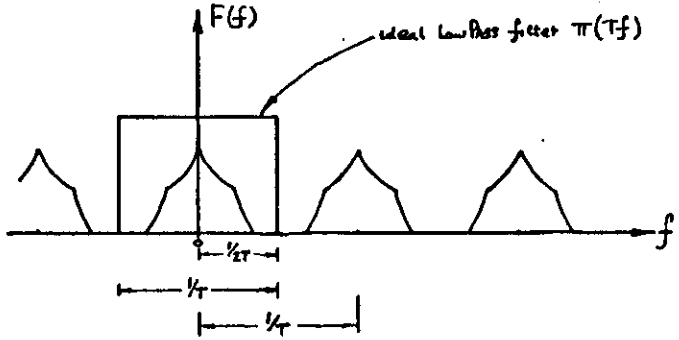
Its clear that the result in this case is diminished tails, i.e., with undersampling the odd part is reduced and the even part increased and the overall result is more even than the original function was.

This aliasing is an inevitable result of undersampling because the higher frequency components can never be retrieved from the resulting spectrum and of course gives an error in the low frequency result.

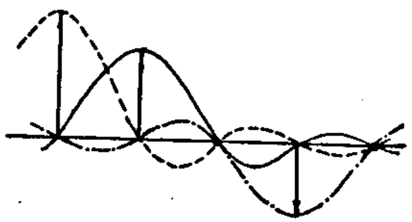
4. INTERPOLATION

If a function has been sufficiently sampled (the Nyquist criterion) so that all original information has been retained (i.e., at greater than or equal to twice the highest frequency component) then it is possible to mathematically reproduce every point of the original analogue signal (between the samples). This process is known as interpolation.

If this set of signal samples is passed through an ideal low pass filter (unity gain, linear phase) which has a cut off frequency of exactly half the replication frequency then in the frequency domain this is equivalent to multiplication by a gate of width 1/T i.e., $\pi(Tf)$. With such a mathematically perfect low pass filter this removes all the replications and retains only the original signal:

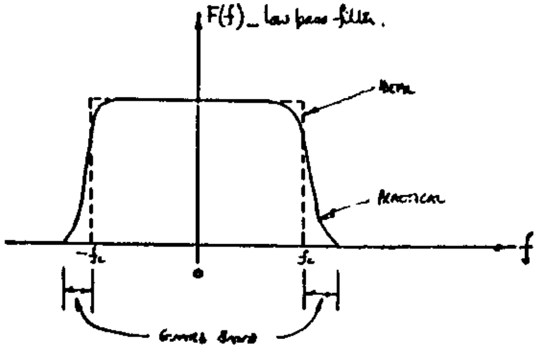


This is exactly equivalent in the time domain to convolving the the input sample set with the inverse gate transform which is a sinc function. This is done by erecting a sinc function on each sample value whose central value is equal to that sample. Choosing a low pass filter which cuts off precisely at a frequency of 1/2T ensures that the zero values of the sinc function occur at all the other sample points, i.e.,:



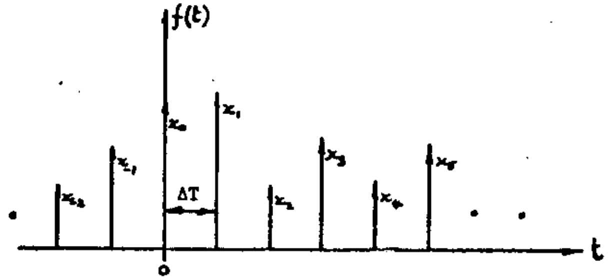
Thus the sample values are unaffected while the regions <u>between</u> sample values are exactly reconstructed provided the sampling rate is greater than or equal to the Nyquist rate.

Direct analogue retrieval can also be carried out. However it is not possible to construct a filter with such steep edges <u>and</u> linear phase shift, and a guard band allowance must be made.



6. THE DISCRETE FOURIER TRANSFORM

The move towards the Discrete Fourier Transform (DFT) is really very simple. Consider a series (which shall be assumed to be infinite for the moment) of discrete data values which are equally spaced.



then

$$f(t) = \ldots + x_0\mu_0(t) + x_1\mu_0(t-\Delta T) + x_2\mu_0(t-2\Delta T) + \ldots\ldots \equiv Ш(t/\Delta T)\cdot f(x)$$

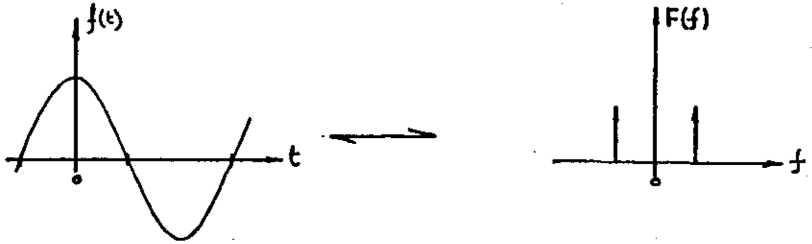
where ΔT is the sampling interval. ------------------------------(26)

The Fourier Transform can be directly written as:

$$F(f) = x_0 + x_1e^{-j2\pi f\Delta T} + x_2e^{-j2\pi f2\Delta T} + x_3e^{-j2\pi f3\Delta T} + \text{etc.} \qquad \text{------(27)}$$

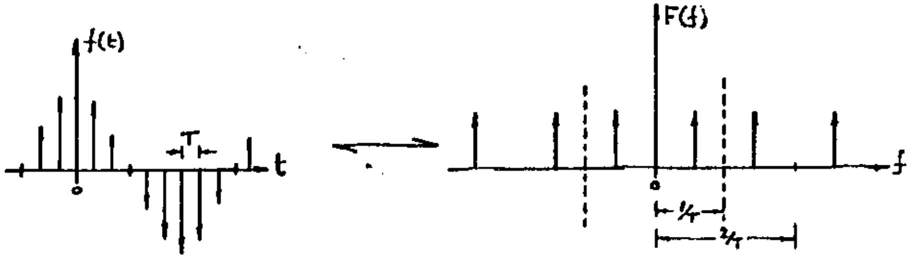
This expression is in fact a "discrete" Fourier Transform - but in this simple form it is still a <u>continuous</u> function of frequency f. However in a digital computer this function can only be stored as a set of discrete values. Intuitively it is clear that each sampling impulse contains wideband information so whether or not the sampled signal is wideband there is no doubt that the resulting spectrum will be wideband. i.e., the sampled signal will be more wideband than the original signal was!

Consider then a simple cosine wave and its Fourier Transform



If this simple signal is now sampled with interval T the sample set and its spectrum become:

i.e., the basic Fourier Transform is endlessly replicated



The Simple Discrete Fourier Transform (DFT) formulation

Consider the set of N equispaced samples of a signal:

Notice that a particular sample represents the period in time of one half the sample width to each side of the sample.

The N samples extend then over total time $T = N*\Delta T$ and the continuous Fourier Transform of this function would be defined as:

$$F(f) = \int_{-T/2}^{T/2} f(t)\, e^{-j2\pi ft}\, dt \qquad \text{------------(28)}$$

Since total $T = N*\Delta T$: $\Delta T = T/N$.

Particular sampling points occurs at $t = k\Delta T$, where $k = 0, 1, \ldots\ldots, N-1$.

The 'fundamental' frequency (of the Fourier series form) is:

$$\Delta f = \frac{1}{T} = \frac{1}{N\Delta T} \qquad \text{------------(29)}$$

The frequency scale will be made up of 'harmonics' of this 'fundamental'.

i.e., $f = m\Delta f = \frac{m}{N\Delta T}$ for $m = 0, 1, \ldots\ldots, N-1$.

The continuous Fourier integral Equation (28) can be approximated as a summation:

$$F(m\Delta f) = \sum_{-\frac{N-1}{2}}^{\frac{N-1}{2}} f(k\Delta T)\, e^{-j2\pi \frac{m}{(N\Delta T)}(k\Delta T)} * T/N$$

$$= \frac{T}{N} \sum_{-\frac{N-1}{2}}^{\frac{N-1}{2}} f(k\Delta T)\, e^{-j2\pi \frac{mk}{N}} \qquad \text{------------(30)}$$

Obviously the approximation will be better as ΔT gets smaller.

Normally the exponential term $e^{-j2\pi\left(\frac{mk}{N}\right)}$ is called a "Fourier Weight"

$e^{j\frac{2\pi}{N}}$

and given the symbol W_N.

So in the final form:

$$F(m\Delta f) = \frac{T}{N} \sum_{-\frac{N-1}{2}}^{\frac{N-1}{2}} f(k\Delta T)(W_N)^{-mk} \qquad \text{------------------------------------(31)}$$

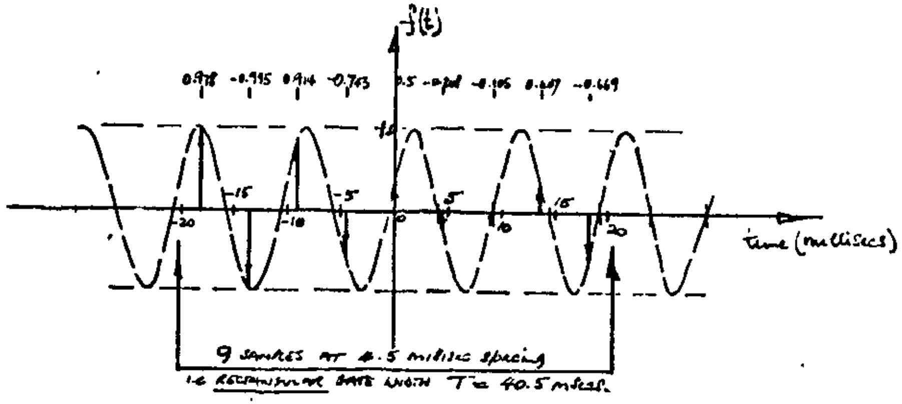
Unless specifically needed the term T/N is set to unity.

This is the form of the DISCRETE FOURIER TRANSFORM.

Consider a simple numerical example:

A simple numerical example illustrates the mechanics of performing a Discrete FourierTransform (DFT) and quite incidentally some of the difficulties involved with interpreting the result.

Consider a 100 Hz sine wave which sampled 9 times at 4.5 msec. intervals.



i.e., input data:

$f_{-4} = 0.978$

$f_{-3} = -0.995$

$f_{-2} = 0.914$

$f_{-1} = -0.743$

$f_0 = 0.5$

$f_1 = -0.028$

$f_2 = -0.105$

$f_3 = 0.407$

$f_4 = -0.669$

and T = 40.5 millisecs.

i.e., frequency resolution $= \frac{1}{T} = 24.69$ Hz.

Then the Discrete Fourier Transform of these samples is obtained from equation (31) as:

$$F(m\Delta F)_{\substack{m=-4 \text{ to} \\ +4 \text{ by } +1}} = \sum_{k=-4}^{+4} f(k\Delta T) W_N^{-mk} \quad : \text{where } W_N = e^{j\frac{2\pi}{9}}; \quad F = \frac{1}{40.5(\text{msecs})} = 24.69 \text{ Hz.}$$

i.e.,

$$F(-4\times 24.69) = (.978 e^{-j\frac{2\pi}{9}.-4.-4} + (-.995) e^{-j\frac{2\pi}{9}.-4.-3} + (.914) e^{-j\frac{2\pi}{9}.-4.-2} + (-.743) e^{-j\frac{2\pi}{9}.-4.-1} + (.5) e^{-j\frac{2\pi}{9}.4.0} + (-.208) e^{-j\frac{2\pi}{9}.-4.1}$$

$$+ (-.105) e^{-j\frac{2\pi}{9}.-4.2} + (.407) e^{-j\frac{2\pi}{9} -4.3} + (-.669) e^{-j\frac{2\pi}{9} -4.4}$$

$= 4.37\underline{/57.3^o} = 2.36 + j3.67$

F(-3) $= 0.19\underline{/-155^o} = -.17 - j.08$

F(-2) $= 0.11\underline{/13^o} = +.105 + j.024$

F(-1) $= 0.09\underline{/-174^o} = -.08 - j.008$

F(0) $= 0.08$ (simply the sum of the samples)

F(1) $= 0.09\underline{/174^o}$

F(2) $= 0.11\underline{/-13^o}$

F(3) $= 0.19\underline{/155^o}$

F(4) $= 4.37\underline{/-57.3^o}$

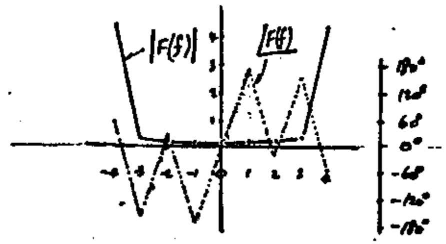
F(5) $= 4.37\underline{/57.3^o}$

F(6) $= 0.19\underline{/-155^o}$

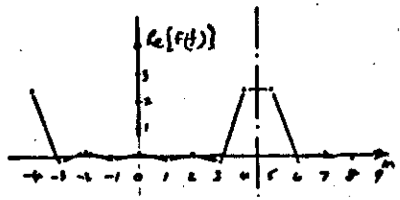
F(7) $= 0.11\underline{/+13^o}$

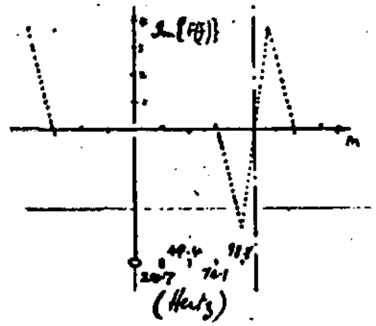
F(8) $= 0.09\underline{/-174^o}$

Now we can plot this result in magnitude and phase:



or we can separate this result into real and imaginary parts as:

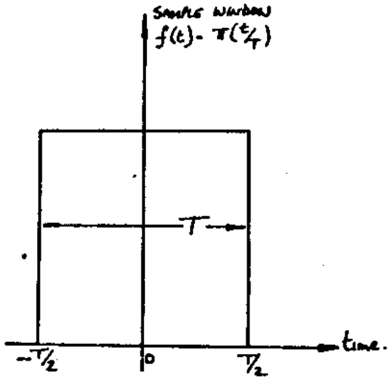


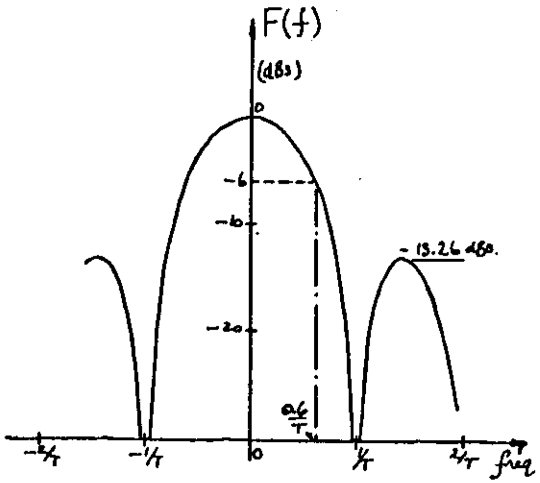


This latter display is much more revealing for sinusoidal components because we can directly estimate the amount of cosine (even component) and sine (odd component) contained in the original. It is perhaps not surprising that the DFT most efficiently picks out sinusoidal components since they are the basis of its formulation.

The 9 samples constitute a narrow window of only 40.5 msecs. so that we expect the result in the frequency domain to be convolved with a sinc function and this is very evident in the DFT result shown here. The sinc broadening of the lines in the frequency sprectrum is largely due to the sharp edges of the gate which has been used in the time domain.

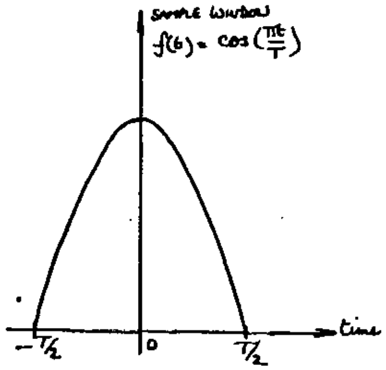
In effect each frequency contained in the input waveform is convolved with the sinc function shown in Fig. (page 7).

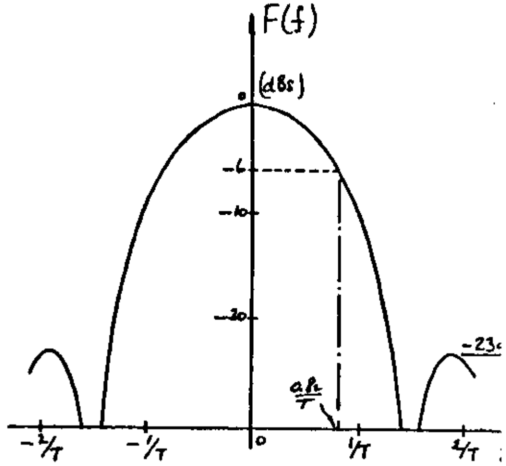




If only a single frequency is present in the time samples (as in this numerical example) the resulting frequency result is then a sampled version of this sinc function. If the input signal frequency happens to be an exact multiple of the discrete output frequency (=1/T) a perfect result will seem to be obtained as the frequency samples then occur at all the zeroes of the sinc function. For all other cases and when more than one frequency is present this "leakage" effect of whichever window is used will be evident. The possible frequency resolution of the transform is also limited by the sinc function and is simply given by the frequency spacing between points where the sinc drops to ½ of its maximum value. The frequency resolution in this case is given be $2 \times \frac{0.6}{T} = 1.2 \times (1/T)$.

The effects of such spectrum spreading can be reduced by applying carefully shaped windows to the input data. The simplest example of a window is the simple cosine function shown in Fig. following (this is also called the HANNING window of zero order).





Its immediately obvious from this transform pair that the level of the frequency domain sidelobes has been substantially reduced in comparison to the sinc frequency function in Fig. (page 31). Another important change is that the frequency width to ½ level (i.e., the resolution ability of the transform) has also been significantly widened. The frequency resolution in this case is given by $2 \times \frac{0.82}{T} = 1.64 \times (1/T)$.

Unsymmetical data windows

In our discussion so far the DFT has been expressed in the form where the window of samples is perfectly symmetrical about the origin. This ensures the purity of the phase of the resulting transform i.e., the DFT operation does not (by itself) introduce any errors in the phase of the result.

The more common formulation of the DFT is expressed as :

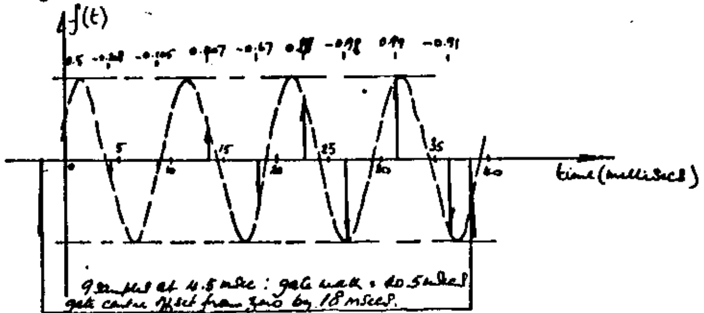
$$F(m) = \sum_{k=0}^{N} f(k)\, e^{\frac{-j2\pi km}{N}}$$

or

$$= \sum_{k=0}^{\omega} f(k)\, w_N^{mk} \qquad \text{where as usual we define } w_N = e^{\frac{-j2\pi}{N}} \qquad \text{....EQN. (32)}$$

where N is the total no. of samples.

If we take, for example, the time domain samples for the same 100 Hz sine wave which we considered previously, but with the samples now starting from zero time:



Drawing the window about these sample set it is clear that the centre of the window has a "offset from zero". This offset effectively applies a linear phase to the transform result i.e., comparing this transform result with that obtained previously:

Transform result with symmetrical window		Transform result for this offset window.	Phase modification due to unsymmetrical window.
$F_{-4} = 4.37/\underline{57^o}$	compared with:	$F_{-4} = 4.46/\underline{235^o}$	+4 x 134.5°
$F_{-3} = 0.19/\underline{-155^o}$		$F_{-3} = 0.24/\underline{283^o}$	+3 x 146°
$F_{-2} = 0.11/\underline{13^o}$		$F_{-2} = 0.14/\underline{312^o}$	+2 x 149.5°
$F_{-1} = 0.09/\underline{-174^o}$		$F_{-1} = 0.11/\underline{336^o}$	+1 x 150°
$F_0 = 0.08/\underline{0^o}$		$F_0 = 0.11/\underline{0^o}$	0°
$F_1 = 0.09/\underline{174^o}$		$F_1 = 0.11/\underline{24^o}$	-1 x 150°
$F_2 = 0.11/\underline{-13^o}$		$F_2 = 0.14/\underline{48^o}$	-2 x 149.5°
$F_3 = 0.19/\underline{155^o}$		$F_3 = 0.24/\underline{77^o}$	-3 x 146°
$F_4 = 4.37/\underline{-57^o}$		$F_4 = 4.46/\underline{125^o}$	-4 x 134.5°

The component magnitudes in the two transform results are virtually identical - any difference being attributable to the small number of samples being considered and that half the samples change their value in the second case. If, however, the phase results in the cases are compared it is clear that the result in the second case (with the unsymmetrical window) has been considerably modified. In fact a phase taken which is close to linear has been applied to the result. Since the window has been offset from zero by 18 milliseconds in this example the exact phase modification for a continuous signal would have been given by $e^{-j2\pi Tf}$ i.e., a phase taper of -6.48 f degrees. In this example the discrete frequency step is 24.7 Hz. giving an exact phase taper of -160 N degrees. The digital result is thus obviously of the correct order. With more samples and larger gate widths the correlation between discrete and exact results would be very much more accurate.

So long as one is aware of the artifact it is a simple matter to make allowances for its effect in the program detail.

For the remainder of this note we shall make use of the standard formulation of Eqn. 32 in our development.

(THERE IS NO PAGE 36)

It is immediately obvious from this transform pair that the level of the frequency domain sidelobes has been substantially reduced in comparison to the sinc frequency function in Fig. . Another important change is that the frequency width to ½ level (i.e., the resolution ability of the transform) has also been significantly widened. The frequency resolution in this case is given by

$2 \times \frac{0.82}{T} = 1.64 \times (1/T)$ (cf 1.2 × 1/T for the sinc function!)

. . . . .

In the discussion so far the DFT has been expressed in the form where the window of samples is perfectly symmetrical about the origin. This ensures the purity of the phase of the resulting transform i.e., the DFT operation does not (by itself) <u>introduce</u> any errors in the phase of the result.

The more common formulation of the DFT is expressed as:

$$F(m) = \sum_{k=0}^{N-1} f(k)\, e^{\frac{-j2\pi km}{N}}$$

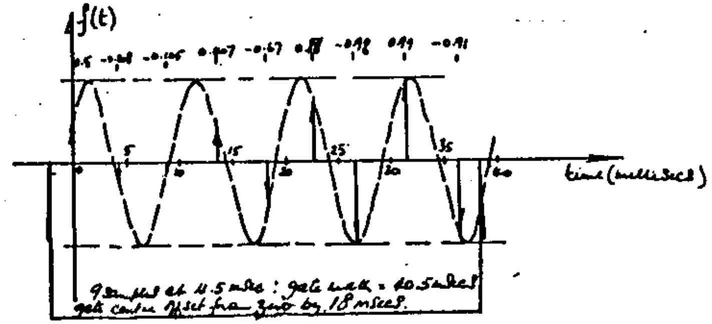
or

$$= \sum_{k=0}^{N-1} f(k)\, W_N^{-mk}$$

where as usual we define $W_N = e^{\frac{+j2\pi}{N}}$

where N is the total No. of samples. ...EQN.

If we take, for example, the time domain samples for the same 100 Hz sine wave which we considered previously, but with the samples now starting from zero time:



Drawing the window about this sample set it is clear that the centre of the window has a "offset from zero". This offset effectively applies a "linear" phase to the transform result. i.e., comparing this transform result with that obtained previously:

Transform result with symmetrical window		Transform result for this offset window	Phase modification due to unsymmetrical window
$F_{-4} = 4.37\angle 57^\circ$	Compared with:	$F_{-4} = 4.46\angle 235^\circ$	+4 x 134.5°
$F_{-3} = 0.19\angle -155^\circ$		$F_{-3} = 0.24\angle 283^\circ$	+3 x 146°
$F_{-2} = 0.11\angle 13^\circ$		$F_{-2} = 0.14\angle 312^\circ$	+2 x 149.5°
$F_{-1} = 0.09\angle -174^\circ$		$F_{-1} = 0.11\angle 336^\circ$	+1 x 150°
$F_0 = 0.08\angle 0^\circ$		$F_0 = 0.11\angle 0^\circ$	0°
$F_1 = 0.09\angle 174^\circ$		$F_1 = 0.11\angle 24^\circ$	-1 x 150°
$F_2 = 0.11\angle -13^\circ$		$F_2 = 0.14\angle 48^\circ$	-2 x 149.5°
$F_3 = 0.19\angle 155^\circ$		$F_3 = 0.24\angle 77^\circ$	-3 x 146°
$F_4 = 4.37\angle -57^\circ$		$F_4 = 4.46\angle 125^\circ$	-4 x 134.5°

The component magnitudes in the two transform results are very similar - any difference being attributable to the small number of samples being considered and that half the samples change their value in the second case. Comparing the phase results, however it can be seen that the result in the second case (with the unsymmetrical window) has been considerably modified. In fact a phase taper which is close to linear has been applied to the result. Since the window has been offset from zero by 18 milliseconds in this example the expected phase modification for a continuous signal would be $e^{-j2\pi(\text{gate offset})f}$ i.e., a phase taper of -6.68f degrees. In this example the discrete frequency step is 24.7 Hz. giving an exact phase taper of -160m. degrees. The result for the 9 sample set is of the correct order. With more samples and larger gate widths the phase taper in the discrete case would closely approach the theoretical (continuous) result.

So long as one is aware of this artifact it is a simple matter to make allowances for its effect in the program detail.

COMPLEX SAMPLE INPUT

The DFT copes equally well with complex data. An illustrative example is given by the components of a rotating unit vector: of 100 Hz. with a leading phase of 30°:

i.e., input cos (100 Hz. + 30°) as the real part | at 2.5 msec. intervals.
and sin (100 Hz. + 30°) as the imaginary part |

The input data set is:

-0.2079 + j0.9781
-0.9135 - j0.4067
0.9945 + j0.1045
-0.9781 + j0.2079
0.8660 + j0.5000
-0.9781 - j0.2079
0.9945 - j0.1045
-0.9135 + j0.4067
-0.3079 - j0.9781

The result is:

F_{-6}	2.6896 $\angle 169.3^{\circ}$
F_{-5}	7.7085 $\angle 3.7^{\circ}$
F_{-4}	2.4967 $\angle 11.6^{\circ}$
F_{-3}	1.2117 $\angle 24.4^{\circ}$
F_{-2}	0.9602 $\angle 148.6^{\circ}$
F_{-1}	1.5383 $\angle 19^{\circ}$
F_{0}	1.4340 $\angle 159.6^{\circ}$
F_{1}	0.7653 $\angle 40.8^{\circ}$
F_{2}	0.8406 $\angle 143.5^{\circ}$
F_{3}	2.6896 $\angle 143.5^{\circ}$
F_{4}	7.7085 $\angle 3.7^{\circ}$
F_{5}	2.4967 $\angle 11.6^{\circ}$
F_{6}	1.2117 $\angle 24.4^{\circ}$

THE DEVELOPMENT OF THE FAST FOURIER TRANSFORM (FFT)

Now consider the computing requirements for N input samples to the Discrete Fourier Transform then the useful output is specified by:

$$F(m\Delta f)\Big|_{m=0 \text{ to } \frac{N-1}{2}} = \sum_{k=0}^{N-1} f(k)\, W_N^{km}$$

(f(k) and W_N^{km}: both complex numbers)

i.e., it requires approximately $\frac{N}{2}$ N complex multiplications plus some additions. Since multiplications are orders of magnitude more time consuming than additions in terms of computing effort it can be said that the time required to compute the DFT is proportional to N^2 simple multiplications.

Now suppose that the N sample set is divided into two subsets each of length N/2. Clearly each subset will require $(\frac{N}{2})^2$ multiplications and the total computing requirement is then $2 \times (\frac{N}{2})^2 = \frac{N^2}{2}$ which is half the original! Further subdivision must obviously continue to reduce the number of multiplications required. If the number of samples N which is used is a power of 2 then the number of multiplications required by the subdivision method is $N \log_2 N$. When N is large the difference between N^2 and $N \log_2 N$ implies a dramatic saving in computing time! This then is the basis of the FFT.

Really where the economy in multiplication comes from is the fact that the Fourier weights are complex exponentials which are cyclical and simply repeat themselves.

This whole process is best illustrated by considering a set of 8 $(=2^3)$ time samples f_0 .... f_7.

assume DFT form : $F(m)_{m=0,\ldots,N} = \sum_{k=0}^{N} f(k) W_N^{mk}$ where $W_N^{mk} = e^{+j\frac{mk}{N}\cdot 2\pi}$

The transform result may be directly written as :

$$F(0) = W^0(f_0+f_1+f_2+f_3+f_4+f_5+f_6+f_7)$$
$$F(1) = W^0f_0+W^1f_1+W^2f_2+W^3f_3+W^4f_4+W^5f_5+W^6f_6+W^7f_7$$
$$F(2) = W^0f_0+W^2f_1+W^4f_2+W^6f_3+W^8f_4+W^{10}f_5+W^{12}f_6+W^{14}f_7$$
$$F(3) = W^0f_0+W^3f_1+W^6f_2+W^9f_3+W^{12}f_4+W^{15}f_5+W^{18}f_6+W^{21}f_7$$
$$F(4) = W^0f_0+W^4f_1+W^8f_2+W^{12}f_3+W^{16}f_4+W^{20}f_5+W^{24}f_6+W^{28}f_7$$
$$F(5) = W^0f_0+W^5f_1+W^{10}f_2+W^{15}f_3+W^{20}f_4+W^{25}f_5+W^{30}f_6+W^{35}f_7$$
$$F(6) = W^0f_0+W^6f_1+W^{12}f_2+W^{18}f_3+W^{24}f_4+W^{30}f_5+W^{36}f_6+W^{42}f_7$$
$$F(7) = W^0f_0+W^7f_1+W^{14}f_2+W^{21}f_3+W^{28}f_4+W^{35}f_5+W^{42}f_6+W^{49}f_7$$
$$F(8) = W^0f_0+W^8f_1+W^{16}f_2+W^{24}f_3+W^{32}f_4+W^{40}f_5+W^{48}f_6+W^{56}f_7$$

where:

$$\frac{2\pi}{N} = \frac{2\pi}{8} = 45^\circ$$
$$W^x = e^{-jx\,.\,45^\circ}$$

$x = 0 \quad W^0 = 1$ (always)

$x = 1 \quad W^1 = e^{-j45^\circ} = \frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}$

$x = 2 \quad W^2 = e^{-j90} = -j$

$x = 3 \quad W^3 = e^{-j135} = -\frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}$

$x = 4 \quad W^4 = -W^0$

$x = 5 \quad W^5 = -W^1$

$x = 6 \quad W^6 = -W^2$

$x = 7 \quad W^7 = -W^3$

$x = 8 \quad W^8 = W^0$

$x = 9 \quad W^9 = W^1$ etc.,

i.e., simply subtract multiples of N=8 as often as possible.

So rewriting ——————— and grouping with care

$F(0) = W^0(f_0+f_1+f_2+f_3+f_4+f_5+f_6+f_7) = \underline{W^0(f_0+f_4)}+W^0(f_1+f_5)+W^0(f_2+f_6)+W^0(f_3+f_7)$

$F(1) = W^0f_0+W^1f_1+W^2f_2+W^3f_3-W^0f_4-W^1f_5-W^2f_6-W^3f_7 = W^0(f_0-f_4)+W^1(f_1-f_5)+W^2(f_2-f_6)+W^3(f_3-f_7)$

$F(2) = W^0f_0+W^2f_1-W^0f_2-W^2f_3+W^0f_4+W^2f_5-W^0f_6-W^2f_7 = \underline{W^0(f_0+f_4)}+W^2(f_1+f_5)-W^0(f_2+f_6)-W^2(f_3+f_7)$

$F(3) = W^1f_0+W^3f_1-W^2f_2+W^1f_3-W^0f_4-W^3f_5+W^2f_6-W^1f_7 = W^0(f_0-f_4)+W^3(f_1-f_5)-W^2(f_0-f_6)+W^1(f_3+f_7)$

$F(4) = W^0f_0-W^0f_1+W^0f_2-W^0f_3+W^0f_4-W^0f_5+W^0f_6-W^0f_7 = \underline{W^0(f_0+f_4)}-W^0(f_1+f_5)+W^0(f_2+f_6)-W^0(f_3+f_7)$

$F(5) = W^0f_0-W^1f_1+W^2f_2-W^3f_3-W^0f_4+W^1f_5-W^2f_6+W^3f_7 = W^0(f_0-f_4)-W^1(f_1-f_5)+W^2(f_2-f_6)-W^3(f_3-f_7)$

$F(6) = W^0f_0-W^2f_1-W^0f_2+W^2f_3+W^0f_4-W^2f_5+W^0f_6+W^2f_7 = \underline{W^0(f_0+f_4)}-W^2(f_1+f_5)-W^0(f_2+f_6)+W^2(f_3+f_7)$

$F(7) = W^0f_0-W^3f_1-W^2f_2-W^1f_3-W^0f_4+W^3f_5+W^2f_6+W^1f_7 = W^0(f_0-f_4)-W^3(f_1-f_5)-W^2(f_2-f_6)-W^1(f_3-f_7)$

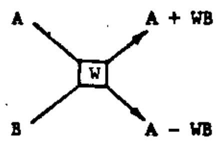
$F(8) = W^0f_0+W^0f_1+W^0f_2+W^0f_3+W^0f_4+W^0f_5+W^0f_6+W^0f_7 = F(0)$

So the four basic weights W^0, W^1, W^2 and W^3 are used repeatedly and since $W^0 = 1$(always) only 3 of the 4 weights are complex.

The RHS of this diagram clearly illustrates the repetitions of the same factor (four such factors are underlined).

The final form on the right hand side here can be written in terms of what are called "Butterfly" operations.

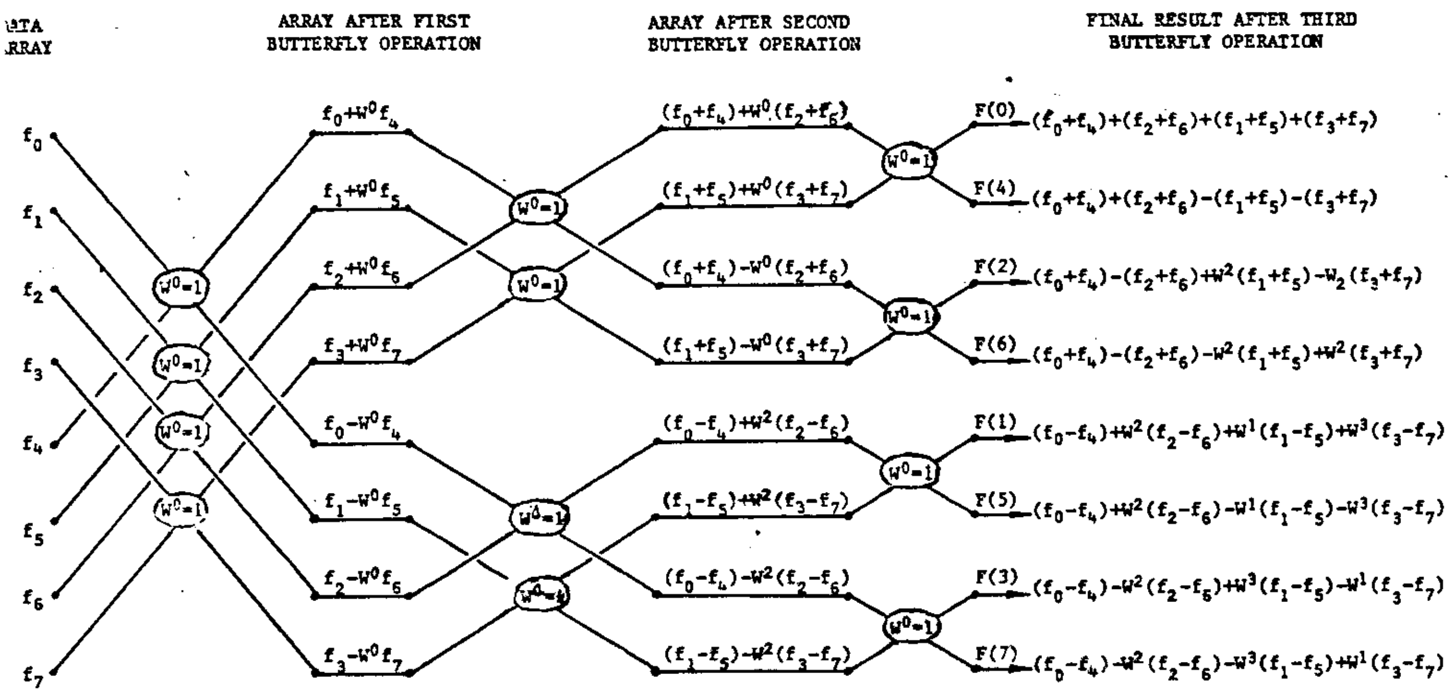
A single butterfly implies the following operation:



The input numbers A and B which may be complex and the operation by the weight W which in general is also complex gives the two outputs

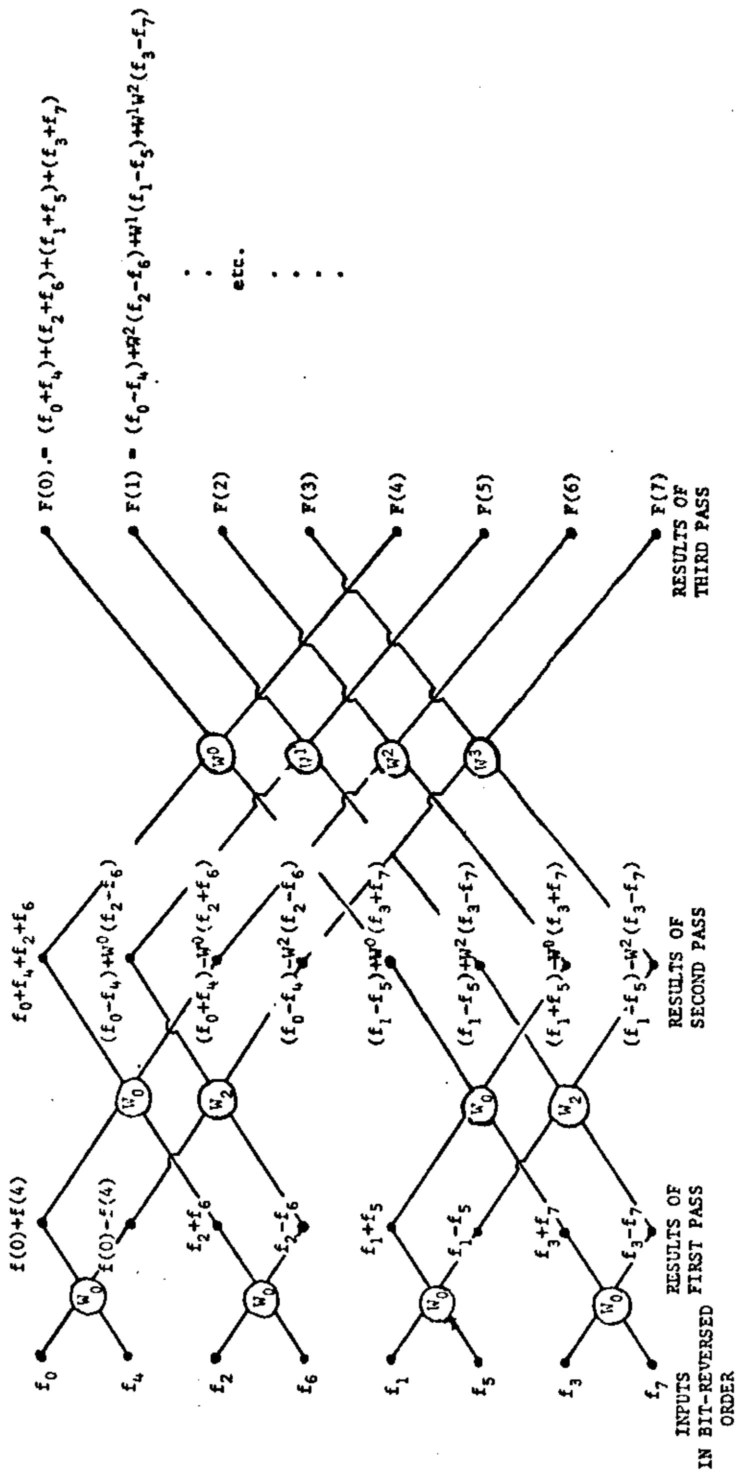
A + WB and A - WB

The above equations can be described using these butterflies as:



This entirely avoids duplicate operations and is a neat operation but its only disadvantage is that the output data array is shuffled and has to be reorganised before it can be used. It turns out that this shuffled data is in "bit-reversed" order i.e., for an eight sample data the 3rd sample in binary would be 011 - bit reversing this becomes 110 i.e., position 6.

It turns out that if we arrange the <u>input</u> samples (f_0 - f_7) in bit-reversed order that the output result emerges in natural order i.e.,



The 'weights' are simply exponentials so that for example $W^1W^2 = W^3$ and $W^0 = 1$

Three passes (or recombinations) of the data are necessary and it can be seen that only a single storage array is needed - i.e., the calculations can be performed 'in place'. The number of passes required is simply the power of 2 which gives the total number of samples, i.e., for 8 samples a total of 3 passes are required.

During the first (or 0th) pass the Butterfly spans adjacent samples ($2^0=1$). During the second (or 1th) pass the Butterfly spans samples by 2 ($2^1=2$). i.e., the 'span' is given by $2^{(\text{number of the pass-counting from zero})}$.

Now to arrange the weights which are used in each butterfly we simply have to remember that at each pass we are effectively combining different group lengths of data. e.g., in the 1st pass N=2 and the single weight is given by $e^{\frac{-j2\pi}{2}\cdot k}$ - where k only takes the value zero. At the second pass we are combining groups of data of length 4 so the two weights are given by $e^{\frac{-j2\pi}{4}\cdot k}$ - where k takes the values 0 and 1. So that at each successive pass we are effectively combining a further power of 2 sample length and the weights follow straight forwardly. Depending on the organisation of the particular computer being used considerable economies in computing effort may be achieved by first evaluating a "look up" table of Fourier weights (W^0, W^1, W^2 ..... etc.) which may then be called as required during the FFT evaluation.

It is worth emphasising that the result of the FFT caluculation is absolutely identical to that of the DFT. It does not introduce any further aberrations.

Two Dimensional Transforms

An important modern application of Fourier Transforms is in the area of picture processing by digital means - by implication two dimensional fields of real data are then implied. Modern Topics such as feature enhancement, filtering etc., have increasingly important application and it is then necessary to carry out 2.D discrete transforms and also to be able to interpret the results.

The extension of the Fourier Transform (i.e., the DFT or FFT) to more than one dimension is entirely straightforward. Consider the two dimensional arrays of time sample values f(I,J) shown. Where I = 0 to M-1 and J = 0 to N-1.

M SAMPLES →

N SAMPLES ↓					
	f(0,0)	f(1,0)	f(2,0)		f(M-1,0)
	f(0,1)	f(1,1)	f(2,1)	f(3,1)	--------
	f(0,2)	f(1,2)	f(2,2)	f(3,2)	--------
	f(0,3)	f(1,3)	f(2,3)	f(3,3)	'
	f(0,4)	f(1,4)	f(2,4)	f(3,4)	'
	'	'	'	'	'
	'	'	'	'	'
	'	'	'	'	'
	'	'	'	'	'
	f(0,N-1)	'	'	'	f(M-1,N-1)

The 2 dimensional Fourier Transform is then quite straightforwardly given by:

$$\sum_{J=0}^{N-1} \left\{ \sum_{I=0}^{M-1} f(I,J)\, e^{\frac{-j2\pi IK}{M}} \right\} e^{\frac{-j2\pi JK}{M}}$$

Notice there is no limitation whatever on the values of M and N so that we are <u>NOT</u> restricted to consideration of a square array.

To give an example of the result of using the 2 dimensional transform: Consider a 2 dimensional set of real input samples which are similar to the previous one dimensional examples.

i.e., $\sin(\underline{x}\ 108I + 50^{\circ}) * \cos(\underline{y}\ 108J + 15^{\circ})$

then the 9 <u>square</u> set of real input data samples (normalised to the largest value) is:

DATA INPUT (real numbers)

-0.212	0.584	-0.137	-0.479	0.433	0.212	-0.584	0.137	0.479
-0.245	0.652	-0.158	-0.554	0.500	0.245	-0.652	0.158	0.554
0.383	-0.987	0.234	0.822	-0.742	-0.383	0.987	-0.234	-0.822
0.020	-0.054	0.013	0.046	-0.042	-0.020	0.054	-0.013	-0.046
-0.378	1.000	-0.243	-0.850	0.768	0.378	-1.000	0.243	0.850
0.212	-0.584	0.137	0.479	-0.433	-0.212	0.584	-0.137	-0.479
0.245	-0.652	0.158	0.554	-0.500	-0.245	0.652	-0.158	-0.554
-0.383	0.987	-0.234	-0.822	0.742	0.383	-0.987	0.234	0.822
-0.020	0.054	-0.013	-0.046	0.042	0.020	-0.054	0.013	0.046

To simplify the transform result in this case the time samples have been taken to be symmetrical about the origin in both dimensions (i.e., the central point of this data set is effectively (0,0) in "time"!)

The output result in this case is:

2D-DFT outputs magnitude and phase

0.04∠20	0.15∠170	0.07∠150	0.03∠-35	0.03∠132	0.03∠-60	0.07∠105	0.15∠80	0.04∠-122
0.20∠-125	1.00∠20	0.47∠0	0.24∠174	0.20∠-10	0.24∠140	0.47∠-40	1.00∠-65	0.20∠07
0.15∠-115	0.50∠37	0.27∠10	0.14∠-177	0.12∠-0	0.14∠150	0.27∠-30	0.50∠-50	0.15∠97
0.00∠70	0.32∠-137	0.15∠-157	0.07∠0	0.00∠170	0.07∠-17	0.15∠149	0.32∠129	0.00∠-70
0.07∠-100	0.20∠47	0.13∠27	0.07∠-107	0.00∠0	0.07∠107	0.13∠-27	0.20∠-47	0.07∠100
0.00∠70	0.32∠-129	0.15∠-149	0.07∠17	0.00∠-170	0.07∠-0	0.15∠157	0.32∠137	0.00∠-70
0.15∠-97	0.50∠50	0.27∠30	0.14∠-150	0.12∠0	0.14∠177	0.27∠-10	0.50∠-37	0.15∠115
0.20∠-07	1.00∠65	0.47∠40	0.24∠-140	0.20∠10	0.24∠-174	0.47∠-0	1.00∠-20	0.20∠125
0.04∠122	0.15∠-80	0.07∠-105	0.03∠60	0.03∠-132	0.03∠35	0.07∠-150	0.15∠-170	0.04∠-20

The central point is a 2 dimensional DC term for this "frequency" result and the interpretation in either the horizontal or vertical direction is just the same as in the one dimensional case. For sinusoidal inputs the phase information can be retrieved in exactly the same way as in the one dimensional case. Here for example careful analysis of the transform result suggests that the waveform was generated by $\sin(180I + 44^\circ) \times \cos(108J + 19^\circ)$.

There are 9 x 9 = 81 numbers being input to the transform so just as with the one dimensional case there should be 81 useful output numbers in the result. With the sine wave result above the useful number of outputs is masked by the natural symmetries which occur.

Inputting 9 square real random numbers through the transform as follows:

DATA INPUT (real numbers)

0.234	0.067	0.000	0.480	0.084	0.578	0.808	0.552	0.621
0.010	0.019	0.740	0.088	0.762	0.548	0.254	0.685	0.633
0.903	0.288	0.507	0.625	0.169	0.036	0.024	0.939	0.564
0.714	0.177	0.770	0.274	0.504	0.250	0.318	0.508	0.617
0.443	0.794	0.460	0.248	0.960	0.383	0.330	0.601	0.992
0.294	0.322	0.875	0.742	0.730	0.637	0.262	0.411	0.883
0.468	0.980	0.201	0.958	0.016	0.205	0.314	0.141	0.492
0.157	0.952	0.657	0.080	0.028	0.298	0.689	1.000	0.020
0.572	0.439	0.427	0.334	0.968	0.109	0.581	0.165	0.677

Gives the following result:

2D-DFT outputs magnitude and phase

0.11∠-113	0.04∠-67	0.09∠-68	0.09∠-102	0.03∠-128	0.01∠-7	0.07∠-83	0.05∠78	0.03∠-140
0.07∠118	0.07∠48	0.08∠88	0.03∠104	0.05∠37	0.09∠173	0.08∠-177	0.11∠-10	0.07∠28
0.14∠-43	0.03∠135	0.05∠58	0.08∠57	0.00∠61	0.04∠-48	0.00∠14	0.05∠-49	0.02∠42
0.00∠2	0.04∠102	0.10∠-67	0.03∠34	0.04∠-44	0.07∠92	0.06∠7	0.05∠-138	0.07∠-137
0.03∠-64	0.06∠20	0.03∠-93	0.11∠-153	1.00∠0	0.11∠153	0.03∠93	0.06∠-20	0.03∠64
0.07∠137	0.05∠138	0.06∠-7	0.07∠-92	0.04∠44	0.03∠-34	0.10∠67	0.04∠-102	0.00∠-2
0.02∠-42	0.05∠49	0.00∠-14	0.04∠48	0.00∠-61	0.08∠-57	0.05∠-58	0.03∠-135	0.14∠49
0.07∠-28	0.11∠10	0.08∠177	0.03∠-173	0.05∠-37	0.03∠-104	0.08∠-88	0.07∠-48	0.07∠-118
0.03∠140	0.05∠-78	0.07∠83	0.01∠7	0.03∠128	0.09∠102	0.09∠68	0.04∠67	0.11∠113

Clear symmetrics are again evident but careful consideration again reveals 81 useful outputs in terms of both magnitude and phase.

For the straightforward DFT the computing effort is $\sim(NM^2+MN^2)$ multiplications and intuitively one can picture the repeated use of exponential multipliers which allowed the development of the FFT in the one dimensional case. The application of 'FAST" techniques to the 2 dimensional case will again massively reduce the computing effort particularly when large arrays are involved.

The economies of the FFT are just as relevant to the 2D Discrete Fourier Transform case and it is a straightforward mattet of application of the FFT to successive rows storing the complex results "in place" and then applying the FFT successively in column order.

The problem that more generally arises with the 2 dimensional situation is the immense quantity of data which can be involved. For example to provide reasonable resolution with a digitised satellite picture could require millions of data points. The problem then resolves to storing the gross quantity of data onto disc and recalling the data into the computer in small lots.

Significant efficiencies of scale can be achieved in such a situation if when we recall individual rows to perform the "row FFT" we can arrange enough storage to cope with 2 of these rows of data in the computer at one time. Then by simply choosing the rows carefully we can perform the first Butterfly stage of the 'column' FFT at the same time as we are doing the 'row FFT". For example suppose we had 8 immensely long rows such that we can only fit there into the active computer area 2 rows at a time we could then arrange to choose rows in the usual bit reversed order i.e., first rows 0 and 4 performing the complete row FFT and also the first butterfly combination of the column FFT. We would then deal similarly with rows 2 and 6 and so on. Many similarly efficient schemes have been reported [Ref.]

Appendix 5. The Concept of Convolution

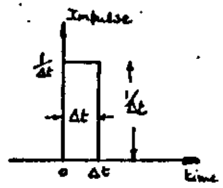
Convolution is widely used in the manipulation of Fourier Transforms and it is therefore essential that the process and the reasons for its use must be clearly understood.

In a Fourier analysis the assumption is generally made that systems are 'linear'. This means that the sampled data being used may have been measured at the output of a system whose characteristics can be represented by a linear differential equation. The most important consequence of this is that superposition then applies i.e., signals may be split into components which will be dealt with in the same way by the system.

The Unit Impulse Response

One of the most fundamental measurements which can be made on a system is to find its response to stimulus by an impulsive input.

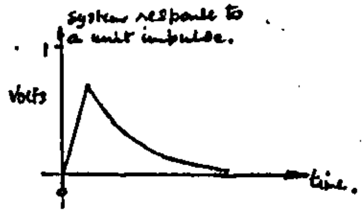
The ideal impulse is a spike of infinite height which lasts for an infinitesimally short time. i.e., it is obtained from the



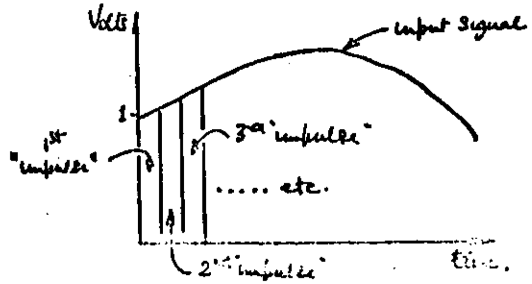
rectangular pulse shown when $\Delta t \to 0$. The 'value' of the impulse is defined as its area - in this case $\Delta t \times \frac{1}{\Delta t} = 1$ i.e. a unit impulse.

A good practical approximation to the ideal 'unit' impulse can be readily developed.

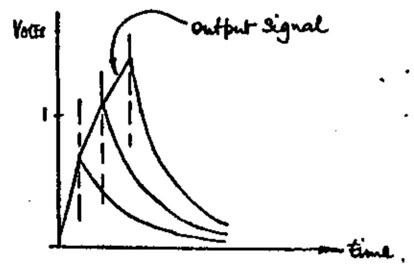
The result which is obtained when the unit impulse is applied is called the unit impulse response of the system and the measurement in fact completely characterises the system. Assuming that the impulse response is known i.e.,:



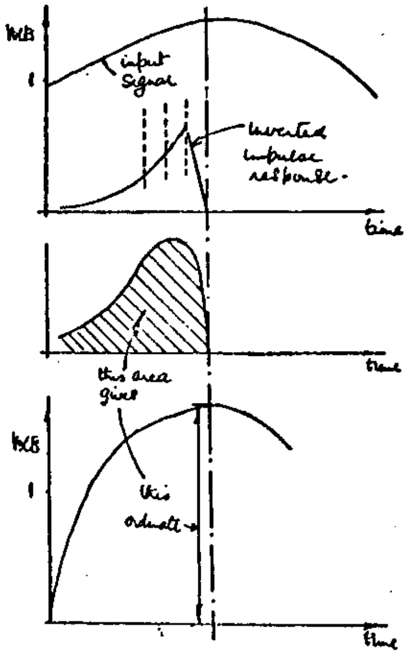
Then we can conceptually take any input signal and divide it up into a very large number of impulses i.e.,



Then the output at a time t_1 can be obtained by simply adding up all the constituent impulse responses:



This approach would appear to be a difficult and unwieldy task however there is an extremely simple way of arriving at the same result. At the desired output time t_1 erect an inverted unit impulse response as:



If the product of each vertical slice value of the two curves is then plotted as a product curve it can be reasoned that the area under this curve then gives the same result as adding up all the individual impulse responses. This area is finally plotted as an output ordinate at the t_1.

Sliding this inverted impulse response along the time axis simply

allows the output to be obtained. This process of inverting one curve and successively multiplying it by another is known as convolution. It must always result in a smoother answer than either of the component curves.

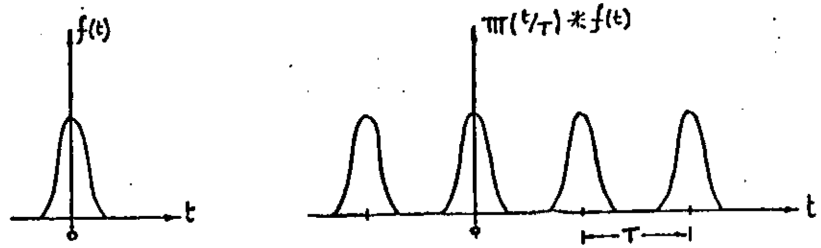
If the input curve is expressed as f(t) and the impulse response as h(t) then the convolution is defined by the integral [Ref.].

$$\text{output}\Big|_{t_2} = \int_{-\infty}^{t_1} f(t)\, h(t-\tau)\, d\tau$$

Example: The convolution of the sampling function and a finite duration is given by:

$$Ш\left(\frac{t}{T}\right) * f(t) = \sum_{n=-\infty}^{\infty} f(t-nT)$$

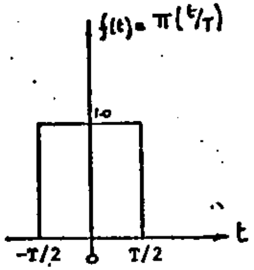
i.e.,



Clearly from the right hand side of this diagram overlapping will occur if the sampling interval T is shorter than the duration of f(t).

Appendix 3. The Gate Function

The last special function which is needed before considering sampling is the gate function which is given the special symbol $\pi(\frac{t}{T})$.



then

$$\pi(\tfrac{t}{T}) = \{\mu_1(t + T/2) - \mu_1(t - T/2)\}$$

where $\mu_1(x)$ is the unit positive step function commencing at $x = 0$.

$$FT\{\pi(t/T)\} = \int_{-\infty}^{\infty} \big(\mu_1(t + T/2) - \mu_1(t - T/2)\big)\, e^{-j2\pi ft}\, dt$$

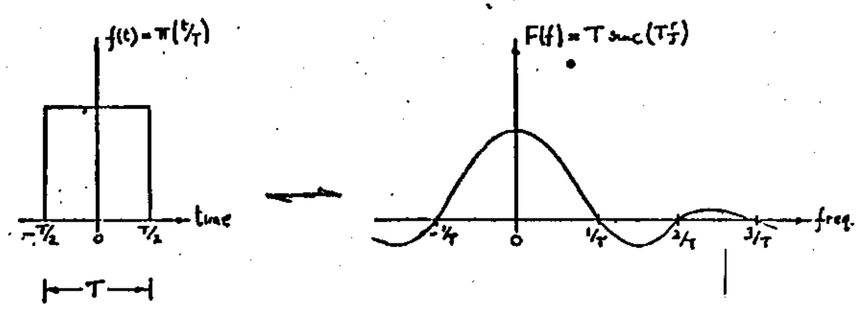
$$= \int_{-T/2}^{T/2} 1\cdot e^{-j2\pi ft}\, dt$$

$$= T \cdot \frac{\sin(2\pi f \cdot T/2)}{(2\pi f \cdot T/2)}$$

$$= T\,\frac{\sin(\pi fT)}{\pi fT}$$

$$= T\ \mathrm{sinc}(Tf)$$

Then we have the transform pair in pictorial form:



or alternatively