Challenges of Modeling VLSI Interconnects in the DSM Era

Narain D. Arora
Simplex Solutions, Inc., Sunnyvale, CA, USA, arora@simplex.com

ABSTRACT

As VLSI technology shrinks to deep sub-micron (DSM) geometries (below 0.18µm), the parasitic due to interconnects are becoming a limiting factor in determining circuit performance. An accurate modeling of interconnect parasitic resistance (R), capacitance (C) and inductance (L) is thus essential in determining various chip interconnect related issues such as delay, cross-talk, IR drop, power dissipation etc. Solutions that provide extreme accuracy, though suitable for very small design with few hundred of devices, become impracticable and unrealistic when applied to the full chip with millions of devices, a common place in today's DSM era. In this paper we will review practical methods of accurately estimating R, C, and L of a given circuit layout that maximizes the accuracy while minimizing the time and resources that such accuracy demands. The paper then covers other related issues as model order reduction and silicon validation of the models.

Keywords: Interconnect parasitic extraction, interconnect modeling, silicon validation of interconnect models.

1 INTRODUCTION

With ever increasing circuit density, operating speed, and high level of integration in deep sub-micron (DSM) chip designs, an accurate and proper interconnect modeling is a must to assure the performance and functionality of multi-million transistor VLSI circuits. An interconnection can be described by means of its three electrical parameters, resistance (R), capacitance (C) and inductance (L), and today most of the interconnect parasitic extraction and delay analysis tools are limited to interconnect R and C only. The values of these parameters depends on the physical and geometric description of the wire. However, faster on-chip rise times, longer wire lengths, use of low resistance Cu interconnects and low-K dielectric insulators have necessitated the modeling of wire inductive (L) effects which were ignored in the past. The most complete description of the wire is now given by RLC model. Selection of the model in terms of both simplicity and accuracy must be made in relation to values of wire parameters, driver parameters and the frequency bandwidth of the signal that the wire must transmit [1].

In section, 2 a brief description of the various techniques used in modeling R,C, L of a wire are presented. In terms of solution methods both the domain methods that solves Maxwell's equations and techniques that precharacterizes environment of a wire in two-dimensional (2D), quasi-3D (or 2.5D) and 3D at the chip level are covered. Section 3 will cover characterization of on-chip interconnect using actual silicon measurement and briefly discuss process variation followed by conclusions of the paper.

2 INTERCONNECT MODELING

It is well known that in the DSM regime interconnect delays dominate gate delays (Figure 1). Historically interconnect was modeled as a single lumped capacitance using the well known parallel plate capacitance formula. With the continuous scaling of technology the wire cross-sectional area decreased, while at the same time wire length increased due to increase in the die area, resulting in a wire resistance that becomes significant. This leads to the development of the RC delay model first as a lumped RC circuit (Figure 2i) and then as a distributed RC model (many sections of lumped RC) to improve accuracy. With faster on-chip rise time, higher clock frequencies and use of Cu wire as interconnect has necessitated the use of RLC models as a distributed network (Figure 2). The computation of R, L and C matrices of a multiport, multi-conductor VLSI interconnects involves solution of quasi-static electro-magnetic system of equations (Maxwell Equations) [2]

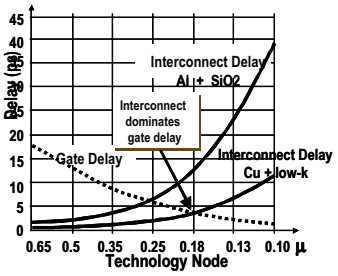

Figure 1. Gate vs. interconnect delay (SIA roadmap 1997).

$$\operatorname{curl}\mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \qquad \operatorname{div}\mathbf{D} = \rho \qquad (1)$$

Ampere's Theorem:

$$\operatorname{curl}\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \operatorname{div}\mathbf{B} = 0$$

Faraday's Law:

Constitutive Equations:

$$\mathbf{B} = \mu\mathbf{H}, \quad \mathbf{D} = \varepsilon\mathbf{E}, \quad \mathbf{J} = \sigma\mathbf{E} \qquad (3)$$

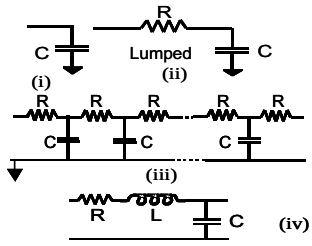

Figure 2. Interconnect delay model. Lumped (i) capacitor model, (ii) RC model. (iii) Distributed RC model, and (iv) RLC model.

to be solved with initial and boundary conditions, where symbols have their usual meaning. Choice of numerical techniques to solve the partial differential equations such as finite difference (FD), finite element (FE), boundary element method (BEM) etc. leads to different methods of discretization of the domains

by elements of simple shapes (meshes) [2]. In all these methods differential equations are converted into algebraic equations which in turn are solved by different methods such as iterative, direct or multigrid methods to calculate potential and fields. Interpolation or integration then leads to interconnect parameter of interest such as R, L, and C.

Though each of these methods give similar results in a dense layout system, they often give different results in a sparse layout system depending upon the boundary conditions used and problem set-up. Also each method has its own advantages and disadvantages. The most commonly used field solvers in the public domain are FASTCAP [3] for capacitance and FASTHENRY [4] for inductance calculation, though many other solvers are also available [5]. More recently the Monte Carlo based statistical field solver, called Quickcap [6], is fast becoming the industry standard for capacitance calculations because the method does not require meshing of the domain and as such one can easily calculate capacitance of a long net with less of a memory. Table 1 shows the capacitance of a orthogonal net shown in Figure 3 using 3 different methods. Due to the complexity of the N1 with many small corners(vias), FEM will take least compute time due to lower number of mesh elements (unknown) These meshes could be of any shape, unlike in FD method which could have only fixed shape meshes. A regular structure such as three parallel lines will have almost the same number of mesh elements in both FEM and FD method resulting in similar compute time.

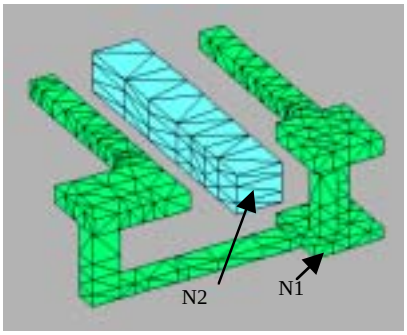

Figure 3. Two nets structure showing meshes.

Table 1. Different methods of capacitance calculations.

Method	FEM		FDM		Qcap	
	net1	net2	net1	net2	net 1	net2
Unknowns	26,322		111,456			
Time (s)	26		258		80	
Capacitance (fF)	2.1	1.0	2.09	.98	2.01	.99

With the ever increasing complexity in process technology (e.g. low-K, conformal dielectrics, air bubble dielectric, non-vertical conductor cross-section, shallow trench isolations etc) the interconnects are no longer just conductors separated by an insulator. In fact today's wires are multi-level (5 and more metal levels are common), multi-layer and multi-dielectric. In addition, use of chemical-mechanical polishing (CMP) requires certain minimum density of the metal in a given area (say 100x100 μm^2) thereby necessitating the use of metal fills (dummy metal). Impact of these metal fills is to change the line capacitance by almost 3-10% depending upon the shape, and size of the metal fills and how dense is the net under consideration. The challenge for the field solvers is to handle these process related effects very efficiently and accurately.

Although numerical methods discussed above provide an accurate electromagnetic field solutions of a complex geometry, they are too compute intensive and can not be used at the chip level extraction of wire RLC. In fact these methods are not practicable for circuits containing larger than few tens of transistors. At the chip level, an estimation of wire RLC requires an approach that is very efficient and has reasonable accuracy to say within 10% (of the field solver value) for each net in a chip containing millions of nets. Different modeling approaches of computing RLC at the chip level has resulted in different types of tools [6] namely (1) Rule based that used Boolean operations, (2) pattern matching using look-up table and (3) context based that looks at each conductor with 3D surrounding. All these approaches use field solvers to precalculate R and C for different structures. With table look-up, the memory requirement grows very quickly due to an increase in the number of parameters required for each configuration. Sophisticated interpolation are used to reduce the amount of data that needs to be stored. Analytical models execute quickly providing faster extraction, but needs sophisticated model development, particularly for C calculation, because capacitance is complex function of layout parameters [7]. For R calculation, it is the analytical modeling approach that is more common, while for L and C both approaches are used. The software tool that reads the layout geometry and calculates the corresponding RLC of the wires is known as Layout to Parasitic Extractor (LPE), or simply parasitic extractor.

2.1 Resistance Extraction

Computation of R is the simplest of the three interconnect parameters because R is a function of the conductor geometry and conductivity only, being independent of adjoining conductors. To calculate R, interconnections are approximated by 2D homogeneous regions (metals, polysilicon, diffusion area) that are connected to each other along surfaces representing vias between these regions. Resistance of each wire is then obtained simply by multiplying sheet resistance, r_s, by ratio of the length to width of the wire. Due to interconnect being multi-layer, sheet resistance of a wire, particularly Cu wires, now becomes line width and pattern dependent. The lower the width, higher the resistance. New processes require slotting of the wide wires and

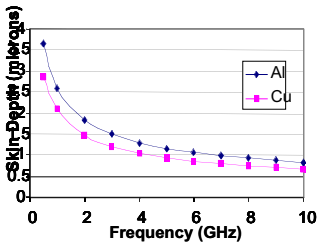

Figure 4. Skin depth vs frequency for Al and Cu wire.

as such one needs to use 2D as solver to precalculate impact of the slots on wire resistance. As signal frequency increases the penetration depth (skin depth, $\delta=1/\sqrt{\pi f\mu\sigma}$) of the electromagnetic field in the conductor decreases thereby increasing conductor resistance. At 1 GHz, for Al, δ=2.8μm. As clock frequency exceeds 1 GHz, impact of skin effect must be considered. However, modeling R with skin effect is not straight forward [8].

2.2 Capacitance Extraction

The capacitance estimation of a wire is more complex compared to the resistance estimation. This is because capacitance between two wires is affected by the proximity of the other wires that could be above, below, or adjacent to the wire under consideration. Historically, the wire was modeled as a parallel plate capacitance ($C=\varepsilon\ \ell\ W/H$) where ℓ and W are length and width, respectively, of the wire and H is height of the wire from the substrate such that W<< H and wire thickness T is negligible. This, the so called 1D formula, became inadequate with the scaling of the technology when lateral capacitance (C_{12} and C_{13} in Figure 2) became significant part of the total line capacitance. In the 2D modeling approach all capacitive effects (area plus lateral) are modeled as long routing lines with per unit length values. This 2D model though more accurate compared to the 1D, still gives significant error while calculating capacitance of two crossing lines which is a 3D problem. To address the 3D effects the so called 2.5D model is proposed which in effect combines two orthogonal 2D models [7]. However, for DSM technologies a true 3D model is needed, where capacitance of a 3D pattern is calculated. Remember that 3D model is not a trivial extension of a 2D model.

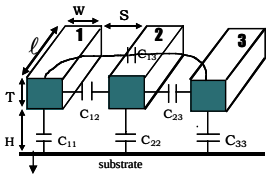

Figure 5. Capacitance of wire 2 is the sum of lateral plus ground capacitance.

Following is the block diagram of a 3D capacitance extractor typical of a today's parasitic extractor such as Simplex Fire and ICE [8]. Given the description of the process cross-section, the first step is to generate

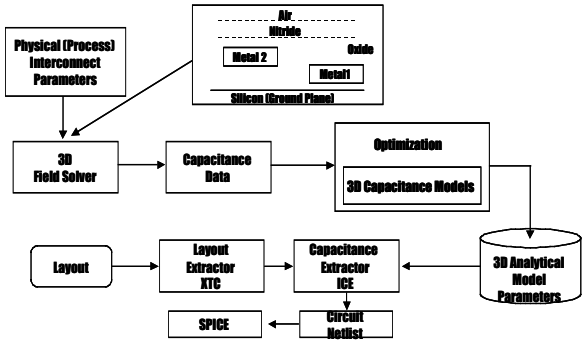

Figure 6. Block diagram of a typical capacitance extraction

using 3D field solvers for tens of thousands of structures. This data is generated once for each technology as a function of line width, W, and spacing, S. The resulting data is then fitted into physics based empirical models using non-linear optimization method. The model parameter coefficients are then stored. In some other approaches the data is stored as a look-up table (also called a pattern library). Layout geometry is fractured first into stripes (window) and then into elemental area with polygons. Capacitance is then calculated for each vertical profile containing these polygons using stored model equations or look-up table. For more details refer to [10].

1.3 Inductance Extraction

Inductance of a wire is much more complicated to model compared to resistance or capacitance because inductance is defined only for a loop of a wire. In a VLSI chip, calculation of wire inductance will therefore require knowledge of the current return path(s). However, often the return path is not easily identified from the layout as it is not necessarily through the silicon substrate. Calculation is further complicated by the fact that not all the return currents follow DC paths as some are in the form of displacement current through the interconnect capacitances. Furthermore, unlike capacitive coupling, inductive coupling is much stronger (has long range effects). As such, localized windowing (nearest neighbor approximation), commonly used for capacitance calculation, is not valid for inductance calculation as it leads to unstable interconnect models [11].

If the return path is known, calculating L using the following equation (Maxwell)

$$\frac{J}{\sigma}+\frac{j\omega\mu}{4\pi}\int_{v}\frac{J(r')}{\|r-r'\|}dr'=V \qquad (4)$$

is straight forward. Knowing current I, the inductance L could be calculated using Φ/I. Note that some times inductance matrix [L] is derived directly from the capacitance matrix [C] such that $[L]=[C]^{-1}/v_0^2$ where v_0^2 is phase velocity of the medium. However, for on-chip inductance this relation does not hold and will result in overestimation of the wire inductance [12].

On a chip it is not clear which conductor forms the loop. The concept of partial inductance is introduced which allows algebra to take care of determining the loops [13]. Each partial inductance assumes current return at infinity so that "infinity return" paths cancel out when we do the subtraction. Based on this approach one can easily calculate self and mutual inductance of two parallel lines of length l, width w, thickness t, separated by distance d

$$L_{self}=\frac{\mu_0}{2\pi}\left[\ell.\ln\left(\frac{2\ell}{w+T}\right)+\frac{\ell}{2}+0.2235(w+T)\right]$$

$$M=\frac{\mu_0\ell}{2\pi}\left[\ln\left(\frac{2\ell}{d}\right)-1+\frac{d}{\ell}\right] \qquad (5)$$

Most of the chip level inductance extractors use a closed form expression of the type above (based on partial inductance) that employs a quasi-rule based scheme and is applicable for long wide lines found in upper level routing such as power lines and buses. Note that while wire capacitance is important for any length of a wire, the wire inductance is significant only for a certain ranges of wire length. For the case where the driver and line impedances match, it has been shown that the line inductance is important only if the length ℓ falls within the following inequalities [12]

$$t_r / \sqrt{LC} < \ell < 4\sqrt{L/C} / R$$

where t_r is the rise time of the signal. Inductance is not an issue when $t_r = 4L / R$.

Extracting wire inductance at the chip level for all nets is not practicable due to large amount of data generated (unknown return path). As such, inductance calculation is restricted to clock lines and buses only. Furthermore today's commonly used RC delay calculators do not efficiently handle inductance and work is in progress [12]. For these reasons inductance often is reduced by process design using metal plates or interdigitated shields.

2.4 Model Order reduction

As VLSI is moving towards system-on-chip (SOC) design, with the result that the number of extracted parasitic becoming increasing more than 100M with database in GB range. A typical parasitic extractor often takes 24-48 hours (single CPU) to just extract R and C. To use this huge parasitic database for downstream tools such as timing analysis, signal integrity etc, the data needs to be reduced. The reduced model must be accurate (preserve both DC and AC characteristic), stable and scalable. The most commonly used model order reduction technique is the asymptotic waveform evaluation (AWE) with varying order moment matching (Elmore delay being first order matching). However, accurate computation of these models can be done using algorithms like Pade via Lanczos (PVL) or Arnoldi based algorithms [14].

3 MODEL VALIDATION

The on-chip interconnect modeling mostly focuses on the generation of 2D/3D interconnect model libraries using 2D/3D field solvers based on interconnect geometry and material characteristics from the foundry's electrical design rules. To validate these models, test structures are designed, fabricated on silicon, measured and characterized to extract the parameters of interconnect geometry (width W, thickness T, inter-layer dielectric, ILD, thickness), material characteristics (sheet resistance) and model libraries. These structures should also include practical line structures such as global lines or clock nets so as to make the calibration more meaningful. Though it is expensive and time consuming, it is the only way to do correct model validation. Methods of measurements on such test structures are discussed in reference [15].

Due to process variation, interconnect process parameters vary substantially as shown in Figure 7. Variation is especially large in ILD thickness and metal line width, almost 20%. In fact resistance of the drawn lines (layout) also changes due to optical proximity correction (OPC) required for proper processing of DSM lines. These variations finally result in variation of the signal delay and cross-talk noise and as such must be characterized and modeled.

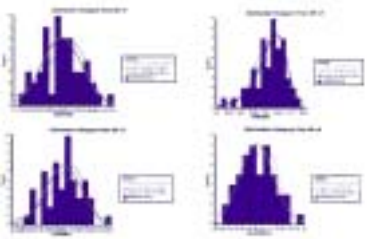

Figure 7. Statistical variation of process parameters

4 CONCLUSIONS

Accurate modeling and characterization of interconnect is essential as it strongly impacts circuit performance. Different modeling approaches for estimating and extracting interconnect parameters - resistance, capacitance and inductance - have been reviewed. Due to the large data base generated using parasitic extractors, model order reduction techniques are briefly reviewed. All efforts should be made to create realistic test structures that truly reflects actual IC products and their process variation studied and modeled.

REFERENCES

[1] C.K. Cheng, J. Lillis, S. Lin and N. Chang, "Interconnect Analysis and Synthesis", John Wiley & Sons, 2000.

[2] M.N.O. Sadiku, "Numerical techniques in Electromagnetics", CRC press, 1992.

[3] K. Nabors and J. White, IEEE Trans. Computer-Aided Design, 10, 1447, 1991.

[4] M. Kamon, M.J. Tsuk, and J. White, IEEE Trans. Microwave Theory and Technique, 42, 1750, 1994.

[5] N.D. Arora, *Tutorial on Interconnect Modeling*, ISQED March 2001.

[6] Y.L. LeCoz and R.B. Iverson, Solid-State Electron. 35, 1005, 1992.

[7] W.H. Kao, C.Y. Lo, M. Basel and R. Singh, Proc. IEEE, 89, 729 (2001).

[8] B. Krauter and S. Malhotra, Proc. 35th DAC, 303, 1998.

[9] Fire & ICE, User's Guide, 2000 http//simplex.com

[10] N.D. Arora, K.V. Raol, R. Schumann, and L.M. Richardson, IEEE Trans., Computer-Aided Design, 15, 58, 1996.

[11] M. Beattie and R. Pileggi, Proc. 36th DAC, 915 (1999).

[12] Y. Ismail, and E. Friedman, "On chip inductance in High speed integrated circuits", Kluwer Academic Publishers, 2001.

[13] A.E. Ruehli, IBM J. Res. Dev., 16, 470 (1972).

[14]J.R. Philips, E.Chiprout, and D. Ling, Proc. 33rd DAC, 377, 1996.

[15] D. Sylvester and C. Hu, Proc. IEEE, 89, 634 (2001).