A Modeling Approach for Addressing Power Supply Switching Noise Related Failures of Integrated Circuits

Chandra Tirumurti, Sandip Kundu, Susmita Sur-Kolay[†], Yi-Shing Chang
3600 Juliette Lane, Santa Clara, CA 95052, USA
Contact Author: *Sandip.Kundu@intel.com*

Abstract

Power density of high-end microprocessors has been increasing by approximately 80% per technology generation, while the voltage is scaling by a factor of 0.8. This leads to 225% increase in current per unit area in successive generation of technologies. The cost of maintaining the same IR drop becomes too high. This leads to compromise in power delivery and power grid becomes a performance limiter. Traditional performance related test techniques with transition and path delay fault models focus on testing the logic but not the power delivery. In this paper we view power grid as performance limiter and develop a fault model to address the problem of vector generation for delay faults arising out of power delivery problems. A fault extraction methodology applied to a microprocessor design block is explained.

1. Introduction

Semiconductor industry has been driven by Moore's law for almost a quarter century. Miniaturization of device size has allowed more transistors to be packed into an area while the improved transistor performance has resulted in significant increase in frequency.

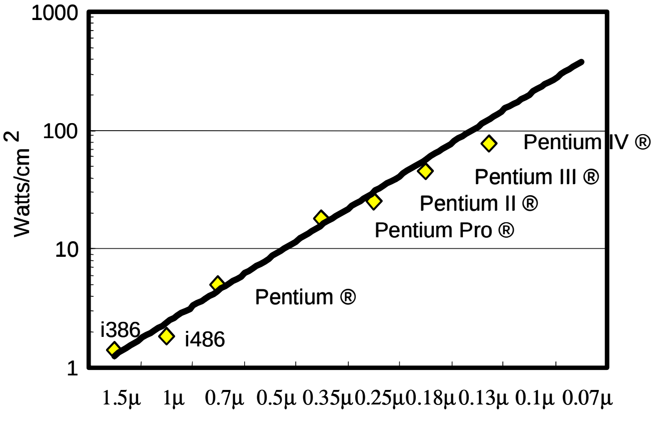

Figure 1 Power density by technology

Increased density of switching devices and rising frequency has lead to a power density problem. In Figure 1 we show how the power density of leading edge microprocessors has progressed over multiple technology generations.

The rise in power density with a simultaneous reduction in power supply voltage leads to a large increase in the amount of current that needs to be delivered.

Non-uniform pattern of power consumption across a power distribution grid causes a non-uniform voltage drop. Instantaneous switching of nodes may cause localized drop in power supply voltage, which we call as droop. This instantaneous drop in power supply at the point of switching causes excessive delay and a speed path problem. The goal in testing is to create excessive switching around nodes with slack ranging from none to very little, so that any extra delay creates a speed-path.

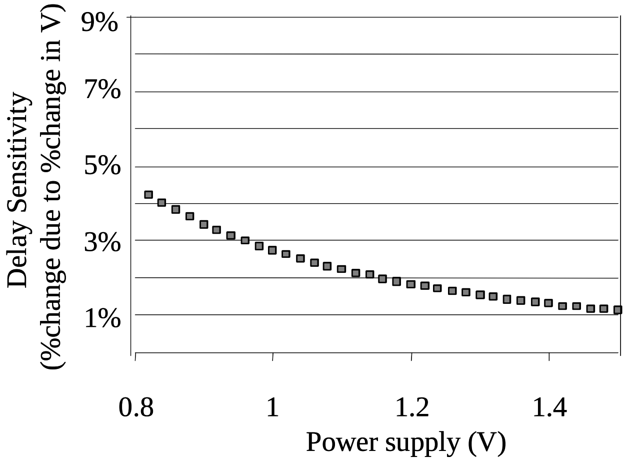

Figure 2 Delay sensitivity as a function of Vdd

Power supply voltage scales with technology. Figure 2 shows delay sensitivity as a function of supply voltage for a 90nm technology. Sensitivity rises with declining voltage and

[†] The author is with Indian Statistical Institute. The work was performed while she was a visiting professor at Intel Corporation.

1530-1591/04 $20.00 (c) 2004 IEEE

at 0.9V, 1% change in power supply voltage will cause nearly 4% change in delay for most static CMOS gates in this technology. Thus we have a compounding problem. Sharper switching rates cause greater droop while delay sensitivity to power supply is getting worse.

The objective in testing is to (1) identify regions that may experience large droop and (2) judge criticality of droop induced delay by comparing against slack to create a list of targets. For lack of a better term, we will call it a *fault - power supply switching fault* to be precise. Once we have such a fault list, fault simulation and pattern generation can be performed. Power supply switching faults are described using Generalized Fault Model (GFM) [1]. In section 5, we will describe fault model in more detail.

There has been work in this area for generating patterns for delay testing and timing analysis with worse case power supply noise [2]. Here the authors propose a test generation approach for paths including power supply noise induced delays.

In our analysis we use a *vector-less* approach. Actual simulation of circuits using extracted power supply model and transistors against input vectors is complex and can only be performed for small circuits. Thus, vector-less approach is imperative for scalability of solution.

However, we use actual simulation on small circuits for *learning* purposes. This data is maintained as a *library of knowledge*. We refer to this during analysis phase.

The basic divide-and-conquer philosophy is based on identifying a small cluster of transistors called *cells*. Since the fault simulation and ATPG are run on a gate level model, the smallest cells are *gates* such as NAND, NOR gates. For cells that perform more complex functions, we deploy *two definitions* for cells. The first definition is for combinational gates. For a combinational gate, a cell is a channel connected transistor graph. In literature this has also been called a Channel Connected Component (CCC) or a Channel Connected Switching Network (CCSN). For sequential gates, a cell may comprise of more than one CCSN. Consequently, they are identified manually using a template based approach.

The key elements in our library of knowledge are (a) peak current consumption by a cell and (b) how the current distributes itself through a power distribution network. We use *superposition* rule for adding cell currents to identify the impact of multiple gate switching. This is the key principle used in analysis phase.

For analyzing how switching current at the node of consumption distributes itself through a power distribution network, we use circuit simulation on extracted power grid using a current source. The input in this simulation is the switching current source. Simulation is performed for varying input waveforms with different slopes and peak magnitudes. The observations from these simulations are stored in a tabular form and referred during actual analysis. We call this *the power grid characterization phase*. Further details appear in section 2.

For characterizing peak current consumption of a cell, a vector based circuit simulation is used to find the worst case current. This is treated in more details in section 3.

The final analysis phase breaks up a design block into cells. The peak current consumption of a cell and its geometric position are used to derive a geometric distribution of current around the cell. Geometric distribution of current from individual cells is tallied up to obtain global ranking of the regions. From this ranking a list of corresponding switching nodes is collected. This list is then processed with information from static timing analysis to produce GFM. This will be described in section 4.

2. Power Grid Characterization

The power grid is characterized for obtaining a geometric distribution of switching current at the via locations. The total via current is used as an indicator for grid saturation/droop. Packaging delivers the current to the C4 pads located at the upper most level of the die. Current delivered to Vdd nodes flows down through successive metal layers to the transistors and flows back up to the corresponding C4 pads that connect to Ground at the package level. This is somewhat simplified because it ignores AC current return path at intermediate metal layers. However for the purpose of power starvation calculation this is quite adequate because, most of the power starvation effects come from switching transistors. In Figure 3 we show a simplified power distribution grid.

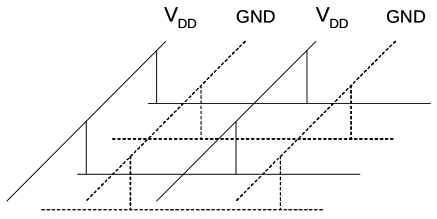

Figure 3 Simplified View of the Power Grid

The power grid is composed of alternate metal lines of power (Vdd) and ground (GND) in each layer. The upper and lower metal layers are connected by vertical vias. Transistor Vdd and GND terminal connect to the lower metal layer vias. Currents and voltages can be analyzed at the vias. Simultaneous switching of transistors connected to the via causes maximum current flow through the via and a droop in voltage. Voltage droop at the via slows the switching of transistors connected to it. Since the transistors are directly connected to the lower metal layers, vias at the lower metal layers are important for analysis of current and voltage droop.

2.1 Power Grid analysis

An internally developed power grid analysis tool is used to derive an RLC model of the power grid for circuit simulation. The tool uses process specific technology libraries for reistance and capacitance values, uses built-in inductance and de-cap models. This tool also allows us to stitch a current source at a location of user's choosing. Voltage and current values at each via for each metal layer can be obtained from simulation. For a small analysis region, this analyzer is very fast, allowing hundreds of runs for characterization.

Starting with a realistic power grid used in microprocessor design, we used the analysis tool to study (1) magnitude of the droop and (2) the spatial distribution of current at power vias for various current waveforms. The following conclusions were made about the power grid.

2.2 Droop Magnitude

We found that the voltage droop magnitude at the vias is a function of peak current of the source and is maximum at the lower metal layers. As expected, higher current corresponds to higher voltage droop. The results are shown in Figure 4.

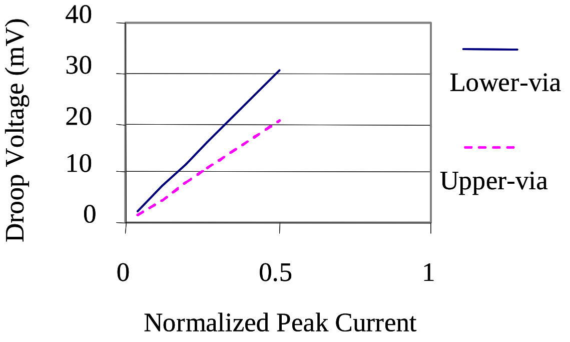

Figure 4 Voltage droop by source current

2.3 Current Distribution

It is also observed that the spatial distribution of the droop is independent of *Ipeak*, the peak source current. As expected, current at the via drops off as you move away from the source placement. This drop off is a strong function of the via layer. Numerically, the current does not quite go to zero. So we need a cut-off. We use 10% of Ipeak as cut-off threshold for deciding the region of influence. Based on this definition, a typical spatial distribution of droop current by via-layer is shown in Figure 5.

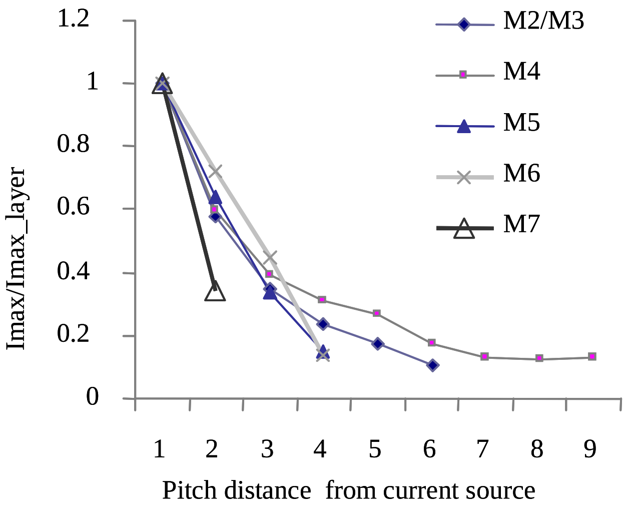

Figure 5 Current contribution of vias by distance

This characterization method allows for the estimation of current and voltage droop at a via, given the distance and the magnitude of the current source. It is easy to work with via currents instead of voltage droop, since currents can be summed independently at the vias from the current sources. The characterization data is maintained as a library for use by the fault extraction tool.

3. Peak current computation

Given a logic block in the design, an estimate of the peak switching current is needed, so that via currents can be estimated using the characterization library described above.

One way to do this for standard-cell blocks is to use a circuit simulator to create a table of peak Vdd and Vss currents for different values of output load and input slope. A typical current waveform for an inverter is shown in Figure 6.

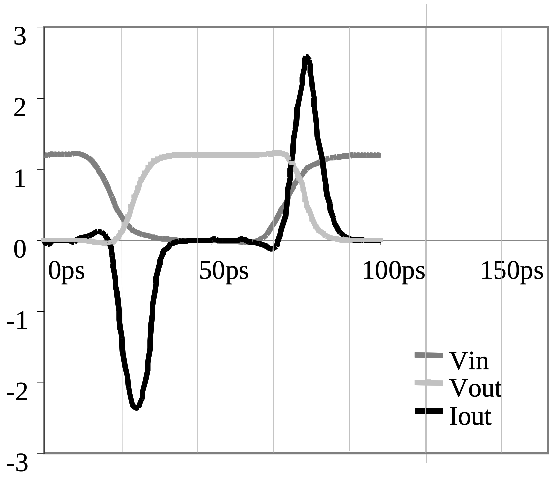

Figure 6 A Current waveform for an Inverter

This step can be made part of cell library characterization. This information is stored in a tabular form. Regression analysis on tabulated results can produce parameterized equations for peak current computation based on the type of cell, its load and input slope conditions.

A second approach is to create analytical equations for peak current based on total charge (C*Vdd). This requires knowledge of waveform shape. Kayssi [3] and Wang [4] have derived analytical models for these waveform shapes. Based on our simulations we have found that for small loads, the current waveform can be approximated to sine^2 . For medium loads, a piece-wise linear (PWL) triangular wave form is needed while for much larger loads trapezoidal waveforms need to be used.

4. Analysis for Fault Extraction

The first extraction step involves computing the worst-case currents at each via of the power grid, given a fully placed and routed design. Based on a minimum current threshold, a list of problem vias is then extracted.

4.1 Layout information

Physical placement of the logic blocks along with the coordinates of the power grid vias is required for spatial distribution of the currents. Local routing information is needed to go from a via to the logic blocks fed by the via.

4.2 Timing information

The slack on logic block outputs is used to identify potential fault sites for power supply switching fault model. Other information such as load capacitance and slope are useful for picking the aggressors and estimating peak current.

4.3 Worst-case Via current computation

Once the peak currents are computed for each logic block, as explained in section 3. They are spatially distributed to the neighboring vias. The spatial distribution is obtained from a look-up table that is created during power grid characterization phase. The Vdd or Vss currents are summed separately and maintained at each via.

```
 for_all_cells (cell)

    Ipeak = cell.get_peak_current() ;
    region = grid.get_spatial_spread() ;

    for_all_vias_in (region)
       dist = | via_x_y - cell_x_y | ;
       Iratio = grid.get_ratio (dist) ;
       via.current += Ipeak * Iratio ;
       via.aggressors.add(cell) ;
    endfor // vias

endfor // cells
```

Figure 7 Algorithm for finding peak via current

A list of logic blocks that contributed current to a specific via is also maintained. We call this list the feed region for the via. A list of logic blocks that are directly connected to this via is also maintained. It should be noted that this list is a smaller subset of the first. Algorithm for current summation and feed list construction are shown in Figure 7.

A minimum current threshold is chosen based on voltage droop magnitude. Vias, whose total current is below the current threshold are discarded. A maximum slack threshold is also determined based on the design frequency. From the list of logic blocks that are directly connected to the via, if the slack on the logic block output exceeds the slack threshold, it is removed from this list. After this step, if the list is empty, the via can be discarded, since the voltage droop at this via does not affect any critical path. Fault extraction steps are shown in Figure 8.

```
// algorithm for fault extraction
for_all_vias (via)

  type = <VDD | VSS> ;
  Ivia = via.current() ;

  if (Ivia > current_threshold)

     via.victims = 0;
     for_all_cells_fed_by_via(cell)
         if(cell.slack <= slack_threshold)
              via.victims.add(cell) ;
        endif
     end_for

    if(via.victims != 0)
     new Fault (type, via.victims, via.aggressors) ;
    endif

  endif // Ivia > current_thresold

end_for // for all vias
```

Figure 8 Algorithm for fault extraction

5. Fault Modeling

The effect of voltage droop caused by simultaneous switching of logic blocks is modeled as a delay fault. Since this droop is a local phenomena around the saturated via, the signal transitions in the region will be slowed down. This fault can be modeled in the fault simulator as STR (slow-to-rise) or STF (slow-to-fall).

5.1 Overview of Generalized Fault Model (GFM)

GFM was developed in our earlier work [1] to repressent faults in a simple, yet flexible manner that is suitable for fault simulation and ATPG. In GFM construct, a fault refers to either a physical defect such as a bridge defect or a problematic behavior such as a cross-talk fault.

Similar generic fault formulations have been proposed in pattern fault [5] and fault-tuples [6].

A GFM fault consists of one or more fault atoms. A fault atom represents a facet of the defective behavior. For example, a power supply switching fault on a fan-out stem may be detected at either of the fan-out sink nodes or a combination of those nodes. We call out each behavior as a separate atom. Therefore, by definition, if a fault atom is detected, then the fault is detected. The fault atoms within a fault are ranked in terms of their analog behavior. For example, if one atom represents 30 mA of via current at certain via and another atom represents 25 mA, then detecting the first atom gives a test of better quality. Thus we transform the analog quality to a sorted priority order among atoms.

5.2 Power supply switching fault modeling

Each via that is considered for fault modeling has a list of logic blocks that are current sinks. Let us call this *aggressor list*. The via also contains a list of logic blocks that are in its feed region. We will call this *victim list*. In the extraction step, we have already discarded some vias based on total aggressor current and victim slack.

We can form several fault atoms by combinations of aggressors that meet this threshold, provided that the victims are part of the combination. The atoms are sorted according to the total via current. The first atom in this fault is the best choice in terms of fault simulation or test generation, since it creates the worst droop.

For each atom, a condition list is constructed from the aggressors, requiring a transition (0->1 or 1->0) depending on the via type (Vdd or Vss). The victims will now have transition fault (STR or STF). It should be noted here that victims are part of condition list by construction.

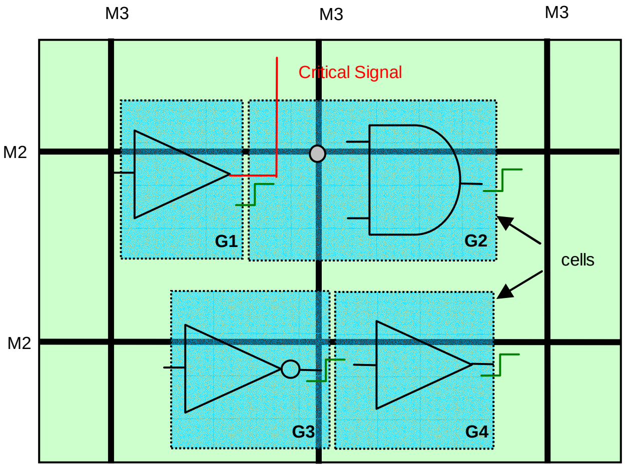

Figure 10 Example showing cells against power grid

The example in Figure 10 shows the placement of four cells on a power grid. The horizontal metal layer is M2 and the vertical layer is M3. For the sake of simplicity we will ignore the ground wires and consider only the power (Vdd). The via of interest is shown as a gray circle under cell bounding box of G2 and connects M2 and M3 vertically. Gates G1 and G2 are powered directly by this via. Gates G3 and G4 also influence this via, but not directly connected to it.

To maximize the droop in this via, all four cells must have a 0->1 transition, as shown in the figure by a rising pulse. The impact is assigned to only G1, since it has a small slack (on critical path).

A GFM representation of the Example is shown in Figure 11. There are several combinations of G1,G2,G3 and G4 are possible, out of which fault atom 1 provides the combination that induces maximum droop on the target via. If we ignore the cells G3,G4 that are farther away and have a smaller contribution to the via current, fault atom 2 can be formulated which only considers cells G1 and G2.

Fault atom 1:(total current=12mA)
mandatory conditions:
G1=01,G2=01,G3=01,G4=01
impact: G1=slow-to-rise, delay=2

Fault atom 2:(total current=8mA)
mandatory conditions: G1=01,G2=01
impact: G1=slow-to-rise, delay=2

Figure 11 GFM Representation of Power Supply Switching Fault

6. Tool Flow

The fault extraction and simulation flow is shown in Figure 12. The procedure described in this paper has been implemented for standard cell based design.

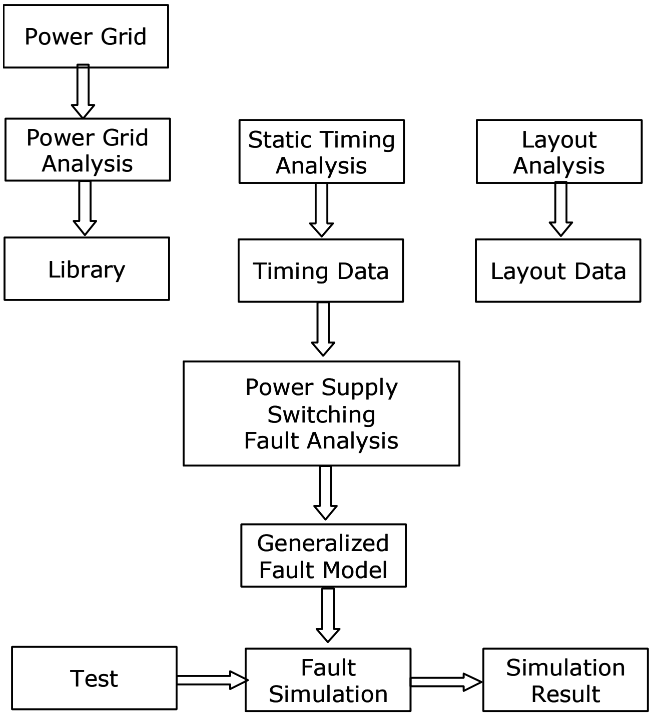

Figure 12 Tool flow for fault model extraction and simulation

7. Results

The tool flow was exercised on a standard cell based microprocessor design block consisting of approximately 128,000 cells covering an area of 2300X1000 microns.

Fault extraction was performed on the Vdd rail for the lowest via layer, via2 (M2-M3). The maximum via current distribution is shown in Figure 13.

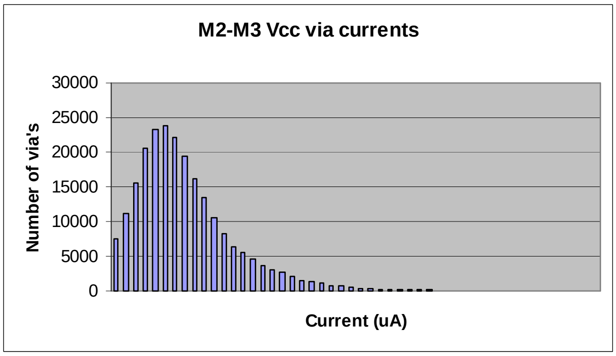

Figure 13 Histogram of peak via current over all vias

It can be seen from the figure that the distribution has a long tail with a small number of vias that have large current. This is the focus for fault extraction.

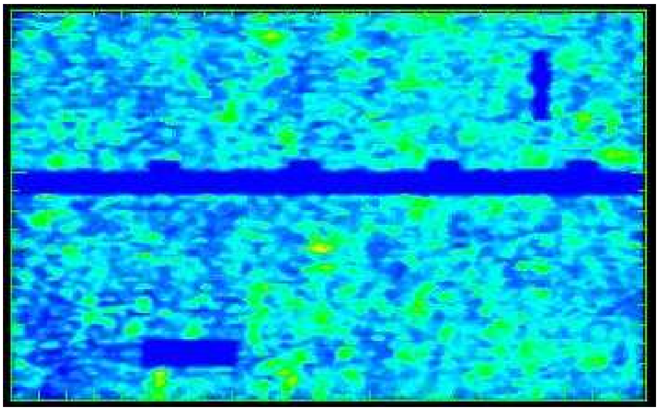

Figure 14 Color coded visualization of via currents

Figure 14 shows the same via current data as a thermal map on the actual design block floor plan. The red and yellow regions correspond to the high current vias.

A number of fault extractions were performed by varying the extraction conditions. Typically the via current threshold was set at 80% of the maximum. We also pruned any cell that contributed less than 10% of the total via current from the aggressor list. When the slack threshold was set at 140ps, the extracted fault list consisted of only 69 faults. In this fault list, the the largest number of aggressors for a fault was 13 and smallest 4 while an the average number of agreesors for a fault was 6.9. Setting a higher threshold results in fewer faults (33 faults at 85%, 14 faults at 90% and 4 at 95%) that represent the tail of the distribution in Figure 13. Reducing slack tolerance increases the number of faults. At 70ps, only 82 faults are selected. This attests to high degree of localization and selectivity of the proposed modeling technique. Further, given the relatively small number of aggressors, test pattern generation for power droop is more likely to be successful.

8. Summary

In this paper it was shown that power supply switching noise is getting worse with technology scaling. We illustrated with data that delay sensitivity to power supply noise is also deteriorating. Compounding of these trends create unintended delays that are not currently modeled in static timing analysis. A methodology for analyzing where such excessive power supply noise may be located was presented. Since, power droop is also related to board and package level inductances, in an ideal solution, they should be incorporated as well. However, such droop does not typically have large spatial distribution. Thiis technique provides spatial localization for problem areas. In absence of proper modeling, one needs to inspect millions of power supply switching nodes. Analysis and pruning techniques presented here yield a small set of critical faults. An encoding technique of the faults to a general construct called GFM was used for simulation and ATPG purposes. Untestability analysis, Fault simulation and ATPG infrastructure is under development. Initial results on fault extraction were reported. Silicon validation of results is under progress.

Acknowledgement: The authors would like to acknowledge contributions from several members of the staff at Intel Corporation. In particular, contributions made by Sujit Zachariah and Souvik Ghosh is greatly appreciated.

9. References

[1] S.T.Zachariah, Yi-Shing Chang, S.Kundu, C.Tirumurti, “On modeling Cross-talk Faults”, Design, Automation and Test in Europe, pp. 490-495, March 2003.

[2] Angela Krstic, Yi-Min Jiang and Kwang-Ting Cheng, “Pattern Generation for Delay Testing and Dynamic Timing Analysis Considering Power-Supply Noise Effects”, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 20, No. 3, March 2001.

[3] A. I. Kayssi, K. A. Sakallah, and T. M. Burks, “Analytical transient response of CMOS inverters,” Trans. Briefs, IEEE Trans. on Circuit and Systems, Vol. 39, pp.43-45, January 1992.

[4] J. H. Wang. Current Waveform Simulation for CMOS VLSI circuits considering Event-Overlapping, IEICE Trans. Fundamentals, Vol. E83-A, No. 1, January 2000.

[5] Brion Keller et. al, “Hierarchical pattern faults for describing logic circuit failure mechanisms”, US Patent 5,546,408

[6] K. D. Dwarakanath and R. D. Blanton, "Universal Fault Simulation," Proc. 37th Design Automation Conference, Los Angeles, CA, June 2000.