Benchmarks of the PC-UNIX Computer with Electronic Structure Calculation

Jen-Shiang K. Yu[†]

Department of Nuclear Science, National Tsing Hua University, Hsinchu, Taiwan 300

Chin-hui Yu*

Department of Chemistry, National Tsing Hua University, Hsinchu, Taiwan 300

Received April 9, 1997[⊗]

Several ab initio and density functional theory computations were performed on PC-UNIX machines with the GAUSSIAN 94 program to compare the utilized central processing unit (CPU) time with workstations and supercomputers. The performance of the Intel PentiumPro 200 CPU was about half to one-fifth the speed of high-level workstations for most types of computations. Utilizing the *RAM DISK* feature developed in the Linux operating system, we assigned the scratch files to be processed in the main memory; the total CPU utilization of small jobs, such as Hartree−Fock calculations, showed little difference from the calculation where the scratch files were processed in the magnetic disk. The result indicates that the I/O architecture of the PC is not the main bottleneck for nongigantic computation. Using a PC with Intel PentiumPro 200 CPU to run the electronic structure calculation can be at worst 12 times slower than using a fast superscalar model, such as the SGI PowerChallenge R8000, while processing the geometry optimization for formyl cyanide at the MP2/6-311G** level of theory. Nevertheless, PCs finish the jobs within reasonable time for the majority of the tested cases. Considering the ratio of price−performance, the PC is an attractive alternative platform for certain types of electronic structure calculations.

INTRODUCTION

Workstations with RISC architecture have reduced the gap between supercomputers and workstations, while the development of personal computers (PCs) has improved dramatically in recent years. The increasing speed of PCs has blurred the division between personal computers and workstations. Enhanced hardware performance and decreasing price, together with the numerous public domain or free software, make personal computers competitive with workstations. Programs for molecular modeling have been ported and run on PCs, and highly satisfactory results have been obtained.[1] However, another branch in computation chemistry, the determination of quantum electronic structure of molecules, has not yet been systematically performed on PCs. Simple tests of ab initio calculations on PCs with non-UNIX operating system[2] have been run, and remarkable improvement of the central processing unit (CPU) speed has been observed. However, the comprehensive testing of computations at the research level is still absent.

This article reports a series of benchmarks of ab initio and density functional calculations with different methods and basis sets carried out on various supercomputers, RISC workstations and Intel-based personal computers running UNIX-based operating systems.

BENCHMARKING DETAILS

The benchmark of electronic structure calculations involved both ab initio and density functional methods. Ab initio computations, including Hartree−Fock (HF), second order Møller−Plesset perturbation theory (MP2), and quadratic configuration interaction with single and double excitations (QCISD),[3] were performed with the molecule formyl cyanide and one of its transition states of the unimolecular dissociation.[4] The 6-311G** basis set was employed at all levels of theory. Geometry optimization and harmonic vibrational frequencies were calculated with the above levels of theory. Single-point energy at the level of QCISD(T,E4T) was also carried out. Density functional theories including BP86,[5] BLYP,[6] and B3LYP[7] were carried out using the aug-cc-pVTZ[8] basis set for the molecule SO_3.[9]

System hardware and software configurations for the computers are shown in Table 1. All of the computations were done with the GAUSSIAN 94 program[10] compiled on each computer system using its native FORTRAN compiler and the default optimization option offered in the makefile of the GAUSSIAN program. GAUSSIAN 94 computations on each computer were performed with the same input file to provide a consistent computing environment such as memory usage and maximum disk access. The single CPU usage was enforced during each calculation.

The Intel Pentium 100 CPU was installed on an ASUS P/I/P55T2P4 motherboard with 512kB of on-board pipeline burst L2 cache memory. Four banks of an NEC-60ns 72-pin 32MB fastpage memory module were equipped to make the system 128MB main memory on board. The slackware v3.0 Linux system was installed, and the kernel was upgraded to version 2.0.16. The PentiumPro 200 system running Linux v2.0 was equipped with an ASUS P/I/P6NP5 motherboard. Another system running FreeBSD v3.0 was installed on an Elite Group P6FX1-A motherboard. The on-chip L2 cache memory of this CPU was 256kB. The same RAM modules as those for the Pentium 100 were installed on a PentiumPro 200. Due to the lack of noncommercial native FORTRAN compiler written for PC-UNIX, a FORTRAN-to-C translator, f2c, was used to translate the FOR-

* Author to whom correspondence should be addressed.

[†] Present address: Department of Chemistry, National Tsing Hua University.

[⊗] Abstract published in *Advance ACS Abstracts,* September 1, 1997.

S0095-2338(97)00284-9 CCC: $14.00

Table 1. System Configuration of Machines Used for Benchmark Test

MFR	model	CPU	ext freq[a]	RAM	cache L1/L2	OS	Linpack benchmark[b]	Fortran77 compiler
Fujitsu	VPP300/2			2G	64k/−	UXP/V 4.1	203/1936	
SGI	Power ChallangeXL	R8000	75	2G	32k/4M	Irix 6.1	104/261	
Cray	J916/8	YMP	100	1G		Unicos8.0.4.2	106/203	6.0.4.3
IBM	9076 SP2	Power2	66.7	128M	160k/−	AIX 3.2.5.4	130/236	3.2.2
IBM	390	Power2	66.5	256M		AIX 3.2.5.1	53/181	
IBM	37T	Power2		64M		AIX 3.2.5	N/A	
DEC	3000/900	21064A	275	256M	32k/2M	OSF1/v3.2	52/193	3.8
Convex	SPP1000/CD	PA-RISC7100	100	1G	2M/−	HP-UX A.09.03	48/123	
HP	9000/755	PA-RISC7100	99	185M	384k/−	HP-UX A.09.05A	41/120	
TwinHead	Sparc20Model 50	TI-SuperSPARC	50	256M	36k/−	SunOS 4.1.3_U1	N/A	
Intel	ASUS/P55T2P4 [c]	Pentium	100	128M	16k/512k	Linux 2.0.16	12/NA	f2c 19950112
Intel	ASUS/P6NP5	PentiumPro	200	128M	16k/256k	Linux 2.0.16	38/NA	with gcc 2.7.2.1
Intel	Elite G./P6FX1-A[d]	PentiumPro	200	128M	16k/256k	FreeBSD3.0	38/NA	/ g77 0.5.18 [e]

[a] In the unit of megahertz. [b] Linpack benchmarks from http://performance.netlib.org/performance/html/PDSreports.html. The unit is MFLOPS. The results of solving for matrices of dimension 100 × 100/1000 × 1000 are listed, using a single CPU. [c] The detailed information of the motherboard can be found at http://www.asus.com.tw. [d] The detailed information of the motherboard can be found at http://www.esc.com.tw. [e] The GAUSSIAN 94 program was compiled on Linux and FreeBSD with an f2c compiler and g77 0.5.18 individually, both with a gcc2.7.2.1 back end.

Table 2. CPU Time of Each Computer for the Geometry Optimizations, Harmonic Vibrational Frequency Calculations, and Single-Point Energetics with Various Levels of Theories Using 6-31G** Basis Sets[a]

machine	HF/opt[b]	HF/freq	MP2/opt	MP2/freq	QCISD/opt	QCISD(T,E4T)/sp
Fujitsu	00:09:09.9	00:07:52.0	00:06:30.5	00:20:59.1	00:38:34.9	00:56:57.4
SGI/PC	00:11:41.9	00:10:15.8	00:11:18.5	00:38:15.1	03:28:42.5	00:40:08.3
Cray/J916	00:18:24.5	00:13:33.6	00:06:37.6	00:45:44.2	04:26:40.0	00:48:37.8
IBM/SP2	00:14:15.0	00:10:49.2	00:11:47.9	00:33:01.6	04:28:21.6	00:31:41.2
IBM/390	00:21:31.1	00:16:19.3	00:44:16.3	00:49:01.2	06:30:57.5	00:47:56.6
IBM/37T	00:31:12.6	00:23:35.8	01:33:54.0	01:18:04.2	06:52:59.6	01:12:53.3
DEC3000	00:16:13.6	00:14:04.4	00:27:11.5	00:57:39.6	07:20:00.9	00:57:58.8
ConvexSPP	00:25:32.7	00:20:31.9	00:32:48.0	01:28:19.5	06:34:13.2	01:27:12.1
HP9000/755	00:33:52.8	00:26:11.5	00:56:49.6	02:21:54.6	12:28:51.1	02:00:00.6
SunSparc20	01:14:04.2	00:55:11.3	03:34:57.6	03:34:14.4	18:49:31.0	04:45:03.7
P-Pro 200(L)[c]	00:49:32.3	00:39:44.9	02:30:42.1	02:56:45.3	14:28:26.3	05:12:58.5
P-Pro 200(B)[d]	00:47:55.7	00:26:34.7	02:15:33.8	02:52:35.7	14:54:13.1	05:43:40.5
Pentium 100	01:34:45.7	01:17:10.4	05:40:28.6	06:03:31.9	28:31:45.5	08:59:54.1

[a] The calculations were limited on a single CPU and four MegaWord of memory usage. The sample molecule is formyl cyanide. [b] Input keywords: HF/opt, SCF=DIRECT OPT; HF/freq, SCF=DIRECT GEOM=CHECK GUESS=READ FREQ; MP2/opt, SCF=DIRECT GEOM=CHECK GUESS=READ OPT; MP2/freq, SCF=DIRECT GEOM=CHECK GUESS=READ FREQ; QCISD/opt, SCF=DIRECT GEOM=CHECK GUESS=READ OPT; QCISD(T,E4T)/single point, SCF=DIRECT GEOM=CHECK GUESS=READ. The above jobs were performed sequentially in the listed order. [c] (L) means with the Linux operating system installed. [d] (B) means with the FreeBSD operating system installed.

TRAN source codes of GAUSSIAN 94 to C language, and the generated C codes were compiled with the GNU C compiler (GCC, version 2.7.2.1, with an additional optimization flag −*O3* and a new option −*m486* to optimize for the Intel 80486+ CPU based machines). There is an unofficial release of GCC which supports −*mpentium* optimization option for pentium CPUs. However, it is not used in this work due to the concern of the compiler's stability and the reliability of the results.

RESULTS AND DISCUSSIONS

The CPU time of the various hardware for computations of geometry optimization, vibrational frequencies and single-point energies of formyl cyanide and its transition state for unimolecular dissociation are listed in Tables 2 and 3. The CPU time of geometry optimizations for SO_3 using density functional theories are presented in Table 4. The results, such as optimized structures, vibrational frequencies, and the energetics, were verified against those in refs 4 and 9. The deviations of the results among different computers are negligible.

Table 2 summarizes the CPU time of various types of ab initio calculations of formyl cyanide. For Hartree−Fock geometry optimization, the speed of the PentiumPro 200 is about 20% of that of the SGI PowerChallenge with a R8000 CPU (SGI R8000) and is about 40% of that of an IBM 390 workstation. For vibrational frequency computation at the Hartree−Fock level, the PentiumPro 200 runs one-third to one-fourth the speed of the SGI R8000 and is about half the speed of an IBM 390 workstation. For MP2 computations, the PentiumPro 200 is slower than the SGI R8000 by about 12 times for geometry optimization and by 4.5 times for frequency calculation, whereas it is slower than the IBM 390 by about 3.5 times for both optimization and frequency calculations. For QCISD optimization, the PentiumPro is about 4.5 times slower than R8000 and is around 2.5 times slower than the IBM 390. For the QCISD(T,E4T) single-point energy calculation, the PentiumPro 200 is an eighth of the performance of the SGI R8000 and IBM 390 machines.

Table 3 presents the transition-state search results. The PentiumPro 200 is about one-third of the speed of the SGI R8000 and IBM 390 for HF transition-state optimization and is about one-fourth of the performance of the SGI R8000

Table 3. CPU Time of Each Computer for the Geometry Optimizations, Harmonic Vibrational Frequency Calculations, and Single-Point Energetics with Various Levels of Theories Using 6-311G** Basis Sets[a]

machine	HF/opt[b]	HF/freq	MP2/opt	MP2/freq	QCISD/opt	QCISD(T,E4T)/sp
Fujitsu	00:11:36.2	00:04:26.0	00:06:15.4	00:19:20.1	03:33:16.8	01:02:05.7
SGI/PC	00:18:21.4	00:06:27.9	00:24:57.2	00:36:06.1	04:53:34.9	00:45:18.5
Cray/J916	00:11:21.0	00:05:20.4	00:17:48.1	00:44:21.1	06:29:07.0	00:53:15.1
IBM/SP2	00:14:09.1	00:05:39.1	00:20:22.6	00:32:29.6	02:39:26.8	00:38:14.6
IBM/390	00:20:38.2	00:08:48.2	00:25:39.1	00:48:27.4	06:50:18.7	00:55:00.2
IBM/37T	00:32:15.5	00:14:10.7	00:31:32.9	01:18:53.0	06:38:50.4	01:29:11.4
DEC3000	00:20:38.3	00:10:30.8	00:34:41.5	00:56:03.3	04:29:20.3	01:06:08.7
ConvexSPP	00:29:06.5	00:14:42.7	02:02:38.6	01:20:55.2	09:32:33.8	01:33:56.9
HP9000/755	00:39:45.2	00:18:35.2	00:40:50.4	02:22:10.5	19:02:21.1	02:32:23.9
SunSparc20	00:55:59.4	00:31:38.8	01:11:53.8	03:33:54.7	15:40:55.3	05:13:35.3
P-Pro 200(L)[c]	00:59:23.3	00:24:18.8	00:58:49.7	02:40:25.7	12:41:56.8	05;43:13.3
P-Pro 200(B)[d]	00:40:19.8	00:16:57.4	00:48:57.7	02:50:45.9	13:05:59.9	07:10:54.0
Pentium 100	02:41:22.2	00:45:26.6	01:22:52.2	05:56:55.8	35:44:44.7	09:39:01.9

[a] The calculations were limited to a single CPU and eight MegaWord of memory usage. The sample molecule is a transition state for the unimolecular dissociation of formyl cyanide. [b] Input keywords: HF/opt, SCF=DIRECT OPT=(calchffc,noeigentest,z-mat); HF/freq, SCF=DIRECT GEOM=CHECK GUESS=READ FREQ; MP2/opt, SCF=DIRECT GEOM=CHECK GUESS=READ OPT=(readfc,noeigentest,z-mat); MP2/freq, SCF=DIRECT GEOM=CHECK GUESS=READ FREQ; QCISD/opt, SCF=DIRECT GEOM=CHECK GUESS=READ OPT=(readfc,noeigentest,z-mat); QCISD(T,E4T)/single point, SCF=DIRECT GEOM=CHECK GUESS=READ. The above jobs were performed sequentially in the listed order. [c] (L) means with the Linux operating system installed. [d] (B) means with the FreeBSD operating system installed.

Table 4. CPU Time of Each Computer of the Geometry Optimizations for SO_3 Using BP86, BLYP, and B3LYP Density Functional Theories with the aug-cc-pVTZ Basis Set[a]

machine	BP86[b]	BLYP	B3LYP
Fujitsu	08:34:36.9	06:33:09.9	10:53:38.2
SGI/PC	23:29:30.1	17:24:21.9	20:58:27.0
Cray/J916	23:11:58.6	17:52:47.9	21:08:24.5
IBM/SP2	18:25:12.6	13:43:21.2	05:19:28.9
IBM/390	26:54:10.6	19:49:01.1	24:05:28.2
IBM/37T	43:21:05.5	39:53:49.0	32:42:08.1
DEC3000	36:09:27.3	26:45:36.4	10:07:30.6
ConvexSPP	49:52:20.6	*e*	*e*
HP9000/755	61:03:33.9	45:22:46.3	53:05:07.4
SunSparc20	131:02:17.8	96:47:29.7	114:44:40.8
P-Pro 200(L)[c]	115:10:51.9	84:59:54.9	98:32:47.3
P-Pro 200(B)[d]	77:33:40.8	56:47:01.0	59:03:06.4

[a] The calculations were limited to a single CPU and eight MegaWord of memory usage. [b] Input keywords: BP86, SCF=DIRECT OPT=(calcall,tight,z-mat); BLYP, SCF=DIRECT GEOM=CHECK GUESS=READ OPT=(calcall,tight,z-mat); B3LYP, SCF=DIRECT GEOM=CHECK GUESS=READ OPT=(calcall,tight,z-mat). The above jobs were performed sequentially: BP86, followed by BLYP, and then B3LYP. [c] (L) means with the Linux operating system installed. [d] (B) means with the FreeBSD operating system installed. [e] The BLYP and B3LYP calculations on the Convex SPP 1000 cannot be completed due to the policy of account rearrangement in the National Center for High-Performance Computing of the Republic of China.

for HF frequency computation. For MP2 geometry optimization, the SGI R8000 is double the speed of the PentiumPro 200. For frequency calculation at the MP2 level, the IBM 390 and the SGI R8000 is three and four times faster than the PentiumPro 200, respectively. For QCISD transition-state optimization, the PentiumPro 200 is about two and half times slower than the SGI R8000 and is about two times slower than the IBM 390. For single-point energy calculation at QCISD(T,E4T), the PentiumPro 200 is about eight times slower than the SGI R8000 and is six and half times slower than the IBM 390.

The CPU time of the larger-scale density functional theory (DFT) computations are listed in Table 4. The speed of the PentiumPro 200 varies significantly with the equipped operating system. FreeBSD runs faster than Linux for these DFT jobs by 50%. The PentiumPro 200 with FreeBSD runs about one-third the time of the SGI R8000 and is about 2.4−2.8 times slower than the IBM 390. The Pentium 100 is too slow for jobs of this size and was not tested. From the above data, it is evident that the PentiumPro 200 can practically be employed to execute the computation of first principle electronic structure theories. Its performance is better than some middle-level workstations and is about half to one-third the speed of several fast workstations. However, considering the lower price, the PC is an economical alternative with satisfactory performance.

A significant portion of the computation of electronic structure theory is solving simultaneous linear equations of huge dimensions. LINPACK[11] is a code written in FORTRAN to solve linear equations. The LINPACK benchmark[12] in Table 1 was done by solving 100 × 100 and 1000 × 1000 matrices with double precision. The results showed that the change of the dimension of the matrices affected the performance of the computers significantly, and this trend is consistent with that in the calculations of electronic structure theory.

To prevent possible I/O delay during the data transfer between the main memory and the magnetic disk, an additional calculation of formyl cyanide on the Intel PentiumPro 200 with Linux was performed using Hartree−Fock theory. Out of the total of 128MB of main memory, 64MB were created into a *RAM DISK file system* using the RAM DISK feature of the Linux operating system. The scratch files generated by GAUSSIAN 94 were assigned to be processed inside this *RAM DISK file system*; that is, a virtual hard disk with 60 ns access time was used. Thus, the computations were done with an *in memory* stage to eliminate the disk I/O delay. However, the results showed only seconds of difference in CPU time compared with the results listed in Table 2. Hence, the difference in I/O architecture between PCs and workstations is a minor factor in small jobs such as HF calculations.

To assess the execution rate, another version of the FORTRAN translator, GNU FORTRAN 77 (g77, version 0.5.18), was used to compile the GAUSSIAN 94 program. The g77-compiled GAUSSIAN 94 executables generated correct results compared with f2c-compiled executables. However, the CPU time consumption was almost the same as that of the f2c version. The f2c is a translator for

FORTRAN to C; then it uses GCC to compile the translated C codes.[13] For GNU's compilers such as GCC and G77, they all use the same *back-end compiler* (*GBE*, refers to GNU back end) to generate the machine code. For example, the GCC drives the source C file sequentially through the preprocessor (*cpp*), the C compiler (internally named *cc1*), the assembler (usually *as*), and the linker (*ld*) to produce the executable program. G77 is similar: it uses its internal *front end* (named *f771*), which interprets FORTRAN codes to determine what they intend to do, and then communicates that knowledge to the *GBE* for actual compilation of these programs, calling the *as*, and lastly *ld*, in analog to GCC. Hence, these two compilers share many common demands; the quality of generated codes (in terms of speed and size) is similar because the back-end compilation was carried out by the shared *GBE*. Accordingly, f2c- and g77-compiled codes utilize nearly the same CPU time for identical jobs.

The LINPACK benchmark report[12] supports the aforementioned point of view. An Intel PentiumPro 200 running SunSoft Solaris 2.5 operating system with g77 version 0.5.5 obtains 38 MFLOPS solving 100 × 100 LINPACK matrices, while the same computer running Microsoft FORTRAN under the Microsoft WindowsNT operating system reaches 62 MFLOPS solving an identical problem. In conclusion, a native FORTRAN compiler with an optimized back-end compiler for its own platform and operating system may be an important factor affecting the efficiency of the execution rate of the codes.

CONCLUDING REMARKS

It is possible to run FORTRAN coded electronic structure calculation programs on Intel CPU based personal computers under a UNIX environment. Porting those programs into C language or obtaining a high-performance FORTRAN compiler optimized for Intel CPUs may generate more efficient executables to evaluate the best performance of the PC.

ACKNOWLEDGMENT

This work was supported by the National Science Council of the Republic of China under Grant NSC 86-2113-M-007-018. We thank the National Center for High-Performance Computing of the Republic of China and the Computer and Communication Center of the National Tsing Hua University for offering the computation resource.

REFERENCES AND NOTES

(1) Tirado-rives, J.; Jorgensen, W. L. *J. Comput. Chem.* **1996**, *17*, 1385−1386.

(2) *Gaussian News*, 1996 Summer Issue; Gaussian, Inc: Pittsburgh, PA, 1996.

(3) Pople, J. A.; Head-Gordon, M. ; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968−5975.

(4) Chang, N. Y.; Yu, C.-h. *Chem. Phys. Lett.* **1995**, *242*, 232−237.

(5) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100. (b) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824.

(6) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(7) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789. (c) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(8) Kendall, R. A.; Dunning, Jr., T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806.

(9) Jou, S. H.; Shen, M. Y.; Yu, C.-h.; Lee, Y. P. *J. Chem. Phys.* **1996**, *104*, 5745−5753.

(10) *Gaussian 94*, Revision D.2; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 1995.

(11) Dongarra, J. J.; Moler, C. B.; Bunch, J. R.; Steward, G. W. *LINPACK Users' Guide*; Society for Industrial and Applied Mathematics: Philadelphia, 1979.

(12) Benchmark from http://performance.netlib.org/performance/html/PDSreports.html.

(13) (a) Aho, A. V.; Sethi, R.; Ullman, J. D. *Compilers: Principles, Techniques, and Tools*; Addison-Wesley: Reading, MA, 1988. (b) The *GNU Task List* of GCC and G77.

CI970284F