Analysis of Buffered Hybrid Structured Clock Networks

Yi Zou, Qiang Zhou ,Yici Cai and Xianlong Hong

Department of Computer Science and Technology
Tsinghua University
Beijing, 100084, China

Sheldon X.-D. Tan

Department of Electrical Engineering
University of California, Riverside
CA 92521, USA

Abstract - This paper presents a novel approach for fast transient analysis of buffered hybrid structured clock networks. The new method applies structure reduction and relaxed hierarchical analysis methods to reduce the circuit complexity and speedup the simulation. A simple controlled sources model is used for modeling clock buffers to deal with nonlinearity in the buffered clock trees. Our experiment results show that the proposed algorithm is about two orders of magnitude faster than HSPICE without loss on accuracy and stability. The relatively errors on delay times are within a few percent of the exact ones.

I. INTRODUCTION

Clock distribution has become an increasingly challenging problem for VLSI designs，and design of clock networks is a critical step in physical implantation of high-performance VLSI circuits. The major issue concerning with clock network design is to achieve low delays and low skews. Tree-like clock networks have been widely used in the past to meet zero skew constructions [12]. However, as the new deep sub-micron technologies become more subject to variations, the portion of the clock skews introduced by the process variations on the wire width and the clock buffer can no longer be ignored. Comparing with tree-only structures, hybrid structures that incorporate both tree and mesh structures are more tolerant of process variations [1] and are becoming more popular in the topology design of the high-performance clock networks [9][13].

However, comparing with tree structured clock networks, a hybrid structured clock network that consists of both tree and mesh is more difficult for timing analysis and synthesis. The root of the complexity stems from the more complicated topology as more interconnects are present at cross nodes and a large number of loops exist, which makes traditional analysis method inefficient as analysis is in the inner loops of the synthesis processes. Moreover, as the clock frequency climbs to GHz range, the inductance effect can no longer be ignored especially when the chip has faster on-chip rise times and long wires behaves more like distributed transmission lines. Higher order RLCM interconnects models are typically used by analysis tools to obtain better accuracy. Furthermore, a large number of clock buffers are typically inserted in the clock network to balance load capacitance and trim the long interconnects. Clock buffers, which are typically composed of cascaded CMOS inverters, are nonlinear in nature, which further complicates the analysis process. With VLSI integration toward system on a chip and shrinking feature size below 100nm, the complexities of clock networks due to extracted parasitic become too large to be analyzed efficiently by current transistor-level SPICE-like simulation tools. A fast yet accurate analysis method beyond Elmore delay model is urgently needed for the design, analysis and optimization of the hybrid structured clock networks.

Elmore model is widely used for fast delay approximation. It is the first moment of the interconnect impulse response and can be computed for mesh structured circuits by finding the DC solution of the circuits [3, 6]. But moment matching methods cannot cope with active devices such as cascaded CMOS inverters thus many empirical models such as k-factor model [13] were used.

For hybrid-structured networks, transient simulations are required in general. High-order moment matching methods such as AWE [3], PRIMA [4] can be used to calculate propagation delay. But they can't deal with nonlinear circuits such as buffered clock networks and their efficiency will also degrade when coping with mesh-structured networks.

Another efficient way for speeding up transient analysis is by means of structure reduction [11]. It exploits the special structures of clock networks where trees and ladder chain circuit modeling the distributed transmission lines are commonly seen. But this method fails to consider clock buffers either.

In this paper, we propose an efficient method to analyze the buffered hybrid structured clock networks. Different from the existing approaches, the new approach explicitly considers the buffers in the clock trees. We model the clock buffers as voltage controlled current sources. The new method combines structure reduction method [11] with a hierarchical analysis scheme to speed up the analysis of reduced order mesh network. The resulting algorithm leads to a fast yet more accurate analysis for buffered hybrid structured clock networks.

The clock topologies we consider in this paper are

This work is supported by National Science Foundation of China No 60176016 and 90307017 and Hi-Tech Research & Development (863) Program of China 2002AA1Z1460 and UC Regent's Faculty Fellowship.

mesh-tree or tree-mesh-tree structures [8]. There are many underlying tree structures driven by the overlying mesh. Clock buffers are inserted in the underlying tree structures. In our structure reduction, each tree branch structure is reduced into an equivalent transient current source in parallel with a conductor. The nodes on the mesh structure just connecting to resistors and inductors are also eliminated by Y-Δ transformation. After the structure reduction, we partition the reduced grid into small local grids, which are connected via a global grid. At each simulation step, Nodal Analysis (NA) is formulated for the global grid and the local grids and they are solved by preconditioned conjugate gradient iterative method [7], which shows an almost linear performance in practice. Clock buffers laying at the underlying tree structures are modeled as voltage controlled current sources to drive the fan-out gates and interconnects. Later the delay time between the sink nodes can be obtained by interpolation. Experiments results show that our method demonstrates two orders of magnitude faster than HSPICE with the marginal errors.

We organize this paper as follows: Section II provides our simulation model for interconnects and clock buffers and the formulation of Nodal Analysis in the time domain. Section III gives a brief review of the model order reduction via equivalent circuit model. Section IV describes the hierarchical analysis scheme for overlying mesh. Section V describes the incorporation of clock buffers into the simulating process. At the end of this paper, experimental results and conclusion are presented.

II. Description of Simulation Model

A. Hybrid Structured Clock Network

Our topology is a tree-mesh-tree hybrid structure which is very suitable in high-performance SOC technology [8], where a clock source which is treated as a step or ramped voltage source in transient simulation drives a global H or X tree which directly drives the mesh structure. And the local trees get the clock signal from the mesh structure and delivery clock signals to the sink nodes. Clock buffers are inserted in the bottom of the tree structures to balance the load capacitance and reduce the clock tree delays. The clock drivers at top-level tree structures will be considered. Figure 1 shows a tree-mesh-tree hybrid clock distribution network.

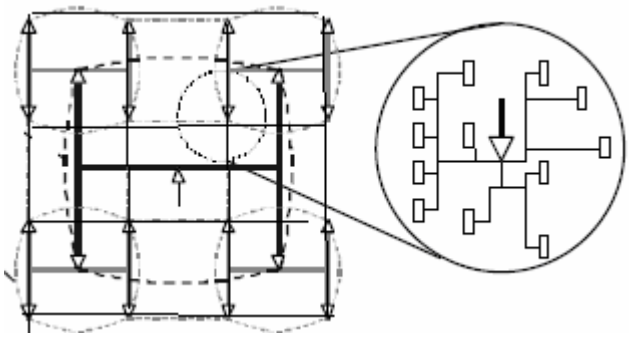

Fig.1. Tree-Mesh-Tree Clock Distribution Network

B. RLC Simulation Model for Interconnects

We use dumped RLC circuits to model the electromagnetic property of clock network. This model treats distributed parameters of interconnect wire between any two nodes in the topology as serial connected resistor and inductor with capacitor between the wire and ground. With sufficient RLC ladder sections, the remission line effects will be modeled accurately enough.

B. Model of Clock Buffers

In our implementation, we use cascaded CMOS inverters to form clock buffers.

For a typical inverter, Spice Level 1 model is used to calculate the channel current of MOSFET. Here is the expression for channel currents:

$$\text{Ids} = \begin{cases} 0 \ (\text{Vgs}<\text{Vth}) \\ \text{K}(\text{Vgs} - \text{Vth} - 0.5\text{Vds})\text{Vds} \ \ (\text{Vds} < \text{Vgs} - \text{Vth}) \\ 0.5\text{K}(\text{Vgs} - \text{Vth})^2 \ \ \ (\text{Vds} >= \text{Vgs} - \text{Vth}) \end{cases} \quad (1)$$

In order to get the output current of a CMOS inverter from the channel currents

$$I_p - I_n = C_M\left(\frac{dV_{in}}{dt} - \frac{dV_{out}}{dt}\right) + C_L\left(\frac{dV_{out}}{dt}\right) \quad (2)$$

Here the C_M is the capacitance between gate and drain including gate-to-drain overlap capacitance etc. C_L stands for the load capacitance or the equivalent capacitance of the fan-out gates and fan-out RLC interconnects.

The output current of the CMOS inverters can be derived by the above equation:

$$I_{out} = C_L\left(\frac{dV_{out}}{dt}\right) = \frac{C_L}{C_L + C_M}(I_p - I_n) + \frac{C_L C_M}{C_L + C_M}\left(\frac{dV_{in}}{dt}\right) \quad (3)$$

The utilization and integration of the model of the clock buffers in the simulation process is described in section V.

Note alpha-power model and other statistical or analytical model of MOS gates can also be used in this simulation framework to obtain a better accuracy in sub-micro technology.

III. Structure Complexity Reduction

A. Transient Nodal Analysis

We first transform the differential equations into algebraic equations by using trapezoidal discretization method: Suppose h be the analysis time step. For capacitors and inductances, their finite difference equations are

$$I_{c,k+1} = \frac{2C}{h}V_{c,k+1} - \frac{2C}{h}(V_{c,k} + I_{c,k}) \quad (4)$$

$$I_{c,k+1} = \frac{h}{2L}V_{c,k+1} + \frac{h}{2L}(V_{c,k} + I_{c,k}) \quad (5)$$

The two equations lead to Norton companion models of L and C components shown in figure 2.

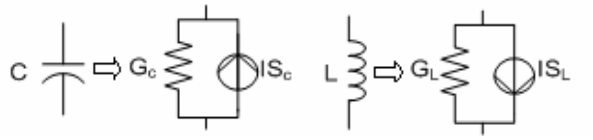

Fig.2. Linear Model of L and C Components.

After replacing all the capacitances and inductances in the network with their linear comparison models, a pure resistor equivalent network is obtained. We can then use Nodal Analysis to formulate the resulting resistor network. Note that the transient conductance of L and C shown in equation (4) and (5) is stamped in G_{nxn} below and transient current of L and C is stamped in the right-hand side (RHS) of the equations (6).

$$G_{nxn} \cdot V_n^t = I_n^t \quad (6) \qquad g_{ii} = - \sum_{j=1, j \neq i}^{n} g_{ij} \quad (7)$$

For a resistor-only network with current sources, the resulting circuit matrix formatted by NA is symmetric positive definite, which will be exploited by iterative linear solution approach shown later.

B . Topology reduction methods

After stamping R and L components as their correspondent linear models, topology reduction methods [11] are first employed to reduce the problem size of nodal analysis.

The basic idea of topology reduction methods is to transform the bottom of the tree structures or chain structures etc into a equivalent conductance and current source by local Y-Δ transformation.

Circuits that have parallel or series connected elements named type 1 and 2 can be simplified to equivalent ones by simple algebraic operations and these two kinds of circuits can change to each other, as shown in Figure 3. .

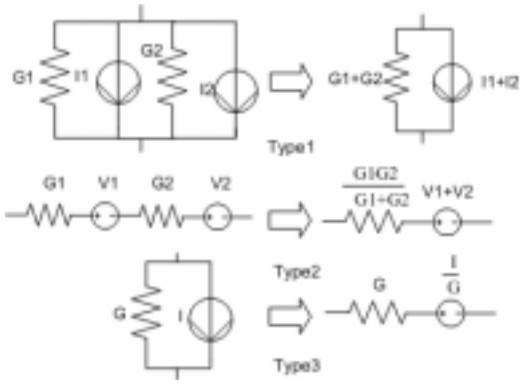

Fig.3. Reduction of series or parallel components.

Based on the reduction of series or parallel components, the tree structure can be easily simplified. Take a typical reduction of a tree branch for example. First, we calculate the bottom conductance using operations for parallel connected shown Figure 4. Then we calculate the conductance using operations for series connected components twice. The current source or voltage source is also simplified in the process. Note that the simplification process goes from the leaf nodes of the tree to the root nodes on the mesh structure. After the simplification, a tree branch is reduced to a current source which can be stamped into the RHS of nodal analysis and a parallel connected resistor which can be stamped into matrix G_{nxn} of the reduced network. Figure 4 shows the reduction of a tree branch.

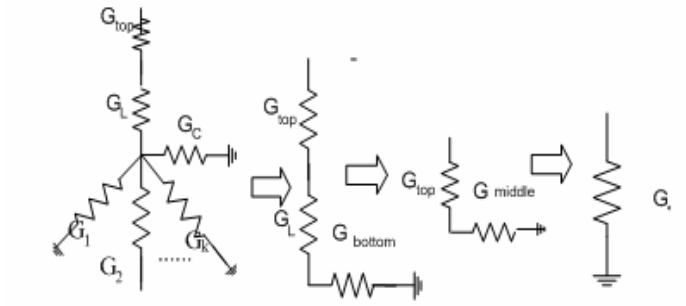

Fig.4 Reduction of a tree branch.

The idea can also be applied to chain circuits using Y-△ transformation and R-L nodes in the mesh structure [11].

C. Topology Reduction of buffered network

Clock buffers are inserted in the underlying tree structure to balance load capacitance and trim the long interconnects. The original tree structure is thus partitioned into many sub-trees by the clock buffers.

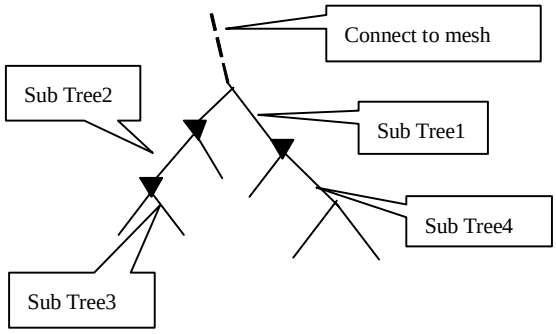

Fig.5. Segmentation of the tree by clock buffers.

We define the *upstream sub-tree* of a clock buffer in which a node directly drives that buffer. Also we define the *downstream sub-tree* for a buffer the sub-tree in which a node is directly driven by that buffer. Figure 5 shows the segmentation of the tree by clock buffers. For the upstream sub-tree of one clock buffer, the load input capacitance that can be calculated by SPICE parameters [13] is modeled for the topology reduction of the upstream sub-tree. Note that a sub-tree can only be driven by one buffer due to the property of the tree structure. We perform the topology reduction for each sub-tree consequentially. The reduction also goes from the leaf nodes of the sub-tree (including the nodes which drive the buffer) to the root node of the sub-tree, which is

also the buffer driven node.

After the reduction, the sub-tree, which directly connects to the mesh structure, is reduced to a current source and conductance connected in parallel at the root node of the sub-tree. The downstream sub-tree of a buffer is reduced to a current source and conductance connected in parallel at the buffer driven node of the sub-tree.

Note that the refreshing of the current source depends on the nodal voltage of each node. The computation of the nodal voltage of a buffered network is discussed in section V.

IV. HIERARCHICAL ANALYSIS

The topology reduction proposed in part III aims to provide reduction in the underlying tree structure. However, in some circumstances, there is a large clock mesh driving few local trees, so the reduction rate could be small. To further improve the simulation efficiency, we further explore the regularity of mesh structures by using a hierarchical analysis method, which makes use of partitioning and divide-and-conquer strategy.

Nodal analysis for a reduced mesh network permits preconditioned conjugate gradient (PCG) method to solve the linear equations. The sparsity of the coefficient matrix makes the PCG algorithm a close-to-linear performance. However, the performance of the algorithm will degrade when the number of nodes is very large. We propose to use partitioning and divide-and-conquer strategy to further reduce the computational and memory cost.

A. Topology Partition

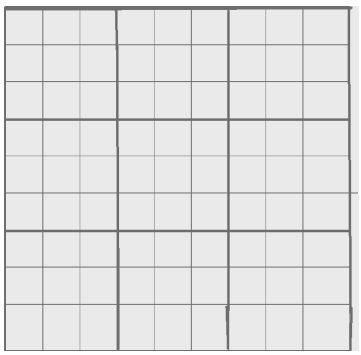

Fig.6. Global Grid and local grids

First, we partition the remaining grid into a global grid and many local grids, just as shown in the figure 6.The thick lines represents the global grid while the remaining grids are local grids. The nodes in the global grid partitions the whole interconnect graph into many portions of local grids such that all the nodes in the local grids are surrounded by the nodes in the global grid. As a result, we can clearly see that there are no direct connections between different local grids. So the local grid can only connect to itself or to global grid. We use this property to speed up the solving of the system equations of the nodal analysis of the whole grid.

B. Partitioning of the System Equations

When the nodal analysis equation of the whole reduced mesh topology after partition is expressed in the formula (6), we transform the expression into a block matrix as shown in (8). The matrix block expressing the connection between different local grids is blocks containing only zero.

$$\begin{pmatrix} G_{00} & G_{01}\cdots & G_{0k} \\ G_{01}^T & A_1\cdots & 0 \\ \vdots & \ddots & \ddots \\ G_{0k}^T & 0\cdots & A_k \end{pmatrix}\begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_k \end{pmatrix}=\begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix} \quad (8)$$

In this equation, partition *0* stands for global grid, block matrix G_{0i} stands for the nodal equivalent admittance between global grid and local grids numbered *I*, A_i stands for the nodal equivalent .admittance matrix of local grid numbered *i*. *x* stands for the solution vector and *b* stands for the RHS of the nodal analysis equation.

C. Hierarchical Relax Method

The basic idea of our hierarchical relax method is to solve the global grid and local grid sequentially as shown in Fig.7. The global grid gives out the boundary condition of the local grids, and after solving out the nodal equations for the global grid, the equations for the local grid can hereafter be solved.

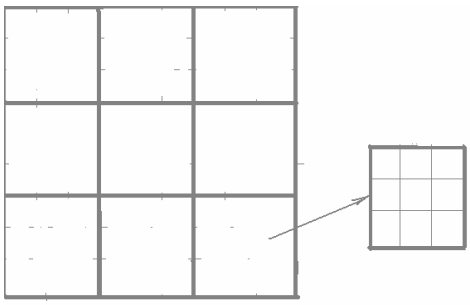

Fig.7. Solve the global grid and local grid sequentially

There are many ways to solve the grid hierarchically. If we solve the block matrix expression (8) using block Gauss Elimination, the original network is reduced as

$$(G_{00}-\sum G_{0i}A_i^{-1}G_{0i}^T)x_0=(b_0-\sum G_{0i}A_i^{-1}b_i) \quad (9)$$

$$x_i=A_i^{-1}(b_i-G_{0i}^T x_0) \quad (10)$$

The major difficulty of expression (10) lies in the inversion of matrix A_i, which is very time-consuming. Besides, the inversion of matrix A_i will generate a full matrix, which is not memory efficient. Thus a method without explicitly using matrix inversion is more desirable.

If we shift the block matrixes, which are not among the diagonal blocks into RHS, the following equations are

derived:

$$G_{00}x_0 = b_0 - \sum G_{0i}x_i \quad (11)$$

$$A_i x_i = b_i - G_{0i}^T x_0 \quad (12)$$

If we could accurately know the solution vector, the shifting will not introduce any model error. However, we cannot know the solution vector before solving at the simulation step. A typical strategy is to use the solution vector of the one simulation step before the current step as the initial solution for current time step.

$$G_{00}x_0^{t+h} = b_0 - \sum G_{0i}x_i^t \quad (13)$$

$$A_i x_i^{t+h} = b_i - G_{0i}^T x_0^{t+h} \quad (14)$$

This actually is a relaxation process for hierarchical analysis. Comparing with the sparsification method in the hierarchical method proposed in [14], the error that may incurred due to sparsification method is proportional to the absolute value of the solution vector while our relax method is proportional to the difference between solution vector at two subsequent steps. Also the error can be further reduced if more iteration is carried out.

D. Using Multiple-Step Method to Reduce Model Error.

Since the voltage waveform we capture will typically be a rising or a falling waveform, using the solution vector of the previous step to calculate the nodal voltages of the current step in the global grid will cause the model error to accumulate. We then use interpolation of the solution vectors of several previous simulation steps to forecast and approximate the current solution vector, and use this vector to finish the solving of the linear system of the global grid and the local grid respectfully. For example, a two-step forecast method is demonstrated as follows:

$$G_{00}x_0^{t+h} = b_0 - \sum G_{0i}((1+\beta)x_i^t - \beta x_i^{t-h}) \quad (15)$$

$$A_i x_i^{t+h} = b_i - G_{0i}^T x_0^{t+h} \quad (16)$$

β is a parameter to demonstrate the linearity of the output vector, if the linearity is good then β can be larger. In our implementation, we choose β =1 for simplicity.

Besides, the solving of the local grids can be carried out in parallel (in multiple processors and network of stations platform) in order to increase the computational efficiency. However, unlike the hierarchical method proposed in [14], the proposed algorithm is also about 2 times faster than PCG algorithm without parallel implementation while the algorithm in [14] shows slightly performance degradation when running in single CPU.

Note that the multiple iterations can also be used to increase the accuracy of the hierarchical analysis but the analysis time will rise as the number of iterations increases.

V. DEALING WITH CLOCK BUFFERS

The node reconstruction process [11] transfers the nodal voltages of each simulation step in the reduced network into the original network using the information collected in the simplification process and the transient current is captured for the next simulation step. For a buffered network, we need to transfer the nodal voltages through the clock buffers.

A. Using Inverter Model In Time Domain

Equation (3) gives the output current of a CMOS inverter that is the element of clock buffers. At each simulation step, the input voltage is already obtained by nodal analysis of mesh structure and its driven interconnects or the analysis of the buffer driven interconnects in the upper level. Also we use the output voltage in the previous simulation step to calculate the channel currents.

The capacitance between gate and drain can be calculated by device parameters according to [13] or looked up by a pre-constructed table for buffers. And the load capacitance of the fan-out gates and interconnects can be calculated via the difference equation of the definition of capacitance.

$$C_{tran}^t = \frac{I_{out}^t}{\frac{dV_{out}^t}{dt}} = \frac{h\bullet(I_{out}^t + I_{out}^{t-h})}{2(V_{out}^t - V_{out}^{t-h})} \quad (17)$$

In some circumstances, the load capacitances are much larger than the capacitance between gate and drain, thus the expression (3) can be approximated as follows:

$$I_{out} = C_L(\frac{dV_{out}}{dt}) = (I_p - I_n) + C_M(\frac{dV_{in}}{dt}) \quad (18)$$

This equation is convenient since the calculation of the transient load capacitance will not be needed.

B. The Analysis of Clock Buffers Driving Interconnects

Since clock buffers shield the lower level interconnects and gates, we only need to consider the input capacitance of the clock buffer when we are analyzing a buffer driven interconnect.

First we perform the topology simplification of the fan-out interconnects as in section III, for a tree-structured interconnect, an equivalent conductance Geq and equivalent circuit Ieq can be obtained after topology simplification. Then we can use the output current I_{out} to drive the simplified load and calculate the output voltage via the following equation:

$$G_{eq}V_{out}^{t+h} = I_{out} - I_{eq} \quad (19)$$

The voltages of other nodes are calculated via node reconstruction process.

Note that if the clock buffer or driver is driving a non-tree topology on an upper level, we can simply stamp the output current of the clock driver into the RHS of the system equation of the reduced mesh network.

VI. EXPERIMENT RESULTS

Table 1 Caparison with HSPICE in terms of speed and accuracy

Mesh Size	Total Nodes	Running time		Speedup	Delay of test Node (Spice)	Delay of test Node (Proposed Algorithm)	Relatively error
		HSPICE	Proposed algorithm				
10*10	1166	25.22	0.55	45.85	2.027	2.041	0.69%
20*20	1822	175.83	4.45	39.51	8.082	8.068	0.17%
50*50	23135	2250	46.32	48.57	28.72	28.616	0.36%
100*100	148332	NA	510.17	NA	NA	NA	NA
150*150	401091	NA	2839.0	NA	NA	NA	NA

Our algorithm is implemented in C language. Statistical information is shown in table 1. All the experimental results are obtained on Sun Ultra Sparc workstation with 4 250 MHz CPUs, 8 GB memory. Elmore model for hybrid structure is used to capture a worst-case maximum delay, which set the upper bound for simulation time. Simulation step are determined according to the desired accuracy. We analyze the circuit till the delay times of all the sink nodes are captured.

Test circuits of hybrid structured clock network are generated randomly, clock buffers are randomly inserted at the merging node of children branches. We analyze the test circuit of different sizes to demonstrate its scalability of the proposed method. From the results we can clear see that the proposed algorithm has 40X speedup over HSPICE with marginal errors. For the running time of test circuits, we record both the time of simplification and the time of matrix building to make a fair comparison with HSPICE.

VII. CONCLUSIONS AND FUTURE WORK

This paper presents a novel approach for efficient analysis of buffered hybrid structured clock networks. The new method combines both structure reduction and relaxed hierarchical analysis methods to reduce circuit complexity and speedup the simulation. A simple controlled sources model is used for modeling clock buffers to deal with nonlinearity in the buffered clock trees. Our experimental results show that the resulting algorithm gives about two orders of magnitude speedup over HSPICE for analyzing buffered clock trees with marginal accuracy loss.

In the future, we would like to take power-supply variation impacts on the hybrid structure into consideration, which can have a strong influence on clock skew verification. Also equivalent waveform concept [15] and look-up table model for clock buffer could be incorporated to speed up the calculation and ensure the compatibility.

REFERENCES

[1] H. Su and S. S. Sapatnekar, "Hybrid structured clock network construction", Proc. ICCAD 2001, pp. 333-336

[2] P.J. Restle, et al, "A clock distribution network for microprocessors", IEEE Journal of Solid-State Circuits, Vol.36, No. 5, May 2001 pp. 792-99

[3] L. T. Pillage and R. A. Rohrer, "Asymptotic waveform evaluation for timing analysis", IEEE Trans. on Computer-Aided Design，vol. 9，No. 4 , April 1990, pp.352-366

[4] A. Odabasioglu, M. Celik and L. T. Pilleggi, "PRIMA: Passive reduced-order interconnect macromodeling algorithm", IEEE Trans. on Computer-Aided Design，vol. 17，Aug. 1998, pp. 645-654.

[5] Y. I. Ismail. "Improved model-order reduction by using spatial information in moments"，IEEE Transactions on VLSI Systems, vol. 11, Oct. 2003, pp. 900 – 908

[6] M. P. Desai, R. Cvijetic, J. Jensen, "Sizing of clock distribution networks for high performance CPU chips", Proceedings DAC, June, 1996 pp.389-394,

[7] T. Chen and C.C. Chen, "Efficient large-scale power grid analysis based on preconditioned Krylov-subspace iterative method", Proc. DAC, pp. 559-562, June, 2001.

[8] C. C. P. Chen, and E. Cheng "Future SOC design challenges and solutions", Proc. International Symposium on Quality Electronic Design, 2002, March 2002, pp. 534 – 537.

[9] P. J. Restle, and A. Deutsch, "Design the best Clock network", Proc IEEE VLSI Circuit Symposium 1998 pp. 2-5

[10] D. Kucar, and A. Vannelli, "Interconnection modeling using distributed RLC models", Proc IEEE international workshop on System-on-Chip for Real-Time Applications, July 2003, pp. 32-35

[11] Y. Zou, Y. Cai, Q. Zhou, X. Hong, S. X.-D. Tan, "A fast delay analysis algorithm for the hybrid structured clock network", in *Proc. Int. Conf. Computer Design* (ICCD), pp. 344-349, 2004.

[12] K. D. Boese and A. B. Kahng， "Zero skew clock routing trees with minimum wire length", Proc. IEEE Int. Conf. ASICs，Sept. 1992, pp.1.1.1-1.1.5

[13] N. H. E. Weste and K. Eshraghian, Principles of CMOS VLSI Designs, 2nd Edition, Addison-Wesley Publishing Company, 1993

[14] M. Zhao, R. V. Panda, S. S. Sapatnekar, T. Edwards, R. Chaudhry, D. Blaauw "Hierarchical analysis of power distribution networks", Proc. DAC, 2000., June, 2000, pp150-155

[15] M. Hashimoto, Y. Yamada, Y, H. Onodera, "Equivalent waveform propagation for static timing analysis", Proc ICCAD-2003. Nov. 2003, pp. 169-175