An Alternative ARAM Cell Design to Increase Storage Time

REGŐ KOZMA, CSABA HORVÁTH, FERENC KOVÁCS, GÁBOR HOSSZÚ
Department of Electron Devices
Budapest University of Technology and Economics
Budapest, 1111 Goldman Gy. tér 3
HUNGARY

Abstract: Long-term data storage in analogue signal processing is hindered by the errors of the data storage units generally used. This paper proposes a possible solution that is capable of storing high precision analogue data for a relatively long time. This solution has numerous uses in CNN applications and in custom signal processing devices.

Keywords: Analog RAM, Long storage time, CNN

1 Introduction
The problem that no analogue data storage unit capable of storing high precision data for a long time period exists was encountered while developing a custom signal processing device [1]. Since the data that was to be processed was fetal heart sound [2], the nature of the input requires data samples to be stored for as long as 1-2 seconds. The same problem was encountered by CNN device developers[3]. The solution proposed by CNN researchers utilized switched-capacity circuitry [4][5], but is capable of short-term storage only. Since CNN technology requires storage times of a few hundred milliseconds, that solution was suitable[6]. Using the aforementioned CNN storage unit as the base of our research, we improved it to a device that is capable of storing data for much longer periods.

2 Implementation

2.1 The basic storage unit
In the development of the base cell of the analogue RAM unit, previous work laid down in [7] were used, since these discuss the sample-holder shown on Fig. 1., that is the base unit of analogue data storage, in detail.

![Fig. 1. Analogue RAM base unit](image)

Fig. 1. Analogue RAM base unit

Practically the sample-holder circuit works as a low-pass filter in this case.

![Fig. 2. Voltage on capacity](image)

Fig. 2. Voltage on capacity

Fig. 2 shows how the sample works: chart 1 shows the input signal, chart 2 shows the signal driving the sample-holder circuit, chart 3 shows how the circuit should work ideally, and chart 4 shows how the circuit really works. On chart 4 one can clearly see that the sampled signal shown in bold never reaches the level of the sampled input signal. This deviance is caused by different error factors. The first factor is the acquisition error ($e_a$) where the maximum sampling frequency also must be taken into account (bandwidth error). The method to calculate this error is shown on (1), with $V_{\text{max}}$ and $V_{\text{min}}$ being the maximal and minimal voltages on the capacitor, respectively.

$$V_{\text{max}} - V_{\text{min}} = \frac{V_{\text{max}} - V_{\text{min}}}{1 + \frac{R}{C}}$$
\[ \varepsilon_a = -(V_{\text{max}} - V_{\text{min}})e^{-\frac{t}{\tau}} = -Ae^{-\frac{t}{\tau}} \] (1)

At the end of the sampling phase, two other errors show up causing a slight drop in the signal level: the error caused by charge injection and the feedthrough error (noted as \( \varepsilon_{ch} \) and \( \varepsilon_f \), respectively). The approximation of \( \varepsilon_f \) is shown on (2),

\[ \varepsilon_f \approx \frac{C_{ov}}{C_1 + C_{ov}} \left[ V_{\text{write}} - V_{S,0} - V_{th}(V_{S,B}) \right] \] (2)

where \( C_{ov} \) is the overlapping capacitor from the writing gate controlled by \( V_{\text{write}} \), \( V_{S,0} \) is the voltage on the source of writing transistor at turning off, and \( V_{th}(V_{S,B}) \) is the threshold voltage as function of the voltage difference between the source and the bulk.

To approximate the error \( \varepsilon_{ch} \), (3) should be used.

\[ \varepsilon_{ch} = -\rho \frac{C_{ov}WL}{C_1} \left[ V_{\text{DD}} - V_{\text{write}} - V_{th}(V_{S,B}) \right] \] (3)

\( V_{\text{DD}} \) is the maximum driving voltage of the transistor, \( W, L, C_{ov} \) are technology parameters, and \( \rho \) is the ratio of the charge flowing into the storage capacitor. Because of certain nondeterministic properties the error \( \varepsilon_{ch} \) cannot be expressed exactly, and the parameter \( \rho \) was introduced to represent their effect.

After the sampling phase the charge stored in the capacity leaks slowly away over time.

### 2.2 Non-compensated ARAM

The sample-holder circuit discussed before can be built into read/write circuitry as shown on Fig. 3.

![Diagram of Basic analogue RAM unit](image)

**Fig. 3. Basic analogue RAM unit**

The actual storage is handled by the C1, C2, ..., Cn capacities, and the capacities are addressed using the S1, S2, ..., Sn transistors.

Writing into the storage is enabled by the transistor "Write". The function of the "Write" transistor is the decoupling of the O1 amplifier to reduce the error caused by feed-forward. The driving signal of "Write" must slightly precede the signal of "Write".

The transistor named Read enables or disables reading out stored data.

The following equations describe how the above circuitry functions:

\[ V_{C_k} = V_{In} - V_{offset} - V_{ref} \] (4)

where \( V_{C_k} \) is the voltage of the k-th capacity

\( V_{In} \) : input voltage

\( V_{ref} \) : reference voltage of amplifier

\( V_{offset} \) : offset voltage of amplifier

\[ V_{Out} = V_{C_k} + V_{offset} + V_{ref} \] (5)

where \( V_{C_k} \) is the voltage of the k-th capacity

\( V_{Out} \) : output voltage

\( V_{ref} \) : reference voltage of amplifier

\( V_{offset} \) : offset voltage of amplifier

Based on the above equations \( V_{Out} = V_{In} \), thus the entered values can be regained.

The above circuitry is the analogue RAM cell used in CNN technology. Its main flaw is the already mentioned charge leakage that limits the maximal storage time to a few hundreds of milliseconds since there is no compensation when values are read out. CNN applications do not require long storage times though, since the time elapsed between writing and reading the cell is small, causing an error margin of a mere 1-1.5% that limits accuracy to 7-8 bits.

By this point, the parameters describing capacities and transistors have been determined and thus the circuitry could be built. One must still notice though that every read and write operation requires the use of two transistors, which in turn affects error parameters. Our results show that the change of transistor parameters from \( L=0.4 \, \mu \text{m} \) and \( W=1.0 \, \mu \text{m} \) to \( L=0.4 \, \mu \text{m} \) and \( W=2.0 \, \mu \text{m} \) does not make a difference in the electronic parameters of the circuitry.

### 2.3 Compensated ARAM unit

A method of error compensation has been developed that lessens the effects of charge leakage by using an auxiliary capacity array thus lengthening the possible storage time. The error compensation method is based on the assumption that the decrease in the level of the stored signal value is linear if the value stored in the capacity is
refreshed very often. Along with leakage error compensation, refreshing of the stored data could also be done thus further lengthening the storage time. The block diagram of the implementation is shown on Fig. 4.

Fig. 4. The complete ARAM device

As one can see, there are two capacity arrays for storage, but the capacities of the second array (auxiliary error compensation capacity array) are charged to \( V_{ref} = 1 \) V voltage instead of the input voltage. Charging of the second array is synchronized to the main storage array.

We must allow for the fact that the rate of leakage is not constant because of the exponential discharge characteristics. This error can be minimized by three different methods. Firstly, the regions where the discharge characteristics are flat must not be used. The \( V_{ref} \) voltage must be set to the middle of the working region of the characteristics. The values stored in the capacities must also be refreshed quickly.

Upon reading values from the cell, the value of the capacity in the upper array is \( V_{in} - V_{leak} \) while the value of the compensating capacity is \( V_{ref} - V_{leak} \). This leakage can be compensated by using the following equation:

\[
V_{out} = V_{in} - V_{leak} - (V_{ref} - V_{leak}) + V_{ref} \tag{6}
\]

When reading, the value of the compensating capacity \( V_{ref} - V_{leak} \) is channeled into an adder circuit with a negative sign in order to determine by how much the stored value has decreased. Since the value of \( V_{leak} \) will be positive, it can simply be added to the value read out to compensate leakage error. The value that has been read out can then be used normally in the signal processing procedure and will be used to refresh the values stored in the ARAM unit, too. Refreshing of the values is scheduled by using the transistors M1, M2, M3. These transistors handle refreshing via a tracking mode op. amplifier. The inverter marked A3 can be left out from the circuit, if the compensating capacities are charged to \(-V_{ref}\) instead of \(V_{ref}\).

During the planning of the actual device, one must aim to place the storage and compensation capacities in pairs that are as close to each other as possible. This should ensure that the parameters that affect leakage (like temperature and material quality) should be almost the same in the two capacities.

In the above circuit, the error caused by the offset of the op. amplifiers can be lessened if the reference voltage at the first adder circuit is decreased by the calculated offset voltage. The A1, A2, A3 adder and multiplier circuits have been built from the op. amplifiers discussed before, using their basic circuit blocks.

3 Results

An analogue RAM device as shown on Fig. 4 has been developed. The ARAM device used in CNN applications has been improved using a compensation method.

The new device has been tested using SPICE simulation.

Fig. 5 shows 2.3V being stored into the new ARAM device (x: storing time in ms, y: storing value in V). The continuous and the broken lines show the results acquired from the non-compensated and the compensated ARAM cells respectively.

One can clearly see that the error is much smaller in the compensated module.
Fig. 6 shows a case where the compensator circuit overcompensates the error. The continuous line shows the result from the non-compensated cell, and the broken line shows the same from the compensated ARAM cell. The overcompensation is caused by the nonlinearity of the leakage.

Fig. 7 shows input data voltages and the deviation of the capacity's actual voltage from the originally entered value after 100 ms. The continuous line represents the results of a non-compensated cell. The broken line with squares shows the effects of compensation, while the broken line with triangles shows the effect of refreshing with 3 ms refresh interval. Refreshing lessens the effect of nonlinearity, since in a short time the leakage profile does not deviate much from linear.

If the nonlinearity of the charge leakage is to be taken into account, the compensation should be carried out using a finely adjusted voltage-controlled amplifier.

These results show that the compensation method is a feasible way to increase the maximum retention time of ARAM cells, and that further improvements are possible in this field. In order to attain higher accuracy, the leakage errors must be further decreased either by improving circuitry, or by developing a non-linear amplifier model to allow us to apply more accurate compensation values.

References: