IN-32 TM
56186
26P

# DATA COMMUNICATION REQUIREMENTS FOR THE ADVANCED NAS NETWORK

*Eugene Levin, RIACS*

*C.K. Eaton and Bruce Young,*
*General Electric Western Systems,*
*San Jose, California 95134*

November, 1986

Research Institute for Advanced Computer Science
NASA Ames Research Center
Moffett Field, California 94035

RIACS TR 86.22

# RIACS

**Research Institute for Advanced Computer Science**

# DATA COMMUNICATION REQUIREMENTS
# FOR THE ADVANCED NAS NETWORK

E. Levin*
Research Institute for Advanced Computer Science
Moffett Field, California 94035

C. K. Eaton and Bruce Young
General Electric Western Systems
San Jose, California 95134

## ABSTRACT

*The goal of NASA's Numerical Aerodynamic Simulation (NAS) Program is to provide a powerful computational environment for advanced research and development in aeronautics and related disciplines. The present NAS system consists of a Cray 2 supercomputer connected by a data network to a large mass storage system, to sophisticated local graphics workstations and by remote communications to researchers throughout the United States. The program plan is to continue acquiring the most powerful supercomputers as they become available. In the 1987/1988 time period it is anticipated that a computer with 4 times the processing speed of a Cray 2 will be obtained and by 1990 an additional supercomputer with 16 times the speed of the Cray 2. This paper describes the implications of this 20-fold increase in processing power on the data communications requirements. The analysis was based on models of the projected workload and system architecture. The results are presented together with estimates of their sensitivity to assumptions inherent in the models.*

# DATA COMMUNICATION REQUIREMENTS
# FOR THE ADVANCED NAS NETWORK

**INTRODUCTION**: The Numerical Aerodynamic Simulation (NAS) Program was initiated by NASA to establish a national resource for advanced research and development in aeronautics and related disciplines. To achieve this goal the NAS Program is to, "act as the pathfinder in advanced, large-scale computer system capability through systematic incorporation of state-of-the-art improvements in computer hardware and software technologies". The justification, historical background, technical objectives and long term plans of the NAS Program have been presented previously [1, 2, 3].

The first major milestone of the NAS Program has now been achieved and the initial operating configuration of the NAS Processing System Network (NPSN) located at the NASA Ames Research Center in California is shown schematically in Figure 1. (See [4] for detailed specifications). The centerpiece is a Cray 2 supercomputer with 250 million (64-bit) words of memory and a *sustained* performance rated at 250 MFLOPS (250 Million FLoating-point OPerations per Second) as measured on a set of NAS benchmark tests for optimized large-scale computational aerodynamic application codes. This impressive capability is enabling researchers in the aerophysics field to address previously unsolved problems and to gain insight into complex aerodynamic phenomena. However it is only the beginning. In recognition of the on-going nature of the NAS program, the Cray 2 is designated HSP-1 (High Speed Processor 1).

The need for much more powerful processors can be seen from Figure 2 [1, 2, 3] which depicts the estimated speed and memory requirements for various levels of approximation to the governing fluid-dynamics equations for three levels of geometric complexity; an airfoil, a wing and a complete aircraft. Note, for example, that if viscosity effects are included by using the Reynolds-averaged Navier-Stokes equations, a three dimensional solution for a wing requires about 100 times the computing speed of a comparable inviscid solution and only now with the Cray 2 is it feasible to perform highly repetitive design optimization studies for such cases. Furthermore, if still more realistic large eddy effects are to be considered, a further factor of about 1000 is required for runs cf the same duration. Finally, Figure 2 indicates that (with 1985 algorithms), a single 15 minute run including large eddy effects for a complete aircraft would require computer speed in excess of $10^{12}$ floating point operations per second and an estimated random access memory of $10^{11}$ bytes!

In consideration of these computational needs, the NAS Program plan is to continue to acquire the most powerful supercomputers as they become available. In the 1987/1988 time period it is anticipated that a "one GigaFLOPS" computer (HSP-2) with 4 times the speed of a Cray 2 will be obtained and by 1990 an

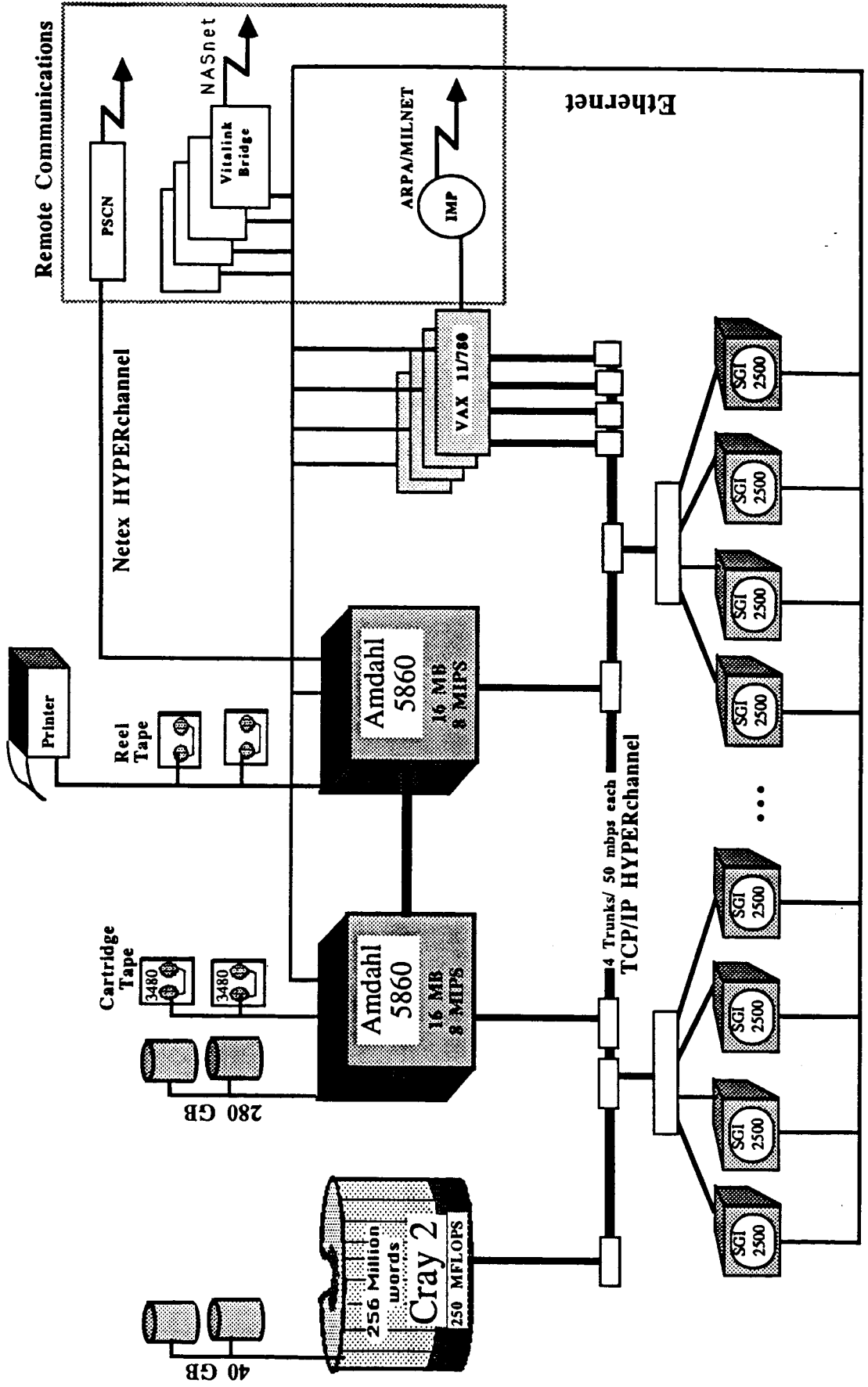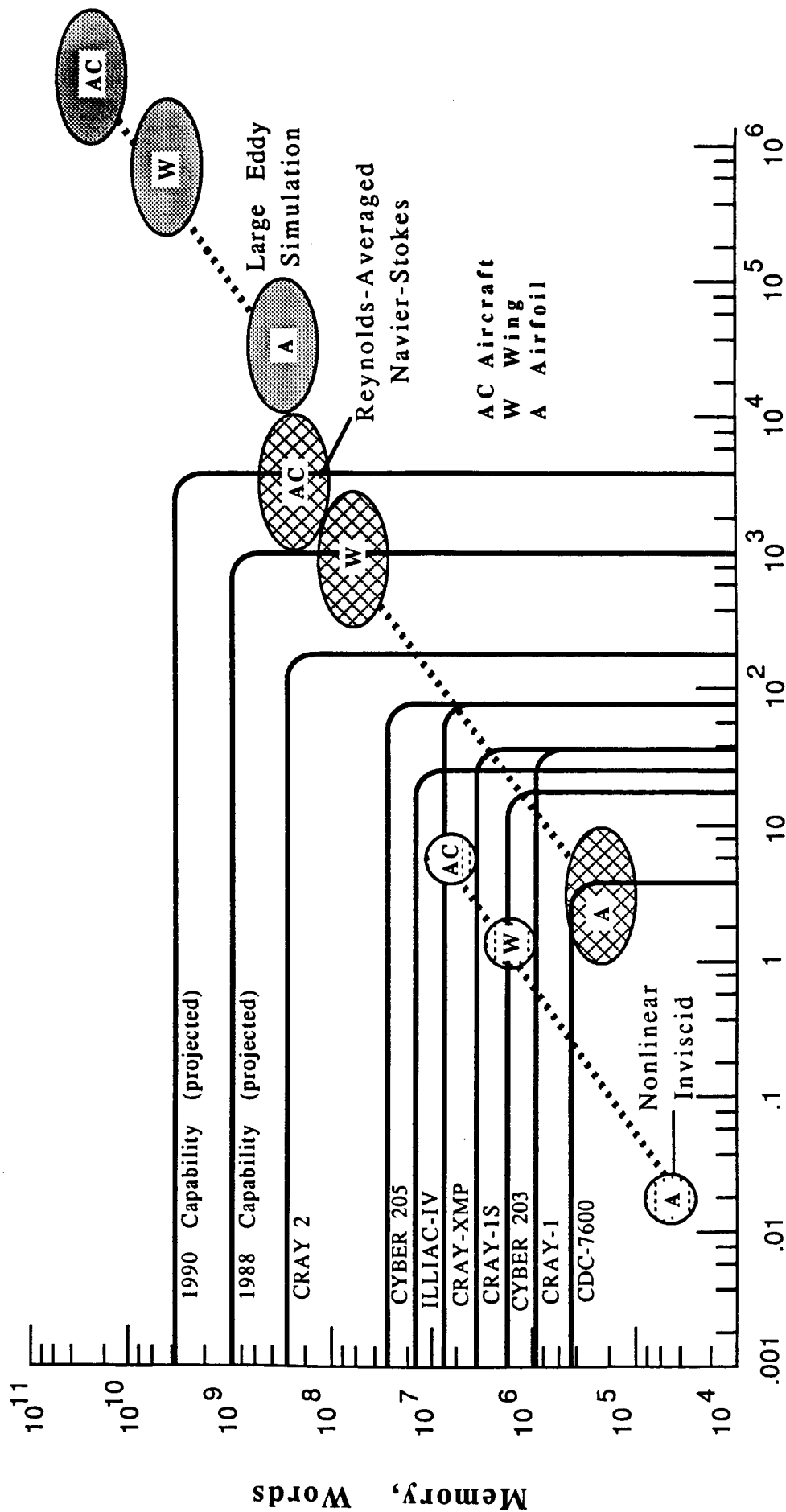# Figure 1. NAS Initial Operating Configuration

# Figure 2. Computer Speed and Memory Requirements

## 15 Minute Computational Aerodynamics Runs; 1985 Algorithms



Large Eddy Simulation

Reynolds-Averaged Navier-Stokes

AC  Aircraft
W   Wing
A   Airfoil

1990 Capability (projected)

1988 Capability (projected)

CRAY 2

CYBER 205
ILLIAC-IV
CRAY-XMP
CRAY-1S
CYBER 203
CRAY-1
CDC-7600

Nonlinear Inviscid

$10^{11}$
$10^{10}$
$10^{9}$
$10^{8}$
$10^{7}$
$10^{6}$
$10^{5}$
$10^{4}$

Memory, Words

.001   .01   .1   1   10   $10^{2}$   $10^{3}$   $10^{4}$   $10^{5}$   $10^{6}$

Sustained Computer Speed, MFLOPS

additional supercomputer (HSP-3) with 16 times the Cray 2 speed (4 GigaFLOPS, i.e., 4000 MFLOPS) will be added. To assure that the user can fully exploit this increased power, it is essential to examine the total supporting system-level infrastructure. In particular, it is critical to provide sufficient capacity for the very large data files characteristic of computational fluid dynamics computations and to scale the bandwidth of the communication systems to handle the increased traffic. This paper reports the results of an analysis of the implications of this 20-fold increase in processing speed on the network data communications requirements.

**RELATED WORK:** Several authors have recently considered the "balances" needed between computer speed, storage requirements and data communications. Kung [5] presents a model of balanced computer architectures for particular classes of computation; however his work is primarily concerned with the characteristics of the processor itself and not the total environment. The same limitation is true of "Amdahl's rules of thumb" (see [13]) which Worlton [6] states in the form, "One byte of main memory is required to support each instruction per second," and, "One bit of I/O is required to support each instruction per second."

Thorndyke [7] does consider the support environment for the ETA-10 supercomputer and views the mass storage subsystem as part of a memory hierarchy consisting of central processor memory, shared memory, local disks, and mass storage. He concludes that the capacity ratios between each level of this hierarchy should increase by a factor of 16:1. He also proposes that data communication rates be matched to disk transfer rates of 10 Megabytes/sec. Ewald and Worlton [8] note that the Cray XMP 48 also exhibits a 16:1 ratio between local "disk" storage (SSD in this case) and main memory. Furthermore, they note that historically the requirements at all levels of the storage hierarchy have been roughly proportional to the speed of the computer. For future scaling, they propose that on-line disk capacity requirements should grow at about 2/3 of the performance growth of the supercomputer and that transfer rates be increased to achieve a balanced system.

The 1986 work of Wallgren [9] uses similar theoretical scaling laws based on past and current experience to project the supporting environment needed for future supercomputers. His primary focus is on storage and not on data communications requirements. He correctly notes that the results are dependent on the assumed system architecture and on the usage profile. Wallgren's extrapolations extend over more than a decade in time and a factor of as much as $10^4$ increase in supercomputer speed.

**SCOPE:** The present study is far less ambitious and examines the expected impact on the data communications of a 20-fold increase in computer speed over a 3-4 year time period and specifically assumes an architecture (Figure 3) which is structurally the same as the existing initial operating configuration of Figure 1.

# Figure 3.  1990 Model Configuration



Figure 3.  1990 Model Configuration
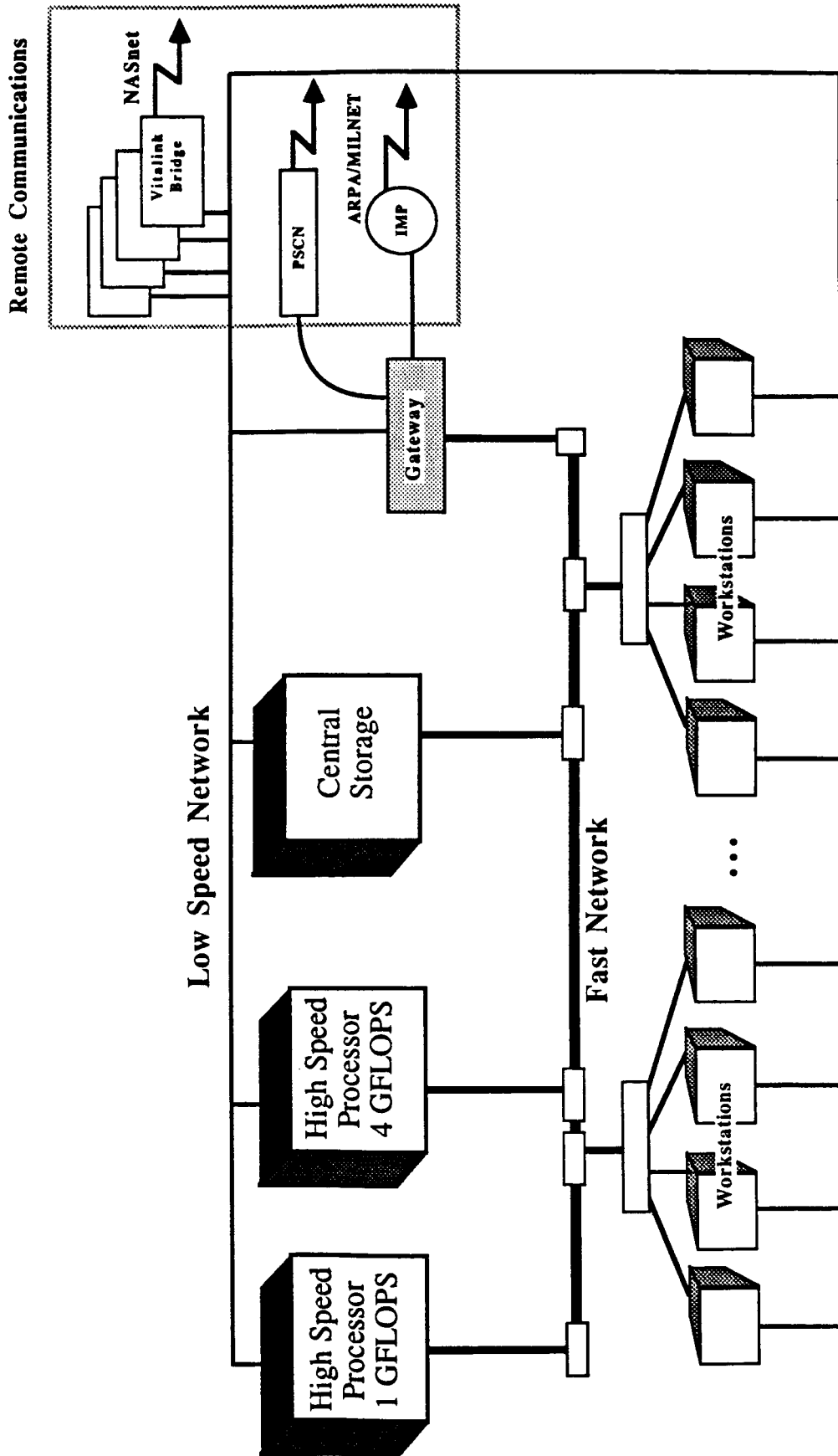
Furthermore, the user population is very well defined since the NAS Program is not a general purpose scientific computing center but is devoted specifically to aeronautics studies and applications; the program plan calls for some 90% of the total time to be used for computational fluid dynamics. This permitted a very detailed workload model to be developed. The results were obtained by analysis of this model and by computer runs of a discrete simulation program.

The data communication topology assumed for 1990 is shown in Figure 3 and consists of a primary high speed data network between two supercomputers and a central mass storage system; a secondary network for communication between the local workstations and the remaining subsystems; and a remote communication subsystem. The primary object of the study was to determine the requirements for the "backbone" high speed data network for the movement of large files (typically from 5-80 million words in size) to and from the high speed processors. (The existing high speed data network of Figure 1, consisting of four parallel HYPER-channel trunks, is capable of providing an *effective* transfer rate in excess of 6 megabits per second. This is expected to be adequate for the first scheduled upgrade in 1987/1988 to a 1 Gigaflop processor.) Both the remote communication subsystem and the local area network serving the workstations are scheduled to increase in data rates to handle the projected increase in usage. Their sizing is determined by factors *other* than the speed of the supercomputers.

**REMOTE COMMUNICATIONS:** The NAS remote communications system currently supports Arpanet/Milnet, NFSnet, NASnet, and the NASA-wide Program Support Communication Network (PSCN). Both terrestrial links (at 56 Kilobits/sec) and satellite T1 links (at 1.544 megabits/sec to the NASA centers) are presently provided. The currently active remote sites and proposed future sites are shown in Table 1. In addition to the high speed service shown in the Table, the 1990 system is to support up to 100 dial-up users at selectable rates from 1.2-9.6 Kilobits/sec.

Access to the NAS network from remote sites is shown in Figure 4. Vitalink bridges manage the inter-network connection and monitor all traffic on the ethernet local area network passing only those messages that have remote destination addresses to the appropriate Vitalink unit.

Although the long term goal is to provide equivalent services to both local and remote NAS users, remote communication bandwidths are constrained by existing technology (and funding) limitations. Upgrade to T2 rates (6.2 megabits/sec) for selected NASA centers is planned during the 1988/1989 time period. Additional improvements under investigation include implementing class of service protocols (such as distinguishing bulk file transfers from interactive traffic) and techniques for reducing the volume of remote data communications (such as remote block editors and distributed graphics services).

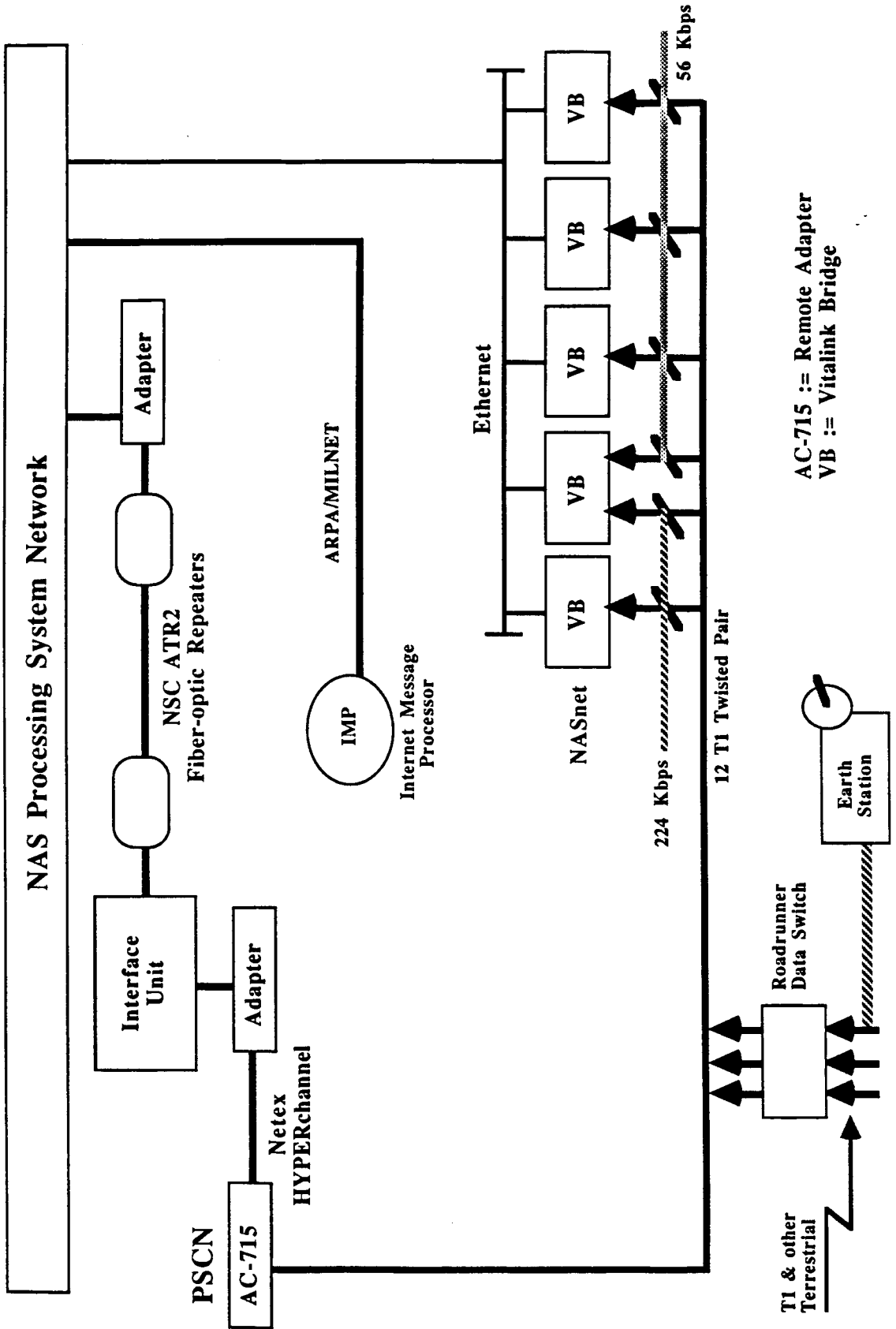# Table 1. High Speed Remote Communications

## Table 1.A  Sites Currently On-line

| Site | Location | Rate | On-Line Date | T1/T2 |
|---|---|---|---|---|
| Lockheed | Marietta GA | 56 Kbps | Sep 86 | Nov 88 |
| Grumman | Bethpage NY | 56 Kbps | Oct 86 | Jan 88 |
| United Technologies | East Hartford CN | 56 Kbps | Sep 86 | Dec 88 |
| McDonnel Douglas | Long Beach CA | 56 Kbps | Sep 86 | Jan 89 |
| McDonnel Douglas | Berkeley MO | 56 Kbps | Oct 86 | Mar 88 |
| Rockwell | Thousand Oaks CA | 56 Kbps | Oct 86 | Jun 88 |
| Northrop | Hawthorne CA | 56 Kbps | Oct 86 | Apr 88 |
| General Dynamics | Fort Worth TX | 56 Kbps | Oct 86 | Nov 87 |
| Lockheed | Burbank CA | 56 Kbps | Oct 86 | Feb 88 |
| Boeing | Seattle WA | 56 Kbps | Nov 86 | Oct 87 |
| Rocketdyne | Canoga Park CA | 56 Kbps | Oct 86 | Feb 88 |
| JSC | Houston TX | 56 Kbps | Oct 86 | Oct 87 |
| MSFC | Huntsville AL | 56 Kbps | Oct 86 | Feb 89 |
| GSFC | Green belt MD | 56 Kbps | Oct 86 | Sep 88 |
| Colorado State | Fort Collins CO | 56 Kbps | Jan 86 | N/A |
| LaRC | Hampton VA | 56/224 Kbps | Jan 86 | Jun 88 |
| LeRC | Cleveland OH | 56/224 Kbps | Jan 86 | Jun 88 |

## Table 1.B  Future Sites

| Site | Location | Rate | On-Line Date | T1/T2 |
|---|---|---|---|---|
| Flow Research | Kent WA | 56 Kbps | Mar 87 | N/A |
| General Electric | Cincinnate OH | 56 Kbps | Jul 87 | Oct 88 |
| Gates Learjet | Wichita KS | 56 Kbps | Sep 87 | N/A |
| Pratt-Whitney | East Hartford CN | 56 Kbps | Oct 87 | May 88 |
| Rohr | Chula Vista CA | 56 Kbps | Oct 87 | N/A |
| Vought | Dallas TX | 56 Kbps | Oct 87 | N/A |
| United Technologies | West Palm Beach FL | 56 Kbps | Apr 88 | Mar 89 |
| Detroit Diesel | Indianapolis IN | 56 Kbps | Mar 88 | N/A |

# Figure 4.  NAS Remote Communications

## January 1987



NAS Processing System Network

Adapter

NSC ATR2
Fiber-optic Repeaters

ARPA/MILNET

IMP

Internet Message
Processor

Interface
Unit

Adapter

Netex
HYPERchannel

PSCN

AC-715

Ethernet

VB   VB   VB   VB   VB

56 Kbps

NASnet

224 Kbps

12 T1 Twisted Pair

Roadrunner
Data Switch

Earth
Station

T1 & other
Terrestrial

AC-715 := Remote Adapter
VB := Vitalink Bridge

**LOCAL WORKSTATION COMMUNICATIONS:** The principal local users access the NAS system through powerful graphics workstations. The Silicon Graphics Model 2500 Turbo includes a fast 32-bit microprocessor, a floating-point accelerator, 2.5 megabytes of memory, a 474 megabyte disk and a special purpose microprocessor for hardware-based graphics manipulations such as hidden surface removal, rotation, zooming and clipping of raster images. The display resolution is 1024x780 and the user may select up to $2^{24}$ colors.

The workstation is an essential element of the system to permit graphical analysis and interpretation of the extremely large output files generated by computational fluid dynamics batch programs. It also enables the user to work interactively with the supercomputer. For example, a user may designate locations in the vicinity of an aerodynamic surface on his display, the flow patterns representing particle traces from those locations are computed on the supercomputer and the results returned to the user's display. The data communications services to the workstation must be sufficient to enable the user to make maximum use of both the batch and interactive workstation/supercomputer capabilities.

There are three modes in which graphics can be displayed at the workstation. It is possible to send down the solution files (or major subsets of these files) for computation and display at the local workstation. Although the workstation has the processing power comparable to a Vax 11/780 and the special purpose graphics hardware expedites the generation of vectors, the display of polygons, and performs up to 500.000 coordinate transformations per second (for rotation, translation, scaling), there are many limitations to this mode of operation including the I/O bandwidth of the workstation and limitations of main memory and disk space. The preferred modes use the supercomputer for the computationally intensive tasks and send down either display lists or pixel data to the workstation. Both modes may be used, however with the current workstations it is difficult to use the pixel mode of interaction effectively. Future workstation upgrades will alleviate this limitation and the model assumes that both the display list mode and the pixel mode of interaction will be used extensively.

Based on the above description of operational use, the data communications requirements to the workstation re primarily sized according to the graphics needs of the user and the projected capabilities of future workstations. The present *effective* rate of approximately 2 megabits per second is appropriate for the current workstation. Some representative display capabilities at 8 megabits per second (as planned for 1990) are shown in Table 2. (For comparison, the data rate needed to support real time, interactive, color graphics with animation at 30 frames/sec. would be about 800 megabits/sec. The NAS Program is funding a prototype of such a bus and considering prototyping an advanced graphics workstation with a massive frame buffer and very high resolution graphics capabilities. This is not included in the 1990 system model.)

# Table 2. Display of Simple* Color Images

### (* 8000 Polygons, includes Hidden Surface Removal, Flat Shading)

| Mode of Operation | Typical File Size | Channel Time @ 8 Mbps | Time for (current) Workstation to Generate Display | Remarks |
|---|---|---|---|---|
| Send Solution File from Supercomputer to Workstation | 400 Megabits | 50 seconds | approx. 25 seconds | Near capability of workstation storage |
| Send Display List to Workstation | 2.6 Megabits | 0.3 seconds | approx. 5 seconds | Approx. 1 sec. if no hidden surface removal |
| Send Minimal Pixel File to Future Workstation | 8 Megabits (1k x 1k x 8 bits) | 1 second | approx. 0 seconds | Requires external access to frame buffer |
| Send Fully Rendered, Very High Quality, Pixel File to Future Workstation (100,000 Polygons) | 400 Megabits (2k x 2k x 24 bits) | 6 seconds | approx. 0 seconds | "Finished product" graphics |

**HIGH SPEED "BACKBONE" COMMUNICATIONS:** The high speed data communication requirements were derived by analysis of the results of a model of the expected workload and checked using a discrete simulation program. The workload model assumed that the processing power of 5 Gigaflops available from the two 1990 high speed supercomputers (Figure 3) was fully utilized. A detailed profile (for computational fluid dynamics) of types of user tasks and expected frequencies had been developed over a six year time period and was updated in 1986 to reflect current algorithms, projected increased interactive usage and substantially increased use of graphics [10, 11, 12].

There are over two hundred model parameters including: numbers and types of local and remote users, number of host processors, protocol delays, amount of disk storage attached to the supercomputers, distribution of batch and interactive work, frequency of task execution, probability of abortive runs, etc. Scripts were developed to represent characteristic delays associated with "think time" to separate user-initiated sequential processes. These asynchronous processes compete for system resources. The high speed data communications capability was initially taken as unbounded and the workload was progressively increased until the full capability of the supercomputers was saturated.

A simplified listing of the classes of work initiated by the users is summarized in Table 3. The principal execution runs represent various types of numerical aerodynamic simulations used to solve fluid flow problems. These generally involved repeated iterations of difference equations over a three-dimensional grid which was assumed to consist of one millic grid points. The simple steady state design simulations used an inviscid potential and required approximately 20,000 calculations per grid point of result file. The more complex steady state design simulations used the Reynolds-Averaged form of the Navier-Stokes equations and required approximately 600,000 calculations per grid point. The comparable unsteady solutions typically required over 4 million computations per grid point. From Table 3 it is evident that this represented the dominant load on the supercomputers.

The workload data base contained detailed estimates of the sizes of the various input/output files for each task and a model of how they would be utilized. This provided a means for determining (for each class of work) the resultant numbers of files generated and where they would be moved across the network. For example, Table 4 shows the network traffic load for the dominant Unsteady Design Simulation task.

The source of most of the data communications load comes from the output of the high speed processors and the major traffic load is the movement of large files between the supercomputers and the central mass storage system. Figure 5 shows the results of the model study for this traffic. It should be noted that the study assumes that there are 400 Gigabytes of disk storage local to the high speed
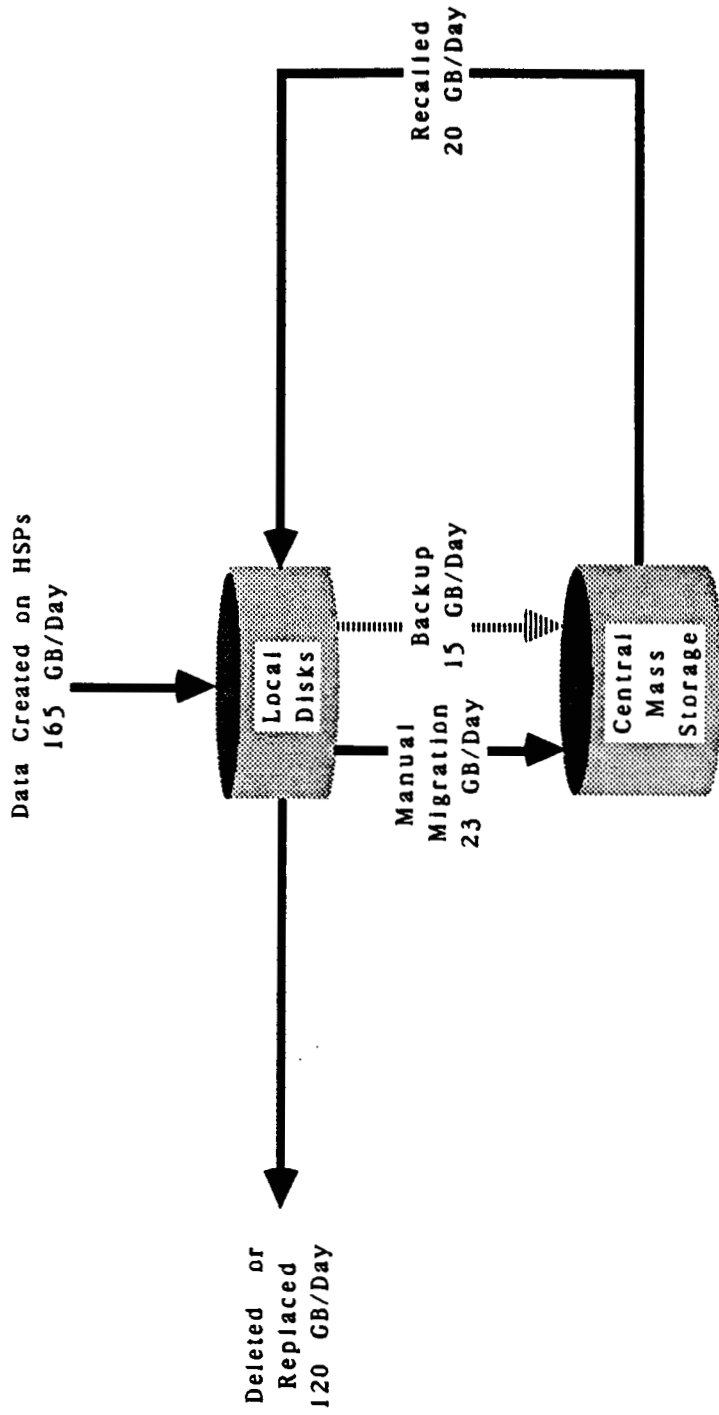
# Table 3. Estimated Processor Load of User Initiated Runs

| TASK | GFLOP/DAY |
|------|----------:|
| CODE & PARAMETER PREPARATION | 66 |
| PATCH GENERATION | 43 |
| GRID GENERATION | 96 |
| METHOD AND CODE DEVELOPMENT | 13,003 |
| SIMPLE STEADY STATE DESIGN SIMULATIONS | 5,732 |
| COMPLEX STEADY STATE DESIGN SIMULATIONS | 31,834 |
| COMPLEX UNSTEADY STATE DESIGN SIMULATIONS | 250,978 |
| FLUID RESEARCH LARGE EDDY SIMULATIONS | 53,472 |
| RESULT EDITING & VIEWING | 17,465 |
| DOCUMENT PREPARATION & USER COMMUNICATIONS | 31 |
| TOTAL | 372,719 |

**Table 4.   Network Load for Unsteady Design Simulations in MB/day - 1990 Model**

| Sources / Destinations | MASS STORAGE | HIGH SPEED PROCESSORS | TOTAL |
|---|---|---|---|
| MASS STORAGE | - | 4,428 | 4,428 |
| LONG-HAUL COMMUNICATIONS | 0.2 | 2,048 | 2,048 |
| WORK-STATIONS | 0.1 | 1,810 | 1,810 |
| HIGH SPEED PROCESSORS | 1,555 | - | 1,555 |
| TOTAL | 1,555 | 8,286 | 9,841 |

# Figure 5. Daily Supercomputer File Activity

Data Created on HSPs
165 GB/Day

Local
Disks

Recalled
20 GB/Day

Backup
15 GB/Day

Manual
Migration
23 GB/Day

Central
Mass
Storage

Deleted or
Replaced
120 GB/Day

Note: Supercomputer Disks Daily Net Accumulation = 42 GB

processors. This value affects the migration, backup and number of files recalled per day. It may also be observed that there is a net accumulation predicted on the local disks of 42 Gigabytes per day. This accumulation must be moved (by system-managed file migration) to the central mass storage system so that the supercomputers' disk space will be available for current work. This additional load on the data communications can, however, be distributed over the periods of lightest user activity.

Figure 6 summarizes the major result of the study and illustrates both the hourly distribution of traffic as well as the average daily traffic including migration. To avoid significant queueing delays during times of peak activity, and to handle periodic bursts within an hour efficiently, the bandwidth for the 1990 system should be sized well in excess of the peak rates shown on Figure 6. The design value recommended was 100 megabits per second [14]. This was regarded as reasonable in view of projected technological advances.
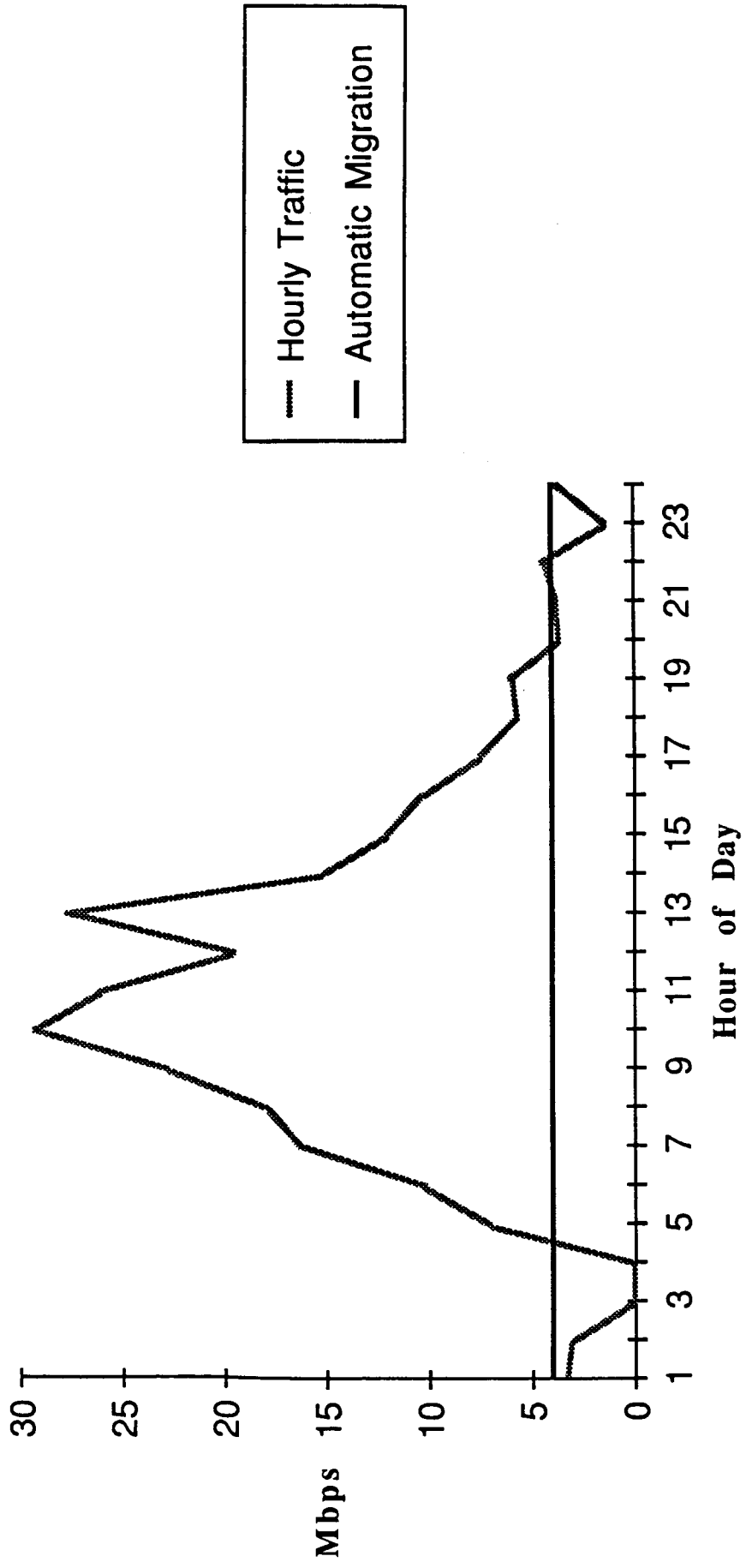
## CRITIQUE OF RESULTS

An essential (but often neglected) step in studies of this type is to estimate the sensitivity of the results to the assumptions inherent in the model. The succeeding sections provide estimates of the effects of some of the major assumptions on the projected data communications requirement.

**USE OF GRAPHICS**: The need for sophisticated graphics to analyze the very large data files typical of computational fluid dynamics is unquestioned and a significant allowance for this type of processing was included in the model. However, this was a projection and might not correspond to the actual future usage. Among other factors it is dependent on the capabilities of future workstations.

The model provided a breakdown into various types of output. The three principal classes created were raw solution results, graphics display lists and pixel files. It was found that a very useful measure for scaling data communications with processing power was the ratio of bytes of output to MFLOPs (millions of floating point operations) required to generate that output. This *characteristic output ratio* will be designated $\beta_M$ with units of Bytes/MFLOP. The importance of this parameter is that if the *only* deviation from the assumed model is processing power, the results can (within reasonable limits) be scaled directly. Furthermore, if the mix of classes of output varies, it is possible to estimate the effect on this critical parameter and hence (within modest bounds) on the final results.

Over the distribution and frequency of tasks initiated by users, the model showed that raw result files (including debug runs) were found to generate $\beta_M$=130 B/M (Bytes per MFLOP). In contrast, the generation of graphics display lists generate $\beta_M \approx 8,000$ B/M and pixel files generate $\beta_M \approx 5000$ B/M. In the model, 5% of the

# Figure 6. Data Transfer Rates



Legend:
— Hourly Traffic
— Automatic Migration

Mbps
30
25
20
15
10
5
0

Hour of Day
1  3  5  7  9  11  13  15  17  19  21  23

total available processing power (FLOPs) was devoted to graphics. Approximately 75% of the graphics FLOPs were utilized for pixel files, however much more processing is required for pixel files. Hence, this distribution corresponded to approximately three graphics display lists for each pixel file produced. Since the supercomputer power of the 1990 model system was sized at 5 GFLOPS, the 5% utilized for graphics corresponds to 250 MFLOPS, equivalent to the total processing capability of the 4 CPU's of a CRAY 2!

For the assumed usage distribution, the overall value of the characteristic output ratio for the 1990 model was found to be $\beta_M \approx 400$ Bytes/Mflop. Figure 7 shows the variation in this parameter with changes in the relative amount and type of graphics usage. The abscissa represents the proportion of processing power used for graphics processing (both display list and pixel data) and the one-parameter family of lines represents various mixes of pixel output and display list output. The reference point corresponding to the model is marked. It may be noted that increased use of pixel data relative to display list output has a significant effect in reducing $\beta_M$. Display list output with a $\beta_M$ of 8,000 B/M (compared with pixel file output with a $\beta_M$ of 5000) utilizes less of the processing power of the supercomputer but places a greater load on the data communications (and on the workstations).

**THE USER PROFILE:** The workload model was specifically designed for solving problems in computational fluid dynamics. Table 5 summarizes a 1986 study of major users of supercomputers at the NASA Ames Research Center. The first three columns represent supercomputers of the Ames Central Computing Facility and only the Cray 2 is a NAS supercomputer. Furthermore, the two sampling periods shown in Table 5 occurred shortly after the Cray 2 was initially installed, in a pre-operational time period, and do not represent the *intended* operational distribution of users. Nevertheless, it appeared that computational chemistry utilized large amounts of supercomputer time and might represent a future NAS system user that was not reflected in the workload model.

Much of the current work in computatonal chemistry at NASA Ames Research Center is devoted to determining the electron density distribution of molecular structures by solving the Schroedinger equation. The methods used involve evaluation of many multi-dimensional integrals and manipulation of very large matrices to find eigenvectors and eigenvalues. These methods are quite different than those used in computational aerodynamics studies.

Consequently, a study was made of the character of the output files generated by these supercomputer users. The results are shown in Figure 8 in terms of the parameter $\beta_M$. The prediction of the 1990 model study is shown for comparison. The characteristic output of computational chemistry was found to be comparable to the major computational aerodynamics users. Hence, even if computational chemistry were in the future to become a significant component in the mix of NAS

# Figure 7. Variation of characteristic output ratio as a function of graphics processing
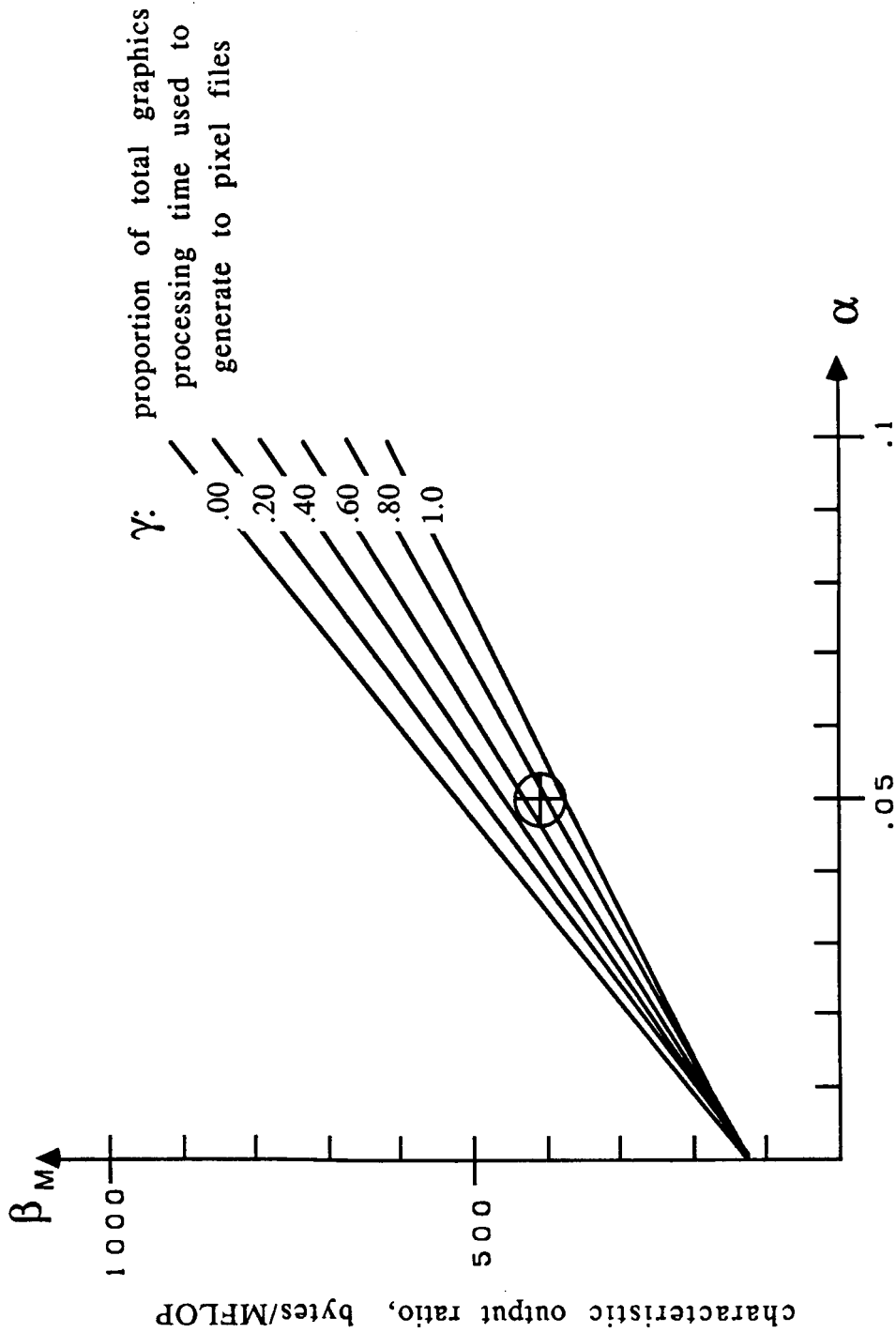


$\gamma$: proportion of total graphics processing time used to generate to pixel files

.00
.20
.40
.60
.80
1.0

$\alpha$

.05

.1

relative amount of HSP computing power used for graphic processing

$\beta_M$

1000

500

characteristic output ratio, bytes/MFLOP

# Table 5. Nasa Ames Vector Processor Utilization
## Percentage of User Time by Organization

| USERS | CRAY X-MP 22 FY 85 | CYBER 205 FY 85 | CRAY X-MP 48 3/86-10/86 | CRAY 2 | |
|---|---|---|---|---|---|
| | | | | 3/86-6/86 | 7/86-10/86 |
| Remote | -------- | -------- | -------- | 42.2 % | 31.3 % |
| Advanced Aerodynamics Concepts | 2.3 % | -------- | 7.1 % | -------- | -------- |
| Applied Computational Fluids | 22.7 % | 0.1 % | 24.1 % | 1.9 % | 12.2 % |
| Experimental Fluid Dynamics | 8.0 % | 0.7 % | 6.1 % | -------- | -------- |
| Computational Fluid Dynamics | 25.1 % | 25.0 % | 24.1 % | 17.8 % | 9.9 % |
| Computational Chemistry | 21.0 % | 67.9 % | 16.0 % | 23.5 % | 14.8 % |
| TOTAL | 79.1 % | 93.7 % | 77.4 % | 85.4 % | 77.6 % |

Figure 8. Comparison of Current Usage and 1990 Model
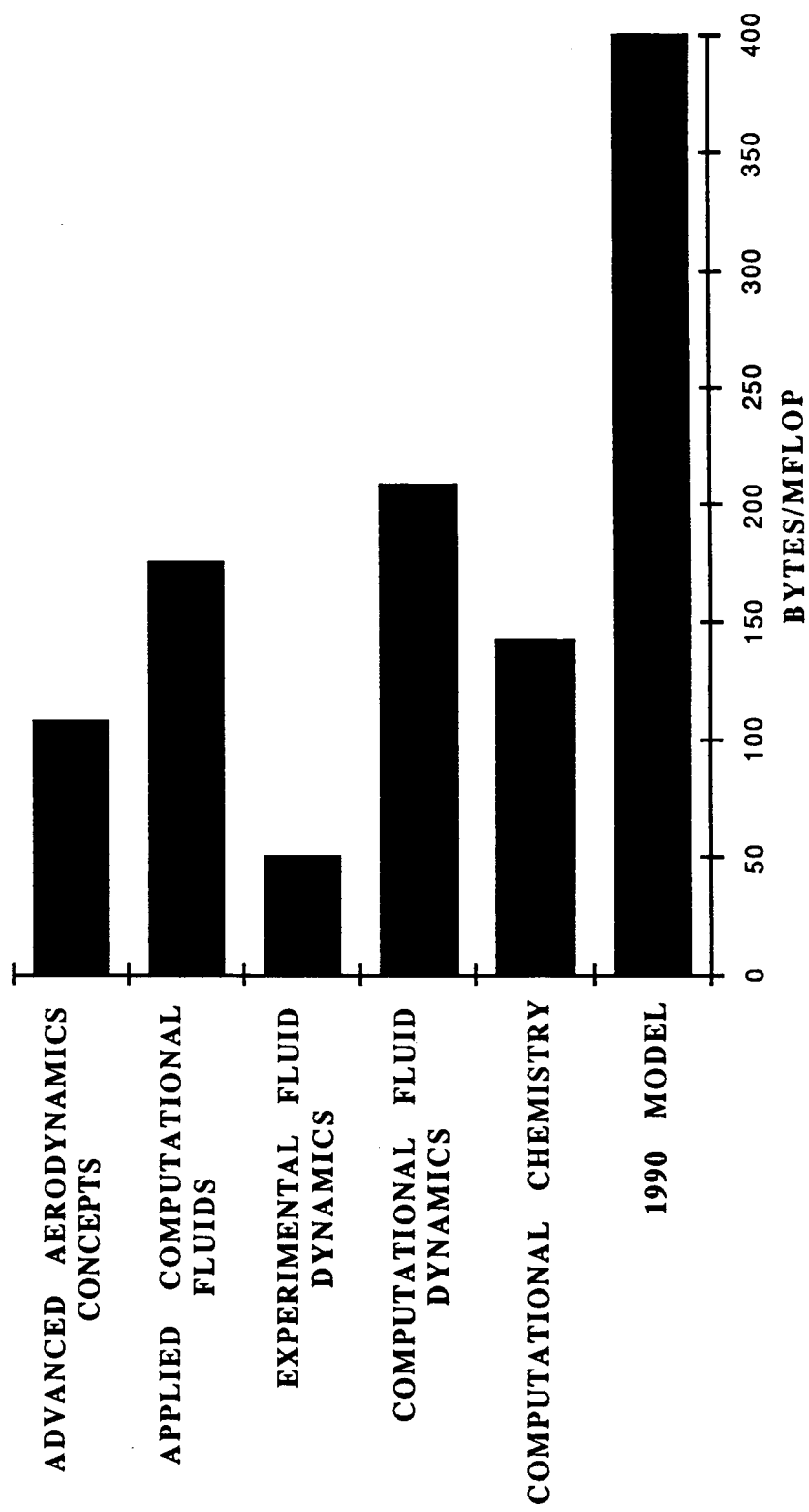
users, it **was** concluded that this would not invalidate the results.

It is also of considerable interest to note the relationship of the predicted 1990 usage compared to actual 1986 usage. It is not surprising that the predicted value of $\beta_M$ is almost twice the 1986 measured values in Figure 8 since extensive use of graphics was just beginning. Figure 7 confirms that increased graphics use results in an increase in $\beta_M$.

**AMOUNT OF LOCAL DISK STORAGE**: Locality of files increases the hit rate and reduces the volume of data traffic (see [15]). If local storage were unlimited, there would be no need to move files to a central storage facility. With finite local storage, there are two major effects to consider; (1) the ability to mask automatic migration, and (2) the effect of aging of files.

The two supercomputers of the 1990 system were modelled with a total of 400 Gigabytes of local storage. From Figure 5 the daily production from the supercomputers is 165 Gigabytes/day (with a net accumulation of 42 Gigabytes/day). Since more than half the 400 Gigabytes of local disk space is needed for current working space each day, it was found that the average age of files before automatic migration to central storage was about 3.5 days. This permitted the forced migration shown on Figure 6 to be spread over times of minimum activity. Unless the deviation from the assumed 400 Gigabytes dropped to as low as about 250 Gigabytes, the automatic migration could still be masked.

The main effect of aging is to reduce the volume of files recalled from storage. If files could be maintained local to the supercomputers for about 10 days (two working weeks) instead of 3.5 days, it was estimated that the recall load would be reduced from 20 Gigabytes per day to less than 10 Gigabytes per day. Aging would also increase the number of files deleted and decrease the amount of manual migration shown in Figure 5. All of these effects decrease the traffic on the network. It was estimated that increasing the local disk storage by an additional 250 Gigabytes could reduce the traffic between the supercomputers and central storage by 30-50%.

**OTHER EFFECTS:**

The model assumed a typical grid size of $10^6$ points. Increasing the geometrical refinement by using more grid points does not have any affect on $\beta_M$ provided the computations needed per grid point remain constant. However, in many cases not all the final data needs to be saved. The finer grid not only yields results at more spatial locations but also permits a more accurate approximation to the differential equation. It is often only necessary to save the results on a much coarser grid (except for critical regions of interest) and recompute the interior points if needed. For example, saving every fourth grid point in each dimension would reduce the output by a factor of 64. Similarly, it may not always be necessary to save 64-

bits of precision in the output data.

The realism of the physics was discussed briefly in connection with Figure 2 where it was noted that solving the more complex equations such as large eddy simulation involved an enormous increase in the number of computations. There would certainly not be a proportionate increase in the output, hence the trend towards removing the approximations would result in a sharp decrease in the ratio $\beta_M$ and alleviate the data communications load. This was a factor in the 1990 model and will become increasingly significant as more powerful supercomputers become available.

It was assumed that the algorithms used in the 1990 time frame would be the same as those presently known or in use. This will certainly not be the case since algorithm development is a fertile, on-going activity. Prior trends have been in the direction of progressively decreasing $\beta_M$, but it is impossible to predict the implications of undiscovered algorithms. If there is a *radical* change from deterministic approaches to statistical techniques (for example, gas lattice automata methods, [16]) this would clearly require much more computational effort without a corresponding increase in output, resulting in a sharp decrease in $\beta_M$.

## SUMMARY

The results of an analysis of the 1990 "backbone" data communications requirements for the advanced Numerical Aerodynamic Simulation Program showed that an *effective* bandwidth of 100 megabits per second would be appropriate.

A critical examination of the various assumptions inherent in the model indicated that this was a safe, conservative result.

Among the most sensitive assumptions was the projected amount and type of interactive graphics expected to be used.

Increasing the amount of disk storage local to the supercomputers would result in a significant decrease in the data communications requirements.

It was found that an extremely useful parameter in the scaling study was the characteristic output ratio $\beta_M$ representing a measure of bytes of output relative to megaFLOPs of processing.

This parameter made it possible to estimate readily the effects of (modest) variations in the model assumptions on the data communications requirements.

# References

[1] Peterson, V.L., Ballhaus, W.F., Jr., and Bailey, F.R., "Numerical Aerodynamic Simulation (NAS)", *Large Scale Scientific Computation*, ed. by Seymour V. Parter, Academic Press, Inc., 1984

[2] Bailey, F.R. and Ballhaus, W.F., "A Large Scale Computer Facility for Computational Aerodynamics", IEEE Transactions on Nuclear Science, Vol. NS-32, No. 1, 1985.

[3] Bailey, F.R., "Status and Projections of the NAS Program", Symposium on Future Directions of Computational Mechanics, ASME Annual Meeting, Anaheim, CA., Dec. 1986 (Also available as NASA Technical Memorandum 88339, July 1986)

[4] "Numerical Aerodynamic Simulation Program NPSN System Specification for IOC", NASA Ames document PS-1100-01-C00, 27 February 1984

[5] Kung, H.T., "Balanced Computer Architectures", J. Complexity, V. 1, 1985 (Also available as "Memory Requirements for Balanced Computer Architectures", Carnegie-Mellon University, CMU-CS-85-158)

[6] Worlton, J., "Some Notes on Storage Systems for Supercomputers", Presentation at Los Alamos National Labs., June 1985

[7] Thorndyke, L.M., "Supercomputers and Mass Storage: The Challenges and Impacts", IEEE Symposium on Mass Storage Systems, Tucson, Arizona, Nov. 1985

[8] Ewald, R.H., and Worlton, W.J., "A Review of Supercomputer Installation Mass Storage Requirements", IEEE Symposium on Mass Storage Systems, Tucson, Arizona, Nov. 1985

[9] Wallgren, K., "Supercomputing and Storage", SPIE 30th Annual International Symposium, August 1986, San Diego, Ca.

[10] "NPS System Specification, Appendix II, Multi-Task Model", NASA Ames document PC-320-02, 15 September 1980

[11] "Revisions to Multi-Task Model for NAS Program", NASA Ames document PD-1000-01-C00, 6 February 1984

[12] "Workload Data Base Overview", NASA Ames document PR-1101-00-N00, 18 February 1986

[13] Sieworek, D.P., Bell, C.G. and Newell,A., *Computer Structures: Principles and Examples*, McGraw Hill, New York, 1982

[14] "Numerical Aerodynamic Simulation Program NPSN System Specification for the Extended Operating Configuration", NASA Ames document PS-1101-01-C00, 13 September 1986

[15] Smith, A.J., "Long Term File Migration: Development and Evaluation of Algorithms", Communications of the ACM, V. 24, No. 8., August 1981

[16] Frisch, U., Hasslacher, B., and Pomeau, Y., "Lattice Gas Automata for the Navier-Stokes Equation", Phys. Rev. Lett., 56(1986) 1505