
The Role of Graphics Super-Workstations in a Supercomputing Environment

E. Levin

September 1989

Research Institute for Advanced Computer Science
NASA Ames Research Center

RIACS Technical Report 89.37

NASA Cooperative Agreement Number NCC 2-387

(NASA-CR-188652) THE ROLE OF GRAPHICS
SUPER-WORKSTATIONS IN A SUPERCOMPUTING
ENVIRONMENT (Research Inst. for Advanced
Computer Science) 14 p

CSSL 09B

N92-10306

Unclas
0043038

G3/61

RIACS

Research Institute for Advanced Computer Science
An Institute of the Universities Space Research Association

11/2/89
43038
P-14

The Role of Graphics Super-Workstations in a Supercomputing Environment

Dr. E. Levin, Deputy Director
Research Institute for Advanced Computer Science
NASA Ames Research Center, Mail Stop 230/5
Moffett Field, California 94035, USA

KEYWORDS/ABSTRACT: graphics workstations/scientific visualization/supercomputing environments/interactive computing/ distributed processing

A new class of very powerful workstations has recently become available which integrate near-supercomputer computational performance with very powerful and high quality graphics capability. These "graphics super-workstations" are expected to play an increasingly important role in providing an enhanced environment for supercomputer users. Their potential uses include; off-loading the supercomputer (by serving as stand-alone processors, by post-processing of the output of supercomputer calculations, and by distributed or shared processing), scientific visualization (understanding of results, communication of results), and by real time interaction with the supercomputer (to "steer" an iterative computation, to abort a bad run, or to explore and develop new algorithms).

INTRODUCTION

The term "graphics super-workstation" is defined here to refer to a category of workstations introduced in 1988 which combine very powerful computational capability with high quality graphics in a tightly coupled, *architecturally integrated* system. Typically, such workstations provide from $\approx 1/10$ to $1/100$ the floating point speed of the most powerful current supercomputers, they have relatively large main memories plus specialized buffers and caches, and are capable of generating and manipulating realistic, three-dimensional graphics displays. Furthermore, these workstations have been carefully structured to provide for very rapid movement of data between memory resources, computational resources and graphics resources and are designed to enable the efficient, concurrent utilization of these resources. Hardware and software have been incorporated to facilitate networking to supercomputers and to other workstations or computational resources. The extensive system and application software provides users with a powerful and convenient working environment. The primary purpose of this paper is to explore the possible future role and impact of such workstations in a supercomputing environment. This is preceded by an overview of the current status.

CURRENT STATUS

The principal vendors that provide true graphics super-workstations include Ardent Computer Corporation (the Titan), Silicon Graphics Inc.(the high end of the Power Series family), and Stellar Computer Inc. (the GS-1000 and more recently the GS-2000). (The continued presence of the Apollo 10000 Series as a competitor at the very top end is less clear since the acquisition of Apollo by Hewlett-Packard.) Many other workstations have varying degrees

This work supported by Cooperative Agreement NCC 2-387 from the National Aeronautics and Space Administration (NASA) to the Universities Space Research Association (USRA).

of processing and graphics capability (including Sun, DEC, Tektronix, etc.) and there are attached processors using signal processing technology that are capable of generating very impressive graphics approaching the quality of photo-realism (such as the Pixar and the AT&T Pixel machine). There is also a recent trend by vendors of mini-supercomputers (notably Alliant) to add specialized hardware and software to their mainframes to support graphics applications. However, the distinguishing feature of the graphics super-workstation is the efficient, effective and convenient integration of the very powerful computational and graphics capabilities into a single-user system. Typical graphics generation and display manipulation characteristics are summarized in Table 1.

Table 1. Typical Graphics Capabilities	
Screen Display	<p>19" Color Monitor Resolution 1280 x 1024</p>
Image Bit-Planes (+ Control and Overlay Planes)	<p>16-32 Bit Z-Buffer 24 (+) Bit Color Planes Provision for Double Buffering</p>
Display Speed	<p>500,000 3-D Vectors/Sec 150,000 Gouraud-shaded, Z-buffered Triangles/Sec 30,000 Phong-shaded, Z-buffered Triangles/Sec</p>
Surface Geometry Approximations (Primitives)	<p>Polygons, Triangular Strips, Meshes NURBS</p>
Lighting/Shading/Rendering	<p>Flat, Phong, Gouraud Shading, Texturing Transparency, Specular Highlighting, True Ray Tracing</p>
Support for Animation	<p>Playback at 60 frames/sec (Stellar: 74 frames/sec) NTSC Compatible, RS 170 out 10,000 Gouraud-shaded triangles/frame at 15 frames/sec</p>
Stereo	<p>Full Color Multiple Viewers</p>

Representative values of the hardware and processing characteristics are shown on Table 2. The range given for the vector processing capabilities reflects both differences between the various vendors as well as the usual obfuscation regarding how the performance is to be measured. (The 64 MegaFlops Peak is for a 4-processor Ardent Titan.) Since current prices range from about \$60,000 to \$250,000, depending on the vendor and the configuration, the enthusiastic claims of outstanding price/performance seem well justified.

It is interesting to note how much technology has changed the economics of computing. During the 1960's, Grosch's "law" that processing performance was proportional to the square of the cost favored very large central processors. This is no longer the case from the point of view of a single user *provided* there is adequate memory to handle the problem without excessive I/O and runs can be completed in wall clock times that fit human work schedules and limits of patience.

Table 2. Representative Hardware and Processing Characteristics
Number of vector, floating-point processors: 1-4 (8?)
Vector Processing: 5-30 MegaFlops { 64 MFlops Peak }
Integer Processing: 10-80 Mips
Main Memory: 16-128 MegaBytes { Access = 300 MegaBytes/sec }
Cache: ≈ 1 MegaByte { Rapid Access ≈ 1 GigaByte/sec }
Input/Output: VME Bus Bandwidth 80-100 Megabits/sec
Disc Storage: 375-2,000 MegaBytes

Typical system software offered by essentially all the vendors is listed in Table 3. In addition to the system software, a very large number of sophisticated application software packages are offered in such areas as:

- Computational Fluid Dynamics
- Computational Chemistry
- Image Processing and Image Synthesis (e.g. Medical)
- Geophysical/Seismic Visualization
- Electrical Computer Aided Design
- Computer-Aided Engineering, Computer-Aided Manufacturing
- Math Libraries (Mathematica, Matlab)
- Animation/Visual Simulation

Table 3. Typical System Software

UNIX

AT&T System V.3

Berkeley 4.2/4.3 Extensions

LANGUAGES/COMPILERS

Fortran 77 + Extensions

C

Graphics "*Languages*" (e.g. Phigs+, Dore)

NETWORKING/INTERFACING

RPC

NFS, TCP/IP

Ethernet/Pronet

FDDI

Windowing {X-Windows or NEWS (SGI)}

A potentially serious software problem is the lack of agreement on standards for three-dimensional graphics both in terms of the libraries and primitives provided as well as the structure of the "language" to call and concatenate library entities. This lack of agreement limits the portability of code which has the undesirable side-effect of either locking a user into a specific vendor (or set of vendors) or, alternately, requiring time consuming re-write of the code. It also inhibits the communication of graphics between dissimilar workstations.

This is not a new problem. Unfortunately the graphics community has historically failed to reach consensus on such standards. During the 1970's there was the early CORE "standard" that achieved some reasonable degree of acceptance. However, the situation rapidly deteriorated with the adoption of CGI as the ANSI 2-D standard and GKS as the ISO 2-D standard.

Some of the key issues regarding present graphics standards are highlighted in Table 4. PHIGS (the Programmer's Hierarchical Interactive Graphics Standard) has now been formally adopted as the ANSI two dimensional Graphics Standard while GKS continues as the ISO 2-D standard. The evolution of GKS to a three dimensional ISO standard is currently under development and PHIGS+ is out for final comments prior to likely adoption as the ANSI standard. PHIGS+ provides extensions of PHIGS for surfaces, lighting and texturing. There is some need to strengthen the language bindings, however it is very probable that PHIGS+ will be adopted by ANSI.

There are also graphics support systems which might be termed *vendor standards* in that

they have been developed by and are maintained by specific vendors rather than defined by national or international committees. To some extent all vendors provide certain unique features such as specialized libraries, customized macros, etc. to provide convenience to the users. For example, Stellar offers AVS (Application Visualization System) as part of the StellarVision™ environment, Silicon Graphics provides an extensive GL™ graphics library and performs most graphics operations in hardware. These vendor packages do not conflict with PHIGS+ which is supported by both Stellar and Silicon Graphics. The RENDERMAN™ system by PIXAR is somewhat different and is proposed as a *general* interface between the geometry and rendering processes. It provides for rendering operations on surfaces defined by NURBS (Non-Uniform Rational B-Splines). NURBS has also become a generic term for surface elements bounded by a broad class of space curves.

DORE™ (Dynamic Object-Rendering Environment) is quite another matter. Ardent is aggressively promoting DORE as a *vendor standard* alternative to PHIGS+. DORE is available on CRAY and SUN computers and Ardent provides relatively easy and inexpensive access to the source code. There are many powerful and convenient attributes incorporated into the concept and design of DORE, yet the incompatibility with PHIGS+ further exacerbates the issue of 3-D graphics standards. (The recent announcement in August of 1989 of the proposed merger of Ardent and Stellar may resolve this incompatibility.)

Table 4. Graphics Standards

GKS { ISO 2D-Standard Exists, 3D Under Development }
PHIGS { ANSI 2D Graphics Standard }
PHIGS+ { <i>Almost</i> ANSI 3D Standard }
PHIGS Extension for Surfaces, Lighting, Texturing
Language Bindings Not Well-Defined
PEX { PHIGS+ Extended to X Windows, "3D X-Windows" }
RENDERMAN (Pixar)
Separates Rendering from Geometry (Provides Interface)
Shading/Texturing of NURBS-defined Surfaces
GRAPHICS SUPPORT SYSTEMS (<i>Vendor Standards</i>)
GL (Silicon Graphics)
AVS/StellarVision (Stellar Computer)
DORE (Ardent Computer)
Written in C, User Extensible
Easy to Use, Powerful Functionality { >PHIGS+ }
Not Compatible with PHIGS+

Ideally, scientific or industrial users should never have to deal with graphics primitives at all nor be concerned with many of the issues discussed above. There should be specialized user environments with appropriate libraries and "natural" languages for each of the various disciplines. If necessary, it should be possible for sophisticated users to modify, extend or customize their libraries and command environments. To some extent this is already taking place but the portability issue between various vendor products continues to be an outstanding problem.

CURRENT ROLE AND A LOOK AHEAD

Supercomputing is currently in the *gigaworld* era. We enjoy GigaFlops of processing power, GigaBytes of main memory, and GigaBits/sec of (local) data communication rates (e.g. UltraBus). Unfortunately, we may also be faced with GigaBytes of output to be digested. This may occur not only from scientific problems that deal with very large amounts of observational input data, but also as output data from solutions to mathematical equations representing some physical process.

A recent study [1] presented at the Third IFIP International Conference on Data Communication Systems and Their Performance reported a stubborn ratio of approximately 10^2 Bytes of output per MegaFlops of calculation. This empirical result was obtained for two quite dissimilar types of large scale calculations: computational fluid dynamics involving repeated iterations over a spatial grid, and *ab initio* computational chemistry involving determination of eigenvalues of very large sparse matrices and multi-dimensional integrations. Furthermore, the time period of the study spanned the transition from Cray 1 or Cyber 205 supercomputing to the Cray XMP era. It may be argued that not all of the output was needed, however the fact remains that it was requested by the users and it is reasonable to believe that such habits will be slow to change. Hence, as we approach processing power of ≈ 10 GigaFlops, this empirical ratio (for a moderate length run of 1000 seconds) would predict a GigaByte of output:

$$(10^2 \text{ Bytes/MFlops}) \times (10^4 \text{ MFlops/Sec}) \times 10^3 \text{ Sec} = 10^9 \text{ Bytes} = 1 \text{ GigaByte.}$$

We are also seeking far greater capabilities to be enable us to use computational methods to solve some of the challenging problems of science and engineering. Figure 1. depicts schematically three of the principal factors that serve as driving forces for increased processing speed and main memory; physics realism, increased dimensionality and data volume. Realistic representations of physical phenomena may increase geometric complexity or eliminate simplifying approximations, e.g., incorporating non-linearities or chaotic behaviour in the mathematical models. The issue of dimensionality is not limited to the three physical dimensions and time but more generally represents the number of degrees of freedom that must be considered as in the number of grid points in a CFD problem or the number of electrons included in an *ab initio* solution of the Schroedinger equation. The driving factor of data volume arises *primarily* from anticipated massive increases in observational data but large output files from computer runs add to the load. The need to process this data in reasonable time places severe demands both on computational speed and memory size.

At a recent conference, Peterson [2] expressed the need in Computational Fluid Dynamics for an ExaFlop processor with TeraWord main memory and noted that NASA has established a high performance computing goal for the next decade of a TeraFlop. At the same conference, Bretherton [3] stated that global monitoring in the 1990's would produce at

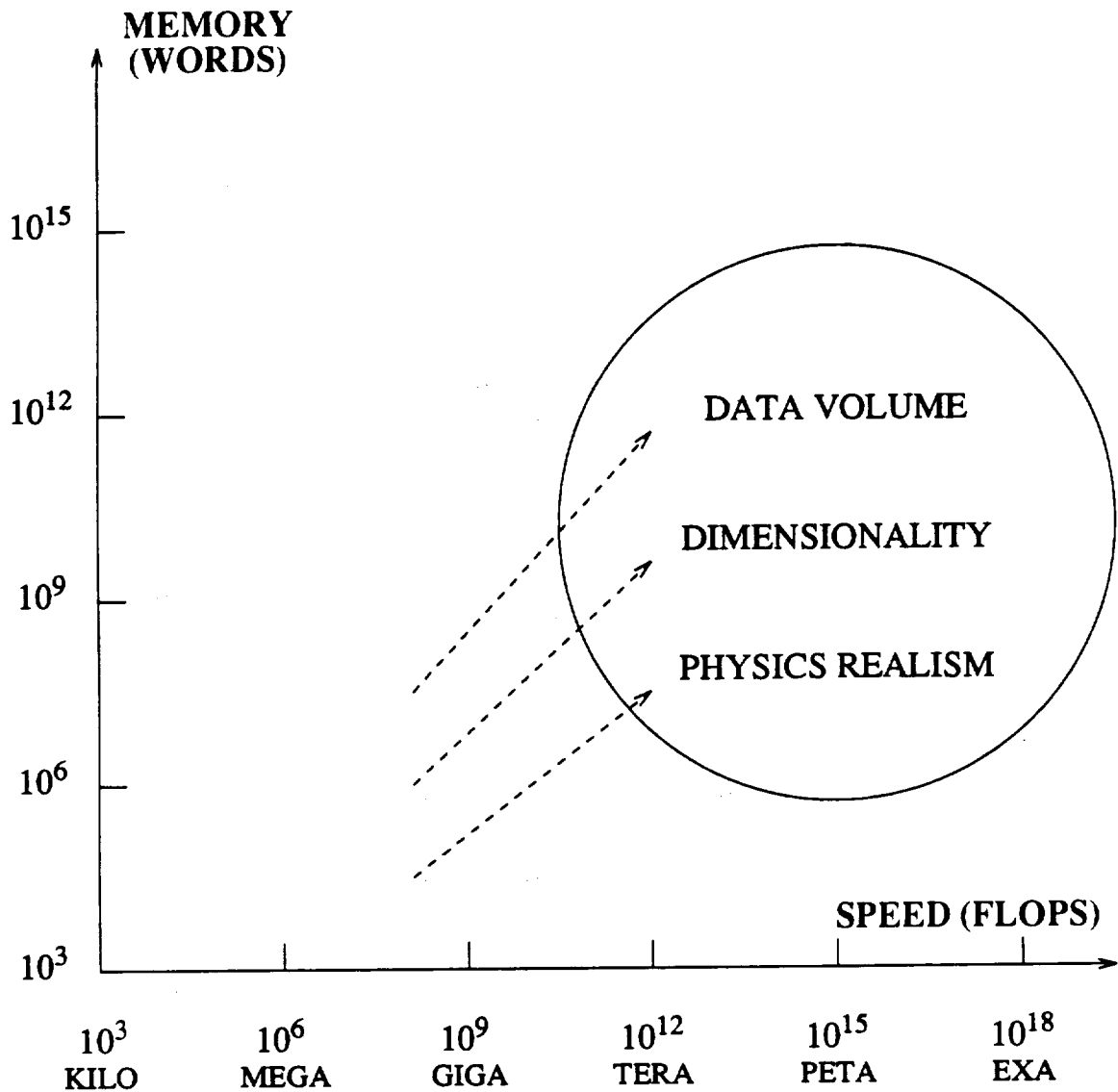


Figure 1. Driving Forces for Increased Computational Capability

least one TeraByte/day of observational data, hence three years of such data gathering would produce a PetaByte of data (if retained). Some of the scientific problems that are driving forces for the indicated increases in computational power are:

- Physics/Chemistry (molecular dynamics, *ab initio* quantum chemistry, surface chemistry, statistical mechanics, relativistic physics, cosmology, astrophysics),
- Computational Fluid Dynamics (including chaos/turbulence and coupling of aerodynamics with structural mechanics and propulsion system behaviour),
- Biology/Pharmacology (genome sequencing, genetic engineering, enzyme activity, cell modelling, drug design),
- Material Science (superconductivity, materials by design),
- Medicine (modelling of human organs and bones, surgical simulator/trainer),
- Planetary Science (global weather, environmental modelling, seismology).

It is a mistake to attribute the need for great increases in computational capability to the aggregate demands of many users each of whom may need only a modest amount of supercomputer time. The supercomputers of the future are needed to solve important scientific problems that presently *cannot be done at all* and should be designed and used for this purpose. The new generation of powerful workstations provide a logical, cost-effective and *user-time-effective* alternative to shared supercomputers and indeed this is one of their appropriate and important roles in a supercomputing environment. It is anticipated that this role of off-loading the supercomputer for problems of modest size as well as post-processing of supercomputer output will become increasingly important in the future. For graphics processing they are already a superior alternative to the supercomputer.

Along with processing speed and larger main memories, there is an implied concomitant need for increased storage and communication bandwidth. The present limitations on local communication rates of about a Gigabit/second are marginally adequate for now, however long-haul (wide-area) communications are totally inadequate and the problem will be severely exacerbated as computational capability and supercomputer usage increases. Senator Gore [4] has introduced legislation to establish a 3 Gigabit/sec national fiber optic network which, if implemented, would provide temporary relief. Fiber-optic links offer a theoretical bandwidth of a Terabit/sec but at this transmission rate the bits are spaced about 0.25mm apart and practical implementation beyond 10 Gigabits/sec is questionable. The problem of mass storage is even more limiting and there are fewer promising technological developments on the horizon.

In my view, this system-level problem is being addressed at the wrong end! The only hope for a successful solution is to change the way in which we make use of supercomputers so as to effect an enormous reduction in the amount of data that needs to be stored or communicated. This requires a fundamental shift in the approach to handling information in the chain from physical problem to final presentation of results. The emergence of graphics super-workstations offers an opportunity to enable that essential change in methodology.

Figure 2. depicts the steps in the typical solution sequence. The most primitive form of graphics utilization is to prepare a diagram, chart or graph to present the findings after completing the analysis and reduction of the results by computational means. This is denoted as the "old paradigm" on Figure 2. For very large output files, the post-processing itself becomes a major computational problem and a much more powerful approach is to use graphical techniques to search through the data and assist in the analyses by presenting visual representations to the scientist of the content. Only those portions of the data that are "interesting" may then need to be subjected to more detailed analysis and the final preparation of findings. This approach is part of the "new paradigm" indicated on Figure 2. It does not of itself necessarily reduce the volume of data but does facilitate the analysis.

Graphics super-workstations can be utilized effectively further "upstream" in the process as part of the calculation itself. Their specialized hardware and architectural properties can be used effectively in conjunction with the supercomputer in a distributed processing system. Although this augments the computational power available, it does not necessarily alleviate the overall problem of excessive data volume. However, the availability of X-Windows on workstations and on supercomputers not only enables a scientist to view the progress of a computation (and, if necessary, abort a clearly bad run) but permits the user to interact with the computation *in process* (e.g., by modifying parameters such as step size, grid spacing,

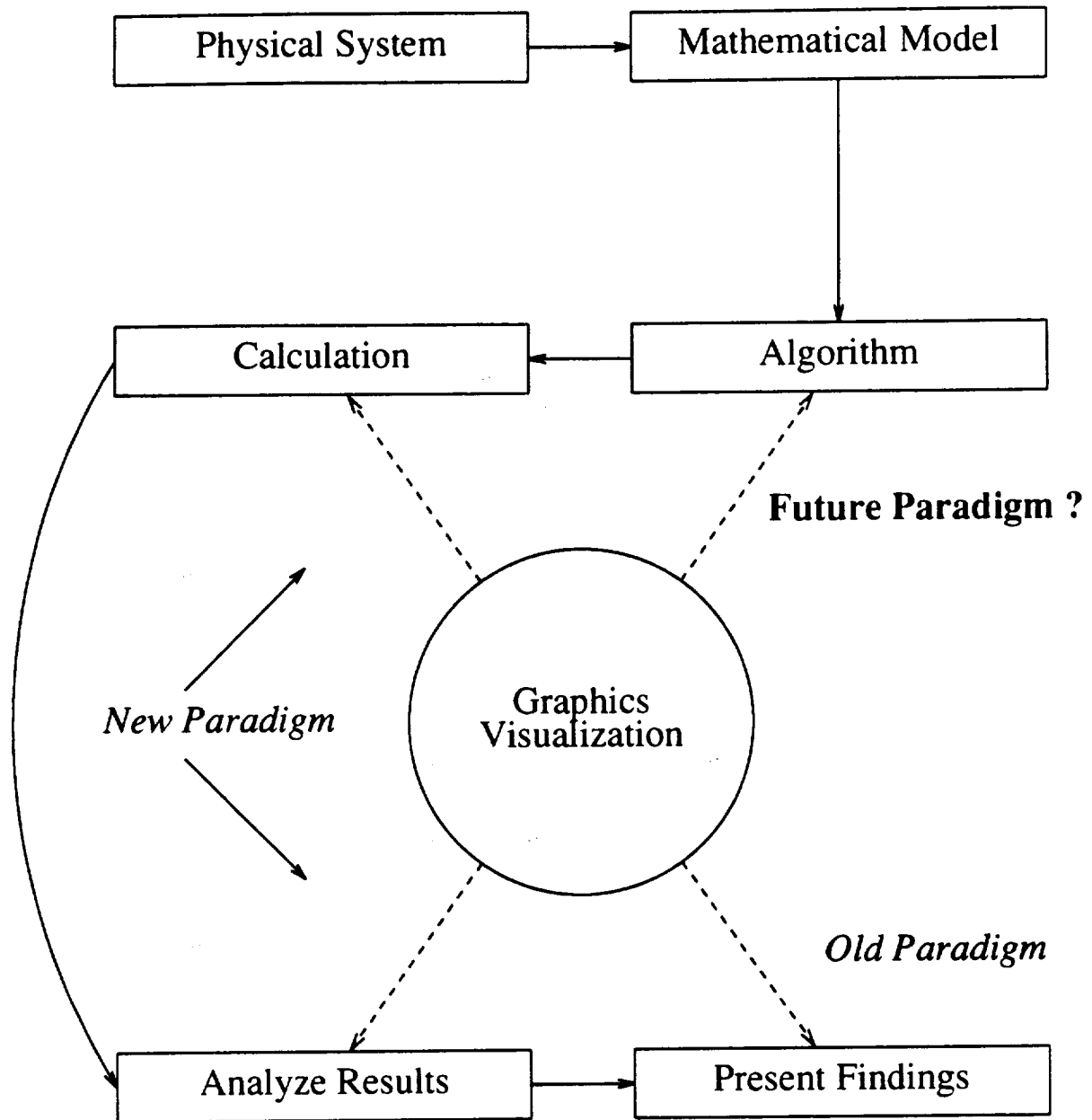


Figure 2. Paradigms for Use of Graphics Visualization

damping terms, etc.). This is not a new idea but the current realization of the concept is indeed new. The possibility was recognized quite early by John von Neumann who said [5], "....if he wishes...he can instruct the machine to present to him the relevant characteristics of the situation, continuously or in discrete succession, as the calculation progresses...then intervene whenever he sees fit." This aspect of the "new paradigm" not only expedites complex computations but has the potential to reduce markedly the volume of output data to be stored and transmitted. Indeed, in many cases, the scientist may have already determined his "answer" (or understanding) and no output at all is requested.

There are entirely new opportunities to explore. It has been estimated [6] that as much as 90% of the supercomputer time used for code development is wasted due to lack of good

interactive visual feedback. This is illustrated in Figure 2. as the "future paradigm". Human interaction with the supercomputer by means of powerful graphics workstations will also enable or facilitate the solution of computationally difficult problems where the intervention of a human is a key (or possibly essential) part of the process. Examples include algorithms with "controllable" instabilities, problems involving transition boundaries of unknown location, systems containing multiple extrema where a global extremum is desired, and iterative processes that converge very slowly. In these examples, *the scientist becomes an essential part of the algorithm* and is empowered to act at a very high level by serving as part of a complicated non-linear feedback loop, by using knowledge of physical behaviour not reflected in the computer code or by detecting trends permitting a leap ahead to anticipate the converged state.

Graphics super-workstations also create new opportunities at the terminal end of the process shown in Figure 2., i.e., for the presentation of results in the form of "electronic journals". At present, the delay from completion of a report to publication is often 18 months or more. Advanced workstations could be used to communicate scientific results electronically including both still and animated graphics to enhance the timeliness and quality of information exchange between colleagues. Obviously, sending pixel-level graphics would be a gross misuse of precious communication bandwidth and (hopefully) one of the important attributes of graphics super-workstations of the next decade will be the development of standard or compatible means to communicate (or reconstruct!) graphical information.

This attribute of integration/networking is listed in Table 5. which also presents a prediction of other characteristics of the graphics super-workstation one decade hence. Just as supercomputers will achieve great increases in computational power through massive parallelism, so too will workstations. Supercomputer vendors and super-workstation vendors are already adopting many common architectural concepts and technological advances. Massively parallel workstations are a particularly logical development for the graphics processing functions as well as for enhanced computational capability. The predicted 100/1 ratio of processing speed would imply that a TeraFlop supercomputer would co-exist with 10 GigaFlop workstations! This might be achieved by 1000 10MF processors or by 100 processors each of which has a 100MF processing capability. Neither of these models would require a fundamental breakthrough in technology. (It is not unreasonable to expect that the cost of such a workstation might be about the same as present costs as measured in 1989 dollars.)

Note that Table 5. lists ray tracing quality instead of ray tracing since there are already more efficient means to obtain excellent approximations to ray tracing without calculating individual rays. These techniques include radiosity and spatial decomposition. Similarly, there are also relatively simple tools in existence and much more powerful ones under development to facilitate real-time animation.

Of all the attributes predicted, the two most important are integration/networking and ease of use. The integration issue has been addressed previously and Upson [7] clearly captures the essence of the problem of ease of use as follows: ".To date, most software has been developed for programmers and not end-user scientists and engineers.....Designing software for scientists and engineer end-users introduces several new constraints....The most important of these is the need for computational environments that are easy to use and require little documentation....scientists, in general, have exhibited little desire to learn the detail necessary to use graphics.."

Table 5. Attributes of the 1999 Graphics Workstation

WELL-INTEGRATED

With Supercomputer(s)
With Other Workstations
Into Local and Wide-area Networks

SUPERCOMPUTER/WORKSTATION RELATIONSHIPS

Processing Speed $\approx \frac{100}{1}$

Main Memory $\approx \frac{1000}{1}$

Similar Parallel Architectures

SOPHISTICATED SOFTWARE

Powerful Application-Specific Packages

Simple, Real-Time Animation

"Almost" Ray-Tracing Quality

Radiosity

Spatial Decomposition

EASY TO USE BY SCIENTISTS AND ENGINEERS

"Natural" Human Interface

Discipline-Specific Languages

SUMMARY

An overview of the characteristics and capabilities of currently available graphics superworkstations showed that one of the most important uses in a supercomputing environment was the analysis and interpretation of large masses of complex information. This scientific visualization function was not only important in processing the often overwhelming files of output data or observational data, but could be utilized in a much more powerful mode by interacting with a computation in progress.

As the computational power of supercomputers increases during the next decade, it is predicted that the graphics superworkstations will show a proportionate increase and a TeraFlop supercomputer might co-exist with 10 GigaFlop workstations. Both supercomputers and workstations are expected to use similar technology; in particular, future graphics superworkstations will be based on massively parallel architectures. It is suggested that future supercomputers be reserved for attacking the grand challenges of computational science and the computational needs of scientists with less demanding problems be off-loaded to stand-alone superworkstations.

At least two major improvements are needed to fully exploit the great potential of graphics super-workstations. There is an urgent need to resolve incompatibilities between the various vendor products. These differences limit both the portability of graphics code and, of equal importance, the ability of workstations to communicate efficiently over both local and wide-area networks with supercomputers and with other workstations. Probably the single most important improvement needed is to provide for ease of use by scientists who should not be required to learn a new discipline in order to make use of these valuable tools.

In summary, the role of graphics super-workstations in a supercomputing environment includes:

Off-loading the Supercomputer

- Stand Alone Primary Processing
- Post-Processing of Output
- Graphics Processing
- Program Development
- Distributed (Heterogeneous, Shared) Processing

Scientific Visualization

- Understanding of Results (Scientific Insight)
- Communication of Results

Real-Time Interaction

- "Steering" a Computation, e.g., Convergence
- Aborting a Bad Run

It is the view of this author that *true* real-time interaction with a supercomputer by means of a well-integrated graphics super-workstation has the greatest potential pay-off. Thus

The most important role of the graphics super-workstation in a supercomputing environment is to empower the human as an essential component of the solution process.

REFERENCES

1. Levin, E., Eaton, C.K., and Young, B.: Scaling of Data Communications for an Advanced Supercomputer Network. *Data Communication Systems and Their Performance*. pp. 107-122. North-Holland 1988
2. Peterson, V.: Address at the Conference on Grand Challenges to Computational Science. January 1989. (Proceedings to be published in *Future Generations Computer Science* by Elsevier Publishers, Amsterdam)
3. Bretherton, F.: Address at the Conference on Grand Challenges to Computational Science. January 1989. (Proceedings to be published in *Future Generations Computer Science* by Elsevier Publishers, Amsterdam)
4. Gore, Albert, Jr.: Senate Bill S. 1067, *National High-performance Computer Technology Act of 1989*

5. Goldstine, H., von Neumann, J.: *John von Neumann Collected Works, Design of Computers, Theory of Automata and Numerical Analysis* (A. H. Taub, ed.): Pergamon Press 1963
6. McCormick, B. H.: *Issues in Scientific Visualization*. Presented at NATO Advanced Research Workshop on Supercomputing. June 1989. (Proceedings to be published by Springer-Verlag in 1990)
7. Upson, C.: *Scientific Visualization Environments for the Computational Sciences. Proceedings of the 34th IEEE Computer Society International Conference--Compcn Spring 1989*. pp.322-327. IEEE Computer Society Press 1989



RIACS

**Mail Stop 230-5
NASA Ames Research Center
Moffett Field, CA 94035
(415) 694-6363**

The Research Institute for Advanced Computer Science
is an Institute of the
Universities Space Research Association
The American City Building
Suite 212
Columbia, MD 21044
(301) 730-2656