2

Trends in Research and Development of Next-generation Si-MOS Devices

ATSUSHI OGASAWARA
Information and Communications Research Unit

2.1 Introduction

Preceding the VLSI Symposium (held from June 12), which is one of the major international meetings for research findings on semiconductor devices that are fundamental elements of IT, important announcements that indicated future directions for Si-MOS devices were made by IBM (on June 8) and Intel (on June 9).

The June 8 announcement made by IBM focused on their achievement of increasing the operating speed of a MOS transistor by extending Si crystal lattices to increase the mobility of the electrons in the transistor's channel without changing the design rule. The other announcement by Intel focused on their achievement in further scalinging the gate length of a MOS device from 30nm to 20nm (0.13μm, or 130nm, with the existing process).

According to newspapers and other mass media, the two announcements have drawn much attention since they exceeded the Si-MOS transistor's capacity, which was estimated to reach the limit in the first decade of the 2000s, and they made it possible to hold Moore's Law - the number of transistors per square inch on integrated circuits doubles every 24 to 48 months - until the 2010s.

However, this article will not explain the operating speed of a single transistor, but will examine the two announcements by IBM and Intel and discuss their significance.

2.2 IBM's and Intel's Announcements

2.2.1 IBM

IBM announced that it developed a transistor with increased mobility of electrons by extending the Si crystal lattice to eliminate the main cause of electron dispersions through combining the SiGe layer and the Si layer.

In comparison to existing MOS transistors under the same design rule, the IBM transistor's speed can be increased by approx. 70% in electron's mobility and 35% in chip performance, according to IBM.

In other words, the announcement indicated that the performance can be increased by 35% without further scaling transistors.

2.2.2 Intel

In December 2000, Intel developed a transistor with an effective gate length*[1] of 30nm (70nm under the process rule), which shook device researchers. However, only 6 months later, the company developed a MOS transistor that has an effective gate length of 20nm (45nm under the process rule) and an oxide film thickness of 0.8nm (equivalent of 3 atoms), and is capable of turning on/off 1,500 billion times per second. This represents 1,000 times the switching speed of existing MPU (micro processing unit) transistors. Intel plans to commercialize an MPU operating at 20GHz using this transistor by 2007.

2.3 Concept of High-speed Transistors

As stated earlier, IBM increased the operating speed by improving the mobility of electrons, while Intel achieved it by further reducing the gate length. However, the realization of high-speed transistors does not necessarily mean the realization of a logic circuit operating at high speed or the realization of faster LSIs such as MPUs.

In fact, the strained silicon technology announced by IBM is the same technology that supported the phenomenon observed in thin-film SOI MOSes produced from direct bonding SOI*2 or SIMOX (oxygen ion impregnated SOI) between the late 1980s and the early 1990s. Although the two technologies differ in the cause of stress imposed on Si, i.e., the Si-SiO_2 interface or the Si-SiGe interface, they equally form strained lattices in principle.

With the strained SOI MOS device, like the transistor announced by IBM, the speed of the n-channel MOS, which uses electrons as the carrier, increased by more than 10% as a result of reducing the effective mass of the electron and eliminating the causes of dispersion. On the contrary, with the p-channel MOS, which uses holes as the carrier, the effective mass of holes increased, causing the mobility to drop and decreasing the operating speed to a level slower than ordinary p-channel MOSes, which comply with the same design rule.

This stirred a heated debate among SOI researchers, including those at Stanford University, UCB (University of California Berkley), Hewlett-Packard and Xerox, as well as scientists at MIT in the east coast and IMEC in Europe.

In many logic circuits we use as LSIs, so-called CMOS (complementary metal oxide semiconductor), a combination of an n-channel MOS that uses electrons (negative charge) as the carrier and a p-channel MOS that uses holes (positive charge) as the carrier, plays an important role. This means that in many logic circuits, if only the n-channel is fast, the overall performance may not necessarily be fast. Of course, it is not impossible to make up a circuit with n-channel transistors only. However, if such a circuit makeup had been expected to become the mainstay, the GaAs transistor*3, which was several times faster than Si even at that time, would have dominated the market.

The debate came to a conclusion that although the faster n-channel MOS would be attractive, it was more important to balance between the n-channel and p-channel transistors and keep both components fast. (At that time, other than this problem, there were many more problems caused by factors such as the Si-SiO_2 interface, local stress on the transistor, and high field strength. Thus, no high-speed transistor was put into practical use.)

2.4 Course of Direction for High-speed Transistor

From the viewpoints outlined above, it may appear that the announcement by IBM lacks originality because it is based on conventional technologies discussed in past debates. However, the announcement is still notable in it has clearly established a design concept of aggressively controlling the effective mass and the mobility of a Si-MOS transistor by using strained lattices.

In the field of semiconductor lasers, design technologies for the strained quantum well structure rapidly advanced over the past 10 years, and the strained quantum well structure has been applied to the laser units of most CD, MD and DVD players available on the market today. Even in this field, there was also a fierce debate in the beginning on reliability and other factors in straining crystals, and it was pointed out that commercialization would be difficult since it was necessary to control multiple layers at the nano level. However, these challenges have been overcome today, and the strained quantum well structure is incorporated even in mass-manufactured game machines and inexpensive laser pointers priced at less than 1,000 yen. Taking this into account, the approach of making the most of strained lattices to control the solid-state properties is commercially feasible and a preferable alternative.

However, unlike laser units, which are frequently used as discrete devices, transistors are required to serve as large-scale integrated circuits. And, it is

problematic that the future of p-channel MOS transistors, which are the other components for the devices, and of CMOS, which combines both, is not mentioned.

By contrast, Intel has steadily advanced miniaturization by taking advantage of the cutting-edge EUV technology reported in the May 2001 issue of "Science and Technology Trends (Japanese version)," and thus, it can be said that Intel has been on the high road. This approach is basically the same as the conventional methods and seems to have only few problems. However, is it really correct to only take that approach?

Concerning circuit performances, even with the present n-channel and p-channel, MOSes have different drive capacities and, in fact they are unbalanced. When comparing the transistor areas between the n-channel MOS and the p-channel MOS, the latter is 3 times larger than the former. There is an opinion that in order to increase the overall speed of the circuit, it is preferable to raise the speed of the p-channel MOS while slightly reducing the speed of n-channel MOS (priority is given to increasing the mobility of the hole rather than the electrons) because it will increase the rate-determining unit's performance. Based on this opinion, Ohmi Lab of Tohoku University has studied the formation of transistors on the (111) crystal surface of Si (existing MOS LSIs are formed on the (100) surface). In addition, if the compression stress is imposed on Si as opposed to the IBM announcement, a similar result may be obtained. Thus, there may be other alternatives as well. In fact, at the recent IEDM (IEEE International Electron Device Meeting, December 2001), UCB presented an effort to increase the performance of the p-channel MOS by straining SiGe. Although such approaches may seem inconspicuous, they should be considered very important.

2.5 Conclusion

Instead of focusing on the speed of a single transistor, it is desirable to discuss how to achieve maximum performance in terms of the whole circuit and the whole LSI, and what should be done for that. The June issue of "Science and Technology Trends (Japanese version)" also mentioned the problem of wiring in high-speed MPUs, indicating that it is increasingly important to study next-generation Si-MOS devices from diversified points of view.

Note:

On June 25, IBM announced a high-speed transistor that operates at 210GHz. This is a type of bipolar transistor called HBT (hetero bipolar transistor). It is mainly used for RF communication and also as a component of super computers. Bipolar transistors were the mainstay when most super computer architectures used vector processors. However, amid the current situation where multiple general-purpose processors operate in parallel (multi processors) and multiple computers operate in parallel (cluster) as the mainstay architecture, MOS transistors are becoming the mainstay component for high-speed computing. Therefore, they are not mentioned in this article.

Glossary

*1 Effective gate length
The physical length of a MOS transistor's gate electrode is called the gate length. However, elements that form the source and drain electrode areas immediately beneath the gate are partially dispersed, making the area that actually acts as the gate shorter than the gate electrode. This actual gate length is called the effective gate length. Since this gap is large with a miniaturized MOS, the effective gate length is used in many cases.

*2 SOI (silicon on insulator)
With Si crystals used in most LSIs, Si cannot be used as an insulator, and multiple transistors are electrically separated with the pn junction. However, it generates a parasitic current that also acts to suppress high-speed operation of a transistor. Accordingly, the SOI structure where a Si layer is formed as the operating layer on the SiO_2 insulation to produce a transistor has been tested.

*3 GaAs transistor
This represents gallium arsenide transistor. This transistor was considered the favorite for high-speed devices until 10 years ago. It is a chemical compound semiconductor consisting

of Ga (gallium) and As (arsenic). Since the speed of its electrons is several times faster than Si, research has been conducted on this transistor as a component of super computers. However, it has lost its position as an element for digital circuits for several reasons such as the difficulty in obtaining quality insulation films (oxide films) like SiO_2 for Si MOS, difficulty in producing a mutually complementary circuit like CMOS by producing a p-channel transistor, and the inability to obtain a specific advantage as comparable to the gap in electron mobility because the element was not significantly downsized as opposed to Si. This component is widely used in analog RF communication applications such as for satellite broadcast reception and mobile phones.

(Original Japanese version: published in July 2001)