Technical Progress Report 8/1/94 – 10/31/94
Construction of a Connectionist Network Supercomputer
University of California, Berkeley
ONR URI Grant No. N00014-92-J-1617

1 Abstract

We have made progress in several areas this quarter. Highlights include:

- High-level Software: Public release of Sather 1.0, and major progress on pSather.
- Low-level Software: Design of "bare die" test software to be used for actual chip testing.
- Applications: A complete speech recognition application running with the Torrent instruction simulator.
- Network Hardware: A new network interface test chip running at 300 Mbits/second.
- System Hardware: Completion of the SPERT circuit board design and layout.
- Analog VLSI: A new communications protocol for connecting multiple chips efficiently.

In addition to the scientific and technical work, we had the occasion to present a summary of our work at two events:

- Dedication day for the new U.C. Berkeley Computer Science building, Soda Hall. (October 24) Over 50 visitors from academia, industry and government stopped by our four exhibits showing highlights of the CNS-1 project.
- ONR site visit and review. (November 3) Attended by Cliff Lau from ONR and Dave Andes from China Lake, this day long presentation by the CNS-1 project team included talks, posters and informal discussion.

The project continues to have a significant impact on the education of graduate and undergraduate students at our institution. There are currently 15 Ph.D., 1 M.S. and 2 B.S. students associated with the project (some are paid through supporting agencies other than the ONR). One M.S. student graduated and took a position with Digital Equipment Corporation as a VLSI designer.

2 Technical Status

2.1 Software and Applications

High-level software. Significant progress was achieved in the high-level software area of the project. The public release of Sather 1.0 has been successfully ported to a number of platforms and acceptance is very good. There are now hundreds of sites runnning the system and many are contributing to its development. A tutorial on Sather will soon be published which should further extend its acceptance.

The parallel version, pSather, has also made excellent progress; a complete syntax checker for the language has been incorporated into the Sather 1.0 system and is in use. The run time system has been implemented and a proof of its safety properties completed. The group was invited to present a paper on pSather to the POOMA conference on parallel object-oriented languages.

An important aspect of the software part of the project is the detailed analysis of the CNS architecture for a variety of problems of interest. A paper on this was presented at the major meeting in this area [mueller]. We are also continuing the study of how the Torrent architecture and CNS can be applied to image understanding. The results were very encouraging and we are extending these studies under funding from B. Yoon of Arpa. We have also linked this work to the pSather developments described above. Ben Gomes has completed a version of ICSIM, a Sather-based simulator for connectionist networks. This is intended to be one of the main programming systems for CNS.

Low-level software. Continued effort was put into supporting the VLSI development. Extra detail was added to the instruction set simulator, allowing a wider range of supervisor mode tests to function identically on the simulator and hardware models. Also, an environment for running "bare die" tests was created. This uses the T0 instruction cache for memory and the serial interface as I/O, allowing test code to be run on the processor without connecting expensive SRAM. To aid test code development, the same facilities were included in the register level and instruction level simulators.

The library development focused on routines needed for the phoneme probability estimator program (see below). Specifically, effort was concentrated on the functions used to introduce non-linearities into neural net computations.

Speech application. An important milestone towards the porting of the speech application to the Torrent processor was achieved. A version of the neural net phoneme estimator was coded, linked to the fixed point libraries, and run on the Torrent instruction set simulator. This function was then integrated with our speech recognition demonstration, the Berkeley Restaurant Project (BeRP). Although the recognition performance of the demo was restricted by using a simulator (vs. the actual Torrent chip), several important pieces of the final design were designed and debugged in the process. This project proved the sta-

bility of the development tools and simulators, and was useful for exercising and verifying all parts of the system.

We have also been working on modifications to our connectionist speech recognition training paradigm to make it more robust and accurate but which will require considerably more computing than we currently have available. Significant progress was made during the quarter in the development of basic theory in this area. The theory shows how to use neural networks in an interative procedure that results in a classifier that is statistically optimal, but that also takes significant inspiration from some known properties of human perception. The theory is now complete (resulting in an algorithm called REMAP [bourlard]), and we have begun some preliminary experiments.

2.2 Hardware Development

Network communication test chip. The fourth chip designed to try new network ideas was fabricated and successfully tested. The chip works at 150 MHz, or 300 Mbits/second, almost 50% higher than the design target. The margins and phase tolerance were checked and appear to be sufficient for reliable operation. Work has started on another iteration of the communications chip using a new delay-locked loop design to replace the PLL used previously.

SPERT board & T0 Chip issues. The SPERT board logic design was completed during the quarter, including a design review by an outside consultant. The circuit board has been laid out, and will be fabricated during the next few weeks.

2.3 Analog VLSI pre-processors

Several advances have occured over the past three months in our work on auditory pre-processors. Many of the advances have concerned an extension to the communications protocol used to send neural representations off chip, the address-event representation protocol (AER). In the spring we designed an extension to AER which permits many chips to send out neural representations on a shared bus efficiently.

In September, fabricated test chips incorporating the AER extension were received from MOSIS. The chips were tested and found to be functional; we subsequently built a system of seven AER senders sharing a common bus, which also worked. A paper describing this work was submitted to the Advanced Research in VLSI conference in early October [lazzarol].

The AER extension has also been incorporated into a new version of our auditory preprocessor chip described in earlier reports. This design was sent out for fabrication in early September. With this improved design, we will be able to build auditory pre-processing systems that have many different representations of sound, each computed by a different chip.

We also gave an invited talk on our auditory pre-processor work at the Biomedical Engineering Society Meeting in Tempe Arizona in mid-October [lazzaro2].

3 Publications

[anguita] Anguita, D. and Gomes, B., "MPB on T0: mixing floating- and fixed-point formats in BP learning," ICSI Technical Report, August 1994.

[adamo1] Adamo, J. M., "Development of Parallel BLAS with ARCH Object-Oriented Parallel Library, Implementation on CM-5," ICSI Technical Report, TR-94-045, August 1994.

[adamo2] Adamo, J. M. and Anguita, D., "Object Oriented Design of a BP Neural Network Simulator and Implementation on the Connection Machine (CM-5)," ICSI Technical Report, TR-94-046, September 1994.

[bourlard] H. Bourlard, Y. Konig, and N. Morgan, "REMAP: Recursive Estimation and Maximization of A posteriori Probabilities", ICSI Technical Report, In Prep.

[formella] Formella, A., "Some MPEG Decoding Functions on Spert: An Example for Assembly Programmers," ICSI Technical Report, TR-94-027, October 1994.

[lazzaro1] Lazzaro, J. and Wawrzynek, J., "A Multi-Sender Asynchronous Extension to the AER protocol," submitted, 1995 Advanced Research in VLSI conference (copy enclosed).

[lazzaro2] Lazzaro, J. P., Wawrzynek, J., and Kramer, A, "Systems technologies for silicon auditory models," Fall Meeting of the Biomedical Engineering Society, published in the abstract supplement to the Annals of Biomedical Engineering, 22:1, p. 53, 1994.

[mueller] S. Mueller and B. Gomes, "Efficient Mapping of Randomly Sparse Neural Networks on Parallel Vector Supercomputers", Sixth IEEE Symposium on Parallel and Distributed Processing, 1994.