SNS COLLEGE OF TECHNOLOGY

An Autonomous Institution

Coimbatore-35

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A+' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

19ECB212 – DIGITAL SIGNAL PROCESSING

II YEAR/ IV SEMESTER

UNIT 4 – FINITE WORD LENGTH EFFECTS

TOPIC – QUANTIZATION NOISE,COEFFICIENT QUANTIZATION ERROR AND PRODUCT QUANTIZATION ERROR

QUANTIZATION OF INPUT DATA

- For processing of analog signal using a digital signal the analog signal has to be digitized by A/D (Analog to Digital) Converter
- The A/D Converter consists of sampler and quantizer
- The sampler will sample the value of analog signal at uniform intervals to produce a sequence of unquantized values of the signal
- The quantizer will quantize the analog value and produce the corresponding binary codes. The process of assigning binary number to quantized analog value is called coding

QUANTIZATION OF INPUT DATA

- The two types of errors that are produced by A/D Conversion process are quantization errors and saturation errors
- The quantization error is due to representation of the sampled signal by a fixed number of digital levels (quantization levels)
- The saturation error occurs when the analog signal exceed the dynamic range of A/D converter
- In analog to digital conversion, when B-bits binary code (including sign bit) is selected, we can generate 2^B different binary numbers

QUANTIZATION OF INPUT DATA

- If the range of analog signal to be quantized is R then the quantization step size q is given by

Quantization Step Size , q = $R/\ 2^{B} = R\ /\ 2^{b+1}$

- Where, B=Size of binary including sign bit
- b = B-1 = Size of binary excluding sign bit
- Usually the analog signal is scaled such that the magnitude of quantized signal is less than or equal to one. In such case the range of analog signal to be quantized is -1 to +1, therefore R=2

QUANTIZATION OF INPUT DATA

- Let x(n) = Unquantized sample of the signal
- x $_q$(n) = Quantized sample of the signal
- Now the quantization error is defined as

Quantization error, e(n) = x $_q$(n) – x(n)

- In A/D Converters the quantization can be performed by truncation and rounding. But the quantization by rounding is preferred in A/D converters due to zero mean value of quantization error and low variance when compared to truncation

QUANTIZATION OF INPUT DATA

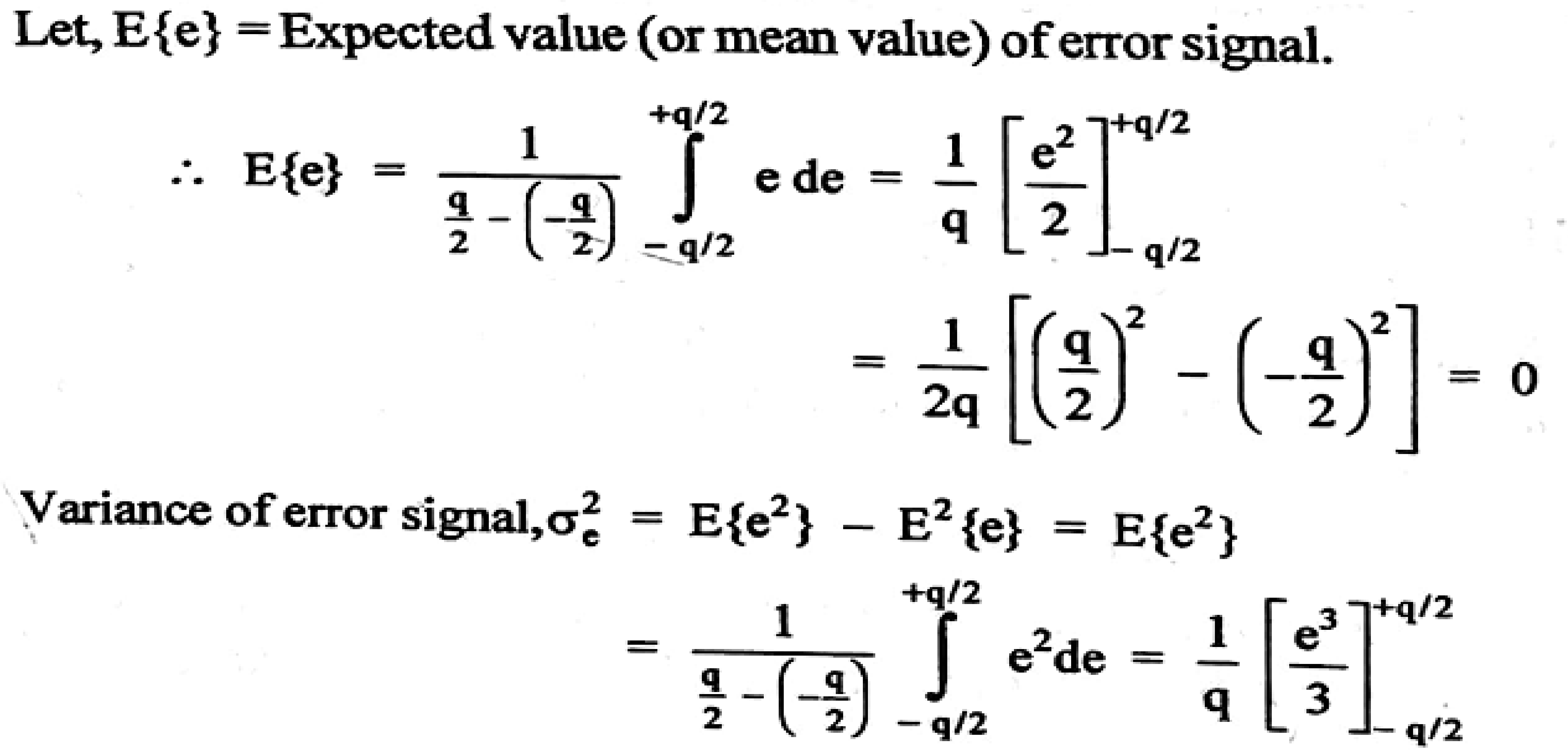
- The quantization error for rounding will be in the range of –q/2 to +q/2
- Assume that all errors are equiprobable and so the mean value of error is zero. The error due to rounding is treated as a random variable
- For a uniformly distributed random variable “x” in the interval $(x_{1,}\ x_2)$ the expected value (or mean value) and variance are given by

$$\text{Expected value, } E\{x\} = \frac{1}{x_2 - x_1}\int_{x_1}^{x_2} x\,dx$$

$$\text{Variance, } \sigma^2 = E\{x^2\} - E^2\{x\}$$

- If the random variable x is uniform in the interval (-c,c) then E{x} =0, i.e., mean value is zero and variance $\sigma^2 = E\{x^2\}$

QUANTIZATION OF INPUT DATA



Let, $E\{e\}$ = Expected value (or mean value) of error signal.

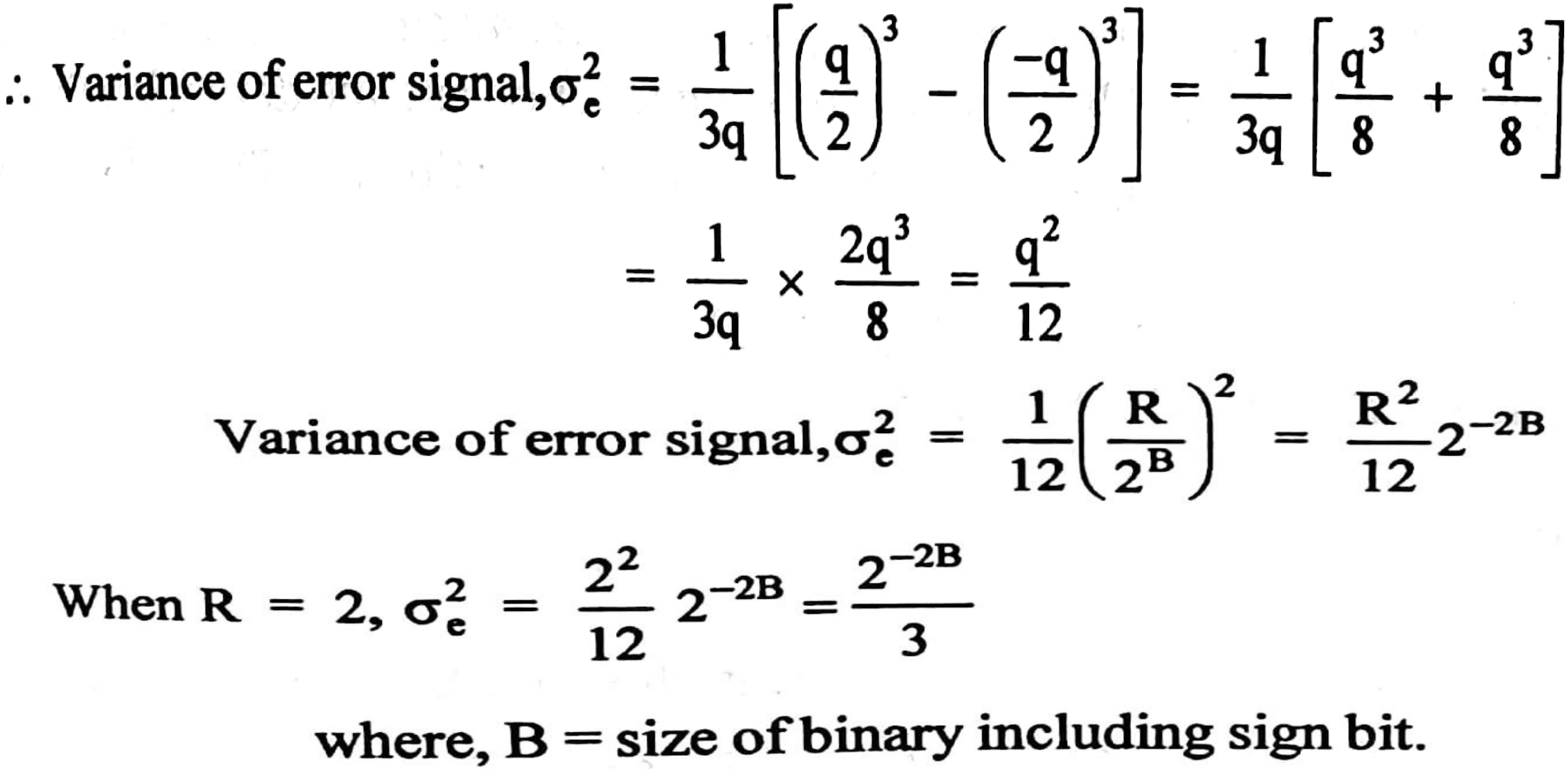
$$\therefore\ E\{e\} = \frac{1}{\frac{q}{2}-\left(-\frac{q}{2}\right)}\int_{-q/2}^{+q/2} e\,de = \frac{1}{q}\left[\frac{e^2}{2}\right]_{-q/2}^{+q/2}$$

$$= \frac{1}{2q}\left[\left(\frac{q}{2}\right)^2 - \left(-\frac{q}{2}\right)^2\right] = 0$$

Variance of error signal, $\sigma_e^2 = E\{e^2\} - E^2\{e\} = E\{e^2\}$

$$= \frac{1}{\frac{q}{2}-\left(-\frac{q}{2}\right)}\int_{-q/2}^{+q/2} e^2 de = \frac{1}{q}\left[\frac{e^3}{3}\right]_{-q/2}^{+q/2}$$

QUANTIZATION OF INPUT DATA



$$\therefore \text{ Variance of error signal}, \sigma_e^2 = \frac{1}{3q}\left[\left(\frac{q}{2}\right)^3 - \left(\frac{-q}{2}\right)^3\right] = \frac{1}{3q}\left[\frac{q^3}{8} + \frac{q^3}{8}\right]$$

$$= \frac{1}{3q} \times \frac{2q^3}{8} = \frac{q^2}{12}$$

$$\text{Variance of error signal}, \sigma_e^2 = \frac{1}{12}\left(\frac{R}{2^B}\right)^2 = \frac{R^2}{12}2^{-2B}$$

$$\text{When } R = 2,\ \sigma_e^2 = \frac{2^2}{12}2^{-2B} = \frac{2^{-2B}}{3}$$

where, B = size of binary including sign bit.

QUANTIZATION OF INPUT DATA

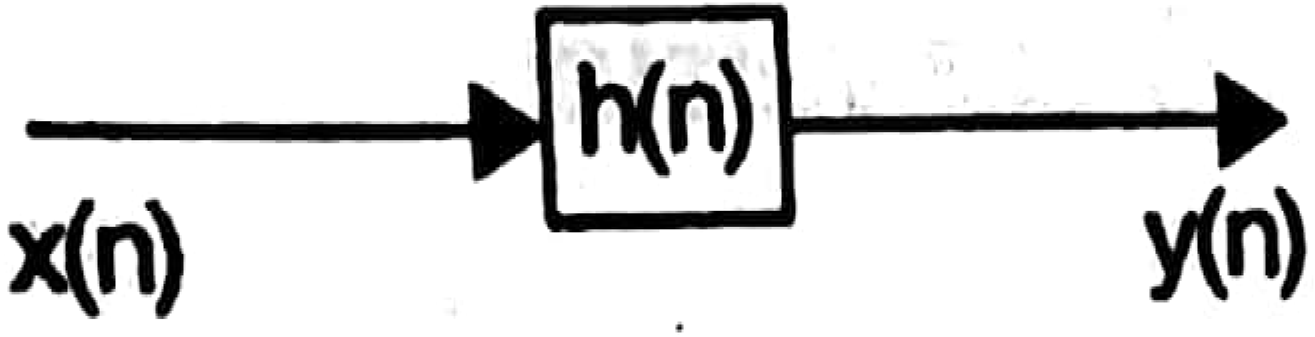
- The variance of error signal is also called steady state noise power due to input quantization. In variance of error signal, the steady state noise power tends to zero as B tends to infinity
- The value of B is infinite only if A/D converter has infinite precision, which is not practically possible. When a large number of bits are used to digitize such a signal, then the analog noise are well represented on the digitized signal
- Increasing the bits in A/D converter beyond a certain limit merely increases the accuracy by which an analog noise is represented
- Therefore the word length of an A/D converter also depends on the type of signal to be converted

STEADY STATE OUTPUT NOISE VARIANCE (POWER) DUE TO THE QUANTIZATION ERROR SIGNAL

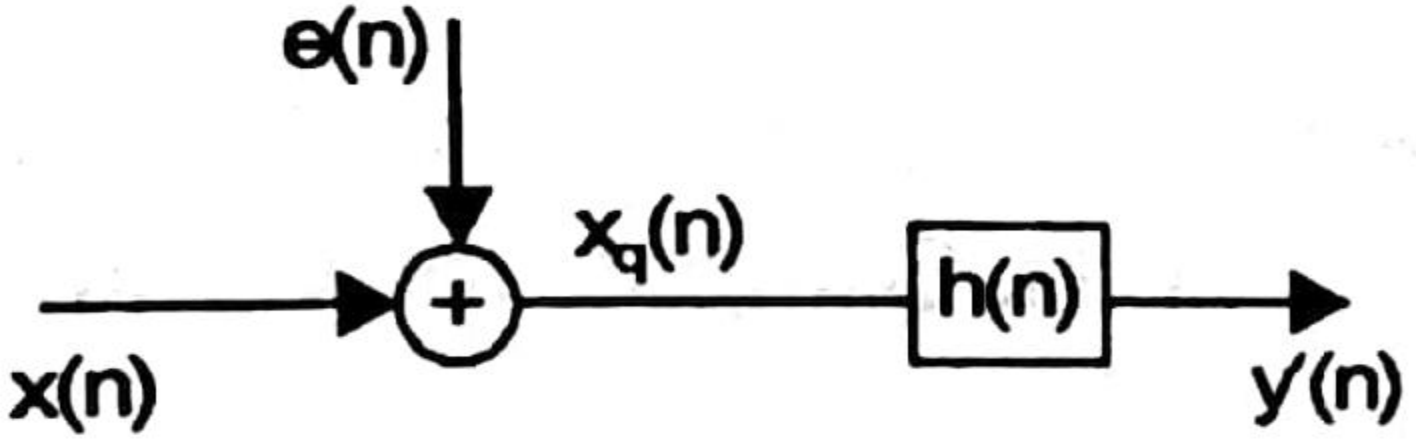
- The quantized input signal of a digital system can be represented as a sum of unquantized signal x(n) and error signal e(n) as shown

$$\mathbf{x_q(n) = x(n) + e(n)}$$

LTI System with Unquantized Input



LTI System with quantized Input



Representation of input quantization noise in an LTI System

STEADY STATE OUTPUT NOISE VARIANCE (POWER) DUE TO THE QUANTIZATION ERROR SIGNAL

- h(n) is the impulse response of the system and y'(n) is the response or output of the system due to input and error signal. The response of the system is given by convolution of input and impulse response

- For linear systems using distributive property of convolution the response y'(n) can be written as **$y'(n) = x_q(n) * h(n)$**

$$y'(n) = [x(n) + e(n)] * h(n)$$

$$y'(n) = [\, x(n) * h(n)] + [e(n)*h(n)]$$

$$y'(n) = y(n) + \varepsilon(n)$$

- Where y(n) = x(n)*h(n) = Output due to input signal x(n)

- ε (n) = e(n)*h(n) = = Output due to error signal e(n)

STEADY STATE OUTPUT NOISE VARIANCE (POWER) DUE TO THE QUANTIZATION ERROR SIGNAL

- The variance of the signal ε(n) is called output noise power or steady state output noise power (or variance) due to the quantization error signal. Using autocorrelation function and the definition for variance of a discrete time signal, the expression for output noise power shown as

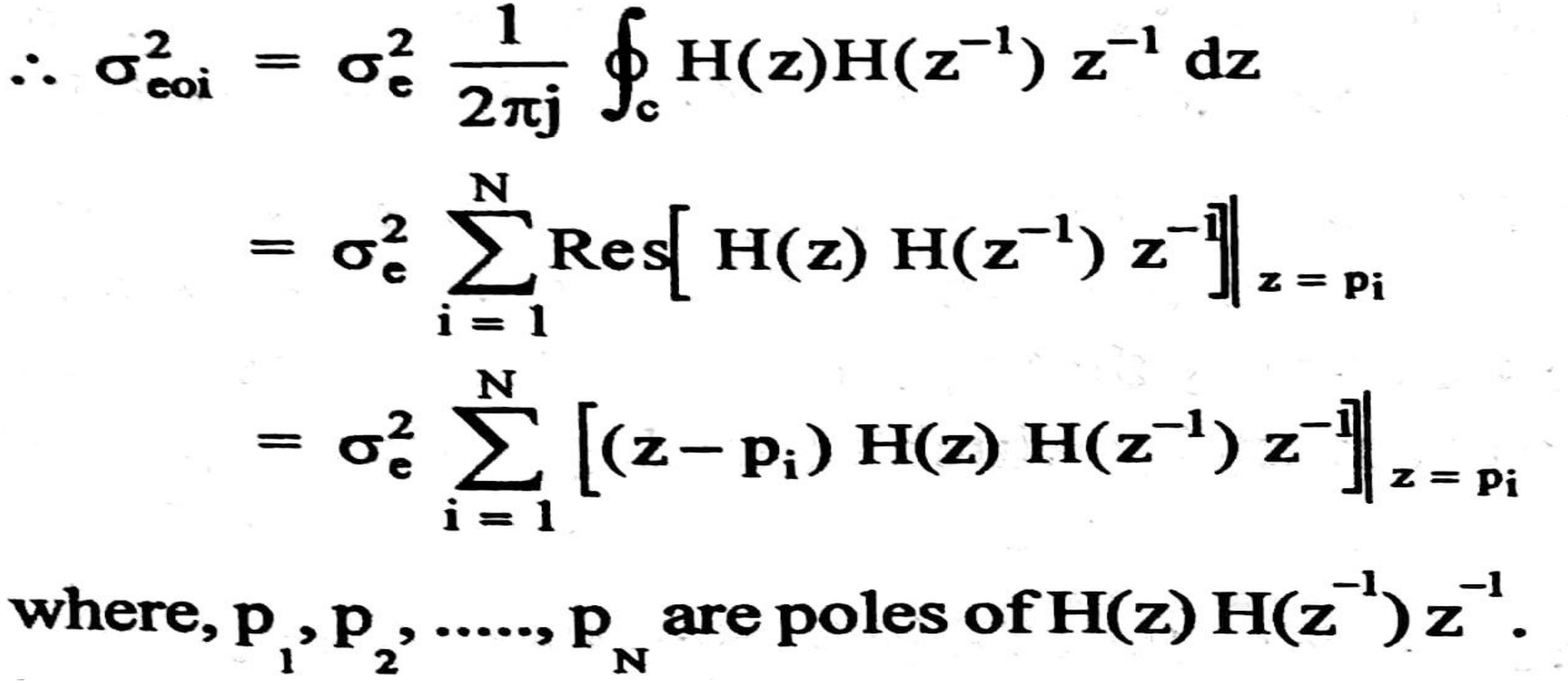
$$\left.\begin{array}{l}\text{Steady state output noise power}\\ \text{due to input quantization errors}\end{array}\right\}\sigma_{eoi}^{2}=\sigma_{e}^{2}\sum_{n=0}^{\infty}h^{2}(n)$$

- The variance of error signal $\sigma_e{}^2$ can be evaluated and the summation of h^2 (n) can be evaluated using parseval's theorem

$$\therefore\ \sigma_{eoi}^{2}=\sigma_{e}^{2}\sum_{n=0}^{\infty}h^{2}(n)=\sigma_{e}^{2}\,\frac{1}{2\pi j}\oint_{c}H(z)\,H(z^{-1})\,z^{-1}\,dz$$

STEADY STATE OUTPUT NOISE VARIANCE (POWER) DUE TO THE QUANTIZATION ERROR SIGNAL

- The closed contour integration of above equation can be evaluated using residue theorem of Z transform



$$\therefore\ \sigma_{eoi}^2 = \sigma_e^2 \frac{1}{2\pi j} \oint_c H(z)H(z^{-1})\, z^{-1}\, dz$$

$$= \sigma_e^2 \sum_{i=1}^{N} \text{Res}\left[H(z)\, H(z^{-1})\, z^{-1} \right]\Big|_{z=p_i}$$

$$= \sigma_e^2 \sum_{i=1}^{N} \left[(z-p_i)\, H(z)\, H(z^{-1})\, z^{-1} \right]\Big|_{z=p_i}$$

where, $p_1, p_2, \ldots\ldots, p_N$ are poles of $H(z)\, H(z^{-1})\, z^{-1}$.

- Since the closed contour integration in above equation is around the unit circle |z| =1, only the residues for the poles that lie inside the unit circle in z plane

QUANTIZATION OF FILTER COEFFICIENTS

- In the realization of FIR and IIR filters in hardware or in software, the accuracy with which filter coefficients can be specified is limited by the word length of the register used to store the coefficients
- Usually the filter coefficients are quantized to the word size of the register used to store them either by truncation or rounding
- The location (or the value) of poles and zeros of the digital filters directly depends on the value of filter coefficients. The quantization of the filter coefficients will modify the value of poles and zeros, and so the location of the poles and zeros will be shifted from the desired location

QUANTIZATION OF FILTER COEFFICIENTS

- This will create deviations in the frequency response of the system. Hence we obtain a filter having a frequency response that is different from the frequency response of the filter with unquantized coefficients
- The sensitivity of the filter frequency response characteristics to quantization of the filter coefficients is minimized by realizing the filter having a large number of poles and zeros as an interconnection of second-order sections
- This leads to the parallel form and cascade form realization in which the basic building blocks of first –order and second-order sections. It is possible to prove that the coefficient quantization has less effect in cascade realization when compared to parallel realization

PRODUCT QUANTIZATION ERROR

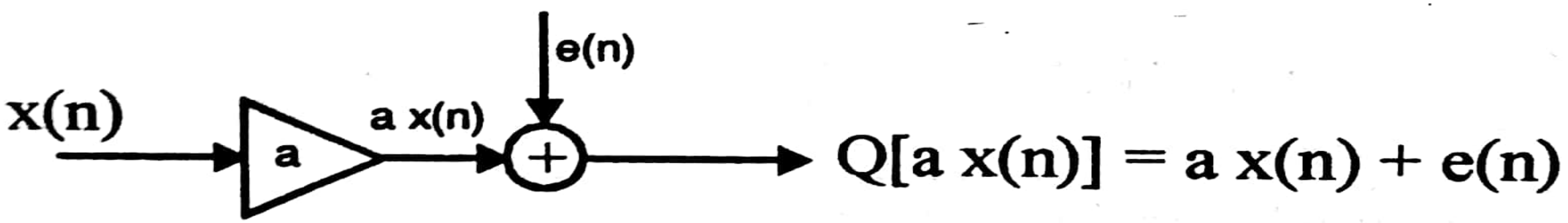
- In realization structures of IIR Systems, multipliers are used to multiply the signal by constants. The output of the multipliers i.e., the product are quantized to finite word length in order to store them in registers and to be used in subsequent calculations
- In fixed point arithmetic, the multiplication of two b-bit numbers results in a product of length 2b-bits. If the word length of the register used to store the result is b-bits then it is necessary to quantize the product (result) to b-bits. The error due to quantization of the output of multiplier is referred to as **Product Quantization Error**

PRODUCT QUANTIZATION ERROR

- In digital system the product quantization is performed by rounding due to the following desirable characteristics of rounding
- In rounding the error signal is independent of the type of arithmetic employed
- The mean value of error signal due to rounding is zero
- The variance of error signal due to rounding is the least
- The analysis of product quantization error is similar to the analysis of quantization error due to A/D process
- But in product quantization error analysis it is necessary to define the noise transfer function, which depends on the structure of the digital network

PRODUCT QUANTIZATION ERROR

- The Noise Transfer Function (NTF) is defined as the transfer function from the noise source to the filter output (i.e., NTF is the transfer function obtained by treating the noise source as actual input)
- The model of the multiplier of a digital network using fixed point arithmetic as shown. The multiplier is considered as an infinite precision multiplier. Using an adder the error signal is added to the output of the multiplier so that the output of adder is equal to the quantized product
- Therefore the output of finite word length multiplier can be expressed as



PRODUCT QUANTIZATION ERROR

Quantized Product = Q[a x(n)] = a x(n) + e(n)

- Where a x(n) = Unquantized Product
- e(n) = Product quantization error signal
- The product quantization error signal is treated as a random process with uniform probability density function. The following assumptions are made regarding the statistical independence of the various noise sources in the digital filter
- Any two different samples from the same noise source are considered
- Any two different noise sources, When considered as random processes are uncorrelated
- Each noise source is uncorrelated with the input sequence

ASSESSMENT

1. The A/D Converter consists of -------------- & ---------------

2. The two types of errors that are produced by A/D Conversion process are ----------- and ----------------

3. Define quantization error

4. The variance of the signal $\varepsilon(n)$ is called -----------------------

5. What is meant by product quantization error?

6. Define Noise Transfer Function.

7. Filter coefficients are quantized to the word size of the register used to store them either by --------------- and ---------------

THANK YOU