PVM IMPLEMENTATIONS OF ADVECTION-CHEMISTRY MODULES OF AIR POLLUTION MODELS

B. S. Andersen* **P. Kaae†** **C. Keable‡**
W. Owczarz* **J. Waśniewski*** **Z. Zlatev§**

1 THE PVM PACKAGE

The Parallel Virtual Machine (PVM) Package is available now at many sites and can be used to develop truly heterogeneous programs for multiple computer systems. The PVM Package is a software package developed at Oak Ridge National Laboratory and Emory University. It enables a heterogeneous collection of UNIX computers that are linked in a network to function as a single parallel computer. PVM applications can be run across a wide variety of architectures. PVM handles automatically all message conversions required when the linked computers use different data representations. Participating computers can be distributed anywhere in the world and linked by a variety of networks. It should be possible to distribute the tasks in an optimal way: to send to each computer in the network tasks that are most suitable for its architecture.

*UNI•C, Building 305, DTU, DK-2800 Lyngby, Denmark
†Convex Computer, Struergade 22, DK–2635 Tåstrup, Denmark
‡Cray Research, Oldbury, Bracknell, Berkshire RG12 4TQ, England
§National Environmental Research Institute, Frederiksborgvej 399, DK-4000 Roskilde, Denmark

PVM is a public domain code. The PVM source code and user's guide can be obtained by electronic mail. The source code has been tested on many different computers. Visualisation and digging facilities are available in the PVM program. More details can be found in **Geist et al., [1]**.

PVM has been used to run some modules of a large air pollution model on several different networks of processors. The problems solved will briefly be described in Section 2. An attempt to minimize the communications has been carried out; this will be discussed in Section 3. Some numerical results will be given in Section 4.

2 THE AIR POLLUTION PROBLEMS

Air pollution is an important environmental problem. Mathematical models can successfully be used to study the air pollution in different areas and to make decisions concerning its reduction. The most important parts of an air pollution model are the advection part (that describes the transport due to the wind) and the chemical part (where the chemical transformations during the transport are handled). It is worthwhile to test the performance of the model by testing modules that contain advection and chemistry. Such modules were tested by the PVM program and the results will be reported in this paper.

Assume that the space domain is discretised on a (32×32) equidistant grid. Assume that a simplified chemical scheme containing 35 air pollutants (chemical species) is implemented in the model. Then the discretisation of the models leads to the solution (at each time-step) of two systems of ODE's. The number of equations in each of these two systems is $32 \times 32 \times 35 = 35840$. The typical number of time-steps is several hundreds; the number of steps actually used in this paper is 576.

This information is quite sufficient for our purposes in this paper. More detailed on the advection-chemistry modules can be found in **[4, 5]**. The whole air pollution model is described in **[2, 3]**.

3 PARALLEL TASKS AND REDUCTION OF THE COMMUNICATIONS

Consider the advection part of the module. This is a system of $32 \times 32 \times 35 = 35840$ (see the previous section). However, the advection process for each air pollutant can be carried out independently of the advection processes for the other air pollutants. This means that the number of independent tasks is equal to the number of air pollutants (in our particular case to 35). In other words, 35 systems of ODE's, each of them containing (32×32) equations, are to be solved at every time-step. In many cases this is quite satisfactory (the tasks are big and sufficiently many). If this is not the case then parallelism on a lower level could be achieved (the tasks become smaller, but much more when a device based on a parallelism on a lower level is applied).

Consider now the chemical part of the module. This part is also a system of $32 \times 32 \times 35 = 35840$ (see the previous section). The chemical reactions that take place at a given grid-point depends only on the air pollutants (the chemical species) at this point. This shows that the big system is in fact split into 32×32 small systems (each of them containing 32 equations). Thus, the chemical problem contains as many parallel tasks as the number of grid-points (in our case $32 \times 32 = 1024$). In most of cases the number of parallel tasks is quite sufficient. In fact, in many modern air pollution models one uses finer grids (96×96 grids being often used at present). Therefore, in the real cases there will be even more parallel tasks.

Consider the arrays that are used in the advection part. They have dimensions (1024,35). This means that the different processors receive columns of these arrays. Consider the arrays used in the chemical part. These arrays have dimensions (35,1024). The different processors receive columns of this arrays. This short description of the arrays used indicate that at each step the data that is used in the advection part and has to be further used in the chemical part have to be reshaped (the arrays have to be transposed). Also the data used in the chemical part that have to be further used (at the next time-step) in the advection part have to be reshaped (transposed). It is quite clear that one should try to minimize these transpositions of arrays and sending data to the different processors. A careful examination of the data needed in the different parts show that in fact only one array (the array containing the concentrations of the air pollutants) must be reshaped on the main processor and distributed to the other processors (accepting some a small amount of calculations that have to be replicated on all processors). In this way the communications are reduced very considerably.

4 NUMERICAL RESULTS

Tests with the PVM version of the air pollution module have been carried out on four architectures: KSR, a cluster of HP workstations, the CONVEX SPP and CRAY T3D computers.

The KSR Computer was used (in this case) for testing our PVM version of the Air Pollution program. We do not present, here, any KSR performance results.

A cluster of HP735 workstations has been used in this experiment. The results are given in Table 1.

Nr of proc	Chemistry only			Advection & chemistry		
	Total time	Calc time	Trans time	Total time	Calc time	Trans time
1	**720**	**720**		**1203**	**1203**	
5	**259(2.8)**	**153**	**106**	**475(2.5)**	**251**	**224**
7	**217(3.3)**	**108**	**109**	**421(2.9)**	**181**	**240**

Table 1: **Computing times (in seconds) on the cluster of HP workstations using the PVM system. The speed-up achieved when more than one processor is in use is given in brackets (only for the total cpu time).**

A CONVEX SPP1000 computer has been used in the runs. The results are given in Table 2.

Nr of proc	Chemistry only			Advection & chemistry		
	Total time	Calc time	Trans time	Total time	Calc time	Trans time
1	598	598		1069	1069	
5	203(2.9)	121	82	384(2.8)	220	164
7	170(3.5)	85	85	334(3.2)	157	177

Table 2: **Computing times (in seconds) on Convex SPP1000 using the PVM system. The speed-up achieved when more than one processor is in use is given in brackets (only for the total cpu time).**

Some preliminary experiments have been carried out on a CRAY T3D massively parallel computer. Only the chemical submodel was run on this computer. The results are given in table 3. Both the computing time and the communication time are taken into account. In this case, computations for two grids, 32^2 and 64^2 were carried out.

Nr of proc	32^2 grid		64^2 grid	
	Total time	speed up	Total time	speed up
16	14		60	
32	7	2	32	2
64	4	4	17	4

Table 3: **Computing times and speed-ups obtained in runs with different number of processors, for two grids on a the CRAY T3D (only chemistry is used).**

References

[1] **A. Geist, A. Beguelin, J. Dongarra, W. Jiang, R. Manchek and V. Sunderam:** *"PVM 3 user's guide and reference manual"*. Report No. ORNL/TM-12187. Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA, **1994**.

[2] **Z. Zlatev, J. Christensen and A. Eliassen:** *"Studying high ozone concentrations by using the Danish Eulerian Model"*. Atmospheric Environment, **27A(1993)**, 845-865.

[3] **Z. Zlatev, J. Christensen and Ø. Hov:** *"An Eulerian air pollution model for Europe with nonlinear chemistry"*. J. Atmos. Chem., **15(1992)**, 1-37.

[4] **Z. Zlatev, J. Waśniewski, M. Venugopal and J. Moth:** *"Optimizing Air Pollution Models on Two Alliant Computers"*. In: **"PARALLEL COMPUTATION"** (A. E. Fincham and B. Ford, eds.), pp. 115-133. Clarendon Press, Oxford, **1993**.

[5] **Z. Zlatev & J. Waśniewski:** *"Running air pollution models on the Connection Machine"*. Mathematical and Computer Modeling; to appear.