Technology Roadmap for 22nm and Beyond

Hiroshi Iwai
Frontier Research Canter, Tokyo Institute of Technology
4259, Nagatsuta, Midori-ku, Yokohama 226-8502, Japan
Phone: +81-45-924-5471 Fax: +81-45-924-5584 Email: iwai.h.aa@m.titech.ac.jp

***Abstract*- Logic CMOS technology roadmap for '22 nm and beyond' is described with ITRS (International Technology Roadmap for Semiconductor) as a reference. In the ITRS 2008 Update published just recently, there has been some significant change in the trend of the gate length. The predicted trend has been amended to be less aggressive from the ITRS 2008-Update, resulting in the delay in the gate-length shrinkage for 3 years in the short term and 5 years in the long term from those predicted in ITRS 2007. Regarding the downsize limit, it would take probably 20 to 30 years until we reach the final limit, because the duration between the generations will become longer when approaching the limit. In order to suppress the off-leakage current, double gate (DG) or fin-FET type MOSFETs are the most promising. Then, it is a natural extension for DG FETs to evolve to Si-nanowire MOSFETs as the ultimate structure of transistors for CMOS circuit applications. Si-nanowire FETs are more attractive than the conventional DG FETs because of higher on-current conduction due to their quantum nature and also because of their adoptability for high-density integration including that of 3D. Then, what will come next after reaching the final limit of the downsizing? The answer is new algorithm. In the latter half of this century, the application of algorithm used for the natural bio system will make the integrated circuits operation tremendously high efficiency. Much higher performance with ultimately low power consumption will be realized.**

I. Introduction

The down-scaling of MOSFETs has been the most important and effective way for achieving the high performance and low power consumption for LSIs, and thus, the shrinking trends of the gate-length has been kept many years. Now, the power consumption became the limiting factor [1], and clock frequency and chip area have not very much increased recently. Furthermore, the concern for the difficulty of the downsizing is stronger than the past, facing the tremendous cost increase in lithography, difficulties in developing new technologies, and expected large variations of electrical characteristics of smaller geometry MOSFETs. However, still, the downsizing is the 'royal road' and the effort of downsizing will be continue by all means towards the limit until several more generations or 20 to 30 years, even though the duration between the generations would become longer. This paper describes a roadmap for high-performance logic CMOS technology for '22 nm and beyond' with ITRS 2008 Update [2] as a reference.

II. Roadmap for 22 nm

So far, the physical gate length of the logic CMOS has been much smaller than the half pitch of the lithography, however, the trends of the physical gate-length shrinkage predicted by recent versions of the ITRS have been even further aggressive for the most advanced semiconductor companies to catch up. Thus, the future trend has to be adjusted to be less aggressive in the ITRS 2008-Update, resulting in the delay in the gate-length shrinkage for 3 years in near future and even 5 years in the middle term as shown in Fig. 1 [2,3].

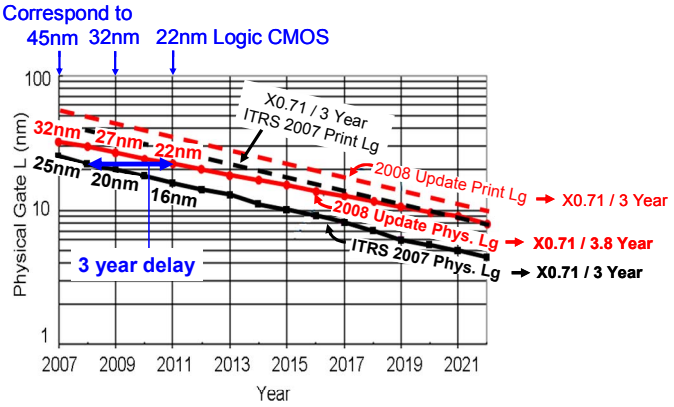

Fig.1 Comparison of ITRS 2007 and 2008 Update for the trends of printed (resist) and physical gate lengths

Due to the recent serious economical depression which started last year, all the semiconductor companies except Intel reduced the investment for R & D significantly. Thus, there is a high possibility that the gate-length shrinkage trend delays further. Corresponding to the delay in the gate length downsizing, the downsizing trend of EOT (Equivalent Oxide Thickness) of the gate insulator and junction depth will delay with the same pace. The critical dimension control or variation control of the gate length for 22 nm node becomes easier because of the gate length increase and red background of the column for the 22 nm node – which means no solution for the gate dimension control – turned to white and yellow. Also, the pace of the introduction of new technologies becomes slower. For example, introduction of DG or fin-gate structure will delay with 4 years, and 22 nm logic CMOS – which is expected to start production in 2011~12 –, can be made with the planer bulk CMOS , of course, as shown in Fig. 2 [2,3]. In other words, planar bulk CMOS will have a much longer life than expected by ITRS 2007. In ITRS 2007, it was expected that bulk planar structure is replaced by DG structure during 32nm node and that the silicon channel is replaced by high μ (mobility) materials such as Ge and GaAs from 22 nm node as shown in Fig.3 [2,3]. However, now, introduction of those structures and materials are thought to delay significantly.

IEEE Catalog Number: CFP0926C
ISBN: 978-1-4244-3832-7
Library of Congress: 2009900354

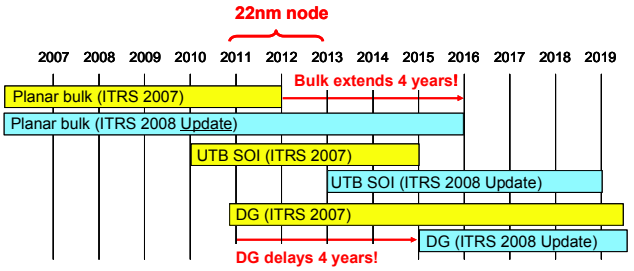

Fig.2 Comparison of ITRS 2007 and 2008 Update for MOSFET strcutures

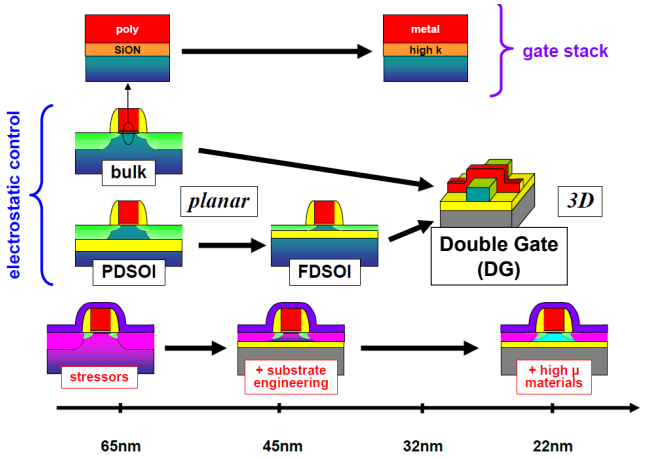

Fig. 3 Structure and technology innovation for MOSFETs

The clock frequency had kept increase until it reached 3 GHz. However, recent trend is that it even decreased slightly down to 1-2 GHz when introducing the multi-core scheme. Too high clock frequency too much increases the power consumption and resulted heat generation. This is not a wise way, and ITRS predicts only a small increase of the clock frequency as the entire chip operation. However, local on chip clock frequency, or the core clock frequency is expected to keep increase in the ITRS 2007. Already SRAM operation at 6 GHz was confirmed experimentally [4] and it is forecasted that the local clock frequency keep to increase with the same rate as before in the figure and even though 8% increase per year in IRTS 2008 Update with 6.3 GHz in 2011. It is not sure if the clock frequency can keep such an increase even it is a core frequency for a medium and long term.

Regarding the other issues for 22 nm technology node, 450 mm wafer is predicted to be introduced still in the 22 nm node from 2012 in ITRS 2008 update. However, most of the people do not think it can be introduced such a near future. For low-k, there is a slight retardation in k value with 0.1 to 0.3 in ITRS 2008 Update. It is really difficult to realize the low k value predicted by ITRS in production. In fact, k value kept retardation in almost every new version of the ITRS.

III. Low Supply Voltage and Off-Leakage Current

The increasing power consumption is the limiting factor of the logic CMOS, and lowering the supply voltage is the most effective way to decrease the dynamic power consumption. However, in order to decrease the supply voltage, the threshold voltage has to be reduced. This results in the significant increase in the 'off-leakage' current because of the significant increase of the subthreshold leakage current with low threshold voltage, as shown in Fig.4 [3].

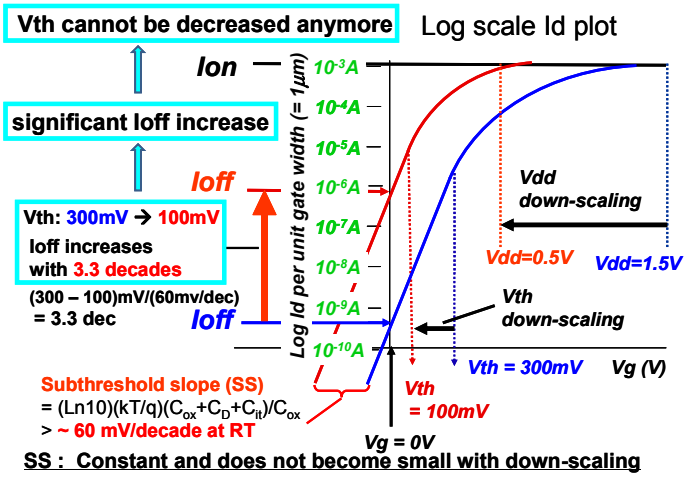

Fig.4 Increase in off-leakage current at low supply voltage

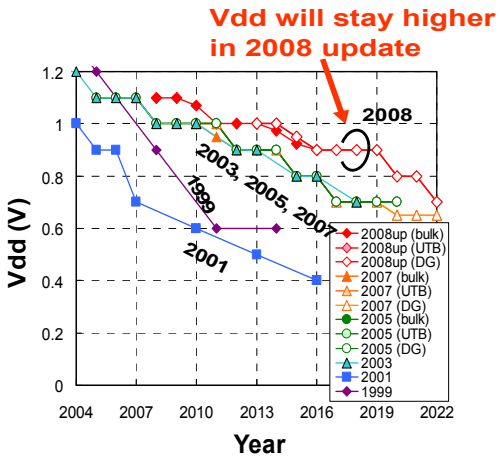

Fig.5 Trend of supply voltage for various versions of ITRS.

Thus, the threshold and hence, the supply voltages cannot be scaled-down easily. Their values are supposed to stay above 0.1 and 0.9 V, respectively for next 10 years in ITRS 2008 Update as shown in Fig.5 [2,3]. This kind of improper down scaling, — with keeping higher supply voltage and larger gate oxide thickness —, is the solution for the downsizing for the moment. However, the improper scaling enhances the short channel effects, resulting in the larger off-leakage current and larger variation of the thereshold voltage. The ITRS trends for the EOT of the gate insulator saturate at 0.5 nm. This will cause the increase in the off-leakage current and threshold variation in a future small geometry MOSFETs. Regarding the future possibility for the EOT below 0.5 nm, we have already experimentally confirmed a good operation of MOSFETs with EOT of 0.37 nm using the La_2O_3 gate insulator [5]. Thus, in

future the EOT value predicted in the roadmap will decreases a little further to solve the problems, and then, the supply voltage would decrease further in order to suppress the short channel effects.

IV. SRAM SCALING

SRAM composes a significantly important part of logic devices as cache memories and its occupying area is quite large. Even a small off-leakage current of a single MOSFETs in a SRAM cell makes a large off-leakage in the entire chip, hence, it is especially difficult to decrease the gate length and supply voltage of the SRAM cell. Thus, the gate length and supply voltage used in the SRAM cell are often designed to be larger than those used in the logic part of the chip. Nevertheless, the experimental fabrications of the SRAM cell shows the same reduction trend – with the shrink rate from 1/2 to 2/3 for every generation – until the 22 nm node as shown in Fig.6 [3]. In order to realize the 32 nm and 22 nm SRAM cells, new techniques are introduced. One is a double lithography to realize square endcap of the gate pattern [6,7] and another is high-k/metal gate stack in order to suppress the threshold voltage variation with reducing the EOT [6].

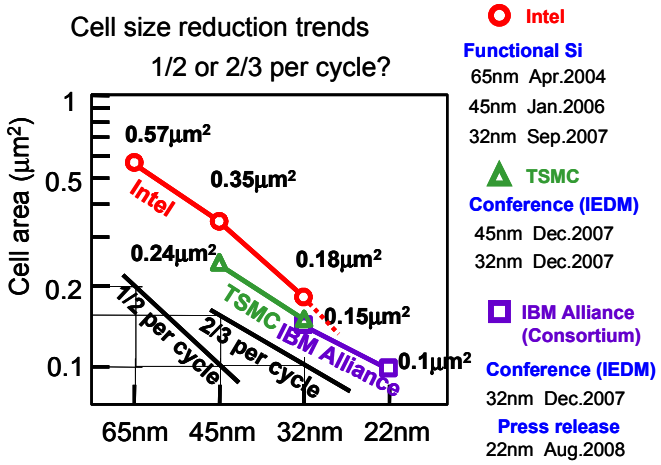

Fig. 6 Trend of experimentally fabricated SRAM Cell size.

It should be noted that the reduction of the supply voltage in SRAM cell degrades the data retention of the cell. In order to improve the data retention, it is necessary to improve both the read and write voltage margin of the cell. However, it is difficult to optimize the read and write voltage margins at the same time, because the optimum gate width ratio of MOSFETs between the SRAM latch and transfer gate parts are opposite for read and write margins. In order to solve this, in the design of Intel's new multi-core microprocessor, Nehalem, they use 8T cell for L1 and L2 cache in a core [8], separating the read and write bit lines to chose optimum gate width ratio for the read and write, paying a penalty of the cell area increase of about 30%. In this way, the supply voltage for the cores can be decreased with keeping high data retention. The high density L3 cache commonly used in the entire chip level still uses the 6T cell in order to suppress the increase of the chip area, paying a penalty of higher supply voltage than the core. In some future further beyond the 22 nm node, introduction of new cell structure such as DG-FET cell or D-RAM capacitor cell will keep the cell size reduction rate.

V. ROADMAP FOR FURTHER FUTURE

The logic CMOS will encounter its downsizing limit sometime in 2020-2030 around the gate length of 5 nm [9], presumably due to the huge off-leakage current in the entire chip. Thus, probably we will have 6 more generation until then. Two types of FET's have been recently recognized as the emerging devices which could replace current planer bulk CMOS [3]. They are the Si-nanowire FET and the alternative channel (such as GaAs and Ge) FET. They are quite different from the current planar type Si CMOS devices, in terms of structure and material, respectively. Considering the compatibility with current Si CMOS process technologies, Si-nanowire FETs would be easier for production and more promising. Even if the alternative channel FET would become the main stream, the channel shape should be a wire type, because of the strong demands for the suppression of the off-leakage current.

The Si nanowire FET has higher on-current conduction due to their quantum nature and also because of their adoptability for high-density integration including that of 3 dimensional stacked layer structure [10]. Because the nanowire pattern itself is simple, nano-inprint technology will be used for future high-density lithography with extremely small pitch. If the ideal one dimensional ballistic conduction is realized for the nanowire, the nanowire itself has basically a high quantum conduction with 77.8 μS per wire regardless of the wire diameter and the channel length. In addition, the channel current is multiplied with the number of the quantum channel available for the conduction. However the increase of the quantum channel degrade the conduction because of the carrier scattering between the conduction bands, and there is a trade off relation between the one-dimensional ballistic conduction and the number of quantum channel in terms of the nanowire width. Smaller wire diameter is desirable for one-dimensional ballistic conduction and larger diameter is desirable for the umber of quantum channel.

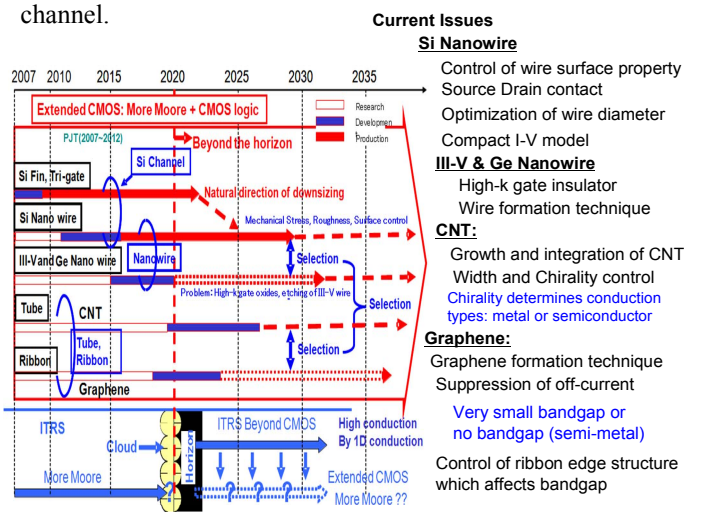

Fig. 7 Long range roadmap for logic CMOS transistor research for next 30 years.

Figure 7 [9] shows a long range roadmap including the period which ITRS does not covers. The Si nanowire FETs is the most promising candidate as explained. III-V and Ge nanowire FETs are the 2nd candidate. However, technical barrier for the fabrication process is much higher compared with the Si nanowire. In further future, CNTs (Carbon Nanowire Transistor) and graphene ribbon FETs could be candidates to replace the Si nanowire FETs. However, they are still too far at this moment because of no substantial idea for the integration method of so huge number transistors in a chip and bandgap control for high on/off ratio.

What will come next, after reaching the final limit of the downsizing? Probably, there will be the innovation or revolution in the production method of LSIs, and the LSIs will be produced with much cheaper cost. Then, the next step is to use new algorithm. In the latter half of this century, the application of algorithm used for the natural bio system such as the brains of insects as shown in Fig. 8 [11] and even those of human will make the integrated circuits operation tremendously high efficiency. Just for example, brain of the mosquito make the real time 3D flight control with image processing equipped with many sensors such as infrared and CO_2 with extremely small brain volume and extremely small energy consumption. The performance of dragonfly's brain is much higher. Today's performance and energy consumption of the microprocessor are not comparable to those of insect brains, at all. Introduction of the algorithm of the bio system will be the ultimate method in the roadmap.

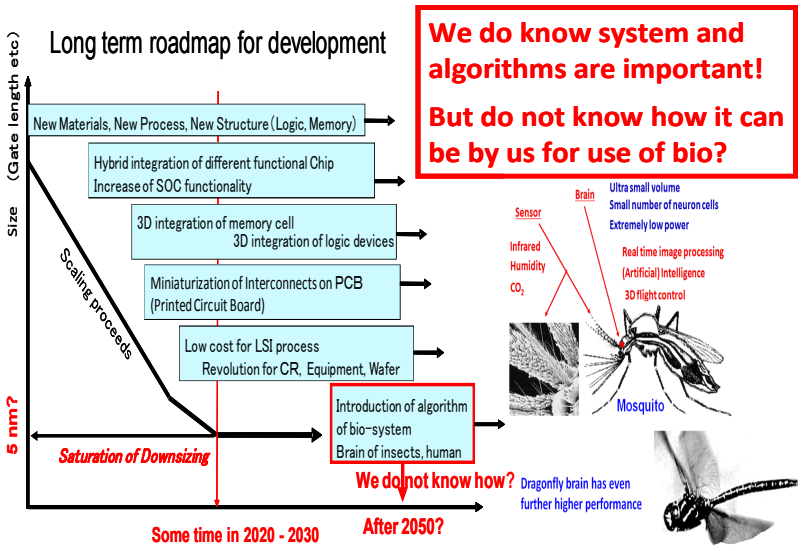

Fig. 9 Long term roadmap in this century

VI. Summary and Conclusion

A roadmap for high-performance logic CMOS technology for '22 nm and beyond' with ITRS 2008 Update as a reference. The predicted trend of gate-length reduction in the past version of the ITRS was too aggressive for the industry to catch up and thus, the pace of the reduction in the gate length become less aggressive from the ITRS 2008 update. Corresponding this, introduction of the new technologies, structures, and materials will delay and 22 nm logic CMOS will be made with planar bulk MOSFETs. The supply voltage reduction is a very difficult item for the next 10 years, because of the difficulty in reducing the threshold voltage any more, and the supply voltage stays at 0.9V even in 2019 in the ITRS 2008 update. 22 nm SRAM cell for cache application can be made with planar MOSFETs with introduction of new technologies.

In the long term, Si nanowire FETs are the most promising candidate because of process compatibility with the current planar CMOS LSIs, and also because of its small off-leakage current and high on-current. In further future, introduction of the algorithm of bio system will be the key for further improvement of the performance and energy consumption.

Acknowledgements

This study was partially supported by the 'Innovation Research Project on Nanoelectronics Materials and Structures' sponsored by Ministry of Economy, Trade and Industry, Japan. This paper is based on a lecture given at the IEDM Short Course given in San Francisco, US on December, 2008. The author would like to thank Drs. H. Ishiuchi of Toshiba M. Saito, Y. Urakawa, and T. Yabe of Toshiba, K. De Meyer of IMEC, Dr. Bohr, T. Ghani, and P. Gargini of Intel, B.S. Haran of IBM, Profs. K. Natori and K. Shiraishi of Tsukuba University, Profs. K. Yamada and K. Ohmori of Waseda University, and Profs. K. Kaukshima and P. Ahmet, and Messrs. T. Kawanago, S. Sato, and Y. Lee of Tokyo Institute of Technology for the useful information and discussion for the preparation of materials.

References

[1] H. Iwai and S. Ohmi, Microelectronics Reliability vol. 42 p.1251, 2002
[2] http://www.itrs.net/reports.html
[3] H. Iwai, SC, IEDM 2008
[4] J. Pille, C. Adams, T. Christensen, S. Cottier, S. Ehrenreich, F. Kono, D. Nelson, O. Takahashi, S. Tokito, O. Torreiter, O. Wagner, and D. Wendel, ISSCC Dig. Techs, p. 322, 2007
[5] K. Kakushima, K. Okamoto, K. Tachi, P. Ahmet, K. Tsutsui, N. Sugii, T. hattori, and H. Iwai, IWDTF, p.9, 2008
[6] K. J. Kuhn, IEDM Tech. Dig. p.471, 2007
[7] M. Bohr, Proc. ICSICT. P.13, 2008
[8] Intel Developer Forum 2008
http://www.intel.com/idf/index.htm
[9] H. Iwai, IWJT, p.1, 2008
[10] Y. Lee., T. Nagata., K. Kakushima., K. Shiraishi, and H. Iwai, IWDTF, 83, 2008
[11] H. Iwai, H. S. Momose, M. Saito, M. Ono, and Y. Katsumata, INFOS, p.147, 1995
[12] H. Iwai, IPFA, p.1, 2006