A Self-adjusting Scheme to Determine the Optimum RBB by Monitoring Leakage Currents

Nikhil Jayakumar* Sandeep Dhar‡ Sunil P Khatri*

* Department of EE, Texas A&M University, College Station TX 77843
‡ National Semiconductor, Longmont, CO 80501

ABSTRACT

Reverse body biasing (RBB) is often used to reduce the leakage power of a device. However, recent research has shown that if this applied RBB is too high, the leakage power can actually increase due to the contribution of Band-to-Band Tunneling (BTBT) currents. Hence, there exists an optimal RBB value at which the leakage is minimum. This optimum point can vary with temperature and process variations. In this paper we show that it is desirable to operate at the optimal RBB point which minimizes total leakage. We present a scheme that monitors the total leakage current (the sum of the sub-threshold, BTBT and gate leakage) of an IC with a representative leaking device and, using this monitored value, *automatically* finds the optimum RBB value across temperature and process corners, using a self-adjusting circuit. Our approach has a modest placed-and-routed area utilization, and a low power consumption.

Categories and Subject Descriptors: B.7.1 [Integrated Circuits]: VLSI

General Terms: Design, Measurement

Keywords: Leakage power, Body-biasing, Self-adjusting

1. INTRODUCTION

Leakage power is expected to exceed dynamic power consumption in the near future [1]. One of the methods to reduce leakage power is by increasing the threshold voltages (V_T) of the device. This is done either statically (through use of multi-threshold devices) or dynamically (through RBB).

The sub-threshold leakage (cut-off) current of a transistor decreases with greater applied RBB. Reverse Body Biasing affects V_T through body effect, and sub-threshold leakage has an exponential dependence on V_T.

However, while the sub-threshold leakage decreases, there are other components to the leakage current that have to be considered as well. Two of these are bulk Band-to-Band-Tunneling (BTBT) and surface BTBT. Bulk BTBT is commonly referred to as simply BTBT while surface BTBT is commonly called Gate Induced Drain Leakage (GIDL) [2, 3]. While GIDL does not play a major role at RBB [2], BTBT increases with applied RBB [2, 4, 5, 6]. This means that there is an optimum RBB voltage at which the total leakage power (the sum of the sub-threshold leakage, the gate leakage, BTBT and GIDL) is minimum [2, 4, 5, 6]. In modern processes this optimum point is reached before the upper limit of the RBB (based on the voltage at which the bulk-drain / bulk-source junction breaks down). Hence, in any scheme where leakage power reduction is desired through RBB, there is a need to find this optimum point, in order to avoid applying too high an RBB and inadvertently *increasing* total leakage.

This paper describes a scheme that monitors the leakage current variation of a NMOS/PMOS device with applied Reverse Body Bias (RBB) and finds the optimum RBB that yields the lowest leakage. In section 2 we discuss the motivation behind our work. Section 3 discusses previous approaches to dynamically adjust body-bias. Section 4 describes our approach to dynamically self-adjust the RBB of PMOS and NMOS devices in order to obtain a minimum total leakage, along with experimental results that support the utility of our scheme. Conclusions and future work are discussed in section 5.

2. MOTIVATION

In this paper we are concerned with minimizing the total leakage current (the sum of the sub-threshold, BTBT and gate leakage) through a non-conducting (turned-off) device in a static CMOS design. In the case of an NMOS device this would mean we are concerned with the leakage through an NMOS device when its drain terminal is at VDD, its source and gate terminals are at GND and bulk terminal (p- well)of the device is at a certain RBB value. In such a scenario, the leakage current measured at the drain of the device is due to three sources – (i) the sub-threshold leakage from the drain to the source of the device, (ii) the gate leakage current from the drain to the gate and (iii) the drain-bulk junction current. The drain-bulk leakage current has three main components – bulk BTBT (or simply BTBT), surface BTBT (or GIDL) and the classical reverse biased PN junction current [7, 2]. The two BTBT currents dominate the reverse bias PN junction current. While the sub-threshold leakage decreases with increased RBB (due to the increase in V_T of the device), bulk BTBT current increases with RBB. The drain-gate leakage current does not change appreciably with applied RBB [4]. Also, at RBB, bulk BTBT dominates GIDL [2]. Hence it is mainly the sub-threshold and the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
DAC 2005, June 13–17, 2005, Anaheim, California, USA.
Copyright 2005 ACM 1-59593-058-2/05/0006 ...$5.00.

BTBT component of the leakage currents that change with applied RBB. Also, since these two components behave differently with respect to RBB, there exists an optimal RBB value [5, 2, 4]. We performed experiments on a test-chip manufactured using the TSMC 0.13μm triple well process. The test chip had one large PMOS (W_{eff} = 676mm, L_{eff} = 0.13μm) and one large NMOS (W_{eff} = 504mm, L_{eff} = 0.13μm) device. When a device is turned-off, the current measured at the source represents the sub-threshold leakage current from the drain to the source (I_{ds}), the current measured at the gate represents the gate leakage from the drain to the gate (I_{dg}) and current measured at the bulk contact represents the drain/source to bulk current (I_{db},I_{sb}). Since the drain is at VDD, most of the bulk current is from the drain (i.e. I_{db} dominates I_{sb}). The current measured at the drain of the device (I_{leak}) was found to be approximately the sum of the currents measured at the gate, source and bulk terminals confirming that I_{sb} is very small in practice.

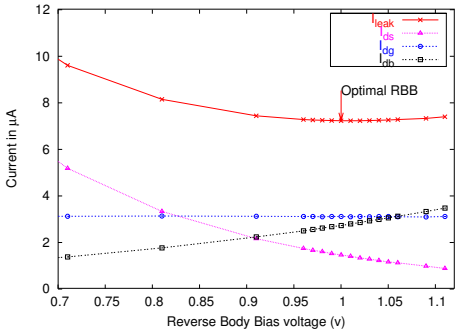

Figure 1: Leakage current components for large NMOS device at 25°C

Figure 1 shows measurements taken from our manufactured test chip for a non-conducting NMOS device at a temperature of 25° C with the RBB being swept from 0.7v to 1.1v below the source terminal. The VDD used was 1.2v. In this case the optimal RBB value is 1.0v.

The optimum RBB value can shift with temperature and process variations. Table 1 shows the penalty (in terms of percentage of leakage power increase from optimum) for the large NMOS device, with temperature and process variations. In columns 1 and 2, we describe the leakage penalty variation if the RBB is fixed to the optimum value (1.015V) for one particular temperature (25° C). Columns 3 and 4 report the leakage penalty variation with process. Column 3 reports the variation from typical (T) process for V_{T_N} and l_{eff} respectively [1]. A "S" label indicates a slow corner, and an "F" label indicates a fast corner. Column 4 reports the leakage penalty variation if the RBB is fixed to the optimum value (1.015V) for the typical corner for both V_{T_N} and l_{eff}.

Table 1 proves that fixing the RBB at a particular value may not be a good idea if we are interested in reducing leakage over all temperature and process variations.

We hence need a scheme by which we can monitor the leakage current of a chip and automatically self-adjust the RBB value of the PMOS and NMOS devices to keep the leakage power as low as possible. The problem of monitoring the optimum point is compounded by the fact that the total leakage current can vary by as much as 3 orders of magnitude over temperature and RBB variations. The leakage monitor must be able to find the optimum RBB point, over this wide range of currents.

[1]The variation of V_{T_N} is ±8% from nominal, and the l_{eff} variation is ±10nm from nominal.

Temp (°C)	Lkg penalty	Corner	Lkg penalty
-40	23.38%	TS	16.15%
0	6.99%	TF	4.02 %
25	0%	FF	10.73%
70	35.29%	SS	58.3 %
125	163.55%	ST	20.77%

Table 1: Leakage penalty variation

3. PREVIOUS WORK

In [4], a simple circuit is presented that helps find the optimal RBB value. The accuracy of this circuit is dependent on the assumption that gate leakage can be neglected (or is very small) and that sub-threshold leakage is negligible when compared to the BTBT current in a stack of 2 non-conducting devices. Under these assumptions, the authors claim that the optimal RBB value occurs at the point where the leakage current through two stacked non-conducting devices is primarily BTBT current, and is equal to half the leakage through a single non-conducting device. However, experiments with our test chip show that these assumptions are significantly inaccurate.

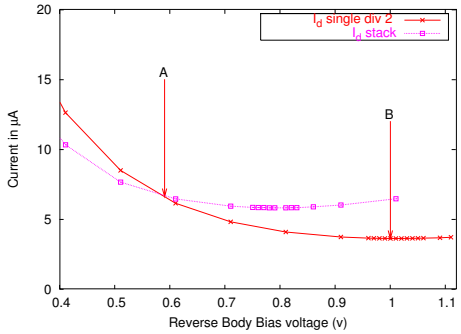

Figure 2: Leakage Current for Stacked and Single Devices

Figure 2 shows a plot of *half the leakage current* through a single non-conducting NMOS device on our test-chip (labeled as 'Id single div 2') and the leakage current through a stack of two non-conducting NMOS devices (labeled as 'Id stack'). The currents were measured at a temperature of 25° C. The arrow labeled 'A' shows the optimal RBB value as would be suggested by the circuit in [4] while the arrow labeled 'B' shows the actual optimal RBB value for a single non-conducting NMOS device at 25°C. We found that if the RBB value marked by A was used as the "optimal" RBB instead of the RBB value pointed by B, the leakage current for a single non-conducting NMOS device (at 25°C) would be 70% higher than optimum.

In [8] and [9] the authors suggest sensing the voltage dropped by a leaking device towards the goal of adjusting the body bias and thus controlling the leakage. To amplify the leakage current, the gate bias is set to a value such that the leaking device is still cut-off but has a high enough leakage current to drop a significant voltage. This voltage is sensed and if it crosses a certain threshold, RBB is applied.

The authors of [10] suggest a similar mechanism as a way of stabilizing sub-threshold CMOS logic. However, [8, 9, 10] do not target the problem of finding the optimum RBB value.

No method known to the authors of this paper tracks the variation of total leakage with RBB dynamically and finds the exact optimum RBB value using a self-adjusting scheme.

4. LEAKAGE MONITORING / SELF-ADJUSTING SCHEME

Our leakage monitoring scheme is based on measuring the time taken for the leakage current to discharge (for monitoring the leakage of a leaking NMOS device) a capacitive load. For a leaking PMOS device, the time taken for charging-up the load is considered. A higher leakage would be indicated by a shorter time to discharge the load while a longer time to discharge the load would indicate a lower leakage. To monitor the leakage current of an NMOS device, the capacitively loaded node is initially pre-charged to a logic-high value. The leakage current is estimated by measuring the time taken to discharge this node. Similarly, for a leaking PMOS device, the capacitively loaded node is initially pre-discharged and the leakage current is estimated based on the time taken to charge this node to a logic-high value.

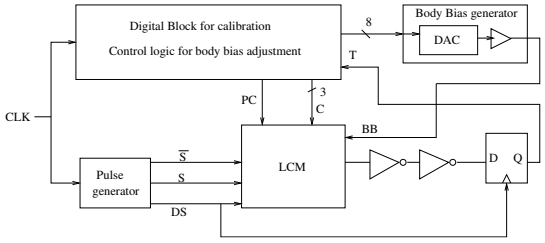

Figure 3: LCM Scheme Block Diagram(for NMOS)

The leakage monitoring scheme is conceptually illustrated in Figure 3 (for NMOS bulk control). A similar structure is used to control the PMOS bulk node. The 3 main blocks of the leakage monitoring scheme are: (i) a leakage current monitoring (LCM) block that contains a representative leaking device, (ii) a digital block to interface with the LCM and control the body bias voltage and (iii) a programmable body bias voltage generator to translate the body bias control value from the digital block into a body bias voltage value. In this paper we deal with the leakage monitoring block and the digital control block. Details of the bias generator are omitted in this work.

4.1 Leakage Current Monitoring Block (LCM)

In this section the design and operation of the LCM block will be discussed. We use the LCM for NMOS devices as an example. Our objective is to track the variation of total leakage current through a circuit with applied RBB. However, placing a current monitoring device in series with the IC supply and circuit power rails of the logic devices is not an option since the addition of such a device would increase the delay of the circuit. Hence we choose a representative device to model the leakage of the entire circuit. The optimal RBB value is smaller for stacked devices when compared to single (unstacked) devices. This is because sub-threshold leakage is lower for stacked devices and hence BTBT dominates at a lower RBB value. However, it is infeasible to have separate substrates for stacked and non-stacked devices. In our scheme we chose a non-stacked device as the representative leaking transistor based on the intuition that for most ICs the dominant source of leakage is from unstacked devices. However, if we were to design a leakage monitor to track the leakage of a chip with stacked devices being the dominant source of leakage, the leakage monitor would have to use stacked devices as the representative leaking transistors.

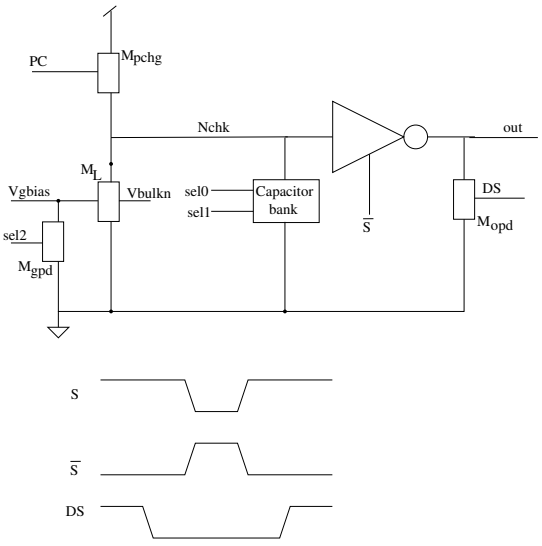

Figure 4: LCM for NMOS devices

The leakage current variation of NMOS and PMOS devices is monitored separately. Figure 4 shows the circuit that implements the leakage current monitoring block for NMOS devices. In Figure 4, device M_L is the representative leaking transistor. Transistor M_{pchg} is the device that precharges the node Nchk. M_L and M_{pchg} are sized relatively so that the leakage of M_L dominates the leakage of M_{pchg}. The leakage monitoring scheme is based on the idea that the time taken for the leaking transistor M_L to discharge the node Nchk would represent the leakage current through M_L and hence the leakage current through the entire circuit.

In Figure 4, the capacitor bank and the device M_{gpd} allow the LCM to work over a wide range of leakage currents. If the leakage current is too low, it needs to be magnified for the LCM to work effectively. This is done by first disconnecting the capacitor bank from Nchk (to speed up the rate of discharge of the node Nchk). Further magnification of the leakage current is achieved by turning off M_{gpd} and hence increasing the gate bias of M_L (in a similar manner as in [8, 9]) to a value of about 0.1V above GND (such that M_L is still in the sub-threshold/cut-off mode). The circuit that generates this low gate bias voltage is designed such that its output voltage decreases with an increase in temperature. Without this feature, the current in M_L increases too rapidly with increasing temperature when M_{gpd} is off.

The LCM works by 'sampling' (turning on the tri-stateable inverter at the output of the LCM) the node Nchk at regular intervals. During this sampling, the output pulldown device, M_{opd} is turned off. Note that the sampling period is

short, which keeps the power consumption of the LCM low. If the node Nchk has fallen low enough, the output of the LCM goes high and this output is buffered and then latched in a D flip-flop. The DFF output (shown as 'T' in Figure 3) triggers the digital block. The purpose of this trigger signal will be explained in the following sub-section.

The LCM for PMOS devices is implemented in a manner similar to that of the LCM for NMOS devices.

4.2 Digital Control Block

The Digital Control Block contains an 8-bit counter that counts up till either the end of the count is reached or till it receives a trigger signal from the DFF at the output of the LCM. When a trigger signal is received, the value of the 8-bit counter is stored. This counter value is representative of the time taken for the transistor M_L to discharge the node Nchk and is hence a measure of the leakage current of M_L. Next, the node Nchk is precharged (signal 'PC' goes low) and held in this precharged state till a new body-bias is set. The applied RBB value is increased till the point at which the new counter value is smaller than the previous counter value (the point at which the leakage current starts increasing with applied RBB). If the end of the count is reached before a trigger signal is received, this implies that the total leakage is too low. In such a situation, control signals from the digital block are applied to the LCM to magnify the leakage current. The digital block sends appropriate signals (shown as 'C' in Figure 3 and sel0, sel1, sel2 in Figure 4) that control the capacitor bank and M_{gpd} in the LCM to achieve this magnification, as described in Section 4.1.

In summary, our leakage monitoring scheme works by essentially converting the problem of sensing the total leakage current into one of measuring the time taken for a representative leaking transistor to discharge a purely capacitive load. The time taken is measured using a counter and the applied RBB is increased in linear steps till the time measured by the counter for a particular body-bias value, is shorter than the time measured by the counter for a previous body-bias value used. The LCM is designed for correct operation over a wide range of leakage currents.

The accuracy of the scheme can be improved by increasing the frequency of the clock and hence increasing the frequency of sampling of the node Nchk. We utilize a clock with a period of 2ns. Simulations showed the proposed scheme has a very small power consumption of 11.4μA. Of this, the LCM block consumes about 4μA, while the digital control block consumes about 6μA. Note that simulations were done at 1.2V at 125°C (to model the worst-case power consumption) for a TSMC 0.13μm process. The digital block was synthesized using a 0.13μm process standard-cell library.

Cell	Width(μm)	Height(μm)	Area(μm^2)
LCM NMOS	77.87	3.285	255.7
LCM PMOS	86.41	3.285	283.86
Pulse generator	38.22	3.285	125.55
Total	-	-	665.11

Table 2: Size of the standard-cell implementations of the LCMs and pulse generator

We also created layout macro-cells for the pulse generator (that generates the S and DS signals for the LCM block), the LCM block for NMOS leakage monitoring and the LCM block for PMOS leakage monitoring. The LCM blocks include the circuitry required to generate the low Vgbias voltage. Table 2 shows the placed-and-routed size of each cell in the layout.

5. CONCLUSION

In this paper, we have described an automatic, self-adjusting mechanism to find the optimal RBB value to minimize total leakage. Our method consists of a leakage current monitor, and a digital block that senses the discharging (charging in the case of a PMOS transistor) of a representative NMOS device in the design. Based on the speed of discharge, which is faster for leakier devices, an appropriate RBB value is applied.Our technique is able to find the optimal RBB point, and incurs very reasonable placed-and-routed area and power penalties in its operation. In the near future, we plan to fabricate this concept to characterize its efficacy on silicon.

6. REFERENCES

[1] "The International Technology Roadmap for Semiconductors." `http://public.itrs.net/`, 2003.

[2] A. Keshavarzi, S. Narendra, S. Borkar, C. Hawkins, K. Royi, and V. De, "Technology scaling behavior of optimum reverse body bias for standby leakage power reduction in CMOS ICs," in *Proceedings International Symposium on Low Power Electronics and Design*, pp. 252–254, Aug 1999.

[3] C. Neau. Personal communication, June 2004.

[4] C. Neau and K. Roy, "Optimal body bias selection for leakage improvement and process compensation over different technology generations," in *Proceedings of the International Symposium on Low Power Electronics and Design*, pp. 116 – 121, Aug 2003.

[5] X. Liu and S. Mourad, "Performance of submicron CMOS devices and gates with substrate biasing," in *The IEEE International Symposium on Circuits and Systems*, vol. 4, (Geneva), pp. 9–12, May 2000.

[6] Y.-S. Lin, C.-C. Wu, C.-S. Chang, R.-P. Yang, W.-M. Chen, J.-J. Liaw, and C. Diaz, "Leakage scaling in deep submicron CMOS for SoC," *IEEE Transactions on Electron Devices*, vol. 49, pp. 1034–1041, June 2002.

[7] S. Mukhopadhyay, H. Mahmoodi-Meimand, C. Neau, and K. Roy, "Leakage in nanometer scale CMOS circuits," in *International Symposium on VLSI Technology, Systems, and Applications*, pp. 307–312, 2003.

[8] T. Kuroda, T. Fujita, S. Mita, T. Nagamatsu, S. Yoshioka, K. Suzuki, F. Sano, M. Norishima, M. Murota, M. Kako, M. K. M. Kakumu, and T. Sakurai, "A 0.9-v, 150-mhz, 10-mw, 4 *mm* 2, 2-d discrete cosine transform core processor with variable threshold-voltage (VT) scheme," *IEEE Journal of Solid-State Circuits*, vol. 31, pp. 1770–1779, Nov 1996.

[9] T. Kobayashi and T. Sakurai, "Self-adjusting threshold-voltage scheme (SATS) for low-voltage high-speed operation," in *Proceedings of the IEEE Custom Integrated Circuits Conference*, pp. 271–274, May 1994.

[10] H. Soeleman, K. Roy, and B. Paul, "Robust subthreshold logic for ultra-low power operation," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 9, no. 1, pp. 90–99, 2001.