A Review of Sharp-Switching Devices for Ultra-Low Power Applications

SORIN CRISTOLOVEANU1 (Fellow, IEEE), JING WAN1, AND ALEXANDER ZASLAVSKY2
1 IMEP-LAHC Laboratory, INP–Grenoble, Grenoble 38016, France
2 School of Engineering, Brown University, Providence, RI 02912, USA
CORRESPONDING AUTHOR: S. CRISTOLOVEANU (e-mail: sorin@minatec.grenoble-inp.fr)

This work was supported in part by the Nanosciences Foundation of Grenoble, and in part by the European Projects Compose3, WayToGoFast, and Reminder. The work of A. Zaslavsky was supported by the U.S. National Science Foundation under Award ECCS-1068895.

ABSTRACT The reduction of the supply voltage is standard MOSFETs is impeded by the subthreshold slope, which cannot be lowered below 60 mV/decade, even in ideal fully-depleted devices. We review selected CMOS-compatible devices capable of switching more abruptly than MOSFETs, and discuss their merits and limitations. Tunneling FETs (TFETs) are reverse-biased gated PIN diodes where the gate controls the electric field in the interband tunneling junction. Technological solutions for improved performance will be described, including alternative channel materials and geometries, as well as a proposed paradigm shift of increasing the current drive by internal amplification in the bipolar-enhanced TFET. Other emerging sharp-switching mechanisms are reviewed, including the abrupt change in the polarization of ferroelectric materials, mechanical contact in nano-electro-mechanical systems, energy filtering of injected carriers, etc. Recently proposed band modulation feedback transistors are conceptually different from MOSFETs or TFETs. They have similar gated-diode configuration, but are operated in forward-bias mode. Electrostatic barriers are formed (via gate biasing) to prevent electron/hole injection into the channel until the gate or drain bias reaches a turn-on value. Due to bandgap modulation along the channel, these devices can switch abruptly (<1 mV/decade) to a high current.

INDEX TERMS Sharp switching device, TFET, Z2-FET, BET-FET, FE-FET, band modulation, CMOS, SOI, subthreshold slope.

I. INTRODUCTION

The key words for future CMOS devices are “low power” and “RF”. While MOSFETs will continue to improve by adopting alternative semiconductors (Ge, III-V), shorter dimensions and FDSOI configuration, it is clear that innovative switches based on different physical mechanisms are a must. The quest for sharp-switching transistors corresponds to this demand.

The rationale for this strategy is evident from the subthreshold transistor characteristics illustrated in Fig. 1. The lower limit of the operating voltage is governed by the threshold voltage $V_T$. In a conventional bulk MOSFET (curve 1), $V_T$ cannot be reduced without increasing the off-state current $I_{OFF}$, and hence the static power consumption. A shift in $V_T$ by 200 mV causes a variation in leakage current by 3 orders of magnitude. A transistor with steeper subthreshold slope (curve 2) would allow a lower $V_T$ for a fixed $I_{OFF}$ value. Alternatively, one may opt to maintain $V_T$ constant for the benefit of a much-reduced $I_{OFF}$ (curve 3).

The inversion charge in a MOSFET depends exponentially on the surface potential: $Q_{inv} \sim \exp(q\Psi_S/kT)$. An increase by one decade of the inversion charge (or drain current, if the impact of a changing $\Psi_S$ on the mobility is ignored) requires a change in surface potential of $\Delta\Psi_S = (kT/q)\ln10$, which is roughly 60 mV at room temperature. In the ideal MOSFET world, the gate voltage sweep translates fully into surface potential modification, $\Delta V_G = \Delta\Psi_S$, explaining the unbreakable limit of 60 mV per decade of drain current. Such perfect gate-to-surface potential conversion occurs only in fully depleted transistors like planar FDSOI, FinFETs or nanowires. In bulk or partially depleted MOSFETs, the
II. TUNNELING FIELD EFFECT TRANSISTORS (TFETs)

The most popular sharp-switching device that in principle can beat the 60 mV/decade switching limit at room temperature without abandoning compatibility with the dominant CMOS materials and processing is the tunneling FET (TFET). The simplest planar device layout of a Si-based TFET is essentially a gated \( P^+ - I - N^+ \) diode operated in reverse bias, [1] shown schematically in Fig. 2a (although sidewall-gated vertical layouts [2], [3] and gated \( P^+ N^+ \) diodes without a channel were discussed early on). Electrons can tunnel from the valence band of the \( P^+ \) region to the conduction band of the \( N^+ \) region as soon as these band edges are energetically aligned. However, when the diode body is fully depleted (\( V_G \approx 0 \)), the distance between the \( N^+ \) and \( P^+ \) terminals is too large for tunneling to occur. The gate voltage \( V_G \) fills the channel with either electrons or holes, depending on polarity, exactly as in a standard MOSFET, leading to an effectively heavily-doped junction between the channel and one of the electrodes. This junction can pass a tunneling current in reverse bias, just like the Esaki diode (see arrow in Fig. 2a) [5]. The magnitude of the tunneling current is a sensitive function of the electric field in the junction, the magnitude and direction of which generally depends on both \( V_G \) and \( V_D \). If the device is symmetric, as in Fig. 2a with both source and drain made of Si with equally sharp doping profiles, then both \( N \) and \( P \)-type TFET operation can be accomplished in the same layout. The disadvantage is a high \( I_{OFF} \). More typically, there is an asymmetry built into the layout (either intentionally, through the use of a heterostructure or gate underlap, or unintentionally due to unequal dopant density and diffusion profile), leading to preferential operation in either \( N \) or \( P \) mode.

Figure 2b shows the band diagram in the channel of an asymmetric heterostructure TFET, with a \( P^+ \)-Ge drain, \( P^- \)-Si channel, and \( N^- \)-Si source (including bandgap narrowing). The device is operated in \( N \) mode (turned on by a positive \( V_G \)) and illustrates the difference in carrier injection between a TFET and a MOSFET, emphasizing the TFET’s possible advantage as a sharp switch. The barrier to carrier injection into the channel is set not by the gate-controlled surface potential \( \Psi_S \), but by the bandgap of the material at the tunneling junction and the magnitude of the electric field \( F \). While quantitative formulae for the tunneling current in indirect-gap materials [6], [7], where interband tunneling is a phonon- or impurity scattering-assisted process, should be approached with caution, expressions for the tunneling current density \( J_{TUN} \) typically follow a modified exponential dependence on the maximum electric field \( F_{MAX} \) [8]:

\[
J_{TUN} \sim F_{MAX} \exp(-F_0/F_{MAX})
\]

where \( F_0 \approx E_G^{3/2} m^{*1/2} \) depends on the bandgap \( E_G \) and the tunneling effective mass \( m^{*} \) (the latter treated as a fitting parameter). Expressions analogous to Eq. (2) are typically used for TCAD simulations of TFET currents and assumptions of heavily-doped and abrupt junctions have produced predictions of sharp \( SS < 60 \) mV/decade switching, dating back a decade [9].

Research interest in Si-compatible TFETs was substantially increased by the first experimental reports of sub-60 mV/decade switching [10], [11]. An example of the measured transfer characteristics of an SOI-based all-Si TFET is shown in Fig. 3a. The \( SS \approx 53 \) mV/decade region is evident at low \( I_D \), unambiguously demonstrating that TFETs can overcome the standard transistor 60 mV/decade switching limit. At the same time, the \( I_{ON}/I_{OFF} \) ratio is somewhat low, at \( \sim 10^4 \), and the \( I_{ON} \) for moderate \( V_D = V_G \approx 0.5 \) V remains in the \( \sim 1 \mu A/\mu m \) range, too low for practical applications.

---

**FIGURE 1.** Illustration of the sharp-switching transistor paradigm. Degraded subthreshold characteristics of bulk-Si MOSFET (curve 1); fully-depleted (FD) MOSFET with 60 mV/decade slope and low \( V_T \) (curve 2) or reduced \( I_{OFF} \) (curve 3); and “ideal” sharp-switching device (curve 4). The dots indicate the operating voltages (\( V_{DP} \) and 0).
Attempts to increase the all-Si TFET $I_{\text{ON}}$ while maintaining $SS < 60 \text{ mV/dec}$ over several orders of magnitude in $I_D$ proved disappointing, with even the best results [12], [13] falling well below what could be achieved in standard CMOS. Given the $F_0 \sim E_G^{3/2}$ dependence in Eq. (2) and the relatively large Si bandgap $E_G \sim 1.1 \text{ eV}$, an obvious avenue for increasing $I_{\text{ON}}$ without abandoning Si compatibility was to position the tunneling junction in a lower bandgap material such as SiGe or even pure Ge, ideally without going to an all-Ge device, where $I_{\text{OFF}}$ would also be enhanced [11], [14]. An asymmetric Ge-source TFET using selective Ge regrowth was reported by Kim et al. [15]: while the $I_{\text{ON}}/I_{\text{OFF}}$ ratio improved to $>10^6$, the $I_{\text{ON}}$ did not exceed 1 $\mu A/\mu m$ at $V_D = 0.5 \text{ V}$. A trigate Ge/Si heteronanowire TFET was reported by Le et al. and shown in Fig. 3b [16]. In this device the $I_{\text{ON}}$ slightly exceeded 1 $\mu A/\mu m$ (when normalized to nanowire diameter of $\sim 50 \text{ nm}$), but again remained uncompetitive compared to standard CMOS (or nanowire FETs). Finally, higher $I_{\text{ON}}$ was reported in SiGe-based TFETs, up to $\sim 400 \mu A/\mu m$, but at relatively elevated operating voltages [17] and with $SS > 100 \text{ mV/dec}$ throughout [18].

A. SEARCH FOR ENHANCED $I_{\text{ON}}$: ALTERNATIVE MATERIALS

The relatively disappointing $I_{\text{ON}}$ reported in indirect-bandgap Si and SiGe-based TFETs has stimulated a great deal of research into alternative materials, whether based on III-V semiconductors (either combined with Si, such as InAs/Si hetero-TFETs, or entirely based on III-V compounds) or on even more exotic materials, such as graphene and alternative atomically thin 2D crystals. The key parameters at stake are the bandgap $E_G$ and the tunneling effective mass. In addition, if the full range of various III-V heterostructures is assumed available, there are interesting possibilities involving the staggered bandgap of InAs/GaSb or InAs/AlGaSb [19]. The number of enthusiastic simulation-based publications on the superior performance of such TFETs, quoting low $SS$ values combined with high $I_{\text{ON}}$, is extremely large. However, the experimental reports of alternative material-based TFETs are not nearly as common and generally lag far behind theoretical predictions in terms of performance. This can be seen, for example, in a recent compendium of TFET results [20], where many (although certainly not all) publications claiming low $SS$ are compared to both theoretical predictions and to FinFET CMOS transistors. In addition to the assumptions of highly abrupt and uniform tunnel junctions that underpin most if not all TFET simulations that promise CMOS-beating performance, alternative material TFETs also suffer from optimistic assumptions regarding the successful deployment of ultrathin high-$\kappa$ dielectric gate stacks on various non-Si materials – much easier in simulation than in reality – as well as defect-assisted tunneling that is undoubtedly present in non-lattice-matched heterojunctions.

The profusion of theoretical and experimental results has led to interest in TFET benchmarks that would make for a fair comparison of actual TFET performance, including both the $SS$ and $I_{\text{ON}}$ figures of merit, as well as the operating voltage. One simple benchmark, proposed in a 2013 publication by an IMEC group, classified TFETs by the $I_{60}$ current value, corresponding to the maximum current at which the $SS$ remained below the 60 mV/decade value [21]. Given that low-standby power CMOS operated at $V_{DD} \sim 0.5 \text{ V}$ can provide $I_{\text{ON}} \sim 100 \mu A/\mu m$ and $SS \sim 75 \text{ mV/decade}$, the value of $I_{60}$ required to make the TFET competitive would have to fall in the $1-10 \mu A/\mu m$ range, far above the best experimental reported result of $I_{60} < 10^{-2} \mu A/\mu m$ [22].

A more recent benchmark, proposed in 2015 by an IBM group [23], does not presume the existence of an $SS < 60 \text{ mV/decade}$ region in the transfer curve, making it more generally applicable to alternative material-based TFETs, where much of the recent experimental progress has focused on improving $I_{\text{ON}}$. The procedure is illustrated in Fig. 4a and the results from a number of experimental publications on non-Si III-V [24]–[27] and Si or InAs-Si TFETs [28]–[32] are collected in Fig. 4b. As shown schematically in Fig. 4a, the average SS is calculated taking the exponentially increasing $I_D-V_G$ TFET characteristic between points $I_{\text{DMIN}}$ and $I_{\text{DMAX}}$ for some constant $V_D$. It is then plotted vs. $I_{\text{DMAX}}/V_D$, where normalization by $V_D$ favors devices that provide significant current drive at low $V_D$. A TFET with $SS < 60 \text{ mV/decade}$ over several orders of magnitude in $I_D$ that also offered $I_{\text{DMAX}}/V_D > 100 \mu A/\mu m$
per V of $V_D$ would fulfill the original TFET promise of sharper switching than standard CMOS. Unfortunately, an inspection of Fig. 4b reveals that alternative material TFETs, while making some progress towards higher $I_{ON}$ compared to Si or SiGe devices, remain quite far from the lower right quadrant of Fig. 4b, despite nearly a decade of research. In fact, if the original rationale for TFET research was the existence of an $SS < 60$ mV/decade subthreshold transfer characteristic, it is worrisome that the best results in terms of low $SS$ in Fig. 4b are largely group-IV based and date back several years.

**FIGURE 4.** (a) Schematic $I_G$–$V_G$ characteristic indicating the points used to extract the average $SS$ from the exponential section of the transfer curve between $I_{ONMIN}$ and $I_{ONMAX}$ (after Cutaia et al. [23]); (b) average $SS$ vs. normalized $I_{ONMAX}$ plot, with results from Ref. 23 supplemented by additional points from Ge/Si [16] and SiGe/Si [17] devices and a low-standby power CMOS point for comparison.

### B. ALTERNATIVE GEOMETRIES FOR ENHANCED $I_{ON}$

In addition to attempting to enhance the TFET $I_{ON}$ by pursuing alternative channel materials and heterostructure combinations, there have been proposals to increase the effective tunneling area while maintaining a compact device footprint. In the simplest planar implementation of the TFET as a counterdoped MOSFET shown in Fig. 2, the tunneling area is small. Increasing the channel thickness is of little help, since the tunneling current is exponentially concentrated in the region of highest $F_{MAX}$, near the gate. At least two possible geometric enhancements intended to line up the tunneling direction with the $V_G$-induced $F_{MAX}$ over a larger area have been suggested, as illustrated in Fig. 5.

The use of a raised Ge electrode region adjacent to the gate (Fig. 5a), would in principle permit tunneling into an inversion layer created by $V_G$, resulting in a larger tunneling area (an arrangement sometimes referred to as line tunneling [33]). Proposed in a simulation study with a raised Ge source [34], this arrangement was more recently experimentally realized with a SiGe source [35] – the $I_{ON}$ current varied linearly with $L_G$, as expected for tunneling into a $V_G$-controlled inversion layer, but the values of $I_{ON}$ remained disappointingly small.

A more aggressive attempt to realize a TFET with the tunneling aligned with the $V_G$-created field over the entire $L_G$ area has been published in a series of papers on electron-hole bilayer (EHB) TFETs [36], [37]. The basic idea is illustrated in Fig. 5b for a planar implementation. In the active mode, the channel contains both electron and hole 2D electron gases near the top and bottom insulator interfaces, controlled by the $V_G$ and $V_{BG}$, respectively. The tunneling direction is vertical and the concept can, in principle, function for a number of channel materials, including Si, Ge [38], and more exotic materials like graphene [39] (it should be noted that similar unipolar devices, based on resonant tunneling between parallel independently-gated 2D electron gases date back almost two decades [40]). The experimental realization of high-current EHB-TFETs has been held back by a number of issues [41]. First, the body needs to be quite ultrathin to ensure a high vertical $F_{MAX}$ and hence high $J_{TUN}$ in the on-state. However, the same high $F_{MAX}$ effectively adds quantization energies of both electron and hole subbands to $E_G$, increasing the tunnel barrier. Secondly, in semiconductors with anisotropic or multiply-branched band structures, the lowest subbands of both 2D carrier gases are set by the heavier effective mass and it is the same heavy effective mass $m^*$ that enters into the vertical tunneling expression (unlike the standard planar MOSFET or the planar TFET of Fig. 2a, which benefit from a lighter in-plane transport or tunneling $m^*$). Finally, the simultaneous coexistence of high-density electron and hole 2D gases in the same ultrathin body may be impeded by the super-coupling effect, with one type of carrier gas displacing the other [42].

Another possible challenge to both types of TFET devices in Fig. 5 is the voltage drop arising from the ordinary current flow along the oxide interface. For a significant tunneling current, this voltage drop would lead to a nonuniform $F_{MAX}$ as a function of position along $L_G$, an effect quite similar to emitter crowding in a narrow-base heterojunction bipolar transistor (HBT). Given the exponential dependence of $J_{TUN}$ on $F_{MAX}$, the increased effective tunneling area promised by the EHB-TFET would appear to require a negligible in-plane resistance for inversion layer in Fig. 5a and either a negligible in-plane resistance or a perfectly matched in-plane resistance for both 2D carrier gases in Fig. 5b. In simulation, this issue can generally be overcome by assuming very high mobilities in graphene or monolayer 2D crystal materials [43], [44], but experimental demonstrations of high-current EHB TFETs in any material system have been lacking to date.

### C. BIPOLAR AMPLIFICATION FOR ENHANCED $I_{ON}$

A different approach for improving TFET $I_{ON}$ has been proposed and simulated in SiGe/Si materials by combining a standard TFET and a Si/SiGe HBT in a single device [45]. The concept of the bipolar-enhanced TFET (or BET-TFET) is schematically illustrated in Fig. 6a: the $V_G$-controlled TFET current biases the emitter-base junction of a monolithically integrated HBT, so the tunneling current is multiplied by the usual HBT current gain $\beta > 100$, leading to a high $I_{ON}$. When the TFET current is off, the HBT base is effectively floating, leading to a negligibly small
II. I OFF. A vertical current flow version of the BET-FET is illustrated in Fig. 6b. From top to bottom we have an N+ source (collector), the tunneling SiGe region with a surrounding gate, an undoped Si buffer layer, a P+ SiGe base, and the N+ drain (emitter). Using a Si0.7Ge0.3 HBT base provides $\beta \sim 10^3$ and produces $SS < 60$ mV/decade over 7 orders in the drain current [45]. Planar BET-FET layouts have also been simulated, including an asymmetric design with a Si/SiGe N+/P emitter-base heterojunction on one side of the gate only (an asymmetry analogous to Fig. 5a) that predicts $SS < 60$ mV/decade over 10 orders in $ID$ and $I_{ON} > 1000 \mu A/\mu m$, at the cost of a more challenging fabrication involving selective epitaxy [46]. A planar layout with a lateral Si/SiGe heterojunction in the channel, adjacent to a top-gate controlled TFET that is, in principle, compatible with current advanced FD-SOI process flow is shown in Fig. 6c. Again, the simulated performance is quite promising, with $I_{ON} \sim 260 \mu A/\mu m$ [47], but needs experimental validation. Similar layouts will be discussed later, in Section IV of this review, in the context of band modulation devices.

III. A SYMPHONY OF ALTERNATIVE SHARP SWITCHES
A. FERROELECTRIC FET
The concept of the FE-FET consists in integrating the ferroelectric material in the gate stack, as shown in Fig. 7a. A ferroelectric material (FE) exhibits an S-shaped polarization versus electric field characteristic (Fig. 7b). Since the polarization is related to charge and the field to voltage, the $Q-V$ curve is also S-shaped with a negative capacitance region (Landau switch). As the gate bias increases, the voltage drop across the FE decreases, boosting the for-
tance region (Landau switch). As the gate bias increases, the FE field is accelerated by the con-
comitant decrease in threshold voltage. The principle is

clear, buttressed by countless simulations. Unfortunately, proofs are scarce at the experimental level, as most fabric-
cated devices feature subthreshold swing much higher than 60 mV/decade. An indirect hint of the FE mechanism is shown in Fig. 7c. While the subthreshold swing is expected to increase with temperature, the opposite effect is noted in a narrow temperature range (300–350 K) and attributed to FE action [50].

More recently, sub-$kT/q$ swing was demonstrated in long MOSFETs with HfZrO-based gate dielectrics, operated in deep subthreshold and saturation regions ($V_G \ll V_T$ and $V_D > 0.5 V$) [51]–[53]. The sharp switch disappeared in the Ohmic region and there was no evidence for negative capacitance effect [51]. Lee et al. [53] fabricated metal-gate Fe-MOSFET with 5 nm thick HfZrO (including a monolayer of SiO2 as an intermediate dielectric) directly grown by ALD on silicon. Hysteresis was observed with a steeper slope for
reverse sweep. The device suffered from a drop in $I_{ON}$ due leakage through the thin FE dielectric.

The technology of FE-based negative capacitors is progressing [54]. Besides technical problems there are reliability issues (cycling, fatigue). A more fundamental limit is set by the thickness of the FE layer, which should not leak and yet must be ultrathin. Otherwise, the penetration of the fringing field from drain into the gate dielectric will trigger short-channel effects (SCEs) that will definitely compromise, if not completely destroy, the steep switching characteristic. A partial solution is to add an external FE capacitor in series with the gate terminal, rather than integrated in the gate dielectric [55]. A single-crystal BiFeO$_3$ capacitor externally connected to the gate of a conventional FinFET led to 8 mV/decade swing over 8 decades of current [56]. At the circuit level, designers would have to live with the inherent hysteresis effect in FE materials.

**B. RESISTIVE-GATE AND RESISTIVE-SOURCE FETS**

The principle of this device (Re-FET) is borrowed from resistive memory. A metal-insulator-metal MIM structure is able to switch back and forth between high- and low-resistance states via the formation of current filaments. The Re-FET was fabricated by depositing the MIM layer on top of the gate stack, as shown in the inset of Fig. 8 [57]. In subthreshold region, the gate voltage is entirely absorbed by the high-resistance MIM layer, which blocks the growth of the inversion charge and maintains the device in the OFF state. As soon as the voltage drop across the MIM is sufficient to switch it into the low-resistance state, the gate voltage becomes fully available to sustain the inversion charge. The drain current increases suddenly from $I_{OFF}$ to $I_{ON}$ state with an 8 mV/decade swing.

Further optimization will be required to overcome insufficient $I_{ON}/I_{OFF}$ ratio (only ∼2 decades in Fig. 8), relatively high operating voltage of ∼2 V, and reliability (cycling). A serious limitation is hysteresis: for reverse sweep, the device remains in the ON state unless a reset operation is performed at $V_G < 0$. The Re-FET principle can be extended to any material used in resistive memories, and even to spin torque or insulator-to-metal transition (IMT) structures, that are able to switch between two highly distinct resistive states. The crucial point is, again, the thickness of the final gate stack to prevent short-channel effects.

An alternative approach is to embed the switching material in the source terminal rather than in the gate stack. Shukla et al. [58] demonstrated such a transistor, where vanadium oxide VO$_2$ is used as an IMT layer deposited on the source. The gate modulates the current flowing through the series combination of MOSFET and VO$_2$ layer, triggering an abrupt phase transition. Operated at relatively high drain and gate bias (∼5 V), this device showed very steep switching over a limited range of current (∼2 decades) and hysteresis. The bias could potentially be reduced by scaling the device size and using InGaAs or Ge channel FinFETs [58].

In the OFF state, the gate is pulled up, lowering $C_{OX}$ and increasing $V_T$. In the ON state, the gate is pulled down while increasing $V_G$, which further boosts the channel formation. This mechanism is analogous to the negative capacitance in a mechanical Landau switch. The proof of concept is shown in Fig. 9, where the swing is as low as 2 mV/decade [59]. The thick air-gap (200 nm) explains the excessive $V_G$ needed to operate the device, as well as the wide hysteresis. An analytical model has been developed to describe the basic device operation and suggest guidelines for optimization [60].

A fundamental issue is that the air-gap is increased in the OFF state, potentially leading to SCEs that could jeopardize $I_{OFF}$. The solution is to use accumulation-mode or junctionless FETs, where the required gate displacement is in the opposite direction [61]. Here, the gate is pulled down
in the OFF state to fully deplete the channel (and better control SCEs). Increasing $V_G$, the gate is pulled up, leaving the channel undepleted. Numerical simulations of SG-FET with 1 nm air-gap, $10^{18}$ cm$^{-3}$ doping, 10 nm thick body, and 25 nm gate length promise outstanding performance: 13 times higher $I_{ON}$ and 5 orders lower $I_{OFF}$ compared with a fixed gate device. Nevertheless, we wonder when, if ever, a 1-nm air-gap will be fabricated while avoiding sticking, sealing and variability issues.

A more pragmatic implementation of a NEM-FET may be the reciprocal configuration, here the gate is fixed while the suspended channel moves. In such a device the channel could consist of a carbon nanotube [62], a nanowire [63], or a 2D material.

**D. NEM RELAY**

The NEM relay is a very simple mechanical switch, shown in Fig. 10. A metal cantilever beam forms the “channel” and there is no inversion charge to control. The gate just provides the electrostatic force to move the beam up and down, thereby connecting or disconnecting the source and drain. The $I_D$-$V_G$ characteristics are abrupt (Fig. 10) and show hysteresis induced by the contact forces (sticking) [64]. Complementary logic gates with energy recycling have been demonstrated: the potential energy stored in the bent beam serves to oscillate the beam between OFF and ON states [65].

Drastic reduction in operating voltage ($<1$ V) may be achieved by scaling the size of the NEM relay FET down to a 5 nm air-gap, 10 nm-thick cantilever, and 100 nm gate length [66]. The challenge is the contact reliability, which depends on how gently the beam lands on the drain without sticking or damaging the material (the so-called “Hammer effect”). Since logic circuits require very fast operation, a less ambitious and more realistic application of the NEM relay may be nonvolatile memory with energy-recycling capability [67], [68].

**E. SUPERLATTICE NANOWIRE FET**

This conceptual transistor aims at selecting carriers with appropriate energy passport that will be eventually authorized to pass the frontier from source into the channel. High-energy carriers are not suitable because they are likely to overcome the barrier and contribute to the drift-diffusion current with a 60 mV/decade limit. The brilliant idea for carrier discrimination is to create forbidden energy zones by using a superlattice (SL) embedded in the source. According to the Krönig-Penney model, a periodic barrier-well structure results in minibands with well-defined energies. Since the lateral growth of a thin SL is impossible, the proposed device (SL-FET) is a vertical gate-all-around nanowire, shown in Fig. 11a [69]. A carefully designed SL (GaAs/GaAlAs, InGaAs/InAlAs, etc.) with at least 7 periods generates minibands that are narrow enough (for energy confinement) and sufficiently separated from each other, as illustrated in Fig. 11b. Only one conduction miniband ($E_{C,MB}$) is selected to release carriers into the channel and contribute to the drain current. Electrons with energy below the bottom level of the conduction miniband cannot enter the channel. Conversely, high-energy carriers cannot enter the narrow miniband, nor can they reach the upper miniband that lies much higher in energy.

**FIGURE 11.** (a) Schematic layout of the superlattice-FET; (b) band diagram showing the superlattice and minibands at the source (left) and the channel barrier (right); and (c) simulated transfer characteristics at $V_G = 100$ mV for the following parameters: $L_G = 10$ nm, undoped SL (curve A); $L_G = 10$ nm, $N_A = 10^{19}$ cm$^{-3}$ in SL quantum wells (curve B); $L_G = 20$ nm, $N_A = 10^{19}$ cm$^{-3}$ in SL quantum wells (curve C) (adapted from Gnani et al. [69]).

The SL-FET is still a field-effect transistor, based on the thermionic injection of energy-filtered electrons. For low bias ($V_G \ll V_T$), the source-channel barrier is high, located well above the upper boundary of the conduction miniband, which inhibits the carrier injection. Increasing $V_G$, the barrier height is gradually lowered below the miniband upper edge enabling more and more carriers to be injected into the channel and reach the drain. The analogy with a water lock controlling the water flow from a shallow lake comes to mind. Only a limited lock displacement ($V_G$ range) enables the whole lake (miniband) to discharge into the channel. Sharp switching is nothing but a narrow $V_G$ range needed to transition between OFF and ON states.

Outstanding characteristics with 15 mV/decade swing have been simulated (curve C in Fig. 11c). The SL-FET is very sensitive to the SL composition and thickness. For example, curves A and B in Fig. 11c correspond to devices that were insufficiently optimized. Tailoring a superlattice with sub-Å resolution is a difficult task for technologists, which explains why the device has not been fabricated yet despite its intrinsic elegance.

Source and drain engineering has also been proposed to form a resonant-barrier tunneling FET [70]. A very thin nanowire (2–4 nm in diameter) is constricted at the gate edges by introducing $\sim$1 nm dielectric pockets in the source.
and drain regions. Quantum simulations of the resulting double-barrier structure confirm that carriers penetrate into the channel by resonant tunneling, which causes the $I_D(V_G)$ characteristic to feature a 45 mV/decade swing over a limited range of $V_G$.

F. SOI MOSFET IN LATCH MODE

Regular SOI MOSFETs are capable of sharply switching from OFF to ON state as a result of combined floating-body and impact ionization effects. At relatively high drain bias, carriers experience impact ionization even in very weak inversion. While the generated minority carriers (electrons) are collected by the drain, the majority carriers (holes) tend to accumulate in the body. This extra positive charge increases the body potential and reduces $V_T$, which means that more minority carriers are available to enjoy impact ionization. A feedback mechanism makes the transistor switch abruptly, see Fig. 12. The turn-off characteristic is delayed because the positive charge opposes the necessary drop of the body potential. In the extreme case of high $V_D$, the transistor remains locked in ON state (transistor latch-up). The subthreshold $I_D(V_G)$ characteristic is actually not “vertical” but exhibits a snapback with negative values for both the transconductance and the output conductance.

Detailed measurements on primitive SOI MOSFETs have led to a model where the instability (snapback) is treated with phase-transition theory [71]. This mechanism has recently been rediscovered in the context of ultrathin FDSOI MOSFETs [72]. The hysteresis (see curve B in Fig. 12) offers a good option for floating-body capacitorless one-transistor DRAM (1T-DRAM) memory. On the other hand, the sharp switching occurs for too high a drain voltage and is not appealing for low-power logic circuits.

IV. BAND MODULATION DEVICES

Band modulation is a relatively recent concept in the sharp-switching device arena. The band modulation device is essentially a forward biased $P^+-I-N^+$ diode with two field-controlled gates. It achieves sharp switching behavior by using the feedback mechanism between electron and hole injections. This leads to exceptionally abrupt switching, potentially useful for logic, as well as gate-controlled hysteresis in the output characteristics and charge sensitivity that may find other applications (memories, sensors).

A. OPERATING PRINCIPLE OF BAND MODULATION DEVICES

Band modulation devices can take different forms, but the operating principles are quite similar. Taking the $Z^2$-FET (for zero subthreshold swing and zero impact ionization FET) as an example, schematically shown in Fig. 14a, the carrier injection barriers are formed close to the $N^+$ drain and $P^+$ source by the negatively biased front gate ($V_G < 0$) and positively biased back gate ($V_G > 0$), respectively, see the bands plotted in Fig. 14b for $V_D = 0$ [75]. These two barriers emulate the virtual $NPNP$ thyristor structure and block the flow of electrons and holes under low negative $V_D$. As $V_D$ becomes more negative, the channel potential rises accordingly, until it is pinned by the gate potential, at which point a further increase in $|V_D|$ begins to reduce the electron injection barrier, see the $V_D = -1.5$ V diagram in Fig. 14b. Given a lower barrier, electrons are injected from the drain into the channel and flow to the source. The electron current creates
a potential drop at the source junction and reduces the hole injection barrier, causing hole injection from the source into the channel. The holes, in turn, flow to the drain and reduce the electron barrier. This creates positive feedback, where higher electron current leads to higher hole current and vice versa, finally triggering an abrupt collapse of both electron and hole injection barriers, shown in Fig. 14b at $V_D = -2$ V. The device turns on sharply, see Fig. 14c. In the ON state, high densities of electrons and holes injected into the channel screen the electric field from the gate. Hence, after turn-on, the device cannot be turned off unless $|V_D|$ is reduced below a certain low level, where the injection of carriers becomes too low to sustain the electron-hole plasma in the channel. This can be seen from Fig. 14c showing the $I_D-V_D$ characteristics of $Z^2$-FET. Large hysteresis is observed between sweeping $V_D$ forward and backward, with turn-on voltage $V_{ON}$ linearly controlled by the front gate voltage $V_G$ [76].

Unlike the thyristor, which also works with the positive feedback mechanism, the band modulation device relies on the gate-controlled injection barriers without involving the impact ionization. Thus, the band modulation device has advantages of better temperature stability and lower operating voltage compared to conventional thyristor.

**FIGURE 14.** (a) Schematic view of the $Z^2$-FET; (b) evolution of potential profile from source to drain under different $V_G$ values for $V_D = -2$ V and $V_{G2} = +2$ V; (c) $I_D-V_D$ characteristics at room temperature of a $Z^2$-FET built on an SOI substrate with $L_G = 400$ nm, $t_{HOx} = 500$ nm, Si channel thickness of 20 nm, 3 nm HfO$_2$ gate oxide, and $f_{Box} = 140$ nm.

**B. VARIANTS OF BAND MODULATION DEVICES**

Depending on the number of front gates and on how the electron and hole injection barriers are formed, different band modulation devices have been reported, such as the field effect diode (FED) [77], the $Z^2$-FET [75], [76], and the zero front-gate, zero subthreshold swing and zero impact ionization FET ($Z^2$-FET) [78].

The FED features two front gates divided by a gap [77], [79], schematically shown in Fig. 15. The two front gates on top of the channel are biased oppositely to form the electron and hole injection barriers adjacent to the $N^+$ and $P^+$ doping regions, respectively. The $I-V$ characteristics of the FED show sharp switching and hysteretic behavior, see Fig. 15. The turn-on voltage is linearly controlled by $V_G = V_{G1} - V_{G2}$, the difference between the separately biased gate voltages. The FED can be used for electrostatic discharge (ESD) protection, as it exhibits high $I_{ON}$ and fast switching. Though the design with two front gates allows the flexibility to control the electron and hole barriers separately, the FED suffers from some disadvantages. Due to the gap needed between the two front gates for isolation, the device occupies more area. In addition, this ungated and lightly-doped gap contributes extra channel resistance, reducing the $I_{ON}$. Besides, the two front gates will have high parasitic capacitance through the narrow gap affecting the high frequency performance.

![FIGURE 15. (a) Schematic view and $I_D-V_G$ characteristics vs. $V_C = V_{G1} - V_{G2}$ of an FED device built in a 70 nm Si channel SOI substrate with $L_G = 500$ nm (adapted from Salman et al. [77]).](image1)

Unlike the FED, the $Z^2$-FET uses a single front gate combined with a backgate to create the injection barriers, see Fig. 14a. In the $Z^2$-FET, the uncovered part of the channel is controlled by the backgate only, whereas the rest of the channel is governed by both gates. The $V_{ON}$ of the $Z^2$-FET is still linearly controlled by the front gate, as shown in Fig. 14c. Due to the thick buried oxide compared to front gate oxide, the impact of $V_{G2}$ on $V_{ON}$ is negligible unless the device is fabricated in an advanced FD-SOI technology where ultra-thin buried oxide is used to enable aggressive scaling. In this case, higher $V_G$ is needed to counteract the effect from $V_{G2}$ and restore the $V_{ON}$.

Compared to FED, the $Z^2$-FET occupies a smaller footprint and has better scaling capability down to 20 nm [76]. At the same time, it has higher $I_{ON}$ and lower parasitic capacitance due to the absence of the gap between the gates. So far, the $Z^2$-FET has been demonstrated as sharp switch with subthreshold swing $SS < 1$ mV/dec and $I_{ON}$ at the 1000 $\mu$A/µm level, exceeding other sharp-switching devices under moderate voltages – see the comparison in Fig. 16a [76]. A number of $Z^2$-FET variants, both N and P mode, have also been demonstrated without an actual back gate, since the injection barrier in the ungated region can also be created using surface charge [76], [80], channel doping or a ground plane [81]. The local ground plane variant makes the $Z^2$-FET a three-terminal device that is fully compatible with FD-SOI CMOS.

Moreover, extensive studies have been performed to explore the application of the $Z^2$-FET as high performance single-transistor capacitorless dynamic and static random access memory (1T-DRAM and SRAM), see Fig. 16b [82]. Unlike the conventional 1T-1C DRAM, where the stored charge is directly used to distinguish the logic states, the $Z^2$-FET 1T-DRAM uses the charge $\Delta Q_{G}$ stored under the gate (memory state “1”) as the seed to generate a discharge current $\Delta Q_{G}/\Delta t$ that triggers the positive feedback and turns on the device under a fast $V_D$ pulse. Thus, it requires less
stored charge and features a low supply voltage, high access speed and long retention time. If no charge is stored under the gate (memory state “0”), there is no discharge current and the device remains blocked. In addition, the $Z^2$-FET has been used in ESD protection, which requires high $I_{ON}$, low $I_{OFF}$, gate-controlled $V_{ON}$ and fast response, showing excellent performance [81], [83].

Finally, the newly-invented $Z^3$-FET takes the geometry one step further and fully relies on the back gates without any front gate, see Fig. 17a [78]. In the $Z^3$-FET, there are two ground plane regions doped oppositely beneath the channel. These two ground planes are biased to keep their junction in reverse bias ($V_{GBP}<0$ and $V_{GBN}>0$, see Fig. 17a) and to form injection barriers in the channel above. The device performs very well with pronounced sharp switching and backgate-controlled hysteresis, as can be seen from the output characteristics in Fig. 17b. The design of the $Z^3$-FET, with back gates instead of front gates, has several advantages. The back gate control through the high quality and thick BOX instead of the front gate oxide can sustain higher voltages. This is especially useful in applications such as ESD protection and high power devices. Figure 17c shows the TLP test of $Z^3$-FET used in ESD protection, confirming high $I_{ON}$ in response to a short pulse. Besides, thanks to the free top surface, the $Z^2$-FET may be a candidate for ion- and light-sensing applications.

From the outset, TFETs were assumed to be the primary choice. However, the past few years have witnessed some incremental progress, but not more than that. The challenge is to increase the $I_{ON}$ current, by turning to alternative materials and enlarging the tunneling area via architectural modifications, and to maintain sub-$kT/q$ swing over a wide range of drain current. In this respect, the frequently reported record values of point subthreshold swing (measured at the most favorable $V_G$ bias, at very low $I_D$) do not constitute an informative figure of merit. The introduction of Ge and III-V semiconductors with narrow bandgap could certainly improve the $I_{ON}$, but the optimization of the swing still requires an independent approach: suppression of trap-assisted tunneling and junction inhomogeneities. The in-situ amplification of the tunneling current, such as proposed in the BET-FET structure, is appealing and deserves consideration in the clean room. The succession of two different mechanisms (tunneling first, then bipolar amplification) may hinder applications in the RF domain and needs further investigation.

Since TFETs are still far away from competitive with fully depleted MOSFETs (FDSOI or FinFET), a number of alternative avenues have materialized. A variety of astonishing devices – Landau switches, NEM-based FETs and relays, resistive-gate FETs, transistors with selective-injection mechanisms – have been proposed. They testify to bright creativity and deep mastery of device physics, but most will not overcome the veto from technologists. The association of an external switching element (ferroelectric capacitor or MIM resistor) is a hybrid solution that we contemplate with mixed feelings, as a compact single device is preferable.

On the other hand, band modulation devices are fully compatible with CMOS in FDSOI or FinFET versions. They exhibit experimentally measured abrupt switching and attractive features for memory, ESD protection and sensing applications. Low-power logic circuits require sharp switch in both ON-OFF and OFF-ON directions without hysteresis, a key issue still to be addressed.

One of these devices will come first but there is room for more. Maybe we just missed the best.

**REFERENCES**


SORIN CRISTOLOVEANU received the Ph.D. and Doctoratés-Sciences degrees from the Grenoble Institute of Technology, in 1976 and 1981, respectively. He has authored over 1100 contributions in the area of electrical characterization and modeling of semiconductor materials and devices. His work is mainly focused on SOI technology and innovative MOS devices for memory and sharp switching. He is the Director of Research CNRS, a fellow of ECS, a Distinguished Lecturer of the Electron Device Society, and an Editor of Solid-State Electronics.

JING WAN was born in Nanchang, China. He received the B.Sc. degree in electronic engineering from Nanchang University, China, in 2005, the M.Sc. degree in microelectronics from Fudan University, China, in 2009, and the Ph.D. degree in Nanoelectronics and Nanotechnology from the University of Grenoble, France, in 2012. His research interests cover the fabrication, electrical characterization, and TCAD simulation of semiconductor devices, including MOSFET, TFET, 2Z-FET, and other novel devices.

ALEXANDER ZASLAVSKY received the Ph.D. degree from Princeton University, USA, in 1991. He was a Post-Doctoral Scientist with IBM Research, Yorktown Heights, NY, USA, from 1991 to 1993. In 1994, he joined Brown University, USA, where he is currently a Professor of Engineering and Physics. He was a recipient of the Sloan Fellowship, and the NSF and ONR Young Investigator Awards. He is an Editor of Solid-State Electronics.