Keep on Shrinking Interconnect Size: is it Still the Best Solution?

Denis Deschacht, Sébastien de Rivaz, Alexis Farcy, Thierry Lacrevaz, Bernard Flechet

Keep on shrinking interconnect size: is it still the best solution?

D. Deschacht[1], S. de Rivaz[2], A. Farcy[3], T. Lacrevaz[2], B. Flechet[2]

[1]Université Montpellier 2, LIRMM, UMR CNRS 5506, CC 477, 161 rue Ada,
34 095 Montpellier Cedex 5 – France

[2]Université de Savoie, IMEP-LAHC, UMR CNRS 5130, 73376 Le Bourget du Lac – France

[3]STMicroelectronics, 38920 Crolles – France

Tel : 33 4 67 41 86 42, Fax : 33 4 67 41 85 00

Denis.Deschacht@lirmm.fr

Abstract

According to the evolution between each new technological generation of CMOS ICs, ITRS suggests a reduction in interconnect sizes by a factor of around square root of 2. In this paper a reference design rule is based on a perfectly controlled technology of the CMOS 45 nm node, with interconnects width equal to their separation space. Our works are focused on the impact on signal transmission speed and delay along interconnects of decreasing the space or width. To avoid new industrial manufacturing constraints on cost and reliability, this study is performed without modifying process and materials used in the BEOL of CMOS 45 nm IC. We will study interconnects of 50 nm width, with a 50 nm space between lines in accordance with CMOS 32 nm FEOL requirements. In the second time we will relax geometrical constraints to enlarge the scope of application.

Introduction

According to Moore's law, each new generation of ICs has to target a high level of integration. With the dimension shrink and the clock frequency increase, power consumption, delay time and parasitic crosstalk level require a strong decrease in order to maintain high interconnect performance. ITRS [1] recommends 17 % of improvement on performance when technology scales down from one node to the other. Thus, the achievement of efficient interconnect networks in the Back End Of Line (BEOL) is a difficult challenge for designers. They are confronted to an increase of distributed interconnect resistances. This effect comes from the reduction of the cross-section of copper wires, causing also a higher resistivity of copper and a higher ratio between metallic barrier and copper thicknesses when dimensions fall [2], [3]. Moreover performance suffers from increase of capacitance and especially mutual capacitance with the interconnect space decrease. To get round this problem, large drivers with fast responses can be chosen by designers. Nevertheless, these drivers generate costly growth of silicone area in the IC and are not always sufficient to guaranty required performance, especially for critically long interconnects (few hundred of μm). The only remaining degree of freedom for the designer consists in relaxing interconnects density in one metal level because process, materials and technological stacks are laid down.

In a first step different EM simulations are carried out on interconnects from the 45 nm to the 32 nm generations in order to extract RLCG parameters [4] and access sensibility of interconnect width-space pair on line parameters, without modifying the various technological stages of manufacture, nor the materials used. Because of the high density of the BEOL network, not only self RLCG but also neighbouring mutual parameters between interconnects are considered. Moreover, access resistances at the interface due to contacts and inter levels vias are also taken into account. In a second part we will evaluate the interconnect network performance inside the same metal level. Two main features are concerned: the worst-case crosstalk level and the worst-case delay. Time domain simulations are achieved to obtain performance according to different driver sizes. In a third part, we will look at if by relaxing interconnect density in some dedicated intermediate metal level, interconnects can be advantageously spaced or widened without significantly going against integration, in addition satisfy the performance gain required and enlarge the scope of application.

1. Description of the simulation conditions.

For this study, we consider a configuration of parallel coupled lines, representative of intra-level interconnects networks, and a typical geometry of an intermediate metal-layer. Table 1 illustrates parameters specification of the 45nm technology node.

TABLE I.   TYPICAL 45 NM INTERCONNECT PARAMETERS

BEOL	Dimensions	Materials
Metal	w=s≈70 nm t ≈ 140 nm	copper σ ≈ 35 MS/m
Dielectric SiOCH	h ≈ 120 nm	$\varepsilon'_r \approx 2.6$ $\varepsilon''_r = 0$
Metal barrier TaN/Ta	$t_{mb} \approx 7$ nm	σ ≈ 1,4 MS/m
Dielectric barrier SiCN	$t_{db} \approx 40$ nm	$\varepsilon'_r \approx 5$

The Cu interconnects stand between two very dense metal layers, so that perfect metallic walls are taken into account on both sides of the wires. We used these dedicated ground planes as return paths. Self and coupled RLCG parameters are accurately evaluated by electromagnetic simulations. These last take into account interconnect width W, space S and thickness t, dielectric high h as well as technologic stack including material properties, metallic and dielectric barriers, as illustrated on Figure 1.
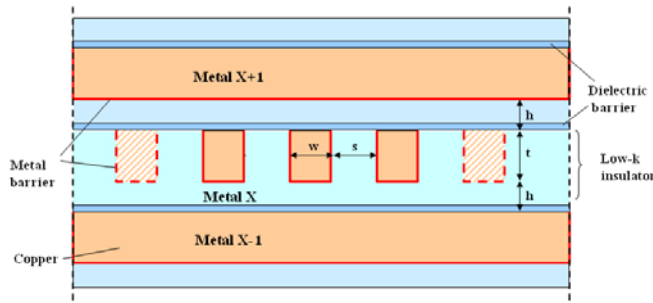
Figure1: Illustration of intermediate metal levels modeling in BEOL

Numerous parameters such as thickness, dielectric high as well as barriers thicknesses and materials strictly depend on the technology stack. So, in this work the only remaining degree of freedom to lower resistances or whole capacitances consists in relaxing width or/and space between lines. Electromagnetic simulations give self and coupled RLCG parameters for different configurations width and space.

For the actual case, the conductance G is negligible and must not be incorporated in the model. These RLC values are used to build the distributed π-RLC model (we use 40 cells) in the electrical simulations under HSPICE software. For each line, the driver is modelled as a linear resistor (i.e. Thevenin model) with an input slope of 10 ps. Loading capacitance of 0.284 fF representative of realistic CMOS transistor gate is added at the far-end of the line. The power supply is 1 V. We consider three adjacent lines driven by inverters as illustrated on Figure 2.
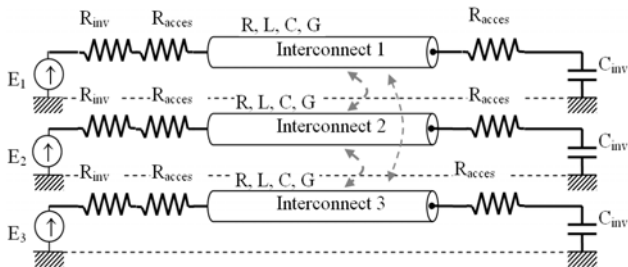

Figure 2: Illustration of 3 coupled interconnects with respective loads and excitations

2. Performance evaluation for configuration: W = S = 50 nm.

According to the reduction of their dimensions, the increase of interconnect resistance significantly degrades the signal propagation. To highlight this phenomenon, we will study the worst-case propagation delay: interconnect in the middle is driven from low to high level while two adjacent ones are driven from high to low level. The propagation delay is the duration between the time of excitation (50 % of voltage level) at the input of the upstream inverter and the time of arrival at the input of the downstream inverter. Figure 3 shows the shape of the signal at the output of the interconnection for the same input signal for different length of line.
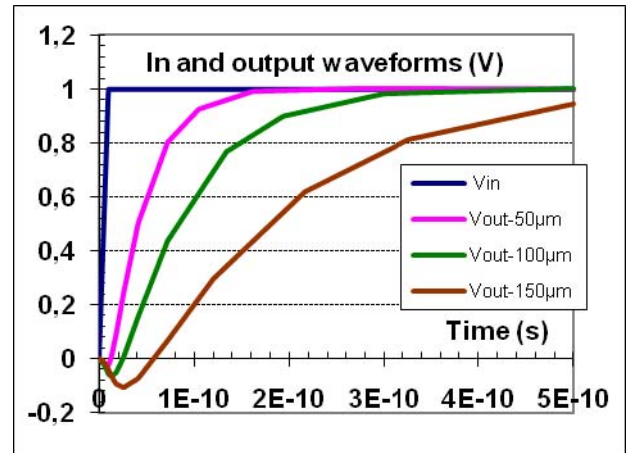

Figure 3: Illustration of the rise of transmission delay with interconnect length

The increase of the transmission time from a low level to a high level will limit the rate of signals to be transmitted. One can consider that the logical level is stable and perfectly transmitted as long as the transition time should not exceed the quarter of the period of the signal to be transmitted. From this consideration we can thus determine the frequency of maximum use according to the length of the lines (Figure 4).
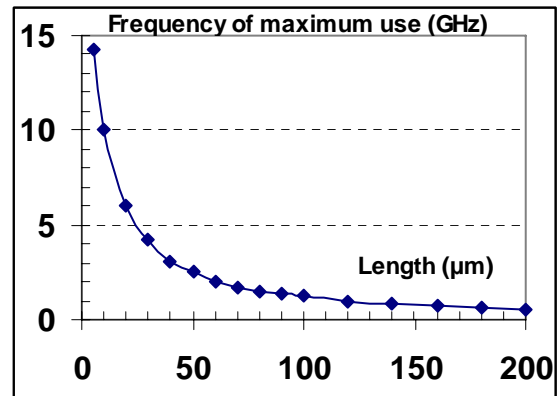

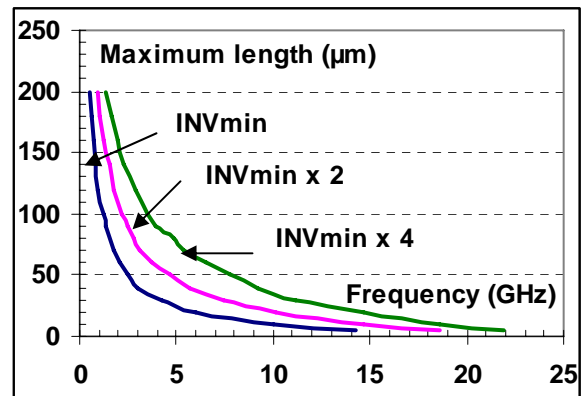Figure 4: Frequency of maximum use versus length of lines.

Figure 5: Maximum interconnects reachable lengths versus transmission frequency for different buffer size (nominal interconnect space S = 50 nm)

One knows that higher the frequency of the signals to be transmitted is, shorter interconnects must be. To increase this one, the minimum buffer size (INV_{min}) can be increased by a factor two or four. Figure 5 illustrates the variations obtained for these three different buffer sizes.

It is worthy to note that a limitation exists due to the crosstalk voltage which appears on the central line at a constant logic level when the two adjacent lines commutate with the same transition. This limitation is around 220 µm for INV_{min} and 140 µm for INV_{min} x 4. In the following it will be questioned if it is better to increase the width of the lines to decrease their resistances or to increase the space between the lines to decrease the coupling capacitances. The aimed performance metric will be the maximum reachable interconnect length meeting a given specified transmission rate.

3. Interconnect density relaxation.

The variations given on Figure 6 show that the width increase (W = 70 nm) does not help to increase the maximum frequency of the signals to be transmitted. Nevertheless this configuration allows the use of longer interconnects, since it involves also a reduction of the crosstalk voltage.
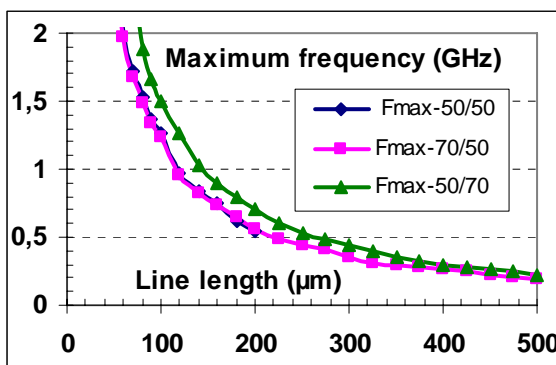

Figure 6: Maximum frequency of the signals to be transmitted versus interconnect length, for different W/S configurations.
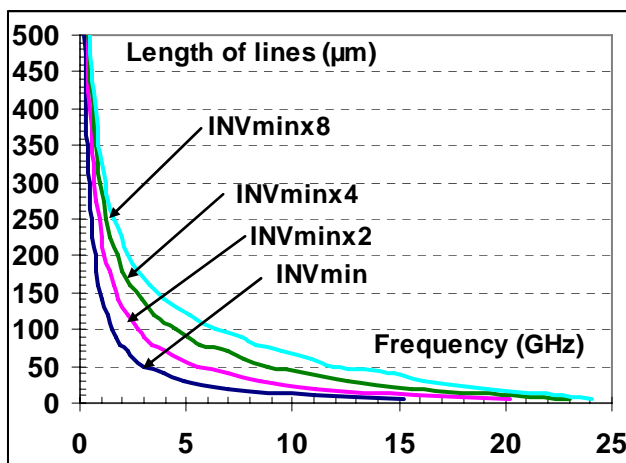

Figure 7: Maximum interconnect reachable length versus transmission frequency for different buffer size (relaxed interconnect space S = 70 nm)

On the other hand, increasing the space between the lines (S = 70 nm) allows for the same frequency to obtain larger reachable interconnect lengths. For high frequency (F > 3 GHz), the lines remains short. Figure 7 shows the maximum interconnect length with this relaxed geometrical configuration for various buffer sizes.

Rate performance (maximum frequency range and minimum delay) are plotted on Figure 8 for a typical 45 nm technology and a minimum buffer size. The dotted curve gives the aimed improvement of 17% on performance recommended by ITRS in comparison with performance in the actual 45 nm configuration (W/S = 70/70 nm). This last reference curve is compared with the performance curve of the goaled configuration (W/S = 50/50 nm). For the small lengths (L<140 µm), limit imposed by the crosstalk voltage becomes too high (V > 0.45 V). Nevertheless it is sufficient to double the buffer size to lower the crosstalk under harmful levels. Moreover the structure 50/50 with a buffer twice larger fully satisfied the improvement conditions necessary when technology scales down from 45 nm to 32 nm. Indeed the maximum frequency to be used is in average 50 % higher and the worst-case delay always 25 % faster, on this length range (Figure 8).
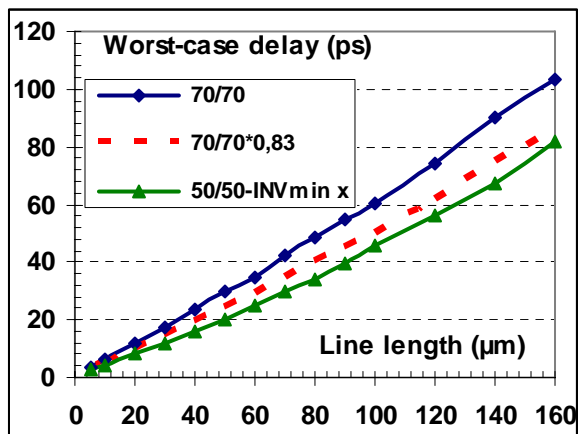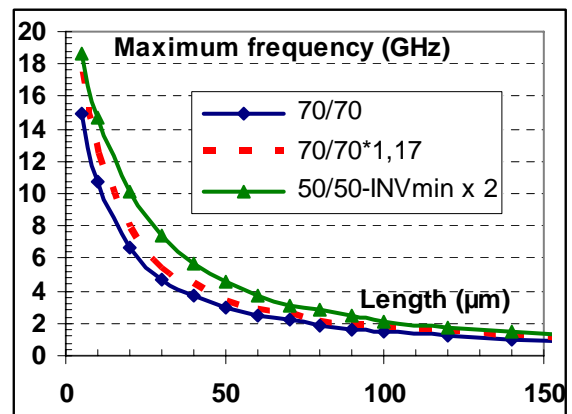

Figure 8: Performance (maximum frequency range and minimum delay) comparison for two different pitches (small lines).

For the higher lengths, configuration 50/70 associated with a buffer twice larger will have to be chosen. Profit in maximum frequency to be used reaches 75 % for L = 150 µm, to decrease towards 50 % for L = 350 µm, 40 % for L = 400

µm to reach finally the 17 % targeted for L = 750 µm. For the worst-case delay, the targeted performance improvement is only satisfied for L < 650 µm (Figure 9).
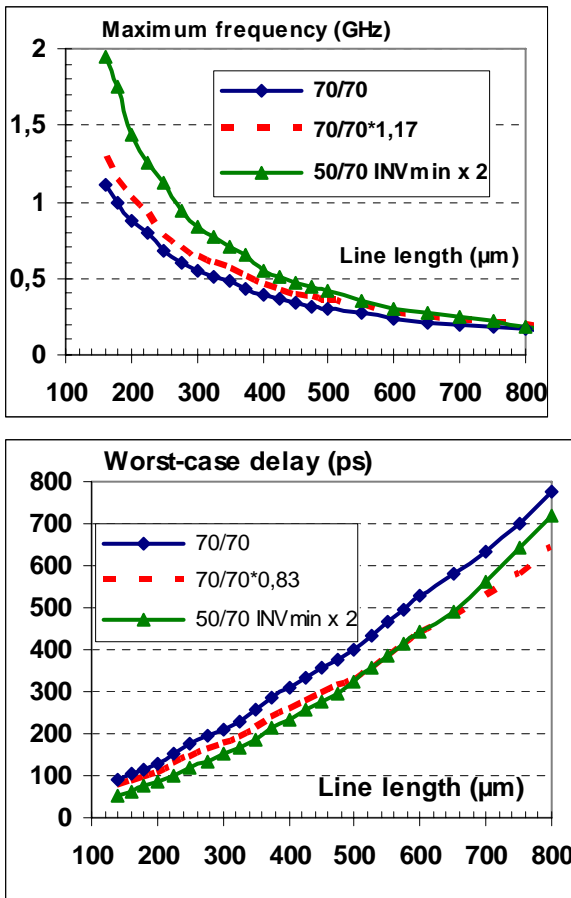

Figure 9: Performance (maximum frequency and minimum delay) comparison for two different pitches (higher length)

For the line length higher than 650 µm, it is necessary to increase the buffer's size or to use another optimization method like splitting up the lines into two parts.

4. Conclusion

We showed in this paper that high density interconnects implies critical increase of the transition time of high speed signals. Therefore the frequency of the signals to be transmitted is limited. From a well tried 45 nm technological node, we determined the conditions of designing less wide and less spaced interconnects (W=S=50 nm), while answering to the required improvements expected for the 32 nm technological node in term of performance. This solution is adequate for most current interconnect lengths lower than 150 µm and for the first levels of metallization. Higher interconnect lengths, which are infrequent in BEOL, require a relaxed constraints on space between lines (S increases from 50 to 70 nm). By following these design rules, interconnects delays and circuit rates are strongly improved without modifying the well tried 45 nm BEOL process, a very cost efficient way is in addition guaranteed.

References

[1] ITRS International Technology Roadmap for Semiconductors, http://public.itrs.net (2009).

[2] A. Farcy, M. Gallitre, V. Arnal, J. Torres, B. Fléchet, P. Ancey, "Evolution and challenges of interconnect technologies and performance", Invited paper, 12th IEEE Signal Propagation on Interconnects, May 12-15, 2008, Avignon, France.

[3] J. Gambino, F. Chen, and J. He, "Copper interconnect technology for the 32nm node and beyond," Proceeding of IEEE Custom Integrated Circuits Conference, CICC'09, p.141-148, Sep. 2009, San Jose, CA, USA.

[4] F. Charlet, C. Bermond, S. Putot, G. Le Carval, B. Flechet, « Extraction of (RLCG) interconnect parameters in 2D transmission lines using fast and efficient numerical tools », Proceedings of Simulation of Semiconductor Processes and Devices International Conference, p.87-89, 2000, Seattle, Washington, USA.