

# Electrical Engineering Department

## Telecommunications & Signal Processing Seminar

*Fast Parallel Algorithms for Universal Lossless Source Coding*

By

Dr. Dror Baron  
Department of Electrical and Computer Engineering  
Rice University

Date: February 4, 2004  
Time: 11:00 A.M. – 12:00 P.M.  
Place: Room 342 Zachry

### **Abstract**

Most data compression research in recent years has focused on lossy compression for audio, images, and video, but lossless source coding is still important for compressing text files, executables, financial and medical data, etc. When the statistics of the source are unknown, a universal method that estimates a model for the source must be used. This talk focuses on fast (and hardware-implementable!) parallel algorithms for universal lossless source coding.

We first provide improvements to the previous state of the art in serial (non-parallel) universal compression. By combining efficient prefix tree constructions with a two-pass approach, we obtain an  $O(N)$  non-sequential encoder whose redundancy with respect to any (unbounded depth) tree source is  $O(1)$  bits per state above Rissanen's bound on the best universal compression.

The talk then considers parallel universal compression. We first present the use of two-part codes for distributed and parallel universal compression of i.i.d. sequences, and then describe our parallel compression algorithm for tree sources, which is our main contribution. We partition a length- $N$  input into  $B$  blocks, accumulate statistical information on all  $B$  blocks in parallel, estimate the single minimum description length (MDL) source underlying all  $B$  blocks, and encode the blocks in parallel. We provide an  $O(N/B)$  complexity parallel algorithm that compresses almost as well as the best serial algorithms.