Comp. Appl. Arch. 13, 1985.

# THE HARRIS MATRIX AS A PARTIALLY ORDERED SET

**John Haigh**

School of Archaeological and Mathematical Sciences,
University of Bradford, Richmond Road, Bradford BD7 1DP

## Introduction

Among the many areas where attempts have been made to apply computers to archaeology, one of the more frequently discussed is site stratigraphy. Orton (1980) gives a good summary of the differing viewpoints on the subject. He surmises that the relationships between the contexts of a site form what, in mathematical terminology, is a partially ordered set, or lattice.

In practice, most attempts to computerise the problem of site stratigraphy assume, explicitly or implicitly, that the relationships form a partially ordered set. If the application of computers is to be of archaeological value, then that assumption must be examined very carefully. Orton's surmise, for example, could be interpreted as indicating that a lattice is simply an alternative mathematical term for a partially ordered set. This is certainly not the case, since lattices form an important subclass within the larger class of partially ordered sets. Therefore to say that relationships between contexts form a lattice is a much more precise statement, in mathematical terms, than to say that they form a partially ordered set.

The aim of this paper is to examine in detail the correspondence between site stratigraphy and mathematical concepts and to discuss how it may be employed in the construction of computer programs.

## Partially Ordered Sets

The concept of a partially ordered set or poset is set out in a variety of mathematical texts. A classic example is provided by Maclane and Birkhoff (1967), and a possibly more readable example by Simmons (1963). The precise definition of poset varies between texts, but the following is widely accepted. A poset is a set of elements x, y, z . . . n, in which there exists a binary relation x<=y, satisfying the three rules:

```
For all x,    x<=x               (reflexivity)
If  x<=y  and  y<=x  then  x=y   (antisymmetry)
If  x<=y  and  Y<=z  then  x<=z  (transitivity)
```

The BASIC-like symbol <= is used for the binary relation as a matter of typographical convenience, but it is not necessary to associate any particular meaning to it, provided that the three rules are satisfied. For archaeological purposes, a natural meaning would be is earlier than, which in turn summarises physical relationships such as underlies or is cut by. Such an interpretation would contradict the reflexive rule, since a context cannot be earlier than itself. Therefore it is necessary to introduce an interpretation of the type is earlier than or equivalent to.

The question is then raised as to how one should interpret the = sign in the antisymmetric rule. Does = imply actual equality, in the sense of identity, or may it be treated as an equivalence such as is of the same age as or is contemporary with? The mathematical texts are not entirely specific on this point, but it appears that actual equality is implied. Thus any attempt at stratification based entirely on analogy with a poset must effectively exclude same date relationships. It is interesting to note that attempts to include same date relationships in early versions of the STRATA program (Wilcock and Bishop, 1976) met with considerable difficulties.

Although posets are widely discussed in the mathematical literature, there is never a great deal said about them. Simply because the concept of poset is a very general one, it does not have many particular properties to enumerate. Further rules must be added to the list given above, if more detailed properties are to be produced. Most of the general properties of posets have direct analogues in archaeological stratigraphy.

For instance, an element (or context) P is said to cover the element Q if $q<p$, but there is no other element X such that $q<x<p$. Here $<$ must be interpreted as the strict relationship is earlier than. When a poset is represented in a diagram, each covering relationship is shown as a downward line, but other relationships are not shown explicitly. In archaeology such a diagram is called a Harris matrix. It should be noted that the mathematical diagram of a poset does not include horizontal lines. Therefore the same date relationships shown in Figure 3.5 of Orton (1980) cannot be included if the strict analogy to the poset is to be maintained.

If there is an element A such that $a<=x$ for every element X, then A is called the least element of the poset. Likewise an element B such that $x<=b$ for every element X is called the greatest element. Natural soil or bedrock can always be regarded as the least element of an archaeological stratigraphy, and topsoil or air as the greatest.

The concept of a chain is essentially the same in both mathematical and archaeological senses. The height of an element is the length of the shortest chain connecting it to the least element. Likewise the depth of an element may be regarded as the length of the shortest chain connecting it to the greatest element. The concepts of height and depth play an important part in the arguments later in this paper.

Lattices

According to its mathematical definition, a lattice is a poset in which every pair of elements x,y has a meet and a join. The meet M or greatest lower bound of x,y has the properties:

$m<=x$, $m<=y$, but that if $w<=x$ and $w<=y$, then $w<=m$

The join 'j' or least upper bound of x,y has the properties:

$x<=j$, $y<=j$, but if $x<=w$ and $y<=w$, then $j<=w$

Lattices possess a much richer set of properties than posets in general, and at least one classic text has been devoted to them (Birkhoff 1967). The most familiar form of lattice is the outline of a diagonal garden trellis, illustrated in Figure 1, from which the mathematical use of the term lattice probably derives. It is rather a special form of mathematical lattice, with a number of properties additional to those implied by the definition above, and is known as a distributive lattice. Some of its properties will be discussed below.

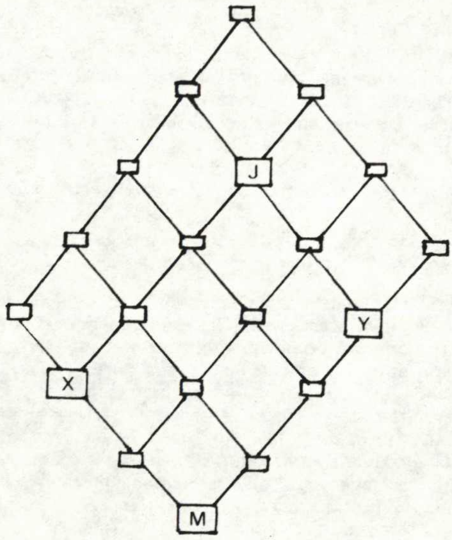


Figure 1: A simple diagonal lattice. J is the join of X and Y. M is their meet.

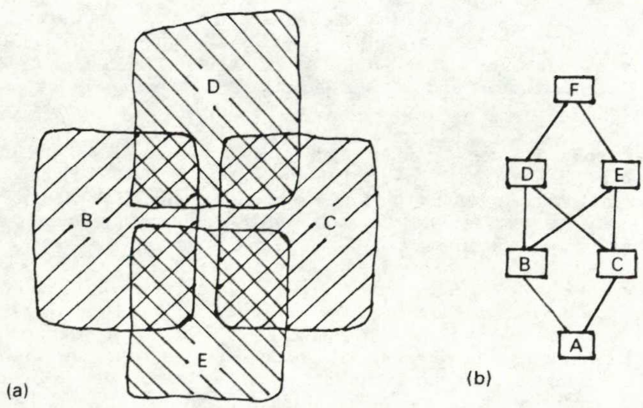


Figure 2: (a) Plan of a hypothetical site and (b) its Harris matrix.

Do relationships between archaeological contexts form a lattice, as suggested by Orton (1980)? In answer to that question consider the very simple site, illustrated in Figure 2(a), in which two contexts B and C (hatched bottom-left to top-right) overlie natural soil A; they are then overlain by contexts D and E (hatched top-left to bottom-right), which are in turn overlaid by topsoil. The Harris matrix for this hypothetical site is shown in Figure 2(b). Now B and C both underlie D, and also both underlie E, but D does not underlie E and E does not underlie D. Hence neither D nor E is the join of B and C and so, since there is no other candidate, B and C have no join. Similarly, D and E have no meet. Therefore this particular stratigraphy does not form a lattice. It is not true in general that relationships between archaeological contexts form lattices.

## The Layer Sorting Algorithm

The author was first introduced to layer sorting as a basis for the construction of a Harris matrix by Rains (1984), and later independently by Cheetham (1984). The technique is to separate contexts which do not appear as the lower member of any relationship into the first layer of the stratification, together with the relationships in which they appear as the upper member. Any contexts which do not appear as the lower member of any of the remaining relationshps are then separated into the second layer. The process is repeated for the third layer, the fourth layer, and so on, until either all the contexts and relationships have been allocated to layers, or else a number of relationships remain which cannot be allocated to further layers. The latter case indicates that there are cyclic relationships among the remaining contexts, in contradiction of the transitive rule above. It is then up to the user to determine which of the relationships are at fault, and to correct them. This rapid validation of data is one of the great advantages of the layer sorting technique.

The effect of layer sorting is that all the contexts are categorised by depth, as defined above, the depth of any context being one less than the number of the layer to which it is assigned. Since it is known that each context is related to at least one context in the layer above, it is quite simple to start with the deepest contexts and to trace the principle chains up through successive layers.

Up to this point, the calculation is relatively straightforward, and the author had little difficulty in modifying a version of the program STRATA (Bishop & Wilcock 1976; Haigh & Wilcock 1984) to accept the new algorithm. The modified program is quite efficient. Written in Microsoft FORTRAN 80, using integer arithmetic and running on a Research Machines 380Z microcomputer, it takes about 1min to produce a preliminary Harris matrix from 100 relationships, and about 15min to work on 800 relationships, including time on a slow printer.

There is a problem however, with relationships that do not connect adjacent layers. Such a multilayer relationship may be either essential, in that it uniquely defines stratigraphic information, or inessential, in that it can be compounded from other relationships by the reflexive rule. It considerably simplifies matters if the user is prepared either to dispense with all multilayer relationships, or else to accept them all. A reasonable compromise is to reject all relationships which can be made up as chains of single layer steps, but to accept all other multilayer relationships.

The point is that multilayer relationships are not really part of the layer-sorted model. To ascertain whether or not such a relationship is essential requires considerable programming effort, probably more than was involved in setting up the model in the first place. It is also expensive in computer time, although this may not be apparent when simply working with small data sets. With large and complicated data sets, however, the checks for essential relationships may take far longer than the basic sorting and chaining procedures.

Figure 3 is based on a typical output from the author's stratification program, relating to a small portion of a site at Tours in France. Although this is a comparatively small data set, containing some 85 relationships on 62 contexts, it gives rise to fairly complicated stratification. The program prints out a preliminary form for the Harris matrix, making no attempt to draw in the links between the contexts, but giving clear indications of where the links should be drawn. Each chain is drawn in a separate column, without any attempt to minimse the total width of the diagram. The cross-links 4/28, 37/16 and 67/20 each occupy a complete column of the original output, as do any other essential multilayer cross-links between different chains.

Apart from the improvement in speed, it was hoped that the new version of the program would give a rather more compact output than the older versions based directly on the program STRATA (Haigh & Wilcock 1984). Unfortunately this has not proved to be the case. A reasonably large stratification may spread over several hundred columns. With such a wide spread of results, it is extremely difficult to present the user with the particular section of information required. For this reason, little effort has yet been devoted to improving the visual presentation of the results.

## Alternative Models for the Data

It has been demonstrated above that the layer sorting algorithm does not provide a good model for multilayer relationships, this being particularly apparent for large data sets. The difficulty does not arise in the case of a diagonal lattice, as shown in Figure 1, which has the property that chains connecting the same two elements of the lattice must be equal in length. Consequently there are no multilayer relationships in the diagonal lattice, and every covering relationship links elements in adjoining layers. The absence of multilayer relationships is not a property possessed by every lattice, but only by the limited subclass called modular lattices, of which distributive lattices in turn form a smaller subclass.

It has been shown in Figure 2 that archaeological stratigraphies cannot generally be described as lattices. A fortiori it is impossible to describe them as modular lattices, so that the problem of multilayer relationships cannot be circumvented by applying such a description. Is there, then, any mathematical structure which exactly describes an archaeological stratigraphy, and whose properties can be used to create a Harris matrix? Before attempting to answer that question, one must examine what properties may be ascribed to the Harris matrix.

Figure 3 was derived from computer output with the aim of removing as many crossing lines as possible. Orton (1980) states categorically that if the relationships have been correctly expressed then there is never a need for the lines to cross, but Figure 2(b) shows the Harris matrix of a hypothetical site where the lines quite definitely cross. The author challenges anyone to show that Figure 2(b) cannot occur as the matrix of an actual site or, otherwise, to rearrange Figure 2(b) so that the crossing is removed. Figure 3, the stratification of a real site, demonstrates that Figure 2(b) is not merely a

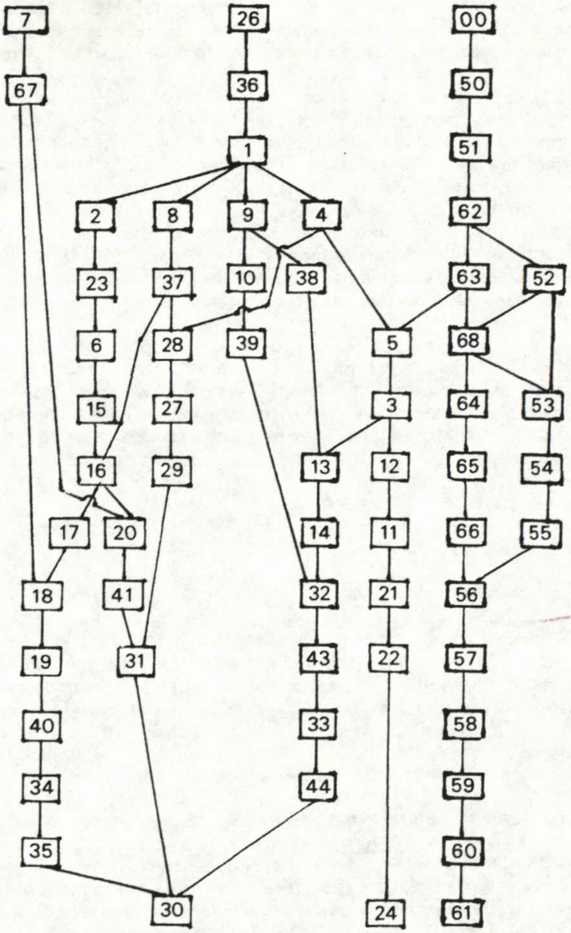


Figure 3: Harris matrix of an actual site.

hypothetical case. The situation shown in Figure 2(b) occurs three times, with the role of context B being played by contexts 18, 20, 67, and 16, of C by contexts 28, 13, 1 and 4, of D and E by contexts 28, 32, 1, and 4. Consequently Figure 3 contains three irremovable crossovers.

Figure 2(b) is often quoted in mathematical texts as a classic example of a diagram of a poset which is not also a lattice. There are also examples of lattice diagrams which contain irremovable crossovers. The idea that a Harris matrix should not contain crossovers seems to have arisen from some of the early examples, which were usually based on drawings of sections. Since a section is two-dimensional, there is no need for crossovers when representing it as a two-dimensional diagram. On the other hand, if a site is fully recorded in three dimensions, then the creation of the Harris matrix involves a compression into two dimensions. As anyone who has used multi-dimensional scaling must be aware, compressing the number of dimensions almost inevitably causes a strain among the data. In the case of the Harris matrix, the strain manifests itself in the crossovers, a given context can have several neighbours in different horizontal directions on a complex site, but in the Harris matrix only two neighbours are allowed.

Thus it becomes increasingly difficult to give a precise generic description to archaeological stratigraphy, it is not a lattice, and it is not true that there should be no crossovers in the Harris matrix. It has been shown that the layer sorting algorithm, which assumes a description no more specific than that of a poset, leads to difficulties with multilayer relationships. One cannot overlook such relationships, since they may be of crucial significance in determining the relative sequence of separate chains within the stratigraphy, a problem which archaeologists sometimes describe as floating chains. Additional light may be thrown on such problems by sorting by height, starting with the separation of the lowest layer of contexts and working upwards through the other layers. A height sort and a depth sort can together give a lot of information about floating contexts, but it is possible that the optimum arrangement archaeologically may correspond to neither of the two sorts.

Another possibility is to separate the overall stratigraphy into portions that are small enough to make a convenient calculation and perhaps to be displayed on a single screen. The problems here are that the stratigraphy cannot be divided into independent portions, there will always be some relationships connecting the different portions, and that it must be the responsibility of the user to select the contexts to be included in each calculation. A reasonable variation would be to sort the entire stratigraphy by depth, before the preliminary matrix is constructed for the selected portion, so that each context would appear in its correct layer. If the user were to make a bad selection of contexts, he could be faced with an overwhelming number of multilayer links.

A third alternative is to take note of other archaeological information beyond the basic stratigraphical relationships. For instance it may be possible to assign a feature number to each of the contexts. The term feature need not be used in any strict archaeological sense, but should rather signify a group of contexts which are expected to appear in the same sector of the Harris matrix. When faced with a choice of links during the chain-building process, the algorithm should give preference to links between contexts within the same feature. This preference will not interfere with the mathematical logic of the algorithm, but it should produce a preliminary matrix in a form which is much more satisfactory to the archaeologist. Since the majority of contexts should be grouped into their features, it would be a relatively easy task to manipulate the Harris matrix into its preferred form, with or without the aid of computer graphics.

## An alternative to the Harris matrix

At this point, it is worth asking what is the purpose of a Harris matrix. Does it have any use other than as a status symbol, whereby the importance of a site is rated by the area of wall covered by its Harris matrix? Archaeologists who have made extensive use of Harris matrices are willing to admit that the principal advantage comes during the actual preparation of the matrix, since that cannot be done without developing a sound knowledge of the site stratigraphy. The completed matrix may well be too extensive and convoluted to be conveniently used for reference. The problem of complexity is essentially the same as that noted above, but on a much larger scale, since the author has not attempted to computerise a matrix representing anything more than 20% of a typical large site.

If a Harris matrix is more useful during preparation than after completion, then there is some danger of a computer program taking away the one feature that is likely to be of advantage to an archaeologist. The problem of the Harris matrix may lie in the fact that, by linking contexts into chains, it emphasises vertical structures, whereas horizontal relationships may be of greater interest to the archaeologist. An archaeologist's primary concern is to link contexts into phases, through which it is possible to begin to work out the temporal sequences within a site.

It may therefore be more profitable to address computers directly to the problem of phasing the site, rather than to the formation of a Harris matrix. A possible means to that end is to set up a model in which the stratigraphical relationships are signified as pointers between data items representing the contexts. Certainly care would have to be taken to ensure that no cyclic relationships were incorporated into the model, and possibly that no redundant relationships were included. Such a model can be set up very easily with a symbolic program language such as PROLOG, although there may be a severe size restriction on many implementations. It should also be suitable for structured languages such as Pascal or C.

Once a model has been set up, the archaeologist could supply the computer with a list of contexts C which are suspected to belong to the same phase. The computer would then divide the remaining contexts into four classes:

(a) those that lie above C.
(b) those that are unrelated to C.
(c) those that lie between members of C.
(d) those that lie below C.

If class (c) includes contexts which definitely do not belong to C, then the archaeologist would have to remove some contexts from the list. Otherwise she or he would have the option of adding further contexts to C, in order to improve the overall division of the site. With large sites the quantity of information might be so great that there is difficulty in presenting it to the archaeologist, in which case the information would have to be restricted to contexts fairly closely related to C.

Once the archaeologist is satisfied with one phase, it is possible to ask the computer to list contexts contiguous with C, either immediately above or immediately below. Some care must be taken with regard to the definition of contiguity, since two contexts may be contiguous in the archaeological sense without one actually covering the other in the mathematical sense. The archaeologist can then build the contiguous contexts into adjoining phrases, and hence create layer by layer a phased picture of the entire site. One advantage of this approach is that the problem of using same date relationships is largely avoided. The archaeologist may choose the contexts within each layer to be as closely related or as loosely related as desired.

## Summary and Conclusions

The layer-sorting algorithm provides an efficient means to create a preliminary Harris matrix from a small set of relationships. It is able to detect very readily the presence of cyclic relationships and to isolate them fairly effectively. By assigning both a depth and a height for each context, it is capable of giving much useful information about site stratigraphy. On the other hand, when applied to a large set of relationships, it tends to produce a Harris matrix with the same wide horizontal spread that earlier methods gave, and the discrimination between multilayer relationships may be expensive in computer time.

The algorithm implicitly assumes that a stratigraphy may be described as a partially ordered set. Although such a description is not very specific mathematically, it precludes the use of same-date relationships as an inherent part of the stratigraphy. It is difficult to find any description which is both appropriate and more specific. The term lattice is certainly inappropriate.

It should be possible to improve the preliminary matrix by superimposing additional archaeological information on to the strict definition of a poset, thereby grouping contexts from the same feature of the site into the same portion of the preliminary matrix. There may be, however, an inherent weakness in the Harris matrix when applied to large sites, since the emphasis is on vertical structures or chains, rather than on horizontal structures or phases which are of archaeological interest.

Because the Harris matrix is so closely equivalent to the diagram of a poset, a radically different approach may be required for phasing information to be obtained satisfactorily. One possible approach has been described in the section above, whereby the computer allows the archaeologist to examine the relationships between a group of contexts which are seen as one phase and the remaining contexts. Such an approach should be incorporated into an interactive program, so that the user is able to see the effect of small adjustments to his suggested structure. With a large site, it will be difficult to arrange that all necessary information is presented to the user at the appropriate times.

It is important that stratigraphy should be an integral part of overall computing strategy. When site records are properly computerised, they should incorporate all relevant stratigraphical information, and the user should be able to gain access to it without detailed intervention on his part. Because of the conceptual complexities of stratigraphy, archaeologists will achieve little overall gain from computerisation if it involves them in long-winded input routines or in the preparation of complicated data files.

## Acknowledgements

The author gratefully acknowledges the helpful suggestions and constructive criticism of Mike Kelly and Paul Cheetham, who have helped to formulate some of the ideas expressed in this paper, but willingly accepts responsibility for any

faults or misinterpretations.

References

BIRKHOFF, G. 1966 Lattice theory (3rd ed.). American Mathematical Society, Providence, Rhode Island.

BISHOP, S. & WILCOCK, J.D. 1976 Archaeological context sorting by computer: the STRATA program. Science and Archaeology 17: 3-12.

CHEETHAM, P. 1984 Unpublished BSc dissertation, University of Bradford.

HAIGH, J.G.B. & WILCOCK, J.D. 1983 Some further developments of the STRATA system. Computer Applications in Archaeology 11: 189-90.

HARRIS, E.C. 1979 Principles of Archaeological Stratigraphy. Academic Press, London.

MACLANE, S. & BIRKHOFF, G. 1967 Algebra. MacMillan, New York.

ORTON, C.R. 1982 Mathematics in Archaeology. Cambridge University Press.

RAINS, M.J. 1984 Home Computers in Archaeology. Computer Applications in Archaeology 12: 15-26. University of Birmingham Press.

SIMMONS, G.F. 1963 Topology and Modern Analysis. McGraw-Hill/Kogajusha, Tokyo.