CRPC Application Review

Geoffrey Fox

CRPC-TR91151
February 1991

Center for Research on Parallel Computation
Rice University
P.O. Box 1892
Houston, TX 77251-1892

CRPC - TR91151
SCCS - 106

CRPC Application Review
February 11-14, 1991

Initial Report
25 February 91
G. Fox

Overview

The application research in CRPC can be classified into three broad categories:

I: Development of core parallel algorithms of generic importance.

II: Provision of ongoing evaluation and testing of other CRPC activities; in particular the parallel software (PCN, Fortran D) and computer facilities.

III: Outreach activity developing the parallel computing in a particular scientific field with either computational kernels or complete applications.

Of course, a given activity may have aspects of all three categories but this simple classification is still helpful. As a result of the application review, we have developed a "Grand Challenge Consortium" which will allow us to seek additional funding to increase our outreach (category III) activities.

Brief Description of Activities

R: Rice

R1: Parallel Algorithms for Modeling Flow in Porous Media -- Mary Wheeler

This is an outreach (III) activity to the geoscience community involving University of Texas, Rice and Houston. It is also developing new higher order discreditizations of general interest for Solution for Partial Differential Equations (category I). Cellular automata approaches to this problem are under development at Los Alamos.

Benchmark codes: Multigrid (Gary Doolen)
Domain Decomposition

Later UTCHEM (15,000 lines) and UTCOMP (30,000 lines) may be benchmark codes.

R2: Seismic Inversion and Parallel Computation -- Symes

This is also outreach (III) to geoscience. There are two projects involving hypercube and CM-2 respectively. There is a collaboration with Dennis on optimization for hypercube. The CM-2 could be a benchmark.

R3: Airline Crew Scheduling -- Bixby

This is outreach (III) to integer programming community. A public domain problem of crew scheduling from American Airlines has been looked at on the Cray with impressive results. As linear algebra negligible, this is easier than some such problems.

The problem could be part of a benchmark but not the implementation.

R4: Pilot Plant Design -- Dennis

This is outreach (III) to geoscience community and Dennis is working with major vendor suppling code for this plant optimization area.

A 5,000 line nonlinear programming code could be part of benchmark and is part of base optimization (category I) effort.

R5: Optimal Well Placement -- Dennis, Lewis

This is outreach (III) to geoscience community. It can also be viewed as a test case for the integration of simulation and optimization. It is a good candidate for use of PCN to control modules containing Fortran simulation. The project has just started and could produce a very interesting benchmark.

R6: Computational Biophysics -- Phillips

This is an outreach (III) to computational biology involving Karplus's code CHARMM. The original plan to use PCN has been shelved and a "quick and dirty" approach to parallelizing computational bottleneck adopted. We will continue talking to Karplus and we will revisit this then.

CHARMM is a standard 10^5 line molecular dynamics code.

R7: Parallel Weather simulation -- Mellor-Crummey

The base code from the Oklahoma "Storm" Science and Technology Center CAPS has been implemented on the CM-2. This can be viewed as both outreach (III) and as testing software for automatic vectorization (II).

The 3,000 line Fortran code ARPS is a possible benchmark.

C: **Caltech**

Here we do not discuss the linear algebra work which we decided was better discussed as part of CRPC's core algorithm research.

C1: Parallel Spectral Methods for CFD -- Fisher

This continues work with Patera at MIT where original parallel code was developed. This is outreach (III) to CFD community and also will stress CCSF facilities with large production runs. (Category II) Not clear if 10,000 line Fortran (plus message passing) code is a useful benchmark; partly because of "ownership" issues.

C2: Vortex Flow -- Harrar, Taylor

This project is mainly aimed (category II) at testing PCN for this class of finite difference code which has 800 lines of PCN, 4,000 lines of Fortran.

C3: Barnes-Hut Approach to Astrophysics -- Salmon

This can be viewed as outreach (III) to astrophysics community and as providing a challenge (II) to software part of CRPC to design appropriate parallelization languages and tools.

C4: Clustering Model for Vortex Method -- Leonard

This can be viewed again as either outreach (III) to CFD community or as in C3, a challenge (II) to software community.

C3, C4 can also be viewed as developing the new (Greengard) clustering approach to particle dynamics which is a generic algorithm (I). Ding in chemistry at Caltech has followed up on this work by applying to chemical molecular dynamics (type of problem addressed by CHARMM).

C4: Cellular Automata -- B. Sturtevart

This is again outreach (III) to CFD with an easily parallelizable cellular automata approach to CFD. The code is notable for only involving integer arithmetic in the inner loops.

C5: Plasma Physics and Particle in the Cell (PIC) Methods -- P. Liewer (Caltech), Brackbill (Los Alamos)

This is a collaboration with Jerry Brackbill at Los Alamos and Dawson's group at UCLA. (Category III) There are four codes; 1D Electromagnetic PIC; 2D electrostatic PIC; 2D electromagnetic PIC; 2D FLIP-MHD 50,000 line code fro Los Alamos. The last uses an adaptive grid. There are several benchmark codes here and they may provide good Fortran 90D tests. PIC also is challenging to software (II) as two distinct decompositions.

L: **Los Alamos**

L1: QCD -- Gupta

This is outreach (III) to the computational QCD community which is major university use of Cray time at NSF and DoE centers. This project now runs on CM-2 at 6 gigaflops and is a good candidate for benchmark set. It also has been the first serious application to run on both original Caltech hypercubes and on CM-2 and so has historically helped develop these systems (II).

L2: Molecular Dynamics -- Holian

This is outreach (III) to computational physics at Los Alamos. This study of metals with local interactions could be a benchmark.

L3: Global Ocean Modeling -- Smith

This is outreach (III) to climate field with collaborations with Malone (Los Alamos), Semtner (Naval postgraduate school), and Chervin (NCAR). The 2,000 line CMFortran code can be a benchmark.

L4: High Speed Multimedia Flow -- Smith

This is outreach (III) to computational Physics at Los Alamos. The large 24,000 line code Pagosa is not available for distribution.

L5: Adaptive Refinement -- Saltzman

This is developing a SIMD adaptive refinement algorithm of general applicability (I). This project is just starting.

L6: Multigrid on CM-2 -- Dendy

This is developing a SIMD multigrid algorithm which would of general applicability (I). The project revealed flaws in CMFortran compiler which is of interest to CRPC software (II). The projects L5, L6 by investigating irregular adaptive problems on SIMD architectures will help our understanding of FortranD for expressing irregular loosely synchronous problems.

S: Syracuse

S1: Climate Modelling -- Fox, Mills

This is a pilot outreach program to the climate community involving a collaboration with TRW and the UCLA atmospheric science group headed by Michael Ghil. We have compared various languages including Fortran77, Fortran90, C and C*. This program helped design FortranD and falls in categories II and III.

S2: Clustering Models -- Apostolakis, Coddington, Edelsohn, Marinari

We are particularly interested in applications that are "difficult" to express and parallelize in current approaches. The irregular "loosely-synchronous" class is particularly interesting especially for extending FortranD. The work of Leonard and Salmon in parallel hierarchical clustering approach to vortex and particle dynamics is in this class. We are continuing this at Syracuse for Monte Carlo applications to statistical physical. All these problems are hard to parallelize in a high level language and hard to implement on SIMD machines. We are investigating both issues - expressing these algorithms for SIMD machines is closely related to issue of expressing in a high level language (category II).

S3: Fortran D Benchmarks -- Wu

The initial benchmark set for FortranD is

a) 3 versions of Gaussian Elimination
b) FFT
c) N-body
and we are adding
d) Sparse matrix from Linear Programming
e) Irregular finite element
f) Quicksort
g) Barnes Hut algorithm (C3)
h) Cluster algorithm (S2)

S4: Linear Programming -- Li, Wu

The distributed memory simplex method is being tested on linear programming benchmarks from SIAM and Rice.

S5: Financial Modeling - Mills, Vinson

This is an outreach (III) to the management school at Syracuse and involves parallelizing in CMFORTRAN an original IBM3090 code used to model stock option pricing. They are looking at new algorithms (including dynamical systems and neural networks) and will explore other contracts in financial world - both on Wall Street and in academia.

Consortium

On the basis of trip, we generalized the outreach (III) activities with the following draft program. Syracuse and Rice submitted a proposal like this in the area of climate modelling to DoE. Caltech and Rice would be very interested in a chemistry outreach program like this.

CRPC Grand Challenge Consortium

Purpose:

The Center for Research in Parallel Computation (CRPC) is an NSF funded Science and Technology Center centered on developing and using parallel computing software algorithms and hardware. The Grand Challenge Consortium is an outreach program to integrate parallel computing into specific application areas. It is a collaboration between CRPC and researchers within an application area. Funding is separate from the NSF STC but leverages this base funding. Consortium members include industry, government and university projects.

Program:

This will be tailored for each consortium member and would typically be a three year project. Components of the program are

1) Annual workshop bringing together appropriate computer science and application scientists. This would initially isolate key algorithm and software issues and monitor progress.

2) Parallelization of selected kernels and complete applications.

3) Application visitors at CRPC sites. CRPC visitors to consortium member sites.

4) Training (short courses) and access to parallel computing facilities.

5) CRPC will provide software engineering expertise for parallelization; parallel algorithm expertise based on its basic applied mathematics research and experience from other applications.

Organization:

CRPC is a distributed center. For each consortium member, CRPC will designate a lead researcher and one or more sites to be actively involved. The make-up of personnel will vary but is typically:

1 Senior researcher and 1 graduate student in application area
1 Senior and 1 graduate student in parallel software
1 graduate student in algorithms

The consortium membership budget would also fund the workshops, facilities, expenses, and management. The annual funds needed vary between $250,000 and $400,000.

Project Summary

On the basis of trip, we outlined a few key questions recorded below. Each question was to be answered in one or two sentences by each application.

1) Scientific or other goal
   e.g. "find proton mass"
   "test PCN"
   "develop adaptive mesh"
2) Parallelization or other issues
   e.g. "won't run on SIMD"
   "makes Fortran 90 compiler crash"
3) CRPC collaborators within or outside institution
4) Use of CRPC facilities
5) CRPC funding; past, present or none
6) Any outreach/contacts/technology transfer to other groups in field outside CRPC
   e.g. "CRPC parallel application helped persuade oil company use CM-2"

R1: Flow in Porous Media -- Wheeler

Scientific or other goal:

Parallelization of reservoir simulation; higher order discreditization algorithms; grand challenges in geoscience; benchmark set.

Parallelization or other issues:

Presently in pde's we don't know what are the best type of machines for flow in porous media problems.

CRPC collaborators within or outside institution:

We have collaborated with T. Dupont at University of Chicago and Joe Dendy at Los Alamos.

Use of CRPC facilities:

We have used the INTEL machine at Rice and the Oak Ridge Intel RX. We plan to use the Delta machine as soon as available.

CRPC funding; past, present or none:

My funding the first year of CRPC was 2 graduate students; I believe budget was about 30K. Beginning Feb. 1990, my NSF budget was 130K. We hired Weiser halftime and then he became full-time on the project. Other personnel included 1 graduate student and a postdoc. The Texas money which is $200K per year, began in November.

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

R. Glowinski at University of Houston has also been another collaborator as well as the UT (University of Texas at Austin) petroleum engineers (list on slides). With respect to Outreach we will provide info to the UT Affiliates Program--I gave Ken a brochure. They have over 45 companies. Also there is the Workshop in March that we are having. Danny can provide names of the attendees, if needed, which include ten or so companies.

R2: Seismic Inversion - Symes

Scientific or other goal:

My research has as its goal the provision of tools to determine the feasibility and value of seismic inversion, as a tool for exploring the earth's subsurface. The construction of such tools requires the definition and resolution of many mathematical questions, and most likely the use of massively parallel computing equipment offering performance beyond that available.

Parallelization or other issues:

The core of my codes are simulators of seismic wave propagation and related modules; these offer numerous levels of parallelism.

CRPC collaborators within or outside institution:

I am currently collaborating with Virginia Torczon and Preston Briggs in work on the Intel iPSC machines.

Use of CRPC facilities:

I have made extensive use of the CM2s at LANL and Rice, and continue to do so. I plan to use the iPSC/RX at Rice and the Delta at CIT in the near future.

CRPC funding; past, present or none:

I receive no funding from CRPC. My work is partly supported by the Geophysical Parallel Computation Project (State of Texas).

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

I am circulating a consortium proposal to the oil industry.

R3: Airline Crew Scheduling - Bixby

Scientific or other Goal:

Crew-scheduling models, and other related scheduling models, are really integer programs, not linear programs. However, to solve them we need to be able to quickly solve a large number of associated linear programming problems. That's the goal of this work, to solve these associated LPs very fast. Currently the sequential implementation takes 750 Cray YMP seconds to solve an 837 x 12.75 million instance. That's impressive, but not impressive enough if we have to solve 1000 of them.

Parallelization or other issues:

SIMD is probably out of the question. Depending on the algorithm distributed for shared memory MIMD is appropriate. Parallel reads are important (and not so straightforward on a Cray), unless a machine is available with 10^9 bytes of memory. Use results to direct further work in automatic parallelization.

CRPC Collaborators within or outside institution:

Irv Lustig, Ann Kilgore, Virginia Torczson.

Use of CRPC facilities:

None at the moment. Cray research has provided access to a YMP. Eventually, we plan to use a LANL machine for Cray work. When we start distributed memory work, then we will be dependent upon CRPC facilities.

CRPC Funding; past, present or more:

Irv Lustig and Anne Kilgore are funded.

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

Essentially no one who works in this area uses parallelism. The main benefit will be to change that, both in industry and in the research community. As concerns concrete technology transfer, and as noted above, the airlines really do use these models to do their scheduling, at present using less than optimal solutions. If optimal solutions can be found that are 10 to 20 percent better than current solutions, then they will buy machines.

R4: Pilot Plant Design - Dennis

Scientific or other goal:

To use parallel computation to make feasible real-time identification and control of nonlinear processes.

Parallelization or other issues:

Current algorithms are expensive and unreliable. Our approach seems more efficient and more reliable on the Sun4, the Sequent Symmetry, and the Cray. An iPSC/860 implementation is underway. We believe it is possible to automatically recast problems of this type to suit the particular parallel architecture.

CRPC Collaborators within or outside institution:

Collaborators: Phuong Vu of Cray Research
Automatic Differentiation Group at ANL
Parascope Group

Use of CRPC facilities:

Couldn't solve "real" problems without access to CRPC's powerful machines (i.e. Sequent, Cray, Hypercube). Besides, industrial users of this technology are very interested in using parallel machines to solve larger problems practically and to solve current problems faster.

CRPC Funding:

Past and present CRPC funding. We have 1.5 postdocs.

Any outreach/contacts/technology transfer to other groups

Dynamic Matrix Control Corporation is advising us on properties of industrial problems. Dr. Phuong Vu of Cray Research ported our code to the Cray. Professor Biegler of the CMU Chemical Engineering Department has provided us with test problems. The automatic differentiation group at ANL is cooperating with us to interface their system with our code. Karen Williamson and our code are serving as a guinea pig for the Parascope project.

R5: Optimal Well Placement and Management - Dennis

This is part of the Texas Geophysical Parallel Computation Project.

Scientific or other goal:

The scientific goal of this project is to develop the mathematical tools needed for the determination of optimal strategies for the development and management of enhanced oil recovery processes for oil reservoirs.

Parallelization or other issues:

The central issue in our research is how best to take advantage of parallelism in the integration of optimization algorithms with the solution of p. d. e.'s. In this case, p. d. e.'s are involved in the simulation of flow in porous media phenomena.

CRPC collaborators within or outside institution:

John E. Dennis, Jr.
Robert Michael Lewis
Phuong Vu

Use of CRPC facilities:

Our target machine is the iPSC i860 hypercube. Our initial implementation will be carried out on the machine at Rice.

CRPC funding; past, present or none:

The money for the Texas project was received in November 1990. Our project began in January, 1991; the Texas project is currently funded for four years.

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

At present, we do not have any connections to groups outside of Rice.

R6: Computational Biophysics - Phillips

Scientific Goal:

To carry out extended simulations of molecular dynamics in protein molecules and crystals.

Parallelization issues:

Many body problem, with additional conjugate gradient and FFT algorithms. Existing packages have 100,000+ lines of Fortran code.

CRPC facilities:

As part of the Keck Center for Computational Biology, we help fund the local CM-2 and Intel machines, and make regular use of each. We plan to use other CRPC machines after codes are developed locally.

CRPC Collaborators:

We are working with John Mellor-Crummey to some degree, and occasionally interact with others.

CRPC funding; past, present or none:

None, except use of local machines.

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

Codes adapted to Intel machine are being distributed to other Intel users.

R7: Parallel Weather Simulation - Mellor-Crummey

Scientific Goal:

Analysis and Prediction of Severe Storms

Parallelization issues:

Evaluate suitability of the coding strategy employed in the development of the ARPS weather modeling code (written in Fortran 77) for automatic parallelization targeted at both MIMD and SIMD architectures. Suggest adjustments to coding strategy to make automatic parallelization more effective.

Evaluate the effectiveness of automatic parallelization tools under development in the CRPC for automatic parallelization of this code. Use results to direct further work in automatic parallelization.

CRPC facilities:

An early version of the weather code is running on the 8K processor CM-2 installed at Rice.

CRPC Collaborators:

Paul Havlak (graduate student), Joel Castllanos (staff programmer)

CRPC funding:

No separate funding

Outreach/Contacts:

The ARPS weather prediction code is currently under development at the Center for Analysis and Prediction of Storms (CAPS -- an NSF STC) at the University of Oklahoma. The key contact at CAPS is

Kelvin Droegmeier
Deputy Director for Research
Center for Analysis and Prediction of Storms
University of Oklahoma
401 East Boyd
Norma, OK 73019-0515
(405) 325-6561

C2: Parallel Computation of Wavy-Vortex Flows

Scientific Goal:

Compute transition from Taylor-vortex to wavy-vortex flow in one configuration of the Taylor experiment. Construct appropriate bifurcation diagrams for wavy-vortex flows. Seek period doubling. Show wavy-vortices develop from Hopf bifurcation. Use PCN in large application program.

Parallelization issues:

Use PCN to compose Fortran subroutines.

CRPC facilities:

Use many different parallel computers including the Intel Touchstone Delta machine.

CRPC Collaborators:

Dr. David Harrar II and Prof. Herbert B. Keller
Department of Applied Mathematics
California Institute of Technology

Prof. Stephen Taylor
Department of Computer Science
California Institute of Technology

CRPC funding; past, present or none:

One half Postdoc.

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

None.

C3: Barnes Hut Method for Astrophysics - Salmon

Scientific Goal:

Understand the evolution and formation of galaxies by simulating the growth of mass fluctuations in an expanding universe. In addition: develop new techniques based on clustering and hierarchical data structures to speed such simulations by many orders of magnitude.

Parallelization issues:

There is no efficient implementation of fast, hierarchical N-body techniques on SIMD computers. We have an implementation that run on hypercubes, or any reasonably well interconnected MIMD system (the Intel Delta qualifies as such). Efficiencies are in the neighborhood of 50%-90%, for realistic, interesting problems.

CRPC facilities:

We have done most of our large calculations on the nCUBE and MarkIII machines at Caltech.

CRPC Collaborators:

Mike Warren and Wojciech Zurek are at Los Alamos, in T-6 division. I'm not sure if they are formally affiliated with CRPC. Geoffrey Fox was my thesis advisor.

CRPC funding; past, present, or none:

Currently, I am under an NSF fellowship, but I think some of my overhead costs are being covered by CRPC.

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

No technology transfer yet.

C4: Computational Fluid Mechanics by Direct Simulation of Molecular Motions - Sturtevant

Scientific Goal:

Conduct numerical experiments to determine the validity and performance of particle methods ideally adapted for massively parallel computation.

Parallelization issues:

Methods have been applied on all parallel computers at Caltech Computer Science from the original Cosmic Cube to the Symult 2010 and Intel Gamma, and preparations are being made to continue the work on the Intel Delta.

CRPC facilities:

See "Parallelization issues" above.

CRPC Collaborators:

C. Seitz, CIT
G. Doolen, LANL
J. Moss, NASA Langely

CRPC funding; past, present or none:

1989	$38,000
1990	$40,000

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

We have exchanged our programs with NASA Langely. They have visited CIT to use the Computer Science parallel machines and measure performance of their codes based on our techniques.

S1: Climate Modeling - Mills

Scientific Goal:

Develop an advanced climate model (coupled atmosphere/ ocean/biosphere model) using highly parallel computers.

Parallelization issues:

Exploit existing parallel software/hardware systems and evaluate both performance and implementation effort; examine language, communication, load balancing, SIMD vs MIMD compatibility, and algorithm development issues.

CRPC facilities:

Intel delta touchstone prototype
128 node Intel hypercube
64K CM-2
Butterfly TC2000

CRPC Collaborators:

None at present. Argonne group works in this area.

CRPC funding; past, present or none:

None.

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

UCLA Atmospheric Science Department

S2: Clustering Models - Coddington

Scientific Goal:

a) Study the physics of spin models using new non-local Monte Carlo algorithms (cluster algorithms).
b) Investigate these algorithms, and try to generalize them to more complicated models such as spin glasses and lattice guage theory.
c) Find efficient parallel algorithms for labeling the clusters.

Parallelization issues:

Main computational requirement is connected component labeling of large irregular regions, which is difficult to parallelize efficiently, especially on fine grained SIMD machines. We have invented a number of parallel algorithms for both MIMD and SIMD machines.

CRPC facilities:

Workstations, Encores and Connection Machine at Syracuse
Symult, nCUBE-1 and DELTA at Caltech
nCUBE-1, nCUBE-2 and Connection Machine at Sandia
Connection Machine at Rice
Butterfly at Michigan State

CRPC Collaborators:

Ofer Biham, Alan Middleton, John Apostolakis, Allen Su and Leping Han from Syracuse.

CRPC funding; past, present or none:

CRPC funding at Caltech and Syracuse for one postdoc and one student.

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

Collaborating with Clive Baillie, in the physics department at the University of Colorado, Boulder; Enzo Marinari from Rome University Connected component labeling has other possible applications, e.g. in image analysis.

S4: Linear Programming - Wu

Scientific Goal:

Efficiently solve linear programming problems on parallel supercomputers.

Parallelization issues:

Parallelize simplex algorithms on SIMD and MIMD parallel computers. Design and implement sparse matrix algorithms for large applications.

CRPC facilities:

CM-2.

CRPC Collaborators:

R. Bixby, G. Li, Rice University

CRPC funding; past, present or none:

One Postdoc and one student.

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

None at present.

S5: Financial Modeling - Mills, Vinson

Scientific Goal:

Understand potential role of high performance computing in financial community. Parallelize example codes, look at new algorithms and organize seminars.

Parallelization issues:

Current codes seem suitable for SIMD and may also run "embarrassingly parallel" on a network of workstations.

CRPC facilities:

None at present. Uses NPAC's 32K CM-2.

CRPC Collaborators:

Finucane and Diz; two faculty from Syracuse Management School; Mills from NPAC; Vinson from Physics Department.

CRPC funding; past, present or none:

None at present.

Any outreach/contacts/technology transfer to other groups in field outside CRPC:

Zenios from Wharton School.