THE DEVELOPMENT OF A COMPUTER DISPLAY SYSTEM FOR ARCHAEOLOGICAL PROSPECTING

by

Alister D. H. Bartlett
Summary
This thesis describes the development of computer programs for use by the Ancient Monuments Laboratory of the Department of the Environment in the display and interpretation of geophysical data from surveys of archaeological sites.

The relative significance of archaeological and other contributions to the instrument response and the characteristics by which archaeological features are recognized in the results are discussed with examples from magnetic and resistivity surveys. Significant anomalies may in part be defined in terms of their lateral extent and so may be emphasised or extracted through two-dimensional spatial filtering techniques as employed in image processing. Filtering procedures have been investigated and applied, and examples of results are described. It is proposed that filtering is usually necessary for resistivity readings which have a wide spatial response, but is of less relevance for magnetic surveys using the fluxgate gradiometer. This detects only at close range so that the output is effectively pre-filtered, and is the standard instrument at the AM Laboratory. The main requirement for processing magnetic results is simply a clear display based on a sufficiently close sample of the signal to allow detailed manual interpretation.

Additional numerical operations are required during processing to interpolate to scale and to define the range of display levels, and their effect on image quality is discussed.

Two sets of programs have been written, one of which is implemented on a time-sharing system and the other on a minicomputer. Each is interactive so that the treatment may be refined through tests on sections of data, and they have both been used for production processing. Design work now continues with the aim of setting up an extended system with capacity to process very large surveys efficiently. Proposed solutions to problems of program organization, control, and data structure are described.
## Introduction

### Chapter 1  Survey methods and interpretation of results

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Magnetic surveying</td>
<td>6</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Magnetometers</td>
<td>6</td>
</tr>
<tr>
<td>1.1.2</td>
<td>Recording equipment</td>
<td>8</td>
</tr>
<tr>
<td>1.1.3</td>
<td>Magnetic anomalies</td>
<td>10</td>
</tr>
<tr>
<td>1.1.4</td>
<td>Sources of noise and error</td>
<td>16</td>
</tr>
<tr>
<td>1.2</td>
<td>Resistivity surveying</td>
<td>27</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Equipment and technique</td>
<td>27</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Interpretation</td>
<td>30</td>
</tr>
<tr>
<td>1.2.3</td>
<td>Sources of noise and error</td>
<td>32</td>
</tr>
<tr>
<td>1.3</td>
<td>Other detection methods</td>
<td>33</td>
</tr>
<tr>
<td>1.4</td>
<td>Survey processing</td>
<td>35</td>
</tr>
<tr>
<td>1.4.1</td>
<td>Geophysical interpretation methods</td>
<td>35</td>
</tr>
<tr>
<td>1.4.2</td>
<td>Presentation of survey results</td>
<td>37</td>
</tr>
<tr>
<td>1.4.3</td>
<td>Computing requirements</td>
<td>40</td>
</tr>
</tbody>
</table>

### Chapter 2  Image processing - theory and application

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>42</td>
</tr>
<tr>
<td>2.2</td>
<td>Filtering</td>
<td>44</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Spatial filtering</td>
<td>48</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Frequency domain filtering</td>
<td>65</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Filtering: related techniques</td>
<td>68</td>
</tr>
<tr>
<td>2.3</td>
<td>Sampling</td>
<td>72</td>
</tr>
<tr>
<td>2.4</td>
<td>Edge matching</td>
<td>74</td>
</tr>
<tr>
<td>2.5</td>
<td>Interpolation</td>
<td>77</td>
</tr>
<tr>
<td>2.6</td>
<td>Statistics and plotting range</td>
<td>84</td>
</tr>
<tr>
<td>2.7</td>
<td>Display</td>
<td>92</td>
</tr>
<tr>
<td>2.7.1</td>
<td>Numerical output</td>
<td>92</td>
</tr>
<tr>
<td>2.7.2</td>
<td>Pictorial display</td>
<td>93</td>
</tr>
</tbody>
</table>

### Chapter 3  Computer implementation of processing system

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>100</td>
</tr>
<tr>
<td>3.2</td>
<td>Honeywell programs</td>
<td>104</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>3.4 Design proposals for a high-capacity production system</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>Chapter 4 Case studies: examples of the display and interpretation of survey results</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>4.1 Wharram Percy</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>4.2 Kenchester</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>4.3 Tarraby</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>129</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>Appendix: List of computed surveys</td>
<td>135</td>
<td></td>
</tr>
</tbody>
</table>
Geophysical prospecting is the only non-destructive and reproducible method of detailed site investigation available to archaeology. It is a technique which extends and complements other sources of fieldwork evidence but which has been relatively little exploited.

As a means of large-scale reconnaissance geophysics cannot offer the economy or extended coverage of aerial photography, but in detailed resolution the results are often superior. The condition of cropmarks and soilmarks varies with weather and season so that photographic information should ideally be integrated over a period of years (Hampton et al, 1977), but geophysical evidence may in principal be recovered through a single survey.

Fieldwalking is perhaps the method of site location most widely used by archaeologists, and the surface finds may provide cultural and chronological information not otherwise available to fieldwork. By plotting debris and artefacts on a measured grid it may be possible to locate features and activities within the site in general terms (Foard, 1978), but geophysics allows detailed and precise planning.

Fieldwork studies based on information from all such sources together with topographical and historical evidence are of increasing concern in archaeology. In part this development represents a move towards research into the history of the complete landscape rather than of individual sites in isolation, but such work is also a necessary aid to conservation. Few archaeological sites other than those specifically preserved will outlast this century in the face of current wholesale destruction and sites may only be defended or protective measures enforced where archaeology is known to exist. The only satisfactory means of conservation is to leave the archaeology undisturbed.
destructive. To dig on a scale corresponding to the needs of conservation would be quite impossible, and both information and objects are far more secure underground than they would be in the care of overextended archaeologists and museums.

Geophysical detection then offers potential contributions in many areas of archaeology, but so far few archaeologists are in a position to make use of it. For the moment it provides in many cases the most immediate and comprehensive means of establishing and recording the character and extent of a site. It is a major function of the Geophysics Section of the Ancient Monuments Laboratory at the Department of the Environment to supply such information to the Inspectorate of Ancient Monuments for administrative purposes in enforcing the Ancient Monuments Acts. Where necessary surveys are also provided as a guide to intended excavations, and sometimes as a means to complete the plan of a partially excavated site.

Suitable geophysical methods for archaeological use are by now technically well established, but a need remains for further practical developments to allow their efficient application on a scale which begins to meet the archaeological demand. The detection methods used (principally magnetic and electrical resistivity surveying) are well known in other areas of geophysical prospecting, but the vastly different scale of operation in archaeology requires a distinctive approach both in survey procedure and the interpretation of results. Developments in these areas are necessarily a concern of the Ancient Monuments Laboratory which is responsible for much practical archaeological surveying.

The object of the work described here was to investigate the application of computer techniques to the processing and display of survey data,
The survey data originates either as a grid of discrete values or as an analogue signal plotted along regularly spaced traverses. The readings are subject to noise and extraneous disturbances of various kinds and the final interpretation of features of archaeological interest is a matter of subjective assessment. A clear graphical presentation of the data is required as a basis for interpretation, and to prepare this manually is highly laborious. Much of the present work of the AM Laboratory is based on directly recorded analogue plots but numerical processing and computer display could offer greater power and flexibility. The archaeological features detected in general lie close beneath the ground surface and the data represents approximately the spatial variation in two dimensions of the quantity measured. Treatment of the data is therefore primarily a matter of image processing in which techniques for the enhancement of degraded images may be applied.

The exact treatment required to give a satisfactory visual effect varies for each survey, and so the processing parameters are best determined empirically by means of test plots produced through an interactive package of computer programs.

Two sets of programs have so far been completed and are described in this thesis. One operates on the Honeywell/General Electric Mark III timesharing system through a terminal at the AM Laboratory, and the other on the Data General Nova 800 minicomputer at the University of Surrey Physics Department. They represent two experimental approaches to the problem and so vary in structure and capabilities. They have both been used for some time in routine processing work, notably of
the use of analogue plots for the bulk of the laboratory output. Experience has by now shown however that replacement of the analogue plots with computed plots is possible and desirable. A further expanded system of programs which should allow computer treatment of the complete laboratory workload is therefore now in preparation.

The results of this work are presented as follows:

In chapter 1 survey procedures are outlined, and the factors, archaeological and otherwise, which affect the data are described. Examples of results are given to show methods of interpretation and presentation. The extent to which computer techniques are of value is then considered. In chapter 2 relevant image processing theory and its application to the problem are described. Other numerical operations carried out on the data are also discussed. The computer programs written for this project are described in chapter 3, and the requirements and proposals for the final extended processing system are explained. Some case studies of the results obtained from particular surveys with notes on the processing methods used are given in chapter 4, and finally surveys from which results have been processed using the programs described are listed in an appendix.
Chapter 1

Survey methods and interpretation of results

Techniques of archaeological geophysics are reviewed in the books by Aitken (1974) and Tite (1972), and the equipment and procedures used at the Ancient Monuments Laboratory have been the subject of papers by Clark and Hadden Reece (1972) and Clark (1975). Accounts of computer methods of data treatment by Scollar (eg 1969b, 1974, 1973), and also by Linington (1968, 1969) are referred to in more detail in chapters 2 and 3.

In this chapter fieldwork techniques are outlined to indicate the nature of the physical response on which the survey results depend and the consequent sources of both the archaeological 'signal' and the non-archaeological 'noise' which must be distinguished in the final interpretation. The methods of magnetic and resistivity surveying employed at the AM Laboratory are described and other techniques mentioned briefly.

Interpretation of survey plots, whether manually prepared or computed, is by no means an exact science and in some cases assessments by different interpreters may vary. Any report giving a full account of the results of a survey must therefore include representations of both the geophysical findings and the archaeological interpretation derived from them, together with an estimate of the reliability or significance of the results. Examples will be given of both initial and interpreted plots to show the nature of the problem.

Archaeological survey interpretation based on the 2-dimensional spatial variation in the data differs fundamentally from interpretation in large scale geophysical prospecting where the aim is often to reconstruct the source of the detected anomaly in 3 dimensions. The possible relevance of such an approach in archaeology is considered in section 1.4.1.
The practical requirements for a computer system to be employed in archaeological survey processing are mentioned briefly in section 1.4.3.

1.1 Magnetic Surveying

The measurement of local variations in magnetic field strength has been used as an archaeological prospecting technique since the introduction of a transistorised proton magnetometer (Aitken et al 1958), and is the most widely applied of the available methods. Magnetic detection is particularly effective in archaeology because the chemical and physical processes which give rise to detectable local magnetic anomalies are to some extent promoted by human agency and activity.

1.1.1 Magnetometers

The magnetometer preferred for routine surveys at the AM Laboratory is the fluxgate gradiometer. The archaeological use of an instrument of this type was reported by Alldred (1964), and the current standard model is that described by Philpot (1972). In this instrument each of the fluxgate detectors contains a pair of parallel high permeability (Permalloy) metal cores. These are wound with drive coils in series opposition, and the complete unit is surrounded by a sense coil. An applied drive current saturates the core thus excluding the magnetic flux of the ambient field, but produces an effectively zero resultant field on the sense winding. A pulsed DC drive current therefore causes the ambient flux to be 'gated' intermittently through the core, and in the process it cuts and induces a detectable voltage in the sense coil.

An individual detector measures the field component parallel to its axis and so is highly direction sensitive, but this is overcome by mounting a pair of detectors coaxially one at each end of a one metre long tube. The final output then represents the local gradient of the
This arrangement has the archaeological advantage that the instrument is most sensitive to features immediately beneath the detector tube (when carried upright). There is a weaker horizontal response and the range of detection is limited because with increasing distance the field strength at the two detectors tends to the same value. This means that it is possible to work to within a few metres of a metal fence and the instrument is not affected by diurnal changes in the earth’s field, nor by such interference as DC electric railways.

The integration time of the output is short enough (less than 1/5 second) for the signal to be regarded as a continuous and instantaneous measure of the field gradient. This continuous output governs the display and interpretation methods used and affects the strategies adopted in the computer work to be described below.

The proton magnetometer also remains in archaeological use. The precession frequency of protons (i.e. hydrogen nuclei in a sample of water or organic liquid) after they have been partially aligned by application of a polarizing field is proportional to the external field strength, and may be measured through the voltage induced in a detector coil (Aitken, 1974, ch 7). A single detector may be used as an absolute instrument to measure the total field strength, but the output is then subject to external field variations. These may be excluded by using a pair of detectors in differential mode with one kept at a base station, which is the preferred method of the Bonn Landesmuseum (e.g. Scollar 1974), or by mounting two detector bottles on one staff in a gradiometer configuration. In all cases discrete readings must be taken because of the several seconds required for the typical polarization and relaxation cycle. The practical sensitivity of both a single proton and fluxgate magnetometer is about 1 gamma (where $10^5$ gamma = 1 oersted in the emu system and 1 gamma = $(10/4\pi)$ milliamp/metre in SI units).
High sensitivity optically pumped magnetometers have also been used in archaeology, notably by the University Museum, Philadelphia. These are continuous reading absolute instruments and so are used differentially (Ralph et al, 1968). The instrument may be sensitive to variations of as little as 0.005 gamma but practical archaeological noise levels far exceed this.

1.1.2 Recording Equipment

The continuous signal of the fluxgate magnetometer allows the instrument output to be plotted directly on a chart recorder, and this is the method used in the standard AM Laboratory field recording system. Regularly spaced traverses are plotted using a string and pulley driven potentiometer to supply a distance signal to the X-axis of the recorder (Clark and Haddon-Reece, 1972). In exceptional cases this gives plots which may be presented as clear archaeological evidence with no further interpretation (eg fig 1.1), although more usually significant features are selected and outlined for clarity. For computing the signal is sampled at 1 m intervals and logged in digital form on a magnetic cassette for later transcription to punched tape. A new recording system shortly to be commissioned will use an optical encoder for accurate sampling. The readings are recorded as 8 bit binary integers, and may (at present) be set so that the maximum value of 255 corresponds to a reading of either 100 or 300 gamma.

On the few occasions when proton magnetometer data has been processed the readings were simply written down in the field and later typed to punched tape. Scollar (1974) uses a logging system carried in a van and under minicomputer control.
Fig 1.1 Magnetometer survey plot of Iron Age enclosure with plan of excavated houses at Groundwell Farm, Swindon.
The magnetic properties of soils have been extensively studied, initially by Le Borgne who first observed the generally increased magnetic susceptibility of topsoil relative to subsoil and proposed mechanisms to account for it. His results are discussed by Scollar, (1966) and Graham (1976), and the subject is reviewed at length in the thesis by Mullins (1974).

Soils typically contain a few percent of dispersed iron oxides, notably haematite \((\alpha \text{Fe}_2 \text{O}_3)\) which is weakly ferrimagnetic and occurs in single domain grains, but also goethite \((\alpha \text{FeOOH})\), magnetite \((\text{Fe}_3 \text{O}_4)\), and maghaemite \((\gamma \text{Fe}_2 \text{O}_3)\). Haematite may itself be produced by mild heating and then be converted through reduction to magnetite and then by reoxidation to the multidomain maghaemite, which is about 1000 times more strongly ferrimagnetic. This process has been reproduced experimentally by heating soil samples in hydrogen and then air (Tite and Mullins 1971), or in nitrogen with flour as an organic reducing agent and then air (Graham 1976). The susceptibility change after heating is sufficient to account for observed topsoil values in terms of the accumulated effects of occasional fires over historical timescales.

A fermentation process in which the reduction takes place during anaerobic decay in wet conditions and the reoxidation in dry conditions may also contribute to susceptibility enhancement. This process is more difficult to simulate than oxide conversion by heating (Aitken 1974, p 225), but rubbish pits with rich organic fill are often found to give strong magnetic anomalies.
A magnetic moment (of strength proportional to the susceptibility) is induced progressively in the soil as the magnetic moments rotate or realign parallel to the applied field direction. Detection therefore depends largely on susceptibility contrast between soil layers or features. Any variation in the relative depth of layers of different susceptibility, for example the presence of a silted ditch or pit cut into the subsoil, causes a local change in magnetic field strength which is in principle detectable. Contrast of this kind occurs generally but the effect is often much increased on past occupation sites where both burning and decay (in refuse pits or middens) are likely to have been concentrated. In some cases strong magnetic anomalies are detected close to a settlement, but the response diminishes away from the site as the susceptibility falls off to its natural value.

Measurements of the fractional conversion of iron oxides in soil samples have confirmed the frequent association of archaeology and enhanced magnetic susceptibility. The ratio of the initial susceptibility to the value after full conversion of the oxides to magnetite by heating in reducing conditions indicates the proportion of available oxides which are present in converted or magnetically enhanced form, and the ratio is usually found to be higher in archaeological samples than others (Tite and Mullins, 1971; Tite 1972).

There is an additional process which contributes to the observed magnetic anomaly of a burnt structure. Any such feature which survives intact will in addition to the induced field retain a strong and stable thermoremanent magnetisation (TRM). Any iron oxides present are, as mentioned previously, liable to be converted on heating to haematite and then magnetite. On cooling from the Curie point (675°C for haematite, 580°C for magnetite) a fractional excess of domains will
Fig 1.2  Magnetic survey of Romano-British kiln site at Mancetter, Warwickshire.

1:500

Vertical scale 40 gamma/cm
align with the applied field. Each grain has a distinct blocking temperature below which the direction of magnetization remains fixed so that the bulk material acquires a stable remanence proportional to and in the direction of the external field (Barbetti, 1976). Remanence is of major significance in rock magnetism and a detailed treatment is given by Stacey and Banerjee (1974). For archaeological surveying purposes it is simply the cause of strong and well-defined anomalies which are usually easily recognized, as in the plot from Mancetter (fig 1.2) where the kiln anomalies have a strength of 150γ compared to 10–40γ for other features. Hearth, the remains of metal working, etc are likely to give a similarly strong response.

It has been suggested (e.g., by Graham 1976, and Mullins, 1974 ch 4) that susceptibility readings from different soil types and locations should be used as a guide to the likely success of a magnetic survey. It is our experience that local variations are sufficient to make general prediction hazardous and that few sites are completely unresponsive. Results depend more on Le Borgne contrast (between topsoil and subsoil) than on the overall value and so a range of samples from topsoil, subsoil and fill are needed, and they are rarely available from unexcavated sites. The response from sites having any given value of topsoil susceptibility may vary widely: Ditches at Mucking, Essex (on gravel) or at Dean Bottom, Marlborough (chalk; AM Laboratory report G28/77) where the susceptibility was in each case $14 \times 10^{-6}$ emu/gm* failed to

* Quoted figures are AC bridge readings (Scollar 1968). They represent approximately half the total value of the susceptibility if the long-term viscous component of the magnetization is taken into account, but are useful for comparative purposes. The total susceptibility may be measured using, eg, a translation magnetometer (Graham, 1976). Readings given here are in emu where $1 \text{emu/gm} = 4\pi \cdot 10^{-3} \text{SI/kg}$. 
respond, but at other sites with the same value (Lake Farm, Wimborne, AM Laboratory report G11/76; Happpower, Dorset, report G16/76), on gravel and limestone respectively the results were satisfactory.

The topsoil susceptibility at Groundwell Farm (fig 1.1) was $29 \times 10^{-6}$ emu/gm, and the contrast with the chalk subsoil (less than $5 \times 10^{-6}$) was very strong. By comparison, at Raginnis, Cornwall, on granite, the topsoil value was higher ($34 \times 10^{-6}$), but there was only a very fragmentary response to ditches known to exist from cropmarks (see Fig 1.3).

It may be possible to exclude a few sites on the basis of low susceptibility values (eg Hambledon Hill, Dorset, topsoil reading less than $5 \times 10^{-6}$ emu/gm, was unresponsive), but with no great confidence. Even on the weakest sites there may be detectable thermoremanent or other strongly magnetic features (Saxon graves and swords were found at Pewsey, Wiltshire on a site with soil conditions similar to those of Hambledon; the plot is reproduced in Clark, 1975). The best way to evaluate a site is usually with a rapid magnetometer scan on arrival; with resistivity in reserve it is rare that a survey fails to produce at least valid negative evidence.

Measurements of variation in soil susceptibility across a site may provide evidence of the presence of occupied areas where strong enhancement is found, and this may be a useful prospecting technique, particularly where features are too eroded to give magnetic anomalies (Clark, 1977). The readings may be taken using either pulsed induction meter (PIM) or soil conductivity meter (SCM) type electromagnetic instruments, which respond to magnetic viscosity or susceptibility respectively (Mullins 1974), or by collecting gridded soil samples.
Fig 1.3 Raginnis, Cornwall - magnetic plot compared with cropmarks and plan of stone wall around field.
1.1.4 Sources of noise and error in magnetic surveying

Using the terminology of Scollar (1970a and b) all the non-archaeological contributions to the observed survey response may be regarded as noise of one of two types: There is uncorrelated noise which varies randomly for each reading and correlated noise which systematically affects more than one reading. Sources of uncorrelated noise include the following:

(i) Instrument noise and soil noise

These are difficult to distinguish and together give an irregular background to the plotted traces with displacements of (typically) about $\pm 2\gamma$. Much of the instrument noise can be damped with a series resistor in the input to the chart recorder, but any sudden jolt or knock to the fluxgate magnetometer will cause a noticeable deflection. Soil irregularities may cause rather larger randomly placed anomalies, and these are well seen in fig 1.4 (Moody's Down long barrow, Hants) which represents a survey in a ploughed field with a relatively high noise level of about $\pm 5\gamma$.

The soil noise sets a lower limit to the survey resolution. To be recognised in the interpretation pits must either be large enough to affect more than one traverse (more than 1 metre across), or contain a sufficiently magnetic fill to give an anomaly significantly greater than the noise level. The lower size limit in this case is about $\frac{1}{2}$ metre. Features of strength of a similar order to the noise are only recognisable through continuity, and this was the criterion used to fix the slightly idealised outline of the long barrow side ditches shown in fig 1.4.
Fig 1.4 Moody's Down, Barton Stacey, Hants. Survey to locate side ditches of long barrow. Plot showing magnetic interpretation compared with location of resistivity anomalies. 1:500

Magnetic scale 40 gamma/cm
Soil noise causes most difficulty in interpretation for relatively weak sites (Moody’s Down is on chalk). For the great majority of sites most of the significant anomalies are well above the noise level. In rare extreme cases a strongly magnetic topsoil (eg Coupar Angus, AM Laboratory report G5/75; topsoil susceptibility $21 \times 10^{-5}$ emu/gm, subsoil $25 \times 10^{-6}$ emu/gm) will cause surface noise which is seriously obtrusive even in the presence of quite strong anomalies.

(ii) Recording Errors

The locational accuracy of the survey depends on various systematic and random factors which include; precision in measuring the site grid to field boundaries or national grid (we are rarely allowed to leave the field pegged out for relocation); the magnetometer carrier who must hold the instrument in the correct position relative to the potentiometer string and follow a straight line; adjustment of the zero position of the chart recorder pen, etc. Errors are difficult to estimate but the combined effect is probably well within the $\frac{1}{2}$ metre resolution of the system (or 1 m resolution of logged computer data).

One error of similar magnitude is visible in the traces; a (supposed) linear feature is sometimes laterally offset in alternate traverses. This is probably a combined fault of time delay in the magnetometer and sluggishness due to overdamping in the chart recorder (or play in the drive). The effect is visible in fig 1.5 which shows the initial traces next to a partially smoothed interpretation.

Much larger lateral errors are liable to be introduced into computer data by failures in the logging system. Each missed reading causes a possible error of 1 m, and only an arbitrary correction is possible (see notes on INPUT programme, section 3.2). The new field recording system (section 1.1.2, above) should improve matters.
Fig 1.5
Leckford, Hants. Magnetic plot of ditch (note lateral offset of anomalies in alternate traverses).
Fig 1.6
(iii) **Interference**

The most common cause of random small scale interference is the presence of scattered nails, horseshoes and other small iron objects. One advantage of the fluxgate plots compared with coarsely sampled discrete data is that most such objects give characteristic 'iron spikes' immediately recognisable by their shape and by the presence of a distinct negative deflection. Only rusty or more deeply buried iron is likely to resemble an isolated archaeological feature. In a very bad case surface iron may obscure the archaeology but it need not be confused with it; see fig 1.6. This survey at Reawla, Guinear, Cornwall was on waste ground behind some houses, but the main outlines of an Iron Age enclosure system are visible through the anomalies caused by old prams and builder's rubbish.

On certain soils, particularly boulder clay and some gravels, erratic igneous rocks may cause anomalies not unlike iron spikes. The rocks vary in size and depth so that some give distinct spikes but other anomalies are weaker and broader. This is a form of soil noise which differs from that mentioned in (i) above in being the result of variation in composition rather than distribution of the soil. In such conditions the identification of isolated archaeological features is again a very uncertain process. Fig 17 (Milfield, Northumberland) shows a survey in which archaeological anomalies are not clearly distinguishable from those due to igneous stones in the gravel subsoil.

Some of the location errors listed in (ii), above, act at random but others affect complete traverses and so satisfy Scollar's criteria for correlated noise. These errors can only be minimised by accurate fieldwork (except that alternately offset traverses may be smoothed), but there are other sources of correlated noise which are visible in the plots and which may be compensated for through interpretation or
Fig. 1.7
Hilfield, Northumberland
Survey to locate ditches of possible Anglo-Saxon settlement. Plot shows anomalies caused by small volcanic rocks dispersed in gravel subsoil.
(i) **Systematic Instrument Error**

Two faults may affect the zero setting of the fluxgate and cause location errors on the chart: First the output falls off and the traverses are compressed together as the batteries run flat; and secondly alternate traverses are displaced vertically to give wide and narrow gaps in turn. This second effect occurs either when the person carrying the magnetometer is magnetic or when the fluxgate detectors are out of alignment. Each of these problems is best cured in the field but can be corrected if necessary by tracing traverses or anomalies on to a correctly spaced grid.

(ii) **Background changes and geology**

These are of little significance in gradiometer surveys. Any change in field strength originating more than a few metres from the instrument affects both detectors similarly and is not recorded. This excludes all variations in the earth’s field; diurnal changes, micropulsations and magnetic storms, and also any effects from underlying geology. Occasionally small geomorphological features of a magnitude intermediate between soil noise and geology and detected. Two possible cases are shown in fig 1.8. At Thorpe, Egham (plot (i)). There are broad undulating anomalies which are plainly not archaeological and perhaps represent variations in the depth to the gravel; and at Levington, Ipswich (ii) in similar conditions there is a strange apparently circular feature. The general character of this is archaeological but it fails to fit any likely category. It is too irregular for a barrow ditch, too large and isolated for a hut circle, too small for an enclosure, and it produced only clean sand when augered to a depth of 1 m. The evidence seems to indicate that it
Magnetic surveys on gravel showing anomalies of possible geomorphological origin.

1:400
Magnetic scale 30 gamma/cm
the anomaly is particularly strong and distinct for a natural feature.

There is occasional but less frequent response to small-scale natural features on soils other than gravels; for example, pockets of clay-with-flints may be detected on chalk.

(iii) Interference

There is no distinction in principle from the cases mentioned in the previous section, but I use this term to refer to effects of a different order of magnitude from the archaeology. The main sources of interference of a correlated kind are buried pipes and igneous geology. A small cast iron pipe will give an anomaly of perhaps 100\gamma, but a large one, for example a high pressure gas pipeline, causes an anomaly of several thousand gamma and deflects readings up to 20 m away to either side.

Not all igneous geology interferes with surveys and some granite sites have responded well. Geologically recent volcanic rocks are however capable of producing magnetic gradients steep enough to affect the readings. Fig 1.9 shows a survey on a headland above high shale cliffs cut by vertical intrusive dykes, and was plotted in the hope of locating a ditch (enclosing Marisco Castle on Lundy Island) despite the very strong geological anomalies. There are some weaker ditch-like features to the N of the survey but they are wrongly orientated to be part of the fortifications and probably represent more deeply buried igneous intrusions.

The practical requirements for a survey processing system are determined largely by the methods adopted both to compensate for the sources of error listed here, and to extract archaeologically significant results. The methods used for magnetic and resistivity data and their consequences for the design of a processing system are considered in section 1.4.
Fig 1.9 Lundy Island: magnetic survey showing anomalies caused by igneous dykes.
The measurement of variations in soil resistivity was first used as a method of archaeological detection in 1946 (Atkinson 1963), and the technique remains a necessary complement to the magnetometer. Current conduction through the soil is by electrolysis and the resistivity measurements effectively represent the water content. Resistivity is thus a useful method for detecting such features as wall footings and roads which retain little moisture but are not necessarily magnetic, and also for ditches and pits that happen to give a poor magnetic contrast, or in conditions too disturbed for magnetic surveying. The necessity of inserting probes for each reading makes resistivity surveying much slower than magnetic work and so it is usually reserved for specific tasks where magnetic surveying is unsuitable.

1.2.1 Equipment and Technique

A feature of most practical surveying instruments is an AC current source to avoid time dependent polarization effects at the electrodes. The Martin-Clark meter commonly used in archaeology operates at a frequency of 67 Hz. Separate pairs of current and potential probes are used, as with other instruments, to reduce contact resistance effects. This is further minimised by taking readings at a null balance point obtained by backing off the output of the potential probes through a transformer against a potentiometer through which the ground current passes. The ground/current is zero when the reading is taken.

There are various standard geophysical configurations of current (C) and potential (P) probes (see eg Telford et al, 1976, pp 655-661). The two which are used routinely in archaeology are the Werner configuration (CPPC), which gives a double peak because detection occurs between each current-potential pair, and the double dipole (CGPP) which effectively combines the two peaks in the centre. The
for probe spacing $a$ and resistivity $\rho$ in a homogenous medium the measured resistance $R$ is given by $R = \rho / 2\pi a$ for Wenner and $R = \rho / 6\pi a$ for dd. A useful compromise is therefore to take Wenner and dd readings at each position along a traverse so that the presence of a single dd peak between a double Wenner peak gives a repeated check on the presence of an anomaly. The probes are connected to a rotary switch on the meter with a fifth spare position in the cycle so that each probe in turn can be moved at the same time as a reading is taken with the other four to save time. This method is generally used for preliminary investigation by making long traverses at intervals across a site.

For detailed planning of an area the twin electrode configuration of Aspinall and Lynam (1970) is preferred. The two halves of the Wenner array are effectively separated with one OP pair fixed at a distance at least 30 times the probe spacing from the other. Readings taken with the second pair represent local features with only a negligible contribution from any subsurface variation in the ground between the two pairs. The depth of penetration is rather greater than for Wenner and dd at the same probe spacing. The sensitivity is also less than for Wenner, but usually sufficient in practice. The advantages are relative speed of working and simplicity of interpretation. Each anomaly gives a single peak and readings taken on a regular grid may be computed and displayed as for a magnetic survey. Readings are usually taken at 1m separation to give resolution similar to logged magnetic data. To cover a given area still takes some six times longer than a magnetic survey and resistivity is used on a correspondingly small scale. This might change if ever resistivity is successfully mechanised.
Fig 1.10 Lundy Island: Plot of resistivity traverses located on plan of magnetic survey squares.
Vernier and double dipole readings from a given traverse are usually plotted on the same axes for comparison and then interpreted qualitatively in a similar manner to magnetic traces. In fig 1.10 a series of traverses are superimposed on the plan of the magnetic squares shown in fig 1.9. The soil was shallow and stony and the results are very confused, but the only possible continuation of the line of the extant rampart and ditch appears to be through the anomalies at the E of traverses V, VI and VII. If so the ditch lies further to the E than expected and the absence of any likely ditch from the magnetic squares is confirmed. Traverse XII was surveyed in case the ditch curved to the SE but extreme high readings from near-surface rocks obscure any response. Readings were taken with probe spacing 1m, but with additional 2m readings along traverses V and VI. The depth of penetration is of the same order as the probe spacing, but any broad or general change in resistance within that depth is recorded (eg traverse X) and is not reduced to a constant base level as magnetic measurements would be if taken with a gradiometer.

There is an example of an isolated high resistance feature with the dd peak centered between two Wenner peaks at 25m in traverse VII. Here it probably represents a rock or small outcrop but the effect is useful for identifying walls on less disturbed sites.

The ditch, if it has been detected at all, appears in this survey as a series of high readings or positive anomalies. This is not always the case and a ditch with silted fill often gives a negative anomaly. Other ditches, depending on fill, soil conditions and weather, change seasonally from positive to negative (Clark 1975).

The Lundy results show the effects of geological noise in an extreme case, but the criteria for selecting possible archaeological features
continuity between the traverses and the archaeological plausibility of the plan. Continuity is more evident in closely spaced data and so twin electrode readings on a regular grid are taken wherever time allows. The numbers are written down (but could quite simply be logged), and it is here that there is no realistic alternative to computer display.

Fig 1.11 shows a first example of results:

![Diagram showing survey results](image)

Fig 1.11 Saltby, Leicester: Resistivity survey of round barrow showing ditch between concentric stone kerbs. (Data supplied by Leicester museum.)

The conventional means of presentation of a computed survey is as a plot in which the density of the display varies according to the strength of the reading. Either dot-density plots (fig 2.1), or (as here) symbol plots in which characters are overprinted on the terminal to give the required density range may be used. These and other display techniques are discussed further in section 2.7.2 below.

In the survey shown here (Saltby, Leicester) the mound of the round barrow and the surrounding ditch both give negative anomalies (blank on plot). The positive anomalies represent a stone kerb between the mound and ditch, and another outside the ditch. The excavation showed that the inner kerb was cut by a robber trench at the top centre of the plot.
1.2.3 Sources of noise and error in resistivity surveys

If it is assumed that contact resistance and polarization effects are minimised in the design of the meter, the following sources of operational error remain:

(i) Soil noise and geology

These are listed jointly because variations of any spatial extent may occur from random changes in single readings to overall drift across the complete survey. Resistivity results are less subject to individual wild values than magnetic data, but as the Lundy plots showed, extreme readings are not impossible in dry conditions.

(ii) Experimental accuracy

The main source of error here (if possible meter reading and recording errors are neglected) is in positioning the probes. The reading varies in direct proportion to the probe spacing and errors of a few cm are significant. The effect should be random and represents an increase in general background noise not distinguishable from small scale soil noise.

(iii) Changes in base level

Apart from any overall geological trend there may be a shift in the level of readings if there is rain during the course of a survey, especially if the ground was previously very dry. There is also a change if the remote probes are moved part way through a twin electrode survey.

(iv) Ground currents

These may cause meter instability but the operating frequency is chosen to minimise their effect.
1.3 Other Detection Methods

A variety of geophysical methods which have been experimented with in archaeology are listed by Aitken (1974) and Tite (1972). Some have achieved a measure of success for particular problems, but others are unsuitable because the scale of the parameter measured is not appropriate to archaeology (e.g., gravity, seismic, soil radioactivity methods), or offer no general advantage over established techniques.

Electromagnetic instruments suffer from limited penetration because the combined attenuation of transmitted and return signal means that sensitivity falls off as \((\text{depth})^{-6}\), compared with \((\text{depth})^{-3}\) for an (absolute) magnetometer. For a low frequency instrument (e.g., the SCM at 4 KHz) response varies with the induced magnetic moment in phase with the transmitter voltage, but at high frequencies (greater than 100 KHz) the out of phase eddy current response dominates and provides a measure of resistivity (Tite 1972, p36). Commercial instruments of this kind are available, and results have also been obtained using long wave radio stations as the source of the transmitted signal (Tabbagh, 1972). Any release from resistivity probe pushing would be welcome, but whether these alternatives offer equivalent resolution or penetration is uncertain.

The in-phase magnetic response of the SCM is proportional to magnetic susceptibility which may indicate the presence of occupation sites (as mentioned in section 1.1.3 above), and the instrument is also capable of locating variations in topsoil thickness and the presence of pits etc to a depth of about \(\frac{1}{2}\) metre. Chemical tests of soil phosphate concentration (Schwarc 1967) are also useful in prospecting, especially in conjunction with susceptibility measurements. Comparative results from the two methods may be of greater value in locating occupied areas than either used independently (Clark 1977).
An alternative to resistivity detection is the induced polarization system of Aspinall and Lynam (1970). This may give improved response to some features but would be no less laborious to use.

All of these techniques have in common that they rely on the identification of local variations in spatially distributed readings. None of them (except for experiments in susceptibility and phosphate sampling) are used in routine work at the AM Laboratory, but if necessary they could be used and processed in an identical manner to magnetic and resistivity data.

There is another category of technique in which multiple reflections are recorded at each position and which could be used for sounding in depth rather than for two-dimensional planning. Ground penetrating radar (Bevan and Kenyon, 1976), and a recent acoustic device which uses a computer to reconstruct cross sections at different depths from an array of reflected waveforms (Ozawa and Matsuda, 1979) are of this type. They may have applications for specialised studies but not for the extensive surveys which are the subject of this project.
1.4 Survey Processing

1.4.1 Geophysical Interpretation Methods

It has already been stated that the interpretation of archaeological surveys is a task essentially different from that of reconstructing geological features in three dimensions, as may be done in large-scale geophysical prospecting. The simplifying assumptions that are required in calculating such results are rarely applicable in archaeology where features may be assumed to lie close to the surface but are much more varied and irregular in form, and where far more detailed interpretation is required.

The standard geophysical techniques for magnetic surveys described in Telford et al (1976) are much concerned with correction for large scale terrain effects. These are a minor problem in archaeology where the variations over the small areas covered are usually slight. There is therefore no need for such techniques as analytical continuation in which anomalies are projected from an irregular surface on to an arbitrary plane (Telford p 147-150).

The most commonly applied method of constructing a geological interpretation from magnetic data is to match the observed anomalies with those calculated for various idealized geometrical forms. The shapes used in calculations include two-dimensional bodies (i.e. prisms of infinite length), planes and steps. They are variously simulated by lines of dipoles or single poles, or results are derived by summation of elements.

This approach has found some use in archaeology and Linnington (1972) has published a lengthy description of the method with results for the calculated anomalies for various bodies. He compares his results for a triangular ditch with those from a simulation programme described
by Scollar (1969) and with the model studies of Aitken and Alldred (1964), finding reasonable agreement.

Usually for detailed calculation more information is required about the detected anomalies than is available in a magnetic survey; for example the susceptibility and depth of the features and the relative contributions of induced and remanent magnetization. Such calculations are a necessary aid to understanding the source and nature of anomalies, but are not directly useful in routine survey interpretation.

In practice satisfactory results are obtained through qualitative interpretation based on the plan, distribution and strength of the anomalies.

The physical nature of the magnetic response is sufficiently taken account of in the rules given by Aitken (1974):

(1) The maximum anomaly is displaced south of the source by a distance approximately \( \frac{1}{3} \) the depth of the feature.

(2) The depth or width of the feature (whichever is greater) equals the width at half the maximum value.

(3) The reverse anomaly is 10% of the maximum anomaly and lies north of the feature at a distance equal to the depth.

The observed anomaly is the vector sum of the earth's field at that point with the field due to the feature. The displacement to the south and the reverse anomaly therefore depend on the angle of dip and vary with latitude.

In plotting the anomalies the distortions caused by (2) and (3) may be corrected by appropriate choice of the range of values to be included, and if necessary a slight shift in the origin of the plot.
will correct for (1). For many shallow features the displacement may be disregarded and treated as negligible in comparison with other surveying and recording errors. The anomalies may then be read as a direct representation of the plan of the features causing them.

Similar considerations apply to resistivity surveys. In large scale geophysics resistivity is often used for investigation in depth at a single point and much of the interpretation theory relates to this. In archaeology the spatial variation of the readings is almost always the main concern and, except for the double peaks of the Wenner array, this is directly apparent in a plot of the data.

1.4.2 Presentation of Survey Results

Any evaluation of the final archaeological significance of a survey must usually be based upon not only the measured readings but also the nature and context of the site and its known or possible archaeological character. At some stage of the overall process a distinction must therefore be drawn between the presentation of the measured data and the derived interpretation. Geophysical surveying is an extension to the techniques of field archaeology and often the problems which are presented to the Geophysics Section can only be answered by proposing an archaeological hypothesis. This may sometimes be tested by excavation, but in other cases (where to draw the boundary to the scheduled area?) action must be taken without any possibility of further verification. To preserve the distinction between data and hypothesis direct presentations of results are, where appropriate, included alongside interpreted plans in the reports.

The object in preparing the data, whether by computer methods or not, is therefore to produce a clear and comprehensive display for interpretation. It is not necessary to extract the archaeological features
response prior to interpretation, although it may be useful to do so, by computing or otherwise, in a final published plan. This approach differs from that evident in some of the published work of Scollar (eg 1970a and b), where so far as possible archaeological characteristics are comprehensively predefined and the features then extracted automatically.

The emphasis on display in survey processing follows from the use of analogue plots as the principal medium of recording and interpreting AM Laboratory magnetic surveys. These plots allow detailed interpretation of isolated or continuous features, and recognition of continuous features against a noisy background. The task perhaps compares in some respects with that of interpreting aerial photographs; experience in recognizing faint or diffuse features is required in either case. Errors may be allowed for or corrected, for example by ignoring iron spikes or doubtful geological features, smoothing staggered traces etc. The difficulties are first that the method is inflexible, the scale and sensitivity of the plot must be preset during fieldwork and cannot then be modified without re-surveying the site. Secondly and more importantly, production of a publishable plan from the initial sections of field chart involves a long sequence of laborious operations. First the traces are glued in place on a scale grid on a light table with correction if necessary for incorrect traverse spacing. All blemishes in the lines are whited out and the section photocopied. The copies are assembled into a complete plan of the survey and margins and grid points redrawn. This plan is sent for photographic reduction to a manageable size for final copying (usually to 1:500 from 1:200) and meanwhile an interpretation is worked out on the initial charts. This is either redrawn on the final reduced plot or made the basis of a separate plan. A redrawn plan (eg fig 1.12(ii)) is not a substitute
Fig 1.12 Barton Stacey, Hants. Magnetic survey of Iron Age settlement.
for the traces (fig 1.12(i)) because much detailed evidence is lost. It was for example possible to deduce the sequence of construction of the trackway (parallel ditches) and enclosure by close examination of the traces at the intersection. In other cases an area of increased soil noise in the chart may indicate the presence of occupation, and this would be lost in any less detailed display.

The survey charts therefore have virtues of detail and resolution which should be preserved in any computer based processing system.

1.4.3 Computing Requirements

The processing methods described so far allow archaeological information to be extracted with reasonable thoroughness, but slowly. The time taken by chart preparation is disproportionate to that required for working out the interpretation or drafting the report. The available time between surveys is very limited and so it is difficult to maintain the survey programme required without a large and enduring backlog of reports.

The most substantial benefits of computing therefore follow simply from its use as a drafting method to speed up the preparation of plots for later interpretation. For resistivity data similar considerations apply even more strongly because no initial field plot is available.

This may be the primary aim, but a computer is far more than a drafting tool and once facilities are available there is every reason to process the data further to try to eliminate as much as possible of the manual work. The aim as stated is not necessarily to isolate all the archaeology, although this may occasionally be achieved, but rather to arrive at an interpretation representing a synthesis of the available evidence. To this end the use of computer techniques to
clarify, enhance or extract particular types of feature, and the comparison of results from several such procedures may all be relevant.
2.1 Introduction

In a comprehensive survey display system facilities are required to manipulate and display any stated class of feature or section of the survey, and these operations may be performed using the techniques of digital image processing.

The survey readings may be regarded as the elements of a digitized image; they are scalar quantities which vary as a function of position and are significant in terms of their spatial variation and their relative rather than absolute values. The features to be sought or enhanced may be specified in terms of their spatial extent or numerical magnitude. The final interpretation of the display is qualitative and is to be judged by visual criteria.

The nature of the interpretation process differs essentially from that in geological prospecting where quantitative 3-dimensional results may be required. There are however techniques used generally in geophysics in common with image processing, notably those used to extract regional trends or residual features from gridded data, and it is here that methods most directly applicable to archaeological surveys are found.

A distinct body of image processing theory and techniques has developed recently as demands (especially in space and medical imaging) and computing capabilities have converged. Techniques exist to control the contrast or signal to noise ratio, to restore or reconstruct degraded or blurred images, as well as to modify variously defined features within the image. The methods in part have a formal basis
largely derived by extending the linear systems theory used for
1-dimensional time-varying signal analysis into two dimensions, and
are partly ad hoc empirical procedures applicable to particular tasks.

The subject has been reviewed by Hunt (1975) and a series of literature
surveys published by Rosenfeld (eg 1973). Texts include treatments
by Andrews (1970) and Gonzalez and Wintz (1977). The methods and
literature of image processing are distinct from those of computer
graphics, in which the image is first generated by the computer, and
from pattern recognition, although this subject is closely related.
Pattern recognition methods have been applied to the prediction of
uranium deposits from prospecting data (Briggs and Press, 1977), but
would not be remotely justified for the small volumes of data available
in archaeology.

The fundamental operation in image analysis or enhancement is the
separation of spatial frequency components through filtering. This may
be done indirectly on the initial data by convolution with arrays of
filter coefficients having the required transfer characteristics, or
directly by multiplication of components of the frequency spectrum
obtained by taking the Fourier transform of the data.

In practice filtering is only one of a series of numerical operations
required to prepare and display the data. The operations are described
below in sequence (except that sampling, which occurs during field
data collection, is discussed after filtering). Examples of results
from the two sets of programs currently operational are given where
appropriate, and modified or extended facilities to be incorporated in
the forthcoming production system are noted. The overall design and
operation of each set of programs is discussed in chapter 5.
2.2 Filtering

A digital (or optical) imaging system may be assumed to behave as a linear system, so that the total response is equal to a linear combination of more elementary responses (Andrews 1970, Goodman 1968), and a superposition property applies. This means that the response to a given input may be found by decomposing the input into a number of elementary inputs, calculating the response to each, and superimposing the responses to give the total output.

One possible decomposition of the input is to express it as a sum of displaced Dirac delta functions (ie, \( \int_{-\infty}^{\infty} \delta(x, y) \, dx \, dy = 1; \delta(x, y) = 0 \) for \( x, y \neq 0 \)).

The delta function has the fundamental property that:

\[
\begin{align*}
  f(x, y) &= \int_{-\infty}^{\infty} f(p, q) \delta(x-p, y-q) \, dp \, dq \\
  \text{(for continuous } f) 
\end{align*}
\]

and this provides the decomposition directly as a linear combination of displaced and weighted delta functions.

The system may be regarded as a mapping of a set of input functions into a set of output functions and represented by an operator \( S \), so that for input \( f \) and output \( g \),

\[
  g(x, y) = S \left\{ f(x, y) \right\} 
\]

(2)

Substituting for \( f \) from equation (1) then gives:

\[
  g(x, y) = S \left\{ \int_{-\infty}^{\infty} f(p, q) \delta(x-p, y-q) \, dp \, dq \right\} 
\]

(3)

If the function \( f(p, q) \) is regarded simply as a weighting factor the superposition property of a linear system may be applied to give:

\[
  g(x, y) = \int_{-\infty}^{\infty} f(p, q) S \left\{ \delta(x-p, y-q) \right\} \, dp \, dq 
\]

(4)

where \( S \left\{ \delta(x-p, y-q) \right\} \) is called the impulse response or point spread function of the system. This may be rewritten as a function \( h \) of the position variables;

\[
  \text{i.e., } h(x, y; p, q) = S \left\{ \delta(x-p, y-q) \right\} 
\]

(5)
In the case of a space-invariant image system the output image moves but is not otherwise modified as an input point source changes position. This means the impulse response of the system is a function of \( x-p, y-q \) only. Therefore,

\[
h(x, y; p, q) = h(x-p, y-q), \tag{6}
\]

and the superposition integral (4) becomes:

\[
g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(p, q) h(x-p, y-q) \, dp \, dq, \tag{7}
\]

which is a 2-dimensional convolution of the object function with the impulse response of the system, and may be written:

\[
g(x, y) = f(x, y) * h(x, y).
\]

Convolution is commutative, therefore:

\[
g(x, y) = h(x, y) * f(x, y) \tag{9}
\]

and eqn (7) may be rearranged:

\[
g(x, y) = \int_{-\infty}^{\infty} h(p, q) f(x-p, y-q) \, dp \, dq \tag{10}
\]

For the practical case in which the input function is represented by regularly spaced discrete readings the integral is replaced by a sum:

\[
g(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} h(m, n) f(x-m, y-n). \tag{11}
\]

The impulse response function is here represented by a matrix of coefficients \( h(m, n) \). (eg Zurflueh 1967, Gonzalez and Wints 1977).

A survey plot contains features of varying lateral extent, or spatial wavelength. The wavelength characteristics may be modified by digital convolution using equation (11) to give a set of output data containing those wavelength from the input data which fall within a range specified by the impulse response function, (or point-spread function) represented by the array of filter coefficients \( h(m, n) \). This is the process of spatial or convolution filtering.

There is no simple or direct method of generating a coefficient array which corresponds to any given impulse response, and so in practice the function must either be approximated or derived indirectly.
The indirect derivation makes use of the alternative decomposition of the input function (or data) into elementary periodic functions which is provided by the Fourier transform:

For any arbitrary (one dimensioned) function \( f(x) \) the Fourier transform \( F(u) \) is defined by the equation:

\[
F(u) = \int_{-\infty}^{\infty} f(x) \exp(-j2\pi ux) \, dx, \quad (12)
\]

and the inverse Fourier transform by:

\[
f(x) = \int_{-\infty}^{\infty} F(u) \exp(j2\pi ux) \, du. \quad (13)
\]

\( F(u) \) represents the frequency content of \( f(x) \) explicitly. It may be expanded into real and imaginary parts:

\[
F(u) = \int_{-\infty}^{\infty} f(x) \cos 2\pi ux \, dx - j \int_{-\infty}^{\infty} f(x) \sin 2\pi ux \, dx, \quad (14)
\]

showing that it is composed of an infinite sum of periodic terms and each value of \( u \) determines the frequency of that sine - cosine pair.

The Fourier transform represents the limit of a discrete Fourier series expansion of a periodic function obtained when the frequency interval between the terms of the series tends to zero and the fundamental frequency of the function tends to infinity (so that it is no longer periodic) (Stearns, 1975). The transform retains the property of the Fourier series that the component at each frequency, or value of \( u \), is a least squares fit to the function. The initial function may be reconstructed from its frequency components using eqn (13) where the term \( F(u) \) acts as an infinite set of coefficients of the complex waveforms given by \( \exp(j2\pi ux) \).

The Fourier transform may be extended to a function of 2 variables:

\[
F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp(-j2\pi(ux + vy)) \, dx \, dy \quad (15)
\]

and then modified for the practical case of discrete sampled data:

\[
F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \exp \left (-j2\pi \left( \frac{ux}{M} + \frac{vy}{N} \right) \right) \quad (16)
\]

for \( u = 0, 1 \ldots M-1, \ v = 0, 1 \ldots N-1 \).
The terms $F(u, v)$ of the transform are complex and so must be displayed in the form of the amplitude spectrum given by the modulus of the equivalent vector at each point. If eqn (16) is separated in the same manner as equation (14) into a real part $R(u, v)$ and an imaginary part $I(u, v)$, then the amplitude spectrum is given by

$$|F(u, v)| = \left[ R^2(u, v) + I^2(u, v) \right]^{1/2} \quad (17)$$

The transform is displayed with long wavelengths (low frequencies, small $u, v$) plotted at the centre of the frequency $(u, v)$ plane, and short wavelengths at the edges. The edge occurs where the wavelength equals the sampling interval and is called the Nyquist limit (Scollar 1970b). The transform is in fact periodic with period $N$ (for square data $N \times N$), but is completely specified by a single period.

Once an array of data has been transformed it may be filtered by direct modification of the terms of the frequency spectrum. A frequency domain filter takes the form of an array of coefficients, each of which modifies one element of the spectrum. Coefficients between 0 and 1 will either attenuate or transmit the corresponding frequency components when multiplied. The values in the coefficient array represent the transfer function of the filter, $H(u, v)$. Given a set of transformed data $F$ the effect of a filter is given by

$$G(u, v) = H(u, v) \cdot F(u, v) \quad (18)$$

and the final filtered data may be recovered by taking the inverse transform of the output function $G$.

The convolution theorem however states that the Fourier transform of the convolution of 2 functions equals the product of the Fourier transforms of the 2 functions (eg Papoulis 1962). If therefore $g, h, f$ are the inverse transforms of $G, H, F$ the operation

$$g(x, y) = h(x, y) * f(x, y) \quad (19)$$
in the space domain is equivalent to eqn (18) in the frequency domain. Eqn (18) is identical to equation (9) where \( h(x, y) \) represents the impulse response of the convolution filter. It is therefore possible to obtain a set of coefficients to represent a convolution filter having specified impulse response characteristics by taking the inverse transform of a suitable frequency domain transfer function.

Either of the two eqns (18) and (19) may be used as the basis of a practical filtering system, and the choice of multiplication in the frequency domain or convolution in the space domain is one to be made on practical as well as theoretical grounds. Frequency domain methods have found wide application in geophysics and elsewhere and are of particular direct relevance to the analysis of time-varying seismic data. Their application to archaeological survey data is described in the extensive work of Scollar (var), but convolution techniques remain in use both in archaeology and in other image processing applications. Filtering of the Mariner 9 Mars pictures in 1971 was limited to convolution, usually in one dimension along the lines of the image (Open University, 1978), and frequency domain techniques are not necessarily required in medical tomographic imaging (Jackson et al, 1978). The two methods and their archaeological applications are described in turn below:

2.2.1 Spatial Filtering

The practical realization of various types of high and low pass convolution filters (ie, to pass high and low spatial frequencies), and their application to geophysical survey data is described by Dean (1958) and by Zurflueh (1957). Zurflueh gives an example of a filter coefficient array with a cut-off wavelength of 8 grid units derived by taking the inverse transform of the equivalent transfer function. The ideal
inverse transform is infinite and so a practical compromise giving an
approximation to the required impulse response and cut-off frequency
with a minimum number of coefficients is found by trial and error.
A reasonable correspondence between the Fourier transform of the impulse
response and the required transfer function is obtained for a filter
array of width $1\frac{1}{2}$ times the cut-off wavelength. Execution of the filter
for each output value is simply a matter of summing the neighbours of
the corresponding input value, each multiplied by the appropriate
coefficient as in eqn (11). Filters of different cut-off wavelength
are obtainable by interpolating the 8-unit filter coefficients to a
larger or smaller array.

A high-pass filter results if the low-pass filtered output is
subtracted from the original data, and this may be done directly by
inverting the sign of the coefficients and including a central positive
spike value of magnitude equal to the sum of the other coefficients.
If the low-pass coefficients are normalized to a total magnitude of
1 (−1 when inverted) the central spike is +1.

An approximation to the process of multiplying by a normalized
coefficient array which is simple and economical in execution is
obtained by taking the mean of specified neighbouring values, usually
those forming a ring of stated radius around each reading in turn.
In the case of a high pass filter the mean is then subtracted from the
current central value (or it may be added in the case of a low-pass
filter).

\[ g(x, y) = f(x, y) - \frac{1}{N} \sum f(n, m) \]

where \( f(n, m) \) is summed over \( N \) neighbours as defined in the program
parameters. Such filters have a precise cut-off frequency but the
impulse response is distorted by 'ringing' as described in the next
section (Filtering Subroutines), below.
'Ring residual' filters of this kind have a long history of archaeological application (Scollar and Kruckeberg 1966, Linnington 1969-9), and are also used in the only commercial geophysical display package of which I have details (SIA, 1973). At the AM Laboratory they were incorporated in the program used for early computer bureau work undertaken by CI Data Ltd, and then included in later programs.

Spatial Filtering - Applications

The essential requirement in archaeological work is for high-pass filters capable of discriminating narrow linear features such as ditches or walls, or isolated features such as pits, from broad background trends. There is rarely any call for low-pass (or band-pass) filtering because the sampling interval of the survey is in general coarse compared with the size of detectable features. The currently used survey techniques and logging system are not capable of recording noise at a higher frequency than the significant data. Random 'uncorrelated' noise cannot then be separated in terms of frequency characteristics in sampled data (although it may be in the analogue plots), and can only be defined in terms of amplitude (see 2.6, Plotting Range).

The main object of filtering is therefore to exclude those sources of interference listed as correlated noise in magnetic data in section 1.1.4, and the geological and background changes in resistivity. There is in practice only a very limited need for this in processing magnetic surveys because the fluxgate magnetometer is very little affected by correlated noise. Neglecting instrument faults and gross magnetic interference which obscures the archaeology in any case, the only remaining source of correlated interference is from local geomorphology. The instrument has a narrow range of spatial sensitivity
so that the output is effectively pre-filtered to include only features with archaeologically possible dimensions. The natural variations in soil conditions which are detected are those of wavelength similar to the archaeology, and to distinguish them is the task of manual interpretation.

A secondary effect of high-pass filtering is to sharpen the image. The analogue magnetic charts are again in little need of this. The gradiometer signal falls off with distance faster than the inverse cube relationship of an absolute magnetometer, giving minimal spread in the detected anomaly. To sample the data at 1m intervals through the data logs is effectively to blur this signal. Subsequent filtering may partly compensate for the visual effect of the blurring, but not for the loss of information content. The practical effect of image sharpening is limited by the resolution of the display, which for some current plots is no better than that of the initial data.

Filtering is therefore a process to be applied to gradiometer surveys for the sake of marginal gains in clarity, and perhaps to exclude features at the broad end of the detectable wavelength spectrum, rather than for any fundamental reasons. Fig 2.1 shows part of an Iron Age enclosure containing several clusters of pits. Display is in the form of a dot-density plot in which each reading is represented by a group of randomly offset dots corresponding in number to the strength of the reading. An 8-element high pass ring filter of radius 3 (grid units) was used. There is random interference, mostly from iron spikes which are not easily recognised at this resolution, along a track above the enclosure, and the strong anomalies to the bottom right of the plot are caused by an iron pipe.
Fig 2.1 Guiting Power, Glos.

Dot-density plot of magnetic survey showing enclosure and pits.

Filter radius 3; range mean to mean + 2 standard deviations in 25 levels.
Fig 2.2 SHALFORD MANOR MAGNETOMETER SURVEY

(i) UNFILTERED PLOT
MEAN TO MEAN + 1.5 ST DEV

(ii) FILTER RADIUS 2
MEAN TO MEAN + 1.5 ST DEV

SCALE 1:200
Fig 2.2 shows how limited the effects of a filter may be for data with only narrow significant anomalies. Shalford manor is unusual in having iron slag incorporated in the buried foundations which are therefore detectable magnetically as well as by resistivity. The unfiltered data (i) and filtered data (ii) were plotted at different scales, but there is no other noticeable change in the anomalies.

Filtering is altogether more important in the case of resistivity surveys. Here there is no initial discrimination against long wavelengths and some overall drift, mostly from changes in soil depth, is usually present. The response compares more with that from an absolute or differential magnetometer than with a gradiometer. The differential use of two magnetometers with one at a fixed base station, as in the arrangement described by Scollar (eg 1970 C, 1974), excludes the background changes in total field strength which would be detected by a single absolute instrument, but anomalies of wavelength up to the size of the survey are admitted. Filtering is an essential stage in the processing of all such data, and computer treatment, correspondingly, of more fundamental importance.

The unfiltered data from the resistivity survey at Shalford is shown in fig 2.5. The high readings are initially all at one side of the plot, but after filtering (fig 2.4) the outline of a building, more complete than in the magnetic survey, is visible. A narrow filter (i) gives a sharp but rather broken outline and the building may be more easily recognised in the slightly blurred but more continuous plot from a wider filter (ii).

Fig 2.5 gives examples of dot-density plots which similarly show the resolution of local features through filtering. Here a filter of radius 4 was used to extract from the initial data anomalies representing large burning pits marking the positions of destroyed stones in the northern inner circle at Avebury.
RESISTIVITY SURVEY
UNFILTERED PLOT
MEAN TO +1.5 ST DEV

PLOT NO. 055
SCALE 1:200

AM LAB 02/06/78

Fig 2.3
Fig 2.4  Shalford Manor, Guildford  Resistivity survey  1:200

(i) FILTER RADIUS 1
MEAN TO MEAN + 1.5 ST DEV

(ii) FILTER RADIUS 3
MEAN TO MEAN + 1.5 ST DEV
(i) Unfiltered plot

(ii) Filter radius 4

Resistivity survey showing burning pits which indicate positions of destroyed stones.

Fig 2.5 Avebury
Filtering Subroutines

Two alternative implementations of symmetrical high pass (residual) filters were included in the Honeywell programs (see AM Laboratory report G 1/79). In one of them (SFILTER2) the mean value subtracted from the central readings represents either 4 readings at stated positions along the principal axes, or 8 readings with intermediate positions included, as shown in fig 2.6.

The output of a spatial filter is the convolution of its impulse response with the data. The impulse response of a 4-element filter (ie output from a zero array with central value = 1) is shown in fig 2.6(i). The output is a plan of the filter itself centered on the impulse, and this causes spurious negative anomalies in the region of any high readings in the data. The magnitude of the negatives may be reduced by including more elements in the filter, and so in the other subroutine (SFILTER1) all values which lie close to the stated radius from the central value are located and included in the mean. At the small radii most often used in practice (2-4 readings) the effect of this filter is not noticeably different from that of an 8 element filter, and only the 4 and 8 element options were included in the Surrey programs (Julien and Stannard 1976). As the impulse response (fig 2.6 (ii)) shows, 8 filter elements are not enough to suppress the negatives (although the magnitude of each is \(-\frac{1}{2}\) rather than \(-\frac{3}{2}\)), and 'ringing' (Gonzalez and Wints, 1977, p 145) still occurs. The practical result of ringing is clearly visible in the blank strip to each side of the enclosure ditch in fig 2.1, and also in fig 2.5 where the circular shape of the anomalies makes the effect of the convolution directly apparent.
(a) Plan of filter

(b) Impulse response

(i) 4 ELEMENT FILTER
MIN TO MAX IN 8 LEVELS

(ii) 8 ELEMENT FILTER
MIN TO 0.8 IN 8 LEVELS

Fig 2.6 Spatial filters with plots showing impulse response reading included in mean value.
Ringing occurs to some extent with all practical filters. It is possible to reduce the distortion of the impulse response by using extended and tapered coefficient sets (as in Zurflueh 1967), but at the cost of a less sharp cut-off frequency. There are equivalent difficulties in frequency domain filter design where a sharp cut-off frequency causes ringing in the image (after inverse transformation) similar to that seen here. Filter design is a compromise between sharpness of cut-off and impulse response (Scollar 1970 b), and the application of suitable tapering or 'apodizing' functions to define the edge of the filter transfer function is the subject of considerable literature (Gonzalez and Wintz 1977, Stearns 1975). This tapering is usually done in the frequency domain and convolution arrays, if required, are found indirectly by taking the inverse transform.

In practice ringing does not seriously impair the information content of the plots and may be considered an aid to clarity by improving the local contrast of features in dot-densities. A wider choice of filter arrays would however allow some control of the degree of ringing acceptable in a given plot, and means of implementing this are considered further below.

**Edge Effects**

In convolution with readings at the edge of the data the filter array will partially extend beyond the edge, and either the filter or data must therefore be modified. Possible remedies include surrounding the data with zeroes, extrapolating the data, or simply truncating the output which falls within half the width of the filter array from the edge of the input data (Black and Scollar, 1969). The problem does not arise in frequency domain filtering where the wave numbers are represented directly. The method used in SFILTER1 and SFILTER2 is to modify the filter by testing whether each location falls within the
data so that only values from a partial neighbourhood are included at the edges. The filter response is less well-defined at the edges than elsewhere but the effect is not noticeable in practice, and some loss of accuracy is preferable to a complete loss of data. Similar edge effects will occur between sections (unless they overlap) when a data file exceeds the core space and is read to the program in pieces. This happens with the existing Honeywell program for filtering large files (FILTER), but the virtual data structure of the new programs will avoid the problem (section 3.4).

Blank values within the data (identified by a preset marker) are also tested for in all the current programs and similarly excluded.

**Non-circular Filters**

Any arbitrarily defined set of elements may be averaged and subtracted in a convolution filter, and not just the circular groups described above. To subtract a non-circular group may be useful particularly when linear features of a given orientation are sought. The response is then strengthened by defining a linear array of filter elements directed at right angles to the expected features. The filter must still be symmetrical about the central value or the position of features will be displaced in the output, but the symmetry need not be circular. The effect of such filters has been investigated by Linnington (1970, 1971).

A variation on this principle is the baseline filter subroutine (BFILTER) included as a third option in the Honeywell program. This was intended to correct for displacement of fluxgate traverses caused by faulty detector alignment or operator interference (section 1.1.4) through fitting a local baseline to each traverse. A least-squares fit
(i) Initial data

(ii) Data - mode of each traverse

(iii) Data - running mean of 10 readings

Fig 2.7 Bullock Down, Eastbourne

One-dimensional filtering to correct for diurnal drift in proton magnetometer data.
to the data would tend to be displaced by anomalies, and the method used is to find the mode of the data. This is effective except where an anomaly parallel to the traverse occupies more than half the traverse. Because the values are held as real numbers the frequency of occurrence of each number cannot be counted directly, and instead a count is made of the number of readings falling within each of a stated number of steps (defined between the mean of the data ±2 standard deviations). The lower bound of the step with the highest count is taken as the undeflected base level for the traverse and subtracted from each of the readings. This is effectively a process of one-dimensional filtering to correct for horizontal discontinuities.

One application was to a set of proton magnetometer readings for which no base station reference readings for diurnal drift correction were available. A plot of the initial data (fig 2.7 (i)) shows strong drift during each of the two days of the survey. Subtracting the mode of each traverse removed the overall drift between traverses (ii) but there was still sufficient variation within each traverse to affect the plot. Each reading was therefore compared with a running mean of 10 neighbours along the same traverse to give the result in (iii). This is equivalent to 1-dimensional high-pass filtering with a filter wide enough to admit anomalies but not wide enough to be affected by drift.

Alternative Implementation of Spatial Filters

In the subroutine described above the coordinates of the filter elements are generated by the program according to fixed parameters, and only the radius of the group may be varied. A differently structured program is therefore required for each different type of filter configuration. It would be possible to achieve a simpler and more flexible realisation of the process by making the convolution operation
(equation 11, above) explicit in the program coding. An array of coefficients representing any type of data could then be entered and only one simple routine would be required to calculate its convolution with the data. Coefficient arrays to represent high-, low- or band-pass, 1- or 2-dimensional, symmetrical or asymmetrical filters could then be defined by other routines or independent programs and held in files to be read into the processing program as required. The arithmetic required to locate the filter elements is therefore removed from the processing routine which is thereby greatly simplified, and transferred to a quite separate filter creation programme.

It would be equally possible to use either large accurately tapered coefficient arrays, or small arbitrarily defined ones without modification to the program. The existing filters are easily expressed as arrays: a 4-element residual filter for example is obtained by summing the products of the coefficients across an array with a central value of 1, coefficients of $-\frac{1}{2}$ for the 4 elements, and zeroes elsewhere. The coefficients should be normalized so that their sum equals one for a regional (low pass) filter to preserve the total amplitude of the data, or equals zero (as here) for a residual filter. The normalization could be provided in the filter creation programs. Filter arrays could be modified using the same processing routines as any other data and scaled, interpolated or rotated as necessary. The new programmes are designed to process multiple arbitrarily dimensioned arrays simultaneously, and so execution with a coefficient array of any reasonable size will be possible.

The increased number of arithmetic operations required to calculate the convolution with the total array rather than stated non-zero elements as in the current routines will cause some increase in execution time, but this is at present negligible and an increase would
not be a great problem. Most practical filters would fit into arrays of 25-121 elements and the increase would be slight. Existing routines will in any case be retained as options. Scollar (1970 b) criticizes convolution filtering on the grounds of excessive execution time, and for large coefficient arrays the time does become significant. Convolution with an array the same size (N) as the data would require \( N^2 \) operations, which is of the order \( 10^8 \) for even a modest survey. The advantage of frequency domain techniques is that they allow filtering of similar precision with the number of operations proportional to only \( N \log_2 N \) (Cooley et al. 1968). For small convolution arrays however the balance of practical advantage probably lies with space domain methods. Experience has shown that useful results are possible with simple spatial filters, and archaeological data are probably not usually of a precision that would make accurate filter design important. If clearly defined filter characteristics are ever required in particular cases then computational advantage and simplicity of filter design would both favour frequency domain processing.

### 2.2.2 Frequency Domain Filtering

Once the frequency spectrum of the data has been calculated it can be modified directly by multiplication of the coefficients as in equ (18):

\[
G(u, v) = H(u, v) \cdot F(u, v)
\]

It is in principle much simpler to specify the transfer function \( H(u, v) \) than to generate an array of convolution coefficients. Standard fast Fourier transform routines operate by reordering the data to allow initial computation of a set of two point transforms. These are combined into higher order transforms through \( \log_2 N \) successive doublings until a single transform for that row of data is obtained. The algorithm was first published for computer application
by Cooley and Tukey (1965), and subroutine listings are given by Cooley et al (1969) and Stearns (1975).

The execution time is proportional to $N \log_2 N$ (for $N$ data points) rather than $N^2$ for a routine coded directly from equation (16) above. For practical data arrays $\log_2 N$ is likely to be 6-8, so that even after filtering and taking the inverse transform the number of operations compares with convolution for a small filter array and is much fewer than for convolution with a large array. A transform routine will be included in the new program, but certain practical complications are likely to limit its usefulness:

Most published FFT routines require the length of each side of the data array to be a power of two (although Black and Scollar, 1969, describe a scheme in which it need be only a multiple of 4), and implementation is simplified for a square array. The data could first be interpolated to any required dimensions using facilities to be provided in the programs in any case, but to divide the data into square sections will be troublesome for large surveys, as experience with the core-limited versions of the present programs has shown. A 2-dimensional transform is found by transforming first the rows and then the columns. Methods of transposing the rows and columns of the intermediate results when the number of items exceeds the core space are described in section 2.5 (Interpolation).

There will also be problems in processing incomplete blocks of data, as well as those that are irregularly shaped. Scollar (1970b) suggests filling in missing readings by interpolation. This is provided for in the existing programs, but the usual practice there is to test for blank marker values at each stage of processing and transfer them to the output so that blanks appear in the final plot. To achieve this with transformed data it may be necessary first to copy a map of the
blanks to file, interpolate and process, and then replace the blanks in the output. This should be possible in the new programs which are designed to allow arithmetic or logical operations between multiple input arrays. This data structure will also allow large filter arrays to be read from file alongside the data. Once this is done filtering is a simple matter of multiplying corresponding elements in the two arrays, but other programs to prepare the filter arrays are also required.

Standard functions for tapering the edge of the pass-band to reduce ringing are discussed by Scollar (1970b) and Gonzalez and Wints (1977). They include the normal or Gaussian curve which is completely free of ringing but gives a very blurred cut-off frequency, sines, decaying exponentials, filters with trapezoidal cross section, and Butterworth filters constructed from the transform of a first-order Bessel function. In frequency domain filtering the problem of edge effects does not arise, although spurious wavelengths may be introduced if there is a step at the edge of the data.

More elaborate filtering schemes are discussed by Scollar, and include the use of asymmetrical (e.g. elliptical) coefficient arrays as direction-sensitive filters, and optimum or Wiener filtering in which the normalized complex conjugate of the transform of expected anomaly is used to provide the filter coefficients. Because the correlation of two space domain functions is equivalent to multiplication of the transform of one by the complex conjugate of transform of the other:

$$f(x, y) * g(x, y) \Leftrightarrow F(u, v) * G^*(u, v)$$

(where * represents the complex conjugate), Wiener filtering effectively calculates the cross-correlation between the actual and assumed response.

Probably too much previous information.
about the detected features is required for this method to be of use in practice.

The Fourier transform is the most generally used image transform, but others based on decomposition of the data into various alternative elementary orthonormal function sets are available. Some (eg Haar and Walsh transforms which take values ±1) offer possible computational advantages. Both the transform algorithm and filtering techniques here are similar in principle to those of the Fourier transform (Gubbins et al, 1971, Gonzalez and Wints, 1977).

A further alternative to computing the Fourier transform would be to make use of optical filtering as described in Goodman (1968). The diffraction pattern formed in the focal plane of a lens is the Fourier transform of the object, and may be filtered directly, eg with apertures of different sizes to represent lowpass filters, and the results observed in the image plane. Scoller (1970b) suggests filtering transparencies representing sections of survey data in this way, but problems of experimental accuracy and the physical realization of filters are likely to limit the usefulness of the technique. Work of this kind with Landsat image data has been reported, however (Barnett and Harnett, 1975).

2.2.3 Filtering; Related Techniques

There are various other processes involving numerical operations on the data which are closely related to those required in filtering:

Smoothing

A standard technique for suppressing individual extreme values or reducing random background noise is to replace each reading with a weighted mean of neighbouring values.
This process in fact is simply a particular application of low-pass filtering by convolution and could be implemented directly using the general convolution routine proposed in section 2.2.1. It is not therefore a process generally applicable to magnetic data for which the sample interval is coarse compared with the noise frequencies. It could be relevant in cases where comparatively broad features are sought, perhaps for example the barrow ditches in fig 1.4, and has been used at times in plotting resistivity traverses to help distinguish broad or deeper-lying features from surface noise.

**Differentiation**

This provides a means of emphasising edges and discontinuities in the data and could provide an effective display for certain types of feature. The gradient at a given point is defined by Gonzalez and Wintz (1977) as the magnitude of the resultant vector with components given by the partial derivatives along the two axes. In practice the derivatives are approximated by differences between the neighbouring values. This is closely related to high-pass filtering but differs from a simple convolution in that either the mean square or the total absolute value of the differences is found, rather than the simple difference from the neighbours.

To test for gradients which exceed a stated threshold would probably be the most efficient means of extracting iron spikes from magnetic data, and is effectively the method used at present in visual interpretation of the survey charts. Again, more closely sampled data would be needed for a reliable computer implementation of such a test because most iron spikes are too narrow to be recognized clearly in the currently available sampled data.
The second derivative of the data is often required in quantitative
geophysical interpretation and may be approximately determined either
by convolution or frequency domain techniques. The second derivative
of potential field readings gives an indication of the form of the
buried structure and there is considerable literature on methods of
calculation and related techniques for finding upward and downward
continuations and the anomalies from arbitrary bodies (e.g. Bhattacharyya
1978, Hahn 1976). As noted in section 1.4.1 quantitative results of
this kind are of little direct relevance in most archaeological survey
interpretation.

De-spiking

There is no certain method of excluding the readings caused by scattered
pieces of iron and they form a major hazard in processing magnetic
data. The various possible alternatives to testing the gradient
(above) would again only be partially effective with a coarse sample
of the data.

One simple possibility for a survey lacking strong archaeological
anomalies would be to provide an option to set readings which exceed
a stated value to zero rather than to the maximum plotting level.
This would exclude strong iron spikes, but there is no reliable test
for those of magnitude within the range of the archaeological features.
Genuine pits may affect only a single reading and in a dot-density plot
pits and spikes look very much alike (compare the iron noise along the
track with the pits within the enclosure in fig 2.1). Low-pass
filtering would spread the effect of the spikes but also blur the
picture.

Given a closer sample (e.g. 0.5 metre) it would be possible to test either
the gradient with more reliability, or else to test for some empirical
combination of height and width characteristics, and substitute other
values where necessary before the main processing. The substituted value could simply be the mean of an arbitrary set of neighbours or, as suggested by Scollar (1970b), the values could be taken from a low order fitted surface or found by very low-pass filtering.

Most of the practical work so far has been with resistivity data where wild values do not usually occur.
The minimum sampling interval required for complete reconstruction of an image is found by considering the effect of the sampling interval on the Fourier transform. If the sampling interval is $\Delta x$ the transform of the sampled data is periodic and repeats at intervals $1/\Delta x$. Adjacent periods of the transform will therefore overlap unless $1/\Delta x$ exceeds twice the maximum frequency of the transform, or $\Delta x$ is less than half the minimum wavelength. This result is known as the Whittaker-Shannon sampling theorem (Gonzalez and Wintz, 1977).

Only one period of a Fourier transform is needed to reconstruct an image, and if the periods overlap then the transform in any interval is corrupted at the edges by adjoining periods so that high frequency information is lost and the image blurred.

It follows from the sampling theorem that to obtain the maximum information from a survey the initial field data collection should ideally be arranged so that readings are taken at intervals corresponding to half the width of the smallest detectable feature. The limit of resolution for the fluxgate magnetometer is about $\frac{1}{2}$ metre for a shallow feature in good conditions, which would mean a sample spacing of $\frac{1}{4}$ metre. This resolution is not achieved at all sites but there are few surveys in which features 1 metre wide are not visible in the traces. To produce a computer plot of quality equal to the traces for a site giving 1 metre resolution would therefore require readings recorded on a $\frac{1}{3}$ metre grid. In other site conditions the interval might be increased. Ralph et al (1968) describe the survey procedure used to locate remains of a Greek city buried under 6m of flood plain alluvium on the Plain of Sybaris in southern Italy. The calculated half width of typical anomalies (from fired clay tiles and walls) was 6m which would require sampling at 3m spacing. In fact readings were taken at 2m spacing on traverses
The fluxgate would be quite insensitive in such conditions but we have never in this country not archaeology deeper than the immediate subsoil (except occasionally under sand dunes), and much finer resolution is usually possible. The actual traverse spacing used is in practice more often a function of the size of the site and the number of traverses that can be walked in a week than of the anomaly width. Where possible 1m spacing is used but 2m and, more desperately, 5m have been tried. Interpretation of the charts is still possible but detail rapidly diminishes and at 5m only a general indication of the presence or absence of the archaeology is obtained. Such methods are therefore appropriate in cases where the initial archaeological problem is posed in such terms. More commonly we are asked to supply as much detail as possible, but at present with readings logged at 1m spacing the full potential detail is not available in the computer plots.

With the existing field equipment the primary record of the survey is provided by the paper chart recorder, with an auxiliary digital tape recorder which is limited by its speed of response to recording at 1m intervals. Once it is practical to complete all routine survey results other arrangements might be preferable. The immediate chart trace would not be essential so long as some form of on-site display took its place. Readings could instead be recorded in a solid-state memory which could be triggered as often as required, perhaps in the simplest case by markers on a string stretched across the traverse. Such a system would have no moving parts and so be completely non-magnetic. It could therefore be carried with the magnetometer, so abolishing the very troublesome trailing lead and greatly speeding the work. At intervals the reading could be read from the memory to a micro-processor in the
and transferred to magnetic tape for later computing.

In a system of this kind fieldwork and processing would be integrated so that fully detailed computer plots could be produced with no additional labour. To construct such equipment would be a logical continuation of the present project, but for the moment the computer plots are limited to a theoretical resolution of about 2 metres.

2.4 Edge Matching

The survey readings often originate in small sections which are not necessarily convenient for processing. After each break in the survey there is likely to be some variation in instrument setting or response which will cause a change in plot density unless corrected, and so a program is needed to combine sections of equalized data into a processing file.

The correction most often required is a simple additive shift in the level of the readings. The Flurgate has no fixed zero and is liable to drift slightly. The setting within a 30 metre survey square may be regarded as constant, but the level for the day will depend on the arbitrary initial adjustment. On the paper charts a change in the zero setting simply offsets the origin of the plot, and this is corrected when the pen is positioned, but the setting still affects the logged data.

Similar changes occur between sections of a resistivity survey. The cause may be a difference in soil moisture content due to overnight rain etc, but more often the break happens simply when the remote twin-electrode probes must be moved to keep within range of the survey. Variation in water content should cause a change in sensitivity of detection and require correction by proportional multiplication rather
different times this is difficult to confirm. The magnitude of anomalies must vary in the course of the seasonal disappearance or reversal effects described in Clark (1975), but this does not usually seem to be noticeable in the few days required for a survey.

In the Honeywell survey processing system a separate data pre-treatment program separate from the main processing program was provided to equalize and combine the sections of data. This program (COMBINE; See AM Laboratory Report G1/79, sections 1.4.1 & 2.4) accepts a list of up to 10 initial data files and transfers the readings to stated locations in a single output file, which may then in principle be processed complete without the need to run the programs repeatedly for the different sections.

The readings in the files may be equalized according to mean values calculated either from each complete file or from the readings in the row or column at the edge adjoining the next file. Either the difference in mean values is added to the file to be corrected, or the readings are multiplied by the ratio of the mean values.

If the mean values of the complete files (rather than edges) are to be equated, then the input files may be located in the output file by coordinates relative to an arbitrary origin to the bottom left of all the data. The program creates an output file to contain all the sections of input data and fills gaps with blank markers (999's). The input sections (files) may overlap and there are options to add, subtract or replace superimposed readings.

If the adjoining edges are to be equalized the files must be entered consecutively in row or column order, and the correction is made progressively between each succeeding pair of squares. This arrangement
is not very satisfactory because neighbouring rows cannot safely be assumed to be similar, but it was included in the program because to collect extra readings from duplicated edge traverses would complicate fieldwork procedures, especially for magnetic surveys.

The complete program is core independent. Files are input and output by records and not held in core, and so there is no practical limit on their size.

Only restricted use has so far been made of the program. Simple overall correction is probably sufficient for changes in instrument setting in magnetic surveys, but in resistivity results instrumental and geological effects are combined, and any overall shift is as likely to create as to correct any discontinuities. Practical limitations to the capacity of the processing programs have also prevented work with large surveys, and so there has been little actual need to set up combined working files. With the Honeywell programs it is possible to plot large surveys, but the system is too slow and expensive for large-scale practical use. The Surrey programs are severely core limited and so surveys must be plotted in small sections in any case.

These are major restrictions which the new production system is intended to overcome. To do so facilities equivalent to those of COMBINE will be required as part of a more flexible system of data management.

Rather than separating data equalization in a pre-treatment program it would be better included in the processing systems. Additive correction is useful in some cases but a more generally effective method of equalization is simply to filter the readings in each initial section. Alternatively subtraction of a least-squares fitted polynomial surface would have a similar effect. Scollar (1970b) suggests this as a form of pre-treatment for Fourier transform data
Subtraction of a surface is in any case a form of high pass filtering, and sometimes used as such in geophysical processing. The closer the fit of the surface, or the higher the polynomial order, then the narrower is the effect of the filter.

The facility for grid-to-grid arithmetic included in COMBINE also needs to be transferred in an extended form to the main processing system. Various processes require arithmetic or logical operations between multiple parallel arrays, and this is not allowed for in the data structure of the previous programs.

Equalization and edge matching are therefore processing options to be included in the facilities of a comprehensive program, but to do so it will be necessary to accept initial data from a series of external files rather than a single file. This complex input procedure was in the Honeywell system included in a program separate from the main processing routine to avoid excessive complexity. With the experience gained it should now be possible to integrate all the requirements in a single system.

2.5 Interpolation

Modification of the sample interval of the data by interpolation is an essential facility in a processing system and provides a means of controlling the scale and resolution of the plot. A scale plot using a display of fixed character size is only possible after initial interpolation to an array of suitable dimensions. For output on a teletype the data dimensions must be expanded to scale and be corrected in the ratio 5:3 to allow for the proportions of the typeface, and interpolation will also be required for any device with a fixed increment size, such as the Tektronix VDU. The pen plotter currently
used at Surrey has continuous scale adjustment, but only over a limited range. A simple routine to extend the range by interpolation of intermediate values so that the data dimensions are doubled is included in the programs, but a more comprehensive interpolation facility would be preferable and would allow control of the dot size and (to some extent) the resolution independently of the scale of the plot.

The interpolation routines in the Honeywell system provide a possible model for the production programs. The examples in fig 2.8 show something of the effect of interpolation on image quality. For visual clarity plots should be large compared with the elements of the display, and the coarseness of a teletype means that plots as large as possible are required. The fragmented anomalies at the top of the LH square in 2.8 (i) were therefore replotted at twice the scale as seen in 2.8 (ii). The filter radius was increased from 3 to 4 to emphasize continuity, and the saturation plotting level reduced from 2 to 1.5 standard deviations to strengthen the anomalies, but the quite convincing stepped outline of a building which results would not have been resolved without the change of scale.

The output values are calculated by direct linear interpolation between readings. This is simple in principle and the final calculation is done in a single program statement, but to incorporate the procedure in a workable program requires some elaboration. All the subroutines involved may be called either from the core - limited interactive system or, as for each of the other processing options, from a separate core-independent program, in this case INTERP, which allows files of unlimited size to be processed. (See AM Laboratory Report G1/79; sections 1.5.2, 1.6.2, for operating instructions and 2.5.4, 2.6.2 for program notes).
(i) Complete survey; filter radius 3
range mean to mean + 2 standard deviations

(ii) Section from top of LH square interpolated to 2 x scale.
Filter radius 4, range mean to mean + 1.5 standard deviations

Fig 2.8 Fountains Abbey; resistivity survey
done by the subroutine SCALES on the basis of the character size, separation of original readings and the stated plot scale. If the final dimensions exceed the core space in the interactive version there is an option either to respecify the scale or to take a smaller section of the initial data. Because of the length of this subroutine it is held in a separate overlay file from the processing routines.

In the processing overlay the final interpolation subroutines are called from a control routine STRETCH. This extracts readings by rows or columns as necessary from the main 1-dimensional data array and transfers them to a processing buffer. The interpolation subroutine is called and then the output is redistributed to the output section of the main array. A different procedure is needed in the core independent version (below).

The final interpolation occurs in one of two alternative routines which differ in their treatment of blank values:

INTER1 interpolates readings across blanks (marked usually by 999's) from the nearest non-blank neighbours. These are located by a search procedure before interpolation and the distances from the output position to the input readings adjusted as necessary. If the blanks occur at the end of the row previous non-blank values are extrapolated. This subroutine may be used to fill gaps in the data even if there is no change in the dimensions.

In INTER2 the blanks are retained in the output. Each output position whose nearest neighbour in the input is blank is set to blank and the remaining values interpolated.

The additional complication when the subroutines are called from the core-independent routine INTERP is that intermediate results must be
It could be done by reading the complete file for each column but the number of file accesses would then be excessive and equal to the number of readings. A better procedure is to use a more efficient routine to rotate the file $90^\circ$ and then to process the columns as rows.

Hunt (1972) considers this problem of transposing a matrix larger than the core space in connection with Fourier transforms which are similarly calculated by rows and then columns. He suggests reading a $1024 \times 1024$ array to his available 262K of core in several sections, rotating each section in core, and recombining in an output file. He also gives a reference to a transposition technique which requires random access to only two rows of the matrix at a time (Edlundh 1972), although this works only for a square matrix of side length $2^n$.

The solution adopted in the Honeywell programs with only 32K of core was to use a system command to read the data in 1260 word blocks rather than by individual records. Input and output blocks are held in separate buffers and values transferred between coincident positions. For $N$ readings the number of file accesses is $(N/1260)^2$ which is efficient so long as $N/1260$ remains fairly small, which it does in practice. For $N = 1260^2$ the number of accesses would be $N$, and impossibly large, but that is two orders of magnitude greater than current survey dimensions.

The rotation subroutine TWIST is called automatically by INTERP, but may also be called through a main program ROTATE to transpose any other required file.

These subroutines allow accurate control of the plotting scale for teletype display, and provide for interpolation to any larger (or smaller) scale limited only by the dimensions of the data buffers (set at 500 readings). Their main disadvantage is that for large changes of scale
Fig 2.9 Shalford Manor: Resistivity survey after interpolation for plotting at large scale (reduced here from plot at 1:48 scale). Filter radius 2; range mean to mean + 1.5 standard deviations.
of features. In fig 2.9 the survey of Shalford manor was plotted at
twice the scale of fig 2.4 and 4 times that of fig 2.3. The individual
characters are less obtrusive in the larger plot but the additional
interpolation has rounded the corners and caused the edges to fade
gradually through lower readings. Sharpness could perhaps be partially
restored by filtering after interpolation but the effect would be
confused by the asymmetry of the output array and the number of
readings to be filtered is greatly increased. A better solution
would be to use an interpolation procedure which gives a more accurate
representation of variations when the data is substantially expanded.

A higher order general polynomial interpolation would perhaps improve
on the special case of linear interpolation, and could be found through
Lagrange's interpolation formula which fits an N-1 degree polynomial
through each N data points (eg Dorn & McCracken, 1972; McCalla, 1967).
The generally recommended method for geophysical use however appears
to be interpolation with cubic spline functions. Scollon (1970b)
suggests this as a method of extending data arrays to avoid edge effects
in filtering, and the more general geophysical applications of the
technique are described at length by Bhattacharyya (1969). The
procedure involves fitting a cubic polynomial piecewise through the
data but maintaining continuity of first and second derivatives at each
of the initial data points. The continuity of curvature makes it the
smoothest function which interpolates the readings, and its action is
comparable to that of a draftsman's mechanical spline. There is a
description of the method and a specimen Fortran program listing in
Before an array of data is displayed values corresponding to the density levels of the plot must be assigned to the readings. Adjustment of the range and distribution of the levels allows considerable control of the final quality and content of the plot.

Occasionally the full range of the data is simply divided into the required number of levels and plotted, but more usually maximum and minimum cut-off levels are specified. Readings which exceed the maximum are assigned to the highest (or most dense) display level, and those below the minimum to the lowest or lightest. The cut-off values may be defined directly by their numerical values but are often more conveniently stated in statistical terms.

The readings will usually form a roughly normal distribution slightly distorted by the presence of anomalies. Significant anomalies are usually positive, but may be negative (relative to the mode of the distribution), especially in some resistivity surveys. The peak of the distribution may therefore be slightly skewed in the direction of the anomalies or, if the anomalies are very distinct the distribution may become bimodal. Anomalies at the mode of the distribution cannot be distinguished from background noise in terms of magnitude, but may in principle be separated through filtering if they have distinct spatial characteristics. In practice a plotting range which discriminates in favour of anomalies at the expense of noise can usually be defined.

The range may be specified in terms of standard deviations about the mean. For a normal distribution 95% of the readings lie within ±2 standard deviations, and this provides a useful maximum cut-off level for initial test plots which is then if necessary refined. A plot of the full range would be filled at the medium display levels by values
from a value near the mean in the direction of the anomalies. Features may then be seen against a relatively clear background but two plots are necessary if positive and negative anomalies are to be displayed. In each version of the program there is a routine to calculate the statistical results needed to set the plotting range.

In the Honeywell programs there are two versions of the statistical subroutine; one of them (STATS; AM Laboratory Report G1/79 sections 1.5.4 and 2.5.6) forms part of the interactive processing system, and the other (STATS2, section 2.6.4) maintains accumulated totals through repeated file accesses and is called from the core-independent program INTEGER (G1/79 sections 1.6.5 and 2.6.5). STATS2 may also be run under a separate main program SIGMA (G1/79 sections 1.6.4 and 2.6.4.) to return statistical results on any file independently of the processing programs. The values returned are for the maximum, minimum, mean and standard deviation of the data (together with the number of blank values excluded). The standard deviation (or r.m.s. variance about the mean) is found from the standard formula

\[ \sigma = \left( \bar{x}^2 - \bar{x}^2 \right)^{1/2} \]

where \( \bar{x} \) denotes the mean of the data. The subroutines which set the plotting levels from the statistics are described in sections 1.5.5 and 2.5.7 of the report.

The histogram program which is required to allow the plotting range to be defined from explicit and detailed knowledge of the data distribution reached the stage of a first draft, but has since waited for inclusion in the forthcoming production programs rather than with the existing system.
Fig 2.10  Danby Wiske, Yorks.

Magnetic survey of Iron Age settlement showing hut circles and enclosures.

Filter radius 3
Range mean to mean + 1.5 standard deviations
Statistical results similar to those of the Honeywell programs are calculated and printed by the Surrey plotting range routine, but the values refer only to data currently in core. This causes difficulty in a plot such as fig 2.10 in which a standard treatment was given to each of the several sections which were later joined to form the complete plot. There are strong anomalies in part of the plot and enclosures and hut circles are visible, but none in other sections. The effect of applying the same statistical limits to all the sections, but using locally calculated statistics, is to define a narrower plotting range where anomalies are absent. This raises the random background noise into the plotting range and gives a much noisier plot than in the sections with anomalies. This could have been avoided by making test plots of each section and visually adjusting the noise levels, but it would be far less laborious to calculate the statistics from the complete survey.

It is one of the main objects of interactive processing to define the plotting range empirically in this way, but the practical scope of interactive experiment is limited by the speed of the display.

It is only through adjustment of the minimum cut-off level that random 'uncorrelated' background noise which (as stated in section 2.2.1. above) is spatially identical to the smaller anomalies can be excluded. Fig 2.11 shows the visual effects of different plotting ranges. The object here was to test whether the anomalies at the centre of the resistivity survey might represent the remains of a Roman villa. In the initial plot (i) with a plotting range from mean to mean plus 2 standard deviations the anomalies are rather weak. They are strengthened in (ii) with the maximum cut-off reduced to 1 standard deviation but are still fragmentary. In (iii) the lower cut-off is lowered to mean - ½
Fig 2.11 Sarratt, Herts. Resistivity survey of suspected Roman building. 1:400

(i) Range mean to mean + 2 st. dev. (ii) Range mean to mean + 1 st. dev. (iii) Range mean - ½ st. dev. to mean + 1 st. dev.
continuity of the features. The noise level is increased in (iii) by
decreased by the mean of the distribution, but the plan of the central
anomalies is still very diffuse. It appears that any remains which
occur are probably then in a very poor state of preservation.

A rise in the minimum plotting level will in general reduce the noise
level but at risk of losing anomalies, and a reduction in the maximum
will saturate the plot but reduce the discrimination between different
strengths of feature. Further adjustment of the response is possible
by using non-linear plotting levels as determined by a suitable
function. A plot of the square root or log of the readings will
compress high values in relation to low, and exponentials have the
opposite effect.

In the Surrey programs there is a facility to expand the data
exponentially as one of the options in the plotting level routine
(written by Dr L Julian). This substitutes the value

$$X' = X \exp (R(X - X_{\text{max}}))$$

for each reading $X$, where $X_{\text{max}}$ is the maximum value in the data. The
rate of increase may be controlled by adjusting the value of $R$. This
option provides an effective and precise alternative to raising the
minimum cut-off level for supressing background noise, but again careful
adjustment is needed if features are not to be lost. The effect of
this procedure on the contrast of the plot is shown in fig 2.12.

In the Honeywell programs an experimental non-linear option was included
in which the need to call a function routine for each reading is
avoided by matching the values against a pre-set non-linear list of
plotting ranges. It is also possible to state arbitrary limits to each
density level. The program requires modification and the extra
complexity of this approach is probably not justified by the possible
Fig 2.12 Avebury; resistivity survey

Plots showing effect of exponential expansion on noise level.
A standard procedure used in image processing to obtain contrast variations similar to those discussed here is histogram modification (described in Gonzalez & Wintz, 1977). In this the elements to be plotted at any given density level are mapped to another level. They may be redistributed according to a stated probability density function or assigned arbitrarily to directly specified values. The object is usually to expand the contrast range of the image, either uniformly or selectively for light and dark levels. An equivalent result is obtained in the Honeywell and Surrey programs when the stated plotting range is divided into levels which necessarily correspond to the full contrast range of the display. Scholler uses histogram modification for enhancement of aerial photographs (1978).
Facilities to display results at any stage of processing in either numerical or pictorial form are provided in each version of the programs.

2.7.1. **Numerical output**

Numerical listings may be required both of the current values during interactive processing and of data in saved files. Some operating systems may allow files to be listed directly using system commands but with Honeywell this is not possible for the 'random binary' files used for numerical data. Instead, files must be read under program control. The output is then formatted and printed using the same subroutines for both current and saved data. The file containing the printing subroutines (NUMBERS, AM Laboratory Report G1/79, sections 1.5.3 and 2.5.5) may therefore be loaded as part of the interactive system or with the independent program FPRINT (G1/79, sections 1.6.3 and 2.6.3) for printing large saved files.

These programs use Honeywell character handling facilities to accept a run-time format descriptor as a program parameter. The readings are held as real values which may vary greatly in magnitude at different stages of processing, and so the system allows a format suitable for the current values to be used each time results are printed. The program also tests for blank marker values and substitutes any character string entered by the user. Gaps in the data may therefore be indicated by dots, spaces etc. These refinements allow clear and ordered presentation of the data and are useful for such purposes as printing readings at a spacing which corresponds to the scale of a site plan or plot, but they would be very cumbersome to program in standard Fortran.

The program also divides any block of data which exceeds the width of the terminal into vertical strips and centres the block (according to the format descriptor) across the required number of strips. Edge values
on this sectioning arrangement will be needed in the production programs.

2.7.2 Pictorial Display

In all the work so far various forms of density display have been used for all final output. These are not necessarily the most desirable for all purposes and possible alternatives are discussed further below. Only displays using the available equipment and copying facilities are considered. The plots are to be circulated in departmental reports and a first requirement is that they should withstand office photocopying. This would exclude all colour or continuous tone displays.

Symbol Plots

The Honeywell programs were written for a time-sharing system accessed by teletype and output is limited to printed characters or symbols. This is not an impossible restriction and adequate working plots may be produced, especially when the scale is large enough to give good resolution. A list of characters of density corresponding to each plotting level is held in core, or others may be substituted by the user. Up to three characters may be overprinted in each position to give a plot of the required range of densities. It is important that the final visible density levels are reasonably linear if the calculated distribution of plotting levels is to be achieved. At present the programs only allow for 8 levels on the assumption that fine gradations are not possible by this method, but Dewdney (1976) suggests otherwise. Symbol density plots of this kind are quite widely used in image processing work and he describes a method of selecting characters to give a subjectively linear grey-scale of 50 levels by overprinting 4 characters at each level. An adaptation of his character set would be useful in future work.
Comparison of displays

Fig 2.1 Cotton, Yorks. Resistivity survey of square barrow ditches

Dot-donut (Tektronix display)
from core-independent programs PLOT, G1/79, sections 1.6.6 and 2.6.6) will section a large output file into strips to fit the terminal similarly to the data printing routines (but without overlap). They also print a border around the plot with scale marks every 5 (initial) readings, and the sequence is maintained across the strips. Headings and captions are easily printed on the teletype and this saves much work in lettering and finishing the plans.

An experimental option was included in the program to reduce the possible coarse and repetitive effect of the printed symbols by choosing them at random from a list of alternatives at each density level. The result is quite effective except that discrimination between levels is reduced. Probably a similar but more accurate result could be obtained by using an extended plotting range such as that of Dewdney so that fewer readings coincide at each level.

In fig 2.13 plots of a pair of barrow ditches are compared. Two characters were overprinted for each reading in plots (i) and (ii), but for (ii) they were selected randomly. Fig (iii) shows a dot-density of the same data. These results show that symbol plots can compare quite favourably in detail and resolution with dot-densities. They are also potentially very fast to produce on a high-speed printer, and are an option worth retaining for test plots and intermediate results in future programs.

Dot-Density Plots
This form of plot in which dots corresponding in number to the strength of the reading are offset randomly within each display cell to give a simulated half-tone effect has been used in all work so far with the Surrey system. The technique seems to have been invented by Scollar (Scollar and Kruckenberg 1966). It has the advantages that such plots
self-explanatory to users. A disadvantage is that they are slow to produce on the pen plotter. For some early work at Surrey a Tektronix storage tube display was used for interactive display of test plots. The results were quite effective (fig 2.13 iii), but the display was found to be very little faster than the pen plotter. Also, accurate interpolation is required for scale plots, and so it was abandoned. The present configuration allows small test plots to be displayed on an oscilloscope but all permanent output is on the pen plotter.

The visual effect of dot-density plots depends on several variables including the number and size of the dots and the degree of offset given to each. The present program uses 9 dots per reading which is the minimum practical number and allows rapid plotting, but at the cost of difficulty in obtaining a saturated black. The dots may be enlarged into finite triangles or diamonds, but with the present plotter they then appear distorted, and to increase the number to 16 or 25 would extend the plotting time proportionately. This would matter less if dot-densities were used only for a small amount of high quality finished work as proposed below and not for routine plotting.

Weak linear features are easily recognised in a dot-density, but small isolated anomalies are not. A 2-dimensional display is effectively a low-pass filter which limits the resolution to the cell size of the plot. This applies both to the density plots considered here and to contour plots which are discussed in terms of their filtering characteristics by Wren (1975). The apparent resolution could be improved by interpolation to increase the number of display cells, but no density plot could represent the full detail of the analogue traces.
The graphical traces of magnetic traverses plotted by the field recorder allow the anomalies to be interpreted on the basis of gradient and profile, as well as the magnitude and extent which are the only quantities represented in a density plot. As explained in chapter 1 it is desirable to extend computer processing to these surveys but if interpretation to the present level of detail is to continue the resolution of the plots must be retained, and this could best be done by replotting traces from data sampled at sufficiently close intervals. The use of computer-generated plots would reduce the great manual labour of preparation and provide much-needed flexibility in display. Fig 2.14 shows the effect of a simple change of plotting scale. Plot (i) was recorded at a sensitivity of 20 gamma/cm, the setting needed to detect small background features, but the kilns run off the paper. The square was then resurveyed at 150 gamma/cm to give the very clear representation of the kilns seen in (ii), but it would obviously be better to make such adjustments in the course of processing rather than fieldwork.

A routine to plot traces similar to those recorded in the field would be simple to implement using existing equipment, and they would be much quicker to draw than dot densities. It is therefore proposed to make this a central feature of the production system. Dot-densities are preferable for publication and would ideally be used in place of such manually drawn plots as fig 1.12 (ii), but their significance would be assessed on the basis of computer-drawn traces. If no detailed interpretation is required the processing parameters could be established through initial symbol plots. Replotted traces would also provide a suitable means of computer display for such
Fig 2.14  Brampton, Norfolk
Magnetic survey of a group of pottery kilns
wider than 1 metre.

Contours

Contouring is a quite effective form of display except for data which is very noisy. It has not been included in either set of programs, but occasional jobs have been done using programs available elsewhere. These programs do not seem able to locate the contours when there is a plateau in the data, and require continuous gradients for acceptable results. I have planned a routine which should at least be capable of following each contour (by the principle of solving a maze - always turn in the same direction), and this is waiting to be included in the new system.
3.1 Introduction

The subject of this chapter is the overall structure and design of the processing system. The operational consequences of this are at least as significant as the capabilities of the various individual processing routines described in chapter 2.

This project began with the initial object of providing a more flexible and immediate alternative to the slow and expensive plotting of surveys by sending data to a batch processing bureau. Only a small proportion of surveys were plotted through the bureau; either resistivity surveys for which no alternative display was available, or magnetic surveys for which a dot-density plot was specifically required. The two initial sets of programs (Honeywell and Surrey) referred to throughout chapter 2 have been used as intended to maintain output in these restricted areas, but have also provided experience on which the design of a comprehensive processing system can now be based. The bulk of the Ancient Monuments Laboratory output has always relied on analogue traces and at least 95% (in terms of 30 m survey squares) still does. These charts in many ways provide an effective and immediate presentation of the survey findings, and the work of preparing them can only be transferred to the computer once clear practical advantages over alternative methods are to be gained. The limitations of the present programs which prevent this and the ways in which they can be overcome are discussed in the course of this chapter.

The need for program development work of this kind might have been avoided had suitable programs been available elsewhere, but the
circumstances and survey techniques of the AM Laboratory differ sufficiently from those of other workers in this area to make an independent approach worthwhile. If any programs which met the Laboratory requirements did exist and could be borrowed they would probably be very difficult to implement. Standardization of computer languages between manufacturers is minimal, and for operating systems it is non-existent. Any complete working system must therefore be at least partly machine dependent, and experience in this project has shown that the easiest way to convert programmes is often to rewrite them.

One commercially available geophysical processing system which offers some of the facilities needed is the SACK (Surface Approximations and Contour Mapping) package from the SIA bureau (SIA, 1975). This offers most of the (spatial domain) processing facilities required but output is only by contour plots and the system is not designed for interactive use. It would also be very expensive.

The only published work of direct archaeological application is that of Linnington and Scoller. The Home Computer System described by Linnington (1968, 1969) was written for an IBM computer with punched card input and line printer display for output. The facilities provided compare in many respects with the AM Laboratory Honeywell programs but processing capacity is limited to a 101 x 101 array of readings.

The work of Dr Irwin Scollar has been referred to throughout previous chapters and has been the source of many of the ideas discussed. He has dominated the literature of archaeological magnetic surveying for many years and has published comprehensive accounts of the survey and interpretation techniques used at the Rheinisch Landesmuseum, Bonn.
He has applied a full range of image processing techniques to survey results, and recently using a new dedicated computer and display system has extended the area of application to include digitized aerial photographs (Scollar 1978). No work as ambitious as this has been attempted elsewhere in archaeology, nor would it be possible here. There are also, as explained in the earlier discussion of filtering, basic differences of approach. Soil conditions and the shallow depth of most English archaeology allow the use of the fluxgate gradiometer which gives an output restricted to spatial wavelengths within a range the whole of which may be of potential archaeological significance. Anomalies may only be further identified by visual evaluation rather than computation, and much of the processing to extract significant features which is of central importance in the methods used by Scollar is in many cases unnecessary. Facilities to extract spatially defined features are needed in a comprehensive processing system but at the AML Laboratory are of secondary priority to a full range of display options and the capacity to process large volumes of data.

A processing system of some elaboration is therefore needed to display the data from magnetic surveys, irrespective of whether any numerical enhancement operations are applied at all. Conversely, once the framework of a data management and display system is established extra processing options may be added with very little extra difficulty. The ideal would be a system capable of plotting a large magnetic survey in a session of perhaps half a day at the computer, and so replace several days of manual chart preparation. This is possible in principle using present equipment, but not with the existing programs. The requirements are first a more powerful control system so that the program will accept and retain instructions for a complete series of test plots, rather than requiring all parameters to be re-entered interactively for each plot. Facilities for direct interactive
working must however be retained to allow final adjustment of parameters after
the initial series of standard tests. It must then be possible to produce
a single final plot for the complete survey, rather than process it as
a mosaic of small sections. It is worth investing extra programming
effort to transfer as many incidental tasks as possible to the computer.
Minor time-consuming jobs at each stage of processing such as sectioning
the data to fit the program or adding captions by Letraset to finished
plots, when multiplied by the number of surveys become a critical
factor in containing the backlog or losing control of it. Computer
time is far more freely available than manpower and should be fully
exploited. The means of doing so are discussed in the section on
program design for a production system below.

Neither the Honeywell nor the Surrey program systems were constructed
for large scale working, although the Honeywell one goes some way
towards that end. They would each require fundamental change rather than
simple extension to make such working possible, and so the design of
each was deliberately 'frozen' in a minimum working configuration when
they were completed in 1976. The extended system which is to take
their place now exists in partial draft but final implementation will
depend as always on the demands of fieldwork.

The Honeywell and Surrey programs have provided experience and the
necessary background for further developments. The Surrey programs
are relevant to problems of minicomputer operation and graphical
display, and the Honeywell ones to program structure and data
management. The two systems are described in turn below together with
the design proposals for the new programs.
3.2 Honeywell Programs

These programs were written to allow immediate processing of results through the terminal at the AM Laboratory, and have been used regularly for this since they were completed in mid-1976. The Honeywell/General Electric (now renamed GEISCO) time-sharing system is a major international network accessible from around the world and with a satellite link to the main computer in America. It is in many respects over-elaborate for the requirement of this project, but it was all the DOE could provide at the time.

The intention was to provide a reasonably complete set of processing facilities with immediate display as a teletype plot, but with provision to output processed data on punched tape for final plotting elsewhere if necessary. Readings were transferred in this way in some cases before the Surrey programs were operational. The data for fig 2.1 for example was prepared through the Honeywell programs and sent to CI Data Centre for plotting. An additional object was to investigate techniques of programming and processing that would be relevant in future minicomputer applications. The results have been applied in part in the existing Surrey programs, but much of the experience gained will not be fully put to use until the new production system is operational.

The system is fully documented in AM Laboratory Report G1/79 (AM Lab, 1979). In Part I there are instructions which should be sufficient to explain the system to a user, together with program notes and flowcharts, and program listings are given in Part II. The specific processing facilities of the system were described in chapter 2 above and will not now be discussed further.
The system includes a number of input and output and data preparation programs as well as the main processing routines and other incidental programs.

The input and output of data may appear to be a trivial task but proved to be a matter of some practical complexity. Magnetometer readings logged in the field are translated from digital cassette tapes to punched paper tape in binary format. There are system commands which allow data of this kind to be entered directly and held in file, but to set up a correctly ordered file this must be done under program control. The tape is therefore entered through the program INPUT (G1/79, sections 1.2.1 and 2.2.1) which tests for and interprets the various marker values recorded with the data. These markers are recorded in the field but may also be added during subsequent hand editing of the tape. They delimit traverses, blocks of traverses, comments, and also mark a traverse or pair of traverses to be deleted. The significance of a marker may vary with its position on the tape, and action at any point depends on the logical combination of preceding markers. The program must also truncate or fill out traverses of incorrect length, and there are options to do this interactively as well as to make other editing corrections and delete or split traverses as they are transferred to the output file. Traverses are recorded in the field in alternate directions and so any change in their number during processing inverts the remainder of the file left to right, which further complicates the program.

The corrected readings are finally held in a random access unformatted file ('random binary' in Honeywell terminology) and saved under a name generated by the program.
Such files are the standard type used for numerical data, but may only be accessed through program read and write statements and not through keyboard commands which are only available for formatted character files. A program PRINT is therefore needed to print data from file, and others are required to check individual records, RCHECK and RECORD. The correction routines called from INPUT are held on a separate subroutine ALTER which may also be called from another programme EDIT, which allows corrections to be made to previously created files.

Initial data, mostly from resistivity surveys, which is not automatically logged in the field is typed at the terminal to give a tape of formatted 'ASCII' characters. This may be entered and edited directly using standard system commands but must then be translated through a program to give a random binary working file. The program for this, CONVERT, is relatively simple because no correction to the data is required. The only complication is that a traverse must usually be split when typed to fit the width of the terminal. The program therefore combines a stated number of records from the ASCII input file in each record of the output file. All working files are of real (floating point) type. There is no core space penalty for this because the system uses a 32 bit word and holds one item of either real or integer data per word.

There are other programs which provide equivalent options on output. TAPEDLIST allows any random binary file to be listed to punched tape according to a stated run-time format descriptor entered by the user, and OUTPUT lists the file in a packed binary format similar to the initial logged tapes. The binary tapes are compact and convenient for storage and provide much quicker input and output through the terminal. Arithmetic options in the program allow the data to be scaled to fit the available range 0-255.
An additional data preparation program COMBINE allows initial sections of data to be assembled into a processing file representing the complete survey, or any other group of files to be joined or superimposed. The detailed operation of the program is described in section 2.4 (Edge matching), above.

Once the data files have been set up, which is usually by far the most lengthy part of any job, they may be processed interactively through the main program (sections 1.5 and 2.5 in report). This allows processing and output options to be applied in any sequence to any data file or stated section of a file within the capacity of the system. The processing subroutines are held in overlay files which are loaded as necessary and may be called in turn indefinitely to produce a series of plots if required. Fig 3.1 shows the structure of the system.

The stated block of readings is read from file at the start of processing and then (in its various successive processed forms) held in a core buffer until it is overwritten by a new section of data, or until the end of the program run. A copy of the current values may be written to an output file at any stage. The file dimensions correspond to those of the actual survey but the core array does not. Items are addressed during processing by their actual survey coordinates but the values are located by mapping into a 1-dimensional array. This has the advantage over the use of redimensioned 2-dimensional data arrays that the relative sizes of the arrays need not be predetermined and core may be freely re-allocated between input and output data. This is particularly important for interpolation where initial, intermediate and output arrays of different dimensions must all be accommodated.

The core buffer has a capacity of 10,400 words and so will only accept data from 2-3 30 m survey squares when interpolated for plotting at 1:200 scale. A file representing a typical magnetic survey would be
Fig 3.1 Interactive Processing System
(Honeywell version)
Block diagram showing main subroutine calls

Control returns to SUBSET from the initial subroutine in each Link (or from either subroutine in Link1)

- Subroutine loaded in overlay file
- Subroutine not in overlay

C = CORNERS
S = STATS
10 times this size and so other facilities are required for efficient working in such cases:

To process files which exceed the core space through the main system would mean setting up working files to accept the data between each pair of subroutine calls, and then reading and writing all data in sections. This seemed overcomplicated at the time the programs were written and so an additional set of main programs was provided, each of which calls a single processing routine but is capable of sectioning the input and output data to process files of unlimited size (G1/79 sections 1.6 and 2.6). It was not intended that these be run in a continuous interactive sequence, but rather that they should be run once each to produce a single final plot once the parameters have been selected through the main interactive program. The practical use of the programs in this way has been restricted by what was found to be the excessive cost of processing large files on the time-sharing system. This is a problem which will only be solved by limiting future work to a minicomputer where the cost rises only with time, and not in proportion to the volume of data or number of file accesses.

In addition to the input, output and processing programs, there is a record keeping system. All transactions which result in an output file or plot are noted in a file, together with a program identifier, the job heading entered by the user, and the date. The program INDEX when run prints a table of this information. This was intended as an aid to keeping track of the proliferation of output files which are created and named by the system.

The Honeywell programs have provided a firm basis of experience from which to develop the next version of the system, but are not themselves suitable for extension. This is partly because time sharing
was only used to begin with for the lack of alternative, but more importantly because Honeywell Fortran is non-standard and almost totally incompatible with any other. The language contains not only many special system routines, but also non-standard data types and these make conversion to a standard form almost impossible. This has been a great hindrance to past progress and is a problem to be rigorously avoided in future work.

3.3 Surrey Programs

These programs represent a partial implementation of the Honeywell system modified for use on a minicomputer. They were written under the terms of a contract between the DOE and the University of Surrey by Dr L Julien and Christine Stannard of the Physics Department and are described in their report (Julien and Stannard, 1976). They were designed to allow routine working on the Physics Department's Nova 800 minicomputer with access to a plotter for output.

The processing facilities are similar to those of the Honeywell system in essentials but have been simplified in various respects. There is a data input program ARF2 which accepts binary tapes and writes them to a disc file as with the Honeywell system. It does not have correction facilities and so tapes are usually prepared at the AM Laboratory and corrected copies then brought to Surrey. Alternative versions of the program are provided which either invert alternate traverses or read them all from left to right.

The main processing program operates interactively with parameters entered as required, and is structured similarly to that of the Honeywell system with subroutines in overlay files which are called as necessary. The routines which filter the data and set the final plotting range are similar in principle to the Honeywell equivalents,
but could not be adapted directly because of language difficulties. Output is by dot-density plot and because the plotter is adjustable no interpolation is needed in principle for plotting to scale. A routine which expands the data block to twice the size by interpolating intermediate values is however included to increase the range of adjustment.

One major point of difference from the Honeywell program is that the readings are held during processing in two-dimensional input and output arrays of fixed size. This has advantages of simplicity but means that not only are the relative maximum dimensions of input and output data arrays restricted, but also that the permissible shapes of array are limited. A survey is as likely to be long and narrow as square and so this is the cause of some inconvenience. The available core space is not sufficient to hold data which exceeds the capacity of the printer or plotter and so there is no need for sectioning the output.

The data type used throughout the programs is fixed rather than floating point. This is because two 16 bit words are needed for each floating point value, and so their use would waste too much core space. Given the limited range of the initial data and the small number of final display levels the use of fixed point data is not a serious restriction. Most of the processing routines operate quite well with integer arithmetic and a perfectly workable system is possible. There is sometimes a loss of resolution when a very narrow plotting range is required, but this could easily be overcome by arithmetic scaling. Integer processing will probably be retained in the production system.

All the dot-density plots included in this thesis except fig 2.1 were produced at Surrey and the experience of using these programs for routine plotting has shown the great advantages of a minicomputer for
this application. It is economical for processing work (although not necessarily so for data preparation, program testing and file editing at current costs), and has the overwhelming advantages of rapid input and output and freedom from a slow and unreliable telephone link. These advantages must be exploited to the full if the aim of routine processing of all surveys is to be realised.

3.4 Design Proposals for a high-capacity production system

It has been established through progress and experience with this project so far that survey processing on a scale corresponding to the Ak Laboratory workload is possible even with a modest minicomputer rather than a large main frame provided the programs are designed with efficiency of throughput as an overriding concern.

Neither of the existing processing systems has the capacity required. The Honeywell programs are restricted by the telephone link and the cost of using the system, and the Surrey ones by their limited data capacity. Additionally both sets of programs are fully interactive, which means that all parameters are entered in response to keyboard requests and must be re-entered for each section of every survey. This means that operator response time far exceeds the processing time and roughly equals that required for final plotting. There are merits in interactive working which must be retained, but some less laborious means of specifying standard or repeated operations is required.

The combined effect of these limitations is that taking account of all stages from data preparation to finishing the plots it takes longer to compute a magnetic survey than to process it manually using the analogue charts, and in consequence computing has been limited to experimental jobs with only small resistivity surveys plotted routinely.
In designing a system to overcome these problems the actual nature of the processing and display operations is of minor concern. All the image processing and display techniques discussed in chapter 2 are required, but initially it is sufficient to retain flexibility so that additional overlay subroutines may be added without alteration to the remainder of the system. The most urgent task is to develop a set of control and data management routines having sufficient capacity and which impose no restrictions on the processing routines.

Most of the requirements for such a system have already been stated but they may be summarised as follows:

(1) Parameters are usually selected empirically on the basis of test results. The system must therefore be capable of applying a series of different processes to a section of data in the initial tests and then processing and plotting the complete survey when the optimum treatment has been chosen.

(2) There may be 50,000 or more readings in a magnetic survey, and perhaps 5 times as many if ever it becomes possible to sample the signal more closely. The amount of data will be multiplied again at various stages of processing and would exceed the core capacity of even a large computer. The survey must therefore be sectioned for processing and experience with existing programs has shown that to do this manually causes error and delay not acceptable in production work. The final plot must also be sectioned if necessary to fit the output device and the sections required will not necessarily correspond to those suitable for processing. There must be no restrictions on the shape of the data array.
The data from a given survey is likely to originate in arbitrary pieces which require adjustment and equalization for consistent treatment. It must therefore be possible to specify the input to the system in terms of a series of blocks of data held in different files, to read and variously modify the data from each, and to locate them in a single output file.

Items must be extracted from the data in various sequences for different processes. For example:

1. Sections of data must overlap to avoid edge effects in convolution filtering.

2. Access is required in random sequence for contouring.

3. It must be possible to extract columns of data from file for interpolation and Fourier transform routines.

4. If traces are to be plotted from the data with a hidden line removal option this would be most easily done by reading the file records in reverse order.

The total number of parameters required to define a complete process is very large and so they must be held in file so that only those that differ for a given job need be modified. In the Honeywell system these requirements were met in part through different programs all of which accepted parameters interactively. Multiple input files were transferred to a single output file by COMBINE; small test plots could be produced through the interactive system calling different processing routines, and it was intended that large files could be processed and plotted through a series of distinct programs. This approach has operational and structural disadvantages and the various functions would be better combined in a single system. As explained in section 2.4 there is no clear distinction to be drawn
between initial data equalization and subsequent processing, and the
programme COMBINE provides a partial model of the data input facilities
required in a processing system. Additionally any distinction between
large and small surveys is arbitrary and machine-dependent, and is not
a matter which should concern the user. The program should deal with
either as appropriate.

A solution to the conflicting requirements of data access at different
stages of processing is found in a fundamental restructuring of the
program. Instead of reading the values from file to a core array
before calling the processing subroutines as has been done in all
previous programs, the sequence is reversed. In place of a subscripted
array reference a function subroutine with the coordinates of the
required item as arguments is called for each item to be processed,
eg ITEM (I, J, K). This allows each operational routine to be written
as if it was addressing locations in an unlimited three-dimensional
virtual array, and to do so in any order. The third dimension is
needed to allow grid-to-grid operations such as combining superimposed
input files, or multiplying by the coefficients in a frequency domain
filter array.

The function ITEM checks whether the specified reading is present in
the core array and if so returns the value. If not it calls subroutine
IREAD to read that file record and enough subsequent records to fill
the data array. This means that for a file processed in reasonably
consecutive order no unnecessary or repeated file accesses are made and
the function simply locates each item in the core array through a simple
calculation of position as in the existing Honeywell programs.

All readings are held in a single one-dimensional array which may be
freely divided and re-allocated between as many input files as required,
with a section reserved for the output data. Readings may be written
to file after each stage of processing and then read for the next stage if necessary, or they may remain in core throughout if the number of items is less than the total core space. The program will therefore in comparable circumstances make no more file accesses than the existing program. There will be some increase in execution time because of the need to call a subprogram for each reading but the addition to the amount of calculation already required is slight. Any increase in processing time which reduces manual effort elsewhere is in any case justified. The general scheme of the program is shown in outline in the block diagram fig 3.2.

The use of this input/output procedure will greatly simplify the problem of accepting data from multiple input files to be processed in parallel and written to a single output file (requirement 5 above). The coordinates within each file and the number of the file in the sequence are simply addressed directly as arguments of the data function ITEM.

A converse problem is to apply multiple processes in sequence to any given block of readings. This is necessary for production of series of test plots (requirement 1 above), and may be achieved by enclosing the whole system in an outer loop, during each cycle of which a complete sequence of process and plotting instructions are executed. The loop is located in the subroutine CONTROL.

In any given run of the program it will therefore be possible to produce a combined plot or output file containing data from a number of input files located in relation to each other, or a series of different plots from any given block of data. The final display routines may call readings in whatever order they require using the same data function as other routines, and so section the plot to fit the output device.
Fig 3.2 New processing system with virtual data structure

Block diagram

- Subroutine calls interactive routines to request parameters and writes values to file.

- Subroutine directs control to each plot in sequence.

- Reads parameters for this plot from file and calls processing and display subroutines.

One subroutine to set values for each processing or display routine

- Subroutine loaded in overlay file

- Subroutine not in overlay

- Subroutine called from each subroutine

Function ITEM returns value for specified location in virtual program array from actual array.

Function IOUT transfers values from virtual program array to actual output array.

IF: ITEM (I,J,K)

F: IREAD

I: IREAD

K: IREAD

L: IREAD

M: IREAD

N: IREAD

O: IREAD

P: IREAD

Q: IREAD

R: IREAD

S: IREAD

T: IREAD

U: IREAD

V: IREAD

W: IREAD

X: IREAD

Y: IREAD

Z: IREAD

Function IOUT transfers values from actual core arrays to overlay file.
The instructions for a complete program run are held in a control file, which for each plot contains a list of initial data blocks and list of references to processes required for that plot. The actual parameters for each process are held in other files and read only when that process is to be executed. This allows flexibility in permutating previously defined processes and data blocks. A control file may be defined interactively and run immediately so that for the user the system operates much as do the present programs. The advantage comes when a plot is to be repeated but modified in detail, or the same series of test plots is to be produced from different data. In either case only the relevant parameters need be modified and this is done by editing the existing control file.

The separation of input and output routines and the use of virtual core addressing means that not every item read from the input array need necessarily be returned to the corresponding position in the output array. The system may therefore also be used quite simply for the interactive editing and correction of initial data and could take the place of the Honeywell program INPUT. The complete initial tape with its marker values etc is read to the input array. It is then examined, tested and edited as necessary by the operator, and only those items which are accepted as valid data are written to the output file.

To preserve future flexibility in the implementation of the system the incompatibility between different manufacturer's versions of Fortran which has proved troublesome in the past must be strictly guarded against. This can only be done by working to the most rigidly conservative definition of the language, and it is therefore proposed to keep to a subset of the 1966 standard drawn up with this aim in view (HECB, 1976). Initial implementation will be on the Data General machine, which requires only minor departures from the standard. All
non-standard instructions (eg graphics and file creation and opening commands) should however be kept isolated in clearly annotated subroutines.

If a program of this proposed size is ever to work at all a rigorous approach to design and implementation is essential, and this is best maintained through the methods of structured programming. Here instead of directing control within the program by means of GOTO statements, each stage of processing is identified and isolated in a separate program unit, which in Fortran takes the form of a subroutine or function subprogram. It is not possible to avoid GOTO commands completely in Fortran which lacks IF THEN ELSE, DO WHILE and DO UNTIL statements, but they should be used sparingly and only where necessary to code local groups of statements equivalent to these constructs. GOTO statements should not be used to switch control between different tasks or to return to earlier distinct sections of coding.

By this method the main program... are reduced to little more than lists of subroutine references. The subroutines themselves should be kept simple, and according to Kernighan and Plauger (1978) each should perform only one task, clearly. In the course of program design the program units are defined progressively starting with the most general high level ones and all detailed processing postponed to the lowest possible level. The resulting inverted tree has no cross links between the outer branches, although these may be introduced in the form of shared subroutines as an optimization measure (Jackson, 1975).

There are numerous advantages. Most of the work is done at the design stage and final coding reduces to a trivial exercise. There is an enormous gain in clarity. It should be possible to produce program text which is directly intelligible (with comments) without the copious documentation that was needed for the Honeywell programs. Flowcharts
are unnecessary. They are a device to indicate the flow of control within a program unit, and by the nature of the structured approach this is always sequentially downwards through successive statements. Any program complex enough to need a flowchart for comprehension should be further broken down into distinct units. Instead only simple block or tree diagrams illustrating the hierarchy of subprograms are required.

Debugging should cease to be a major concern. It increases in difficulty very rapidly with the number of GOTO statements and once a program reaches the logical complexity of, for example, INPUT it becomes almost impossible to debug. Such programs are also very dangerous to modify. In a structured program by contrast the individual routines should be simple enough to be thoroughly hand checked and easy to modify.

The Honeywell programs were written traditionally with GOTO statements and flowcharts. Structured design has only come into widespread use since they were written, but I had begun to discover some of the benefits from experience (compare the notes on INPUT, section 2.2.1 in report G1/79, which was written at the beginning with those for the plotting subroutines, section 2.5.8 written at the end). Experience so far with the new programs has shown enormous gains in clarity and simplicity of coding, and the final bulk and complexity of the program should be less than they were for the Honeywell system, although the new system will be far more flexible and powerful.
Case Studies: Examples of the display and interpretation of survey results

A number of plots of survey results have been included in previous chapters to illustrate general points of theory or application. In this chapter plots are used instead to demonstrate the methods applied and results obtained in processing data from particular surveys.

4.1 Wharram Percy, Yorks

Here the ditches and banks of a field system on high ground near to the deserted medieval village of Wharram Percy were surveyed. The site lies on the Yorkshire Wolds (grid ref SE 855 644) and the geology is chalk.

The banks defining this system of enclosures were still faintly visible above ground and the survey was required to test for the presence of associated ditches. These were detected and are clearly visible as anomalies in the plot (fig 4.1). Some 12 30 m squares were covered which makes this only of moderate size for a magnetic survey. Much of the time on site was taken up with resistivity traverses which were also made across a number of the features.

Fig 4.1 shows the magnetic results as plotted using the Surrey programs. The readings were processed with a filter radius 3 and positive anomalies plotted in the range mean to mean plus 2 standard deviations in 9 levels. This is one of very few magnetic surveys of any size so far processed, and was plotted in 10 sections which were then assembled into the plan.

Most of the features visible in the initial traces can be identified in the plot, although some weaker ones tend to merge with the background.
Fig 4.1 Plot of field system with boundary ditches

Filter radius 3
Range mean to mean + 2 standard deviations in 9 levels
Fig 4.2 Wharram Percy

Unfiltered plot of top LH square of magnetic survey (as shown in fig 4.1).

Range mean to mean + 2 standard deviations in 8 levels.
The plot could be improved by a greater range of plotting levels, or of dots per reading, saturating in a solid black, but this would increase the already considerable time taken to draw each plot.

The right-angled section of ditch in the top left-hand square is also shown in the symbol plot fig 4.2. This plot was produced using the random character selection option in the Honeywell display routine and shows the detail obtainable in a symbol plot provided the scale is large enough for resolution not to be limited by the size of individual characters.

Comparison with the dot-density demonstrates how, given a clear display, there is little to be gained by filtering fluxgate magnetometer data. The unfiltered symbol plot is as clear and sharp as the filtered dot density (even at its original scale of 1:400), and lacks the blank shadow caused by convolution of the filter impulse response with the anomaly.

4.2 Kenchester, Hereford

This single 30m twin electrode resistivity square shows a square outline which was assumed by the archaeologist on the basis of aerial photographs to be that of a Roman temple. The site lies close to the line of the Roman road, now a track, leading to the E gateway of the Roman town of Kenchester some 4 miles W of Hereford (grid ref 80 445 426). A magnetic survey of the area showed, as is usually the case for building foundations, only confused anomalies with no indication of plan.

The plots in fig 4.3 show first the unfiltered data plotted in the range mean to mean plus 2 standard deviations. There is an area of high readings in the top half of the square corresponding to a rise
(i) Unfiltered plot, mean to mean + 2 standard deviations

(ii) Filter radius 3, mean to mean + 2 standard deviations

1:400

Fig 4.3 Kenchester, Hereford

Resistivity survey of Roman building.
This outline is very much stonger in the filtered plot where a clear and regular anomaly is visible. For this plot a filter radius 3 and range again from mean to mean plus 2 standard deviations were used.

The right hand wall of the building is missing and so a further plot with range mean - ½ to mean plus 1½ standard deviations was tried (not shown). This gave a noisier background with no improvement to the completeness of the plan and so was not used in the report.

Later excavation of the site showed that the square outline did not represent a temple, but rather part of the foundations of a villa. The rise in the ground next to it was largely formed of rubble and debris from the building, but in the absence of any distinct surviving structure this appeared as an area of undifferentiated high readings in the plot. This does perhaps illustrate the frequent need to compare, and to include in the report, results from different processing treatments.

4.5 Terraby, Carlisle

The problem in this survey was to locate a milecastle on a section of Hadrian's Wall which has been completely destroyed near Carlisle.

The milecastle (number 65) was thought to lie within the 250m length of the field. It is marked by the Ordnance Survey mid-way along the field but there are possible alternative sites on level ground towards either end. The site is on boulder clay with outcrops of shale close to the surface. The line of the wall follows the hedge along the upper (NW) side of the plot.

Initial resistivity traverses 5m apart and parallel to the hedge for the full length of the field were not conclusive. They showed a general noisy background with various possible anomalies. At the SW
Fig 4.4 Survey to locate milecastle on Hadrian's Wall

Filter radius 4; range mean to mean + 1.5 standard deviations
Fig 4.5 Alternative plots of squares 1 and 2 (as numbered in fig 4.4).
the two traverses. More convincing results were needed and so the
greater part of the field was covered with a twin-electrode area
survey. Fig 4.4 shows the results plotted with filter radius 4 and
range mean to mean plus \( \frac{1}{2} \) standard deviations.

The only linear or rectangular pattern found is in squares 1 and 2
where the plan seems to relate to the anomalies noted in the traverses.
There is a dense central strip with weaker features to each side.
The survey is noisy with anomalies of similar strength to those of
square 1 and 2 elsewhere, but the pattern is random and there are no
other alignments which appear to relate to the direction of the wall.

The plots of squares 1 and 2 are shown again in fig 4.5. The unfiltered
plot (i) shows the broad central anomaly clearly but little other
details. The filtered plot (ii) is taken from the complete plot for
comparison and shows anomalies which could be interpreted as a weak
square enclosure around the central feature. This interpretation is
marked in dotted outline in plot (iii). Whether it would be correct
to interpret the plan in terms of a central paved area within the
outer walls is still uncertain, but a later trial excavation did
confirm the presence of the milecastle at this position.
Acknowledgements

This work was undertaken as a collaborative project between the University of Surrey and the DOE Ancient Monuments Laboratory, and sponsored by the DOE Training Branch. Thanks are due first to my joint supervisors, Mr R Whorlow of the University of Surrey and Mr A J Clark of the AM Laboratory for their patient encouragement. I must also thank Dr A Crocker, Christine Stannard, Dr L Julien and his predecessor Mr K Knight of the Physics Department for their advice and help, and Christine Stannard and Dr Julien particularly for the work done in writing a full set of programs to run on the Surrey computer.

In writing this thesis I have drawn on the results of surveys undertaken by my colleagues for examples and illustrations, and thanks are therefore due to A David, D Haddon-Reece, P Griffiths and others who have worked at the AM Laboratory.


Atldred J C (1964) A fluxgate gradiometer for archaeological surveying. Archaeometry 7, 14-19

Part I Operating Instructions and Program Notes
Part II Loading Instructions and Program Listings
A Bartlett, DOE


Bovan B & Kenyon J (1976) Ground penetrating radar for historical archaeology. MASCA Newsletter, Univ of Pennsylvania


Dean W C (1958) Frequency analysis for gravity and magnetic interpretation. Geophysics 23, 97-127


Julien L S & Stannard C (1976) Report on work carried out on project CSP 4041, 28.10.76. Dept. of Physics, University of Surrey


Linington R E (1971) Further tests on non-symmetrical filtering systems. Prospezione Archeologiche 6, 9-20

Linington R E (1972) A summary of simple theory applicable to magnetic prospecting in archaeology. Prospezione Archeologiche 7-8, 9-59


Schwarz G T (1967) A simplified chemical test for archaeological field work. Archaeometry 10, 57-63

Scollar I & Kruckeberg P (1966) Computer treatment of magnetic measurements from archaeological sites. Archaeometry 9, 61-71

Scollar I (1966) Recent developments in magnetic prospecting in the Rhineland. Prospetione Archeologiche 1, 43-52


Scollar I (1969a) A program for the simulation of magnetic anomalies of archaeological origin in a computer. Prospetione Archeologiche 4, 59-63

Scollar I (1969b) Some techniques for the evaluation of archaeological magnetometer surveys. World Archaeology 1, 77-89


Scollar I (1970b) Fourier transform methods for the evaluation of magnetic maps. Prospetione Archeologiche 5, 9-41


SIA (1973) Surface Approximations and contour mapping; user's guide. Applications Consultants Inc.
Magnetism. Elsevier, Amsterdam


Tabbagh A (1972) Méthode de prospection électromagnétique S.G.D. utilisation de deux sources. Prospezione Archaeologiche 7-8, 125-133 B.G.


List of computed surveys

Results from each of the following geophysical surveys have been processed at the Ancient Monuments Laboratory or the University of Surrey using the computer programs described in this thesis:

<table>
<thead>
<tr>
<th>Site</th>
<th>Survey Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cottam, Yorks</td>
<td>(pre - 1972)</td>
</tr>
<tr>
<td>Usk</td>
<td>22/73</td>
</tr>
<tr>
<td>Stenness, Orkney</td>
<td>23/73</td>
</tr>
<tr>
<td>Housesteads</td>
<td>21/75</td>
</tr>
<tr>
<td>Avebury</td>
<td>23/75</td>
</tr>
<tr>
<td>Benbecula, Hebrides</td>
<td>26/75</td>
</tr>
<tr>
<td>Orted, Surrey</td>
<td>30/75</td>
</tr>
<tr>
<td>Easton Down, Winchester</td>
<td>31/75</td>
</tr>
<tr>
<td>Cotington Power, Glos</td>
<td>3/76</td>
</tr>
<tr>
<td>Bent Farm, Congleton</td>
<td>4/76</td>
</tr>
<tr>
<td>Bullock Down, Eastbourne</td>
<td>6/76</td>
</tr>
<tr>
<td>Corbridge, Northumberland</td>
<td>7/76</td>
</tr>
<tr>
<td>Winklebury, Essex</td>
<td>9/76 (1)</td>
</tr>
<tr>
<td>Lake Farm, Wimborne</td>
<td>11/76</td>
</tr>
<tr>
<td>Wharram Percy, Yorks</td>
<td>13/76</td>
</tr>
<tr>
<td>Nancester, Warks</td>
<td>21/76</td>
</tr>
<tr>
<td>Groundwell Farm, Swindon</td>
<td>23/76 (2)</td>
</tr>
<tr>
<td>Tarraby, Carlisle, 3rd survey</td>
<td>33/76</td>
</tr>
<tr>
<td>Kenchester, Hereford</td>
<td>6/77</td>
</tr>
<tr>
<td>Alice Holt, Epsom, Hascombe</td>
<td>11/77</td>
</tr>
<tr>
<td>Edinburgh Airport</td>
<td>12/77</td>
</tr>
<tr>
<td>Elise, Chyd</td>
<td>13/77</td>
</tr>
<tr>
<td>Danby Wiche, Yorks</td>
<td>14/77</td>
</tr>
<tr>
<td>Maison Dieu, Capringe</td>
<td>15/77</td>
</tr>
<tr>
<td>Maldon Priory, Essex</td>
<td>22/77</td>
</tr>
<tr>
<td>Bagley Hall, Manchester</td>
<td>24/77</td>
</tr>
<tr>
<td>Iona</td>
<td>35/77</td>
</tr>
<tr>
<td>Catagora, Somerset</td>
<td>37/77</td>
</tr>
<tr>
<td>Stafford Castle</td>
<td>19/78</td>
</tr>
<tr>
<td>Landy Island</td>
<td>20/78</td>
</tr>
<tr>
<td>Shalford E Manor, Guildford</td>
<td>23/78</td>
</tr>
<tr>
<td>Richmond, Surrey</td>
<td>26/78</td>
</tr>
<tr>
<td>Hailes Abbey, Glos</td>
<td>28/78</td>
</tr>
<tr>
<td>Uley, Glos</td>
<td>31/78</td>
</tr>
<tr>
<td>Fountains Abbey</td>
<td>34/78</td>
</tr>
<tr>
<td>Ribchester, Lancs</td>
<td>(for University of Lancaster)</td>
</tr>
<tr>
<td>Salby, Leics</td>
<td>(for Leicester museum)</td>
</tr>
<tr>
<td>Dorchester, Dorset</td>
<td>(surveyed by Central Excavation Unit)</td>
</tr>
</tbody>
</table>

Notes:

(1) Data prepared at AM Laboratory; results plotted by CI Data Centre Ltd

(2) Data has been prepared and corrected but not yet plotted