Chapter 8

Advanced FFT Topics and Applications

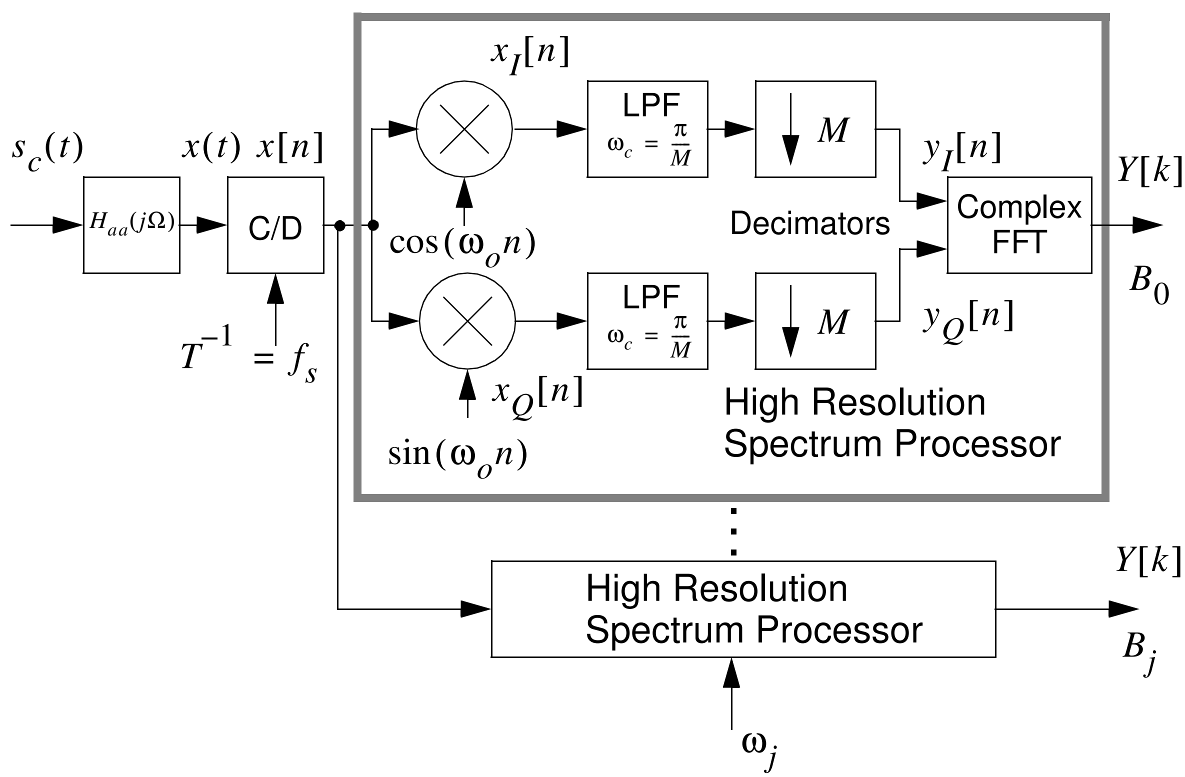
Contents

.

In this chapter we will begin by introducing a technique for improving frequency resolution in an FFT based spectral analysis system. The second topic will be averaged periodogram spectral estimation. The third topic will be a brief look at quantization effects in the computation of the DFT (FFT).

8.1 Zoom Spectral Analysis

To improve the frequency resolution in a spectral analysis system in general requires increasing the DFT length N. Given a fixed band of frequencies must be spectral analyzed, an efficient means of obtaining high resolution is to subdivide the band into smaller bands (most likely contiguous bands). High frequency resolution can be obtained in each subband without the DFT length becoming too large. The block diagram of a *zoom* spectral analysis system is shown below.



- The DFT analysis of a particular subband centered at Ω_o (f_o or ω_o) begins by frequency translating the subband to a spectral location centered about $\Omega = 0$ using a complex demodulator. The required complex demodulation may be performed in either continuous- or discrete-time.

- Suppose $x(t) \stackrel{\mathcal{F}}{\longleftrightarrow} X(j\Omega)$ and $x[n] \stackrel{\mathcal{DFT}}{\longleftrightarrow} X[k]$, then

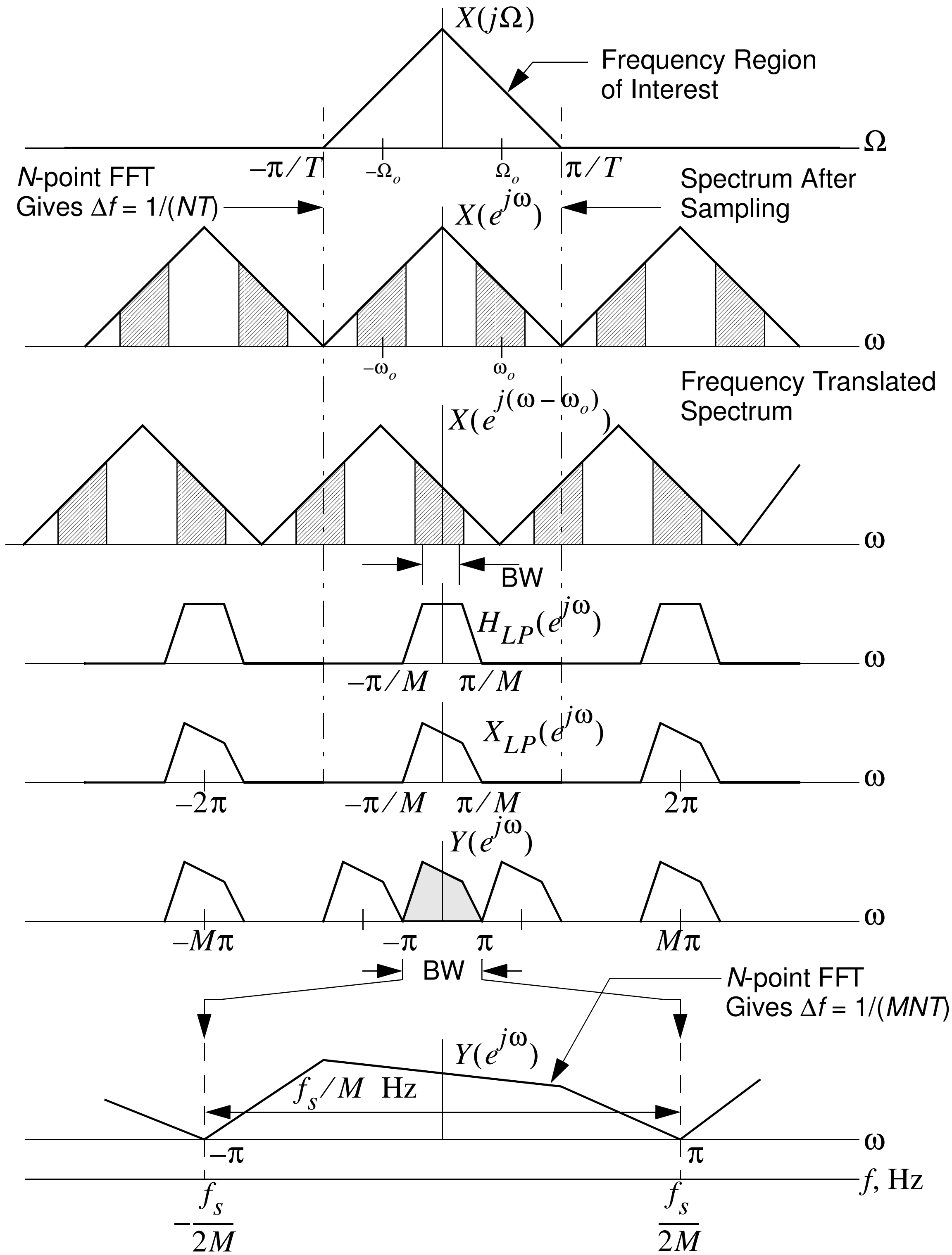
$$
\begin{aligned}
x(t)e^{-j\Omega_o t} &\stackrel{\mathcal{F}}{\longleftrightarrow} X(j\Omega + j\Omega_o) \\
x[n]e^{-j2\pi k_o/N} &\stackrel{\mathcal{DFT}}{\longleftrightarrow} X[k + k_o],\ \ 0 \leq k \leq N-1
\end{aligned}
$$

  Note that in either case the output of the complex demodulator is a complex signal (i.e. $x[n] = x_I[n] + jx_Q[n]$)

- Following complex demodulation the signal is lowpass filtered and then decimated (downsampled) by a factor M

- The lowpass filter (in practice a digital filter) prior to decimation is needed to prevent aliasing when the effective sampling rate is reduced by M. Assuming an ideal lowpass filter, the lowpass cutoff frequency, ω_c, must be less than π/M

- On the following two pages appropriate frequency domain waveforms within the zoom spectrum analyzer are given



Frequency spectra from the input through the decimator output

- With the sampling rate reduced by M, the frequency band analyzed by an N-point DFT has total width $1/(MT) = f_s/M$ Hz (assuming the initial sampling rate is f_s samples/sec)
- The minimum frequency spacing has been reduced from f_s/N to $f_s/(MN)$
- To analyze a different subband all we need to do is change the value of f_o (k_o)
- In the study of windows it was pointed out that to increase frequency resolution the window length must increase. For the zoom transform discussed above the signal capture- or collect-time, is identical to using an MN-point DFT.
- With the zoom transform fewer points are actually collected (only one out of M samples is saved) and hence less computation is required, since only a narrow band of frequencies is being analyzed. The signal is compressed and the frequency axis in effect expands.

Example 8.1: Designing for 10 Hz Resolution

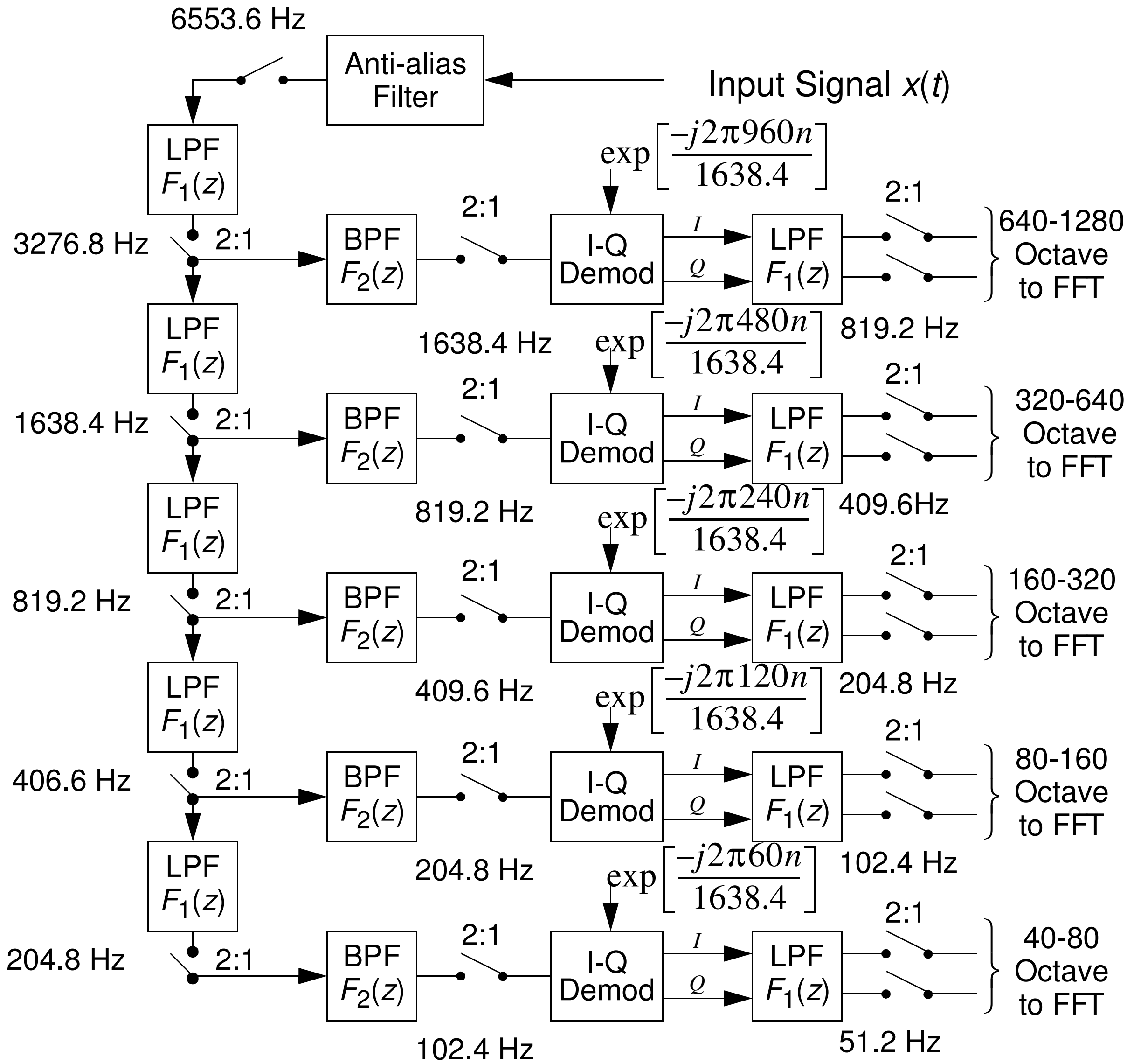
Given $f_s = 10$ MHz find N and M to spectrum analyze a 100 kHz subband with $\Delta f \leq 10$ Hz

$$\text{BW} = \frac{f_s}{M} \Rightarrow M = 100$$
$$\Delta f = \frac{f_s}{MN} \Rightarrow N \geq 10,000 \; (N = 2^{14} = 16,384)$$

Example 8.2: Using Octave Subbands [1]

Parameter	Requirement
A/D output frequency	6553.6 Hz obtained by dividing down by 300 a 1.96608 MHz crystal oscillator
FFT outputs	5 octaves between 40 and 1280 Hz
FFT resolution	400 filters per octave yielding resolutions of 1.6 Hz in the 640-1280 Hz octave, 0.8 Hz in the 320-640 Hz octave, and so on.
FFT size	512 points
Maximum passband ripple due to all analog and digital filtering operations	± 0.5 dB
Minimum rejection of sinusoids aliased into the DFT analysis band	-43 dB

[1]D.F. Elliot and K.R. Rao, *Fast Transforms Algorithms, Analysis, Applications*, Academic Press, Inc., 1982.



System block diagram for 40 – 1280 Hz coverage

8.2 Averaged Periodogram Spectral Estimation

For continuous-time signals the power spectrum of a wide-sense stationary random process $x_c(t)$ is defined as

$$P_{xx}(f) = \mathcal{F}\{R_{xx}(\tau)\}$$

where $R_{xx}(\tau)$ is the autocorrelation function of $x_c(t)$. For $x_c(t)$ in general a complex signal, the autocorrelation is given by

$$R_{xx}(\tau) = E\{x(t+\tau)x^*(t)\}$$

- The Fourier transform pair used here is

$$\begin{aligned} X(f) &= \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,dt \\ x(t) &= \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}\,df \end{aligned}$$

- An alternate definition for the power spectrum is

$$P_{xx}(f) = \lim_{T_L\to\infty} \frac{E\{|X_{T_L}(f)|^2\}}{T_L}$$

  where $X_{T_L}(f) = \mathcal{F}\{x_{T_L}(t)\}$ and

$$x_{T_L}(t) = \begin{cases} x(t), & |t| \le T_L/2 \\ 0, & \text{otherwise} \end{cases}$$

- An estimator for the power spectrum is the *periodogram*

$$I_{xx}(f) = \frac{1}{T_L}|X_{T_L}(f)|^2$$

  which has units of Watts/Hz

- The discrete-time equivalent to $X_{T_L}(f)$ is the L-point DFT (FFT) $V[k]$, $0 \leq k \leq L-1$ and the associated periodogram

$$\begin{aligned} I_{xx}[k] &= I_{xx}(\omega_k), \omega_k = \frac{2\pi k}{L} \\ &= \frac{T}{L}|V[k]|^2,\ 0 \leq k \leq L-1 \end{aligned}$$

  where the factor T/L has been included to give $I_{xx}[k]$ units of power spectral density (i.e. Watts/ Hz)

  Note: $I_{xx}[k]$ is defined in terms of $v[n] = x[n]w[n]$ where, as shown in notes chapter 7, $w[n]$ selects an L sample segment of $x[n]$

- For analysis purposes it may be easier to deal with the continuous frequency quantities $V(e^{j\omega})$ and $I_{xx}(\omega)$

- For a nonuniform window, $w[n]$, we define the *modified periodogram* as

$$I_{xx}[k] = \frac{T}{LU}|V[k]|^2$$

  where U is a bias normalization factor depending upon the window function

8.2.1 Periodogram Properties

Since $x[n]$ is a random process the periodogram is also a random process in ω or index k. The quality of the periodogram estimator can be determined from the mean and variance of $I_{xx}(\omega)$. Ideally we desire an estimator that is *unbiased* and *consistent*. An estimator is unbiased if the mean is the quantity we are trying to estimate.

An estimator is consistent if as the number of samples used in the estimate increases the variance goes to zero.

- The expected value of the periodogram can be shown to be

$$E\{I(\omega)\} = \frac{T}{2\pi LU} \int_{-\pi}^{\pi} P_{xx}(\theta) C_{ww}(e^{j(\omega-\theta)})\, d\theta$$

  where $P_{xx}(\omega)$ is the true power spectrum and $C_{ww}(e^{j\omega})$ is the magnitude squared of the Fourier transformed window function, $w[n]$ (i.e. $C_{ww}(e^{j\omega}) = |W(e^{j\omega})|^2$)

- As L increases $W(e^{j\omega})$ approaches a delta function, and $E\{I_{xx}(\omega)\}$ becomes $P_{xx}(\omega)$ provided U is properly chosen

- By setting the area under $\frac{1}{U}|W(e^{j\omega})|^2$ on $[-\pi, \pi)$ equal to one we can find the proper value for U

$$U = \frac{1}{L} \sum_{n=0}^{L-1} (w[n])^2$$

- The variance of $I_{xx}(\omega)$ is difficult to calculate in general. In Jenkins and Watts[2] it is shown that for a wide range of conditions the variance is on the order of $P_{xx}^2(\omega)$

- The conclusion is that although the periodogram is asymptotically unbiased, the estimator is inconsistent

[2]G.M. Jenkins and D.G. Watts, *Spectral Analysis and Its Applications*, Holden Day, San Francisco, 1968

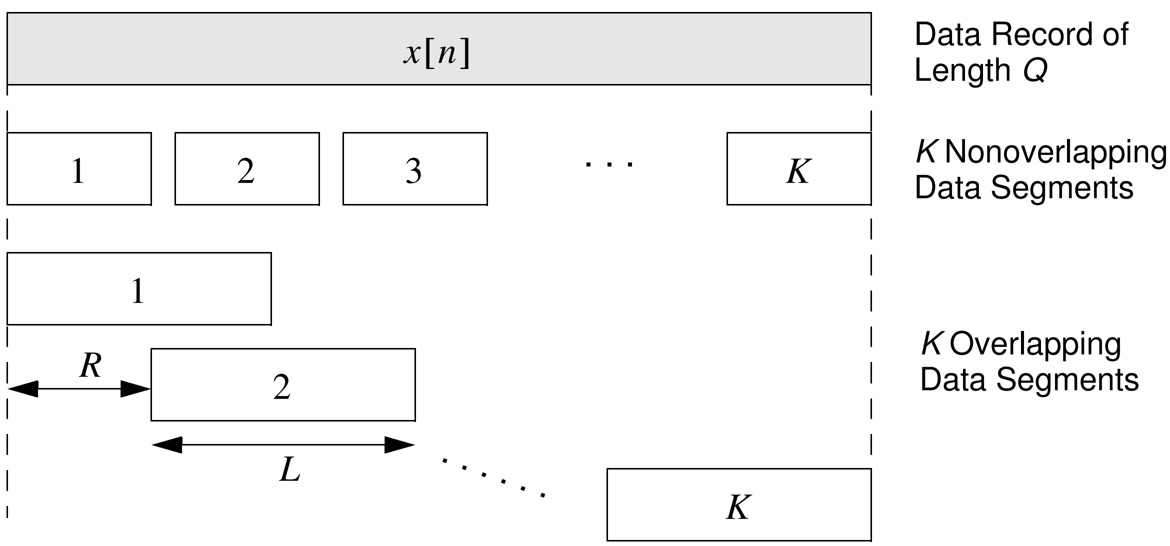
8.2.2 Averaged Periodograms

To reduce the variance of the periodogram spectral estimate we may average periodograms computed over K, possibly overlapping, data segments. Assume the available data sequence is $x[n]$, $0 \leq n \leq Q-1$.

- Divide $x[n]$ into K segments each of length L

$$x_r[n] = x[rR+n]w[n], \;\; 0 \leq n \leq L-1$$

  For $R < L$ the segments overlap, and if $R = L$ they are contiguous



Data record segmentation

- The number of segments K is the largest integer that gives $(K-1)R + (L-1) \leq Q-1$

- The rth segment periodogram is given by

$$I_{xx}^r(\omega) = \frac{T}{LU}|X_r(e^{j\omega})|^2$$

- The averaged periodogram is defined by

$$\bar{I}_{xx}(\omega) = \frac{1}{K}\sum_{r=0}^{K-1} I_{xx}^{r}(\omega)$$

- Since each periodogram is windowed with $w[n]$ the mean of the averaged periodogram is also biased

- If as an approximation we assume independence between periodogram samples, then the variance of the averaged periodogram is

$$\text{var}[\hat{I}_{xx}(\omega)] \simeq \frac{1}{K}P_{xx}^2(\omega)$$

  Thus the variance will approach zero as K increases

- In summary if both K and L are made large as $Q \to \infty$, then both the bias and variance can approach zero. In practice Q is finite so there is a tradeoff between bias (resolution) and variance

---

Example 8.3: **Two Sinusoids in Noise**

Estimate the power spectrum of the signal

$$x[n] = A\cos[2\pi(n/8)] + \frac{A}{2}\cos[2\pi(3n/8)] + w[n]$$

where $w[n]$ is a white Gaussian noise sequence with zero mean and variance σ_w^2. The sinusoid signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = \frac{A^2}{2\sigma_w^2}$$

Note that the SNR of the sinusoid with amplitude $A/2$ is 6 dB lower in power than the sinusoid with amplitude A.

- For a single sinusoid in white noise a wide-sense stationary random process model is

$$z[n] = A\cos(\omega_o n + \theta) + w[n]$$

where θ is an independent random variable uniform on $[0, 2\pi)$

- It can be shown that the corresponding power spectrum of $z[n]$ is

$$P_{zz}(\omega) = \frac{A^2\pi}{2}\big[\delta(\omega - \omega_o) + \delta(\omega + \omega_o)\big] + \sigma_w^2$$

  - The only difference here being a factor of 2π in just the spectral line components

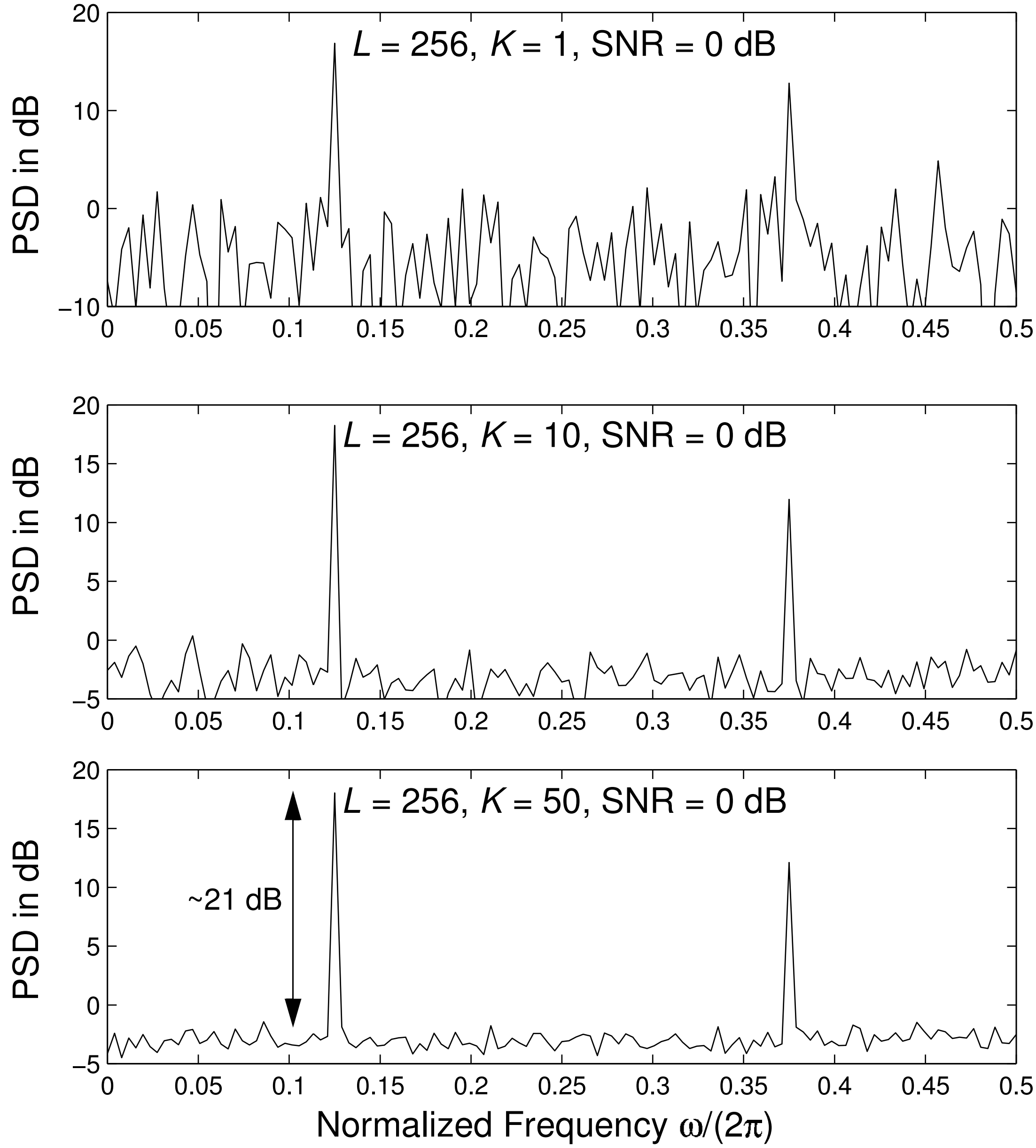
- Assuming we wish to display the above power spectrum using frequency f in Hz, and $T = 1$ s, we can write

$$P_{zz}(f) = \frac{A^2}{4}\big[\delta(f - f_o) + \delta(f + f_o)\big] + \sigma_w^2$$

  - The only difference here being a factor of 2π in just the spectral line components

- For the first set of example results let $A = 1$, SNR = 0 dB, and $L = 256$

- We then consider $K = 1$, 10, and 50



Averaged periodogram of two sinusoids in white noise at SNR = 0 dB

- As K increases the variance is clearly reduced
- The theoretical power spectrum should have a noise spectral density of

$$\sigma_w^2 = \frac{1^2}{2 \cdot 10^{0/10}} = \frac{1}{2} \text{ or -3.01 dB}$$

- The spectral line at $f = 1/8$ should in theory be a delta function with area equal to $1^2/4$ or -6.02 dB

- The spectral lines in the averaged periodogram estimate are both well above 0 dB! why?

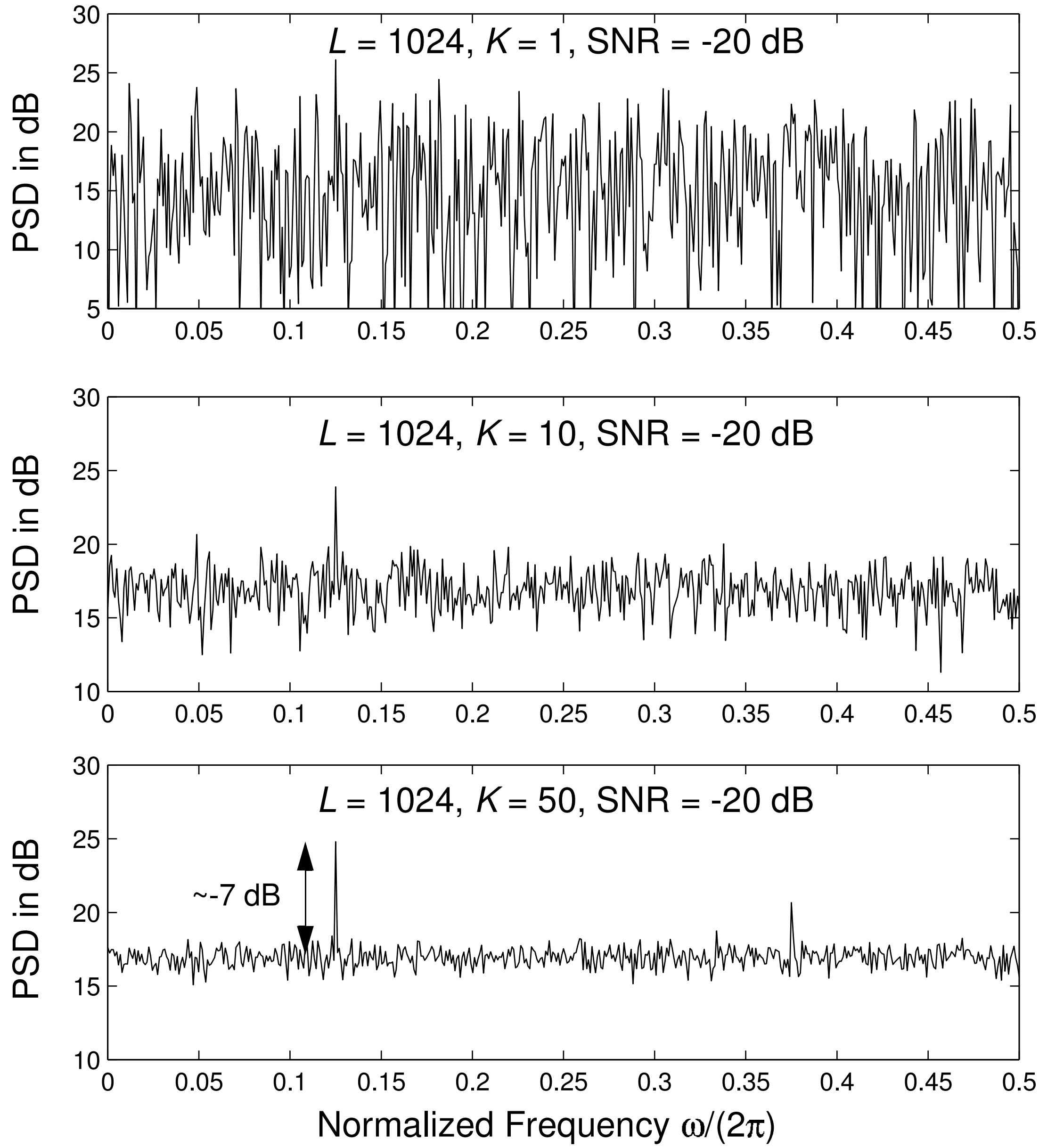
- As in a true analog spectrum analyzer we cannot calibrate the display to correctly show both a continuous spectral density and the power in a spectral line

- The height of the spectral peak at $f = 1/8$ is actually the signal power divided by the resolution bandwidth or FFT bin width, which in this case is $\Delta f = f_s/L = 1/256$ Hz

- The spectral peak in dB should be $10\log_{10}[A^2/4] + 10\log_{10}[L] = -6.02 + 24.08 = 18.06$ dB

- The difference between the spectral peak and the noise spectral density is about 18.06 - (-3.01) = 21.07 dB

- As another special case consider $A = 1$, $L = 1024$ and SNR = -20 dB



Averaged periodogram of two sinusoids in white noise at SNR = -20 dB

- It seems an impossible task to discern a sinusoid at a -20 dB SNR
- The increased spectral resolution makes this possible

- We expect the noise spectral density to be at

$$\sigma_w^2 = \frac{1^2}{2 \cdot 10^{-20/10}} = 50 \text{ or } 17.0 \text{ dB}$$

- The spectral peak in dB should be $10\log_{10}[1/4] + 10\log_{10}[1024] = -6.02 + 30.10 = 24.08$ dB

- The difference between the spectral peak and the noise spectral density is about 24.08 - (17) = 7.08 dB

8.3 Quantization Effects in DFT Computations

In this section we will study the effects of finite precision arithmetic in the direct computation of the DFT and also in a radix-2 FFT. The source of error initially considered is due to rounding after multiplication. For a $B + 1$ bit word length (one sign bit) multiplication produces a $2B + 1$ bit word which we assume is rounded back to $B + 1$ bits. A simplified analysis will be performed by assuming that at each point in the algorithm where coefficient multiplication occurs we place an additive white noise source.

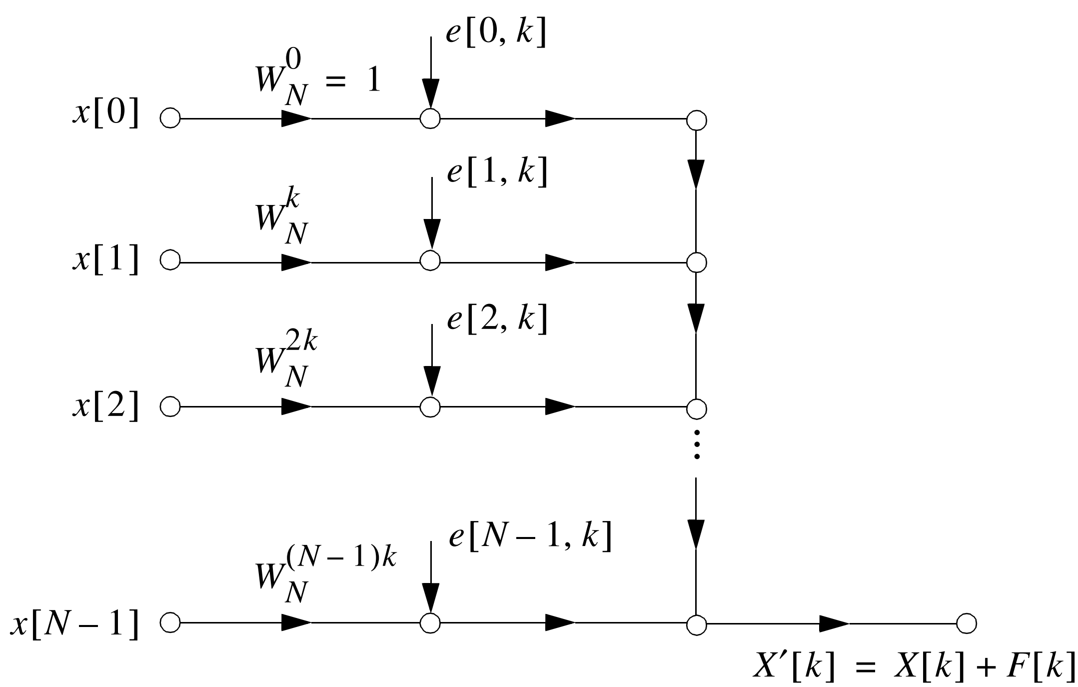
8.3.1 Quantization Errors in Direct Computation of the DFT

For the sequence $x[n]$, $0 \leq n \leq N - 1$ the DFT is defined as

$$X[k] = \sum_{n=0}^{N-1} x[n] W_N^{kn}, \quad k = 0,\ 1,\ \ldots,\ N - 1$$

where as defined earlier $W_N = e^{-j(2\pi/N)}$.

- If we assume that $x[n]$ is a complex valued sequence, then the product $x[n]W_N^{kn}$ requires four real multiplications
- The total complex error due to a twiddle factor multiplication is denoted $e[n,k]$



Linear noise model roundoff noise contributions in the kth point of an N-point DFT

- As a result of each multiplication/rounding operation an error sequence $e_i[n,k]$, $i = 1, 2, 3, 4$ is generated
- The following assumptions are made about each $e_i[n,k]$
  - The errors are uniform over the interval $[-\Delta, \Delta]$, where $\Delta = 2^{-B-1}$ assuming $B + 1$ bit signed fractions. The variance is $2^{-2B}/12$

  - The error sequences are mutually uncorrelated from each other over the 4N multiplies required per DFT point
  - The errors are uncorrelated with the input signal and hence also the output transform

- The complex error has zero mean and variance

$$\text{var}\{e[n,k]\} = 4\frac{2^{-2B}}{12} = \frac{1}{3}2^{-2B}$$

- The quantization error variance associated with each DFT point is just

$$\sigma_F^2 = \frac{N}{3}2^{-2B}$$

- To prevent overflow we must limit the input dynamic range. To avoid overflow in addition suppose that we have $|X[k]| < 1$, then

$$|X[k]| \leq \sum_{n=0}^{N-1} |x[n]| < N, \quad k = 0, 1, \ldots, N-1$$

  Thus in the worst case we must scale each input by N to avoid overflow (assuming initially that $|x[n]| < 1$).

- Further suppose that the input is a white sequence uniform on $[-1/N, 1/N]$

- The variance of the input is

$$\sigma_x^2 = \frac{(2/N)^2}{12} = \frac{1}{3N^2}$$

- The variance of the DFT output $X[k]$ is just

$$\sigma_X^2 = N\sigma_x^2 = \frac{1}{3N}$$

since the twiddle factors have magnitude of one

- Finally the output SNR is

$$\frac{\sigma_X^2}{\sigma_F^2} = \frac{2^{2B}}{N^2}$$

Example 8.4: 1024-Point DFT

In the computation of a 1024-point DFT an SNR of at least 30 dB is required. Find how many bits of precision are required.

- Using the results presented above

$$\begin{aligned} 30 &= 10\log_{10}\left[2^{(2B-20)}\right] \\ &= 3.01(2B-20) \Rightarrow \ B \geq 15 \end{aligned}$$

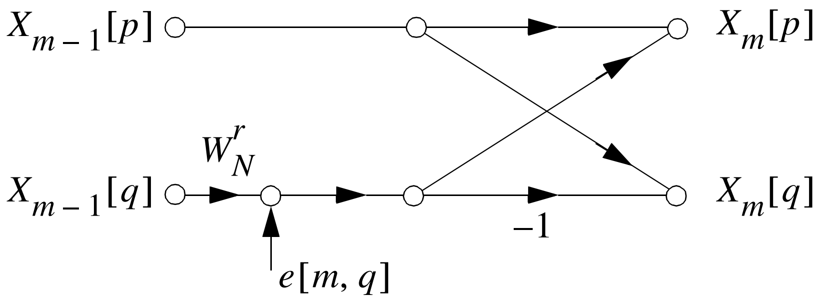
8.3.2 Quantization Errors in the Radix-2 FFT

The radix-2 FFT algorithm requires significantly fewer multiplications than the direct DFT. From this fact one might initially conclude that there is less quantization error as well. This is not true unless stage-by-stage scaling is used.

Consider the equations for a decimation-in-time butterfly

$$\begin{aligned} X_m[p] &= X_{m-1}[p] + W_N^r X_{m-1}[q] \\ X_m[q] &= X_{m-1}[p] - W_N^r X_{m-1}[q] \end{aligned}$$

The linear noise model for this butterfly is shown below.



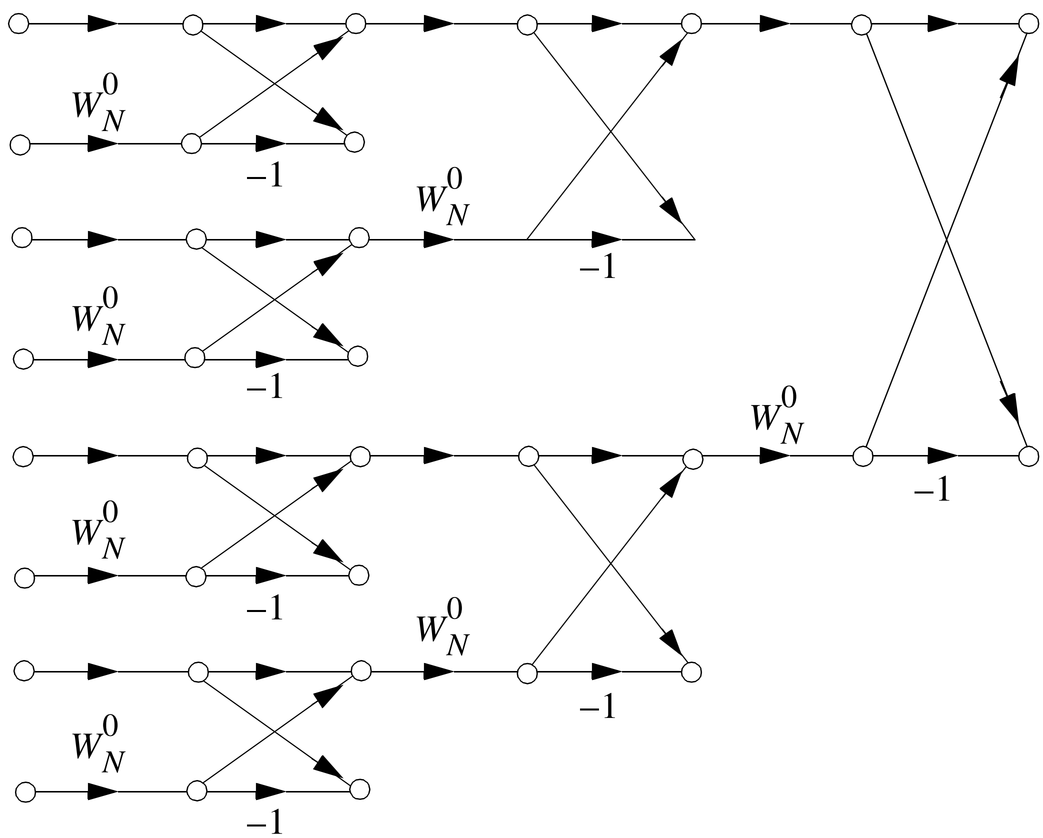
Decimation-in-time butterfly with roundoff noise source $e[m, q]$

- Each butterfly noise source is assumed to be zero mean and white with variance

$$\mathrm{var}\{e[n,k]\} = \frac{1}{3}2^{-2b} = \sigma_B^2$$

- Each butterfly noise source is transmitted to the butterfly outputs with unity gain, since all the flow graph transmission coefficients have unity magnitude

- Note that in the radix-2 FFT each output point (node) is connected directly or indirectly to $N-1$ butterflies. The 7 butterfly connections for the point $X[0]$ when $N = 8$ are shown below.



$N = 8$ butterfly connections for computing $X[0]$

- Each butterfly noise source propagates to each output node with variance σ_B^2. The variance due to quantization noise is thus

$$\sigma_F^2 = (N-1)\sigma_B^2 \simeq N\sigma_B^2 \qquad \text{for } N \text{ large}$$

  Note this is the same result as obtained for the direct DFT

- To insure there is no overflow we can again scale each input point by N

- An improved scaling method can be devised by first considering the overflow constraints at each butterfly:

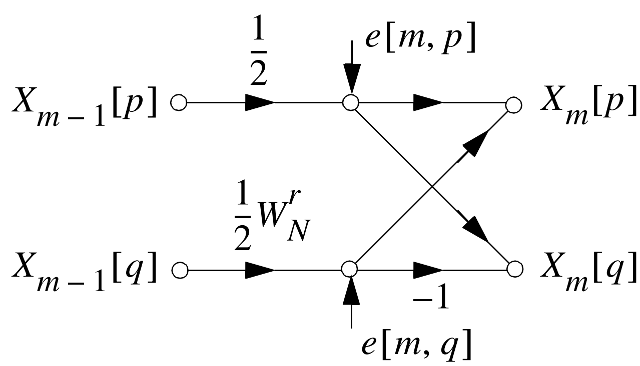
$$\max(|X_{m-1}[p]|, |X_{m-1}[q]|) \leq \max(|X_m[p]|, |X_m[q]|)$$

and

$$\max(|X_m[p]|, |X_m[q]|) \leq 2\max(|X_{m-1}[p]|, |X_{m-1}[q]|)$$

Note that the magnitude is nondecreasing from stage to stage.

- Assuming that the FFT outputs must be less than unity and noting that the maximum magnitude increase from stage to stage is at most a factor of 2, overflow scaling can be accomplished by:
  - Requiring $|x[n]| < 1$
  - Incorporating an attenuation factor of 2 into each butterfly (overall the scaling is still $1/N$)



Radix-2 butterfly with overflow scaling and roundoff noise

- The noise variance that each butterfly contributes to the total output noise, is still σ_B^2
- The total noise variance contributed by all $N-1$ butterflies will be less than the $1/N$ scaling case since each noise source passes through various levels of attenuation before arriving at an output node

- There are v butterfly stages, where $N = 2^v$
- A general formula which takes into account the 1/2 scale factors is

$$\begin{aligned}
\sigma_F^2 &= \sigma_B^2 \left\{ \underbrace{\frac{N}{2}\left(\frac{1}{4}\right)^{v-1}}_{\text{1st stage}} + \underbrace{\frac{N}{4}\left(\frac{1}{4}\right)^{v-2}}_{\text{2nd stage}} + \cdots + \underbrace{1}_{\text{vth stage}} \right\} \\
&= \sigma_B^2 \left\{ \left(\frac{1}{2}\right)^{v-1} + \left(\frac{1}{2}\right)^{v-2} + \cdots + 1 \right\} \\
&= \sigma_B^2 \sum_{m=0}^{v-1} (0.5)^m \\
&= 2\sigma_B^2 \left[1 - \left(\frac{1}{2}\right)^v \right] \simeq 2\sigma_B^2 = \frac{2}{3} 2^{-2B}
\end{aligned}$$

- If we assume that the input is a white sequence uniform on $[-1, 1]$. The variance of the input is

$$\sigma_x^2 = \frac{1}{3}$$

- The variance of the DFT output $X[k]$ is still $\sigma_X^2 = 1/(3N)$ since the twiddle factors have magnitude of one. Hence the output SNR is

$$\frac{\sigma_X^2}{\sigma_F^2} = \frac{2^{2B}}{2N}$$

Example 8.5: 1024-Point Radix-2 FFT

In the computation of a 1024-point DFT an SNR of at least 30 dB is required. Find how many bits of precision are required.

- Using the results presented above

$$
\begin{aligned}
30 &= 10\log_{10}\left[2^{(2B-11)}\right] \\
&= 3.01(2B-11) \Rightarrow \; B \geq 11
\end{aligned}
$$

- This compares with $B \geq 15$ required for the direct DFT or FFT with all the scaling performed on the input