DOE/ER/3072-58
February 12, 1990

Proposal to the SSC Laboratory for Research and Development of a Parallel Computing Farm

J.G. Heinrich, D.R Marlow and K.T. McDonald*
Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544

L.D. Gladney, N.S. Lockyer, and K.J. Ragan
University of Pennsylvania, Philadelphia, PA 19104

Abstract

We propose an R&D program to study the suitability of a tightly networked parallel computer farm for triggering and data acquisition for an SSC experiment. We will perform this work with the aid of a 32-node farm of Intel i860 processors that will be purchased shortly (at a cost of $900k) by the Program in Applied and Computational Mathematics at Princeton University. We request **$80k** for dedicated I/O devices that would permit the SSC R&D to proceed in parallel with the main use of the farm by the Computational Mathematics Program.

This document is a Supplement to our proposal of the same title submitted on September 29, 1989.

* Contactperson

In an SSC experiment such as the Bottom Collider Detector where the signature for B-meson decays is not as distinctive as, say, $H^0 \rightarrow Z^0 Z^0$ it may prove advantageous to implement the second-level trigger in on-line programmable processors. This strategy is encouraged by projections that RISC processors costing less than \$1k in the year 2000 will have an excess of 500 MIPS computing power. A farm of 1000-10,000 of such processors could perform trigger decisions on 10^5-10^6 events per second. This rate is typical of many hard-wired trigger processors in use today. A trigger based on a processor farm would have the advantage of being fully programmable, and that the farm could serve as the off-line computer for the experiment as well.

Already many groups are implementing farms of processors as on-line or off-line computing facilities, typically with straightforward networking architectures. But as noted by W. Dally,[1] in very large farm architectures the wiring of the network is the limiting factor in performance. For example, if we desire to process 10^5 events per second of 1 MByte size each, the dataflow rate is 100 GByte/sec. A single high-speed data link will have a bandwidth of only 10-100 MByte/sec, so that at least 1000 internal high-speed data paths will be needed.

Among commercial vendors of processor farms, Intel Scientific Computers appears to have the most advanced planning for implementation of a network architecture capable of controlling such a massive dataflow. Their Touchstone project,[2] funded by DARPA, has already delivered its first product in the form of farms of up to 128 processor nodes in a hypercube architecture,[3] and has an evolutionary path leading to multi-teraFLOP computer farms by the year 2000.

We wish to evaluate this approach to parallel computing for on-line applications in high energy physics experiments. Several issues would be addressed in our initial studies:

- Single-processor studies: Can triggering algorithms based on pattern recognition in various tracking detectors be efficiently implemented in RISC processors such as the Intel i860? This device has an internal pipeline structure that may be advantageous for track-finding algorithms typical in high energy physics. Such studies continue our work reported in Ref. 4.

- Multi-processor studies: Could an event be more effectively analyzed by assigning several processors to it, each taking on a task with code and data sizes well matched to the on-chip cache memory of the processor?

- Network studies: How well can the network
  - control the use of several processor per event?
  - route events to relevant clusters of processors where code resides to treat a particular trigger type?
  - route the output from successful triggers to the storage devices while the input data streams are also present on the network?
  - perform the 'event-building' function of combining the multiple data streams off the detector into separate events?

Such questions should be answered before a major investment is made in a particular computer farm. Yet it is hard to obtain the answers without having access to a significant example of the farm in question. We are fortunate that the Program of Applied and Computational Mathematics at Princeton University has decided to purchase a 32-processor farm (described in the Appendix) from Intel, and that it can be made available to us. An agreeable arrangement (with S.A. Orszag of the Computational Mathematics Program) is that there would be no cost to the high energy physics groups for use of up to 5% of this machine, if we provide our own dedicated I/O devices as needed for our studies. Hence we request a total of **$80k** for:

- A second **System Resource Manager; $18k.** This device is an 80386-based host computer on which the user prepares code to be implemented on the farm, and which supervises the execution of the code. It supports ethernet communication to user workstations. The price includes a one-year maintenance contract.

- Upgrades to the **Ethernet Communications; $5k**, between Jadwin Hall and Fine Hall at Princeton. These include a new ethernet port for our CISCO router, and a new cable. The upgrade will permit access to the Intel farm from the physics building, and provide connection to HEPnet for users at U. Penn and elsewhere.

- An additional **Hypercube I/O node, four 640-MByte hard disks, crate and rack; $57k.** High-speed I/O with the farm itself must be accomplished via special processor nodes that reside on the hypercube network. We will use the disks as the source of simulated input events, and as the simulated output devices. The configuration of the farm to be purchased by the Computational Mathematics Program is such that any expansion requires both a new I/O crate and a new 19″ rack. The I/O processor has the largest cost of these items. The price includes a one-year maintenance contract.

Appendix: The Princeton Computational Mathematics Processor Farm

As an example of the state of negotiations between Intel and the Program of Computational Mathematics, Steven Orszag (609-258-6206) gave us permission to reproduce the following letter to him, dated January 29, 1990. As of today, 2/12/90, the leasing arrangements discussed below have been satisfactorily resolved.

Following up on our meeting last Thursday, January 25, we have several answers to questions you asked, and a revised quotation for the system package.

We have results back on the benchmark you discussed with Dr. Michael Barton. This fast Fourier transform ($128 \times 128 \times 128$) was run on the NASA Ames system for several different configurations, with the following results with single precision (32 bits):

Nodes	Time (sec)	MFLOPS
16	3.69	59.7
32	1.89	116.6
64	0.98	223.8

These results reflect what you can realistically expect when using the system because they are conservative (using i860 processors at 32 MHz rather than the 40 MHz you will have) and they use the FFT routine in our Veclib library.

The FFT routine is handcoded rather than compiled with Fortran – the library will be Fortran compiled for March customer shipments – but we will make the routine available to you at the time you take delivery of the system. Eventually all such often-used routines will be handcoded and inserted into the Veclib to replace the Fortran-compiled routines. By the way, handcoding increased the performance of that particular routine by about a factor of 3.

The attached quotation has been revised per our discussion last week. It is now a Model 32, fully configured with standard I/O and a total of 16 MBytes of memory on each of the 32 computational nodes. The upgrade of the computational nodes from the standard 8 MBytes of memory per node to 16 MBytes of memory per node would be scheduled for early July.

This configuration has a list price of $1,117,695, and we are proposing a discount of $279,423 in consideration for Intel cited as a sponsor in all papers and other published

documents pertaining to research on the iPSC/860 system. The price of the system then, is $838,272.

The maintenance contract on this system remains the originally quoted $21,160 annually.

In regards to your concerns about the leasing arrangements, we are making arrangements to propose a leasing company later this week. We understand your funding of $225,000 per year, and we'll try to work within these bounds when negotiating with the leasing company. Without knowing the exact terms required by the leasing company, we cannot tell you whether your funding can support a system of this size and price. We plan to work with you and the chosen leasing company to resolve this within the next three weeks.

We urge you to proceed with the letter of intent for this purchase as soon as possible. With this in hand, we can proceed with negotiations on the lease, and if those can be completed by the end of February, there is a strong possibility your system could be installed by mid-April. Otherwise we will probably be forced into early July for installing the system.

We realize we have not yet provided the guarantee you have requested regarding consulting time for one of our applications programmers to help optimize your code. Before putting this into print we need to review the issue again with the manager of application programmers. We'll contact you in the next few days to conduct this brief meeting over the phone. We do not expect any problems in making such a guarantee, we simply need to make sure it is realistic and both parties are clear on goals.

If you have any further questions, please do not hesitate to contact either Gary MacDonald in Boston at (508) 747-5773 or myself here in Oregon (503) 629-7720.

Regards,

David Billstrom
Manager, Science and Research
Intel Scientific Computers
15201 N.W. Greenbrier Parkway
Beaverton, Oregon 97006

References

1. W.J. Dally, *A VLSI Architecture for Concurrent Data Structures* (Kluwer, Hingham, MA, 1987); also, *Performance Analysis of k-ary n-cube Interconnect Networks*, to appear in IEEE Trans. on Computers.
2. J. Rattner, *Intel-DARPA Touchstone Project*, in *Copies of transparencies from presentations at the Workshop on B Physics in p-p Collisions at the SSC* (DeSoto, TX, 6-8 June, 1989), p. 409; see also, Federal Computer Week (April 10, 1989).
3. See, for example, Digital Review (January 15, 1990).
4. L.D. Gladney *et al.*, *Initial Experience with the Intel i860 Microprocessor*, U. Penn preprint UPR-0184E (Jan. 13, 1990).