Applied Physics Letters ARTICLE scitation.org/journal/apl

Impact ionization-induced bistability in CMOS transistors at cryogenic temperatures for capacitorless memory applications

Cite as: Appl. Phys. Lett. **119**, 043501 (2021); doi: 10.1063/5.0060343
Submitted: 17 June 2021 · Accepted: 8 July 2021 ·
Published Online: 29 July 2021

A. Zaslavsky,[1,2,a)] C. A. Richter,[2] P. R. Shrestha,[2,3] B. D. Hoskins,[2] S. T. Le,[2,3] A. Madhavan,[2,4] and J. J. McClelland[2]

AFFILIATIONS

[1]School of Engineering, Brown University, Providence, Rhode Island 02912, USA
[2]Nanoscale Device Characterization Division, Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA
[3]Theiss Research, La Jolla, California 92037, USA
[4]University of Maryland, College Park, Maryland 20742, USA

[a)]Author to whom correspondence should be addressed: Alexander_Zaslavsky@brown.edu

ABSTRACT

Cryogenic operation of complementary metal oxide semiconductor (CMOS) silicon transistors is crucial for quantum information science, but it brings deviations from standard transistor operation. Here, we report on sharp current jumps and stable hysteretic loops in the drain current as a function of gate voltage V_G for both *n*- and *p*-type commercial-foundry 180-nm-process CMOS transistors when operated at voltages exceeding 1.3 V at cryogenic temperatures. The physical mechanism responsible for the device bistability is impact ionization charging of the transistor body, which leads to effective back-gating of the inversion channel. This mechanism is verified by independent measurements of the body potential. The hysteretic loops, which have a $>10^7$ ratio of high to low drain current states at the same V_G, can be used for a compact capacitorless single-transistor memory at cryogenic temperatures with long retention times.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0060343

There is currently a great interest in cryogenic operation of complementary metal oxide semiconductor (CMOS) circuits for quantum information science and quantum computing — where CMOS devices are expected to provide the control and readout circuitry for solid-state sensors and qubits operated at ultra-low dilution refrigerator temperatures.[1,2] These applications require operating CMOS transistors either in the sub-100 mK regime or near 4.2 K, depending on the design architecture. Since the required circuits are typically relatively small, for reasons of cost and foundry availability, they are often designed in legacy bulk sub-0.25 μm technologies. As a result, several groups have studied submicrometer CMOS transistor operation at cryogenic temperatures,[3,4] investigating such figures of merit as the subthreshold slope (SS), which becomes sharper, and the current drive I_{ON}, which becomes higher due to improved mobility. These studies have also noted deviations from standard transistor curves, such as current jumps and hysteresis caused by physical phenomena like incomplete dopant activation or impact ionization (II) currents. Much of the focus has been on developing cryogenic models compatible with simulation software that could accurately capture the changes caused by new physical mechanisms at low temperatures.[3–5] However, these deviations from standard characteristics can also confer potentially useful device functionality.

In this Letter, we report on bistability and hysteretic loops that appear in the current-voltage characteristics $I_D(V_G)$ of foundry-fabricated, 180-nm-process CMOS transistors operated at low temperatures $T < 30$ K. These effects appear when the transistors are operated at drain voltages V_D sufficient to initiate impact ionization (II) at the drain (all terminal voltages are referred to the grounded source). The hysteretic loop can be magnified by leaving the substrate contact floating or connected to an ultra-high input impedance electrometer, which makes it possible to measure the body potential V_B during the $I_D(V_G)$ sweep. Once V_B is created by the small nanoampere-scale II current, it persists when V_G is scanned back to zero, resulting in a stable, repeatable hysteretic loop. This behavior can be used for a one-transistor (1 T) capacitorless memory with a long retention time (on the scale of minutes at $T \cong 3$ K).

While the exact process parameters for the foundry-made CMOS are not available, the transistors were made in a single-poly twin-well

process with an SiO_2-based gate dielectric. The transistors were wire-bonded, and their electrical properties were characterized in a variable-temperature cryostat capable of base $T \cong 3$ K (as measured by using a calibrated resistor mounted in place of the sample). The $I(V_D,V_G,V_B)$ characteristics were measured using the source-measure units (SMUs) of a commercial parameter analyzer. However, when measuring V_B in the floating substrate configuration, an electrometer with $>10^{14}$ Ω input impedance was used, to avoid providing a current path to ground. A schematic diagram of the NMOS transistor is shown in Fig. 1(a), including the choice of substrate connections; the corresponding PMOS would have the substrate and contact doping

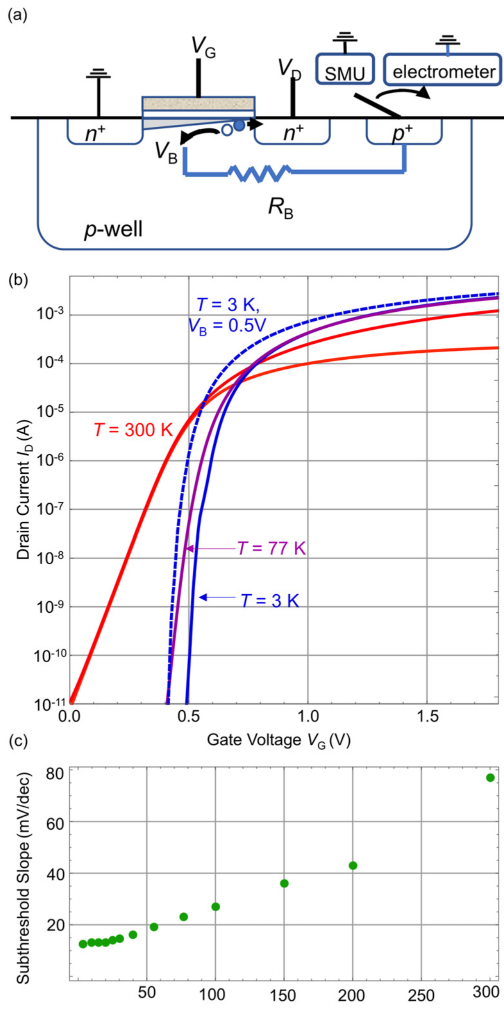

FIG. 1. (a) Schematic diagram of the transistor and measurement set-up. The substrate contact is either a low-impedance signal measurement unit (SMU) that captures the body current I_B or an ultra-high input impedance electrometer that measures the body potential V_B; (b) NMOS $I_D(V_G)$ characteristics at $V_B = 0$ for $V_D = 0.1$ and 1 V ($T = 300$ K) and $V_D = 1$ V ($T = 77$ and $\cong 3$ K). The dashed line shows $I_D(V_G)$ at $V_D = 1$ V and $T \cong 3$ K, but for $V_B = 0.5$ V; (c) subthreshold slope (SS) vs T extracted from the 5–500 pA I_D range. The device size is $L_G/W = 1.44/10$ μm.

polarities reversed. The II current, which becomes significant when the inversion channel has formed ($V_G > V_T$) and V_D is large enough to accelerate carriers to sufficient energy for electron-hole pair creation, is also illustrated. In the case of the NMOS transistors, the II-created electrons would flow to the drain and contribute to I_D, but the holes charge up the body, eventually flowing to ground via either the body contact, leading to a measurable I_B if the body contact is monitored with a parameter analyzer SMU, or the forward-biased source-body *np* junction, if the body contact is floating or connected to the high-impedance electrometer.

The $I_D(V_G)$ characteristics of a long-channel NMOS with a gate length $L_G = 1.44$ μm and width $W = 10$ μm measured at $T = 300$ K ($V_D = 0.1$ and 1 V), 77, and 3 K ($V_D = 1$ V only) with the substrate grounded ($V_B = 0$) are shown in Fig. 1(b). At room temperature, the transistor characteristics are standard, with the SS ≅ 80 mV. As T is lowered, the switching becomes sharper, I_D increases slightly for the same gate overdrive because of higher mobility, and V_T shifts to higher values. The decrease in the SS slows below $T \cong 40$ K, as illustrated in the evolution of the SS extracted from the 5 to 500 pA range of I_D in Fig. 1(c). As seen in other studies of low-T CMOS, the SS ceases to be proportional to kT at cryogenic temperatures because of surface state and band-tailing effects.[3,4,6,7] In our case, the SS saturates at ≅13 mV/decade, which is typical for SiO_2-based gate dielectric CMOS and actually somewhat steeper than what has been seen in more modern high-k dielectric/metal gate nodes.[8] Finally, as expected for an NMOS transistor, a positive $V_B = 0.5$ V shifts V_T to lower values, as shown by the dashed line in Fig. 1(b), which corresponds to the $I_D(V_G)$ of the same transistor at $T \cong 3$ K and $V_D = 1$ V. The V_B-lowered V_T increases I_D by orders of magnitude in the subthreshold regime for a given $V_G < V_T$.

Standard $I_D(V_G)$ characteristics for the same device at higher $V_D = 1.3$, 1.5, and 1.7 V, measured with the substrate grounded through the parameter analyzer SMU, are shown in Fig. 2(a). Once V_D exceeds 1.3 V, the transfer curve exhibits sharp current jumps and a bistable characteristic when V_G is swept up to 1.8 V (the prescribed maximum power supply voltage for this technology) and then down. The physical explanation of this effect dates back three decades.[9–11] At low T, the resistance of the p-well becomes large due to carrier freeze out. The II-created holes produce a significant positive body potential V_B (note that V_B is limited to ∼1 V by the turn-on of the forward-biased source-body junction). This V_B, in turn, lowers the V_T of the transistor, leading to an increase in I_D, analogous to a directly applied V_B, as illustrated in Fig. 1(b). In turn, the increase in I_D leads to more II and a higher V_B. The resulting positive feedback produces an extremely sharp current jump slightly above the I_D turn-on threshold at ≅0.46 V. When the V_G is swept down from a high value, the body is already charged, so the I_D does not drop until the lower V_B-influenced threshold is reached. The simultaneously measured I_B traces corresponding to the $I_D(V_G)$ sweeps at constant V_D are shown in Fig. 2(b). The II-hole current is tiny relative to I_D, peaking at ≈5 nA for the $V_D = 1.7$ V sweep, but, due to the very high body resistance R_B at cryogenic T, it creates a large substrate voltage V_B. It should be emphasized that R_B depends not only on the temperature but also on the current and the details of the substrate doping, and it is, therefore, very challenging to model accurately.[4,12]

A larger and better-controlled hysteretic loop is obtained either by leaving the substrate contact floating or attaching it to an electrometer with ultra-high ($>10^{14}$ Ω) input impedance — these two

Published under an exclusive license by AIP Publishing

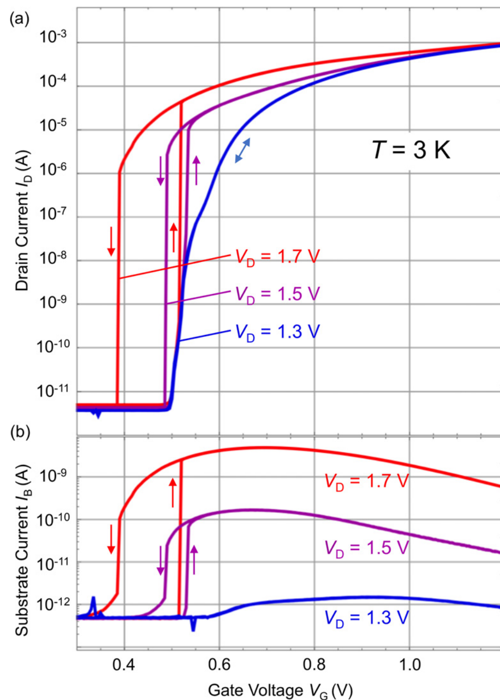

FIG. 2. (a) NMOS $I_D(V_G)$ for $V_D = 1.3$, 1.5, and 1.7 V at $T \cong 3$ K and $V_B = 0$, showing current jumps and hysteresis developing for $V_D > 1.3$ V; (b) corresponding I_B for $V_D = 1.3$, 1.5, and 1.7 V, measured by using the SMU substrate contact. Note that even at $V_D = 1.7$ V, impact ionization I_B remains < 10 nA.

measurement schemes produce identical characteristics, but the latter makes it possible to measure V_B independently. The $I_D(V_G)$ characteristics for $V_D = 1.2$, 1.4, and 1.7 V are shown in Fig. 3(a), measured at a slow scan rate of ≈0.1 V/minute to probe steady-state behavior. For the two higher values of V_D, we observe sharp current jumps and hysteretic loops > 0.2 V wide, with the ratio of the high-state current to the low-current states exceeding 10^7 at $V_G = 0.3$ V, near the middle of the bistable loop. The corresponding $V_B(V_G)$ traces, measured with the electrometer, are shown in Fig. 3(b). At $V_D = 1.2$ V, we find that V_B remains low throughout. But as soon as V_D exceeds the II threshold, we observe a sharp jump in V_B to ≅1.07 V at the same V_G value (≅0.48 V) that produces the current jump in Fig. 3(a). The subsequent behavior of V_B, as determined by the evacuation of II-induced hole charge via the source-body *np* junction, is shown in Fig. 3(b). Using the $V_D = 1.7$ V curve as an example, we see that V_B decays slightly, down to 0.91 V, as V_G is increased to the maximum value of 1.8 V, consistent with somewhat reduced II current due to electron mobility reduction at high vertical fields at large V_G.[13] When the V_G is swept back, V_B goes through the same 1.07 V value at $V_G = 0.48$ V and increases slightly to 1.17 V at $V_G = 0.15$ V and then drops – but not to zero. Rather, V_B remains at ≅0.75 V as V_G returns to zero, and this memory effect is quite persistent in time. Figure 3(c) shows the time evolution of the body voltage V_B, charged by scanning V_G up to 0.8 V and back down to zero at $V_D = 1.7$ V, and then measured for 10^5 s at $V_D = 1.7$ and $V_G = 0$. The evolution of V_B is well described by an exponential decay with a time constant on the order of 13 min over

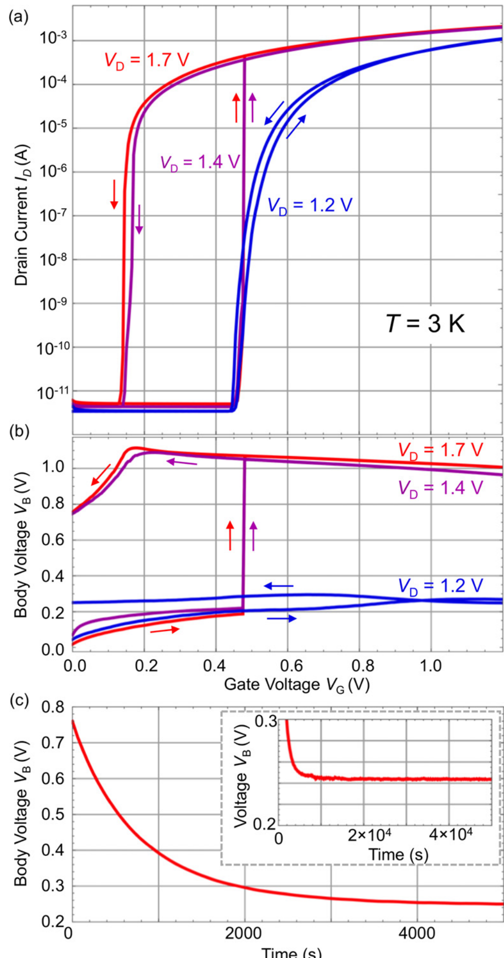

FIG. 3. (a) NMOS $I_D(V_G)$ for $V_D = 1.2$, 1.4, and 1.7 V at $T \cong 3$ K with substrate connected to the electrometer (no I_B due to $>10^{14}$ Ω input impedance); (b) corresponding $V_B(V_G)$ showing the complex evolution of the body potential, note that for large V_D, V_B remains finite as V_G is swept down to zero; (c) evolution of V_B after the body is charged, with the device maintained at $V_G = 0$ and $V_D = 1.7$ V, at $T \cong 3$ K. The decay of V_B is exponential with a time constant of 795 s. Inset shows the longer-term behavior of V_B out to 5×10^4 s.

the first 5000 s (the time constant is ≈795 s, with $R^2 = 0.999$) down to a very slowly decreasing $V_B \cong 0.25$ V background—the longer-term evolution of V_B out to 5×10^4 s is shown in the inset. The decay of V_B is not strongly dependent on V_D: a similar V_B retention measurement at $V_D = 0.1$ V yielded a slightly shorter time constant of ≈755 s, which is reasonable since the drain-body junction is also forward-biased at low V_D, providing an additional channel for discharging V_B. As a result, while a memory utilizing this effect cannot be truly termed non-volatile, on the scale of the anticipated duration of quantum computation a memory that keeps its state for minutes would not require refreshing. We also note that gate leakage I_G would provide another path to discharging V_B, but it was negligible in all of our devices at cryogenic temperatures. To fully discharge V_B, it suffices to momentarily ground the substrate contact, which was done immediately prior

Published under an exclusive license by AIP Publishing

to the start of each measurement to obtain the $I_D(V_G)$ curves in Fig. 3(a) and the corresponding $V_B(V_G)$ curves in Fig. 3(b) for different constant V_D values.

The data in Fig. 3 confirm that this bulk CMOS transistor is suitable as a capacitorless 1 T memory at cryogenic temperatures. It is quite similar to capacitorless 1T-DRAM-like memories fabricated in silicon-on-insulator devices for room-temperature operation, where the body potential was used to store the two memory states, and the resulting V_T shift was used to readout the memory.[14–16] The key difference is that here the memory is based on a standard bulk CMOS transistor, and the retention time is very long, on the order of minutes at $T \cong 3$ K, because the leakage of the source-body junction for $V_B < 0.8$ V is small.

Figure 4 shows analogous bistability in a PMOS transistor fabricated in the same technology with $L_G = 0.18\,\mu$m and $W = 10\,\mu$m, measured with the body contact attached to the electrometer: the $I_D(V_G)$ characteristics are shown in Fig. 4 for $V_D = -1.4$, -1.6, and -1.7 V at $T \cong 3$ K. Evidently, bistable behavior can be obtained from both transistor polarities. In the PMOS case, the hysteretic loop begins to develop for $|V_D| > 1.5$ V, since a slightly larger drain voltage is needed to produce a II current due to the lower mobility of holes. The measurement at $V_D = -1.6$ V shows a gradual charging of the body, as confirmed by the $V_B(V_G)$ trace shown in the inset. When $V_D = -1.7$ V (even more negative), the switching becomes abrupt, and the hysteretic loop resembles the NMOS results in Fig. 3. The magnitude of the drain voltage V_D required to produce a measurable II current in the PMOS transistor is somewhat higher, but the operating mechanism is essentially identical. Note that the device in Fig. 4 has the smallest $L_G = 0.18\,\mu$m available in this technology, so the full range of L_G in both transistor polarities can be employed for the capacitorless DRAM based on our effect.

The capacitorless single-transistor DRAM[15–17] was intended to compete with a standard DRAM as a general-purpose compact volatile memory, but the continued success in scaling the standard DRAM

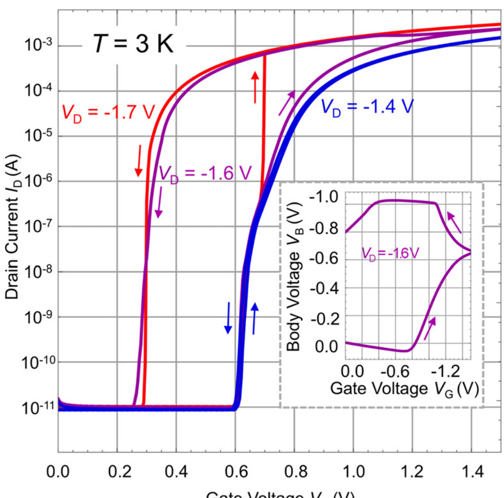

FIG. 4. PMOS $I_D(V_G)$ characteristics of an $L_G/W = 0.18/10\,\mu$m device at $T = 3$ K, showing similar hysteretic curves at $V_D = -1.4$, -1.6, and -1.7 V. At $V_D = -1.6$ V, the body potential charges up gradually, as shown in the inset, whereas at $V_D = -1.7$ V, the current and V_B switch sharply, as in the NMOS device of Fig. 3.

capacitor precluded the capacitorless DRAM's widespread adoption. The cryogenic bulk capacitor-less memory presented here is intended for a niche application: due to the long retention times demonstrated in Fig. 3 and its fabrication in a standard bulk CMOS process, it can be used to provide nonvolatile memory for silicon control circuitry intended for quantum computation or sensing at cryogenic temperatures fabricated in the same process as the digital logic. At the same time, digital CMOS transistors operated as on/off switches would not be greatly affected by this bistable behavior, as the effect of V_B on I_{ON} or I_{OFF} is relatively slight.

In conclusion, we have demonstrated bistable behavior at cryogenic temperatures in bulk CMOS transistors fabricated in a 180 nm foundry process. The physical basis for this behavior is the charging of the local body potential by small impact ionization currents that appear at modest drain voltages (<1.8 V). The body potential can be as large as ≈1 V; it produces a current ratio of $>10^7$ between the high- and low-current states, and its retention time at cryogenic temperatures is sufficiently long, to provide memory capability without refreshing on the time scales predicted for quantum computation. These features make it promising for a compact, cryogenic memory that is effectively nonvolatile if low temperature is maintained.

This work was supported by the National Science Foundation (award QII-TACS-1936221). A. Zaslavsky gratefully acknowledges sabbatical support from Federal Award No. 70NANB18H160 (Brown/NIST PREP Gaithersburg program), while A. Madhavan acknowledges support under the Cooperative Research Agreement Award No. 70NANB14H209 through the University of Maryland.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] B. Patra, R. M. Incandela, J. P. G. van Dijk *et al.*, "Cryo-CMOS circuits and systems for quantum computing applications," IEEE J. Solid-State Circ.uit **53**, 309 (2018).
[2] S. J. Pauka, K. Das, R. Kalra, A. Moini, Y. Yang, M. Trainer, A. Bousquet, C. Cantaloube, N. Dick, G. C. Gardner, M. J. Manfra, and D. J. Reilly, "A cryogenic CMOS chip for generating control signals for multiple qubits," Nat. Electron. **4**, 64 (2021).
[3] A. Beckers, F. Jazaeri, and C. Enz, "Characterization and modeling of 28-nm bulk CMOS technology down to 4.2 K," J. Electron Dev. Soc. **6**, 1007–1018 (2018).
[4] R. M. Incandela, L. Song, H. Homulle, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Characterization and compact modeling of nanometer CMOS transistors at deep-cryogenic temperatures," J. Electron Dev. Soc. **6**, 996–1006 (2018).
[5] J. Ning, M. Schormans, and A. Demosthenous, "Towards an improved model for 65-nm CMOS at cryogenic temperatures," in Proceedings of the 2020 IEEE International Symposium Circuit Systems (ISCAS) (2020).
[6] A. Beckers, F. Jazaeri, and C. Enz, "Theoretical limit of low temperature subthreshold swing in field-effect transistors," IEEE Electron Dev. Lett. **41**, 276 (2020).
[7] F. Balestra and G. Ghibaudo, "Physics and performance of nanoscale semiconductor devices at cryogenic temperatures," Semicond. Sci. Technol. **32**, 023002 (2017).
[8] W. Chakraborty, K. Ni, S. Dutta, B. Grisafe, J. Smith, and S. Datta, "Cryogenic response of HKMG MOSFETs for quantum computing systems," in *2019 Device Research Conference (DRC)* (IEEE, 2019), pp. 115–116.

Published under an exclusive license by AIP Publishing

[9]L. Deferm, E. Simoen, and C. Claeys, "The importance of the internal bulk-source potential on the low temperature kink in NMOST's," IEEE Trans. Electron. Devices **38**, 1459 (1991).
[10]F. Balestra, L. Audaire, and C. Lucas, "Influence of substrate freeze-out on the characteristics of MOS transistors at very low temperatures," Solid State Electron. **30**, 321 (1987).
[11]*Device and Circuit Cryogenic Operation for Low Temperature Electronics*, edited by F. Balestra and G. Ghibaudo (Springer, New York, 2001).
[12]A. Akturk, M. Peckerar, M. Dornajafi, N. Goldsman, K. Eng, T. Gurrieri, and M. S. Carroll, "Impact ionization and freeze-out model for simulation of low gate bias kink effect in SOI-MOSFETs operating at liquid He temperature," in 2009 Interntional Conference on Simulation of Semiconductor Processes and Devices (SISPAD) (2009).
[13]Y. Taur and T. Ning, *Fundamentals of Modern VLSI Devices*, 2nd ed. (Cambridge University Press, New York, 2013).
[14]T. Ouisse, G. Ghibaudo, J. Brini, S. Cristoloveanu, and G. Borel, "Investigation of floating body effects in silicon-on-insulator metal-oxide-semiconductor field-effect transistors," J. Appl. Phys. **70**, 3912 (1991).
[15]S. Okhonin, M. Nagoga, J. M. Sallese, and P. Fazan, "A SOI capacitor-less 1T-DRAM concept," in *2001 IEEE International on SOI Conference* (IEEE, 2001), pp. 153–154.
[16]C. Hu, T.-J. King, and C. Hu, "A capacitor-less 1T-DRAM cell," IEEE Electron Device Lett. **23**, 345–347 (2002).
[17]An up-to-date overview of SOI-based floating-body mechanisms and memories is available in S. Cristoloveanu, *Fully Depleted Silicon-on-Insulator: Nanodevices, Mechanisms, and Characterization* (Elsevier, Cambridge, MA, 2021).

Published under an exclusive license by AIP Publishing