Title
1991 Conference Summary on Computing in High Energy Physics

Author
Loken, S.C.

Publication Date
1991-03-01

LBL-30906

Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

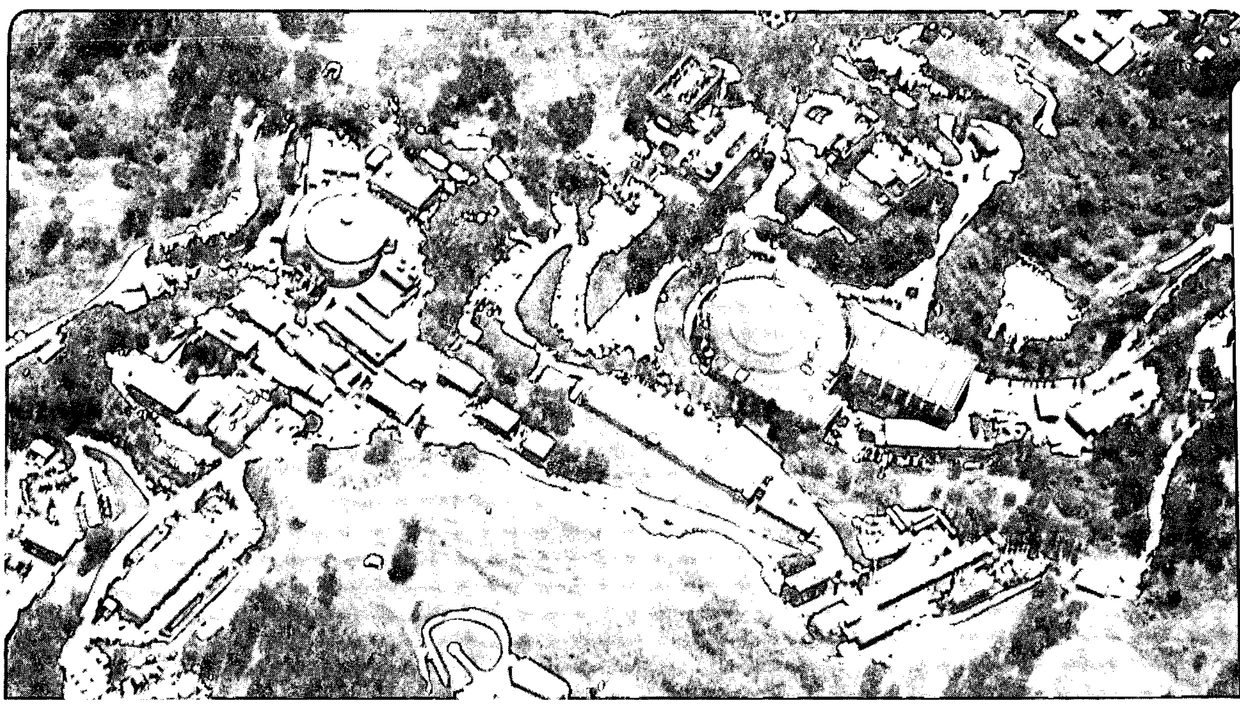
Information and Computing Sciences Division

Presented at the 1991 Conference on Computing in High Energy Physics, Tsukuba City, Japan, March 11–15, 1991, and to be published in the Proceedings

1991 Conference Summary on Computing in High Energy Physics

S.C. Loken

March 1991



Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

LBL-30906

1991 Conference Summary on Computing in High Energy Physics

Stewart C. Loken

Information and Computing Sciences Division
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

March 1991

This work was supported by the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Conference Summary

Stewart C. Loken
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

Abstract

The papers presented at the Conference cover a wide range of important issues in software engineering and management. They indicate a trend toward more use of commercial systems and standards. This trend will likely have a significant influence on plans for future systems.

Introduction

In any review that covers many parallel and plenary sessions, it is necessary to omit important contributions. I have tried to select those areas that seem to indicate new directions for our community and that will be of interest to us in the planning of future experiments.

The issues that I will address here are the following:

- Architectures
- Operating systems
- Networking
- User interfaces
- Software engineering
- Data management

In preparing this summary, I have omitted important contributions in other areas and I apologize to those who have been left out.

Architectures

Many of the presentations at the conference describe approaches to the problem of architecture. They cover the range from the very small up to the very large. The small include special–purpose chips and trigger components. The very large range from individual computers to data acquisition systems and computer networks to address the needs of a single experiment or a full High Energy Physics laboratory.

At the small–system end of the spectrum, there has been considerable progress in special–purpose chips to address the requirements of experiments at the SSC/LHC. At higher levels, neural networks are beginning to be used as part of trigger experiments in the current generation of experiments.

Although High Energy Physics has traditionally used scalar computer systems, we are now seeing some effort to utilize more complex architectures. Vector supercomputers have had some impact on the field but that has been limited to a few areas such as shower simulation. Future supercomputers are likely to built as massively parallel systems. These systems are now receiving the attention of the HEP community. Many groups have reported work on both SIMD and MIMD systems. The results are promising although much remains to be done before we see wide application for these new systems.

The current generation of experiments has invested a great deal of effort in the architecture of data acquisition and data analysis systems as well. The systems are built of large numbers of computers, disks and tapes as well as a network to deliver the data to the physicists. It is clear that these systems are major engineering efforts that require design efforts that are comparable to many other aspects of the experiment. These systems are also critical to the success of the experiment.

In engineering these systems it is important to recognize that they are a complex combination of hardware and software components. Frequently simulations are used to optimize the design of these systems and predict their performance. To develop systems that can be upgraded easily as new components become available it is important that the systems be modular and that hardware and software standards are used wherever possible.

Operating systems

This conference perhaps marks the emergence of UNIX as a part of the High Energy Physics vocabulary. Even those in the community who hoped, or believed, that UNIX would fade out of existence now agree that it will be an important part of the future.

This change is the result of the use of UNIX on virtually every major workstation and computer server. With the rapid evolution of computing technology, vendors have had to adopt a (nearly) standard portable operating system to avoid significant delays in bringing new products to market. This will continue for the foreseeable future.

At this time, most of the use of UNIX systems has been in computer farms and special processors. For these, the number of users has remained relatively small and most of the High Energy community has not needed to learn the details of UNIX. With the increasing use of RISC workstations, however, the situation is changing rapidly.

The proliferation of workstations presents a new set of management challenges. The systems need to be updated and maintained on a regular basis. The community is only beginning to address these issues.

While UNIX will have an important role in High Energy Physics the VMS and VM systems are still used by a majority of the community. At most laboratories, those systems are saturated while the UNIX workstations are idle much of the time. Physicists will use systems that make them productive. They will move slowly unless the UNIX environment can provide the same productivity. Efforts to make the UNIX environment more familiar to the VM or VMS users may help but more likely new, high-level interfaces will change the way physicists work with their computers.

Networking

Networking has become more important as the size and geographical extent of collaborations has increased. The network is now critical to the success of most experiments and even of much of the theoretical work.

In the last few years, the speed of network links has increased significantly. In 1987, the typical link operated at 9.6 kbps. Today, the principal links operate at 1.5 or 2 Mbps within the United States and Europe. There is a 1.5 MBPS link between the United States and CERN. Other international lines typically operate at 64 kbps.

Even with this increase, many local and wide area networks are approaching saturation. This is a result of the use of workstations for local data analysis and the need to look at larger data sets. Local Ethernets are gradually being replaced by FDDI and the speed of wide area networks continues to increase. Even with these improvements, however, there will be places where special–purpose networks will be needed to deliver large data samples to remote computers. These network links will be point–to–point and will likely run non–standard protocols.

There have been other significant changes in networking as well. In 1987, DECNet was the main protocol for the High Energy Physics community. This provided good functionality among the VAX systems that were used by most groups. Today, with the increasing use of UNIX workstations from a variety of vendors, the Internet Protocol, TCP/IP, has replaced DECNet as the dominant network protocol.

At the same time new capabilities are being added to the networks. An example of this is video conferencing. Current technology permits conferencing over links operating at 128 kbps. These systems use video compression and decompression units (CODECs) to provide quasi–full–motion video. Industry and research groups are now working on ways to include video into the same packets that now transmit data. When these systems are operational, video conferencing and multi–media messages will be as ubiquitous as electronic mail is today.

As the complexity and functionality of networks increases it will be even more important to design network interfaces so that the details of routing and services are invisible to the user. The network should make remote services such as data storage systems and computing systems appear as though they are resident on a local computer.

User interfaces

One of the keys to building systems to enhance physicist productivity is the user interface. This is especially true for the casual user who must interact with a new analysis program or control system. There have been significant improvements in the quality of displays and of input techniques.

The improvement in user interfaces is due, in large measure, to the ease with which the interfaces and visualization systems can be developed. Many computer systems now have interface builders or visualization systems. Some examples presented at this meeting include the NeXT Interface Builder and SuperCard for the Macintosh. The Application Visualization System (AVS) from Stardent has been used to develop powerful visualization programs.

Many of the interface building programs are associated with Object Oriented Programming systems (OOPs). The availability of these programs contribute to the growing popularity of OOPs, but is not the only reason to give them serious consideration.

Software engineering

Software engineering is the biggest challenge facing High Energy Physics today. The challenge results, in part, from the number lines of code that must be developed, tested and maintained, and from the volume of data and the complexity of the detector. The most significant factors, however, are the number of people who contribute to the software, the fact that they reside at many different institutions around the world, and that few of them are trained in computer science or programming. These factors would make most programming experts give up before they start.

For some time, there was a belief that the engineering problem was simply one of choosing the right design tools. Some groups adopted Structured Analysis/Structured Design (SA/SD) in the hopes that they would be able to generate correct and maintainable code without doing any extra work. In many cases, they were disappointed and they dropped the methodology.

It is now widely recognized that software engineering is not just a matter of writing code. It is necessary to have a complete system to design, implement, test and maintain programs. It is essential to have a set of tools that support all aspects of the software cycle.

An important issue in software engineering is the question of the programming language. The limitations of FORTRAN have been known for sometime, especially in the handling of data records. Memory managers such as ZEBRA and BOS have been invented and used in many experiments to address the limitations. Looking to the future, perhaps FORTRAN 90 will provide the alternative and eliminate the need for memory managers. It may be, however, that the time has come to look seriously at other languages. In applications such as data acquisition where a smaller group of people contribute to the code, C and other languages are already having a significant impact. Now that most systems support calling modules written in different languages, it is possible to choose the language based on the needs of the application.

Perhaps the most promising alternative to FORTRAN is Object Oriented Programming. Presentations at this meeting show that progress in this area is continuing and new converts have joined the ranks. The REASON project at SLAC served as a demonstration of the potential benefits and a new effort, GISMO, shows that the approach can be useful for detector simulation. The techniques have also been successfully applied to problems as varied as the reconstruction of B–mesons and the slow–control system for L3.

In reviewing this work, the most visible benefit has been the ability to develop very high–quality user interfaces. The real benefit, however, is in the use of classes and inheritance to simplify the description of complex systems. For example, the layers or cells of a chamber inherit the characteristics of the chamber itself. In a similar way, particle properties can be transferred among families with the same quantum numbers.

There does seem to be a significant threshold that people must overcome before they are able to benefit from Object Oriented Programming. So far, a small number of converts have learned to use the techniques and have been very effective. In an experiment, however, more people must become familiar with the methods. Even if they do not write programs they must be able to review designs and code. Another potential problem for experiments is the lack of tools to support software development. Most groups that adopted Structured Analysis/Structured Design found that the lack of tools was the biggest problem in using this methodology.

Although the software engineering effort is directed toward making it easier to write better programs, there is a growing community who feel that the effort should be directed instead at doing physics without writing programs. In particular, they suggest that event analysis can be done with only a graphical user interface. Although there is very important work being done in this area it will be some time before the search for new particles is carried out by clicking a mouse.

Data management

This is, perhaps, the heart of the software engineering problem. The problem for experiments is fundamentally one of handling data and delivering it to the scientists. The present generation of experiments produce terabytes of data and the SSC/LHC experiments will have much more. The data must be delivered to hundreds of physicists and many sites around the world. Depending on the type of analysis, the physicist may want a few variables for millions of events or a complete display of a few events as seen in the detector. The data handling system must support both types of analysis.

Much effort has been devoted to developing systems to address the needs of experiments and to provide general tools. These include data presentation programs as well as databases and file management software. The specification of the IEEE mass storage reference model provides a useful framework for the development of these systems. The IEEE model will likely emerge as the standard for future systems.

There are a number of other issues in the data management area. These include data modeling tools that can assist in the design of data handling systems and query languages that can make data access methods more portable,

In the near future, new approaches such as object oriented databases or extensible databases may make it easier to add relationships among data elements as the understanding of that database improves. As groups begin to address the expected needs of the SSC/LHC experiments there are likely to be dramatic changes in the management of data.

Conclusions

We are beginning to see a major change in the way that physicists do their computing. Computing is now recognized as a major component of experiments and of theoretical work. There is significant effort at the over–all–system level and at the level of individual hardware and software components. As present experiments accumulate larger event samples and as the community prepares for the next generation, the effort of computing systems will become even more intense.

LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
INFORMATION RESOURCES DEPARTMENT
BERKELEY, CALIFORNIA 94720