7.4 FinFET — A Quasi-Planar Double-Gate MOSFET

Stephen H. Tang, Leland Chang, Nick Lindert, Yang-Kyu Choi, Wen-Chin Lee[1], Xuejue Huang, Vivek Subramanian, Jeffrey Bokor, Tsu-Jae King, Chenming Hu

Dept. of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA
[1]Now with Intel Corp., Hillsboro, OR.

Scaling conventional CMOS transistors much below 50nm is difficult. Control of leakage currents requires gate dielectrics so thin and bodies doped so heavily that a process window sufficiently large for manufacturing might not be found. Double-gate MOSFET structures can overcome these and other limitations to transistor scaling. By placing a second gate on the opposite side of the device, the gate capacitance to the channel is doubled and the channel potential is better controlled by the gate electrode, thus limiting the leakage current.

The FinFET is a double-gate device that experimentally achieves 820μA/μm pMOS I_{dsat} and 18nm gate length (L_g) with a 2.5nm gate oxide [1]. As illustrated in Figure 7.4.1, it is a device in which a thin, fin-shaped body is straddled by the gate, forming two self-aligned channels that run along the sides of the fin. Although the gates protrude up out of the wafer plane, current flows in a plane parallel to the wafer plane, making the FinFET a quasi-planar structure.

Other double-gate structures are proposed using techniques such as pattern-constrained selective-epitaxy [2, 3], surround-gate [4], or fins on bulk silicon [5]. The challenge is to find a manufacturable structure. The FinFET is developed with special emphasis on process simplicity and compatibility with conventional planar CMOS technology. Every process module used to fabricate the FinFET is in widespread use today.

FinFET width can be increased easily by drawing multiple fins in parallel. The FinFET is not fully planar, so its current per unit width must be carefully calculated. Since gate straddles the fin, the active device width is actually the height of the fin. To achieve a fair comparison with standard fully-planar structures, the consumption of chip real estate must be considered; the total current flowing in the fin divided by the fin pitch (thickness of the fin plus the space between fins) gives the current per unit width of real estate. Figure 7.4.2 shows that when the fin pitch equals the fin height, the FinFET has the same active width as that of a fully planar device consuming the same chip area. Note, though, that the FinFET gives twice the current per unit width because of its double-gate nature. Thus, the break-even point in terms of current is when the fin pitch is twice the fin height. Figure 7.4.2 also illustrates how fin pitch can be half the lithography pitch by using spacers as the fin-etching mask.

FinFET layout is essentially the same as conventional transistor layout except that the active area is composed of fins rather than a single rectangle. Schematic layouts of a conventional device and a FinFET are compared in Figure 7.4.3 above an SEM picture of a FinFET in process. The SEM shows fins that are 20nm wide and spaced 160nm apart. Figure 7.4.3 also shows that instead of conventional contact holes, the fins may be contacted together by strapping them all with metal. This allows source/drain contact to be made along the sides and ends of the fin. As a result, contact area can be larger than real estate consumed and increases proportionally with fin height.

To investigate the potential circuit performance of FinFET technology, the 2D device simulator MEDICI [6] is used to perform mixed-mode simulation of various benchmark circuits. It is demonstrated that MEDICI simulation matches experimental FinFET data quite well [1]. Therefore, 2D mixed-mode circuit simulation should yield reasonable results. The FinFET is compared with the leading competitors to the double-gate MOSFET: the ground-plane (GP) [7, 8] and ultra-thin body (UTB) MOSFETs [9]. Cross sections of these various structures are shown in Figure 7.4.7. For these simulations, all electrical gate oxides are 20Å thick and gate lengths are 50nm. All devices are designed to give the same leakage current by changing the gate workfunction. Each circuit is connected in a ring oscillator arrangement so that input waveforms are consistent with output waveforms.

The benchmark circuits include a fan-out of 4 (FO4) inverter, NAND pull-down stack, and pass-gate multiplexer. The circuit schematics are shown in Figure 7.4.8. The FO4 inverter delay is a standard technology benchmark used to predict delay of more complex circuits. The NAND and pass-gate structures consider the impact of the body effect on performance, since in both these circuits, the source nodes of various transistors drift from the supply rails. Figure 7.4.4 shows the delay of the various circuits. The FinFET outperforms the other two structures for all supply voltages and circuits. The GP device, with its high body effect and non-ideal subthreshold swing, is significantly slower than FinFET and UTB.

GP and double-gate are the best candidates for scaling below 50nm gate length [7]. Although the GP device does not have two channels like the double-gate, it does have the feature of a tunable threshold voltage controlled by the voltage applied to the ground-plane. As in Reference 8, the ground-plane is assumed to be p^+/n^+-poly for nMOS/pMOS devices. However, for this study the back-gate oxide is assumed to be the same thickness as the front oxide. This makes V_t a strong function of the bias applied to the ground-plane.

Another set of simulations are run with FinFET and GP designed with 35nm gate length and 12Å gate oxide. This time, fan-out of 2 (FO2) inverters is used and capacitors are added between each stage of the oscillator to represent the wiring capacitance (the previous simulations assumed no wiring capacitance). Figure 7.4.5 shows that even though the GP devices are drawn twice as large as the FinFETs to compensate for their double-gate nature, the FinFET delay is less sensitive to load. If the back-gate bias of the GP device is used to lower the V_t and decrease delay, leakage in these low-V_t devices must also increase. Figure 7.4.6 shows that the GP leakage must be increased by 60 times over that of the FinFET to achieve the same delay in the 3fF wire load case (absolute GP leakage in this case is 1.3μA/μm). Note that the plot also shows that output swing degrades as V_t is lowered. Even though FinFET lacks the tunable V_t of the GP device, large leakage must be tolerated if the GP is to have the same performance.

Acknowledgments:
This research is supported by the DARPA AME program under contract N66001-97-1-8910.

References:
[1] X. Huang et al, "Sub 50-nm FinFET: PMOS," 1999 IEDM Technical Digest.
[2] H.-S. P. Wong, et al., "Self-Aligned (Top and Bottom) Double-Gate MOSFET with a 25 nm Thick Silicon Channel," 1997 IEDM Technical Digest.
[3] D. Monroe and J. Hergenrother, "The Vertical Replacement-Gate (VRG) Process for Scalable General-Purpose Complementary Logic," ISSCC Digest of Technical Papers, Feb. 2000.
[4] C. P. Auth and J. D. Plummer, "Vertical, Fully-Depleted, Surrounding Gate MOSFETs on sub-0.1 mm Thick Silicon Pillars," 1996 DRC Technical Digest.
[5] D. Hisamoto et al., "Impact of the vertical SOI DELTA Structure on Planar Device Technology," IEEE Transactions on Electron Devices, vol. 38 no. 6, June 1991.
[6] MEDICI 4.1.0, Avant! Corp. and TMA, Inc., Fremont, CA.
[7] H.-S. P. Wong, rt al., "Device Design Considerations for Double-Gate, Ground-Plane, and Single-Gated Ultra-Thin SOI MOSFET's at the 25 nm Channel Length Generation," 1998 IEDM Technical Digest.
[8] I. Yang et al, "Back-Gated CMOS on SOIAS for Dynamic Threshold Voltage Control," IEEE Transactions on Electron Devices, vol. 44 no. 5, May 1997.
[9] Y.-K. Choi et al, "Ultrathin-Body SOI MOSFET for Deep-Sub-Tenth Micron Era," IEEE Electron Device Letters, vol. 21 no. 5, May 2000.

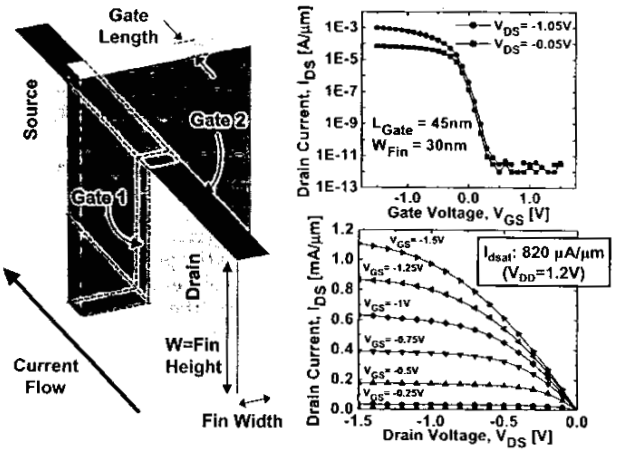


Figure 7.4.1: FinFET structure and experimental I-V curves [1].

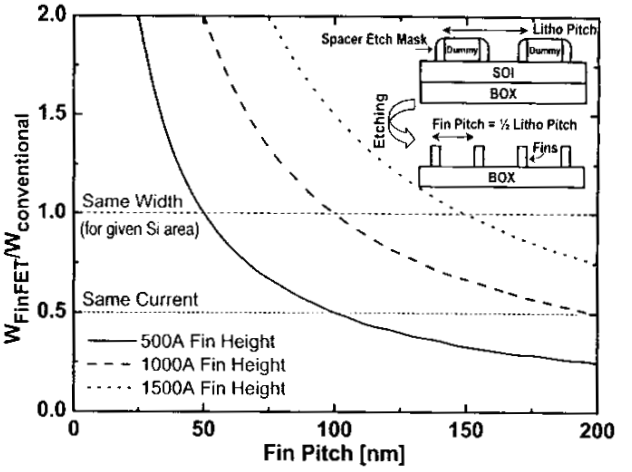


Figure 7.4.2: Width ratio for given silicon area vs. fin pitch.

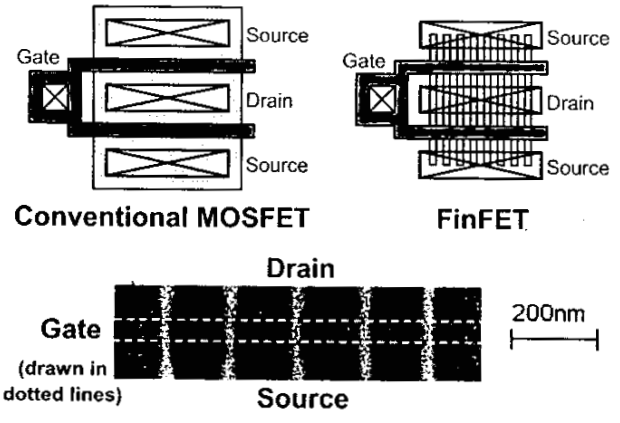


Figure 7.4.3: Transistor layout comparison and FinFET SEM.

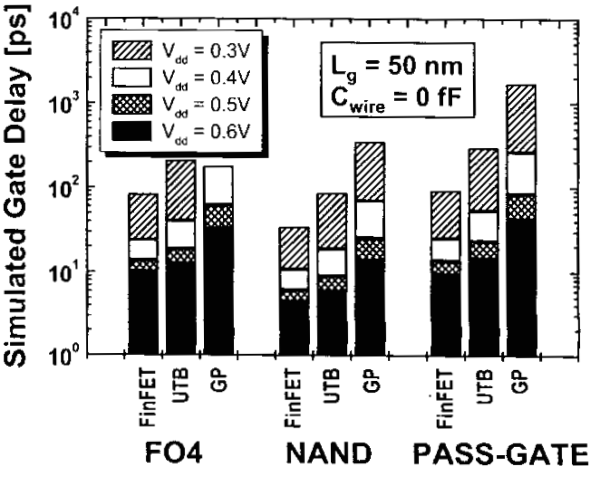


Figure 7.4.4: Comparison of unloaded benchmark circuit delay.

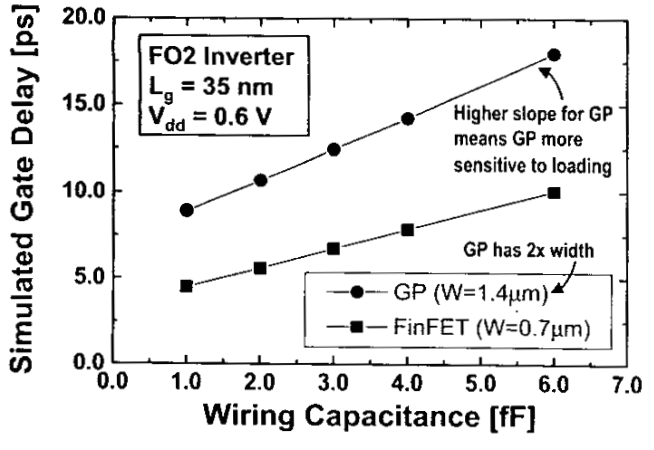


Figure 7.4.5: Delay of FO2 inverter with wiring capacitance added.

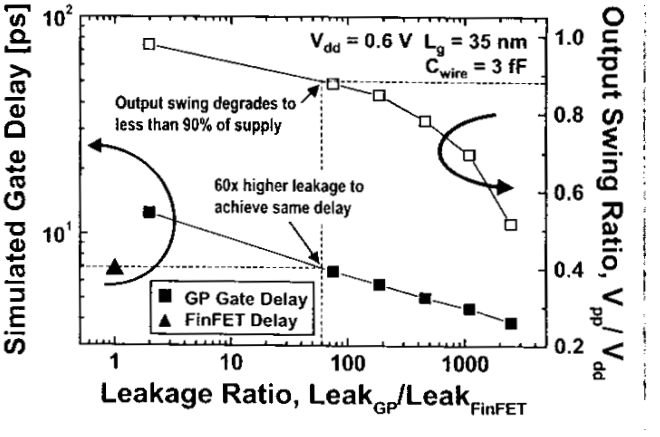


Figure 7.4.6: Delay and output swing as a function of leakage allowed.

Continued on Page 437

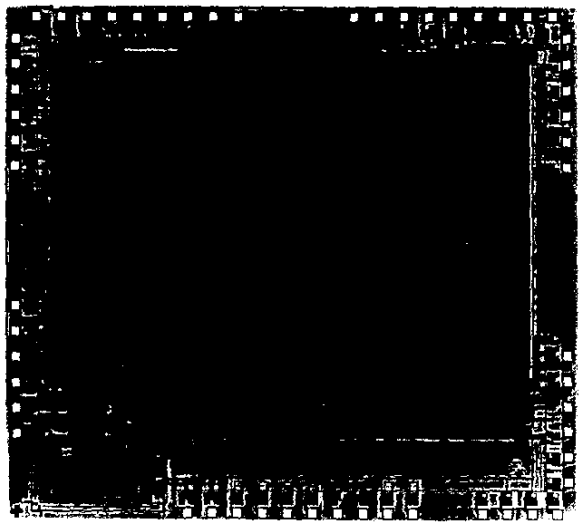


Figure 6.6.7: Micrograph of signal-processing CMOS image sensor.

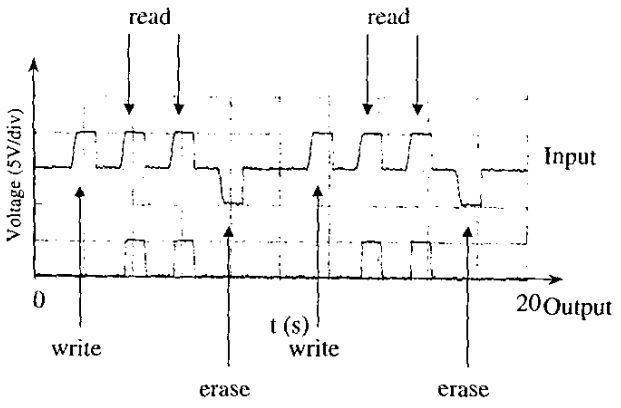
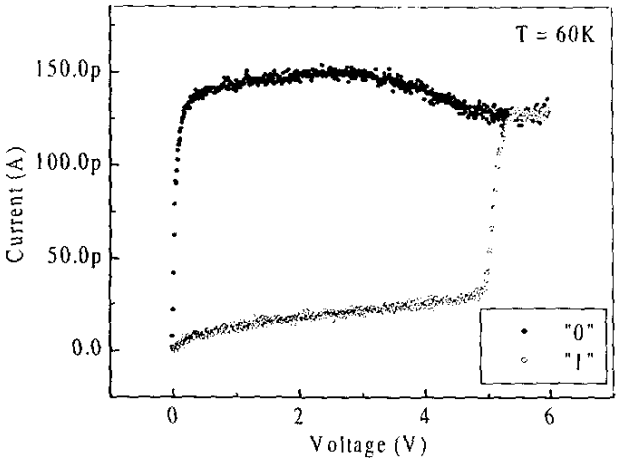


Figure 7.2.6. Timing diagram of a molecular random access memory.



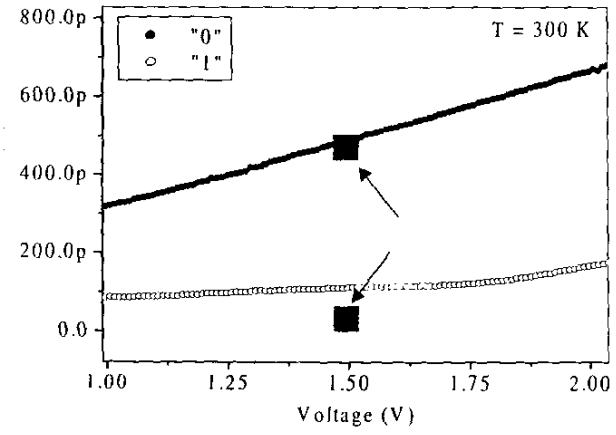


Figure 7.2.5. I(V) characteristics of stored and initial/erased states in Au-(2)-Au at (a) 60K and (b) ambient temperature (300K). The device area is ≤ 2×10^{-11} cm^2. Setpoints for the circuit of Figure 7.2.6.

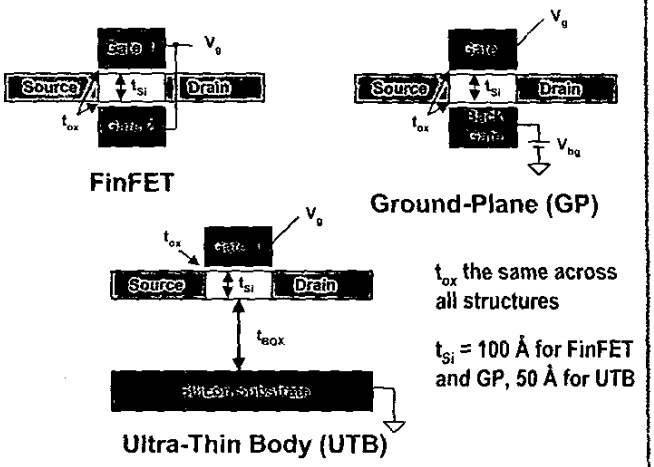


Figure 7.4.7: Advanced MOSFET cross-sections.

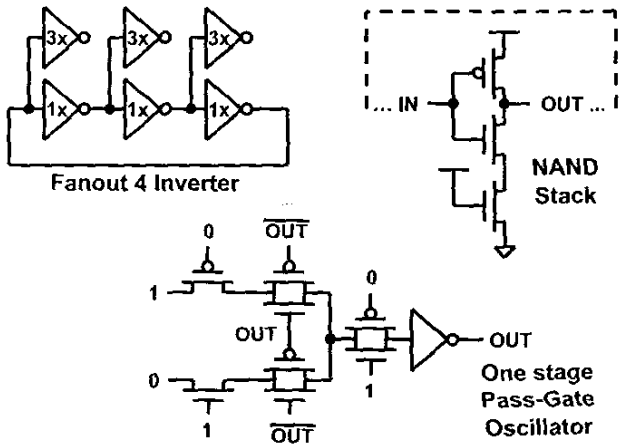


Figure 7.4.8: Schematics of benchmark ring oscillator circuits.