Boosted Gate MOS (BGMOS): Device/Circuit Cooperation Scheme to Achieve Leakage-Free Giga-Scale Integration

T. Inukai[1], M. Takamiya[1], K. Nose[1], H. Kawaguchi[1], T. Hiramoto[1,2] and T. Sakurai[1,3]

[1]Institute of Industrial Science, the University of Tokyo, Roppongi, Minato-ku, Tokyo 106-8558, Japan
Phone: +81-3-3402-6231(ex.2337), Fax: +81-3-5411-0695, E-mail: inukai@nano.iis.u-tokyo.ac.jp
[2]VLSI Design and Education Center, the University of Tokyo, Hongo, Bunkyo-ku, Tokyo 106-8656, Japan
[3]Center for Collaborative Research, the University of Tokyo, Roppongi, Minato-ku, Tokyo 106-8558, Japan

Abstract

This paper proposes a new device and circuit scheme that drastically suppresses the stand-by leakage current for the deep sub-0.1μm era while maintaining the circuit speed. Applying boosted gate voltage on the low leakage switches with higher V_{th} and thicker T_{ox}, extremely low stand-by power for battery type application is achieved, while degradation of circuit performance and an increase of area overhead are sufficiently suppressed. The combination with a negative gate voltage scheme and the application of the boosted voltage scheme to SRAMs are also discussed.

Introduction

The performance of VLSI chips will further increase by continuous device scaling. However, the present device scaling method will certainly confront the practical limits caused by increased stand-by power. In the deep sub-0.1 μm generation, the MOS device is no longer an ideal device and the on/off ratio would be severely degraded. The off-current will consist of (i) subthreshold current due to very low threshold voltage (V_{th}) that is required by low supply voltage (V_{dd}), (ii) gate direct tunnel current due to very thin gate oxide (T_{ox}), and (iii) band-to-band tunnel leakage current due to heavily-doped halo. Therefore, stand-by current rapidly increases as the device is scaled. Fig. 1 shows the variation of leakage current with device scaling. The leakage current will become far above the requirement for battery-operated chips.

Several proposals have been reported to suppress the stand-by power, including VTCMOS [1] and MTCMOS [2]. In VTCMOS, however, gate tunnel current cannot be suppressed and junction leakage becomes even worse by substrate bias [3]. In MTCMOS, gate tunnel current and junction current cannot be suppressed and area penalty rapidly increases as the supply voltage is scaled. On the other hand, gate dielectrics with high k will suppress the gate tunnel current, but the subthreshold and junction leakage can not be reduced. The stand-by power problem will not be solved by the conventional device and circuit scheme. Moreover, no solutions have been proposed for stand-by leakage of SRAMs.

In this paper, we propose a new device/circuit cooperation scheme, the boosted gate MOS (BGMOS) scheme, that drastically suppresses the stand-by current while maintaining circuit performance. The effectiveness of the BGMOS scheme is confirmed by means of device simulation as well as circuit simulation. This scheme can be applied to the whole circuits in the chip including flip-flop, SRAMs, and I/O circuits. This scheme would be the only solution, ever reported, that achieves high speed and low stand-by current at the same time.

Table 1 List of the device parameters for simulations[4].

Tech. Node	180nm	130nm	100nm	70nm	50nm
L_g (nm)	140	100	70	50	35
V_{dd} (V)	1.8	1.5	1.2	0.9	0.6
T_{ox} (nm)	3.5	3	2	1.5	1
X_j (nm)	50	35	30	20	15
$N_{s/d}$ (cm^{-3})	1x10^{19}	5x10^{19}	1x10^{20}	1x10^{20}	1x10^{20}
Side Wall Width (nm)	100	70	30	10	7

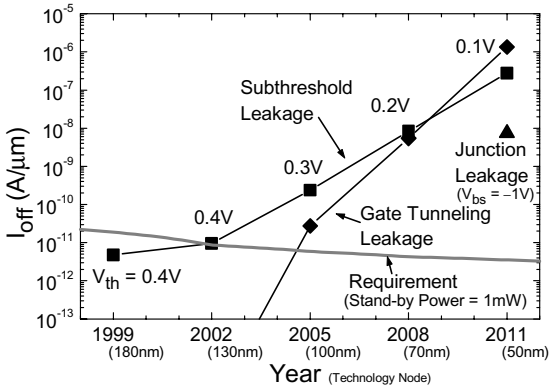

Fig. 1 The trend of off-current of MOSFETs. Three types of off-current will become problematic: (i) subthreshold leakage, (ii) gate tunneling leakage and (iii) junction leakage due to band-to-band tunneling with substrate bias. Parameters assumed are listed in Table 1.

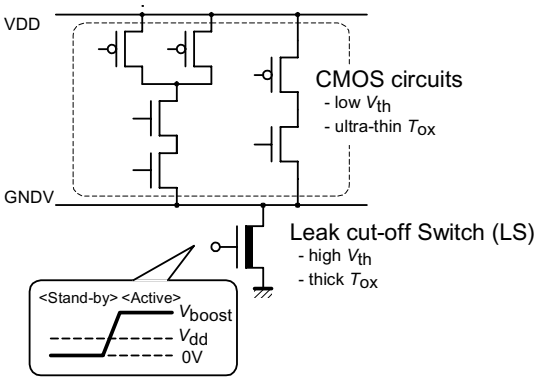

Fig. 2 A schematic of the proposed BGMOS scheme. Logic circuits consist of high performance MOSFETs with ultra-thin T_{ox} and low V_{th}. In stand-by mode, leak cut-off switches (LSs), whose T_{ox} is thick enough, V_{th} is high, and off-current is extremely low, suppress the stand-by current sufficiently. On the other hand, in active mode, LSs are driven by boosted gate voltage and the area penalty is well suppressed.

Boosted Gate MOS (BGMOS)

In order to solve the stand-by power issue with maintaining the circuit speed, two types of devices are necessary. One is an ultra-high speed, large leakage device with low V_{th} and thin T_{ox} that follows the roadmap [4] and is used for logic circuits. The other is an extremely small stand-by current device with high V_{th} and thick T_{ox} that is used for the leak cut-off switches (LSs), SRAM cells, and I/O circuits. As the supply voltage is reduced, however, the drive current of such low leakage switches is extremely low, which causes an increase of area overhead if used for the LSs. Therefore, the improvement of the drive current of LSs is essential for the suppression of the area overhead. The proposed device scheme makes it possible by applying higher gate voltage (V_{boost}) than V_{dd} in the active mode.

Fig. 2 shows a schematic of the proposed BGMOS scheme. Logic circuits consist of high-performance devices with ultra-thin T_{ox} and aggressively low V_{th}. The supply voltage is reduced to suppress the ac power. On the other hand, LSs consist of leakage-free devices with high V_{th} and thick T_{ox}. LSs are inserted in series and the stand-by power is completely suppressed. The gate voltage of LS is boosted to reduce the area penalty.

Design of Leak Cut-off Switches

The optimal condition for LS, which acts as a leakage-free device, is discussed. With thick T_{ox}, the gate voltage can be boosted. However, LS with thicker T_{ox} cannot suppress short channel effect sufficiently and needs longer L_g. To take the short channel effects and other device issues into consideration, two dimensional device simulation was performed. Fig. 3 shows the drain conductance in the linear region and required L_g for suppression of short channel effect in the 50nm technology node. The condition is constant V_{th} lowering by DIBL of 50mV. Boosted gate voltage is varied proportional to T_{ox} , assuming constant gate electric field (5MV/cm). With boosted gate voltage, drain conductance of LS is drastically improved without sacrificing gate oxide reliability. The optimal T_{ox} and boosted-gate voltage are found to be 4nm and 2V respectively.

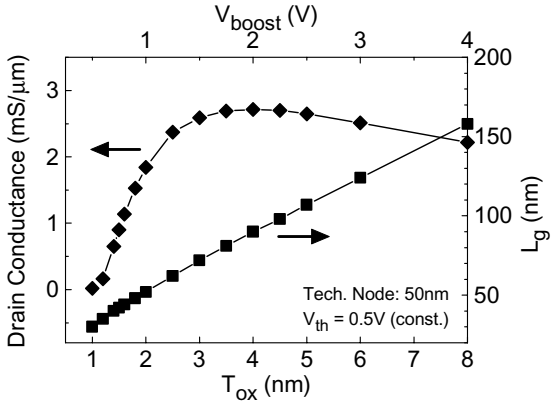

Fig. 3 Drain conductance of LS in the linear region. Boosted gate voltage is proportional to oxide thickness (T_{ox}) and gate length is derived from the condition of constant short channel effect (DIBL). Drain conductance has its maximum value when T_{ox} is about 4nm. Moreover, this optimal condition is applicable to other technology nodes (not shown).

SPICE Simulation

The effect of boosted gate and the area penalty of LS are discussed by SPICE simulation, where the SPICE parameters are derived from the results of device simulation. Fig. 4 shows the dependence of active performance of 2-way NAND on the gate width of LS. When the gate is not boosted (conventional MTCMOS), the voltage drop across LS is large and logic performance is degraded, which means that the conventional scheme needs extremely large area overhead. Fig. 5 compares the area overheads. Using boosted gate, area penalty is kept less than 10 % in each technology node, while the area penalty rapidly increases without boosted gate. On the other hand, Fig. 6 shows the trend of stand-by power of a chip. The BGMOS scheme also drastically reduces the stand-by current. This low stand-by current cannot be achieved by conventional CMOS without LS or conventional MTCMOS.

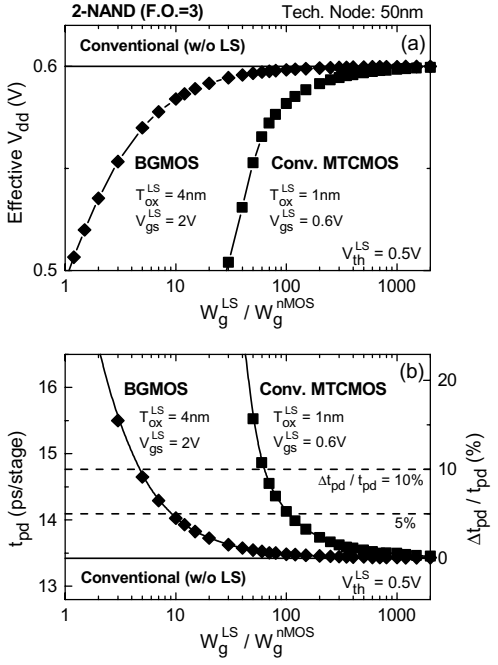

Fig. 4 The active performance of proposed BGMOS compared with conventional MTCMOS. (a) effective supply voltage and (b) propagation delay time. In conventional MTCMOS (without boosted gate), LS conductance is too small and extraordinarily large gate width is required.

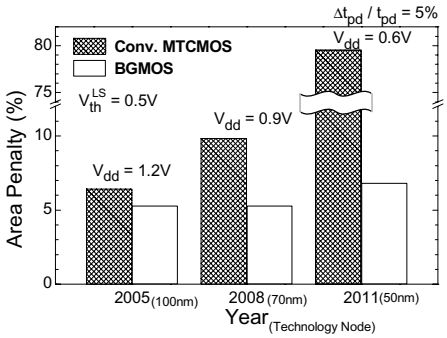

Fig. 5 The trend of area penalty due to the insertion of LS for constant Δtpd. While the area penalty in conventional MTCMOS is rapidly increasing with technology node, the area penalty in the proposed BGMOS is increasing slightly.

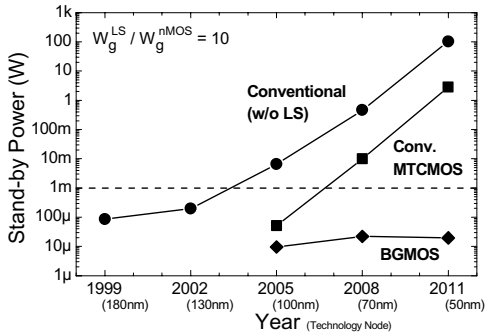

Fig. 6 The variation of stand-by power at constant area penalty. Stand-by power will increase and become intolerable for battery type operation in conventional CMOS (without LS) and conventional MTCMOS. On the other hand, proposed BGMOS can suppress tunneling leakage and subthreshold leakage, thus reducing the stand-by power efficiently.

Combination with Super Cut-off Scheme

Another solution to the degradation of circuit performance and increase of area overhead might be super cut-off CMOS (SCCMOS) scheme [5]. In SCCMOS, LSs consist of low V_{th} devices and the gate is negatively biased in stand-by mode in order to suppress stand-by subthreshold leakage as low as high V_{th} devices. With low V_{th}, LSs can get higher drive current in active mode and the degradation of circuit performance is suppressed well. However, applying negative bias in stand-by mode causes higher voltage across gate oxide resulting in the reliability problem.

BGMOS can solve this problem. Using thicker T_{ox} for LSs in SCCMOS, the oxide reliability is improved and boosted gate voltage in active mode becomes possible. Fig. 7 shows the effect of this combined scheme and bias conditions in stand-by and active modes. Applying gate voltage to the limit determined from oxide reliability in both modes, the drain conductance of LSs is maximized. While BGMOS uses the maximized voltage only in active mode, SCCMOS uses it only in stand-by mode. Therefore, the combined scheme enables further suppression of area overhead, compared with BGMOS and SCCMOS, which is confirmed by SPICE simulation as shown in Fig. 8.

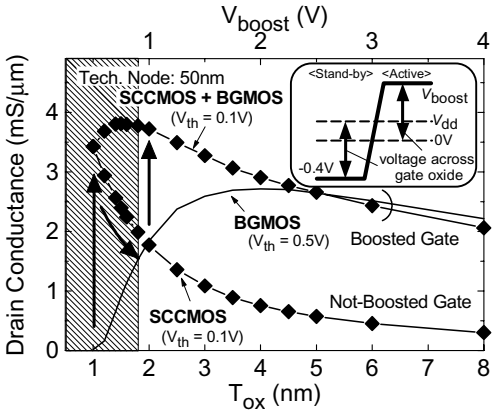

Fig. 7 Drain conductance of LS in SCCMOS combined with BGMOS scheme. Using SCCMOS scheme, drain conductance is drastically improved. However voltage across gate oxide in stand-by mode is over V_{dd} and oxide reliability becomes an issue (grayed region). BGMOS scheme does not only solve this problem but also enable further improvement of drain conductance of LS.

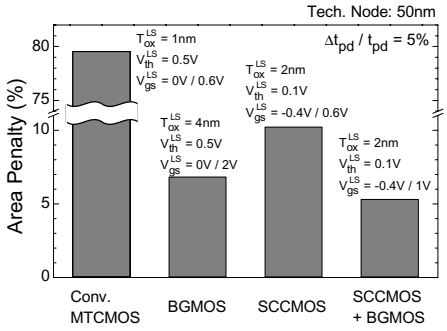

Fig. 8 Comparison of area penalty for various circuits schemes. SCCMOS scheme needs thicker T_{ox} for oxide reliability, which makes its area penalty a little larger than BGMOS's. On the other hand, SCCMOS combined with BGMOS scheme shows the smallest area penalty.

Application to SRAMs and F/F

For SRAM cells, leak cut-off switch scheme cannot be used because the stored data are lost in the stand-by mode. To reduce the stand-by power, leakage free devices with high V_{dd} and high V_{th} (equal to V_{boost}) should be used in SRAMs. On the other hand, the data in flip-flop (F/F) circuits, which consist of high performance devices and connect with LS, would be lost in stand-by mode. This problem can be solved by the circuits that combine low V_{dd} F/F and high V_{dd} SRAM, as shown in Fig. 9, where stored data in F/F are transferred to high V_{dd} SRAMs and stored temporally in stand-by mode [5]. Moreover, this circuit scheme is applicable to I/O circuits that are conventionally consisted of devices with high V_{dd} for compatibility with peripheral equipments. Fig. 10 shows a schematic of the proposed circuit scheme.

The device for the proposed SRAM cells can be common to that for LS, except that high V_{dd} is applied to drain in SRAM. Therefore, L_g of SRAM MOSFET cannot be so much scaled as conventional devices because of large DIBL, as shown in Fig. 11. On the other hand, Fig. 12 shows the speed (V_{dd} / I_{on}) of the proposed SRAM MOSFET compared with the conventional SRAM MOSFET assuming constant V_{th} and constant I_{on} respectively. While the proposed device is much faster than the conventional device with constant V_{th}, it is slower than conventional device with constant I_{on}. This degradation of speed is due to longer L_g and higher V_{dd} in the proposed SRAM. Therefore, some measures should be taken to avoid the speed degradation of SRAM. Dividing macrocells into small macros is one of the good examples. The area penalty of the proposed SRAM cells increases with technology node as shown in Fig. 13. However, when we take T_{ox} = 3 nm (hence V_{dd} = 1.5 V), for both LSs and SRAMs, the total area penalty is suppressed. Although the area penalty increases, this device and circuit scheme would be one of the best solutions for the suppression of stand-by power in SRAMs.

Conclusion

We have proposed a new device and circuit scheme for sub-0.1µm generation to achieve extremely small stand-by power without degrading circuit performance. This scheme would be only solution ever reported to several problems, which conventional schemes will face, that is, stand-by power, speed degradation and area overhead.

Acknowledgment

The authors thank T. Kuroda of Toshiba corp. for fruitful discussion. This work is partly supported by the JSPS Research for the Future Program.

References

[1] T. Kuroda, T. Fujita, S, Mita, T. Nagamatsu, S. Yoshioka, K. Suzuki, F. Sano, M. Norishima, M. Murota, M. Kako, M. Kinugawa, M. Kakumu, and T. Sakurai, "A 0.9V, 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, 1996.

[2] S. Mutoh, T. Douseki, Y. Matsuya, T. Aoki, S. Shigematsu, and J. Yamada, "1-V power supply high-speed digital circuit technology with multithreshold-volgate CMOS", *IEEE Journal of Solid-State Circuits*, vol. 30, pp. 847-854, 1995.

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

[4] International Technology Roadmap for Semiconductors, 1998 update.

[5] H. Kawaguchi, K. Nose, and T. Sakurai, "A CMOS scheme for 0.5V supply voltage with pico-ampere standby current", *ISSCC Digest of Technical Papers*, pp. 192-193, 1998.

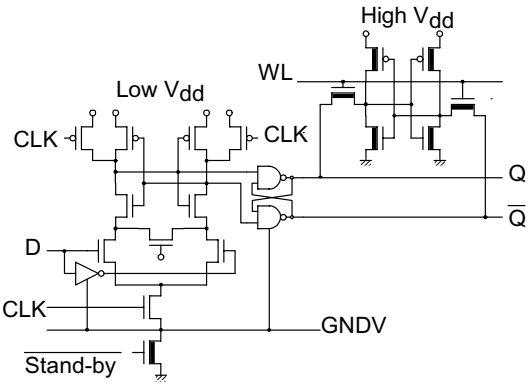

Fig. 9 A flip-flop (F/F) is an important element of logic circuits. If LS is inserted in series with the conventional F/F and the switch turns off, the conventional F/F's lose stored information. A F/F proposed in this figure consists of F/F with LS in series and an SRAM cell with the higher V_{th} and V_{dd}. The information stored in the F/F is written to the SRAM by activating the WL when standby signal is asserted. In the wakeup process, the information in the SRAM is written back again to the F/F and thus the information is not lost. The delay overhead is 5%, since the speed is determined by the F/F with the BGMOS scheme. The area overhead is about 30% in 2011, since the SRAM area is 20% of the total F/F area when the same transistors are used for both F/F and SRAM.

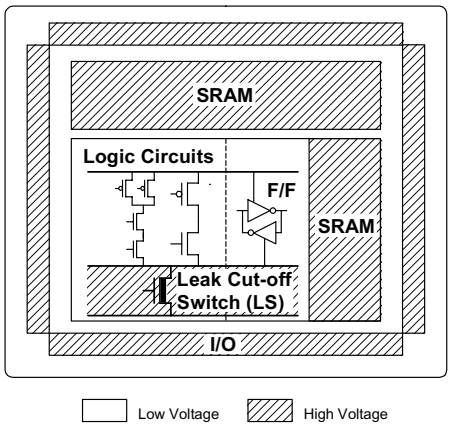

Fig. 10 A schematic of the proposed circuit scheme. In addition to logic circuits with LS, as shown in Fig. 2, this scheme is applicable to whole circuits including F/F circuits, SRAM and I/O circuits.

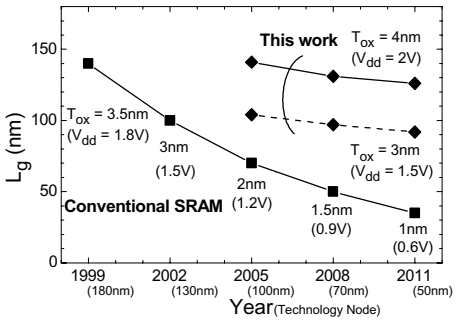

Fig. 11 The trend of gate length for conventional transistor and SRAM transistor. SRAM transistor can not be so much scaled due to large DIBL. If we take T_{ox} = 3nm (V_{dd} = 1.5V), further scaling of L_g becomes possible.

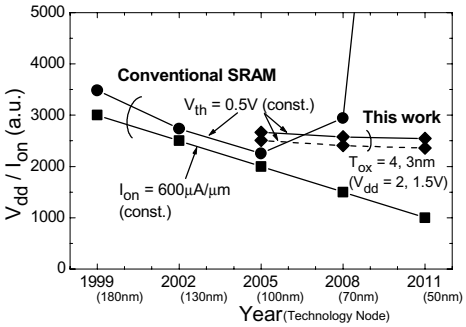

Fig. 12 The trend of device speed (V_{dd} / I_{on}). Although the speed of proposed SRAM design is degraded compared with conventional transistor whose on-current is kept constant, but this design can keep off-current sufficiently low and keep its speed faster than conventional transistor with constant off-current.

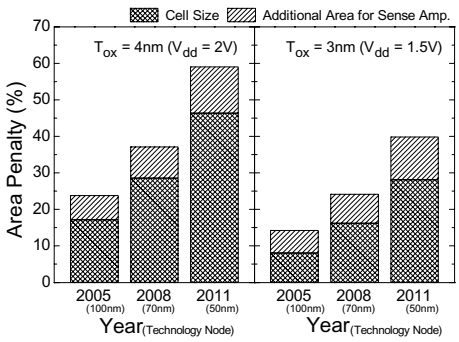

Fig. 13 The trend of area penalty for the proposed SRAM design. Area penalty has two origins: (i) an increase of cell area due to the increased Lg and (ii) an increase of additional area for sense amplifiers due to division of macro-cell into small macro in order to keep high speed.