The UCSC Kestrel Sequence Analysis Server

David Dahle, Mark Diekhans, Leslie Grate, Eric Rice, Richard Hughey
Department of Computer Engineering
Jack Baskin School of Engineering
University of California, Santa Cruz, CA 95064
rph@cse.ucsc.edu

The UCSC Kestrel Project has built a programmable parallel processor with 512 processing elements (PEs) on a single PC-compatible board. The array features a 1.4 million transistor chip composed of 64 byte-wide PEs, each with its own memory. The system architecture was designed specifically to optimize a large variety of computational biology algorithms.

In this poster, we present an update on the Kestrel system, including discussion of the Kestrel WWW servers for Smith & Waterman (SW) and hidden Markov model (HMM) sequence searches.

The Kestrel performance on the Smith & Waterman algorithm, which can be run from our WWW page, is 20 times faster than a 433-MHz Alpha. The Kestrel WWW interface includes SwissProt and NRP searches, and provides sequence alignments and links to MedLine entries.

Kestrel is also designed for using and training the linear (profile) hidden Markov models introduced by the UCSC computational biology group. Kestrel HMM search is also 20 times faster than a 433-MHz Alpha.

We are currently integrating Kestrel as a back-end for the SAM-T98 structure prediction WWW servers. These servers build HMMs particularly suited for homology-based structure prediction. SAM-T98 is one of the most promising fully automatic homology-based protein structure prediction methods, and with Kestrel models can be built and used in minutes rather than hours.

We are also developing a computational chemistry application for conformational fingerprinting. In this application, the thousands of possible conformations of each molecule are mapped to the Kestrel array. Each group of PEs calculates the presence or absence of specific triangles of atoms, and these results are combined to form a 6726-bit fingerprint of the conformational space. Fingerprints on Kestrel are generated 35 times faster than on a single processor of an SGI Origin 2000.

The Kestrel prototype board currently runs at 20 MHz. We plan to redesign the system for a factor of 2 speedup. Future implementations of the Kestrel architecture could produce a system 10–20 times faster than the current system.

The UCSC Kestrel group includes: Elizabeth Avila, Albion Baucom, David Dahle, Kevin Delaney, Andrea di Blas, Mark Diekhans, Leslie Grate, Cricket Grice, Jeff Hirschberg, Kevin Karplus, Hansjörg Keller, Rachel Karchin, Jennifer Leech, Justin Meyer, David Pease, Eric Rice, Osama Salem, Don Speck, Ebin Warner, Raymond Wheeler, and Douglas Williams. This project is supported by NSF grant MIP-9423985 and the Affymax Research Institute. Additional information and references are available from

http://www.cse.ucsc.edu/research/compbio
http://www.cse.ucsc.edu/research/kestrel