Silent Data Corruption – Myth or Reality?

Cristian Constantinescu, AMD
Ishwar Parulkar, Sun Microsystems
Rick Harper, IBM Research
Sarah Michalak, LANL

Abstract

The higher complexity of the hardware and software employed by modern computing systems, as well as semiconductor technology scaling, are increasing the likelihood of Silent Data Corruption (SDC). SDC occurs when incorrect data is provided to the user, e.g., written to the memory or I/O system, and no error is triggered. Such events may have catastrophic effects, in the case of life critical applications, and represent a significant cost penalty for businesses. The purpose of this panel is to provide real examples of silent corruption, and discuss solutions for avoiding it. The presentations address SDC generated at the semiconductor device level, as well as the virtualization software level. Techniques for reducing SDC, from the circuit to system level, will be covered. Results of an extensive SDC study, carried out at Los Alamos National Laboratory (LANL) on high-performance computing (HPC) platforms are also given.

Silicon Faults and Silent Data Corruption

Cristian Constantinescu
Advanced Micro Devices

Scaling of semiconductor devices, as well as the increased complexity of the integrated circuits, has led to higher error rates due to transient and intermittent faults. These faults may primarily occur as a result of neutron and alpha particle strikes, ultra-thin oxide breakdown, crosstalk, process variations, manufacturing residuals, electromigration, and negative bias temperature instability (NBTI). Intermittent faults may be also activated by environmental conditions and workload. Errors induced by transient and intermittent faults are capable of silently corrupting the data processed by the affected devices.

This presentation provides several examples of SDC due to hardware transient and intermittent faults. For instance, the impact of high energy neutrons and electrostatic discharge is discussed.

High volume off the shelf integrated circuits, in general, and microprocessors, in particular, rely on error detection and correction circuitry, for avoiding hardware induced SDC. Parity, ECC and CRC, for instance, have become ubiquitous in today computing systems. However, the increased complexity of the fault-tolerance mechanisms has led, sometimes, to new error sources. These errors particularly manifest when circuits operate in "corner cases", difficult to validate. Examples of design bugs, responsible for silent corruption of the data are provided and solutions for avoiding such bugs are discussed.

SDC - Enterprise Data Centre Customers' Worst Nightmare

Ishwar Parulkar
Sun Microsystems

We define SDC as incorrect data being generated in hardware at some point during execution, and data being passed on to the software layer, without being detected for a period of time, that could be damaging to a customer (it might get detected eventually). Sun's position is that SDC, as defined above, is a reality and investment in design features to prevent or reduce SDC is a necessity. However, there are two key factors to be considered in determining how much to invest:

1) Global factors: The application space for which the hardware is targeted - e.g. enterprise servers versus game consoles, etc.

1-4244-2398-9/08/$20.00 ©2008 IEEE

2) Local factors: The nature of the circuits, logic, microarchitecture which determines the probability of silent data corruption - e.g., SRAM cell versus AND gate, arithmetic computation logic versus branch prediction logic, etc.

SDC though often associated with soft errors from cosmic particle hits, can also arise from other sources. The different sources of SDC are: 1) soft errors resulting from cosmic particle hits, 2) process/design marginalities excited only under very specific environmental conditions, 3) hard errors resulting from latent defects due to aging in silicon, and 4) corner case functional bugs that occur very rarely. We give examples of how each type of source can cause SDC.

We have a comprehensive approach to minimizing SDC in our servers including irradiation test and measurement, circuit level techniques, microarchitectural/architectural techniques at the hardware layer and software techniques. We also model SDC rates in various subsystems to provide quantitative guidance in implementing features to prevent or minimize SDC. We give a brief overview of the different solutions deployed in various subsystems of our servers - the microprocessors, associated ASIC chipsets, DRAM memory subsystem, system interconnect and IO subsystem.

Silent Data Corruption in Virtualized Environments

Rick Harper
IBM Research

This panel presentation is on silent data corruption in virtualized environments. The usage of virtualization in the commercial computing market is growing extremely rapidly, and is challenging our traditional dependability techniques. This presentation briefly introduces server, networking, and storage virtualization technology, and then outlines the new opportunities that such a virtualized environment offers for silent data corruption. For example, an entirely new "virtualization layer" of software, which is in itself quite complex, is used to provide an emulation of a hardware environment. This layer itself has the potential of corrupting data, but it also comprises an architecturally promising point to detect and correct data corruption, especially that produced by the hardware. Therefore the presentation would then move to new opportunities to detect and prevent silent data corruption using this virtualization layer. Moreover, the predominant systems health management technologies for virtualized environments are "out of band," that is, they do not use agents inside a virtual machine, but instead deduce its health based on metrics that are only visible external to the virtual machine. Special techniques are therefore necessary to determine that a virtual machine may be corrupted. The presentation concludes with concrete examples of data corruption that have been observed in this environment, and present a challenge to the audience to develop better mechanisms to reduce silent data corruption in virtual environments.

Silent Data Corruption Research at Los Alamos National Laboratory

Sarah Michalak
Statistical Sciences Group
Los Alamos National Laboratory

The potential occurrence of SDC on HPC platforms and associated systems is of interest to LANL HPC community. While SDC is presumed to be very rare, because of the number of components in our HPC platforms and the scale of the scientific calculations performed on them, it is possible that scientists' calculations occasionally are affected by SDC.

To address this issue, LANL is funding a Strategic Capability research program that is rigorously investigating SDC via experiments conducted on 1) production HPC platforms experiencing typical operating conditions, 2) decommissioned systems which can be subjected to harsher test conditions and 3) test systems operated under laboratory conditions in which temperature, voltage, and frequency may be controlled. The goals of this research are to estimate the SDC rate of production and decommissioned platforms and to estimate the relationship between temperature, voltage and frequency and the SDC rate for one or more test systems. Anecdotal historical evidence is also being compiled. At this writing, SDC testing has been undertaken on one decommissioned system, and preparations for SDC testing on a second decommissioned system, a production system and on test systems under laboratory conditions are in progress.

An overview of LANL's SDC research program and results are presented. Results include a compilation of the data from thousands of SDC tests conducted via Linpack, which has known correct answer. These tests have been conducted under nominal conditions, elevated temperatures, and voltages that have been set to their upper margin or lower margin. Based on these data, bounds on the SDC rates of one or more LANL HPC platforms are presented.

1-4244-2398-9/08/$20.00 ©2008 IEEE