INTERIM

IN 34-CR

O CIT

16378

7P

Semi-Annual Report
Grant No. NAG-1-1529
January 1, 1994 - June 30, 1994

RESEARCH IN PARALLEL COMPUTING

Submitted to:

National Aeronautics and Space Administration
Langley Research Center
Hampton, VA 23681-0001

Attention:

Dr. J. J. Lambiotte, Jr., M/S 125

Submitted by:

James M. Ortega
Charles Henderson Professor and Chair

SEAS Report No. UVA/528474/CS95/102
July 1995

DEPARTMENT OF COMPUTER SCIENCE

SCHOOL OF

ENGINEERING
& APPLIED SCIENCE

University of Virginia
Thornton Hall
Charlottesville, VA 22903

Semi-Annual Report
Grant No. NAG-1-1529
January 1, 1994 - June 30, 1994

RESEARCH IN PARALLEL COMPUTING

Submitted to:

National Aeronautics and Space Administration
Langley Research Center
Hampton, VA 23681-0001

Attention:

Dr. J. J. Lambiotte, Jr. M/S 125

Submitted by:

James M. Ortega
Charles Henderson Professor and Chair

Department of Computer Science
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VA 22903-2442

Copy No._____

This report summarizes work on NASA Grant NAG-1-1529 for the period January 1, 1994 to June 30 , 1994.

During this period, the grant supported three PhD students in Computer Science: Carmen Pancerella was supported through May, at which time she completed her PhD; she has accepted a position with Sandia-Livermore Laboratory. Bronis DeSupinski replaced Carmen on the grant and was supported for the month of June. Michael Delong was supported for the entire period.

Short summaries of the research of these three students follow.

Highly Parallel Preconditioners
Michael Delong, PhD Candidate in Computer Science
Advisor: James Ortega, Professor of Computer Science

Many problems in fluid dynamics and other application areas lead to large sparse nonsymmetric systems of linear equations. A model problem, which we have been using in our numerical experiments, is the convection-diffusion equation

$$\nabla^2 u + a u_x + b u_y + c u_z = 0 \tag{1}$$

where a, b and c are functions of the spatial variables x, y and z. When (1) is discretized by finite differences (or finite elements), a very large linear system arises, nonsymmetric because of the first derivative terms. Since the system is nonsymmetric, the conjugate gradient (CG) method cannot be used but several extensions of the CG method for nonsymmetric systems have been developed in recent years. Some of the most promising of these are GMRES, CGS, QMR and BiCGSTAB. All of these methods require preconditioning and the most common preconditioner for serial machines, incomplete LU factorization, suffers from the need to solve large sparse triangular systems. If the unknowns in the discretization of (1) are ordered naturally, these triangular systems have very little parallelism, while if red/black or multicolor orderings are used, the systems may be solved efficiently in parallel but the rate of convergence of the preconditioned conjugate gradient type method is degraded.

We have been investigating another approach: using several iterations of the SOR method as a preconditioner. With the use of multicolor or red/black orderings the SOR steps may be carried out efficiently in parallel. Using

1

GMRES as the CG-type method, and (1) in two-dimensions as a model problem, our experiments to date on a SUN have shown the following:

- The use of the red/black ordering does not degrade the rate of convergence of the SOR preconditioned GMRES method.

- SOR is a much better preconditioner than ILU.

- The rate of convergence depends on the relaxation factor ω but much less sensitively than in SOR by itself. This allows the possibility of choosing a suitable ω in a fairly easy way.

- The overall efficiency of the method depends on the number of SOR steps each GMRES iteration, and 5 steps seems to be a good number.

Mr. Delong is now developing a parallel code for the Intel Paragon at NASA-Langley.

Target-Specific Parallel Reductions
Carmen Pancerella, PhD, Computer Science
Advisor: Paul F. Reynolds, Jr., Associate Professor of Computer Science

Many parallel computations are characterized by a high potential, but low actual, need for synchronization among processes. The cost of supporting the potential need can impair the effectiveness of employing parallelism. We have investigated the use of asynchronous, parallel reduction networks as a means of supporting the potential need (and, therefore, the actual need) with virtually no cost to a parallel computation. We have built a novel asynchronous, parallel reduction network which supports the dissemination of "target-specific" reductions: near-perfect state information.

Parallel reduction networks support binary, associative operations, thus reducing a common set of network inputs to a single value. For example, in a parallel iterative technique such as Jacobi Iteration, the minimum step value among all processes can be computed in a reduction network. The minimum over all processes would be a global minimum. For a given process, the minimum over all of that process' immediate neighbors would be a target-specific minimum. A network that could support the concurrent computation

2

of target-specific minima for all processes would be a traget-specific reduction network. We have built such a network.

Our target-specific network is not ideal because its concurrency comes largely from pipelining. We have explored truly concurrent target-specific networks and found there exists a classical time-space tradeoff. For n processes we can compute cn target-specific values (c values for each of n process) concurrently in $\log(n)$ time using a reduction network of $O(n^2)$ complexity. We can perform the same cn reductions in $O(n \log n)$ time using a network of $O(n)$ complexity. We have established a continuum of time / network-complexity results between these two extremes. Also, we have have investigated the sequential cost (time*space) of performing cn parallel reductions as a way of placing a theoretical cap on the optimal time and space concurrent result that can be achieved. Again, we have established a continuum of theoretical bounds for time-space results. This analysis incorporated recent results in sorting networks. An important consequence of our theoretical explorations is that sequential target-specific reductions can be done in sub-quadratic time and space, suggesting a similar parallel result exists. That parallel result has yet to be identified.

We have conducted performance studies, both on the parallel reduction network we built and on simulations of that network. Our application has been parallel descrete event simulation. We have found in aggressive (Time Warp) simulations that use of the reduction network can lead to one to two order of magnitude reductions in space saving costs. In preliminary studies we have also found that we can reduce the wall clock time of a Time Warp simulation by 50%. Further performance studies are planned.

Simulation of Delta-Cache Protocols
Bronis DeSupinski, Masters/PhD Student in Computer Science
Advisor: Paul F. Reynolds, Jr., Associate Professor of Computer Science

With the growing gap between processor speed and memory speed, the need for effective caches becomes more critical. This need is amplified in parallel systems. Delta-cache protocols are a new approach to concurrent caching, having the speed advantages of snoopy caching and scalability of directory- based caching. Delta-cache protocols are based on isotach timing systems, which can guarantee critical properties such as atomicity and

3

sequential consistency. These properties are important to the proper maintenance of cache consistency. Delta cache protocols have the potential to improve parallel processor performance by an order of magnitude or more, as has been observed in perfromance studies of isotach timing systems.

We are conducting a simulation-based performance study of delta-cache protocols. This study is a direct comparative analysis with conventional caching schemes, which employ locking to guarantee atomicity and consistency. A number of interesting problems had to be addressed, including the nature of the workload model (since all previous studies had assumed locking as a way of guaranteeing atomicity). The simulator is complete and studies of different workloads have just begun. We expect to have a final analysis within the next few months.

DISTRIBUTION LIST

1 - 3   National Aeronautics and Space Administration
Langley Research Center
Hampton, VA 23681-0001

Attention: Dr. J. J. Lambiotte, Jr.
M/S 125

4   NASA/LaRC
Acquisition Division
Hampton, VA 23681-0001

Attention: Mr. N. Price
Grants Officer
M/S 126

5 - 6*   NASA Scientific and Technical Information Facility
P.O. Box 8757
Baltimore/Washington International Airport
Baltimore, MD 21240

7 - 8   H. Earnhardt, Clark Hall

9 - 10   J. M. Ortega

**   SEAS Postaward Administration

11   SEAS Preaward Administration Files

*One unbound copy

**Cover letter

JO#5875:ph