Discrete Time Linear System

skimu@me.com

February 16, 2026

Abstract

From z-transform to switched cap delta-sigma modulator.

1 Discrete Time Linear System

We are going to study discrete time linear systems. In such systems, we only care discrete time points nT_s, where n is an integer and T_s is the sampling period. For time dependent quantities, we use square bracket to point time points of the quantity:

$$x[n] = x(nT_s),$$

where $x(t)$ is a function of the continuous time t. In a linear system, relation between input and output is linear, which means it satisfies principle of superposition. Suppose that an input of $x_1[n]$ gives output of $y_1[n]$ and that another input $x_2[n]$ gives output $y_2[n]$, then the output will be $a\,y_1[n] + b\,y_2[n]$ if input $a\,x_1[n] + b\,x_2[n]$ is given. Let $F\{\cdot\}$ denote the relation between the input and the output of the system, we write above relation between x_1, x_2 and y_1, y_2 as

$$y_1[n] = F\{x_1[n]\}, \quad y_2[n] = F\{x_2[n]\}.$$

When the system is linear, the principle of superposition can be expressed as

$$F\{a\,x_1[n] + b\,x_2[n]\} = a\,y_1[n] + b\,y_2[n],$$

where a and b is arbitrary constant.

1.1 Impulse response and frequency response

Let's decompose x into impulse functions:

$$x[n] = \sum_{m=-\infty}^{\infty} x[m]\,\delta[n-m].$$

where $\delta[n]$ is the impulse function:

$$\delta[n] = \begin{cases} 1 & (n = 0), \\ 0 & (n \neq 0), \end{cases}$$

and put it into $y = F\{x\}$:

$$y[n] = F\{x[n]\},$$
$$= \sum_{m=-\infty}^{\infty} F\{x[m]\,\delta[n-m]\}.$$

Here $x[m]$ is just a parameterized coefficient. We can bring it out from F:

$$y[n] = \sum_{m=-\infty}^{\infty} x[m]\,F\{\delta[n-m]\}.$$

Exchanging the order of multiplication and using $h[n]$ for impulse response $F\{\delta[n]\}$, we see $h[n-m]$ is contribution weight of input at time m to the output at n:

$$y[n] = \sum_{m=-\infty}^{\infty} h[n-m]\,x[m], \qquad h[n] = F\{\delta[n]\}.$$

The system will not react before the impulse is given to the input, therefore $h[n] = 0$ for $n < 0$, and it should be enough to take summation up to n for $y[n]$:

$$y[n] = \sum_{m=-\infty}^{n} h[n-m]\,x[m].$$

Now we would like to take a look at frequency response of the system. Output $y[n]$ for complex sine wave input, $x[n] = x_f\,e^{i\,2\pi f\,nT_s}$ will be

$$y[n] = \sum_{m=-\infty}^{n} h[n-m]\,x_f\,e^{i\,2\pi f\,mT_s} = \sum_{m=-\infty}^{n} h[n-m]\,x_f\,e^{i\,\frac{2\pi f}{F_s}m},$$

where we defined $F_s = 1/T_s$. With a new variable, $m' = n - m$, $(m = n - m')$, we can get rid of n from summation boundary and summand, i.e.,

$$y[n] = \sum_{m'=\infty}^{0} h[m']\,x_f\,e^{i\,\frac{2\pi f}{F_s}(n-m')},$$

$$= x_f\,e^{i\,\frac{2\pi f}{F_s}n} \sum_{m=0}^{\infty} h[m]\,e^{-i\,\frac{2\pi f}{F_s}m}.$$

The summation is constant (does not depend on n), i.e., sine wave input gives sine wave output of the same frequency, which is anticipated result from a linear system. If we write $y[n] = y_f \exp(i\frac{2\pi f}{F_s}n)$, we see frequency response is Fourier transform of impulse response, (since $h[n] = 0$ for $m < 0$, we can take summation from $-\infty$) and we call it H.

$$\frac{y_f}{x_f} = \sum_{n=0}^{\infty} h[n]\,e^{-i\,\frac{2\pi f}{F_s}n} = \sum_{n=-\infty}^{\infty} h[n]\,e^{-i\,\frac{2\pi f}{F_s}n} \equiv H(e^{i\,\frac{2\pi f}{F_s}}).$$

Inverse Fourier transform of will be,

$$h[n] = \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} H(e^{i\,\frac{2\pi f}{F_s}})\,e^{i\,\frac{2\pi f}{F_s}n}df.$$

Since $h[n] = 0$ for $n < 0$, there must be corresponding condition for a complex function H to be a frequency response function. We will come back to this condition later in Section 1.4. However, we will see the reason why I write $H(e^{i\,\frac{2\pi f}{F_s}})$ rather than $H(f)$ in the next page.

3

1.2 Frequency response and z-Transform

We have just learnt that impulse response is Fourier transform of frequency response:

$$h[n] = \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} H(e^{i\frac{2\pi f}{F_s}}) \, e^{i\frac{2\pi f}{F_s}n} df, \qquad H(e^{i\frac{2\pi f}{F_s}}) = \sum_{n=-\infty}^{\infty} h[n] \, e^{-i\frac{2\pi f}{F_s}n}.$$

With $z = e^{i\frac{2\pi f}{F_s}}$, above can be written as

$$h[n] = \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} H(z) \, z^n \, df, \qquad H(z) = \sum_{n=-\infty}^{\infty} h[n] \, z^{-n}.$$

We see that z makes a circle as f goes from $-F_s/2$ to $F_s/2$, therefore the integral with respect to f become contour integral on complex plane.

$$h[n] = \frac{1}{2\pi i} \oint_{|z|=1} H(z) \, z^{n-1} dz,$$

where we used $dz = \frac{2\pi i}{F_s} e^{i\frac{2\pi f}{F_s}} df$.

Now, we found a transformation pair. This transform from n space to z space is called z-Transform and we use $\mathcal{Z}\{\cdot\}$ to denote this transform:

$$\mathcal{Z}\{h[n]\} = \sum_{n=0}^{\infty} h[n] \, z^{-n} = H(z).$$

Note that summation starts from 0 since $h[n] = 0$ for $n < 0$.

The inverse transform $\mathcal{Z}^{-1}$ is,

$$\mathcal{Z}^{-1}\{H(z)\} = \frac{1}{2\pi i} \oint_{|z|=1} H(z) \, z^{n-1} dz = h[n].$$

This transform has similar properties as Fourier/Laplace transform. For example,

$$y[n] = a\,x_1[n] + b\,x_2[n] \quad \rightarrow \quad Y(z) = aX_1 + bX_2,$$

$$y[n] = \sum_{m=-\infty}^{n} h[n-m]\,x[m] \quad \rightarrow \quad Y(z) = H(z)X(z),$$

$$H(z^*) = H(z)^*.$$

And z acts as time increment operator:

$$\mathcal{Z}^{-1}\{zH(z)\} = h[n+1], \qquad \mathcal{Z}^{-1}\{z^{-1}H(z)\} = h[n-1].$$
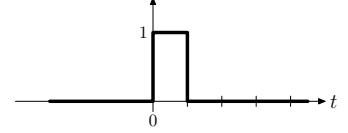
And $H(e^{i\frac{2\pi f}{F_s}})$ gives frequency response. Since $e^{i\frac{2\pi f}{F_s}}$ is a periodic function of period F_s, $H(e^{i\frac{2\pi f}{F_s}})$ is also a periodic function of the same period. This is anticipated result from a discrete time system where frequency f and f plus any integer multiple of F_s gives the same time progression.

4

1.3   z-Transform of basic functions

Let's take a look at z-Transform of a few discrete time functions, and calculate its inverse transform to see if it gives the original function.

1.3.1   Impulse function

$$x[n] = \begin{cases} 1 & (n = 0), \\ 0 & (n \neq 0). \end{cases}$$

Taking z-Transform yields,

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n=0}^{\infty} x[n]\, z^{-n} = x[0]\, z^{-0} = 1.$$

z-Transform of impulse function is 1.

The inverse transform is calculated as follows:

$$\mathcal{Z}^{-1}\{X(z)\} = \frac{1}{2\pi i} \oint_{|z|=1} X(z)\, z^{n-1} dz = \frac{1}{2\pi i} \oint_{|z|=1} z^{n-1} dz.$$

Using

$$z = e^{\frac{i2\pi f}{F_s}}, \qquad dz = \frac{i2\pi}{F_s} e^{\frac{i2\pi f}{F_s}} df,$$
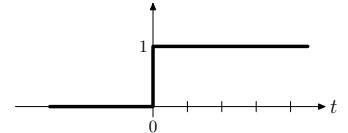
the contour integral on complex plane becomes usual integral on real axis, i.e.,

$$\mathcal{Z}^{-1}\{X(z)\} = \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} e^{\frac{i2\pi f}{F_s} n} df = \begin{cases} 1 & (n = 0), \\ 0 & (n \neq 0). \end{cases}$$

We see that inverse z-Transform of 1 is impulse function.

1.3.2   Step function

$$x[n] = \begin{cases} 0 & (n < 0), \\ 1 & (n \geq 0). \end{cases}$$

z-Transform:

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n=0}^{\infty} x[n]\, z^{-n} = \sum_{n=0}^{\infty} z^{-n} = \frac{1}{1 - z^{-1}}.$$

5

Inverse transform:

$$\mathcal{Z}^{-1}\{X(x)\} = \frac{1}{2\pi i} \oint_{|z|=1} X(z)\, z^{n-1}\, dz = \frac{1}{2\pi i} \oint_{|z|=1} \frac{z^{n-1}}{1-z^{-1}}\, dz = \frac{1}{2\pi i} \oint_{|z|=1} \frac{z^n}{z-1}\, dz.$$

We have a pole on top of $|z| = 1$, but we get the correct answer if we take integral over slightly larger circle, at least for $n \geq 0$,

$$\mathcal{Z}^{-1}\{X(x)\} = \lim_{\delta \to +0} \frac{1}{2\pi i} \oint_{|z|=1+\delta} \frac{z^n}{z-1}\, dz = 1. \qquad (n \geq 0)$$

We will cover case of $n < 0$ in the next section.

1.3.3 Geometric progression

Now, let's take a look at a geometric progression.

$$x[n] = \begin{cases} 0 & (n < 0), \\ a\,\gamma^n & (n \geq 0). \end{cases}$$

Step function is a special case when $a = 1$ and $\gamma \to 1$. z-Transform is

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n=0}^{\infty} a\gamma^n z^{-n} = \frac{a}{1-\gamma z^{-1}}.$$

Let's calculate inverse transform of $X(z)$ to see if it gives the original function: $a\,\gamma^n$.

$$\mathcal{Z}^{-1}\{X(z)\} = \frac{1}{2\pi i} \oint_{|z|=1} X(z)\, z^{n-1}\, dz = \frac{1}{2\pi i} \oint_{|z|=1} \frac{a\,z^n}{z-\gamma}\, dz$$

For $n \geq 0$, the numerator of the integrand is holomorphic, the integral have non-zero value if $|\gamma| \leq 1$, i.e.,

$$\mathcal{Z}^{-1}\{X(z)\} = \frac{1}{2\pi i} \oint_{|z|=1} \frac{a\,z^n}{z-\gamma}\, dz = \begin{cases} a\,\gamma^n & (|\gamma| \leq 1,\ n \geq 0), \\ 0 & (|\gamma| > 1,\ n \geq 0). \end{cases}$$

Note that we take slightly larger circle if $|\gamma| = 1$ just as we did in the last section.

For $n < 0$, let's start with $n = -1$,

$$\mathcal{Z}^{-1}\{X(z)\}\big|_{n=-1} = \frac{a}{2\pi i} \oint_{|z|=1} \frac{1}{z(z-\gamma)}\, dz$$

Recalling that

$$\frac{1}{z(z-\gamma)} = \frac{1}{\gamma}\left(\frac{1}{z-\gamma} - \frac{1}{z}\right),$$

we get

$$\mathcal{Z}^{-1}\{X(z)\}\big|_{n=-1} = \frac{a}{\gamma \, 2\pi i} \oint_{|z|=1} \frac{1}{z-\gamma} - \frac{1}{z} \, dz.$$

Therefore

$$\mathcal{Z}^{-1}\{X(z)\}\big|_{n=-1} = \begin{cases} 0 & (|\gamma| \le 1), \\ -a\,\gamma^{-1} & (|\gamma| > 1). \end{cases}$$

If we repeat this procedure for $n = -2, -3, ...$, we will find

$$\mathcal{Z}^{-1}\{X(z)\}\big|_{n=-2} = \gamma^{-1}\,\mathcal{Z}^{-1}\{X(z)\}\big|_{n=-1},$$
$$\mathcal{Z}^{-1}\{X(z)\}\big|_{n=-3} = \gamma^{-1}\,\mathcal{Z}^{-1}\{X(z)\}\big|_{n=-2}, \dots$$

Therefore

$$\mathcal{Z}^{-1}\{X(z)\} = \begin{cases} 0 & (|\gamma| \le 1, \ n < 0), \\ -a\,\gamma^{n} & (|\gamma| > 1, \ n < 0). \end{cases}$$

Combining $n \ge 0$ case for $|\gamma| \le 1$

$$\mathcal{Z}^{-1}\{X(z)\} = \begin{cases} 0 & (n < 0), \\ a\,\gamma^{n} & (n \ge 0). \end{cases}$$

And for $|\gamma| > 1$

$$\mathcal{Z}^{-1}\{X(z)\} = \begin{cases} -a\,\gamma^{n} & (n < 0), \\ 0 & (n \ge 0). \end{cases}$$

The inverse transform gives different answer depending on absolute value of γ. The inverse transform gives the original function if only $|\gamma| \le 1$ and $|\gamma| > 1$ gives non-zero value for negative time points, which is not a valid response function.

Sine wave generator   Let's consider a case that $\gamma = e^{i\frac{2\pi f}{F_s}}$, $a = 1$. And use $\theta = \frac{2\pi f}{F_s}$ to make equations look simpler.

$$x[n] = e^{i\theta n}, \qquad X(z) = \frac{1}{1 - e^{i\theta}\,z^{-1}}.$$

Recalling that $\cos\theta$ and $\sin\theta$ is, respectively, real and imaginary part of $e^{i\theta}$,

$$c[n] = \cos\theta n = \operatorname{Re} x[n], \qquad s[n] = \sin\theta n = \operatorname{Im} x[n].$$

Therefore, z-Transform of cosine and sine will be

$$C(z) = \mathcal{Z}\{\cos n\theta\} = \mathcal{Z}\left\{\frac{e^{in\theta} + e^{-in\theta}}{2}\right\} = \frac{X(z) + X(z)^*}{2} = \frac{1 - z^{-1}\cos\theta}{1 - 2\cos\theta \cdot z^{-1} + z^{-2}},$$

7

and

$$S(z) = \mathcal{Z}\left\{\sin n\theta\right\} = \mathcal{Z}\left\{\frac{e^{in\theta} - e^{-in\theta}}{2i}\right\} = \frac{X(z) - X(z)^*}{2i} = \frac{z^{-1}\sin\theta}{1 - 2\cos\theta \cdot z^{-1} + z^{-2}}.$$

Therefore for cosine,

$$\left(1 - 2\cos\theta \cdot z^{-1} + z^{-2}\right) C(z) = 1 - z^{-1}\cos\theta.$$

Bringing this back to n-space yields

$$c[n] = 2\cos\theta\, c[n-1] - c[n-2] + \delta[n] - \cos\theta \cdot \delta[n-1].$$

or
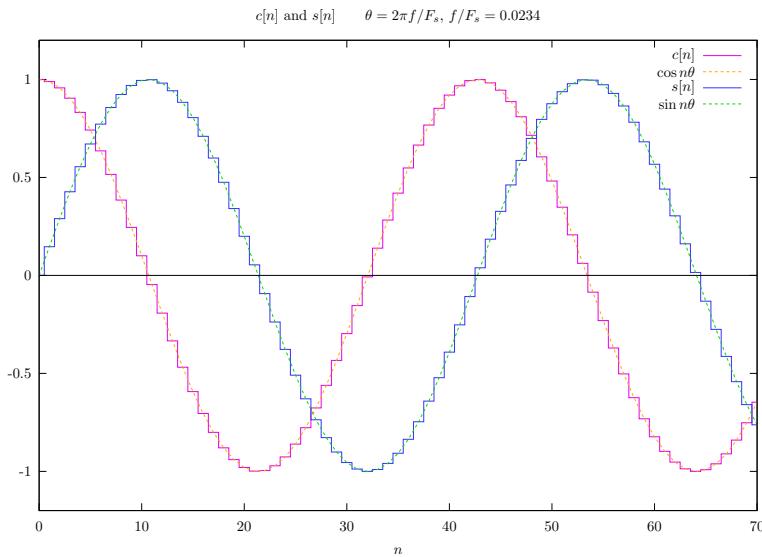
$$c[0] = 2\cos\theta \cdot c[-1] - c[-2] + \delta[0] - \cos\theta \cdot \delta[-1] = 1,$$
$$c[1] = 2\cos\theta \cdot c[0] - c[-1] + \delta[1] - \cos\theta \cdot \delta[0] = \cos\theta,$$
$$c[n] = 2\cos\theta \cdot c[n-1] - c[n-2]. \qquad (n \geq 2)$$

$c[n]$ should be equal to $\cos n\theta = \cos(2\pi f/F_s \cdot n) = \cos(2\pi f\, nT_s)$.

As for sine, only initial values are different:

$$s[0] = 0,$$
$$s[1] = \sin\theta,$$
$$s[n] = 2\cos\theta \cdot s[n-1] - s[n-2]. \qquad (n \geq 2)$$

Below compares $c[n]$ and $s[n]$ with the system math function, $\sin(\theta n)$ and $\cos(\theta n)$.

8

1.4 Causality condition

We are interested in causal systems. Impulse response of such a system should be zero before the impulse is given. Now we would like to find the causality condition, i.e., impulse response is zero for all negative time points. That is $h[-n] = 0$ for $n \geq 1$.

$$h[-n] = \frac{1}{2\pi i} \oint_{|z|=1} H(z)\, z^{-n-1}\, dz \;=\; 0. \qquad (n \geq 1)$$

To avoid pole at the origin (which is within the unit circle), we bring this integral to $\zeta = z^{-1}$ space.

$$\zeta = z^{-1}, \qquad d\zeta = -(z^{-1})^2 dz.$$

Since ζ rotates opposite direction to z, the minus sign in $d\zeta$ will be canceled with $\oint$. The integral becomes

$$\frac{1}{2\pi i} \oint_{|z|=1} H(z)\, z^{-n-1}\, dz = \frac{1}{2\pi i} \oint_{|z|=1} H(z)\, (z^{-1})^{n-1}(z^{-1})^2\, dz,$$

$$= \frac{1}{2\pi i} \oint_{|\zeta|=1} H(\zeta^{-1})\, \zeta^{n-1}\, d\zeta.$$

Therefore the wanted condition is that $H(\zeta^{-1})$ does not have pole inside the unit circle. The region inside the unit circle in ζ-space is mapped to the region outside the unit circle in z-space. Therefore the same condition can be said that $H(z)$ does not have pole outside the unit circle in z-space.

In the geometric progression example we saw in the last section, $H(z)$ have pole at $z = \gamma$ and $|\gamma|$ has to be less than 1 to meet $h[-n] = 0$.

Formal solution for unstable system When we find $H(\zeta^{-1})$ to have poles inside the unit circle as a result of a difference equation (See next section), we can still find causal impulse response by making integration path smaller so that all poles are located outside the circle. However, $H(e^{i\frac{2\pi f}{F_s}})$ will lose its meaning as frequency response function. Indeed, such a solution grows indefinitely, there's no Fourier transform of such function.

In the geometric progression with $|\gamma| > 1$, for example, if we use a path which encircles the point $z = \gamma$, we will get causal result $\mathcal{Z}^{-1}\{X(z)\} = a\gamma^n$ for $n \geq 0$. However it does not have reasonable frequency response.

In this document, we are mainly concerned with stable systems, i.e., response functions are bound and have reasonable frequency responses. We are going to use $|z| = 1$ exclusively.

1.5   z-Transform and difference equation

1.5.1   General form of response function

Consider a linear system which maps input $x[n]$ to output $y[n]$. Output y at time point n should be determined from the past states (input and output) and the current input $x[n]$. Therefore $y[n]$ should be a linear combination of those.

$$y[n] = \sum_{j=1}^{D} a_j\, y[n-j] + \sum_{j=0}^{N} b_j\, x[n-j],$$

where D, N determines how far past point makes influence to the current state of the system and a_j, b_j are constant coefficients. Taking z-Transform yields

$$Y(z) = \sum_{j=1}^{D} a_j\, z^{-j}\, Y(z) + \sum_{j=0}^{N} b_j\, z^{-j}\, X(z). \tag{1.1}$$

Therefore transfer function $H(z)$ can be written with coefficients of difference equations as follows.

$$H(z) = \frac{Y(z)}{X(z)} = \frac{b_1 z^{-1} + b_2 z^{-2} + ... + b_N z^{-N}}{1 - a_1 z^{-1} - a_2 z^{-2} - ... - a_D z^{-D}}$$

As we have just seen in the previous section, the condition for $H(z)$ being causal as well as $H(e^{i\omega t})$ being reasonable frequency response will be that polynomial $1 - a_1\zeta - a_2\zeta^2 - ... - a_D\zeta^D$ does not have root inside the unit circle.

Suppose that γ_n are the roots of the denominator, $H(z)$ can be decomposed into partial fractions like this

$$H(z) = \frac{\alpha_1}{1 - \gamma_1 z^{-1}} + \frac{\alpha_2}{1 - \gamma_2 z^{-1}} + ... + \frac{\alpha_D}{1 - \gamma_D z^{-1}} + P(z^{-1}),$$

where $P(\zeta)$ is a polynomial of the order $(N-D)$, and it will be zero if $(N-D) < 0$. We would like to become more familiar with the transfer function of $\alpha/(1 - \gamma z^{-1})$.

1.5.2   Integrator

Let's take a look at a simplest case of one term with $\alpha = 1$, $\gamma = 1$, first. That is

$$Y = H(z)X, \qquad H(z) = \frac{1}{1 - z^{-1}}.$$

Therefore

$$\left(1 - z^{-1}\right) Y = X.$$

10

Bringing this back to n space yields corresponding difference equation:

$$y[n] - y[n-1] = x[n].$$

Therefore

$$y[0] = y[-1] + x[0] = x[0],$$
$$y[1] = y[0] + x[1] = x[0] + x[1],$$
$$y[2] = y[1] + x[2] = x[0] + x[1] + x[2],$$

$$\dots$$

$$y[n] = x[0] + x[1] + x[2] + \dots + x[n] = \sum_{i=0}^{n} x[n].$$

$y[n]$ is the sum of all the input from the past. $H(z) = 1/(1 - z^{-1})$ is an integrator.
    Frequency response is

$$H(e^{i\frac{2\pi f}{F_s}}) = \begin{cases} \dfrac{1}{i2\pi f/F_s} & (f/F_s \ll 1), \\[2ex] \dfrac{1}{1+i} & (f/F_s = 1/4), \\[2ex] \dfrac{1}{2} & (f/F_s = 1/2). \end{cases}$$

Frequency response diverges as $f \to 0$, which means the input can not have DC component (average value has to be zero) otherwise output diverges.

1.5.3 Low-pass filter

Now we turn into the general case.

$$Y = H(z)X, \qquad H(z) = \frac{\alpha}{1 - \gamma z^{-1}}.$$

Frequency Response Frequency response is

$$H(e^{i\frac{2\pi f}{F_s}}) = \begin{cases} \dfrac{\alpha}{1-\gamma} \cdot \dfrac{1}{1 + i2\pi f \cdot \gamma/(1-\gamma) \cdot T_s} & (f \ll F_s), \\[2ex] \dfrac{\alpha}{1+i\gamma} & (f = \frac{1}{4}F_s), \\[2ex] \dfrac{\alpha}{1+\gamma} & (f = \frac{1}{2}F_s). \end{cases}$$

Gain changes from $\alpha/(1-\gamma)$ at $f = 0$ to $\alpha/(1+\gamma)$ at $f = F_s/2$. In case γ is real and $f \ll F_s$, this behaves as first order low-pass filter of time constant $T_s\gamma/(1-\gamma)$, or cut-off frequency of $F_s/2\pi \cdot (1-\gamma)/\gamma$.

11

Let's see the difference equation. $Y = H(z)X$ becomes

$$(1 - \gamma z^{-1})Y = \alpha X \quad \rightarrow \quad y[n] = \gamma\, y[n-1] + \alpha\, x[n].$$

If we set $\alpha = (1 - \gamma)$ to get unity gain at $f = 0$,

$$y[n] = \gamma\, y[n-1] + (1 - \gamma)\, x[n].$$

New y is at the point which which previous y and x by the ratio of $(1 - \gamma) : \gamma$.

Impulse response Recalling that impulse function is unity in z-space

$$Y = H(x)X \quad \rightarrow \quad Y = \frac{\alpha}{1 - \gamma z^{-1}} \cdot 1.$$

We have seen the inverse z-Transform of this before. It was a geometric progression.

$$y[n] = \alpha\gamma^n.$$

This is either exponential decay if γ is real, or damped oscillation if γ has non-zero imaginary part. If γ has non-zero imaginary part, $H(z)$ has to have a term with its complex conjugate γ^* to get real output from real input. This is guaranteed by the fact that γ is a root of a polynomial of real coefficient.

Step response Step response is a special case of sine wave response we will see next. The result is

$$y[n] = \frac{\alpha}{1 - \gamma}\left(1 - \gamma^{n+1}\right).$$

Sine wave response With $\theta = 2\pi f/F_s$,

$$Y = \frac{\alpha}{1 - \gamma z^{-1}} \cdot \frac{1}{1 - e^{i\theta}z^{-1}}.$$

Recalling that

$$\frac{1}{1 - \gamma_1 z^{-1}} \cdot \frac{1}{1 - \gamma_2 z^{-1}} = \frac{1}{\gamma_1 - \gamma_2}\left(\frac{\gamma_1}{1 - \gamma_1 z^{-1}} - \frac{\gamma_2}{1 - \gamma_2 z^{-1}}\right),$$

we get

$$Y = \frac{\alpha}{\gamma - e^{i\theta}}\left(\frac{\gamma}{1 - \gamma z^{-1}} - \frac{e^{i\theta}}{1 - e^{i\theta}z^{-1}}\right),$$

$$= \frac{\alpha\, e^{i\theta}}{e^{i\theta} - \gamma} \cdot \frac{1}{1 - e^{i\theta}z^{-1}} - \frac{\alpha\gamma}{e^{i\theta} - \gamma} \cdot \frac{1}{1 - \gamma z^{-1}}.$$

12

In n-space

$$y[n] = \frac{\alpha\, e^{i\theta n}}{1 - \gamma\, e^{-i\theta}} - \frac{\alpha\gamma^{n+1}}{e^{i\theta} - \gamma}.$$

If we put $\theta = 0$, we get step response. After long run where n is sufficiently large, we get $y[n] = H(e^{-i\theta})\, e^{i\theta n}$, which is anticipated from the frequency response.

How about the step response where $\gamma = 1$ and $\alpha = 1$, do we get integrator?

$$y[n] = \frac{e^{i\theta n}}{1 - e^{-i\theta}} - \frac{1}{e^{i\theta} - 1} = \frac{1 - e^{i\theta(n+1)}}{1 - e^{i\theta}} = \sum_{n=0}^{n} e^{i\theta n}.$$

And for $\theta \ll 1$, using $\exp(i\theta) \sim 1 + i\theta$, we get

$$y[n] = \frac{e^{i\theta n}}{i\theta} - \frac{1}{i\theta} = \int_{0}^{n} e^{i\theta n}\, dn.$$

1.5.4 FIR approximation

Consider following Taylor expansion

$$\frac{1}{1 - \gamma z^{-1}} = 1 + \gamma z^{-1} + \gamma^2 z^{-2} + \gamma^3 z^{-3} + \dots$$

We see that contribution weight from the past is getting less and less. Therefore we can cut-off higher order terms

$$\frac{1}{1 - \gamma z^{-1}} \sim 1 + \gamma z^{-1} + \gamma^2 z^{-2} + \dots + \gamma^m z^{-m}.$$

The good news is now response function is polynomial of z^{-1}, which does not have pole outside the unit circle, no instability to worry about. It is guaranteed that bound input gives bound output. The bad news is we need to memorize many past inputs and calculate many terms to update y. In contrast, we only needed to memorize one last state and to calculate only two terms in the original case. Impulse response is

$$y[n] = \begin{cases} 0 & (n < 0), \\ \alpha\gamma^n & (0 \le n \le m), \\ 0 & (n > m). \end{cases}$$
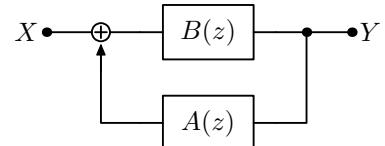
This approximation is equivalent to cutting off impulse response within finite number of time points. (Finite Impulse Response).

1.5.5 Feedback equation

Eq. (1.1) can be written

$$Y = A(z)\, Y + B(z)\, X.$$

$$Y = \frac{B(z)}{1 - A(z)}\, X.$$

13

1.6   Relation with continuous time response

For continuous time linear systems, response function $H_c(s)$ is Laplace transform of impulse response $h_c(t)$ of the system and they can be written as follows:

$$H_c(s) = \sum_{k=1}^{N} \frac{a_k \, \tau_k}{1 + s \, \tau_k}, \qquad h_c(t) = \sum_{k=1}^{N} a_k \, e^{-t/\tau_k}.$$

Here, we would like to convert this response function into z-space.

Recalling that

$$y(t) = \int_{-\infty}^{t} h_c(t - \tau) \, x(\tau) \, d\tau,$$

in case signals are moving slowly compared to T_s, we get, by rewriting integral to summation,

$$y(nT_s) = T_s \sum_{m=-\infty}^{n} h_c(nT_s - mT_s) \, x(mT_s).$$

Comparing this with

$$y[n] = \sum_{m=-\infty}^{n} h[n - m] \, x[m],$$

we find

$$h[n] = T_s \, h_c(nT_s).$$

Therefore for signals moving slowly compared to T_s,

$$h[n] = T_s \, h_c(nT_s) = T_s \sum_{k=1}^{N} a_k \, e^{-nT_s/\tau_k} = \sum_{k=1}^{N} a_k \, T_s \left(e^{-T_s/\tau_k} \right)^n.$$

In z-space,

$$H(z) = \sum_{k=1}^{N} \frac{a_k \, T_s}{1 - e^{-T_s/\tau_k} \, z^{-1}}.$$

Example
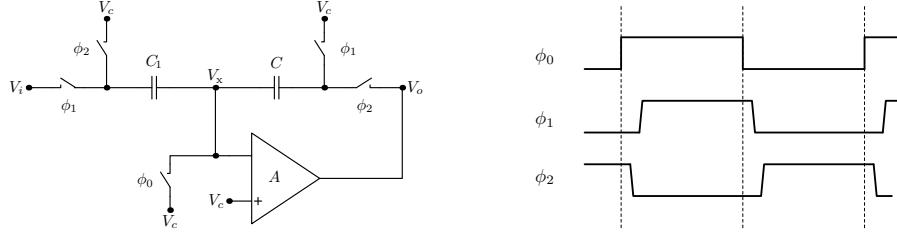
$$H_c(s) = \frac{a \, \tau}{1 + s \, \tau} \quad \rightarrow \quad H(z) = \frac{a \, T_s}{1 - e^{-T_s/\tau} z^{-1}}.$$

In case $a\tau = G$,

$$H(z) = G \cdot \frac{T_s/\tau}{1 - e^{-T_s/\tau} z^{-1}}.$$

1.7 Discrete time analog components
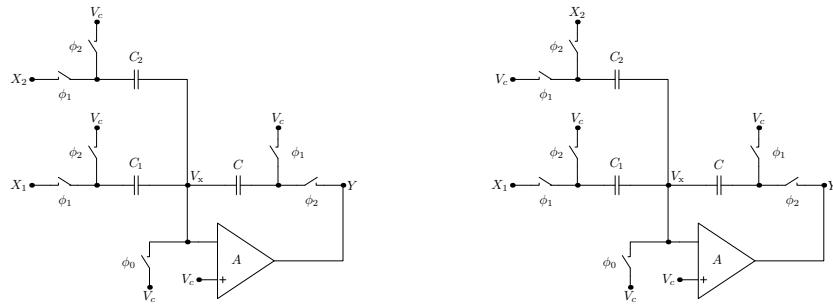
We'd like to obtain response functions of discrete time analog components to find implementations of systems described in z-space. These components have two operating phases, one is sampling (or tracking) phase and the other is amplification (or hold) phase and we take output $y[n]$ at the end of hold phase. We customarily call sampling and hold phase ϕ_1 and ϕ_2, respectively. Below shows a typical sample and hold circuit. ϕ_0 is the time reference, charge at C_1 is sampled at the falling edge of ϕ_0.

The problem is that we take output $y[n]$ at the end of ϕ_2, but input x is sampled at the end of ϕ_1, only half clock cycle before the current time point n. We use $x[n-1/2]$ in n-space, and $z^{-1/2}X$ in z-space to denote this half clock cycle delay. While there is no such thing like frequency response of $z^{-1/2}$, there is a way to get reasonable response function (rational function of z^{-1}) for a system as a whole. For example, response function of a simple sample-and-hold circuit would be $z^{-1/2}$ (simple half clock delay), however for two cascaded sampled-and-hold driven by $\bar{\phi}$ and ϕ, the one would be z^{-1}, which has reasonable response function.

1.7.1 Sample and hold
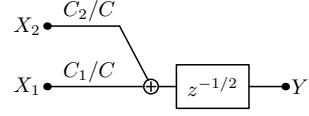
Opamp adjusts its output voltage to keep its input terminal at the ground level (V_c). Therefore the output voltage y is determined by the charge on the capacitor placed in between opamp's input and output in ϕ_2. Let's say capacitance of this capacitor is C. This capacitor is discharged in ϕ_1 and charged by the opamp in ϕ_2. Capacitors connected to inputs, C_1, C_2, are discharged in one phase and charged to corresponding input level in the other phase. Here we show two input case. For one input, we can simply put $C_2 = 0$. Extending three or more inputs is straightforward.


15

In case both C_1 and C_2 is charged at ϕ_1 (left):

$$y[n] = \frac{1}{C}\left(C_1\,x_1[n-1/2] + C_2\,x_2[n-1/2]\right),$$
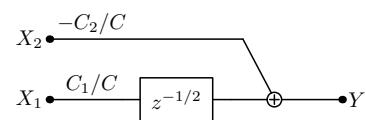
$$Y = z^{-1/2}\left(\frac{C_1}{C}\,X_1 + \frac{C_2}{C}\,X_2\right).$$

In case C_2 is charged at ϕ_2 (right):
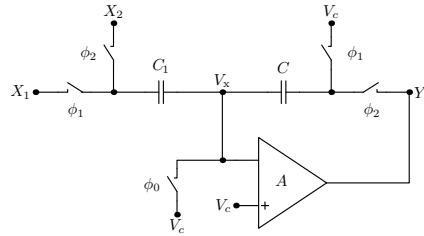
$$y[n] = \frac{1}{C}\left(C_1\,x_1[n-1/2] - C_2\,x_2[n]\right),$$

$$Y = z^{-1/2}\frac{C_1}{C}\,X_1 - \frac{C_2}{C}\,X_2.$$
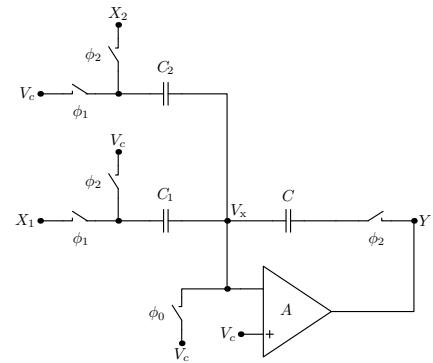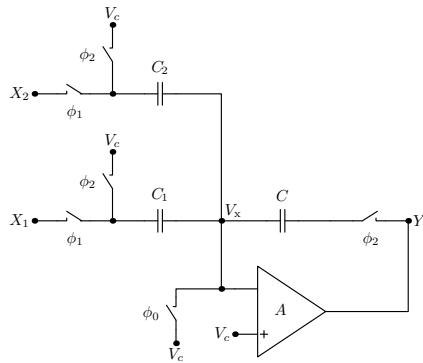
If $C_1 = C_2$, they can be merged into one.

$$Y = \frac{C_1}{C}\left(z^{-1/2}X_1 - X_2\right).$$

1.7.2 Integrator

In integrator, we do not discharge output capacitor C by disconnecting ϕ_1 switch from C. Charge in C kept unchanged during ϕ_1. Therefore output voltage $y[n]$ is its previous value plus input.


16
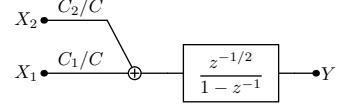
In case both C_1 and C_2 is charged at ϕ_1 like shown in above left:

$$y[n] = y[n-1] + \frac{1}{C}\left(C_1\, x_1[n-1/2] + C_2\, x_2[n-1/2]\right),$$

$$Y = z^{-1}\, Y + \frac{z^{-1/2}}{C}\left(C_1\, X_1 + C_2\, X_2\right).$$
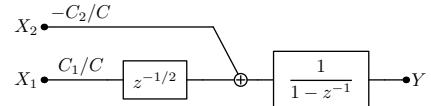
Therefore

$$Y = \frac{z^{-1/2}}{1-z^{-1}} \cdot \frac{C_1}{C}\, X_1 + \frac{z^{-1/2}}{1-z^{-1}} \cdot \frac{C_2}{C}\, X_2.$$
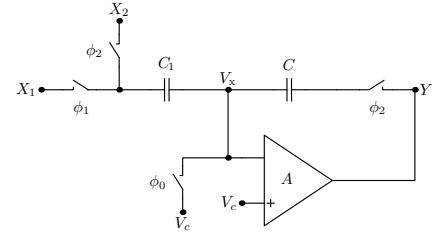
Similarly, in case C_2 is charged at ϕ_2 like shown in the right:

$$Y = \frac{z^{-1/2}}{1-z^{-1}} \cdot \frac{C_1}{C}\, X_1 - \frac{1}{1-z^{-1}} \cdot \frac{C_2}{C}\, X_2.$$

And in case $C_1 = C_2$, they can be merged into one.

$$Y = \frac{C_1}{C} \cdot \frac{1}{1-z^{-1}}\left(z^{-1/2}\, X_1 - X_2\right).$$

1.7.3   AD converter
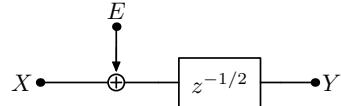
Input voltage x is digitized to output y. We think quantization error $e = y - x$ is added to x.

$$y[n] = x[n-1/2] + e[n-1/2].$$

In z-space,

$$Y = z^{-1/2}(X + E).$$

Signal flow graph is
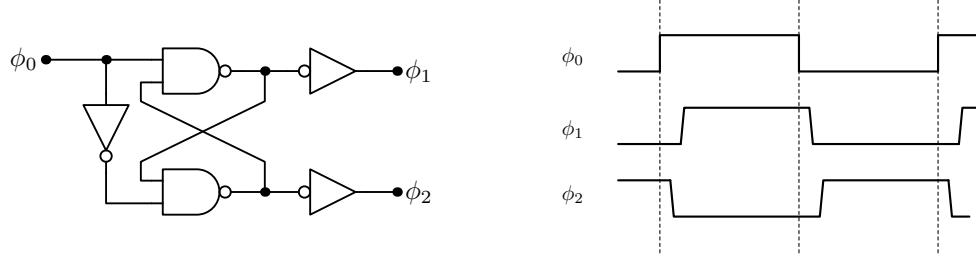
Mean square of quantization error $\langle e^2 \rangle$ is $\triangle^2/12$, where $\triangle$ is the least significant bit (LSB) of the AD converter.

It is widely believed that as the number of quantization level of the AD converter is getting larger the correlation between X and E is getting less and that E can be treated as white noise.

17

1.7.4 Non-overlap clock generator

Below left shows a circuit which generates non-overlap clock shown in the right. In a practical implementation, negative logic `INV` gate may be replaced by negative logic `NAND` gate for "enable" function.

For high speed application, we likely need carefully sized transister level implementation for desired driving ability and non-overlap time.

1.8   Our signals

We have been concerned with response functions of discrete time linear systems, where Fourier transform of impulse response, frequency response function, are always well defined. Frequency response functions are always periodic with the period of F_s, because we cannot tell frequency f from f plus any integer multiple of F_s. Therefore, we do not want to treat such two frequencies at the same time. We limit our signal's frequency component within a band of F_s centered at zero or integer multiple of F_s. In this section, we will discuss how to construct such band-limited signals.

1.8.1   Interpolation formula

Suppose that $x(t)$ does not have any frequency component outside $-F_s/2 < f < F_s/2$, in other words $x(t)$ can be written with its Fourier transform $X_c(f)$ like this:

$$x(t) = \int_{-\infty}^{\infty} X_c(f)\, e^{i2\pi ft}\, df = \int_{-F_s/2}^{F_s/2} X_c(f)\, e^{i2\pi ft}\, df$$

Comparing this with discrete time Fourier transform pair $(T_s = 1/F_s)$,

$$x(nT_s) = \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} X(f)\, e^{i\frac{2\pi f}{F_s}n}\, df, \qquad X(f) = \sum_{n=-\infty}^{\infty} x(nT_s)\, e^{-i\frac{2\pi f}{F_s}n},$$

we find, for frequencies $-F_s/2 < f < F_s/2$,

$$X_c(f) = \frac{1}{F_s}\, X(f).$$

Discrete time Fourier transform differs continuous time Fourier transform only by a scaling factor F_s. We can reproduce $x(t)$ completely from $X(f)$ which can be obtained from discrete time points, $x(nT_s)$, only. It is more convenient to write above like this

$$X_c(f) = \frac{1}{F_s} R(f)\, X(f), \qquad R(f) = \begin{cases} 1 & (-F_s/2 < f < F_s/2), \\ 0 & (\text{otherwise}). \end{cases}$$

So that $x(t)$ can be written with discrete time Fourier transform $X(f)$ like this.

$$x(t) = \int_{-\infty}^{\infty} X_c(f)\, e^{i2\pi ft}\, df = \frac{1}{F_s} \int_{-\infty}^{\infty} R(f)\, X(f)\, e^{i2\pi ft}\, df$$

$R(f)$ can also be written with its inverse transform $r(t)$.

$$R(f) = \int_{-\infty}^{\infty} r(t)\, e^{-i2\pi ft}\, dt$$

19

Therefore, by inserting this and definition of $X(f)$ into above $x(t)$ formula, we get

$$x(t) = \frac{1}{F_s} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r(t') e^{-i2\pi f t'} \, dt' \sum_{n=-\infty}^{\infty} x[n] \, e^{i\frac{2\pi f}{F_s} n} \, e^{i2\pi f t} df,$$

$$= \frac{1}{F_s} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} r(t') \, x[n] \int_{-\infty}^{\infty} e^{i2\pi f(t - nT_s - t')} \, df \, dt',$$

$$= \frac{1}{F_s} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} r(t') \, x[n] \, \delta(t - nT_s - t') \, dt',$$

$$= \frac{1}{F_s} \sum_{n=-\infty}^{\infty} r(t - nT_s) \, x[n].$$

Inverse transform of $R(f)$ can be obtained easily[1]

$$r(t) = \int_{-\infty}^{\infty} R(f) \, e^{-i2\pi ft} \, dt = \int_{-F_s/2}^{F_s/2} e^{-i2\pi ft} \, dt = \frac{\sin(\pi F_s t)}{\pi t}. \tag{1.2}$$

Therefore, $x(t)$ can be calculated from discrete time points $x[n]$ by

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \cdot \frac{\sin \pi F_s(t - nT_s)}{\pi F_s (t - nT_s)} = \sum_{n=-\infty}^{\infty} x[n] \cdot \frac{\sin \pi(F_s t - n)}{\pi(F_s t - n)}. \tag{1.3}$$

Recalling that when k is an integer,

$$\frac{\sin(\pi k)}{\pi k} = \begin{cases} 1, & (k = 0) \\ 0. & \text{(otherwise)} \end{cases}$$

We see that $x(nT_s) = x[n]$ and that $x(t)$ connects these discrete time points smoothly. Eq. (1.3) is called interpolation formula.[1] Since $|\sin(\pi k)/\pi k|$ becomes smaller and smaller as $|k|$ goes larger and larger, we can cut-off summation with a finite number of terms to get approximated value of $x(t)$.

Noise generator If we set $a[n]$ at random with gaussian distribution,

$$x(t) = \sum_{n=-\infty}^{\infty} a[n] \cdot \frac{\sin \pi(Ft - n)}{\pi(Ft - n)}.$$

will give a band limited noise waveform within $\pm F/2$.

---

[1]Since $r(t)$ has finite value for $t < 0$, this filter is not causal.

Down sampling Similarly, if frequency component of $x(t)$ is concentrated within $kF_s - F_s/2 < f < kF_s + F_s/2$, where k is integer, we get

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \cdot e^{-i2\pi k(F_s t - n)} \cdot \frac{\sin \pi(F_s t - n)}{\pi(F_s t - n)}.$$

In this case, $x(t)$ can not be real, since frequency component of a real function has to be complex conjugate of that of frequency of opposite sign, but here, we only have frequency component on one side. To get real $x(t)$ we have to add image, (complex conjugate at the frequency of opposite sign). The result is

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \cdot \cos(2\pi k(F_s t - n)) \cdot \frac{\sin \pi(F_s t - n)}{\pi(F_s t - n)}.$$

1.8.2   Interpolation formula for periodic signal

In case $x(t)$ is periodic with period of T, i.e., $x(t + T) = x(t)$. It only has discrete frequency components:

$$x(t) = \frac{1}{T} \sum_{k=-\infty}^{\infty} X_c(k/T) \, e^{i\frac{2\pi k}{T}t}, \qquad X_c(k/T) = \int_{-T/2}^{T/2} x(t) \, e^{-i\frac{2\pi k}{T}t} \, dt.$$

Suppose that $X_c(k/T)$ is zero outside $-N/2 \le k \le N/2 - 1$,

$$x(t) = \frac{1}{T} \sum_{k=-N/2}^{N/2-1} X_c(k/T) \, e^{i\frac{2\pi k}{T}t}, \qquad X_c(k/T) = \int_{-T/2}^{T/2} x(t) \, e^{-i\frac{2\pi k}{T}t} \, dt.$$

Let's sample at $t = nT_s$, where $T = NT_s$, and use $\int dt \simeq \sum T_s$,

$$x(nT_s) = \frac{1}{T} \sum_{k=-N/2}^{N/2-1} X_c(k/T) \, e^{i\frac{2\pi k}{NT_s}nT_s}, \quad X_c(k/T) \simeq \sum_{-N/2}^{N/2-1} x(nT_s) \, e^{-i\frac{2\pi k}{NT_s}nT_s} \, T_s$$

Comparing this with discrete time Fourier transform,

$$x(nT_s) = \frac{1}{N} \sum_{k=-N/2}^{N/2-1} X(k) \, e^{i\frac{2\pi kn}{N}}, \qquad X(k) = \sum_{n=-N/2}^{N/2-1} x(nT_s) \, e^{-i\frac{2\pi kn}{N}},$$

we find,

$$\frac{1}{T} X_c(k/T) = \frac{1}{N} X(k) \quad \rightarrow \quad X_c(k/T) = T_s \, X(k) = \frac{1}{F_s} X(k),$$

21

where we are sampling coherently, i.e., $T = NT_s$. Therefore, $x(t)$ can be expressed with discrete time Fourier coefficients as

$$x(t) = \frac{T_s}{T} \sum_{k=-\infty}^{\infty} R(k/T) \, X(k) \, e^{i\frac{2\pi k}{T}t} = \frac{1}{F_s} \sum_{n=-N/2}^{N/2-1} x[n] \cdot r(t - nT_s),$$

where $R(f)$ is periodic ideal filter ($R(f) = 1$ for $|f| < F_s/2$, otherwise it is zero), and $r(t)$ is its Fourier transform:

$$R(k/T) = \int_{-T/2}^{T/2} r(t) \, e^{-i\frac{2\pi k}{T}t} \, dt, \qquad r(t) = \frac{1}{T} \sum_{k=-\infty}^{\infty} R(k/T) \, e^{i\frac{2\pi k}{T}t},$$

$$= \frac{1}{T} \sum_{k=-N/2}^{N/2-1} e^{i\frac{2\pi k}{T}t} = \frac{1}{T} \cdot \frac{\sin\frac{\pi N t}{T}}{\sin\frac{\pi t}{T}}.$$

This result matches Eq. (1.2) with $T \to \infty$ while keeping $N/T = F_s$ constant.

1.9　Power Spectral Density and Correlation

1.9.1　Power spectral density

As we have discussed earlier, if signal is infinite stream of data, it cannot have Fourier transform, since such signals are not square integrable, i.e., $\sum_n x^*[n]\,x[n]$ diverges. However, we know our signals have meaningful mean square. Let's start our discussion from mean square of N points.

$$\left\langle x^2 \right\rangle_N = \frac{1}{N} \sum_{n=-N/2}^{N/2-1} x^*[n]\,x[n]$$

For finite number of points, we have Fourier transform pair:

$$x[n] = \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} X_N(f)\,e^{i\frac{2\pi f}{F_s}n}\,df, \qquad X_N(f) = \sum_{n=-N/2}^{N/2-1} x[n]e^{-i\frac{2\pi f}{F_s}n}.$$

Inserting this into above yields

$$\left\langle x^2 \right\rangle_N = \frac{1}{N} \sum_{n=-N/2}^{N/2-1} \frac{1}{F_s^2} \int_{-F_s/2}^{F_s/2} \int_{-F_s/2}^{F_s/2} X_N^*(f')\,X_N(f)\,e^{i\frac{2\pi(-f'+f)}{F_s}n}\,df'df.$$

Recalling that $\sum_n e^{i\frac{2\pi f}{F_s}n}$ has sharp peak at $f = 0$ and it becomes delta function with $N \to \infty$:

$$\delta(f) = \frac{1}{F_s} \sum_{n=-\infty}^{\infty} e^{i\frac{2\pi f}{F_s}n},$$

we find

$$\lim_{N\to\infty} \left\langle x^2 \right\rangle_N = \lim_{N\to\infty} \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} \frac{X_N^*(f)\,X_N(f)}{N}\,df.$$

Let's take ensemble average, assuming time average is equal to ensemble average,

$$\left\langle x^2 \right\rangle = \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} \lim_{N\to\infty} \left\langle \frac{X_N^*(f)\,X_N(f)}{N} \right\rangle df = \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} S_x(f)\,df.$$

We call

$$S_x(f) = \lim_{N\to\infty} \left\langle \frac{X_N^*(f)\,X_N(f)}{N} \right\rangle$$

power spectral density. When we have meaningful mean square, we will have meaningful power spectral density. Note that $X_N^*(f)\,X_N(f)$ is $|X_N(f)|^2$, but we leave it as it is for later convenience. Since $x(t)$ is infinite stream, i.e., there is no distinct reference time point, we have lost phase information.

1.9.2 Autocorrelation

Let's calculate following quantity

$$\lim_{N \to \infty} \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} \frac{X_N^*(f) \, X_N(f)}{N} \, e^{i \frac{2\pi f}{F_s} m} \, df$$

$$= \lim_{N \to \infty} \frac{1}{N F_s} \int_{-F_s/2}^{F_s/2} \sum_{n,n'} x^*[n'] \, e^{i \frac{2\pi f}{F_s} n'} \, x[n] \, e^{-i \frac{2\pi f}{F_s} n} e^{i \frac{2\pi f}{F_s} m} \, df,$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{n,n'} x^*[n'] \, x[n] \, \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} e^{i \frac{2\pi f}{F_s}(m-n+n')} \, df,$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{n,n'} x^*[n'] \, x[n] \, \delta[m - n + n'],$$

$$= \lim_{N \to \infty} \frac{1}{N} \sum_{n} x^*[n - m] \, x[n],$$

$$= \lim_{N \to \infty} \langle x^*[n - m] \, x[n] \rangle_N \,,$$

where we have used

$$\delta[m] = \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} e^{i \frac{2\pi f}{F_s} m} \, df.$$

If we take ensemble average of this quantity, we see that Fourier transform of power spectral density is autocorrelation and we use $C_x[m]$ for it.

$$C_x[m] = \langle x[n] \, x^*[n - m] \rangle = \lim_{N \to \infty} \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} \left\langle \frac{X_N^*(f) \, X_N(f)}{N} \right\rangle e^{i \frac{2\pi f}{F_s} m} \, df,$$

$$= \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} S_x(f) \, e^{i \frac{2\pi f}{F_s} m} \, df.$$

Inversely, Fourier transform of autocorrelation is power spectral density.

$$S_x(f) = \sum_{m=-\infty}^{\infty} C_x[m] \, e^{-i \frac{2\pi f}{F_s} m}$$

$C_x[0]$ gives mean square by definition.

$$C_x[0] = \langle x^2 \rangle$$

In case $x[n]$ is real, $C_x[m]$ and $S_x(f)$ is even function,

$$C_x[m] = C_x[-m], \qquad S_x(-f) = S_x(f).$$

1.9.3 Correlation

Similarly, with

$$S_{xy}(f) = \lim_{N \to \infty} \left\langle \frac{Y_N^*(f) \, X_N(f)}{N} \right\rangle.$$

we can show follows.

$$C_{xy}[m] = \langle x[n] \, y^*[n-m] \rangle = \frac{1}{F_s} \int_{-F_s/2}^{F_s/2} S_{xy}(f) \, e^{i \frac{2\pi f}{F_s} m} \, df$$

$$S_{xy}(f) = \sum_{m=-\infty}^{\infty} C_{xy}[m] \, e^{-i \frac{2\pi f}{F_s} m}$$

1.9.4 Composition of power spectral density

Suppose that we are observing a signal $y[n]$ through a transfer function H and that spectral density of y is known. Let's say spectral density of y is $S_y(f)$,

$$S_y(f) = \lim_{N \to \infty} \left\langle \frac{Y_N^*(f) \, Y_N(f)}{N} \right\rangle,$$

and what we are observing is x, which is output of H.

$$X = H Y$$

Power spectral density $S_x(f)$ is

$$S_x(f) = \lim_{N \to \infty} \left\langle \frac{X_N^*(f) \, X_N(f)}{N} \right\rangle = \lim_{N \to \infty} \left\langle \frac{H^*(f) Y_N^*(f) \, H(f) Y_N(f)}{N} \right\rangle.$$

Since $H(f)$ is characteristic of the system and it is not subject to ensemble average, we can bring it out.

$$S_x(f) = |H(f)|^2 \lim_{N \to \infty} \left\langle \frac{Y_N^*(f) \, Y_N(f)}{N} \right\rangle = |H(f)|^2 \, S_y(f).$$

Similarly, when x is sum of two signals Y and Z which went through transfer function F and G.

$$X = F Y + G Z$$

Power spectral density $S_x(f)$ will be

$$S_x(f) = |F(f)|^2 \, S_y(f) + |G(f)|^2 \, S_z(f) + \text{Re} \left\{ F(f) G(f) \, S_{zy}(f) \right\}$$

In case Y and Z is independent, i.e., $S_{zy} = 0$,

$$S_x(f) = |F(f)|^2 \, S_y(f) + |G(f)|^2 \, S_z(f).$$

1.10 Delta Sigma Modulators

Let us consider an AD converter.

$$Y = X + E,$$

where X, Y, E is the analog input, the digital output and the quantization error, respectively. If we, somehow, could get quantization error of the previous sample and subtract it from the analog input, above becomes

$$Y = (X - z^{-1}E) + E \qquad \text{or} \qquad Y = X + (1 - z^{-1})E.$$
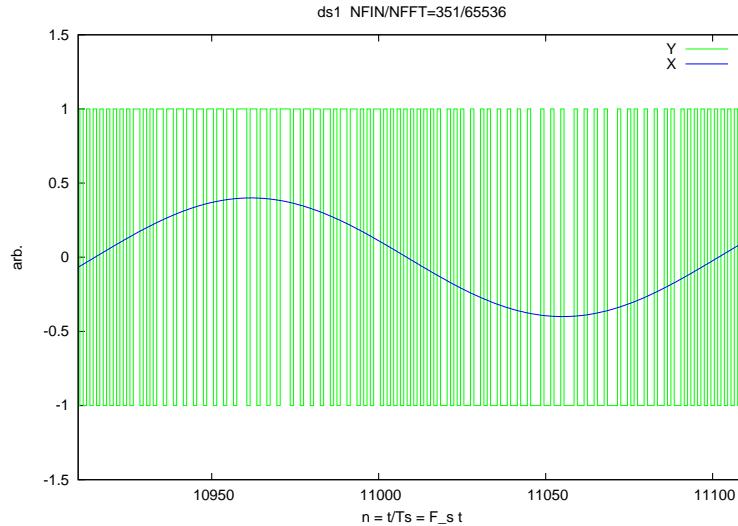
In the frequency domain, above becomes

$$Y(f) = X(f) + (1 - e^{-i\frac{2\pi f}{F_s}})E(f).$$

Suppose that E is white noise, i.e., no frequency dependence and no correlation to X,

$$|Y(f)|^2 = |X(f)|^2 + 4\sin^2\frac{\pi f}{F_s} \cdot |E|^2.$$

Quantization error is suppressed at low frequencies. A devastating fact is that even if Y is only 1bit, you can reproduce the input in digital domain. Below is a waveform of X and Y.

Y is vibrating rapidly as a result of high frequency component from E, low frequency component should be identical to X as we will see it shortly. This is called delta-sigma modulation. We place z^{-1} to the input for later convenience, z^{-1} by it self is just a one clock delay, it does not change waveform of the input X.

$$Y = z^{-1}X + (1 - z^{-1})E.$$

1.10.1 Noise shaping and oversampling

For L-th order, above would be

$$Y = z^{-L} X + (1 - z^{-1})^L E.$$

Let us rewrite above as follows.

$$Y = H_{\mathrm{x}} X + H_{\mathrm{e}} E, \qquad H_{\mathrm{x}} = z^{-L}, \quad H_{\mathrm{e}} = (1 - z^{-1})^L.$$

H_{x} is called signal transfer function. H_{e} is called noise transfer function. Frequency response of those are

$$H_{\mathrm{x}}(e^{i\frac{2\pi f}{F_s}}) = e^{-i\frac{2\pi f}{F_s}L}, \qquad H_{\mathrm{e}}(e^{i\frac{2\pi f}{F_s}}) = \left(1 - e^{-i\frac{2\pi f}{F_s}}\right)^L.$$

$$\left|H_{\mathrm{x}}(e^{i\frac{2\pi f}{F_s}})\right|^2 = 1 \qquad \left|H_{\mathrm{e}}(e^{i\frac{2\pi f}{F_s}})\right|^2 = \left|\left(1 - e^{-i\frac{2\pi f}{F_s}}\right)^L\right|^2 = \left(2\sin\frac{\pi f}{F_s}\right)^{2L}.$$

H_{x} changes only phase, but H_{e} is high-pass filter, low frequency component of quantization error is suppressed. $|H_{\mathrm{e}}|$ is proportional to f^{2L} for $f \ll F_s$, and it is 2^{2L} at $f = F_s/2$. If we limit signal within a narrow band, say $\pm F_h$, we can filter out quantization error.

Let's calculate available signal resolution. Here we assume that quantization error can be treated as white noise. Which means that it does not have correlation with the input, otherwise correlation term will show up in the power spectral density of the output, and that power spectral density of quantization error $S_e(f)$ is flat. Mean square of quantization error $\langle e^2 \rangle$ is $\triangle^2/12$, where $\triangle$ is least significant bit of the quantizer and its power spectral density is

$$S_e(f) = \frac{\triangle^2}{12} \cdot \frac{1}{F_s}.$$

Power of quantization error shows up at y within $\pm F_h$ is calculated as follows.

$$P_{\mathrm{e}} = \int_{-F_h}^{F_h} \left|H_{\mathrm{e}}(e^{i\frac{2\pi f}{F_s}})\right|^2 S_e(f)\, df,$$

$$= \int_{-F_h}^{F_h} \left|\left(1 - e^{-i\frac{2\pi f}{F_s}}\right)^L\right|^2 df \cdot \frac{\triangle^2}{12} \cdot \frac{1}{F_s},$$

$$= \int_{-F_h}^{F_h} (2\sin(\pi f/F_s))^{2L}\, df \cdot \frac{\triangle^2}{12} \cdot \frac{1}{F_s}.$$

Using the fact that $F_h \ll F_s$ and that $\sin x \sim x$ for $x \ll 1$ yields

$$P_{\mathrm{e}} \sim \int_{-F_h}^{F_h} (2\pi f/F_s))^{2L}\, df \cdot \frac{\triangle^2}{12} \cdot \frac{1}{F_s} = \frac{\triangle^2}{6} \cdot \frac{1}{2\pi} \cdot \frac{(2\pi F_h/F_s)^{2L+1}}{2L+1}.$$

Ratio between the width of the modulator's nyquist band ($\pm F_s/2$) and the width of this limited band ($\pm F_h$) is called oversampling ratio $M = F_s/2F_h$. With oversampling ratio M, above becomes
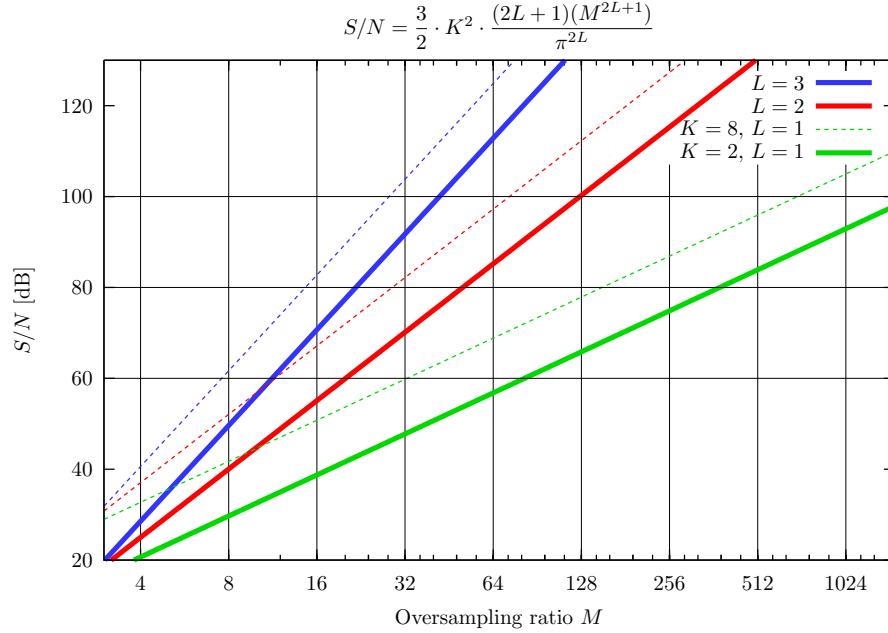
$$P_\text{e} \ \sim \ \frac{\triangle^2}{12} \cdot \frac{\pi^{2L}}{2L+1} \cdot \frac{1}{M^{2L+1}}.$$

Recalling that full swing sinusoidal signal power, P_s, is

$$P_\text{s} = \left(\frac{\triangle \cdot K}{2\sqrt{2}}\right)^2,$$
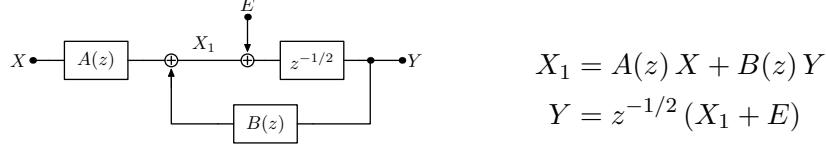
where K is number of quantization levels and $\triangle \cdot K$ is full-scale of the quantizer, signal to noise ratio is calculated to be

$$\text{Signal to noise ratio} \ = \ \frac{P_\text{s}}{P_\text{e}} \ = \ \frac{3}{2} \cdot K^2 \cdot \frac{(2L+1)(M^{2L+1})}{\pi^{2L}}.$$

$$S/N = \frac{3}{2} \cdot K^2 \cdot \frac{(2L+1)(M^{2L+1})}{\pi^{2L}}$$

1.10.2 Switched capacitor implementation

Now we would like to implement this modulator using switched capacitor circuits, namely an AD converter and delayed integrators. First, Y is digital output, it must be the output of an ADC and we would connect it to X and Y though response function A and B like shown below.

$$X_1 = A(z)\, X + B(z)\, Y$$
$$Y = z^{-1/2}\,(X_1 + E)$$

Eliminating X_1 yields

$$Y \;=\; \frac{z^{-1/2}\, A(x)\, X}{1 - z^{-1/2}\, B(z)} + \frac{z^{-1/2}\, E}{1 - z^{-1/2}\, B(z)}.$$

Here $z^{-1/2}E$ is the quantization noise half clock cycle prior to Y, i.e., $z^{-1/2}E$ is the quantization noise actually injected. Therefore noise transfer function is the coefficient of $z^{-1/2}E$, and we want it to be $(1 - z^{-1})^L$.

$$\frac{1}{1 - z^{-1/2}\, B(z)} = (1 - z^{-1})^L.$$

Therefore signal transfer function, coefficient of X, will be

$$H_\mathrm{x}(z) = z^{-1/2}(1 - z^{-1})^L\, A(z),$$

and we want it to be z^{-L}, (or at least $|H_\mathrm{x}(z)| = 1$). Therefore

$$A(z) = \frac{z^{1/2} z^{-L}}{(1 - z^{-1})^L}, \qquad B(z) = z^{1/2}\left(1 - \frac{1}{(1 - z^{-1})^L}\right).$$

and

$$X_1 \;=\; \frac{z^{1/2} z^{-L}}{(1 - z^{-1})^L}\, X + z^{1/2}\left(1 - \frac{1}{(1 - z^{-1})^L}\right) Y.$$

First order modulator   Inserting $L = 1$ yields,

$$X_1 = \frac{z^{-1/2}}{1 - z^{-1}}\,(X - Y).$$

Second order modulator   Similarly, inserting $L = 2$ yields,

$$X_1 = \frac{z^{-1/2}\, z^{-1}}{(1 - z^{-1})^2}\, X + \frac{-2z^{-1/2} + z^{-1/2}z^{-1}}{(1 - z^{-1})^2}\, Y.$$

29

From here factor out delayed integrator one after another to yield feedback equation:

$$X_1 = \frac{z^{-1/2}}{1 - z^{-1}}\left(\frac{z^{-1}}{1 - z^{-1}}X + \frac{-2 + z^{-1}}{1 - z^{-1}}Y\right),$$

$$= \frac{z^{-1/2}}{1 - z^{-1}}\left(\frac{z^{-1/2}}{1 - z^{-1}}\left(z^{-1/2}X + z^{-1/2}Y\right) - \frac{2}{1 - z^{-1}}Y\right),$$

$$= \frac{z^{-1/2}}{1 - z^{-1}}\left(\frac{z^{-1/2}}{1 - z^{-1}}\left(z^{-1/2}X - z^{-1/2}Y\right) - 2Y\right),$$

where we used,

$$\frac{1}{1 - z^{-1}} = 1 + \frac{z^{-1}}{1 - z^{-1}}$$

to get the last expression. This can be

$$X_3 = z^{-1/2}X, \qquad\qquad \text{(Sample and hold)}$$

$$X_2 = \frac{z^{-1/2}}{1 - z^{-1}}(X_3 - Y_1), \qquad \text{(1st stage)}$$

$$X_1 = \frac{z^{-1/2}}{1 - z^{-1}}(X_2 - 2Y), \qquad \text{(2nd stage)}$$

$$Y_1 = z^{-1/2}Y. \qquad\qquad \text{(Latch)}$$

We have a S/H for X and a latch for Y before going to the first stage integrator, which uses Y at ϕ_1.

Another possibility is

$$X_1 = \frac{z^{-1/2}}{1 - z^{-1}}\left(\frac{z^{-1/2}}{1 - z^{-1}}z^{-1/2}X - \frac{1}{1 - z^{-1}}z^{-1}Y - 2Y\right),$$

or

$$X_3 = z^{-1/2}X, \qquad\qquad\qquad \text{(Sample and hold)}$$

$$X_2 = \frac{z^{-1/2}}{1 - z^{-1}}X_3 - \frac{1}{1 - z^{-1}}Y_2, \qquad \text{(1st stage)}$$

$$X_1 = \frac{z^{-1/2}}{1 - z^{-1}}(X_2 - 2Y),$$

$$= \frac{z^{-1/2}}{1 - z^{-1}}X_2 - \frac{1}{1 - z^{-1}}\cdot 2Y_1, \qquad \text{(2nd stage)}$$

$$Y_1 = z^{-1/2}Y, \qquad\qquad\qquad \text{(Latch)}$$

$$Y_2 = z^{-1/2}Y_1, \qquad\qquad\qquad \text{(Latch)}$$

30
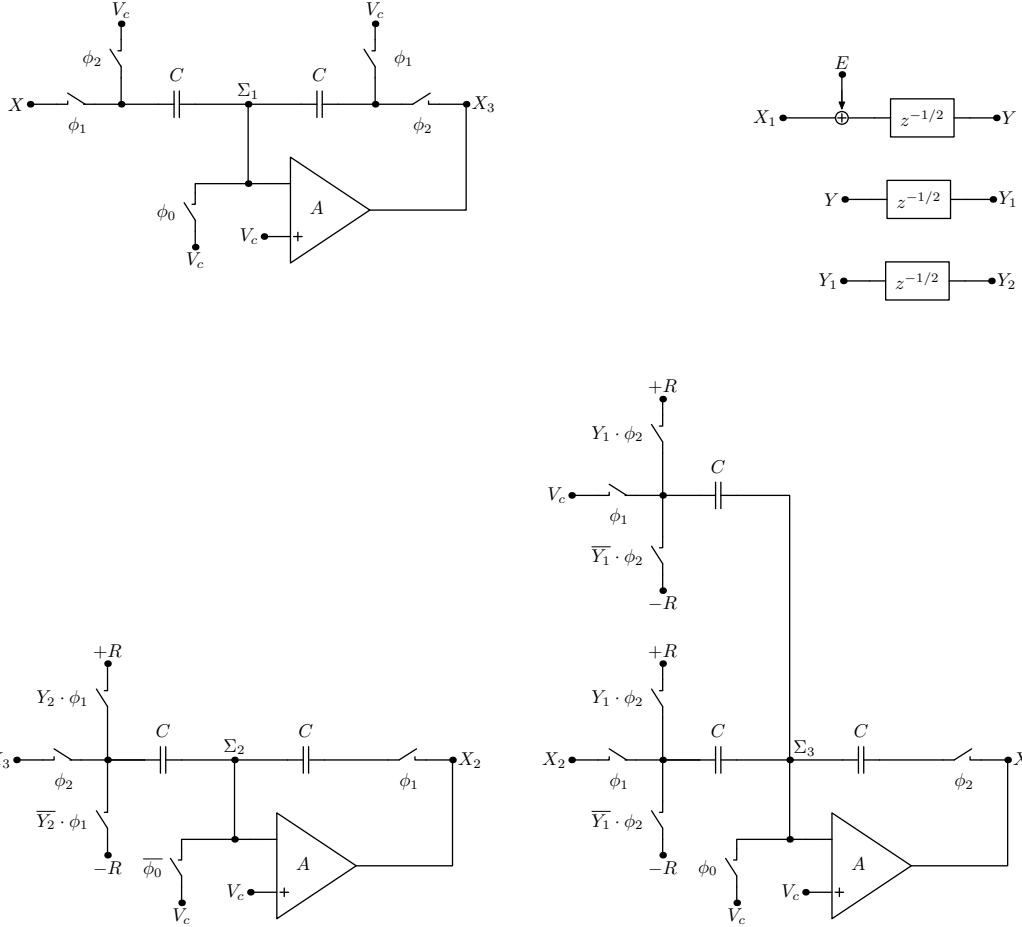
Figure 1: A switched capacitor implementation of second order delta-sigma modulator. X_1 is followed by a 1bit ADC of which output is Y. Y_1 and Y_2 is half clock and full clock delay of Y, respectively.

Circuit diagrams are shown in Figure 1. X_1 is followed by an ADC and its output is Y. Node Y has valid data on ϕ_1. Y_1 and Y_2 is half clock cycle delay and full clock cycle delay of Y, respectively.

Yet another possibility is

$$X_3 = z^{-1/2} X, \qquad\qquad\qquad\qquad \text{(Sample and hold)}$$

$$\frac{1}{2} X_2 = \frac{1}{2} \left(\frac{z^{-1/2}}{1 - z^{-1}} X_3 - \frac{1}{1 - z^{-1}} Y_2 \right), \qquad \text{(1st stage)}$$

$$X_1 = \frac{z^{-1/2}}{1 - z^{-1}} \left(X_2 - 2\,Y \right),$$

$$\qquad = 2 \left(\frac{z^{-1/2}}{1 - z^{-1}} \cdot \frac{1}{2} X_2 - \frac{1}{1 - z^{-1}} \cdot Y_1 \right), \qquad \text{(2nd stage)}$$

$$Y_1 = z^{-1/2} Y, \qquad\qquad\qquad\qquad \text{(Latch)}$$

$$Y_2 = z^{-1/2} Y_1, \qquad\qquad\qquad\qquad \text{(Latch)}$$

Circuit diagrams are shown in Figure 2. Switch network gets simplified compared to the first one while signal swing of the second stage gets half. Reduced swing potentially impacts thermal noise performance (power consumption), while we'd expect better linearity of the second stage output. Choice of two implementation would depend on external requirements (or priority).

1.10.3 Simulation

In the noise shaping calculation, we have assumed quantization error can be treated as white noise without basis. In fact this assumption is only reasonable when we have many levels of quantization. Apparently when our ADC has only few quantization levels there will be good correlation between the input of the ADC and the quantization error. Here we simulate the first and the second order modulator with only two levels of quantization ($K = 2$, or 1bit). We will see that there still be large correlation in the first order modulator. However it gets much less in the second order modulator.
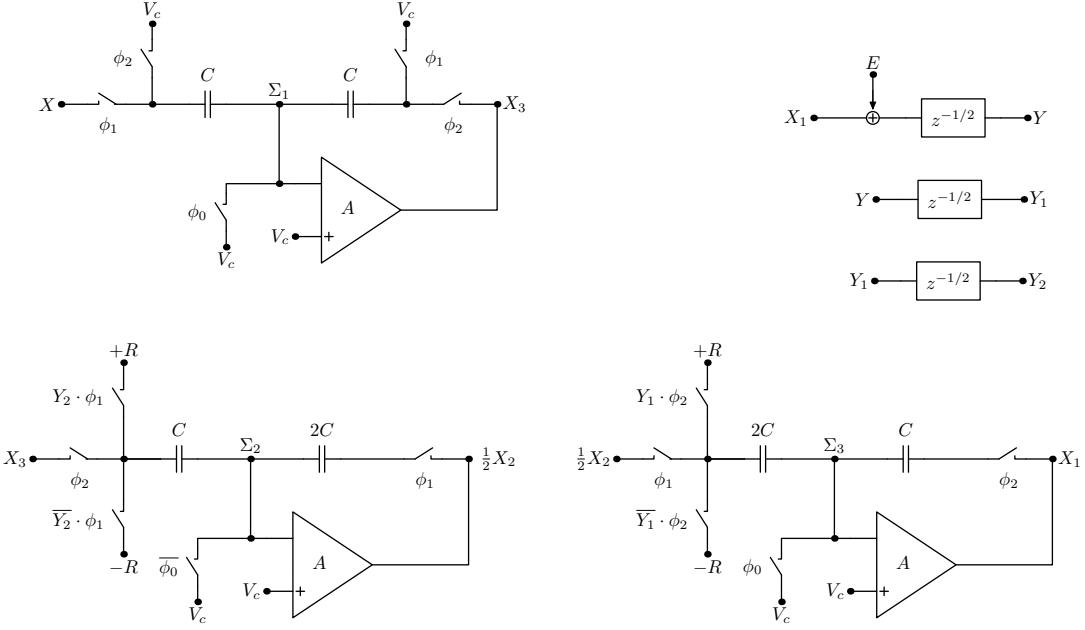
Figure 2: Second possibility of switched capacitor implementation of second order delta-sigma modulator. X_1 is followed by a 1bit ADC of which output is Y. Y_1 and Y_2 is half clock and full clock delay of Y, respectively. Switch network gets simplified compared to the first one while signal swing of the second stage gets half. Reduced swing potentially impacts thermal noise performance (power consumption), while we'd expect better linearity of the second stage output. Choice of two implementation would depend on external requirements (or priority).
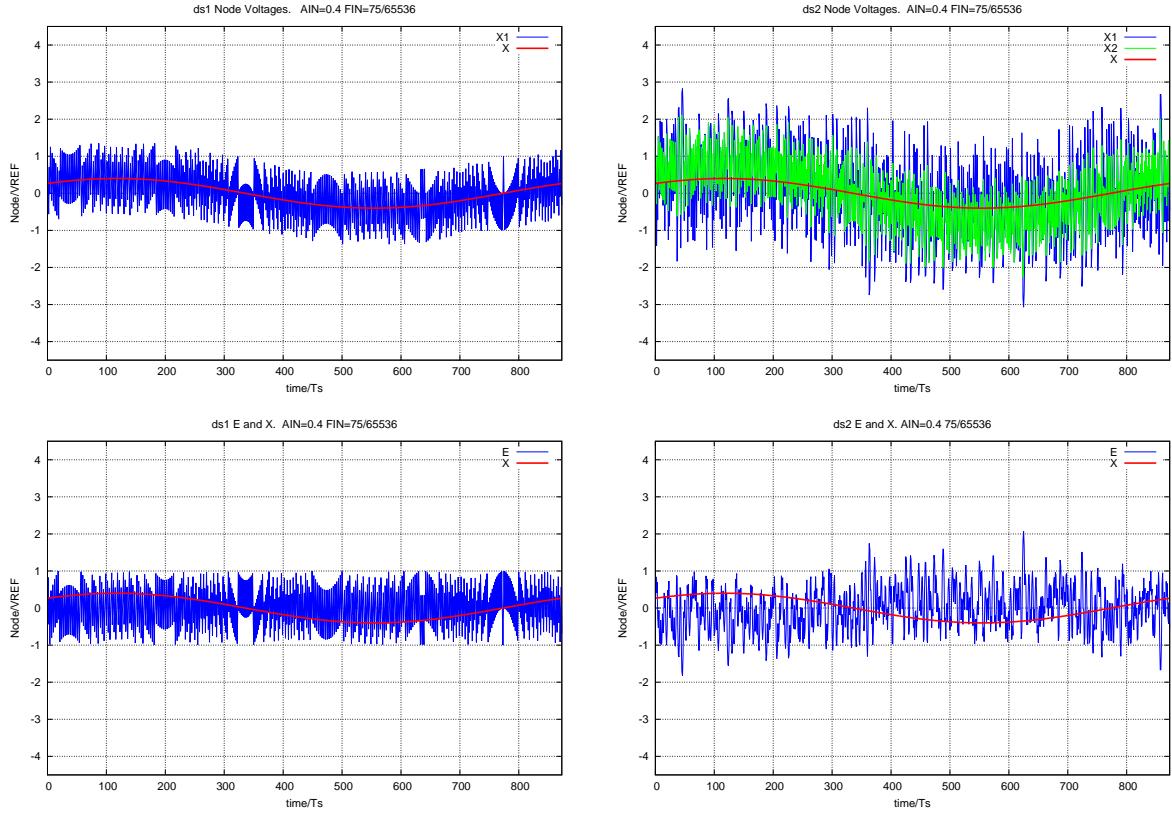
Figure 3: Simulated waveform for first order (left, ds1) and second order (right, ds2) delta-sigma modulators. Top plots are node voltages, $x[n]$, $x_1[n]$, etc. Bottom plots are $e[n]$. With the same input swing and reference level, ds2's node voltages swing much wider than that of ds1. $e[n]$ is vibrating rapidly for both cases. However there's obvious harmonic content in ds1, while ds2 has fairly randomized error with obvious fundamental content.

34

Power Spectral Density ds1 AIN=0.4 FIN=75/65536Fs 1 sample.

f^{2L}

$S_y(f)$ [dB]

0
-20
-40
-60
-80
-100
-120
-140
-160
-180

1/2048 FIN 1/512 1/256 1/128 1/64 1/32 1/16   1/2

f/F_s

Power Spectral Density ds2 AIN=0.4 FIN=75/65536Fs 1 sample.

f^{2L}

$S_y(f)$ [dB]

0
-20
-40
-60
-80
-100
-120
-140
-160
-180

1/2048 FIN 1/512 1/256 1/128 1/64 1/32 1/16   1/2

f/F_s

Power Spectral Density ds1 AIN=0.4 FIN=75/65536Fs 100 samples.

f^{2L}

$S_y(f)$ [dB]

0
-20
-40
-60
-80
-100
-120
-140
-160
-180

1/2048 FIN 1/512 1/256 1/128 1/64 1/32 1/16   1/2

f/F_s

Power Spectral Density ds2 AIN=0.4 FIN=75/65536Fs 100 samples.

f^{2L}

$S_y(f)$ [dB]

0
-20
-40
-60
-80
-100
-120
-140
-160
-180

1/2048 FIN 1/512 1/256 1/128 1/64 1/32 1/16   1/2

f/F_s

Figure 4: Power spectral density $S_y(f)$ for first order (left, ds1) and second order (right, ds2) delta-sigma modulators. Top plots are single shot input. Bottom plots are average of 100 Monte Carlo samples (phase of input sine wave is randomly chosen). While ds2 shows tight and nice f^{2L} behavior in the bottom plot, ds1 still has a lot of grasses, which suggests strong correlation between input X and quantization error E.

Appendix A Useful Formula

Parallel impedance operator

$$(r_1//r_2) = \frac{1}{1/r_1 + 1/r_2} = \frac{r_1 r_2}{r_1 + r_2}, \quad (r_1//r_2//r_3) = \frac{1}{1/r_1 + 1/r_2 + 1/r_3} = \frac{r_1 r_2 r_3}{r_1 + r_2 + r_3}$$

$$(r_1//r_2) = (r_2//r_1), \quad \frac{1}{(r_1//r_2)} + \frac{1}{r_3} = \frac{1}{(r_1//r_2//r_3)}, \quad (r_1//c_1) = \frac{r_1}{1 + s\, r_1 c_1}$$

Minimum value

$$\min\left(\frac{A}{x} + Bx\right) = 2\sqrt{AB}, \qquad x_{\min} = \sqrt{\frac{A}{B}}.$$

Integral

$$\int_0^\infty \frac{d\omega/2\pi}{1 + (\omega/\omega_0)^2} = \frac{\omega_0}{2\pi}\tan^{-1}\left(\frac{\omega}{\omega_0}\right)\Big|_0^\infty = \frac{1}{4}\omega_0$$

$$\int_{\omega_s}^\infty \frac{d\omega}{\omega(1 + (\omega/\omega_0)^2)} = \frac{1}{2}\ln\frac{(\omega/\omega_0)^2}{1 + (\omega/\omega_0)^2}\Big|_{\omega_s} = \frac{1}{2}\ln\frac{1 + (\omega_s/\omega_0)^2}{(\omega_s/\omega_0)^2}$$

Two pole response function and its impulse response

$$v_o/v_i = \frac{1}{1 + s\,b + s^2\,a} = \frac{1}{(1 + s\,\tau_\oplus)(1 + s\,\tau_\ominus)} \quad (a > 0,\ b > 0)$$

$$1/\tau_{\oplus,\ominus} = \frac{b \pm \sqrt{b^2 - 4a}}{2a} = \frac{b}{2a}\left(1 \pm \sqrt{1 - 4a/b^2}\right)$$

Discriminant $4a/b^2$:

$$
\begin{aligned}
4a/b^2 < 1 &\to \quad \text{Exponential settling (overshooting)} \\
= 1 &\to \quad \text{Critical damping} \\
> 1 &\to \quad \text{Ringing}
\end{aligned}
$$

If $4a/b^2 \ll 1$,

$$1/\tau_\oplus = b/a - 1/b, \qquad 1/\tau_\ominus = 1/b$$

Canonical form of two pole amplifier

$$A(s) = \frac{N}{Q + s\,B + s^2\,A} = \frac{A_0}{(1 + s\,\tau_A A_0)(1 + s\,\tau_\oplus)}$$

If $4AQ/B^2 \ll 1$:

$$A_0 = N/Q, \quad \tau_A = B/N, \quad 1/\tau_\oplus = B/A - Q/B$$

Laplace transform

$$\mathcal{L}\{\delta(t)\} = 1, \qquad \mathcal{L}\{1\} = \frac{1}{s}, \qquad \mathcal{L}\{t\} = \frac{1}{s^2}, \qquad \mathcal{L}\{e^{-t/\tau_1}\} = \frac{\tau_1}{1 + s\,\tau_1}$$

$$\mathcal{L}\{t/\tau_1\,e^{-t/\tau_1}\} = \frac{\tau_1}{(1 + s\,\tau_1)^2}$$

$$\frac{1}{(1 + s\,\tau_1)(1 + s\,\tau_2)} = \frac{1}{\tau_1 - \tau_2}\left(\frac{\tau_1}{1 + s\,\tau_1} - \frac{\tau_2}{1 + s\,\tau_2}\right)$$

$$\frac{s}{(1 + s\,\tau_1)(1 + s\,\tau_2)} = -\frac{1}{\tau_1 - \tau_2}\left(\frac{1}{\tau_1}\cdot\frac{\tau_1}{1 + s\,\tau_1} - \frac{1}{\tau_2}\cdot\frac{\tau_2}{1 + s\,\tau_2}\right)$$

$$\frac{1 + s\,\tau_3}{(1 + s\,\tau_1)(1 + s\,\tau_2)} = \frac{1}{\tau_1 - \tau_2}\left(\frac{\tau_1 - \tau_3}{\tau_1}\cdot\frac{\tau_1}{1 + s\,\tau_1} - \frac{\tau_2 - \tau_3}{\tau_2}\cdot\frac{\tau_2}{1 + s\,\tau_2}\right)$$

$$\frac{s}{(1 + s\,\tau_1)^2} = \frac{1}{\tau_1^2}\left(\frac{\tau_1}{1 + s\,\tau_1} - \frac{\tau_1}{(1 + s\,\tau_1)^2}\right)$$

$$\frac{1}{s\,(1 + s\,\tau_1)} = \frac{1}{s} - \frac{\tau_1}{1 + s\,\tau_1}$$

$$\frac{1}{s^2\,(1 + s\,\tau_1)} = \frac{1}{s^2} - \frac{\tau_1}{s} + \frac{\tau_1^2}{1 + s\,\tau_1}$$

$$\frac{1}{s\,(1 + s\,\tau_1)(1 + s\,\tau_2)} = \frac{1}{s} - \frac{\tau_1}{\tau_1 - \tau_2}\cdot\frac{\tau_1}{1 + s\,\tau_1} + \frac{\tau_2}{\tau_1 - \tau_2}\cdot\frac{\tau_2}{1 + s\,\tau_2}$$

$$\frac{1}{s\,(1 + s\,\tau_1)^2} = \frac{1}{s} - \frac{\tau_1}{1 + s\,\tau_1} - \frac{\tau_1}{(1 + s\,\tau_1)^2}$$

$$\frac{1 + s\,\tau_3}{s\,(1 + s\,\tau_1)(1 + s\,\tau_2)} = \frac{1}{s} - \frac{\tau_1 - \tau_3}{\tau_1 - \tau_2}\cdot\frac{\tau_1}{1 + s\,\tau_1} + \frac{\tau_2 - \tau_3}{\tau_1 - \tau_2}\cdot\frac{\tau_2}{1 + s\,\tau_2}$$

Approximation If $\tau_1 \gg \tau_2$,

$$\frac{s}{(1 + s\,\tau_1)(1 + s\,\tau_2)} \sim \frac{1}{\tau_1\tau_2}\left(\frac{\tau_2}{1 + s\,\tau_2} - \frac{\tau_2}{\tau_1}\cdot\frac{\tau_1}{1 + s\,\tau_1}\right)$$

$$\frac{1}{s\,(1 + s\,\tau_1)(1 + s\,\tau_2)} \sim \frac{1}{s} - \frac{\tau_1}{1 + s\,\tau_1} + \frac{\tau_2}{\tau_1}\cdot\frac{\tau_2}{1 + s\,\tau_2}$$

References

[1] C. E. Shannon, *A Mathematical Theory of Communication*, The Bell System Technical Journal **27** (1948), 379–423, 623–656.