NTNU

Det skapende universitet

Department of Electronics and Telecommunications

Examination paper for TTT4120 DIGITAL SIGNAL PROCESSING

Contact during examination: Magne H. Johnsen

Tel.: 930 25 534

Examination date: Friday, December 9th, 2015

Examination time: kl. 09.00 - 13.00

Permitted examination support material:
D– No calculators allowed.
No printed or handwritten materials allowed.

Other information:

- The examination includes 4 problems. The weight of each subproblem is given in parenthesis at problem start. The total possible point score is 83.
- Some basic formulas are given in the appendix
- The course responsible will visit you twice, the first time around 10.00 o'clock and the second time around 12.00.

Language: English

Number of pages: 5

Number of pages enclosed: 3

Checked by:

Dato Sign

Problem 1 : (3+5+4+4+6)

Given a stable LTI-system $H(z)$ in the form $H(z) = H_1(z)H_2(z)$ where

$$H_1(z) = \frac{1+z^{-1}}{1-\frac{1}{3}z^{-1}} \quad (1)$$

$$H_2(z) = \frac{1}{1+\frac{1}{3}z^{-1}} \quad (2)$$

1a) Are the systems $H_1(z)$ og $H_2(z)$ causal and/or do they have minimum phase? State the reason for your answers!

Answer :
The pole values are $\frac{1}{3}$ and $-\frac{1}{3}$, i.e inside the unit, thus both systems are causal. A minimum phase system is invertible, i.e all poles and zeros must be on the **inside** of the unit circle. The zero at $z = -1$ is however on the unit circle, thus $H_1(z)$ does not have minimum phase. $H_2(z)$ does not have zeros (except for $z = 0$) and has therefore minimum phase.

1b) Show that $H(z)$ can be written in the following DF2 and parallel form

$$H(z) = \frac{1+z^{-1}}{1-\frac{1}{9}z^{-2}} \quad (3)$$

$$H(z) = H_3(z) + H_4(z) = \frac{2}{1-\frac{1}{3}z^{-1}} - \frac{1}{1+\frac{1}{3}z^{-1}} \quad (4)$$

Answer :
The DF2 form is easily derived by multiplying the denominators in $H_1(z)$ og $H_2(z)$ This gives eq 3.

The parallel form is derived by residual calculation :

$$H(z) = \frac{A}{1-\frac{1}{3}z^{-1}} + \frac{B}{1+\frac{1}{3}z^{-1}} = \frac{(A+B)+\frac{1}{3}(A-B)z^{-1}}{1-\frac{1}{9}z^{-2}}$$

Comparing constants

$$A + B = 1$$

$$\frac{1}{3}(A-B) = 1$$

easily lead to $A = 2$ and $B = -1$

1c) Show that the unit pulse responses of $H_1(z)$ and $H(z)$ fullfil the following

$$h_1(n) = \begin{cases} 1 & n = 0 \\ 4(\frac{1}{3})^n & n > 0 \end{cases} \tag{5}$$

$$h(n) = 2(\frac{1}{3})^n - (-\frac{1}{3})^n \quad n \geq 0 \tag{6}$$

Answer :
We first define $G_1(z) = \frac{1}{1-\frac{1}{3}z^{-1}} \Leftrightarrow g_1(n) = (\frac{1}{3})^n u(n)$. Further rewriting $H_1(z) = \frac{1}{1-\frac{1}{3}z^{-1}} + \frac{z^{-1}}{1-\frac{1}{3}z^{-1}} = G_1(z) + z^{-1}G_1(z)$ we have $h_1(n) = g_1(n) + g_1(n-1) = (\frac{1}{3})^n u(n) + (\frac{1}{3})^{n-1}u(n-1) = (\frac{1}{3})^n u(n) + 3(\frac{1}{3})^n u(n-1)$

For $n = 0$ the second term is zero thus :

$$h_1(n) = \begin{cases} 1 & n = 0 \\ 4(\frac{1}{3})^n & n > 0 \end{cases}$$

The unit pulse response in eq 6 is directly given by the parallel form in eq 4.

1d) Show that the unit pulse response $h(n)$ of the total system fullfils the following

$$\sum_{n=0}^{\infty} |h(n)| = \frac{9}{4} \tag{7}$$

$$r_{hh}(0) = \frac{81}{40} \tag{8}$$

Answer :
We have to inspect $h(n)$ (eq 6) both for even $n = 2m$ and odd $n = 2m + 1$ indexes, i.e $\sum_n |h(n)| = \sum_{neven} |h(n]| + \sum_{nodd} |h(n)| = \sum_{m=0}^{\infty} g_e(m) + \sum_{m=0}^{\infty} g_o(m)$ where

$$\begin{aligned} g_e(m) = h(2m) &= 2(\frac{1}{3})^{2m} - (-\frac{1}{3})^{2m} = 2(\frac{1}{9})^m - (\frac{1}{9})^m = (\frac{1}{9})^m \quad m = 0, \ldots, \infty \\ g_o(m = h(2m+1) &= 2(\frac{1}{3})^{2m+1} - (-\frac{1}{3})^{2m+1} = \frac{2}{3}(\frac{1}{9})^m + \frac{1}{3}(\frac{1}{9})^m = (\frac{1}{9})^m \quad m = 0, \ldots, \infty \end{aligned}$$

Thus we have

$$\sum_{n=0}^{\infty} |h(n)| = \sum_{m=0}^{\infty} (\frac{1}{9})^m + \sum_{m=0}^{\infty} (\frac{1}{9})^m = 2\sum_{m=0}^{\infty} (\frac{1}{9})^m = 2\frac{1}{1-\frac{1}{9}} = \frac{9}{4}$$

For the energy of the system we get

$$r_{hh}(0) = \sum_{n=0}^{\infty} h^2(n) = \sum_{m=0}^{\infty} g_e^2(m) + \sum_{m=0}^{\infty} g_o^2(m) = 2\sum_{m=0}^{\infty} ((\frac{1}{9})^m)^2 = 2\frac{1}{1-\frac{1}{81}} = \frac{81}{40}$$

Alternatively $r_{hh}(0)$ can be found by squaring and summing eq. 6.

1e) Sketch the filter structurs of the DF2 form (eq 3) and the parallel form (eq 4). In the latter structure the gain 2 (in the first term of eq 4) shall be placed in front of the corresponding feedback loop.

How many different structures can you choose among if you instead want to use a *cascade structure* of first order sections?

Answer :

Note that the figures include noise sources due to quantization! This is not default required from the students!

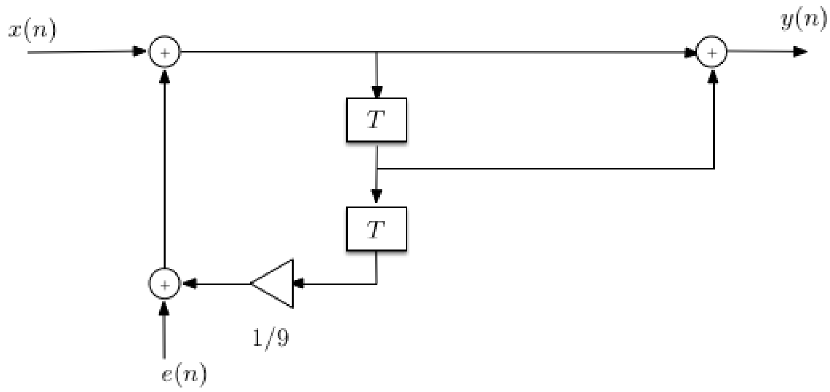

Figure 1: DF2-structure

If we define $G_3(z) = 1 + z^{-1}$ and use the definitions in 1c for $G_1(z)$ we have $H(z) = G_1(z)H_2(z)G_3(z)$. This gives us a total of six combinations for cascading the three sub-filters. In addition we can use $H_1(z) = G_3(z)G_1(z)$ in a DF2 form and we then have two new options in $H(z) = H_1(z)H_2(z) = H_2(z)H_1(z)$. Alternatively we can merge $G_3(z)H_2(z)$ to a DF2-form and cascade with $G_1(z)$ to get two extra alternatives. Thus we have totally ten possible alternatives for a cascade structure!

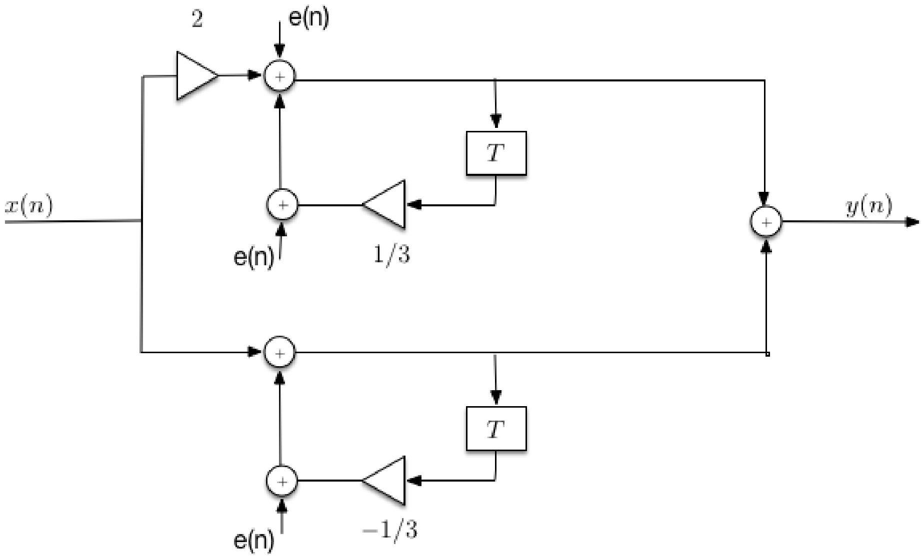

Figure 2: Parallel structure with gain in front of loop

Problem 2 : (4+5+4+5+5)

The filter $H(z)$ given by eq 3 and 4 in problem 1 shall be implemented in fixed point representation using $B+1$ bits and dynamic range $[-1, 1)$. Rounding (quantization) is performed after each multiplication and the rounding error, $e(n)$, can be regarded as white noise with power σ_e^2.

The filter input $x(n)$ has a uniform amplitude density; i.e. $x_{max} = \max_n |x(n)| = 1$.

2a) Find the resulting noise power at the output as a function of σ_e^2 for the DF2 structure.

Answer :
The DF2-structure is shown in task 1e. There is only one multiplication and the unit pulse response from there to the output is given by $h(n)$. Thus the noise at the output is given by $\sigma_y^2 = \sigma_e^2 r_{hh}(0) = \frac{81}{40}\sigma_e^2$

2b) Find the necessary scaling factor at the input of the DF2 structure in order to avoid overflow.

Find the reduction in signal-to-noise ratio (S/N) at the output due to scaling.

Answer :
We have two summation nodes. One is at the output, i.e. the unit pulse response from the input is $h(n)$. From task 1d we have that $\sum_n |h(n]| = \frac{9}{4}$.

The other node is at the input, i.e the transfer function from the input is

$$Q(z) = \frac{1}{1 - \frac{1}{9}z^{-2}} = \frac{1}{(1 - \frac{1}{3}z^{-1})(1 + \frac{1}{3}z^{-1})} = \frac{1}{2}[\frac{1}{1 - \frac{1}{3}z^{-1}} + \frac{1}{1 + \frac{1}{3}z^{-1}}]$$

where the last parallel form is easily found by doing the same exercise as in task 1b. Thus the corresponding unit pulse response is

$$q(n) = \frac{1}{2}[(\frac{1}{3})^n + (-\frac{1}{3})^n]$$

Thus $q(n) = (\frac{1}{3})^n$ for n even and $q(n) = 0$ for n odd. Using $n = 2m$ this results in

$$\sum_n |q(n]| = \sum_m (\frac{1}{9})^m = \frac{9}{8}$$

Thus the output node decides the scaling factor to be $\frac{9}{4}$.
The signal power, and thus the S/N is reduced by the square of the scaling factor, i.e by a factor of $\frac{81}{16} \approx 5$.
Note that the intro of a scaling factor is a new multiplication. Thus, to be absolutely correct, one should add a corresponding rounding noise source after the scaling. However, most previous exam solution have ignored this. Thus both kind of answers should be accepted as correct!

In this solution (2d and 2e) this last noise source is omitted.

2c) Find the resulting noise power at the output as a function of σ_e^2 for the parallel structure.

Answer :
We now have three multiplications, one in each loop and the gain of 2 in the upper branch. As $(-\frac{1}{3})^{2n} = (\frac{1}{3})^{2n} = (\frac{1}{9})^n$ all three noise sources will contribute by the same amount of power at the output, i.e.

$$\sigma_y^2 = 3\sigma_e^2 \sum_n ((\frac{1}{3})^n)^2 = 3\sigma_e^2 \sum_n (\frac{1}{9})^n = 3\sigma_e^2 \frac{9}{8} = \frac{27}{8}\sigma_e^2$$

2d) Find the necessary scaling factor at the input of the parallel structure in order to avoid overflow.

Find the reduction in signal-to-noise ratio (S/N) at the output due to scaling.

Answer :
We now have three summation nodes. One is at the output and we have calculated the corresponding scale factor 9/4 in task 2b. The other two are in the feedback loops. The unit pulse responses from the input are given by the two terms in eq 6. Given that the gain is in front of the corresponding loop we have $\sum_n 2(\frac{1}{3})^n = 2\frac{1}{1-\frac{1}{3}} = 3$ for that node and $\sum_n (\frac{1}{3})^n = \frac{3}{2}$ for the other loop. Thus the largest of the three values is 3.

The reduction of the S/N corresponding to the downscaling by 3 is $3^2 = 9$.

2e) An option for the parallel structure is to move the gain 2 to *after* the feedback loop.

Find the necessary scaling factor at the input of the corresponding parallel structure.

Find the reduction in signal-to-noise ratio (S/N) at the output due to scaling.

Which of the three scaled structures are best with respect to signal-to-noise ratio (S/N) at the output?

Answer :

We have also now three summation nodes. One is at the output and we have calculated the corresponding scale factor 9/4 in task 2b. The other two are in the feedback loops. The unit pulse responses from the input are given by respectively $(-1/3)^n$ and $(1/3)^n$ These two give the same scaling need, i.e 3/2. Thus the largest of the three values is 9/4.

The reduction of the S/N corresponding to the downscaling by is $(9/4)^2 = 81/16 \approx 5$ (as for the DF2 structure).

We now have a new situation wrt rounding noise. The noise in the upper branch of Figure 2 (task 1e) is now multiplied by 2, i.e the power by 4, while the noise from the gain multiplication is at the output. Thus we have :

$$\sigma_y^2 = \sigma_e^2 \sum_n ((2\frac{1}{3})^n)^2 + \sigma_e^2 \sum_n ((-\frac{1}{3})^n)^2 + \sigma_e^2 = 5\sigma_e^2 \frac{9}{8} + \sigma_e^2 = \frac{53}{8}\sigma_e^2$$

Let us denote the signal power at the output (before scaling) by P_y The corresponding output power values after downscaling of the input are respectively $16P_y/81$, $P_y/9$ and $16P_y/81$ for the DF2 and the two parallel structures. Thus the S/N at the output after scaling is

$$\begin{aligned}
S/N_{DF2} &= \frac{16P_y/81}{81\sigma_e^2/40} = \frac{P_y}{\sigma_e^2}\frac{16*40}{81^2} \approx 0.0975\frac{P_y}{\sigma_e^2} \\
S/N_{2par} &= \frac{P_y/9}{27\sigma_e^2/8} = \frac{P_y}{\sigma_e^2}\frac{8}{27*9} = \frac{P_y}{\sigma_e^2}\frac{8*3*9}{81^2} \approx 0.0329\frac{P_y}{\sigma_e^2} \\
S/N_{par2} &= \frac{16P_y/81}{53\sigma_e^2/8} = \frac{P_y}{\sigma_e^2}\frac{16*8}{81*53} \approx 0.0298\frac{P_y}{\sigma_e^2}
\end{aligned}$$

Thus the scaled DF2 structure gives approximately three times higher S/N than the two parallel structures. Further it is not a good idea to move the branch gain of 2 to after the loop.

Problem 3 : (3+6+6+5)

3a) Define respectively an ARMA[P,M], AR[P] and MA[M] process.

Answer :
An ARMA[P,M] process is the output when white noise with power σ_w^2 is filtered through an IIR-filter with transfer function $H(z) = \frac{B_M(z)}{A_P(z)}$ where the subindexes give the order of the nominator (number of zeros) and denominator (number of poles).
In the AR[P]-process the nominator polynom is $B_M(z) = 1$. This correponds to $M = 0$, i.e no zeros (a so called allpole-filter).
In the MA[M]-process the denominator polynom is $A_P(z) = 1$. This correponds to $P = 0$, i.e no poles (FIR-filter).
The AR[P]-model is often preferred as the best compromize between modeling capacity and mathematical complexity.

3b) Given a physical signal $x_a(t)$. We measure N samples $x(n) = x_a(nT),\ n = 0, N-1$ from the signal.

Explain which approximations we have to use in order to be able to estimate the frequency spectrum of an autoregressive process.

Answer :
We need to approximate the signal $x(n)$ by a statistical model such that we can estimate the frequency content from the measured samples. The approximations needed for this are :

1. Model the signal as a nonstationary process
2. Approximate the above as a short-time stationary/ergodic process. First we must define the terms ergodic and short-time
3. In a stationary process the statistical properties (acf and the frequency specter) do not change with time
4. The statistical properties of an ergodic process can be derived from a single sequence $x(n)$
5. Estimates of the statistical properties of an ergodic process can be derived from a finite number of samples $x(n)$ $n = 0, \ldots, N-1$
6. A short-time ergodic process will have (approximately) constant properties over a number N of samples. Thus we can estimate new (short-time) properties which change for every new set of N samples we measure.
7. For many practical applications/signals the short-time ergodic estimation technique gives satisfactory results

8. Parametric models like an AR[P] model typically gives better spectrum estimates than the classical nonparametric periodogram.

3c) Assume that we have estimated the first four autocorrelation values of $x(n)$ to be

$\hat{\gamma}_{xx}(0) = 1.25$, $\hat{\gamma}_{xx}(1) = 0.75$, $\hat{\gamma}_{xx}(2) = 0.25$ og $\hat{\gamma}_{xx}(3) = -0.10$.

Finn the best AR[1]- and the best AR[2]-model for $x(n)$.

Answer :

For the AR[1]-modell we have from the Normal equations (see appendix) given $m = 1 = P$:
$a_1\gamma_{xx}(0) = -\gamma_{xx}(1) \quad \Rightarrow \quad a_1 = -\gamma_{xx}(1)/\gamma_{xx}(0)$.
Inserting for the acf-estimates we get : $\hat{a_1} = -0.75/1.25 = -3/5 = -0.6$
Using the Y-W equations for $m = 0$ we get :
$\sigma_f^2 = \gamma_{xx}(0) + a_1\gamma_{xx}(1)$
Inserting the estimates we get $\hat{\sigma_f^2} = 1.25 - 0.6 * 0.75 = 0.8$

For the AR[2]-modell we have from the Normal equations (see appendix) given $m = 1$ and $m = 2 = P$:

$$\begin{aligned} a_1\gamma_{xx}(0) + a_2\gamma_{xx}(1) &= -\gamma_{xx}(1) \\ a_1\gamma_{xx}(1) + a_2\gamma_{xx}(0) &= -\gamma_{xx}(2) \end{aligned}$$

Inserting the acf-estimates and doing the standard calculation of two unknowns in two equations we get :
$\hat{a_1} = -9/16 = -3/4$ and $\hat{a_2} = +1/4$
Using the Y-W equations for $m = 0$ we get :
$\sigma_f^2 = \gamma_{xx}(0) + a_1\gamma_{xx}(1) + a_2\gamma_{xx}(2)$
Inserting the estimates we get $\hat{\sigma_f^2} = 1.25 - 0.75 * 3/4 + 0.25 * 1/4 = 0.75$

3d) Explain the principle for design of a Wiener filter. If possible use a sketch.

Answer :

A physical (ergodic) signal $s(n)$ with known properties is contaminated by white

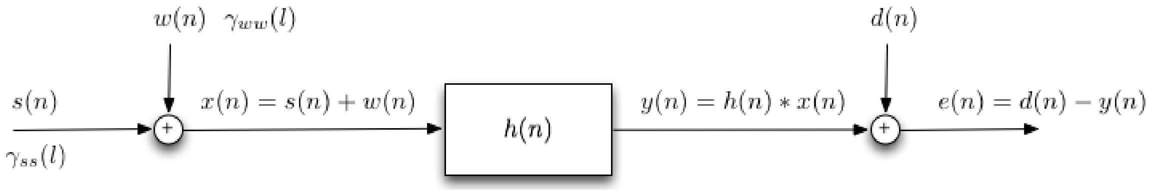

Figure 3: Wiener-filter system

noise $w(n)$ with known power. We observe the sum $x(n) = s(n) + w(n)$ and want to use a filter to convert $x(n)$ into a signal $y(n)$ which is "as similar as possible" to another ergodic signal $d(n)$ with known properties. We choose to use mean square error (MSE) $E[e^2(n)]$ where $e(n) = d(n) - y(n)$ as a measure for similarity. Closed form solutions for minimum MSE can be found both for the noncausal IIR case and for the causal IIR and FIR cases. The most used application is $d(n) = s(n)$, i.e. noise reduction.

Problem 4 (3+6+5+4)

4a) Set up the formula for a N-point Diskret Fourier Transform (DFT) for a sequence $x(n)$ of finite length M

Also set up the formula for the inverse DFT (IDFT). How must N be chosen if one wishes reproduce $x(n)$ from the DFT values?

Answer :

$$X(k) = \sum_{n=0}^{M-1} x(n) e^{-2j\pi nk/N} \quad k = 0, N-1$$

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{2j\pi nk/N} \quad n = 0, M-1$$

Formula for $x(n)$ is only correct if $N \geq M$ (to reproduce $x(n)$ from $X(k)$)

4b) One wants to filter an infinitely long sequence $x(n),\ n = -\infty, \infty$ by a FIR-filter $h(n)$ of length L.

Explain how the filtering can be performed in the frequency domain by using the so called "overlap-add" method.

Compare the"overlap-add" method to standard time domain filtrering with respect to the number of multiplications and additions per output sample

Answer :
One splits the input sequence into consecutive segments $x_i(n)$ of lengths M

$$x_i(n) = x(n + iM) \quad \text{for} \quad n = 0, \ldots, M-1 \text{ and } i = -\infty, \infty$$

This leads to :

$$y(n) = h(n) * x(n) = h(n) * \sum_i x_i(n) = \sum_i h(n) * x_i(n) = \sum_i y_i(n)$$

where the output segments $y_i(n)$ have lengths $M+L-1$. Thus two consecutive output segments overlap by L samples, but are easily summed to achieve $y(n)$.
The calculation of $y_i(n)$ can be done in the frequency domain as both the input segment and the filter have finite lengths. We choose $N = 2^R \geq M + L - 1$ (in order to use the FFT and be able to reproduce $y_i(n)$). Further given that $H(k)$ $k = 0, N-1$ is calculated only once, i.e precalculated, the algorithm is as follows : For each segment $i = -\infty, \infty$

- Calculate $X_i(k)$ $k = 0, N-1$ from $x_i(n)$
- Calculate $Y_i(k) = H(k)X_i(k)$ $k = 0, N-1$
- Calculate $y_i(n)$ $n = 0, M+L-1$ from $Y_i(k)$
- Calculate $y(n)$ from two consecutive segments $y_i(n)$

To produce N output values of $y(n)$ by time domain filtering we need M*N mults+adds. Using the frequency domain method and FFT we need $N*R+N+N*R = N*(2R+1)$ where $R = log_2(N)$. Thus for any N where $(2log_2(N)+1) < M$ the overlap-add technique with FFT should be used.

4c) One wants to use DFT to perform a frequency analysis of an infinitely long sequence $x(n),\ n = -\infty, \infty$. In real life one has to base the analysis of a finite segment of length K of the sequence.

Discuss the problems regarding frequency resolution and frequency "leakage" (sidelobes) as a function of the segment length K.

How can one manage to achieve a compromise with respect to the two non-idealities in the frequency domain?

Answer :
Just using a segment of length K is mathematically equivalent to using a rectangular window of length K. The two nonidealities can be explained by looking at the frequency content of a segment of a harmonic (sinus). Ideally a harmonic is a dirac pulse but using only a segment (i.e. a window) gives a sinc-like function; i.e. a bandwidth and sidelobes. The bandwidth is proportional to $C*(1/K)$ (where C is a constant) and obviously gives us the frequency resolution. The sidelobes give us the frequency leackage. For a rectangular window the constant C is relatively small, i.e the frequency resolution is high. However the sidelobes are relatively high and do only converge towards a finite level (-26 dB relatively to the main lobe) for $K \to \infty$.
A tapered window will decrease the frequency resolution (larger value of C) but also increase the sidelobe attenuation. Different tapering form will give minor differences in this compromize.

4d) The radix-2 Fast Fourier Transform (FFT) is a fast algorithm for calculating the DFT of a sequence when the length N is a power of 2, i.e. $N = 2^R$

Explain briefly the *principle* of the radix-2 FFT algorithm.

Answer :

The main principle for the FFT is that one can implement a $N = 2^R$ point DFT by using 2 $N/2$ point DFTs and N multiplications. And it is easily shown than the latter leads to fewer mults+adds (m+a). A general N point DFT use N^2 m+a. Thus $N^2 \geq 2 * (N/2)^2) + N = N^2/2 + N$ for all values $N > 2$! The difference gets large for typical values of N, i.e. $N = 64, 128, 256, 512, 1024, ...$

The radix-2 FFT successively splits the DFTs into smaller such that one ends up with a structure consisting of $(N/2) * log_2(N)$ 2-point DFTs (so called butterflies), which each requires maximum 2 m+a.