Transactions of the Materials Research Society of Japan 29 [3] 717-722 (2004)

Technology Requirements for 45nm Node CMOS and beyond

Tatsuo Noguchi, Yoichi Takegawa, Amane Ooishi and Masaaki Iwai
System LSI Div., Semiconductor Company, TOSHIBA Corporation
8, Shinsugita-cho, Isogo-ku, Yokohama, 235-8522, Japan
Fax: 81-045-770-3194, e-mail: tatsuo.noguchi@toshiba.co.jp

The performance limitation factors for CMOS technology of 45nm node and beyond are clarified. The expectations of the introduction of new materials and technologies to break through those limitations are also discussed. In addition to these, the performance trend of the CMOS technology is summarized.
Key words: CMOS, Roadmap, MOSFET, parasitic capacitance, parasitic resistance, Interconnection, Cu, Low-k

1. INTRODUCTION

With the advent of the internet era, the required rate of miniaturization of CMOS technology has been accelerated not only from the scaling demand of system on a chip (SoC), but also by the need to manage huge data at high speed. Up to the 0.25um node, a new generation technology was released every three years [1]. Since the 0.18um node, this rate of introduction accelerated to two years [2]. From the 90nm node we expect a new generation to be released every 1.5 years.

In this paper, we clarify, the key technology components and challenges of MOSFET's and interconnect required to continue to achieve performance improvement down to the 45nm node and beyond. In addition, the feasibility of the 45nm node and beyond is studied.

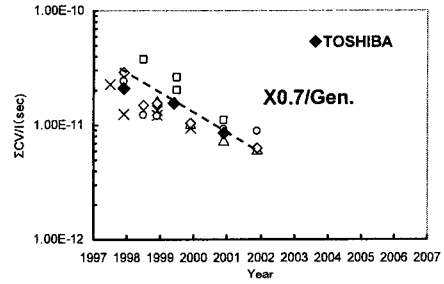


Fig.1 Trend of CMOS performance

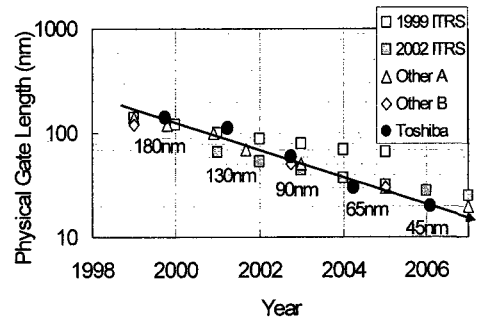


Fig.2 Gate Length Scaling Trend

2. IMPROVEMENT OF MOSFET's CHARCTERISTICS

2-1.Trend of CMOS technology and the target performance of 45nm node and beyond

Figure 1 shows the performance trend of CMOS technology. In this figure each plot shows the normalized performance (CV/I) of various organizations from various technical papers. C is the MOSFET parasitic capacitance, V is the applied power, and I is the MOSFET drive current. When the rate of introduction of new generations accelerates to every one and half year, the rate of the performance enhancement should be 70% per generation. To achieve this target, we have to not only increase MOSFET drive current but also decrease both parasitic capacitance and resistance of the transistor.

2-2.Analysis of the parasitic capacitance of MOSFET

In general, the empirical scaling factor of the two dimensional patterns, such as active area pitch or metal pitch, is from 70% to 80% between each generations. But scaling of the MOSFET is more aggressive than that of two-dimensional patterns. The shrink factor of gate length is 65% per generation as shown in Figure 2, because of the requirement of the SoC performance improvement. In this paper, a scaling factor of two dimensional pattern is defined as 74% per generation, and the shrink rate for both a MOSFET's physical gate length and effective gate oxide thickness (EOT) is defined as 65% per generation. The power supply voltage is assumed to reduce to 90% per generation.

To continue a 70% performance improvement rate for every generation, the required parasitic capacitance reduction from the CV/I equation is 58% per generation, where the MOSFET drive currents are kept constant even when the power supply voltage is reduced to 90% , and gate width is scaled 74% per generation. The components of the parasitic capacitance are shown schematically in Figure 3. Figure 4 shows results of the parasitic capacitance in each generation based on the assumption of simple scaling. In this figure Ctot is simply calculated as follows

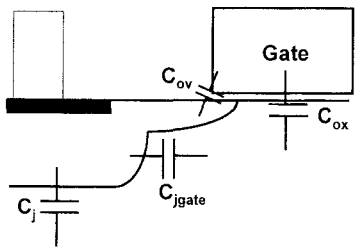


Fig.3 Components of Parasitic Capacitance

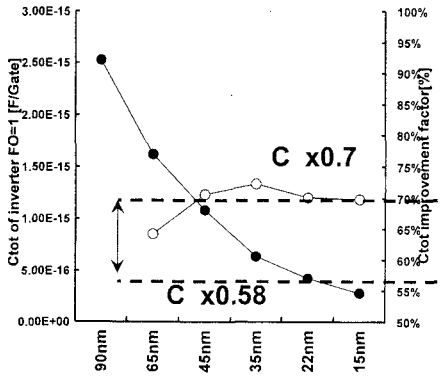


Fig.4 Scaling of Parasitic Capacitance

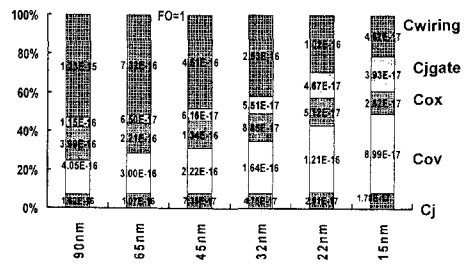


Fig.5 Effects of the parasitic capacitance for monotonous scaling

$$Ctot = Cov + Cox + Cj + Cjgate + Cint \quad (1)$$

where

Cov = constant

Cj $\propto$ W * (X - SW)

W = Channel Length * 0.74 / generation

X = Distance between gate edge to active area edge * 0.74 / generation.

Cj and Cjgate are varied by modification of the channel impurity concentration in each generation. There is a mismatch between the target and the calculated reduction factor of approximately 70% per generation. To analyze this mismatch, the capacity trend of each component is investigated, as shown in Figure 5. In this calculation, the wiring capacitance (Cwiring) is assumed to shrink 74% per generation. It is clear that the overlap capacitance between the gate electrode and drain junction (Cov) becomes dominant below the 45nm node. It is also shown that the share of capacitance between the gate electrode and deep junction (Cjgate) is increasing gradually, but the junction to substrate

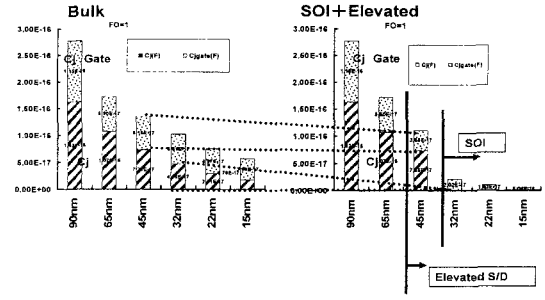


Fig.6 Effects of SOI and Elevated S/D for parasitic Capacitance

capacitance (Cj) remains at the same level.

From the technology point of view, adopting an elevated S/D process or SOI technology is very effective [3] to reduce Cj and Cjgate. Figure 6 shows the improvement of the parasitic capacitance with elevated S/D process after the 45nm node and SOI technology after the 32nm node. The elevated S/D process decreases Cjgate by almost half compared to the original structure and SOI technology decreases Cj to almost zero. But even when we adopt both technologies mentioned above, the average scaling factor is about 65%, which is still higher than target. If we cannot identify any technologies which reduce parasitic capacitance more, we must increase MOSFET drive current at a greater rate than the initial assumption

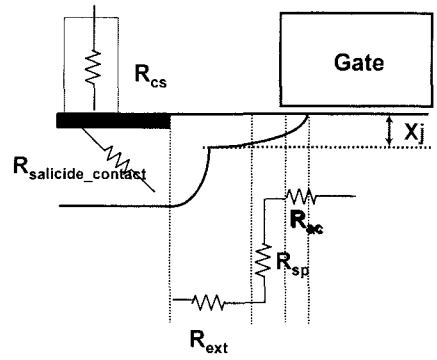


Fig.7 The Parasitic Resistance Elements

2-3.Decreasing the MOSFET parasitic resistance

In discussing improvement of MOSFET performance, the parasitic resistance cannot be ignored beginning with the 90nm node. Figure 7 schematically shows the elements of the MOSFET parasitic resistance, where Rac is the accumulation resistance, Rsp is the spreading resistance, Rext is the extension resistance and Rcs is the contact plug resistance. The calculated trend of the parasitic resistance elements for NMOSFET is shown in Figure 8. Assumptions in this calculation are that the gate width (Wg) is kept 1um, and other physical parameters are decreased to 74% per generation. This calculation is calibrated using measurement data from the 90nm node. To ignore this resistance, it must be sufficiently low compared with the MOSFET turn-on resistance. The target is 15% of NMOSFET turn-on resistance in each generation, and is shown by the thick

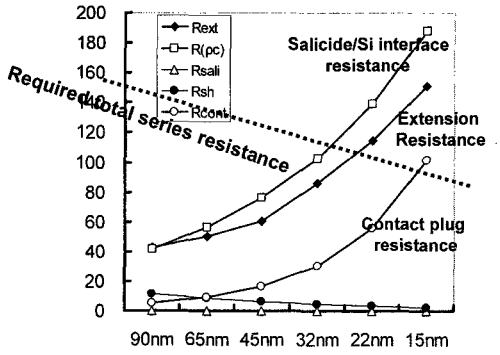


Fig.8 The trend of parasitic resistance

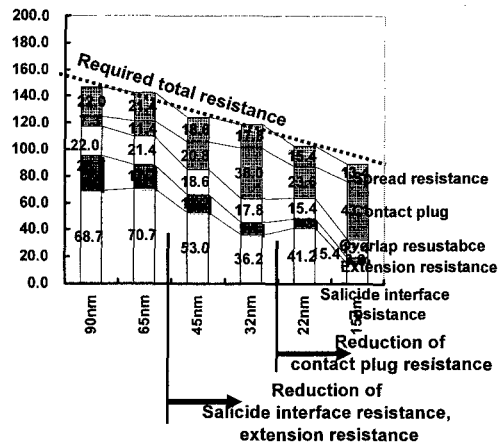


Fig.9 Budget of the parasitic resistance elements to keep performance trend

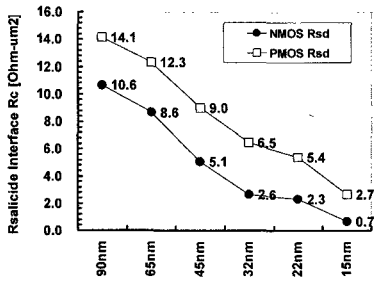


Fig.10 Trend of Salicide to Si interface resistance for NMOS and PMOS

dotted line in the same figure. It is found that the total parasitic resistance will exceed the target value even at the 45nm node. It is also clear salicide-to-Si interface resistance, contact plug resistance and extension resistance are increasing at each generation shift, and become dominant factors after the 45nm node and beyond.

To meet the target value defined above, the calculated budget for the parasitic resistance elements is shown in Figure 9. To keep the resistance within budget, new technology should be introduced to reduce salicide-to-Si interface resistance and extension resistance from the 45nm node. In addition, contact plug resistance should be reduced from the 22nm node. Figure 10 summarizes the required salicide-to-Si interface resistance for both NMOSFET and PMOSFET. In this figure the vertical axis is normalized by the unit interface area. To keep parasitic resistance less than 15% of the MOSFET turn-on resistance, salicide-to-Si interface resistance should be reduced by almost half every two generations. Figure 11 shows the trend of contact plug resistance. W plug can be applicable until the 32nm node by redacting the thickness of pre-Metal Dielectric (PMD) material.

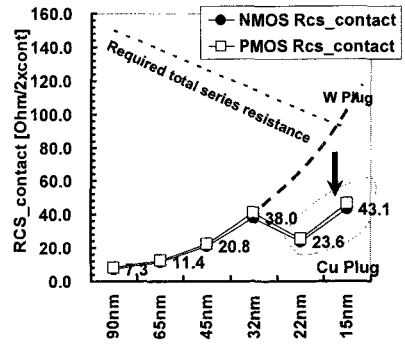


Fig.11 The trend of contact resistance

By using NG's equation [4], the extension resistance of source and drain regions are expressed as

$$R_{sp} = R_0 + C\int_{x1}^{x2} \frac{e^{-CKx}dx}{x} \qquad (2)$$

$$R_{ac} = F(K,N) \qquad (3)$$

The required abruptness of the junction profile of the extension regions under the assumption of the scaling of junction depth will follow 85% per generation as shown in Figure 12 [5]. In this figure, the requiremed improvement of the abruptness should be 50 to 60% per generation. To achieve this abruptness, innovative improvement of the thermal treatments to activate impurity will be the critical challenge below 45nm node.

2-4.Improvement of MOSFET drivability

If parasitic capacitance or resistance is successfully reduced to the target values, the MOSFET saturation current should be increased 10% per generation when

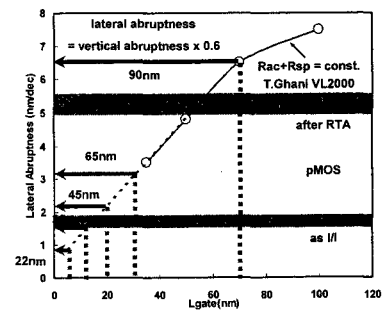


Fig.12 The requirement of abruptness of the extension region and required technology

power supply voltage is reduced 10% per generation. This means 30 to 50% in mobility enhancement is required after the 45nm technology, as shown in Figure 13. Stress engineering such as Strained Si technology will be very important from the 45nm node [6].

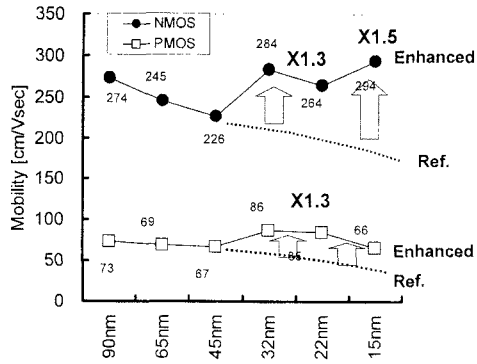


Fig.13 Target of Mobility Enhancement

3. OPTIMIZATION OF THE MULTILEVEL INTERCONNECT STRUCTURE

3-1.Model of a signal line on a chip

As an example to consider the technology requirement of multilevel interconnects, MOSFET parameters are summarized in Table 1, where the assumption of performance improvement is 70% per generation, and the reduction in power density is 50% per generation. The model to estimate performance of the signal lines is illustrated in Figure 14. The signal starts from point A, propagates through multilevel interconnect, and arrives at point B of the different mega cell in the same chip. Sakurai's model[7] is used to estimate wiring capacitance as shown Figure 15.

Node	Vdd (V)	Tinv (nm)	Lgate (nm)	Idn (uA/um)	Idp (uA/um)	Logic Depth	Gate#
90nm	1.00	2.10	50	846	355	18	1.00E+07
65nm	0.90	1.80	33	893	375	14	2.05E+07
45nm	0.80	1.60	21	937	393	11	4.35E+07
32nm	0.70	1.40	14	984	413	8	8.81E+07
22nm	0.70	1.20	9	1033	433	6	1.85E+08
15nm	0.60	1.10	6	1085	455	6	3.76E+08

↑ (2NAND)

Circuit design improvement is expressed in Logic Depth.

Table1 Assumption of Circuit Simulation

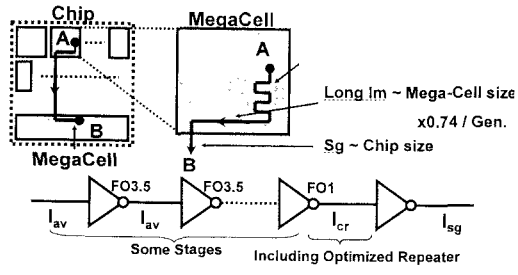
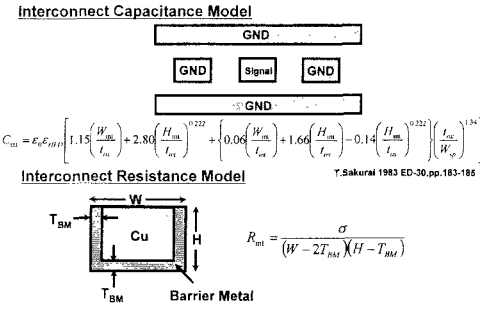


Fig.14 Model for Interconnection performance

3-2.Optimization of the total number of interconnect layers



Fig,15 Wire Capacitance molel

To minimize chip size and wiring delay, multilevel interconnect is commonly used in SoC chips. The necessary total number of signal wiring layers calculated as a function of device generation is shown in Figure 16. This calculation is performed using Davis' model [8], where the parameters of this model where calibrated to meet our existing SoC chips. The performance is improved monotonically as a function of metal layers for few metal layers because both chip size and wiring length can be shrunk with the number of layers in this region. It is shown that the performance improvement will saturate at a certain number of metal layers. This means that the additional metal layers provide no further advantage for shrinking the chip size or wiring length. The saturation point of the number of layers increases with alternating technology generations. In this figure, one additional layer is required for every two generations to optimize performance of the chip. This estimation is grovel wiring, so in the case of actual SoC, a few additional layers are used for local and power wiring and design flexibility.

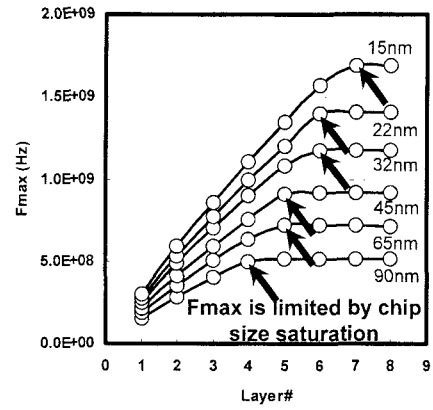


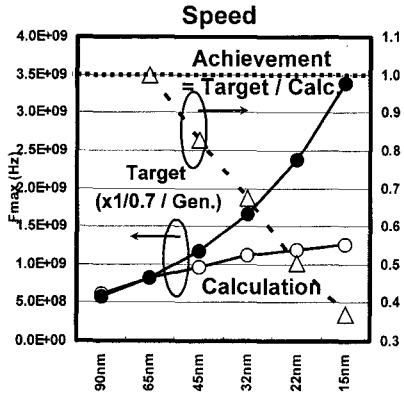
Fig.16 Wiring layers dependence of Performance

3-3.Improvement of performance and power consumption

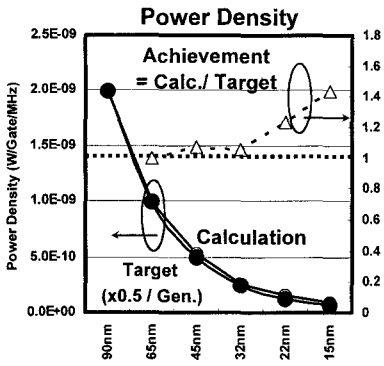
To understand the impact of scaling on performance and power consumption, the case of monotonically scaling only physical dimensions 74% per generation is calculated as shown in Figure 17, where (a) and (b) are a performance and a power consumption trend, respectively. It is clear that very little performance gain is observed with only just physical scaling.

(Received October 10, 2003; Accepted March 20, 2004)

To understand the key technology elements to enhance performance and reduce power density, sensitivity analysis of the interconnect elements was performed. The elements used in this analyze were relative dielectric constant (k) of the intermediate dielectrics (ILD), metal resistance, physical parameters of the interconnect, and thickness of the barrier metal of the wiring. Figures 18-1 to 18-5 show the results of the sensitivity analysis. Key parameters to improve performance are lowering k value, lowering sheet resistance, and thinning barrier metal thickness. It is also found that there is an optimum point in metal width because narrowing the wire increases wiring resistance , and narrowing the space increases coupling capacitance. Thicker metal improves performance, but too thick metal increases coupling capacitance and reduces the performance improvement.



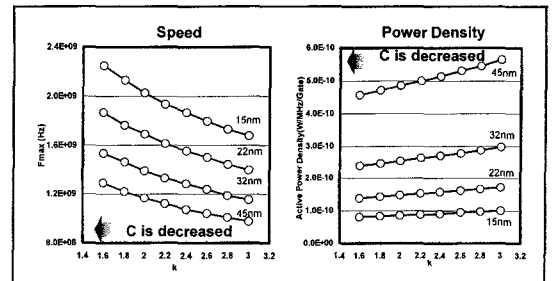
(a)



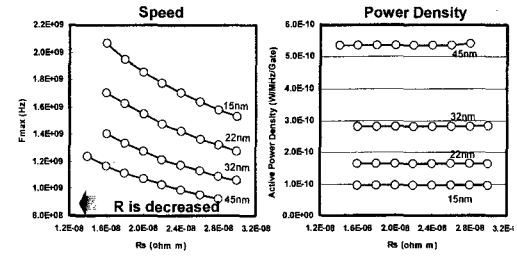
(b)

Fig.17 Impact of Scaling for performance and Power Consumption

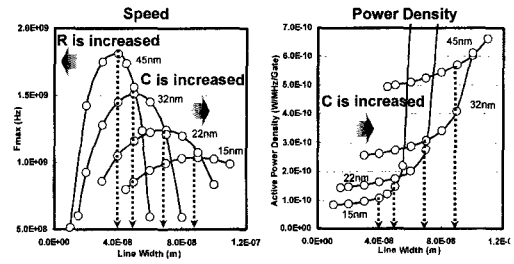
Because the k value is most sensitive for performance, calculation of the target performance improvement by just lowering k value was performed, and is shown in Figure 19 with a comparison of the case k values from ITRS 2002 [2]. In this case, the k value trend is separate from the ITRS value and is unrealistic. So, it is clarified that lowering wiring resistance is one of the keys to achieve the target. Table 2 summarizes realistic values of the k and physical dimensions to calculate the target wiring resistance. Figure 20 shows the performance trend using the interconnect elements parameters in Table 2. There still exist a difference between target and calculated value. To eliminate this gap, – the result of



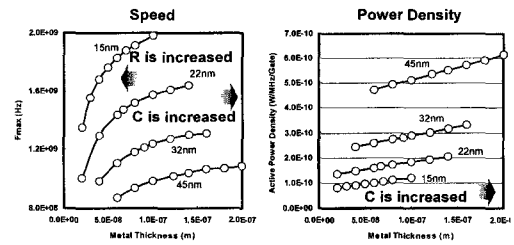
(1) ILD k value



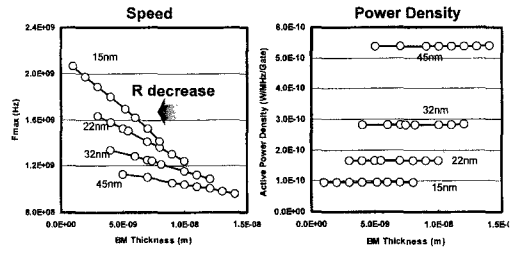
(2) Metal Resistance



(3) Wiring Line width



(4) metal thickness



(5) BM thickness

Fig.18 Sensitivity analysis of Interconnection parameters

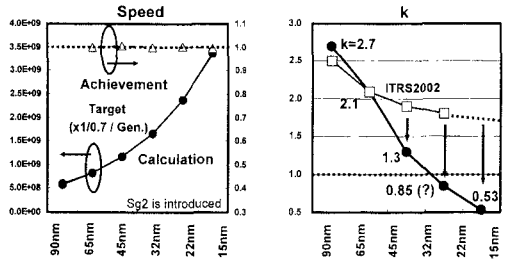


Fig.19 reduction of k value to achieve performance targets

Im

Node	L/S(nm)	T Metal(nm)	T ILD(nm)	keff	T BM(nm)
45nm	70 / 70	125	125	2.4	4.0
32nm	52 / 52	100	100	2.1	3.0
22nm	38 / 38	76	76	1.9	2.2
15nm	28 / 28	56	56	1.8	1.6

Sg

Node	Sg k
45nm	2.8
32nm	2.4
22nm	2.1
15nm	1.9

Sg2

L / S = 280nm / 280nm

T_{Metal} = 500nm

k_{eff} = 3.5

Table 2 Realistic Target of Interconnection parameters

target optimization of wiring resistance is shown in Figure 21. In this figure other parameters are the same as previous figure. It is found that to achieve the performance improvement for the trend, both lowering k and wiring resistance are the essential from the 45nm node and beyond.

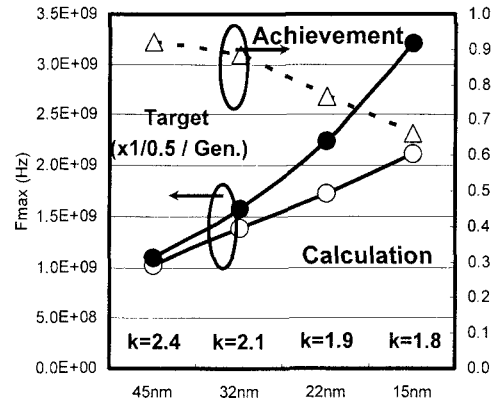


Fig.20 performance calculation from Table2

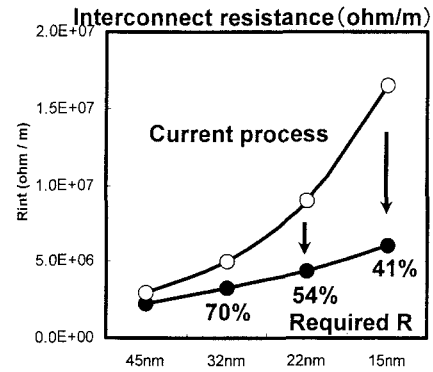


Fig.21 Resistance reduction to catch up Target

4. SUMMARY

The key technologies of the performance improvement at the 45nm node and beyond were discussed for both MOSFET and interconnect. It is very important to enhance MOSFET performance, and managing both the parasitic capacitance and resistance is essential. Lowering both k and wiring resistance are the challenges for interconnect technology.

5. REFERENCES

[1] National Technology Roadmap for Semiconductors, 1997

[2]International Technology Roadmap for Semiconductors, Update2002

[3] S.Narasimha et.al., "High Performance Sub-40nm CMOS Device on SOI for the 70nm Technology" IBM Microelectronics Reserch Center, NY, USA Technical Digest of IEDM 2001

[4] S.D.Kim et.al., "Detailed Modeling of Source/Drain Parasitic and Their Impact on MOSFET scaling" University of Electrical Engineering, University of California, LA, USA Abstract of International Workshop on Junction Technology, 2002

[5] NG kk et.al., "Analysis of the gate-voltage-dependent series resistance of MOSFETs" IEEE Trans. Electron Devices vol.ED-33, p.p.965-972

[6] K.Rim et.al., "Strained Si MOSFETs for High Performance CMOS Technology" IBM Watson Research Center, NY, USA Abstract of VLSI symposium 2001

[7] T.Sakurai and K.Tamaru, "Simple Formulas for Two-and Three-Dimensional Capacitance" IEEE Trans. Electron Devices, vol. ED-30 ,pp.183-185,1983

[8] J A. Davis et.al., "A Stochastic Wire-Length Distribution for Gigascale Integration (GSI) –Part I: Derivation and Validation" IEEE Trans. Electron Devices, vol. ED-45 ,pp.580-589,1998

(Received October 10, 2003; Accepted March 20, 2004)