Interconnect Modeling for High Speed Digital Circuits - the Role of RLC Coupling

R. Suaya
Mentor Graphics Corp.
roberto_suaya@mentor.com

ABSTRACT

The modelisation of capacitance and inductance of wires for timing and noise simulation of digital circuits is discussed. Different domains characterized by the length of the wires and the driving strength require different approximations in the representation of the parasitic parameters. A modelisation methodology for interconnect parasitics is emerging.

1 INTRODUCTION

The complexity associated with model generation and characterization of interconnect parameters has increased substantially over the past few years. For dimensions down to $250nm$ a representation of local interconnect wires in terms of its lumped parameters R_{wire} and C_w total resistance and total capacitance of a wire suffice. The total capacitance of a local interconnect wire is dominated by the capacitance to substrate. For global interconnect wires a replacement of the lumped parameters with distributed per unit length parameters, r_w, c_w, and L the length, for each homogeneous subsection of the wire is an adequate representation of wires for timing simulation. Each signal wire can be analyzed independently of its surrounding signal wires. Effects of cross coupling among wires, either capacitive or inductive are negligible. The extraction of the relevant parameters from the layout data is straightforward. The timing simulation is controlled by the logic, and the effect of the wires is predictably small for realistic values of the wire parameters. The traditional approximation to delay holds:

$$delay = R_{driver}(C_w + C_{load} + C_{diff}) + r_w c_w L^2 \quad (1)$$

with C_{diff} the parasitic capacitance associated with the source and drain regions of the load device, and the last term is typically negligible.

With the onset of Deep Submicron below $250nm$ the last term in (1) becomes dominant for long wires, giving rise to two separate regimes. Keutzer an Sylvester [1] successfully argued that blocks of nearly 50K gates can be built with a wire load model that neglects r_w, model that is conventionally used during synthesis. In fact, -under scaling-for blocks up to 50K gates interconnect contribution to delay falls relative to logic delay. For each process generation one can compute the critical distance L_{cr} beyond which wire delay dominates the delay of a minimum size transistor T_{mst}

$$\tau = \frac{1}{2} r c L_{cr}^2 = T_{mst} \quad (2)$$

The value of $L_{cr} = 0.6mm$ at $250nm$ decreasing by 0.75 in going from one generation to the next. Wires whose length exceed L_{cr} require special treatment that includes buffer insertion and wire widening [2]. The fraction of wires whose length is larger than L_{cr} is roughly 0.5% at $150nm$, growing to nearly 2% at $50nm$. These wires are used for global interconnect, their analysis impact the simulation, verification and synthesis strategy. To control the growth in r_w, technologists use reverse scaling of metal thickness. A side effect manifest in c_w The dominant contribution to c_w for minimum wire separation is the cross coupling capacitance to nearest neighbors c_{cc}. The ratio c_{cc}/c_{total} can reach 90% as the technology approaches $50nm$ for nearby wires separated by the minimum distance. Driver's non linearity and the relative timing of nearby transitions play a fundamental role in the accurate determination of time delay. Circuit level simulations with the dynamic information of the signal activity are strongly recommended. Direct application of (1) must account for the Miller effect and are not particularly reliable in the presence of almost simultaneous nearby transitions. To fully characterize the system, both c_w and c_{cc} need to be known with high accuracy. Unduly pessimistic estimates for the net effect of cross coupling on time delay were advanced in the 1997 NTRS [3]. They were based on the dominance of the last term of (1), assuming minimum distance separation among signal wires for computing the delay. Reality is somewhat more forgiving. On local interconnect, c_w is in fact dominated by c_{cc} for density considerations, but the average length is much smaller than L_{cr}, the last term in (1) is small. To ensure the continuation of dominance of the load factors to the interconnect components demand proper resizing of transistor ratios W/L_{tr}. For global wires, of lengths $0(L_{cr})$ the net effect of c_{cc} on

timing could be large, if one insists in keeping minimum separation among signal wires (factors of 5 to 6 difference in timing estimates according to the activity of nearby wires). Since global wires run on less dense upper metal layers, wire separation can increase so as to ameliorate the influence of c_{cc}. Wires related effects on noise are even more pronounced. Noise voltage ratios to power supply were derived [4] for local and global interconnect.

$$\frac{V_n}{V_{dd}} = 0.51\frac{r_w c_{cc} L^2}{T_{rise}} \quad \text{for} \quad T_{rise} >> \tau, \tag{3}$$

$$\frac{V_n}{V_{dd}} = 0.50\frac{c_{cc}}{c_{gr} + c_{cc}} \quad \text{for} \quad T_{rise} < \tau \tag{4}$$

with T_{rise} is the driver switching time: $R_{dr}.C_{gate}$, and τ given in (2). Safe noise limits $V_n < \alpha V_{dd}$ with $\alpha \leq 0.2$ demand stricter requirements on local interconnect than timing. Under scaling noise margins scale inversely to the scaling factor,posing additional burden on timing verification. Noise in addition to timing modelisation demand accurate and independent estimates of c_{cc} and c_{tot}.

2 CAPACITANCE MODELISATION

Capacitance coupling modelisation and extraction has improved substantially over the last few years. Accuracy (to within 10% for nominal values of technology parameters) can be efficiently accounted with today's commercial extractors, among them Mentor Graphics xCalibre. The major concern on today's extraction technology is the monumental size of the database. Proper use of hierarchy and data reduction is actively being pursued. Screening effects play an important role in the computation of capacitance couplings thus permitting to reduce the problem to that of computing only nearby interactions, among them the coupling to the nearest wires in the same metal layer c_{cc} the cross over coupling to wires in the closest occupied metal layer and the substrate contribution. The total c_w can be approximated by the sum of these terms. The value of each one of these terms depend on the overall configuration. Meaning the capacitance matrix is not separable. At variance with the potential vector computation that is separable. In practice the correction to separability can be approximated by simple scale factors on the underlying functional form. We illustrate the modelisation. Consider a configuration consisting of three equidistant parallel wires on the same plane, in the presence of a plate on a nearby plane. The capacitance of the middle wire in the three wires plane c is given by: $c = 2c_{cc} + c_r$, with c_{cc} near-body contribution c_r cross over capacitance to plate. To choice of functional form is dictated by experience with 3D simulations: The following parameterization for c_{cc} is often used at Mentor Graphics:

$$c_{cc}(s, w) = \left(e^{-\alpha_1 s} + \beta_2 s^{-\alpha_2}\right) \beta_1 w^{\alpha_3}. \tag{5}$$

The parameters are configuration dependent. Both in terms of the topology and the geometry, e.g., it varies with the metal layers under consideration. They can be obtained for example using nonlinear least square fit [5] to observations (one could use experimental values or 3D Laplace solver simulations.) In figures 1 and 2 we display the results of one particular fit. In an obvious notation $c_{cc}(s, w) = c1(s) * c2(w)$, with s being the wire separation and w the common width. The number of configurations to be analyzed is large.

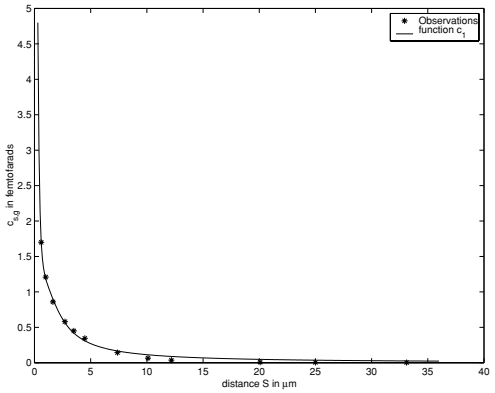

Figure 1: fit of c_1, and chosen observation values.

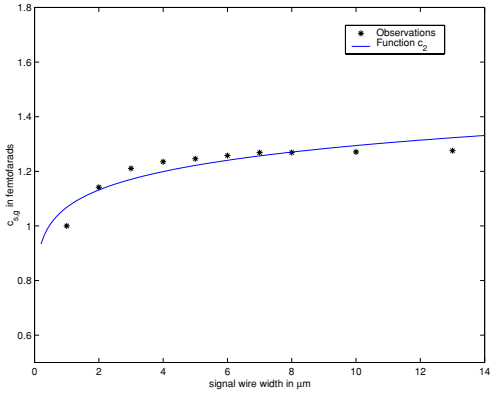

Figure 2: fit of c_2, and chosen observation values.

2.1 Validation

Given the sensitivity of timing and noise to cross coupling capacitance values it is mandatory to find appropriate validation methods to the capacitance extraction problem. In two accompanying papers we discuss in detail proper validation methods. [6], [7]. In [6] we describe a simple experimental method to independently measure the c_{cc} and c_w to a high level of precision, for a comprehensive library of capacitance structures, while in [7] we discuss a validation methodology to accurately calibrate a 3D simulator using a statistical validation procedure that incorporates accurate experimental data.

The entire methodology has been tested in research environment at LETI and in a leading $130nm$ process at TSMC. Three observations are noteworthy: simulator calibration is important for more reasons than the verification of the accuracy of a particular algorithm. Nature does not quite satisfy some of the geometrical and layer simplifications that 3D simulators are forced to impose for expediency of the computation. The second observation is related to the importance to introduce process related variations in the computation. Capacitance fluctuations due to process variations are larger than the corresponding variations in transistor parameters. Finally, an analysis based on worst case process corners for parasitic extraction is unduly pessimistic.

3 INDUCTANCE EFFECTS

The traditional time delay analysis, (1) does not take into account possible inductance effects. Inductance coupling effects on interconnects is an emerging concern in high performance digital integrated circuits. Inductance effects characterized by the loop inductance per unit length l become appreciable for global signals wires driven by low impedance drivers. It is mainly a high frequency concern. The serial impedance of a line $Z(\omega) = R + j\omega lL$ will become sensitive to inductance starting at a frequency ω where $\omega lL \approx R$.

The frequency content of a signal is dictated by its rise time T_{rise}, the signal spectrum will have appreciable Fourier components up to:

$$\omega_{max} \approx 1/T_{rise} \tag{6}$$

The effect first manifests on low resistivity signals: Clock signals and data busses. Clocks signals run on global upper metal layers, with reversed scaled thickness. Typical technology parameters for clock Cu wires are: $c = O(10^{-10})F/m$, $r = O(5000)\Omega/m$ and $l = O(10^{-7})H/m$. For a $50ps$ rise time achievable today at $130nm$ the reactance will become 50% of the resistance at the maximum frequency, obtained from (6) $20GHTz$. Notice that the reactance will in fact become the dominant component of the Impedance as we scale down towards $50nm$. This is due to the scaling of the maximum frequency: ω_{max} scales as S. It is therefore imperative that high performance digital clock wiring timing analysis includes RCL networks to describe these (and other) signal wires. A study of Clock signals in the high frequency domain including inductance is given in an accompanying paper [8]. Our approach takes a different twist, rather than searching configurations that minimize inductance, we search for configurations that guarantee transmission line behavior. Maximum speed propagation. We restrict ourselves to present in this talk some salient results. Davis and Meindl (DM) [9] identified the domain of electrical parameters that characterize transmission line behavior, that is to say linear delay with distance x,$\tau = \sqrt{lc}x$ with x the distance that a step pulse has traveled in time τ, in terms of l and c.

$$\frac{rL}{Z_0} \leq 2ln(\frac{4Z_0}{R_{tr}+Z_0}) \tag{7}$$

$$\frac{R_{tr}}{3} < Z_0 \tag{8}$$

R_{tr} the driver's impedance, $Z_0 = \sqrt{l/c}$, and r is the effective loop resistance per unit length of the wire ground configuration.

A second region characteristic of diffusion like behavior (rcL^2) is present when:

$$\frac{rL}{Z_0} \geq 2ln(\frac{4Z_0}{R_{tr}+Z_0}) \tag{9}$$

$$\frac{R_{tr}}{3} > Z_0 \tag{10}$$

Qualitatively the regions can be described as follows: For very short interconnect, the driver and the load dominates the delay, for very long interconnect the rcL^2 factor dominates. The intermediate region where inductance effects at high frequencies become appreciable happens on a window of lengths values in the $0.5mm \leq L \leq 10mm$. The size of the window depends on several geometrical factors discussed in the accompanying paper, and growths with scaling. Notice that the typical lengths are of the same order as L_{cr}. This has two ramifications namely, for wires whose length are near L_{cr} time delay computations and associated buffer insertion are not particularly reliable in the absence of rlc analysis. Moreover inductance effects can be used to advantage in properly sizing clock trees such that the time of flight regime dominates [8]. In other words the length scale over which delay is linear with length can be made to increase with the corresponding reduction in the number of required buffers, with positive side effects on power consumption. The l values that enter into the equivalent transmission line are the values that arise from computing the loop inductance of the equivalent circuit composed by the signal wire and the ground return path. For clock trees, a sensible layout representation consists of sandwich balanced tree structures. The return path of the the signal is localized to the closest ground wires. At variance with the extraction of capacitance that is short range in nature but computationally expensive, the extraction of inductance can be long range, but computationally simple. This simplification is particularly useful when we know the return path of the currents. It is therefore recommended for synthesis to stress the advantage of sandwich configurations (ground, signal, ground) to localize the current return path. Signal wires on buses in the same length regime can be treated in the same fashion, provided that

local ground wires are placed as sandwich configuration to minimize cross coupling and effectively control inductance. The main goal needs not to be to minimize inductance but rather to control it such that two end results occur: No overdrive, so as to minimize side failures, and this is controlled by the strength of the driver, and minimum cross coupling and this can be controlled by inclusion of nearby ground wires. Commercial microprocessor approach of including extra ground planes, are significantly less advantageous.

3.1 Data Reduction

The compactness of the representation in capacitance and inductance comes about through data reduction.The size of the database of r,l,c values, can be quite large, since at every change of environment for a wire a new multiplet of capacitance and inductance values must be added. A proper way to deal with the underlying complexity is to use model reduction techniques. For rc networks as well as for rlc networks that are not overdriven ($R_{tr} \leq Z_0$) AWE type methods that exploit the monotonicity of the response are appropriate [11].

4 CONCLUSIONS

A drastic evolution of complexity occurs in descending from from $250nm$ to $100nm$. See figure 3. We did not include the effects of inductance on noise in coupled signal lines, it can be significant. Neither did we treat the general case of geometries where length disparities would demand a full 3 dimensional treatment of rlc coupled networks. Such general treatment demands theoretical attention today. Finally the emergence of new regimes in wire behavior both in terms of lengths and relevant frequencies put a word of caution on reuse methodologies of IP.

REFERENCES

[1] D. Sylvester and K. Keutzer, *Rethinking Deep Submicron Circuit Design.* IEEE Computer, November 1999, pp. 25-33.

[2] J. Cong and L. He, *Theory and algorithm of local-refinement-based optimization with application to device and interconnect sizing.* IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol.18, no.4, pp. 406-420, April 1999.

[3] *The National Technology Roadmap for Semiconductors*, 1997.

[4] J. Davis et al., *Interconnect Limits on Gigascale Integration (GSI) in the 21st Century.* Proceedings of the IEEE - special issue on limits to semiconductor technology, Vol. 89, No. 3, March 2001, pp. 305-324.

[5] G. Golub and V. Pereyra, *The Differentiation of Pseudo-Inverses and Nonlinear Least Squares Problems whose Variables Separate.* SIAM J. Numer. Anal. 10, pp. 413-432. April 1973.

[6] M. Belleville and R. Suaya, *Methodology and Implementation of Interconnect of Cross Coupling and Total Wire Capacitance Measurements in SOI Technologies.* Journal of Modeling and Simulation of Microsystem. Special Issue 2002, to be published.

[7] A. Doganis and R. Suaya, *Statistical Validation Procedure for Interconnect Capacitance Modeling.* Journal of Modeling and Simulation of Microsystem. Special Issue 2002, to be published.

[8] R. Escovar and R. Suaya, *Optimal Design of Clock Trees for Leading Edge Microprocessors.* Journal of Modeling and Simulation of Microsystem. Special Issue 2002, to be published.

[9] J. Davis and J. Meindl, *Compact Distributed RLC Interconnect Models.* IEEE Transactions on Electronic Devices, Vol. 47, No.11, pp. 2068-2087, November 2000.

[10] D. Sylvester, *Measurement Techniques and Interconnect Estimation.* Proceedings of System-Level Interconnect Prediction Workshop, pp. 79-81, 2000.

[11] L.T. Pillage and R.A. Rohrer, *Asymptotic waveform evaluation for timing analysis.* IEEE Trans. CAD, vol 9, April 1990.

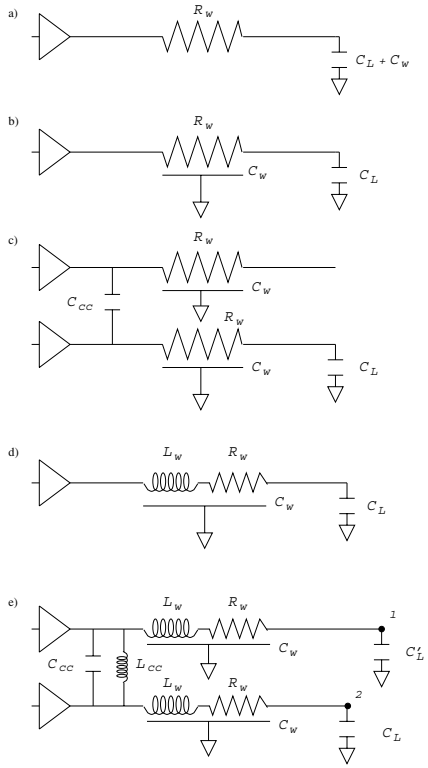

Figure 3: Relevant wire representations