RC23480 (W0501-018) January 6, 2005
Electrical Engineering

IBM Research Report

A Gate Leakage Reduction Technique for CMOS Driver Circuits

J. B. Kuang, Hung C. Ngo, Kevin J. Nowka
IBM Research Division
Austin Research Laboratory
11501 Burnet Road
Austin, TX 78758

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

A Gate Leakage Reduction Technique for CMOS Driver Circuits

J.B. Kuang, Hung C. Ngo, and Kevin J. Nowka

IBM Austin Research Laboratory
Austin, TX 78758

Abstract

We present a low-overhead circuit technique that mitigates gate leakage for sub-90nm technologies. In a circuit path, the effective driver strength is retained but divided into always-on and gate-stress-relieved branches. During stand-by or after switching in the active mode, large devices are turned off or placed in a floating state that collapses over time, thus reducing the gate current. Total leakage reduction up to 54% can be observed for typical driver circuits.

Keywords: low power, gate leakage, CMOS, digital circuit.

Introduction

As the gate oxide thickness continues to scale down with technology migration, tunneling currents becomes a noticeable concern for VLSI designers. Gate leakage power can potentially be the dominant component at stand-by in the sub-90nm technologies that use conventional gate insulator materials [1-3]. Therefore, controlling this leakage component is very desirable for stand-by power management and leakage-induced yield loss prevention. Circuit level solutions, such as [4] and [5], that timely modify drive strength and turn off certain devices have been proposed. In this paper, we present a general circuit technique that retains the attractive features of these solutions while minimizing device area overhead and active power penalty at the same time.

To illustrate this circuit method, hardware calibrated [6-7] 90nm PD/SOI technology device models with 1.05nm gate oxide thickness are used in the analyses. High Vt devices are selected for the driver designs in this paper to minimize the leakage contribution from the subthreshold component as well as to better characterize the effects of gate currents. Driver/buffer style circuits are identified because of their substantial sizes and contributions to chip leakage.

Circuit Topology

The schematic diagram of the base topology for a two-stage low gate leakage driver circuit is shown in Fig. 1 where the total output device drive strength is large. The two middle inverters are always kept on and intentionally sized small. Most of the circuit drive strength is allocated to the large output PFET and NFET whose gate terminals (internal nodes H and L) can be conditionally placed in a collapsing floating state, i.e., decaying from the initial ground and vdd potential, respectively, to an intermediate voltage level. At the first stage (Stage i), an NFET footer is inserted to the inverter driving node H and a PFET header is inserted to the inverter driving node L. Table 1 describes the state diagram for gate stress relief with respect to the two control signals. When full gate stress relief is asserted, control signal "cut_n" is set to low and "cut_p" is set to high, rendering H or L conditionally floating from the initial ground or vdd potential to a collapsed voltage level. The voltage stress across the gate oxide becomes significantly decreased, thus reducing the gate current. When nodes H or L is not in the floating state, and when the pull-up or pull-down leg that is not gated by the sleep device in Stage i is on, the large PFET or NFET in the second stage (Stage i+1) is completely turned off, which also results in low gate leakage. This circuit topology incurs small device width overhead; while the output-stage drive strength is kept constant by device width reassignment, the input-stage device width penalty is small due to the use of smaller devices.

To improve circuit robustness, floating nodes should be physically protected from noise interference. In layout, nodes H and L are shielded by the adjacent vdd and ground wires. Fig. 2 shows a circuit sizing example where the always-on (middle) portion of the circuit is set at 1/20th of the total output drive strength. The choice of this ratio depends on the specific operating condition and noise immunity requirements. Circuit timing is matched to a logically equivalent two inverter driver at a stage gain of 4 and identical output load. Fig. 3 compares the rising switching transients of the conventional and low-gate-leakage driver. Clean voltage characteristics are observed for all internal nodes.

For nodes favoring a skewed voltage transition, gate oxide stress relief can be applied preferentially to only one pull-up or pull-down leg. Fig. 4 shows a three stage circuit path that is composed of two skewed two-stage circuit blocks, A and B, connected sequentially with one shared inverter stage. During power-down periods, the two floating nodes are benign because the always-on path preserves the correct logic state. More complex circuits can be topologically arranged in a similar fashion.

Implementation Examples

Control inputs "cut_n" and "cut_p" can be locally generated or globally provided. When they are globally distributed signals, timing granularity should be coarse, i.e., infrequent swapping between active and stand-by mode, such that both control and distribution circuit remain simple and small. When dynamic clock gating is also employed for the design, static leakage control signals can be derived from the clock gating signal, and be assigned with slower slews to minimize area as well as control circuit power overhead. Leakage power control is generally chosen to be less timing critical than active power control.

Fig. 5 shows a circuit example of automatically engaging the low gate leakage mode by using 3 serially connected minimum-sized feedback inverters in the two-stage driver, where the output state transition triggers the turning off of header or footer device at the first stage. The self-generated local control scheme to relieve gate stress allows efficient and flexible employment of this circuit in the driver/buffer applications. Fig. 6 depicts the switching transients of the circuit that is calibrated to have the same delay as the conventional two stage inverter driver at the same load. The upper graph in Fig. 6 shows the transmission of feedback control signals. The lower graph in Fig. 6 shows the three middle nodes H, M, and L. Table 2 compares the total leakage power of this circuit with the conventional driver. A substantial leakage improvement is observed at all corners, and most prominently at corners with lower subthreshold leakage current.

The local control circuitry for automated leakage reduction can be shared to minimize design overhead. Fig. 7 is an example of sharing the feedback inverters between a skewed low driver A and skewed high driver B serially connected with one common stage. The control circuit size is kept at minimum because the feedback timing is not speed critical. By using the circuit of Fig. 8, leakage saving characteristics can be independently quantized for the large output PFET and NFET. The gate voltages of the output NFET and PFET can be independently controlled using two external enable pins, en_low_pleak and en_low_nleak.

Considering device sizing, it is advantageous to generalize the topology to deeper stage length such that area and active power overhead can be further minimized. Fig. 9 shows the schematic diagram of a three-stage driver. Nodes M1 and M2 are in the always-on path. Nodes H1 or L1 is conditionally precharged to ground or vdd by a small NFET or PFET keeper in a non-timing critical fashion. Nodes H2 or L2 is conditionally placed at floating low or high, which eventually brings the circuit to a low gate leakage state. Table 3 compares total (subthreshold and gate) leakage for the design of Fig. 9 and a conventional three-stage inverter driver (stage gain set to 4) with equal timing, output drive strength, and output load under three corner conditions. Significant leakage reduction is observed at all corners. Also, more leakage saving is noted in the three-stage than in the two-stage design because the aggregated device width overhead is lower. Table 4 shows the active power comparison for two switching activity factors (SF's). A high SF of 50% is characteristic of clock circuits while an SF of 10% is typical in average static CMOS circuits. While some configurations may see a slight increase in instantaneous active power, the availability of various power-down modes in most high-performance low-power systems effectively makes the average active power lower.

Fig. 10 shows an alternative three-stage design where, instead of the use of collapsed floating nodes, wide leaky gates are turned off after state transition to reduce gate leakage. The devices for the feedback inputs to NAND and NOR at the first stage are skewed small to conserve device width and to have slower transitions whereas the forward path is optimized for speed. This option exhibits slightly higher active and leakage power than the collapsed floating node method but is more immune to noise because the floating nodes are removed.

Supply gating is an effective means that reduces the stand-by leakage in low-power designs [8]. Gate stress relief can be used in conjunction with supply gating applications. To ensure adequate performance, sleep transistors may take up a sizable device area. During power on, charging of the entire circuit block requires a larger amount of current than in normal operations. When virtual rails are established, the required current for circuit operation becomes lower. This difference in power supply current is particularly significant for SRAM and register file style circuits. To minimize gate leakage of sleep transistors during normal operations, circuits shown in Figs 11 and 12 can be adopted. Figs 11 and 12 are the example of footer and header style application, respectively. After the completion of virtual rail charging (i.e., block power on), a portion of the gate terminals of the header or footer devices are placed at collapsing floating state (Fig. 11) or turned off strongly (Fig. 12), thus reducing the gate leakage of sleep transistors.

Conclusion

A gate leakage mitigation circuit method is proposed for sub-90nm technologies with conventional gate oxide materials. Sources of large amount of gate currents can be effectively suppressed. Judicious use of this technique can result in overall leakage power reduction and wafer yield enhancement.

Acknowledgments

The authors thank Lorraine Herger and Michael Rosenfield for management support. This work was supported in part under Defense Advanced Research Project Agency Contract NBCH30390004.

References

[1] Semiconductor Industry Assoc, International Technology Roadmap for Semiconductors, 2002 Update.

[2] N.S. Kim, et al., "Leakage current: Moore's law meets static power," IEEE Computer, 2003, pp. 68-75.

[3] A. Marras, I. De Munari, D. Vescovi, and P. Ciampolini, "Effects of gate-currents on CMOS circuit behaviour," ESSSC, 2003, pp. 577-580.

[4] K. Nii, et al., "A 90-nm low-power 32-KB embedded SRAM with gate leakage suppression circuit for mobile applications," IEEE J. Solid-State Circuits, 2004, pp. 684-693.

[5] K. Bernstein, P.E. Cottrell, E.J. Nowak, N.J. Rohrer, and D.W. Stout, "Transient gate tunneling current control," US Patent 6577178, Jun 10, 2003.

[6] D. Boerstler, et al., "A 10+ GHz low jitter wide band PLL in 90 nm PD SOI CMOS technology," Dig. Tech. Papers, Symp. VLSI Circuits, 2004, pp. 228-231.

[7] D. Pham, et al., "The design and implementation of a first-generation CELL processor," Dig. Tech. Papers, ISSCC, 2005.

[8] S. Mutoh, S. Shigematsu, Y. Matsuya, H. Fukuda, T. Kaneko, and J Yamada, "A 1-V multithreshold-voltage CMOS digital signal processor for mobile phone application," IEEE J. Solid-State Circuits, 1996, pp. 1795-1802.

Table 1 Control state diagram for the circuit of Fig. 1

cut_n=0; cut_p=1	
In=1	PU=float
	PD=0
In=0	PU=1
	PD=float
cut_n=1; cut_p=1	
Gate stress relief for output pull-down	
cut_n=0; cut_p=0	
Gate stress relief for output pull-up	
cut_n=1; cut_p=0	
No gate stress relief applied	

Table 2 Leakage power for the driver of Fig. 5 and a conventional 2-stage buffer of equal timing and output drive strength

	Input State	Leakage Power (μW) This Work	Conventional Driver
Corner 1	0	11.2	13.7
	1	8.3	12.0
Corner 2	0	2.7	3.8
	1	2.0	3.3
Corner 3	0	1.5	2.3
	1	1.1	2.0

Corner 1: vdd=1.2V, T=25C, and fast process condition
Corner 2: vdd=1.0V, T=85C, and nominal process condition
Corner 3: vdd=0.9V, T=85C, and nominal process condition

Table 3 Leakage power for the driver of Figs 9 and a conventional 3-stage buffer of equal timing and output drive strength

	Input State	Leakage Power (μW) This Work	Conventional Driver
Corner 1	0	7.2	12.0
	1	9.4	14.4
Corner 2	0	1.9	3.5
	1	2.1	4.0
Corner 3	0	1.0	2.1
	1	1.1	2.4

Corner conditions: same as in Table 2

Table 4 Active power and device overhead for the driver of Fig. 5 and a conventional 3-stage buffer of equal timing and output drive strength

OVERHEAD	This Work	Conventional Driver
	Active Power (mW)	
50% SF	0.361 (1.2x)	0.300
10% SF	0.064 (1.1x)	0.058
	Normalized Total Device Width	
PFETs	234w (1.11x)	210w
NFETs	119w (1.13x)	105w

Active power measured at Corner 2 of Table 2.
50% SF (switching activity factor) = clock at 2.5GHz
10% SF = mimicking typical datapath circuit behavior

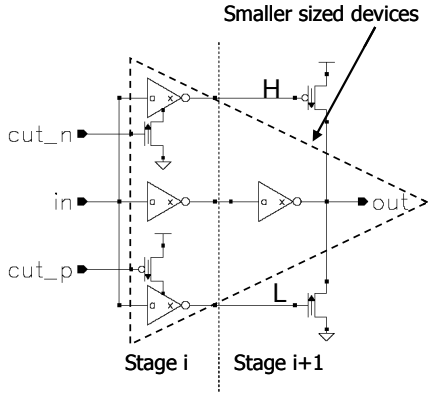

Fig. 1 Base topology of the low gate leakage driver circuit

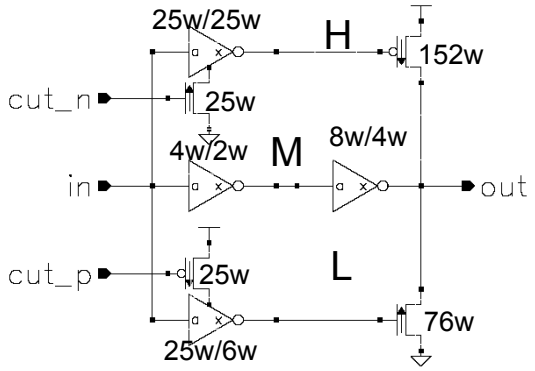

Fig. 2 A circuit sizing example with gate widths expressed in units of variable w.

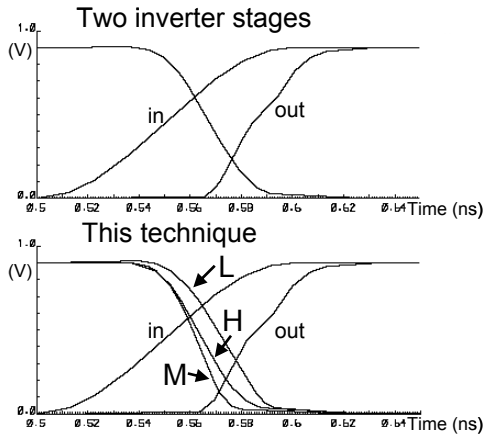

Fig. 3 Transient waveform comparison of the circuit shown in Fig. 2 and an equivalent driver (inverter size of 40w/20w and 160w/80w, respectively) and equal FO=4 load.

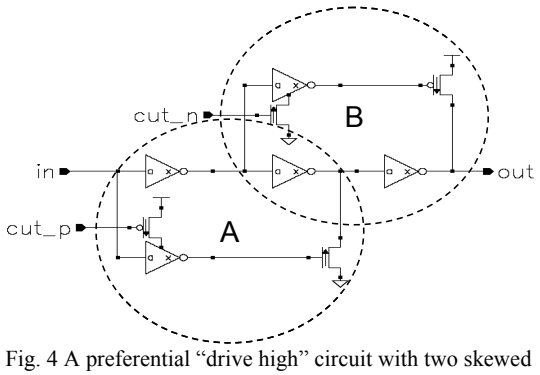

Fig. 4 A preferential "drive high" circuit with two skewed driver components, A and B, connected in an sequentially overlapped fashion.

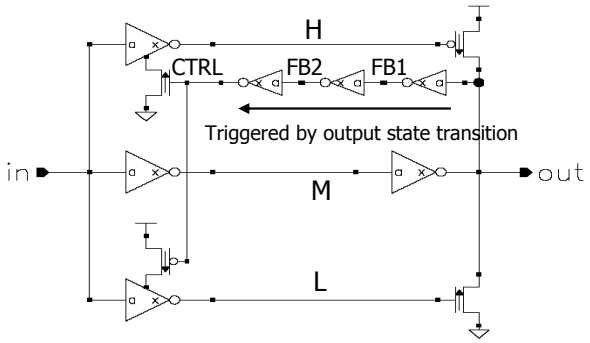

Fig. 5 Circuit schematic diagram where the low gate leakage mode is automatically triggered. Odd number of small inverters are inserted between the output and header/footer control gates.

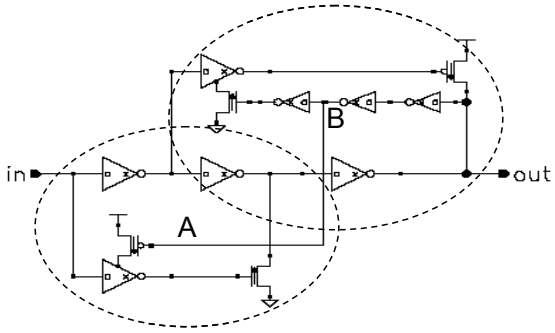

Fig. 7 Automated low leakage enabling scheme for the circuit of Fig. 5 by using small feedback inverters

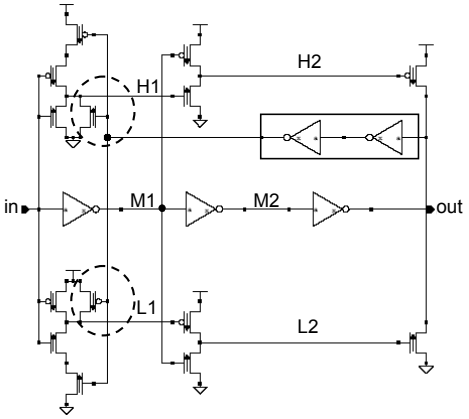

Fig. 9 A three-stage driver. Even number of small inverters are inserted between the output and control gates. Encircled small keepers bring nodes H1 and L1 to high or low states.

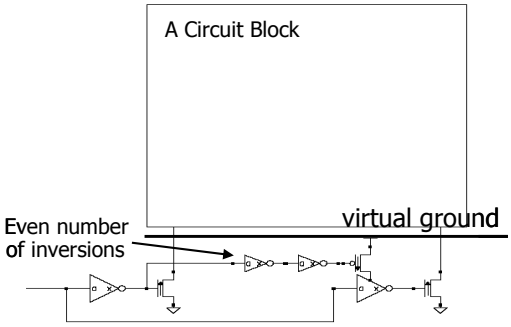

Fig. 11 Virtual ground resource management to a circuit block by using the floating node gate leakage reduction technique

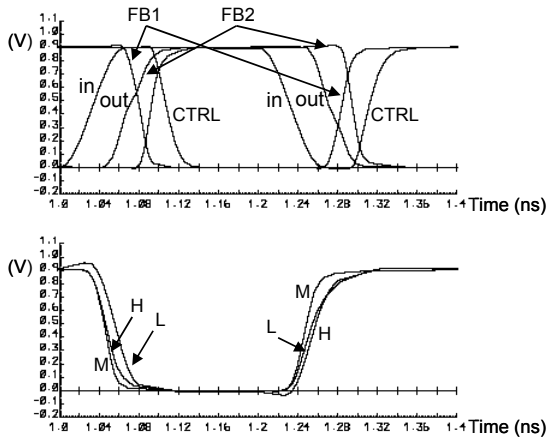

Fig. 6 Transient waveforms of the circuit shown in Fig. 6

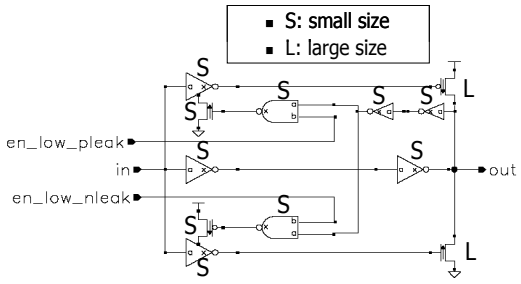

Fig. 8 Characterization method for the low gate leakage circuit. S and L designate device size choices.

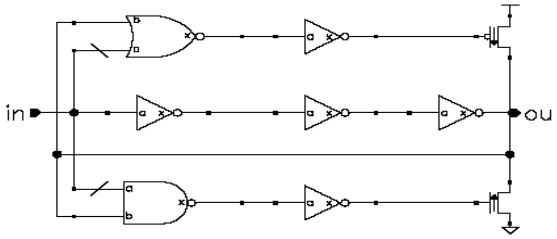

Fig. 10 An alternative three stage driver where every node in the circuit is placed in a known state. Similar to the circuit in Fig. 9, devices with keeper purposes are made small in the NAND2 and NOR2 gates.

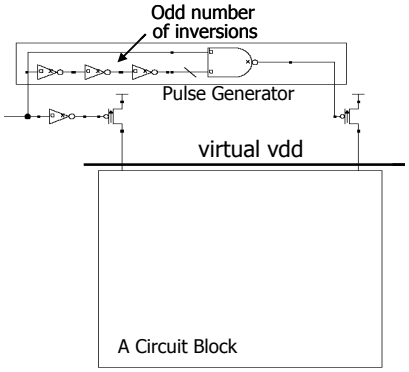

Fig. 12 Virtual vdd resource management to a circuit block by using the hard turn-off gate leakage reduction technique