

Optimum Repeater Insertion to Minimize the Propagation Delay into 32nm RLC Interconnect

Denis Deschacht

► To cite this version:

Denis Deschacht. Optimum Repeater Insertion to Minimize the Propagation Delay into 32nm RLC Interconnect. EDAPS'11: The IEEE Electrical Design of Advanced Packaging & Systems, Hangzhou, China. IEEE, pp.1330-1350, 2012. <lirmm-00661461>

HAL Id: lirmm-00661461

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00661461>

Submitted on 19 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimum repeater insertion to minimize the propagation delay into 32nm RLC interconnect.

D. Deschacht

LIRMM, UMR Université Montpellier II - CNRS, CC 477, 161 rue Ada, 34 095 Montpellier cedex5 – France
Denis.Deschacht@lirmm.fr

Abstract— When high speed integrated digital circuits technology scales down from one node to the other as ITRS recommends, a significant gain is obtained on signal speed, consumption and area of CMOS transistors. Nevertheless a specific issue occurs from the 45 nm technology node. The obtained gain on active devices is foiled by an increase of interconnect propagation delays in the Back-End of Line (BEOL). This issue especially concerns relatively long (few hundred of μm) interconnects of the intermediate metal level. By introducing drivers (repeaters) in order to divide long interconnect in shorter sections and choosing optimal drivers sizes, speed can be maximized. This paper proposes a new optimal buffer sizing, and maximum length to be used for repeater networks, to optimize propagation delay for long interconnect of the 32nm technology, by taking into account, for the first time, the input transition time at each stage.

I. INTRODUCTION

According to Moore's law and ITRS recommendations [1], each new generation of integrated circuits (ICs) has to target a high level of integration. Nevertheless, from generation of the CMOS 65 nm technology node, the IC's speed increase gained on active devices is partially loosed. This is mainly due to interconnects delays increase as dimensions of interconnects are shrunk to satisfy integration requirements [2]. With technology scaling, more and more functionality is being integrated on-chip which results in an increase in the die size in spite of the reduction in minimum feature size [3]. As a result, the number of long global lines and the length of these global lines increase with technology scaling. Traditionally, repeaters are inserted into RC lines to partition an interconnect line into shorter sections [4], thereby reducing the total propagation delay. In [5] the same idea is apply to the general case of an RLC line. Their delay models are too simplistic to catch the real performance. In these papers it is considered that since the sections are each equal, the total delay can be expressed as the product of the propagation delay of a single section by the number of sections. The delay of each section is not the same, because due to the attenuation factor, the transition time increases from a section to the following one, and the propagation time is input slope dependent and also increases. An illustration will be given in part III of this paper. To neglect the impact of input transition time on the delay

causes large error in circuit analysis and can seriously compromise the validity of the optimized parameters. An accurate delay model not only reduces the number of buffers, but also the buffer size. This motivates this work to propose a new algorithm for RLC buffer insertion.

In this paper, a new methodology is developed to calculate the number and the repeater size and interconnect length which minimize the total propagation delay for RLC intermediate and global interconnect. The paper is organized as follows: in part II, we will describe the environment of the interconnection and the simulation conditions. In part III, we will show the impact of the transition time at each stage on the propagation delay. Then we will determine the length of a section which necessitates a fragmentation. Part IV describes the methodology we use to calculate the number and the buffer's size to optimize the propagation delay of long interconnect up to 600 μm . In the last part we will show the temporal and frequency gain obtained with the optimal solutions.

II. DESCRIPTION OF THE SIMULATION CONDITIONS

For this study, we consider a configuration representative of intra-level interconnects networks, and a typical geometry of an intermediate metal-layer. 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. 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 Fig. 1 and Table 1. In accordance with ITRS, the half pitch of interconnects of the intermediate level, equal to distance s between adjacent interconnects, is 50 nm while their thickness t is exceeded hundred nm for the 32 nm node technology. Electromagnetic simulations give self RLCG parameters for different width and space. In this case, the conductance G is negligible (10^{-6} mS/cm) and must not be incorporated in the model, $R=6.11$ ohms/ μm , $C=118\text{aF}/\mu\text{m}$ and $L=542\text{fH}/\mu\text{m}$. These RLC values are used to build the distributed π -RLC model (we use 40 cells) in the electrical simulations under HSPICE simulator.

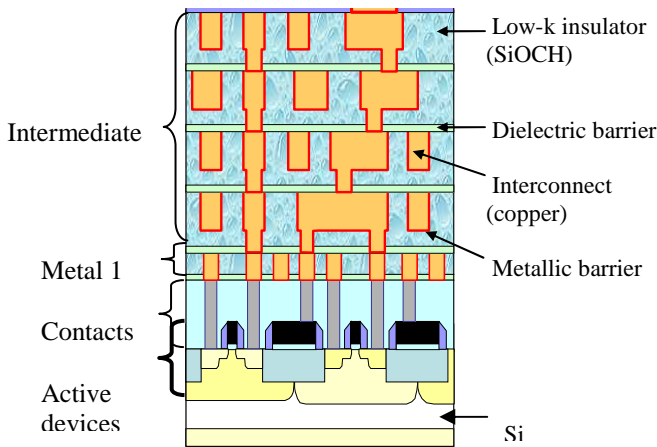


Figure 1. IC cross section illustrating hierarchical interconnects levels

TABLE I. TYPICAL 32 NM INTERCONNECT PARAMETERS [6]

BEOL	Dimensions	Materials
Intermediate Metal Level	$w = s = 50$ nm $t \approx 140$ nm	copper $\sigma \approx 35$ MS/m
Dielectric SiOCH	$h \approx 120$ nm	$\epsilon'_r \approx 2.6$ $\epsilon''_r = 0$
Metal barrier TaN/Ta	$t_{mb} \approx 7$ nm	$\sigma \approx 1,4$ MS/m
Dielectric barrier SiCN	$t_{db} \approx 40$ nm	$\epsilon'_r \approx 5$

The minimum driver size is call INVmin. Driver size is expressed in multiple of INVmin : $INVmin * x$. The driver is modelled as a linear resistor (i.e. Thevenin model), the generator with an input slope of 10ps for the first section and a load capacitance of 0.284fF, corresponding to the input capacitance of a minimum buffer, is added at the far-end of each section to minimize the transition time. This permits also to ensure proper signal polarity whatever the number of repeaters n , which is also of practical utility. At last access resistances at the interface due to contacts and inter levels vias must be taken into account. According to the reduction of their dimensions the resistance of interconnections takes an increasingly significant part in the propagation of the signals. 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. The propagation delay being input slope dependent, it is absolutely necessary to determine the delay at each stage. The increase in the transition time, from a low level to a high level for the same input signal for different length of line, will limit the periodicity of the signals to be transmitted. Consequently, fast signal transmissions (up to 5 GHz) becomes impossible on relatively long interconnects. Repeaters must be introduced to speed up signals propagation thanks to interconnect division into smaller sections.

III. DETERMINATION OF THE OPTIMAL LENGTH OF A SECTION.

Beyond critical length of interconnect sections, delays become so high that high speed signal (few GHz or more) transmission are prevented. Thereby length of interconnect

sections must be limited at few hundreds of μm . On the other hand, as few paths could reach up to around 1 mm in the BEOL, repeaters must be introduced to section path into shorter interconnect sections. But the cost linked to path cutting, including drivers insertion, contacts and vias, becomes prohibitive if interconnect sections are too short. In order to refine these assertions, this study focus on nominal lengths of interconnects sections comprised between few tens of μm and few hundreds of μm . First let us consider an interconnection of 400 μm , cut in 4 sections of 100 μm . For every section, we report in Table II, the propagation time (t_p) and the transition time (t_r) for different buffer's size. We observe that the propagation time dependent of the buffer input transition time increases from a section to the following one. The transition time also increases after every degradation due to the interconnection. A new approach is necessary.

TABLE II: ILLUSTRATION OF THE NECESSITY TO DETERMINE THE PROPAGATION DELAY AT EACH STAGE.

INVmin*x	x=2	x=4	x=6	x=8	
1 st section	t_p	12.46	8.38	6.94	6.26
	t_r	45.34	30.43	24.86	23.1
2 nd section	t_p	16.32	12.17	11.5	11.76
	t_r	72.53	47.73	40.84	39.24
3 rd section	t_p	17.47	12.73	12.1	12.35
	t_r	94.4	61.09	51.76	51.4
4 th section	t_p	18.04	12.92	12.2	12.44
	t_r	112	71.29	57.7	56.84

The effect of interconnect lengths on performance is evaluated for different driver sizes ($INVmin * 2$ up to $INVmin * 8$), and different number of sections n . We call $tp1$ the propagation time of a section of length L , $tp2$ that one of the same length but split in two sections, $tp3$ for three sections... We determine for each given dimension of buffer $INVmin * x$, the length for which we obtain : $tp2 = tp1$, $tp3 = tp2$.. and so on. We illustrate this study in this paper with an example given Fig. 2, for a buffer equal to $INVmin * 4$.

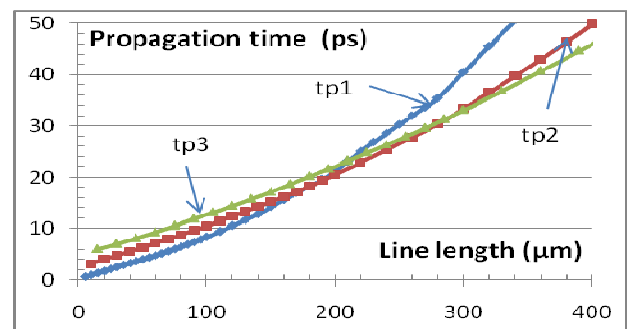


Fig. 2: Determination of the optimal length of a section

For the different buffer sizing, we have observed that the length of a section is almost constant. The results are reported Table III. $L(1 \rightarrow 2)$ means the length obtained for $tp1 = tp2$, that is the length from which a division becomes necessary.

TABLE III: LENGTH FROM WHICH A DIVISION BECOMES NECESSARY.

INVmin*x	L (1->2)	L (2->3)	L (3->4)	L (4->5)
x = 2	165	255	340	410
x = 3	180	280	375	455
x = 4	180	295	380	480
x = 5	190	310	420	510
x = 6	200	330	450	
x = 7	210	345	470	
x = 8	220	360	495	

We calculate the loading factor, defined as the total output capacitance divided by the input capacitance of the buffer :

$$F = (C_{line} + C_{INVmin}) / (C_{INVmin} * x)$$

corresponding to this average section's length. The variations are given Fig. 3. We notice that the variation of the length of an optimal section behaves with the dimension of the buffer in an almost linear way, while the loading factor (Fig. 4) decreases with the increase of the buffer's width.

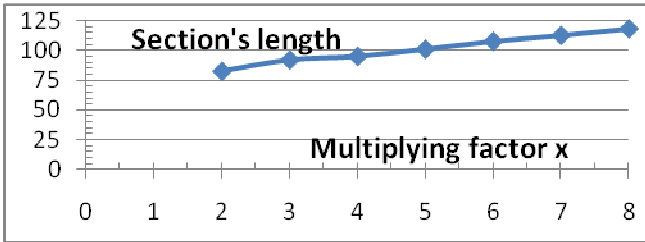


Fig. 3: Length of the optimal section versus the multiplying factor of the buffer's width.

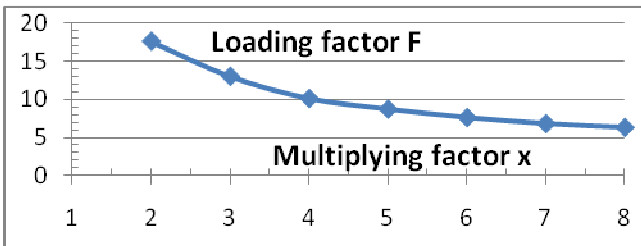


Fig. 4: Loading factor of the optimal section versus the multiplying factor of the buffer's width.

For a given length, the optimization will thus be strongly dependent on the width of the buffer.

IV. SPLIT OPTIMIZATION FOR A GIVEN LENGTH.

For every given length of line, we draw the variation of the propagation time according to the width of the buffer, for the

configuration one line, a line divided by 2, by 3 ... Examples are shown Fig.5a for L = 300 μm, and Fig.5b for L = 600 μm.

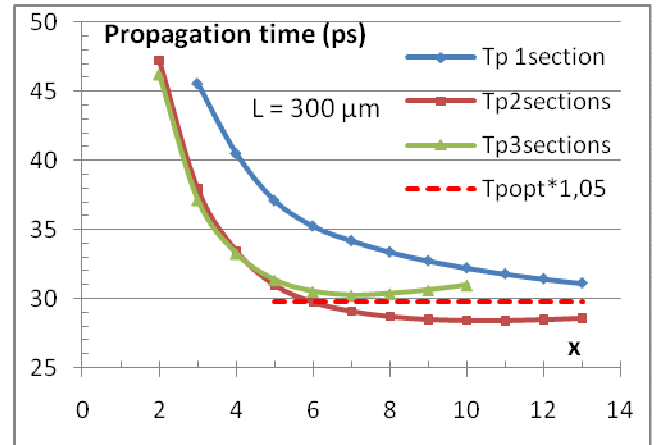


Fig. 5a: Variation of the propagation time versus the multiplying factor of the buffer's width for different configurations. L = 300 μm

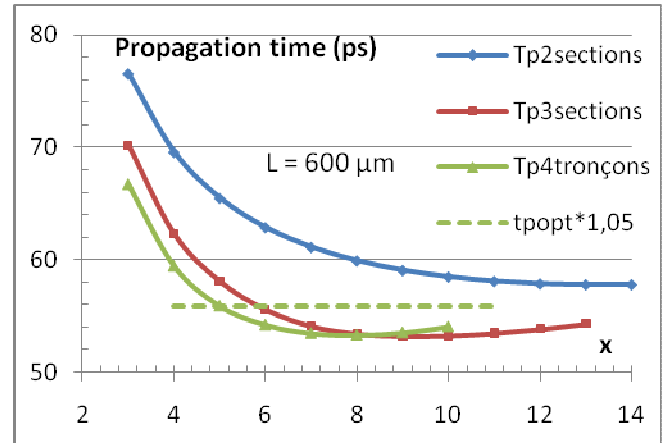


Fig. 5b: Variation of the propagation time versus the multiplying factor of the buffer's width for different configurations. L = 600 μm

The propagation time has a minimum value T_{popt} for n sections and for the multiplying factor x of the buffer's width. We observe that whatever the length of the line, the variation of the propagation time around this minimum is low, at the price of a strong increase of the buffer's size. If we accept an increase of 5% of this minimum value, we can use a much smaller buffer. On the example shown in this paper for L = 300 μm, the minimum propagation time is obtained for x=11. An increase of less than 5% allows to reduce the width of the buffer to 6 (Table 3). From 400 μm, several solutions are satisfactory. For L = 600μm, three configurations are possible. All the results are summarized Table III.

The solution satisfying our criterion of $T_{popt} * 1.05$ present an appreciably constant load factor, whatever the length is. If we take a loading factor equal to 10.5 the problem of optimization means determining the values of n and x such as:

- If $C_{line} \gg C_{INVmin}$

$$n \cdot x = (C_{line} / \mu m * L) / (C_{INVmin} * 10.5)$$

We have then to determine two integers with $n < x$ and such as the product $n \cdot x$ is the closest possible of the calculated $n \cdot x$. Results are given Table IV.

TABLE III: SECTIONS NUMBER N AND BUFFER DIMENSION X FOR DIFFERENT LENGTH.

Length(μm)	n	x	F
250	2	6	8.82
300	2	6	10.55
350	2	7	10.53
400	2	8	10.51
	3	5	11.27
450	3	6	10.55
	4	5	9.55
500	3	6	11.71
	4	5	10.59
600	3	8	10.51
	4	6	10.55
	5	5	10.17

TABLE IV: DETERMINATION OF OPTIMAL VALUES OF n AND x.

Length (μm)	nx calculated	n	x
200	7.9	2	4
250	9.89	2	5
300	11.87	2	6
350	13.85	2	7
400	15.83	2	8
450	17.81	3	6
500	19.79	4	5
600	23.74	4	6

By using the values of n and x (Table IV), we can determine the temporal gain (Fig. 6) obtained by comparing the propagation time obtained with this distribution with the propagation time for the line not split with the same dimension of buffer. We showed in this paper that the electrical parameters of the interconnections imposed an important degradation of the transition time at the output, therefore limited the frequency of the signals to be transmitted. With our optimization technique, the frequency performances are strongly improved (Fig. 6).

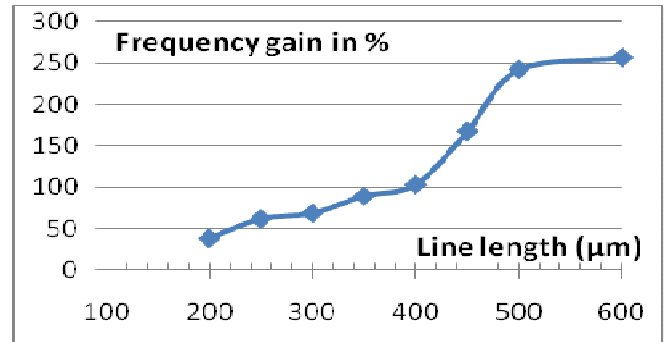
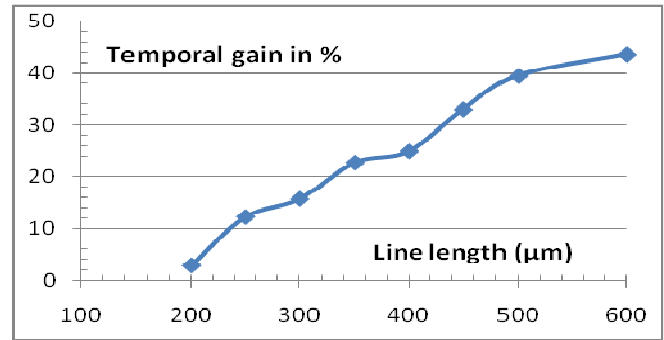


Fig. 6: Temporal and frequency gain in percent obtained with the optimal values of n and x.

V. CONCLUSION

It has been shown that very high speed circuits of the 32nm generation require both short sections (around 150 μm) and relatively large drivers. The optimal section number and driver's size have been determined with a new methodology including the impact of the transition time at each stage. A simple analytical equation gives the number of repeaters and the size of the driver that must be used to optimize the propagation delay. This new optimization technique improves considerably the performances and makes possible the use of interconnections of 50nm in the upper and intermediate metal levels.

REFERENCES

- [1] ITRS International Technology Roadmap for Semiconductors, <http://public.itrs.net> (2009).
- [2] 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, Sep. 2009, San Jose, CA, US, p.141-148
- [3] V. De and S. Borkar, "Technology and design challenges for low power and high performance" in Proc. Int. Symp. Low power Electronics and Design, 1999, pp.163-168.
- [4] H. Shah, P. Shiu, B. Bell, M. Aldredge, N. Sopory and J. Davis, "Repeater Insertion and Wire Sizing Optimization for Throughput-Centric VLSI Global Interconnects", Proc. ICCAD, San Jose, USA, 2002, pp. 280-284.
- [5] Y. Ismail and E. Friedman, "Effects of Inductance on the Propagation Delay and Repeater Insertion in VLSI Circuits", IEEE Transactions on VLSI Systems, Vol. 8, no. 2, April 2000, pp.195-206.