Built-In-Chip Testing of Voltage Overshoots in High-Speed SoCs

Amir Attarha and Mehrdad Nourani

Center for Integrated Circuits & Systems
The Univ. of Texas at Dallas
Richardson, TX 75083-0688

Dept. of Electrical and Computer Eng.
The Univ. of Tehran
Tehran 14399, Iran
attarha/nourani@utdallas.edu

ABSTRACT

We present a methodology to detect and measure the signal overshoots occurring on the interconnects of high-speed system-on-chips. Overshoots are known to inject hot-carriers into the gate oxide which cause permanent degradation of MOSFET transistors' performance over time. We propose a built-in chip mechanism to detect overshoots, collect the occurrence information and scan them out efficiently and inexpensively for built-in self-test, reliability analysis and diagnosis.

I. INTRODUCTION

A. Motivation

With fine miniaturization of the VLSI circuits and rapid increase in the working frequency (gigahertz range) of digital system-on-chips (SoC), the signal integrity becomes a major concern for design and test engineers. Although various parasitic factors for transistors can be well controlled during fabrication, the parasitic capacitances, inductances and their cross coupling effects on the interconnects play a significant role on the proper functionality and performance of high-speed SoCs.

Core-based system-on-chip (SoC) design strategies help companies to significantly shorten the time-to-market and reduce the design cost of their new products. A core is a highly complex logic block which is fully defined, in terms of its behavior, and is also predictable and reusable [ChPa96].

In recent years, among different factors with adverse effect on signal integrity, such as delay, crosstalk and overshoot, the least attention was paid to detecting overshoots. The voltage overshoots play a major role in final performance, reliability and lifetime of a sub-micron GHz+ chip [Leb96] and thus should be considered in the testing process of such chips.

B. Voltage Overshoot – Cause and Effects

Distributed and coupling capacitances and inductances on the interconnect may cause a signal with very short rise time \( T_r \) (e.g. a high-frequency signal) to momentarily exceed \( V_{dd} \). This phenomenon, called overshoot, is pictured in Figure 1.

The effect of overshoot is not only rising signals to the values larger than \( V_{dd} \) but also contributing to delay [Brew91], noise [KuGh97] and hot-carrier damage [MHPF94][CKMC94][CTFH99] in MOS transistors. Among these factors, the hot-carrier degradation may leave a permanent scar on MOS devices [WBK91][MHP94]. In sub-micron ranges, because of the significant increase of the horizontal and vertical electric fields in the channel region of MOSFET transistors, electrons and holes gain high kinetic energies. These high-energy electrons and holes, also called hot-carriers, may penetrate the gate oxide and cause permanent changes in the oxide charge distribution, and ultimately degrade the current-voltage characteristics of the transistors [Leb96]. Such performance degradation over time creates serious reliability concern [Leb96].

Authors in [FTCH98] [FaLY94] and [CTFH99] present empirical and simulation evidences showing the impact of overshoots on the transistor's reliability and performance. For example, [FTCH98] shows that many circuits that are under hot-carrier attack reach an unacceptable speed degradation over time (e.g. after 10 to 1000 hours). The experimental results in [CTFH99] show that an inverter delay increases up to 49% under severe overshoot (e.g. large number or large spike value) occurrence.

Note that the overshoots and undershoots do not propagate via static CMOS logic [SiGB99]. However, damping noise using such gates (e.g. CMOS buffers) on long interconnects is limited because of the performance drawback. Moreover, even for the I/O buses that can be observed directly, it is extremely difficult to see/test the overshoots by probing the pads using test equipments since the CMOS logic (e.g. buffers used in pads or the parasitic RLC of the probes) suppresses the overshoots.
C. Related Work

Many of the existing gate and transistor level simulators can not analyze the signal integrity (e.g. crosstalk or voltage overshoot) in a circuit accurately. An accurate analysis requires exact extraction of capacitance and inductance values from the layout, considering correlation among the nearby signals (e.g. coupling capacitance and inductance between wires), precise computation of the arrival time and shape of signals [KuGh97][NTOG98]. Recently, some researchers addressed different issues about crosstalk. Accurate analysis [ChKu99], test generation for crosstalk noise [ChGB98], [LeNA98], and fault modeling [CDBJ99] are some examples.

Almost all of the works related to the signal overshoots discuss the physical cause [FTCH98], analytical model using parasitic parameters [Leb96] or accurate simulation and analysis [KuGh97][NTOG98]. The impact of overshoots on device hot carrier reliability and delay were also investigated in [CTFH99][FaLY94][FTCH98]. BERT simulator, discussed in [TRCL93], analyzes the hot-electron degradation in MOSFET, bipolar and BiCMOS transistors and predicts circuit failure rate due to the oxide breakdown and electromigration. BERT works in conjunction with a circuit simulator, such as SPICE in order to simulate reliability for actual circuits. In [NTOG98], the authors proposed a technique to solve the transmission line equations for accurate simulation of the interconnect between cores. [SIGB99] and [GZPY00] proposed a test generation procedure for generating test vectors to detect functional errors caused by overshoots, undershoots and crosstalk.

Our main contribution in this work is in proposing an on-chip mechanism to detect the voltage overshoots for high-speed SoCs. Such built-in test strategy does not require any external probing or signal waveform monitoring. Instead, the overshoot detector cells monitor signals received by a core (e.g. from the system bus) and record the occurrence of overshoots over a period of operation. The overshoot information accumulated by these cells can be compressed and eventually scanned out for final test and reliability judgment.

II. INTERCONNECT MODEL AND TRACING NOISE

Aggressive improvement in deep submicron technology and passing gigahertz frequency created concerns regarding the signal integrity degradation which may lead to low reliability. One of these integrity elements is signal overshoot. As the feature size (λ) and the interconnect thickness (t) decrease and the chip size (A) and clock frequency (f) increase we expect to frequently see signal overshoots and undershoots with larger magnitude relative to $V_{dd}$ [GZPY00][NTOG98].

To consider the overshoots/undershoots (also called noise) accurate simulation tools are needed to trace noise in high frequency systems. Such simulation tools use the distributed model that runs much slower than the lumped model because they deal with too many parasitic R, L, C values and their couplings [NTOG98].

A. Interconnect Model

A VLSI interconnect can be modeled in a number of ways with various accuracy and computational overhead. Generally, when an interconnect wire becomes sufficiently long or the circuits become sufficiently fast, the inductance of the wire starts to dominate the delay behavior, and it should be considered in the interconnect modeling. In spite of considering R, L and C, lumped model can not demonstrate the real behavior of the interconnects, therefore the distributed model has been proposed.

The distributed RLC model of a wire has become one of the most accurate approximation of its actual behavior [Veen98]. Considering coupling capacitances is an inseparable issue in accurate modeling of interconnects. Unintentional couplings from the neighboring interconnects affect the signal integrity significantly and cannot be ignored in high frequency [NTOG98][Shee99][SKEW96].

There are many good distributed models in the literature [CLL00][Veen98]. We do not advocate any specific interconnect model since the accuracy of the interconnect line model is beyond the scope of this paper. However, for the purpose of reporting the experimental results we use the distributed RLC model with coupling capacitances, reported in [Shee99]. This model is shown in Figure 2.

Figure 3 shows the SPICE simulation [TISP94] results for a 10 mm interconnect line affecting a signal with $T_{r} = 0.05 \text{ ns}$ ($f \approx 1.5 \text{ GHz}$) using the lumped and distributed RLC models. SPICE does not reveal noise (overshoots, undershoots etc.) for the lumped model because such simplistic model does not show the true behavior of a high frequency signal traveling along the interconnect.

III. DETECTING THE VOLTAGE OVERSHOOT

A. Review of Sense Amplifiers

Sense amplifiers are widely used in memory architectures (both DRAM's and SRAM's) for performance speedup, power reduction and signal restoration (e.g. refreshing in DRAM's) [Rah96][GrMo93]. Differential sense amplifiers present numerous advantages over single-ended counterparts but they are

Fig. 2. An interconnect model.

Fig. 3. Spice simulation using lumped and distributed models.
directly applicable in SRAM memories only in which data and data both are available [Raba96]. Figure 4 shows three conventional differential sense amplifiers. In all three types NMOS transistor $T_2$ is the current source (when $SE = 1$) and PMOS transistors $T_3$ and $T_4$ are loads. The positive feedbacks (drain-gate connection between $T_3$ and $T_4$) allow amplification in these structures. In-depth details can be found in [Raba96][GrMe93].

B. A CMOS Overshoot Detector (OD) Cell

The modified cross-coupled PMOS differential sense amplifier is our choice to detect voltage overshoots as pictured in Figure 5. In this figure for simplicity we showed only one bit of an interconnect (point-to-point or bus) between Core $i$ and Core $j$. The overshoot detector (OD) cell sits physically near the receiving core and samples the actual signal plus noise received by Core $j$. $SE$ is connected to test mode to create a permanent current source in the test mode and input $x$ is connected to $V_{dd}$ to define the threshold level for sensing $V_s$, i.e. the voltage received in $x$. The inverter, formed by $T_6$ and $T_7$, stabilizes the voltage levels in the output of OD cell.

By adjusting the size of the PMOS transistors (i.e. $W$ and $L$), the DC current through transistors $T_1$ and $T_2$ are set to different values. Combining this with the feedbacks between PMOS transistors creates threshold voltages to turn the transistors on or off. Figure 6 shows signals on input and output (points b and c) of the cell to validate the behavior of our overshoot detector cell.

Suppose that $V_s$ is initially low. Transistors $T_2$ and $T_3$ both are in linear region and behave as "on" switches. Transistors $T_1$ and $T_4$ are in cut-off that make $V_{th} = 0$ and $V_{sc} = 1$. As $V_s$ increases $V_{th} > V_s$, $T_1$ goes to saturation region and the current through $T_1$ starts increasing. On the other hand, $T_3$ is in linear region and its current is equal to the current of $T_1$. Therefore, when current of $T_3$ increases, based on the large model of MOSFET transistors [AIHo87], the resistance between drain and source decreases. The transistors stays in their regions, as discussed, until the $V_s$ passes $V_{th}$. After passing $V_{th}$, transistors $T_1$, $T_2$ and $T_4$ are in linear region and $T_3$ is in cut-off and therefore $V_{th}$ is forced to "1" by $T_4$ and therefore $V_s = 0$.

Although the current in cut-off region is generally negligible, sub-threshold current plays an important role in this circuit. When $V_s$ starts decreasing, $V_{dd}$ of $T_2$ decreases and we expect to see lower current flowing through $T_1$. However, $T_3$ is in cut-off region and controls the current. When $V_s$ decreases the channel length also decreases and eventually the resistance of channel increases and it results in slightly voltage increase in the drain of $T_1$ and $T_3$. Therefore, the gate voltage of $T_3$ increases and $V_{dd}$ decreases and the current of $T_4$ and the drain voltage (the resistance of channel of $T_3$ x current) of $T_4$ and $T_2$ decrease. When $V_s$ passes $V_{th}$, the $T_4$ and $T_3$ go to cut-off and linear regions, respectively, making $V_{th} = 0$ and $V_{sc} = 1$. The $V_{th}$ threshold voltage can be regulated based on transistor sizes, which cause different channel resistances and eventually different voltage drops at the drains and gates of $T_3$ and $T_4$.

The above analysis and the waveforms in Figure 6 confirm that the OD cell shows a hysteresis (Schmitt-trigger) property which implicitly indicates a (temporary) storage behavior. To confirm this we ran a DC analysis on the OD cell to get the hysteresis curve shown in Figure 7. In this figure, for example, the solid-line curve shows that the switching threshold voltages are $V_s = 2.75$ and $V_{sc} = 1.60$ when $V_{dd} = 2.5$. The hysteresis property is quite desirable since it means the OD cell not only captures the overshoots but also it filters out the signal bounces (noise) before settling down.

The unacceptable level of overshoot can be a matter of reliability debate. Researchers estimated that 0.1$V_{dd}$ or more overshoot values create hot-carriers and thus may lead to a permanent damage [CLLC00]. A nice feature of our OD cell is that for any $V_{dd}$ the overshoot threshold (i.e., $V_{th}$ of hysteresis) can be tuned by changing the layout size of the PMOS transistors (mainly $W$'s of $T_3$ and $T_4$). This is also shown in Figure 7 in which two set of transistor widths ($W = set 1$ and $W = set 2$ for $T_3$ and $T_4$) and two $V_{dd}$ values (3.3 and 2.5 volts) have been used. Analytical analysis (see [Raba96]) or a simulation-based
approach can be used for such tuning.

IV. TEST MECHANISM

Recording the occurrence of overshoot was a crucial step that is performed by the OD cells explained in the previous section. Each time that overshoot occurs (i.e., $V_k > V_R$), the OD cell generates a "0" signal that remains unchanged until $V_k$ drops below $V_R$.

The OD cells are not expensive – seven transistors per cell as Figure 5 shows. The test architecture to read out the information stored in the OD cells is a DFT decision which depends on the overall SoC test methodology, testing objective and cost consideration. In what follows we discuss some alternatives.

A. Using Compressors

In Figure 8, a compressor unit is used to compact test information (i.e., overshoot occurrence) so that $n$ bit data is compacted into $\lceil \log_2 (n + 1) \rceil$ bits. The compression unit is a simple combinational circuit that outputs total number of "0"s generated by OD cells as indication of overshoots occurring on the bus that appear on its inputs. Similar 4:2 compressor units have been extensively used in multiplier designs [ElAE97]. The circuit shown in Figure 8(a) can measure the overshoot occurrences on the $n$-bit bus over one data-transfer cycle. The result is stored in a $\lceil \log_2 (n + 1) \rceil$-bit register. If this register becomes a part of the scan chain, its content can be scanned out in $\lceil \log_2 (n + 1) \rceil$ cycles. If the overshoots are required to be measured over a period of $m$ transfer cycles (i.e., $m$ patterns) a $\lceil \log_2 (mn + 1) \rceil$-bit adder can be used as shown in Figure 8(b).

This is the size of adder and register capable of storing numbers between 0 and $mn$. This is a very pessimistic worst case scenario in which we assume all lines are subject to overshoots in all $m$ cycles. In reality, a much smaller adder and scannable register can be used to keep the statistics.

B. Using Flip-Flops

Figure 9(a) demonstrates a flip-flop based test architecture, which is able to record the occurrence of overshoot and transfer it to the output. When at least one overshoot occurs, the output of the NAND gate (OD-flag) becomes 1 which is stored in a flip-flop. This architecture only reports the occurrence of overshoot on the bus and is not able to detect a particular defective line. For diagnosis, we propose to use scannable flip-flop (OD-FF's) architecture, as shown in Figure 9(b). In the test mode, first the overshoot flag signal (OD-flag) is transferred, through MUX, to the test controller. If an overshoot occurs (OD-flag=1), the content of flip-flops (OD-FF’s) are scanned out through $S_{out}$ for further reliability and diagnosis analysis.

The very pessimistic worst case scenario in terms of test time is a case in which all lines are subject to overshoots in all $m$ cycles. This situation requires overall $m \cdot n$ cycles for readout. In practice, a much shorter time (e.g., $k \cdot n$, where $k << m$) is sufficient since the presence of defects causing overshoots is quite limited.

C. Using Counters

Figure 10 shows two other test architectures to collect information about overshoot occurrence. Figure 10(a) the output of the $n$-input NAND gate will be 1 if at least one overshoot
occurs on the n-bit input port of a core (attached to an n-bit bus) for a specific test vector. The output of the NAND gate is connected to the clock line of a counter to record number of times that the core input is exposed to overshoots. If the overshoot occurrence is measured over m cycles (m patterns) then we need a \(\lceil \log_2 (m + 1) \rceil\)-bit counter in the worst case.

If the cost is justified, we can get more accurate statistics about overshoot occurrence on each line of the bus by assigning each line to a dedicated counter as shown in Figure 10(b). This architecture will be \(n\) times costly compared to the one in Figure 10(a) but it keeps much more information about individual lines that may help in testing, diagnosis and reliability analysis. Grouping \(n\) lines into \(g\) groups in order to use \(g\) counters is a compromise architecture between the two extremes. Ultimately, the content of the counter(s) can be scanned out through dedicated scan chain or through the main scan chain.

V. Simulation Results

The experimental results here are reported for a two-core DSP processing SoC communicating through a 16-bit high-speed bus. We report the results, all obtained using SPICE [TISP94], for three important factors that affect overshoots: 1) rise time (that implies average frequency), 2) the line RLC segments (that implies the wire length); and 3) the input sets (which implies signal coupling).

In this experiments we let the two cores interact (as in normal mode) through the bus for 500 cycles (i.e. 500 patterns transferred from core 1 to core 2 or vice versa). We read out the overshoot statistics after every \(m = 100\) cycles. This means we have statistics for five signal sets each of size 100 vectors. To be conservative in recording the worst case (maximum occurrence) of overshoots we used a counting structures of Figure 8(b) and Figure 10(a) using \(\lceil \log_2 (100 + 16 + 1) \rceil = 12\)-bit adder and \(\lceil \log_2 (100 + 1) \rceil = 7\)-bit counter, respectively. The results are tabulated in Figure 11 and Figure 12.

Figure 11 is a histogram showing the overshoot percentages, i.e. the number of times (out of 100) that the receiving core detects overshoots in at least one (out of 16) lines. In this figure, the overshoot percentages are reported for different rise time (0.25 to 1.0 ns) and wire lengths (5, 10 and 20 mm) for the same set of signal data (\(S_1\)). As expected, for shorter rise time and longer wire lengths the overshoot occurrence increases.

Figure 12 shows the overshoot percentage for a fixed rise time (\(T_r = 0.5\) ns) and wire length (10 RLC segments corresponding to about 10 mm) but five different signal sets (\(S_1\) through \(S_5\)) traveling through the interconnect. As seen here, due to the effect of coupling between adjacent lines, some signal sets (i.e. \(S_2\) and \(S_4\)) cause much more overshoots than the other two.

Our second experimentation was to show the effect of overshoot on an overall system. We analyzed five main buses (data, address, control and two internal) of the famous 8051 microprocessor [Inte94]. In our implementation the 7 cores communicate through these buses and are potentially subject to overshoot in high frequency. For experimentation purpose, we assumed that the 8051 system runs in 1 GHz. This is not the actual working frequency of the 8051. We used our 30 MHz 8051 to obtain the patterns traveling on its interconnects. Then, we have applied the same patterns to the interconnects assuming that they run under 1 GHz frequency. The statistics are summarized in Table I. As shown in Table I, the average occurrence of overshoots in a presumed 1 GHz 8051 system will be 32.44% and it will cause severe damages on chip over time.

Table II compares the cost and time of different alternatives for the test architectures proposed in section IV. We assumed that the total number of \(m\) patterns are applied to test each bus. The cost of Compressor+Adder (Figure 8(b), OD-FF+MUX (Figure 9(b)) and Multiple Counter (Figure 10(b)) are more than other three architectures since they use wider components to keep the overall statistics of overshoot occurrence in all wires through the entire test session. Among these three, the Multiple Counter architecture is the most expensive one. In terms of test time for readout, the capture time is the same for all architectures since all use the same number and type of OD-cells. However, OD-FF+MUX architecture is the slowest one.

To quantify the formulas given in Table II, we show the statistics for the 16-bit address bus in 8051 reported by SYNOPSIS design analyzer toolset [Syno00] when 100 patterns are applied. All costs are expressed in terms of 2-input NAND gates. The results are summarized in Table III.
TABLE II
COST AND TIME OVERHEAD FOR DIFFERENT TEST ARCHITECTURES

<table>
<thead>
<tr>
<th>Test Architecture</th>
<th>Component Size [bit]</th>
<th>Time Test [cycle]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Compressor</td>
<td>Flip-Flops</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>[log2(n+1)]</td>
</tr>
<tr>
<td></td>
<td>Compressor+Adder</td>
<td>[log2(mn+1)]</td>
</tr>
<tr>
<td></td>
<td>OD-FF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OD-FF+MUX</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Single Counter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multiple Counters</td>
<td></td>
</tr>
</tbody>
</table>

TABLE III
TEST OVERHEAD FOR 8051 16-BIT ADDRESS BUS.

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Cost [NANDs]</th>
<th>Time [cycle]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compressor</td>
<td>222</td>
<td>500</td>
</tr>
<tr>
<td>Compressor+Adder</td>
<td>335</td>
<td>11</td>
</tr>
<tr>
<td>OD-FF</td>
<td>18</td>
<td>100</td>
</tr>
<tr>
<td>OD-FF+MUX</td>
<td>127</td>
<td>1600</td>
</tr>
<tr>
<td>Single Counter</td>
<td>47</td>
<td>7</td>
</tr>
<tr>
<td>Multiple Counters</td>
<td>784</td>
<td>112</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

We have shown that much care needs to be given to preserve signal integrity on the interconnects in high-speed SoCs. Specifically, overshoots cause hot-carrier degradation that may lead to permanent damage and reliability loss. Accurate parasitic extraction and simulation to detect the interconnect problems is very time consuming and very sensitive to the circuit characteristics and thus is not practical for large SoCs. We proposed a built-in-chip mechanism to detect, count, compress and scan out the overshoot occurrences for reliability analysis.

REFERENCES


116