社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
DSP2003-135, ICD2003-133, IE2003-95(2003-10)

高リーク環境における Self-Timed Cut-Off 法を利用した統計的なリーク電流削減手法

許　蛍雪　　崔　珍赫　　宮崎　隆行　　川口　博　　桜井　貴康

東京大学生産技術研究所　〒153-8505　東京都目黒区駒場 4-6-1

E-mail:　{xyx, jinhchoi, tmiyazak, kawapy, tsakurai}@iis.u-tokyo.ac.jp

あらまし　高リーク世代において、Self-Timed Cut-Off に基づくブロックレベル活性化法は、待機時のみならず動作時においてもリーク電流を減らすことができ有効である。その基本動作は統計情報に基づいている。Self-Timed Cut-Off スイッチを使い、一定時間動いてないブロックをスリープモードに入れる。本提案の有効性を 8 ビット RISC マイクロプロセッサの Verilog HDL によるシミュレーションとともに、0.25um の SOI プロセスで試作された 64 ビットの carry look-head 加算器の測定を通して検証した。

キーワード　高リーク電流，低電力，ブロック活性化，統計的

Statistical Leakage Current Reduction by Self-Timed Cut-Off Scheme for High Leakage Environments

Yingxue Xu　Jin-Hyeok Choi　Takayuki Miyazaki　Hiroshi Kawaguchi　and　Takayasu Sakurai

Institute of Industrial Science, University of Tokyo　4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505 Japan

E-mail:　{xyx, jinhchoi, tmiyazak, kawapy, tsakurai}@iis.u-tokyo.ac.jp

Abstract　This paper describes a statistical leakage current reduction scheme that can reduce leakage current even if the chip is in an active mode. The scheme utilizes a self-timed cut-off switch that puts a given block into a sleep mode if the block is not used for a certain number of cycles. The effectiveness of the proposed scheme is verified by an 8-bit RISC microprocessor using Verilog HDL, and demonstrated by a 64bit carry look-ahead adder fabricated with dual-V_{TH} SOI technology.

Keyword　High Leakage Current, Low Power, Block Activation, Statistical

1. Introduction

As semiconductor technology scales down, the leakage power increases exponentially because the threshold voltage is reduced to compensate for the performance degradation caused by the reduced supply voltage to guarantee the device reliability. It is anticipated that the leakage power will become comparable to the dynamic power in near future [1], and will become the dominant component in active power [2]. This means that the reduction of leakage is getting more and more important. In order to reduce the sleep leakage current, power-down techniques by entering a sleep mode using body bias applying [3-4] and Multi-Threshold CMOS (MTCMOS) [5-8] have been proposed and extensively investigated. Other than the power-aware design for the sleep mode just mentioned, an adaptive supply voltage scheme is reported [9] for active dynamic power reduction. However, only little attention has been given to handle the leakage current in an active mode.

This paper presents a new switching methodology that can give a statistical leakage current reduction even if the chip is in an active mode as well as in a sleep mode. One of the major obstacles to put a given block to a sleep mode is long wake-up time. Recently, a new cut-off switching scheme called Zigzag Super Cutoff scheme (ZSCCMOS) is proposed to shorten the wake-up time by almost an order of magnitude [10], so that it can make the block usable in the next clock cycle. The biggest issue in the leakage cut-off scheme by inserting the cut-off switch is that the sleep and the wake-up processes need dynamic power to turn on and off the cut-off switch so that frequent activation and de-activation of a block will rather increase the total energy although leakage power may be reduced. If the block is to be activated very soon, it is wise not to sleep. The bad news is that we cannot predict the future but the good news is that we found that there is a statistical behavior in the sleep duration. Thus

the proposed approach set a certain number of cycles before the sleep by using a self-timed circuit.

The feasibility of the proposed method is verified by an 8-bit RISC microprocessor using Verilog Hardware Description Language (HDL). It is proved that the proposed leakage reduction scheme can give a leakage saving in an active mode by using block activation data of real applications and simulations. The proposed scheme is also demonstrated to be effective by the leakage measurement of a 64-bit carry look-ahead adder (CLA) with self-timed cut-off switch, which is fabricated with dual-V_{TH} SOI technology.

2. Concept of Self-Timed Cut-off Scheme

The concept of the proposed method is based on the fact that not all the circuit blocks on a chip are operating even though the chip is in an active mode. For example, when a CPU is doing a data transfer from on-chip memory to external equipments, a shifter and multiplier blocks are not working but still consuming leakage energy. One more important observation is that the block activation probability is closely related to the previous activation history like cache memory hit/miss characteristic. The activation of a certain block is burst-like and has some locality in time.

A block diagram of the proposed self-timed cut-off scheme is shown in Fig. 1. Any sleep switching structures, such as MTCMOS, VTCMOS, BGMOS and ZSCCMOS, can be applicable with the proposed switching scheme. The self-timed cut-off switch makes the block go to the low leakage sleep mode by turning-off the switch if condition of entering sleep mode was met. The details of the self-timed cut-off switch are discussed later in Fig. 8.

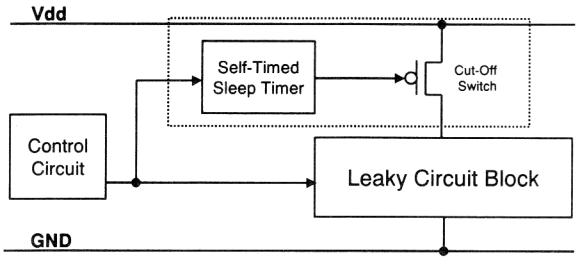


Fig. 1. Block diagram of the self-timed cut-off switch operation applied to MTCMOS structure.

Figure 2 depicts the consumed energy behaviors by leakage current with two different cut-off time intervals, which sets the basis of this paper. The energy overhead associated with the sleep and wake-up processes is shown in the figure. The energy consumptions associated with the turning-off and turning-on of the block are expressed as $E_{turn\text{-}off}$ and $E_{turn\text{-}on}$, respectively. The energy saved by turning-off a given block for a duration of cut-off time, can be expressed as $(P_{lkg}\text{-}P_{sleep})T_{sleep}$, where P_{lkg} and P_{sleep} are leakage power in an active and sleep mode. The cut-off time of A can be defined as *leakage energy equivalent time*, T_{lkg_eqv}, which means the active energy consumed by the switching is the same with the leakage energy saved by cutting off the block for that time. If a circuit block under a sleep mode is woken up after T_{lkg_eqv}, energy is saved by cutting off the block. Otherwise, it consumes energy by toggling of switch.

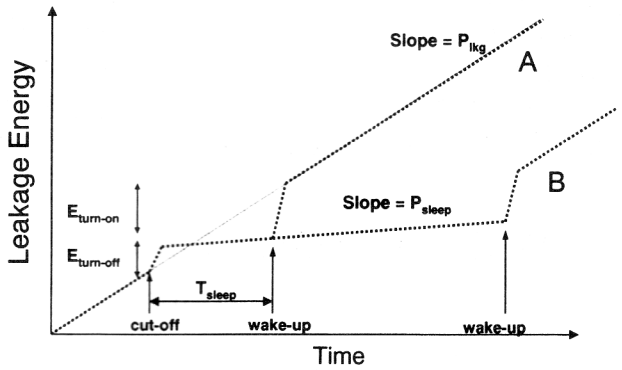


Fig.2. Consumed energy behaviors by leakage current and turn-on/off the switch with two different cut-off time intervals.

Figure 3 shows the simulation results of the 64-bit carry look-ahead adder (CLA) with various sleep durations. Zigzag cut-off scheme is assumed. The 100nm MOS model parameters by Berkeley Predictive Technology Model are used with modified V_{TH}. The simulated results show that the turning-on/off of 1.2 ns interval causes energy consumption increase, while that of 20 ns interval achieves energy saving. The leakage energy equivalent time in this condition is 9.5 ns. This simulation results indicate that the turning-off of a given block can save the leakage energy if that block will not be re-activated within T_{lkg_eqv}. At least, if the probability of the re-activate within T_{lkg_eqv} is less than 0.5, the system can save leakage energy statistically. Figure 4 shows the trend of T_{lkg_eqv} with the decrease of V_{TH} for 100nm and 70nm model parameters. Exponential decrease of the T_{lkg_eqv} can be seen with the decreases of threshold voltage and technology scaling.

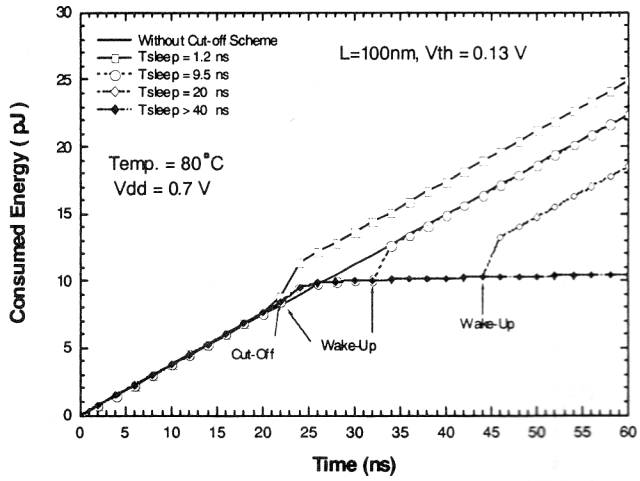


Fig. 3. Simulation results of the 64-bit carry look-ahead adder with respect to various sleep times.

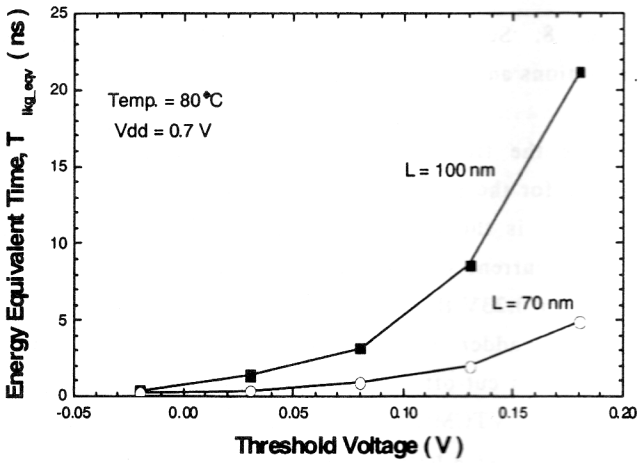


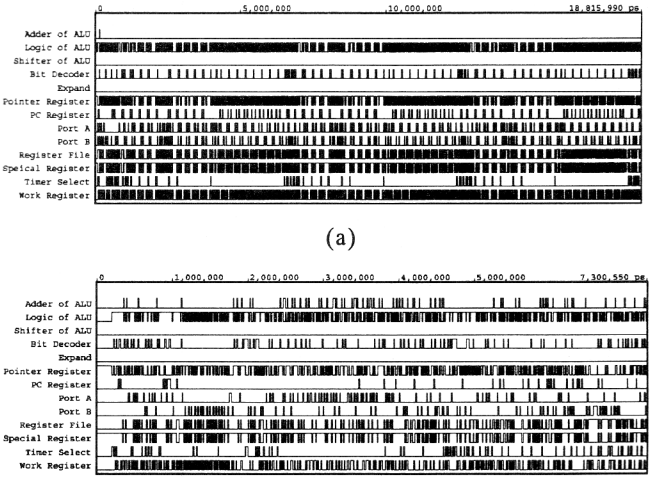
Fig. 4. Estimated leakage energy equivalent times with the variation of threshold voltage and channel length.

3. Implementation of Proposed Scheme

3.1. Block Control Signal

A block enable signal is needed for each block to realize the proposed scheme but the preparation of the clock enable signal is an easy part because basically the block enable signal is the same as a clock gating signal which is widely used currently. An 8-bit RISC CPU is described using Verilog HDL to verify whether block activation signals are efficiently generated for each block in the CPU. Verilog HDL is also used to analyze distributions of block activation duration and sleep duration in real applications. Real application programs for a LCD driver (A) and a toy controller are run (B).

Figure 5 shows the behaviors of the important block enable signals for two different applications. The adder in ALU is hardly working in the application A, while operating frequently in application B. The shifter in ALU does not work at all in two applications. This real application results indicate that even if the chip is in an active mode, some parts of the chip are not working or hardly working, and still consume current by leakage.



(a)

(b)

Fig.5. Block enable signals for two different application of 8-bit RISC microprocessor. (a) Application A: LCD driving, (b) Application B: Toy control.

3.2. Probability Distributions

Distributions of activation interval are needed to verify that the proposed scheme can save the leakage power statistically. Figure 6 shows the distributions for the two applications. Event interval signifies the number of clocks between two adjacent block activations. The extracted statistical results show that the operation probability of a block is strongly dependent on previous event, like cache memory behavior. In other words, it is more probable for a block to be operated if the block is activated recently. The criterion to enter the block sleep mode is important to achieve the statistical energy saving effectively. Let's define $p(T_{lkg\_eqv} \mid T_{lkg\_eqv})$ as the conditional probability of activation within T_{lkg_eqv} on condition that the activation signal has not been issued for T_{wait}. It can be written as

$$p(T_{lkg\_eqv} \mid T_{wait}) \equiv \frac{p[T_{wait} \le t \le (T_{wait} + T_{lkg\_eqv})]}{1 - p(0 \le t \le T_{wait})} \qquad (1)$$

where $P(t)$ means the probability of block activation within time interval of t, and T_{wait} stands for the time interval between sleep mode and latest block activation.

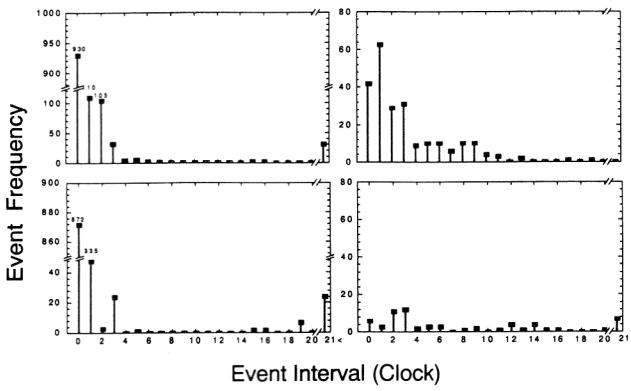


Fig. 6. Extracted frequency of event as a function of event interval (the number of clock between two adjacent commands of activation). A and B mean two applications shown in Fig.5.

Figure 7 shows the $p(T_{lkg\_eqv} \mid T_{wait})$ as a function of waiting time for an adder and a registerfile with two different T_{lkg_eqv}. $p(10|5)$ is 2% for the adder unit and 8% for the register file. This means when the adder and the register file enter the sleep mode after 5-clock waiting time, the blocks will be reactivated within T_{lkg_eqv} with the probability of 2% and 8%. The activation probability decreases when decreasing T_{lkg_eqv} since time window of $p(T_{lkg\_eqv} \mid T_{wait})$ is shortened in (1). This implies that the reduction of leakage power by the proposed scheme is getting more probable as technology advances.

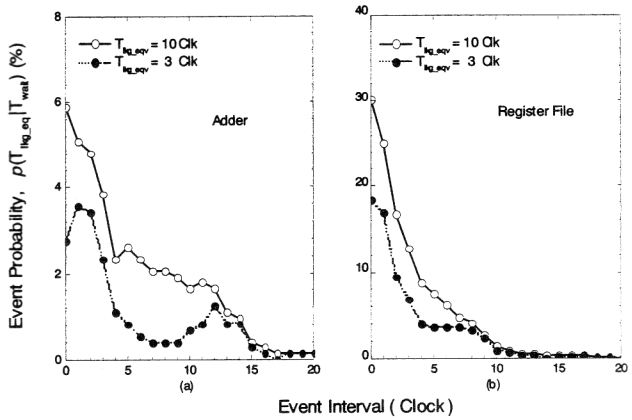


Fig. 7. Activation probability as a function of waiting time; (a)Adder, (b)Register File of application B.

3.3. Simulation Results

Self-timed cut-off controller used in the simulation and experiments is shown in Fig. 8. The cut-off controller consists of a leakage integrator as a timer, a gated clock to suppress the conventional dynamic power by the clock, and one flip-flop to synchronize the sleep signal to the clock pulse and eliminate a glitch at the transition. Since entering the sleep mode is not a timing critical sequence, high V_{TH} is used to minimize the leakage except the leakage-monitor PMOS and the capacitor which acts as a timer.

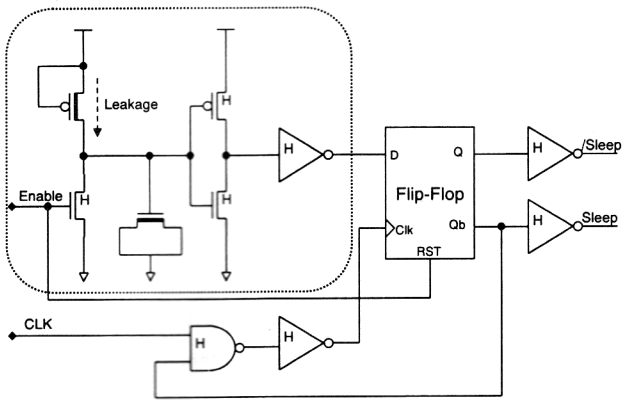


Fig. 8. Self-timed cut-off controller used in the simulations and experiments.

Since the flip-flop and the NAND gate are needed anyway for the gated-clock, pure design overhead of the controller is the circuits in dotted area in Fig. 8. The leakage current caused by this controller is less than 0.1μA for 0.23V threshold voltage, i.e. less than 0.1% of the total adder leakage. Of course other than this controller, a cut-off switch itself is needed depending on MTCMOS, VTCMOS, ZSCCMOS and other schemes. Figure 9 shows a part of simulated leakage behavior of a 64-bit CLA with self cut-off controller shown in Fig. 8. Waiting time of 27ns and 5ns are assumed. The enable signal pattern of the self-timed cut-off controller is taken by the Verilog simulation using the program of application B. Simulation result shows that the sleep signal generated by the self-timed cut-off controller is effective. Re-activation of the CLA just after entering a sleep mode also can be seen, but statistically speaking, the average energy consumption is decreased.

Figure 10 shows average energy consumption as a function of waiting time before sleep for 100nm and 70nm generation. The proposed scheme is very effective in active energy reduction for high leakage environments, but the efficiency is decreased with the leakage decrease. It is clearly seen that there is an optimum point for the waiting time before sleep. The criterion of the effectiveness of the proposed switch scheme is derived by the fact that the leakage energy for one clock (E_{lkg}) multiplied by average number of clock interval (T_{av}) should be larger than activation energies of the block

(E_{act}) and switch (E_{swact}) to achieve a benefit of energy saving by cut-off switch. On the other hand, if E_{lkg} is larger than $E_{act}+E_{swact}$, the proposed self-timed cut-off scheme is not needed because the block can be turned off every non-operating condition without the self-timed controller.

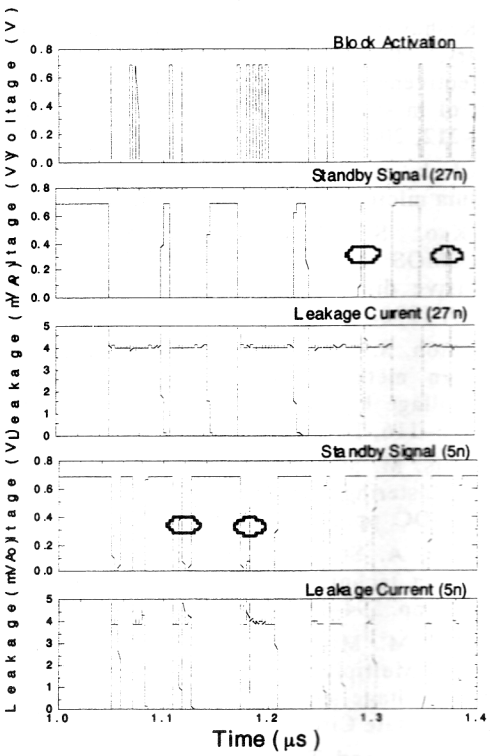


Fig. 9. Simulated leakage current. Time conditions marked by circle lose energy, the others save energy.

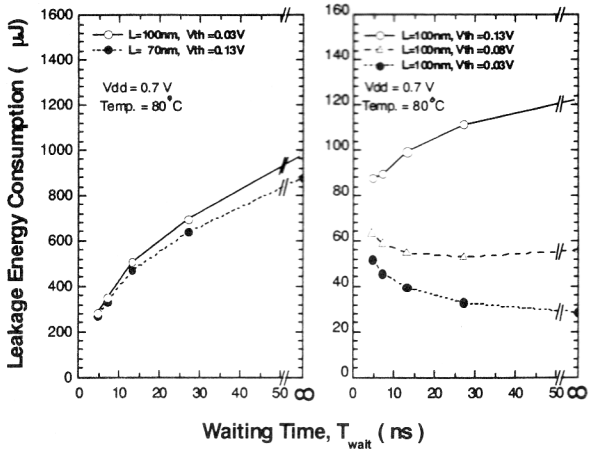


Fig. 10. Simulated leakage energy consumption of 16-bit CLA for application B with various device conditions. (1500ns duration).

Let's define $\eta=E_{lkg}/E_{act}$ that represents leakage energy ratio of the block with respect to active energy, and $a=E_{swact}/E_{act}$ that means the overhead of active energy caused by the controller. The leakage criterion (η) can be expressed as

$$\frac{1+\alpha}{T_{av}} \le \eta < 1+\alpha \qquad (2)$$

3.4. Waking-Up within One Clock

If a block enable signal is asserted in a sleep mode, the circuit block should wake up without delay. Recently, Zigzag Super Cutoff CMOS (ZSCCMOS) is proposed to shorten the waking-up time [10]. In this paper, however, PMOS and NMOS cut-off switches have the higher V_{TH} instead of the super cut-off scheme originally proposed. Simulated wake-up time of the 64-bit carry look-ahead adder is 0.60ns for 100nm device technology and the add operation can be executed just after the wake-up cycle.

3.5. Experimental Results

A test chip is fabricated using 0.25μm dual-V_{TH} SOI technology. A microphotograph of the fabricated chip is shown in Fig. 12. The test chip consists of a 64-bit carry look-ahead adder of MTCMOS structure whose gate is controlled by a built-in self-timed cut-off controller. High and low threshold voltages of NMOS are 0.15V and 0V, and those for PMOS are –0.15V and –0V, respectively. The layout size of the timer is 1% of the 64-bit MTCMOS full adder, but the total layout size is not increased since it is placed under the interconnections and power line as shown in Fig. 11. The block enable signal is generated by a data generator. The gate bias of the PMOS leakage-monitor in Fig. 8 is changed to modify the waiting time for the measurement purpose.

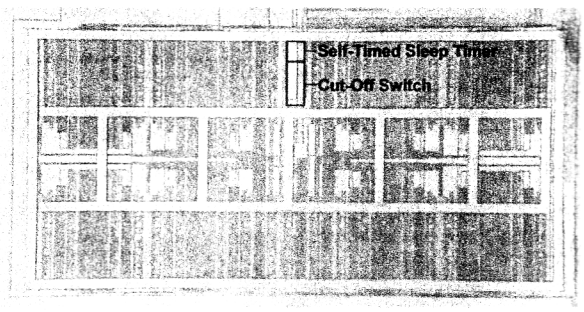


Fig. 11. Microphotograph of fabricated test chip of 64-bit CLA adder with self-timed cut-off switch.

Figure 12 shows measured average energy consumption as a function of event interval with respect to various waiting times, T_{wait}. The measured result indicates that 3-clock wait time with 20-clock activation interval saves

leakage by 64% with proposed method while that with 4-clock activation interval shows energy consuming. The measured results demonstrate that the self-timed cut-off controller can optimize the total energy consumption by tuning the waiting time before sleep. If event interval is distributed, the statistical energy saving can be determined by (2). In reality, the various sized PMOS like size 1, 2, 4 and 8 units can be placed in parallel which are switch on and off according to the desired waiting time before sleep.

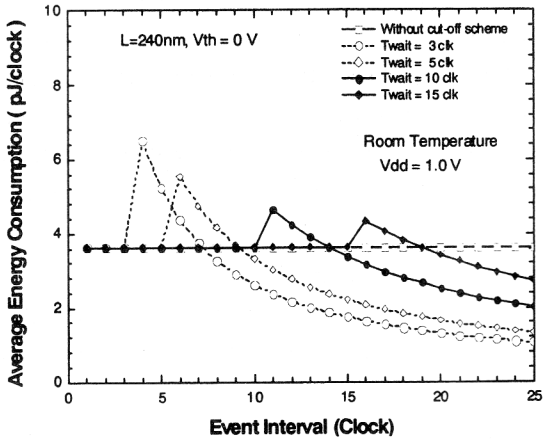


Fig. 12. Measured average energy consumption as a function of event interval with respect to various waiting times.

4. Conclusion

A new leakage reduction scheme with self-timed cut-off controller is proposed as an effective leakage power reduction approach in an active mode as well as in a sleep mode. The proposed scheme gives statistical leakage reduction by adding waiting time before sleep. The feasibility and effectiveness of the proposed scheme are verified by an 8-bit RISC microprocessor using Verilog HDL. The simulated data shows a 75% active leakage reduction in 70nm channel length with 0.13V threshold voltage for an adder unit. Measured power saving of a 64-bit carry look-ahead adder fabricated on 0.25μm SOI also proves that the proposed method is efficient to optimize the total energy consumption in an active mode.

5. Acknowledgment

The authors would thank to New Energy and Industrial Technology Department Organization for chip fabrication.

References

[1] G. Moore, "No exponential is forever: but forever can be delayed," Tech. Digest of ISSCC, pp.20–23, 2003.

[2] T. Sakurai, "Perspectives on Power-aware electronics," Tech. Digest of ISSCC, pp.26–29, 2003.

[3] A. Keshavarzi, S. Ma, S. Narenda, B. Bloechel, K. Mistry, T. Ghani, S. Borkar, and V. De, "Effectiveness of reverse body bias for leakage control in scaled dual Vt CMOS ICs," ISLPED, pp. 207–212, 2001.

[4] L. Clark, et al., "Sleep power management for a 0.18um microprocessor," ISLPED, pp. 7–12, 2002.

[5] J. Kao, S. Narenda, and A. Chandrakasan, "MTCMOS hierarchical sizing based on mutual exclusive discharge patterns," Proc. of DAC., pp. 15–19, 1998.

[6] S. Mutoh, S. Shigematsu, Y. Gotoh, and S. Konaka, "Design method of MTCMOS power switch for low-voltage high-speed LSIs," Proc. of ASP-DAC., pp. 113–116, 1999.

[7] M. Anis, M. Mahmoud, and M. Elmasry, "Efficient gate clustering for MTCMOS circuits," Proc. of ASIC/SOC, pp. 34–38, 2001.

[8] P. Meer, A. Staveren, "Effectivity of sleep-energy reduction techniques for deep sub-micron CMOS," ISCAS, pp. 594–597, 2001.

[9] J. Kao, M. Miyazaki, and A. Chandrakasan, "A 175-mV Multiply-accumulate unit using an adaptive supply voltage and body bias architecture," IEEE J. of Solid State Circuits., pp. 1545–1554, 2002.

[10] K. S. Min, and T. Sakurai, "Zigzag Super Cut-off CMOS (ZSCCMOS) block activation with self-adaptive voltage level controller: An alternative to clock-gating scheme in leakage dominant era," in Tech. Digest of ISSCC, pp. 400–401, 2003.