

# ornl

ORNL/TM-13180

RECEIVED

APR 15 1996

OSTI

**OAK RIDGE  
NATIONAL  
LABORATORY**

LOCKHEED MARTIN 

**CRADA ORNL 91-0046B Final Report:  
Assessment of IBM Advanced  
Computing Architectures**

G. A. Geist

MANAGED AND OPERATED BY  
LOCKHEED MARTIN ENERGY RESEARCH CORPORATION  
FOR THE UNITED STATES  
DEPARTMENT OF ENERGY

ORNL-27 (3-96)

**MASTER**

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED *cat*

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831; prices available from (615) 576-8401, FTS 626-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Computer Science and Mathematics Division

Mathematical Sciences Section

**CRADA ORNL 91-0046B FINAL REPORT: ASSESSMENT OF IBM  
ADVANCED COMPUTING ARCHITECTURES**

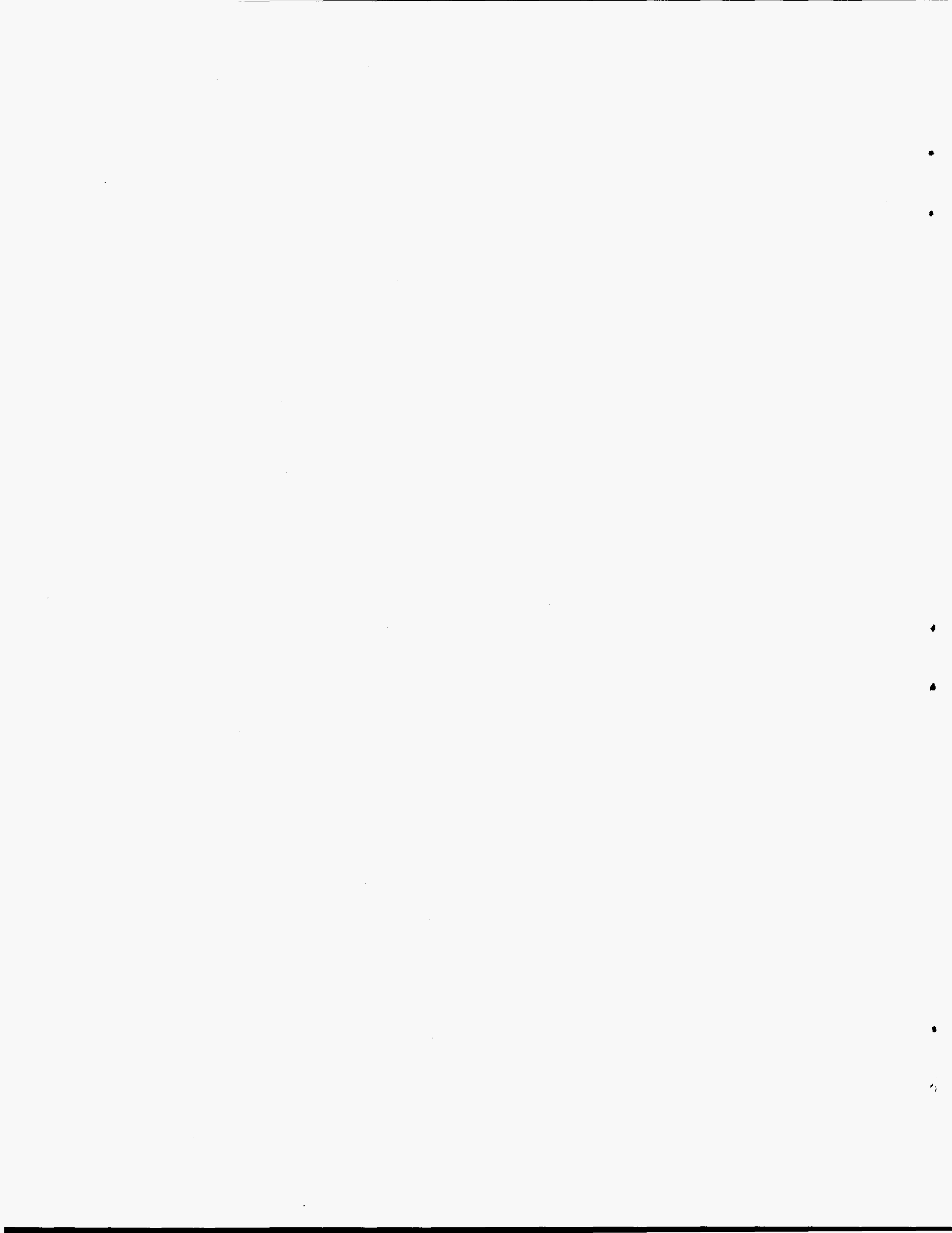
G. A. Geist

Mathematical Sciences Section  
Oak Ridge National Laboratory  
P.O. Box 2008, Bldg. 6012  
Oak Ridge, TN 37831-6367

Date Published: February 1996

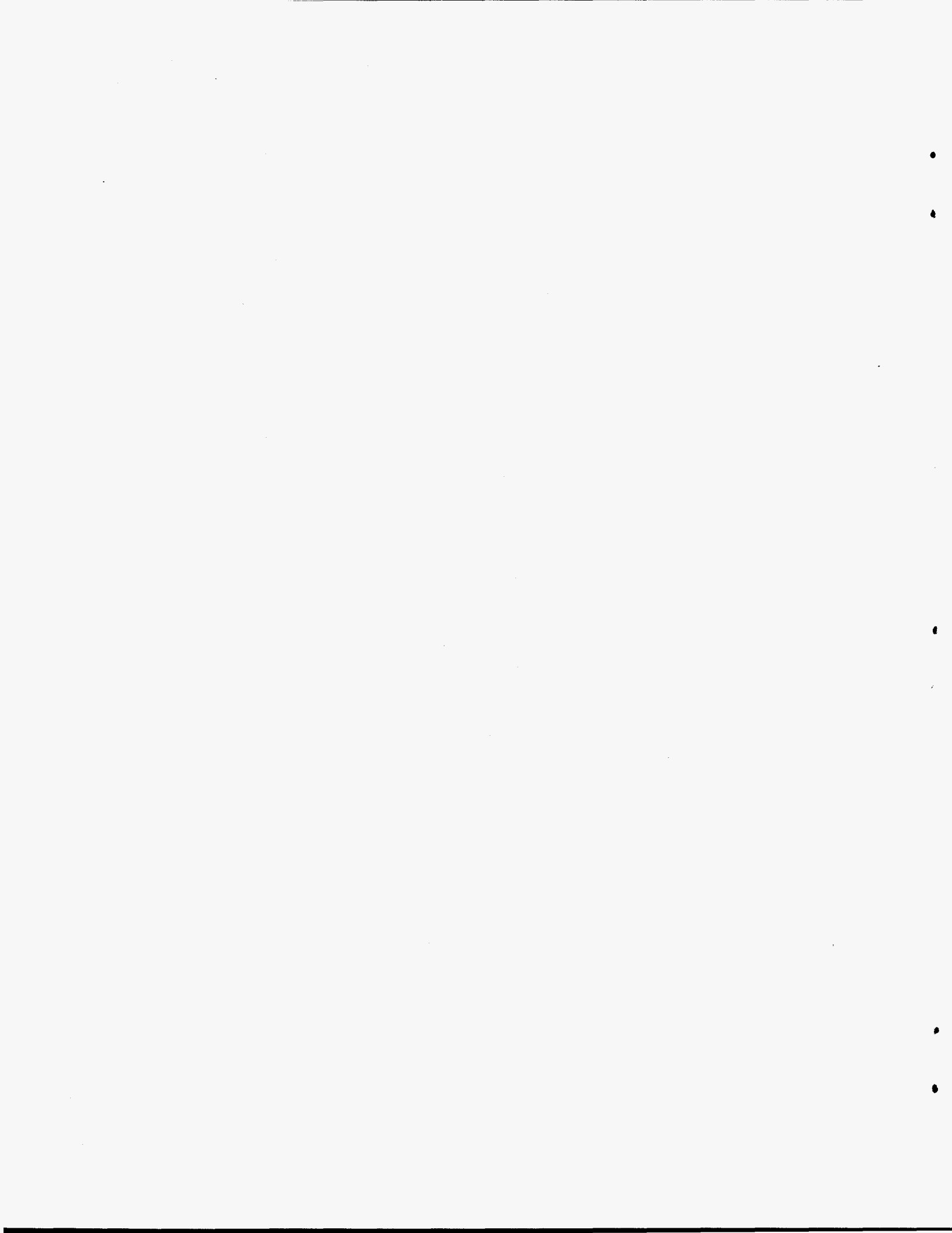
Research was supported by the Applied Mathematical Sciences Research Program of the Office of Energy Research, U.S. Department of Energy.

Prepared by the  
Oak Ridge National Laboratory  
Oak Ridge, Tennessee 37831  
managed by  
Lockheed Martin Energy Research Corp.  
for the  
U.S. DEPARTMENT OF ENERGY  
under Contract No. DE-AC05-96OR22464



## Contents

1	Objectives of the CRADA . . . . .	1
2	Task Summary . . . . .	1
	2.1 Lab Accomplishments . . . . .	1
	2.2 Partner Accomplishments . . . . .	2
3	Benefit to DOE . . . . .	2
4	Technical discussion . . . . .	3
	4.1 POWER/4 . . . . .	3
	4.2 POWER/2 . . . . .	5
5	Inventions . . . . .	7
6	Commercialization . . . . .	7
7	Future collaboration . . . . .	7
8	Conclusions . . . . .	8
9	References . . . . .	9



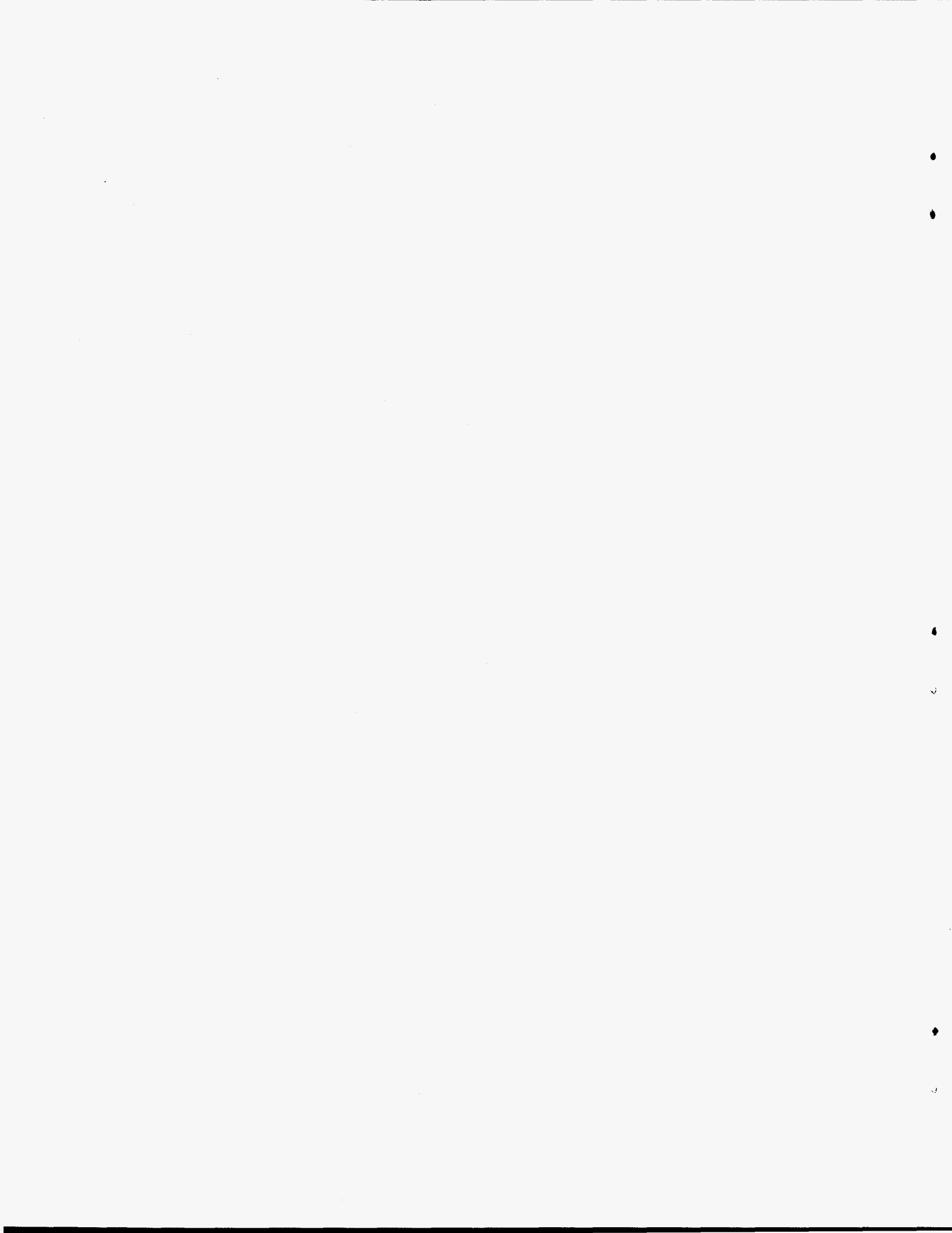
**CRADA ORNL 91-0046B FINAL REPORT: ASSESSMENT OF IBM  
ADVANCED COMPUTING ARCHITECTURES**

G. A. Geist

**Abstract**

This report presents a summary of the work performed under ER-LTT project number: ORNL90-0046b. This was a Cooperative Research and Development Agreement (CRADA) with IBM to assess their advanced computer architectures. Over the course of this project three different architectures were evaluated: The POWER/4 RIOS1 based shared memory multiprocessor, The POWER/2 RIOS2 based high performance workstation, and the J30 PowerPC based shared memory multiprocessor. In addition to this hardware several software packages were beta tested for IBM including: ESSL scientific computing library, nv video-conferencing package, Ultimedia multimedia display environment, FORTRAN 90 and C++ compilers, and the AIX 4.1 operating system. Both IBM and ORNL benefited from the research performed in this project and even though access to the POWER/4 computer was delayed several months, all milestones were met.

This technical report serves as the final report required by the CRADA agreement.





## 1. Objectives of the CRADA

This project involved the evaluation of prototype computer architectures developed by IBM, in particular, the POWER/4 and POWER/2 architectures. The goal was to determine the strengths and weaknesses of these architectures and their associated software in order to determine their commercial viability. As part of the project the PVM [3] and LAPACK [2] software, which ORNL helped develop, was tested on the architectures. The associated beta software that we evaluated included: Fortran 90 compiler, C++ compiler, Ultamedia, AIX 4.1, and ESSL.

In addition we supplied documentation, expertise, and early software to assist IBM in porting PVM to the SP-1 architecture as part of our cooperative research.

## 2. Task Summary

### 2.1. Lab Accomplishments

- Task 1. ORNL will evaluate the performance and viability of the POWER/4 architecture and extrapolate to larger systems. The evaluation showed poor performance of architecture.
- Task 2. ORNL will port PVM to the IBM shared memory machine and evaluate the performance of PVM on this system. ORNL will supply advanced specifications and access to the PVM development team to assist IBM's port of PVM to SP-1. The port of PVM to POWER/4 uncovered a fatal flaw in POWER/4 operating system. IBM decided not to make POWER/4 a product, so the flaw was never fixed. We delivered assistance and materials for SP-1 port.
- Task 3. ORNL will evaluate the performance of the IBM machines for solving applications of interest to DOE. We found that our DOE applications ran poorly on POWER/4, well below expectations, but the same applications ran better than expected on POWER/2. Results on a POWER/4 at a different site were moderately better, suggesting that the expected performance is sensitive to OS level and setup.
- Task 4. ORNL will port and verify compatibility and evaluate performance of the LAPACK linear algebra library executing on POWER/2. ORNL will inform IBM of

any problems in porting or testing and write a report describing the POWER/2 evaluation. The LAPACK port to POWER/2 was completed. The ORNL Technical report ORNL/TM-11768 describes the performance of the RS6000 system (general name for the POWER series) on linear algebra operations. Performance was excellent and attributed to the optimizing Fortran compiler. The C++, Fortran 90, and multicast kernel software packages were evaluated on the POWER/2. IBM was kept informed of all problems that were found.

## **2.2. Partner Accomplishments**

- Task 1. IBM will supply architecture documents on the new multiprocessor. Documents were delivered.
- Task 2. IBM will provide Internet access to one of the POWER/4 machines for the duration of the project. Problems building POWER/4 prototypes delayed access by nine months, but access was finally arranged in August 1993.
- Task 3. IBM will supply software manuals and technical references for extended AIX, parallelizing Fortran, and their parallel debugger. IBM will provide access to a software architect for 18 months to support the porting efforts. The manuals were delivered. The software architect provided excellent support.
- Task 5. IBM will provide ORNL with a POWER/2 system and the technical resources required to maintain the hardware and software in an operational mode. IBM will provide ORNL with the following software: AIX 3.2.5 (beta), X-windows, TCP/IP, C, C++, and Fortran compilers. IBM will use problem reports to correct product deficiencies and will make fixes available to ORNL as time permits. POWER/2 system and listed software were provided to ORNL. Technical assistance to maintain systems was provided. Fixes were made available to ORNL in a timely manner.

## **3. Benefit to DOE**

The cooperative research venture has benefited both parties. IBM got a steady stream of evaluation/problem reports used to improve the commercial acceptance of its prod-

ucts. ORNL had the opportunity to evaluate DOE computational problems on state-of-the-art prototype hardware.

There are many benefits of this research to DOE. First, part of DOE's High Performance Computing and Communication (HPCC) initiative involves the evaluation of state of the art computer hardware. This CRADA provided an opportunity for ORNL to gain early access to three different IBM architectures and evaluate them. Second, ORNL was able to get access to beta software packages to apply to DOE research. Research often progresses because of the availability of new technology. By having access to beta technology ORNL has the opportunity to be "first" in research breakthroughs. Having early access allows ORNL to influence software developers to make sure the software has the features required by DOE problems. Finally, by having advanced knowledge of the stability and performance of certain software packages, the ORNL PI has acted in the role of advisor to production staff members at the Lab about whether a particular software product should be upgraded lab-wide.

#### **4. Technical discussion**

The goal of this project was to determine the commercial viability of prototype architectures, discover and correct software problems in the supporting software prior to general market availability of the product, and enhance the marketability of the architectures by porting PVM and LAPACK.

##### **4.1. POWER/4**

IBM POWER/4 architecture is four RS/6000 model 530 workstations interconnected by an atomic memory complex and several megabytes of shared memory. We evaluated if such a shared memory architecture could effectively solve the computational problems of the user community.

The IBM POWER/4 architecture as conceived in 1991 is a hybrid of the symmetric multiprocessor system, where everything is shared, and the cluster multiprocessor system, where nothing is shared. In the POWER/4 architecture, nothing is shared except where explicitly needed. It integrates existing uniprocessor complexes, processors, and I/O, together with a shared, global memory. An atomic complex is designed to provide interprocessor communication, process synchronization and a monotonic

counter/incrementer for event ordering. In addition, the POWER/4 memory subsystem provides local memory for each processor to reduce accesses to the shared global memory. The local memory feature is designed to enhance the system scalability.

The theory behind the POWER/4 design is that with its shared memory support and the load/store model preserved, it can inherit most of the shared memory processor (SMP) software technology in particular parallel compilers. And at the same time, with aggressive use of local memory, the POWER/4 can scale beyond the current limits of SMP, particularly if the shared memory is used only as a fast communication path. Shared memory as a communication channel has low latency, and it can have bandwidth close to memory copy speeds.

The POWER/4 system is composed of four POWERserver processors each running at 42 Mhz in a single box. Each box has a the potential of 320 Mflops peak. The operating system is AIX 3.2 with kernel extensions to accommodate the atomic complex. Local memory can be configured from 16 MB to 64 MB per processor, and the global memory can be configured from 64 MB to 448 MB per box.

Initially, ORNL obtained accounts on a half-speed POWER/4 system located at Florida State University. This system was used for learning about the system and to create the initial design for a PVM port to the POWER/4. There was a nine month delay in the manufacture of full-speed systems due to hardware problems and IBM lawyers approving access to outside beta testers. ORNL got an account on a full-speed POWER/4 system located at the IBM site in Austin Texas.

A full shared memory port of PVM was completed for the POWER/4 architecture. Several OS problems arose during this development. They were reported to IBM who fixed them. The beta version of AIX 3.2 used on the POWER/4 underwent several updates over the course of the CRADA. By the end of the POWER/4 evaluation period, the PVM port was finished and available to IBM, but there was still one outstanding bug in the AIX version running on the Austin POWER/4.

Two DOE application codes were run on the Austin POWER/4 to evaluate its compiler performance, CPU performance, and interactive performance. One code calculated the electronic density-of-states of high temperature superconductors. These calculations were done from first principles requiring the solution of thousands of linear systems per iteration until self-consistency was reached. The other code simulated

a laser ablation process. In laser ablation, a beam of light heats and then vaporizes the surface of one material forming a plasma plume which deposits a thin layer of the material on nearby surfaces. The simulation studied how the shape, direction, and density of the plume could be controlled using the shape and intensity of the laser and thus indirectly controlling the deposited film. The ability to coat a surface with a material that normally would not adhere to it makes laser ablation of great interest to DOE.

Both of these application codes were run on IBM RIOS-1 workstations (model 530 and 320) for comparison. At the time of the tests, these workstations were considered the state-of-the-art in terms of performance and compiler technology. The performance of the Austin POWER/4 was much lower than the IBM 530, in some tests by as much as a factor of two. No other users were logged onto the system during the tests, and the tests were run at different times over the course of several weeks. These results were not consistent with what was expected. To verify, the same tests were repeated on the half-speed machine at FSU. The FSU POWER/4 machine ran the tests at the same speed as the IBM 530. This is consistent with the facts that applications ran with speedup approximately one on the POWER/4 and that the IBM 530 is 25 Mhz (about half the POWER/4 speed).

One conclusion drawn from these tests was that the Austin POWER/4 had a problem in its OS version or perhaps a hidden compute-intensive process running on it continuously.

The performance of the POWER/4 shared memory system was found to be much lower than expected, this and other production problems led to IBM's decision not to make the POWER/4 a commercial product.

#### **4.2. POWER/2**

The POWER/2 Architecture is the follow-on of the RIOS-1 design. The design has been enhanced by adding dual fixed point execution units, dual floating point execution units and larger instruction and data cache sizes, 32 KB and 256 KB respectively. The POWER/2 has the ability to do quad-word floating point load and store and hardware square root. Also new on the POWER/2 are hardware performance monitors accessible from software. The architecture is built from five separate custom designed VLSI chips:

a fixed point unit, an floating point unit, instruction cache unit, data cache unit, and a storage control unit.

The fixed point chip contains the general purpose registers and two arithmetic logic units, which execute all fixed point instructions. It provides the directories and control for the data cache and controls the extension of both fixed and floating point load/store instructions.

The floating point chip contains two full 64-bit double-precision floating point units which conform to IEEE 754 binary floating point standard and performs IEEE 64 bit double-precision operations. The floating point and fixed point are equal priority and independent functional units. At a given cycle, two fixed point and two floating point instructions can be executed simultaneously. A dual quad-word data bus allows four floating point registers to be loaded during each cycle. Each floating point unit can perform a multiply/add on each cycle. Therefore, up to four floating point instructions can be executed per cycle.

The instruction cache chip contains a two-way set associative 32 KB cache with a cache line size of 128 bytes. This chip processes branch and condition register operations, removes them from the instruction stream, and dispatches the rest of the instructions to the fixed and floating point units. In most cases, the fixed and floating point units receive an uninterrupted instruction sequence and do not see the effect of the branches. When the outcome of the branch is determined, the branch-not-taken path instructions are canceled. The branch-taken path is fetched from the instruction cache array and dispatched as target instructions.

The POWER/2 has a four-way set associative dual port 256 KB data cache divided into four separate data cache unit chips. The cache line size is 256 bytes, and the cache is implemented as a store-back cache to minimize the memory bus traffic.

The storage control chip contains the controls and configuration registers for the system memory, controls for the system I/O bus, and controls the interface between the system memory and the data cache units. This chip also controls the interface to the system read-only memory used for initialization of the CPU.

The first tests we ran on the POWER/2 were matrix kernels and scientific benchmarks. The theoretical peak performance of the POWER/2 is 266 Mflops, given it can do four operations per cycle and has a clock rate of 66.6 Mhz. Amazingly, we were

able to achieve performance of 236 Mflops from the Linpack benchmark solving a linear system with 1000 unknowns. The benchmark is written completely in Fortran 77. Normally, hand tuning and assembly language routines would be required to get this close to peak performance on a workstation. The POWER/2 was able to achieve this through extensive optimization routines in the xlf Fortran compiler. Achieving this performance occurred only after a process of evaluating several beta compiler options and helping IBM decide which options should be available and which should be default settings. A detailed study of the POWER architecture on linear algebra operations was documented in an ORNL technical report [1].

Scientific applications from materials science, climate modeling, and contaminate-transport simulations were compiled and run on the POWER/2. The high quality compiler optimization made these applications run in the 50-150 Mflop range, which was much better than any other workstation at the time.

## **5. Inventions**

No inventions were produced as part of this cooperative research effort.

## **6. Commercialization**

The POWER/2 architecture is now being marketed by IBM across a full range of products from desktop computers like the Model 390 up to massively parallel computers like the 512 processor SP-2 at Cornell.

## **7. Future collaboration**

Our relationship with IBM was quite good at the end of the CRADA period. IBM asked if we would be interested in continuing to evaluate prototype hardware and beta software as part of our mutual research interests. IBM shipped us \$80,000 worth of hardware and software to continue the research. We drafted a CRADA extension to match their contributions, describing our work in evaluating the IBM J30 symmetric multiprocessor architecture. Due to staff turnover at IBM, it was several months before IBM had evaluated, negotiated, and signed the agreement. The signed agreement was mailed back to Martin Marietta on April 5, 1995. On April 12, the Martin Marietta

Office of Technology Transfer sent a letter to IBM's Software Contracts and Licensing office saying: "Amendment for CRADA Number ORNL 91-0046B we sent you on October 31, 1994, has been declined by the Department of Energy (DOE) due to the extreme time delay in renewing the Agreement."

Since this letter was sent after MMES should have had the signed agreement, the impression it gave of ORNL's (lack of) desire to establish CRADA's was much worse than if the letter had been sent during the long negotiation phase. Neither the ORNL PI nor the IBM PI had knowledge that such a letter was being drafted by the Office of Technology Transfer. They learned of it only after it had been sent, so they were not able to soften the impact of this impression. I particularly felt for the IBM PI who had worked for months pushing this agreement through the slow IBM legal system.

Our relationship with IBM cooled significantly after the letter was sent. The ORNL and IBM PI's have remained on good terms, and IBM has recently asked if ORNL could help in the research and evaluation of their AIX 4.2 operating system. This research would involve no exchange of funds and would be completed under a standard IBM beta tester agreement.

## **8. Conclusions**

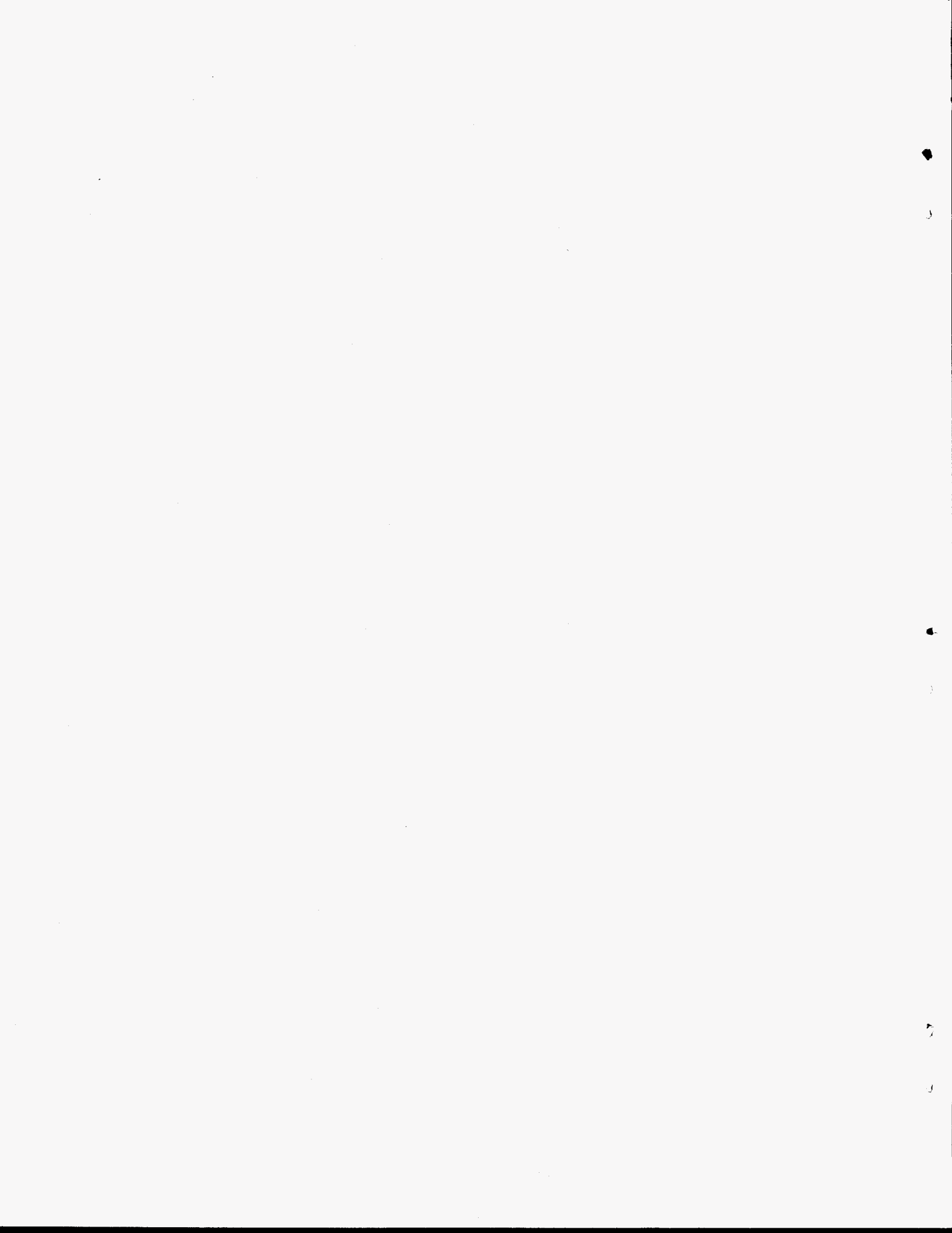
Due to problems with the POWER/4 architecture, some of which were brought to light by this CRADA, IBM decided not to develop the POWER/4 architecture as a product. Instead, IBM is focusing its marketing on the POWER/2 architecture, which our evaluations showed performed even better than expected. The cooperative research venture has benefited both parties. IBM received a steady stream of evaluation/problem reports that it used to improve the commercial acceptance of its products. ORNL had the opportunity to evaluate DOE computational problems on state-of-the-art prototype hardware.

In contrast the POWER/2 architecture has exceeded performance expectations and has demonstrated its ability to solve large computational problems of interest to the user community. IBM is now focused on marketing POWER/2 systems and developing an entire line of computers based on this technology.



## 9. References

- [1] J.J. Dongarra, Peter Mays, and Giuseppe Radicati di Brozolo. The IBM RISC system/6000 and linear algebra operations. Technical Report ORNL/TM-11768, Oak Ridge National Laboratory, January 1991.
- [2] E. Anderson, et al. *LAPACK Users' Guide*. SIAM, 1992.
- [3] Al Geist, Adam Beguelin, Jack Dongarra, Weicheng Jiang, Robert Manchek, and Vaidy Sunderam. *PVM: Parallel Virtual Machine*. MIT press, 1994.



**INTERNAL DISTRIBUTION**

- |        |               |        |                                 |
|--------|---------------|--------|---------------------------------|
| 1.     | T. S. Darland | 20-24. | R. F. Sincovec                  |
| 2-6.   | G. A. Geist   | 25.    | Central Research Library        |
| 7-11.  | M. R. Leuze   | 26.    | K-25 Applied Technology Library |
| 12.    | C. E. Oliver  | 27-28. | Laboratory Records Department   |
| 13.    | W. P. Painter | 29.    | Laboratory Records Dept. - RC   |
| 14-18. | S. A. Raby    | 30.    | Patent Office                   |
| 19.    | T. M. Rosseel | 31.    | Y-12 Technical Library          |

**EXTERNAL DISTRIBUTION**

32. Roger W. Brockett, Wang Professor of EE and CS, Div. of Applied Sciences, 29 Oxford St., Harvard University, Cambridge, MA 02138
33. Office of Assistant Manager for Energy Research and Development, U.S. Department of Energy, Oak Ridge Operations Office, P.O. Box 2001, Oak Ridge, TN 37831-6269
- 34-35. Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37830

