MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Computation Structures Group Memo 94

A Petri Net Version of Rabin's Undecidability Proof for Vector Addition Systems

by

Michel Hack

This work was supported by the National Science Foundation under research grant GJ-34671.

December 1973

Introduction

R. Karp and R. Miller [7] introduced Vector Addition Systems to answer certain decidability questions about their Parallel Program Schemata, and M. Rabin showed that a particular problem about Vector Addition Systems was undecidable: is the Reachability Set of one Vector Addition System a subset of the Reachability Set of some other given Vector Addition System. Rabin's first proof in 1967 used exponential polynomials [2]; at that time Hilbert's 10th Problem [4] had not yet been shown to be undecidable.

In 1970, Matijasevič [9] proved that Hilbert's 10th Problem was undecidable, and thus permitted a technically simpler proof of Rabin's result. Rabin never published his proof, but in 1972 he presented his new proof in a talk at MIT, an account of which can be found in [1].

Vector Addition Systems and Petri nets can fully represent each other [3,8] thus Rabin's result also gives us an undecidable problem about Petri nets. Furthermore, we believe that the graphical character of the Petri net model permits an easier exposition of the undecidability result.

Theorem: Given two Petri nets having the same number of places, each with a given initial marking, it is undecidable in general whether every marking reachable in one net is also reachable in the other.

Proof: We show that, given an arbitrary polynomial $P(x_1, \ldots, x_r)$ of r variables with integer coefficients, there exists a pair of Petri nets such that the set of reachable markings of one is a subset of the reachable markings of the other if and only if the polynomial P has an integral root. Thus, if we could decide for any two Petri nets whether in fact the set of reachable markings of one is a subset of the reachable markings of the other, we could also decide whether an arbitrary polynomial with integral coefficients has an integral root. But this is Hilbert's 10th Problem, which has been shown to be undecidable by Matijasevič.

Actually, we use the following equivalent form of Hilbert's 10th problem:

Lemma a: Given two polynomials of r variables with non-negative integer coefficients $P(\bar{x})$ and $Q(\bar{x})$ such that, $\forall \bar{x} \in \mathbb{N}^r$: $P(\bar{x}) \geq Q(\bar{x})$, it is undecidable whether there exists a solution $\bar{x} \in \mathbb{N}^r$ to $P(\bar{x}) = Q(\bar{x})$.

Proof of Lemma a: Let $R(\bar{x})$ be an arbitrary polynomial with r variables. Then $R(\bar{x}) = 0$ has a solution in $\mathbb{Z}^r$ if and only if one of the 2^r polynomials obtained from R by replacing some of the variables by their negative has a root in $\mathbb{N}^r$. Thus a finite number of tests for non-negative integer roots is enough to find any integer root of R.

Now, let $R_1(\bar{x})$ be a polynomial for which we check for roots in $\mathbb{N}^r$. Let $R_2(\bar{x}) = (R_1(\bar{x}))^2$. Then we have:

$\forall \bar{x} \in \mathbb{N}^r$: $R_2(\bar{x}) \geq 0$, and the roots of R_2 are clearly roots of R_1 and vice versa.

Now, we separate positive and negative coefficients of R_2:

$$R_2(\bar{x}) = P(\bar{x}) - Q(\bar{x}) \geq 0$$

where P and Q are polynomials with non-negative coefficients and clearly satisfy the conditions of the Lemma.

First, we shall show how to get a Petri net to behave like a polynomial.

Lemma b: Given a polynomial with non-negative integer coefficients of r variables, $P(x_1, \ldots, x_r)$, there exists a Petri net with $r+1$ distinguished places such that the set of all markings reachable in these distinguished places is the set $\{(x_1, \ldots, x_r, z) | x_i \in \mathbb{N} \ \& \ 0 \leq z \leq P(x_1, \ldots, x_r)\}$

There may be many more places in this Petri net than just these distinguished places, but for the moment we disregard their markings.

As an example, consider the following net, which can be seen to correspond to the polynomial of one variable $P(x) = x+1$:

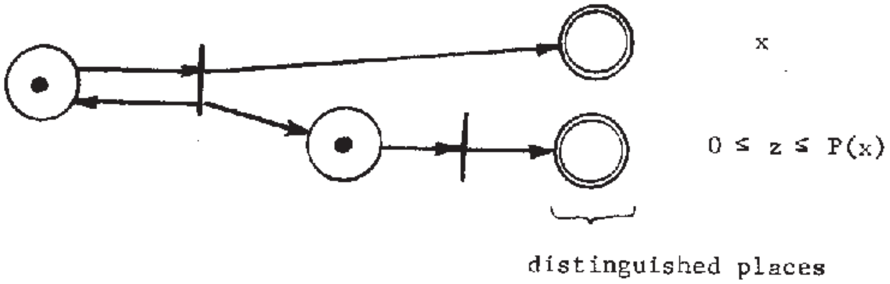


Figure 1

The possible markings for the distinguished places are:

x z

⟨0, 0⟩	⟨1, 0⟩	⟨2, 0⟩	
⟨0, 1⟩	⟨1, 1⟩	⟨2, 1⟩	
	⟨1, 2⟩	⟨2, 2⟩	
		⟨2, 3⟩	etc.

The relation to the graph of P(x) is obvious: The reachable markings can be represented by the integral points below or on the graph:

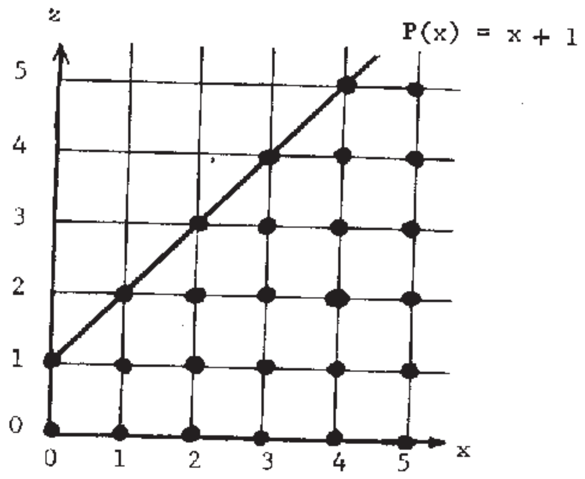


Figure 2

<u>Proof</u> <u>of</u> <u>Lemma</u> <u>b</u>: We shall show how to construct such a net, given a polynomial P with r variables $x_1, \ldots, x_r$.

The general structure is shown below:

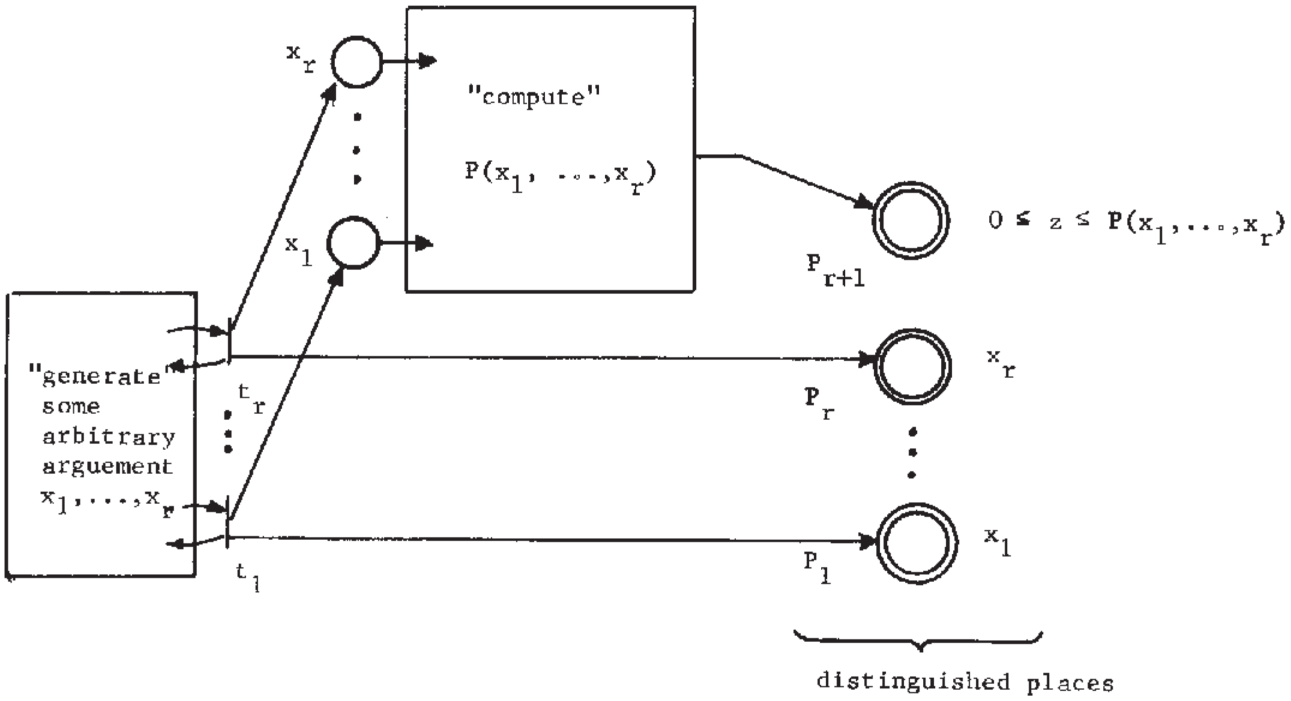


Figure 3

The generation part is easy to build:

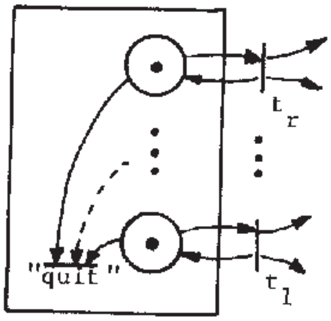


Figure 4. "generate"

Each transition t_i fires some number (possibly zero) of times, generating a value for x_i in two copies (one for the "computer," one for the corresponding distinguished place), then the "generator" quits. The "argument" part of the distinguished marking is now established, and will not be altered.

The "computer" is a Petri net which, for a given "argument" $x_1, \ldots, x_n$, tries to compute $P(x_1, \ldots, x_n)$. However, for its output place z, $P(x_1, \ldots, x_r)$ is only an upper bound: No firing sequence can possibly put more tokens on z, but there exists a firing sequence which does put $P(x_1, \ldots, x_r)$ tokens on z. It does not matter if some other firing sequence kills the net before the bound is reached.

Rabin calls such a computation by upper bounds "weak computation," and we are about to show that polynomials with non-negative coefficients are weakly computable by Petri nets.

Polynomials are computed by the operations of addition of two numbers, multiplication of two numbers, and substitution of previous results into one or several new additions or multiplications. Now, since, for positive integers, each of the operations add, multiply, copy is non-decreasing as a function of its arguments, if we substitute a reachable upper bound for its arguments, the result will also be a reachable upper bound.

Also, we shall make sure that the reachable upper bound can be approached one token at a time, so that the possible markings of the "result" place include all integers from zero to the bound included.

The add and copy operations can be represented by a Petri net as follows:

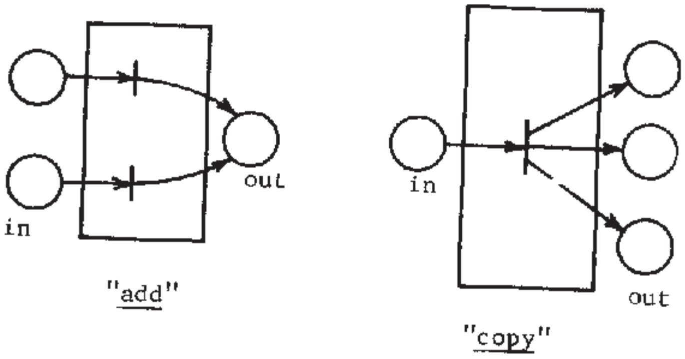


Figure 5

And the following Petri net has a reachable upper bound of x • y in its output place:

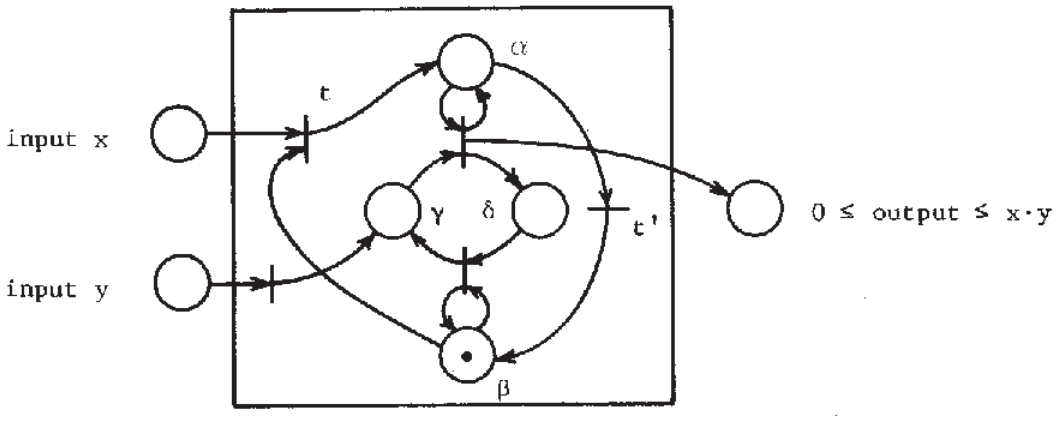


Figure 6. "multiply"

It can be seen that the following strategy yields $x \cdot y$ tokens at the output, and that this cannot be exceeded, though it is possible to exhaust x and thus grind to a halt by firing only t and t', not producing any tokens at the output. The maximum output strategy is: Transfer all y tokens into γ, fire t, transfer all of γ into δ (at this point we have y tokens at the output, $x - 1$ at the input), then firing t' and bring all y tokens back to γ, and repeat this for the remaining $x - 1$ tokens. t can fire only x times, and at most y tokens can be transferred to the output between firings of t.

Having thus shown that addition, multiplication and substitution are weakly computable by Petri nets (and argued that substitution in fact preserves weak computability), we can now construct a Petri net that weakly computes a polynomial, say $3x^2 + 2xy + y^3$, by interconnecting the Petri nets weakly computing add, copy, and multiply, as shown in Figure 7.

Example: Compute

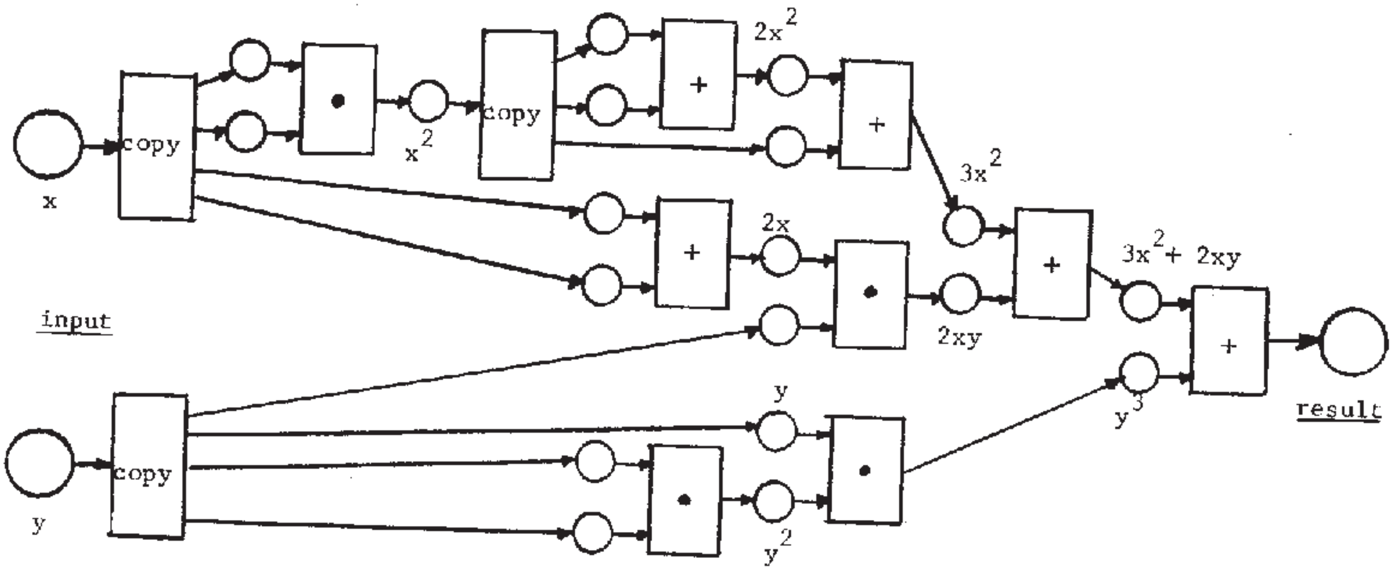


Figure 7. "Compute" $3x^2 + 2xy + y^3$

QED

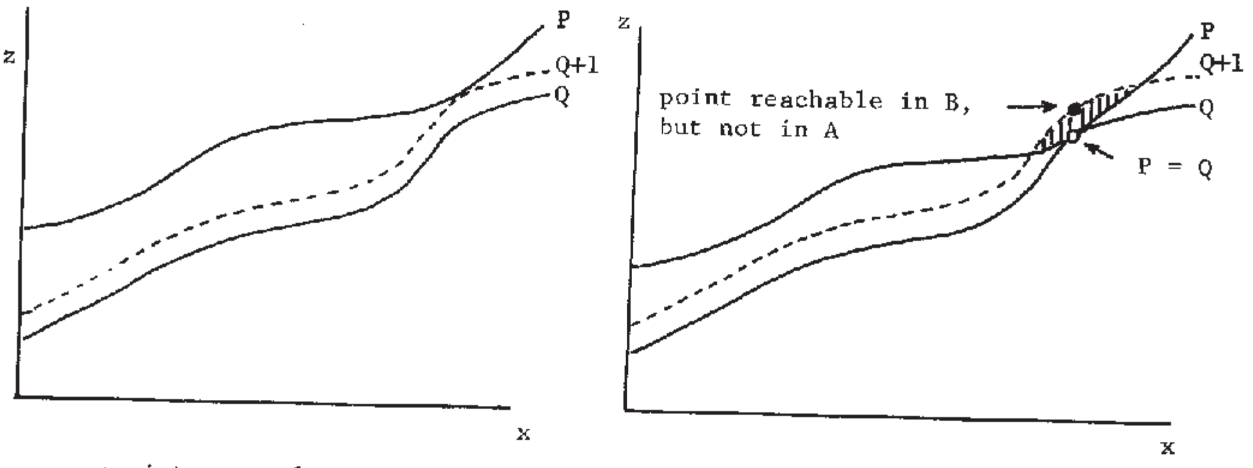
Now we will show how to construct two Petri nets, A and B, such that every marking reachable by A is also reachable by B if and only if there exists a collection of non-negative integers $x_1, \ldots, x_r$ such that, for two given polynomials P and Q as described in Lemma a, we have:

$$P(x_1, \ldots, x_r) = Q(x_1, \ldots, x_r)$$

Since $P(\bar{x}) \geq Q(\bar{x})$, we have:

$$(\forall \bar{x} \in \mathbb{N}^r) \quad (P(\bar{x}) = Q(\bar{x}) \Longleftrightarrow P(\bar{x}) < Q(\bar{x}) + 1)$$

As far as the graphs of P and Q + 1 in (r + 1)-space are concerned, it means that the graph of P "dips under"[*] the graph of Q + 1 if and only if P = Q has a solution:



no integer solution

integer solution

Figure 8

Now let A' and B' be Petri nets corresponding to the polynomials P and Q + 1 according to Lemma b. Every marking of the set of r + 1 distinguished places of B' is reachable as a marking of the distinguished set of r + 1 places of A' except if the graph of P "dips under" the graph of Q, i.e. if there is an integer solution to the equation P = Q. Yet we want to have two Petri nets A and B where every marking of B is reachable by A if and only if there is no solution to P = Q; we want to compare the markings of two complete nets, not just for a subset of the places.

What remains to be done is to modify A' and B' into two nets A and B of same number of places n, such that every marking of B is reachable in A except if the

[*] Enough for the "dip" (shaded area in Fig. 8) to contain an integral point.

marking of the distinguished places of B' cannot be reached by the distinguished places of A'.

As a first step, we add enough extra blank places, not connected to any existing transition, to one of the nets, in order to get two nets of the same number of places $n-2$, then we add two more places α,β to each net. These are all the places in A resp. B. In B, let α be blank and β be marked with one token; neither place is connected to any transition. This completes B, which thus differs from B' in only a few disconnected places. In A, however, we insert a transition from α to β, and we let place α be in a self-loop on every transition of A. We let α be originally marked with one token, and β be initially blank. Thus, as long as the token is in α, A behaves just like A', but when the token transfers to β, all transitions become permanently disabled, and in particular, the marking in the $r+1$ distinguished places will be frozen.

Now, for each of the $n-2-(r+1)$ undistinguished places of A, we add two transitions, one of which puts a token on the place, the other removes a token from it; then we put all these new transitions in self-loops on place β. Thus, after the token from α is transferred to β, any marking can be reached in the undistinguished places of A by firing these extra transitions a suitable number of times.

To see how this construction works, let us see under what conditions every marking reachable in B can also be reached in A.

Let us label the places as follows: $p_1, \ldots, p_r$ are the places containing the argument for the polynomial, p_{r+1} contains a partial result of the computation. These are the $r+1$ distinguished places. For the sake of argument, let the number of places of B' be the smaller number k, and the number of places of A' be $n-2>k$. We add $n-2-k$ undistinguished places to B'. Let us label the undistinguished places of A and B $p_{r+2} \cdots p_{n-2}$, and let us label α and β, p_n and p_{n-1}, respectively. (See figure 9.)

For comparing markings in A and B, we pair the places according to their labels p_i. Now, any marking of B will be, by construction, of the following form, where $z \le Q(x_1, \ldots, x_n)+1$, and the y_i could have any values.

$$\begin{array}{ccccccccc} p_1 & \cdots & p_r & p_{r+1} & p_{r+2} & \cdots & p_{n-2} & p_{n-1} & p_n \\ \langle x_1, & \ldots, & x_r, & q, & y_1, & \ldots, & y_{n-r-3}, & 1, & 0\rangle \end{array}$$

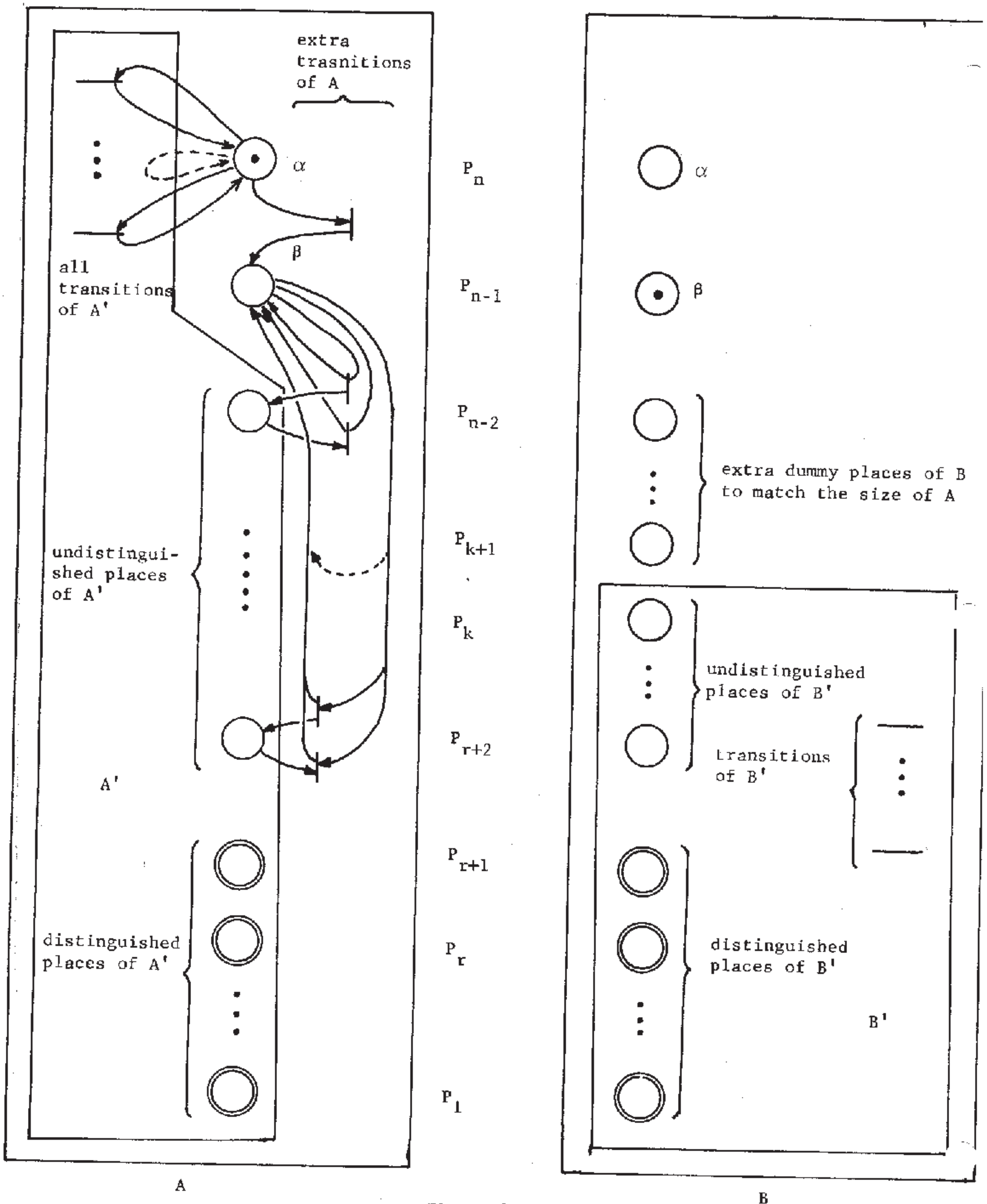

extra
trasnitions
of A
α
β
all
transitions
of A'
undistingui-
shed places
of A'
A'
distinguished
places of A'
A
P_n
P_{n-1}
P_{n-2}
P_{k+1}
P_k
P_{r+2}
P_{r+1}
P_r
P_1
α
β
extra dummy places of B
to match the size of A
undistinguished
places of B'
transitions
of B'
distinguished
places of B'
B'
B

Figure 9.

To reach this marking in A, we must first try to match $p_1, \ldots, p_{r+1}$, since after we match p_{n-1} and p_n, we will have frozen the marking of the distinguished places of A. Therefore, we first generate the argument $x_1, \ldots, x_r$ for polynomial P, then partially compute $P(x_1, \ldots, x_r)$ in a way that, if completed, would actually yield $P(x_1, \ldots, x_n)$ tokens in p_{r+1} of A. But we stop as soon as we reach z, the marking we try to match in p_{r+1} of B. This is possible if and only if $P(x_1, \ldots, x_r) \geq z$, which in turn could fail only if $z = Q(x_1, \ldots, x_r) + 1$ and in fact $P(x_1, \ldots, x_r) = Q(x_1, \ldots, x_r)$. Suppose we could reach z in p_{r+1} of A. As soon as we do, we switch off all transitions of A' by transferring the token from $\alpha(p_n)$ to $\beta(p_{n-1})$, at the same time matching the marking in these two places to the one in B. But now, we can reach any marking we wish in $p_{r+2}, \ldots, p_{n-2}$ of A, by firing the extra transitions of A a suitable number of times; in particular, we can match $y_1, \ldots, y_{n-r-3}$, thus reaching in A the proposed marking of B. As we pointed out, this can be carried out for all markings of B except one where we have:

$$z = Q(x_1, \ldots, x_r) + 1 = P(x_1, \ldots, x_r) + 1$$

But such a marking is reachable in B if and only if the above equation does have a solution in non-negative integers. Thus:

$$(\forall \bar{x} \in \mathbb{N}^r)\, P(\bar{x}) \neq Q(\bar{x}) \iff \text{every marking reachable in B is also reachable in A}$$

QED

References

1. Baker, H. G., Jr. Rabin's Proof of the Undecidability of the Reachability Set Inclusion Problem of Vector Addition Systems. CSG Memo 79, Project MAC, MIT, July 1973.

2. Davis, M., H. Putnam, and J. Robinson. The decision problem for exponential diophantine equations. Annals of Math., Vol. 74, No. 3 (November 1961), pp 425-436.

3. Hack, M. Decision Problems for Petri Nets and Vector Addition Systems, CSG Memo 95, Project MAC, MIT, in preparation.

4. Hilbert, D. Mathematische Probleme. Vortag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris 1900. Nachr. K. Ges. Wiss. Göttingen, Math.-Phys. Kl. 1900, pp 253-297. Translation: Bull. Amer. Math. Soc., 8 (1901-1902), pp 437-479.

5. Holt, A. W. Final Report of the Information System Theory Project. Technical Report RADC-TR-68-305, Rome Air Development Center, Griffiss Air Force Base, New York, 1968.

6. Holt, A. W., and F. Commoner, Events and Conditions. Applied Data Research, New York 1970.

7. Karp, R. M. and R. E. Miller. Parallel program schemata: A mathematical model for parallel computation. IEEE Conference Record, Eighth Annual Symposium on Switching and Automata Theory, October 1967, pp 55-61.

8. Keller, R. M. Vector Replacement Systems: A Formalism for Modeling Asynchronous Systems. Technical Report 117, Princeton University, Computer Science Laboratory, December 1972.

9. Matijasevič, Ju. V. Enumerable sets are diophantine. Soviet Math. Dokl. Vol. 11, No. 2 (1970), pp 354-357.

10. Petri, C. A. Communication With Automata. Supplement 1 to Technical Report RADC-TR-65-377, Vol. 1, Griffiss Air Force Base, New York, 1966. Originally published in German: Kommunikation mit Automaten, University of Bonn, 1962.

11. Rabin, M. Private communication, Fall 1972.