# Advanced AD/DA converters

## First-Order ΔΣ Modulator (MOD1)

**Pietro Andreani**
Dept. of Electrical and Information Technology
Lund University, Sweden

# Overview

- Basic concepts
- 1st-order ΔΣ modulator
- Idle tones, stability
- Finite opamp gain, dead zones
- Decimator filter

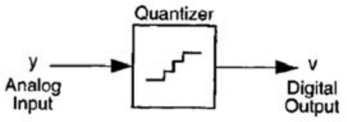
# A/D conversion – basic concepts

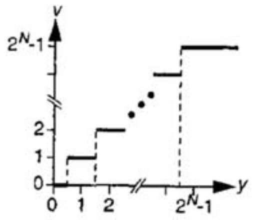
Continuous-time analog to discrete-time digital → 1) analog signal is sampled (with a period T, usually constant), and 2) the resulting signal is eventually quantized (→ assumes one of a finite number of values)

Quantization is usually (but not always) uniform → two adjacent levels differ by a quantity Δ

Quantization is performed by a quantizer (ideal A/D converter), assumed memoryless → static input-output *y*-*v* curve

Here: only positive input values, unipolar quantizer, Δ =1

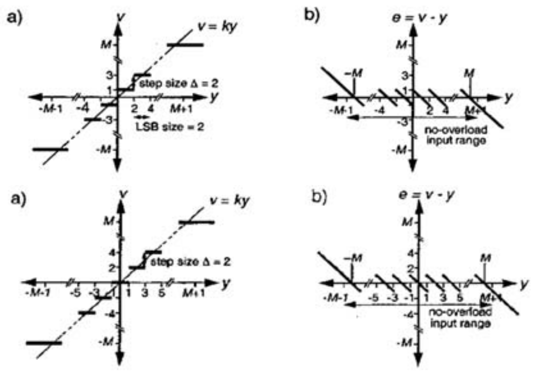




# A/D conversion – basic concepts

Below → bipolar quantizers with associated error functions (e = v-y)
top: a step occurs for y=0; bottom: y=0 is in the middle of a step; in both cases, Δ=2 → q-levels are integer in both cases (odd in the first, even in the second)

T=1 and Δ=2 are assumed in the following (Schreier's book)



# A/D conversion – basic concepts

As long as *y* is between –(M+1) and (M+1) the error is between 1 and -1 → this is the (no-overload) input range

The difference between maximum and minimum level is the full scale (FS) of the quantizer; the table summarizes other notable properties

| Parameter | Value |
|---|---|
| input step size (LSB size) | 2 |
| output step size | 2 |
| number of steps | $M$ |
| number of levels | $M+1$ |
| number of bits | $\lceil \log_2(M+1) \rceil$ |
| no-overload input range | $[-(M+1), M+1]$ |
| full-scale | $2M$ |
| input thresholds | $0, \pm 2, \ldots, \pm(M-1)$, $M$ odd<br>$\pm 1, \pm 3, \ldots, \pm(M-1)$, $M$ even |
| output levels | $\pm 1, \pm 3, \ldots, \pm M$, $M$ odd<br>$0, \pm 2, \pm 4, \ldots, \pm M$, $M$ even |

# A/D conversion – basic concepts

Quantizer is deterministic → *v* and *e* are fully determined by *y*

However, if *y* stays within the FS and changes by sufficiently large amounts from sample to sample, its position within a quantization interval is essentially random → *e* is assumed to be a white noise process with samples uniformly distributed between –Δ/2 and Δ/2 → *e* has zero mean, and power $\sigma_e^2 = \Delta^2/12$

This noise is simply added to the scaled input, to give $v = ky + e$

In fact, this equation is always true → the approximation is in the assumed nature of *e*

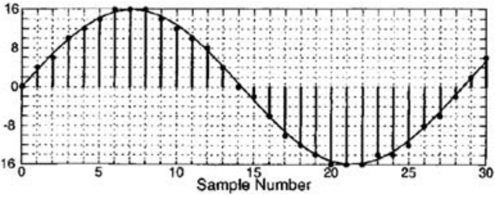
A white-noise-like error *e* is not produced in a number of cases, such as with a constant input, or a periodic input harmonically related to T (even worse if *y* changes from sample to sample by rational fractions of Δ)

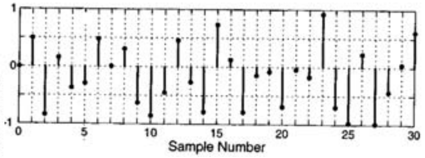
# Quantization error

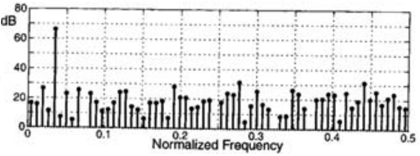
Below: full scale sinusoid sampled by a 16-step quantizer → sampling frequency is relatively high, and in no simple relation to signal frequency → q-errors appear to be quite random (even if in reality they are not random (i.e. uncorrelated) at the peaks of the sinusoid)

The mean value of the error is 0.30, which is very close to the white-noise-approximation expected value of $2^2/12 = 1/3$

The Fourier transform also shows a white-noise-like noise floor

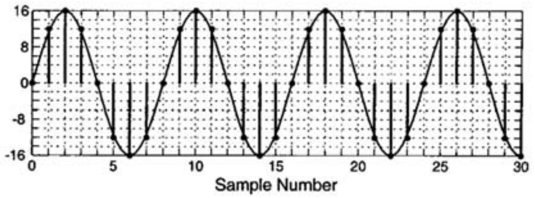


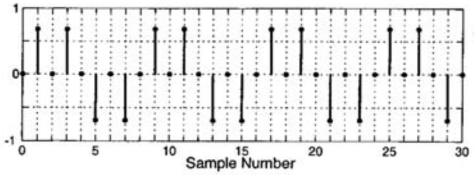


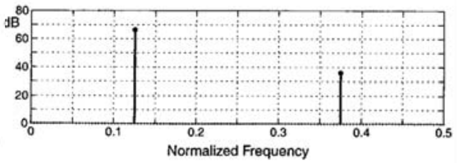


# Quantization error

Below: full scale sinusoid sampled at 8 times its own frequency → q-error is periodic and assumes only 3 values → far from uniformly distributed. The mean value of the error is only 0.23, and the FFT shows only two spectral components, fundamental and 3rd harmonic!

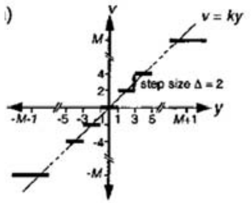




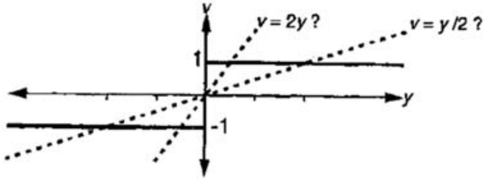


# Quantization error

As clear from the figure below, the gain *k* of the linear model for a multi-bit quantizer is given by the ratio of the step size to the distance between adjacent thresholds



In a single-bit quantizer, however, this is not possible, since there is only one threshold! Below: the two lines are both possible, both resulting in the same maximum error of Δ/2, although with different no-overload ranges



# Quantization error

If the statistical properties of the input *y* are known, a criterion for determining *k* is to minimize the mean square value (i.e., the power) of the error sequence *e*; this value is defined as the expected value of *e*²:

$$\sigma_e^2 = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{\infty} e^2(n) \equiv \langle e,e\rangle \quad \text{, with} \qquad \langle a,b\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{\infty} a(n)b(n) = E[a,b]$$

$$\sigma_e^2 = \langle e,e\rangle = \langle v-ky, v-ky\rangle = \langle v,v\rangle - 2k\langle v,y\rangle + k^2\langle y,y\rangle$$

This is minimized by

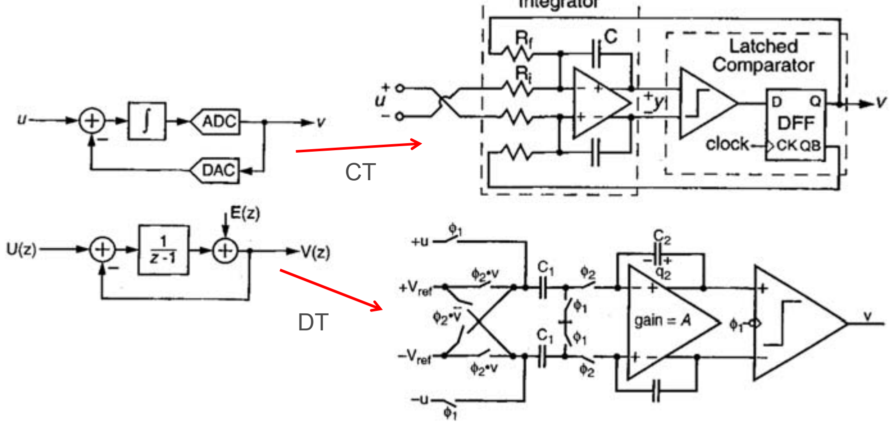
$$k = \frac{\langle v,y\rangle}{\langle y,y\rangle} = \frac{\langle \mathrm{sgn}[(y)],y\rangle}{\langle y,y\rangle} = \frac{E[|y|]}{E[y^2]}$$

Sanity check: if some *k* corresponds to a given *y*, then if we define *y'*=10*y,* it is immediate to check the *k'*=*k*/10, which makes sense since the output does not change. When a system containing a binary quantizer is replaced by a linear model, the estimate of the quantizer gain *k* should be found from extensive simulations; otherwise, misleading results may follow.

# MOD1 modulator

Simplest ΔΣ modulator, MOD1, either analog or digital → one integrator, one 1-b DAC, and one 1-b DAC → transforms continuous or finely-quantized input into coarsely-quantized output with noise-shaped spectrum

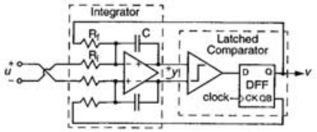
Here: CT and DT implementations

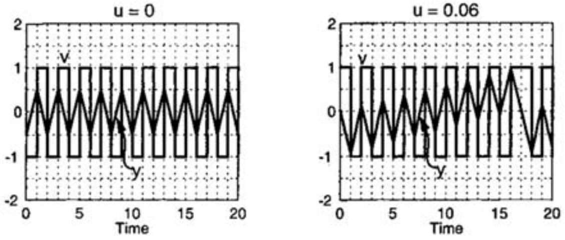


# MOD1 as A/D converter

Below: *u*=0 → output toggles between +1 and -1; *u*=0.06 → output toggles until the integrator output is so high that two adjacent +1 are produced

The circuit forces the average of *v*/R$_f$ to be equal to the average of *u*/R$_i$

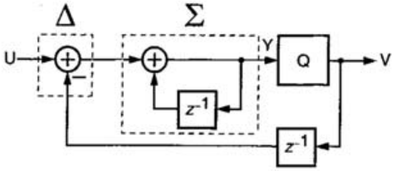




## MOD1

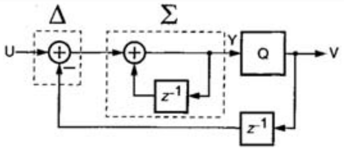
Remarkable feature: for a constant input, arbitrarily high accuracy can be achieved (in principle!). Further: a common source of non-linearity in Nyquist converters is component mismatch – here not an issue!! In fact, changing the integrating capacitor in the CT modulator, or the ratio $C_1/C_2$ in the DT modulator, simply scales the output, but does not change its sign, which is the only thing the 1-b ADC detects! Similarly, comparator hysteresis, opamp offset and component mismatch have no impact on the linearity. At worst, the input full-range scale will shift and the input-referred offset will be non-zero.

These qualitative results can be retrieved by a simple analysis on the equivalent circuit below



## MOD1 analysis

$$v(n)=\text{sgn}\left[y(n)\right]$$
$$y(n)=y(n-1)+u(n)-v(n-1)$$



Thus, combining the equations for n=0,1,2,…N, we have

$$y(N)-y(0)=\sum_{n=0}^{N}\left[u(n)-v(n-1)\right]$$

If $y$ is bound, we have $\lim_{N\to\infty}\frac{y(N)-y(0)}{N}=0$ , which from the above equation implies that the average of the input samples $u$ is equal to the average of the digital output $v$. The average of $v$ is recovered by cascading a digital low-pass filter to the modulator

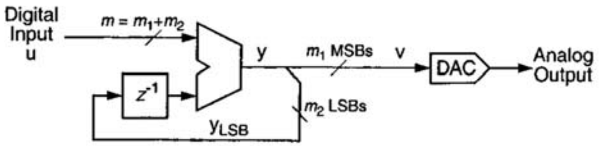
MOD1 uses a 2-level quantizer → DAC capable of perfect linearity – in general, 2nd-order effects such as dependence of quantizer threshold and/or reference and/or power supply on input signal introduce some non-linearity

## MOD1 in a D/A converter

Example: we want to drive an 8-bit DAC with a 16-bit data stream → one could discard the 8 LSBs, with an awful loss of resolution

Better → correct the truncation error in the current sample by introducing the opposite error in the next sample → truncation averaged over time!

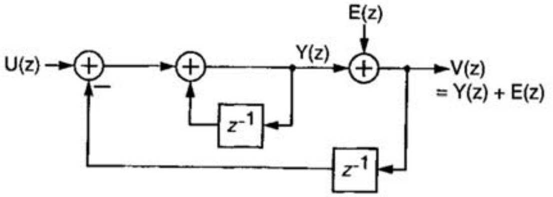
This approach is shown in the 1st-order digital modulator here, used in a ΔΣ DAC!



$$y(n)=u(n)+y_{LSB}(n-1);\qquad y_{LSB}(n-1)=y(n-1)-v(n-1)$$
$$y(n)=u(n)+y(n-1)-v(n-1)$$

This is the MOD1 equation of the previous slide → m-bit resolution can be preserved! It is also easy to show that the accumulated difference between input and output is always less than one DAC LSB

## MOD1 in the z-domain



$$y(n)=u(n)+y(n-1)-v(n-1)\quad\rightarrow\quad Y(z)=z^{-1}Y(z)+U(z)-z^{-1}V(z)$$

$$\begin{aligned}V(z)&=Y(z)+E(z)=z^{-1}Y(z)+U(z)-z^{-1}V(z)+E(z)\\&=U(z)+E(z)-z^{-1}\left[V(z)-Y(z)\right]\\&=U(z)+E(z)-z^{-1}E(z)\\&=U(z)+\left(1-z^{-1}\right)E(z)\end{aligned}$$

DC value for z=1: if E is limited, we recover the fact that V and U have the same DC value → arbitrarily high resolution at DC

# MOD1 in the z-domain

The equation $V(z)=U(z)+\left(1-z^{-1}\right)E(z)$ can be written in the following general form:

$$V(z)=STF(z)U(z)+NTF(z)E(z)$$

where the **STF** is the signal transfer function, and **NTF** the noise transfer function

To find the in-band power of the q-noise, we evaluate the NTF for $z=e^{j2\pi f}$

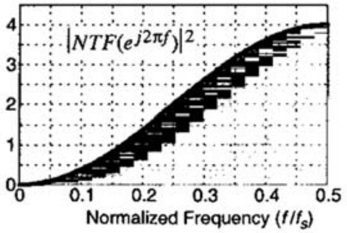
At low frequencies (high OSR) we have

$$\left|NTF\left(e^{j2\pi f}\right)\right|^2=\left|1-e^{-j2\pi f}\right|^2=\left|e^{-j\pi f}\left(e^{j\pi f}-e^{-j\pi f}\right)\right|^2=\left|2\sin(\pi f)\right|^2\approx(2\pi f)^2$$

High-pass response → q-noise suppressed at and near DC (where we have the input signal) and amplified out-of-band (at or near $0.5f_s$)

This noise-shaping action is crucial for the effectiveness of the ΔΣ approach

# MOD1 in the z-domain



Quantizer error from internal ADC is white → its power is $\Delta^2/12$ = 1/3, and its power is found between DC and Nyquist = $f_s/2$ (=0.5 normalized to the sampling frequency) → thus, the spectral power density is

$$S_e(f)=\frac{\Delta^2/12}{1/2}=\frac{2}{3}$$

# SQNR

Thus, the in-band noise power of the output *v* is given by

$$\sigma_q^2=\int_0^{f_B}(2\pi f)^2S_e(f)\,df=\int_0^{\frac{1}{2OSR}}(2\pi f)^2S_e(f)\,df=\frac{\pi^2}{9(OSR)^3}$$

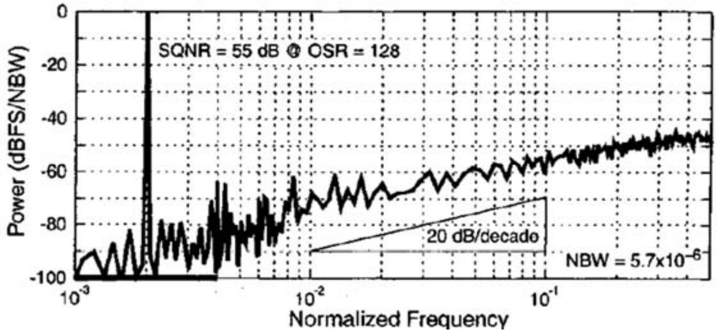
If the input signal is a full-scale sinusoid with peak amplitude M, with STF=1 we have that the output signal power is $\sigma_u^2=M^2/2$, resulting in a signal-to-quantization-noise ratio (SQNR) of

$$SQNR=\frac{9M^2(OSR)^3}{2\pi^2}$$

This means that the SQNR increases by 9dB if the OSR is doubled → this is not much, and the SQNR is usually relatively low (less than 70dB even for OSR as high as 256, M=1)

# Simulations

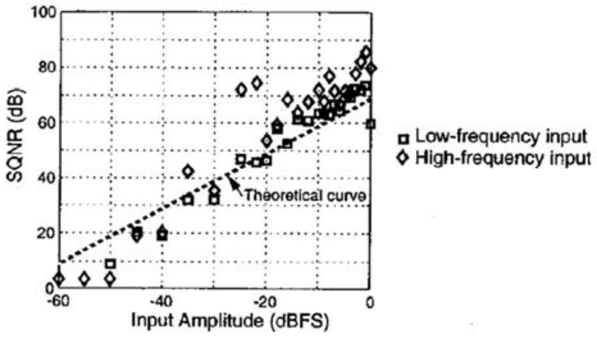
Spectrum of the output of MOD1 for a full-scale input → noise shaping is close to the expected 20dB/decade; the optimal *k* is 0.9 (in the sense explained earlier), very close to the *k*=1 assumed in the linear analysis. However, SQNR = 55dB for OSR=128 and M=1, while 60dB was expected!



# Simulations

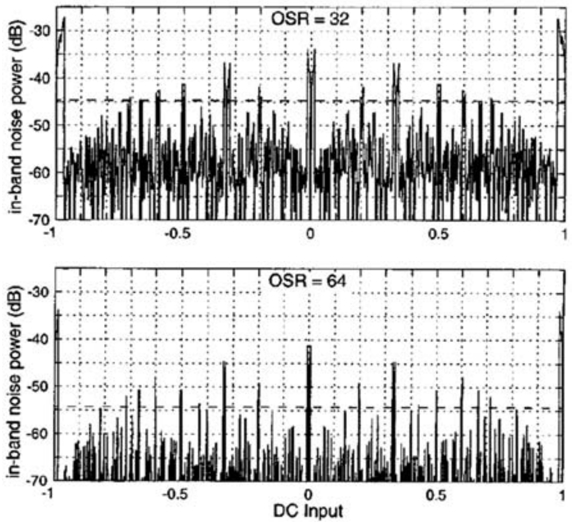
In fact, the SQNR behavior is quite erratic vs. signal amplitude and frequency! (here: OSR = 256)

Obviously, high-frequency signals very often yield a higher SQNR than low-frequency signals



# Strange behavior

In-band noise is particularly ill-behaved for DC inputs, where it is very large for specific values of the input (e.g., 0 and ±1, ±1/2, ±1/3, etc)



# Idle tones

If the input to the modulator is a DC value, the hypothesis of large and random variations at the input of the quantizer (a necessary condition for a white q-error) are not met any more!

Recall the MOD1 equation:

$$y(n) = y(n-1) + u(n) - v(n-1)$$

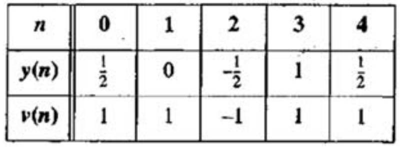
$$v(n) = \mathrm{sgn}\left[y(n)\right]$$

We also assume (to avoid ambiguity) that $v(n) = 1$ if $y(n) = 0$

This yields

$$y(n) = y(n-1) + u(n) - \mathrm{sgn}\left[y(n-1)\right]$$

Assuming u=1/2 and y(0)=1/2, we have a periodic output with period 4: so-called idle tone!



| $n$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $y(n)$ | $\frac{1}{2}$ | 0 | $-\frac{1}{2}$ | 1 | $\frac{1}{2}$ |
| $v(n)$ | 1 | 1 | −1 | 1 | 1 |

# Idle tones

An idle tone (also called *limit cycle*) at $f_s/4$ is relatively easy to filter out. However, in general, if *u*=*n*/*m*, *n* and *m* odd integers, then the limit cycle has period *m*; if *n* or *m* are even, then the limit cycle has period 2*m.*
If *m* is large, the fundamental of the limit cycle and maybe some of its harmonics fall into the signal band, even if OSR is high

If input is a rational number → limit cycles; otherwise no limit cycles; however, limit cycles appear also in the presence of slowly varying inputs that stay near a critical level long enough for a limit cycle to appear

In digital audio idle tones cannot be tolerated, since the human ear can detect tones that are 20dB below the white noise floor!

Higher-order modulators are much less prone to generate idle tones → a strong reason for using them!

Dithering (i.e. adding a pseudo-white-noise to the modulator input, or to the quantizer input) is also an effective way of disrupting limit cycles

# Stability

A linear approach would predict an unconditional stability, as the phase of the loop transfer function is -90º at all frequencies (first-order system, one pole) – however, the modulator is not linear!

Assume DC input → clearly, *y* becomes unbounded if |*u*|>1, since the feedback signal is only +1 or –1

If |*u*|<1 and |*y(0)*|≤2, the loop remains stable, with |*y*|≤2 – this is true also for time-varying inputs; in fact, using the MOD1 equation

$$y(n) = y(n-1) + u(n) - \operatorname{sgn}\left[y(n-1)\right]$$

if |*y(0)*|≤2, then $\left|y(0) - \operatorname{sgn}\left[y(0)\right]\right| \leq 1$, and since |*u*|<1, we obtain

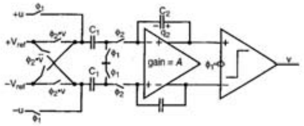
$$\left|y(1)\right| = \left|u(1) + y(0) - \operatorname{sgn}\left[y(0)\right]\right| \leq 2$$

By induction, the same equation applies for all time samples.
If |*y(0)*|>2 and |*u*|<1, the modulator will produce a string of +1 or –1 (depending whether *y(0)*>2 or *y(0)*<2), until |y|<2; at this point, we are back to the previous case → <u>MOD1 is stable with arbitrary |*u*|<1, and can recover from any initial condition</u>

# Finite opamp gain

There are several 2nd-order effects that affect the behavior of modulators – perhaps the most obvious is that opamps do not provide infinite gain! If the DC gain of the opamp is A, the difference equation for the charge in the integrating capacitance $C_2$ becomes

$$q_2(n) = q_2(n-1) + C_1\left(u(n) - v(n-1) - \frac{q_2(n)}{C_2(A+1)}\right)$$



Assuming $C_1$=$C_2$ and A>>1, the voltage across $C_2$ (= $q_2/C_2$) becomes

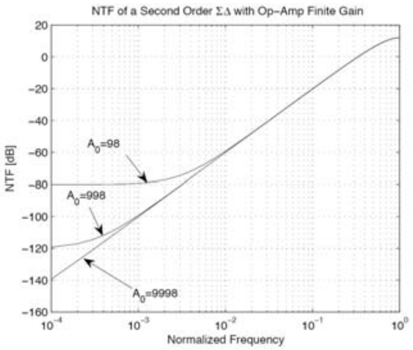
$$Y(z) = p\frac{zU(z) - V(z)}{z-p} \qquad \text{with} \quad p = 1\Big/\left(1+\frac{1}{A}\right) \approx 1 - \frac{1}{A}$$

The integrator does not provide an infinite gain at DC any more (*leaky* integrator); the NTF becomes

$$NTF(z) = 1 - pz^{-1}$$

# Finite opamp gain

The zero of the NTF shifts from DC (z=1) to z=p, inside the unit circle – the NTF gain at DC changes from 0 to 1-p=1/A – no longer infinite precision for DC signals → the NTF “fills-in” close to DC



# Finite opamp gain

The additional noise power caused by the filling-in of the NTF at low frequencies can be estimated by integrating the new NTF,

$$\left|NTF\left(e^{j\omega}\right)\right|^2 = \left|1 - pe^{-j\omega}\right|^2 \approx \left|1 - p(1 - j\omega)\right|^2 \approx \left|A^{-1} + j\omega\right|^2 = A^{-2} + \omega^2$$

between 0 and the signal band limit, i.e. $\pi/OSR$, and comparing with the ideal case of an infinite A. If A>OSR, the SNR loss is less than 0.2dB → negligible

However, this assumes a <u>*linear*</u> gain A – a low opamp gain can be a problem if the gain is sufficiently non-linear!

(Also, a finite gain is a big issue in MASH (cascaded) modulator architectures)

# Non-linear analysis – dead zones

Let us assume a small positive DC input *u*, with infinite A, and with y(0)=0

→ $$y(n)=y(n-1)+u(n)-\operatorname{sgn}\left[y(n-1)\right];\qquad y(0)=0$$

We obtain for the first samples:

$$y(1)=y(0)+u-\operatorname{sgn}\left[y(0)\right]=u-1<0$$

$$y(2)=u-1+u+1=2u>0$$

$$y(3)=2u+u-1=3u-1<0$$

Thus, initially the output alternates; for the kth sample we have

$$y(k)=\begin{cases} ku-1, & k \text{ odd} \\ ku, & k \text{ even} \end{cases}$$

The effect of *u*>0 occurs in the first odd sample for which *ku*–1>0, at which time *v(k)* is positive, and therefore *v(k)*=1 appears twice in a row → this occurs with a frequency of *u* (i.e. after 1/*u* periods)

# Dead zones

If now lossy integrator: $y(n)=py(n-1)+u(n)-\operatorname{sgn}\left[y(n-1)\right];\qquad y(0)=0$

$$y(1)=py(0)+u-\operatorname{sgn}\left[y(0)\right]=u-1<0$$

$$y(2)=p(u-1)+u+1=(1+p)u+(1-p)>0$$

$$y(3)=p\left[(1+p)u+(1-p)\right]+u-1=\left(1+p+p^2\right)u-\left(1-p+p^2\right)<0$$

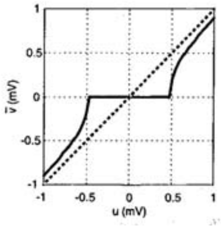
kth sample: $$y(k)=\sum_{i=0}^{k-1}p^i u+(-1)^k\sum_{i=0}^{k-1}(-p)^i$$

The effect of *u* occurs if sgn(*y*) is positive for some odd value of *k*; for *k* large, this requires

$$\frac{u}{1-p}>\frac{1}{1+p}\quad\to\quad u>\frac{1-p}{1+p}=\frac{1/A}{2-1/A}\approx\frac{1}{2A}$$

# Dead zones

Thus, inputs smaller than 1/(2A) will have no effects on the output! e.g. with A=1000 and $V_{ref}$=1V, a DC signal less than 0.5mV will have no effect! → dead zone (or dead band) around *u*=0



Dead zones exist around all rational values of *u*, and are narrower (except those around ±1) than that around 0

Dead zones have an impact on limit cycles as well – in an ideal MOD1 the limit cycles are unstable or non-attracting, since an arbitrarily small change in the input results eventually in large signals after the integrator and therefore in a disruption of the limit cycle – but if the integrator is leaky, the limit cycles are stable (attracting), since a small change in the input will lead to a small change in the integrator output, and consequently to no change in the output pattern – this is of course a very detrimental effect! (by the way, if the NTF zeros are outside the unit circle, the limit cycles are repelling, and the modulator becomes "chaotic" – good for breaking limit cycles)

# Decimation filter for MOD1

Very large q-noise at high frequencies → out-of-band noise must be removed by a digital lowpass filter; afterwards the signal may be decimated, thereby reducing the sampling rate to the Nyquist limit $2f_B$

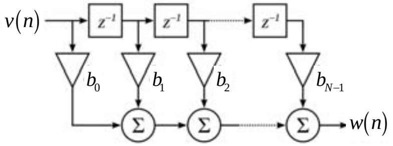
The gain of the lowpass filter must be flat and large in-band, and very small between $f_B$ and $f_s/2$; furthermore, it is often desirable to have a flat group delay response in the signal band; this can be accomplished using a linear-phase finite-impulse-response (FIR) filter

In a single-bit modulator, it may be practical to use a single-stage high-order FIR filter, since there are no actual multiplications involved between signal samples and coefficients of the filter taps. However, it is more efficient to carry out filtering and decimation in stages. The stages most often used are the so-called sinc filters

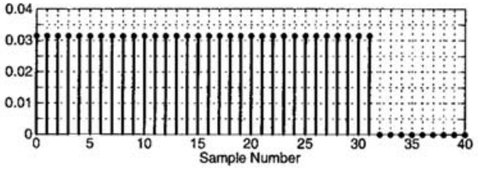
A sinc filter is an FIR filter with N-1 delays and N equal-valued tap weights – it computes the running average of the input data stream *v(n)*

# Decimation filter

FIR filter ($b_i$ = 1/N), FIR output, and impulse response (with N=32):



$$w(n)=\frac{1}{N}\sum_{i=0}^{N-1}v(n-i)\quad\rightarrow\quad h_1(n)=\begin{cases}1/N, & 0\le n\le N\text{-}1\\ 0, & \text{otherwise}\end{cases}$$



In the z-domain: $$H_1(z)=\frac{1}{N}\left(1+z^{-1}+...+z^{-N+1}\right)=\frac{1}{N}\frac{1-z^{-N}}{1-z^{-1}}$$

# Decimation filter – II

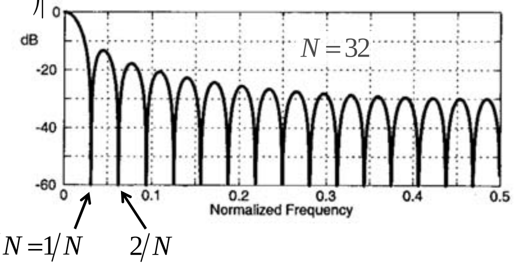
In the frequency domain:

$$H_1\left(e^{j2\pi f}\right)=\frac{1}{N}\frac{1-e^{-j2\pi fN}}{1-e^{-j2\pi f}}=\frac{1}{N}\frac{e^{-j\pi fN}}{e^{-j\pi f}}\frac{e^{j\pi fN}-e^{-j\pi fN}}{e^{j\pi f}-e^{-j\pi f}}$$
$$=\frac{1}{N}e^{-j\pi f(N-1)}\frac{\sin(\pi fN)}{\sin(\pi f)}=e^{-j\pi f(N-1)}\frac{\operatorname{sinc}(fN)}{\operatorname{sinc}(f)}$$

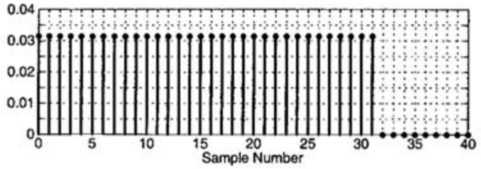
linear phase

where the sinc function is defined as $\operatorname{sinc}(f)=\dfrac{\sin(\pi f)}{\pi f}$

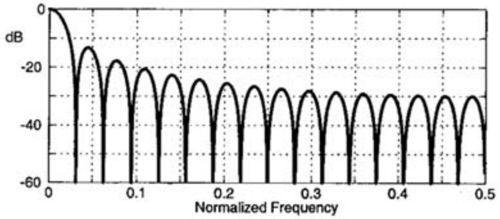
$$\left|H_1\left(e^{j2\pi f}\right)\right|$$



# Sinc impulse/frequency response



$$h_1(n)=\begin{cases}1/N, & 0\le n\le N\text{-}1\\ 0, & \text{otherwise}\end{cases}$$



$$\left|H_1\left(e^{j2\pi f}\right)\right|=\frac{\operatorname{sinc}(fN)}{\operatorname{sinc}(f)}$$

Gain close to 1 at baseband; close to 0 near $f_s$/N and its harmonics → if the sampling rate at the output is reduced by N, causing the noise around $f_s$/N, 2$f_s$/N, …, to fold into the baseband, the energy of the aliased noise will be reduced by the attenuation of the sinc filter → N=OSR should be chosen

# Decimation and folding

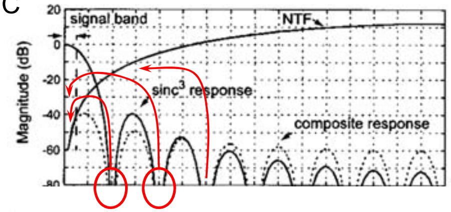
We can see the decimated function $f_D$ as the original function *f*, multiplied by a rectangular function *g* having a duty cycle of 1/N:

$$f_D(t)=f(t)\cdot g(t)\qquad\qquad g(t)=\begin{cases}1, & 0\le t\le T_s\\ 0, & T_s<t\le NT_s\end{cases}$$

Obviously, *g(t)* has a period of NT$_s$, i.e. a normalized period of N → the fundamental harmonic of *g* is 1/N

The spectrum of $f_D$ is the convolution of the spectrum of *f* with the spectrum of *g:* $$f_D(t)=f(t)\cdot g(t)\quad\rightarrow\quad F_D(\omega)=F(\omega)*G(\omega)$$

→ the harmonic content of *F* at frequency 1/N and its harmonics is folded (scaled by the corresponding tone in G) on top of the spectrum of *F* at and close to DC



example with sinc$^3$ filter

# Decimation filter – III

How much residual noise is left after the FIR filter, compared to an ideal lowpass filter? → the noise at the FIR output is

$$Q_1(z)=H_1(z)\,NTF(z)\,E(z)=\frac{1}{N}\frac{1-z^{-N}}{1-z^{-1}}\left(1-z^{-1}\right)E(z)=\frac{1}{N}\left(1-z^{-N}\right)E(z)$$

which in the time domain becomes

$$q_1(n)=\frac{1}{N}\left[e(n)-e(n-N)\right]$$

Assuming *e(n)* and *e(n-N)* to be uncorrelated, and each with an rms value of $\sigma_e$, the q-noise power at the sinc output is

$$\sigma_{q_1}^2=\frac{2\sigma_e^2}{N^2}$$

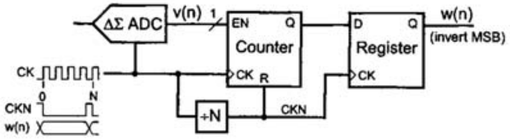
while the output noise power for a MOD1 followed by an ideal LPF with unity gain at DC is, with OSR=N:

$$\sigma_q^2=\frac{\pi^2\sigma_e^2}{3N^3}$$

# Decimation filter – IV

Thus, a single sinc filter is N times less effective than an ideal LPF → therefore, the sinc is usually only one stage in a multi-stage decimator filter

By the way, in single-bit modulators the sinc decimator can be realized very easily

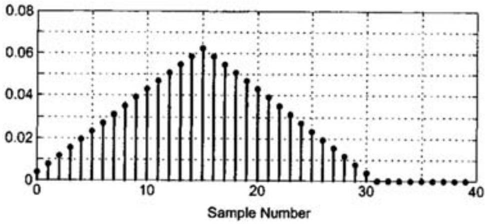


the output is a down-sampled count of the number of 1s over the last N clock cycles → accumulate and dump

If N=$2^k$, the counter produces a k-bit output, which may be interpreted as a binary fraction between 0 and 1 in a unipolar topology. In a bipolar ADC, the MSB is inverted and the data is interpreted in a 2's complement form

# Decimation filter – V

Better decimation → two cascaded sinc filters → sinc² filter → the impulse response is obtained by convolving the sinc response with itself



In frequency domain, the response is just the square of the sinc response:

$$H_2(z)=\left(\frac{1}{N}\frac{1-z^{-N}}{1-z^{-1}}\right)^2 \rightarrow H_2\left(e^{j2\pi f}\right)=e^{-j2\pi f(N-1)}\left(\frac{\operatorname{sinc}(fN)}{\operatorname{sinc}(f)}\right)^2$$

# Decimation filter – VI

The residual q-noise at the sinc² output is given by

$$Q_2(z)=H_2(z)\,NTF(z)\,E(z)=\frac{1}{N^2}\left(\frac{1-z^{-N}}{1-z^{-1}}\right)^2\left(1-z^{-1}\right)E(z)$$

$$=\frac{1}{N^2}\frac{1-z^{-N}}{1-z^{-1}}\left(1-z^{-N}\right)E(z)=\frac{1}{N}H_1(z)\left[\left(1-z^{-N}\right)E(z)\right]$$

$H_1$ performs an N-sample average in the time domain, while $\left(1-z^{-1}\right)E(z)$ corresponds to $e(n)-e(n-N)$ →

$$q_2(n)=\frac{1}{N^2}\sum_{i=0}^{N-1}e(n-i)-e(n-N-i)$$

If $e$ is white with power $\sigma_e^2$, the noise power at the output of the sinc² becomes

$$\sigma_{q_2}^2=\frac{2N\sigma_e^2}{N^4}=\frac{2\sigma_e^2}{N^3}$$

# Decimation filter – VII

sinc² noise:

$$\sigma_{q_2}^2 = \frac{2\sigma_e^2}{N^3}$$

ideal noise:

$$\sigma_q^2 = \frac{\pi^2\sigma_e^2}{3N^3}$$

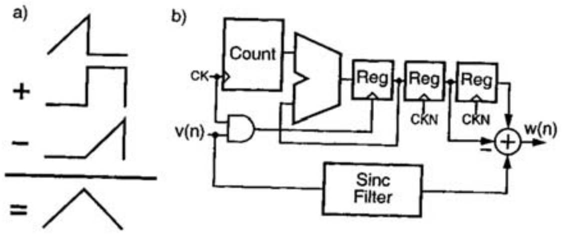
Thus, the sinc² noise is even lower than the ideal!!

However, we have to take into account the passband droop for the desired signal (which can be equalized with a post-filter if necessary), which reduces somewhat the signal energy as well → all in all, the SNR is slightly lower than in the ideal case – small difference, sinc² ok!

In general, a sinc$^{L+1}$ LPF is sufficient for an L$^{th}$-order loop

# Decimation filter – VIII

Finally, the triangularly-weighed sum needed for the sinc² can be generated more efficiently than cascading two sinc filters, as shown e.g. in the solution below


a)
+
–
=
b)
Count
CK
Reg
Reg
Reg
CKN
CKN
v(n)
w(n)
Sinc Filter