Effects of Temperature in Deep-Submicron Global Interconnect Optimization in Future Technology Nodes

Mariagrazia Graziano *, Mario R. Casu, Guido Masera, Gianluca Piccinini, Maurizio Zamboni

Politecnico di Torino, Dipartimento di Elettronica, C.so Duca degli Abruzzi 24, I-10129 Torino, Italy

Abstract

The resistance of on-chip interconnects and the current drive of transistors are strongly temperature dependent. As a result, the interconnect performance in Deep-Submicron technologies is affected by temperature in a substantial proportion. In this paper we evaluate thermal effects in global *RLC* interconnects and quantify their impact in a standard optimization procedure based on repeaters insertion. By evaluating the difference between a simple *RC* and an accurate *RLC* model, we show how the temperature induced increase of resistance may reduce the impact of inductance. We also project the evolution of such effects in future CMOS technologies, according to the semiconductor roadmap.

Key words: Interconnects, Temperature, Delay optimization, Inductance, Scaling.

1 Introduction

Long on-chip interconnects have been usually modeled as *RC* distributed lines and *ad hoc* optimization rules have been consequently developed [1][2]. Nowadays, clock frequencies on the order of and higher than 1 GHz require a suitable *RLC* modelization because the inductance of wires is no more negligible. A length-based classification of interconnects that explains when inductance effects have to be taken into account is proposed in [3]. In a recent paper,

* Corresponding author.
Email addresses: mariagrazia.graziano@polito.it (Mariagrazia Graziano), mario.casu@polito.it (Mario R. Casu).

Ismail and Friedman presented the formulation of the interconnect delay as a compact closed-form function of RLC interconnect, driver and receiver parameters [4]. In particular the formula captures the entire range where the interconnects behave as RC or RLC lines by combining the effects of the various parameters. In the same paper, new design formulas for the computation of the optimal number of repeaters and the length of wire segments between repeaters are also proposed. These are expressed in a form such that when the line behaves as RC instead of RLC, the classic formulation by Bakoglu still holds [1].

If thermal effects are taken into account, the interconnect resistance must be expressed as a function of the temperature. Usually a linear model is accurate within the range of on-chip operating temperatures. Since the amount of wire resistance may change the operating regime of the interconnect from pure RC to moderate RLC or "quasi-LC", to achieve an accurate modelization it is important to know the operating temperature of wires. The transistor properties like current drive, on-resistance, off-current and threshold voltage are temperature dependent as well. Since the optimal number of repeaters and the optimal wire sizing depend on both wire and device on-resistance, it is important to incorporate such effects in the design formulas. Some work in this direction has been done for pure RC interconnect taking into account the temperature of the line [5][6], but analyses for global RLC lines are still lacking in the literature.

The width of interconnects is still expected to scale in future technology nodes [1] as described in the International Technology Roadmap of Semiconductor (SIA roadmap) [7]. The wire resistance will increase and will make interconnects more susceptible to thermal effects. The analysis of this phenomenon and its trend in future technologies is carried out in this work.

In section 2 we introduce the equations of the delay of a RLC line and incorporate the temperature dependency in the interconnect and device parameters. Then we evaluate the impact on non-optimized global lines of various lengths in a current VLSI technology. We show the importance of taking inductance into account for accuracy and to avoid timing underestimations. In section 3 the global lines are optimized by a suitable repeater insertion whose optimum number and sizing depend on temperature effects. Then an estimation of the trend in future technologies is proposed in section 4. Finally the conclusion summarizes the achievements of this work.

[1] The SIA roadmap organizes scaled future CMOS processes in Technology Nodes (TN), where each "node" is the minimum metal pitch used on any product of the same generation, for example TN=130 nm is either DRAM half pitch or Metal1 half pitch in logic/microprocessor.

2 Thermal effects in RLC on-chip interconnects

Let's consider a global RLC interconnect of length l driven by a CMOS buffer of resistance R_r and charged by the input capacitance C_r of the output buffer. The 50% delay can be computed with good approximation by the following empirical equation developed by Ismail and Friedman [4]

$$T(\zeta) = \frac{1}{\omega_n}\Big(e^{-2.9\zeta^{1.35}} + 1.48\,\zeta\Big) \tag{1}$$

where

$$\omega_n = \frac{1}{\sqrt{Ll(Cl + C_r)}},\;\; \zeta = \frac{1}{2\sqrt{1+\frac{C_r}{Cl}}}\Big(\frac{Rl}{2Z_0} + \frac{R_r}{Z_0} + \frac{(R_r + Rl)C_r}{\sqrt{LCl^2}}\Big),\;\; Z_0 = \sqrt{\frac{L}{C}}$$

L, R and C are per unit length values ($[L]$=H/m, $[R]$=Ω/m, $[C]$=F/m). For $L \to 0$ (and so $\zeta \to \infty$) eq. (1) converges to the usual delay of a RC line [1][2].

It is clear that since the temperature affects all resistive parameters, i.e. line and drivers, its effect on the interconnect optimization should be accounted for. We have taken interconnect and transistor data from the 2001 SIA roadmap [7] and added typical temperature dependence of interconnect resistance and transistor current both fitted in a linear expression:

$$R(T) = \frac{\rho(T_0)}{WH}\Big[1 + \alpha(T - T_0)\Big] \tag{2}$$

$$I_{dsat}(T) = I_{dsat}(T_0)\Big[1 + \beta(T - T_0)\Big] \tag{3}$$

where W and H are width and thickness of the line and ρ is the resistivity ($[\rho]$=Ωm). ρ and α are known being respectively $1.68 \times 10^{-8}\,\Omega$m and $0.4°\mathrm{C}^{-1}$ for bulk Copper at 20°C. However, the effective resistivity is higher for the effect of the Cu barrier and is about $2.2{\times}10^{-8}\,\Omega$m [7]. The temperature dependency changes as well and we derived a higher value of $0.53°\mathrm{C}^{-1}$ from [8]. $I_{dsat}(T_0)$ is one of the roadmap's specifications and is $900\,\mathrm{Am}^{-1}$ (i.e. current per device width unit) at $T_0 = 25°$C for high performance devices [7]. This value will be constant in future technology nodes. Using Hspice and the BSIM3v3 MOSFET model we derived an approximate value for β of $-1.1\,\mathrm{Am}^{-1}\mathrm{K}^{-1}$ (negative, because the current decreases as temperature increases). The device resistance is given by

$$R_r(T) = \gamma \cdot V_{dd}/I_{dsat}(T) \tag{4}$$

where γ is a fitting coefficient [9]. The other LC parameters can be evaluated by means of proper expressions for the typical configuration of a wire embedded between two other wires of the same metal layer and sandwiched

between two ground planes [11][12]. Let's consider a line of variable length l implemented in a 130 nm roadmap's technology node [7]. Its parameters are about R=$10^5\,\Omega$/m at 300 K and $1.4{\cdot}10^5\,\Omega$/m at 400 K, C=$2{\cdot}10^{-10}$F/m and L=$2{\cdot}10^{-6}$H/m. In figure 1 the RLC and the RC delays of equation (1) is reported, for a typical driver-load pair, as a function of length and at various temperatures, where RC delay is obtained neglecting the inductance L. The percentage difference between the two models is plotted in figure 2, while the delay as a function of temperature is in figure 3. Two cases are reported in each figure: on the top graph (1.A, 2.A, 3.A) a minimum width wire, according to the minimum pitch rule of the 130 nm technology; on the bottom graph (1.B, 2.B, 3.B) a wire with 5 times the minimum width is considered.

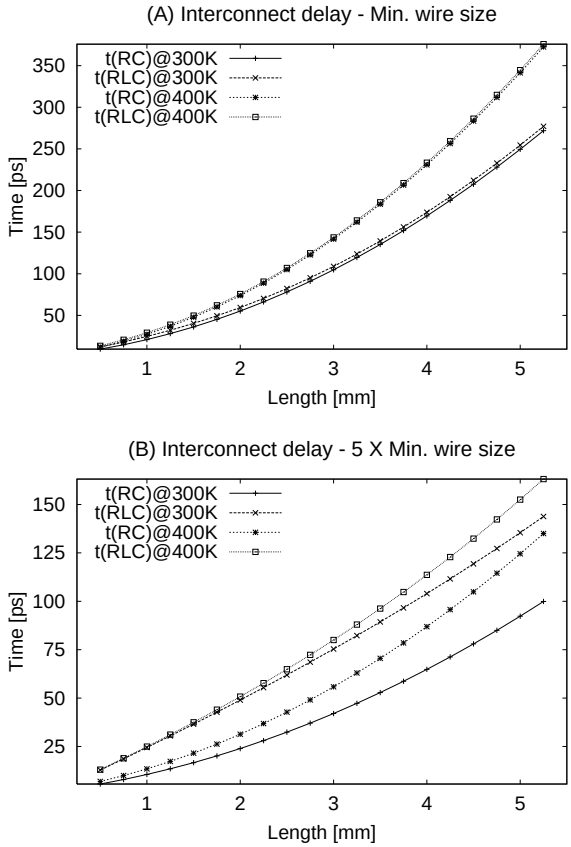

Fig. 1. RC and RLC delays (t(RC) and t(RLC) respectively) a function of length and at different temperatures; minimum wire width (A) and 5× minimum width (B).

The minimum width wire behaves as a RC line except for short lengths where the error RC vs. RLC is on the order of 20% at low T and lower than 20% at high T. In the larger line the resistance is reduced and the inductance effect becomes preeminent. The error is much higher and tends to reduce as T and l increase.

The temperature variation leads to a relevant delay change when the inter-

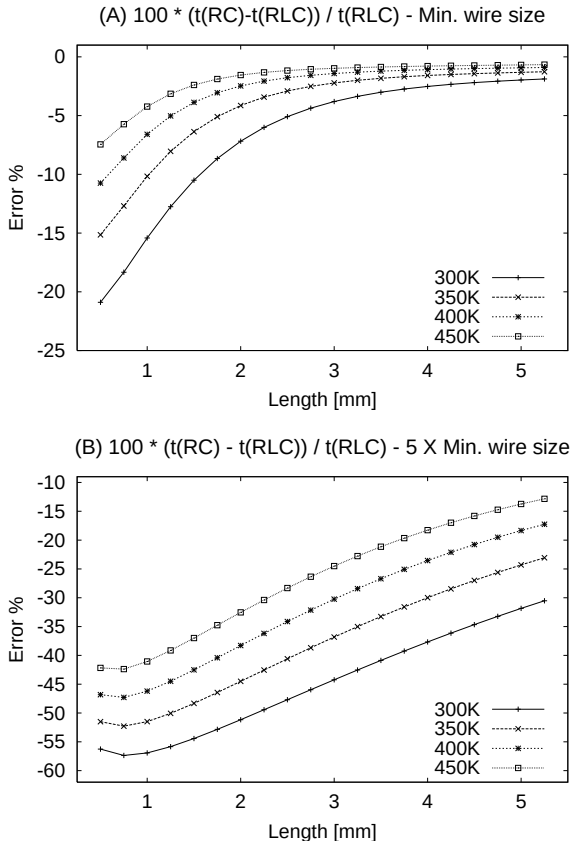

Fig. 2. RC vs. RLC delay error as a function of length and at different temperatures; minimum wire width (A) and 5× minimum width (B).

connect behaves as a RC line, on the order of 50% over the range 300-400 K for minimum width "almost-RC" lines. For the larger "true-RLC" line it is less than 10% over the same range. A proper wire sizing will be then useful not only for the delay reduction, but also for the delay variance minimization if the operating temperature is not known or not precisely controlled.
An accurate estimation of the temperature is needed to improve the correct timing estimation. Since the delays are monotone increasing functions of temperature, a worst case approach could also be followed, but this would result in an overdesign cost that cannot be acceptable for high performance designs or that might impact the global power dissipation.

The RC delay is always smaller than the RLC one regardless of the temperature and the line length. Therefore an accurate RLC modelization is needed to avoid timing underestimations. The RC delay is a stronger function of l at a given temperature. This is not surprising because the RC delay is $\propto l^2$ while the delay of a pure LC lossless transmission line is $\propto l$ (time of flight $\tau_{OF} = l\sqrt{LC}$). The usual technique of placing buffers between RC line segments is very effective because of the square nature of the RC delay with l. In a lossless pure LC line this technique will be ineffective [4]. If we evaluate the

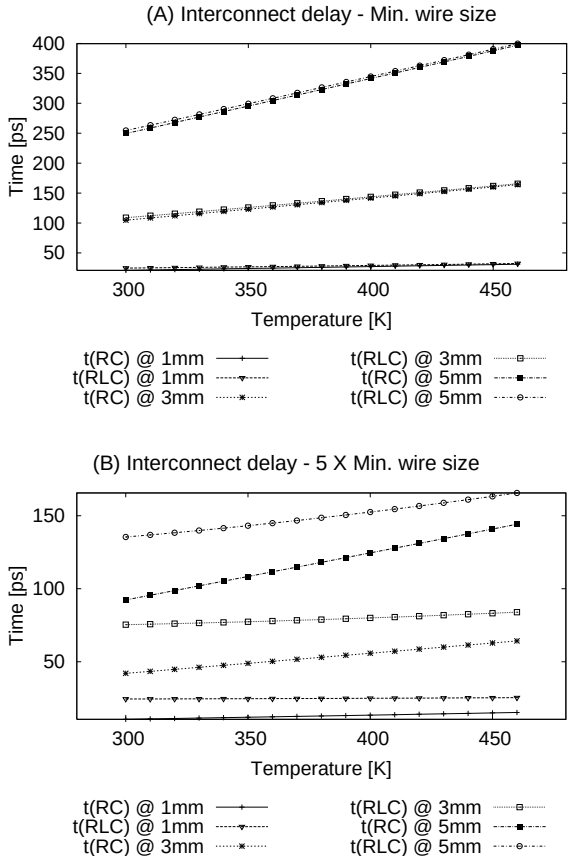

Fig. 3. RC and RLC optimal delays using repeater insertion (t(RC) and t(RLC) respectively) as a function of temperature and at different length; min. wire width (A) and 5× min. width (B).

optimum number of repeaters we can expect that a lower number is needed for the RLC case with respect to the RC one. However, as the temperature increases, the number of repeaters will approach the optimum number found for the RC case.

3 Delay optimization via repeaters insertion

Optimum size and number of repeaters in RLC lines, W_{opt} and $N_{opt} = \lceil n_{opt} \rceil$, are approximated by the following expressions [4]

$$n_{opt} = \frac{\sqrt{\dfrac{CRl^2}{R_{r0}C_{r0}}}}{\left(1 + 0.18\, L^3/(RR_{r0}C_{r0})^3\right)^{0.3}} \qquad W_{opt} = \frac{\sqrt{\dfrac{CR_{r0}}{C_{r0}R}}}{\left(1 + 0.16\, L^3/(RR_{r0}C_{r0})^3\right)^{0.24}}$$

where R_{r0} and C_{r0} are resistance and capacitance of a minimum width driver. The optimum device resistance is given by $R_{ropt} = R_{r0}/W_{opt}$. The above for-

mulas tend to the classic Bakoglu's formulas for RC lines when $L \to 0$ [1].

In figures 4 and 5 the optimum number of repeaters is reported as a function of length and temperature for the 130 nm global interconnect, assuming that the output load is a fraction of the input capacitance of the repeater itself.

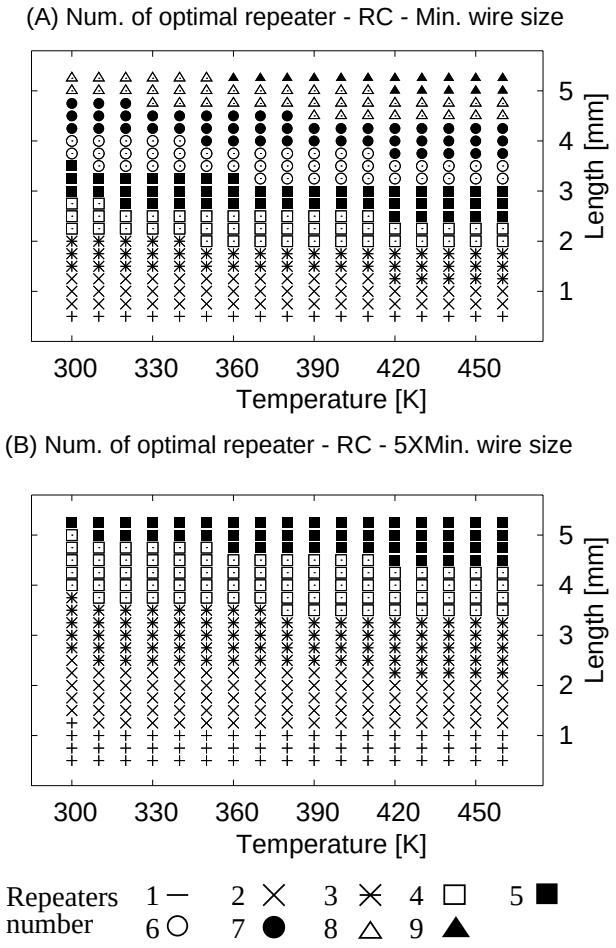

Fig. 4. Sub-optimum number of repeaters for a 130 nm wire modeled as a RC line as a function of length and temperature; min. wirewidth (A) and 5× min. width (B).

In both figures for the top graph (A) the minimum width has been used, while 5× the minimum width in the bottom graph (B). On figure 4 the line has been optimized as if it was a RC line (equation (3) with $L \to 0$) while on figure 5 as a RLC line (same equation with $L \neq 0$). The minimum width wire requires a high number of repeaters because the line is "almost-RC", that is unaffected by the small differences between 4.A and 5.A. For true RLC lines much less repeaters are needed (figure 5.B). The erroneous RC evaluation of the actual RLC delay leads to an excessive number of repeaters (4.B vs 5.B). This is on the one hand sub-optimum under the delay point of view but on the

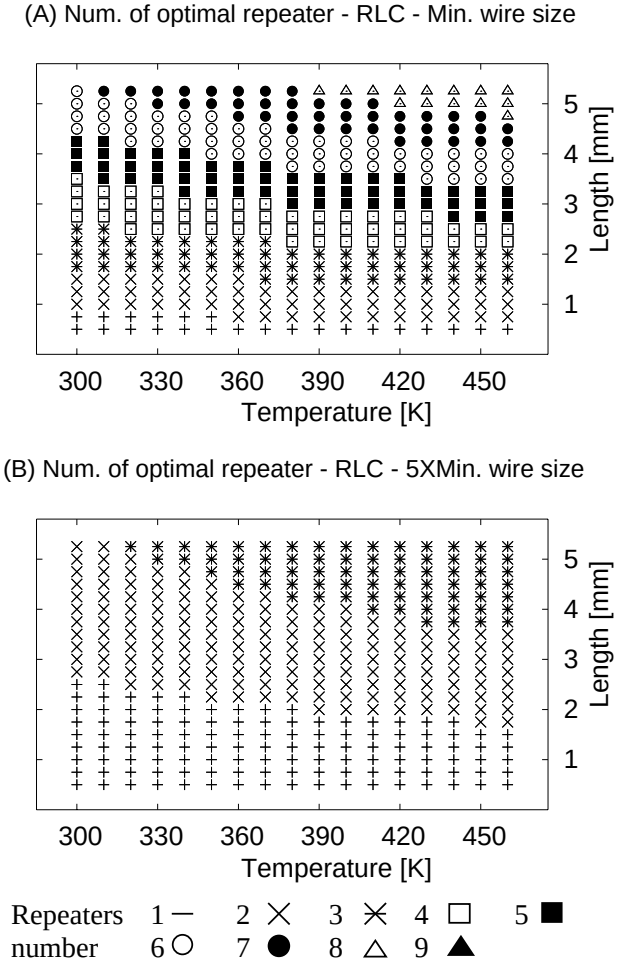

Fig. 5. Optimum number of repeaters for a 130 nm wire correctly modeled as an RLC line as a function of length and temperature; min. wirewidth (A) and 5× min. width (B).

other hand detrimental for the power consumption. The lower sensitivity to temperature variations of the *RLC* line already observed before is reflected in the optimum number of repeaters that do not vary as for *RC* lines at different T. In table 1 the values of W_{opt} obtained during the optimization process are reported. The ranges result from the variation of l and T.

Table 1
W_{opt} size of optimized repeaters for the 130 nm technology node.

	Min wire size Min wire size W_{opt} [μm]	5XMin wire size 5XMin wire size W_{opt} [μm]
RC model	18.4-21.2	51.1-58.7
RLC model	17-18.6	32.6-38.4

Once again the minimum width line is confirmed being "almost-RC" with respect to the larger line as shown by the similar values in the first column of table 1. On the contrary for the 5× minimum width lines the inaccurate RC model largely overestimates the sizing of the repeaters (second column).

If we plot the delay after optimization we observe that it is now almost linear with length. It is interesting to evaluate the true RLC delay of a line optimized as if it was a RC line and to compare it with the delay of the RLC optimized line. The corresponding delay is higher because of the capacitance of the additional not useful buffers. In figure 6 the two delays are reported as a function of length and at various temperatures. Again the minimum width (6.A) and the larger wire (6.B) are considered.

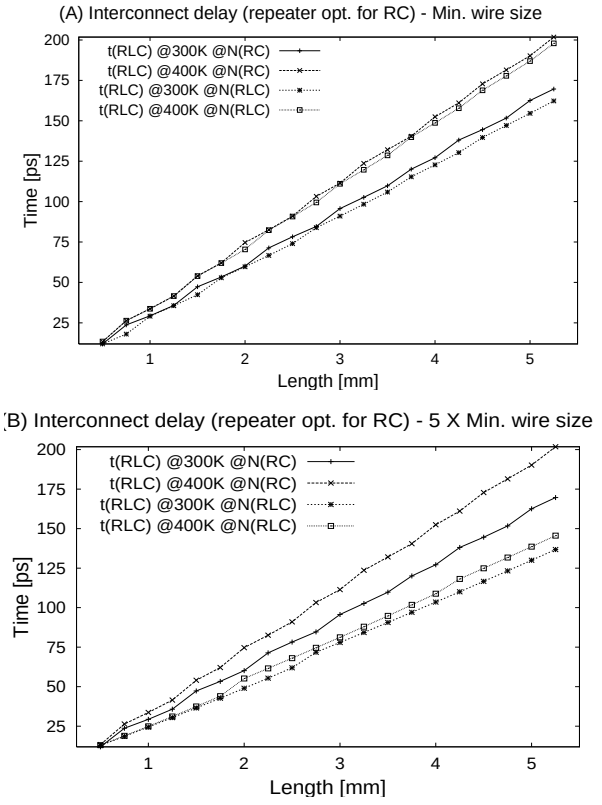

Fig. 6. RLC delay of a line designed using an optimum number N of repeaters (t(RLC) @ N(RLC)) compared to the RLC delay of a sub-optimal optimization using the number of repeaters obtained using a RC model (t(RLC) @ N(RC)). Minimum wire width (A) and 5× minimum (B).

The correct estimation of the interconnect temperature as an input for the optimization may be difficult. In addition thermal gradients are possible such that the temperature is non-uniform along the interconnect length [5][6]. Therefore we have analyzed what happens if the interconnect is optimized at a given temperature and its delay is evaluated at a different temperature. Among the obtained results the most significant ones are reported in figure 7

for the 5 mm case, minimum width (7.A) and 5× minimum (7.B).

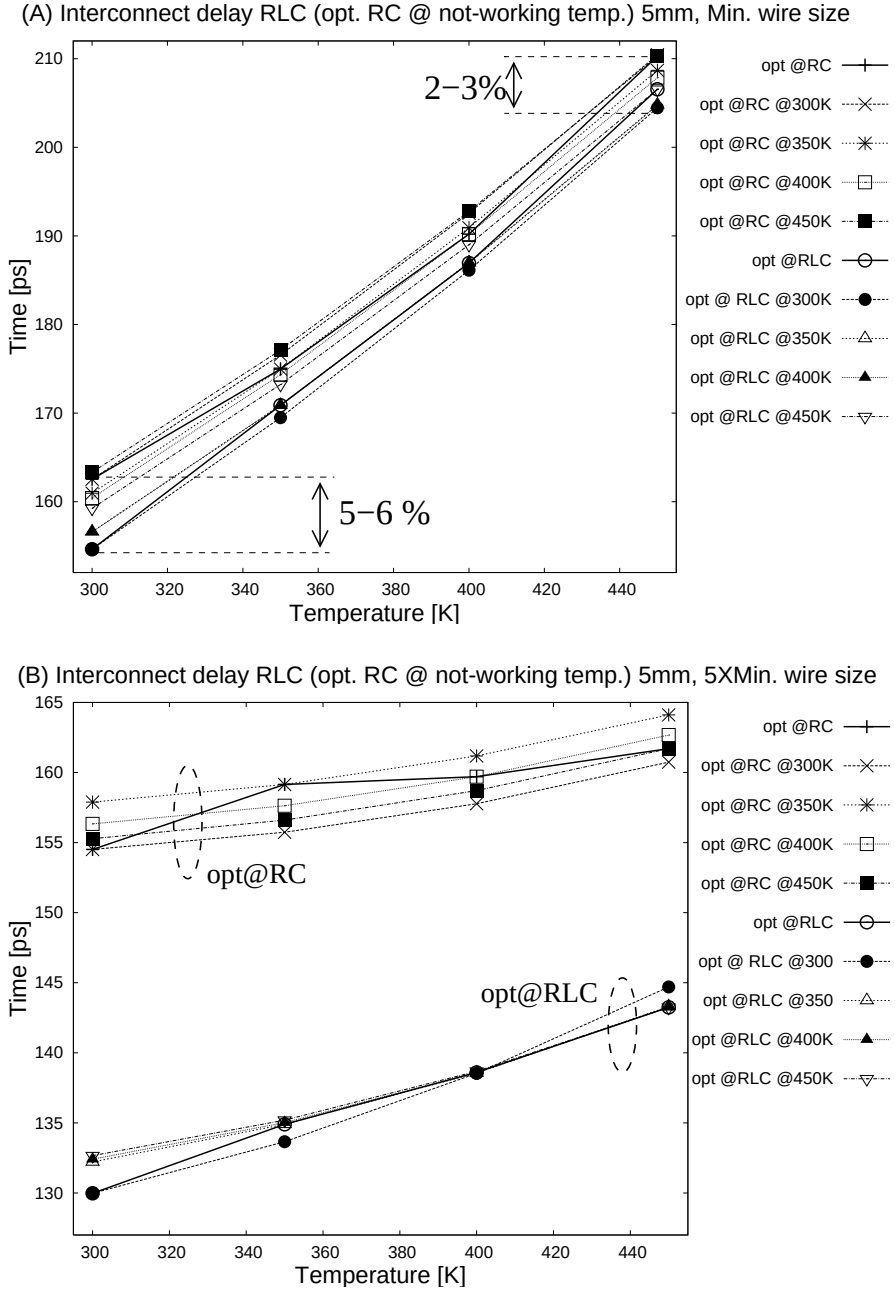

Fig. 7. RLC delays for minimum width (A) and 5× minimum (B) lines optimized at a given temperature and evaluated at other temperatures within the range. Optimal repeater insertion and sizing using a *RLC* model (opt@RLC) and sub-optimal solution using a *RC* model (opt@RC).

In the same graphs the true *RLC* delays of lines correctly optimized with lines optimized as if they were *RC* interconnects are compared as well. The thicker solid lines are the optimum delays (both *RC* and *RLC*) all over the temperature range while the other curves are optimized only at one temperature

and thus are suboptimum over the entire range. The thick curve, of course, crosses the other curves at temperatures where suboptimum lines have been optimized. The curves in figure 7.A are really close to each other, meaning that for a narrow "almost-*RC*" wire the use of a *RC* or a *RLC* model does not imply a large error (less than 5%). On the contrary, as already shown in figure 6, the "true-*RLC*" line correctly optimized using the *RLC* model presents a much smaller delay than the same lines optimized as *RC* (graph 7.B). Sometimes suboptimum is slightly better than optimum because of the integer number of buffers obtained from $\lceil n_{opt} \rceil$. The true optimum would be obtained by using a non-integer n_{opt} value which is obviously impossible. We observe again that the "almost-*RC*" narrow wire is more sensitive to thermal effects such that a line optimized at 300 K presents a delay increase of about 30% at 400 K ("×" dotted curve in figure 7.A) while the variation for the larger line is 10% ("×" in figure 7.B). Moreover the difference between *RC* optimized and *RLC* optimized is slight, less than 5% for minimum wire width (graph 7.A).

The difference between lines optimized at different temperatures is small meaning that the optimization can be safely done at a reference temperature. The minimization of the variation over the entire range is obtained by setting the optimization temperature at about the middle of the range. Such approach can be beneficial for the non-critical paths, provided that the maximum variation does not make them critical. On the contrary critical paths have to be treated using a worst case approach. Therefore the designer shall evaluate the maximum temperature over the entire range and optimize the line consequently. Since the delay is a monotonically increasing function of temperature, this approach will ensure that the delay constraints will be satisfied at all range.

4 Interconnect optimization in future technology nodes

It is interesting to study the trend of our previous analysis in future roadmap's technologies. We evaluated the *RLC* delay of global interconnect lines optimized in the near term technologies from 130 nm to 65 nm [7]. The parameters for the 65 nm line are about R=$4.8{\cdot}10^5\,\Omega$/m at 300 K, C=$1.6{\cdot}10^{-10}$F/m and L=$2.2{\cdot}10^{-6}$H/m. In figure 8 the *RC* and *RLC* delays as a function of length and at various temperatures are reported for the previous 2001 130 nm and for the 2007 65 nm nodes. Both minimum (8.A) and 5× the minimum width lines (8.B) are reported. The error between *RC* and *RLC* delays is shown in the two graphs of figure 9, that correspond to the two graphs of previous figure 8). The number of repeaters for the same *RLC* lines are reported in figure 10 calculated at different temperatures (in the range 300-450 K) (again, results for two possible widths are in the 9.A and 9.B graphs).

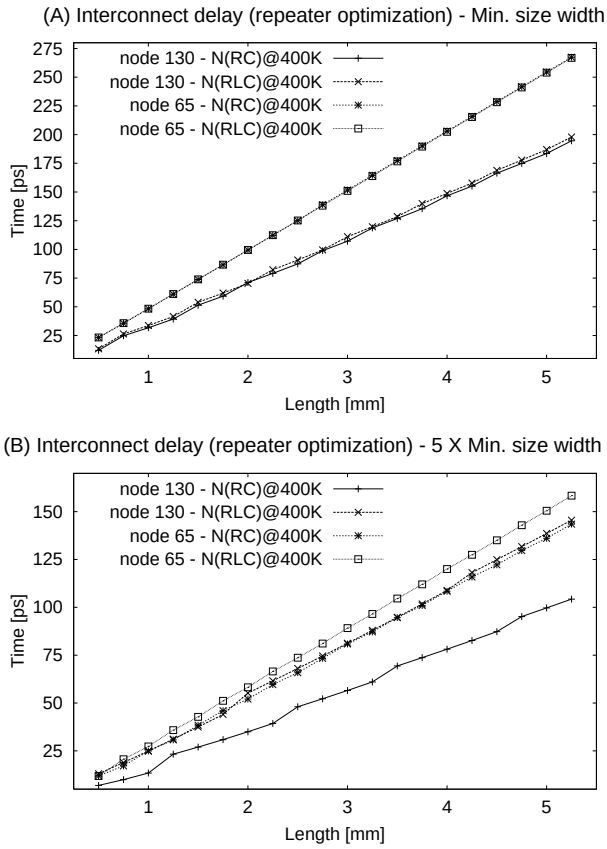

Fig. 8. RLC and RC delays of 130 nm and 65 nm lines: minimum width (A) and 5× minimum width (B) interconnects. Solutions found using optimum number of repeater (N(RLC)) and suboptimum number of of repeater (N(RC)) are reported.

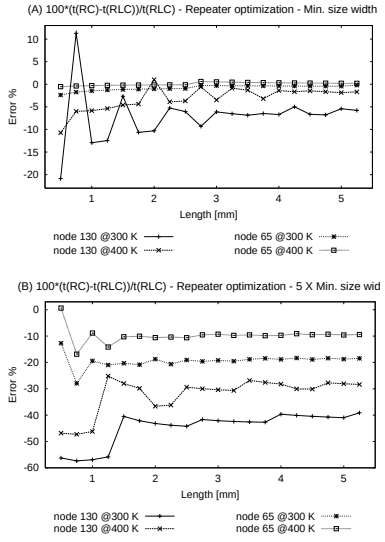

Fig. 9. RLC and RC error (A) of 130 nm and 65 nm lines: minimum width (A) and 5× minimum width (B) interconnects.

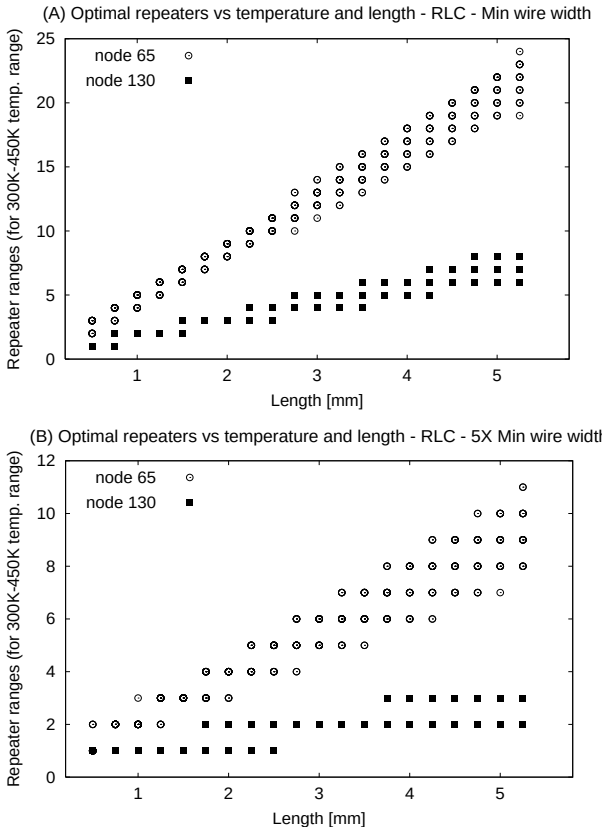

Fig. 10. RLC number of repeaters of 130 nm and 65 nm lines: minimum width (A) and 5× minimum width (B) interconnects.

The minimum width line in the new technology is much more resistive than the 130 nm line as shown by the complete overlap of RC and RLC delay curves in figure 8.A. As for the larger line the difference is more appreciable but still much lower than for the 130 nm case (8.B). We can also observe that at a given length the delay of the 65 nm node is worse than the corresponding 130 nm delay. One could argue that such a comparison is not correct since scaled interconnect lengths should be compared. In this case we should compare the 130 nm delay at length l to the 65 nm delay at length $l \times 65/130 = l/2$ therefore observing a reduction in delay. However the average length of global interconnects is strictly related to the chip size that, according to the SIA roadmap document, does not scale with technology nodes [7]. Thus the comparison at a given length is correct. The overall result is not new because the non-scalability of global wires is a well-known problem [10] but we observe that the impact of both non-scaled length and high temperature make them more RC than RLC as clear from figures.

For what concerns the optimum number of repeaters, in figure 10 we see that a higher number of repeaters is needed at a given length for the 65 nm line. This is also true at scaled lengths ($l \rightarrow l/2$). The reason is again that the scaled line is much more resistive than the 130 nm line.

In figure 11 the delay trend is reported for all the intermediate nodes from 130 to 65 nm at a length of 5 mm, while in figure 12 the delay as a function of temperature is plotted for the same technologies. Figure 13 reports the error *RC* vs. *RLC* for all the nodes from 130 to 65 nm at a length of 5 mm.

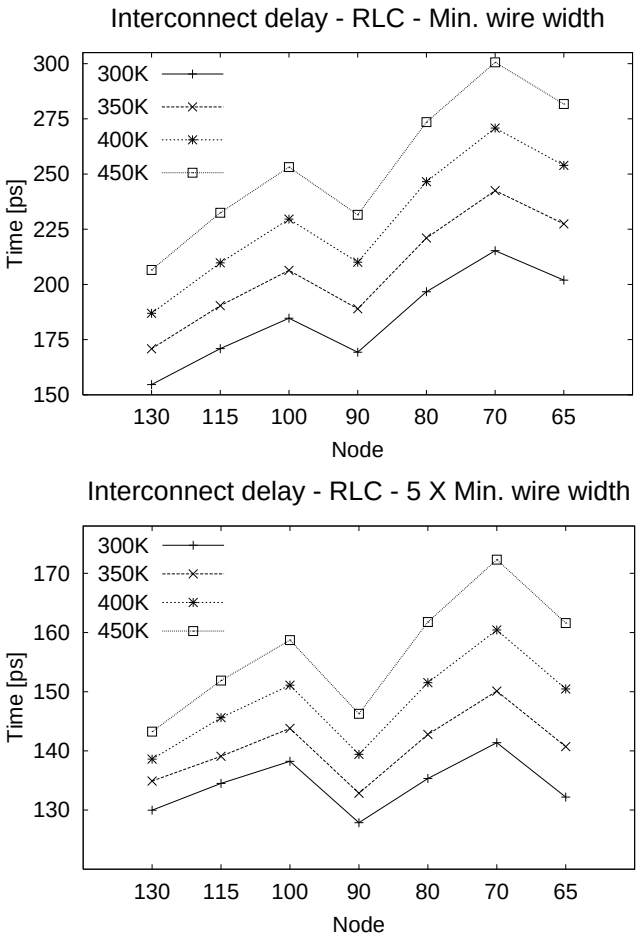

Fig. 11. RLC delay of 5 mm global interconnects as a function of the SIA roadmap's technology node, from 130 nm. to 65 nm.: min. width (A), 5X min.n width (B).

The delay in figure 11 tends to increase and presents an abrupt reduction at 90 nm due to the foreseen introduction of a new low-k material that results in a reduction of capacitance. The difference between the *RC* and *RLC* models (figure 13) tends to diminish for higher temperatures as already shown before and also for scaled technologies approaching the last node. Therefore we conclude that contrarily to the common opinion, the inductance effect are less dominant than expected in future technology nodes because global wires are more resistive. This phenomenon is exacerbated by the temperature effect that increases the wire resistance.

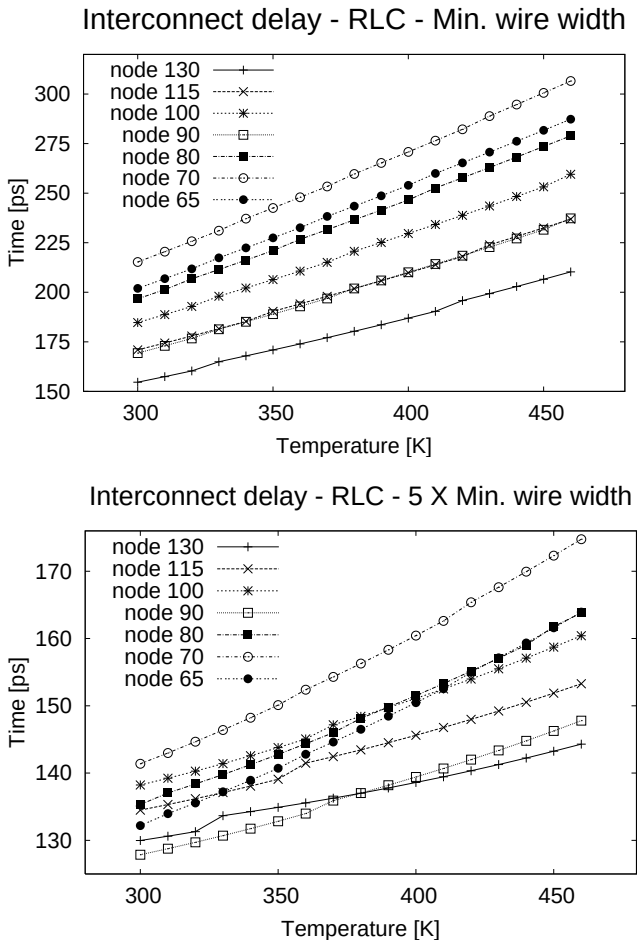

Fig. 12. RLC delay as a function of temperature fro technology nodes from 130 nm to 65 nm: min. width (A), 5X min.n width (B).

5 Conclusions

In this paper we have shown how the temperature dependence of interconnect and driver resistance impact the behavior of global wires in scaled VLSI technologies. The effects of temperature are particularly important in *RC* interconnects. We have seen that if inductance effects are not negligible so that a *RLC* modelization is needed, they tend to mitigate the impact of temperature. We showed how temperature may change the behavior of a line from *RC* to *RLC* by modifying the resistance value. Therefore it is very important to incorporate thermal effects into analysis and design. We also showed that the optimization of interconnect performance, by using repeater insertion and sizing, may give rise to strongly different results over the operating range of temperature.

As technology improves and lithography allows to scale wire widths, the resistance per unit length increases and tend to shield the inductance effects. As a

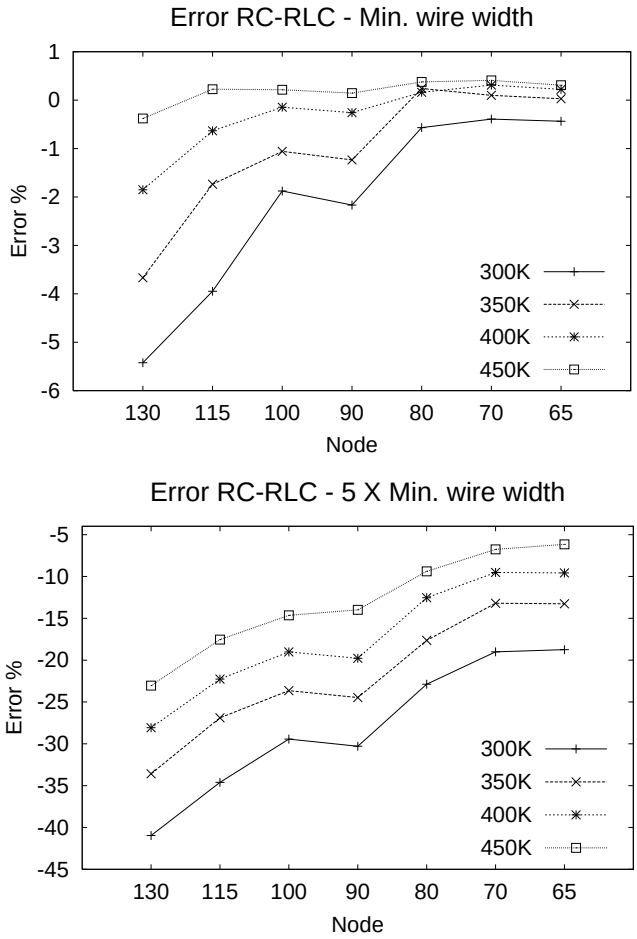

Fig. 13. Difference between RC and RLC models of 5 mm global interconnects as a function of the SIA roadmap's technology node, from 130 nm. to 65 nm.: min. width (A), 5X min.n width (B).

result, we foresee that in future wires the inductance impact will be mitigated and that the effects of temperature will be more and more important.

6 Acknowledgements

This work was funded by the Politecnico di Torino center of excellence for the development of multimedia radio-communications (CERCOM).

References

[1] H.B. Bakoglu, *Circuits, Interconnections and Packaging for VLSI.* Reading, MA: Addison-Wesley, 1990.

[2] T. Sakurai, "Closed Form Expressions for Interconnection Delay, Coupling, and Crosstalk in VLSI's," *IEEE Trans. on Electron Devices,* vol. 40, no. 1, pp. 118-124, Jan. 1993.

[3] A. Deutsch *et alii,* "On-chip Wiring Design Challenges for Gigahertz Operation," *Proc. IEEE*, Vol. 89, No. 4, pp. 529-555, April 2001.

[4] Y.I. Ismail and E.G. Friedman, "Effects of Inductance on the Propagation Delay and Repeater Insertion in VLSI Circuits," *IEEE Trans. on VLSI,* vol. 8, no. 2, pp. 195-206, April 2000.

[5] A.H. Ajami *et alii,* "Analysis of Non-Uniform Temperature-Dependent Interconnect Performance in High Performance ICs," Proc. DAC 2001, pp. 567-572.

[6] A.H. Ajami *et alii,* "Analysis of Substrate Thermal Gradient Effects on Optimal Buffer Insertion," Proc. of ICCAD, Nov. 2001, pp. 44-48.

[7] The international technology roadmap for semiconductors, 2001, SIA.

[8] P. Kapur *et al.*, "Technology and Reliability Constrained Future Copper Interconnects–Part I: Resistance Modeling," *IEEE Trans. on Electron Devices*, vol. 49, no. 4, pp. 590-597, Apr. 2002.

[9] D. Sylvester and C. Hu, "Analytical Modeling and Characterization of Deep-Submicrometer Interconnect," *Proc. IEEE*, vol. 89, no. 5, pp. 634-664, May 2001.

[10] D. Sylvester and K. Keutzer, "Impact of Small Process Geometries on Microarchitectures in Systems on a Chip," *Proc. IEEE*, vol. 89, no. 4, pp. 467-489, April 2001.

[11] J.H. Chern *et alii,* "Multilevel Metal Capacitance Models for CAD Design Synthesis Systems, IEEE Electron Device Letters, vol. 13, pp. 32-34, Jan. 1992.

[12] F.W. Grover, *Inductance Calculations: Working Formulas and Tables,* New York: Van Nostrand, 1946.