An On-Chip Load Model for Off-Chip PDN Analysis Considering Interdependency Between Supply Voltage, Current Profile and Clock Latency

Jun Chen
Osaka University
Osaka, Japan
j-chen@ist.osaka-u.ac.jp

Toshiki Kanamoto
Hirosaki University
Aomori, Japan
kana@hirosaki-u.ac.jp

Hajime Kando
Murata Manufacturing Co., Ltd.
Kyoto, Japan
kando@murata.com

Masanori Hashimoto
Osaka University
Osaka, Japan
hasimoto@ist.osakau.ac.jp

Abstract—Simple yet accurate on-chip load model is demanded for off-chip power delivery network (PDN) design and verification. Conventionally, a current source that represents a short chip operation period is used for this purpose, but it cannot consider the interdependency between supply voltage, load current and clock latency. The ignorance of this interdependency could mislead off-chip PDN design causing over- and under-design. To address this issue, this paper proposes an on-chip load model with Verilog-A that can replay the load current and clock latency under dynamic supply noise. The model is expanded to support different chip operation modes, and it can be used as a sub-model to construct a large chip model. Experiment shows over 200X run-time improvement comparing with full SPICE netlist simulation. We also confirm that the current profile, power consumption, and clock latency are closely correlated.

I. Introduction

Chip power dissipation and consequent operation current increase dramatically as the technology node advances to nanometer era. Clock gating, frequency scaling, and dynamic voltage scaling introduce considerable load current variations. Such chip current variations arise large supply voltage variance due to imperfect power delivery network (PDN) [1].

On-chip load model is widely used for off-chip PDN design and verification. Conventionally, a current source that represents a chip operation in a short period is used as the load model [2], [3]. On the other hand, the current source has a limitation by nature that the load current is constant irrelevant to voltage variance. In actual circuits, the load current becomes smaller as the supply voltage drops, and hence the dynamic noise is naturally mitigated. However, the current source model cannot consider this interdependency, and then the supply noise is likely to be overestimated. To avoid over- and under-design of PDN, the interdependency between supply voltage and load current must be reproduced.

Besides, the chip peak current is dominated by clock switching, and the clock timing is affected by the supply noise. When the supply voltage is high, the clock latency from the clock source to FFs becomes short. At a lower supply voltage, conversely, the clock latency becomes longer. Therefore, to accurately reproduce the load current, the clock latency model that can consider supply voltage is demanded. Reference [4] points out that the interdependency between supply voltage and clock latency needs to be accurately reproduced for high-speed I/O design and verification.

Meanwhile, the off-chip PDN design aims to guarantee correct chip operation. In digital circuits, one of the primary concerns is timing errors since the supply noise affects path delay [5], [6]. Reference [7] reported over 8% timing impact under supply voltage noise after 90-nm technology node. Conventionally, however, it is difficult for off-chip PDN designers to access on-chip timing information. Therefore, the on-chip load model that can provide the information on the timing variation is highly desirable for off-chip PDN designers since the relation between the off-chip PDN design and on-chip timing behavior becomes visible and can guide the PDN design. For addressing these issues mentioned above, this paper proposes an on-chip load model with Verilog-A that can replay the load current, power consumption, and clock latency under dynamic supply noise.

II. Related Works

A naive approach to analyze PDN under dynamic noise taking into account the interdependencies between supply voltage, load current and clock latency is to simulate PDN together with the full transistor-level SPICE netlist of the chip with actual input vectors. However, one simulation run may take weeks or even years to finish. Extensive design exploration is totally infeasible.

For CPU time reduction, current source load model and a number of simplified on-chip load models are proposed with different accuracy targets. To address the interdependency between voltage variance and load current, some models improve the accuracy based on traditional current source load. References [2], [3] distribute current sources in PDN and introduce scaling factors to migrate simulation error, and multiple current sources are prepared for different operation modes. However, even with their detailed characterization effort, current source is inherently irrelevant to voltage variance. Other models in [8], [9] utilize RC elements to describe the chip load. This approach, however, requires careful tuning effort on RC parameters to match both the current peak and
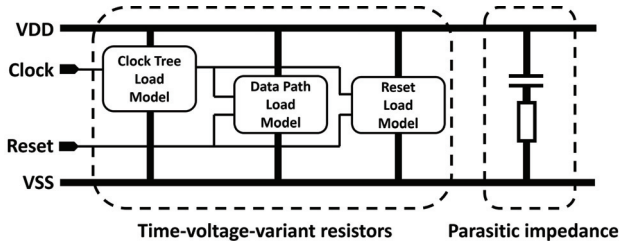
Fig. 1. Overall structure of on-chip load model.

Fig. 2. Comparison of equivalent switching resistance.

Fig. 3. Delay under constant supply voltage.

width. On the other hand, due to the non-linear characteristics of MOS transistors, the resistance in the RC model depends on the supply voltage. Without the consideration of the non-linearity, the simulation error is tolerable only for a small range of voltage variation.

As for the interdependency between supply voltage and clock latency, both time domain and frequency domain modeling methods are developed. The frequency domain modeling methods are usually based on voltage-latency sensitivity table [5]. When a noise spectrum is given, these models are useful to predict the jitter/latency probability distribution, but they have difficulties in time-domain estimation. On the other hand, the time domain model [6] estimates the jitter/latency, but this approach requires intensive extraction efforts for pin-to-pin delay or remarkable computation efforts for delay propagation.

III. PROPOSED ON-CHIP LOAD MODEL

A chip circuit load can be modeled by variable resistors [1], [8], [9]. This model can be further improved with a time-voltage-variant impedance to reproduce chip load behavior both in time and frequency domain. Here, we propose a new load model that consists of two parts as shown in Fig. 1.

The left side is time-voltage-variant resistor part, which is responsible for reproducing the switching current in time domain. In Fig. 1, the sub-modules of chip clock tree and data path are modeled separately. Also, two modes of normal and reset operation are offered. A reset signal is inputted to enable or disenable the sub-model for different operation modes. Apparently, this structure is expandable for additional modes and sub-modules. The right side is parasitic impedance part, which is responsible for reproducing the voltage-current response in frequency-domain.

The most challenging point for this structure is to model the resistance interdependency between time and supply voltage. This challenge is addressed by proposing a scaled resistance profile (RP) method, which is explained in the first subsection. The second subsection will describe the model characterization process. The last subsection will present simulation methodology for mixed-signal simulation.

A. Time-Voltage-Variant Resistor Modeling

This section proposes the scaled profile method to model the time-voltage-variant resistor part. First, we define the RP element by $\begin{pmatrix} t_n(V_{DD}) & r_n(V_{DD}) \end{pmatrix}$, where t_n is time in simulation and r_n is the chip equivalent resistance. t_n and r_n 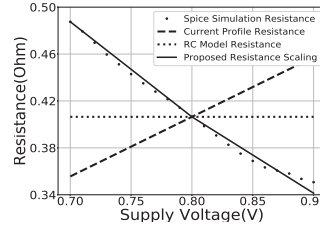are functions of supply voltage V_{DD}. The simulator will update the resistance to r_n at t_n according to V_{DD}, and naturally deduce current by Ohm's law. Supposing chip load behavior is composed of N RP elements, we define **RP** as
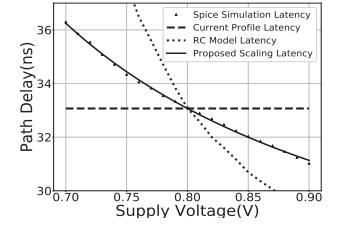
$$\mathbf{RP} = \begin{pmatrix} \mathbf{T}_N & \mathbf{R}_N \end{pmatrix}, \tag{1}$$

where $\mathbf{T}_N$ and $\mathbf{R}_N$ are time and resistance vectors, respectively. Each RP element consists of $t_n \in \mathbf{T}_N$ and $r_n \in \mathbf{R}_N$.

1) Resistance Vector Modeling: Given a sub-module circuit, N_{tr} transistors are conductive. Suppose V_{DS} over a conductive transistor is small, and supply voltage $V_{DD} \approx V_{GS}$. Then, the equivalent resistance $r(V_{DD})$ can be expressed by

$$r(V_{DD}) = \frac{V_{DD}}{\sum_{i=1}^{N_{tr}} I_i} \approx (\sum_{i=1}^{N_{tr}} \frac{(V_{DD} - V_T)}{k_i} \cdot \left(\frac{W_i}{L_i}\right))^{-1}, \tag{2}$$

where I_i, k_i, L_i and W_i are drain current, conductivity factor, channel length, and channel width of individual transistors, respectively, and V_T is threshold voltage. We use a piecewise linear function to fit this voltage-variant equivalent resistance. Then, the resistance can be expressed with a scaling factor by

$$r(V_{DD}) = r(V_0) \cdot SR(V_{DD}), \tag{3}$$

where V_{DD} is supply voltage, $r(V_0)$ is the equivalent resistance derived from current profile at nominal supply voltage V_0, and $SR(V_{DD})$ is the piecewise resistance scaling function fit from (2).

Fig. 2 shows the advantage of this scaling method over conventional methods. A four-stage clock tree is selected for demonstration. SPICE simulation resistance is derived from the current profile of the clock tree, and it is the reference. The current source load and an RC load model are prepared at nominal voltage. We can see that the RC model and current source model underestimate the resistance at low supply voltage and overestimate it at high supply voltage, and hence current and power will also be misestimated. On the other hand, the proposed scaling resistance can correlate closely with SPICE simulation result as expected.

2) Time Vector Modeling: Suppose a given path delay D is divided into N intervals and Δt_n denotes the n-th interval. Assuming intervals are sufficiently short, the interval duration is determined by average voltage V_{An} during the interval since the interval is impacted by transistor switching speed. This transistor switching includes RC charging and discharging processes, and hence the interval can also be scaled by time scaling function similar to resistance vector elements.

$$\Delta t_n(V_{An}) = \Delta t_n(V_0) \cdot ST_n(V_{An}), \tag{4}$$

where $ST_n(V_{An})$ is the time scaling function for n-th interval. When the intervals are evenly distributed along the path, we use a single time scaling function $ST(V_{An})$ as the representative. Hence, the path delay is expressed as

$$D = \sum_{n=1}^{N} (\ \Delta t_n(V_0) \cdot ST(V_{An})\ ). \tag{5}$$

Then, the time vector element t_n becomes

$$t_{n+1} = t_n + \Delta t_n(V_0) \cdot ST(V_{An}). \tag{6}$$

At a constant supply voltage V_{DD}, path delay (5) can be simplified as

$$D(V_{DD}) = D(V_0) \cdot ST(V_{DD}) = \sum_{n=1}^{N} \Delta t_n(V_0) \cdot ST(V_{DD}). \tag{7}$$

Time scaling function $ST(V_{DD})$ can be extracted from the circuit simulation or static timing analysis with libraries at different voltages. With (6) and (3), we can scale the resistance profile (1), and deduce the clock latency under both constant supply voltage and dynamic supply noise by (7) and (5).

Fig. 3 shows the estimated latency of the four-stage clock tree. RC model and current source model are derived at nominal voltage. We can see these two conventional models either over- or under-estimate the path delay under different supply voltage. The proposed scaled latency, on the other hand, can correlate closely with SPICE simulation result.

B. On-chip Load Model Characterization

Next, we present general characterization process for constructing the on-chip load model. For a given input logic vector, the time-voltage-variant resistor is characterized as follows.

Step 1: Generate current profile at nominal voltage.
Step 2: Translate current profile into resistance profile form.
Step 3: Simulate and extract metrics (equivalent resistance, latency, and current) at different supply voltages.
Step 4: Run fitting process and generate scaling functions.
Step 5: Generate resistance profile with scaling functions.

Given a sub circuit module, we need one round of full time power simulation at nominal supply voltage at Step 1. In Step 2, the resistance profile form refers to $\begin{pmatrix} t_n & r_n \end{pmatrix}$, where t_n is time in simulation and r_n is equivalent resistance under nominal voltage. Step 3 requires hundreds of clock cycles simulation to extract necessary metrics for the fitting process in Step 4. In Step 5, final resistance profile is composed by time and resistance vectors described in (6) and (3), respectively.

For the parasitic impedance part, the model is characterized by small signal analysis with equivalent circuit shown in Fig. 4, where C_1 and R_1 represent the parasitic impedance and R_2 is chip leak resistance. By sweeping frequency of small AC signal (typically $0.001\,\mathrm{V}$ from $1\,\mathrm{kHz}$ to up to $100\,\mathrm{GHz}$), the equivalent impedance is obtained as Fig. 5. Then, the parameter R_1, C_1, and R_2 can be extracted by least squares fitting. When leakage current is included in RP, we remove R_2 and keep only C_1 and R_1 as parasitic impedance part.
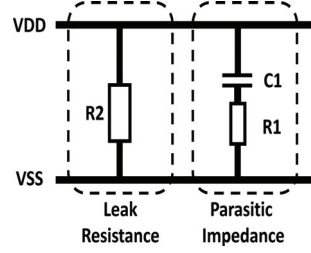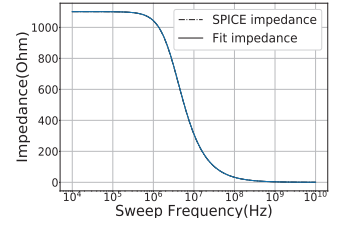

Fig. 4. Small signal equivalent circuit.

Fig. 5. Small signal impedance.

Algorithm 1 RP Simulation Algorithm

Input: V_{DD}, V_{in_signal}, $V_{shutdown\_signal}$
Output: I, V_{out_signal}
    Initialization :
1: Set leak resistance
    Main Routine :
2: **if** $V_{shutdown\_signal}$ is enabled **then**
3:     Set the equivalent resistance under shutdown mode
4: **else**
5:     **if** V_{in_signal} is enabled **then**
6:        **for** $n = 1$ to N **do**
7:          Obtain r_n and t_n from RP
8:          Calculate the resistance and time interval
9:          Schedule the resistance update event at t_n
10:        Add the interval to $V_{in\_signal} - V_{out\_signal}$ transition, and deduce the current
11:       **end for**
12:     **end if**
13: **end if**

C. Simulation Procedure

We present the simulation procedure in Algorithm 1. This algorithm can be implemented with Verilog-A. Verilog-A is supported by mainstream mixed-signal simulators, and hence our model can be co-simulated with Verilog and SPICE modules. By applying the similar approach to other sub circuit modules or modes, we can expand the on-chip load model.

IV. EXPERIMENT EVALUATION

For the experiment, we prepared a 32-bit OpenRISC processor synthesized with NanGate 15nm Open Cell Library. The number of cells is over 17k, the maximum clock frequency for the core processor logic is $1.2\,\mathrm{GHz}$, and the average clock latency is $114.88\,\mathrm{ps}$ at $0.8\,\mathrm{V}$ supply voltage. A CRC checksum program is given to OpenRISC as workload. The characterization for 500-cycle operation finished within two hours in this test case.

First, we compare current profile of full SPICE netlist simulation and the proposed on-chip load model. Fig. 6 shows the load current waveforms at nominal voltage of $0.8\,\mathrm{V}$. We can see that the load current waveform of the proposed model is close to the transistor-level SPICE simulation result.

Next, we evaluate the accuracy at different supply voltages from 0.7 to $0.9\,\mathrm{V}$. The results are listed in Table I. This evaluation simulated for 200 clock cycles. For the current peak
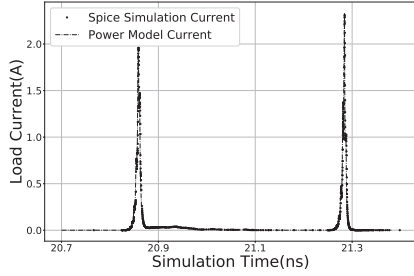
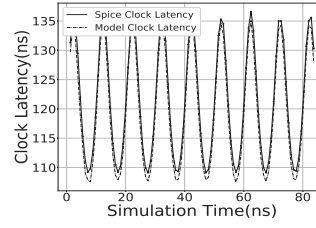Fig. 6. Current waveform comparison within one clock cycle.

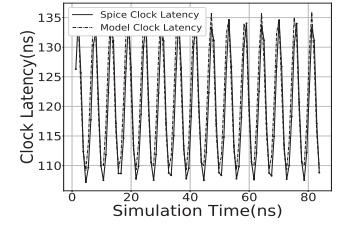Fig. 7. Clock latency estimation with 100MHz supply noise.
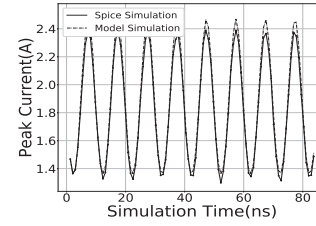

Fig. 8. Clock latency estimation with 1GHz supply noise.

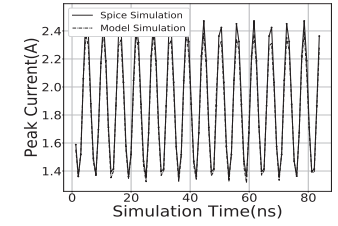Fig. 9. Peak current estimation with 100MHz supply noise.

Fig. 10. Peak current estimation with 1GHz supply noise.

TABLE I
LOAD CURRENT AND CLOCK LATENCY AT VARIOUS SUPPLY VOLTAGES.

Supply Volt.(V)	Avg. Peak Curr.(A)		Err(%)	Avg. Latency (ns)		Err(%)
	SPICE	Model		SPICE	Model	
0.70	1.378	1.361	1.37%	131.73	132.24	0.44%
0.73	1.537	1.495	2.73%	125.83	126.06	0.27%
0.77	1.744	1.699	2.69%	119.02	119.38	0.43%
0.80	1.911	1.873	2.01%	114.88	115.25	0.39%
0.83	2.097	2.033	3.20%	111.40	111.70	0.33%
0.87	2.350	2.271	3.30%	107.45	107.67	0.26%
0.90	2.497	2.471	1.17%	104.89	105.08	0.24%
Avg.	-	-	**2.35%**	-	-	**0.33%**

evaluation, we calculated the errors for 400 current peaks and computed the average of them, where 400 peaks are 200 clock cycles multiplied by two peaks per clock cycle. The average error for individual peak currents is 2.35%. On the other hand, conventional current source and RC model cannot attain such an accuracy. The average peak current errors are 17.63% and 10.45%, respectively. For the clock latency evaluation, the average latency error is 0.33%, whereas the average errors for current source and RC model are 6.33% and 11.41% respectively. Especially, the current source model suffered up to 38.53% error in peak current estimation, and RC model suffered up to 39.17% error in latency estimation.

Then, to validate the model under dynamic supply noise, we injected a sinusoidal noise with 100 mV amplitude whose frequency ranged from 100 MHz to 1 GHz, where 100 MHz is roughly 10x lower and 1 GHz is almost similar to the clock frequency. We simulated 100 clock cycles for both full-SPICE netlist and the proposed on-chip load model. Figs. 7 and 8 show the clock tree latency comparison. We can see both the clock latencies are well correlated. The average latency errors are 1.46% for 100 MHz noise, and 2.62% for 1 GHz noise. The peak current under dynamic noise is also compared in Figs. 9 and 10. The average peak current errors are 2.32% for 100 MHz noise and 2.15% for 1 GHz noise.

Finally, we conducted the entire PDN simulation that included DC/DC converter, PCB, package, and the on-chip load model. Simulation time was 500 ns. Our on-chip load model took 136 seconds, which is comparable with 122 seconds of current source. Full-SPICE netlist load took 29,632 seconds, which means over 200X runtime reduction. Note that this runtime reduction is more significant when the circuit under

modeling is larger.

V. CONCLUSION

In this paper, we proposed an on-chip load model that can replay the load current and clock latency under dynamic supply noise. The model can be further expanded for different chip operation modes and be integrated to larger circuit scale. The experiment shows accurate estimation results regarding clock latency and peak current. Over 200X runtime reduction is achieved compared with full SPICE netlist simulation.

REFERENCES

[1] S. Lin and N. Chang, "Challenges in power-ground integrity," *Proc. ICCAD*, pp. 651-654, 2001.
[2] W. Cui, P. Parmar, J. Morgan, and U. Sheth, "Modeling the network processor and package for power delivery analysis," *Proc. EMC*, vol. 3, pp. 690-694, 2005.
[3] L. Zheng, Y. Zhang, and M. Bakir, "Full-chip power supply noise time-domain numerical modeling and analysis for single and stacked ICs," *IEEE Trans. Electromagnetic Compatibility*, vol. 63, no. 3, pp. 1225-1231, 2016.
[4] H. Shi, G. Liu, A. Liu, A. Pannikkat, K. Ng, and Y. Yew, "Simultaneous switching noise in FPGA and structure ASIC devices, methodologies for analysis," *Proc. ECTC* , pp. 229-236, 2006.
[5] Y. Shim and D. Oh, "System level modeling of timing margin loss due to dynamic supply noise for high-speed clock forwarding interface," *IEEE Trans. EC*, vol. 58, no. 4, pp. 1349-1358, 2016.
[6] G. Bai, S. Bobba, and I. N. Hjj, "Static timing analysis including power supply noise effect on propagation delay in VLSI circuits," *Proc. DAC*, pp. 295-300, 2001.
[7] M. Saint-Laurent and M. Swaminathan, "Impact of power-supply noise on timing in high-frequency microprocessors," *IEEE Trans. Advanced Packaging*, vol. 27, no. 1, pp. 135-144, 2004.
[8] H. Chen and J. Neely, "Interconnect and circuit modeling techniques for full-chip power supply noise analysis," *IEEE Trans. Components, Packaging, and Manufacturing Technology: Part B*, vol. 21, no. 3, pp. 209-215, 1998.
[9] Y. Ogasahara and M. Hashimoto, "Validation of a full-chip simulation model for supply noise and delay dependence on average voltage drop with on-chip delay measurement," *IEEE Trans. Circuits and Systems II: Express Briefs*, vol. 54, no. 10, pp. 868-872, 2007.