STABLE TRANSVERSALS AND STOCHASTIC FUNCTIONS IN POLYOMINOES

Samuel W. Bent

Computer Sciences Technical Report #520

November 1983

Stable Transversals and Stochastic Functions in Polyominoes

Samuel W. Bent *
Computer Science Department
University of Wisconsin-Madison

* Research partially supported by National Science Foundation Grant MCS 8203238.

Abstract: We show that not every polyomino has a stochastic function (a labelling of its cells by nonnegative real numbers so that the labels in every maximal rectangle sum to 1). We also show that determining whether a polyomino has a stochastic function can be done in polynomial time, but that determining whether it has a stable transversal (a stochastic function with integer labels) is NP-complete. This settles some open questions posed by Berge, Chen, Chvatal, and Seow.

Keywords: polyomino, stochastic function, stable transversal, NP-complete, circuit design.

AMS subject classifications (1980): 05B50, 05C65, 68C25.

1. Introduction

A *polyomino* is a finite set of cells in the infinite planar square grid. Polyominoes have an ancient tradition as a game or puzzle [4], but recently they have attained new importance in digital image processing and in circuit design. An image or a circuit layout can be thought of as a polyomino for some purposes, and combinatorial properties of polyominoes, such as the minimum number of rectangles whose union equals a given polyomino, influence the effeciency with which an image or circuit can be represented or processed in some way.

Berge et al. surveyed many combinatorial results about polyominoes, and posed many more open questions [2]. This paper answers two of these questions, as well as a third related question that Berge et al. did not explicitly pose.

Any polyomino can be thought of as a hypergraph in a natural way. The cells of the polyomino correspond to the vertices of a hypergraph; its maximal rectangles correspond to the edges. (A maximal rectangle of a polyomino P is simply any rectangle contained in P that is not strictly contained in some larger rectangle within P.) Using the language of hypergraphs, define a ***transversal*** of a polyomino P to be a set of cells of P that has at least one cell in common with each maximal rectangle of P. The set is a ***stable transversal*** if it contains exactly one cell in common with each maximal rectangle.

Equivalently, a stable transversal is a function X of the cells of P that maps each cell c to $\{0,1\}$ in such a way that

$$\sum_{c \varepsilon R} X(c) = 1$$

for each maximal rectangle R in P. Allowing X to take on non-integer values yields a *stochastic function*, namely a function X mapping the cells of P to $[0,1]$ such that

$$(1) \qquad \sum_{c \varepsilon R} X(c) = 1$$

for each maximal rectangle R in P. Clearly a stable transversal is a special case of a stochastic function.

Berge et al. give an example of a polyomino that has no stable transversal, although it does have a stochastic function. They pose the following open questions:

Q1 Is there a polyomino with no stochastic function?

Q2 How difficult is it to determine whether a polyomino has a stable transversal?

If the answer to the first question is "yes", a third question follows naturally:

Q3 How difficult is it to determine whether a polyomino has a stochastic function?

We provide the answers "yes", "NP-complete", and "polynomial", respectively, to the three questions.

2. Notation and a useful lemma

Let P be a polyomino equipped with a stochastic function X. Number the n cells of P with the integers {1,2,...,n}. A *top cell* of P is a cell whose upper neighbor is not in P. Similarly, a *bottom* (respectively *left*, *right*) *cell* is a cell in P whose lower (respectively left, right) neighbor is not in P. A rectangle is determined by the locations of two diagonally opposite corners, so let $<a,b>$ denote the rectangle with corners numbered a and b. (In most cases, we will only refer to $<a,b>$ when it is a maximal rectangle of P with a as its upper left corner.) Let x_a denote the value of X at cell a, and let $x_{<a,b>}$ denote

$$\sum_{c\varepsilon<a,b>} X(c).$$

In Figure 1, the top cells are 1, 2, and 5; the right cells are 2, 5, and 7; and the maximal rectangles are <1,4>, <2,6>, <3,5>, and <4,7>. Thus we must have $x_{<2,6>} = x_2+x_4+x_6 = 1$.

(Figure 1)

The following lemma formalizes and slightly generalizes an argument used by Berge et al. to exhibit a polyomino with no stable transversal.

Lemma 1 (Berge et al. [2]). Let P be a polyomino with stochastic function X. Suppose R_1 (with corners a, b, c, and d, reading clockwise from upper left) and R_2 (with corners e, f, g, and h) are rectangles in P such that

(i) $c\,\varepsilon R_2$ and $e\,\varepsilon R_1$,

(ii) $x_{<a,c>} = x_{<e,g>} = 1$,

(iii) $x_{<b,h>} = x_{<d,f>} = 1$.

Then $x_q=0$ for all cells q in $(R_1 \cup R_2)-(<b,h> \cup <d,f>)$. (See Figure 2.)

(Figure 2)

Proof: By (i), $<b,h>$ and $<d,f>$ are actually rectangles within P, so we are allowed to refer to them in (iii). We have

$$x_{<a,c>}+x_{<e,g>} = 2, \text{ by (ii)},$$
and
$$x_{<b,h>}+x_{<d,f>} = 2 \text{ by (iii)}.$$

The first equation counts cells in $R_1 \cup R_2$, counting cells in $R_1 \cap R_2$ twice; the second counts cells in $<b,h> \cup <d,f>$, also counting cells in $R_1 \cap R_2$ twice. Subtract the second from the first and recall that $x_q \geq 0$ for all cells q to complete the proof.

We will almost always apply Lemma 1 in situations where $<a,c>$, $<e,g>$, $<b,h>$, and $<d,f>$ are maximal rectangles, although in one case $<b,h>$ and $<d,f>$ will be contained in larger maximal rectangles whose cells are known to have value 0 outside the region shown in Figure 2.

If s is a real number, let $\bar{s}$ denote $1-s$.

3. Wires, signals, and gates

The constructions in the main theorems are best described by analogy with digital circuits. This section describes the building blocks of circuit design via polyominoes. Each piece of circuit is described as if it were part of a large polyomino P with stochastic function X.

A ***wire*** is a series of overlapping 2x2 rectangles, as in Figure 3(a). By Lemma 1 applied to the maximal rectangles <1,4> and <4,7>, we find that $x_1 = x_7 = 0$. Applying Lemma 1 to each pair of overlapping 2x2 rectangles, we find successively that $x_4 = x_{10} = 0$ and $x_7 = x_{13} = 0$. It is not necessary that cell 1 be a top or left cell, or that cell 13 be a bottom or right cell; as long as there are at least three 2x2 rectangles, we can extend the wire as far as we like, knowing that the center cells must have value 0.

A wire propagates a ***signal*** s along one side as follows. If $x_3=s$, then because $x_{<1,4>}=1$ and $x_1=x_4=0$, we must have $x_2=\bar{s}$. This forces $x_6=s$, because $x_{<2,6>}=1$ and

$x_4=0$. Similarly this forces $x_5=\bar{s}$ and $x_9=s$, forcing $x_8=\bar{s}$ and $x_{12}=s$, finally forcing $x_{11}=\bar{s}$. These values satisfy equation (1) for all the rectangles in the wire. Thus the wire propagates signal s down the lower left side, or, equivalently, $\bar{s}$ up the upper right; the signal is determined by facing in the direction of propagation and reading the vaule on the right of the wire. We symbolize a wire as in Figure 3(b).

(Figure 3)

A *tab* is a group of three 2x2 rectangles attached to a wire as in Figure 4(a). Its purpose is to restrict the signal on the side of the wire nearest the tab to be at least 1/2; this restriction is symbolized as in Figure 4(b). It does this as follows. In the wire, Lemma 1 shows that

$$x_1=x_4=x_7=x_{10} = x_{18}=x_{24}=x_{30}=x_{35} = 0;$$

note that x_{13} is not included because it doesn't belong to a 2x2 maximal rectangle. In the tab, Lemma 1 shows that

$$x_{32}=x_{27}=x_{21}=x_{16} = 0.$$

Suppose $x_3=s$ and $x_{33}=t$. Then as before we find that

$$x_6=x_9=s; \qquad x_2=x_5=x_8=x_{11}=\bar{s};$$

$$x_{28}=x_{22}=t; \qquad x_{26}=x_{20}=x_{15}=\bar{t}.$$

But then

$$0 = x_{<10,17>}-x_{<12,18>} = (x_{10}+x_{11})-(x_{14}+x_{18}) = \bar{s}-x_{14},$$

so

$$x_{14}=x_{19}=x_{25}=x_{31}=\bar{s}; \qquad x_{23}=x_{29}=x_{34}=s.$$

Thus the tab doesn't stop signal s from propagating. However, by looking at part of the rectangles <8,22> and <15,19>, we see that

$$x_8+x_{22} = \bar{s}+t \leq 1$$

and $$x_{19}+x_{15} = \bar{s}+\bar{t} \leq 1.$$

Adding, we find that

$$2\bar{s}+1\leq 2,$$

which implies that $s\geq 1/2$, as claimed. The remaining cells (12, 13, and 17) can be assigned values in many ways, the easiest being $x_{12}=x_{17}=0$, $x_{13}=s$. This forces $t=s$, but the value of t really doesn't matter.

(Figure 4)

The next building block consists of a central 3x3 rectangle with three 2x2 rectangles attached to each corner, as shown in Figure 5(a). These attachments can be extended as wires in adjacent pairs to form a *turn* (Figure 5(b)), in opposite pairs to form an *inverter* (Figure 5(c)), or in threes (Figure 5(d)). If signal s enters on the upper left wire, then signal $\bar{s}$ leaves on any other attached wire. To see this, use Lemma 1 to find that

$$x_1=x_6=x_{12}=x_{18} = x_4=x_9=x_{15}=x_{20} = x_{42}=x_{37}=x_{31}=x_{26} = x_{45}=x_{40}=x_{34}=x_{28} = 0.$$

Then if $x_5=s$, the signal propagates toward the center, yielding $x_{17}=s$, and $x_{13}=\bar{s}$. Now $x_{17}+x_{21}\leq 1$, so $x_{21}\leq 1-x_{17}=\bar{s}$, which forces $x_{14}=1-x_{21}\geq s$. Similarly, $x_{33}\leq 1-x_{14}\leq\bar{s}$, forcing $x_{29}\geq s$, forcing $x_{25}\leq\bar{s}$, and finally forcing $x_{32}\geq s$. But since $x_{32}\leq 1-x_{13}=s$, we must have $x_{32}=s$, which forces all the other inequalities to be equalities too. In short,

$$x_{17}=x_{14}=x_{29}=x_{32}=s, \qquad x_{13}=x_{21}=x_{33}=x_{25}=\bar{s}.$$

These signals then propagate down their respective wires as claimed. The remaining values must be $x_{19}=x_{22}=x_{24}=x_{27}=0$ and $x_{23}=1$.

(Figure 5)

A *crossover* (Figure 6(a)) allows two wires to cross each other without interference, except that the signals are inverted as they cross. It consists of two elongated overlapping inverters, symbolized as in Figure 6(b). As in the previous paragraph, a signal entering the "vertical" inverter, say $x_3=s$, starts a chain reaction forcing $x_3=x_2=x_{15}=x_{16}=s$ and $x_1=x_4=x_{17}=x_{14}=\bar{s}$. A signal entering the "horizontal" inverter, say $x_7=t$, forces $x_7=x_6=x_{11}=x_{12}=t$ and $x_5=x_8=x_{13}=x_{10}=\bar{t}$. The two signals never need to interact, since we can set $x_9=1$ to take care of the large central rectangles.

(Figure 6)

A NAND *gate* (Figure 7(a), symbolized in Figure 7(b)) takes two input signals s and t, and produces an output signal r. Besides the usual restriction that these signals must lie in [0,1], a NAND gate enforces the restriction

(2) $\quad r+s+t \le 2,$

but allows any values that obey this restriction. In other words, if $s=t=1$ then r must be 0; otherwise r can be nonzero. In section 5 we will see why the name NAND is appropriate. To verify its behavior, suppose $x_1=s$, $x_{19}=t$, and $x_{14}=r$. By Lemma 1 applied to the input and output wires,

$$x_2=x_3=x_{11}=x_{15}=x_{17}=x_{18} = 0,$$

hence

$$x_4=\bar{s},\ x_{16}=\bar{t},\ \text{and}\ x_{12}=x_8=\bar{r}.$$

Applying Lemma 1 to <6,11> and <11,15> yields

$$x_6=x_7=x_{15} = 0.$$

We must have $x_{10}\le 1-x_1=\bar{s}$, and $x_9\le 1-x_{19}=\bar{t}$. But since

$$1 = x_{<6,11>} = x_8+x_9+x_{10} \le \bar{r}+\bar{t}+\bar{s},$$

equation (2) must hold. If r decreases from $2-(s+t)$, then $\bar{r}$ increases, and x_9 or x_{10} must decrease. Increasing x_{13} or x_5 compensates for this, allowing smaller values for r.

(Figure 7)

Finally, a *supertab* (Figure 8(a), symbolized in Figure 8(b)) forces entering signals to be 1 and departing signals to be 0. Wires may be attached at any of the corners. To verify its behavior, apply Lemma 1 to <2,36> and <10,44> to find that

$$x_2=x_3 = x_{43}=x_{44} = 0.$$

By symmetry,

$$x_{13}=x_{19} = x_{27}=x_{33} = 0.$$

Now Lemma 1 applies to <1,10> and <8,38>, even though <7,12> is not maximal, because $x_{13}=0$. Thus

$$x_1 = x_{17}=x_{18}=x_{24}=x_{25}=x_{31}=x_{32}=x_{37}=x_{38} = 0.$$

By symmetry, every cell has value 0 except for x_4, x_5, x_7, x_{20}, x_{23}, x_{26}, x_{39}, x_{41}, and x_{42}. This implies that $1 = x_{<7,13>} = x_7$, and by symmetry that

$$x_5=x_7=x_{41}=x_{39}=1,$$

which in turn forces

$$x_4=x_{20}=x_{26}=x_{42}=0, \text{ and } x_{23}=1.$$

(Figure 8)

Since wires run at a 45° angle from the coordinate axes of the planar grid, it is convenient to rotate the schematic diagrams of circuits by 45°. The remaining figures are drawn with this convention.

4. A polyomino with no stochastic function

Berge et al. give an example of a polyomino with no stable transversal. Using the components of the previous section, we can view their example as a wire with tabs on opposite sides (Figure 9). Let s be the signal propagating to the right. the first tab forces $s \leq 1/2$ and the second forces $s \geq 1/2$, so we must have $s = 1/2$. A stable transversal has only integer values, so this polyomino cannot have one.

(Figure 9)

Similar reasoning shows that the polyomino shown schematically in Figure 10 cannot have a stochastic function. If the wire begins by propagating signal s to the right, the first supertab inverts s to $\bar{s}$ and forces $s=1$, then the second inverts $\bar{s}$ to s and forces $\bar{s}=1$. But it is impossible to have both $s=1$ and $\bar{s}=1$.

(Figure 10)

5. NP-completeness and polynomial algorithms

A polyomino may or may not have a stable transversal or a stochastic function. Theorems 1 and 2 below show that determining whether it has a stable transversal is NP-complete, whereas deterining if it has a stochastic function can be done in polynomial time. For definitions and standard results about NP-completeness, consult the excellent book by Garey and Johnson [3].

Theorem 1. Determining whether a polyomino has a stochastic function can be done in polynomial time.

Proof: Let P be a polyomino with n cells. Since a rectangle is determined by two opposite corners, there are at most n^2 rectangles in P. Determining whether P has a stochastic function is equivalent to determining whether the system of linear inequalities

$$\sum_{c \varepsilon R} x_c = 1, \text{ for all maximal rectangles } R \text{ in } P;$$

$$0 \le x_c \le 1, \text{ for all cells } c \text{ in } P$$

has a feasible solution $(x_1, \ldots, x_n)$. This system has at most $2n^2+2n$ inequalities, n variables, and integer coefficients. A feasible solution can be found, if it exists, in polynomial time using (for example) Khachiyan's ellipsoid algorithm for linear programming [1].

Theorem 2. Determining whether a polyomino has a stable transversal is NP-complete.

Proof: Clearly the problem is in NP, since given a polyomino with n cells we can guess the n vaules of a stable transversal X and verify that equation (1) holds for each maximal rectangle. There are at most n^2 such rectangles (as noted in the proof of Theorem 1), each of size at most n, so the verification can easily be done in time $O(n^3)$.

We prove the problem is NP-hard by giving a reduction from 3SAT [3]. Let F be a boolean formula in 3CNF with n variables and m clauses. We construct a polyomino P with $O(m^2)$ cells that has a stable transversal if and only if F is satisfiable.

For each variable v in F, form a "variable component" consisting of a short wire leading into a sequences of inverters, as in Figure 11. Use as many inverters as there are occurrences of v in F, leaving enough room between them to allow the descending wires to contain turns. Call the signal entering from the left v; then signal v leaves from the even numbered inverters, $\bar{v}$ from the odd. In a stable transversal, signal v must be either 0 or 1, which we interpret as *false* and *true* to obtain a truth assignment to the variable v.

(Figure 11)

For each clause $(a \vee b \vee c)$ in F, form a "clause component" consisting of two NAND gates, input wires labelled $\bar{a}$, $\bar{b}$, and $\bar{c}$, and an output wire with a tab attached, all connected with turns and inverters as shown in Figure 12. The tab forces the output signal to be at least 1/2, so it must be 1, representing *true*. A NAND gate with *true* output forces at least one of its inputs to be *false*, by equation (2). Thus the component forces at least one of the three input signals $\bar{a}$, $\bar{b}$, and $\bar{c}$ to be *false*, which means at least one of a, b, and c must be *true*.

(Figure 12)

Finally, connect the $3m$ wires descending from the variable components to the $3m$ wires ascending from the clause components, using turns, crossovers, and inverters as necessary to ensure that the signal leaving a variable component arrives at a clause component with the correct value. Figure 13 shows a possible connection for the formula $(x \vee \bar{y} \vee z)(\bar{x} \vee \bar{z} \vee w)$. By the observations here and in section 4, the resulting polyomino has a stable transversal if and only if F is satisfiable.

(Figure 13)

The strategy used in Figure 13 to connect the wires partitions the space between the variables and the clauses into $3m$ layers, each used to route one wire to the right until it is above its destination. There are $3m$ vertical strips near variables and another $3m$ near clauses. Each layer and strip has constant thickness (enough to hold a crossover or turn), so the whole construction lies within a rectangle of area $O(m^2)$. Thus the number of cells in the resulting polyomino is $O(m^2)$, and the construction can be done easily in polynomial time.

6. Acknowledgment

I am grateful to Li Qiao and Richard Brualdi for bringing these problems to my attention.

7. References

1. Bengt Aspvall and Richard E. Stone, "Khachiyan's linear programming algorithm". *J. Algorithms* 1 (1980), 1-13.

2. C. Berge, C. C. Chen, V. Chvatal, and C. S. Seow, "Combinatorial properties of polyominoes". *Combinatorica* 1 (1981), 217-224.

3. Michael R. Garey and David S. Johnson, *Computers and Intractability: A Guide to NP-Completeness.* W. H. Freeman, San Francisco, 1979.

4. Solomon W. Golomb, *Polyominoes.* Charles Scribner's Sons, New York, 1965.

Figures

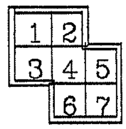


Figure 1
A simple polyomino

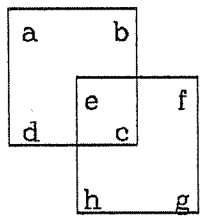
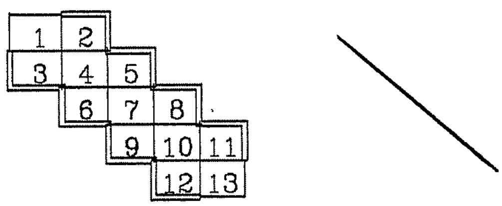
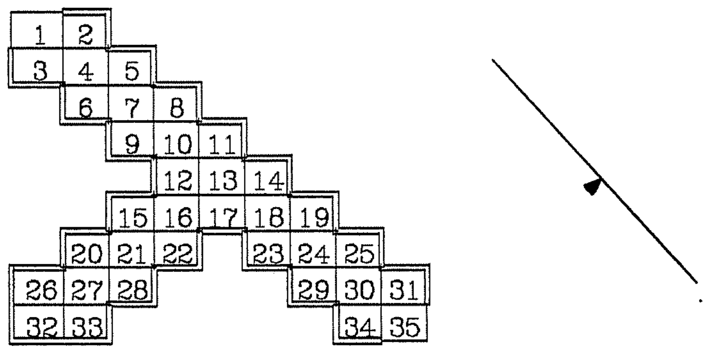


Figure 2
Overlapping rectangles



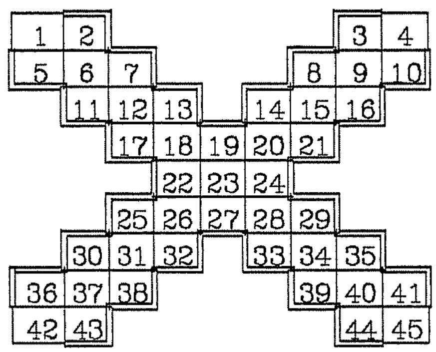
(a) (b)

Figure 3
Wire

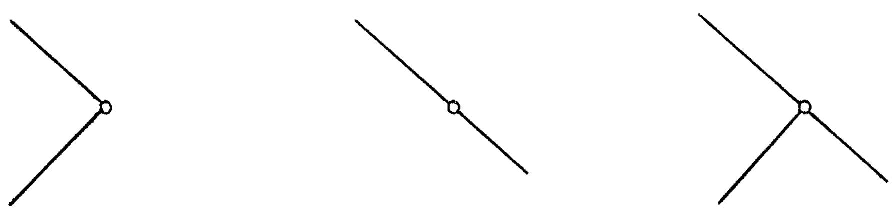


(a) (b)

Figure 4
Tab

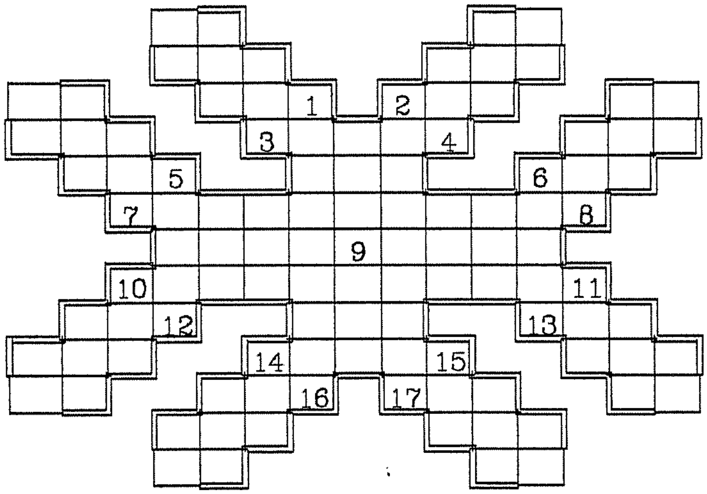


(a)

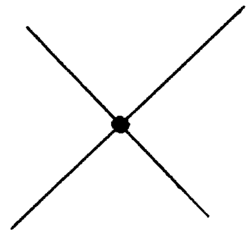


(b) (c) (d)

Figure 5
Turn/Inverter

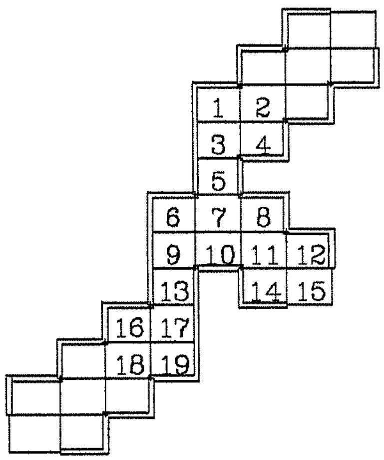


(a)

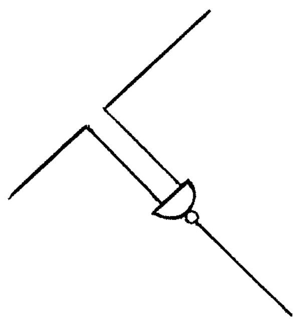


(b)

Figure 6
Crossover

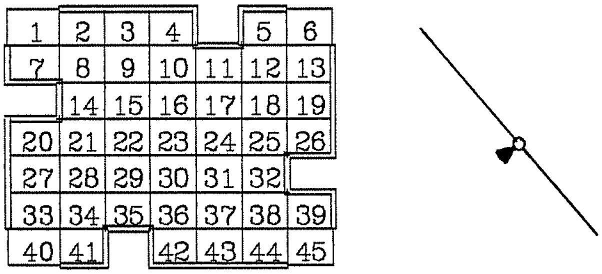


(a)



(b)

Figure 7
NAND gate



(a) (b)

Figure 8
Supertab

Figure 9
A polyomino with
no stable transversal

Figure 10
A polyomino with
no stochastic function

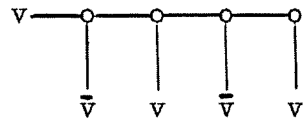


Figure 11
Variable component

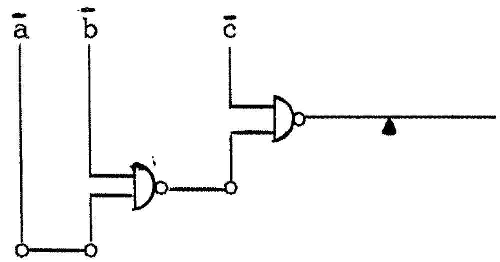


Figure 12
Clause component

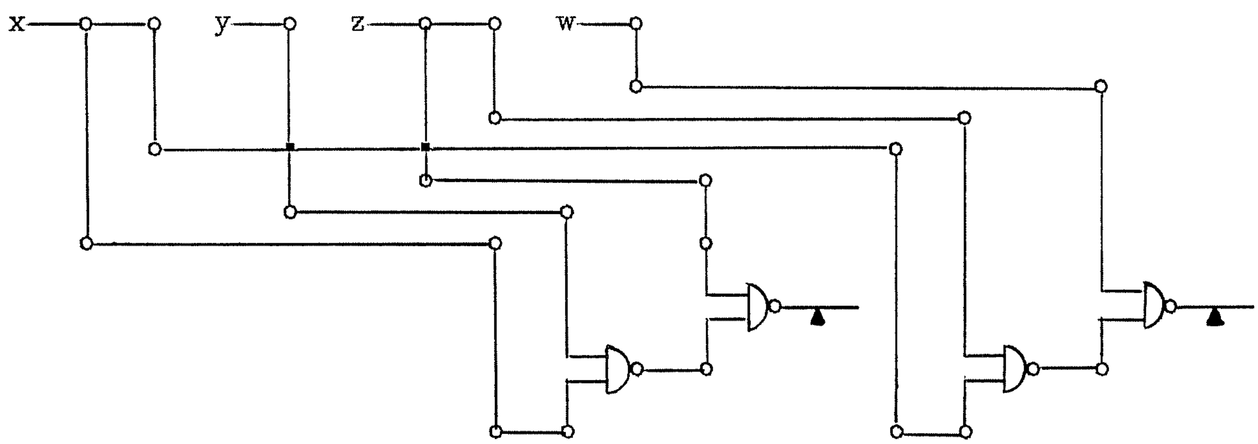


Figure 13
Polyomino constructed for
$(x \vee \bar{y} \vee z)(\bar{x} \vee \bar{z} \vee w)$