

On-chip Interconnect Modeling by Wire Duplication *

Guoan Zhong, Cheng-Kok Koh, and Kaushik Roy
School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN 47907-1285
{zhongg,chengkok,kaushik}@ecn.purdue.edu

Abstract

In this paper, we present a novel wire duplication-based interconnect modeling technique. The proposed modeling technique exploits the sparsity of the L^{-1} matrix, where L is the inductance matrix, and constructs a sparse and stable equivalent RLC circuit by windowing the original inductance matrix. The model avoids matrix inversions. Most important, it is more accurate and more efficient than many existing techniques.

1 Introduction

With the continual increasing of clock frequency and global interconnect length and decreasing of signal transition time, accurate modeling of inductance effects becomes increasingly more important. The partial inductance matrix L obtained from the PEEC model [7] is extremely large and dense. Direct simulation of the full L matrix is very time-consuming and memory-consuming. To make the simulation more efficient, sparsification of L and L^{-1} matrices has been considered in [6, 4, 3, 5, 1, 2].

One sparsification approach is to discard the mutual coupling terms that are below some threshold. However, the resulting inductance matrix may not be positive definite; that leads to an unstable circuit. The shift-truncate method proposed in [6, 4] can guarantee that the generated sparse inductance matrix is positive definite. However, the accuracy is not satisfactory [3, 5].

[3] demonstrates the locality of L^{-1} . Hence, the L^{-1} matrix can be easily sparsified by dropping small entries while stability is guaranteed. Thus, modeling the inductance with the truncated L^{-1} matrix (denoted by \mathbb{L}^{-1}) instead of the L matrix can reduce the number of coupling elements and speed up the simulation. In [5], a new circuit element K , which is defined as the inverse of inductance, is introduced

and is incorporated in a simulation tool (known as the K method). To avoid the K element in simulation, The truncated L^{-1} matrix can be inverted to obtain a new inductance matrix (denoted by \mathbb{L}). As \mathbb{L} is also a dense matrix, direct simulation of \mathbb{L} (referred to as the \mathbb{L} method) is not efficient. [1] performs sparsification on the \mathbb{L} matrix (known as the double-inverse inductance model). Essentially, the double-inverse inductance model requires two approximation (sparsification) steps. [2] calculates the sparse inductance matrix directly by using exponential potentials and matrix inversions are avoided.

In this paper, we present a novel interconnect modeling technique based on *wire duplication*. This technique is motivated by the mathematical property that only a subset of the entries of the the \mathbb{L} matrix is required to reconstruct the \mathbb{L}^{-1} matrix. Consequently, we can construct a circuit that is equivalent to the \mathbb{L}^{-1} matrix out of the subset of \mathbb{L} by *wire duplication*. It is stable, sparse and as accurate as the K method [3, 5]. Furthermore, we can apply the wire duplication technique to the original inductance matrix L directly. Thus, matrix inversions are avoided. Most important, the accuracy is improved.

We use the following notation in the paper:

- L : The original partial inductance matrix.
- L^{-1} : The inverse of L .
- \mathbb{L}^{-1} : Truncated L^{-1} .
- \mathbb{L} : the inverse of truncated L^{-1} .
- \mathbb{L} method: The method that uses \mathbb{L} instead of L in the simulation.
- WD/ \mathbb{L} : The wire duplication model using the \mathbb{L} matrix.
- WD/ L : The wire duplication model using the original inductance matrix L .

*This research is supported in part by SRC (99-TJ-689), NSF (CAREER Award CCR-9984553), and a grant from Intel Corporation.

where V_{lk} and I_k refer to the voltage drop due to the inductance and the current in the wire k respectively.

We can rewrite the preceding equation in terms of the \mathbf{L} matrix:

$$\begin{bmatrix} V_{11} \\ V_{12} \\ V_{13} \\ V_{14} \\ V_{15} \\ V_{16} \\ V_{17} \end{bmatrix} = \begin{bmatrix} 6.73 & 4.20 & 2.49 & 1.48 & 0.90 & 0.57 & 0.38 \\ 4.20 & 6.32 & 3.75 & 2.23 & 1.35 & 0.86 & 0.57 \\ 2.49 & 3.75 & 5.92 & 3.52 & 2.13 & 1.35 & 0.90 \\ 1.48 & 2.23 & 3.52 & 5.81 & 3.52 & 2.23 & 1.48 \\ 0.90 & 1.35 & 2.13 & 3.52 & 5.92 & 3.75 & 2.49 \\ 0.57 & 0.86 & 1.35 & 2.23 & 3.75 & 6.32 & 4.20 \\ 0.38 & 0.57 & 0.90 & 1.48 & 2.49 & 4.20 & 6.73 \end{bmatrix} \frac{d}{dt} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \\ I_7 \end{bmatrix}$$

Now, we shall show how an equivalent circuit can be constructed out of the windows of the \mathbf{L} matrix. For example, if we take the window corresponding to the (2:4,2:4) sub-matrix of the \mathbf{L} matrix, and apply them to wires 2, 3 and 4, we have:

$$\begin{bmatrix} V_{12} \\ V_{13} \\ V_{14} \end{bmatrix} = \begin{bmatrix} 6.32 & 3.75 & 2.23 \\ 3.75 & 5.92 & 3.52 \\ 2.23 & 3.52 & 5.81 \end{bmatrix} \frac{d}{dt} \begin{bmatrix} I_2 \\ I_3 \\ I_4 \end{bmatrix}, \quad (3)$$

or

$$\frac{d}{dt} \begin{bmatrix} I_2 \\ I_3 \\ I_4 \end{bmatrix} = \begin{bmatrix} 2.53 & -1.60 & -0.00 \\ -1.60 & 3.65 & -1.60 \\ -0.00 & -1.60 & 2.69 \end{bmatrix} \begin{bmatrix} V_{12} \\ V_{13} \\ V_{14} \end{bmatrix}. \quad (4)$$

Among the three circuit equations for I_2 , I_3 , and I_4 in Eqn. (4), only the following equation

$$\frac{dI_3}{dt} = -1.60V_{12} + 3.65V_{13} - 1.60V_{14} \quad (5)$$

matches that in Eqn. (2). Hence, we can model wire 3 correctly, provided that V_{12} and V_{14} are correct. However, the equations for I_2 and I_4 in Eqn. (4) do not match those in Eqn. (2), i.e., wires 2 and 4 are not correctly modeled. Thus, their voltages V_{12} and V_{14} are incorrect. To provide a remedy to this problem, we can model these two wires correctly somewhere else and use the correct V_{12} and V_{14} values for the modeling of wire 3 here.

Figure 1 shows the modeling of signal 3. In this figure, the symbol \diamond stands for voltage controlled voltage source (VCVS) element. The two VCVSs provide the correct voltages for L_2 and L_4 . The inputs of the VCVSs come from the correct modelings of L_2 and L_4 . Since L_2 and L_4 here are controlled by their corresponding correct modelings, they are just dummy copies. We call such copies *dummy wires* and draw them in dashed lines. In contrast, if a wire is correctly modeled, we call it a *real wire* and draw it in solid lines. Here wire 3 is a real wire and wires 2 and 4 are dummy wires. The real wire and the dummy wires form a *group*. The total number of wires in a group is called the *group size* or *window size*. Figure 1 shows such a group, which models wire 3 correctly. Similarly, we can construct a group that includes dummy copies of wire 1 and wire 3 to model wire 2 correctly. In the group that models wire 4 correctly, dummy copies of

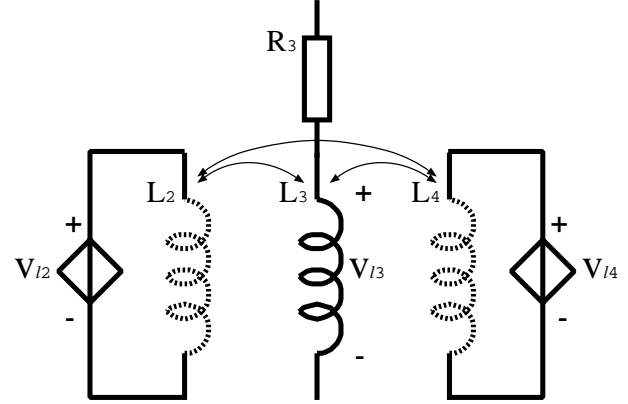


Figure 1. Modeling of wire 3.

wire 3 and wire 5 are included. Real wires 2 and 4 in these two groups provide the correct voltages V_{12} and V_{14} for the VCVSs in Figure 1.

In general, only one wire is correctly modeled in one group; so we need N groups for N wires in the simulation. There are one real wire and $2b$ dummy wires in each group if the bandwidth of \mathbf{L}^{-1} is $2b + 1$ (the groups at the two ends are special cases).

In each group, every pair of wires (including both real and dummy ones) are inductively coupled, and there is no inductive coupling among groups. Let $\tilde{\mathbf{L}}$ be the partial inductance matrix for the wire duplication model, then $\tilde{\mathbf{L}}$ is block diagonal and each block corresponds to one group. $\tilde{\mathbf{L}}^{-1}$ is also block diagonal. If we remove the rows for dummy wires and utilize the fact that dummy wires have the same voltages as the corresponding real wires, we get back the \mathbf{L}^{-1} matrix, which is positive-definite [5, 1, 2]. Thus, the circuit obtained by wire duplication is stable.

We use HSPICE to verify the correctness of this *wire duplication* model. We refer to the wire duplication model using the \mathbf{L} matrix as the WD/ \mathbf{L} model. Two sets of simulation are carried out: one set uses the full \mathbf{L} matrix (the \mathbf{L} method), the other uses the WD/ \mathbf{L} model. They produce *exactly* the same results (See Section 6 for details). It indicates that simulation with the WD/ \mathbf{L} model is also numerically equivalent to the \mathbf{L} method. This is expected, as they are physically and mathematically equivalent. Since the \mathbf{L} method and K method [3] are equivalent, simulation with the WD/ \mathbf{L} model is also equivalent to the K method [3].

4 Optimizing the Group Size

In the wire duplication model described in the previous section, there are $2b + 1$ wires in each group (less than $2b$ at the two ends). There are altogether about $N \cdot (2b + 1) \cdot b$

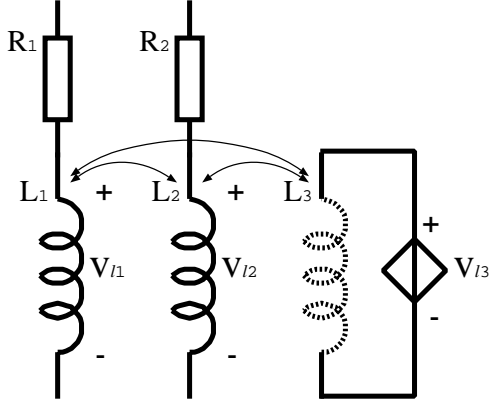


Figure 2. Modeling of wires 1 and 2.

inductive couplings, whereas the full inductance matrix contains $N \cdot (N - 1)/2$ couplings. If $b \ll N$, the wire duplication technique will produce an equivalent circuit of a smaller size.

There are two methods to reduce the circuit size even further. The first method merges the groups at the ends. The following window captures the modeling of wire 2:

$$l = \mathbf{L}(1 : 3, 1 : 3) = \begin{bmatrix} 6.73 & 4.20 & 2.49 \\ 4.20 & 6.32 & 3.75 \\ 2.49 & 3.75 & 5.92 \end{bmatrix},$$

$$l^{-1} = \begin{bmatrix} 2.54 & -1.68 & 0.00 \\ -1.68 & 3.65 & -1.60 \\ 0.00 & -1.60 & 2.70 \end{bmatrix}.$$

We can see that wire 1 is also correctly modeled. It means that wire 1 and 2 can share one group, as shown in Figure 2. Similarly, wires $N - 1$ and N can share one group.

Such an improvement is marginal; the second method, which uses larger windows, can achieve more reduction. For example, if we use a window of size 4, for the (1:4,1:4) window, the corresponding matrices are:

$$l = \mathbf{L}(1 : 4, 1 : 4) = \begin{bmatrix} 6.73 & 4.20 & 2.49 & 1.48 \\ 4.20 & 6.32 & 3.75 & 2.23 \\ 2.49 & 3.75 & 5.92 & 3.52 \\ 1.48 & 2.23 & 3.52 & 5.81 \end{bmatrix},$$

$$l^{-1} = \begin{bmatrix} 2.54 & -1.68 & 0.00 & -0.00 \\ -1.68 & 3.65 & -1.60 & -0.00 \\ 0.00 & -1.60 & 3.65 & -1.60 \\ -0.00 & -0.00 & -1.60 & 2.69 \end{bmatrix}.$$

In this case, wires 1, 2, and 3 are correctly modeled in this group with a dummy copy of wire 4.

With larger windows, fewer groups are needed to model all the wires. However, this reduction is achieved at the expense of more couplings within each group; the number of

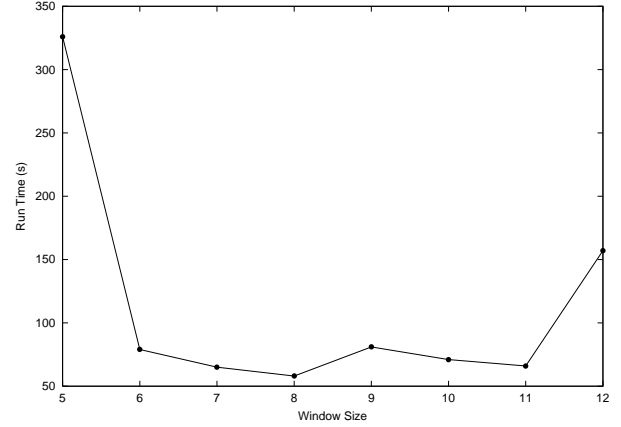


Figure 3. Simulation times for different window sizes.

couplings in each group increases quadratically with the window size. We discuss the trade-off in the remainder of this section.

For simplicity, we assume that all the groups are of the same size, B . The number of wires commonly found in two adjacent groups should be $2b$. Let n be the number of groups needed. Then,

$$n \cdot B - 2b(n - 1) = N \Rightarrow n = \frac{N - 2b}{B - 2b} \quad (6)$$

The number of total wires used is $B \cdot n$; the number of total couplings is $\frac{B \cdot (B - 1)}{2} n$. The purpose of this study is to build an equivalent circuit with a smaller size. It includes both wires and the coupling elements. As a rough estimate, we use $B^2 \cdot n/2$ as the circuit size and try to minimize it. We can easily conclude that $\frac{B^2 \cdot n}{2} = \frac{B^2(N - 2b)}{2(B - 2b)}$ is minimized when

$$B = 4b, \quad (7)$$

and the minimal value of $B^2 \cdot n/2$ is

$$(B^2 \cdot n/2)_{min} = 4b(N - 2b). \quad (8)$$

For the circuit example in Section 6, we set the bandwidth of L^{-1} to be 5, i.e., $b = 2$, and perform wire duplication simulations for different window sizes (from 5 through 12). The run times are shown in Figure 3. We can see that the circuit obtained with a window size of 8, which is $4b$, has the smallest run time. That coincides with our estimation.

Although the circuit sizes at different window sizes are different, all the resulting circuits are physically and mathematically equivalent. The simulation results for all of them are *exactly* the same.

When $B = 4b$, there are about $(4b - 1)(N - 2b)$ inductive couplings and $2(N - 2b)$ wires ($N - 4b$ of them are dummy

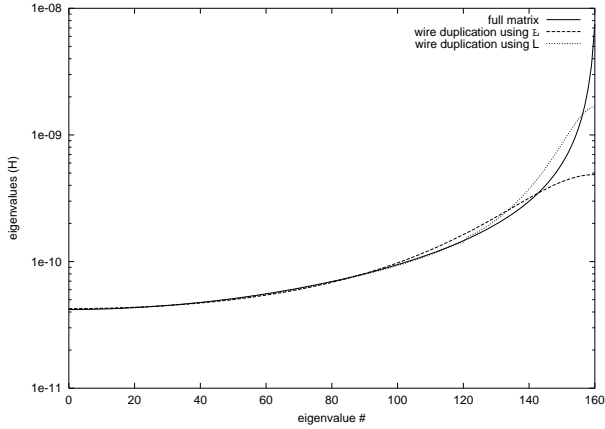


Figure 4. Eigenvalues for full matrix, wire duplication using \mathcal{L} matrix, and wire duplication using L matrix.

wires). Note that the K method has about $b \cdot N$ couplings and requires no dummy wires. Therefore, the wire duplication technique uses about four times couplings required in the K method and an additional $N - 2b$ dummy wires. This is the price that we pay for RLC simulation instead of RKC simulation.

5 Wire Duplication Using L Matrix

In the previous sections, we use the \mathcal{L} matrix, which is obtained by two matrix inversions, in the wire duplication model. It turns out that we can use the original inductance matrix L directly in the modeling to avoid matrix inversions. An additional and more significant benefit is that the accuracy is also improved. We refer to the wire duplication model using the L matrix as the WD/ L model.

The use of the windowed \mathcal{L} matrix is based on a strict mathematical property. That mathematical property can also explain the validity of using the windowed L matrix to a certain extent, since L^{-1} is almost a band matrix. We have shown that WD/ \mathcal{L} is stable. For WD/ L , we can also prove its stability in a similar way.

There is another benefit of using the L matrix instead of the \mathcal{L} matrix: It is more accurate. If we use the $4b - 1$ inductive couplings between each wire and its neighbors. In contrast, the WD/ \mathcal{L} model can only capture $2b$. Although some of the extra coupling captured by WD/ L may not be very accurate, it is better than ignoring them. Figure 4 plots the eigenvalues for the full matrix and the equivalent inductance matrices for WD/ \mathcal{L} , and WD/ L . The bandwidth of \mathcal{L}^{-1} and window size used in the wire duplication methods are 5

($b = 2$) and 8 ($= 4b$), respectively. We can see that the eigenvalues of the equivalent inductance matrix for WD/ L match those of the full matrix better; the eigenvalues of the equivalent inductance matrix for WD/ \mathcal{L} diverges earlier. Simulation results also validate this conclusion (see Section 6 for details).

In the wire duplication method, using the L matrix is more accurate than using the \mathcal{L} matrix only when the window size B is larger than its minimum value $2b + 1$. If $B = 2b + 1$, using L does not capture more couplings. Indeed, it may not capture the correct \mathcal{L}^{-1} values due to the small window size. As pointed out in Section 4, using minimum window size ($B = 2b + 1$) for simulation is not efficient. As it is preferred to a large window size for simulation, we should always use the original inductance matrix L instead of \mathcal{L} .

6 Experiment Results

We demonstrate the wire duplication technique on a bus with 128 signals. Shields are inserted after every four signals. The wire length is 1mm, the cross-section is $1 \times 1\mu\text{m}$, and the separation between wires is $1\mu\text{m}$. The wires are divided into 5 segments along the length. The driver resistance is 30Ω and the load capacitance is 40fF .

Different simulation techniques are studied: the full matrix method, the \mathcal{L} method, WD/ \mathcal{L} , WD/ L , simple truncation, shift-truncate [6, 4], and the double-inverse inductance model [1, 2]. Because Ksim [5] is not available to us, we cannot perform K method directly. Instead, we use the \mathcal{L} method, which is mathematically equivalent to the K method. (Note that \mathcal{L} method is different from the double-inverse inductance model). The bandwidth of \mathcal{L}^{-1} is 5 ($b = 2$), and the window size for wire duplication is 8 ($= 4b$).

A 1V 20ps ramp input is applied to the first signal, and the rest are quiet. The waveforms for the second signal obtained from different methods are shown in Figure 5. Results from the \mathcal{L} method and the wire duplication methods are shown on the left; and results from the shift-truncate and double-inverse methods are shown on the right. As a reference, the result from full matrix modeling appears in both plots. The simple truncation method is not stable and not shown.

The first conclusion we can draw from Figure 5 is that WD/ \mathcal{L} is equivalent to the \mathcal{L} method. These two methods match *exactly*. The second conclusion is that WD/ L is more accurate than WD/ \mathcal{L} .

In the double-inverse method [1, 2], the same cutoff percentage is used for the L^{-1} matrix and the \mathcal{L} matrix. Both the double-inverse method with 2.2% cutoff and the shift-truncate method have similar number of mutual inductances as the wire duplication methods, while the double-inverse method with 1% cutoff has about twice as many mutual inductances. We can see that they perform worse than WD/ L .

Table 1 lists the memory and time usage for different

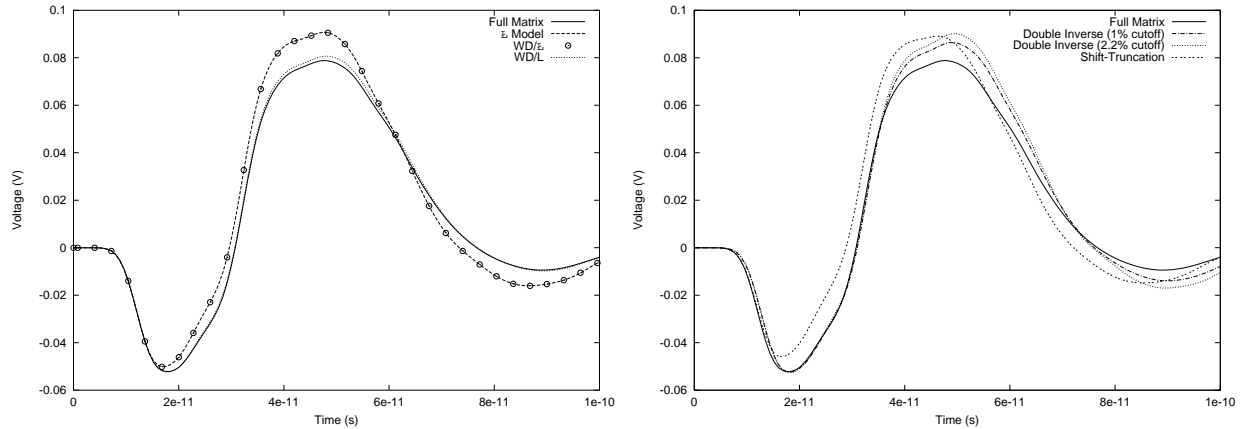


Figure 5. Simulation results for signal 2.

Table 1. Run time and memory usage.

Method	Memory(MB)	Run Time(s)
full L matrix	812.3	4.74×10^5
\mathbb{L} method	230.2	2.30×10^4
double-inverse (1% cutoff)	97.1	897
double-inverse (2.2% cutoff)	44.3	357
shift-truncate	38.5	283
wire duplication	15.3	58

methods. Both WD/ \mathbb{L} and WD/L methods use the same amount of time and memory. We can see that although there are additional dummy wires in wire duplication method, it uses much less memory and runs much faster. Although the number of wires and couplings contribute directly to run time of simulation, the convergence rate, which depends on how well-conditioned the matrices are, is also very important. We can see that the wire duplication methods are faster and require less memory even than the shift-truncate method and the double-inverse method (2.2% cutoff), despite the fact that their circuit sizes are actually larger (due to the additional dummy wires). One possible explanation is that the inductance matrices of the wire duplication model are well-conditioned, resulting faster simulation times.

7 Conclusion

In this paper, we propose a new interconnect modeling technique—wire duplication. With this technique, we can generate stable, sparse and yet accurate inductance models for on-chip interconnects. The wire duplication model using the \mathbb{L} matrix is physically and mathematically equivalent to K method. However, it avoids using the new circuit element K . Using the original inductance matrix L in the wire dupli-

cation model instead of the \mathbb{L} matrix is even more accurate. Moreover, it avoids matrix inversions.

References

- [1] M. Beattie and L. Pileggi. Efficient inductance extraction via windowing. In *Proc. Design Automation and Test in Europe Conf.*, pages 430–436, 2001.
- [2] M. Beattie and L. Pileggi. Modeling magnetic coupling for on-chip interconnect. In *Proc. Design Automation Conf.*, pages 335–340, 2001.
- [3] A. Devgan, H. Ji, and W. Dai. How to efficiently capture on-chip inductance effects: introducing a new circuit element K . In *Proc. Int. Conf. on Computer Aided Design*, pages 150–155, 2000.
- [4] Zhijiang He, Mustafa Celik, and Lawrence T. Pileggi. SPIE: Sparse partial inductance extraction. In *Proc. Design Automation Conf.*, pages 137–140, 1997.
- [5] H. Ji, A. Devgan, and W. Dai. KSim: a stable and efficient RKC simulator for capturing on-chip inductance effect. In *Proc. Asia South Pacific Design Automation Conf.*, pages 379–384, 2001.
- [6] B. Krauter and T. L. Pileggi. Generating sparse partial inductance matrices with guaranteed stability. In *Proc. Int. Conf. on Computer Aided Design*, pages 45–52, 1995.
- [7] A. E. Ruehli. Inductance calculation in a complex integrated circuit environment. *IBM Journal of Research and Development*, pages 470–481, September 1972.
- [8] G. Zhong, C.-K. Koh, and K. Roy. On-chip interconnect modeling by wire duplication. In *Purdue University Technical Report, ECE-02-04*, 2002.