Undecidability and Nonperiodicity
for Tilings of the Plane

RAPHAEL M. ROBINSON (Berkeley)

§1. Introduction

This paper is related to the work of Hao Wang and others growing out of a problem which he proposed in [8], § 4.1. Suppose that we are given a finite set of unit squares with colored edges, placed with their edges horizontal and vertical. We are interested in tiling the plane with copies of these tiles obtained by translation only. The tiles are to be placed with their vertices at lattice points, and abutting edges must have the same color. Wang raised the question whether there is a general method of deciding which finite sets of colored squares can be used to tile the plane in this way. He also discussed the relation of this problem to the decision problem for certain classes of formulas of the predicate calculus, but we shall consider only the geometrical problem here.

Suppose that we have a tiling of the plane of this type which has a horizontal period. That is, we assume that the tiling remains invariant under a certain horizontal translation. There will then be a vertical strip which can be repeated to cover the plane. This strip has only a finite number of different horizontal cross sections, and hence has two which are alike. Thus the same tiles may be used to construct a tiling which has a vertical period as well as a horizontal period.

A similar argument can be used even when the given period is not horizontal. That is, if a set of tiles permits a periodic tiling, then it also permits a doubly periodic tiling. In any such tiling, we can find equal horizontal and vertical periods, and hence can find a square of some size which repeats to cover the plane.

Wang made the conjecture, since proved false, that any set of tiles which permits a tiling of the plane also permits a periodic tiling. He pointed out that if this conjecture were true, then we would have a decision method for an arbitrary set of tiles. Indeed, it would be sufficient to form all possible squares from the given set of tiles, starting with the smaller squares and working up, until we either reach a square which can be repeated periodically, or we find a square of a certain size which cannot be tiled at all. The latter will always happen if tiling of the whole plane

is impossible. Another formulation of the result just proved is that if the tiling problem is undecidable, then there must be finite sets of tiles which permit only nonperiodic tilings of the plane.

In the preceding paragraph, we used the fact that if squares of all sizes can be covered with the given tiles, then the whole plane can be tiled. This is an easy application of a method used by König [5]. Start with any tile which allows extension of the tiling arbitrarily far in all directions. Of the various tiles which we can adjoin to this tile along a prescribed edge, choose one for which the pair admits such extensions. Continue adding tiles one at a time in some convenient pattern, being sure at each step that the portion of the tiling thus formed admits extension in all directions as far as we please. In this way, we obtain a tiling of the whole plane.

It follows in particular that a given set of tiles can be used to tile the whole plane if and only if it can be used to tile the first quadrant.

Various constraints on the tiling of the first quadrant were considered, in which either the tile at the origin, the tiles along one axis, or the tiles along the diagonal were restricted. That is, only a certain tile was allowed at the origin, or only tiles from a certain subset were allowed along the axis or diagonal. The origin-constrained problem is the easiest to settle, and was proved undecidable by Wang, as explained in [9]. A closely related result was obtained independently by Büchi [2], and was applied to the decision problem for the predicate calculus. The result which was particularly needed in this connection, the undecidability of the diagonal-constrained problem, was proved by Kahr, Moore, and Wang [4]. The proof of the undecidability of the row-constrained problem is similar, but slightly simpler. A description of the development of the subject from 1960 to 1962 is given by Wang [9].

It should perhaps be mentioned that with the constraint that a particular tile is to be used at the origin, it is no longer true that any set of tiles which can be used to tile the first quadrant can also be used to tile the whole plane. However, the proof that there is no decision method is similar in the two cases.

The problem without any constraints is no doubt the one of greatest geometrical interest. It remained unsolved for several years, and was then proved undecidable by Robert Berger [1]. As noted earlier, this means that there must be some set of tiles which can cover the plane, but cannot cover the plane periodically. Actually, Berger found it necessary to construct such a set in the course of his proof. This set contains over twenty thousand tiles, though Berger points out how this number can be considerably reduced.

After reading Berger's paper, I became interested in the problem of reducing this number as far as possible. How far this reduction can be

carried depends on exactly what rules are used. We first discuss some alternative rules, and then describe the results which were obtained.

Wang imposed the condition that abutting tiles should have edges of the same color. An equivalent formulation is the following. Let all colors be arranged in complementary pairs, and insist that abutting edges should have complementary colors. The tiling problem with this condition is exactly equivalent to Wang's, since we need only replace the colors of the right and upper edges of each tile by the complementary colors. However, the two problems are no longer equivalent when rotation is allowed, and indeed the problem with matching colors becomes trivial, since copies of any one tile can be used to cover the plane, but the problem with complementary colors does not.

The tiling problem with complementary colors remains undecidable when translation and rotation are allowed. Indeed, let any finite set of square tiles with colored edges be given with which we wish to tile the plane, using translation only and insisting on matching colors. We may suppose that the horizontal colors are different from the vertical colors. Change the right and upper edges to the complementary colors, where the complementary colors are new colors. Then the color of any edge identifies the position of the edge (left, right, bottom, top), so that rotation is useless. Hence the desired tiling is possible if and only if the plane can be tiled with these new tiles so that complementary colors abut, translation and rotation being allowed. By Berger's theorem, there cannot be a decision method for this problem.

Still another variant of Wang's problem is obtained if we notch the edges in such a way that the bumps and dents on an edge fit the dents and bumps on an abutting edge. Here we insist that the tiles should cover the plane without overlapping. The notches can be chosen so that they force alignment of the squares. Assuming that there are no bumps or dents at the corners, this problem is equivalent to the problem with matching colors or with complementary colors if only translation is used, or to the problem with complementary colors if translation and rotation are used. However, if reflection is also allowed, the problem with complementary colors is no longer the same as the problem with notches, since the latter allows an unsymmetrical edge whereas the former does not.

The tiling problem with notched edges remains undecidable even when translation, rotation, and reflection are allowed. Indeed, let any set of square tiles with notched edges be given, and suppose that we want to tile the plane using translation only. It is easy to add notches which do not affect the fit of tiles obtained by translation, but which make rotation or reflection useless. Thus the undecidability of the new problem is reduced to Berger's theorem. A somewhat more elaborate

13*

argument shows that the tiling problem with complementary colors also remains undecidable when translation, rotation, and reflection are allowed.

In 1967, I found a set of seven square tiles with notched edges which permit only nonperiodic tilings of the plane when translation, rotation, and reflection of the tiles are allowed. This result was announced in [7]. Reflecting these seven tiles produces a set of 13 tiles which force non-periodicity when translation and rotation are used. Rotating these 13 tiles produces a set of 52 tiles which force nonperiodicity when only translation is used. Thus the number of tiles needed for Berger's non-periodicity theorem was reduced to 52.

In 1969, I found a set of six polygonal tiles which force nonperiodicity when translation, rotation, and reflection are used. These are shown in Fig. 1. They have both advantages and disadvantages as compared to
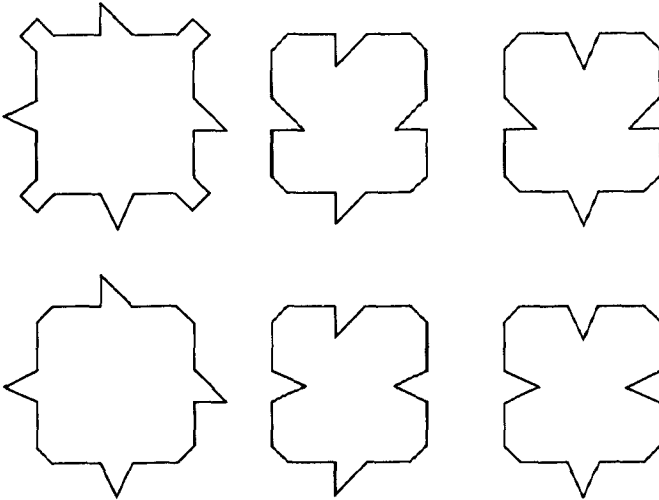

Fig. 1. Six tiles which force nonperiodicity

the previous set. As before, they are notched squares, but here corner bumps and dents are also used, which makes a transformation into tiles with colored edges difficult. In any case, this is the smallest set of tiles which I have been able to find which can tile the plane using translation, rotation, and reflection, but for which no such tiling is periodic, or admits a translation onto itself. It will also turn out that no such tiling admits a rotation onto itself. These six tiles will be studied in this paper.

One fact may be noted at once. As is easily seen, if the plane is tiled with these notched squares, then the unnotched squares also tile the plane, the squares being arranged in rows and columns.

The two tiles in the first column of Fig. 1 are symmetric to a diagonal, and the two in the last column are symmetric to a vertical line. Thus if we adjoin the mirror images of the other two tiles, we obtain a set of eight tiles which force nonperiodicity when translation and rotation are used. These in turn lead to a set of 32 tiles which force nonperiodicity when only translation is used. However, because of the corner markings, these cannot be thought of as tiles of Wang's type.

A closely related set of ten notched squares without any corner bumps or dents is also shown to force nonperiodicity when translation, rotation, and reflection are used. By reflection and rotation, these lead to a set of 56 tiles which can be thought of as squares with colored edges and which force nonperiodicity when only translation is used. Although this estimate is not quite as good as the one announced in [7], it is all that will be proved in this paper about this problem.

By a more detailed analysis of the tiles studied in this paper, I have succeeded in reducing the number of tiles needed for Berger's non-periodicity theorem to 35. This analysis also justifies the seven tiles announced in [7], and has led to another set of six polygonal tiles which force nonperiodicity. In this case, five of the six tiles are notched squares, and the sixth is small, being chosen so that it can fill in the gap when two abutting notched squares both have dents. In a different direction, I have shown how Berger's undecidability theorem may be used to prove that there is no decision method for the problem of tiling the plane with copies of an arbitrary finite set of convex hexagons or convex pentagons whose vertices lie at lattice points, even when translation, rotation, and reflection are allowed. I hope to publish these various results later.

A problem which I do not know how to solve is whether similar results can be found for the hyperbolic plane. Is there a general method of deciding whether a set of polygonal tiles can be used to tile the hyperbolic plane, using arbitrary motions (or motions and reflections)? Can we find a set of polygonal tiles which can tile the hyperbolic plane, but for which no such tiling admits a translation onto itself? It is also not clear that there is any longer a connection between the two questions.

This paper does not assume any knowledge of tiling problems. The undecidability proofs require some knowledge of Turing machines, but the portions of the paper concerned with nonperiodicity do not. The three main objectives of the paper, and the relevant sections, are as follows.

(a) Proof that the six polygonal tiles of Fig. 1 allow only nonperiodic tilings of the plane (§§ 2, 3). The same result is obtained for the closely related set of ten tiles mentioned above.

(b) Proof that there is a set of 36 square tiles with colored edges for which the completion problem is undecidable, translation only being used (§§ 4, 5, 6). By the completion problem is meant the problem of

deciding whether an arbitrary finite partial tiling of the plane can be completed.

(c) Simplified proof of Berger's theorem that the plane tiling problem is undecidable (§§ 2, 3, 4, 7).

Notice that (a) and (b) are completely independent, but that (c) uses (a) and part of (b). Some supplementary information on the tilings is given in §8. In particular, the first part of §8 gives a description of all possible tilings using the tiles of Fig. 1, and may be read immediately after §3.

§2. The Five Basic Tiles and Their Modifications

We shall show how the six tiles of Fig. 1, and another set of ten tiles which were previously mentioned and will be important in later sections, can both be obtained as modifications of a certain basic set of five tiles. These are obtained by deleting the corner markings from the tiles in Fig. 1, thus making all of the corners square. The two tiles in the first column then become identical, and we obtain the five basic tiles.

We shall study tilings of the plane by copies of the five basic tiles obtained by translation, rotation, and reflection, subject to certain constraints. As was the case for the six tiles of Fig. 1, any tiling of the plane by these notched squares forces alignment of the unnotched squares. We may also note that three of the basic tiles have an axis of symmetry. Hence the tiles which can be obtained from the five basic tiles by translation, rotation, and reflection, can be obtained from just seven tiles by translation and rotation, or from 28 tiles by translation alone.

The five basic tiles may be represented symbolically as in Fig. 2, where the arrow heads represent bumps and the tails represent dents.
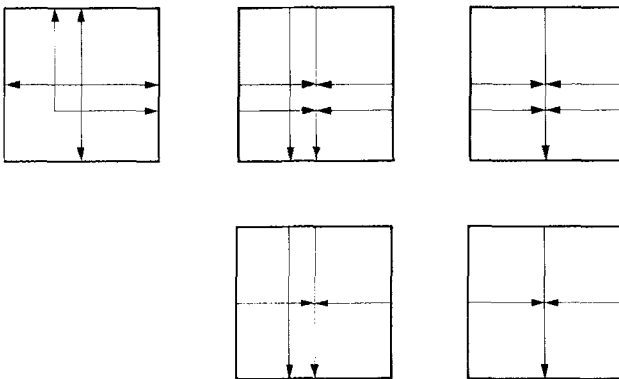

Fig. 2. The five basic tiles

In fitting the tiles together, arrow heads must meet arrow tails. Notice that the symmetrical bumps and dents are represented by central arrows. The unsymmetrical bumps and dents are represented by double arrows, one central and one to the side. Alternatively, we could use just the side arrow, but the double arrow proves more convenient.

The first tile in Fig. 2, which has arrow heads on all four sides, will be called a cross. As drawn, the cross is said to face up and to the right. The other basic tiles will be called arms. Every arm has a principal arrow, the central arrow which runs across the tile from one side to the opposite side. An arm is said to point in the direction of its principal arrow. Every arm also has central in arrows at right angles to the principal arrow. If there are side in arrows as well, then they are toward the head of the principal arrow. This fact will be essential for determining the orientation of arms.

The abbreviated symbolism of Fig. 3 will also be used. The first square represents a cross with unspecified orientation. The second square represents any arm whose principal arrow is as shown.
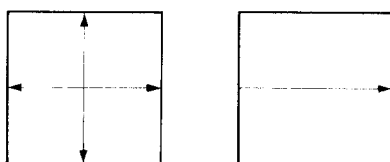

Fig. 3. Abbreviated notation for cross and arm

On the other hand, starting with the six tiles of Fig. 1, and deleting the side markings, we obtain just the two polygons in Fig. 4, which we may describe as a bumpy square and a dented square. If the plane is
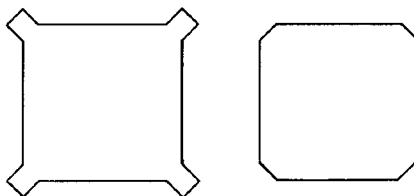

Fig. 4. Bumpy and dented squares

tiled with these squares, then one bumpy square and three dented squares must meet at each corner. No other restriction is imposed. One possibility is that we have a completely regular pattern, with the bumpy squares lying in alternate columns and in alternate rows. If the columns and rows are suitably numbered, then these will be the odd-numbered

columns and rows, or what we may call the odd-odd positions. Besides the regular pattern, there is the possibility of shifting some of the rows containing bumpy squares one unit to the right. This gives a possible pattern in every case. In a similar way, some of the columns could be shifted, instead of some of the rows. Is there any other possibility?

Notice that if one of two adjacent rows of squares consists of alternate bumpy and dented squares, then the other must consist solely of dented squares, and vice versa. Thus if we have any row consisting of alternate bumpy and dented squares, then alternate rows must be of this type. There are two choices for the position of each of these rows. Similarly, one column in which bumpy squares occur in alternate positions force bumpy squares to occur in alternate positions in alternate columns.

Suppose now that we have any pattern which is not the completely regular pattern first mentioned. We can then find three consecutive squares in some row or column of which the first is bumpy and the other two dented. Let these three squares be placed in the center column of Fig. 5, with the bumpy square at the bottom. Then the two bumpy
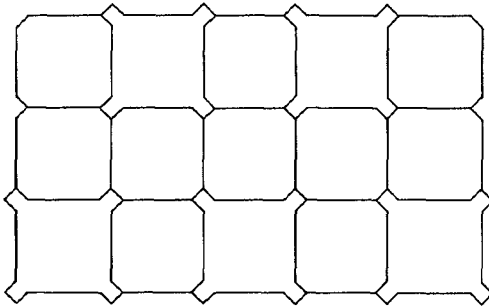

Fig. 5. A pattern of bumpy and dented squares

squares at the top are forced, and these in turn force the outer bumpy squares at the bottom, and so forth. Thus bumpy squares must occur in alternate positions in both the bottom and top rows. As noticed earlier, this forces bumpy squares to occur in alternate positions in alternate rows throughout the plane. There is no other possibility, except to interchange rows and columns.

If these corner markings are combined with the five basic tiles, then the effect is to force the cross to occur in alternate positions in alternate rows, or in alternate positions in alternate columns. The cross may occur in other positions as well. Thus tiling the plane with the six tiles of Fig. 1 is equivalent to tiling the plane with the five basic tiles subject to this constraint.

An alternative to the corner markings is furnished by the parity markings shown in Fig. 6. The arrows here are in a different position than for the basic tiles. They may be understood as a symbolic notation for bumps and dents at new locations. All of the parity tiles are symmetric to vertical and horizontal axes. The tile at the upper left can be obtained from the tile at the lower right by rotating through 90°, or by reflection in a diagonal. Thus there are really just three different parity tiles.
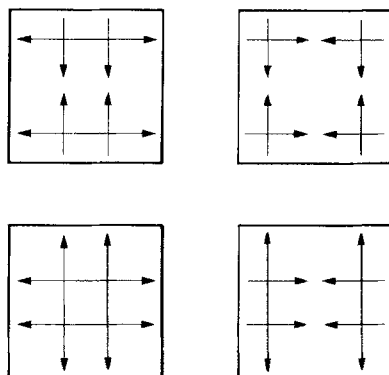

Fig. 6. Parity tiles

If the plane is tiled with parity tiles, then these must alternate both horizontally and vertically in the order shown in Fig. 6. By a suitable numbering of the columns and rows, the lower left tile will occur in just the odd-odd positions.

Parity markings will be added to the five basic tiles as follows. The cross will be combined with the parity tile at the lower left in Fig. 6. Vertical arms will be combined with the parity tile at the lower right. Equivalently, horizontal arms will be combined with the parity tile at the upper left. All of the basic tiles will be combined with the parity tile at the upper right. This gives a total of ten basic tiles with parity markings.

Use of these ten tiles forces the cross to occur in alternate columns and in alternate rows, say in the odd-odd positions. No further restriction is imposed, since if the odd-odd positions are filled with crosses, then the even-odd positions must be filled with vertical arms and the odd-even positions with horizontal arms, and suitable parity markings were allowed for this. Also, the parity markings allow all of the five basic tiles in the even-even positions, no matter what the orientation.

These ten tiles act just like the six tiles of Fig. 1, except for a slightly stronger restriction on the occurrence of crosses. The advantage is that no corner markings are used. Since the ten tiles retain all of the symmetry

of the five basic tiles, they give rise to just 14 tiles by reflection, or 56 by reflection and rotation. The results of §3 show that these 56 tiles force nonperiodicity for tilings obtained by translation alone. This yields the estimate 56 for the number of tiles needed in Berger's nonperiodicity theorem.

§3. Nonperiodic Tilings

We shall show in this section that copies of the five basic tiles obtained by translation, rotation, and reflection can be used to tile the plane in such a way that the cross appears in alternate positions in alternate rows, and perhaps elsewhere, but that no such tiling is periodic. This establishes at the same time the nonperiodicity of tilings by the six tiles with corner markings and by the ten tiles with parity markings.

I have made a complete analysis of tilings using the five basic tiles, but this cannot be included in this paper. In particular, it turns out that without any constraint on the appearances of the cross, periodic tilings are possible.

A few results about the five basic tiles without constraints will be proved here. Consider any tiling of the plane by the five basic tiles. It is important to analyze what happens in some direction from a cross, say to the right, before another cross is reached. We have a sequence of arms, of which at most one can be vertical, since a vertical arm forces the tiles to its left and right, if arms, to point toward it. Thus, in general, the cross will be followed by a finite sequence of right arms, a vertical arm, a finite sequence of left arms, and then the next cross. Either of the finite sequences may be empty. An exception to this pattern would occur if either of the sequences were infinite; in this case, the subsequent portion would be missing.

We can see that two consecutive crosses in the same row or column must either face each other or be back to back. Suppose, for example, that we are looking at two consecutive crosses in the same row, and that the first cross faces up and to the right. Then the next vertical arm must point down, since there cannot be in arrows near the tail of the principal arrow, and the next cross must face up and to the left. A possible pattern is shown in Fig. 7. In any case, two facing crosses must be mirror images. If the two crosses are not face to face, then they must be back to back. In this case, the vertical arm may point either up or down, and
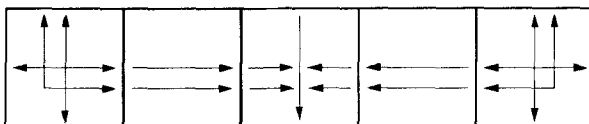

Fig. 7. Facing crosses

the second cross may be the mirror image of the first, or it may be inverted. A possible pattern is shown in Fig. 8.
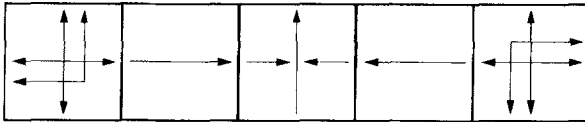

Fig. 8. Back-to-back crosses

In the last two figures, we have used a symbolism intermediate between the complete notation of Fig. 2 and the abbreviated notation of Fig. 3, all horizontal arrows having been drawn. Notice that if the arms had been drawn in the abbreviated form, the extra arrows could have been supplied mentally.

We can show that the distance between facing crosses, measured center to center, must be even. Suppose, for example, that the crosses are in the same row, and face up. Between them, there are horizontal arms and one down arm. Since these tiles have arrow tails at the top, the tiles immediately above them must have arrow heads at the bottom. Hence they must be alternately crosses and down arms. But a similar argument applies to columns, so we must begin and end with a cross, as in Fig. 9. In particular, a cross facing up and to the right forces another cross to its upper right. The evenness of the distance between facing crosses is now apparent. However, if the crosses are back to back, we cannot prove that the distance between them must be even.
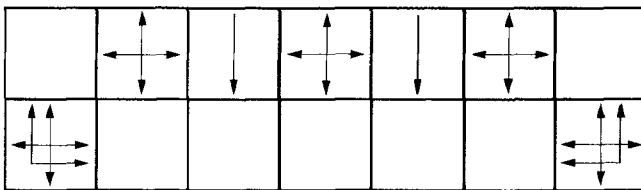
Fig. 9. Pattern forced by facing crosses

We shall now impose on the five basic tiles the constraint that crosses appear in alternate positions in alternate rows. These crosses are the bumpy crosses or the odd-odd crosses, depending on whether we are using corner markings or parity markings. Any such cross will be called a 1-square.

If we start with any 1-square, then it will face another cross two units to its right or left. This pair of crosses and the intermediate arm will constitute the bottom or top row of Fig. 10. The central cross is
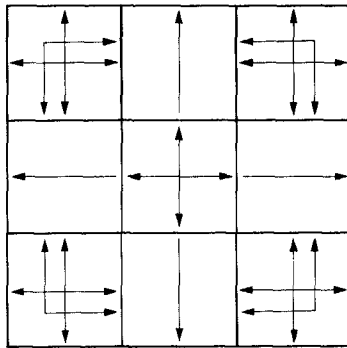
Fig. 10. A 3-square

forced, and prevents slippage of the last row, so the 3-square shown in Fig. 10 must be completed. The orientation of the central cross is unknown. The 3-square is determined by the 1-square in any one of its four corners.

We can extend any 3-square in the directions faced by its central cross. Suppose, for example, that the central cross faces up and to the right. The 3-square may be taken at the lower left of Fig. 11. The central
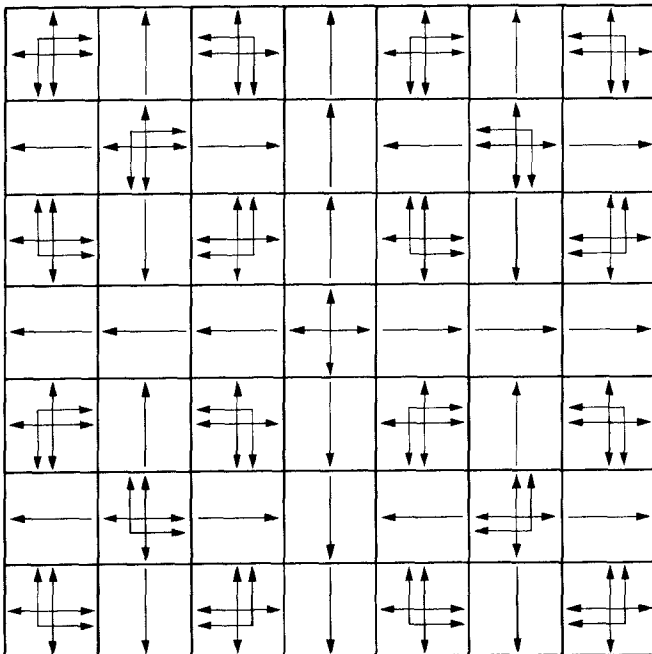
Fig. 11. A 7-square

cross of this 3-square forces the arms extending to the left and down from the central cross of the 7-square, and these in turn force the central cross of the 7-square to be there. This cross prevents any slippage of the row above; that is, the crosses must occur in the positions shown. We then see that the crosses to the upper left, upper right, and lower right of the central cross of the 7-square must have the orientations shown. Each of these crosses forces a 3-square, so the 7-square must be completed. It is determined, except for the orientation of the central cross, by the 3-square in any one of its four corners.

In a similar way, any 7-square can be extended in the directions faced by its central cross to a 15-square, consisting of a central cross, four lines of arms radiating from it, and 7-squares in the four corners whose central crosses face each other. Of the arms flanking the given 7-square and radiating from the central cross of the 15-square, the place to start the argument is with the third arms out. These are forced by the central cross of the 7-square. They in turn force the second arms, then the first arms, and finally the central cross of the 15-square. This cross prevents slippage of the row above. The central cross of the 15-square is thus surrounded diagonally by crosses, which must all be back to back. Each of the three new crosses forces a 7-square, so the 15-square must be completed. The 15-square, in turn, can be extended in the directions faced by its central cross to a 31-square, and so forth.

In any tiling of the plane by the five basic tiles subject to the constraint which we imposed, every 1-square uniquely determines the 3-square, 7-square, 15-square, 31-square, etc., in which it lies. The union of this expanding sequence of squares is either a quarter plane, a half plane, or the whole plane, depending on the successive orientations of the central crosses. No two of these union figures can overlap. In any tiling, the union figures so obtained must consist of either (a) the whole plane, (b) two half planes, (c) a half plane and two quarter planes, or (d) four quarter planes. Adjacent half or quarter planes must be separated by a corridor consisting of a single row or column of arms. The corridor separating two half planes may be a fault; that is, the two half planes need not be symmetric to it. However, if one or both of the half planes are divided into quarter planes, no additional fault is possible.

Since $(2^n - 1)$-squares can be constructed for every n, it follows that the plane can be tiled. (A description of all possible tilings is given in §8. In finding this description, no use is made of §§4–7.) However, no such tiling is periodic, since, for every positive integer n, there exist crosses which face other crosses at the distance 2^n, namely the central crosses of $(2^n - 1)$-squares.

In a certain sense, the tilings are almost periodic. Indeed, ignoring fault lines, the 1-squares repeat horizontally and vertically with period 4,

the 3-squares with period 8, the 7-squares with period 16, etc. Thus at least $[(2^n - 1)/2^n]^2$ of the tiles repeat horizontally and vertically with the period 2^{n+1}. Hence we can choose a period large enough so that an arbitrarily small fraction of the tiles fail to repeat.

§4. Turing Machines and Tiling

The next three sections are independent of the last two. However, §§ 2–3 as well as this section are needed in §7. We return now to Wang's tiling problem, using squares whose colored edges must match, and using translation only.

The undecidability of the tiling problem is proved in §7. Here we consider the easier completion problem. Suppose that a finite portion of the tiling is given. Can the tiling be completed? The completion problem does not seem to have been studied previously, except for the case where just one tile has been laid down. In this case, we simply impose the constraint that a particular tile must be used at least once. This makes the proof of undecidability much simpler. It was essentially this problem, the origin-constrained problem, which was first solved by Wang, as explained in [9]. The problem is considered there for a quadrant, but the method is the same in either case. We shall present the proof in full here, since it is needed for the further development.

The undecidability of this problem is based on the undecidability of the halting problem for Turing machines. We shall use the following description of Turing machines. The tape will be infinite in both directions. There will be a finite number of states $q_0, q_1, \ldots,$ of which q_0 is the initial state, and a finite number of symbols $s_0, s_1, \ldots,$ of which s_0 is the blank. At any instant, the reading and writing head is scanning one square of the tape. The action of the machine will be determined by some quintuples of the forms

$$q_i \, s_j \, s_k \, L \, q_l, \qquad q_i \, s_j \, s_k \, R \, q_l,$$

which indicate that if the machine is in state q_i and scanning symbol s_j, it will overprint s_k, move left or right, and go into state q_l. No two quintuples start with the same $q_i \, s_j$. If the machine is in state q_i and scanning symbol s_j, and there is no quintuple starting $q_i \, s_j$, then the machine halts. This formulation of Turing machines is almost that originally used by Turing, and agrees with that used by Minsky [6]. In some other accounts, the tape is finite to the left, or overprinting and left or right moves are done separately.

We shall use two undecidable halting problems, both discussed by Minsky [6], Chapter 8. (a) There is no method for deciding whether a given universal Turing machine will halt when started on an arbitrary

tape. (b) There is no method for deciding whether an arbitrary Turing machine will halt when started on a blank tape.

We must describe how the operation of a Turing machine will be related to tilings of the plane. The tape of the Turing machine will be represented horizontally and time vertically, increasing upward. The complete configuration (tape with scanned square, and state) of the Turing machine at two consecutive instants will be represented on the lower and upper edges of a row of tiles. In this, we follow Berger [1], except that there time increased downward. In contrast, Wang [9] represented the complete configurations by rows of tiles. However, Berger's method seems simpler.

The coloring of the edges of the tiles will be indicated by labelled arrows. The colors will be associated with the label and with the direction of the arrow (rather than with the heads or tails), so that matching colors are indicated by an arrow head abutting a tail with the same label. The arrows will point up, left, or right. The tiles may be thought of as transmitting signals in the indicated direction.

The operation of the machine will be represented by tiles of several sorts. The alphabet tile in Fig. 12 transmits the symbol s_k unchanged. It is used for all k. The merging tiles in Fig. 13 combine a state q_i with a symbol s_j. They may be allowed for all values of i and j, though some may not be needed. The first or second action tile in Fig. 14 is used only if the quintuple $q_i\, s_j\, s_k\, L\, q_l$ or $q_i\, s_j\, s_k\, R\, q_l$ is present.
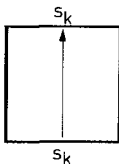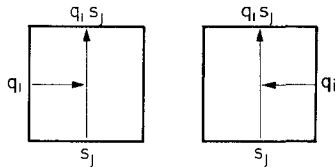

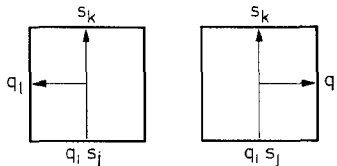Fig. 12. Alphabet tile   Fig. 13. Merging tiles

Fig. 14. Action tiles

Suppose that we are given a row of tiles whose upper edges represent the complete configuration of the Turing machine at time t. One edge will have an up arrow labelled $q_i\, s_j$, and the others will have up arrows with labels of the form s_k. If we use tiles of the types shown in Figs. 12–14,

then the row of tiles above the given row will be uniquely determined, and its upper edge will represent the complete configuration at time $t+1$. Of course, the tiling will be impossible if the machine halts at time t. If all the merging tiles are provided, then this is the only exceptional case. However, if we wish to use as few merging tiles as possible, then we will omit the ones leading to a halting state-symbol pair $q_i s_j$, and the tiling of the next row will also be impossible if the machine halts at time $t+1$.

Assume that the Turing machine is started on a blank tape. We may use the three tiles in Fig. 15 to represent the starting configuration. The second tile is the one which is to be laid down initially. It forces an
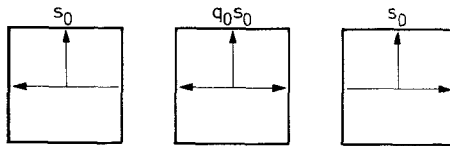
Fig. 15. Starting tiles for blank tape

infinite sequence of copies of the first tile to appear at its left, and an infinite sequence of copies of the third tile to appear at its right. The upper edge of this row of tiles represents the complete configuration of the machine starting on a blank tape. In addition, a blank tile is provided with which the lower half plane can be tiled. (It must be noted that the blank symbol on the tape is not represented by the blank edge of a tile.)

We see that the tiling of the plane can be completed with the tiles mentioned, after the initial tile has been laid down, if and only if the Turing machine never halts. Thus there is no method of deciding whether an arbitrary set of tiles can be used to tile the plane, when the constraint is imposed that a particular tile must be used at least once. A similar argument can be given for a quadrant, using a Turing machine with a tape infinite only to the right. As mentioned above, this was done by Wang.

In this problem, we cannot put any bound on the number of different tiles used, since to do so would allow only a finite number of different sets of tiles, except for renaming the colors, and hence decidability would be trivial. But for the general completion problem, this is not the case. Indeed, we shall find in §6 a set of 36 tiles for which the completion problem is undecidable, even when we restrict the given portion of the tiling to consist of a finite number of tiles arranged in a horizontal row.

We shall limit ourselves here to proving the existence of a fixed finite set of tiles for which the completion problem is undecidable. Let any universal Turing machine be given. There will be some symbol s_h

such that there is no method of deciding whether the machine will halt when started on a tape which is arbitrary except that the scanned symbol is s_h. Besides the tiles of Figs. 12–14 which correspond to this machine, we allow a blank tile and the five tiles of Fig. 16.
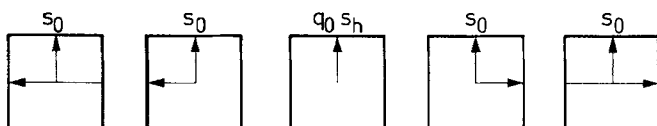
Fig. 16. General starting tiles

Corresponding to any initial tape of the Turing machine with s_h in the scanned square, the given portion of the tiling is determined as follows. A segment of the initial tape containing all of the printed squares is chosen, the corresponding alphabet tiles are lined up, the tile corresponding to the scanned square is replaced by the third tile in Fig. 16, and the second and fourth tiles in Fig. 16 are placed at the left and right ends of the row.

The leftmost tile then forces an infinite sequence of copies of the first tile in Fig. 16 to appear to its left, and the rightmost tile forces an infinite sequence of copies of the fifth tile in Fig. 16 to appear to its right. The upper edge of this row of tiles represents the initial complete configuration of the Turing machine. The tiling of the upper half plane can be completed if and only if the Turing machine never halts. The tiling of the lower half plane is trivially possible, using only blank tiles and alphabet tiles. Thus we have found a finite set of tiles for which the completion problem is undecidable, even if we use only a finite row of tiles as the initial position.

§ 5. A Generalized Turing Machine

In this section, we shall discuss a type of generalized Turing machine. A particular machine of this type will be constructed, and will be applied in § 6 to estimate the number of tiles needed to make the completion problem be undecidable. The results of these two sections will not be used elsewhere in the paper, so the reader may proceed directly to § 7 if he prefers.

A Turing machine is usually considered to require one unit of time to complete each move. It will be more convenient for us to use instant action instead. By this, we mean that all action of the Turing machine which consists of motion of the head in one direction along the tape is to happen instantaneously. On the other hand, whenever there is a change in direction, one unit of time is needed. This does not affect the

manner in which the machine operates, but merely means that we are keeping time with a peculiar clock. If the head ultimately makes an infinite sequence of moves in the same direction, then with instant action the operation of the machine is completed in a finite time, even though the machine does not halt.

We shall now introduce a type of generalized Turing machine with several heads, each acting independently. Furthermore, we agree that a head, when in a certain state and scanning a certain symbol, may disappear, or may split into two heads, one of which moves to the left and the other to the right. We also agree that each head uses instant action. With this degree of generality, some convention about colliding heads would be needed. However, the machine which we actually construct will be much more special. There will be just one permanent head. From time to time, it will split, but one of the resulting heads will be transient; that is, it will perform a certain instant action and then disappear instantaneously. Thus collision of heads will be impossible.

In describing the machine, it will be convenient to agree that each state can be entered only by a right move, or only by a left move. Thus the states can be classified as right states $R0, R1, \ldots$, and left states $L0, L1, \ldots$. Naming the new state will automatically tell the direction of motion on the move. We shall construct such a machine with an undecidable halting problem. This machine will use four symbols and seven states, of which four are right states and three left states. It bears a strong resemblance to the 4-symbol 7-state universal Turing machine constructed by Minsky. A description of that machine may be found in [6], § 14.8.

Our machine, like Minsky's, will represent the action of a system of Post tag with deletion number 2. In such a tag system, we are given a finite alphabet, and a finite number of Post normal productions of a special kind. For each production, the transform of any word of two or more letters which starts with a prescribed letter is obtained by deleting the first two letters, and adding certain letters at the end. There is at most one production for each initial letter. For this type of production, we shall use the abbreviated notation

$$a \to a_1 a_2 \ldots a_n$$

to indicate that if the first letter of a word is a, then we may apply this production to the word, and it will have the effect of deleting the first two letters and adjoining $a_1 a_2 \ldots a_n$ at the end. For example, the production $a \to a$ applied to abc yields ca.

Suppose that any word of the tag alphabet is given. We apply the given productions as long as possible. We must stop if the initial letter of the word is a halting letter, that is, a letter to which no production

corresponds, or if the length of the word is less than 2. The last word so obtained is considered as the answer. If the productions may be applied indefinitely, then no answer is obtained.

It was shown by Cocke and Minsky [3] that the action of a Turing machine using two symbols may be translated into the action of a Post tag system with deletion number 2. The proof may also be found in Minsky [6], §14.6. The number of letters added at the end varies from 1 to 4. In their construction, the tag word always has length at least 4, so that halting can be caused only by halting letters. The number of halting letters may be reduced to one, if we like. If we start with a universal Turing machine, we obtain a universal tag system. Minsky's 4-symbol 7-state universal Turing machine was obtained by a retranslation of this tag system.

It is easily seen that we may modify the tag system so as to eliminate the halting letters by adding some new productions, and that this may be done in such a way that the halting problem remains undecidable. That is, there will be no way of deciding whether the process will terminate when we apply the productions to an arbitrary tag word. Indeed, in the Cocke-Minsky proof, each complete configuration of the given Turing machine is made to correspond to a word in a four-letter alphabet, a different alphabet being used for each state-symbol pair $q_i s_j$. Suppose that a certain $q_i s_j$ causes the Turing machine to halt. Let the corresponding alphabet for tag words be A, a, B, b. Adjoin the four productions $A \to A$, $a \to a$, $B \to B$, $b \to b$. Successive applications of these productions will reduce the tag word to length 1, producing a halt. If four such productions are added for each $q_i s_j$ which causes the Turing machine to halt, then, even without using halting letters, we have made halting for the tag system correspond exactly to halting for the given Turing machine.

We now construct a generalized Turing machine which reflects the action of such a tag system, and therefore has an undecidable halting problem. We shall use a notation similar to Minsky's, to make it easy to compare our machine to his. Let the alphabet of the tag system be $a_1, a_2, \ldots, a_m$, and suppose that the productions are

$$a_1 \to a_{11} a_{12} \ldots a_{1n_1},$$
$$a_2 \to a_{21} a_{22} \ldots a_{2n_2},$$
$$\ldots \ldots \ldots \ldots \ldots$$
$$a_m \to a_{m1} a_{m2} \ldots a_{mn_m}.$$

That is, if the first of the two deleted letters is a_i, then the letters $a_{i1} a_{i2} \ldots a_{in_i}$ are to be added at the end of the tag word. It is understood that each a_{ij} is some a_k. We may suppose that $1 \leq n_i \leq 4$.

14*

The symbols used by the Turing machine will be 0, 1, 2, 3, of which 0 is the blank. We put

$$N_i = 1 + (n_1 + 2) + \cdots + (n_{i-1} + 2),$$

so that, in particular, $N_1 = 1$. Any nonempty tag word $a_r a_s \dots a_z$ will then be coded as

$$S = 2^{N_r} \ 1 \ 2^{N_s} \ 1 \dots 1 \ 2^{N_z}.$$

The productions will be represented in a somewhat different way. Indeed, the right side of the i-th production will be coded as

$$P_i = 1 \ 1 \ 0^{N_{i n_i}} \ 0 \ 1 \dots 0 \ 1 \ 0^{N_{i2}} \ 0 \ 1 \ 0^{N_{i1}} \ 0 \ 1.$$

It is understood that if $a_{ij} = a_k$, then $N_{ij} = N_k$. If the initial tag word is coded as S, then the initial tape of the Turing machine will be taken in the form

$$\dots 0 \ 0 \ P_m \dots P_2 \ P_1 \ 1 \ 0 \dots 0 \ S \ 0 \ 0 \dots,$$

where we may insert any number of zeros preceding S. The different initial tapes which need to be considered differ only in the values of S.

Notice that N_i is the number of 1's between P_i and S. Thus the tag letter a_i is coded in a form 2^{N_i} which contains the information needed to locate the corresponding production P_i. This idea is borrowed from Minsky's treatment of his 4-symbol 7-state universal Turing machine, although the details are slightly different here.

On the tape of the Turing machine, we think of 2 and 3 as disguised forms of 0 and 1. At any time, the tape, excluding the blank tails at either end, may be divided into three parts, left, middle, and right, using the symbols $(0, 1)$, $(2, 3)$, and $(2, 1)$, respectively. That is, the middle part is completely disguised, but the right part is only half disguised. To begin with, the middle is empty and the right part consists of just S.

Machine table

	0	1	2	3
$R\,0$	—	$2L1$	$(L0)R$	—
$R\,1$	1	R	R	R
$R\,2$	2	R	R	R
$R\,3$	—	$0R0$	$0R$	$1R$
$L\,0$	$2L$	3	L	L
$L\,1$	$2(R2)L$	$3L2$	L	L
$L\,2$	$2(R1)L1$	$R3$	—	—

Our machine will use four right states, $R\,0$, $R\,1$, $R\,2$, $R\,3$, and three left states, $L\,0$, $L\,1$, $L\,2$. Each entry in the machine table shows the action when a head is in a given state and scanning a given symbol. For example, for a head in state $L\,2$ and scanning 0, the entry $2(R\,1)L\,1$ means that the 0 is changed to 2, and that the head splits into two heads, one in state $R\,1$ and one in state $L\,1$. The first of these is transient, so it is more convenient to say that the given head emits a transient head in state $R\,1$ and itself goes into state $L\,1$. The transient state is indicated in parentheses. In the absence of parentheses, no transient head is emitted. Also, in the machine table, the initial digit is omitted when the symbol is unchanged, and the final digit when the state is unchanged. But the omission of the L or R as well as the final digit indicates disappearance of the head. The dash indicates a halt. However, the only essential case is when the head is in state $R\,0$ and is scanning 0. The other four dashes occur in situations which will never be reached if we start with a tape of the sort described.

The machine is started with the permanent head in state $R\,0$ and scanning the first symbol of S, which must be 2. A transient head in state $L\,0$ is emitted, and seeks out the first 1 to the left and changes it to 3, changing 0's to 2's on the way. The transient head then disappears. As the permanent head traverses the 2^{N_r} at the beginning of S, this process is repeated N_r times, which suffices to change all the 0's and 1's between P and S to 2's and 3's. Thus this part of the tape has been put in the disguised form mentioned earlier, and is now considered as the middle of the tape.

Assume first that the given tag word $a_r a_s \ldots a_z$ has length at least 2. We are now ready for the production. The first 1 in S is changed to 2, and the permanent head goes into state $L\,1$. This head moves left to the 1 at the right end of P_r, changes it to 3, goes into state $L\,2$, finds the 0 preceding the 1, changes it to 2, emits a transient head in state $R\,1$ which writes a 1 to the right of S, and goes back into state $L\,1$. Each additional 0 encountered is changed to 2, and causes a 2 to be written to the right of S. If the pair 01 is encountered again, then the whole process is repeated. Finally 11 is encountered, which causes the permanent head to go into state $R\,3$. The desired addition has been made to the right of S.

In state $R\,3$, the productions $P_r, \ldots, P_1$ are restored to their original form, and the following 1 is also restored. Any original 0's preceding S are changed back into 0's, and the original $2^{N_r}1\,2^{N_s}$ at the beginning of S, which is now $2^{N_r}2\,2^{N_s}$, is also changed to 0's. Following this, there will be a 1, whether it was there to begin with or not. This will be changed to 0, and the head will go into state $R\,0$ scanning the next symbol, which is indeed the first symbol of the new S.

The above description assumes that the tag word has length at least 2. If the tag word consists of a single letter a_r, then $S = 2^{N_r}$. We locate P_r as before, but we do not use it. After the 2^{N_r} is passed, we find 0 instead of 1. The head is in state $R\,0$, and the machine halts. Thus the Turing machine halts if and only if the tag word is reduced to length 1, that is, if and only if the tag productions are halted.

We add the remark that if the machine is started with an initial tape of the sort described, then no matter whether the machine halts or runs forever, it will never scan a square to the left of the initially printed portion of the tape. Thus a one-way infinite tape could be used just as well as a two-way infinite tape.

§ 6. The Completion Problem

We shall now show how to find a set of 36 tiles for which the completion problem is undecidable. This is the smallest set of tiles which I have been able to find with this property.

We sketch briefly the stages by which this result was obtained, but make a precise count of pieces only at the final stage. In the first place, we could apply the methods of § 4 to Minsky's 4-symbol 7-state universal Turing machine. Since most states can be entered from only one side, many of the merging tiles from Fig. 13 can be omitted.

A considerable improvement can be made using instant action. The two tiles in Fig. 17 are replaced by the single tile in Fig. 18. A similar replacement is made when the motion is to the left. However, when the direction of motion changes, a unit of time elapses, and the merging and action tiles must be kept separate.
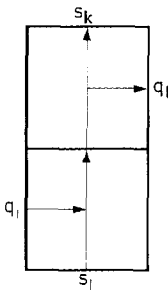

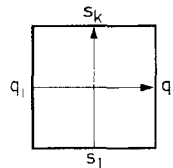Fig. 17. Merging and action tiles   Fig. 18. Instant action tile

With this new system of instant action, it will still be true that the lower and upper edges of a row of tiles represent the complete configurations of the Turing machine at two consecutive instants. However, many moves of the Turing machine may have happened in between. We

capture the complete configurations of the machine only at those times when the head reverses its direction of motion.

The following exceptional situation can occur for some Turing machines, although it does not actually arise when Minsky's machine is used. Suppose that, for some initial tape, the head ultimately makes an infinite sequence of moves in the same direction. In this case, although the Turing machine does not halt, its action is completed in a finite time. In the corresponding tiling, only alphabet tiles are used above a certain level. However, it is still true that the tiling can be completed if and only if the machine does not halt.

Now consider generalized Turing machines of the sort discussed at the beginning of §5. If the state-symbol pair $q_i s_j$ requires the head to split, then one unit of time is required. Besides a merging tile, we need a new tile of the sort shown in Fig. 19. If the state-symbol pair $q_i s_j$ causes the head to disappear, then instant action is used. No merging tile is required, but we need a new tile like the one in Fig. 20, or its mirror image in a vertical line.
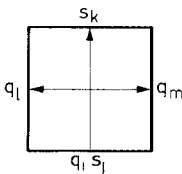

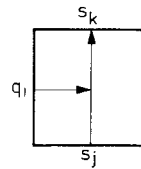Fig. 19. Splitting head   Fig. 20. Disappearing head

Here also, the lower and upper edges of any row of tiles will represent the complete configurations of the machine at two consecutive instants. Assuming that there is just one head which is not transient, each complete configuration indicates just one scanned square. We capture the complete configurations at those times when the permanent head splits or reverses its direction of motion. The exceptional situation mentioned earlier occurs if this head ultimately makes an infinite sequence of moves in the same direction.

The best estimate which I have found for the number of tiles needed to make the completion problem undecidable is obtained in this way using the machine with an undecidable halting problem which we constructed in §5.

The tiles corresponding to the machine table are easily counted. We need two for each entry involving a turn or a split, and one for each other entry except a dash. The entries requiring two tiles occur for the state-symbol pairs $R\,0\,1, R\,0\,2, L\,1\,0, L\,2\,0, L\,2\,1$. These five are just balanced by the five dashes, so that 28 tiles are required. One of the 28 tiles is

shown in Fig. 21. It is of the splitting-head type introduced in Fig. 19, and represents the first action of the machine when it is started in state $R\,0$ scanning 2.
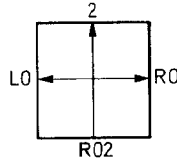

Fig. 21. The first action

In addition, we need 4 alphabet tiles and some starting tiles. A possible set, more economical than those used in §4, consists of the blank tile and the 3 tiles shown in Fig. 22. This makes 36 tiles in all.
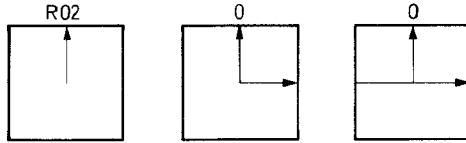

Fig. 22. New starting tiles

It may be of interest to count how many colors we are using. For these tiles, arrows with a given label always point in the same direction; in particular, unlabelled arrows always point to the right. If we introduce the color B for blank edges, A for unlabelled arrows, and otherwise use the label as a color, then the tiles will fit if and only if the abutting colors match. We see that there are nine colors, B, A, $R\,0$, $R\,1$, $R\,2$, $R\,3$, $L\,0$, $L\,1, L\,2$, used on vertical edges, and ten colors, B, 0, 1, 2, 3, $R\,0\,1$, $R\,0\,2$, $L\,1\,0$, $L\,2\,0, L\,2\,1$, used on horizontal edges. Since the two sets of colors are completely independent, we may use ten colors in all.

The initial position for the completion problem is obtained as follows. We line up the alphabet tiles corresponding to the tape segment $0\,P_m \ldots P_2\,P_1\,1\,S$. This includes the printed portion of the initial tape as described in §5, where we have chosen the option of inserting no 0's before S. The tile corresponding to the first 2 in S is replaced by the first tile in Fig. 22, indicating that the 2 is scanned in state $R\,0$. The initial position is completed by adjoining a blank tile at the left end of the row, and the second tile in Fig. 22 at the right end. This row of tiles may be represented symbolically as $B\,U\,C\,V\,D$, where B is the blank tile, C and D are the first two tiles in Fig. 22, U is a fixed string of alphabet tiles 0 and 1, and V is a variable string of alphabet tiles 1 and 2. Only six different tiles are used in the initial position.

The rightmost tile in this row forces an infinite sequence of tiles like the third tile in Fig. 22 to appear to its right. We shall also show that the blank tile and the 0 tile at the left end of the row force infinite columns of similar tiles to appear above them. We recall that the 0 represented by the 0 tile is never scanned by the Turing machine. Hence there is no intrusion into the column of 0 tiles from the right. Thus as long as we use blank tiles in the first column, we must use 0 tiles in the second. The first time that we do not use a blank tile in the first column, we must use one of the tiles in Fig. 22. The second and third tiles are impossible, since no tile would fit to the right. Thus the first tile must be used, and a 0 tile will appear to its right. However, at the next level, the tile in Fig. 21 will appear in the first column, and no tile can fit to its right. Indeed, the machine table shows that the machine halts if it is in state $R\,0$ scanning 0. Consequently no merging tile or instant action tile is provided for this case.

So far, we have a horizontal row of tiles extending infinitely far to the right, and two vertical columns extending infinitely far upward. The horizontal row represents the printed portion of the initial tape and the blank portion to its right, which are the only portions of the tape used by the Turing machine. The two vertical columns protect the first quadrant from intrusions from the left. Thus the operation of the Turing machine is exactly mirrored in the first quadrant, and this quadrant can be completely tiled if and only if the machine never halts. The tiling of the rest of the plane can always be completed in a trivial way, using just blank tiles and alphabet tiles. Since the machine has an undecidable halting problem, this set of 36 tiles has an undecidable completion problem, even when we restrict the initial position to a finite row of tiles of the special sort described above.

§7. New Proof of Berger's Theorem

We now return to the considerations of §3, and develop these ideas a bit further. At the end of this section, we combine the results obtained with those of §4 to obtain a new proof of Berger's undecidability theorem.

We again consider tilings using copies of the five basic tiles obtained by translation, rotation, and reflection, subject to the constraint that crosses appear in alternate positions in alternate rows. As we saw in §3, every $(2^{n+1}-1)$-square has a cross at the center with arms radiating out, and four (2^n-1)-squares in the corners. At the centers of these (2^n-1)-squares are four crosses, each of which faces two of the others at the distance 2^n. These crosses together with the arms in between them form a hollow square of outer dimension 2^n+1 and inner dimension

$2^n - 1$. This hollow square will be called a border of side 2^n, or a 2^n-border. Every pair of crosses which face each other at the distance 2^n will lie at two corners of a 2^n-border.

The only cross in the $(2^{n+1} - 1)$-square which does not face another cross within this square is the one at the center. Thus each of the other crosses determines a 2^k-border with $k \leq n$, and we see that $k = n$ only for the 2^n-border described above. Thus this 2^n-border does not intersect any other 2^n-border, and the only larger border which it intersects is the 2^{n+1}-border one of whose corner crosses is at the center of the 2^n-border. It follows that two borders can intersect only if the side of one is twice the side of the other. To every border, there is another border whose side is twice as long and which has one corner at the center of the given border, the larger border passing through the middle of two sides of the smaller border. On the other hand, every border of side at least 4 has four borders whose sides are half as long which are centered at its four corner crosses. They intersect it one-quarter and three-quarters of the way along the sides.

We shall modify the five basic tiles by coloring the side arrows red or green. There is, however, one tile with nothing to color. For each of the other four tiles, we allow two colorings, according to the following rules. At most one color may be used horizontally, and at most one color may be used vertically. For the cross, the same color shall be used in both directions. For the arm which has side arrows both ways, one color shall be used horizontally and the other color vertically. The nine tiles obtained in this way will be called the colored basic tiles.

We now consider tilings of the plane using copies of the nine colored basic tiles obtained by translation, rotation, and reflection, subject to the constraint that green crosses appear in alternate positions in alternate rows, and perhaps elsewhere. If we ignore the colors, then the tilings will be of the type previously considered.

Notice that the colors will go completely around the borders, so that each border is either red or green. Intersecting borders must be of different colors. But the constraint on the green crosses forces 2-borders to be green. Hence 4-borders are red, 8-borders green, etc. That is, every 2^n-border is green when n is odd, but is red when n is even.

Conversely, if we are given any tiling of the plane by the five basic tiles subject to the constraint that crosses appear in alternate positions in alternate rows, then the side arrows may be colored red or green in such a way that we use only the nine colored basic tiles and satisfy the constraint mentioned above. We simply color the 2^n-borders green when n is odd and red when n is even. If, outside of all the borders, there are one or two corridors containing side arrows, then these may be colored in either of two ways.

We can realize the constraint on the green crosses by using either corner markings or parity markings. We shall describe both of these, though only the parity markings will be useful for the proof of Berger's theorem. In the case of the corner markings, we allow all nine tiles to have dented corners, but only the green cross can have bumps at the corners. This makes a total of ten tiles. In the case of the parity markings, the constraint on the green crosses is enforced by allowing the green cross to have either the odd-odd or the even-even marking, as was done previously for the cross, but forcing the red cross to have the even-even marking. The two tiles with red in arrows may also be restricted to the even-even marking, since they can occur only at the middle of the edges of red borders, or at the junction of corridors. The other tiles are allowed two parity markings as usual. Thus there are fifteen tiles in all.

The green borders will now be forgotten, only the red borders being considered. Red borders have sides of the form 4^n. No two red borders can intersect, although one may lie completely within another. Every tiling of the plane contains arbitrarily long sequences of borders of sides $4, 4^2, 4^3, \ldots, 4^n$, each lying within the next. Hence there are some tilings of the plane which contain infinite sequences of borders of sides $4, 4^2, 4^3, \ldots$, each lying within the next. For such a tiling, every tile will lie within some red border.

The region within a red border but outside of all red borders within it will be called a board. By what we have just said, some tilings of the plane consist solely of boards and borders. On a board, it will be important to locate the free rows and columns of tiles, that is, the ones which run completely across the board, from outer border to outer border, without running into any of the smaller borders inside.

First we count the number of free rows or columns. Let F_n be this number for a board of side $4^n - 1$. Now the positions of the 4^k-borders repeat with the period $2 \cdot 4^k$, hence the pattern of free rows or columns of a $(4^n - 1)$-board is exactly repeated in the middle of a $(4^{n+1} - 1)$-board. In addition, if we omit the center row or column, the remaining half patterns are repeated at the sides. Thus we find that $F_{n+1} = 2F_n - 1$. But $F_1 = 3$, hence $F_n = 2^n + 1$ in general. It may be noted that, aside from the center row and column, all of the other free rows and columns have odd numbers.

To locate the free rows and columns, we shall use a new type of marking. It is convenient to think of these markings as a signal which travels along the rows and columns. We shall call this the obstruction signal, since its purpose is to determine whether there is any obstruction along the line. The obstruction signals will be transmitted unchanged along the rows and columns of the board. They will be emitted and absorbed by the red borders. Specifically, if a tile is part of a red border,

then we agree that each edge of the tile along the outer boundary of the border must either emit or absorb an obstruction signal, whereas an edge along an inner boundary may absorb but cannot emit an obstruction signal.

We see that no obstruction signals run along the free rows. Now look at any tile on the board which is not in a free row. It will have the outer boundary of a red border to either its right or left, or both. If there is an outer boundary on one side and an inner boundary on the other, then there will be an obstruction signal running from the outer boundary to the inner boundary. If there are outer boundaries on both sides, then the sense of the obstruction signal is ambiguous; it may run either from left to right or from right to left. In any case, a tile lies in a free row if and only if no obstruction signal runs through it horizontally. Similarly, a tile lies in a free column if and only if no obstruction signal runs through it vertically. The obstruction signals may be indicated by means of symmetrical markings which can be added to our ten tiles with corner markings or our fifteen tiles with parity markings before they are rotated or reflected.

We take the tiles just constructed, using the variant based on parity markings (and hence involving no corner markings), and rotate and reflect them in all possible ways. Hereafter, we allow only translation.

We are now ready to give a proof of Berger's theorem that there is no decision method for the problem of tiling the plane with copies of a finite number of tiles with colored edges so that abutting edges match, translation only being used. The proof will be based on the above results and on §4. As in §4, we shall show how to make a set of tiles correspond to each Turing machine in such a way that the plane can be tiled with these tiles if and only if the Turing machine never halts if started on a blank tape. The new difficulty in the proof is to find a part of the plane in which to picture the operation of the machine. Just as was the case with Berger [1], what we actually do is to find parts of the plane in which we can picture arbitrarily large finite portions of the machine operation. For this purpose, we shall use the boards constructed above.

Let an arbitrary Turing machine be given. On each board tile which is free both horizontally and vertically, we superpose each of the Turing machine signals in Figs. 12–14. These signals, which were indicated in the figures by central arrows, can be shifted so as not to interfere with the other markings on these tiles. The board tiles which are free in one direction but not in the other will simply transmit signals unchanged in the free direction. The board then acts exactly as if it were a square of side $2^n + 1$ with the free rows and columns being contiguous.

We need some boundary conditions. We agree that every arm with a horizontal red marking below the center which does not absorb an obstruction signal along its upper edge shall emit a Turing machine signal there, and that this signal shall be s_0 unless we have red in arrows, in which case it shall be $q_0 s_0$. This will force the center tile along the lower side of a red border to emit $q_0 s_0$ from its top edge, and the other tiles at the bottom of free columns to emit s_0. In a similar way, we mark the other red arms so the left, right, and upper parts of the border may absorb any Turing machine signals along their inner edges.

With these agreements, we shall be able to tile arbitrarily large boards if and only if the Turing machine never halts. Since some tilings of the plane consist solely of boards and borders, we can tile the plane if and only if the Turing machine never halts. This completes the new proof of Berger's theorem.

An examination of the possibilities when some tiles are outside of all borders shows that no matter what tiling of the plane we make with the tiles which we had before the Turing machine signals were added, these signals can be added if and only if the machine never halts. There may be a complete model of the action of the Turing machine, but there is no way to force such a model to exist. More details about this are given in § 8.

§ 8. Tilings and 2-adic Numbers

We showed in § 3 that any tiling of the plane by the five basic tiles with the constraint that crosses appear in alternate positions in alternate rows must be nonperiodic. This result may be applied both to the six tiles with corner markings and to the ten tiles with parity markings. In this section, we shall determine all possible tilings of the plane using these tiles.

It will be convenient to think of the tiles placed with their centers at lattice points, and to locate a tile by giving the coordinates of its center. When we speak, for example, of tiling the region $x > 0$, $y > 0$ (the first quadrant), we shall mean that a tile is to be placed at each lattice point (x, y) with $x > 0$, $y > 0$. Notice that the union of these tiles, ignoring the notches, will cover just the point set $x \geqq \frac{1}{2}$, $y \geqq \frac{1}{2}$.

First we look at a special tiling of the first quadrant. Take an expanding sequence of $(2^n - 1)$-squares, in the sense of § 3, determined by $0 < x < 2^n$, $0 < y < 2^n$, for $n = 1, 2, 3, \ldots$. The center cross of each square will face up and to the right. The union of these squares will determine a tiling of the first quadrant, $x > 0$, $y > 0$. In this tiling, it is easily seen that there is a cross at (x, y) if and only if x and y contain the same power of 2. Conversely, this condition completely determines the tiling. The constraint is satisfied as soon as we put down the odd-odd crosses. The

fact that there are crosses for $x \equiv y \equiv 2 \pmod 4$ determines the orientation of the odd-odd crosses, etc. The arms are completely determined by the crosses.

The characterization of the tilings in general requires the use of 2-adic integers. A 2-adic integer A has the form

$$A = a_0 + a_1 \cdot 2 + a_2 \cdot 2^2 + a_3 \cdot 2^3 + \cdots,$$

where each a_n is 0 or 1. The value of A is determined as soon as we know in which residue class mod 2^n it falls for each n. The ordinary integers are included among the 2-adic integers. Indeed, the natural numbers are obtained by taking $a_n = 0$ for large n, and the negative integers are obtained by taking $a_n = 1$ for large n. In particular, $-1 = 1 + 2 + 2^2 + 2^3 + \cdots$.

Now consider the union of any expanding sequence of $(2^n - 1)$-squares. We know that this union is either a plane, a half plane, or a quarter plane. Determine 2-adic integers A and B by the condition that the next columns of tiles to the left and right of the $(2^n - 1)$-square satisfy $A + x \equiv 0 \pmod{2^n}$, and that the next rows of tiles below and above the $(2^n - 1)$-square satisfy $B + y \equiv 0 \pmod{2^n}$. If the union has a vertical boundary, it will lie along the line $A + x = 0$, and hence A will be an ordinary integer. If the union has a horizontal boundary, then it will lie along the line $B + y = 0$, and hence B will be an ordinary integer. In any case, crosses will appear at just those positions in the union for which $A + x$ and $B + y$ contain the same power of 2.

Conversely, suppose that any two 2-adic integers A and B are given. If A is an ordinary integer, also choose one of the two inequalities $A + x > 0$ or $A + x < 0$. If B is an ordinary integer, also choose one of the two inequalities $B + y > 0$ or $B + y < 0$. The tiling of the plane, half plane, or quarter plane defined by these inequalities will be uniquely determined if we insist that crosses shall appear at just those positions for which $A + x$ and $B + y$ contain the same power of 2. If $A = B = 0$, and we choose the inequalities $x > 0$, $y > 0$, then this tiling is just the one first considered.

If A and B are not ordinary integers, we obtain at once a tiling of the whole plane. On the other hand, if A is an ordinary integer but B is not, then we obtain a tiling of one of the half planes $A + x > 0$ or $A + x < 0$. We could define a tiling of the right half plane $A + x > 0$ using some B_1, and a tiling of the left half plane $A + x < 0$ using some B_2. If $B_1 \neq B_2$, then the line $A + x = 0$ is a fault. If corner markings are used, then B_1 and B_2 are arbitrary, but if parity markings are used, then $B_1 - B_2$ must be even. (This is the first time that a difference between the two problems has appeared.) A sequence of arms can be filled in along the corridor $A + x = 0$. Either all of the arms will point up, or else all will point down. Besides the central vertical arrows, there may be no other vertical

arrows, or each tile in the corridor may have a vertical arrow at the left, or each may have one at the right.

Either of the numbers B_1 or B_2 may be an ordinary integer. We would then have an additional corridor along $A + x > 0$, $B_1 + y = 0$, or along $A + x < 0$, $B_2 + y = 0$, but this cannot be a fault. The corridors can be filled with arms, which point out at one end of one corridor, and point in elsewhere. The only case when there is any other possibility is when $B_1 = B_2 = B$, and B is an ordinary integer. Suppose, for example, that $A = B = 0$. In all four quadrants, we place a cross at (x, y) if and only if x and y contain the same power of 2. There are corridors along the coordinate axes. The tile at the origin is completely arbitrary. After it has been chosen, all of the other tiles along the axes are determined. If the tile at the origin is a cross, then arms radiate out in all four directions.

It may be noted that the last tiling considered is invariant under rotation through 90° about the origin until the tiles along the axes are added, but that this symmetry is then destroyed. Indeed, it is easily seen that no tiling can have rotational symmetry. On the other hand, the tiling may be symmetric to one of the axes or to the line $y = \pm x$, depending on which tile is used at the origin. No other tiling has greater symmetry than this, so the group of isometries of the tiling has order at most 2. We can prevent nontrivial isometries by using more tiles.

The only other possible tilings besides those described above are the ones in which the roles of horizontal and vertical are interchanged.

We shall now make a further study of the red borders introduced in §7. We want to know whether the red borders and the boards bounded by them completely fill the plane, or whether there are some tiles outside of all the red borders. If so, we also ask whether there are complete columns or complete rows outside the red borders.

Consider a tiling with no fault, so that the same 2-adic integers A and B are used throughout the plane to determine the tiling. If the tiling has a fault, then the considerations below may be applied separately to the half planes on either side of the fault.

Any border of side 4^k is centered at a point (x, y) for which $A + x$ and $B + y$ are odd multiples of 4^k, say $A + x = (2u+1) \cdot 4^k$ and $B + y = (2v+1) \cdot 4^k$. The border together with its interior is then defined by the inequalities

$$|2(A + x) - (4u + 2) \cdot 4^k| \leqq 4^k,$$

$$|2(B + y) - (4v + 2) \cdot 4^k| \leqq 4^k.$$

Now expand $2A + 2x$ and $2B + 2y$ as 4-adic integers; that is, we write

$$2A + 2x = a_0 + a_1 \cdot 4 + a_2 \cdot 4^2 + a_3 \cdot 4^3 + \cdots,$$

$$2B + 2y = b_0 + b_1 \cdot 4 + b_2 \cdot 4^2 + b_3 \cdot 4^3 + \cdots,$$

where $0 \leq a_n \leq 3$ and $0 \leq b_n \leq 3$ for all n. The first inequality above will be satisfied for some u if $a_k = 1$ or 2, or if $a_k = 3$ and $a_0 = a_1 = \cdots = a_{k-1} = 0$, but in no other case. Column x will be outside of all red borders if this condition is not satisfied for any $k > 0$. This will be the case if and only if either $A + x = 0$ or else the unit digit in the 4-adic expansion of $2A + 2x$ is 2 and each other digit is 0 or 3.

If A is given, we want to know in what ways x can be chosen so that this last condition is satisfied. If A is an ordinary integer, we can choose x so that $A + x = 0$, and we can choose x in infinitely many ways so that the other alternative is satisfied, all but a finite number of the digits of $2A + 2x$ being 0, or all but a finite number being 3. If A is not an ordinary integer, then adding $2x$ to $2A$ can change only a finite number of the digits of $2A$. Thus we cannot satisfy the condition unless the 4-adic expansion of $2A$ contains only a finite number of 1's and 2's. If this happens, then infinitely many values of x may be chosen.

It follows that there are some complete columns outside of all the red borders if and only if the 4-adic expansion of $2A$ contains only a finite number of 1's and 2's. In this case, there are indeed infinitely many columns outside of the red borders. Similarly, there are some complete rows (and then infinitely many) outside of all red borders if and only if the 4-adic expansion of $2B$ contains only a finite number of 1's and 2's.

There is a peculiar case which should be noted. If A is an ordinary integer, then although the column x defined by $A + x = 0$ is outside of all red borders, it may actually be red. Indeed, it may look like part of the left or right side of a red border. Similarly, if B is an ordinary integer, then the row y with $B + y = 0$ may look like part of the bottom or top side of a red border.

There may be some tiles outside of all red borders without there being any complete columns or any complete rows outside. The tile (x, y) will be inside or on some border of side 4^k if and only if

$$a_k = 1 \text{ or } 2, \quad \text{or} \quad a_k = 3 \quad \text{and} \quad a_0 = a_1 = \cdots = a_{k-1} = 0,$$

and

$$b_k = 1 \text{ or } 2, \quad \text{or} \quad b_k = 3 \quad \text{and} \quad b_0 = b_1 = \cdots = b_{k-1} = 0.$$

We can find such a value of k for every tile (x, y) if and only if there are infinitely many positions in which the 4-adic expansions of both $2A$ and $2B$ contain 1 or 2. In other words, some tiles will lie outside of all red borders just in case there are only a finite number of positions in which both expansions contain 1 or 2. However, if each expansion contains infinitely many 1's or 2's, then there will not be any complete column or complete row outside of the red borders.

Finally, we examine briefly the problem of adding Turing machine signals to the tiles outside of the red borders. If there are no complete

columns or rows outside, then each outside tile must have vertical and horizontal obstruction signals, and no Turing machine signals are to be added. Even when there are complete columns or rows outside, we may add such obstruction signals. Thus no matter what tiling of the plane by the colored basic tiles with parity markings is given, we can add the obstruction and Turing machine signals if and only if the Turing machine never halts. No action of the Turing machine need be shown outside of the red borders.

But we can say more. Suppose that we are given a tiling of the plane with the obstruction signals already added in any permissible way. Then we can still add Turing machine signals if and only if the machine never halts. This is trivial for the portion of the plane outside of the red borders unless there is a down arm with red in arrows there with no obstruction signal absorbed on its upper edge. Only if there is such a tile will we have a complete or partial model of the operation of the Turing machine outside of the red borders. A case in which a complete model of the Turing machine will appear is that in which we use $A = B = 0$ throughout the plane, place a down arm with red in arrows at the origin, and do not introduce any unnecessary obstruction signals.

References

1. Berger, R.: The undecidability of the domino problem. Mem. Amer. Math. Soc. no. **66**, (1966).
2. Büchi, J.R.: Turing-machines and the Entscheidungsproblem. Math. Ann. **148**, 201–213 (1962).
3. Cocke, J., Minsky, M.: Universality of tag systems with $P = 2$. J. Assoc. Comput. Mach. **11**, 15–20 (1964).
4. Kahr, A.S., Moore, E.F., Wang, H.: Entscheidungsproblem reduced to the ∀∃∀ case. Proc. Nat. Acad. Sci. U.S.A. **48**, 365–377 (1962).
5. König, D.: Über eine Schlußweise aus dem Endlichen ins Unendliche. Acta Litt. Sci. Szeged **3**, 121–130 (1927).
6. Minsky, M.L.: Computation: Finite and infinite machines. Englewood Cliffs, N.J.: Prentice-Hall 1967.
7. Robinson, R.M.: Seven polygons which permit only nonperiodic tilings of the plane (abstract). Notices Amer. Math. Soc. **14**, 835 (1967).
8. Wang, H.: Proving theorems by pattern recognition – II. Bell System Tech. J. **40**, 1–41 (1961).
9. – Dominoes and the AEA case of the decision problem, Mathematical theory of automata, p. 23–55. Brooklyn, N.Y.: Polytechnic Press 1963.

Raphael M. Robinson
Department of Mathematics
University of California
Berkeley, Calif. 94720, USA