Low-leakage zero-static power consumption analogue CMOS switch

Giuseppe Sciortino,✉ Alireza Mesri, Fabio Toso, Francesco Zanetto, iD and Giorgio Ferrari iD

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, 20133, Italy

✉Email: giuseppe.sciortino@polimi.it

Fig. 1 *(a) Schematic of an analogue memory based on a pMOS switch. Cross section of (b) pMOS and (c) deep n-well nMOS transistors showing the main leakage current components*

Fig. 2 *(a) Standard switch and (b) the proposed multi-stage architecture. (c) Cross section scheme of the basic cell and corresponding layout (top view)*

The architecture of a low-leakage switch in complementary metal-oxide-semiconductor (CMOS) technology developed to implement long-term analogue memories for low-power applications is presented. The switch reduces the leakage current using a cascade of modified pass-transistors without active circuits, thus producing a negligible power consumption. The technique is simple, modular, robust, and reduces the leakage current down to tens of aA at room temperature using a TSMC 0.18 μm technology.
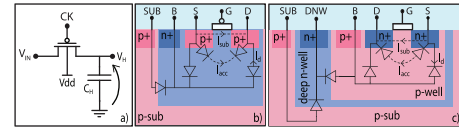
Introduction: In standard complementary metal-oxide-semiconductor (CMOS) circuits, analogue memories can be implemented by accumulating a charge on a capacitor C_H and by using a MOS switch to access it, as shown in Figure 1(a). Analogue switches with a very low leakage in OFF condition are required for an accurate and long-term storage of the analogue information. They find application in a wide range of circuits, such as low power sample and hold and switched-capacitor circuits [1–4] and synaptic memory in neuromorphic architectures [5]. In biomedical implantable devices, for example, the combination of small bandwidth of the signals and very low power consumption drives the design of slow analogue-to-digital converters and switched capacitors filters, that require the capability to hold a charge on a capacitor for tens of milliseconds with high accuracy [6].

The hold time of the analogue memory shown in Figure 1(a) is limited by the leakage current I_{OFF} of the MOS switch that changes the stored charge, giving a voltage error that grows with the retention time. A simple way to extend the maximum hold time is by increasing the C_H value, that is the stored charge. However, this solution is often unacceptable due to the increase of silicon area, the higher dynamic power consumption, and the slower write time of the analogue memory. A common technique to reduce I_{OFF} requires instead an additional amplifier properly connected to force to zero the source-drain V_{sd} and/or source-bulk V_{sb} voltages of the MOS transistor [7–10]. Although the leakage current of the MOSFET is strongly reduced to sub-fA values, both the static power consumption and the occupied area are increased by the additional amplifier, one for each low-leakage switch used in the circuit. Alternatively, the I_{OFF} can be compensated with an opposite contribution given by a replica switch with [11] or without [12] an active negative feedback network. However, the resulting leakage is still above 1 fA, limited by the matching between the MOS switch and the replica used for the leakage compensation.

Here, we present a simple approach to reduce the leakage current I_{OFF} and strongly extend the maximum retention time of the analogue memory. The technique is fully passive, that is it does not require static power consumption, compact, and without a significant penalty in the speed of the writing operation of the memory. A detailed analysis of an implementation in TSMC 0.18 μm technology is presented, as well as layout guidelines.

Main discharging mechanism in CMOS switches: The leakage current of transistors in off-state is mainly given by four mechanisms [7, 10]: sub-threshold current I_{sub}, accumulation current I_{acc}, diode reverse current I_d, and gate leakage (Figure 1(b, c)). For a technology node of 130 nm or higher, as commonly used in slow low-power analogue applications and in this work, the gate leakage is negligible [10] and it will not be further considered.

The sub-threshold current I_{sub} is a drain-source current with an exponential dependence on the gate-source voltage V_{gs} and on the drain-source voltage V_{ds}. For a nMOS transistor, it is modelled by [7]:
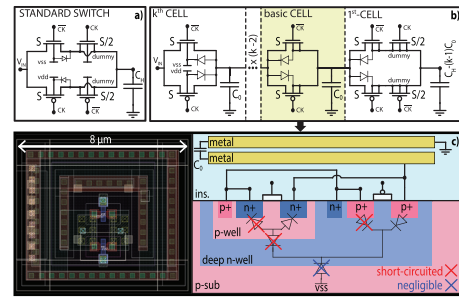
$$I_{sub} = I_0 \frac{W}{L} e^{\frac{V_{gs}-V_T}{nV_{th}}} \left(1 - e^{-\frac{V_{ds}}{V_{th}}} \right) \tag{1}$$

where I_0 is a process-dependent current, n is the sub-threshold slope, V_T is the MOS threshold voltage and V_{th} is the thermal voltage. I_{sub} is commonly the dominant contribution for a positive V_{gs} and V_{ds} greater than few mV. It is in the order of few pA for minimum size transistors at $V_{gs} = 0$ V in CMOS 0.18 μm and 0.35 μm technologies [13, 14].

The sub-threshold current is reduced by operating the nMOS transistor with a negative V_{gs} voltage [13] and it is negligible in accumulation mode ($V_{gs} \ll 0$ V). In this regime, however, the source-drain current is caused by the interaction of the two source/drain-bulk diodes in the region below the accumulation layer. This accumulation current I_{acc} is reduced to few tens of aA by keeping V_{db} equal to V_{sb} [7, 8].

Finally, an additional leakage term is given by the reverse-biased p-n junctions of the MOS transistor. By considering, as an example, a pMOS inside an n-well (Figure 1(b)), the source/drain-well diodes give the dominant leakage currents for $V_{sg} \ll 0$ V, with simulated currents of tens of fA for a minimum-size transistor. Given the high doping level of the n and p regions, the reverse current of these p-n junctions is dominated by the band-to-band tunnelling mechanism, significantly increasing the leakage with respect to the minority carrier diffusion and the electron–hole pair generation in the depletion region [15]. The low doping level of substrate avoids band-to-band tunnelling in the well-substrate junction allowing a much lower reverse current than the source/drain-well junctions. Measurements reported in [16] on a 0.13 μm CMOS process show a well-substrate junction leakage current as low as few tens of aA/μm^2. In agreement with these experimental values, the simulation models provided by TSMC for the 0.18 μm technology give a leakage current of few aA/μm^2 at room temperature.

Low-leakage switch architecture: Based on the previous discussion, the leakage current of a switched-off transistor is reduced by keeping: (i) $V_{ds} \cong 0$ V to limit the sub-threshold and accumulation-mode currents; (ii) the source/drain-well p-n junctions short-circuited. These conditions are not fulfilled in a standard switch. As an example, Figure 2(a) shows a rail-to-rail switch implemented by a transmission gate. For a general discussion, the switch includes two dummy transistors connected to the hold capacitance C_H and with half the size of the transmission gate transistors. They are commonly used to reduce the effect of charge injection during the switching by nominally extracting the same charge as the one injected into C_H [17]. During the hold time, the drain-source voltage of the MOSFETs implementing the transmission gate is not restricted to zero. Therefore, they produce significant sub-threshold and accumulation-mode currents. Moreover the source-well diodes are reversed biased with a large voltage, which is the worst condition in terms of leakage current.
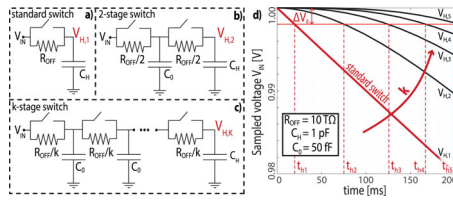
Fig. 3 *(a) Single pole, (b) double poles and (c) general multi poles simplified schemes of the proposed switches. (d) Transient of the voltage across the hold capacitor C_H as a function of the number k of stages in the switch*

Fig. 4 *Improvement factor of the maximum hold time of the proposed architecture in comparison to the standard switch as a function of the desired relative error (bottom axis) or equivalent number of bits (top axis)*

The proposed low-leakage switch adds a cascade of k-1 modified transmission gates, each one with a hold capacitor C_0, as shown in Figure 2(b). The hold capacitance of the proposed switch is reduced to $C_{H,k} = C_H - (k-1)C_0$ in order to have the same total capacitance of the standard switch. Excluding the transistors connected to the input voltage, all the others have the well and source terminals short-circuited, thus cancelling the reverse current of their diodes. During the writing/sampling phase of the memory, the C_0 capacitors and $C_{H,k}$ are all charged to the same voltage. Therefore, at the beginning of the hold phase all the transistors, except those of the first stage connected to the input voltage, have $V_{ds} = 0$ V, minimizing the sub-threshold and accumulation-mode currents. Since the drain, source and well are initially kept to the same voltage, the reverse current of the drain-well diodes is minimized as well.

The C_0 capacitor of the k-th stage is the first to be discharged due to the leakage current of the transistors connected to the input voltage V_{IN}, that in general have a $V_{ds} \neq 0$ V. Moreover, the wells of the nMOS and pMOS in this k-th stage show a leakage current due to the respective connections to V_{ss} and V_{dd}, required to prevent a forward biased drain-well junction. When the voltage drop on the C_0 of the k-th stage reaches few tens of mV, a significant leakage current arises in the second transmission gate, discharging the C_0 capacitor of the next stage. The discharge of the hold capacitor $C_{H,k}$ will approximately start only when all the C_0 capacitors of the previous $k-1$ stages have been sufficiently discharged, with a significant delay with respect to the case of a standard switch and a corresponding increase of the maximum hold time for a given voltage error. Note that the leakage current of the well-substrate diode is the only component unaffected by the proposed architecture. However, it is order of magnitude smaller than the other leakage currents, as previously discussed.

The advantages offered by the proposed cascade of modified switches can be analysed by referring to the simplified schematics shown in Figure 3. By neglecting the well-substrate diode, the leakage current I_{OFF} of an open switch is modelled by a simple resistor R_{OFF} between the source and drain terminals. A cascade of identical stages would increase the total resistance and capacitance of the switch with a beneficial effect on the hold time and a detrimental effect on the sampling time. To limit these effects and better understand the advantages of the proposed architecture, the theoretical analysis considers the total resistance and capacitance of the switch independent of the number k of cascaded stages. This is obtained by selecting each stage with $R_{OFF,k} = R_{OFF}/k$, where R_{OFF} is the resistance of the standard single-cell switch (Figure 3(a)), the same hold capacitance C_H for all structures and $(k-1)C_0 \ll C_H$. Given these conditions, the dominant time constant of the network is independent of k and approximately equal to:

$$\tau_k \approx C_H \cdot \sum_{i=1}^{k} R_{OFF,k} = C_H \cdot R_{OFF} = \tau_L \quad (2)$$

The same approximation is valid when the switch is turned on to charge the hold capacitance C_H and leads to a dominant time constant for the sampling phase $\tau_C = C_H R_{ON}$, where R_{ON} is the overall resistance when the switch is closed. Single and multi-pole RC circuits with the same dominant pole have similar step response for an observation time much longer than the dominant time constant. Thus, the time required to fully charge the hold capacitance during the sampling phase, much longer than τ_C, is approximately independent of the number of cascaded stages.

On the contrary, the maximum hold time is determined by the transient of the stored voltage on a time scale much shorter than τ_L, where the effect of the non-dominant poles is not negligible. Indeed, the step
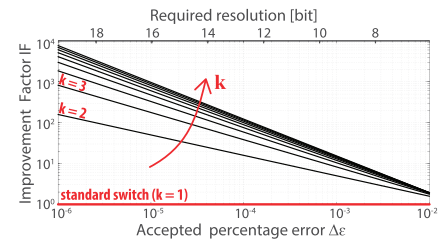
response of a RC network with k real poles and no other singularities has a zero derivative up to the order k-1 at the starting time of the transient. Consequently, the initial transient is drastically slowed down by the increase of the number of poles, as it is obtained with the proposed multi-stage structure.

This is shown in Figure 3(d), that reports the discharge transient in the hold phase of the standard switch (Figure 3(a)) and of the proposed modified switches with 2 (Figure 3(b)), 3, 4 and 5 cascaded stages. The standard switch is a single pole system characterized by a linear discharge for $t \ll \tau_L$ with a slope of $V_{H,1}(0)/\tau_L$, where $V_{H,1}(0)$ is the initial voltage sampled on the hold capacitor. By cascading two stages, the circuit has two poles, $\tau_0 \approx \tau_L$ and $\tau_1 \approx C_0 R_{OFF,2}/2$, therefore the system response has zero time derivative at $t = 0$ and evolves quadratically in time for $t \ll \tau_1$. Fixed a small enough discharge error ΔV_ϵ, an increase in the maximum allowed hold time is evident, as shown in Figure 3(d).

The time required to reach a given percentage error $\Delta\epsilon = \Delta V_\epsilon/V_H(0)$ on the stored voltage as a function of the number k of RC stages is approximately given by:

$$t_{h,k} \approx \left(\Delta\epsilon \cdot k! \cdot \frac{R_{OFF}}{k} \cdot C_H \right)^{1/k} \left(\frac{R_{OFF}}{k} C_0 \right)^{(k-1)/k} \quad (3)$$

This has to be compared with the time required by a standard single-stage switch to reach the same error:

$$t_{h,1} \approx \Delta\epsilon \cdot R_{OFF} \cdot C_H \quad (4)$$

By combining Equations (3) and (4), we can define an improvement factor (IF) of the multi-pole solution in comparison to the standard switch:

$$IF = \frac{t_{h,k}}{t_{h,1}} = (k-1)!^{1/k} \left(\frac{C_0}{k\Delta\epsilon C_H} \right)^{(k-1)/k} \quad (5)$$

Figure 4 shows IF as a function of the hold phase accuracy expressed in terms of percentage error $\Delta\epsilon$. The plot refers to a switch with $C_H = 1$ pF and a small C_0 of 50 fF. The hold time can be increased of 2 orders of magnitude with respect to the simple switch just introducing two stages ($k = 2$). The improved switch is particularly suitable for high accuracy analogue circuits, as the hold time can be increased of orders of magnitude without losing resolution. Note that IF is independent of R_{OFF} in our simplified analysis. This suggests an increase of the hold time mainly given by the architecture of the switch, with a limited sensitivity to the CMOS process parameters that are summarized in R_{OFF}.

Simulation results: The circuits in Figure 2 have been designed using the TSMC 0.18 μm CMOS process operated at 1.8 V with a hold capacitor C_H of 1 pF and a capacitor C_0 of 50 fF in each basic cell. Dimension of the transmission gates transistors are $S = 0.5\,\mu\text{m}/0.2\,\mu\text{m}$, whereas the dummy transistors are $S/2 = 0.25\,\mu\text{m}/0.2\,\mu\text{m}$. The basic cell uses a limited area of $8 \times 8\,\mu\text{m}^2$, about one-eighth of C_H, thanks to the placement of C_0 on top of the transistors, made possible by the use of a metal-insulator-metal (MIM) capacitor, as shown in Figure 2(c). Several implementations of switches with different number k of identical stages have been designed and each one has been compared with the standard switch of Figure 2(a).

Figure 5(a) shows the variation of the stored voltage, after 1 s hold time, as a function of the initial input voltage V_{IN} for a standard switch and for the proposed multi-stage architecture. As expected, the discharging error during the hold phase drastically decreases by just introducing few stages. The improvement is particularly large for sampling voltages in the ± 0.6 V range, where the dominant discharging contribution for
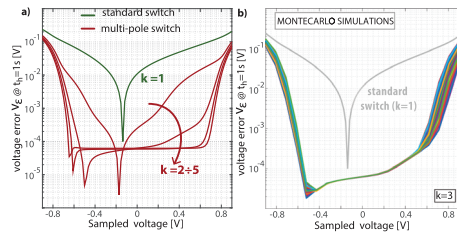
Fig. 5 *(a) Voltage error V_ϵ on the capacitance C_H after 1 s hold time as a function of the sampled voltage. V_{IN} is fixed to 0 V during the hold phase. (b) Montecarlo simulation (100 runs) of V_ϵ for a three-stage switch architecture*
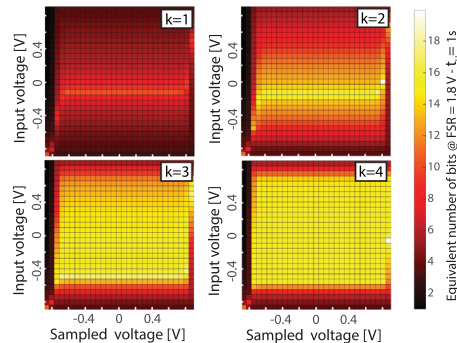
Fig. 6 *Voltage error after 1 s of hold time expressed as equivalent number of bits referred to the 1.8 V input rail, as a function of the voltage on C_H and of the input voltage during the hold phase. The simulations have been performed for the standard switch ($k = 1$) and for the improved architecture with $k = 2, 3, 4$*

the standard switch is given by the reverse current of the highly doped source/drain-well p-n junctions. The proposed architecture reduces the error V_ϵ down to a minimum value given by the well-substrate reverse leakage current. A sampled voltage near to the power supply value unavoidably reduces the gate-source voltages of the transistors, which increases the sub-threshold leakage current. The simulations are performed at a temperature of 300 K. By increasing the temperature all the leakage currents of the switches increase, in particular the well-substrate reverse current, causing an overall degradation of the voltage errors reported in Figure 5(a).

The Montecarlo post-layout analysis for the three-stage implementation of the switch ($k = 3$) is shown in Figure 5(b). Regardless of process and mismatch variations, the error voltage V_ϵ of the improved solution remains orders of magnitude smaller than the standard pass-transistor. Noteworthy, such a large improvement is obtained with an increase of the silicon area occupied by the analogue memory as limited as +34% and with an increase of the sampling time of a factor 2.7.

Previous simulations were performed by sweeping V_{IN} on the entire full input range during the sampling phase and by keeping V_{IN} fixed to 0 V in the hold phase. To validate the proposed architecture in all the possible operating conditions, the sampled voltage stored on C_H and the input voltage condition during the hold phase have been both swept in the full ± 0.9 V range. The discharging error V_ϵ was calculated after an hold time t_h of 1 s, for each combination of voltages. The results are represented in the Figure 6, where V_ϵ is expressed as the achievable equivalent number of bits resolution on the entire full scale range of 1.8 V. By increasing the number of stages k of the improved architecture, the high resolution operating area increases evenly over almost all the possible operating conditions, demonstrating the effectiveness of the proposed architecture for the reduction of the leakage current of the switch.

Conclusion: A modular low leakage switch architecture has been proposed. It is composed by simple general blocks that could be used for very low-power switched circuits operated at low frequency and for long-term and high-accuracy analogue memories. The leakage current of the switch is reduced down to few tens of aA over a wide range of input voltages at room temperature, with the residual leakage given mainly by the well-substrate junction reverse current. Such a small leakage current is comparable to the best solutions reported in the literature [7–9]. Unlike these solutions, the proposed technique does not require an ad-

ditional amplifier allowing a zero static power consumption and a small silicon area occupation ($8 \times 8\ \mu\text{m}^2$ for the basic cell). The circuit has been designed in TSMC 0.18 μm technology and from accurate simulations an equivalent resolution of 15 bits over almost all the operating conditions has been obtained, after 1 s hold time, with a switch made of four cascaded basic cells. Theoretical analysis and Montecarlo simulations validate the robustness to process and matching parameters variations of the proposed architecture.

References

1 Mao, W., et al.: A low power 12-bit 1-ks/s SAR ADC for biomedical signal processing. *IEEE Trans. Circ. Syst. 1: Reg. Papers* **66**, 477–488 (2019)

2 Harikumar, P., Wikner, J.J., Alvandpour, A.: A 0.4-V Subnanowatt 8-Bit 1-kS/s SAR ADC in 65-nm CMOS for wireless sensor applications. *IEEE Trans. Circ. Syst. II: Exp. Briefs* **63**(8), 743–747 (2016)

3 Zhang, D., Bhide, A., Alvandpour, A.: A 53-nW 9.1-ENOB 1-kS/s SAR ADC in 0.13-μm CMOS for medical implant devices. *IEEE J. Solid-State Circuits* **47**(7), 1585–1593 (2012)

4 Mayr, C. et al.: A biological-realtime neuromorphic system in 28 nm CMOS using low-leakage switched capacitor circuits. *IEEE Trans. Bio. Circ. Syst.* 243–254, (2016)

5 Rovere, G. et al.: Ultra low leakage synaptic scaling circuits for implementing homeostatic plasticity in neuromorphic architectures. In: IEEE International Symposium on Circuits and Systems (ISCAS), Melbourne (2014)

6 Wong, L.S.Y., et al.: A very low-power CMOS mixed-signal IC for implantable pacemaker applications. *IEEE J. Solid-State Circuits* **39**(12), 2446–2456 (2004)

7 Micah, O., Sarpeshkar, R.: A 10-nW 12-bit accurate analog storage cell with 10-aA leakage. *IEEE J. Solid-State Circuits* **39**(11), 1985–1996 (2004)

8 O'Halloran, M., Sarpeshkar, R.: An analog storage cell with 5e-/sec leakage. In: IEEE International Symposium on Circuits and Systems (ISCAS), Kos (2006)

9 Su, J.J., Demirci, K.S., Brand, O.: A low-leakage body-guarded analog switch in 0.35μm BiCMOS and its applications in low-speed switched-capacitor circuits. *IEEE Trans. Circ. Syst. II, Expr. Briefs* **62**(10), 847–951 (2015)

10 Jiangtao, X., et al.: Low-leakage analog switches for low-speed sample-and-hold circuits. *Microelectr. J.* **76**, 22–27 (2018)

11 Wong, L.S., Hossain, S., Walker, A.: Leakage current cancellation technique for low power switched-capacitor circuits. In: ISLPED'01: Proceedings of the International Symposium on Low Power Electronics and Design (IEEE Cat. No. 01TH8581) Huntington Beach (2001)

12 Zou, L. et al.: Sample-and-hold circuit with dynamic switch leakage compensation. *Electr. Lett.* **49**(21), 1323–1325, (2013)

13 Linares-Barranco, B., Serrano-Gotarredona, T.: On the design and characterization of femtoampere current-mode circuits. *IEEE J. Solid-State Circuits* **38**(8), 1353–1363 (2003)

14 Lin, Y.-S., et al.: Leakage scaling in deep submicron CMOS for SoC. *IEEE Trans. Electron Devices* **49**, 1034–1041 (2002)

15 Roy, K., Mukhopadhyay, S., Mahmoodi-Meimand, H.: Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. *Proc. IEEE* **91**(2), 305–327 (2003)

16 Ratti, L., et al.: Front-end performance and charge collection properties of heavily irradiated DNW MAPS. *IEEE Trans. Nucl. Sci.* **57**(4), 1781–1789 (2010)

17 Suarez, R., Gray, P., Hodges, D.: All-MOS charge-redistribution analog-to-digital conversion techniques. *II. IEEE J. Solid-State Circuits* **10**(6), 379–385 (1975)