Impact of electron device miniaturization on computing and world after the end of miniaturization

Hiroshi Iwai
Institute of Innovative Research,
Tokyo Institute of Technology
Yokohama
JAPAN
Iwai.h.aa@m.titech.ac.jp

Abstract: - The downsizing of electron devices has always been the key to promote the computer performance and energy consumption for the past 70 years since its beginning. However, it should be noted now that the downsizing of the electron devices cannot continue forever, and expected to reach its limit about 10 years from now. Thus, no big progress expected in the microprocessor level any more, but the rapid growth of IoT technologies will connect many different kinds of devices, resulting in the significant enhancement of the processing abilities in our society. Also, AI constructed on the semiconductor computer will become much wiser than human brain in specific purpose use, helping our society promote to higher stage. In a long term, however, even with the tremendous progress of AI technologies, the system fabricated on the solid-state devices and their network cannot compete with the existing bio system constructed on organic materials in efficiency and maintenance. Thus, the technology how to use the bio system better, or how to combine electron and bio systems will become very important.

Key-Words: - end, limit, downsizing, electron devices, computing, scaling, IoT, AI, biology, neuron, synapse

1 Introduction

Electronics started with the invention of vacuum tube at the beginning of 20th century. Among the inventions of vacuum tubes, especially that of the triode in 1906 was the most important because it has a similar functions of the transistor and can modify (or control) the current between the cathode and anode by the change of the grid bias like a switch or an amplifier. With the appearance of the vacuum tubes, wirelessly communicating, computing and many other things became possible. Today, electronics are regarded to be the most important invention in the 20th century because almost all the activities in our society cannot be done without the support of electronics. The progress of the electronics, especially that of computing technology, has been conducted by the downsizing of the electron devices such as vacuum tubes and semiconductor transistors. The downsizing decreases the capacitance and operation voltage of the device, resulting in the decrease in switching time and power consumption. Also, the downsizing can realize more device integration, resulting in the cost reduction per device, and also the functionality increase in the integrated systems.

Although the continuous downsizing of the electron devices has significantly contributed to the promotion of computers, it is believed now that the downsizing will reach its limit in ten years because of several reasons. In this paper, the reasons for the limitation of the downsizing are explained and the world after we reach the limit is predicted.

2 Impact of downsizing

The downsizing of electron devices has always been the key to promote the computer performance for the past 70 years since its beginning. The downsizing has increased the number of the components loaded in the computer, resulting in the enhancement of its functions and parallel processing. At the same time, the downsizing has reduced the capacitance of the components, resulting in the reduction of the processing time and energy. Indeed, the size of the recent MOSFETs decreased 10 million times from that of the vacuum tubes used in the first computer, ENIAC, and, as the result, the performance per Watt of a recent smart phone made of semiconductor transistors is 1 trillion times higher than that of ENIAC.

Today, 1T bit memories (128 GB SD cards) are popular in the market with a reasonable price for individual use. However, if we assume that there had been no downsizing of electron devices and that we had to use 1 trillion old vacuum tubes to construct a 1T bit memory, what would be the situation? Then, the price would be an order of 1 trillion Japanese yen, which is equivalent with yearly Japanese government budget, the weight would be 100 Mton, which corresponds with the weight of 2 million elephants, the volume would be $1 \times 0.5 \times 0.5$ km^3 , which is larger than that of the world tallest building in Dubai, the power consumption would be 50 TW, which is as large as the power generation of 5,000 nuclear reactors, and yearly energy consumption would be 4×10^{20} calorie, which is as big as the 2,000 times of the yearly total human consumption on the earth.

3 Limit of downsizing

However, it should be noted now that the downsizing of the electron devices cannot continue forever, and that we will do face the limit near future. What is the limit of the downsizing?

2.1 Ultimate limit

The ultimate limit of the downsizing of the device elements is about 0.3 nm, which is defined by atomic distance of the materials. No one can make an electron device structure smaller than the atomic size. Thus, there is no pico-meter technology in future.

2.2 Fundamental limit

Before reaching the ultimate limit, there is the fundamental limit for the operation of the devices, which is about 3 nm (Fig.1).

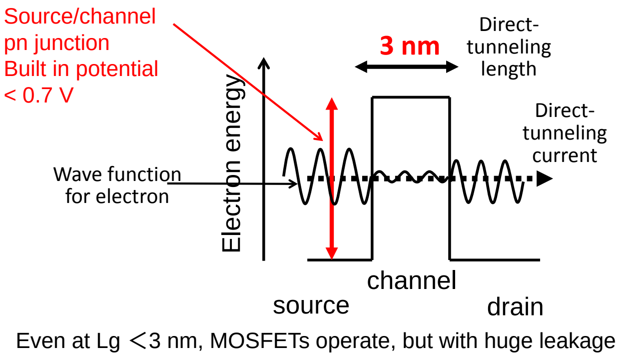

Fig.1 Direct-tunneling leakage during switch off

When the element size becomes less than 3 nm, quantum mechanical tunneling phenomenon makes no electrical insulation between the device elements, resulting in no switching-off of the devices.

2.3 Practical limit

Furthermore, the downsizing will stop even before reaching the fundamental limit if the demerit of the downsizing becomes bigger than the merit. This is the practical limit. The demerits could be cost increase, yield and reliability degradation and even performance degradation. Considering that the CMOS technology is now in the stage of shifting from 14 nm technology to 10 nm, there are not so many generations left until reaching the practical or fundamental limit, and it is expected that CMOS will reach its limit about 10 years from now.

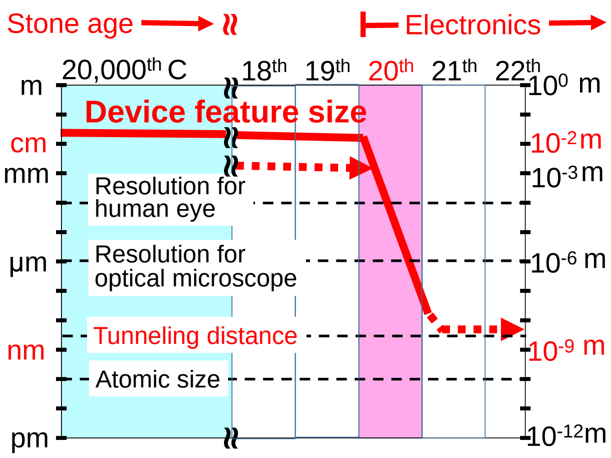

Fig.2 Device size downsizing trend in human history

Fig.2 shows the past trend of device downsizing in human history. We can recognize that 20th century is characterized by the name of ‘century of downsizing’. It is difficult to precisely predict the number for the practical limit, but from the figure, it is easily recognized that there is only a few rooms left for the further downsizing until to reach the fundamental limit.

Unfortunately at this moment, there are no really promising electron device candidates after the CMOS reaches the limit. Thus, in near future, the tremendous progress of microprocessors in performance and power consumption cannot be expected any more.

4 World after reaching limit

4.1 Future for middle term

After reaching the limit of the downsizing, there is no more Moore's law [1] and we have to accept the fact that such tremendous merits as we enjoyed in the past obtained by the downsizing cannot be expected.

Then, what will happen for computing technology? No big progress can be expected any more at the microprocessor or the semiconductor chip level. However, on the stage of the cluster and cloud computing level, there will be still rooms for the progress for another decades. Especially, rapid growth of IoT technology connects not only the microprocessors but also many different kinds of devices through the network, -- such as sensors, photovoltaic devices, power devices, batteries, super capacitors, memories, etc. --, resulting in the significant enhancement of the processing abilities in our society. This in a virtual extension of Moore's low by net. Experience and knowledge obtained by the development of Si CMOS device in in the past many years will be useful for the development of the new solid-state devices with different kinds.

On the software side, AI technologies utilizing deep learning scheme on the neural network created on the semiconductor devices will make tremendous growth of the processing abilities. Recently, 'Go' game software, 'alpha-Go', defeated the professional champions of the game, and the development of AI seems to make the computer much wiser than human brain for specific purposes use or applications.

It is sure that AI will execute specific jobs better than human. AI on the software on semiconductor devices can learn many experiences in a short time and can evolve the neural networks on the program very rapidly. While, the human brain needs much more time to learn and evolve its neural networks because the physical connection change of the neurons are required. For example, 'Go' game software needs days to months to become a champion level, but the human brain needs at least a decade and even only very limited persons can reach the champion. This is the big advantage of AI over human brains.

The near future will become an exciting period and many technological innovations will be born. Everyone may have a big chance to contribute, because the most of the tools will be already ready for use; AI program, nanofabrication tools, simulators from system design to material engineering. Only the necessary thing is an excellent idea for the application and a successful business model.

4.2 Future for long term

However, one thing to notice is that AI on semiconductor devices needs large power consumption. Indeed, Alpha-go used 1,000 CPU's and 200 GPU's to defeat Mr. Lee, who is a world-top class professional 'Go' player. The consumed AI energy in this case was much more than that of the human brain which is only tens of Watt. Thus, in terms of power consumption, AI constructed on semiconductor devices does not seem to compete with bio systems at all. If we assume to make a dragonfly robot using semiconductor brain and eyes, probably extremely huge power to operate super computers would be necessary to emulate the sophisticated real time 3D flight control of the dragonfly. Furthermore, the CMOS image sensor equivalent with dragonfly's big compound eyes and real time processing of the image will consume another big energy. Now, physical analysis of neuron networks in animal brains including that of insects and humans is going on aggressively, and in some future, we may figure out the brain algorithm. Installing the brain algorithm on semiconductor devices could reduce the power consumption significantly. However, this kind of 'semiconductor brain' will still not able to compete with bio system or real brains because of the three reasons.

One reason is that the vacuum tube/semiconductor computers have a history of only 70 years, while the bio systems have that of billions of years of natural selection. In other words, if the semiconductor computers were more efficient for performance and energy consumption, animals could have been certainly equipped with a semiconductor brain with metal nerves during our long time bio evolution. But, in reality, it does not become so. The second reason is that there are huge number of interconnects between the neurons in the brain. It is said that one neuron have thousands of connections to other neurons through synapses. It is almost impossible to construct 1,000 contacts for each transistor in the semiconductor integrated circuit case, because there is no space for the small geometry transistor to provide so many contacts. Even if we could assume to allocate 1,000 connections to a semiconductor transistor, then, the parasitic capacitance and resistance of the interconnects would become too huge, and each transistor should be a huge power transistor to drive the capacitive and resistive huge load. The third reason is the extremely low voltage pulse of several tens of milli-volts used for the neuron signals. For the case of MOSFETs, the operation of sub-100 mV is very difficult because of significant increase of the subthreshold leakage current. When the threshold voltage becomes too small such as below 100 mV, the portion of electrons which has higher energy than the threshold potential barrier increases dramatically, and as the result, the subthreshold leakage current due to the electrons overflowing the potential barrier becomes too huge (Fig.3). In the case of neuron, the signal pulse is created only when receiving neurotransmitters or chemical messengers from other neuron through the synapses and thus, there is no constant voltage applied to the neuron, resulting in no leakage at the off state.

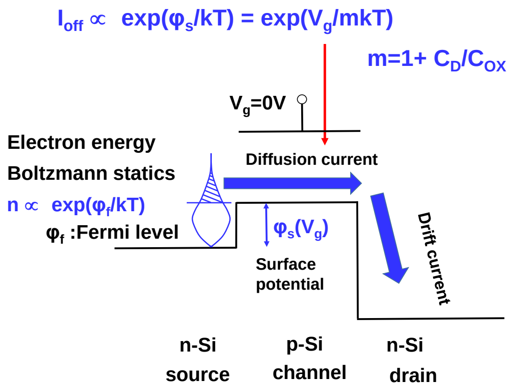

Fig. 3 Subthreshold leakage current

Implementing brain algorithm into the semiconductor devices could enhance the AI ability and create certain big merits. But in terms of the energy consumption, AI realized on semiconductor devices cannot complete the bio systems. This means that the AI constructed on the semiconductor computers will not superior to human brains comprehensively, although AI will be superior for many specialized tasks. Thus, in a long term, we may have to think about using the organic bio system as a vehicle of AI. However, designing complicated artificial brain using neurons and fabricate it from scratch seem to be almost impossible Even, utilizing DNA, optimizing the artificial organic brain by fabrication experiments will take huge number of years such like the time required the evolution of real brains.

Thus, using the existing brains of animals, -- such as insects, fishes, birds and dogs -- seems to be the good solution. Furthermore, considering the maintenance of the brain, it will not be a wise selection to use the separated brains from the body. We will use brains with entire animal bodies being alive, although ethical issues should be carefully considered. Indeed, we have already used the dogs for so many years as very good friends of humans in many purposes, such as – drug sniffing, hunting, sledding, policing, rescuing people, seeing eye, stock framing, and pet. We have also used hawks for hunting, cormorants for fishing, and pigeons for sending messages. However, in future, we could use more animals with more reliable manner. For example, some insects such as ants, dragonflies, and fishes, and birds could do very good surveillance of the places where human and machine are very difficult to reach with low energy consumption. This will be very helpful to check the degradation of constructions -- such as roads, bridges, tunnels, buildings from the surface, sky, ground and water. In order to make the operation reliable, interface technologies to directly communicate with the animal brain or neurons will become very important. Energy and information have been the two important elements for human society during its long history. Before the electricity became utilized at the end of 19th century, only biological, chemical and mechanical measures were used for human activities for energy and information. For example, for movement, human's walk, horse riding, horse-driven carriage use biological energy. Sail boat and steam locomotive use physical energy. For processing information, human brains have been the major measure. Sometimes, dogs and other animals helped human as already described. Thus, biology was the major measure for the information processing, although abacuses and written documents were used as the means of subsidy.

Advantage of the electricity over the biological, chemical and mechanical measure is that electricity can transfer huge volume of energy and information instantaneously with wired or wireless media even for a long distance. Because of this advantage, motors driven by electricity has been the major source of the motive force since 20th century, in addition with engines which can create huge power by chemical explosion. Also, because of the electricity advantage mentioned in the above, the physical processing speed of information in computer is extremely high and that is the reason why the computer can make calculations much faster than human. Motor is thought to be more clean than engine from the ecological point of view, and thus, electricity has been believed to be the almighty as the future measure for human activities for both energy and information. Indeed, significant portion of the measures are going to be replaced from biological and chemical to electrical ones. AI on the computer is one example for the information processing, and EV (Electric Vehicle) is the one for the motive force. This will certainly improve the quality of human society.

However, electric measure is not necessarily to be the best choice in terms of energy consumption depending on the situation as we saw at the discussion for the AI and brain. It is interesting to note that animals or humans do not use electric motor but muscle made of protein for the movement. If everyone has electric motors to move the body, total electric power consumption of 7 billion people on the earth should be extremely huge.

3 Summary and conclusions

The downsizing or miniaturization of electron devices has been the major driving force for the tremendous progress of computing in the past 70 years since its beginning.

However, the downsizing will reach its limit about 10 years from now. Thus, there will be no more Moore's law, and tremendous improvement in performance and energy consumption of microprocessors will not be expected in near future. However, rapid growth of IoT technology will connect not only the microprocessors but also many different kinds of devices through the network, resulting in the significant enhancement of the processing abilities in our society. This in a virtual extension of Moore's low.

AI on the neural network constructed on semiconductor integrated circuits can learn many experiences in a short time, and make evolution very rapidly. Thus, AI will become much wiser than human brains for specific purpose use.
However, energy consumption of AI on the semiconductor devices is huge compared with animal brains including those of insects and humans. Thus, for a long term, utilizing existing bio system or animal brains will be considered seriously. In order to guarantee reliable control of the bio systems interface technology to directly make communication between electron devices and animal brains or neurons will become important. The good combination of electron and bio devices will be the best solutions.

References:

[1] G. Moore, "Cramming more components onto integrated circuits" Electronics, pp. 114-117, April 19, 1965.