Data Converters

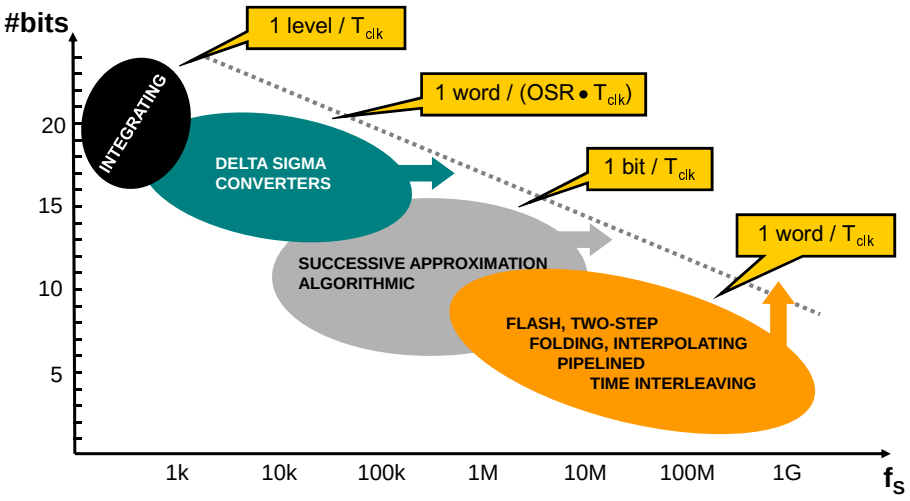
Oversampling and Low-Order ΔΣ Modulators

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

Overview

- Principle of oversampling
- Noise shaping
- 1st-order ΣΔ modulator
- 2nd-order ΣΔ modulator
- Effect of op-amp non-idealities

Speed vs. accuracy of ADCs



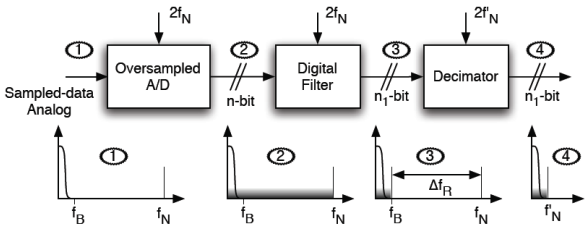
Principle of oversampling

Key feature: if signal band occupies only a small fraction of Nyquist, it is possible to remove the large fraction of quantization noise outside the signal band, improving the SNR; assuming white q-noise spectrum, we obtain

$$V_{n,B}^2 = \frac{\Delta^2}{12}\frac{2f_B}{f_s} = \frac{V_{ref}^2}{12\cdot 2^{2n}}\cdot\frac{1}{OSR}$$

where OSR= $(f_s/2)/f_B$ is the oversampling ratio. The ENOB becomes

$$ENOB = n + 0.5\log_2 OSR$$

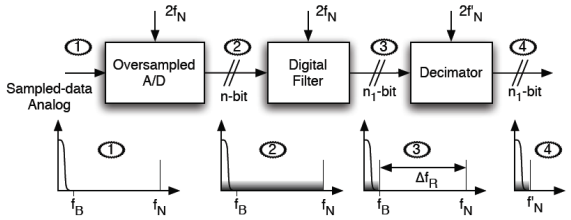


Principle of oversampling – II

An increase of the OSR by 4 yields one extra bit in resolution – not dramatic; however, if oversampling is used to relax the anti-aliasing filter, the improvement comes for free!

Oversampling is very effective in the analog world, but is a waste of power in the digital → sampling rate is reduced by decimation

Decimation by *k* → one out of *k* samples is used → equal to down-sampling → high-frequency regions of the Nyquist band are aliased into the "reduced-by-*k*" base-band → digital noise at those high frequencies must be filtered off to obtain the SNR improvement – such anti-aliasing filter (prior to decimation, running at the ADC frequency) would be needed anyway to remove the HF noise

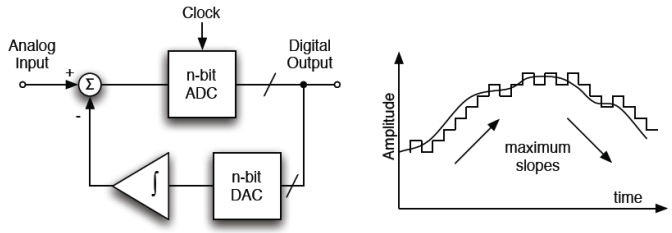


Delta modulator

Originally: oversampling not for q-noise spreading, but for improving pulse-code modulation (PCM) → high sampling rate to transmit the change (delta) between samples instead of the whole sample

Below → delta modulation if 1-bit, differential PCM with multi-bit → sampling rate and quantization step should be large enough to allow tracking

No significant info at the output for DC signals → high-pass response



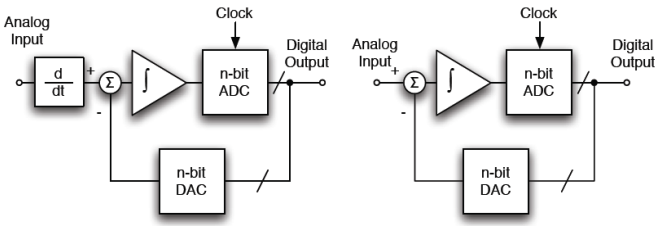
Evolution

Left → equivalent to previous algorithm; Right → derivative at input has been removed → integrator operates on signal error, not on estimated signal → response changed from high-pass to low-pass

Right → algorithm performs an integration (sum, *sigma*) of the difference (*delta*) at its input → ΣΔ (or ΔΣ) modulator

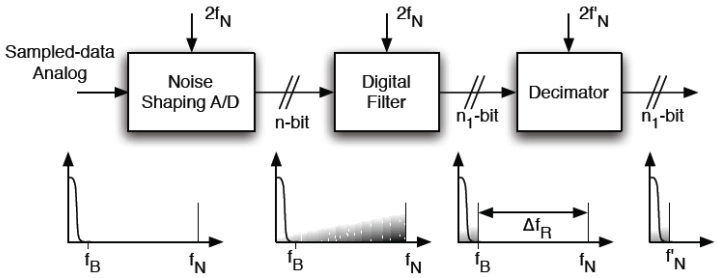
More exactly, this is a first-order ΣΔ modulator, as it uses only one integration

The key advantage of ΣΔ modulators is that they shape q-noise, greatly improving the SNR



Noise shaping

Oversampling becomes more effective if we can shift most of the q-noise towards high frequencies (where it can be filtered off), decreasing it in the signal band → noise shaping → SNR largely improved → the ENOB can greatly exceed what would be allowed in terms of pure component matching



Noise shaping – II

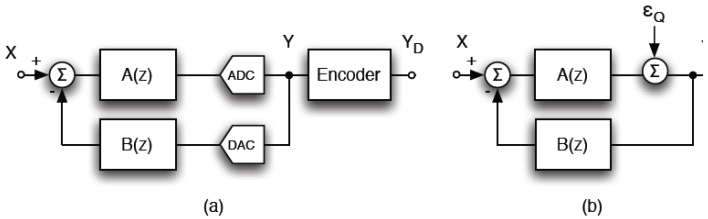
General model (left) → linearized with additive q-noise (right)

$$\left[X - Y\cdot B(z)\right]A(z) + \varepsilon_Q = Y \quad \rightarrow \quad Y = \frac{X\cdot A(z)}{1+A(z)B(z)} + \frac{\varepsilon_Q}{1+A(z)B(z)}$$

signal transfer function (STF) — noise transfer function (NTF)

$$Y = X\cdot S(z) + \varepsilon_Q \cdot N(z)$$

STF should be low pass, and NTF high pass – often B=1 → A must be integrator-like



First-order modulator

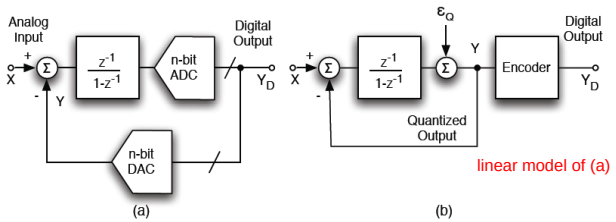
Integration → $H(z) = \frac{z^{-1}}{1-z^{-1}}$ (Euler-forward in this case)

$$Y(z) = \left[X(z) - Y(z)\right]\frac{z^{-1}}{1-z^{-1}} + \varepsilon_Q(z) \quad \rightarrow \quad Y(z) = X(z)\cdot z^{-1} + \varepsilon_Q(z)\cdot\left(1-z^{-1}\right)$$

STF is just a one-sample delay, while the NTF is

$$NTF(\omega) = 1 - e^{-j\omega T} = 2je^{-j\omega T/2}\frac{e^{j\omega T/2} - e^{-j\omega T/2}}{2j} = 2je^{-j\omega T/2}\sin(\omega T/2)$$

→ at low frequencies the NTF is very small (but *x*2 at maximum → *x*4 in power) → q-noise is high-frequency shaped!



First-order modulator and noise

Calling $v_{n,Q}^2$ the q-noise power spectral density, the q-noise power inside the band f_B is

$$V_n^2 = v_{n,Q}^2\int_0^{f_B} 4\sin^2(\pi fT)\,df \approx v_{n,Q}^2\int_0^{f_B} 4(\pi fT)^2\,df = v_{n,Q}^2\frac{4\pi^2}{3}f_B^3T^2$$

However, $v_{n,Q}^2 = \frac{\Delta^2}{12(f_s/2)}$, $T = \frac{1}{f_s}$ → $V_n^2 = \frac{\Delta^2}{12}\frac{\pi^2}{3}\left(\frac{f_B}{f_s/2}\right)^3 = \frac{\Delta^2}{12}\frac{\pi^2}{3}(OSR)^{-3}$

If the ADC has *k* thresholds (which means that the DAC generates *k*+1 levels between V_{ref} and 0), the quantization step is

$$V_{DAC}(i) = i\frac{V_{ref}}{k},\quad i = 0...k;\quad \Delta = \frac{V_{ref}}{k}$$

At full scale, we have $\frac{\Delta^2}{12} = \frac{V_{ref}^2}{12k^2}$, $V_{\sin}^2 = \frac{V_{ref}^2}{8}$, and maximum SNR is

$$SNR_{\Sigma\Delta,1} = \frac{12}{8}k^2\frac{3}{\pi^2}OSR^3$$

First-order modulator and noise – II

Setting $n' = \log_2 k$ = "extra bits", we obtain

$$SNR_{\Sigma\Delta,1}\big|_{dB} = 6.02n' + 1.76 - 5.17 + 9.03\log_2(OSR)$$

→ every doubling of the sampling frequency yields 1.5 bits

ADC output is binary → # of bits n_Q sent to the digital filter is the rounding of $\log_2(k+1)$ to the next integer

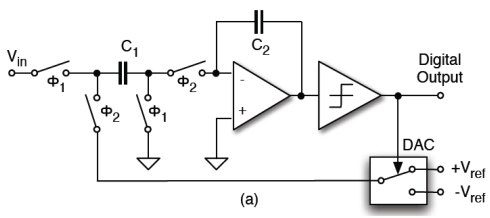
Table 6.1 - SNR improvement with Multi-level Quantizers

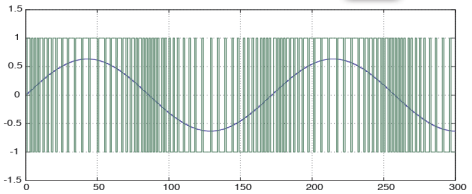
ADC Thresholds	DAC Levels	n_Q	n' extra bits	ΔSNR [dB]
1	2	1	0	0
2	3	2	1	6.02
3	4	2	1.58	9.54
4	5	3	2	12.04
5	6	3	2.32	13.97
6	7	3	2.58	15.56
7	8	3	2.81	16.84
8	9	3	3	18.03
15	16	4	3.91	23.52

1st order single-bit SC modulator

Samples the input during Φ_1, and injects the difference between input and DAC output during Φ_2; ADC is a comparator, DAC connects to either $+V_{ref}$ or $-V_{ref}$; the plot shows the ±1 output sequence for an input sine with amplitude 0.634 – output is mainly +1 (-1) when the input is close to maximum (minimum); when the input is close to 0, the two output states are equally represented – in general, the output looks very different from the input, but nevertheless the average of the bit stream follows the input

It is also intuitively clear that a large amount of high-frequency noise is generated by this output sequence

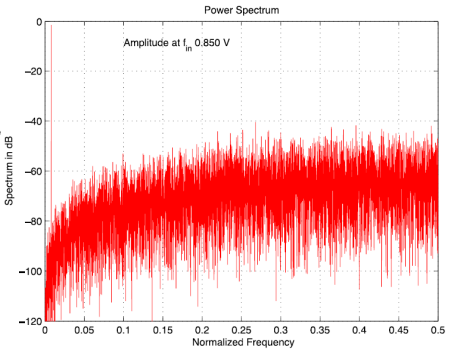




Example 6.1

V_{FS}=1V, 3b quantizer → $\Delta = 1/8$ → q-noise power of $\Delta^2/12 = 0.0013$ → with an FFT with 2^{14} samples, the power in each of the $2^{14}/2$ bins is $1.6 \cdot 10^{-7}$ → close to Nyquist, the power is 4 times higher

DC input → for some critical values, q-noise is not well shaped, but rather displays large tones with some shaping in between

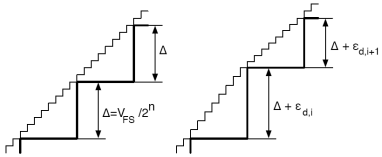


Qualitative considerations

1) integrator output is bounded only if input is (on average) zero → DAC output tracks (on average) the modulator input
2) While q-noise is zero at DC, the total q-noise power is actually doubled by shaping! (but, to repeat, most of it is filtered off)
3) Oversampling improves SNR adopting a sampling rate much higher than required by Nyquist → smart dynamic averaging performed on very many signal samples, disregarding higher frequencies
4) Input amplitude between two consecutive q-levels → output changes between these two levels, in such a way as to give an average output equal to the input – it does so (hopefully) without repeated patterns, since the input changes during the conversion – anyway, the operation can be seen as an interpolation between the two levels – virtually, the modulator adds extra steps in the input-output transfer

Qualitative considerations – II

5) if DAC non-linearity affects two (large) consecutive steps → resolution is still very good, but linearity does not improve (see right)
6) any limit affecting the digital signal produced by the ADC (e.g., noise and errors on the thresholds) is much alleviated by the feedback loop – indeed, the ADC output must be referred to the integrator input, and then to the modulator input → divided by the integrator gain, very high in the band of interest
7) this is not true for the DAC, which is in the feedback path → errors injected directly at the modulator input → DAC linearity is not relaxed! (i.e. method reduces # of levels, but not their accuracy requirement) – 14 bits often targeted → DAC linearity is bottleneck



1-bit quantization

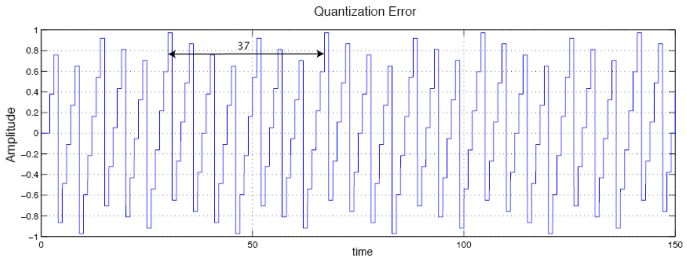
A line connecting many points is typically broken (i.e., non-linear), but a line connecting only two points is surely straight!! → if DAC has only two levels, no linearity problem arises → we need a 1-bit ADC (i.e. a comparator) and two reference levels (0 and V_{ref}, or $-V_{ref}$ and V_{ref})

However, problematic for two reasons: 1) q-step is as large as whole dynamic range, and converter relies only on OSR for high SNR → OSR must be very high; 2) one fundamental condition for assuming white q-noise, i.e. many q-levels, is not met → in fact, q-noise often appears concentrated at a few frequencies only, which may fall into the signal band

Quantization error and idle tones

Assume a first-order 1-b ΔΣ modulator with a DC input signal of amplitude $\Delta \cdot n/m$, where Δ is the quantization error and *n,m* are integers, *m>n* → the modulator output is a pattern of *n* 1's with a period of *m* clock cycles → spurious tones at f_s/m and its multiples → so-called *idle tones*

The quantization error is also periodic, see below, with *m*=37 and *n*=23 (V_{ref} = ±1V)



Dithering

Higher-order modulators and a busy input (as it normally happens, instead of a DC) make things less critical → however, the risk remains, especially for 1-bit quantization

Tones → limit cycles in the state space of the modulator (oscillations)

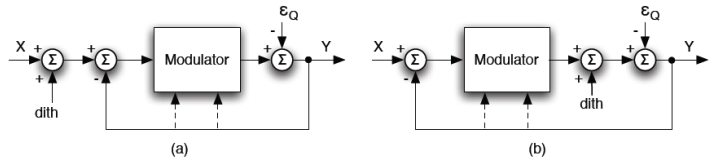
Noise → forces a chaotic behavior, may break limit cycles

Auxiliary input to inject a signal able to break the limit cycles (without affecting the SNR etc, at least ideally) → *dithering*

Two possibilities: 1) inject a small out-of-band sine/square wave, which is removed by filtering together with the quantization noise; this signal must be as low as possible, since it reduces the dynamic range at the input; 2) inject a noise-like signal, whose contribution should not degrade the SNR (shaped spectrum); the electronic noise may be sufficient by itself

Dithering – II

The dithering signal is usually a bipolar signal, $\pm V_{dith}$, with constant amplitude and sign controlled by a pseudo-random bit-stream generator



a) injection at input → necessary to shape the bit stream with a high-pass filter $\left(1-z^{-1}\right)^p$

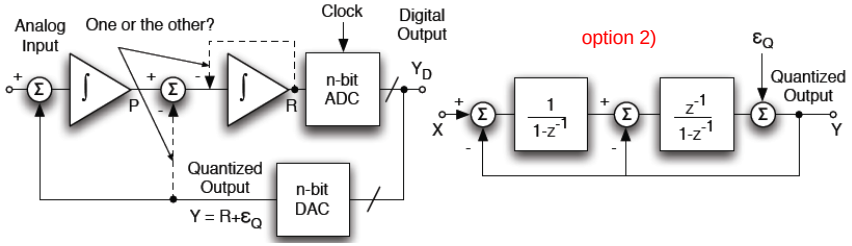
b) injection at output → the bit stream is shaped by the modulator itself; since the power of dither is $V_{dith}^2/12$ (as it is white-noise-like), it is enough to use a dither amplitude $V_{dith} < \Delta$

2nd order modulators

1st order → 1.5b for an OSR doubling, and sometimes large noise tones → we can do better with 2nd order → two cascaded integrators cause instability → one must be damped – two options: 1) conventional approximated integrator; 2) longer path that includes quantizer → option 1) and 2) yield respectively

1) $$R=\frac{P-R}{s\tau} \quad \rightarrow \quad Y=R+\varepsilon_Q=\frac{P}{1+s\tau}+\varepsilon_Q$$

2) $$Y-\varepsilon_Q=\frac{P-Y}{s\tau} \quad \rightarrow \quad Y=\frac{P}{1+s\tau}+\frac{s\tau\varepsilon_Q}{1+s\tau}$$



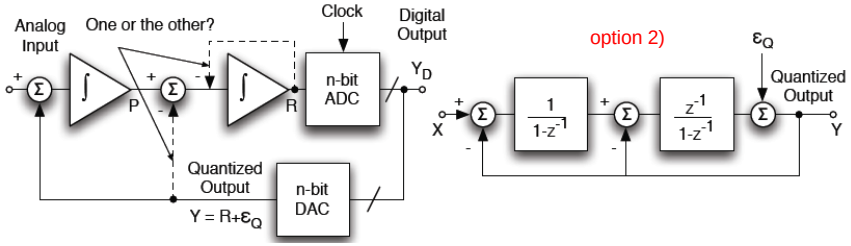
2nd order modulators – II

Thus, $$Y=\frac{P}{1+s\tau}+\varepsilon_Q \quad \text{or} \quad Y=\frac{P}{1+s\tau}+\frac{s\tau\varepsilon_Q}{1+s\tau}$$

In the first case, q-noise is left unchanged; in the second, it is high-pass filtered; since the other integrator introduces one more zero, the second circuit secures a double zero in the NTF → very advantageous

Circuit on the right → two different integrator, with and without delay → optimal STF

$$\left[(X-Y)\frac{1}{1-z^{-1}}-Y\right]\frac{z^{-1}}{1-z^{-1}}+\varepsilon_Q=Y \quad \rightarrow \quad Y=X\cdot z^{-1}+\varepsilon_Q\left(1-z^{-1}\right)^2$$



2nd order modulators – III

STF → simple delay; NTF → square of the 1st order NTF, as expected
If the integrators have gain errors → imperfect cancellations in the previous equation → parasitic denominators appear in both STF and NTF → negligible for small gain errors

NTF on unit circle is $NTF(\omega)=\left(1-e^{-j\omega T}\right)^2=-4e^{-j\omega T}\sin^2(\omega T/2)$, and the noise power becomes (assuming again $\omega T/2$ small)

$$V_n^2=v_{n,Q}^2\int_0^{f_B}16\sin^4(\pi fT)\,df\approx v_{n,Q}^2\int_0^{f_B}16(\pi fT)^4\,df=v_{n,Q}^2\frac{16\pi^4}{5}f_B^5T^4$$

$$v_{n,Q}^2=\frac{\Delta^2}{12(f_s/2)}, \quad T=\frac{1}{f_s} \quad \rightarrow \quad V_n^2=\frac{\Delta^2}{12}\frac{\pi^4}{5}\left(\frac{f_B}{f_s/2}\right)^5=\frac{\Delta^2}{12}\frac{\pi^4}{5}(OSR)^{-5}$$

$$SNR_{\Sigma\Delta,2}=\frac{12}{8}k^2\frac{5}{\pi^4}OSR^5 \qquad SNR_{\Sigma\Delta,2}\big|_{dB}=6.02n'+1.76-12.9+15.05\log_2(OSR)$$

Although we start with a "loss" of 12.9dB, every doubling of the OSR yields a 2.5b improvement in the SNR → great!

Circuit design issues – op-amp offset

The offset of the first integrator and of the 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 ore more integrators when it is referred to the input → negligible impact → opens up the possibility of positioning the ADC thresholds at optimal voltage levels

Circuit design issues – finite op-amp gain

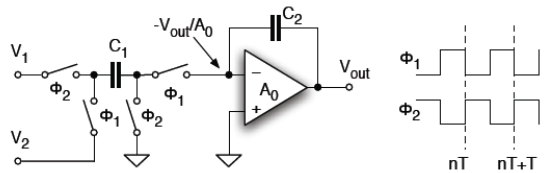
The DC gain of the op-amp is not infinite → we obtain

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

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

→ 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

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

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

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

$$\frac{1+A_0}{2+A_0}e^{-s_cT}=1 \quad\rightarrow\quad e^{s_cT}=\frac{1+A_0}{2+A_0} \quad\rightarrow\quad s_cT=\ln\left(\frac{1+A_0}{2+A_0}\right)$$

$$\omega_cT=-s_cT=\ln\left(\frac{2+A_0}{1+A_0}\right)=\ln\left(1+\frac{1}{1+A_0}\right)\approx\frac{1}{1+A_0}$$

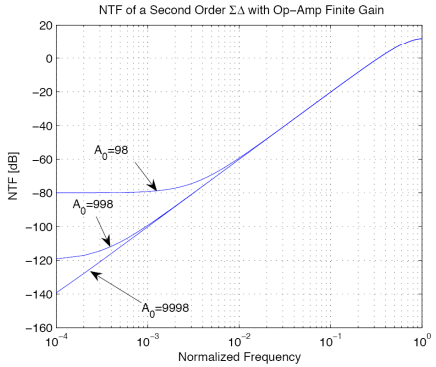
$$f_c=\frac{1}{2\pi T}\frac{1}{1+A_0}=\frac{f_s}{2\pi}\frac{1}{1+A_0}$$

Finite op-amp gain – III

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

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

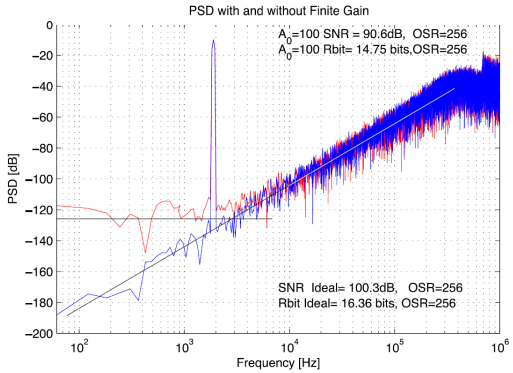
resulting in a very relaxed op-amp gain demand for modulators with medium OSR

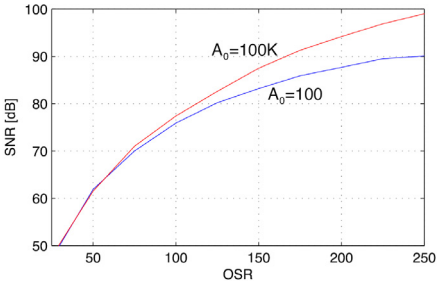


Circuit design issues – finite op-amp gain

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

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





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

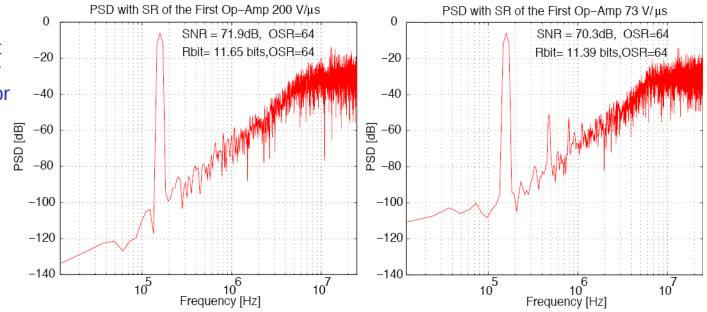
All these equations can be used in a behavioral simulator to enormously speed up the study of the combined impact of finite bandwidth and finite slew rate for the op-amp

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 $SR_1>\Delta V_{out,1}/(T/2)\approx 75V/\mu s$ $SR_2>\Delta V_{out,2}/(T/2)\approx 321V/\mu s$

Ideally, SNR=72dB with op-amp's βf_T=100MHz, OSR=64, f_{in}=160kHz
if $SR_1=325V/\mu s$, $SR_1=78V/\mu s$, 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



ADC/DAC non-idealities

Static/dynamic limitations on the ADC degrade performances:

$$V_{ADC,out}=V_{ADC,in}+\varepsilon_Q+\varepsilon_{ADC}$$

However, the modulator shapes ε_{ADC} as well → if $\varepsilon_{ADC}<\varepsilon_Q$, which is easily accomplished, the ADC does not limit the overall performances

The DAC, on the other hand, lies in the feedback path → its non-idealities are at the modulator input → not shaped

DAC non-linearity is a big concern → 1-bit DAC is inherently linear

The DAC is often implemented with switched capacitors → kT/C issue

If we assume that the sampled noise is white up to Nyquist, the minimum value for C_{in} in a 2nd-order modulator is (assuming that out-of-band noise is filtered off)

$$v^2_{n,kT/C}=\frac{kT}{OSR\cdot C_{in}}<\frac{V^2_{ref}}{12\cdot k^2}\frac{\pi^4}{5}\frac{1}{OSR^5}$$

Single-bit vs. multi-bit

High SNR with single-bit ΔΣ → high-order modulators (stability issue) and/or high OSR

High OSR, and bandwidth of the op-amps has to be higher than clock frequency → ok for audio or instrumentation applications

Usable V_{ref} with single-bit is a small fraction of the supply voltage, since the swing at the op-amp outputs is rather large

Assuming that the dynamic range at the op-amp output is αV_{DD}, and that a -6dB$_{FS}$ sine gives rise to a swing of $\pm\beta_{swing} V_{ref}$ at the output of the first integrator → the maximum V_{ref} is then given by

$$|V_{ref}| < \frac{\alpha V_{DD}}{2\beta_{swing}}$$

For low supply voltages, α may be only ≈0.7 and $\beta_{swing} = 2$, resulting in

$$|V_{ref}| = 0.175 V_{DD}$$

Single-bit vs. multi-bit – II

Such a low value of V_{ref} is problematic, because of the constraints on the kT/C noise and op-amp thermal noise ($\gamma kT/C_L$), especially for the first op-amp → 1-bit quantization is convenient only with medium-high supply voltages

Slew-rate issue → input of first integrator is the difference between analog input and DAC output; DAC output follows the input with an accuracy dependent on the DAC resolution (and input bandwidth) → reasonable to assume that the maximum difference is 2Δ → if 1-bit, this becomes $2V_{ref}$ → either very high SR, or low V_{ref} → with multi-bit, integrator input is reduced by the number of quantization levels

Multi-bit → additional power in ADC

However, increasing the resolution by 2.5 bits in a second-order modulator requires doubling the clock frequency → optimal use of power entails a trade-off between increased speed in op-amps and more comparators in quantizer

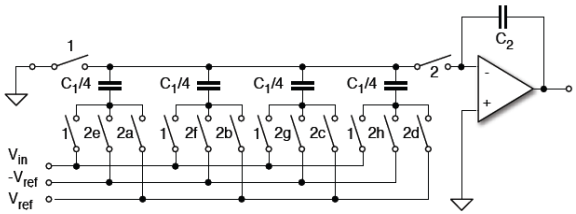
Single-bit vs. multi-bit – III

Rule-of-thumb: power used by comparator is 1/20 that used by op-amp, operated at the same speed

More comparators also means more complexity, multi-bit digital signal processing in the decimator filter, and extra logic for digital calibration and dynamic element matching (if needed)

Typically, 3 to 15 comparators are used

Multi-bit DAC → usually implemented as a capacitive MDAC

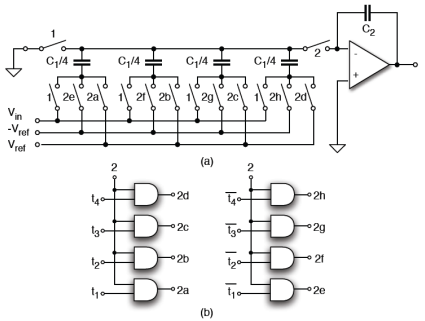


Multi-bit DAC

Below: 2-b MDAC → C_1 is split into 4 units, pre-charged to the input signal during Φ_1, and connected to $+V_{ref}$ or $-V_{ref}$ under the control of the thermometric code t_1-t_4 during Φ_2

C1 is used for both input signal and feedback → good, feedback factor for the op-amp is not decreased as it would with separate capacitors

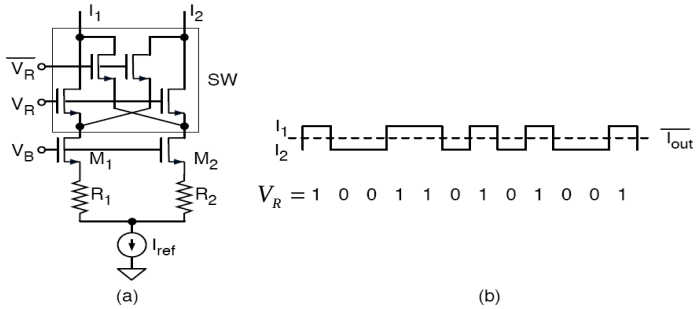
Drawback → charge delivered by Vref is a non-linear function of the input: if the control of the DAC is $k(n) \approx V_{in}(n-1)/\Delta$, then $k(n)$ capacitors already charged to $V_{in}(n)$ are connected to V_{ref} The output resistance of V_{ref} must be very low, to avoid distortion in the delivered charge Q_{ref} :



$$Q_{ref}(n) = k(n)\left[V_{ref} - V_{in}(n)\right]$$

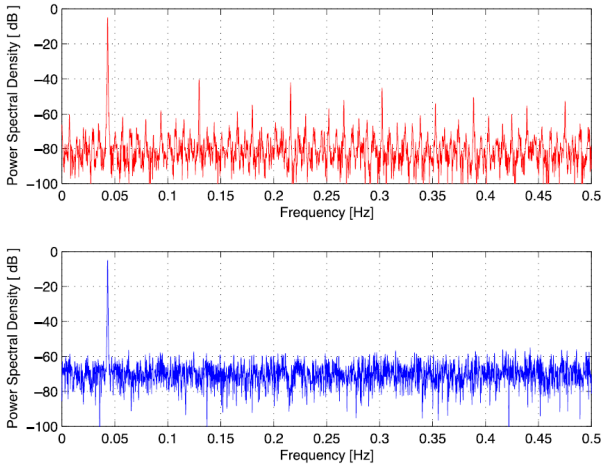
Dynamic element matching (DEM)

Components are made equal on average, instead of performing a static correction → good for cancelling temperature and aging effects – below: I_{ref} is split into two equal parts by M_1 and M_2, R_1 and R_2 improve matching by reducing the impact of the MOS threshold mismatch - however, resistor mismatch impacts as well → the four switches multiply I_{ref} on average 50% of the time with +1, and 50% with –1 with a pseudo-random sequence → mismatch becomes noise like – if only a fraction of Nyquist is used, noise shaping improves further the technique



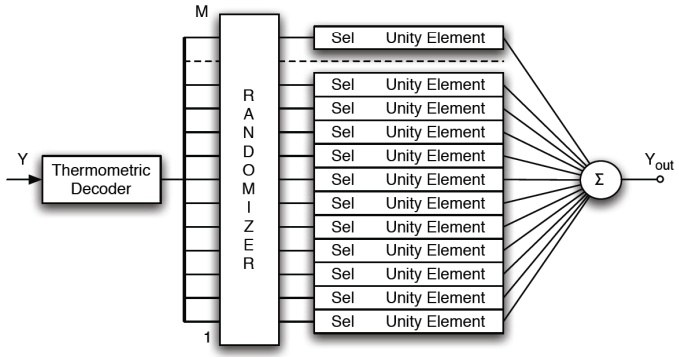
DEM – example 8.1

7-b DAC, binary weighted elements with current splitting as in previous slide, matching with large variance to make impact more clear → DEM reduces the tones due to INL, but these tones are turned into noise → DEM increases the noise floor, as is clear from the simulations below



Butterfly randomization

Control of DEM in DACs with thermometric selection of unit elements can be problematic → typically, randomization as below: randomizer receives N thermometric 1s out of M input lines, and generates a scrambled set of M controls, N of which are 1s – the number of possible scrambled outputs is M! → huge number: 5040 for M=7, and 3,628,800 for M=10 → however, this is overkill; it is enough to avoid frequent repetition of the same (or similar) code

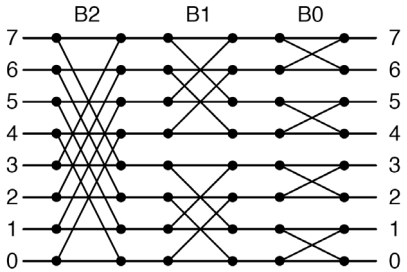


Butterfly randomization – II

A simple solution is to use an M-port barrel shifter which rotates one increment every clock – more effective is the butterfly randomizer → the use of $\log_2$M stages (see below) ensures that any input can be connected to any output – more stages increase the number of possible connections – the control of the butterfly switches can use $\log_2$M bits from a k-bit random number generator, or, more simply, by the successive division by 2 of the clock (clocked averaging)

If the value of the N elements in the set is X_i, their average is $\overline{X} = \frac{1}{M}\sum_1^M X_i$ while the addition of N random elements yields

$$Y(N) = \sum_1^M d_i X_i$$



where d_i is 1 if X_i is selected – the error on Y is given by

$$\varepsilon_Y(N) = \sum_1^M d_i X_i - N\overline{X}$$

$$= \sum_1^M d_i X_i - \frac{N}{M}\sum_1^M X_i$$

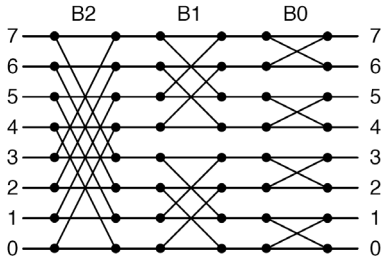
Randomization and noise

Assume that $X_i = \overline{X} + \delta X_i$, that the variance of δX_i is $\overline{X}^2 \sigma_X^2$, and that the various δX_i are uncorrelated with each other → the variance of the error becomes

$$\sigma_Y^2 = E\left\{\varepsilon_Y^2(N)\right\} = \left(N - \frac{N^2}{M}\right)\overline{X}^2 \sigma_X^2$$

dependent on input amplitude, zero for N=0 or N=M, and maximum for N=M/2

mismatch in space is transformed into mismatch in time → if randomizer works properly, trades discrete tones with additional white noise



Therefore, if all amplitudes are equally probable, the mismatch noise power is

$$P_{mism} = \frac{M}{6}\overline{X}^2 \sigma_X^2$$

Randomization and noise – II

The peak-to-peak amplitude of the output signal is $M\overline{X}$ → the power of a full-scale sine wave is $M^2\overline{X}^2/8$ → the SNR determined only by the mismatch error and OSR becomes

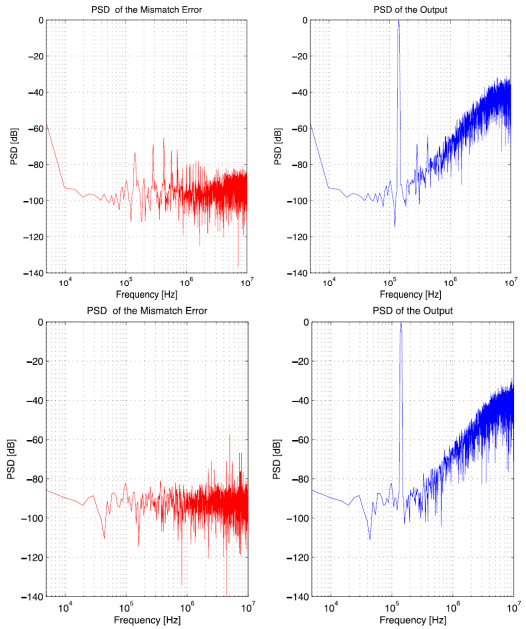
$$SNR = \frac{3M}{4\sigma_X^2} OSR$$

If M=8, OSR=1 (Nyquist-rate converter), and σ_X =2·10^{-3} → SNR=62dB
If M=8, OSR=32, and σ_X =2·10^{-3} → SNR=77dB

The white-noise assumption depends on how effective the randomizer is – with *b* butterfly stages, the clocked averaging repeats the same pattern every 2^b clock periods, introducing tones at $f_s/2^b$

– a pseudo-random number generator requires more hardware, but is more effective, especially when *b* is low

Example 8.2



2nd-order 3-bit ΣΔ with OSR=20 and 0.5% random mismatch in the 8 DAC elements → ideally, SNR=69dB with input = -2dB$_{FS}$

top: mismatches introduce non-linearities → tones clearly visible above the noise floor → SFDR≈60dB (unfiltered)

bottom: butterfly randomizer → tones are actually still present, but pushed higher up in frequency, where they are below the noise floor – however, the noise floor in the signal band has clearly increased → SNR≈SNDR is approx. 60dB

Randomization and noise – III

Randomization turns tones into white-like noise – however, the total error power caused by mismatches is not reduced → for Nyquist-rate converters, the SNDR remains almost constant, while the SFDR improves – for oversampled converters, the SNR improves, but only by 3dB for an OSR doubling, as in plain oversampled architectures

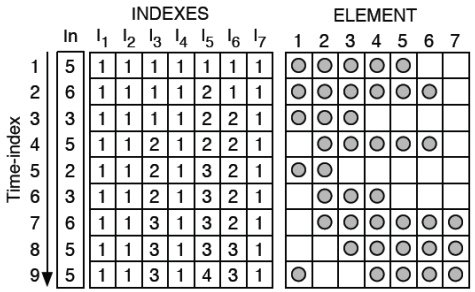
In ΣΔ converters, on the other hand, it would be very advantageous to shape the mismatch noise towards higher frequencies, where it can be filtered off together with quantization noise

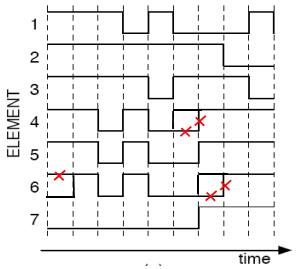
Basically, the approach to mismatch noise shaping is to use all the elements in the array in fast cycles, as this gives rise to high-frequency noise terms

Individual level averaging (ILA)

The goal is to use each of the M elements with equal probability for each digital input code – use of indexes $I_k(i)$, where k = input code, and i = time – the elements used when k is applied are those indexed by $I_k(i)$, $I_k(i)+1, \ldots, I_k(i)+k-1$ (with wrap-around when this exceeds M)

Rotation approach → I_k is increased by 1 every time code k is used – below, we see indexes and elements used with the input sequence {5 6 3 5 2 3 6 5 5} (all indexes start with value 1) – right: busy elements, good spreading of mismatches into white-like noise

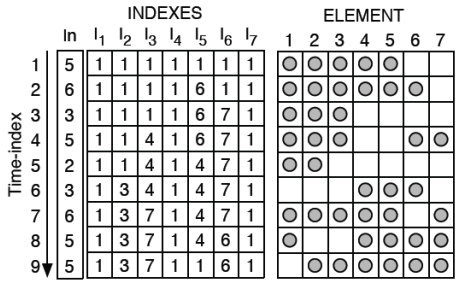


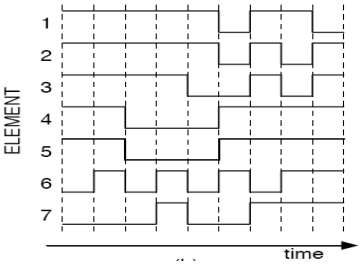


ILA – II

Addition approach → I_k is increased by k (modulo M) every time the code k is used – below, we see indexes and elements used with the input sequence {5 6 3 5 2 3 6 5 5}

All elements are even more busy than with the rotation approach – however, the effectiveness of the methods should be assessed via extensive computer simulations





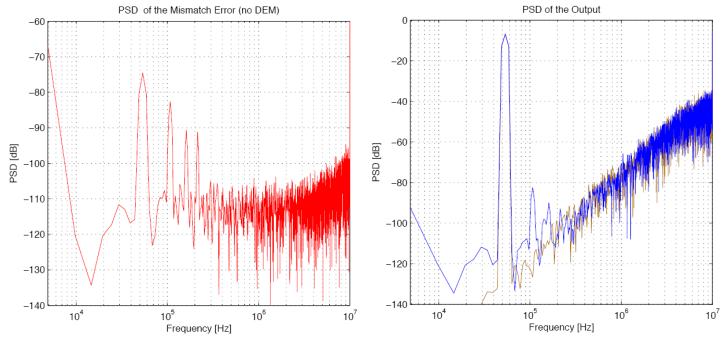
Example 8.3

2nd order 3-bit ΣΔ with OSR=64 → with input at -6dB$_{FS}$, we have ideally:

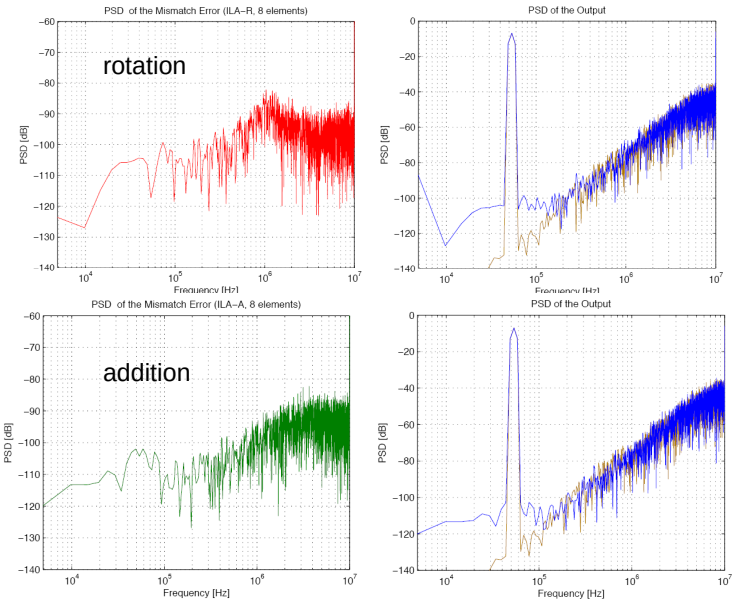
$$SNDR = -6 + \left(6.02 \cdot 3 + 1.76 - 12.9 + 15.05 \cdot \log_2\left(OSR\right)\right) \approx 91\text{dB}$$

→ a 0.2% mismatch results in more noise and discrete tones, with an SNDR=75dB (i.e., a deterioration as large as 20dB)

Next slide → both ILA methods remove the tones – however, the rotation methods achieves an SNDR of 84dB, while the addition method is more effective in shaping the noise, and yields SNDR=87dB

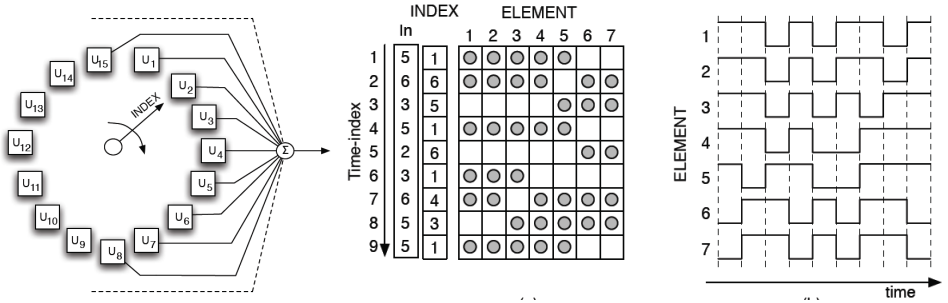


Example 8.3 – II



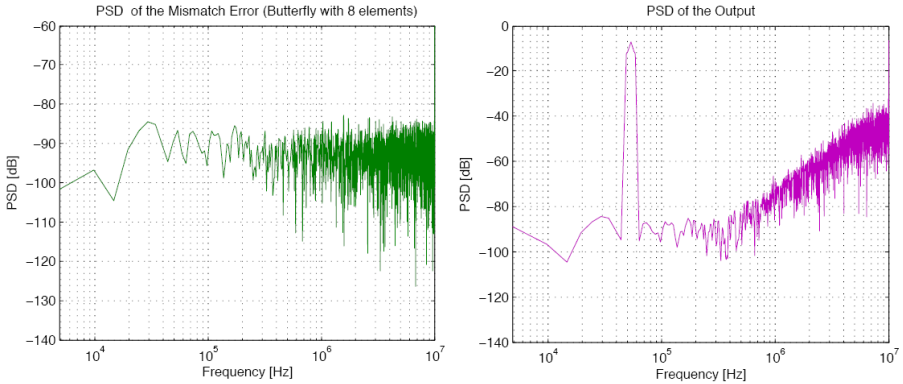
Data weighted averaging (DWA)

Uses only 1 index, updated by adding the new input code to its content → very fast, changes at every clock period – the same sequence {5 6 3 5 2 3 6 5 5} results in the indexing and element usage as below – very busy – both ILA and DWA perform noise shaping; however: simulations suggest that ILA is better for a small M, while DWA is better for M > 7



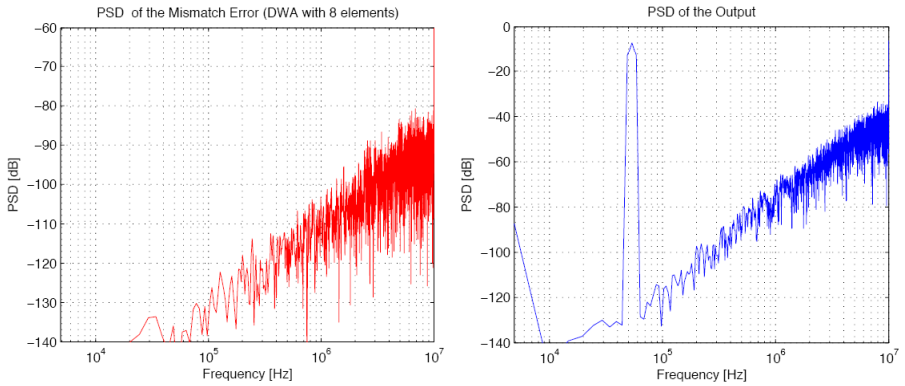
Example 8.4

2nd order 3-bit ΣΔ with OSR=64, f_S=20MHz → f_B=156kH → with input at -6dB$_{FS}$ and 0.4% mismatch, Butterfly randomization results in a flat spectrum up to 400kHz → very significant spectrum degradation → SNR=70dB



Example 8.4 – II

DWA → mismatch noise is 1st order shaped → 20dB/dec slope also in the signal band → no degradation of the SNR with respect to the ideal case with SNR=91dB! (compare the plots below with previous simulations referring to the same ideal converter)



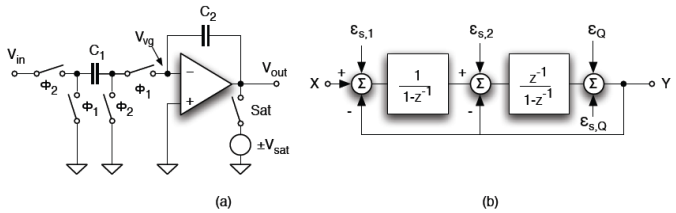
Integrator dynamic range – I

In general, both signal and q-noise are present in the modulator → the dynamic range of both integrators and quantizer must be larger than the reference

When the integrator output exceeds the op-amp dynamic range → loss of feedback, signal clipping, distortion

(a) below → if C_1 is still loaded with Q_{res} when V_{out} reaches saturation, the final charge on C_1 is $Q_{res} \cdot C_1/(C_1+C_2)$ → the input-referred voltage error becomes:

$$\varepsilon_s = \frac{Q_{res}}{C_1 + C_2}$$



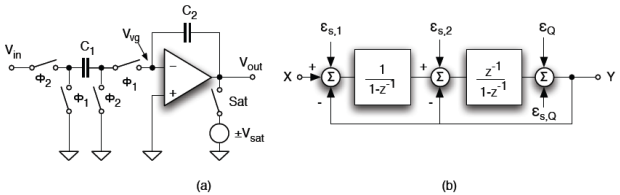
Integrator dynamic range – II

Error depends on how close to saturation the output is before each new charge transfer, and sign of charge → (almost) unpredictable → (hopefully) white spectrum

Exceeding the limits of the quantizer in the flash ADC (over-range or under-range) also gives a quantization error similar to the op-amp saturation → modeled as a white noise $\varepsilon_{s,Q}$

For the 2nd-order modulator in (b), we have in total

$$Y = Xz^{-1} + \varepsilon_{s,1}z^{-1} + \varepsilon_{s,2}\left(1-z^{-1}\right) + \left(\varepsilon_Q + \varepsilon_{s,Q}\right)\left(1-z^{-1}\right)^2$$



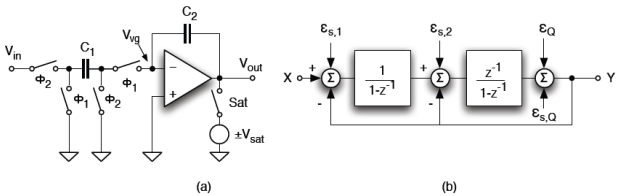
Integrator dynamic range – III

$$V_n^2 = \frac{V_{n,1}^2}{OSR} + V_{n,2}^2\frac{\pi^2}{OSR^3} + \left(V_{n,Q}^2 + \frac{\Delta^2}{12}\right)\frac{\pi^4}{5\cdot OSR^5}$$

$$V_{n,1}^2 = \varepsilon_{s,1}^2 \cdot f_B; \qquad V_{n,2}^2 = \varepsilon_{s,2}^2 \cdot f_B; \qquad V_{n,Q}^2 = \varepsilon_{s,Q}^2 \cdot f_B;$$

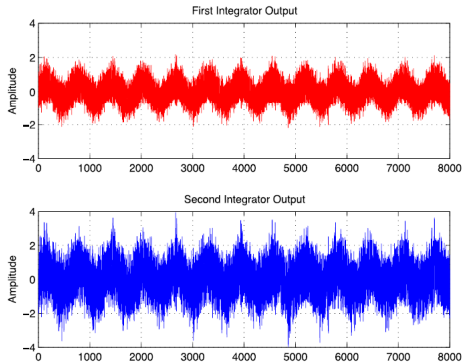
If OSR=64 → $V_{n,1}^2$ is reduced by 64, $V_{n,2}^2$ by 79682, and $V_{n,Q}^2$ by 55.100.000

Thus, saturation in the first integrator is most critical; over-range in the quantizer matters only when errors are comparable with Δ



Example 6.4 – I

Previous modulator, with 1b-DAC, V_{ref}=±1V, and a -6dB$_{FS}$ input → combination of signal + feedback determines max peaks as high as 2.18V and 3.96V (almost 4 times the reference)

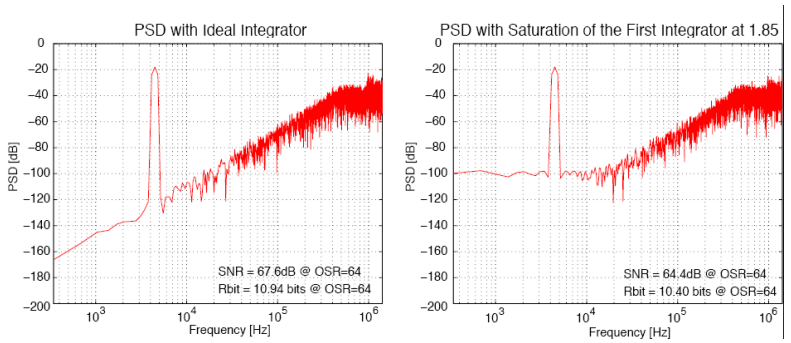


Example 6.4 – II

-10dB$_{FS}$ input → max. peaks still at 1.9V and 3.1V

Ideal modulator → SNR=67.6dB, not far from the 69.2dB predicted by equation on slide #19; this deterioration is caused by over-range in the quantizer; spectrum slope is 40dB/decade, as it should

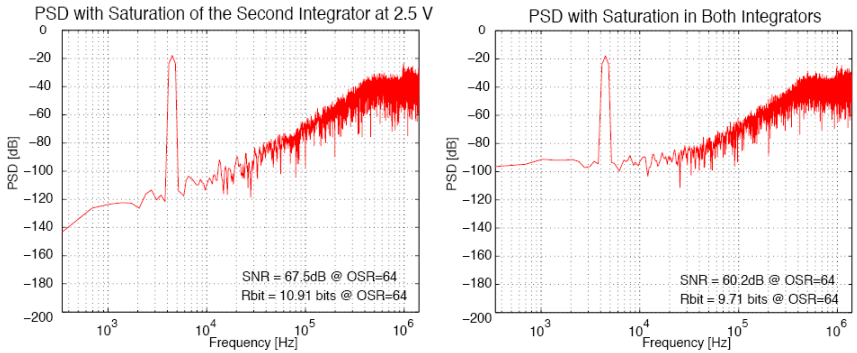
First integrator with saturation at 1.85V output → white noise floor appears, SNR drops to 64.4dB



Example 6.4 – III

Second integrator clipping at 2.5V → introduces white noise floor which is first-order shaped → 20dB/decade slope, SNR drops a negligible 0.2dB

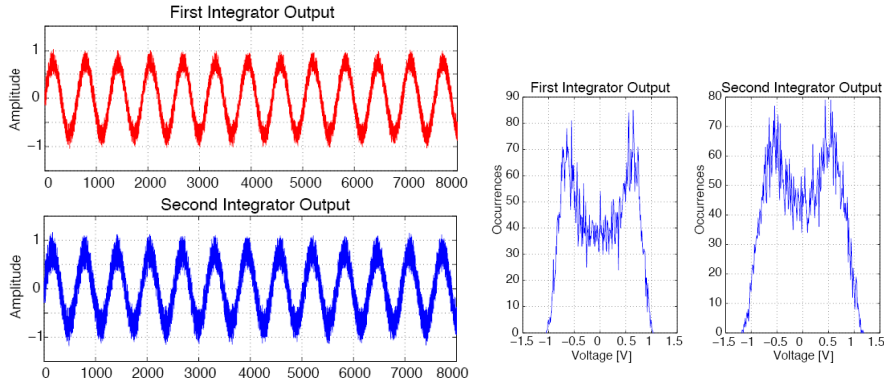
Finally, with both integrators clipping, the SNR drops by more than 5dB to 60.2dB



Example 6.4 – IV

Now, DAC with 7 thresholds, V_{ref}=±1V, and a -2.4dB$_{FS}$ input (0.758V) → max. peaks at 1.037V and 1.17V; histograms show the number of times the outputs reached a given max level

Simulated SNR of 94.0dB → ideally 93.6dB, given by the sum of 76.8dB (1-b DAC) plus 16.84dB = 6.02·log$_2$(7)

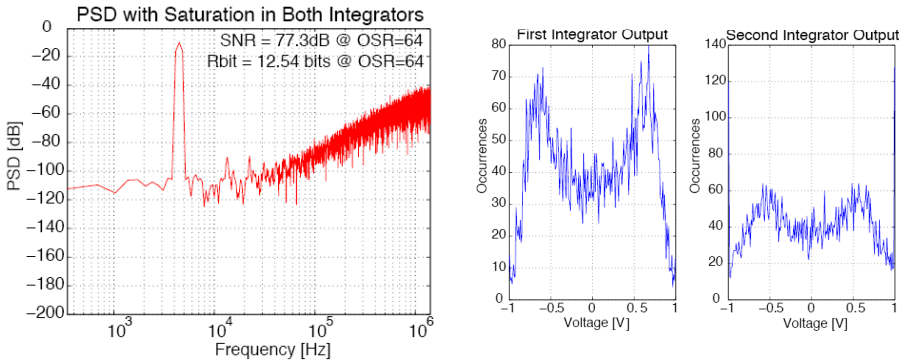


Example 6.4 – V

First amplifier clipping at 1V → SNR drops to 79.1dB

Both amplifiers clipping at 1V → SNR=77.3dB; further, IM3 and IM5 of approx. -80dBc

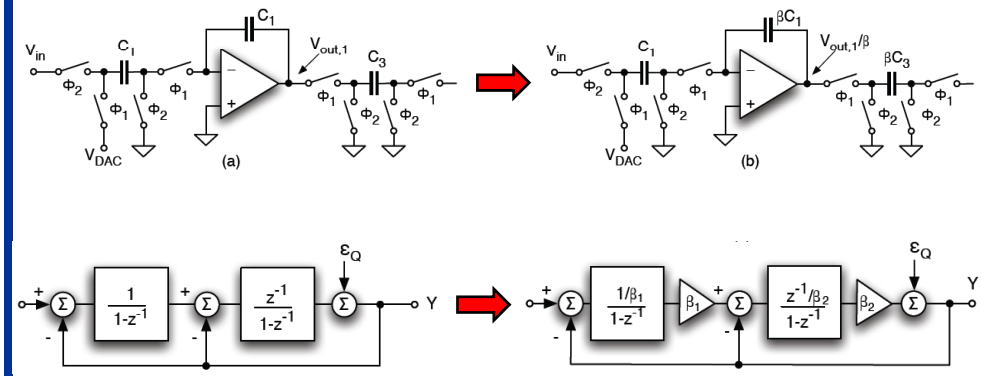
From histograms → saturations spreads out the signal distribution, compensating the reduced output range – also decorrelates (somewhat) input and output)



Optimization of dynamic range

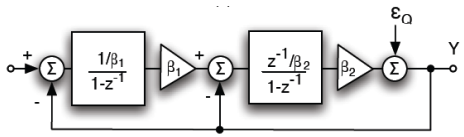
Dynamic range should be high enough to avoid clipping, but not too high, in order to minimize the electronic noise → solution: attenuation (or amplification) of the integrator output, compensated by an inverse amplification (or attenuation) at the input of the next stage(s)

Below: application of the principle in SC-design and in 2nd order modulator



Optimization of dynamic range – II

Scaling at the output of the second integrator → instead, ADC thresholds can be scaled down by β_2 (1-b ADC only detects zeros and scaling is not needed)



2nd-order modulator

In the 2nd-order modulator below, both integrators are delaying 1 clock cycle → benefit of 1 extra clock period for the feedback signal

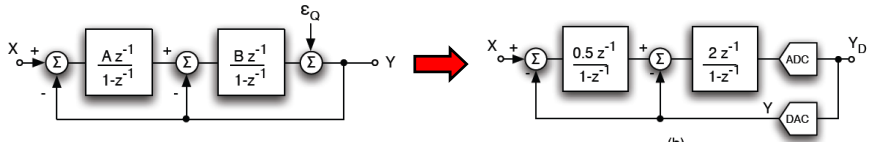
Circuits analysis yields

$$\left[(X-Y)\frac{Az^{-1}}{1-z^{-1}}-Y\right]\frac{Bz^{-1}}{1-z^{-1}}+\varepsilon_Q=Y \quad \Rightarrow \quad Y=\frac{XABz^{-2}+\varepsilon_Q\left(1-z^{-1}\right)^2}{1-(2-B)z^{-1}+(1-B+AB)z^{-2}}$$

Signal gain =1 if AB=1; if then B=2 (i.e., A=1/2), denominator=1, we obtain

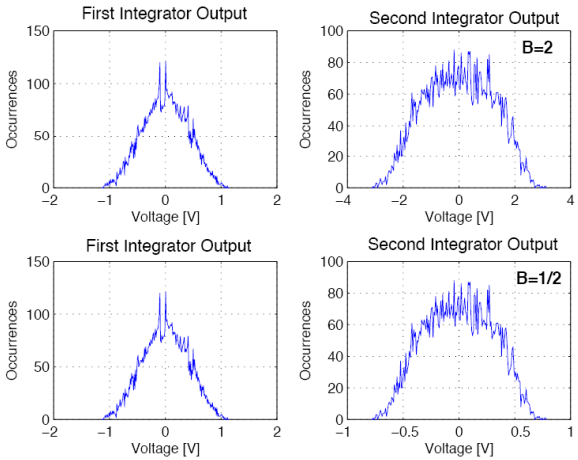
$$Y=Xz^{-2}+\varepsilon_Q\left(1-z^{-1}\right)^2$$

which is the optimal transfer function already found, apart from an extra delay on the signal path



2nd-order modulator - simulations

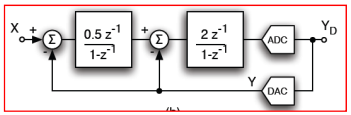
1b-DAC, OSR=64, V_{ref}=±1V, -10dB_{FS} input, A=1/2, B=2 or 0.5 → as expected, the dynamic range at the output of the second integrator is reduced by a factor 4



2nd-order modulator – dynamic range

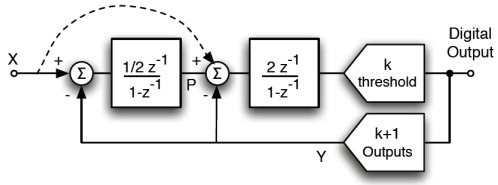
The output P of the first integrator is given by

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



With a multi-level DAC, P is dominated by the first (signal) term (if the signal is large), since the second term is at most as large as Δ

Feedforward can be used in multi-level modulators to reduce the dynamic range of P, as in the architecture below



2nd-order modulator – dynamic range – II

The feedforward branch is expressed, referred to the input, as $2X\left(1+z^{-1}\right)/z^{-1}$

The output become then

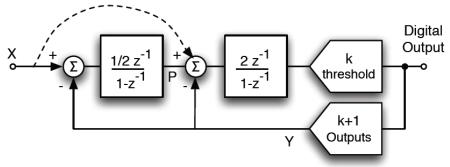
$$Y = X = \left(z^{-2} + 2z^{-1}\left(1-z^{-1}\right)\right) + \varepsilon_Q z^{-1}\left(1-z^{-1}\right)^2$$

and P is now

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

which shows that P is much reduced, since Z is high-pass filtered, which gives rise to a large attenuation in the signal band. The STF shows now a high pass term, which is however usually negligible:

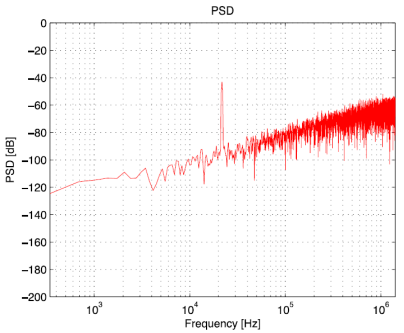
$$STF = z^{-2} + 2\left(1-z^{-1}\right)$$

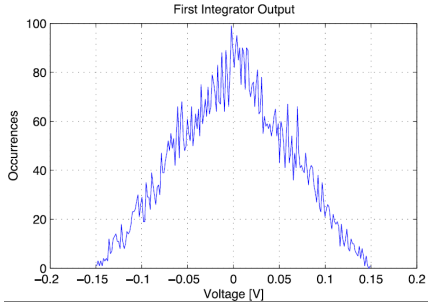


Example 6.6

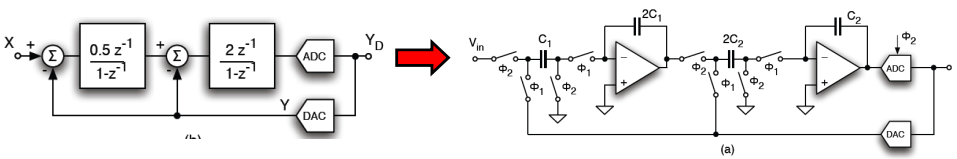
7-comparator DAC, OSR=64, V_{ref}=±1V, -3dB$_{FS}$ input → SNR=93dB, almost unchanged by feedforward. However, now the output of the first integrator is very low, see below.

Very close to the bandwidth limit (f_s/128.3) the signal gain is only 0.02dB higher than unity





SC circuit implementation



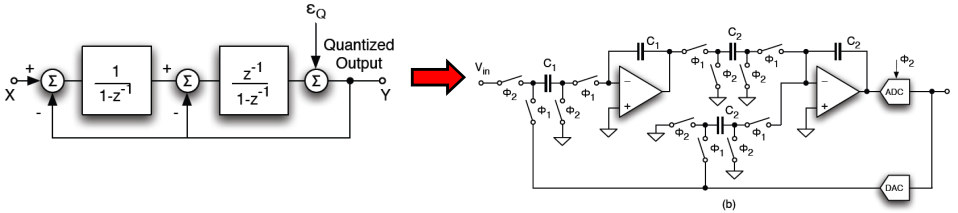
Both integrators inject the charge into the virtual ground at the beginning of Φ_1

Integrators have Φ_1 to settle; sampling occurs during Φ_2

Subtraction of signal and DAC feedback is obtained for both integrators by pre-charging in a non-inverting way the sampling capacitors during Φ_2, while the DAC signal sees an inverting integration

Easy to check that there is a delay of one sampling period in the loop going from the output of the second integrator to the input of the same integrator, while there is a delay of two sampling periods along the outer loop → correct implementation of the block circuit

SC circuit implementation – II



Here, delay of only one sampling period along the outer loop, since the first integrator immediately samples and injects the DAC feedback into the second integrator (upper SC circuit)

The ADC lathes are activated by the rising edge of Φ_2, leaving this entire phase for the digital conversion and the pre-setting of the DAC

Limitation: the two op-amps work in cascade → limits the max. clock sampling frequency

Feedback factor is 1/2 for both integrators; in the previous modulator, it is 2/3 and 1/3 → op-amps with different gain-bandwidths

Noise analysis

Electronic noise in any ΔΣ modulator is caused by the op-amps noise and by the kT/C noise in the capacitors

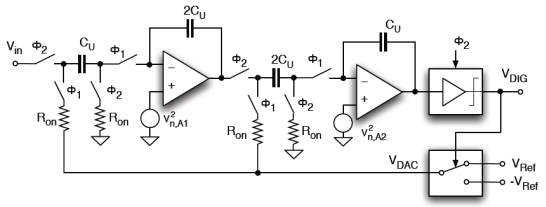
The noise injected in each capacitor during each of the two phases must be calculated (colored noise spectra in general)

The following sampling results in almost white spectra, because of noise folding into the base band

The superposition of the noise power of all noise sources, integrated over the signal band, yields the total noise power

Noise calculations

2nd-order ΔΣ modulator with two delaying integrators



One on-resistance for each pair of switches is included; the input-referred white noise of the op-amps is

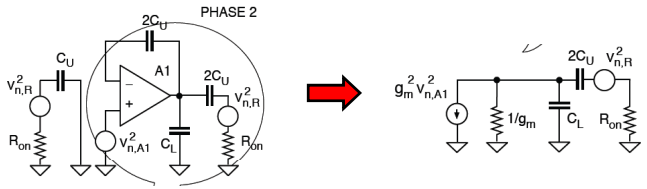
$$v_{n,A1}^2 = \gamma_{A1}\frac{4kT}{g_{m,A1}} \qquad v_{n,A2}^2 = \gamma_{A2}\frac{4kT}{g_{m,A2}}$$

1) During Φ_2: the signal is sampled on C_U → the noise power on C_U is

$$v_{n,R}^2 = \frac{kT}{C_U}$$

Noise calculations – II

2) The output of the first op-amp charges the input cap. of the second op-amp (2 C_U). The first op-amp (A1) is in unity-gain configuration during Φ_2 → the equivalent model is the following



where $g_m = g_{m,A1}$ is the output conductance as well. The transfer function from input-referred noise to colored noise across $2C_U$ is

$$H_{A1,in2} = \frac{v_{n,2C_U}}{v_{n,A1}} = \frac{1}{1+s(\tau_0 + 2\tau_0 C_U/C_L + \tau_R) + s^2\tau_0\tau_R}$$

where

$$\tau_0 = \frac{C_L}{g_m}, \qquad \tau_R = 2C_U R_{on}$$

Noise calculations – III

3) Two poles → if R_{on} is small and $2C_U/C_L<1$, the dominant pole is at

$$\omega_T = g_m/C_L$$

and the noise power across $2C_U$ is

$$V_{n,A1,in2}^2 = \gamma_{A1}\frac{kT}{C_L}$$

4) if $2C_U/C_L>1$, the dominant poles moves at slightly lower frequencies and improves noise shaping – benefit not larger than 1dB, though

5) The noise spectrum $v_{n,R}^2$ if filtered by the transfer function

$$H_{R,in2} = \frac{1+\tau_0}{1+s(\tau_0 + \tau_0 2C_U/C_L + \tau_R) + s^2\tau_0\tau_R}$$

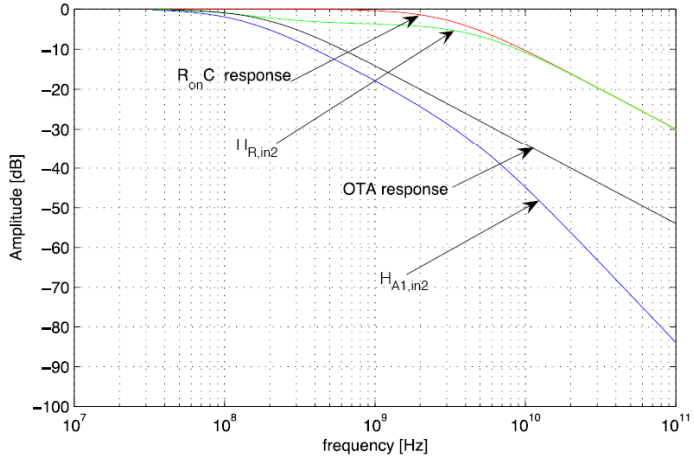
If $2C_U/C_L<1$, zero and dominant pole cancel out, and leave the other pole at $\tau_R = 2C_U R_{on}$, resulting in the noise power

$$V_{n,R,in2}^2 = \frac{kT}{2C_U}$$

Plots

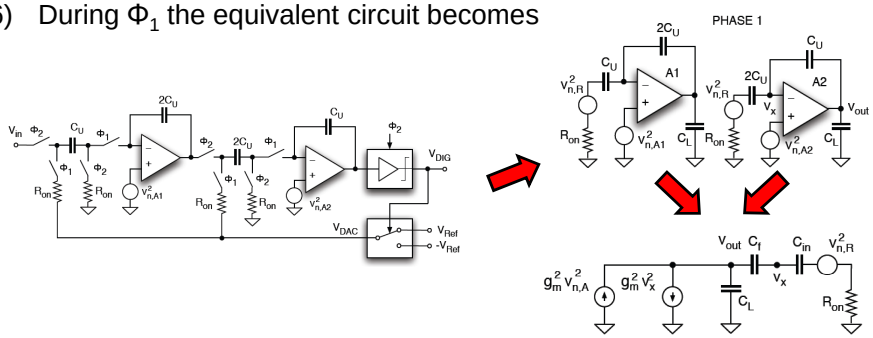
$$f_T = 200MHz, \quad C_L = 1pF, \quad 2C_U = 0.5pF, \quad R_{on} = 100\Omega$$

In this case, $2C_U/C_L$=0.5, and HR,in2 shows a somewhat flat region → slight noise improvement (1-2dB at most)



Noise calculations – IV

6) During Φ_1 the equivalent circuit becomes



where the small-signal circuits applies to both integrators, where v_x is the voltage at the input of the op-amp. Nodal analysis yields

$$g_m\left(v_{n,A}-v_x\right)=v_{out}sC_L+\left(v_{out}-v_x\right)sC_f \qquad v_{C_{in}}=\left(v_x-v_{out}\right)\frac{C_f}{C_{in}}$$

$$\left(v_{out}-v_x\right)C_f s+v_{n,R}\left(\frac{C_{in}s}{1+R_{on}C_{in}s}\right)=v_x\left(\frac{C_{in}s}{1+R_{on}C_{in}s}\right)$$

Noise calculations – V

which result in

$$v_{C_{in}}=\frac{C_L}{C_f}\frac{-v_{n,A}+\left(1+\tau_0\right)v_{n,R}}{1+s\left(\tau_0/\beta+\tau_0 C_{in}/C_L+\tau_R\right)+s^2\tau_0\tau_R}$$

with $\tau_0 = C_L/g_m, \quad \tau_R = C_{in}R_{on}, \quad \beta = C_{in}/\left(C_{in}+C_f\right)$

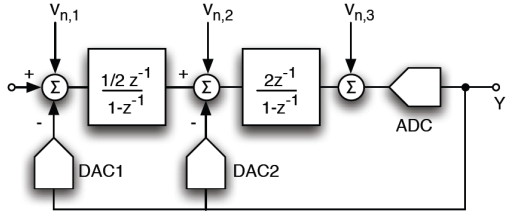
Thus, also during Φ_1 the op-amp noise sees two poles, while the switch noise sees a zero as well. With the same procedure as before, we get

$$V^2_{n,A,C_{in}}=\gamma_{A1}\frac{kT}{C_L} \qquad V^2_{n,R,C_{in}}=\frac{kT}{2C_U}$$

7) Finally, the second integrator (whose output is sampled by the quantizer at the rising edge of Φ_2) also contributes sampled noise on the ADC capacitance, C_{ADC}

Noise summary

The noise contributions from the various sources is summarized in the table/circuit below:



Phase	Source	V^2_{n1}	V^2_{n2}	V^2_{n3}
Φ_2	$4kTR_{on}$	kT/C_U	$kT/(2C_U)$	kT/C_{ADC}
Φ_2	$\gamma_{Ai}\,4kT/g_m$	–	$\gamma_{A1}kT/C_L$	$\gamma_{A2}kT/C_L$
Φ_1	$4kTR_{on}$	kT/C_U	$kT/(2C_U)$	–
Φ_1	$\gamma_{Ai}\,4kT/g_m$	$\gamma_{A1}kT/C_L$	$\gamma_{A2}kT/C_L$	–

Noise power/spectrum

We can now use the fact that the various noise source are uncorrelated, and that the whole power is white from DC to Nyquist → the white noise power spectral density (to be used in simulations and calculations) becomes

$$v_{n,1}^2 = 2T_s\left(\frac{2kT}{C_U}+\gamma_{A1}\frac{kT}{C_L}\right) \qquad v_{n,2}^2 = 2T_s\left(\frac{kT}{C_U}+\gamma_{A1}\frac{kT}{C_L}+\gamma_{A2}\frac{kT}{C_L}\right)$$

$$v_{n,3}^2 = 2T_s\left(\frac{kT}{C_{ADC}}+\gamma_{A2}\frac{kT}{C_L}\right)$$

The noise power spectrum at the output is then

$$v_{n,out}^2 = v_{n,1}^2\left|z^{-2}\right|^2 + v_{n,2}^2\left|2z^{-1}\left(1-z^{-1}\right)\right|^2 + v_{n,3}^2\left|\left(1-z^{-1}\right)^2\right|^2$$

The contribution of $v_{n,1}$ is not shaped (apart from OSR), while the other two are first-order and second-order shaped