Advanced AD/DA converters

Second-Order ΔΣ Modulators

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

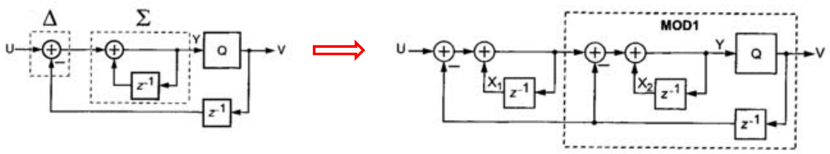
Overview

- Basic MOD2 modulator
- Non-linear behavior and dead bands
- Stability
- Alternative designs of MOD2
- Decimator filter

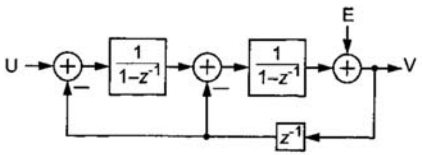
Basics

For most applications, MOD1 is not enough, in terms of resolution and idle-tone generations – MOD2 is a much better choice!

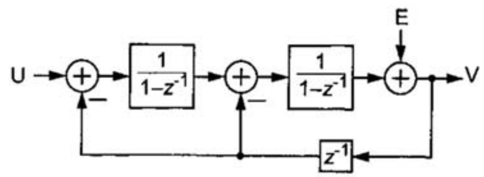
Straightforward way of implementing MOD2: replace the quantizer in MOD1 with another copy of MOD1:



Linearized circuit:



MOD2 linear equations



$$V(z)=E(z)+\frac{1}{1-z^{-1}}\left[-z^{-1}V(z)+\frac{1}{1-z^{-1}}\left(-z^{-1}V(z)+U(z)\right)\right]$$

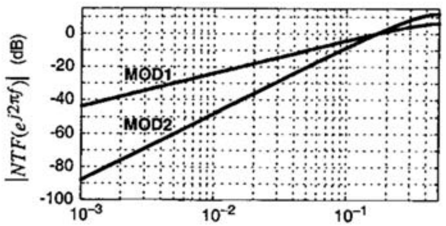
$$=\frac{\left(1-z^{-1}\right)^{2}E(z)-\left[\left(1-z^{-1}\right)z^{-1}+z^{-1}\right]V(z)+U(z)}{\left(1-z^{-1}\right)^{2}}$$

$$\boxed{V(z)=U(z)+\left(1-z^{-1}\right)^{2}E(z)}$$

NTF

STF=1, and NTF is the square of MOD1's NTF → much higher attenuation of in-band q-noise!

$$NTF(z)=\left(1-z^{-1}\right)^2 \rightarrow \quad \left|NTF\left(e^{j2\pi f}\right)\right|^2=\left|e^{-j2\pi f}\left(e^{j\pi f}-e^{-j\pi f}\right)^2\right|^2=\left[2\sin(\pi f)\right]^4 \approx (2\pi f)^4$$



Notice that the out-of-band gain of the q-noise is higher in MOD2 → needs a more selective decimation filter (in fact, the total power of the q-noise is higher in MOD2)

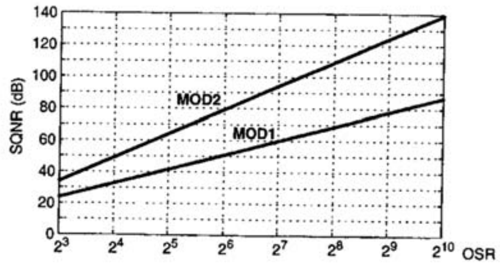
Noise power

Following the same steps used in the derivation of the in-band q-noise for MOD1, we obtain

$$\sigma_q^2=\int_0^{\frac{1}{2OSR}}(2\pi f)^4\cdot 2\sigma_e^2=\frac{\pi^4\sigma_e^2}{5\cdot OSR^5}$$

If again $\sigma_e^2=1/3$, and assuming an input sine wave of amplitude M, we obtain the SQNR expression

$$SNQR=\frac{M^2/2}{\sigma_q^2}=\frac{15\,M^2 OSR^5}{2\pi^4}$$

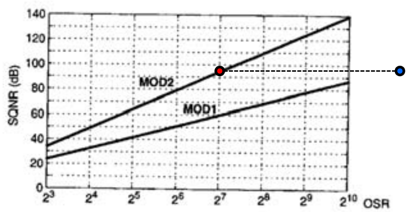


SQNR

Double OSR → SQNR increases by a factor 32 = 15dB = 2.5 bits!

This is much better than the 1.5 bits given by MOD1

Example: MOD2, OSR=128, M=1 (single-bit DAC) → SQNR=94.2dB → resolution of almost 16 bits (if audio is targeted, f_B=20kHz and f_s=5.12MHz, not an issue)



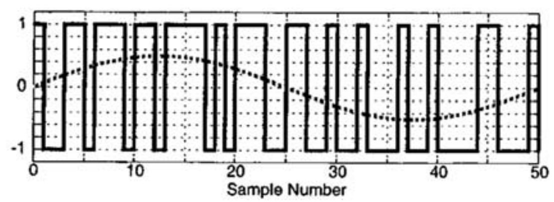
With MOD1, we need OSR=1800 and f_s=72MHz

Single-bit quantizer → same inherent linearity as already discussed for MOD1

MOD2 simulations

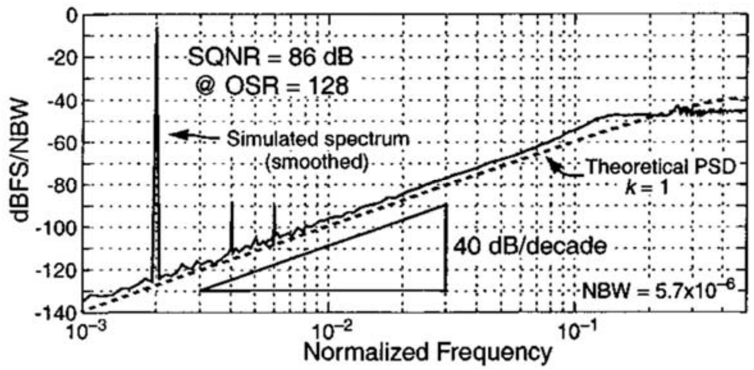
MOD2 is, as MOD1, intrinsically non-linear → time-domain simulations are more reliable than the linearized model (but, alas, also much more difficult to generalize!)

The actual time-domain plot of the MOD2 output, however, provides very little insight about the quality of the output signal!



Simulations

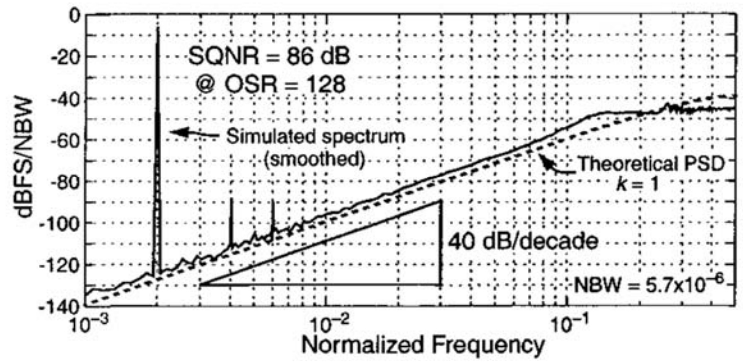
The spectrum of the MOD2 output signal, for an input of -6dBFS, shows a 40dB/decade noise shaping, as expected, which yields an SQNR of 86dB, extrapolated to 92dB at full scale, in close agreement with the 94dB predicted by the theory (previous SQNR curve)



Simulations

There are, however, two difficulties. The first is that there are 3rd-order and 5th-order harmonics in the spectrum, which cannot be accounted for by a linear system and white quantization noise model!

The second is that the theoretical NTF (scaled by $2\sigma_e^2 NBW$ – more on the *NBW* later) does not align with the simulated one: it is lower at low frequencies and higher at high frequencies

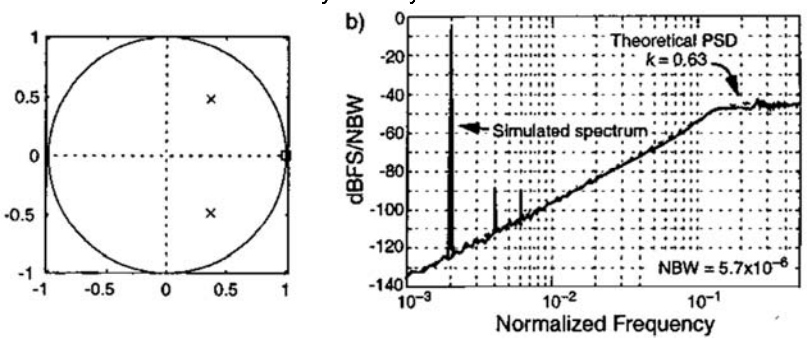


Simulations

We start with the second discrepancy, by calculating (as already done in the MOD1 case) the effective quantizer gain *k* as

$$k = E\left[|y|\right] \big/ E\left[y^2\right]$$

Using the simulation data, we obtain *k*=0.63 – a recomputation of the NTF with *k*=0.63 yields new poles/zeros (the zeros are actually still at DC) for the NTF, and now the theoretical quantization noise matches time-domain simulations very closely



Quantizer gain

However, if the input signal amplitude is lowered, simulations show that the optimal quantizer gain increases (slightly) to *k*=0.75 (for inputs below -12dBFS)

The following formula gives the NTF for any value of *k*, once the NTF for *k*=1 is known:

$$NTF_k(z) = \frac{NTF_1(z)}{k + (1-k)\,NTF_1(z)}$$

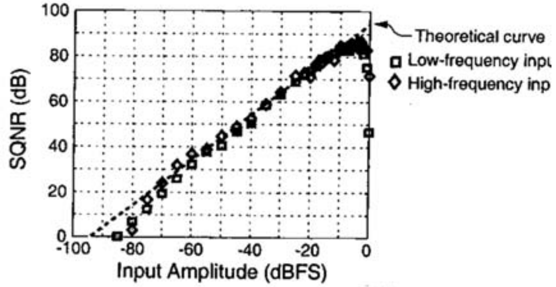
where NTF_1 is the NTF for *k*=1; for *k*=0.75 we obtain

$$NTF_{0.75}(z) = \frac{(z-1)^2}{z^2 - 0.5z + 0.25}$$

This NTF has an in-band gain 1/0.75 = 2.5dB higher than NTF_1 → in-band noise in time-domain simulations should be 2.5dB higher than that predicted by NTF_1

Non-linear quantizer gain

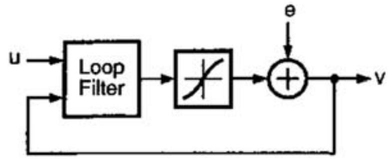
When the input amplitude exceeds -6dBFS the optimal *k* decreases, degrading further noise shaping. Below you can see the SQNR plots for two different frequencies, comparing them with the theory in the ideal case of *k*=1. In the middle range of the amplitudes the match is quite good; for small amplitudes the SQNR is lower; for large amplitudes the SQNR peaks at -5dBFS, then drops abruptly → the largest degradation is for large low-frequency signals, which apply large signals for a longer time. Finally, it is clear that the SQNR here is much more well behaved than in MOD1



Non-linear effects

We have seen that *k*=0.75 for small input signals, and *k*=0.63 for larger signals → the binary quantizer shows a non-linear effective gain → the distortion we have wondered about should in fact be expected!!

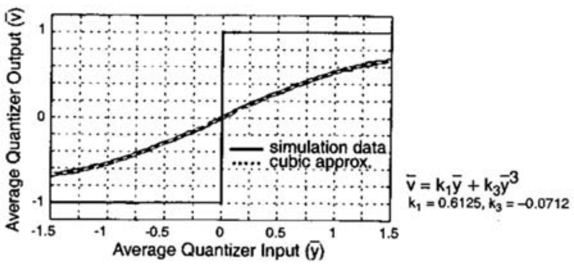
We can apply a quantitative approach by replacing the quantizer with the (weakly) non-linear quantizer transfer curve (QTC) + additive noise (as usual)



QTC is determined by computing the average quantizer output as a function of the average quantizer input, while the input signal amplitude is swept

Quantizer transfer curve

QTC is compressive, meaning that *k* drops for large inputs (we already knew this) → cubic approximation, with coefficients k_1 = 0.6125 and k_3 = -0.0712



We proceed by first determining the NTF with k=k_1 :

$$NTF(z,k_1)=\frac{(z-1)^2}{z^2-0.775z+0.3875}$$

Quantizer transfer curve

Now, the distortion term k_3y^3 is added at the loop output → same place as q-noise → shaped by the NTF → distortion is largest where the NTF is largest → at the edge of the passband

For a small low-frequency sinusoidal input of amplitude A, the average of the output is also a sinusoid of amplitude A (since STF=1), and, in the linear model, the average of the quantizer input is also a sinusoid, of amplitude A/k_1 → the amplitude of the 3rd harmonic generated by the QTC is therefore

$$A_{3rd}=k_3\cdot\frac{1}{4}\frac{A^3}{k_1^3}$$

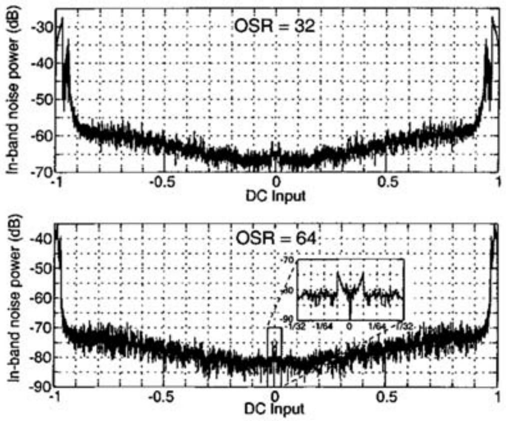
from which the 3rd-order harmonic distortion is easily calculated as

$$HD_3=\left|\frac{A_{3rd}\cdot NTF(z,k_1)}{A}\right|=\left|\frac{k_3A^2}{4k_1^3}NTF(z,k_1)\right|$$

where $z=e^{j2\pi(3f)}$, since the NTF must be evaluated at the frequency of the 3rd harmonic → with A=0.5 (= -6dBFS) and f=1/500, same conditions as the previous FFT, we obtain HD_3 of -87dB, while we simulated -82dB → fair approximation, but no more than fair!

Quantization noise

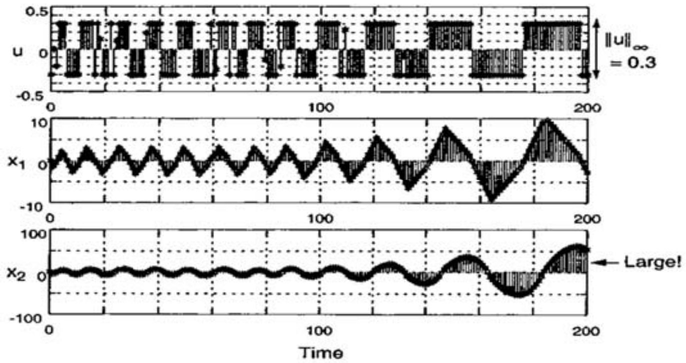
In-band q-noise power (still with binary quantization) vs. DC input level → much better behaved than MOD1, but tones are still present!
Large input signals result in large q-noise (we know that *k* drops with increasing signal amplitude) → q-noise is not stationary, but rather time-variant, as it is a function of the input signal!



Stability of MOD2

It can be shown that MOD2 is stable for all DC inputs below the reference (assumed to be 1V, or better 1)

Since the modulator tracks the low-frequency part of its input, we might be tempted to conclude that all time-varying signals remaining below 1 lead to stable operation. But this is not true, see below! Luckily, this input waveform is not likely to appear ☺



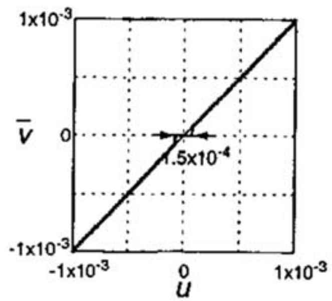
Stability of MOD2

It is known that MOD2 is stable for arbitrary inputs lower than 0.1 – the upper limit of the input amplitude still guaranteeing stable operation is not known!

It is a good idea to keep the input amplitude below 0.9 or 0.8 (which is the one drawback compared to MOD1, which can accept input signals with amplitude up to 1), and also to be able to detect unwanted large states, to force the modulator back into a stable state

Dead bands

Finite gain A of the opamps is again resulting in dead bands; however, the fact that there are two cascaded integrators results in much smaller dead bands, with width $1.5/A^2$ instead of $1/A$ as in MOD1 → much smaller problem! Simulations confirm this (A=100)



Circuit design issues – op-amp offset

The offset of the first integrator and of the input DAC are added to the input signal and cause equal offsets at the output

The offset of the second integrator is referred to the input by dividing it by the gain of the first integrator, which is very large at DC → negligible impact

The ADC offset is also divided by the gain of one or more integrators when referred to the input → negligible impact → opens up the possibility of positioning the ADC thresholds at optimal voltage levels

This and next several slides from Maloberti's book

Circuit design issues – finite op-amp gain

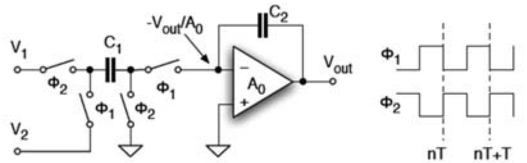
The DC gain of the op-amp is not infinite → we obtain

$$C_2 V_{out}(n+1)\cdot\left(1+\frac{1}{A_0}\right) = C_2 V_{out}(n)\cdot\left(1+\frac{1}{A_0}\right) + C_1\left[V_1(n) - V_2(n+1) - \frac{V_{out}(n)}{A_0}\right]$$

$$\frac{V_{out}}{V_1 - z^{-1}V_2} = \frac{C_1}{C_2}\left(\frac{A_0}{1+A_0+C_1/C_2}\right)\frac{z^{-1}}{1-\frac{(1+A_0)C_2}{C_1+(1+A_0)C_2}z^{-1}}$$

→ gain error of $A_0/(1+A_0)$, and pole inside the unit circle:

$$z_p = (1+A_0)/(1+A_0+C_1/C_2)$$



Finite op-amp gain – II

STF is only marginally affected; however, the NTF is not longer zero at DC, becoming

$$NTF = \left(1 - z_{p1}z^{-1}\right)\left(1 - z_{p2}z^{-1}\right)$$

and, at DC (i.e. $z=1$) $NTF(DC) = \left(1-z_{p1}\right)\left(1-z_{p2}\right)$

If the two gains and the two caps are equal, we obtain $NTF = \left(1 - \frac{1+A_0}{2+A_0}z^{-1}\right)^2$

Corner frequency at

$$\frac{1+A_0}{2+A_0}e^{-s_c T} = 1 \;\rightarrow\; e^{s_c T} = \frac{1+A_0}{2+A_0} \;\rightarrow\; s_c T = \ln\left(\frac{1+A_0}{2+A_0}\right)$$

(s_c is negative, as it should in a stable system)

$$\omega_c T = -s_c T = \ln\left(\frac{2+A_0}{1+A_0}\right) = \ln\left(1 + \frac{1}{1+A_0}\right) \approx \frac{1}{1+A_0}$$

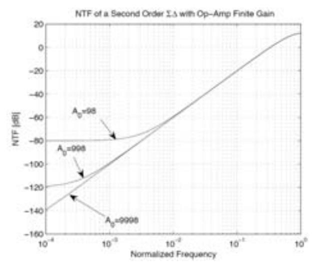
$$f_c = \frac{1}{2\pi T}\frac{1}{1+A_0} = \frac{f_s}{2\pi}\frac{1}{1+A_0}$$

Finite op-amp gain – III

The finite op-amp gain does not affect the NTF as long as $f_B >> f_c$ → both gain and OSR must be set to satisfy the condition

$$f_B >> f_c \;\rightarrow\; f_B >> \frac{f_s}{2\pi}\frac{1}{1+A_0} \;\rightarrow\; \frac{f_s}{2}\cdot\frac{1}{OSR} >> \frac{f_s}{2\pi}\frac{1}{1+A_0} \;\rightarrow\; \pi(1+A_0) >> OSR$$

resulting in a very relaxed op-amp gain demand for modulators with medium OSR

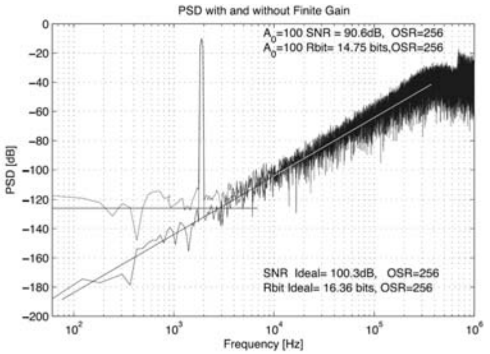


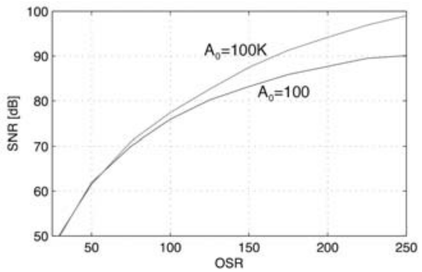
However, this assumes a *linear* gain A – a low opamp gain can be (and often is) a problem if the gain is sufficiently non-linear

Circuit design issues – finite op-amp gain

Simulations on a 2nd-order single-bit ΔΣ modulator with op-amps with A_0=100 and sampling frequency of 2MHz → corner frequency at 3kHz, in very good agreement with theory

Furthermore, as long as the condition $\pi(1+A_0) \approx 320 >> OSR$ is fulfilled, no SNR penalty is paid, compared to having A_0=100k; however, 10dB are lost if OSR=250, see below





Circuit design issues – finite op-amp bandwidth

Assuming a single-pole response, we have

$$V_{out}(nT+t) = V_{out}(nT) + \Delta V_{out}\left(1-e^{-t\beta/\tau}\right)$$

with $\beta = C_2/(C_1+C_2)$. The integration phase stops at T/2, causing an error on the final output of

$$\varepsilon_{BW} = \Delta V_{out} e^{-T\beta/2\tau}$$

The error is proportional to the signal itself → bad for linearity

Finite op-amp slew-rate and bandwidth

Assuming an ideal SC integrator, an input step of $-V_{in}$ would result in an output step of $\Delta V_{out} = V_{in} \cdot C_1/C_2$; in contrast, a real op-amp has a slewing time of

$$t_{slew} = \frac{\Delta V_{out}}{SR} - \tau$$

At t=t_{slew}, the output voltage differs from the final value by $\Delta V = SR \cdot \tau$, and evolves exponentially in the remaining fraction of T/2; at T/2, the error on the output voltage is

$$\varepsilon_{SR} = \Delta V e^{-(T/2 - t_{slew})/\tau}$$

Thus, also in this case the error depends on the step itself → possible impact on linearity

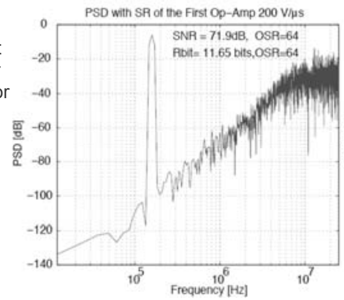
All these equations can be used in a behavioral simulator to enormously speed up the study of the combined impact of finite bandwidth and finite slew rate for the op-amp

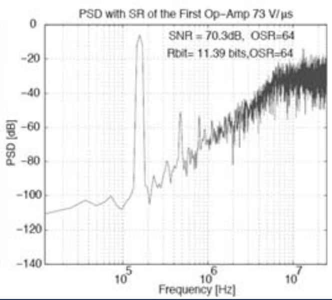
Finite op-amp slew-rate and bandwidth – II

Ideal simulations show that the maximum changes at the output of the 1st and 2nd integrators are 0.749V and 3.21V → with an f_s of 50MHz, we have $SR_1 > \Delta V_{out,1}/(T/2) \approx 75V/\mu s$ $\quad SR_2 > \Delta V_{out,2}/(T/2) \approx 321V/\mu s$

Ideally, SNR=72dB with op-amp's βf_T=100MHz, OSR=64, f_{in}=160kHz
If $SR_1 = 78V/\mu s, SR_2 = 325V/\mu s$, the SNR does not change significantly; if $SR_1 = 73V/\mu s$, the SNR is not much affected, but the non-linear output response gives rise to harmonic tones – finally, simulations show (as expected) that a performance degradation on the 2nd integrator has a lower impact than on the 1st

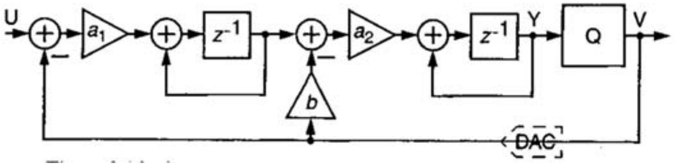
single-bit
2nd-order
modulator





Alternative 2nd-order modulators – Boser-Wooley

Two delaying integrators → good, because they can settle independently of each other, relaxing speed requirements (and therefore power consumption) in DT designs



Assuming k=1:

$$STF(z)=\frac{a_1a_2z^{-2}}{D(z)};\qquad NTF(z)=\frac{\left(1-z^{-1}\right)^2}{D(z)}$$

with:

$$D(z)=\left(1-z^{-1}\right)^2+a_2bz^{-1}\left(1-z^{-1}\right)+a_1a_2z^{-2}$$

Other 2nd-order modulators – Boser-Wooley – II

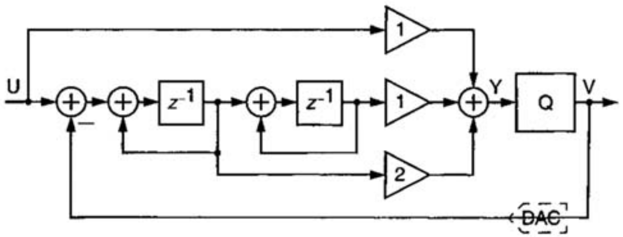
We desire: $STF(z)=z^{-2}$ $\qquad NTF(z)=\left(1-z^{-1}\right)^2$

→ the conditions are: $a_1a_2=1,\qquad a_2b=2$

Infinite number of possible solutions → in reality dynamic range scaling (more about it later) removes the ambiguity in selecting the three parameters

Silva-Steensgaard

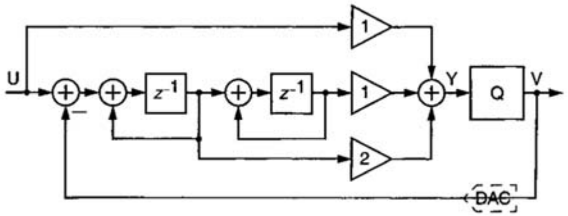
Direct feedforward from input to quantizer, and single feedback from the digital output



The output is, as before:

$$V(z)=U(z)+\left(1-z^{-1}\right)^2E(z)$$

Silva-Steensgaard



However, the input signal to the loop contains now only the shaped quantization noise!

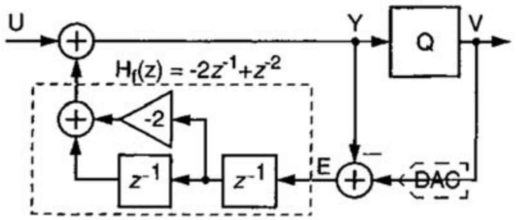
$$U(z)-V(z)=-\left(1-z^{-1}\right)^2E(z)$$

The loop does not process the signal → does not have to be very linear, great advantage!

Output of the second integrator is $-z^{-2}E(z)$, which can be directly used if the modulator is the first stage in a MASH architecture

Error-Feedback

An example of a modulator that looks simple and therefore attractive, but which is not suitable for analog applications!



The q-error is obtained in analog form by subtracting the input of the internal ADC from the DAC output → we obtain

$$V(z)=Y(z)+E(z)=U(z)+H_f(z)E(z)+E(z)=U(z)+\left(-2z^{-1}+z^{-2}+1\right)E(z)$$

$$\rightarrow \qquad STF(z)=1, \qquad NTF(z)=\left(1-z^{-1}\right)^2$$

Error-Feedback

Very nice, but also very sensitive to parameter variations in the feedback path!

For instance, if the multiplication by 2 has a 0.5% error, resulting in 2.01, the NTF will become

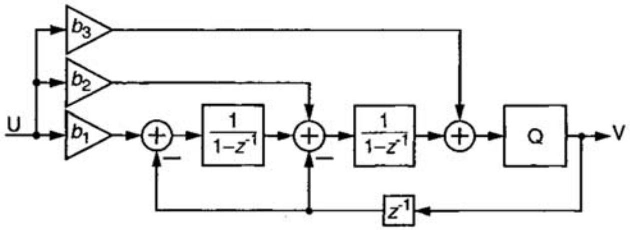
$$NTF(z)=\left(1-z^{-1}\right)^2-0.01z^{-1}$$

Thus, at very low frequencies the NTF will be 0.01 instead of 0, or equivalently -40dB → typically, ENOB not larger than 10, even with high OSR – comparable mismatches still allow ENOB=18 for all other 2nd-order modulators discussed so far!

On the other hand, the error-feedback modulator is very useful in the digital loop required in ΔΣ DACs, since the numbers 1 and 2 (and many others!) are easily realized with arbitrary precision in the digital domain!

More general 2nd-order structure

The input is fed also to the input of the second integrator and of the quantizer

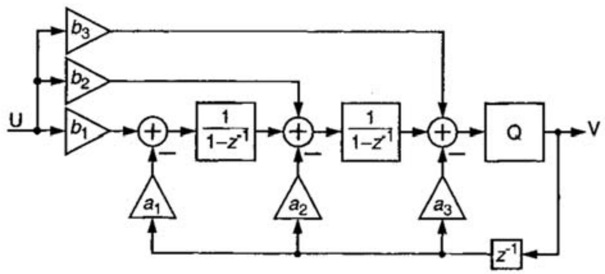


The STF becomes $STF(z)=b_1+b_2\left(1-z^{-1}\right)+b_3\left(1-z^{-1}\right)^2$ →

a free 2nd-order FIR pre-filter can be incorporated into the modulator!

General 2nd-order structure

All possible feedforward and feedback coefficients!



$$STF(z)=\frac{B(z)}{A(z)} \qquad NTF(z)=\frac{\left(1-z^{-1}\right)^2}{A(z)}$$

$$B(z)=b_1+b_2\left(1-z^{-1}\right)+b_3\left(1-z^{-1}\right)^2$$

$$A(z)=1+\left(a_1+a_2+a_3-2\right)z^{-1}+\left(1-a_2-2a_3\right)z^{-2}+a_3z^{-3}$$

General structure

$$NTF(z)=\frac{\left(1-z^{-1}\right)^2}{A(z)}$$

$$A(z)=1+\left(a_1+a_2+a_3-2\right)z^{-1}+\left(1-a_2-2a_3\right)z^{-2}+a_3z^{-3}$$

The a_3 feedback term increases the NTF order to 3, but does not increase the number of in-band NTF zeros → not used in discrete-time applications – however, it is often useful in continuous-time modulators! (as we shall see later)

Multiple feedforward/feedback yields more flexibility → enhanced stability, improved dynamic range

Optimal 2nd-order modulator

So, which architecture is the best?? This actually boils down to optimizing the NTF! Of course, also the topology used is important, but more because of practical considerations (such as robustness to mismatches etc) than fundamental limits

The STF only filters the signal, and plays no major role in the SQNR

Let us find the 2nd-order NTF yielding the highest SQNR → minimizes the in-band quantization noise – we start from

$$NTF(z)=\frac{\left(1-z^{-1}\right)^2}{A(z)}$$

Now, assuming A(z) almost constant in the passband, for a high value of OSR the magnitude of the NTF in the passband is

$$\left|NTF(\omega)\right|=\left|\frac{\left(1-e^{-j\omega}\right)^2}{A\left(e^{j\omega}\right)}\right|_{\omega\text{ small}}\approx\frac{\omega^2}{A(1)}\equiv K\omega^2$$

Optimal 2nd-order modulator – II

Shifting the zeros from $z_z=e^{\pm j0}=1$ to $z_z=e^{\pm j\alpha}$, we obtain the new NTF:

$$\left|NTF(\omega)\right|=K(\omega-\alpha)(\omega+\alpha)=K\left(\omega^2-\alpha^2\right)$$

The integral over the passband of $|NTF|^2$ is a measure of the in-band noise → the following integral should be minimized with respect to α:

$$I(\alpha)=\int_0^{\omega_B}\left(\omega^2-\alpha^2\right)^2d\omega$$

yielding

$$\alpha_{opt}=\frac{\omega_B}{\sqrt{3}}\quad\rightarrow\quad\frac{I(0)}{I\left(\alpha_{opt}\right)}=\frac{9}{4}\quad\rightarrow\quad SNR_{opt}=SNR+3.5dB$$

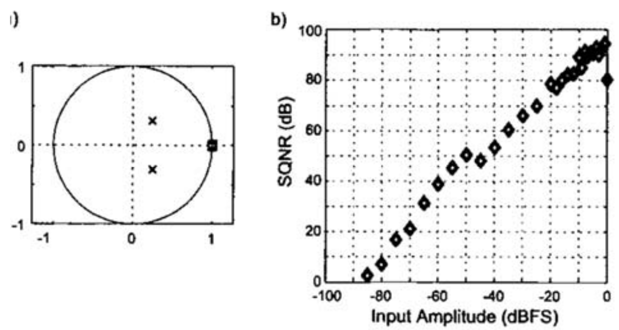
To repeat, this assumes white q-noise and A(z)=A(1) over the passband

Optimal 2nd-order modulator – III

An exhaustive search of the NTF design space for the highest SQNR yields the NTF with denominator

$$A(z)=1-0.5z^{-1}+0.16z^{-2}$$

Compared to our standard MOD2, this SQNR is more linear and supports signals closer to full scale without saturating. The peak SQNR is about 94dB, which is approximately 6dB higher than that of MOD2

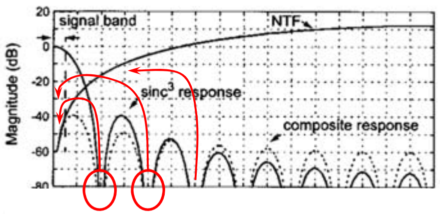


Decimation

Cascade of sinc filters → to find the order K of the sinc^K filter for an L^{th}-order modulator, we have to consider the following:

1) Around f_B, the filter should cut off at a faster rate than the modulator NTF rises, so that very little out-of-band noise is left unsuppressed around f_B after decimation
2) The gain of the filter around f_s/OSR (and its harmonics) should be less steep than the NTF around DC, so that the folding of the noise from the bands around f_s/OSR, $2f_s$/OSR, etc, after decimation adds negligibly to the in-band noise

Both conditions require K > L; usually, K = L+1 is already enough

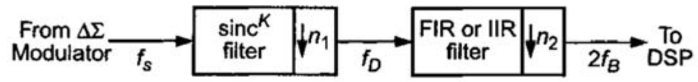


example with sinc^3 filter

Decimation

2^{nd}-order modulator → since NTF grows with order 2, the order of the sinc should be 3 – higher is not necessary

If high resolution is needed → two-stage decimator

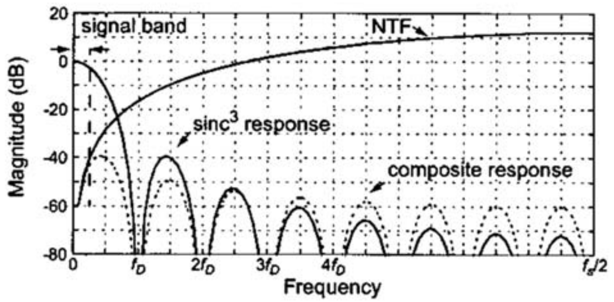


The first sinc^3 (sinc^K in general) stage is clocked at f_s and is used to reduce the sampling rate from f_s to an intermediate frequency f_D. The second stage (FIR or IIR) suppresses the remaining noise between f_B and $f_D/2$, allowing a further reduction of the sampling rate to $2f_B$.

The second filter may be implemented as a cascade of so-called halfband FIR filters, and may also incorporate compensation for the in-band droop of the sinc^K filter

Decimation

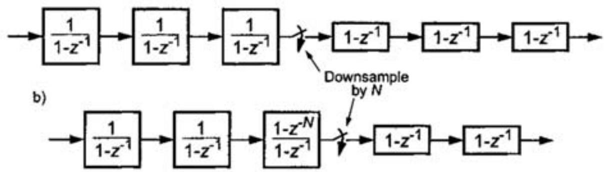
Below: composite frequency response of NTF + sinc^3 first stage of the decimator, with $f_D=4f_B$.



Lower f_D → speed and complexity reduction for the second filter, but if $f_D/2f_B$ drops below 4, the droop of sinc^3 at f_B becomes large; also, too much noise may be folded into the baseband, i.e., the second assumption for good decimation (see previous two slides) may not be fulfilled → intermediate OSR of 4 seems to be optimal

Hogenauer sinc^3 architecture

$$\boxed{H(z)=\frac{1}{N^3}\left(\frac{1-z^{-N}}{1-z^{-1}}\right)^3=\frac{1}{1-z^{-1}}\frac{1}{1-z^{-1}}\frac{1}{1-z^{-1}}\frac{1-z^{-N}}{N}\frac{1-z^{-N}}{N}\frac{1-z^{-N}}{N}}$$



Accumulators and differentiators!

The N-period delay in the differentiators is implemented with a single-delay block; N is usually a power of 2 and therefore division by N does not require any arithmetic operation

Accumulators will overflow, but with wrap-around arithmetic (such as 2's complement) the correct output will be obtained!

For M-bit input → M + K $\log_2 N$ bits (here, K=3) at all nodes are sufficient (fewer bits can be used in the internal nodes because of noise shaping!)