Operational Amplifiers part 6: auto zero OPAMP

June 23, 2019

In the last post (part 5) I described the chopper stabilized OPAMP. One of the observations was that the offset cancellation is determined by the accuracy of the duty cycle of the clock used. Auto zero topologies are an alternative approach to get rid of offsets without the high requirements for the clock generator. Auto zero topologies can offer high performance especially in systems that don't need to measure permanently and that offer certain time intervals during which a nulling can be done without affecting the measurement.

1 Types of Auto Zero Amplifiers

Different from a chopper stabilized amplifier the signal is not shifted to an other frequency band . The signal remains in the base band. Only the offset of the amplifier is removed. But the 1/f noise remains in the signal chain.

One of the advantages of auto zero amplifiers is the lower bandwidth required compared to a chopping amplifier. The second advantage is the low frequency the switches are operated with. This leads to less clock feed through (in terms of average clock power).

The offset correction can operate incremental (This applies especially for the in the loop offset correction). The more correction steps the better the offset correction gets. This means the charging of the capacitor storing the correction voltage can be made slow (even switched capacitor solutions to pump discrete portions of charge into the capacitor can easily be implemented). This way the storage capacitor can intentionally be made the dominant pole of the regulation loop.

On the negative side is the auto zero signal, that is somewhere in the middle of the use bandwidth. So the switching noise is visible in the signal. The way this switching is visible depends on the topology used. It depends on the application which kind of distortion can be accepted.

There are several approaches of building auto zero amplifiers:

1. Zeroing while not in use
2. Ping-Pong design
3. In the loop auto zero

Zeroing while not in use is the cheapest approach. The amplifier is not permanently used. While the amplifier is not used the inputs are shorted and nulled. There is a variant of the concept: 2 amplifiers that are periodically swapped (ping pong). The one that is swapped out of the signal path is getting nulled. "Zeroing while not in use" can be done in open loop applications too (This technique is frequently used for comparators of $\Delta\sigma$ ADCs).

In the loop auto zero requires two amplifiers for nulling and for measuring. The total effort is like a ping pong topology. In the loop auto zero provides nicely continuous signals. It only works in systems with a feedback (OPAMPs in closed loop application). It doesn't work for open loop systems such as comparators.

1.1 Storage of the nulling

The adjustment must be stored in one or another way. The most classical analog approach is capacitive storage of the correction value. The capacitive storage must be refreshed regularly because of leakage. Since leakage is temperature dependent the refresh cycles must be more frequent for high temperature applications. Regarding simulation tools and simulation time analog storage is faster to design.

With digital processing getting cheaper and cheaper digital storage is getting more common. The advantage is that digital storage doesn't suffer from leakage. Repetition rates can be kept lower. The price is that and ADC (analog to digital converter) is needed. Design of a system with digital storage requires significantly more computation power for the simulation (either a high number of digital transistors to be simulated with the analog part or mixed signal simulation will be required.)

For ease of understanding let's start with analog storage.

2 Zeroing while not in use

In this "poor man's auto zero amplifier" the main amplifier is only used part of the time. While the main amplifier is not used the inputs are disconnected by switches sw1 and sw2. sw0 short circuits the inputs of the amplifier. sw3 connects the amplifier measuring the output voltage to the adjust input adj. The correction amplifier will tune the adj node until the output voltage reaches 0V. (Literature [1] shows even more simple designs using a floating capacitor right at the amplifier input. But I didn't get happy with it due to the miller capacity of the switches significantly disturbing the circuit - at least as long as I use only some hundred fF to store the charge. Eventually my amplifiers all use a correction at the current mirror and the correction voltage is in the range of hundreds of millivolts. This way the gate charge of the switch becomes negligible)

When the amplifier is returned to operational mode the switches sw0 and sw3 open. The adjust voltage is capacitively stored. switches sw1 and sw2 connect the amplifier to inp and inn again.

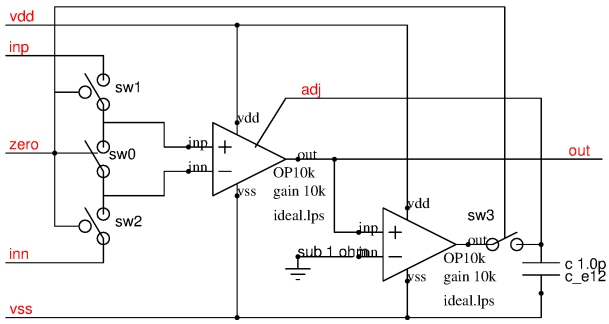

Fig.2.1: Auto zero circuit performing the auto zero while the amplifier is not operated

Since the nulling amplifier compares the already amplified signal with a reference the gain of the nulling amplifier can be kept low. This leads to a very simple design of the nulling amplifier. The main silicon real estate is in the main amplifier while the nulling amplifier often can be neglected.

This approach has limited use for applications in which the signal is only needed at certain times. Typical examples are current sense amplifiers that are only needed while the power transistor is on.

2.1 Signal folding of the simple auto zero amplifier

The simple auto zero amplifier multiplies the signal with the gating clock. Since there is no polarity change the equation looks like this:

$$V_{out} = k * D * 0.5 * (1 + rect(t))$$

K is an amplification factor. In most applications K will be determined by a resistor feedback network like for other OPAMPs in closed loop operation as well. The factor D is the duty cycle. rect(t) is a rectangular function that is +1 for 50% of the time and -1 for the other 50% of the time. In the time domain the signals look like this:

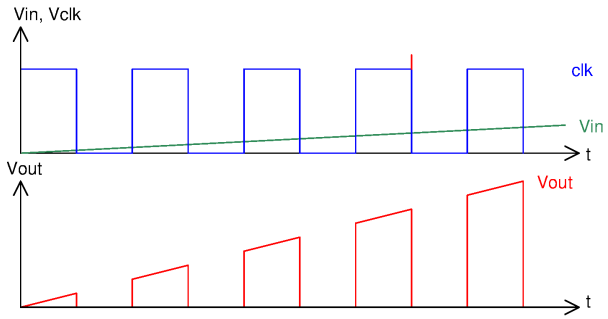

Fig.2.1.1: Time domain signals of the most simple auto zero amplifier

In the multiplication done in the time domain leads to a folding in the frequency domain. The colors in the frequency domain are the same as the colors used in the time domain:

- Input signal: green
- clock signal: blue
- output signal: red

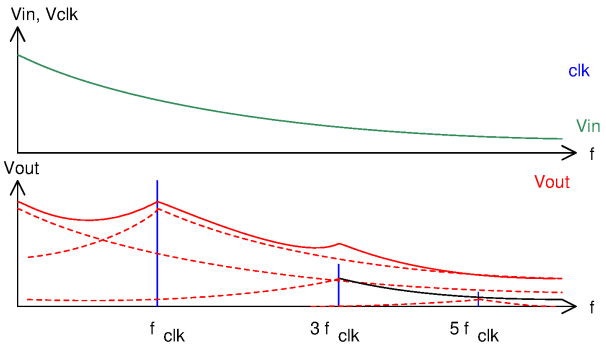

Fig.2.1.2: Frequency domain representation of the signals

The clock becomes visible as discrete frequencies at fclk and odd multiples of fck. The original spectrum of the input signal gets superimposed by the spectrum of the input signal folded around the clock frequency and it's harmonics. So for continuous time systems this kind of amplifier creates a lot of distortions.

Things get a bit more usable if we interpolate the signal when the cock is zero. Basically all we have to do is store the last value seen at the falling edge of the clock until we get a new usable value at the next rising edge. A ample and hold circuit can do this job in the analog world. In the digital world this can be implemented using an ADC and a register.

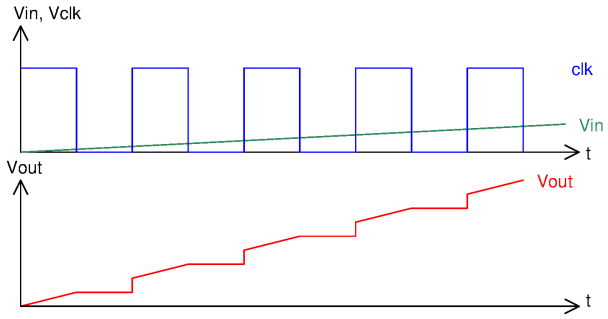

Fig.2.1.3: Time domain signal interpolating while the clock is 0

The question is, what are we doing in the frequency domain? The interpolation rejects the clock and all it's harmonic frequencies (well, not perfectly. There is

a triangular error signal left over, similar to ADC quantization noise). For the spectrum it means we have added a notch filter at N*fclk with N=1,3,5,7.... The resulting output spectrum is very similar to the input signal except for the notches in the frequency domain.

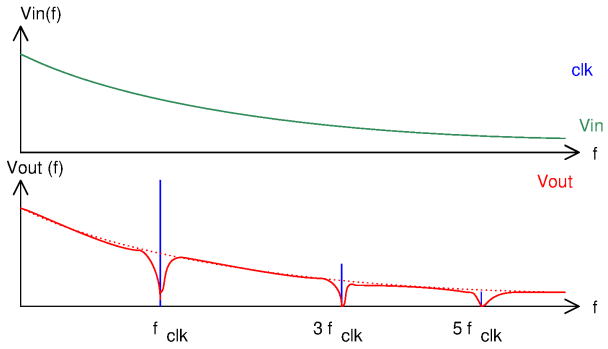

Fig.2.1.4: Spectrum of the simple auto zero amplifier if the signal change is ignored while the clock is 0

For many applications retrieving the original spectrum with the exception of the notches is good enough. This is especially true if the information during the time the clock is 0 isn't needed. (For instance applications that are simply not operating while the clock is 0.)

3 Ping-pong auto zero amplifier

In open loop systems without any "inactive time" a ping pong design may make sense. In a ping-pong system there are two amplifiers that are periodically changing places. One is acting as an amplifier while the other one is nulled. This double the effort but the system is permanently available.

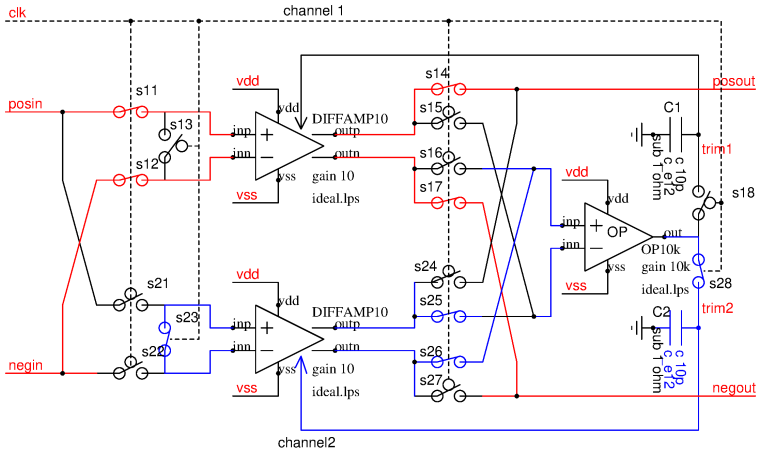

Fig.3.1: simplified concept of a ping pong auto zero amplifier

In the figure above the operating path is drawn in red color and the nulling path is shown in blue color.

One of the problems of the ping pong design is the gain matching of the two channels. Even the smallest gain mismatch of the two amplifier becomes visible as a change of the output signal with the clock frequency. This disturbance increases with the input signal. For this reason the spectrum at the output of a ping pong amplifier usually holds the clock frequency and it's harmonics. As long as mainly the offset is of interest but not (mostly rectangular) clock signal superimposed on the output this can be accepted. Usually the clock is suppressed in the range of 40dB to 60dB corresponding a gain mismatch of 0.1% to 1% between the two channels.

A second effect that cant be fully avoided is the slewing of the amplifier back into the operating point when the two amplifiers are swapped. The amplifier coming from the nulling starts with an output voltage close to zero. Assuming we want to amplify a sine wave the signal looks similar to the following figure.

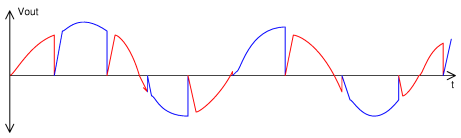

Fig.3.2: Output signal of a ping pong auto zero amplifier

I admit I overemphasized the transitions between the two amplifiers. The amplifier is operating correctly most of the time. Furthermore in mixed signal applications

the output signal of the amplifier usually is sampled (for instance by an ADC). Reasonably the sample time of the ADC will be chosen somewhere in the middle between the swapping events.

In addition the settling time can be reduced dramatically if the frequency compensation is not swapped. This way the frequency compensation capacitor more or less hands over the last operating point from the active amplifier (that now enters the nulling procedure) to the amplifier taking over the signal (that was nulled before).

The signal can be interpreted as a multiplication of two signals in the time domain. One of them is the sine wave, the other one is a (1-trinagular_pulse).

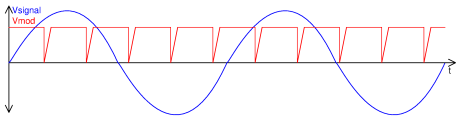

Fig.3.3: Representation of the amplifier output signal by the two functions to be multiplied

These two signals that are multiplied in the time domain get folded in the frequency domain. The constant (DC) part of the red signal folds most of the sinewave directly back into the base band. The short triangular pulses are getting folded with the sine wave. Since the sine wave is symmetrical the frequency of the triangular signal gets suppressed.

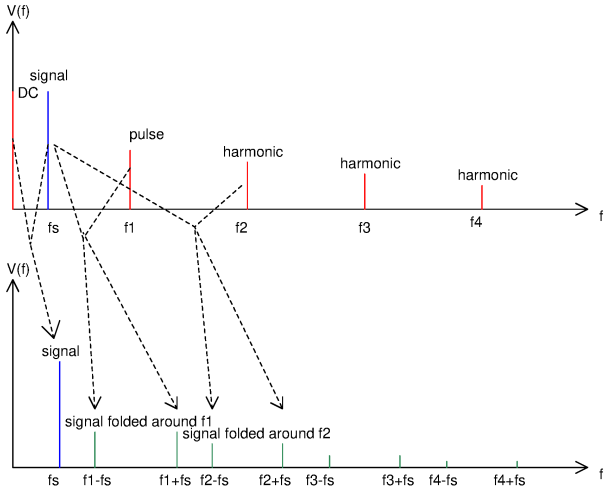

Fig.3.4: Folding of the signal (fs) around the DC component back into the base band and around the triangular pulses into side bands of double the swapping frequency

The unwanted spectral components at $fx \pm fs$ can be minimized reducing the energy of the triangular pulses as far as possible. But in a ping pong design these distortions can't be completely eliminated. In the following figure the simulation of

the output of a ping_pong auto zero amplifier (without storing the operating point in the frequency compensation) is shown. In the example the signal was 1kHz and the swapping frequency was 9.8kHz (so there is a disturbance every 51μs)

The choice of swapping period of 102μs was intentional to obtain a signal that doesn't exactly repeat for a long time.

Simulation was pure transient without noise. For this reason there is no noise floor visible (Using "transient noise" the noise floor would be visible at about 10μV instead of 10nV)

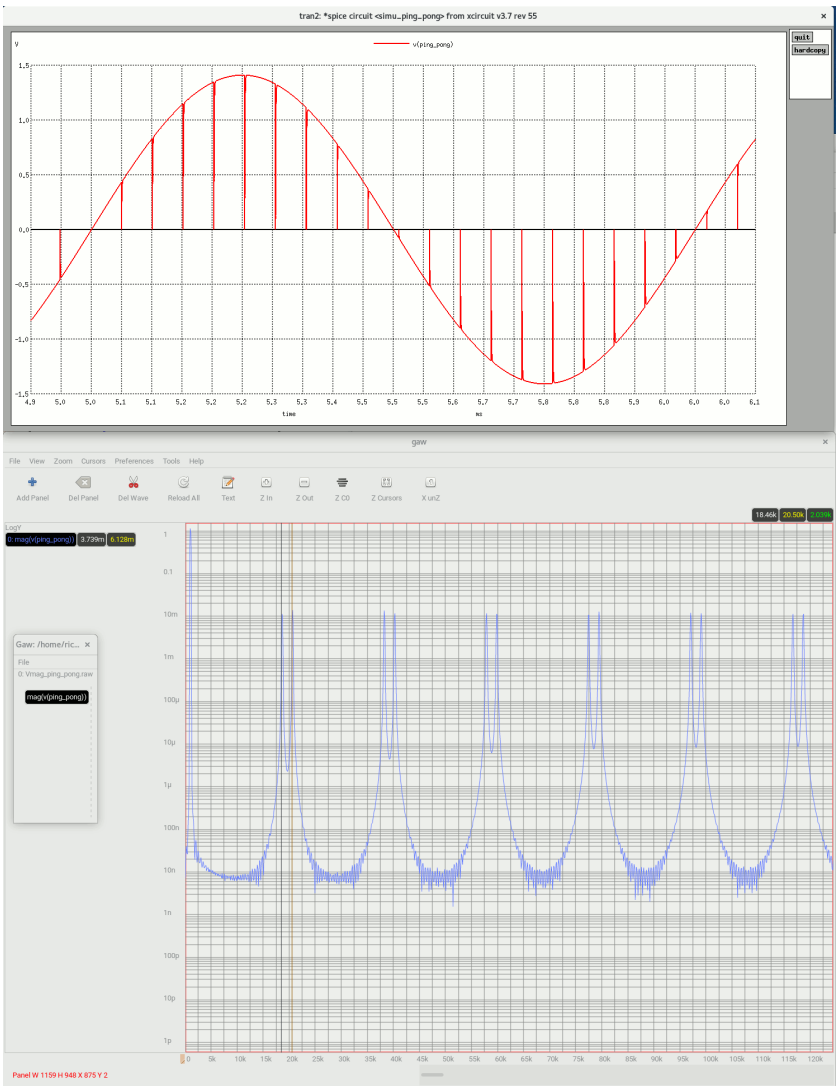

Fig.3.5: Simulation of a ping pong auto zero amplifier and the spectrum obtained

As expected the carrier at 9.8kHz is invisible and due to the folding the double

carrier frequency of 19.6kHz is rejected while the two side bands are visible in the spectrum.

Typical applications are detection of fast signals with requirements for low offset but low gain accuracy requirements. One example is the fast detection of short circuits using very low resistive sense resistors or detecting open loads of a power stage. (Often both tasks are performed by the same amplifier. The open load detection requires a low offset voltage while the short detection must be fast.)

Like in the "adjust while not in use" topology the nulling amplifier OP observes the already amplified signal. The requirements of this nulling amplifier usually are low. Nevertheless the effort is almost double now because we need two amplifiers with a high gain.

4 In the loop auto zero

In the loop auto zero is the most preferred auto zero technique. It offers the advantage that the amplifier is permanently available similar to the ping pong amplifier. The ping pong design has the problem that switching between the channels can produce significant noise due to channel mismatch. In the loop auto zero topologies always use the same amplifier channel for the signal path. The switching noise gets eliminated by concept. (Practical implementations still have some switching noise due to coupling via the gate capacities of the switches. But in most cases this parasitic coupling noise is lower than the noise of a ping pong design.)

One of the first in the loop auto zero amplifiers was the ICL7650 of Intersil [3]. The basic idea is that an ideal operational amplifier always makes the (offset free) input voltage equal 0V. If the amplifier doesn't reach the 0V across it's input it obviously has an offset and need to be corrected. The in the loop auto zero consists of two amplifiers both having a high gain. The nulling amplifier needs a gain of the same magnitude as the main amplifier because it monitors the input instead of the output of the amplifier it has to adjust. For this reason the effort of an in the loop auto zero amplifier is the same as the effort of a ping-pong design. Open loop operation of an in the loop auto zero isn't possible.

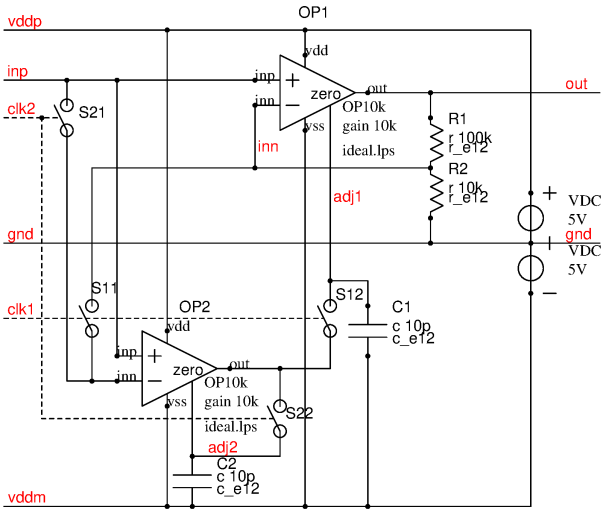

Fig.4.1: Auto zero amplifier with in the loop adjustment

The basic idea of in the loop adjustment is that an ideal OPAMP (with zero offset and infinite gain) will always bring the differential voltage between its inputs inp and inn to 0V. The signal path is running via OP1. OP2 is used for adjustment. The adjustment is done by the following steps:

1. Short circuit the inputs of OP2 closing S21
2. Close S22 to let OP2 correct it's own offset.
3. Open S22 to store correction of OP2 in C2
4. Open S21 and close S11 to measure the input offset of OP1
5. Close S12 to adjust OP1
6. Open S12 to store correction of OP1 in C1
7. Restart at step 1.

5 Beyond all measures

Theoretically it is possible to use an in the loop auto zero topology inside a copping amplifier. The auto zero amplifier reduces the offset to about $10\mu V$ and the chopper amplifier further reduces the offset to about 100nV and removes the $1/f$ noise. But what is the benefit? As soon as we have to work with signals coming from outside the chip there are many interconnects interfacing different metals at different temperatures. On low power chips the temperature differences can be kept under control to a certain extent. On high power chips the temperature gradients can already become significant.

The following table holds the Seebeck coefficients of some commonly used materials [4] at room temperature versus platinum (Pt) . Ideally each junction exists as a pair. As long as these junctions have the same temperature the Seebeck voltages cancel.

The exact numbers (e.g. of semiconductors) depend on doping and the exact alloys used. Furthermore the Seebeck constants depend on temperature. So these numbers are approximate values for room temperature only.

material	usage/position	Seebeck const.
Si	gate contact, source contact	$\approx 440\mu V/K$
Ge	emitter contacts	$\approx 300\mu V/K$
NiCr	Nichrome	$25\mu V/K$
Fe	Pin of ICs	$19\mu V/K$
W	Tungsten vias	$7.5\mu V/K$
Au	bond wires	$6.5\mu V/K$
Ag	RF inductors	$6.5\mu V/K$
Cu	wires, bond wires	$6.5\mu V/K$
Pb	lead solder	$4.0\mu V/K$
Al	pad, wires on chip	$3.5\mu V/K$
C	Carbon resistors	$3.0\mu V/K$
Pt	reference	0
Ni	Nickel	$-15\mu V/K$
	Constantan resistors	$-35\mu V/K$
Bi	Bismuth solder	$-72\mu V/K$

Especially the solder joints on the board are a source of errors. Two solder joints with temperature difference of only 1K will already lead to a measurement error of some $10\mu V$. In other words designing an OPAMP that can be connected to a IC pin for offset voltages below $10\mu V$ doesn't make sense because the Seebeck voltages on the board will contribute errors that are much higher than those of the amplifier itself!

On chip it may be a different situation. On low power chips the temperature gradients are low. Amplifiers optimized for reading HAL sensors placed on the same chip as the amplifier are in fact designed to reach offset voltages of $1\sigma = 100nV$ or even better (for an electronic compass you need a resolution of about $1\mu T$or better. Typical HAL sensors calculated with the data of [2] have a sensitivity of $100nV/\mu T$ to $200nV/\mu T$). However as soon as power stages are on the same chip the temperature gradients are higher. Combining HAL sensors with power stages on the same chip will easily degrade performance by one magnitude.

References

[1] Randall L. Geiger, Phillip E. Allen, Noel R. Strader, "VLSI Design Techniques for Analog and Digital Circuits", McGraw Hill, 1990

[2] R. Mueller, “Grundlagen der Halbleiter Elektronik”, Springer, 1987

[3] “Intersil SE Spezial-Electronic”, Spezial-Electronic, 1976

[4] “Thermoelektrische Spannungsreihe”, https://de.wikipedia.org/wiki/Thermoelektrische_Spannungsreihe, snapshot June 19 2019