

Fun Size Your Data: Using Statistical Techniques to Efficiently
Compress and Exploit Benchmarking Results

David J. Lilja
Department of Electrical and Computer Engineering
University of Minnesota

The wide variety of benchmark programs available today makes it relatively easy to generate copious amounts of performance data. Actually understanding all of that data, though, often can be rather challenging. Fortunately, there are many statistical techniques that have been developed to aggregate large amounts of data into meaningful information. The common mean and variance, for instance, compress a collection of data into a simple two-number summary. However, this compression throws away a great deal of useful information. This talk will describe how statistical design-of-experiments techniques can efficiently summarize complex benchmark results to extract new insights and to exploit the results in unexpected ways. As examples, we demonstrate how these techniques can be used to classify benchmark programs according to their impact on a system, and to perform microarchitecture-aware physical floor-planning.