Interconnection from Design Perspective

Takayasu Sakurai

Center for Collaborative Research and Institute of Industrial Science, University of Tokyo,
7-22-1 Roppongi, Minato-ku, Tokyo, 106-8558 Japan
Phone: +81-3-3402-6226, Fax +81-3-3402-6227, tsakurai@iis.u-tokyo.ac.jp

Summary

If we look into the scaling law carefully, we find that three crises exist in realizing VLSI's of the coming years: namely power crisis, interconnection crisis, and complexity crisis (Fig.1). As for the power crisis, the current crisis as is shown in Fig.2 is more important from the viewpoint of interconnections. The IR voltage drop in Fig.3 may demand thicker metal layers in Fig. 4 and new 3-dimensional assembly schemes such as System-in-Package can solve the problems.

The interconnection crisis is depicted in Fig.5. Not MOSFET's but interconnections will be determining cost, delay, power, reliability and turn-around time of the future LSI's. Some of the design issues for the deep submicron interconnects are summarized in Fig.6. Here, signal integrity is becoming one of the major design issues due to the increased coupling capacitance between interconnects (Figs. 7, 8, 9, 10). The increased coupling capacitance relative to grounding capacitance is due to a higher aspect ratio of deep submicron interconnects. Interconnect delay is another big headache of scaled-down interconnects (Figs.11, 12), which can be mitigated by using a buffer insertion technique as shown in Fig. 13. The delay can be reduced by the technique but the power is increased by about 70% due to the inserted buffers as shown in Figs.14 and 15. Another way to decrease the interconnect delay without increasing power is to use a thicker and wider metal layer as in Fig.16 like super-connects described below.

It is just impossible to design LSI's with 100 million transistors from scratch. The complexity crisis can only be solved by the sharing and re-use of design data. So-called IP-based System-on-Chip design style will be preferable. The virtual components are put together on a silicon die to build billion-transistor VLSI's, which can be compared to the present system implementation with printed circuit boards (PCB) and separately packaged VLSI components. However, issues in System-on-Chip are getting clear such as undistributed IP's (i.e. CPU, DSP of a certain company), huge initial investment for masks and development, IP testability, upfront IP test cost, process-dependent memory IP's, difficulty in high precision analog IP's due to noise, and process incompatibility with non-Si materials and/or MEMS. The mask count increases so much if different types of technologies are to be embedded on a single chip (Fig.17). Moreover, the embedding technologies should be developed for each generation and if the types of technologies are diverse, the required engineering efforts are almost impossible to spare.

In order to cope with these issues, a new type of 3-dimensional assembly called System-In-Package has been proposed as shown in Fig.18. The System-In-Package will be using the 'super-connect' technology as shown in Fig.19 with the interconnect thickness of the order of 10µm. The super-connect technology will fill up technology vacuum between the design rule of 1µm order for on-chip interconnects and that of 100µm order for off-chip interconnects. The super-connects in a package used in cooperation with on-chip interconnects will solve the IR drop problem, the clock distribution problem and other problems of the future VLSI's. The co-design of on-chip interconnects and the super-connects in a package is important including the development of a new set of EDA tools.

The super-connect technology fills a gap between off-chip interconnects and on-chip interconnects not only in terms of design rules but also in terms of power, bandwidth, area, cost and turn-around-time as is shown in Fig.20. The major issue in realizing the System-in-Package, however, is to establish a method to select known good dies before assembly. It is very difficult to test a chip at an operating speed at a wafer level without a package, since probing needles used for the wafer test cannot handle signals more than several hundred MHz. These days, however, a new test method using a semi-package called an interposer has been proposed. By using the semi-package, it is possible to carry out at-speed testing of a chip, which may solve the known good die problem. The assembly and packaging technology is becoming vital to VLSI's as the following passage from ITRS shows: "There is an increased awareness in the industry that assembly and packaging is becoming a differentiator in product development."

The overall future perspective of VLSI's in 2014 is shown in Fig. 21.

Here, I would like to add one important piece of information as to the RC delay of interconnect and its behavior due to scaling. It is known that by inserting buffers (or sometimes they is called repeaters), the delay of a long interconnect can be lowered. Let us think about the delay of the buffered interconnect system. The delay can be approximately expressed as below.

$$t_{05} \approx 0.377 R_{INT} C_{INT} + 0.693 (R_T C_T + R_T C_{INT} + R_{INT} C_T)$$

C_{INT} is capacitance of interconnect, R_{INT} is resistance of interconnect, C_0 is gate capacitance of minimum width MOSFET and R_0 is gate effective resistance of minimum width MOSFET. If the interconnect is divided into k sections and (k-1) buffers are inserted, the total delay of the buffered interconnect system is expressed as follows.

Delay

$$\approx k\left[p_1\,\frac{R_{INT}}{k}\,\frac{C_{INT}}{k} + p_2\left(\frac{R_0}{h}hC_0 + \frac{R_0}{h}\frac{C_{INT}}{k} + \frac{R_{INT}}{k}hC_0\right)\right]$$

, where h denotes the gate size of a buffer. What should be optimized here are k and h to minimize the Delay expressed in the above formula. By differentiating in terms of h and k, and setting the derivative equal to zero, it is easy to obtain the optimum h, h_{OPT}, and optimum k, k_{OPT}, as follows.

$$\frac{\partial \text{Delay}}{\partial h} = 0 \rightarrow h_{OPT} = \sqrt{\frac{C_{INT}R_0}{R_{INT}C_0}}$$

$$\frac{\partial \text{Delay}}{\partial k} = 0 \rightarrow k_{OPT} = \sqrt{\frac{p_1}{p_2}}\sqrt{\frac{R_{INT}C_{INT}}{R_0C_0}}$$

Then the optimized delay is expressed like below.

$$\text{Delay}_{OPT} = 2\left(\sqrt{p_1p_2} + p_2\right)\sqrt{R_{INT}C_{INT}R_0C_0}$$
$$\approx 2.4\sqrt{\tau_{INT}\tau_{MOS}}$$

Here, τ_{INT} $(=R_{INT}C_{INT})$ is a time constant of the interconnect and τ_{MOS} $(=R_0C_0)$ is a time constant of the gate, which is proportional to a logic gate delay of a certain technology node. P_1 is 0.377 and p_2 is 0.693 for the case of the delay from zero to a half of V_{DD}, but if these values are modified the optimization can be possible for zero to $0.9V_{DD}$ delay and other intermediate values and in this sense, the formula is quite general. The above expression is interesting in that the delay of the buffered interconnect system is a geometric of the interconnect delay itself and the logic gate. Since the scaling factor of the interconnect delay is almost constant and the delay of the logic gate is supposed to improve very rapidly as technology advances, scaling of the delay the buffered interconnect system is supposed to improve slowly.

In the optimum buffered interconnect, the capacitance of the system increases due to the inserted buffers. The total gate capacitance of buffers is expressed as follows.

$$\text{Cap. of gates} = k_{OPT}h_{OPT}C_0 = \sqrt{p_1/p_2}\,C_{INT} = 0.73C_{INT}$$

This means that the total capacitance is increased by 73% compared with the system without buffers. The increase in capacitance in turn increases power consumption.

Transistors		Scaling coefficients
V_{DD}	[V]	$1/k$
Tr. dimensions	[x]	$1/k$
Drain current	[I~1/x x/x V^1.3]	$1/k^{0.3}$
Gate capacitance	[C~1/x xx]	$1/k$
Tr. delay	[d~CV/I]	$1/k^{1.7}$
Tr. power	[P~VI~CVV/d]	$1/k^{1.3}$
Power density	**[p~P/x/x]**	$k^{0.7}$
Tr. density	**[n~1/x/x]**	k^2

Interconnects			
Type		**Local**	**Global**
Scaling scenario		**Scaled**	**Anti-scaled**
Line thickness	[T]	$1/k$	k
Width	[W]	$1/k$	k
Separation	[S]	$1/k$	k
Oxide thickness	[H]	$1/k$	1
Length	[L]	$1/k$	1
Resistance	[R_{INT}~L/W/T]	k	$1/k^2$
Capacitance	[C_{INT}~LW/H]	$1/k$	k
RC delay/Tr. delay	**[D~$R_{INT}C_{INT}$/d]**	$k^{1.7}$	–
Current density	**[J~pWL/V /W/T]**	–	$k^{0.7}$
DC noise / V_{DD}	**[N~JWTR/V]**	–	$k^{1.7}$

Fig. 1 Scaling law. When a device is shrunk to a half, the scaling variable, k, is equal to 2. The hatched quantities will have problems as a technology advances and a device size is miniaturized.
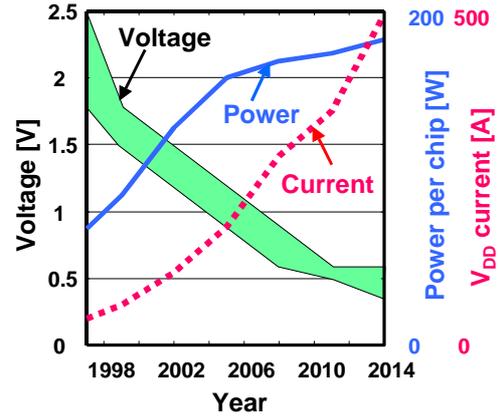
Fig.2 Trend of voltage, power and current taken from ITRS [1]. Since the voltage is scaled down and power is increasing, the current is increasing rapidly, which cause IR drop problem and electro-migration related reliability problem.

$$V_{i,j} = (V_{i-1,j} + V_{i+1,j} + V_{i,j-1} + V_{i,j+1})/4 - r\,I_{i,j}$$

$$\sum I_{i,j} = I,\ \text{Sheet resistance} = R$$

Take IR as unity voltage drop
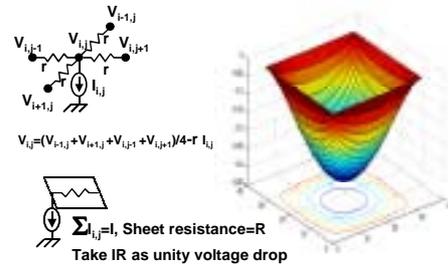
Fig.3 IR voltage drop. If pads are placed around a chip and the chip consumes current uniformly over area, the peak IR drop is observed at the center of the

chip and the amount of the IR drop is about 30% of IR, where I is the total current consumed and R is sheet resistance of the power supply metal sheet.

Unscaled / anti-scaled
- Clock
- Long bus
- Power supply

Scaled interconnect
- Signal
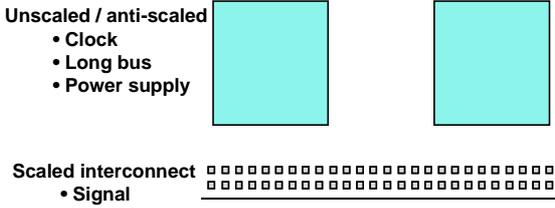
1V 20W → 20A current
2% noise on VDD & VSS → ~0.02V / 20A → ~10μm thick Cu
Thick layer interconnect, area pad, package are co-designed.

Fig. 4 Interconnect cross-section and noise. Even for intermediately power consuming chip needs very thick metal like 10um. This type of thick metal will be in a package and area pads and co-design of VLSI and a package become necessary.
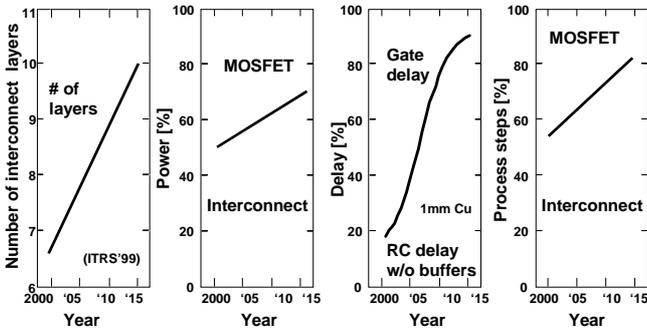
Fig.5 Trend in power, delay, area, and TAT of interconnect. Usually a figure of merit of VLSI is measured by PDAT, where P is power, D is Delay, A is area and T is Turn-around-time. Then the figure of merit of VLSI will be determined by interconnect designs rather than transistor designs.

Larger current
   IR drop (static and dynamic)
   Reliability (electro-migration)

Smaller geometry / Denser pattern
   RC delay
   Signal Integrity
   Crosstalk noise
   Delay fluctuation

Higher speed
   Inductance
   EMI

Fig. 6 Issues in deep submicron interconnect design
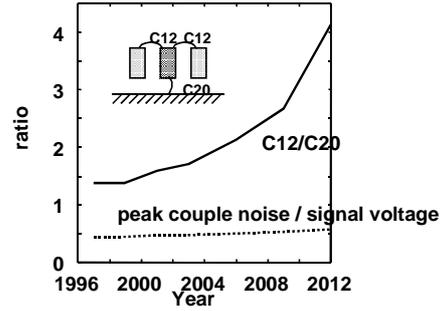
Fig.7 Trend in coupling capacitance. The higher aspect ratio of interconnects increases the coupling capacitance among lines (C12), in relative to the grounding capacitance (C20).

H.Kawaguchi and T.Sakurai, "Delay and Noise Formulas for Capacitively Coupled Distributed RC Lines," ASPDAC, Digest of Tech. Papers, pp.35-43, Feb. 1998.
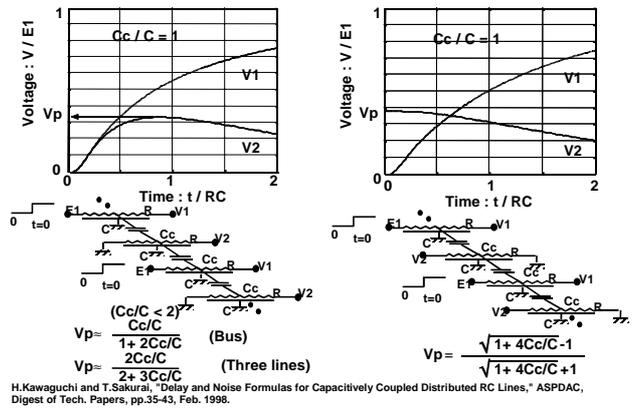
Fig.8 Noise due to coupling capacitance.
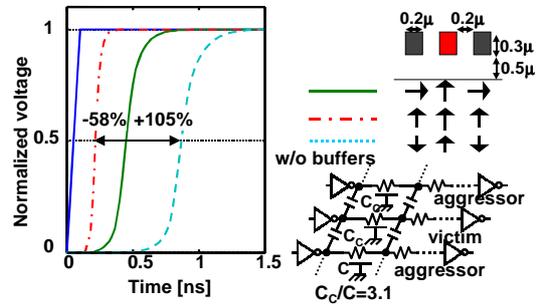
$$\frac{t_{pd}}{RC} = \frac{1}{2} + 2\eta - \sqrt{1 + 8\eta + 6\eta^2} \; \frac{1}{\sqrt{6}} \log \frac{e}{2}$$
$$\approx 1.63\eta + 0.37 \, (\eta \leq 2) \; (\eta = C_c / C)$$

Fig. 9 Delay fluctuation due to coupling capacitance. The delay of interconnect may fluctuate about a factor of 4 between in-phase drive and anti-phase drive of adjacent lines. This phenomenon makes a designer to think about voltage behaviors of adjacent lines other than the delay of the interconnect, which is a nightmare.
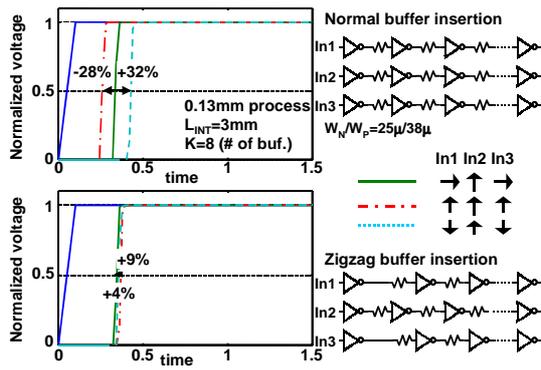
Fig.10 Delay fluctuation due to coupling capacitance can be mitigated by using buffer insertion technique. The fluctuation will be reduced more by staggering the location of buffers even in the worst case [2].
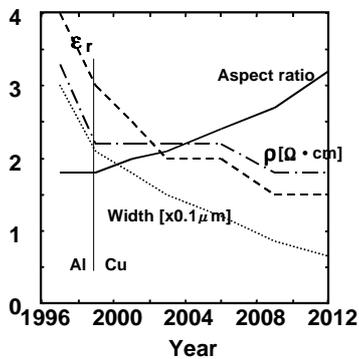
Fig.11 Interconnect parameters trend from ITRS'97.
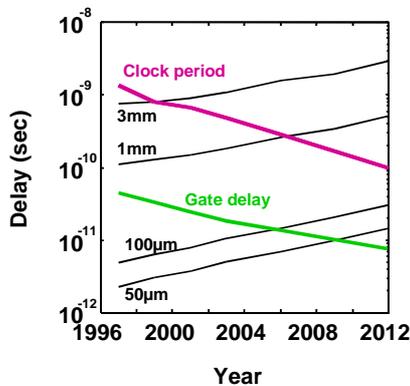
Fig.12 RC delay and gate delay. If we use the minimum size (cross-section wise) interconnect, the signal can not be propagated to 1mm distance in one clock cycle.
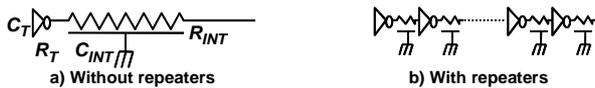
Fig.13 Delay reduction by buffer insertion
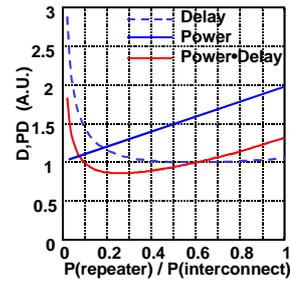
Delay optimized
→P: P(repeater)=0.73 P(interconnect)
Power•Delay optimized
→D: 1.09 Dopt
→P: P(repeater)=0.26 P(interconnect)
→PD: 0.86 of Dopt case

Fig.14 Delay and power optimization for buffers. If delay is to be minimized, the inserted buffers increase the capacitance of interconnect by 73%, while if power-delay product is to be minimized, the capacitance increase is reduced to 26%
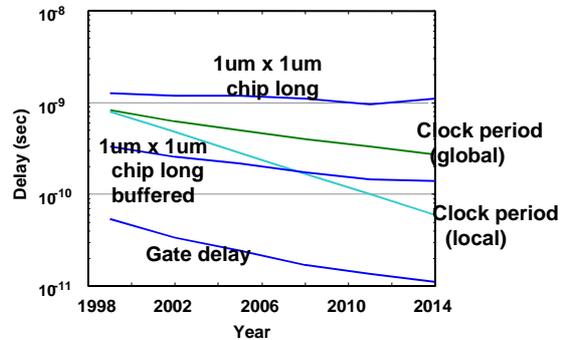
Fig.15 Trend in interconnection delay with and without buffer insertion.
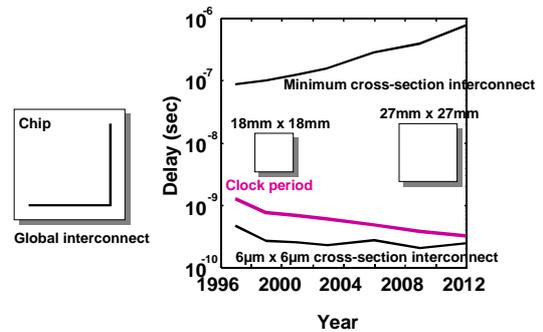
Fig.16 RC delay of global interconnects. If a thick metal layer is available which could be a layer in a package, by using 6µm x 6µm interconnect, the RC delay can be reduced to the point where the signal can propagate within a chip in a clock cycle. This approach does not increase capacitance and hence power in contrast to the buffer insertion approach.
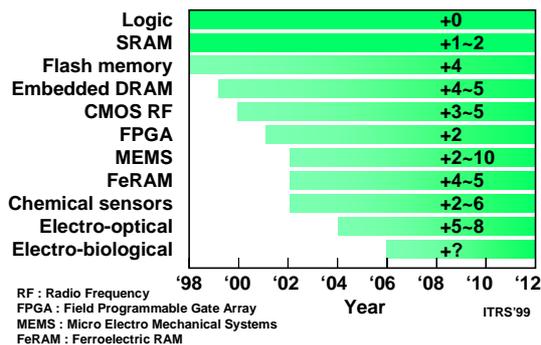
Fig.17 Technologies integrated on a chip. The numbers in the bars show the increase of mask steps extra to a logic process.

- Smaller area
- Shorter interconnect
- Optimized process for each die (Analog, DRAM, MEMS…)
- Good electrical isolation
- Through-chip via

- Heat dissipation is an issue
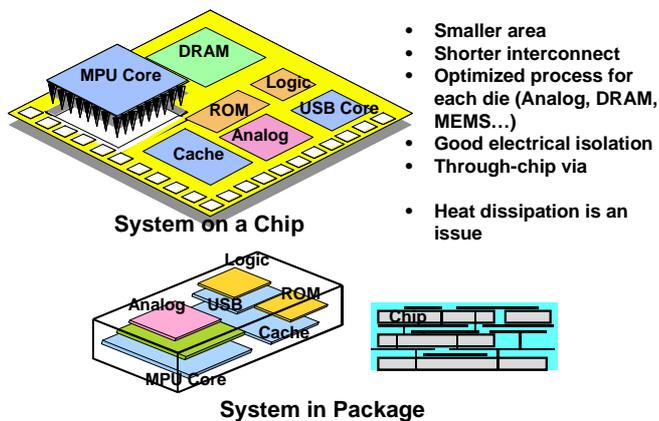
Fig.18 System-on-Chip vs. System-In-Package.
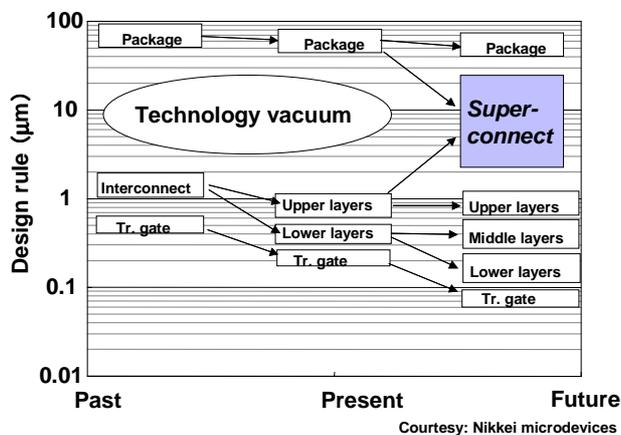
Fig.19 Super-connect technology which fills up a technology vacuum between on-chip and package interconnects.
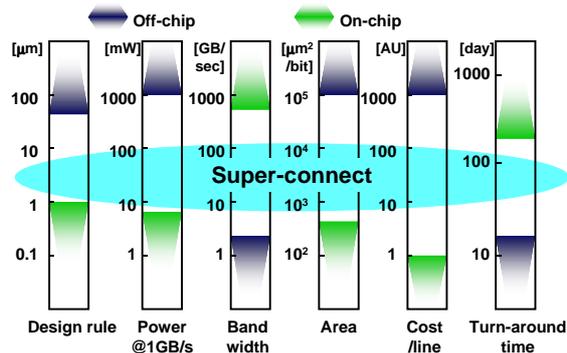
Fig.20 Performance gat between off-chip and on-chip interconnects. The super-connects fill a gap.

Year	Unit	1999	2014	Factor
Design rule	µm	0.18	0.035	0.2
Tr. Density	/cm2	6.2M	390M	30
Chip size	mm2	340	900	2.6
Tr. Count per chip (µP)		21M	3.6G	170
DRAM capacity		1G	1T	1000
Local clock on a chip	Hz	1.2G	17G	14
Global clock on a chip	Hz	1.2G	3.7G	3.1
Power	W	90	183	2.0
Supply voltage	V	1.5	0.37	0.2
Current	A	60	494.6	8
Interconnection levels		6	10	1.7
Mask count		22	28	1.3
Cost / tr. (packaged)	µcents	1735	22	0.01
Chip to board clock	Hz	500M	1.5G	3.0
# of package pins		810	2700	3.3
Package cost	cents/pin	1.61	0.75	0.5

Fig.21 VLSI's in 2014

References

[1]   International Technology Roadmap for Semiconductors: 1999 edition. Austin, TX: International SEMATECH, 1999.

[2]   Koichi Nose, private communications.