Process-Induced Power-Performance Variability in Sub-5nm III-V Tunnel FETs

Yang Xiang, *Student Member, IEEE*, Anne S. Verhulst, Dmitry Yakimets, *Member, IEEE*, Bertrand Parvais, Anda Mocuta and Guido Groeseneken, *Fellow, IEEE*

Abstract—We examine the power-performance variability of a projected sub-5nm GaAsSb/InGaAs vertical tunnel FET considering various process control tolerances in the state-of-the-art device integration and propose countermeasures in device design. Nominal and three-sigma-corner device characteristics generated in quantum-mechanical/TCAD simulations are used to calibrate a semi-empirical compact model, based on which the nominal and variability-inclusive energy-delay landscapes are extracted from ring-oscillator circuit simulations at sub-500mV supply voltages. Variations in four parameters are identified as of major impact on the worst-case speed loss and iso-speed energy penalty: dopant pocket thickness, gate work function, hetero-band offset and body thickness (in descending order). Variability-resilient device options are explored against pocket thickness variation, including 1) pocket de-sensitization with increased thickness and reduced doping concentration; 2) broken-gap tunnel FET with a negative effective band gap. Re-engineered devices achieve $< 18\times$ speed loss and $< 3\times$ energy penalty for (0.1 – 1) ns gate delay with respect to the nominal corner.

Index Terms— Circuit simulation, heterojunctions, III-V semiconductor materials, logic circuits, tunnel FETs (TFETs), power-performance, process variability

I. INTRODUCTION

REDUCING energy consumption has become a major challenge in present day's MOSFET scaling due to the difficulty in supply voltage (V_{DD}) reduction and leakage current (I_{OFF}) suppression [1][2]. Among the proposed solutions, tunnel FETs (TFETs) with III-V materials based heterojunction structures ("hetero-TFETs") [3][4] have shown the promise of combining the desired sub-60 mV/dec subthreshold swing (SS) and high on-current (I_{ON}) in experimental devices [5][6][7], which has further translated to an outstanding energy efficiency in circuits/systems from predictive simulations [8][9].

Given the band-to-band tunneling (BTBT) mechanism and steep slope of TFETs [2] that hint a sensitive response of current to voltage, the power-performance (PP) variability of TFETs merits close attention seeing the multitude of parameter variations in aggressively scaled devices [10], including global

parameter deviation due to finite processing tolerance (*process*) and local statistical parameter fluctuations (*statistical*) [11]. Particularly, study of *process*-induced PP variability provides guidelines to process control budget in device integration and insights to parameter optimization in device design; moreover, it enables accurate modeling of *statistical* variability which depends highly on the former at scaled dimensions [11].

To date, while device-level variability in III-V (hetero-) TFETs has been extensively explored [12][13][14][15], PP variability study is only reported for a few scattered *statistical* parameter variations [14][15]. Meanwhile a comprehensive *process*-induced PP variability is yet lacking, which becomes especially pressing considering the paradigm shift in device integration driven by various scaling boosters in TFETs such as confined vertical nanowire [6][7] and dopant pocket [5][6].

Here we report on a predictive PP variability study of a sub-5nm-node GaAsSb/InGaAs vertical p-n-i-n TFET over comprehensive *process* variations. The paper is organized as follows. In section II we introduce the device structure, nominal device characteristics, *process* variation assumptions and device-level impact. In section III we elaborate on the methodology in propagating parameter variations to the energy-delay relations in a ring-oscillator circuit and quantify the impact of each parameter on the worst-case speed loss and iso-speed energy penalty. In section IV, we propose device design options against the most sensitive parameter, pocket thickness, namely a re-engineered pocket design in the GaAsSb/InGaAs TFET, as well as alternative material systems with a broken-gap band alignment such as GaSb/InAs heterojunction.

II. DEVICE BACKGROUND AND DEVICE-LEVEL VARIABILITY

A. Nominal Device and Characteristics

The nominal GaAsSb/InGaAs n-TFET (Fig. 1(a)) has a staggered-gap (StG) band alignment and a vertical architecture [3][16], which is fully optimized at device level with its I_{60} , the maximum current attainable at sub-60-mV/dec-SS [3][4] reaching $9.2 \mu\text{A}/\mu\text{m}$ at $V_{DD} = 0.3 \text{ V}$ (Fig. 2). The complementary p-TFET is assumed to be an electrically mirrored n-TFET, as a symmetric n-p-i-p pTFET does not simultaneously deliver the desired low SS and high I_{ON} due to the large source degeneracy at high doping concentration [14]; this allows to focus exclusively on the impact of process variability and exclude

Manuscript received March 3, 2019. This work was supported by imec's Industrial Affiliation Program.

Y. Xiang and G. Groeseneken are with imec, Leuven 3001, Belgium and the Department of Electrical Engineering, KU Leuven, Leuven 3001, Belgium (e-mail: Yang.Xiang@imec.be; Guido.Groeseneken@imec.be).

A. S. Verhulst, D. Yakimets and A. Mocuta are with imec, Leuven 3001, Belgium (email: Anne.Verhulst@imec.be; Dmitry.Yakimets@imec.be; Anda.Mocuta@imec.be).

B. Parvais is with imec, Leuven 3001, Belgium and Vrije Universiteit Brussel, Brussels 1050, Belgium (email: Bertrand.Parvais@imec.be).

Fig. 1. (a) Top: schematic of the cross section of the vertical TFET; bottom: the band alignment of the bulk heterojunction. X_g , L_g , L_s , L_{ch} , L_d , t_{po} , t_{body} , N_s , N_{po} , N_d , EOT and $E_{g,eff}$ denote gate position (from top electrode), gate length, source extension length, channel length, drain extension length, dopant pocket thickness, body thickness, source doping, pocket doping, drain doping, equivalent oxide thickness and effective band gap (in bulk), respectively. (b) Top: the layered layout of a complementary-TFET inverter cell showing the $49 \times 10 \text{ nm}^2$ NSH channel and the cell ground rules: cell height (CH) = 10 metal tracks (10 tr.), contact gate pitch (CGP) = 32 nm, metal pitch (MP) = 24 nm; bottom: schematic of device parasitic capacitance and resistance model up to the first metal layer M1 (not drawn to scale).

factors related to the lack of perfect p-n symmetry. A nanosheet (NSH) channel is chosen to maximize drive strength [17]. The standard cell ground rules (Fig. 1(b)) enable a similar cell area to that of a 3nm-node lateral Si/SiGe FinFET [18]¹; the latter will be referred to as “3nm MOSFET” in the paper.

A full-zone 15-band two-dimensional quantum mechanical (QM) solver [16] is used to calculate the BTBT current in the TFET, with the input of electrostatic potential generated by the TCAD tool Synopsys SDevice [19] using a calibrated density-gradient model; the terminal charges are also calculated based on SDevice. A semi-empirical compact model (CM) of TFET [20] is then calibrated to the I-V and Q-V characteristics simulated by QM/TCAD (see Fig. 2), on top of which a three-dimensional device parasitic RC model [21] is used to capture the characteristics of the full device (Fig. 1(b)). It turns out that while the impact of parasitic resistance is negligible, the device capacitance is increased by $\sim(1.5 - 2) \times$ due to the parasitic capacitance (Fig. 2).

B. Process Variation Assumptions

Table I gives the 3σ values of device parameters in Fig. 1(a) under various global *process* variations (assuming a normal distribution), which are based on the learnings from our in-house integration facilities and are representative of state-of-the-art process control capabilities. In particular, the vertical dimensions t_{po} , L_{ch} , L_g and X_g are defined in terms of film/layer thickness and subject to its nonuniformity across the wafer. Note that local *statistical* variations (Pelgrom mismatch [14], random dopant effects [10][14], line edge roughness [14][15], etc.) that exist on top of *process* variations are beyond the scope

Fig. 2. I_{DS} - V_{GS} (left) and C_{GG} - V_{GS} (right, derived from Q_G - V_{GS}) characteristics of the TFET in Fig. 1 (nominal) and of the 3nm MOSFET ([18], nominal). The TFET is simulated by “QM/TCAD” (quantum-mechanical solver/TCAD), “CM (intrinsic)” (compact model [20]) and “CM (device)” (compact model [20] with parasitic RC [21] per Fig. 1(b)). The MOSFET is simulated by a compact model with device parasitic RC [18]. All I-V and C-V curves are shifted to $I_{OFF} = 50 \text{ pA}$ @ $V_{DS} = 0.3 \text{ V}$. The TFET shows an $I_{60} = 0.45 \mu\text{A}$ ($= 9.2 \mu\text{A}/\mu\text{m}$) at $V_{DS} = 0.3 \text{ V}$.

of this paper and left for follow-up studies. Furthermore, for a full picture of TFET nonidealities, possible parasitic conduction (trap assisted tunneling [22], band tails [23], etc.) that potentially degrade the desired steep slope are yet to be included based on well-calibrated parameter assumptions (e.g., trap density, density of band-tail states) with respect to experimental data.

C. Device-Level Variability

Since the semi-empirical compact model [20] does not support precise physical scalability, full Monte-Carlo circuit simulation is generally not available (except for Φ_g variation²).

TABLE I
PROCESS PARAMETERS: NOMINAL AND $\pm 3\sigma$ CORNER VALUES

Parameter	Defining process step	Nominal value	$\pm 3\sigma$ corner value
t_{po} (nm)	ALD ^a or CVD ^b	2.5	± 1
EOT (nm)	ALD ^a	0.6	± 0.2 ^c
L_{ch} (nm)	CVD ^b	28	± 2 @ source (“ $\Delta L_{ch@S}$ ”) ^d ± 2 @ drain (“ $\Delta L_{ch@d}$ ”) ^d
L_g (nm)	Deposition + etch	16	± 2 @ source (“ $\Delta L_{g@S}$ ”) ^d ± 2 @ drain (“ $\Delta L_{g@d}$ ”) ^d
X_g (nm)	Deposition + etch	32	± 2
t_{body} (nm)	EUV Double Patterning	10	± 1
N_s (cm ⁻³)	In-situ doping	1×10^{20}	± 20 %
N_{po} (cm ⁻³)	In-situ doping	1×10^{20}	± 20 %
N_d (cm ⁻³)	In-situ doping	5×10^{18}	± 20 %
$E_{g,eff}$ (eV; bulk)	GaAsSb epitaxy (CVD ^b)	0.27	± 0.05 ^e
Φ_g (eV) ^f	Metal tuning in RMG ^g	($\Phi_{g,0}$) ^f	± 0.07

^aALD: atomic layer deposition.

^bCVD: chemical vapor deposition.

^cFrom $\pm 1 \text{ nm}$ physical thickness variation of high-k dielectrics (dielectric constant ~ 20).

^d L_{ch} and L_g lengthening or shortening at either the source side or the drain side, hence the different cases of L_{ch} and L_g variations.

^eFrom ± 0.05 in GaAs_{1-y}Sb_y and/or In_xGa_{1-x}As stoichiometry change and the resultant band structure change and local strain.

^f Φ_g : gate work function, the nominal value of which (“ $\Phi_{g,0}$ ”) is determined after specifying the nominal I_{OFF} (or equivalently V_t) at a given V_{DD} , hence not a fixed number.

^gRMG: replacement metal gate, during which Φ_g is determined by the thickness of work function metal layers (“metal tuning”).

¹ Inverter cell area: this vertical TFET = $10 \text{ tr.} \times 24 \text{ nm} \times 32 \text{ nm} = 0.0077 \mu\text{m}^2$, 3nm FinFET in [16] = $5.5 \text{ tr.} \times 21 \text{ nm} \times 42 \text{ nm} \times 2 = 0.0097 \mu\text{m}^2$, where the factor 2 comes from the dummy gate in lateral FinFET.

² Φ_g variation can be simulated by generating random Φ_g values in Monte-Carlo circuit simulations, but this is not used for drawing conclusions in order to keep the methodology consistent.

To avoid inaccuracy in device characteristics extrapolation, separate QM/TCAD simulations are performed for the $\pm 3\sigma$ corner cases³. Hence for each parameter variation, two additional model cards are generated for the $\pm 3\sigma$ corner devices after re-calibrating the model to QM/TCAD simulation results.

Fig. 3 illustrates the device-level I-V and C-V variability associated with the following parameter variations: Δt_{po} , Δt_{body} , $\Delta \Phi_g$, $\Delta E_{g,eff}$, $\Delta L_{ch@S}$ and ΔX_g . For parameters not plotted in Fig. 3, $\Delta L_{g@S}$ quantitatively resembles $\Delta L_{ch@S}$ in C-V variability but shows only minor influence on I-V; while ΔEOT , $\Delta L_{ch@d}$, $\Delta L_{g@d}$, ΔN_s and ΔN_d are found of negligible impact. ΔN_{po} is expected to alter I-V in a similar style to Δt_{po} , namely by changing the built-in field at the junction, but quantitatively it is estimated that the effect of ± 1 nm Δt_{po} from $t_{po} = 2.5$ nm is comparable to $\pm 75\%$ ΔN_{po} , as the potential drop across the exactly fully-depleted pocket [4] is more sensitive to t_{po} ; consequently the effect of $\pm 20\%$ ΔN_{po} can be represented as a reduced effect of ± 1 nm Δt_{po} , and we hence choose to probe the overall pocket variability by studying ± 1 nm Δt_{po} without performing dedicated simulations on ΔN_{po} .

Clearly, Fig. 3 indicates that I-V variability constitutes the main aspect of device-level variability and C-V variability is less prominent. In detail, we observe that:

- 1) I-V variability (except for Δt_{po}) commonly appears as the shift of threshold voltage V_t symmetrically in the $\pm 3\sigma$ corners, which alters I_{OFF} without significantly modifying SS and shows diminishing impact at increased V_{GS} ; a maximum increase in I_{OFF} is found in $\Delta \Phi_g$ ($\sim 300\times$). For Δt_{po} , the V_t shift is lopsided to the -3σ corner; meanwhile the slope degrades in both $+3\sigma$ and -3σ corners, leading to a notable -3σ corner I_{ON} degradation ($\sim 20\times$ at $V_{DD} = 0.3$ V). This indicates the weakened junction band bending due to insufficient t_{po} at a given N_{po} and the resultant increase in the tunneling path length.
- 2) C-V variability is less pronounced compared to I-V, with a maximum of 15 % due to $\Delta L_{ch@S}$ or $\Delta L_{g@S}$ (not shown). Disproportionate variabilities in I-V and C-V with a given parameter (e.g., Δt_{body} impacts noticeably on I-V but marginally on C-V; ΔX_g as the opposite case) result from

Fig. 3. Simulated I_{DS} - V_{GS} ($V_{DS} = 0.3$ V) and C_{GG} - V_{GS} ($V_{DS} = 0.05$ V) curves corresponding to the $\pm 3\sigma$ corners of selected process variations $\{\Delta t_{po}, \Delta t_{body}, \Delta \Phi_g, \Delta E_{g,eff}, \Delta L_{ch@S}, \Delta X_g\}$. The nominal $I_{OFF} = 50$ pA @ $V_{DD} = 0.3$ V, which corresponds to $\Phi_{g,0} = 4.461$ eV; this applies to both I_{DS} - V_{GS} and C_{GG} - V_{GS} curves.

³ Except for $\pm 3\sigma$ corners of Φ_g , which are equivalently done by offsetting the I_{DS} - V_{GS} and C_{GG} - V_{GS} characteristics by ± 70 meV per the assumption in Table I.

different contributions of parasitics in device I-V and C-V characteristics (Fig. 2): device I-V variability is channel dominated, while device C-V variability is a net result of terminal charge variation and parasitic capacitance alteration due to device dimension variation which do not scale in proportion.

III. CIRCUIT-LEVEL POWER-PERFORMANCE VARIABILITY

A. Nominal Circuit-Level Power-Performance

The power-performance of the TFET is simulated in Cadence Spectre [24] on a 15-stage ring oscillator (RO) circuit made of fan-out-3 inverter gates (INV-FO3), with an interconnect wire load of $50\times$ CGP (wire resistance ~ 700 Ω , capacitance ~ 0.4 fF, interconnect RC delay $\tau_{intc} \sim 0.3$ ps). This benchmark is chosen as representative of a critical path in applications like system-on-chip (SoC) [25] which requires good energy efficiency; PP metrics of TFET in this test case is thus deemed highly relevant for both nominal and variability situations.

The nominal PP of TFET is studied similar to [14] (Fig. 4). First, V_t is swept as a free variable by tuning the nominal Φ_g value, “ $\Phi_{g,0}$ ”, to target a nominal off-current “ $I_{OFF,0}$ ” between 1 pA and 10 nA at a given V_{DD} between 0.1 V and 0.5 V. Next, the total energy consumption E is calculated in accordance with the logic activity factor (AF ; 10 % and 0.1 % chosen here), i.e.,

$$E = AF \cdot E_{dyn} + E_{stat} = (AF \cdot P_{dyn} + P_{leak}) \cdot \tau \quad (1)$$

where E_{dyn} , E_{stat} , P_{dyn} , P_{leak} and τ stand for “dynamic energy”, “static energy”, “dynamic power”, “leakage power” and “delay”, respectively (all normalized to a single RO stage).

Finally, the energy-delay landscape is obtained by applying Pareto optimization [14] to all swept solutions which gives the minimum energy at a given delay (“Pareto-landscape”); meanwhile a parallel exercise is done on a 3nm-MOSFET-based [18] INV-FO3 ($\tau_{intc} = 0.4$ ps) (Fig. 4).

Consistent with [8][25], TFET is found to have remarkable iso-performance energy saving than MOSFET at low speed and low AF (up to $\sim 10\times$ at $\tau = 1$ ns) where leakage energy

Fig. 4. The Pareto-optimized energy-delay landscapes (solid envelope curves) for the TFET in Fig. 1 and the 3nm MOSFET [18] at $AF = 10\%$ (left) and 0.1% (right); all solutions on the Pareto landscapes fulfill the constraint noise margin ($NM > 0$) to guarantee correct logic operation; wire load (W.L.) is $50\times$ CGP (CGP = 32 nm for TFET, 42 nm for MOSFET [18]). The scattered dots denote solutions to $(V_{DD}, I_{OFF,0})$ sweep: $V_{DD} \in [0.10, 0.50]$ V, $I_{OFF,0} \in [10^{-12}, 10^{-8}]$ A. The V_{DD} and $I_{OFF,0}$ values at $\tau = 100$ ps and 1 ns are shown for both MOSFET and TFET.

Fig. 5. Illustration of the simulated delay triplet $(\tau_0, \tau_{+3\sigma}, \tau_{-3\sigma})$ (left panel) and energy triplet $(E_0, E_{+3\sigma}, E_{-3\sigma})$ (right panel) due to $\Delta\Phi_g$ at $V_{DD} = 0.3$ V, $I_{OFF,0} = 50$ pA (accordingly $\Phi_{g,0} = 4.461$ eV), $\pm 3\sigma = \pm 70$ meV; the probability density function f of the Gaussian distribution of $\Delta\Phi_g$ is plotted on top. The histograms under f represent normalized local population inside the bins centered at $\pm 3\sigma, \pm 2.5\sigma, \dots, \pm 0.5\sigma, 0$. Energy pertaining to bins centered at $\pm 2.5\sigma, \pm 2.0\sigma, \dots, \pm 0.5\sigma$ are interpolated from simulated energy $(E_0, E_{+3\sigma}, E_{-3\sigma})$ in a logarithmic manner. The variability-inclusive delay τ_{var} and energy E_{var} are calculated per the formulae in the insets.

dominates – thanks to its steep slope that enables ultra-low I_{OFF} . Accordingly, for PP variability study the hotspot of TFET, namely, $\tau \sim (0.1 - 1)$ ns at low AF situation (0.1 %) will be the focus given its relevance for projected applications.

B. Variability-Inclusive Delay and Energy

For any arbitrary $(V_{DD}, I_{OFF,0})$ pair from the sweep, the deviated delay and energy due to variation in parameter Y can be estimated by substituting the $\pm 3\sigma$ -corner model cards for the nominal card in RO simulations, where all instances in the RO assume identically “fast” or “slow” TFETs considering the global, inter-chip range of *process*-induced variations that remain invariant across a single chip. This gives a delay-/energy-triplet $(\tau_0, \tau_{+3\sigma}, \tau_{-3\sigma})$ and $(E_0, E_{+3\sigma}, E_{-3\sigma})$ (see Fig. 5).

Next, we construct the variability-inclusive metrics (τ_{var}, E_{var}) at a given $(V_{DD}, I_{OFF,0})$ with respect to ΔY as follows:

- 1) Performance-wise, τ_{var} takes the maximum of $(\tau_0, \tau_{+3\sigma}, \tau_{-3\sigma})$ as the gate-level delay target must be met even in the worst case (Fig. 5(a)).
- 2) Energy-wise, consider the global distribution of RO characteristics due to different Y values from chip to chip, we report an approximated statistical mean energy E_{var} where parameter variation ΔY obeys a normal distribution $\Delta Y \sim N(0, \sigma^2)$ with a probability density function f ; here binning is used to approximate the integral (Fig. 5(b)).

It is worth mentioning that despite the intention of using E_{var} to approximate a “statistically averaged energy”, and while $E_0 < E_{var} < \max(E_0, E_{+3\sigma}, E_{-3\sigma})$, the logarithm-style interpolation from $(E_0, E_{+3\sigma}, E_{-3\sigma})$ in Fig. 5 might have overestimated the energy consumption, especially if the simulated energy triplet $(E_0, E_{+3\sigma}, E_{-3\sigma})$ happens to be biased towards the high-energy end as in Fig. 5. Nevertheless, the (τ_{var}, E_{var}) pairs are only indicative of the variability-inclusive metrics of an INV gate in an application-representative (e.g., wireload, FO) benchmark, while the critical path delay and switching energy in digital blocks depend also on other design details, e.g., logic depth [9].

C. Variability of Energy-Delay Landscape

Below we assess PP variability of the nominal Pareto-

Fig. 6. Illustration of the workflow to include variability in the energy-delay landscape. “Nominal Pareto”: the nominal Pareto-landscape of the TFET in Fig. 4 ($AF = 0.1\%$); “nominal $\{(V_{DD}^{(P)}, I_{OFF,0}^{(P)})\}$ ”: the $(V_{DD}, I_{OFF,0})$ combinations that give “nominal Pareto”, where $I_{OFF,0}^{(P)}$ contains the information of its corresponding $\Phi_{g,0}$; “corrupted Pareto”: the ensemble of (τ_{var}, E_{var}) taken at each nominal $(V_{DD}^{(P)}, I_{OFF,0}^{(P)})$; “variability-inclusive Pareto”: the envelope obtained from Pareto-optimization of $\{(\tau_{var}, E_{var})\}$ across the entire $\{(V_{DD}, I_{OFF,0})\}$ sweep, which are operated at their respective $\{(V_{DD}^*, I_{OFF,0}^*)\}$.

landscape ($AF = 0.1\%$) of the TFET in two aspects (Fig. 6):

- 1) *Worst-case speed loss*. This is done by comparing τ_{var} and τ_0 at a certain $(V_{DD}^{(P)}, I_{OFF,0}^{(P)})$ on “nominal Pareto” and calculate the speed loss due to variability, which is visualized as the shift to “corrupted Pareto” in Fig. 7.

Table II shows that the speed loss $\tau_{var}^{(P)}/\tau_0^{(P)}$ due to various parameter variations, with the four most significant sensitivity being $\Delta t_{po} \gg \Delta\Phi_g \gg \Delta t_{body} \geq \Delta E_{g,eff}$ (Fig. 7). In particular, the $\sim 10^3\times$ speed loss due to Δt_{po} poses potential timing violations in synchronous circuits that can lead to circuit functional failures [26], which challenges the current process control budget and demands more stringent, sub-nm tolerance during ALD or CVD pocket growth for higher yield. In contrast, other vertically defined dimensions such as $L_{ch@S}$ show lesser sensitivity ($< 2\times$) and are not of major concern. Meanwhile, Table II indicates that speed loss is dominated by effective resistance (R_{EFF}) increase that characterizes drive current loss; this originates from the notable device I-V variability at ultra-low V_{DD} (< 200 mV) and low-to-medium $I_{OFF,0} \sim (10^0 - 10^1)$ pA (Fig. 3). In comparison, the impact of node effective capacitance (C_{EFF}) is marginal as it varies little ($\leq \pm 10\%$ in -3σ

TABLE II
SUMMARY OF POWER-PERFORMANCE VARIABILITY OF PROCESS PARAMETERS

Variation source	Speed loss $\tau_{var}^{(P)}/\tau_0^{(P)}$ $@\tau_0^{(P)} =$	Max. R_{EFF} increase $@\tau_0^{(P)} =$	Energy penalty $E_{var}^*/E_0^{(P)}$ $@\tau_T =$	Operating point V_{DD}^* (mV) ^a $@\tau_T =$
	0.1ns 1ns (1x) (1x)	0.1ns 1ns (1x) (1x)	0.1ns 1ns (1x) (1x)	0.1ns 1ns (188) (125)
(nominal)				
Δt_{po}	410x 6300x	430x 5400x	13.3x 5.2x	500 275
$\Delta\Phi_g$	9.1x 110x	9.5x 110x	2.7x 3.2x	300 212
$\Delta E_{g,eff}$	2.2x 5.3x	2.3x 5.5x	1.7x 1.6x	225 150
Δt_{body}	2.7x 5.5x	2.7x 5.7x	1.6x 1.5x	225 150
$\Delta L_{ch@S}$	1.4x 1.9x	1.4x 2.0x	1.2x 1.2x	200 138
$\Delta L_{ch@d}$	1.1x 1.3x	1.1x 1.3x	1.0x 1.1x	188 125
$\Delta L_{g@S}$	1.0x 1.1x	1.0x 1.0x	1.0x 1.0x	188 125
$\Delta L_{g@d}$	1.3x 1.7x	1.2x 1.6x	1.1x 1.2x	200 138
ΔX_g	1.2x 1.5x	1.2x 1.5x	1.1x 1.1x	188 125
ΔEOT	1.2x 1.6x	1.2x 1.6x	1.1x 1.2x	188 138
ΔN_s	1.1x 1.1x	1.1x 1.1x	1.0x 1.0x	188 125
ΔN_d	1.0x 1.0x	1.0x 1.0x	1.0x 1.0x	188 125

^aSince V_{DD} is swept in finite steps of 12.5 mV and rounded to integers of mV, V_{DD} change smaller than 13 mV is not detected.

Fig. 7. The “nominal Pareto”, “corrupted Pareto” and “variability-inclusive Pareto” pertaining to parameters giving the most visible variability: $\{\Delta t_{po}, \Delta t_{body}, \Delta \Phi_g, \Delta E_{g,eff}\}$. Orange arrows indicate the speed loss $\tau_{var}^{(P)}/\tau_0^{(P)}$ at $T_0^{(P)} = 0.1$ ns and 1 ns at their respective $(V_{DD}, I_{OFF,0})$ point: 0.1 ns @ $(V_{DD} = 188$ mV, $I_{OFF,0} = 56$ pA), 1 ns @ $(V_{DD} = 125$ mV, $I_{OFF,0} = 5.6$ pA). Magenta arrows indicate the energy penalty $E_{var}/E_0^{(P)}$ that is needed to reach target delay $\tau_T = 0.1$ ns and 1 ns.

corner) due to the dominant non-varying wire load capacitance (~ 0.4 fF) over the FO3 gate input capacitance ($\sim 3 \times 0.05$ fF).

2) *Iso-speed energy penalty*. Here a “variability-inclusive Pareto” (Fig. 7) is constructed by Pareto-optimizing all (τ_{var}, E_{var}) pairs obtained in the $(V_{DD}, I_{OFF,0})$ sweep for a given parameter (Fig. 6), and we compare the energy increase with respect to the “nominal Pareto” at a certain target delay τ_T (Table II).

Fig. 7 shows that “variability-inclusive Pareto” always lies at the left-lower envelope of “corrupted Pareto” and rejects non-minimum-energy solutions at a given delay, indicating that the former gives the attainable most energy-efficient solutions in the presence of variability. In terms of iso-speed penalties, Fig. 7 and Table II show that trend in energy and V_{DD} penalty across various parameter variations are in line with their respective sensitivities in speed loss, with the most significant four being $\Delta t_{po} \gg \Delta \Phi_g > \Delta E_{g,eff} \geq \Delta t_{body}$. Most evidently, the energy penalty due to Δt_{po} (up to 13x) poses a marked challenge in achieving the desired energy benefit of TFET, since (0.1 – 1) ns FO3 delay maps to (10 – 100) MHz processor clock frequency (assume a scaling factor of 100 [8]) that is typical of targeted low-power applications like Internet-of-Things [27]; additionally, the (275 – 500) mV V_{DD}^* needs to be reduced in order to be comparable to typical CMOS-based low-power designs which operate at (200 – 400) mV [27].

IV. VARIABILITY-RESILIENT DEVICE OPTIONS

The striking sensitivity of circuit performance to Δt_{po} seen in section III demands explicit variability-awareness in device optimization. Nominal-oriented device design tends to prefer a small t_{po} (and accordingly a high N_{po}) that maximizes gate control of junction band bending to reduce SS in the off-state and minimizes tunneling distance to enhance BTBT current in the on-state [4], which leads to the design of the TFET in Fig. 1(a) (below referred to as “StG1”). However, as the ± 1 nm Δt_{po} ($\pm 40\%$ of t_{po}) can degrade device drive current and circuit speed by $(10^2 - 10^3)$ x, an optimal pocket design (t_{po}, N_{po}) needs to be re-centered at a trade-off point between BTBT enhancement (small t_{po}) and t_{po} -sensitivity reduction (large t_{po}). Here we

Fig. 8. (a) Comparison of I_{DS} - V_{GS} variability (aligned to $I_{OFF,0} = 50$ pA at $V_{DD} = 0.3$ V): staggered-gap TFETs “StG1” and “StG2” (left), broken-gap TFETs “BG1” and “BG2” (right); nominal curves in solid, $+3\sigma$ in dashed and -3σ in dotted. Nominal I_{60} are: 9.2 $\mu\text{A}/\mu\text{m}$ (StG1), 3.5 $\mu\text{A}/\mu\text{m}$ (StG2), 13.0 $\mu\text{A}/\mu\text{m}$ (BG1) and 12.7 $\mu\text{A}/\mu\text{m}$ (BG2), respectively. (b) Comparison of the respective “nominal Pareto” (solid) and “variability-inclusive Pareto” (dashed) of StG1/StG2/BG1/BG2, and the energy saving relative to StG1 after variation (at $\tau_T = 0.1$ ns and 1 ns).

propose an alternative design “StG2” at $t_{po} = 7$ nm, with a lowered $N_{po} = 2 \times 10^{19} \text{ cm}^{-3}$ for best gate control of the junction and minimal SS.

In addition, TFETs with broken-gap (BG) band alignment have a negative $E_{g,eff}$ and intrinsically shorter tunneling path at the junction [4], which enable a lower reliance on the dopant pocket for BTBT boosting and hence reduced t_{po} sensitivity. Here we adopt the GaSb/InAs BG TFETs ($E_{g,eff} = -0.14$ eV in bulk) similar to [25], including a nominal-oriented design “BG1” ($N_{po} = 10^{20} \text{ cm}^{-3}$, $t_{po} = 2$ nm) parallel to StG1 and pocket-redesigned “BG2” ($N_{po} = 5 \times 10^{19} \text{ cm}^{-3}$, $t_{po} = 3$ nm) parallel to StG2; both BG1 and BG2 have $t_{body} = 8$ nm.

The device-level I-V variability in Fig. 8(a) shows that compared to StG1, StG2 has a lower nominal I_{60} , but notably less worst-case I_{ON} loss which allows consistently higher drive current in the -3σ corner than StG1. At circuit level, despite its less favorable “nominal Pareto”, StG2 enables up to 2.8x better energy efficiency (Fig. 8(b)), (200 – 350)x less speed loss and a relaxed V_{DD} (< 400 mV) (Table III) than StG1. The re-centered pocket design hence substantially suppresses the PP sensitivity to t_{po} and mitigates variability, although the exact numeric values of StG2 might be over-optimistic as an increasing impact of ΔN_{po} is foreseen at reduced N_{po} , which can partially offset the gain from Δt_{po} -desensitization. Nevertheless, the parameter re-centering methodology in principle can also be generalized to other less sensitive parameter variations.

In Fig. 8 and Table III BG1 exhibits much less pronounced drive current reduction, speed loss and energy penalty than StG1 even at an identical N_{po} and a reduced t_{po} , which confirms

TABLE III
PP VARIABILITY BENCHMARKING: NOMINAL-ORIENTED DEVICE DESIGN VS VARIABILITY-RESILIENT DEVICE OPTIONS

Device option	Speed loss ^a		Energy penalty ^b		Operating point	
	$\tau_{var}^{(P)}/\tau_0^{(P)} @ \tau_T = 0.1$ ns	$\tau_{var}^{(P)}/\tau_0^{(P)} @ \tau_T = 1$ ns	$E_{var}/E_0^{(P)} @ \tau_T = 0.1$ ns	$E_{var}/E_0^{(P)} @ \tau_T = 1$ ns	V_{DD}^* (mV) @ $\tau_T = 0.1$ ns	V_{DD}^* (mV) @ $\tau_T = 1$ ns
StG1	410x	6300x	13.3x	5.2x	500	275
StG2	2.0x	18x	2.5x	2.9x	375	212
BG1	37x	150x	3.5x	3.8x	300	200
BG2	12x	41x	3.0x	2.5x	288	200

^{a,b}With respect to the nominal corners of StG1/StG2/BG1/BG2.

that BG TFET is inherently less sensitive to the perturbation of Δt_{po} as expected. Further pocket-engineered BG2 presents more favorable energy benefit and more relaxed V_{DD} than StG2 without having to suffer from the potentially higher impact of ΔN_{po} at overly reduced N_{po} . Overall, the preliminary results in BG TFETs confirm that enhanced intrinsic performance allows better harnessing of the performance-boosting dopant pocket without developing excessive sensitivity to it; from this perspective BG TFET indeed seems a promising variability-resilient device option, although issues such as phonon-thermalization-induced I_{OFF} have yet to be solved [4].

V. CONCLUSION

We studied the power-performance variability of a sub-5nm-node GaAsSb/InGaAs vertical TFET in logic circuits across comprehensive process variations based on state-of-the-art processing assumptions. Worst-case speed loss and iso-speed energy penalty are derived from QM/TCAD/Spectre-based 3 σ -corner analyses for sub-500mV V_{DD} , which suggests the demand for most critical process control budget in dopant pocket thickness (t_{po}) and to a lesser extent, in gate work function (Φ_g), effective band gap ($E_{g,eff}$) and body thickness (t_{body}). Variability-resilient device options are explored with respect to the most detrimental t_{po} variability, where re-centered pocket parameters offer down to 350x less speed loss and 5x less energy penalty than nominal-oriented device design; meanwhile, broken-gap TFETs show inherently low sensitivity to t_{po} variability and better post-variability power-performance when combined with proper pocket design. Still, for (0.1 – 1)ns gate delay, $\sim 18x$ worst-case speed degradation and $\sim 3x$ iso-speed energy penalty is possible due to t_{po} variation, but an unambiguous conclusion on the seriousness of such variability can be only made after a parallel exercise has been done on the iso-node MOSFET, which is yet to be studied.

ACKNOWLEDGMENT

The authors gratefully acknowledge Pieter Schuddinck and Pieter Weckx for thought-provoking discussions.

REFERENCES

- [1] R. G. Dreslinski, M. Wiecekowsky, D. Blaauw, D. Sylvester and T. Mudge, "Near-Threshold Computing: Reclaiming Moore's Law Through Energy Efficient Integrated Circuits", *Proc. IEEE*, vol. 98, no. 2, pp. 253-266, Feb. 2010, DOI: 10.1109/JPROC.2009.2034764.
- [2] A. M. Ionescu and H. Riel, "Tunnel field-effect transistors as energy-efficient electronic switches", *Nature*, vol. 479, no. 7373, pp. 329-337, Nov. 2011, DOI: 10.1038/nature10679.
- [3] A. S. Verhulst, D. Verreck, Q. Smets, K-H. Kao, M. Van de Put, R. Rooyackers, B. Sorée, A. Vandooren, K. De Meyer, G. Groeseneken, M. M. Heyns, A. Mocuta, N. Collaert and A. V-Y. Thean, "Perspective of tunnel-FET for future low-power technology nodes", in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, Dec. 2014, pp. 30.2.1-30.2.4, DOI: 10.1109/IEDM.2014.7047140.
- [4] D. Verreck, G. Groeseneken and A. Verhulst, "The Tunnel Field-Effect Transistor", in *Wiley Encyclopedia of Electrical and Electronics Engineering*, J. G. Webster, Ed. John Wiley & Sons, Inc., Nov. 2016, DOI: 10.1002/047134608X.W8333.
- [5] G. Dewey, B. Chu-Kung, J. Boardman, J. M. Fastenau, J. Kavalieros, R. Kotlyar, W. K. Liu, D. Lubyshev, M. Metz, N. Mukherjee, P. Oakey, R. Pillarisetty, M. Radosavljevic, H. W. Then, and R. Chau, "Fabrication, Characterization, and Physics of III-V Heterojunction Tunneling Field Effect Transistors (H-TFET) for Steep Sub-Threshold Swing", in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Washington, DC, USA, Dec. 2011, pp. 33.6.1-33.6.4, DOI: 10.1109/IEDM.2011.6131666.
- [6] E. Memisevic, J. Svensson, M. Hellenbrand, E. Lind and L-E. Wernersson, "Vertical InAs/GaAsSb/GaSb tunneling field-effect transistor on Si with $S = 48$ mV/decade and $I_{on} = 10 \mu A/\mu m$ for $I_{off} = 1$ nA/ μm at $V_{ds} = 0.3$ V", in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, Dec. 2016, pp. 19.1.1-19.1.4, DOI: 10.1109/IEDM.2016.7838450.
- [7] A. Alian, S. E. Kazzi, A. Verhulst, A. Milenin, N. Pinna, T. Ivanov, D. Lin, D. Mocuta and N. Collaert, "Record 47 mV/dec top-down vertical nanowire InGaAs/GaAsSb tunnel FETs", in *Symp. on VLSI Technol. Dig. Tech. Papers*, Honolulu, HI, USA, June 2018, pp. 133-134, DOI: 10.1109/VLSIT.2018.8510619.
- [8] S. Strangio, F. Settimo, P. Palestri, M. Lanuzza, F. Crupi, D. Esseni and L. Selmi, "Digital and analog TFET circuits: Design and benchmark", *Solid-State Electron.*, vol. 146, pp. 50-65, Aug. 2018, DOI: 10.1016/j.sse.2018.05.003.
- [9] K. Vaidyanathan, D. H. Morris, U. E. Avci, H. Liu, T. Karnik, H. Wang and I. A. Young, "Improving Energy Efficiency of Low-Voltage Logic by Technology-Driven Design", *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 4, no. 1, pp. 10-18, June 2018, DOI: 10.1109/JXDC.2018.2812242.
- [10] K. J. Kuhn, "CMOS transistor scaling past 32nm and implications on variation", in *IEEE/SEMI Adv. Semicond. Manuf. Conf. (ASMC)*, San Francisco, CA, USA, July 2010, pp. 241-246, DOI: 10.1109/ASMC.2010.5551461.
- [11] X. Wang, B. Cheng, A. R. Brown, C. Millar, J. B. Kuang, S. Nassif, and A. Asenov, "Interplay Between Process-Induced and Statistical Variability in 14-nm CMOS Technology Double-Gate SOI FinFETs", *IEEE Trans. Electron Devices*, vol. 60, no. 8, pp. 2485-2492, Aug. 2013, DOI: 10.1109/TED.2013.2267745.
- [12] U. E. Avci, R. Rios, K. Kuhn and I. A. Young, "Comparison of Performance, Switching Energy and Process Variations for the TFET and MOSFET in Logic", in *Symp. on VLSI Technol. Dig. Tech. Papers*, Honolulu, HI, USA, June 2011, pp. 124-125. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5984671&isnumber=5984315>.
- [13] V. Saripalli, S. Datta, V. Narayanan and J. P. Kulkarni, "Variation-Tolerant Ultra Low-Power Heterojunction Tunnel FET SRAM Design", in *Proc. IEEE/ACM Int. Symp. on Nanoscale Archit. (NANOARCH)*, San Diego, CA, USA, June 2011, pp. 45-52, DOI: 10.1109/NANOARCH.2011.5941482.
- [14] U. E. Avci, D. H. Morris, S. Hasan, R. Kotlyar, R. Kim, R. Rios, D. E. Nikonov and I. A. Young, "Energy efficiency comparison of nanowire heterojunction TFET and Si MOSFET at $L_g=13$ nm, including P-TFET and variation considerations", in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Washington, DC, USA, Dec. 2013, pp. 33.4.1 - 33.4.4, DOI: 10.1109/IEDM.2013.6724744.
- [15] C. Chen, Y. Chen, M. Fan, V. P. Hu and C. Chuang, "Impacts of work function variation and line-edge roughness on TFET and FinFET devices and logic circuits", in *Proc. SOLI-3D-Subthreshold Microelectron. Technol. Unified Conf. (S3S)*, Millbrae, CA, USA, Oct. 2014, DOI: 10.1109/S3S.2014.7028225.
- [16] D. Verreck, A. S. Verhulst, M. Van de Put, B. Sorée, W. Magnus, A. Mocuta, N. Collaert, A. Thean, and G. Groeseneken, "Full-zone spectral envelope function formalism for the optimization of line and point tunnel field-effect transistors", *J. Appl. Phys.*, vol. 118, no. 13, Oct. 2015, Art. no. 134502, DOI: 10.1063/1.4931890.
- [17] D. Yakimets, "Vertical Transistors: A slippery path towards the ultimate CMOS scaling," Ph.D. dissertation, Dept. Elect. Eng., KU Leuven, Leuven, Belgium, 2016.
- [18] D. Yakimets, M. Garcia Bardon, D. Jang, P. Schuddinck, Y. Sherazi, P. Weckx, K. Miyaguchi, B. Parvais, P. Raghavan, A. Spessot, D. Verkest, and A. Mocuta, "Power aware FinFET and lateral nanosheet FET targeting for 3nm CMOS technology", in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, Dec. 2017, pp. 20.4.1-20.4.4, DOI: 10.1109/IEDM.2017.8268429.
- [19] *Sentaurus™ Device User Guide, 2016.12*, Synopsys, Mountain View, CA, USA, Dec. 2016.
- [20] H. Lu, W. Li, Y. Lu, P. Fay, T. Ytterdal, and A. Seabaugh, "Universal charge-conserving TFET SPICE model incorporating gate current and noise", *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 2, pp. 20-27, Dec. 2016, DOI: 10.1109/JXDC.2016.2582204.
- [21] D. Yakimets, G. Eneman, P. Schuddinck, T. H. Bao, M. Garcia Bardon, P. Raghavan, A. Veloso, N. Collaert, A. Mercha, D. Verkest, A. V.-Y. Thean, and K. De Meyer, "Vertical GAAFETs for the Ultimate CMOS

- Scaling", *IEEE Trans. Electron Devices*, vol. 62, no. 5, pp. 1433-1439, May 2015, DOI: 10.1109/TED.2015.2414924.
- [22] E. Memisevic, M. Hellenbrand, E. Lind, A. R. Persson, S. Sant, A. Schenk, J. Svensson, R. Wallenberg, and L.-E. Wernersson, "Individual Defects in InAs/InGaAsSb/GaSb Nanowire Tunnel Field-Effect Transistors Operating below 60 mV/decade", *Nano Letters*, vol. 17, no. 7, pp. 4373-4380, June 2017, DOI: 10.1021/acs.nanolett.7b01455.
- [23] J. Bizindavyi, A. S. Verhulst, Q. Smets, D. Verreck, B. Soree and G. Groeseneken, "Band-Tails Tunneling Resolving the Theory-Experiment Discrepancy in Esaki Diodes", *IEEE Journal of the Electron Devices Society*, vol. 6, pp. 633-641, May 2018, DOI: 10.1109/jeds.2018.2834825.
- [24] *Virtuoso® Spectre® Circuit Simulator Reference, V.15.1*, Cadence, San Jose, CA, USA, Mar. 2016.
- [25] D. Verreck, A. S. Verhulst, Y. Xiang, D. Yakimets, S. E. Kazzi, B. Parvais, G. Groeseneken, N. Collaert, and A. Mocuta, "Built-In Sheet Charge As an Alternative to Dopant Pockets in Tunnel Field-Effect Transistors", *IEEE J. Electron Devices Soc.*, vol. 6, pp. 658-663, May 2018, DOI: 10.1109/JEDS.2018.2835501.
- [26] N. Verma, J. Kwong and A. P. Chandrakasan, "Nanometer MOSFET Variation in Minimum Energy Subthreshold Circuits", *IEEE Trans. Electron Devices*, vol. 55, no. 1, pp. 163-174, Jan. 2008, DOI: 10.1109/TED.2007.911352.
- [27] W. Dehaene and A. S. Verhulst, "New devices for internet of things: a circuit level perspective", in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Washington, DC, USA, Dec. 2015, pp. 25.5.1-25.5.4, DOI: 10.1109/IEDM.2015.7409766.