Three-dimensional global interconnect based on a design window*

Qian Li-Bo(钱利波), Zhu Zhang-Ming(朱樟明)†, and Yang Yin-Tang(杨银堂)

Microelectronics School, Xidian University, Xi'an 710071, *China*

(Received 29 May 2011; revised manuscript received 21 June 2011)

Based on a stochastic wire length distributed model, the interconnect distribution of a three-dimensional integrated circuit (3D IC) is predicted exactly. Using the results of this model, a global interconnect design window for a giga-scale system-on-chip (SOC) is established by evaluating the constraints of 1) wiring resource, 2) wiring bandwidth, and 3) wiring noise. In comparison to a two-dimensional integrated circuit (2D IC) in a 130-nm and 45-nm technology node, the design window expands for a 3D IC to improve the design reliability and system performance, further supporting 3D IC application in future integrated circuit design.

Keywords: three-dimensional integrated circuit, design window, wiring resource, bandwidth

PACS: 84.30.–r, 84.30.Bv **DOI:** 10.1088/1674-1056/20/10/108401

1. Introduction

The unprecedented development of computer and information technology demands ultra-large-scale integrated (ULSI) circuits with increasing functionality and performance at minimum cost and power consumption. The feature size of transistors is being aggressively shrunk to meet this demand. However, in turn, this has introduced some very serious problems. Interconnect effects such as signal delay, distortion and crosstalk are the dominant factors limiting overall performance.[1−3] Wiring requirements, particularly that of a global interconnect, is fast becoming a key bottleneck to system performance and reliability. Therefore, it is imperative to optimize global interconnects for present and future giga-scaled integrated systems.

Many methods have been suggested to relieve the interconnect limits in interconnect systems for giga-scale integration. IBM firstly employed the copper and low-k insulator instead of an aluminum interconnect to improve interconnect delay. However, the linear thickness and resistivity limited the advantage of copper interconnects for future technology generation. In addition, the implementation of low-k material was difficult because of reliability issues.[4] Repeater insertion methodology reduced the delay caused by a long interconnect by changing the quadratic delay-length dependence to linear proportionality.[5,6] However, for a large high-performance design, the number of such repeaters would be prohibitively high and can cost significant silicon area and power consumption. Utilizing an on-chip network (NOC) instead of a global interconnect decreased global routing congestion and the efficiency of sending packets in a single global network rather than the signal in global wires was improved, but required a new design architecture different from conventional microprocessor architecture.[7] However, on-chip optical interconnection had high bandwidth and consumed less power for long-distance communication, but was incompatible with standard complementary metal-oxide-semiconductor (CMOS) technology.[8] A new design technique should be developed to overcome 2D interconnect design limitation. Three-dimensional (3D) integration is emerging as a promising solution that can form highly integrated system by vertically stacking and connecting various active device layers as shown in Fig. 1.[9] One of the most important benefits of 3D chip over a traditional 2D design is the reduction in global interconnect and total wire-length, thereby providing higher performance, lower power consumption, and larger design window.

Based on a stochastic wire length distribution

*Project supported by the National Natural Science Foundation of China (Grant Nos. 60725415 and 60676009) and the Natural Science and Technology Major Project of the Ministry of Science and Technology of China (Grant No. 2009ZX01034-002-001-005).

†Corresponding author. E-mail: zmyh@263.net

© 2011 Chinese Physical Society and IOP Publishing Ltd

model, this paper establishes a global interconnect design window for giga-scale SOC by evaluating the constraints of wiring requirements, bandwidth and crosstalk noise. The rest of the paper is organized as follows. In Section 2, a stochastic wire length distribution model for a 3D-SOC is rigorously derived as an extension for the previous 2D wiring distribution model. Then the interconnect design window for a 3D IC is established by considering the constraints of wiring requirements, bandwidth and crosstalk in Section 3. Comparative results with 2D-SOC in Section 4 reveals 3D integrated technology advantages in system performance and design reliability. Finally, a conclusion is presented in Section 5.

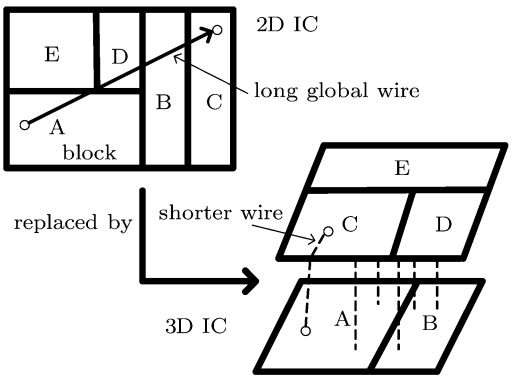


Fig. 1. Interconnect architecture of 2D and 3D IC.

2. Interconnect stochastic wire-length distribution model

System-on-chip clock frequency, power consumption and chip size are largely determined by wiring requirements. It is, therefore, imperative to gain a thorough understanding of wire-length distribution in circuit front-end design that can be used to optimize the design architecture and provide rigorous reference for IC layout design. In the early days, empirical formulas related to chip size were often employed to estimate the interconnect distribution, but that easily led to great error between the assumption and the realistic design. With the new advancements in computer-aided-design (CAD) tools, some new wiring distribution methodologies are introduced. Based on Rent's Rule, Davis firstly presented an interconnect density function (idf) and derived a complete wire-length distribution model[10]

$$I(a < l < b) = \int_a^b i(l)\,\mathrm{d}l, \tag{1}$$

where $I(l)$ is idf, defined as the interconnect number between length $l = a$ and $l = b$; l is the interconnect length in unit of gate pitch; $i(l)$ is the discrete wire-length distribution that gives the expected interconnect number for a given length l,

$$i(l) = M_t(l) I_{\exp}(l), \tag{2}$$

where $I_{\exp}(l)$ is the expected number of interconnect connecting gate pairs that are separated by length l; $M_t(l)$ is the number of gate pairs separated by a length l.

However, the wire length model presented by Davis is limited to traditional 2D architecture. For a 3D integrated circuit with the number of active layer $S \geq 2$, the wire length distribution model consists of horizontal and vertical interconnects as follows:[11]

$$\begin{aligned} i_{3\mathrm{D}}(l) &= M_{t_3\mathrm{D}}(l) \cdot I_{\exp}(l) = i_h(l) + i_v(l), \\ i_h(l) &= I_{\exp}(l) \sum_{v=0}^{S-1} M_z(v) M_s(l - rv), \\ i_v(l) &= I_{\exp}(l) \sum_{v=0}^{S-1} N_s(S - v)\delta(l - rv), \end{aligned} \tag{3}$$

where $M_{t_3\mathrm{D}}(l)$ is the number of gate pairs separated by l gate pitch in 3D IC. $M_z(v)$ is the number of die pairs separated by v vertical pitch. $M_s(l)$ and N_s are the number of gate pairs separated by l gate pitch and the number of gate pairs in a die, respectively, and r is the die-to-gate pitch ratio.

Based on the wire-length distribution model mentioned above, the interconnect distributions for 2D-SOC and 3D-SOC consisting of one-thousand of logic gates are compared, where Rent's coefficient and exponent are 3 and 0.75, respectively. The distribution curves in Fig. 2 show a significant reduction of the number of 3D global interconnects due to a vertical interconnect instead of a long global net. Similarly, as the number of active layers increases, the length of the longest interconnects decreases greatly. This reduction results from both shorter corner-to-corner distances and from each gate having more neighbouring gates in 3D IC. The relationship between the number of active layer S and the longest global interconnect I_{longest} is given as

$$I_{\mathrm{longest}} = 2\left(\sqrt{\frac{N_{\mathrm{t}}}{S}} - 1\right) + r(S - 1), \tag{4}$$

where N_{t} is the total number of gates in the SOC. Assuming that the system is sufficient large such that

$N_t \gg r$, the length of the longest global interconnect scales inversely with the square root of the number of active layer. Thus the longest interconnect in 3D IC can be reduced in length by 30% by adding only one active layer to a traditional 2D architecture. For the SOC being dominated by the interconnect effect, this 30% reduction denotes a decrease in wire delay about 50%, thus improving the system performance significantly.

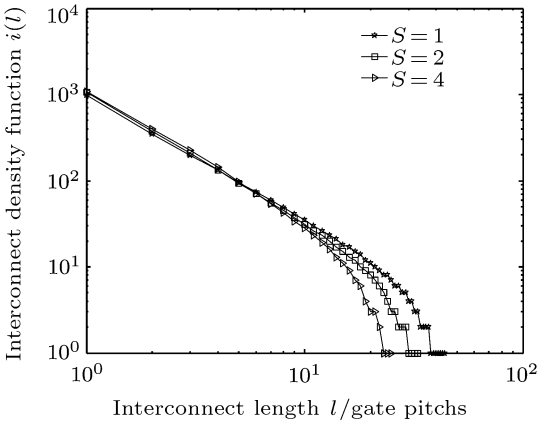


Fig. 2. Interconnect wire-length distribution for 2D and 3D SOC.

3. Global interconnect design methodology

The global interconnect design is one of the main barriers to achieving giga-scaled integration for the present and future giga-scaled integrated (GSI) system, that has a huge impact on the clock frequency, power consumption and chip size. Figure 3 illustrates the global interconnect architecture that consists of the signal, clock and power-supply networks, where w, h, and s are signal interconnect width, height and space, respectively; t_{ox} is the dielectric thickness. In determining the values of w, h, s, and t_{ox} for optimal interconnect design, three design constraints for global interconnects are introduced: 1) wiring resource, 2) wiring bandwidth and 3) wiring noise.

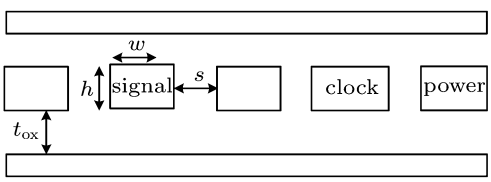


Fig. 3. Cross-section schematic diagram of a global interconnect.

3.1. Wiring resource

The wiring resource requirement is the primary design constraint for an interconnect in a GSI system that consists of local interconnects used for connecting in a functional block and a global interconnect between these blocks. As the local interconnect pitch is relatively small, it can be ignored in the early design stage. Then the global wiring resource can be expressed as follows, which is largely dependent on the system routing, placement and power consumption, and often limited by the SOC area, A_c,

$$A_{\text{signal}} + A_{\text{power}} + A_{\text{clock}} \leq n_{\text{level}} A_c, \tag{5}$$

where n_{level} is the number of metal wire layers in a multilevel wiring network; A_{signal}, A_{clock}, and A_{power} are the wiring area for the global signal, clock and power-supply networks, respectively. The area required for the global signal routing is given by[12,13]

$$A_{\text{signal}} = \frac{(w+s)L_{\text{tot}}}{\eta}, \tag{6}$$

where η is signal routing efficiency. L_{tot} is the total length of a global interconnect that can be obtained by the stochastic length-net distribution model mentioned above

$$L_{\text{tot}} = \int_{l=L_{\text{loc}}}^{l_{\text{longest}}} l i(l) \mathrm{d}l, \tag{7}$$

in which, L_{loc} and l_{longest} are the lengths of the longest local and global interconnects, respectively.

The wiring area of a power-supply network has a strong relationship with power dissipation and IR drop in SOC. Based on the optimal power-supply design model provided by Joyner,[14] the minimum power-supply network satisfying the IR drop limitation can be given by

$$A_{\text{power}} = n_{\text{level}} A_c \frac{P_{\text{tot}}\rho}{16\delta V_{\text{dd}}^2 h n_{\text{pad}}} \left(2 - \frac{P_{\text{tot}}\rho}{16\delta V_{\text{dd}}^2 h n_{\text{pad}}}\right), \tag{8}$$

where P_{tot} is the total chip power, ρ is the metal resistivity, n_{pad} is the number of power and ground pads, and δ is the ratio of the worst-case IR drop to power supply voltage V_{dd}.

Since the ratio of the wiring area for the clock network to total chip area is about 1%, the A_{clock} is negligible in this paper. According to the definition of the wiring area, the global interconnect dimension

from the wiring area constraint can be determined by substituting Eqs. (6), (7), and (8) into Eq. (5)

$$w + s \leq \eta \frac{n_{\text{level}} A_{\text{c}}}{L_{\text{tot}}} \left(1 - \frac{P_{\text{tot}} \rho}{16 \delta V_{\text{dd}}^2 h n_{\text{pad}}}\right)^2. \quad (9)$$

3.2. Wiring bandwidth

Since clock signal is the fastest signal in SOC, it requires a high bandwidth interconnect to preserve the rise time and transmission integrity when the clock signal passes through the interconnect. Therefore, the bandwidth of global interconnects is imposed by clock frequency as[15]

$$f_{\text{clk}} \leq \frac{1}{2 t_{\text{rslow}}}, \quad (10)$$

where f_{clk} is the system clock frequency, t_{rslow} is the slowest transition time of the distribution RC interconnect, calculated by the Sakurai delay model[16]

$$t_{\text{rslow}} = \frac{t_{90\%} - t_{10\%}}{0.8} = 1.1 rc \cdot l_{\text{longest}}^2 \quad (11)$$

in which l_{longest} is the longest global interconnect, r and c are the interconnect resistance and capacitance unit length, respectively, that can be expressed by Eqs. (12) and (13)

$$r = \frac{\rho}{wh}, \quad (12)$$

$$c = 2(c_{\text{gnd}} + c_{\text{m}}) = 2\left(\frac{\varepsilon_0 \varepsilon_{\text{r}} w}{t_{\text{ox}}} + \frac{\varepsilon_0 \varepsilon_{\text{r}} h}{s}\right), \quad (13)$$

where ε_0 and ε_{r} are the permittivity of free space and the relative permittivity, respectively.

Substituting Eqs. (11), (12), and (13) into Eq. (10), the global interconnect dimension from the wiring bandwidth constraint can be given as

$$w \cdot s \geq \left(\frac{1}{4.4 f_{\text{clk}} l_{\text{longest}}^2 \rho \varepsilon_0 \varepsilon_{\text{r}}} - \frac{1}{h \cdot t_{\text{ox}}}\right)^{-1}. \quad (14)$$

3.3. Wiring noise

Interconnect coupling noise or crosstalk is a major design constraint for global wires, caused by the couple capacitance and inductance with electromagnetic interaction. For signal integrity and reliability, the wiring noise should satisfy this constraint as shown below[17,18]

$$\frac{V_{\text{peak}}}{V_{\text{dd}}} \leq P_{\text{noise}}, \quad (15)$$

where V_{peak} represents peak noise, determined by numerous factors such as rise time and the width of input pulse, interconnect length and repeater parasitic resistance. P_{noise} is the maximum noise allowed as a percent of V_{dd}. Based on the distributed RC crosstalk noise model, the interconnect peak noise can be expressed simply by[16]

$$\frac{V_{\text{peak}}}{V_{\text{dd}}} = \frac{1}{2} \frac{c_{\text{m}}}{c_{\text{m}} + c_{\text{gnd}}}. \quad (16)$$

Applying the parallel plane capacitance model in Eq. (13), and substituting Eq. (16) into Eq. (15), we can obtain the global interconnect dimension from the wiring noise constraint

$$ws \geq \frac{1 - 2P_{\text{noise}}}{2P_{\text{noise}}} h t_{\text{ox}}. \quad (17)$$

3.4. Global interconnect design window

As demonstrated above, the three design constraints, 1) wiring resource, 2) wiring bandwidth, and 3) wiring noise, establish a number range for the optimization of global interconnect parameters, which are defined as the global interconnect design window. Assuming $w = s$, $t_{\text{ox}} = h$ and using the technology parameters at 130-nm technology node listed in Table 1,[19] a new global interconnect design window for 2D SOC with four metal wire levels is plotted in Fig. 4, where interconnect width w and height h serve as x and y axes, respectively. It is shown that the central triangular region is the design window for which all three constraints are satisfied simultaneously, and the other region beyond limitation is called the forbidden zone for circuit design. Three corners of the window A, B, and C correspond to the minimum 1/2 wire pitch, minimum aspect ratio and maximum clock frequency, respectively. The D point in the centre of the design window

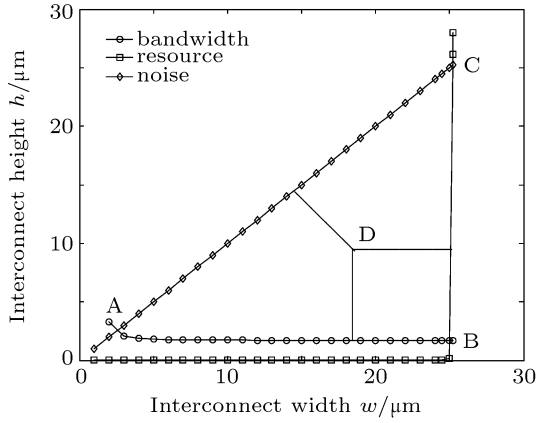


Fig. 4. Interconnect design window in wiring resource, bandwidth and noise constraints.

Table 1. Key interconnect technology parameters used for SOC design.

	130 nm	45 nm
$V_{\rm dd}$/V	1.5	0.6
$A_{\rm c}$/mm^2	450	817
$f_{\rm clk}$/GHZ	1.2	3
$P_{\rm tot}$/W	130	174
$l_{\rm longest}$/cm	4.2	5.7
$l_{\rm tot}$/m	17.82	63.87
$P_{\rm noise}$/%	20	20
η/%	50	50
$n_{\rm pad}$	1536	1536
δ/%	5	5
$N_{\rm t}$/M	20	100
$\varepsilon_{\rm r}$	3.3	2.1

is regarded as the maximum reliability design position. Utilizing this design methodology reasonably predicts critical physical parameters for a global interconnect network in an early design stage, satisfying SOC design constraints for wiring resource, bandwidth and noise. Its simplified physical model can be used to guide SOC design as a simple quantitative tool.

4. Comparison and validation

The maximum clock frequency and power dissipation aggressively increase with advances in process technology, that shrinks the interconnect design window continuously and has been a tremendous challenge for global interconnect design in the nanometer generation. Though, to some extent, multi-level metal wire technology can mitigate the design stress, while resulting in a cost increase and yield problem. Since the total length of wire and the longest global wire scale inversely with the square root of the number of active layers, as shown in Fig. 2, the wiring resource demand and bandwidth limitation can be relaxed by use of more active layers. The 3D integrated technology will be an efficient and flexible design strategy for present and future GSI systems. Combinations of process parameters in Table 1 and three constraints, Figs. 5 and 6 illustrate global interconnect design windows for systems of 1, 2, and 4 active layers at 130-nm and 45-nm technology nodes. The comparative results indicate that the design window shrinks with advancements in process technology; the window for a single active layer at the 45-nm node is significantly smaller than that at the 130-nm node, implying that increasing restriction is placed on global design and instability due to process variation. For a 3D IC with multi-active layers, the design window

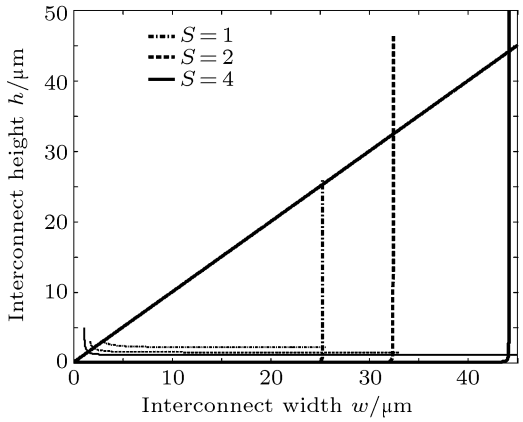


Fig. 5. Interconnect design window for systems of 1, 2, and 4 active layers at 130-nm technology node.

is expanded greatly, increasing the width of the safe band that separates the central design point D from the design constraint, and enhances the tolerance of larger deviations in interconnect dimension due to process variation, further improving the design margin and circuit reliability.

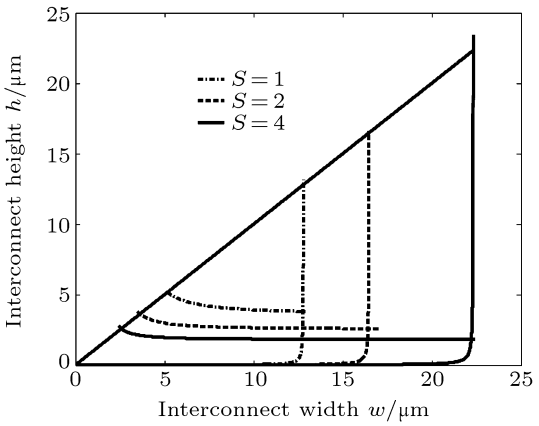


Fig. 6. Interconnect design window for systems of 1, 2, and 4 active layers at 45-nm technology node.

In addition, 3D integrated architecture can efficiently improve system performance by increasing the number of active layers. As shown in Figs. 5 and 6, the maximum frequency corner C shifts upwards with the increase of the number of active layers, indicating the higher speed limit of the interconnect architecture. To find the relationship between the maximum clock frequency $f_{\rm max,clk}$ and the number of active layers S, the wiring area for the power-supply network is ignored for the purpose of simplicity of calculation,

then solving Eqs. (9) and (17) for interconnect width and height

$$w = \frac{\eta \cdot n_{\text{level}} A_{\text{c}}}{2L_{\text{tot}}}, \tag{18}$$

$$h = \sqrt{\frac{2P_{\text{noise}}}{1 - 2P_{\text{noise}}}}. \tag{19}$$

Substituting these values into Eq. (14), the maximum clock frequency expression can be written as

$$f_{\text{max,clk}} \approx \frac{\eta^2 \cdot n_{\text{level}}^2 \cdot A_{\text{c}}^2 \cdot P_{\text{noise}}}{8.8 \cdot \rho\varepsilon_{\text{r}}\varepsilon_0 \cdot L_{\text{tot}}^2 \cdot l_{\text{longest}}^2} \propto S. \tag{20}$$

As the longest global wire l_{longest} is inversely proportional to the square root of the number of active layers, the maximum clock frequency exhibits the linear proportion to S. Taking into account the reduction in the total interconnect length further, the maximum clock frequency for 3D-SOC with two active layers is more than two times that of 2D-SOC, that improves the system performance significantly without adoption of additional circuit block and design innovation.

Even though 3D integrated architecture seems to be an ideal alternative to alleviate the interconnect effect in 2D-SOC, there are several key obstacles to the implementation of 3D-SOC. First of all, multiple heat sources are generated with vertical stacking, substrate thinning technology further yields the poorer heat dissipation and thermal gradient in 3D, leading to design challenge in timing and reliability. New high-density packing technology is demanded to satisfy the requirement of high bandwidth and high density I/Os. Furthermore, TSV technology is moving past the feasibility phase into the commercial phase, but the industry still lacks of a uniform standard for TSV such as tsv fabrication and filling method.[20] Despite of these disadvantages, the potential benefit in performance and reliability enables 3D integration to be the main-stream development trend for future integrated circuit.

5. Conclusion

Based on a stochastic wire length distributed model, the interconnect distribution of three dimensional integrated circuit is predicted exactly. Using the results of this model, a global interconnect design window for giga-scale system-on-chip is established by evaluating the constraints of 1) wiring resource demand, 2) wiring bandwidth requirement, 3) wiring noise limit. In comparison with a two-dimensional integrated circuit in 130-nm and 45-nm technology nodes, the design window expands for a 3D IC to improve the design reliability and system performance, further supporting 3D IC application in future integrated circuit design.

References

[1] Zhu Z M, Xiu L P and Yang Y T 2010 *Chin. Phys.* B **19** 077802

[2] Dong G, Yang Y, Chai C C and Yang Y T 2010 *Chin. Phys.* B **19** 110202

[3] Zhu Z M, Qian L B and Yang Y T 2009 *Acta Phys. Sin.* **58** 2631 (in Chinese)

[4] Heidenreich J and Edelstein D 1998 *Proceedings of IEEE International Interconnect Technology Conference*, April 5–6, 1998 San Francisco, USA p. 151

[5] Banerjee K and Mehrotra A 2002 *IEEE Trans. Electron Dev.* **49** 2001

[6] Zhu Z M, Qian L B and Yang Y T 2009 *Chin. Phys.* B **18** 1188

[7] Agarwal A, Iskander C and Shankar R 2009 *Journal of Engineering, Computing and Architecture* **3** 1423

[8] Ohashi K, Nishi K and Shimizu T 2009 *Proceeding of the IEEE* **97** 1186

[9] Pavlidis V F and Friedman E G 2009 *Three-Dimensional Integrated Circuit Design* (Boston: Morgan Kaufmann) p. 166

[10] Davis J A, Vivek K D and Meindl J D 1998 *IEEE Trans. Electron Dev.* **45** 580

[11] Joyner J D 2003 *Opportunities and Limitations of Three-dimensional Integration for Interconnect Design* (Atlanta: Georgia Institute of Technology, Electrical and Computer Engineering)

[12] Davis J A and Meindl J D 2003 *Interconnect Technology and Design for Giga Scale Integration* (Boston: Kluwer Academic) p. 160

[13] Zhu Z M, Wan D J and Yang Y T 2010 *Chin. Phys.* B **19** 097803

[14] Joyner J W, Payman Z H and Meindl J D 2004 *IEEE Trans. VLSI System* **12** 367

[15] Chen G Q and Friedman E G 2006 *IEEE Trans. VLSI System* **14** 161

[16] Sakurai T 1993 *IEEE Trans. Electron Dev.* **40** 118

[17] Wang X, Shan Z Y, Zhu Y T and Shao B X 2006 *Acta Electron. Sin.* **34** 214 (in Chinese)

[18] Zhu Z M, Hao B T, Yang Y T and Li Y J 2010 *Chin. Phys.* B **19** 127805

[19] Semiconductor Industry Association http://www.itrs.net

[20] Motoyoshi M 2009 *Proceedings of the IEEE* **97** 43