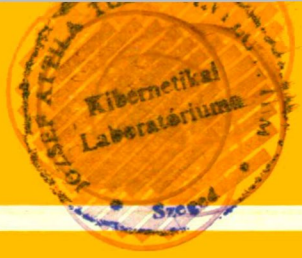


Tomus 4. Fasciculus 2.

ACTA CYBERNETICA

FORUM CENTRALE PUBLICATIONUM
CYBERNETICARUM HUNGARICUM

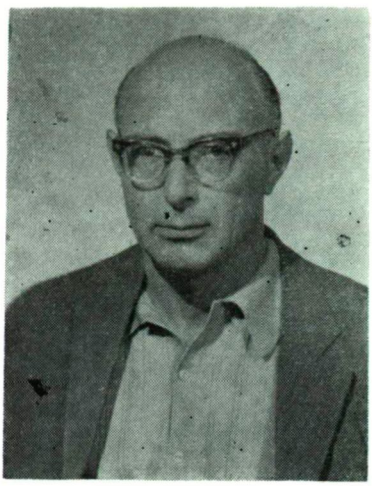
Szeged, 1979
Curat: Universitas Szegediensis de Attila József nominata

4. kötet 2. füzet

ACTA CYBERNETICA

A HAZAI KIBERNETIKAI KUTATÁSOK
KÖZPONTI PUBLIKÁCIÓS FÓRUMA

Szeged, 1979. március
A Szegedi József Attila Tudományegyetem gondozásában



TAMÁS FREY
1927—1978

To our regret we learned that Tamás Frey, member of the editorial board of these Acta prematurely died on April 9, 1978. He held the Chair of Matmatics at the Budapest Technological University and was a Doctor of Matematical Sciences. In him we lost an interesting and many-sided individual.

He obtained significant results in several branches of mathematics. His first area of interest was classical analysis; later he turned to the theoretical problems of abstract automata and computer science.

We cherish his memory.

On some types of incompletely specified automata

By M. K. Chirkov

1. Preliminaries

In this paper the most general definition of an incompletely specified (or *partial*) finite automaton (generalized, probabilistic and deterministic) is proposed and some special classes of such automata are introduced. The conceptions of this paper are the further development of the author's ideas, stated in the book [1]. The known notions of partial finite automata (for example [1], [2], [3] and [4]) are included in the proposed definitions as exeptional cases. For the notations and notions that will not be defined here, the author refers to the books [1] and [5].

First of all it is useful to recall some definitions of the completely specified finite automata theory [1] and [5], and introduce some further notations.

By an *alphabet* X we mean a finite non-empty ordered set of elements. A finite sequence $X^{(t)}=X_{s_1}X_{s_2}\dots X_{s_t}$ $(X_{s_i}\in X,\ t\geqq 0)$ is called a *word* over X, and $t=|X^{(t)}|$ is the *length* of $X^{(t)}$. We use the notations X^* and X^t for the set of all words over X and for the set of all words of length t over X, respectively. Besides the following notations are used for the sets of all real numbers, vectors and matrices:

$$\mathscr{R}=(-\infty,\infty),\quad \mathscr{R}^m=\{r|r=(r_1,r_2,\dots,r_m),\ r_i\in\mathscr{R},\ i=\overline{1,m}\},$$

$$\mathscr{R}^{m,n}=\{R|R=(r_{ij})_{m,n},\ r_{ij}\in\mathscr{R},\ i=\overline{1,m},\ j=\overline{1,n}\}.$$

A vector is called *stochastic* (or *probabilistic*) if all its entries are non-negative and the sum of its entries is equal to 1. A matrix is called *stochastic* (or *probabilistic*) if all its rows are stochastic vectors. A stochastic vector is called *degenerate* if one of its entries is 1 and the other are equal to 0. A stochastic matrix is *degenerate* if all its rows are degenerate stochastic vectors. The following notations are used for the sets of all stochastic (degenerate stochastic) m-dimensional vectors and $(m\times n)$-matrices:

$$\mathscr{P}^m=\{p|p=(p_1,p_2,\dots,p_m),\ p_i\in[0,1],\ i=\overline{1,m},\ \sum_i p_i=1\},$$

$$\mathscr{D}^m=\{d|d=(d_1,d_2,\dots,d_m),\ d_i\in\{0,1\},\ i=\overline{1,m},\ \sum_i d_i=1\},$$

$$\mathscr{P}^{m,n}=\{P|P=(p_{ij})_{m,n},\ p_{ij}\in[0,1],\ \sum_j p_{ij}=1,\ i=\overline{1,m},\ j=\overline{1,n}\},$$

$$\mathscr{D}^{m,n}=\{D|D=(d_{ij})_{m,n},\ d_{ij}\in\{0,1\},\ \sum_j d_{ij}=1,\ i=\overline{1,m},\ j=\overline{1,n}\}.$$

Let $X=\{X_1, X_2, \ldots, X_n\}$, $A=\{A_1, A_2, \ldots, A_m\}$, $Y=\{Y_1, Y_2, \ldots, Y_k\}$ be the alphabets of inputs, states and outputs, respectively. Then a *finite generalized automaton* is a system

$$\mathbf{A}_{\text{gen}} = \langle X, A, Y, r^{(0)}, R\rangle \tag{1}$$

where $r^{(0)} \in \mathscr{R}^m$ is the initial vector and R ($\in \mathscr{R}^{nm,km}$) is the transition-output matrix, which presents a mapping of $X\times A\times Y\times A$ into the set of real numbers $\mathscr{R}$. The matrix R is usually represented by a combination of its nk square submatrices $\{R(X_s, Y_l)\}$ such that

$$R = \begin{pmatrix} R(X_1, Y_1) & R(X_1, Y_2) & \ldots & R(X_1, Y_k) \\ \ldots & \ldots & \ldots & \ldots \\ R(X_n, Y_1) & R(X_n, Y_2) & \ldots & R(X_n, Y_k) \end{pmatrix}.$$

In this case it may be said that R presents a mapping of $X\times Y$ into $\mathscr{R}^{m,m}$. The domain of this mapping is extended from $X\times Y$ to $(X\times Y)^t$ ($t=1, 2, \ldots$), where

$$(X\times Y)^t = \{(X^{(t)}, Y^{(t)}) | X^{(t)} \in X^t,\ Y^{(t)} \in Y^t\}$$

and

$$R(X^{(t)}, Y^{(t)}) = \prod_{i=1}^{t} R(X_{s_i}, Y_{l_i}),$$

with

$$X^{(t)} = X_{s_1} X_{s_2} \ldots X_{s_t},\quad Y^{(t)} = Y_{l_1} Y_{l_2} \ldots Y_{l_t}.$$

The *generalized mapping* Φ induced by a generalized automaton $\mathbf{A}_{\text{gen}}$ (in notation: $\Phi \vdash \mathbf{A}_{\text{gen}}$) is the mapping of

$$(X\times Y)^* = \{(X^{(t)}, Y^{(t)}) | X^{(t)} \in X^t,\ Y^{(t)} \in Y^t,\ t = 0, 1, \ldots\},$$

into $\mathscr{R}$ defined by

$$\Phi(X^{(t)}, Y^{(t)}) = r^{(0)} \prod_{i=1}^{t} R(X_{s_i}, Y_{l_i}) e,$$

where e is the m-dimensional column vector whose each entry is 1.

Hereafter we use the term automaton to mean a finite automaton.

A *probabilistic automaton*

$$\mathbf{A}_{pr} = \langle X, A, Y, p^{(0)}, P\rangle \tag{2}$$

is a generalized automaton (1) such that $r^{(0)} = p^{(0)} \in \mathscr{P}^m$ and $R = P \in \mathscr{P}^{nm,km}$. $p^{(0)}$ is called the *initial probabilistic distribution* on the state set A and P is called the *transition-output probability matrix* of the automaton $\mathbf{A}_{pr}$. The elements of P are treated as

$$p_{si,lj} = Pr(Y_l A_j | X_s A_i).$$

A probabilistic automaton $\mathbf{A}_{pr}$ induces the *probabilistic mapping* Φ of $(X\times X)^*$ into the closed real interval [0, 1] defined by

$$\Phi(X^{(t)}, Y^{(t)}) = Pr(Y^{(t)} | X^{(t)}) = p^{(0)} \prod_{i=1}^{t} P(Y_{l_i} | X_{s_i}) e,$$

where $P(Y_{l_i}|X_{s_i})$ is the proper square submatrix of R.

A *deterministic automaton*

$$\mathbf{A}_{\text{det}} = \langle X, A, Y, d^{(0)}, D \rangle$$

is a probabilistic automaton (2) such that $p^{(0)} = d^{(0)} \in \mathscr{D}^m$ and $P = D \in \mathscr{D}^{nm,km}$. If $d^{(0)} = (d_1, d_2, \ldots, d_m)$, $d_j = 1$, $d_i = 0$, $i \neq j$, then A_j is called the initial state of $\mathbf{A}_{\text{det}}$. A deterministic automaton $\mathbf{A}_{\text{det}}$ with the initial state A_j induces the *deterministic mapping*

$$\Phi_j \colon X^* \to Y^*$$

given by

$$\Phi_j(X^{(t)}) = Y^{(t)} \Leftrightarrow d^{(0)} \prod_{i=1}^{t} D(X_{s_i}, Y_{l_i}) e = 1.$$

2. Partial vectors, matrices and automata

Hereafter we use the term "partial" to mean "incompletely specified". In accordance with the classical automata theory an automaton $\mathbf{A}_{\text{gen}}$ ($\mathbf{A}_{pr}$ or $\mathbf{A}_{\text{det}}$) is partial if some of the elements of $r^{(0)}$, R ($p^{(0)}$, P or $d^{(0)}$, D) are undefined and represented by "—" ([2], [3] and [4]). The conditions under which this occurs are usually treated as "don't care conditions" when either some combinations of input and present state never occur or the output (the next state) is of no concern for some combinations of input and present state. Such an incomplete specification is usually interpreted to mean that the designer may use these incomplete specifications in arbitrary way to his advantage in obtaining a completely specified automaton. It is clear that such an interpretation of partial automata is not universal and does not embrace many interesting (as theoretical, so practical) cases. For example, there are many such problems that an incomplete specification of an automaton is the result of our ignorance of its exact structure or is the effect of the opportunity to choose its structure from a certain restricted class of structures. As a rule in practice there are not free choises of the indeterminate elements of $r^{(0)}$, R ($p^{(0)}$, P or $d^{(0)}$, D) and the various ways of their specification are closely interdependent. Thus it will be usefull to offer the most general interpretation of partial automata.

Some more general classes of partial probabilistic vectors, matrices and automata were proposed and studied by the author in the book [1]. Now we are going to make the furthermost generalization of the concept of partial vectors, matrices and automata. The main idea of this generalization is that any partial object (vector matrix, automaton) may be treated as a set of completely specified objects (vectors, matrices, automata) which are the results of various ways of its specification. Thus it is possible to describe this partial object by means of a set of objects and to investigate this set.

We shall now introduce the following general definitions. Any non-empty subset $\tilde{r}$ of the set $\mathscr{R}^m$ is called a *partial m-dimensional vector*. Any non-empty subset $\tilde{R}$ of the set $\mathscr{R}^{m,n}$ is called a *partial* $(m \times n)$*-matrix*. For instance, the partial $(m \times m)$-matrix

$$\tilde{R} = \{R | R \in \mathscr{R}^{m,m},\ |R| \in (0, 2]\}$$

is the subset of those $(m \times m)$-matrices whose determinants have values lying in the interval $(0, 2]$.

A *partial generalized automaton* is a system

$$\tilde{\mathbf{A}}_{\text{gen}} = \langle X, A, Y, \tilde{r}^{(0)}, \tilde{R} \rangle \tag{3}$$

where X, A, Y are as usual the alphabets of inputs, states and outputs, $\tilde{r}^{(0)}$ ($\subseteq \mathscr{R}^m$) is a partial initial vector and $\tilde{R}$ ($\subseteq \mathscr{R}^{nm,km}$) is a partial transition-output matrix. A partial generalized automaton (3) defines the set of completely specified generalized automata (1) such that

$$\mathbf{A}_{\text{gen}} \in \tilde{\mathbf{A}}_{\text{gen}} \Leftrightarrow r^{(0)} \in \tilde{r}^{(0)} \quad \& \quad R \in \tilde{R}.$$

By the *partial generalized mapping* $\tilde{\Phi}$ induced by $\mathbf{A}_{gen}$ we mean the following set of mappings of $(X \times Y)^*$ into $\mathscr{R}$:

$$\tilde{\Phi} = \{\Phi | \Phi \vdash A_{\text{gen}}, A_{\text{gen}} \in \tilde{A}_{\text{gen}}\}.$$

3. Partial p-vectors, p-matrices, p-automata

In accordance with above definitions any non-empty subset $\tilde{p}$ of the set $\mathscr{P}^m$ is called a *partial probabilistic vector*, or shortly, a *partial p-vector*. Any non-empty subset $\tilde{P}$ of the set $\mathscr{P}^{m,n}$ is called a *partial probabilistic* $(m \times n)$*-matrix*, or shortly, a *partial p-matrix*. Thus, any partial vector $\tilde{r}$ (matrix $\tilde{R}$) is a partial p-vector (p-matrix) if and only if all $r \in \tilde{r}$ ($R \in \tilde{R}$) are stochastic.

A *partial probabilistic automaton* (a *partial p-automaton*) is a system

$$\tilde{\mathbf{A}}_{pr} = \langle X, A, Y, \tilde{p}^{(0)}, \tilde{P} \rangle$$

where $\tilde{p}^{(0)} \subseteq \mathscr{P}^m$, $\tilde{P} \subseteq \mathscr{P}^{nm,km}$ and

$$\mathbf{A}_{pr} \in \tilde{\mathbf{A}}_{pr} \Leftrightarrow p^{(0)} \in \tilde{p}^{(0)} \quad \& \quad P \in \tilde{P}.$$

So far we have said nothing about methods of specification of $\tilde{r}^{(0)}$, $\tilde{p}^{(0)}$, $\tilde{R}$, $\tilde{P}$. As it was shown in [1] some problems of abstract theory of partial automata may be investigated without indication of such a concrete specification method. But there are many problems which may be solved only if this method is given. Many different types of partial vectors, matrices and automata may be constructed by various methods of specification of $\tilde{r}^{(0)}$, $\tilde{p}^{(0)}$, $\tilde{R}$ and $\tilde{P}$. Some of them will be introduced hereinafter.

4. Partial f-vectors, f-matrices, f-automata

Let $\xi_1, \xi_2, \ldots, \xi_q$ be q independent parameters and $\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_q$ be their domains. Let $f_i(\xi_1, \xi_2, \ldots, \xi_q)$ $(i = \overline{1, m})$ be real single-valued functions. Then a partial vector

$$\tilde{r} = \{r | r = (r_1, r_2, \ldots, r_m),\ r_i = f_i(\xi_1, \xi_2, \ldots, \xi_q),\ i = \overline{1, m},\ \xi_v \in \tilde{\sigma}_v,\ v = \overline{1, q}\} \tag{4}$$

is called a *partial f-vector* and is presented as

$$\tilde{r} = (f_1(\{\xi_v\}), f_2(\{\xi_v\}), \ldots, f_m(\{\xi_v\}))\ (\xi_v \in \tilde{\sigma}_v,\ v = \overline{1, q})$$

where

$$f_i(\{\xi_v\}) = f_i(\xi_1, \xi_2, \ldots, \xi_q).$$

Accordingly, a partial matrix

$$\tilde{R} = \{R | R = (r_{ij})_{m,n},\ r_{ij} = f_{ij}(\{\xi_v\}),\ i = \overline{1, m},\ j = \overline{1, n},\ \xi_v \in \tilde{\sigma}_v,\ v = \overline{1, q}\},$$

where f_{ij} is a real single-valued function $(i = \overline{1, m},\ j = \overline{1, n})$, is called a *partial f-matrix* and is presented as

$$\tilde{R} = (f_{ij}(\{\xi_v\}))_{m,n}\ (\xi_v \in \tilde{\sigma}_v,\ v = \overline{1, q}). \tag{5}$$

For example,

$$\tilde{R} = \begin{pmatrix} \xi_2 + \sin \xi_1 & \sqrt{1 + \xi_1 \xi_2} \\ \xi_1 - \xi_2^2 & 2 \end{pmatrix} \quad \left(\xi_1 \in \left[\frac{1}{2}, 1\right), \quad \xi_2 \in \{0, 1, 2\}\right)$$

is a partial square *f*-matrix of order 2.

By substituting the different values of the parameters into f_i or f_{ij}, the various completely specified vectors or matrices of $\tilde{r}$ or $\tilde{R}$ may be found.

We say that a function $f(\{\xi_v\})$ *essentially depends* on the parameter ξ_v if there exist $b_1,\ b_2 \in \tilde{\sigma}_v$ such that

$$f(\xi_1, \ldots, \xi_{v-1}, b_1, \xi_{v+1}, \ldots, \xi_q) \neq f(\xi_1, \ldots, \xi_{v-1}, b_2, \xi_{v+1}, \ldots, \xi_q)$$

holds.

A partial *f*-vector (4) *essentially depends* on ξ_v if some of its elements essentially depends on ξ_v. Two partial *f*-vectors are called *indepedent* if there is no such parameter on which both *f*-vectors essentially depend.

If every two rows of a partial *f*-matrix are independent partial *f*-vectors then this matrix is called a *partial f-matrix with independent rows* and it may be represented in the form

$$\tilde{R} = (f_{ij}(\{\xi_v^{(i)}\}))_{m,n} \quad (\xi_v^{(i)} \in \tilde{\sigma}_v^{(i)},\ i = \overline{1, m},\ v = \overline{1, q_i}),$$

where all parameters are independent.

If every two columns of a partial *f*-matrix are independent partial *f*-vectors then this matrix is called a *partial f-matrix with independent columns*. Such a matrix may be represented in the form

$$\tilde{R} = (f_{ij}(\{\xi_v^{(j)}\}))_{m,n} \quad (\xi_v^{(j)} \in \tilde{\sigma}_v^{(j)},\ v = \overline{1, q_j},\ j = \overline{1, m}).$$

For example,

$$\tilde{R} = \begin{pmatrix} \xi_2^{(1)} + \sin \xi_1^{(1)} & \sqrt{1 + \xi_1^{(1)} \xi_2^{(1)}} \\ \xi_1^{(2)} + \cos \xi_2^{(2)} & 3\, \xi_1^{(2)} \xi_2^{(2)} \end{pmatrix},$$

where

$$\xi_1^{(1)} \in \left[\frac{1}{2}, 1\right), \quad \xi_2^{(1)} \in \{0, 1, 2\},$$

$$\xi_1^{(2)} \in \left[2, 3\frac{1}{4}\right), \quad \xi_2^{(2)} \in \left[\frac{\pi}{8}, \frac{\pi}{4}\right],$$

is a partial square *f*-matrix of order 2 with independent rows.

In accordance with above definitions a *partial generalized f-automaton* is a system

$$\tilde{\mathbf{A}}_{\text{gen}} = \langle X, A, Y, \tilde{r}^{(0)}, \tilde{R} \rangle, \tag{6}$$

$$\tilde{r}^{(0)} = (f_1(\{\xi_\nu\}), f_2(\{\xi_\nu\}), \ldots, f_m(\{\xi_\nu\})),$$

$$\tilde{R} = (f_{si,lj}(\{\xi_\nu\}))_{nm,km}, \quad \xi_\nu \in \tilde{\sigma}_\nu, \quad \nu = \overline{1,q}$$

where $f_i, f_{si,lj}$ are real single-valued functions defined on all $\xi_\nu (\in \tilde{\sigma}_\nu, \nu = \overline{1,q})$ and $\tilde{\sigma}_\nu$ $(\nu = \overline{1,q})$ are specified.

Let $\tilde{R}(X_s, Y_l)$ be a partial square f-submatrix of $\tilde{R}$ defined by

$$\tilde{R}(X_s, Y_l) = (f_{si,lj}(\{\xi_\nu\}))$$

$$i = \overline{1,m}, \quad j = \overline{1,m}.$$

Then the partial generalized mapping $\tilde{\Phi}$ (the set of mappings of $(X \times Y)^*$ into $\mathscr{R}$ induced by the partial generalized f-automaton (6) may be defined by

$$\tilde{\Phi}(X^{(t)}, Y^{(t)}) = \tilde{r}^{(0)} \prod_{i=1}^{t} \tilde{R}(X_{s_i}, Y_{l_i}) e \quad (\xi_\nu \in \tilde{\sigma}_\nu, \nu = \overline{1,q}).$$

5. Partial *pf*-vectors, *pf*-matrices, *pf*-automata

A partial f-vector (4) is probabilistic if and only if

$$0 \leqq f_i(\{\xi_\nu\}) \leqq 1 \quad \text{and} \quad \sum_i f_i(\{\xi_\nu\}) = 1 \quad (\xi_\nu \in \tilde{\sigma}_\nu, \nu = \overline{1,q}). \tag{7}$$

Such a partial f-vector is called a *partial pf-vector*. A partial f-matrix (5) is a *partial pf-matrix* if

$$0 \leqq f_{ij}(\{\xi_\nu\}) \leqq 1 \quad \text{and} \quad \sum_j f_{ij}(\{\xi_\nu\}) = 1 \quad (\xi_\nu \in \tilde{\sigma}_\nu, \nu = \overline{1,q}, i = \overline{1,m}). \tag{8}$$

For example,

$$\tilde{P} = \begin{pmatrix} \sin^2 \xi & \cos^2 \xi \\ \dfrac{2\xi}{\pi} & 1 - \dfrac{2\xi}{\pi} \end{pmatrix} \quad \left(\xi \in \left[\frac{\pi}{4}, \frac{\pi}{2}\right)\right)$$

is a partial square *pf*-matrix of order 2.

It is clear that there are no partial *pf*-matrices with independent columns, but we shall say that a partial *pf*-matrix $\tilde{P}$ is a *partial pf-matrix with minimal dependent columns* if there is a partial f-matrix $\tilde{R}$ with independent columns such that for every completely specified stochastic $(m \times n)$-matrix P,

$$P \in \tilde{P} \Leftrightarrow P \in \tilde{R}$$

holds.

A *partial probabilistic f-automaton* (i. e., a *partial pf-automaton*) is a system

$$\tilde{\mathbf{A}}_{pr} = \langle X, A, Y, \tilde{p}^{(0)}, \tilde{P} \rangle,$$

where

$$\tilde{p}^{(0)} = (f_1(\{\xi_v\}), f_2(\{\xi_v\}), \ldots, f_m(\{\xi_v\})),$$
$$0 \leqq f_i(\{\xi_v\}) \leqq 1, \quad \sum_i f_i(\{\xi_v\}) = 1 \tag{9}$$

and

$$\tilde{P} = (f_{si,lj}(\{\xi_v\}))_{nm,km},$$
$$0 \leqq f_{si,lj}(\{\xi_v\}) \leqq 1, \quad \sum_{lj} f_{si,lj}(\{\xi_v\}) = 1, \tag{10}$$
$$\xi_v \in \tilde{\sigma}_v, \quad v = \overline{1,q}, \quad s = \overline{1,n}, \quad i,j = \overline{1,m}, \quad l = \overline{1,k}.$$

6. Partial *l*-vectors, *l*-matrices, *l*-automata

A partial *f*-vector defined as

$$\tilde{r} = (f_1, f_2, \ldots, f_m), \quad f_i = \sum_v a_i^{(v)} \xi_v,$$
$$i = \overline{1,m}, \quad \xi_v \in \tilde{\sigma}_v, \quad v = \overline{1,q}$$

where $a_i^{(v)}$ ($v = \overline{1,q}$, $i = \overline{1,m}$) are real coefficients, is called a *partial l-vector*. A partial *f*-matrix defined as

$$\tilde{R} = (f_{ij})_{m,n}, \quad f_{ij} = \sum_v a_{ij}^{(v)} \xi_v,$$
$$i = \overline{1,m}, \quad j = \overline{1,n}, \quad \xi_v \in \tilde{\sigma}_v, \quad v = \overline{1,q},$$

is called a *partial l-matrix*. For example,

$$\tilde{R} = \begin{pmatrix} \xi_1 + 2\xi_2 & \xi_2 - \xi_1 & 3\xi_2 \\ 4 & 2\xi_1 + 1 & \xi_2 \end{pmatrix},$$
$$\xi_1 \in \left(\frac{3}{4}, 2\frac{1}{2}\right], \quad \xi_2 \in \left[2, 7\frac{1}{3}\right).$$

A *partial generalized l-automaton* is a system

$$\tilde{\mathbf{A}}_{\text{gen}} = \langle X, A, Y, \tilde{r}^{(0)}, \tilde{R} \rangle,$$

where

$$\tilde{r}^{(0)} = \left(\sum_v a_1^{(v)} \xi_v, \sum_v a_2^{(v)} \xi_v, \ldots, \sum_v a_m^{(v)} \xi_v\right),$$
$$\tilde{R} = \left(\sum_v a_{si,lj}^{(v)} \xi_v\right)_{nm,km},$$
$$\xi_v \in \tilde{\sigma}_v, \quad v = \overline{1,q}.$$

Accordingly, a partial *l*-vector (*l*-matrix, generalized *l*-automaton) is a partial *pl*-vector (*pl*-matrix, *pl*-automaton) if for all its entries $f_i(f_{ij}, f_i, f_{si,lj})$ the conditions (7) ((8), (9), (10)) hold. Some examples of partial *pl*-automata may be found in [1].

7. Partial $\tilde{\sigma}$-vectors, $\tilde{\sigma}$-matrices, $\tilde{\sigma}$-automata

A partial l-vector in form

$$\tilde{r} = (\xi_1, \xi_2, \ldots, \xi_q) \quad (\xi_v \in \tilde{\sigma}_v, \ v = \overline{1, q})$$

where $\tilde{\sigma}_v$ $(v = \overline{1, q})$ are defined subsets of $\mathscr{R}$, is called a *partial vector with independent elements*, or more briefly, a *partial $\tilde{\sigma}$-vector* and is specified as

$$\tilde{r} = (\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_m). \tag{11}$$

A partial l-matrix in form

$$\tilde{R} = (\xi_{ij})_{m,n} \quad (\xi_{ij} \in \tilde{\sigma}_{ij}, \ i = \overline{1, m}, \quad j = \overline{1, n}),$$

where $\tilde{\sigma}_{ij}$ $(i = \overline{1, m}, \ j = \overline{1, n})$ are defined subsets of $\mathscr{R}$, is called a *partial $\tilde{\sigma}$-matrix* and is specified as

$$\tilde{R} = (\tilde{\sigma}_{ij})_{m,n}, \tag{12}$$

i.e., in form of matrix whose elements are defined subsets of $\mathscr{R}$. For example,

$$\tilde{R} = \begin{pmatrix} \left[-\frac{1}{2}, 1\right] & \{0, 2\} & \left\{\frac{1}{4}, \frac{1}{2}\right\} \cup \left(\frac{2}{3}, 1\right] \\ \frac{3}{4} & \left(-\frac{1}{2}, \frac{1}{2}\right) & \left\{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}\right\} \\ 1 & 0 & \left\{\xi \mid \xi = \frac{1}{2^t}, \quad t = 1, 2, \ldots\right\} \end{pmatrix}.$$

It is useful to notice that each partial f-matrix with independent rows and columns may be represented in form of a partial $\tilde{\sigma}$-matrix.

Accordingly with these definitions a *partial generalized $\tilde{\sigma}$-automaton* is a system

$$\tilde{\mathbf{A}}_{\text{gen}} = \langle X, A, Y, \tilde{r}^{(0)}, \tilde{R} \rangle,$$

$$\tilde{r}^{(0)} = (\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_m), \quad \tilde{R} = (\tilde{\sigma}_{si,lj})_{nm,km}$$

where $\tilde{\sigma}_i$, $\tilde{\sigma}_{si,lj}$ are defined subsets of $\mathscr{R}$. If

$$\begin{gathered} \mathbf{A}_{\text{gen}} = \langle X, A, Y, r^{(0)}, R \rangle \\ r^{(0)} = (r_1, r_2, \ldots, r_m), \quad R = (r_{si,lj})_{nm,km} \end{gathered} \tag{13}$$

is a completely specified generalized automaton then

$$\mathbf{A}_{\text{gen}} \in \tilde{\mathbf{A}}_{\text{gen}} \Leftrightarrow r_i \in \tilde{\sigma}_i \quad \& \quad r_{si,lj} \in \tilde{\sigma}_{si,lj} \quad \text{for all } s, i, l, j.$$

8. Partial $p\tilde{\sigma}$-vectors, $p\tilde{\sigma}$-matrices, $p\tilde{\sigma}$-automata

A *partial p-vector with minimal dependent elements* is a subset of $\mathscr{P}^m$ defined as

$$\tilde{p} = \{p | p = (p_1, p_2, \ldots, p_m), p_i \in \tilde{\sigma}_i \subseteq [0, 1], \sum_i p_i = 1\}.$$

Such a partial p-vector is called a *partial $p\tilde{\sigma}$-vector* and is specified in form

$$\tilde{p} = (\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_m) \tag{14}$$

where $\tilde{\sigma}_i$ $(i=\overline{1, m})$ are defined subsets of [0, 1] and the condition $\sum_i p_i = 1$ is omitted as obvious.

A partial p-matrix defined as

$$\tilde{P} = \{P | P = (p_{ij})_{m,n},\ p_{ij} \in \tilde{\sigma}_{ij} \subseteq [0, 1], \sum_j p_{ij} = 1,\ i = \overline{1, m},\ j = \overline{1, n}\}$$

may be specified in form

$$\tilde{P} = (\tilde{\sigma}_{ij})_{m,n} \tag{15}$$

where $\tilde{\sigma}_{ij}$ $(i=\overline{1, m}, j=\overline{1, n})$ are defined subsets of [0, 1] and the conditions $\sum_j p_{ij} = 1$ $(i=\overline{1, m})$ are omitted as obvious. Such a partial p-matrix is called a *partial $p\tilde{\sigma}$-matrix*. It is clear that each partial pf-matrix with independent rows and minimal dependent columns may be specified in form of a partial $p\tilde{\sigma}$-matrix.

We say [1] that a partial $p\tilde{\sigma}$-vector (14) is *correctly specified* if $\tilde{\sigma}_i \neq \emptyset$ $(\tilde{\sigma}_i \subseteq [0, 1],\ i=\overline{1, m})$ and for each $p_j \in \tilde{\sigma}_j$ there exists $p_i \in \tilde{\sigma}_i$ $(i \neq j)$ such that $\sum_{s=1}^{m} p_s = 1$ $(j=\overline{1, m})$. A partial $p\tilde{\sigma}$-matrix is *correctly specified* if each of its rows is a correctly specified partial $p\tilde{\sigma}$-vector. For example,

$$\tilde{P} = \begin{pmatrix} \left\{0, \frac{1}{4}\right\} & \left[\frac{1}{4}, \frac{1}{2}\right) & \left(\frac{1}{4}, \frac{3}{4}\right] \\ 0 & \left\{\frac{1}{2}, \frac{3}{4}\right\} & \left\{\frac{1}{4}, \frac{1}{2}\right\} \end{pmatrix}$$

is a correctly specified partial $p\tilde{\sigma}$-matrix.

A *partial $p\tilde{\sigma}$-automaton* is a system

$$\tilde{\mathbf{A}}_{pr} = \langle X, A, Y, \tilde{p}^{(0)}, \tilde{P} \rangle$$

where $\tilde{p}^{(0)} = (\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_m)$ is a correctly specified partial $p\tilde{\sigma}$-vector (a partial probabilistic distribution on the state set) and $\tilde{P} = (\tilde{\sigma}_{si,lj})_{nm,km}$ is a correctly specified partial transition-output $p\tilde{\sigma}$-matrix.

9. Partial *i*-vectors, *i*-matrices, *i*-automata

Let us propose the following notations, where $\alpha, \beta \in \{0, 1\}$:

$$\underset{\alpha}{|} = \begin{cases} (& \text{if } \alpha = 0, \\ [& \text{if } \alpha = 1 \end{cases} \qquad \overset{\beta}{|} = \begin{cases}) & \text{if } \beta = 0, \\ ] & \text{if } \beta = 1. \end{cases}$$

A partial $\tilde{\sigma}$-vector (11) is called a *partial vector with interval elements* (a *partial i-vector*) if in (11)

$$\tilde{\sigma}_i = \underset{\alpha_i}{|} a_i, b_i \overset{\beta_i}{|} \quad (i = \overline{1, m})$$

where $\alpha_i, \beta_i \in \{0, 1\}$, $a_i, b_i \in \mathscr{R}$, $a_i < b_i$ if $\alpha_i \beta_i = 0$, $a_i \leqq b_i$ if $\alpha_i \beta_i = 1$. Thus a partial *i*-vector is a partial $\tilde{\sigma}$-vector such that each of its elements is an interval (closed or unclosed).

Accordingly, a *partial i-matrix* is a partial $\tilde{\sigma}$-matrix (12) such that

$$\tilde{\sigma}_{ij} = \underset{\alpha_{ij}}{|} a_{ij}, b_{ij} \overset{\beta_{ij}}{|} \quad (i = \overline{1, m}, j = \overline{1, n})$$

where $\alpha_{ij}, \beta_{ij} \in \{0, 1\}$, $a_{ij}, \beta_{ij} \in \mathscr{R}$, $a_{ij} < b_{ij}$ if $\alpha_{ij}\beta_{ij} = 0$, $a_{ij} \leqq b_{ij}$ if $\alpha_{ij}\beta_{ij} = 1$. For example,

$$\tilde{R} = \begin{pmatrix} \left[-\frac{1}{2}, 1\right] & \left(\frac{2}{3}, 1\right] \\ \left(-\frac{1}{2}, \frac{1}{2}\right) & \left[\frac{1}{8}, \infty\right) \end{pmatrix}.$$

A *partial generalized i-automaton* is a system

$$\tilde{\mathbf{A}}_{\text{gen}} = \langle X, A, Y, \tilde{r}^{(0)}, \tilde{R} \rangle, \tag{16}$$

$$\tilde{r}^{(0)} = \left(\underset{\alpha_1}{|} a_1, b_1 \overset{\beta_1}{|}, \underset{\alpha_2}{|} a_2, b_2 \overset{\beta_2}{|}, \ldots, \underset{\alpha_m}{|} a_m, b_m \overset{\beta_m}{|} \right),$$

$$\tilde{R} = \left(\underset{\alpha_{si,lj}}{|} a_{si,lj}, b_{si,lj} \overset{\beta_{si,lj}}{|} \right)_{nm,km}.$$

A partial generalized *i*-automaton (16) defines a set of completely specified generalized automata such that

$$\mathbf{A}_{\text{gen}} \in \tilde{\mathbf{A}}_{\text{gen}} \Leftrightarrow r_i \in \underset{\alpha_i}{|} a_i, b_i \overset{\beta_i}{|} \;\&\; r_{si,lj} \in \underset{\alpha_{si,lj}}{|} a_{si,lj}, b_{si,lj} \overset{\beta_{si,lj}}{|} \text{ for all } s, i, l, j$$

where $\mathbf{A}_{\text{gen}}$ is defined by (13).

10. Partial *pi*-vectors, *pi*-matrices, *pi*-automata

A *partial pi-vector* (a partial probabilistic vector with interval elements) is a partial $p\tilde{\sigma}$-vector (14) such that

$$\tilde{\sigma}_i = \left| \underset{\alpha_i}{a_i}, b_i \right|^{\beta_i} \subseteq [0, 1] \quad (i = \overline{1, m}).$$

A *partial pi-matrix* is a partial $p\tilde{\sigma}$-matrix (15) such that

$$\tilde{\sigma}_{ij} = \left| \underset{\alpha_{ij}}{a_{ij}}, b_{ij} \right|^{\beta_{ij}} \subseteq [0, 1] \quad (i = \overline{1, m},\ j = \overline{1, n}).$$

For example,

$$\tilde{P} = \begin{pmatrix} [0;\ 0{,}3) & [0{,}2;\ 0{,}4] & (0{,}3;\ 0{,}8] \\ (0{,}1;\ 0{,}2] & (0{,}3;\ 0{,}5] & [0{,}3;\ 0{,}6) \\ [0{,}2;\ 0{,}3] & [0{,}5;\ 0{,}6] & 0{,}2 \end{pmatrix}$$

is a correctly specified partial square *pi*-matrix of order 3.

A *partial pi-automaton* is a system

$$\tilde{\mathbf{A}}_{pr} = \langle X, A, Y, \tilde{p}^{(0)}, \tilde{P} \rangle$$

where $\tilde{p}^{(0)}$ is a correctly specified partial m-dimensional *pi*-vector and $\tilde{P}$ is a correctly specified partial *pi*-matrix of size $nm \times km$. In the case of closed intervals the problem of partial *pi*-automata minimization was studied in [1].

11. The conditions of correct specification

Now we are going to find the conditions which must be satisfied for correct specification of a partial *pi*-vector (*pi*-matrix, *pi*-automaton). Such conditions in case of $\alpha_i = \beta_i = 1$ $(i = \overline{1, m})$ were found in [1].

Theorem. *Let p be a partial pi-vector defined as*

$$\tilde{p} = \left(\left| \underset{\alpha_1}{a_1}, b_1 \right|^{\beta_1}, \left| \underset{\alpha_2}{a_2}, b_2 \right|^{\beta_2}, \ldots, \left| \underset{\alpha_m}{a_m}, b_m \right|^{\beta_m} \right) \tag{17}$$

Vhere

$$\left| \underset{\alpha_i}{a_i}, b_i \right|^{\beta_i} \neq \emptyset, \quad \left| \underset{\alpha_i}{a_i}, b_i \right|^{\beta_i} \subseteq [0, 1], \quad i = \overline{1, m}.$$

ahen p is correctly specified if and only if the following conditions hold for $j = \overline{1, m}$*:*

(a)
$$a_j \geqq 1 - \sum_{i \neq j} b_i \tag{18}$$

and

$$a_j = 1 - \sum_{i \neq j} b_i \ \& \ \exists\, i : i \neq j, \quad \beta_i = 0 \Rightarrow \alpha_j = 0, \tag{19}$$

(b)
$$b_j \leqq 1 - \sum_{i \neq j} a_i \tag{20}$$

and

$$b_j = 1 - \sum_{i \neq j} a_i \;\&\; \exists i : i \neq j, \quad \alpha_i = 0 \Rightarrow \beta_j = 0. \tag{21}$$

Proof. For the proof of the necessity let $\tilde{p}$ be a correctly specified partial *pi*-vector. Since $\tilde{p} \neq \emptyset$ thus $\sum_i a_i \leqq 1$, $\sum_i b_i \geqq 1$, and for every j,

$$b_j \geqq 1 - \sum_{i \neq j} b_i, \quad a_j \leqq 1 - \sum_{i \neq j} a_i. \tag{22}$$

Assume now that the condition (18) does not hold for any j and $b_j - a_j > 0$,

$$a_j < 1 - \sum_{i \neq j} b_i. \tag{23}$$

Then we take

$$p_j = \frac{a_j + 1 - \sum_{i \neq j} b_i}{2}. \tag{24}$$

Since (22) and (23) hold thus $a_j < p_j < b_j$ and $p_j \in |\,a_j, b_j\,|_{\alpha_j}^{\beta_j}$. Since $\tilde{p}$ is correctly specified thus there must be a probabilistic vector $p = (p_1, p_2, \ldots, p_m) \in \tilde{p}$ such that p_j has a value (24). Then for p,

$$\sum_i p_i = 1 = \frac{a_j + 1 - \sum_{i \neq j} b_i}{2} + \sum_{i \neq j} p_i \leqq \frac{a_j + 1 - \sum_{i \neq j} b_i}{2} + \sum_{i \neq j} b_i$$

holds. This implies that

$$a_j \geqq 1 - \sum_{i \neq j} b_i$$

which contradicts our assumption (23). Therefore in the case $b_j - a_j > 0$ the condition (18) holds.

In exeptional case when $\tilde{\sigma}_j = [a_j, a_j] = a_j$, every probabilistic vector $p \in \tilde{p}$ has $p_j = a_j$ and, therefore,

$$\sum_i p_i = 1 = a_j + \sum_{i \neq j} p_i \leqq a_j + \sum_{i \neq j} b_i,$$

i.e., the condition (18) also holds.

Assume now that $a_j = 1 - \sum_{i \neq j} b_i$ (i.e., $a_j + \sum_{i \neq j} b_i = 1$) and there is an $s \neq j$ such that $\beta_s = 0$ but $\alpha_j = 1$. Since $\tilde{p}$ is correctly specified thus in this case there is a probabilistic vector $p \in \tilde{p}$ such that $p_j = a_j$, $p_s < b_s$. Then for the vector p,

$$\sum_i p_i = 1 = a_j + \sum_{i \neq j} p_i < a_j + \sum_{i \neq j} b_i$$

holds. But this contradicts our assumption. Therefore $\alpha_j = 0$ and the condition (19) holds. This ends the proof of the necessity of the conditions (a).

The necessity of condition (b) can be shown similarly.

Conversely, assume that conditions (a) and (b) hold for $\tilde{p}$. We prove that $\tilde{p}$ is correctly specified. Let us take any j and any $p_j \in |\,a_j, b_j\,|_{\alpha_j}^{\beta_j}$. It follows from (18) and (20) that

$$1 - p_j \geqq 1 - b_j \geqq \sum_{i \neq j} a_i, \quad 1 - p_j \leqq 1 - a_j \leqq \sum_{i \neq j} b_i. \tag{25}$$

We take for $i\neq j$ the following elements of a vector $p=(p_1, p_2, \ldots, p_m)$

$$p_i = a_i + \frac{1-p_j-\sum\limits_{i\neq j} a_i}{\sum\limits_{i\neq j}(b_i-a_i)}(b_i-a_i) \quad (i\neq j), \tag{26}$$

where $\sum\limits_{i\neq j}(b_i-a_i)>0$ (if $\sum\limits_{i\neq j}(b_i-a_i)=0$ then $a_i=b_i$ $(i=\overline{1,m})$ and $\tilde{p}$ is a completely specified probabilistic vector). Then for the vector p,

$$\sum_i p_i = p_j + \sum_{i\neq j}\left(a_i + \frac{1-p_j-\sum\limits_{i\neq j} a_i}{\sum\limits_{i\neq j}(b_i-a_i)}(b_i-a_i)\right) = 1,$$

i.e., p is a probabilistic vector. Now we shall prove that $p\in\tilde{p}$.

From (25) and (26) we have that $p_i\cong a_i$ $(i\neq j)$ and for any $i\neq j$,

$$p_i = a_i \Leftrightarrow b_i = a_i \vee 1-p_j = \sum_{i\neq j} a_i.$$

If $b_i=a_i$ then $\tilde{\sigma}_i=[a_i, a_i]=a_i$. If $1-p_j=\sum\limits_{i\neq j} a_i$ then in accordance with (25), $p_j=b_j$ $\beta_j=1$, $1-b_j=\sum\limits_{i\neq j} a_i$ and it follows from (21) that $\alpha_i=1$ $(i\neq j)$. Thus if $p_i=a_i$ for any $i\neq j$ then $\alpha_i=1$ and $p_i\in\tilde{\sigma}_i=[a_i, b_i|^{\beta_i}$. Moreover, it follows from (25) that

$$1-p_j-\sum_{i\neq j} a_i \leqq \sum_{i\neq j}(b_i-a_i).$$

Therefore, $p_i\leqq b_i$ $(i\neq j)$, and for any $i\neq j$,

$$p_i = b_i \Leftrightarrow b_i = a_i \quad \vee \quad 1-p_j-\sum_{i\neq j} a_i = \sum_{i\neq j}(b_i-a_i).$$

If $b_i=a_i$ then $\tilde{\sigma}_i=[a_i, a_i]=a_i$. If $1-p_j-\sum\limits_{i\neq j} a_i=\sum\limits_{i\neq j}(b_i-a_i)$ then $1-p_j=\sum\limits_{i\neq j} b_i$ and, in accordance with (25), $p_j=a_j$, $\alpha_j=1$. In this case the condition (19) implies $\beta_i=1$ $(i\neq j)$. Thus if $p_i=b_i$ for any $i\neq j$ then $\beta_i=1$ and $p_i\in\tilde{\sigma}_i=|a_i, b_i]$. Finally, if for any $i\neq j$, $a_i<p_i<b_i$ then $p_i\in|a_i, b_i|^{\beta_i}_{\alpha_i}$. Thus, we proved that the constracted vector p is probabilistic and $p\in\tilde{p}$. Therefore, $\tilde{p}$ is correctly specified. This completes the proof of the Theorem.

12. Partial *b*-vectors, *b*-matrices, *b*-automata

A partial *f*-vector

$$\tilde{r} = (f_1(\{\xi_v\}), f_2(\{\xi_v\}), \ldots, f_m(\{\xi_v\})) \quad (\xi_v\in\{0, 1\}, \quad v=\overline{1,q}), \tag{27}$$

where $f_i(\{\xi_v\})$ $(i=\overline{1,m})$ are boolean (logical) functions, is called a *partial boolean vector* (a *partial b-vector*). A partial *f*-matrix

$$\tilde{R} = (f_{ij}(\{\xi_v\}))_{m,n} \quad (\xi_v\in\{0, 1\}, \ v=\overline{1,q}) \tag{28}$$

where $f_{ij}(\{\xi_v\})$ $(i=\overline{1,m},\ j=\overline{1,n})$ are boolean functions, is a *partial b-matrix*. For b-vectors and b-matrices the domain of every parameter is $\{0, 1\}$, therefore it may be omitted. For example,

$$\tilde{R} = \begin{pmatrix} \xi_1 \vee \xi_2 & \xi_1 \bar{\xi}_2 & \bar{\xi}_1 \xi_2 \\ \xi_2 \xi_3 & \bar{\xi}_2 \vee \xi_3 & \xi_1 \bar{\xi}_2 \vee \xi_3 \end{pmatrix}.$$

If a partial b-vector (b-matrix) is a partial $\tilde{\sigma}$-vector ($\tilde{\sigma}$-matrix) then its elements may be 0, 1 or $\{0, 1\}$. In this case it is convenient to replace $\{0, 1\}$ by "—". For example,

$$\tilde{R} = \begin{pmatrix} 0 & - & 1 \\ - & 0 & - \\ 1 & 1 & - \end{pmatrix}.$$

A *partial generalized b-automaton* is a system

$$\tilde{\mathbf{A}}_{\text{gen}} = \langle X, A, Y, \tilde{r}^{(0)}, \tilde{R} \rangle,$$

$$\tilde{r}^{(0)} = (f_1, f_2, \ldots, f_m), \quad \tilde{R} = (f_{ij})_{m,n}$$

where $f_i = f_i(\{\xi_v\}), f_{ij} = f_{ij}(\{\xi_v\})$ are boolean functions of the parameters $\xi_1, \xi_2, \ldots, \xi_q$ $(\xi_v \in \{0, 1\},\ v=\overline{1,q})$.

13. Partial d-vectors, d-matrices, d-automata

If a partial b-vector (27) is also a partial p-vector then $\tilde{p} \leqq \mathscr{D}^m$ and

$$f_i f_j \equiv 0 \quad (i \neq j), \quad \bigvee_i f_i \equiv 1. \tag{29}$$

Such a partial vector is called a *partial d-vector*. Thus if a partial b-matrix (28) is also a partial p-matrix then it is of form

$$\tilde{D} = (f_{ij})_{m,n}, \quad f_{ij} f_{il} \equiv 0 \quad (j \neq l), \quad \bigvee_j f_{ij} \equiv 1 \quad (i = \overline{1,m}) \tag{30}$$

and $\tilde{D} \leqq \mathscr{D}^{m,n}$. Such a partial matrix is called a *partial d-matrix*. It is useful to notice that any subset of $\mathscr{D}^m (\mathscr{D}^{m,n})$ may be specified as a partial d-vector (d-matrix). For example,

$$\tilde{D} = \begin{pmatrix} \xi_1 \vee \xi_2 & \bar{\xi}_1 \bar{\xi}_2 & 0 \\ \bar{\xi}_1 & \xi_1 \bar{\xi}_2 & \xi_1 \xi_2 \\ \bar{\xi}_2 \xi_3 & \xi_2 \vee \bar{\xi}_3 & 0 \end{pmatrix}$$

is a partial square d-matrix of order 3.

If a partial d-matrix is a partial $p\tilde{\sigma}$-matrix then $\{0, 1\}$ may also be replaced by "—", but it is necessary to keep in mind the conditions (30).

A *partial deterministic automaton* (a *partial d-automaton*) is a system

$$\tilde{\mathbf{A}}_{\text{det}} = \langle X, A, Y, \tilde{d}^{(0)}, \tilde{D} \rangle$$

where $\tilde{d}^{(0)} = (f_1, f_2, \ldots, f_m)$ is a partial d-vector and $\tilde{D} = (f_{si,lj})_{nm,km}$ is a partial d-matrix.

14. Automata programming

Above the most general definitions of incompletely specified finite automata were proposed and some special classes of such automata were introduced. For these automata all classical problems of the automata theory may be formulated. Some of such problems were investigated, for example in [1]—[4] for certain partial p-automata, partial pi-automata and partial d-automata. But a partial automaton is a more interesting object for investigation than a completely specified automaton and there are many special important problems in its theory. One class of such problems which we shall call "the problems of automata programming" may be formulated in the following way.

Let $\tilde{\mathbf{A}}_{\text{gen}}^{(1)}, \tilde{\mathbf{A}}_{\text{gen}}^{(2)}, \ldots, \tilde{\mathbf{A}}_{\text{gen}}^{(q)}$ be partial automata (for example generalized) and Ψ be a mapping

$$\Psi : \tilde{\mathbf{A}}_{\text{gen}}^{(1)} \times \tilde{\mathbf{A}}_{\text{gen}}^{(2)} \times \ldots \times \tilde{\mathbf{A}}_{\text{gen}}^{(q)} \to \mathscr{R}.$$

It is necessary to find partial automata $\tilde{\mathbf{A}}_{\text{gen}}^{(1)\prime}, \tilde{\mathbf{A}}_{\text{gen}}^{(2)\prime}, \ldots, \tilde{\mathbf{A}}_{\text{gen}}^{(q)\prime}$ such that

$$\tilde{\mathbf{A}}_{\text{gen}}^{(i)\prime} \subseteq \tilde{\mathbf{A}}_{\text{gen}}^{(i)} \quad (i = \overline{1, q})$$

and

$$\mathbf{A}_{\text{gen}}^{i} \in \tilde{\mathbf{A}}_{\text{gen}}^{(i)\prime} \quad (i = \overline{1,q}) \Leftrightarrow \Psi(\mathbf{A}_{\text{gen}}^{(1)}, \mathbf{A}_{\text{gen}}^{(2)}, \ldots, \mathbf{A}_{\text{gen}}^{(q)}) = \Psi_{\max}$$

where

$$\Psi_{\max} = \max_{\substack{\mathbf{A}_{\text{gen}}^{(i)} \in \tilde{\mathbf{A}}_{\text{gen}}^{(i)} \\ i=\overline{1, q}}} \Psi(\mathbf{A}_{\text{gen}}^{(1)}, \mathbf{A}_{\text{gen}}^{(2)}, \ldots, \mathbf{A}_{\text{gen}}^{(q)}).$$

Such problems, for example, are very important for optimization of automata or some systems and processes which may be described in terms of automata. One such problem concerned with automata reliability was solved in [1].

Finally the author wishes to express his deep gratitude to the University of Budapest which provided him with ideal working conditions during two months. The author is especially indepted to Prof. I. Kátai, Dr. I. Peák and Dr. L. Hunyadvári who helped to prepare this work for publication.

ŹDANOV STATE UNIVERSITY
LENINGRAD

References

[1] Чирков, М. К., Основы общей теории конечных автоматов, Издательство Ленинградского университета, Ленинград, 1975.
[2] Bernard W. Lovell, The incompletely-specified finite-state stochastic sequential machine equivalence and reduction, Annual Symp. on Switching and Automata Theory, 10th IEEE Conf. Record, Waterloo, 1969, New York, N. Y., 1969, pp. 82—89.
[3] Muller, D. E. and R. E. Muller, A generalization of the theory of incompletely specified machines, *J. Comput. System Sci.*, v. 6, 1972, pp. 419—447.
[4] Tomescu, I., A method for minimizing the number of states for a restricted class of incompletely specified sequential machines, *An. Univ. Bucureşti Mat.—Mec.*, v. 21, 1972, pp. 97—108.
[5] Paz, A., *Introduction to probabilistic automata,* Academic Press, New York—London, 1971.

(Received Dec. 27, 1977)

The solvability of the equivalence problem for deterministic frontier-to-root tree transducers

By Z. ZACHAR

1. Introduction

In this paper we deal with effective solvability of the equivalence of frontier-to-root tree transducers. T. V. Griffits has shown in [2] that the equivalence problem is unsolvable for λ-free nondeterministic generalized sequential machines which are special frontier-to-root tree transducers, so the equivalence of the nondeterministic frontier-to-root transducers is unsolvable, too. Then in a natural way one can raise the question whether the equivalence of deterministic frontier-to-root tree transducers is solvable. We show the answer is in the affirmative. The proof is based on the proof of the solvability of equivalence problem for λ-free deterministic generalized sequential machines given by F. Gécseg (unpublished result). M. Steinby has called the author's attention to the fact that this result can be employed for minimalization of deterministic frontier-to-root tree transducers. In section 4 we give an algorithm for the minimalization.

A systematic summary of further results concerning frontier-to-root and root-to-frontier tree transducers can be found in [1], where they are called bottom-up and top-down tree transducers, respectively.

2. Notions and notations

Let $X=\{x_1, \ldots, x_n, \ldots\}$, $Y=\{y_1, \ldots, y_m, \ldots\}$ and $Z=\{z_1, \ldots, z_k, \ldots\}$ be countable sets of variables kept fix in this paper. Denote by X_n the subset $\{x_1, \ldots, x_n\}$ of X. Consider a nonvoid set F and a mapping v of F into the set of all nonnegative integers. The pair (F, v) is called a type. Then the set $T_F(X)$ of *polynomial symbols* over X of type F is defined in the following way:

(a) for each x $(x\in X)$, $x\in T_F(X)$,

(b) if $f\in F$, $v(f)=k(\geqq 0)$, and $p_1, \ldots, p_k\in T_F(X)$ then $f(p_1, \ldots, p_k)\in T_F(X)$,

(c) the polynomial symbols over X of type F are those and only those which we get from (a) and (b) in finite number of steps.

Now we define the *depth* $d(p)$ of $p\in T_F(X)$ as follows:

(a) if $p=x$ $(x\in X)$ then $d(p)=0$,

(b) if $p=f$ $(f\in F)$ and $v(f)=0$ then $d(p)=0$,

(c) if $p=f(p_1, \ldots, p_k)$ $(v(f)=k>0)$ then $d(p)=\max(d(p_i)|i=1, \ldots, k)+1$.

In the literature elements of $T_F(X)$ are called trees, or, in a more detailed form, F-trees.

Next we define the *frontier* $\mathrm{fr}(p)$ of a tree $p\in T_F(X)$ in the following way:

(a) if $p=x$ $(x\in X)$ then $\mathrm{fr}(p)=x$,

(b) if $p=f(p_1, \ldots, p_k)$ $(v(f)=k)$ then $\mathrm{fr}(p)=\mathrm{fr}(p_1)\ldots\mathrm{fr}(p_k)$.

We notice that if $p=f$ and $v(f)=0$, then $\mathrm{fr}(p)=\lambda$, where λ denotes the empty word over X.

We can define the set $\mathrm{sub}(p)$ of *subtrees* of $p\in T_F(X)$ as follows:

(a) if $p=x$ $(x\in X)$ then $\mathrm{sub}(p)=\{x\}$,

(b) if $p=f(p_1, \ldots, p_k)$ $(v(f)=k)$ then
$\mathrm{sub}(p)=\bigcup(\mathrm{sub}(p_i)|i=1, \ldots, k)\cup\{p\}$.

Let $\overline{\mathrm{sub}}(p)=\mathrm{sub}(p)\setminus\{p\}$ be the set of proper subtrees of a tree $p\in T_F(X)$.

Next we define the concept of a substitution. Let $p\in T_F(X_n)$ be an arbitrary tree and $T_1, \ldots, T_n\subseteq T_F(X_n)$. Then $p[T_1\to x_1, \ldots, T_n\to x_n]$ is the set of trees obtained by replacing every occurrence of $x_1, \ldots, x_n$ by a tree in $T_1, \ldots, T_n$, respectively. Formally,

(a) if $p=x_i$ $(x_i\in X_n)$ then $p[T_1\to x_1, \ldots, T_n\to x_n]=T_i$,

(b) if $p=f(p_1, \ldots, p_k)$ $(v(f)=k)$ then $p[T_1\to x_1, \ldots, T_n\to x_n]=$
$=\{f(\bar{p}_1, \ldots, \bar{p}_k)|\bar{p}_i\in p_i[T_1\to x_1, \ldots, T_n\to x_n],\ i=1, \ldots, k\}$.

Let $T_1, T_2\subseteq T_F(X_n)$ be arbitrary subsets and $x_i\in X_n$. Then the *x_i-product* $T_1\cdot x_i T_2$ of T_1 by T_2 is the set of trees which can be obtained by replacing every occurrence of x_i in some tree from T_2 by a tree in T_1.

Let $T_1^{0,x_i}=\{x_i\}$ and for every $k>0$

$$T_1^{k,x_i}=T_1^{k-1,x_i}\cdot x_i T_1.$$

Obviously,

$$T_1\cdot x_i T_2=\{p[\{x_1\}\to x_1, \ldots, \{x_{i-1}\}\to x_{i-1},\ T_1\to x_i,\ \{x_{i+1}\}\to x_{i+1}, \ldots, \{x_n\}\to x_n]|p\in T_2\}.$$

Let us note that a singleton will also be denoted by its element.

Let (F, v) and (G, μ) be fixed finite types. Moreover, let A be a finite set of states.

A *frontier-to-root rewriting* (FR) *rule* is determined by a triple of the following two forms:

(a) (x, a, q), where $x\in X$, $a\in A$ and $q\in T_G(Y)$,

(b) $(f((a_1, z_1), \ldots, (a_k, z_k)), a, q)$, where $f\in F$, $v(f)=k$,
$(a_i, z_i)\in A\times\{z_i\}$ $(i=1, \ldots, k)$, $a\in A$ and $q\in T_G(Y\,|\,|\,Z_k)$.

In the sequel we write the FR rules in the form $x\to aq$ and $f(a_1z_1, \ldots, a_kz_k)\to aq$, respectively.

A *root-to-frontier rewriting* (RF) *rule* is given by a triple of the following forms:

(a) (a, x, q) where $a\in A$, $x\in X$ and $q\in T_G(Y)$,

(b) $(a, f(z_1, \ldots, z_k), q)$ where $a\in A$, $f\in F$, $v(f)=k$ and $q\in T_G(Y\cup A\times Z_k)$.

Further on we write the RF rules in the form $ax\to q$ and $af(z_1, \ldots, z_k)\to q$, respectively.

By a *frontier-to-root tree* (FRT) *transducer* we mean a system $\mathfrak{A}=(F, A, G, A', \Sigma)$, where A' is a subset of A called the set of final states and Σ is a finite set of FR rules. Since Σ is finite thus there is a number n such that the set of symbols x, for which

there exists a rule in Σ with left hand side x, is a subset of X_n. Similarly, there exists a number m such that right hand sides of rules from Σ get into $A \times T_G(Y_m \cup Z)$. Then we can restrict ourselves to X_n and Y_m.

For each $a \in A$ and $p \in T_F(X_n)$, the set of all *a-translations of* p, denoted by $\mathfrak{A}_a(p)$, is defined as follows:

(a) if $p = x_i$ $(1 \leqq i \leqq n)$, then $\mathfrak{A}_a(p) = \{q | x_i \to aq \in \Sigma\}$,

(b) if $p = f(p_1, \ldots, p_k)$ $(\nu(f) = k)$ then

$\mathfrak{A}_a(p) = \{q | f(a_1 z_1, \ldots, a_k z_k) \to a\bar{q} \in \Sigma,\ q \in \bar{q}[\mathfrak{A}_{a_1}(p_1) \to z_1, \ldots, \mathfrak{A}_{a_k}(p_k) \to z_k]\}$.

An FRT transducer $\mathfrak{A}$ is *deterministic* (DFRT transducer) if

(a) for all $x_i \in X_n$, there is at most one rule with left hand side x_i,

(b) for all $f \in F$ and $a_1, \ldots, a_k \in A$, there is at most one rule with left hand side $f(a_1 z_1, \ldots, a_k z_k)$.

By a *root-to-frontier tree* (RFT) *transducer* we mean a system $\mathfrak{A} = (F, A, G, A', \Sigma)$, where A' $(\subseteq A)$ is the set of initial states and Σ is a finite set of RF rules. Similarly, in this case we can be restricted to X_n and Y_m for some n and m.

For each $a \in A$ and $p \in T_F(X_n)$, the set of all *a-translations of* p, denoted by $\mathfrak{A}_a(p)$, is defined as follows:

(a) if $p = x_i$ $(1 \leqq i \leqq n)$ then $\mathfrak{A}_a(p) = \{q | ax_i \to q \in \Sigma\}$,

(b) if $p = f(p_1, \ldots, p_k)$ $(\nu(f) = k)$ then

$\mathfrak{A}_a(p) = \{q | af(z_1, \ldots, z_k) \to \bar{q}(\ldots, \bar{a} z_i, \ldots) \in \Sigma,\ q \in \bar{q}[\ldots, \mathfrak{A}_{\bar{a}}(p_i) \to \bar{a} z_i, \ldots]\}$.

An RFT transducer $\mathfrak{A}$ is *deterministic* (DRFT transducer) if

(a) for all $x_i \in X_n$ and $a \in A$, there is at most one rule with left hand side ax_i,

(b) for all $f \in F$ $(\nu(f) = k)$ and $a \in A$, there is at most one rule with left hand side $af(z_1, \ldots, z_k)$,

(c) A' is a singleton.

Let $\mathfrak{A} = (F, A, G, A', \Sigma)$ be a FRT (RFT) transducer and $p \in T_F(X_n)$. The *translations of* p *induced by* $\mathfrak{A}$, denoted by $\mathfrak{A}(p)$, is the set $\bigcup(\mathfrak{A}_a(p) | a \in A')$.

We define the *transformation induced by* $\mathfrak{A}$ to be the relation $\{(p, q) | p \in T_F(X_n), q \in \mathfrak{A}(p)\}$ from $T_F(X_n)$ into $T_G(Y_m)$.

If $\mathfrak{A}$ is a deterministic FRT (RFT) transducer, then for each $p \in T_F(X_n)$ at most one element is in $\mathfrak{A}(p)$. Therefore, the transformation induced by $\mathfrak{A}$ is a (partial) mapping from $T_F(X_n)$ into $T_G(Y_m)$, and it is denoted by $\mathfrak{A}$, too. This mapping is called the *mapping induced by* $\mathfrak{A}$.

Let $\mathfrak{A} = (F, A, G, A' \Sigma_A)$ and $\mathfrak{B} = (F, B, G, B', \Sigma_B)$ be FRT (RFT) transducers. We say that $\mathfrak{A}$ and $\mathfrak{B}$ are *equivalent* if and only if $\mathfrak{A}$ and $\mathfrak{B}$ induce the same transformation. The FRT (RFT) transducer $\mathfrak{A}$ is *minimal* if and only if for all FRT (RFT) transducer $\mathfrak{C} = (F, C, G, C', \Sigma_C)$ equivalent to $\mathfrak{A}$, $|A| \leqq |C|$ holds.

We say that $\mathfrak{A}$ is a *minimal transducer belonging to* $\mathfrak{B}$ if and only if $\mathfrak{A}$ and $\mathfrak{B}$ are equivalent and $\mathfrak{A}$ is minimal.

3. The equivalence of deterministic frontier-to-root tree transducers

Let $\mathfrak{A} = (F, A, G, A', \Sigma_A)$ and $\mathfrak{B} = (F, B, G, B', \Sigma_B)$ be deterministic frontier-to-root tree transducers such that the mappings induced by $\mathfrak{A}$ and $\mathfrak{B}$ are from $T_F(X_n)$ into $T_G(Y_m)$. Let us construct, for the states $a \in A$ and $b \in B$, two DFRT transducers

$$\mathfrak{A}^a = (F, A, B, A', \Sigma_A \cup \{\# \to a\#\})$$

and

$$\mathfrak{B}^b = (F, B, G, B', \Sigma_B \cup \{\# \to b\#\}).$$

Then $\mathfrak{A}^a$ and $\mathfrak{B}^b$ induce mappings from $T_F(X_n \cup \{\#\})$ into $T_G(Y_m \cup \{\#\})$.

We define the $\#$-*depth* $\bar{d}(p)$ of a tree $p \in T_F(X_n)$ in the following way:

(a) if $p = x_i$ $(1 \leqq i \leqq n)$ then $\bar{d}(p)$ is undefined,

(b) if $p = \#$ then $\bar{d}(p) = 0$,

(c) if $p = f(p_1, \ldots, p_k)$ $(\nu(f) = k)$ and $\bar{d}(p_i)$ $(i = 1, \ldots, k)$ are undefined then $\bar{d}(p)$ is undefined,

(d) if $p = f(p_1, \ldots, p_k)$ $(\nu(f) = k)$ and one of $\bar{d}(p_i)$ $(1 \leqq i \leqq k)$ is defined, then $\bar{d}(p) = \max(\bar{d}(p_i) | \bar{d}(p_i)$ is defined, $1 \leqq i \leqq k) + 1$.

Let T be the set of all trees $p \in T_F(X_n)$ for which both $\mathfrak{A}(p)$ and $\mathfrak{B}(p)$ are defined.

Take a tree $p \in T$ and an arbitrary subtree $\bar{p} \in \text{sub}(p)$. Let $\bar{\bar{p}} \in T_F(X_n \cup \{\#\})$ be the tree obtained by replacing a fix occurrence of $\bar{p}$ by $\#$. Obviously, $\bar{\bar{p}}$ contains exactly one symbol $\#$ on its frontier and $p = \bar{\bar{p}} \cdot \bar{p}$, where $\bar{\bar{p}} \cdot \bar{p}$ denotes the $\#$-product of $\bar{\bar{p}}$ by $\bar{p}$. Since $p \in T$, there exist exactly one state of A and B denoted respectively by $A_{\bar{p}}$ and $B_{\bar{p}}$, such that both $\mathfrak{A}_{A_{\bar{p}}}(\bar{\bar{p}})$ and $\mathfrak{B}_{B_{\bar{p}}}(\bar{\bar{p}})$ are defined.

The following two lemmas hold under these notations.

Lemma 1. For each $p \in T$ and $\bar{p} \in \text{sub}(p)$,

$$\mathfrak{A}(p) = \mathfrak{A}_{A_{\bar{p}}}(\bar{\bar{p}}) \cdot \mathfrak{A}^{A_{\bar{p}}}(\bar{p})$$

and

$$\mathfrak{B}(p) = \mathfrak{B}_{B_{\bar{p}}}(\bar{\bar{p}}) \cdot \mathfrak{B}^{B_{\bar{p}}}(\bar{p})$$

hold.

Proof is obvious.

Next let $|A| = M$ and $|B| = N$.

Lemma 2. Let $p \in T$ be an arbitrary tree and $\bar{p} \in \text{sub}(p)$. Then there exists a tree $t \in T_F(X_n \cup \{\#\})$ containing exactly one symbol $\#$ on its frontier such that $\bar{d}(t) < MN$, $d(t) < 2MN - 1$ and $\bar{p} \cdot t \in T$.

Proof. First we give a tree $\bar{t}$, for which $\bar{d}(\bar{t}) < MN$. Construct a sequence $t_1, \ldots, t_s, \ldots$ of trees as follows: Set $t_0 = \bar{p}$. Then consider the sequence $q_0, \ldots, q_l$ of maximal length, for which $q_0 = t_s$, $q_l = \#$ and $q_i \in \overline{\text{sub}}(q_{i-1})$ $(i = 1, \ldots, l)$. If $l < MN$ then $\bar{d}(t_s) < MN$, and in this case let $\bar{t} = t_s$. Otherwise, we can find two indices j and k such that $0 \leqq j < k \leqq l$ and $A_{q_j} = A_{q_k}$, $B_{q_j} = B_{q_k}$. Then let t_{s+1} be the tree obtained from t_s by replacing the subtree q_j in t_s by q_k. It is clear that $\bar{d}(t_{s+1}) < \bar{d}(t_s)$. Thus, continuing this process in a finite number of steps we arrive at the desired tree $\bar{t}$. If $d(\bar{t}) < 2MN - 1$ then let $t = \bar{t}$. In the opposite case there exists a sequence $q_0, \ldots, q_l$ of subtrees of $\bar{t}$ with $l \geqq MN$, $\# \notin \text{sub}(q_0)$, $q_l \in X_n$ and $q_i \in \overline{\text{sub}}(q_{i-1})$ $(i = 1, \ldots, l)$. We construct a tree $\bar{\bar{t}}$ from $\bar{t}$ by means of the sequence $q_0, \ldots, q_l$ in the same way as $\bar{t}$ has been constructed from $\bar{p}$. The tree $\bar{\bar{t}}$ contains less occurrences of symbols from F than $\bar{t}$ does. It follows that the procedure can be continued till the depth of the resulting tree is not less than $2MN - 1$. The constructed tree satisfies the conclusions of Lemma 2.

Notice that if the frontier of $\mathfrak{A}^{A\bar{p}}(\bar{p})$ contains the symbol #, then it occurs in the frontier of $\mathfrak{A}^{A\bar{p}}(t)$. Similar statement is valid for $\mathfrak{B}^{B\bar{p}}(\bar{p})$ and $\mathfrak{B}^{B\bar{p}}(t)$.

Lemma 3. Let $p\in T$ and $d(p)\geqq 4MN$. Then there exist trees $p_1, p_2, p_3, p_4, p_5, p_6\in$ $\in T_F(X_n\cup\{\#\})$ such that p_2, p_3, p_4, p_5, p_6 contain exactly one symbol # in their frontiers. Moreover, $p=p_1\cdot p_2\cdot p_3\cdot p_4\cdot p_5\cdot p_6$, $d(p_i)\geqq 1$ $(i=2, 3, 4, 5)$ and $d(p_1\cdot p_2\cdot p_3\cdot p_4\cdot p_5)\leqq 4MN$. Finally, the following equations hold:

$$A_{p_1} = A_{(p_1\cdot p_2)} = A_{(p_1\cdot p_2\cdot p_3)} = A_{(p_1\cdot p_2\cdot p_3\cdot p_4)} = A_{(p_1\cdot p_2\cdot p_3\cdot p_4\cdot p_5)} = a,$$

$$B_{p_1} = B_{(p_1\cdot p_2)} = B_{(p_1\cdot p_2\cdot p_3)} = B_{(p_1\cdot p_2\cdot p_3\cdot p_4)} = B_{(p_1\cdot p_2\cdot p_3\cdot p_4\cdot p_5)} = b,$$

$$\mathfrak{A}(p) = \mathfrak{A}_a(p_1)\cdot\mathfrak{A}^a_a(p_2)\cdot\mathfrak{A}^a_a(p_3)\cdot\mathfrak{A}^a_a(p_4)\cdot\mathfrak{A}^a_a(p_5)\cdot\mathfrak{A}^a(p_6),$$

$$\mathfrak{B}(p) = \mathfrak{B}_b(p_1)\cdot\mathfrak{B}^b_b(p_2)\cdot\mathfrak{B}^b_b(p_3)\cdot\mathfrak{B}^b_b(p_4)\cdot\mathfrak{B}^b_b(p_5)\cdot\mathfrak{B}^b(p_6).$$

Proof. Let $\bar{p}$ be an arbitrary subtree of p with depth $4MN$. Then there exists a sequence $q_0, \ldots, q_{4MN}$ of trees with $q_0=\bar{p}$ and $q_i\in\overline{\mathrm{sub}}(q_{i-1})$ $(i=1, \ldots, 4MN)$. Consider the pairs of states (A_{q_i}, B_{q_i}) $(i=0, \ldots, 4MN)$. Obviously, there exist indices j_1, j_2, j_3, j_4, j_5 $(4MN\geqq j_1>j_2>j_3>j_4>j_5\geqq 0)$ having the same pairs of states.

Let $p_1=q_{j_1}$. Construct the tree p_k by replacing the subtree $q_{j_{k-1}}$ in the tree q_{j_k} by the symbol # $(k=2, 3, 4, 5)$. Finally, let p_6 be the tree obtained from p by replacing its subtree q_{j_5} by #. From the construction and Lemma 1, it is clear that the trees $p_1, p_2, p_3, p_4, p_5, p_6$ constructed in this way satisfy the conditions of Lemma 3.

Let $L=\max\left(d(\mathfrak{A}(p)),\ d(\mathfrak{B}(p))\middle|p\in T,\ d(p)\leqq 6MN\right)$ and $K=4(L+2)MN$.

Lemma 4. Take a tree $p\in T$. Moreover, let $p_1, p_2, p_3, p_4, p_5, p_6\in T_F(X_n\cup\{\#\})$ be trees and $a\in A$ and $b\in B$ states satisfying the conditions of Lemma 3. If $\mathfrak{A}(p)\neq\mathfrak{B}(p)$ and $\bar{d}(\mathfrak{A}^a(p_4\cdot p_5\cdot p_6))$ is undefined, then there is a tree $\bar{p}\in T$, for which $d(\bar{p})<K$ and $\mathfrak{A}(\bar{p})\neq\mathfrak{B}(\bar{p})$.

Proof. Let S be the set of trees with minimal depth satisfying the conditions of Lemma 4. Let $p(\in S)$ be a tree which has minimal number of occurrences of symbols from F among all trees in S. Assume that $d(p)\geqq K$.

The #-depth of the tree $\mathfrak{B}^b(p_3\cdot p_4\cdot p_5\cdot p_6)$ is defined and $\bar{d}(\mathfrak{B}^b_b(p_3))>0$, for otherwise

$$\mathfrak{A}(p_1\cdot p_3\cdot p_4\cdot p_5\cdot p_6) = \mathfrak{A}(p)\neq\mathfrak{B}(p) = \mathfrak{B}(p_1\cdot p_3\cdot p_4\cdot p_5\cdot p_6)$$

or

$$\mathfrak{A}(p_1\cdot p_2\cdot p_4\cdot p_5\cdot p_6) = \mathfrak{A}(p)\neq\mathfrak{B}(p) = \mathfrak{B}(p_1\cdot p_2\cdot p_4\cdot p_5\cdot p_6)$$

holds, which contradicts the minimality of p. Next we define a tree t, for which

$$d(t)<3MN-1 \text{ and } \bar{d}(t)<2MN-1.$$

First we consider the sequence $q_0, \ldots, q_l$ of subtrees with maximal length for which $q_0=p_4\cdot p_5\cdot p_6$, $q_l=\#$ and $q_i\in\overline{\mathrm{sub}}(q_{i-1})$ $(i=1, \ldots, l)$. Then for each q_i there is exactly one state $a_i\in A$ such that $\mathfrak{A}^a_{a_i}(q_i)$ is defined. Let i be the maximal index, for which $\bar{d}(\mathfrak{A}^a_{a_i}(q_i))$ is undefined. Since $\mathfrak{A}^a_{a_0}(q_0)=\mathfrak{A}^a(p_4\cdot p_5\cdot p_6)$, $\bar{d}(\mathfrak{A}^a(p_4\cdot p_5\cdot p_6))$ is undefined and $\mathfrak{A}^a_{a_l}(q_l)=\#$ thus $0\leqq i\leqq l-1$ holds. Now we consider the tree t_2 given by Lemma 2 for the tree p and the subtree $p_1\cdot p_2\cdot p_3\cdot q_i$. Let $q_i=f(r_1, \ldots, r_k)$

$(v(f)=k)$. Then there exists an index j $(1\leqq j\leqq k)$ such that $r_j=q_{i+1}$. Let us construct the tree $\bar{r}_j$ from r_j in exactly that way as the tree $\bar{t}$ has been constructed from the tree $\bar{p}$ in the proof of Lemma 2.

Furthermore, let t_1 be the tree arising from the tree $f(r_1, \ldots, r_{j-1}, \bar{r}_j, r_{j+1}, \ldots, r_k)$ in the same way as the tree t has been obtained from the tree $\bar{t}$ in Lemma 2. Let $t=t_1\cdot t_2$.

Consider the tree $q=p_1\cdot p_3^{L+1}\cdot t$, where $p_3^{L+1}=\{p_3\}^{L+1,\#}$. It is clear that $q\in T$, and

$$\mathfrak{A}(q)=\mathfrak{A}^a(t)$$

and

$$\mathfrak{B}(q)=\mathfrak{B}_b(p_1)\cdot(\mathfrak{B}_b^b(p_3))^{L+1}\cdot\mathfrak{B}^b(t)$$

hold by Lemma 1. Since $d(\mathfrak{A}(q))\leqq L$ and $d(\mathfrak{B}(q))>L$ thus $\mathfrak{A}(q)\neq\mathfrak{B}(q)$. But $d(q)<K$, which contradicts the minimality of p.

Lemma 5. Let $p\in T$ be a tree for which $\mathfrak{A}(p)\neq\mathfrak{B}(p)$. Assume that there exist trees $p_1', p_2', p_3', p_4', p_5', p_6'\in T_F(X_n\cup\{\#\})$ and states $a\in A$ and $b\in B$ satisfying the conditions of Lemma 3. If $\bar{d}(\mathfrak{A}^a(p_4'\cdot p_5'\cdot p_6'))$ is defined, then there exists a tree $\bar{p}\in T$ such that $d(\bar{p})<K$ and $\mathfrak{A}(\bar{p})\neq\mathfrak{B}(\bar{p})$.

Proof. Let S be the set of trees with minimal depth satisfying the conditions of Lemma 5. Let $p(\in S)$ be a tree which has minimal number of occurrences of symbols from F among all trees in S. Assume that $d(p)\geqq K$.

Let t be the tree given by Lemma 2 to the tree p and the subtree $p_1'\cdot p_2'\cdot p_3'\cdot p_4'\cdot p_5'$. We introduce the following notations:

$$p_1=p_1'\cdot p_2'\cdot p_3',\quad p_2=p_4',\quad p_3=p_5',\quad p_4=p_6'$$
$$\mathfrak{A}_a(p_1)=q_1,\quad \mathfrak{B}_b(p_1)=r_1,$$
$$\mathfrak{A}_a^a(p_2)=q_2,\quad \mathfrak{B}_b^b(p_2)=r_2,$$
$$\mathfrak{A}_a^a(p_3)=q_3,\quad \mathfrak{B}_b^b(p_3)=r_3,$$
$$\mathfrak{A}^a(p_4)=q_4,\quad \mathfrak{B}^b(p_4)=r_4,$$
$$\mathfrak{A}^a(t)=\bar{q}_4,\quad \mathfrak{B}^b(t)=\bar{r}_4.$$

First let us illustrate the idea of the proof in a special case. Assume that $v(f)=1$ and $\mu(g)=1$ for all $f\in F$ and $g\in G$. Then the DFRT transducers $\mathfrak{A}$ and $\mathfrak{B}$ may be considered as deterministic generalized sequential machines.

In Figure 1 we indicate the trees p, $\mathfrak{A}(p)$, $\mathfrak{B}(p)$. Now let us consider the trees $=p_1\cdot p_2^l\cdot t$ and $\mathfrak{A}(t_l)$, $\mathfrak{B}(t_l)$ $(l=1, \ldots, L+1)$ (see, Figure 2).

Since $\mathfrak{A}(t_l)=\mathfrak{B}(t_l)$ $(l=1, \ldots, L+1)$, thus Figure 2 shows that the same tree is constructed in two different ways. As it appears from Figure 2, and it can be readily verified, too, $q_2=\bar{r}_2\cdot\bar{q}_2$ and $r_2=\bar{q}_2\cdot\bar{r}_2$. The idea behind the proof of Lemma 5 is similar, but more involved.

The $\#$-depth of $\mathfrak{B}^b(p_4'\cdot p_5'\cdot p_6')$ is defined, for otherwise, by Lemma 4, there exists a tree $\bar{p}\in T$, for which $d(\bar{p})<K$ and $\mathfrak{A}(\bar{p})\neq\mathfrak{B}(\bar{p})$ hold contradicting the minimality of p. Since both $\bar{d}(\mathfrak{A}^a(p_2\cdot p_3\cdot p_4))$ and $\bar{d}(\mathfrak{B}^b(p_2\cdot p_3\cdot p_4))$ are defined thus all the trees q_2, q_3, q_4 and r_2, r_3, r_4 contain the symbol $\#$ in their frontiers. Moreover, by the note following Lemma 2, the frontiers of the trees $\bar{q}_4$ and $\bar{r}_4$ contain it, too.

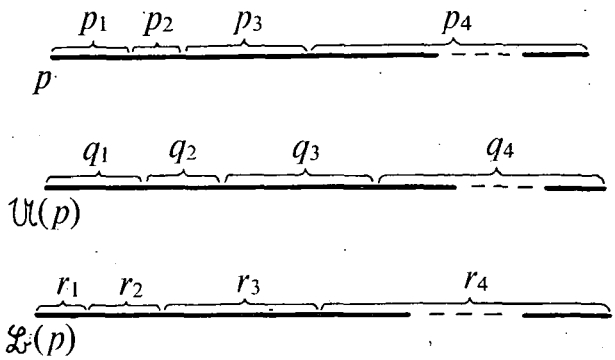


Fig. 1

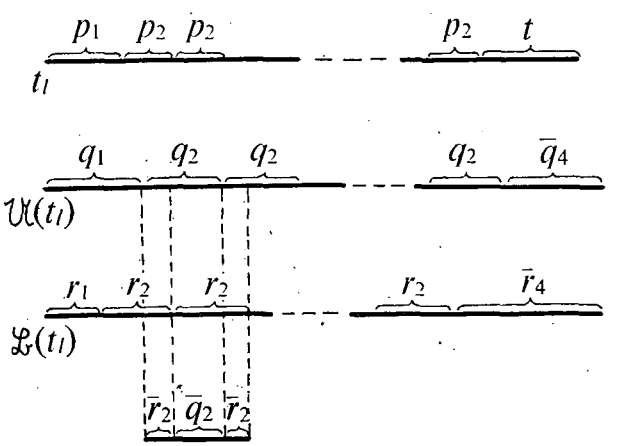


Fig. 2

Assume that $\bar{d}(q_2)=\bar{d}(r_2)=0$. Then

$$\mathfrak{A}(p) = q_1 \cdot q_3 \cdot q_4 = \mathfrak{A}(p_1 \cdot p_3 \cdot p_4)$$

and

$$\mathfrak{B}(p) = r_1 \cdot r_3 \cdot r_4 = \mathfrak{B}(p_1 \cdot p_3 \cdot p_4).$$

i.e. $\mathfrak{A}(p_1 \cdot p_3 \cdot p_4) \neq \mathfrak{B}(p_1 \cdot p_3 \cdot p_4)$, which is a contradiction.

In the same way we obtain that if $\bar{d}(q_3)=\bar{d}(r_3)=0$, then $\mathfrak{A}(p_1 \cdot p_2 \cdot p_4) \neq \mathfrak{B}(p_1 \cdot p_2 \cdot p_4)$, which is impossible.

Now we consider the trees

$$t_l = p_1 \cdot p_2^l \cdot t \quad \text{and} \quad s_l = p_1 \cdot p_3^l \cdot t \quad (l = 0, \ldots, L+1).$$

By Lemma 1, it follows that

$$\mathfrak{A}(t_l) = q_1 \cdot q_2^l \cdot \bar{q}_4, \quad \mathfrak{B}(t_l) = r_1 \cdot r_2^l \cdot \bar{r}_4,$$

$$\mathfrak{A}(s_l) = q_1 \cdot q_3^l \cdot \bar{q}_4, \quad \mathfrak{B}(s_l) = r_1 \cdot r_3^l \cdot \bar{r}_4 \quad (l = 0, \ldots, L+1).$$

Since $d(t_l), d(s_l)<K$ thus $\mathfrak{A}(t_l)=\mathfrak{B}(t_l)$ and $\mathfrak{A}(s_l)=\mathfrak{B}(s_l)$ $(l=0, \ldots, L+1)$. If exactly one of $\bar{d}(q_2)$ and $\bar{d}(r_2)$ is equal to zero, say $\bar{d}(q_2)=0$ and $\bar{d}(r_2)>0$, then $d(\mathfrak{A}(t_{L+1}))<d(\mathfrak{B}(t_{L+1}))$, consequently, $\mathfrak{A}(t_{L+1})\neq\mathfrak{B}(t_{L+1})$, which contradicts the minimality of p. It means that the following equalities are true:

$$d(\mathfrak{A}(t_l)) = \bar{d}(\bar{q}_4)+(l-1)\bar{d}(q_2)+d(q_1\cdot q_2)$$

and

$$d(\mathfrak{B}(t_l)) = \bar{d}(\bar{r}_4)+(l-1)\bar{d}(r_2)+d(r_1\cdot r_2) \quad (l=L, L+1).$$

This implies that $\bar{d}(q_2)=\bar{d}(r_2)>0$. Similarly, we get that $\bar{d}(q_3)=\bar{d}(r_3)>0$.

The tree $\mathfrak{A}(t_{L+1})$ is obtained from the trèe $\bar{q}_4$ by replacing all occurrences of the subtree # by the tree $q_1\cdot q_2^{L+1}$, while $\mathfrak{B}(t_{L+1})$ is given by replacing all occurrences of # in $\bar{r}_4$ by the tree $r_1\cdot r_2^{L+1}$.

We have that $d(\bar{q}_4)\leqq L$, $d(\bar{r}_4)\leqq L$ and $d(q_1\cdot q_2^{L+1})>L$, $d(r_1\cdot r_2^{L+1})>L$. Thus the equality $\mathfrak{A}(t_{L+1})=\mathfrak{B}(t_{L+1})$ implies that $r_1\cdot r_2^{L+1}\in\mathrm{sub}\,(q_1\cdot q_2^{L+1})$ or $q_1\cdot q_2^{L+1}\in\in\mathrm{sub}\,(r_1\cdot r_2^{L+1})$.

Assume that $r_1\cdot r_2^{L+1}\in\mathrm{sub}\,(q_1\cdot q_2^{L+1})$. Let j be the minimal number, for which $r_1\cdot r_2^{L+1}\in\mathrm{sub}\,(q_1\cdot q_2^j)$. Since $r_1\cdot r_2^{L+1}\in\mathrm{sub}\,(q_1\cdot q_2^{L+1})$ and $d(r_1\cdot r_2^{L+1})>d(q_1\cdot q_2)$ thus $2\leqq j\leqq L+1$.

Let $\bar{q}_2$ be the tree obtained from the tree $q_1\cdot q_2^j$ by replacing all occurrences of $r_1\cdot r_2^{L+1}$ by the symbol #. Therefore, $r_1\cdot r_2^{L+1}\cdot\bar{q}_2=q_1\cdot q_2^j$ and $r_1\cdot r_2^{L+1}\notin\mathrm{sub}\,(\bar{q}_2)$. Since j is minimal, it follows that $r_1\cdot r_2^{L+1}\notin\mathrm{sub}\,(q_1\cdot q_2^{j-1})$. On the other hand $r_1\cdot r_2^{L+1}\cdot\bar{q}_2=$ $=q_1\cdot q_2^{j-1}\cdot q_2$ and $r_1\cdot r_2^{L+1}\notin\mathrm{sub}\,(q_2)$. Therefore, $q_1\cdot q_2^{j-1}\in\mathrm{sub}\,(r_1\cdot r_2^{L+1})$.

Let $\bar{r}_2$ be the tree given from $r_1\cdot r_2^{L+1}$ by replacing all occurrences of $q_1\cdot q_2^{j-1}$ by the symbol #. Thus $q_1\cdot q_2^{j-1}\cdot\bar{r}_2=r_1\cdot r_2^{L+1}$ and $q_1\cdot q_2^{j-1}\notin\mathrm{sub}\,(\bar{r}_2)$. It means that $q_1\cdot q_2^{j-1}\cdot\bar{r}_2\cdot\bar{q}_2=q_1\cdot q_2^{j-1}\cdot q_2$.

Next we show that $q_1\cdot q_2^{j-1}\notin\mathrm{sub}\,(\bar{r}_2\cdot\bar{q}_2)$ holds, too. Indeed, if $q_1\cdot q_2^{j-1}\in\in\mathrm{sub}\,(\bar{r}_2\cdot\bar{q}_2)$, then $q_1\cdot q_2^{j-1}\in\mathrm{sub}\,(\bar{q}_2)$ because of $q_1\cdot q_2^{j-1}\notin\mathrm{sub}\,(\bar{r}_2)$ and $\bar{r}_2\notin\mathrm{sub}\,(q_1\cdot q_2^{j-1})$. Thus, in $q_1\cdot q_2^{j-1}\cdot q_2$ there exists a subtree $q_1\cdot q_2^{j-1}$, which is not a subtree of $r_1\cdot r_2^{L+1}$. But this is impossible since in this case one can show that $r_1\cdot r_2^{L+1}\in\mathrm{sub}\,(q_1\cdot q_2^{j-1})$. Therefore, one have

$$\bar{r}_2\cdot\bar{q}_2 = q_2.$$

Since $\mathfrak{A}(t_{L+1})=\mathfrak{B}(t_{L+1})$ thus

$$r_1\cdot r_2^{L+1}\cdot\bar{r}_4 = q_1\cdot q_2^{L+1}\cdot\bar{q}_4 = q_1\cdot q_2^j\cdot q_2^{L+1-j}\cdot\bar{q}_4 = r_1\cdot r_2^{L+1}\cdot\bar{q}_2\cdot q_2^{L+1-j}\cdot\bar{q}_4.$$

Furthermore, $r_1\cdot r_2^{L+1}$ is not a subtree of any of the trees $\bar{q}_4, q_2, \bar{q}_2, \bar{r}_4$. Thus the preceding equality implies

$$\bar{r}_4 = \bar{q}_2\cdot q_2^{L+1-j}\cdot\bar{q}_4.$$

We have $\mathfrak{A}(t_0)=\mathfrak{B}(t_0)$. Thus $q_1\cdot\bar{q}_4=r_1\cdot\bar{r}_4=r_1\cdot\bar{q}_2\cdot q_2^{L+1-j}\cdot\bar{q}_4$. Therefore,

$$q_1 = r_1\cdot\bar{q}_2\cdot q_2^{L+1-j}.$$

Using the equality $\mathfrak{A}(t_1)=\mathfrak{B}(t_1)$ we get

$$q_1\cdot q_2\cdot\bar{q}_4 = r_1\cdot\bar{q}_2\cdot(\bar{r}_2\cdot\bar{q}_2)^{L+1-j}\cdot(\bar{r}_2\cdot\bar{q}_2)\cdot\bar{q}_4,$$

$$r_1\cdot r_2\cdot\bar{r}_4 = r_1\cdot r_2\cdot\bar{q}_2\cdot(\bar{r}_2\cdot\bar{q}_2)^{L+1-j}\cdot\bar{q}_4.$$

This implies that $r_1 \cdot \bar{q}_2 \cdot \bar{r}_2 = r_1 \cdot r_2$. Furthermore, from the equalities $\mathfrak{A}(t_l) = \mathfrak{B}(t_l)$ $(l=0, \ldots, L+1)$, by induction, we obtain $r_1 \cdot (\bar{q}_2 \cdot \bar{r}_2)^{L+1} = r_1 \cdot (\bar{q}_2 \cdot \bar{r}_2)^L \cdot r_2$. Since $2 \leqq d(\mathfrak{A}(p_1 \cdot p_2 \cdot p_3 \cdot t)) \leqq L$, thus $d(r_1 \cdot (\bar{q}_2 \cdot \bar{r}_2)^L) > d(r_1 \cdot \bar{q}_2 \cdot \bar{r}_2) = d(r_1 \cdot r_2) \geqq d(r_2)$. Therefore, $r_1 \cdot (\bar{q}_2 \cdot \bar{r}_2)^L \notin \mathrm{sub}\,(r_2)$, implying

$$\bar{q}_2 \cdot \bar{r}_2 = r_2.$$

Now consider the trees $s_l = p_1 \cdot p_3^l \cdot t$ $(l=0, \ldots, L+1)$. Then $r_1 \cdot r_3^{L+1} \in \mathrm{sub}\,(q_1 \cdot q_3^{L+1})$ because of $\bar{r}_4 = \bar{q}_2 \cdot q_2^{L+1-j} \cdot \bar{q}_4$. In the above way we get that there are trees $\bar{q}_3, \bar{r}_3$ and a number i $(2 \leqq i \leqq L+1)$ such that

$$q_1 = r_1 \cdot \bar{q}_3 \cdot q_3^{L+1-i},$$

$$q_3 = \bar{r}_3 \cdot \bar{q}_3,$$

$$r_3 = \bar{q}_3 \cdot \bar{r}_3.$$

Since p is minimal thus

$$\mathfrak{A}(p_1 \cdot p_4) = \mathfrak{B}(p_1 \cdot p_4) \quad \text{and} \quad \mathfrak{A}(p_1 \cdot p_2 \cdot p_4) = \mathfrak{B}(p_1 \cdot p_2 \cdot p_4),$$

i.e.,

$$q_1 \cdot q_4 = r_1 \cdot r_4 \qquad \text{and} \qquad q_1 \cdot q_2 \cdot q_4 = r_1 \cdot r_2 \cdot r_4.$$

The first equality implies that $r_1 \cdot r_4 = r_1 \cdot \bar{q}_2 \cdot (\bar{r}_2 \cdot \bar{q}_2)^{L+1-j} \cdot q_4$. Consequently, r_4 can differ from $\bar{q}_2 \cdot (\bar{r}_2 \cdot \bar{q}_2)^{L+1-j} \cdot q_4$ in the tree r_1 only, i.e. whenever $\#$ is a subtree in one of them then the corresponding subtree in the other one should be r_1 or $\#$. By the above second equality we get

$$r_1 \cdot \bar{q}_2 \cdot \bar{r}_2 \cdot r_4 = r_1 \cdot \bar{q}_2 \cdot (\bar{r}_2 \cdot \bar{q}_2)^{L+1-j} \cdot (\bar{r}_2 \cdot \bar{q}_2) \cdot q_4.$$

Thus r_4 and $\bar{q}_2 \cdot (\bar{r}_2 \cdot \bar{q}_2)^{L+1-j} \cdot q_4$ can differ only in $r_1 \cdot r_2$. Thus, by $r_1 \cdot r_2 \neq r_1$, we have

$$r_4 = \bar{q}_2 \cdot (\bar{r}_2 \cdot \bar{q}_2)^{L+1-j} \cdot q_4.$$

Similarly, using the trees $p_1 \cdot p_4$ and $p_1 \cdot p_3 \cdot p_4$, we obtain

$$r_4 = \bar{q}_3 \cdot (\bar{r}_3 \cdot \bar{q}_3)^{L+1-i} \cdot q_4.$$

Therefore, $\bar{q}_2 \cdot (\bar{r}_2 \cdot \bar{q}_2)^{L+1-j} \cdot q_4 = \bar{q}_3 \cdot (\bar{r}_3 \cdot \bar{q}_3)^{L+1-i} \cdot q_4$ implying

$$\bar{q}_2 \cdot (\bar{r}_2 \cdot \bar{q}_2)^{L+1-j} = \bar{q}_3 \cdot (\bar{r}_3 \cdot \bar{q}_3)^{L+1-i}.$$

Finally, using the above equalities, we get

$$q_1 \cdot q_2 \cdot q_3 \cdot q_4 = r_1 \cdot \bar{q}_2 \cdot (\bar{r}_2 \cdot \bar{q}_2)^{L+1-j} \cdot (\bar{r}_2 \cdot \bar{q}_2) \cdot q_3 \cdot q_4 =$$

$$= r_1 \cdot (\bar{q}_2 \cdot \bar{r}_2) \cdot \bar{q}_2 \cdot (\bar{r}_2 \cdot \bar{q}_2)^{L+1-j} \cdot q_3 \cdot q_4 = r_1 \cdot r_2 \cdot \bar{q}_3 \cdot (\bar{r}_3 \cdot \bar{q}_3)^{L+1-i} \cdot (\bar{r}_3 \cdot \bar{q}_3) \cdot q_4 =$$

$$= r_1 \cdot r_2 \cdot (\bar{q}_3 \cdot \bar{r}_3) \cdot \bar{q}_3 \cdot (\bar{r}_3 \cdot \bar{q}_3)^{L+1-i} \cdot q_4 = r_1 \cdot r_2 \cdot r_3 \cdot r_4,$$

i.e., $\mathfrak{A}(p) = \mathfrak{B}(p)$ contradicting our assumption.

Similarly, we arrive at a contradiction by assuming

$$q_1 \cdot q_2^{L+1} \in \mathrm{sub}\,(r_1 \cdot r_2^{L+1}).$$

This means that the depth of p is smaller than K ending the proof of this lemma.

Theorem 6. The equivalence problem of deterministic frontier-to-root tree transducers is effectively solvable.

Proof. Consider two arbitrary DFRT transducers $\mathfrak{A}=(F, A, G, A', \Sigma_A)$ and $\mathfrak{B}=(F, B, G, B', \Sigma_B)$. The set of all trees p, for which $\mathfrak{A}(p)$ and $\mathfrak{B}(p)$ are defined, is a regular set of trees, which can be given effectively (see, Corollary 3.12. in [1]). Thus, the problem whether or not the domains of mappings induced by $\mathfrak{A}$ and $\mathfrak{B}$ are equal is solvable. If they are not equal, then the transducers are not equivalent. In the opposite case, by Lemmas 4 and 5 it is sufficient to check whether their translations coincide on a finite number of trees. This ends the proof of Theorem 6.

Finally, we present a result concerning the equivalence problem in a special class of deterministic root-to-frontier tree transducers.

Let $\mathfrak{M}$ be the set of deterministic root-to-frontier tree (DRFT) transducers $\mathfrak{A}=(F, A, G, A', \Sigma)$ with the following property: if $af(z_1, \ldots, z_k)\to q$ is in Σ $(v(f)=k, k>0)$, then there are states $a_1, \ldots, a_k\in A$ such that $q\in T_G(Y\cup\{(a_i, z_i)|i=1, \ldots, k\})$. For such DRFT transducers one can prove Lemmas 1—5. Thus we have

Theorem 7. The equivalence problem of DRFT transducers in $\mathfrak{M}$ is effectively solvable.

4. Minimalization of DFRT transducers

Take a DFRT transducer $\mathfrak{A}=(F, A, G, A', \Sigma_A)$ such that the mapping induced by $\mathfrak{A}$ is from $T_F(X_n)$ into $T_G(Y_m)$. Moreover, let p be an arbitrary tree, for which $\mathfrak{A}(p)$ is defined, i.e., $p\in\mathfrak{A}^{-1}(T_G(Y_m))$. In this case for any $\bar{p}\in\mathrm{sub}(p)$ of the form $\bar{p}=f(p_1, \ldots, p_k)$ or $\bar{p}=x_i$, there is exactly one rule in Σ_A, denoted by $\sigma(\bar{p})$ such that if $\sigma(\bar{p})=f(a_1z_1, \ldots, a_kz_k)\to A_{\bar{p}}q$ then

$$\mathfrak{A}_{A_{\bar{p}}}(\bar{p}) = q[\mathfrak{A}_{a_1}(p_1)\to z_1, \ldots, \mathfrak{A}_{a_k}(p_k)\to z_k],$$

and

$$\mathfrak{A}_{A_{\bar{p}}}(\bar{p}) = q \quad \text{if} \quad \sigma(\bar{p}) = x_i \to A_{\bar{p}}q.$$

Lemma 8. Let $p\in\mathfrak{A}^{-1}(T_G(Y_m))$ and $\bar{p}\in\mathrm{sub}(p)$ be arbitrary. Then there exist a $p'\in\mathfrak{A}^{-1}(T_G(Y_m))$ and a $\bar{p}'\in\mathrm{sub}(p')$, such that $\sigma(\bar{p})=\sigma(\bar{p}')$ and $d(p')<2|A|$.

Proof. Let $\bar{p}$ denote the tree obtained by replacing the subtree $\bar{p}$ in p by $\#$. Let $\bar{p}'$ be the tree given by Lemma 2 to the tree p and its subtree $\bar{p}$. Assume, that $\bar{p}=f(p_1, \ldots, p_k)$. Let us construct the tree $\bar{p}_i$ from p_i $(i=1, \ldots, k)$ in exactly that way as the tree $\bar{t}$ has been constructed from the tree $\bar{p}$ in the proof of Lemma 2 $(i=1, \ldots, k)$. Let $\bar{p}'=f(\bar{p}_1, \ldots, \bar{p}_k)$ and $p'=\bar{p}'\cdot\bar{p}'$. From the construction it is clear, that the trees p' and $\bar{p}'$ satisfy the conditions of Lemma 8. A similar argument can be used in the case $\bar{p}=x_i$.

Let $L=\max\left(d(\mathfrak{A}(p))|p\in\mathfrak{A}^{-1}(T_G(Y_m)),\ d(p)<2|A|\right)$.

Lemma 9. There exists a minimal DFRT transducer $\mathfrak{B}=(F, B, G, B', \Sigma_B)$ belonging to $\mathfrak{A}$ such that if $x_i\to bq$ or $f(b_1z_1, \ldots, b_kz_k)\to bq$ is in Σ_B then $d(q)\leqq L$.

Proof. Let $\mathfrak{B}$ be a minimal DFRT transducer belonging to $\mathfrak{A}$. Assume that there exist $p\in\mathfrak{B}^{-1}(T_G(Y_m))$ and $\bar{p}\in\mathrm{sub}(p)$ such that the depth of the right hand side of

$\sigma(\bar{p})$ is greater, than L. We show that $\bar{d}(\mathfrak{B}(\bar{p}))$ is undefined, where $\bar{p}$ is obtained by replacing $\bar{p}$ in p by $\#$.

Indeed, by Lemma 8, there exist trees p' and $\bar{p}'$, $\bar{\bar{p}}'$, for which $p'=\bar{p}'\cdot\bar{\bar{p}}'$, $\sigma(\bar{p})=\sigma(\bar{p}')$, $p'\in\mathfrak{B}^{-1}(T_G(Y_m))$ and $d(p')<2|B|\leqq 2|A|$.

By the note following Lemma 2, if $\bar{d}(\mathfrak{B}(\bar{p}))$ is defined then so is $d(\mathfrak{B}(\bar{\bar{p}}'))$. But $d(\mathfrak{B}(p'))\geqq\bar{d}(\mathfrak{B}(\bar{p}'))+d(\mathfrak{B}(\bar{\bar{p}}'))$. Furthermore, by our assumption $d(\mathfrak{B}(\bar{\bar{p}}'))>L$. Thus $d(\mathfrak{B}(p'))>L$ which is a contradiction since $\mathfrak{B}(p')=\mathfrak{A}(p')$ and $d(p')<2|B|\leqq\leqq 2|A|$.

Now for all $\sigma=f(b_1z_1,\ldots,b_kz_k)\to bq$ and $\sigma=x_i\to bq$ with $d(q)>L$, let us replace σ in Σ_B by $\bar{\sigma}=f(b_1z_1,\ldots,b_kz_k)\to by_1$ and $\bar{\sigma}=x_i\to by_1$, respectively, and denote the resulting set of rules by $\bar{\Sigma}_B$. Then the DFRT transducer $\bar{\mathfrak{B}}=(F,B,G,B',\bar{\Sigma}_B)$ is equivalent to $\mathfrak{B}$, completing the proof of Lemma 9.

Theorem 10. There exists an algorithm for determining to any DFRT transducer $\mathfrak{A}=(F,A,G,A',\Sigma_A)$ a minimal DFRT transducer belonging to $\mathfrak{A}$.

Proof. Let $|A|=M$ and $L=\max\left(d(\mathfrak{A}(p))|p\in\mathfrak{A}^{-1}(T_G(Y)),\ d(p)<2M\right)$. Then for a minimal DFRT transducer belonging to $\mathfrak{A}$, it holds that the number of its states is less than or equal to M. Furthermore, by Lemma 9, we can assume that the depths of right hand sides of rules of a minimal DFRT transducer belonging to $\mathfrak{A}$ are less than or equal to L. But there is only a finite number of DFRT transducers satisfying these two assumptions. This means that it is enough to check only for finitely many DFRT transducers whether they are equivalent to $\mathfrak{A}$.

After determining all such DFRT transducers equivalent to $\mathfrak{A}$, we choose one of them with minimal number of states.

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF SZEGED
H—6720 SZEGED, HUNGARY
ARADI V. TERE 1.

References

[1] ENGELFRIET, J., Bottom-up and top-down tree transformations —.a comparison, *Math. Systems Theory*, v. 9, 1975, pp. 198—231.

[2] GRIFFITHS, T. V., The unsolvability of the equivalence problem for λ-free nondeterministic generalized machines, *J. Assoc. Comput. Mach.*, v. 15, 1968, pp. 409—413.

(Received Feb. 22, 1978)

A note on deadlocks

By Z. LABORCZI

Introduction

A set of processes uses resources of several types concurrently. We assume that there is only a limited number of resources from each type. The number of resources can be characterized by a vector **t**, where t_i is the total number of resources of type i.

Another constraint is that the processes cannot be forced to release resources they currently use.

If we know nothing about the behaviour of the processes, then the only possible way for scheduling the processes is the strictly sequential ordering. Real concurrency could not be allowed because each concurrent process may request all the resources at any time, and this request cannot be fullfilled if resources are allocated to other processes.

Therefore, we must have information on the behaviour of the processes in form of some kind of limitation the processes comply with.

One possible limitation among others [4] is the following: on entering the system, a process p has to announce a vector $\mathbf{goal}(p)$ declaring that it will not use more than $goal_i(p)$ resources from the i-th resource type. In order to be able to satisfy other requests, processes are not allowed to work forever, that is, if we place $\mathbf{goal}(p)$ resources at p's disposal and wait, p will terminate in finite time and return all the resources allocated to it. When p starts, it usually does not need all $\mathbf{goal}(p)$ resources immediately, and if we want to describe the current state of p, we have to introduce the vector $\mathbf{alloc}(p)$ which tells us how many resources have been allocated to p. It is clear that $\mathbf{alloc}(p) \leqq \mathbf{goal}(p)$, and the difference

$$\mathbf{need}(p) = \mathbf{goal}(p) - \mathbf{alloc}(p)$$

shows how many resources p still needs in order to complete. We may assume that there is no process in the system with $\mathbf{need}(p) = \mathbf{0}$ for if that is the case, we wait until p completes and continue the examination of the system only after the completion.

1. Graphical representation of processes

We describe a graphical representation of competing processes, which will turn out to be very useful later. The current need of a process may be represented by a point of the l-dimensional space, where l is the number of different resource types. A process p starts from **goal**(p) and currently stays at **need**(p), thus it is very natural to think of p as an arrow from **goal**(p) to **need**(p).

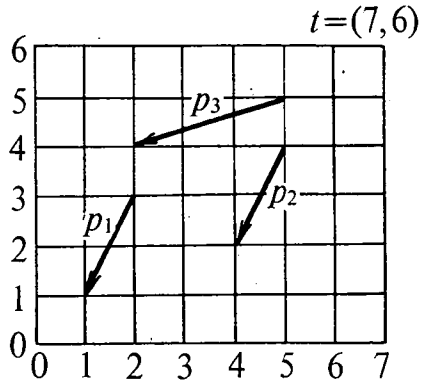


Figure 1

Figure 1 visualizes three processes competing for resources of two different types.

The number of resources in the system equals (7, 6), furthermore,

goal$(p_1)=(2, 3)$,
need$(p_1)=(1, 1)$,
goal$(p_2)=(5, 4)$,
need$(p_2)=(4, 2)$,
goal$(p_3)=(5, 5)$,
need$(p_3)=(2, 4)$.

This kind of representation is applicable only if $l=2$. For greater values of l one needs to be highly imaginative.

2. The definition of deadlock

Informally speaking, a system of concurrent processes is in deadlock, if there is no guarante that every process can complete. In other words the system is free of deadlocks if all the processes can finish, even if they request all their **needs** immediately. A formal definition of the latter assertion is the following:

Definition. The set of processes π is said to be free of deadlock (or deadlock-free) if there exists a permutation $p_1, p_2, \ldots, p_k$ of the processes in π such that

$$\mathbf{need}(p_i) \leqq \mathbf{t} - \sum_{j=i}^{k} \mathbf{alloc}(p_j)$$

for $i=1, 2, \ldots, k$.

This inequality means that if $p_1, p_2, \ldots, p_{i-1}$ have completed and returned the resources they used, then the **need** of p_i does not exceed the amount of the currently available free resources.

The following theorem is sometimes stated as another definition of the deadlock.

Theorem 1. The set of processes π contains a deadlock (or is in deadlock) if and only if there exists a nonempty subset π' of π such that the following inequality holds for every p in π':

$$\mathbf{need}(p) \not\leqq \mathbf{free}(\pi') \tag{1}$$

where $\mathbf{free}(\pi')=\mathbf{t}-\sum_{q \text{ in } \pi'} \mathbf{alloc}(q)$ is the amount of the resources currently not used by the processes in π'.

3. A condition and an algorithm

The definition of the deadlock and Theorem 1 speak about permutations and subsets of the process set. It is desirable to find a necessary and sufficient condition in which these notions do not occur, or in other words in which every process is mentioned only once, and not as a member of a permutation or a subset. This is accomplished by Theorem 2.

Theorem 2. Let $\mathbf{n}$ be a vector describing some amount of resources and $\mathbf{n}<\mathbf{t}$ ($\mathbf{t}$ is the total number of resources)[1]. A set of processes π is deadlock-free if and only if for every such $\mathbf{n}$ we have:

$$\sum_{\substack{p \text{ in } \pi \text{ and} \\ \mathbf{need}(p) \not\leqq \mathbf{n}}} \mathbf{alloc}(p) \not\geqq \mathbf{t}-\mathbf{n}. \tag{2}$$

For a deadlock, we can not only state that we can find an $\mathbf{n}$ such that $\geqq$ holds in (2) instead of $\not\geqq$, but we can replace $\geqq$ by $=$, that is the following assertion holds: π is in deadlock if and only if there is an $\mathbf{n}<\mathbf{t}$ such that

$$\sum_{\substack{p \text{ in } \pi \text{ and} \\ \mathbf{need}(p) \not\leqq \mathbf{n}}} \mathbf{alloc}(p) = \mathbf{t}-\mathbf{n} \tag{3}$$

Proof. It is sufficient to prove that

I. if π is deadlock-free, then (2) holds for each $\mathbf{n}$;

II. if π contains a deadlock, then (3) holds for at least one $\mathbf{n}$.

Proof of I. Let π be deadlock-free and $\mathbf{n}<\mathbf{t}$. We define

$$\Phi(\mathbf{n}) = \{q \text{ in } \pi | \mathbf{need}(p) \not\geqq \mathbf{n}\}.$$

$\Phi(\mathbf{n})$ contains exactly those processes for which (2) forms a sum, so if $\Phi(\mathbf{n})$ is empty, (2) is true.

Otherwise we apply Theorem 1 for $\Phi(\mathbf{n})$ and find a p in $\Phi(\mathbf{n})$ for which

$$\mathbf{need}(p) \leqq \mathbf{t} - \sum_{q \text{ in } \Phi(\mathbf{n})} \mathbf{alloc}(q) \tag{4}$$

For p as a member of $\Phi(\mathbf{n})$ we have also $\mathbf{need}(p) \not\leqq \mathbf{n}$, and replacing the left hand side of this inequality by the right hand side of (4) we get (2).

Proof of II. Let π be in deadlock and let π' be a maximal subset of π which statisfies Theorem 1, i.e., if we put a new element to π', (1) will not be true. We prove that (3) holds for

$$\mathbf{n} = \mathbf{t} - \sum_{q \text{ in } \pi'} \mathbf{alloc}(q)$$

by showing that $\Phi(\mathbf{n})=\pi'$.

Assuming that p is in $\Phi(\mathbf{n})$, we recall the definition of Φ and $\mathbf{n}$

$$\mathbf{need}(p) \not\geqq \mathbf{t} - \sum_{q \text{ in } \pi'} \mathbf{alloc}(q).$$

[1] $\mathbf{n}<\mathbf{t}$ means that $\leqq$ holds for every component and $<$ holds for at least one component.

It is now easy to verify that (1) holds for the subset $\pi' \cup \{p\}$ and as π' is a maximal subset in deadlock, p is in π'.

Starting from the other end, assume that p is in π'. From (1) we get

$$\mathbf{need}(p) \not\leqq \mathbf{t} - \sum_{q \text{ in } \pi'} \mathbf{alloc}(q),$$

that is $\mathbf{need}(p) \not\leqq \mathbf{n}$ and this means that p is in $\Phi(\mathbf{n})$.

After completing the proof of Theorem 2, we append the following remark to the last step of the proof:

At the very end we showed that π' is a subset of $\Phi(\mathbf{n})$. This means that

$$\mathbf{t} - \mathbf{n} = \sum_{p \text{ in } \pi'} \mathbf{alloc}(p) \leqq \sum_{p \text{ in } \Phi(\mathbf{n})} \mathbf{alloc}(p),$$

and this is exactly the negation of (2). If π happens to be a maximal subset, then the equality will hold.

If we write the formula of Theorem 2 in the following way

$$\mathbf{t} - \sum_{p \text{ in } \Phi(\mathbf{n})} \mathbf{alloc}(p) \not\leqq \mathbf{n} \tag{5}$$

we might formulate the meaning of Theorem 2 in terms of the arrows introduced earlier. Choosing an $\mathbf{n}$ we select the members of π which (as arrows) lie outside the rectangle consisting of the points less than or equal to $\mathbf{n}$. This set is $\Phi(\mathbf{n})$. (5) states that if we start at $\mathbf{t}$ and decrease our coordinates by $\mathbf{alloc}(p)$ for each p in $\Phi(\mathbf{n})$, we eventually reach a point lying still outside the rectangle.

For a deadlock state there must be an $\mathbf{n}$ such that the resultant point is not only within the rectangle but is identical to $\mathbf{n}$.

Exploiting these facts we may devise an algorithm to decide whether a set of processes is in deadlock or not. The algorithm runs as follows:

I. We define the function $\mathbf{f}$:

$$\mathbf{f}(\mathbf{n}) = \mathbf{t} - \mathbf{n}$$

for every $\mathbf{n} < \mathbf{t}$.

II. For every p in π, decrease the value of $\mathbf{f}$ by $\mathbf{alloc}(p)$ in the points of its domain for which

$$\mathbf{n} \not\geqq \mathbf{need}(p) \quad \text{and} \quad \mathbf{n} \leqq \mathbf{t} - \mathbf{alloc}(p) \tag{6}$$

III. Test after each decrease, whether the new value of $\mathbf{f}$ is 0. If so, we have a deadlock situation, otherwise if no 0 occured while performing II the system is free of deadlocks.

Condition (6) in step II needs some explanation. Requiring $\mathbf{n} \not\geqq \mathbf{need}(p)$ guarantees that p is in $\Phi(\mathbf{n})$. However, not all such $\mathbf{n}$-s have to be taken into consideration, because the decrease of $\mathbf{f}$ may result in 0 only for $\mathbf{n}$-s for which $\mathbf{n} \leqq \mathbf{t} - \mathbf{alloc}(p)$ is also true.

The small circles on Figure 2 indicate the points for which $\mathbf{f}$ has to be decreased in connection with p_3. For the point $\mathbf{n} = (3, 3)$ we have initially $\mathbf{f}(\mathbf{n}) = (4, 3)$. On executing step II for p_2 and p_3, $f(\mathbf{n})$ becomes 0, so these two processes are in deadlock independently of the existence of p_1.

The algorithm as described above detects deadlock within a system. A slight

modification makes it capable of handling the deadlock avoidance problem. Let us assume that π is deadlock-free and a process p requests some more resources. If we update **f(n)** by performing step II and step III for p only, it can be decided whether the fulfilling of the new request leads to a deadlock state.

Unfortunately this algorithm is not suitable for being incorporated into a real system, because it requires a great amount of space and time. Let us assume that we have 256 memory pages and 4—4 peripheral devices from two different types as resources at our disposal. In this case **t** becomes (256, 4, 4) and the number of **n**-s is $(256+1)*(4+1)*(4+1)-1=6424$.

Therefore we need almost 100,000 bits for representing the function **f**, and we have mentioned nothing about the time needed to update such amount of information. Therefore, we conclude that the general algorithms for detecting and avoiding deadlock [2, 3] have to be applied in the general case.

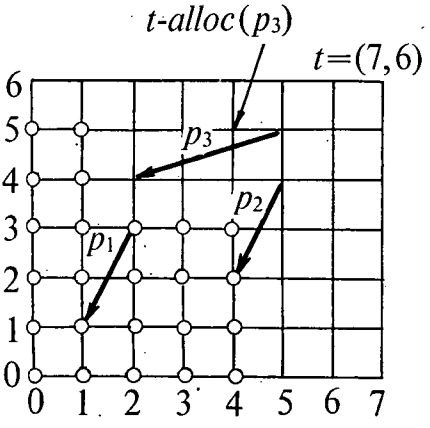


Figure 2

In one dimension, that is for one resource type, however, everything becomes very simple. There are no vectors, thus we may replace $\not\leq$ and $\not\geq$ by $>$ and $<$ respectively. In addition **f** becomes a scalar to scalar function. The updating of f is also less complicated: for a process p, f has to be decreased in those n-s where $n<need(p)$ holds. In fact we have arrived at a modified version of Habermann's theorem on detecting deadlocks for one resource type [1], thus this paper generalizes his results.

Conclusion

On investigating whether a simple theorem on detecting deadlock with one resource type can be generalized to more resource types, we found that the answer is yes, but the new theorem still needs further works to develop a practically usable algorithm.

Acknowledgement. I am indebted to Mr. J. Somogyi for many fruitful discussions on the subject of this paper and related areas.

Abstract

According to a paper by A. N. Habermann there is a simple necessary and sufficient condition whether a set of processes utilizing a limited number of resources of the same type is in deadlock. We show that this condition may be generalized to the case of more than one resource types. This result can be illustrated graphically in a clear way. The question of constructing algorithms on the basis of the extended condition is considered as well.

RESEARCH INSTITUTE FOR
APPLIED COMPUTER SCIENCES
H—1536 BUDAPEST, HUNGARY
P. O. BOX 227.

References

[1] HABERMANN, A. N., A new approach to avoidance of system deadlocks, Operating Systems, Proceeding of an International Symposium, *Lecture Notes in Computer Science*, v. 16, 1974.
[2] HABERMANN, A. N., Efficient deadlock avoidance algorithms, Unpublished paper, Intended to present at the Winter School, Budapest, 1977.
[3] HOLT, R. C., Some deadlock properties of computer systems, *Operating System Review*, v. 6, June 1972.
[4] DEVILLERS, R., Game interpretation of the deadlock avoidance problem, *Comm. ACM*, v. 20, No. 10, 1977.

(Received April 14, 1978)

Use of Petri nets for performance evaluation

By J. SIFAKIS

Introduction

Petri nets [1], [2] have been found a simple and elegant formalism for the description of asynchronous systems with concurrent evolutions. According to the adopted interpretation, they can be used to model flow phenomena of information, of energy and of materials [3], [4] and [5]. However, this model is not complete enough for the study of system performances since no assumption is made about the firing of a transition as far as its duration and the moment at which it takes place after the transition has been enabled.

Timed Petri nets have been introduced by C. Ramchandani [6] by associating firing times to the transitions of Petri nets. He studied the steady state behavior and gave methods for calculating the throughput rate for certain classes of Petri nets. The results given in this paper are applicable to the class of pure [7] Petri nets and generalize, in some sense, those presented in [6]. The litterature on timed Petri nets is very poor: to the author's knowledge, the only works on this subjet are the Ramchandani's thesis and a paper by S. Ghosh [8] comparing the properties of boundedness and liveness for timed Petri nets and unrestricted Petri nets.

I. Definitions

Definition 1. A Petri net (PN) is a quadruple $\mathcal{N}=(P, T, \alpha, \beta)$ where
P is a set of *places*, $P\neq\emptyset$
T is a set of transitions, $T\neq\emptyset$, $P\cap T=\emptyset$
α: $P\times T\rightarrow\mathbf{N}$ forward incidence function
β: $P\times T\rightarrow\mathbf{N}$ backward incidence function
($\mathbf{N}$ represents the set of natural numbers: 0, 1, 2, 3, ...).

REPRESENTATION. To a PN one can associate a digraph the nodes of which are the places and the transitions, represented respectively by circles and dashes. There is a directed edge from the place p_s to the transition t_j iff $\alpha(p_s, t_j)=n_{sj}\neq 0$. This edge is labeled by the value n_{sj}, called weight of the edge. There also is a directed edge from the transition t_r to the place p_w iff $\beta(p_w, t_r)=n_{w_r}\neq 0$. This edge is labeled by the weight n_{w_r}.

Definition 2. Let $\mathcal{N}=(P, T, \alpha, \beta)$ be a PN. We adopt the following notations:
For $t\in T$, ${}^{\bullet}t=\{p\in P|\alpha(p, t)\neq 0\}$ and $t^{\bullet}=\{p\in P|\beta(p, t)\neq 0\}$.
For $p\in P$, ${}^{\bullet}p=\{t\in T|\beta(p, t)\neq 0\}$ and $p^{\bullet}=\{t\in T|\alpha(p, t)\neq 0\}$.
We call ${}^{\bullet}t\,(t^{\bullet})$ *set of input (output) places* of t and by analogy, ${}^{\bullet}p\,(p^{\bullet})$ *set of input (output) transitions* of p. These notations are extended to subsets of T and P; for example, if $P_1\subset P$ then, ${}^{\bullet}P_1=\bigcup_{p_k\in P_1}{}^{\bullet}p_k$.

Definition 3. A marking M of a PN $\mathcal{N}=(P, T, \alpha, \beta)$ is a mapping of P into $\mathbf{N}: P\xrightarrow{M}\mathbf{N}$. When $|P|=n$, one can represent a marking M by a vector $M\in\mathbf{N}^n$ such that its i-th entry $m_i=M(p_i)$.

Definition 4. A transition t of a PN is *enabled* for a marking M iff:

$$\forall p\in{}^{\bullet}t\Rightarrow\alpha(p, t)\leqq M(p).$$

Definition 5. Let $\mathcal{M}_t$ be the set of markings for which a transition t of a PN is enabled. The *firing* of the transition $t(F(t))$ is a mapping of $\mathcal{M}_t$ into the set of the markings $\mathcal{M}$ defined as follows: if $F(t)[M_i]=M_j$ then

$$M_j(p)=\begin{cases}M_i(p), \forall p\notin{}^{\bullet}t\cup t^{\bullet},\\ M_i(p)-\alpha(p, t), \forall p\in{}^{\bullet}t-({}^{\bullet}t\cap t^{\bullet}),\\ M_i(p)+\beta(p, t), \forall p\in t^{\bullet}-({}^{\bullet}t\cap t^{\bullet}),\\ M_i(p)+\beta(p, t)-\alpha(p, t), \forall p\in{}^{\bullet}t\cap t^{\bullet}.\end{cases}$$

Definition 6. Let $\mathcal{N}=(P, T, \alpha, \beta)$ be a PN and M_0 one of its markings. Consider a sequence of transitions $\sigma=t_{j_1}t_{j_2}\ldots t_{j_s}$. We say that σ is a *simulation sequence* or a *firing sequence* from M_0 iff there exists a sequence of markings $M_1, M_2, M_3, \ldots, \ldots, M_s$ such that $F(t_{j_i})[M_{i-1}]=M_i$ for $i=1, 2, 3, \ldots, s$. We note that $M_0\xrightarrow{\sigma}M_s$. M_s is the marking *attained* by applying σ from M_0. We denote by $\vec{M}_0$ the set of markings that can be attained from M_0. The *firing vector* of σ is a vector $R(R\in\mathbf{N}^m, m=|T|)$ such that its k-th entry is equal to the number of occurrences of the transition t_k in σ.

Definition 7. An *ordinary* PN is a PN $\mathcal{N}=(P, T, \alpha, \beta)$, such that $\alpha: P\times T\to\{0, 1\}$ and $\beta: P\times T\to\{0, 1\}$. A *marked graph* is an ordinary PN such that $\forall p\in P$, $|{}^{\bullet}p|\leqq 1$ and $|p^{\bullet}|\leqq 1$. A *state graph* is an ordinary PN such that $\forall t\in T$, $|t^{\bullet}|\leqq 1$ and $|{}^{\bullet}\tau|\leqq 1$.

Definition 8. Take a PN and one of its markings M_0.
We say that a *place p is bounded for M_0* iff $\exists k\in\mathbf{N}$ such that $\forall M\in\vec{M}_0$, $M(p)<k$. A PN is *bounded for M_0* iff all its places are bounded. We say that a *transition t is live* for M_0 iff for every marking $M, M\in\vec{M}_0$, there exists a sequence $\sigma, \sigma\in T^*$ such that σt is a firing sequence from M. A net having all its transitions live for a marking M_0 is called *live for M_0*.

Definition 9. A *pure* PN is a PN such that $\forall t\in T$, $\{{}^{\bullet}t\}\cap\{t^{\bullet}\}=\emptyset$.
For a pure PN $\mathcal{N}=(P, T, \alpha, \beta)$ ($|P|=n$, $|T|=m$) one can define the matrices:

$$C=[c_{ij}]_{n\times m}\quad\text{with}\quad c_{ij}=\begin{cases}\beta(p_i, t_j) & \text{if}\quad \beta(p_i, t_j)\neq 0,\\ -\alpha(p_i, t_j) & \text{if}\quad \alpha(p_i, t_j)\neq 0,\\ 0 & \text{otherwise.}\end{cases}$$

C is called *incidence matrix* of the net [7].

$$C^+ = [c_{ij}^+]_{n\times m} \quad \text{with} \quad c_{ij}^+ = \begin{cases} \beta(p_i, t_j) & \text{if} \quad \beta(p_i, t_j) \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$C^- = [c_{ij}^-]_{n\times m} \quad \text{with} \quad c_{ij}^- = \begin{cases} \alpha(p_i, t_j) & \text{if} \quad \alpha(p_i, t_j) \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

REMARK. $C=C^+-C^-$.

II. Timed Petri nets

II. 1. Definitions

Definition 10. A *timed Petri net* (TPN) consists in giving:
a Petri net $\mathcal{N}=(P, T, \alpha, \beta)$,
$T=(\tau_1, \tau_2, \ldots, \tau_i, \ldots)$ an increasing sequence of real numbers called *time base*,
a mapping v: $P\times T\to T$ such that $\forall\,(p, \tau_i)\in P\times T$: $v(p, \tau_i)=\tau_j \Rightarrow \tau_j \geqq \tau_i$.

SIMULATION RULES

a) A marker in a TPN may be in one of the two following states: *available* or *unavailable*. Initially each place p contains $M_0(p)$ available markers.

b) A transition t is enabled iff every place $p_s(p_s\in {}^\bullet t)$ contains $\alpha(p_s, t)$ available markers at least.

c) The firing of a transition t has to take place instantaneously as soon as t is enabled. It consits in removing $\alpha(p_s, t)$ available markers from each place p_s $(p_s\in {}^\bullet t)$ and in placing $\beta(p_w, t)$ markers in each place $p_w\in t^\bullet$.

d) A marker remains unavailable in a place p_s during the time interval between the instant of its arrival τ_i and the instant $v(p_s, \tau_i)$; then it becomes available.

REMARK. According to the above definition, firings in a TPN take place only at moments of T.

In what follows we study the behaviour of pure TPN's such that $\forall\,(p_i, \tau)\in$ $\in P\times T$: $\forall\,(p_i, \tau)-\tau=z_i=$constant. That is, each marker is delayed by z_i in the place independently of the instant of its arrival.

Definition 11. Take a TPN and let $M_1 M_2 \ldots M_s$ be the markings attained successively from an initial marking M_0 by applying a firing sequence $\sigma=t_{i_0}, t_{i_1}, \ldots,$ $\ldots, t_{i_{s-1}}$ and $\tau_{i_0}, \tau_{i_1}, \ldots, \tau_{i_{s-1}}$ the moments of firing of the transitions $t_{i_0}, t_{i_1}, \ldots, t_{i_{s-1}}$ respectively. The marking of the net at a moment τ_{i_k} will be by definition the marking of the net in the interval $\tau_{i_{k-1}}\leqq\tau'<\tau_{i_k}$. Generally, the marking of a TPN with $T=(\tau_0, \tau_1, \tau_2, \ldots, \tau_i, \ldots)$ at a moment $\tau_i\in T$ will be the marking of the net at the interval $\tau_{i-1}\leqq\tau<\tau_i$, $i=1, 2, 3, \ldots$ The marking at τ_0 corresponds to the initial marking. For a TPN with n places we define a general temporal variable $Q^t(\tau)=$ $=[q_1(\tau), q_2(\tau), \ldots, q_n(\tau)]$ such that $\forall\tau_i\in T$, $Q(\tau_i)=M$ where M is the marking at the moment τ_i (Q^t denotes the transpose of a matrix Q). The variable $Q(\tau)$ will be called *charge variable.*

Let M_s be a marking attained from a marking M_0 by applying a sequence σ, $(M_0 \xrightarrow{\sigma} M_s)$ in a PN defined by its incidence matrix C. Then

$$M_s = M_0 + CR \tag{I}$$

where $R \in \mathbf{N}^m$, $m = |T|$, is the firing vector of σ. Equation (I) can be written for a TPN

$$Q(\tau) = Q(\tau_0) + CR(\tau). \tag{II}$$

Let us suppose now, that $\tau \neq \tau_0$ and put $\Delta\tau = \tau - \tau_0$. We have from (II)

$$\frac{\Delta Q(\tau)}{\Delta\tau} = \frac{Q(\tau) - Q(\tau_0)}{\Delta\tau} = C\frac{R(\tau)}{\Delta\tau} = CI(\tau) \Rightarrow \frac{\Delta Q(\tau)}{\Delta\tau} = CI(\tau) \tag{III}$$

where

$\frac{Q(\tau)}{\Delta\tau}$ is a vector representing the mean variation of the charge of the net in the interval $\Delta\tau$,

the k-th entry of the vector I, $i_k = \frac{r_k(\tau)}{\Delta\tau}$ represents the mean frequence of firing of the transition t_k during $\Delta\tau$.

The vector $I(\tau)$ will be called *current vector*, and evidently $\forall \tau_j \in T$, $I(\tau_j) > 0$.

II. 2. Description of the behavior for constant currents

II. 2.1 — **General case**. We are interested in the cases of functioning with constant currents for which the total charge of the net remains bounded. This amounts to searching for solutions of the equation

$$CI = 0 \quad (I > 0). \tag{IV}$$

Those solutions correspond to cyclic firing sequences in the net as it is shown in [6]. We give additional relations that the current vector I must satisfy in terms of the initial marking and of the delays associated to the places.

Definition 12. Let C be a matrix of order $n \times m$ on Q. We denote by $\mathscr{C}$ ($\mathscr{C}^t$) the set of non negative solutions of $CX = 0$ ($C^t X = 0$). A *generator* of $\mathscr{C}(\mathscr{C}^t)$ is a set of vectors $\{X_j\}_{j=1}^s$ with $X_j \in \mathbf{N}^m$ ($X_j \in \mathbf{N}^n$) such that any element X_0 of $\mathscr{C}(\mathscr{C}^t)$ could be expressed as the linear combination of elements of $\{X_j\}_{j=1}^s$ with non-negative rational coefficients, i.e., $X_0 = \sum_{j=1} \lambda_j X_j$, where λ_j are non-negative rational numbers.

If we assign constant currents to the transitions of a bounded TPN, we have a periodic functioning, and let $Q(\tau_{k_0})$, $Q(\tau_{k_1})$, ..., $Q(\tau_{k_s})$ be the successive markings of the net during a period. Then, the mean value $\bar{Q}$ of the charge variable $Q(\tau)$ is given by

$$\bar{Q} = \frac{Q(\tau_{k_0}) + Q(\tau_{k_1}) + Q(\tau_{k_2}) + \ldots + Q(\tau_{k_s})}{s+1}.$$

If we multiply this last equation by $J_0^t (J_0^t \in \mathscr{C}^t)$, we obtain

$$J_0^t \bar{Q} = J_0^t Q(\tau_0). \tag{Va}$$

But the mean value $\bar{q}_w$ of the charge of a place p_w satisfies the inequality

$$\bar{q}_w \geqq z_w C_w^+ I$$

where C_w^+ is the w-th line of the matrix C^+, and the product $C_w^+ I$ represents the mean frequence of the arrivals of markers at the place p_w.

Let Z be the following square matrix of order n:

$$Z = \begin{bmatrix} z_1 & 0 & 0 & 0 & 0 \dots 0 \\ 0 & z_2 & 0 & 0 & 0 \dots 0 \\ 0 & 0 & z_3 & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 \dots z_n \end{bmatrix}$$

Then, the set of the inequalities $\{\bar{q}_w \geqq z_w C_w^+ I\}_{w=1}^n$ can be written in the form

$$\bar{Q} \geqq ZC^+ I = ZC^- I. \tag{Vb}$$

Let J_0^t be a positive solution of $J^t C = 0$. From (Va) and (Vb) one can obtain

$$J_0^t Q(\tau_0) \geqq J_0^t ZC^+ I = J_0^t ZC^- I. \tag{Vc}$$

This last inequality establishies a relation between the initial marking, the current vector and the delays associated to the places of a TPN.

Let $\mathscr{J} = \{J_1^t, J_2^t, \dots, J_k^t\}$ be a generator of $\mathscr{C}^t$. If $J_0^t \in \mathscr{C}^t$ then any inequality $J_0^t Q(\tau_0) \geqq J_0^t ZC^+ I$ can be expressed as a linear combination of the set of inequalities $\{J_s^t Q(\tau_0) \geqq J_s^t ZC^+ I\}_{s=1}^k$.

The relations

$$CI = 0 \quad (I > 0) \tag{IV}$$

$$\{J_s^t Q(\tau_0) \geqq J_s^t ZC^+ I\}_{s=1}^k \tag{V}$$

describe the functioning of a timed Petri net for constant currents.

II. 2.2 — **Functioning of TPN at its natural rate.**

Definition 13. Given a TPN by its incideme matrix C and its delay matrix Z. We say that it functions at its *natural rate* for a given current vector I_0 iff I_0 satisfies the equations $CI = 0$ (IV) and $\{J_s^t Q(\tau_0) = J_s^t ZC^+ I\}_{s=1}^k$ (VI), where $\{J_s^t\}_{s=1}^k$ is a generator of $\mathscr{C}^t$.

Proposition 1. There exist at most n linearly independent equations describing the functioning at natural rate of a TPN with n places.

Proof. Suppose that the rank of C is equal to ϱ. Then (IV) contains at most ϱ linearly independent equations. Also, the dimension of the space of the solutions of $J^t C = 0$ is $n - \varrho$. Thus (VI) has at most $n - \varrho$ linearly independent equations, and consequently there exist at most n linearly independent equations in the system (IV) and (VI).

Example 1. Consider the TPN of Figure 1. We want to calculate the current vectors (if there exist any) corresponding to functionings at natural rate. $Q(\tau_0)$ and Z are supposed given.

$$C = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ -3 & 1 & 0 & 0 \\ 3 & -1 & 0 & 0 \end{bmatrix}$$

Solution of $CI=0$: we find $i_2=i_4=3i_1$, $i_1=i_3$

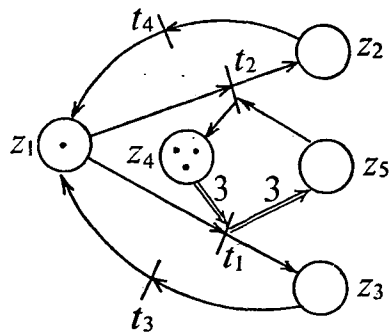


Figure 1

A generator of $\mathscr{C}^t$ is $\{J_1^t=[1\ 1\ 1\ 0\ 0],\ J_2^t=[0\ 0\ 0\ 1\ 1]\}$. Thus

$$J_1^t ZC^+ I = J_1^t Q_0 \Rightarrow q_{0_1}+q_{0_2}+q_{0_3} = z_1 i_3 + z_1 i_4 + z_2 i_2 + z_3 i_1$$

and

$$J_2^t ZC^+ I = J_2^t Q_0 \Rightarrow q_{0_4}+q_{0_5} = i_2 z_4 + 3 i_1 z_5$$

The condition for the existence of a solution is

$$\frac{q_{0_1}+q_{0_2}+q_{0_3}}{4z_1+3z_2+z_3} = \frac{q_{0_4}+q_{0_5}}{3(z_4+z_5)} \quad (\alpha).$$

In this case

$$i_1=\frac{q_{0_4}+q_{0_5}}{3(z_4+z_5)},\quad i_2=\frac{q_{0_4}+q_{0_5}}{z_4+z_5}.$$

Suppose now that we have $z_1=z_2=z_3=z_4=z_5=1$ and $Q_0^t=[1\ 0\ 0\ 3\ 0]$. Then the equation (α) is not verified and there is no functioning at natural rate. The inequalities (V) give

$$q_{0_1}+q_{0_2}+q_{0_3} \geqq z_1(i_3+i_4)+z_2 i_2+z_3 i_1 \Rightarrow 1 \geqq 8i_1,$$

$$q_{0_4}+q_{0_5} \geqq i_2 z_4+3i_1 z_5 \Rightarrow 3 \geqq 6i_1$$

yielding

$$i_{1\max}=\min\left\{\frac{1}{8},\frac{1}{2}\right\}=\frac{1}{8} \text{ and } i_{2\max}=\frac{3}{8}.$$

III. Solution of $CI=0$ and $J^tC=0$ decomposition

In this section we present some results relative to the properties of non-negative solutions of $CI=0$ and $J^tC=0$, where C is the incidence matrix of a PN. Many authors have used linear equations for the study of the properties of PN's ([6], [7], [9], [10], [11], [12] and [14]). In particular, a part of the results on the decomposition of PN's exposed in this section have been developped independently by Memmi [10], Crespi—Reghizzi and Mandrioli [9] and the author [14]. Also, similar results, in a less restrained context, are well known since several years (see, for example [13]). Our contribution consists in making evident the ralations between the structure of the net (decomposability into consistent and invariant components) and the solutions of $CI=0$ and $J^tC=0$. We borrowed the terms "consistent" and "invariant" from [6] and [7], respectively, and the term "support" from FulKerson [13], although it is used in a slightly different sense. This study is limited to pure and strongly connected PN's. Pureness is imposed by the fact that we use the incidence matrix for representing PN's and strong connexity by the fact that it is a necessary condition for a net to be bounded [6]. In what follows, the term "PN" denotes a strongly connected and pure PN.

Definition 14. For a PN $\mathcal{N}=(P, T, \alpha, \beta)$ a *subnet* of $\mathcal{N}$ is a PN $\mathcal{N}_1=(P_1, T_1, \alpha_1, \beta_1)$ such that $P_1 \subset P$, $T_1 \subset T$, moreover α_1: $P_1 \times T_1 \to \mathbf{N}$ and β_1: $P_1 \times T_1 \to \mathbf{N}$ are restrictions of α and β, respectively.

Definition 15. The *union* of two subnets $\mathcal{N}_1=(P_1, T_1, \alpha_1, \beta_1)$ and $\mathcal{N}_2=(P_2, T_2, \alpha_2, \beta_2)$ of a PN $\mathcal{N}=(P, T, \alpha, \beta)$ is a subnet $\mathcal{N}_3=(P_3, T_3, \alpha_3, \beta_3)$ of $\mathcal{N}$ with $P_3=P_1 \cup P_2$ and $T_3=T_1 \cup T_2$.

Definition 16. Let $\mathcal{N}=(P, T, \alpha, \beta)$ be a PN and $\mathcal{S}=\{\mathcal{N}_i=(P_i, T_i, \alpha_i, \beta_i)\}_{i=1}^{k}$ a set of subnets of $\mathcal{N}$. $\mathcal{N}$ is *covered* by S or S is a *decomposition* of $\mathcal{N}$ if

$$P = \bigcup_{i=1}^{k} P_i \quad \text{and} \quad T = \bigcup_{i=1}^{k} T_i.$$

III. 1. Non-negative solutions of $CI=0$. Decomposition into consistent components

Definition 17. Let $\mathcal{N}=(P, T, \alpha, \beta)$ be a PN. Then a set $T_1 \subset T$ defines a *t-complete subnet* $\mathcal{N}_1=(P_1, T_1, \alpha_1, \beta_1)$ of $\mathcal{N}$ if $P_1={}^\bullet T_1=T_1^\bullet$.

Proposition 2. Let C be the incidence matrix of a PN and $I_0 \in \mathscr{C}$. Then the set $F_1=\{t_j | i_{0_j} \neq 0\}$ defines a t-complete subnet of the net having C as incidence matrix.

Proof. Consider the subnet with $T_1=\{t_j | i_{0_j} \neq 0\}$ and $P_1={}^\bullet T_1 \cup T_1^\bullet$. Then each place p of P_1 has at least one input transition or one output transition (by construction of the set P_1). Suppose that a place p_w ($p_w \in P_1$) has the input transitions $t_{i_1}, t_{i_2}, \ldots, t_{i_r}$ but no output transition in the subnet defined by P_1 and T_1. Then we have $\sum_j i_{0_{ij}} \beta(p_w, t_{ij})=0$ where $i_{0_{ij}}$ and $\beta(p_w, t_{ij})$ are positive rational numbers which is absurde. Thus p_w must have an output transition belonging to T_1. In the same

manner one can prove that if a place p_w has an output transition belonging to T_1 then it has an input transition belonging to T_1.

Definition 18. Let C be the incidence matrix of a PN $\mathcal{N}$. A *consistent component* of $\mathcal{N}$ is any t-complete subnet $\mathcal{N}_1$ defined by the set of transitions corresponding to the positive entries of a vector I_1 ($I_1 \in \mathscr{C}$). $\mathcal{N}_1$ is the *support* of I_1, (we note $\mathcal{N}_1 = S(I_1)$). If there exists I_0 such that $S(I_0) = \mathcal{N}$ then we say that $\mathcal{N}$ is *consistent*.

Definition 19. Let a PN be given with an initial marking M. A firing sequence σ is a *cyclic* firing sequence from M iff $M \xrightarrow{\sigma} M$.

Proposition 3 [6]. A PN having a live and bounded marking is concistent.

Proposition 4. [6]. Let $\mathcal{N}_1 = (P_1, T_1, \alpha_1, \beta_1)$ be a consistent component of a net $\mathcal{N}$. Then $\mathcal{N}_1$ has a marking M from which there exists a cyclic firing sequence $\sigma = t_{k_1} t_{k_2} \dots t_{k_s}$ such that $\bigcup_{j=1}^{s} t_{k_j} = T_1$. Inversely, each cyclic firing sequence $\sigma = t_{k_1} t_{k_2} \dots t_{k_s}$ in $\mathcal{N}$, defines a consistent component of $\mathcal{N}$ having $T_1 = \bigcup_{j=1}^{s} t_{k_j}$ as set of transitions.

Proposition 5. The union of two consistent components of a net is a consistent component.

Proof. Let I_1 and I_2 be two elements of $\mathscr{C}$ defining two consistent components $S(I_1)$ and $S(I_2)$. Then $I_1 + I_2 \in \mathscr{C}$ defines the consistent component $S(I_1) \cap S(I_2)$.

Definition 20. Let $I_1 \in \mathscr{C}$ where C is the incidence matrix of a PN $\mathcal{N}$. Then $S(I_1)$ is an *elementary* consistent component of $\mathcal{N}$ iff there exists no I_2 ($I_2 \neq 0$, $I_2 \in \mathscr{C}$) such that $S(I_2) \subset S(I_1)$. A vector I_1 defining an elementary consistent component $S(I_1)$ is called *elementary vector* of $\mathscr{C}$.

Proposition 6. If C is the incidence matrix of a PN and I_1 and I_2 are two elementary vectors of $\mathscr{C}$ such that $S(I_1) = S(I_2)$ then I_1 and I_2 are linearly dependent.

Proof. Suppose that T_1 and I_2 are linearly independent and $S(I_1) = S(I_2)$. Let $\lambda = \min_{i_{2_j} \neq 0} \left\{ \frac{i_{1_j}}{i_{2_j}} \right\}$ and $I_3' = I_1 - \lambda I_2$. Then $I_3' \neq 0$ and there exists a scalar μ such that $I_3 = \mu I_3' \in \mathscr{C}$. We have $S(I_3) \subset S(I_1)$. Thus $S(I_1)$ is not elementary.

Proposition 7. Every consistent PN $\mathcal{N}$ can be decomposed into a set of elementary consistent components.

Proof. Let $I_0 \in \mathscr{C}$, $S(I_0) = \mathcal{N}$ and suppose that $\mathcal{N}$ is not elementary. Then, there exists a consistent component $\mathcal{N}_1 (\mathcal{N}_1 \subset \mathcal{N})$ and I_1 such that $S(I_1) = \mathcal{N}_1$. Let $\lambda = \min_{i_{1_j} \neq 0} \left\{ \frac{i_{0_j}}{i_{1_j}} \right\}$ and $I_2' = I_0 - \lambda I_1$. Then it is easy to verify that there exists a scalar μ such that $I_2 = \mu I_2' \in \mathscr{C}$ and if $\mathcal{N}_2 = S(I_2)$ then $\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2$.

Corollary 1. The set of elementary vectors of $\mathscr{C}$ is a generator.

Definition 21. Let a PN $\mathcal{N}$ with incidence matrix C be given and S a set of elementary vectors of $\mathscr{C}$. Then S is a *t-base* of $\mathcal{N}$ iff S is a generator (of $\mathscr{C}$) of minimal cardinality.

Proposition 8. Let $B=[I_1, I_2, \ldots, I_s]$ be a matrix of order $m \times s$ such that $\{I_j\}_{j=1}^{s}$ is a t-base of a PN. Then the rank of B is less than or equal to $m-\varrho$ where ϱ is the rank of the incidence matrix of the net. Furthermore, if the net is consistent then the rank of B is equal to $m-\varrho$.

Proof. The fact that rank $[B] \leqq m-\varrho$ is obvious because the space of the solutions of $CI=0$ is of dimension $m-\varrho$. In order to prove that rank $[B]=m-\varrho$, in the case of a consistent net, it is sufficient to prove that any solution I_0 of $CI=0$ can be expressed as the linear combination of $I_1, I_2, \ldots, I_s$ (columns of B). If $I_0>0$ then this is always possible according to corollary 1. If not, one can obtain from I_0, a vector $\tilde{I}$ ($\tilde{I}>0$) such that $\tilde{I}=\sum_{j=1}^{s} \beta_j I_j + I_0$ where the β_j's are non-negative rational numbers. But $C\tilde{I}=0$ and $\tilde{I}$ defines a consistent component. Thus, according to the Corollary 1, we can write $\tilde{I}=\sum_{j=1}^{s} \gamma_j I_j$. This gives $I_0=\sum_{j=1}^{s} (\gamma_j-\beta_j) I_j$.

REMARK. $CB=0$. For any current vector $I \in \mathscr{C}$, $I=BI_b$, the k-th entry of I_b being the "loop current" associated to the elementary component corresponding to the k-th column of B.

III. 2. Non-negative solutions of $J'C=0$. Decomposition into invariant components

The following definitions and propositions are dual of those in III. 1.

Definition 22. Let $\mathscr{N}=(P, T, \alpha, \beta)$ be a PN. Then a set P_1 ($P_1 \subset P$) defines a *p-complete* subnet $\mathscr{N}_1=(P_1, T_1, \alpha_1, \beta_1)$ of $\mathscr{N}$ if $T_1={}^{\cdot}P_1=P_1^{\cdot}$.

Proposition 9. Let C be the incidence matrix of a PN and $J_0^t \in \mathscr{C}^t$. Then the set $P_1=\{p_i | j_{0_i} \neq 0\}$ defines a p-complete subnet of the net having C as incidence matrix.

Definition 23. Let C be the incidence matrix of a PN $\mathscr{N}$. An *invariant component* pf $\mathscr{N}$ is any p-complete subnet $\mathscr{N}_1$ defined by the set of places corresponding to the oositive entries of a vector J_1^t ($J_1^t \in \mathscr{C}^t$). $\mathscr{N}_1$ is the *support* of J_1^t, (we note $\mathscr{N}_1=S(J_1^t)$). If there exists J_0^t such that $S(J_0^t)=\mathscr{N}$ then we say that $\mathscr{N}$ is *invariant*.

Proposition 10. The union of two invariant components is an invariant component.

Definition 24. Let $J_1^t \in \mathscr{C}^t$ where C is the incidence matrix of a PN $\mathscr{N}$. Then $S(J_1^t)$ is an *elementary* invariant component of $\mathscr{N}$ iff there exists no J_2^t ($J_2^t \neq 0$, $J_2^t \in \mathscr{C}^t$) such that $S(J_2^t) \subset S(J_1^t)$. A vector J_1^t defining an elementary invariant component $S(J_1^t)$ is called *elementary vector* of $\mathscr{C}^t$.

Proposition 11. If C is the incidence matrix of a PN and J_1^t and J_2^t are two elementary vectors of $\mathscr{C}^t$ with $S(J_1^t)=S(J_2^t)$ then J_1^t and J_2^t are linearly dependent.

Proposition 12. Every invariant PN $\mathscr{N}$ can be decomposed into a set of elementary invariant components.

Corollary 2. The set of elementary vectors of $\mathscr{C}^t$ is a generator (of $\mathscr{C}^t$).

Definition 25. Let $\mathcal{N}$ be a PN with incidence matrix C and S a set of elementary vectors of $\mathscr{C}^t$. Then S is a *p-base* of $\mathcal{N}$ iff S is a generator (of $\mathscr{C}^t$) of minimal cardinality.

Proposition 13. Let $D^t=[J_1, J_2, \ldots, J_s]$ be a matrix of order $n\times s$ such that $\{J_i^t\}_{i=1}^s$ is a p-base of a PN. Then the rank of D is less than or equal to $n-\varrho$ where ϱ is the rank of the incidence matrix of the net. Furthermore, if the net is invariant then rank $[D]=n-\varrho$.

III. 3. Particular cases: state graphs and marked graphs

Proposition 14. Let C be the incidence matrix of a pure and strongly connected state graph with n places and m transitions. Then the following statements are well known:

a) rank $[C]=n-1$,

b) the space of solutions of $CI=0$ is of dimension $m-n+1$,

c) the space of solutions of $J^tC=0$ is of dimension 1 and the vector $J_0^t=[1\ 1\ 1\ldots1]$ is a base of this space.

REMARKS. A t-base for a state graph is a circuit base,
C in Proposition 14 expresses the fact that any strongly connected state graph is an elementary invariant component.

Proposition 15. Let C be the incidence matrix of a pure strongly connected marked graph with n-places and m transitions. Then we have the dual of Proposition 14:

a) rank $[C]=m-1$,

b) the space of the solutions of $CI=0$ is of dimension 1 and the vector $I_0^t=[1\ 1\ 1\ \ldots\ 1]$ is a base of this space,

c) the space of solutions of $J^tC=0$ is of dimension $n-m+1$.

REMARKS. A t-base for a marked graph is a circuit base,
b) in Proposition 15 expresses the fact that any strongly connected marked graph is an elementary consistent component.

IV. Resolution of the equations (IV) and (VI) with given $Q(\tau_0)$ and Z

In this section we show that the problem of determining the currents of a TPN for functioning at natural rate, when we know $Q(\tau_0)$ and Z may have either several solutions or no solution at all. The extreme cases correspond to state graphs and marked graphs.

Example 2. For the TPN of Figure 2 the system of the equations (IV) and (VI) generally has a unique solution for I. We have:

$$B^t=\begin{bmatrix}2&1&1&0&1&1\\2&1&1&1&2&0\end{bmatrix} \text{ and } D=\begin{bmatrix}1&1&1&1&0&0\\0&0&0&0&1&1\end{bmatrix}.$$

If i_x and i_y are the currents associated respectively to the two elementary consistent components of the net, we have, $I=B\begin{vmatrix} i_x \\ i_y \end{vmatrix}$.

On the other hand, we have two equations expressing the charge conservation in the state graphs defined by the lines of D:

$$(i_x+i_y)2z_0=1 \quad \text{(we put } z_0=z_5=z_6\text{)}$$

and

$$2(i_x+i_y)z_1+2(i_x+i_y)z_2+(i_x+i_y)z_3+(2i_y+i_x)z_4=1.$$

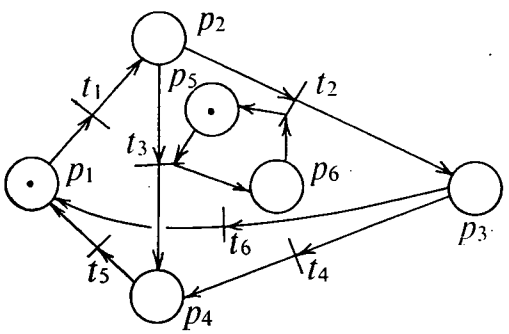


Figure 2

By resolving this system, we obtain

$$i_x=\frac{2(z_1+z_2+z_4)+z_3-2z_0}{2z_0z_4}, \quad i_y=\frac{2z_0-2(z_1+z_2)-z_3-z_4}{2z_0z_4}$$

where i_x and i_y must satisfy the inequalities $i_y>0$ and $i_x+i_y>0$. The second inequality is always verified, and the first gives the condition

$$z_0>\frac{2(z_1+z_2)+z_3+z_4}{2}.$$

TIMED MARKED GRAPHS. In this case we have $n\geqq m$ ($n=|P|$, $m=|T|$, the equality is verified only if the marked graph is a circuit). Thus, the currents determined by solving m equations among the n equations (IV) and (VI) must satisfy the remaining $n-m$ equations in order to have a functioning at natural rate. If not, it is sufficient to search for the solutions of

$$\{J_r^tQ_0 \geqq J_r^tZC^+I\}_{r=1}^{n-m+1}$$

where $\{J_r^t\}_{r=1}^{n-m+1}$ is a p-base (base of circuits in this case) and $I^t=[ii\ldots i]$ a solution of $CI=0$.

The r-th inequality can be written in the form $\sum_{k_r} q_{0_i} \geqq (\sum_{k_r} z_i) i$, where $\sum_{k_r} q_{0_i}$ is the sum of the markers in the circuit K_r, $K_r = S(J_r^!)$ and $\sum_{k_r} z_i$ is the sum of the delays associated to the places of this circuit. Therefore,

$$i_{\max} = \min_{r=1}^{n-m+1} \{(\sum_{k_r} q_{0_i})/(\sum_{k_r} z_i)\}.$$

This result is given in [6].

TIMED STATE GRAPHS. In this case $m \geqq n$, and for I it is always possible to solve the system (IV) and (VI). One can construct a system having a unique solution for I by giving additional equations imposing a constant ratio between the currents of the transitions having the same input place. There exist exactly $m-n$ linearly independent equations of this kind for any state graph.

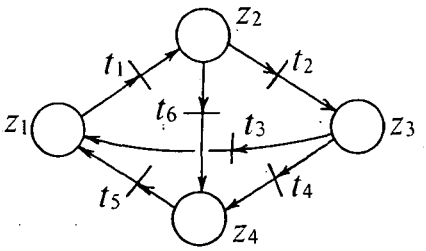


Figure 3

Example 3. Consider the timed state graph of Figure 3. The solution of $CI=0$ gives

$$i_1 = i_3+i_4+i_6,$$
$$i_2 = i_3+i_4,$$
$$i_5 = i_6+i_4.$$

The equation of conservation of the charge in the graph is

$$(i_6+i_3+i_4)(z_1+z_2)+(i_3+i_4)z_3+(i_4+i_6)z_4 = \sum_{j=1}^{4} q_{0_j}.$$

If we impose $\frac{i_2}{i_6}=\lambda_1$ and $\frac{i_3}{i_4}=\lambda_2$ we have

$$i_2 = \frac{\lambda_1 i_1}{\lambda_1+1}, \quad i_3 = \frac{\lambda_1\lambda_2 i_1}{(1+\lambda_1)(1+\lambda_2)}, \quad i_4 = \frac{\lambda_1 i_1}{(1+\lambda_1)(1+\lambda_2)},$$

$$i_5 = \frac{(1+\lambda_1+\lambda_2) i_1}{(1+\lambda_1)(1+\lambda_2)}, \quad i_6 = \frac{i_1}{1+\lambda_1}.$$

We can now uniquely determine the currents in terms of Q_0, Z and parameters λ_1 and λ_2. For example, for i_1 we obtain

$$i_1\left[(z_1+z_2)+\frac{\lambda_1}{1+\lambda_1} z_3+\frac{1+\lambda_1+\lambda_2}{(1+\lambda_1)(1+\lambda_2)} z_4\right] = \sum_{j=1}^{4} q_{0_i}.$$

REMARK. One can construct a system having a unique solution for I, from the system of the equations (IV) and (VI) by imposing the additional constraint that the sum of the charge of each circuit of a base of circuits of the state graph is constant. In this case we have $(n-1)$ linearly independent equations from the system $CI=0$ and $m-n+1$ linearly independent equations by application of this constraint. Thus

we have m equations describing the behaviour of the net. (The equation [1 1 1 ... 1] $Q_0=[1\ 1\ 1\ \ldots\ 1]\ ZC^+I$ can be obtained as the linear combination of $n-m+1$ equations). The analogy with the electrical circuits is obvious. The $m-n+1$ equations express the application of the Kirchhoff's voltage law: the sum of $i_j \cdot z_j$ (voltage drops) for a circuit is equal to its total charge (electromotive force).

V. Applications

Example 4. *Producer-consumer system.* Consider the producer-consumer problem with a buffer of bounded capacity N_0. We suppose that the producer and the consumer do not try to access the buffer at the same time. The producer deposits items in the buffer as long as it is not full and the consumer does not try to take an item from the buffer when it is empty. Items are produced, deposited, taken and consumed one by one.

The TPN of Figure 4 describes the system producer-consumer with a possible initial marking. Interpretation of the delays associated to the places:

z_p means time of producing an item,
z_d means time of depositing an item,
z_t means time of taking an item,
z_c means time of consuming an item.

We suppose that the z_i' s associated to the other places are equal to zero. That is, the producer and the consumer are functioning at maximum speed: the producer is allowed to deposit an item right after having produced one and he always finds the access to the buffer free. Also, the consumer is allowed to take an item right after having consumed one and he always finds the access to the buffer free.

By solving the equation $CI=0$ we find that the same current i must be associated to all the transitions. Also, a cover by elementary invariant components (state graphs in this case) is given in Figure 5.

PROBLEM. Considering as initial marking the marking given in Figure 4, find conditions for functioning at natural rate.

The inequality (V) applied for SG1, SG2, SG3, SG4 gives, respectively,

$$i \leqq \frac{1}{z_p+z_d}, \quad i \leqq \frac{1}{z_d+z_t+z_s}, \quad i \leqq \frac{1}{z_c+z_t}, \quad i \leqq \frac{N_0}{z_d+z_t+z_a}$$

which yield: $i_{\max}=\min\left\{\frac{1}{z_p+z_d}, \frac{1}{z_d+z_t+z_s}, \frac{1}{z_c+z_t}, \frac{N_0}{z_d+z_t+z_a}\right\}$.

Conditions for functioning at natural rate are

$$z_s = z_p-z_t=z_c-z_d > 0 \quad \text{and} \quad N_0-1 = \frac{z_a-z_s}{z_p+z_d} = \frac{z_a-z_s}{z_c+z_t}$$

CONCLUSION. The producer's and consumer's periods must be equal: $z=z_p+z_d=$ $=z_c+z_t$. Also, z_s, the mean time between two successive accesses, is given by $z_s=$ $=z_p-z_t=z_c-z_d>0$. From $N_0-1=\frac{z_a-z_s}{z}$ we deduce that:

a) for $z_a < z_s$, a functioning at natural rate is impossible,
b) if $z_a = z_s$, a minimum capacity $N_0 = 1$ is necessary,
c) if $z_a > z_s$, a minimum capacity of $N_0 = 1 + \frac{z_a - z_s}{z}$ is necessary.

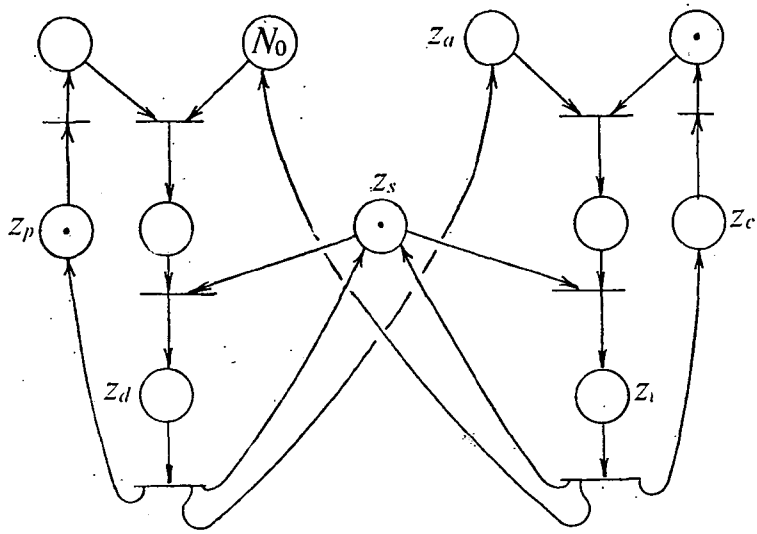


Figure 4

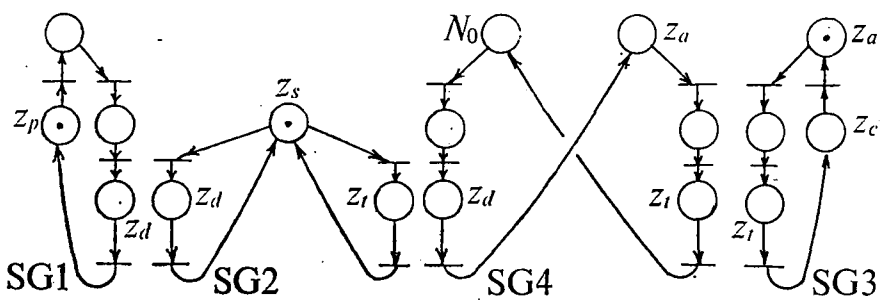


Figure 5

Example 5. *System of r producers and w consumers.* Let a system of r producers and w consumers be connected with a buffer of capacity N_0. The simultaneous access to the buffer is not allowed. We consider for the delays associated to the places the same notations as in the preceding example by adding an index in order to distinquish the producers and the consumers among them. Thus, z_{d_i} is the time for the deposit of an item by the i-th producer and z_{c_j} is the time of consuming an item by the j-th consumer. We consider the case in which procuders and consumers are functioning at maximum speed, which implies zero waiting times before the deposit or before taking an item (Figure 6).

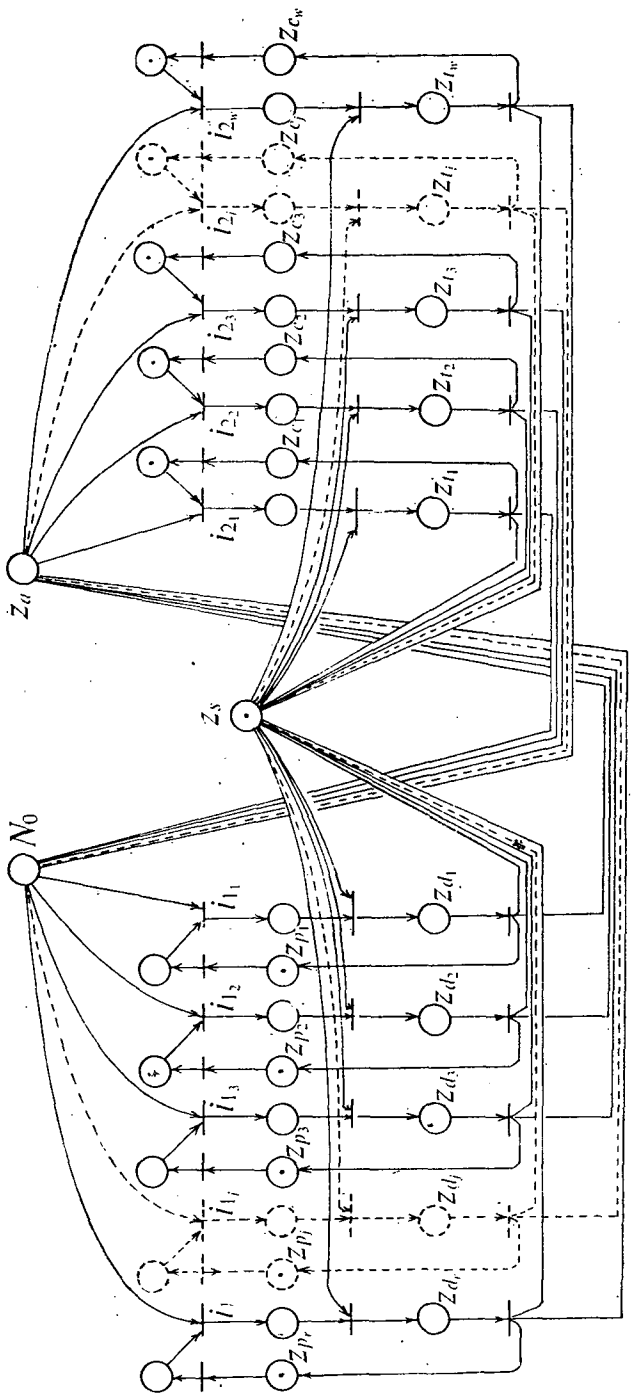

z_{c_w}
z_{l_w}
i_{2_w}
z_{c_j}
z_{l_j}
i_{2_j}
z_{c_3}
z_{l_3}
i_{2_3}
z_{c_2}
z_{l_2}
i_{2_2}
z_{c_1}
z_{l_1}
i_{2_1}
z_w
z_s
N_0
i_{1_1}
$z_{p_{11}}$
z_{d_1}
i_{1_2}
$z_{p_{12}}$
z_{d_2}
i_{1_3}
$z_{p_{13}}$
z_{d_3}
i_{1_j}
$z_{p_{1j}}$
z_{d_j}
i_{1_r}
$z_{p_{1r}}$
z_{d_r}

Figure 6

In Figure 7, we give a decomposition of the PN representing the system into elementary invariant components.

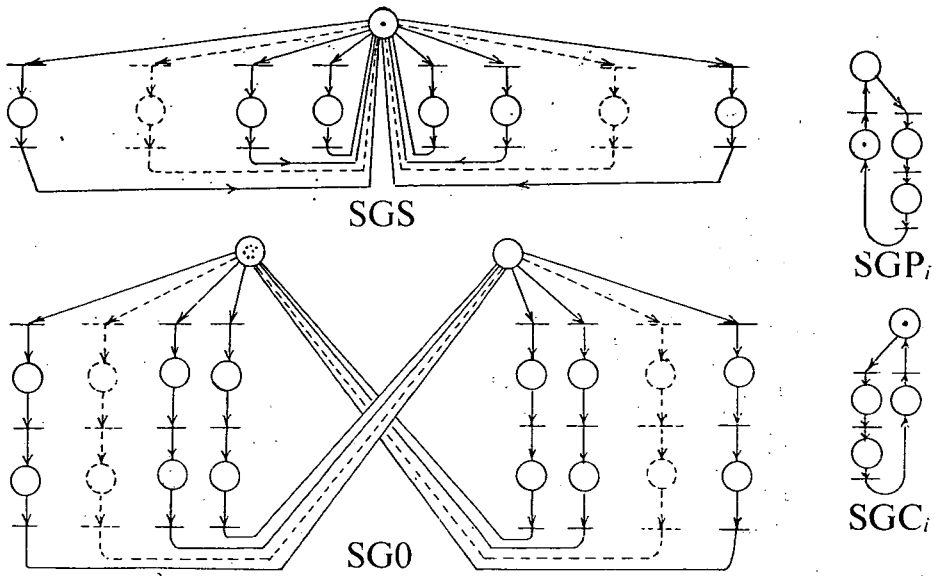


Figure 7

If i_{0_j} and i_{0_j} are the currents associated to the cycles of the j-th producer and j-th consumer, respectively, we have:

$$\left\{i_{1_j} \leqq \frac{1}{z_{d_j}+z_{p_j}}\right\}_{j=1}^{r} \quad \text{and} \quad \left\{i_{2_j} \leqq \frac{1}{z_{c_j}+z_{t_j}}\right\}_{j=1}^{w}.$$

Furthermore, $\sum_{j=1}^{r} i_{1_j} = \sum_{j=1}^{w} i_{2_j} = i_0$ where i_0 is the current throughout the buffer. Thus, $i_0 = \min\left\{\sum_{j=1}^{r} \frac{1}{z_{d_j}+z_{p_j}}, \sum_{j=1}^{w} \frac{1}{z_{c_j}+z_{t_j}}\right\}$.

The equation of conservation of the charge for SGS is:

$$\sum_{j=1}^{r} i_{1_j} z_{d_j} + \sum_{j=1}^{w} i_{2_j} z_{t_j} + i_0 z_s = 1 \Rightarrow z_s = \frac{1 - \sum_{j=1}^{r} i_{1_j} z_{d_j} - \sum_{j=1}^{w} i_{2_j} z_{t_j}}{i_0}. \tag{a}$$

But,

$$\sum_{j=1}^{r} i_{1_j} z_{d_j} \leqq \sum_{j=1}^{r} \frac{z_{d_j}}{z_{d_j}+z_{p_j}} \quad \text{and} \quad \sum_{j=1}^{w} i_{2_j} z_{t_j} \leqq \sum_{j=1}^{w} \frac{z_{t_j}}{z_{c_j}+z_{t_j}}.$$

From the two preceding inequalities and (a) we get

$$z_s \geqq \frac{1}{i_0}\left(1 - \Sigma \frac{z_{d_j}}{z_{d_j}+z_{p_j}} - \Sigma \frac{z_{t_j}}{z_{t_j}+z_{c_j}}\right). \tag{b}$$

Finally, for SGO (Figure 7) we have

$$\sum_{j=1}^{r} i_{1_j} z_{d_j} + \sum_{j=1}^{w} i_{2_j} z_{t_j} + z_a i_0 = N_0 \Rightarrow 1 - i_0 z_s + i_0 z_a = N_0 \Rightarrow N_0 - 1 = (z_a - z_s) i_0.$$

From this last equation and the inequality (b) we obtain

$$z_a \geqq \frac{1}{i_0}\left(N_0 - \sum_{j=1}^{r} \frac{z_{d_j}}{z_{d_j} + z_{p_j}} - \sum_{j=1}^{w} \frac{z_{t_j}}{z_{t_j} + z_{c_j}}\right). \qquad \text{(c)}$$

The inequalities (b) and (c) give least bounds for the mean time between two successive accesses to the buffer (z_s) and for the mean waiting time (z_a) of an item in the buffer.

Example 6. Consider the TPN of Figure 8. One could imagine that it represents the functioning of an enterprise of car location having customers of two types. Customers of type 1, whose number is N_1, have a mean location time z_1 and a mean time between two succesive demands for location z_{a_1}. Also, customers of type 2, whose number is N_2, have a mean location time z_2 and a mean time between two successive demands for location z_{a_2}. We suppose that the total number of cars of the enterprise is N_0 and that after location, a service of mean duration z_s is done to each car. We finally admit that a car ready for location waits during z_0 before a customer demands it.

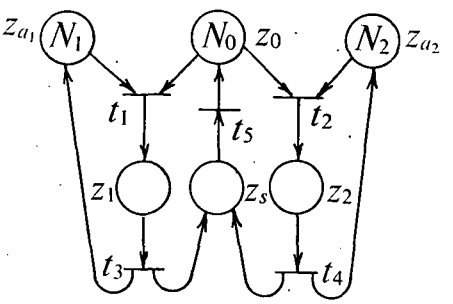


Figure 8

By solving $CI=0$, we get $i_1=i_3$, $i_2=i_4$, $i_5=i_1+i_2$. Furthermore, the resolution of $J'C=0$ gives a decomposition into state graphs (Figure 9).

PROBLEM. Knowing N_1 and N_2 as well as the delays associated to the places, determine N_0 such that a functioning at natural rate will be possible.

The equations of charge conservation for SG1 and SG2 are, respectively,

$$i_1 = \frac{N_1}{z_1 + z_{a_1}} \quad \text{and} \quad i_2 = \frac{N_2}{z_2 + z_{a_2}}.$$

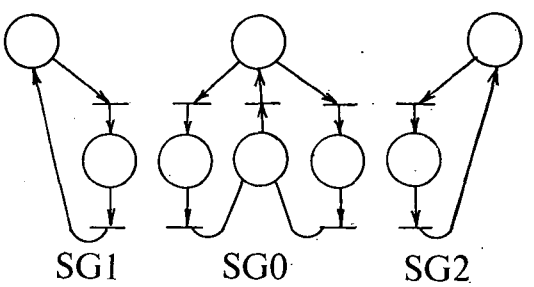


Figure 9

For SG0, we have

$$N_0 = (i_1+i_2)(z_0+z_s)+i_1z_1+i_2z_2 \Rightarrow N_0 = \frac{N_1(z_0+z_1+z_s)}{z_{a_1}+z_1}+N_2\frac{(z_0+z_s+z_2)}{z_{a_2}+z_2}.$$

N_0 is the minimum number of cars to satisfy the demands of the (N_1+N_2) customers.

Abstract

We study the behavior of pure timed Petri nets for constant current assignments. It is given a set of relations describing the behavior of a timed Petri net and it is shown that its maximum computation rate can be calculated by solving a set of n linear equations where n is the number of its places. These relations are established between the currents, the initial marking and the delays of the network. Also, in order to better understand and use these relations, we give some results on the decompositions of a Petri net, obtained by studying the types of solutions of the equations $CJ=0$ and $J'C=0$ where C is the incidence matrix of the net. It is shown, that every consistent (resp. invariant) Petri net can be decomposed into a set of consistent (invariant) "elementary" subnets. We finally give some examples in order to illustrate the use of timed Petri nets in the study of the dynamic behavior of the systems.

INSTITUT DE MATHÉMATIQUES APPLIQUÉES
INFORMATIQUE
GRENOBLE, FRANCE

References

[1] PETRI, C. A., Communication with automata, Techn. Rep. n° RADC—TR—65—377 v. 1, Rome Air Develop. Center, N. Y., Jan. 1966.
[2] HOLT, A. W., Information systems theory project, Applied Data Research Incorporated Princeton, N. J., AD. 676.972, Sept. 1968.
[3] HACK, M., Analysis of production schemata by Petri nets, MIT—Project MAC, TR—94, 1972.
[4] PETRI, C. A., Concepts of net theory, Proc. of Symp. on Math. Foundations of Comp. Science, High Tatras, Sept. 1973, pp. 137—146.
[5] PETRI, C. A., Interpretations of net theory, Interner Bericht 75—07, G. M. D. Bonn, July 1975.
[6] RAMCHANDANI, C., Analysis of asynchronous concurrent systems by timed Petri nets, PhD. Thesis MIT, Sept. 1973.
[7] LAUTENBACH, K. and H. A. SCHMID, Use of Petri nets for proving correctness of concurrent process systems, IFIP 1974, North Holland Publ. Co., 1974, pp. 187—191.
[8] GHOSH, S., Some comments on timed Petri nets, Journées sur les Réseaux de Petri, March 1977, Paris, pp. 151—163.
[9] CRESPI-REGHIZZI, S. and D. MANDRIOLI, Some algebraic properties of Petri nets, Alta Frequenza, v. 45, N° 2, pp. 130—137.
[10] MEMMI, G., Semiflows and invariants. Applications in Petri nets theory, Journées sur les Réseaux de Petri, Paris. March 1977, pp. 145—150.
[11] LIEN, Y. E., Termination properties of generalized Petri nets, S.I.A.M. J. Comput. v. 5, N° 2, June 1976, pp. 251—265.
[12] MURATA, T. and R. W. CHURCH, Analysis of marked graphs and Petri nets by matrix equations, R. R. N° M.D.C. 1.1.8 Univ. of Illinois, Nov. 1975.
[13] FULKERSON, D. R., Networks, frames, blocking systems, in Mathematics for decision sciences, v. 11 of Lectures in Applied Mathematics, 1968, the American Mathematical Society, Providence, Rhode Island.
[14] SIFAKIS, J., Etude du comportement permanent des réseaux de Petri temporisés, Journées sur les Réseaux de Petri, Paris, March 1977, pp. 165—185.

(Received Dec. 27, 1977)

Strongly connected digraphs in which each edge is contained in exactly two cycles

By B. ZELINKA

In [1] A. ÁDÁM suggests a problem to characterize strongly connected digraphs without cut vertices with the property that each edge of such a graph is contained at most in two cycles. (See Problem 2, p. 189 in [1].) In this note we do not solve this problem in general, but we consider a particular case when each edge is contained exactly in two cycles. We consider finite digraphs without loops and without pairs of equally oriented edges joining the same pair of vertices.

We start by a definition.

DEFINITION. Let $A_1, A_2, \ldots, A_n$ for $n \geqq 2$ be pairwise disjoint cycles. On each A_i for $i=1, \ldots, n$ choose two distinct vertices a_i, b_i. Then identify b_i with a_{i+1} for all $i=1, \ldots, n-1$ and b_n with a_1. The class of all digraphs obtained in this way will be denoted by $\mathscr{A}$ (Fig. 1).

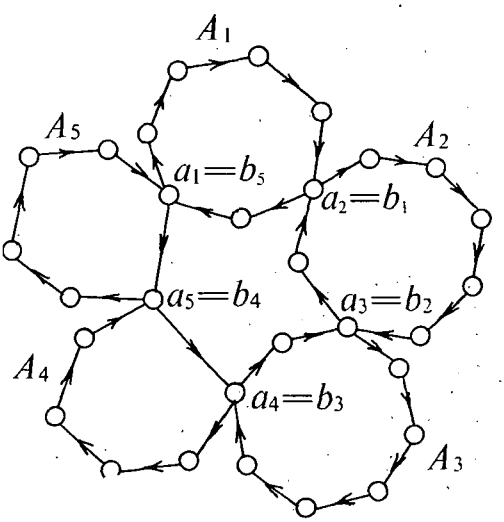


Fig. 1

Further, by a diagonal path of a cycle C we shall mean a directed path whose initial and terminal vertices are in C, while its edges and inner vertices (if any) are not.

THEOREM. *Let G be a strongly connected finite digraph without cut vertices. Then the following two assertions are equivalent:*

(i) $G \in \mathscr{A}$.

(ii) *Each edge of G is contained in exactly two cycles of G.*

PROOF: (i)⇒(ii). Let $G \in \mathscr{A}$. Let e be an edge of G. The edge e is contained in some cycle A_i for $1 \leqq i \leqq n$. The cycle A_i is the union of two directed paths $P_1^{(i)}$, $P_2^{(i)}$,

where $P_1^{(i)}$ is the path from a_i into b_i in A_i and $P_2^{(i)}$ is the path from b_i into a_i in A_i; these two paths are edge-disjoint. If e belongs to $P_1^{(i)}$ then, evidently, each cycle containing e contains the whole $P_1^{(i)}$, therefore, it must contain also a directed path from b_i into a_i in G. There are exactly two such paths; one of them is $P_2^{(i)}$, the other is the union of all $P_1^{(j)}$ for $1 \leqq j \leqq n, j \neq i$, where $P_1^{(j)}$ is defined analogously as $P_1^{(i)}$. Therefore, there are exactly two cycles in G which contain e. For the case when e is in $P_2^{(i)}$ the proof is analogous, obtained from this proof by interchanging subscripts 1 and 2.

(ii)$\Rightarrow$(i). Let G satisfy (ii). Let C_0 be a cycle in G. Let $\overrightarrow{u_1u_2}$ be an edge of C_0. As $\overrightarrow{u_1u_2}$ must be contained in two cycles, there exists a cycle C_1 containing $\overrightarrow{u_1u_2}$ and distinct from C_0. Evidently, there exists the longest directed path P_1 which contains $\overrightarrow{u_1u_2}$ and is contained in both C_0 and C_1. Let this path go from a vertex u_3 into a vertex u_4. Let P_1' be the path in C_1 from u_4 into u_3. Suppose that P_1' contains a vertex u' of C_0 distinct from u_3 and u_4; let u_1' be the first vertex of P_1' with this property. Then there exists a cycle which is the union of P_1, the subpath of P_1' from u_4 into u_1' and the path in C_0 from u_1' into u_3. This cycle is evidently distinct from both C_0 and C_1 and contains $\overrightarrow{u_1u_2}$, which is a contradiction. Thus P_1' is a diagonal path of C_0. Let u_5 be the terminal vertex of the edge of C_0 whose initial vertex is u_4. There exists a cycle C_2 distinct from C_0 and C_1 which contains the edge $\overrightarrow{u_4u_5}$. Let P_2 be the longest path which contains $\overrightarrow{u_4u_5}$ and is contained in both C_0 and C_2, let it go from a vertex u_6 into a vertex u_7. Let P_2' be the path in C_2 from u_7 into u_6; it is a diagonal path of C_0. Suppose that P_1' and P_2' have a common inner vertex; and let v be the first inner vertex of P_2' belonging to P_1'. If $u_7 \neq u_3$, then any edge belonging to the intersection of the paths in C_0 from u_6 into u_4 and from u_3 into u_7 belongs to three cycles, namely C_0, C_1 and the cycle which is the union of the path from u_3 into u_7 in C_1, the subpath of P_2' from u_7 into v and the subpath of P_1' from v into u_3, which is a contradiction. An analogous contradiction will be obtained for $u_6 \neq u_4$. Therefore P_1' and P_2' can have a common inner vertex only if $u_7 = u_3$ and $u_6 = u_4$; this case will be denoted by (*), the opposite case by (**).

Consider the case (*). Each edge of the path in C_0 from u_3 into u_4 is contained in C_0 and C_1, each edge of the path in C_0 from u_4 into u_3 is contained in C_0 and C_2. Let v_1 be the first vertex of P_1' distinct from u_4 and belonging to P_2'. The subpath of P_1' from u_4 into v_1 and the subpath of P_2' from v_1 into u_4 form a cycle D_1. Each edge of D_1 is contained in two cycles only, therefore, an inner vertex neither of the subpath of P_1' from v_1 into u_3, nor of the subpath of P_2' from u_3 into v_1 can belong to D_1. If $v_1 \neq u_3$, we repeat this consideration with the subpath of P_1' from v_1 into u_3 instead of P_1' and with the subpath of P_2' from u_3 into v_1 instead of P_2', and analogously as we have obtained v_1 and D_1 we obtain v_2 and D_2. Thus we proceed further, until we obtain $v_k = u_3$ for some k (this will be performed after a finite number of steps). The cycles $C_0, D_1, \ldots, D_k$ correspond to the cycles $A_1, A_2, \ldots, A_n$ from the definition of $\mathscr{A}$. The graph G evidently cannot contain further vertices or edges, because then (ii) would be violated. Therefore $G \in \mathscr{A}$ (Fig. 2).

Now consider the case (**). Suppose that $u_6 \neq u_4$. As C_2 must contain $\overrightarrow{u_4u_5}$, the vertex u_4 lies on the path in C_0 from u_6 into u_7. As $u_6 \neq u_4$, also the edge of C_0 whose terminal vertex is u_4 is contained in this path and in the cycle C_2. Then this edge is contained in three cycles C_0, C_1, C_2, which is a contradiction. Therefore, $u_6 = u_4$. If u_7 is an inner vertex of P_1, then an arbitrary edge of the path in C_0 from u_3 into u_7 is contained in C_0, C_1 and the cycle which is the union of P_2', P_1' and the

path in C_0 from u_3 into u_7, which is a contradiction. Therefore, u_7 lies on the path in C_0 from u_4 into u_3. We see that C_1 and C_2 have only one common vertex u_4. Thus we may proceed further and we obtain further cycles $C_3, \ldots, C_k$. The cycles $C_1, C_2, \ldots, \ldots, C_k$ then correspond to the cycles $A_1, A_2, \ldots, A_n$ from the definition of $\mathscr{A}$. As G cannot contain further vertices and edges, we have $G \in \mathscr{A}$ (Fig. 3).

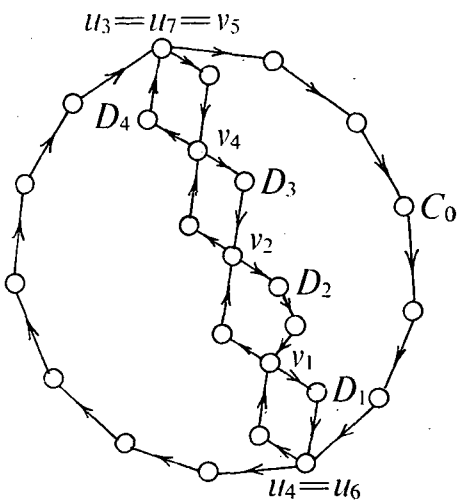


Fig. 2

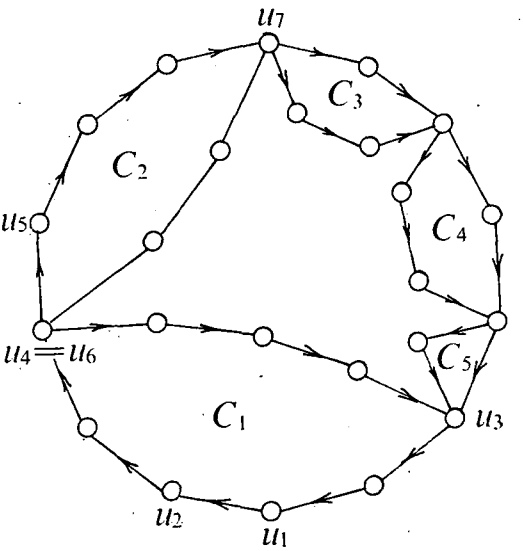


Fig. 3

Сильно связные орграфы, в которых каждая дуга принадлежит точно двум циклам

В статье характеризован класс всех конечных сильно связных ориентированных графов, в которых каждая дуга принадлежит точно двум циклам. Это является частичным решением одной проблемы предложенной А. Ádám-ом.

DEPARTMENT OF MATHEMATICS,
INSTITUTE OF MECHANICAL
AND TEXTILE TECHNOLOGY,
KOMENSKÉHO 2.
460 01 LIBEREC, CZECHOSLOVAKIA

Reference

[1] ÁDÁM, A., On some open problems of applied automaton theory and graph theory (suggested by the mathematical modelling of certain neuronal networks), *Acta Cybernetica*, v. 3, 1977, pp. 187—214.

(Received Feb. 21, 1978)

Some remarks on the chromatic number of the strong product of graphs

By K. VESZTERGOMBI

Let G_1, G_2 be two graphs. Let $V(G_1)=\{x_1, \ldots, x_n\}$, $V(G_2)=\{y_1, \ldots, y_m\}$ be the sets of points, $E(G_1)$, $E(G_2)$ the sets of edges. The definition of the strong product $H=G_1\times G_2$, is the following:

$$V(H) = \{(x_i, y_j) | x_i \in V(G_1),\ y_j \in V(G_2)\}$$

$$E(H) = \{((x_i, y_j), (x_k, y_l)) |\ \text{either}\quad x_i = x_k \quad \text{and} \quad (y_j, y_l) \in E(G_2)$$
$$\text{or} \quad (x_i, x_k) \in E(G_1) \quad \text{and} \quad y_j = y_l$$
$$\text{or} \quad (x_i, x_k) \in E(G_1) \quad \text{and} \quad (y_j, y_l) \in E(G_2)\}.$$

The sets $\{(x_i, y_j) | x_i \text{ fixed}, y_j \in V(G_2)\}$ will be called *rows*, the sets $\{(x_i, y_j) | x_i \in V(G_1), y_j \text{ fixed}\}$ will be called *columns*. There are some trivial estimations for the chromatic number of the product. Let $\chi(G)$ denote the chromatic number of the graph G, we have then the following inequality (see e.g. [1], [2])

$$\max(\chi(G_1), \chi(G_2)) \leqq \chi(H) \leqq \chi(G_1) \cdot \chi(G_2).$$

The upper bound is sharp, in the sense, that we have equality in many cases, for instance if in both G_1 and G_2, the chromatic number equals to the clique number. (The clique number is the maximum cardinality of complete subgraphs.) In this sense the lower bound is not sharp. In the following we give a better lower estimation.

Let us denote by K_2 the single edge, we have then:

Theorem 1. $\chi(K_2\times G) \geqq \chi(G)+2$.

Proof. We give an indirect proof. Let $\chi(G)=k$. Let us suppose that we have coloured the product $K_2\times G$ in $k+1$ colours. $1, 2, \ldots, k+1$. In this case we can colour G as follows. Let us denote the points of K_2 by a and b. Then we can colour $x_i \in V(G)$ with the smaller one of the colours of (a, x_i) and (b, x_i) if this minimum is smaller than k. If this minimum equals k, then we colour x_i by $k-1$. This colouring is a good colouring of G. In fact, the pairs with minimum k or $k-1$ cannot be adjacent, since this would give a complete graph on 4 points, coloured in three colours, which is a contradiction. This way we have a colouring of G with $k-1$ colours and this contradicts our assumption that $\chi(G)=k$. So we have proved the theorem.

In the case when $\chi(G)=2$ our lower bound coincides with the upper bound, so this is a trivial case.

COROLLARY. If both G_1 and G_2 have at least one edge, then

$$\chi(G_1\times G_2) \geqq \max(\chi(G_1), \chi(G_2))+2.$$

Next we examine the case $\chi(G_1)=\chi(G_2)=3$.

Theorem 2. The product of any two odd circuits longer than 3 can be coloured with 5 colours.

Proof. Let us denote the circuit of length m by C_m. One can easily see that the colouring of $C_5\times C_5$ shown on Fig. 1 with 5 colours is a good colouring.

1	2	3	4	5
4	5	1	2	3
2	3	4	5	1
5	1	2	3	4
3	4	5	1	2

Fig. 1

For $C_5\times C_{2l+1}$ ($l>2$), we can do the following (see Fig. 2).

1	2	3	4	5					
4	5	1	2	3					
2	3	4	5	1					
5	1	2	3	4					
3	4	5	1	2					

$l-1$ times

Fig. 2

The first 5 columns are coloured in the same way as in $C_5\times C_5$, and we repeat the colouring of the 4th, and 5th columns $l-1$ times. This trivially gives a good colouring. In the case of $C_{2k+1}\times C_{2l+1}$ ($k>2$) first we colour $C_5\times C_{2l+1}$, then we repeat the colouring of the 4th and 5th rows $k-1$ times.

Remark 1. Consider the graph $K_2\times C_5$. It has 10 points. In a 5-colouration of this graph, a colour can occur at most twice, therefore each colour must occur exactly twice. So if one row contains all 5 colours then so does the other. Moreover one can easily check that if one row is coloured say (1 2 3 4 5) then the other is either (3 4 5 1 2) or (4 5 1 2 3).

Theorem 3. $C_5\times C_5$ can be coloured with 5 colours essentially uniquely.

Proof. We show, that we can colour $C_5\times C_5$ with 5 colours only so, that in every row we use all the 5 colours. Suppose indirectly that we have found a colouring,

in which for instance in the first row colour 5 does not occur. Then in the second row colour 5 must occur twice by Remark 1. Continuing the colouring, in the third row we cannot have 5, in the fourth row we must find it twice, in the fifth row we cannot have number 5, which is impossible.

So suppose that the first row is coloured 1 2 3 4 5 by Remark 1 we may assume that the second row is coloured (4 5 1 2 3). The third row is therefore either (1 2 3 4 5) or (2 3 4 5 1). The first possibilty cannot occur, because the above argument applies to the columns as well, therefore the colours of the first column must be different. Going on similarly we get that the fourth and fifth rows are (5 1 2 3 4), (3 4 5 1 2).

In the sequel we present a characterization of graphs which give a five-colourable product with C_5. Before stating the theorem we need the following definition.

A homomorphism of G into H is a mapping $\varphi: V(G) \to V(H)$ for which we have that whenever $(x, y) \in E(G)$ then $(\varphi(x), \varphi(y)) \in E(H)$.

Theorem 4. Let G be a graph, for which $\chi(G) > 2$. Then $\chi(G \times C_5) = 5$ if and only if there is a homomorphism of G into C_5.

Proof. I. We know that we have a 5-colouring for $C_5 \times C_5$. Let us take a homomorphism φ of G into C_5. Let $v \in V(G)$, then $\varphi(v) \in V(C_5)$ and we colour the row $v \times C_5$ in the same way as the $\varphi(v)^{\text{th}}$ row of $C_5 \times C_5$. This colouring is obviously good because of the definition of homomorphism.

II. For the proof of the "only if" part we introduce the *5-colouration graph* of C_5. We define the *k-colouration graph* of G in the following way. Let $k \geqq \chi(G)$. The vertices of the k-colouration graph are the different colourings of G with k given colours (which need not occur all in the colouring) and two vertices a, b are adjacent if and only if $G \times K_2$ can be coloured with k colours so that the colouring of the first row corresponds to a, and the colouring of the second row corresponds to b.

Lemma 1. The 5-colouration graph of C_5 has the following structure. There are 5! colouration with 5 colours in which every colour occurs exactly once. Those colourations form 4! pentagons. The remaining colourations form a bipartite graph.

The proof of this Lemma is straightforward from the proof of Theorem 3.

Continuing the proof of Theorem 4, assume that we have a 5-colouring of $G \times C_5$. Every row $v_i \times C_5$ expresses a 5-colouring of the pentagon and this 5-colouring corresponds to a point in the 5-colouration graph. We take the mapping ψ for which $\varphi(v_i)$ is this point of the 5-colouration graph, this mapping φ will be a homomorphism of G into the 5-colouration graph. Lemma 1 gives the structure of the 5-colouration graph of the pentagon and this graph has a homomorphism ψ into the pentagon (obviously). Now the mapping $\psi\varphi$ is a homomorphism of G into the pentagon, and this is what we wanted to show.

From the generalization of the above theorem, one can get the next theorem:

Theorem 5. Let $k \geqq \chi(G)$, and assume that G has at least one edge. Then for a graph H we have $\chi(G \times H) \leqq k$ if and only if H has a homomorphism into the k-colouration graph of G.

Proof. I. The rows of the product are copies of the graph G. So if $H \times G$ is k coloured then the colouring of a row $v \times G$ corresponds to some vertex of the k-colouration graph. This defines a homomorism of H into the k-colouration graph of G.

II. Conversely, assume that H has a homomorphism φ into the k-colouration graph of G. Then colour the row $v \times G$ as in the colouration $\varphi(v)$.

A very simple argument shows that if we take the product $K_n \times C_{2k+1}$ (K_n is the complete n-graph) then the chromatic number of this product decreases as k increases, but for $k > n$ we always get $\chi(K_n \times C_{2k+1}) = 2n+1$. In particular, $\chi(K_3 \times C_{2k+1}) = 7$. One can have the feeling that if we take the product of odd circuits with 3-chromatic graphs with girth (the length of the shortest circuit in the graph) larger than 3, we can get smaller chromatic number for the product, probably only 5. We show some examples which contradict to this tendency.

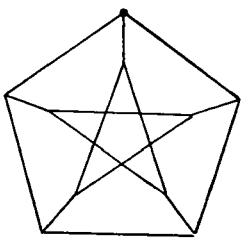


Fig. 3

Lemma 2. Let $\alpha(G)$ denote the maximum number of independent points in G. Then for any G and k, we have the inequality:

$$2\alpha(G \times C_{2k+1}) \leqq (2k+1)\alpha(G).$$

Proof. The maximum number of independent points in $G \times K_2$ is the same as in G. Summing this for all edges of C_{2k+1} the inequality follows.

Let P_5 denote the Petersen-graph shown in Fig. 3. It is obvious that $\chi(P_5) = 3$. Then we have the following theorem.

Theorem 6. The chromatic number of the product $P_5 \times C_{2l+1}$ is always greater than 5.

Proof. We have $\alpha(P_5) = 4$. From Lemma 2 we get the inequality:

$$\alpha(P_5 \times C_{2l+1}) \leqq 2(2l+1).$$

We show that here the equality cannot hold. Suppose it does. Then equality holds in the proof of Lemma 2 on all edges of C_{2l+1}.

Let us consider the independent vertices in the first row of the product $P_5 \times C_{2l+1}$. Let its number be f ($f = 0, 1, 2, 3, 4$). We can choose from the second row at most $4-f$ independent vertices. Since equality holds in the proof of Lemma 2, we have precisely $4-f$ vertices from the second row. From the third row we must choose f independent vertices again. Continuing this procedure from the $(2l)^{\text{th}}$ row we must choose $4-f$ vertices. From the $(2l+1)^{\text{st}}$ row we must choose f vertices because this row is the neighbour of the $(2l)^{\text{th}}$ row, but this is the neighbour of the first row too, se we must choose $4-f$ vertices from the $(2l+1)^{\text{st}}$ row. This is possible only if $f = 4-f$, $f = 2$. It can be easily seen that if we take in P_5 two independent vertices, then they uniquely determine the maximal independent set which contains these vertices. In the case $f = 2$ this gives that in every $(2i+1)^{\text{st}}$ row ($i < l$) we have the same two vertices. But this excludes any vertices in the $(2l+1)^{\text{st}}$ row because it is the neighbour of both of the first and $(2l)^{\text{th}}$ row. Thus we get

$$\alpha(P_5 \times C_{2l+1}) < 2(2l+1).$$

From this and from the well-known inequality

$$\frac{|V(G)|}{\alpha(G)} \leqq \chi(G). \tag{1}$$

We get that $\chi(P_5 \times C_{2l+1}) > 5$ for any l.

Let us take the "generalized Petersen-graph", P_{13}, seen in Fig. 4.

One can easily see that $\chi(P_{13}) = 3$. The length of the shortest circuit is 7. For this graph we have the following theorem:

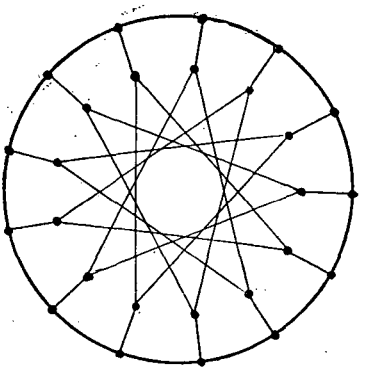


Fig. 4

Theorem 7. For any k $\chi(P_{13} \times C_{2k+1}) > 5$.

Proof. Let us consider the maximal number of independent vertices in P_{13}. In the outer and the inner circuit there can be at most six independent vertices. We show that if we have in the outer circuit six independent vertices, then we can have in the inner circuit at most four vertices. The outer six independent vertices exclude their six inner neighbours (seei Fg. 5a).

It is essentially unique to choose six independent vertices in the outer circuit. The remaining seven vertices in the inner circuit consist of one isolated vertex and from three independent edges and from this graph we can choose at most four independent vertices (as it is indicated in Fig. 5a).

We can argue similarly in the case when we have six independent vertices in the inner circuit (see Fig. 5b). Thus $\alpha(P_{13}) = 10$. Using Lemma 2 we have the following inequality:

$$\alpha(P_{13} \times C_{2k+1}) \leqq (2k+1) \cdot 5.$$

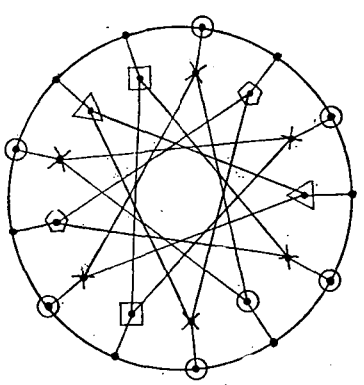


Fig. 5a

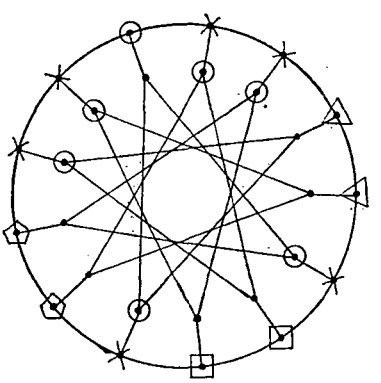


Fig. 5b

From this and using inequality (1) we have:

$$\chi(P_{13}\times C_{2k+1}) \geqq \frac{2\cdot 13\cdot(2k+1)}{5(2k+1)} > 5.$$

So we have proved the theorem.

Problems. 1. Give a better lower bound for the $\chi(G_1\times G_2)$ if $\chi(G_1)$, $\chi(G_2)$ are larger than 3.

2. Prove that for any large g one can find G_1 and G_2, for which $\chi(G_1)=\chi(G_2)=3$, the girth of both graphs is larger than g but $\chi(G_1\times G_2)\geqq 6$.

3. It would be interesting to determine the structure of the k-colouration graphs of some classes of graphs, to get similair results as in Theorem 4.

Reference

[1] BERGE, C., *Graphs and hypergraphs*, North Holland, 1973.
[2] BOROWIECKI, M., On chromatic number of products of two graphs, *Colloq. Math.*, v. 25, 1972, pp. 49—52.

(Received April 12, 1978)

On Sperner families in which no 3 sets have an empty intersection

By H.-D. O. F. GRONAU

1. Introduction

Let $\mathcal{G}(r, k)$ denote the set of all Sperner families $\mathcal{F}$ (i.e. $X \not\subset Y$ for all different $X, Y \in \mathcal{F}$) on $R=[1, r]$ (the interval of the first r natural numbers with $r \geqq 3$) satisfying $\bigcup_{i=1}^{k} X_i \subset R$ for all $X_i \in \mathcal{F}$ $(i=1, \ldots, k)$ where $\subset$ is used in the strong sense. Furthermore we use the following notations:

$$\mathcal{G}^1(r, k) = \{\mathcal{F}: \mathcal{F} \in \mathcal{G}(r, k), \bigcup_{X \in \mathcal{F}} X = R\},$$

$$\mathcal{G}^0(r, k) = \{\mathcal{F}: \mathcal{F} \in \mathcal{G}(r, k), \bigcup_{X \in \mathcal{F}} X \subset R\},$$

$$n(r, k) = \max_{\mathcal{F} \in \mathcal{G}} |\mathcal{F}|, \; n^1(r, k) = \max_{\mathcal{F} \in \mathcal{G}^1} |\mathcal{F}| \quad \text{and} \quad n^0(r, k) = \max_{\mathcal{F} \in \mathcal{G}^0} |\mathcal{F}|.$$

We notice that $\mathcal{G}^1(r, k)=\emptyset$ holds for $k \geqq r$.

$n(r, 2)$ was determined by E. C. MILNER [6] (for the dual case) and later by A. BRACE and D. E. DAYKIN [1], and $n(r, k)$ with $k \geqq 4$ was determined by the author [3].

For $n(r, 3)$ the following two configurations are known:

$$n(r, 3) = \left(\left[\frac{r-1}{\left[\frac{r-1}{2}\right]} \right] \right) + 1 \tag{1}$$

and

$$n(r, 3) = \left(\left[\frac{r-1}{\left[\frac{r-1}{2}\right]} \right] \right). \tag{2}$$

P. FRANKL [2] proved (1) for large enough even r (e.g. for $r>1000$) and (2) for large enough odd r (e.g. for $r>300$). The author [3] showed (1) for $r=7$ and even

$r>400$, and (2) for all odd r with the exception of the following 12 values: 7, 11, 13, 17, 19, 23, 25, 29, 31, 35, 37 and 43.

In the present paper we prove

(1) for $r=4, 6, 114$ and even $r\geqq 120$ and

(2) for $r=11, 17, 23, 29, 35, 43$.

We observe that exchanging all $X\in\mathscr{F}$ by $R\setminus X$ we get analogous results for Sperner families in which no 3 sets have an empty intersection.

We shall sharpen Theorem 5 of [3] in the case $k=3$. There we divided a maximal family $\mathscr{F}\in\mathscr{G}(r, 3)$ to two families $\mathscr{F}_0$ and $\mathscr{F}_1$, and showed

$$|\mathscr{F}_0|\leqq\binom{r-1}{\left[\frac{r-2}{2}\right]}\quad\text{and}\quad|\mathscr{F}_1|\leqq\binom{r-1}{\left[\frac{r-1}{3}\right]-1}.$$

In fact $|\mathscr{F}_1|$ depends on $|\mathscr{F}_0|$. For $k=3$ and even r, $|\mathscr{F}_0|=\binom{r-1}{\left[\frac{r-2}{2}\right]}$ implies $|\mathscr{F}_1|=1$.

In section 2 we shall present our main results and give a new type estimation of families of sets, which will be used in section 3 to prove a theorem analogous to Theorem 5 [3]. Finally, in section 4 we shall prove our main result.

2. Main results

Throughout this paper let $a=\left[\frac{r-2}{2}\right]$ and $b=\left[\frac{r-1}{3}\right]$.

Theorem 1. 1° $n(r, 3)=\binom{r-1}{\left[\frac{r-1}{2}\right]}+1$ *for* $r=4, 6, 114$ *and even* $r\geqq 120$,

2° $n(r, 3)=\binom{r-1}{\left[\frac{r-1}{2}\right]}$ *for* $r=11, 17, 23, 29, 35, 43$.

Let $r\geqq 4$. Then $n(r, 3)$, $n^1(r, 3)$ and $n^0(r, 3)$ exist and it holds $n(r, 3)=\max(n^1(r, 3), n^0(r, 3))$.

For $\mathscr{F}\in\mathscr{G}^0(r, 3)$ there is an element $v\in R$ such that $\mathscr{F}$ is a Sperner family on $R\setminus\{v\}$, and it follows by SPERNER's theorem [7]:

Lemma 1. $n^0(r, 3)=\binom{r-1}{\left[\frac{r-1}{2}\right]}$.

We shall use the following lemma shown in more general form in [3] (Lemma 2).

Lemma 2. Let $\mathcal{F}\in\mathcal{G}^1(r,3)$ such that $|\mathcal{F}|=n^1(r,3)$ and $\max\limits_{X\in\mathcal{F}}|X|$ is minimal. Then $|X|\leqq a$ holds for all $X\in\mathcal{F}$.

Lemma 3. Let $s\leqq\frac{r}{2}$ be an integer and let $\mathcal{F}_s$ denote an arbitrary family of different s-element subsets of R. Finally, let $\mathcal{F}_{2s}^*$ denote the largest family of $(2s)$-element subsets of R such that for every $X\in\mathcal{F}_{2s}^*$ there is at least one pair (Y,Z) of subsets of $\mathcal{F}_s$ satisfying $Y\cup Z=X$. Then

$$|\mathcal{F}_{2s}^*|\geqq\frac{\binom{r-s}{s}}{\binom{2s-1}{s}}|\mathcal{F}_s|-\binom{r}{2s}.$$

Proof. Let us consider the following families:

$$\overline{\mathcal{F}_s}=\{X: X\subset R,\ |X|=s,\ X\notin\mathcal{F}_s\},$$

$$\overline{\mathcal{F}_{2s}^*}=\{X: X\subset R,\ |X|=2s,\ X\notin\mathcal{F}_{2s}^*\}.$$

Then for any $X\in\overline{\mathcal{F}_{2s}^*}$ there is no pair (Y,Z) of sets of $\mathcal{F}_s$ with $Y\cup Z=X$. For every such $X\in\overline{\mathcal{F}_{2s}^*}$ there exist exactly $\frac{1}{2}\binom{2s}{s}=\binom{2s-1}{s}$ unordered pairs (Y,Z) with $|Y|=|Z|=s$ and $Y\cup Z=X$. All these sets are mutually disjoint, i.e., at least $\binom{2s-1}{s}$ s-element subsets belong to $\overline{\mathcal{F}_s}$ for every $X\in\overline{\mathcal{F}_{2s}^*}$.

On the other hand for every s-element set Y of R there exist exactly $\binom{r-s}{s}$ disjoint s-element sets Z. Hence

$$|\overline{\mathcal{F}_{2s}^*}|\binom{2s-1}{s-1}\leqq|\overline{\mathcal{F}_s}|\binom{r-s}{s}.$$

Using $|\overline{\mathcal{F}_{2s}^*}|=\binom{r}{2s}-|\mathcal{F}_{2s}^*|$ and $|\overline{\mathcal{F}_s}|=\binom{r}{s}-|\mathcal{F}_s|$ we obtain the inequality of Lemma 3. □

3. An upper bound for $n^1(r,3)$

Let $\mathcal{F}\in\mathcal{G}^1(r,3)$ such that $|\mathcal{F}|=n^1(r,3)$ and $\max\limits_{X\in\mathcal{F}}|X|$ is minimal. By Lemma 2, we have $|X|\leqq a$ for all $X\in\mathcal{F}$. The numbers $p_i=|\{X: X\in\mathcal{F},\ |X|=i\}|$ $(i=0,\ldots,r)$ are called parameters of the family $\mathcal{F}$. $\mathcal{SF}$ denotes the canonical Sperner family (see A. J. W. HILTON [4]).

Now we decompose $\mathcal{F}$ to the subfamilies $\mathcal{D}$, $\mathcal{E}$ and $\mathcal{H}$ defined as follows.
— $\mathcal{D}$ is a subfamily of $\mathcal{F}$ with $\mathcal{SD}=\{X: X\in\mathcal{SF},\ r\notin X\}$.
— $\mathcal{E}=\{X: X\in\mathcal{F}\setminus\mathcal{D},\ |X|\leqq r-2a-1\}$.
— $\mathcal{H}=\{X: X\in\mathcal{F}\setminus\mathcal{D},\ |X|\geqq r-2a\}$.

1. It has been proved by A. J. W. HILTON [4] that all $X\in\mathcal{F}$ with $|X|>b$ belong to $\mathcal{D}$. $\mathcal{S}\mathcal{D}$ is a Sperner family on $R\setminus\{r\}$. Using $\binom{r-1}{|X|}\leqq\binom{r-1}{a-1}$ for $|X|\leqq a-1<\frac{r-1}{2}$, by LUBELL's inequality [5] we obtain

$$\sum_{X\in\mathcal{S}\mathcal{D}}\frac{1}{\binom{r-1}{|X|}}=\sum_{\substack{X\in\mathcal{S}\mathcal{D}\\|X|=a}}\frac{1}{\binom{r-1}{a}}+\sum_{\substack{X\in\mathcal{S}\mathcal{D}\\|X|\leqq a-1}}\frac{1}{\binom{r-1}{|X|}}\leqq 1,$$

$$\frac{p_a}{\binom{r-1}{a}}+\frac{|\mathcal{S}\mathcal{D}|-p_a}{\binom{r-1}{a-1}}\leqq 1$$

and

$$|\mathcal{D}|=|\mathcal{S}\mathcal{D}|\leqq\frac{a}{r-a}\binom{r-1}{a}+\frac{r-2a}{r-a}p_a.$$

2. $\mathcal{J}=\{X\colon X\cup\{r\}\in\mathcal{S}(\mathcal{D}\cup\mathcal{E}),\ r\notin X\}$ is a Sperner family of cardinality $|\mathcal{E}|$ on $R\setminus\{r\}$ and $|X|\leqq r-2a-2$ holds for all $X\in\mathcal{J}$.

By LUBELL's inequality [5] we obtain

$$\sum_{X\in\mathcal{J}}\frac{1}{\binom{r-1}{|X|}}\leqq 1,\quad\frac{|\mathcal{J}|}{\binom{r-1}{r-2a-2}}\leqq 1\quad\text{and}\quad|\mathcal{E}|=|\mathcal{J}|\leqq\binom{r-1}{r-2a-2}.$$

3. Let $\mathcal{F}_{2a}^{**}=\{X\colon R\setminus X\in\mathcal{F}_{2a}^{*}\}$. Then $\mathcal{D}\cup\mathcal{H}\cup\mathcal{F}_{2a}^{**}$ is a Sperner family. We notice that $|X|\geqq r-2a$ holds for all $X\in\mathcal{D}\cup\mathcal{H}$ [1] and $|X|=r-2a$ holds for all $X\in\mathcal{F}_{2a}^{**}$. Clearly, $\mathcal{D}\cup\mathcal{H}$ and $\mathcal{F}_{2a}^{**}$ are Sperner families themselves. We have only to show that there is no pair (Y,Z) with $Y\in\mathcal{F}_{2a}^{**}$ and $Z\in\mathcal{D}\cup\mathcal{H}$ satisfying $Y\subseteq Z$. Let us assume the contrary. Then there are two sets Y_1, $Y_2\in\mathcal{D}$ with $Y_1\cup Y_2=R\setminus Y$. Hence for the sets Y_1, Y_2, $Z\in\mathcal{F}$ it follows $Y_1\cup Y_2\cup Z=(R\setminus Y)\cup Z\supseteq(R\setminus Y)\cup Y=R$, which is impossible for $\mathcal{F}\in\mathcal{G}(r,3)$.

$\mathcal{J}'=\{X\colon X\cup\{r\}\in\mathcal{S}(\mathcal{D}\cup\mathcal{H}\cup\mathcal{F}_{2a}^{**}),\ r\notin X\}$ is a Sperner family on $R\setminus\{r\}$. If q_i, q_i' and q_i'' are the parameters of the families $\mathcal{J}'$, $\mathcal{H}$ and $\mathcal{F}_{2a}^{**}$, respectively, then $q_i=q_{i+1}'+q_{i+1}''$ holds. By LUBELL's inequality [5], using $\binom{r-1}{|X|}\leqq\binom{r-1}{b}$ for $|X|\leqq b<\frac{r-1}{2}$, we get

$$\sum_{X\in\mathcal{J}'}\frac{1}{\binom{r-1}{|X|}}\leqq 1,\quad\sum_{X\in\mathcal{H}}\frac{1}{\binom{r-1}{|X|-1}}+\sum_{X\in\mathcal{F}_{2a}^{**}}\frac{1}{\binom{r-1}{r-2a-1}}\leqq 1$$

and

$$\frac{|H|}{\binom{r-1}{b-1}}+\frac{|\mathcal{F}_{2a}^{**}|}{\binom{r-1}{r-2a-1}}\leqq 1.$$

By Lemma 3 using $|\mathcal{F}|=n^1(r,3)$ and the estimations for $\mathcal{D}$, $\mathcal{E}$ and $\mathcal{H}$ we obtain

[1] as $\min_{X\in\mathcal{D}\cup\mathcal{H}}|X|\leqq r-2a-1$ would imply $\mathcal{H}=\emptyset$ and, together with 1. and 2., the estimation given in Theorem 2.

Theorem 2.

$$n^1(r,3) \leqq$$

$$\leqq \max_{p_a}\left(\frac{a}{r-a}\binom{r-1}{a}+\frac{r-2a}{r-a}p_a+\binom{r-1}{r-2a-2}+\binom{r-1}{b-1}\frac{2(r-a)}{r-2a}\left(1-\frac{p_a}{\binom{r-1}{a}}\right)\right).$$

4. Proof of Theorem 1

Clearly, $n(r,3)=\max\left(n^1(r,3), \left(\left[\frac{r-1}{2}\right]\right)\right)$ holds by Lemma 1.

1°. Let r be even. Then all a-element subsets of $R\setminus\{r\}$ and the set $\{r\}$ form a family $\mathscr{F}\in\mathscr{G}(r,3)$ having the cardinality $\binom{r-1}{a}+1$. So we have only to show that the right side of the inequality of Theorem 2 has the value $\binom{r-1}{a}+1$, too.

For $r=4$ it is easy to see that $n^1(4,3)=4$ holds.

Now let $r=6$, 114 or $r\geqq 120$.

The function $f(p_a)$, of which we consider the maximum in Theorem 2, is a linear function in p_a. We have to take the maximum over the interval $\left[0, \binom{r-1}{a}\right]$, as an immediate consequence of A. J. W. HILTON's result [4] which we used in the definition of $\mathscr{D}$. We have $f\left(\binom{r-1}{a}\right)=\binom{r-1}{a}+1$. We have only to show that the factor of p_a in $f(p_a)$ is positive (or equal to o), i.e., using $r-2a=2$,

$$\frac{2}{r-a}-(r-a)\frac{\binom{r-1}{b-1}}{\binom{r-1}{a}}>0. \tag{3}$$

(3) is equivalent to

$$M(r)=\frac{2(r-b)(r-b-1)\ldots(r-a+1)}{(r-a)a(a-1)\ldots b}>1. \tag{4}$$

$$M(6)=\frac{5}{4} \quad \text{and} \quad M(124)=\frac{35047435882784}{34511088479301}>1.$$

Furthermore,

$$\frac{M(6t+10)}{M(6t+4)}=\frac{2^{10}}{3^6}\,\frac{t+\frac{7}{4}}{t+2}\,\frac{t+\frac{3}{2}}{t+2}\,\frac{t+\frac{5}{4}}{t+\frac{5}{3}}\left(\frac{t+1}{t+\frac{4}{3}}\right)^2\frac{t+\frac{1}{2}}{t+\frac{2}{3}}=g(t)$$

is monotonically increasing, because $\frac{t+x}{t+y}$ is monotonically increasing for fixed x and y with $x<y$.

For $t \geqq 20$ we obtain $g(t) \geqq g(20) = \frac{127766373}{99866624} > 1$. By induction it follows that $M(6t+4) > 1$ for $t \geqq 20$.

Moreover we have

$$\frac{M(6t+2)}{M(6t+4)} = \frac{9}{8} \frac{t+1}{t+\frac{3}{4}} \frac{t+1}{t+\frac{2}{3}} \frac{t+\frac{1}{3}}{t} > \frac{9}{8} > 1$$

and

$$\frac{M(6t)}{M(6t+4)} = \frac{3^4}{2^6} \frac{t+1}{t+\frac{3}{4}} \frac{t+1}{t+\frac{1}{2}} \frac{t+\frac{2}{3}}{t-\frac{1}{2}} > \frac{81}{64} > 1,$$

which proves $M(2t) > 1$ for $t \geqq 60$.

Finally we complete our proof by $\frac{M(114)}{M(124)} = \frac{59025914157}{53793208352} > 1$.

2°. In [3] the author proved the following estimation for $|\mathscr{E} \cup \mathscr{H}| : |\mathscr{E} \cup \mathscr{H}| \leqq \leqq \binom{r-1}{b-1}$. Using our estimation for $|\mathscr{D}|$ we obtain $|\mathscr{F}| \leqq \binom{r-1}{a} \frac{a}{r-a} + \frac{r-2a}{r-a} p_a + + \binom{r-1}{b-1}$. Both, this estimation and the bound given in Theorem 2 are valid for each $|\mathscr{F}|$. It suffices to show that for every p_a one of our upper bounds is less than $\binom{r-1}{a+1}$, because in this case r is odd, i.e. $\left[\frac{r-1}{2}\right] = a+1$. We distinguish the following cases.

1. $p_a < \frac{a+3}{3} \left\{ \binom{r-1}{a+1} - \binom{r-1}{a-1} - \binom{r-1}{b-1} \right\}$. Then $|\mathscr{F}| < \binom{r-1}{a+1}$ follows from our last estimation.

2. $p_a \geqq \frac{a+3}{3} \left\{ \binom{r-1}{a+1} - \binom{r-1}{a-1} - \binom{r-1}{b-1} \right\} = \frac{2}{3} \frac{2a+3}{a+1} \binom{r-1}{a} - \frac{a+3}{3} \binom{r-1}{b-1}$.

Then we use the estimation of Theorem 2. First we prove that the factor of p_a in $f(p_a)$ is negative, i.e.

$$\frac{r-2a}{r-a} - \frac{2(r-a)}{r-2a} \frac{\binom{r-1}{b-1}}{\binom{r-1}{a}} < 0. \tag{5}$$

(5) is equivalent to

$$N(r) = \frac{9(r-b)(r-b-1)\ldots(a+4)}{2(a+3)a(a-1)\ldots b} < 1.$$

We have that

$$\frac{N(6t+5)}{N(6t-1)} = \frac{2^{10}}{3^6} \frac{t+\frac{3}{4}}{t+\frac{4}{3}} \frac{t+\frac{1}{2}}{t+\frac{4}{3}} \frac{t+\frac{1}{4}}{t+\frac{2}{3}} \frac{t-\frac{1}{2}}{t-\frac{1}{3}} = g'(t)$$

is monotonically increasing by our remark above.

For $2 \leqq t \leqq 5$ we obtain $g'(t) \leqq g'(5) = \frac{6072}{6137} < 1$. From $N(11) = \frac{3}{7}$, $N(6t-1) < 1$ follows by induction for $2 \leqq t \leqq 6$. Finally, we get $N(43) = \frac{10179}{59432} < 1$. $f(p_a)$ takes the maximum in the described interval at $p_a = \frac{a+3}{3}\left\{\binom{r-1}{a+1} - \binom{r-1}{a-1} - \binom{r-1}{b-1}\right\}$, consequently. We will complete our proof by showing the following inequality.

$$\left\{\frac{3}{2}\frac{2a+3}{a+1}\binom{r-1}{a} - \frac{a+3}{3}\binom{r-1}{b-1}\right\}\left\{\frac{3}{a+3} - \frac{2}{3}(a+3)\frac{\binom{r-1}{b-1}}{\binom{r-1}{a}}\right\} + \frac{a}{a+3}\binom{r-1}{a} +$$

$$+ (r-1) + \frac{2}{3}(a+3)\binom{r-1}{b-1} < \binom{r-1}{a+1}.$$

This inequality is equivalent to

$$w(r) = \binom{r-1}{b-1}\left\{1 + \frac{2(a+3)^2}{9(a+1)}\left(1 - (a+1)\frac{\binom{r-1}{b-1}}{\binom{r-1}{a}}\right)\right\} - (r-1) > 0.$$

$$w(11) = 112 > 0.$$

Furthermore we prove the inequality $w'(r) = \frac{(a+3)(a+1)}{(2a+13)}\frac{\binom{r-1}{b-1}}{\binom{r-1}{a}} \leqq \frac{1}{2}$ for $r = 17, 23, 29, 35, 43$ by referring to the following table:

r	17	23	29	35	43
$w'(r)$	$\frac{140}{297}$	$\frac{1}{2}$	$\frac{154}{323}$	$\frac{442}{1035}$	$\frac{9044}{19981}$

Using this estimation of $w'(r)$ we get first

$$w(r) \geqq \binom{r-1}{b-1}\left(1 + \frac{2(a+3)^2}{9(a+1)}\left(1 - \frac{2a+13}{2(a+3)}\right)\right) - (r-1)$$

$$\geqq \binom{r-1}{b-1}\frac{2(a-6)}{9(a+1)} - (r-1),$$

then $w(17) \geqq \frac{311}{9} > 0$. $r \geqq 17$ implies $\frac{a-6}{a+1} \geqq \frac{1}{8}$ and for $2 \leqq i \leqq b-1$ we have $\frac{r-b-1+i}{i} > 3$. Hence for $r \in \{23, 29, 35, 43\}$:

$$w(r) \geqq (r-1)\frac{2(a-6)}{9(a+1)} \prod_{i=2}^{b-1} \frac{r-b-1+i}{i} - (r-1)$$

$$\geqq (r-1)\frac{2}{9}\frac{1}{8}3^{b-2} - (r-1)$$

$$\geqq (r-1)\frac{1}{4}3^3 - (r-1)$$

$$> 0 \quad \text{follows.} \ \square$$

5. Concluding remark

The author conjectures that (1) holds for the remaining even r and (2) holds for the remaining odd r, i.e. 13, 19, 25, 31 and 37.

WILHELM—PIECK-UNIVERSITÄT
SEKTION MATHEMATIK
UNIVERSITÄTSPLATZ 1
DDR—25 ROSTOCK

References

[1] BRACE, A. and D. E. DAYKIN, Sperner type theorems for finite sets, Combinatorics, Proc. Conf. Combinatorial Math. Inst. Oxford, 1972, pp. 18—37.
[2] FRANKL, P., On Sperner families satisfying an additional condition, *J. Combinatorial Theory*, v. 20, 1976, pp. 1—11.
[3] GRONAU, H.-D. O. F., On Sperner families in which no k sets have an empty intersection, *J. Combinatorial Theory*, to appear.
[4] HILTON, A. J. W., Some results on Sperner families, *J. Combinatorial Theory*, to appear.
[5] LUBELL, D., A short proof of Sperner's lemma, *J. Combinatorial Theory* v. 1 1966, p. 299.
[6] MILNER, E. C., A combinatorial theorem on systems of sets, *J. London Math. Soc.*, v. 43, 1968, pp. 204—206.
[7] SPERNER, E., Ein Satz über Untermengen einer endlichen Menge, *Math. Z.*, v. 27, 1928, pp. 544—548.

(Received March 6, 1978)

A possible new model of neurons and neural processes based on the quantum-mechanical theory of measurement

By F. JÓLESZ* and M. SZILÁGYI**

Abstract

A new model of neurons and neural processes is proposed which aims at providing a framework for treating the phenomena of statistical nature in the nervous system. Its conceptual base is the quantum-mechanical theory of measurement and some general characteristics of the interactions between systems. The simpler form of the model takes into account two observables, the stimulus intensity and length, measured by individual neurons, with only one threshold for each. By considerations analogous to those of quantum mechanics an uncertainty relation is derived between the possible accuracy of the measured intensity and time length values. The model is extended to the case of many thresholds and to measurements made by neuron populations which, in fact, generally occur in the nervous system.

Introduction

The paradigms of various sciences, particularly that of biology, have always shown characteristic relationships to the current theories of physics. Approaches to the problem of neural processes have in all ages, and today as well, depended on the generally accepted physical model of the world.

In classical considerations the changes of state of the units are always related to some interactions — "stimulation — excitation" — but there is no attention paid for the unique character of this interaction, namely that it means a kind of measurement, too. In our present knowledge, the measuring interaction is in general not negligible to the interaction determining the change of state. The principle of strict individuality of neural objects results, however, in an essentially dynamic treatment in which the neural units — like the organism as a whole — behave as the subject in the interaction with their own environment and remain separable from it during the process. Therefore, the states and their transitions belong to the units themselves and not to the interactions.

In constrast to this dynamic picture the recent use of some methods of statistical

mechanics in the theoretical approach to neural systems was a serious step forward (Wiener, 1965; Cowan, 1968; Cowan, 1970, Amari, 1974).

Today the application of techniques developed in statistical mechanics seems unavoidable in the study of the nervous system and these techniques meet quite general acceptance. One can, nevertheless, expect that in the light of further experiences these statistical concepts would not be satisfactory enough, and the group of available theoretical methods should again be enlarged by developing ideas more departed from those of classical physics.

In our opinion three important features of the neural phenomena point to this direction. First, the discrete character of the structural and functional organization of the nervous system on various levels. Second, the probabilistic character of the distribution of activity in space and time. Third, last but not least, the existence of the above mentioned "measuring process" itself.

Generalization of the concept of "measurement"

Before trying to describe any hypothetical "measuring process" in the nervous system it is worth discussing the meaning of some general terms to be used. If we want to generalize the concept of measurement we must be aware that, in the definition and quantitative characterization of any measurable quantity — i.e. any observable —, the task of finding an appropriate device for the measurement can not be rejected. At the first steps of generalization, however, we need not identify these devices immediately with some concrete physical objects, in particular if we start with empirical experiences in the cases when objects and events are not separable from each other.

Let us try first to find a mathematical model that is fairly general to serve as a framework for any possible structure of events. Mathematically, the concept of an event is considered a primitive notion that is not otherwise defined; our ultimate aim is to get to a formalism for treating the systems of events taking place in the nervous system. Thus every object will be defined only by the system of events belonging to it.

Fortunately, there is already existing a general model of physical systems which applies not only to physical but any other systems as well and is adequate for the description of the measuring process (Mackey, 1963). Let us now sum it up briefly.

Suppose the structure of our system is not changing in time and the values of all the observable quantities are real numbers. The distribution of any of these quantities can be determined by measurements, i.e., by processes which select out a subset of the sample set given for the observables. The observables can have several different distributions; the state α of the system determines which one of them would result as the outcome of the measurement. Thus a real-numbered random variable φ belongs to each observable. Mathematically, φ is a Borel function mapping an (Ω, S, P) measurable space into the set R of real numbers:

$$\varphi: \Omega \to R \tag{1}$$

where (Ω, S) is the set of events being subsets of the sample set Ω and P is a probability measure defined on (Ω, S).

Now, let φ be assigned to a given observable 0 and let E be a given subset of the σ-algebra B of the Borel sets in the set R of real numbers, i.e., $E \in B$. In simple cases E is an interval. The question, then, arises: what is the probability that the value of φ falls into the set E?

According to the definition of φ, there is a subset $\varphi^{-1}(E)$ in (Ω, S) which is mapped to E by φ (Fig. 1). $\varphi^{-1}(E)$ is, of course, an event. As P is the probability measure defined in the same space (Ω, S, P) the probability we asked is $P[\varphi^{-1}(E)]$.

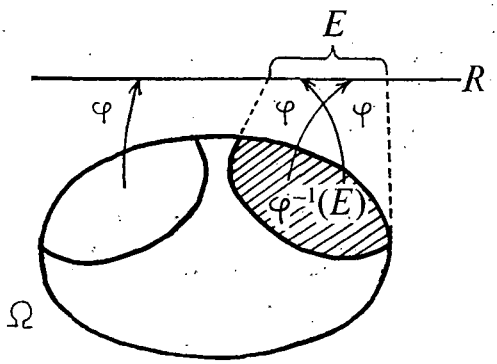


Fig. 1

The distribution of the values of φ given in this way can be called *the distribution induced by P*. We shall denote it by P_φ^α expressing the dependence of this distribution on the state α of the system and on the observable φ Therefore, every system has a family of distributions P_φ^α (Fig. 2).

Mathematically, if we consider several different (Ω, S, P) probability spaces the induced distributions P_φ^α may also be different. Physically, however, we expect that if the state of the system, denoted by α, is the same while the measured observables $\varphi_1, \varphi_2, \ldots, \varphi_n$ are different, the distributions $P_{\varphi_1}^\alpha, P_{\varphi_2}^\alpha, \ldots, P_{\varphi_n}^\alpha$ should be induced by the same probability measure $P = P^\alpha$ depending only on α in a fixed (Ω, S) meas-

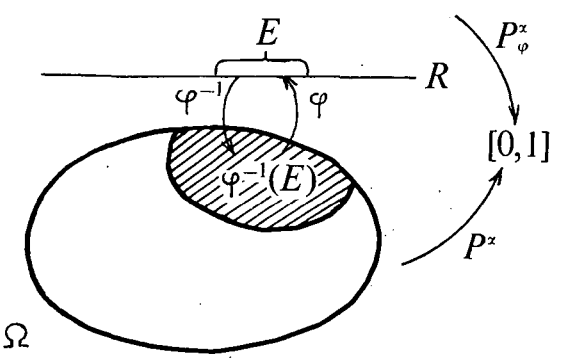


Fig. 2

urable space. In other words, as we want to characterize the system by the simultaneous description of the different observables — i.e., physical quantities — we must have an event space common to all of them. This common event space, then, represents the system by representing the states that determine the induced distributions.

As it is well known in probability theory, the most general event space in which a probability measure can be defined — i.e., in which the events are all compatible with each other — is the Boolean σ-algebra. Accordingly, Boolean σ-algebra can be an adequate structure for the common space outlined above *if all combinations*

of simultaneous events are physically possible in the system. For example, in this case any set of predetermined values $f_1, f_2, \ldots, f_n$ for the observables $\varphi_1, \varphi_2, \ldots, \varphi_n$ can be given as an outcome of a measurement. The structure of the event space in classical physics — involving statistical mechanics and such extensions as e.g. the present models of neural systems — is therefore Boolean σ-algebra but in quantum mechanics where, in fact, it is not possible for certain events to occur simultaneously, another structure must be chosen.

In quantum mechanics the subspace lattice of the Hilbert space H(of infinite dimensions) is used as a common event space. Thus the mapping φ^{-1} operates from the intervals of real numbers — more generally, the subsets of the set of Borel sets of real numbers — to the subspaces of H:

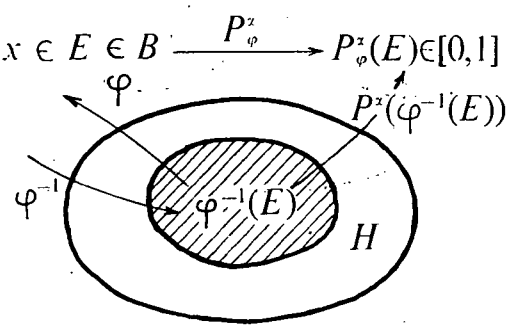


Fig. 3

$$\varphi^{-1}: B \to H \tag{2}$$

where $\varphi^{-1}(E) = \{\varphi^{-1}(x) | x \in E\}$, and x is a real number (that is taken by the random variable φ at an element of H) (Fig. 3).

It is obvious that the mapping

$$\varphi^{-1}: B \to H$$

is a homomorphism.

Now we can construct a one-to-one correspondence I between the subspaces of H and the operators P projecting to these subspaces. Then the mapping

$$I \circ \varphi^{-1}: B \to P \tag{3}$$

operating from the intervals of real numbers to the set of projection operators will be a so-called projection measure. As it is known, each projection measure is equivalent to a self-adjoint operator; the theory of self-adjoint operators (Neumann, 1932; Araki and Yanase, 1960) then provides us with the formalism adequate for deducing all the consequences essential in quantum mechanics.

The operators corresponding to the observables in classical physics are all commutative. Mathematically this follows from the fact that the common event space in this case is a distributive lattice; in quantum mechanics, on the contrary, the subspace lattice of H is not distributive, thus the operators do not always commute. From the point of view of measurement the non-commutativity involves the existence of observables whose values are not measurable simultaneously. Therefore, if we want to decide whether a system can be described by means of a formalism of classical type or not, we examine the physical possibilities for simultaneous determination of any set of values of the various observables.

The aim of this paper is to suggest and outline a model of abstract neural objects in which the common event space is a non-distributive lattice; i.e., the formalism of treatment is analogous to that of quantum mechanics. Apart from the mathematical construction described above one can have another, more general, possibility to approach the problem that in which cases an essentially probabilistic view of a given system is necessary.

Measurement and complexity

It is trivial that any process by which information can be obtained — i.e., any measurement — involves an interaction between at least two systems, say A and B. A and B are connected together in such a way that as a result of the measuring interaction the states $S_1^A, S_2^A, \ldots, S_n^A$ of A will inevitably correspond to some states $S_1^B, S_2^B, \ldots, S_n^B$ of B and vice versa. Thus, if $O^{(A)}$ is an observable belonging to A and $O^{(B)}$ is another belonging to B, any given value of $O^{(A)}$ (determined by the state of A) corresponds to a given value of $O^{(B)}$ (determined by the corresponding state of B). Let now A be the system to be measured and B the measuring one. Assume that both A and B have more or less complex structure. The notion of complexity does not need a strict definition here; it is enough to consider that the more complex a system is, the more complex the changes of its states and the parameters describing these changes will be. In the case of the measuring interaction between A and B there are two basic possibilities.

If A and B have equal complexity or B is more complex than A then the state changes in B can reflect in an adequate way the state changes in A. If, however, B is less complex than A, B does not have a large enough number of states for this purpose and in this case the measurement can lead to only a probabilistic description of A via the parameters of the state changes of B.

In practice, there is possible an important compromise. Namely, if though the system A is the more complex one, but it does not take part, as a whole, in the interaction, then the description may be dynamic. The necessary condition for this is that the part of A interacting with the measuring system B should not be more complex than the totality of B. H is possible only in this case — that all observables describing A are measurable simultaneously.

The quantum-mechanical concept of measurement is, therefore, the adequate tool for studying systems exhibiting non-negligible complexity in their interactions. In our opinion the nervous system does have this property. The application of the theory of measurement for this branch of biology is possible because the interactions between the nervous system and its environment, or between the parts of the nervous system itself, can be viewed as measuring processes (Jólesz and Gyöngy, 1975). In addition, any measurement has an aspect regarding to information, as the result of the measurement appears as a given state of the system (or its parameters), and the same is true for the new information obtained by the measurement. If we observe the measuring processes in the nervous system are of statistical nature we accordingly tend to discard the dynamic principles that would uniquely determine all details of the interactions occuring in the system.

A simple model of the neural measuring process

The simple abstract model of the neuron, discussed here, is somewhat similar to the so-called formal neuron, but a few of its properties are essentially different. As the mathematical concepts and procedures are all well known from quantum mechanics, for the sake of conciseness we will confine the treatment to a brief outline.

The basic postulates are as follows:

1. The system is characterized by a wave function Φ; the state of the system is fully determined by Φ. Φ is an element of the Hilbert space H.

2. The properties of the system are described by giving the possible values (the so-called eigenvalues) of the observables and by associating with each of them one or more state-functions in the Hilbert space, termed eigenfunctions. In addition, each wave function Φ can be expanded as a linear combination of the eigenfunctions of any observable. Thus, the observables $O_1, O_2, \ldots, O_n, \ldots$ are characterized by the appropriate sequences of real numbers:

$$O_1\colon k_1^{(1)}, k_2^{(1)}, \ldots, k_r^{(1)}, \ldots$$

$$O_2\colon k_1^{(2)}, k_2^{(2)}, \ldots, k_r^{(2)}, \ldots$$

$$\ldots$$

To each observable a set of probability values is assigned:

$$O_1\colon W(k_1^{(1)}, \Phi),\ W(k_2^{(1)}, \Phi), \ldots, W(k_r^{(1)}, \Phi), \ldots$$

$$O_2\colon W(k_1^{(2)}, \Phi),\ W(k_2^{(2)}, \Phi), \ldots, W(k_r^{(2)}, \Phi), \ldots$$

$$\ldots$$

where the probabilities W depend also on Φ.

The wave functions are elements of the space H and can be demonstrated as vectors in the Euclidean space of infinite dimensions. The observables are operators in the space of functions or matrices in the Euclidean space. Matrices and operators are both linear mappings of vector spaces.

3. The probability of that any given observable O takes on a value from the given interval (k', k'') can be determined in any state. (See Fig. 4.)

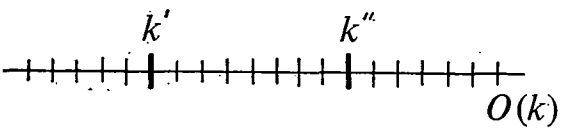


Fig. 4

4. The probability of that the values of two observables O_1 and O_2 falls simultaneously into the intervals (k', k'') and (l', l''), respectively, can be computed for some observables but can not be done so for others. The operators of these latter are not commutable:

$$O_1 O_2 \neq O_2 O_1 \tag{4}$$

or

$$O_1 O_2 - O_2 O_1 = cI \tag{5}$$

where c is an imaginary number and I is the unit operator or matrix.

In the model we assume that from the point of view of the measurement only two observables are relevant: *the stimulation intensity I* and *the time length T during which the stimulation operates*. In other words we assume that the stimulation has the only physical properties of intensity and length. As for both these observables the number of possible values is infinitely large we need some further simplifications.

In accordance with the existence of intensity and length thresholds — i_0 and t_0 — in real neurons suppose that both these observables can take on only two values each (Fig. 5).

To be more clear, let the states be represented by vectors of unit length on the plane. Let φ_1 and φ_2 be the state vectors associated with the two possible values

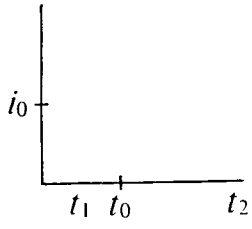


Fig. 5
Where i_1 and t_1 denote values below, while i_2 and t_2 above the threshold

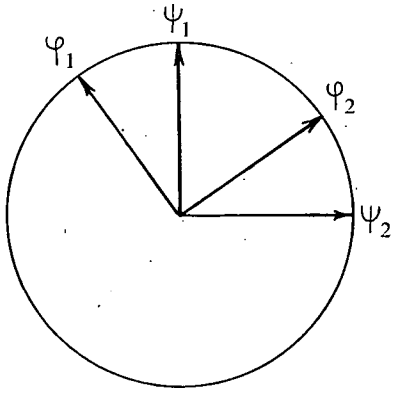


Fig. 6

i_1 and i_2 of the observable I and ψ_1 and ψ_2 the state vectors similarly for the observable T (Fig. 6).

It may be worth noting that all the eigenfunctions are mutually orthogonal, i.e., the scalar product of all pairs of them is equal to zero.

Now we have the symbols necessary for working with the model: the stimulus intensity I and length T as quantities to be measured, the possible values i_1, i_2 and t_1, t_2 of them, being equal either 0 or 1:

$$I\begin{cases} i_1 = 0\,(\varphi_1) \\ i_2 = 1\,(\varphi_2) \end{cases} \quad T\begin{cases} t_1 = 0\,(\psi_1) \\ t_2 = 1\,(\psi_2) \end{cases}$$

Operators, probabilities, expectations and variances

Let the projection of an arbitrary state Φ to the direction of the eigenstates φ_1 and φ_2 be a_1 and a_2, respectively, and to the direction of ψ_1 and ψ_2 be b_1 and b_2 (Fig. 7).

Thus, Φ can be written as

$$\Phi = a_1\varphi_1 + a_2\varphi_2 = b_1\psi_1 + b_2\psi_2.$$

The expansion of Φ as a linear combination of the eigenstates of I (or T) is called the I (or T) representation of Φ.

The operators projecting to a given state σ will be denoted by P_σ. Thus applying P_{φ_1} to Φ we get

$$P_{\varphi_1}\Phi = a_1\varphi_1.$$

It is obvious that projecting φ_1 to itself it remains unchanged

$$P_{\varphi_1}\varphi_1 = \varphi_1,$$

and projecting to a direction orthogonal to it the result will be zero

$$P_{\varphi_2}\varphi_1 = 0.$$

In the simple model presented here we consider only two operators A and B representing the observables I and T, respectively. As it was assumed, I and T can take on two values each and with every one of these values a corresponding state is associated. Note that in quantum mechanics this kind of characterization of the states and observables is quite general but the number of states is usually infinite. Therefore, the operators representing I and T can be written as

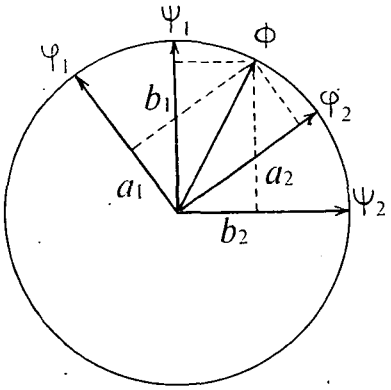


Fig. 7

$$I \to A = i_1 P_{\varphi_1} + i_2 P_{\varphi}$$

and

$$T \to B = t_1 P_{\psi_1} + t_2 P_{\psi_2}$$

If we want to compute the probability of obtaining the various values of the observables in a given state Φ we should multiply by the proper projection operators and then take the scalar product of the result with itself:

$$\text{Probability } \{\text{The measurement yields the value } k\} = (P_\varkappa \Phi, P_\varkappa \Phi) \tag{6}$$

where $\varkappa$ is the eigenstate associated with k.

Applying the above procedure to i_1, i_2, t_1 and t_2 the probabilities essential in our model are

$$\begin{aligned} W(i_1, \Phi) &= (P_{\varphi_1}\Phi, P_{\varphi_1}\Phi) = a_1^2, \\ W(i_2, \Phi) &= (P_{\varphi_2}\Phi, P_{\varphi_2}\Phi) = a_2^2, \\ W(t_1, \Phi) &= (P_{\psi_1}\Phi, P_{\psi_1}\Phi) = b_1^2, \\ W(t_2, \Phi) &= (P_{\psi_2}\Phi, P_{\psi_2}\Phi) = b_2^2. \end{aligned} \tag{7}$$

Note that

$$a_1^2 + a_2^2 = 1 \tag{8}$$

and

$$b_1^2 + b_2^2 = 1 \tag{9}$$

as in both cases one of the two possible values will certainly be obtained with probability 1.

As two important special cases, it is apparent that obtaining the value i_2 in the state φ_1 is impossible

$$W(i_2, \varphi_1) = (P_{\varphi_2}\varphi_1, P_{\varphi_2}\varphi_1) = 0, \tag{10}$$

and obtaining the value i_1 in the state φ_1 is certain

$$W(i_1, \varphi_1) = (P_{\varphi_1}\varphi_1, P_{\varphi_1}\varphi_1) = 1, \tag{11}$$

In general, if the state of the system coincides with one of the eigenstates of a given operator, the observable corresponding to that operator has a uniquely determined value by the measurement; and in any other state the probabilities of obtaining any permissible value of that observable can be computed in the way outlined above. If, however, we want to determine the probability of that in an arbitrary state Φ the observable I takes on the value i_1 *and* the observable T takes on the value t_1 we may easily get to a confusing result. In this case, Φ should first be projected to φ_1 belonging to i_1

$$P_{\varphi_1}\Phi = a_1\varphi_1, \tag{12}$$

then the result of this projection should be projected to ψ_1 belonging to t_1

$$P_{\psi_1}(P_{\varphi_1}\Phi) = P_{\psi_1}(a_1\varphi_1) = a_1 P_{\psi_1}\varphi_1. \tag{13}$$

The sought probability according to (6) is

$$(a_1 P_{\psi_1}\varphi_1, a_1 P_{\psi_1}\varphi_1) = a_1^2(P_{\psi_1}\varphi_1, P_{\psi_1}\varphi_1) = a_1^2(\varphi_1\psi_1)^2. \tag{14}$$

But, if we follow the reverse order of this procedure, namely projecting first to ψ_1 and then to φ_1

$$P_{\psi_1}\Phi = b_1\psi_1, \tag{15}$$

$$P_{\varphi_1}(P_{\psi_1}\Phi) = P_{\varphi_1}(b_1\psi_1) = b_1 P_{\varphi_1}\psi_1, \tag{16}$$

$$(b_1 P_{\varphi_1}\psi_1, b_1 P_{\varphi_1}\psi_1) = b_1^2(P_{\varphi_1}\psi_1, P_{\varphi_1}\psi_1) = b_1^2(\psi_1, \varphi_1)^2. \tag{17}$$

It is trivial that the two computed values for the same probability are not equal:

$$a_1^2(\varphi_1, \psi_1)^2 \neq b_1^2(\psi_1, \varphi_1)^2 \tag{18}$$

as because of the result of the scalar product being independent of the order of the factors

$$(\varphi_1, \psi_1) = (\psi_1, \varphi_1) \tag{19}$$

and thus

$$a_1^2 \neq b_1^2 \tag{20}$$

causes the inequality to hold.

Accordingly, it is not possible to make a unique assertion about the probability of obtaining simultaneous values for I and T. The order of applying the operators corresponding to I and T to a given state function Φ is not commutable:

$$AB\Phi \neq BA\Phi \tag{21}$$

or

$$AB\Phi - BA\Phi = C\Phi \tag{22}$$

where $C \neq 0$ is called commutator. C is independent of the state, i.e., of the wave function Φ.

In the model it is of paramount importance to compute the expectation value and the variance of the stimulus intensity and length. The expectations in a given state Φ are

$$\bar{I} = i_1 a_1^2 + i_2 a_2^2 = i_1(P_{\varphi_1}\Phi, P_{\varphi_1}\Phi) + i_2(P_{\varphi_2}\Phi, P_{\varphi_2}\Phi) =$$

$$= i_1(P_{\varphi_1}\Phi, \Phi) + i_2(P_{\varphi_2}\Phi, \Phi) = ([i_1 P_{\varphi_1} + i_2 P_{\varphi_2}]\Phi, \Phi) = (A\Phi, \Phi) \tag{23}$$

and

$$\bar{T} = t_1 b_1^2 + t_2 b_2^2 = (B\Phi, \Phi). \tag{24}$$

The expectation value thus can be given without knowing the analytic form of the wave function and the operator representing the observable to be measured:

$$\bar{A} = (A\Phi, \Phi) \tag{25}$$

$$\bar{B} = (B\Phi, \Phi) \tag{26}$$

In a quite similar way we get the variances:

$$(\Delta I)^2 = ([A - \bar{A}\mathbf{I}]^2 \Phi, \Phi) \tag{27}$$

$$(\Delta T)^2 = ([B - B\mathbf{I}]^2 \Phi, \Phi) \tag{28}$$

where **I** is the unit operator.

Appearance of the uncertainty relations

In the light of the above considerations a question arises concerning the meaning of the simultaneous measurement of the stimulus intensity and time length. It is obvious that the neuron can be regarded as a physiological device for measuring the intensity and length of various stimuli and not less obviously this measuring process relates to some threshold conditions. In our simple model what consequences can be drawn if, as we have just seen, there is an inherent ambiguity in the process of simultaneous determination of the probabilities for I and T?

If the measurement of one of the observables, e.g. the stimulus intensity, can yield two permitted values i_1 and i_2 according to the existence of a threshold, the states φ_1 and φ_2 associated with the eigenvalues i_1 and i_2 respectively assign the probability 1 to the corresponding i_1 or i_2 values of the observable I. If by the same measurement the neuron does determine the length of the stimulus, too, the measurement yields either the value t_1 or t_2 for the observable T and, consequently, we can be sure that the system was either in the state ψ_1 or ψ_2. Thus, in a simultaneous measurement one of the eigenstates of I would be the same as one of the eigenstates of T. This involves the commutativity of the operators belonging to I and T. In the model suggested here, however, like in quantum mechanics, neither φ_1 or φ_2 is equal to ψ_1 or ψ_2. The contradiction disappears only if we accept that in the measuring process of the neuron there are measurable quantities whose statistics can not be correlated with each other. In other words, as there are no common eigenstates of the intensity and length of the stimulus, these two observables can not be measured simultaneously however obey them, separately, quite well-defined probabilistic laws.

The variances of the observables relevant in our model have a relationship to each other similar to that between the variances of canonical conjugate variables

described by the Heisenberg uncertainty relation. This relation expresses, in fact, that the two observables have no common eigenstates.

The farer is the state of a neural object from the eigenstate, for a given observable, the more uncertain the value of that observable; the variance is zero only in an eigenstate. If the equation (21)—(22) holds true, the variances of the two quantities in that equation can not be zero simultaneously. Of course, this means that the simultaneous measurement of these quantities can not, even theoretically, be arbitrarily accurate. We can ask only to what extent the $(\Delta A)^2$ and $(\Delta B)^2$ variances can be simultaneously lowered.

To see this let us introduce two auxiliary operators:

$$A' = A - \bar{A}\mathbf{I}$$

and

$$B' = A - \bar{B}\mathbf{I} \tag{29}$$

As it is obvious the commutation relations remain true for A' and B'

$$A'B' - B'A' = C.$$

The variances

$$\begin{aligned}(\Delta A)^2 &= (\Phi, A'^2\Phi) = (A'\Phi, A'\Phi)\\ (\Delta B)^2 &= (\Phi, B'^2\Phi) = (B'\Phi, B'\Phi).\end{aligned} \tag{30}$$

Now let

$$\begin{aligned}f &= A'\Phi\\ g &= B'\Phi.\end{aligned} \tag{31}$$

As f and g are quadratically integrable functions, according to the Schwartz inequality

$$|(f, g)|^2 \leqq \|f\| \cdot \|g\|. \tag{32}$$

Substituting from (30) into (32):

$$(\Delta A)^2(\Delta B)^2 \geqq |(A'\Phi, B'\Phi)|^2 = |(\Phi, A'B'\Phi)|^2 \tag{33}$$

With some trivial transcriptions:

$$\Delta A \Delta B \geqq \frac{1}{2}|\bar{C}| \tag{34}$$

This inequality characterizes the correlation between the uncertainties in the measured value of observables corresponding to non-commutable operators. The choice of the state Φ of the model neuron does not affect the validity of (34).

For the measurements of stimulus intensity and length

$$\Delta I \Delta T \geqq C. \tag{35}$$

This relationship is analogous to the Heisenberg uncertainty relation with the important difference of C being not a universal constant but only one independent of the state. (35), (27), (28) assert that in the proposed model the accurate simultaneous measurement of the stimulus intensity and length has an absolute (theoretical) limitation (Jólesz and Szilágyi, 1974.). This does not mean that the neuron could not

measure either the intensity or the length of a given stimulus; it states only that it can not measure them simultaneously with an arbitrary precision. In other words, if the state function Φ is not an eigenvector of a given operator A, by the measurement of the observable corresponding to A we can be given several different results and the farer is the state of the neuron from an eigenstate of A, the more uncertain the value of the given observable.

To sum up the process of measurement by the model neuron, the most important points, in our opinion, are the following. The neuron measures the property I or T of the stimulus, or rather their thresholds; in the process of this measurement its original state Φ turns into φ_1 or φ_2 (associated with the numbers i_1 and i_2 characterizing the value of I being above or below the threshold) in the case of measuring I, and into ψ_1 or ψ_2 (associated with t_1 or t_2) in the case of measuring T:

$$\Phi \to \begin{matrix}\varphi_1 \\ \varphi_2\end{matrix} \quad \Phi \to \begin{matrix}\psi_1 \\ \psi_2\end{matrix} \tag{36}$$

This state change is called the measuring process. It corresponds to the projection process of the state vector Φ onto the direction of one of the eigenvectors. In the measuring process the object being in the state Φ turns into one of the eigenstates φ_j of the operator A of the measured variable (I or T). The process itself does not require any description; what is relevant are only the probabilities of the occurrence of the various possible φ_j final states.

Denoting the projection operator projecting onto the direction of φ_j by P_{φ_j} we obtained that the probability of the

$$\Phi \to P_{\varphi_j}\Phi = (\Phi, \varphi_j)\varphi_j \tag{37}$$

transition is $|(\Phi, \varphi_j)|^2$ as a result of the measuring process.

As the state vectors are normalized to unity the multiplicative coefficient of φ_j can be eliminated by normalization. Thus the original state Φ becomes completely vanished from the expression of the final state. The original state takes part only in the expression of the transition probability. The measuring process has a representative only in the set of the projection operators.

Apart from the measuring process — which is in some sense a singular one — another process is existing in the model: the spontaneous change of state of the undisturbed system. This process can be described by a continuous rotation:

$$\Phi \to U(\alpha)\Phi \tag{38}$$

where $U(\alpha)$ a unitary operator with the rotation parameter α.

It is obvious that both processes outlined are well identifyable in the case of our model and of real neurons as well. In contrast to the continuous transition in spontaneous processes, the measuring process represents a discrete change of state.

Finally, it may be worth noting that it is possible to draw conclusions about the state before a measurement. One can do so by the measurement itself, because the probability distribution of the measured spectrum reflects just the distribution of the possible states before the measurement. This latter but characterizes a real state for, in the case of neurons as analogous systems to micro-objects in quantum mechanics, it is the probability distribution of the possible eigenvalues which contains the whole information about any given state independently of the measurement.

An extension of the model

As we have stated before, the neuron can be viewed as a device for measuring the stimulus intensity and length. The neuron can have, however, not only one threshold for the measured observables, but many different ones for each. It involves the need for extending the previous simple model to be able to treat a series of eigenvalues $\{i_1, i_2, \ldots, i_n, \ldots\}$ and $\{t_1, t_2, \ldots, t_n, \ldots\}$.

Consider the points k_i on the real line permitted for a given observable R to take on as measured values. Thresholds can be taken into account by leaving out certain points from the set of the possibilities; mathematically this is done by means of a projection operator $E(k)$ which increases in the permitted points and remains constant elsewhere. The whole set of the permitted values will be called the spectrum of R.

It is obvious that operators having discrete series of eigenvalues are adequate for the description of observables whose permitted values constitute also a discrete series. Thus, this kind of operators can be used in a neuron model with a number of discrete thresholds.

For theoretical derivation of the possible values of a given observable — e.g. thresholds of intensity — we need to know the operator corresponding to that observable. It is enough to determine the two observables characterizing the stimulus because all the other observables — e.g. which relates to the speed of the stimulus intensity change in time — and consequently their operators can be deduced from these.

The formal neuron (McCulloch and Pitts, 1943) and the variations of it can also be regarded as devices for measuring the stimulus intensity (Lábos, 1975). In some experiments (Lábos, 1973; Sclabassi, Lábos et al., 1973) neurons have yielded response characteristics the analysis of which by means of model frequency code points towards the idea of the neuron with more than one threshold. A similar system of thresholds can be obtained from the Hodgkin—Huxley model (successive current thresholds, Lieberstein, 1973) as well.

According to Lábos (1975) any neuron having response characteristic with generalized distribution function can be regarded as a measuring device. The response characteristics which refer to more than one threshold generate discrete Lebesgue—Stieltjes measurable spaces. In addition, Lábos stated that neurons have various different sets of thresholds depending on the length of the stimulus: the shorter the stimulus in time, the fewer levels of intensity can be distinguished. As it was mentioned before, the neuron should be represented by not only some thresholds in intensity but in length as well. Therefore, the outlined model should be extended.

In the most general case there must be a solution both for discrete and continuous eigenvalues. Thus, let i be an arbitrary real number (being one of the eigenvalues of the stimulus intensity as an observable) and $E(i)$ an operator with the argument i. $E(i)$ is a generalization of projection operators of the simple model, projecting to different subspaces depending on the value i. Let f and g be two arbitrary elements of the domain of $E(i)$. Taking the inner product

$$(f, E(i)g) = (E(i)f, g) \tag{39}$$

if the following Lebesgue—Stieltjes integral exists:

$$\int_{-\infty}^{\infty} id\,(f, E(i)g) \tag{40}$$

where

$$E(-\infty) = 0, \quad E(+\infty) = \mathbf{I} \tag{41}$$

$\mathbf{I}$ being the unit operator, then an operator A can be constructed:

$$(f, Ag) = \int id(f, E(i)g) \tag{42}$$

$$A = \int idE(i) \tag{43}$$

$E(i)$ is called the spectral decomposition of the operator A. In the case of discrete spectrum $E(i)$ depends on i in the following way

$$E(i) = \sum_{i_n \leqq 1} P_n. \tag{44}$$

The intervals where $(f, E(i)g)$ is constant may be excluded from the domain of integration. These values do not belong to the spectrum of A; to this spectrum do belong only the values i whose corresponding product $(f, E(i)g)$ is changing. Where the change is continuous, so is the spectrum, while the points where there is an abrupt change in $(f, E(i)g)$ constitute the point spectrum of A. In this way the existence of thresholds may be taken into consideration. There can be no projection operator attributed to the isolated points in the domain of any continuous spectrum. On the contrary, to the interval (i', i'') the following projection operator belongs

$$E = E(i'') - E(i'). \tag{45}$$

The commutation relations for the operators are closely related to those for their spectral decompositions. Let

$$A = \int idE(i) \tag{46}$$

and

$$B = \int tdF(t). \tag{47}$$

If

$$E(i)F(t) = F(t)E(i) \tag{48}$$

then

$$AB = BA. \tag{49}$$

It would hold true only, however, if both A and B were functions of the same operator but, according to our assumptions, now this is not the case.

In our opinion the extended neuron model is characterizable by the set of thresholds of stimulus intensity and length. The operators associated with these observables are in one-to-one correspondence with some subspaces of H and hence the equality $EF = FE$ would be valid only if EF and FE projected onto the same subspace. As but the system of eigenvectors and so the subspaces in question are not common the operators do not commute.

According to the outlined model, neural objects are represented by the Hilbert space; events are represented by its subspaces or the projection operators being in one-to-one correspondence with the subspaces.

The relevance of measurement should be stressed in particular. The outcome of the measurement is affected by chance; the probability of any given transition from a state before to another after the measurement depends on the beginning state and the measured observable. The beginning state Φ of the object is an element of the Hilbert space representing the object in question. The resulting state φ_n is always an eigenstate of the operator of the measured observable. The measured value of the observable is the i_n eigenvalue (in the case of intensity measurement) belonging to the eigenstate φ_n. The distribution function determined by the transition probabilities, i.e., the probability of that the value i_n of I is not greater than a given value i is as follows

$$W(i_n \leqq i\,|\Phi) = (\Phi, E(i)\,\Phi) \tag{50}$$

or, with some trivial transcriptions

$$W(i_n \leqq i|\Phi) = (f, E(i)\,\Phi) = (\Phi, E^2(i)\,\Phi) = (E(i)\,\Phi, E(i)\,\Phi) = \|E(i)\,\Phi\|^2. \tag{51}$$

Knowing the probability distribution, the expected value is easily computed

$$\bar{A} = \int id(\Phi, E(i)\Phi). \tag{52}$$

The scatter

$$\overline{(A-\bar{A}\mathbf{I})^2} = (\Phi, (A-\bar{A}\mathbf{I})^2\,\Phi) = |(A-\bar{A}\mathbf{I})\,\Phi, (A-\bar{A}\mathbf{I})\,\Phi| = \|(A-\bar{A}\mathbf{I})\,\Phi\|^2. \tag{53}$$

As regards to the measuring process, the following are important. In the case of the measuring interaction the object being in the state Φ turns into another state φ_n. φ_n is one of the eigenvectors of the operator of the measured observable, i.e., I, where I can be expanded by the projection operators P_n:

$$\begin{gathered} A = \sum i_n P_n = \int id(E(i)) \\ P_n\varphi_n = \varphi_n, \quad A\varphi_n = i_n\varphi_n \\ E(i) = \sum_{i_n \leqq i} P_n \end{gathered} \tag{54}$$

The distribution function of the transition probabilities by transcribing (51)

$$W(i_n \leqq i|\Phi) = \|E(i)\Phi\|^2 = \sum_{i_n \leqq i} |(\varphi_n, \Phi)|^2 \tag{55}$$

where $|(\varphi_n, \Phi)|^2$ is the probability of the transition $\Phi \to \varphi_n$.

The measurement yields a correct value for I if φ_n uniquely determines i_n and vice versa. The eigenvalue i_n is really the correct value, because in the state φ_n the deviation for the operator I is zero:

$$\begin{aligned} (\varphi_n, [A-\bar{A}\mathbf{I}]^2\varphi_n) &= ([A-\bar{A}\mathbf{I}]\varphi_n, [A-\bar{A}\mathbf{I}]\varphi_n) = \\ &= \|(A-\bar{A}\mathbf{I})\varphi_n\|^2 = \|A\varphi_n - \bar{A}\varphi_n\|^2 = 0 \end{aligned} \tag{56}$$

Another case occurs if the observable to be measured has a continuous spectrum (i.e. set of eigenvalues). Let the spectral decomposition of I be $E(i)$. A measurement with given correctness means that the outcome i falls into a given interval (i', i''). The probability of this is

$$W(i' < i \leqq i''|\Phi) = (\Phi, E(i)\,\Phi) \tag{57}$$

where Φ is the state before the measurement. The length of the interval (i', i'') may be arbitrarily small the value of $W(i' < i \leqq i'' | \Phi)$ remains finite; it follows then that the probability of the case of absolute correct measurement is zero

$$W(i' < i \leqq i'' \,|\, \Phi) = 0. \tag{58}$$

Consequently, in the continuous range of the spectrum (according to the model suggested) there exist measurements only with non-zero uncertainty.

Probabilistic interpretation of the neural measuring process

Theoretical interpretations regarding to the operation of the nervous system — in spite of that they contradict to each other in some respects — have the common feature of accepting (at least at present) the probabilistic nature of the neural processes. In this respect the opinions are diverging in whether this nature is the same as that of other disciplines in physics (classical physics, thermodynamics, quantum mechanics) or is inherently different.

In our model the spontaneous and the measuring process may be in close connection with the probabilistic interpretation. The measuring process influences the state of the system, so obtaining information is connected to the state change. When we state that the new information, i.e., the result of the measurement, is reflected in the new state of the system, we lay stress on the statistical meaning of the state. Namely, while during the spontaneous process the state transition $\Phi \to \Phi'$ is not statistic (the system turns from $U = P_\Phi$ into $U = P_{\Phi'}$ in a continuous way), the measuring process causes the state Φ to transform into one of the eigenstates $\varphi_1, \varphi_2, \ldots$, this transformation being only stochastically determined: the probabilities $|(\Phi, \varphi_1)|$, $|(\Phi, \varphi_2)|, \ldots$ of the states $\varphi_1, \varphi_2, \ldots$ are uniquely determined and not so is the final state itself. During the measuring process the states turn into mixed form

$$U = P_\Phi \to U' = \sum_{n=1}^{\infty} |(\Phi, \varphi_n)|^2 P_{\varphi_n}. \tag{59}$$

In the language of information theory the measurement is a kind of mixing processes hence it is necessarily irreversible.

The basic difference between the spontaneous and measuring processes is that while in the time interval between two measurements the variation of the state vector is determined and continuous, the variation owing to the measurement is sudden and discontinuous. This latter can be described only by probability laws.

After the measurement the state of the system is a compound consisting of the eigenstates of the operator of the measured observable. All these statements are of importance if one considers a measurement taken by the neuron: from the point of view of neural networks, the outcome of the measurement means that a group of neurons is not in a homogeneous state but its members have different states with different probabilities. Any combination of these states can be a measurement outcome if the measuring device is the given group. After the measurement the group can yield only probabilistic relationships.

In the field of theoretical neurobiology relatively large area is occupied by statistical mechanics (Wiener, 1958; Cowan, 1968; 1970; Amari, 1974). Regarding

to the origin of the probabilistic laws applied, the analysis may show different levels of deepness (Griffith, 1971). In the application of the theory of random processes to macroscopic neural networks one can disregard even the existence of the network structure. It causes, then, the treatment to confine itself to statistical fluctuations. However, there are a number of theories of considerable efficiency by the utilization of probabilistic concepts of neural processes.

In addition to the use of the methods of statistical mechanics some examples of the use of quantum statistics can also be found (Winograd and Cowan, 1963; Cowan, 1965; Agin, 1963; Michalov, 1967, 1968). In our opinion this way is very promising. By means of the formalism of quantum mechanics essential features of the neural measuring process may become known. Moreover, we can extend the borders of the probabilistic interpretation by taking into account that only the statistics of the observables are really "observable". The connection between the state functions and the observables has an inherent statistical nature because, in the model suggested here, in the case of a system with k degrees of freedom the states are characterized by a function $\Phi(g_1, g_2, \ldots, g_k)$ which is an element of the Hilbert space; therefore, with even a full giving of Φ one can make only statistical assertions about the system. (Obviously it is irrelevant that the probability of the truth of these assertions lies sometimes near 0 or 1).

Characterization of the state of a neural micro-object by a state function yields the possibility of making probabilistic statements, but the validity of the statements can be checked only on some groups of micro-objects, i.e., on neural populations. This means also that this formalism expects immediately the measurement to be made by neuron populations or, equivalently, it expects the recording of the statistics of the measured observables. In classical statistical mechanics the question concerning the probability of finding a given neuron from the population in a given state can always be asked and answered as well. On the contrary, the probabilistic expressions in the formalism of quantum theory give possibility only of answering the question about the probabilities of a given value to fall into the interval (i', i'') or the interval (t', t'') separately. There is no probability measure common to both intervals because there is no common state in which both probabilities can be measured.

Assuming that populations of neurons are generally not in pure state (i.e., all neurons are not in the same state) we should consider mixtures. If a measurement is made on a system in mixed state it forces the system to turn into an eigenstate. The sudden change of the state function at the moment of the measurement can be described by means only of probabilistic relationships. In the theory of measurement (Neumann, 1932) it is generally assumed that measuring an observable on a single object R is not all that possible and the measurement even should be made on a system of very many objects. With the attitude of measurement theory the large dimensionality of the nervous system and also the concept of redundancy can gain a new interpretation. Namely, according to this attitude, it is advisable to make measurements on large statistical groups consisting of a number of micro-objects $R_1, R_2, \ldots, R_N$, where N is a large number. On such a group the distribution of the values of the measured observable is determined. The advantage of this procedure is that though the measurement disturbs the object on which it was made, the disturbance of the population as a whole may be arbitrarily small if N is large enough. Furthermore, though two observables having non-commuting operators can not

be measured simultaneously to any degree of accuracy, in the population their probability distributions can be determined with an arbitrarily small error. It is enough to measure a part of the whole system, if the number of elements M of that part is large, i.e. $M \gg 1$, but it is much smaller than $N (M \ll N)$. In this case the measurement affects only the M/N part of the total system. Measuring another observable on another part, made up by K elements, of the system, the two measurements do not interfere if $(K+M) \ll N$ and $(K+M)/N \ll 1$. These requirements can easily be fulfilled if N is large enough; in this case K and M may also be large.

In the nervous system by means of statistically large populations of neurons there is a possibility of objective measurements being independent of occasional disturbations and of that any single neural object is unable to make simultaneous accurate measurements of two non-commuting observables. However, as it will soon be demonstrated, the measurement can not be absolutely accurate even in this case.

Consider, for example, the simpler one of the models described, in which the observable I can take on the values i_1 and i_2 only. Let us measure I on a population $\{R_1, R_2, \ldots, R_n\}$; then we get i_1 as a result at a part $\{R'_1, \ldots, R'_{N1}\}$ of the population and i_2 at another part $\{R''_1, \ldots, R''_{N2}\}$. As a consequence of the measuring process, however, the neuron states will change in both parts of the population and, for this reason, if we measure another observable T on the same population (with possible values t_1 and t_2) it is no more possible to obtain scatterless results because the eigenstates of I can not be eigenstates of T. Accordingly, simultaneous measurements can not produce pure populations.

It is generally stated that probability is a property of certain classes of populations. This permits, however, to apply probability calculus to some individual (e.g. neural) processes, if we know that probability calculus is applicable if the process at hand leads to statistical populations. In the study of neural processes the reason for probability to play role is twofold. On the one hand, in the starting mixed state of the system certain properties — e.g. intensity and length thresholds — have statistical distributions and this causes a statistical distribution of the measured values as well. On the other hand, during the measurement the nervous system passes through a series of interactions which gives rise to statistical distributions of certain parameters. In both cases it is plausible that the individual objects making up the population (i.e. the neurons) obey some dynamic laws different from the statistical ones. Therefore, the concept of the so-called measuring interaction requires an adequate framework to be placed in, in order to be distinguishable from another types of interactions.

Acknowledgement

The authors have indebted to their colleagues István Gyöngy and Zoltán Vassy for their worthy advices and discussions.

Summary

The purpose of the present paper is to investigate some properties of measuring processes performed by single nerve cells and neural nets. We applied the formalism of quantum mechanics and the quantum-mechanical concept of measurement. We used the subspace lattice of the Hilbert space as a common event space. Two neuron models were analyzed in which we assumed that only

two observables (stimulus intensity and time) are relevant. In these models we considered only two operators representing the observables.

The first model was characterizable by one threshold, the second one by the set of thresholds. In the proposed models the simultaneous measurement of the observables has an absolute limitation and the variances of them have a relationship which is analogous to the Heisenberg uncertainty relation, with the important difference of C being not a universal constant. Statistical properties of the neuronal measuring processes were examined.

The mathematical methods for dealing with neuronal systems that we have described in this paper seem to have many advantages over the methods usually used. There is a strong analog between these methods and the techniques generally used for physical systems. Although we have limited ourselves in this paper to two neuron models, the presented new method is generalizable to neuronpopulations that are composed of elements with n thresholds and in which not only two observables are measured.

*NATIONAL INSTITUTE OF NEUROSURGERY
BUDAPEST, HUNGARY

**"KANDÓ KÁLMÁN" COLLEGE OF ELECTRICAL ENGINEERING
BUDAPEST, HUNGARY

Literature

AGIN, D., Some comments on the Hodgkin—Huxley equations, *J. Theoret. Biol.*, v. 5, 1963, p. 161.
AMARI, S., A method of statistical neurodynamics, *Kybernetik*, v. 14, 1974, p. 201.
ARAKI, G., M. YANASE, Measurement of quantum mechanical operators, *Phys. Rev.*, v. 120, 1960, p. 622.
COWAN, J. D., The problem of organismic reliability, *Progr. Brain Res.*, v. 17, 1965, p. 9.
COWAN J. D., Statistical mechanics of nervous nets, in Caianello, "Neural Networks". Springer Verlag, Berlin, 1968.
COWAN, J. D., A statistical mechanics of nervous activity, Some Mathematical Questions in Biology, Rhode Island: American Mathematical Society, 1970.
GRIFFITH, F., *Mathematical neurobiology*, Academic Press, New York, 1971.
JÓLESZ, F. A., I. GYÖNGY, The use of quantum mechanical measuring theory in the study of nervous system, International Conference on Bionics, Varna, 1975.
LÁBOS, E., Neural nets with measure-like coding, *Neumann Colloquium*, Szeged, 1973, (in Hungarian).
LÁBOS, E., The significance of wiring and language in the behavior and information processing of neural nets, Proc. of the Dept. of Biology of the Acad. Sci. Hung., v. 18, No 3, 1975, p. 325, (in Hungarian).
LIEBERSTEIN, H. M., *Mathematical physiology*, American Elsevier Publ. Co. Inc., New York, 1973.
MACKEY, H., *The mathematical foundation of quantum mechanics*, Benjamin X, New York, 1963,
MCCULLOCH, W. S., W, H, PITTS, A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biophys.*, v. 5, 1943, p. 115.
MICHALOV, J., Contribution to the question of analogy between excitation and damped oscillatory motion, *Biologia*, Bratislava, v. 22, 1967, p. 81.
MICHALOV, J., Physics of the excitable membrane: Physical statistics model of ionic conductances, *Biologia*, Bratislava, v. 23, 1968, p. 3.
NEUMANN, J. v., *Matematische Grundlagen der Quantenmechanik*, Springer Verlag, Berlin, 1932.
SCLABASSI, R. J., E. LÁBOS, B. MAGELHAES-CASTRO, B. E. STEIN, L. KRUGER, Ann. Meeting of Society for Neurosciences, Bethesda, Maryland, 1973.
WIENER, N., *Non-linear problems in random theory*, MIT Press, Cambridge, 1958.
WIENER, N., Perspectives in Cybernetics, *Progr. Brain Res.*, v. 17, 1965, p. 398.
WIGNER, E., Theorie der Quantenmechanischen Messung, *Physikertagung*, Wien, 1962.
WINOGRAD, J., J. D. COWAN, *Reliable computation in the presence of noise*, MIT Press, Cambridge, 1963.
JÓLESZ, F., M. SZILÁGYI, The stability of dynamical biological systems and the measuring process, *Neumann Colloqium*, Szeged, 1974, (in Hungarian).

(Received Nov. 8, 1977)

INDEX — TARTALOM

ISSN 0324—721 X

Felelős szerkesztő és kiadó: Gécseg Ferenc
A kézirat a nyomdába érkezett: 1978. július hó
Megjelenés: 1979. március hó
Példányszám: 1025. Terjedelem: 8, (A/5) ív
Készült monószedéssel, íves magasnyomással
az MSZ 5601 és az MSZ 5602—55 szabvány szerint
78-3174 — Szegedi Nyomda — F. v.: Dobó József igazgató