Q. Jl R. astr. Soc. (1982) 23, 485–504

STARLINK

M.J.Disney
Astronomy Department, University College, PO Box 78, Cardiff, CF1 1XL

P.T.Wallace
Computing Division, Rutherford and Appleton Laboratories, Chilton, Didcot, OX11 0QX

(Received 1982 May 22)

1 INTRODUCTION AND HISTORY

STARLINK is a project set up and funded by the SERC to provide first-rate image-processing facilities for all British astronomers over the next decade or so. Because it proved possible to get the project going very rapidly there has been little time to publish anything accessible about it so far. We hope however that STARLINK will have a considerable impact on British astronomy in the years to come and if astronomers are to make the best use of the project it is important that they understand what they can and cannot expect of it. We here describe the history, philosophy, present status and likely future development of STARLINK.

The increasing number of telescopes and the dramatic and continuing improvement in detectors and plate-scanning machines provide astronomers with a swelling flood of data, mostly in digital form. We know of astronomers who already return to the UK from a single observing run with *more* than 10^{10} bits of digital data. Faced with the problem of converting these data into interesting astrophysics a number of groups in the UK applied to the SERC for funds to purchase various types of computer-based image-processing equipment to meet their own parochial needs. Worried by the increasing expense and by the necessarily piecemeal approach the SERC set up a panel in 1978 to look into the matter across the whole wavelength range. The panel, of which one of us (MJD) was chairman, consisted of four astronomers and two independently appointed computer experts. It consulted a wide range of astronomers and computer people both from at home amd abroad and reported to the relevant SERC committee in early 1979. The panel found that the astronomical demand on one side and practical constraints on the other left between them a very narrow range of sensible choices. It recommended that a network of image-processing nodes be set up around the UK, each based on common equipment, and all managed by a central staff directed from the Rutherford–Appleton Laboratory. The purpose of the commonality and of the low speed line links between nodes was to share in the development and use of a core of common software that would be easily comprehensible and accessible to all UK astronomers.

Having accepted the report the ASRB assigned a sum of about £2 million and 17 staff to the project. Most of the equipment was installed and connected in early 1980 and the project was officially opened by the Minister of State for Science in 1980 October and named STARLINK. However, much remains to be done before the system reaches full maturity.

2 THE IMAGE-PROCESSING PROBLEM IN ASTRONOMY

Most often these days the end result of an observing run will be one or more inscrutable magnetic tapes. On these tapes will be a series of one-, two- or three-dimensional images stored in digital form. The images may come from a diverse variety of sources which include: optical spectrograms or CCD pictures taken with a ground-based optical telescope like the AAT; the output of similar instruments on satellites like *IUE* and the *Space Telescope*; X-ray 'photographs' taken with *Einstein* or *Exosat*; radio pictures either in the continuum from synthesis telescopes or as a series of super-imposed 21-cm images taken in different velocity channels from an instrument like Westerbork; digitized photoelectric scans of Schmidt plates originating for instance in the COSMOS or Kibblewhite machines; TV frames of planets taken by *Voyager*; and three-dimensional pictures obtained with imaging Fabry–Pérot systems like TAURUS. Not only are these tapes inscrutable but the images they contain are frequently crude and dirty. Before they can be turned into useful astrophysics they must first be calibrated, corrected and cleaned. And even these relatively clean data may still require further processing which may include foreground and background subtraction, filtering, contouring and integration under contours, contrast enhancement and so on before they are usable for the real purpose at hand. Finally, the data must be converted into a form where they can be directly compared with the output of various theoretical modelling processes, and then prepared for publication. This whole procedure is what we refer to as image processing. Since it generally requires from the astronomer a far greater investment of effort and time than the mere acquisition of the observations, we clearly have to do the best we can to make the whole procedure as efficient and streamlined as possible. Equipped with a good image-processing system the astronomer is able to concentrate on what he likes and can do best: without it, he may become ensnared in a nightmare tangle of programs he cannot make run, of machine printout he cannot make sense of and of data he cannot publish.

To size up the problem it is worth examining the various formats in which images or frames may come, and then estimate the likely total input of frames to the UK community from different sources.

Analysis shows that the data come, and will continue to come, in three main *formats*; these formats are largely determined by the detecting devices and it is fortunate, and fortuitous, that present radio and X-ray formats resemble those in the optical. They are:

(a) *Spectral* format, composed of a few one-dimensional lines of data, basically $1024 \times 8 \times 10$ bits, e.g. IPCS spectra.

(b) *Television* format, which is basically $512 \times 512 \times 16$ bits, e.g. CCD pictures.

(c) *Grand* format which is typically 2000 × 2000 either in (i) 8 bits from photographic plate-scanning machines, or (ii) 16 bits from CCD data, e.g. Space Telescope Wide Field Camera.

Next one can roughly estimate the 24-hr bit-rate from instruments of various kinds:

Instrument	Format	Pixel/ frames	Bits/ pixel	Frames/ day	Bits/day
Optical photographic	Grand	$(10^4)^2$	8	10	8 G-bits
Optical television	TV	$(512)^2$	16	10	40M
Radio ap. synthesis	TV	$(512)^2$	10	2	5M
Optical spectroscopy	Spec.	10^4	10	10	1M
Imaging X-ray satellite	Grand	$(10^3)^2$	16	10	160M
Satellite planetary camera	TV	$(512)^2$	16	channel capacity	1G
Ground-based IR	—	—	—	—	Very small

where the above figures refer only to useful astrophysical frames and not to calibration frames, which may be far more numerous.

Convolving the above estimates with the suite of instruments to which UK astronomers will have regular access over the next five to seven years one easily concludes that by far the largest source of data will come from ground-based optical astronomy. If we can handle, including calibration data, an anticipated 100 TV frames/day from the optical and five grand format photographic frames/day (plate scanning), we should be easily able to cope with the additional load from other parts of the spectrum.

Of course not all programmes imply the same processing requirement per unit of raw data. Radio-source identification, for instance, may require very little. Equally there are projects like galaxy photometry implying an enormous effort. The radial light distributions of galaxies are such that much, perhaps most, of the light lies at very low surface brightness levels where it can only be detected, and the foreground and background contamination allowed for, by painstaking analysis, experiment and re-anlysis. Our hesitant judgment is that averaged overall an astronomer will need 2–5 hr of terminal time per provided hour of observing time.

No reduction system will be satisfactory unless it fits naturally into one's research time. From talking to numerous astronomers it is clear that:

(a) The average customer wants *flexibility* and *informality* in the reduction process. He must fit his terminal time into a multitude of other duties and he wants unpredictable amounts of terminal time separated by unpredictable intervals for reflection and experiment depending on the outcome so far.

(b) At many stages it will be essential to *interact* with data on a more or less real time basis. This general requirement is especially strong in astronomy because: (i) one needs to use one's judgment and experience in dealing with marginal and complex data, often affected by spurious background or foreground sources; (ii) the required information may be patchily distributed, e.g. specific stars in a cluster; (iii) data may be affected by changing meteorological conditions; (iv) the eye is better than the machine in detecting blemishes; (v) batch processing algorithms can only be written to deal with the foreseen. The unforeseen may be scientifically more exciting.

(c) The astronomer wants a great deal of *software support* for experience has taught him he may otherwise spend the majority of his 'research time' writing and debugging programs.

Fortunately it transpires that most reductions are made up of a string of relatively small operations that are common across a wide field. Such operations include the display and inspection of frames with zooming and object identification using a cursor; division of one frame by another for calibration; removal of geometric distortions and registering of frames on top of one another; the subtraction of frames to determine interesting differences and addition to see more detail; smoothing and filtering; contouring and integration under contours; background determination and subtraction using interpolations; profile fitting and so on.

And at a higher level the reduction of one galaxy spectrum or one H–R diagram may be much like another. So although the software effort required may be enormous the commonality of techniques and of final aims may be so similar between astronomers with quite different astrophysical motives that there is an opportunity, even a necessity, to share this software effort out, and avoid what would otherwise be a great deal of unnecessary duplication.

It follows from the above that the requirements of the average astronomer will not be met until there is available, close to his office, an interactive computer terminal (with, as it turns out, some rather sophisticated colour display devices) which he can use without much prior booking, and which is provided with an extensive and easy-to-use software library.

Before setting up STARLINK a thorough survey was made of image-processing methods both at home and abroad, not only in astronomy but in other fields such as crystallography, engineering and earth resource science. Although much was learned we could find no ideal model appropriate to our needs. The crux of the matter appears to be the potentially enormous and rather specialized software effort involved and the consequent compulsion to share as much of it as possible. Other people's experience shows that such sharing will not work well in an informal environment because:

(a) Quite small differences in hardware and in operating systems make translation from one machine to another extremely difficult and time-consuming. And even when compatibility has been achieved and maintained it is more than likely to deteriorate or vanish as system developments are made at either end.

(b) Even supposing compatibility can be achieved and maintained software cannot be shared unless it is carefully documented. There must exist, between the writer and the potential user, a well-defined documentation stage.

In setting up STARLINK we become more and more conscious of the software problem. Unfortunately, it is almost impossible to describe the situation in any quantative terms. Nevertheless those in the positions to know best emphasized that it ought to be our cardinal consideration. And so it has been.

3 HARDWARE CONSIDERATIONS

The size and speed of the computing machines and line-links we need for image processing can be worked out from rather simple considerations. These in turn impact the costs and the sort of service one can aim for.

One can think of image processing as an astronomer sitting in front of a console linked to a processor. Data processing will consist of a number of stages in each of which: (a) the human takes T_H s to think and execute; (b) the data set is sent to the processor by link taking T_L s; (c) the processor completes its calculation in T_P s; (d) the data returns for display at the console in T_L s. The total time for each stage is then $\sim T_H+T_P+2T_L$. In an efficiently matched situation $T_H \sim T_P \geqslant 2T_L$ for otherwise the human, the processor, or both, will be idle. Let

F = no. of picture elements (pixels)/frame,
b = no. of bits/pixel,
$\langle N \rangle$ = average no. of calculations/pixel/processing stage,
P = processor speed in calculation s^{-1},
L = line speed in bits/s^{-1}.

We decided to aim for $T_H \sim 30$ s, $\langle N \rangle \sim 10$ (arrived at by averaging out a number of typical operations) and $b = 32$ bits. (To address pixels directly in frames above 248^2 32-bit words are necessary.) Then

$$T_P \leqslant T_H \text{ implies } P \geqslant \frac{F\langle N \rangle}{T_H}, \qquad (1)$$

and

$$2T_L \leqslant T_H \text{ implies } L \geqslant \frac{2Fb}{T_H}. \qquad (2)$$

Then:

(a) For TV frames these imply that $P \geqslant 100$ k calculation s^{-1} and $L \geqslant 500$ k bit s^{-1};

(b) For Grand Format TV frames these equations imply $P \geqslant 1 \cdot 3$ million calculation s^{-1} and $L \geqslant 8$ Mbits s^{-1}.

(c) If the processor is to address and manipulate Grand Format frames effectively it will need a word length of 32 bits and a fast memory size of at least a megaword and preferably more. To handle several such frames at once, which is a desirable requirement, it will need a virtual memory system which will automatically swap data from disc to memory and back again in such a way that the programmer is unaware of the limited memory size.

The implications of these sums for UK astronomical image processing are as follows:

(1) The very high line speeds cannot at present be sustained any great distance. Either astronomers must travel to the processor, or else processors must be distributed at the main sites where astronomers work.

(2) The intermittent demands on the processor are too large to be supplied as one of the tasks on a normal main frame machine like a university computer which has a large number of other customers to satisfy.

(3) Conventional 16-bit minicomputers fall short of the processing requirements in several important respects; principally in memory size, addressing capability and calculation speed.

(4) Modern 32-bit mega-mini machines with several megabytes of core and memory access rates of about 8 Mbyte s^{-1} *will* provide the necessary power

but at about £250k each. To fit the cost inside a feasible budget implied that no more than about half a dozen such machines could be supplied in the UK. If they are installed at the six sites where the largest number of astronomers are to be found then 80 per cent of the 200-strong community will be serviced and 20 per cent will remain unsatisfied.

Each computer will need at least one sophisticated image display system which can store and display more than one TV frame and which can, independent of the host machine, provide simple and fast picture manipulations which include pan, zoom, overlay, blink and false colour representations. Colour is desirable for the display of a third dimension (for instance velocity or polarization information) and to overcome the limited dynamic image of a normal 7-bit black and white display. Such displays are not cheap, costing between £20k and £30k each.

4 THE STARLINK DESIGN

This was arrived at after considering a number of alternatives which ranged between a single national centre to which astronomers would travel and a completely *laissez-faire* system in which separate groups would each set up local systems to meet their own requirements. The national centre concept was ruled out because the vast majority of users would have a substantial journey to reach it and hence it would fail to provide the informally interactive system required by the very nature of the research. Likewise, *laissez-faire* was rejected because of the unnecessary duplication of software effort and for its failure to provide for astronomers at the smaller institutions.

The finally adopted design is a linked minicomputer network in which a number of compatible computers are provided by the SERC and connected by a low speed ($\leqslant 9.6$ kbits s^{-1}) communications link to a centre. Such a network is referred to as being 'star-linked', hence the name. Each node has an obligation to some smaller institutions nearby and the whole system is managed from the centre. The centre purchases, runs and services the network, pays and hires personnel, provides and updates systems software, ensures documentation standards and coordinates the provision of applications software. Such a system should meet the needs for an interactive facility close at hand for a substatial majority of astronomers and also provide a service for smaller institutions. Because system software will be written at a single centre and transferred to other nodes by links, compatibility can be ensured and maintained, and hence software sharing will be encouraged.

The existence of the dedicated links means that:

(a) software can be sent rapidly up and down the line;

(b) associated documentation can reside in a single updated file and be interrogated as necessary;

(c) astronomers and software people can communicate easily with one another and with their collaborators around the country;

(d) software fault repairs can be disseminated rapidly;

(e) evolutionary changes can be introduced and the effects on compatibility measured almost in real time;

(f) some data can be sent through the system, e.g. star catalogue data, and this will be a step towards a common data base;

The choice of sites was based on the following principles in order of importance:

(a) contiguity to the maximum number of user astronomers;

(b) ease of access for non-local users;

(c) accessibility for non optical/UV astronomers. A survey of applicants to PATT and *IUE* quickly settled the choice of the first six upon Cambridge, London (UCL), RGO, ROE, RAL and Manchester. We emphasize this choice is dictated by national rather than parochial requirements. One more site has recently been installed in Durham. Were a further site to be founded in the future, it would, for the same reasons, probably go to the Midlands.

The centre was placed inside the computing division at RAL because of the experiece they already have in procuring and operating a wide range of computers, because of their ambitious communications facilities which include satellite links and because of their systems expertise which is mostly lacking elsewhere in the SERC. Their previous experience in setting up the Interactive Computing Facility, which is rather similar to STARLINK, but which is based on smaller machines, has proved invaluable. We should not forget that there are also some astronomers on site mostly connected with projects like *IUE* and IRAS.

There is a STARLINK staff of 17 directed jointly by a Project Manager (PTW) and a Project Scientist at RAL. There are five operations staff and programmers at RAL, six site managers and four application programmers distributed round the network. While these staff maintain and coordinate the system *we should emphasize that they are not intended* to produce the majority of common application software that will be needed. This can only come from active astronomers and programmers around the country who have the expertise, the motivation and the data. The central staff must concentrate on systems, on communication, on coordination, on documentation and on long-term support of the most urgently needed and most widely used software packages.

5 THE EXISTING NETWORK

This has been working in various degrees of completion for two years and there are already 400 or so accredited users.

STARLINK currently has seven VAX computers built by the Digital Equipment Corporation. Six are the VAX 11/780 model, each with typically four Mbytes of main memory and over one Gbyte of disc storage. The remaining machine is the slightly less powerful VAX 11/750 (at Durham), with about half the memory and disc capacity of the others. VAXs were selected for four main reasons. First, they met our detailed specification and peformed well on image-processing benchmark tests which we prepared. Secondly, they run a powerful and well-documented operating system called VMS. Thirdly, they are machines near the beginning of their design life with a good deal of potential for expansion. Fourthly, they appear to be the preferred choice of other astronomical groups abroad; this will facilitate the exchange of software. For instance, the ST Institute will have four VAXS, the ST Wide Field camera has two, the Faint Object Camera on ST has one at ESTEC, ESO will have two, the AAO has two, Mt Stromlo has one, the DAO in Canada has one, Meudon has one, Westerbork has one, and the Italians are building up a network of eight VAXS called ASTRONET, rather similar to STARLINK. Our VAXS have proved very reliable and satisfactory.

The image display systems proved harder to choose if only because there is a large variety spanning wide choices of performance and price. After a world-wide evaluation, we picked Sigma ARGS machines as the preferred displays for our use. Two of them are now installed at each of the nodes (one at Durham) and connected to the VAXS by 8 Mbyte s^{-1} channels. There are in addition all the usual peripherals at each site.

The outlying nodes are connected to the centre at Rutherford, and hence to each other, by dedicated lines running at between 2·4 and 4·8 kbits s^{-1}. We use the propriety DECNET communication protocol at present. However, networking is in a generally fluid state and we expect shortly to switch over to the newly defined international X-25 packet switching protocol which is already running in experimental form on the SERC network. As anticipated the network facility has proved to be invaluable, and not only for its original purpose of management and software exchange. For instance, collaborating astronomers find it a very convenient channel for mail and for small data-streams. So far the low line speeds have generally proved adequate.

All equipment remains the property of the SERC and all full-time staff are hired by SERC through Professor J.T.Houghton of the Rutherford–Appleton Laboratory who is the Accounting Officer in charge of the STARLINK project. However, management is decentralized as much as possible with each node having its own Local Management Group, and it is these groups, via the local manager, which allocate time and resources on the local node to individual users. This decentralization ensures that any spare capacity is used to the best advantage for non image-processing purposes. The VAXS power is adequate for what are usually mainframe activities. For instance, at Cambridge Dr Aarseth runs his many-particle gravitational simulations on the VAX at night and weekends. When nodes become saturated it will be up to the Local Management Groups to assign priorities, bearing in mind the original purposes of STARLINK.

The overall scientific goals and priorities are set by the Scientific Advisory Group (SAG) which meets about four times a year. SAG, which is largely composed of user astronomers, is presently chaired by one of us (MJD) and is advisory to the ASR Board. SAG advisers on matters like the development of software, the purchase and siting of new equipment, the hiring and replacement of staff, international collaboration and the procurement of funds.

Astronomers not sited at a node are at a grave disadvantage at present and it is SAG's highest priority to improve their position so far as possible. Accredited users can travel to use the most convenient node, and indeed they should be represented on their Local Management Group. Travel funds are available within STARLINK for this.

Nevertheless, travelling is still a handicap and where possible we are trying to link in remote users directly by lines; STARLINK can allocate some terminals and hard-copy devices for this purpose. While such links will greatly facilitate software development and the performance of other valuable tasks, such as spectral reduction, it is as well to remember that the linespeeds do not approach the 8 Mbyte s^{-1} rates which are sometimes necessary and available within each node. An alternative is to try to run the STARLINK system on some of the many non-STARLINK VAXs which are springing up around the community. Intending new users should explore the possibilities with the management (PTW) at RAL.

We should emphasize that STARLINK is intended to be an expanding and evolutionary system and up-to-date information is circulated to users either via the network or through the newsheet *Enterprise* which is published every four months.

6 SOFTWARE

STARLINK's software challenge was and is immense. Perhaps surprisingly, this comes about not because of the intricate numerical manipulations involved in areas like image enhancement, but instead because of the first of three the main requirements discussed in Section 2 – that the system must be flexible and informal. These properties are hard to design into a system, because they can be only vaguely specified and because the user's perception of them may differ radically from that of the implementors.

Those members of the astronomical community who advise the STARLINK project occupy a continuum of opinion which extends from ultra-pragmatists (who believe that little if any supporting software over and above the applications routines themselves is required) to idealists (who cannot brook the slightest departure from complete portability across machine types or any suggestion that implementation questions be allowed to influence the system as seen by the user). Those at the pragmatic end of the continuum see nothing wrong in selecting a powerful computer, with a friendly and secure operating system and a lot of good basic software, installing one at several sites and then standing well back – expecting vast amounts of useful software to be written by the astronomers. Those at the other extreme consider it essential that an elaborate, powerful and independent machine-supervisory sub-system be designed and then implemented on the chosen machine type; only then can application programming (and indeed astronomical use) begin.

At various times before STARLINK both these approaches have been tried and have always failed to produce an optimum result. The first fails because the programs that are written, apart from overlapping in function, are usually undocumented, highly specific to the particular problem which led to the program being written and to the local peculiarities of the computer, and are incompatible with one another (in data formats especially). The second approach fails because no one can agree on what the ideal system is, and anything that is actually implemented is too late and grossly inefficient. The most successful systems seem to be those that have their own software environment but are written by (or at least conceived by) one individual, an approach difficult (and indeed dangerous) to apply to STARLINK.

All this was known at the time the STARLINK software effort was about to begin, when a group of about twenty people were invited to attend an international workshop at Appleton Laboratory in 1980 June to try to decide how the development of STARLINK software should proceed. The 2-day meeting was attended by representatives from the SRC (now SERC) and several UK universities and from Europe, the USA and Australia. Included on the comprehensive agenda were discussions covering both ideal future systems and experiences with existing ones (e.g. the Kitt Peak IPPS). There was a surprisingly high degree of consensus (arguably due to a comparative lack of FORTH and UNIX enthusiasts). The main conclusion was that STARLINK should provide an overall supervisory system within which applications programs would be run. It was agreed that external data interchange should be via the FITS tape format (Wells, Greisen & Harten 1981); a similarly self-defining internal format should be used, possibly based on the AAO/MSSSO proposals for hierarchical data structures. The supervisor should, like the Groningen HERMES system, include a user-friendly command language operable in both interactive and pre-programmed nodes and including on-line 'help' facilities. To simplify applications progams and to foster standardization, interface routines would be provided, enabling programs to access parameters, data and graphics devices in a uniform and convenient manner. Once these interfaces had been defined, applications could be written which would survive major changes in the supervisory system and might even be portable to other machines. These recommendations led to the development of the first 'STARLINK Software Environment' (now in use as the basis of the ASPIC picture processing system) which allowed some STARLINK applications programming to begin in late 1980.

In the early stages of planning STARLINK, the question was addressed of whether to provide dedicated applications programming effort and if so where to site the programmers. The two extreme approaches are:

(a) to have no official programmers but instead to exploit the work of astronomers and research students, and (b) to have a large team of programmers at the central location. The STARLINK Scientific Advisory Group decided that it was desirable to have at least four STARLINK programmers, who though centrally managed should work closely with astronomers. The initial four were appointed (at various times and for various periods) at Durham (STARLINK node), Manchester, UCL and Cambridge.

An important interface in STARLINK is the one between those implementing software – users as well as those employed by the project – and the astronomical community at large. Accordingly, a small number of Special Interest Groups (SIGs) were created to monitor progress in specific areas and to advise the project. The current list of SIGs is as follows:

(1) software environment,
(2) spectroscopy,
(3) database,
(4) *IUE*,
(5) 2-D photometry (picture processing).

One of the first duties of the spectroscopy SIG was to select, from several candidates, an existing spectrum reduction and analysis system to act as an interim facility (at least for IPCS data) while the official system was developed. The one selected was SPICA, written by K.Shortridge and others at UCL, and this has since been much used and developed.

The picture processing SIG, in the absence of any existing VAX image-processing packages, recommended that some programs written within the emerging STARLINK Software Environment be collected together and released as one package. The result, ASPIC, has now been in use for almost a year and currently includes over 100 programs. It also features a surprisingly powerful command language called DSCL which is implemented as an extension to the standard VAX control language DCL.

The *IUE* SIG were quick to endorse the selection of the DRP/STAK/TRAK programs written by J.Giddings and S.Wright at UCL. These programs enable extraction of the *IUE* échelle spectra and a wide variety of reduction and analysis facilities.

It is impossible to list here the many other powerful and useful systems developed for STARLINK and distributed to all nodes, let alone those used at one location only.

Let us return now to the question of why it is necessary for STARLINK to provide any software at all over and above astronomical applications programs. To examine the problem we will consider the case of an astronomer programmer who wishes to develop software for the interactive reduction and analysis of optical spectra obtained with IPCS detector. What happens if we give him a friendly computer (namely a VAX) and let him get on with it?

First he will define a standard way of storing the spectra on disc, probably one observation per file, with various parameters (notes, date and time, telescope position, name of source, etc.) stored in known places along with the main data array. In the case of the IPCS he will probably allow the data array to vary in size, but for other instruments may be tempted to fix the size.

He will then write a program which, driven by a dialogue with the user, carries out operations on these spectra, generating new ones and producing a variety of reports. There will, of course, be functions for reading the data from magnetic tape and probably also ones for archiving reduced spectra back on to tape. There will also be a variety of graphical operations including plotting on display screens (in some cases with interaction using a joystick-controlled cursor or the equivalent) and on to hardcopy devices.

We thus have a comprehensive and probably highly efficient system for dealing with IPCS spectra. What, then, can possibly be wrong with such an approach?

First, a single program which contains the entire repertoire of functions is apt to be huge. This can lead to problems of address space (which have to be tackled in highly host-dependent ways). Also, link times can grow to unacceptable levels, with the most minor program change requiring many minutes of computer activity. There are also difficulties when the user wishes to provide his own functions; he has to modify a copy of the standard system and link the whole lot.

Secondly, it can be difficult to allow for changes in the data format. Changes are always needed sooner or later, either because of developments in the instrument or because various 'reduced' forms of the data need to be accessed. What tends to happen is that the original raw data format, which was lovingly crafted to hold just the right information in the most natural and convenient way, has to be 'fudged', For example, it might be necessary to squeeze in an extra instrument parameter by sacrificing two characters at the the end of the 'notes' field to hold the new field and then adding 24 to the UTC hours to flag that this has been done. Such arcane adjustments, not surprisingly, lead to endless troubles. Furthermore, the requirement will inevitably arise to accommodate spectra from other instruments. A set of elegant programs for extracting astrophysics from calibrated IPCS spectra should be equally applicable, for example, to *IUE* spectra and there will come a time when *IUE* data have to be fitted into the IPCS format to allow this to be done.

Thirdly, each reduction and analysis package (IPCS, *IUE*, CCD, electronography, PDS, etc.) will have its own idiosyncratic style. Some implementors will provide a 'menu' system. Others will grant instant access to all functions to any user who can memorize several hundred four-character strings. Others will use question/answer dialogues. Some will provide lavish on-line 'help' facilities. Some will include algebraic expression handling. Some will use positional parameters, others keywords, others a stack. Some will require all parameters to be entered; others will offer defaulting to sensible values. Some systems will crash unless spoken to in the right tone of voice; others will have the most comprehensive protection imaginable. And each system will have its own set of devotees who cannot comprehend the stupidity and lack of taste of those who favour the rival systems. There will also be considerable duplication of effort, both in the 'system' areas (argument decoding, expression parsing, function activation, etc.) and in the application areas, where a large set of basic utility functions would have to be provided separately by all the systems.

Fourthly, there are few standards for graphics and any system developed by a small group is likely to be idiosyncratic in style and may even depend on local facilities (e.g. special printer/plotter spooling arrangements). In the graphics area, for example, it is not unusual for a poor degree of device independence to be present, with quite different routines called for plotting on, respectively, the terminal, a pen plotter and a printer/plotter, all of which complicates life for any astronomer who wishes to write his own programs when he has a special problem to solve.

Finally, we have yet to enter the era of truly portable software and it is all too easy to write programs in a supposedly standard language like FORTRAN and then to discover that they are highly dependent on one machine type or even on one specific machine.

STARLINK hopes to overcome all these difficulties by providing a '*software environment*' (SSE) within which easy-to-write, efficient and portable applications programs can be run. The environment has three main areas: the *supervisor*, the *data system*, and the *graphics system*. Much work on all three areas has been done but a lot more remains to be done.

The *supervisor* includes a *command language*, which is the interface to the astronomer, and a mechanism for running applications programs. The latter are single-function separately linked programs, and it will be possible for the user to draw on his own libraries of such programs without the supervisor having explicit knowledge of them. Thus adding a private function will be easy, fast, natural and safe. Mechanisms may be provided to permit multiple programs to be run simultaneously. The command language will offer both interactive and pre-programmed modes of operation. There will be simple expression evalution, looping and conditional constructs and macro- and procedure facilities. There will be several syntaxes for program activation, including parameter passing by position, keyword or stack. Prompting for parameters and the substitution of default values will be handled automatically without the application program having to take explicit action. Extensive on-line 'help' facilities will be provided. It is important to understand that the command language will not necessarily be portable from one machine type to another and will strive for convenience and efficiency by cooperating with the VAX/VMS operating system in many ways (for example the astronomers will be aware of and will be allowed to exploit the relationship of his data to the machine filing system). This is in contrast to the applications programs, which will contain no VAX/VMS dependent features above a tiny, innocuous and more or less irreducible minimum. However, when it becomes necessary to implement the STARLINK system on another machine type (which will happen sooner or later), the command language will in fact be transferred with its appearance largely unchanged. The minor compromise in portability in the applications programs (specifically, relying on a large address space, and on a memory pointer facility in FORTRAN) and the rather more major compromises in the command language are there: (a) to avoid writing software already provided by the computer supplier (e.g. disc-filing systems including user quotas, protection, automatic garbage collection, etc.), and (b) to gain efficiency. The latter is an extremely important issue.

The *data system* puts off the problem of predefining image formats for every purpose by using self-defining structures. This is the approach used in the tape interchange standard FITS. However, in its disc formats STARLINK is going further than FITS in allowing multi-level hierarchies of data. Each *data object* in a hierarchy has a *name* (by which that particular instance of an object is referenced), a *type* (which identifies that object as belonging to a certain class with a known underlying structure) and a *shape* (i.e. dimensionality). A non-primitive data object is made up of data objects recursively; the primitive data objects have the familiar types easily accessible in FORTRAN 'character', 'integer', 'single precision real' and so on. Apart from the primitive types, STARLINK will have to maintain centrally lists of reserved high-level types to avoid clashes between implementors (although problems will only arise if incorrect data of a reserved type are offered to an application program which cannot cope with the unexpected varieties). The rules for any given type may be rigid – a list of sub-objects, all mandatory – or quite loose. In an analogous way, applications will range from ones which insist on a very specific data type to others which can search any structure the user points to and make intelligent guesses (for example a 2-D plot program could exhaustively search a structure and then plot the largest 2-D array of numbers that it had encountered). Furthermore, there will be a basic set of utility applications which work on primitive data objects; the astronomer will tell the program, via parameters, where he wants his graphics to go and the program will, to a large extent, not have to worry about the details. STARLINK has selected the proposed ISO standard GKS (Graphical Kernel System) as its central graphics system. Using GKS may enhance the portability and long-term supportability of STARLINK applications and will ensure that certain advanced graphics techniques not yet widely used by astronomers will be available when needed. As well as permitting direct access to GKS routines from applications programs, STARLINK is providing several layers of useful software above GKS, ranging from low-level routines which simplify straightforward graph plotting to high-level facilities capable of plotting attractively presented and fully annotated complete graphs. In fact, there are plans to go further than device independence between physical workstations and to operate in terms of 'virtual devices' which will frequently be subsections of physical devices. Each virtual device will be of one of two sorts – line or picture – and some devices will be able to support either of these concepts, sometimes simultaneously (for example the Sigma ARGS display). The system will retain a knowledge of what has been displayed, so that enquiries via trackerball controlled cursors, etc., can yield coordinate and data information. These arrangements solve two problems commonly experienced in image-processing systems which are composed of many separate programs (especially written by different people). The first is that the user would like to be able to split up his display devices to suit the astronomical problem, not the applications program. So one day he wants to see one picture on his screen, the next day he wants four pictures, and the day after two spectra one above the other. The STARLINK system will allow him to arrange his displays as an activity separate from running the applications programs, which will plot equally well on the top half of a Tektronix 4010 screen or the bottom right-hand quadrant of an ARGS overlay plane. The second problem that will be easier to solve with the STARLINK system is how one program can interact with a picture plotted by another without explicit cooperation. It would be a pity, for example, if a new and sophisticated display program could not be used to set up the screen ready for the running of an old established interactive stellar photometry application.

Most of the subroutine interfaces between the SSE and the applications program have yet to be finalized although the data system calls are essentially complete and the low level graphics routines already exist. There is, in addition, a programming standard which deals with details of the programming language (to date the only allowed language is USA Standard FORTRAN 77, with one or two extensions and a number of prohibitions), techniques and style. The standard recognizes the heterogeneous nature of the STARLINK community of programs by being rather generously worded with considerable scope for personal taste.

All STARLINK applications software will run within the SSE, although many rather straightforward utilities (e.g. programs to do astrometry or compute guide star positions) will not lean heavily on the SSE and will run more or less unchanged under VAX/VMS. Four categories of software can be identified:

(i) data acceptance,
(ii) data reduction, manipulation and analysis,
(iii) utilities, and
(iv) database.

Data acceptance programs will translate external data tapes into the STARLINK internal form. FITS format will be the preferred form (although even then the astronomer will usually have to say where the tape came from in order to secure the most appropriate translation). Native data will be acceptable from many sources – IPCS, IDS, other spectral formats, *IUE*, plate scanners, CCDs and other non-photographic detectors, synthesis radiotelescopes, X-ray imaging satellites, and the *Space Telescope*.

Data reduction, manipulation and analysis facilities will include both highly specific programs (e.g. spectrum extraction from raw *IUE* échellograms, full automatic calibration of IPCS data) and general purpose programs (*n*-D image–image and image–scalar arithmetic, smoothing and other filtering, various forms of editing, background/continuum estimation, display and plotting, function fitting, spatial warping, etc.).

Utilities will include both astronomical and non-astronomical programs. Examples of the former are celestial coordinate transformations, starfield overlay plotting, topocentric position/velocity, time/date conversions, predictions of celestial phenomena, and astrometry. Sometimes more than one grade will be provided – for instance Earth velocities will be calculable either fast and fuss-free (but to a limited accuracy) or to the utmost accuracy (but involving precomputed ephemeris files and other complications). In all cases, the performance will be very carefully assessed and documented; the most refined routines will be definitive. Non-astronomical facilities will include mathematical, statistical and graphical libraries, and of course a wealth of system-oriented programs (for example provision for reading, writing, copying and analysing 'foreign' magnetic tapes).

A major rôle for STARLINK will be the provision and maintenance of various databases and the tools for accessing them. A start has already been made in this direction, with a useful set of magnetic tape star catalogues provided at each node (and many more available on request) together with software to access them. It is hoped to have a unified and powerful database management system eventually with extensive and varied data archives – not just star catalogues, but atomic and molecular spectroscopy data, plate catalogues, observing logs, and more. It will very likely become necessary to store each really huge database at an appropriate single location, to be accessed remotely. Fundamental astrometric catalogues and so on will continue to be available on every machine.

7 THE ASTRONOMER AND STARLINK

The astronomer has three rôles within STARLINK: user, adviser and implementer.

What does the astronomer experience as a STARLINK user? Having applied for and been given a STARLINK account, he will be given a kit containing documents about both the VAX and STARLINK. With the help of the node manager (and usually his colleagues) he will get to grips with the machine either through using a standard software package and beginning to perceive dimly what is going on, or alternatively by pretending to be a non-astronomical VAX user and experimenting (with the aid of the vendor-supplied primer) with text editors, FORTRAN, disc files, magnetic tapes, and so on. The VAX is very friendly and few astronomers, if any, experience serious difficulties. He will then go on to use whichever of the astronomical software packages are relevant to his work, and perhaps to develop programs himself. In the fullness of time, most of his work will be done within the STARLINK Software Environment and the general flavour of this way of doing things can be judged now by using the ASPIC package. Having initiated the SSE (possibly automatically at LOGIN time) the astronomer will run applications programs by typing commands with parameters expressed in a way which suits him. He may omit parameters, causing explanatory prompts to appear or – in appropriate cases – sensible default values to be taken. HELP facilities, including brief lists of programs and information on the command language will be instantly and easily accessible, allowing effective use of the system to be made from the very first encounter. Pre-programmed sequences may be set up and then called on command; 'batch' facilities will be available, which will permit processing after the astronomer has relinquished his terminal. He will be able to run programs written by many different people, including himself, in a uniform manner and with information flowing naturally from one application to another. The essence of the approach is that the applications programs will appear to be cooperating with one another to solve the astronomer's particular problem despite there having been no collusion between the authors; the individual programs will be seen as tools rather than self-contained applications.

Once the astronomer is an active user of STARLINK he naturally develops opinions on what changes should be made to the software and what new software should be written. There are a number of ways he can express himself. The first is to approach the programmer directly. This is perfectly proper, and the opportunity for day-to-day interaction between the user and the programmer is the crucial factor in siting applications programmers at the nodes rather than in a central team. Another way is to try to persuade local experts (notably the site manager) to implement a local solution. Another is to make a report to the central management, which will duly be passed on to the appropriate programmer. Finally, major changes and new initiatives can be brought to the attention of the appropriate Special Interest Group, who may then recommend them to STARLINK management (and in practice directly to the programmer).

It is expected, in fact, that most users of STARLINK will be perfectly capable of writing their own applications (bearing in mind that an important function of the SSE will be to *simplify* this task) and in many cases software potentially of lasting value will be developed in this way.

The fundamental challenge of STARLINK is to share out both the hard work of writing and the benefit of using the vast amount of software most of us are going to need. Having only four programmers of its own there is no way that the STARLINK system itself can provide more than a very modest fraction of the necessary effort. The STARLINK management's job is to encourage, advise and support the astronomers who are willing and able to contribute their own efforts and their own expertise to the task. The benefits of sharing the effort are so great that they hardly need to be underlined. But the difficulties must not be underestimated either. While many astronomers and programmers have been and are contributing very generously, and benefiting in return from the enterprise so far, others are, and always will be, tempted to go their own way. *In the short run* it is always easier to write programs which you alone can understand. Taking the trouble to find what others have done and interpreting their programs is time-consuming and irritating. Documenting and sharing one's own programs may *in the short term* seem a thankless task, especially if other people, with no corresponding effort, can immediately make use of them.

However, the disadvantages of separate development should be glaringly evident to any active astronomer. The contemporary observationalist should be able to make use of a wide variety of observing instruments, from radio-telescopes, to spectrographs, to satellites. There is no way he can hope, on his own, or even with several colleagues, to develop software at his own institution to make the best use of all the opportunities. There *is no alternative* to sharing software development and sharing requires alertness, openness, generosity and a strong spirit of compromise. Therefore: (1) Before an astronomer undertakes a major piece of software development it is in his own interest to find out, via the STARLINK system, what is already available or can be adapted to his purposes. (2) If he finds, perhaps at some foreign observatory, software which might be of general benefit to STARLINK users he should alert the STARLINK management. (3) Finally. if he does undertake some major program development himself, he should advise the STARLINK team. They may, if the program looks to be of sufficient general interest, help and collaborate with him in producing and documenting his program. If the program is subsequently widely used, arrangements can be made, through the management, to see that the authorship of the software is properly acknowledged. It may even be that the author feels he has a right to co-authorship in papers where his work forms a major part of the whole effort. It is a task of the STARLINK project staff, and of the SIGs, to try to solve problems like this, as they arise. We feel that in general cooperation is flourishing, and will continue to, so long as most users are getting more out of the system than they put in. With 400 users this is not too difficult to ensure so long as everyone gives at least a modest push to the common wheel.

8 THE FUTURE OF STARLINK

This will be examined in the short, intermediate and long term.

In the short term, we foresee the hardening of the software environment and the gradual proliferation of astronomer-invented application packages. More off-site astronomers will be linked in from their own terminals, and STARLINK software will run on more non-STARLINK VAXs. We have made it a policy that all STARLINK software will be available to outsiders on negotiation and it is likely that line-links will gradually build up with ESA, ESO, ASTRONET in Italy, and later on farther afield. At the same time we can expect the logical up-grading of the present nodes to deal with bottle-necks as they arise.

Several developments can be seen as likely in the intermediate term. Line speeds may gradually increase to the point where data networking becomes slightly more feasible. A database will be introduced and will expand. Interfaces must at some level be built to the 32-bit Perkin–Elmer instrument computers at the LPO. We may introduce TV cameras at each node which can feed in some pictorial data directly via A to D converters. The foreseen development of further software-compatible mini-VAX machines may lead to the purchase of additional and cheaper nodes.

In the longer term, it is worth examining the likely implications and opportunities for astronomy offered by the continuing information revolution. In particular, cheap high-bandwidth ($\geqslant 1$ Mbit s^{-1}) links via satellites and optical fibres may lead to profound changes in the way we all work.

Fig. 1 shows a schematic model of the relationship between the astronomer and the main resources of his working environment. The way astronomy is done in any epoch depends very much on the speed and convenience of the various channels between one resource and another. Consider for instance the impact of the cheap plane ticket over the past 20 years. The backyard telescope and the staff astronomer have given way to the remote National Facility and the guest observer. In the process, astronomers have become problem- rather than technique-oriented. The research team based around

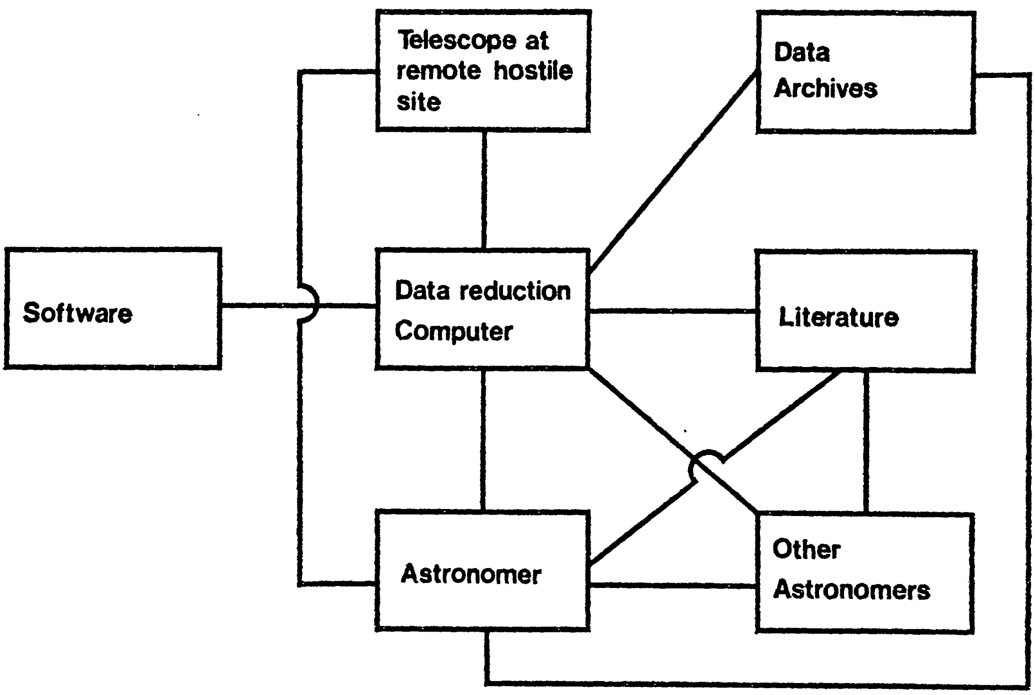

Fig. 1

community of locations is giving way to the international team based on community of interest. The autocratic leader is being replaced by the suave diplomat, the triennial IAU gathering by more specific international colloquia.

Without being too visionary one can see high bandwidth links ushering in some quite profound changes but we make no guesses at the time scales or costs involved. For instance:

(1) The hostility, remoteness and complexity of telescopes, be they on mountain tops or in the sky, will lead to the same sort of remote operation we presently employ on satellites. Rather than travelling, the astronomers will tend to sit before their consoles at home, controlling the observations by keyboard (e.g. Robinson, Schechter & James 1982).

(2) Face to face astronomer–astronomer interactions will be augmented and partially superseded by console-to-console communication. But the voice and alphanumeric channels alone are far too clumsy to make this very attractive at present. Dramatic improvements will come about when we can introduce analogue TV cameras into the link. Then equations, drawings, papers, data and pictures will pour freely and informally back and forth. It will be as if all linked astronomers live in the same electronic corridor. Electronic interactions, electronic collaborations, even electronic seminars will become a possibility. Most of the advantages of a large institute will then spread to a wider community.

(3) Data in different degrees of preparedness, all the way from raw data to the finished scientific paper will find their way up and down the corridor. Scientific preprints may whistle round the community penetrating to potentially valuable and unforeseen recipients. And readers' comments may attach themselves to these circulating preprints providing an almost instant peer-group review. What price refereed journals then?

(4) Not only will software and data be more easily shared but we can enhance and speed up the availability of various theoretical models and tools. For instance, one can imagine a synchrotron radiation package with a simple interface for the simple observer. The crucial confrontation between theory and observation will then be easier to make though not necessarily more sound.

What have all these imaginative developments got to do with STARLINK? Looking back to Fig. 1 it is interesting to note how many of these evolving or accelerating channels will switch through the data reduction resource. STARLINK is therefore a natural start for the evolution of an electronic corridor within our own community. Being seeded in the Rutherford–Appleton Laboratory, it is provided with strong potential roots. The Laboratory is already very active in satellite communication links, computer networking and automatic control. If there is a funded demand from the astronomical community RAL can no doubt provide for us the same sort of facilities it already provides for particle physicists. Their experimental project 'Universe' already runs through the laboratory connecting CERN and other big European establishments via 1 Mbit s^{-1} satellite links. There is no reason for astronomers to lag behind, and there is indeed every incentive and opportunity for UK astronomers to take a lead in this exciting and unavoidable race.

ACKNOWLEDGMENTS

STARLINK is the brainchild of many, but in particular of Igor Aleksander, Jim Alty, Alec Boksenberg, Peter Casey, Bob Dickens, Peter Davey, Richard Ellis, Bob Fosbury, Ken Hartley, Ed Kibblewhite, Mike Lawden and Cliff Pavelin.

REFERENCES

Wells, D.C., Greisen, E.W. & Harten, R.H. 1981. *Astr. Astrophys. Suppl. Ser.*, 44, 363–770.
Robinson, W., Schechter, P. & James, C., 1982. National New Technology Telescope, *Progr. Rep. No. 2*, Kitt Peak National Observatory.