Assessment of InAs/GaSb Tunnel-FET Virtual Technology Platform for Low-Power Digital Circuits

Sebastiano Strangio, Pierpaolo Palestri, Senior Member, IEEE, Marco Lanuzza, Member, IEEE, Felice Crupi, Senior Member, IEEE, David Esseni, Fellow, IEEE, Luca Selmi, Fellow, IEEE

Abstract — In this work, a complementary InAs/Al0.05Ga0.95Sb tunnel field-effect-transistor (TFET) virtual technology platform is benchmarked against the projection to the CMOS FinFET 10-nm node, by means of device and basic circuit simulations. The comparison is performed in the ultra-low voltage regime (below 500 mV), where the proposed III-V TFETs feature on-current levels comparable with scaled FinFETs, for the same low-operating-power (LOP) off-current. Due to the asymmetrical n- and p-type I-Vs, trends of noise margins and performances are investigated for different Wp/Wn ratios. Implications of the device threshold voltage variability, which turned out to be dramatic for steep slope TFETs, are also addressed.

Index Terms— TFET, VLSI, III-V, Full-Adder.

I. INTRODUCTION

THE growing need for energy efficient electronics is calling for alternative device concepts, based on different operating principles and semiconductor materials with respect to the conventional silicon MOSFET. The tunnel field-effect-transistor (TFET) realized with III-V broken-gap/staggered heterostructures represents an interesting option due to its potential sub-60 mV/dec operation at suitable current levels [1-8]. Some experimental InAs/GaSb TFETs have already shown evidence of a relatively high on-current for reduced supply voltage (>100 μA/μm for VDD = 500 mV) [6-8], whereas the reported sub-threshold swing (SS) levels are still unattractive due to detrimental aspects related to the immaturity of the fabrication process [9-11]. Considering that atomistic full-quantum simulators can dependably predict the quantum effects underlying the TFET operation, sophisticated numerical simulations are widely used, as a cost effective alternative to device fabrication and characterization, in order to scrutinize the most suitable heterostructure materials and to optimize the key design parameters. In this context, a virtual TFET technology platform consisting of InAs/Al0.05Ga0.95Sb nanowires has been recently designed by Baravelli et al. [1-2].

The assessment of a technology can be carried out by evaluating device-level figures-of-merit (on- and off-currents for a certain VDD, SS, intrinsic capacitances, etc.), which can be translated through simplified models in circuit performances. A preliminary benchmark against the projections to the 10-nm node for CMOS FinFETs (i.e. the predictive technology models for the 10-nm node multi gate transistors [12,13]) has already been shown in [2]. The comparison was carried out by considering the static and dynamic behavior of an inverter, as obtained from a post-processing of the drain current characteristics and effective gate capacitance of the p- and n-type TFETs (Tφ and TN, respectively).

On the other hand, a comparison with conventional CMOS transistors, whose different operating mechanisms lead to different I-Vs and C-Vs trends, can lead to questionable conclusions if based exclusively on figures-of-merit for devices and/or inverters. A more systematic benchmark should include a circuit-level analysis with the evaluation of related static and dynamic figures-of-merit [14-19]. To this purpose, basic logic circuits such as inverters and ring-oscillators are employed in this work to investigate the effects of the p-/n-type device asymmetry and to identify the best Wp/Wn ratio. Then, the conventional 28T full-adder, identified as a relevant vehicle circuit representative of the digital logic environment, is analyzed in detail; performance and energy figures-of-merit, extracted for both TFET and FinFET implementations, are compared and discussed. Furthermore, the implications of device-to-device process variations are also considered, since they become a major concern in the ultra-low voltage scenario.

The reminder of this paper is organized as follows. Section II describes the simulation approach along with the calibration of the device models against full-quantum simulations; the simulation methodology for the digital circuits is presented too. Section III presents preliminary TFET simulations based on simple circuits, to address the p-/n-type device imbalance and to find a suitable Wp/Wn ratio. In Section IV, the simulation results on full-adders are presented along with the Vth variability effects. Finally, Section V concludes the paper.

II. SIMULATION METHODOLOGY

The simulation analysis has been carried out within the Cadence environment, by employing for the TFETs Verilog-A
compact models based on lookup tables (LUTs) with the device characteristics predicted by the full-quantum simulator [1,2]. Differently, -spice models [12,13] for the FinFETs were used. In the procedure for generating the LUTs, the drain current \( I_D \) and the gate capacitances \( C_{GS} \) and \( C_{GD} \) need to be computed for each biasing point of a discretized bidimensional domain represented by the \( V_{GS} \) and \( V_{DS} \) variables, where the widths of the \( \Delta V_{GS} \) and \( \Delta V_{DS} \) steps determine the tradeoff between computation burden and accuracy. In order to improve the computational efficiency in the generation of the \( I_D-V_{GS}-V_{DS} \), \( C_{GS}-V_{GS}-V_{DS} \) and \( C_{GD}-V_{GS}-V_{DS} \) LUTs and with marginal accuracy losses (voltage steps of 10 mV), we have used the TCAD simulator Sentaurus Device (adevice) [20] to reproduce the results from [1,2] through a fine calibration of the TCAD models, as described below.

### A. TCAD model calibration and Verilog-A compact model verification

The complementary InAs/AlGaSb TFETs proposed in [1,2] feature a 20 nm gate-length, a 7×7 nm² square cross-section and an EOT of 1 nm. The TFET design consisted in the careful selection of several key-aspects, such as the nanowire geometry (cross-section and source-gate-drain length), the material parameters (i.e. the Al mole fraction in the AlGaSb), and the doping levels. These parameters play a key-role in the device electrical behavior, since all of them affect the heterostructure band-diagram lineup, which is clearly crucial for the band-to-band-tunneling (BiBT) mechanism.

TCAD models have been recalibrated in order to fit the \( T_P \) and \( T_N \) transfer-characteristics of the full-quantum results, as reported in Fig.1a. Although such calibration was performed just by focusing on the \( I_D-V_{GS} \), the agreement between the TCAD simulations and the data from [2] is also satisfactory for the output-characteristics \( I_D-V_{DS} \) and gate capacitance characteristics \( C_{GG}(V_{GS}) \), as evidenced in Fig.1b \( (T_P \text{ and } T_N \text{ and } I_D-V_{DS}) \) and in Fig.1c-d \( (C_{GG}(T_P \text{ and of } T_N \text{, respectively})\).

The calibration\(^1\) was conducted in the following steps: (1) matching of the InAs/AlGaSb heterostructure band diagram, by adjusting the energy gap \( E_G \) and the electron affinity \( \chi \) to take into account for the effects of size-induced quantization [1], (2) calculation of the \( A_{\text{direct}} \) and \( B_{\text{direct}} \) constants for the dynamic non-local-path BiBT model [20] from the effective masses of bulk InAs and GaSb [21], (3) trimming of the effective valence/conduction band density of states \( N_V \) and \( N_C \) to improve the match (that is reasonable since the corresponding effective density of states under quantization is larger than in the bulk case for semiconductors with strongly non-parabolic energy dispersion).

Fig.2 compares the voltage-transfer-characteristics (VTC) of a TFET inverter and the switching current as a function of the input voltage simulated in Verilog-A with the one reported in [2], which have been calculated with the load-line method by using the \( T_P \) and \( T_N \) device output characteristics from the full-

\(^1\) Al\(_{0.48}\)Ga\(_{0.52}\)Sb (InAs) calibrated parameters: BiBT model: \( A = 1.51 \cdot 10^{20} \) \((1.44 \cdot 10^{-8})\) cm\(^3\) s\(^-1\), \( B = 9.54 \) (2.94) MV/cm. Effective density-of-states: \( N_C \) = \( 1.26 \cdot 10^{20} \) \((5.22 \cdot 10^{-17})\) cm\(^3\), \( N_V \) = \( 1.8 \cdot 10^{15} \) \((6.6 \cdot 10^{-11})\) cm\(^3\).

### B. Threshold voltage variability

Device-to-device variability is one of the major issues for circuits operating in the sub-threshold voltage regime. From the conventional MOSFET point of view, the various variability sources can be modeled together, considering that their combined effect leads to a variability of the threshold voltage \( V_{th} \) \((\text{V})\). The statistical variation of the FinFETs \( V_{th} \) can be estimated through the Pelgrom law [22], which relates the standard deviation of the \( V_{th} \) to the square root of the effective gate area of the device\(^2\). Concerning to TFETs, although a few theoretical studies have included the various process variation sources by means of statistical simulations at device level (including random dopant fluctuations (RDF), line

\(^2\) \( \sigma_{Vth} = 1.95 \cdot 10^{-10} \) \((W_{eff}L_{eff})^{-1/2} \)
edge roughness (LER) and metal gate work-function variation (VFW)) [23-28], due to the poor maturity of the TFET technology, the estimate of the device sensitivity obtained either by simulations or by experimental data is similarly impractical. For this reason, we have simply assumed the same \( V_{th} \) variability for both TFETs and FinFETs, which ensures a fair comparison between the two devices when performing the variability analysis at circuit level. In circuit simulations, the VFW has been modeled by adding a random voltage generator (with zero mean value) in series with the gate of each device, whose standard deviation is set to 35 mV (estimated for the FinFET architecture from (2), assuming \( \lambda_{vt} = 0.95 \text{ mV} \cdot \mu \text{m} \) [16,29,30]). It is worth mentioning that the same VFW affects the FinFETs and TFETs transfer-characteristics in a different way, due to their different shapes. The different SS (and \( g_{m}/I_{D} \)) of TFETs and FinFETs leads to a different sensitivity of the leakage [18] against a horizontal shift of the \( I_{D} \) vs \( V_{GS} \), which is a dramatic concern for TFETs due to their steeper characteristics at current levels close to \( I_{off} \). In Fig.3, the transfer-characteristics of the four devices under investigation are shown (\( T_{P} \) and \( T_{N} \) are the p- and n-type FinFET, respectively). Each figure shows the nominal curve (black solid line), the curves obtained from 1000 Monte-Carlo (MC) simulations by considering the VFW (grey lines) and the corresponding average value (dashed black line). In order to perform a comparison for similar leakage power, the nominal transfer characteristics (i.e. before activating the VFW), has been realigned so that the average value (\( \mu \)) of the \( I_{off} \) in

Fig. 3. Transfer-characteristics considering the \( V_{th} \) variability (VFW) of the (a) \( T_{P} \), (b) \( T_{N} \), (c) \( M_{P} \) and (d) \( M_{N} \). Compared to curves in Fig. 1a, the nominal \( I_{off} \) (\( V_{GS} \) @ \( I_{off,target} = 35 \text{ pA} \)) of the n(p)-type devices (i.e. without VFW) are preventively increased (decreased): \( V_{off}(T_{N}) = 45 \text{mV}, V_{off}(M_{N}) = 20 \text{ mV}, V_{off}(M_{P}) = 20 \text{ mV}, \) resulting in lower nominal \( I_{off} \) (\( I_{off} = 0 \text{V} \) and \( |V_{GS}| = 400 \text{ mV} \)). This \( I_{off} \) setup ensures an average \( I_{off} \) coinciding with the \( I_{off,target} \) when VFW is activated.

Fig. 4. (a) TFET and (b) FinFET inverter VTCs, for \( W_{P}/W_{N} = 1/1 \): nominal simulation (black line) and 1000 MC simulations (grey lines). (c) TFET inverter and (d) FinFET inverter variability of the VTC maximum gain (both for \( W_{P}/W_{N} = 1/1 \)). (e) TFET inverter noise margins (NMH and NML) as function of the \( W_{P}/W_{N} \) ratio (the ones of the FinFET inverter for \( W_{P}/W_{N}=1/1 \) are reported too -red symbols-). \( V_{DD} = 400 \text{mV} \).

presence of VFW is the same for all the devices, that is \( \mu(I_{off}) = 35 \text{ pA} \) at \( V_{DD} = 400 \text{mV} \), as shown in Fig.3.

III. \( W_{P}/W_{N} \) SIZING

Unlike the \( M_{P} \) and \( M_{N} \) FinFETs, which are essentially symmetric, the \( T_{P} \) and \( T_{N} \) TFETs feature asymmetric characteristics, given that the on-current is approximately 4 times larger for the \( T_{P} \) at \( V_{DD} = 400 \text{ mV} \) for a given transistor size. This explains the asymmetric TFET inverter VTC in Fig. 4a, as opposed to the FinFET one that is mirrored with respect to the line \( V_{OUT} = V_{IN} \) with a logic threshold at \( V_{DD}/2 \). For this reason, inverter-based circuits are studied in this section with focus on noise margins, performance and energy trade-offs for various \( W_{P}/W_{N} \) conditions.

A. Static noise margins

In Fig.4a-b, the VTCs obtained for 1000 MC instances are reported along with the nominal cases. They result in the probability density function (PDF) in Fig.4c-d, consisting of the values of the maximum voltage gain for the TFET and FinFET inverter at the logic threshold, respectively, and for \( W_{P}/W_{N} = 1/1 \). The TFET inverter features a larger gain (\( \mu=58.43, \sigma=17.32 \)), but also a larger variability than the FinFET one (\( \mu=12.43, \sigma=0.15 \)). This difference can be understood by considering the nominal VTCs in Fig.4a and b: for the TFET inverter, the transition from the high- to the low-state is very sharp only close to the logic threshold; this is not the case of the FinFET inverter, where the slope of the \( V_{OUT} \) (\( V_{IN} \)) curve is almost constant in the whole transition but relatively lower. Despite the larger gain, the noise margins
related to the high and low logic state (NMH and NML, respectively), are quite lower for the TFET inverter, as shown in Fig.4e. This issue is in part due to different Tp and TN I-Vs, and in part to the combined effects of the superliner output-characteristics at very low VDS (Fig.1b) and of the almost saturated transfer-characteristics at high current levels (Fig.1a), which lead to a wider transition for the TFET inverter VTC [14]. For this reason, although the NMH can be balanced with an adequate sizing (e.g. Wp/Wn=3/1), they still remain lower than the ones of the FinFET inverter with Wp/Wn=1/1 (Fig.4e).

B. Performance optimization

The basic logic blocks of Fig.5 have been simulated with nominal TFETs in order to investigate the performance trends for different Wp/Wn conditions.

The TFET inverter either is loaded by an equal stage (self-loading inverter in Fig.5a) or by 4 equal stages (FO4 inverter in Fig.5b). For these configurations, rise and fall times are extracted and summarized in Fig.6 as a function of the Wp/Wn ratio, by assuming near to ideal input signals (i.e. negligible rise/fall times). The rise and fall times are defined as the delay from the 10% to the 90% of the transition of the output waveform. For both loading conditions, a Wp/Wn ratio larger than 1 allows to reduce the rise time (the minimum rise time correspond to a Wp/Wn=2/1 and 3/1 for the self-loading and FO4 inverters, respectively). On the other hand, the average time monotonically increases with the increasing Wp/Wn ratio, due to the trends of the fall-transitions which are dominant. Clearly, such results depend on the input waveform, that is an ideal square waveform in this case, and thus the impact of the device input capacitance is not appropriately accounted for. In order to circumvent this issue, we have considered also the ring-oscillator (Fig.5c).

The critical path of a digital circuit sets a limit for the maximum frequency at which the circuit can operate. It usually depends on the device technology, the logic depth, the sizes of transistors, the load capacitance and the considered VDD. The ring-oscillator is conventionally used for benchmarking purposes, because the ratio between the oscillation period (Tosc) and the critical path delay of a generic circuit is, at a first order, independent from the technology, transistor sizing, temperature and VDD [31]. For this reason, we have used a 11-stage ring oscillator to optimize the Wp/Wn ratio with respect to the Tosc, which is strongly correlated with the critical path delay of a generic circuit, and to the energy per Tosc which in turn is correlated with the energy per operation when the same digital circuit is operated at the maximum frequency (for a particular VDD). In Fig.7a, the minimum Tosc corresponds to Wp/Wn=1/2, that is for symmetric Tp and TN capacitances (and not for symmetric currents), whereas the minimum energy in Fig.7b corresponds to the minimum device area (Wp/Wn=1/1), that is to the condition of minimum overall intrinsic capacitance.

In conclusion, although symmetric TFET drive-currents lead to improved noise margins, from the performance and energy consumption point of view, symmetric and as small as possible intrinsic capacitances are required.

IV. FULL-ADDCERS BENCHMARK

The conventional 28T full-adder [32,23] has been implemented exploiting TFET and FinFET solutions, and the two designs have been compared in terms of performance and energy figures-of-merit. The average energy per cycle as a function of VDD for both TFET and FinFET 32-bit ripple carry adders (RCAs) is also considered, where for each VDD the cycle is set to the minimum clock period limited by the critical path given by 31 tprop+max{tprop+sum} [32], where tprop and tsum are the propagation delay (carry-in to carry-out delay) and the
sum delay of the single-bit full-adder, respectively. Static and dynamic energy contributions are decoupled in order to be investigated independently. At high $V_{DD}$, the dynamic energy (proportional to $V_{DD}$) is dominant. When $V_{DD}$ decreases toward ultra-low voltage levels, despite the linear proportionality of the static power to $V_{DD}$, there is a severe increase of the static energy due to the longer critical path delay. As a result, the minimum energy per cycle corresponds to the $V_{DD}$ value where the dynamic and static components are well balanced.

Thus, although performance is not the main factor on which we should base the comparison of technologies operated in the ultralow voltage regime, it is anyhow convenient to investigate the trends of the critical path delay with $V_{DD}$, considering that it plays a fundamental role in the static energy consumption (proportional to the static power and to the duration of the cycle). For the sake of completeness, results obtained for nominal simulations [19] are also reported (Fig.8). The TFET full-adder becomes faster (i.e. lower propagation-delay) than the FinFET counterpart for $V_{DD}$ below ~350 mV (Fig.8a). In Fig.8b, the reduced performance degradation for the TFET circuit translates in a reduced increase of the static energy component for lower $V_{DD}$. This is the basic reason allowing the TFET circuit to reach a lower minimum energy point (135 aJ/cyc vs. 976 aJ/cyc) at a lower $V_{DD}$ (150 mV vs. 250 mV). In fact, for nominal simulations, the 32-bit RCA implemented with TFETs is ~7.23x more energy efficient than the FinFET counterpart when both operate at the $V_{DD}$ corresponding to the minimum energy point [19].

### A. Evaluation of the minimum $V_{DD}$

As opposite to symmetric M$_p$ and M$_n$ FinFETs, the $T_P$-$T_N$ asymmetry may affect the correct operation of the TFET full-adder at low $V_{DD}$ when the VtV is considered. Thus, for each of the 8 possible input combinations (i.e. $\{A,B,C\} = \{0/1,0/1,0/1\}$), we have simulated 100 MC instances in order to evaluate the minimum $V_{DD}$. These points have been used to obtain the probability density functions (PDFs) in Fig.9, where the "μ+3σ" values delineate the boundary between a region with a high error probability and a safe operating region. Interestingly, despite the $W_P/W_N$ asymmetry at $V_{DD}$ = 400mV, the distributions of the minimum $V_{DD}$ for TFET and FinFET full-adders are very similar (with practically the same "μ+3σ" boundary). For this reason, we have identified the $W_P/W_N$=1/1 condition as the most suitable for the rest of the analysis. This choice is also supported by argument that the lower energy condition is met for the 1/1 ratio, as shown in section III-B for the ring oscillator test circuit.

### B. Performance and energy degradation with VtV

The same figures-of-merit in Fig.8 have been evaluated by including the VtV (with the $I_{off}$ realigned as shown in Fig.3), and then extracting both the mean (μ) values as well as the "μ+3σ" values for the related performance metric. From the propagation delay point of view, the variability is strongly related to the sensitivity of the on-current to the VtV, which is less variable for the TFET circuit for almost the entire range of investigation as evidenced by Fig.10: by comparing the "μ" with the "μ+3σ" lines, it is easy to relate the propagation delay variability with the normalized transconductance ($g_{m(D,ON)}$) at the corresponding $V_{DD}$. In fact, conversely to FinFETs, which are always in the sub threshold region in the investigated voltage range, the almost saturated $I_{DS}$-$V_{GS}$ of TFETs at large $V_{GS}$ directly translates in a lower variability of the delay.
On the other hand, the TFET steep characteristics at low $V_{GS}$ becomes dramatic in terms of leakage variability, which affects the static component of the energy per cycle, as shown in Fig.11a. The variability of the dynamic energy components in Fig.11b of both TFET and FinFET implementations is practically negligible if compared to the ones of the static energy, since they are related just to the variability of device capacitance characteristics ($E_{DD} \propto C_{eq} V_{DD}^2$). Fig.11c shows the effect of VtV on the overall energy per cycle. Considering the average values, the TFET RCA has a minimum energy point of 263 aJ/cyc at 200 mV (note that the $V_{DD}$ cannot be further scaled down due to variability issues, as demonstrated by Fig.9), whereas the one of the FinFET implementation is 1194 aJ/cyc at 300 mV. This means that the energy improvement that is achieved with the TFET RCA is of $-4.54x$ with respect to the case with FinFETs (this factor is lower than the one estimated from nominal simulations [19]). When considering the "$\mu \pm 3\sigma$" trends, the energy saving with the TFET circuit further decreases, particularly due to the strong variability of the TFET static energy, that leads to a rightward shift of the minimum energy point (572 aJ/cyc at 300 mV), whereas the minimum energy point for the FinFET case is still at 300 mV (as for the "$\mu$" case) but increases to 1440 aJ/cyc.

In summary, the VtV tends to lower the energy improvements of TFET circuits when compared to their FinFET counterparts. This basically means that research efforts for steeper and steeper devices should be accompanied by improvements of the reproducibility of fabricated device characteristics, considering that the VtV becomes an increasing issue with the increasing $\mu_{eff} I_{D}$ (i.e. with the decreasing SS) at current levels close to the $I_{off}$ target.

It is worth noting that these results have been achieved with the devices transfer-characteristics aligned at the same $\mu_{(I_{off})}$, performed at the beginning of this study to partially counteract the different sensitivity of the TFET and FinFET leakage with respect to the VtV. Fig.12 shows the same plot of the energy per cycle versus $V_{DD}$ as in Fig.11c, but considering the device transfer-characteristics aligned at the same $I_{off}$ for nominal TFET and FinFET devices, but different $\mu_{(I_{off})}$ when the VtV is considered. In this case, the TFET advantages in terms of energy are almost completely lost due to the larger impact of the leakage variability for TFET devices.

V. CONCLUSION

In this study, we have compared two virtual complementary platforms (namely, the $T_P$ and $T_N$ heterojunction III-V TFETs and the $M_P$ and $M_N$ FinFETs) in the ultra-low voltage regime, with emphasis on the implications of the device $V_{th}$ variability on the dynamic and static characteristics of digital circuits. Unlike the $M_P$ and $M_N$ FinFETs, the $T_P$ and $T_N$ TFETs feature an asymmetry in the drive-current. This limit can be partially addressed by a careful sizing of the $W_P/W_N$ ratio, at least from the noise margins point of view. Nevertheless, from the speed and energy point of view, effects as the unbalancing of the $T_P$-$T_N$ intrinsic capacitances and the increase of the overall effective capacitance, tend to be more detrimental than the improvements which are obtained thanks to the balanced drive-currents. The 32 bit ripple carry adders designed using 28T full-adder blocks, have been used to benchmark the TFET and FinFET technologies, by considering the same condition of variability for the $V_{th}$. Due to different slope of the turn-on characteristics, it is recommended to preventively consider the $V_{th}$ variability, by aligning the device $I_{D2}$-$V_{DS}$ so as to ensure an iso-leakage condition when the VtV is considered (same average value of the $I_{off}$). This guideline allows to keep the static energy below the dynamic energy component at higher $V_{DD}$, so as that the TFET implementation can reach a lower minimum energy point (as for nominal simulations). Even so, by assuming the same $V_{th}$ variability, the energy improvement which can be obtained with TFETs tends to be lower than the one estimated from nominal simulations. This means that a steep device intrinsically needs a superior control of process variation items in order to completely exploit the possible advantages related to the low SS.