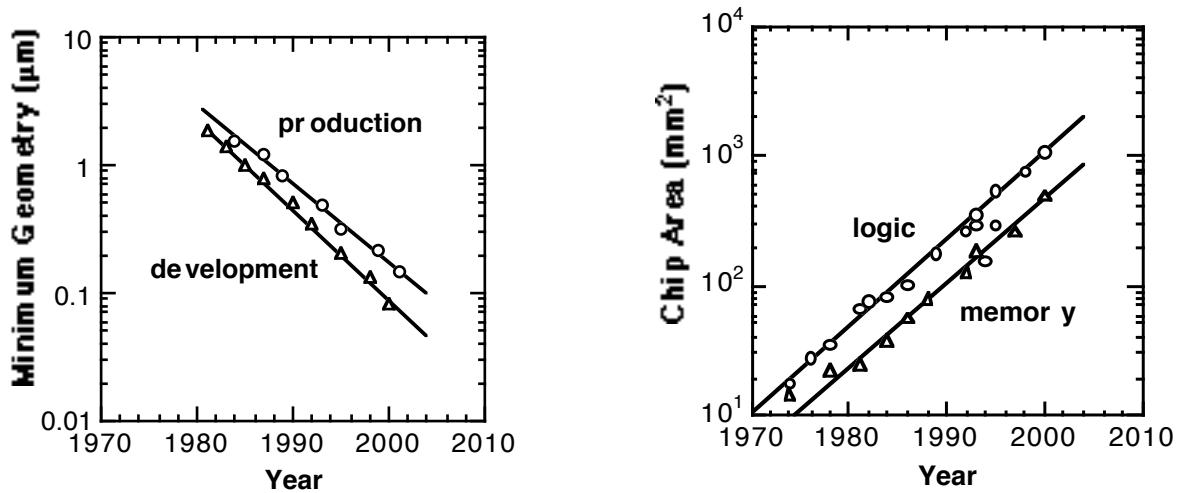# Scaling of Interconnections

Scaling of Minimum Feature size and Chip Area

Ref. A. Loke, PhD Thesis, Stanford Univ. 1999

Once the active devices and regions are fabricated they must be electrically connected to each other to make circuits. They must also be connected to the outside world through their inputs and outputs on bonding pads. Making these connections is the job of contacts, vias and interconnects. Separating the interconnects from each other is the job of dielectric layers. All of these components are part of the "metallization" or "backend" structure. The figure below is a schematic diagram showing these components in a typical integrated circuit structure. In recent times the relative importance of the backend structures has greatly increased, and will likely to continue as integrated circuit technology progresses.
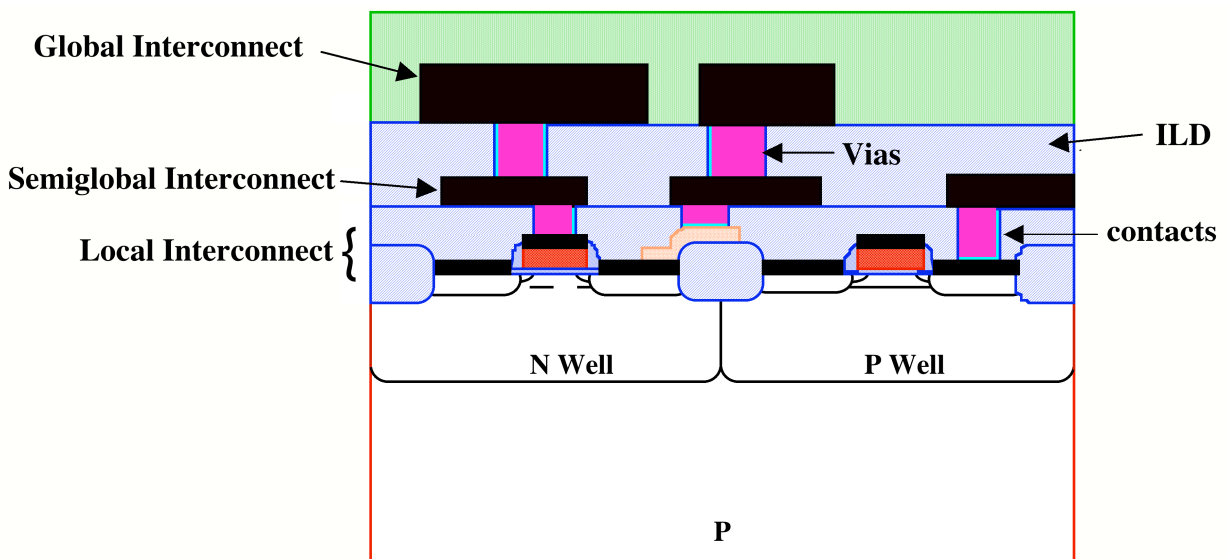
Fig. 2 Schematic cross-section of backend structure, showing interconnects, contacts and vias, separated by dielectric layers.

Interconnects can either be global or local. In general, local interconnects are the first, or lowest, level of interconnects. They usually connect gates, sources and drains in MOS technology, and emitters, bases, and collectors in bipolar technology. In MOS technology a local interconnect, polycrystalline silicon, also serves as the gate electrode material. Silicided gates and silicided source/drain regions also act as local interconnects. In addition, TiN, a by-product of a silicided gate process, can be used as a local interconnect, and W is sometimes used as well. Local interconnects can afford to have higher resistivities than global interconnects since they do not travel very long distances. But they must also be able to withstand higher processing temperatures. Global interconnects, mostly made of Al, are generally all of the interconnect levels above the local interconnect level. They often travel over large distances, between different devices and different parts of the circuit, and therefore are always low resistant metals. Local and global interconnects are shown in the multilevel schematic diagram in Figure 2 and 3.
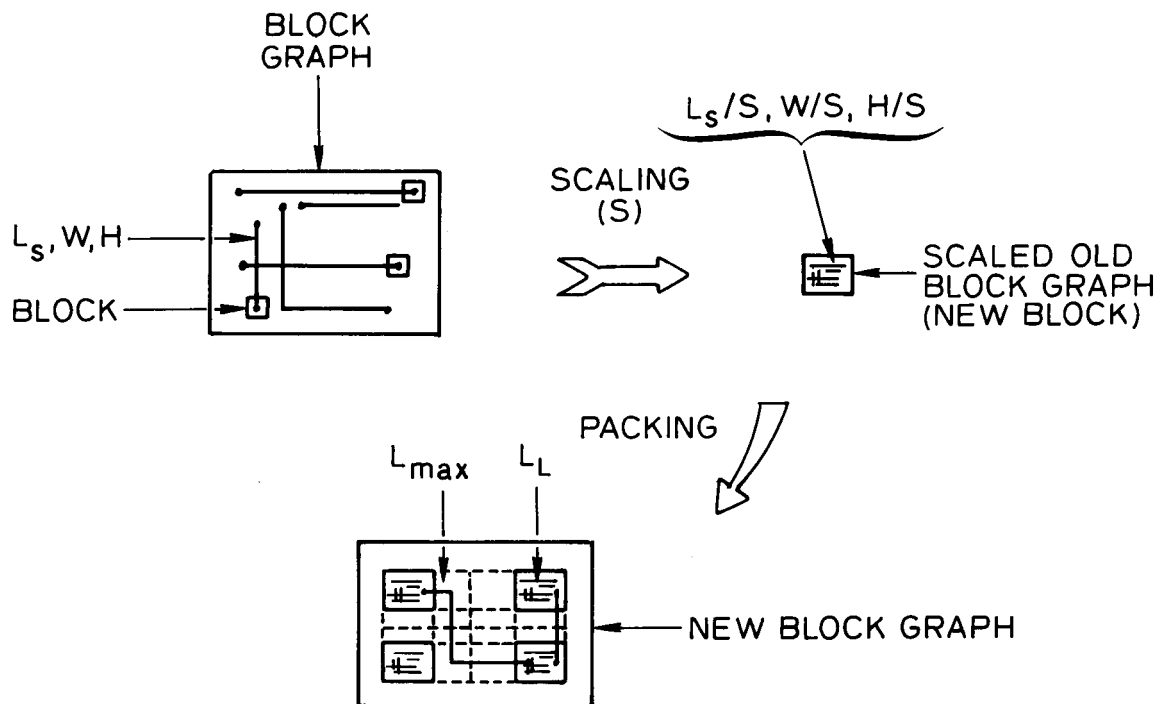

Fig. 3 Scaling of a chip and interconnections

In modern ICs as the complexity of interconnects is increasing an additional level, semiglobal interconnects between local and global has been introduced. The hierarchy of interconnects is as follows:

- Local interconnects – used for very short interconnects at the device level

- Semiglobal interconnects – Used to connect devices within a block
- Global interconnects – Used to connect long interconnects between the blocks, including power, ground and clocks.

As the device dimensions are scaled down so are that of the interconnects. The scaling trends of the height and width of the interconnects are shown below.

Technology node (μm)
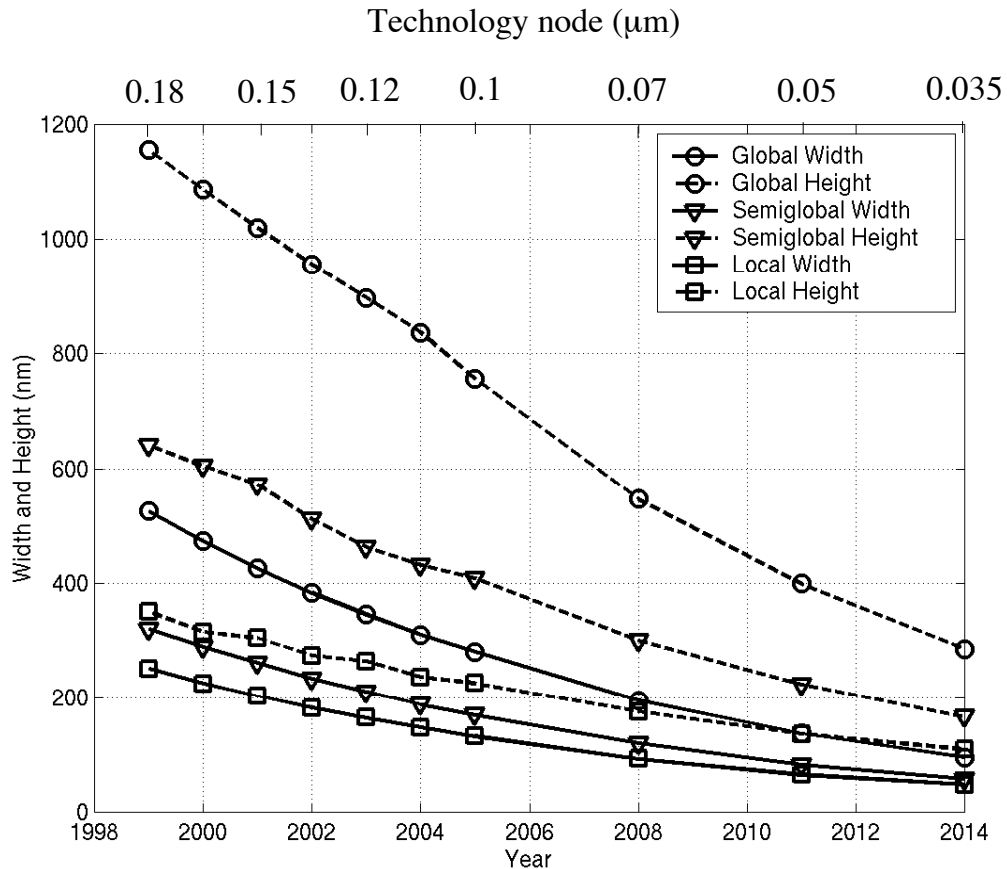
ITRS '99 dictated dimensions for local, semiglobal and global interconnects

Ohmic contacts connect an interconnect with active regions or devices in the silicon substrate. A high resistivity dielectric layer, usually silicon dioxide, separates the active regions from the first level global interconnect, and electrical contact is made between the interconnect and the active regions in the silicon through openings in that dielectric layer. At the same time in the processing, contact can be made between the first level global interconnect and the local interconnects, since they are also separated by that same dielectric layer. This is shown in Figure 2, with contacts from the lowest level global interconnect to: 1. a silicided source region in the silicon: 2. a TiN local interconnect: and 3. a silicided gate. Connections between two levels of global interconnects are usually given a different name: vias. They are made through openings in the different levels of the intermetal dielectrics. These are also illustrated in the figure. Whether they are contacts or vias, the goal of them is to electrically connect

different levels of conducting layers, and to contribute as little as possible to the total electrical resistance.

Separating interconnects from each other and from the active areas and devices are dielectric materials. Those dielectric layers separating one global interconnect level from another are called intermetal dielectrics, or IMD. (Some call these interlevel dielectrics, or ILD.) Vias connect interconnects through these layers. The layer separating global interconnects from both the substrate and local interconnects, through which the contacts are made, are given a variety of names: first level dielectric, dielectric-1, poly-gate dielectric, pre-metal dielectric, etc. - anything but intermetal dielectric. We will usually use the term "first level dielectric". Examples of both kinds - intermetal dielectrics and first level dielectrics - are shown in Figure 2.

There are several reasons why metallization structures are getting much more attention than before. One is because of their increased importance as circuit elements decrease in size and circuit areas increase in size. Straightforward electrical analysis, following that of Saraswat and Mohammadi [ref. 1] and based on the simple structure in Figure 4, can demonstrate this. The line resistance of one of the interconnects is given by:

$$R = \rho \frac{L}{WH}$$

$$C_{OX} = K_{OX}\varepsilon_O \frac{WL}{X_{OX}}$$
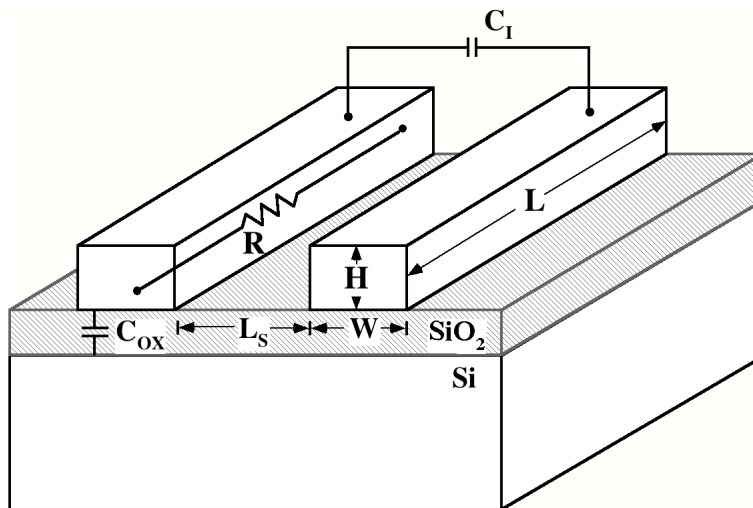
$$C_I = K_{OX}\varepsilon_O \frac{HL}{L_S}$$

Fig 4. Interconnect structure for RC analysis. The two lines on top are the metal interconnects of dimensions W, L, and H, sitting on a SiO$_2$ layer. There is also SiO$_2$ between the metal lines.

$$R = \rho \frac{L}{WH} \qquad\qquad (1)$$

where ρ is the interconnect's resistivity, and L, W, and H are the interconnect's length, width and height, respectively. The line-to-substrate capacitance is:

$$C_{ox} = K_{ox}\varepsilon_o \frac{WL}{X_{ox}} \qquad (2)$$

where X_{ox} and K_{ox} are the oxide thickness and dielectric constant, respectively, and ε_o is the permittivity of free space. The capacitance between the two interconnect lines is:

$$C_I = K_{ox}\varepsilon_o \frac{HL}{L_S} \qquad (3)$$

where L_S is the distance between the two lines.   For this simple structure, the total line capacitance can be approximated [2] by:

$$C = K_I(C_{ox} + C_I) \qquad (4)$$

where K_I takes into account the fringing fields to the substrate and is approximately equal to 2. (In a multi-level interconnect structure a factor of 2 is also used for the total capacitance, in that case to account for the total capacitance above and below and on both sides of an interconnect line.) The time delay of this structure can be calculated by treating it as a distributed-parameter transmission line. It can be shown that for such a system, the time delay, τ_L, is approximately equal to 0.89 RC.   Substituting in Equations 1-4 leads to:

$$\tau_L = 0.89 \cdot K_I K_{ox}\varepsilon_o\rho \frac{L^2}{WH}\left(\frac{W}{X_{ox}} + \frac{H}{L_S}\right) \qquad (5)$$
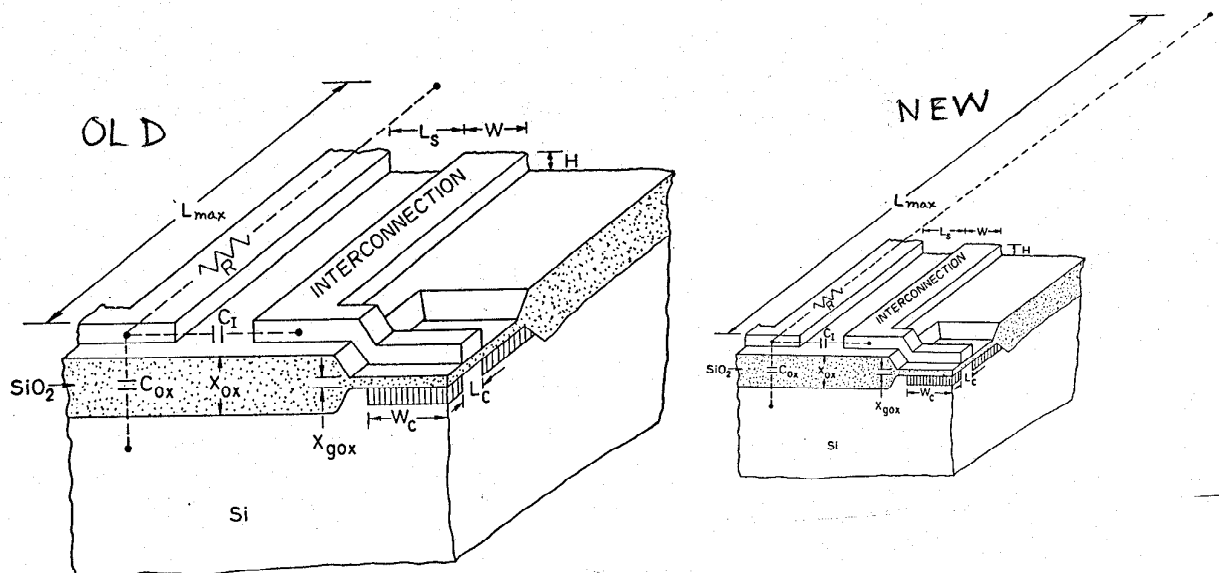

Fig. 5. Scaling of global (long distance) interconnections

5

Now we must consider what happens as the technology progresses in regards to the dimensions of the structure. First, L_S and W are usually close to the smallest possible dimensions possible, dictated by the lithography and etching capabilities, and designated by λ. As the technology progresses, λ gets smaller. Currently λ is 0.25 to 0.35 microns. Both X_{ox} and H shrink as λ shrinks in this ideal scaling scheme, keeping the aspect ratio, H/W, constant. It is assumed in our analysis that X_{ox} and H are equal to λ. (In the original analysis by Saraswat and Mohammadi, it was assumed that X_{ox} and H equaled 0.35λ and 0.25λ, respectively, but as we will discuss, the thicknesses have not decreased as quickly in recent years, especially for the global interconnects. We will still assume ideal, linear scaling here, but with a higher prefactor value of one.) Plugging these into Equation 5, and with K_I equaling 2, Equation 5 becomes:

$$\tau_L = 3.56 \cdot K_{ox} \varepsilon_o \rho \frac{L^2}{\lambda^2} \qquad (6)$$

We have said that λ gets smaller as the technology progresses. But what about L, the interconnect length? For local interconnects, L usually also shrinks as λ shrinks. Therefore, the net result is that the time delay for local interconnects stays approximately constant according to this scaling scheme.   However, for global interconnects, it is a different story. As the technology progresses, the length of the global interconnects usually increases rather than shrinks as shown in Fig. 5. This is because the chip area of each new technology generation usually keeps increasing as shown in Fig. 1, forcing the global interconnects to increase in length to connect up all the areas of devices. The average length of the longest global interconnect in a circuit can be approximated by [ref. 1]:

$$L_{max} = \frac{\sqrt{A}}{2} \qquad (7)$$

where A is the chip area. Equation 7 can be used to calculate the scaling factor S_c with the use of the data in Fig. 1. Inserting this into Equation 6 for global interconnects we get:

$$\tau_L = 0.89 \cdot K_{ox} \varepsilon_o \rho \frac{A}{\lambda^2} \qquad (8)$$

. As technology progresses, λ shrinks and A usually gets larger. By Equation 8 both of these lead to an increase in the time delay in global interconnects. Even if A stays constant, the time delay will increase due to λ shrinking. Shown in Figure 6 is Equation 8 plotted as a function of chip area for different values of λ and different interconnect materials, updated from the original figure in Ref. 1 to use the new expressions for X_{ox} and H and to include the $\lambda = 0.25$ micron graph. The interconnect delay is based on the longest global interconnect using Equation 7.   Also plotted is a typical gate delay, τ_G, for that minimum feature size, which is dependent only upon λ

and not A and thus decreases as the minimum feature size decreases.   One can see that for large λ's, the gate delay dominates over the Al global interconnect delay. (And one can also see why one would not want to use polysilicon as a global interconnect.) But for smaller critical dimensions and larger chip areas, the global interconnect delay can have a large impact on the circuit performance.

Table 1. Scaling of local interconnects

Interconnection Parameter	Scaling Factor* (S)
Interconnection dimensions L, H, W, L_s, X_{ox}	1/S
Line resistance $R = \rho L/WH$	S
Line capacitance $C_{ox} = K_{ox}\, \varepsilon_0\, L_s W/X_F$	1/S
Interelectrode capacitance $C_I^{\iota} = K_{ox}\, \varepsilon_0\, L\ H/L_s$	1/S
Line response time 0.89 RC	1
Line voltage drop IR	1
Line current density	S

*Scaling factor S > 1.

Table 2. Scaling of global (long distance) interconnects

Interconnection Parameter	Scaling Factor* (S)
Long-distance interconnection dimensions H, W, L_s, X_F	$1/S$
Long-distance interconnection dimension L_{max}†	S_c
Line resistance $R = \rho L_{max}/WH$	$S^2 S_c$
Line capacitance $C_{ox} = K_{ox} \varepsilon_o L_{max} W/X_{ox}$	S_c
Interelectrode capacitance $C_I = K_{ox} \varepsilon_o L_{max} H/L_s$	S_c
Line response time $0.89\ RC$	$S^2 S_c^2$
Line voltage drop IR	SS_c
Line current density	S

Today, λ is commonly 0.25 to 0.35 microns, chip areas are typically 200-350 mm^2 (corresponding to a maximum global interconnect length of 7-9 mm) and gate delays as measured in ring oscillators are generally about 30-70 picoseconds. The global interconnect delay by this analysis is about 2-4 nanoseconds, or about 60 times that of the gate delay. This does not mean that the total interconnect delay of a circuit will be 60 times that of the total gate delay. Remember, the interconnect delay we calculated is for the longest line in the circuit. The specific circuit layout and the distribution of line lengths will determine the net interconnect delay of the circuit.   What is important are the trends that these and other analyses predict: while the gate delays will keep getting smaller, the local interconnect delays will stay about the same, and the global interconnect delays will get larger and have a larger impact on the circuit performance.

Each new technology generation represents a 0.7x reduction in feature size (going from 0.5 to 0.35 to 025 to 0.18 microns, for example). By Bohr's results, as well as Equation 6 above, this corresponds to a 2x increase in interconnect delay per generation for a given line length. As noted by Bohr and mentioned above, the metal line aspect ratio is not typically kept constant but is increased with each technology generation largely due to the fact that the metal line thickness has stayed rather

constant in recent times rather than shrinking, as mentioned above. (The aspect ratio for Intel has increased from about 0.4 to 1.4 over the last several generations.) This means that the interconnect delay does not increase quite as fast, and has lead to an actual increase in interconnect delay per generation of about 1.3x for recent Intel logic technologies - still a significant increase. While the interconnect delay is increasing with each generation, Bohr found that the gate delay in current technologies is reduced by about 0.7x per generation, corresponding to the 0.7x decrease in feature size. As a result, the delay due to the interconnects is becoming a larger percentage of the total circuit delay time.   Bohr [ref. 5] and others [ref. 4] estimate that in today's CMOS technology the delay due to the interconnects is currently about 33% of the total circuit delay or clock cycle, up from about 15-20% just a year or two previously. In the next technology generations, the interconnect delay could be more than 50% of the total circuit delay.
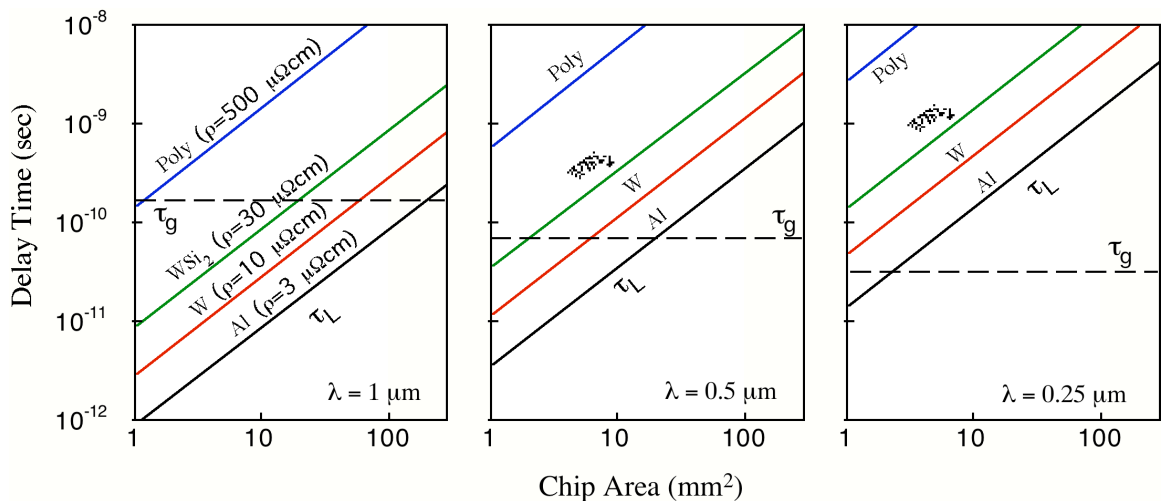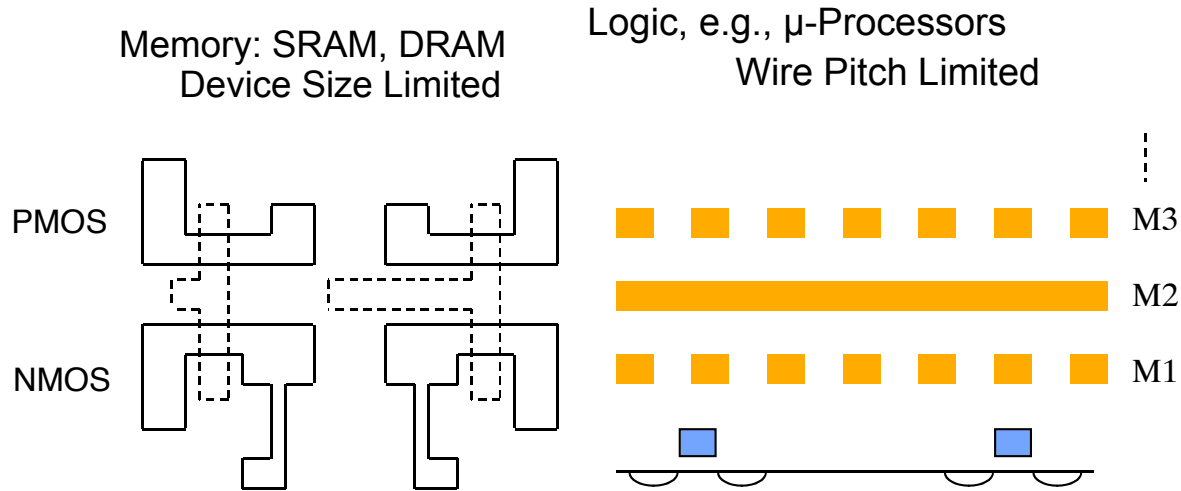

Fig. 6. Global interconnect and gate time delay versus chip area and minimum feature sizes for various interconnect materials, based on the analysis and scaling presented in the text. Interconnect delay corresponds to longest global interconnect in circuit. As the technology progresses, the interconnect delay becomes more important relative to the gate delay of a circuit. [after ref. 1]

Chip Size

Memory: SRAM, DRAM
Device Size Limited

Logic, e.g., µ-Processors
Wire Pitch Limited

PMOS

NMOS

M3

M2

M1

Memories have very regular and dense structure. They don't need too many interconnects. As a result the device packing density is very high. Logic chips are more irregular and are dominated by communication requirements. Device packing density is *not* very high. These chips generally have larger number of interconnects and thus need more levels of them.
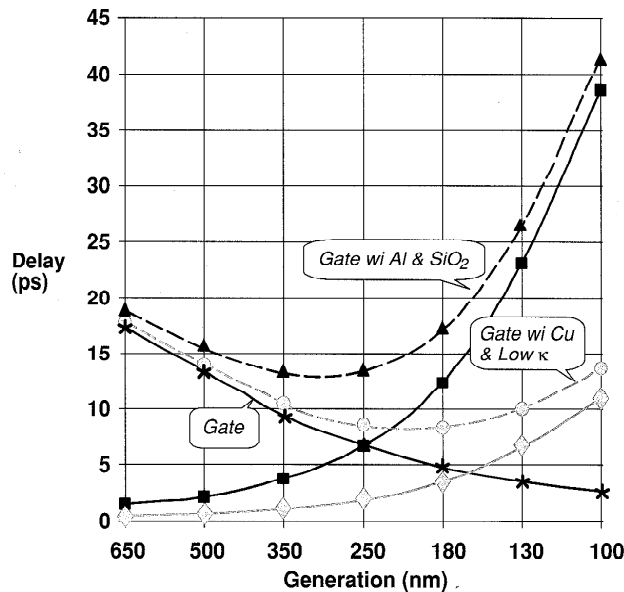

Fig. 7 Improvement in global interconnect time delay versus technology node for advanced materials (Cu and low k dielectrics)

The analysis so far has assumed that all interconnections have same area of cross section and follow same scaling riles. In a multilevel scheme that may not be true. Different layers may be optimized for different requirements. For example, for longer interconnections we may use bigger area of cross section and lower resistivity material to reduce the resistance. Smaller length interconnections could be placed in a layer of smaller area of cross section and higher resistivity material if that material has other desirable properties.
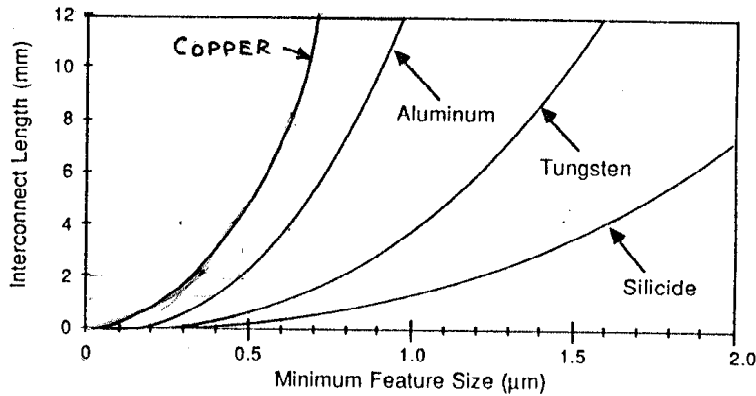
Fig. 8. Scaling long-distance interconnections using four commonly used materials resistivity materials in VLSI. The metal line width used is 1.5 times the minimum feature size. (Ref. Gardner, Meindl and Saraswat, IEEE Trans Electron Dev., Vol. ED-34, No. 3, pp. 633—643, March 1987.)

- Maximum length limited by $\tau_G = \tau_I$

- Longer wires must be fabricated with higher conductivity material – copper
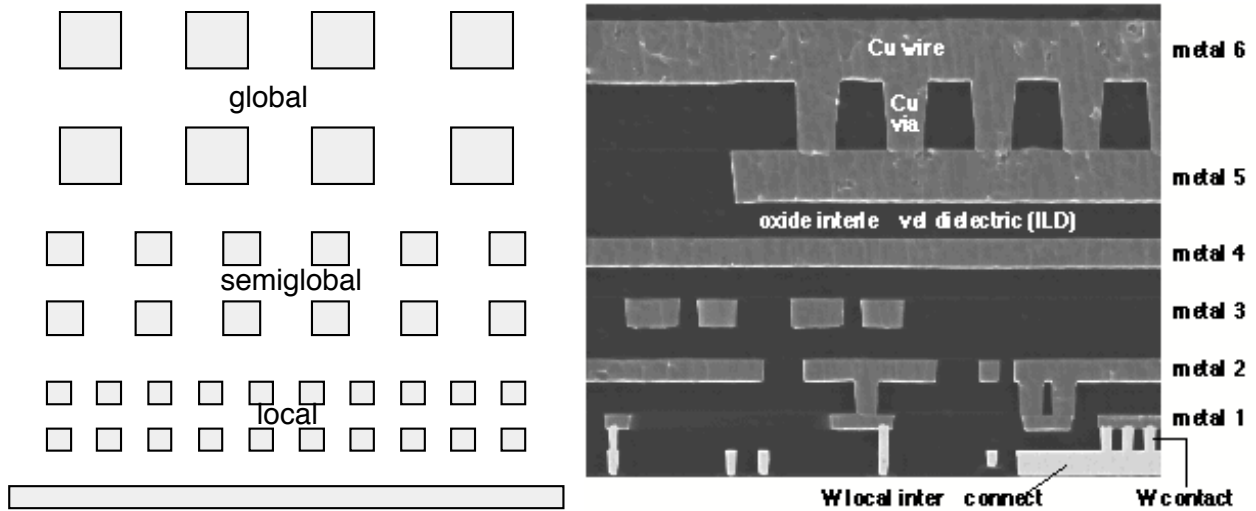
- Shorter (local) wires can be fabricated with lower conductivity material – silicides, tungsten

Fig. 9. (Left) A 3 tier interconnect architecture , (right) SEM of interconnect architecture with six levels of Cu wires/vias demonstrations by IBM.
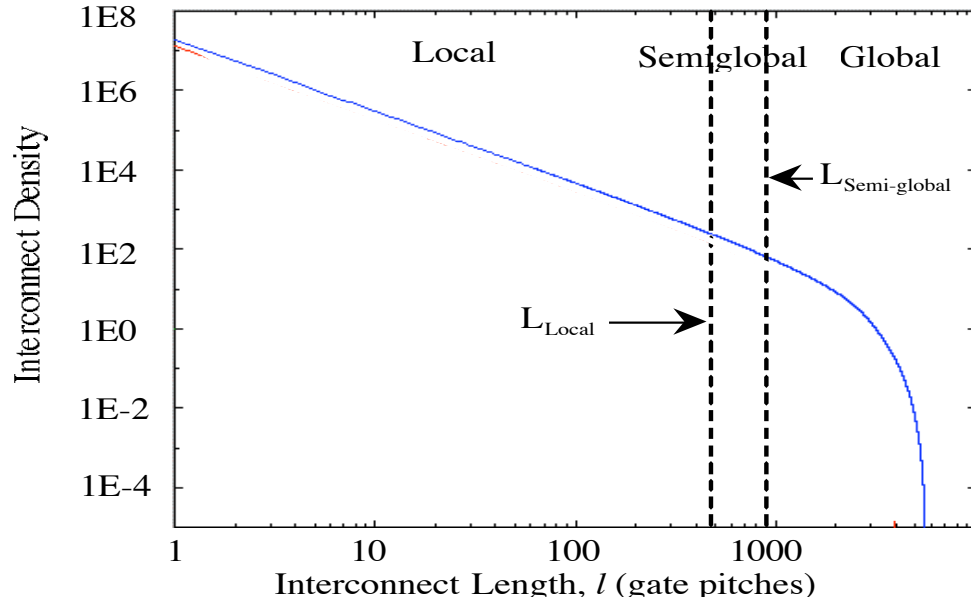
Fig. 10. Wire-length distributions (in terms of gate pitches) for a logic circuit with 180 million gates. Metal tiers determined by L_{Loc} and $L_{Semi-global}$ boundaries defined by maximum allowable delay.

- Longer wires for, clocks, power, ground and long distance communication can be placed in the global level.

- Mid length wires for communication within a block are placed in the semiglobal level.

- Very short wires where delay is primarily device governed are placed in the local level.

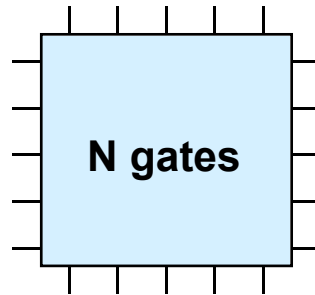**Use of Rent's Rule for Wire-length distributions calculations**

Fig. 11

Rent's Rule correlates the number of signal input and output (I/O) pins *T*, to the number of gates *N*, in a random logic network and is given by the following expression:

$$T = k\, N^P \tag{12}$$

where
k = avgerage I/O's per gate
P = Rent's exponent. It denote the degree of wiring complexity (with *p*=1 representing the most complex wiring network)
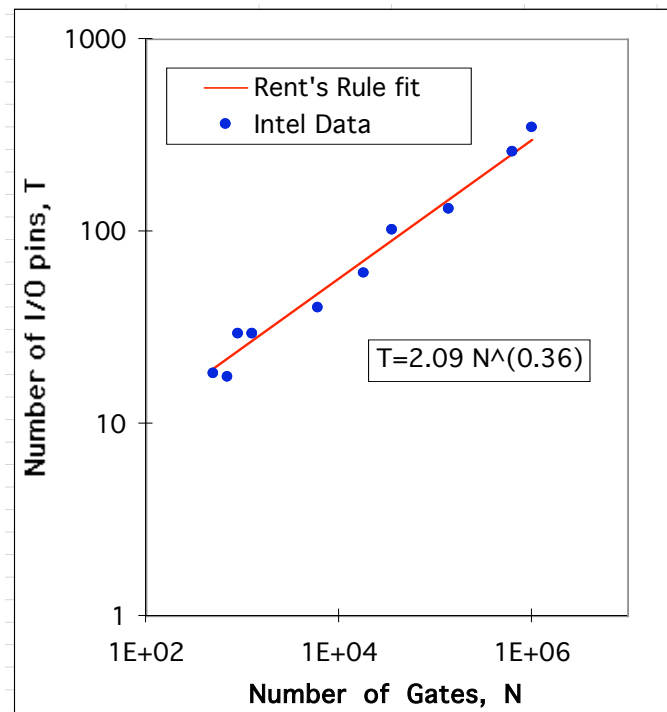

Fig. 12

The wire length distribution can be determined through the recursive application of Rent's Rule throughout an entire logic system. To determine all the shortest wires in a logic system, the recursive property of Rent's Rule is used, where the logic system is

divided into logic gates and Rent's Rule is applied to the interconnects between closest neighbor gates. This determines the number of interconnections between the closest logic gates. The longer wires are similarly determined by clustering the logic gates in growing numbers until the longest interconnects are found.

The wire length distribution can be described by *i(l)*, an Interconnect Density Function (i.d.f.), or by *I(l)*, the Cumulative Interconnect Distribution Function (c.i.d.f.) which gives the total number of interconnects that have length less than or equal to *l* (measured in gate pitches), and is defined as,

$$I(l) = \int_1^l i(x)\,dx \qquad (13)$$

where *x* is a variable of integration representing length and *l* is the length of the interconnect in gate pitches. The derivation of the wire-length distribution in an IC is based on Rent's Rule. To derive the wire-length distribution, *I(l)* of an integrated circuit, the latter is divided up into *N* logic gates, where *N* is related to the total number of transistors, N_t, in an integrated circuit by $N = N_t /$, where is a function of the average fan-in (*f.i.*) and fan-out (*f.o.*) in the system. The gate pitch is defined as the average separation between the logic gates and is equal to $\sqrt{A_c / N}$ where A_c is the area of the chip.
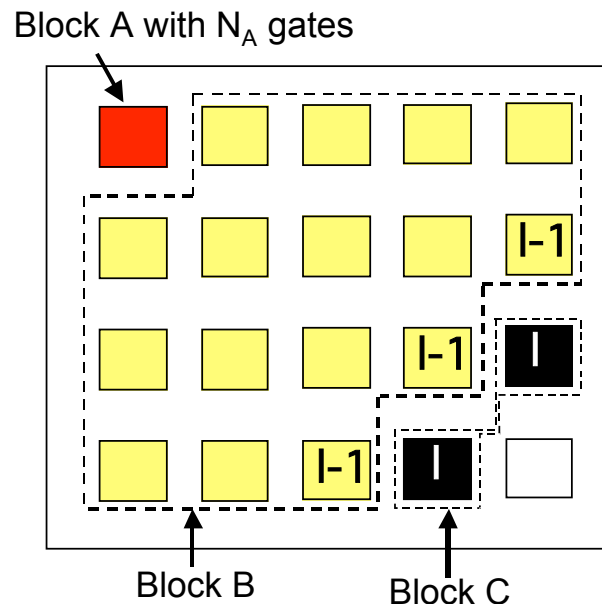

Fig. 13 Schematic view of logic blocks used for determining wire length distribution.

In order to derive the complete wire length distribution for a chip, the stochastic wire length distribution of a single gate must be calculated. The number of connections from the single logic gate in Block A to all other gates that are located at a distance of *l* gate pitches is determined using Rent's Rule. The gates shown in Fig. 13 are grouped into three distinct but adjacent blocks (A, B, and C), such that a closed single path can encircle one, two, or three of these blocks. The number of connections between Block A and Block C is calculated by conserving all I/O terminals for blocks, A, B, and C,

which states that terminals for blocks A, B, and C are either inter-block connections or external system connections.

Hence, applying the principle of conservation of I/O pins to this system of three logic blocks shown in Fig. 13 gives,

$$T_A + T_B + T_C = T_{A-to-C} + T_{A-to-B} + T_{B-to-C} + T_{ABC} \tag{14}$$

where T_A, T_B, and T_C are the number of I/Os for blocks A, B, and C respectively. T_{A-to-C}, T_{A-to-B}, and T_{B-to-C} are the numbers of I/Os between blocks A and C, blocks A and B, and between blocks B and C respectively. T_{ABC} represents the number of I/Os for the entire system comprising of all the three blocks. From conservation of I/Os, the number of I/Os between adjacent blocks A and B, and between adjacent blocks B and C can be expressed as,

$$T_{A-to-B} = T_A + T_B - T_{AB} \tag{15}$$

$$T_{B-to-C} = T_B + T_C - T_{BC} \tag{16}$$

Substituting (15) and (16) in (14) gives,

$$T_{A-to-C} = T_{AB} + T_{BC} - T_B - T_{ABC} \tag{17}$$

Now the number of I/O pins for any single block or a group of blocks can be calculated using Rent's Rule. If we assume that N_A, N_B, and N_C are the number of gates in blocks A, B, and C respectively, then it follows that,

$$T_B = k\left(N_B\right)^p \tag{18}$$

$$T_{AB} = k\left(N_A + N_B\right)^p \tag{19}$$

$$T_{BC} = k\left(N_B + N_C\right)^p \tag{20}$$

$$T_{ABC} = k\left(N_A + N_B + N_C\right)^p \tag{21}$$

where $N = N_A + N_B + N_C$. Substituting (18)-(21) in (17) gives,

$$T_{A-to-C} = k\left[\left(N_A + N_B\right)^p - \left(N_B\right)^p + \left(N_B + N_C\right)^p - \left(N_A + N_B + N_C\right)^p\right] \tag{22}$$

The number of interconnects between Block A and Block C (I_{A-to-C}) is determined using the relation,

$$I_{A-to-C} = \alpha\, k\left(T_{A-to-C}\right) \tag{23}$$

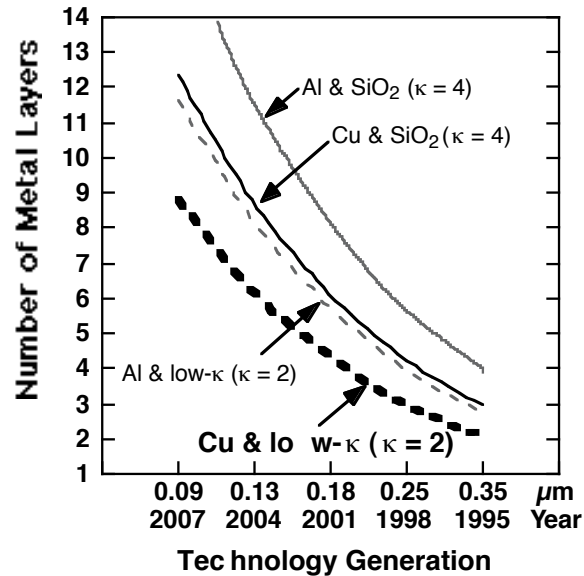Here α is related to the average fan-out (*f.o.*) by,

$$\alpha = \frac{f.o}{1 + f.o} \tag{24}$$

Equation (23) can be used to calculate the number of interconnects for each length *l* in the range from one gate pitch to $2\sqrt{N}$ gate pitches, to generate the complete stochastic wire-length distribution for the logic gate in Block A.
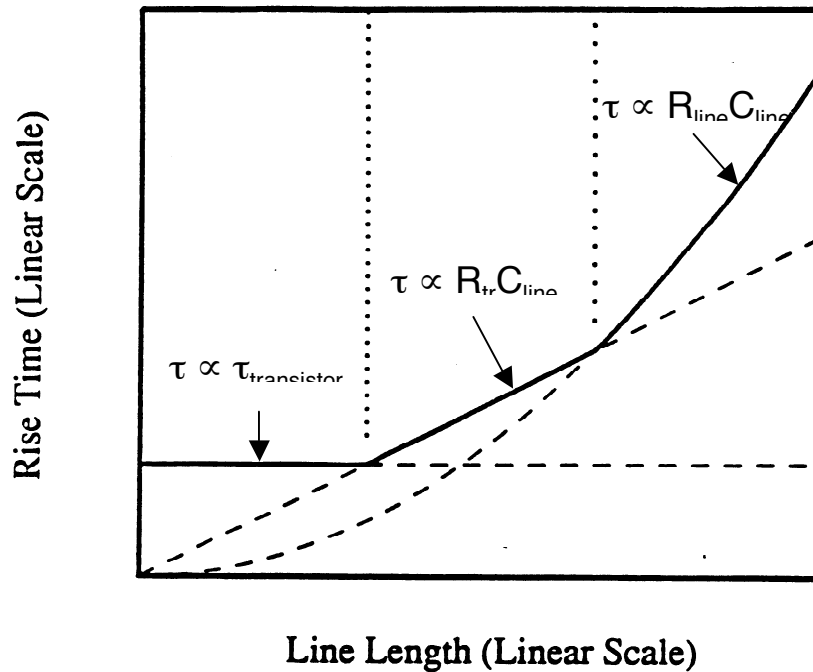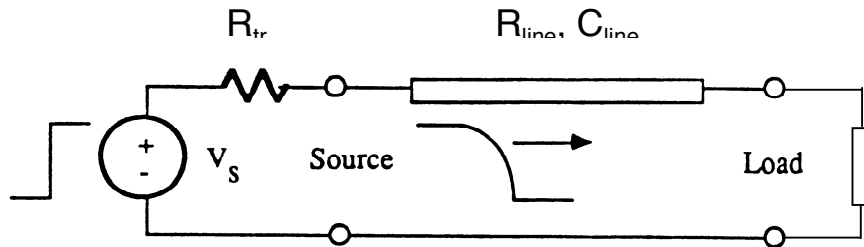
Interconnect Technology Roadmap

Metal Levels (Logic)

RC Delay (fF-Ohm/μm^2)
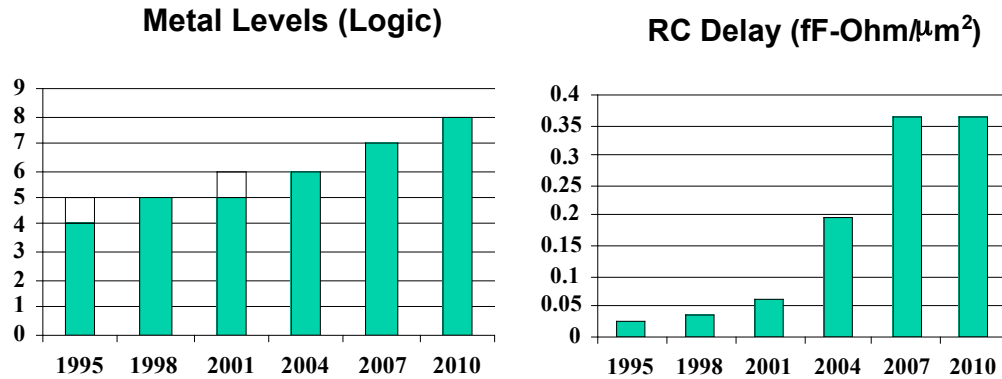
Number of metal layers for various interconnect materials options

Al & SiO$_2$ ($\kappa = 4$)

Cu & SiO$_2$ ($\kappa = 4$)

Al & low-κ ($\kappa = 2$)

Cu & lo w-κ ($\kappa = 2$)

Number of Metal Layers

| 0.09 | 0.13 | 0.18 | 0.25 | 0.35 | μm |
| 2007 | 2004 | 2001 | 1998 | 1995 | Year |

Tec hnology Generation

- By using lower resistivity material wire resistance can be minimized
- By using lower K dielectric capacitance can be minimized
- An interconnect made with Al and SiO$_2$ placed at the global level can be placed at the semiglobal level if made made with Cu and a low-K material
- Interconnect density is larger at the semiglobal level and thus smaller number of levels are needed.

Circuit Speed Dependence on Transistor and Interconnect

R_{tr} R_{line}, C_{line}

V_S Source   Load

Rise Time (Linear Scale)

$\tau \propto R_{line}C_{line}$

$\tau \propto R_{tr}C_{line}$

$\tau \propto \tau_{transistor}$

Line Length (Linear Scale)

1. <u>Small line length</u> - Transistor speed governs the circuit speed.

2. <u>Medium line length</u> - Transistor resistance and line capacitance govern the circuit speed

3. <u>Long line length</u> - Line resistance and line capacitance govern the circuit speed

Note that inductance has been ignored here. At higher frequencies it will become very important and would not be neglected.

Reduce resistivity of the conductor and dielectric constant of the dielectric material

Repeaters

A first option for reducing RC delays is to use better interconnect materials when they are available and appropriate. However, for very long wires the delay can be substantially larger than the gate delay. For instance, in a 0.25 µm technology the gate delay is about 25 psec, however, a 5 mm long aluminum wire with 0.25 µm x 0.25 µm cross section has a delay of 1 nsec. It is possible to reduce the propagation delay by introducing intermediate buffers, generally known as *repeaters,* in the interconnect line.
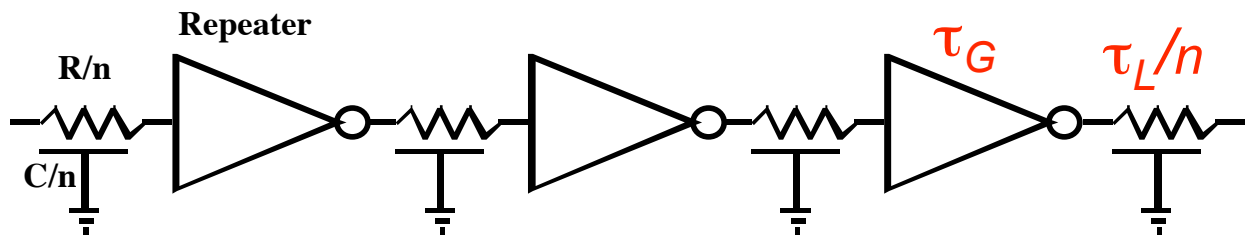
From Eq. 6 the delay of a line is

$$\tau_L = \frac{3.56 \cdot K_{ox} \varepsilon_o \rho}{\lambda^2} L^2 \tag{6}$$

By breaking a long interconnect line into n smaller lines the propagation delay of each line is reduced quadratically

$$\tau_{L/n} = \frac{3.56 \cdot K_{ox} \varepsilon_o \rho}{\lambda^2} \left(\frac{L}{n}\right)^2 \tag{9}$$

The total wire delay is thus

$$\left(\tau_{L/n} + \tau_G\right) n = \frac{3.56 \cdot K_{ox} \varepsilon_o \rho}{\lambda^2} \left(\frac{L^2}{n}\right) + n \tau_G \tag{10}$$

As long as the gate delay is small the total wire delay is reduced substantially. This gain results at the cost of increased chip area occupied and extra power consumed by the repeaters.

Effect of lowering temperature

INTRINSIC RESISTIVITY
OF PURE METALS
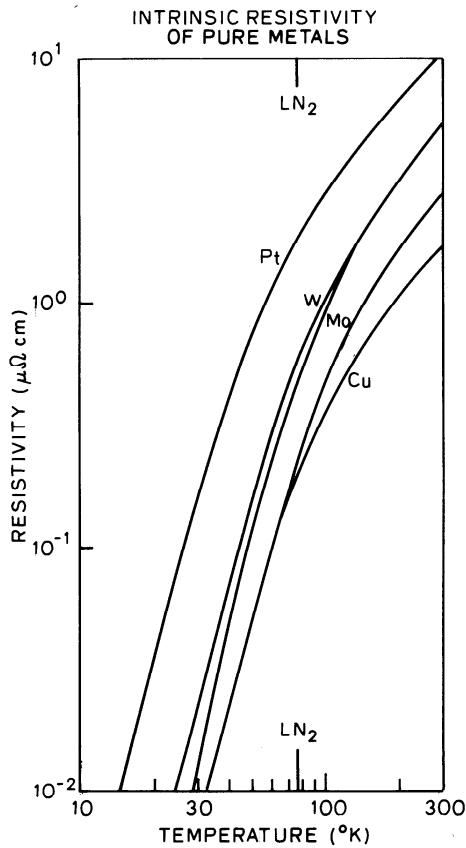
Cooling from room temperature to 77K reduces the resistivity by about an order of magnitude.
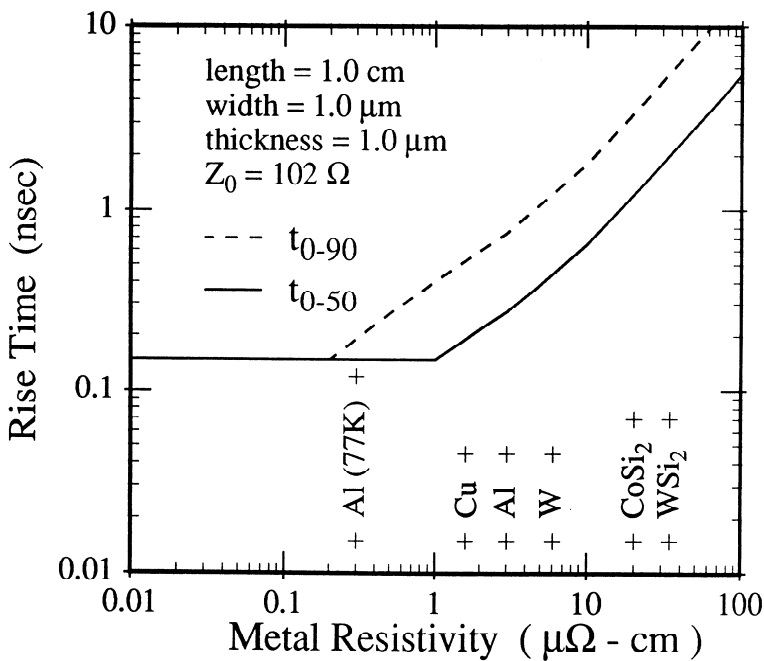
These simulations were done for an impedance matched line. At low values of line resistance the transistor impedance dominates the delay.

Ref:
Schrever, Nishi & Saraswat

Will superconductors really improve the circuit speed?
Cooling conventional conductors to 77°K may be sufficient.

Thin Film Effects

•Effect of Electron Scattering => Reduced mobility as dimensions decrease
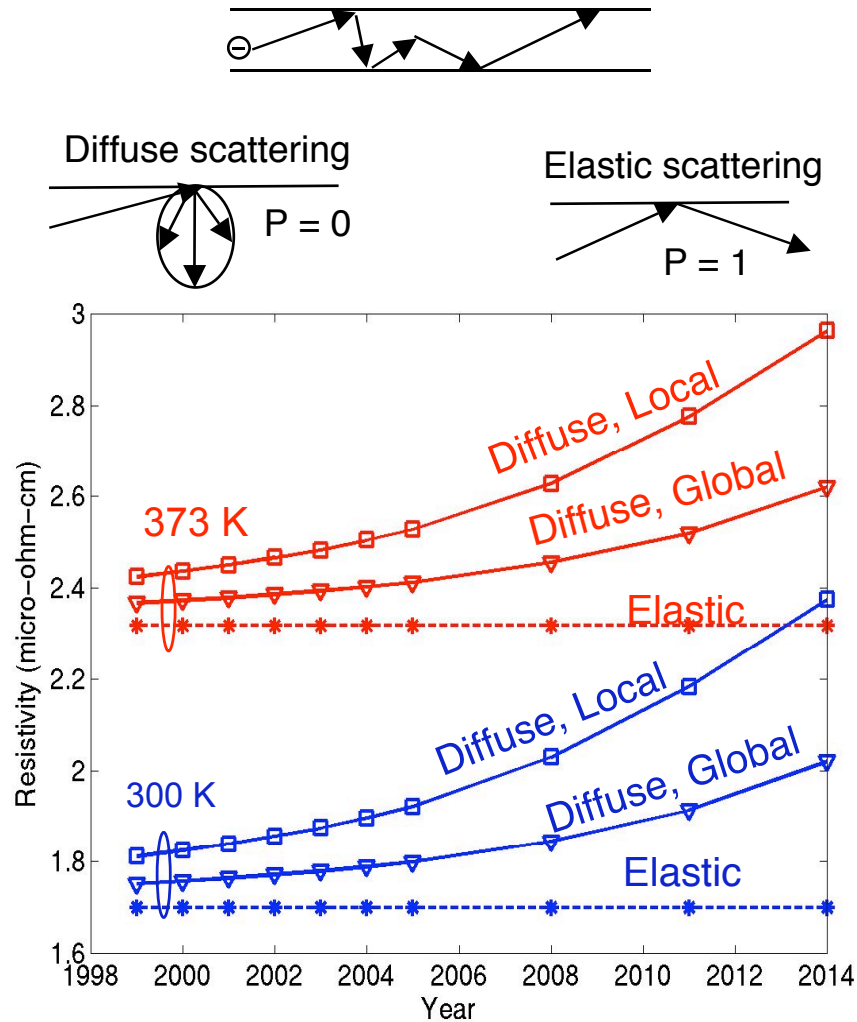
Resistivity of Cu wires for dimensions recommended by ITRS'99

- Surface electron scattering increases resistivity.
- Most metals today exhibit P ≈ 0.5
- Real chips operate at higher temperatures
- With scaling the wire cross section is decreasing. As a result the surface scattering is becoming more serious.
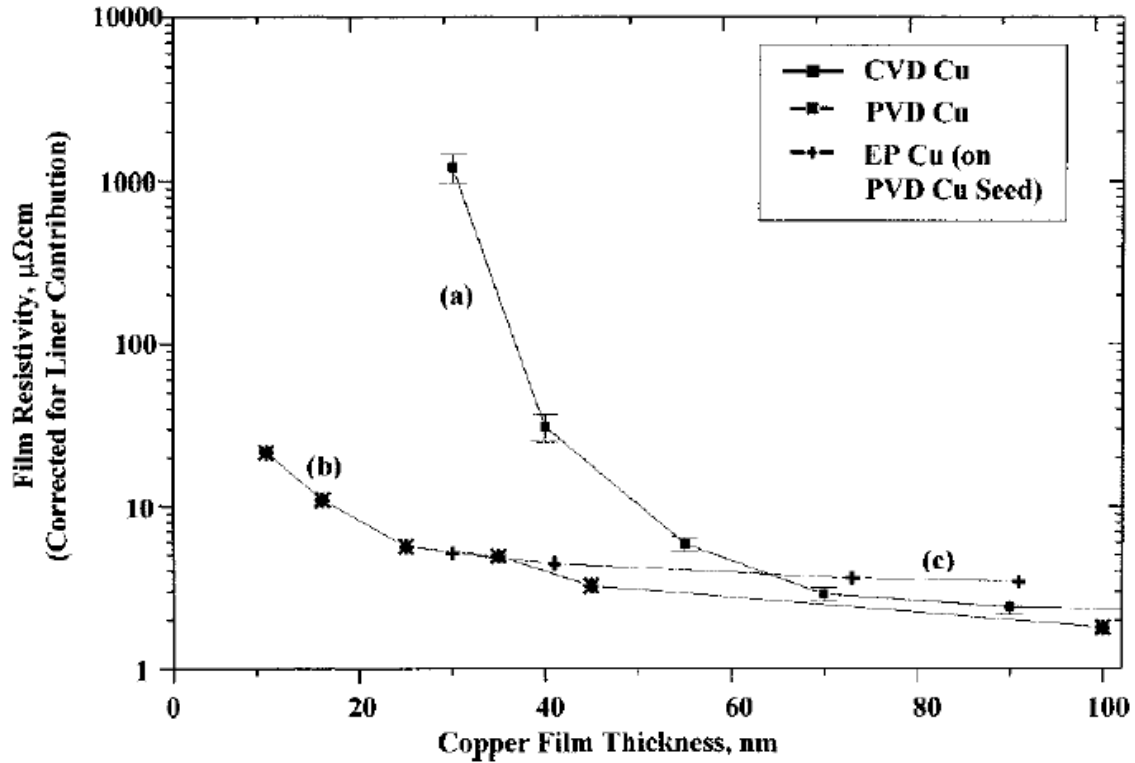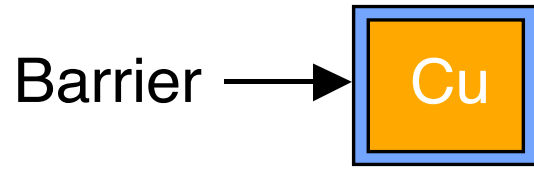
(Ref: Meindl et al. Proc IEEE, March 2001)

- Effect of Higher Frequencies - Carriers confined to outer skin increasing resistivity. These effects will become increasingly more important as the chips are operated in the GHz regime.
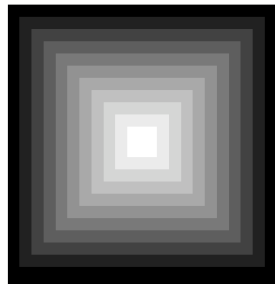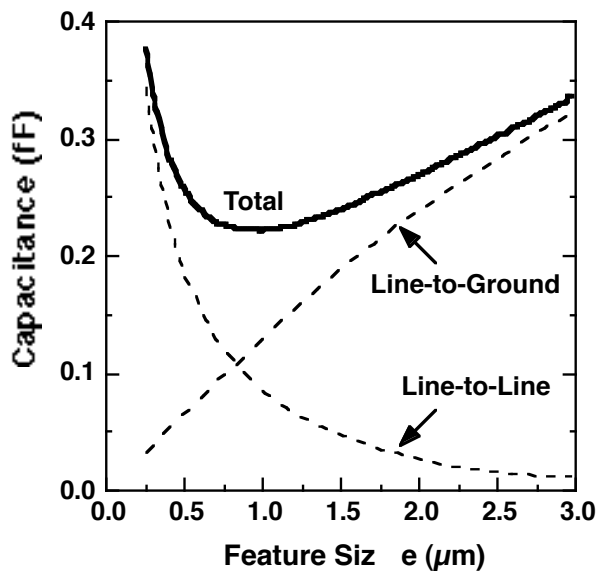

- Effect of Cu diffusion Barrier
    - Barriers have higher resistivity
    - Barriers can't be scaled below a minimum thickness

Barrier ⟶ Cu

Crosstalk due to capacitive coupling
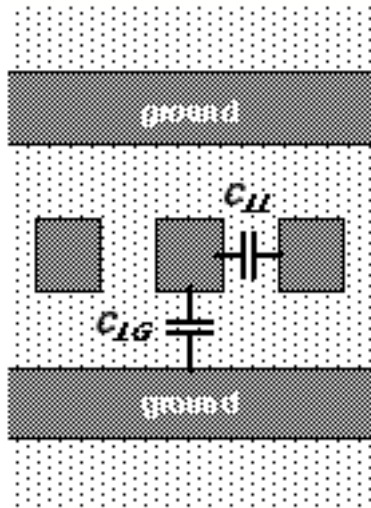
Higher Packing Density
⇓
Decreased Space
Between Interconnects
⇓
Higher crosstalk

Due to increased line-to-line capacitance there can be coupling of signal from one line to the adjacent line

Reduce dielectric constant of the dielectric material

Dielectric Materials	Dielectric Constant
undoped plasma SiO2	4.0-4.5
fluorinated SiO2	3.5
spin-on glasses (silsesquioxanes)	2.2-3.0
organic polymers (e.g., polyimides)	2.0-3.9
fluorinated amorphous carbon (a-C:F)	2.1-2.3

nanoporous dielectrics (e.g., xerogels)   1.2-2.2
Air   1

Requirements of the interconnection materials

- Low resistivity of conductors

- Low capacitance => low dielectric constant
    - Low RC delay
    - Low cross talk
    - Low power dissipation (CV^2f loss)

- Resistance to electromigration

- Ease of deposition of thin films of the material

- Ability to withstand the chemicals and high temperatures required in the fabrication process

- Ability to be thermally oxidized

- Good adhesion to other layers - low physical stress

- Stability of electrical contacts to other layers

- Ability to contact shallow junctions and provide low resistance

- Good MOS properties

- Ability to be defined into fine patterns - dry etching

Useful References:

1. K. C. Saraswat and F. Mohammadi, "Effect of scaling of interconnections on the time delay of VLSI circuits," IEEE Trans. Electron Dev., vol ED-29, p. 645, 1982.
2. R. L. M. Dang and N. Shigyo, "A two-dimensional simulation of LSI interconnect capacitance," IEEE Electron Device Lett., vol EDL-2, p. 196, 1981.
3. D. Gardner, J. D. Meindl and K. C. Saraswat, IEEE Trans Electron Dev., Vol. ED-34, No. 3, pp. 633--643, March 1987.
4. M. T. Bohr, "Interconnect scaling - the real limiter to high performance ULSI," IEDM 95, p. 241, 1995.
5. M. T. Bohr and Y. A. El-Mansy, "Technology for advanced high-performance microprocessors," IEEE Trans. Electron Dev., vol. 45, p. 620, 1998.
A. K. Stamper, V. McGahay, and J.P. Hummel, Intermetal dielectrics - A five year outlook," Proc. of Diel. for ULSI Multilevel Inter. Conf., , p. 13, 1997.
6. Y. Yi, et al., "Electromigration-induced integration limits on the future ULSI's and the beneficial effects of lower operation temperatures," IEEE Trans. Electron Dev., April 1995, pp. 683-688.
7. B. Davri, R. Dennard, and G. Shahidi, "CMOS scaling for high performance and low power - Next ten years," Proc. IEEE, April 1995, pp. 595-606.
8. H. B. Bakoglu and J. D. Meindl, "Optimal interconnect circuits for VLSI," IEEE Transactions on Electron Devices, vol. ED-32, no. 5, pp. 903–909, May 1985.
9. S.-Y. Oh and K.-J. Chang, "2001 needs for multi-level interconnect technology," IEEE Circuits and Devices, vol. 11, no. 1, pp. 16-21, 1995.
10. D. H. Cho, Y. S. Eo, M. H. Seung, N. H. Kim, J. K. Wee, O. K. Kwon, and H. S. Park, "Interconnect capacitance, crosstalk, and signal propagation delay for 0.35 μm CMOS technology," International Electron Device Meeting Technical Digest, pp. 619-622, Dec. 1996.
11. J. A. Davis, R. Venkatesan, A. Kaloyeros, M. Beylansky, S. J. Souri, K. Banerjee, K. C. Saraswat, A. Rahman, R. Reif, Fellow, Ieee, And J. D. Meindl, Interconnect Limits on Gigascale Integration(Gsi) In The 21st Century", Proc. IEEE, Vol. 89, No. 3, March 2001, pp. 305-324.
12. P. Kapur, J. P. McVittie and K. C. Saraswat. "Technology and Reliability Constrained Future Copper Interconnects -- Part I: Resistance Modeling, IEEE Trans. Electron Dev., Vol. 49, No. 4, April 2002, pp. 590-597.
13. P. Kapur, G. Chandra, J. P. McVittie and K. C. Saraswat. "Technology and Reliability Constrained Future Copper Interconnects -- Part II: Performance Implications," IEEE Trans. Electron Dev., Vol. 49, No. 4, April 2002, pp. 598-604.
14. K. Banerjee, S. J. Souri, P. and K. C. Saraswat, "3-D ICs: A Novel Chip Design for Improving Deep Submicron Interconnect Performance and Systems-on-Chip Integration," Proc. IEEE, Vol. 89, No. 5, May 2001, pp. 602-633.