Accurate design of a MOS-based resistive network for time-controlled diffusion filtering

Jorge Fernández-Berni, Ricardo Carmona-Galán
Institute of Microelectronics of Seville (IMSE-CNM)
Consejo Superior de Investigaciones Científicas y Universidad de Sevilla
C/ Américo Vespucio s/n, 41092, Seville, Spain
Email: berni@imse.cnm.es

***Abstract*—This paper analyses a MOS-based resistive network suitable for massively parallel image processing. The inclusion of MOS transistors biased in the ohmic region instead of true resistors permits certain control over the underlying spatial filtering while reducing the required area for VLSI implementation. However, it also leads to nonlinearities and thereby to errors with respect to an ideal resistive grid. By studying an elementary network composed of only two nodes we determine the guidelines to be followed in order to minimize the error according to the selected signal range. These guidelines are then extrapolated to larger networks demonstrating that pretty accurate networks can be achieved even for relatively wide signal ranges. Simulations are employed to validate the extrapolated results. The numerical examples will also allow to visualize how the insertion of MOS transistors affects the spatial filtering carried out by the grid.**

I. Introduction

Resistive networks present very good features for power-efficient focal-plane image filtering [1]. They are passive networks performing massively parallel processing from certain initial conditions [2]. Their VLSI implementation thus results specially interesting for low-power vision applications with tight timing requirements. But a true linear resistive grid is difficult to implement in VLSI because resistors need large areas in CMOS. Furthermore, a simple resistor does not permit electronic control over the underlying processing.

These drawbacks can be overcome by using MOS transistors. With their symmetric drain and source terminals, they can replace the resistors one by one. MOS transistors can achieve larger resistances with less area than resistors made with polysilicon or diffusion strips. Besides, by controlling their gate voltage, it is possible to operate over the processing realized by the grid. However, the inclusion of MOS transistors also entails other problems like their intrinsic nonlinearities. Indeed the substitution of resistors by MOS transistors in resistive networks has been previously studied. In [3] linearity of currents through resistive grids is achieved by means of MOS transistors in weak inversion. The value of the resistance associated to each transistor is directly controlled by the corresponding gate voltage. This property of current linearity is also applicable even if the transistors leave weak inversion as long as all of them share the same gate voltage [4]. In this case the resistance of the transistors would not already be electronically tunable but determined by their geometry. On the contrary, linearity is not so easy to reach when signals are encoded by voltages. It is the case of the resistive network depicted in Fig. 1 where a spatial filtering is realized along time over the initial voltages at the capacitors. The use of MOSFETs operating in the ohmic region instead of resistors is apparently the most simple option [5]. However the intrinsic nonlinearity in the $I-V$ characteristic has leaded to more ellaborated alternatives in which the nonlinear term is cancelled [6] or to transconductor-based implementations [7].

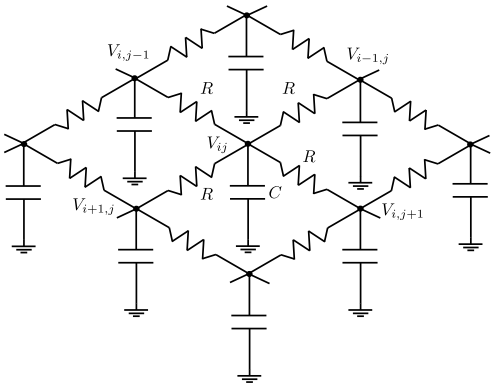

Fig. 1. Resistive network for spatial filtering

In this paper we demonstrate that, for moderate requirements of accuracy, the resistive network in Fig. 1 can be implemented by replacing every resistor by a MOS transistor biased in the ohmic region despite the unavoidable nonlinearity of its $I-V$ characteristic. If the equivalent resistance which the MOSFET network is compared to is adequately chosen, we will see that errors are under a reasonably low limit.

II. Spatial filtering by a resistor network

In order to define the spatial filtering performed by the grid in Fig. 1, let us consider the initial voltage at the capacitor of every node as the value of the corresponding pixel. If we permit the network to evolve from this initial state, the equation satisfied at each node is:

$$\tau\frac{dV_{ij}}{dt} = -4V_{ij} + V_{i+1,j} + V_{i-1,j} + \\ +V_{i,j+1} + V_{i,j-1} \quad (1)$$

where $\tau = RC$. Applying the DFT to this equation we obtain:

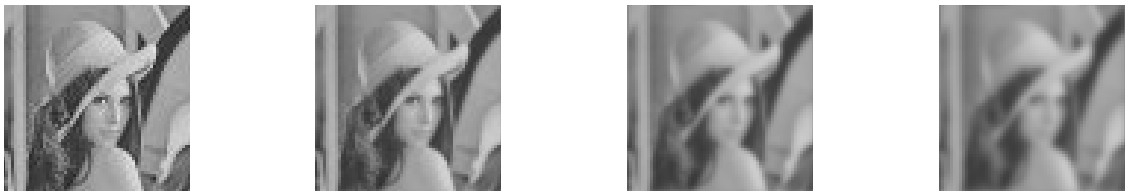

Fig. 2. Original image (64 × 64 px) and outcome of applying the filtering defined by Eq. (5) for $t = 50$ns, $t = 150$ns and $t = 250$ns with $\tau = 300$ns, respectively

$$\tau\frac{d\hat{V}_{uv}}{dt} = -4\hat{V}_{uv} + e^{\frac{2\pi iu}{M}}\hat{V}_{uv} + e^{\frac{-2\pi iu}{M}}\hat{V}_{uv} + \\ + e^{\frac{2\pi iv}{N}}\hat{V}_{uv} + e^{\frac{-2\pi iv}{N}}\hat{V}_{uv} \quad (2)$$

where we have considered an array whose size is $M \times N$ pixels. Eq. (2) can be rewritten as:

$$\tau\frac{d\hat{V}_{uv}}{dt} = [-4 + 2cos(\frac{2\pi u}{M}) + 2cos(\frac{2\pi v}{N})]\hat{V}_{uv} \quad (3)$$

and solving now in the time domain we obtain:

$$\hat{V}_{uv}(t) = \hat{V}_{uv}(0)e^{\frac{2t}{\tau}[cos(\frac{2\pi u}{M})+cos(\frac{2\pi v}{N})-2]} \quad (4)$$

where $\hat{V}_{uv}(0)$ represents the DFT of the image defined by the initial voltages at the capacitors and $\hat{V}_{uv}(t)$ is the DFT of the image defined by those same node voltages after a certain time interval t since the network started to evolve. A transfer function can be defined as follows:

$$\hat{H}_{uv}(t) = \frac{\hat{V}_{uv}(t)}{\hat{V}_{uv}(0)} = e^{\frac{2t}{\tau}[cos(\frac{2\pi u}{M})+cos(\frac{2\pi v}{N})-2]} \quad (5)$$

which describes the filtering process undergone by the initial image as the network evolves. As an example, consider the Fig. 2 where we have applied this filtering for $t = 50$ns, $t = 150$ns, $t = 250$ns with $\tau = 300$ns.

III. Error on a 2-node MOS-based resistive grid

To analyse the error committed when resistors are replaced by MOS transistors biased in the triode region, we are going to thoroughly compare the circuits in Fig. 3. They represent a 2-node ideal resistive grid and its corresponding MOS-based implementation. For purposes of clarity, we will confine the analysis to n-channel MOS transistors, but it applies equally well for p-channel transistors. The gate voltage V_G is fixed and we will assume, without loss of generality, that the initial conditions of the capacitors fulfill $V_1(0) > V_2(0)$, being $V_1(0) = V_1'(0)$ and $V_2(0) = V_2'(0)$. We also assume that the transistor is biased in the triode region for any voltage at the nodes, with:

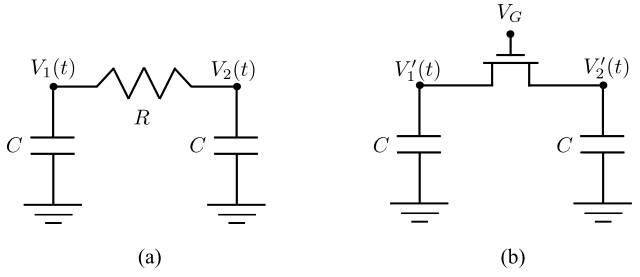

Fig. 3. 2-node ideal resistive grid (a) and its MOS-based implementation (b)

$$\begin{cases} V_1'(t) \in [V_{min}, V_{max}] \\ V_2'(t) \in [V_{min}, V_{max}] \end{cases} \quad (6)$$

The error committed at each node of the MOS-based implementation with respect to the ideal grid after a certain time interval t since the moment in which both networks started to evolve can be expressed as:

$$e(t) = \frac{1}{C}\int_0^t \left[\frac{V_1(\zeta) - V_2(\zeta)}{R} - \frac{V_1'(\zeta) - V_2'(\zeta)}{R_M(\zeta)}\right] d\zeta \quad (7)$$

where R_M represents the instantaneous resistance associated to the transistor while the MOS-based network is evolving. Thus we can write that:

$$\begin{cases} V_1'(t) = V_1(t) + e(t) \\ V_2'(t) = V_2(t) - e(t) \end{cases} \quad (8)$$

The equivalent differential form of Eq. (7) is:

$$\tau\frac{de}{dt} = [V_1'(t) - V_2'(t)]\left[1 - \frac{R}{R_M(t)}\right] - 2e(t) \quad (9)$$

where we have applied the relations defined by Eq. (8). As $e(0) = 0$ and $e(\infty) = 0$, by making $de/dt = 0$ we obtain the extreme of $e(t)$:

$$e_{ext} = \frac{1}{2}(V_1' - V_2')\left(1 - \frac{R}{R_M}\right) \quad (10)$$

which depends on the value of V_1', V_2' and R_M at the time instant in which such an extreme is reached. Besides, it depends on the value of the resistance R that is the elementary resistance in the linear network which we are comparing the MOSFET network behaviour. Our objective is therefore to determine the resistance R which minimizes $\max|e_{ext}|$. In other words, we want to find the ideal resistive grid which is emulated with minimum error by the MOS-based grid within the signal range established by Eq. (6). This will permit us to

build the design equations for a MOS-based resistive network that emulates a resistor grid with minimum error.

As a first step to achieve this objective, we are going to determine the instantaneous resistance of the MOSFET. The drain current of the transistor in Fig. 3(b) is [8]:

$$I_D(t) = \beta\{2\left[V_G - V_2'(t) - V_{T_n}\right]\left[V_1'(t) - V_2'(t)\right] - \\ - \left[V_1'(t) - V_2'(t)\right]^2\} \quad (11)$$

where $\beta = (1/2)\mu_n C_{ox}'(W/L)$ and V_{T_n} represents the threshold voltage. By definition $R_M(t) = \left[V_1'(t) - V_2'(t)\right]/I_D(t)$ what, from Eq. (11), leads to:

$$R_M(t) = \frac{1}{\beta\left\{V_C - \left[V_1'(t) + V_2'(t)\right]\right\}} \quad (12)$$

where $V_C = 2(V_G - V_{tn})$. We are neglecting the body effect at this point in order to simplify the reasoning. The defined V_C is therefore supposed to be constant. As can be seen, this resistance depends on the node voltages and thus varies constantly during the evolution of the diffusion. The variation range of $R_M(t)$, according to Eq. (6), is:

$$R_M(t) \in \left[\frac{1}{\beta(V_C - 2V_{min})}, \frac{1}{\beta(V_C - 2V_{max})}\right] \quad (13)$$

Note that this range also determines the possible values of the equivalent resistance R implemented by the transistor. It would not make any sense to compare the MOS-based grid with an ideal resistive grid where the value of the resistor is never reached by the transistor throughout the signal range. In this way, we say that:

$$R = \frac{1}{\beta(V_C - V_E)} \quad (14)$$

where V_E must be that value within the interval $[2V_{min}, 2V_{max}]$ defining an equivalent resistance R which minimizes the error committed by the MOS-based resistive grid when compared to an ideal grid with such an elementary resistance.

Inserting the previous definitions of R_M and R into Eq. (10), we obtain the following function:

$$e_{ext} = \frac{1}{2}(V_1' - V_2')\left(\frac{V_1' + V_2' - V_E}{V_C - V_E}\right) \quad (15)$$

where V_C, V_E, and the signal range of V_1' and V_2' are design parameters. Notice that the larger V_C, that is, the larger the gate voltage, the smaller $|e_{ext}|$. It makes sense as, according to Eq. (11), the larger the gate voltage, the less the influence of the nonlinearity of the transistor and therefore the better the behaviour of the MOSFET as a resistor. For the variables V_1' and V_2' there are no local extremes of e_{ext}. There is only a saddle point. However, taking into account that both V_1' and V_2' belong to the interval $[V_{min}, V_{max}]$, the absolute maximum and minimum of e_{ext} can be calculated. And the maximum absolute error is the maximum of the absolute values of these absolute extremes:

$$\max|e_{ext}| = \max\left[\frac{1}{8}\frac{(2V_{max} - V_E)^2}{V_C - V_E}, \frac{1}{8}\frac{(V_E - 2V_{min})^2}{V_C - V_E}\right] \quad (16)$$

which reaches its minimum value for $V_E = V_{min} + V_{max}$, resulting in:

$$\min\left(\max|e_{ext}|\right) = \frac{1}{16}\frac{(V_{max} - V_{min})^2}{V_G - V_{T_n} - \left(\frac{V_{max}+V_{min}}{2}\right)} \quad (17)$$

It means that the minimum of the maximum absolute error, which happens for this selection of V_E, depends, once the V_G has been also selected, on both the average value of the signal range and the signal range itself. The smaller the signal range and its average value, the smaller the error. Hence, for the generic signal range defined by Eq. (6), a 2-node MOS-based resistive grid achieves the best emulation of a 2-node ideal resistive grid when the value of the resistance in this ideal grid is:

$$R = \frac{1}{\beta[V_C - (V_{min} + V_{max})]} \quad (18)$$

In other words, if we want a MOS-based network to emulate a linear resistive grid with an elementary resistance given by R, then V_G, V_{max}, V_{min} and (W/L) must be chosen to hold:

$$\left(\frac{W}{L}\right) = \frac{1}{\mu_n C_{ox}' R\left(V_G - V_{T_n} - \frac{V_{max}+V_{min}}{2}\right)} \quad (19)$$

and the upper bound for the error committed is given by Eq. (17).

Finally, a crucial point to be remarked from Eq. (15) is that, according to the simplified model of the MOSFET used, the geometry of the transistor does not affect the error. It implies that for a prescribed absolute error, a wide range of values of R can be implemented by varying (W/L).

IV. Extrapolation to larger networks

We have applied the results obtained for the 2-node case to the design of a 64×64 MOS-based resistive grid in HSPICE. The used models belong to a standard 0.35μm CMOS 3.3V process. The signal range at the nodes is [0V,1.5V], wide enough to demonstrate that the nonlinearities of the MOSFETs do not necessarily introduce too much error in the spatial filtering. The gate voltage for all the transistors is established at 3.3V. Our aim is to implement a time constant of $\tau = 300$ns as in the ideal example of Fig. 2. In principle, we have infinite combinations of R and C that make this τ. We will select a transistor of $W = 0.4\mu$m and $L = 10\mu$m. This initially arbitrary value should be employed to balance the the sizes of the transistor and the capacitor according to the process parameters. For this geometry, the HSPICE model of the transistor, with $V_D = 1.5$V and $V_S = 0$V, renders an equivalent resistance of $R = 139\text{k}\Omega$. Notice that the operating conditions of the transistor in the simulation performed to calculate this equivalent resistance hold the relation expressed

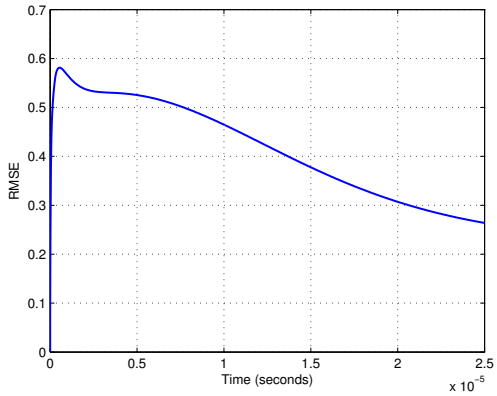

Fig. 4. RMSE of the 64×64 MOS-based grid with respect to the ideal case

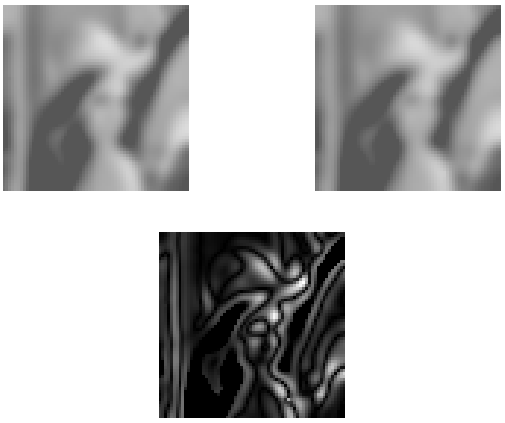

Fig. 5. Output image for an ideal resistive grid and for the NMOS grid, respectively, at the time instant in which the RMSE in the NMOS grid reaches its maximum. The last image corresponds to their difference normalized by a maximum observed error between pixels of 2.28%

in Eq. (18). Finally, by choosing a capacitance $C = 2.16$pF, we are emulating a 64×64 ideal grid with $\tau \simeq 300$ns. Let us initialize this MOS-based grid with the original image in Fig. 2. The RMSE of the pixels during the evolution of the network with respect to the ideal grid is depicted in Fig. 4. As can be seen, the maximum RMSE is below 0.6%, that is, an equivalent resolution between 6 and 7 bits. To visualize this accuracy, we have represented in Fig. 5 the output images for an ideal resistive grid and for the implemented NMOS grid at the time instant at which the RMSE in the NMOS grid reaches its maximum value. We can see that the outcome is perceptually equivalent. We have also represented their difference normalized by a maximum observed error between pixels of 2.28%. Finally, we have depicted in Fig. 6 the evolution of the RMSE in the MOS-based grid for 30 simulations where independent gaussian deviations with $\sigma = 10\%$ of W, L, μ_0 and t_{ox} from their nominal values are introduced [9]. We can see that the RMSE is always less than 0.8%. These results point out the robustness and accuracy of a MOS-based resistive grid under mismatch conditions.

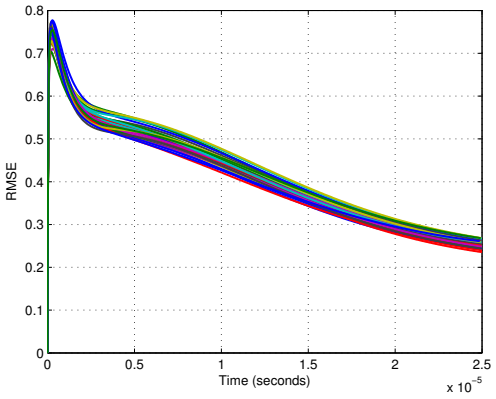

Fig. 6. RMSE for 30 simulations of the 64×64 MOS-based grid introducing independent gaussian deviations of W, L, μ_0 and t_{ox}

V. Conclusions

In this paper we have demonstrated that moderate accuracy can be achieved in MOS-based resistive grids despite the non-linearities associated to MOS transistors in the ohmic region. The key is to find the resistor which is being emulated with minimum error by the MOSFET. In the reverse formulation, finding the relation between the MOSFET parameters and the desired elementary resistor, we have the design equations for a minimum error implementation. By analysing the 2-node case we have found an analytical expression for this equivalent resistor which has been applied to the design of a larger network. An accuracy between 6 and 7 bits is achieved even under mismatch conditions.

Acknowledgment

This work is funded by CICE/JA and MICINN (Spain) through projects 2006-TIC-2352 and TEC 2006-15722 respectively

References

[1] L. Raffo, S. Sabatini, G. Bo, and G. Bisio, "Analog VLSI circuits as physical structures for perception in early visual tasks," *IEEE Trans. Neural Netw.*, vol. 9, no. 6, pp. 1483–1494, 1998.

[2] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.

[3] E. Vittoz and X. Arreguit, "Linear networks based on transistors," *Electronic Letters*, vol. 29, no. 3, pp. 297–299, 1993.

[4] K. Bult and J. Geelen, "An inherently linear and compact most-only current division technique," *IEEE J. Solid-State Circuits*, vol. 27, no. 12, pp. 1730–1735, 1992.

[5] L. Vadasz, "The use of MOS structure for the design of high value resistors in monolithic integrated circuits," *IEEE Trans. Electron Devices*, vol. 13, no. 5, pp. 459–465, 1966.

[6] H. Kobayashi, J. White, and A. Abidi, "An active resistive network for gaussian filtering of images," *IEEE J. Solid-State Circuits*, vol. 26, no. 5, pp. 738–748, 1991.

[7] K. Hui and B. Shi, "Distortion in analog networks for image filtering," *IEEE Trans. Circuits Syst. I*, vol. 46, no. 10, pp. 1161–1171, 1999.

[8] D. Johns and K. Martin, *Analog Integrated Circuit Design*. Wiley, 1996.

[9] P. Drennan and C. McAndrew, "Understanding MOSFET mismatch for analog design," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 3, pp. 450–456, 2003.