Nonlinear Analysis: Modelling and Control, 2002, Vol. 7, No. 2, 113–121

PC Cluster Possibilities in Mathematical Modeling in Quantum Mechanical Molecular Computations

R.J. Rakauskas[1], J. Šulskus[2], S. Vošterienė[3]

[1]Gen. J. Žemaitis Lithuanian Military Academy, Šilo 5A, Vilnius, Lithuania
rimantas.rakauskas@tmk.lka.lt

[2]Gen. J. Žemaitis Lithuanian Military Academy, Šilo 5A, Vilnius, Lithuania
Vilnius University, Saulėtekio 9, Vilnius, Lithuania
juozas.sulskus@ff.vu.lt

[3]Gen. J. Žemaitis Lithuanian Military Academy, Šilo 5A, Vilnius, Lithuania
svajone.vosteriene@tmk.lka.lt

Received: 11.09.2002
Accepted: 20.11.2002

Abstract. We present the PC cluster built in the Department of Applied Sciences of Lithuanian Military Academy. The structure of the cluster is described and the performance is evaluated by solving of linear algebra testing tasks and nonlinear quantum chemistry molecular electronic structure computations.

Keywords: PC cluster, parallel computations, quantum chemistry.

1 Introduction

Computational quantum mechanical methods are widely used for investigation of molecular electronic structure, dynamical properties of many atomic structures, chemical reactions. From the mathematical point these investigations are nonlinear mathematical modeling problems. The basis of these investigations is quantum mechanics, but in order to achieve quantitative results comparable with experimental we need supercomputer's power or parallel computers cluster. In this way such tasks are solved during the last decade of 20th century. Clusters based on PCs running Linux have become the cheapest "supercomputers" in the academic and commercial field. We created such a cluster in Applied Sciences Department in Lithuanian Military Academy (LMA). The cluster's performance have been tested by solving linear equations systems of different order. Created computational facilities were applied to investigate electronic structure and vibrational spectra of trinitrotoluene molecule by means of non-empirical quantum mechanical computational methods. The tasks solved were used to compare the possibilities of our cluster with other clusters and supercomputers. The accurate nonempirical quantum chemical computations (molecular electronic structure and vibrational spectra investigations) are essential in development of spectroscopic methods for detection of small amounts of materials On the other hand the highly sophisticated quantum chemical computations have been become the routine tests of supercomputers performance.

2 Personal Computer Cluster

With the power and low prices of today's PCs and the availability of 100 Mb/s Ethernet interconnect, it makes possible to combine them to build High-Performance-Computing and Parallel Computing environment. Today, there is a wide range of switches available, ranging from 8 to over 100 ports, some with one or more Gigabit modules that lets you build large systems by interconnecting such switches or by using them with a Gigabit switch. Switches have become inexpensive enough, so there is not much reason to build your network by using cheap hubs or by connecting the nodes directly in a hypercube network.

The local area PC cluster was made in Lithuanian Military Academy for parallel calculations. PC cluster of 22 nodes: 2 of them dual AMD 1.2 GHz processors, 5 – dual 800 MHz Pentium III processors, 8 – dual 733 MHz Pentium III processors; all have 512 MB RAM, a 20 GB disk drives and the rest 7 nodes have 233 MHz AMD processors, 32 MB RAM, 2 GB disk drives. Machines have been installed with the standard RedHat 7.1, Windows NT 4.0 and Solaris 8 OS (parallel cluster runs in Linux OS at this time). All PCs are assumed to boot from a their own hard drive and have a fast Ethernet network (100 Mb/s) connection to a switch controlling the private cluster network. The suggested range of addresses for a private network is from 192.168.100.1 to 192.168.100.20. Nobody is able to connect directly to a compute node from outside this network anyway. This keeps normal traffic from interfering with inter-node communication and vice versa.

Total peak performance of the cluster is 30.96 Gflops; total memory – 7.9 GB; total disk space – 300 GB. Our cluster's "Tauras" parameters are comparable with other ones located in academic areas. "Tauras" is registered in the top500 clusters list [1].

We use the SCore Cluster System Software - a high-performance parallel programming environment for workstation and PC clusters [2]. Score uses the PM II high performance communication library a dedicated communication library for cluster computing (it allows using many types of networks). Communication library used is faster than usual TCP protocol. In addition a trunking is possible (installation and use of up to 4 NICs). SCore, users are not aware whether or not a system is a cluster of single/multi-processor computers or a cluster of clusters. Score system has check point function. It allows restart function and improves a reliability of the system. Available compilers are gcc, g77 and Portland Group compilers C, C++, Fortran and HPF.

The cluster differs from the network of workstations in security, application software, administration, booting and file systems. Application software uses underlying message passing systems like Message Passing Interface (MPI) [3] and Parallel Virtual Machine (PVM). There are many ways to express parallelism, but message passing is the more effective and more modern.

For administering the cluster we have “login node” with a keyboard, monitor and mouse. Other nodes can be “headless” (no keyboard, mouse, or monitor) but in our way there are all machines like “login node”, because our cluster is made on the basis of teaching class.

For testing calculations only first 9 nodes with CPUs faster than 700 MHz have been used.

Testing *ab initio* calculations of geometry optimization of trinitrotoluene molecule in Hartree-Fock (HF) approximation in 6-311G* basis [4] were performed with different numbers of CPUs using program package GAMESS [5]. In table 1 we give the load of different processors during that task.

We can see that 1.2 GHz AMD processor’s performance is about two times greater than 733 MHz Pentium III, and the ratio in MHz is only 1.6. GAMESS code evaluates all computers performance, not only processor. Evaluating peak performance of clusters the AMD processors performance (nodes 2, 3) is usually evaluated as their MHz multiplied by 2 in order to achieve peak performance comparable with Intel CPUs. It's clear that CPU's load will be more equal and cluster's performance is greater when nodes characteristics are the same. During our test (table 1) AMD processors where idle about half the time.

Node	0	1	2	3	4	5	6	7	8
CPU MHz	733	800	1200	1200	800	733	733	733	800
CPU time, min.	3132	2839	1537	1538	2848	3073	3073	3085	2845

Table 1. Dependence of CPU load from the single processor performance

Another is a LINPACK test of solving of linear equations system (table 2.) [6]. The order of linear equations system was changed from 2000 to 15000. 512 MB of RAM per node was not enough to solve system of 20000 equations without memory swapping. The reduction of the performance (column 2) when linear equations orders greater than 8000 may be explained by slower network performance. On other hand node performance plays the same role (columns 4 and 6).

Linear equation system order	8 nodes (LMA) 100 Mbs network - 1 NIC per node	8 Duals Intel PIII 550 Mhz (512 Mb) – Myrinet [6]	4 nodes (1.2 GHz AMD CPUs) (LMA)	4 AMD Athlon K7 500 Mhz (256 Mb) - (2x) 100 Mbs Switched - 2 NICs per node (channel bonding) [6]	4 nodes (PENTIUM III 733 MHz) (LMA)
2000	0.62	1.76	1.51	1.28	0.79
5000	1.49	2.32	2.54	1.73	1.03
8000	1.87	2.51	3.01	1.89	1.11
10000	1.98	2.58	3.13	1.95	1.06
15000	1.91	2.72			

Table 2. Dependence of performance of different size clusters from the order of solved linear equations systems (in Gflops)

The important feature of the cluster that it may be upgraded or expanded without any essential rebuilds. Our cluster speedup (the time of the task with one CPU divided by the time necessary to solve the same task with n CPUs) is shown in figure 1. The task solved was single point trinitrotoluene molecule electronic structure determination in HF approximation using 6-31G* basis set.

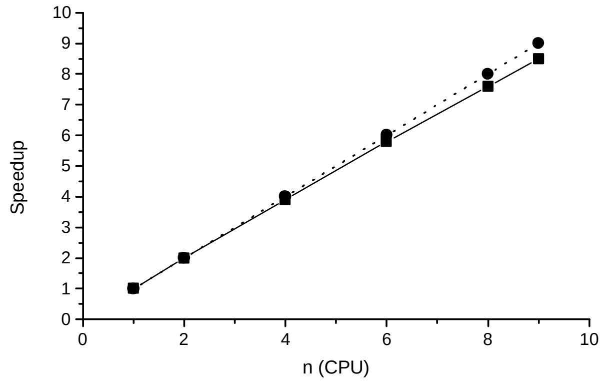

Fig. 1. Speedup against the number of CPUs (n) during trinitrotoluene molecule electronic structure determination in HF approximation (solid line, solid squares; 6-31G* basis). Dashed line (solid circles) is pure linear dependence between n and speedup

The dependence is almost linear like in similar clusters [7]. Of course strait physical expansion of the cluster slows down the performance of the whole cluster. The slow down depends on the problem solved and on the performance of the network communications speed.

3 Cluster in Quantum Chemical Computations

As was stated in the introduction the accurate computations of trinitrotoluene molecule may be of interest for experimental spectroscopy. In this section we briefly describe the complexity of the investigation of electronic structure of molecules and present results and amount of necessary computational facilities in order to achieve reliable results for trinitrotoluene molecule.

Theoretical investigations have been performed by means of *ab initio* quantum mechanical calculations using GAMESS computer code [5]. The basis of all such calculations is solving of multidimensional time independent Shrodinger equation.

$$\widehat{H}\Psi_i = E_i \Psi_i, \tag{1}$$

here $\widehat{H}$ – Hamilton operator of investigated quantum particle system; Ψ_i – wave function of the system; E_i – alowable energy value. Hamiltonian operator and wavefunction of the system depend from the coordinates of all particles (electrons and nucleus).

We use non-relativistic Hamilton operator for electronic part of the problem [4, 9]

$$\widehat{H}_e = \sum_{p=1}^{N} h_p + \sum_{p>q}^{N} g_{pq} , \tag{2}$$

here

$$h_p = \frac{1}{2}\nabla_p^2 - \sum_{\mu=1}^{M} \frac{Z_\mu}{r_{p\mu}} , \tag{3}$$

is one electronic part of Hamiltonian operator; M is the number of nucleus; Z_μ – the charge of μ-th nuclei; $r_{p\mu}$ – the distance between μ-th nuclei and p-th electron;

$$g_{pq} = \frac{1}{r_{pq}} \tag{4}$$

is the two electronic part of Hamiltonian.

We use the atomic units system in all formulae.

In one particle approximation the molecular wave function for closed shell electronic system (with odd number of electrons) is presented by Slatter determinant

$$\Psi = \frac{1}{\sqrt{N!}} \det\left|\psi_1 \overline{\psi}_1 \psi_2 \overline{\psi}_2 \ldots \psi_{N/2} \overline{\psi}_{N/2}\right| , \tag{5}$$

where N is the total number of electrons;

$$\psi_i = \varphi_i \alpha \ , \quad \overline{\psi}_i = \varphi_i \beta , \tag{6}$$

ψ_i and $\overline{\psi}_i$ are one electron wave functions depending from i-th electron coordinates: φ_i – spatial part of electron wave function; α and β – spin orbitals.

Molecular orbitals (MO) φ_i are often expanded in the basis of Cartessian Gaussian atomic orbitals:

$$\varphi_i = \sum_k \chi_k T_{ki} , \tag{7}$$

$$\chi_{\mathrm{GO}}(n,l,m) = N(\alpha;l,m,n) x^l y^m z^n \mathrm{e}^{-\alpha r^2} , \tag{8}$$

or in the matrix form

$$\varphi = \chi \mathrm{T}, \tag{9}$$

here $N(\alpha; l, m, n)$ – normalization coefficient; T – matrix with expansion coeficients for MO in the columns.

The first order variation of the full energy with additional requirement of orthogonality between MOs leads to matrix equation for MO coefficients matrix T and one elctronic energies matrix ε.

$$\widehat{H}^F \mathrm{T} = \mathrm{S}\mathrm{T}\varepsilon, \tag{10}$$

where S is the matrix of overlap integrals between AO ($S_{ij} = \int \chi_i \chi_j \, d\tau$).

$$\widehat{H}^F = \widehat{h} + \sum_n (2\widehat{J}_n - \widehat{K}_n), \tag{11}$$

Coulomb and exchange operators are defined accordingly

$$\widehat{J}_n \varphi_m(\mu) = \int \frac{|\varphi_n(\nu)|^2}{r_{\mu\nu}} d\bar{r}_\nu \varphi_m(\mu); \tag{12}$$

$$\widehat{K}_n \varphi_m(\mu) = \int \frac{\varphi_n^*(\nu)\varphi_m(\nu)}{r_{\mu\nu}} d\bar{r}_\nu \varphi_m(\mu). \tag{13}$$

As operators $\widehat{J}$ and $\widehat{K}$ depend themselves from MO φ_i, equation (10) is usually solved by iterational self consistent field method (SCF).

So all the problem is concerned with evaluation of great number of integrals with (11) operator and the use of linear algebra matrix computational methods.

One of the methods to go beyond the one electron approximation during investigation of molecular electronic structure is use of multiconfigurational wave function of the form

$$\psi = \sum_I C_I \phi_I, \tag{14}$$

where ϕ_I – one determinant function of the form (5).

In that case during SCF procedure coefficients C_I from (14) and expansion coefficients from (7) are determined at the same time. It is very computational resources consuming methods, but when expansions in (7) and (14) are big enough it gives the results comparable with experimental ones.

Different basis sets were used during investigations: from 6-31 G* (with polarization *d* functions on second period elements; all in all 250 basis functions) to 6-311 G** (with two polarization *d* functions on second period elements) plus one *f* function and diffuse *s* and *p* functions on second period atoms and diffuse functions on hydrogen atoms (all in all 770 basis functions).

The basis functions used are necessary in order to achieve reliable results in such investigations During multiconfigurational selfconsistent field (MCSCF) [5, 8] calculations performed for trinitrotoluene we left unfrozen 10 highest occupied MO and 10 lowest unoccupied MO from HF Slater determinant. All configurations generated from unfrozen MOs with spin projection zero were included in MCSCF procedure. We had more than 65000 configurations.

Full geometry optimization was performed for trinitrotoluene molecule in Hartree-Fock and in MCSCF approximations. During investigation the C_1 symmetry point group of the molecule was assumed. Figure 2 shows significant change of electronic charges of atoms and dipole momentum for trinitrotoluene molecule. The results in figure 2 were achieved by the same cluster [9]. Figure 2

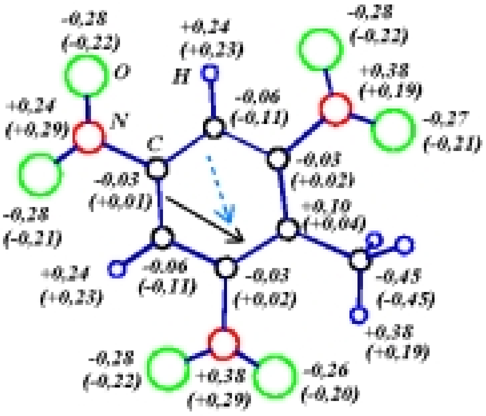

Fig. 2. Charges of the atoms in trinitrotoluole molecule (in a.u.) calculated in HF approximation in 6-31G* basis set and by MCSCF method (in parenthesis) [9]. Arrows show approximate direction of dipole moment calculated in Hatrtee-Fock approximation ($D_{HF} = 1.759\ D$) and by MCSCF method ($D_{MCSCF} = 1.444\ D$, dashed arrow)

show significant influence of electron correlation and necessity of MCSCF method for investigation of optical properties of investigated molecule. multiconfigurational properties.

The results of trinitrotoluene geometry optimization are presented in table 3. Even with such comparatively small cluster we can perform all necessary computations at HF level. It cam be stated that created PCs cluster allows solving of modern quantum chemical problems at necessary level.

Basis set	Numb. of basis functions	$E_{tot.\ HF} + 880$, a.u.	Dipole moment, D	Computational time, min. (h)
6-31G* (1d)	250	-0.11701925	1.7586	146.9 (2.45)
6-311G** (1d)	319	-0.32609999	1.7849	247.4 (4.12)
6-311G** (2d)	415	-0.36559238	1.7020	555.0 (9.25)
6-311G** (2d+1f)	575	-0.40653911	1.7112	1511.7 (25.20)
6-311G** (3d+1f)	671	-0.42604909	1.7277	5349.8 (89.16)

Table 3. Geometry optimization results and necessary CPU time dependence from basis set used (number of basis functions). Computations were performed in Hartree-Fock approximation. The cluster of 9 CPUs was used

References

1. http://clusters.top500.org/db
2. Ishikawa Y. et al. In: *Proceedings of the 5th Annual Linux Expo* **55**, 1999
3. Gropp W., Lusk E., Skjellum A. *Using MPI,* The MIT Press, p. 371, 1999
4. Fudzinaga S. *Molecular* Orbitals *Method*, Moscow, Mir, p. 461, 1983 (in Russian)
5. M.W. Schmidt et al. General Atomic and Molecular Electronic System, *J. Comput. Chem*., **14**, p. 1347, 1993
6. High Performance *Computing Linpack Bentchmark (HPL) HPL 1.0,* www.netlib.org/bentchmark/hpl, September 27, 2000
7. Grimme S., Muck-Lichtenfeld C. Lyra: Construction of a Parallel Computer from Standard PC Components New Impulses for Quantum Chemical Calculations, www.uni-muenster.de/Chemie/OC/research/grime/lyra/lyra-e.html
8. Wilson S. *Electron* Correlation *in Molecules*, Moscow, Mir, p. 304, 1987 (in Russian)
9. Cicėnas S., Rakauskas R.J., Vošterienė S., Šulskus J. Theoretical investigation of the vibrational spectra of trinitrotoluene and 2,4,6-trinitrophenol molecules, *Lithuanian Journal of Physics,* **41**, p., 221, 2001