

Received July 29, 2020, accepted August 12, 2020, date of publication August 19, 2020, date of current version September 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3017756

Chip Power-Frequency Scaling in 10/7nm Node

PHIL OLDIGES, (Senior Member, IEEE), REINALDO A. VEGA, (Member, IEEE), HENRY K. UTOMO, NICK A. LANZILLO¹, THOMAS WASSICK, (Member, IEEE), JUNTAO LI, JUNLI WANG, (Member, IEEE), AND GHAVAM G. SHAHIDI¹, (Fellow, IEEE)

IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA

Corresponding author: Ghavam G. Shahidi (shahidi@us.ibm.com)

ABSTRACT The 10/7nm node has been introduced by all major semiconductor manufacturers (Intel, TSMC, and Samsung Electronics). This article looks at the power-performance benefit of the 10/7nm node as compared to the previous node (14nm). Specifically, we track the power-performance in high performance space, using Intel's Core-i7 (Intel's highest performance consumer microprocessor that uses the highest performance CMOS technology node) manufactured in Intel's 10nm. The paper first looks at the scaling of the device power-performance from the Intel 14++nm node to Intel 10nm, using 3D TCAD simulation with dimensions obtained from actual product cross-sections, and also scaling of the interconnect capacitance node-to-node. Next, the paper does a comparison of industry 10/7nm node technologies (from Intel, TSMC, and Samsung Electronics). The paper argues that for Intel, in the 10nm nodes, the total chip power at constant frequency (energy-per-operation) has scaled by a much lower amount vs. the 14++ node, as compared to the 14++ vs. the previous (22 nm) node. The lack of power scaling can be traced to a reduction in current per device perimeter (caused by the increased device parasitic resistance and the reduced device and fin pitch) and to an increase in capacitance per fin (caused by an increase in the FinFET height). Proper scaling of the device is critical for chip power scaling (energy-per-operation) at upcoming nodes, especially as it applies to high performance microprocessors and for the data analyzed here this is not the case.

INDEX TERMS Computer performance, CMOS scaling, FinFET, Moore's Law, MOSFET, power dissipation, scaling, technology node.

I. INTRODUCTION

Key benefits of CMOS scaling have been density improvement (*i.e.* more transistors per area), chip frequency improvement (*i.e.* for single thread tasks), and power reduction at a given frequency [1], [2]. In recent nodes, frequency scaling has slowed down. Focus has shifted to adding cores and functionality with migration to the new node. The extra functionality was enabled by the density increase and drop in power at constant frequency; however, in the most recent nodes, there has been a slowdown in chip power scaling at a given frequency (*i.e.* energy-per-operation). This has been compensated by design improvements (architecture, place and route, *etc.*) to maintain power scaling [3], [4].

In order to evaluate the benefit of CMOS scaling in the high-performance space, we had previously followed the evolution of the node-to-node power-performance benefit for Intel's highest performance consumer microprocessor, Intel Core-i7, across many technology nodes [5] (highest

performance prior to the introduction of Core-i9 in 14++ nm, which is limited to desktop applications). That work showed that there has been marked reduction in power-performance gain in recent technology nodes (*i.e.* 22nm through 14nm) as compared to the earlier technology nodes (250nm through 32nm nodes). By the time that work was published, there was only one announced Intel 10nm part. That 10nm part did not show any power-performance gain. Since then, Intel has announced several 10nm parts. The goal of this study is to compare the power-performance in the newly announced parts, and to do a systematic TCAD study in order to understand the power-performance behavior as we transition from a well-designed 14 nm technology to a 10/7 nm node. Furthermore, we performed a systematic comparison of Intel's 10nm vs. the other industry 7nm nodes with similar ground rules (TSMC and Samsung 7nm nodes), to determine if the observed behavior of any 10/7 node is unexpected.

II. METHODOLOGY AND APPROACH

This work uses total chip power at a given frequency to track the evolution of the total power (or total

The associate editor coordinating the review of this manuscript and approving it for publication was Muhamamd Aleem¹.

energy-per-operation) through node transitions. Total chip power, P_{Total} is roughly about $fCV^2 + P_{Leakage}$, where f is the frequency and C is the effective switching capacitance, and $P_{Leakage}$ is the stand-by leakage power (power at stopped clock). $P_{Leakage}$ is a function of total device widths on the chip, device off currents (*i.e.* threshold voltage), and operating voltage. In this study, focus is on high performance CMOS, and specifically higher performance microprocessors. In recent high-performance microprocessors, active power dominates, and the leakage power (*i.e.* deep sleep power) is about 5-20% of the total power [8], [9]. For the total chip power, the thermal design power (TDP), which is the highest steady amount of power that the chip can generate while running applications, is used. To keep the number of cores and the amount cache the same throughout this study, the TDP-frequency of 4 core/8 MB cache product family or 2 cores/4 MB or 6 cores/12 MB, all scaled to 4 core/8 MB, are the focus here. The total number of processor core transistors have been fairly constant (about 800 million) and while the graphic engine device count has increased dramatically over many generations, the power of the graphic engine is only a few percentages of the total power especially at high frequencies (circuit details are in [3]).

In this article we use device technology computer aided design (TCAD) to simulate the expected power drop at a given frequency based on device structure. Device structures are obtained from the cross section of high-performance microprocessors products in a given node [10]. We try to match the currents reported for 14++ and 10 nm nodes. We extract the device capacitance and resistance as it scales node-to-node and use that (along with the metallization capacitance and resistance) to predict the power performance at the circuit level.

III. PROBLEM STATEMENT

Intel introduced its first 10nm processor (Core-i3) in early 2019. That part (core-i3 8130U, pink data point) had a noticeable higher power at a given frequency (Figure 1), and a lower turbo frequency [6], as compared to the equivalent 14nm part. The degraded power-performance were reflected in the software benchmarks [7]. Late in 2019, Intel released several parts in 10nm, covering the entire Core-i3 through Core-i7 family [6]. Focusing on the 10nm core-i3 parts (marked red in Figure 1), they still show degraded power performance as compared to 14nm (even the 10th generation 14, marked in purple), and a reduced turbo frequency.

In order to do a more systematic study of the 10nm parts, and evaluate the benefit of the CMOS scaling in the high-performance space, in this work we focus on the evolution of the node-to-node benefit for the Intel Core-i7 processor for multiple generations (similar to our previous work [5]). Intel introduced the 10th generation of the Core-i7 in both the 14++ node as well as in 10nm node. The 10th generation parts in 14++ and 10nm are equivalent, except for some minor difference in the graphic core. The graphic core power is a small portion of the total power. Figure 2(a) is a

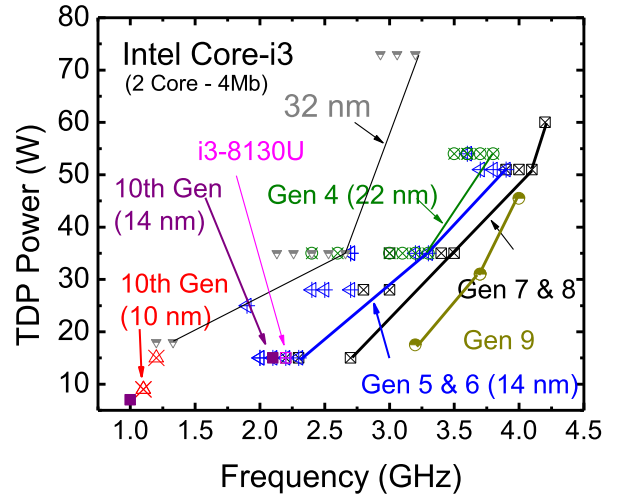


FIGURE 1. Power (TDP vs. frequency) for Intel Core-i3 covering many product and technology generations, for the first 10nm node (pink) release, and the second 10nm release (red), and the earlier generations.

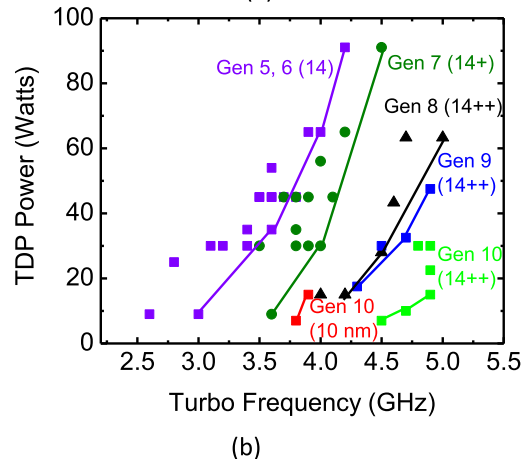
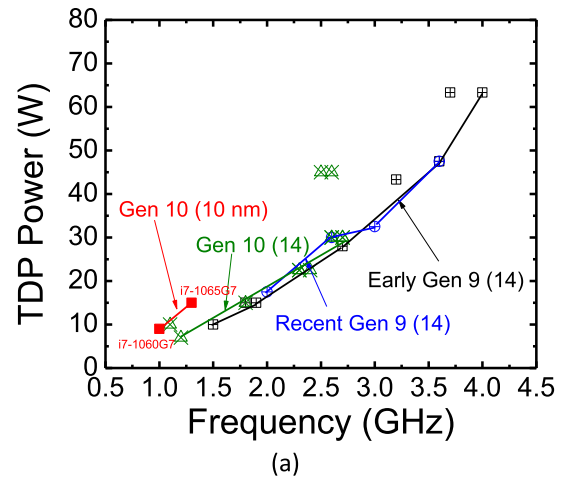


FIGURE 2. (a) Power (TDP) vs. frequency for Intel Core-i7, generation 9 (14++nm) and 10 (14++ and 10nm) (b) Turbo frequency vs. power for Intel core-i7 for generations 5-9 (14-14++nm) and 10 (14++nm and 10 nm).

plot of TDP power vs. frequency for Intel Core-i7 for the 9th (in 14++ node) and 10th (in 14++ and 10nm nodes)

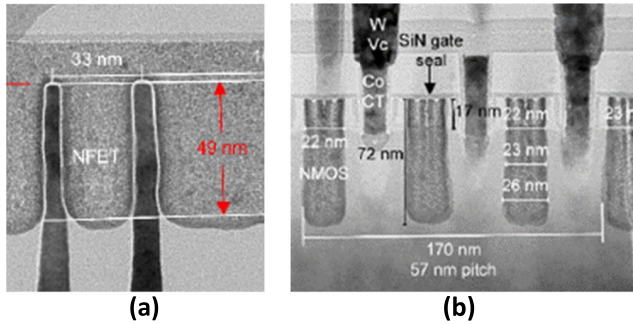


FIGURE 3. Cross section through FinFET under the gate and through PC from Intel 10nm microprocessor [10].

TABLE 1. Device feature comparisons between Intel 14++nm node and 10nm node technologies [10].

	Product Used in Construction Analysis	Fin Pitch	Fin Height	Fin Width (Middle of Fin)	Fin Width (Bottom of the Fin)	CPP	PC Width (Top of the Fin)	PC Width (Bottom of the Fin)
Intel 14++nm	Core i7-10510U	42	44	6	12	69	20	26
Intel 10 nm	Core i7-1065G7 (formerly Ice Lake)	34	48	5	10	54	18	24

generations. It can be deduced that power-frequency has been constant for the 9th and 10th generation design built in the 14++ node. For the 10th generation design in the 10nm node, there has been an increase in chip power at the same frequency. Further, to consider the technology node-to-node benefit, we look at the turbo frequency (Figure 2-b). The turbo-frequency of the 10th generation Core-i7 in 10nm shows slight degradation as compared to the 8th and 9th generations Core-i7 in 14++ node. The 10th generation Core-i7 in 14++ shows some improvement as compared to the 8th and 9th generations design.

Based on the early parts released by Intel in 10nm, it seems that there may be an issue regarding improving power-performance node-to-node as they scaled from 14++ to the 10nm node. This contrasts with the expected trends in the previous node.

The next question that we tried to address, is whether there is a fundamental difference between Intel’s 10nm and other industry 7nm technologies. We carried out a structural comparison as well as TCAD simulation. We found that the 10/7 nm node technologies (Intel 10, and TSMC and Samsung 7) are remarkably similar in terms of structures and device features. Furthermore, to within a few percent, they have similar current drive as well as device capacitance.

IV. TCAD DEVICE SIMULATIONS

We use TCAD to understand the performance limitations of the Intel 10 nm and 14 nm products. Initially we used an ideal FinFET structure with nearly vertical sidewall FinFET with the dimensions for fin height, pitch, and width as described in Table 1. The Synopsys TCAD Suite [11] was used to model these devices. Although we do not know much of the finer details of the Intel structures, the process and device simulation deck that we used was based on and

Intel 10 nm simulation data. tapered gate with fin foot

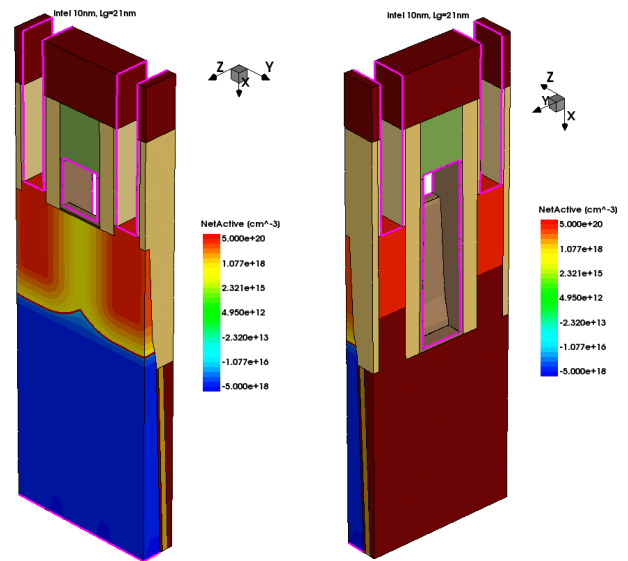


FIGURE 4. The FinFET structure used in TCAD Simulation: Left figure is cut through the fin under the gate. The right diagram is the rotated figure with the gate metal removed. The taper in the Fin and the gate are included.

validated against IBM Research’s 7 nm FinFET technology reported in [12]. The conventional Drift-Diffusion model was used, including quantum corrections via a density gradient method found within the Synopsys tool. A mobility model appropriate for thin body MOS devices was used that takes into account mobility degrade due to thin body and high-K effects as well as mobility enhancements due to stress, which also includes the impact of surface orientation and transport direction. Lumped resistances on the source and drain contact were included to model the effect of the MOL resistance including contact resistance. The simulation deck was parameterized such that the fin dimensions from Table 1 could be fed into the process simulation tool and provide a reasonable representation of the Intel FinFET structures. The gate work function in our device simulation deck was adjusted to obtain $\sim 10\text{nA}/\mu\text{m}$ off current for a nominal gate length device of $\sim 20\text{nm}$. Gate work functions of 4.46eV and 4.49eV were used for the intel 10 nm and 14 nm devices respectively. Internal device simulations using NEGF-type device transport tools have shown us that the peak internal carrier velocity of such short channel devices exceeds the conventional saturation velocity that is used in Drift-Diffusion models. In spite of this, we have also found that scaling of the saturation velocity allows us to match device on currents. We therefore scaled the saturation velocity from 1.0×10^7 cm/sec to $\sim 3.0 \times 10^7$ cm/sec in order to match the reported measured on current for the 10 nm Intel device [13]. For the 14 nm device [14], [15], the MOL resistance is expected to be less than that for the 10 nm device because the CA area is $\sim 2\text{X}$ larger although CA height is higher. For this reason, we reduced the lumped external resistance values while keeping the transport parameters the same as those for the 10nm device and we were able to

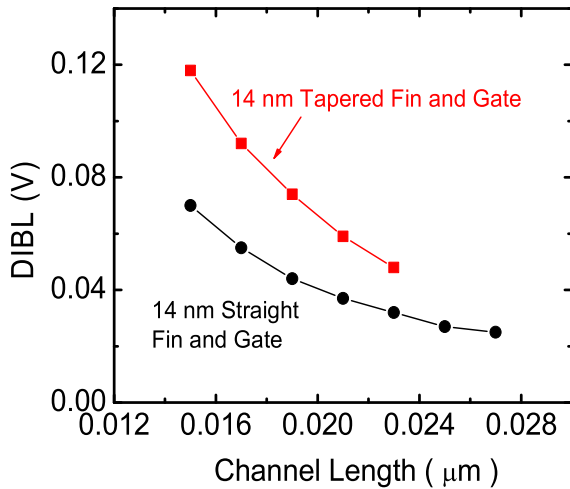


FIGURE 8. DIBL obtained from TCAD for Intel’s 14++ for ideal vertical fin and gate, as well as device with taper in fin and gate.

TABLE 3. Simulated device capacitance for Intel 14++ and 10nm structures.

	Fin Height (μm)	Fin Pitch (μm)	Cgs		Cgate (L=21nm, Ioff=10nA/μm)		~Ceff (Cgate+Cgs) per Fin	
			(ff/μm)	(aF/fin)	(ff/μm)	aF/fin	(aF/Fin)	Normalized to 14 nm
14nm	42	42	0.217	20.0	0.914	84.1	104	1
10nm	48	34	0.170	17.5	0.919	94.7	112	1.08

The impact of having a foot and tapered gate on device performance was studied. Figure 6 compares the simulated Ioff/Ion characteristics of the intel 10nm and 14nm devices for the idealized fin structures and the more realistic structures. For the 14 nm Intel node, the impact of the large foot at the bottom of the fin is immediately apparent. The gate cannot control the off current well, so the poor electrostatics reduces the drive current. For the Intel 10 nm node, it appears that the fin foot also degrades the drive current, but not as significantly as in the 14nm node. This small impact can be explained by recalling that the gate is tapered, top to bottom. For each of these non-ideal device structures, the gate length near the top of the fin is smaller than the gate length at the bottom of the fin. For the 14nm Intel node, the gate is not wide enough to control the leakage in the fin foot region. For the 10 nm Intel node however, the fin process control is seen to be better, so the added gate width at the bottom of the fin yields better electrostatics.

Table 3 summarizes the device capacitance node to node. The reduction of fin pitch (effectively reduced gate height), causes a reduction of the PC to PC and junction capacitance. Because of an increase in Fin height (or lack of scaling of the Fin height), the device width does not scale node to node. If the number of fins per function is not reduced by the scaling factor (and in fact, there can even be an increase in the fin height by about 10% node-to-node). This results in a net capacitance increase, per fin node-to-node.

TABLE 4. BEOL capacitance and resistance change node to node.

	Pitch at 14 nm	Pitch at 10 nm	Scaling Factor	Aspect Ratio
M0		40		1.59
M1	62	40	0.645161	1.19
M2	61	36	0.590164	1.15
M3	55	44	0.8	1.15
M4	52	44	0.846154	1.15
M5	52	52	1	1.6

(a)

	Relative C/μm	Relative R/μm	RC Delay (per μm ²)	Scaled RC Delay
14nm	1.0	1.0	1.0	1.0
10nm	0.84	2.83	2.40	1.2

(b)

V. BEOL SCALING

RC scaling from node-to-node is driven by several factors including pure dimensional scaling, as well as structural and materials-based changes. Comparing capacitance between 14nm and 10nm, there are several structural changes apparent which individually can act to either increase or decrease total capacitance. For example, the height-to-width aspect ratio of metal lines appears to increase from approximately 1.0 to 1.15 for 10 nm metal lines relative to 14 nm metal lines, which increases line-to-line capacitance. In addition, the via height increases from approximately 0.5 to 0.6-0.7, while the line width-to-pitch ratio decreases from approximately 0.7 to 0.6. Both factors act to decrease line-to-line capacitance. Since the metal pitch and corresponding line width both decrease in going from 14 nm to 10 nm, the metal resistivity and line resistance must also increase. This is compounded due to a change in minimum-pitch conductor from 52 nm in 14 nm to 36 at 10 nm. In the end, the non-linear increase in line resistance dominates the net RC trend, which is projected to increase by ~2.4x for 1 μm lines and by about 1.2X for scaled lines, assuming a 0.7X average shrink factor for 10nm relative to 14 nm. This is summarized in Table 4.

VI. POWER-PERFORMANCE SCALING INTO THE 10nm NODE

Two benefits have been associated with scaling: Density and power-performance. In recent nodes there has been a marked decrease of improvement in power-performance. Nevertheless, the industry has been able to take advantage of density scaling, and at the same time obtain some power-performance benefit at the product level. In the transition from 14++ to 10 nm, there has been noticeable challenge in power-performance scaling in the high performance space: In Intel’s 10 nm node (for the high performance processors) in transition from 14++, there was a full node-to-node density scaling (CPP in 14 was 42 nm), about 15% drop in Ion/μm device width (from 0.96 to 0.818 μA/μm), and slight increase in fin height resulting in about 8% device capacitance

TABLE 5. Comparisons of device features for various industry 14nm generation technologies (16-10nm nodes).

	Product Used in Construction Analysis	Fin Pitch	Fin Height	Fin Width (Middle of Fin)	Fin Width (Bottom of the Fin)	CPP	Logic nFET PC Width (Top of the Fin)	Logic nFET PC Width (Bottom of the Fin)
Intel 14	Broadwell SR217 Core™ M-5Y10	48	39-42	7-11		69	22-24	
TSMC 16 nm FinFET Plus	HiSilicon Kirin 960 (Hi3660GFC)	48	41-42	5-7	10-12	96	27 (27-33)	
TSMC 12FFN	NVIDIA GV100-400-A1	48	42	7-9		90	24	34
GF 12 nm	Ryzen 7 2700X (9500A Die)	48	44	9	18	78	28	30
Samsung 14 nm LPP	Qualcomm Snapdragon 820	48	42	8	17	78	26	
Samsung 10nm LPP	Samsung Exynos9810	42	45	5-6	12	68	21	26

increase per fin (Table 3) The BEOL capacitance has scaled properly (by about 70% for scaled lines). In critical path circuits, for the reasons of performance and noise, the circuit capacitance ratio of device to BEOL is split in the range of 70/30 to 60/40. With slight increase in the device capacitance, and a more significant decrease in the BEOL capacitance, one would expect near-constant CV^2 nod-to-node. However, since both the device and BEOL resistance have increased, in order to obtain the same performance, it is necessary to increase the operating voltage of the device resulting in power-performance degradation and reduction in turbo-frequency at the same chip power. The 10nm power-performance behavior is very similar to what was observed in transition from 22nm to 14nm: where there was noticeable challenge in power-performance: Intel’s 14 nm node, was a full node to node density shrink from the previous node (22 nm), i.e. CPP in 14 nm was 70 nm vs. 90 in 22 nm, with full BEOL shrink. There was also 20% drop in Ion/ μm device width [5], [14], and slight increase in fin height. Nevertheless, power-performance node-to-node (and software benchmarks) were flat, as shown in Figure 1 for generation 5 and 6 of Core-i3 (and Core-i7) [15]. Only by the introduction of 14+ and 14++ and increasing the Ion [16], it was possible to obtain ~25% drop in chip power at the same frequency [17] for generations 7 and 8 of the Intel Core family [5].

VII. COMPARISONS OF INDUSTRY 7nm NODE TECHNOLOGIES

As the data in the prior section indicate, it has been difficult for Intel to obtain noticeable gain in power-performance in migrating from their 14++ nm technology to their 10 nm node. The reason for this difficulty is possibly because Intel’s 14 nm exhibits particularly high performance: Table 5 has the key data on several 16-10 nm technologies as practiced by the leading manufacturers in the industry. It can be observed that Intel has the shortest channel length for that generation of the technology, the thinnest fin dimensions, and the most straight sidewalls. Intel’s 14++ device is very similar to Intel’s 10 nm node. As shown in the previous section (and expected), without scaling channel length or device width, and at the same time shrinking device (CPP) and fin pitch, it is difficult to obtain power-performance benefit.

TABLE 6. Comparisons of technologies used in AMD’s processors for 14 and 7nm generation products.

	Product Used in Construction Analysis	Fin Pitch	Fin Height	Fin Width (Middle of Fin)	Fin Width (Bottom of the Fin)	CPP	Logic nFET PC Width (Top of the Fin)	Logic nFET PC Width (Bottom of the Fin)
GF 12 nm	Ryzen 7 2700X (9500A Die)	48	44	9	18	78	28	30
TSMC 7 nm	Ryzen 5 3600X	33	49	5	10	57	22	26

(a)

		FH (nm)	FP (nm)	Fin Perimeter (nm)	Ion		Rext	
					($\mu\text{A}/\mu\text{m}$)	($\mu\text{A}/\text{fin}$)	($\Omega\text{-}\mu\text{m}$)	(Ω/fin)
12nm GF	ideal	44	48	97	0.89	86	406	4190
	non-ideal	49	30	103	0.82	84	450	4400
7nm TSMC					0.825	85	500	4800

(b)

		FH (nm)	FP (nm)	Fin Perimeter (nm)	Cgs	Cgate	Ceff	
					(ff/ μm)	(ff/ μm)	(ff/ μm)	(aF/fin)
12nm GF	ideal	44	48	97	0.239	1.081	1.32	128
	non-ideal	49	30	103	0.158	0.774	0.933	96
7nm TSMC					0.171	0.798	0.969	100

(c)

Even though Intel appears to have challenges in power-performance benefit in migrating their products to 10 nm, Advanced Micro Devices Corporation (AMD), which has a similar product family to that of Intel, has obtained about 50% power reduction in migrating their microprocessors from their 14 nm node to the 7nm node [4]: They attributed 9% of the power reduction to a drop in AC capacitance (C_{AC}), and about 12% of the power drop to “7nm Timing”, both enabled by technology migration [4]. AMD’s 14nm generation product was manufactured using Global Foundries 12nm FinFET technology. AMD’s 7nm product generation was manufactured using TSMC’s 7nm technology. The key features of these technologies are listed in Table 6. It is observed that TSMC’s 7 nm has a thinner fin (5-10nm vs. 10-18nm) and shorter channel length (22-26 nm vs. 28-30 nm). In the GF 12nm to TSMC 7 nm migration the subsequent drop in fin thickness and channel length is significantly larger than observed in Intel’s 14 nm to 10 nm migration. Thus, it appears reasonable to assume a larger drop in device capacitance and in drain induced barrier lowering (DIBL) which results in higher performance. Migration from a relatively low performance technology (as compared to Intel’s 14nm) seems to be the reason for AMD being able to obtain noticeable power drop at the same frequency attributable to technology migration.

It is worthwhile to notice that for all the 10/7 nm nodes technologies, implemented by the three major semiconductor fabricators (Intel, TSMC and Samsung Electronics), the device features are remarkably similar. Table 7 summarizes the fin dimensions and channel lengths for a 10/7nm node technologies. Looking at the images of the fin and the

TABLE 7. Comparisons of device features for the 10/7nm technologies in production by the leading manufacturers.

	Product Used in Construction Analysis	Fin Pitch	Fin Height	Fin Width (Middle of Fin)	Fin Width Bottom of the Fin	CPP	Logic nFET PC Width (Top of the Fin)	Logic nFET PC Width (Bottom of the Fin)
Intel 10 nm		34	48	5	10	54	18	24
TSMC 7FF	Ryzen 5 3600X	33	49	5	10	57	22	26
TSMC 7FF+	HiSilicon Hi3690	33	48-50	5	7	57	21	26
Samsung 7LPP	ExynosOcta 9825	35	48	5	10	68	21	28

TABLE 8. Ion Ioff and DIBL for straight and tapered fin.

		FH (nm)	FP (nm)	Fin Perimeter (nm)	Ion		Rext	
					($\mu\text{A}/\mu\text{m}$)	($\mu\text{A}/\text{fin}$)	($\Omega\text{-}\mu\text{m}$)	(Ω/fin)
10nm Intel	ideal	48	34	103	0.837	86	420	4070
	non-ideal				0.825	85	457	4440
7nm SEC	ideal	48	27	101	0.82	83	450	4440
	non-ideal				0.86	87	445	4410
7nm TSMC	ideal	49	30	103	0.82	84	450	4400
	non-ideal				0.825	85	500	4800

(a)

		FH (nm)	FP (nm)	Fin Perimeter (nm)	Cgs	Cgate	Ceff	
					(ff/ μm)	(ff/ μm)	(ff/ μm)	(af/fin)
10nm Intel	ideal	48	34	103	0.177	0.835	1.013	104
	non-ideal				0.17	0.919	1.089	112
7nm SEC	ideal	48	27	101	0.147	0.754	0.901	91
	non-ideal				0.152	0.846	0.999	101
7nm TSMC	ideal	49	30	103	0.158	0.774	0.933	96
	non-ideal				0.171	0.798	0.969	100

(b)

gate, it is noticed that there is always some taper at the bottom of the fin and increase in channel length L at the bottom of the gate. For these technologies, the channel length is about 19-20 nm at the top of the FinFET and about 25-26 nm at the bottom of the FinFET. Furthermore, the fin width is about 5 nm at the middle of the fin, and about ~10nm at the bottom of the fin. There can be a number of reasons for the taper in fin or gate: One method to address the leakage at the bottom of the fin, can be increasing L in that region. The taper in the gate can also be driven by the process requirements of the gate-last high K process. The taper at the bottom of the fin, can be driven by the requirements for the spacer etch. In order to shut down the punch-through current below the fin in the Si bulk, the very bottom of the fin usually has high doping and slightly higher threshold. Figure 7 and 8 are the TCAD study, highlighting the effect of the taper at the bottom of the fin on the channel length scaling: It cause 4-5 nm shift in L scaling (*i.e.* shift in DIBL or threshold roll of vs. L). It results in higher off current at the same I_{ON} and L, and larger source/drain resistance. The gate capacitance also increases because of the foot at the bottom of the fin. The taper in the gate and the fin, seen across 10/7nm node technologies and observed across the three leading manufacturers, raises the question if taper is a requirement for FinFET technologies, needed to control the punch through, and maybe related to why Intel has been unable to scale channel length;

Industry has announced technology roadmaps for 7 nm (Intel) and 5 nm (TSMC and Samsung) nodes and 3 nm nodes. The challenges faced by Intel in scaling into 10 nm, bring out the question of whether power-performance can be maintained, let-alone improved, with scaling in the high-performance space. One central issue is how to scale channel length. Scaling L (and “contacted poly pitch”, CPP) is critical for both performance and density scaling. It is not clear if the lack of L scaling for 10/7nm nodes relative to the previous nodes is driven by the short channel issues (caused by the FinFET taper at the bottom of the Fin), or by process related issues (*i.e.* the gate last removal and re-fill, gate doping, *etc.*). Fin width is probably at its limit. Fin taper at the bottom of the fin is one area that may be improved.

VIII. DISCUSSIONS AND SUMMARY

Going from 14nm node to 10nm node, for a very well designed 14nm node, the node-to-node intrinsic drivability (current per channel perimeter) is dropping and is compensated by an increase in fin height, principally to satisfy timing requirements for critical circuit paths. However, this increases capacitance and is apparently not adequately offset by fin pitch and gate length scaling, resulting in a net power/performance degradation at the product level for the first iteration of that node. Subsequent versions of the same node (10+, 10++, *etc.*) are expected to primarily improve the intrinsic drivability (*e.g.*, R_{ext} reduction, transport), and then later reduce device capacitance (*e.g.*, larger CPP), to achieve a net power/performance improvement, assuming history repeats itself. The broader question is whether this is sustainable going forward to 7nm, 5nm, *etc.* Increasing fin height has diminishing returns owing to increase in capacitance (and lack of device width scaling), and R_{ext} which itself increases as the fin pitch is scaled. This points to a floor in fin pitch scaling (R_{ext} -limited), which also limits capacitance scaling as an offset to increasing the fin height. Short of a breakthrough in carrier transport improvement or R_{ext} engineering, future FinFET nodes may plausibly be defined by the essentially fixed FET structure with only some wiring and ground rule tricks to achieve density scaling without any tangible performance increase.

ACKNOWLEDGMENT

The authors are grateful to Prof. Dimitri Antoniadis of MIT for technical discussions.

REFERENCES

- [1] G. E. Moore, “Progress in digital integrated electronics,” in *IEDM Tech. Dig.*, Dec. 1975, pp. 11–13.
- [2] G. Baccarani, M. R. Wordeman, and R. H. Dennard, “Generalized scaling theory and its application to a 1/4 micrometer MOSFET design,” *IEEE Trans. Electron Devices*, vol. 31, no. 4, pp. 452–462, Apr. 1984, doi: 10.1109/T-ED.1984.21550.
- [3] E. Fayneh, M. Yuffe, E. Knoll, M. Zelikson, M. Abozaed, Y. Talker, Z. Shmueli, and S. A. Rahme, “14 nm 6th-generation core processor SoC with low power consumption and improved performance,” in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Jan. 2016, pp. 72–73, doi: 10.1109/ISSCC.2016.7417912.

- [4] T. Singh, S. Rangarajan, D. John, R. Schreiber, S. Oliver, R. Seahra, and A. Schaefer, "Zen 2: The AMD 7 nm energy-efficient high-performance x86-64 microprocessor core," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2020, pp. 42–44, doi: [10.1109/ISSCC19947.2020.9063113](https://doi.org/10.1109/ISSCC19947.2020.9063113).
- [5] G. G. Shahidi, "Chip power scaling in recent CMOS technology nodes," *IEEE Access*, vol. 7, pp. 851–856, 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8570735/references#references>
- [6] Intel. (Apr. 2020). *Intel's 8th Generation Core-i7 Data*. [Online]. Available: <https://ark.intel.com/> and <https://ark.intel.com/products/series/122593/8th-Generation-Intel-Core-i7-Processors>
- [7] I. Cutress. (Jan. 2018). *Intel's 10 nm Cannon Lake and Core i3-8121U Deep Dive Review*. [Online]. Available: <https://www.anandtech.com/show/13405/intel-10nm-cannon-lake-and-core-i3-8121u-deep-dive-review>
- [8] I. Cutress. (Oct. 2018). *The Intel 9th Gen Review: Core i9-9900K, Core i7-9700K and Core i5-9600K Tested*. [Online]. Available: <https://www.anandtech.com/show/13400/intel-9th-gen-core-i9-9900k-i7-9700k-i5-9600k-review/21>
- [9] H. Wong. (Oct. 2012). *A Comparison of Intel's 32 nm and 22 nm Core i5 CPUs: Power, Voltage, Temperature, and Frequency*. [Online]. Available: <http://blog.stuffedcow.net/2012/10/intel32nm-22nm-core-i5-comparison/>
- [10] Techinsight. *Product Structural Features, Images, Shapes, and Key Dimensions Were Obtained From the Product Specific Technical Reports*. Accessed: Mar. 2020. [Online]. Available: <https://www.techinsights.com/>
- [11] Synopsys *Sentaurus TCAD Ver. L-2016.03*, Synopsys, Inc., Mountain View, CA, USA, 2016. [Online]. Available: <https://www.synopsys.com/home.aspx>
- [12] R. Xie et al., "A 7 nm FinFET technology featuring EUV patterning and dual strained high mobility channels," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2016, pp. 2.7.1–2.7.4, doi: [10.1109/IEDM.2016.7838334](https://doi.org/10.1109/IEDM.2016.7838334).
- [13] 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," in *IEDM Tech. Dig.*, Dec. 2017, pp. 29.1.1–29.1.4, doi: [10.1109/IEDM.2017.8268472](https://doi.org/10.1109/IEDM.2017.8268472).
- [14] S. Natarajan et al., "A 14 nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size," in *IEDM Tech. Dig.*, Dec. 2014, pp. 3.7.2–3.7.5, doi: [10.1109/IEDM.2014.7046976](https://doi.org/10.1109/IEDM.2014.7046976).
- [15] R. Brain. (Mar. 2017). *14 nm Technology Leadership*. [Online]. Available: <https://newsroom.intel.com/newsroom/wp-content/uploads/sites/11/2017/03/Ruth-Brain-2017-Manufacturing.pdf>
- [16] I. Cutress. (Aug. 2015). *The Intel 6th Gen Skylake Review: Core i7-6700K and i5-6600K Tested*. [Online]. Available: <https://www.anandtech.com/show/9483/intel-skylake-review-6700k-6600k-ddr4-ddr3-ipc-6th-generation>
- [17] I. Cutress. (Oct. 2018). *The Intel 9th Gen Review: Core i9-9900K, Core i7-9700K and Core i5-9600K Tested*. [Online]. Available: <https://www.anandtech.com/show/13400/intel-9th-gen-core-i9-9900k-i7-9700k-i5-9600k-review/21>



PHIL OLDIGES (Senior Member, IEEE) received the B.A. degree in physics from the Thomas More College, Ft. Mitchell, KY, in 1981, the B.E. degree in electrical engineering from the University of Dayton, Dayton, OH, in 1981, and the M.S. degree in electrical engineering and the Ph.D. degree in electrophysics from Cornell University, Ithaca, NY, in 1984 and 1988, respectively.

From 1984 to 1986, he was a Visiting Researcher with Toshiba Corporation, investigating alpha-particle induced charge collection dynamics. From 1988 to 1993, he developed and implemented physical models for high-field transport in FET devices at Sony Corporation. From 1993 to 1998, he has worked at Digital Equipment Corporation, where he developed models for front end process simulation and tools for soft error evaluation. Since 1998, he has been with IBM Research, where he has been the focal point for soft error modeling, and the Manager of the IBM Research TCAD Group for 15 years, responsible for front end process and device models for the 7-nm node and beyond. His current research interests include neuromorphic device modeling and device performance limitations of beyond 7-nm technology.



REINALDO A. VEGA (Member, IEEE) received the B.S. and M.S. degrees in microelectronic engineering from the Rochester Institute of Technology (RIT), Rochester, NY, in 2004 and 2006, respectively, and the Ph.D. degree in electrical engineering from the University of California at Berkeley, in 2010.

In 2010, he joined the Semiconductor Research and Development Center, IBM, East Fishkill, NY, as a CMOS Device Design Engineer. He has been actively involved in CMOS technology development from the 20-nm node onwards, spanning SOI and bulk planar and non-planar FET architectures. In 2015, he moved to IBM Research, focusing heavily on modeling and performance benchmarking of FinFET and beyond-FinFET architectures. He is currently a Senior Engineer and an IBM Master Inventor, having authored or coauthored over 100 patents. His research interests include advanced CMOS device design, AI and neuromorphic computing devices, competitive analysis, and developing representative benchmarking techniques.

Dr. Vega has served as a Reviewer for the IEEE ELECTRON DEVICE LETTERS, the IEEE TRANSACTIONS ON ELECTRON DEVICES, the IEEE TRANSACTIONS ON NANOTECHNOLOGY, and the IEEE TRANSACTIONS ON EDUCATION. He was a recipient of the First Place Award at the 2004 RIT IEEE Student Design Contest for his earlier work on Schottky barrier MOSFETs. He was also a recipient of the 2004 Professor I. Renan Turkman Scholarship for Outstanding Achievements in Semiconductor Device Engineering, the 2004 RIT Undergraduate Research Symposium Award of Excellence, the 2005 RIT Intellectual Property Productivity Award, the Honorable Mention for the 2005 NSF Graduate Research Fellowship Program (GRFP), the 2006 IBM/Global Research Corporation Ph.D. Fellowship, the Best in Session at SRC TECHCON 2008 and 2009, the 32 IBM Invention Plateaus, and other IBM-internal awards.



HENRY K. UTOMO received the B.S. degree in chemical engineering and the M.S. degree in electrical engineering from the University of Minnesota, Minneapolis, MN, USA, in 1999 and 2001, respectively.

Since 2001, he has been a Process Integration Engineer with IBM, where he was involved in the development of CMOS and memory technology.



NICK A. LANZILLO received the B.S. degree in physics from Syracuse University, Syracuse, NY, in 2009, and the Ph.D. degree in physics from the Rensselaer Polytechnic Institute, Troy, NY, in 2014.

He is currently a Research Staff Member with IBM Research, Albany, NY. He has authored or coauthored more than 30 technical publications and holds more than ten patents. His research interests include materials science, semiconductor physics, interconnect integration, design-technology co-optimization (DTCO), and machine learning.



THOMAS WASSICK (Member, IEEE) received the B.S. degree in biomedical engineering and the M.S. degree in materials engineering from the Rensselaer Polytechnic Institute, in 1980 and 1982, respectively.

He is currently a Senior Technical Staff Member in the Packaging Development organization within the IBM Systems Group, East Fishkill, NY. In his years at IBM, he was worked in multiple technology areas all related to electronics packaging technology, most notably in the areas of thin film wiring and repair technology and laser materials processing. He is also responsible for electrical diagnostics and failure analysis for IBM's chip packaging technology and is the focal point for interconnect electromigration activities at IBM. He holds over 40 U.S. Patents and has been the author or a coauthor on numerous journal publications tied to his work at IBM.

Mr. Wassick is a member of IMAPS.



JUNLI WANG (Member, IEEE) received the Doctor of Engineering degree in applied science for electronics and materials from Kyushu University, Fukuoka, Japan, in 2001.

From 2001 to 2003, he was a Postdoctoral Researcher with Kyushu University. From 2003 to 2008, he was with SONY Corporation, Japan, as a Semiconductor Research Engineer, focusing on advanced CMOS device integration. In 2008, he joined IBM Research, Albany, NY, as a Research Staff Member. Since then, he has been working on advanced logic CMOS research, focusing on FINFET technology, especially FEOL integration, metal gate process, process assumptions, DTCCO, and future technology node definitions. He has authored or coauthored more than 30 publications. He has applied and filed more than 260 patents.



JUNTAO LI received the B.S. and M.S. degrees in electronic engineering from Southeast University, Nanjing, China, in 2003, and the Ph.D. degree in nanoscale engineering from the College of Nanoscale Science and Engineering, University at Albany-SUNY, Albany, NY, USA, in 2010. He joined IBM Research at Albany Nanotech, Albany, NY, USA, in 2010, where he is currently a Senior Engineer with the Department of Inline Characterization and Advanced Failure Analysis.

His current research interests include advanced analytical characterizations of emerging CMOS and non-volatile memory technologies. He is the author or a coauthor of more than 100 technical journal articles and conference proceedings, and one book chapter. He is a Master Inventor of IBM with over 250 U.S. patents and patent applications.



GHAVAM G. SHAHIDI (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from MIT. In 1989, he joined the IBM Thomas J. Watson Research Center, where he initiated the SOI Research Program which resulted in the first use of SOI in mainstream CMOS. In starting 2000, he led the development of multiple CMOS technology generations as the Director of High-Performance Logic Development with the IBM's Semiconductor Research and Development

Center. From 2003 to April 2015, he worked on CMOS technology scaling down to single digit nodes, as the Director of Silicon Technology with the IBM Research Division. He is currently with IBM Research. He is an IBM Fellow and a Foreign Member of the National Academy of Engineering.

...