

# On-Chip Diagnosis for Early-Life and Wear-Out Failures



Matthew Beckler



Shawn Blanton

The impact of manufacturing variations at future technology nodes is expected to have a significant impact on yield, both at the time of manufacturing and later in the field. As a result, chips and systems must be designed with robustness in mind to overcome these challenges. The Stanford-Carnegie-Mellon approach achieves robustness by (i) on-chip testing at runtime, (ii) identifying the location of any faults, and (iii) repairing the affected modules. One way to locate faults is through the use of fault dictionaries. A fault dictionary contains pre-computed circuit responses in the presence of modeled faults, and can be used to identify specific faults affecting the circuit under test through a dictionary-lookup operation. Conventional fault dictionaries require too much memory for on-chip use however, and are limited to locating only simple faults. A new X-based fault model has been developed for creating highly-compacted on-chip fault dictionaries for localizing early-life chip and wearout failures that occur during runtime. Additionally, on-chip hardware for diagnosing failures has been developed.

Initial experiments use module-level representations of the ISCAS-85 benchmarks as well as core/uncore components from the OpenSPARC T2 design (e.g., L2B). Fig. 1 compares fault dictionary sizes across the compaction techniques in use. Dictionary size is expressed in bits, plotted on a vertical log scale, normalized against the full-response dictionary size (blue). Converting the full response to a single pass/fail response bit reduces the dictionary size (red bar). Due to the focus on module-level fault localization, a number of faults can be eliminated from consideration due to their equivalence with or subsumption of other faults in the same module (green).

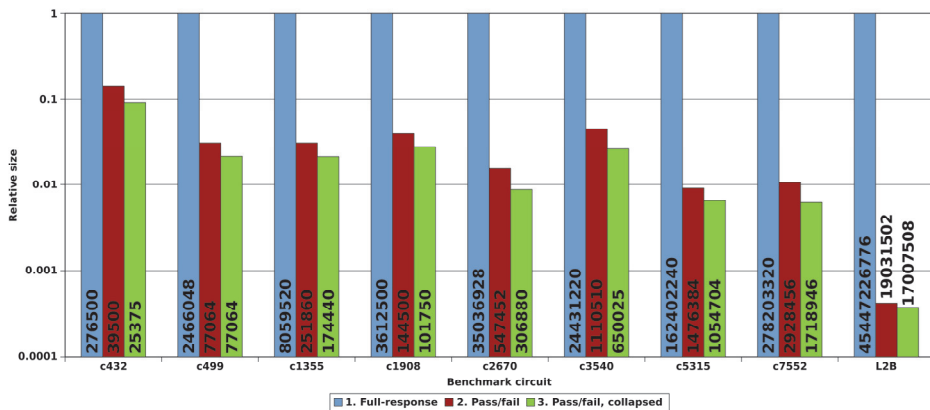


Fig. 1: Fault dictionary comparisons based on a relaxed, repair-level of diagnostic resolution.