System LSI: Challenges and Opportunities

Tadahiro Kuroda

Department of Electronics and Electrical Engineering Keio University, Yokohama, Japan

Abstract

End of CMOS scaling has been discussed in many places since the late 90's. Even if the end of CMOS scaling is irrelevant, it is for sure that we are facing a turning point in semiconductor business. In this paper, challenges and opportunities of system LSI are discussed from three levels of perspectives.

Transistor Level (Physics)

What will happen if we continue to scale a CMOS device? Gate controllability will be degraded. If gate oxide thickness is kept constant in order to suppress gate leakage, and if channel length is scaled further, potential of silicon surface near the source that is associated with $V_{\rm TH}$ is lowered by depletion layer capacitance of the drain ($C_{\rm D}$), as well as gate capacitance ($C_{\rm G}$). When $C_{\rm D}$ is not small enough compared to $C_{\rm G}$, MOSFET behaves like a resistor, not a transistor. In other words, MOSFET does not turn off sufficiently. On top of that, leakage from gate is rapidly increasing. It should be noted that $I_{\rm off}$ dependence on temperature is diminishing as technology scales. This indicates that the leakage is caused by tunneling leakage that depends on barrier thickness, not height. In the 45nm technology node, $I_{\rm off}$ will reach to an unacceptable level.

How to cope with the leakage problem? There are two approaches. One is to change material. High-*k* gate dielectric together with metal gate is a must. Another important change in material is strained silicon. Mobility is increased in NMOS by pulling silicon body, while in PMOS by pushing it. I_{on} can be increased by 10 to 25 percent for the same leakage current, or I_{off} can be reduced by 85 to 100 percent for the same I_{on} , or active power dissipation can be reduced by 15%. An advantage of the strained silicon technology is that its associated penalty is small. Two approaches are investigated, substrate-strained silicon and process-induced strained silicon.

The second approach to improve I_{on}/I_{off} is to change structure of a device. Degradation in I_{on}/I_{off} is caused by poor gate controllability. Leakage current flows deep in body where gate control does not reach sufficiently. Therefore, by thinning body, the leakage current can be suppressed. If body is as thin as about 1/3 of channel length, leakage current is sufficiently suppressed. Ultra thin body approach, however, has difficulty in scaling. Double gate can ease body thickness to 2/3 of channel length. Non-planer structure, such as FinFET, is widely investigated.

In this way, CMOS scaling will continue, but not in the same fashion as before. We need to introduce various new technologies, namely technology boosters. Even though each technology booster does not continue to improve device performance by scaling, various technology boosters are currently investigated and they will be introduced continuously. Device technology will further be evolve and revolve. 3D integration, carbon nanotube-FET, single-electron devices, molecular devices, DNA computing, and spintronics will emerge.

IC Level (Electronics)

Leakage problem of a transistor shall be solved by intelligence and enthusiasm of device engineers. Difficult problems are shifting from transistor level (physics) to integrated circuit level (electronics). Two issues will stand in our way: powerwall and variations.

A) Power Aware Computing

CMOS power dissipation is increasing as a natural result of device scaling. It was increased four fold every three years until the mid 90's, as a result of the constant voltage scaling. Recently, it is increased double every six years under the constant field scaling. Since it is very difficult to scale V_{TH} further, active power dissipation may increase rapidly again due to the constant voltage scaling, or active leakage current may increase due to further V_{TH} scaling. In any way, tradeoff between speed and power may be more and more difficult. Since it is based on a device scaling theory, which is the physical principle of semiconductor devices, it is very difficult to cope with this problem. Opportunities may be found in IC design level.

Best practices for power aware CMOS design are summarized as the following ten tips [1]:

- Tip 1: Optimize and control V_{DD} and V_{TH} .
- Tip 2: Total power is minimum when $P_{\text{leakage}}/P_{\text{active}} = 30/70$.
- Tip 3: If you don't need to hustle, relax and save power.
- Tip 4: Utilize surplus timing with multiple V_{DD} 's and V_{TH} 's.
- Tip 5: Total power is minimum when $V_{\text{DDL}}/V_{\text{DDH}} = 0.7$.
- Tip 6: Two types are sufficient.
- Tip 7: Adapt to the change with variable V_{DD} and V_{TH} .
- Tip 8: Two levels are sufficient.
- Tip 9: Cooperate across various levels of design hierarchy.
- Tip 10: Right circuit for the right job.

For instance, a tiny microprocessor is embedded in a next generation Intel microprocessor, Montecito (Itanium), that controls power supply voltage and clock frequency by monitoring power dissipation, temperature and supply voltage. Montecito outperforms the previous design named Madison with less power dissipation. In this way, dynamic voltage control and frequency control will be carried out more sophisticatedly and in finer granularity.

B) Power Aware Communications

As VLSI technology scales, the gap between chip performance and I/O bandwidth is widening. In order to fill the gap between the chip performance and pin bandwidth, various circuit techniques have been employed, such as pre-emphasis in a transmitter and equalization in a receiver, but at the cost of increase in power dissipation. Communications between chips may scale as well when distance between the chips scales, from board ranges to chip ranges, and beyond.

Obviously, System-on-a-Chip (SoC) is an ultimate solution. However, SoC is a very expensive technology. Good compromise between power, speed, and cost, turn-around-time, may be found in a System-in-a-Package approach, or SiP.

Through-Si-Via technologies with high density vertical interconnections through a chip have been investigated by many researchers for many years. They are not practical applications yet, though. The reasons are from the following issues and drawbacks of the approach. 1) yield degradation due to difficulty in screening a known good die. 2) cost increase caused by addition in process complexity, as well as reliability issues caused by mechanical contacts, and 3) scaling beyond 30µm may be very difficult in the mechanical contact.

Research on wireless inter-chip data communications has attracted attentions. All the above mentioned issues can be solved. Furthermore, the thinner the chip thickness, the shorter the communications distance, and the better channel performance. Thinning a chip has not been challenged before, just because there was no such a demand. A great opportunity may be found. A scaling scenario may be found where cost/performance ratio can be improved by scaling a chip not only horizontally but also vertically.

Two technologies have been investigated: capacitive coupling [2-3] and inductive coupling [4-9]. In the inductive coupling, chips are stacked, whether they are faced up or down, and inductively coupled by metal spiral inductors. Current change in a transmitter is transferred to a receiver as a voltage signal. Some of recent research achievements will be discussed in the presentation.

Business Level (Economics)

Mr. Gorden Moore from Intel mentioned in a plenary talk of ISSCC2003 that no exponential is forever, but 'forever' can be delayed! Prof. Chemming Hu of U.C. Berkeley presented a grand view of CMOS technology and industry at the plenary talk at CICC2004, entitled "CMOS for one more century". The message was CMOS must and will meet world's needs for intelligent devices through 21st century. Dr. Kazuo Yano from Hitachi mentioned in a workshop of ISSCC2002 that wide ranges of the semiconductor technology from high-end microprocessors to tiny RFID chips will be used for a wide variety of new applications. Prof. Takayasu Sakurai of Univ. of Tokyo mentioned in his plenary talk of ISSCC2003 that poweraware electronics opens up vast new applications and markets. Dr. Tsuguo Makimoto, ex-CTO of Sony, gave a message in his luncheon talk at CICC2003 that robotic chip will be a technology driver in the coming decades.

I think that computers will be invisible behind broadband networks as servers, while terminals will come closer to us as wearable and implantable devices, more friendly devices with intelligent Human Computer Interactions. I think three key technologies are essential: short range wireless interface and intelligent Human Computer Interactions for access networks, and broadband high-speed links for backbone networks. IC chips will be implanted everywhere so that things can think and talk for distributed information processing. For example, it will be possible to automatically detect who enters a room so that air conditions can be locally adjusted according to his/her preference.

Summary

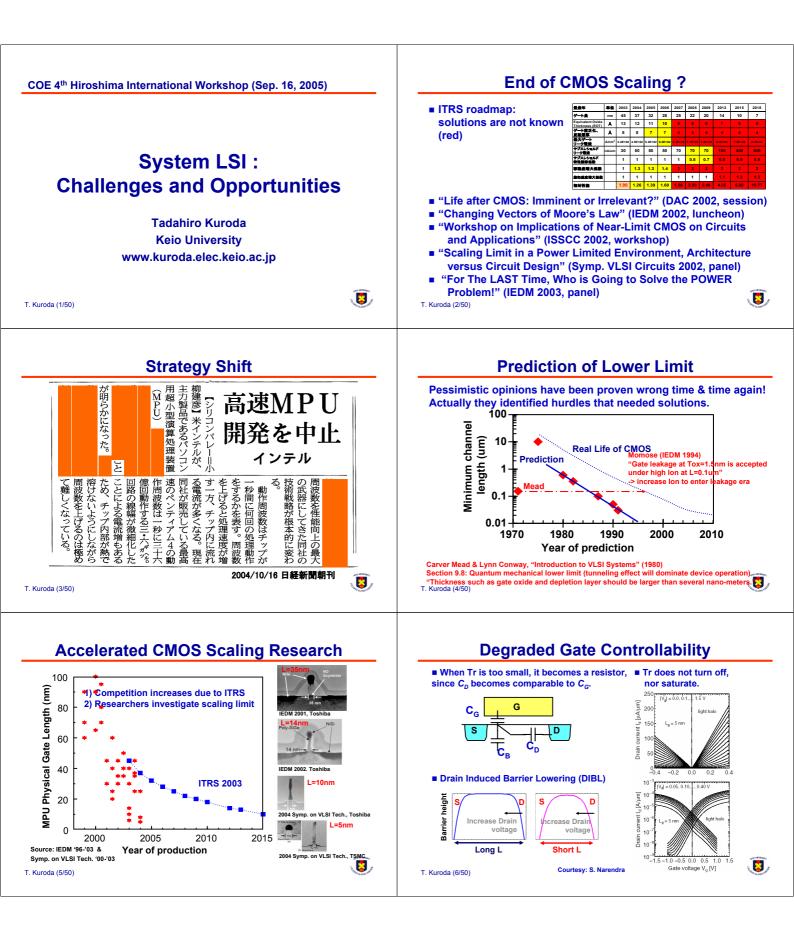
Scaling of CMOS integrated circuits is becoming difficult, due to increase in power dissipation and device variations. Post-CMOS device for mass production, however, is not on the horizon. Even though speed of technology improvement may be slowed down compared to that before, the semiconductor industry and technology will continue to develop and improve remarkably. We will enter the golden era of opportunities for both technology and industry. Computers and communications will be scaled further, merged together, and materialized in consumer applications. In order to make this ubiquitous society possible, intelligent interface based on low-power, high-speed CMOS design is inevitable. Especially, three key technologies are essential: short range wireless interface for access networks, broadband high-speed serial links for backbone networks, and intelligent Human Computer Interactions. In addition, integrated design engineering through system, circuit, device, process, lithography, and packaging, will be essential, as well as extended learning and knowledge such as nanotechnology, biology, and robotics. The era will transit from "Know-How (craftsman's age)" to "Know-What (visionary's age)", and further to "Know-Who (producer's age)".

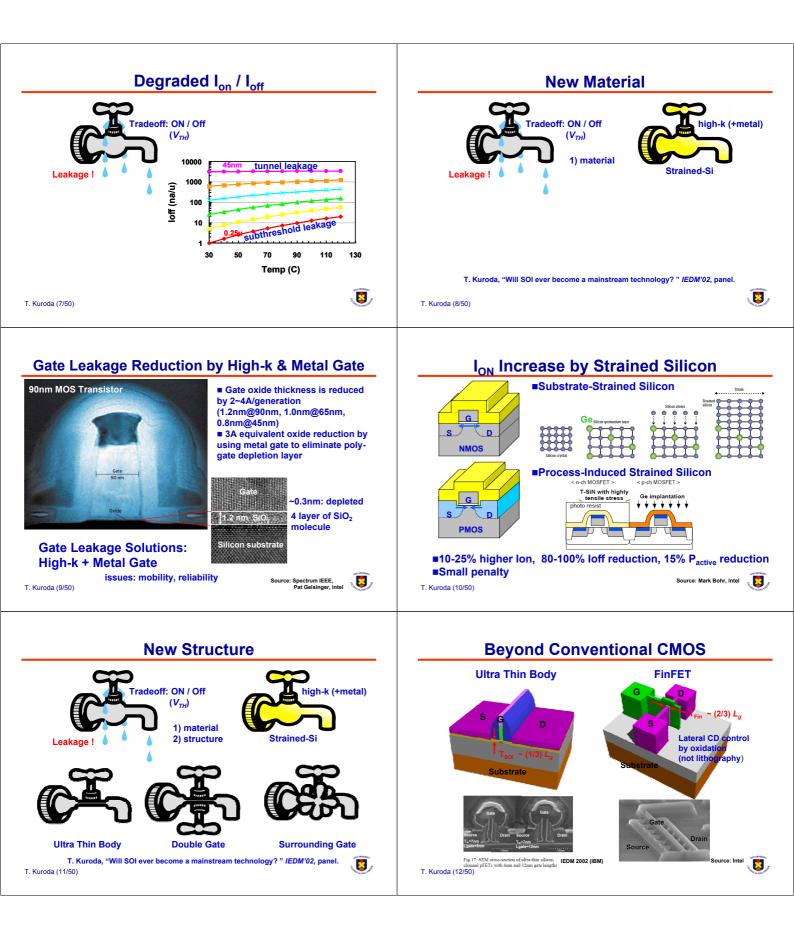
Acknowledgements

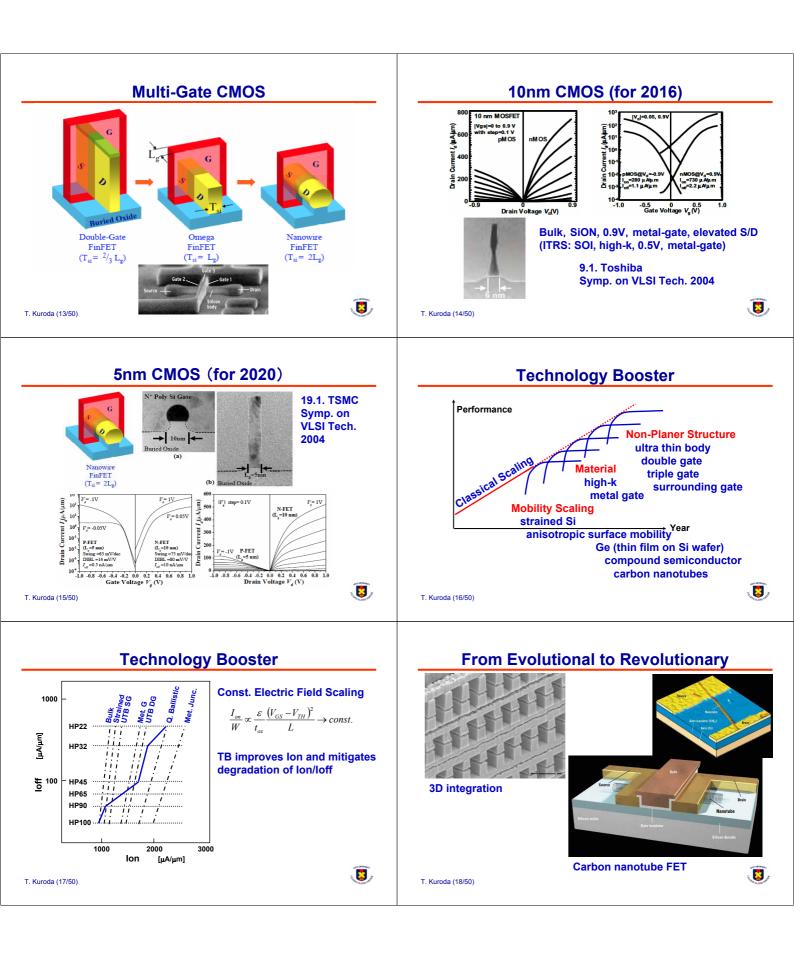
The author would like to acknowledge Prof. Iwata, Prof. Yokoyama, and Prof. Kikkawa for encouragement.

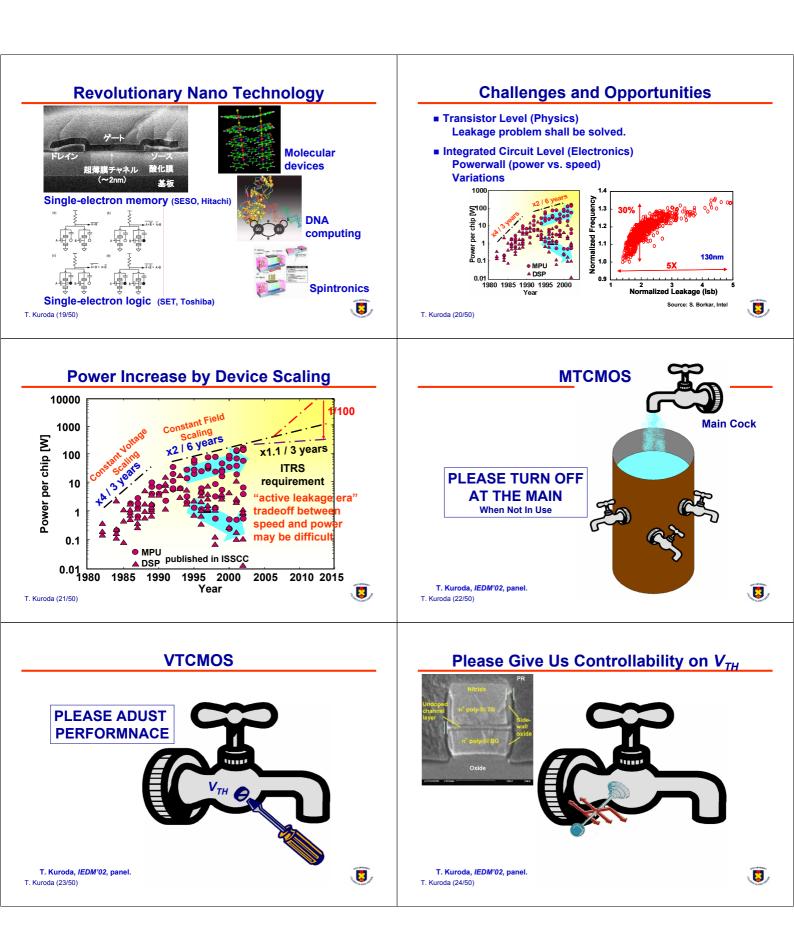
References

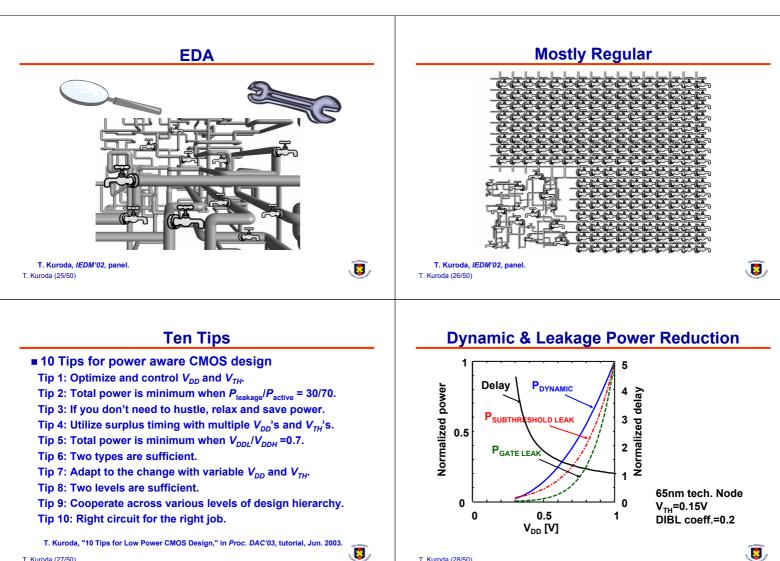
- [1] T. Kuroda, "10 Tips for Low Power CMOS Design," DAC, June 2003.
- [2] K. Kanda, et al., "A 1.27Gb/s/ch 3mW/pin Wireless Superconnect (WSC) Interface Scheme," *ISSCC*, pp.186-187, Feb. 2003.
 [3] R. J. Drost, et al., "Proximity Communication," *CICC*, pp.469-472, Sep. 2003.
- [4] D. Mizoguchi, et al., "A 1.2Gb/s/pin Wireless Superconnect Based on Inductive Inter-chip Signaling (IIS)," ISSCC, pp.142-143, Feb. 2004.
- [5] N. Miura, et al., "Analysis and Design of Transceiver Circuit and Inductor Layout for Inductive Inter-chip Wireless Superconnect," Symposium on VLSI Circuits, pp. 246-249 June 2004.
- [6] N. Miura, et al., "Cross Talk Countermeasures in Inductive Inter-chip Wireless Superconnect," CICC, pp.99-102, Oct. 2004.
- [7] A. Iwata, et al., "A 3D Integration Scheme utilizing Wireless Interconnections for Implementing Hyper Brains," ISSCC, pp.262-263, Feb. 2005.
- [8] N. Miura, et al., "A 195Gb/s 1.2W 3D-Stacked Inductive Inter-Chip Wireless Superconnect with Transmit Power Control Scheme," ISSCC, pp.264-265, Feb. 2005.
- [9] M. Sasaki, et al., "A 0.95mW/1.0Gbps Spiral-Inductor Based Wireless Chip-Interconnect with Asynchronous Communication Scheme," Symposium on VLSI Circuits, pp. 348-349, June 2005.



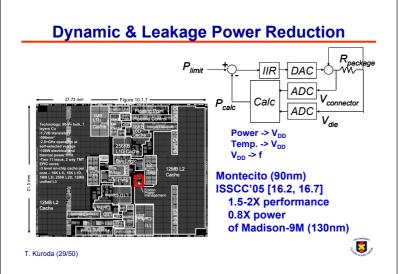




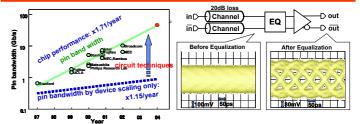




T. Kuroda (27/50)



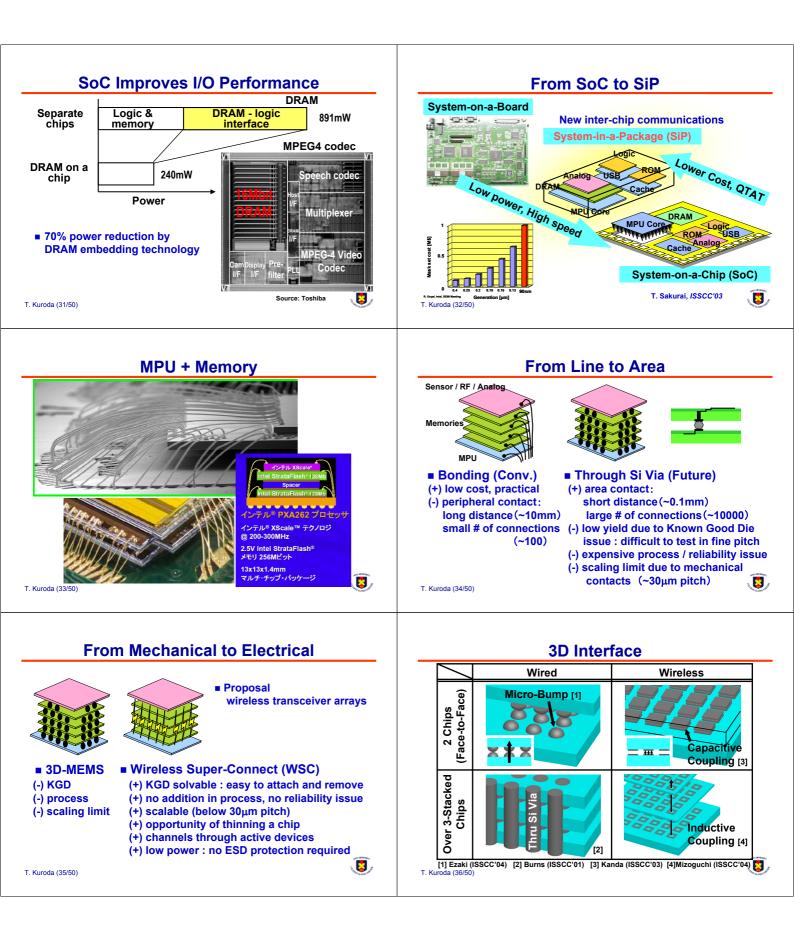
Challenges in Inter-Chip Interface

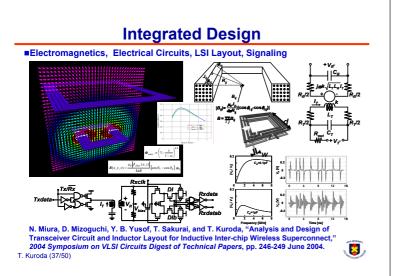


I/O is a primary bottleneck of system performance. Communications does not scale as computation scales by device scaling. Circuit techniques (e.g. equalization) consumes large power dissipation. Communications scale by distance scaling from board, to chip, and beyond. Moor's law in inter-chip communications is expected.

T. Kuroda (30/50)

T. Kuroda (28/50)

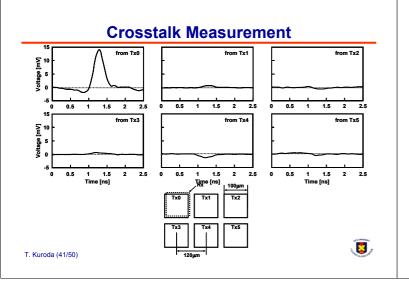


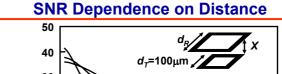


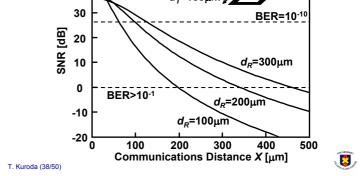
Scaling Scenario

Process	Chip Thickness	Inductor Diameter	Transmit Power	Data Rate	Power
0.35µm	240µm	300µm	43mW	1.25Gb/s/ch	46mW
90nm	30µm	300µm	1mW	1.25Gb/s/ch	1.1mW
90nm	30µm	50µm	21mW	1Tb/s/mm ²	4.2W

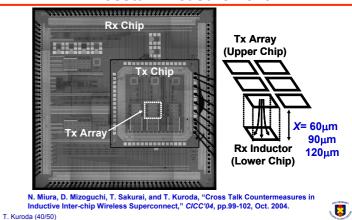
T. Kuroda (39/50)



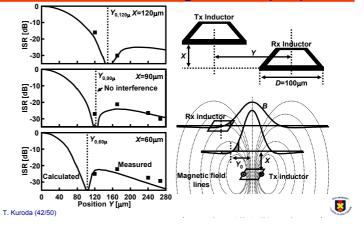


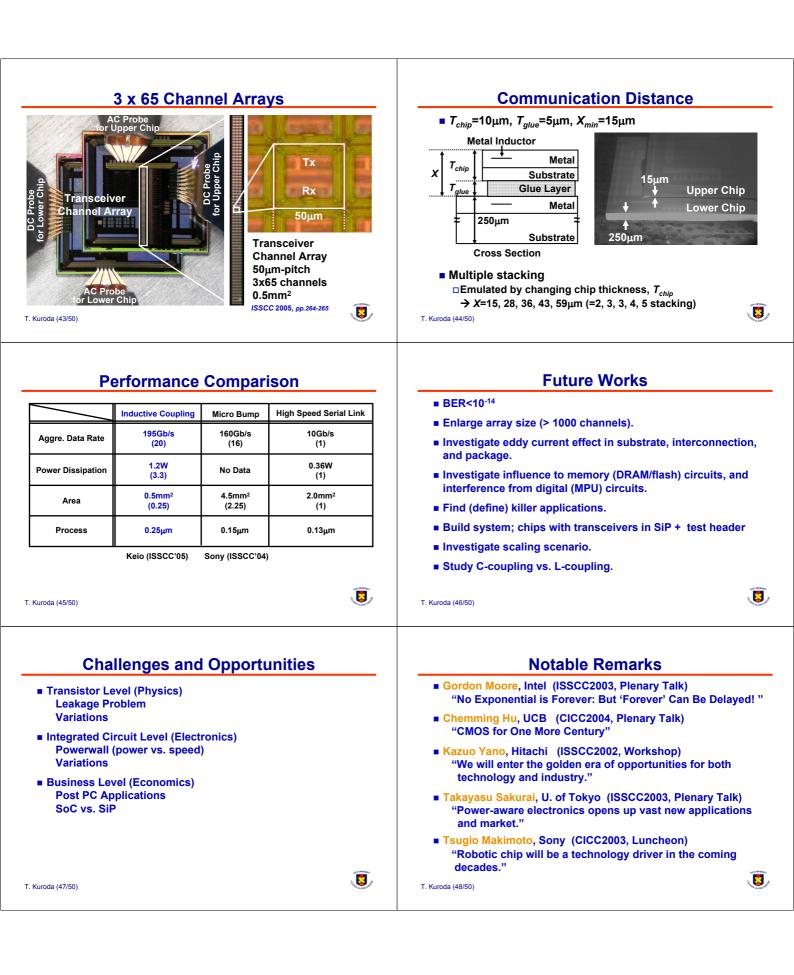


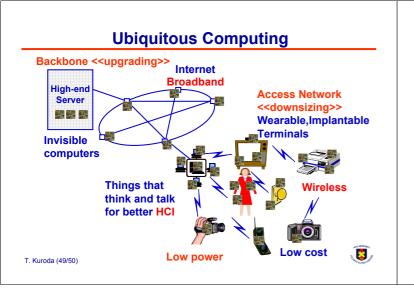
Crosstalk Measurement



Interference to Signal Ratio (ISR)







Summary

Scaling of CMOS integrated circuits is becoming difficult, due to increase in power dissipation and device variations. Post-CMOS device for mass production, however, is not on the horizon. Even though speed of technology improvement may be slowed down compared to that before, the semiconductor industry and technology will continue to develop and improve remarkably for more than a century. We will enter the golden era of opportunities for both technology and industry. Computers and communications will be scaled further, merged together, and materialized in consumer applications. Computers will be small terminals that are implanted in every thing. They will be connected to a network by a wireless technology so that they can be used comfortably wherever you may go. In order to make this ubiquitous society possible, intelligent interface based on low-power, high-speed CMOS design is inevitable. Especially, three key technologies are essential: short range wireless interface and intelligent HCI for access networks, and broadband high-speed links for backbone networks. In addition, integrated design engineering through system, circuit, device, process, lithography, and packaging, will be essential, as well as extended learning and knowledge such as nanotechnology, biology, and robotics. The era transited from 'Know-How (craftsman's age)' to 'Know-What (visionary's age)', and will transit to 'Know-Who (producer's age)'. Role of academia will be more and more important.

T. Kuroda (50/50)