Energy-Efficient FinFET-Versus TFET-Based STT-MRAM Bitcells

Ariana Musello, Santiago Perez, Marco Villegas, Luis Miguel Procel, Ramiro Taco, Lionel Trojman

Energy-Efficient FinFET- Versus TFET-Based STT-MRAM Bitcells

Ariana Musello *, Santiago S. Pérez *, Marco Villegas *, Luis Miguel Prócel *, Ramiro Taco *, Lionel Trojman †*
* Instituto de Micro y Nanoelectrónica (IMNE), Universidad San Francisco de Quito (USFQ), Quito 170901, Ecuador
† Institut Supérieur d'Électronique de Paris (ISEP), 10 rue de Vanves, Issy les Moulineaux, 92130, France.

***Abstract*—This paper explores STT-MRAM bitcells based on double-barrier magnetic tunnel junctions (DMTJs) at the circuit-level, benchmarking TFET- against FinFET-based bitcells. Different bitcell configurations are tested to find optimal minimum energy design points using both technologies at a range of ultralow supply voltages. TFETs were found to be the optimal access device for supply voltages under or equal to 0.4V because of their significantly more robust behavior and lower write energy consumption, albeit higher write delays and bigger area for higher voltages.**

***Index Terms*—STT-MRAM, double-barrier magnetic tunnel junction (DMTJ), tunnel FET (TFET), FinFET, ultralow voltage**

I. Introduction

Spin-transfer torque magnetic random access memories (STT-MRAMs) have the potential of leading non-volatile on-chip memory technology due to their potential in offering high write/read speed, small area footprint, and low power consumption [1], [2]. In addition to applications, such as on-chip memories, and logic-in-memory computing, STT-MRAMs can also be exploited in cryogenic electronics [3]. In particular, for cache memory applications, STT-MRAM technology has been shown to be an attractive alternative to overcome the constant increase in leakage power of classical CMOS-based memory technology. The basic bit element of conventional STT-MRAM is the single-barrier magnetic tunnel junction (SMTJ), which typically presents high writing currents. The diminution of write switching currents has become a critical factor to improve the performance of STT-MRAMs, since it is required for lower energy and higher speed memories [2]. In this regard, the use of double-barrier MTJ (DMTJ) is an effective strategy that offers reduced switching currents down to few μA [4].

In the above context, this work investigates STT-MRAM bitcells based on DMTJs at the circuit-level, exploring two technologies for the access devices: a calibrated 10nm-FinFET technology model [5] and a Tunnel FET technology model. The latter is an emergent technology that offers very low operating bias supply [6]. While the technologies reported in other studies [4], [6] are based on numerical and predictive models, this work considers models based on experimental studies [5], [7] for FinFET devices and that include parasitic effects for TFET devices. A critical design issue in STT-MRAM bitcells is their high write energy [1], which is why this work specifically focuses on the write operation and on determining which technology would work as the most energy-efficient access device. Additionally, the comparison is carried out at operation voltages in the ultra-low voltage domain attending to this work's energy-efficiency perspective.

The paper is organized as follows. Section 2 presents a device-level description of the DMTJs, FinFETs and TFETs. Section 3 gives a circuit-level description of the bitcell configurations and presents the write operation results in terms of current, delay, and energy. Section 4 concludes this paper.

II. Device Modeling

A. Double-Barrier Magnetic Tunnel Junction (DMTJ)

A SMTJ is made up of two ferromagnetic (FM) layers sandwiching a thin oxide layer [8]. While one FM layer - the reference layer (RL) - has a fixed magnetization, the free layer (FL) magnetization can be either parallel or antiparallel depending on the magnetization direction. Therefore, the SMTJ has two stable states, '0' and '1', corresponding to low and high resistance states. The STT mechanism enables the switching between these two states when the applied switching current (I_{write}) is greater than the critical switching current (I_{c0}) [9], which is typically high for conventional SMTJ-based STT-MRAM. To deal with this, the use of DMTJ allows reduced switching currents thanks to the presence of an additional RL that enhances the total torque acting on the FL [10].

In this work, at the device-level, DMTJ device has been described by macrospin-based Verilog-A compact models [11] used in the Synopsys Custom-Compiler tool.

B. FinFET and Tunnel-FET (TFET) devices

TFETs are nanowire shaped P-i-N junctions surrounded by gate oxide. As a result, the current is based on the band-to-band-tunneling (BTBT). This effect enables TFETs to reach a sub-threshold swing (SS) lower than 60 mV/dec and very high output resistance [6]. Hence, TFETs can deliver higher on-currents at lower supply voltages (hundreds of mV), along with very low power dissipation.Thus, their highest performance is within the ultra-low voltage domain ($0.1V \leq V_{DD} \leq 0.4V$) in many analog and digital applications [12]. As an emergent technology, the choice of the material for the fabrication process is still under study, and there are the issue of ambipolarity and unidirectionality of the current that affect their performance [6]. The TFETs used in this paper

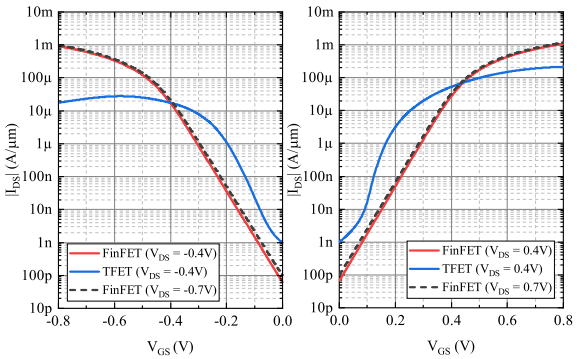

Fig. 1. Transfer characteristics of FinFET and TFET devices.

are the complementary III-V heterojunction TFET nanowires (NWs) of AlGaSb/InAs proposed by [6]. Note also that we simulate the TFETs with the extrinsic parasitic component using the compact model (CM) described in [13]. The considered physics model is run with TCAD-Sentaurus (Synopsys) which provides I-V and C-V curves stored in look-up tables (LUTs), while the CM considers the geometrical structure of the TFET to extract the parasitic values added to the LUTs. The final model reports a higher off-current associated with leakage current.

FinFETs are a recent transistor technology in the industry replacing bulk planar CMOS technology for nodes lower than 28nm. The principal difference refers to the geometry, as the FinFETs gate is elevated over the wafer forming a "fin"-shaped channel with the oxide wrapped around the fin (as a trigate) forming a double port structure. Such structure offers a better control of the channel by the gate and a larger current area.In fact, the drive current is higher than in planar CMOS, so the commutation time is also higher. [14]. In this work, the 10-nm FinFET technology was simulated using a Predictive Technology Model (PTM) proposed in [15] and the $I_D - V_{GS}$ characteristics were calibrated using the data and physical parameters reported in [5], [7]. Indeed, for the calibration we chose the model for low-power (LP) applications: at the operation voltage of $V_{DD} = 0.7V$ and 1 finger, both the n-FinFET and p-FinFET off-currents were fixed at $10pA$ by shifting the gate work function consistently with [5].

The $I_{DS} - V_{GS}$ curves for each device at $V_{DS} = 0.4V$, normalized by the device perimeter $W_{perimeter}$, are presented in Fig. 1. For FinFETs, $W_{perimeter} = 2H_{fin} + W_{fin}$; for TFETs, $W_{perimeter} = 4L_S$. The curve for FinFETs at $V_{DS} = 0.7V$ is also shown. On one hand, at $V_{DS} = 0.4V$, *n*-FinFETs and *p*-FinFETs feature the same off-current (I_{off}) of 67.14 $pA/\mu m$, and on-currents (I_{on}) of 27.34 $\mu A/\mu m$ and 17.01 $\mu A/\mu m$, respectively. On the other hand, *n*-TFETs and *p*-TFETs feature an I_{off} of 1.01 $nA/\mu m$ and 0.98 $nA/\mu m$, and an I_{on} of 52.75 $uA/\mu m$ and 17.46 $uA/\mu m$, respectively. It can be observed that FinFETs and TFETs (both p-type and n-type) have similar normalized on-currents at $V_{GS} = 0.4V$. This leads us to choose $V_{DD} = 0.4V$ as the operation voltage for an initial analysis, as it implies a fair comparison.

The main device characteristics of both transistor technologies and DMTJs are presented in Table I. The footprint area per

TABLE I
DEVICE PARAMETERS AND CHARACTERISTICS

Parameter	Description	Value
DMTJ		
d	MTJ diameter	22 nm
t_{FL}	Free layer thickness	1 nm
$t_{OX,T}$	Top barrier thickness	0.85 nm
$t_{OX,B}$	Bottom barrier thickness	0.4 nm
RA	Resistance-area product	$\sim 7\ \Omega\mu m^2$
R_H	High resistance state at 0V	44.6 kΩ
R_L	Low resistance state	20 kΩ
I_{c0}	Critical switching current	$\sim 3\ \mu A$
FinFET		
L_G	Gate Length	18 nm
H_{fin}	Fin height	46 nm
t_{fin}	Fin width	7 nm
EOT	Equivalent Oxide Thickness	0.7 nm
V_{th}	Threshold voltage (n-type / p-type)	0.48 V / 0.50V
TFET		
L_G	Gate length	20 nm
L_S	Length of nanowiere square cross-section	7 nm
EOT	Equivalent Oxide Thickness	1 nm

device for TFETs (one NW) is $\sim 135nm^2$, whilst for the FinFETs (one finger) it is $\sim 165nm^2$. Even though the difference in area is relatively small, it must be noted that for TFETs, $W_{perimeter} = 28nm$, and for FinFETs $W_{perimeter} = 99nm$. This means that at $V_{DD} \approx 0.4V$, in order to have a similar on-current with both TFETs and FinFETs, for every 1 FinFET finger, 3 to 4 TFET NWs would be necessary. This implies that, in general, TFET-based bitcells will require more parallel-connected devices than FinFET-based bitcells.

III. BITCELL-LEVEL ANALYSIS AND RESULTS

Fig. 2 (a)-(h) provides the set of bitcell configurations considered in this work. These configurations can use either a standard connection (SC) - access transistors (AT) connected to the RL_T, as in (b), (d), (f) and (h) in Fig. 2 - or a reverse connection (RC) - AT connected to the RL_B as in (a), (c), (e) and (g) in Fig. 2 [10]. Moreover, each configuration can be based on just n-type transistors - as in (a), (b), (e) and (f) -, or use complementary n-type and p-type devices - as in (c), (d), (g) and (h). Note that due to TFET unidirectionality, two transistors are always needed as AT for TFET-based bitcells.

First, we conduct a bitcell-level analysis at nominal conditions (*i.e.* without considering process variations), evaluating the 8 aforementioned configurations at $VDD = 0.4V$. We

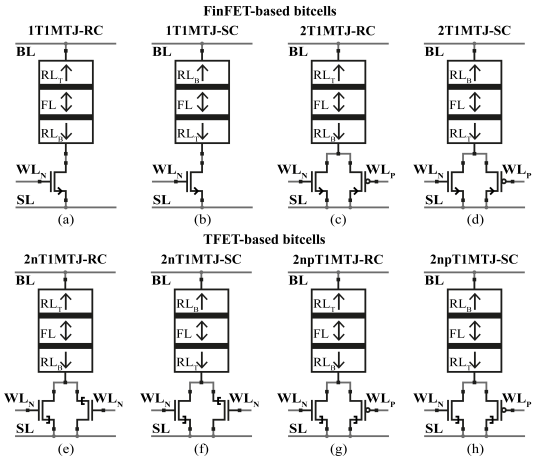

Fig. 2. FinFET-based (a-d) and TFET-based (e-h) bitcell configurations.

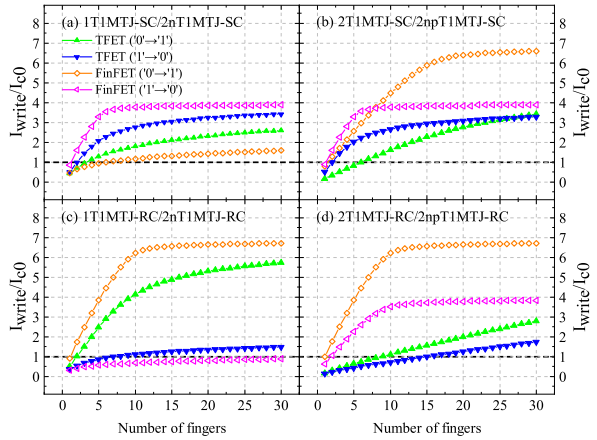

Fig. 3. I_{write}/I_{c0} ratios as a function of the number of parallel-connected access devices for TFET and FinFET-based bitcells: (a) 1T1MTJ-SC / 2nT1MTJ-SC, (b) 2T1MTJ-SC / 2npT1MTJ-SC, (c) 1T1MTJ-RC / 2nT1MTJ-RC, and (d) 2T1MTJ-RC / 2npT1MTJ-RC.

look for an optimal design point for each technology in order to compare them when their performance is best. Figs. 3(a)-(d) show the write current I_{write} for every bitcell configuration as a function of the number of fingers and NWs of the FinFETs and TFETs used. The results are expressed in terms of the I_{write}/I_{c0} ratio, and correspond to both the '0'→'1' and '1'→'0' switching transitions. Note that source degeneration and the different resistances associated with the MTJ's '0' and '1' states cause asymmetries in the write current curves for the two switching transitions [10]. Ideally, $I_{write}/I_{c0} \geq 3$ to have a robust write operation [4]. Under that criteria, 1T1MTJ-RC/2nT1MTJ-RC would be discarded as optimal bitcell configurations, as well as 2npT1MTJ-RC for TFET-based ones and 1T1MTJ-SC for FinFET-based ones. Interestingly, the write current curves for the two transitions of FinFET (TFET)-based bitcells in the 2T1MTJ-SC (2npT1MTJ-SC) configuration intersect at 8 fingers (26 NWs); this parity of write current is desirable.

Fig. 4 shows write delay and energy results as a function of the number of fingers/NWs for (a)(c) TFET-based and (b)(d) FinFET-based bitcells. The delay refers to the write pulse width at a write-error-rate (WER) of 10^{-7}, taking into account the worst-case transition. The energy refers to the average write energy per access for the two transitions. Clearly, the optimal bitcell configuration that allows the lowest write delay and energy is 2npT1MTJ-SC when using TFETs and 2T1MTJ-SC when using FinFETs.

The minimum-energy point (MEP) in these configurations, marked in Fig. 4, is chosen as the optimal design point [10]. When using TFETs, it corresponds to 26 NWs, at which the write delay is $1.77ns$, the write energy is $6.63fJ$ and the write currents for the two switching transitions are $\sim 3.2\mu A$. When using FinFETs, it corresponds to 8 fingers, at which the write delay is $1.45ns$, the write energy is $6.31fJ$ and the write currents are $\sim 3.7\mu A$. According to the nominal results, TFET-based bitcells present a 5.17% higher write energy and a 21.72% larger delay at their optimal design point than their FinFET-based counterparts.

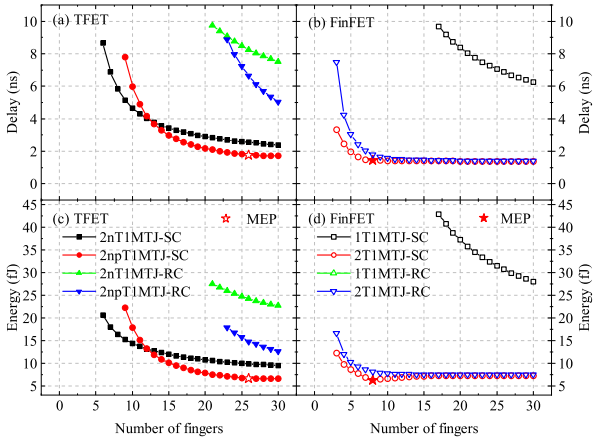

Fig. 4. Delay and energy results as a function of the number of parallel-connected access devices for (a)(c) TFET-based bitcells and (b)(d) FinFET-based bitcells.

However, process variability must be contemplated, both on the DMTJs and the transistors. For DMTJs, Gaussian-distributed variations are considered for area and the DMTJ FL and top/bottom oxide layers [4]. For the transistors, a Gaussian-distributed variation is considered for the threshold voltage. As reported in [4], the maximum deviation $\sigma_{v_{th}}$ that ensures a 100% yield is $\sigma_{v_{th}} = 10mV$.

The Monte Carlo (MC) results for write energy and delay are presented in Fig. 5. We can observe that while the design point with TFETs presents a very clear Gaussian distribution, that is not the case for the design point with FinFETs. For the latter, it can be observed that the energy and delay standard deviations are much higher, and the energy and delay means are rather different than the nominal values found before. Considering the MC mean values - and thus process variability - it can be observed that TFET-based bitcells present 8.13% more write delay but actually consume 6.17% less energy than their FinFET-based counterparts.

To further understand the behavior of TFET and FinFET-based bitcells at their optimal design points based on the MEP, the same previous analysis was made at $V_{DD} = 0.35V$ and $V_{DD} = 0.45V$, using the same optimal bitcell topologies as in $V_{DD} = 0.4V$ (2T1MTJ-SC for FFs and 2np1MTJ-SC for TFs). Lower supply voltages were discarded because FinFETs could not provide enough write current, while TFETs could.

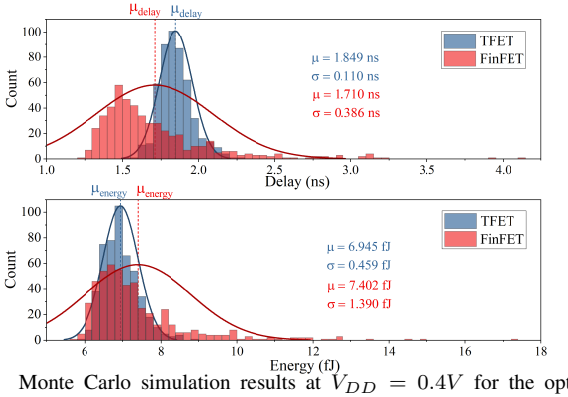

Fig. 5. Monte Carlo simulation results at $V_{DD} = 0.4V$ for the optimal design point of (a) TFET- and (b) FinFET-based bitcells.

TABLE II
OPTIMAL DESIGN POINTS BASED ON NOMINAL RESULTS

	VDD (V)	#Fs/NWs	Area (nm^2)	Energy (fJ)	Delay (ns)
TFET	0.35	28	3780	6.236	2.275
	0.40	26	3510	6.633	1.771
	0.45	28	3780	7.013	1.400
FinFET	0.35	29	4785	5.846	1.853
	0.40	8	1320	6.307	1.455
	0.45	3	495	6.989	1.202

TABLE III
MONTE CARLO RESULTS FOR WRITE ENERGY AND DELAY

		Write energy			Write delay		
	V_{DD}	μ (fJ)	σ (fJ)	CV	μ (ns)	σ (ns)	CV
TFET	0.35 V	6.657	0.523	0.079	2.414	0.173	0.072
	0.40 V	6.945	0.459	0.066	1.849	0.110	0.059
	0.45 V	7.320	0.444	0.061	1.458	0.082	0.056
FinFET	0.35 V	7.230	1.777	0.246	2.301	0.661	0.287
	0.40 V	7.402	1.390	0.188	1.710	0.386	0.226
	0.45 V	7.510	0.819	0.109	1.296	0.175	0.135

Table II presents the MEP points - and therefore the optimal design points - for each voltage and access device type, including the number of fingers/NWs (#Fs/NWs), area, the associated worst-case write delay, and average write energy.

Just as with only $V_{DD} = 0.4V$, the nominal results could lead to the conclusion that FinFET-based bitcells have a write energy-consumption and write delay advantage over TFET-based bitcells at every voltage.

However, once process variability is taken into account the conclusions are different. The MC statistical results for the three supply voltages are presented in Table III, along with the coefficient of variation (CV). For $V_{DD} = 0.35V$, $0.4V$ and $0.45V$, TFET-based bitcells present mean write energy values that are 7.93%, 6.17%, and 2.53% smaller than for the FinFET-based cells, respectively. Regarding mean write delay at those voltages, FinFET-based bitcells display 4.68%, 7.52% and 11.11% shorter delays than when TFETs are used. As V_{DD} increases, the energy-efficiency of using TFETs decreases, while the velocity-efficiency of using FinFETs increases.

Considering these energy-delay trade-offs, the real benefit of using TFETs from an energy-efficiency perspective happens for $V_{DD} \leq 0.4V$. Moreover, at $V_{DD} = 0.35V$ TFET usage presents a particular advantage: they occupy 21% less area than FinFETs. At $V_{DD} = 0.45V$, the energy reduction that TFETs could provide does not compensate for the increase in write delay and the 3780 nm^2 area they occupy - in contrast to the 495 nm^2 area of FinFETs at their optimal design point.

Nevertheless, based on the FinFET and TFET technology models we are using, the much larger coefficients of variation ($CV = \sigma/\mu$) of FinFET-based bitcells show that they are significantly less robust than their TFET-based counterparts. This is because the FinFETs are being operated at supply voltages lower than their V_{th}: in the subthreshold region, they are more susceptible to process variations. As a matter of fact, if we considered the 'worst-case' and compared the energy and delay as $\mu + 3\sigma$ results, then the especially high standard deviations of FinFET-based bitcells would cause TFET-based bitcells to be both more energy efficient and faster for every voltage, with considerable advantages in each case except for area occupation at $V_{DD} = 0.45V$ and $0.40V$.

IV. CONCLUSION

In this work, the write operation in STT-MRAM bitcells was compared using TFETs and FinFETs as access devices at supply voltages of 0.35V, 0.4V and 0.45V. Considering process variability, TFET-based bitcells at optimal design points present lower mean write energy-consumption, albeit higher write delays. The advantages related to energy-efficiency are best appreciated for $V_{DD} \leq 0.4$. Regarding area, TFET-based bitcells present a clear advantage at $V_{DD} = 0.35V$. FinFET-based bitcells are significantly more sensitive to process variations than TFET-based bitcells, as they are being operated at voltages below V_{th}. Hence, the benefit of higher write velocity when using FinFETs could be overshadowed by the lack of robustness - particularly for $V_{DD} \leq 0.4$ - and the energy-efficiency benefits of using TFETs would become prevalent. Therefore, for $V_{DD} \leq 0.4$ and considering the design challenge in STT-MRAMs of a relatively high write energy, as compared to FinFETs-based alternative, the TFET-based solution presents lower write energy consumption and a robust behavior, at the cost of higher write delays and bigger cell area at $V_{DD} \geq 0.4$.

REFERENCES

[1] X. Fong *et al.*, "Spin-Transfer Torque Memories: Devices, Circuits, and Systems," *Proc. IEEE*, vol. 104, p. 1449–1488, Jul. 2016.

[2] E. Garzón *et al.*, "Assessment of STT-MRAM performance at nanoscaled technology nodes using a device-to-memory simulation framework," *Microelectronic Engineering*, vol. 215, p. 111009, 2019.

[3] E. Garzón *et al.*, "Exploiting STT-MRAMs for cryogenic non-volatile cache applications," *IEEE Trans. Nanotechnol.*, vol. 20, pp. 123–128, 2021.

[4] E. Garzón *et al.*, "Ultralow Voltage FinFET-Versus TFET-Based STT-MRAM Cells for IoT Applications," *Electronics*, vol. 10, no. 15, p. 1756, 2021.

[5] J. S. Yoon *et al.*, "Bottom Oxide Bulk FinFETs Without Punch-Through-Stopper for Extending Toward 5-nm Node," *IEEE Access*, vol. 7, pp. 75 762 – 75 767, Jun. 2019.

[6] S. Strangio *et al.*, "Digital and analog TFET circuits: Design and benchmark," *Solid-State Electronics*, vol. 146, 05 2018.

[7] C. Auth *et al.*, "A 10nm high performance and low-power cmos technology featuring 3rd generation finfet transistors, self-aligned quad patterning, contact over active gate and cobalt local interconnects," *2017 IEEE IEDM*, pp. 29.1.1–29.1.4, 2017.

[8] Z. Diao *et al.*, "Spin transfer switching in dual MgO magnetic tunnel junctions," *Applied Physics Letters*, vol. 90, no. 13, p. 132508, 2007.

[9] G. Hu *et al.*, "STT-MRAM with double magnetic tunnel junctions," in *2015 IEEE IEDM*, 2015, pp. 26.3.1–26.3.4.

[10] E. Garzón *et al.*, "Assessment of STT-MRAMs based on double-barrier MTJs for cache applications by means of a device-to-system level simulation framework," *Integration*, vol. 71, pp. 56–69, Mar. 2020.

[11] R. De Rose *et al.*, "Compact modeling of perpendicular STT-MTJs with double reference layers," *IEEE Trans. Nanotechnol.*, vol. 18, pp. 1063–1070, 2019.

[12] B. Sedighi *et al.*, "Analog Circuit Design Using Tunnel-FETs," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, pp. 39–48, 2015.

[13] F. Horst, "Compact DC Modeling of Tunnel-FETs," Ph.D. dissertation, 11 2019.

[14] C.-Y. Su *et al.*, "Transistor reliability characterization and modeling of the 22FFL FinFET technology," in *2018 IEEE IRPS*, 2018, pp. 6F.8–1–6F.8–7.

[15] ASU, "Predictive Technology Model (PTM). Available on: http://ptm.asu.edu." [Online]. Available: http://ptm.asu.edu