Nanoscale CMOS Circuit Leakage Power Reduction by Double-Gate Device

Keunwoo Kim, Koushik K. Das, Rajiv V. Joshi, Ching-Te Chuang
IBM T. J. Watson Research Center
Yorktown Heights, NY 10598
(914) 945-1336
{kkim, kkdas, rvjoshi, ctchuang}@us.ibm.com

ABSTRACT

Leakage power for extremely scaled (L_{eff} = 25 nm) double-gate devices is examined. Numerical two-dimensional simulation results for double-gate CMOS device/circuit power are presented from physics principle, identifying that double-gate technology is an ideal candidate for low-power applications. Unique double-gate device features resulting from gate-gate coupling are discussed and effectively exploited for optimal low-leakage device design. Design trade-offs for double-gate CMOS power and performance are suggested for low-power and high-performance applications. Total power consumptions of static and dynamic circuits and latches for double-gate device are analyzed considering state dependency, showing that leakage current is reduced by a factor of over 10X, compared with conventional bulk-Si counterpart.

Categories and Subject Descriptors

B.7.1 [**Hardware**]: Integrated Circuits–*Types and Design Style*

General Terms

Design, Theory, Verification

Keywords

Double-gate device, Leakage power, Short-channel effect

1. INTRODUCTION

Continuous CMOS scaling has been the main driving factor of silicon technology advancement to improve the performance. However, the sustained scaling process in the sub-nanometer regime (< 25 nm) requires ultimately scaling down of the gate-oxide thickness and increasing channel/body doping densities to overcome severe short-channel effects (SCEs) [1]. With a thinner gate oxide, more gate tunneling current occurs. For gate-oxide thickness below 1 nm, high-K dielectric materials with thicker layer than SiO_2 must be used to reduce gate leakage [2,3,4]. However, it significantly reduces carrier mobility, thereby degrading CMOS performance [5]. Circuit design techniques to mitigate the impact of gate leakage would be much less efficient than the use of high-K material since gate leakage is a stronger function of process-induced oxide thickness fluctuation as compared to change in V_{DD} and threshold voltage [4]. In addition to the gate-oxide scaling issue, higher doping concentrations would degrade subthreshold swing (S). Furthermore, V_{DD} scaling necessitates threshold voltage (V_t) reduction, which exponentially increase I_{off} [1]. I_{off} reduction is critical where chips are often in standby mode of operation, and even during active operation, acceptable I_{off} is required since leakage power consumption is rapidly increasing at a much faster rate compared to dynamic power. Leakage power is projected to overwhelm dynamic power in the next few technology generations [6].

ISLPED'04, August 9-11, 2004, Newport Beach, California, USA.

Double-gate (DG) MOSFET can potentially overcome this hurdle of leakage power by its unique property of electrical coupling of the two gates [7,8,9]. Furthermore, undoped or lightly-doped body and relatively thicker oxide could be employed to defer high-K material [1,7]. DG inverter delay and power with different back-gate oxide thickness have been studied without any comparison to conventional technologies [9]. In this paper, we present the comparative analysis of DG device/circuit power with respect to bulk Si. Superior DG device characteristics are analyzed with a significant reduction of CMOS circuit leakage. Feasible optimal 25 nm DG device design is presented. Our study/simulations are based on physically-calibrated model parameters in a two-dimensional (2-D) numerical simulator [10] vis-a-vis Monte Carlo simulation data [11,12]. Low-power DG circuit design points/trade-offs, as compared to bulk-Si technology, are analyzed by reducing leakage current while maintaining or even improving performance. Unique DG device stack effects in logic blocks and latches are studied for the first time. Optimum DG device/circuit design methodologies are suggested for low-power and high-performance applications.

2. LOW LEAKAGE 25 NM DG DEVICE

Figure 1 shows the schematic cross section of double-gate (DG) nFET. Electrically-coupled front and back gates effectively control SCEs, resulting in reduced drain-induced barrier lowing (DIBL) and improved S [13]. Due to this ideal device characteristics, DG technologies are most suitable for low-power circuits by reducing standby power even with an increase in performance.

For 25 nm device design, we first simulated a 30 nm symmetrical DG nMOSFET [11] using the 2-D numerical device simulator [10] to compare/calibrate the results with respect to (more rigorous, physical) Monte Carlo-predicted data [11]. The DG device has abrupt source/drain junction (i.e., the effective channel length is assumed same as the metallurgical channel length: $L_{eff} = L_{met}$ = 30 nm [3]). Since

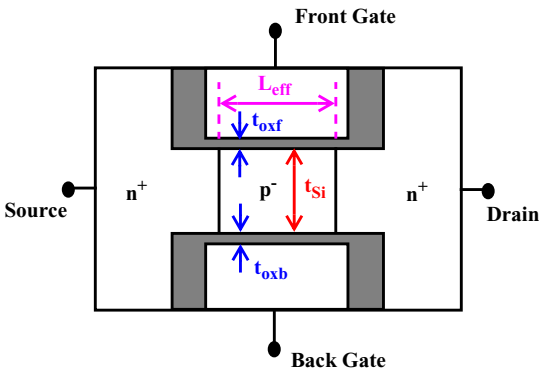

Figure 1. The double-gate nFET structure: L_{eff} = effective channel length, t_{oxf} = front-gate oxide thickness, t_{oxb} = back-gate oxide thickness, t_{Si} = Si-film thickness. For the symmetrical DG device with undoped or lightly-doped body, the gates must have near-mid-gap work functions for V_t control [7].

drift/diffusion transport model could underestimate the drain saturation current, hydrodynamic (HD) transport option is used, effectively considering quasi-ballistic transport [15] by adjusting energy relaxation time. The device structure, shown in Figure 1, is designed according to [11], with lightly doped ($N_A = 10^{15}$ cm^{-3}) thin (t_{Si} = 5 nm) Si-film and relatively thick ($t_{oxf} = t_{oxb}$ = 3 nm) front- and back-gate oxides. The gate material, or work function, is arbitrary here; we used a "mid-gap" gate material ($\Phi_M = X_{Si} + 0.5E_{g(Si)}$) [11]. In the 30 nm DG device with the HD model in MEDICI simulation, we estimate the energy relaxation time ($\tau_n = \tau_p$ = 0.56 ps) from the experimental values of energy relaxation length ($\lambda_{w(n)} = \lambda_{w(p)}$ = 65 nm) [16] and saturation velocity ($v_{sat(n)} = v_{sat(p)} = 7x10^6$ cm/s) [17]. Figure 2 shows MEDICI-simulated current-voltage characteristics for the 30 nm symmetrical DG nMOSFET, with comparison to Monte-Carlo predicted results [11]. The excellent agreement testifies for the physical parameter values used.

In order to enhance device performance by scaling, symmetrical DG nFET is re-designed with $L_{eff} = L_{met}$ = 25 nm and $t_{oxf} = t_{oxb}$ = 1.5 nm, based on the HD models extracted for the 30 nm DG nFET shown in Figure 2. Relatively thicker oxide for extremely short L_{eff} is used to mitigate gate-oxide tunneling current (I_{Gate}); it is observed that $I_{off} >> I_{Gate}$ in this study. Due to lightly-doped body in DG device, reverse/forward-bias source/drain-to-body junction band-to-band and trap-assisted tunneling currents are negligibly small [3]. Quantum-mechanical (QM) confinements of carriers, which significantly impact on 25 nm devices by increasing V_t and decreasing effective gate capacitance, are fully considered in the device design by solving one-dimensional (1-D) self-consistent Shrödinger-Poisson equations [18].

Figure 3 shows MEDICI-predicted I_{DS}-V_{GS} characteristics for 25 nm bulk-Si and DG nFETs. Both devices are made with equal drive current around $V_{DD} = V_{GS} = V_{DS}$ = 1 V and comparable V_t at V_{DS} = 1 V. Since DG device has two identical channel at the front and back surfaces, the sum of the device width of the two gates is assumed the same as the bulk-Si device width for fair comparison. DG MOSFET exhibits far superior device characteristics as compared to bulk-Si counterpart with much lower S (= 65 versus 90 mV) and 3X lower DIBL, defined as V_t difference between V_{DS} = 0.05 and 1 V, (= 35 versus 105 mV). These offer over a 10X reduction in I_{off}. Note that the model parameters in 2-D device simulations are physically calibrated against Monte Carlo-predicted data using full band structure of Si with consistently calculated scattering rate and Fermi-Dirac statistics [11,12]. Therefore, MEDICI-predicted results would be representative for device/circuit performance and power projections. The substantial reduction of either DIBL or S by bulk-Si device design would not be possible in the sub-nanoscale arena since the reduced DIBL worsens S due to the decreased depletion width and the improved S degrades DIBL due to the reduced channel-to-body coupling [14]. By lowering both S and DIBL, DG device would significantly reduce I_{off}.

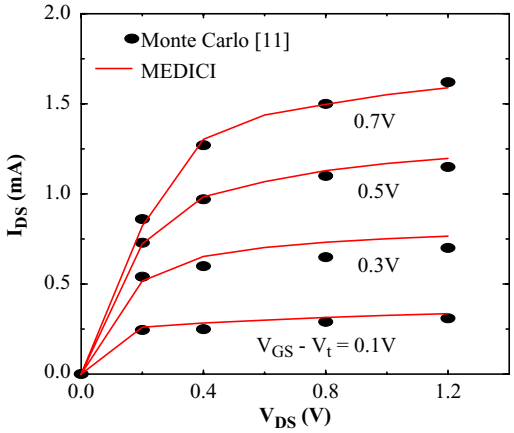

Figure 2. MEDICI-predicted I_{DS} versus V_{DS} for 30 nm symmetrical DG nFET with t_{Si} = 5 nm, $N_A = 10^{15}$ cm^{-3}, and $t_{oxf} = t_{oxb}$ = 3 nm, confirmed with Monte Carlo-simulated data.

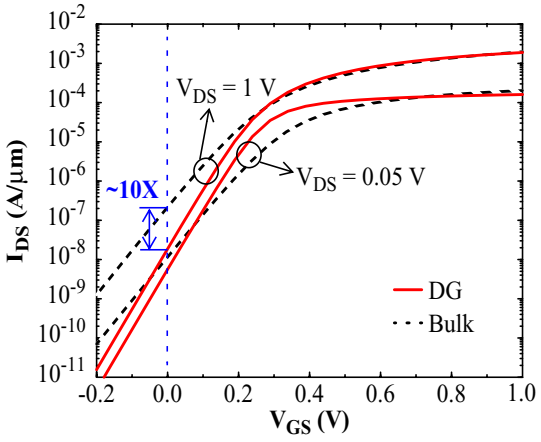

Figure 3. MEDICI-predicted I_{DS}-V_{GS} characteristics at V_{DS} = 0.05, 1 V for 25 nm bulk-Si and DG nFETs. DG nFET is re-designed with L_{eff} = 25 nm, $t_{oxf} = t_{oxb}$ = 1.5 nm, t_{Si} = 5 nm, and $N_A = 10^{15}$ cm^{-15}.

3. POWER ANALYSIS OF DOUBLE CMOS CIRCUITS

We analyze in detail leakage power consumptions for inverter, typical static and dynamic CMOS circuits, and latches with 25 nm double-gate (DG) devices, compared with bulk-Si devices. Circuit leakage dependency on input pattern is also discussed.

3.1 CMOS Inverter

Figure 4 shows MEDICI-predicted unloaded inverter delay versus I_{off} with varying V_{DD} (= 0.8, 0.9, 1.0, 1.1 V) for 25 nm bulk-Si and DG devices. For equal delay at V_{DD} = 1.0 V, I_{off} for DG is noticeably reduced by ~12X. Both delay and I_{off} are less sensitive to V_{DD} variation for DG device due to less DIBL. Figure 5 shows inverter delay versus V_{DD} for bulk-Si and DG devices designed for equal I_{off} at V_{DD} = 1 V by adjusting V_t, compared with DG devices designed for equal delay with bulk Si at V_{DD} = 1 V. Since V_t for DG device in the equal I_{off} is lower due to lower S, DG CMOS is faster, and the speed improvement is larger for lower V_{DD} due to the more (relatively) enhanced I_{on} by less DIBL. For V_{DD} = 1 V, DG is faster by ~15%. Note that V_{DD} for DG could be reduced by ~0.15-0.22 V for the equal CMOS performance shown in Figure 5, which could significantly reduce dynamic power consumption by ~28-54%. Since DIBL is 3X lower in the DG device, $V_{t(lin)}$ (V_t defined at V_{DS} = 0.05 V) is lower for equal I_{off}. Lower $V_{t(lin)}$ would improve performance and noise margin for the circuits such as pass gate, stack devices, and SRAM read/write pass transistor where devices spend substantial amount of time in linear operation during switching transient [19].

Figure 6 shows MEDICI-predicted I_{off} at V_{DD} = 1 V versus L_{eff} for bulk-Si and DG CMOS inverters, which have the equal delay for L_{eff} = 25 nm. I_{off} reduction is dramatically increased as L_{eff} is scaled due to much less SCEs. As L_{eff} is scaled from 25 nm to 19 nm, I_{off} increases by only ~20X for DG, as opposed to ~250X for bulk Si. Due to a lesser increase of I_{off} for shorter L_{eff}, DG devices will be much more scalable. Since 2-D SCEs are numerically calculated based on 2-D potential/charge distributions for L_{eff} variations, the predicted results are quite representative. Figure 7 shows I_{off} versus temperature (T) for bulk-Si and DG inverters. I_{off} is lower in DG device by ~12-26X, and as temperature increases, the difference becomes more prominent. It is to be noted that I_{off} for DG CMOS is much less sensitive to temperature variation. Since the depletion width in the DG device is independent of temperature due to the fully-depleted (ultra-thin) body (5 nm) [20,21], V_t of DG device is much less reduced with respect to the increase of temperature, which leads to a further relative reduction of I_{off} at higher temperature.

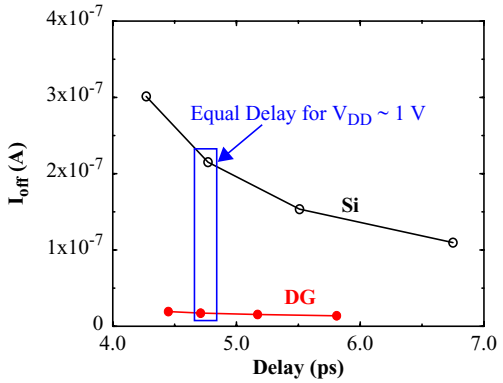

Figure 4. Delay of unloaded inverters versus I_{off} with varying V_{DD} = 0.8, 0.9, 1.0, 1.1 V for 25 nm bulk-Si and DG devices.

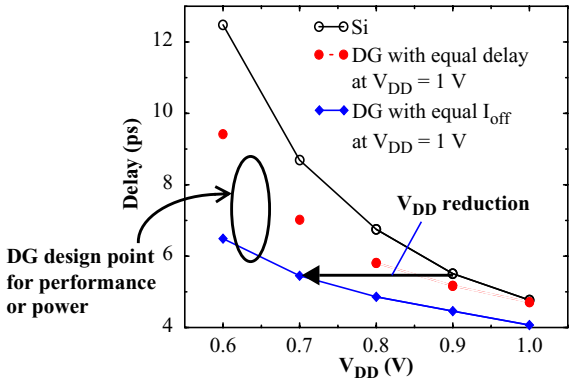

Figure 5. Inverter delay versus V_{DD} for bulk-Si and DG devices designed for equal I_{off} or equal delay at V_{DD} = 1 V. DG CMOS design trade-off for performance and power is indicated.

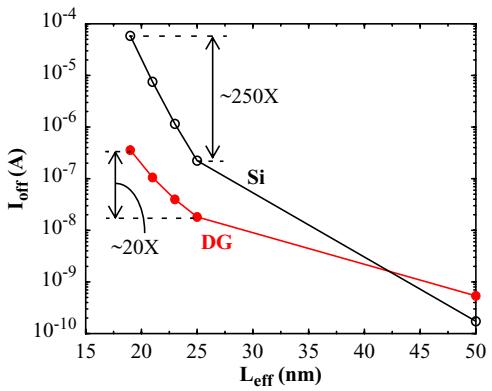

Figure 6. I_{off} versus L_{eff} at V_{DD} = 1 V for bulk-Si and DG inverters. The delay is equal for L_{eff} = 25 nm.

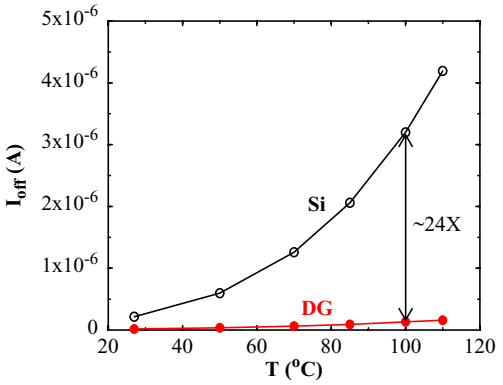

Figure 7. I_{off} versus temperature (27-110°C) at V_{DD} = 1 V for bulk-Si and DG inverters. The delay is equal for T = 27°C.

3.2 Static NAND

One of efficient ways to reduce I_{off} in CMOS circuit is to stack/turn off the devices rather than to use a single off device [22,23]. Figure 8 shows the leakage current characteristics for 2-way NAND with bulk-Si and DG devices for input patterns (AB) = (11), (01), (10), (00). Both bulk-Si and DG circuits are designed with near-equal delay at the critical path by adjusting V_t of DG devices [7]. V_t for DG pFETs could be further increased since pFETs have less impact on (worst) delay estimation, thereby much reducing I_{off} by ~65X for (00), i. e., when two pFETs are off. Due to less DIBL and S, the leakage currents in stack devices are significantly lower in DG by over 9X for all input patterns. I_{off} for DG is almost same in (01) and (10) due to reduced DIBL, but I_{off} for bulk-Si case is varied by ~3X between (01) and (10). Because of less DIBL, I_{off} for DG 2-way NAND is less sensitive to input patterns; the difference between the best and worst case is ~8X for DG and ~55X for bulk Si. This makes DG CMOS very suitable for nanoscale circuit design by making the standby power very insensitive to the input vector patterns. The uncertainty of power estimation, done in the early phase of the design cycle in stack gates, will be substantially reduced. Note that the same leakage current flows through the stacked nFETs since gate tunneling current is much smaller due to relatively thicker oxide. Note also, in stacked devices, that reverse-body effects will not occur in DG device since the body is floating, which could reduce I_{off} but would substantially degrade performance [1,3].

Figure 9 shows the leakage current characteristics for 3-way NAND with bulk-Si and DG devices for (ABC) = (111), (011), (101), (110), (001), (010), (100), (000). Note that both bulk-Si and DG devices have near-equal performance in the critical path. The leakage currents are significantly lower in DG by over ~7X for all input patterns. I_{off} for DG is almost same in (011), (101), and (110) where only one nFET is off due to reduced DIBL, but I_{off} in bulk-Si case is varied by ~3X. I_{off} difference between the best and worst case is ~20X for DG and ~175X for bulk Si. Therefore, as stacking of gates increases, leakage power variation for input patterns does not increase as much in DG as it does for bulk Si. Based on the study for the dependence of leakage current on the stack height, leakage power optimization by stacking will be relatively less efficient for DG technology. The similar comparative results for bulk Si and DG stacked pFETs are observed in NOR circuits; however, the results are omitted for brevity

Figure 10 shows clock frequency versus total power for bulk-Si and DG 2-way NAND circuits for two different switching activities, 1% and 10%, at V_{DD} = 1 V and T = 27°C. For the time interval when the circuit is not switching, its input pattern is assumed to be the worst case (11) combination. For activity = 1%, total power is reduced by ~35-1.2X for the clock frequency range (f = 3.9 MHz-1.67 GHz). For activity = 10%, total power is reduced by ~8-1X for f = 3.9 MHz-1.67 GHz. For lower activity, total power reduction by DG is increased due to the higher impact of leakage power on total power consumption. Although not shown for the brevity, DG

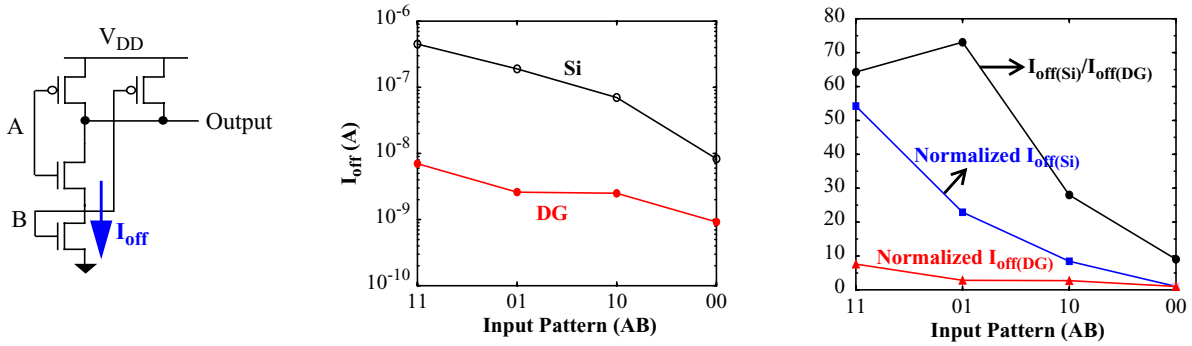

Figure 8. I_{off} characteristics for 2-way NAND with bulk-Si and DG devices at V_{DD} = 1 V for (AB) = (11), (01), (10), (00). Normalized I_{off} is defined as I_{off} for (AB) divided by I_{off} for (00), i. e., I_{off} ratio for each input pattern with respect to the lowest I_{off}. Both bulk-Si and DG devices have equal performance when the two inputs are switching from 0 to V_{DD}.

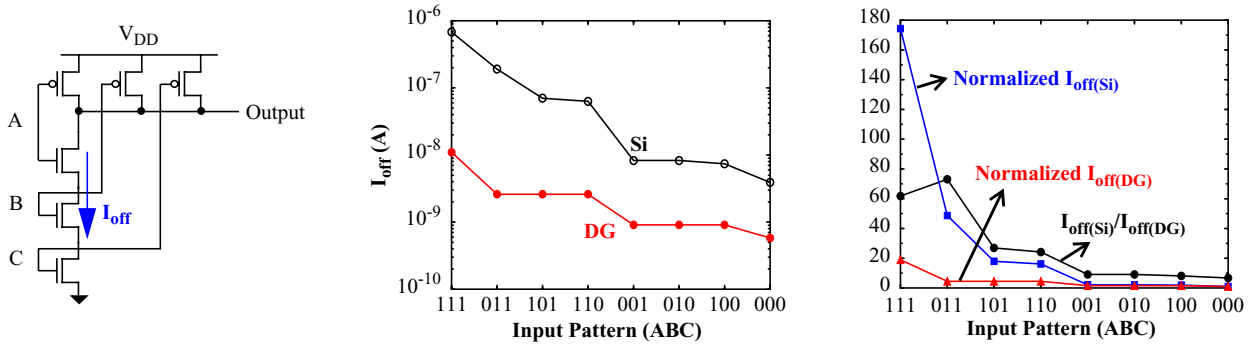

Figure 9. I_{off} characteristics for 2-way NAND with bulk-Si and DG devices at V_{DD} = 1 V for (ABC) = (111), (011), (101), (110), (001), (010), (100), (000). Normalized I_{off} is defined as I_{off} for (ABC) divided by I_{off} for (000), i. e., I_{off} ratio for each input pattern with respect to the lowest I_{off}. Both bulk-Si and DG devices have equal performance when the three inputs are switching from 0 to V_{DD}.

technology would offer the significant advantage for other standby input patterns. Figure 11 shows normalized total power for (11) with respect to (00) in 2-way NAND versus activity of circuit for clock frequency (f = 62.5 MHz, 1.67 GHz) at V_{DD} = 1 V and T = 27°C with bulk-Si and DG devices. (11) and (00) refer to the worst and base case of standby power, respectively. Relative contribution of leakage on total power is very little in DG even when activity and clock frequency are reduced; normalized total power will stay close to unity in the figure. Therefore, the DG circuit exhibits state-independent total power consumption.

Figure 12 shows clock frequency versus total power for bulk-Si and DG 2-way NAND circuits for worst case of leakage power at V_{DD} = 1 V, T = 100°C, and activity = 1%, 10%. For activity = 1%, total power is reduced by ~115-4X for f = 3.9 MHz-1.67 GHz. For activity = 10%, total power is reduced by ~60-1.3X for f = 3.9 MHz-1.67 GHz. As temperature increases, DG circuit power does not increase as much as it does for bulk Si, due to less variation of I_{off} as predicted in Figure 7. This also makes DG device an ideal candidate for a variety of low-power battery applications with long standby intervals. Although not shown for the brevity, state-dependency of circuit power to activity is also less dependent in DG than bulk Si for T = 100°C.

3.3 Dynamic Circuits

Dynamic circuits are used for improving circuit performance. However, they suffer from severe leakage current and noise problems, especially at nanoscale technologies. To resolve this, DG CMOS could be resorted. Figure 13 shows schematic of Domino circuit and leakage power consumption versus V_{DD} for bulk-Si and DG devices designed with comparable circuit performance. The leakage power ranges are shown as the error bar due to the different patterns for inputs (AB). DG consumes ~10-12X lower leakage power for V_{DD} = 0.8-1.0 V. The DG Domino circuit leakage power is insensitive to the input patterns due to less DIBL. It is to be noted that DG Domino circuit power even for the worst case is much less than the best case of bulk-Si counterpart. Due to this noticeable reduction of leakage power, DG technology would be more effectively used in dynamic circuits.

3.4 Latches

Latches are very important circuit elements due to their storage ability, and power consumptions in latches form a significant portion of total chip power. We studied leakage power for pulse latch as an example. Figure 14 shows schematic of pulse latch and leakage power consumption versus V_{DD} for bulk-Si and DG devices designed with comparable circuit performance. The leakage power ranges are shown as the error bar due to the variations for A and Clk. DG has ~8-12X lower standby power for V_{DD} = 0.8-1.0 V. The DG latch leakage power is insensitive to the input patterns due to less DIBL. Note that DG latch leakage power even for the worst case is much less than the best case of Si counterpart.

Although not shown for the brevity, it is observed that 25 nm DG circuit performance of static NAND, Domino, and pulse latch relative to bulk Si is enhanced by more than 15% for the equal I_{off} around 200 nA/μm at V_{DD} = 1 V.

4. CONCLUSIONS

Physics-based numerical simulation-aided study with optimal 25 nm DG device, calibrated against Monte Carlo-predicted

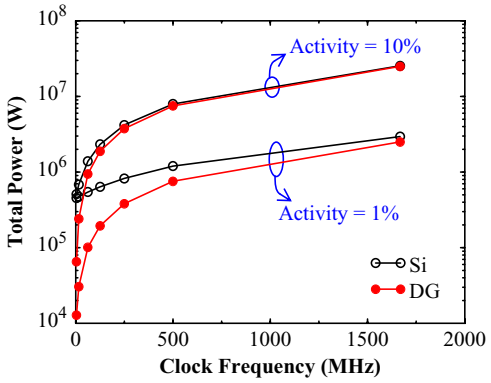

Figure 10. MEDICI-predicted clock frequency versus total power for bulk-Si and DG 2-way NAND circuits at V_{DD} = 1 V and T = 27°C.

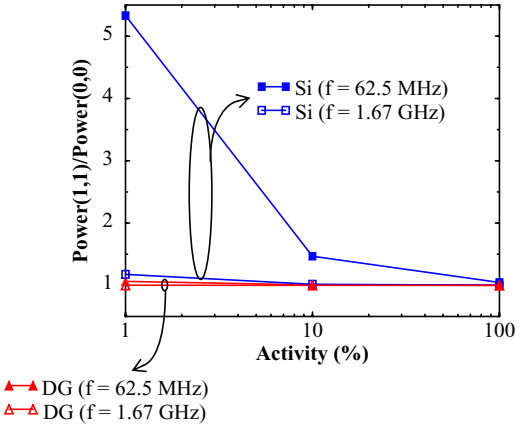

Figure 11. MEDICI-predicted normalized total power for (11) with respect to (00) in 2-way NAND versus activity of circuit for clock frequency (f = 62.5 MHz, 1.67 GHz) at V_{DD} = 1 V and T = 27 °C for bulk-Si and DG devices. (11) and (00) refer to the standby state when the circuit is not switching.

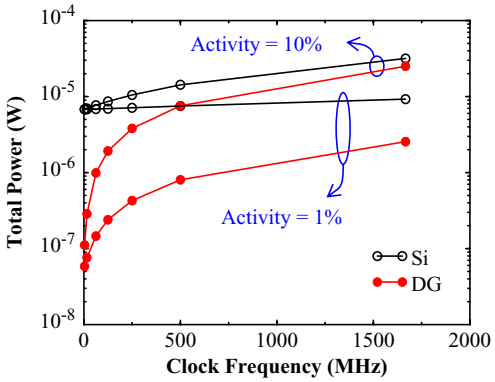

Figure 12. MEDICI-predicted clock frequency versus total power for bulk-Si and DG 2-way NAND circuits at V_{DD} = 1 V and T = 100°C.

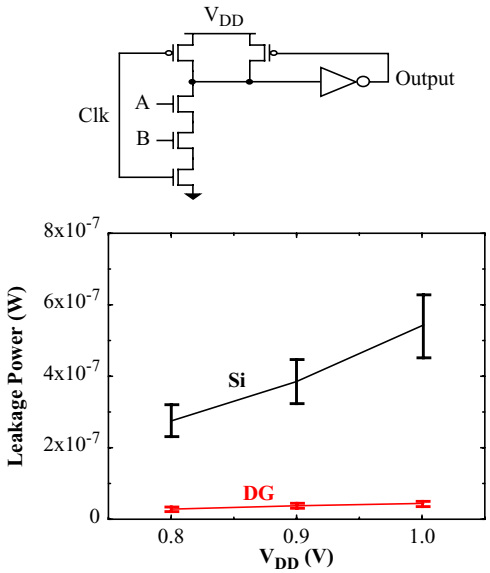

Figure 13. Schematics and leakage power versus V_{DD} for DG and bulk-Si Domino circuits in comparable performance. The error bar shows the sensitivity to input patterns and Clk.

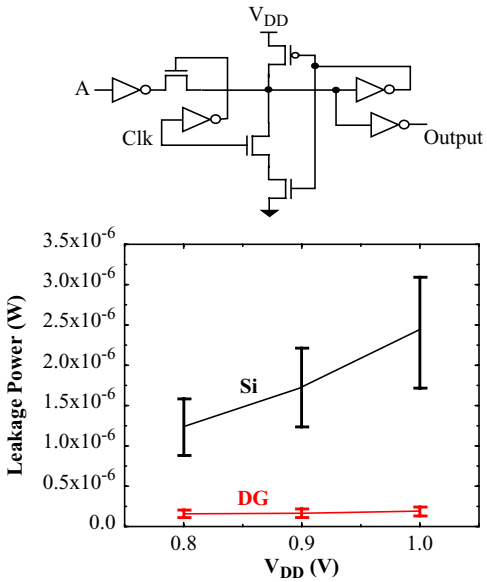

Figure 14. Schematics and leakage power versus V_{DD} for DG and bulk-Si pulse latches in comparable performance. The error bar shows the sensitivity to variations of A and Clk.

data, is presented, considering quasi-ballistic carrier transport and quantum-mechanical effects. In ultimately scaled technology, CMOS circuit leakage power would be significantly reduced by DG devices. Considering power and performance trade-off in circuit design, the 25 nm DG inverter could offer ~10X lower leakage power or 15% faster performance due to near-ideal S and 3X lower DIBL. Due to a lesser I_{off} increase for L_{eff} reduction, DG device will be much more scalable. For higher temperature and lower activity, power saving is increased with DG technology. For T = 100^{o}C and activity = 1%, the total power in static 2-way NAND is reduced by ~115-4X for clock frequency range (3.9 MHz-1.67 GHz). It is also observed that 25 nm DG CMOS technology would offer much lower leakage power by ~10X for dynamic circuits and latches, compared with the bulk-Si counterpart technology. We studied and analyzed the leakage dependence on input patterns in detail for all kinds of circuits, concluding that DG is very state-insensitive, which would make sub-nanoscale CMOS circuit design more flexible.

5. REFERENCES

[1] H.-S. P. Wong, et al., *Proc. IEEE*, pp. 537-570, April 1999.

[2] H. Iwai, et al., *Tech Digest IEDM*, pp. 625-628, 2002.

[3] Y. Taur and T. H. Ning, *NY: Cambridge Univ. Press*, 1998.

[4] Y. Nakagome, et al., *IBM J. Res. Devel.*, pp. 525-552, Nov., 2003.

[5] M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, *J. Appl. Phys.*, pp. 4587-4608, 2001.

[6] International Technology Roadmap for Semiconductor, 2003 (http://public.itrs.net).

[7] K. Kim and J. G. Fossum, *IEEE Trans.*, pp. 294-299, Feb. 2001.

[8] S. H. Tang, et al., pp. 118-119, *ISSCC* 2001.

[9] R. Zhang, K. Roy, and D. B. Janes, pp. 213-218, *ISLPED* 2001.

[10] Taurus-Medici: Industry-standard simulation tool. Mountain View, CA: Synopsis, Inc., 2003.

[11] D. J. Frank, S. E. Laux, and M. V. Fischetti, *Tech Digest IEDM*, pp. 553-556, 1992.

[12] S. E. Laux, M. V. Fischetti, and D. J. Frank, *IBM J. Res. Devel.*, pp.466-494, 1996.

[13] K. Kim, J. G. Fossum, and C. T. Chuang, pp. 223-226, *IEEE Int. Conf. SISPAD* 2003.

[14] E. J. Norwak, et al., *IEEE Circuits Devices Mag.*, pp. 20-31, Jan/Feb. 2004.

[15] M. Lundstrom, *IEEE Electron Device Lett.*, pp. 361-363, July 1997.

[16] J. W. Slotboom, et al., *Tech Digest IEDM*, pp. 127-130, 1991.

[17] Y. Taur, et al., *Solid-State Electron.*, pp. 1085-1087, 1993.

[18] D. Vasileska and Z. Ren, Arizona State Univ., Tempe, AZ, and Purdue Univ., West Lafayette, IN, Feb. 2000.

[19] C. T. Chuang, et al., *IEEE Circuits Devices Mag.*, pp. 6-19, Jan/Feb. 2004.

[20] G. Groeseneken, et. al., *IEEE Electron Device Lett.*, pp. 329-331, Aug. 1990.

[21] D. Jeon and D. E. Burk, *IEEE Trans.*, pp. 2101-2111, Sep. 1991.

[22] S. Narendra, et al., *ISLPED*, pp. 195-200, 2001.

[23] M. C. Johnson, et al., *IEEE Trans. VLSI Syst.*, pp. 1-5, Feb. 2002.