A sampled voltage reference

C. T. Rooijers

TUDelft
Delft
University of
Technology

**Electronic
Instrumentation
Laboratory**

Challenge the future

A SAMPLED VOLTAGE REFERENCE

by

C. T. Rooijers

in partial fulfillment of the requirements for the degree of

Master of Science
in Microelectronics

at the Electronic Instrumentation Laboratory,
Department of Electrical Engineering,
Delft University of Technology,
to be defended publicly on Monday January 25, 2016 at 10:00 AM.

Supervisors:	Prof. dr. K. A. A. Makinwa	
	Prof. dr. J. H. Huijsing	
Thesis committee:	Prof. dr. K. A. A. Makinwa	TU Delft
	Dr. M. Pertijs	TU Delft
	Dr. V. Giagka	TU Delft
	Prof. dr. J. H. Huijsing	TU Delft

This thesis is confidential and cannot be made public until December 31, 2020.

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TUDelft Delft
University of
Technology

ABSTRACT

This thesis describes the implementation of a sampled voltage reference. A sampled voltage reference aims to achieve both low-power and low-noise by storing the output of a voltage reference on a capacitor for a long time. This allows the reference to be switched off during the hold period, which leads to lower average power. At the same time, all the voltage reference's noise is pushed down into a bandwidth determined by the refresh frequency, while the buffer can be made low-noise by auto-zeroing.

The design of a continuous-time auto-zeroed buffer with low-noise and low-offset is presented. Various techniques have been used to reduce the transients created by auto-zeroing. In simulation the transients are below 1 μV peak-to-peak, but the effectiveness of the techniques could not be evaluated in measurements.

The design of a low leakage sample and hold circuit is also presented. This uses bootstrap techniques to maintain zero potential across critical parasitic diodes. It is shown to be effective, resulting in a drift of about 5 μV per second. A mechanism is found which explains how the buffer's residual offset is transferred to the hold capacitor. A special slow-chopping technique is presented and implemented to try to reverse the leakage due to this residual offset.

The final implemented design suffers from a large coupling between the high frequency reference clock and the output. It is shown that this causes a much higher residual offset than expected, which in turn increases the leakage of the hold capacitor. With a measured residual offset in the order of several 100 μVs and an auto-zeroing frequency of 2 kHz, the leakage in 1 s is 34 mV. Via simulation it is shown that with a lower residual offset, the leakage can be greatly reduced.

Methods of improving the current design have been investigated. A new clocking scheme is proposed and simulated. Improvements to the buffer are also proposed, which should lower its 1/f noise corner. This would allow for a lower auto-zeroing frequency, which in turn will further reduce the leakage.

Acknowledgement

CONTENTS

LIST OF FIGURES

1

INTRODUCTION

The voltage reference is a commonly used building block in many analog and mixed-signal systems. In a measurement system, the voltage reference provides the standard against which other voltage measurements are referred. To be useful, the reference voltage should be insensitive to environmental variations such as: temperature, supply voltage and time. The application of voltage references include: sensor conditioning, analog-to-digital conversion (ADC) and digital-to-analog conversion (DAC), as shown in Figure 1.1.
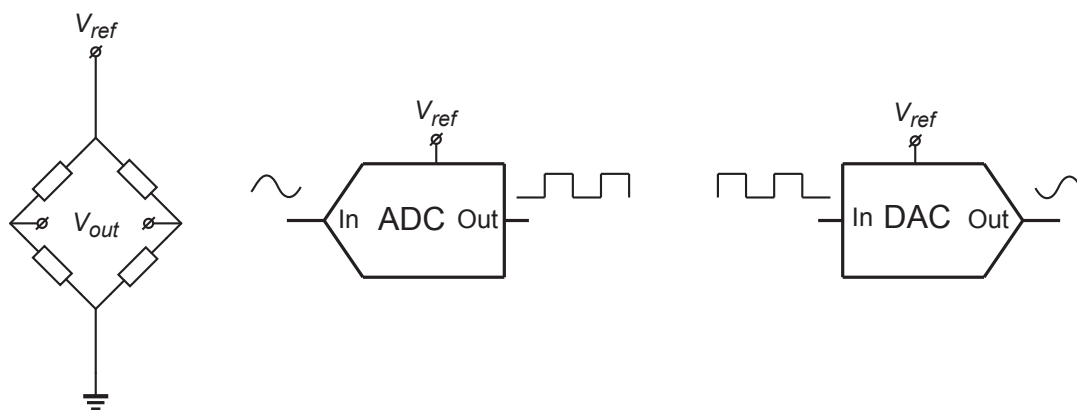
Figure 1.1: Some typical applications for a voltage reference. From left to right: sensor conditioning, analog-to-digital conversion and digital-to-analog conversion.

For these applications some key parameters for the voltage reference have been determined:

- Initial accuracy
- Temperature coefficient
- Noise
- Long-term drift
- Power supply rejection ratio

1.1. LITERATURE/HISTORY

Voltage references have been around for a long time. The first voltage references were the Clark cell and Weston cell developed in the 19th century [2]. These bulky electrochemical cells where mostly used as laboratory standards.

The first semiconductor voltage reference was the zener diode introduced in the late 1950s [3]. The zener reference behaves like a non-linear divider. It works by reverse biasing a PN junction until it reaches the breakdown voltage where the voltage over the junction changes very little with supply current. The breakdown effect is due to two different effects namely: the avalanche effect and the zener effect. The avalanche effect gives a negative temperature coefficient while the zener effect gives a positive temperature coefficient. The residual temperature coefficient of a Zener diode is slightly positive. This positive temperature coefficient can be combined with the negative temperature coefficient of a forward biased diode, to create a temperature compensated Zener diode. In the 1970's [4, 5] zener diodes were placed deeper in the silicon, to get rid of the $1/f$ noise which is mostly a surface effect.

Even though buried zener references with superior temperature coefficient (0.05 ppm/°C using temperature stabilization) and long term-stability ($2\mu V/\sqrt{kHr}$) are around [6], they are no longer widely used due to their high supply voltage requirement (>7V) and large current consumption.
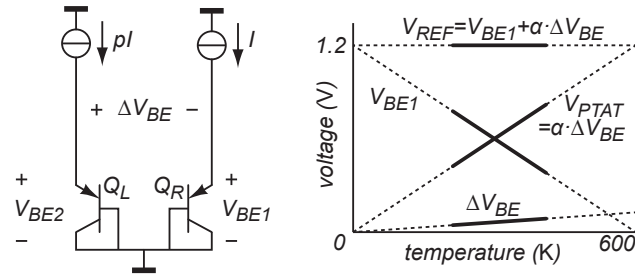
Figure 1.2: The bandgap reference principle.

In 1964 [7] a new principle was introduced by Hilbiber to create a stable voltage reference, this principle is refered to as the bandgap principle. In the bandgap principle, shown in Figure 1.2, two voltages with different temperature coefficients are combined to get a temperature independent output. Most commonly the base-emitter voltage (V_{BE}) is used which has a negative temperature coefficient (CTAT), combined with the difference between two base-emitter voltage (ΔV_{BE}) which has a positive temperature coefficient (PTAT). After the introduction by Hilbiber, several circuits have been published that use this principle to create a reference voltage [8–10].

The first bandgap references were first-order temperature compensated, higher-order effects give rise to residual temperature dependency. This results in a typical temperature coefficient of around 0.1 %/°C. To get rid of second-order effects, second-order compensation was introduced [11, 12]. This improves the temperature coefficient by an order of magnitude, to typical values of around 50 ppm/°C. Using two-point trimming [13], the temperature coefficient can be lowered to typical values of 3 ppm/°C. A bandgap reference typically contains a CMOS opamp to combine V_{BE} with a scaled ΔV_{BE}. The spread of V_{BE} is mainly PTAT, while the offset of a CMOS opamp typically has a non-PTAT temperature drift. Therefore, a single-point trim cannot compensate for both. Using chopping to minimize the offset of the opamp, the temperature coefficient can be below 10 ppm/°C with a one-point trim [14].

1.2. STATE-OF-THE-ART / RECENT TECHNIQUES

With the increasing trend towards lower supply voltage and low-power applications. In recent years many publications have been written about bandgap references with an output below the usual 1.2 V [15, 16] and nW-power bandgap references [17, 18]. Both of these nW-power bandgap references achieve low power by reducing the current used in the bandgap core. Although not confirmed by published results, this will come at the cost of increased noise.

Another approach [19, 20] to getting low-power bandgap references is by using a sampled bandgap reference, to reduce the average required current. By holding the reference voltage created by a bandgap reference on a capacitor, the bandgap core can be switched off, reducing the average required current. The key to getting low power is reducing the leakage to get a long hold time. [20] aims to achieve a long hold time by measuring the leakage of a small capacitor (50 fF) and using this to reduce the leakage of a large hold capacitor. The hold time is dynamically changed, based on the measured voltage droop. This is shown to be effective with an average power consumption of 2.98 nW at room temperature. However, the allowed voltage droop is rather large with 100 μV. Furthermore, the hold time drops dramatically for higher temperatures (100 s at 40 °C, 10 s for 50 °C and 0.5 s for 80 °C), which also dramatically increases the average required power. [19] aims to achieve a long hold time by using a buffered hold capacitor voltage to reduce the leakage of the critical junctions. It is shown to be effective with an average power of 170 nW. The hold time is rather low with 100 ms and no mention is made of the voltage droop. Large output transients of 20 mV peak-to-peak also occur during the charging of the hold capacitor. Low noise is claimed, but no data is supplied to support this claim. Without chopping and auto-zeroing, the noise will be limited by the buffer. [21] combines a power-efficient ADC with a low-power reference. The reference only consumes 25 nW and this can be reduced to 2.5 nW by using duty-cycling. The duty-cycling is done with a cascade of three sample-and-hold stages, where each stage lowers the voltage across the switch and reduces the leakage. A duty-cycle of 10% is achieved, with a maximum ripple of 100 μV and no performance loss in the ADC.

1.3. MOTIVATION

The design of a voltage reference is, just like the design of other analog circuits, a trade-off between different design parameters. The initial accuracy is a reflection on the sensitivity to process variation. Voltage references with very high initial accuracy require special circuit techniques such as dynamic element matching and costly trimming techniques. References with very low temperature coefficients require second-order curvature compensation. Low noise usually comes at the cost of increased power consumption. While references with a low power consumption usually have degraded initial accuracy, temperature coefficient and noise performance.

For these reasons, commercially available low-power voltage references, such as the LT6656 shown in table 1.1, usually have relatively high noise and poorer temperature coefficients. While the really accurate and low-noise voltage references, such as the LTC6655, are relatively power hungry.

	LTC6655 [22]	MAX6126 [23]	LM4140 [24]	LT6656 [25]	X60008 [26]
Initial accuracy (Max %)	0.025	0.02	0.1	0.05	0.01
Noise peak-to-peak 0,1 to 10 Hz (ppm)	0.25	0.63	2.15	30	6
Temp co. (ppm/°C)	2	3	3	10	1
Supply current (Typ μA)	5000	380	375	0.85	0.5
Technique	Bandgap	Bandgap	Bandgap	Bandgap	Floating gate

Table 1.1: Several commercially available voltage references compared, ordered from high to low supply current.

A voltage reference that seems to break this trend is the X60008. This reference has the lowest current consumption of all the parts in Table 1.1 and at the same time achieves the lowest initial accuracy and lowest temperature coefficient. The peak-to-peak noise is higher than most of the other references, but it is still five times better than that of the second lowest power reference, the LT6656. The X60008 voltage reference is based on the floating-gate voltage reference developed by Xicor/Intersil in 2005 [27]. They produced a very low leakage capacitor by using a technology similar to the one used for EEPROM digital memory. They then pre-charged this very low leakage capacitor to a precise reference voltage and used a standard buffer to drive a load. References based on this technology are able to achieve a very low power consumption but the noise performance is still significantly higher than other low noise voltage references.

1.4. KEY IDEA

The key idea of this work is to make a voltage reference by storing a reference voltage on a hold capacitor. This hold capacitor is then buffered so that it can supply current to a load without draining the hold capacitor. This approach is similar to the one used in the X60008. However, the goal of this work is to realize a reference in a

standard CMOS technology, in this case the TSMC 0.18 μm 1.8V technology. This obviates the need for an exotic ultra-low-leakage High Voltage technology and results in a voltage reference that is generally applicable, also as a part of a system-on-a-chip (SoC). A standard CMOS technology does not allow for a very low leakage capacitor, such as the one used in the Intersil/Xicor product. Therefore a normal leaky capacitor is used, which is recharged periodically to keep the output reference voltage within a certain margin from the desired value. This creates a sampled voltage reference, shown in Figure 1.3.
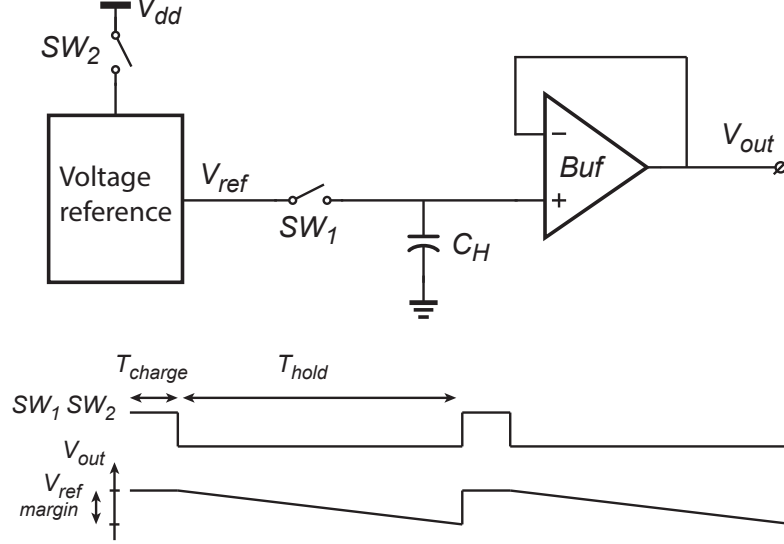
Figure 1.3: The key idea of a sampled voltage reference.

Keeping the reference voltage stored on a hold capacitor with a long hold time, has two advantages. First, once the hold capacitor is charged to the reference voltage, the voltage reference can be switched off. In this way the average power required by the sampled voltage reference can be lowered.

The average power (P_{avg}) required by the sampled voltage reference, is given by:

$$P_{avg} = dP_{Ref} + P_{Buf} = \frac{T_{charge}}{T_{charge} + T_{hold}} P_{Ref} + P_{Buf}. \tag{1.1}$$

Where d is the duty cycle, P_{Ref} is the power consumed by the reference and the buffer of the and P_{buf} the power of the buffer. To benefit from the lower average power consumption, the power consumed by the buffer should be considerably lower than the power consumed by the voltage reference. Furthermore the capacitor should be charged by the reference quickly and the capacitor should hold the voltage for a long time. All to get a low duty cycle, which will reduce the average power consumption.

The second advantage is that the sampling action will shape the noise. All the noise of the voltage reference will be pushed down into a frequency band below the refresh frequency. The system effectively behaves like a switched capacitor filter with a very large time constant. Key to getting the low-noise from the reference is in making the leakage low enough so that the refresh frequency can be low, while the leakage induced error is small enough. During the hold period the only noise present on the output is due to the buffer amplifier. By using dynamic offset compensation techniques this noise can be made very low without using a large amount of power.

Ultimately this should lead to a very precise voltage reference that achieves low noise while at the same time having a very low power consumption. In this thesis the focus is on the design of a low-leakage sample and hold circuit and an offset compensated buffer.

2

SYSTEM LEVEL DESIGN

This chapter presents the system level design of the sampled voltage reference. First the design of the offset compensated buffer is presented. Of the different offset compensation techniques, the most appropriate for this application is chosen. Specifications for the buffer are determined, based on a commercially available high performance voltage reference. Next techniques are described to achieve low switching transients and low residual offset. The noise is analyzed for a sampled reference, followed by the design of the low-leakage sample and hold circuit.

Ideally the buffer should provide an exact copy of the reference voltage stored on the capacitor. However, several non-idealities of the buffer work against this. The design challenge is then to keep these non-idealities as small as possible.

Challenges of the buffer design:

- Low offset and low offset drift

- Low unity-gain error

- Low noise

- Small switching transients

- Low influence on the hold capacitor.

2.1. OFFSET AND OFFSET DRIFT CANCELLATION

The buffer should have a low offset and a low offset drift over temperature and time. Any offset will deteriorate the initial accuracy of the reference and any offset drift with temperature will degrade the temperature coefficient of the voltage reference. Finally, any offset drift over time will influence the reference's long-term stability. In CMOS technology an offset of around 1-10 mV is not uncommon [28]. Careful layout and large input transistors can only help to reduce the offset to the millivolt level. However, commercially-available high performance bandgap references, have initial accuracies of ± 0.025 % and temperature coefficients of 2 ppm/°C [22]. For a reference with a nominal output voltage of 1.250 V this translate into an initial accuracy of ± 312.5 μV and a temperature coefficient of 2.5 μV/°C over a temperature range of -40 to 125 °C. To keep the buffer from negatively influencing the performance of a high performance reference, its offset and offset drift over temperature should at least be of the same order or preferably an order of magnitude better. Therefore a way to reduce the offset and the offset drift from the normal mV-level is desirable, to keep the performance of the overall system similar to a high performance bandgap reference.

2.1.1. TRIMMING

To reduce the offset, two basic techniques can be distinguished namely: static offset reduction and dynamic offset reduction. Static offset reduction, or trimming, works by driving an extra set of inputs with a trimming voltage that compensates for the current generated by the offset voltage, as shown in Figure 2.1. Trimming can easily

reduce the offset to below 312.5 μV but it only effectively compensates for offset drift in bipolar devices and is less effective in the case of MOS devices. To get an offset drift below 2.5 μV/°C, two point trimming or special circuit techniques together with a one-point trim are necessary [29]. Furthermore, trimming does not reduce $1/f$ noise and is costly to use, due to the increased production time. In a system combining buffer and bandgap reference, trimming only needs to be done in the bandgap, removing the extra burden of trimming the buffer. Dynamic offset compensation techniques include chopping and auto-zeroing. These techniques use additional circuitry that continually works to reduce the offset. Therefore any slow changes in the offset and noise will also be compensated for.
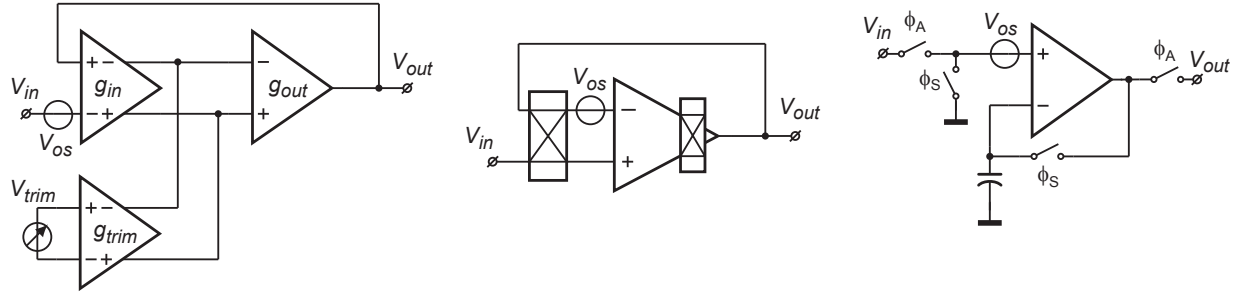
Figure 2.1: Offset compensation techniques. From left to right: trimming, chopping and input sampled auto-zeroing.

2.1.2. Chopping
Chopping can reduce the offset by using two pairs of polarity-reversing switches or choppers. The input chopper modulates the input signal away from DC. An output chopper can then demodulate the signal back to DC, while the offset and $1/f$ noise is modulated to the chopping frequency and can in principle be filtered out. The modulated offset will be a triangular waveform after being integrated by the output stage, this will result in a triangular ripple on the output. Furthermore the act of chopping at the input introduces spikes which, due to the limited bandwidth of the amplifier and the demodulation at the output, have a residual DC component. For chopper amplifiers, residual offsets of ± 10 μVs and ripples with amplitudes of 10 mVs are not uncommon [30], for a voltage reference buffer the offset is acceptable but the ripple is undesirably high. Several techniques exist to reduce the ripple caused by the offset and the spikes. The techniques include nested chopping [31], chopper stabilized amplifiers, ripple reduction loops and switched capacitor notch filters. These techniques can reduce the offset to 100 nVs but the ripple is still in the 10 μVs range [30].

2.1.3. Auto-zeroing
Auto-zeroing is based on the sampling of the offset in one phase (ϕ_s) and compensating the offset in the other phase (ϕ_A). The simplest forms of auto-zeroing are discrete time based, such as the circuit shown in Figure 2.1. For a generally applicable voltage reference a continuous-time solution is required, since most applications require the output to be continuously available. The advantage of auto-zeroing is that it doesn't use modulation to get rid of the offset, so it also doesn't suffer from the ripple associated with the modulation. Of course there is still some switching involved and therefore spikes still occur at the output. Furthermore, the sampling action of auto-zeroing leads to noise fold-back, which makes it harder to make a low-noise auto-zeroed buffer. So for a sampled voltage reference we have to find ways to realize continuous-time auto-zeroing, while avoiding excessive noise fold back.

2.1.4. Reducing the unity-gain error
The buffer should have a low unity-gain error. Therefore the buffer should have a large open-loop gain. Any unity-gain error will reduce the initial accuracy of the reference.

2.1.5. Conclusion
In conclusion, for a voltage reference buffer, two offset reduction techniques are suitable namely: trimming and auto-zeroing. Trimming is more costly to use, unless it is combined with the trimming of the bandgap. It doesn't benefit from the lower $1/f$ noise achieved by dynamic offset compensation and thus requires more power to achieve low noise. The big advantage is that no switching is involved and therefore no spikes will occur at the output and, as will be discussed later, the buffer's input current will be essentially zero. The second candidate is auto-zeroing which can achieve low residual offset, low noise and at the same time low power. Therefore for

a buffer with low residual offset, low noise, small switching transients and low power, an auto-zeroed buffer is probably the best choice.

2.1.6. SPECIFICATIONS

Switching transients	$\leq 1\ \mu$Vpp
Offset	$\leq 300\ \mu$V
Offset drift	$\leq 2.5\ \mu$V/°C
Noise (0.1 to 10 Hz)	$\leq 500\ n$Vrms
Power	as low as possible

Table 2.1: The specifications of the auto-zeroed buffer.

2.2. A CONTINUOUS-TIME AUTO-ZEROED BUFFER

Now that it has been established that an auto-zeroed buffer is the best option. Next it is time to look at how a continuous-time auto-zeroed amplifier can be created. For general voltage reference applications, a continuously available output voltage is desirable. Several topologies exist for continuous-time auto-zeroed amplifiers. The first is based on using two auto-zeroed amplifiers in a ping-pong technique [32], as shown in Figure 2.2. While one of the amplifiers is being auto-zeroed, the other can amplify the signal. The second continuous-time auto-zeroed amplifier is commonly referred to as an auto-zero offset-stabilized amplifier [1]. In this auto-zeroing technique, the buffer stays in the signal path, making it suitable for continuous time applications. The offset of the buffer is removed by a loop, that switches between nulling the offset of the loop itself and nulling the offset of the buffer. Since the ping-pong technique relies on interchanging the two amplifiers, which causes spikes [28], the offset-stabilized architecture is preferred.
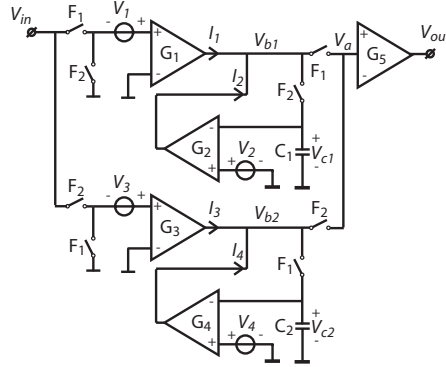
Figure 2.2: A ping-pong continuous-time auto-zeroed amplifier [1].
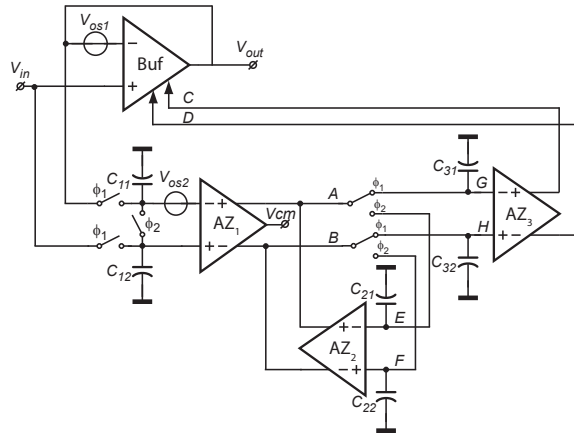
Figure 2.3: A conventional continuous-time auto-zeroed buffer.

In Figure 2.3 a conventional continuous-time auto-zeroed buffer is shown. The buffer has an offset V_{os1} which has to be removed. To remove this offset a compensation loop is created by the operational transconductance amplifiers (OTA's) AZ1, AZ2 and AZ3. This loop measures the offset of the buffer, stores it on capacitor C_{31} and C_{32} and compensates for it. To measure the offset of the buffer, first, the offset of the sensing OTA AZ1 has to be removed. In the next phase this basically offset-less AZ1 can sense the offset of the buffer and compensate for it. These two actions leads to two different phases. Phase ϕ_2, in which the offset of AZ1 is measured and stored by AZ1 and AZ3. Phase ϕ_1, in which the offset of the buffer is measured and stored. Both phases will be analyzed separately.

For the auto-zeroing loop a differential structure is preferred, above a single-ended structure. This is because the charge injection of two switches will than partially cancel and we only have to deal with their charge-injection mismatch.

2.3. Switching transients

In the conventional continuous-time auto-zeroing approach shown in Figure 2.3, switching transients occur during the transition from the sampling of the buffer offset (V_{os1}) to the sampling of the AZ1 offset (V_{os2}) and vice versa. Some of the transients are due to the limited bandwidth of the OTA's, while some are due to the non-idealities of the switches connected at the input and output of AZ1. These two contributions can be minimized separately. Since output transients are generally undesirable for voltage reference applications and due to the relatively high auto-zeroing frequency required to get rid-off the $1/f$ noise (explained in section 2.5), an effort is done to keep them below $1\,\mu$V.

The dominant source of the output switching transients is due to the limited bandwidth of the OTA's. Therefore this effect is first studied and several techniques are investigated to reduce the transients.

Switching transients due to limited bandwidth

In the auto-zeroed buffer shown in Figure 2.3, the output current of AZ1 (I_{AZ1}) needs to switch between $g_{m\_AZ1}V_{os1}$ and $g_{m\_AZ1}V_{os2}$. Where g_{m_AZ1} is the transconductance of AZ1. Due to the finite bandwidth of AZ1, the transition between these two current levels will take some time, as shown in Figure 2.4. This settling behavior from one output current level to the next will introduce spikes on the nodes GH and EF. Since the voltage on nodes GH directly influence the output, spikes on GH will also be present on the output. The spikes on GH will be attenuated by the ratio of the AZ3 and buffer gain, $\alpha_1 = \frac{A_{AZ3}}{A_{Buf}} = 30$, which will be explained in more detail in Section 2.4. This means that for switching transients below $1\,\mu$V peak-to-peak, the switching transients at GH should be below $30\,\mu$V.
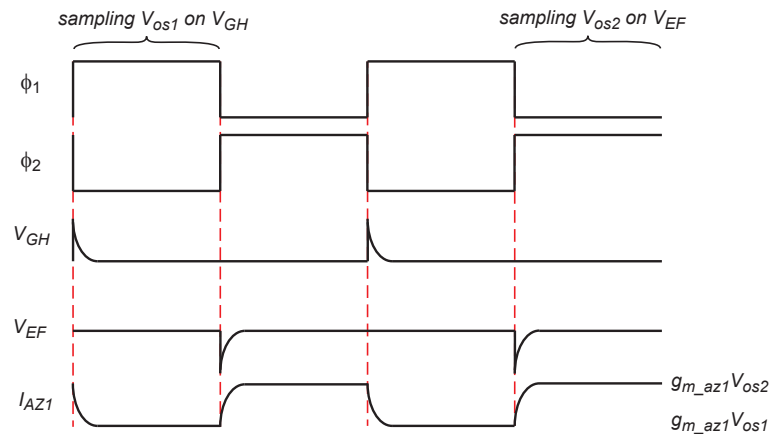
Figure 2.4: A sketch of the effect of a limited bandwidth of AZ1 on the output current and the voltage on EF and GH. Drawn for the case that $V_{os2} > V_{os1}$.

The spikes in Figure 2.4 can be minimized by keeping the offset level of AZ1 and the buffer small and hence close to each other. This can be done by careful layout of the input transistors of AZ1 and the buffer. Furthermore the parasitic capacitance at the output of AZ1 can be minimized, which translates in to a high bandwidth of AZ1. However, these measures alone will not reduce the output switching transients below 1 μV peak-to-peak. In the next two sections, techniques will be introduced to further lower these transients.

Switching transients due to switching non-idealities

Besides the transients due to the limited bandwidth of AZ1, some of the transients are due to the non-idealities of the switches. In CMOS switches, there are two non-ideal effects that cause transients, namely: charge injection and clock feedthrough. During the on-state of a MOS switch, a channel exists between the source and drain. During the transition to the off-state, this charge escapes the channel to the source, drain and bulk. The total charge in the channel can be expressed as:

$$Q = WLC_{ox}(V_{gs} - V_{th}). \tag{2.1}$$

A first obvious way to minimize the charge injection is to take minimum size transistors as switches. This can be done since speed is not important in this design. A second way would be to minimize the effective gate-source voltage $(V_{gs} - V_{th})$.

Besides charge injection, clock feedthrough also creates switching transients. Due to a slight overlap capacitance between gate-drain and gate-source will lead to capacitive coupling of the clock signal. This parasitic capacitance will couple any transition on the gate to the source and drain.

Three well-known techniques exist to minimize the above described effects, shown in Figure 2.5. These techniques are: dummy switches, transmission gates and differential sampling.
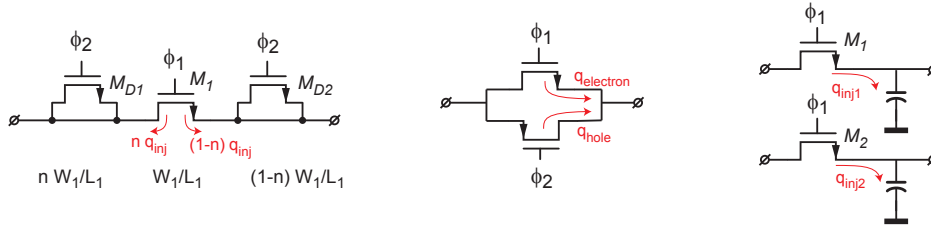
Figure 2.5: From left to right: dummy switches, transmission gates and differential sampling.

The use of dummy switches assumes that a predictable fraction of the channel charge will flow to the source and drain. The source and drain then have dummy switches that are sized according to this fraction and switch with the complementary clock. In theory, the charge needed to build the channel in these dummy switches will compensate for the charge of the original switch. In practice the charge injection depends on the impedance seen by the drain and source and the clock rise and fall time [33, 34]. This makes it difficult to predict the precise behavior of the charge, making it hard to implement a proper compensation.

The use of transmission gates relies on the cancellation of the holes injected by a PMOS and the electrons injected by a NMOS. These two charge contributions will only cancel for one precise input voltage. Furthermore the overlap capacitance of a PMOS and a NMOS will generally not match. Previous implementations with dummy switches and transmission gates, seem to favor transmission gates slight above dummy switches [35]. Simulations done on the two compensations schemes give the same result. Furthermore, with overlapping clocks the instantaneous injection of the PMOS and NMOS can cancel before affecting the output. This instantaneous cancellation depends on the simultaneity of the switching of the PMOS and NMOS. Any skew between the clock driving the PMOS and NMOS will reduce the effectiveness of this technique [36].

The final approach is differential sampling. This relies on the fact that the signal can be stored as a differential

voltage on two capacitors and that the charge injection of two transistors into these capacitors is roughly similar. The charge injection transient will therefor occur as a common-mode step on both capacitors. A small differential transient will remain due to the charge injection mismatch of the two switches, which is much smaller than the charge injection of a single switch.

Since the switches in front of AZ1 are directly connected to the output they will directly create transients. If the switches go from on to off, charge needs to go from the channel to a low impedance node. This will lead to a temporarily increased output, which will slowly be corrected by the buffer. If the switches go from off to on, charge will be needed to fill the channel. This will lead to a temporarily decreased output, which will slowly be corrected by the buffer. These switches are necessary for the auto-zeroing operation. To measure the offset of AZ1, the input of AZ1 has to be shorted in phase ϕ_2. Therefore the transients due to these switches can only be minimized and not circumvented. These switches are implemented with transmission gates to minimize the charge injection.

The goal of the design is to keep switching transients below 1 μV peak-to-peak. Several techniques are used to achieve this, which are explained in the rest of this paragraph.

2.3.1. Active integration
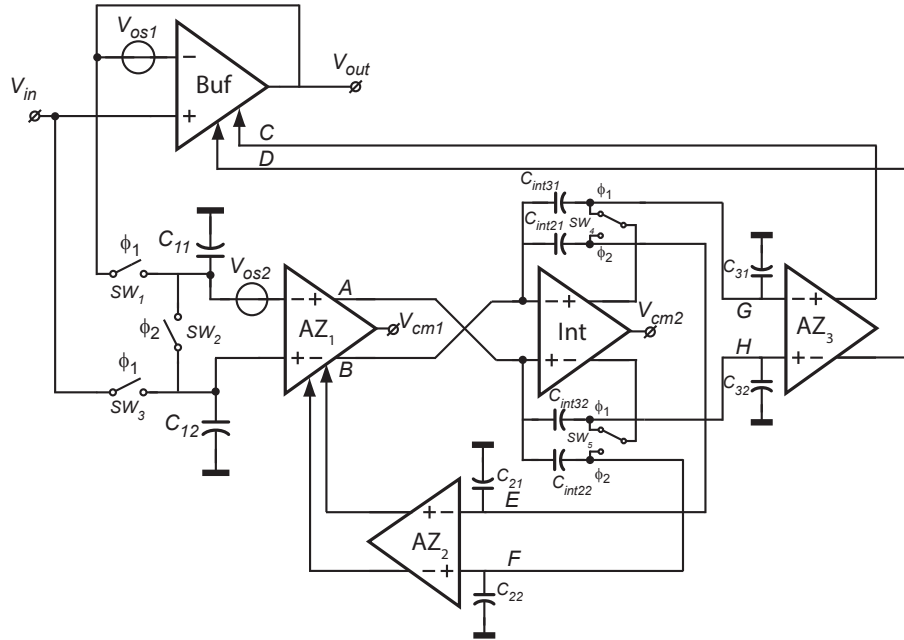
Figure 2.6: A continuous-time auto-zeroed buffer with active integration.

The previous section explained how the limited bandwidth of AZ1 can create output transients. To reduce these transients, the circuit shown in Figure 2.6 may be used. This incorporates an active integrator rather than a passive one. The output current of AZ1 is now steered into a virtual ground. Any transients on this virtual ground will be actively suppressed by the integrator. This results in the transients now being fully determined by the step response of the integrator. Since AZ1 is basically a copy of the buffer to keep the offset and noise contributions about the same, the focus is then on the design of a fast integrator.

The active integrator is present in both phase ϕ_1 and ϕ_2 of the offset nulling loop. The DC-gain added by the integrator will therefore help to reduce the residual offset, both due to the offset of AZ1 and the buffer. For the implementation of the active integration two options were investigated, shown in Figure 2.7. The first option is to use two integrators, one for each auto-zeroing phase. However, both integrators have to be auto-zeroed, to minimize the mismatch in offset between the two integrators. The second option shares one active integrator for both phases. In a single integrator implementation, the offset of the integrator is sampled in phase ϕ_2 together

with the offset of AZ1 and the same offset is compensated for in phase ϕ_1. This omits the need of an auto-zeroed integrator, while at the same time being more power efficient.
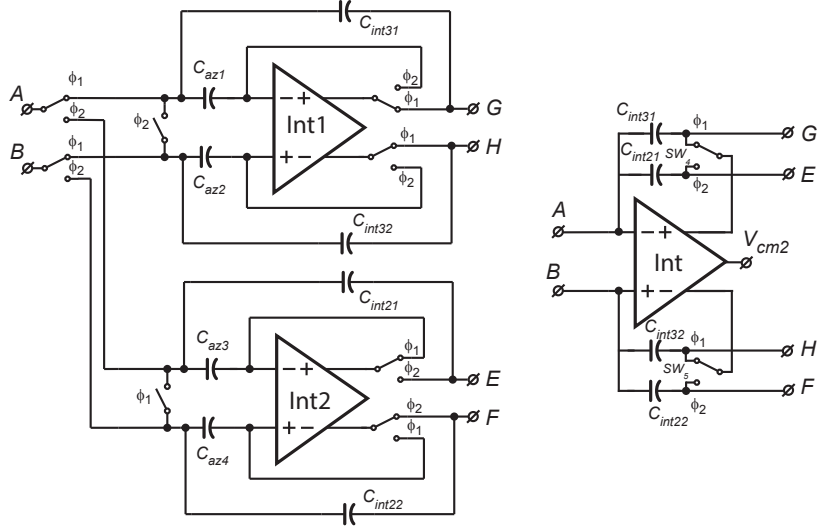
Figure 2.7: The two implementations of the active integrator. On the left two auto-zeroed integrators and on the right a shared integrator.

2.3.2. DEAD TIME

The active integration introduced in the previous section still has limited bandwidth and therefore will only be able to partially reduce the transients. To further mitigate the effect of the limited bandwidth of the integrator, a special clocking scheme is used shown in Figure 2.8, similar to the one used in [37].
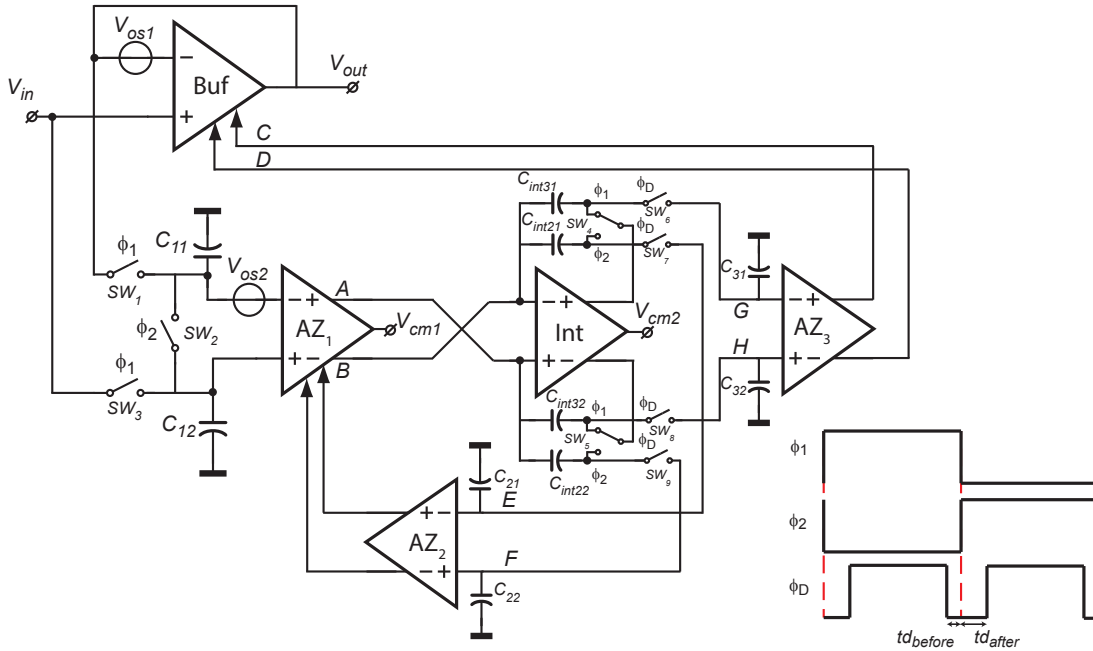
Figure 2.8: A continuous-time auto-zeroed buffer with active integration and dead time with the timing of the dead time (ϕ_d) and the auto-zeroing phases.

As shown in Figure 2.8, an extra clock is introduced (ϕ_D) besides the regular auto-zeroing clocks (ϕ_1 and ϕ_2). This extra clock is referred to as the dead time clock. The dead time clock has a dead band around the auto-zeroing moments. This clock signal controls four switches (SW_6, SW_7, SW_8 and SW_9) around the active integrator, shown

in Figure 2.8. These switches open a little time before (td_{before}) the auto-zero switching moment and close a little time after the auto-zeroing switch moment (td_{after}). This results in four different phases that the auto-zeroing circuit goes through, which are shown in Figure 2.9.

Due to the fact that the potentials at nodes GH and EF are continuously held by capacitors C_{31}, C_{32}, C_{21} and C_{22}, the output of the integrator can settle to almost the correct value during the time that the dead-time switches are open. A sufficient large td_{after}, allows the integrator to settle to the right value and then reconnect the integrator output to the appropriate node, either EF or GH. This circumvents the direct propagation of the transients due to the settling behavior of the integrator.
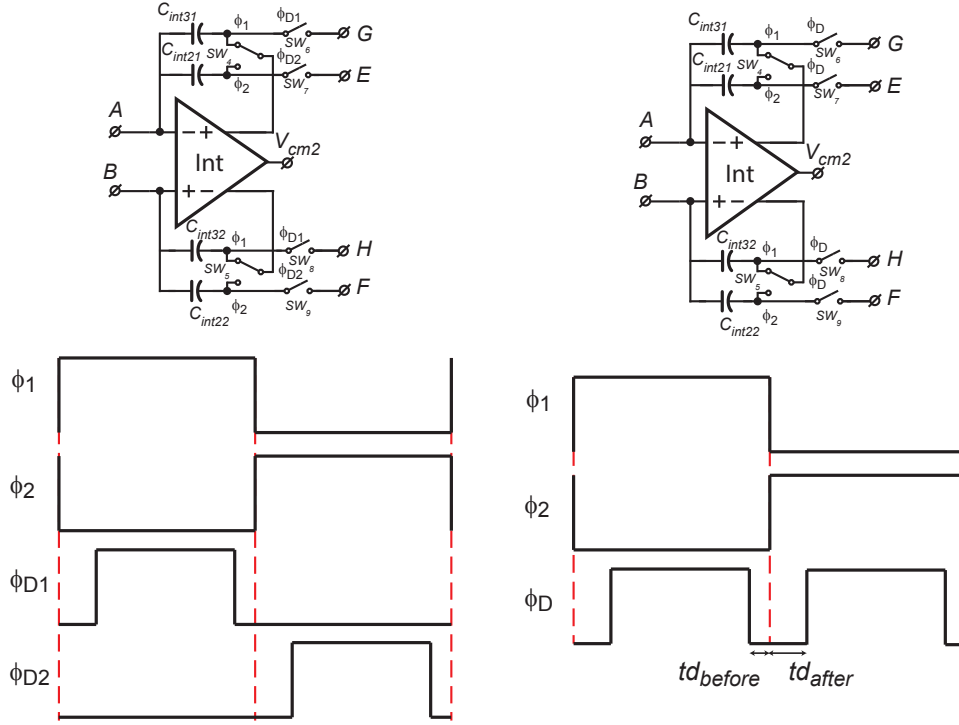
Figure 2.10: Two options for the dead time implementation. On the left the option with two deadtime clocks and on the right with a single shared deadtime clock.

Two options for the dead time were investigated, as shown in Figure 2.10. The first option is to have a deadtime clock for each auto-zeroing phase. Since during phase ϕ_2 capacitors C_{31} and C_{32} are not refreshed, they are simply left disconnected from the integrator during this phase. The second option is to have one shared deadtime clock for both phases. During phase ϕ_2 capacitors C_{31} and C_{32} are now reconnected to the integration capacitors during phase ϕ_2 even though they are not recharged. This increases the switching activity and the number of spikes appearing at the output, but since these spikes are attenuated by a factor 30, this is not a real problem. The advantage of the single dead time, is that it lowers the complexity of the clock generation circuitry. Therefore the option of one shared deadtime clock was picked for the final implementation.

Even though more switches are required by the introduction of a dead time, this will not lead to an increased charge injection. The charge injection due to the switches controlled by ϕ_1 and ϕ_2 will not propagate to the output due to the dead time. The remaining charge injection is due to the charge injection mismatch of the deadtime switches.

2.3.3. Source degeneration
For the OTA's AZ2 and AZ3 source degeneration is used. This increases the input voltage range over which the transconductance of AZ2 and AZ3 behave linearly and decreases their transconductance. Therefore large voltage spikes at the input of AZ2 and AZ3 will lead to smaller spikes in the output current domain. This will help to reduce the output spikes due to the spikes at GH and help to improve the settling behaviour.

(a) Phase ϕ_2 and dead time switches open.

(b) Phase ϕ_2 and dead time switches closed.

(c) Phase ϕ_1 and dead time switches open.

(d) Phase ϕ_1 and dead time switches closed.
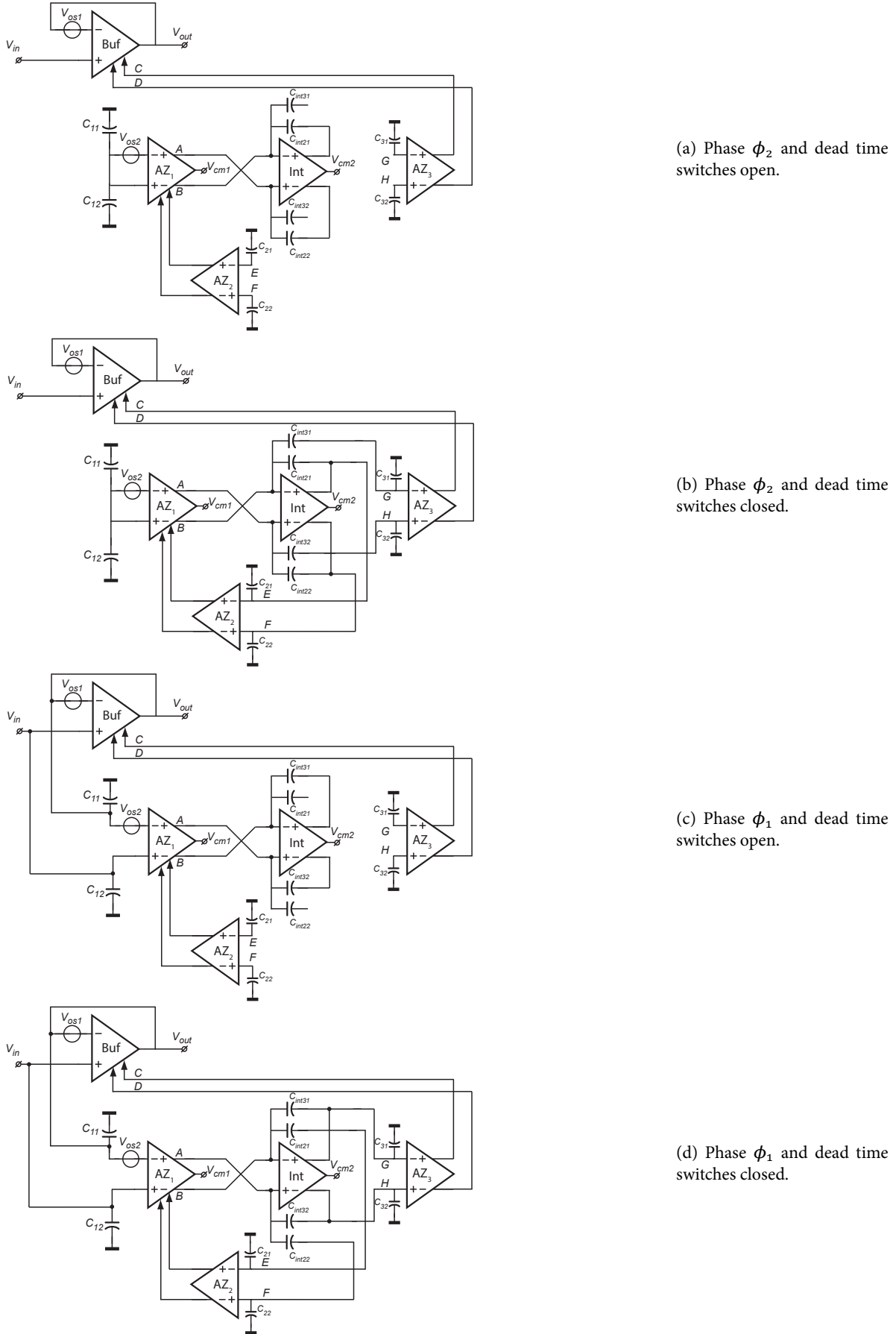
Figure 2.9: The different phases that the continuous-time auto-zeroed buffer with active integration and dead time goes through.

The source degeneration will also lower the transconductance of AZ2 and AZ3. This will increase the ratio between the AZ1 and AZ2 gain and AZ3 and buffer gain. As will be explained in the next section, this will reduce the sensitivity to the charge injection mismatch and will result in lower offset and lower transients.

The transconductance of AZ2 and AZ3 with source degeneration is given by:

$$g_{m,deg} = \frac{g_m}{1 + g_m R_s} \tag{2.2}$$

Where g_m is the transconductance of the input transistors and R_S is the magnitude of the source degeneration resistor.

With a source degeneration resistor of 100 kΩ and the transconductance of AZ2 and AZ3, equal to $g_{m\_az2} = g_{m\_az3} = 11.5\ \mu$S. The degenerated transconductance will be around 5.35 μS. With an AZ1 and buffer transcondutance of about 143.5 μS, the ratio between the buffer and AZ3 gain will be about 30 times.

2.4. Residual offset analysis

In this section the residual offset is analyzed, which can be expected from the circuit in Figure 2.6. As mentioned earlier, the auto-zeroing consists of two phases. Phase ϕ_2 in which the offset of AZ1 (Vos2) is measured and stored and phase ϕ_1 in which the offset of the buffer (Vos1) is measured and stored.

Phase ϕ_2
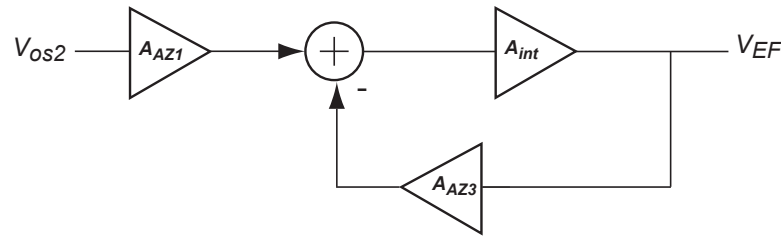
Figure 2.11: The equivalent block diagram of the loop measuring Vos2.

In phase ϕ_2 the equivalent block diagram of the circuit measuring the offset of AZ1 (Vos2) is shown in Figure 2.11. This loop settles at :

$$V_{EF} = \frac{A_{AZ1}}{\frac{1}{A_{int}} + A_{AZ3}} V_{os2}. \tag{2.3}$$

Where A_{AZ1} is the gain of AZ1, A_{AZ3} is the gain of AZ3 and A_{int} is the gain of the active integrator. Since $A_{int} \gg 1$ and the output impedance seen by A_{AZ3} and A_{AZ1} are the same, the final value of EF is determined by the ratio of the transconductance of AZ1 and AZ3.

$$V_{EF} = \frac{g_{m\_az1}}{g_{m\_az3}} V_{os2} \tag{2.4}$$

At the end of phase ϕ_2, V_{EF} is sampled and the charge injection associated with this sampling will introduce an error ΔV_{EF}, therefore the sampled value is:

$$V_{EF} = \frac{A_{AZ1}}{\frac{1}{A_{int}} + A_{AZ3}} V_{os2} + \Delta V_{EF}. \tag{2.5}$$

Phase ϕ_1
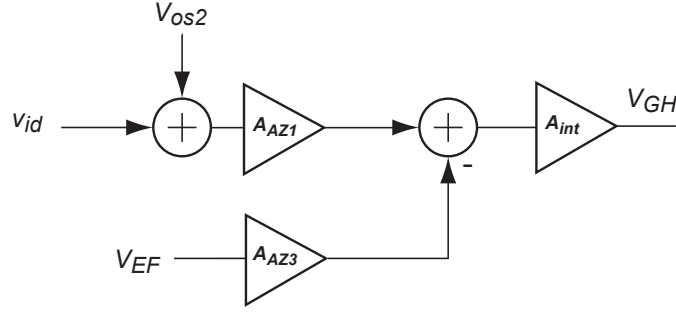
Figure 2.12: The equivalent block diagram of the loop measuring Vos1.

In phase ϕ_1 the equivalent block diagram of the circuit measuring the offset Vos1 of the buffer is shown in Figure 2.12. The voltage on V_{GH} is given by:

$$V_{GH} = A_{AZ1}(v_{id} + V_{os2}) - A_{AZ3}V_{EF}. \tag{2.6}$$

Where v_{id} is the differential voltage seen at the input of the buffer. Combining Equation 2.6 and the previously found equation for V_{EF} (2.5) leads to:

$$V_{GH} = A_{AZ1}A_{int}v_{id} + \frac{A_{AZ1}A_{int}}{1 + A_{AZ3}A_{int}}V_{os2} - A_{AZ3}A_{int}\Delta V_{EF}. \tag{2.7}$$
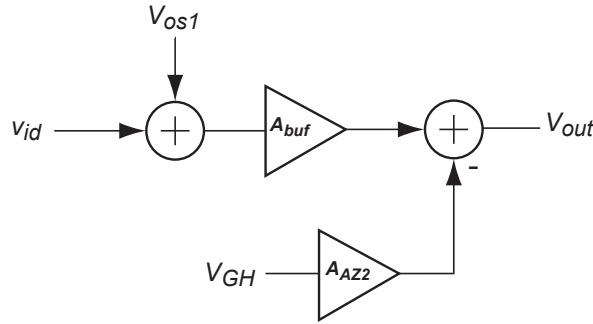
Figure 2.13: The equivalent block diagram of the circuit removing the offset of the buffer.

The equivalent block diagram of the circuit continuously working to remove the offset of the buffer is shown in Figure 2.13. The output voltage of the buffer is given by:

$$V_{out} = (v_{id} + V_{os1})A_{buf} - V_{GH}A_{AZ2}. \tag{2.8}$$

Combing this with the previously found equation for V_{GH} (2.7), gives:

$$V_{out} = (A_{buf} + A_{AZ1}A_{AZ2}A_{int})v_{id} + A_{buf}V_{os1} + \frac{A_{AZ1}A_{AZ2}A_{int}}{1 + A_{AZ3}A_{int}}V_{os2} - A_{AZ2}A_{AZ3}A_{int}\Delta V_{EF}. \tag{2.9}$$

According to the definition of residual offset: $v_{id} = V_{os\_res}$ and $V_{out} = 0$. Using this definition on the previously found expression of V_{out}, gives the following approximate expression for the residual offset:

$$V_{os\_res} \approx \frac{A_{buf}}{A_{AZ1}A_{AZ2}A_{int}}V_{os1} + \frac{1}{A_{AZ3}A_{int}}V_{os2} + \frac{A_{AZ3}\Delta V_{EF}}{A_{AZ1}}. \qquad (2.10)$$

This assumes that the gain of the offset reduction loop ($A_{AZ1}A_{AZ2}A_{int}$) is much larger than the gain of the buffer alone. If we now define the gain ratios: $\alpha_1 = A_{buf}/A_{AZ2}$ and $\alpha_2 = A_{AZ1}/A_{AZ3}$, this leads to:

$$V_{os\_res} \approx \frac{\alpha_1}{A_{AZ1}A_{int}}V_{os1} + \frac{\alpha_2}{A_{AZ1}A_{int}}V_{os2} + \frac{\Delta V_{EF}}{\alpha_1}. \qquad (2.11)$$

This equation is valid for phase ϕ_1, but for phase ϕ_2 V_{GH} is sampled which leads to a sampling error ΔV_{GH}. This sampling error changes output by: $\Delta V_{out} = A_{AZ2}\Delta V_{GH}$, which results in:

$$V_{os\_res} \approx \frac{\alpha_1}{A_{AZ1}A_{int}}V_{os1} + \frac{\alpha_2}{A_{AZ1}A_{int}}V_{os2} + \frac{\Delta V_{EF}}{\alpha_1} + \frac{\Delta V_{GH}}{\alpha_2} \qquad (2.12)$$

for phase ϕ_2. From equation 2.12 it is clear that when $A_{int} = 1$ (which is the case for the conventional circuit) it is not trivial to get low offset. For a low sensitivity to charge injection, the gain ratios α_1 and α_2 should be large. At the same time the first two terms of Equation 2.12 require low ratios α_1 and α_2. The addition of active integration allows the first terms to be minimized by making a large integrator gain (A_{int}). While at the same having a low sensitivity to charge injection by having large gain ratios α_1 and α_2. The offset of AZ2, AZ3 and the integrator were not taken into account in this analysis but since they are also present in the sampling and compensation loops they will also be compensated for by the loops.

With Equation 2.12 and some typical gain values, the expected residual offset can now be calculated. The typical gain of the integrator (A_{int}) is around 105 dB. The typical gain of OTA AZ1 (A_{AZ1}) is around 55 dB, with a typical offset for the buffer of OTA AZ1 around 5 mV and charge-injection step of 3 μV. The residual offset will be:

Term 1 and 2:	Term 3 and 4:	Total:
$\frac{\alpha_{1,2}}{A_{AZ1}A_{int}}V_{os1,2}$	$\frac{\Delta V_{EF}}{\alpha_1}$ and $\frac{\Delta V_{GH}}{\alpha_2}$	
= 1.5 nV	= 3 μV/30 = 100 nV	= 203 nV

Table 2.2: The calculated residual offset contribution for typical values of the gain.

From Table 2.2 it is clear that the residual offset will typically be around 203 nV for phase ϕ_2 and around 103 nV for phase ϕ_1. The main contributor to the residual offset is the due to charge injection mismatch of the deadtime switches.

2.5. NOISE ANALYSIS

A low-noise bandgap voltage reference usually requires significant current to lower the noise generated by the bandgap core and large load capacitors to limit the bandwidth. A sampled voltage reference can reduce the bandwidth of the output noise by sampling at a very low frequency, while the same action will also lower the average required supply current. The output noise above the sampling frequency is determined by the noise of the buffer, which can be made low-noise and low-power by using dynamic offset compensation.

In this paragraph two noise reduction mechanisms are explained. The first is the reduction of the noise bandwidth, by the noise shaping caused by sampling and holding at a low frequency. The second is the reduction of the low frequency noise above the refresh frequency, by auto-zeroing the buffer which reduces $1/f$ noise.

2.5.1. GETTING LOW NOISE IN THE 0.1 TO 10 HZ BANDWIDTH

For voltage references used in low frequency application the most important noise is in the 0.1 to 10Hz range. The sample and hold function will not reduce the total integrated noise, but it will shape the noise. Due to the low

refresh frequency, the input noise will be undersampled. This will cause all the noise of the input to foldback to the frequency band below the refresh frequency.

Assuming that the hold time is much larger than the time needed to charge the hold capacitor, the sample and hold circuit can be seen as an ideal sampler combined with a zero-order hold. In this case it can be deduced [38] that the PSD of the hold voltage noise is given by:

$$S_h(f) = \text{sinc}^2(\pi f T_s) \sum_{n=-\infty}^{\infty} S_{in}(f - \frac{n}{T_s}). \tag{2.13}$$

In equation 2.13 the left term is the noise shaping due to the hold function and the right term the noise folding due to the under-sampling of the input noise.

Assuming that the input noise is ideally low-pass filtered white noise with a noise level of S_0 and bandwidth B, equation 2.13 reduces to [38]:

$$S_h(f) = \frac{2B}{f_s} S_0 \, \text{sinc}^2(\pi f T_s). \tag{2.14}$$

In equation 2.14 the first term is the under-sampling ratio. This equation shows that the noise is reduced in bandwidth by the sinc function, but at the same time it is increased in that smaller bandwidth due to the undersampling.
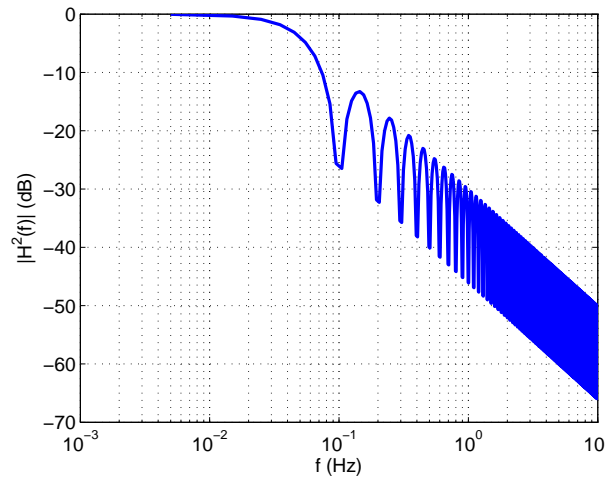
Figure 2.14: The sinc function for a refresh frequency of 10s.

If the hold time of the sample and hold circuit can be 10s or higher than all the noise will be folded to a bandwidth below 0.1Hz, due to the noise shaping of the sinc function. For a refresh frequency of 10s the noise shaping done by the zero-order hold is shown in Figure 2.14. From the figure it is clear that the main lobe stops at 0.1Hz, focusing the majority of the noise below this frequency.

The sampling action via the on-resistance of the switches will also introduce kT/C noise. For a 30 pF capacitor the kT/C noise will be 12 μV_{rms}.

2.5.2. AUTO-ZERO NOISE

During the hold period, the noise above the refresh frequency is fully determined by the output noise of the buffer. Since the buffer is implemented in CMOS, the output noise will be dominated by $1/f$ noise up until a few kHz, such as shown in Figure 2.15. To get a low $1/f$ noise and offset, auto-zeroing is employed for the buffer. The advantage of auto-zeroing is that the low frequency noise is removed by the auto-zeroing action, due to the correlation between the sampled noise that is subtracted from the real-time noise. The disadvantage is that due to the sampling action,
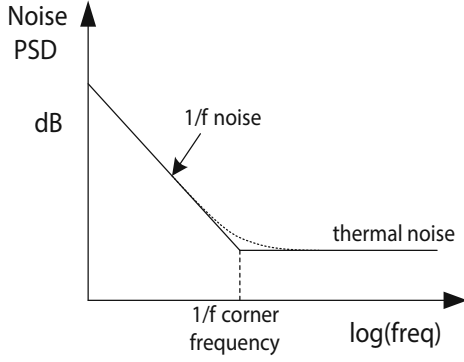
Figure 2.15: Noise PSD of a typical CMOS amplifier [1].
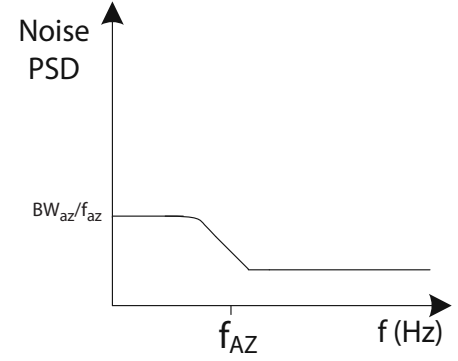
Figure 2.16: Noise PSD of an auto-zeroed amplifier [1].

all noise above the auto-zeroing frequency, is folded back due to aliasing. Most auto-zeroing amplifiers take the bandwidth of the auto-zeroing loop to be about five times larger than the auto-zeroing frequency. This allows the auto-zeroing loop to fully settle within one period. The resulting output noise spectrum is shown in Figure 2.16 .

To achieve a low output noise two aspects are important. One, the auto-zeroing frequency should be higher than the $1/f$ corner of the non-auto-zeroed buffer. Otherwise the $1/f$ noise will fold back. Secondly, for minimal fold back the bandwidth of the auto-zeroing loop should be well below the auto-zeroing frequency. As shown in figure 2.16, the foldback is determined by the ratio of the auto-zeroing loop bandwidth (BW_{az}) and the auto-zeroing frequency.

The potential drawback of having a bandwidth of the auto-zeroing loop which is much lower than the auto-zeroing frequency is that the auto-zeroing now requires several auto-zeroing cycles to settle. However, for most voltage reference applications taking a while after startup to settle is acceptable. The auto-zeroing loop bandwidth (BW_{az}) is limited by the integrator bandwidth. The integrator bandwidth can be set by picking a value of the integration capacitors, that ensures that the integrated noise over the 0.1 to 10 Hz bandwidth is within the required specification of 500 nV_{rms}. Taking integration capacitors that ensure that this requirement is met, allows the auto-zeroing loop to settle as fast as the noise requirement allows.

Figure 2.17 show the output noise spectrum of the buffer. It shows that the $1/f$ noise start to flatten above 200 kHz. This shows that a rather high auto-zeroing frequency is necessary, to keep the $1/f$ noise from folding over too much, due to the relatively small area of the input transistors of the buffer and AZ1.

The transconductance of the buffer is around $g_m = \frac{I}{nV_t}$, where nV_t = 60 mV. With a current of 10 μA, the transconductance will be about 167 μS. This will give a noise spectral density of: $V_n = \sqrt{16kT/gm} = 20 \, nV/\sqrt{Hz}$.

Figure 2.18 shows the output noise spectrum of the buffer with auto-zeroing. The auto-zeroing frequency is chosen as 100 kHz. In the output spectrum some folding can be observed. For low frequency the noise power spectral density is 140 nV/$\sqrt{Hz}$. The integrated output noise over the 0.1 to 10 Hz bandwidth is with 440 nV_{rms} below the required 500 nV_{rms}.

The noise is within specifications but could actually be a lot lower. Due to the fact that the auto-zeroing frequency is chosen slightly lower than the actual $1/f$ corner, some $1/f$ noise will be folded back. The top contributors are flicker noise from AZ1 and the buffer. The auto-zero bandwidth is defined by $BW_{az} = g_{m\_int}/2\pi C_{int}$ which is around 240 kHz for this design. This makes the ratio of auto-zero bandwidth and auto-zeroing frequency, around a factor 2.4. Therefore the contribution of the folded white noise of AZ1 and the buffer is around 70 nV/$\sqrt{Hz}$. This is about half of the achieved 140 nV/$\sqrt{Hz}$. Increasing the size of the input transistors in AZ1 and the buffer could further reduce the noise.
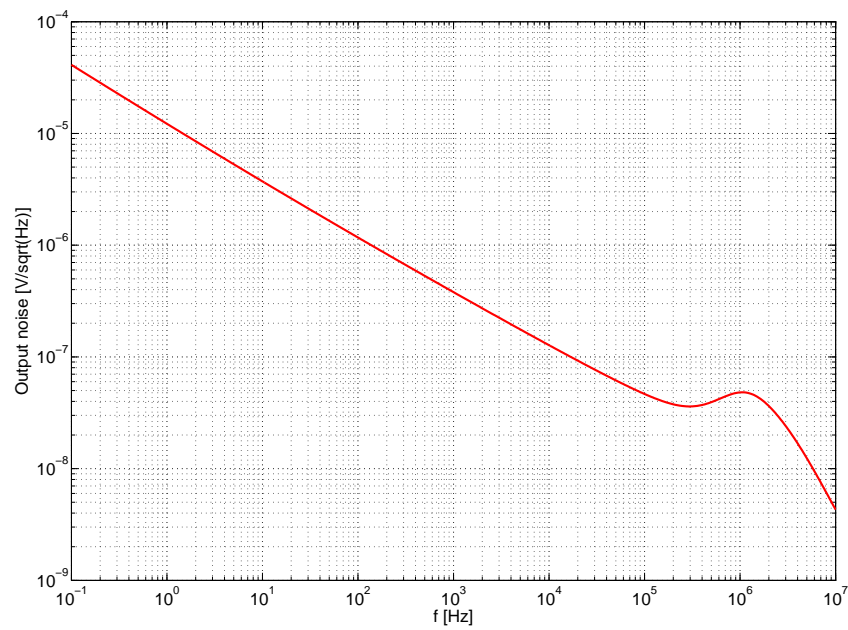
Figure 2.17: The output noise spectrum of the buffer without auto-zeroing, with logarithmic y and x-axis.
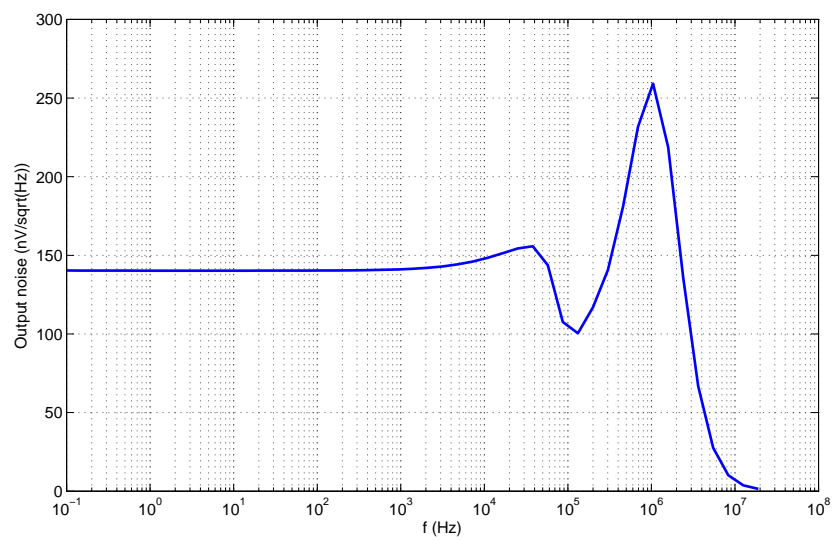
Figure 2.18: The output noise spectrum of the buffer with auto-zeroing at 100 kHz.

2.6. Achieving a long hold time

The key to realizing a sampled voltage reference with both low-noise bandwidth and low-power, lies in achieving a long hold time. From the previous section it is clear that, to push all the noise below 0.1 Hz, a hold time of at least 10 s is necessary. This long hold time is achieved by reducing the leakage of the hold capacitor. In this paragraph several leakage mechanisms in CMOS switches are explained and several techniques are described to reduce the leakage.

The leakage of the hold capacitor can be split into two different components: one is the leakage due to the sample and hold circuit and two is the leakage due to the auto-zeroed buffer connected to the hold capacitors. These two components can be individually examined and optimized for low leakage.
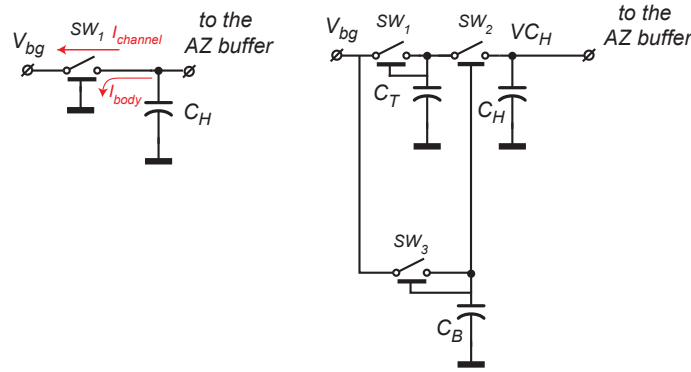
Figure 2.19: On the left a normal S&H circuit and on the right the special low leakage S&H circuit.

In a normal sample and hold circuit, shown in Figure 2.19, the main source of leakage is due to the leakage of the sample and hold switches in the off state. The effects that mainly contribute to this are the leakage of the different junctions, such as the body diode leakage (I_{body}) and the channel leakage ($I_{channel}$).

The leakage of the auto-zero circuit is mainly due to the input switches. For chopper amplifiers it is known that adding switches to the input increases the input current [39, 40]. For auto-zero amplifiers this effect is not well documented. Therefore in the rest of this chapter the input current due to auto-zeroing will also be discussed.

There are several effects that contribute to the hold capacitor leakage, that will be discussed in the following sections.

2.6.1. Body diode leakage

The body and the channel of a MOS transistor form a parasitic PN-junction diode, which is reversed biased under normal conditions. The current through this reversed biased diode can be described by:

$$I = I_S \left[\exp \left(\frac{V_d}{NV_t} \right) - 1 \right].$$

(2.15)

From equation 2.15 it is clear that the diode current equals zero if there is no voltage across the diode. This can be achieved by biasing the body at a bias voltage equal to the drain and source voltage. For a p-substrate process with a n-well, the PMOS body can be biased at a different potential easily by the availability of an isolated n-well. However switches SW_1, SW_2 and SW_3 in Figure 2.19 are created with transmission gates, containing both NMOS and PMOS transistors.

For a NMOS devices in a p-substrate process, all the bodies are connected together by the shared substrate. The use of a triple-well technology allows the creation of a p-substrate area (p-sub) which is isolated from the rest of the substrate. This isolation is achieved by placing a deep n-well around a part of the substrate, which created an

electrically isolated p-substrate area. The availability of this isolated p-sub allows the body of the NMOS used in the transmission gate to be biased at a different potential.
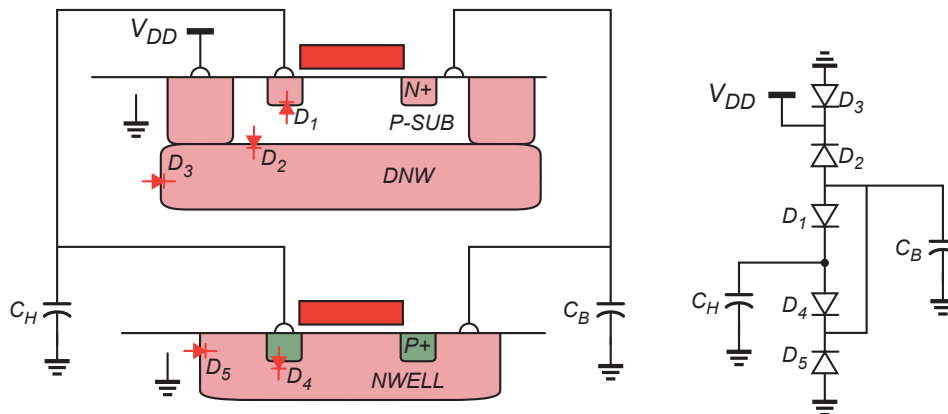
Figure 2.20: The different parasitic diodes of the T-gate implementation of switch SW2 in a tripple well process

In Figure 2.20 the cross-section of such a transmission gate implementation is shown, together with the parasitic diodes. For the PMOS, there is a diode between the source/drain and the NWELL (D_4) and between the substrate and the NWELL (D_5). The NMOS implementation adds a diode between the drain and the p-substrate (D_1), a diode between the p-sub and the DNW (D_2) and a diode between the DNW and the substrate (D_3).

By using the body biasing, as long as the hold capacitor voltage (Vch) is equal to the body capacitor voltage (Vcb), diode D_1 and diode D_4 will not leak any current. This effectively protects the hold capacitor from the body diode leakage. If Vcb is slightly higher than Vch (Vcb>Vch) then diode D_1 will become forward biased and diode D_4 will become reversed biased. If Vcb is slightly lower than Vch (Vcb<Vch) then diode D_1 will become reversed biased and diode D_4 will become forward biased. Assuming that the reverse saturation current is similar for the NMOS and PMOS and since the PN junction cross section areas have been chosen equal, these two components cancel each other for small differences.

Diode D_2 and diode D_5 are reverse biased diodes. The reversed saturation current of diode D_2 will charge capacitor C_B, while the reversed saturation current of D_5 will discharge capacitor C_B. Again assuming that the reverse saturation current is similar for both diodes and assuming that the PN junction cross section area is kept similar in layout, then these currents will cancel if the two diodes are biased with the same voltage. Since diode D_2 is biased at VDD-Vcb and diode D_4 is biased at Vcb they will only cancel for Vcb = Vdd/2.

Another option is to buffer the body capacitors, so that the remaining leakage current is supplied by the buffer and not by the capacitors. This however creates another problem. Any offset of this buffer will bias diodes D_1 and D_4 and since the characteristic of a diode is quite steep and inherently asymmetrical, due to the exponential, this offset has to be very small to keep it from creating a considerable charging current. Off course there is a low offset buffered version of the hold capacitor voltage available at the output. However the spikes on the output will be rectified by the diode D_1, which will in turn lead to a charging effect.

A last option is to refresh the body capacitor more often than the hold capacitor. This helps in case that the body voltage capacitor runs down too fast.

2.6.2. Channel leakage
Even when a MOS transistor is turned off there will still be a slight conduction between the drain and source. A common way to reduce this effect is to use long channel devices or high threshold devices, to create a large off resistance. However both of these options are undesirable, due to the increased charge injection. Another

technique is to keep the voltage between the source and drain zero, to keep charge from leaking. When the bandgap reference is turned off, after charging the hold capacitor, there will be a large potential difference across switches SW_1 and SW_3 in figure 2.19. The temporary capacitor (C_t) will help to keep the potential difference over switch SW_2 small, but only temporary as the channel leakage of switches SW_1 and SW_2 will quickly drain capacitor C_T.
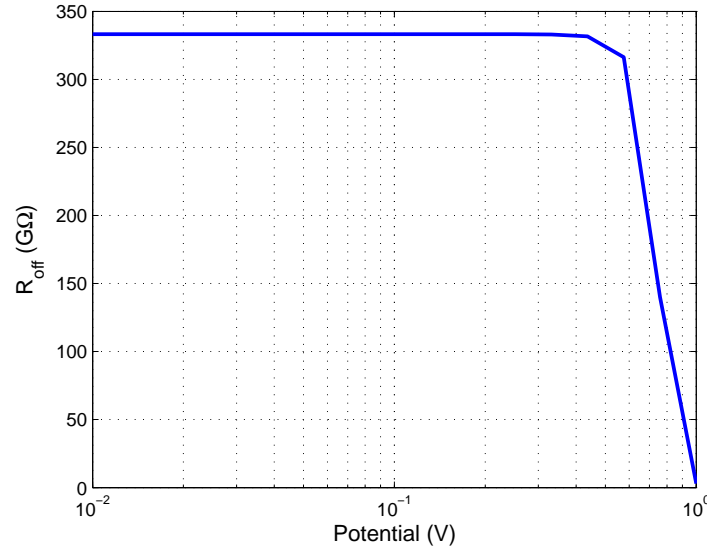
Figure 2.21: The off resistance against the potential across the switch.

Figure 2.21 shows the off resistance for different potentials across the switch. The switch is the transmission gate implementation used in switch SW_1, SW_2 and SW_3 of the sample and hold, with minimum size PMOS and NMOS transistors. It can be seen that for large potentials (> 300 mV) the off resistance decreases. For small potentials the off resistance is around 333 GΩ. If the full reference voltage (1.024 V) is left over the switch, the off-resistance is reduced to 3 GΩ. This will drain the whole hold capacitor in less than 100 ms. Therefore additional measures have to be taken to keep the potential across the switch small.

A good solution, is shown in Figure 2.22. Here a buffered version of the hold capacitor voltage is applied at the input during the moment that the bandgap reference is shutdown. However, in this case the offset of the buffer determines the leakage of the hold capacitor.

For a typical offset of several mVs, the hold time can be several 100 ms, while keeping the voltage droop within several μVs. For a 1 μV offset the voltage droop per unit time is around 0.1 μV/s. This will allow for a hold time in the order of tens of seconds, while the error on the hold capacitor is still in the μV-range. This trend is confirmed by [41]. To achieve this low-offset, auto-zeroing can also be applied for this buffer. Since the noise is not an issue in this case, the auto-zeroing frequency can be much lower. The current that the buffer needs to supply is around 3 pA per switch. The main buffer might be used for this purpose, however, care must be taken that the switching transients do not cause a charging effect. Due to the limited design time this option was not fully investigated and therefore not implemented in the final design, but it is certainly a good option for a future design. This option to reduce channel leakage, has been successfully applied in [19].
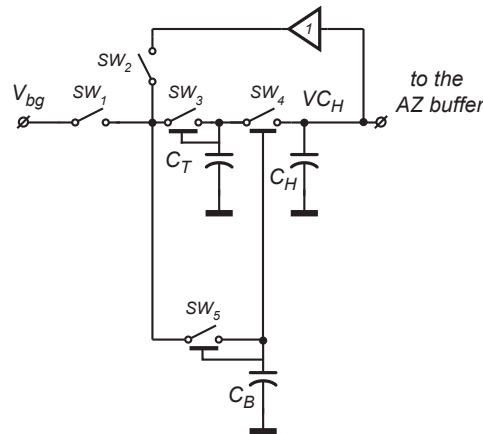
Figure 2.22: Low leakage S&H circuit with a buffer to supply the channel leakage current.

2.6.3. CAPACITOR SELECTION

A typical standard CMOS technology allows for three different capacitor types: fringe capacitors, metal-insulator-metal (MIM) capacitors and metal-oxide-semiconductor (MOS) capacitors. In the TSMC mixed signal 0.18 μm technology used in this design, only a standard cell is supplied for a MIM capacitor. The MIM capacitor is fabricated using a special metal layer, which allows the top and bottom plate of the MIM capacitor to be as little as 38 nm apart. This allows for a high capacitance per unit area but at the same time the specified leakage, $< 1 \frac{fA}{\mu m^2}$, is quite high. At these small distances with high electrical field densities, the tunneling current is a significant component of the capacitor leakage [42]. Furthermore the standard cell uses the top metal layer to create the top and bottom plate, which doesn't allow for a top shielding layer to be added.

To make sure that the tunneling current was not a problem a fingered fringe capacitor is used in this design. Since the normal metal layers are being separated by an insulator with a thickness of 0.26 μm, the tunneling current is no longer a significant component in the leakage current. Since no standard cell existed for a fingered fringe capacitor, a custom layout was used. For the creation of the fingered fringe capacitor, metals 1 to 5 were used which left metal 6 as a top shield. Together with a substrate with enough substrate ties, this allows for a shielded capacitor.

2.6.4. CHARGE INJECTION/ CLOCK FEEDTHROUGH

Switch SW_2 in Figure 2.23 is directly connected to the hold capacitor. Once the hold capacitor is charged to the right voltage and switch SW_2 opens, some charge will be injected into the hold capacitor. This charge will cause the final sampled voltage to be slightly off from the desired value. As described earlier since this switch is critical to the functioning of the circuit, the charge injection and clock feedthrough can only be compensated for and cannot be completely circumvented. For compensation, transmission gates are used. Since switch SW_2 only closes once per refresh period, this error should not be very problematic.

2.6.5. INPUT CURRENT OF THE BUFFER AND AZ1

The positive input terminal of the buffer amplifier is directly connected to the hold capacitor. Any input current of the buffer will therefore directly drain the hold capacitor. For CMOS input transistors the input current is generally very small. This leakage also marks the minimal achievable leakage due to the buffer.

2.7. CHARGE KICKBACK

The buffer is making a loadable copy of the hold capacitor voltage. However due to several reasons this copy is not perfect and there is a slight error on the output, in regards to the input.

This error might be due to:

• Unsettled auto-zeroing

• Offset remainder

- Change in load

- Gain error
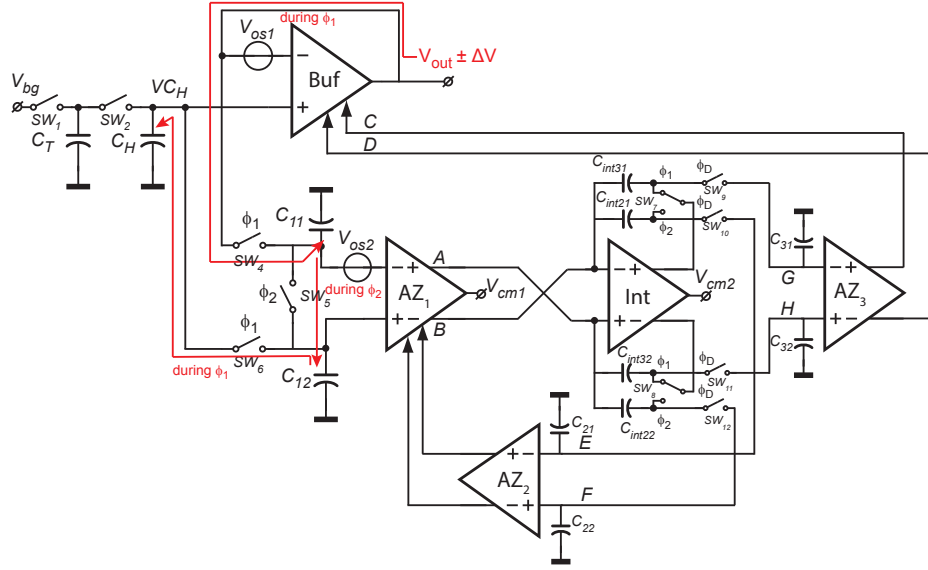
Figure 2.23: Charge kickback caused by an error on the output.

As shown in figure 2.23, the switching scheme of the auto-zeroing switches transfer the error on the output towards the hold capacitor. In the first ϕ_1 state the output voltage $\pm$ the error is sampled on C_{11}. In the next ϕ_2 state this sample is transferred to C_{12} and finally, in the next ϕ_1 state, it is transferred on to the hold capacitor.
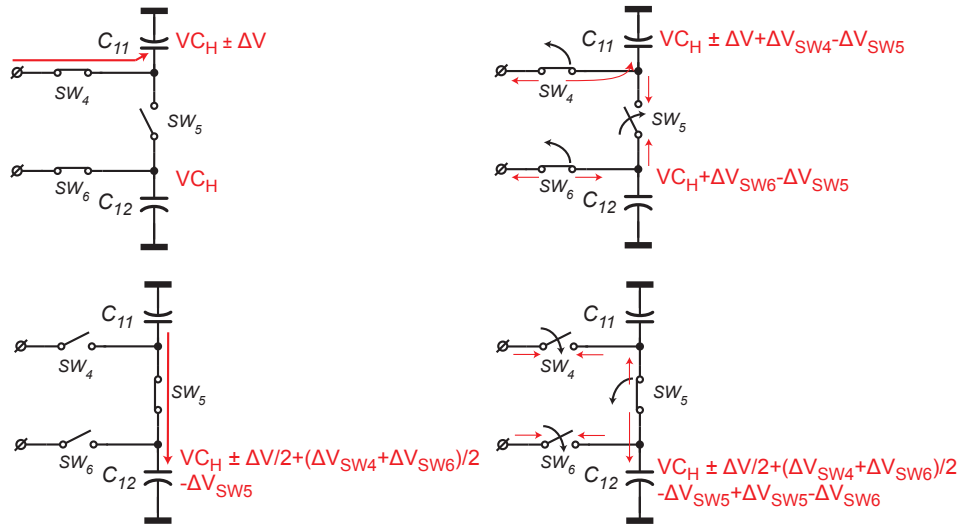
Figure 2.24: Charge kickback with charge injection taken into account.

The above story doesn't take into account any errors due to charge injection. Figure 2.24 shows the charge kickback with charge injection taken into account. After the sampling of $V_{in}\pm\Delta V$ on C_{11}, SW_4 will add charge to C_{11} and SW_5 takes charge. The same is done at C_{12}, where V_{in} is sampled and SW_6 will add charge and SW_5 will take charge. When SW_5 is closed, an average of the voltage on the capacitors will be left. Next, SW_5 will add charge while SW_6 takes charge. Overall, the opening and closing of SW_5 will mostly cancel out, due to the symmetry of drain and source side. The contribution of SW_4 and SW_6 will be somewhat similar. Both switches have a 1 pF

capacitor (C_{11} and C_{12}) on one side and a 30 pF capacitor on the other (C_H and C_{load}). Furthermore, due to the fast transitioning auto-zero clock, the charge will be split equally regardless of the symmetry of source and drain impedance [34]. The transmission gate implementation of the switches, might make the effect of charge kickback very dependent on clock parameters, such as: skew and rise/fall time [36].

Even with very small values for the residual offset, the effect of charge kickback, together with a high auto-zeroing frequency can create a significant leakage current. Therefore the residual offset and auto-zeroing frequency should be as low as possible. The voltage droop rate due to charge kickback is given by:

$$\frac{\Delta V}{\Delta T} = \frac{C_{11} V_{os} f_{AZ}}{2 C_H}. \tag{2.16}$$

This problem is especially apparent during startup when the auto-zeroed loop is not settled yet. The unsettled residual offset is much higher and can have a large effect on the hold capacitor voltage, when it is transferred through the charge kickback described above. This problem can be circumvented by keeping the voltage reference connected to the hold capacitor during startup. This allows the auto-zeroing to remove the offset during several auto-zeroing phases and settle to the final output level, after which the voltage reference can be disconnected again.

2.8. SLOW-CHOPPING
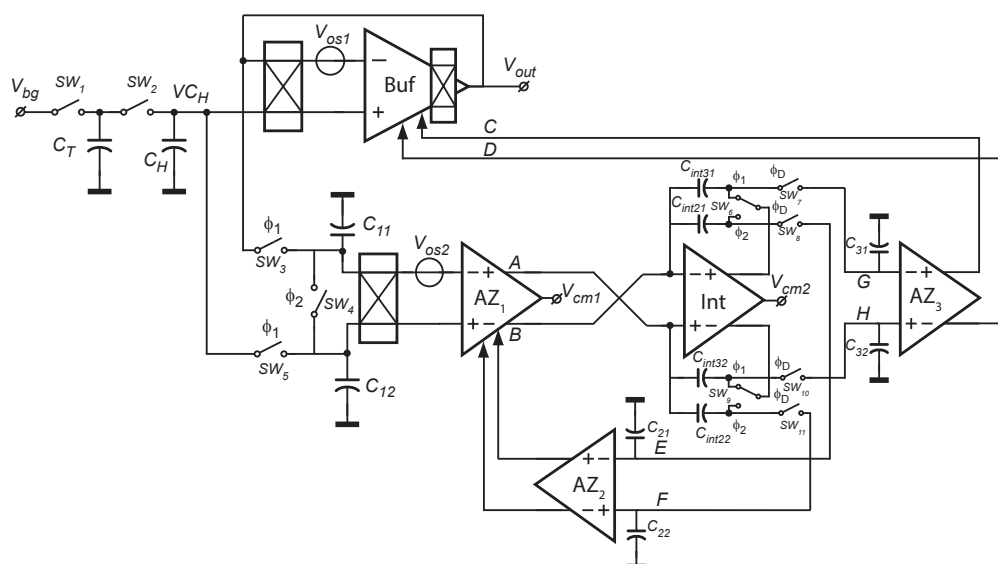
Figure 2.25: Continuous time auto-zeroed buffer with active integration and leakage reversing slow chopping

A special slow-chopping techniques is used to chop the charge kickback leakage described in the previous section. As shown in Figure 2.25, choppers are added at the input of the buffer and OTA AZ1 to reverse the polarity of respectively V_{os1} and V_{os2}. To keep the phase of the direct path via the buffer and offset reduction path the same phase, a chopper is also placed inside the buffer in between the input stage and the output stage.

Even though the polarity of the offset has changed at the output, the polarity remains the same for the auto-zeroing loop and nodes EF and GH. This is important, since the auto-zero loops doesn't have to settle to a new value for a different slow-chopping phase.

The idea behind the slow-chopping is to reverse the polarity of the residual offset from negative to positive and vice versa. The leakage due to the charge kickback of residual offset will then all so change from a leaking to a charging effect and vice versa.

The effectiveness of this technique depends on the fact that the residual offset is equal in magnitude in both chopping phases and only changes in polarity. In section 2.4 the residual offset was analyzed. From equation 2.12 it is clear that the residual offset has a term due to the reduced offset of the buffer and AZ1 and a term due to charge injection. The charge injection term will not be changed by the slow-chopping, so for this technique to be effective this term has to be small. Also any offset introduced by the output stage of the buffer should not be too large, since this is not chopped.

Furthermore the spikes introduced by the chopping itself should not be too high, although the lower chopping frequency makes this less problematic. Also the charge injection and clock feedthrough of the chopper into the hold capacitor should not be too large, otherwise this might circumvent a properly working leakage reversal.
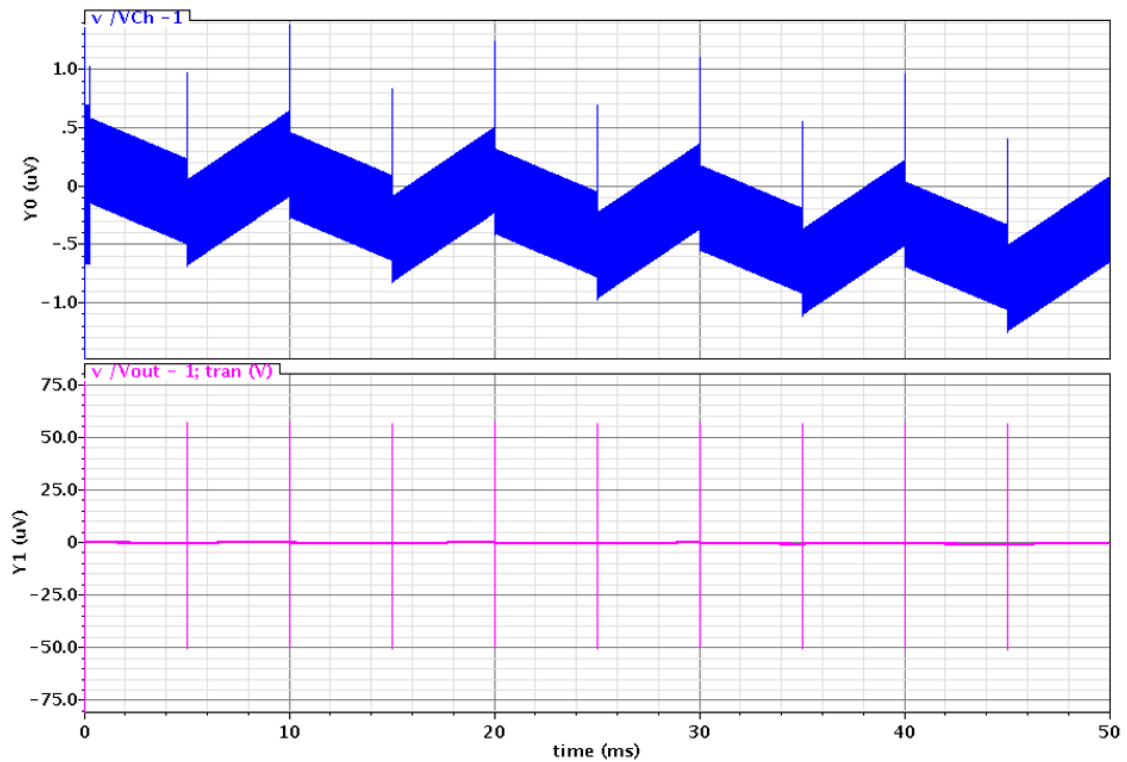
Figure 2.26: Simulation result of the slow chopping.

In figure 2.26 the simulation result of the slow chopping is shown. Every 5 ms the sign of the residual offset is changed by chopping the two offsets. On the hold capacitor a charging and discharging effect can be seen. Still an overall slight discharging effect is present. This is mainly due to the fact that some of the residual offset cannot be chopped. Still in simulation, this reversal of the leakage allows for an overall longer holding time. The bottom trace in figure 2.26 shows that the spikes due to the chopping on the output are around 50 μV.

3

Transistor level implementation details
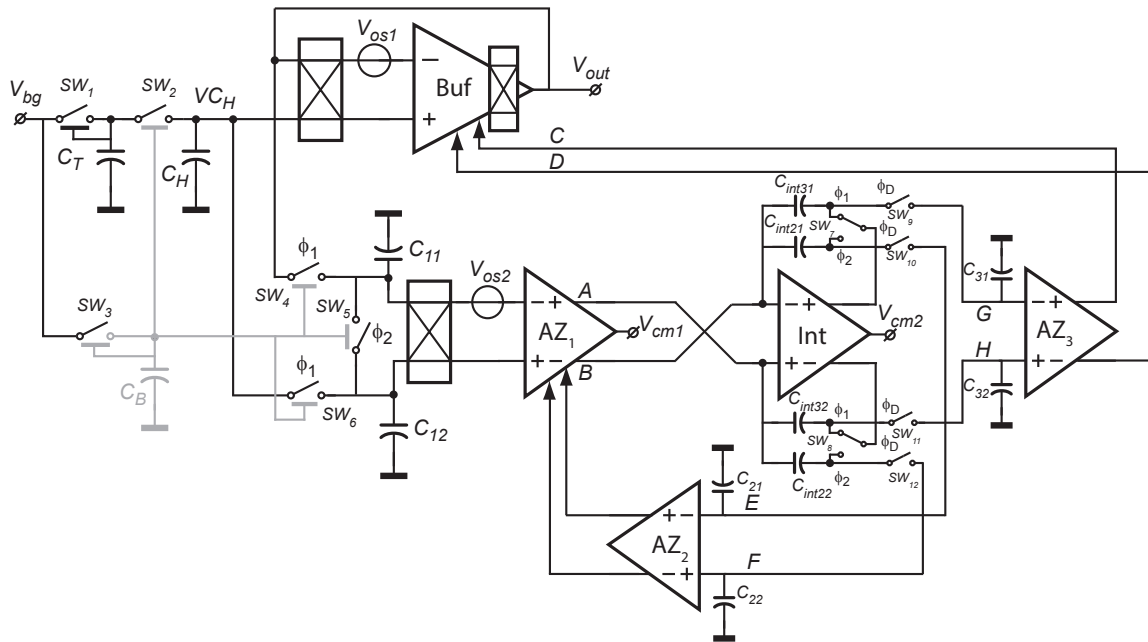
Figure 3.1: The complete circuit with continuous-time auto-zeroed buffer and low-leakage sample and hold.

In Figure 3.1 the complete circuit with continuous-time auto-zeroed buffer and low-leakage sample and hold is shown. In this chapter the transistor level implementation details of the buffer, AZ1, AZ2+AZ3 and integrator are presented. Also some details are given about the digital circuits and layout.

3.1. Buffer transistor implementation
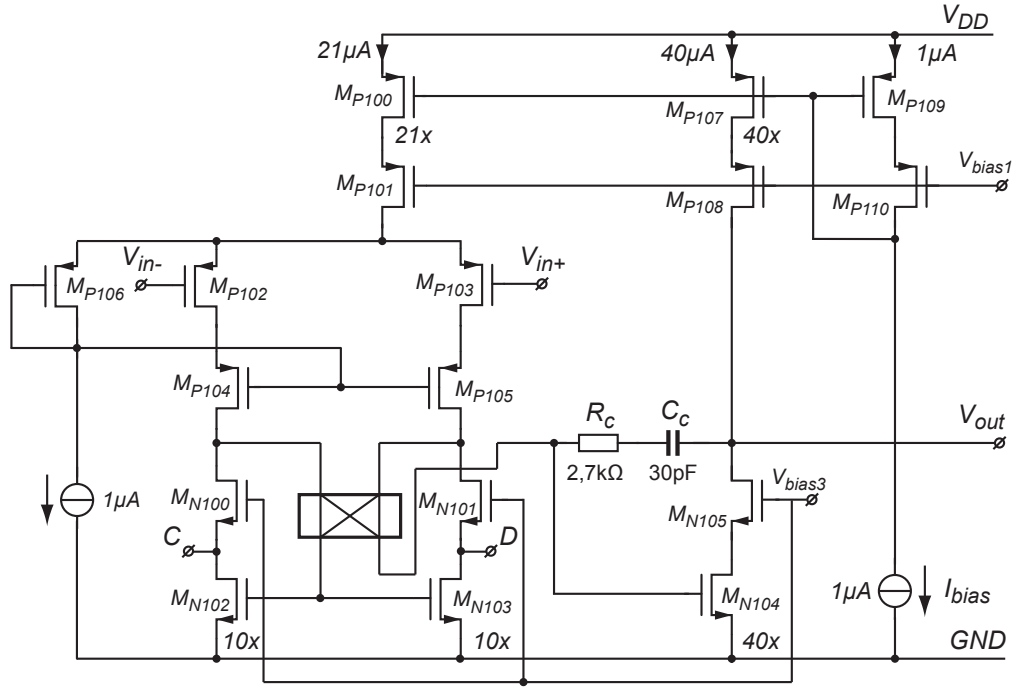
Figure 3.2: Buffer amplifier, two stage: a telescopic amplifier and a class A output stage.

Figure 3.2 show the implementation of the buffer. For the buffer a two-stage design is used. The first stage is a telescopic amplifier, this is chosen above the folded cascode structure for its better power efficiency. From the top, transistors M_{P100} and M_{P101} form a cascode current mirror for a better power supply rejection ratio. For the input transistors: M_{P102} and M_{P103}, PMOS devices are used for their better $1/f$ noise performance. Furthermore the input transistors are medium threshold devices to make sure that they have enough voltage headroom. With a supply voltage of 1.8 V and two times a V_{DS} for M_{P100} and M_{P101}, the sources of the input transistors are at 1.4 V. With an input level of 1 V this only leaves 0.4 V for the V_{GS} of the input transistors, this is enough for the medium threshold transistors. The input transistors are cascoded by M_{P104} and M_{P105}. M_{P106} makes sure that the gates of the cascode devices is at a fixed voltage below the sources of the input devices so there is enough room for a $V_{DS} + V_{GS}$.

The bottom current mirror is formed by M_{N102} and M_{N103}, which is cascoded by M_{N100} and M_{N101}. M_{N102} and M_{N103} are biased in strong inversion to minimize the noise and offset contribution. The auto-zeroing loop is completed by the feedback at C and D, which connects to the output of AZ3.

The output chopper is implemented in between the input stage and the output stage. This chopper is purely there to chop the residual offset, for the reasons described in section 2.8. The input chopper is in front of the input transistors not shown here.

The output stage consists of a cascoded common source stage. The cascode transistor M_{N105} is added and special attention is paid to the sizing to make sure that the chopper switches between to similar voltage levels, which will make the spike at the output due to the copping as small as possible. The maximum output current that can be delivered is set by the top current source at 40 μA.

For a generally applicable buffered reference the maximum output current is rather low. Furthermore due to the class A structure of the output stage, the power efficiency is also low. To improve the power efficiency and maximum current driving strength, the use of a Monticelli class AB output stage has been investigated. This has an increased maximum driving current and power efficiency but problems were encountered with the transients

Figure 3.3: The implementation of AZ1, a telescopic OTA.

introduced by the slow-chopping. Therefore this was not used in the final tapeout.

The total current consumed by the buffer with biasing is 69 μA. The DC-gain of the buffer is around 118 dB. The compensation ensures a phase-margin of 70 Deg and a gain-margin of 40 dB with a load of 30 pF.

3.2. AZ1 IMPLEMENTATION

Figure 3.3 shows the implementation of the OTA AZ1. The structure of AZ1 is basically the same as the buffer. The differences are that this is a single stage implementation and the output is differential rather than single ended. To set the common-mode of the output, a common-mode feedback loop is used which feeds back at M_{N202} and M_{N203}. This continuous-time common-mode feedback ensures that the output common-mode is around 600 mV. The offset compensation loop for the offset of AZ1 feeds back at nodes I and J, which is connected to the output of AZ2. The sizing and bias currents are kept similar to the buffer, to make sure that the offset and noise contributions of the two amplifiers are similar.

The total current consumed by the OTA AZ1 with biasing and common-mode regulator is 32 μA. The DC-gain of AZ1 is around 55 dB.

3.3. AZ2 AND AZ3 IMPLEMENTATION

Figure 3.4 show the implementation of AZ2 and AZ3, which are two telescopic OTA's. The bias current of the input transistors is a factor ten times smaller, to get a transconductance which is a factor ten times smaller than AZ1 and the buffer. Furthermore source degeneration is used, for the reasons mentioned in section 2.3.3. The source degeneration resistor is taken as $1/g_m$ of the input transistors M_{P4} and M_{P5}. This makes the overall transconductance of AZ2 and AZ3 a factor two times smaller, which makes the transcondutance of AZ2 and AZ3 twenty times smaller than the AZ1 and the buffer.

The total current consumed by AZ2 and AZ3 combined is 12 μA. The DC-gain of AZ2 and AZ3 is around 35 dB.

Figure 3.4: The implementation of AZ2 and AZ3.

3.4. INTEGRATOR

Figure 3.5: The integrator implementation

Figure 3.5 show the implementation of the active integrator. As shown in the residual offset analysis (section 2.4), a high DC-gain of the integrator helps to reduce the residual offset. The integrator is a telescopic amplifier with gain-boosting for added DC-gain. The gain-booster amplifiers are implemented differentially, to reduced the required amplifiers from four to two. The gain booster amplifiers are folded cascode amplifiers. The bottom

gain-booster has a PMOS input transistors to accommodate the input common-mode level, near the lower part of the supply range. The top gain-booster has NMOS input transistors to accommodate the input common-mode level, near the upper part of the supply range. The gain-booster both have common-mode regulation which sets the biasing point of M_{P304}, M_{P305}, M_{N300} and M_{N301}. Small transistors are used to allow for a fast integrator.

Care must be taken to ensure that no slow-settling occurs due to the gain-boosting. When slow-settling occurs the td_{after} has to be unnecessary long to allow for sufficient settling.

The total current consumed by the integrator with gain-boosters and common-mode regulator is 44 μA. The DC-gain of the integrator is about 105 dB. Adding up all the contributions mentioned in the previous sections and adding 4 μA for the biasing generator, a total current consumption of 161 μA results.

3.5. Digital design

All switches and choppers in this design are implemented with transmission gates, apart from the switches around the integrator. The ability of these devices to reduce the transients due to their charge injection and clock feedthrough is highly dependent on the synchronism of the complementary clock edges, as described in Section 2.3.

The synchronized complementary signals are generated on-chip by converting single-ended clock signals supplied by an external source. This way was chosen, since it results in a very simple circuit. The circuit used to generate the complementary signals is shown in Figure 3.6, from now on referred to as the complementary clock generator. This complementary clock generator generator is placed close to the corresponding switch, to make sure that both the normal and the complementary signal see the same parasitics.
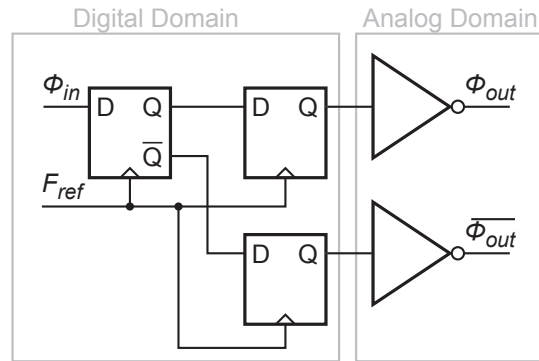
Figure 3.6: The digital circuit used to generate synchronized complementary clock signals from a single-ended input clock.

The complementary clock generator generator consists of three flip-flops and two inverters. The first flip-flop makes sure that the input clock signal (ϕ_{in}) is re-clocked with reference clock (f_{ref}) and generates the complementary version of the input clock signal. After this there will still be a slight delay between the normal clock signal (at Q) and the complementary version (at Q_{bar}), due to a slight difference between the D to Q delay and the D to Q_{bar} delay of a flip-flop. The second and third flip-flop resynchronize both the input clock signal and the complementary version with the reference clock. This will finally lead to a very synchronous ϕ_{out} and $\overline{\phi_{out}}$ signal, generated from a single ϕ_{in} signal. Finally the signals are passed through an inverter fed by the analog supply, to make sure that the digital supply is not coupled directly to the switch that it is steering.

The downside of this approach is that the frequency of the reference clock (f_{ref}) has to be rather high to allow for a high auto-zeroing frequency. Furthermore it has a high change of coupling signals in the analog domain with signal in the digital domain.

Figure 3.7: A block diagram of the complete digital circuit.

In Figure 3.7 a block diagram is shown of the complete digital circuit. The clock signals ϕ_{AZ_in}, ϕ_{D_in}, f_{ref_in}, $fref_{bar\_in}$ and ϕ_{CH_in} are created off-chip. The clock signal ϕ_{AZ_in}, ϕ_{D_in} and ϕ_{CH_in} are all passed through one or several complementary clock generators to generate synchronized complementary signals. For the other signals an on-chip shift-register (SR) is used. This shift register controls switches SW_1, SW_2 and SW_3 from the sample and hold circuit, shown in Figure 3.1. These can be easily controlled with a shift-register, since these are low-frequency signals. The shift-register also controls two test signals: AZ_enable and Int_{out_enable}. The AZ_enable signal allows the auto-zeroing loop to be broken at the feedforward point at nodes C and D, shown in Figure 3.1. The Int_{out_enable} signal controls two transmission gates, that allow the differential output of the integrator to be routed to two bondpads. This allows the integrator output to be measured, when necessary. Finally the shift-register controls a startup signal, that can force the bias circuit to start when it doesn't do this on it's own.

3.6. LAYOUT

The design has been implemented in a TSMC mixed-signal 0.18 μm technology. Even though considerable effort was done to combine the sample and hold circuit and the auto-zeroed buffer with an integrated switched capacitor bandgap reference, in this design it was chosen to omit the integrated reference. This concentrates the attention on testing the hold time of the sample and hold circuit and the switching transients of the auto-zeroed buffer. The reference voltage will be externally supplied by an off-chip reference. Figure 3.8, show a micrograph of the final implemented design. Highlighted in red area some key components of the design.

The active area of the chip is 765 μm by 525 μm. From figure 3.8 it is clear that most of the area is consumed by the different capacitors in the design. Some extra space at the top is used to include decoupling capacitors for the analog supply line. Also a small error in the layout can be spotted at AZ1. Due to a missing exclusion layer, the structure of AZ1 cannot be seen, meaning that the transistors are covered by dummy tiling. This means that the offset of AZ1 will be slightly higher due to the asymmetry caused by the dummy tiling.
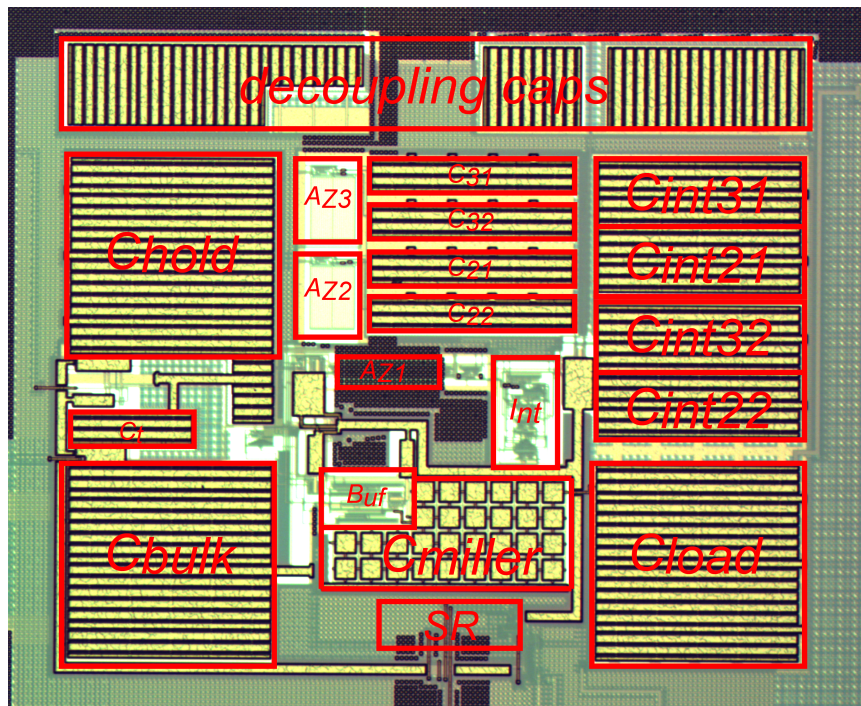
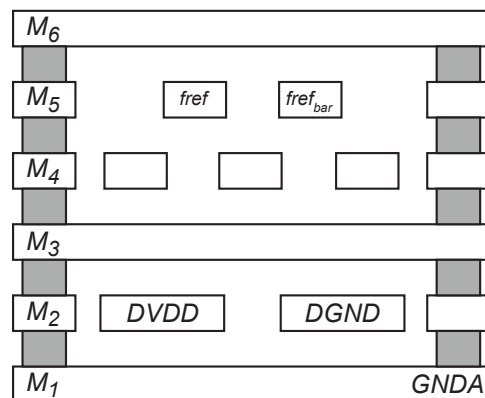Figure 3.8: A micrograph of the implemented chip.

Figure 3.9: Cross section of the shielded clock signals in the 6 Metal TSMC process

The clock signals have to be routed to the complementary clock generators, which are placed close to the sensitive analog signals. Therefore all clock signals can couple easily to the sensitive analog signals. To keep this coupling to a minimum all clock signals are routed towards the complementary clock generators in shielded coax cables on-chip, as shown in Figure 3.9. The downside of routing the clock signals through the shielded coax is that this introduces a large parasitic capacitor between the clock signal and the shield ground. To keep the rise and fall times of the clock signals from increasing to much, strong clock buffers have to be used to drive the shielded clock lines.

For the high frequency reference clock which is routed throughout the chip, also a complementary version is brought to the chip ($fref_{bar}$). This complementary version has no real function on the chip, it is purely routed along side f_{ref} with the idea to cancel any coupling of f_{ref} by an equal coupling of $fref_{bar}$.

<div style="text-align: right">

4

</div>

Results and Future work

This chapter begins by showing the simulated switching transients for the different circuit configurations discussed in section 2.3. After this, the measurement setup is explained and the measurement results are presented. The effect of the reference clock on the output is shown, followed by the measured hold time without auto-zeroing. Next the simulation of the hold time with auto-zeroing is shown, which is compared with the measured hold time. This chapter continues by showing how the different clock signals influence the leakage. Finally, some proposals are made to improve the design. For every measurement a comparison is made with the simulated values whenever possible.

4.1. Simulation of the switching transients

The target specification for the switching transients is that they should be below $1\,\mu$V peak-to-peak. To achieve this several techniques are used to reduce the switching transients. These are active integration, inclusion of a dead-time and source degeneration. The switching transients and residual offset of the different circuit configurations are shown below. It will be shown that with each circuit iteration, the residual offset and switching transients decrease.

For all the following switching transient simulations, the auto-zeroed buffer is first allowed to settle. To make sure of this the data is taken after $100\,\mu s$ from the start of the simulation. The auto-zeroing frequency is 100 kHz. The offset of the buffer (V_{os1}) is taken to be around -4 mV. The offset of AZ1 (V_{os2}) is taken to be around 3 mV.

4.1.1. Conventional continuous-time auto-zeroed buffer

The conventional continuous-time auto-zeroed buffer, shown in Figure 2.3, has no measures to lower the switching transients. This simulation can therefor be seen as a baseline, to see how well the techniques to reduce the switching trasients work.
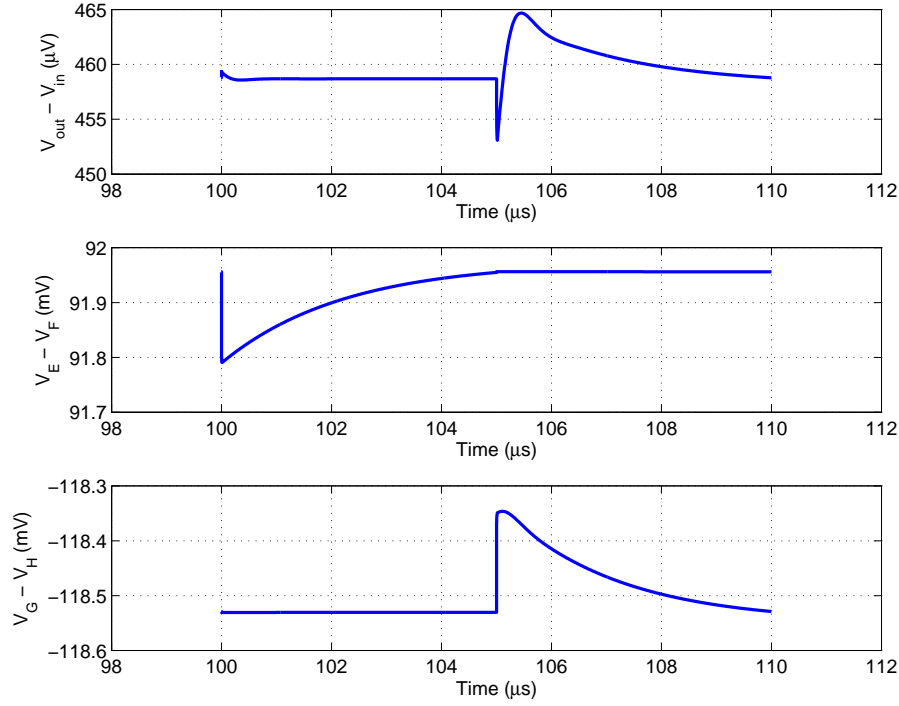

Figure 4.1: The simulated switching transients and residual offset of the conventional continuous time auto-zeroed buffer.

In Figure 4.1 the simulated switching transients on the output versus the input, $V_E - V_F$ and $V_G - V_H$ of the conventional continuous time auto-zeroed buffer (Figure 2.3) are shown. The switching transients due to the limited bandwidth of AZ1, as explained in Section 2.3, can clearly be seen. At t = 100 μs phase ϕ_2 starts, which samples the offset of OTA AZ1 (V_{os2}) on $V_E - V_F$. The settled values will be around V_{os2} x α_1, as explained in Section 2.4, which corresponds to 3 mV x 30 = 90 mV. At t = 100 μs a clear switching transient can be seen on $V_E - V_F$ due to the limited bandwidth of AZ1, which take the full 5 μs to settle. The switching transient is down, since the value of $V_G - V_H$ is lower, which is sampled in the state before.

At t = 105 μs phase ϕ_1 starts, which samples the offset of the buffer (V_{os1}) on $V_G - V_H$, which cause a switching on $V_G - V_H$. The settled values will be around V_{os1} x α_2, which corresponds to 4 mV x 30 = 120 mV. Since $V_G - V_H$ is directly connected to the output of the buffer, the transients on $V_G - V_H$ also causes a spike on the output at t = 105 μs. The effect of the spike on $V_E - V_F$ is barely visible on the output, since this node doesn't directly effect the output. The spike on $V_G - V_H$ is around 200 μV. This causes a spike at the output of around 200 $\mu V/\alpha_1$ = 6.7 μV, as explained in section 2.4. The residual settled offset is rather large with 458 μV. This is due to the limited gain of AZ2 and AZ3. The switching transients are 12 μV peak-to-peak.

4.1.2. Continuous-time auto-zeroed buffer with active integration

The continuous-time auto-zeroed buffer with active integration, shown in Figure 2.6, uses an active integrator to lower the switching transients.

Figure 4.2: The simulated switching transients and residual offset of the continuous time auto-zeroed buffer with active integration.

In Figure 4.2 the simulated switching transients of the continuous time auto-zeroed buffer with the shared active integrator (Figure 2.6) are shown. Due to the added gain of the active integrator, the residual offset is considerably lower: about 140 nV when settled. Also the switching transients are much smaller: 3 μV peak-to-peak.

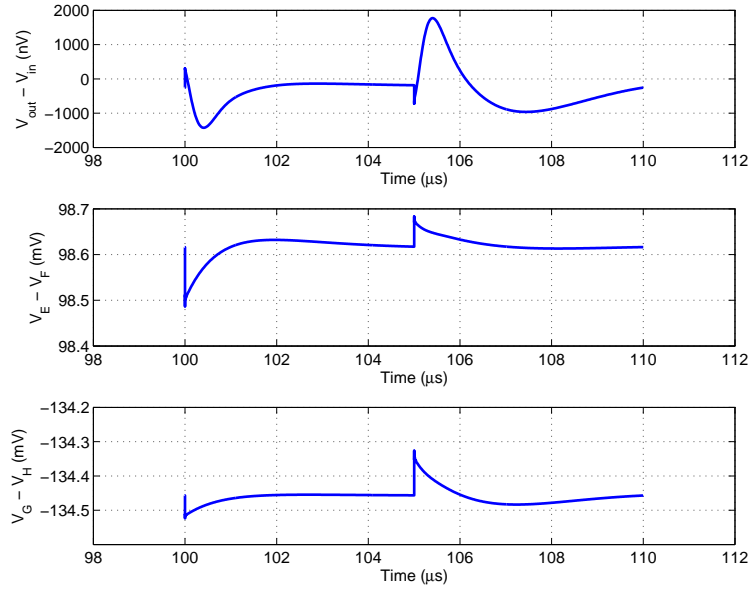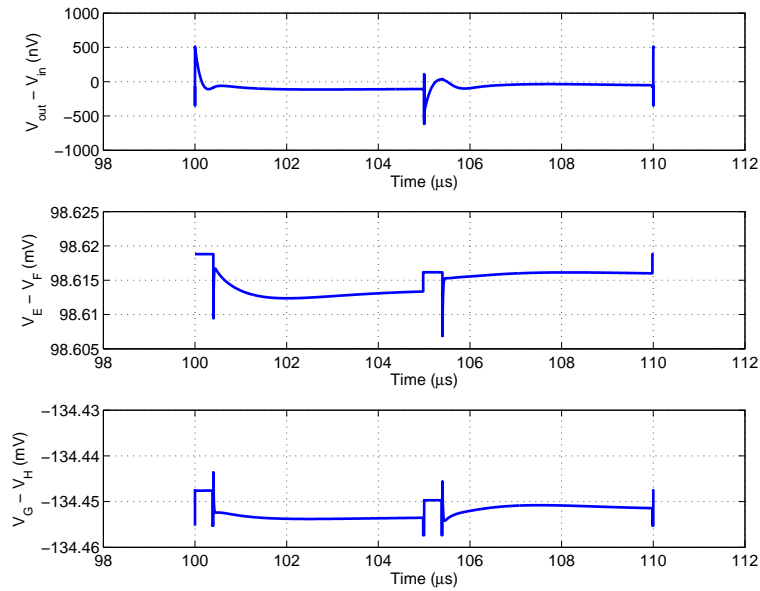### 4.1.3. CONTINUOUS-TIME AUTO-ZEROED BUFFER WITH ACTIVE INTEGRATION AND DEADTIME

Figure 4.3: The simulated switching transients and residual offset of the continuous time auto-zeroed buffer with active integration and deadtime.

In Figure 4.3 the simulated switching transients of the continuous time auto-zeroed buffer with the shared active integrator and dead time are shown. The residual settled offset between 100 and 105 μs is 115 nV, which is the ϕ_2 phase. Between 105 and 110 the residual offset is 35 nV, which corresponds to the ϕ_1 phase. The small difference in residual offset between the two phases is explained in section 2.4. The switching transients have been reduced to the required 1 μV peak-to-peak and a great reduction in the settling speed after the transients can be seen, when compared to Figure 4.2.
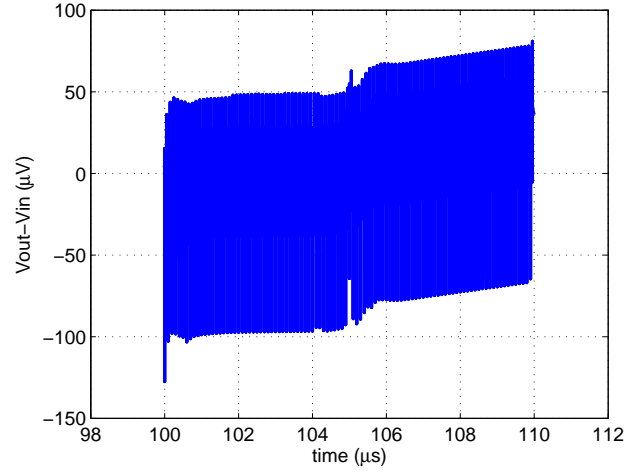
4.1.4. POST-LAYOUT

Figure 4.4: The simulated switching transients and residual offset of the extracted post-layout circuit.

After the tape-out some thorough simulations were done on the extracted post-layout circuit. The extracted post-layout circuit contains all the extracted parasitic capacitances. In Figure 4.4, the simulated switching transients and residual offset are shown. A large increase in the transients can be seen, due to the coupling of reference clock f_{ref}, required by the complementary clock generator. Even with some thorough on-chip shielding, the coupling remains strong. The switching transients have increased to 150 μV peak-to-peak. Due to this large coupling the buffer and auto-zero loop are continuously kicked, which cause a residual error on the output that is not steady. This can be problematic if the sampled value at the end of phase ϕ_1 is large and causes a large charge kickback towards the hold capacitor. At the end of phase ϕ_1 the error on the output is about 30 μV.
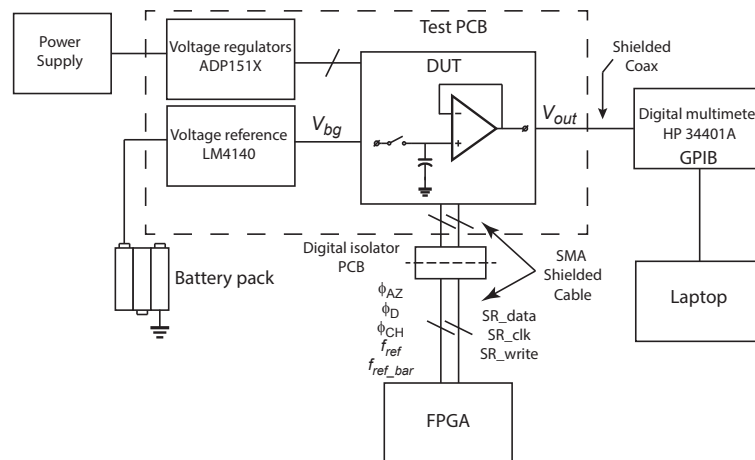
4.2. THE MEASUREMENT SETUP

Figure 4.5: A block diagram of the measurement setup.

The measurement setup is shown in Figure 4.5. To measure the manufactured chip of the sampled voltage reference, a Printed Circuit Board (PCB) has been designed. This PCB has several voltage regulators to generate the different supply voltages from a single power supply. It also contains a voltage reference which supplies the sampled voltage reference with the input reference voltage. The LM4140 is used as an external reference with a nominal output voltage of 1.024 V. The LM4140 is fed by a battery pack, which provides a very low noise supply. To load the shift register and periodically refresh it to recharge the hold capacitor, a Field Programmable Gate Array (FPGA) development board (Altera DE0-Nano) is used. The FPGA also generates the clock signals for auto-zeroing (ϕ_{AZ}), dead time (ϕ_D), chopping (ϕ_{CH}), the reference frequency (f_{ref}) and the complementary reference frequency ($fref_{bar}$). The signals from the FPGA are routed through digital isolators, to isolate the signals from the chip and cut any ground loops from the FPGA development board. All the clock signals are connected with shielded SMA cables, to prevent coupling. The output of the design under test (DUT) is measured by a HP 34401A digital multimeter. The connection is made with a shielded coax cable to make sure no outside coupling can occur. This digital multimeter has a GPIB interface, which is used to send the data to a laptop.

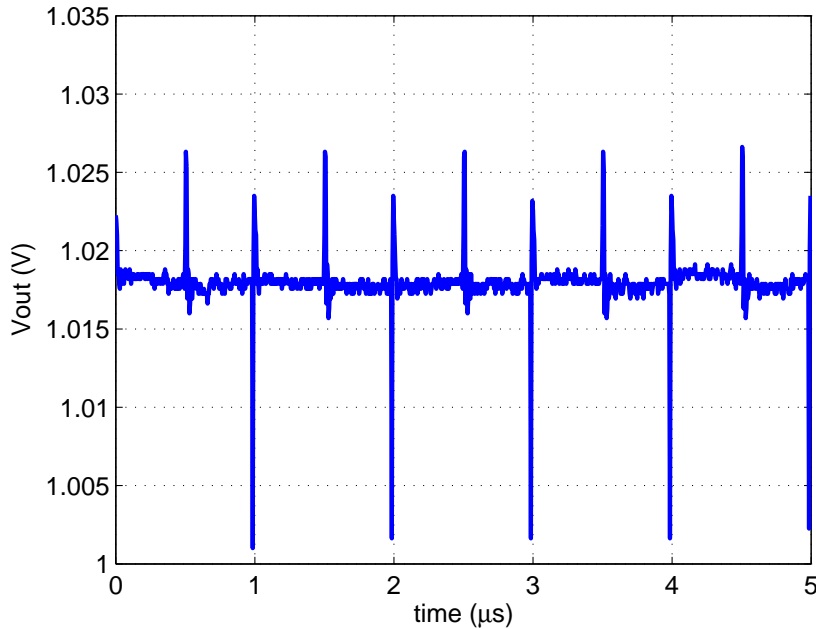## 4.3. The measured effect of the reference clock on the output

Figure 4.6: The measured effect of the reference frequency on the output of the buffer.

The first thing to note is that the reference clock (f_{ref}), used by the complementary clock generator described in section 3.5, has a rather high coupling to the output. To measure the effect on the output a measurement was done using a high-speed oscilloscope, the auto-zeroing clock is enabled, the reference clock frequency is set to 1 MHz and the external reference is permanently connected.

In Figure 4.6 the effect of the reference frequency on the output can be seen. Even though substantial effort has been put in shielding this clock line inside and outside the chip, the resulting spikes at the output have a value of 25 mV peak-to-peak. Due to the large spikes on the output it is rather tricky to do measurements on the chip, since the signal of interest should be measured in between the spikes. Unfortunately the effectiveness of the measures to lower the output switching transients due to the auto-zeroing, such as active integration, dead time and source degeneration, could not be evaluated properly due to these spikes.

Another important observation is that in between the spikes caused by the reference clock, occurring every 0.5 μs, the output of the buffer is not steady. The average value of the output is around 1.018 V, but in between the

spikes the output changes about 1 mV above and below this average. This can be explained by the fact that the kick given by the reference clock coupling, causes the buffer and auto-zero loop to change values and slowly settle back. Therefore the auto-zeroing loop never fully settles and never reaches the low residual offset that is reached in simulation. A higher residual offset will also be problematic for the leakage. Through charge kickback the increased residual offset will be transferred to the hold capacitor. Due to the large kicks of the reference clock, the error on the output when sampled at the end of phase ϕ_1 is in in the order of several 100 μV.

4.4. Measured hold time without auto-zeroing

As a first test to determine how long the voltage on the capacitor can be held, a measurement is done without auto-zeroing. This will illustrate how well the low leakage sample and hold circuit works, without the additional leakage caused by auto-zeroing. To do this the auto-zeroing loop is disconnected from the buffer and the auto-zeroing clock is shutoff. To measure the hold time of the circuit, the FPGA is programmed to periodically update the on-chip shift-register, which in turn periodically refreshes the hold capacitor. After the refresh of the hold capacitor, the sample switches are opened and the shift-register doesn't have to change over the rest of the hold period. Therefore during the hold period the signals controlling the shift-register are switched off, to make sure that no coupling occurs and no extra leakage is introduced. The reference clock is still continuously on since is it necessary to clock through the signals controlling the sample and hold switches. A reference clock frequency of 1 MHz is used for all the measurements. The external voltage reference is not turned off during the hold time due to the fact that the channel leakage is too high in this design, as explained in section 2.6.2.

Due to the fact the the parasitic diodes of the transistors are not properly modeled, a proper comparison between the measured result and simulation cannot be made.
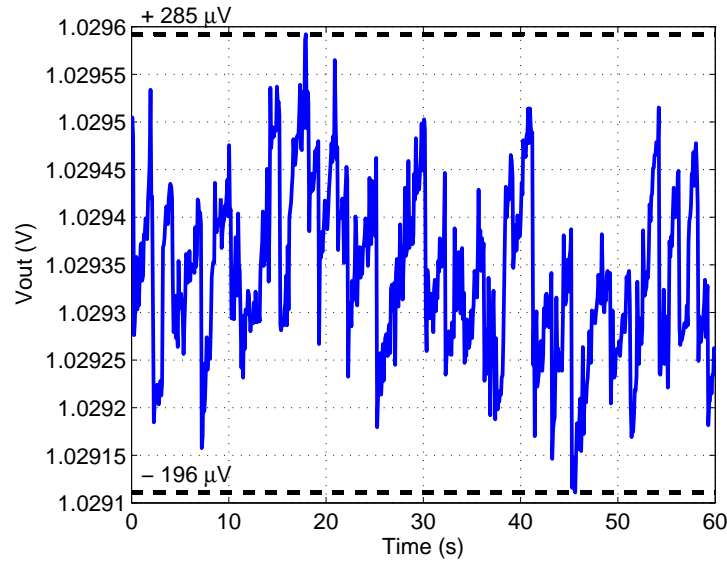

Figure 4.7: Leakage measurement without auto-zeroing, with a refresh period of 1s.

Figure 4.7 shows the result with a refresh period of 1 s. The average output voltage is around 1.0293 V, while the output of the external reference is around 1.024 V ($\pm$0.1%). This indicates that the offset of the buffer is around 5.3 mV for this sample. The noise is rather large with a value of about 481 μV peak-to-peak, or about 72.9 μV_{rms}. Since no auto-zeroing is applied in this case, there is much more low frequency noise than the specified 500 nV_{rms} (0.1 to 10 Hz). Also the sample and hold action will increase the noise below 1 Hz, due to the kT/C noise and the undersampled reference noise. Therefore in the output voltage some jumps can be seen with a 1 Hz frequency. No clear charging or leaking effect can be distinguished, which means that any charging or leaking effect is of the same order or is below the noise.
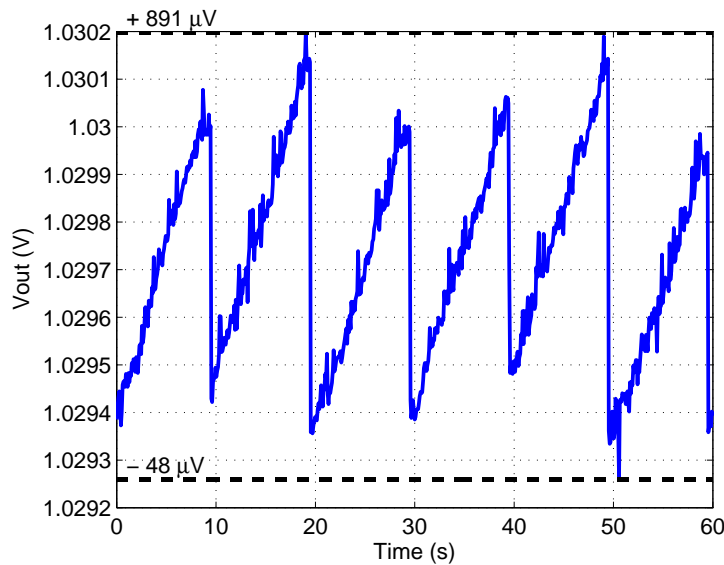
Figure 4.8: Leakage measurement without auto-zeroing, with a refresh period of 10s.

Next the refresh period is increased to 10 s, the result can be seen in figure 4.8. A clear charging effect can be seen. In the hold period from 30 to 40 s, an increase from 1.0294 V to about 1.03 V can be observed. This indicates that for a 10 s refresh frequency the voltage increase on the capacitor is about 600 μV. This would give an error rate of about 60 μV per second. The value of the reference voltage at the start of the hold period is not constant, due to the difference in sampled noise.

The increase in output voltage over a 10 s hold time, indicates that there is a net charging current somewhere. In section 2.6.1 it was explained that the body diode leakage might perform better for a body capacitor voltage level of Vdd/2. For the case of a body capacitor voltage of 1.024 (reference voltage), diode D_2 will be biased at around 0.8 V and diode D_5 will be biased at around 1V. This would indicate that the reverse saturation current of diode D_5 will drain the body capacitor faster than D_2 will charge (assuming equal diode parameters). However for dissimilar diode parameters the charging current of D_2 might very well be large. Further measurements indicate that this charging effect is indeed influenced by the body capacitor voltage level.
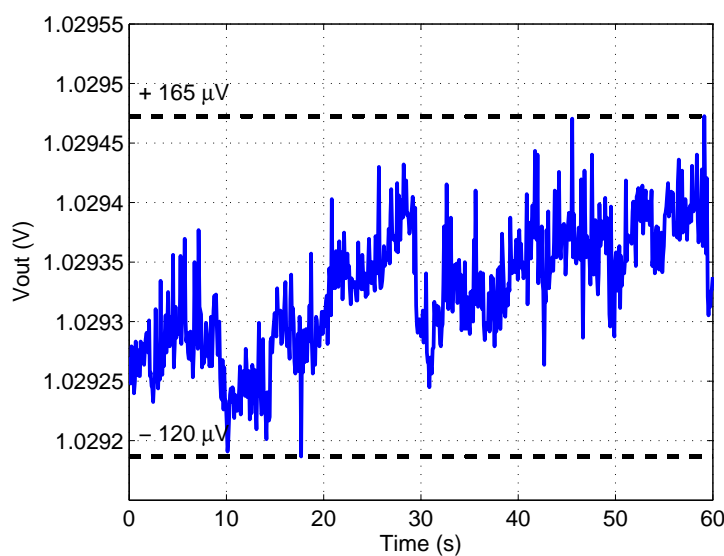

Figure 4.9: Leakage measurement without auto-zeroing, with a refresh period of 10s and the body capacitor voltage set at 0.9 V.

Figure 4.9 shows the leakage measurement, with a refresh period of 10s, when the body capacitor voltage level is set close to Vdd/2. This is done by using a seperate power supply which can be set with a precision of about 5 mV. No clear leakage is observed, when compared to the measurement in Figure 4.8, where the body capacitor voltage level is at the reference potential. Some jumps with a frequency of 0.1 Hz can be observed due to the difference in sampled noise.
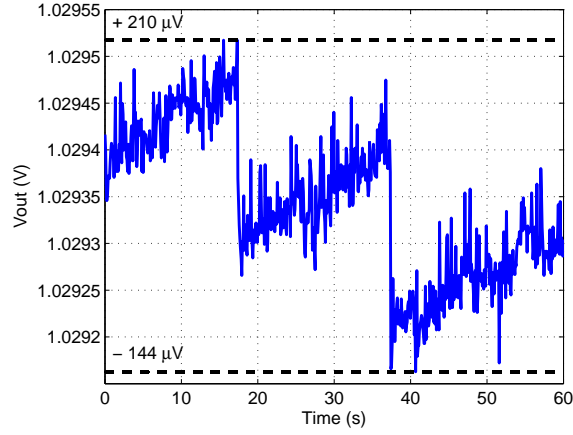

Figure 4.10: Leakage measurement without auto-zeroing, with a refresh period of 20s and the body capacitor voltage set at 0.9 V.

If the refresh period is increased to 20 s, the result can be seen in figure 4.10. A clear charging trend can be seen from 18 to 38 s. The increase in output voltage is about 100 μV for a 20s hold time. This would give an error rate of about 5 μV per second. In this case the body capacitor voltage is set with a precision of about 5 mV. If this precision could be improved, the leakage could be reduced further.

4.5. Hold time with continuous auto-zeroing

To see the effect that auto-zeroing has on the leakage of the hold capacitors, both simulations and measurement were done. The results of the simulation is compared with the measurement results. The measurements are shown for the pre-layout circuit which doesn't include the parasitic coupling capacitors, and the extracted post-layout result which does included all parasitic capacitors.
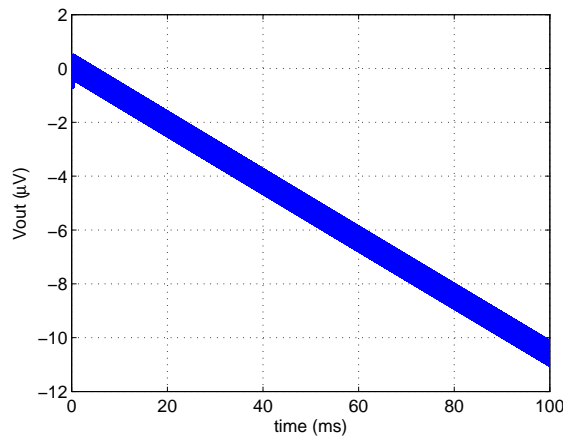
4.5.1. Pre-layout simulation

Figure 4.11: A simulation of the leakage for the pre-layout circuit.

In Figure 4.11 the results are shown for a simulation of the leakage pre-layout. First the auto-zero loop is allowed to settle for 0.1 ms, after which the external reference is disconnected. An auto-zeroing frequency of 100 kHz is used. After this the external reference is left disconnected for 100 ms and the resulting droop in this hold period is about 10.7 μV. If this leakage is trend is extrapolated to 1s, the resulting droop would be 107 μV. This gives an error rate of about 107 μV per second.

The residual offset at the end of phase ϕ_1 is around 60 nV. The voltage droop due to charge kickback will be $\frac{\Delta V}{\Delta T} = \frac{C_{11} V_{os} f_{AZ}}{2C_H} = 100 \ \mu$V/s. This matches the simulated result very well.
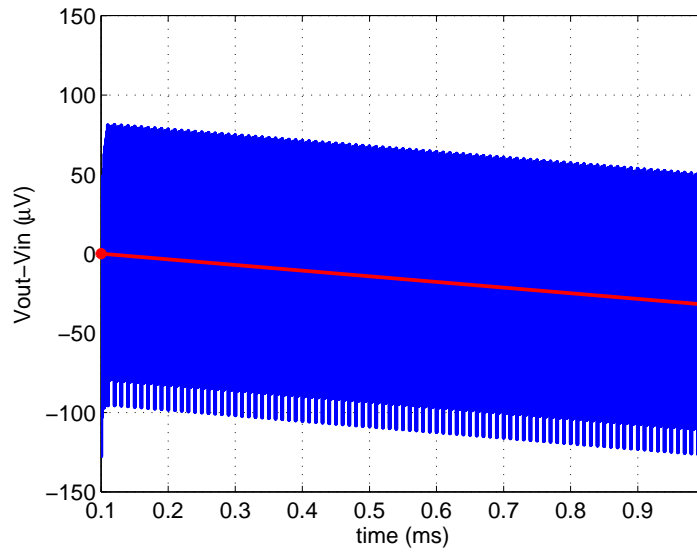
4.5.2. Post-layout simulation

Figure 4.12: A simulation of the leakage for the post-layout circuit.

Figure 4.12 shows a simulation of the leakage post-layout. First the auto-zero loop is allowed to settle for 0.1 ms, after which the external reference is disconnected. An auto-zeroing frequency of 100 kHz is used. Due to the long run time required for a post-layout simulation, the result is only shown till 1 ms. The same increased spikes as in Figure 4.4 can be seen. The average between the peaks at the start of the hold period (at 0.1 ms), is around 0 V. After 1 ms the voltage on the hold capacitor has dropped, to about 32 μV. If the same trend would continue till 1 s, a voltage droop of about 32 mV would remain.

The leakage observed in the post-layout simulation, is much larger than the result of the pre-layout simulation. This can be explained by the increased residual offset, caused by the coupling of the reference clock (f_{ref}). At the end of phase ϕ_1, this increased residual offset will be sampled and transferred towards the hold capacitor, via charge kickback. The residual offset at the end of phase ϕ_1 of the pre-layout simulation is about 60 nV and for the post-layout it increases to about 30 μV. Since the residual offset is about 500 times larger in the post-layout simulation, about the same increase in voltage droop can be observed between pre-layout and post-layout simulation. The calculated voltage droop is $\frac{\Delta V}{\Delta T} = \frac{C_{11} V_{os} f_{AZ}}{2C_H} = 50$ mV/s, which matches the simulated result of 32 mV/s very well.

4.5.3. Measured

Next the leakage of the hold capacitor is measured for the final implemented system. For this the auto-zeroing loop is connected and the auto-zeroing clock is turned on, to see the effect of the auto-zeroing on the discharging of the hold capacitor. For this measurement an auto-zeroing frequency (f_{az}) of 2 kHz is used, together with a reference frequency of f_{ref} 1 MHz. The $fref_{bar}$ signal is disabled. The hold capacitor is refreshed every second.
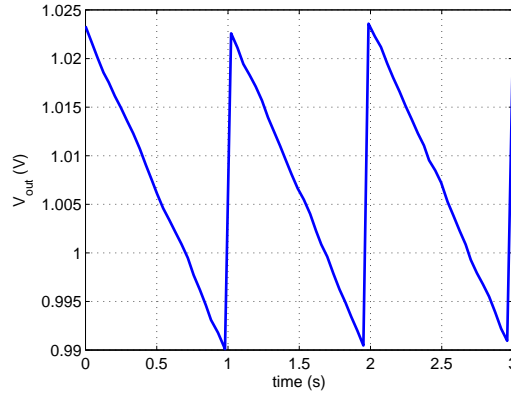
Figure 4.13: Leakage measurement with auto-zeroing with a frequency of 2 kHz and a refresh period of 1 s.

Figure 4.13 shows the measured leakage with auto-zeroing enabled, a large increase in the droop of the hold capacitor voltage can be seen. Every second the refresh of the hold capacitor can be seen, where the output goes to the 1.024 V of the external voltage reference. After 1 s, the voltage on the output drops to about 0.99 V. For an auto-zeroing frequency of 2 kHz, a droop in the output voltage of about 34 mV can be observed.

To improve the leakage, the slow-chopping technique explained in section 2.8 was tested. With auto-zeroing enabled no clear leakage reversal could be observed. This indicates that the major leakage component is not due to the charge-kickback of residual offset or that the residual offset cannot be properly switched in polarity by chopping. If for example the coupling of clock signals into the auto-zeroing clock is the major cause of residual offset, then chopping will not reverse the polarity.

4.6. The effect of clock signals on the measured leakage

Throughout the measurements it became clear that the leakage of the hold capacitor is very sensitive to different parameters of the clock signals. When the rise and fall time of the reference clock were changed or the reference clock was slightly delayed compared to other clocks the leakage changed a lot. Below the results are shown, when the complementary reference clock is enabled and when capacitive loading is applied to the reference clock.

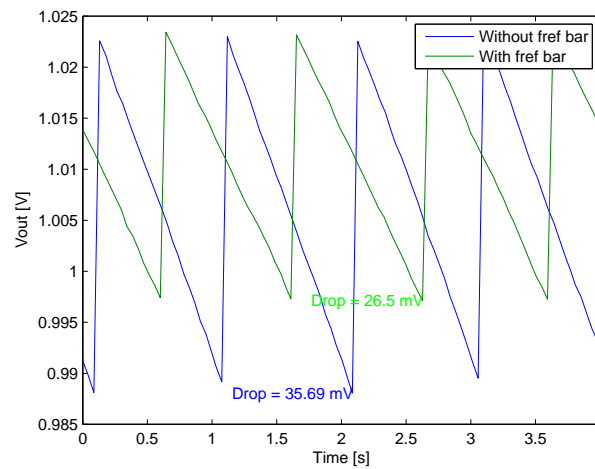### 4.6.1. The effect of the complementary reference clock

Figure 4.14: Leakage measurement with auto-zeroing with a frequency of 2 kHz, with a refresh period of 1 s and with $fref_{bar}$.

When the same leakage measurement is performed as Figure 4.13 with auto-zeroing and a refresh period of 1 s, but with complementary reference clock $fref_{bar}$ enabled, Figure 4.14 appears. The complementary reference clock $fref_{bar}$ was routed throughout the chip next to the reference clock, to cancel coupling effects. From figure 4.14 it is clear that with $fref_{bar}$ enabled, the droop on the output decreases to about 26 mV. This seems to indicate that the coupling of f_{ref} is causing at least part of the leakage and trying to cancel this coupling by enabling $fref_{bar}$ is helping to decrease the leakage.

The reduced coupling effect of the reference clock, by enabling $fref_{bar}$, will allow the auto-zeroed buffer to settle at a lower residual offset value. When at the end of phase ϕ_1, this lower residual offset is sampled and transferred towards the hold capacitor via charge kickback, this will allow for a lower leakage.

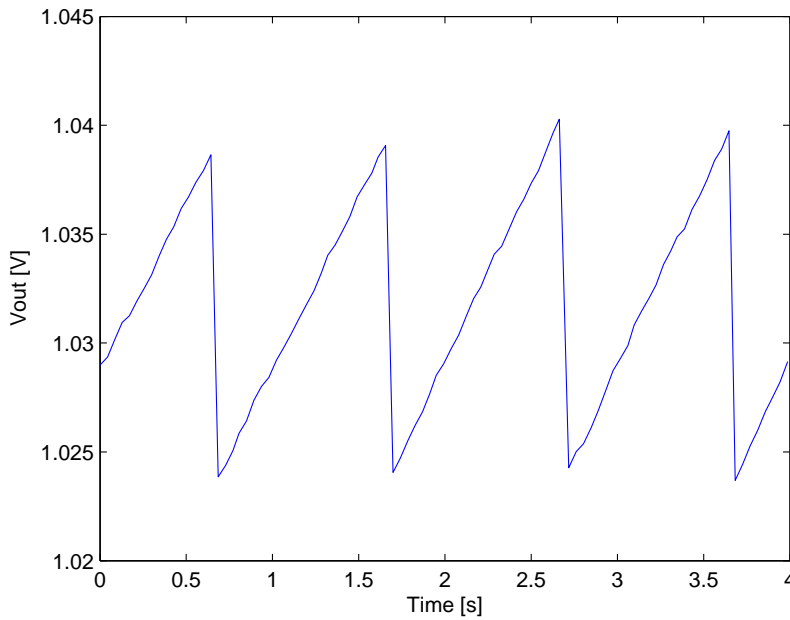## 4.6.2. The effect of loading the reference clock

Figure 4.15: Leakage measurement with auto-zeroing with a frequency of 2 kHz, with a refresh period of 1 s and with capacitive loading of $fref_{bar}$.

When the same leakage measurement is performed as Figure 4.14 with auto-zeroing, a refresh period of 1 s, complementary reference clock $fref_{bar}$ enabled and with some added capacitive loading, Figure 4.15 appears. In this measurement rather than a leaking effect after the refresh of the hold capacitor at 1 s, a charging effect is observed. The voltage on the hold capacitor increases about 12 mV over the refresh period.

When the f_{ref} signal is capacitively loaded, the rise and fall time will decrease. This will both lower the coupling and change the moment when the auto-zeroing clock switches. Since the auto-zeroing clock is made with the complementary clock generator, the rise and fall time of the reference clock will also change the rise and fall time of the auto-zeroing clock. A slower fall time will mean that the sampling moment of the residual offset will change and therefore the amount of leakage via charge kickback is changed.
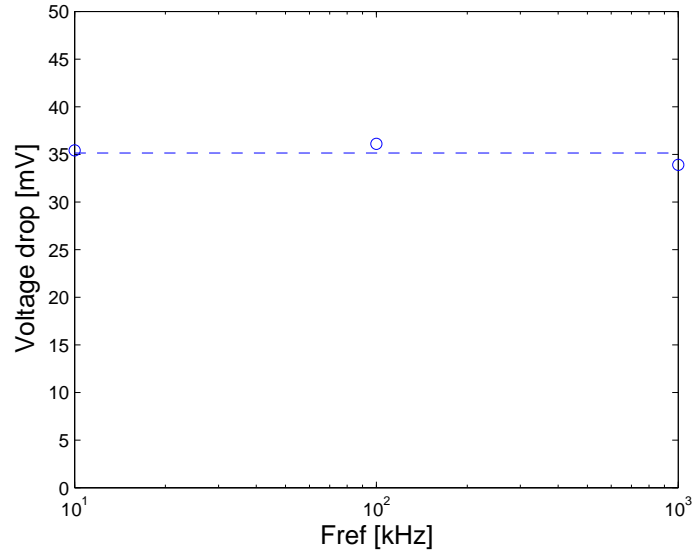
Figure 4.16: Leakage measurement for different reference frequencies.

To see the effect that reference clock frequency has on the voltage droop on the hold capacitor a measurement is done were the reference frequency is changed from 10 to 100 kHz and finally 1 MHz. For the measurement the refresh period is 1 second and the auto-zeroing frequency is held constant at 2 kHz. The result is shown in Figure 4.16. No clear trend between the voltage droop of the hold capacitor and the reference clock frequency can be seen.

Even for the lowest reference frequency of 10 kHz a voltage droop of around 35 mV appears. Even though the auto-zeroed buffer should have more time to settle at this lower reference frequency, a similar voltage droop appears. This is due to the fact that the buffer and auto-zeroing loop always get a kick right before the transition of the auto-zeroing clock, which always causes a similar charge kickback.
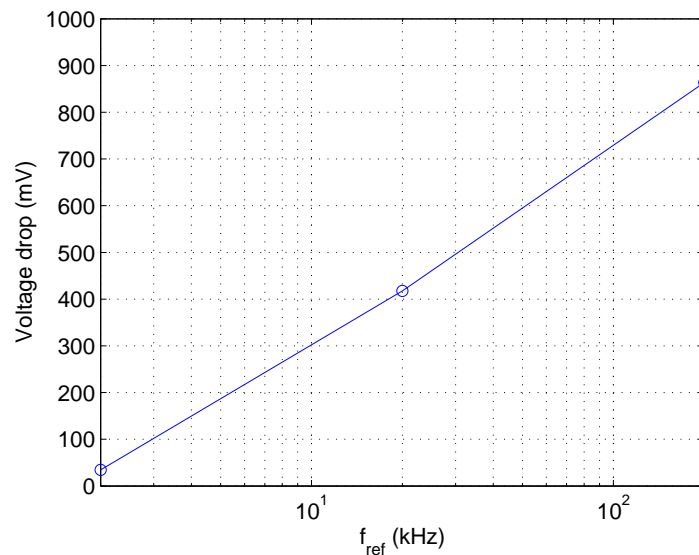

Figure 4.17: Leakage measurement for different auto-zeroing frequencies.

A similar measurement is done where the auto-zeroing frequency is changed and the reference clock frequency is

held constant, to see the effect that the auto-zeroing clock has on the voltage droop on the hold capacitor. For this measurement, the reference frequency is held constant at 1 MHz and the auto-zeroing frequency is changed from 2 to 20 and finally 200 kHz. The result is shown in Figure 4.17. Here a clear trend between auto-zeroing frequency and the voltage droop of the hold capacitor can be seen.

This can be explained by the fact that for a higher auto-zeroing frequency, the charge kickback happens more frequently. This will result in a faster leakage of the hold capacitor.

4.7. Future work

From the measurements, it is clear that this design has a much higher leakage due to the charge kickback caused by auto-zeroing, than is expected from the simulation results. With a pre-layout offset of about 60 nV, the leakage in 1 s with an auto-zeroing frequency of 100 kHz is 107 μV. With a post-layout offset that is about 500 times larger, with 30 μV, the leakage in 1 s with an auto-zeroing frequency of 100 kHz also increases by about the same order to 32 mV. In the final design, the leakage increases to about 34 mV in 1 s with an auto-zeroing frequency 2 kHz. Equation 2.16, can explain this increase in voltage droop when the sampled residual offset increases 100 μV. Figure 4.6 shows that such an increase could be due to the coupling of clock signals.

This increase in residual offset is caused by the coupling of the reference clock which kicks the buffer and auto-zeroing loop. A future design should therefore focus on bringing the residual offset back to the level that was originally designed for in the pre-layout simulation. Since this design already has some rigorous shielding to prevent the coupling of the reference clock, a future design should remove the need for a reference clock to completely remove any coupling. Besides the charge kickback caused by residual offset, some of the charge kickback is caused by charge injection mismatch, as described in section 2.7. The charge injection mismatch with transmission gates depends on clock parameters such as: clock skew and rise/fall time. Therefore, part of the increased leakage might also be explained by this.

Several things can be done to lower the leakage due to charge kickback:

- Lowering the residual offset sampled on C_{11}. This definitely requires a different clocking scheme. Furthermore the gain of the integrator could be increased further and the ratio of buffer and AZ3 gain could be increased.

- Lowering the auto-zeroing frequency. This lowers the frequency at which charge kickback occurs.

- Increasing the size of the hold capacitor. This lower the influence of injected charge on the hold capacitor.

- Decreasing the size of capacitors C_{11} and C_{12}. This lowers the amount of charge that is transferred. Care must be taken that this doesn't increase the leakage due to charge injection mismatch.

If capacitors C_{11} and C_{12} are decreased with a factor ten to 0.1 pF, the residual offset is reduced to about 50 nV, the auto-zeroing is reduced to 10 kHz and the hold capacitor is increased to 50 pF, then according to Equation 2.16 a voltage droop of 0.5 μV/s could be achieved.

4.7.1. Changing the clocking scheme

To remove the reference clock from the design a different clocking scheme has to be used. The current complementary clock generator circuit relies on a high frequency clock, to allow for a synchronous switching action. The complementary clock generators are placed close to the switches, to keep the path of the two clock signals identical. This requires the high frequency clock to be routed throughout the chip close to sensitive signal, increasing the change of coupling. A good alternative for the current complementary clock generator is to use an SR-latch. This can achieve simultaneous transitioning clock signals without using an external reference clock.
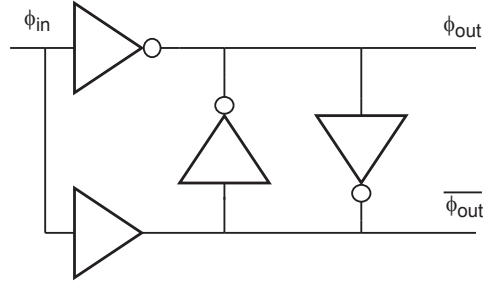
Figure 4.18: The implementation of the SR-latch

The SR-latch is implemented with the circuit shown in Figure 4.18. A single clock signal comes in, the bottom buffer creates the normal output signal and the top inverter creates the complementary signal. The cross coupled inverters perform the latching. The bottom buffer and top inverter are sized to have equal rise and fall times.
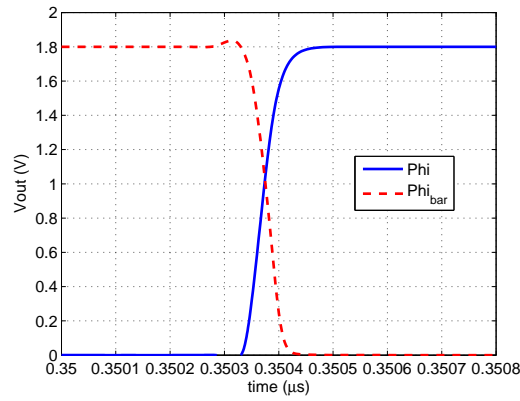
Figure 4.19: Phi and Phi_{bar} generated by the complementary clock generator.
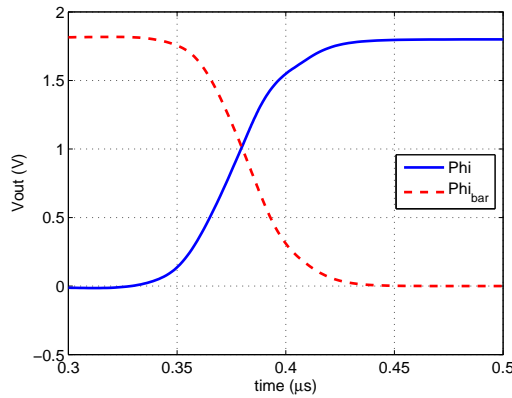
Figure 4.20: Phi and Phi_{bar} generated by the SR-latch.

Figure 4.19 shows the simulation result of the simultaneous transitioning signals obtained by the complementary clock generator used in this design and Figure 4.20 shows the simulation result obtained by using the SR-latch. They achieve comparable results in synchronicity. The rise and fall times of the SR-latch are considerably longer and this could be changed by increasing the driving strength of the inverters. Finally more careful simulation of the SR-latch over corners has to be done.

The reference clock is the highest frequency clock in this design and therefore also causes the most problems when coupled into the output. Some other clock signals still remain that can couple, these include: the auto-zero clock (ϕ_{az}), the dead-time clock (ϕ_d), the chopper clock (ϕ_{ch}) and the clocks controlling sample and hold switches SW_1, SW_2 and SW_3. Since the first three are all brought in from outside, they still have a high change of coupling, through for example the bondwires.

The auto-zeroing clock and dead time clock both have frequencies in the kHz-range and therefore they can still cause problems when coupling. It would be a good idea to create these high frequency signals as much as possible on-chip, to isolate the problem of coupling to a localized area of the chip.
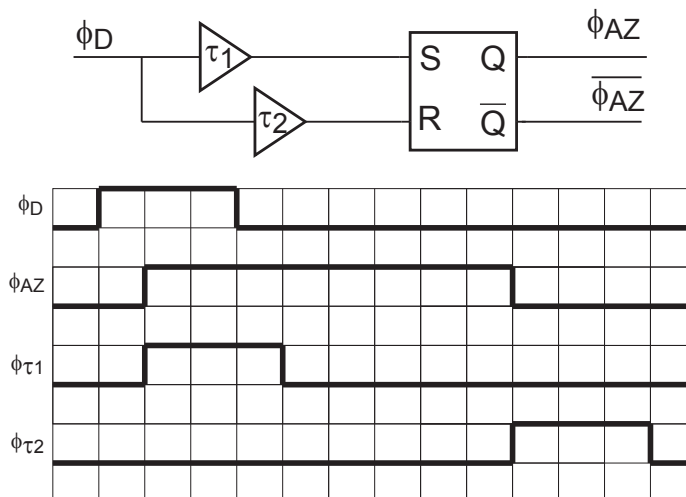

Figure 4.21: Deriving the auto-zero clock from the dead-time clock.

An option is to derive the auto-zeroing clock from the dead-time clock and only bring the dead-time clock from outside the chip. This circuit should not use clocked elements. The auto-zeroing clock always switches a fixed delay after the up-going flank of the dead-time clock and a fixed delay after the down-going flank. If two delays are made using RC-times together with a comparator which are both fed to a SR-latch, the auto-zero clock can be derived from the dead-time clock. The resulting circuit is shown in Figure 4.21.

An other option is to make the auto-zeroing clock on-chip with an oscillator and derive the dead-time clock from this. The big advantage of this solution is that all the high frequency signals are localized in one area of the chip, making it easier to shield it properly. Furthermore no signals are coming from outside the chip, which remove the change of signals coupling via the bondwires. The downside is that these digital block should be properly simulated, to ensure that the wholes chip doesn't fail due to a failing clock generation circuit.

The hold switches SW_1, SW_2 and SW_3 are controlled via the shift-register. The coupling of the clock and data signals controlling the shift register are minimal, since they are only updated during the refresh and switched off during the hold period. The coupling of the clock signals of SW_1, SW_2 and SW_3 themselves are not a problem since they only switch during refresh. The chopper clock is low frequency and should therefore also not be a problem.

4.7.2. LOWER THE REQUIRED AUTO-ZEROING FREQUENCY

Besides removing the reference clock, several other improvements can be made to reduce the leakage. Since the frequency of the auto-zeroing clock determines the frequency with which the charge-kickback occurs and therefore the speed of the leakage, lowering the auto-zeroing frequency will help to reduced the leakage. This design required an auto-zeroing frequency of 100 kHz to get rid off the $1/f$ noise, as explained in section 2.5. If the $1/f$ corner frequency of the buffer amplifier and OTA AZ1 is reduced, it will allow the auto-zeroing frequency to be reduced as well. Using more area for the input-transistors of the buffer and AZ1, will reduce reduce the $1/f$ noise corner.
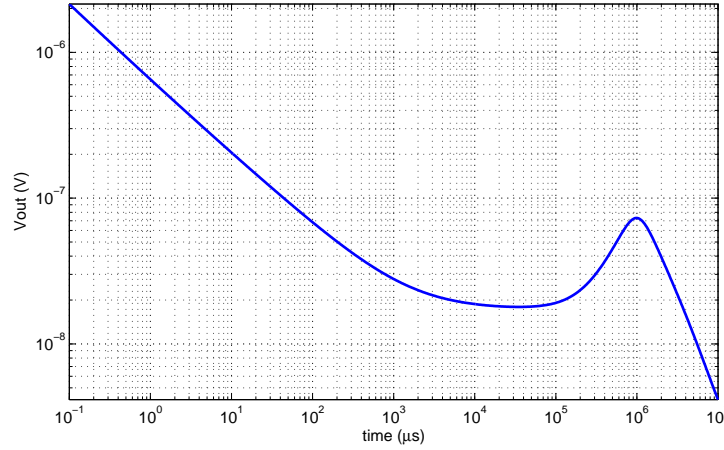
Figure 4.22: The output noise spectrum of the improved buffer without auto-zeroing.

The dominant low-frequency noise producers are the input transistors and the bottom current mirror transistors. If the size of the input transistors M_{P102} and M_{P103} of the buffer in Figure 3.2 is increased by a factor 15 and the bottom current mirror transistors M_{N102} and M_{N103} as well, the low frequency noise reduces. The resulting noise plot is shown in Figure 4.22. The $1/f$ noise corner has reduced to about 10 kHz. If the corresponding transistors of OTA AZ1 (M_{P202}, M_{P203}, M_{N202} and M_{N203} in Figure 3.3) are scaled up by a factor 15 as well, the auto-zeroing frequency to be reduced ten times. This will also reduce the leakage of the hold capacitor by a factor ten.

4.7.3. Using trimming to compensate the leakage current
When the above options fail or the leakage due to charge injection mismatch turns out to be to large, trimming can be used to lower the leakage. In [20] the hold circuit is replicated with a small hold capacitor and the leakage of this small hold capacitor is used to compensate a larger hold capacitor. The leakage is reduced by an order of magnitude over a limited temperature range: -20 to 40 C.

In [43] a special input current trimming technique is used, which aims to trim the mismatch of clock injection. They use a very high chopping (1.2 MHz) and auto-zeroing frequency, while reducing the input current from the usual 1.5 nA to 150 pA. Using this same technique to get a low input current with chopping and auto-zeroing frequencies in the tens of kHz range, might be very promising.

4.7.4. Other improvements
Beside the urgent improvement of the leakage due to auto-zeroing, several other less urgent improvements can be made:

- Lowering the power consumption of the AZ buffer. Since the buffer is never switched off, the dominant contribution to the power consumption is the power required by the buffer. The reference before the sample and hold can be duty-cycled to reduce the average power consumption, but the buffer has to stay on. The current consumption of this design is 161 μA, from a 1.8 V supply. This design has not been optimized for low power, using more low power techniques could significantly reduce the current consumption.

- Investigating the use of a class AB output stage for the buffer. In this design the use of a class AB output stage for the buffer has been described and briefly been investigated. In the final design it was not used due to the increased output switching transients and the large spikes generated during the slow chopping with the AB output stage. For a increased current driving strength and power efficiency a class AB output stage is highly desirable. More investigation is necessary to reduce the switching transients and spikes with a class AB output stage.

- The noise could be further reduced by using a combination of chopping and auto-zeroing. In [37] a combination of chopping and auto-zeroing is used, to eliminate the increased low frequency noise caused by

auto-zeroing. Care must be taken to the input current this combination gives, but the fact that the chopping can be done at a lower frequency then the auto-zeroing is promising.

- Implementing the channel leakage reduction scheme, introduced in section 2.6.2. In this scheme, the channel leakage is reduced by applying a buffered version of the hold capacitor voltage across the switch. The offset of this buffer will determine the residual channel leakage.

- Using two capacitors to average the noise. In section 2.5 two noise reduction mechanisms were explained. A third noise reduction mechanism is using two capacitors average the noise.
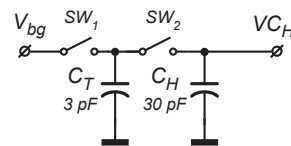

Figure 4.23: The switched capacitor network used for averaging the noise.

Due to the noise of the external source, the hold capacitor will be charged to a slightly different value each time. To reduce the noise in the bandwidth below the refresh frequency, averaging can be used between two capacitors. Before the hold capacitor a ten times smaller temporary capacitor is added (C_T), leading to the configuration shown in Figure 4.23. Sampling the voltage reference on this C_T and switching back and forth between C_T and C_H ten times, can in theory lower the noise by a factor ten. Some promising simulations have been done using this techniques with a switched capacitor bandgap, but for a successful implementation the effect of charge injection has to be further evaluated.

- Finally, this sample and hold circuit with buffer could be combined with a voltage-reference in a single chip. This will help to remove any leakage effect caused by the bondpad. Furthermore, the idea of this sample voltage reference is to be combined with an on-chip bandgap reference. Simulations have been done on a bandgap core, with a switched capacitor integrator to combine the V_{BE} with fifteen times a ΔV_{BE}, to get a temperature independent output. The noise folding will then be determined by the integrator, since the sampling starts there. The noise performance of such a switched capacitor bandgap, combined with a sample and hold and auto-zeroed buffer could be very low (below 2 μV peak-to-peak).

4.7.5. SWITCHING TO AN OTHER APPROACH

This design aims to achieve a low-noise, low-offset and low-power buffer by using auto-zeroing. The success of a future design with this approach depends on how far the leakage due to auto-zeroing can be reduced, by the options mentioned above. A design with auto-zeroing just during the charge time or trimming might also be used to get low-offset. This will mean that current and area will have to be spent in the buffer to get low noise and this will still lead to a higher $1/f$ noise. Such a design will have to lower the requirements on power and noise.

To test the option of auto-zeroing just during the charge time, a measurement has been done on the current design. Figure 4.24, 4.25 and 4.26 show the results for a hold time of 1s, 10s and 10s with Vcb at VDD/2 respectively. This is done for the same sample as the leakage measurements in section 4.4. The offset of 5.3 mV is removed by the auto-zeroing and similar leakage performance can be seen. Even though the auto-zeroing circuit is not really made for long holding, the offset doesn't change much with time.
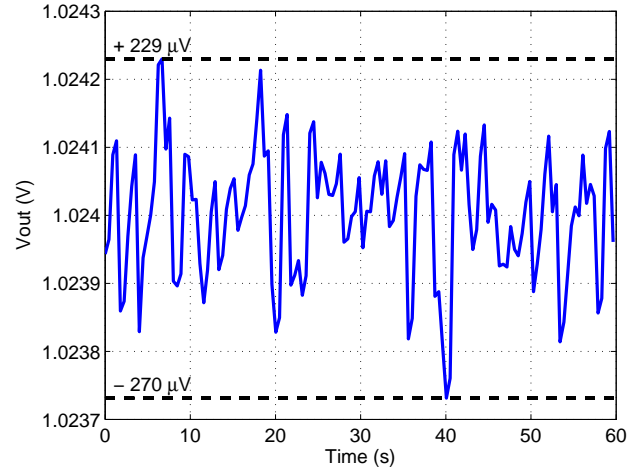
Figure 4.24: Leakage measurement for auto-zeroing just during the charge time, with a 1s hold time.
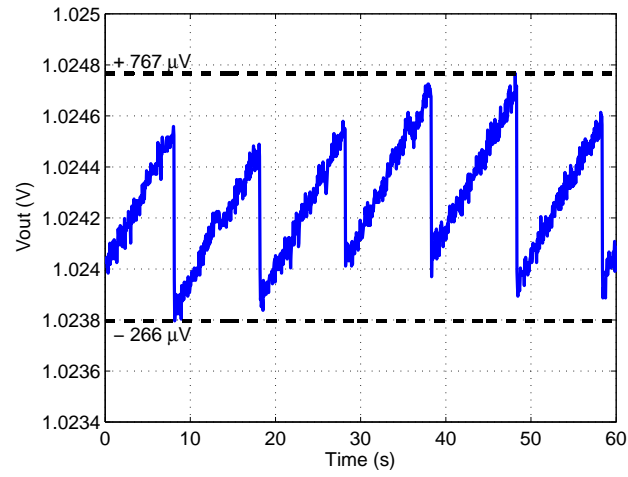

Figure 4.25: Leakage measurement for auto-zeroing just during the charge time, with a 10s hold time.
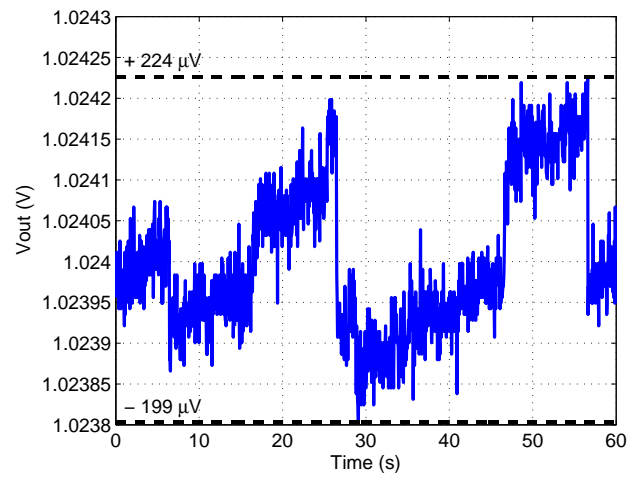

Figure 4.26: Leakage measurement for auto-zeroing just during the charge time, with a 10s hold time and VCb = Vdd/2.

5

CONCLUSIONS

In this thesis the design, implementation and measurement of a sampled voltage reference is presented. It aims to achieve both low-power and low-noise by storing the voltage generated by a reference on a capacitor for a long time. This then allows the reference to be switched off during the hold period, which leads to a lower average power. At the same time all the noise is limited to a bandwidth determined by the refresh frequency.

This design uses standard CMOS technology to create a low leakage sample-and-hold circuit and a continuous-time auto-zeroed buffer. The continuous-time auto-zeroed buffer allows for low-noise and low-offset, combined with low-power. For the buffer, special care has been taken to reduce the transients created by the auto-zeroing, by using a shared active integrator, a dead time clock and source degeneration. This allows the transients due to the auto-zeroing to be as low as 1 μV peak-to-peak in simulation, although the effectiveness of the techniques could not be evaluated in measurements.

The low leakage sample and hold circuit tries to minimize the body diode and channel leakage by nulling the voltage across the junctions, of the critical diodes. It is shown that the process of auto-zeroing transfers residual offset on the output to the hold capacitor, via a mechanism referred to as charge kickback. A special slow-chopping technique is presented and implemented to try to reverse the leakage due to the residual offset at the output.

The low leakage sample and hold circuit is shown to be effective, with no clear leakage for a 1 second hold time and 600 μV of voltage drop for 10 seconds of hold time. When the body capacitor voltage is lowered a 10 second hold time with no clear leakage can be achieved. For a 20 second hold time the increase is about 100 μV, this gives a resulting in a drift of about 5 μV per second. The final implemented design suffers from a large degree of coupling between the high frequency reference clock and the output. It is shown that this cause a higher residual error than originally designed for, which in turn increase the leakage via charge kickback. With a measured residual error in the order of several 100 μVs, the leakage in 1 s with an auto-zeroing frequency 2 kHz is 34 mV. Via simulation it is shown that with a lower residual offset, the leakage can be greatly reduced.

Changes are proposed to improve the current design. A change in the clocking scheme is proposed and a circuit is shown with simulations to replace the high frequency reference clock. Improvements to the buffer and AZ1 are shown, to lower the $1/f$ noise corner. This would allow for a lower auto-zeroing frequency, which would reduce the leakage.

Bibliography

[1] J. F. Witte, K. A. Makinwa, and J. H. Huijsing, *Dynamic Offset Compensated CMOS Amplfiers* (Springer, 2009).

[2] *Voltaic cell,* (1893), uS Patent 494,827.

[3] D. H. Smith, *The suitability of the silicon alloy junction diode as a reference standard in regulated metallic rectifier circuits,* American Institute of Electrical Engineers, Part I: Communication and Electronics, Transactions of the **73**, 645 (1955).

[4] G. Bosch and J. Koster, *Integrated thermostat voltage stabilizer,* in *Solid-State Circuits Conference. Digest of Technical Papers. 1971 IEEE International*, Vol. XIV (1971) pp. 144–145.

[5] J. Lieux and K. Lance, *An ultra-stable voltage reference using a new type of bulk zener diode,* in *Solid-State Circuits Conference. Digest of Technical Papers. 1972 IEEE International*, Vol. XV (1972) pp. 148–149.

[6] L. Technology, *Ltz1000,* ().

[7] D. Hilbiber, *A new semiconductor voltage standard,* in *Solid-State Circuits Conference. Digest of Technical Papers. 1964 IEEE International*, Vol. VII (1964) pp. 32–33.

[8] R. Widlar, *Some circuit design techniques for linear integrated circuits,* Circuit Theory, IEEE Transactions on **12**, 586 (1965).

[9] A. Brokaw, *A simple three-terminal ic bandgap reference,* Solid-State Circuits, IEEE Journal of **9**, 388 (1974).

[10] K. Kuijk, *A precision reference voltage source,* Solid-State Circuits, IEEE Journal of **8**, 222 (1973).

[11] C. Palmer and R. Dobkin, *A curvature corrected micropower voltage reference,* in *Solid-State Circuits Conference. Digest of Technical Papers. 1981 IEEE International*, Vol. XXIV (1981) pp. 58–59.

[12] G. C. Meijer, P. C. Schmale, and K. van Zalinge, *A new curvature-corrected bandgap reference,* in *Solid State Circuits Conference, 1981. ESSCIRC '81. 7th European* (1981) pp. 208–210.

[13] D. Spady and V. Ivanov, *A cmos bandgap voltage reference with absolute value and temperature drift trims,* in *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on* (2005) pp. 3853–3856 Vol. 4.

[14] G. Ge, C. Zhang, G. Hoogzaad, and K. Makinwa, *A single-trim cmos bandgap reference with a 3☒inaccuracy of ±0.15% from - 40 c to 125 c,* Solid-State Circuits, IEEE Journal of **46**, 2693 (2011).

[15] H. Banba, H. Shiga, A. Umezawa, T. Miyaba, T. Tanzawa, S. Atsumi, and K. Sakui, *A cmos bandgap reference circuit with sub-1-v operation,* Solid-State Circuits, IEEE Journal of **34**, 670 (1999).

[16] A.-J. Annema and G. Goksun, *A 0.0025mm2 bandgap voltage reference for 1.1v supply in standard 0.16 x03bc;m cmos,* in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2012 IEEE International* (2012) pp. 364–366.

[17] J. M. Lee, Y. Ji, S. Choi, Y.-C. Cho, S.-J. Jang, J. S. Choi, B. Kim, H.-J. Park, and J.-Y. Sim, *5.7 a 29nw bandgap reference circuit,* in *Solid- State Circuits Conference - (ISSCC), 2015 IEEE International* (2015) pp. 1–3.

[18] A. Shrivastava, K. Craig, N. Roberts, D. Wentzloff, and B. Calhoun, *5.4 a 32nw bandgap reference voltage operational from 0.5v supply for ultra-low power systems,* in *Solid- State Circuits Conference - (ISSCC), 2015 IEEE International* (2015) pp. 1–3.

[19] V. Ivanov, R. Brederlow, and J. Gerber, *An ultra low power bandgap operational at supply from 0.75 v,* Solid-State Circuits, IEEE Journal of **47**, 1515 (2012).

[20] Y.-P. Chen, M. Fojtik, D. Blaauw, and D. Sylvester, *A 2.98nw bandgap voltage reference using a self-tuning low leakage sample and hold,* in *VLSI Circuits (VLSIC), 2012 Symposium on* (2012) pp. 200–201.

[21] M. Liu, P. Harpe, R. van Dommele, and A. van Roermund, *15.4 a 0.8v 10b 80ks/s sar adc with duty-cycled reference generation,* in *Solid- State Circuits Conference - (ISSCC), 2015 IEEE International* (2015) pp. 1–3.

[22] L. Technology, *Ltc6655,* ().

[23] M. Integrated, *Max6126,* .

[24] T. Instruments, *Lm4140,* .

[25] L. Technology, *Lt6656,* ().

[26] Intersil, *X60008,* .

[27] B. Ahuja, H. Vu, C. Laber, and W. Owen, *A very high precision 500-na cmos floating-gate analog voltage reference,* Solid-State Circuits, IEEE Journal of **40**, 2364 (2005).

[28] C. Enz and G. Temes, *Circuit techniques for reducing the effects of op-amp imperfections: autozeroing, correlated double sampling, and chopper stabilization,* Proceedings of the IEEE **84**, 1584 (1996).

[29] M. Bolatkale, M. Pertijs, W. Kindt, J. Huijsing, and K. Makinwa, *A single-temperature trimming technique for mos-input operational amplifiers achieving 0.33 µv/ c offset drift,* Solid-State Circuits, IEEE Journal of **46**, 2099 (2011).

[30] J. Huijsing, *Operational Amplifiers Theory and Design* (Springer, 2011).

[31] A. Bakker, K. Thiele, and J. Huijsing, *A cmos nested-chopper instrumentation amplifier with 100-nv offset,* Solid-State Circuits, IEEE Journal of **35**, 1877 (2000).

[32] C.-G. Yu and R. Geiger, *An automatic offset compensation scheme with ping-pong control for cmos operational amplifiers,* Solid-State Circuits, IEEE Journal of **29**, 601 (1994).

[33] C. Eichenberger and W. Guggenbuhl, *Charge injection of analogue cmos switches,* Circuits, Devices and Systems, IEE Proceedings G **138**, 155 (1991).

[34] G. Wegmann, E. Vittoz, and F. Rahali, *Charge injection in analog mos switches,* Solid-State Circuits, IEEE Journal of **22**, 1091 (1987).

[35] J. Xu, Q. Fan, J. Huijsing, C. Van Hoof, R. Yazicioglu, and K. Makinwa, *Measurement and analysis of current noise in chopper amplifiers,* Solid-State Circuits, IEEE Journal of **48**, 1575 (2013).

[36] P. Van Peteghem and W. Sansen, *Single versus complementary switches: a discussion of clock feedthrough in s.c. circuits,* in *Solid-State Circuits Conference, 1986. ESSCIRC '86. Twelfth European* (1986) pp. 143–145.

[37] M. Pertijs and W. Kindt, *A 140db-cmrr current-feedback instrumentation amplifier employing ping-pong autozeroing and chopping,* in *Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International* (2009) pp. 324–325,325a.

[38] C.-C. Enz, *High precision CMOS micropower amplifiers*, Ph.D. thesis, Lausanne (1989).

[39] Q. Fan, J. Huijsing, and K. Makinwa, *Input characteristics of a chopped multi-path current feedback instrumentation amplifier,* in *Advances in Sensors and Interfaces (IWASI), 2011 4th IEEE International Workshop on* (2011) pp. 61–66.

[40] J. Xu, Q. Fan, J. Huijsing, C. Van Hoof, R. Yazicioglu, and K. Makinwa, *Measurement and analysis of current noise in chopper amplifiers,* Solid-State Circuits, IEEE Journal of **48**, 1575 (2013).

[41] V. Ivanov, *Analog techniques for nano-power circuits.* in *Tutorial Solid- State Circuits Conference - (ISSCC), 2015 IEEE International* (2015).

[42] J. Racko, M. Mikolášek, L. Harmatha, J. Breza, B. Hudec, K. Fröhlich, J. Aarik, A. Tarre, R. Granzner, and F. Schwierz, *Analysis of leakage current mechanisms in ruo2–tio2–ruo2 mim structures,* Journal of Vacuum Science Technology B **29** (2011), http://dx.doi.org/10.1116/1.3534022.

[43] Y. Kusuda, *A 60 v auto-zero and chopper operational amplifier with 800 khz interleaved clocks and input bias current trimming,* Solid-State Circuits, IEEE Journal of PP, 1 (2015).