Analog Integrated Circuits and Signal Processing 2, 281-295 (1992)

Full Analog CMOS Integration of Very Large Time Constants for Synaptic Transfer in Neural Networks

P. KINGET AND M. STEYAERT
ESAT-MICAS, K. U. Leuven Kard. Mercierlaan 94, B-3001 Heverlee, Belgium
J. VAN DER SPIEGEL
Department of Electrical Engineering, University of Pennsylvania, Philadelphia, PA 19104-6390

Abstract. A method for the full on-chip analog implementation of large time constants in a CMOS technology is presented. These time constants are used for delayed synaptic transfer in neural networks for signal processing. For real-time speech recognition time constants from 1 to 500 ms are necessary, that vary logarithmically with a 4-bit digital code. An *RC*-type circuit is used to make a continuous time implementation. A modified operational transconductance amplifier (OTA) in a negative feedback configuration is used as a resistor and a gate-oxide capacitor is used as load. The output current of the OTA is divided by current dividers to obtain an electronic multiplication of the time constant. An adapted current mirror design is applied. Much attention is paid to offset-voltage reduction. This new multiplication scheme has several important advantages compared to other schemes. A global area minimization procedure is explained. The total active are of a 3-μm CMOS chip for a 500-ms time constant is 0.5 mm^2. The realized circuit has a typical measured offset voltage of 130 mV. The area efficiency and the offset sensitivity are shown to be one order of magnitude better than in other designs.

1. Introduction

In the design and integration of artificial neural networks for signal processing applications, time constants (1 to 500 ms) for synaptic transfer are required. They are necessary building blocks for the temporal integration, differentiation and generation of real-time signals in applications such as speech recognition, motion detection and neurocontrol [1, 2]. Of course many other signal processing circuits, e.g., in audio, control and biomedical applications, use large integrated time constants as well.

In the integration of neural networks a minimal chip area for the building blocks is the main concern, since a high circuit density is necessary for real-world applications. The adaptability of the time constants is desired for learning purposes. Due to the fault tolerance and the adaptive nature of neural networks there is no need for large accuracies. The use of a standard circuit technology is of course preferred for the compatibility with other circuits and for economical reasons.

P. Kinget is a research assistant of the National Fund for Scientific Research, Belgium.

In this paper the analog integration of very large time constants is addressed. In the next section an overview of the desired features and the design restrictions for the circuit are given. In Section 3 different time constant circuits are examined for their aptitude for implementing large time constants. In Section 4 electronic multiplication of time constants is discussed and a new multiplication scheme is presented. Section 5 considers the tunability of the time constants. In Section 6 the total area optimization and layout considerations are given. Experimental results are presented in Section 7. In Section 8 a comparison with other published designs is made. Finally some conclusions are drawn in Section 9.

2. Features and Restrictions

At present, work is carried out for the development of a general-purpose analog neural computer [1]. It is a modular machine consisting of neuron-, synapse-, and switch chips. The neuron characteristic, the synaptic weights and time constants and the interconnection topology are all controlled by a digital host computer. The presented design has to be compatible with the

existing designs of the other components of the analog neural computer. As a result the following specifications are required:

- Signal range 0–3 V
- *Vdd/Vss* +5–−5 V
- CMOS technology (no external components allowed)
- Logarithmic variation of the time constant with 4-bit digital code (16 time constants) from 1 to 500 ms
- Accuracy: error $<20\%$

From the experience with a discrete neural network and from biological inspiration a simple *RC* structure is proposed for the time constant circuit. This makes the circuit a first order system. Hence the time constant (τ) is independent of the signal amplitude and the frequency response is a first-order low-pass filter with a cutoff frequency of $1/(2\pi \times \tau)$.

The design problem discussed in this paper is the realization of a chip, with no external components, for a building block with a time constant as high as 500 ms. It is this specification which introduces the bottleneck toward a full on-chip integration of the building block.

For the integration of neural networks *analog* circuits are often preferred for the high integration density that can be achieved. The low accuracy of analog circuits is compensated by the adaptiveness and fault tolerance of neural networks.

The use of a standard CMOS process makes the circuits compatible with, e.g., digital circuits and is economically cheap. Moreover, CMOS circuits have a low power consumption which favors high integration densities.

3. Review of Time Constant Circuits for Large Time Constants

The on-chip implementation of large time constants has to deal with the limited availability of pasive components. A full passive implementation of a 500-ms time constant results easily in a 7-mm^2 chip area for a 3-μm CMOS technology. An alternative using active elements has to be investigated.

3.1. Area of a Time Constant Circuit

Assuming that a resistor R_{area} and a capacitor C_{area} per unit area are available the total area A_{TOT} of the circuit is

$$A_{\text{TOT}} = \frac{R}{R_{\text{area}}} + \frac{C}{C_{\text{area}}} \tag{1}$$

For a given time constant ($\tau = RC$) the optimal resistance value can be computed by substituting C by τ/R in equation (1) and computing the derivative with respect to R. The following results are then obtained for the optimal resistor (R_{opt}), the optimal capacitor (C_{opt}) and the optimal total area ($A_{\text{TOT}_{\text{opt}}}$):

$$\begin{aligned} R_{\text{opt}}(\tau) &= \sqrt{\frac{R_{\text{area}}}{C_{\text{area}}}}\,\tau \\ C_{\text{opt}}(\tau) &= \sqrt{\frac{C_{\text{area}}}{R_{\text{area}}}}\,\tau \\ A_{\text{TOT}_{\text{opt}}}(\tau) &= 2\sqrt{\frac{\tau}{R_{\text{area}}C_{\text{area}}}} \end{aligned} \tag{2}$$

For this optimal realization it is easily verified that the optimal resistor and the optimal capacitor each occupy *half of the total area*. To obtain a small total area a good resistor structure (high R_{area}) and capacitor structure (high C_{area}) have to be used.

3.2. Time Discrete Realization

Using switched capacitor techniques, time constant circuits can be designed using a switched capacitor as resistor [3]. However these circuits are discrete-time systems and hence continuous time antialiasing input and output filters are necessary. To obtain large time constants low clock frequencies have to be applied which results in low-frequency continuous time low pass antialiasing filters. Considering the total area of switched capacitor circuit, clock oscillator and antialiasing filters, a continuous time circuit looks more area efficient. This will also be proven in Section 8, where several designs are compared toward chip area.

3.3. Capacitor Structure

On chip two types of capacitors are available: *gate-oxide* and *double-poly* capacitors. Both have a large voltage range but the linearity of a double-poly capacitor is better than that of a gate-oxide capacitor. However the capacitance of gate-oxide capacitors is typically twice the capacitance of a double-poly capacitor. Since our first concern is area minimization a gate-oxide capacitor is selected.

3.4. Resistance Structure

A linear V-to-I converter is necessary to charge the capacitor with a fixed time constant.

The most straightforward way for realizing a V-to-I converter on chip is using a *diffusion resistor* or a *MOS transistor in its linear region*. Both have the same limitations. Their resistance values are (for a constant voltage range) determined by their size, which means hardly tunable. They have a low R_{area} on the order of 50 $\Omega/\mu m^2$ for the diffusion resistor and 2 $k\Omega/\mu m^2$ for the MOS resistor (cf. Appendix I). For an optimal total area the size of the resistor is equal to the size of the capacitor. This means that the parasitic capacitance to the bulk is of the same order of magnitude as the load capacitance which results in a distributed line behavior. Moreover, this distributed line effect results in a transfer function of higher order [4].

An *operational transconductance amplifier* (OTA) in a negative feedback configuration has an output current that is proportional to the voltage difference between noninverting input and output. This structure can thus be used as a resistor and is therefore called an OTA-R (figure 1). The equivalent resistance, when the input transistors operate in *strong inversion*, equals

$$R_{eq} = \frac{1}{gm} = \frac{(V_{GS} - V_T)_{DP}}{I_{BIAS}} = \frac{1}{\sqrt{(\mu C_{ox} W/L)_{DP} I_{BIAS}}} \quad (3)$$

with I_{BIAS} the tail current of the differential pair [5]. The $(V_{GS} - V_T)_{DP}$ is determined by the desired linear voltage range. The attainable R_{area} is in the range of 1 $k\Omega/\mu m^2$ (see Appendix I). For a given OTA-R the resistance value is tunable through the bias current. However this is limited by the change of the linear range of the OTA-R.

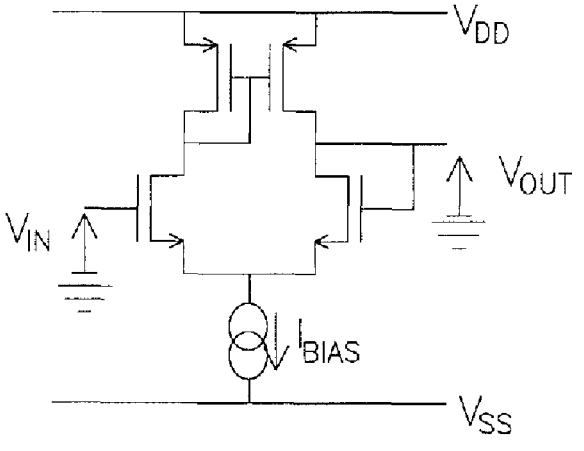


Fig. 1. Schematic of a simple OTA in a negative feedback configuration used as an OTA-R.

When the input transistors are operated in *weak inversion*, the equivalent resistance is very large but the linear input range is limited to about ±70 mV.

From the above discussion it is clear that for the purpose of integrating large *tunable* time constants, with a *first-order behavior*, the use of an OTA-R operated in strong inversion is the best choice. The R_{area} is lower than for a MOS resistor but can be increased through electronic multiplication. Using this approach a high resistance value is obtained as will be shown in table 4 of Section 8.

4. Electronic Multiplication of Time Constants: A New Scheme

4.1. General Considerations

A general model for a time constant circuit and for the electronic multiplication of a time constant is presented in figure 2a and b [6]. The output voltage for 2b can be expressed as:

$$V_{out} = \frac{V_{IN} + V_{E_1} + BV_{E_2}}{1 + B\tau s} \quad (4)$$

The time constant is multiplied with a factor B, but the influence of errors (offset voltages, mismatches, leakage currents, ...) that occur *after* the multiplication (V_{E_2}), is also multiplied with the same factor B. The influence of the errors occurring *before* the multiplication (V_{E_1}), is not amplified by the multiplication factor. This implies that the placement of the multiplication scheme has to be considered very carefully in order to obtain a reasonably accurate circuit.

4.2. New Multiplication Scheme for an OTA-R

In figure 3a a new multiplication scheme for an OTA-R based on output current division, is presented. In contrast to classical OTA designs a *current division* (B = 40,000) is used instead of multiplication (B = 0.3-1) and the *placement of the current dividers is nonsymmetric*. A model for the circuit is drawn in figure 3b and the output voltage can be expressed as

$$V_{OUT} = \frac{V_{IN} + V_{os} + BRI_{leak}}{1 + BRCs} \quad (5)$$

- V_{os} includes the errors that occur *before* the current dividers (mismatches differential pair transistors, mismatches between current dividers).

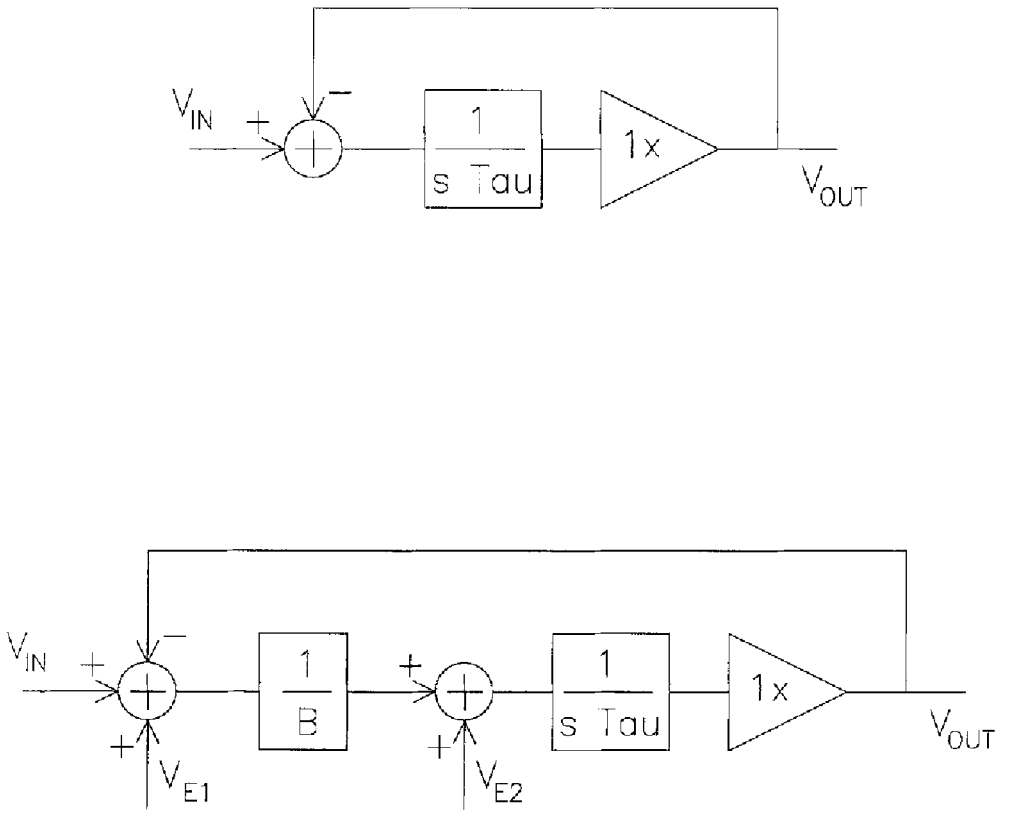


Fig. 2. (a) General model for a time constant (tau) circuit. (b) General model for electronic multiplication circuits for time constants. V_{E_1} and V_{E_2} represent the error sources that occur in the circuit.

- I_{leak} includes the errors that occur *after* the current dividers (mainly leakage currents from drain-bulk diodes of output transistors $M6$ and $M8$).

It is clear that errors after the current dividers will contribute most to the offset of the filter. This justifies the nonsymmetric placement of the current dividers. The offset of the OTA, V_{os} will increase due to lack of symmetry. However, *only in one branch* very small currents are flowing, which reduces largely the effect of the second term. Moreover, the leakage currents of the two bulk diodes compensate each other a little. The temperature dependence of the offset voltage is also limited to the influence of the leakage currents in the output branch, which is a clear advantage over other schemes.

Using a large B factor makes it possible to use a "large" input bias current such that a linear input range of several volts is possible. This allows also to bias current mirrors in *strong inversion*, which minimizes mismatch errors in the current mirrors themselves. Moreover, in Section 6 a *special current divider layout* is presented to improve the matching in the large current divider circuits.

4.3. Comparison with Other Multiplication Schemes

4.3.1. Capacitance Multiplication [7]. Figures 4a and b show two possible principles for capacitance multiplication. Scheme 4(a) is only useful for small signal ranges when using large A. In scheme 4(b) the leakage currents have a contribution to the offset equal to $AI_{leak}R$, which is comparable to our circuit. The implementation of the current multiplier results also in a current mirror structure with a large current multiplier factor. However this method has an additional problem: the small charging and discharging capacitor currents must be measured without disturbing the capacitor voltage itself. This is an extra problem which does not arise with the presented technique.

4.3.2. Transconductance Reduction Techniques [8, 9]. Transconductance techniques are commonly used to increase the slew-rate of operational amplifiers. In all circuits the reduction is also realized through dimension ratios of input transistors, but the problem of leakage currents is never concentrated in one branch. Consequently, these techniques suffer for this application from

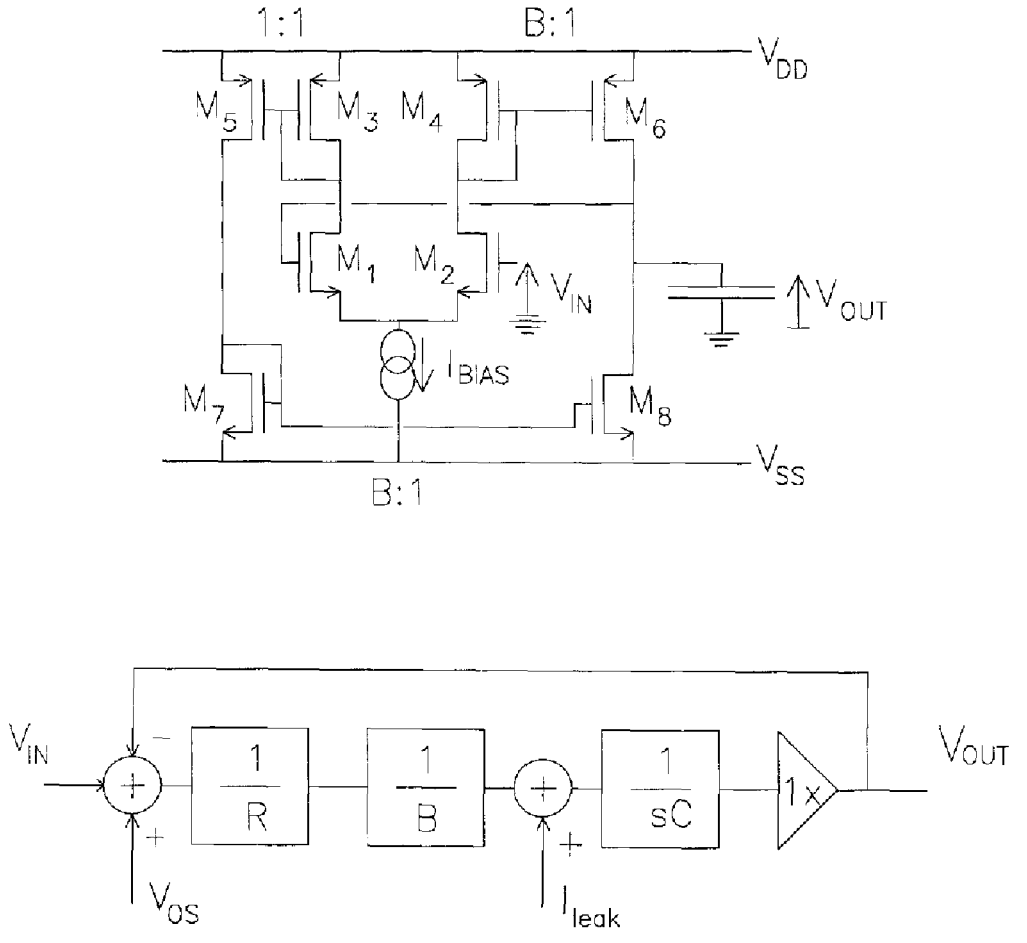


Fig. 3. (a) Principle transistor structure of the new multiplication scheme for an OTA-R based time constant circuit. Note the placement of the current dividers. (b) Model for the circuit of figure 3a. V_{os} represents the leakage current at the output node.

high offset. Moreover, at the input extra transistors are added to achieve a current subtraction and thus a *gm* reduction, but the noise of the extra transistors is *added* to the noise of the original transistors. This circuit has a lower noise performance than the presented approach.

4.3.3. Input and Output Voltage Dividers. This is the most straightforward application of the general scheme of figure 2b. A fraction of the input voltage and the output voltage is connected to the inputs of the OTA. The voltage dividers are easily implemented using MOS transistors which implies that the multiplication factor is again determined by transistor dimension ratios. However in this circuit the effect of the leakage currents at the output *and* of the offset of the OTA are amplified, which is a serious drawback.

In all multiplication schemes the multiplication factor is determined by dimension ratios of MOS transistors. Our new scheme has the important advantage that the amplified errors are localized in only one circuit branch. Thanks to its simplicity the circuit realization can be better optimized and is more reliable.

5. Tunability

Several time constants must be obtained with one circuit. Since the integration of large time constants is most difficult, an optimal circuit is designed for the largest time constant. This circuit is then modified to make the time constant tunable toward smaller time constants.

The time constant for the presented circuit (figure 3a) is

$$\tau = \frac{BC}{gm} = \frac{BC}{\sqrt{(\mu C_{ox} W/L)_{DP} I_{BIAS}}} \tag{6}$$

Three possible parameters are available to tune the time constant: B, C, and I_{BIAS}. To reduce the time constant

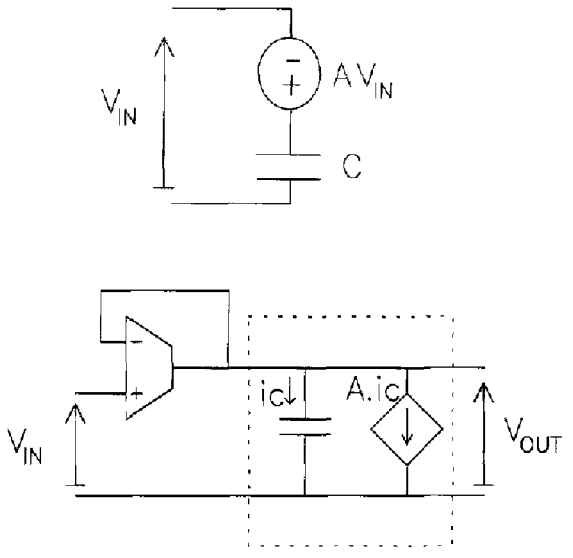


Fig. 4. (a) Capacitance multiplication scheme based on a VCVS. The equivalent capacitance is $(A + 1)C$. (b) Capacitance multiplication scheme based on a CCCS. The equivalent capacitance is $(A + 1)C$.

the B or C values have to be decreased or the bias current has to be increased.

5.1. Changing Capacitor C

The capacitor C can be split in a parallel connection of capacitors and with a few switches a variable load capacitor is realized. Two limitations have to be taken into account:

- The load capacitance must provide the necessary load for the stabilization of the negatively fed-back OTA.
- The switches may not cause extra leakage paths at the output node.

We split the load capacitor in four parallel capacitors. The smallest one is permanently connected to the output and is large enough to stabilize the OTA. The three other capacitors can be separately switched in to change the time constant. In this way four different time constants are obtained.

5.2. Changing the Current Division Factor B

By splitting the large current division into a succession of several smaller current divisions, which can each be bypassed, the current division factor is alterable. But now small currents are flowing in several circuit branches, which is detrimental for the leakage current immunity. This procedure can only be used when the intermediate dc currents are still large enough. This implies that only a small fraction of the B factor can be split off.

In the presented realization the large B factor was split in two cascaded current dividers: a 22 current divider followed by a $B/22$ current divider. The transistors $M4$, $M6$ of figure 3a are replaced by the circuit represented in figure 5 and the transistors $M7$, $M8$ are replaced by a similar structure. The current in the intermediate branch is still large enough compared to the leakage currents and no considerable extra offset voltage is generated. When the switch (SW) is closed, the 22 current divider is bypassed. Two different time constants can be realized.

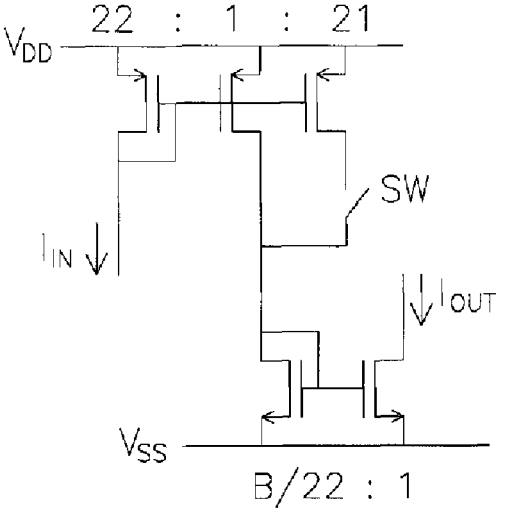


Fig. 5. Schematic of cascaded current divider: switch SW open, current division factor B; switch SW closed, current division factor $B/22$.

5.3. Changing the Bias Current I_{BIAS}

Changing the bias current of the differential pair is the only way for continuously fine-tuning the time constant. The dynamic range of this timing is however seriously limited by two factors:

- An increase of the bias current results in an increase of the V_{GS} of the input transistors and consequently the voltage drop over the bias current source decreases.
- A reduction of the bias current results in a reduction of the V_{GS} of the input transistors, which implies a smaller linear input range.

We use an external reference for the bias current. This allows us to use two different bias currents to obtain two different time constants.

None of the above tuning techniques is flexible enough to realize a logarithmic adjustable time constant in the range from 1 to 500 ms. We use a combination of them to get 16 different time constants in this range.

6. Area Optimization and Layout

In the previous sections a principal circuit was developed for the integration of large time constants (figure 3a). In this section the optimization of the total area (OTA-R and load capacitor) of the circuit is established. It is very important to aim for a minimal *total* area and not for a minimal capacitor area. Using a very large resistor and a small capacitor is only sensible when the area efficiency of the resistor implementation is better than the area efficiency of the capacitor implementation. Otherwise, the same time constant can be obtained on a smaller area using a larger capacitor and a smaller resistor. To get a minimal total area, resistor area and capacitor area have to be optimized together.

First an expression of the area of an OTA-R as a function of the resistance (R) must be derived. Assuming that all current mirrors are biased in strong inversion (i.e., $V_{GS} - V_T > 200$ mV) and that the $V_{GS} - V_T$ of the input transistors is determined by the desired linear input range, a simple expression for the area of an optimal OTA-R ($A_{\text{OTA-R}}$) is obtained (cf. Appendix II):

$$A_{\text{OTA-R}}(R) = A_1 + \frac{R}{R_{\text{area}}} \tag{7}$$

The constant A_1 represents the area of the OTA, the bias circuitry included but the area of the current dividers excluded. The optimal bias current is independent of the resistance (R) and is defined by process parameters. The constant R_{area} is determined by the area efficiency of the current dividers realization.

The area of the capacitor is

$$A_C = \frac{C}{C_{\text{area}}} \tag{8}$$

C_{area} equals the process parameter C_{ox} for gate-oxide capacitors.

Now the optimal R_{opt} and C_{opt} values and the optimally achievable total area ($A_{\text{TOT}_{\text{opt}}}$) can be easily calculated for a time constant τ and are

$$R_{\text{opt}}(\tau) = \sqrt{\frac{R_{\text{area}}}{C_{\text{area}}}\,\tau}$$

$$C_{\text{opt}}(\tau) = \sqrt{\frac{C_{\text{area}}}{R_{\text{area}}}\,\tau}$$

$$A_{\text{TOT}_{\text{opt}}}(\tau) = A_1 + 2\sqrt{\frac{\tau}{R_{\text{area}}\,C_{\text{area}}}} \tag{9}$$

These results are identical to the case of a simple RC implementation except for the constant A_1 term in the total area. For this circuit the area of the current dividers and the area of the capacitor are equal. From the optimal resistance value and the optimal bias current the value of the B factor is calculated. In table 1 the optimal values for our 3-μm CMOS process are given.

Table 1. Optimal design parameters for a 500-ms time constant circuit in a 3-μm CMOS process.

R_{opt}	6600 MΩ
C_{opt}	75.8 pF
I_{BIASopt}	15 μA
B_{opt}	46670

With these optimal design parameters the actual transistor sizing can be done. The design of the differential pair, bias current source and capacitor switches is straightforward. An extra measuring buffer is added at the output to avoid insertion of extra external leakage paths when performing measurements.

The design and layout of the large current dividers (division factor $B' = B/22$) demands special consideration. The integration of the current mirror factor based on simple relative transistor dimension ratios results in a short transistor and an extremely long transistor, with no matching between both. A first solution is to implement the longest transistor as a series connection of B' unit transistors with their gates connected and of the size of the shortest transistor. The division factor is then defined by the number of transistors in series. The matching of the individual transistors can be improved by using a centroid layout. However this introduces B' intermediate extra source-drain areas with a considerable increase in leakage current at the output node. When the short transistor is implemented as a parallel connection of N-times longer unit transistors and the long transistor as a series connection of B'/N of these unit transistors, the number of leakage areas is reduced by N. Using this idea a compromise between extra area and leakage current reduction can be found. Figure 6 is a schematic of the large nMOS current divider. In table 2 the transistor size of the current dividers are summarized.

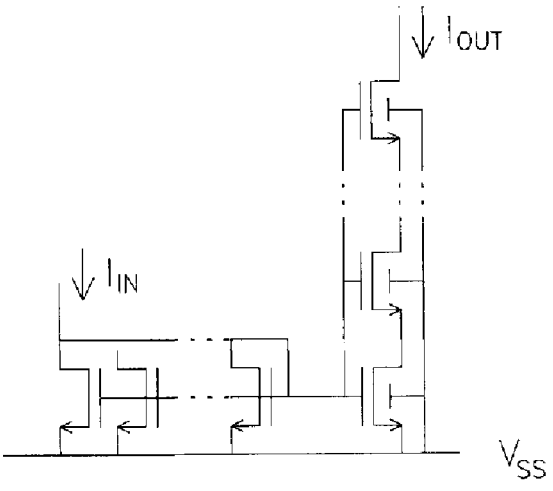


Fig. 6. Circuit schematic for large nMOS current divider. All transistors are unit transistors. The current division factor is the number of parallel connected transistors times the number of series connected transistors.

Table 2. Transistor dimensions for current dividers.

	nMOS Current Divider	pMOS Current Divider
Unit transistor *W/L*	3 μm/50 μm	3 μm/25 μm
Number of *series* connected unit transistors for *long* transistor	265	212
Number of *parallel* connected unit transistors for *short* transistor	8	10
Division factor *B'*	265 × 8 = 2120	212 × 10 = 2120

Figure 7 is a microphotograph of the layout. The capacitor block consisting of four capacitors and the centroid structure of the two output current dividers are well distinguished.

7. Experimental Results

A 500-ms time constant was fabricated in a 3-μm double metal, n-well CMOS process. Some measured circuit characteristics are summarized in table 3.

Figure 8 is a frequency response of the circuit for the largest time constant. The cutoff frequency is 0.220 Hz (720 ms) which is slightly out of the 20% accuracy range from 500 ms. This deviation is mainly due to a bad characterization of the C_{ox} value for large capacitors. A clear first-order behavior is observed. At higher frequencies an extra zero is introduced by the capacitance feedthrough from input to output through the gate-source capacitances of the input transistors. The effect of this zero can be neglected.

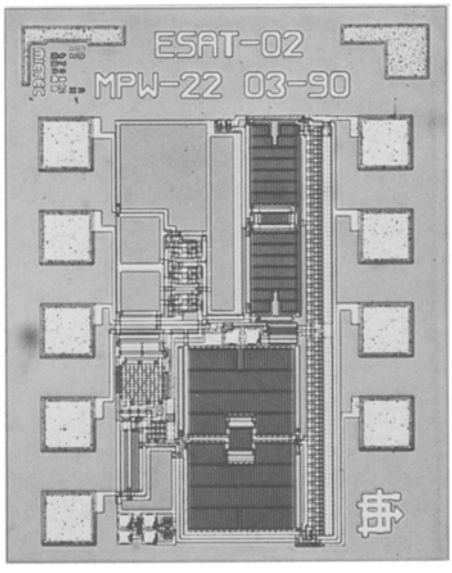


Fig. 7. Microphotograph of the realized circuit. In the upper left corner the capacitor block is located. The current dividers are situated on the right side.

Table 3. Measured circuit characteristics.

Supply voltages	+5–−5 V
Total current (largest time constant)	135 μA
Mean offset voltage (largest time constant)	130 mV
St. dev. offset voltage (largest time constant)	81 mV
Integrated output noise voltage (1 Hz–1 kHz)	
Largest time constant	30 μV_{rms}
Smallest time constant	100 μV_{rms}

Figure 9 presents the output response for the largest time constant to block voltages of 1, 1.5, 2, 2.5, and 3 V. Clearly, no slewing occurs in the voltage range from 0 to 3 V. The time constants for the positive step are respectively 720, 736, 746, 752, and 772 ms, which are all contained in a 20% accuracy range. The circuit performs similarly well for the negative step.

Figure 10 illustrates the programmability of the time constants with the digital tune code. There is a good logarithmic agreement between digital code and time constants. Figure 11 shows the time response to a constant step input voltage for different time constants.

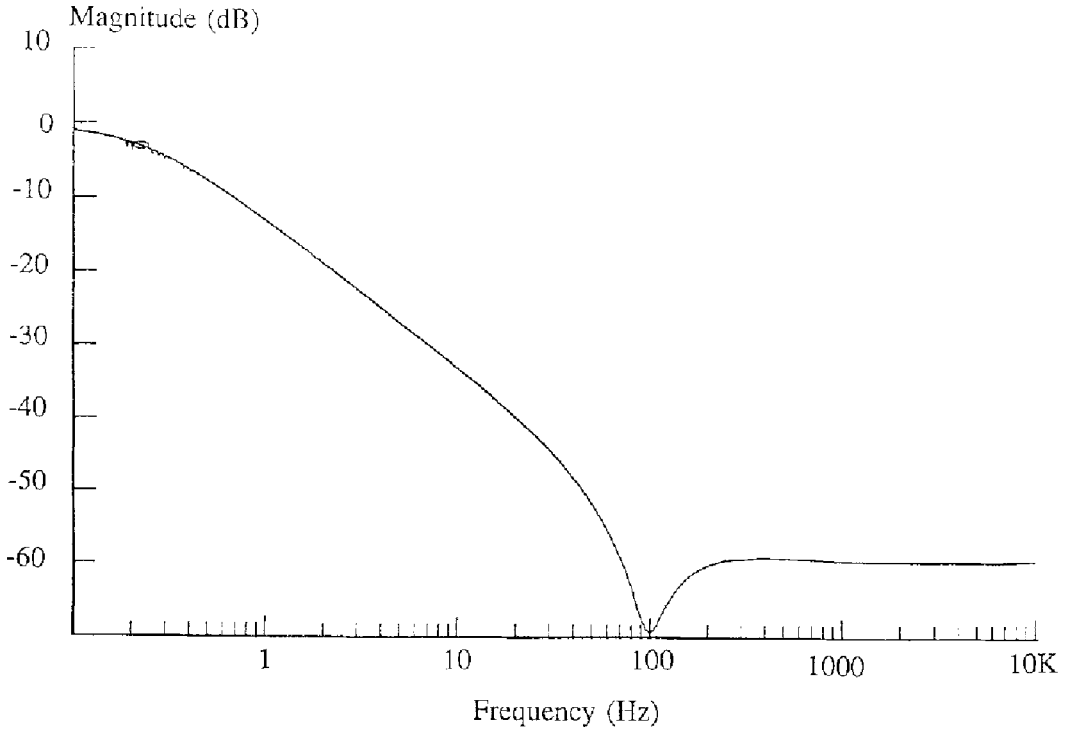


Fig. 8. Frequency response for the largest time constant. Marker is at −3.0 dB and 0.22 Hz. A first-order behavior is observed. The effect of the extra zero at higher frequencies can be neglected.

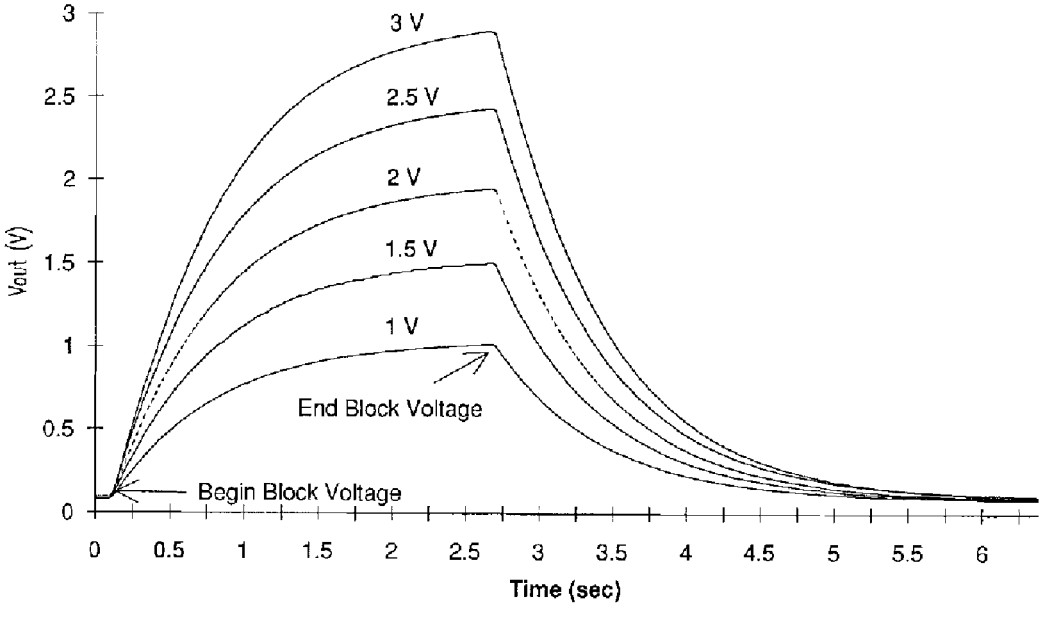


Fig. 9. Output response for the largest time constant to several block input voltages.

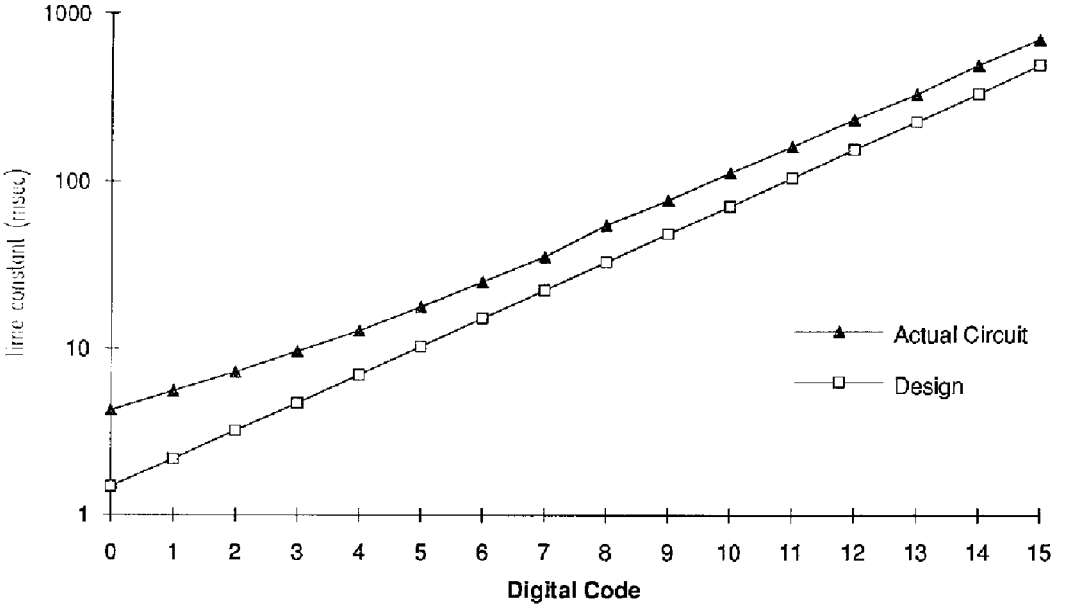


Fig. 10. Variation of the time constant with the digital tuning code. A good logarithmic dependence is observed.

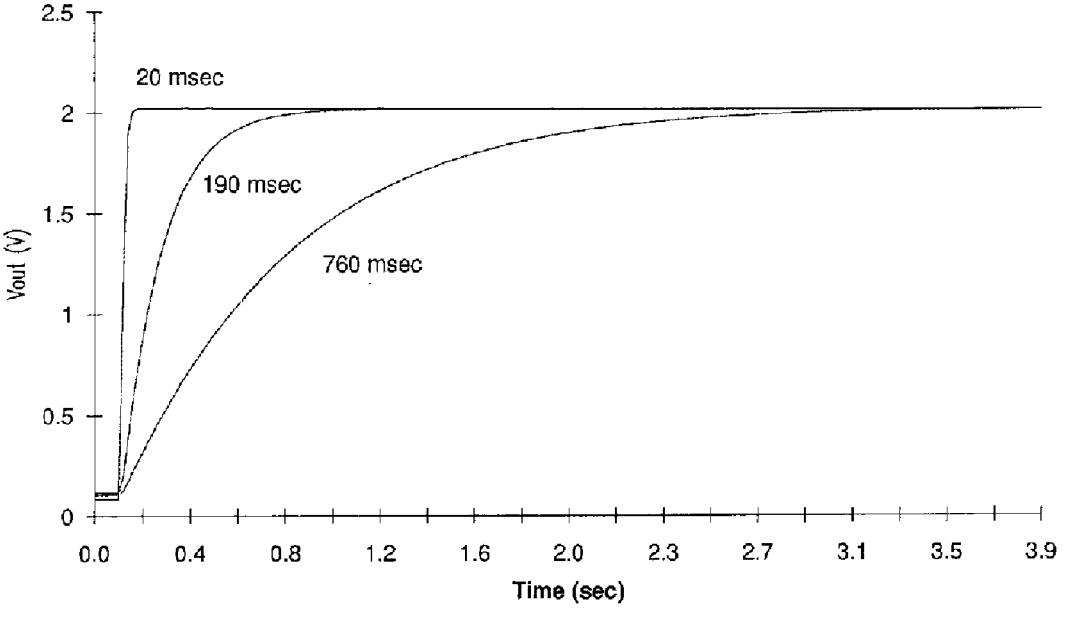


Fig. 11. Output response to a constant step voltage for three different time constants.

8. Evaluation and Comparison

8.1. Evaluation of Optimization

For the optimization procedure a few simplifications were made, concerning the circuit layout and integration. A simplified layout for the current dividers was used, the switches for the capacitor selection were omitted and cascode transistors were excluded. These simplifications introduce certain deviations between the optimal circuit and the real circuit. In table 4 the optimization results are compared with the real circuit implementation.

The big difference for the R_{area} value can be explained from the simplification made for the current mirror layout. The use of a centroid layout and of a

Table 4. Comparison optimization results with actual circuit implementation.

	Optimization	Layout
R_{area}	70 kΩ/μm²	37 kΩ/μm²
C_{area}	0.8 fF/μm²	0.6 fF/μm²
A_1	2000 μm²	70,000 μm²
Total area	0.2 mm²	0.4 mm²

series connection for the long and a parallel connection for the wide mirror transistor introduce an area overhead of about the same size as the simple current mirror implementation itself. This results in an area efficiency reduction of about 2.

The difference for the C_{area} value is due to the area overhead of the switches and of extra area needed for splitting the capacitor in a parallel connection. For the bias circuit (A_1) the omission of cascode transistors introduce also a considerable deviation from the optimized values.

The real total area is the double of the optimized value due to all the above variations in parameters. The area of the current dividers is about 1.4 times the area of the capacitors because the deviation of the R_{area} is bigger than for the C_{area} parameter.

When the optimization procedure is repeated with the actual R_{area} and C_{area} values for a 500-ms time constant, an optimal resistor of 5400 MΩ and an optimal capacitor of 93 pF are to be used. The total area is then 0.3 mm².

It can be concluded that the optimization procedure yielded a good indication for the optimal design parameters and that based on the actual values next designs will be more area efficient.

8.2. *Comparison with Other Realizations*

The CMOS circuit is compared with other very low frequent low-pass filter realizations from literature. To make a correct comparison of these circuits with different cutoff frequencies (f_c), quality factors for the area efficiency and for the offset sensitivity are defined.

From equations (2) and (9) for $A_{TOT_{opt}}$ it is clear that the area of a circuit for large time constants is proportional to the square root of the time constant and thus inversely proportional to the square root of the cutoff frequency. Consequently, the factor Area $\cdot \sqrt{f_c}$ expresses the area efficiency of the implementation independent of the cutoff frequency.

The offset of a time constant circuit grows linearly with the amount of electronic multiplication used as can be seen from equations (4) and (5). For larger time constants (or smaller cutoff frequencies), larger multiplication factors have to be used with larger offset voltages as a result. The factor $V_{os}f_c$ is thus a measure for the sensitivity of the circuit to offset voltage problems.

The results are summarized in table 5. The presented design is compared to two very low frequency low-pass filters from literature. The filter of [6] is a BICMOS circuit. Bipolar transconductance circuits (I_{BIAS} = 1 nA) are used as resistances and electronic multiplication is used to increase the time constant.

The second design is a swtiched capacitor filter using a T-cell structure. Capacitive T-cells result in the most compact switched capacitor integrators with very large time constants [3]. The filter is realized with a 5-μm CMOS process. It was designed for a clock frequency of 100 kHz and a cutoff frequency of 22 Hz. Using the minimal clock frequency, which is limited to 10 kHz by leakage currents, a cutoff frequency of 2.2 Hz is obtained. The active area of the chip does *not* include the areas of the antialiasing filters and of the clock oscillators.

The presented CMOS design has approximately one order of magnitude smaller chip area. This confirms our approach of total circuit area minimization instead of capacitor area minimization. Also the offset voltage is one order of magnitude smaller for the same cutoff frequency. The special care taken to concentrate the offset generating error sources in only one circuit branch is justified by this result.

8.3. *Useful Time Constant Range for Presented Circuit*

The constant term A_1 in the equation for $A_{TOT_{opt}}$ (9) implies that the OTA-R time constant circuit is only area efficient for "larger" time constants, because the area

Table 5. Measured performances of different designs for large time constants.

Structure	f_c (Hz)	Active Area (mm²)	V_{offset} (mV)	$f_c \cdot V_{offset}$ (Hz · mV)	Area $\cdot \sqrt{f_c}$ (mm² · √Hz)
Presented (CMOS)	0.23	0.5	130	29	0.12
Deguelle [6] (Bipolar)	10	0.4	20	200	1.25
T-cell [3] (switched capacitor)	2.3	0.8	100	230	1.21

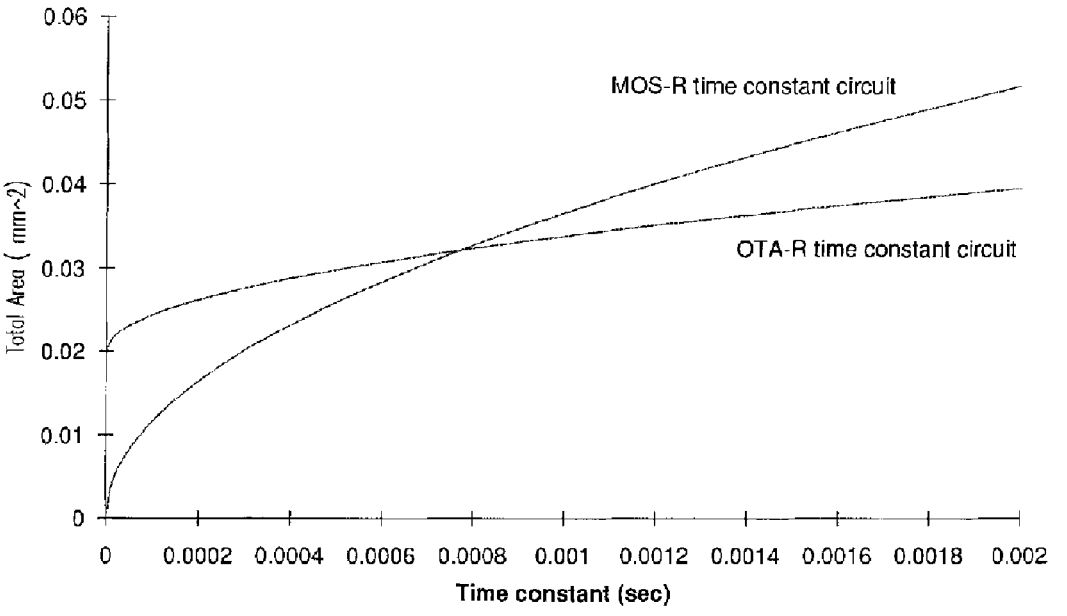


Fig. 12. Comparison of the area of a MOS-R time constant circuit and an OTA-R time constant circuit as a function of the realized time constant.

of a MOS-R or diffusion resistor time constant circuit has no constant term (cf. equation (2)). In figure 12 the area as a function of the time constant is displayed for an OTA-R and a MOS-R circuit. In our process, the OTA-R is more efficient for time constants larger than about 1 ms. A diffusion resistor circuit has a larger area than a MOS-R circuit. It is always less area efficient than a MOS-R circuit and only more efficient than an OTA-R circuit for time constants smaller than 4 μs. Of course these limit values depend on the value of A_1 and thus process parameters and the bias circuit used (cf. Appendix II).

9. General Conclusions

In this paper a new circuit is presented for the full on-chip integration of large time constants. The circuit is used in a neural net implementation so the primary design objective is area minimization. High accuracy is only of secondary importance. The presented circuit includes a new multiplication scheme which has several interesting advantages. A large signal range is obtained. The error sources introducing offset voltages are all localized in only one node, and the resulting offset voltages are very acceptable when compared to previously realized circuits. Moreover, due to the simplicity of the design a good area optimization can be performed which results in a small circuit for the total circuit. The design was fabricated in a 3-μm CMOS technology and the measurements agree well with the expectations.

10. Appendix I: Calculation of the R_{area} Values

In this appendix the calculation of the R_{area} values (resistance per unit area $\Omega/\mu m^2$) for the different resistor circuits is done and some indicating values are given.

10.1. Diffusion Resistor

The resistance values of diffusion resistors are commonly expressed in terms of sheet resistance ($R_\square$). When the minimal width (W_{min}) is used, the expression for R_{area} is

$$R_{area} = \frac{R_\square}{W_{min}^2}$$

In our 3-μm CMOS n-well process a well resistor has the highest R_{area}—50 $\Omega/\mu m^2$.

10.2 MOS Transistor in Linear Region

A linear approximation for the resistance of a MOS transistor biased in its linear region, is derived in [10]:

$$R = \left(\frac{L}{W}\right) \frac{1}{\mu C_{ox}(V_{gate} - V_T)}$$

In order to keep the transistor in its linear region for input and output voltages ranging from 0 to V_{max}, the $V_G - V_T$ must be larger than V_{max}. Using a minimal width W_{min}, the R_{area} is

$$R_{\text{area}} = \frac{1}{\mu C_{\text{ox}} V_{\max} W_{\min}^2}$$

This is an overestimation of the area efficiency since the area of the nonlinearity canceling circuitry [10] is not included in this approximation.

10.3 OTA Resistor

For an OTA resistor the equivalent resistance is equal to $1/gm$ of the input transistors. The $V_{GS} - V_T$ of the input transistors must be set to $V_{\max}$ for a signal range of 0–$V_{\max}$. The resistance of an OTA-R is then

$$R = \left(\frac{L}{W}\right) \frac{1}{\mu C_{\text{ox}} V_{\max}}$$

which is the same as for a MOS resistor. A first approximation for the area efficiency is

$$R_{\text{area}} = \frac{1}{2\mu C_{\text{ox}} V_{\max} W_{\min}^2}$$

which is half of the area efficiency of a MOS resistor. Here only the area of the two input transistors is taken into account, so that the area efficiency is overestimated.

For the 3-μm CMOS n-well process the values for the area efficiencies and the process parameters are given in table 6.

Table 6. Resistance area efficiencies for a 3-μm n-well CMOS process.

Process Parameters		Area Efficiencies	
$W_{\min}$	3 μm	Diffusion resistor	50 $\Omega/\mu m^2$
μC_{ox} (pMOS)	15 $\mu A/V^2$	MOS resistor	2500 $\Omega/\mu m^2$
$V_{\max}$	3 V	OTA resistor	1250 $\Omega/\mu m^2$

11. Appendix II: Expressions for the Area of an OTA-R

In this appendix the formula for the area of the OTA-R as a function of its resistance is derived. For these calculations following constraints are defined:

- All transistors are operated in strong inversion, i.e., $(V_{GS} - V_T)_{\min} = 200$ mV.
- The W/L of the input transistors is defined by the $V_{\max}$ of the signals, i.e., $(V_{GS} - V_T)_{\text{DP}} = V_{\max}$.
- The equivalent resistance of the OTA-R is

$$R = B \frac{V_{\max}}{I_{\text{BIAS}}}$$

First an expression for the area of an individual transistor is given. For a given resistance the I_{BIAS} and the B factor have to be chosen optimally. An expression for the total area of the OTA-R, with I_{BIAS} and B as a parameter, can easily be determined. Then the optimal bias current for the OTA-R is calculated.

11.1. Expression for the Area of One Transistor

The area of a transistor is approximated by the product of its width (W) and length (L). For a given transistor current, the W/L value is determined by

$$\frac{W}{L} = \frac{I_{DS}}{(\mu C_{\text{ox}}/2)(V_{GS} - V_T)^2}$$

Two possibilities can occur:

$W/L > 1$: a wide transistor with a minimal length;
$W/L < 1$: a long transistor with a minimal width.

This implies that depending on the current a minimal width or a minimal length transistor is used and that a general expression for the transistor area contains a conditional. To simplify the expressions some constants are defined:

$$I_N = \frac{\mu_{\text{NMOS}} C_{\text{ox}}}{2} (V_{GS} - V_T)_{\min}^2$$

$$I_P = \frac{\mu_{\text{PMOS}} C_{\text{ox}}}{2} (V_{GS} - V_T)_{\min}^2$$

$$I_{\text{DP}} = \frac{\mu_{\text{NMOS}} C_{\text{ox}}}{2} (V_{GS} - V_T)_{\text{DP}}^2$$

The area of an NMOS transistor (NTOR) can be written as ([condition] is 0 if condition is false and [condition] is 1 if condition is true)

$$\text{Area}_{\text{NTOR}}(I_{DS}) = [I_{DS} \geq I_N] \frac{I_{DS}}{I_N} L_{\min}^2 + [I_{DS} < I_N] \frac{I_N}{I_{DS}} W_{\min}^2$$

with $L_{\min}$ and $W_{\min}$ the minimal transistor length and width. For the PMOS transistors (PTOR) and for the input transistors (DP) analog expressions ($\text{Area}_{\text{PTOR}}(I_{\text{DS}})$, $\text{Area}_{\text{DP}}(I_{\text{DS}})$) are valid using respectively the constants I_P and I_{DP}.

11.2. Expression for the Total Area of the OTA-R

The area of the individual transistors is under the above constraints only a function of their current. Two situa-

tions occur in the OTA-R (figure 3a). Transistors $M1$, $M2$, $M3$, $M4$, $M5$, and $M7$ draw a current of $I_{BIAS}/2$. Their area can be expressed as a function of the bias current only and is represented by the term $A_{11}(I_{BIAS})$. The area of the bias circuitry can also be included in this term. The transistors $M6$ and $M8$ have a current of $I_{BIAS}/2B$. Using the equation of the equivalent resistance, their area can be expressed as $A_{22}(R)$. The total area of the OTA-R is then

$$\text{Area}_{\text{OTA-R}} = A_{11}(I_{BIAS}) + A_{22}(R)$$

From this equation it is concluded that the optimal bias current, and thus the optimal sizing of $M1$–$M5$, $M7$ and the biasing circuitry, is independent of the resistance and is only determined by process parameters. From the expression $A_{11}(I_{BIAS})$ the optimal bias current is calculated, which is process dependent. For our process the optimal current is about 15 μA.

The sizes of the transistors $M6$ and $M8$ are fully determined by the desired resistance value. As a first approximation—the error introduced by this approximation is evaluated in Section 8—it can be assumed that the current dividers are implemented as simple current mirrors so that the equations for the transistor area remain valid. For the given bias current and for the desired time constants $M6$ and $M8$ are long transistors. The area of the OTA-R is then

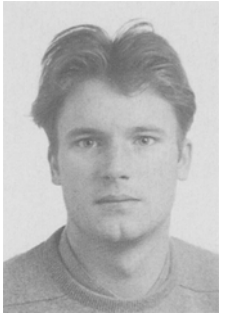
$$\text{Area}_{\text{OTA-R}} = A_1 + \frac{R}{R_{\text{area}}}$$

with

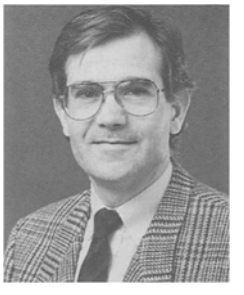
$$R_{\text{area}} = \frac{V_{\max}}{2W_{\min}^2(I_N + I_P)} \quad \text{and} \quad A_1 = A_{11}(I_{BIASopt})$$

References

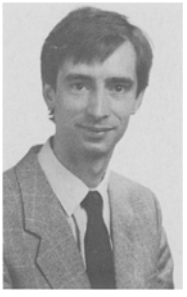
1. J. Van der Spiegel, P. Mueller, D. Blackman, P. Chance, C. Donham, R. Etienne-Cummings, and P. Kinget, "An analog neural computer with modular architecture for real-time dynamic computations," *IEEE J. Solid-State Circuits*, Vol. SC-27, pp. 82–92, 1992.
2. M. Steyaert, P. Kinget, W. Sansen, and J. Van der Spiegel, "Full integration of extremely large time constants in CMOS," *Electron. Lett.*, Vol. 27, pp. 790–791, 1991.
3. P. Van Peteghem, "An area-efficient approach to the design of very-large time constants in switched-capacitor integrators," *IEEE J. Solid-State Circuits*, Vol. SC-19, pp. 772–780, 1984.
4. A. Grebene, *Bipolar and MOS Analog Integrated Circuit Design*, Wiley: New York, 1984.
5. P.R. Gray and R.G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley: New York, 1984.
6. W.H.G. Deguelle, "Limitations on the integration of analog filters for frequencies below 10 Hz," in *Proc. IEEE-ESSCIRC*, pp. 131–134, 1988.
7. L.J. Stotts, "Introduction to implantable biomedical IC design," *IEEE Circuits Devices Mag.*, pp. 12–18, Jan. 1989.
8. P. Garde, "Transconductance cancellation for operational amplifiers," *IEEE J. Solid-State Circuits*, Vol. SC-12, pp. 310–311, 1977.
9. J.C. Schmoock, "An input stage transconductance reduction technique for high-slew rate operational amplifiers," *IEEE J. Solid-State Circuits*, Vol. SC-10, pp. 407–411, 1975.
10. Y. Tsividis, M. Banu, and J. Khoury, "Continuous-time MOSFET-C filters in VLSI," *IEEE Trans. Circuits Syst.*, Vol. CAS-33, pp. 125–139, 1986.



Peter R.D. Kinget was born in Buffalo, NY, in 1967. He received the M.Sc. degree in electrical and mechanical engineering in 1990 from the Katholieke Universiteit Leuven, Belgium. Currently, he is a research assistant of the National Fund for Scientific Research, Belgium, at the ESAT-MICAS laboratory, Katholieke Universiteit Leuven. He is working toward a Ph.D. degree on analog VLSI implementation of neural networks.



Jan Van der Spiegel received his engineering degree in electromechanical engineering, and a Ph.D. degree in electrical engineering from the Katholieke Universiteit of Leuven, Belgium, in 1974 and 1979, respectively. From 1980 to 1981, he was a postdoctoral fellow at the University of Pennsylvania, where he became an assistant professor of electrical engineering in 1981. Currently, he is an associate professor of electrical engineering and of material science and engineering at the University of Pennsylvania. He holds the Bicentennial Chair of the Class of 1940, and is director of the Center for Sensor Technologies at the University of Pennsylvania. Professor Van der Spiegel's research interests are in artificial neural networks, CCD sensors, microsensor technology, and low-noise, low-power, high-speed analog integrated circuits. He is editor for North and South America of *Sensors and Actuators*, and is on the editorial board of the *International Journal of High Speed Electronics*. He is a senior member of the IEEE, a member of the International Neural Network Society, Phi Beta Delta and Tau Beta Pi.



Michel S.J. Steyaert was born in Aalst, Belgium, in 1959. He received his Ph.D. degree in electrical and mechanical engineering from the Katholicke Universiteit Leuven, Heverlee, Belgium in 1987. From 1983 to 1986 he obtained an IWNOL fellowship which allowed him to work as a research assistant at the Laboratory ESAT-K.U. Leuven. In 1987 he was responsible for several industrial projects in the field of analog micropower circuits at the Laboratory ESAT-K.U. Leuven as an IWONL project researcher. In 1988, he was a visiting assistant professor at the University of California, Los Angeles. Since 1989 he has been appointed as a NFWO research associate at the Laboratory ESAT–K.U. Leuven, where he has been an associate professor since 1990. His current research interests are in high-frequency analog integrated circuits for telecommunications and integrated circuits for biomedical purposes.