Accurate Microarchitecture-Level Fault Modeling for Studying Hardware Faults

Man-Lap Li, Pradeep Ramachandran, Ulya R. Karpuzcu, Siva Kumar Sastry Hari, Sarita V. Adve
Department of Computer Science
University of Illinois at Urbana-Champaign

HPCA 2008

Ankur Sharma
Apr23, 2013

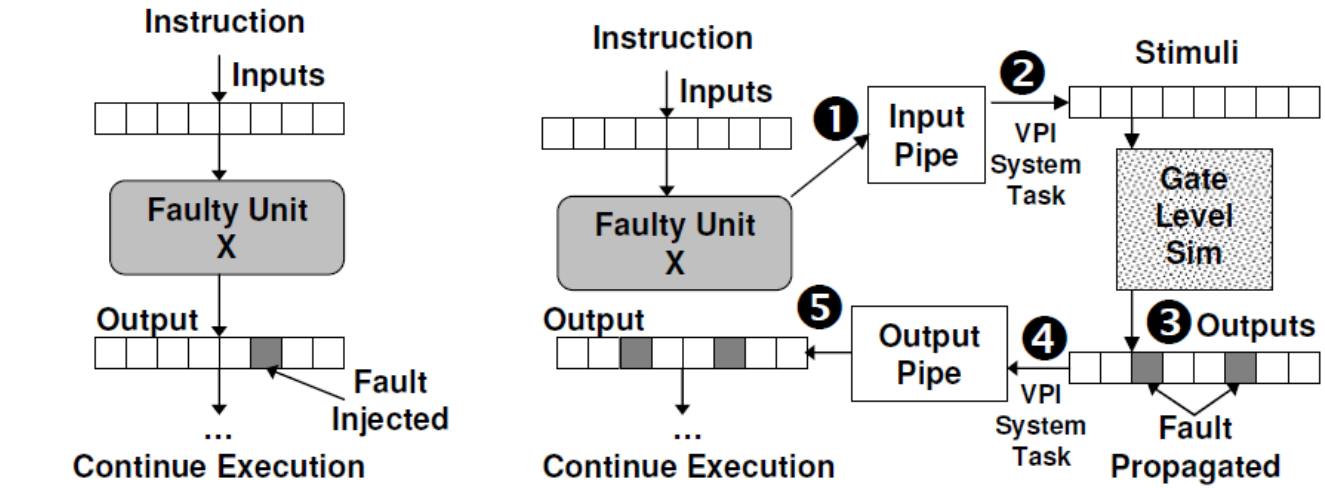
Motivation

- To evaluate the efficacy of uarch-level fault tolerance solutions, it is essential to capture the expected behavior of the fault at the level at which the solution is implemented.
- Gate-level simulations: Accurately capture low-level faults but slow
- uarch-level simulations: Less accurate but fast
- Best of both the worlds ??

Contributions

- Swat Sim:
  - a novel fault injection infrastructure for studying system-level effects of gate-level permanent faults
    - selective and on-demand gate-level simulation
    - repeated invocations of gate-level simulator during run-time
    - gate-level timing simulator coupled with arch-level simulator to model gate-level timing faults
- Show that, in general, arch-level stuck-at faults do not result in similar system-level fault manifestation as gate-level stuckat or delay faults
- Derive two probabilistic fault models, the *P-model and the PDmodel,* for gate-level stuck-at and delay faults (not covered in this presentation)

SWAT-SIM



(a) µarch-Level Fault Simulation

(b) SWAT-Sim Fault Simulation

Experimental Methodology

- Swat- Sim coupled with NC Verilog
- Faulty Components:
  - ALU, AGEN, Decoder
- Fault models:
  - uarch stuckat 0 and 1
  - gate level stuckat 0 and 1
  - gate level delay
    - gate delay increased by 1 cycle
    - gate delay increased by 0.5 cycle
- Possible outcomes: uarch-mask; arch-mask; app-mask; detected; detected > 10M; SDC
- Coverage = **percentage of unmasked** faults that are detected within 10 million instructions

Results - Performance Overhead

Unit	Fault Model	Max	Avg
ALU	Gate Stuck-At	2.20	1.56
	Gate Delay	2.65	1.93
AGEN	Gate Stuck-At	1.59	1.26
	Gate Delay	1.89	1.35
Decoder	Gate Stuck-At	2.91	2.12
	Gate Delay	5.10	2.91

Table 2. Slowdowns of SWAT-Sim when compared to pure μarch-level simulation.

Results – Accuracy of Uarch-Level Model

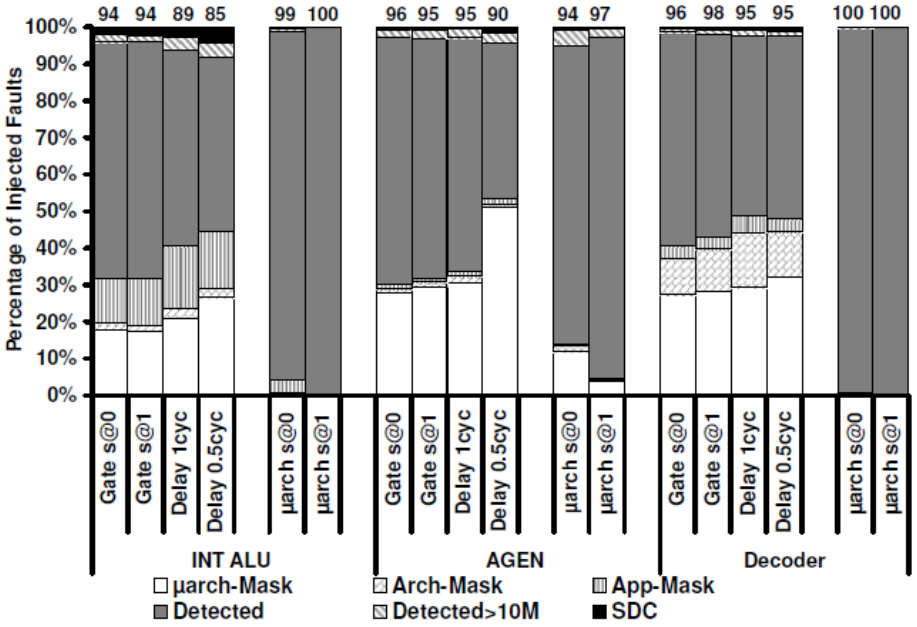


Figure 2. Efficacy of the SWAT fault detection scheme [12] under different fault models for the ALU, AGEN, and Decoder. Depending on the fault model and the structure, the μarch-level fault may or may not capture the system-level effects of gate-level faults accurately, as indicated by the differences in coverage.

- Lower coverage for gate-level models
- High masking rate for gate-level models
  - fault may not be activated
  - fault may not propagate
- uarch-level models directly change the latch value

Differences Between Fault Models
Fault Activation Rate

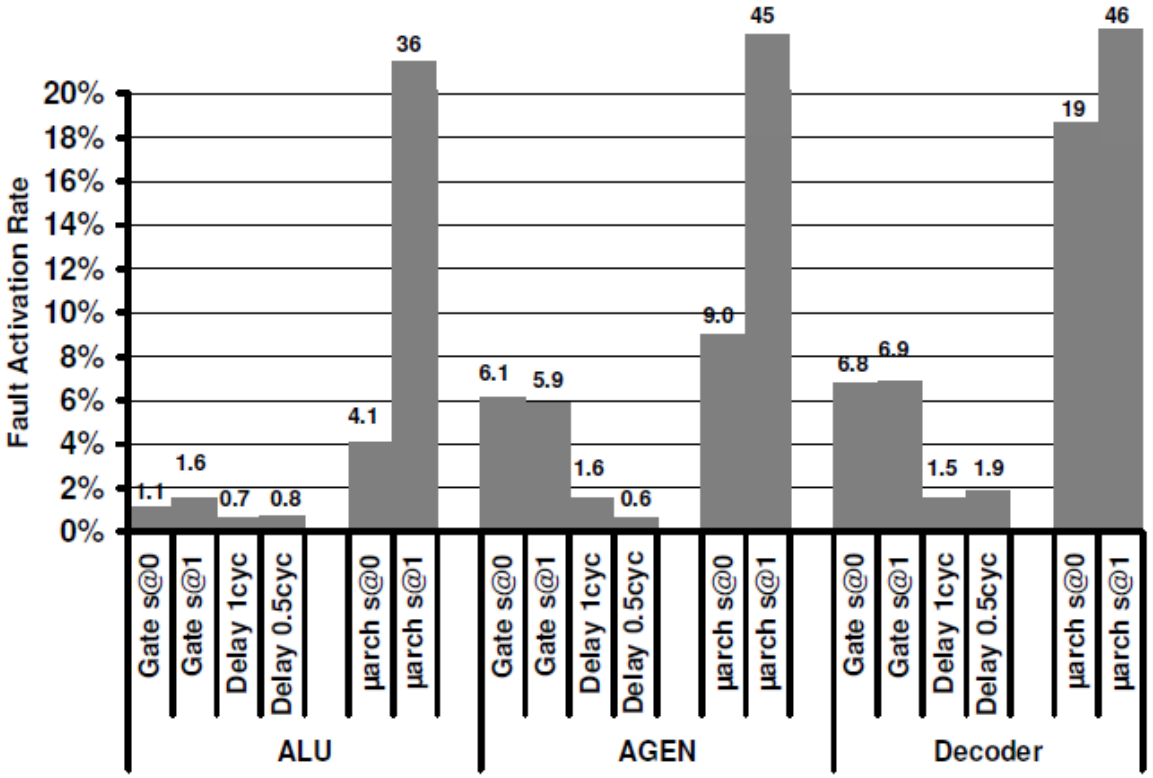


Figure 4. Mean fault activation rate for the different fault models as a percentage of the number of instructions.

Fault activation rate: % instructions that are corrupted by the fault among all instructions using the hardware.

- Activation rate lower for gate-level
  - Excite and propagate
  - uarch-level fault directly injecte
- Delay gate-level activation rate lower than stuck-at gate level
  - lower excitation probability.

Differences Between Fault Models
Bit Corruption Pattern

	ALU					AGEN					Decoder				
Bits	1	2	4	8	9+	1	2	4	8	9+	1	2	4	8	9+
μarch	100.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	72.5%	0.2%	4.8%	8.9%	13.4%
Gate s@1	91.1%	4.7%	1.2%	1.1%	1.9%	87.1%	6.8%	5.0%	1.0%	0.1%	66.1%	14.9%	10.5%	6.2%	2.3%
Gate s@0	84.4%	4.6%	2.8%	1.1%	7.1%	75.5%	8.4%	8.6%	7.4%	0.0%	60.8%	22.3%	12.2%	2.6%	2.2%
Delay 1cyc	90.4%	3.9%	1.4%	1.1%	3.2%	90.5%	4.1%	3.7%	1.5%	0.2%	71.7%	11.1%	12.5%	1.7%	2.9%
Delay 0.5cyc	75.0%	5.8%	2.2%	3.9%	13.1%	83.7%	7.9%	3.1%	2.4%	2.8%	68.2%	12.8%	4.3%	2.7%	12.0%

Table 3. Percentage of bits incorrect at the output latch.

For ALU and AGEN:
uarch-level faults injected in atmost 1 output latch bit but gate-level model can corrupt multi-bits. Most common case: only 1 bit in the output latch is corrupted.
For Decoder:
uarch-level faults injected in input latch corrupt 8+ bits in >22% cases due to large output cone.

Probability of bit flip

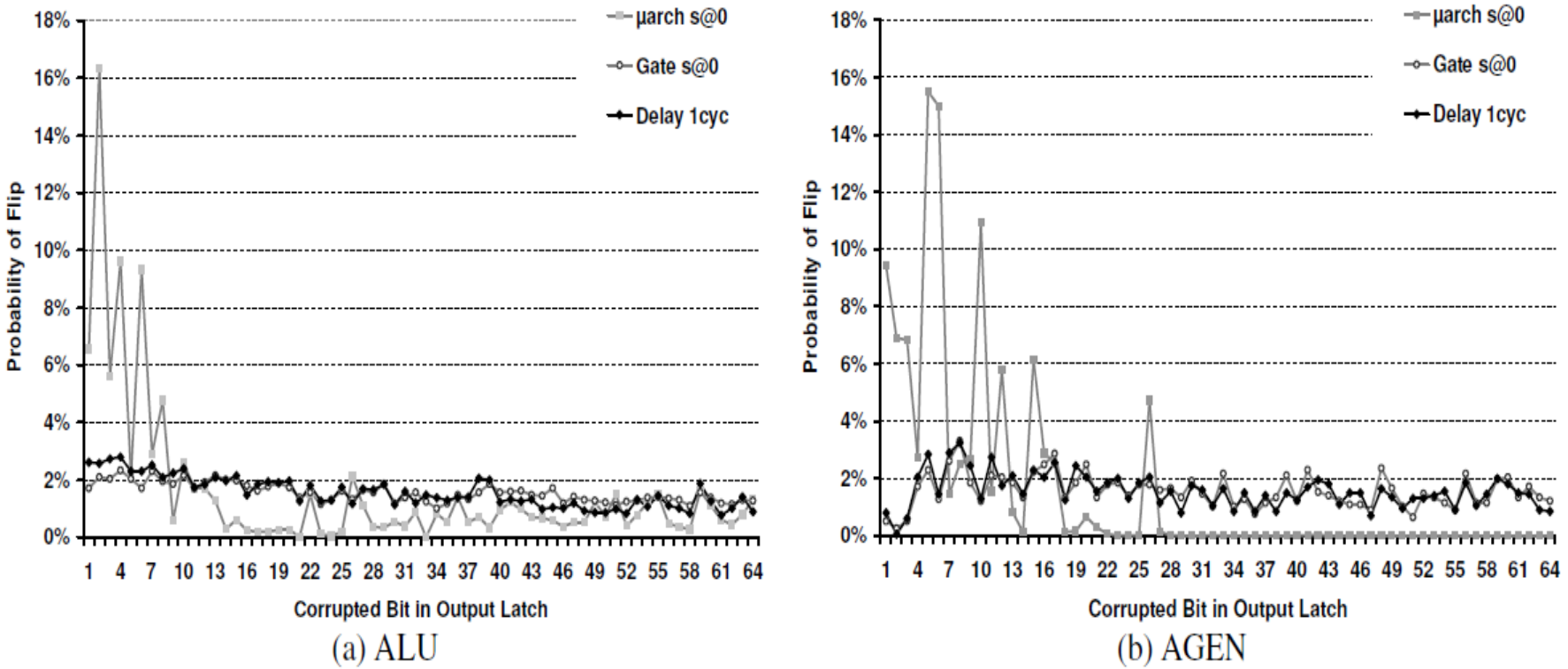


Figure 5. Probability of corrupting each bit of the output latch, under μarch-level s@0, gate level s@0, and gate level delay models.

Another source of discrepancy of the arch-level model to represent gate-level faults:

- uarch-level model corrupts an output bit with much different probability than gate-level model

- Higher activation rate implies more chances of detection => better coverage. Therefore, uarch-level models have better coverage.
- Multi-bit corruption improves the coverage even for low activation rates. Therefore, coverage of gate-level model can be similar to uarch-level model