ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XV, 1 (1976)

J. ŁĄSKI (Gdańsk)

SYNTHESIS OF PROBABILITY TRANSFORMERS

1. Introduction. In this paper we discuss the problem of synthesis of a multi-input, multi-output combinational network as a probability transformer. This work is the continuation of previous papers which where based upon the spectral analysis of logical functions. The previous results are briefly recapitulated in Section 3 and are reformulated into the discrete probability domain in Section 4, where also the main theorems concerning the synthesis problem are proved. Section 5 is devoted to the problem of enumeration of equivalence classes of acceptable solutions.

In Section 2 we shortly discuss the main problems concerning the synthesis of switching circuits as models of stochastic processes. It has been shown that in most situations the problem of synthesis of the combinational probability transformer is of a special importance.

2. Switching circuits as models of stochastic processes. In many applications there arises a need of simulating a given stochastic process with prescribed probability characteristics. The scope of problems which need simulation as a useful or only possible research tool seems practically unlimited and is fairly widely described in the literature (see [15] and [16]).

The generation of a random variable with given probability distribution is one of the most important problems of stochastic simulation and presently is established as a well-defined discipline. The existing methods of forming a distribution are mostly of algorithmical nature and, therefore, are well-suited for computer applications. However, there are situations (e.g. in the case of field laboratory) when a physical device yielding the defined process is required. One of the approaches is to design a sequential circuit which, when supplied with a binary random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ with distribution $P(\xi) = P(x_1, x_2, \ldots, x_n)$, produces at its outputs the random vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_m)$ which is to simulate the required process $\mathcal{X}(t)$ (Fig. 1). In other words, the discrete stochastic process $\boldsymbol{Y}(\tau)$ $(\tau = 0, 1, \ldots)$ is considered as a digital model of $\mathcal{X}(t)$. The degree of adequacy between the process and the model depends upon the synthesis goals and is basically related to the assumed accuracy of description of $\mathcal{X}(t)$. Consequently, the structure of the circuit is synthesized to meet the assumed level of accuracy.

Namely, if the stochastic process $X(t)$ is characterized by the one-dimensional probability density $p(x)$, then the digital model $Y(\tau)$ is the output of a combinational network, i.e. memoryless one. The output distribution $P(\lambda) = P(y_1, y_2, \ldots, y_m)$ should then approximate the given density $p(x)$.

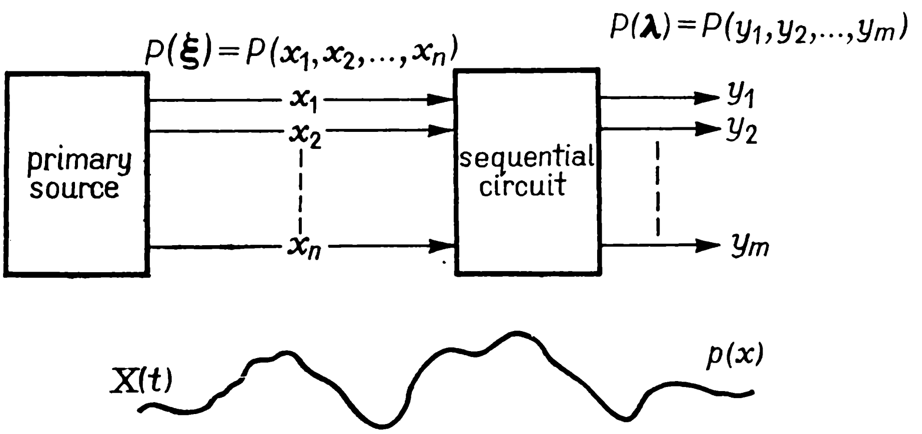

Fig. 1

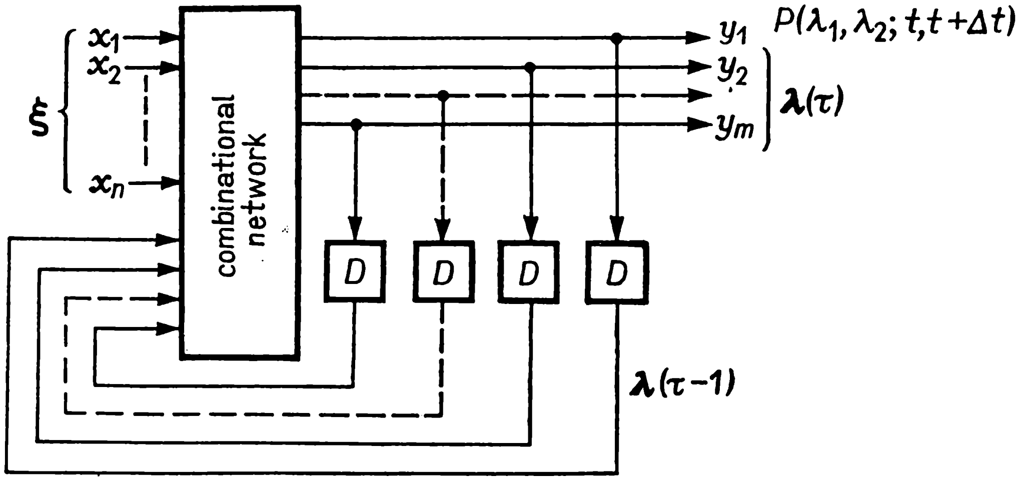

Fig. 2. Sequential circuit as a generator of history-dependent Markov stochastic process; D — delay element

On the other hand, if the process $X(t)$ is determined by the k-dimensional density $p(s_1, s_2, \ldots, s_k; t_1, t_2, \ldots, t_k)$, where t_i for $i = 1, 2, \ldots, k$ are the discrete moments of time, then the circuit contains memory elements necessary for storing the last $k-1$ values of $\lambda(\tau)$ (see Fig. 2 for $k = 2$). Therefore, the circuit is, in fact, an autonomous stochastic automaton and the generated process $Y(\tau)$ belongs to the class of k-dimensional Markov processes. The output distribution

$$P(y_1, y_2, \ldots, y_k; t_1, t_1+\Delta t, \ldots, t_1+k\Delta t),$$

where Δt is the delay or the "clock" interval, is now to approximate the continuous multidimensional density $p(s_1, s_2, \ldots, s_k)$.

Let us note that in both cases the synthesis problem reduces to the determination of combinational circuits realizing the required probability transformation. The same is true when the process to be simulated is determined by its one-dimensional density $p(x)$ and correlation function $R(\tau)$. In this case we deal with the one-dimensional Markov model, but the joint distribution $P_2(y_1, y_2; t, t+\Delta t)$ may be defined somewhat arbitrarily with only restrictions

$$\sum_{\{\lambda_i\}} P_2(\lambda_1, \lambda_2) = P_1(\lambda), \qquad \sum_{\{\lambda_i\}\times\{\lambda_j\}} P_2(\lambda_1, \lambda_2) v(\lambda_1) v(\lambda_2) \cong R(\Delta\tau),$$

where $v(\lambda)$ denotes the numerical interpretation (valuation) of λ (note that λ is a Boolean vector).

As far as the primary random vector $X(\tau)$ is concerned, the practical reasons suggest the white source at the input, i.e. with position- and history-independence and the uniform distribution $P(x_1, \ldots, x_n) = 2^{-n}$ $=$ const for all n-tuples $\langle x_1, x_2, \ldots, x_n\rangle$. However, it has been shown that

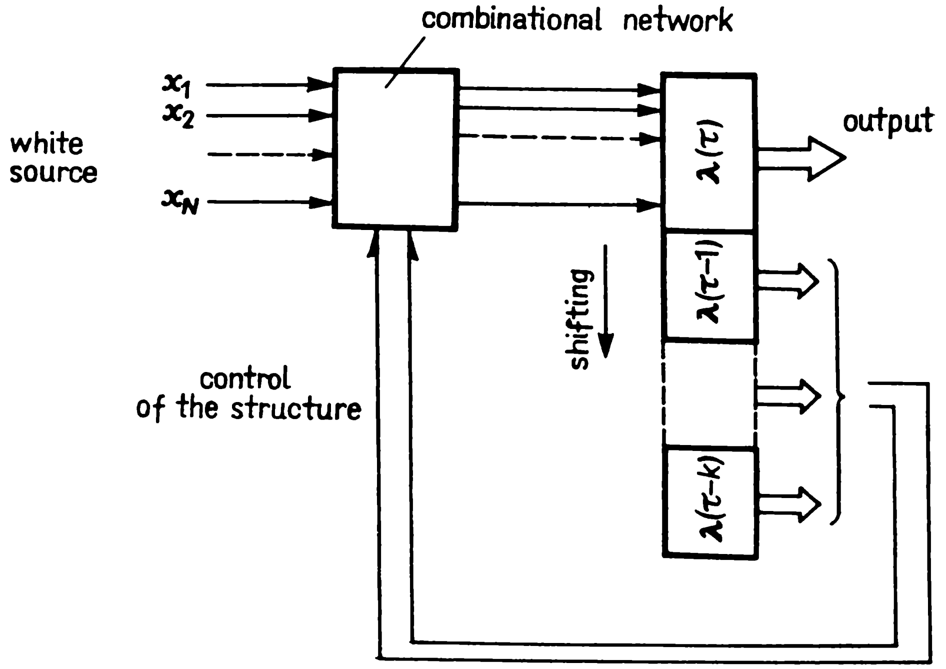

Fig. 3. An alternative structure-varying form of the automaton from Fig. 2

only a restricted class of distributions may be obtained in this way (see [7], [8], [13] and [14]); such distributions are referred to as binary realizable ones (cf. [7] and [8]).

It is clear that in the case of an automaton with memory (Fig. 2) the inputs of the combinational network $\xi(\tau), \lambda(\tau-1), \lambda(\tau-2), \ldots$ $\ldots, \lambda(\tau-k)$ are not generally the realization of the white random source. However, it is possible to adopt an alternative structure-varying form of the automaton (Fig. 3) when the memory is represented by k shifting registers storing the last k values $\lambda(\tau), \lambda(\tau-1), \ldots, \lambda(\tau-k)$ of the output.

In this case we deal with a finite number of the combinational networks, each of which realizes the conditional transformation $P[\lambda(\tau)|\lambda(\tau-1), \ldots$ $\ldots, \lambda(\tau-k)]$; the choice of an actual network depends upon the $(k-1)$-st step history of the process.

We shall not discuss the formal conditions to be satisfied to assure the equivalence of the two forms of automata since our purpose is to emphasize the importance of synthesis of a combinational probability transformer fed with the white binary vector. However, there is a need of studying this problem with emphasis laid upon such aspects as the state equivalence, state minimization, number of input variables, methods of passing from one description to the other, etc. Let us only note that the structure in Fig. 2 belongs to the class of so-called S-automata ([4] and [6]), while the structure in Fig. 3 represents an M-automaton, and that any M-automaton may be reformulated as an S-automaton, though the inverse operation is not always possible.

Let us also note that the possibility of synthesis of a stochastic automaton as a deterministic one fed with a random variable has been proved by many authors (see, e.g., [1]-[3]); some aspects of realization have been discussed in [3] and [11]. However, the problem of synthesis of a combinational transformer has not been paid many attention to in the literature (cf. Warfield [13] and [14]).

Recapitulating thus our considerations, we may divide the synthesis problem into the following five steps:

1° Determination of the principal goals of the simulation problem at hand, i.e. what characteristics of the process to be simulated should be taken into account in the discrete model.

2° Approximation of the continuous density (densities) by the discrete distribution (distributions).

3° Approximation of the above-mentioned probabilities by the binary-realizable ones.

4° Determination of the conditions which are to be satisfied by any set of logical functions $y_1, y_2, \ldots, y_m$ to realize the required probability transformation.

5° Choosing — perhaps on the base of an additional criterion — the best solution.

Step 1° is entirely based upon the nature of the problem to be solved by the simulation method and, therefore, we shall treat it as a given design specification.

Step 2° has been solved in [10].

Step 3° has been solved in the recent paper [8] for the one-dimensional density case but it may be easily adopted to handle also the multidimensional case.

The spectral approach as a tool for handling the problem of analysis and synthesis of a probability transformer has been developed in the recent papers [7] and [8]; this paper contains, as their continuation, the solution of step 4°.

The last step — an optimization of the solution — is to be left aside, since it strongly depends upon the assumed criterion of optimality.

Let us also mention that in [9] we have developed an algorithmic method (a generating function) for determining the orthogonal expansions of logical functions; this approach proved to be useful as a convenient tool for translating the results obtained in Section 3 on the base of spectral approach into the discrete probability domain (Section 4).

3. Synthesis. The problem of synthesis of probability transformers has been formulated in [7] in the following way.

Given binary random vectors

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n) \quad \text{and} \quad \boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_m)$$

with distributions

$$P(\xi) = P(x_1, x_2, \ldots, x_n) \quad \text{and} \quad P(\lambda) = (y_1, y_2, \ldots, y_m),$$

respectively, we have to find a set of Boolean functions $y_i = f_i(x_1, x_2, \ldots, x_n)$ $(i = 1, 2, \ldots, m)$ that transforms the random variable $\boldsymbol{X}$ into $\boldsymbol{Y}$.

The approach applied was based upon the orthogonal expansions of the logical functions and probability distributions into the Walsh series. According to the definitions introduced previously (see [8]), the set of Fourier coefficients $a = (a_0, a_1, \ldots, a_{2^n-1})$ of a logical function

$$f_a(x_1, x_2, \ldots, x_n) = f(\xi),$$

where

$$a_i = 2^{-n} \sum_{\{\xi_j\}} W_i(\xi) f(\xi), \qquad f_a(\xi) = \sum_{i=0}^{2^n-1} a_i W_i(\xi),$$

or, symbolically, $a = \psi(f_a)$ with $f_a = \psi^{-1}(a)$, will be referred to as the structure of the function $f(\xi)$. Similarly, for any discrete distribution $P(\xi)$ there exists a uniquely determined set of its orthogonal expansion coefficients $\sigma = (c_0, c_1, \ldots, c_{2^n-1})$ referred to as the spectrum of $P(\xi)$.

The main results presented in [7] establish the necessary and sufficient conditions for any ordered m-tuple of structures to represent the set of m logical functions $y_1, y_2, \ldots, y_m$ realizing the required transformation. They have been formulated as the set of the equations

$$\text{(1)} \qquad \begin{aligned} a^{(i)} \odot a^{(i)} &= a^{(i)} \quad (i = 1, 2, \ldots, m), \\ g_j &= \varphi_j(a^{(1)}, a^{(2)}, \ldots, a^{(m)}) \quad (j = 0, 1, \ldots, 2^m - 1), \end{aligned}$$

where $a^{(i)}$ denotes the structure of $y_i = f_i(\xi)$, and g_j — the elements of the spectrum of the output distribution $P(\lambda) = P(y_1, y_2, \ldots, y_m)$.

Let $\odot$ denote a binary operation such that if

$$\alpha = (a_0, a_1, \ldots, a_{2^n-1}) \quad \text{and} \quad \beta = (b_0, b_1, \ldots, b_{2^n-1})$$

are 2^n-element vectors, then $\varepsilon = \alpha \odot \beta$ (see [7] and [9]) is the vector $\varepsilon = (e_0, e_1, \ldots, e_{2^n-1})$, where

$$e_i = \sum_{j=0}^{2^n-1} a_j b_{j\oplus i} = \sum_{j=0}^{2^n-1} a_{j\oplus i} b_i .$$

The function $\varphi_j(a^{(1)}, a^{(2)}, \ldots, a^{(m)})$ has been defined implicitly by recurrent formulas for the spectrum of the joint distribution $P(\xi, \lambda)$.

In the general case we deal with a set of m-tuples that constitute the solutions of (1), rather than with one unique solution. Since these solutions belong to equivalence classes defined by the probability transformation relation, the set of all possible m-tuples is structured in a natural way by the conditions established by (1).

We assume that the necessary conditions for the existence of a solution are satisfied (see [7] and [13]). It means that the output distribution $P(\lambda)$ belongs to the class of binary-realizable distributions with its stochastic degree being less than n. Let us recall here that the *stochastic degree* of $P(\lambda)$ is defined as the length of the binary expansion of $P(\lambda_j)$ for $j = 0, 1, \ldots, 2^m-1$.

It has been shown in [7] that the spectrum $\delta^{(m)}$ of the joint input-output distribution $P(\xi, \lambda)$ consists of 2^m segments ϱ_i $(i = 0, 1, \ldots, 2^m-1)$,

$$\delta^{(m)} = \bigcup_{i=0}^{2^m-1} \varrho_i ,$$

where $\varrho_i \circ \varrho_j$ denotes the concatenation of segments ϱ_i and ϱ_j. Every segment corresponds to the subset of the output variables y_j $(j = 1, 2, \ldots, m)$ and has the form

$$\varrho_i = \sum_{l=1}^{r_i} \sum_{j=1}^{\binom{r_i}{l}} \frac{(-1)^l}{2^{m-l}} \sigma \odot a^{(i_1)} \odot a^{(i_2)} \odot \ldots \odot a^{(i_l)} + \frac{\sigma}{2^m} , \tag{2}$$

where $j = 2^{i_1-1} + 2^{i_2-1} + \ldots + 2^{i_l-1}$, and r_i $(i = 1, 2, \ldots, m)$ is the number of outputs associated with the segment ϱ_i, called the *rank of* ϱ_i, and structures $a^{(i_1)}, a^{(i_2)}, \ldots, a^{(i_l)}$ exhaust all possible combinations of the chosen subset of output functions.

For example, if $m = 3$, we have

$$\delta^{(3)} = (\tfrac{1}{8}\sigma)\circ(\tfrac{1}{8}\sigma-\tfrac{1}{4}a^{(1)}\odot\sigma)\circ(\tfrac{1}{8}\sigma-\tfrac{1}{4}a^{(2)}\odot\sigma)\circ$$
$$\circ(\tfrac{1}{8}\sigma-\tfrac{1}{4}a^{(1)}\odot\sigma-\tfrac{1}{4}a^{(2)}\odot\sigma+\tfrac{1}{2}a^{(1)}\odot a^{(2)}\odot\sigma)\circ$$
$$\circ(\tfrac{1}{8}\sigma-\tfrac{1}{4}a^{(3)}\odot\sigma)\circ(\tfrac{1}{8}\sigma-\tfrac{1}{4}a^{(1)}\odot\sigma-\tfrac{1}{4}a^{(3)}\odot\sigma+\tfrac{1}{2}a^{(1)}\odot a^{(3)}\odot\sigma)\circ$$
$$\circ(\tfrac{1}{8}\sigma-\tfrac{1}{4}a^{(2)}\odot\sigma-\tfrac{1}{4}a^{(3)}\odot\sigma+\tfrac{1}{2}a^{(2)}\odot a^{(3)}\odot\sigma)\circ$$
$$\circ(\tfrac{1}{8}\sigma-\tfrac{1}{4}a^{(1)}\odot\sigma-\tfrac{1}{4}a^{(2)}\odot\sigma-\tfrac{1}{4}a^{(3)}\odot\sigma+\tfrac{1}{2}a^{(1)}\odot a^{(2)}\odot\sigma+$$
$$+\tfrac{1}{2}a^{(1)}\odot a^{(3)}\odot\sigma+\tfrac{1}{2}a^{(2)}\odot a^{(3)}\odot\sigma-a^{(1)}\odot a^{(2)}\odot a^{(3)}\odot\sigma).$$

It has been also shown [7] that the joint spectrum $\delta^{(m)}$ can be determined recurrently as the concatenation of two segments $\delta_1^{(m)}$ and $\delta_2^{(m)}$,

$$\delta^{(m)} = \delta_1^{(m)}\circ\delta_2^{(m)}, \tag{3}$$

where

$$\delta_1^{(m)} = \tfrac{1}{2}\delta^{(m-1)}, \tag{4}$$
$$\delta_2^{(m)} = \delta_1^{(m)}-2\delta_1^{(m)}\odot a^{(m)} = \tfrac{1}{2}\delta^{(m-1)}-\delta^{(m-1)}\odot a^{(m)}.$$

We have assumed that the operation $\odot$ is distributive with respect to the operation of concatenation $\circ$, i.e., if

$$\delta^{(m)} = \bigcup_i \varrho_i,$$

then

$$\delta^{(m-1)}\odot a^{(m)} \overset{\mathrm{df}}{=} \bigcup_i \varrho_i\odot a^{(m)}.$$

Introducing the convention

$$(a\circ\beta)+(\gamma\circ\delta) \overset{\mathrm{df}}{=} (a+\gamma)\circ(\beta+\delta),$$

we may rewrite (2) in the following form:

$$\delta^{(m)} = \bigcup_{i=0}^{2^m-1} 2^{-m}\sigma+\sigma\odot\bigcup_{i=0}^{2^m-1}\varepsilon_i. \tag{5}$$

The joint spectrum $\delta^{(m)}$ contains the spectrum γ of the output distribution

$$P(\lambda) = \sum_{\{\xi_j\}} P(\xi,\lambda).$$

LEMMA 1. *The coefficients g_k of the output spectrum $\gamma = (g_0, g_1, \ldots, g_{2^m-1})$ of the output distribution $P(\lambda) = P(y_1, y_2, \ldots, y_m)$ are proportional to the first elements of the segments ϱ_k, namely, if*

$$\varrho_k = (d_{k\cdot 2^n}, d_{(k+1)\cdot 2^n}, \ldots, d_{(k+2^n-1)\cdot 2^n}),$$

then

$$g_k = 2^n d_{k\cdot 2^n}, \quad k = 0, 1, \ldots, 2^m-1.$$

Proof. Let $i = j\cdot 2^n + l$ for $j = 0, 1, \ldots, 2^m - 1$; $l = 0, 1, \ldots, 2^n - 1$. Then $W_i(\xi, \lambda) = W_j(\lambda) W_l(\xi)$ and

$$P(\lambda) = \sum_{\{\xi_j\}} P(\xi, \lambda) = \sum_{j=0}^{2^m-1} \sum_{l=0}^{2^n-1} d_{l+j\cdot 2^n} \sum_{\{\xi_j\}} W_j(\xi) W_l(\xi).$$

Since

$$\sum_{\{\xi_j\}} W_l(\lambda) = \begin{cases} 2^n & \text{if } l = 0, \\ 0 & \text{otherwise}, \end{cases}$$

we have

$$P(\xi) = \sum_{j=0}^{2^m-1} d_{j\cdot 2^n} W_{j\cdot 2^n}(\lambda)\cdot 2^n \Rightarrow g_j = 2^n d_{j\cdot 2^n}.$$

Corollary. *If $P(\xi) = 2^{-n}$, then $\sigma = (1, 0, 0, \ldots, 0)\cdot 2^{-n}$ and from* (5) *we have*

$$g_i = 2^n\{2^{-m}\cdot 2^{-n} + 2^{-n} e_0\} = 2^{-m} + e_0^{(i)}$$

with $e_0^{(i)}$ — the first element of the segment ε_i,

$$(6) \qquad e_0^{(i)} = \mathbf{1}_n \cdot \sum_{l=1}^{r_i} \sum_{j=1}^{\binom{r_i}{l}} \frac{(-1)^l}{2^{m-l}} a^{(i_1)} \odot a^{(i_2)} \odot \ldots \odot a^{(i_l)},$$

where $\mathbf{1}_n = (1, 0, \ldots, 0)$, the 2^n-element vector being the identity with respect to $\odot$, and $\mathbf{1}_n \cdot \varepsilon_i$ denotes the scalar multiplication of $\mathbf{1}_n$ and ε_i.

It has been shown [9] that the operation $\odot$, defined in the set $A = \{a^{(j)}\}$ $(j = 1, 2, \ldots, 2^{2^n})$ of all structures of order n, corresponds to the logical multiplication in the set $F = \{f_j\}$ of all logical functions of n variables. Therefore, the coefficients g_i are functions of the coefficients of structures of functions $y_j = f_j(x_1, x_2, \ldots, x_n)$ and all their possible products (for $l = 1, 2, \ldots, r_i$).

The first coefficient a_0 of a structure a is just proportional to the weight of the function $f_a(x_1, x_2, \ldots, x_n)$; clearly,

$$a_0 = \frac{w(f_a)}{2^n},$$

where the weight $w(f_a)$ of the function f_a is defined as the number of ones in the truth table for f_a (see [5]). In view of the one-to-one correspondence $a \Leftrightarrow f_a$ we write either $w(f_a)$ or $w(a)$.

Since the terms on the right-hand side of (6) of the form

$$\mathbf{1}_n \cdot (a^{(i_1)} \odot a^{(i_2)} \odot \ldots \odot a^{(i_l)})$$

denote the zero element of the vector $(a^{(i_1)} \odot a^{(i_2)} \odot \ldots \odot^{(i_l)})$, the coefficients g_i are linear combination of the weights of the products $a^{(i_1)} \odot a^{(i_2)} \odot \odot \ldots \odot a^{(i_l)}$. Clearly,

$$g_i = 2^{-m} + 2^{-n} \sum_{l=1}^{k} \sum_{j=1}^{\binom{k}{l}} \frac{(-1)^l}{2^{m-l}} w_j (a^{(j_1)} \odot a^{(j_2)} \odot \ldots \odot a^{(j_l)})$$

$$(i = 2^{i_1-1} + 2^{i_2-1} + \ldots + 2^{i_k-1};\ j = 2^{j_1-1} + 2^{j_2-1} + \ldots + 2^{j_k-1}).$$

Let $w_i^{(k)}$ $(i = 0, 1, \ldots, 2^{m-1};\ k = 1, \ldots, m)$ denote the weight of order k, i.e. the weight of the function $f_i(x_1, x_2, \ldots, x_n)$ being the product of k functions

$$f_{i_1}(\xi), f_{i_2}(\xi), \ldots, f_{i_k}(\xi)$$

$$(i_1, i_2, \ldots, i_k = 1, 2, \ldots, m;\ i = 2^{i_1-1} + 2^{i_2-1} + \ldots + 2^{i_k-1}).$$

Thus the actual numbers of functions contributing to the product are determined by the positions of ones in the binary expansion of i. Thus the synthesis conditions may be formulated as

$$w_i^{(k)} = q_i^{(k)} \quad (i = 1, 2, \ldots, 2^m - 1),$$

where the q_i are to be calculated on the base of the prescribed distribution or, equivalently, on the base of the output spectrum $\gamma = (g_0, g_1, \ldots, g_{2^m-1})$. The weights $w_i^{(k)}$ $(i = 1, 2, \ldots, 2^m - 1)$ can be determined in the following way:

(i) On the base of the first-order coefficients $g_i^{(1)}$ it is possible to find the first-order weights $w_i^{(1)}$ by the formula

$$w_i^{(1)} = (g_i^{(1)} - 2^{-m})(-2^n \cdot 2^{m-1}) = (-1) \cdot 2^{n+m-1}(g_i^{(1)} - 2^{-m})$$

$$(i = 2^{j-1};\ j = 1, 2, \ldots, m).$$

(ii) On the base of the second-order coefficients $g_i^{(2)}$ we are able to determine the second-order weights $w_i^{(2)}$ by the formula

$$w_i^{(2)} = (g_i^{(2)} - 2^{-m}) \cdot 2^{n+m-2} \cdot (-1)^2 + \tfrac{1}{2}(w_{i_1}^{(1)} + w_{i_2}^{(1)})$$

$$(i = 2^{i_1-1} + 2^{i_2-1};\ i_1, i_2 = 1, 2, \ldots, m).$$

Consequently, it is easy to extend this idea to the case of the higher-order coefficients using the following relationship:

$$w_i^{(k)} = (-1)^k (g_i^{(k)} - 2^{-m}) \cdot 2^{n+m+k} - \sum_{l=1}^{k-1} \sum_{j=1}^{\binom{k}{l}} \frac{(-1)^l}{2^{m-l}} w_j^{(l)}$$

$$(i = 2^{i_1-1} + 2^{i_2-1} + \ldots + 2^{i_k-1}).$$

Since the weights w_i are uniquely determined on the base of the spectrum of the output distribution, we have the following

THEOREM 1. *If the input distribution of the transformer is uniform, i.e.* $P(\xi) = 2^{-n}$, *and if there exist two ordered m-tuples of logical functions* $\langle f_1, f_2, \ldots, f_m\rangle$ and $\langle h_1, h_2, \ldots, h_m\rangle$ *such that the weights of the corresponding functions and the weights of all their possible products are equal, then the probability transformations yielded by these m-tuples are equivalent.*

If the set of the weights $\{w_i\}$ $(i = 1, 2, \ldots, 2^m-1)$ is determined, the synthesis problem can be defined as finding such an m-tuple of logical functions $f_1, f_2, \ldots, f_m$ which have $w_1^{(1)}, w_2^{(1)}, \ldots, w_{2m-1}^{(1)}$ ones in their truth tables and $w_i^{(l)}$ ones in their logical multiplication for all such possible products.

These results have been obtained on the base of the spectral method. In the next section we shall consider the problem again in the discrete-probability domain which will give us a deeper insight into many important questions.

4. Discrete probability analysis. To reformulate the results obtained in the last section into the discrete-probability language we need the concept of the generating function developed recently in [9]. Although it was primarily a method for finding orthogonal expansions of logical functions, it can be easily extended to the case of pseudo-Boolean functions, i.e. functions defined on the set

$$\{\xi_j\} = \{x_1^{(j)}, x_2^{(j)}, \ldots, x_n^{(j)}\}, \quad \text{where } x_i^{(j)} = 0, 1,$$

with arbitrary, not necessarily zero-one values.

The basic idea of the generating functions may be shortly summarized as follows. Let $H_n = \{h_i^{(n)}\}$ denote the set of functions belonging to the space L_n^2 (cf. [7]), and let $M_n = \{\mu_i^{(n)}\}$ be the set of orthogonal expansions of the h_i. Any element $h_i^{(n)} \in H_n$ can be represented as the concatenation of two elements $s_0^{(n-1)}$ and $s_1^{(n-1)}$, where $s_0, s_1 \in H_{n-1}$. We have then

$$h_i^{(n)} = s_0^{(n-1)} \circ s_1^{(n-1)}$$

and

$$h_i^{(n)} = (x_1, x_2, \ldots, x_n) = \begin{cases} s_0^{(n-1)}(x_1, x_2, \ldots, x_{n-1}) & \text{for } x_n = 0, \\ s_1^{(n-1)}(x_1, x_2, \ldots, x_{n-1}) & \text{for } x_m = 1. \end{cases}$$

In other words, if the functions $s_0^{(n-1)}$ and $s_1^{(n-1)}$ are given in tabular forms, then the function $h_i^{(n)} = s_0^{(n-1)} \circ s_1^{(n-1)}$ is defined as the concatenation of the two tables with $x_n = 0$ for the upper half (i.e. for s_0) and with $x_n = 1$ for the lower half (i.e. for s_1). This may be transliterated into the spectral analysis domain in the following way. Let

$$\eta_n = \psi(h^{(n)}), \quad \sigma_0 = \psi(s_0^{(n-1)}), \quad \sigma_1 = \psi(s_1^{(n-1)}) \quad (\sigma_0, \sigma_1 \in M_{n-1};\ \eta_n \in M_n).$$

If $h^{(n)} = s_0^{(n-1)} \circ s_1^{(n-1)}$, then

$$(7) \qquad \eta_n = \varphi(s_0^{(n-1)}, s_1^{(n-1)}) \overset{\text{df}}{=} \tfrac{1}{2}(s_0^{(n-1)} + s_1^{(n-1)}) \circ \tfrac{1}{2}(s_0^{(n-1)} - s_1^{(n-1)}).$$

Consequently, for any pair $(s_0^{(n-1)}, s_1^{(n-1)})$ of orthogonal expansions of order $n-1$ (i.e. of $n-1$ variables), there exists a uniquely determined orthogonal expansion of order n defined by the generating function φ in (7). Conversely, for any $\varepsilon_n \in M_n$, there exists a uniquely defined pair of expansions of order $n-1$, namely

$$(q_1^{(n-1)}, q_2^{(n-1)}) = \varphi^{-1}(\varepsilon_n) = (\varepsilon' + \varepsilon'', \varepsilon' - \varepsilon''), \quad \text{where } \varepsilon_n = \varepsilon' \circ \varepsilon''.$$

Now we can apply the concept of the generating function to the case of the probability transformer. Consider the equation

$$\delta^{(m)} = \delta_1^{(m)} \circ \delta_2^{(m)},$$

where $\delta^{(m)}$ is the spectrum of the joint input-output distribution $P^{(m)}(\xi, y_1, y_2, \ldots, y_m)$, i.e. $\delta^{(m)} = \psi(P^m)$. Then, representing $P^{(m)}$ as the concatenation of two segments $P_0^{(m)}$ and $P_1^{(m)}$ (note that, generally, $P_0^{(m)}$ and $P_1^{(m)}$ are not probability distributions),

$$P^{(m)} = P_0^{(m)} \circ P_1^{(m)},$$

and making use of expressions (3) and (4), we have

$$\varphi^{-1}(\delta^{(m)}) = (\delta_1^{(m)} + \delta_2^{(m)}, \delta_1^{(m)} - \delta_2^{(m)}) = (\delta^{(m-1)} - \delta^{(m-1)} \odot \alpha^{(m)}, \delta^{(m-1)} \odot \alpha^{(m)}).$$

Since

$$P_0^{(m)} = \psi^{-1}(\delta^{(m-1)} - \delta^{(m-1)} \odot \alpha^{(m)}) \quad \text{and} \quad P_1^{(m)} = \psi^{-1}(\delta^{(m-1)} \odot \alpha^{(m)}),$$

we have

$$P^{(m)} = (P^{(m-1)}[1_n - f_m]) \circ (P^{(m-1)} f_m) = P^{(m-1)}[(1_n - f_m) \circ f_m], \tag{8}$$

where 1_n is the logical function identically equal to 1.

The last expression is to be interpreted in the following way. If $P^{(m-1)}$ is the joint distribution of $n+m-1$ variables, then its tabular representation consists of 2^{m-1} segments h_j, each of which contains 2^n elements. Thus we have

$$P^{(m-1)} = \bigcup_{j=0}^{2^{m-1}-1} h_j$$

and

$$P^{(m-1)} f_m = \bigcup_{j=0}^{2^{m-1}-1} (h_j f_m), \tag{9}$$

where

$$(h_j f_m)(x_1, x_2, \ldots, x_{n-1}) \overset{\text{df}}{=} h_j(x_1, x_2, \ldots, x_{n-1}) f_m(x_1, x_2, \ldots, x_{n-1})$$

for all possible values of arguments $(x_1, x_2, \ldots, x_{n-1})$. Thus the distribution $P^{(m)}$ consists of 2^m segments divided into two parts: the first being $P^{(m-1)}$ multiplied by $(1_n - f_m)$, and the second — $P^{(m-1)}$ multiplied by f_m; the multiplication is performed according to (9).

It is easy to show that the normalization condition for $P^{(m)}$ is always satisfied if only it is satisfied for $P^{(m-1)}$. Let $P^0(\xi) \equiv P(\xi)$. Then

$$P^{(1)} = P(\xi)[\bar{f}_1 \circ f_1],$$
$$P^{(2)} = P^{(1)}[\bar{f}_2 \circ f_2] = P(\xi)[\bar{f}_1\bar{f}_2 \circ \bar{f}_1 f_2 \circ f_1\bar{f}_2 \circ f_1 f_2].$$

It is easy to show by induction that expression (8) is equivalent to

$$P^{(m)} = P(\xi) \bigcup_{j=0}^{2^m-1} (f_1^{j_1} f_2^{j_2} \dots f_m^{j_m}) \quad (j_i = 0, 1), \tag{10}$$

where

$$f_i^{j_i} = \begin{cases} f_i & \text{if } j_i = 1, \\ \bar{f}_i & \text{if } j_i = 0. \end{cases}$$

The output distribution $\pi^{(m)}(y_1, y_2, \dots, y_m)$ is then easily calculated by the formula

$$\pi^{(m)} = \sum_{\{\xi\}} P^{(m)}(\xi, y_1, y_2, \dots, y_m). \tag{11}$$

In the case of the uniform distribution, i.e. $P(\xi) = 2^{-n}$, by (10) and (11) we have

$$\pi^{(m)} = 2^{-n} \bigcup_{j=0}^{2^m-1} w_j(f_1^{j_1} f_2^{j_2} \dots f_m^{j_m}) = 2^{-n} \bigcup_{j=0}^{2^m-1} v_j, \tag{12}$$

where $v_j \stackrel{\text{df}}{=} w_j(f_1^{j_1} f_2^{j_2} \dots f_m^{j_m})$ are the weights of minterms of the output variables $y_1, y_2, \dots, y_m$. By a *minterm* of a set of Boolean variables $x_1, \dots, x_k$ we mean any expression of the form $x_1^{j_1} x_2^{j_2} \dots x_k^{j_k}$ $(j_l = 0, 1)$, where $x_l^{j_l} = x_l$ if $j_l = 1$ and $x_l^{j_l} = \bar{x}_l$ if $j_l = 0$. These results can be summarized in the form of the following two theorems:

THEOREM 2. *If the input distribution of the binary probability transformer is uniform, then the value of the output distribution $\pi^{(m)}(y_1, y_2, \dots, y_m)$ for concrete values of arguments $y_1^{j_1}, y_2^{j_2}, \dots, y_m^{j_m}$ is proportional to the weight v_j $(j = 2^{j_1-1} + 2^{j_2-1} + \dots + 2^{j_m-1})$ of the function being the j-th minterm of output variables.*

THEOREM 3. *In the case of the uniform distribution at the input any two ordered m-tuples of logical functions $\langle f_1, f_2, \dots, f_m \rangle$ and $\langle h_1, h_2, \dots, h_m \rangle$ yield the same probability transformations if and only if the weights of all minterms of the output variables are equal, i.e. if*

$$w(f_1^{j_1} f_2^{j_2} \dots f_m^{j_m}) = w(h_1^{j_1} h_2^{j_2} \dots h_m^{j_m}).$$

In order to show that Theorems 1 and 3 are equivalent it is sufficient to replace each f_i in (10) by $1 - f_i$ and perform the straightforward multiplications. For example, if $m = 2$, then

$$P^{(2)} = [2^{-n}](1_n - f_1 - f_2 + f_1 f_2) \circ (f_1 - f_1 f_2) \circ (f_2 - f_1 f_2) \circ (f_1 f_2),$$

where $[2^{-n}]$ denotes the vector consisting of 2^n elements equal to 2^{-n}, and 1_n is the logical function identically equal to 1.

If $m = 3$, then

$$P^{(3)} = [2^{-n}](1_n - f_1 - f_2 - f_3 + f_1f_2 + f_1f_3 + f_2f_3 - f_1f_2f_3)\circ$$
$$\circ(f_1 - f_1f_2 - f_1f_3 + f_1f_2f_3)\circ(f_2 - f_1f_2 - f_2f_3 + f_1f_2f_3)\circ(f_1f_2 - f_1f_2f_3)\circ$$
$$\circ(f_3 - f_1f_3 - f_2f_3 + f_1f_2f_3)\circ(f_1f_3 - f_1f_2f_3)\circ(f_2f_3 - f_1f_2f_3)\circ(f_1f_2f_3).$$

Therefore, the output distribution $\pi^{(m)}$ may be expressed in terms of the weights of the functions f_i and their products, namely, for $m = 2$,

$$\pi^{(2)} = 2^{-n}(2^n - w_1 - w_2 + w_{12})\circ(w_1 - w_{12})\circ(w_2 - w_{12})\circ(w_{12}),$$
$$\text{where } w_{ij} \stackrel{\Delta}{=} w(f_if_j).$$

It is easy to show that the j-th element of the distribution $\pi^{(m)} = (\mu_0^{(m)}, \mu_1^{(m)}, \ldots, \mu_{2^m-1}^{(m)})$ is determined by the expression

$$\mu_j = 2^{-n}\sum_{k=j}^{2^m-1}(-1)^l w_k^{(l)} \tag{13}$$

$$(j = 0, 1, \ldots, 2^m - 1;\ k = 2^{j_1-1} + 2^{j_2-1} + \ldots + 2^{j_l-1}),\quad w_0^{(0)} \equiv 2^n.$$

It is possible to find a method for determining the weights w_i on the base of the output distribution $\pi^{(m)} = (\mu_0, \mu_1, \ldots, \mu_{2^m-1})$; in this case one has to start with the highest-order weight w_{2^m-1}, since $w_{2^m-1} = \mu_{2^m-1}\cdot 2^n$. The $(m-1)$-order weights are then easily determined by the formula

$$w_i^{(m-1)} = \big(\mu_i + (-1)^m w_{2^m-1}\big)\cdot 2^n,$$

and so on up to the first-order weights.

5. Number of solutions. Since the output probabilities μ_j $(j = 0, 1, \ldots, 2^m - 1)$ are non-negative, equations (12) and (13) establish conditions satisfied by an arbitrary m-tuple of logical functions $\langle f_1, f_2, \ldots, f_m\rangle$ of n variables. The following two equivalent sets of conditions (14) and (15) are obtained from (12) and (13), respectively:

$$\sum_{j=0}^{2^m-1} w(f_1^{j_1}f_2^{j_2}\ldots f_m^{j_m}) = 2^n = \sum_{j=0}^{2^m-1} v_j,\qquad v_j \geqslant 0, \tag{14}$$

$$\sum_{k=j}^{2^m-1}(-1)^l w_k^{(l)} \geqslant 0\qquad (k = 2^{j_1-1} + 2^{j_2-1} + \ldots + 2^{j_l-1}),\ w_0^{(0)} \equiv 2^n. \tag{15}$$

It is easy to verify that the normalization condition is always satisfied by $\pi^{(m)}$ if its elements are determined by (13), since

$$\sum_{j=0}^{2^m-1}\mu_j = \sum_{j=0}^{2^m-1} 2^{-n}\sum_{k=j}^{2^m-1}(-1)^l w_k^{(l)} \equiv 1$$

for arbitrary values $w_i^{(l)}$ satisfying conditions (15). Since conditions (14) and (15) are equivalent, there exist one-to-one relationships between two sets of weights $\{w_i\}$ and $\{v_i\}$. Since v_i are proportional to the output probabilities μ_i, they can be easily computed from (13).

The problem of the existence of a defined number of solutions is easier tractable using conditions (14) laid upon the weights of the minterms of the output variables. Let us note that if the number of input variables is n, then the stochastic degree of the output variable does not exceed n. Conversely, for any given distribution with stochastic degree k, the number of input variables must be greater than or equal to k to satisfy conditions (14). Assume thus that the conditions are satisfied and consider the number of solutions of the synthesis problem.

It is easy to see that an arbitrary solution may be simply obtained by representing the j-th minterm of the output variables $y_1, y_2, \ldots, y_m$ as the logical sum of v_j minterms ($j = 0, 1, \ldots, 2^m-1$) of the input variables $x_1, x_2, \ldots, x_n$:

$$\lambda_j = y_1^{j_1} y_2^{j_2} \ldots y_m^{j_m} = \bigvee_{\{i:\ i=s_1,\ldots,s_{v_j}\}} x_1^{i_1} x_2^{i_2} \ldots x_n^{i_n}.$$

Taking then into account those input minterms for which the output variable y_k equals 1, we obtain the canonical expression for y_k.

Consider the first output minterm λ_0. It can be assigned v_0 input minterms in $\binom{2^n}{v_0}$ ways. To the next minterm λ_1 its v_1 input minterms can be assigned in $\binom{2^n - v_0}{v_1}$ ways. Consequently, the total number of different assignments N (i.e. the number of different m-tuples realizing the required probability transformation) is expressed by the following formula:

$$N = \binom{2^n}{v_0} \prod_{j=1}^{2^m-2} \binom{2^n - \sum_{i=0}^{j-1} v_i}{v_{k+1}}. \tag{16}$$

The last minterm λ_{2^m-1} has not been included in this formula, since the sequence of weights $\{v_i\}$ is linearly dependent and there is no freedom in the choice of the v_{2^m-1} input minterm.

Note that in the case where the number of input variables n' is greater than the output stochastic degree n, say $n' = n+k$, we have to assign $2^k v_i$ input minterms to each output minterm λ_i.

Finally, let us consider the set $\mathcal{F}$,

$$\mathcal{F} \stackrel{\text{df}}{=} F_n \times F_n \times \ldots \times F_n = (F_n)^m,$$

being the m-th Cartesian power of the set F_n containing 2^{2^n} logical functions of n binary variables.

The probability transformation relation R_π defined in F by

$$\langle f_1, f_2, \ldots, f_m\rangle R_\pi \langle g_1, g_2, \ldots, g_m\rangle \Leftrightarrow \forall j < 2^m, \quad j = \sum_{l=1}^{m} 2^{j_l - 1},$$
$$w(f_1^{j_1} f_2^{j_2} \ldots f_m^{j_m}) = w(g_1^{j_1} g_2^{j_2} \ldots g_m^{j_m}),$$

is an equivalence relation in $\mathscr{F}$. Therefore, it divides $\mathscr{F}$ into mutually exclusive equivalence classes containing elements yielding the same probability transformation of the input random variable. The number of these classes is equal to the number of partitions of 2^n by 2^m, since each partition is equivalent to the choice of 2^m weights v_i of the output minterms. Although a direct formula for the number of classes N is not known, it may be found indirectly as the coefficient at t^n in the following expression (see [5] and [12]):

$$\prod_{i=1}^{\infty} \frac{1}{1-t^i}.$$

The number of elements in the class is given by (16) since the set of weights $\{v_i\}$ (alternatively – the set $\{w_i\}$) uniquely defines the equivalence class. Therefore, if we denote by $N[\{v_i^{(j)}\}]$ the number of elements in the equivalence class determined by the set of output minterms of weights $\{v_i\}$ for a given partition p_j ($j = 1, 2, \ldots, N_c$), then

$$\sum_{j=1}^{N_c} N[\{v_i^{(j)}\}] = \sum_{j=1}^{N_c} \binom{2^n}{v_0} \sum_{s=1}^{2^m-2} \binom{2^n - \sum_{i=0}^{s-1} v_i^{(j)}}{v_{s+1}} = 2^{m\cdot 2^n}.$$

6. Conclusions. The results obtained in this paper are due to the assumption that the input distribution of the transformer is uniform. In the case of an arbitrary input distribution both the class of realizable probabilities and the structure of the transformer are subject to changes. Namely, the analysis of equations (10) and (11) shows that the probability of an output minterm is the sum of probabilities of those input minterms for which the output minterm takes on logical 1. Consequently, the output stochastic degree cannot exceed the input stochastic degree and we may conclude that the class of realizable distributions consists of such distributions which can be represented as a combination of input minterm probabilities.

Therefore, the conditions laid upon the weights of the output minterms are not still valid and have to be replaced by new ones; one may suspect that they will be strongly dependent on the form of the input distribution.

References

[1] G. N. Cercvadze (Г. Н. Церцвадзе), *Некоторые свойства и методы синтеза стохастических автоматов*, Автоматика и Телемеханика 24 (1963), p. 341-352.

[2] M. L. Cetlin (М. Л. Цетлин), *О поведении конечных автоматов в случайных средах*, ibidem 22 (1961), p. 1345-1354.

[3] El-Ghoroury and S. C. Gupta, *Realization of stochastic automata*, IEEE-TC C-20 (1971), p. 889-893.

[4] A. Gill, *Introduction to the theory of finite-state machines*, McGraw-Hill, New York 1962.

[5] M. A. Harrison, *Introduction to switching and automata theory*, McGraw-Hill, New York 1965.

[6] J. Klir, *A note on the basic block diagrams of finite automata from the engineering point of view*, IEEE EC-16 (1967), p. 223-224.

[7] J. Łąski, *The analysis and synthesis of probability transformers*, Journal of Engineering Mathematics 7 (1973), p. 289-295.

[8] — *Analiza kombinacyjnych przekształtników prawdopodobieństwa*, Arch. Automatyki i Telemechaniki 18 (1973), p. 308-319.

[9] — *An algorithm for generating the orthogonal expansions of logical functions*, Bull. Acad. Pol. Sci., Sér. math. astr. phys., 23 (1975), p. 209-213.

[10] — *Algorytmiczne metody oceny funkcji gęstości prawdopodobieństwa*, Arch. Automatyki i Telemechaniki 14 (1969), p. 439-457.

[11] A. A. Lorentz (А. А. Лоренц), *Элементы конструктивной теории вероятностных автоматов*, Рига 1972.

[12] J. Riordan, *An introduction to combinatorial analysis*, J. Wiley, New York 1958.

[13] J. N. Warfield, *Synthesis of switching circuits to yield prescribed probability relations* in *Symposium on switching and logical design*, Ann Arbor, Mich., Oct. 6-8, 1965, p. 303-309.

[14] — *Switching networks as models of discrete stochastic processes* in J. T. Tou (Editor), *Applied automata theory*, Academic Press, 1968.

[15] R. Zieliński, *Metody Monte-Carlo*, WNT, Warszawa 1970.

[16] — *Generatory liczb losowych*, WNT, Warszawa 1972.

INSTITUTE OF INFORMATICS
TECHNICAL UNIVERSITY, GDAŃSK

Received on 23. 1. 1974

J. ŁĄSKI (Gdańsk)

SYNTEZA PRZEKSZTAŁTNIKÓW PRAWDOPODOBIEŃSTWA

STRESZCZENIE

Praca poświęcona jest zagadnieniu syntezy wielowejściowo-wielowyjściowej sieci logicznej, jako przekształtnika prawdopodobieństwa, i stanowi kontynuację poprzednich prac autora ([7]-[10]).

W rozdz. 2 omówiono problem syntezy sieci przełączających, jako modeli dyskretnych procesów przypadkowych, i pokazano, że głównym zadaniem jest tu synteza kombinacyjnych sieci przełączających, realizujących dane przekształcenie prawdopodobieństwa. W rozdz. 3, opierając się na metodzie widmowej, sformułowano warunki, jakie muszą spełniać wszystkie układy funkcji logicznych, aby zachodziła żądana transformacja prawdopodobieństwa. Wyniki te zostały szerzej omówione w rozdz. 4, gdzie wykorzystano analizę dyskretnych, dwójkowo realizowalnych rozkładów prawdopodobieństwa; podstawą takiej analizy był aparat funkcji generujących, wprowadzony w [9]. W rozdz. 5 obliczono liczebność klas równoważności, zawierających m-ki funkcji logicznych o tych samych właściwościach transformujących.