Present and Future of FinFETs in CMOS Scaling

Evan Dorsky

***Abstract*—FinFETs offer many advantages over modern planar MOSFETs, while requiring only minor modifications to CMOS processes. They are currently manufactured and commercialized. At nanoscale nodes, FinFETs are subject to quantum effects that have the potential to both enhance and degrade device behavior. FinFETs remain difficult to manufacture and scale, yet are now replacing planar MOSFETs as device sizes, densities and performance continue to advance.**

***Index terms*—Semiconductors, FinFETs, SOI, nanotechnology, next-generation technologies.**

I. Introduction

THE push to scale CMOS through microscale and into nanoscale dimensions, while improving device characteristics, has led to a series of significant process and design modifications. In the past few generations, gate-length scaling has lagged behind total device scaling [1]. To continue the scaling progress and improve device performance and uniformity, innovative methods have been used. These methods include: channel strain, to increase carrier mobility [2]; and novel metal alloy gates with high-κ dielectric materials using rare-earth elements, to improve the control of the gate on channel carriers [1]. One promising method for device scaling is the development of the FinFET, a transistor design that increases the capacitive coupling between the channel and gate by exposing more channel surface area to the gate within the same planar area. FinFETs have been in consumer devices for years [3], and they continue to offer promise in performance improvements, such as reduced power consumption and higher density, compared to planar CMOS.

A. History

The first FinFET was fabricated in 1998, with a process based on planar silicon on insulator (SOI). A thin fin of silicon (width of 15 nm, height of 50 nm) was created on top of the insulating oxide layer of an SOI wafer [1, 4]. FinFET implementation faced numerous challenges before achieving successful production.

One challenge is the fact that the widths required to make effective fins are too narrow for present lithographic methods. Spacer lithography (illustrated in Fig. 1) was developed to allow for smaller fin pitch (distance between fins) and narrower fins [1]. The fin thickness is determined by the thickness of the mask layer, so this thickness will be uniform even if the fin edges themselves are jagged or rippled [1]. Because mobile carrier distributions in thin fins vary significantly with the thickness of the fin (as discussed in Section II-B1), the device operation is more predictable and reproducible when the fin widths are uniform.

Submitted to Dr. Sherra Kerns, April 24, 2015. Accepted April 24, 2015
E. Dorsky is a student in the electrical and computer engineering program at the Franklin W. Olin College of Engineering, MA, 02492 USA e-mail: evan.dorsky@gmail.com.

Hydrogen annealing at temperatures around 1000°C has been used to smooth the fin edges and corners. At annealing temperatures in the presence of pure hydrogen gas, Silicon atoms on the surface of the fins will flow into lower-energy (smoother) configurations. Lower surface roughness decreases carrier scattering, increasing effective carrier mobility. Hydrogen annealing also significantly decreases leakage current [5].

In 2004, FinFETs with high-κ dielectric materials and metal gates were introduced to further improve gate-channel capacitive coupling, increasing transistor speed and reducing leakage current [1]. FinFETs with high-κ dielectrics and metal gates are currently in production, but as they scale further, additional challenges emerge. Process parameter variations lead to nonuniform threshold voltage, V_T. In an ideal MOSFET,

$$V_T = 2\phi_F + \frac{\kappa_S x_o}{\kappa_O}\sqrt{\frac{4qN_A}{\kappa_S \varepsilon_0}\phi_F}\,, \qquad (1)$$

where κ_S is the semiconductor dielectric constant; κ_O is the oxide dielectric constant; N_A is the p-type substrate doping concentration $\left[\frac{acceptors}{cm^3}\right]$; and $\Phi_F = \frac{1}{q}[E_i(bulk) - E_F]$ [6]. Equation (1) shows that V_T depends on the oxide thickness (x_o) and the channel doping (with N_A and Φ_F) both of which are difficult to control at modern device dimensions. Variations in these parameters between transistors need to be matched, otherwise circuit performance degrades [7]. Parasitic capacitances not present in planar CMOS are also caused by coupling between the fin-top hard mask and gate, as well as other elements of the FinFET structure [1].

The principles governing FinFET operation are related to those of the planar MOSFET. Section II offers a brief discussion on the limits of MOSFET before focusing on FinFETs. In Section III, the advantages of the FinFET over

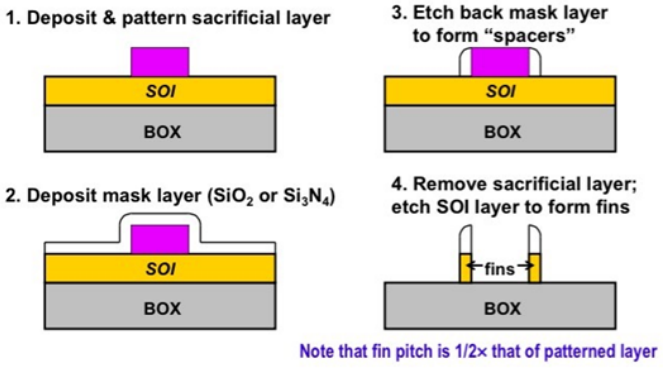

Fig. 1. Process flow of spacer lithography [1]

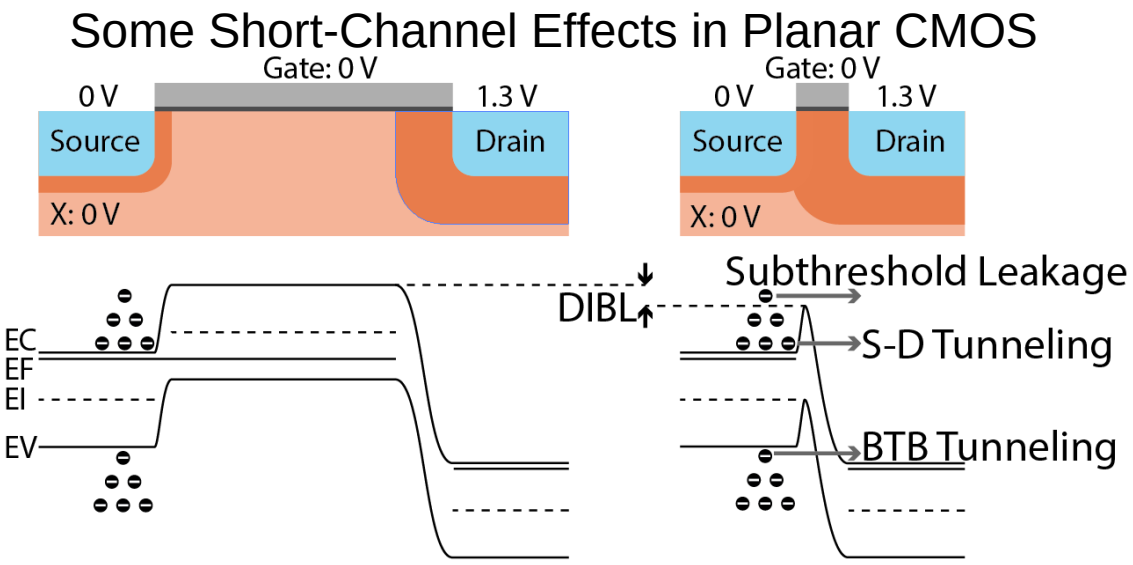

Fig. 2. Diagram showing how a short-channel MOSFET (right) has sources of leakage current that a longer-channel MOSFET (left) does not, even when $\boldsymbol{V_{GS}}$ is 0 V. The band diagrams correspond to the simplified device diagrams. After [8]

conventional planar CMOS are explained, as well as the current challenges and potential solutions that will help FinFETs continue to scale.

II. Basic Device Facts

A. *Limits of Planar MOSFETs*

In MOSFETs, depletion regions characteristic of p-n junctions form around both the drain and source terminals. When the channel becomes short, these depletion regions can interact, diminishing the gate's control over the channel (Fig. 2). Much of the work required to scale devices during the last decades has been devoted to minimizing this interaction and the accompanying "subthreshold leakage" when the gate voltage is below V_T. In nanoscale devices, electrons can tunnel straight through the channel from source to drain as shown in Fig. 2. By introducing forward leakage current, these effects cumulatively lower the threshold voltage of the MOSFET, diminishing the gate's control over the channel current. At channel lengths below 50 nm, quantum effects on V_T can no longer be ignored [9]. Novel device geometries (like those of FinFETs) have been shown to be effective in combatting short-channel effects.

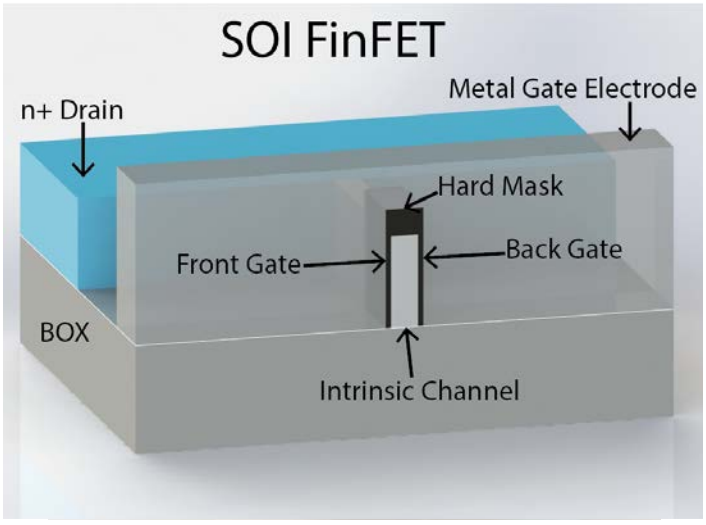

Fig. 3. The SOI FinFET with important elements labeled. The transistor is mirrored about the cutaway plane, with the source on the other side. After [10]

B. *The FinFET*

The FinFET operates on many of the same basic principles as the planar MOSFET, especially with a fin thickness above 20 nm [11]. In a FinFET, the channel is raised above the substrate in a narrow fin-like structure as shown in Fig. 3. As each vertical face of the fin provides one width equal to the height of the fin (H_{FIN}), the effective gate width is doubled:

$$W_{FIN} = 2 * H_{FIN}. \quad (2)$$

This geometry increases the surface area of and capacitive coupling between the channel and gate. The top face of the fin is often masked over to keep the gate from influencing it, simplifying the manufacturing process [1]. Greater coupling increases the I_{ON}/I_{OFF} ratio, reducing leakage current for a given ON current and increasing switching speeds [1].

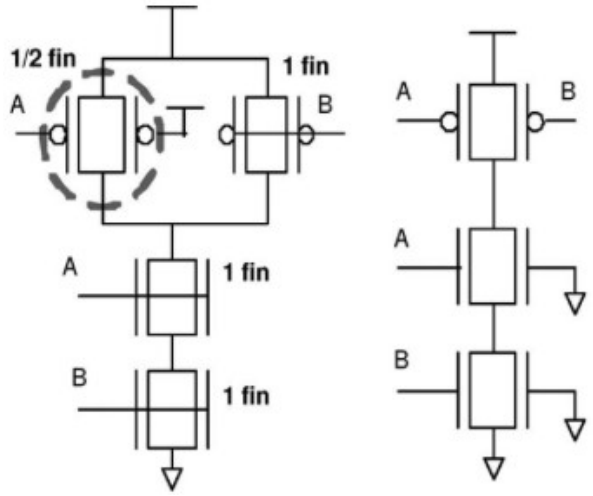

Fig. 4. Diagram of a CMOS NAND gate showing how independent control of each gate of a DG FinFET allows two parallel fins (top of left circuit) to be merged into one (top of right circuit) [14]

Equation (2) holds when the two gates are electrically connected, the most common configuration for FinFETs operating as replacement devices in planar MOSFET designs. However, the two gates of the FinFET can also be electrically independent, allowing for post-process V_T tuning or novel logic gate designs as seen in Fig. 4 [12-14]. The dual gate structure and thin fin design of the FinFET affect the quantum-mechanically "allowed states" for carriers moving within the fin [11].

1) Quantum Effects on FinFET operation: Classically, the inversion layer in the channel of an ON MOSFET is extremely thin and located at or near the channel-gate boundary. In ultra thin body (UTB) FinFETs (with fin thickness below 10 nm), the allowed energy levels for mobile carriers are quantized, rather than in "bands". Classical models do not apply; allowed states and carrier densities depend strongly on fin thickness, as shown in Fig. 6 [11]. For wide fins (in ON devices), the gates on either side of the fin draw carriers near the fin edges. Since the fin-gate interfaces exacerbate carrier scattering, mobility is effectively lowered. UTB fins move carriers through the center of the channel, decreasing scattering and increasing mobility. The quantization of conduction band energies also increases the device threshold voltage, in part because the minimum conduction state energy is higher than the classically predicted conduction band energy.

III. Advantages and Challenges

The geometry of modern FinFETs offers both advantages and challenges.

A. Speed

Dual gate (DG) MOSFETs, of which FinFETs are one type, have been shown to have a lower subthreshold swing than comparable planar bulk MOSFETs [12]. The subthreshold swing is the inverse of the slope of the transistor I_{DS} vs. V_{GS} curve with units $\left[\frac{mV}{decade}\right]$. Subthreshold swing is inversely proportional to switching speed, because the more current flowing through the channel at a given V_{GS}, the more quickly the MOSFET will "switch" states. The speed increase

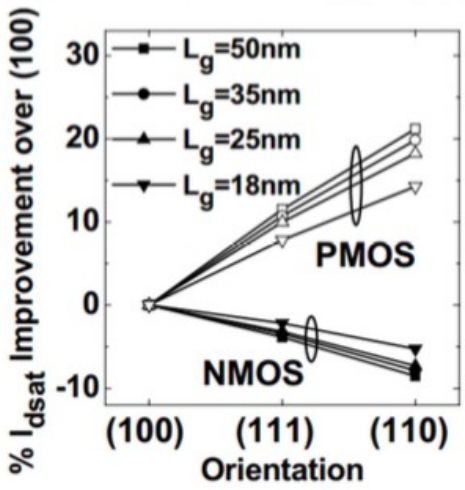

Fig. 5. Effect of lattice channel lattice direction on carrier mobility in FinFETs [1]

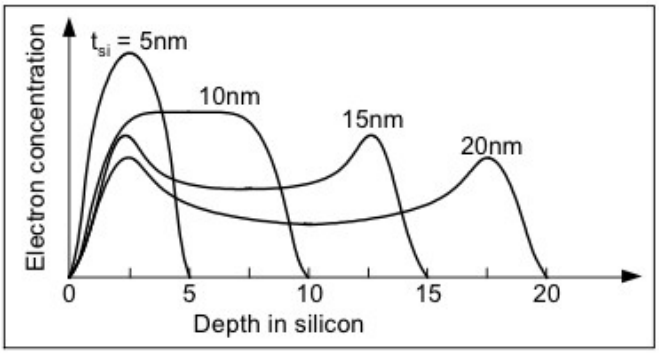

Fig. 6. Mobile carrier concentration in the center of the fin increases as fin width decreases [11]

in DG MOSFETs arises from: the greater channel control, from better capacitive coupling between the channel and gate; and the higher carrier mobility, arising from fewer carrier surface interactions compared to single-gate planar MOSFETs.

1) Optimal Lattice Direction: The optimal lattice directions for p-channel and n-channel FinFET fins are not the same [1]. As shown in Fig. 5, electron mobility is highest in the (100) direction, while hole mobility is highest in the (110) direction. Integration of (100) and (110) channel direction FinFETs would allow for higher energy efficiency, yet also increases the footprint of a given circuit. Fabricating p-channel devices in different orientations relative to n-channel devices requires different circuit layouts than have been traditionally used in advanced CMOS, if total circuit area is to be minimized.

B. Density

According to the 2013 International Technology Roadmap for Semiconductors (ITRS), FinFETs had a half-pitch minimum smaller than that of the cutting-edge logic in 2013, at the "16/14" nm technology node. They are predicted to continue this trend in 2015, with a half-pitch of 24 nm [15]. However, while FinFETs are at the forefront of CMOS scaling, they face physical limitations (Section III-B1) and practical considerations (Section III-B2, Section III-B3). Further technology innovations will be required to continue device scaling.

1) Random Dopant Fluctuations: All nanoscale MOSFETs suffer from random dopant fluctuations (RDF) to varying degrees [1]. Nanoscale MOSFETs have so few atoms that even the "heavily doped" source and drain regions actually have a small number of dopant atoms as seen in Fig. 7. A difference of just a few dopant atoms between transistors could lead to significant discrepancies in important parameters, such as threshold voltage, and such variations are extremely difficult to control. For complex ICs to operate, the

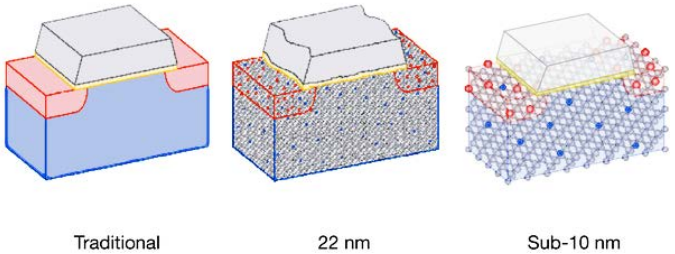

Fig. 7. Diagram illustrating the effect of device size on sensitivity to minor variations in dopant atom location [16]

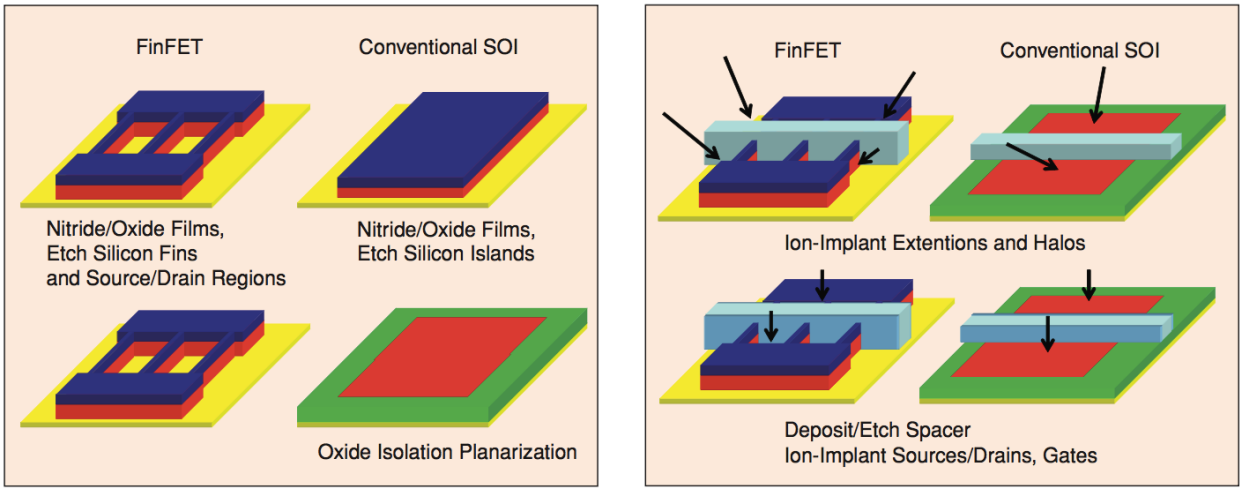

Fig. 8. FinFET and planar MOSFET process comparison [12]

total distribution of parameters must be very small. Threshold voltages must be tightly matched, and margins for variation are ratiometric, so they decrease with feature size. The lightly doped channel is even more sensitive to RDF than the source or drain, because a few atoms in the channel represent a larger percentage of the (smaller) desired number [15].

Many modern processes use intrinsic semiconductor channels along with the SOI process, known as FD-SOI, to mitigate RDF sensitivity. An intrinsic channel also allows for a lower V_T, as the channel is essentially already in the "depleted" phase at zero gate voltage. Adding charge to the gate will immediately increase the concentration of oppositely-charged carriers at the channel-gate interface, allowing current to flow through the channel.

2) Device Width Quantization: In planar CMOS processes, transistor driving strengths are related to transistor dimensions. The driving strength S of a transistor is defined as

$$S = \frac{W}{L} \tag{3}$$

where W and L are the width and length of the transistor. The strength acts as a multiplier of the channel current: a transistor with a strength of 2 will have twice the channel current of a unit-strength transistor when all terminal voltages are equal. Many digital and analog circuits operate based on strength ratios between transistors. The effective width W of a FinFET is equal to $2 * H_{fin}$ (Equation (2)), twice the height of the fin. Because the height of fins is set by an etching step in the process, it is uniform across the die. So the effective width of a FinFET can only be changed by giving one transistor multiple fins. The equivalent expression for FinFET strength, S, is

$$S = \frac{N*2H_{fin}}{L} \tag{4}$$

where N is the number of fins connected in parallel. This relationship restricts the potential widths of FinFETs to a discrete set of values. As this represents a departure from the conventional method of sizing transistors, circuit designers have adopted new design techniques and software to build circuits that take width quantization into account [12].

3) Process Variation Sensitivity: As FinFETs have scaled, they have become even more sensitive to process variations than planar MOSFETs. Small, random variations in the process can lead to device failure, but even in cases when the devices themselves work, entire circuits can still fail if the relative transistor lengths vary enough to upset a critical driving strength ratio between two transistors, as can be the case in SRAM [7].

C. Manufacturability and Cost

SOI FinFET manufacturing requires only modest modifications to existing planar CMOS SOI processes (if crystal orientation optimization is ignored) [12], contributing to the current prominence of FinFETs. A simplified diagram of the process flows of planar CMOS and FinFETs is shown in Fig. 8. In both cases, semiconductor and insulator films are first applied to the body oxide and etched. In the FinFET process, the fins are etched at this stage. UTB processes can use spacer lithography (Section I-A) here to etch ultra-thin fins. Next, conventional SOI requires an additional step in which the oxide isolating individual devices is deposited and planarized. Following this step, in both processes, the gate is deposited and the source and drain are doped. FinFETs require angled doping to reach all of the source/drain extension regions. The fin pitch limits the doping angle [1], as adjacent fins can occlude one another. Specialized doping also differs with device geometry, but overall the processes are remarkably similar. Though modern high-κ and metal-gate FinFETs require expensive materials like hafnium [3], these materials are just as necessary in conventional CMOS, which faces even greater challenges with capacitive coupling than FinFETs. In recent years, FinFETs have been successfully demonstrated as the successors to planar CMOS.

D. Market

Consumer applications of FinFETs are now commonplace. Intel uses them throughout its Ivy Bridge processors [3]. Significant research has gone into using FinFETs to build standard 6T and 8T SRAM cells, and width quantization-related problems have been identified and largely solved with new design techniques and tools [7, 17, 18].

IV. Conclusion

FinFETs have met with significant success since their introduction, as they are much less prone to the deleterious short-channel effects of planar devices. Still, CMOS devices will only be able to scale so far. The thermodynamic subthreshold swing limit for MOSFETs ($\frac{60mv}{decade}$) limits switching speed and I_{ON}/I_{OFF} ratio. FinFETs have been demonstrated with channels as thin as 5 nm, so there is still some room for the FinFET to scale. It may be the dominant technology for the next few nodes to come [15].

Acknowledgements

The author would like to thank Stephanie Northway, Dr. Sherra Kerns, Dr. David Kerns, the Wayback Machine, and the entire Olin Semiconductor Devices Spring 2015 class, especially review group A (Cameron Anderson, Emma Regan and Shaun Ong).

References

[1] T. K. Liu, “Finfet history, fundamentals and future,” Department of Electrical Engineering and Computer Sciences University of California, Berkeley, CA, Tech. Rep., June 2012, http://www.eecs.berkeley.edu/~tking/presentations/KingLiu_2012VLSI-Tshortcourse.

[2] S. Flachowsky, A. Wei, R. Illgen, T. Herrmann, J. Hontschel, M. Horstmann, W. Klix, and R. Stenzel, "Understanding strain-induced drive-current enhancement in strained-silicon n-mosfet and p-mosfet," Electron Devices, IEEE Transactions on, vol. 57, no. 6, pp. 1343-1354, June 2010.

[3] D. James, "Intel ivy bridge unveiled — the first commercial tri-gate, high-k, metal-gate cpu," in Custom Integrated Circuits Conference (CICC), 2012 IEEE, Sept 2012, pp. 1-4.

[4] D. Hisamoto, W.-C. Lee, J. Kedzierski, E. Anderson, H. Takeuchi, K. Asano, T.-J. King, J. Bokor, and C. Hu, "A folded-channel mosfet for deep-sub-tenth micron era," in Electron Devices Meeting, 1998. IEDM

[5] Y.-K. Choi, L. Chang, P. Ranade, J.-S. Lee, D. Ha, S. Balasubramanian, A. Agarwal, M. Ameen, T.-J. King, and J. Bokor, "Finfet process refinements for improved mobility and gate work function engineering," in Electron Devices Meeting, 2002. IEDM '02. International, Dec 2002, pp. 259-262.

[6] R. F. Pierret, Semiconductor Device Fundamentals. Addison-Wesley, 1996, pp. 572, 581, 617.

[7] B. Raj, A. Saxena, and S. Dasgupta, "Nanoscale finfet based sram cell design: Analysis of performance metric, process variation, underlapped finfet, and temperature effect," Circuits and Systems Magazine, IEEE, vol. 11, no. 3, pp. 38-50, thirdquarter 2011.

[8] A. Whitcombe, "Multigate transistors: a nanoscale alternative to planar cmos," 2011.

[9] D. Zhang, Z. Yu, and L. Tian, “Threshold voltage correction model for quantum short channel,” in *Solid-State and Integrated Circuits Technology, 2004. Proceedings. 7th International Conference on*, vol. 2, Oct 2004, pp. 946–949 vol.2

[10] S. Chaudhuri and N. Jha, "Finfet logic circuit optimization with different finfet styles: Lower power possible at higher supply voltage," in VLSI Design and 2014 13th International Conference on Embedded Systems, 2014 27th International Conference on, Jan 2014, pp. 476-

[11] 482. J.-P. .Coligne, FinFets and Other Multi-Gate Transistors, ser. Series on Integrated Circuits and Systems. Springer Science+Business Media, LLC, pp. 30, 2008.

[12] E. Nowak, I. Aller, T. Ludwig, K. Kim, R. Joshi, C.-T. Chuang, K. Bernstein, and R. Puri, "Turning silicon on its edge [double gate cmos/finfet technology]," Circuits and Devices Magazine, IEEE, vol. 20, no. 1, pp. 20-31, Jan 2004.

[13] M. Masahara, K. Endo, Y. Liu, S. O'uchi, T. Matsukawa, R. Surdeanu, L. Witters, G. Doornbos, V. Nguyen, G. Van den Bosch, C. Vrancken, M. Jurczak, S. Biesemans, and E. Suzuki, "Four-terminal finfet device technology," in Integrated Circuit Design and Technology, 2007. ICICDT '07. IEEE International Conference on, May 2007, pp. 1-4.

[14] A. Datta, A. Goel, R. Cakici, H. Mahmoodi, D. Lekshmanan, and K. Roy, "Modeling and circuit synthesis for independently controlled double gate finfet devices," Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on, vol. 26, no. 11, pp. 1957-1966, Nov 2007.

[15] I. W. Group, "2013 overall roadmap technology characteristics (ortc) table," International Technology Roadmap for Semiconductors, Tech. Rep., 2013.

[16] I. L. Markov, "Limits on fundamental limits to computation," Nature, vol. 512, no. 7513, pp. 147-154, 08 2014. [Online]. Available: http://dx.doi.org/10.1038/nature13570

[17] M. Bayoumi and A. Dutta, "Finfet based sram design: A survey on device, circuit, and technology issues," in Electronics, Circuits and Systems (ICECS), 2014 21st IEEE International Conference on, Dec 2014, pp. 387-390.

[18] M. Turi and J. Delgado-Frias, "An evaluation of 6t and 8t finfet sram cell leakage currents," in Circuits and Systems (MWSCAS), 2014 IEEE 57th International Midwest Symposium on, Aug 2014, pp. 523-526.

Evan Dorsky was born in Avon, CT, in 1994. He is currently pursuing a BSECE from Olin College in Needham, MA. His areas of interest include web development, undergraduate engineering education, and power electronics.