[Scientific Innovation Series 6]

반도체 기술의 미래 녹취록

행사일: 2021 년 4 월 16 일 (송출)

기록자: 박유원, 김지수

Introduction

신창환:

Distinguished guests, ladies and gentlemen. Welcome to the Chey Institute's Scientific Innovation Series. I am Changhwan Shin from SKKU.

Today, with world-class scholars, I am going not only to highlight semiconductor technology as a key driver for future scientific innovations but also to explore challenges and opportunities in the sector of semiconductor materials, devices, and circuits. Semiconductors have become indispensable building blocks for modern-day life. The competition in industry is intensifying, as nations aim to take technological leadership, and the recent disruptions in the semiconductor supply chain show that semiconductor supply chain is essential to the world's businesses.

For today's webinar, I am going to introduce four distinguished speakers. We have the pleasure of inviting Professor Tsu-Jae King Liu again. She has delivered her plenary talk about a year ago in the 2nd Chey Scientific Innovation Conference. Today, she is going to talk about emerging chip technologies for the age of ambient intelligence, such as next-generation transistors, milli-volt & ultra-sensitive switches, optical interconnects & Photonics integration, flexible electronics, embedded non-

volatile memory, Silicon photonics + CMOS integration, and CMOS + NEMS integration. And, I'm very excited to have three more new speakers, Professors Sayeef Salahuddin in UC Berkeley (who will talk about ultra-thin ferroelectrics on silicon and its application for energy-efficient logic and memory devices, mostly focusing on negative capacitance field effect transistor), Prof. Changhwan Choi in Hanyang University (who will talk about heterogeneous integration, including monolithic-3D semiconductor processing technology published in International Electron Device Meetings 2020), and Prof. Jae-duk Han in Hanyang University (who will talk about automated circuit design methodology).

I hope that today's webinar featuring the world-renowned speakers serves to be very resourceful for audience to have an insight for future semiconductor technology. Please stay with us by the end of today's Webinar.

Plenary Session

신창환:

Let me introduce the first speaker in the plenary session.

Tsu-Jae King Liu has received her B.S., M.S. and Ph.D. degrees in electrical engineering at Stanford University in 1984, 1986 and 1994, respectively. From 1992 to 1996, she was member of research staff at the Xerox Palo Alto Research Center (PARC). In 1996, she joined the faculty of the Department of Electrical Engineering and Computer Sciences (EECS) at the University of California, Berkeley, where she is now Dean of the College of Engineering.

Liu is internationally known in academia and industry for her innovations in semiconductor devices and technology, and is highly regarded for her achievements as an instructor, mentor and administrator. She is a fellow of the Institute of Electrical and Electronics Engineers (IEEE), an elected member of the U.S. National Academy of Engineering, a fellow of the U.S. National Academy of Inventors, and a board member for Intel Corporation. She has received many awards and honors. One of them is the Defense Advanced Research Projects Agency (DARPA) Significant Technical Achievement Award for her role in the development of the FinFET. The FinFET is an advanced transistor design used in high-end computer chips that you are using now.

Let’s take a listen to her talk.

Tsu-Jae King Liu:

Greetings, everyone. It is my pleasure to participate in the Chey Institute's Scientific Innovation Series event today to share with you my perspective on emerging technologies for the age of ambient intelligence. I would like to thank President Park of the Chey Institute for Advanced Studies for inviting me to participate in today's virtual conference.

Shown here on this slide is a vision of the future during the age of ambient intelligence. Advancements in information and communication technologies have given rise to cloud computing and the Internet of Things, which, together with advancements in machine learning, are giving rise to artificially intelligent systems, which can be used to manage critical infrastructure to realize future smart cities and also can be used to manage intelligent medical devices for personalized health care

and medicine for future smart hospitals. Artificially intelligent systems also can be used to automate cars and to optimize traffic flow, to realize smart highways and also to optimize factory operations and logistics for future smart factories.

Now, artificially intelligence systems require real-time processing of large quantities of data. This is possible today because the capabilities of computing devices have improved dramatically over the past one hundred and twenty years. Today they are rapidly approaching the capability of the brain. Kurzweil's Law states that the performance of the most advanced computing device measured in the number of calculations per second per dollar increases exponentially with time. This is because technological innovation itself is accelerated by advances in information technology.

Underlying Kurzweil's Law is Moore's Law, which states that the number of transistors on the most advanced microprocessor chip doubles every two years. Steady advancement in chip manufacturing technology over the past 50 years has enabled smaller and smaller transistors to be fabricated, so that more and more of them can be integrated together on a single die to provide for improved functionality at lower cost per function. This, in turn, has enabled new electronics products with greater information processing capability, resulting in market growth. Semiconductor companies generally invest a portion of their revenues or profits into advancing manufacturing technology to enable this virtuous cycle to continue. Today, the most advanced microprocessor products comprise almost 40 billion transistors on a die. Now, underlying Moore's Law are advancements in the materials, processes and structures used for transistors that have enabled them to be miniaturized to 20 nanometer feature sizes in the most advanced chips today. In the 2000s, Intel was the first to incorporate silicon germanium material in the source and drain regions

of p-channel transistors to induce a mechanical strain in the channel region to facilitate the flow of electronic charge between the source and drain regions and thereby enhance transistor performance. Intel was also the first to incorporate its metallic gate materials and high permittivity insulating materials to be used in high volume production to acquire better control by the gate over the channel electric potential to better suppress unwanted off-state leakage current as we scale down the gate length. Starting in the year 2011, advanced transistor structures were adopted in high volume manufacturing, first by Intel Corporation. Today, the most advanced microprocessors utilize transistors that have fin-shaped channel region because this provides for scalability to 10 nanometer gate length. These fin-shape transistors are referred to as FinFETs. Now if we want to scale the gate length further below 10 nanometers, then we need to be able to wrap the gate electrode completely around the channel region in order to suppress unwanted off-state leakage current and in order to achieve good layout area efficiency, more current for unit area, we need to stack multiple channel regions on top of each other. So, in the future, we expect that stacked nanosheet transistors will be used in high volume production to scale gate lengths to below 10 nanometers.

Now, the transistors in a microprocessor chip generally are used as simple on off electronic switches to perform digital logic functions. However, modern electronic information devices such as smartphones and tablets with functionality beyond computing, communicate information wirelessly using radio frequency signals, and they have various sensing capabilities, such as sound, motion, and light detection. These functionalities have been made possible with technological advancements in thin-film resonators and microelectromechanical devices, or MEMS, and photonic devices.

There are many technological challenges that must be overcome in order for ambient intelligence to become a reality. These challenges present exciting opportunities for innovation. For example, the electricity consumed by computing devices has increased exponentially with the proliferation of information and communication technology. To avoid a power crisis in the future, new energy efficient computing devices and information communication technologies must be developed together with innovations in integrated circuit design and chip architecture. The Internet of Things requires wireless sensing devices to be non-intrusive, easily deployed, and inexpensive in order to become a pervasive technology. For example, devices used for monitoring health and activity must be wearable and fully self-contained, heterogeneously integrating sensing, information processing, communication and storage, as well as energy scavenging and energy storage functions into a compact form factor. Finally, with increasing collection of data, there is an increasing need to be able to store and process larger data sets in real time to derive actionable information.

To address all of those challenges and opportunities, the Berkeley Emerging Technologies Research Center, or the BETR Center, at the University of California, Berkeley, is driving innovations in materials, processes and solid-state devices to enable future ubiquitous information systems. Assuming that artificially intelligent systems must always be awake, interactive and networked across many devices, it is imperative that future electronic devices must be much more energy efficient in order to be ubiquitous and compatible with flexible substrates in order to be wearable. The Berkeley Center is a hub for physical electronics research, and it involves a diversity of Berkeley professors shown here, who collaborate across the disciplines of electrical engineering, computer science and material science to build technological foundations for future electronic devices and information systems.

The BETR Center is also a nexus for interactions with companies for long-term research collaborations and knowledge transfer. Member companies have early access to innovative ideas and research results and interact directly with the students and faculty, and the students are future prospective employees. University researchers benefit from these interactions by gaining insight into problems faced by industry and society and by seeing the application of their research solve real world problems. We are pleased with how SK Hynix, a leading manufacturer of memory chips, among BETR Center corporate members. During the Center's semi-annual workshops, employees from all of these companies have the opportunity not only to meet with the researchers at UC Berkeley, but also to exchange information between companies across the worldwide semiconductor ecosystem.

The research conducted under the BETR Center is organized into various thrusts shown here. For the remainder of my presentation, I will provide a brief overview of some research projects under these thrusts.

First, research on devices for energy efficient, integrated systems in the BETR Center is conducted by Professors Jeffrey Bokor, Ali Javey, myself, Sayeef Salahuddin, and Eli Yablonovitch, and encompasses the search for alternative switch designs and transistors that can operate with much lower voltages. So, rather than with voltages close to 1 Volt, we would like to develop digital logics that can operate with millivolts. Also, we incorporate theoretical analysis and technology computer aided design tools to study, stimulate, and optimize the design of these new switches before they are fabricated in the nano-fabrication laboratory.

Examples of millivolt switches that we have developed include electromechanical switches developed by my research group. The advantage of a mechanical switch is that in the off state, they conduct zero off-state leakage current,

which is very advantageous for zero standby power consumption. Mechanical switches can also be used to implement functions with far fewer switches, so here is the circuit diagram of a 2:1 multiplexer circuit, which comprises only two relays, whereas with conventional transistor technology, you would need at least eight transistors to implement this function. One in the middle here is a micrograph of the 2-relay circuit and shown on the right are measured voltage waveforms, input voltage waveforms here, voltage versus time and measured output voltage waveform, showing the functionality of this 2:1 multiplexer operating with voltage signals that are only 50 millivolts in magnitude. So, this is approximately ten times smaller voltage than required for transistors today. So that translates to a tremendous dynamic power consumption deduction. Also, in the BETR Center, we have researchers in Professor Bokor's group developing thin-film graphene nanoribbon transistor devices. So, this is a micrograph of a very narrow width graphene nanoribbon that is less than one nanometer across, and it's fabricated with perfectly atomically smooth edges to achieve uniform bandgap energy. So, we can have good control of the transistor properties. Now, Professor Bokor's group has demonstrated transistors fabricated with these nanoribbons of graphene. This is a plot of current versus gate voltage that's showing the first functional graphene nanoribbon transistors. And because these devices have a very large surface to volume ratio, they're ideally suited for very chemical sensitive applications.

Next, I'd like to tell you a little bit about the research being done in photonics under the BETR Center. Now, silicon photonics technology today has been adopted for high speed communication of information between servers within data centers. This is because optical signals can propagate faster and with better energy efficiency than voltage signals over long distances. Light also can be used to transmit information across the chip through silicon microstructures that act as waveguides.

However, we still need tremendous improvements in the efficiency of miniaturized light emitters and also in the sensitivity of photo detectors in order for optical interconnects on a chip to be more energy efficient than electrical interconnects. So, BETR Center researchers in the groups of Professors Ali Javey and Vladimir Stojanovic, Ming Wu, and Eli, are addressing these needs by investigating the incorporation of optical antennas to enhance the spontaneous emission rate of light emitting diodes and by exploring novel approaches to alleviate the trade-offs between photo detector speed, capacitance, and optical absorption. Research in the BETR Center also entails package-level electronic and photonic device integration, the development of fast optical switches and integrated circuit design breakthroughs to increase the communication bandwidth for silicon photonic chips.

So, as an example, Professor Ming Wu's research group has developed two dimensional, or 2D, random access optical beam steering systems, illustrated conceptually on the left here, and shown in the middle is a die photo. These devices are key components for many applications, including light detection and ranging, or LiDAR, and for free space optical communications. This system is implemented using a silicon on insulator wafer substrate and comprises a 20 by 20 focal plane switch array, which is implemented with microelectromechanical switches that allow for row and column addressing to achieve the random access 2D beam steering functionality, achieving a very wide field of view, 32 degrees by 32 degrees, and very small beam diversions, switching with sub-microsecond reconfiguration time.

Professor Ali Javey has developed a very novel light emitting device design that can achieve electroluminescence across a wide range of wavelengths using a wide range of semiconductor materials. So, he does this by applying a sufficiently

large gate voltage across a metal oxide semi-conductor structure, that is alternating positive and negative, and this achieves energy band bending in the semiconductor film, so that holes and electrons, positive and negative charge, can be injected from a source electrode into the semiconductor layer. So, by injecting both holes and electrons, we create the opportunity for electrons and holes to recombine in the semiconductor material and emit light. Now, a novel aspect of this device is the use of carbon nanotubes to form a porous electrode that allows the semiconductor material to be formed last over this contact electrode. This relaxes the process integration constraints for the semiconductor material to allow a wider variety of semiconductor materials to be used as light emitters. And shown here on the right are experimental demonstrations of light emitting from this kind of a novel structure with different wavelengths, using different semiconductor materials. The range of semiconductor materials and the wavelength of light ranges from infrared to ultraviolet, so the entire optical spectrum. And this new device technology creates interesting opportunities for applications such as electroluminescence spectroscopy as a metrology, and sensing technique.

Now, in the area of flexible electronics, we know that new manufacturing and deployment paradigms are necessary in order for these flexible electronic devices to be highly dispersed and interactive. And such devices include information displays, sensors with embedded control logic circuitry. So, spearheaded by Professor Ana Arias and Professor Ali Javey, researchers at UC Berkeley are developing wearable and flexible electronics and this is enabled by the development of tools, processes, materials for role-to-role processing, for layer transfer, for high-resolution printing, and packaging.

Now, a fundamental challenge for achieving high-performance semiconductor devices on flexible substrates is that it is difficult to form defect-free crystalline semiconductor layers on plastic materials because the plastic materials have no crystalline structure that can serve as a template for the growth of crystalline materials. And also, plastic substrates generally cannot withstand very high process temperatures. Now, Professor Javey's group recently reported the breakthrough development of a technique called templated liquid phase (TLP) crystal growth, and this technique enables direct growth of shape-controlled single-crystal semiconductor material on amorphous substrates. They demonstrated the growth of crystalline Indium phosphide on polyimide, using a temperature that did not exceed 220 degrees C, so that in the future we can form this high-performance crystalline semiconductor devices on flexible substrates. Now, Professor Arias' group is developing a variety of sensors that can be fabricated on plastic using low cost, roll-to-roll printing, blade coating, or organic binding techniques. They have leveraged the phenomenon of potentiometric mechano-transduction to develop stretchable mechanical sensors that are simple, requiring only a single electrode so that you can have improved pixel density, so resolution of sensing, and therefore better data acquisition speed compared with traditional dual-electrode-mode electronic skins. Now, their devices shown here on the right were fabricated by an all-solution processing technique, and these devices exhibit ultra-low power consumption, high tunability, and a good capability to detect both static and dynamic mechanical stimuli, which shows promise for applications in robotics and prosthetics and healthcare.

Now, mobile devices and Internet of Things devices need to be able to store information in non-volatile form. This leads to the need for the development of new embedded non-volatile memory device technology. To meet this need, the research

groups of Professor Bokor, myself, Ramamoorthy Ramesh, Sayeef Salahuddin, and Professor Vladimír Stojanovic, have worked to innovate, new nonvolatile memory devices that can be monolithically integrated with digital logic circuitry. And such devices include nano-electromechanical switches, nanometer-scale magnetic and ferroelectric devices that can be fabricated with relatively low process temperatures for the CMOS circuitry.

Now, for example, Professors Bokor and Salahuddin have collaborated to develop electric current driven ultra-high-speed magnetic elements so using a current, you can switch the state of a small magnet, which is used to store information in a non-volatile manner. So, what they've demonstrated is that an electric pulse of a current that's only six picoseconds in duration can be used to switch the magnetization of a small nanoscale magnet. And if we can scale that nanomagnet to a 20-nanometer dimension similar to transistors today, then this magnet can switch with only femtojoules (fJ) of energy for very low power consumption. Professor Salahuddin's group also recently made a groundbreaking discovery of spin orbit work generated by an amorphous iron silicide and cobalt layered structure, both stated here in the cartoon on the right. So, the significance of this discovery is that we can use conventional silicide materials, which are in high volume CMOS technology today, to implement spintronics devices integrated easily with CMOS. So, this is an exciting development in the area of embedded nonvolatile memory technology.

Now, future information and communication systems set a broad context for research conducted by Professors Bokor, myself, Sophia Shell, and Vladimir Stojanovic. We have collaborative projects that bridge innovations in physical electronics and integrated circuit design or co-optimization of new solid-state device

technologies and computer architectures. Through device modeling and simulation of integrated systems, tradeoffs between energy efficiency and performance can be optimized for the design of testbeds to experimentally demonstrate the benefits of new device technologies for specific applications.

Now, one example is that Professor Stojanovic's group has developed an optical phased array technology based on wafer scale three-dimensional integration of photonics and CMOS circuitry as illustrated here. Now, 3D integration, the photonics in electronics allows the photonics devices to be highly customized independently of the electronics, which opens up unlimited opportunities for integrated free space system design. Flexible and ultra-dense connections between these two layers with through-oxide vias allow for high density of 3D interconnections and achieve a large active array aperture within a compact die. So, the optical phased array prototype achieves wide-range 2D steering while consuming only 20 milliwatts per element of average power.

Now, in collaboration with me, Professor Stojanovic has developed compact architectures for implementing low power functions. One example is shown here, where we implement a lookup table, which is more energy efficient and faster than a computational circuit. So, what we've done is we developed nano-electromechanical switches implemented using standard back-end-of-line interconnect metal layers in a conventional CMOS process. Shown here on the left is a scanning electron micrograph cross section of a chip that was fabricated at TSMC using its standard 16 nanometer process technology. We have FinFETs at the bottom here and then multiple layers of metal that then pattern to form movable switches that can be activated to form reconfigurable interconnects. So, my graduate student demonstrated these reconfigurable interconnects that retain their state in nonvolatile

manner, integrated together with CMOS circuitry to implement a lookup table, and functionality is demonstrated by the measured voltage waveforms on the right. So, basically, this technology, CMOS plus NEMS technology implemented with a conventional high-volume manufacturing process enables very low power, very high-speed circuitry for looking up data to avoid a need for energy consumption, to calculate information, and also can be used for data searching operations and decoding in the future.

The rise of artificial intelligence has precipitated the development of chip architectures that are specialized for machine learning in order to maximize performance and energy efficiency. So, these professors here are investigating hardware accelerators that are specialized for large scale matrix computations that are used for deep neural networks. Researchers in the BETR Center are also developing new hardware approaches for solving combinatorial optimization problems, such as those found in operations research, finance and circuit design. Examples include nano-electromechanical switch arrays and other novel switches and architecture-aware network tuning techniques and analog machines that can solve NP-hard optimization problems without the need for complexity of quantum bits.

Shown here on the left is an example from Professor Yablonovitch's group who has been investigating the use of continuous time analog circuits for combinatorial optimization. As an example, they designed and analyzed an electrical LC oscillator-based Ising machine, as shown here on the left. This system is a physical implementation of the well-known method of Legrange multipliers in optimization theory, which has application in control systems, operations research, as well as artificial intelligence. As another example, Professor Sayeef Salahuddin and his group has successfully used the Restricted Boltzmann Machine (RBM) as a

stochastic neural network capable of solving NP-hard combinatorial optimization problems efficiently. Specifically, the RBM structure and sampling algorithm were demonstrated for an Ising model problem as shown here to the right. These results open the possibility of using parallel stochastic computing to solve NP-hard and NP-complete problems with far reaching consequences in fields like logistics, scheduling, and resource allocation.

So, I hope I've given you the idea that even as the practical limits for transistor miniaturization are reached, alternative approaches for improving chip functionality and energy efficiency are being developed to meet the growing demand for information and communication technology. There's still much room for innovations in materials, devices and architectures to sustain the virtuous cycle of technology advancement, to usher in the age of ambient intelligence. In the future, we can all look forward to information technology that will be ubiquitous and to realize ambient intelligence that will enhance the health, safety and quality of life for all members of our global society. Thank you very much for your kind attention.

신창환:

Let me introduce the second speaker in the plenary session.

Sayeef Salahuddin is the TSMC Distinguished professor of Electrical Engineering and Computer Sciences at the University of California Berkeley. Salahuddin received his B.Sc. in Electrical and Electronic Engineering from BUET (Bangladesh University of Engineering and Technology) in 2003 and PhD in Electrical and Computer Engineering from Purdue University in 2007. He joined the faculty of Electrical Engineering and Computer Science at University of California, Berkeley in

2008. His work has focused on conceptualization and exploration of novel device physics for low power electronic and spintronic devices.

Salahuddin has received the Presidential Early Career Award for Scientist and Engineers (PECASE), the highest honor bestowed by the US Government on early career scientist and engineers. Salahuddin also received a number of awards including the NSF CAREER award, the IEEE Nanotechnology Early Career Award, the Young Investigator Awards from the Air Force Office of Scientific Research (AFOSR) and the Army Research Office (ARO).

Salahuddin is a co-director of the Berkeley Device Modeling Center and Berkeley Center for Negative Capacitance Transistors. He served on the editorial board of IEEE Electron Devices Letters (2013-16) and was the chair the IEEE Electron Devices Society committee on Nanotechnology (2014-16). Currently, Salahuddin is a fellow of the IEEE and the APS.

Let’s take a listen to his talk.

Sayeef Salahuddin:

I wanted to start by thanking the organizers for giving me this opportunity to present my work. This is truly an honor for me to be here today. So, what I'm going to talk about today is ultrathin ferroelectrics as a functional material and its application for energy efficient logic and memory devices.

So, the computing workload that we see changes with time. Today, the most dominant computing workload is artificial intelligence. We are all aware of it. But if we look back at history and we try to ask that question, what enabled this AI

revolution that we see today, here is something that I found in Wikipedia. So, on 30 September 2012, AlexNet achieved a top-5 error of 15.3%. So, this is how it all started. This was made feasible due to the utilization of graphics processing units or GPUs, an essential ingredient of the deep learning revolution. So, what this is saying is that, you know, the A.I. algorithms were enabled by this continuous advancement in hardware, which enabled the graphics processing units, and that enabled solving these large model problems in artificial intelligence. If we, of course, look into the GPU a little bit more, we see that, for example, NVIDIA Volta GPU today is built up on 21 billion transistors. In fact, it is a build-up on the FinFET technology which came out of Berkeley, which we are very proud about.

So, this shows that the computing workload that we see today is really enabled by the advancement in the hardware. And we can look into that advancement by looking at, for example, how the transistor density and SRAM bit density have increased over time. So, for example, in this chart, we are looking at both transistor density and SRAM bit density, and what we see is that both of these have increased exponentially over the last almost five decades. And this is what has enabled this tremendous computing revolution that we see today. Here, I'm showing a slide that has been adopted from a presentation from Lisa Sue, who is CEO of AMD. She gave this talk in the ERI DARPA Summit in 2019, and what this shows is that if you look at the complete performance gain over the last decade, more than 50 percent of it has come from innovations in device materials and processing. So, this shows that improvement in hardware has remained to be a very critical component of the computing revolution that we see today. And that is the motivation for people like us who work on materials and devices to continue innovation in this space.

So, in that context, what my group at Berkeley works on is ferroelectric materials as a functional material. We believe that ferroelectrics can add new functionalities to computing devices like transistors and memory devices. So, in a very simple sense, a ferroelectric material is a dielectric, which has permanent dipoles. So, this is what I'm trying to show here pictorially. For example, a ferroelectric can have up and down dipoles and therefore, a ferroelectric insulator can have two states when all the dipoles are up and when all the polls are down. So, in a polarization-versus-voltage characteristic, we see these two different states. And to go from one state to the other will require some voltage.

So, when we have a material which has a characteristic like this, where there are two different polarization states and to go from state to the other, we need to apply a voltage, its energy landscape, which is the potential energy as a function of polarization, looks something like what is shown here. It's like a "W" because the two polarization states have the same energy, so it can be either here or there. So, there are these degeneracy in the energy states and the fact that I need a voltage to go from one to the other means that there is an energy barrier between these two states. So, we have to apply this much energy from outside so that the material can switch its state and go from one to the other.

In the recent days, there has been a lot of different devices that have been discussed in literature in terms of potential applications of ferroelectric materials. For example, from Berkeley, we work on negative capacitance FET, which we think can reduce the supply voltage requirement of advanced transistors. One can also think about ferroelectric tunnel junctions and I will discuss its operating principle in a few minutes. There is also the application as a ferroelectric memory, where a ferroelectric is put on the gate of a transistor and the dipoles are used as the

memory states. There are also applications where people have discussed how a ferroelectric could be used for neuromorphic applications. So, in the recent years, a number of these applications have been discussed, and so, this is quite exciting for the electric community as a whole. However, there is a fundamental challenge in terms of adopting ferroelectricity for advanced devices. So, what I'm showing here is the micrograph of FinFETs, and what we know is that as we are advancing in our electronic devices, the pitch between them is going down and down and down. For example, for the 40 nanometer FinFET transistor, the fin pitch is around 43 nanometers. Today, we are thinking about going down to the three nanometer node transistors where the fin pitch will be very, very small. And when the fin pitch is that small, or the space between two fins is that small, there is not too much space that one can allocate for many different materials that needd to go in between those things. For example, the space in between those fins has to account for the gate dielectric, the work function metal, and then the gate metal itself. In fact, when we talk to our industry colleagues, they even say we can hardly have a 2nm gate dielectric there. So, if you are thinking about using ferroelectric materials in the gate of our advanced transistors, we cannot allocate more than two nanometers for that space. Now, on the other hand, if you look at the history of ferroelectricity, in fact, we have known about electricity for more than 100 years. Just last year, we actually celebrated the centennial of ferroelectricity. And in this very long history, one thing that we have learned is that we cannot make the conventional ferroelectrics very thin. So, for example, this is showing the trend of the ferroelectricity as a function of its thickness for conventional perovskites ferroelectrics. So, here we can see that by the time we are below 20nm ferroelectricity, it starts to become weaker. And by the time we are around 6-7nm, the ferroelectricity is very, very weak. Below5nm, it is almost gone. On the other hand, we need to be around 2nm thickness for advanced

devices. So, this seems to be a fundamental challenge which one needs to overcome if ferroelectricity has to be used for advanced devices. So, that is something that we wanted to look at, especially for new types of ferroelectrics. For example, it is now known that doped hafnium dioxide can be a ferroelectric. So, these materials are process-compatible with silicon transistors. So, we wanted to look into these materials and see how thin we make them before they lose their electricity.

So, in that pursuit, what we have seen is that, in fact, even just 1nm or just two unit-cells of ferroelectric materials, of hafnium dioxide, can be made ferroelectric. So, what you see here is an amorphous silicon dioxide sitting on silicon, and on top of that is zirconium-doped hafnium dioxide of the thickness of around 1nm, and what we have seen is that this 1nm film is a ferroelectric material. So, a ferroelectric material is also piezoelectric, so we can use its piezoelectric property to actually write phase contrast on the film. So, you can see that this is the "Cal" logo that we actually wrote down on that 1nm film. And this showed us that even 1nm films are very ferroelectric. Now, of course, the proof of ferroelectricity for such thin films cannot be only done using the piezo-force microscopy. So, we actually looked at electrical measurements and in the electrical measurements, we see this very nice butterfly-looking curves. And this hysteretic behavior is something that I just showed you a few slides ago. So, we actually detect very strong electricity in these very thin ferroelectric films.

Now, one thing that is remarkable is that, as I discussed, in conventional ferroelectrics, as we reduce the ferroelectric thickness down, the ferroelectricity gets weaker. However, in these new films, what we have seen is that as we reduce the thickness down, the ferroelectricity actually starts to get stronger and stronger and stronger. So, what this shows is that this new type of ferroelectrics, which is doped

hafnium dioxide material, is fundamentally different from the conventional ferroelectricity. So, therefore, these materials are actually very, very suitable for advanced electronics, because now we have shown that this fundamental challenge of pinning down the ferroelectric, which is critical for advanced electronic devices, does not cause a problem for this new type of related materials. And in fact, as thin as a 1nm film is still a ferroelectric.

So why can we make these films ferroelectrics where the conventional ferroelectric materials lose their ferroelectricity at a very thin thickness? So, what we have, what we believe, and this is our hypothesis, is that we are able to grow these films on an amorphous and atomically smooth SiO2. And so, what happens is, as we are reducing the film thickness, the strain from the top metal becomes stronger and stronger. And this is why as we reduce the thickness of the of the film, the ferroelectricity actually becomes stronger and stronger in these materials compared to conventional perovskites, which are usually grown on an epitaxial substrate. And therefore, the strain is more or less determined by the substrate rather than what is sitting on top. So, as an as an evidence to that hypothesis, what I'm showing here is that, again, this butterfly curves for two different regions on the same film. One region had this metal electrode, which is used as a stressor, and the other region did not have any metal electrode. And what we find is that when we do not have the metal electrode, which can stress the film, we do not really see any ferroelectricity. On the other hand, when the film is capped with the metal, we see the ferroelectric behavior in those regions. However, the exact mechanism still remains to be an open question. On the other hand, the fact is that at this very thin, 2-unit-cell thickness, these materials are very strong ferroelectrics.

So, now that we have a very thin ferroelectric material, we can now start to look at their applications. So, the first application that we wanted to look at is a tunnel junction. We have just 1nm film now, so we can look into tunneling through them. The interesting thing is when a tunneling happens through a ferroelectric material, the tunneling current can depend on the direction of the polarization. This can be understood in this way. Let's say that I have a thin ferroelectric between two metals. So, if you think about the doped hafnium dioxide, it has 6V of bandgap. So, one can look at these ferroelectrics as a high-bandgap semiconductor. And so, if we look at these metal oxide junctions, there is a s Schottky barrier that is present at the metal oxide surface. And the height of that barrier depends on the polarization. If the polarization is pointing from left to right, then we will have the positive charges here, the negative charges here, and that will reduce the barrier in this direction and increase the barrier height on the other junction. And when we switch the polarization, the opposite will happen. And as a result, the tunneling current that flows through that junction will be will be different. So, this is often called ferroelectric tunnel junction. Now, if you think about this, if we are going to use this as a memory, we really want very clean ferroelectrics because if the tunneling oxide is very thick, we are going to get very small current flow into the junction. As a result, it will be very difficult and very slow to read that memory. So, we want the current flowing through this junction to be reasonably high. And because of that, we want very thin oxide. All right. So, if you look at what we see in terms of our one nanometer film sandwiched between two electrodes, what we are doing, this shows the pulse scheme that we use for this memory. We first write a state, then we use a small voltage to read the state. Again, we write the state on the opposite direction. So, this has an opposite polarity. And then, we use again a small grid voltage to read the state. And what you can see that when we do that, we actually see two

different amounts of current flowing for two different states and we can switch the materials from one state to the other by applying the voltage. So, what we have found is that we can get around 10A/cm2 tunneling current with more than 10x on-off ratio. This combination is actually one of the highest that has been seen for any ferroelectric junction. So, we are quite excited about what we are seeing with these very thin ferroelectric films. Again, these films can be directly grown on silicon and back in compatible manner. So, this could actually become a viable memory solution from that point of view.

Now, the second application that I want to talk about is negative capacitance transistors. Now, in the interest of time, I'm not going to go into the details of negative capacitance operation, but in a very brief manner. We already talked about this energy landscape that ferroelectric material has. And so, if you look at this middle region, in this region, the capacitance of the material is actually negative. And to see that, we can recognize that potential energy is E·dp. And so, if we want to look at the polarization versus electric field characteristic, all that we need to do is to take a derivative of the potential energy function. And that's what we have done here on the left. And so, this "W" after we take the derivative, starts to look like an S curve. And remember, this is P, which is charge and this is electric field. So, when we look at the slope of polarization with respect to electric field, which is capacitance, in this region, within this yellow box, that capacitance is negative, which actually corresponds to this middle region. Now, the question is, what can we get if we have a negative capacitance? So, think about the following situation. Let's say that I have a transistor and I used the ferroelectric biased in this negative capacitance region as my gate dielectric. So, MOSFET will then be the series combination of the gate dielectric and the semiconductor capacitance and all other capacitances like the short channel capacitances and others that are represented by

C2. So, if you look at the surface potential, so I can write the change in the surface potential as C1 divided by C1 plus C2 times the change in the supply voltage, and if C1, my gate insulator, was negative, dψs can be written as C1 divided by C1 minus C2 times dVG. In other words, dψs can be larger than dVG and that means that an amplification is possible. So, the surface potential in the channel can change faster than the change in the supply voltage, which means that if a conventional MOFET was giving me this log IDVG of this black curve, for the negative capacitance, we can expect amplification of this surface potential and we will see this red curve. In other words, I could get to the on current at a smaller voltage for the same off current and that will reduce my supply voltage requirement.

So, I have already discussed the fact that for advanced transistors, we need very thin—less than 2nm—of ferroelectrics. So, for the negative capacitance operation, if we want to be in the most relevant regime, we want to use a ferroelectric, which is less than 2nm. And now that we know how to synthesize these 2nm ferroelectrics, we developed these SiO2 transistors with just 1.8nm HZO, which is the zirconium-doped hafnium dioxide. And this is showing the experimentally measured drain current versus gate voltage characteristic. So, these red lines—in fact, it looks like a thick line because we have measured many devices, so lots of devices are present here. So, these red lines are for negative capacitance FET, which has this 1.8nm ferroelectric in there. And these blue lines are for hafnium dioxide, which is the conventional dielectric, and we call that the reference device. What we can see is, compared to this conventional transistor, the negative capacitance FET gives us a better sub-threshold swing, and it also gives us a lot of off current. What you can see is that, if I move these red curves to the left so that both have the same off current negative capacitance on this transistor will give me the same on current at a smaller voltage. OK, now one could look at this and one could say that, "OK, your

sub-threshold swing is better, the slope is better. This basically just says that the electric has a better dielectric. It has just a high K." OK. However, what I want to point out is that this is actually much more than a high K, and to see that we have to get into the device, which is a little bit more.

So, in transistors, the lowest current that we can get is usually determined by gate-induced drain leakage. We often call this the GIDL current. So, the idea is that if I look at the energy band diagram, at the drain side, as we are reducing the gate voltage, this barrier is going up and up and up. And as a result, electrons can tunnel from the body to the drain, giving us these little currents. So, if we use a high-K material, it not only increases the gate control, it also increases the electric field at this drain end, which means that we attain a higher permittivity material, not only the slope will get better, but the off current will actually degrade because we are going to increase the GIDL current. So, this is actually using an industry standard TCAD simulator, which is showing that if this blue line is our reference, with an increased permittivity, we will see this ash line where the swing has improved, but the off current has degraded. Compared to that, what we see in our negative capacitance transistors is that not only the swing has improved, but the off current has actually improved. So, this is not a typical high K. We cannot use classical electrostatics to explain this behavior.

If you look further into this, what we find is the sub-threshold swing for the negative capacitance transistors actually initially improves with increasing drain current. This is again, not explainable by classical electrostatics. For example, if we look at our baseline MOSFET, the sub-threshold swing continuously increases with increasing drain current. And this is exactly what is expected from classical electrostatics. But if we look at the negative capacitance transistors, it initially

improves and then it gets worse. So again, what we are observing here is non-classical electrostatics. On the other hand, in this industry standard TCAD software, if we use a ferroelectric model together with the electron transport in the transistor, we actually get very good agreement of our model with the experimental data. So, for example, this red line and the blue line—this is actually not an average drawn through this experimental data. These are our modelling for these devices. And as you can see, that we get a very good behavior between the data and the model.

So, we calibrate our model to one gate length, but to check the robustness of this model, what we have done here is we have actually plotted the model calculations and experimentally measured data for various gate lengths and for two different drain voltages. So, this is a small drain voltage, and this is a large drain voltage. And as you can see, by calibrating to one gate length, we were able to actually explain behavior seen in many different gate voltages, in many different channel lengths, and for two different drain voltages. So, this gives us the confidence that the model, which we use as a ferroelectric model, is able to capture the main physics of the device.

OK, so this is what I showed a few slides ago. This was our motivation of using ferroelectrics on top of silicon transistors that we will be able to steepen this curve, and as a result, we will reach the on current at a smaller voltage than a conventional transistor. This is one of our better transistors, a small channel length—30nm channel length—transistors with this very thin ferroelectric on top. And we are seeing this effect. We are essentially steepening the curve enough that we can get to the on current at a smaller voltage. But at the same time, our off current is also lower. So, we have essentially improved the on off ratio by orders of magnitude at a smaller voltage. And so, this is the promise of the negative capacitance transistors.

The great thing is that the ferroelectric material that we need to achieve this has already been tested in production tools. So, for example, in 2017, GlobalFoundries presented a result with ferroelectric on top of their 40nm FinFET transistor. So, in terms of very large-scale integration, there has already been a demonstration in that direction. So, again, the point that I want to make is that for continued advancement of our electronic devices, we need to go to very thin dielectric materials. Ferroelectrics can add new functionality, but it needs to do that at a very small thickness. And now, we have demonstrated that that is indeed possible.

So, one last topic that I wanted to present is a ferroelectric memory, which has been a dominant area of research on ferroelectricity over the last few years. And the basic idea is simple. The basic idea is that ferroelectric materials, as we discussed, have permanent dipoles. So, if you put them on top of a transistor and can polarize them enough, then these permanent dipoles will shift the VT of the transistor, and as a result, we will see these two shifted current voltage characteristics depending on which direction the polarization is pointing to. And that can give us a memory. It can give us a very, very dense memory because a single transistor ferroelectric memory is possible.

Now, again, in terms of integration, as we discussed for the NCFET, the material, doped hafnium dioxide-based ferroelectric memory has already been demonstrated in very large-scale integration—for example, in GlobalFoundries, 22nm FDSOI platform. So, we can see that in this scale technology, we see very good memory behavior, reasonably good variation. But one of the big problems in these ferroelectric memory devices is the endurance. So, as you can see that, by the time we have switched the device or written to this device 10,000 times, their behavior starts to significantly degrade. And by the time we are around 105, the memory

window has collapsed. In other words, there is no way anymore to distinguish between two states. Now, again, we can get into the discussion of why this happens. But in the interest of time, let me just say that there is some agreement in the community that there are two different effects that caused this to happen. One is that polarization creates a lot of electric field in the interfacial oxide, and breaks down that interfacial oxide. As a result, the device breaks down. The second is that as we keep cycling these devices, the ferroelectric materials that have been traditionally used even in this paper, for example, are 6-7nm thick. So, as we are writing to these materials, as we are applying voltage, we start trapping charges at the bulk and this bulk charged trapping acts in the opposite direction of the ferroelectric polarization. And so, as we keep going and we push more and more charge into the bulk of the material, at some point it completely nullifies the hysteresis or the memory window that we see from the ferroelectric. So, to address these two problems, recently—this paper is still in review—what we have done is we have used a high-K, high-permittivity interfacial layer. The idea is that the high permittivity will allows us to reduce the amount of electric field that the interfacial layer sees, because that electric field gets scaled down by the permittivity of that material. The second thing that we did is, again, because we now know how to do very thin ferroelectrics, we tried to grow a 4nm ferroelectric rather than 7 or 8nm ferroelectric on top of this material. What this does is it significantly reduces the bulk charge trapping. This is very well understood from high-K gate interfacial layer for advanced transistors. So, when we combine these two, this is data that shows the cycling endurance. In other words, we are writing them back and forth. And you can see that this device gives us pretty good separation between the two states, even beyond 1010 cycles. So, this is almost five orders of magnitude larger than what we have just seen in the previous slide. So, this, again, shows that the thin

ferroelectricity, combined with a better interfacial layer, can actually make this ferroelectric memory to have very high endurance. And in fact, beyond 1010, this memory starts to become interesting for last level cache applications. So, this is, again, an exciting area of application for ferroelectrics.

All right, so with that, let me conclude my talk. What I tried to emphasize is that ferroelectric materials bring the promise of new functional augmentation to our electronic devices. However, one of the major challenges for ferroelectric materials has been the fact that it's very difficult to reduce its thickness. On the other hand, we really need 2-4nm ferroelectric material for our most advanced and scale devices. So, in the last few years, we have been working on this problem and what we have been able to show that, for doped hafnium dioxide ferroelectric materials which are process compatible with silicon, even 1nm thick films are strongly ferroelectric. And we have shown all the major applications of ferroelectricity with this thin ferroelectric. We have shown the ferroelectric tunneling junction, we have shown the negative capacitance transistors with better on-off ratio and a smaller operating voltage. And we have also shown ferroelectric memory devices with very large endurance. So, on that point, I'll just say that we do not see any fundamental bottleneck in terms of thickness for the ferroelectric materials, and therefore we think that this material constitutes a very interesting material system for our future devices. With that, let me stop by thanking my students and post-docs from the last few years and also of Professor Chenming Hu, who has been a longtime collaborator with us on all types of ferroelectrics research. Thank you.

Keynote Session

신창환:

OK. Now, I'm going to invite two scholars in Korea. Let me introduce the first speaker in the keynote session.

Changhwan Choi is a professor and head of the division of materials science and engineering at Hanyang University. He received B.S and M.S degrees in materials science and engineering from Hanyang University and The University of Texas at Austin, respectively. He also attained Ph.D degree in electrical and computer engineering from The University of Texas at Austin. Before joining into school, he had worked for IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA, as a research staff member, where he did research on semiconductor material and device fabrication processes, characterization and modeling for 45-, 32- and 22 nm technology. His main interests are nanoelectronic devices, materials and process developments for advanced memory and logic device applications.

Let's take a listen to his talk.

최창환:

Good afternoon. My name is Changhwan Choi, from Division of Materials Science and Engineering of Hanyang University. First of all, it is great honor to be invited by Chey Institute for Advanced Studies and I am also happy to share my research result. Today I will introduce 3D integration technology that can be one of the alternatives to the scaling limit of semiconductor devices.

This is outline of my talk. I will start with the motivation of 3D integration. Why do we need to shift toward 3D integration? I move on 3D heterogeneous

integration (HI) with some examples, which is already implemented in semiconductor industry. Then, monolithic 3D integration will be introduced. Particularly, I will present one example of M3D application, which was presented at 2020 IEDM from our lab. I will continue to show another research result of M3D or 3D integration from our lab. That is 3D neuromorphic system consisting of CMOS and memory array. Then, I'll wrap up the presentation with summary.

Semiconductor technology have evolved by introducing a variety of materials, process technologies, different device structure, designs and system. This chart shows the number of transistors per unit area as a function of year, released by ERI, 2019. Historical change in semiconductor technology is expressed with Lithography, Materials, Device, Design, and System. In terms of scaling, we can say classical scaling period is down to 90 nm. Beyond that, it belongs to effective scaling period. Let's look at each area in details. For lithography, KrF and ArF have been used and and we have EUV now. Further device scaling is expected. For materials, SiGe at 90nm, High-K/Metal Gate at 45 nm were big impact. Besides these materials, other materials are already implemented, and more materials are to be expected. For device, planar device was changed to 3D FINFET at 22 nm, and improved FINFETs with taller fin are being implemented. Below 5 nm or 3 nm, one more device structure change is expected with gate all around. For design, design technology co-optimization (DTCO) is driven and different computing systems are expected like neuromorphic computing. For system, 3D integration like heterogeneous and monolithic 3D integration are expected within this decade.

In 2020, ERI released another report. Again, performance or number of transistors with respect to years has been divided by four different periods. Geometric scaling [using DUV lithography], BEOL scaling [using low-k/Cu, Bus width

scaling], 3D devices [using FINFET, High-k dielectric, strain engineering, multi-core architecture]. Now, we are in the heterogeneous integration using alternative material, domain-specific function, 3D architecture. They emphasize on 3D integration again. So, 3D integration is still predicted as an important driver in semiconductor industry development.

As we already know, semiconductor devices have improved performance through scaling, but scaling is difficult due to an increase in process complexity. The difficulty of device scaling is not only a technical issue, but also a cost aspect. Cost reduction per transistor start to slow down around 28 nm technology. In terms of device fabrication cost, many fabless companies argue that the gain in performance improvement over investment is not that significant after 28 nm technology node. Cost-effective scaling will continue. Besides scaling strategy, a new approach is still needed.

In addition to scaling semiconductor devices, power scaling should be also considered. As semiconductor technology nodes advance, transistor delay decreases, but interconnection RC delay increases, leading to problems of performance and power consumption. Same trend is on the right chart. Wire-driven delay is more dominant as we go into advanced technology node. So, this wire-driven delay becomes a concern as the wire length increases and 2D scaling cannot alleviate the interconnection problem.

We also need to think about increasing huge data and various functionalities. Every day we generate very large data volume from social network service (SNS) using a variety of electronic products such as high-performance computing, mobile electronics and various sensors. How do we manage these ones? Various

functionalized & diverse novel electronic devices should be connected further. So, we need both device scaling and diverse functions.

This slide shows an example of versatility in various electronic devices. Mobile processor, Mobile-wearable, High performance processing. Stacked DRAM, HBM, RF chip, GPU core and so on. These various types of semiconductor devices are required to perform various functions, and these have been integrated in such a manner as package-on-package (PoP), system in package (SiP), wafer level package (WLP) and Partitioning. All are related to packaging and 3D integration.

So, what pathways should we take in semiconductor development? 4 different ways. Continuous geometric scaling, mainly driven by lithography. For example, Extreme ultra-violet lithography (EUVL). Device scaling. Device structure changes planar, finfet, gate-all-around below 5 or 3 nm. Circuit scaling. System on chip or advanced packaging. In this category, 3D heterogeneous integration belongs. Architecture scaling, for example, memory hierarchy change with storage class memory compared to current architecture. With this regard, I'd like to mention about the importance of 3D integration for the semiconductor development.

As we are trying to solve the residential limitations of urban spaces with high-rise buildings, limitations of 2D scaling can be solved through a three-dimensional integration process. Furthermore, by reducing the length of the interconnection by 3D integration, RC delay and power consumption can be reduced, and the functionality can be maximized by stacking devices with various functions in 3D. Historically, discrete devices have been integrated, and now 2D integrated devices are expected to be integrated in three dimensions. Considering scaling limit, power issue, various device functions, and cost, 3D integration method will become more and more important.

The DARPA-ERI program has started investing heavily in four areas since 2018. New materials & devices. Specialized functions. Design & security. 3D Heterogenous Integration. We can get an insight from their research program. Again, 3D heterogeneous integration is one of them. So, we are on the same page that 3D integration is important.

In general, the 3D integration process is called 3D packaging or heterogeneous integration by those who work on the packaging side. However, in a larger perspective, the 3D integration process can include both 3D packaging and monolithic 3D. Some examples of 3D packaging include Si interposer, memory stacking, and HBM, while M3D includes 3D SOC.

I'd like to show you actual images of semiconductor chips fabricated by 3D heterogeneous integration. This slide is an example of early 3D heterogenous integration in the fields of image sensor, interposer stacking, and memory stacking. As time goes on, you can see that it becomes a more complex 3D structure.

Then, what is the heterogeneous integration? That is Integration of separately manufactured & tested components into higher level assembly, which provides enhanced functionality & improved characteristics. For example, here is an intel agilex FPGA chip, and various types of ICs are integrated with each other. For example, 10nm based FPGA, HBM, other chiplets connected via EMIB called embedded multi die interconnection bridge technology.

Let's go over the definition of 3DIC, package, and board before going for next slides. It depends on the purpose and applications, but when the semiconductor chip is disassembled, the structure is as shown in general. Here are a PCB board and a package board. Looking at the square area. Memory or device is connected via

TSV (Through Silicon Via) and stacking memory is connected to the silicon interposer by μ-bump, and another IC is also connected to the other area. This area is called a 3D IC, and the 3D IC connected to the package substrate by a C4 bump is called a package. These terms are helpful to understand heterogenous integration.

Here are some examples of chiplet integration. Chiplet integration, which connects various chips by placing them on a single board, is the core technology of 3D heterogenous integration. FCMCM, flip chip multi-chip module, two SOCs are placed on the package substrate. 2.1D or 2.5D, different chips like SoC, HBM, RF and so on are placed on the medium called thin film or interposer. Then, they connected each other. EMIB, this is a kind of 2.5D. A method of using only a specific area rather than using all interposers. FOEB or FOMCM. 3D structure. Different chip fabricated using different technology node. In this way, chiplet technology can improve functionality by integrating various types of chips together.

In order to connect various types of chips, we need to increase the interconnection density. This interconnection density is affected by bump pitch and bump density, which influences power consumption. As the pitch length was reduced, a variety of 2D and 3D technologies were introduced. As package changes from standard to 2.5D and 3D heterogeneous structure, bump pitch reduction, bump density increase, and power reduction are possible. In other words, higher interconnection density is required to attain high memory bandwidth and power efficiency. Therefore, companies are developing the way to increase the interconnection density, and each company has own unique technology. From the next couples of slides, I will shortly introduce some technologies reported from Intel and TSMC companies.

Intel reports 3 different technologies of 3D heterogenous integration. EMIB, Foveros, Co-EMIB. In Embedded multi-die interconnection bridge (EMIB), CPU, Memory or other I/O devices are connected by EMIB, this red connection part, on the package substrate. This is a kind of 2.5D interposer technology. In this way, they can increase bandwidth density compared to on-board package. In Foveros, various functional chips connected by face-to-face via μBUMP on the interposer and connected via TSV to package. This is 3D integration structure. Co-EMIB is the hybrid technology using EMIB and Foveros.

Here are actual cross-section images of those 3 technology-based structures. EMIB, No full interposer. FOVEROS, Interconnect density scaling & lower parasitic. Co-EMIB. More functional partitioning.

TSMC also reports two different approaches, but their methods are different from intel. For example, SOC or Memory are die-partitioned and making chiplets. In here, different technology can be used for various chiplets. Two ways, 2.5D approach, CoWoS or InFO. 3D approach, SoIC. I'll shortly explain these ones on the following slides.

CoWoS, Chip-on-Wafer-on-Substrate. HBM connected to Si interposer via μ-bump and silicon interposer connected to package substrate via C4 bump. This technology can provide multiple dies side-by-side on a Si interposer for better interconnect density and performance.

InFO, Integrated Fan-Out. This is also a kind of 2.5D interposer using RDL (Redistribution layer) and TIV (Through InFO). There are two types, InFO-oS and InFO-PoP. Particularly, InFO-PoP has better electric and thermal performance due to

thinner profile. This technology was mainly implemented to mobile networking application. For example, apple iPhone.

SoIC, System on integrated chip. Two types. Wafer on wafer (WoW) stacking, Chip on Wafer (CoW) stacking. It can use wafer-scale FEOL process, similar to SOC. Thus, better form factor and performance are expected. This technology would be implemented this year according to reports. The basic is to increase bump density. Bump density as a function of bump or bonding pitch. As reducing the pitch, higher bump density. From Flip chip to 2.5D/3D IC to SoIC, more bump density is possible. I'm going to introduce alternative 3D integration scheme, monolithic 3D integration, later. That can belong to SoIC+, which is possible to offer the highest bump density.

Depending on applications, TSMC will use different approaches. InFO with SoIC is targeted for mobile application. CoWoS with SoIC is for high performance computing.

As can be seen from the examples from Intel and TSMC, 3D heterogeneous integration provides higher interconnection density, lower power, lower latency. That is why this heterogenous integration has received a great attention in these days. Let me move on the alternative 3D integration topic from the following slides. That is monolithic 3D integration (M3D).

The importance of the 3D integration process was already explained earlier, but a new 3D integration process is still needed because it requires versatility, more data processing and low power consumption.

There are two types of 3D integration. 3D parallel integration, TSV, which is already implemented in 3D heterogenous integration. This is 3D integration using Via connection between devices fabricated on the top and bottom substrates. It is

superior to wire bonding. 3D sequential integration, M3D. Device fabrication on the one substrate, followed by separate substrate formation on the bottom substrate using bonding, recrystallization, EPI or other methods. Then, formation of devices on the upper substrate and connection with device at the bottom substrate monolithically. Also, without upper layer formation, direct device formation is possible on the bottom substrate. These two process techniques are not competing each other. Depending on applications, there is more suitable processes. However, these processes have their own issues. For TSV, hard to fill Via, high resistance, thermal problem, bonding alignment. For M3D, low temperature process should be newly developed not to degrade device at lower layer. Also, intrinsic thermal dissipation in 3D structure.

Here is comparison chart between two schemes again. Main difference between two processes is contact size and pitch size. Compared to TSV, high Via density, small Via size and shorter interconnection length is possible using M3D. Thus, less RC delay and power consumption are expected. But, again, device fabrication on the upper layer should be processed at low temperature, not to degrade the devices at the lower layer. This is quite challenging.

I have emphasized on the usefulness of M3D integration so far. Then, what area can we apply with this M3D technology? Potential applications are to be 3D imagers, high performance, memory/computing (Computing in Memory, CIM) and neuromorphic computing.

When it comes to the formation of the upper layer in M3D scheme, various methods have been reported. For example, Using SOI wafer, flip the wafer and bonding with etch stop. This way is not limited to Si. Possible to compound semiconductor. Instead of substrate transfer, amorphous Si deposition, followed by

laser crystallization. Using ion implantation, Epitaxial Lift-Off (ELO). Of course, M3D without the upper layer formation is also possible depending on the applications.

In M3D research, there are two representative locations. TSRI @ Taiwan. They form amorphous Si first, followed by different laser anneal to attain epi-like Si layer formation. Using this scheme, they reported various device applications. CEA-LETI @ France, they use SOI, silicon-on-insulator, wafer and direct bonding. Also, various applications were reported.

The IEDM conference is the most prestigious conference in the field of semiconductor devices. So, I searched with keyword "monolithic" in 2020 IEDM. It turns out that there are 37 papers related to M3D integration, indicating M3D has garnered a great interest. For example, using M3D integration, CEA-LETI and TSRI demonstrated neuromorphic computing. TSRI also reported BEOL M3D integration. Intel presented n-GaN on the P-Si to maximize GaN FET. IMEC demonstrated p-FINFET on the n-Nanosheet FET. Let's look at more examples of M3D integration schemes reported from different locations on the following slides. Then, we can get an insight more on this technology.

From two major semiconductor conferences, Symposium on VLSI Technology, and International Electron Devices Meeting, last year. I pick up some examples. GaN and Si transistor using M3D heterogeneous from Intel. 3D complementary FET (CFET) from IMEC. GAA CMOS with FeFET memory from TSRI. 1T1R MLC for neural network from CEA-LETI.

Various M3D applications. Si monocrystalline channel in the form of N over P or P over N stacking. Alternative channel orientation to boost mobility like 110 p-Si

over 100 n-Si. Energy efficient computing, N over P with alternative channel like n-III-V/p-Ge.

Alternative computing – quantum computing. Multi layers of CNFET/RRAM/RRAM/Si-CMOS. Thin film transistor on Poly-Si channel. BEOL oxide semiconductor transistor in the metallization level. Stacked 2D materials. So, now we can understand that many research institutes are working on M3D technology and trying to apply it for various application fields. So far, I have introduced M3D and the relevant research results. From the next slides, I will show two results of the research conducted in our laboratory. The first one is CMOS + RRAM + Photosensor system using M3D and the second one is a 3D neuromorphic system using a CMOS+RRAM array.

This result was presented at last year's IEDM conference. A current sensor and a ring-oscillator were formed in the lower part, and a photosensor and a resistive switching memory device were formed on the upper layer, which was prepared by an ion-cut method. Upon light exposure to photodiode, the generated electric current transferred to the lower device and the pulse transferred to the upper resistive switching memory devices. Electrical characteristics of this system was evaluated by changing light intensity and pulse conditions.

Unlike TSRI and CEA-LETI, we adopted ion-cut method to transfer the upper layer. First, hydrogen ion implantation was carried out, then, the wafer was bonded after controlling the surface energy to improve the bonding strength. The following was to do heat treatment for the formation of cleavage, leading to the thin Si layer transfer. Finally, the surface roughness was controlled by subsequent CMP, chemical mechanical polishing. Then, we formed the devices and did electrical characterization. On the following slides, more detail information will be explained.

The transferred Si thickness is determined by the hydrogen ion implantation conditions. We introduced hydrogen ion into Si substrate to induce compressive stress inside. Compressive stress was caused by controlling the amount of hydrogen ion implantation. Increasing dose induces high compressive stress. However, too high quantity will rather cause crack formation.

We also adjusted implantation power and got an optimum. Depth profile changes with implantation power. We can clearly see the cleavage after annealing as a result of hydrogen ion diffusion.

Since two layers are bonded each, bonding strength is quite critical. To improve bonding strength between lower and upper layers, surface treatment was done by plasma. After plasma treatment, much improvement was attained on both bulk-to-bulk substrate bond and bulk-to-CMOS substrate.

Between the upper and lower substrate, the surfaces are oxide. So, bonding strength is surely affected by oxide types. We prepared different oxide types and process before bonding. T+C+H combination process offers the reasonable bonding strength, which was evaluated by surface energy. Carbon within oxide could degrade the surface energy, leading to delamination after bonding. Therefore, careful process and choice is needed.

After cleavage splitting with heat treatment, 3.5 um-thick upper Si layer transferred on the bottom substrate. To improve the surface roughness, delicate Si CMP was carried out and finally 70 nm thick thin Si layer formed on the lower CMOS device wafer.

After ion-cut based M3D integration, we used photodiode, CMOS, and resistive switching memory devices to evaluate whether 3D integrated systems

properly function. On the upper layer, photodiode was verified first. We confirmed that current and capacitance are well-responsive to increasing light intensity. In addition, in the case of M3D, the electrical characteristics are affected by the thickness and length of the metal wiring. So, the current loss was confirmed by changing metal line thickness and length. The current loss rate increases as the length or thickness of the metal wiring increases.

On the lower layer, the device performance characteristics were secured by measuring proper IV characteristics of CMOS and attaining the linear frequency dependence with increasing current in the current sensor and 21 stage ring-oscillator devices. In addition, we analyzed the signal delay due to the increase in the interconnection length. We have ring-oscillators with direct cell and delay cell. In delay cell, additional interconnection line was added. Compared to ring-oscillator with direct cell (Black), the clock frequency operation speed decreases from ring-oscillator with delay cell (Red).

When the photodiode is exposed to light, it generates a current, then converts it to a voltage signal through a current sensor, operates the 21stage ring-oscillator, and amplifies it by the last amplifier so that the output frequency can be checked in the oscilloscope. Compared to weak light (Red), a faster frequency waveform was observed with strong light (Blue), and it was also affected by the current loss parameters mentioned in the previous slide. At the same light intensity, longer interconnection induces the higher the current loss as shown previous slide, leading to the lower the frequency. Interconnection line and width are a matter to be considered when implementing M3D integration.

The output frequency was extracted using the current generated from photo-detector at the upper layer, this output frequency was applied to the resistive

switching memory device (RRAM) on the upper layer. RRAM device consists of Ti/Ta2O5/Pt structure. Under different gate interval pulse conditions like 2 μs, 250 μs, and 1 ms, the amount of current increases as the light intensity increases (Black to Red). The current increase was the highest at the 250 μs gate interval pulse, and the lowest current increase was at the 2 μs interval. It is considered that the output frequency by light and the phase of the 250 μs interval pulse are well-matched, leading to the highest amount of current. So, we demonstrated well-functional ion-cut based M3D integrated system using photodiode, CMOS and resistive switching memory devices.

Let me move on another example from our lab. That is 3D neuromorphic system. The current von-Neumann architecture-based computing causes a bottleneck between memory and processor. That is, in the process of exchanging huge data between the processor and memory, sufficient processing speed cannot be achieved. Therefore, new computing systems such as quantum computing or neuromorphic computing system are required. For the neuromorphic computing, which can provide higher data processing and lower power consumption, various hardware-based artificial synapses and neuron devices are being studied.

Neuromorphic systems are being studied in various types of devices such as RRAM or flash array, and neural networks such as CNN and BNN in terms of circuitry. Also, RRAM-based devices are being studied to implement a 3D neuromorphic system.

Our group implemented a 3D neuromorphic system by stacking an RRAM neuromorphic array on a CMOS wafer through 3D integration scheme, CMOS device at the lower layer generates and drives pulses along the configured TEG and transmits them to the upper synaptic elements (RRAM array). It is possible to select

the desired target element among the special synaptic arrays. In this way, the change in the current level of the device varies according to the weight of the pulse delivered to a specific device, so neuromorphic characteristics can be secured.

Here is the simplified process flow and actual image of 3D neuromorphic chips. The lower part is CMOS device on the 8-inch wafer, consisting of digital controller and device control switch, and the upper part is a 12 x 14 RRAM array. The top and bottom were inter-connected and the system was evaluated after packaging on the board.

Neuromorphic characteristics were verified according to the TEG designed on the CMOS wafer at the lower layer. 12 x 14 RRAM array devices were also evaluated, and diffusion memristor characteristics were confirmed.

From the 3D neuromorphic system, various synaptic characteristics were evaluated. For example, the pulses transmitted through the lower CMOS wafer showed different spike-rate dependent plasticity (SRDP) behavior depending on the weight of the current level. As the synaptic weight increased (Black to Blue direction), the current gradient changed, and through this, a learning curve was obtained. Also, we obtained the forgetting curves as a result of measuring the change in the current level under the read voltage after strong stimulus and the relaxation time. Pulse-paired facilitation (PPF) and Post-tetanic potentiation (PTP) characteristics were also obtained. By analyzing the learning rate according to the synaptic weight, we check the exponentially increasing characteristics, which means that we can imitate the characteristics that appear in the human brain. As a result, when the "H" and "Y" letters were trained on the array and checked after the relaxation time, it was confirmed that the "H" and "Y" letters learned by the strong stimulation appeared clearly.

Ok, here is conclusion of my talk. I would like to emphasize that 3D integration can be an alternative scheme in terms of performance, cost, data and functionalities. Bump pitch, bump density, and power should be considered for the higher interconnection density. Each company has a unique 3D heterogeneous integration (HI), which has advantages/disadvantages, and purse different approaches depending on the mobile or HPC applications. 3D HI is really worthy to focus on the future semiconductor, along with FEOL device/process development. M3D can allow higher interconnection density and lower power consumption. CMOS image sensor, memory in computing, and neuromorphic computing can be real application fields using M3D integration technology

Before finishing my talk, I would like to deeply express my gratitude to current and past group members who have helped me a lot in this research work. Thank you for listening my talk.

신창환:

Let me introduce the second speaker in the keynote session.

Jae-duk Han is an Assistant Professor of Electronic Engineering at Hanyang University. He received the B.S., and M.S. degrees with honors from Seoul National University (SNU) in 2007 and 2009, respectively, and his Ph.D degree from University of California, Berkeley in 2017. From 2007 to 2019, he has held various full-time and internship positions at TLI, Altera (now Intel FPGA), Intel, Xilinx, and Apple.

Dr. Han was a recipient of a KFAS graduate study fellowship in 2009, a KFAS doctoral study fellowship in 2012, the SNU EECS best tutor award in 2008, and the UC Berkeley EECS outstanding course development and teaching award in 2016.

His research interests include design and automatic generation of high-performance integrated circuits, LED lighting systems, silicon photonics, and bio-electronic systems.

Let’s take a listen to his talk.

한재덕:

Ok, let's get started. Thank you very much for attending this online seminar about future semiconductor technology. My name is Jae-duk Han, and I'm Assistant Professor at Hanyang University. In this talk, I am going to present recent progress on the way of implementing integrated circuits using advanced CMOS semiconductor technology. I hope this presentation is useful for continuing innovations on the semiconductor devices and circuits.

As I mentioned, I am going to talk about integrated circuits, which are also called “semiconductor chips” or simply “chips.” Chips are the main components of electronic systems which perform various computation and storage operations in massive scale. The massive scale means more than billions of computations in one second. And in order to perform this massive computation, there are more than billions of transistors and interconnects integrated to one semiconductor chip. Having more transistors always preferred as more transistor means we have more circuits and more computing power per chip. Therefore, people have been working

very, very hard on putting more and more transistors onto semiconductor chips. And actually, people have been doing this for quite a long time, for almost 50 years, and their efforts were quite successful. The empirical pace of transistor integration technology development is called the Moore's Law, which indicates that the number of transistors on integrated circuits doubles approximately every two years, mostly by making the transistors smaller than before. The Moore's Law has held for about 50 years, and now we can integrate billions of transistors on a single chip. As the cost to inclement a semiconductor chip is almost proportional to the area of the chip, in circuit point of view, the Moore's Law means that you can inclement more functionalities with the same cost, or you can reduce the cost of your design simply by migrating your design to a new technology node. Therefore, circuit designers have been utilizing these free fruits for improving their design for almost about 50 years. And now, we can integrate billions of transistors, which means, technically speaking, semiconductor chips can perform more than one billion functions, which is basically a great thing.

Unfortunately, there was a big problem observed around 10 years ago. As the size of transistors becomes too small, smaller than one hundred nanometer range in terms of their channel lengths, various technical challenges started to come into play, slowing down the pace of device scaling. Simple speaking, I will say it is just very hard to make things very small. Therefore, the cost of implementing our transistors started to rise up after a certain point, and many people, including myself, expected that the Moore's Law is dead because we cannot take the advantage of device scaling any more. Smaller feature size does not mean more functionalities with reduced costs any more. So, no more fun things in semiconductor technology. That's what we, actually including myself, expected about 10 years ago. However, after 10 years past, what we observe now is that the Moore's Law is still going on, although

the pace is already slowing down. So, it turns out actually we could keep reducing the transistor fabrication cost and integrate more transistors down to seven nanometers and now we are expecting that we can do well at least down to three nanometers. A similar trend is being observed in the memory side. DRAM cells are continuously scaled down with a similar pace to the logic transistors. Then, what made the difference between what we expected ten years ago and what we are observing now? Well, I think it is very hard to pick a single factor as the continuation of device scaling in this micron era is driven by holistic evolution of technology across multiple fields of electronic engineering. I will say this is a collaborative innovation as researches in various areas have been organized and performed for these great achievements in the past 10 years.

As I mentioned in the previous slide, the technical difficulty level to continue device scaling has increased significantly and tremendous research efforts in multiple areas, including device, process, circuits, and systems have been combined for the technology development. In the device area, new device structures are investigating to achieve the desired transistor characteristics, and we also worked very hard on the process side to integrate the smaller structures. In the meantime, circuit designers also help been adapting on their they're designed to be compatible with the new device and process technology, new device structures that are therefore developed and evaluated to implement this function with the new devices and integration technology. All people in these three different fields interacted and worked together to continue the transistor scaling. This was not an easy thing to do, but so far people have been successfully doing this job, and I am very confident that the technology scaling will continue at this moment.

However, I have a little bit different doubt in my mind, which is can you actually use the great technology to produce useful things on time? The reason I'm worried is that compared to the traditional transistor scaling happened 20-30 years ago, the recent technology advances have a critical problem, which is increased design complexity. As remarkable ideas are applied to the semiconductor technology, they actually ended up with increasing the complexity of the semiconductor technology. As a result, fabricating semiconductor devices in the recent technology node requires hundreds of processing steps and the transistor characteristics and their associated design rules have become quite complicated. And this complication degrades the benefit of device scaling as the cost reduction from the scaling is ruled out by the increased design cost. For example, the graph in this slide indicates that it takes around two hundred million dollars to develop an SoC in a 16-14-millimeter process, as there are not many products that can justify the two hundred-million-dollar expense or the for their development cost. Less and less designs are developed as the semiconductor technology advances. Even if the production volume is large enough to justify the fabrication in such advanced technology nodes, the technology complication is still a big problem as it leads to increased design time and cost.

Therefore, I think now we should pay attention to ways of making things efficiently, in addition to making good things. The technology scaling has been continued to enhance the quality of the final design, but because we are constrained by time and research budgets we have, introducing complications without productivity enhancement will actually degrade the quality of design. Therefore, developing efficient design methodologies to produce the final design with smaller effort becomes as critical as developing techniques to enhance the design quality. And I think automation is the key factor to enhance the design productivity by

delegating iterative and time-consuming jobs to computers and focusing on critical problems, producing our design in more efficient ways.

For the rest of this talk, I am going to introduce recent findings on circuit design automation. And I think one of the most important concepts of the design automation is the generator-based design methodology, which is illustrated in this slide. The generator-based design methodology is a new design paradigm which focuses on certain generators that automatically produce circuit designs from their target specifications, design procedures, and technology parameters. In the generator-based design flows, circuit designers mainly work on the generators here that create final circuit instances for their design targets, instead of directly working on the final instances by themselves. Then, the actual circuit creation—we call instantiation—will be done by running the generators for target specifications and technology parameters. One of the great benefits of this generator-based approach is that we can reuse generators across different targets and processes like this. In conventional design method, if there are any change on design specification or technology parameters, designers have to iterate the whole design flow by themselves, and it was very hard to reuse their design for one part in another part. But in the generator-based methodology, designers can easily adapt their final circuit instances to new targets simply by running the generator with new parameters, which is very useful if you like to reuse previous designs or different parts, saving tremendous time and effort.

So, there are lots of potential advantages of adopting the generator-based methodology, but all the benefits are valid only when the generators are useful enough, and there are several requirements for the generators to be useful. The first requirement is that the generator should be reproducible and reconfigurable, which

means the generator configuration should be easily steered, so that they can produce final instances for different design parameters. For example, if you like to generate the capacitor array from a template schematic like this, the generator should be able to receive various input parameters such as size, number of bits, radix, and connection information, and produce the finite instance corresponding to the input parameters, as shown in this slide. Another example of parameterized generation is shown in the bottom part of this slide as well, which is about generating serializers with various serialization ratios or placement information. By constructing the serializer generator and running it with different parameters, they can generate our serializer circuits for our targeted use cases automatically, instead of manually crafting our end product for individual cases.

Another important aspect of useful generators is that they can generate physical structures of a circuit which are called layouts. The layout creation is one of the most time-consuming steps in circuit design process, especially in advanced semiconductor technology nodes. This is because the transistor structures in the advanced semiconductor processes are quite complicated due to various sophisticated techniques applied to implement decent devices in nanometers scales. The layout generator should be reconfigurable as well to support various use cases like this, and the generator should be designed to be process-portable, which means almost the same code should be used to generate the similar structures in different technologies. However, due to the complex device structures and their likely design rules, implementing process-portable and parametric layout generators is not an easy thing to do. And that's the motivation I have spent my time and effort on implementing the layout generators for several years. From my previous efforts in collaboration with other researchers at UC Berkeley, I developed a layout generation method based on a template- and grid-based approach. The template- and grid-

based layout generation method is a way which utilizes templates, a placement grid, and a routing grid for the automatic layout construction process. In the template- and grid-based design methodology, we create a list of primitive devices for target technology in advance, and the actual layout generation process will be done simply by placing the primitive templates on process-specific placement grid. One good thing about the template-based layout generation methodology is that we can abstract most of the complicated factors with the use of templates because most of the complicated structures and their design rules are related to the primitive structures. So, if you abstract the primitive transistor with templates, we can hide all the technical details of the internal structure of the templates. And if you like to migrate the layout to another process, the porting job can be done simply by replacing the template with the one for the new technology, and re-run the generator code.

In addition to the use of templates, the proposed method achieved the process portability by introducing the concept of placement and routing grids. So, instead of specifying the value of physical coordinates you're going to place your device on templates, we introduced a virtual grid that abstracts the physical coordinate values, so when you describe the device placement command in your generator, you will use this integer-based abstract coordinate numbers, instead of physical ones, as shown in this slide. In this way, you don’t really need to update your placement command when you're migrating to another technology, as long as their virtual grids are compatible with each other.

In addition to the on-grid placement, we can achieve further process portability by utilizing the relative placement technique, which is illustrated in this slide. So, instead of specifying either physical or abstract placement coordinates, you

can describe the placement process by specifying the relative positional relationship between objects. In this way, the description capability and process portability are greatly enhanced as you don't really need to play with numerical numbers, which are determined by the process technology. After placing the primitive devices, metal wires need to be constructed to provide interconnects between the primitive devices. Similar to the placement process, it is desired to make the wiring process to be process-portable and reconfigurable, which can be achieved by playing the same technique with placement. That is, by introducing routing grid and utilizing relative routing information, we can produce the interconnect structures without dealing with the physical parameters, making the routing process technology-agnostic. These simple template- and grid-based techniques greatly simplify the layout iteration process, abstracting the complex design rules related with primitive devices, enabling the process portable and parameterized layout generator descriptions.

In order to verify the validity and usefulness of the generator-based design methodology, we exercised the proposed process over several different circuit designs. The first example is a high-speed wireline transceiver on which the circuitry is produced from the generator-based methodology. High-speed wireline transceivers are widely used to provide high-bandwidth connectivity between integrated circuits, and the high-speed wireline transceivers are usually operating at the fastest frequency among SoC components, as the chip-to-chip data communication bandwidth is mainly determined by their maximal operating frequency. Therefore, wireline transceivers are one of the most challenging circuits to design, and if you can produce critical components in the wireline transceivers from the generator-based design flow, that will actually demonstrate the usefulness of the generator-based design methodology. In order to produce the final instance for the target bandwidth specification related to the data rate, we implemented

circuit generators that capture various effects related to the primitive components, the circuit topology, and their interconnects. Basically, the wireline transceiver generator is constructed based on an equation-based methodology, but we enhance the accuracy of the method by using precise parameters extracted from the layout. Iterative simulation loops are launched after computing the initial circuit size parameters to enhance the accuracy, and all design steps are performed automatically in sequence by launching a generator code without any intermediate intervention by designers. Due to the automatic executions, we can extremely optimize the design with taking much shorter time and effort than traditional design method. Thanks to the automated design flow, which enables extreme optimization, we could enhance the data rate significantly up to 60 gigabits per second a lane, which was the highest number ever reported when the design was presented at ISSCC in 2017. This is because due to the automatic execution, there is basically no limit on the number of iterations of optimization while we can't iterate more than three or four times in conventional design method, which relied on human resources. So, this demonstrates the evident usefulness of the generator-based design methodology.

The second example, in addition to the wireline transceiver design, is the analog-to-digital converter, which is used for capturing radar signals at a high sampling rate. The analog-to-digital converter, which is also called ADC, is composed of two different types of circuits. The front end is made up of an analog circuit for continuous signal acquisition, while the back-end digital circuits perform various signal processing and calibration functions. Therefore, different approaches are used for generating circuits in these two different domains. In the analog side, we used circuit generation framework called the Berkeley Analog Generator or BAG to produce the schematic and layout from the generator-based methodology. In the

digital side, we used a new hardware description language called Chisel, which supports various useful features for parameterized description of digital hardware. The design automation of digital circuits is traditionally achieved by digital synthesis and automatic place-and-route, which convert HDL code to gate-level netlists and physical layouts. However, the industry standard HDLs lack high level description capabilities, which limits the productivity gain from the digital automation. Therefore, we use our newly developed HDL code Chisel, which adopts various advanced programming concepts, such as object-oriented programming and functional programming to enhance the description capability with rich parameterization. For the generation of digital signal processing circuits, we describe generators for circuits in Chisel HDL and the Chisel code will be translated to traditional HDL code or compiled instances depending on the designer's intent and the type of circuit we are generating. In this way, the design productivity is greatly enhanced as any retargeting or reconfiguring of the design can be achieved by launching the transition flow with a different set of parameters. This slide shows the generation result of the analog-to-digital converter in a 16-nanometer FinFET process. There are two great things we found from the execution of the generator-based methodology. The first thing is the design is implemented in a 16nanometer FinFET process, which is a very advanced technology with very sophisticated design rules. Without the use of the template and grid-based design methodology, which abstracted the design rules significantly, we might not be able to complete the design on time. The second thing is the generated circuit instances achieve the figure of merits that are comparable to the design from traditional flows. This means the generator-based methodology is able to deliver high-performance designs as conventional manual design methodologies provided, consuming much smaller effort and time due to the design automation.

The final example is about standard cell generation in DRAM technology. The memory circuits such as DRAM or Flash memories occupy a significant portion in electronic systems. Traditionally, memory systems are designed by full-custom flows, which take a lot of time and effort, and therefore, we applied the generator-based methodology again to produce the standard cells for memory systems. The standard cells are used to construct various circuits in memory systems. So, automating their flow will be very useful for enhancing the overall productivity for memory system development. We utilized our custom layout generation framework called Laygo to build the standard cell generators, and the generation results were very impressive. We were able to create various standard cells for new technology nodes automatically, without trading the quality of the designs. It is estimated that we could achieve around 500% gain in terms of design productivity, reducing the design time significantly.

So, from the previous three examples, we found out a generator-based design methodology could save us from productivity degradation, and this slide concludes my talk. As I mentioned, there are huge demands on enhancing design productivity, and I think the key factor for boosting productivity is automation. The design automation methodology should be reproducible and reconfigurable. You should be able to produce physical layouts as well. We observed from the application examples of the generator-based methodologies that they can greatly enhance design productivity, and I think there will be research and business opportunities in this area. This is the end of my talk. Thank you very much for your attention.

Discussion

신창환:

Now, it is time for a discussion. Discussion cannot develop future semiconductor technology, but without discussion, the future of semiconductor technology can never become true. As you can see on the screen, I have invited Professor Liu and Salahuddin online and Professor Choi in the studio. Unfortunately, today, Professor Han cannot join this session because of his personal health issue, sorry for any inconvenience that it may cause. About a thousand audience have registered this Webinar and have left many, many questions for the panelists. Based on the questions that the audience have left, I have prepared a few questions for the panelists. So, let's move on to the first question. Number one, as you know, there would be two pathways when it comes to technology advancement. Number one, revolutionary pathway. Number two, evolutionary pathway. So, would you please comment on which semiconductor technology is necessary for each way? Please tell us one specific technology for each way. You may pick up any topic that you have introduced in the presentation or any other topic that you did not talk in the presentation. So, Professor Liu, would you please go ahead first?

Tsu-Jae King Liu:

Thank you, Dr. Shin. Yes. So, my answer is that I think is a combination. So, we are looking for innovations that will provide revolutionary improvement in energy efficiency and performance. However, we also would like those revolutionary advances to be practical for manufacturing. And therefore, ideally, they will only require evolutionary advancements in the actual manufacturing technology. So, we

want revolutionary impact, but evolutionary requirement on the manufacturing process.

신창환:

So, I guess so would you please pick up a specific technique to achieve evolutionary pathway and revolutionary pathway together? I guess the audience would like to hear the specific topic from you.

Tsu-Jae King Liu:

Of course. One example that I touch upon in my presentation is to look at the standard fabrication process and take a look at the interconnects. It turns out that the interconnections between the transistors actually determine the functionality of a chip. So, we can actually make those interconnects reconfigurable in a reliable and nonvolatile manner. We can actually enable revolutionary new chip architectures. So, this is an example of how we make very little modification to the existing manufacturing process, we can achieve revolutionary improvements in performance and in energy efficiency.

신창환:

All right. Thank you very much. So, I guess the front end of the line, you know, breakthrough has been done for the last few decades. However, we have missed any

technical breakthrough happened in back end of the line. So, in that sense, Professor Liu has commented on the back end of the line revolutionary, breathtaking technique would be very well welcomed in the upcoming 10 years. And that would take a leadership in bringing the new semiconductor breakthrough. All right. So now, let's move on to Professor Salahuddin. I'd like to hear your opinion. Please pick up two specific techniques for each way. And many audiences would be very happy to hear your opinion. Please go ahead.

Sayeef Salahuddin:

Sure, can you hear me?

신창환 교수:

Yes, I can clearly hear you.

Sayeef Salahuddin:

All right. So, I think, my thinking about evolutionary and revolutionary is very much in line with what Tsu-Jae has said. I think in semiconductor industry, we see the revolution through evolution in the sense that if you are increasing your performance metrics by 15, 20 percent every year, then in four or five years, you double the performance. And then in another four or five years to double the performance again. So, when we look back seven, eight years, we suddenly realize

that we are actually seeing a revolution, although it is an incremental increase year by year. And that, I think, is because of what Tsu-Jae said, is because everything that has to happen has to happen in a very highly scalable manufacturing way. And so, that is very important. So, from that point of view, something suddenly coming and changing everything is very unlikely to some extent. Now, if I had to pick two specific technologies, I would say that, definitely I think that in the front end, I am very much bullish on the fact that we will see a lot of changes in the gate stack on the gate control on the gate oxide and functionality on the gate stack. If we look back two decades, in the last two decades, we have mostly focused on changing the channel material, hoping that we can come up with a new channel material. But that has not really materialized. You could say that for P-type transistors, we are starting to see the use of high mobility channels. But other than that, that effort has not really resulted in too much. But I think that the improvement in the data stack has been stagnant for almost 15 years. And I expect that that will happen in terms of the next decade. And from that point of view, I think that ferroelectrics are very interesting. Both from a negative capacitance point of view - we think that the off current can be reduced - and also from a memory point of view - we think that the embedded memory technology can come. So, I would call that revolution through evolution. That if we can keep making progress, we will see a revolutionary change in the way we do computing. If you say that I have to pick another technology in terms of a real revolution happening, I think that we are starting to see that already through this heterogeneous integration. And I think this silicon interposer and the package that we are seeing, we will start to see new technologies that come embedded into the package itself. So, you could almost think about technologies in the package, which will essentially add functionality and allow us to do new things

as we connect chiplets and try to improve the overall performance. So, those are the two technologies that that comes to my mind. So, let me let me stop there.

신창환:

Thank you very much for your comment. So, I guess by inserting a ultrathin ferroelectric layer within the Gates Stack, that would bring us small change in gate stack, that would make deep change in transistor performance. So, I hope this ultrathin ferroelectric would become a key player in the gate stack engineering in the very near future. All right. What about the spintronics that you're working on? So, any comment on that?

Sayeef Salahuddin:

Yes, so I think that the spintronics, our work has mostly been on magnetic memory, especially embeddable memory. And I think that embedded memory technology is to some extent already commercial. All foundries now have a technology and expect to see chips coming out with embedded magnetic memory soon. And so, to some extent, I think once technology finds its place, it is going to be there. And so, we do expect to see embedded magnetic memory in our chips unless something else like ferroelectric memory shows better performance. But for the next 8-10 years, this is already a commercial technology has gone through the development of high scale manufacturing. So, that technology is going to be there. Our work at Berkeley has been to go beyond the standard spin-transfer torque and trying to see if we can reduce the light energy by using a spin orbit torque. But the

tradeoff is that the spin orbit torque devices have a larger footprint. So, in terms of the applications, it will appeal to different sets of application where light energy or light power is more important than having a very high-density memory added. Thank you.

신창환:

All right, thank you very much for your comment and opinion. Now, let's have a listen to Professor Choi's opinion on the first question. So, would you please comment on the question?

최창환:

Ok, I'm not that revolutionary person, so I'm going to start this evolutionary path. It seems like I'm the only person from the material science engineering, so I would like to mention about some perspective of materials. So, we have focus on the transistor scaling to boost the performance. But without the improvement in the contact resistance or the interconnection, we cannot enjoy that benefit. So, somebody say, even though we have a very nice Porsche or sports car that is stuck in the traffic. Then what we could be enjoying using that that kind of Porsche. So, we have to focus on some contact resistance. For example, think about pitch scaling. So so far, we are using the tungsten. So, when you do the tungsten, you need a certain kind of barrier layer or cladding layer like titanium nitride. And also, you need to some kind of nucleation layer. But when you scale these the pitches, that could be possible do using EUVs. But the room for that kind of cladding layer like

the titanium nitride or tungsten nuclear layer is very tight. And those materials are also not that good to reduce the resistance overall. So, we have to switch to a new process or the new material. So, some companies have their own technologies not using this kind of cladding material like a titanium nitride or the nucleation layer. But eventually, we have to look for some alternative material like molybdenum or ruthenium or cobalt or something like that. Also, this is also linked to development of new processing technology. How do you feel such a small trench? We have to focus on that kind of some interconnection area or the contact areas besides transistor development. Also, I mentioned about 3D integration or the heterogeneity integration, eventually monolithic integration. So, you want enjoy some functionality. So, how do we achieve that kind of functionality? So, it is necessary to fabricate the leading-edge technology to fabricate the CPU or the GPU or HBM and etc. So, some chips have to be fabricated with some legacy technology, not leading technology. So, those kinds of chips have to integrate these together. So, the way we have to integrate is only using those heterogeneous integration or monolithic integration because that kind of system needs more I/O something like this. So that might be some evolutionary way and semiconductor community is a quite conservative. So, when I was a graduate school to 20 years ago, I studied the high-K metal gate. So high-K metal gate was introduced as 45 nanometer technology. At that time, all kind of medias mentioned that the introduction of a high-k metal gate was a revolutionary thing over the last forty decade or something like that. But these days, high-k metal gate is kind of something normal. Because the industry investment cost a lot of they don't want to rely on that kind of equipment that uses revolutionary processes because they can hit some existing equipment in their facility. But if I have to mention about revolutionary path, maybe we have to change some architecture, from the von Neumann architecture to some quantum computing or neuromorphic

computing. But neuromorphic computing got very high attention, so probably within one decade, there might be something because necessity is the mother of the invention. So, if we need something, something will come up. But for quantum computing as far as I know, making the qubit is quite difficult. But it comes following after the neuromorphic computing. So, in summary, the evolutionary path is the contact, the interconnection path, and 3D integration potentially. And then the revolutionary path is new structure like quantum computing and neuromorphic computing. That is my opinion.

최창환:

Thank you very much for your comment. From three panelists that we have collected many specific key ideas as a key enabler for a future of semiconductor technology. So, I guess at this time, since we have invited device professionals and material professionals here, I would like to ask this a follow up question. Any co-optimization would be possible to extend out the lifetime of current CMOS technology platform in the near future? Any comment from panelists? Maybe Professor Liu first? Lady First, please. We cannot hear your voice. Would you please turn on the microphone again?

Tsu-Jae King Liu:

Sure. Thank you. Sorry about that. Yes, I think the semiconductor industry already has been using design technology co optimization for approximately 10 years already or so. So, when we advanced transistor technology, we also have to

optimize the circuit design because transistors, as they scale down, become less ideal as digital switches. So, I think in the future, as we introduce the same negative capacitance transistors, there will indeed need to be some design technology co-optimization to ensure that the full benefit of those innovations can be realized. So absolutely, I think design technology co-optimization is here to stay.

신창환:

Ok, thank you very much. So, DTCO (design technology co-optimization) is very, very necessary and required to extend out the lifetime of CMOS technology. How about Sayeef Salahuddin? Do you agree or disagree that design technology co-optimization or any other type of co-optimization is necessary? Please comment on this.

Sayeef Salahuddin:

I absolutely agree with what Tsu-Jae said. This is already being done and this is yet to stay. To some extent, I could add that when we do exploratory research, if some of that could be also informed by something like DTOC that is done in the industry, that may help focus on certain areas. Right? Because often we chase a matrix on a device or a material which may not have the right impact when it goes into a computing system. So, I'll just give you an example that I was talking about. Right? On magnetic memories. And usually when we do research, the usual idea is that let's try to reduce light energy as much as possible. But what happens is beyond a starting point, beyond a certain number, how much light energy I need in

the memory-bit itself no longer matters because it can be totally swamped by the amount of energy we need to add to charge up the bit-lines and other things. So, it will depend on your block size and other things. So those are the things often, I think because of the history of plug and replacement in a well-known architecture, device researchers are not trained to think in that way. But some of that coming into the research as well, I think would be more impactful.

신창환:

All right, thank you very much for your comment. What about the Professor Choi's comment on this follow up question?

최창환:

So, these days the companies are focusing on PPAC, performance power area costs. So even those investments cost a lot, it's going to be useless if we cannot enjoy some of the benefits. So, the power, performance and area costs. So, besides the development of some of the new technologies, the industry focuses on some DTCO. You guys mentioned about the design technology co-optimization. So, material, process, device and circuit and some system-level that they have to collaborate together to pave the way to the future of the semiconductor. So, I'm not sure that the structure of the current CMOS technology will be changed so dramatically, but I don't believe that they're going to remain there in some time. So maybe next 10 years, next 20 years. They're going to optimize how do we enjoy the benefits by optimizing the PPAC. So, in terms of material processing or the device or

circuit, they're going to develop their own way to meet certain requirements of PPAC. So that is my opinion.

신창환:

Ok, thank you very much for your comment. Let's move on to the next question. Question number two, I guess so many audiences has left this question. In the upcoming ten years from now, what would be the new emerging technology or alternative to current CMOS technology that can restructure the competition in the field of semiconductor business? So, I guess a student has raised up this question. A start startup company brings up breathtaking technology, for example, it becomes a game changer in the upcoming 10 years. So, start-up would replace the CMOS technology or not. Is it possible for any new idea coming from the start-up company would reconstruct the semiconductor business sector? So, please comment on this question. I guess this time, Professor Sayeef Salahuddin, please go ahead first.

Sayeef Salahuddin:

Yeah. You know, it is always easy to say no to answer such a question because from our vantage point, when we look at something, it is very difficult to see that anything can come and suddenly completely usurp the CMOS technology. That looks very unlikely. But in this case, I would say it is more likely so. In fact, one of our faculty at Berkeley EECS, Professor Edward Lee, he has recently written a book where he basically shows by using information theoretical approach, that the way our computing system was developed, it almost follows many similarities with actual

evolution. And based on that, he argues that evolution is very, very robust. It's very difficult to make tremendous changes in an evolutionary object other than if a meteor comes and hits the planet and so the dinosaurs are gone. But other than that kind of very rare event, the objects that come out of evolution are very robust. And based on that, he argues that our computing systems are not going to see a revolutionary change in the sense that one fine morning, you just come up with a new idea that completely changes everything. That's because many of the supply chains and many of the infrastructure has been developed following what we have today. But on the on the same topic, I often discuss with my students that there can be a Tesla moment in semiconductor industry. I see a lot of similarity with the semiconductor industry with car manufacturing. Both are extremely highly capital. It requires a lot of capital to start and it takes a company a long time to stand on its own feet. But if you look at the example of Tesla, they basically had a very good system in mind and they had a very long-term business in mind. And then they came. They basically set up a manufacturing plant under a tent in Fremont and they had a difficult time to survive. But this is where the government policies, especially California state policies, help them to survive the initial days. But now, you see what they have done. And I can see that in semiconductor. If there is somebody extremely smart and visionary person. That person can make that happen. They can come and completely change the way we do manufacturing of computing chips today. And based on a very compelling system-level gadget or system-level solution and completely change the nature of the business. That I think is possible. But can computing go away from CMOS very fast, very soon? That seems to be very unlikely. So, that's a very long answer to your question.

신창환:

Thank you very much for your comment. I guess Tesla moment is necessary to become a future leading scholar and a technologist in the sector of semiconductor business. Maybe the current undergraduate, graduate student should have a Tesla moment in his or her mind. I guess in that sense, Professor Liu, would you please comment on or share your own experience of that kind of Tesla moment in Silicon Valley or in Berkeley? Any comment or any good specific example for that?

Tsu-Jae King Liu:

That's a good question. I agree with Professor Salahuddin's general thinking that, it is difficult to sort of displace CMOS largely because of all the inertia that that that technology has. It's used for so many types of products and there's a lot of design infrastructure. So how do we design how do people design these 10 billion transistor products within a couple of years? Well, there's a lot of automation in the design process. Right? Years of intellectual property built up. So, there's just too much inertia, I think, to completely displace CMOS in the long term. However, if somebody comes up with a technology that can achieve something that CMOS will never be able to achieve, then I think there is that opportunity. So, we just have to keep in mind that CMOS, as Professor Salahaddin said, if it improves 10-15% every year or two, it's going to increase by orders of magnitude within 5-10 years. And so whatever new revolutionary idea you have has to provide for orders of magnitude improvement or actually infinite improvement over CMOS. So whatever CMOS can never do, those are the kinds of innovations I think, that can possibly make a big difference, revolutionary impact in the industry.

신창환:

All right, thank you very much for your comment. I guess all the students, all the audiences may have a chance to hear Moore's Law, the total number of devices in our integrated circuit chip has been doubled every two years. That has, changed everything, the way how humans are living. So, in that sense of Professor Liu has commented on it. So, any new idea should be much, much better than the CMOS technology development advancement over the last few decades. So now, let's provide a chance to Professor Choi. Please comment on the question.

최창환:

So, the CMOS technology is still alive over the next couple of decades. Because these days when the companies are fabricating one FET, how much does it cost? Around 10 billion or 15 billion or something like this. So those kinds of stuff is the basis to fabricate the semiconductor chip. So, they are not going to change the prototype of the CMOS device. But I guess the applications of the CMOS technologies will be expanded. So, for example, Professor Sayeef mentioned about Tesla. Tesla started to fabricate their chips relying on foundry companies, HW 4.0. And then Apple, they are trying to fabricate their chip, M1, for their electrical vehicles. So those kinds of things will proliferate the industry of the CMOS technology. So, I'm not sure that my answer is the proper reply to the question. But what I want to say is, CMOS technologies will just evolve, not disappear.

신창환:

Ok, thank you very much. So, I'm very glad because I'm a CMOS guy. CMOS will never be disappeared in the future. So that sounds excellent. OK, now let's move on to the next question. Number three, the ultimate goal that Chey Institute is looking for is the nexus between geopolitical risk and businesses or scientific innovations. So, I have collected some questions from audience and a few questions are as follows. In the future, which players, which companies are the main competitors of Korean semiconductor companies? And what are the main weak points that the Korean semiconductor industry should overcome in order to raise the awareness of Korean semiconductor technology in the world? So, in order to collect opinions for those questions from the leading scholars, I'm going to use this non-real scenario as follows. Suppose that the Biden administration proposes a trusted semiconductor consortium, TSC, to ensure that critical semiconductor supplies can be sourced wholly inside USA. All the allies have rapidly agreed to the proposal, but Korea did not agree rapidly. This is because of the close trade relationship that Korea enjoys with China. So, Beijing makes clear that trade ties would be put at risk if Korea joins the TSC. In parallel, China offers trade concessions if Korea stays out of the consortium. As a technologist, how to build up the healthy supply chain or ecosystem for a semiconductor business. So please comment on this. I guess Professor Liu have many, many ideas on this scenario. So, would you please comment on this scenario. Please turn on your microphone and please give us your wisdom on this scenario.

Tsu-Jae King Liu:

That's a very interesting scenario, Professor Shin. You know, it's a tough question to answer. I think in general, co-opetition is good to together advance technology for the benefit of all companies, all countries. So, I think there's there can be some interesting ways to structure perhaps joint ventures or a trusted foundry, Perhaps, Korean companies can provide some of the front-end process and then some critical steps can be manufactured in the United States. So, there should be some creative cooperation or coopetition business models that should be possible to satisfy the both sides, China and US requirements for semiconductor manufacturing companies. But I'm interested to hear what the other panelists think.

신창환:

Professor Salahuddin, could you please comment on this scenario.

Sayeef Salahuddin:

I think you have created a quite difficult scenario. Again, I have to agree with what Tsu-Jae said. I think politically and economically what makes sense for a country is beyond the technology itself. So, it's difficult to comment there. But if a situation like that comes, then I'm sure that Korea can figure out a way of participating in a trusted foundry with the US while not violating that trust, having trade partnership with China, I mean, it is definitely possible. In the US, for example, there are foundries, which are both commercial foundries and defense contractors. So, they are able to separate out their business in that way and keep the confidentiality and keep the separation clean enough that the US government is

doing trusted foundry business with them today. So, I'm quite confident that an arrangement like that is possible. The second thing is, you know, competition is always going to be there. And so, from that point of view, I think open innovation is also important. Korea should participate in open innovation and then basically compete on the on the product. Right? So those are the two ideas that I can think of.

신창환:

Right, thank you very much. What about Professor Choi's opinion?

최창환:

We should be friends and not compete each other. So, OK, people worry about the location of chip fabrication over the world. The East Asians, the majority to fabricate the chip in this day. So, I'm not the diplomatic person or the political person. The political persons they have the concern. So, we're not going to rely on some specific location because we realize after the Covid-19 situation that the semiconductor shortages become more and more. So, how do we solve this kind of production location? So, that is why the USA recently, President Biden recently wants to build up some resilience of the semiconductor industry in USA. Intel announces they're going to come back to Foundry. And TSMC and Samsung, they are under pressure to build up some factory in Arizona or Texas. So, the USA is going up to address the importance of the semiconductor chip and then said China wants to also get their own factory. Right? So, there is dispute politically. I think Korea should

not get into that kind of discussion about we should all be the moderators between these two countries. So, Korea has a very strong fabrication. And then we also have to link between these two countries and these two societies. And then the weakest point of the Korean semiconductor industry, we are just strong in fabrication of technology but not strong in circuit or software. The ecosystem is not complete in Korea. So, I think it's a Korean company or the Korean government to have to investigate that kind of some weakest area. So, at any rate, we should be friend. We have to work together.

신창환:

Thank you very much. So, I guess at this point we have missed ASML, which is a key provider of extreme ultraviolet EUV photolithgraphy equipment. So, without it, we cannot implement sub-ten nanometer CMOS technology platform. In that sense, recently new CEO Gelsinger at Intel Corporation has officially announced that they're going to run the foundry business. So maybe their target technology code or node would be 10 nanometers or so. So, in that sense, Intel would need to purchase many, many EUV equipment. So, in Netherland, from EU. Right? So, I guess Professor Liu can answer or give us some, you know, strategy, how Intel is going to be a key player in the section of foundry service. Would you please comment on this?

Tsu-Jae King Liu:

Sure. Well, you know, I think Intel does aim to become a world class foundry. This is part of the IDM 2.0 business model. And as people should know now, the reason why we can afford to advance the technology, because it's very expensive to buy the EUV tools and all the other tools in order to just advance the technology by one manufactured generation. And so, for a company to be able to afford that, they need to be able to make enough profit to have enough revenue so that they can invest further to advance the technology. And so, Intel by becoming a foundry will have more volume of chips manufactured in its fabs. So, we can actually, you know, exercise the fabs and ramp up the technologies more quickly, but also then be able to generate greater revenue to be able to advance the technology. I think one company really cannot fill up the fab adequately to the advancement of new generations of technology. So, this is clearly been one of the advantages of a pure play foundry like TSMC. They have all these customers. They run a lot of wafers with that experience. They can ramp up the technologies faster and recoup the costs and invest in the next generation. Companies like Intel's IDM also need to be able to somehow increase the volume of product that goes through the foundries to be able to keep up with the pace of innovation. So, I think that's the biggest key, is to fill the fabs as much as possible to advance the technology as fast as possible.

신창환:

Ok, thank you very much for your comment. Yes, let's move on to the very last question. So, because all the panelists are working in universities, so to foster the semiconductor business in the upcoming 10 years, which materials which type of classes should we provide and should we teach for undergrad and graduate

students? So, should we teach basics more or should we teach applications more or both? I would like to hear a short comment from Professor Sayeef Salahuddin first.

Sayeef Salahuddin:

Ok, thank you. I feel that, we were talking about DTCO, the co-optimization and all. I think that students of tomorrow need to be taught so that they have a vertical orientation. They can see the entire stack before they actually decide on which direction they want to go in depth. So, I think this is very important, mostly because we know that we are no longer doing drop in replacement of devices or technology into a well-established computing architecture. Right? If we look at it, the people are coming up with new architecture. People are trying to come up with new functional devices. If you just look at the space between DRAM and Flash, there are so many different devices and technologies that people are trying to bring in. But to understand how impactful those will be, a student needs to have that understanding of the entire stack and then they can decide on which direction they will go.

신창환:

Professor Chair, would you please comment on I guess you can give us some class name or others, what kind of classes should we teach for future generation?

최창환:

I think through the panel discussions, we already mentioned about some hints which subject and what stretches what pathway we have to teach students. From the material science and engineering, most students take material-related course. But that is not enough to develop a semiconductor, as you guys agreed. So, we have to bring up some alternative, some comprehensive course. So, I'm teaching some electronic material and electronic device physics. But if possible, we have to push students to take courses in circuit or the other system. These days, the trends to do something for the research, we rely on the holistic approach. So just like that, the teaching also holistic approaches we have to rely on.

신창환:

So, I guess in order to provide a comprehensive understanding in the sector of a semiconductor business from materials, device, up to a circuit system. So, I'd like to hear UC Berkeley's education system from Professor Liu, would you please comment on how you are teaching for Berkeley students.

Tsu-Jae King Liu:

I see well, you know, in the electrical engineering computer sciences department at UC Berkeley, the faculty have thought about revamping the hardware courses really so that students understand early on applications of electronics technology. So, once they understand that it's actually the electrical engineering, electronic engineering that enables these electronic brains to be so capable that the devices that they use, their cell phones, laptop computers, cloud computing really

rely on advances in hardware that will keep their interest. It's consistent with what Professor Choi said. A holistic view. Students need to understand what information systems and devices comprise, including these chips, and then that will peak their curiosity to learn about how they function and how transistors function and how materials and processes make all the difference and how these devices function. So, I think, yes, starting with the big picture and then diving down into details, I think that's a more modern way to teach engineering to attract the best and brightest students to work in this field. Now, the old way of first learning math and then physics and then some circuits, that it takes for three or four years for students to learn what all that knowledge is useful for. And today, the students are not so patient. They need to know exactly what difference it will make for them to work in this field. And we need to tell them that earlier rather than later.

신창환:

Thank you very much for your comments. So, I guess because we are running behind schedule slightly, so now it is time to wrap up the discussion session. So, Gordon Moore, who is the founder of the Intel Corporation, said that the number of devices in an integrated circuit chip has been doubled every two years or every other year. This advancement improvement should dramatically change the way how humans are living. For example, personal computer PC in 1990s, desktops, Internets in 2000s and smartphones in 2010s. What would be the key words for 2020s? Maybe ABCD? A as AI. B as big data. C as cloud. And D as a data analysis or even digital transformation. So, from the lessons in the webinar today, would all the audience find your own insight on semiconductor business sector in upcoming 2020s, even and beyond. I hope you all to find out your own insight for the future

of semiconductor technology. I appreciate your attendance and hope to see you again in the next webinar. Thank you very much for your attendance. Bye bye.