Technical Progress Report 2/1/96 – 4/31/96
Construction of a Connectionist Network Supercomputer
University of California, Berkeley
ONR URI Grant No. N00014-92-J-1617

1 Abstract

This report presents a summary of the technical status for the period 2/1/96–4/31/96.

Use of our SPERT neurocomputing board is becoming more widespread as the boards have now been distributed to six other sites. Along with the hardware and software, we provided a one week training course in Berkeley for programmers from the sites. The high level software effort also moved forward, with the beta version of Sather 1.1 nearing release. In our analog auditory preprocessor research, we have completed characterization of our low power Markov model decoder chip. One of our graduate students has turned in his Ph.D. dissertation on related algorithmic approaches.

2 Technical Status

2.1 The SPERT System

The focus of the SPERT system work this quarter has shifted from design and prototyping to production and quality control. Seven boards have been shipped to 6 sites worldwide. This quarter we expect to send additional boards to these sites, plus one or two new sites.

To reduce the time-to-productivity, we recommend each of our SPERT board sites send someone to Berkeley for a one week training course. To date, five of them have done so, most recently during late February. By having the developers available in the same building, the trainees are able to make a fast start on their particular SPERT applications.

With the testing and qualification phase of the project complete, the remaining 9 SPERT boards have been deployed to users here at Berkeley. Aside from their continuing use in our speech research, the boards are being used by researchers in unrelated fields including MPEG encoding and decoding, Priority Encoding Transmission (PET) for multimedia broadcast, and general purpose DSP applications. Additional detail can be found on the web at:

`http://www.icsi.berkeley.edu/real/spert/t0spert_apps.html`

The next production run of 32 boards is in progress and should be completed during the current quarter. Steps have been taken to improve the yield of the boards at each of the assembly steps. A run of eight new SPERT wafers was completed in April, which yielded 201 good T0 die.

A significant effort this period has been directed towards developing a fully-featured device driver for SPERT. Although not strictly necessary for basic board operation, this has many advantages:

1. Allows effective use of multiple SPERT boards attached to one host
2. Enables higher-performance I/O
3. Will facilitate an RPC-based approach for using SPERT
4. Allows more robust operation

The core driver is complete; performance enhancements and testing will proceed over the coming quater.

The application software for SPERT has made steady progress. The SPERT training and recognition programs are faster and more robust, and are now used by default for new speech experiments at ICSI. The emphasis has now moved to enhancing the usability of the library, making it easier for scientists with differing requirements and more novel approaches to use the SPERT board effectively in their research.

Our article about SPERT (enclosed) appeared in the March issue of IEEE Computer magazine.

2.2 High-Level Software

There was continuing progress in the area of high level programming for parallelism. As reported last quarter, our group has been awarded a grant from Lawrence Livermore National Lab supporting use of their Meiko CS2 super computer. We previously successfully ported pSather to the Meiko and this provides for much greater peak capacity. This quarter we had our first experience using pSather in a regular campus course and this was very promising.

We also prepared the beta version of Sather 1.1, a major new release that will be distributed this summer. There were also major advances in the library organization and design. Significant progress was made on a novel model for data and thread locality by David Stoutamire and presented as part of his qualifying paper and exam. Good progress was made in the coding of the the automatic mapping facilities for the ICSIM connectionist simulator.

2.3 Analog VLSI Pre-processors

Work this quarter focused on characterizing a recently fabricated chip that implements hidden Markov model state decoding in micropower analog VLSI circuits. The chip implements the state decoding portion of a wordspotting speech recognition algorithm developed

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
SEATTLE REGIONAL OFFICE
1107 NE 45TH STREET, SUITE 350
SEATTLE WA 98105-4631

IN REPLY REFER TO:

4330
ONR 247
11 Jul 97

From: Director, Office of Naval Research, Seattle Regional Office, 1107 NE 45th St., Suite 350, Seattle, WA 98105
To: Defense Technical Center, Attn: P. Mawby, 8725 John J. Kingman Rd., Suite 0944, Ft. Belvoir, VA 22060-6218

Subj: RETURNED GRANTEE/CONTRACTOR TECHNICAL REPORTS

1. This confirms our conversations of 27 Feb 97 and 11 Jul 97. Enclosed are a number of technical reports which were returned to our agency for lack of clear distribution availability statement. This confirms that all reports are unclassified and are "APPROVED FOR PUBLIC RELEASE" with no restrictions.

2. Please contact me if you require additional information. My e-mail is *silverr@onr.navy.mil* and my phone is (206) 625-3196.

ROBERT J. SILVERMAN

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

COLLEGE OF ENGINEERING
DEPARTMENT OF ELECTRICAL ENGINEERING
AND COMPUTER SCIENCES
COMPUTER SCIENCE DIVISION
387 SODA HALL #1776

BERKELEY, CALIFORNIA 94720-1776

May 21, 1996

Dr. Nick Bottka
Electronics Division, ONR 312
Office of Naval Research
Scientific Office Code 1114SE
800 North Quincy Street
Arlington, VA 22217-5660

Reference: ONR URI Grant No.: 144-92-J-1617

Dear Dr. Bottka:

Enclosed you will find the quarterly progress report for the research project for Professor John Wawrzynek: "Construction of a Connectionist Network Supercomputer," ONR URI Grant No.: N00014-92-J-1617. This report covers the research done during the performance period 2/1/96-4/30/96. Please direct any mail or questions regarding this contract to me.

Thank you for your continued support of this project.

Sincerely,

Theresa Lessard-Smith
Project Coordinator

cc: Administrative Grants Officer, Seattle
Director, Naval Research Laboratory
Defense Technical Information Center
Sponsored Projects Office - UC Berkeley
file

by Richard Lippmann; the design of our decoding system was done in collaboration with Lippmann.

Our characterization of the decoder test chip shows that the circuit supports input probabilities that from 0.0001 to 1.0000, for a range of 10,000:1. The output likelihoods for the system have a measured range of 40 log (base 10) units of likelihood. This range is sufficient for high-quality speech recognition work. Power consumption is a direct function of the desired input probability range. For a full range of 10,000:1, the system consumes about 3 microwatts of power. When operating with a reduced range of input probabilities, power consumption drops to as low as 141 nanowatts.

Our characterization work on this chip was done in conjunction with dissemination activities. A submission on the chip for the NIPS*96 conference was completed, and a longer journal publication is in progress. We are also investigating patents and industrial licenses of the design with the Berkeley Office of Technology Licensing.

2.4 Speech Recognition Research

We have continued our work in neural network speech recognition algorithms as a possible application target for our vector architectures. In particular, we have been looking at computationally-demanding approaches to the training of speech systems based on transitional information, which have some promise of improving the robustness of speech recognizers to a variety of acoustic degradations.

In the last few months we have completed some milestones in the development of such a system, called REMAP, which is an acronym for Recursive Estimation and Maximization of A Posteriori probabilities. The method provides a path to recognition of speech sequences using probabilities that are discriminant at the local (frame) and global (utterance) levels. In other words, a formalism has been developed that permits estimators to be trained for discrimination between local pieces of an observed feature sequence in such a way that they will also be optimized for discrimination of the correct complete sequence (e.g., sentence) from rival candidates. The training is based on an iterative procedure that is reminiscent of the Baum-Welch procedure used for estimating sequence likelihoods.

In recent months, we have completed experiments in both isolated word and continuous speech recognition using REMAP training. We have been able to show consistent improvement in error rate, while also showing an increase in the posterior utterance probability for the correct models, in accordance with the theory. This work was largely spearheaded by our graduate student, Yochai Konig, whose Ph.D. thesis (REMAP: Recursive Estimation and Maximization of A Posteriori Probabilities in Transition-Based Speech Recognition) was completed last month.

3 Recent Publications

Bilmes, J., Asanovic, K., Demmel, J., Lam, D., Chin, C-W., "Optimizing Matrix Multiply using PHiPAC: a Portable, High-Performance, ANSI C Coding Methodology" Submitted to Supercomputing '96.

Konig, Y., Bourlard, H., and Morgan, N., "REMAP - Experiments with Isolated and Continuous Speech," ICASSP 1996, In Press, 1996.

Konig, Y., Bourlard, H., and Morgan, N., "REMAP : Recursive Estimation and Maximization of A Posteriori Probabilities - Application to Transition-Based Connectionist Speech Recognition," Advances in Neural Information Processing Systems 8, In Press.

Wawrzynek, J., Asanović, K., Kingsbury, B., Beck, J., Johnson, D., Morgan, N., "SPERT-II: A Vector Microprocessor System," *IEEE Computer*, vol. 29, no. 3, pp 79-86, March 1996.