Vol 18 No 3, March 2009
1674-1056/2009/18(03)/1188-06
Chinese Physics B
© 2009 Chin. Phys. Soc.
and IOP Publishing Ltd

A novel interconnect-optimal repeater insertion model with target delay constraint in 65 nm CMOS*

Zhu Zhang-Ming(朱樟明), Qian Li-Bo(钱利波), and Yang Yin-Tang(杨银堂)

Institute of Microelectronics, Xidian University, Xi'an 710071, China

(Received 17 June 2008; revised manuscript received 4 September 2008)

Repeater optimization is the key for SOC (System on Chip) interconnect delay design. This paper proposes a novel optimal model for minimizing power and area overhead of repeaters while meeting the target performance of on-chip interconnect lines. It also presents Lagrangian function to find the number of repeaters and their sizes required for minimizing area and power overhead with target delay constraint. Based on the 65 nanometre CMOS technology, the computed results of the intermediate and global lines show that the proposed model can significantly reduce area and power of interconnected lines, and the better performance will be achieved with the longer line. The results compared with the reference paper demonstrate the validity of this model. It can be integrated into repeater design methodology and CAD (computer aided design) tool for interconnect planning in nanometre SOC.

Keywords: distributed RLC, interconnect power dissipation and area, target delay, lagrangian function

PACC: 6185, 6630Q, 8160C

1. Introduction

As very large-scale integration circuits continue to be scaled aggressively into nanometre technology node, the operating frequency and silicon area increase, and interconnected delay and power become increasingly critical performance bottleneck. Repeater insertion is widely used as a classical solution to decrease delay, which can prevent the capacitive accumulative effect by renewing the potential of interconnect lines, change the delay dependence on the length from quadratic to linear.[1−13] However, along with the increase of the number of long global lines and the length of these global lines with technology scaling, the number and the size of such repeaters will increase correspondingly. According to the report of International Technology Roadmap for Semiconductors 2006, the number of repeaters on-chip will be more than 800 thousands, leading to excessive power dissipation and wastage of silicon area. Hence, how to minimize the power and area while meeting the required performance becomes a concerned issue.

The number and the size of repeaters are presented in Ref.[3] to minimize the interconnected delay by derivative in RC (Resistor–Capacitor) model. The minimum delay–power product is obtained by determining the interconnected lines width in Ref.[5]. In Refs.[2] and [4], silicon area is taken into account to achieve the tradeoff between delay, power dissipation and area. These papers are based on Elmore model,[2−4] inductive effect is not included. The importance of on-chip inductance is continuously increasing with higher operating frequencies and technology scaling trends, especially in clock distribution networks and in upper metal layers, these wires are low resistance lines which can exhibit significant inductive effects,[9] so RLC (Resistor–Capacitor–Inductance) interconnected model has to be introduced. Additionally, leakage power becomes a significant part of total interconnected power dissipation.

In this paper, we propose a novel interconnect-optimal methodology based on RLC model, which minimizes the total interconnected power dissipation and silicon area while meeting given target delay. This paper is organized as follows. In Section 2, we formulate and represent the expressions of delay, power dissipation and area based on RLC model. In Section 3, we integrate power and area into repeater design methodology along with time delay, develop a novel optimum model based on Lagrangian function. Re-

*Project supported by the National Natural Science Foundation of the China (Grant Nos 60676009 and 60776034), the Doctor Foundation of Ministry of Education of China (Grant No 20050701015), and the National Outstanding Young Scientist Foundation of China (Grant No 60725415).

sults of comparing power and area saved for different target delay for various interconnected lines in 65 nm CMOS technology are presented in Section 4.

2. RLC interconnect delay–power–area

2.1. Delay

As shown in Fig.1, a distributed RLC interconnected with length h is evenly divided into n segments with uniformly sized repeaters. r, l and c are respectively, the resistance per unit length, the inductance per unit length, and the capacitance per unit length. The repeater is assumed to be implemented as a CMOS inverter. The output resistance and input capacitance of a repeater are R and C, which can be given by Ref.[14]:

$$R = \frac{k_1}{W_{\mathrm{N}}}, \tag{1}$$

$$C = k_2 W_{\mathrm{N}}. \tag{2}$$

Where W_{N} is the channel width of the NMOS transistor in the repeater. k_1 and k_2 are process dependent constants.

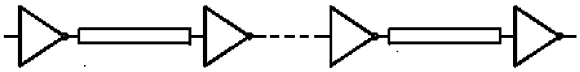


Fig.1. Repeaters inserted in a RLC interconnection.

The time delay for RLC line with repeaters inserted consists of two parts: the time for the signal propagation through the interconnected line, which can be obtained by transfer function of RLC model; and the time to charge up the effective capacitance of repeaters, varied with the range of voltage. When the interconnected circuit as depicted in Fig.1 is excited by step input waveform, the 50% time delay for RLC line is given by Refs.[7] and [8]:

$$t_{\mathrm{rep-rlc}} = \ln 2 \left(\frac{l \cdot n}{r} + \frac{7rch^2}{12n} \right) + 0.693 \left(k_2 W_{\mathrm{N}} rh + k_1 k_2 n + \frac{k_1 ch}{W_{\mathrm{N}}} \right). \tag{3}$$

Setting the derivative of time delay with respect to n and W_{N} equal to 0, gives the optimum number and the size of repeaters:

$$n_{\mathrm{opt}} = \sqrt{\frac{0.404rch^2}{0.693\left(k_1 k_2 + \frac{l}{r}\right)}}, \tag{4}$$

$$W_{\mathrm{N(opt)}} = \sqrt{\frac{k_1 c}{k_2 r}}. \tag{5}$$

Substituting these values in Eqs.(4) and (5) into Eq.(3) we obtain the optimum time delay as

$$t_{\mathrm{rep-rlc}} = \ln 2 \left(\frac{l \cdot n_{\mathrm{opt}}}{r} + \frac{7rch^2}{12n_{\mathrm{opt}}} \right) + 0.693 \left(k_2 W_{\mathrm{N(opt)}} rh + k_1 k_2 n + \frac{k_1 ch}{W_{\mathrm{N(opt)}}} \right). \tag{6}$$

2.2. Power dissipation

It can be interpreted from Eq.(7) that the total power dissipation of interconnection consists of three parts: a) switching power P_{Switch} induced by the charge up and discharge of capacitance, which is the dominant component of power dissipation in typical CMOS circuit; b) leakage power P_{Leakage} caused by leakage current flowing through transistor, which increases exponentially with device scaling and becomes the significant part in nanometre technology;[3] and c) short-circuit power $P_{\mathrm{short\text{-}circuit}}$ incurred when the NMOS and PMOS transistors in an inverter are simultaneously ON. It can be negligible by designing carefully, so it is not considered in this paper. The power dissipation of RLC interconnected line shown in Fig.1 is as follows:

$$P_{\mathrm{repeater}} = P_{\mathrm{switch}} + P_{\mathrm{leakage}} + P_{\mathrm{short\text{-}circuit}}. \tag{7}$$

The switching power of interconnection is given by Ref.[14],

$$P_{\mathrm{switch}} = \alpha(k_2 wn + ch)V_{\mathrm{dd}}^2 f. \tag{8}$$

Where V_{dd} is the power supply voltage, f clock frequency, α switch factor, which is the proportion of repeaters on a chip that are switched in an average clock cycle, taken as 0.15[6] in the paper.

The average leakage power of repeaters in RLC interconnections is given by

$$\begin{aligned} P_{\mathrm{leakage}} &= V_{\mathrm{dd}} I_{\mathrm{leakage}} n \\ &= \frac{1}{2} V_{\mathrm{dd}} \left(I_{\mathrm{off(N)}} W_{\mathrm{N}} n + I_{\mathrm{off(P)}} W_{\mathrm{P}} n \right) \\ &= \frac{3}{2} V_{\mathrm{dd}} I_{\mathrm{off(N)}} W_{\mathrm{N}} n. \end{aligned} \tag{9}$$

Where I_{leakage} is the leakage current flowing through the repeater, $I_{\mathrm{off(N)}}(I_{\mathrm{off(P)}})$ is the leakage current per NMOS (PMOS) transistor.

The factor 1/2 is included because in a long buffered RLC interconnection, on an average, half the inverters will have input of one. Usually $I_{\text{off(N)}} \approx I_{\text{off(P)}}$ and the channel width of the PMOS transistor is two to three times larger than the channel width of NMOS device in an inverter. As $W_{\text{P}} = 2W_{\text{N}}$ can achieve the similar symmetrical effect with $W_{\text{P}} = 3W_{\text{N}}$, and takes less silicon area, so in this paper, we assume $I_{\text{off(N)}} \approx I_{\text{off(P)}}$, and $W_{\text{P}} = 2W_{\text{N}}$.

2.3. Area

Based on nanometre technology, there are a large number of optimal repeaters with n_{opt} and $W_{\text{N(opt)}}$ inserted into interconnected lines, which take up significant fraction of silicon, severely affect the system structure and layout, and have been an un-negligible factor in circuit design. The channel length of MOS transistor is usually assumed the minimum size with the consideration of area in digital circuit, and the characteristic of transistor depends on the channel width parameter W_{N}. The area of repeaters inserted can be simplified as

$$A_{\text{rep}} = 3nW_{\text{N}}. \tag{10}$$

The ITRS 2006 technology and equivalent circuit model parameters for different technology nodes are shown in Table 1.[15] Note that the operating frequency and leakage current increase rapidly with technology scaling.

Table 1. Technology and equivalent circuit model parameters for different technology nodes based on the ITRS 2006.

Tech. Node/nm	90	65	45
k_1/mΩ	8.14×10^{-4}	6.24×10^{-4}	5.94×10^{-4}
k_2/(F/m)	1.22×10^{-8}	1.58×10^{-8}	2.00×10^{-8}
V_{dd}/V	1	0.7	0.6
$I_{\text{off(N)}}/(\mu\text{A}/\mu\text{m})$	3.56	13.4	23.83
f_{clk}/GHz	2.99	6.73	11.51

3. Optimum model of repeater insertion

The minimum time delay of RLC interconnected line t_{opt} can be obtained by repeaters inserted with n_{opt} and $W_{\text{N(opt)}}$ as shown in Eq.(3). However, in most cases, the required target delay $t_{\text{target}} < t_{\text{opt}}$, which indicates that interconnected lines are not performed critically. If we still place optimal repeaters to over-drive the interconnected lines, it could waste silicon area and increase power dissipation. How to find a repeater insertion model which meets the target delay while minimizing the power and area overhead cost is significant. We observe that the power dissipation in Eq.(7) and the area in Eq.(10) are directly proportional to the number n and the size W_{N} of repeaters, and that the time delay is a function with n and W_{N}, then Lagrangian function can be used to satisfy the above goal,

$$F = XA_{\text{rep}} + YP_{\text{rep}} + \lambda(t_{\text{rlc}-\text{rep}} - t_{\text{target}}), \tag{11}$$

where F is a Lagrangian function. X, Y are the weights for area and power overhead respectively, which can be varied to reflect the importance of minimizing one overhead over the other.

Substituting Eqs.(7) and (10) into Eq.(11) we obtain a new expression for F:

$$\begin{aligned} F =& X(3nW_{\text{N}}) + Y(\alpha(k_2W_{\text{N}}n + ch)V_{\text{dd}}^2 f + \frac{3}{2}V_{\text{dd}}I_{\text{off(N)}}W_{\text{N}}n) \\ &+ \lambda\left(\ln 2\left(\frac{l\cdot n}{r} + \frac{7rch^2}{12n}\right) + 0.693\left(k_2W_{\text{N}}rh + k_1k_2n + \frac{k_1ch}{W_{\text{N}}}\right) - t_{\text{target}}\right). \end{aligned} \tag{12}$$

From Lagrange's method, setting the derivative of F with respect to n and W_{N} equal to 0, $\dfrac{\partial F}{\partial n} = 0$ and

$\frac{\partial F}{\partial W_{\mathrm{N}}} = 0$, then we obtain:

$$W_{\mathrm{N}}\left(3X + \alpha Y k_1 V_{\mathrm{dd}}^2 f + \frac{3}{2} V_{\mathrm{dd}} I_{\mathrm{off(N)}}\right) + \lambda\left[0.693\left(\frac{l}{r} + k_1 k_2 - \frac{7rch^2}{12n^2}\right)\right] = 0, \tag{13}$$

$$n\left(3X + \alpha Y k_1 V_{\mathrm{dd}}^2 f + \frac{3}{2} V_{\mathrm{dd}} I_{\mathrm{off(N)}}\right) + \lambda\left[0.693\left(k_2 rh - \frac{k_1 ch}{W_{\mathrm{N}}^2}\right)\right] = 0. \tag{14}$$

Eliminating λ from Eqs.(13) and (14), and substituting the value of target delay into Eq.(3), we obtain

$$\frac{rch^2}{n} = \frac{6}{7}(t'_{\mathrm{target}} - 2k_4 W_{\mathrm{N}}), \tag{15}$$

$$k_5 n - \frac{7rch^2}{12n} = k_4 W_{\mathrm{N}} - \frac{k_3}{W_{\mathrm{N}}}. \tag{16}$$

Where $t'_{\mathrm{target}} = t_{\mathrm{target}}/0.693$, $k_3 = k_1 ch$, $k_4 = k_2 rh$, $k_5 = (l/r) + k_1 k_2$. Solving the above equations, we can calculate the optimum number and width of repeaters with target delay constraint,

$$w = \frac{\left(2k_3k_4 + \frac{t'^2_{\mathrm{target}}}{2} - \frac{7}{6}rch^2k_5\right) - \sqrt{\left(2k_3k_4 + \frac{t'^2_{\mathrm{target}}}{2} - \frac{7}{6}rch^2k_5\right)^2 - 4k_3k_4t'^2_{\mathrm{target}}}}{2k_4t'_{\mathrm{target}}}, \tag{17}$$

$$n = \frac{7rch^2}{6\left[t'_{\mathrm{target}} - \frac{\left(2k_3k_4 + \frac{t'^2_{\mathrm{target}}}{2} - \frac{7}{6}rch^2k_5\right) - \sqrt{\left(2k_3k_4 + \frac{t'^2_{\mathrm{target}}}{2} - \frac{7}{6}rch^2k_5\right)^2 - 4k_3k_4t'^2_{\mathrm{target}}}}{t'_{\mathrm{target}}}\right]}. \tag{18}$$

4. Simulation results

Using the RLC model outlined above, we obtain the analytical calculated results of optimum number and size of repeaters inserted for two different types of interconnected lines in 65 nm CMOS technology, which are presented in Table 2. The ITRS technological parameters required in calculation are given in Table 1. r, l and c are varied with the type of interconnected lines, the corresponding values are obtained by 2.5D Field Solves in Berkeley Predictive Technology Model (BPTM). From Table 2, it is noted that the number of repeaters and their size in intermediate interconnected line decrease with the increase of target delay, which results in the power dissipation and area degrade significantly. The same phenomena can also be observed in global interconnected line.

Table 2. Power and area of repeaters for two types of interconnection in 65 nm CMOS.

Tech. parameter	t_{target}/ps	t_{opt}/ps	n	W_{N}/μm	$A_{\mathrm{rep}}/\mu\mathrm{m}^2$	P_{rep}/μW
intermediate	802	802	19	2.10	119.70	1367.1
h=2 cm	810	802	16	1.82	87.55	1132.6
r=448.9 Ω/mm	830	802	14	1.70	76.68	1053.3
l=2.36 nH/mm	880	802	12	1.37	49.24	853.1
c=49.93 fF/mm	970	802	10	1.13	33.78	740.3
global	2773.8	2773.8	20	10.25	614.82	9843.7
h=10 cm	2800	2773.8	18	8.37	451.95	8655.5
r=40.74 Ω/mm	2850	2773.8	17	7.33	373.61	8312.6
l=2.44 nH/mm	2950	2773.8	15	6.20	320.71	7698.1
c=108.33 fF/mm	3000	2773.8	14	5.82	279.19	7395.2

The area and power dissipation of repeaters are plotted against the target delay for two different types of interconnected lines in 65 nm CMOS technology in Figs.2 and 3. It is easily deduced that area and power decline while the target delay increases, the descending slope of global line is obviously steeper than intermediate line, which indicates that the area and power of global line decrease rapidly in comparison with the intermediate line with delay increase. It is caused by the result that the cardinal number of the number n and the size w of repeaters inserted in global line are greater than those in intermediate line. Moreover, the results also illuminate that the RLC model above is more suitable for the global line, and the better performance will be obtained with the longer lines.

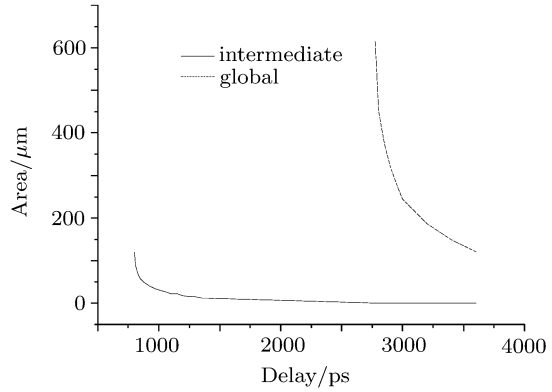


Fig.2. Area versus target delay for two types interconnected lines in 65 nm CMOS.

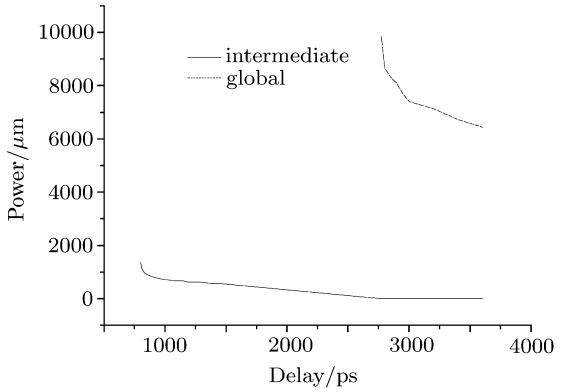


Fig.3. Power versus target delay for two types interconnected lines in 65 nm CMOS.

In order to verify the validity, the proposed model is compared with the Ref.[6] based on global line with a length of 10 centimetres in 65 nanometre CMOS technology. These two models adopt same process parameters, the only difference lies on the inductive effect which is not considered in Ref.[6], the related model parameters and the results of comparison can be seen in Table 3. Where $t_{\rm opt}/t_{\rm target}$ is defined as the normalized performance ratio of interconnected line, γ is the ratio of repeater area in the target delay case to that in the minimum delay case. $P_{\rm rep}/P_{\rm opt}$ is the ratio of interconnected power in the target delay case to that in the minimum delay case. Clearly, the area of repeaters and their power decrease simultaneously with the performance attenuation, but our new repeater insertion solution gives a slightly better performance with less area ratio and less power ratio than those in Ref.[6] under same performance ratio. The direct effectiveness of these two models is displayed in Figs.4 and 5.

Table 3. Comparison results for global line in 65 nm CMOS.

process parameters	$t_{\rm opt}/t_{\rm target}$	γ [proposed]	$\gamma^{[6]}$	$P_{\rm rep}/P_{\rm opt}$ [proposed]	$P_{\rm rep}/P_{\rm opt}^{[6]}$
r=40.74 Ω/mm	1	1	1	1	1
l=2.44 nH/mm	0.9906	0.7351	0.7575	0.8793	0.8863
c=108.33 fF/mm	0.9733	0.6077	0.6232	0.8445	0.8520
k_1=6.24×10^{-4} mΩ	0.9403	0.4541	0.4883	0.8212	0.8374
k_2=21.58×10^{-8} F/m	0.9246	0.3978	0.4444	0.7820	0.7966
$V_{\rm dd}$=0.7 V	0.8668	0.3039	0.3308	0.7513	0.7623
$I_{\rm off(N)}$=13.4 μA/μm	0.7299	0.2426	0.2637	0.7256	0.7311
	0.6935	0.1723	0.1851	0.6828	0.6913
$f_{\rm clk}$=6.73 GHz	0.6164	0.1400	0.1598	0.6549	0.6692

Figure 4 shows the performance ratio of interconnected line along the horizontal axis and the area ratio along the vertical axis. Figure 5 shows the performance ratio along the horizontal axis and the power ratio along the vertical axis. The graphs demonstrate that our model is more efficient than that in Ref.[6], and can be used for the RLC interconnected model compared to Ref.[6], which is limited to the RC model.

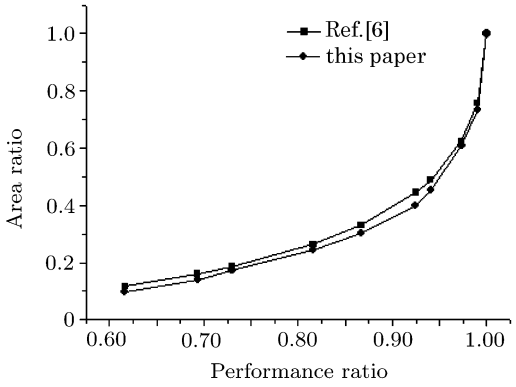


Fig.4. Area ratio versus performance ratio for global line in 65 nm CMOS.

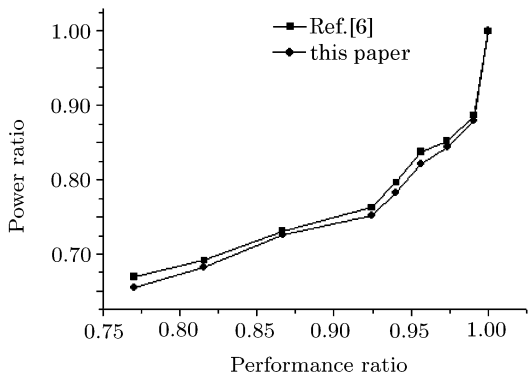


Fig.5. Power ratio versus performance ratio for global line in 65 nm CMOS.

5. Conclusion

In this paper, we develop a novel RLC model to choose the optimum repeaters' number and their size for interconnected lines, which meets the target delay and minimizes the power dissipation and area of repeaters. Using this model, we compute and compare the results for two different types of interconnected lines in 65 nm CMOS technology. It is found that our model can reduce the power and area significantly, and is favourable to optimize interconnected line with target delay constraint. Besides, the better performance will be obtained with longer lines. The comparative results with Ref.[6] demonstrate the validity of this model, and it can be integrated into repeater design methodology and CAD tool for interconnected planning in nanometre SOC.

References

[1] Chen G Q and Friedman E G 2005 *Proc.IEEE Int. Symp. Circuits Syst.* (Kobe, Japan) p596
[2] Nalamalpu A and Burleson W 2001 *IEEE Int. ASIC/SOC Conf.* (Arlington VA) p152
[3] Banerjee K and Mehrotra A 2002 *IEEE Trans. Electron Devices* **49** 2001
[4] Tang M and Mao J F 2006 *Proc. 19th Int. Conf. on VLSI* (Hyderabad, India) p1063
[5] Magdy A and EI-Moursy 2003 *Proc. 13th ACM Great Lakes Symposium on VLSI* (Washington, DC) p176
[6] Garcea G S 2003 *Proc. 29th European Solid-State Circuits Conference* (Estoril, Portugal) p591
[7] Venkatesan R 2003 *IEEE Trans. Electron Devices* **50** 1094
[8] Yu Q J and Kuh E S 2001 *Proc. IEEE* **89** 772
[9] Ismail Y I and Friedman E G 2000 *IEEE Trans. on VLSI* **8** 195
[10] Morgenshtein A and Cidon I 2005 *Proc. IEEE Int. Symp. Circuit Syst.* (Kobe, Japan) p600
[11] Zhang Y Q, Hong X L and Cai Y C 2005 *Chin. J. Electronics* **33** 783 (in Chinese)
[12] Hanchate N and Ranganathan N 2006 *IEEE Trans. Computers* **55** 1011
[13] Liu Z P and Bian J N 2007 *IEEE Computer Society Annual Symposium on VLSI* (Porto Alegre, Brazil) p283
[14] Rabaey J M 2003 *Digital Intergrated Circuits: A Design Perspective* Second Edition (London: Prentice-Hall)
[15] Semiconductor Industry Association 2006 *International Technology Roadmap for Semiconductors*