MCM Interconnect Design Using Two-Pole Approximation

Jianhua Shao and Richard M M Chen
Department of Electronic Engineering
City University of Hong Kong
83 Tat Chee Avenue, Kowloon, Hong Kong
Email: jhshao@ee.cityu.edu.hk

Abstract

In this paper, an optimization scheme is proposed for interconnect design with wire width and series resistance being design variables. Due to the distributed nature of interconnects, poles of such systems are transcendental and infinite in number. First, a two-pole approximation is used to capture the system behavior. Lower-order moments are employed to obtain two approximate dominant poles. Then, the two parameters, damping ratio and natural undamped frequency, are expressed as functions of the two dominant poles. Since the output response is characterized by the two parameters, the parameters are used to define the objective function and constraints, which form a constrained multivariable nonlinear optimization problem. After that, the optimization problem is solved using gradient projection method. One advantage of our approach is the ability to explicitly control the maximum overshoot of the observation points. Two numerical examples are given.

1. Introduction

As system operating frequencies increase, interconnect becomes the dominant factor in determining system performance. Signal delay is largely affected by interconnects than by gates. It has been shown that interconnects of MCM behave like RLC transmission lines [1].

Since inductance makes peaking and oscillations possible, improperly designed interconnects can cause false switching and affect system reliability. Intuition-based approach, which uses time domain simulation while changing line parameters, can be very time-consuming and not promising, unless some effective guidance is provided.

One approach for systematic interconnect design is to approximate the system behavior by poles and zeros. However, due to the distributed nature of interconnects, poles of such systems are transcendental and infinite in number. Previous works have shown the usefulness of two-pole approximation for capturing the system behavior [2] [3] [4] [5] [6] [7].

While previous works use lumped elements to approximate the interconnects and/or have no explicit control of the overshoot of the response, we calculate the two dominant poles using lower-order moments, which are obtained exactly by utilizing the method proposed by Yu and Kuh [8], and explicitly control the overshoot. The output response is characterized by two parameters, and the objective function and constraints, consisting of these two parameters, form a constrained multivariable nonlinear optimization problem.

Deep overdamping results in a slow response and deep underdamping gives rise to a response with much overshoot and a long settling time. Both cases should be avoided in high performance system design. Slightly underdamped design is good to system performance which results in shorter signal delay with small amount of overshoot. For this reason, we design the interconnects to be underdamped, if possible.

In Section 2, the objective function and constraints of the constrained multivariable optimization problem are derived. In Section 3, two design examples are presented. Conclusion is made in Section 4.

2. Formulation of the Optimization Problem

2.1. Two-Pole Approximation

Moments of a time-domain waveform, \( v(t) \), are defined via the Laplace transformation of the waveform as follows:

\[
V(s) = \int_0^\infty e^{-st} v(t) dt = m_0 + m_1 s + m_2 s^2 + \cdots
\]

where \( m_k, k = 0, 1, 2, \cdots \), are the Maclaurin series coeffi-
system behavior, the transfer function is written as:

\[ m_k' = \frac{(-1)^k}{k!} \int_0^\infty t^k v(t) \, dt \]  

(2)

Since interconnects are modeled as transmission line in high performance systems, input signals undergo a pure delay \( \tau = \sum_{i=1}^n d_i \sqrt{L_i C_i} \) to arrive at the output end, where \( n \) is the number of lines on the signal path, \( L_i \) and \( C_i \) are the inductance and capacitance of the \( i \)th line per unit length respectively, and \( d_i \) is the length of the corresponding line.

When rational function is used to approximate the delay in frequency domain, 20 or more poles may be needed [9]. In order to approximate the system behavior better, the pure delay must be excluded from the moments. The new moments have the following relation with original moments:

\[ (m_0 + m_1 s + \cdots) e^{-s\tau} = m'_0 + m'_1 s + \cdots \]  

(3)

which results:

\[
\begin{align*}
m_0 &= m'_0 \\
m_1 &= m'_1 + m'_0 \tau \\
m_2 &= m'_2 + m'_1 \tau + \frac{1}{2} m'_0 \tau^2 \\
&\vdots
\end{align*}
\]

(4)

When two-pole approximation is used to capture the system behavior, the transfer function is written as:

\[ H(s) = \frac{k_1}{s - p_1} + \frac{k_2}{s - p_2} \]  

(5)

where \( p_1, p_2 \) are the two approximate poles and \( k_1, k_2 \) are the corresponding residues. When the system is underdamped, the two poles and residues are conjugate respectively.

Fig. 1 shows a series-terminated interconnection line, driving a capacitive load. In the following sections, we assume that all the loads are capacitive.

\[ \begin{array}{ccc}
R & & \text{V}_{\text{out}} \\
\text{V}_{\text{in}} & & \\
& & C_1
\end{array} \]

Fig. 1 A series-terminated interconnect line.

Since \( h|_{t=0} = 0 \) and the loads are capacitive, the dc gains are unity, we have:

\[
\begin{align*}
k_1 + k_2 &= 0 \\
\frac{k_1}{p_1} - \frac{k_2}{p_2} &= 1
\end{align*}
\]

(6) (7)

While it is possible to obtain the dominant poles by Padé approximation, however, due to its instability, we use a new approach [10]. Let \( p_1 \) and \( p_2 \) be the two approximate poles that we want to find and assume that the two poles are complex conjugate pair. Representing the poles using polar coordinates, we have \( p_1 = re^{i\theta} \) and \( p_2 = re^{-i\theta} \). As a result, the ratios of successive moments becomes:

\[
\frac{m_k}{m_{k+1}} = \frac{1 - e^{i(2k+2)\theta}}{1 - e^{i(2k+4)\theta}}
\]

(8)

Eliminating \( p_1 \) using \( \frac{m_k}{m_{k+1}} \) and \( \frac{m_k}{m_{k+2}} \), we can solve for \( e^{i\theta} \):

\[
\frac{m_k m_{k+2}}{m_{k+1} m_{k+3}} = \frac{(1 - e^{i(2k+2)\theta})(1 - e^{i(2k+6)\theta})}{(1 - e^{i(2k+4)\theta})(1 - e^{i(2k+8)\theta})}
\]

(9)

Then, the magnitude \( r \) can be obtained using (8). Thus, we get both \( p_1 \) and \( p_2 \). We take \( k = 4 \) in our computation.

When the line parameters, resistance and capacitance, are dominant, the ratios of successive moments will converge, the dominant poles are real and the above approach will not work. For this case, we can use the method proposed by Tutuianu instead [11]. The two approximate poles as functions of the moments are:

\[
\begin{align*}
p_1 &= \frac{m_k}{m_{k+1}} \\
p_2 &= \text{polynomial of } \frac{m_k}{m_{k+1}} \frac{m_k}{m_{k+2}} \frac{m_k}{m_{k+3}}
\end{align*}
\]

(10) (11)

We use \( k = 4 \) in (10) and \( k = 3 \) in (11).

Using (6) and (7), we can also write \( H(s) \) as:

\[
H(s) = \frac{\omega_n^2}{s^2 + 2\zeta \omega_n s + \omega_n^2}
\]

(12)

where \( \omega_n \) is the natural undamped frequency and \( \zeta \) is the damping ratio [12]:

\[
\omega_n = \sqrt{p_1 p_2} \\
\zeta = \frac{1}{2} \frac{p_1 + p_2}{\sqrt{p_1 p_2}}
\]

(13)

The damping condition is controlled by \( \zeta \). 0 < \( \zeta < 1 \), \( \zeta = 1 \), and \( \zeta > 1 \) correspond to underdamped, critically damped, and overdamped responses respectively. Next, we discuss the three situations in detail.

i) 0 < \( \zeta < 1 \):

The step response is obtained as:

\[
v(t) = 1 - \frac{1 - e^{-\zeta \omega_n t}}{\sqrt{1 - \zeta^2}} \sin(\omega_n \sqrt{1 - \zeta^2} t + \theta)
\]

(14)

where \( \theta = \cos^{-1} \zeta \).
The first overshoot of the response occurs at:

$$t_{max} = \frac{\pi}{\omega_n \sqrt{1 - \zeta^2}}$$  \hspace{1cm} (15)$$

and the corresponding \( v(t_{max}) \) is

$$v(t_{max}) = 1 + e^{-\frac{\omega_nt}{\sqrt{1 - \zeta^2}}}$$  \hspace{1cm} (16)$$

The percent overshoot is

$$PO = e^{-\frac{\omega_n}{\sqrt{1 - \zeta^2}}}$$  \hspace{1cm} (17)$$

and we can see that the amount of overshoot is only affected by \( \zeta \).

Considering the partial derivatives of \( v(t) \) with respect to \( \omega_n \) and \( \zeta \), we have:

$$\frac{\partial v}{\partial \omega_n} = \frac{e^{-\omega_n t}(\sin(\omega_n \sqrt{1 - \zeta^2} t) - \zeta)}{1 - \zeta^2}$$  \hspace{1cm} (18)$$

which is greater than 0 for \( 0 < t < t_{max} \), and, because of \( \tan(x) > x \) for \( 0 < x < \frac{\pi}{2} \),

$$\frac{\partial v}{\partial \zeta} = \frac{e^{-\omega_n t}(\cos\alpha - \sin\alpha)}{(1 - \zeta^2)^{\frac{3}{2}}}$$  \hspace{1cm} (19)$$

which is less than 0 for \( 0 < t < t_{max} \), where \( \alpha = \omega_n \sqrt{1 - \zeta^2} t \). Hence, larger \( \omega_n \) and smaller \( \zeta \) result in faster response, and consequently, shorter delay time.

Observing \( t_{max} \), it has the same characteristics with respect to \( \omega_n \) and \( \zeta \). Since the output response increases monotonically during \( 0 < t < t_{max} \), we can deduce that smaller \( t_{max} \) leads to shorter delay with \( \zeta \) changing in small range.

\( ii) \) \( \zeta = 1 \):

The step response is written as:

$$v(t) = 1 - e^{-\omega_n t}(1 - \omega_n t)$$  \hspace{1cm} (20)$$

and its derivative with respect to \( \omega_n \) is

$$\frac{dv}{d\omega_n} = \omega_n^2 e^{-\omega_n t} > 0, \text{ for } t > 0$$  \hspace{1cm} (21)$$

Therefore, larger \( \omega_n \) leads to shorter delay time.

\( iii) \) \( \zeta > 1 \):

The step response is obtained as:

$$v(t) = 1 - e^{-\omega_n t}\left( (\frac{\zeta^2 - 1}{\zeta^2 - 1})(cosh(\omega_n \sqrt{\zeta^2 - 1} t) + \sqrt{\zeta^2 - 1} \sinh(\omega_n \sqrt{\zeta^2 - 1} t)) \right)$$  \hspace{1cm} (22)$$

The partial derivatives of \( v(t) \) with respect to \( \omega_n \) and \( \zeta \) are

$$\frac{\partial v}{\partial \omega_n} = \frac{e^{-\omega_n t} \sinh(\omega_n \sqrt{\zeta^2 - 1} t)}{\sqrt{\zeta^2 - 1}}$$  \hspace{1cm} (23)$$

which is greater than 0 for \( t > 0 \), and, because of \( tanh(x) < x \) for \( x > 0 \),

$$\frac{\partial v}{\partial \zeta} = \frac{e^{-\omega_n t}(\sinh(\omega_n \sqrt{\zeta^2 - 1} t) - \omega_n t \cosh(\omega_n \sqrt{\zeta^2 - 1} t))}{\zeta^2 - 1 - \zeta^2}$$  \hspace{1cm} (24)$$

Hence, larger \( \omega_n \) and smaller \( \zeta \) result in shorter delay time.

### 2.2. Objective Function and Constraints

Generally, the line parameters, resistance, inducance, and capacitance are functions of the line width. For an interconnect network, the design objective is to determine the set of line widths and the series resistances with respect to loading conditions so that the output response has short delay.

Based on the discussion of last section, we formulate our objective function for a single output as:

$$q = \begin{cases} \frac{1}{\omega_n \sqrt{1 - \zeta}^2}, & 0 < \zeta < 1 \\ \frac{1}{\omega_n |\lambda - \zeta|}, & \zeta \geq 1 \end{cases}$$

where \( \zeta_u \) is the upper bound of \( \zeta \) and \( \lambda \) is a cost factor to enlarge the costs of critically damped and overdamped cases.

If there are \( N \) observation points, we form a cost vector:

$$Q = [q^{(1)} q^{(2)} q^{(3)} \ldots q^{(N)}]$$  \hspace{1cm} (25)$$

and the cost function is defined as:

$$P = Q^T Q = \sum_{i=1}^{N} (q^{(i)})^2$$  \hspace{1cm} (26)$$

Then the optimization problem is formulated as:

**Objective**

Minimize \( P \)

**Subject to**

$$w_i \leq w_i \leq w_u, \quad 1 \leq i \leq L$$

$$R_i \leq R_j \leq R_u, \quad 1 \leq j \leq M$$

$$\zeta_i \leq \zeta_k \leq \zeta_u, \quad 1 \leq k \leq N$$  \hspace{1cm} (27)$$

where \( L \) is the number of lines, \( M \) is the number of series resistances, and \( N \) is the number of observation points. \( w_i \), \( R_i \), and \( \zeta_i \) are the lower bounds of line width, series resistance, and damping ratio respectively. \( w_u \), \( R_u \), and \( \zeta_u \) are the upper bounds of line width, series resistance, and damping ratio respectively.
3. Numerical Example

In this section, we give two design examples.

Considering the circuit shown in Fig. 2, we first calculate the moments $m_0$, $m_3$, $m_6$ for the two observation points. Then the pure delays are excluded and we get $m_0$, $m_3$, $m_6$. Next, the two approximate poles are obtained and the corresponding $\omega_n$ and $\zeta$ are calculated. The following step is to optimize (27) using gradient projection method [13]. Since there is no explicit expressions for objective function and constraints as functions of design variables, numerical gradient approximations are employed.

The functions for line parameters are taken from [14] and are shown below:

\begin{align*}
R(\Omega/cm) &= 150/w \\
L(nH/cm) &= 1/(0.012w + 0.18) \\
C(pF/cm) &= 0.047w + 0.68
\end{align*} 

(28)

where $w$ is the line width in microns.

If the maximum percent overshoot that we can tolerate is 5%, the corresponding $\zeta$ is about 0.7, which is taken as the lower bound of the damping ratio. Since deep overdamping results in slow response, we set the upper bound $\zeta_u$ to 1.5. The other constraints are shown below:

\begin{align*}
w_1 &= 10\mu m, w_4 = 60\mu m \\
R_1 &= 2\Omega, R_u = 30\Omega
\end{align*}

$L$ is 2, $M$ is 2, and $N$ is 2.

The calculated line widths and termination resistances are $R_1 = 2.0\Omega$, $R_2 = 6.9\Omega$, $w_1 = 59\mu m$, and $w_2 = 18\mu m$.

Fig. 3 shows the simulated output responses at $V_{out1}$ and $V_{out2}$ using Spice3f5 with LTRA (lossy transmission line) model.

Fig. 4 is another example circuit. The lower, upper bounds and line parameter functions are the same as example 1. The calculated line widths and termination resistances are $R_1 = 2.0\Omega$, $R_2 = 2.0\Omega$, $R_3 = 2.0\Omega$, $R_4 = 2.0\Omega$, $R_5 = 8.9\Omega$, $w_1 = 60\mu m$, $w_2 = 56\mu m$, $w_3 = 11\mu m$, $w_4 = 52\mu m$, and $w_5 = 25\mu m$.

Fig. 5 and Fig. 6 show the output responses. Table 1 gives the numerical results. We can see that the overshoot match with the damping ratio quite well.
4. Conclusion

Properly designed interconnects are important part of the implementation of high performance system. We have proposed an optimization scheme to design interconnect using two-pole approximation. The circuit behavior is captured by two parameters, $\omega_n$ and $\zeta$, and the optimization objective and constraints are defined using the two parameters. For efficiency and accuracy, the lower-order moments are utilized to calculate the two approximate poles. For system performance, the damping ratio is employed to obtain an appropriate damping condition. Design high performance interconnect is a complicated task. There are many factors that can affect the design. It is an issue relating technology, circuit structure, loading conditions and line parameters. In the future, we plan to include wire length as a variable to increase the degree of freedom of design.

ACKNOWLEDGMENT

The authors would like to thank City University of Hong Kong for the support to this research work.

References


