A History of Computing at Daresbury Laboratory

Rob Allan, Dave Cable, Tim Franks and Paul Kummer
Daresbury Laboratory

e-Mail: robert.allan@stfc.ac.uk

Abstract

The Science and Technology Facilities Council, STFC, is a research council which operates world class large scale research facilities, provides strategic advice to Government on their development and manages international research projects in support of a broad cross section of the UK research community. In partnership with the other research councils, the STFC sets future priorities to meet UK science needs. The STFC operates several world class research centres formerly known as CLRC, including the Rutherford Appleton Laboratory (RAL) in Oxfordshire, the Daresbury Laboratory (DL) in Cheshire and the Chilton Observatory in Hampshire.

We have collected together some facts and figures which give us a picture of the history of high performance computing to support scientific research at STFC over the last 40 years principally focussing on activities on what is now the Daresbury Science and Innovation Campus but also making reference to related activities at RAL (now the Harwell Science and Innovation Campus). The development of applications and numerical algorithms is in many cases far more important than the evolution of computers. Moore’s Law, which applied originally to the number of transistors on an integrated circuit is widely quoted and has knock on effects for performance of whole computing systems. New applications using our growing knowledge of how to represent scientific phenomena at various length scales can however give orders of magnitude improvement in performance and to tackle complex simulations otherwise un-achievable.

We however concentrate on the computers themselves in this retrospective survey showing how the evolving hardware has been used. Most of the information was originally collected around 1995, before the development of distributed computing and data management now known as e-Science. We have added further information more recently covering the early period. We have some supporting artefacts which are occasionally used for open days and school visits to Daresbury Laboratory, see [http://tardis.dl.ac.uk/computing_history/artefacts/dl.catalog.xml](http://tardis.dl.ac.uk/computing_history/artefacts/dl.catalog.xml).

Links to additional material and photographs appear in a Web version which can be found at [http://tardis.dl.ac.uk/computing_history](http://tardis.dl.ac.uk/computing_history). A PDF version of The History of Computing at Daresbury is here [http://tardis.dl.ac.uk/computing_history/computing_history.pdf](http://tardis.dl.ac.uk/computing_history/computing_history.pdf).

© STFC 2007-19. Neither the STFC Scientific Computing Department nor its collaborators accept any responsibility for loss or damage arising from the use of information contained in any of their reports or in any communication about their tests or investigations.
Contents

1 Background to the Science

2 50 Years of Research Computing Facilities
   2.1 Significant Events in STFC Computing History

3 Computer Hardware available to DL Collaborators
   3.1 The 1960s
   3.2 The 1970s
   3.3 The Cray-1
   3.4 The 1980s
   3.5 The 1990s
   3.6 More recently
   3.7 Year 2000 onwards
   3.8 2005-2012: NW-GRID, the North West Grid

4 Distributed Computing
   4.1 The Importance of Networking
   4.2 The UNIX revolution, 1988-1994
   4.3 The Future?

5 2012 onwards. The Hartree Centre and Big Data
   5.1 The UK’s e-Infrastructure
   5.2 The Hartree Centre
   5.3 50 Years of Big Data Impact

6 Acknowledgements

A Notes on Visit made to Daresbury, 12 February 1974
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>B  The Early Days of CCP4, c.1977</td>
<td>53</td>
</tr>
<tr>
<td>C  Archive Images</td>
<td>56</td>
</tr>
<tr>
<td>D  Computer systems at Daresbury over the Years</td>
<td>56</td>
</tr>
<tr>
<td>E  Exhibitions and Artefacts</td>
<td>59</td>
</tr>
<tr>
<td>E.1 Collection Catalogue</td>
<td>59</td>
</tr>
<tr>
<td>E.2 Exhibitions</td>
<td>59</td>
</tr>
<tr>
<td>F  Interesting Facts</td>
<td>59</td>
</tr>
<tr>
<td>F.1 Why does a Computer need to be so big?</td>
<td>59</td>
</tr>
<tr>
<td>F.2 Need to move data around</td>
<td>63</td>
</tr>
<tr>
<td>F.3 Amdahl's Law</td>
<td>64</td>
</tr>
<tr>
<td>F.4 Moore's Law</td>
<td>64</td>
</tr>
<tr>
<td>F.5 Interesting Facts and Figures</td>
<td>65</td>
</tr>
</tbody>
</table>
1 Background to the Science

Daresbury Laboratory supports a very wide range of research in science and engineering using state of the art experimental and computing equipment. Whatever the question, whether it be about superconducting materials, ceramics, biological tissue, pharmaceuticals, reacting fluids or micro-engineering, Daresbury usually has the technology to find an answer.

A “Theory Group” was formed at Daresbury under the leadership of Prof. A. Donnachie of Manchester University in July 1970. (Source: Quest, vol.3 no.3 July 1970). They worked on nuclear physics and carried on until the Nuclear Structure Facility was closed unexpectedly around 1991, after which most of the group moved to work on projects in other countries.

The introduction of the Cray-1 to the UK in 1979 brought other opportunities.

Mary Culligan, in her editorial to the Cray Channels magazine in 1979 [50] had this to say: *While visiting Daresbury I was extremely impressed with the facilities and struck by the high level of interest and enthusiasm exhibited by the diverse staff.*

*There is a strong esprit de corps among the Daresbury physicists and computer experts, engineers, technical craftsmen, and administrators. Central to this feeling is the knowledge that the work being done at the Lab will help provide scientific advances in many fields.*

*Scientists throughout the British Isles can access the Daresbury computer network, either directly at the Lab or through any of a number of universities and research institutes. Thus the Cray-1 at Daresbury is servicing a multitude of scientific disciplines, providing the computing capability necessary to support the links between the growing experimental programs and the theoretical studies.*

We would like to think that, with more modern equipment, this remains true today.

The research background of the scientific staff at the laboratory is the guiding light in our use of technology. Which technology in particular has changed many times over the years from the original NINA electron synchrotron and Van der Graaf NSF (Nuclear Structure Facility) used for nuclear physics to the dedicated SRS (Synchrotron Radiation Source) electron synchrotron and RUSTI (Research Unit for Surfaces Transforms and Interfaces). These experimental facilities have been complemented by IBM, Perkin-Elmer, DEC, Cray, NAS (National Advanced Systems), Alliant, FPS (Floating Point Systems) and later Convex, Intel and Sun supercomputers and distributed systems for theoretical support and data analysis. The intensity with which we pursue research in collaboration with universities and industrial groups worldwide quickly brings about changes to using the latest technology at the lowest cost to our sponsors. Since the early 1980s Cray and IBM have been widely known as supercomputer providers and a plethora of vendors are now satisfying the commodity cluster market. The Distributed Computing Programme under the service level agreement with EPSRC provides help and advice about the current systems to universities all over the UK.

The Theory and Computational Science Division at Daresbury Laboratory was formed on 1/10/1977 by combining the existing Daresbury theory group with the computational atomic physics, quantum chemistry and crystallography group which moved to Daresbury from the Rutherford Laboratory. A report on the first year’s work of the division was published in 1978 [12]. At that time, the theory group was supporting experimental work carried out on the Synchrotron Radiation Source (SRS)
1 BACKGROUND TO THE SCIENCE

and Nuclear Structure Facility (NSF). The computational science group supported the large projects undertaken in partner universities, the origin of the CCPs described below.

The use of computational methods in research has been a major growth area in science over the last (four plus) decades. Many mathematically intractable theories have now given way to calculation from first principles, experiments are analysed with increasing sophistication and real systems of ever increasing complexity can be modelled computationally. With all this new science comes a continual need for more and more computer power and power deployed in a variety of ways – not just number crunching but also experimental control and data collection, visualisation, interactive working and integration with database systems. Until around 1990 it was standard for academic researchers to access powerful remote supercomputer centres, of which Daresbury was one. This has been very successful and will no doubt continue, but there is a new challenge to these centralised facilities.

The explosive growth of the cost/ performance ratio in the workstation to mini-supercomputer range and most recently the availability of cheap and very powerful parallel supercomputers, has enabled individual research groups to run high performance computing systems dedicated to their own projects. As will be clear from what follows, such developments are common to many countries, but we shall concentrate on activities at Daresbury Laboratory to highlight the benefits and potential pitfalls of the distributed computing approach from the point of view of the researcher and user, as well as its impact on the scientific community involved [24].

Computational science at Daresbury in 1994 was mostly associated with three activities: the Synchrotron Radiation Source (SRS), the Collaborative Computational Projects (CCPs) and “externally” funded contracts (including industrial and European Community projects). In 2006 this is different with a growing focus on the new Diamond Light Source in Oxfordshire requiring both simulation and analysis applications embedded in an e-Science infrastructure, and also the growing hope that similar services will be provided again at Daresbury at some future date.

The SRS was the world’s first dedicated source of X-ray synchrotron radiation; it supplied intense collimated beams of polarised light, tunable from infrared to VUV and hard X-ray wavelengths. It was designed c.1975 and operated for 28 years from 1980-2008 [39]. It supported a wide range of experiments including molecular chemistry, surface science, X-ray spectroscopy, biological spectroscopy, protein crystallography, single crystal and powder diffraction and small angle scattering. Many of these experiments need a high level of theoretical analysis to extract the maximum information and understanding from the data, for example angle and spin resolved photoemission, X-ray absorption and atomic and molecular photoionisation. Such analysis is always founded on a quantum mechanical description of the electronic structure and excitation spectrum of the system. Calculation of these quantities is a task of increasing computational intensity as experiments and theory become more sophisticated. Computer simulation of bulk materials can also play a role in understanding structural experiments in complex systems such as superionic conductors and zeolites, where quantum mechanical calculations are inappropriate or still too difficult due to the large number of atoms involved.

The broad discipline of engineering, both structural using finite element analysis techniques and dynamical, e.g. aerodynamics and multi-phase fluid flow through chemical reactors using computational fluid dynamics techniques, now also have a high profile at the laboratory. An intense collaboration involves a number of research groups brought together by the UK engineering community through CCP12, the international ERCOFTAC organisation and a CEC supported project with ICI plc [8].

The origin of the CCPs lies with decisions made at a meeting of the SERC Science Board on
Table 1: The Collaborative Computational Projects

<table>
<thead>
<tr>
<th>Collaborative Computational Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCP1 The Electronic Structure of Molecules (originally entitled Electron Correlation in Molecular Wavefunctions) (1974-2011)</td>
</tr>
<tr>
<td>CCP2 Continuum states of Atoms and Molecules (1978-2011)</td>
</tr>
<tr>
<td>CCP3 Computational Studies of Surfaces (1979-2011)</td>
</tr>
<tr>
<td>CCP4 Protein Crystallography, now called Nano-molecular Crystallography (1979-)</td>
</tr>
<tr>
<td>CCP5 Computer Simulation of Condensed Phases (originally entitled Molecular Dynamics and Monte Carlo Simulations of Bulk Systems) (1980-)</td>
</tr>
<tr>
<td>CCP6 Heavy Particle Dynamics (1980-2011)</td>
</tr>
<tr>
<td>CCP7 Analysis of Astronomical Spectra (1980-no longer active)</td>
</tr>
<tr>
<td>CCP8 Nuclear Physics (no longer active)</td>
</tr>
<tr>
<td>CCP9 Computational Electronic Structure of Condensed matter. Formerly known as Electronic Structure of Solids</td>
</tr>
<tr>
<td>CCP10 Plasma Physics (no longer active)</td>
</tr>
<tr>
<td>CCP11 Biosequence and Structure analysis (no longer active)</td>
</tr>
<tr>
<td>CCP12 High Performance COmputing in Engineering. Formerly known as Parallel Computing in Fluid Dynamics and Numerical Modelling</td>
</tr>
<tr>
<td>CCP13 Fibre diffraction (no longer active)</td>
</tr>
<tr>
<td>CCP14 Powder and Single Crystal diffraction (-2011)</td>
</tr>
<tr>
<td>CCP-ASEArch Algorithms and Software for Emerging Architectures (2011-)</td>
</tr>
<tr>
<td>CCP-BioSim Biomolecular Simulation</td>
</tr>
<tr>
<td>CCP-EM Electron cro-Microscopy (2011-)</td>
</tr>
<tr>
<td>CCPI Tomographic Imaging (2011-)</td>
</tr>
<tr>
<td>CCPN NMR Spectroscopy</td>
</tr>
<tr>
<td>CCP-NC NMR Crystallography (2011-)</td>
</tr>
<tr>
<td>CCPP Computational Plasma Physics (-2011)</td>
</tr>
<tr>
<td>CCPQ Atomic Physics (2011-, formerly CCP2 and CCP6)</td>
</tr>
</tbody>
</table>

10/10/1973 [12]. Prof. Mason presented input from the Atlas Laboratory Review Panel suggesting that “Meeting Houses” should be selected in various areas of science. Dr. Howlett, Director of Atlas, then set up a steering panel chaired by Prof. McWeeny with Profs. Bransden, Murrell and Burke. This resulted in agreement to form the first three CCPs. By 1980 [13], the CCPs had extended to seven groups.

For some recollections of the early days of CCP4 from Dr. Talapady Bhat, see Appendix [3]

Since these early days, the CCPs have played a key role in the way computational science is coordinated in the UK. Each project brings together the major academic (and industrial) groups in a particular field to pool ideas and resources on software developments too large in scale, or too long term, to be tackled successfully by any individual group. The CCP programme currently includes the following projects:

For more information about the current CCPs see the Web site [http://www.ccp.ac.uk]

As with the SRS community, the theme of ab initio quantum mechanics for atoms, molecules and solids
recurs as does simulation and modelling. The CCPs also maintain and distribute program libraries via
the staff at Daresbury and issue newsletters, organise conferences and workshops and support visits
by an ever increasing number of overseas scientists. It is widely accepted that the projects have been
instrumental in keeping UK groups at the forefront of computational science in many fields.

Whilst the outcome of the work is clearly discovery and innovation, the focus of this report is the
computing equipment which was used in the research process, both experimental (sometimes referred
to as *in vivo*) and theoretical (now often referred to as *in silico* or *in virtuo*).

2 50 Years of Research Computing Facilities

Since the early 1960’s the organisations that now constitute the STFC have provided significant com-
puting facilities to UK science. These have included world class computing power, networking, digital
storage, as well as staff to provide not only professional support for the administration of these facili-
ties, but also to develop applications for users. As computing has become more pervasive throughout
society, and the costs have dramatically reduced, universities, then departments and now individuals
can meet their own basic computing needs. STFC continues to provide for exceptionally large scale
needs for the grand challenges of modern research. Today many of the numerical applications con-
tinue to be developed and associated activities are coordinated by the Computational Science and
Engineering Department. Its staff provide a broad range of services to support science across the UK.

2.1 Significant Events in STFC Computing History

The following list illustrates some events which have had an impact on computing at STFC. Later
sections give more information of systems at Daresbury Laboratory.

1767 Foundation of Her Majesty’s Nautical Almanac Office.

1921 The Radio Research Station (later the Appleton Laboratory) was founded in Slough.

1938 Government Code and Cypher School (GC&CS ) established at Bletchley Park (later GCHQ).

1947 Electronic Delay Storage Automatic Calculator (EDSAC) computer being developed in Cam-
bridge by Douglas Hartree and Maurice Wilkes.

1948 First stored program computer built by Williams and Kilburn at Manchester University, known
as the Manchester Mark I [68].

1957 National Institute for Research in Nuclear Science (NIRNS) set up.

1958 The Rutherford High Energy Laboratory (RHEL) was founded as a NIRNS establishment.

1962 The Daresbury Laboratory founded as a NIRNS establishment.

1962 RHEL installed Ferranti Orion computer which ran until 1967.

1963 Construction work started at Daresbury in November. It was to be home to the new National
Institute Northern Acceleator, NINA and the total cost would be £4.5M.

1964 The Atlas Computing Laboratory was established for civil research at Chilton to operate the world’s most powerful computer, the Ferranti Atlas computer, with staff from GC&CS, GCHQ and Harwell.

1965 The Science Research Council (SRC, to become SERC in 1981) was created and took responsibility for the two NIRNS establishments, the Appleton Laboratory and the Royal Observatories at Greenwich and Edinburgh.

1966 Daresbury laboratory installed an IBM 1800 in June and IBM 360/50 in July. The 1800 was a data acquisition and control system while the 360 was used for data analysis, data being stored on magnetic tape. NINA was the world’s first automated accelerator facility.

1967 The Chilbolton Observatory with its newly built 25 metre radar antenna was opened.

1967 Daresbury Laboratory was officially opened by Rt. Hon. the Prime Minister Harold Wilson on 16th June.

1968 Atlas staff produce the first commercially distributed computer animated film. At this time an IBM 360/75 was in use which had been purchased in 1966 at a cost of approx. £1M.

1968 IBM 360/65 installed at Daresbury in November and ran until January 1973. Prof. Brian Flowers also visited Daresbury that month with the Secretary of State for Education and Science.

1971 RHEL install an IBM 360/195 for £2.5M. With a second processor and series of upgrades, this continued in service until 1982.

1972 Twelve private data lines for remote job entry were linked to RHEL from UK universities and CERN.

1972 SRF, synchrotron radiation facility at Daresbury demonstrated the use of radiation as a spin off from NINA for a range of experiments other than nuclear physics.

1973 Chilton Atlas was replaced by an ICL 1906A in April.

1973 IBM 370/165 installed at Daresbury in January

1973 The IBM 360/195 at RHEL became the first machine outside the USA to connect to the ARPAnet, it was also the most powerful machine on the ARPAnet at the time. The connection was made through University College London and then via Norway to the USA.

1974 The SRS, a dedicated Synchrotron Radiation Source, was proposed in May and approved on 12/5/1975. £3M capital was to be made available.

1975 The Atlas Computing Laboratory was renamed the Atlas Centre and merged with the Rutherford Laboratory.

1975 The Appleton Laboratory relocated from Slough to Chilton and merged with the Rutherford Laboratory to form Rutherford Appleton Laboratory.
1975 SRCnet was established linking the 360/195 at Rutherford with the 370/165 at Daresbury and the ICL 1906A on the Atlas site.

1977 NINA facility closed on 1/4/1977 to make way for the SRS.

1978 Permission was granted for the building of the NSF, Daresbury's Nuclear Structure Facility, at an estimated cost of £4M.

1978 The first Cray-1 supercomputer in the UK is installed at the Daresbury laboratory where it ran for around 5 years.

1980 SRS at Daresbury was established as a dedicated synchrotron radiation facility. The facility was inaugurated by the Rt. Hon. Mark Carlisle, Secretary for State for Education and Science on 7th November.

1981 ICL launch the Perq as the first commercial graphics workstation running software developed at RAL.


1981 Nuclear Structure Facility at Daresbury building commissioned.

1981 A mass storage facility is introduced at RAL with 110 GB.

1982 RAL IBM system replaced by the newer 3032 and 3081D.

1982 NSF startup on 10/1982 was followed by its official inauguration by Sir Keith Joseph on 27/9/1983.

1984 SRCnet is extended to become the first UK national computing network – JANET.

1987 A Cray X-MP/48 was brought into service at RAL.


1989-90 Meiko M10 and M60 parallel computers at Daresbury. An Intel iPSC/2 was also installed in June 1990 and used until 1993.

1992 JANET becomes the highest performance X.25 network in the world.


1992 RAL installs one of the first 50 Web servers in the world.

1994 UKERNA was spun out as a private company from the RAL network group to provide the UK’s education and research network JANET.

1995 The UK Research Councils were re-structured. A new research council, eventually named CCLRC, was formed to operate the Daresbury, Rutherford Appleton and Chilbolton Laboratories.

1996 IBM SP/2 parallel computer installed at Daresbury.
1997 The first regional office of the World Wide Web Consortium (W3C) is established for the UK, based at RAL.

1998 Royal Greenwich Observatory closed and Her Majesty’s Nautical Almanac Office moved to RAL.

1998 Usage of RAL mass storage exceeds enough to transfer the names of everybody on the planet each day.

1998 Re-construction of the Manchester Mark I computer and demonstration of the first working module at the Daresbury Open Day. Completed machine is now in the Manchester Museum of Science and Industry.

1999 The Director General of the Research Councils, John Taylor and STFC staff design the UK e-Science programme to focus funding on computationally intensive science that is carried out in highly distributed network environments. This includes science that uses immense quantities of data that require Grid computing.

2001 CCLRC e-Science Centre established to spearhead the exploitation of e-Science technologies throughout CCLRC’s programmes, the research communities they support and the national science and engineering base.

2002 HPCx starts operation at Daresbury Laboratory as the 5th most powerful computer in the world.

2005 National Grid Service established.

2006 North West Grid established with services at Daresbury and universities of Lancaster, Liverpool and Manchester.

2006 RAL Tier-1 Grid computing service for high energy physics sustains 200 Mbyte/s data transfer to CERN to break world record.

2006 Atlas Petabyte Data Store provides 5 Petabytes of storage.

2007 CCLRC merged with PPARC to become STFC.

2010 End of HPCx service.

2011 Government awarded over £145M to a number of projects nation wide to form the UK e-Infrastructure. This included £37.5M to improve the infrastructure and buy new facilities at Daresbury founding the Hartree Centre.

2012 STFC’s Computational Science and Engineering Department and e-Science Centre merged again to form the Scientific Computing Department.

For a history of parallel computing with a general timeline see [47].

For information about the Atlas 50th celebration 4-6/12/2012 see http://www.cs.man.ac.uk/Atlas50/.

For more general computing history and the influence of women in computer science and computation science see here: https://www.certstaff.com/classes/history-women-computers.html (note takes a few seconds to open).
Figure 1: Rebuilt Manchester Mark I
3 Computer Hardware available to DL Collaborators

3.1 The 1960s

The Flowers Report [21] contains comments on the status of computing in late 1965. The following paragraphs refer to the Science Research Council and Daresbury Laboratory, at that time one of its nuclear physics research institutes.

The S.R.C. is unique in possessing two very large nuclear physics laboratories [Rutherford Laboratory and Daresbury Laboratory], created under the National Institute for Research in Nuclear Science, which are operated for and largely by the universities as central facilities. All of the work of these two laboratories is concerned with large accelerators which cannot be used effectively except in close association with correspondingly large computers operating, in part, on line to the accelerators, either directly or through small control computers. The computing requirements of nuclear physics were recently examined by a joint working party of the former D.S.I.R. and N.I.R.N.S., and it was recommended that both laboratories should acquire large central computers of their own.

In the case of Daresbury the issue is clear because there is no other computer available to them for close, constant and reliable operation with the accelerator. A wealth of experience exists to show that in work of this kind computer demand doubles each year and a large expandable machine is therefore essential. There is no British machine of the required characteristics available at the right time, and the application is for an IBM 360/50 together with on line control computers. We support the application and recommend that the order be placed at once, but envisage that this machine may require upgrading to 360/65 standard within two or three years.

In the case of Rutherford High Energy Laboratory the issue might appear not quite as clear because of the presence of the Chilton Atlas nearby of which the Rutherford Laboratory already has one quarter of the use in addition to its own much smaller Orion. The joint working party examined the technical possibility of using Atlas rather than purchasing a further machine. However, the estimated demand exceeds the total capacity of Atlas especially when allowance is made for the fact that in the years 1966 and 1967 there will be a most serious shortage of computer time for the university film analysis groups most of which are involved in the international programme of C.E.R.N. This shortage alone amounts to at least one third of Atlas and the Rutherford Laboratory have taken the responsibility for meeting this demand. Further the Chilton Atlas is making a vital contribution to general university computing, and that of certain Government establishments, the need for which cannot possibly decline before the programme of new provision recommended in this Report is at least in its third year. The Rutherford Laboratory must therefore have a machine of its own.

The very first IBM computer on the Daresbury site, an IBM 1800 data acquisition and control system swiftly arrived in June 1966 and acted as a data logger and controls computer for the NINA synchrotron. It was rapidly followed by the first IBM mainframe computer at Daresbury, an IBM 360/50 which started service for data analysis in July 1966 [52]. This was replaced by an IBM 360/65 in November 1968 as foreseen in the Flowers Report.

This growth in local computing power led to employment opportunities. A note in Computer Weekly of 3/11/1966 reads as follows. If you are a scientist, mathematician or physicist, but want to get into computers, there is rather a good opportunity for you just south of Liverpool. Daresbury Nuclear Physics Laboratory, near Warrington, Lancs, is setting up a computer group to provide a service for
an experimental physics programme with a 4 GeV electron accelerator.

There are vacancies at all levels for programmers or physicist/programmers. Educational requirements are high. Physics or maths graduates or physicists and mathematicians who have completed their Ph.D. thesis are the standard sought.

Another publication noted that the salary would be between £1,000 and £3,107 p.a.

It is worth noting that other IBM systems were in use for academic research purposes throughout the country, with an IBM 360 at UCL (University College London) and a joint IBM 370/67 service for Newcastle and Durham installed by mid 1967.

In the early years the main task at Daresbury was to provide computational power for the nuclear physics groups. Compared to the present, computing was very different in those days. The normal way of telling the computer what work to do was by punched cards, although some stalwarts were still holding out with 5 hole paper tape. Good old FORTRAN was there – what a marvellous language it seemed then, PL/1 was also sometimes used. Typically one prepared a job on punched cards and placed it on a trolley. Later an operator would come along, bringing back the previous trolley load of punched cards and the line printer output that had been produced. Turn around was measured in tens of minutes at the very least. The mean time between failures towards the end of the 1960s was a day. However these computer crashes were “unseen” by the users who were waiting impatiently for the trolley to re-appear. This was an early form of batch service where jobs ran one after another.

3.2 The 1970s

The IBM 360/65 was replaced by an IBM 370/165 in January 1973 [17]. This had a stunning 12.5 MHz cpu and 3 MB of main memory, which was actually a lot for the time.

1973 also saw the arrival of TSO as an interactive service. It was one of those changes, like seeing colour television for the first time, which was not only here to stay but which could transform the whole character of computing. Several users could simultaneously edit programs interactively (cards no longer required), compile them and submit them to the batch queue.

Some more photos and information about the IBM 370 can be found here.

A PDP-11/05 computer was used as a central network controller at Daresbury from 1974-80. This had a single 16 bit 100 kHz processor and 16 kB micro core memory. They were popular at the time for real time applications and the C programming language was essentially written for this type of computer. The first version of UNIX ran on a PDP-11/20 in 1970. In many ways the PDP-11 could be considered to be the first “modern” computer. In addition to local terminals at Daresbury this one supported remote job entry (RJE) stations at Daresbury, Lancaster, Liverpool, Manchester and Sheffield. There was also a connection over the SRCNet packet switching network to a GEC 4080 computer fulfilling a similar function at the Rutherford site.

In Oct’1977 a cluster of six GEC 4070 computers was bought from Borehamwood at a cost of £170k. These had core memory and were deployed for nuclear physics data analysis and control in the NSF. They did not come on line for users until the early 1980s, but were then kept running and upgraded over a long period. They were particularly suitable for handling large quantities of data streaming from
Figure 2: Original Computer Hall
the experimental facility. Further data analysis was carried out on the IBM 370/165. Four of the GEC machines were connected in pairs, each pair forming a “data station”. One was used interactively and the second for preparation of the subsequent experiment. These machines were connected to printers, plotters, tape units and graphic display units through the other two systems.

Some more information and photos of the PDP-11 and LSI-11 can be found here.

Some more photos of the GEC 4000 cluster and NSF control room can be found here.

Some notes on a visit to Daresbury in 1974 by F.R.A. Hopgood and D.G. House of the Atlas Computing Laboratory is re-produced in Appendix A. This gives a flavour of the environment at that time.

In 1975 the SRCnet was established linking the 360/195 at Rutherford with the 370/165 at Daresbury and the 256 kB ICL 1906A on the Atlas site. Also at that time two Interdata model 85s each with 64 kB memory were purchased at £100k to front end the IBM at Daresbury. All systems were linked using CAMAC. There were also experiments with microwave communication between different Laboratory buildings.

Starting in 1974, the SRS control system [39] was designed to use a two level computer network linked to the Daresbury site central computer, the latter being used for software preparation and applications requiring bulk data storage. Four mini computers were used, one dedicated to each of the three accelerators, the fourth providing controls for the experimental beam lines. These computers with Perkin-Elmer (Interdata) systems models 7/16s and 8/16s with 64 kB memory each. These were 16 bit computers running the proprietary OS/16-MT real time operating system enhanced to support
CAMAC and network communications.

It was suggested [39] that these computer systems would be upgraded and used for eleven years, at which time (1985) the 16 bit machines would be replaced by two 32 bit control systems with a larger 32 bit central computer. These were Perkin-Elmer 3200 series systems (3205 and 3230 respectively) which became available from 1982, Perkin-Elmer had by then spun off Concurrent Computer Corporation which sold these machines. A further upgrade was being planned in 1994 [36].

By the time the SRS commenced operation in 1980 [38], the main control computers were Perkin-Elmer (Interdata) 7/32s with 320 kB memory. These was linked to three or more 7/16s, each of which oversaw elements of the SRS. In total there were two 7/32s and ten 7/16s which were ordered in 1976. The 7/32s initially has 128 kB of memory, 20 MB of disc, a line printer and three VDUs. The 7/16s had only 32 kB memory. CAMAC was also used to interface everything. The 7/32 was a 32 bit machine running OS/32-MT, this machine could also be used for plotting graphs of SRS operations parameters. Applications, mainly for plant control, were programmed using the real time RTL-2 language first used at Daresbury in 1974 in preference to Coral 66. Whilst every effort was made to ensure the whole system was reliable, it was noted that the mean time between failures of a 7/32 was typically a few days. In addition to Perkin-Elmer, a number of Honeywell H-160 computers were in use for data collection and local instrument control.

### 3.3 The Cray-1

The first Cray-1 supercomputer in the UK was installed temporarily at Daresbury Laboratory in November 1977. This was on loan from Cray Inc. and was one of the first Cray vector supercomputers outside the USA [66]. It was actually Cray serial number 1 (SN1) which had been installed at Los Alamos National Laboratory, USA for a six month trial in March 1976. Over the next two years it was upgraded to a Cray-1A and then Cray-1S/500 SN28 which finally became a 1S/1000 with 1 M word (64-bit) memory. The front end system was the IBM 370/165 with the PDP-11 still acting as a packet switching gateway. Not only was the Cray more powerful than the IBM, but was said to be more user friendly, in particular offering metter diagnostics for debugging programs.

The IBM was replaced with the AS/7000 in 1981 fulfilling the same purpose but with an increase in performance and throughput. The PDP was upgraded with the addition of a GEC 4065 in 1981 and GEC 4190 in November 1982. A contemporary account read as follows.

*The Cray-1 system is a high speed computer well suited to the needs of the scientific community which Daresbury Laboratory serves. It is a large scale, general purpose digital computer which features vector as well as scalar processing; it achieves its speed with very fast memory and logic components, and a high degree of parallel operations. The basic clock period is 12.5 nanoseconds, ... the memory cycle is 50 nanoseconds. The machine can operate at speeds in excess of eighty million calculations per second.*

The Cray-1 system at Daresbury consists of a processor with half a million (64 bit) words of memory, a maintenance controller and four large scale disk drives, each capable of storing 2,424 x10**9 bits of data. The system will be linked directly to the SRC network through the IBM 370/165.

---

1RTL was developed by ICI for laboratory and process control, but the Department for Trade and Industry at the time recommended Coral. Daresbury collaborated with ICI on further development.
Areas of science in which the Cray-1 computer system will be used include plasma physics, oceanography and engineering where three dimensional modelling can be undertaken. More complex study than is now possible will be done in protein crystallography, atomic physics and aerodynamics. The enhanced computing capabilities will advance research in astrophysics, nuclear theory and theoretical chemistry.

Users of the system will be drawn from many Universities and research groups through the United Kingdom.

It was reported in the Chester Chronicle on 28/8/1981 that 300 “boffins” from around the world had met at Chester College for a conference entitled Vector and Parallel Processors in Computational Science. It was organised by staff from Daresbury and chaired by Prof. Mike Delves, a mathematician from University of Liverpool. The range and success of applications of the Cray is evident from the collection of papers published in 1982 [14]. The Cray system was thus demonstrated to be of tremendous benefit in physics and chemistry applications, with help from staff at Daresbury tuning codes to run on it, in some cases re-writing in machine code [18]. This was also the start of the Collaborative Computational Projects mentioned above.

The Cray system was bought outright for the UK and moved to the University of London Computer Centre in May 1983 where it was controlled by an Amdahl V8 front end machine. ULCC had been established in 1968 as a consequence of the Flowers Report. A full service for the 738 active users continued with an additional second hand Cray-1 (a Cray-1B?) in 1986 as recommended in the Forty Report [22]. It is possible that the size of memory of the ex. Daresbury Cray was also doubled at this time. The two Crays then became known by users as “the Cray Twins” or “Ronnie and Reggie” and ran until 1989 when they were replaced by a Cray-XMP/28.

Daresbury often pioneered new technology, but others put it into service. SN1 eventually returned to the USA and is now in the Cray Museum at Chippewa Falls, see [http://www.craywiki.co.uk/index.php?title=Cray_museum](http://www.craywiki.co.uk/index.php?title=Cray_museum). Some more information and photos of the Cray-1S can be found here.

In those days data and programs were stored on magnetic tape, there were thousands of them stored for our users in the computer room. Output could be provided on tape or printed on line printer paper which users collected from a hatch. Both storage and user interfaces have come a long way since.

Some more photos of the main Computer Hall at Daresbury can be found here.

### 3.4 The 1980s

It was noted in 1981 [31] that the NSF data handling system which consists of a network of five GEC 4070 processors and associated CAMAC equipment is nearing the point at which it can be tested as a complete system. Initially, there will be two independent data taking stations or event managers into which data from the experimental electronics will be organised and buffered before being transferred via a serial highway to the processors for sorting. There will be five graphics terminals for on line or off line interactive analysis of experimental data. The data handling system is linked, via the Daresbury central IBM 370/165, to workstations situated in user universities.

A National Advanced Systems NAS AS/7000 (IBM “clone”) from Hitachi was installed in June 1981
Figure 4: Cray 1 at Daresbury in 1978. The IBM 370/165 console is in the foreground.
Figure 5: Part of tape store at far end of computer room in 1979.

as a central computer. It gave an enormous increase in power, a significant increase in reliability and facilitated the move to the then modern operating system MVS. The NAS AS/7000 was a Japanese equivalent of an IBM 370 running IBM’s MVS operating system at around 2.7 MIPS. The system was installed in 1981 and had 8 MBytes of main memory, later upgraded to 16 MB. At the same time the network packet switching was upgraded to a GEC 4090, the NAS and GEC together costing some £850k. Dr. Brian Davies, head of Daresbury’s Computer Systems and Electronics Division, explained: *We have had an IBM 370/165 for about eight years now, which has been the workhorse for the laboratory. It is attached to the SERC network which includes most universities in the UK. All the applications are scientific, using Fortran, with a large scientific batch load and about 50 concurrently active terminals working under IBM’s Time Sharing Option.*

FPS-164 and 264 array processors were in use between 1987-91. The FPS-164 had an 11 MHz 64-bit processor and 4 MB memory. FPS systems could be fitted with MAX boards for matrix operations, each giving 22 Mflop/s. Up to 15 of these could be fitted to a single FPS host. Three MAX boards were installed in the FPS-164/MAX at Daresbury. The more general purpose FPS-264 performed at 38 Mflop/s but did not have MAX boards. These FPS systems were installed based on the recommendations of the Forty report [22] which noted the strong demand for special purpose computers at the time.

On 16/12/1988 the AS/7000 user service was terminated, signalling the end of over two decades of IBM compatible computing at Daresbury. Substantial amounts of scientific computation had been completed, so everyone involved in the service joined in a traditional wake, by no means an entirely sad occasion, to mark its ending.
Figure 6: Comparison of 32 kbits of 2 us ferrite store from the IBM 370/165 (left) with 1 Mbits 360 ns MOS store from NAS AS/7000 (right).

Figure 7: The FPS-164/MAX shortly after installation in 1987.
The transputer which came on the market in 1985 was programmed using a special parallel language called Occam. Some exhibits show a selection of the hardware used in transputer based systems, instruction manuals and samples of Occam code.

The Meiko M10 was the first explicitly parallel computer at Daresbury, requiring applications to be significantly re-written. It had 13x T800 transputers which comprised the hardware used at the Laboratory in its status of “Associated Support Centre” to the Engineering Board- DTI Transputer Initiative. The Inmos Transputer was a British invention, hence the interest from the DTI. The system (on loan from the Initiative) was upgraded to its final size in January 1989 with the installation of three “quad boards”. A large amount of software was developed on this system. This included SRS data analysis software MORIA and XANES but also the Occam version of the FORTNET message passing harness which enabled other application codes written in Fortran to be used on Transputer systems. The system was however obsolete and switched off by 1994.

The FPS T20 had 16x T414 transputers with additional vector processing chips to give higher performance. It was operational under Ultrix on a microVax with JANET access. The Occam-2 compiler for FPS was developed at Daresbury by Bill Purvis. A vector maths library was also being written. A high speed interface was developed allowing data to be transferred between the Meiko M10 and the FPS, for instance allowing the FPS to access the graphics hardware in the Meiko. This machine however never realised its contractual upgrade path and when in May 1988 it was no longer supported by Floating Point Systems it was replaced by the Intel iPSC/2.

More information about Meiko parallel machines can be found here.
3.5 The 1990s

Following work on early transputer systems, parallel processing at Daresbury as a service focused around Intel iPSC/2 and iPSC/860 hypercube computers with respectively 32 and 64 processors and a Meiko M60 computing surface with 10 processors used for development work. Each node of the iPSC/860 and M60 consisted of a high performance Intel 64 bit i860 microprocessing chip, memory and an internal network interface to send data to other nodes. For a period of around nine months an 8 processor Alliant FX/2800 system was maintained at the Laboratory which it was developed into a stable system to be installed in the Engineering Department at the University of Manchester. There was a smaller system in use at Jodrell Bank. Descriptions of Parsytec transputer systems and other hardware on test have also appeared from time to time [53].

From 1990 to 1993 the Intel iPSC/860 evolved from an initial parallel development system to a resource acting as the focus of a National Supercomputing Service. The iPSC/860 system had a peak performance of around 2.5 Gflop/s, equivalent to the Cray Y-MP/8I at RAL but at considerably lower capital cost. Some 100 users were registered on the Intel with 140 projects having made use of the system, many of which were surveyed in “Parallel Supercomputing '92” and “Parallel Supercomputing '93. The associated newsletter, “Parallel News”, was circulated to around 1,000 people, providing up to date information on the service, together with a variety of articles, programming hints etc.

More information about Intel parallel machines can be found here.

Daresbury was thus instrumental in demonstrating the successful operation of a national parallel supercomputing service. Another service for the “grand challenge” projects in molecular simulation and high energy physics was operated with a similar 64 processor Meiko M60 system at the University of Edinburgh Parallel Computing Centre. During 1993 it was decided by the Advisory Board to the Research Councils, after much consultation and independent tendering and benchmarking exercises, that a new parallel supercomputer should be bought and installed at Edinburgh. This was a 256 processor Cray T3D system. Again Daresbury had pioneered the use of a novel type of computer system which was later installed at a different site.

Circa 1995, staff at Daresbury and visitors had access to many on site computers and to the Cray Y-MP/8I at RAL. The on site computers were by now linked by a local area network consisting of Ethernet, FDDI and dedicated optical fibres. Access from outside the laboratory was via the JANET service to local hosts such as the Convex and the Sun workstation which was a front end for the Intel hypercube. Some of the systems then in use are listed below.

Multi-user Data Processing and Computation

**Convex**: C220 – this early Convex C2 system was installed in 1988 initially with one processor and 128 MBytes of main memory. It was some four times faster than the AS/7000. Early in 1989 a second 200 MHz processor and further 128 MBytes of memory was delivered. This two processor system was in use until 1994.

**Vax**: 3800, 11/750 and 8350.

**Silicon Graphics**: 4D/420 – quad-processor MIPS 3000 system with dedicated visualisation hardware.
Sun Microsystems: 470/MP – dual-processor system (front end to Intel).

Distributed Memory Parallel Computers

Intel: iPSC/2 – 32 SX/VX nodes were provided by the SERC Science Board in 1988 for this second generation Intel hypercube. Each node had an Intel 80386 plus Weitek 1167 floating point processor set and AMD VX vector co-processors with a total of 5 MByte of memory split between the two processor boards (4 MByte scalar and 1 MByte vector memory). A further 32 SX nodes and concurrent i/o system with 1.5 GBytes of disc were funded by ICI plc. The system was reduced again to 32 nodes to fund ICI’s stake in the iPSC/860. The system became obsolete and too expensive to maintain as a service in 1993, it was however still being used in 1994 for system code development.

Intel: iPSC/860 – 64x Intel i860 nodes. Initially 32 nodes were bought with Science Board funding and money from the trade in of ICI’s iPSC/2 nodes. The system was installed in June 1990. It was upgraded to 64 nodes by the Advisory Board to the Research Councils (ABRC) Supercomputer Management Committee (SMC) in 1993 following the first successful year of a national peer reviewed service.

Meiko: M60 Computing Surface – 4x T800 Transputer nodes were moved from the M10 and 10 Intel i860 nodes were added in a new Meiko M60 chassis funded by Engineering Board in 1990 to support similar systems in a number of UK Universities. This ran until 1993.
When it was first installed, the Intel iPSC/860, much like its predecessor the iPSC/2, remained very much an experimental machine. Indeed it was purchased to investigate the feasibility of parallel computing, but it proved to be a much valued resource by researchers, remained in great demand, despite the introduction of alternative facilities at other UK sites. The rapid conversion of a number of scientific applications to run on the machine provided that parallel computing was a viable technique and was by no means as difficult as some were claiming. Even early teething troubles with the operating system did not deter users from exploiting the power of the machine.

The Intel iPSC/860 was made generally available as a UK wide service in January 1992, when Access to the machine was by peer reviewed grant application, and the take up was immediate, with 40 applications awarded a total of 100,000 node hours in the first year. Subsequent years saw the applications swell to 60, with a total of 320,000 node hours in 1994.

The Intel was essentially a collection of 64 single board computers, which operated in parallel and communicated with each other over a high speed inter-connection network. Although the connection had a fixed hypercube topology, the machine differed from earlier ones by having “intelligent” message routing, so that the programmer did not have to worry about the details of the topology and could pass a message between any pair of processors without significant overhead. Nowadays, all parallel machines use similar techniques and topology has become a detail that concerns only manufacturers and system administrators.

When first delivered, in March 1990, the iPSC/860 consisted of 32 processors, each with 8 MB of memory. This limited node memory provided arguably the major bottleneck in the process of developing parallel applications, a situation aggravated by the Intel developed operating system (NX/2) not supporting virtual memory. Driven by the early impact of the machine, EPSRC’s Science Board
funded an upgrade in October 1991 that increased the memory to 16 MBd per processor. A subsequent upgrade to 64 nodes provided an even greater increase in the power of the machine.

In addition to the processing nodes, a number of additional processors were provided to handle input and output to a collection of disks. Initially, 4 disks each of 750 MB were installed, a configuration that was later upgraded to 8 disks. When the machine was extended to 64 nodes, a simultaneous disk upgrade replaced all 8 disks with 1.5 GB disks with improved performance and reliability (as well as doubling the capacity). The i/o subsystem also included an Ethernet port so that the i/o system could be accessed directly from the network using the FTP protocol.

**Shared Memory Parallel Computers**

**Alliant:** FX/2808 – eight Intel i860 processors sharing memory over a crossbar switch. This was on loan for a period of evaluation before being transferred to University of Manchester.

**Workstation Clusters**

**Apollo:** DN10020 – two systems with a total of five processors running at 18 MHz and a shared disc pool. Apollo was a rival of Sun building graphics workstations in the mid 1980s.

**HP:** 9000 Series Models 750 and 730 – 5 nodes.

**IBM:** RISC/6000 Series Models 530H – 3 nodes.

In the HP cluster, individual machines were connected with an FDDI network. They were used for parallel computing with the MPI, PVM or TCGMSG software, the latter was developed at Daresbury. The “head” node was an HP model 755 with 40 Mflop/s peak performance and 128 MB memory and 4.8 GB disk (of which 2GB was scratch space). “Worker” machines were HP 735 models with 40 Mflop/s performance, 80 MB memory and 3 GB scratch disk. Users would log into the head node and submit jobs to the worker nodes using DQS – the Distributed Queuing System.

**Super Workstations**

**Stardent:** 1520 – (formerly Ardent Titan-2) two systems with a total of three processors. The two processor machine was bought with industrial funding in 1989 to run molecular modelling software (such as BIOGRAF) which requires good graphical facilities, principally for the simulation and modelling of polymers. Both machines were switched off in mid March 1994 as they had become obsolete and too expensive to maintain.

**Silicon Graphics:** SGI 4D/420-GTX – upgraded to four processors at the end of 1992.

**IBM:** RISC/6000 model 530H – dedicated for development and support of the CRYSTAL code.

**DEC:** Alpha – one had been in use for development of advanced networking methods since 1993, two more systems were purchased for the CFD group to support development of code to run on the Cray T3D parallel system (which also uses DEC alpha processors) in 1994.
Desktop Systems

Some 100 systems are in use, mostly Sun Sparc and 3/60 models, Silicon Graphics Indigo and DEC Workstations but also IBM PowerPC models and IBM and Apple PCs.

Beowulf Computers

The first commodity cluster, then known as a Beowulf (from Norse mythology) was built in 1994 and used as a test and development system until 1998. This was built from off-the-shelf PCs linked together with ethernet – a total of 32x 450 MHz Pentium III processors. It occupied a hefty rack and the cables looked like coloured spaghetti. The PCs each had memory and disk but no keyboard or monitor. They were connected by dual fast Ethernet switches – 2x Extreme Summit48, one network for IP traffic (e.g. nfs) and the other for MPI message passing. Additional 8-port KVM switches can be used to attach a keyboard and monitor to any one of the nodes for administrative purposes. The whole cluster had a single master node (with a backup spare) for compilation and resource management.

Some of the applications running on these systems and the national Cray T3E supercomputing facility in Edinburgh, many arising from the work of the CCPs, were showcased at the HPCI Conference, 1998 [7].

Graphical Visualisation and Program Analysis

A contemporary account read: We now have the possibility to log into a graphical workstation to “dial up” another machine with intensive parallel or vectorial capability. We can make it perform a task and display the results as they are calculated, perhaps over a number of computational steps, or modify parameters of the problem between each frame to study the effects. A software package called DISPLAY has been developed. It is intended as a graphical “shell” or “front end” for computational codes of all types and has a simple menu based interface which the user may configure to run his or her application. The application can be run as one component of a parallel task within the distributed...
Figure 12: Sun 470 console in 1989

Figure 13: So called Beowulf Cluster of PCs
Figure 14: Loki Cluster

Harness.

Other graphical front end software was used with Silicon Graphics workstations to ease interactive use of programs maintained by the data analysis group for our SRS users.

Performance analysis of complex parallel programs is also essential to obtain good efficiency. Graphical tools, such as those provided on the Intel hypercube, originally part of the Express programming harness, helped to do this.

3.6 More recently

Loki and Scali Clusters

A more adventurous cluster known as Loki (also a Norse name) was purchased in 1999 and ran until 2003. This eventually had 64x high performance DEC Alpha EV6/7 processors running in 64 bit at 667 MHz. Each processor had 1/2 GB memory. This was a big step up from the earlier systems. The master node had a similar processor but 1GB of memory and a 16GB SCSI disk.
Loki was purchased for departmental use and initially contained 17x dual processor UP2000 boards from API including the master node. The principal interconnect solution was the proprietary Qnet, in which the network interface is based on QSW’s Elan III ASIC. A secondary fast Ethernet was used for cluster management and a further Ethernet was available for communication traffic enabling comparison of the two interconnects on the same hardware platform.

Thanks to support from EPSRC Loki was upgraded to contain 66 Alpha processors including the master node in 2001. The new dual processor nodes were UP2000 and CS20 from API. This has necessitated upgrading the switch from the 16 node version to one with the potential to support up to 128.

The slightly larger Scali cluster was bought in 2003 and used 64x AMD K7 processors, now with a huge 1 GB memory each – remember, this was before the demanding Java era and memory was still relatively expensive.

Both Loki and Scali had high performance switched networks connecting the nodes for parallel computing applications as opposed to the earlier quad or hypercube networks.

IBM SP2

An IBM SP2 machine, also essentially a cluster, was installed in mid-1995 and used intensively for development purposes [4]. It paved the way for the introduction of the HPCx national supercomputing service based at Daresbury.

SP stands for “Scaleable POWERparallel”. This SP2 consisted of fourteen so called IBM Thin Node 2s each having 64 MB memory and two wide nodes each with 128 MB. Each RS/6000 node had a single POWER2 SuperChip processor running at 66.7 MHz and performing at a peak of 267 Mflop/s from its quad instruction floating point core. The nodes, which were the first of their type in the UK, were connected internally by a high performance internal network switch. They were housed in two frames with an additional Power PC RS/6000 control workstation.

HPCx

HPCx started operation at Daresbury in 2002 as the 9th most powerful computer in the world (20th Top500 list, see http://www.top500.org). At that time it was an IBM p690 Regatta system with 1280x 1.3 GHz POWER4 processors and a total of 1.3 TB memory. The HPCx system was located at Daresbury Laboratory but operated by the a consortium legally entitled HPCx Ltd.

By 2007, the HPCx system had been upgraded to IBM eServer 575 nodes for the compute and IBM eServer 575 nodes for login and disk I/O. Each eServer node then contained 16x 1.5 GHz POWER5 processors. The main HPCx service provided 160 nodes for compute jobs for users, giving a total of 2560 processors. There was a separate partition of 12 nodes reserved for special purposes. The peak
computational power of the HPCx system was 15.3 Tflop/s.

The frames in the HPCx system were connected via IBM’s High Performance Switch (HPS). Each eServer frame had two network adapters and there were two links per adapter, making a total of four links between each of the frames and the switch network.

The HPCx service ended on 31/1/2010 after a very successful and trouble free eight years. For more information about HPCx and the work carried out upon it see the Capability Computing newsletters on the Web site [http://www.hpcx.ac.uk/about/newsletter/](http://www.hpcx.ac.uk/about/newsletter/).

Some more information and photos of HPCx can be found here.

### 3.8 2005-2012: NW-GRID, the North West Grid

Computing collaborations in the north west of England go back a long way. Since the start, Daresbury had good links with local universities and research institutions. For instance POL, the Proudman Oceanographic Laboratory in Bidston, Wirral was linked into the Laboratory’s IBM 370/165 in the early 1970s. Their own IBM 1130 was replaced with a new Honeywell 66/20 in 1976. This was used to
set up the British Oceanographic Data Service among other things.

Lancaster University, founded in 1965, was also an important local player. Back on 29/5/1975 Dr. Ken Beauchamp, their director of computer services, was reported in Computer Weekly as saying *users don’t care where their computing power comes from, as long as its available.* At that time the university was operating a six year old ICL 1905F computer with 128 kB memory plus a number of micro-computers in various departments. This was augmented with an ICL 7905 in July 1975. Users also had access to ICL systems at Manchester University and to a new ICL 1906S at Liverpool University, as did Salford and Keele. All these local universities were also using the IBM 370/165 at Daresbury via remote job entry links.

After lengthy discussion and negotiation, NW-GRID was established in April 2005. It initially had four core Sun compute clusters from Streamline Computing funded by NWDA\(^2\) which were accessible using Grid middleware and connected using a dedicated fibre network:

- Daresbury – dl1.nw-grid.ac.uk – 96 nodes (some locally funded);
- Lancaster – lancs1.nw-grid.ac.uk – 192 nodes (96 dedicated to local users);
- Liverpool – lv1.nw-grid.ac.uk – 44 nodes;
- Manchester – man2.nw-grid.ac.uk – 27 nodes

These were installed in 2006. The clusters all had 2x Sun x4200 head nodes each with dual-processor single core AMD Opteron 2.6 GHz (64-bit) processors and 16 GB memory. These had Panasas hardware supported file store of between 3-10 TB per cluster (Manchester had client only). The compute nodes were all Sun x4100 dual processor dual core AMD Opteron 2.4 GHz (64-bit) with 2-4 GB memory per core and 2x 73 GB disks per node. They were connected with Gbit/s ethernet.

Additional clusters and upgrades were procured at the core sites, helped by a second phase of NWDA funding as agreed in the original project plan. Other partners were encouraged to join and share their resources. This resulted in the following configuration in late 2008.

- Daresbury: 96 nodes 2.4 GHz twin dual core CPU;
- Lancaster: 48 nodes 2.6 GHz twin dual core CPU;
- Lancaster: 67 nodes 2.6 GHz twin quad-core CPU;
- Liverpool: 104 nodes, 2.2 GHz twin dual core and 2.3 GHz twin quad-core CPU;
- Liverpool: 108 nodes, 2.4 GHz twin dual core and 2.3 GHz twin quad-core CPU and InfiniPath network;
- Manchester: 25 nodes 2.4 GHz twin dual core CPU plus other local systems;

Daresbury, Lancaster and Liverpool have 8 TB of storage accessed by the Panasas file servers. In addition to this, there are RAID arrays of 2.8 TB at Manchester and 24 TB at each of Lancaster and Liverpool. Nodes are connected by separate data and communications interconnect using Gigabit Ethernet and Liverpool’s second 108 node cluster is connected with InfiniPath.

Around this core are other computer systems that are connected to the NW-GRID.

- Daresbury: IBM BlueGene-L (2048 cores);

\(^2\)NWDA: North West Development Agency was the government regional development agency for the north west of England.
Daresbury: IBM BlueGene-P (4096 cores);
Daresbury: 2560 node IBM 1.5 GHz POWER5 (HPCx – subject to approval);
Daresbury: 32 node Harpertown cluster with Nehalem processors and nVidia Tesla GPUs;
Daresbury 32 node Woodcrest cluster with ClearSpeed accelerators;
Lancaster: 124 node Streamline/ Sun cluster 2.6 GHz twin dual core;
Liverpool: 96 node, 196 core Xeon x86 cluster, contributed by Proudman Oceanographic Laboratory, with 5.7 TB of GPFS storage;
Liverpool: 960 node Dell cluster, Pentium IV processors (Physics);
Manchester: 44 node dual processor Opteron cluster, 2.5 TB RAID storage based on 2 GHz Opterons with 2 GB RAM;
Manchester: SGI Prism with 8 Itanium2 processors and 4x ATI FireGL X3 graphics pipes. There was a similar system at Daresbury;
University of Central Lancashire (UCLan): SGI Altix 3700 with 56x Intel Itanium CPUs and an Intel based cluster with 512 cores of 2.66GHz Xeon processore;
Huddersfield: several Beowulf style clusters plus a share of the enCore service hosted at Daresbury (see below). Huddersfield later acquired part of the original Lancs1 NW-GRID system.

It should also be noted that there are still many separate computer systems in use for academic research in the region. A somewhat separate grid of large compute clusters is dedicated to high energy physics users, specifically for analysis of data from the Large Hadron Collider at CERN. This is known as NorthGrid and has partners from physics departments at Lancaster, Liverpool, Manchester and Sheffield with support from the networking team at Daresbury Laboratory.

Further developments are explained below where we refer to the UK e-Infrastructure.

Sun Cluster

Some more information and photos of the NW-GRID Sun cluster at Daresbury can be found here.

IBM BlueGene

Two of the futuristic BlueGene systems were installed at the Laboratory, a BlueGene/L and a BlueGene/P.

Some more information and photos of BlueGene can be found here.

iDataPlex

Known as SID for STFC iDataPlex, this was a partnership cloud service for University of Huddersfield, STFC staff and for commercial partners via an agreement with OCF plc. The OCF commercial service was referred to as the enCore Compute Cloud, see http://www.ocf.co.uk/what-we-do/encore.

SID comprises a double width compute rack containing 40x nodes. Each node contains 2x six core 2.67 GHz Intel Westmere X56 processors with 24 GB of memory and 250 GB local SATA disk. Two of them have 2x attached nVidia Tesla GPU cards. The nodes are connected by a high performance QLogic Infiniband QDR network switch and GPFS storage.

Like the SP2, this system paved the way for something larger.
Figure 16: NW-GRID Daresbury Sun Cluster.
Figure 17: STFC iDataPlex (SID)
4 Distributed Computing

4.1 The Importance of Networking

Integration of computer systems and other electronic components has long been important for both local and remote users.

As noted in CERN Courier, Aug’1973, an important reason for adopting a distributed computing philosophy is to avoid the need to build large specialised systems for particular experiments which would be expensive in terms of equipment and man hours. The network can also pump data directly into the main computer and for later storage on tape rather than moving tapes from one place to another. Stored data can then be made available across the network for different purposes without duplication. In the 1970s, the Daresbury network was designed to enable many users simultaneous access to equipment without interference between them. Local computers and specialist software are however kept to a minimum and there is no local data storage. Everything was connected to the central computer using twelve high speed links of 12 Mbits/s, multiplexed by the front end IBM 1802. Local equipment comprises of PDP, ARGUS, IBM, etc. computers plus multiple CAMAC units. This network was installed in 1968 and was designed to be easily extended as necessary.

This is still very much the philosophy for grid computing today.

In the 1970s all interfaces were built to the CAMAC, Computer Automated Measurement And Control standard. CAMAC was a modular system originally developed by ESONE: European Standards and Nuclear Electronics under a collaboration of several European nuclear laboratories. CAMAC modules enabled rapid design and build of digital electronic control systems. Up to 600 slim rack mounted modules were available built to internationally agreed standards and available from more than one manufacturer. It was found to be ideal for computer control in many industrial and scientific fields. It was demonstrated at the COMPEC’74 computer peripheral show in London. Daresbury staff provided a display including a PDP-11 for the CAMAC Association stand at the show.

Some more photos and information about CAMAC and associated systems on the SRS can be found here.

As noted above, in 1974 in addition to local terminals at Daresbury a PDP-11 supported remote job entry (RJE) stations at Daresbury, Lancaster, Liverpool, Manchester and Sheffield. There was also a connection over the SRCNet packet switching network to a GEC 4080 computer fulfilling a similar function at the Rutherford site.

These networks used packet switched protocols in which data are bundled up into “datagrams”, blocks of a few kBytes each having extra information about the sending and receiving addresses. Daresbury had one of the eight main packet switched exchanges and was a Network Operations Centre. Protocol for handling datagrams is very sophisticated; individual packets may not necessarily arrive in the order they were sent, perhaps taking different length paths through the network if traffic is high. Each datagram is sent separately, as in a postal system rather than establishing a fixed connection such as through a telephone exchange. Packets may be lost and need to be re-sent. However, error free communication is possible using sophisticated networking software developed over the last 15 years. Our use of computers, and much of our modern way of life, now depends on this capability (e.g. in banks, libraries, mail order, credit cards).
Figure 18: Daresbury Network in 1973.
Experiments with ethernet local area network started at Daresbury as early as Nov’1981. Dr. Brian Davies initiated an evaluation of ethernet in comparison to the currently used rival Cambridge Ring network with the purchase of £20k worth of equipment. This ethernet to connect four work stations was purchased from Thames Systems. At the time there were several different ethernet standards – the Ungermann-Bass system was chosen as it was the one supported by Xerox.

Access to the distributed UNIX programming environment is now available to the academic community through the Joint Academic Network (JANET). This was established in 1984 and preceded similar European initiatives. Almost all HEIs in the UK are now registered on this network, together with some commercial companies. JANET had a link to the main European and North American IBM network (EARN/ BITNET) in the late 1980s, which also provided a world wide electronic mail and data transfer facility, including Japan and Australia.

These services have now largely been replaced by or extended to include networks built upon the internationally standard TCP/IP (Internet Protocol) communications similar to the campus wide Ethernet.

Connection to JANET from major nodes in 1994 was “megastream” based at 2 Mbits/s. Connections to most academic establishments was via “kilostream” with an upper limit of 64 kbits/s. The SuperJANET project was expected to supply 34 Mbits/s connection to participating sites, this technology allowing the SuperJIPS project to proceed. The JANET Internet Protocol Service (JIPS) has successfully provided IP services to the academic community.

Since the protocols in use on local area networks are the same, access to a similar bandwidth, but nationally, offers many exciting possibilities. These include national workstation clustering and parallel execution, interactive visualisation of programs running at the Supercomputing centres using local graphics resources, and uses of remote systems particularly suited to certain types of computation by use of Remote Procedure Calls (RPC). This is now beginning to be realised on grids through the Grid and Web Services.

High level programming software, which encompasses resources on a network, has been developed at Daresbury since the mid 1980s to provide a platform for writing scientific programs and visualising results. We also need to make a single program run efficiently on those compute servers with multiple processors, such as the Intel hypercube and workstation clusters, by dividing it into small pieces which exchange data. Our strategy is therefore to treat the whole environment as one large parallel computer with data passing between the different machines. This is possible with special software to connect Ethernet addresses, with sockets at either end to send or receive data across a link. This is similar to a telephone system with one computer calling another and talking to it when the call is answered. Of course one of the difficulties of parallel computing is when a call isn’t answered, but let us assume that this does not happen very often.

One solution is a harness program which can be invoked from any workstation. It effectively logs into other machines and starts one, or several tasks running and passes data to them. Such harnesses can have an interface standard to all computers needed. At Daresbury a harness called FORTNET (Fortran for Networks) was developed [58]. The only change in moving from one machine to another, apart from speed, is currently the need to recompile the source program. It is even possible to use the same program on a single parallel computer such as the Intel, Meiko, HPCx or a NW-GRID cluster in this portable way after developing it on a workstation cluster. Worldwide activity in message passing software for scientific computing has led to the MPI: Message Passing Interface standard which is now
widely used.

4.2 The UNIX revolution, 1988-1994

The following text is based on that from the article in Physics World [5].

The change from a single AS/7000 mainframe to a Convex C220 running UNIX at Daresbury in 1988 not only meant a 16 fold increase in CPU performance and a 16 fold increase in main memory, but reflected the beginning of the mainstream trend towards distributed computing. Although computational scientists had access to the Cray X-MP at RAL, the differences between the Cray and AS/7000 made software development awkward. The arrival of UNIX was a watershed for Daresbury, enabling the subsequent rapid development of a powerful distributed computing network including super workstations and parallel clusters.

The benefits of this kind of distributed computing are considerable, and qualitatively change the way scientists approach computational problems. Firstly it is a flexible way of providing computer power responsive both to the rapidly changing hardware market place and crucially to the changing requirements of users' scientific applications. New components can be added simply, or upgraded as necessary. All the systems on our network were acquired according to particular needs of the scientific community which we support.

The use of specialised machines can be a spectacularly cost effective route to high performance. The current generation of distributed memory parallel machines can for instance equal and often outperform conventional vector supercomputers at approximately one tenth of the cost. Therefore we have a number of these on our network to provide high performance floating point arithmetic. However our experience has shown that any specialised system including a parallel one is most useful to the scientist when integrated with others of complementary functionality. These facilities must also be easily accessible.

The second feature of distributed computing is the integration of the different servers into one system via the network.

Thirdly users can exploit a kind of "network parallelism" – several machines in a cluster share parts of a single task and work concurrently. Integrating these aspects allows one to use the machines in a complementary was on a single application. For example computationally intensive parts of a problem can be "spawned" to a high performance parallel or vector machine from a graphical workstation environment.

The whole question of supercomputer provision is under continual review. A meeting held at the Royal Institution of Great Britain on 24th September 1992 examined the needs within the UK for further large scale computing resources to tackle both present and future research [15]. Some of Daresbury’s contributions to this provision were highlighted in the Daresbury Laboratory Annual Report.

A trend internationally in supercomputing activities is the rapid growth in “grand challenge” applications or projects. Such projects are characterised by: (i) a large number of collaborating groups, drawn from academia, research laboratories and industry; and (ii) an insatiable appetite for supercomputer cycles. The latter attribute is readily explained by the nature of the problem areas under investigation – examples include: the tracking of pollutants and pollution control in air, ground water
and the ocean; pharmaceutical drug design; chemical reaction and catalysis; structural engineering with applications in, for instance, the building and explosives industries; whole vehicle design and modelling in the aerospace and car industries; and many others. Supercomputers are vital in all these projects, not only to test new theories to high accuracy against the most exact experimental data (e.g. from the Synchrotron Radiation Source), but to enable integration in industrial design, modelling for verification to assess costs and side effects and for optimisation of process risks and costs.

The theme of international collaboration in high performance computing underpins recent programmes from the European Commission, who have announced funding for projects to port a number of major codes to parallel systems driven by consortia involving supercomputer centres, academic establishments and industry.

Many of the software techniques that are underpinning advances in the “grand challenge” arena are also being brought to bear on a new generation of powerful workstations, which may be used by individual researchers and also clustered to support parallel computation and sharing of data via a local area network. Through the availability of such resources, funded in many instances by the Science Board’s Computational Science Initiative (CSI), more researchers can benefit from the software developed and maintained at Daresbury under the auspices of the Collaborative Computational Projects (CCPs).

Computational science of world class status needs commensurate computing resources, and both the SRS and CCP communities make extensive use of the national supercomputing facilities at the Universities of London (Cray X-MP/28 and Convex C380) and Manchester (Amdahl VP1200) and the Joint Research Council’s Cray X-MP/416 at the Rutherford Appleton Laboratory The 64 node Intel iPSC/860 at Daresbury and the 256 processor Cray T3D at the Edinburgh Parallel Computing Centre. Similar facilities are provided at Orsay in France (by the CNRS) several sites in Germany (by the DFG and local Länder) and also in the Netherlands and Italy. These centres provide powerful processors and considerable memory, disk, mass storage and user support resources.

However high performance computing also requires visualisation facilities, a range of programming tools and ease of access and availability. It has long been realised that these supercomputing centres must be supplemented and complemented by quite powerful local facilities in university departments and research centres. This is also true in industry.

In 1986 therefore the SERC established the Computational Science Initiative (CSI) to provide distributed high performance computing for the biological science, chemistry, physics and mathematics communities. The initiative is now in its 7th year and was renamed the Distributed Computing Programme (DisCo) in 1993. It has provided hardware, software and maintenance support as well as staff and studentships to almost 100 research groups in these fields and also to materials researchers and users of the major synchrotron neutron beam and laser facilities at DL and RAL.

Although we only describe our own experiences many other laboratories in Europe and the USA have distributed systems similar to that at Daresbury which therefore should only be considered as an example. In Britain and the rest of Europe many higher education institutions (HEIs) now have their own campus wide networks with workstations providing access to shared resources. One long established example is NUNET linking Durham and Newcastle Universities. In the late 1970s this comprised two IBM mainframes and distributed terminals – it now has a wide assortment of machines including parallel computers and a dedicated cluster of fifteen IBM RS/6000 workstations. Examples in the USA include the national laboratories such as Argonne, NASA Ames, Oak Ridge and Lawrence Livermore. Worldwide collaboration between scientists is helped enormously by having a pool of
applications programs which can be used at key sites within a standard software environment.

The adoption of distributed computing by so many different groups aided by the UNIX operating system and software is indicative of its success. Indeed the widespread acceptance of UNIX and the international standards POSIX and OSF/1 related to UNIX are a crucial issue. Distributed computing and networking encourages interaction between scientists and their software and creates a driving force behind standardisation of hardware interconnection, operating systems, graphical interfaces and computer languages. Much free software is now available in this standard environment and the unrestricted access to information is underlined by the creation of the World Wide Web, by which any UNIX computer user can interrogate repositories of information through an interactive window interface at any registered site in the world. Daresbury is one of these registered sites.

We now consider in more detail the way services are provided for the use by the different components of our distributed computing system, outlining the roles of the network the parallel processing compute servers and graphical workstations.

The key elements in the Daresbury local area network (LAN) are shown in Figure 19. There are "compute servers", high-performance computers for numerical tasks, such as the Intel and Meiko parallel machines, "graphic servers", optimised computers for interactive visualisation and control, the Apollo and Silicon Graphics superworkstations, "file servers", multi-user computers with high storage capacity and good input/output bandwidth also including access to e-mail and network facilities, such as the Convex and Sun servers and netstor, and "clusters" such as the six IBM RS/6000 systems or the five HP series 700 systems with their own high speed interconnexion.

The complete LAN is made up of a number of interconnected Ethernet networks supporting a standard protocol which enables each computer to be identified with a unique name and Ethernet address. There were in 1994 about 100 Sun and Silicon Graphics office workstations as well as IBM and Apple/
Mackintosh PCs. The heavy traffic between all these and the file servers is divided onto a number of branches by bridges from the central “spine”. Software provided with the UNIX operating system can provide the user almost transparent access to both data storage and shared software. This avoids duplication and facilitates maintenance, upgrades and access to all shared resources. Remote log in to any machine on the network is possible with the correct password. Other software supports access to facilities on a remote machine through interactive windows on a workstation.

We note that by 2000 the use of vendor specific flavours of Unix, such as AIX (IBM), Irix (SGI), Solaris (Sun), Ultrix and True64 (DEC) had mostly been replaced by Linux, the widely used open source alternative.

4.3 The Future?

What are the problems associated with distributed computing? It is true that since users have more control over their computing, they also have more responsibility. They certainly need to know more about the hardware and the operating system, as well as their own code to get optimal performance. There is a danger of turning scientists into systems analysts or operators. It is true also that the day-to-day running of the machines, backing up, trouble shooting, etc. needs to be clearly organised – this side of operations is rather different for a distributed system than for one based on a centralised mainframe and more flexibility is needed generally requiring a larger number of dedicated staff.

The most critical element in a distributed system is the network itself. The network must perform well at all times, and a good deal of consideration must be given to its configuration and upgrading as the systems linked to it evolve. This activity also required a high complement of dedicated staff. Current networks become overloaded quickly. As with processors there will be great advances in network bandwidth over the next few years. Huge commercial as well as academic interest is already being shown in the potential offered by video conferencing and integrated media and retrieval facilities. This will give a vital boost for distributed computing. We note that since some of the above text was written in the 1990s, SuperJANET5 has been deployed and is connecting core sites with fibre network of up to 40 Gb/s.

If the potential difficulties are avoided, a distributed computing environment can be very productive for the computational scientist. It offers cost effective and high performance computing complementary to that provided at supercomputer centres, and offers it in a way which is responsive to the user’s research needs with large amounts of compute power under their control. Certainly at Daresbury distributed computing has allowed us to operate at a much higher level than could be envisaged in the mid 1980s. We believe research computing will used distributed computing systems even more in the future.

The above conclusions were written in 1994, but remain largely un-changed today. Our technology has changed, but our ambitions have not!
Figure 20: Computing Scales in North West England
5 2012 onwards. The Hartree Centre and Big Data.

The UK has aspirations to be at the forefront of the so-called “big data revolution” and the Hartree Centre established at Daresbury is a key STFC strength in this area. Hartree is the world’s largest centre dedicated to high performance computing software development and home to Blue Joule, the most powerful supercomputer in the UK. The UK Government has invested over £50 million in the Hartree Centre from 2011-2013 to support the progress of power efficient computing technologies designed for a range of industrial and scientific applications. This is because it is estimated that successful exploitation of high performance computing could potentially increase Europe’s GDP by 2%-3% by 2020.

The Hartree Centre focuses on partnerships with business and academia to unlock the commercial opportunities offered by high performance computing systems. See Web site http://www.stfc.ac.uk/hartree. This has the benefit of utilising contributing to the growth of UK skills in this area and has attracted IBM as the major partner to develop these opportunities. The Centre is working closely with a number of major companies such as Unilever, where we have established formal research partnerships.

5.1 The UK’s e-Infrastructure

For some time before 2012 there had been growing opinion in the research community that a more integrated approach to computing was required in the UK. Reasons stated were to remain on par with similar OECD countries and to aid economic recovery. Discussions came to a head in early 2011. Staff at Daresbury had already formulated a bid for funding known as the Hartree Centre ... What we said was: The Hartree Centre will be a new kind of computational sciences institute for the UK. It will seek to bring together academic, government and industry communities and focus on multi-disciplinary, multi-scale, efficient and effective simulation. The goal is to provide a step change in modelling capabilities for strategic themes including energy, life sciences, the environment and materials.

A 2010 YouTube video of then CSED Director Richard Blake presenting HPCx and the Hartree Centre vision can be found here.

This proved to be very timely and the vision became part of the UK e-Infrastructure [11] for which Government announced funding following the Conservative Party Conference in Manchester, summer 2011. See http://www.stfc.ac.uk/NewsandEvents/37248.aspx

From the DBIS press release of 1/12/2011, the proposed capital spend of £145M was to be as follows.

- £30 million for the Daresbury Science and Innovation Campus, supporting research into the latest product development software. This facility is now known as The Hartree Centre;
- £24 million for high capacity data storage across the Research Councils, ensuring researchers can easily access complex information from experiments;
- £31 million to improve high capacity networks, including JANET, the Higher Education Funding Council for England’s system that helps the higher education community share large amounts
of research data more easily;

- £19 million for specialist supercomputers in areas such as particle physics and astronomy, weather forecasting and climate change, and genome analysis;

- £4.75 million for the UK Space Agency to support the collection and storage of data from satellites; and

- £6.5 million to establish a research fund for collaborative university projects to improve access to e-infrastructure.

The latter became the £8M e-Infrastructure Connectivity Call from EPSRC and coined the term “Tier-2” for shared resources in regional centres of excellence.

Definition of Tiers

**Tier-0:** Europe wide with users from multiple countries, e.g. through PRACE, the European partnership for advanced computing. The UK currently does not have such an HPC machine. HECToR was suggested, but EPSRC were not able to commit the required resources. Tier-0 for the particle physics community is the HPC Data Centre at CERN.

**Tier-1:** typically a national facility. For HPC users this is HECToR, and may shortly also include the Hartree Centre facility at Daresbury. For particle physics users it is the LHC Tier-1 Centre at RAL.

**Tier-2:** with the funding of regional centres for the e-Infrastructure this tier can be defined to be a resource shared by a number of participating institutions. It may be interesting to see how this fits with future plans for the National Grid Service (NGS) and if there are any lessons to be learned.

**Tier-3:** is therefore considered to be the main institutional computing service, either for HPC users, data storage, particle physics or a combination of these. We could propose that it include all resources on a campus. It is the focus of the HPC-SIG and CG-SIG.

5.2 The Hartree Centre

It seems also a good time to recall some of the reasons Richard Blake chose the name Hartree Centre.


In the mid 1920s he derived the Hartree equations, and later V. Fock published the “equations with exchange” now known as Hartree-Fock equation which is the basis of computational chemistry.
1929 saw him become a Professor of Applied Mathematics at the University of Manchester and among other research he built his own differential analyser from Meccano. During the Second World War the subsequent differential analyser at the University of Manchester was the only full size 8 integrator unit in the country and was used to great effect to support the war effort.

In the 1930s he turned his research to radio wave propagation that led to the Appleton-Hartree equation. In 1946 he moved on to Cambridge where he was involved in the early application of digital computers to a number of areas.

From his books [25] you can see that he was instrumental in establishing computational science.

Following a formal tendering process, on 30/3/2012, STFC and IBM announced a major collaboration that will create one of the world’s foremost centres in software development, http://www.stfc.ac.uk/NewsandEvents/38813.aspx. This collaboration is a key component, and marks the launch, of the International Centre of Excellence for Computational Science and Engineering (ICE-CSE) [now known as The Hartree Centre]. Located at STFC’s Daresbury Laboratory in Cheshire, the Centre will establish high performance computing as a highly accessible and invaluable tool to UK industry, accelerating economic growth and helping to rebalance the UK economy.

High performance computing (HPC) has become essential in the modern world, aiding research and innovation, and enabling companies to compete effectively in a global market by providing solutions to extremely complex problems. Breakthroughs in HPC could result in finding cures for serious diseases or significantly improving the prediction of natural disasters such as earthquakes and floods. It will provide the ability to simulate very complex systems, such as mapping the human brain or modelling the Earth’s climate – the data from which would overwhelm even today’s most powerful supercomputer. By the year 2020, supercomputers are expected to be capable of a million trillion calculations per second and will be thousands of times faster than the most powerful systems in use today.

The IDC Corporation report to the European commission in Aug’2010 estimated that the successful exploitation of HPC could lead to an increase in the European GDP of 2-3% within 10 years. In today’s figures this translates into around £25 billion per year in additional revenue to UK Treasury and more than half a million UK based, high value jobs.

The Department of Business Innovation and Skills (DBIS) announced its e-infrastructure initiative in Oct’2011, with £145 million funding to create the necessary computer and network facilities for the UK to access this potential benefit. £30m of this was earmarked for HPC at Daresbury. This was in addition to an earlier government investment of £7.5m into HPC when the creation of an Enterprise Zone at the Daresbury Science and Innovation Campus was announced [now known as Sci-Tech Daresbury]. Following a rigorous tender process as a result of these investments, IBM was named as the successful bidder to form a unique research, development and business outreach collaboration with STFC.

Under the initial 3 year agreement, STFC will invest in IBM’s most advanced hardware systems, most notably the BlueGene/Q and iDataPlex. With a peak performance of 1.4 petaflop/s, which is roughly the equivalent of 1,000,000 iPads, the BlueGene/Q system at Daresbury will be the UK’s most powerful machine by a considerable margin. It is also the most energy efficient supercomputer in the UK, being 8 times more efficient than most other supercomputers.
These systems will help the Centre to develop the necessary software to run on the next generation of supercomputers, thus providing UK academic and industrial communities with the tools they will need to make full use of these systems both now and in the future.

The Centre will target both the increasingly important area of data driven science and continue to target software development for current and future computer systems, due within 5-10 years and could will require entirely new software design. STFC is already a world leading provider of the software engineering skills required to exploit the future growth in available computing power and this is a very exciting time to be collaborating in this way.

The investment into the Centre is being used to upgrade STFC’s existing computing infrastructure to provide the capability to host the next generation of HPC systems which have much higher power densities than existing systems. It will also install an impressive series of internationally competitive computer systems as a software development and demonstration facility, along with a range of advanced visualisation capabilities.

Procurement in early 2012 included refurbishment of existing HPCx computer room (now split into 2, half with water cooling and half with conventional air cooling), and purchase of equipment as follows.

IBM iDataPlex, known as Blue Wonder – 512x nodes with 2.6GHz Intel SandyBridge processors making 8,192 cores. Around half will be running ScaleMP software to allow testing and development of large shared memory applications. Blue Wonder started operation at Daresbury in 2012 as the 114th most powerful computer in the world (39th Top500 list, see http://www.top500.org).

IBM BlueGene/Q, known as Blue Joule – 6 racks with 98,304 1.6 GHz BGC cores and 96 TB RAM + Power7 management and login nodes. Plus BlueGene/Q – 1 rack as above but planned to make this into a prototype data intensive compute server. Blue Joule started operation at Daresbury in 2012 as the 13th most powerful computer in the world (39th Top500 list, see http://www.top500.org).

With data store, backup and visualisation facilities, the initial configuration of the centre is described as follows.

**Hartree Base**: conventional Intel x86 cluster technology (IBM iDataplex). Part of the Blue Wonder system.

**Hartree Data Intensive**: data intensive system. Conventional x86 cluster technology (IBM iDataplex). Part of the Blue Wonder system which will use advanced software from ScaleMP to aggregate nodes into large virtual SMP machines.

**Hartree Advanced**: IBM BlueGene/Q architecture known as Blue Joule.

**Hartree Data Store**: the data store. Uses 8x SFA10k disk arrays from DataDirect Networks, providing 5.7 PB usable disk space, with minimum 15 Gb/s throughput to any of the above compute systems.

**Hartree Archive**: an IBM TS3500 tape library with 12x TS1140 tape drives and 3760 tape slots. This provides around 15 PB tape storage.

**Hartree Viz and Training**: Refurbishment of other laboratory space to create a visualisation and training suite, plus project space also involved purchase of training PCs, top end PCs for development of visual applications and a large 3D immersive visualisation system from Virtalis.
to complement the existing suite used by the Virtual Engineering Centre and the Laboratory lecture theatre. Additional equipment was also installed at ISIC and Atlas on the Harwell site to facilitate joint projects.

**Blue Wonder**

Blue Wonder is a 512-node IBM xSeries iDataPlex. It comprises 228 nodes, each with 2x 8-core Intel Sandybridge processors and 32 GB RAM. Additional 24 nodes with the same spec plus 2x nVidia M2090 GPUs plus 4 high memory nodes with 256 GB RAM each. Infiniband high speed interconnect throughout. A further 256 nodes each have 2x 8-core Intel Sandybridge processors and 128 GB RAM.

[Some more information and photos of iDataPlex clusters can be found here.](#)

**Blue Joule**

When installed in 2012, Blue Joule was a 7 rack BlueGene/Q system. It was the 13th fastest computer in the world at that time. Each rack is around 200 Tflop/s peak performance, so 1.2 Pflop/s overall peak. Each rack has 1,024 16-core processors (16,384 cores). The seventh BlueGene/Q rack was used for more adventurous research projects and now forms part of the Blue Gene Active Storage system.

[Some more information and photos of BlueGene can be found here.](#)

**NextScale**

An IBM NextScale cluster was installed in April 2014. It has 360 nodes, each node has 2 x12 core Intel Xeon Sandybridge processors (E5-2697v2 2.7GHz) and 64GB RAM making a total of 8,640 cores. Interconnect is Infiniband high-speed network from Mellanox.

**Energy Efficient Computing**

[TBA]

### 5.3 50 Years of Big Data Impact

Every day [in 2014] the world creates and shares 2.5 quintillion bytes of data across an increasingly sophisticated global computer network. Fifty years ago it was hard to imagine how much computing would influence how we work, create and share information; and yet in general it is not public demand that drives computing advances but the requirements of researchers to collect, store and manipulate increasingly large and more complex data sets. The power of computing developed to analyse massive and mixed scientific data sets in turn transforms industry and everyday life. As we have seen, STFC and our predecessors have been at the forefront of computing knowhow for the past 50 years. During this time, we have led the way across the whole spectrum of computing capabilities, from high performance computing facilities and digital curation, to graphics and software, from networking to grid infrastructure and the World Wide Web.

In the early 1960s, we developed ground breaking computer graphics and animation technologies to help researchers visualise complex mathematical data sets. This innovative and pioneering approach using computer generated imagery (CGI) caused the Financial Times at the time to pronounce RAL
as the home of computer animation in Britain. STFC’s forebears continued to lead the UK’s CGI field through the next two decades, most notably creating the computer images for the movie Alien in 1979, the first significant film to use CGI and which won an Oscar for best special effects. The success of this film spawned a new sector, with many new companies commercialising the CGI concepts and code developed by STFC and introducing them to new markets. The UK computer animation industry is currently worth £20 billion including £2 billion generated by the video and computer games market.

Increasingly large data sets not only required new techniques but increasingly more powerful computers. In 1964 we were one of three research establishments which hosted an Atlas 1 computer at RAL, then the world’s most powerful computer. In the following years we continued to play a pivotal role in the development and support of the UK’s supercomputing hardware and software development capabilities. Today, STFC super computers such as Blue Joule and DiRAC are at the cutting edge of academic and industrial research, helping to model everything from cosmology to weather systems. Blue Joule, opened in 2013 and situated on the Sci-Tech Daresbury Campus, is the UK’s most powerful supercomputer. It is the foundation of STFC’s Hartree Centre, set up to harness the UK’s leading high performance computing capabilities in academia for the benefit of UK industry. It is estimated that successful exploitation of high performance computing could increase Europe’s GDP by 2-3% by 2020. These activities have re-affirmed the UK’s place as a world leader in high performance computing.

Another pillar in the world’s computing revolution has been connectivity. STFC’s predecessor organisations led the UK’s networking effort many years before the invention of the internet. Twenty-five years ago the Web was established at CERN and is now a fundamental part of our lives: 33 million adults accessed the internet every day in the UK last year and it is worth over £121 billion to the UK economy every year. STFC manages the UK participation in CERN and underpinned the internet’s development in the UK through its early computer networking deployments, hosting the first UK Web site, developing Web standards and protocols, supporting the evolution of the Grid, and spinning out some notable organisations. These include: Nominet, the .co.uk internet registry which manages 10 million UK business domain names; JANET, which manages the computer network for all UK education, the .ac.uk domain and the JISCMail service used by 80% of UK academics (1.2 million users).

Big science projects such as those supported by STFC have consistently pushed the boundaries of data volumes and complexity, serving as “stretch goals” that drive technological innovation in computing capability. In the 1990s, the Large Hadron Collider (LHC) at CERN was the first project to require processing of petabyte scale datasets (a million gigabytes) on an international scale and this led to the development of grid computing. The LHC Grid makes use of computer resources distributed across the UK, Europe and worldwide to process the huge volumes of data produced by the LHC and to identify the tiny fraction of collisions in which a Higgs boson is produced. This technology development was supported by the RCUK e-Science programme and STFC’s GridPP project and the expertise developed is now supporting the UK and European climate and earth system modelling community through the JASMIN facility and the Satellite Applications Catapult through the Climate and Environmental Monitoring from Space (CEMS) facility. This same approach is now widely used by business and academia as part of the Cloud Computing revolution.

As the world becomes increasingly digital, preservation of digital records becomes more and more important across all aspects of daily life; a major task given how quickly innovations in digital media occur. Maintaining access to digital data has been at the heart of STFC science for over 30 years. It is still possible to access the raw data recorded on the ISIS neutron source since its first experiments over
25 years ago. Working through the Consultative Committee on Space Data Systems, STFC helped to derive the standards which have been adopted as the de facto standard for building digital archives and the ISO standard for audit and certification of digital repositories. Working with partners such as the British Library and CERN, STFC has formed the Alliance for Permanent Access to the Record of Science to address issues with long term preservation of digital data.

Looking to the future, the exploding volume of scientific data sets needed for fundamental science will continue to drive innovation. By 2023, the Square Kilometre Array project will generate 1.3 zettabytes of data each month - that’s 1300 billion gigabytes, over 10 times the entire global internet traffic today. Processing such a flood of data will require computers over a thousand times faster than today’s. This is a true stretch goal for the computing industry that may well require a transformative rethink of computer architectures. For this reason industry partners such as IBM and nVidia are closely involved in the current SKA project engineering phase. A related challenge is reducing the energy consumption of computers to well below current levels. Already, a University of Cambridge computer cluster, built to the SKA system design and supported by STFC, is one of the top two most energy efficient supercomputers in the world as ranked by the “Green 500” list. The close connection between SKA and the impact on electronic signal processing, computing and big data is one of the reasons why it is a high priority for STFC and why we are taking a strategic lead in the project.

Whilst we don’t know exactly how these innovations will affect our daily lives, we can be confident of two things: the discovery science projects that STFC supports will continue to drive innovation in information technology; and the sheer pace of change in that industry means that these innovations will very quickly benefit the daily lives of everyone in the country.

In the UK Government’s Autumn Statement of 3/12/2014 it was announced that there will be a further input of £113M to create a Cognitive Computing Research Centre at Daresbury.

6 Acknowledgements

Thanks to the Scientific Computing Department at STFC for allowing us time to do this work and providing funding to house and maintain our collection and make it available for visitors to see. Our work is now part of the Public Engagement work of SCD.

We acknowledge input from various sources, most of which are cited in the text. The various reports from the Atlas Computing Laboratory and Daresbury and Rutherford Appleton Laboratories have been particularly useful, some of which are available from the Chilton Computing Web site. Daresbury Library staff Debbie Franks and Mark Swaisland have sought out a lot of material from their archives for us to complete the picture.

We thank Mike Wilson, Linda Gilbert, Sue Mitchell and Stephen Kill of RAL, Stuart Eyres and Laura Roe of Daresbury for access to additional historical material and Pat Ridley of Daresbury for proof reading and advice. We particularly thank Vic Pucknell, Mark Hancock, Pete James, Bill Smith, Linda Enderby, Mike Miller and Chris Dean and others of DL for loan of exhibits or recounting their experiences.

We thank Bill Purvis for his friendship and collaboration in the past allowing us access to the Manchester Mark I which is now housed at MOSI, the Manchester Museum of Science and Industry.
We thank Prof. Jim Austin (York) for his interest and for sharing information about his Computer Museum.

We thank Mike Bennett, Phil Aikman and others who have shown an interest in this work and donated exhibits.

Thanks to Stephen Hill for information about CAMAC, data collection and control systems.

Thanks to Karl Richardson who worked with us in 2013-14 to help organise and catalogue our collection.

Thanks for Jean de la Taille for contributions on the early Cray computer series.

Thanks to Anna Jones and her students at BookSlide.org for contributions and links.

References


REFERENCES


[19]


D.R. Hartree Calculating Machines: Recent and Prospective Developments and Their Impact on Mathematical Physics (University Press, 1947)


[28] K. Kleese and R.J. Allan Giant machines that make data management a very tall order Scientific Computing World 51 (Feb-Mar’2000) 34-7

REFERENCES


[34] Victoria Marshall *Re-discovering the Chilton Atlas Console* BCS Resurrection 69 (Spring 2015) 23-25


[38] C. Peyton *Radiation Beam Machine will be breaking New Ground* Electronics Times (27/11/1980) 24


[40] C.L. Roberts, MBE *Atlas* (private publication, Apr’1996) copy in RAL library


REFERENCES


[51] *Atlas Computer Laboratory* (Science Research Council, 1967)

[52] *DL Annual Report* series and TCSC Appendix

[53] Series of articles in Parallelogram publication, early 1990s.

[54] *DL Computer Bulletin* series


[57] *CCP Newsletters* regularly published in association with DL

[58] FORTNET publications


[64] *PDP-11.co.uk* [http://www.pdp11.co.uk/tag/pdp-1105/](http://www.pdp11.co.uk/tag/pdp-1105/)

[65] *Computer History Museum* [http://www.computerhistory.org](http://www.computerhistory.org)

[66] *Cray-1 Supercomputer 30th Anniversary* [http://www.youtube.com/watch?v=J9kobkqAicU](http://www.youtube.com/watch?v=J9kobkqAicU)


[68] *Computer 50* [http://www.computer50.org](http://www.computer50.org)

[69] *Centre for Computing History* [http://www.computinghistory.org.uk](http://www.computinghistory.org.uk)

[70] *IBM Archives* [http://www-03.ibm.com/ibm/history](http://www-03.ibm.com/ibm/history)

[71] *Home Computer Museum* [http://www.homecomputer.de](http://www.homecomputer.de)


[73] *The Obsolete Computer Museum* [http://www.obsoletecomputermuseum.org](http://www.obsoletecomputermuseum.org)
Notes on Visit made to Daresbury, 12 February 1974

F.R.A. Hopgood and D.G. House

Introduction

The main purpose of this trip (made by FRAH and DGH) was to obtain some answers to our outstanding questions regarding the Daresbury site. Most of the time was spent with Brian Davies and Trevor Daniels of the Computer Centre.

Location

The countryside around Daresbury seems similar to Berkshire except that most of the towns have significantly more industrial areas associated with them. Outside the towns, there are reasonably unspoilt areas.

Public transport between neighbouring towns and Daresbury is not very good. For example, bus from Warrington runs once every hour.

Schools in the area are comprehensive. There are one or two purpose-built schools. However, most of them have been produced as a result of the amalgamation of existing schools. The change-over took place about 4 years ago.

Most people seem to live south of the site. Northwich, Middlewich and Winsford are quite close (under 10 miles). These towns are similar in size to North Berkshire towns. Industries include an ICI chemical plant, Guinness factory, Power Stations, etc. More select areas are Chester and Lymm. On the other hand, Warrington to the north is rather run down.

House prices range more widely than in North Berkshire. It is possible to buy a terrace house in Warrington for £600. New 3-bedroom houses in some of the towns can be purchased for as little as £7,000. At the other extreme, there are plenty of high class modern and old residences for between £20,000 and £40,000.

Council houses are available in several towns. Runcorn New Town have some with a three week waiting list!

Computing Hardware
The main equipment is:

- 370/165 with 2 Mbytes store
- 2305-2 drum
- 6 3330 exchangeable drives
- 8 2314 exchangeable drives
- 2 1403
- 1 2540
- 1 2501
- 1 Calcomp drum plotter
- 2 7-track tape decks
- 6 9-track tape decks

The eight 2314 drives are to be replaced by 7330s (ITEL) which are 3330 compatible drives.

The 2305, 2314 and 3330 are all attached to the same two 2880 channels.

Three 2860 channels have a 2319, 1800 and 2x 2250s attached to them. All remote teletypes are attached to the 1800.

Most local users seem to use either IBM 2741 typewriters (11), or VISTA alphameric displays (50 of these!). The latter are attached over fast lines. Output rolls up almost instantaneously. These devices cost about £1,000 each.

The Laboratory’s staff are building their own workstations consisting of a PDP-11 with a variety of peripherals attached using CAMAC interfaces. The cost, including a display console, is about £7,000 compared with the GEC 2050’s cost of £15,000.

Computing Workload

The 370/165 is being used 24 hours a day, 7 days a week, contrary to rumours. Most of the workload is generated on site. The major external user is CERN (8%). The NERC workstation attached to the 370/165 uses about 1/2-hour per week [check]. Most of the workload comes from the Nuclear Physics groups at Manchester, Liverpool, Glasgow, Sheffield and Lancaster (80%). Professor Hani (Manchester) provides low priority compute-bound jobs.

The CPU efficiency is about 70% (30% idling). This is similar to the Rutherford 360/195. While we were in the machine room, efficiency rarely dropped below 100%. This was with a number of on-line programs.

On-line programs are each restricted to 70 kbytes. Not more than 5 are allowed at anyone time and averages out at 2 or 3 (140 to 210 Kbytes).

Computing Software

Primary language used is FORTRAN (Level G and Level H). There is a 15% use of PL/I. ALGOL is available but is not used.

Brian Davies appears to do most of the day-to-day running of the Computer side. He seems to split his time between management and applications programming.
The number in the Applications software area is rather vague as some are resident in CERN.

The Systems programmers have been together for about 2 years. They do not seem to have any difficulty in keeping staff.

Operators are recruited locally. Again, there appears to be no trouble in getting or keeping staff.

The Electronics Development section works quite closely with the Systems Group. Most interfacing is done locally.

**Computer Buildings**

The Computer Block and associated offices are housed in a single building. There is both a Think Room with a number of displays and a card punch room. There are about 22 offices in the Computing Block.

To accommodate ACL [the Atlas Computing Laboratory] within the Computer Block is impossible. We probably need 60 to 80 offices for ourselves and 10 to 15 for users.

There are two other office blocks available at Daresbury. The A block is primarily for Administration. The B block currently holds NP staff. This is likely to be used partly for NSF and some space may be available for ACL. For example, there is space that can be used for equipment like the 2050 and microdensitometer. Total number of offices in B block is about 60. It is therefore unlikely that office space will be available for ACL even if a significant part of B block is used.

The position of the Computer Block in the centre of the site makes it difficult (but not impossible) to put up additional accommodation for offices or computer equipment nearby.

The Computing Room was extended to accommodate the 370/165. They do have sufficient space to allow a dual processor 370/168 to replace this. would be some problems. The current raised floor is 18in high in the new part but only 6in in the old. The 370/168 would require the 18in false floor. The old floor would therefore have to be taken out and dug up.

There is no possibility of getting the 370/165 and another comparable machine in the current computer block. Installation of a P4 or STAR-100 would require another computer block. The simplest way of doing this might be to put it on top of the existing one. The structure should stand it if the foundations will.

Note that it is not possible to upgrade a 370/165 to a 370/168. It is possible to upgrade the 370/165 to the Mark 2 version allowing virtual memory. It would not be as fast as a 370/168.

The Operating System is OS/TSO. HASP is not available. However, workstations are run using HASP locally) attached to interleaving with a package (written locally) attached to OS.

Currently, jobs larger than 540 kbytes will not be run in day shift.

**Miscellaneous**

There are no adequate dark rooms on the site. Photographic section only has single frame developing equipment.
The PDP-15 would be connected to the 370/165 via 1800. This could be quite a fast link.

Trunking for teletypes is available in all on-site offices.

Office accommodation is better than at Rutherford, but possibly not as good as at ACL.

**B  The Early Days of CCP4, c.1977**

Dr. Talapady N. Bhat

In 1977 I joined David Blow at Imperial college, London. My job there was to work on the structure determination of Tyrosyl t-RNA synthetase. This was the time when David and Alan Wonacott moved from Cambridge, MRC to London and they were yet to establish a computing facility at Imperial College for protein crystallography. Their initial proposal was to explore the possibility of my using the computers located at Cambridge for my work. A few days after I arrived at Imperial College, David introduced me at MRC to Max Perutz, Bob Diamond, Gerald Bricogne who promised to provide me their support and help to get me started. Max offered me a shared table space in a room full of 3-D ball and stick models of proteins and a computer graphics screen that could be used for real space modeling of proteins using Bob's graphics program. I was very excited to be working in the world's most famous protein crystallography laboratory with an opportunity to learn from world famous scientists. There after for a few weeks, every day I traveled from London to Cambridge for my computing needs. Some weeks later David realized that, going to Cambridge from London every day was not practical to get work done and also that it was expensive. Therefore, he suggested that I needed to look into alternative ways to get computing work done. This was the time when remote dial up technology was at its infancy in the UK, and he suggested that I may use a dial up facility from London to reach the Cambridge computer. A week later, the Imperial college telephone authorities realized that we were using the telephone lines for data transfer and they requested us not to use it any more. Following that David suggested that I may use the dial-up facility from University College London to connect to Cambridge. This mode of using the computer at Cambridge also turned out to be not practical since the dial up facility was very un-reliable and the dial up card reader frequently dropped the line during reading of cards. Furthermore, the very limited disk space (50 to 100 block of 512 bytes) allocated for my use at Cambridge did not make my job any easier.

Around the same time it so happened that, the Daresbury Laboratory was looking for opportunities to support Biological research. With this goal in mind, Dr. Sherman from DL one day visited our group at Imperial to discuss our computing needs and to explore the possibility of providing computing support for us. During his visit I explained to him our frustration in using the Cambridge computer for my computing. He described the excellent computing resource available at Daresbury at that time and then proposed to provide practically unlimited resources (both disc and computing time). He also suggested that we could use the remote terminals located at the high energy physics lab in Imperial college to log in to Daresbury. All these proposals looked too good for us to ignore particularly considering the difficulties we had in using the computers at Cambridge. However, we realized that until that time Daresbury lab did not have any of the protein crystallographic computer programs installed for our use. Though, this is a major issue, we accepted the proposal from DL to support our computing needs. David, thought that this would be a good opportunity to access SERC funds as well for our work and suggested that SERC-DL provide funds for my salary and related expenses.
such as travel to DL whenever needed. Dr. Sherman accepted our request with the condition that he would pay my salary only if I establish a state-of-the-art macro-molecular software suit at DL for general use and also help DL to attract users from other major laboratories such as Birkbeck College, Oxford, MRC, Sheffield and York. We agreed to work towards this goal and talked to Tom Blundel at Birkbeck College about the above proposal from DL. Tom showed interest and support for the idea, though he acknowledged that his computing needs are far less urgent than ours. By that time Tom had a well established computing facility at Birkbeck College. Daresbury identified this support for us by establishing a new funded project called – CCP4.

Following this project award to us by DL-SERC, I started installing protein crystallographic programs at DL. Some of the initial programs installed by me at DL were: a) FFT developed by Lynn Ten Eyck; b) Phase combination program with Gerald Bricogne’s modifications; c) the density modification program and program to refine partially fitted structures written by me; and d) the refinement program, PROLS, by Wayne Hendrickson.

Subsequently, we organized a meeting at Birkbeck College to discuss the possibility of fostering greater participation by other protein crystallographic laboratories in the use of the DL computing facility. Leading computer program developers from several labs came to this meeting. Ian Tickle from Birkbeck College, Phil Evans from Oxford, Eleanor Dodson from York, Phil Bourne from Sheffield, Karle Branden from Sweden, Johann Deisenhofer and W. Steigman from Munich, Alan Wonacott and myself from Imperial College were some of the people who attended this meeting. The Munich group suggested that the best way to establish a complete protein crystallographic program package at DL would be to adopt their program package called “Protein”. They said that they have a 1600 BPI tape ready with them with all the programs and they could give it to us right there. In response to that suggestion, Phil Evans replied that the scaling and phasing program from Oxford is superb and it also got to be part of the package at DL. Eleanor said that the FFT based refinement by A.C. Agarwal has to be a part of the program package at DL. Alan Wonacott replied that Gerald Brocogne’s electron density map skewing and phase combination program is a must for the DL program package and thus in few minutes the list of “the must have programs at DL” started piling up. However, these only operated with their individual file formats and no one knew of a method that would allow to exchange data among the programs. These discussions lead everyone to realize that they had a real serious problem in integrating these must have programs. Then several people started telling horror stories how they mistakenly computed electron density maps with diffraction intensities and refined heavy atoms with wrong data and so on. To solve these problems of data exchange between programs, we considered several models. For instance, the 9A2 by Cambridge and the pre-defined file formats with reserved data columns for each type of values used by Protein package are some of the possibilities that we considered. However, everyone agreed that none of the formats such as 9A2 or the Munich file format system had the features that we were looking for. Subsequently, meeting adjourned with plans to meet again in a month’s time.

Following that meeting, Alan Wonacott and I started to work on the design of a [new] file format that would meet our needs. Alan felt that the rapid sortability of 9A2 format is an essential feature for the new format. I felt that the names of columns of the files need be amenable to machine reasoning such that a user should never have to worry about calculating electron density maps using intensity values or mistaking phases with amplitudes and so on. I also felt that it is also important that a general application program developer need not have to worry about doing input or output to the data files. These requirements led us to develop the Labeled Column Format (LCF). I would consider some of the concepts used by LCF are the early attempts to provide transparent data management support
for program developers [similar to] some of the features currently available in SQL based modern databases. LCF routines practically mask out data from unwanted columns from a user program, and they also make the order in which the data columns are stored in the file irrelevant to their user. The LCF data is also sortable and editable upon request. Columns can be added or deleted as needed in a LCF file.

In a subsequent meeting at Birkbeck College, Alan and I presented these LCF concepts to all members of the team and discussed its features such as what are the minimal columns, what should be their internal storage method, size (bytes) per column, direct access file or sequential access file, what should be the basis on HKL (256 or not) and so on. The idea was quickly picked up by others who were at the meeting. I was a strong proponent of providing common blocks as the means of sharing data between a user program and LCF APIs, but Phil Evans was a strong proponent of using passed parameters while calling the API to share data. We argued over these differences almost for a full morning and then someone, probably Eleanor suggested: since I am going to be developing the LCF routines I may use common blocks to share data among the LCF routines, however, since Phil and other will be developing the macro-molecular crystallographic programs to use the LCF routines that I plan to develop, I need to make provisions in LCF routines to exchange data with external programs by passing data through parameters. That is why the initial documents on LCF routines explain sharing of data both through common blocks and through passed parameters. Since all the programs are expected to be in FORTRAN, it was suggested that all the LCF routines be written in FORTRAN. However, FORTRAN-4 did not allow the use of two character variables for storing data, and therefore an exception was granted to me to use a PLI routine for I/O to the LCF files. The use of direct access for storing data was ruled out since we had only limited disk space available at that time and thus the LCF routines were expected to work from magnetic tape as well.

Soon after that, Daresbury expanded their support for protein crystallography and provided funds for an additional staff member, John Campbell, and a co-ordinator, Pella Machin, for CCP4 both to be stationed at DL. With this additional support, the CCP4 project became fully mature by taking up additional roles such as providing user support, arranging workshops and so on.

C Archive Images

The Daresbury Laboratory reprographics group and library have extensive archives of press cutting and photographs. We are very grateful to them for making this material available for our research. Some of the images are re-produced here.

Follow this link for archive images.

D Computer systems at Daresbury over the Years

Table D lists some of the computer systems which have been used for research work at Daresbury Laboratory.
Table 2: Significant Computer Systems at Daresbury Laboratory

<table>
<thead>
<tr>
<th>Item</th>
<th>Make</th>
<th>Dates installed at DL</th>
<th>procs and speed</th>
<th>memory</th>
<th>total MFlop/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Ferranti Atlas (RAL)</td>
<td>1964-</td>
<td></td>
<td>256 kB</td>
<td>1 MIP/s</td>
</tr>
<tr>
<td>1</td>
<td>IBM 1800</td>
<td>1966-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>IBM 360/50</td>
<td>1966-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>IBM 360/65</td>
<td>1968-</td>
<td>c.1 MHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>IBM 370/165</td>
<td>1973-</td>
<td>12.5 MHz</td>
<td>3 MB</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>PDP-11/05 Perkin-Elmer 7/32</td>
<td>1974-81</td>
<td>100kHz</td>
<td>16kB</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Cray-1S</td>
<td>1978-83</td>
<td>2x 120 MHz</td>
<td>4 MB</td>
<td>480 (115kW per unit)</td>
</tr>
<tr>
<td>7</td>
<td>GEC 4000 cluster NAS 7000</td>
<td>c.1975-1981</td>
<td>100 MHz</td>
<td>8 MB, later 16 MB</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Concurrent 3230 (formerly Perkin-Elmer)</td>
<td>1985-</td>
<td>2x 4 MHz LSI CPU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9a</td>
<td>FPS 164</td>
<td>1984-89</td>
<td>11 MHz</td>
<td>4 MB</td>
<td>11</td>
</tr>
<tr>
<td>9b</td>
<td>FPS 264</td>
<td>1984-89</td>
<td>11 MHz</td>
<td>4 MB</td>
<td>22</td>
</tr>
<tr>
<td>10a</td>
<td>Meiko M10</td>
<td>1986-89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10b</td>
<td>Meiko M60</td>
<td>1990-93</td>
<td>14x T800 transputers and 10x i860 co-processors</td>
<td></td>
<td>560</td>
</tr>
<tr>
<td>11</td>
<td>Convex C220</td>
<td>1988-94</td>
<td>2x 25 MHz custom CMOS CPU</td>
<td>256 MB</td>
<td>72</td>
</tr>
<tr>
<td>12</td>
<td>Intel iPSC/2</td>
<td>1988-94</td>
<td>32x 4 MHz 80386/7, Weitek 1167 and AMD VX co-processors</td>
<td>160 MB</td>
<td>212</td>
</tr>
<tr>
<td>13</td>
<td>Stardent 1520 SGI Power 4D/420</td>
<td>1989-94</td>
<td>2x 32 MHz MIPS R3000</td>
<td>32 MB</td>
<td>16</td>
</tr>
<tr>
<td>14</td>
<td>Intel i860 Alliant FX/2808</td>
<td>c.1990</td>
<td>4x 32 MHz MIPS R3000</td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>15</td>
<td>IBM SP-2/3 Beowulf cluster</td>
<td>1994-98</td>
<td>64x 40 MHz 80860. Top500 no. 210 in June 1993</td>
<td>512 MB</td>
<td>2.56 Gflop/s</td>
</tr>
<tr>
<td>16</td>
<td>Loki cluster</td>
<td>1999-2003</td>
<td>8x 80860</td>
<td></td>
<td>320</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>32x 450 MHz Pentium III</td>
<td>8 GB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Scali cluster</td>
<td>2003-2007</td>
<td>64x 1 GHz AMD K7</td>
<td>64 GB</td>
<td></td>
</tr>
<tr>
<td>19a</td>
<td>IBM Regatta HPCx phase I</td>
<td>2002-2010</td>
<td>1280x 1.3 GHz Power4. Top500 no. 9 in Nov'2002</td>
<td>1.6 TB</td>
<td>6.66 Tflop/s.</td>
</tr>
<tr>
<td>19b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.015 Gflop/s/W</td>
</tr>
</tbody>
</table>
Computer Systems at Daresbury Over the Years

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>19b</td>
<td>IBM Regatta HPCx phase II</td>
<td>1600x 1.7 GHz Power4+</td>
</tr>
<tr>
<td>19c</td>
<td>IBM Regatta HPCx phase III</td>
<td></td>
</tr>
<tr>
<td>19d</td>
<td>IBM Regatta HPCx phase IV</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>NW-GRID</td>
<td>2005-12</td>
</tr>
<tr>
<td>21</td>
<td>BlueGene/L</td>
<td>-2011</td>
</tr>
<tr>
<td>22</td>
<td>Hapu</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Woodcrest</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>BlueGene/P</td>
<td>-2012</td>
</tr>
<tr>
<td>25</td>
<td>CSEHT</td>
<td></td>
</tr>
<tr>
<td>25a</td>
<td>nVidia</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Fujitsu</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>IBM iDataPlex (SID)</td>
<td>2011-present</td>
</tr>
<tr>
<td>28</td>
<td>IBM iDataPlex (Blue Wonder)</td>
<td>2012-present</td>
</tr>
<tr>
<td>29</td>
<td>BlueGene/Q (Blue Joule)</td>
<td>2012-present</td>
</tr>
<tr>
<td>20</td>
<td>BlueGene/L</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Hapu</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Woodcrest</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>BlueGene/P</td>
<td>-2012</td>
</tr>
<tr>
<td>25</td>
<td>CSEHT</td>
<td></td>
</tr>
<tr>
<td>25a</td>
<td>nVidia</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Fujitsu</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>IBM iDataPlex (SID)</td>
<td>2011-present</td>
</tr>
<tr>
<td>28</td>
<td>IBM iDataPlex (Blue Wonder)</td>
<td>2012-present</td>
</tr>
<tr>
<td>29</td>
<td>BlueGene/Q (Blue Joule)</td>
<td>2012-present</td>
</tr>
<tr>
<td>20</td>
<td>BlueGene/L</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Hapu</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Woodcrest</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>BlueGene/P</td>
<td>-2012</td>
</tr>
<tr>
<td>25</td>
<td>CSEHT</td>
<td></td>
</tr>
<tr>
<td>25a</td>
<td>nVidia</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Fujitsu</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>IBM iDataPlex (SID)</td>
<td>2011-present</td>
</tr>
<tr>
<td>28</td>
<td>IBM iDataPlex (Blue Wonder)</td>
<td>2012-present</td>
</tr>
<tr>
<td>29</td>
<td>BlueGene/Q (Blue Joule)</td>
<td>2012-present</td>
</tr>
</tbody>
</table>

* Rpeak MFlop/s quoted for 64-bit arithmetic where possible.

For a history of parallel computing with a general timeline see [47].

Some other events since 1980. (source Google Groups Usenet Timeline [http://www.good-stuff.co.uk/useful/google_usenet_timeline.php](http://www.good-stuff.co.uk/useful/google_usenet_timeline.php)).

May 1981 – first mention of Microsoft
May 1981 – first mention of MS-DOS
Aug’1981 – first review of an IBM PC
Apr’1982 – first mention of Sun Microsystems
Jun’1982 – first mention of a compact disc
Aug’1982 – first mention of the Commodore 64
Aug’1982 – first mention of Apple’s Lisa and Macintosh products
Dec’1982 – announcement of first cell phone deployment in Chicago
Feb’1983 – first mention of a Fax machine
Sep’1983 – Richard Stallman’s announcement of GNU
Nov’1983 – first mention of Microsoft Windows
Aug’1984 – first mention of the Commodore Amiga
Jul’1986 – first mention of Cisco
Mar’1988 – first mention of the term “search engine”
Nov’1988 – first warning about the Morris Internet Worm
Feb’1989 – first mention of Internet Relay Chat (IRC)
Sep’1991 – announcement of Internet Gopher
Oct’1991 – Linus Torvalds’ Linux announcement
Mar’1993 – Marc Andreessen’s Mosaic announcement
Jun’1994 – Announcement of WebCrawler launch
Oct’1994 – Marc Andreessen’s Netscape announcement
Dec’1994 – early mentions of Yahoo! and Lycos
Sep’1995 – eBay founder Pierre Omidyar advertises new auctioning service
Dec’1995 – announcement of AltaVista launch
Mar’1998 – first mention of Google
May’1998 – first mention of Mac OSX

E  Exhibitions and Artefacts

E.1  Collection Catalogue

We list below some of the computer artefacts we have in the collection which are available for open days and other exhibitions. This provides quick links to descriptions in the rest of the Web site.

This catalogue is now on [a separate Web page here](#).

There is a related page for [the Atlas Archive and Collection](#).

E.2  Exhibitions

Exhibitions we have presented so far include the following.

1. Institute of Physics meeting, 14/11/2007
2. BA Science mini-festival, 8/10/2006
3. Daresbury Open Day, 14/9/2008
5. BA Science mini-festival 4/10/2009
6. BA Science mini-festival 30/9/2012
7. Atlas 50 Anniversary, Manchester, 3-6/20/2012. [Slideshow](#) [PDF]
8. Internal Meeting, 7/1/2014. [Presentation slides](#) [PDF]
9. Sci-Tech campus access day 30/10/2014.
F Interestong Facts

Thanks to Andrew Loewe for collecting some of this information.

F.1 Why does a Computer need to be so big?

Computers for Research

Powerful computers have over the last few years given researchers a new way to do science. You can now calculate things rather than doing the equivalent experiment (although scientific theories still have to be validated to ensure the calculations will be realistic). Such calculations can be quicker than an actual experiment if the computer if powerful enough. They can also be safer – some experiments are hazardous. For instance testing of weapons (especially nuclear ones) is prohibited, so some large government establishments particularly in the USA, use comuters instead. The same people might also be looking into nuclear power generation.

We don’t do arms testing, but we do use big computers for physics, chemistry, engineering and environmental studies.

The Knowledge Centre for Materials Chemistry is doing research to develop novel materials for uses in many advanced products. Ask Rick more about what they do and take a look at their Web site: http://www.materialschemistry.org/kcmc/index.html.

If you can do this kind of research faster, by simulating more materials, you can make a big scientific discovery or get your product to market more quickly. The use of computers can thus give a competitive edge. Big computera aren’t just used for scientific research, they are also used in banks and businesses where they can simulate business processes, allow us to test strategies and help with making decisions.

More Processors makes a faster Computer

A modern PC has one, two or four processors (cores) of around 3GHz clock speed (three thousand million ticks per second). Roughly one floating point or integer arithmetic operation can be done per tick, so around 3Gflop/s per core (some special processors do more).

If you add in more processors you get more operations per second. So the fastest computers are equivalent to many PCs connected together. Look at the TOP500 Web site to see what are the fastest ones in the world today: http://www.top500.org.

The biggest computer we had at Daresbury was called HPCx. The biggest one used for academic research in the UK is called HECToR. Can you find out about these?

The world’s largest computers in May 2010.

Information from the TOP500 list, Dec’2009, see http://www.top500.org.
<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Compute Elements</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road Runner</td>
<td>Los Alamos</td>
<td>13,000 cells</td>
<td>[<a href="http://www.lanl.gov/discover/roadrunner_fastest_computer">http://www.lanl.gov/discover/roadrunner_fastest_computer</a>]</td>
</tr>
<tr>
<td>Kraken XT5</td>
<td>National Institute for Computational Sciences/University of Tennessee</td>
<td>99,072 cores</td>
<td>[<a href="http://www.nics.tennessee.edu/computing-resources/kraken">http://www.nics.tennessee.edu/computing-resources/kraken</a>]</td>
</tr>
<tr>
<td>JUGENE</td>
<td>Forschungszentrum Juelich (FZJ)</td>
<td>294,912 cores</td>
<td>[<a href="http://www.fz-juelich.de/jsc/bg-ws10/">http://www.fz-juelich.de/jsc/bg-ws10/</a>]</td>
</tr>
<tr>
<td>BlueGene/L</td>
<td>Livermore</td>
<td>212,992 cores</td>
<td>[<a href="http://www.top500.org/system/7747">http://www.top500.org/system/7747</a>]</td>
</tr>
<tr>
<td>BlueGene/P</td>
<td>Argonne National Laboratory</td>
<td>163,840 cores</td>
<td>[<a href="http://www.top500.org/system/performance/9158">http://www.top500.org/system/performance/9158</a>]</td>
</tr>
<tr>
<td>HECToR</td>
<td>University of Edinburgh</td>
<td>44,544 +12,288 +112</td>
<td>[<a href="http://www.hector.ac.uk">http://www.hector.ac.uk</a>]</td>
</tr>
<tr>
<td>HPCx</td>
<td>Daresbury Lab</td>
<td>2,560 cores</td>
<td>[<a href="http://hpcx.ac.uk">http://hpcx.ac.uk</a>]</td>
</tr>
</tbody>
</table>

**Some supercomputers in the UK in 1997**

This note appeared in HPCProfile Jan’1997 [2].

Browsing the TOP500 list at the University of Mannheim gives useful information about supercomputers installed all over the World. We extracted the current UK situation below. You can compare with other countries by browsing [http://parallel.rz.uni-mannheim.de/top500.html](http://parallel.rz.uni-mannheim.de/top500.html). In releasing the 8th edition of the TOP500 list the authors commented about the growing number of industrial systems, which they imply may indicate a transfer of parallel technology out of the academic world.
Top supercomputers in the UK in 2012

The following list is from the June 2012 Top500.

<table>
<thead>
<tr>
<th>location</th>
<th>system</th>
<th>processors</th>
<th>LINPACK Gflop/s</th>
<th>World rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECMWF Reading</td>
<td>Fujitsu VPP700</td>
<td>46</td>
<td>94.3</td>
<td>10</td>
</tr>
<tr>
<td>EPCC</td>
<td>Cray T3D</td>
<td>512</td>
<td>50.8</td>
<td>32</td>
</tr>
<tr>
<td>UK Met Office</td>
<td>Cray T3E</td>
<td>128</td>
<td>50.43</td>
<td>35</td>
</tr>
<tr>
<td>DRA Farnborough</td>
<td>Cray T3D</td>
<td>256</td>
<td>25.3</td>
<td>62</td>
</tr>
<tr>
<td>EPCC</td>
<td>Cray T3E</td>
<td>64</td>
<td>25.2</td>
<td>†</td>
</tr>
<tr>
<td>AWE Aldermaston</td>
<td>IBM SP2</td>
<td>75</td>
<td>14.38</td>
<td>75</td>
</tr>
<tr>
<td>University of Cambridge</td>
<td>Hitachi SR2201</td>
<td>96</td>
<td>21.3</td>
<td>*</td>
</tr>
<tr>
<td>ECMWF Reading</td>
<td>Cray Y-MP C916</td>
<td>16</td>
<td>13.7</td>
<td>128</td>
</tr>
<tr>
<td>UK Govt Communication</td>
<td>Cray Y-MP C916</td>
<td>16</td>
<td>13.7</td>
<td>136</td>
</tr>
<tr>
<td>Headquarters, Benhall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK Met Office</td>
<td>Cray Y-MP C916</td>
<td>16</td>
<td>13.7</td>
<td>147</td>
</tr>
<tr>
<td>ECMWF Reading</td>
<td>Cray T3D</td>
<td>128</td>
<td>12.8</td>
<td>164</td>
</tr>
<tr>
<td>Ensign</td>
<td>IBM SP2</td>
<td>48</td>
<td>9.53</td>
<td>199</td>
</tr>
<tr>
<td>Fujitsu Uxbridge</td>
<td>Fujitsu VX/4</td>
<td>4</td>
<td>8.6</td>
<td>214</td>
</tr>
<tr>
<td>Western Geophysical</td>
<td>IBM SP2</td>
<td>36</td>
<td>8.2</td>
<td>225</td>
</tr>
<tr>
<td>Western Geophysical</td>
<td>IBM SP2</td>
<td>40</td>
<td>8.05</td>
<td>229</td>
</tr>
</tbody>
</table>

* TOP500 had the Cambridge system listed in place 119 but only counted 64 processors, but if you take into account that it actually has 96 processors it would have been in position 80.

† The EPCC T3E system was acquired by PPARC for installation in early December 1996. Actual number of processors to be installed was unknown at the time of writing, but we assumed 64.
### Blue Joule and Blue Wonder

Blue Joule and Blue Wonder are part of the Daresbury Future Software Centre.

### The Dirac BlueGene system

The Dirac BlueGene system in Edinburgh and the iDataPlex in Durham are part of the STFC funded DIRAC consortium.

### Emerald and the Southampton system

Emerald and the Southampton system are part of the e-Infrastructure South Tier-2 consortium.

### Leeds N8

Leeds N8 provides the service for the Northern-8 Tier-2 consortium.

#### F.2 Need to move data around

To make the processors work together to do a big calculation, e.g. part of a research problem, they need to communicate and share out data. This requires a network consisting of cables and switches. There are several types.

#### What is bandwidth?

Bandwidth refers to how much data you can send through a network or modem connection. It is usually measured in bits per second, or "bps." You can think of bandwidth as a highway with cars travelling on it. The highway is the network connection and the cars are the data. The wider the highway, the more cars can travel on it at one time. Therefore more cars can get to their destinations.
faster. The same principle applies to computer data – the more bandwidth, the more information that can be transferred within a given amount of time.

**What is latency?**

This is the amount of time it takes a packet of data to move across a network connection. When a packet is being sent, there is "latent" time, when the computer that sent the packet waits for confirmation that the packet has been received. Latency and bandwidth are the two factors that determine your network connection speed.

**F.3 Amdahl’s Law**

If you have any mathematical calculation to do it can usually be split into parts. Some parts are independent (can be done in parallel) some have to be done in order (serial). If the time taken for parallel work on one processor is Tp and the time for serial work is Ts then Amdah’s Law predicts the ideal speedup which can be achieved with n processors.

Time on one processor = Ts + Tp

So time on n processors = Tn = Ts + Tp/n

So speedup = T1/Tn = (Ts + Tp) / (Ts + Tp/n)

When n is very large, the maximum speedup is (Ts + Tp)/Ts and the serial part becomes relatively very important.


Unfortunately, the additional movement of data over the network takes extra time and make the actual speedup less than ideal. Why does latency become important when you have a large number of processors?

**F.4 Moore’s Law**

In 1965 Gordon Moore, co founder of Intel, observed that over time the number of transistors that can integrated in a computer chip doubles every two years. This has been called Moore’s Law. The power of the chip is also roughly proportional to the number of transistors. For more information see [http://en.wikipedia.org/wiki/Moore’s_law](http://en.wikipedia.org/wiki/Moore’s_law). The term “Moore’s law” was coined around 1970 by the Caltech professor, VLSI pioneer, and entrepreneur Carver Mead. Predictions of similar increases in computer power had existed years prior. Alan Turing in a 1950 paper had predicted that by the turn of the millennium, computers would have a billion words of memory. Moore may have heard Dougla Engelbart a co-inventor of today’s mechanical computer mouse, discuss the projected downscaling of integrated circuit size in a 1960 lecture. A New York Times article published August 31, 2009, credits Engelbart as having made the prediction in 1959. Moore’s original statement that transistor counts had doubled every year can be found in his publication ”Cramming more components onto integrated circuits”, Electronics Magazine 19 April 1965.
The complexity for minimum component costs has increased at a rate of roughly a factor of two per year... Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least 10 years. That means by 1975, the number of components per integrated circuit for minimum cost will be 65,000. I believe that such a large circuit can be built on a single wafer.

Making things smaller

It is however hard to make very big computer chips, so to get more transistors into the same space each has to made smaller and multiple units provided. A corollary of Moore’s Law is that for the same size chip the transistors must halve in size every two years.

Integrated circuits were made possible by experimental discoveries which showed that semiconductor devices could perform the functions of vacuum tubes, and by mid-20th-century technology advancements in semiconductor device fabrication. The integration of large numbers of tiny transistors into a small chip was an enormous improvement over the manual assembly of circuits using electronic components. The integrated circuit’s mass production capability, reliability, and building-block approach to circuit design ensured the rapid adoption of standardized ICs in place of designs using discrete transistors. See [http://www.scientificamerican.com/article.cfm?id=microprocessor-computer-chip](http://www.scientificamerican.com/article.cfm?id=microprocessor-computer-chip)

Power consumption and heat

Central processing unit power dissipation or CPU power dissipation is the process in which central processing units (CPUs) consume electrical energy, and dissipate this energy by both the action of the switching devices contained in the CPU, such as transistors or vacuum tubes, and via the energy lost in the form of heat due to the impedance of the electronic circuits. Designing CPUs that perform these tasks efficiently without overheating is a major consideration in nearly all CPU manufacturers to date.

Power proportional \( V^2 \times \text{frequency} \)

\( F \) proportional \( V \), so \( P \) proportional \( F^3 \)

This means that a fast single processor consumes a lot of power and therefore gets very hot (because its not perfectly conducting).

What are the limiting factors?

Manufacturing tolerances – can lead to low yield or mal-functions, has an effect on cost Physics – getting transistors and conductors down to the size of atoms is difficult Electronics – high frequency components close together cause interference Heat and stress – build up of localised hot spots and difficulty of heat dissipation

F.5 Interesting Facts and Figures.

Evolution of the Internet worldwide
date -- number of computers connected
1968 start of ARPANet
1969 4
1982 200
1991 1000000
1996 13000000
2001 494320000
2002 568820000
2003 657077000
2004 743257000
2005 814312000
can you complete this table?


**Web browsers**

You are probably using a browser to read this right now. A Web browser, often just called a "browser," is the program people use to access the World Wide Web. It interprets HTML code including text, images, hypertext links, Javascript, and Java applets. After rendering the HTML code, the browser displays a nicely formatted page. Some common browsers are Microsoft Internet Explorer, Firefox, Netscape Communicator, and Apple Safari.

As you can see, since 2005 there has been an increase in the percentage of people using Firefox and a decrease in the percentage using Internet explorer.

**Data Storage on Disc**

1962 IBM disk = 1/2 MByte
5 1/4" discs = initial capacity was 100K, was then lifted to 1.2MB
sundry floppy discs 720kB and 1.44MB
ZIP disc 100MB
CD 650MBytes
1990 DVD = 17GBytes

**Data Storage on Tape**

1/2" tapes, 1600-6250bpi -- nearly 22.5MBytes on a 2400 foot tape (x8 tracks?)
DEC TK50 cartridge, 10GB?
8mm Exabyte 8200 tapes 2.3GB
how much data does a household video tape hold?
1991 QIC (Quarter Inch Cartridge) DC6150 cartridge tapes, 150MB

What do modern tapes hold? The largest tape is the T10000B made by IBM it holds 1000GB of data at 120 MB/s data rate.
**Largest computer ever – 1950-1963**


Each SAGE processor was 250 tons and had 60,000 vacuum tubes and occupied 50x150 feet. Each installation had two CPUs each performing 75 thousand instructions per second, one running and one in standby mode together taking 3MW of power.

As part of the US defense programme in the 1960s there were 24 inter-linked installations in concrete bunkers across the USA and Canada. The whole thing cost in the region of $8-12bn dollars. This was also the beginning of DARPANET, the US Defense network.

**What was SAGE?**

SAGE was the brainchild of Jay Forrester and George Valley, two professors at MIT's Lincoln Lab. SAGE was designed to coordinate radar stations and direct airplanes to intercept incoming planes. SAGE consisted of 23 "direction centers," each with a SAGE computer that could track as many as 400 airplanes.

The SAGE project resulted in the construction of 23 concrete-hardened bunkers across the United States (and one in Canada) linked into a continental air-defense system called "SAGE." SAGE was designed to detect atomic bomb-carrying Soviet bombers and guide American missiles to intercept and destroy them. SAGE was linked to nuclear-tipped Bomarc and Nike missiles. Each of the 23 SAGE "Direction Centers" housed a A/N FSQ-7 computer, the name given to it by the U.S. Military. The SAGE computer system used 3MW of power, and had approximately 60,000 vacuum tubes. It took over 100 people to operate.

**Transistors and Microprocessors (Intel)**

1969 Intel 4004 had 2300 transistors, 4-bit word, 0.06 Mips *
1974 Intel 8080, 6000 transistors, 8-bit word, 0.64 Mips
1978 Intel 8086, 29000 transistors, 16-bit word, 0.66 Mips
1982 Intel 80286, 134000 transistors, 16-bit, 2.66 Mips
1985 Intel 80386, 275000 transistors, 32-bit, 4 Mips
1989 Intel 80486, 1.2M transistors, 32-bit, 70 Mips
1993 Intel 80586, 3.3M transistors, 126-203 Mips
1999 Intel Pentium III, 9.5M transistors, 32-bit
2004 Intel Itanium has over 15 million, 64-bit, 1200 Mips
2007 Intel Xeon E5472, 820 million transistors, 64-bit,
2009 Intel Xeon E5540, 731 million transistors, 64-bit,


* Mips = million instructions per second