Scalability Study of PSANDE: Power Supply Analysis for Noise and Delay Estimation

Sushmita Kadiyala Rao, Bharath Shivashankar, Ryan Robucci, Nilanjan Banerjee and Chintan Patel

CSEE Department, University of Maryland, Baltimore County

Abstract

Variations in the power-distribution network are exacerbated because of scaled supply voltages and smaller noise margins in sub-nanometer designs, which adversely affect performance and yield. Power-Supply noise incurred by excessive simultaneous switching of multiple paths negatively impacts the timing of a circuit. Supply noise is a major issue especially during transition and delay test where test vectors cause increased switching as compared to functional operation resulting in increase in path delays. Test rejects due to excessive noise-induced failures during delay and transition testing negatively impacts yield. Hence there is a need to accurately characterize the resistive and inductive voltage drop caused by excessive switching. To our knowledge, inductive drop has been excluded to simplify noise analysis. In our previous work, we have presented a convolution-based dynamic method (herein referred to as PSANDE) to estimate both IR and Ldi/dt drop on small combinational and sequential circuits. In this paper we show that the effectiveness of the design partitioning technique makes the framework feasible for a larger design. Our dynamic approach involves selectively simulating only extracted switching logic which makes the run-time tractable as compared to prohibitive full-chip SPICE simulations. We also present data to show that PSANDE can accurately predict the power-supply noise due to clock tree switching. Data presented in this paper for power supply noise is based on a large ITC'99 sequential benchmark b17 circuit. with a maximum error of 8.2% in comparison to full-chip SPICE results.

Keywords: Power-Supply Noise, Transition & Delay Testing, IR-drop, Ldi/dt drop, PSANDE

1.0 Introduction

Power-Supply noise continues to plague deep submicron technologies especially during delay testing. Simultaneous switching of multiple paths can degrade the supply voltages affecting circuit performance. The variations in supply voltages increases path delays that can sometimes lead to timing failures. In particular, during test mode extensive switching is incurred by the test vectors that may not occur in functional mode. These test conditions cause a larger voltage droop that results in false failures increasing yield loss. Techniques are needed to accurately model supply noise and the ill-effects it has on path delays so that test vector generation and compaction can be based on it that can alleviate test-related yield loss due to false positives.

Power supply noise (PSN) refers to the variations caused in the supply voltages in the power-distribution network due to considerable switching activity in the circuit. PSN comprises the IR-drop caused by parasitic resistors of signal wires in a circuit and the voltage drop caused by package and parasitic inductance known as Ldi/dt drop. Inductive drop is often excluded to simplify noise analysis leading to inaccuracies in supply noise estimation. In this work we show that the dynamic approach, PSANDE [1][2][3] can be scaled to estimate power supply noise for large circuits. We also predict the contribution of power supply noise from the off-path clock elements.

A survey highlighting the causes and effects of PSN in delay testing is presented in [6]. In [7] a statistical analysis is used to study the impact of IR-drop on critical path delays. Authors in [8] propose an event-driven simulation technique to estimate the worst-case noise where they use genetic algorithms to optimize the noise problem.

The authors in [9] claim that on-chip inductance can be ignored for most of the frequencies since at high-frequencies the effects of inductance are localized in nature and comparatively smaller. But in [10] the authors show that the on-chip self inductance caused the voltage drop to deviate significantly, by more than 50% over that of the IR predicted drop. In [11] the effect of both self and mutual inductance is illustrated from the simulation results of a 4-bit coupled bus. Results show that error for delay predictions can be more than 100% and that for rise time computation can be greater than 70% when the effect of inductance is ignored.

Test pattern generation flow has also been affected by PSN. Authors in [12] predict that the test industry will move towards newer ATPG methods, one being the usage of partition-aware ATPG techniques, by the year 2020 due to the design sizes that increase with shrinking technology. The overkill induced by random filling of don't-care bits during test compaction has been addressed by authors in [13]. Lee et al. [14] proposed a layout-aware test compaction technique such that transition test patterns cause switching activity that is evenly distributed across the chip. In [15] Ma et al. use a layout-aware test pattern generation method to increase switching in the neighborhood cells of the critical path gates to maximize supply noise effects on the critical path under test. The authors in [16] present a fast technique to validate test vectors in the presence of noise caused by IR-drop. In [17] the authors have proposed a virtual circuit partitioning technique in which only test patterns that cause high power dissipation are applied to the subcircuits while the low power test vectors are applied at the full circuit level.

In [18] the authors use genetic algorithms to generate a small set of test patterns that cause higher power-supply noise and use a lower-level simulator to validate them and find the worst case test vector. The authors in [19] propose iterative algorithms to solve linear equations of the power grid that are far more accu-

rate and faster than the random-walk-based algorithm. In [20] the authors have proposed multigrid-like techniques where the grid is reduced for power grid analysis. PSANDE is unique in its focus on prediction of PSN caused by simultaneous switching activity. Using this framework we can accurately predict the increase in path delays due to power consumption in the vicinity of critical paths (or any path) and therefore we can supplement the techniques described above.

2.0 Power Supply Analysis for Noise and Delay Estimation (PSANDE)

In previous work [1][2][3], we have established a fast and accurate convolution-based approach to estimate power-supply noise caused by simultaneous switching of multiple paths. We have also predicted the increase in path propagation delays induced by PSN with SPICE-like accuracy. The basis of PSANDE framework is separating the power grid from the core logic circuitry in any CMOS chip. Analysis of each of these subsytems is explained briefly below.

2.1 Power-Grid Characterization

Since the power grid can be modeled as a linear system, it can be characterized by its Impulse Response (IR) functions. Grid simulations are carried out to compute IRs between input and output locations of the power grid. Inputs are points where standard cells connect to the power ground network and outputs are either the power pads of the chip or any point where standard cells connect to the grid. The Current-to-Current impulse responses (C2C IR) provide the relationship between the current source applied at any input and the corresponding currents measured at any output location on the power grid. For the purpose of this paper we aim to estimate the total transient current at the power pads, though as described in our previous work we can characterize both current and voltage IR responses to any input/output node in the system. Therefore the current responses are measured only at the power pads or C4 bumps. The IR computation is shown in Figure 1.

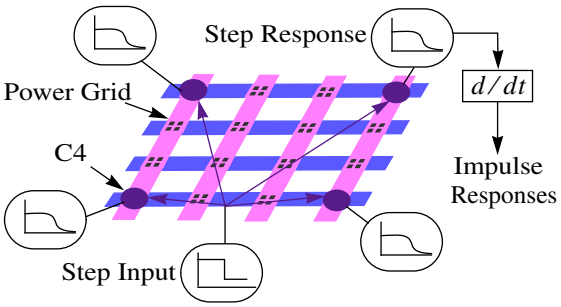

Figure 1. C2C Impulse response [1]

As seen in the figure, a current step input is applied to an input and the output step responses are measured at the power pads are differentiated to compute the C2C IR functions. This pre-characterization of the power grid is carried out only once for a design and accelerates computation.

2.2 Core-Logic Circuit Evaluation

In our work, we carry out isolated path simulations that are much faster in comparison to full-chip simulations accelerating computation. The Value Change Dump (VCD) file obtained from the Verilog simulation is used to identify gates of a switching path. These gates are then extracted from the full-chip circuit netlist with RLC parasitics. Next, the isolated paths are simulated with ideal V_{DD} voltages and the corresponding current transients at the power supply pins of each logic gate in the path are saved. This transient current process is repeated for every test pattern.

To estimate the power supply waveforms at the power pads, the current transients saved during path simulations are convolved with the impulse responses computed during the grid characterization process. Previously, we have published results for estimating PSN and delays for small combinational and sequential benchmark circuits that consisted of about 2000 logic gates.

In this paper, we have applied the PSANDE framework to a much larger sequential benchmark circuit. We have shown that the power supply noise scales well for large designs using the partitioning scheme. Previously we used test circuits that consisted of multiple instances of the c6288 design. Hence, the design partitioning technique divides the design into identical copies of the c6288 circuit. In this work, the partitioning scheme results in several non-identical partitions. These unique partitions are evaluated to estimate power supply noise. We also show that the convolution-based PSN estimation method can be used efficiently to predict the current transients contributed by the on and off-path sequential elements of the clock tree.

This paper presents three novel research contributions that are described in the following sections

1. *Macro-level (full-chip) Design Partitioning into micro-level subsections towards fast full-chip PSN estimation*
2. *Superposition and redistribution techniques to predict macro-level currents using only micro-level simulations*
3. *Accurately quantifying the effect of clock tree switching on full-chip PSN*

3.0 Macro-Design Partitioning Strategy

In this work, we use the ITC'99 b17 sequential benchmark circuit to study the scalability of the PSANDE framework. The b17 circuit is a combination of three 80386 processors (subset) that communicate with each other and to primary inputs/outputs. The layout is implemented in 180-nm technology using a place-and-route tool. The circuit netlist with extracted RLC parasitics consisting of 318,006 transistors; 2,828,786 resistors; 22,721,081 capacitors; and 338,056 parasitic inductors is representative of a large circuit block by today's standards. The circuit is subdivided into nine sections u1-u9 based on the power pad placement as shown in Figure 2. A section of the chip encompassed by power pads or C4s colored in dark orange is

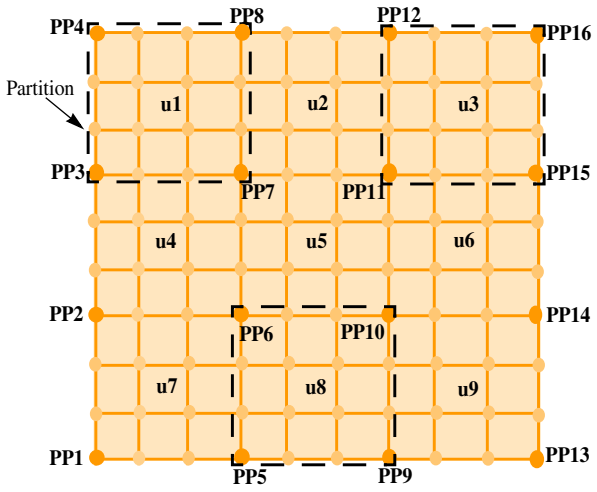

Figure 2. b17 Partitioning

defined as a partition. Current measurements are taken at 16 power pads named PP1-PP16. The SPICE netlist for each of the partition is created using the DEF (Design Exchange Format) file and the full-chip netlist of the b17 design with RLC parasitics.

4.0 Micro-Design Partitioning

The b17 design is broken into the power grid and core logic subsystems. The partitioning is performed at the metal connections between the local V_{DD}/GND and global supply rails called follow-pins. The two independent subsystems are characterized using the grid impulse response and gate current transient computations.

4.1 Grid Impulse Response Computation

In our previous work, we characterized the power grid as a single entity. However, for large circuits even a linear characterization of the power grid would be infeasible. One of the major contributions of this work is to demonstrate the scaling capabilities of PSANDE by partitioning the power grid (which is already separated from the non-linear core logic subsystem) for C2C IR analysis. The number of inputs for the power grid IR computation for each partition is shown in Table 1. Grid simulations are run for each of the partition to obtain the C2C impulse responses for each input in the partition. To reiterate, this pre-characterization process is performed only once for a given design and can be run in parallel. This one time characterization of the power grid can be used to predict full-chip response for *any given input from the core logic circuit* i.e switching paths.

Partition u_X and Number of Inputs					
u1	7370	u4	6908	u7	7093
u2	6799	u5	5724	u8	4658
u3	6212	u6	4697	u9	2866

Table 1: Inputs to the Linear Power Grid for each Partition

4.2 Gate Current Transients Computation

A switching path is identified from the VCD output obtained from the Verilog simulation of the b17 design. The current transient from each switching logic gate is an input to the pre-characterized linear power grid circuit. As the design is partitioned using the placement of power pads, a switching path is also divided into several subpaths across all partitions. This step identifies subpaths that are encompassed by a given partition. Each subpath is then individually simulated in parallel to produce the current waveforms at the power supply pins of the logic gates which are the inputs to the power grid. These subpath simulations take only a few minutes, in contrast to full-chip SPICE simulations which are run-time intensive and therefore prohibitive.

5.0 Comprehensive PSN Estimation

Using the power grid IR characterization and subpath simulations we discuss the technique for estimating full-chip PSN. This involves using logic gate power supply inputs per partition to compute current transients at micro-level power ports and then superimposing and redistributing these transients to predict macro-level full-chip PSN. The gate-level power current transients saved during simulations of each subpath are convolved with the C2C IRs and this step is repeated for all subpaths switching in a given partition. The cumulative sum of current waveforms estimated for all subpaths of a partition result in the micro-level power pad current waveforms for each partition. These partition current waveforms are then superimposed accordingly to generate the current waveforms at the macro-level power pads of the b17 design. For example, to estimate the total macro-level current transient for PP6 shown in Figure 2, we need to superimpose and redistribute currents from each partition. For this estimation, at the micro-level we superimpose currents from upper-left PP of u8, lower-left PP of u5, lower-right PP of u4 and upper-right PP of u7. Additionally we need to redistribute the superimposed micro-level currents to determine the final macro-level PP6 current transients as explained in Section 7.0.

6.0 Clock Tree Transient Estimation

Partitioning the chip based on the power pad placement also necessitates characterizing a fragmented clock tree. In this work we demonstrate a novel technique for predicting the current transients attributed to the global clock signal. It enhances our preliminary method that was previously described in [1], that shows results for clock-tree current transient estimation. Our new approach addresses the scalability of predicting the PSN contributed by the global clock tree as it is impractical to simulate the entire clock distribution circuit of a large design. Therefore we divide the clock tree based on our micro-level partitions and specifically differentiate between clocked elements within a switching path and non-switching sequential elements.

The transients caused by the sequential elements that switch are included in the path extractions and captured during path simulations. Also we must account for the current transients generated by non-switching off-path

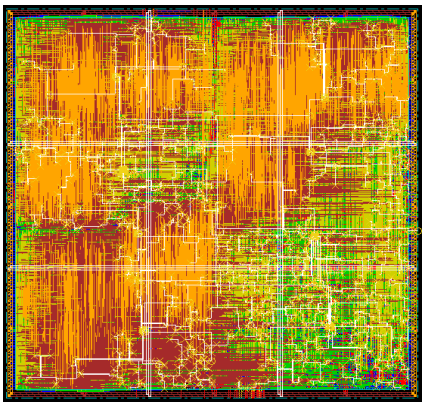

Figure 3. Clock-tree Highlighted (in white) in b17

flip-flops and clock buffers as they are not included in the switching path simulations. We estimate the current transients from non-switching sequential elements and include it along with micro-level computations in the PSANDE framework to estimate the full-chip PSN.

Highlighted in Figure 3 is the entire clock tree of the design.2

7.0 Signal Redistribution Technique

The current waveforms computed at each power pad at the micro-level are superimposed to get an initial estimate of the current transients at the macro-level PPs. Superposition of micro-level current transients only encompasses the effect of switching activity from partitions that a particular PP belongs to. However other power pads on the chip provide current to the PP under consideration that we need to account for. This is accomplished by estimating contributions from other PPs (not included at the micro-level) using a redistribution technique described in [4] and [5]. The technique involves applying an AC current step input at the PP under consideration, measuring the response at all other PPs, creating a transfer-function matrix and using it to redistribute the current contributions. This process is repeated for each PP (under consideration). Each row in the transfer-function matrix refers to one measurement (i.e. a current source applied to one PP at a time) and the columns represent the current measured at all other PPs. Each element in the matrix is normalized by the average of the DC levels of the measured current transient.

8.0 Results

As mentioned in Section 3.0, the ITC'99 b17 benchmark circuit, is used for all the simulations in this work using the Cadence UltraSim simulator. Results for the three major contributions of this work are presented in this section that demonstrate the scalability of the PSANDE framework.

Test Pattern	Combinational Elements	Sequential Elements	Total
Path1	6274	258	6532
Path2	6277	266	6535

Table 2: Switching Gates Count

Test patterns were generated using a commercial ATPG tool and PSN was estimated. The number of both sequential and combinational elements that switch in

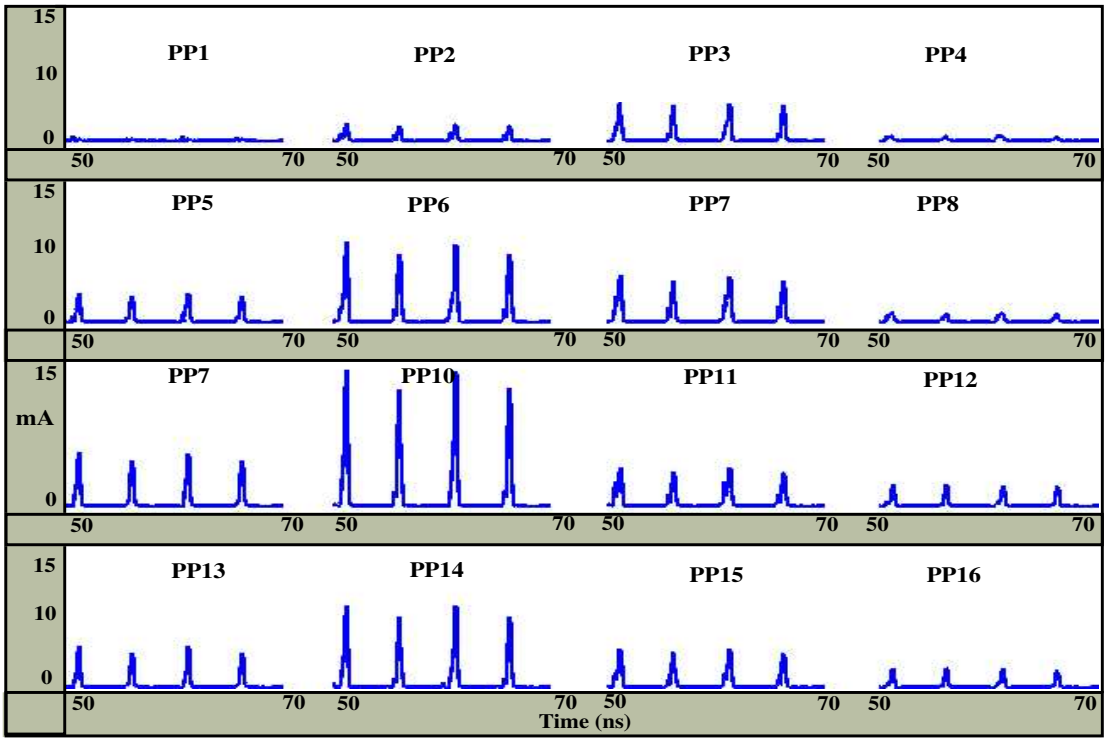

Figure 4. Estimated Path1 Clock Current Transients

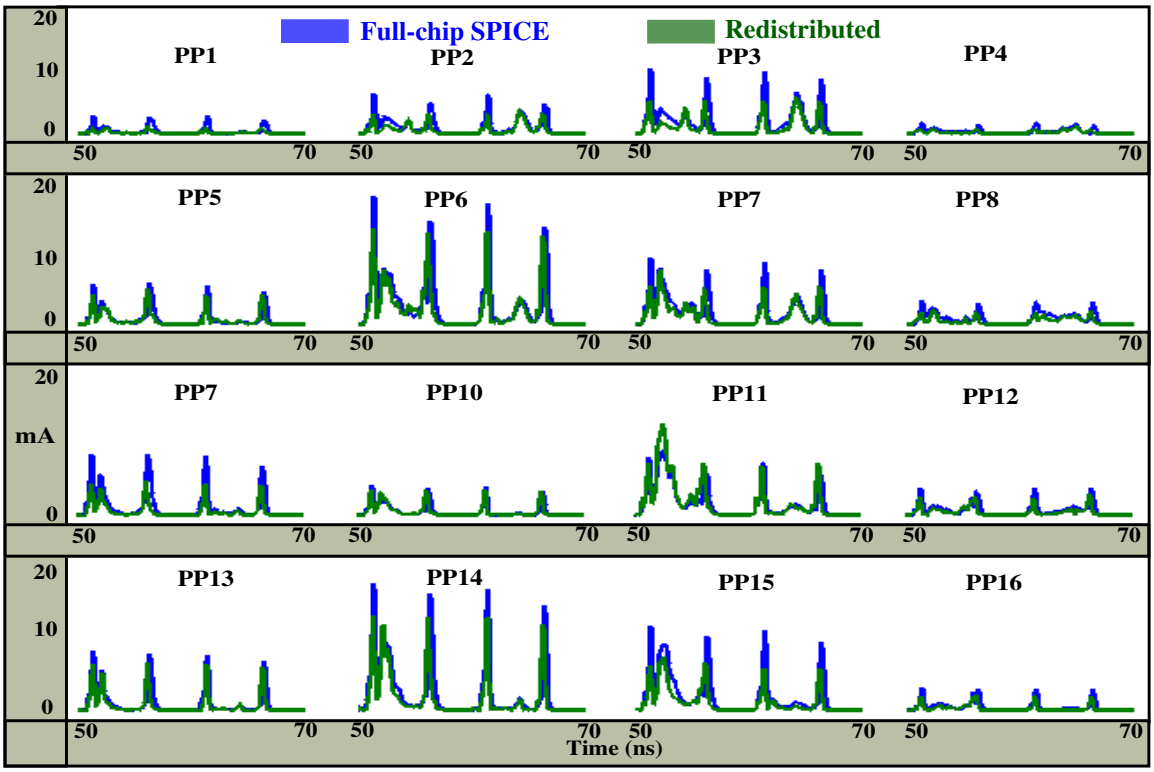

Figure 5. Full-Chip vs. Estimated Path1 Current Transients

each path (or test pattern) are shown in Table 2. Subpaths (on-path sequential and combinational logic gates) in each partition are identified and simulated to obtain the gate current transients which are then convolved with the C2C IR's to estimate the power pad current transients. The same process is repeated for all the off-path sequential elements in the clock tree to predict the current transients contributing to the total power transients measured at the sixteen PPs of the b17 design. The estimated current waveforms for both the clock tree and switching logic gates are presented.

8.1 Clock Current Transients

path	u1	u2	u3	u4	u5	u6	u7	u8	u9
1	76	30	139	168	111	158	52	183	247
2	73	35	137	167	113	156	52	181	245

Table 3: Off-path sequential elements and clock buffer count

The clock transients are estimated for 1164 and 1159 for path 1 and 2 respectively off-path clock buffers and flip-flops in each test pattern spread across the partitions as shown in Table 3. The partition based divide and simulate strategy speeds up the current estimation process since each of the clock path simulation takes less than 5 min. and can be run in parallel. The current waveforms estimated using our method is shown in Figure 4 for one representative path.

8.2 Combinational Current Transients

The off-path clock tree waveforms are superposed with the current waveforms estimated for the switching sequential and combinational logic gates and then redistributed using the transfer-function matrix. These redistributed waveforms are then compared to the full-chip SPICE current waveforms as shown in Figure 5 for this particular path. The figure shows that the estimated waveforms follow the expected full-chip SPICE results closely.

$$\text{NME} = \frac{\left(\frac{1}{N}\sum_{n=0}^{N-1}\left|(I_{\text{Full Chip}}[n] - I_{\text{Estimated}}[n])\right|\right)}{max(I_{\text{Full Chip}}[n])} \times 100\% \quad \textbf{(1)}$$

To quantify the results, we compute the Normalized Mean Error (NME) to compare the full-chip and estimated waveforms as defined in equation 1 shown below. In the equation $I_{\text{Full Chip}}$ is the full-chip transient current, $I_{\text{Estimated}}$ is the estimated transient current and N is the number of datapoints in the waveforms. The NME recorded for two paths is shown in Figure 6. A maximum error of 8.2% is observed.

The scalability and efficiency of PSANDE is illustrated in Table 4. The table shows the comparison of worst-case simulation time required for SPICE simulations of 1) full-chip b17 circuit; 2) each subpath; 3) power grid to compute the C2C IR's per input and; 4)

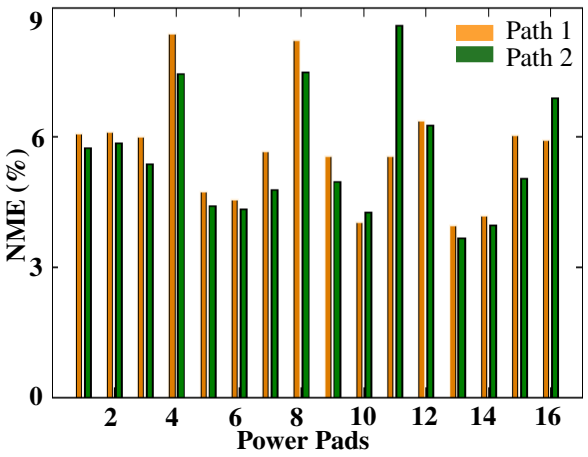

Figure 6. Normalized Mean Error for Two Switching Paths is shown on the Y-axis. The X-axis represents 16 different PPs in the design.

off-path sequential elements to estimate clock-tree transients. In the worst case each subpath and grid simulation take the time shown in the table, where 1415 subpath simulations were performed for the longest path in parallel.

Component Simulated	Time
Full-Chip SPICE	147h 28m
One Subpath	2m 13s
One Grid	1m 42s
One Clock path	4m 16s

Table 4: Simulation Time Comparison

9.0 Conclusion

This paper shows major enhancements to PSANDE that include, complete implementation of macro and micro-level design partitioning, superposition and redistribution techniques to predict macro-level currents using only micro-level simulations and more importantly PSN induced by the clock network. This research specifically demonstrates the scalability of PSANDE. Effect of both switching (on-path) and non-switching (off-path) clocked sequential elements is accurately modeled in the enhanced framework. The accuracy of the PSANDE is also presented and the worst-case error is 8.2%. The run-time for accurately predicting PSN shows improvement from several hours of full-chip simulation to a few minutes using the framework.

10.0 References

[1] Rao, S.K.; Robucci, R.; Patel, C.; , "Simulation Based Framework for Accurately Estimating Dynamic Power-Supply Noise and Path Delay," Journal of Electronic Testing., February 2014, Volume 30, Issue1,pp.125-147.

[2] Rao, S.K.; Robucci, R.; Patel, C.;, "Scalable Dynamic Technique for Accurately Predicting Power-Supply Noise and Path Delays," . VLSI Test Symposium (VTS), 2013 IEEE 31st, vol., no., pp.1-6, April 29 - May 02 2013.

[3] Rao, S.K.; Sathyanarayana, C; Kallianpur,A.; Robucci, R.; Patel, C.; , "Estimating Power Supply Noise and its impact on path delay", VLSI Test Symposium (VTS), 2012 IEEE 30th, vol., no., pp.276-281, 23-25 April 2012.

[4] Abhishek Singh, Jim Plusquellic, Dhananjay Phatak and Chintan Patel, "Defect Simulation Methodology for i_{ddt} Testing", J.Electron Test., 22:255-272, June 2006.

[5] Reza M. Rad and Jim Plusquellic, "A Novel Fault Localization Technique based on Deconvolution and Calibration of Power Pad Transient Signals", J.Electron. Test./ 25:169-185, June 2009.

[6] M.Tehranipoor and K.M. Butler, "Power Supply Noise: A Survey on Effects and Research", IEEE Design & Test of Computers, 27(2):51-67,2010.

[7] Chunsheng Liu, Yang Wu and Yu Hang, "Effect of IR-drop on Path Delay Testing using Statistical Analysis", In Proc. 16th Asian Test Symposium ATS '07, pages 245-250, 2007

[8] S.Zhao, K.Roy and C.K.Koh, "Estimation of Inductive and Resistive Switching Noise on Power Supply Network in Deep Sub-micron CMOS Circuits", In Proc. Int Computer Design Conf, pages 65-72,2000.

[9] Pant, S.; Chiprout, E.; Blaauw, D, "Power Grid Physics and Implications for CAD," Design & Test of Computers, IEEE , vol.24, no.3, pp.246,254, May-June 2007.

[10] Andersson, D.A.; Nilsson, B.; Pihl, J.; Svensson, L.; Larsson-Edefors, P.; , "Supply voltage drop study considering on-chip self inductance of a 32-bit processor's power grid," Signal Propagation on Interconnects, 2009. SPI '09. IEEE Workshop on , vol., no., pp.1-4, 12-15 May 2009

[11] Ismail, Y.I., "On-chip inductance cons and pros," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.10, no.6, pp.685,694, Dec. 2002.

[12] Galivanche, R.: Kapur, R.: Rubio, A.: , "Tesing in the Year 2020," Design, Automation & Test in Europe Conference & Exhibition, 2007. DATE '07 , vol., no., pp.1-6, 16-20 April 2007.

[13] Wang, L.-C.; Walker, D.M.; Xiang Lu; Majhi, A.; Kruseman, B.; Gronthoud, G.; Villagra, L.E.; van de Wiel, P.J.A.; Eichenberger, S., "Modeling Power Supply Noise in Delay Testing," Design & Test of Computers, IEEE , vol.24, no.3, pp.226,234, May-June 2007.

[14] Lee, J.; Narayan, S.; Kapralos, M.; Tehranipoor, M.; , "Layout-Aware, IR-Drop Tolerant Transition Fault Pattern Generation," Design, Automation and Test in Europe, 2008. DATE '08 , vol., no., pp.1172-1177, 10-14 March 2008

[15] Junxia Ma; Lee, J.; Tehranipoor, M.; , "Layout-Aware Pattern Generation for Maximizing Supply Noise Effects on Critical Paths," VLSI Test Symposium, 2009. VTS '09. 27th IEEE , vol., no., pp.221-226, 3-7 May 2009.

[16] Kokrady, A.; Ravikumar, C. P., "Fast, layout-aware validation of test-vectors for nanometer-related timing failures," VLSI Design, 2004. Proceedings. 17th International Conference on , vol., no., pp.597,602, 2004.

[17] Qiang Xu; Dianwei Hu; Dong Xiang; , "Pattern-directed circuit virtual partitioning for test power reduction," Test Conference, 2007. ITC 2007. IEEE International , vol., no., pp.1-10, 21-26 Oct. 2007.

[18] Yi-Min Jiang; Kwang-Ting Cheng; , "Vector generation for power supply noise estimation and verification of deep submicron designs," Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.9, no.2, pp.329-340, April 2001.

[19] Yu Zhong; Wong, M.D.F.; , "Fast algorithms for IR drop analysis in large power grid," Computer-Aided Design, 2005. ICCAD-2005. IEEE/ACM International Conference on , vol., no., pp. 351- 357, 6-10 Nov. 2005.

[20] Kozhaya, J.N.; Nassif, S.R.; Najm, F.N.;,"A multigrid-like technique for power grid analysis," Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on , vol.21, no.10, pp. 1148- 1160, Oct 2002.