ELSEVIER

Parallel Computing 25 (1999) 2015–2037

PARALLEL COMPUTING

www.elsevier.com/locate/parco

Industrial simulation on parallel computers

Clemens-August Thole*, Klaus Stüben

Institute for Algorithms and Scientific Computing (SCAI), GMD – German National Research Center for Information Technology, Schloss Birlinghoven, D-53754 Sankt Augustin, Germany

Abstract

Parallel computers have demonstrated their principle suitability for numerical simulation during the eighties and early nineties. In particular, they were able to provide a cost-effective means of achieving high performance computing (HPC) power. Even so, there was only a limited impact of this technology on industrial computing. In order to foster the take-up of this technology by industrial users, the European Commission launched a number of projects as part of the Esprit programme to parallelize commercial application programs, to demonstrate, document and disseminate the benefits of parallel architectures, and to explore the potential of parallel simulation in new application areas. Large-scale technology transfer initiatives such as Europort,[1] Europort-D and Preparatory Support and Transfer Programme (PST) aimed at helping the industry in Europe to exploit the benefits of HPC, based on parallel computing, thus increasing their competitiveness. This paper gives a review on major activities and highlights their impact on industry by means of some selected examples.

Keywords: Fluid dynamics; Structural mechanics; Parallel computing; Benchmarks; Industrial simulation

1. Overview

The foundation of the journal 'Parallel Computing' in 1984 marks a milestone in the development of parallel architectures, system software and applications: Parallel computing has become a major research topic for the academic community. 1984 marks also the start of several important projects in Europe aiming at the development of parallel hardware architectures such as the German national SUPRENUM project and several European projects focusing on the INMOS Transputer.

* Corresponding author.
E-mail address: thole@gmd.de (C.-A. Thole)

[1] Europort is a registered trademark of Syntel B.V. of Alphen a/d Rijn, The Netherlands. The Europort *initiative* is unrelated to Syntel or any of its software products.

0167-8191/99/$ - see front matter
PII: S0167-8191(99)00065-4

Finally, several research groups at universities were established at this time with a special focus on parallel architectures. However, it still took more than one decade for this technology to become usable by the industrial community.

1.1. Situation in 1993

A first major milestone in using parallel computing in a production environment was in 1993 when the central model of the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) became available in parallel (see Section 2). The parallelization was so successful that the ECMWF started to use this code for daily production runs in 1994, first on a 16-node CRAY-C916 and, since 1996, on a 48-node Fujitsu VPP700. This was the first use of a parallel high performance architecture for parallel simulation in production by a non-academic organization.

In industry, however, the new technology was not yet used. Fig. 1 compares the usage of parallel architectures in 1993 and 1998 by industrial companies in Europe according to the TOP500 list. In 1993 all machines included in this list were used for simulation purposes. Nearly all of them were multiprocessor vector architectures with only a small number of nodes. However, these nodes were not used in parallel but only to increase the throughput of a large number of jobs. Major application areas were automotive, aerospace, oil and chemistry.

Nevertheless, the industrial requirements on numerical simulation were steadily increasing. The progress at BMW in Fig. 2, for instance, shows that in 1993 numerical simulation has become a mature technology at the company. Simulation started to be used in the development process and important design decisions were based on numerical results. Furthermore – rather than developing their own simulation tools – industry started to licence commercial simulation packages. Consequently, the availability of parallel versions of these simulation packages became a pre-requisite for the usage of parallel architectures by industrial companies.

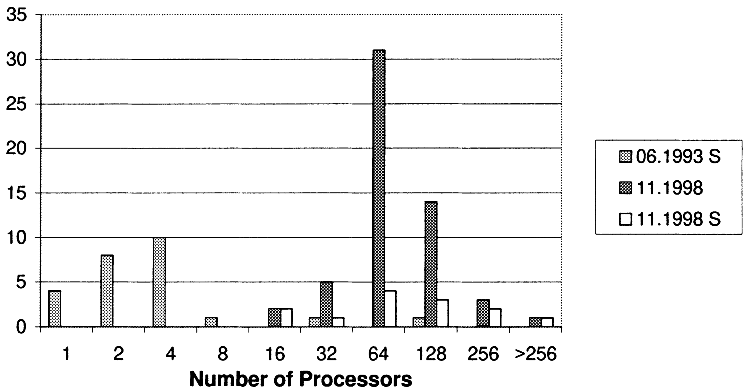

Fig. 1. Number of parallel computers used by industry according to the TOP500 list (S: exclusively used for simulation purposes).

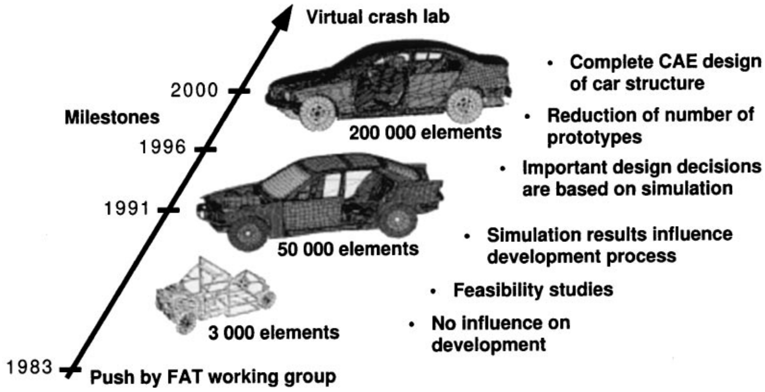

Fig. 2. Numerical crash simulation at BMW.

Against this background, the European Commission decided to start a number of Esprit projects with the goal of making commercial simulation codes available for parallel architectures. The major activities in this direction were Europort, Europort-D and the Preparatory Support and Transfer Programme (PST) initiative.

Europort started at the beginning of 1994, ended mid 1996 and involved some 120 partners europe-wide. The major focus was on porting 38 industrial simulation codes onto parallel architectures targeting various application areas. In 1996, commercial versions of all codes involved in Europort became available. Section 3 contains an overview of Europort and highlights some of its results. As a successor of Europort, the Esprit project *Europort-D* (1996–1997), investigated the impact of parallel computing mainly from a commercial point of view. Section 4 outlines some typical examples of the successful use of parallel computing in industrial companies, in particular, in automotive industry.

Technology transfer on a much larger scale is the aim of the ongoing Esprit PST which started in March 1997 and is running for three years. As part of this programme, the European Commission has established a network of 20 centres all across Europe to help and encourage new users to exploit parallel technology. These centres, called *Technology Transfer Nodes* (TTNs), are coordinating various activities which tackle important industrial applications. In contrast to Europort-D, the TTN-activities are on a much larger scale (altogether there are 175 activities with some 400 industrial partners) and the targets of the resulting experience are primarily small and medium sized enterprises (SMEs). The focus of all activities is on satisfying business requirements, rather than merely promoting technology. Section 5 presents a few characteristic examples.

1.2. Situation in 1999

Compared to 1993, the overall situation has changed dramatically. The chart in Fig. 1 depicts the usage of high performance architectures by industrial companies at the end of 1998. It clearly shows that parallel architectures have become established in European industry. Their proportion on the 500 world-wide fastest machines has increased substantially. The average number of processors is 64. However, a comparison of the proportion of machines which are exclusively used for simulation with the *total* number of machines, also shown in the chart, indicates that most of these machines are used for finance and data mining applications. Only a relatively small fraction is really used for simulation (in the industrial sectors automotive, aerospace, oil and chemistry). Compared to the true development, however, this result gives only an incomplete picture.

The pyramid in Fig. 3 illustrates the trend in using parallel computers for simulation in more detail, based on the situation in automotive industry. The top of the pyramid shows crash simulation, which is currently the most demanding application area in terms of computing requirements. Indeed, most powerful machines are used in the areas of crash and safety simulation. However, this area is very specialized and only a relatively small group of people use this simulation type. The next level of the pyramid refers to computational fluid dynamics (CFD), noise and vibration analysis and virtual reality as application areas. Today, such applications are typically performed on parallel machines which are dedicated to individual simulation groups. For example, in any automotive company, CFD is naturally applied by several groups and, therefore, moderately parallel machines are installed at many places inside the company. However, all these machines do not appear in the TOP500 list and the enormous increase in using parallel computers in automotive industry (and at the many sites of their suppliers) is not reflected in the TOP500 statistics.

One can summarize that, since the early nineties, a vast amount of commercial application codes have migrated from central computing resources to the most suitable platforms. In CFD, for example, all leading simulation codes are now available in parallel. Today, parallel computing is used wherever it turned out to be beneficial. In particular, multiprocessor machines are typically used in parallel and

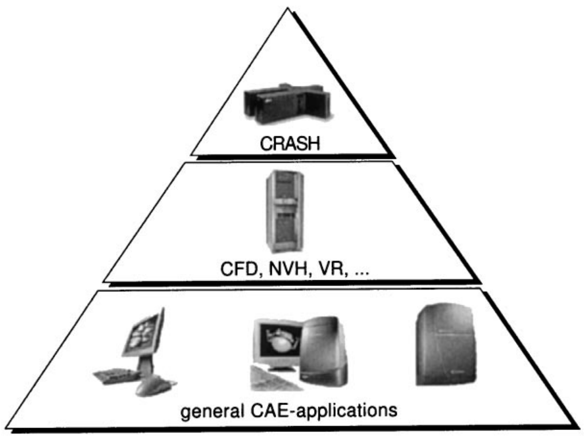

Fig. 3. Usage of parallel architectures in automotive industry [8].

not just to increase throughput. This is a direct consequence of the European projects. In the following sections, we describe the individual steps towards the industrial use of parallel computing in some more detail.

2. A milestone in porting production codes

Since the early nineties there was a discussion about whether or not it makes sense to port very complex production codes to parallel computer architectures. It was not at all clear whether the principal advantages of parallel computing would also be exploitable by simply porting *existing* sequential software and whether the amount of work needed for such a porting would finally pay off. Indeed, porting existing software requires many compromises which may cause unpredictable losses of parallel efficiency.

A pilot project, the parallelization of the production code of the ECMWF, started in 1992. In cooperation with the ECMWF, the central model of the IFS was parallelized by the Institute for Algorithms and Scientific Computing (SCAI) of GMD [1]. The IFS is a very complex simulation code. Starting from measured input data, the predictive variables wind, temperature, humidity and pressure are calculated. At the time being, the numerical model used a 3D computational grid with more than 4 million grid points. All unknowns are computed per time step in each grid point; for a ten-day forecast about one thousand time steps need to be performed leading to a total computational time of 6 h on the ECMWF's CRAY Y-MP8. In order to allow for substantial improvements in weather modelling and forecasting, a drastic increase of computer power was required.

The basic solution method in the IFS is the *spectral transform technique* using triangular truncation. In order to exploit the fact that the spherical harmonics are eigenfunctions of an essential part of the underlying operator, some parts of the calculation are performed in spectral space. Altogether, three different function spaces are involved in the calculations: *grid point space*, *Fourier space* and *spectral space*.

The idea of the parallelization approach, the *data transposition strategy* [1], was to re-distribute the complete data to the processes at various stages of the algorithm such that the arithmetic computations between two consecutive data transpositions could be performed without any further interprocessor communication. This approach seemed feasible since there were only data dependencies within one coordinate direction, this direction being different within the main algorithmic components. Thus, a parallelization within the remaining two dimensions appeared to be suitable for massively parallel systems with a thousand or even more processors. For the IFS, the transposition strategy can be detailed as illustrated in Fig. 4. In grid point space, the data partitioning is over latitudes, whereas during the calculations in spectral and in Fourier space, the data are distributed with respect to the zonal wave numbers. The switch between these two different partitionings is performed in Fourier space. This means that the Fourier coefficients are re-distributed twice in each time step.

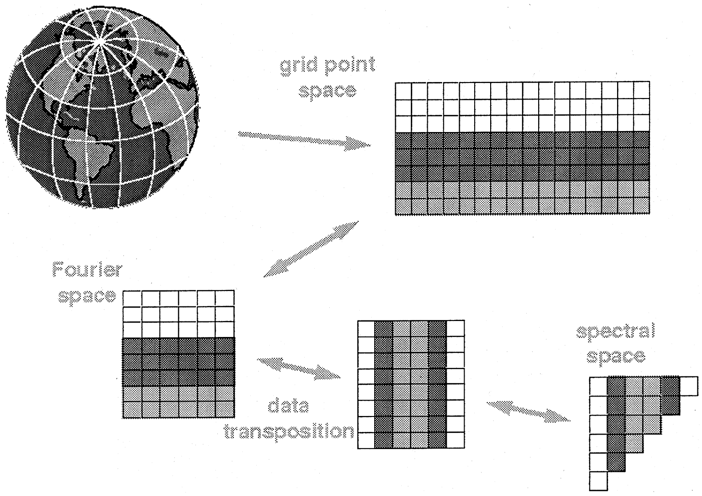

Fig. 4. Parallelization strategy used in migrating the ISF.

Based on this approach, the major part of the porting could be realized within a time scale of less than one year (three scientists). The resulting parallel version was highly scalable and indeed very efficient for hundreds of processors. In the meantime, the parallel weather prediction code (which has been further developed by the ECMWF) is in daily production use on a large parallel VPP700 machine. Even more important, however, is the fact that the results of the project have changed the attitude towards parallel computing in European centres for weather and climate prediction, where parallel systems are now fully accepted (which was not at all the case before). Moreover, the success of this project had significant influence on the decision to start a large-scale porting project with focus on commercial codes used in industry, Europort.

3. Porting of industrial simulation codes (Europort)

Whereas scientists from academia are used to develop their own codes, industrial companies typically use commercially available simulation software. In most cases, such software packages are owned by small companies, and these companies did not see a substantial market perspective for parallel versions of their codes. Moreover, the parallelization of a very complex 10–20 yr old code is a major and risky step for a small company.

In order to overcome this deadlock situation, the European Commission launched the Europort initiative with the goal of parallelizing 38 commercial simulation codes. Table 1 contains a list of these codes and corresponding application areas. The participation of well-known companies such as ABB, Aerospatiale, AGIP, Audi, Bayer, BMW, British Aerospace, CASA, debis, Det Norske Veritas, Dornier, EDF,

Table 1
Application codes ported in Europort

Application areas	Codes
Fluid dynamics	CEL3GR, CFX 4, ESAUNA, NSMB, N3S, PHOENICS, POLYFLOW, STAR-CD
Structural mechanics	FORGE3, LS-DYNA3D, MSC/NASTRAN, PAM-CRASH, PERMAS, SAMCEF
Animation	ANIMO
Chemistry	ADF, GAMESS-UK, GROMOS, MNDO, TURBOMOLE, VAMP
Drug design	MAXHOM
Oil & gas	ECLIPSE 100, FEDEM, FRONTSIM, MUSIC, SEISMIC, SESAM, SWAN
Radiotherapy	RAPT
Bearing simulation	BEARS
Earth observation	CAESAR
Telecommunication	ERLANG
Traffic flow	DSD
Electro magnetic	EMA3D & FAM, TOSCA, SCALA & ELEKTRA

Ericsson, Fiat, Ford, ICI, Mercedes-Benz, Merck, Philips, Rolls-Royce, Saab, SNECMA, Solvay, Statoil, Unilever, Volvo and many others shows the strong interest of industrial users.

The most important result of Europort was certainly the *availability* of parallel versions of the simulation codes already in 1996. The focus of a *systematic benchmarking* was to compare all parallel codes with their sequential analogues (which is of highest interest to an industrial user who is working with a particular code on a daily basis). A typical result was that 3–6 nodes of a parallel IBM SP-2 are sufficient to achieve the same performance as a single headed CRAY-YMP, see Fig. 5. (The CRAY-YMP was the standard system used in industry at the start of Europort.)

With respect to *scalability* (see Fig. 6), in general, structural mechanics and fluid mechanics applications behave differently. For most fluid mechanics applications,

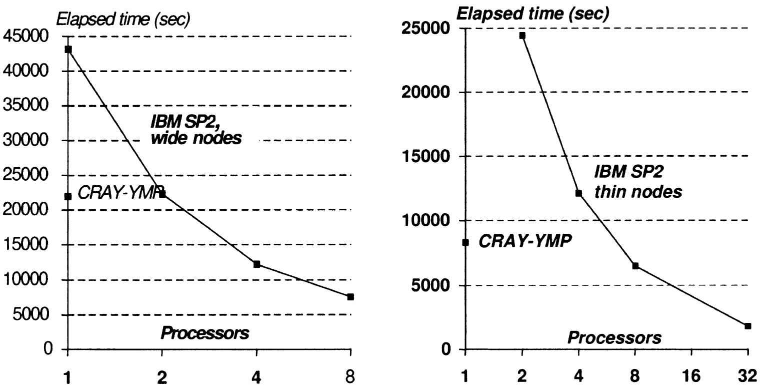

Fig. 5. (a) Sheet metal forming (LS-DYNA3D). (b) Supersonic flow (CEL3GR).

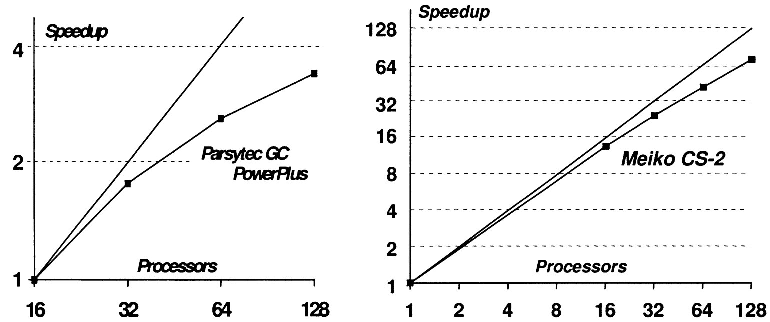

Fig. 6. (a) Crash simulation (PAM-CRASH). (b) Incompressible flow (N3S).

high scalability can be achieved. (Typical applications in industry need at least several hundred thousand elements. About 10,000 elements per node are sufficient to achieve good efficiency.) For typical structural mechanics codes, however, often only up to 16 processors can be used efficiently. In order to use more processors, algorithmic changes are necessary: The *direct* sparse matrix solvers, which are still used today in most cases, have to be replaced by scalable, efficient and robust iterative solvers. (Exceptions here are the explicit metal forming and crash codes which scale well if the contact areas are small.)

At the beginning of Europort, it was decided to base all portings on a standard message passing interface. While the original idea was to ensure *portability* for as many architectures as possible (including shared memory (SM) systems and clusters of workstations), this decision turned out to have additional advantages due to the fact that message passing codes naturally exploit algorithmical locality. The typical benefit can clearly be seen in Fig. 7 where, for the SM architecture SGI

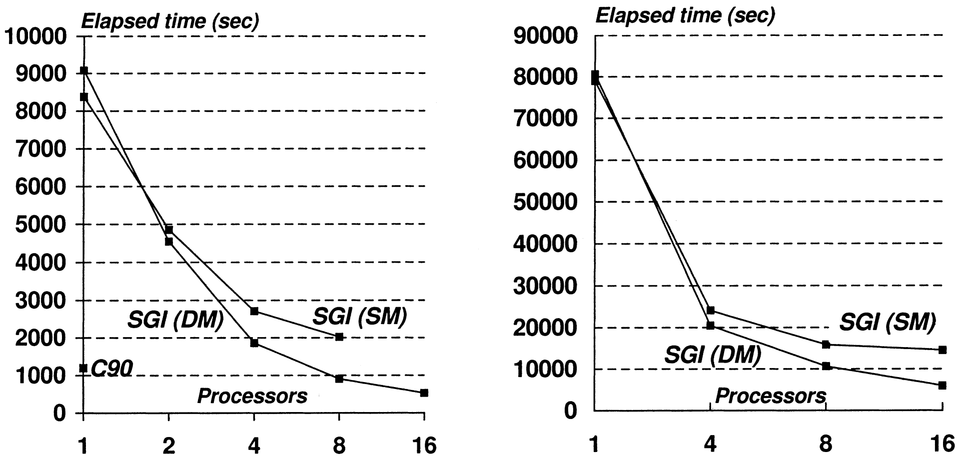

Fig. 7. (a) Flow in a cooling system (STAR CD). (b) Crash simulation (PAM CRASH).

PowerChallenge, distributed memory (“DM”) code versions are compared with their native “SM” counterpart for two largely different codes, namely, the structural analysis code PAM-CRASH and the CFD code STAR-CD. For both codes, on eight nodes, the message-passing version is about twice as fast as the SM version. Moreover, the message passing versions scale well beyond eight nodes.

This short overview can only give some impressions about typical results obtained in Europort. Regarding more results and, in particular, information on typical parallelization strategies used, we refer to [2–4] and the references given therein. Further more detailed articles are found in [7].

Besides the technical results, from a more global point of view, some other aspects need to be mentioned. Europort has demonstrated that it is possible to port – within a limited time frame and with limited resources – large commercial codes in a pragmatic way, but still very efficiently, to parallel platforms. Of course, the inevitable constraints given by parallelizing existing codes conflict with the goal of obtaining the highest parallel performance in a computer science sense. Specifically developed, genuinely parallel codes would certainly perform even better; their development, however, would have taken much more time and manpower. Moreover, Europort acted as an important catalyser for further industry-focused developments. For instance, it fostered the porting of most other European and non-European codes. Moreover, the US funding agency ARPA has launched two projects similar to Europort.

4. Demonstrating industrial benefits (Europort-D)

Europort has been finished mid-1996. By then, many industrial organizations have been made aware of the usability of parallel computing technology and its cost-effectiveness as a means to obtain high performance computing (HPC) performance. More important in convincing industry, though, are real *business benefits*. This gap was closed by the follow-up project Europort-D (June 1996–October 1997). This project consisted of 10 demonstrator sub-projects targeting the application areas *aerodynamics for car design, car crash and safety simulation, cartoon animation production, drug design, forging of machine components, polymer processing, satellite image processing, fire and safety analysis, turbomachinery design* and *vehicle electromagnetic testing*. For each area, new end-users demonstrated the benefit parallel computing technology can provide for their industry. It was expected that the successes enjoyed by these companies would stimulate other organizations to investigate the applicability of the technology also to their business.

Central to the work in Europort-D was the identification of critical issues relevant to each different industrial process and how parallel computing could help. Typical benefits, each of which can be translated into commercial advantage, are:

- *Reduced time for process design*, allowing more design concepts to be examined or the system to be simulated within a timescale critical to the particular process.
- *More accurate process simulation*, providing more confidence in the results, less material wastage, higher quality and safer products.

- *Increased capability*, providing a mechanism for larger or more complex systems to be examined and making new products possible.

In the following sections, we briefly present four selected "business cases" demonstrating the *commercial* relevance of parallel technology. For a more complete description, we refer to [5]. One should emphasize that by far most of the companies involved in Europort-D had no experience with parallel computing before the project. As a result of their participation, all of them experienced substantial benefits and are now strongly committed to the technology.

4.1. Computational fluid dynamics

In most industries employing CFD simulation, there is an ever increasing need to reduce development cycles, minimize resource usage, fulfil global regulations, increase safety, optimize product quality, satisfy customers wishes, etc. Experiments are very expensive, time-consuming, and moreover are often so sophisticated that further design improvements through experiments – e.g. wind tunnel tests in the car industry – are impractical. On the other hand, simulations are highly CPU and memory intensive. Using standard computer technology, many simplifications are being introduced in order to complete numerical simulations in an acceptable time (simplified physical models or geometries, coarse meshes, etc). Although corresponding simulations – often very crude – are still helpful in supporting a design process, a much higher accuracy is required to reach the goal of generally replacing physical experiments by simulation.

Traditional methods using wind tunnel experiments have reached limits such that discovering information about airflow (apart from drag and lift) is extremely difficult and time-consuming. Mercedes-Benz (MB) is convinced that the only realistic way to provide the aerodynamicist with the broad set of data required in order to 'fine-tune' the shape of the car is to exploit the cost-effectiveness of HPC, based on distributed computing, with its ability to scale to virtually unlimited memory capacity. On the other hand, aerodynamicists have always been sceptical regarding the accuracy of CFD simulations, in particular concerning external flows over complete car bodies.

Within Europort-D, MB was able to show that it is possible to compute all physically relevant quantities (drag, lift, airflow separation and re-attachment, pressure at key locations and velocity in the car wake) to the same level of accuracy as typical for wind tunnel tests, i.e. within 10%. The demonstration case was a 1-1 prototype E-Class model, including all exterior details present on the car body (see Fig. 8). The simulation required a mesh of the order of 10 million cells, probably the largest case ever run in automotive industry. Just the sheer size of the model (memory requirement 6 GBytes) required distributed computing.

Only eight processors of an IBM SP2 were available to MB at that time for performing this demanding simulation requiring 21 days of elapsed time. Although this is much too long for practical application, there is evidence from the Europort benchmarking that, for such large problems, the underlying code, STAR-CD, scales

Fig. 8. Airflow visualization around a Mercedes-Benz production vehicle.

nearly linearly to well over 100 processors, enough to obtain the same results in a realistic timescale of between 1 and 2 days by upgrading the machine to 128 processors. As a result of this investigation, MB has purchased departmental parallel machines exclusively dedicated to parallel CFD simulations: a 24-node SP2, a 32-node SGI Origin and several smaller SGI Origins.

To be able to use exterior simulations routinely for virtually all types of optimization of the car body, eliminating the need for most experiments, a further reduction of the turnaround time to the order of hours is needed. Although this is currently not yet practicable with existing computer hardware, it seems realistic to achieve this within the next five years, say. To exploit this, MB will upgrade their departmental computers in terms of number of nodes as well as node performance during the next years.

4.2. Crash and safety

Prototype crash tests can no longer deliver the information required in a timely and cost-effective way; many important details cannot be achieved at all through experiments. In particular, parametric optimizations can only be done by computer simulation because scatter in physical testing does not allow trends to be clearly identified. It is therefore not surprising that industry is increasingly relying on simulation (here based on the PAM-SAFE code) for the design of both the car and its safety equipment. However, the explosion in modelling requirements for vehicles, equipment, airbags and dummies as well as stricter safety requirements are dramatically increasing the computational demands requiring cost-effective high-performance computing. This has become a serious problem both for big car manufacturers and for the dozens of small and SMEs serving the car industry.

Two small suppliers for the car industry, PARS and TRW, have joined Europort-D to investigate the possibility of increasing their computing power by parallel computing. Their computing requirement is tremendous. At TRW, several side-impact airbags are developed per year, each requiring up to 100 different simulation runs (see Fig. 9(b)). Even a reduction of computing time by just a factor of 3–4 would cause a dramatic gain when one considers that each individual run takes between 12 h and 3 days on a single workstation. Similarly, at PARS, up to 180 simulations need to be performed per year for the design of steering wheels, each taking between 3 and 5 days on a fast workstation. The necessary computing times are incompatible with the production time schedule resulting in re-design cost of more than one million DM per year. Thus, for both companies, it is of utmost importance to reduce the computing time.

Being newcomers to parallel computing, both of them were mostly interested in exploiting their available workstations as a cluster. This way, a reduction of computing times by a factor of 3–4 was achieved (using Fast Ethernet). This already allowed the drastic improvement of being able to detect faulty designs and design imperfections before reaching the experimental stage. Both companies are highly convinced about the advantages of parallel computing. They now regard their workstation cluster just as an entry system and have decided to purchase small dedicated systems with 4–8 processors.

At BMW, many thousands of highly complex simulations have to be performed per year. Since, within the next 4–6 yr, vehicles will need to be entirely designed by computers, the computational demands will dramatically increase further. BMW already has half a dozen parallel computers installed with a total of 150 processors. However, these are not able to match the anticipated requirements. Therefore, BMW has investigated the possibility of using their workstations to obtain additional computer power. The Europort-D results indicate that 10 powerful workstations, connected by a 100 MBit network, are able to perform the analysis of a typical 100,000 element fully equipped car model overnight (see Fig. 9(a)). Larger models with 200,000 elements can be run over the weekend.

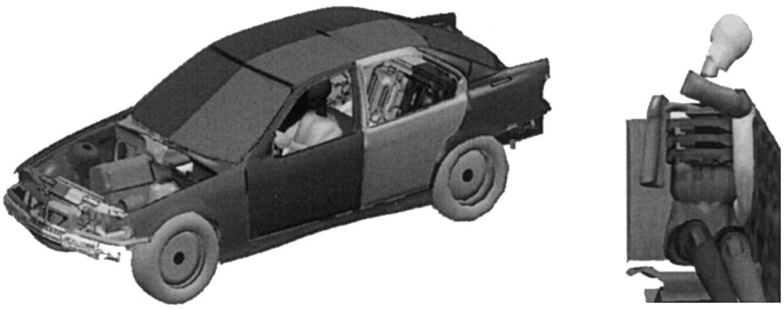

Fig. 9. (a) Fully equipped car model used at BMW for crash and safety simulations. (b) Foam door model, airbag and deformable Eurosid dummy used at TRW.

4.3. Forging

Industrial forging processes require forces equivalent to thousands of tons, very powerful presses and strong dies. Computer simulation of forging allows the optimization of the shape and properties of dies before they are actually produced. In this way, costly re-designs – caused by, for instance, incomplete filling of the dies, folding inside the material or unacceptable die wear – can be avoided. Further cost reduction can be achieved by minimizing material wastage (flash), optimizing the properties of the press and, finally, optimizing the quality of the forged parts (homogeneity, grain size, etc.).

The numerical simulation of complex industrial forging processes is a fairly young discipline. It is only since the early 1990s that it has been possible, using software tools such as FORGE3, to simulate 3D forging processes. Unfortunately, in most cases the simulation time was simply too high for simulation to be considered for routine use. The design of forging processes is most commonly done by engineers close to manufacturing who neither have the money to purchase nor the experience to maintain large powerful computers; their normal computing equipment are workstations. However, on single workstations, the simulation of complex forging processes such as the forging of steering knuckles (see Fig. 10) takes over six weeks – unacceptably high to be practicable. The maximum turnaround time for a simulation is given by the time between ordering the dies and their final production which, for such complex parts, is around two weeks. Due to the scalability properties of the parallel FORGE3 code, this threshold can now easily be crossed by using relatively small parallel systems which are affordable even by small companies.

The steering knuckle is just one example of the complex industrial parts (e.g. lower arms, blades, crankshafts, connectors, ingots) for which cost-effective parallel technology now allows numerical forging simulations in design and optimization where it was practically impossible before. In all these cases, parallel technology helped push numerical simulation across the threshold where it can increasingly replace costly experiments, reduce production time, increase lifetime of the dies and improve the quality of the final products. Many companies are already taking advantage of this possibility. This is directly reflected in the strong increase of sales of the FORGE3 simulation tool as demonstrated in Fig. 11. While the number of sales

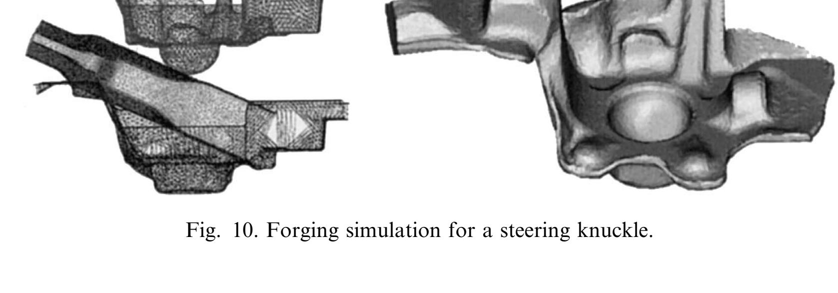

Fig. 10. Forging simulation for a steering knuckle.

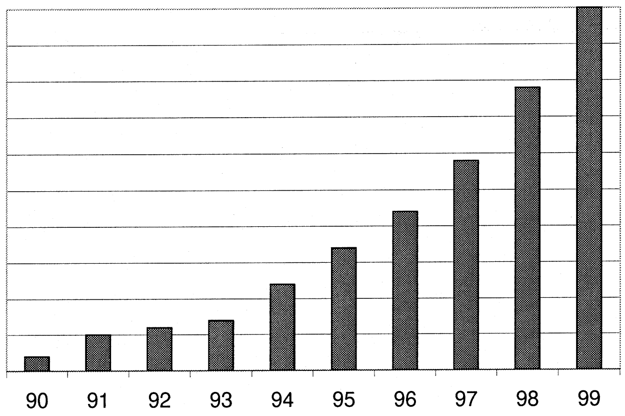

Fig. 11. Sales increase of the simulation code FORGE3 (estimation for 1999).

was stagnating before 1994, it is strongly increasing since the first parallel code version became available (parallelized as part of Europort).

It is estimated that 30% of all forging companies world-wide can directly benefit from this progress in simulation. The French Forging Association conservatively estimates potential savings of over 20 MECU per year for its national industry alone. This takes savings from only reduced material wastage, increased die life and lower prototyping cost into account. Secondary effects (such as higher product quality) are difficult to estimate but may be substantial. Companies like SNECMA and PSA (Peugeot/Citroën) conservatively estimate savings of hundreds of KECU per year due to simulation.

4.4. Drug design

A typical design cycle for developing a new drug may cost in the order of 300 MECU and last up to 10 yr, many months or even a few years of which are required just for the initial design phase, the 'discovery stage'. During this stage, a large number of experiments need to be performed which is not only very expensive and time consuming but also cannot answer all relevant questions. The only alternative, computer simulation of the dynamics of molecules (MD simulation), has, until recently, been perceived by industrial research managers as being of limited value in the design of new drugs. Indeed, MD simulations played only a marginal and sometimes even misleading role. The main reason for this was simple: operational limits on computational time, memory usage, etc. meant that oversimplified models were used. However, reliable data can only be obtained when all relevant biological and chemical interdependencies are taken into account, leading to an increase in complexity by at least a factor of 10. Moreover, in order to make really meaningful analyses, the number of simulation time steps must be increased substantially, requiring trajectories of up to 10 times longer than were commonly performed earlier.

HPC is required to cross the threshold where MD simulation becomes a valuable tool for industry. However, in most pharmaceutical companies, HPC is something very new and supercomputers were simply not available to industrial researchers. With the arrival of affordable high-performance multiprocessor machines and corresponding developments of parallel software, it now becomes possible for industrial researchers to undertake more realistic calculations that were previously out of reach.

Scientists at Novo Nordisk, a large Danish pharmaceutical company, are convinced that this new capability will dramatically change the acceptance of MD simulation as a tool in the design of new ligands (candidate compounds for a new drug). During Europort-D they could, for the first time, study the dynamics of the complex molecular interactions critical for recognition of ligands by their target proteins (see Fig. 12). MD simulations with a turnround of a few days could be achieved even for systems consisting of tens of thousands of atoms (using the DM GROMOS code, parallelized within Europort, on their 18-processor SGI Challenge). Corresponding experiments would require many months of work and much of the information obtained from simulations, such as the details of dynamic behaviour of the bound ligand, can hardly be obtained experimentally at all.

As an alternative to experiments during the discovery stage, MD simulation can now be applied to proposed candidates to see if they have the right recognition properties as observed in their dynamic behaviour when bound to their target protein. Generally, any technology which can reduce the number of syntheses and tests generates savings which, over all development projects, can be in the order of many MECU. The Europort-D results made it evident that MD simulation has the potential for a substantial contribution. Whether and to which extent this is really achievable, can only be seen in the future.

Much more important than potential cost savings during the design phase, however, is the business impact caused by a shortening of the design cycle. As alluded to above, MD simulation offers a new way of testing candidate ligands as to their suitability for further development into drugs. This may create time savings even at the early stages of the design cycle, but perhaps more significantly – by eliminating unwanted characteristics at an early stage – time savings in the later, more costly, developments stages will inevitably accrue.

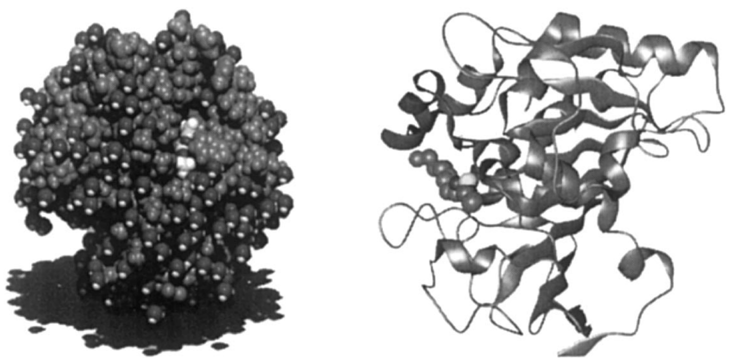

Fig. 12. Ligands binding to a target molecule to modify the behaviour of the protein.

In the face of strong competition, however, even bringing forward a product release date by just a few months, can have a dramatic effect on revenue. Scientists at Novo Nordik believe that with superior design strategies, including the use of parallel computing MD simulations as performed in Europort-D, they can contribute significantly to the selection of a better quality product and even lead to an accelerated launch. The resulting advantages ought to be worth millions of ECU to the enterprise.

5. Parallel computing for small enterprises and other application areas

As in Europort-D, the primary goal of the ongoing PST initiative, introduced in Section 1, is to stimulate "replication effects", that is, to make more industries adopt the technology and enhance their competitiveness. However, the PST initiative operates on a much larger scale, involving some 400 industrial partners. Most importantly, the target of this initiative is primarily small and medium enterprises (SMEs) and/or new application areas. Examples include *toy production, production planning, film restoration, prosthetics, radiography image processing, visual inspection of manufacturing processes,* and many others. In the following, three exemplary projects are outlined each representing one of the following characteristic groups of activities (regarding some further non-traditional CFD applications, see also [6]):

- non-traditional applications of existing parallel simulation codes;
- parallelization of application codes from non-traditional application areas;
- traditional parallel applications for small and medium enterprises.

Before detailing these three cases, it should be stressed that, in an industrial context, HPC is to be understood as "HPC at large" which means that actual performance requirements and improvements have to be seen relative to the target industrial area and its computational history. While for CFD applications in the automotive or aerospace industry, HPC often really means attaining the highest possible performance in terms of Mflops, this is not the case in other areas where often the efficient exploitation of small multiprocessor systems or even networked PCs can yield a substantial benefit for the business. In particular, for most SMEs parallel computing is something very new. For them, clusters of in-house workstations, used during idle night hours or at weekends, provide highly interesting entry-level systems. Moreover, such systems also relieve those memory limitations which often make the use of traditional computers for complex industrial applications impossible. Therefore clusters of workstations are attractive even in addition to small parallel multiprocessor systems in order to increase the total available computing capacity and to deal with applications demanding the most memory.

5.1. Non-traditional applications of existing simulation codes

PAM-CRASH and PAM-SAFE are commercial simulation codes, which have been parallelized as part of the Europort initiative. Both simulation codes are closely

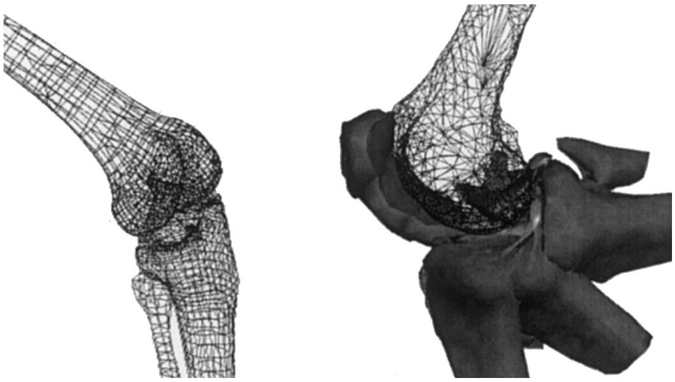

Fig. 13. 3D model of the human knee.

related to each other and are heavily used in automotive industry for crash-analysis and the prediction of occupant safety in the case of accidents (see Section 4.2). Essentially, these codes are able to simulate highly non-linear structural analysis phenomena involving much contact. In particular, such codes are also applicable to the computer modelling of human joints. However, for such new application areas, the physical models for the materials have to be adapted.

Computer modelling of human joints is a lengthy process which was generally restricted to 2D models with one of the limiting factors being the high computational requirements for large models with many contact surfaces. As part of the PST initiative, the University of Sheffield developed a 3D model, which moves and looks exactly like a real knee joint and simulates the bio-mechanical environment in more detail than ever before (see Fig. 13). It has recently been used to study different movements, impacts and the behaviour of a prototype meniscus implant. The results of the simulation were validated by laboratory mechanical testing in which data were collected and compared with simulation results. All output from the model followed the expected movement patterns. Under longitudinal impact loading the peak force transmission varied as expected with the progressive removal of soft tissues, proving that the model really behaved like a real knee joint.

In general, geometrically accurate models require high quality data to be collected by Magnetic Resonance Imaging (MRI) of knee specimens. This enables the reconstruction of the shapes and positions of all the important anatomical features. The PAM-SAFE code then enables the simulation and study of normal and abnormal knee movements. Virtual testing allows the consideration of many what-if scenarios before the expense of a real world implementation of an implant. These possibilities directly benefit many areas of medical and industrial development, for instance:

- *Crash simulation,* more accurate and more detailed information from crash test dummies with lower cost due to the ability of virtual testing.

- *Design of sports and training equipment*, allowing advanced insights into the interaction of products and the human knee joint.
- *Design and testing of orthopaedic implants*, predicting interaction and life expectancy of implants in this harsh environment.

5.2. Non-traditional application areas: Cartoon animation

Cartoon animation studios have always prided themselves on their craft traditions and human skill has been the most coveted commodity in the animation business. However, they are now finding that they too can make use of the mouse and keyboard, as well as the pencil. While many studios adopt computer technology for creative reasons – to achieve new visual effects or a more sophisticated image quality – for most, the decision to invest in computers is driven by the need to compete with imported material produced in the USA or Pacific Rim. For these studios, computer technology represents a way of bringing down production costs, through automation of many of the manual activities.

It was a European software package, Animo, developed by Cambridge Animation (UK), which introduced a combination of task-based modules interlinked by parallel processing capabilities allowing animators to make the most of digital technology and increasing the efficiency of production staff but without significantly changing the traditional animation skills or processes. The relevance and success of these developments, which have been performed within the Europort project, is evident in its takeup by studios worldwide, including Warner Brothers and Dreamworks (Steven Spielberg's studio). Partly due to its distributed facilities and parallel rendering capabilities, Animo has become the world's leading computer animation package (see Fig. 14).

While even the use of standard computer technology is fairly new to most animation studios, the availability of software with the capability of distributed parallel computing across networks of workstations has given rise to dramatic changes; substantially more efficient working practices and new in-house capabilities are now available:

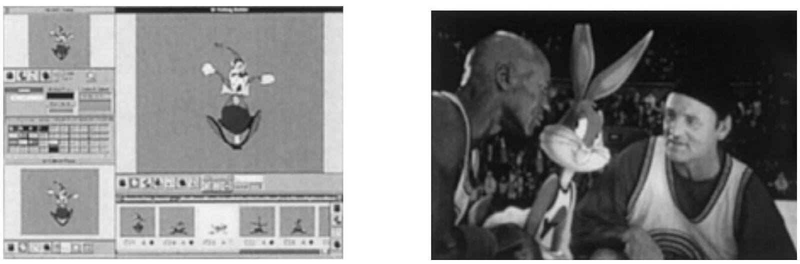

Fig. 14. Computerized cartoon animation (courtesy of Warner Brothers).

- The possibility of using computer networks to distribute work effectively between small European production companies, whilst being able to assure the quality and compatibility of the final products, is changing the manner in which co-production is undertaken between different studios long distances apart. This substantially increases the competitive situation of small European studios when compared to larger overseas competitors.
- The ability to perform rendering – the final production process which is traditionally very labour intensive and now very computationally intensive – in parallel on a network of workstations is enabling small teams to increase drastically their in-house capability and utilization of computational resources. Typically using just five workstations allows work which would have taken two days to complete to be run overnight on otherwise idle systems.

'Cartoon Producción', based in Valencia, is an animation studio leading the move towards digitally based co-production of cartoon films and TV series. The effects on Cartoon Producción's business have been quite dramatic. Within the animation market it is certainly true that in economic terms, bigger is usually better – or at least stronger – and the bigger studios are outside Europe. However by use of leading edge technology and by combining their skills and production capacity with other studios using the same technology, Cartoon Producción is ensuring its own future and helping the European animation community fight back against the giants of the industry.

5.3. Parallel computing for small enterprises

The need for simulation tasks at SMEs has increased dramatically during the last few years. As a result of down-sizing and lean production, many large production companies (in particular in the automotive sector) outsource the production of complete parts to their suppliers all of which are SMEs. Sometimes they even do not provide detailed designs of these products, but only functional specifications. The concrete design of the parts and the manufacturing process is up to the manufacturing company. Such a company typically uses high end workstations for the CAD design of their products. Structural analysis and computational fluid dynamics simulation using commercial packages is standard practice. Although, in principle, HPC platforms are necessary to perform such simulations, corresponding investments are beyond the means of these companies. Similar demands are stated by consulting companies which are specialized in engineering tasks and are either directly subcontracted by large production companies (like car manufacturers) or by their suppliers.

For all these SMEs, it is very attractive to use workstations outside of peak hours as simulation engines. Clustering their workstations will allow them to perform HPC at none or little extra cost, resulting in a substantial improvement of their simulation capabilities. However, if compared to homogenous parallel architectures, the usage of workstation clusters needs several additional features to be supported by the parallel simulation code:

- *Support for heterogeneous nodes.* In a workstation environment, computing resources are grouped together which, in general, have different performance characteristics and originate from different hardware vendors. This requires a simulation package to distribute the workload according to the performance characteristics of a selected set of workstations. For applications from CFD, for instance, the partitioning of the computational grid has to be performed according to the node performances and the respective subgrids have to be assigned to the related nodes. Consequently, at the stage of data partitioning, the performance characteristics of the target nodes have to be available to a simulation code and explicit control on the assignment of subgrids to processors must be possible.
- *Support for low performance interconnect.* Industrial companies are currently upgrading from standard to fast Ethernet (100 Mbit/s) networks. Although, for many applications, fast switched Ethernet interconnect networks are, in principle, fast enough to use efficiently a moderate number of workstations, a standard parallel code typically requires specific tuning and optimization.
- *Resource migration and restart protection against disk failures.* Workstation clusters allow other users to interfere with the actual parallel application by starting additional jobs or by just closing down a workstation. Support is needed for resource migration and restart protection against system and disk failures.
- *Resource scheduling and resource control.* If a workstation cluster contains more than a few workstations and is used by several users for multiple jobs at the same time, a resource management tool – such as LSF (from Platform Computing) and CODINE (from Genias) – is needed for the scheduling of jobs and the assignment of an appropriate share of nodes.

As part of the PST initiative, the parallel simulation codes PHOENICS and STAR-CD (CFD) and PERMAS (structural mechanics) have been optimized for workstation clusters and evaluated by industrial users. As an example, we present some typical results for STAR-CD which was enhanced in order to satisfy all needs of cluster computing as listed above. An interface to LSF allows the code to be used in combination with a resource management tool and therefore the information needed for data partitioning can be derived automatically.

The workstation cluster version of STAR-CD was evaluated by Behr, an automotive supplier in the field of engine cooling and climatization. For the CFD group at Behr, the simulation of HVAC (heating, ventilation and air conditioning) units is the most important task. In particular, the numerical prediction of pressure losses and temperature distributions inside the air distribution chamber is standard during a development process at Behr. Typical model sizes are 300,000–800,000 fluid cells resulting in sequential simulation times of 2–3 days.

For their simulations, Behr normally operates an eight processor SGI Origin 2000. In order to increase their computational capacity further, different workstation clusters out of a pool of 20 workstations (with different performance characteristics) were used for evaluation purposes. The test case was an HVAC unit with 493,900 fluid cells (porous media for filter, evaporator and heat exchanger, see Fig. 15). The

Fig. 15. HVAC unit investigated at Behr.

performance tests were performed for the first 100 iterations using up to six SGI workstations (Indigo2 and Octane with different clock rates) on a switched 100Base-T Ethernet.

It is difficult to evaluate the parallel performance on a heterogeneous workstation cluster. The standard measures, scalability and speed-up, are meaningless because the performance characteristics of the individual nodes used for an application have to be taken into account. (For the particular application at hand, the fastest workstation used was 3.64 times faster – in running STAR-CD – than the slowest one.) The *relative performance* of a node p_j of the cluster is therefore defined as the ratio of the elapsed time for the target application on the slowest node and the elapsed time on node p_j. The *aggregated relative performance* of a set of workstations is the sum of the relative performances of its nodes. This represents the computing power of a given set of workstations for a given application. Of course, different sets of workstations will, in general, have different aggregated relative performance.

Fig. 16 shows the elapsed time versus the aggregated relative performance of the HVAC test case for different sets of workstation clusters. Sets of workstation clusters with the same number of nodes are marked with the same symbol. The figure shows that six workstations can be used effectively and even an additional slow workstation still leads to a performance improvement. This result is of high relevance to Behr which now uses CFD simulations on workstation clusters on a regular basis in order to increase their available computer resources.

6. Final remarks

Although, as discussed in the introduction, parallel computing and parallel architectures have established themselves as a mature technology, this does not imply that all questions raised by parallelization have been solved by now. Various challenges remain to be addressed two of which we want to mention explicitly:

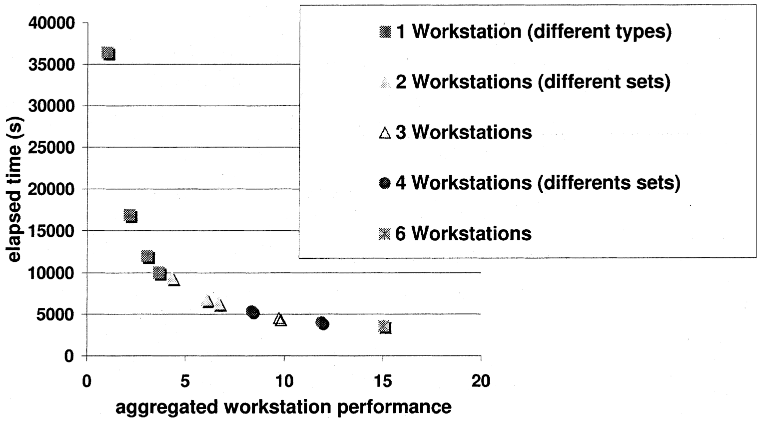

Fig. 16. Elapsed time vs. aggregated workstation performance for the HVAC test case.

- *Dynamic load balancing.* Many industrial applications are time dependent and combine, at each time step, different phases of computation. An example is crash simulation, where contact search/contact treatment, computation of forces and computation of new node positions have to be performed for about 100,000 time steps. Each of these phases relies heavily on the results of the previous phase, which requires, in principle, a general synchronization point. However, the phases have substantially different load distributions among the mesh points which even vary in time. Grid partitioning algorithms have to take this into account and may require a global load re-distribution after a certain number of time steps.
- *Linear sparse matrix solvers.* Although *direct* sparse matrix solvers – often used in structural mechanics codes – are very stable and robust, they have only a limited scalability on parallel computers. Iterative methods still need to be developed which are more suitable for parallelization and still provide a similar robustness.

References

[1] U. Gärtel, W. Joppich, A. Schüller, Parallelizing the ECMWFs weather forecast program: the 2D case, Parallel Computing 19 (1993) 1413–1425.

[2] H. Mierendorff, K. Stüben, C.A. Thole, O. Thomas, Europort-1: Porting industrial codes to parallel architectures, in: Proceedings of HPCN Europe 1995, Springer Lecture Notes in Computer Science 919, Springer, Berlin, 1995, pp. 806–812.

[3] K. Stüben, Industrial parallel computing with real codes, Parallel Computing 22 (1996) 725–737.

[4] K. Stüben, H. Mierendorff, C.A. Thole, O. Thomas, Europort: parallel CFD for industrial applications, in: E. Schiano, S. Periaux (Eds.), Parallel Computational Fluid Dynamics, Elsevier, The Netherlands, 1997, pp. 39–48.

[5] K. Stüben, Europort-D: commercial benefits of using parallel Technology, in: J. D'Hollander, T. Peters (Eds.), Advances in Parallel Computing 12, North-Holland, Amsterdam, 1998, pp. 61–78.

[6] K. Stüben, Industrial HPCN applications in fluid dynamics, in: Proceedings of the Fourth European Computational Fluid Dynamics Conference (ECCOMAS), Athens, vol. 2, 7–11 September 1998, Wiley, New York, pp. 540–545.

[7] A. Colbrook, I. Duff, A. Hey, K. Stüben, C.A. Thole (Eds.), Europort (special issue), Int. J. Supercomputer Applications and High Performance Computing, 11 (4) (1997).
[8] E. Schelke, Digital prototyping am Beispiel der Boxter und 911er – Entwicklung, in: S. Kuschfeld, Th. Ertl (Eds.), Synthese von Berechnung, Simulation und Visualisierung Technik und Kommunikation, Verlag GmbH, Berlin, Germany.