Notes on the Technical Sessions of the Fourth HPCD Workshop

July 17-18, 1997
Rutgers University
New Brunswick, NJ

Brian D. Davison
Keith M. Miyake

October 13, 1997

1 Welcoming Address – Joseph Seneca, University VP for Academic Affairs

Joseph Seneca opened the 4th HPCD workshop by looking over what has happened over the last four years. 40,000 students have graduated from Rutgers, hundreds of millions of dollars have been spent on research, seven million new jobs have been created in the U.S., and the U.S. has become more competitive – a lean industrial nation.

He described some history of government funded research, citing the 18th century effort to accurately compute longitude (which lead to the invention of the chronometer). He also mentioned the recent discovery (with Rutgers scientists) of a new underwater thermal vent.

Dr. Seneca concluded by saying the University continues to encourage cross-disciplinary collaborative work, such as performed within HPCD.

2 Introduction to the HPCD Project and to the Workshop – Saul Amarel

Dr. Amarel thanked Dr. Seneca and the University for their support of HPCD. He then related the goals of the workshop and the themes of HPCD, and introduced several upcoming talks.

The goals of the workshop are as follows:

- Review status
- Bring together the HPCD community
- Summarize experiences
- Identify spin-offs and outgrowths of the HPCD effort

The HPCD project is a four year, large scale, multi-disciplinary, multi-institutional project supported by DARPA. The broad objective of HPCD is to build on advances in

- High performance computing
- Artificial intelligence
- Modeling/Simulation technology
- Software for design automation

to develop a new generation of engineering automation technology that can bring about dramatic gains in industrial productivity.

The focus of the project is on complex engineering design problems. These problems can be computationally intensive, hard to decompose, and have many (possibly conflicting) design goals and constraints. This makes it hard to reason from goals to possible solutions. Examples of the problems addressed in HPCD include:

- Design of inlets and nozzles for jet engines
- Conceptual design of aerospace vehicles
- Design of innovative surface ships
- Design of photolithographic processes for electronic chip fabrication
- Design/Redesign of microprocessors

The project thrust is to:

- Develop methodologies/systems that enable substantial improvement in the solution of such complex engineering design problems with the help of computers.
- Explore increased automation of processes for solving complex engineering design problems.
- Develop prototype systems and experience in specific realistic domains/tasks.
- Shorten the distance between science and engineering with the help of HPC.

Engineering design problems often need computer power at the level of HPC. However, raw computer power alone will not always be sufficient for handling the computational complexity of the design tasks considered. The project investigates:

- Ways of decomposing domain theories into a multiplicity of models at various levels of abstraction and approximation.
- Ways of using models appropriately in design decisions.

The broad theme for the project is to develop a balanced, integrated approach to the solution of computational design problems by bringing to bear in efficient ways a combination of relevant technologies and bodies of knowledge.

3 Design in Computational Fluid Dynamics – Doyle Knight

Dr. Knight gave an overview of the focus of the CFD group. He summarized some past accomplishments, then highlighted current work.

The focus of this research is the development of a design software system for use with inlets and diffusers for aerospace propulsion, and also for particulate flows. Research objectives include:

- To improve both the efficiency and quality of design through the integration of computational fluid dynamics and artificial intelligence into an automated optimal design methodology.
- To develop and broaden methodologies for automated optimal design which extend to other design domains.

Accomplishments include redesign of the P2 and P8 2-dimensional hypersonic inlets. Optimization was performed using a Navier-Stokes code (GASP or NPARC) controlled by a gradient-based optimizer (CFSQP). The resulting designs gave significant improvements based upon the original NASA design objectives. Work was also done on an axisymmetric supersonic mixed-compression inlet for a proposed replacement for the Tomahawk missile. A two level model utilized a fast empirical code (NIDA, from NAWC) along with a Navier-Stokes code (GASP). Several optimization methods including CFSQP and genetic algorithms were used. A 22% increase in total pressure recovery was obtained over the initial hand-optimized design.

Work is in progress on:

- increasing maximum angle of attack of stable operation for a high speed civil transport inlet through optimization of bleed location.
- minimization of total pressure distortion at outflow for a 3-D (S-shaped) subsonic diffuser.
- optimization of a 2-D supersonic missile inlet to perform a given mission profile (in collaboration with Aerospatiale Missiles).
- simulating production of nanomaterials (particles with grain sizes of 1-100 nm).

4 Conceptual Design of Air Vehicles – Don Smith

Dr. Smith outlined the area of conceptual design of air vehicles, and explained the motivation behind creation of tools to enhance utilization of analysis codes.

Starting with algebraic methods used in industry, a model of a plane flying a multi-phase supersonic mission was created. The resulting simulator was used for optimization studies of specific missions. The simulator was also used to study the problems of using automated design tools with existing analysis codes.

In adapting the simulator for automated execution, several sticking points were encountered. Some models were incomplete or used implicit assumptions, which led to strange 'optimized' designs. These problems could be fixed by addition of constraints on input values, or more explicit problem definition. It was also found that some solution methods are friendlier to search routines (e.g. gradients) than others.

From this work, the design interface has emerged. This builds on the idea that search and analysis tools are separate tasks communicating to each other. It provides a standard communications interface so new code can be easily integrated to work with the interface, and old code does not need to be readapted each time a new tool is used. The interface controls process management, so search tools are shielded from runaway or crashed analysis programs. This also allows the use of coarse grain parallelism for execution of analysis programs.

5 A Genetic Algorithm for Engineering Design Optimization – Haym Hirsh

Dr. Hirsh presented details of the Genetic Algorithm for Design Optimization (GADO), which is the dissertation work of Khaled Rasheed.

GADO is a real-valued genetic algorithm optimization method tailored to engineering design tasks. Engineering domains may contain many unevaluable or infeasible points, and the evaluation functions may be expensive. Optimization is performed subject to constraints in a continuous design space. GADO incorporates specialized crossover functions and modules designed to facilitate search in these domains.

GADO has been used in the conceptual aircraft work as well as in supersonic missile domains and compares favorably to other standard optimization methods (CFSQP, genetic algorithms (Genocop III), simulated annealing).

6 Stochastic Structural Optimization for the Design of Crashworthy Structures – Madara Ogot

The focus of this newly formed project is optimization of structures which may undergo multiple impacts. The resulting design space may be discontinuous and may contain multiple local minima. Finite element method (FEM) analysis is prohibitively expensive, so a multilevel strategy is employed.

Simple rigid body models (Level I) will be used explore the design space and locate promising regions. Promising regions are further explored using FEM models (Level II) with modified stochastic optimization methods (to decrease the number of iterations needed).

Level I has been tested for aircraft turbines, where deflection shields are installed to protect key components in the event of rotor failure. The secondary function of the shields is to avoid reflection of debris back into the engine. Work is in progress on incorporation of a hybrid FEM model (already developed) for level II.

7 Visiometrics and Modeling for HPCD – Norman Zabusky

Dr. Zabusky discussed work on 2 and 3-dimensional feature extraction, quantification, and tracking. He briefly mentioned work on data and control flow visualization for applications programs, and a new initiative in collaboration with Los Alamos National Laboratory on "non-equilibrium science".

Feature extraction is useful to gain new insight on problems, and as a tool to help generate reduced simulation models. This technique can be used on 2 and 3-dimensional data from direct observation or numerical simulation. Individual features are identified and described using moment calculations and approximate object representations (skeletons). Using reduced descriptions decreases the time needed when studying large data sets. Objects may be tracked through time, with changes described using the transformation operators continuation, creation, dissipation, bifurcation and amalgamation.

Feature extraction techniques have been applied in association with the voice mimic project for trachea quantification and with the CFD group to visualize the phenomenon of supersonic inlet unstart.

Future work in this area includes extending feature tracking to unstructured and curvilinear grids, including multi-object motion correlation, and improving descriptive parameters associated with objects, such as calculated moments and skeleton information.

8 Perspective from the Aerospace Industry – Raymond Cosner, McDonnell Douglas Aerospace

Dr. Ray Cosner, a member of the Technical Advisory Committee for HPCD, spoke on the design process from the perspective of those in the aerospace industry. His work has been to establish new processes for vehicle design that would make efficient use of resources, incorporate innovative processes, and reduce design time. To do so, it is necessary to coordinate the participation of multiple organizations, since there are always multiple sites, partners, and suppliers. In this effort, incorporating concurrency and risk management into the process has *been primary.*

Over the life cycle of a design (including conceptual design, preliminary design, detailed design, deployment, and operations), there is a great increase in the accuracy required for each phase. Initial conceptual design may be satisfied with 25% accuracy, but at least 5% may be necessary for preliminary design, and 1% for detailed design. Different tools and analyses are required for each design stage.

Today McDonnell-Douglas has strong tools in single disciplines, but poor integration makes multi-discipline trade studies difficult, which stifles innovation. In the future, we want to exploit communications and concurrency, to utilize modeling and simulation, and to formalize the uncertainty/risk associated with designs. Our immediate goal is to design the infrastructure required to develop the integrated process.

The infrastructure goals are to improve process efficiency, allow effective multi-discipline trade studies, permit real-time collaboration, and manage risk. The infrastructure must link tools and processes. It should be modular, reconfigurable, and connect with legacy tools and across heterogeneous computer systems.

We have implemented such a system, called MAX (Multi-Application Executive), and have used it on a NASA grand challenge problem of designing jets with vertical lift. We have also used it in F15 sensor placement optimization.

Overall, we need to reduce risk to an acceptable level (to know the design will be successful). Therefore, we need predictive risk models for each element, including mean performance and variance, to build a rational process. As we look forward, we will need a formal process for uncertainty management, a process integration infrastructure (including architectures, protocols, standards), and process control methodologies (including statistical methods).

Questions and Comments

Saul Amarel asked what work was in process for uncertainty management. He replied that low-level work was being done internally, and mentioned an outside image processing company which assessed the accuracy associated with a given image. He added that much has not been addressed yet.

When Al Despain asked what was the most difficult aspect of the Virtual Design Team, Ray Cosner responded that the top issues were concerning the budget, getting supervisors to agree, and getting people to disconnect from their current projects. To this, Al prodded further, asking about technical difficulties. Ray answered that they found that the individual's skill level was important, so they endeavored to use familiar code to help ensure productivity.

9 Interactive Visual Environment for Design – Tom Ellman

We demonstrated our Interactive Visual Environment for Conceptual Design on a problem of conceptual aircraft design. For this purpose, a Legacy Code Management system was used to exercise a combined aircraft simulation and optimization code and to construct a database of numerical data generated by each module and submodule in the course of an optimization run. The Data Visualizer examined and compared optimizer trajectories corresponding to different starting points. It also identified trajectories which ended prematurely due to an evaluation function failure. The Context Visualizer and Context Signature Visualizer were then used to determine the cause of the failure.

10 Voice Mimic System – Jim Flanagan et al

10.1 Parametric Representation of Speech Signals – Jim Flanagan

The aim of this research is to understand the process of human speech generation to develop an articulatory representation of speech for synthesis, automatic recognition, and low bit-rate speech coding.

10.2 Speech to Articulatory Shape Computation – Samir Chennoukh

The voice mimic system uses two methods to perform the acoustic-to-articulatory mapping: a codebook of matching pairs of vocal shapes and acoustic parameters, and an analytic description of vocal tract parameters as a function of the first two formant frequencies of the speech signal.

A low-bit rate articulatory based speech coding using the analytic mapping technique has been demonstrated. The bit-rate of the coding is determined from the parameter sampling rate and their digital representation, producing an intelligible output speech signal for bit-rates as low as 624 bits/sec using parameter sampling at 48 times/sec.

Questions and Comments

Al Despain asked whether they used Hidden Markov Models. The response was that they used coefficients and table lookups, and dynamic programming to reduce ambiguities.

10.3 Speech Synthesis from Fluid Flow – Dan Sinder

Speech generation is examined as a fluid dynamic process. We have designed and implemented a vocal chord model using a Reynolds Averaged Navier-Stokes solver to compute flow conditions. Static geometries have been tested, and we are developing elastic models which use flow-driven vocal chord motion.

11 Human Centered Systems: A Washington Perspective – Allen Sears, DARPA/ITO

Dr. Allen Sears, a DARPA/ITO Program Manager who provides partial support for the HPCD project, spoke on his perspective on human centered systems, and how this area has come to be important.

In the 1980's, the primary goal was the automation of tasks. Now, we realize there is a partnership, and we need to reduce the level of human effort. Thus, systems should be designed with the human at the center. Recently there has been a multiple department project (including NSF/CIA/NSA/DARPA) called STIMULATE to work on language technology understanding. The consensus is that HuCS (Human Centered Systems) will become more important, and as a result, multiple modality is always important. We call this area of research *humionics* as a parallel to the term avionics for the study of flight.

12 Summer Institute Projects: Center for Computational Design – Doyle Knight et al

The Center for Computational Design (CCD) was established to foster interdisciplinary research in automated computational design and rapid virtual and physical prototyping, and the application of this technology by industry.

Researchers were invited to participate in the summer institute to work closely with Rutgers faculty and students in areas of computational design. The institute objective is to facilitate development of new research initiatives and research proposals to external funding sources.

12.1 Nonlinear Calculations for Designing with Flexibility – Suresh Ananthasuresh, University of Pennsylvania

The objective is design of components which are flexible and strong, instead of the traditional method of designing rigid parts. Given the forces to be applied, determine the resulting topology, size, and shape. This may include having multiple connected flexible parts.

Questions and Comments

When asked about the order of magnitude of computational requirements, Suresh answered that it took 100-1000 iterations to converge.

Elisha Sacks asked whether use of reduced models may speed computation. Dr. Ananthasuresh said one possible method is to model parts as rigid bodies with springs, but that this was not appropriate for the problems in question.

12.2 Integrated Computer-Aided Mechanical Design – Elisha Sacks, Purdue University

Dr. Sacks presented an overview of a gear mechanism for a disposable camera, and a more general discussion of an environment for contact analysis.

12.3 Design Optimization of Supersonic Air-Breathing Missiles – Yan Kergaravat, Aerospatiale Missiles

The purpose of this project is to create methods which begins with an initial specification and outputs an optimal missile inlet and body design. The design process integrates aerodynamics, propulsion, and electromagnetics.

The industrial goals are to decrease development time and cost, improve interdisciplinary integration, and to obtain better designs. Research goals are to implement multi-disciplinary optimization, and to apply analysis and AI techniques to real design.

Challenges faced include computationally intensive simulations, strong coupling between disciplines, conflicting goals, and multiple parameters.

12.4 Design of Force Fields for Predicting Biomedical Interactions – Jay Banks, Columbia University

Predicting interaction (binding) between proteins and ligands (smaller molecules) is important in the design of new drugs. Binding strength can be computed by use of a quantum mechanics model, but the associated computational expense is overwhelming. Work is in progress on reduced models using classical mechanics, which drastically decreases evaluation time.

13 Multi-Level Design of Microprocessors – Louis Steinberg

Our research objective is the study of design across multiple levels of artifact representation, using a domain which has a well-developed set of levels and a CAD technology for each level. Research focused on bottom-up flow of information such as that caused by a change in VLSI technology. We are concerned with how lower level tradeoff and constraint information can be automatically incorporated in decisions at a higher level.

In processor redesign for technology evolution, we considered the actions required to handle an evolutionary change in VLSI technology (such as moving from a .8 micron to a .6 micron process), which would produce changes in processor designs at all levels. This problem description served as a testbed for connections between varying levels. In attempting this, we encountered a huge infrastructure cost in putting differing systems (design levels) together. Expert knowledge about critical resources, as well as implicit knowledge held by human designers were incorporated into the design system. During this process, a special-purpose placement algorithm was built to design regular circuits (i.e. those with standard cells).

Two general methods for controlling the design process were considered. Rational control of multilevel stochastic design takes a utility-based approach to determine the tradeoff between design time and design quality. A highest utility first search method was implemented and tested on a two-level placement problem. This compared favorably with a traditional waterfall control method. The second method combines economic and genetic algorithms in design. It incorporates a market economy to coordinate and optimize many distributed design processes. Our system was able to find good tradeoffs between design time and circuit time/area.

Questions and Comments

Al Despain asked whether an inexpensive estimator (plus calibration) would be better than just a random waterfall. The reply was that estimators were used in the design process.

14 Computer Architecture in HPCD – Alvin Despain

Over the past year work was done on many aspects of the multi-level design effort, including low level libraries, high level design tools as well as innovative architectures and analysis. Studies of speed and area vs. pipeline depth for 32-bit adders, and performance vs. pipeline depth using Spec benchmarks were performed.

We also have designed a new architecture, called HiDISC (Hierarchical Decoupled Instruction Stream Computer), based upon following principles:

- exploit instruction-level parallelism
- tolerate memory latency without increasing pin traffic
- reduce complexity to keep clock frequency high

- prevent speculative execution or prefetches from interfering with necessary computation
- allow compiler to pass information to the hardware

HiDISC uses separate instruction streams for different levels of the memory hierarchy (compute, access, and cache management). Queued communication channels provide synchronization and slip.

Benchmark comparisons done using HiDISC architecture show a speed-up of 2-3, with similar reductions in pin traffic.

Questions and Comments

Saul Levy asked whether data dependencies (in addition to branch dependencies) were modeled, to which Al responded yes. When Elisha Sacks asked what the upper bound on performance improvements with a particular number of stages, the answer was more than double (such as improving branch prediction).

Saul Amarel noted the design space of 10^8 *and asked what the dimensions were. Al responded with a list, including the number of pipelines, the size of cache, cache fault times, clock skew, and probability of cache hits. When asked by Brian Davison, Al confirmed that speedups were for simulated real-time, not MIPS.*

Josh Hall asked whether you get overlap because of more instructions for multiple processors? Al answered that yes, but we still get performance improvements. And not all instructions are executed.

15 Design of Microlithographic Processes – Steve Orszag et al

15.1 Steve Orszag

Dr. Orszag gave an overview of present work and accomplishments in the area of microlithography.

FAIM (Fast Aerial Image Model) models semi-coherent light passsing through a mask and computes the intensity of the projected image on a substrate surface. It uses algorithms specially designed to reduce complexity while more accurately computing the Fourier integral. FAIM is currently used by leading IC manufacturers for making technology decisions. An enhancement called Net-Faim uses FAIM's domain decomposition to allow distributed computation across a network of workstations.

PROCPHASE calculates photoresist deposition based on surface exposure and post-exposure processing. Detailed chemical models are used to describe the exposure, baking, and dissolution steps.

OPTIMASK produces aerial mask designs which can be used to create a desired printed image. Optimizations may be configured by adjusting feature size, illumination models, and other mask properties (such as serifs). It uses FAIM to analyze prospective masks, and iterates to correct problem areas.

Questions and Comments

Al Despain asked why domain decomposition didn't try to avoid splitting features. The response was that decomposition was done automatically.

Bob Lucas asked that since you get super-linear speedup, don't you want to do the domain decomposition at all levels? Steve responded that domain decomposition is always done — even on a single processor.

15.2 Gerry Richter

Dr. Richter presented work on extending the mask exposure calculation in PROCPHASE to use a 3-D simulation.

A 3-D solution requires solving the Helmholtz equations with respect to the photoactive compound (PAC) concentration. The problem may be discretized using the finite difference method (uniform grid), or the finite element method, (non-uniform grid to achieve arbitrarily high accuracy).

Standard solution methods such as Gaussian elimination or nested dissection can be used, but the work required for solution is excessive ($O(n^7)$ and $O(n^6)$ respectively). Iterative methods (e.g. SOR, conjugate gradient, multigrid) do not work on highly indefinite problems such as this.

A transient approach is proposed, where the problem is solved by converging a damped wave equation. This results in an $O(n^4)$ solution for the finite difference method. A new explicit solution is also being studied for the finite element method.

Applications for such an extended simulation include increasing solution accuracy in problem areas, and improving rules for mask redesign.

Questions and Comments

Paul Kantor noted that the results Gerry showed demonstrated some periodicity, but that the methods didn't seem to take advantage of them. Gerry agreed, saying that we don't know how to take advantage of them.

15.3 Don Smith

Dr. Smith spoke on automated VLSI cell redesign using the Circuit Design Associate (CDA).

The circuit design associate redesigns a prototype printed image using localized transformations to ensure manufacturability and to try and meet target electrical specifications.

The CDA consists of two cooperating levels; the higher level Electrical Design Associate (EDA) sets and manages goals based on electrical timing, and the lower level Geometric Design Associate (GDA) manages geometries using knowledge of the fabrication process. The CDA is written in Java and interfaces with existing tools (FAIM/OPTIMASK). It supports several layout representations and is fully networkable.

The CDA has been tested with good result for determining speedup in specific cells obtained by shrinking feature sizes (such as when fabrication technology is improved).

16 Scheduling and Data Mapping Systems – Apostolos Gerasoulis, Tao Yang

16.1 Apostolos Gerasoulis

Dr. Gerasoulis reported on the current status of D-PYRROS and Pyrros+.

Pyrros+ uses the PYRROS scheduling system to perform automatic parallelization. It takes a sequential program with task annotations, and produces a parallel program using the PYRROS scheduler. It predicts parallel performance using program information, processor and network speed estimates. It is the first task-based automatic parallelization and scheduling system.

D-PYRROS uses run-time program computation and communication estimation to perform dynamic scheduling. It uses local or global process clustering to reduce communication overhead, and reschedules based on run-time multiprocessor performance degradation.

16.2 Tao Yang

Dr. Yang presented HPCD spinoff research on RAPID and SWEB.

The SWEB project is designed to improve the scalability of WWW servers (such as the Alexandria Digital Library). Server work and network traffic can be reduced by using the client to process information based on its configuration. A smart client can send information about its CPU power and network latency/bandwidth along with its request. This allows tasks to be broken into separate parts for the client and server.

RAPID is a software tool for parallelizing code which works on sparse/irregular matrices. It uses C functions to specify shared objects and tasks which access these objects. RAPID does runtime parallelization, performing automatic dependence analysis and using PYRROS for automatic scheduling. This results in a space/time efficient scheme for task execution and communication.

RAPID has been applied to various sparse solution methods, and has produced the best performance ever achieved for sparse matrix computations by (semi-)automatically parallelized code.

Questions and Comments

Saul Amarel asked, "What are the hard problems in this area?" Apostolos's answer was that automatic scheduling gets you only a certain performance, and rhetorically asked how do you take advantage of cache and local memory, saying that it was an extremely difficult problem (with a hierarchy of caches). Amarel also asked whether it would be possible to combine these methods with Al Despain's work, to which Apostolos replied that he didn't think you could do it completely automatically. Saul concluded the discussion with the comment that these two activities should try to work together.

17 Impact of Errors on a Quantum Computer Architecture – Alvin Despain

Dr. Despain presented an overview of his quantum computer research (a spinoff of HPCD).

A quantum computer exploits the properties of quantum mechanics. Each quantum bit is a superposition of 0 and 1, and is forced into a single state upon observation. The advantage of such an implementation is based on a hypothesis by Richard Feynman that a quantum computer could provide exponential storage and calculation in non-exponential space and time.

Basic logical operations have been devised for the theoretical computer and adapted for proposed implementation methods. Quantum algorithms calculate all values of a function simultaneously, but only a single value can be observed.

The effects of systematic errors which may occur when using a specific implementation called a Trapped Ion Quantum Computer are presented.

Questions and Comments

Saul Amarel asked Al to compare the state of quantum computers to computing in Babbage's era (1800's). Al agreed, saying we are likely to be at the Babbage stage. He added that he estimates the probability of realization at 1-3%, but that a successful implementation would bring about a huge change.