

The Titanic: What Went Wrong!

Organizer

Sani R. Nassif

IBM

Research, Austin, TX

Abstract:

We often hear about success stories in EDA. We are all justifiably proud of the impact we collectively make on the overall integrated circuit design and manufacturing machine. It is fair to say, however, the one learns far more from failure than one does from success. In this special session we found several brave practitioners who are willing to talk about problems in business-as-usual EDA. These problems include technology related issues; reliability related issues, power issues and even methodology issues – In short, covering a wide swatch of the EDA domain.

Metal Variation-Induced Hold Time Failures

Author: Paul S. Zuchowski - IBM, Burlington, VT

This talk will discuss hold time failures, which occurred sporadically on certain wafer lots, and the improvements to the timing methodology that resulted from the root cause analysis of the fails. The fails were traced to the differential variation in mean metal resistance between two metal layers. The mean resistance values on each layer were within allowable process limits; however traditional ASIC static timing methods did not uncover the design flaw that led to the low yield. In a standard ASIC timing flow, metal parasitics are extracted assuming all metal layers track, meaning if one layer has a high resistance they all do. In reality, the resistance of a metal line is driven primarily by the width and height of the metal on that particular layer. Each metal level is fabricated independently of the other metal levels on the same wafer. They are processed on different days, and potentially on different tools within the fab. Therefore, the resistance of a metal line on one layer has a degree of statistical independence from the resistance of a metal line on a different layer. Using a corner based ASIC timing methodology on a typical design with 8 layers of metal, assuming 2 resistance corners for each level of metal, 16 metal extraction runs would be required. To cover the possible metal variation corners an additional 2^8 times the number of timing runs would be required. As a result of this design, a new timing methodology was developed called Variation Aware Timing. This timing methodology provides a method for simultaneously analyzing all of the various combinations of intra-die systematic metal parameters using a single metal extraction run. The new timing methodology required changes to the ASIC global wire extraction algorithms, the timing optimization tools, the noise analysis tools and of course the timing engine. This new method successfully identified the failing hold time violations on the subject design and is in production on all 130 nm and 90 nm IBM ASIC designs.

Design for Reliability: acknowledging aging effects

Authors: Claude Moughanni, Mohamed Moosa, Gary Anderson - Freescale, Austin, TX

The networking and high performance computing market is driving digital SoC requirements for high performance at low power while applying steady cost reduction pressure. To meet these market demands leading IC design and manufacturing houses, such as Freescale Semiconductor, currently aggressively shrink transistor channel lengths and gate oxide thickness in their manufacturing processes. In the most recent technology nodes, where increasing levels of nitridation are used to enable oxide thickness reductions, products exhibit aging due to phenomena such as Negative Bias Temperature Instability (NBTI) that can significantly constrain the maximum achievable performance and reliability at each scaling point. Where NBTI delta over product life was much less than 1% in older technologies, a delta of 5% or more could be observed in newer technologies, increasing the product performance and reliability guardbands that would be required in the absence of adequate Design For Reliability CAD tools.

NBTI primarily impacts PMOS devices, resulting in an increase in their threshold voltage and a concomitant reduction in their current drive over time. The rate at which devices degrade is found to be strongly related to the supply voltage bias (V_{dd}), Nitrogen content in the gate oxide, and operating temperature. The increase in V_t of the PMOS device impacts P-N ratio sensitive circuits, possibly causing them to drift outside of target specification over the product operational life. In addition, the combination of increased V_t and reduced current drive can adversely impact the maximum frequency of the product.

Empirically, it has been observed that the path limiting a products' maximum frequency changes over time. Moreover, the critical path at any one product age is not the same across the whole product population within a given manufactured lot. A predictive device model for such behavior can prove formidable. Furthermore this type of behavior challenges the conventional SoC integration and validation design flow, potentially requiring that probabilistic critical path analyses be done at several different age points within the expected product lifetime.

Other mechanisms such as hot-carrier injection (HCI), time-dependent dielectric breakdown (TDDB), and interconnect electromigration (EM) further complicate the DFR space since they all contribute to different aging/failure effects and yet are not totally independent from each other. The challenge to the EDA industry and to academia is to ensure the availability of time- and compute-efficient predictive device models, design tools and design flows. The objective is a DFR signoff flow that serves as a quality gate prior to committing products to manufacturing.

Copyright is held by the author/owner(s).

DAC 2005, June 13–17, 2005, Anaheim, California, USA.

ACM 1-59593-058-2/05/0006.

Bridging the Power Reduction & Estimation Gap in the Cell Processor Design Methodology

Author: Stephen D Posluszny - IBM, Austin, TX

Sony, Toshiba and IBM started collaboration on the Cell processor chip in 2001 with some very aggressive goals. Not only were we defining a new high performance microarchitecture, but we had extremely aggressive cycle time targets and power constraints. The cycle time was set at 11FO4 (delay of a chain of 11 inverters, each driving 4 fanout). This high frequency with 234 million transistors would naturally translate to a very high power chip. We had to find ways to reduce power in the architecture, microarchitecture and circuits. The initial set of tools and design methodology was based off the IBM Power 4 microprocessor flow. As you can imagine, power consumption was not the main concern of high performance server processors in the 90's like it is today. We had a gap in our circuits, tools and methodology when it came to controlling power consumption. This panel presentation will cover how the team met these challenges with new circuits, a consistent design methodology, and in particular a new tool to estimate dynamic power.

Out of the 11FO4 cycle time, approximately 2-3FO4 was consumed by the latch delay. Due to their high contribution of delay and power consumption, several different latch topologies were designed and used on the chip, including a low power pulsed-clock latch. A local clock buffer (LCB) was designed with an activate signal to turn on and off the clock, providing fine-grained clock gating. In order to gauge the effective use of the fine-grained clock gating, we required a new power estimation tool which would give cycle-by-cycle power usage. Due to the high number of custom circuits and high frequency of operation, we also wanted the accuracy of a schematic based power estimation tool. Realizing that from a runtime perspective, we could not run thousands of simulation cycles with a circuit simulator on the complete chip, we combined cycle based behavioral simulation at the chip level with a circuit simulation at the macro level. This gave us the runtime performance of chip level RTL simulation with the accuracy of circuit simulation.

Power rules were created for each macro using a high speed circuit simulator. The power rule provided power estimates with zero and 50% input switching factors, with and without clock gating. Global wire capacitance was also used in the tool to estimate power used to switch macro-to-macro signals. For each cycle of RTL simulation, the actual macro input switching and clock activation would be used to approximate each macro's power by extrapolation from its power rule. All the macro power would then be added to the global wire switching power to predict the total chip power for that cycle of operation. Different program traces could then be run to estimate power usage in different circumstances. "StopPower" and "MaxPower" traces were extremely useful in optimizing and checking the clock gating effectiveness. The tool gave logic designers feedback about how much power was burned in particular states ("StopPower" in particular) leading them to find opportunities for further power reduction. The reported cycle-by-cycle power differential was also very helpful in the power distribution analysis. We were also able to use this new tool early in the design process as soon as enough VHDL was complete to run some instructions. The power rule for an unfinished macro was extrapolated from the macro's estimated size and average power density on the chip. Capacitance for unrouted wires could also be estimated from Steiner routes in the floorplan. This new tool helped the design team make the most of new circuit and microarchitecture innovations and keep the power consumption of the Cell microprocessor down to a minimum.

A Series of Unfortunate Events

Authors: Ward Vercruyssen - AMD, Austin, TX

A clear goal, abundant resources, a well staffed project, a ship that was a marvel of technology, far exceeding the requirements - and nothing really unexpected happened. Even the icebergs did what they were supposed to do. However, the mission failed. In this presentation, parallels will be drawn between the series of unfortunate events that led to the Titanic disaster and series of unfortunate events that causes large chip projects to fail.