Supply Switching With Ground Collapse for Low-Leakage Register Files in 65-nm CMOS

Hyung-Ock Kim, Bong Hyun Lee, Jong-Tae Kim, Jung Yun Choi, Kyu-Myung Choi, and Youngsoo Shin

***Abstract*—Power-gating has been widely used to reduce subthreshold leakage current. However, the extent of leakage saving through power-gating diminishes with technology scaling due to gate leakage of data-retention circuit elements. Furthermore, power-gating involves substantial increase of area and wirelength. A circuit technique called supply switching with ground collapse (SSGC) has recently been proposed to overcome the limitation of power-gating. The circuit technique is successfully applied to the register file of ARM9 microprocessor in a 1.2 V, 65-nm CMOS process, and the measured result is reported for the first time. The leakage current is reduced by a factor of 960 on average of 83 dies at 25 °C, and by a factor of 150 at 85 °C. Compared to a register file implemented in conventional power-gating, leakage current is cut by a factor of 2.2, demonstrating that SSGC can be a substitute for power-gating in nanometer CMOS.**

***Index Terms*—Leakage, low-power, power gating, register file, standard cell.**

I. Introduction

Increasing leakage current has been main concern for designing VLSI systems in nanoscale CMOS technology. Power-gating [3], which cuts off power rails when circuits are not in use, is one of the most popular circuit techniques to reduce subthreshold leakage, the largest component of leakage in most technologies, and has been widely used in the semiconductor industry [4]–[6]. However, it has recently been reported [1] that the benefit of power-gating in reducing leakage current diminishes with technology scaling. This is because of gate leakage in extra circuitry to implement power-gating, such as data-retention storage elements, output-isolation circuits, and current switches; power-gating cannot eliminate gate leakage even though it can eliminate most subthreshold leakage. The gate leakage, in fact, can take 70% to 80% of total standby leakage of power-gated circuits in nanometer technologies [1].

Implementing power-gated circuits involves overhead of area and wirelength (17% of area and 18% of wirelength on average of several sequential circuits [2]) due to extra circuitry; the overhead only gets worse with technology scaling, since bigger current switches are needed as V_{dd} is reduced [7]. Among the extra circuitry to implement power-gating, data-retention flip-flop represents the biggest overhead. A data-retention flip-flop [3], [5], [8]–[10] preserves a data in an extra latch, which is fully biased during standby mode since it is never power-gated. Therefore, all the extra latches induce continuous gate leakage, take extra area, and are wired to power-management unit (PMU).

Manuscript received May 13, 2008; revised August 20, 2008 and December 02, 2008. First published July 21, 2009; current version published February 24, 2010. This work was supported by Samsung Electronics.

H.-O. Kim, B. H. Lee, J.-T. Kim, J. Y. Choi, and K.-M. Choi are with Samsung Electronics, Yongin, Gyeonggi-Do 449-711, Korea (e-mail: hokim@dtlab.kaist.ac.kr; bonghyun.lee@samsung.com; jtkim@samsung.com; jungyun74.choi@samsung.com; kmchoi@samsung.com).

Y. Shin is with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea (e-mail: youngsoo@ee.kaist.ac.kr).

Digital Object Identifier 10.1109/TVLSI.2009.2012429

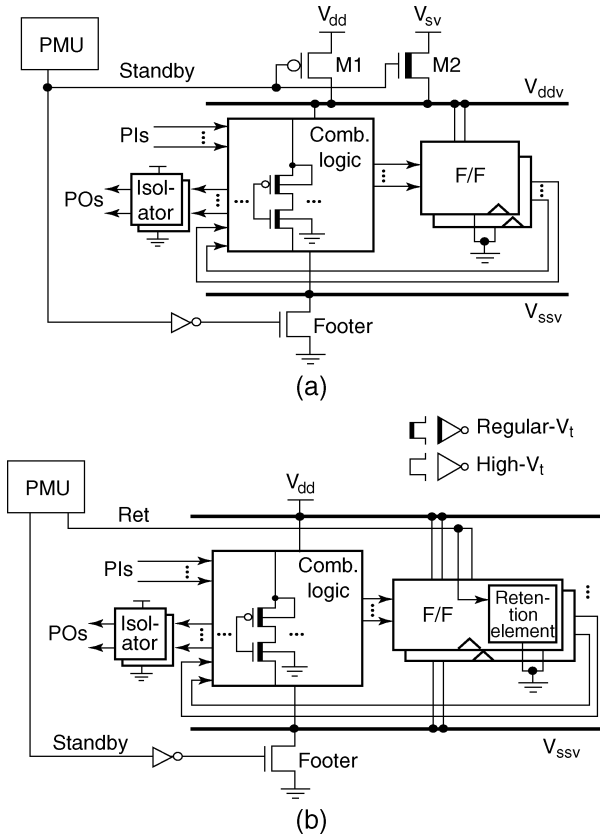

Fig. 1. (a) SSGC circuit and (b) power-gated circuit.

To overcome the limitation of power-gating, a new circuit technique, called supply switching with ground collapse (SSGC) [1], has been proposed. SSGC allows the use of conventional flip-flops with only slight modification on their layouts, thus removes the aforementioned overhead of data-retention flip-flops. The leakage of flip-flops is suppressed by lowering supply voltage, but still high enough to maintain the data of flip-flops; ground is collapsed in the side of combinational sub-circuit, which, together with lowered supply voltage, suppresses the leakage of combinational sub-circuit.

In this paper, SSGC is applied to the register file of ARM9 microprocessor, consisting of 30 registers of 32-bit in total, in a 1.2 V, 65-nm CMOS process; the measured result is reported for the first time. The leakage current is reduced by a factor of 960 on average of 83 dies at 25 °C (150 °C at 85 °C), and by a factor of 2.2 compared to the same register file implemented in conventional power-gating. The turn-on sequence of SSGC is investigated in an effort to reduce transient current spikes; efficient method is proposed and implemented in the register file.

II. Supply Switching With Ground Collapse

Fig. 1(a) illustrates supply switching with ground-collapse. Conventional power-gating is shown in Fig. 1(b) for comparison. When the circuit is in active mode ($\texttt{standby} = 0$), V_{dd} is supplied through a pMOS switch M1 (M2 being turned off) and an nMOS switch footer is turned on. The amount of voltage drop that can be tolerated by the circuit determines the size of M1 and footer, i.e., larger size of M1 and footer if smaller voltage drop is allowed. High-V_t is used for both switches to reduce their subthreshold leakage in standby mode when

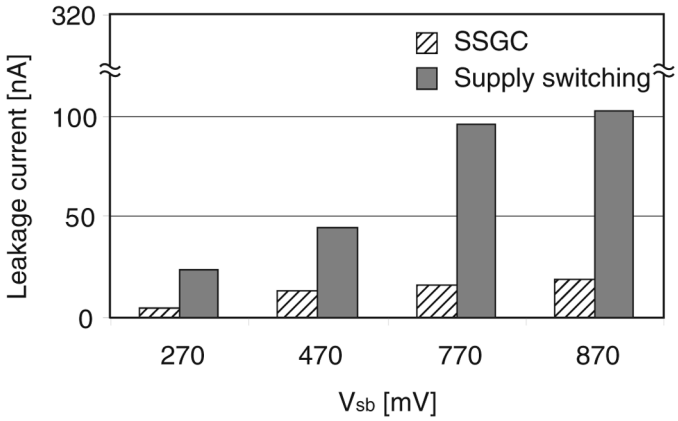

Fig. 2. Comparison of standby leakage current of s3384 with SSGC and supply switching alone in 65-nm technology with 1.2 V V_{dd}. Leakage current of non-power-gated circuit is 320 nA.

they are turned off. Low-V_t (or regular-V_t or multiple-V_t) is used for implementing the circuit itself for high performance.

When the circuit goes to standby mode ($\texttt{standby} = 1$), an nMOS switch M2 is turned on (while M1 and footer are turned off) and standby voltage V_{sv} is supplied to the circuit. The voltage V_{sv} is considerably lower than V_{dd} thereby suppressing gate leakage as well as subthreshold leakage of flip-flops, but still high enough to guarantee the data of the flip-flops to be preserved since they are directly connected to V_{ss} (part of it is connected to V_{ssv} as will be explained shortly) as opposed to V_{ssv} in power-gating shown in Fig. 1(b).

There are two factors that contribute to leakage saving of SSGC: switching to lower supply voltage V_{sv} and power gating through footer. The amount of contribution from each factor depends on V_{sv}. Fig. 2 compares leakage current of s3384, which is one of the ISCAS benchmark circuits, when SSGC is used and switching to V_{sv} alone is used (without using footer) while we vary the value of V_{sv}. When V_{sv} takes small value, switching to V_{sv} is a main factor for reducing leakage, but as V_{sv} increases, it is clear that both factors contribute to leakage saving.

Note that we use nMOS device for M2 switch. At most process corners and at most temperatures, the total leakage current of the circuit when we use nMOS switch is less than that when we use pMOS switch. This is because we determine V_{sv} at maximum temperature, where the leakage is the largest and so is the potential drop across M2, and V_{sv} is larger with pMOS device than with nMOS device for the same V_{ddv}. In case of s3384, 260 mV is the minimum voltage that guarantees the data of the flip-flop to be preserved; V_{sv} has to be at least 458 mV with pMOS device while it can be as low as 270 mV with nMOS device. With V_{sv} set to the aforementioned values, V_{ddv} takes different value at different temperature and at different process corner, but it always takes larger value with pMOS than with nMOS for M2 switch, which is why nMOS produces less leakage. For example of s3384, at 75 °C of nominal process corner, V_{ddv} is 439 mV producing 111 nA with pMOS device; V_{ddv} is 270 mV producing 57 nA with nMOS device.

Fig. 3(a) shows a D-type flip-flop that is used in SSGC. Note that the slave latch is directly connected to V_{ss} while all other components are connected to footer (not shown in the figure) to further reduce leakage. The standby leakage (both subthreshold- and gate-) from the slave latch is reduced by resorting to lowered voltage V_{sv} rather than to footer; the leakage from the remaining part is reduced by footer and V_{sv}. This arrangement allows us to use conventional flip-flops (with slight modification of their layouts due to power rail connection), thereby freeing us from the overhead of data-retention flip-flops used in power-gating [9], which is shown in Fig. 3(b).

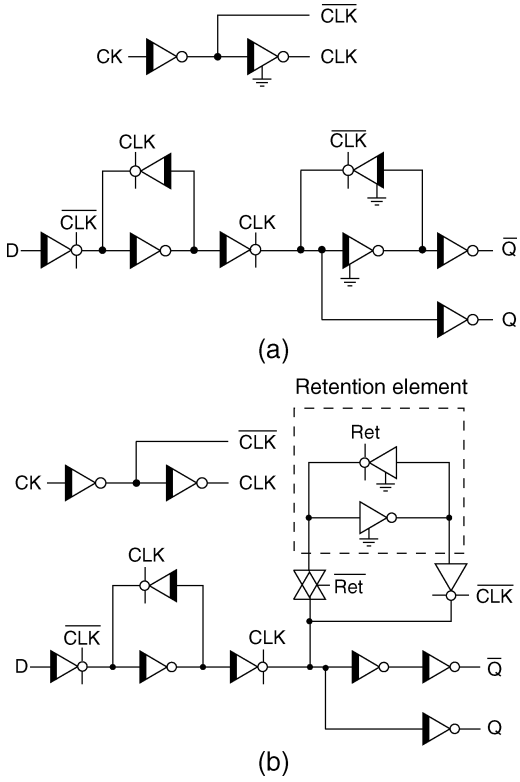

Fig. 3. (a) SSGC flip-flop and (b) data-retention flip-flop.

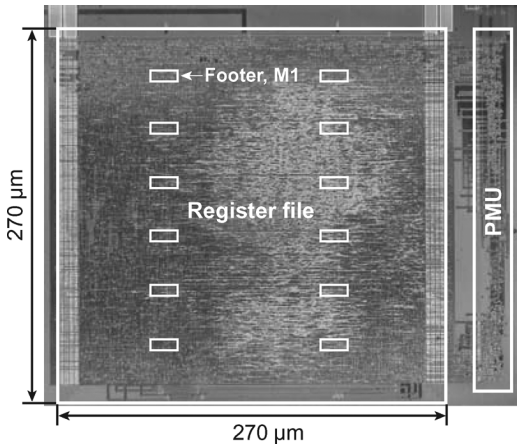

Fig. 4. Chip micrograph.

III. Test Chip Implementation and Results

A. Test Chip Implementation: A Register File

A test chip, along with power management unit, was fabricated with 65-nm bulk digital CMOS process with six metal layers. Fig. 4 shows a chip micrograph. The target circuit is a register file of ARM9 microprocessor, which is illustrated in Fig. 5. The register file consists of a 3-read 2-write register bank and input/output control logic. The register bank consists of 30 registers of 32-bit, which are mapped to 16 logical addresses. The PMU is responsible for deciding one of six operation modes of the register bank (see Fig. 5). Feed-through paths from write ports to read ports provide data forwarding between microprocessor pipeline stages. The register file has 1.2 K flip-flops and 4.4 K logic gates in total.

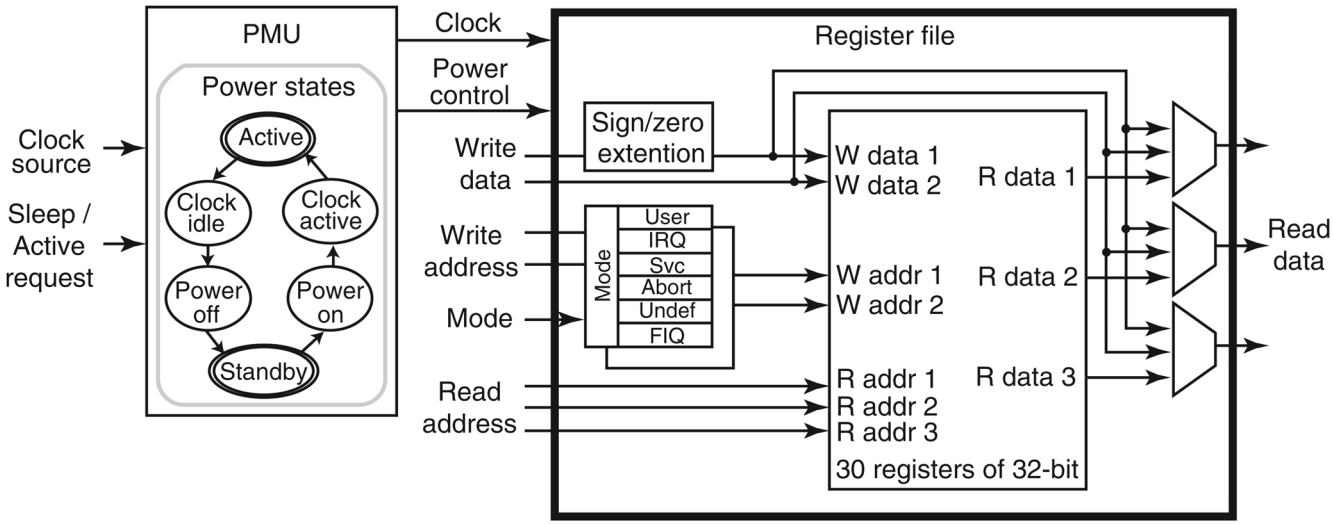

Fig. 5. Block diagram of register file.

TABLE I
COMPARISON OF CRITICAL PATH DELAY, AREA, AND SWITCHING POWER CONSUMPTION (FROM SIMULATION) OF NON-POWER-GATED, POWER-GATED, AND SSGC REGISTER FILES

	Critical path delay (ns)	Area (μm^2)	Switching power (mW)
Non-power-gated	2.98	20,314	1.2
Power-gated	3.00	26,164	1.2
SSGC	3.00	22,712	1.2

The design flow for SSGC [1] was employed for test chip implementation. The size of footer and M1 switch are determined by assuming 5 mV total voltage drop across them, which has negligible impact on critical path delay as shown in Table I. The footer and M1 switch (along with M2 switch) occupy the area of 1442 μm^2 while the footer alone takes 223 μm^2 in power-gating for the same 5 mV voltage drop. This can be understood from the fact that the footer should be sized for the smaller voltage drop in SSGC with the same amount of discharging current as in power-gating, and that M1 switch should be sized for charging current, which is not necessary in power-gating. The footer and M1 switch are placed regularly as shown in Fig. 4.

Since we reserve 45 mV noise margin for power/ground network, the total guardband for this design is 50 mV for 1.2 V V_{dd}. The area of total cells is 22 712 μm^2; the same register file implemented in power-gating occupies 26,164 μm^2, 15% more area, as shown in Table I. This difference is mainly caused by flip-flops: data-retention flip-flop of Fig. 3(b) is 48% larger than SSGC flip-flop of Fig. 3(a) (29% larger for a flip-flop with reset, which is also used in the register file). The total wirelength is 350.8 mm; it is 391.2 mm for power-gating (12% more wires). This difference is caused by gating-control signals such as `standby` and `ret` as shown in Fig. 1 (2 mm for SSGC and 11 mm for power-gating), and corresponding difference of wire congestion.

To determine standby voltage V_{sv}, SSGC flip-flop of Fig. 3(a), as well as the one with reset, is simulated with SPICE at all process corners and at all operating temperatures [1]. It is found out that 260 mV is the minimum voltage that guarantees the data of the flip-flop to be preserved. Since we need 30 mV of drop across M2 switch [see Fig. 1(a)], which is determined by total standby leakage through the circuit, V_{sv} has to be at least 290 mV. We set V_{sv} to 300 mV, which is supplied off-chip, thus is an overhead for this particular implementation.

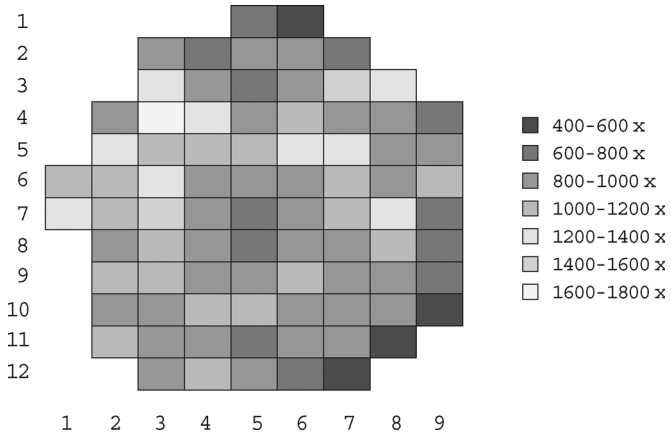

Fig. 6. Distribution of leakage savings over 83 dies on a wafer ($V_{dd} = 1.2$ V, $V_{sv} = 0.3$ V, T = 25 °C).

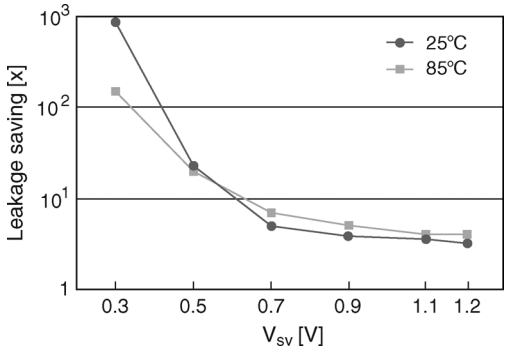

Fig. 7. Average leakage saving of dies with different standby voltage V_{sv} in 25 °C and 85 °C, $V_{dd} = 1.2$ V.

B. Leakage Measurement

By setting V_{sv} to 300 mV (M1 switch and footer being turned off), the standby leakage current of the register file is measured for 83 dies of a wafer at room temperature. We then measure the idle leakage of each die (circuit has no active computation, V_{dd} is supplied, and footer is turned on) assuming that it represents the leakage of non-power-gated

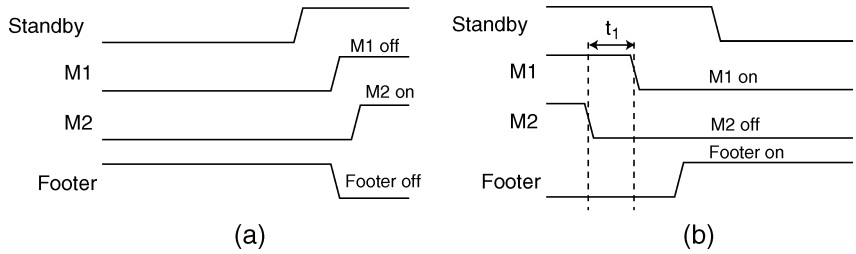

Fig. 8. Power sequencing of SSGC: (a) active to standby and (b) standby to active. The values for M1, M2, and footer correspond to their gate inputs.

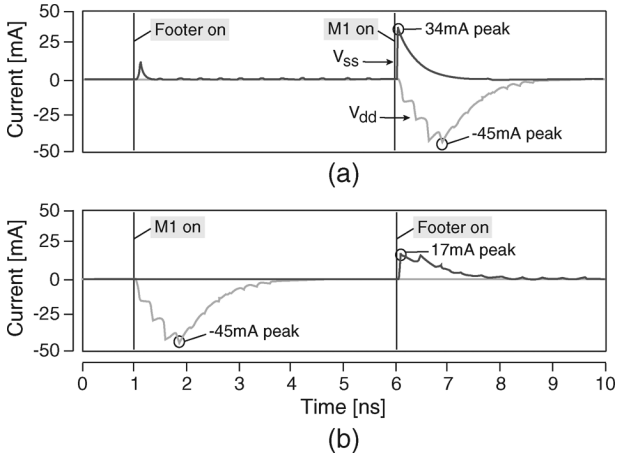

Fig. 9. Turn-on sequences of SSGC: (a) footer turned on before M1 and (b) M1 turned on before footer.

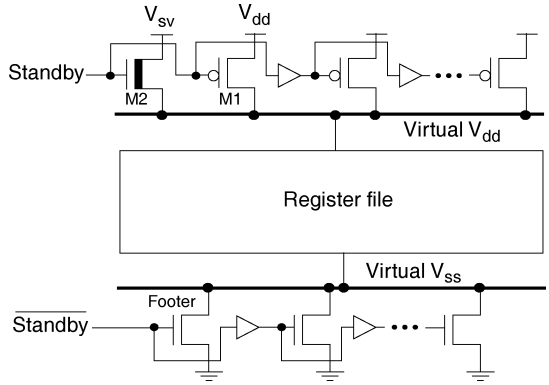

Fig. 10. Footer and M1 switches for register file.

register file. The leakage saving (over idle leakage) of all dies is illustrated in Fig. 6. The saving ranges from a factor of 400 to 1680, with average of 960. We repeat the same measurement in higher temperature of 85 °C. The average leakage saving is reduced to a factor of 150. This is because subthreshold leakage in the slave latch of the flip-flop [see Fig. 3(a)] increases with temperature. The measurement is also performed while we change V_{sv}, which is illustrated in Fig. 7. It is readily seen that the saving decreases with increasing V_{sv}, as it must, but 300 mV was high enough to save and restore the data of flip-flops.

C. Power Sequencing

The sequence of turning on or off switches (M1, M2, and footer) for the transition from active to standby mode is shown in Fig. 8(a). M1 and footer are turned off, followed by turning on M2. The sequence for the transition from standby to active mode is shown in Fig. 8(b). M2 is turned off, which is followed by turning on M1. During the period between the two events, denoted as t_1 [see Fig. 8(b)], V_{ddv} momentarily drops thus can adversely affect the integrity of states preserved in flip-flops; it, however, drops very slowly (about 2.5 mV in 10 ns) thus no care needs to be taken as long as t_1 is kept sufficiently short. The order of turning on M1 and footer affects current spikes at V_{ss}. Fig. 9 shows the simulation results for two different turn-on sequences. It turns out that turning on M1 switch and then footer [see Fig. 9(b)] yields less amount of current spikes at V_{ss}. This is because, by turning on M1, all internal logic states are restored (note that V_{ssv}, which can be charged up to V_{sv} during standby, is closer to V_{ss} rather than to V_{dd}), and the remaining charges trapped in V_{ssv} are subsequently discharged once the footer is turned on.

In our implementation of the register file, M1 and footer are physically grouped into 12 clusters, which are successively turned on as shown in Fig. 10 in an effort to further reduce current spikes [6].

IV. Conclusion

Supply switching with ground collapse has recently been proposed [1] as a substitute for power-gating in nanometer CMOS technology. The circuit technique was successfully applied to the register file of ARM9 microprocessor in 65-nm CMOS, demonstrating less leakage current by a factor of 2.2 compared to the same register file implemented in conventional power-gating. This was achieved with smaller area and shorter wirelength.

References

[1] Y. Shin, S. Heo, H. Kim, and J. Choi, "Supply switching with ground collapse: Simultaneous control of subthreshold and gate leakage current in nanometer-scale CMOS circuits," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 15, no. 7, pp. 758–766, Jul. 2007.

[2] H. Kim and Y. Shin, "Semicustom design methodology of power gated circuits for low leakage applications," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 6, pp. 512–516, Jun. 2007.

[3] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "A 1-V power supply high-speed digital circuit technology with multithreshold-voltage CMOS," *IEEE J. Solid-State Circuits*, vol. 30, no. 8, pp. 847–854, Aug. 1995.

[4] S. V. Kosonocky, M. Immediato, P. Cottrell, and T. Hook, "Enhanced multi-threshold (MTCMOS) circuits using variable well bias," in *Proc. Int. Symp. Low Power Electron. Des.*, Aug. 2001, pp. 165–169.

[5] H.-S. Won, K.-S. Kim, K.-O. Jeong, K.-T. Park, K.-M. Choi, and J.-T. Kong, "An MTCMOS design methodology and its application to mobile computing," in *Proc. Int. Symp. Low Power Electron. Des.*, Aug. 2003, pp. 110–115.

[6] P. Royannez, H. Mair, F. Dahan, M. Wagner, M. Streeter, L. Bouetel, J. Blasquez, H. Clasen, G. Semino, J. Dong, D. Scott, B. Pitts, C. Raibaut, and U. Ko, "90 nm low leakage SoC design techniques for wireless applications," in *Proc. IEEE Int. Solid-State Circuits Conf.*, Feb. 2006, pp. 138–139.

[7] T. Inukai, M. Takamiya, K. Nose, H. Kawaguchi, T. Hiramoto, and T. Sakurai, "Boosted gate MOS (BGMOS): Device/circuit cooperation scheme to achieve leakage-free giga-scale integration," in *Proc. Custom Integr. Circuits Conf.*, May 2000, pp. 409–412.

[8] V. Kursun, S. Tawfik, and Z. Liu, "Leakage-aware design of nanometer SoC," in *Proc. Int. Symp. Circuits Syst.*, May 2005, pp. 3231–3234.

[9] H. Mair, A. Wang, G. Gammie, D. Scott, P. Royannez, S. Gururajarao, M. Chau, R. Lagerquist, L. Ho, M. Basude, N. Culp, A. Sadate, D. Wilson, F. Dahan, J. Song, B. Carlson, and U. Ko, "A 65-nm mobile multimedia applications processor with an adaptive power management scheme to compensate for variations," in *Proc. Symp. VLSI Circuits*, Jun. 2007, pp. 224–225.

[10] Z. Liu and V. Kursun, "New MTCMOS flip-flops with simple control circuitry and low leakage data retention capability," in *Proc. Int. Conf. Electron., Circuits, Syst.*, Dec. 2007, pp. 1276–1279.

A 5-bit 3.2-GS/s Flash ADC With a Digital Offset Calibration Scheme

Ying-Zu Lin, Cheng-Wu Lin, and Soon-Jyh Chang

***Abstract*—In high-speed Flash analog-to-digital converters (ADCs), preamplifiers are often placed in front of a comparator to reduce metastability errors and enhance comparison speed. The accuracy of a Flash ADC is mainly limited by the random offsets of preamplifiers and comparators. This paper presents a 5-b Flash ADC with a digital random offset calibration scheme. For calibration, programmable resistive devices are used as the loading devices of the second-stage preamplifiers. By adjusting the calibration resistors, the input-referred offset voltage of each comparator is reduced to be less than 1/2 LSB. Fabricated in a 0.13-μm CMOS process, experimental results show that the ADC consumes 120 mW from a 1.2-V supply and occupies a 0.18-mm^2 active area. After calibration, the peak differential non-linearity (DNL) and integral non-linearity (INL) are 0.24 and 0.39 LSB, respectively. At 3.2-GS/s operation, the effective number of bits is 4.54 b, and the effective resolution bandwidth is 600 MHz. This ADC achieves figures of merit of 3.07 and 4.30 pJ/conversion-step at 2 and 3.2 GS/s, respectively.**

***Index Terms*—Digital calibration, digitally assisted analog-to-digital converter (ADC), Flash ADC, high-speed data converter, offset calibration.**

I. Introduction

Low-resolution gigasample/second analog-to-digital converters (ADCs) are extensively used in recent communication applications. In a receiver, an ADC converts RF or intermediate-frequency signals into digital codes for baseband processing. A Flash ADC is often utilized in such applications, owing to its high-speed potential and low latency. The accuracy of Flash ADCs is mainly limited by the offsets of the preamplifiers and comparators [1]. Offsets can be categorized as systematic, random, and dynamic offsets. A systematic offset occurs due to improper architecture or circuit design. Its quantity is predictable during the design stage. Process variations during fabrication phases lead to the generation of random offsets. A dynamic offset is associated with the regenerative latch in a comparator. Preamplifiers are often placed in front of a latch-based comparator to suppress dynamic offsets and mitigate kick-back noises. The preamplifiers and comparators, however, contribute random offsets. Enlarging device sizes reduces the random offset since the quantity of the offset voltage is inversely proportional to the square root of the transistor size. However, large-size devices limit the bandwidth of an amplifier and consume larger power. The resistive averaging network is a popular technique for offset reduction. The averaging technique neutralizes random offsets by connecting adjacent preamplifiers together. The output of an averaged preamplifier is not only defined by itself but also by all the preamplifiers nearby. The resistive averaging network has been proven a useful technique for linearity enhancement but requires extra dummy preamplifiers to maintain boundary conditions [2]. As a result, there is a tradeoff between power consumption and averaging efficiency. Calibration techniques are alternative solutions for offset reduction [3]–[7]. In addition to an ADC core, extra circuits are introduced to perform calibration for preamplifiers or comparators.

Digital-to-analog converters (DACs) are utilized in a 4-b 4-GS/s Flash ADC to trim the output voltages of the comparators [4]. However, these DACs induce extra capacitive loading to the comparators. A 4-b 1.25-GS/s Flash ADC uses programmable capacitor arrays to perform offset calibration [5]. The capacitor arrays also contribute extra loading to the comparators. A 5-b 3.5-GS/s Flash ADC directly trims its reference voltages to cancel the offset [6]. A background offset calibration scheme is proposed in a 6-b 1-GS/s two-step ADC [7]. An auxiliary differential pair helps reduce the offset of the main differential pair. This calibration technique requires a reset phase and is therefore not suitable for conventional continuous-time Flash ADCs. This paper proposes a foreground digital calibration scheme. By adjusting the programmable resistive loading devices of the preamplifiers, the random offsets are reduced.

Manuscript received August 25, 2008; revised November 18, 2008. First published May 29, 2009; current version published February 24, 2010. This work was supported by NSC-97-2221-E-006-294-MY2.

The authors are with the Department of Electrical Engineering, National Cheng-Kung University, Tainan 70101, Taiwan (e-mail: tibrius@sscas.ee.ncku.edu.tw; lcw@sscas.ee.ncku.edu.tw; soon@mail.ncku.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2009.2013628

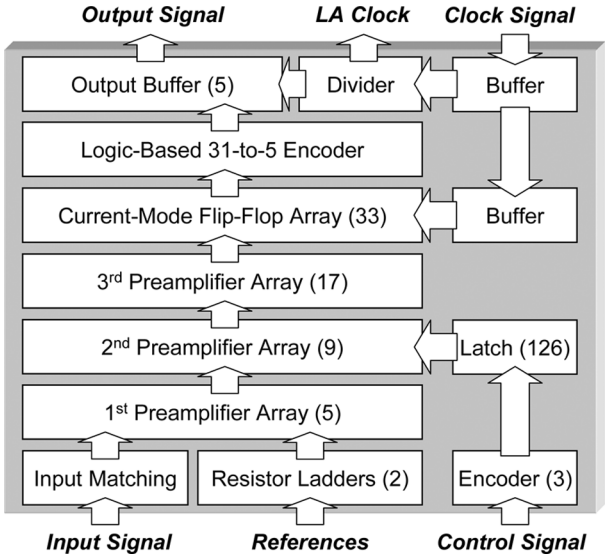

Fig. 1. Simplified block diagram of the proposed ADC.

II. ADC Architecture and Building Blocks

Fig. 1 shows the simplified block diagram of the proposed ADC where *LA Clock* means the trigger signal for the logic analyzer. The ADC architecture mainly follows the design in [8]. This ADC core consists of two reference ladders, three preamplifier arrays, a current-mode flip-flop array, a logic-based 31-to-5 encoder, and digital circuits for offset calibration. For impedance matching, a 100-Ω resistor is connected to the differential input ports to make the amplitude of the input signal stable in the desired frequency range. For measurement, output data of the encoder are sampled by flip-flops clocked at 1/32 sampling frequency.