Finite Wordlength Effects
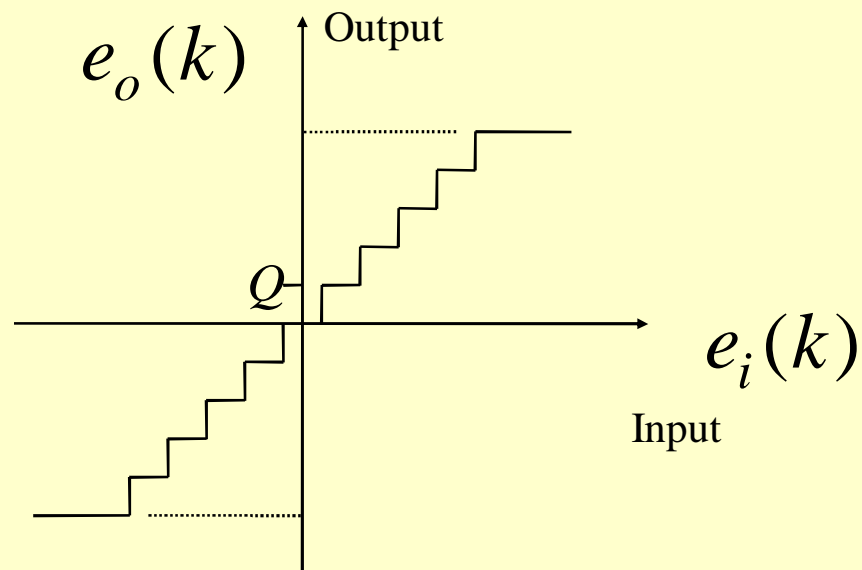
- Finite register lengths and A/D converters cause errors in:-

  (i)   Input quantisation.

  (ii)   Coefficient (or multiplier) quantisation

  (iii) Products of multiplication truncated or rounded due to machine length
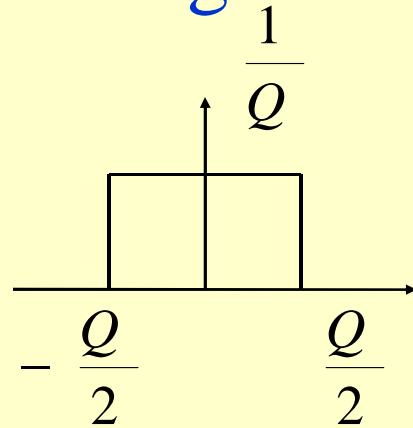
Finite Wordlength Effects

- Quantisation

$$e_o(k) \quad \text{Output}$$

$$Q$$

$$e_i(k)$$

Input

$$-\frac{Q}{2} \leq e_{i,o}(k) \leq \frac{Q}{2}$$

Finite Wordlength Effects

- The pdf for *e* using rounding

- Noise power $\quad \sigma^2 = \int\limits_{-Q/2}^{Q/2} e^2 p(e).de = E\{e^2\}$

  or
  $$\sigma^2 = \frac{Q^2}{12}$$

Finite Wordlength Effects

- Let input signal be sinusoidal of unity amplitude. Then total signal power $P = \dfrac{1}{2}$

- If b bits used for binary then $Q = 2/2^b$

    so that $\sigma^2 = 2^{-2b}/3$

- Hence $P/\sigma^2 = \dfrac{3}{2} \cdot 2^{+2b}$

    or $\text{SNR} = 1.8 + 6b$ dB

Finite Wordlength Effects

- Consider a simple example of finite precision on the coefficients a, b of second order system with poles $\rho e^{\pm j\theta}$

$$H(z) = \frac{1}{1 - az^{-1} + bz^{-2}}$$

$$H(z) = \frac{1}{1 - 2\rho\cos\theta.z^{-1} + \rho^2.z^{-2}}$$

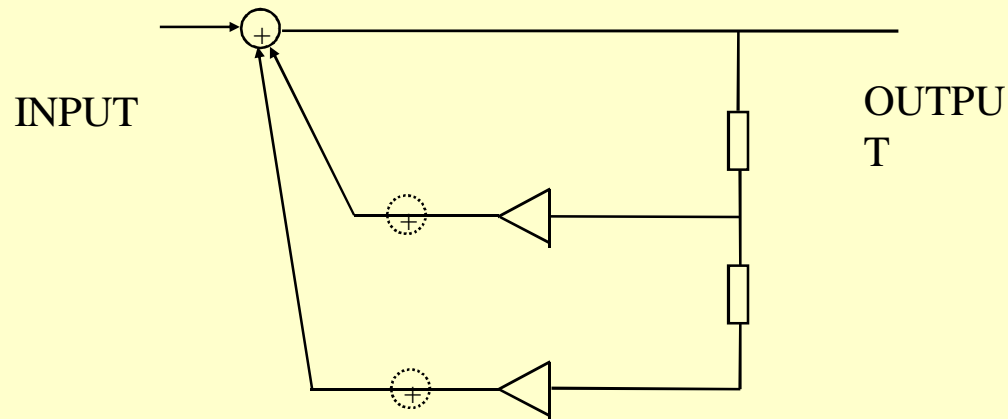- where $\quad a = 2\rho\cos\theta \qquad b = \rho^2$

Finite Wordlength Effects

bit pattern	$2\rho\cos\theta,\ \rho^2$	ρ
000	0	0
001	0.125	0.354
010	0.25	0.5
011	0.375	0.611
100	0.5	0.707
101	0.625	0.791
110	0.75	0.866
111	0.875	0.935
1.0	1.0	1.0

Finite Wordlength Effects

- Finite wordlength computations

Limit-cycles; "Effective Pole" Model; Deadband

- Observe that for

$$H(z) = \frac{1}{(1 + b_1 z^{-1} + b_2 z^{-2})}$$

- instability occurs when

$$|b_2| \to 1$$

- i.e. poles are
    - (i) either on unit circle when complex
    - (ii) or one real pole is outside unit circle.

- Instability under the "effective pole" model is considered as follows

Finite Wordlength Effects

- In the time domain with $H(z) = Y(z) \Big/ X(z)$

-
$$y(n) = x(n) - b_1 y(n-1) - b_2 y(n-2)$$

- With $|b_2| \to 1$ for instability we have $Q[b_2 y(n-2)]$ indistinguishable from $y(n-2)$

- Where $Q[.]$ is quantisation

Finite Wordlength Effects

- With <u>rounding</u>, therefore we have

$$b_2\, y(n-2) \pm 0.5 \qquad y(n-2)$$

  are indistinguishable (for integers)

  or $\quad b_2\, y(n-2) \pm 0.5 = y(n-2)$

- Hence

$$y(n-2) = \frac{\pm 0.5}{1-b_2}$$

- With both positive and negative numbers

$$y(n-2) = \frac{\pm 0.5}{1-|b_2|}$$

Finite Wordlength Effects

- The range of integers $\dfrac{\pm 0.5}{1 - |b_2|}$

  constitutes a set of integers that cannot be individually distinguished as separate or from the asymptotic system behaviour.

- The band of integers $\left(-\dfrac{0.5}{1 - |b_2|}, \quad +\dfrac{0.5}{1 - |b_2|} \right)$

  is known as the "deadband".

- In the second order system, under rounding, the output assumes a cyclic set of values of the deadband. This is a limit-cycle.

Finite Wordlength Effects

- Consider the transfer function

$$G(z) = \frac{1}{(1 + b_1 z^{-1} + b_2 z^{-2})}$$

$$y_k = x_k - b_1 y_{k-1} - b_2 y_{k-2}$$

- if poles are complex then impulse response is given by h_k

$$h_k = \frac{\rho^k}{\sin\theta} . \sin\left[(k+1)\theta\right]$$

Finite Wordlength Effects

- Where $\rho = \sqrt{b_2}$ $\qquad \theta = \cos^{-1}\left(-b_1 \Big/ 2\sqrt{b_2}\right)$

- If $b_2 = 1$ then the response is sinusiodal with frequency

$$\omega = \frac{1}{T}\cos^{-1}\left(-b_1 \Big/ 2\right)$$

- Thus product quantisation causes instability implying an "effective " $b_2 = 1$.
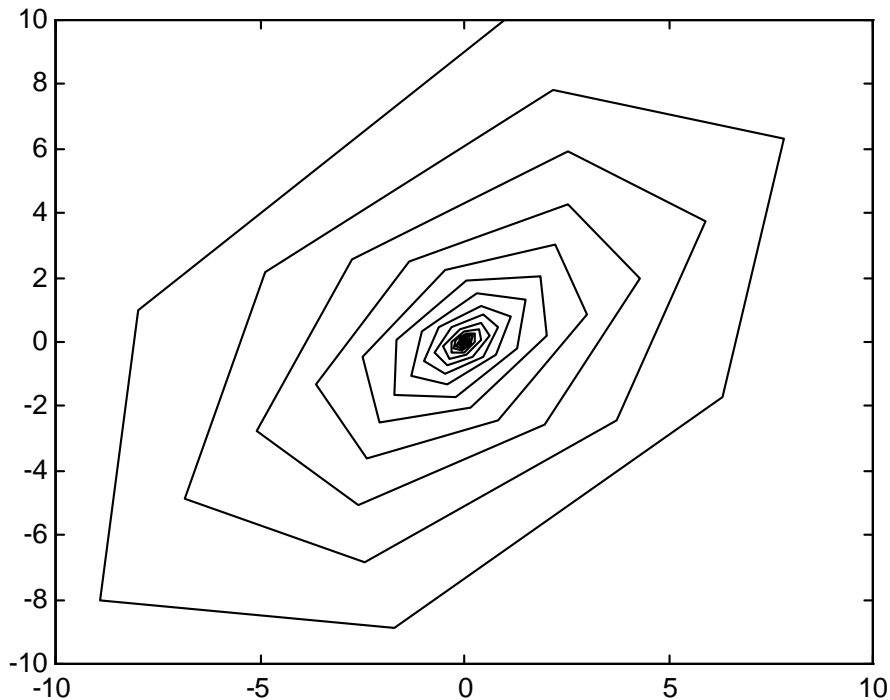
Finite Wordlength Effects

- Consider infinite precision computations for
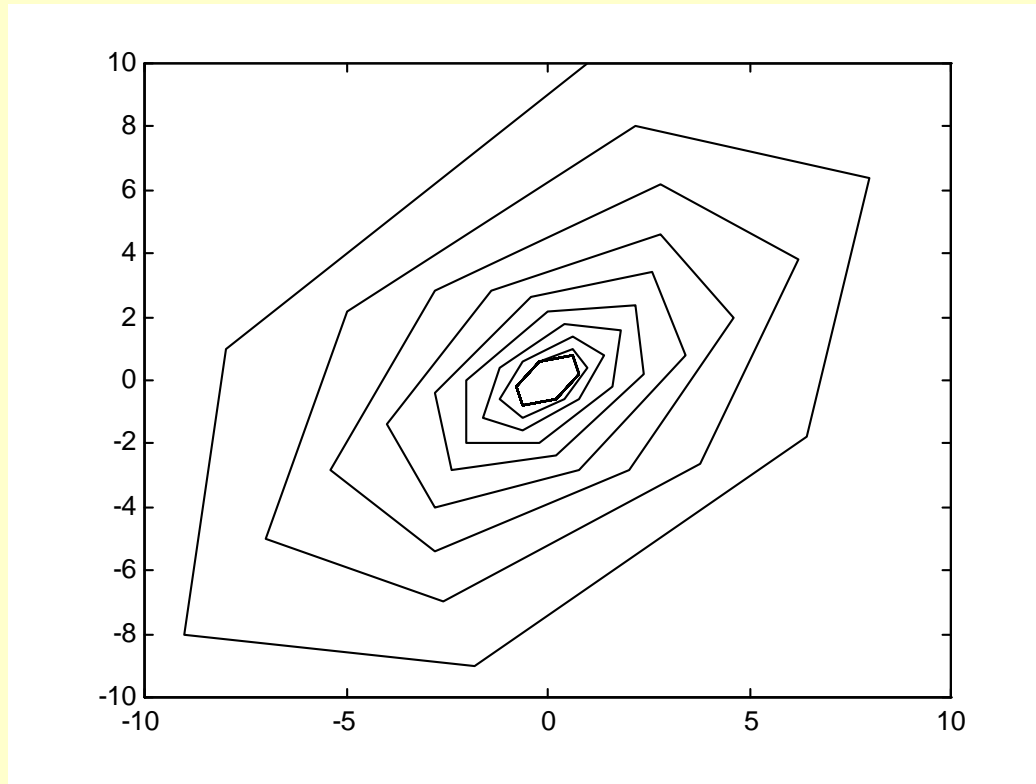
$$y_k = x_k + y_{k-1} - 0.9y_{k-2}$$

$$x_0 = 10$$

$$x_k = 0; \ k \neq 0$$

Finite Wordlength Effects

- Now the same operation with integer precision

Finite Wordlength Effects

- Notice that with infinite precision the response converges to the origin

- With finite precision the reponse does not converge to the origin but assumes cyclically a set of values –the Limit Cycle

Finite Wordlength Effects

- Assume $\{e_1(k)\}$, $\{e_2(k)\}$ ..... are not correlated, random processes etc.

$$\sigma_{0i}^2 = \sigma_e^2 \sum_{k=0}^{\infty} h_i^2(k) \qquad \sigma_e^2 = Q^2 / 12$$

Hence total output noise power

$$\sigma_0^2 = \sigma_{01}^2 + \sigma_{02}^2 = 2 . \frac{2^{-2b}}{12} \sum_{k=0}^{\infty} \rho^{2k} . \frac{\sin^2[(k+1)\theta]}{\sin^2\theta}$$

- Where $Q = 2^{-b}$ and

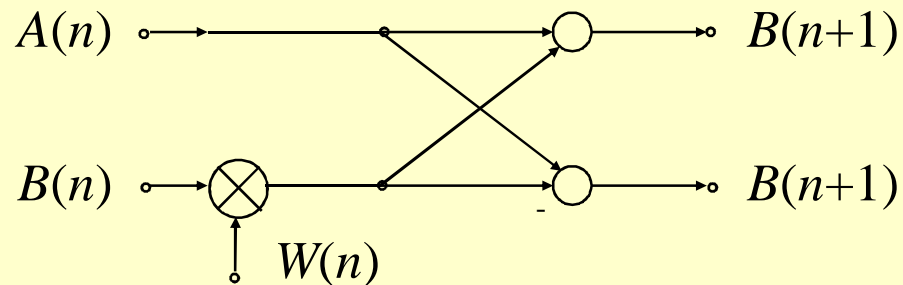$$h_1(k) = h_2(k) = \rho^k . \frac{\sin[(k+1)\theta]}{\sin\theta}; \ k \geq 0$$

Finite Wordlength Effects

- ie

$$\sigma_0{}^2 = \frac{2^{-2b}}{6}\left[\frac{1+\rho^2}{1-\rho^2}\cdot\frac{1}{1+\rho^4-2\rho^2\cos 2\theta}\right]$$

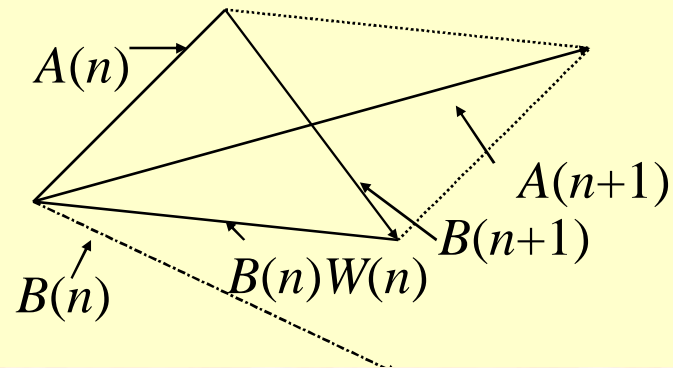# Finite Wordlength Effects

- ## For FFT

$$A(n + 1) = A(n) + W(n).B(n)$$

$$B(n + 1) = A(n) - W(n).B(n)$$

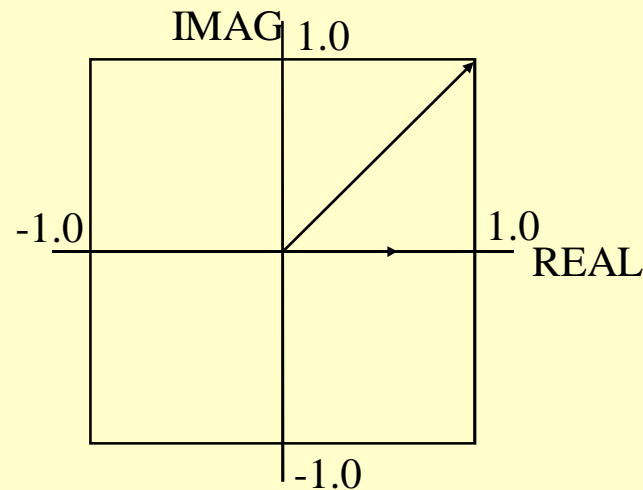# Finite Wordlength Effects

- FFT

$$\left|A(n+1)\right|^2 + \left|B(n+1)\right|^2 = 2$$

$$\left|A(n+1)\right|^2 = 2\left|A(n)\right|^2$$

$$\left|A(n)\right| = \sqrt{2}\left|A(n)\right|$$

- AVERAGE GROWTH: 1/2 BIT/PASS

Finite Wordlength Effects

- FFT

IMAG 1.0

-1.0   1.0
              REAL

-1.0

$$A_x(n+1) = A_x(n) + B_x(n)C(n) - B_y(n)S(n)$$

$$\left|A_x(n+1)\right| < \left|A_x(n)\right| + \left|B_x(n)\right|\left|C(n)\right| - \left|B_y(n)\right|\left|S(n)\right|$$

$$\frac{\left|A_x(n+1)\right|}{\left|A_x(n)\right|} < 1.0 + \left|C(n)\right| - \left|S(n)\right| = 2.414 \ ....$$

- PEAK GROWTH: 1.21.. BITS/PASS

Finite Wordlength Effects

- Linear modelling of product quantisation

$$x(n) \rightarrow \boxed{Q[\cdot]} \rightarrow \tilde{x}(n)$$

- Modelled as

$$x(n) \rightarrow \oplus \rightarrow \tilde{x}(n) = x(n) + q(n)$$
$$q(n)$$

Finite Wordlength Effects

- For <u>rounding</u> operations $q(n)$ is uniform distributed between $-\frac{Q}{2}, \frac{Q}{2}$ and where Q is the quantisation step (i.e. in a wordlength of bits with sign magnitude representation or $\mathrm{mod}\ 2,\ Q = 2^{-b}$).

- A discrete-time system with quantisation at the output of each multiplier may be considered as a multi-input linear system

Finite Wordlength Effects

$$q_1(n)...q_2(n)...q_p(n)$$

$$\{x(n)\} \longrightarrow \boxed{h(n)} \longrightarrow \{y(n)\}$$
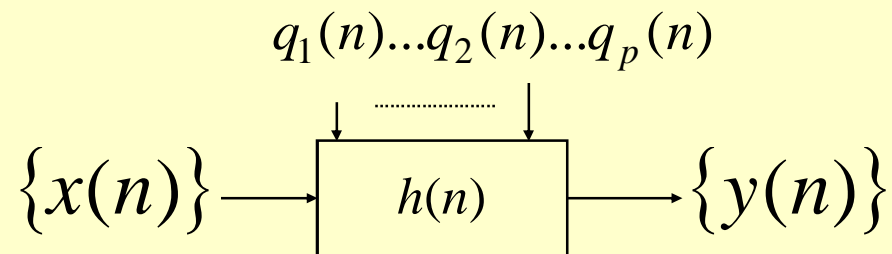
- Then

$$y(n) = \sum_{r=0}^{\infty} x(r).h(n-r) + \sum_{\lambda=1}^{p} \left[\sum_{r=0}^{\infty} q_\lambda(r).h_\lambda(n-r) \right]$$

- where $h_\lambda(n)$ is the impulse response of the system from λ the <u>output</u> of the multiplier to $y(n)$.

Finite Wordlength Effects

- For zero input i.e. $x(n) = 0, \forall n$ we can write

$$|y(n)| \le \sum_{\lambda=1}^{p} |\hat{q}_\lambda| \cdot \sum_{r=0}^{\infty} |h_\lambda(n-r)|$$

- where $|\hat{q}_\lambda|$ is the maximum of $|q_\lambda(r)|, \; \forall \lambda, r$ which is not more than $\dfrac{Q}{2}$

- ie $\quad |y(n)| \le \dfrac{Q}{2} \cdot \sum_{\lambda=1}^{p} \left[\sum_{n=0}^{\infty} |h_\lambda(n-r)| \right]$

Finite Wordlength Effects

- However

$$\sum_{n=0}^{\infty} |h_\lambda(n)| \leq \sum_{n=0}^{\infty} |h(n)|$$

- And hence

$$|y(n)| \leq \frac{pQ}{2} . \sum_{n=0}^{\infty} |h(n)|$$

- ie we can estimate the maximum swing at the output from the system parameters and quantisation level