W dalszych rozważaniach ograniczymy się do przypadku, gdy $f(y) = y^{1-\alpha}$, gdzie α jest liczbą rzeczywistą różną od jedności.

Dzięki oczywistej równości

$$[y^{1-\alpha}(x)]' = (1-\alpha)\,y^{-\alpha}\,y' = \frac{1-\alpha}{y^{\alpha}}\,y'$$

otrzymamy

$$\boxed{\frac{1}{g'(x)}\,\frac{1}{y^{\alpha}}\,y' = \frac{1}{1-\alpha}\,p\,y^{1-\alpha} - \frac{1}{1-\alpha}\,p\,y^{1-\alpha}(0),} \tag{107}$$

gdzie $p = T^{-1}$.

Wzór (107) jest słuszny dla dowolnego $\alpha \neq 1$. W przypadku szczególnym, gdy np. $\alpha = -1$, mamy

$$\frac{1}{g'(x)}\,y\,y' = \frac{1}{2}\,p\,y^2 - \frac{1}{2}\,p\,y^2(0). \tag{107a}$$

W innym zaś przypadku szczególnym, a mianowicie gdy $\alpha = \frac{1}{2}$:

$$\frac{1}{g'(x)}\,\frac{1}{\sqrt{y}}\,y' = 2\,p\sqrt{y} - 2\,p\sqrt{y(0)}. \tag{107b}$$

Wyprowadzone wyżej zależności umożliwiają stosowanie metody operatorowej w niektórych prostych zagadnieniach nieliniowych.

Obliczmy wyniki niektórych operacji.

1. Wprost ze wzoru definicyjnego (106) otrzymujemy:

$$p^{-\nu}(1) = \frac{g^{\nu}(x)}{\nu!}. \tag{108}$$

Istotnie, zakładając we wzorze (106) $y(x) = 1$, otrzymujemy

$$T(1) = \int_0^x g'(\xi)\,d\xi = g(x) - g(0)$$

oraz wobec założenia $g(0) = 0$:

$$T(1) = p^{-1}(1) = g(x).$$

Uogólniając uzyskany wynik otrzymamy (108).

2. Obliczmy wynik operacji

$$\frac{p}{p-\lambda}(1).$$

Podstawiając we wzorze (106) $y^{1-\alpha}(x) = e^{\lambda g(x)}$, otrzymamy

$$T e^{\lambda g(x)} = \int_0^x e^{\lambda g(\xi)} g'(\xi)\, d\xi = \frac{1}{\lambda}(e^{\lambda g(x)} - 1),$$

a stąd, po elementarnych przekształceniach

$$\boxed{e^{\lambda g(x)} = \frac{p}{p-\lambda}(\,)}. \tag{109}$$

Ze wzoru powyższego widzimy, że funkcja będąca wynikiem operacji $\frac{p}{p-\lambda}(1)$ jest uzależniona od funkcji wagi $g(x)$. Jeśli np. założymy, że $g(x) = x$, to otrzymamy znany w rachunku operatorowym wzór określający funkcję $e^{\lambda x}$ jako wynik operacji wykonanej nad funkcją jednostkową. Jeśli w innym przypadku szczególnym położymy $g(x) = \ln x$, to otrzymamy również ustalony poprzednio wzór

$$y^{\lambda} = \frac{p}{p-\lambda}(1); \qquad x \geqslant 1$$

wyprowadzony dla operatorów równań Eulera.

Jeżeli wreszcie założymy np. $g(x) = \ln \operatorname{tg}\left(\frac{x}{2} + \frac{\pi}{4}\right)$, co odpowiada $g'(x) = \frac{1}{\cos x}$, to otrzymamy znów inny wzór, a mianowicie

$$\operatorname{tg}^{\lambda}\left(\frac{x}{2} + \frac{\pi}{4}\right) = \frac{p}{p-\lambda}(1). \tag{110}$$

Można oczywiście z góry obliczyć wyniki niektórych operacji spotykanych w zagadnieniach praktycznych i ułożyć odpowiednie tablice, identyczne do tablic stosowanych w metodzie przekształcenia Laplace'a. Tablice takie ułatwiłyby rozwiązywanie rozważanych równań różniczkowych.

3. Obliczmy z kolei wyniki następujących operacji

$$\frac{1}{p-\lambda}(1) \quad \text{i} \quad \frac{p^{-1}}{p-\lambda}(\,).$$

Otóż korzystając ze wzoru (106) od razu znajdujemy

$$\frac{1}{p-\lambda}(1) = T[e^{\lambda g(x)}] = \frac{1}{\lambda}[e^{\lambda g(x)}] - 1]$$

oraz (111)

$$\frac{p^{-1}}{p-\lambda}(1) = \frac{1}{\lambda^2}[e^{\lambda g(x)} - \lambda g(x) - 1].$$

Wzór (107) można oczywiście uogólnić na pochodną drugiego rzędu $y''(x)$. Uogólnienie to wyprowadzimy dla przypadku szczególnego, gdy funkcja wagi $g(x)$ jest funkcją liniową: $g(x) = x$. Podstawiając mianowicie

$$y^{1-\alpha} = (1-\alpha) y_1^{-\alpha} y_1'$$

we wzorze

$$\frac{1}{y^\alpha} y' = \frac{1}{1-\alpha} p\, y^{1-\alpha} - \frac{1}{1-\alpha} p\, y^{1-\alpha}(0),$$

otrzymamy

$$y_1^{-\alpha} y_1'' - \alpha y_1^{-\alpha-1} y_1' = p\, y_1^{-\alpha} y_1 - p\, y_1^{-\alpha}(0) y_1'(0).$$

Korzystając teraz z zależności (107), w której zgodnie z przyjętym założeniem uwzględniamy $g'(x) = 1$, otrzymamy ostatecznie

$$\boxed{y^{-\alpha} y'' - \alpha y^{-\alpha-1} y' = \frac{1}{1-\alpha} p^2 y^{1-\alpha} - \frac{1}{1-\alpha} p^2 y^{1-\alpha}(0) - p\, y^{-\alpha}(0) y'(0).} \quad (112)$$

Można również w prosty sposób wyprowadzić wzór na pochodną drugiego rzędu $y''(x)$ w założeniu, że funkcja wagi $g(x)$ jest funkcją logarytmiczną

$$g(x) = \ln x; \quad x \geqslant 1.$$

Wówczas we wzorze definicyjnym (106) dolną granicę całkowania należy przesunąć z punktu $x = 0$ do $x = 1$. Po wykonaniu zupełnie elementarnych przekształceń otrzymamy wzór następujący

$$\boxed{x^2 y^{-\alpha} y'' - \alpha x^2 y^{-\alpha-1} y' = \frac{1}{1-\alpha} p(p-1) y^{1-\alpha} - \frac{1}{1-\alpha} p(p-1) y^{1-\alpha}(1) + \\ - p\, y^{-\alpha}(1) y'(1).}$$

(1 3)

Można zauważyć, że w przypadku $\alpha = 0$, z (112) i z (113) otrzymujemy ustalone uprzednio zależności, a mianowicie z (112):

$$y'' = p^2 y - p^2 y(0) - p\, y(0)$$

oraz z (113):

$$x^2y'' = p(p-1)y - p(p-1)y(1) - py'(1).$$

Poniżej podane są przykłady objaśniające sposób praktycznego wykorzystania wyprowadzonych wzorów.

Przykład 1. Wyznaczyć rozwiązanie równania różniczkowego

$$\cos x\, y\, y' + y^2 = 1,$$

spełniające warunek początkowy $y(0) = 2$.

Rozwiązanie. Zakładamy $g(x) = \ln \operatorname{tg}\left(\frac{x}{2} + \frac{\pi}{4}\right)$. Zatem $g'(x) = \frac{1}{\cos x}$ oraz, z uwagi na wzór (107a),

$$\cos x\, y\, y' = \frac{1}{2} p\, y^2 - p\, 2.$$

Mamy więc następujące równanie operatorowe

$$\frac{1}{2} p\, y^2 + y^2 = p\, 2 + 1.$$

Zatem

$$y^2 = 4\frac{p}{p+2}(1) + \frac{2}{p+2}(1)$$

oraz, wobec (109) i (111),

$$y^2 = 4e^{-2g(x)} - e^{-2g(x)} + 1 = 3e^{-2\ln \operatorname{tg}\left(\frac{x}{2}+\frac{\pi}{4}\right)} + 1 = 3\operatorname{ctg}^2\left(\frac{x}{2}+\frac{\pi}{4}\right) + 1.$$

Poszukiwana funkcja dana jest więc przez wzór

$$y(x) = \pm\sqrt{3\operatorname{ctg}^2\left(\frac{x}{2}+\frac{\pi}{4}\right) + 1}.$$

Przykład 2. Wyznaczyć rozwiązanie równania różniczkowego

$$x^2 y\, y'' + x^2 y' - 2\, x\, y\, y' + y^2 = \ln x,$$

spełniające warunki początkowe $y(1) = 1$, $y'(1) = 0$.

Rozwiązanie. Korzystając z następujących zależności

$$x\, y\, y' = \frac{1}{2} p\, y^2 - \frac{1}{2} p\, y^2(1),$$

$$x^2 y\, y'' + x^2 y' = \frac{1}{2} p\,(p-1)\, y^2 - \frac{1}{2} p\,(p-1)\, y^2(1) - p\, y(1)\, y'(1),$$

zamiast zadanego równania różniczkowego rozwiązujemy następujące równanie operatorowe

$$\frac{1}{2} p\,(p-1)\, y^2 - p\, y^2 + y^2 = \ln x + \frac{1}{2} p\,(p-1)\, 1 - p\, 1,$$

które dzięki równości

$$\ln x = p^{-1}(1),$$

daje

$$y^2 = \frac{1}{2p}(1) - \frac{3}{2}\frac{1}{p-2}(1) + \frac{1}{p-1}(1) + 1.$$

Wobec powyższego na mocy wzorów (108) i (111) otrzymamy ostatecznie

$$y(x) = \pm\sqrt{\frac{1}{2}\ln x - \frac{3}{4}x^2 + x + \frac{3}{4}}; \quad x \geqslant 1.$$

Politechnika Warszawska
Katedra Teletransmisji Przewodowej

WYKAZ LITERATURY

1. Bellert S. On Foundations of Operational Calculus. — Bull. Acad. Polon. Sci., Cl.III, 5 (1957), 855—858.
2. Bellert S.: Metoda operatorów liczbowych. — Rozprawy Elektrotechniczne, Tom V, zeszyt 4 (1959), 515—586.
3. Bellert S.: Podstawy rachunku operatorów liczbowych. — Zeszyty Naukowe Politechniki Warszawskiej, Elektryka Nr 3 (1954).
4. Bellert S.: Zastosowanie rachunku operatorów liczbowych do badania układów impulsowych. — Zeszyty Naukowe Politechniki Warszawskiej, Elektryka Nr 4 (1954).
5. Bittner R.: On Certain Axiomatics for the Operational Calculus. — Bull. Acad. Polon. Sci., Vol. VII, No 1 (1959), 1—9.
6. Słowikowski W.: A Generalization of Mikusiński's Operational Calculus. — Bull. Acad. Polon. Sci. Cl.III, 4 (1956), 643—647.
7. Gelfond A. O.: Isczislenije koniecznych raznostiej. Moskwa — Leningrad 1952.

С. БЕЛЛЕРТ

ОПЕРАЦИОННОЕ ИСЧИСЛЕНИЕ В ЛИНЕЙНЫХ ПРОСТРАНСТВАХ

Резюме

В работе представлен метод являющийся обобщением классических методов операционного исчисления. Основы этого метода были опубликованы автором в 1957 году в работе [1], а настоящая статья является расширением метода, так по теоретическим предпосылкам как и по практическим применениям. Целью работы является однообразное изложение операционных методов приспособленных к разным типам вопросов, на пример к решению дифференциальных и разностных уравнений с постоянными коэффициентами, уравнений типа Эйлера, дифференциально-разностных уравнений, уравнений Бернулли и других.

Обсуждаемый метод базируется на основных положениях функционального анализа. Многочисленные примеры помещённые в работе имеют целью познакомить инженерно-технических работников со способом практического использования метода.

S. BELLERT

OPERATIONAL CALCULUS IN LINEAR SPACES

Summary

1. The author presents a method constituting the generalization of the classical methods of operational calculus. The foundations of this method were published by the author in 1957 [1] and the present paper constitutes an extension of the method with respect to theoretical assumptions as well as to practical applications. The paper aims at a uniform treatment of operational methods adopted to different types of problems, for instance, to solving differential and difference equations with constant coefficients, Euler equations, difference-differential equations, Bernoulli equations and other equations.

Such a generalization can be obtained by assuming in every case different interpretations of the space $\mathbf{X}$ and the endomorphism T.

2. Let us assume that in a certain linear space $\mathbf{X}$ over a field $\mathbf{Z}$ of complex numbers, we are given the endomorphism T, and thus the operation satisfying the conditions

$$T(\mathbf{X}) \subset \mathbf{X}, \tag{1}$$

$$T[\alpha x_1 + \beta x_2] = \alpha T x_1 + \beta T x_2, \quad \alpha, \beta \in \mathbf{Z}, x_1, x_2, \in \mathbf{X}. \tag{2}$$

Moreover, let us assume that the set of operation results, $x_i = T^i x$, constitutes a system of elements linearly independent, that is,

$$\sum_{i=0}^{n} a_i T^i x = 0 \supset a_0 = a_1 = \ldots = a_n \quad \text{or} \quad x = 0, \tag{3}$$

where $T^0 \underset{\text{df.}}{=} I = 1$.

From the assumptions taken it follows that the operation Tx has no eigenvalues, that is,

$$Tx - \lambda x = 0 \supset x = 0.$$

3. Let us consider a set $\mathbf{P}$ of polynomial operators

$$W(T) = a_0 + a_1 T + \cdots + a_n T^n. \tag{4}$$

The set $\mathbf{P}$ — in the case of equality and algebraic operations determined analogically as for algebraic polynomials — forms a commutative ring. Such a ring has no zero divisors because of condition (3). (Theorem 1, p. 172).

4. It is then possible by a simple way to extend the ring P to a quotient field with the rational operators of the form

$$\frac{P(T)}{Q(T)} = \frac{a_0 + a_1 T + \ldots + a_n T^n}{\beta_0 + \beta_1 T + \ldots + \beta_m T^m}. \tag{5}$$

The result of the operation

$$\frac{P(T)}{Q(T)}(x) \tag{6}$$

is the element $y \in \mathbf{X}$ satisfying the equation

$$P(T)\,x = Q(T)\,y. \tag{7}$$

On account of condition (3) the result $\frac{P}{Q}\,x$ is determined uniquely, provided that there exists a solution of equation (7). (It is also possible to introduce a generalization in accord with the remark presented on p. 174). The operation $\frac{P}{Q}(T)x$, determined by means of the rational operator, is performable over the set $\mathbf{X}$ if and only if there exists such an element $x_0 \in \mathbf{X}$ that $\frac{P}{Q}T\,(x) = \frac{P^*}{Q^*}(T)x_0$, and the function $\frac{P^*}{Q^*}(z)$ of the complex variable z is holomorphic in the neighborhood of the point $z = 0$.

It may easily be found that, under the assumption of the existence of the corresponding results, the following equalities hold:

$$\frac{P}{Q}[\alpha x_1 + \beta x_2] = \alpha \frac{P}{Q} x_1 + \beta \frac{P}{Q} x_2. \tag{8}$$

$$\frac{P_1}{Q_1} x + \frac{P_2}{Q_2} x = \frac{P_1 Q_2 + P_2 Q_1}{Q_1 Q_2} x, \tag{9}$$

$$\frac{P_1}{Q_1}\left[\frac{P_2}{Q_2} x\right] = \frac{P_1 P_2}{Q_1 Q_2} x. \tag{10}$$

From these equalities it follows that rational operators may be treated as algebraic fractions, which means that the field of rational operators is isomorphic with the field of rational functions. Accordingly, all the laws of algebraic operations concerning franctions (commutative, associative and distributive laws) hold for rational operators. We also may avail ourselves of the theorem of the partial fraction expansion of the rational function, which is important for applications.

The rational operator (5) has thus a unique partial fraction expansion of the form

$$\frac{1}{(T - \alpha)^k}.$$

A rational operator of particular importance is that of the form

$$\frac{1}{T^k}.$$

This operator will also be denoted by the symbol T^{-k} or p^k. In connection with this denotation we obtain the relation:

$$T^k p^k = p^k T^k = 1. \tag{12}$$

The operators T^k and p^k constitute thus a pair of inverse operators.

5. Let $\mathbf{X}$ be a linear topological space, e. g. the Fréchet space L^*. Let us also assume the continuity of the operation T,

$$\lim_n T\,x_n = T \lim_n x_n, \tag{13}$$

and a linear independence of the operation results,

$$\sum_{i=0}^{\infty} a_i\,T^i\,x = 0 \equiv a_0 = a_1 = \ldots = a_i = \ldots 0 \quad \text{or} \quad x = 0 \tag{14}$$

and the following topology will be introduced to the field of the rational operators $\frac{P}{Q}(T)$.

The sequence of the rational operators $\frac{P_n}{Q_n}(T)$ is called a convergent sequence, and we write

$$\lim_n \frac{P_n}{Q_n}(T) = \frac{P}{Q}(T) \quad \text{or} \quad \frac{P_n}{Q_n}(T) \to \frac{P}{Q}(T) \tag{15}$$

if and only if the function sequence $\frac{P_n}{Q_n}(z)$ of the complex variable z is almost uniformly convergent to the rational function $\frac{P}{Q}(z)$, that is,

$$\frac{P_n}{Q_n}(T) \to \frac{P}{Q}(T) \equiv \frac{P_n}{Q_n}(z) \Rightarrow \frac{P}{Q}(z). \tag{16}$$

The sequence of the rational operators $\frac{P_n}{Q_n}(T)$ is called a fundamental sequence, if the sequence $\frac{P_n}{Q_n}(z)$ is almost uniformly convergent to the function $F(z)$ not necessarily rational

$$\frac{P_n}{Q_n}(z) \Rightarrow F(z). \tag{17}$$

The class of the fundamental sequences of rational operators, which has the same limit according to (17), *determines thus uniquely the element* $\frac{P_n}{Q_n}(T) \sim F(T)$ *of the new space* $\mathbf{R}_0$. *This element is given the name of real operator under the assumption of equality, algebraic operations and limit of a sequence in terms of the isomorphism with the functions* $F(z)$.

The result of the operation $F(T)x$ determined by the real operator $F(T) \sim \frac{P_n}{Q_n}(T)$ is called the element y in the linear topological space $\mathbf{X}$ which is determined by the formula

$$y = \lim_n y_n = \lim_n \frac{P_n}{Q_n}\,x; \qquad x, y \in \mathbf{X} \tag{18}$$

under the assumption that

$$\frac{P_n}{Q_n} \to F \supset \frac{P_n}{Q_n}\,x \to F\,x \in \mathbf{X}. \tag{19}$$

Conditions of the existence of the operation $F(T)x$ are specified in the paper for the case when $\mathbf{X}$ is a linear space locally convex with the topology determined by means of the sequence of pseudonorms $\|x\|_k$; $k = 1, 2, 3, \ldots$ that is non-negative functionals satisfying the conditions $i_1 \ldots i_4$ (p. 179).

Denoting

$$\| T \|_k = \sup_{\|x\|_k \leqslant 1} \| T x \|_k \tag{20}$$

we introduce the concept of the operator essentially bounded. The operator essentially bounded is thus the endomorphism T satisfying the following condition

$$\lim_i \sqrt[i]{\|T^i\|_k} = 0 \qquad k = 1, 2, \ldots \tag{21}$$

In the paper there is the following theorem.

Theorem 2. If the endomorphism T is an essentially bounded operator, and there exist such an element $x_0 \in \mathbf{X}$ that $\frac{P_n}{Q_n}(T)\,x = \frac{P_n}{Q_n}(T)x_0$ and if $\lim_{n\to\infty} \frac{P_n^*}{Q_n^*}(z)$ is a holomorphic function in the neighborhood of $z = 0$, then the operation $F(T)x$ determined in $\mathbf{X}$ by means of the real operator $F(T) \sim \frac{P_n}{Q_n}(T)$ is performable in the field $\mathbf{X}$. The proof of this theorem is given on page 180. An essentially bounded operator is for instance the Heaviside operator for which $x(t)$ is a set of continuous functions (or — more generally — integrable) on the real semi-axis T, and

$$T\,x(t) = \int_0^t x(\tau)\,\mathrm{d}\,\tau. \tag{22}$$

The pseudonorm $\|x\|_k$ in this case is determined by

$$\| x \|_k = \sup_{\left(\frac{1}{k},\, k\right)} |\, x(t) \,|; \qquad k = 1, 2, \ldots \tag{23}$$

and the condition that the operator T is essentially bounded results immediately from Cochy's integral (p. 180).

The convergence implied by the sequence of pseudonorms (23) is an almost uniform convergence of the function $x(t)$.

From formula (14) it follows that the set of real operations has no zero divisors, that is,

$$F(T)\,x = 0 \supset F(T) = 0 \quad \text{or} \quad x = 0. \tag{24}$$

The correctness of the relations analogous to relations (10) (p. 172) can easily be verified. To determine certain real operators we may also avail ourselves of Padé's approximation method.

6. The space $\mathbf{R}_0$ of the real operators F is a linear topological space, and thus it is possible to determine in it the endomorphism T_1 satisfying conditions (1) (2), (3) and, by analogy to the previous procedure, to construct the space $\mathbf{R}_1$ of the operators $F^{(1)}(T_1)$. If we assume appropriate conditions for the operator T_1 to ensure the feasibility of the operation, then this process could be continued to obtain the spaces $\mathbf{R}_2$, $\mathbf{R}_3$, ... $\mathbf{R}_n$, ... It is also possible to determine in the space $\mathbf{R}_0$ the differential operation satisfying conditions (50) on page 182 or another operation $S(G) = \mathbf{R}_0$, where G is a linear subspace of the space $\mathbf{R}_0$. An example

of the operation of differential type is the difference operation determined by the formula (51) on page 183, or the differential operation $F'(\lambda) = F'_\lambda (T, \lambda)$ determined over the field of functions of the real variable λ with values in the space $\mathbf{R}_0$ of operators. The derivative $F'(\lambda)$ is determined by the isomorphism with the derivative $F'_\lambda (z, \lambda)$. Such a procedure makes it possible to solve partial equations. (See example on p. 183).

7. It is also possible to consider generalized rational operators

$$\frac{P}{Q}(T) = \frac{A_0 + A_1 T + \ldots + A_n T^n}{B_0 + B_1 T + \ldots + B_m T^m}, \tag{25}$$

where A_i, B_i are endomorphisms commutative with respect to the endomorphism T. If we assume condition (23) from page 175, and condition (22) from page 175, then the ring of polynomial endomorphisms $W(T) = A_0 + A_1 T + \ldots + A_n T^n$ will have no zero divisors and consequently the operator (25) will be determined uniquely, provided that $Q(T) \neq 0$. Such a generalization makes it possible, e.g. to solve equations

$$x^{(n)}(t) + A_1 x^{(n-1)}(t) + \ldots + A_n x(t) = a(t) \tag{26}$$

where $x(t)$ and $a(t)$ are continuous functions of the real variable t with values in a linear topological space $\mathbf{X}$ and A_1, A_2, ..., A_n are continuous endomorphisms of $\mathbf{X}$. It is convenient to assume that the endomorphisms A_1, A_2, ... A_n are the elements of complete commutative algebra.

8. Let us consider the operational equation

$$y - F(T)\, y = 1 \tag{27}$$

determining the operation

$$y = \frac{1}{1 - F(T)}(1). \tag{28}$$

The application of the iterative process to (27) leads to the n-th iteration of the form

$$y_n = F^n[y_0] + F^{n-1}[1] + \ldots + F[1] + 1. \tag{29}$$

If, therefore, in the topological space $\mathbf{X}$ the iterative process (29) is convergent, then by assuming $y_0 = I = 1$ we obtain

$$\frac{I}{[1 - F(T)]^{k+1}} = \sum_{i=0}^{\infty} \binom{k+i}{i} F^i(T). \tag{30}$$

Formula (30) makes it possible to calculate the results of certain operations in an approximate way.

9. Applications

A. Let $\mathbf{X}$ be a linear space of functions $x(t)$ integrable in each interval $[0, t_0]$ and

$$T x(t) = \int_0^t x(\tau)\, d\tau. \tag{31}$$

If the function $x(t)$ has a continuous derivative $x^{(n)}(t)$, then from

$$T x'(t) = x(t) - x(0) \tag{32}$$

we shall obtain

$$x^{(n)}(t) = p^n x(t) - \sum_{k=0}^{n-1} x^{(k)}(0)\, p^{n-k}, \quad n = 1, 2 \ldots \tag{33}$$

The formula (33) permits to solve the linear differential equations with constant coefficients, similarly as the Laplace method and other operational methods. Using the well known Duhamel operational theorem we can also solve non homogeneous equations without any restrictions as to the spead of increase in function.

B. Let

$$T x(t) = \begin{cases} \int_1^t \frac{x(\tau)}{\tau}\, d\tau & \text{for } t \geqslant 1, \\ 0 & \text{for } t < 1. \end{cases} \tag{34}$$

If the function $x(t)$ has a continuous derivative $x^{(n)}(t)$ then from

$$T[t\, x'(t)] = x(t) - x(1) \tag{35}$$

we shall obtain

$$t^n x^{(n)}(t) = p(p-1) \ldots (p-n+1)\, x(t) - p(p-1) \ldots$$
$$\ldots (p-n+1)\, x(1) - \ldots - p(p-1)\, x^{(n-2)}(1) - p\, x^{(n-1)}(1). \tag{36}$$

Similarly as in A, formulas (36) permit to solve the Euler equation

$$a_n t^n x^{(n)}(t) + \ldots + a_1 t\, x'(t) + a_0 x(t) = \varphi(t),$$

where $a_n, a_{n-1}, \ldots, a_0$ are constant coefficients and $\varphi(t)$ is a integrable function for $t \geqslant 1$.

C. Let $\mathbf{X}$ be a space of number sequences $x(n)$ and

$$T x(n) = x(0) + \ldots + x(n-1).$$

From

$$T[\Delta x(n)] = x(n) - x(0), \tag{37}$$

where

$$\Delta x(n) = x(n+1) - x(n), \tag{37}$$

we have

$$\Delta^n x(n) = p^n x(n) - \sum_{k=0}^{n-1} \Delta^k x(0)\, p^{n-k}, \quad n = 1, 2, 3, \ldots \tag{38}$$

The formula (38) enables to solve difference equations with constant coefficients.

D. Let us consider the equation

$$n \Delta y(n) = x(n), \quad n = 1, 2, 3, \ldots \tag{39}$$

For a given initial condition, equation (39) uniquely determines the sequence $y(n)$.

This means that the formula

$$T x(n) = y(n) - y(1), \quad n \geqslant 1 \tag{40}$$

defines the endomorphism T.

If $p = T^{-1}$, we have

$$n \Delta x(n) = p\, x(n) - p\, x(1),$$
$$(n+1)\, n \Delta^2 x(n) = p(p-1)\, x(n) - p(p-1)\, x(1) - p \Delta x(1). \tag{41}$$

These relations enable to solve the difference equations

$$(n+k)\dots(n+1)\,n\,a_0\,\Delta^{k+1}x+\dots+(n+1)\,n\,a_{k-1}\,\Delta^2 x+n\,a_k\,\Delta x+a_{k+1}x=f(n), \tag{42}$$

where $a_0, \dots, a_{k+1}$ are arbitrary constans.

We observe that

$$\frac{p}{p-a}(1)=\prod_{m=1}^{n-1}\left(1+\frac{a}{m}\right). \tag{43}$$

E. Let $\mathbf{X}$ be a space of functions $x(t)$ defined for $t \geqslant 0$.
We consider the difference-differential equation

$$y'(t+1)=x(t). \tag{44}$$

For the initial condition

$$y(t)=f(t) \quad \text{for} \quad t\in[0,1],$$

Eq. (44) uniquely defines the function $y(t)$. This means that the formula

$$Tx(t)=y(t)-y(0,1), \quad t\geqslant 0, \tag{45}$$

where

$$y(0,1)=\begin{cases} f(t) & \text{for} \quad 0\leqslant t<1,\\ f(1)=\text{const} & \text{for} \quad t\geqslant 1,\end{cases}$$

uniquely defines the endomorphism T.

If $p = T^{-1}$, then we shall have

$$x^{(n)}(t+n)=p^n x(t)-\sum_{k=0}^{n-1}p^{n-k}\,x^{(k)}(k,k+1), \tag{46}$$

where $x^{(k)}(k, k+1)$ is the following function:

$$x^{(k)}(k,k+1)=\begin{cases} x^{(k)}(t) & \text{for} \quad k\leqslant t\leqslant k+1,\\ x^{(k)}(k+1)=\text{const} & \text{for} \quad t\geqslant k+1.\end{cases}$$

Rational operations can be calculated from (45). We observe that

$$\frac{p}{p-1}(1)=\frac{t^{[t]}}{[t]!}, \tag{47}$$

where $[t]$ denotes „Entier t".

We put

$$\frac{p}{p-a}(1)=P(a,t). \tag{48}$$

It can be proved that $P(a, t)$ is a continuous function defined by

$$P(a,t)=a^{[t]}\frac{t^{[t]}}{[t]!}-\sum_{n=0}^{[t-1]}a^n(a-1)\frac{t^n}{n!}. \tag{49}$$

Using (46) we can solve difference-differential equations

$$a_n x^{(n)}(t+n)+\dots+a_1 x'(t+1)+a_0 x(t)=0$$

by operational methods.

ANDRZEJ KOBUS

Podstawy geometrycznej teorii dyslokacji w kryształach

Rękopis dostarczono 24.4.1959

Artykuł wprowadza w podstawowe zagadnienia związane z obecnością w kryształach pewnego rodzaju nieprawidłowości, zwanych dyslokacjami. Na wstępie zostały omówione podstawowe pojęcia służące do ilościowego opisu dyslokacji, tj. obieg i wektor Burgersa, oraz modele geometryczne odwzorowujące zasadnicze rodzaje dyslokacji: krawędziową, śrubową i krzywoliniową (złożoną). Następnie omówiono geometryczne modele powstawania i zwielokratniania dyslokacji oraz zjawiska zachodzące w wyniku oddziaływania pól naprężeń wprowadzonych do sieci przez dyslokacje i atomy domieszkowe.

1. WSTĘP

Elektryczne właściwości materiału półprzewodnikowego są określone między innymi przez obecność szeregu nieprawidłowości w sieci krystalicznej. Najważniejszymi nieprawidłowościami, pomijając drgania termiczne sieci krystalicznej, są atomy innych pierwiastków, tzn. „domieszek", stanowiące o oporności właściwej, ruchliwości nośników mniejszościowych i czasie życia. Domieszki obcych atomów są nieprawidłowościami o charakterze punktowym. Rozwijające się intensywnie w ostatnich pięciu latach badania innego typu nieprawidłowości — będącym zaburzeniem periodycznej budowy sieci o charakterze liniowym (jednak nie przez wprowadzenie obcych atomów) — wykazały, że właściwości materiałów półprzewodnikowych zależą również od tego rodzaju nieprawidłowości. Nieprawidłowości te — zwane dyslokacjami — mają szczególnie duży wpływ na czas życia nośników mniejszościowych, co jest związane z wprowadzaniem przez dyslokacje centrów rekombinacji, które ograniczają czas życia.

Ze względu na wymiary nieprawidłowości struktury krystalicznej można podzielić na trzy grupy [1]:

1. zerowymiarowe — do których zalicza się atomy domieszkowe, atomy w pozycjach międzywęzłowych i luki,

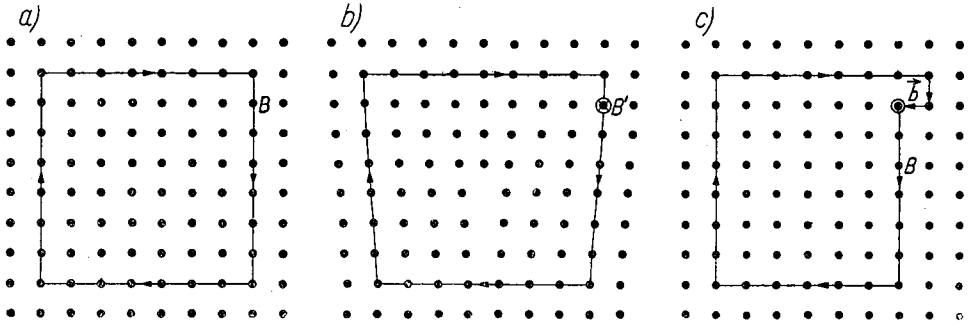
2. jednowymiarowe — to dyslokacje mogące mieć postać linii otwartych lub zamkniętych, powierzchnią oddzielającą obszar materiału idealnego od zakłóconego jest rurka o wymiarach atomowych.

3. dwuwymiarowe — są układami linii dyslokacyjnych, tworzących granicę ziarn[1]). Obszar zdeformowany jest płaszczyzną o grubości atomowej.

W artykule tym będą omawiane nieprawidłowości z drugiej i trzeciej grupy podanego podziału: tj. dyslokacje i ich układy. Zostaną podane również modele przyjęte do przedstawiania tych nieprawidłowości oraz podstawowe wiadomości o naprężeniach związanych z obecnością dyslokacji w krysztale. W dalszym ciągu artykułu zostaną omówione zjawiska związane z dyslokacjami, które mają pewien wpływ na elektryczną strukturę półprzewodników (szczególnie germanu i krzemu) lub będą mogły być wykorzystane do kontroli gęstości dyslokacji.

2. POJĘCIA WSTĘPNE

W celu ilościowego opisu nieprawidłowości liniowych kryształu Burgers [2] wprowadził obieg, zwany od jego nazwiska obiegiem Burgersa.



Rys. 1. Przykład obiegu i wektora Burgersa dla prostej sieci kubicznej
a) obieg Burgersa w krysztale idealnym, b) obieg Burgersa w krysztale nieidealnym, c) obieg Burgersa w krysztale nieidealnym sprowadzonym do idealnego; $\vec{b}$ jest wektorem Burgersa charakteryzującym nieprawidłowość

Obieg ten dla najprostszej sieci w układzie dwuwymiarowym pokazano na rys. 1a (*B*). Jest to linia zakmnięta łącząca sąsiednie atomy w sieci krystalicznej najkrótszymi odcinkami i przebiegająca w idealnym kry-

[1]) Przez granicę ziarn rozumiemy granicę oddzielającą obszary kryształu posiadające niejednakową orientację.

sztale. Na rys. 1b pokazano obieg Burgersa (B') przeprowadzony wokół obszaru zawierającego pewną nieprawidłowość sieci krystalicznej (w tym przypadku jest to dyslokacja krawędziowa, o której będzie mowa dalej). Jeżeli obszar objęty obiegiem Burgersa z rys. 1b(B') przekształcimy w obszar o sieci idealnej przez umieszczenie atomów w pozycjach, jakie zajmowałyby w krysztale idealnym, obieg Burgersa będzie otwarty (rys. 1c). Brakujący odcinek obiegu nazywany jest wektorem Burgersa $(\vec{b})$ i jest miarą zdeformowania obszaru objętego obiegiem Burgersa. Jak więc widać, jest to najmniejsza odległość między atomami w rozpatrywanej płaszczyźnie.

Określenie kierunku wektora Burgersa jest umowne. Jeżeli pewien dowolny kierunek linii dyslokacyjnej obierzemy za dodatni, to kierunek obiegu Burgersa przyjmuje się według reguły śruby prawoskrętnej, a wektor Burgersa skierowany jest od ostatniego punktu obiegu do pierwszego.

Dowodzi się, że jeżeli kilka dyslokacji styka się lub przecina się w pewnym punkcie, zwanym węzłem dyslokacji, to wektory Burgersa spełniają równanie

$$\sum_i \vec{b}_i = 0. \tag{1}$$

Jest to równanie analogiczne do zasady Kirchhoffa dla węzła prądów. Moduł wektora Burgersa nazywany jest natężeniem dyslokacji $|\vec{b}| = b$, i na ogół tą wielkością operuje się w szeregu obliczeń związanych z dyslokacjami.

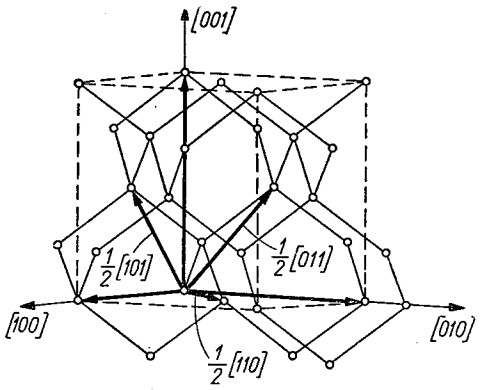
Tablica 1

Moduły wektorów Burgersa dla sieci diamentu

Kierunek wektora Burgersa	$< 110 >$	$< 100 >$
Długość wektora Burgersa a – stała sieci	$\frac{a\sqrt{2}}{2}$	a
b dla Ge ($a = 5{,}65$ Å [3])	4,00	5,65
b dla Si ($a = 5{,}46$ Å [3])	3,86	5,46

Wektor Burgersa może posiadać różne kierunki w zależności od własności aktualnie rozpatrywanej sieci krystalicznej. W przypadku sieci diamentu, w której krystalizują podstawowe materiały półprzewodnikowe (german, krzem, arsenek i antymonek indu), istnieją dwa rodzaje stabilnych wektorów Burgersa [1]; podane są one w tablicy 1. Każdy z rodzajów wektorów może posiadać trzy kierunki, jak na rys. 2. Wszyst-

kie pozostałe wektory Burgersa są niestabilne. Z dwóch podanych również pierwszy z nich, który zapisuje się jako $\frac{1}{2}\langle 110\rangle$, posiada znacznie większe prawdopodobieństwo powstania niż drugi, oznaczany (100),

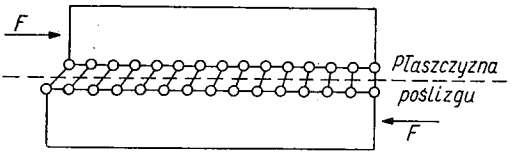


Rys. 2. Wektory Burgersa w sieci diamentu

ponieważ do utworzenia dyslokacji o krótszym wektorze poślizgu potrzebna jest mniejsza energia. Wobec tego omawiając sieć diamentu jako ośrodek, w którym powstają dyslokacje, wystarczy rozpatrywać dyslokacje z wektorem poślizgu równym $\frac{1}{2}\langle 110\rangle$, o czym będzie mowa w następnym rozdziale.

3. PODSTAWOWE TYPY DYSLOKACJI

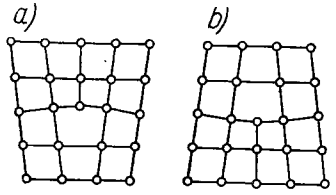
W rozdziale tym zostaną podane geometryczne sposoby przedstawiania dyslokacji. Historycznie biorąc, jako pierwsza do rozważań teorii odkształceń plastycznych została wprowadzona dyslokacja krawędzio-



Rys. 3. Mechanizm powstawania dyslokacji krawędziowej

wa (Taylor, Orowan, Polanyi — 1934 r.). Dyslokacja ta może powstawać w wyniku przyłożenia do kryształu naprężeń ścinających, jak pokazano na rys. 3. Dzięki działaniu naprężeń występuje tzw. poślizg kryształu polegający na przesunięciu części kryształu o odległości ato-

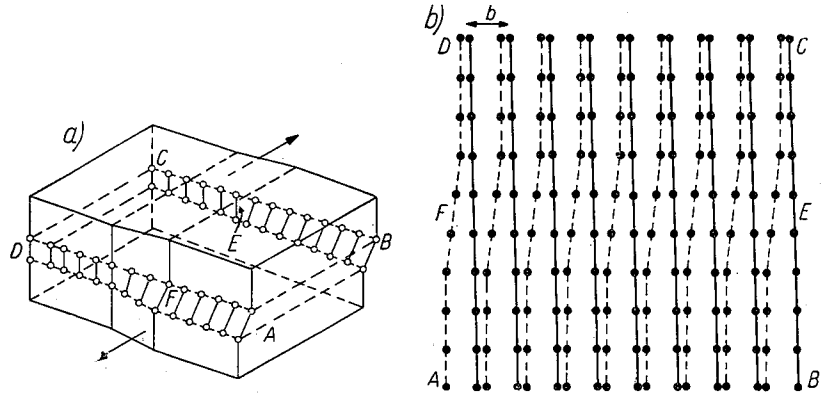
mowe względem pozostałej części kryształu. Płaszczyznę, wzdłuż której odbywa się ruch kryształu, nazywa się płaszczyzną poślizgu. Wskutek takiego przesunięcia powstaje zaburzenie sieci krystalicznej tworzące dyslokację krawędziową, którą w układzie dwuwymiarowym przedstawiono na rys. 1b. Dyslokacja ta jest walcem zdeformowanego obszaru, rozciągającego się w głąb kryształu, prostopadle do kierunku poślizgu. Dyslokację krawędziową określa się jako dodatnią, gdy do-



Rys. 4. Model krawędziowych dyslokacji dodatnich i ujemnych
a) dyslokacja dodatnia, b) dyslokacja ujemna

datkowa półpłaszczyzna atomów znajduje się wyżej płaszczyzny poślizgu (rys. 4a); jeżeli dodatkowa półpłaszczyzna znajduje się poniżej — dyslokacja nazywana jest ujemną (rys. 4b). Jest to definicja umowna, mająca na celu odróżnienie dwóch postaci tego samego rodzaju dyslokacji tak, że będzie ona również obowiązywała, gdy będziemy oglądać kryształ z odwrotnej strony. Jeżeli na jednej płaszczyźnie poślizgu znajduje się dyslokacja ujemna i dodatnia, w wyniku działania pewnych sił, o których będzie mowa dalej, mogą się one znieść.

Drugim typem dyslokacji jest dyslokacja śrubowa, której pojęcie wprowadził Burgers [2] w roku 1939. Powstaje ona również w wyniku

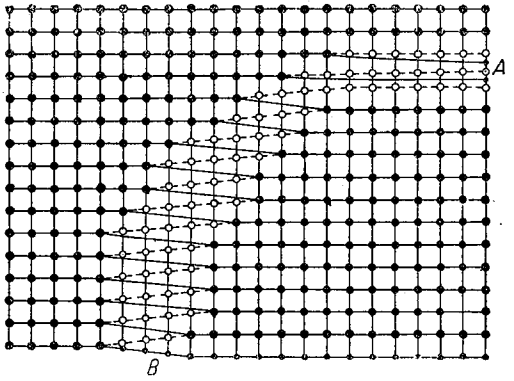


Rys. 5. Model dyslokacji śrubowej
a) mechanizm powstawania dyslokacji, b) układ atomów nad- i pod płaszczyzną poślizgu

naprężeń ścinających, ale w przypadku gdy linia dyslokacyjna jest równoległa do kierunku poślizgu. Mechanizm powstawania dyslokacji śrubowej pokazany jest na rys. 5a. Wskutek przyłożenia sił ścinających

zgodnie z kierunkami strzałek, prawa część kryształu ulega poślizgowi, którego płaszczyzna ograniczona jest punktami *ABEF*. W wyniku tego zjawiska wzdłuż linii *EF* rozciąga się obszar zakłócony, gdy na prawo od tej linii kryształ będzie przesunięty o jedną całą odległość międzyatomową, wobec czego nie będzie zniekształcony. Układ atomów w dyslokacji śrubowej pokazany jest na rys. 5b. Linią ciągłą narysowano atomy w górnej płaszczyźnie z dwóch zaznaczonych na rys. 5a, a linią przerywaną — dolną.

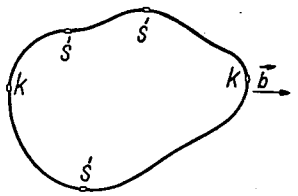
Burgers również udowodnił, że oba omawiane powyżej typy dyslokacji są szczególnymi przypadkami pewnej złożonej linii dyslokacyjnej



Rys. 6. Układ atomów nad- i pod płaszczyzną poślizgu w dyslokacji ogólnej

dowolnie przebiegającej w przestrzeni. Fakt ten jest zilustrowany na rys. 6, na którym pokazano linię dyslokacyjną o charakterze linii złożonej. Na rysunku tym linią ciągłą zaznaczono wiązania między atomami na płaszczyźnie ponad powierzchnią poślizgu, a linią przerywaną — wiązania na płaszczyźnie pod powierzchnią poślizgu.

W punkcie *A* nieprawidłowość sieci ma charakter dyslokacji krawędziowej, gdyż na dwa szeregi atomów na powierzchni górnej przy-

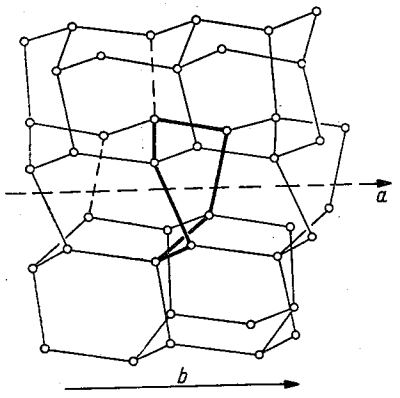


Rys. 7. Przykład ogólnej linii dyslokacyjnej. W punktach oznaczonych *k* dyslokacja posiada orientację krawędziową, a w punktach oznaczonych *ś* — śrubową

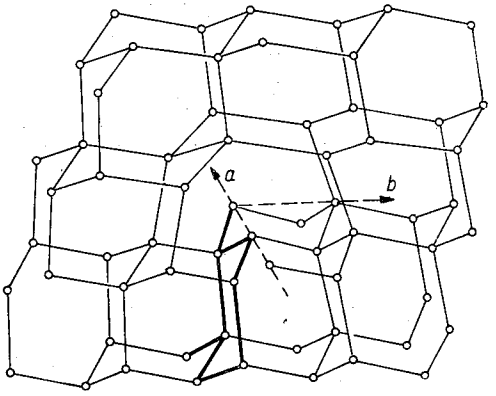
padają trzy szeregi atomów na powierzchni poniżej płaszczyzny poślizgu; w punkcie *B* nieprawidłowość posiada charakter dyslokacji śrubowej; na krzywiźnie linia dyslokacyjna ma charakter pośredni. Widzi-

my więc, że zakwalifikowanie dyslokacji do typu krawędziowego lub ślubowego zależy tylko od orientacji dyslokacji względem kierunku poślizgu. Wobec tego dowolna linia dyslokacyjna będzie miała dwa wyróżnione kierunki, w których będzie określona jako dyslokacja śrubowa lub krawędziowa; w pozostałych odcinkach będzie dyslokacją złożoną. Rysunek 7 przedstawia opisaną sytuację. Tylko w miejscach oznaczonych *k* lub *ś* dyslokacja będzie miała odpowiednio charakter krawędziowy lub śrubowy.

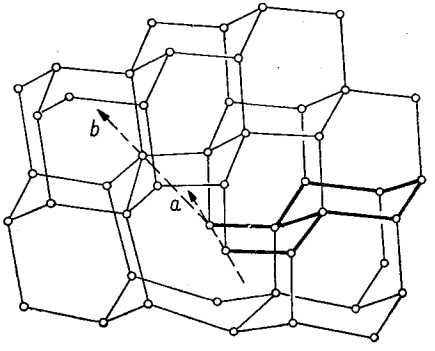
Ogólnie biorąc dyslokacje można określać przez podanie kąta, jaki tworzy linia dyslokacyjna z jej wektorem Burgersa. W przypadku sieci



Rys. 8. Model dyslokacji śrubowej w sieci diamentu [18]



Rys. 9. Model dyslokacji 60-stopniowej w sieci diamentu. Grubymi liniami zaznaczono dodatkową półpłaszczyznę atomów [18]



Rys. 10. Model dyslokacji krawędziowej w sieci diamentu [18]

diamentu, najbardziej nas interesującej, występują trzy rodzaje pełnych dyslokacji posiadających wektor Burgersa równy $\frac{1}{2}\langle 110\rangle$ [18]. Dwa

z nich to dyslokacje typowe: śrubowa, dla której omawiany kąt α jest równy 0°, i krawędziowa — z kątem $\alpha = 90°$. Trzeci rodzaj dyslokacji jest charakterystyczny dla sieci diamentu — jest to tzw. dyslokacja 60-stopniowa. Wszystkie te dyslokacje są równoległe do kierunku ⟨110⟩; ich układ w sieci diamentu pokazano na rysunkach 8, 9, 10. W tablicy 2 podano zestawienie omawianych dyslokacji.

Tablica 2

Dyslokacja w sieci diamentu

Lp.	Symbol osi	Kąt α	Płaszczyzna poślizgu	Ilość przerwanych wiązań na a cm
1	$<110>$	0°	—	0
2	$<110>$	60°	{111}	1,41
3	$<110>$	90°	{100}	2,83 lub 0

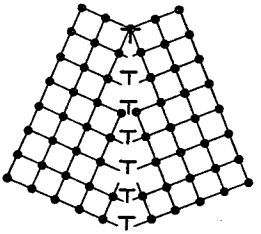
Ponieważ w sieci diamentu najgęściej upakowanymi płaszczyznami atomowymi są płaszczyzny {111}, więc one są płaszczyznami łatwego poślizgu. Wobec tego, jak widać z tablicy 2, w kryształach o sieci diamentu będą uprzywilejowane dyslokacje 60-stopniowe, gdyż leżą one w płaszczyźnie łatwego poślizgu.

W ostatniej rybryce tablicy 2 podano ilość przerwanych wiązań atomowych spowodowanych istnieniem dyslokacji o różnych orientacjach. Wielkość ta określa elektryczne właściwości dyslokacji, gdyż istnienie zerwanego wiązania atomowego stanowi zaburzenie w równowadze elektrycznej kryształu. Wobec tego dyslokacje wprowadzające zerwane wiązania atomowe będą miały wpływ na przewodnictwo elektryczne, a szczególnie na rekombinację nośników w kryształach półprzewodnikowych. Zagadnienie to jednak nie mieści się w ramach niniejszej publikacji.

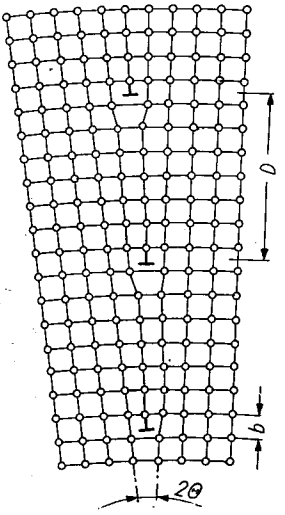
4. DWUWYMIAROWE UKŁADY DYSLOKACJI

W rozdziale tym będą omówione zbiory dyslokacji ułożone w pewien określony sposób. Będą to różne rodzaje granic ziarn w krysztale. Shockley i Read [4] podają dwa sposoby połączenia dwóch ziarn na ich granicy. W pierwszym przypadku kryształ pozostaje nie zniekształcony, a na samej granicy ziarn płaszczyzny atomowe są nie dopasowane (rys. 11). W drugim przypadku może mieć miejsce elastyczne zniekształcenie obu ziarn, które rozciąga się na stosunkowo duże odległości, gdy deformacja plastyczna ogranicza się do małych obszarów na granicy ziarn (zob. rys. 12). Oczywiście pierwszy przypadek będzie odpowied-

niejszym modelem dla granicy ziarn o dużym kącie, gdy drugi — dla granicy ziarn o małym kącie. Szczególnie interesujący, zarówno z punktu widzenia teorii dyslokacji, jak i doświadczenia, jest przypadek granicy ziarn o małym kącie (ang. low-angle boundary). Dyslokacyjny model granicy ziarn o małym kącie został omówiony przez Burgersa [2],



Rys. 11. Model granicy ziarn o dużym kącie



Rys. 12. Model symetrycznej granicy ziarn o małym kącie

Shockleya i Reada w r. 1949 [5]. Prosta granica ziarn o małym kącie jest wynikiem skręcenia dwóch obszarów kryształu o pewien mały kąt. Charakteryzuje się ona tym, że dodatkowe półpłaszczyzny atomowe posiadają ten sam kierunek i zwrot, tzn. są „wsunięte" z tej samej strony. W wyniku tego na granicy skręconych obszarów powstaje szereg dyslokacji krawędziowych. Na rys. 12 pokazano przykład prostej granicy ziarn. Dla kryształu o określonej sieci krystalicznej można wyznaczyć zależność odległości między dyslokacjami krawędziowymi w granicy ziarn a kątem skręcenia obszarów kryształu.

Jeżeli najmniejszą odległość między atomami w rozpatrywanej płaszczyźnie oznaczymy przez b (tj. moduł wektora Burgersa), a odległości między dyslokacjami przez D, to:

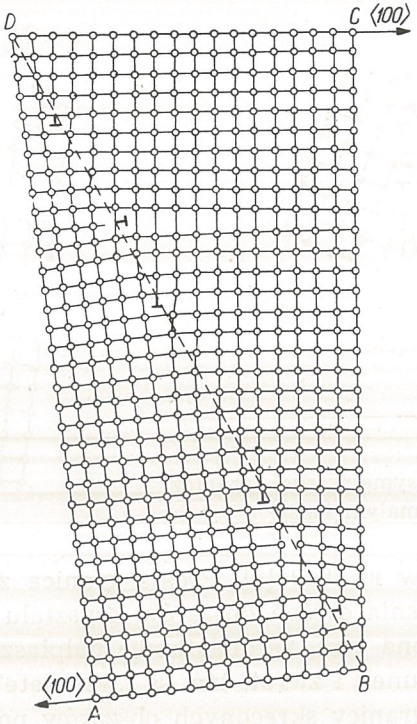
$$\frac{b}{D} = 2 \sin \frac{\Theta}{2}, \tag{2}$$

gdzie Θ jest kątem skręcenia obszarów kryształu po obu stronach granicy. Równanie to dla małych kątów, a takie rozpatrujemy z założenia, można napisać:

$$\Theta = \frac{b}{D}. \tag{3}$$

Rozpatrywany model granicy ziarn jako szereg dyslokacji jest jedynym geometrycznym sposobem przedstawiania granicy ziarn.

Read [6] rozpatruje również ogólniejszy przypadek granicy ziarn, mianowicie niesymetryczną granicę ziarn. Różni się ona tym od prostej granicy ziarn, że istnieją dwie grupy dodatkowych półpłaszczyzn ato-



Rys. 13. Model niesymetrycznej granicy ziarn o małym kącie

mowych skierowanych prostopadle do siebie. Model takiej granicy przedstawiono na rys. 13. Jeżeli jak na rys. 13 kąt między powierzchnią granicy i średnim kierunkiem ⟨100⟩ oznaczony jest przez φ, to z analizy geometrycznej otrzymamy dwie zależności na dwa ciągi dyslokacji:

$$D_{\perp} = \frac{b}{\Theta \sin \varphi}, \tag{4}$$

$$D_{\vdash} = \frac{b}{\Theta \cos \varphi}. \tag{5}$$

W przypadku ogólnym, jeżeli $0 < \varphi < \frac{\pi}{2}$, to otrzymujemy granicę ziarn niesymetryczną, a gdy $\varphi = 0$ lub $\varphi = \frac{\pi}{2}$, otrzymujemy prostą granicę ziarn.

Granice ziarn w kryształach stanowią dogodny materiał do badania właściwości układów dyslokacji, gdy badania pojedynczych linii dyslokacyjnych nastręczają często trudności ze względu na niewielki wpływ tych nieprawidłowości na szereg zjawisk. Szczególnie dobrze do tego celu nadają się granice ziarn o małym kącie, gdzie istnieje wyraźna struktura dyslokacyjna. Wyniki pomiarów właściwości granic ziarn o dużym kącie nastręczają znacznie większe trudności interpretacyjne.

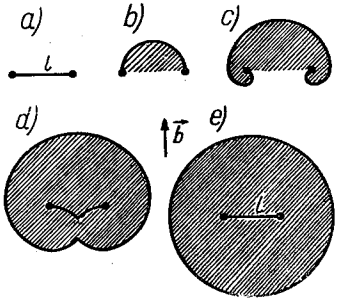
5. POWSTAWANIE I ZWIELOKRATNIANIE DYSLOKACJI

Ogólnie biorąc dyslokacje powstają w wyniku działania naprężeń istniejących w krysztale. Naprężenia w krysztale mogą się tworzyć wskutek nierównomierności rozkładu temperatury w czasie wykonywania kryształu oraz ewentualnych wstrząsów mechanicznych. Pomimo że w praktyce nauczono się otrzymywać monokryształy germanu i krzemu o bardzo wysokiej doskonałości, to jednak mechanizm powstawania dyslokacji nie został zupełnie dokładnie wyjaśniony.

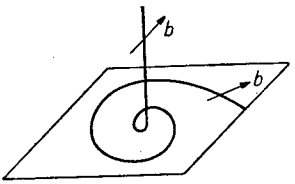
Istnieją dwie hipotezy dotyczące powstawania dyslokacji w kryształach. Pierwsza z nich głosi, że dyslokacje powstają na granicy fazy ciekłej i stałej i następnie są „zamrażane" w fazie stałej. Druga hipoteza mówi, że dyslokacje powstają dopiero w fazie stałej kryształu w wyniku nierównomiernego chłodzenia kryształu. Wydaje się, że obie hipotezy są słuszne i mają swój „zakres działania". W tej chwili brak jest dostatecznych wiadomości na temat słuszności pierwszej hipotezy, ze względu na duże trudności przeprowadzenia badań. Natomiast druga hipoteza została dość szeroko rozwinięta i poparta doświadczeniem.

Wobec tego, że w czasie deformacji plastycznej kryształu stwierdzono, że na typowej płaszczyźnie poślizgu jest ok. 1000 razy więcej elementarnych poślizgów, tzn. przesunięć o jedną odległość, niż wynikałoby to z przejścia pojedynczej dyslokacji, Frank i Read [7] w roku 1950 zaproponowali wyjaśnienie tego zjawiska, zwanego mechanizmem zwielokratniania dyslokacji. Rozwinięcie tego zagadnienia podał również Read w swojej książce [6]. Mechanizm zaproponowany przez Franka-Reada, zwany też źródłem Franka-Reada, omawia ruchy linii dyslokacyjnych, które zostały zakotwiczone, tj. unieruchomione w jednym lub więcej punktach. Jako punkty zakotwiczenia dla dyslokacji mogą służyć węzły dyslokacyjne, powierzchnia boczna kryształu lub obce atomy. Na rys. 14 pokazany jest mechanizm poślizgu podany przez Franka-Reada. Obszar kryształu zawierający dyslokację, której część unieruchomiona, o długości l, jest na rysowana na rys. 14a, w pewnej części swej objętości ulega poślizgowi, wskutek czego występuje ruch linii dyslokacyjnej

(rys. 14b). Obszar zacieniony stanowi powierzchnię, na której odbył się poślizg kryształu. Na rys. 14c pokazano kolejną fazę w procesie poślizgu i następnie, zakotwiczony odcinek dyslokacji „wyemituje" jedną



Rys. 14. Mechanizm poślizgu podany przez Franka-Reada, a-e poszczególne fazy poślizgu



Rys. 15. Przykład powstawania dyslokacji spiralnej w wyniku obrotu dyslokacji typu L dookoła jednego z ramion

pętlę dyslokacyjną, jak na rys. 8d. W dalszej fazie (rys. 14e) układ jest gotowy do powtórzenia całego mechanizmu, a dalszy ruch dyslokacji spowoduje poślizg kolejnych obszarów kryształu o drugą odległość międzyatomową. Mechanizm ten może powtarzać się wielokrotnie, w wyniku czego może powstać z jednego odcinka dyslokacyjnego szereg pętli dyslokacyjnych.

Jeżeli linia dyslokacyjna będzie zakotwiczona w jednym punkcie w wyniku działania źródła Franka-Reada, powstanie dyslokacja spiralna pokazana na rys. 15. Frank i Read [7] wykazali, że źródło dyslokacji zaczyna działać, jeżeli składowa naprężenia ścinającego wzdłuż wektora Burgersa odcinka dyslokacji zakotwiczonego przez węzły dyslokacyjne jest równa lub większa od:

$$\tau = \frac{a \cdot G \cdot b}{l}, \tag{6}$$

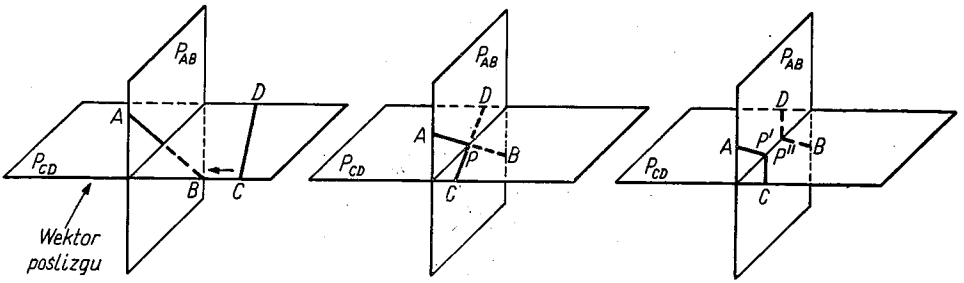
gdzie:

a — współczynnik liczbowy zależny od stałej elastyczności materiału i lekko zależny od rodzaju dyslokacji jest rzędu jedności,
G — moduł ścinania,
b — moduł wektora Burgersa,
l — długość odcinka zakotwiczonego.

Szczególne przypadki źródeł Franka-Reada były badane przez kilku autorów. I tak, Fisher [wg 8] stwierdził, że odcinek linii dyslokacyjnej zakończony z jednej strony na węźle, a z drugiej na powierzchni może działać jako źródło dyslokacji przy naprężeniu dwukrotnie mniejszym

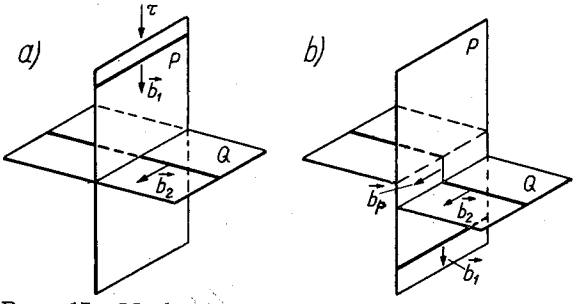
niż w przypadku zakotwiczenia dwustronnego na węzłach. Hart [9] badał przypadek źródeł Franka-Reada zakotwiczonych nie przez węzły dyslokacyjne, a przez atomy domieszkowe i stwierdził, że źródła te również mogą być „uaktywnione" przy niższych naprężeniach niż wynikające z równania [6].

Również interesujący jest mechanizm powstawania dyslokacji typu L podany przez Reada [6]. Rozpatruje się dwie dyslokacje AB i CD



Rys. 16. Przykład powstawania dyslokacji typu L z przecięcia dwóch dyslokacji liniowych

położne na różnych płaszczyznach poślizgu jak na rys. 16a, a posiadające taki sam wektor Burgersa. Dyslokacje poruszają się w kierunkach oznaczonych strzałkami. Przy przecięciu się dyslokacji (rys. 16b) powstaje węzeł dyslokacyjny w punkcie P. Gdy siły działają na dyslokacje w dalszym ciągu, suma długości dyslokacji może się zmniejszać jak na rys. 16c. Powstają dwie dyslokacje typu L. Dyslokacje takie mogą

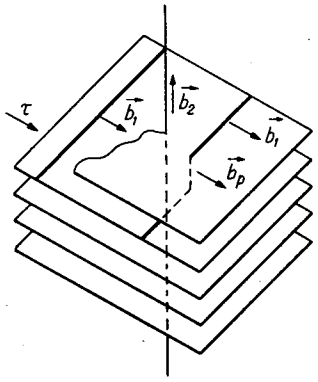


Rys. 17. Mechanizm powstawania progów przy przecięciu dwóch dyslokacji krawędziowych

być następnie źródłem powstawania dyslokacji spiralnych, które powstają przez obrót jednego z ramion dyslokacji typu L w sposób analogiczny do źródła Franka-Reada jak na rys. 15.

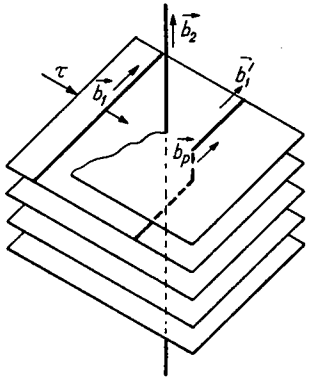
Omówimy teraz powstawanie progów (ang. jogs) w wyniku przecięcia się dwóch dyslokacji. Ruch dyslokacji może odbywać się w wyniku naprężeń termicznych lub naprężeń zewnętrznych przyłożonych do kry-

ształu. Rozpatrzymy najpierw przecięcie się dwóch dyslokacji krawędziowych o wektorach Burgersa $\vec{b}_1$ i $\vec{b}_2$, leżących na dwóch różnych płaszczyznach poślizgu P i Q. Sytuacja ta jest przedstawiona na rys. 17. Jak widać, na dyslokacji drugiej powstanie próg o długości $\vec{b}_1$, którego wektor Burgersa pozostanie bez zmiany, wobec czego próg będzie posiadał orientację krawędziową. Ponieważ kierunek wektora Burgersa



Rys. 18. Mechanizm powstawania progów przy przecięciu dyslokacji krawędziowej ze śrubową

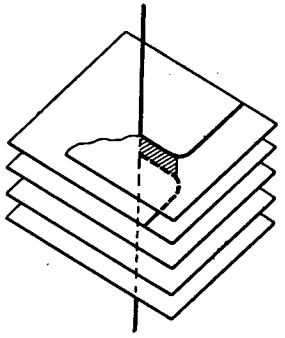
progu $\vec{b}_P$ jest zgodny z kierunkiem poślizgu drugiej dyslokacji $\vec{b}_2$, próg w zasadzie nie będzie przeszkadzał poślizgowi tej dyslokacji. Na następnym rysunku (rys. 18) pokazano sytuację przy przecięciu dysloka-



Rys. 19. Mechanizm powstawania progów przy przecięciu dwóch dyslokacji śrubowych

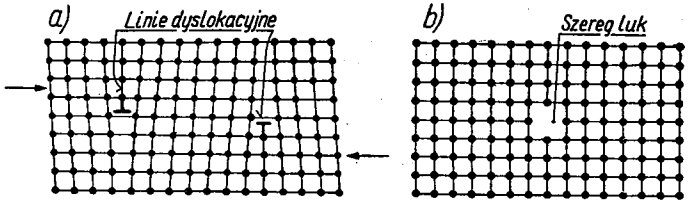
cji śrubowej z krawędziową. Dyslokacja krawędziowa o wektorze Burgersa $\vec{b}_1$ przecinając się ze śrubową tworzy próg o długości $\vec{b}_2$. Wektor poślizgu utworzonego progu pozostaje bez zmian — $\vec{b}_1$. W tym przy-

padku próg ma również orientację krawędziową i kierunek jego jest zgodny z kierunkiem przyłożonej siły powodującej ruch dyslokacji, wobec tego próg również nie przeszkadza poślizgowi dyslokacji krawędziowej. Natomiast w przypadku przecięcia się dwóch dyslokacji śrubowych (rys. 19) sytuacja jest bardziej skomplikowana. W wyniku przecięcia na jednej z dyslokacji (pierwszej) powstaje próg o długości $\vec{b}_2$. Ponieważ kierunek wektora Burgersa progu $\vec{b}_P$ jest prostopadły do kie-



Rys. 20. Deformacja jednej z przecinających się dyslokacji śrubowych

runku siły powodującej poślizg τ, to próg przeszkadza poślizgowi. Spowoduje to zniekształcenie prostoliniowej dyslokacji w sposób podany na rys. 20. Wykrzywione części dyslokacji zmieniają swoją orientację i w końcowej fazie będą odcinkami dyslokacji krawędziowych równoległych do siebie (zakreskowany obszar na rys. 20) i posiadających



Rys. 21. Powstawanie szeregów luk ze zdeformowanej dyslokacyjnej linii śrubowej

przeciwny znak, jak na rys. 21. Jeżeli dyslokacje te spotkają się, powstanie szereg luk lub atomów międzywęzłowych.

Wszystkie omawiane powyżej mechanizmy powstawania, zwielokrotniania i przemiany dyslokacji były potwierdzone doświadczalnie. W przypadku kryształów z materiałów półprzewodnikowych zarówno germanu, jak i krzemu szczególnie dogodne warunki do działania wyżej omówionych mechanizmów istnieją w zakresie temperatur, w których materiały te są plastyczne. Ma to miejsce dla germanu w temperaturach od 500°C do temperatury topnienia (937°C), a dla krzemu —

od 800°C do 1420°C. W kryształach półprzewodnikowych dyslokacje formują się ostatecznie w podanych wyżej zakresach temperatur plastycznych, gdy w temperaturze pracy przyrządów półprzewodnikowych można je uznać za praktycznie niezmienne. Doświadczalne wykazanie obecności takich form dyslokacyjnych zostanie omówione w następnym artykule [10].

6. NAPRĘŻENIA W OTOCZENIU DYSLOKACJI

Dyslokacja wprowadzając zaburzenie periodycznej budowy sieci krystalicznej powoduje również powstawanie pola naprężeń w krysztale. Deformację kryształu w wyniku wprowadzenia dyslokacji dzieli się na dwa zasadnicze obszary: 1) obszar deformacji plastycznej znajdujący się w ośrodku dyslokacji, w którym nie obowiązuje prawo Hooke'a, rozciągający się w promieniu kilku odległości międzyatomowych, oraz — 2) obszar deformacji elastycznej, który rozciąga się na znacznie większe odległości, rzędu nawet 1 cm. Ośrodek dyslokacji jest jak dotychczas mało znany i większość wielkości z nim związanych jest tylko szacowana. Obszar deformacji plastycznej przyjmuje się w szeregu rozważaniach za przekrój czynny dyslokacji i najczęściej szacuje się na trzy odległości międzyatomowe [17]. Energia zgromadzona w ośrodku dyslokacji również nie jest znana, jednak jak wynika z szacowania dokonanego przez Cottrella [11] wynosi ona $\frac{Gb^3}{3}$. Natomiast ogólną energię dyslokacji w krysztale wyznaczyli Leibfried i Lücke [12] w 1949 roku. Otrzymali oni następujący wzór na energię dyslokacji:

$$E \approx \frac{Gb^2}{4\pi(1-\nu)} \ln \frac{R}{r_l}, \tag{7}$$

gdzie:

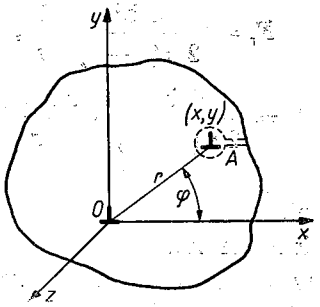
E — energia dyslokacji na jednostkę jej długości,
G — moduł ścinania,
b — moduł wektora Burgersa,
ν — stała Poissona,
r_l — promień obszaru deformacji plastycznej,
R — promień obszaru deformacji elastycznej.

Wzór ten jest ścisły dla dyslokacji śrubowej, a dla dyslokacji krawędziowej — przybliżony. Jak widać, energia dyslokacji jest zależna od parametrów elastycznych i geometrycznych materiału oraz wielkości zdeformowanych obszarów, których określenie nastręcza duże trudności. Porównując wzór (7) z wyrażeniem na energię ośrodka dyslokacji oszacowanym przez Cottrella łatwo można zauważyć, że energia ośrodka dyslokacji jest do pominięcia.

Naprężenia w obszarze deformacji elastycznej zmieniają się odwrotnie proporcjonalnie do odległości od środka dyslokacji. Obszar deformacji plastycznej nie podlega powyższemu prawu, gdyż naprężenia musiałyby osiągnąć wartość nieskończenie wielką dla odległości zerowej. Naprężenia te należałoby określić przez oddziaływanie między sąsiednimi atomami, które są odchylone od położeń równowagi w wyniku deformacji. Jak już podano wyżej, zagadnienie to przedstawia duże trudności. W następnych rozdziałach zostaną omówione następstwa istnienia pola naprężeń wokół dyslokacji: siły oddziaływania między dyslokacjami oraz między dyslokacją i atomami domieszkowymi.

7. ODDZIAŁYWANIE MIĘDZY DYSLOKACJAMI

Gdy będzie rozpatrywany kryształ zawierający dyslokacje, to nawet przy braku naprężeń zewnętrznych wystąpi działanie pewnych sił między dyslokacjami. Oddziaływania te są omówione przez Cottrel-



Rys. 22. Model stosowany do wyznaczania sił między dyslokacjami krawędziowymi

la [11] i Reada [6]. Rozważymy najpierw przypadek oddziaływania między równoległymi dyslokacjami krawędziowymi o równoległych i jednakowych wektorach Burgersa. Przyjmiemy układ jak na rys. 22. Przez punkt O wzdłuż osi Z przebiega dyslokacja krawędziowa o wektorze Burgersa b. Do tego obszaru wprowadzono drugą dyslokację przez deformację ośrodka wzdłuż powierzchni A, która jest równoległa do płaszczyzny poślizgu. W wyniku tego w punkcie o współrzędnych (x, y) znajduje się druga dyslokacja krawędziowa równoległa do pierwszej. Energia oddziaływania między rozpatrywanymi dyslokacjami będzie równa pracy wykonanej przez wprowadzenie drugiej dyslokacji w pole naprężeń pierwszej. Energię oddziaływania można wyznaczyć z następującej całki:

$$E_{12} = \int_A (\vec{F}_1 \cdot \vec{u}_2)\, dA, \tag{8}$$

gdzie:

$\vec{E}_{12}$ — energia oddziaływania między dyslokacjami,

$\vec{F}_1$ — siła związana z polem naprężeń pierwszej dyslokacji, działająca na powierzchnię A,

$\vec{u}_2$ — przesunięcie powierzchni A spowodowane wprowadzeniem drugiej dyslokacji.

Cottrell [11] otrzymał następujące wyrażenia dla składowych sił oddziaływania między równoległymi dyslokacjami krawędziowymi. Dla współrzędnych prostokątnych:

$$F_x = \frac{G b^2}{2\pi(1-\nu)} \cdot \frac{x(x^2-y^2)}{(x^2+y^2)^2}, \tag{9}$$

$$F_y = \frac{G b^2}{2\pi(1-\nu)} \cdot \frac{y(3x^2+y^2)}{(x^2+y^2)^2}. \tag{10}$$

Natomiast we współrzędnych cylindrycznych:

$$F_r = \frac{G b^2}{2\pi(1-\nu)} \cdot \frac{1}{r}, \tag{11}$$

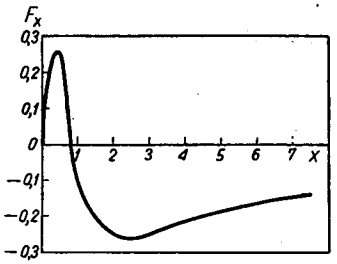
$$F_\varphi = \frac{G b^2}{2\pi(1-\nu)} \cdot \frac{\sin 2\varphi}{r}, \tag{12}$$

gdzie:

G — moduł ścinania,

ν — współczynnik Poissona.

Siły działające między tymi dyslokacjami nie są centralne. Z równania (11) widać, że siła istniejąca między dwoma dyslokacjami działa odpy-



Rys. 23. Wykres zależności składowej F_x (siły oddziaływania między dyslokacjami krawędziowymi) w funkcji odległości między dyslokacjami

chająco wzdłuż linii łączącej je, ale ruch dyslokacji będzie możliwy tylko wzdłuż płaszczyzny poślizgu, tj. równolegle do płaszczyzny xz. Na rys. 23 pokazano zależność składowej F_x od odległości x. Jednostką siły jest wyrażenie $\frac{G b^2}{2\pi(1-\nu)y}$, gdzie y jest odległością między płaszczyznami po-

ślizgu. Z rysunku tego widać, że dwie dyslokacje znajdujące się na różnych płaszczyznach poślizgu odpychają się wzdłuż płaszczyzn poślizgu, gdy $x>y$, a przyciągają się dla $x<y$. Położenia równowagi istnieją dla $x=0$ i $x=y$, lecz tylko pierwsze położenie jest pozycją równowagi trwałej. Z tego wynika, że najbardziej trwały układ dyslokacji ma miejsce, gdy jednoimienne dyslokacje krawędziowe zajmują pozycje jedna nad drugą, prostopadle do kierunku poślizgu. Przypadek ten zachodzi przy granicy ziarn o małym kącie.

Jeżeli będzie się rozpatrywać dwie dyslokacje przeciwnych znaków, przyciąganie będzie miało miejsce dla $x>y$, a odpychanie dla $x<y$, to znaczy odwrotnie niż w przypadku poprzednim. Równowaga trwała ma miejsce dla $x=y$. Natomiast dwie jednoimienne dyslokacje krawędziowe znajdujące się na tej samej płaszczyźnie poślizgu odpychają się dla wszystkich współrzędnych, a różnoimienne przyciągają.

Tablica 3

Zależność siły oddziaływania między dyslokacjami dla różnych odległości

Odległość między dyslokacjami (cm)	10^{-7}	10^{-5}	10^{-3}
Siła na jedn. długości dyslokacji (dyn/cm^2)	600	6	0,06

Cottrell [wg 13] obliczył siły działające między dyslokacjami dla przypadku prostej sieci kubicznej. Wyniki obliczeń podano w tablicy 3.

Gdy na jednej płaszczyźnie poślizgu znajdują się dwie dyslokacje jednoimienne, w warunkach sprzyjających ruchowi dyslokacji (gdy siła działająca między dyslokacjami będzie większa od sił wiązania ośrodka) dyslokacje mogą poruszać się i w efekcie końcowym unicestwić. W tym procesie wyzwala się energia powodująca lokalne ogrzanie kryształu do 1000°C w czasie 10^{-12} sec [wg 13].

W przypadku rozpatrywania sił występujących między dwoma równoległymi dyslokacjami śrubowymi postępowanie jest analogiczne do poprzedniego. Zasadnicza różnica polega na tym, że siły działające między dwoma dyslokacjami śrubowymi są centralne i wskutek tego w układzie cylindrycznym mają tylko składową promieniową. Dla tego przypadku Cottrell [11] otrzymał następujące wyrażenie dla układu cylindrycznego:

$$F_r = \frac{Gb^2}{2\pi r}, \tag{13}$$

a dla układu prostokątnego:

$$F_x = \frac{Gb^2}{2\pi} \cdot \frac{x}{(x^2+y^2)}, \tag{14}$$

$$F_y = \frac{Gb^2}{2\pi} \frac{y}{(x^2+y^2)}. \tag{15}$$

Dyslokacje śrubowe różnych znaków przyciągają się, a jednakowych — odpychają.

Cottrell [11] omówił również siły przyciągania dyslokacji do swobodnej powierzchni kryształu. Z przybliżaniem dyslokacji do powierzchni energia deformacji maleje, gdyż pole naprężeń dyslokacji przenika przez powierzchnię kryształu. W wyniku tego powstaje siła urojona taka, jakby w swobodnej przestrzeni po drugiej stronie powierzchni istniała dyslokacja o znaku przeciwnym. Siła ta w kierunku prostopadłym do powierzchni wyraża się wzorem:

$$F = \frac{Gb^2}{4\pi(1-\nu)} \cdot \frac{1}{r}. \tag{16}$$

Jak widać, jest to postać siły jak przy oddziaływaniu między dyslokacjami, a wielkość jej jest dwukrotnie mniejsza.

Opisane powyżej zjawiska ukazują potencjalną możliwość kontroli gęstości dyslokacji przez odprężanie kryształów. Jednak na temat tego zagadnienia na razie brak obszerniejszych publikacji.

8. ODDZIAŁYWANIE ATOMÓW DOMIESZKOWYCH Z DYSLOKACJĄ

Badając zagadnienie starzenia w metalach zawierających pewne ilości domieszek Cottrell [11] stwierdził istnienie oddziaływania atomów domieszkowych z polem naprężeń dyslokacji. Jak podawano w rozdziale 6, energia dyslokacji zmienia się z odległością od środka nieprawidłowości. Oddziałując z innymi ośrodkami naprężeń w krysztale (a będą to głównie atomy domieszkowe o wymiarach różnych od atomów ośrodka) rozkład energii dyslokacji spowoduje wystąpienie sił, które będą wywoływać segregację atomów domieszkowych w kierunku środka dyslokacji. Atomy te mogą powodować rozluźnienie naprężeń istniejących wokół dyslokacji. Cottrell [11] i [14] podał wzór na energię oddziaływania między liniową dyslokacją krawędziową znajdującą się w punkcie o współrzędnych (0,0) a atomem domieszkowym znajdującym się w punkcie o współrzędnych (R, α) (współrzędne cylindryczne) w następującej postaci:

$$E_I = \frac{4}{3} \cdot \frac{1+\nu}{1-\nu} \cdot \frac{Gb\varepsilon r^2 \sin\alpha}{R}, \tag{17}$$

gdzie:

r — promień atomu ośrodka,

r' — promień atomu domieszki,

$$\epsilon = \frac{r' - r}{r},$$

ν — stała Poissona,

G — moduł ścinania.

Dla atomów domieszkowych o promieniu większym niż atomów ośrodka ($\varepsilon > 0$) energia oddziaływania jest dodatnia nad dyslokacją, a ujemna pod nią. Wobec tego atomy większe od atomów ośrodka będą odpychane od górnej części dyslokacji, przyciągane zaś do dolnej. Dla atomów mniejszych będzie sytuacja odwrotna. Energia oddziaływania, a więc i siła przyciągania domieszek, jak to widać ze wzoru (17), maleje z odległością. W tablicy 4 [wg 15] podano energię oddziaływania dyslokacji z atomami domieszkowymi w germanie i krzemie. Wartość energii w elektronowoltach podano dla odległości 4 Å od środka dyslokacji.

Tablica 4

Energia oddziaływania między dyslokacjami a atomami domieszkowymi w germanie i krzemie

Atom domieszkowy	r'	Energia oddziaływania w odl. 4 Å E_I (eV)	
		German	Krzem
Ni	1,24	0,04	0,12
Fe	1,26	0,07	0,15
Cu	1,28	0,12	0,18

Z powodu występowania omówionych wyżej sił segregacji atomów domieszkowych w krysztale zawierającym dyslokacje, rozkład domieszek nie będzie jednorodny. Atomy domieszkowe będą grupowały się wokół dyslokacji tworząc tzw. „atmosferę domieszek Cottrella". Oczywiście powstawanie tej atmosfery będzie zależało od historii termicznej kryształu. Mianowicie przy podwyższaniu temperatury kryształu zawierającego dyslokacje z atmosferą domieszek, tzn. gdy $E_I < kT$, atomy domieszkowe zaczną opuszczać otoczenie dyslokacji i dyfundować do obszarów o mniejszej koncentracji domieszek. Rozkład koncentracji domieszek odpowiadający temperaturze T i energii oddziaływania w danym punkcie będzie określony zależnością:

$$n = n_0 \exp\left(-\frac{E_I}{kT}\right), \tag{18}$$

gdzie n_0 — koncentracja domieszek w materiale w przypadku braku dyslokacji.

Gdy energia termiczna będzie dużo większa od energii oddziaływania z dyslokacją, rozkład domieszek będzie jednorodny. Jeżeli rozgrzany kryształ zostanie szybko ochłodzony atmosfera domieszek nie powstanie, gdyż w obniżonej temperaturze ruchliwość atomów domieszkowych jest zbyt mała. Wobec tego do utworzenia atmosfery domieszek konieczne jest dostatecznie wolne chłodzenie kryształu. Dla przypadku, gdy $E_I \gg kT$, Cottrell i Bilby [14] podali następujący wzór na ilość atomów domieszkowych docierających do jednostki długości dyslokacji:

$$N(t) = 3\, n_0 \left(\frac{\pi}{2}\right)^{\frac{1}{3}} \left(\frac{A \cdot Dt}{kT}\right)^{\frac{2}{3}}, \tag{19}$$

gdzie:

A — stała, zależna od maksymalnej energii oddziaływania atom domieszkowego z dyslokacją — $E_{I\,\max}$,

D — stała dyfuzji atomów domieszkowych w materiale ośrodka,

k — stała Boltzmanna,

T — temperatura bezwzględna kryształu,

t — czas tworzenia się atmosfery.

Wzór (19) jest słuszny tylko dla początkowej fazy tworzenia się atmosfery domieszek, gdy można pominąć wpływ „nasycenia" obszaru dyslokacji atomami domieszkowymi oraz prawa Ficka. Rozwinięcie teorii powstawania atmosfery domieszek w odniesieniu do pojedynczych dyslokacji jak i granic ziarn o małym kącie podał Bilby w 1955 r. [16]. Wydaje się, że obecność atmosfery domieszek w materiałach półprzewodnikowych odgrywa bardzo ważną rolę, szczególnie w procesie rekombinacji nośników mniejszościowych, jednak zagadnienie to nie zostało jeszcze dostatecznie opracowane.

W zakończeniu pragnąłbym podziękować drowi J. Auleytnerowi z Instytutu Fizyki Uniwersytetu Warszawskiego za życzliwą krytykę niniejszej pracy.

Instytut Podstawowych Problemów Techniki PAN
Zakład Elektroniki

WYKAZ LITERATURY

1. Seeger A.: Theorie der Gitterfehlstellen w „Handbuch der Physik". ed. S. Flügge, Bd. VII, Teil I. Springer-Verlag, 1955.
2. Burgers J. M.: Some Considerations of the Field of Stress Connected with Dislocations in a Regular Crystal Lattice. — Proc. Kon. Ned. Akad. v. Wetensch. 42 (1939), s. 293.

3. Kittel Ch.: Introduction to Solid State Physics, II ed., John Wiley & Sons, Inc., New York 1956.
4. Shockley W., Read W. T.: Dislocation Models of Grain Boundaries w „Imperfections in Nearly Perfect Crystals", W. Shockley i in., ed. John Wiley & Sons, New York 1952.
5. Shockley W., Read W. T. jr: Quantitative Predictions from Dislocation Models of Crystal Grain Boundaries. — Phys. Rev. 75 (1949), s. 692.
6. Read W. T. jr: Dislocations in Crystals. Mc Graw Hill Ltd., New York 1953.
7. Frank F. C., Read W. T.: Multiplication Processess for Slow Moving Dislocations — Phys. Rev. 79 (1950), s. 722.
8. van Bueren H. G.: Influence of Lattice Defects on the Electrical Properties of Cold-Worked Metals. — Philips Res. Rep. 12 (1947), s. 1÷45 i 190.
9. Hart E. W.: The Slip Process During Yield-Point Deformation. — Acta Metallurg. 2 (1954), s. 416.
10. Kobus A.: Chemiczne metody ujawniania dyslokacji. — Rozprawy Elektrotechniczne, tom VI (1960), z. 3.
11. Cottrell A. H.: Dislocations and Plastic Flow in Crystals. Clarendon Press, Oxford, 1953.
12. Leibfried G., Lücke K.: Über das Spannungsfeld einer Versetzung. Zeitsch. f. Phys. 126 (1949), s. 450.
13. Dekeyser W., Amelinck S.: Les dislocations et le croissance des cristaux. Masson et Cie Ed., Paris 1955.
14. Cottrell A. H., Bilby B. A.: Dislocation Theory of Yielding and Strain Ageing of Iron. — Proc. Phys. Soc. A62 (1949), s. 49.
15. Kurtz A. D., Kulin S. A., Averbach B. L.: Effect of Dislocation on the Minority Carrier Lifetime in Semiconductors. — Phys. Rev. 101 (1956), s. 1285.
16. Bilby B. A.: The Migration of Solute Atoms to Dislocation Arrays. — J. Phys. Soc. Japan, 10 (1955), Nr 8, s. 673.
17. Turnbull D., Hoffman R. E., The Effect of Relative Crystal and Boundary Orientation on Grain Boundary Diffusion Rates. — Acta Metallurg. 2 (1954), s. 419.
18. Hornstra J.: Dislocation in the Diamond Lattice. — J. Phys. Chem. Solids. 5 (1958), s. 129÷141.

А. КОБУС

ОСНОВЫ ГЕОМЕТРИЧЕСКОЙ ТЕОРИИ ДИСЛОКАЦИИ В КРИСТАЛЛАХ

Резюме

Статья имеет характер введения в вопросы связанные с присутствием некоторого типа несовершенств в кристаллах, называемых дислокациями. Обсуждаемые несовершенства играют очень важную роль при истолковании механических свойств кристаллов. Кроме того в последние годы найдено, что дислокации имеют значительное влияние на электрические свойства кристаллов вообще и, прежде всего, полупроводниковых кристаллов. В настоящей

работе автором обсуждаются основные явления происходящие в кристаллах при наличии дислокации. В начале представлена классификация несовершенств, а также введены основные понятия служащие для количественного описания дислокации: контур и вектор Бургерса. Далее были описаны геометрические модели разных типов дислокации: то есть краевой дислокации, винтовой дислокации и криволинейной дислокации (сложной дислокации), а также их основные системы: границу зёрн с малым углом и границу зёрн с большим углом. Существование дислокации в кристалле вызывает наличие поля напряжений, которое может воздействовать на другие поля напряжений в кристалле. Они могут быть вызваны другими дислокациями или атомами других элементов. Взаимодействие дислокаций может в соответственных условиях вызвать их движение. Взаимодействие дислокаций и атомов примесей ведёт к появлению сил распределения вызывающих движение этих атомов в направлении дислокации. В значительном результате этого явления возникает вокруг дислокации так называемая „атмосфера примесей Котрелла", которая имеет значительное влияние на механические свойства, а также по всей вероятности, на электрические свойства кристаллов.

A. KOBUS

FOUNDATIONS OF THE GEOMETRICAL THEORY OF DISLOCATIONS IN CRYSTALS

Summary

Fundamental problems connected with the presence in crystals of certain kinds of imperfections, called dislocations are introduced in the paper. In the introduction the author discusses basic concepts to be used in the quantitative description of dislocations, i.e. Burgers' circuit and vector, and geometric models illustrating principal kinds of dislocations, namely, edge-dislocation, screw-dislocation and compound dislocation. Next, geometric models for the arising and multiplication of dislocations are discussed, and phenomena occurring as a result of the influence of stress fields introduced in the lattice by dislocations and impurities are described.

A. KOBUS

LES PRINCIPES DE LA THÉORIE GÉOMETRIQUE DES DISLOCATIONS DANS LES CRISTAUX

Résumé

L'article présente les principaux problèmes liés avec la présence de certaines imperfections dans les cristaux, nommées dislocations. Au commencement on a discuté les principales idées servant à la description quantitative des dislocations, c.à.d. circuit et vector Burgers, et les modèles géometriques représentants les principales espèces de dislocations: dislocation coin, dislocation vis et dislocation courbe (complexe). Ensuite on a discuté les modèles géometriques de la formation et multiplication des dislocations et les phénomènes résultants de l'influence des champs de tension introduits dans le réseau par les dislocations et les atomes étrangers.

A. KOBUS

DIE GRUNDLAGEN DER GEOMETRISCHEN VERSETZUNGTHEORIE IN KRISTALLEN

Zusammenfassung

Der Artikel führt in die mit Anwesenheit einer Gitterfehlstelle in Kristallen gebundene Grundprobleme, sogenannte Versetzungen ein. Am Anfang wurden die Grundbegriffe zur quantitativen Beschreibung der Versetzungen besprochen, d.h. der Burgers Umlauf und Vektor, sowie die geometrischen Modelle, die die Grundarten der Versetzungen abbilden: die Stufenversetzung, die Schraubenversetzung und die allgemeine Versetzunglinie. Demnächst wurden die geometrischen Modelle des Entstehens und der Multiplikation der Versetzungen besprochen, sowie die Phänomene, die als Resultat der Einwirkung der in das Gitter durch Versetzung und Verunreinigungsatome eingeführten Spannungsfelder entstehen.

ANDRZEJ KOBUS

Chemiczne metody ujawniania dyslokacji w monokryształach germanu

Rękopis dostarczono 24.4.1959

Autor podaje przegląd chemicznych metod ujawniania dyslokacji w kryształach germanu. Po odpowiedniej obróbce chemicznej na powierzchni kryształu powstają wgłębienia, które odpowiadają przecięciu linii dyslokacyjnej z powierzchnią. Metody chemiczne pomimo wad (nie wiadomo, czy wszystkie dyslokacje są ujawnione) odznaczają się wielką prostotą i wymagają niewielkiego wyposażenia. Podano szereg odczynników stosowanych do ujawnienia dyslokacji oraz przedyskutowano ich przydatność.

1. WSTĘP

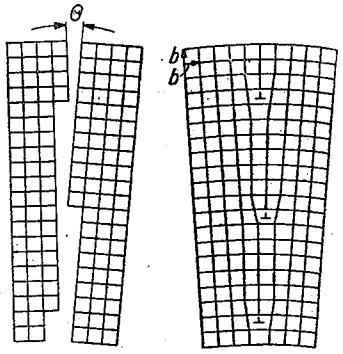
Pomimo że w literaturze światowej opublikowano wiele prac poświęconych chemicznemu ujawnianiu dyslokacji w półprzewodnikach, to jednak dane te są często niejasne lub czasem sprzeczne. Dlatego celem tej pracy było przebadanie szeregu znanych odczynników chemicznych stosowanych przez różnych autorów do ujawniania dyslokacji i usystematyzowanie wyników otrzymanych przez trawienie różnymi metodami. Badanie prcesów ujawniania dyslokacji przeprowadzano na powierzchni (111) monokryształów germanu w większości wykonanych w Zakładzie Elektroniki IPPT PAN [1].

2. ODPOWIENIOŚĆ WGŁĘBIEŃ POSTAŁYCH W WYNIKU TRAWIENIA Z DYSLOKACJAMI

Shockley i Read [2] badając w 1949 roku powierzchnię germanu wyciętego w określonej orientacji krystalograficznej i wytrawioną odpowiednimi odczynnikami chemicznymi stwierdzili, że na powierzchni kryształu istnieją pewne miejsca, które podlegają szybszemu trawieniu niż pozostała część płaszczyzny. Badacze ci sugerowali, że miejsca łatwego trawienia na powierzchni germanu są punktami przecięcia linii dyslokacyjnej z powierzchnią kryształu. Szybsze trawienie punktu przecięcia linii dyslokacyjnej z powierzchnią wynika z faktu, że w obszarze tym

wskutek zakłócenia periodyczności budowy sieci krystalicznej zachodzi również zaburzenie równowagi energetycznej w obszarze dyslokacji. Nadmiar energii w obszarze dyslokacji powoduje przyspieszone działanie odczynnika trawiącego. Zagadnienia związane z energią linii dyslokacyjnej były opracowane przez Cottrella [3] i Reada [4] (patrz również [5]).

Hipotezę jednoznaczności położenia wgłębień powstałych w wyniku trawienia powierzchni germanu z punktami przecięcia linii dyslokacyj-



Rys. 1. Schemat granicy ziarn o małym kącie

nych z powierzchnią udowodniła grupa badaczy amerykańskich w roku 1953 [6]. Rozpatrywali oni kryształ o lekko skręconych względem siebie dwóch obszarach. Trawiąc jego powierzchnię otrzymali oni rząd wgłębień w stałej odległości między sobą, analogiczny do rys. 11. Przyjmując model granicy ziarn o małym kącie, przedstawiony na rys. 1, oraz równanie obowiązujące dla tego układu

$$\Theta = \frac{b}{D}, \tag{1}$$

gdzie:

Θ – kąt skręcenia w radianach,

b – moduł wektora Burgersa,

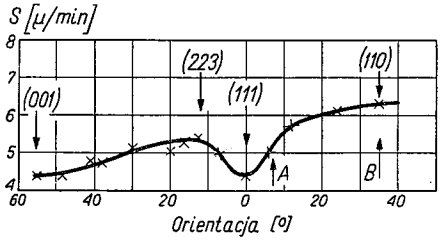
D – odległość między dyslokacjami w granicy ziarn,

wyznaczyli oni kąt skręcenia obu obszarów. Następnie skręcenie tego obszaru określili przy pomocy analizy rentgenowskiej. Różnica wyników dla kątów skręcenia większych od 20 min była mniejsza od 2%, dla kątów ok. 10 min — 10%. Wyniki te należy uznać za potwierdzenie hipotezy Shockleya-Reada.

3. ODDZIAŁYWANIE ODCZYNNIKÓW CHEMICZNYCH Z POWIERZCHNIĄ GERMANU

Mechanizm oddziaływania odpowiedniego odczynnika chemicznego z obszarem dyslokacji na powierzchni półprzewodnika nie jest dosta-

tecznie dokładnie wyjaśniony. Pierwszą publikacją poświęconą wynikom badania mechanizmu trawienia obszaru dyslokacji jest praca B. Battermana z 1957 r. [7]. Autor ten badał szybkość trawienia germanu pewnym odczynnikiem (odczynnik nr 4 wg tablicy 1). Próbki posiadały różne orientacje badanej powierzchni. Z otrzymanych wyników został spo-



Rys. 2. Zależność szybkości trawienia mieszanką nr 5 ($HF—H_2O_2—H_2O$) od orientacji kryształu

rządzony wykres jak na rys. 2. Z doświadczenia wiadomo, że dyslokacje w postaci wgłębień zostają ujawnione przy użyciu wspomnianej mieszanki na powierzchniach (100) i (111), które są jak widać z rys. 2 powierzchniami o minimalnej szybkości trawienia. Natomiast stwierdzono, że na powierzchniach o maksymalnej szybkości trawienia pozostają pagórki ograniczone płaszczyznami krystalograficznymi o mniejszej szybkości trawienia, które nie reprezentują wewnętrznej doskonałości struktury krystalicznej.

Mechanizm tworzenia się wgłębień w punkcie przecięcia się linii dyslokacyjnej z powierzchnią kryształu można w przybliżeniu przedstawić w następujący sposób: wgłębienie powstaje w przypadku, gdy szybkość trawienia obszaru dyslokacji jest większa niż pozostałej powierzchni kryształu. Jeżeli założymy, że szybkość trawienia obszaru dyslokacji zależy przede wszystkim od właściwości samej dyslokacji, np. od pola naprężeń wywołanych jej obecnością, to dla dyslokacji o określonych własnościach szybkość trawienia obszaru zdeformowania będzie stała we wszystkich orientacjach. Dyslokacje będą ujawniane tylko na powierzchniach o dostatecznie małej szybkości trawienia, gdy szybkość trawienia obszaru dyslokacji jest większa od szybkości trawienia badanej powierzchni. Konfrontując wykres z rys. 2 z powyższymi rozważaniami można przyjąć, że dla omawianej mieszanki szybkość trawienia obszaru dyslokacji wynosi około 4,5 μ/min. Omówiona praca rzuca już pewne światło na mechanizm chemicznego ujawniania dyslokacji.

Dane o wynikach ujawniania dyslokacji są pełniejsze, jednak często są niejasne lub czasem sprzeczne. Różni autorzy podają całkiem odmienne dane o gęstości dyslokacji w germanie w zależności od spo-

sobu przygotowania powierzchni próbki. Kurtz, Kulin, Averbach [8], którzy badali monokryształy germanu przy pomocy promieni rentgenowskich oraz trawienia mieszanką CP4 (składy mieszanek będą podane dalej) podawali wartość gęstości dyslokacji na $10^5 \div 10^7$ cm^{-2}, gdy prace Billiga [9], [10] podają gęstości dla germanu $10^3 \div 10^5$ cm^{-2}. Natomiast Ellis [11] badając monokryształy germanu przez trawienie odpowiednich powierzchni różnymi odczynnikami chemicznymi stwierdził istnienie trzech rodzajów dyslokacji. Mianowicie jednego rodzaju, którego gęstość wynosi $10^3 \div 10^4$ cm^{-2}, oraz dwóch innych, które nie zawsze można zdecydowanie zakwalifikować do jednego z dwóch pozostałych typów, a których gęstość wynosi $10^6 \div 10^7$ cm^{-2}. Również Auleytner, Godwood i Kryłow [16] stwierdzili istnienie dwóch rodzajów wgłębień („dużych" i „małych") na powierzchniach kryształów germanu trawionych mieszanką CP4; gęstości ich wynosiły odpowiednio — pierwszych — do 10^5 cm^{-2} i drugich — od 10^6 do 10^8 cm^{-2}. Natomiast w artykule Buerena [12] jest wzmianka o otrzymaniu monokryształu germanu o gęstości dyslokacji $10 \div 10^2$ cm^{-2} [1]).

Jak widać, w literaturze światowej obserwuje się stałe zmniejszanie przytaczanych gęstości dyslokacji, co niewątpliwie wynika z postępu technologii otrzymywania monokryształów, jednak w wyżej wymienionych publikacjach nie wszyscy autorzy podkreślają, czy omawiany sposób ujawniania dyslokacji wykrywa ich część, czy wszystkie.

Praca niniejsza miała na celu porównanie różnych metod chemicznego ujawniania dyslokacji i sprawdzenia przydatności szeregu odczynników. Badania przeprowadzono na szeregu próbek monokryształu germanu na powierzchni (111). Próbki były przygotowywane jak następuje. Najpierw poddawano je szlifowaniu mechanicznemu proszkiem karborundowym # 600 do # 1200 na płycie szklanej (odpowiada to średnicy ziarna odpowiednio od ~14μ do 7μ), a następnie polerowano proszkiem alundowym na filcu i dalej trawiono w szybko trawiącej mieszance w celu otrzymania gładkiej powierzchni bez defektów spowodowanych obróbką mechaniczną, tzn. rys, zadrapań itp. W tablicy 1 podano składy odczynników stosowanych w niniejszej pracy. Jako czynnik polerujący stosowano mieszankę nr 1 lub mieszankę nr 4 (CP4). Mieszanek tych można również używać do ujawniania dyslokacji, jednak należy spełnić pewne warunki, o których będzie mowa w dalszym ciągu. Próbkę wypolerowaną chemicznie poddawano działaniu odpowiedniej mieszanki ujawniającej.

[1]) Wzmianka ta pokrywa się z prywatną informacją otrzymaną przez prof. W. Rosińskiego w Wielkiej Brytanii od przedstawiciela firmy „Marconi" o wykonaniu monokryształu germanu o gęstości dyslokacji 10 cm^{-2}.

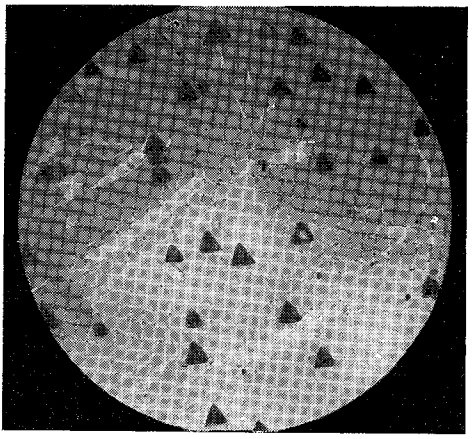
Tablica 1

Mieszanki trawiące stosowane do ujawniania dyslokacji

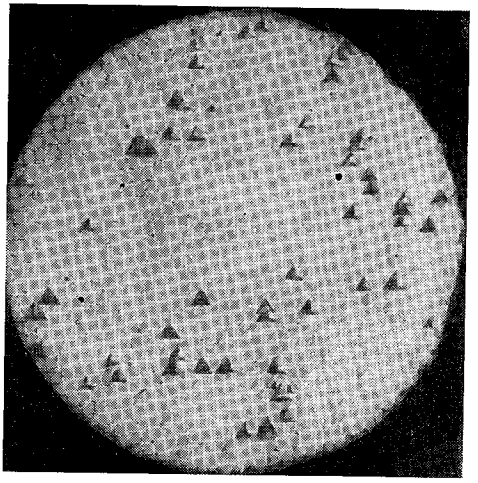
Nr 1		Nr 2 [9]	
HF	50 ccm	$K_3Fe(CN)_6$	8 g
HNO_3	50 ccm	KOH	12 g
		H_2O dest.	100 ml
Czas trawienia 0,5 — 4 min		Próbki gotuje się 3 — 7 min	
Nr 3 (tzw. Westinghouse Silver Etch)		Nr 4 (CP4)	
HF	40 ccm	HNO_3	50 ccm
HNO_3	20 ccm	CH_3COOH	30 ccm
H_2O	40 ccm	HF	30 ccm
$AgNO_3$	2 g	Br (ciekły)	0,6 ccm
Czas trawienia 5 — 8 min		Czas trawienia 5 min	
Nr 5		Nr 6 [11]	
HF	10 ccm	Jest to roztwór wodny mieszanki nr 5	
H_2O_2	10 ccm		
H_2O	40 ccm	Nr 4	1 ccm
		H_2O	56 ccm
Czas trawienia 30 min		Czas trawienia 2 — 16 godz	

Stwierdzono, że mieszanki trawiące podane w tablicy 1 ujawniają dwa zasadnicze rodzaje dyslokacji. Pierwszy z nich tworzy duże wgłębienia o wymiarach liniowych rzędu 20 — 50μ (dla podanych czasów trawienia), o gęstości $10^2 \div 10^4$ cm^{-2} w monokryształach germanu stosowanego do celów produkcyjnych. Obrazy ujawnionych „dużych" wgłębień przy pomocy różnych odczynników podano na rys. 3, 4, 5, 6 i 7. Jak widać ze zdjęć, mieszanki nr 2 i 3 mają jednakowy sposób atakowania powierzchni wytrawiając trójściany, które przy mniejszych powiększeniach wyglądają jak ciemne trójkąty. Natomiast mieszanki nr 4, 5 i 6 ujawniają wgłębienia o kształcie wklęsłych stożków kołowych. W obrazie otrzymanym przez trawienie mieszanką nr 6, jak widać na rys. 7, środek stożka posiada zarys symetrii trójkątnej podobnie jak na rys. 3 i 4, gdy dalsza część wgłębienia ma symetrię kołową.

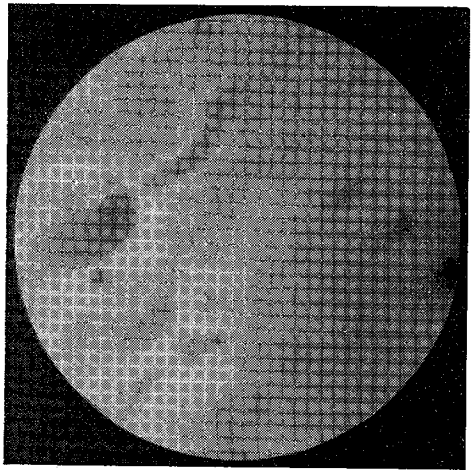
Drugi rodzaj dyslokacji tworzy małe wgłębienia o wymiarach liniowych rzędu 1 — 3μ, o kształcie wklęsłej półkuli. Obrazy tych dyslokacji ujawnione mieszankami nr 4 (CP4) i nr 6 pokazano na rys. 8 i 9. Są one widoczne dopiero przy powiększeniach 300 — 500-krotnych. Ellis [11] podaje, że wśród tych wgłębień można stwierdzić jeszcze drugą odmia-



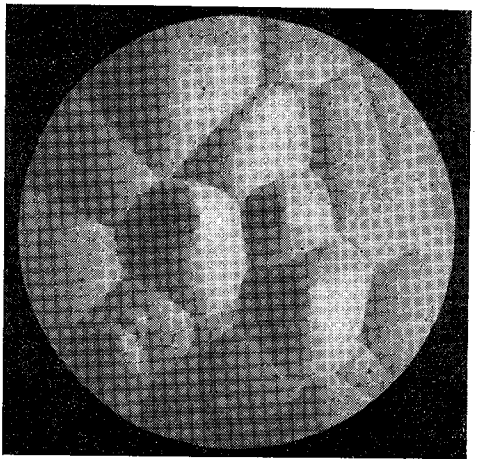
Rys. 3. Obraz „dużych" wgłębień otrzymany w wyniku trawienia mieszanką nr 2 (× 60)



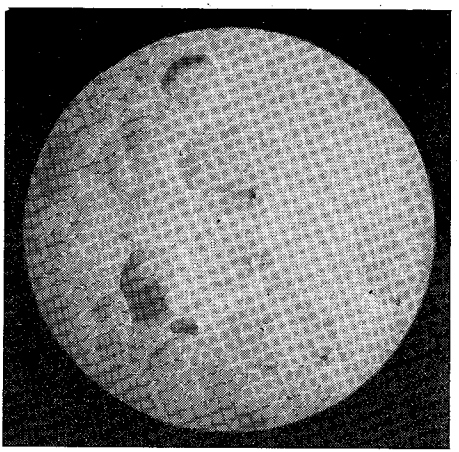
Rys. 4. Obraz „dużych" wgłębień otrzymany w wyniku trawienia mieszanką nr 3 (× 60)



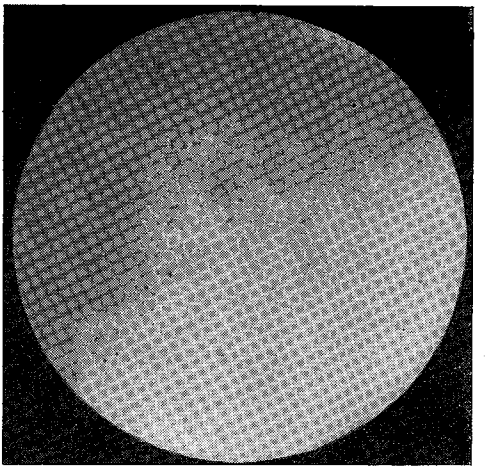
Rys. 5. Obraz „dużych" wgłębień otrzymany w wyniku trawienia mieszanką nr 4 (× 60)



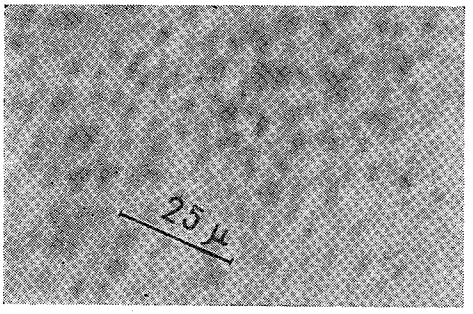
Rys. 6. Obraz „dużych" wgłębień otrzymany w wyniku trawienia mieszanką nr 5 (× 150)



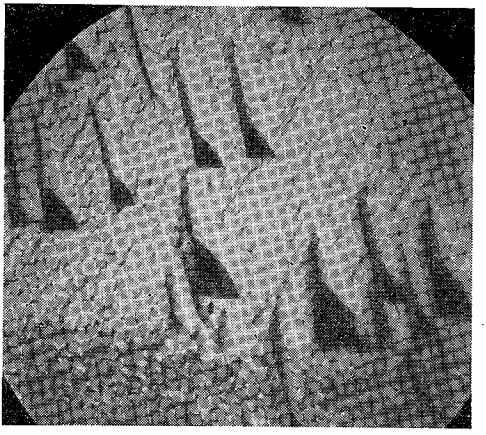
Rys. 7. Obraz „dużych" wgłębień otrzymany w wyniku trawienia mieszanką nr 6 (× 150)



Rys. 8. Obraz „małych" wgłębień uzyskany przez trawienie mieszanką nr 4 (× 300)



Rys. 9. Obraz „małych" wgłębień uzyskany przez trawienie mieszanką nr 6 (× 600)



Rys. 10. Obraz wgłębień na płaszczyźnie różniącej się o ok. 10° od powierzchni (111). Trawienie mieszanką nr 2 (× 300)

nę — posiadającą postać spirali trójkątnej. Jak widać z rys. 8 i 9, pomimo 300 i 600-krotnego powiększenia nie można próbować wyróżnić wgłębień spiralnych. Gęstość dyslokacji drugiego rodzaju zawiera się w granicach 10^5 — 10^7 cm^{-2} również dla monokryształów stosowanych do celów produkcyjnych.

Jak już podano poprzednio, wgłębienia dyslokacyjne są możliwe do wytrawienia tylko na tych powierzchniach, których szybkość trawienia

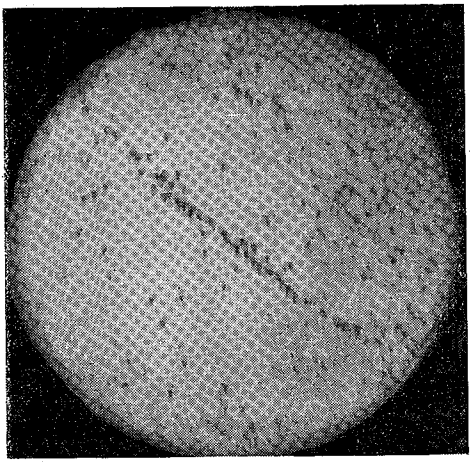
jest mniejsza od szybkości trawienia obszaru dyslokacji. Jak wykazuje doświadczenie, powierzchniami takimi są podstawowe płaszczyzny krystalograficzne, jak (100), (110), (111), na których zazwyczaj przeprowadza się ujawnianie dyslokacji. Gdy badana powierzchnia różni się o pewien kąt od powierzchni o minimalnej szybkości trawienia, to oprócz zmiany szybkości trawienia występuje zmiana kształtu wytrawionego wgłębienia. Na rys. 10 pokazano przykładowo obraz dużych wgłębień wytrawionych na powierzchni różniącej się od płaszczyzny (111) o około 10°. Na rys. 5 obraz dyslokacji również nie jest wytrawiony w dokładnej płaszczyźnie (111), lecz posiada niewielkie odchylenie, rzędu 1 — 2°. Ogólnie biorąc dla dużych wgłębień najłagodniejszy przebieg zależności szybkości trawienia od orientacji kryształu posiada mieszanka nr 2, gdyż wgłębienia znikają przy odchyleniu o ok. 15° od płaszczyzny (111). Mieszanka nr 3 ma podobne właściwości. Natomiast, jak podawano poprzednio, mieszanka nr 5 ujawnia dyslokacje dla orientacji powierzchni (111) $\pm$ 7°. Podobne właściwości posiadają mieszanki nr 4 i 6.

Dyslokacje tworzące małe wgłębienia są bardziej czułe na odchylenie od powierzchni o minimalnej prędkości trawienia i — jak podaje Ellis [13] — znikają przy różnicy orientacji badanej powierzchni o 3° od płaszczyzny (111). Wyżej omówione fakty doświadczalne pozwalają na odtworzenie mechanizmu ujawniania dyslokacji obu rodzajów przy pomocy tego samego odczynnika. Ogólnie biorąc można przyjąć, że obszar małego wgłębienia trawi się wolniej w porównaniu z wgłębieniami dużymi. Wobec tego, aby uprzywilejować ujawnianie małych wgłębień, należy stosować albo zimniejszą kąpiel mieszanki, jak czyniono dla CP4, lub słabszy roztwór trawiący, którym jest mieszanka nr 6 dla odczynnika nr 5. Powoduje to dostatecznie małą prędkość trawienia badanej powierzchni w porównaniu z prędkością trawienia obszaru dyslokacji i możliwość ujawnienia małych wgłębień.

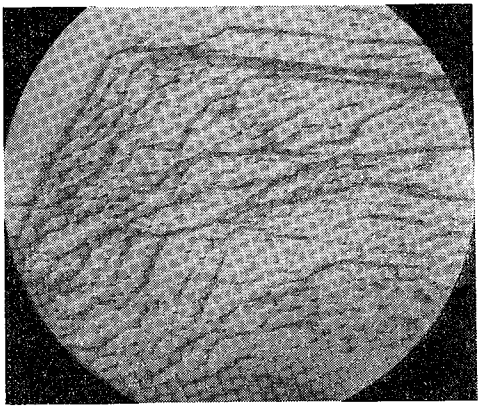
4. UJAWNIANIE UKŁADÓW DYSLOKACJI

W odpowiednich warunkach dyslokacje mogą tworzyć pewne układy nieprawidłowości. Układy dyslokacji oraz mechanizm ich powstawania były omówione w poprzednim artykule [5]. Obecnie zostaną pokazane przykłady ujawnionych układów dyslokacji będące doświadczalnym potwierdzeniem omawianych modeli. Rysune 11 przedstawia obraz granicy ziarn o małym kącie wytrawiony mieszanką nr 2. Wgłębienia leżą w odległościach 10 — 20μ od siebie, co odpowiada kątowi skręcenia 7 — 15′. Rysunek 12 przedstawia silnie zniekształcony obszar w pobliżu przecięcia dwóch granic ziarn o małym kącie. Powierzchnię trawiono

mieszanką nr 2. Gęstość dyslokacji odpowiadających „dużym" wgłębieniom jest bardzo duża i wynosi około 10^5 cm^{-2}. Obecność takich ukła-

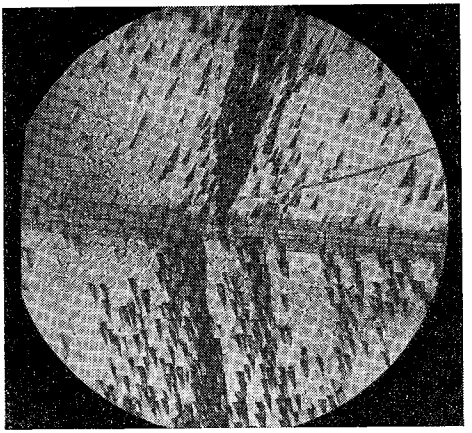


Rys. 11. Obraz granicy ziarn o małym kącie ujawniony mieszanką nr 2 (× 70)

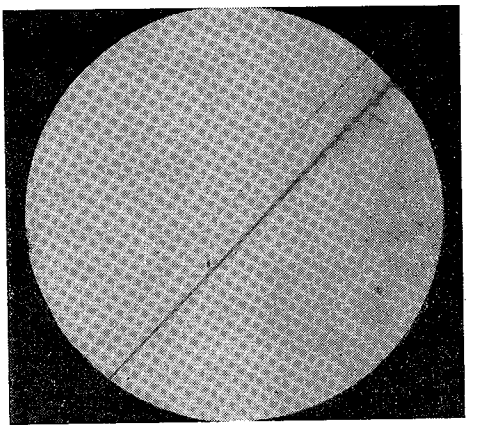


Rys. 12. Obraz silnie zniekształconego obszaru wokół przecięcia dwóch ziarn o małym kącie (× 140)

dów dyslokacji dyskwalifikuje german jako materiał do zastosowań praktycznych. Przykład granicy ziarn (bliźniaków) w monokrysztale germanu



Rys. 13. Obraz granicy ziarn (bliźniaków) ujawniony mieszanką nr 2 (× 60)

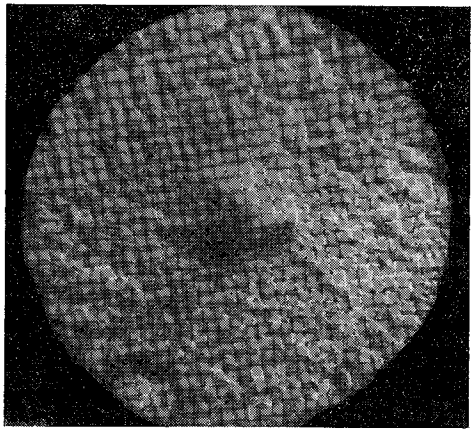


Rys. 14. Obraz granicy ziarn z rys. 13 ujawniony mieszanką nr 1 bez ujawniania dyslokacji (× 60)

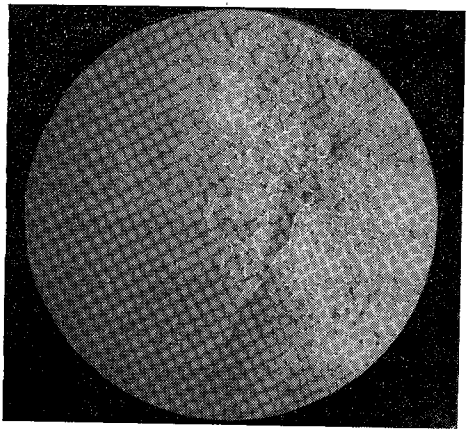
pokazano na rys. 13. Jest to obraz uzyskany również przez trawienie żelazicjankiem potasu. Ze zdjęcia widać, że obszary po obu stronach granicy leżą w przybliżeniu w płaszczyźnie (111), to znaczy oś skręcenia

jest prostopadła do powierzchni obrazu. Widoczne jest również silne zniekształcenie budowy krystalicznej prostopadłe do granicy ziarn. Obraz tej samej granicy wytrawiony tylko mieszanką nr 1 bez ujawniania dyslokacji przedstawiony jest na rysunku 14.

Następna seria zdjęć przedstawia obraz powierzchni monokryształu germanu, którego struktura w pewnym miejscu została silnie zakłóco-



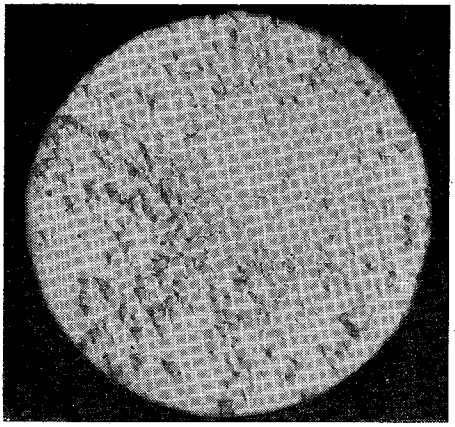
Rys. 15. Obraz silnego zaburzenia sieci ujawniony mieszanką nr 2 (× 150)



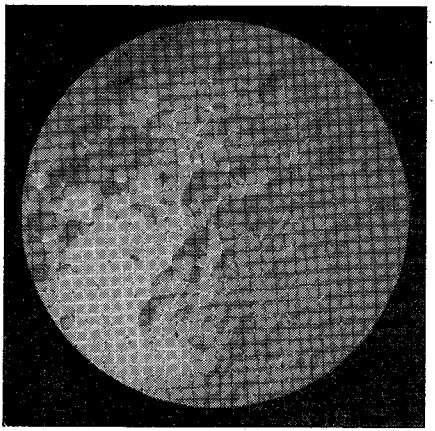
Rys. 16. Obraz silnego zaburzenia sieci ujawniony mieszanką nr 1 (× 150)

na. Ponieważ takie skupiska dyslokacji powstają najczęściej na końcach monokryształu, przypuszcza się, że są one spowodowane silnymi naprężeniami wytworzonymi zmianą średnicy monokryształu i zmienionymi warunkami chłodzenia. Zdjęcia przedstawiają w przybliżeniu ten sam obszar oglądany po trawieniu różnymi mieszankami. Należy zaznaczyć, że zdjęcia nie odpowiadają dokładnie tym samym przekrojom powierzchni, gdyż między poszczególnymi powierzchniami przeprowadzono kolejne polerowanie mechaniczne i chemiczne tak, że różnice między poziomem poszczególnych powierzchni wynoszą około 50 — 100μ. Zdjęcia były wykonywane w następującej kolejności: rys. 15 — trawiony mieszanką nr 2, rys. 16 — trawiony mieszanką nr 1, rys. 17 — trawiony mieszanką nr 3 i rys. 18 — trawiony mieszanką nr 5.

Na zakończenie tego rozdziału zostanie omówiona specjalna technika ujawniania dyslokacji leżących w pobliżu lub równolegle do powierzchni, podana przez Tylera i Dasha [14]. Polega ona na dyfuzji litu do germanu, a następnie trawieniu mieszanką złożoną z HNO_3—HF—CH_3COOH. W wyniku dyfuzji w odpowiednich warunkach termicznych atomy litu przenikają do wnętrza kryształu i tworzą atmosferę domieszek wokół dyslokacji. Następnie mieszanka trawiąca oddziałując

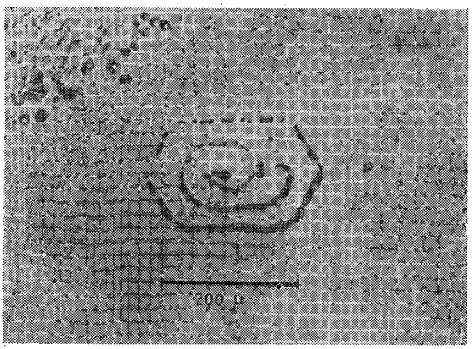


Rys. 17. Obraz silnego zaburzenia sieci ujawniony mieszanką nr 3 (× 150)

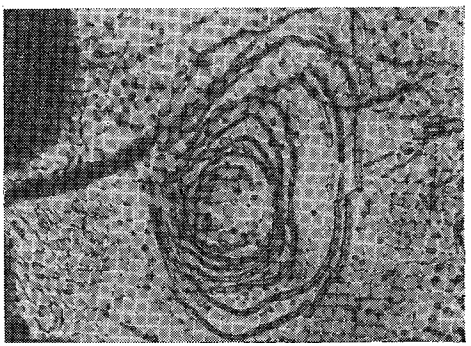


Rys. 18. Obraz silnego zaburzenia sieci ujawniony mieszanką nr 5 (× 150)

przede wszystkim z atomami litu powoduje wytrawianie obszarów o maksymalnej jego zawartości. Autorzy ci badali próbki germanu poddanego deformacji plastycznej w temperaturach 500 — 650°C, na których



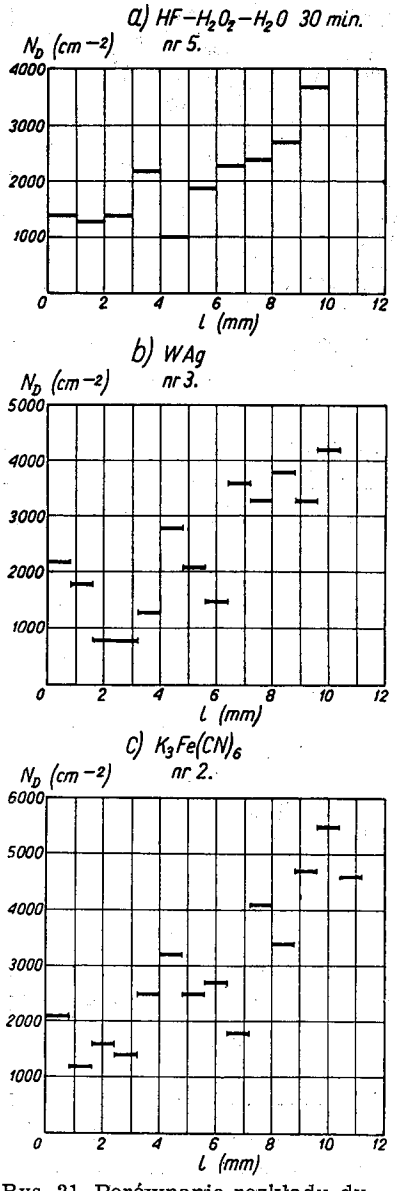
Rys. 19. Obraz źródła Franka-Reada powstałego w wyniku deformacji kryształu germanu w tempearturze 550°C [14]



Rys. 20. Obraz źródła Franka-Reada powstałego w wyniku deformacji kryształu germanu w temperaturze 650°C [14]

następnie przeprowadzano dyfuzję litu i trawienie chemiczne. Metoda ta dała szczególnie dobre obrazy źródeł Franka-Reada w germanie. Przykładem tego jest rys. 19 przedstawiający omawianą nieprawidłowość, która powstała w wyniku deformacji plastycznej kryształu w temperaturze 550°C. Gdy deformacja odbywa się w temperaturach wyższych, na przykład w 650°C, przebieg pętel dyslokacyjnych nie jest tak regularny jak w przypadku poprzednim, co widać na rys. 20. Jak wynika

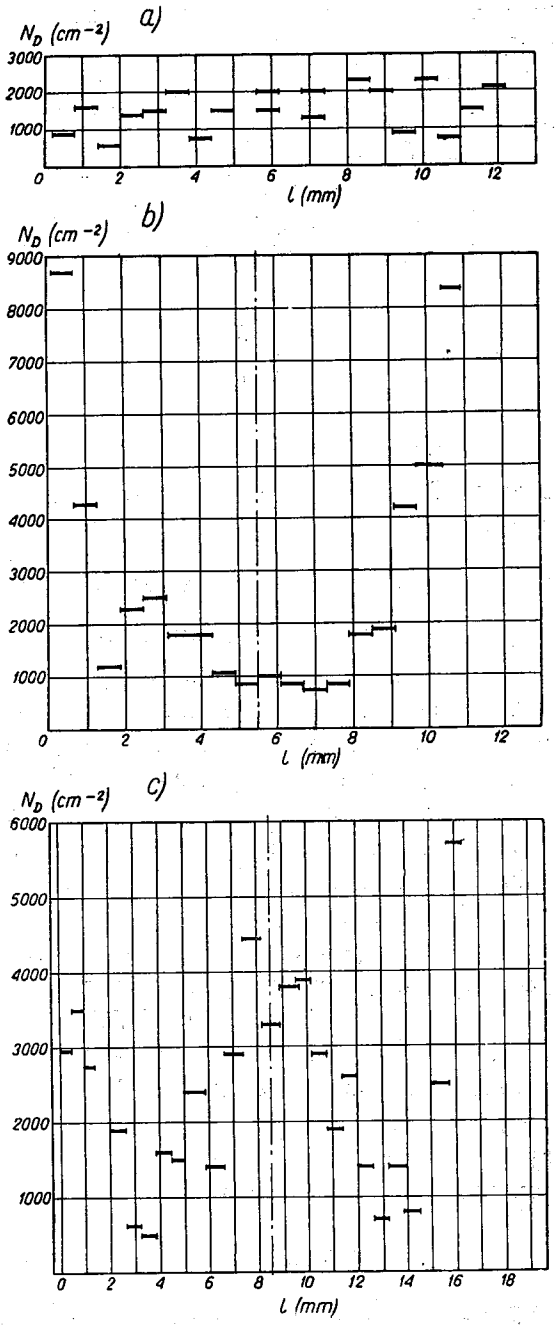
z powyższych zdjęć, ruch dyslokacji w temperaturach poniżej 600°C jest ograniczony pewnymi kierunkami krystalograficznymi. Orientacje poszczególnych odcinków źródła Franka-Reada są kierunkami (110). W próbce odkształconej w temperaturze wyższej od 600°C ruch dyslokacji nie jest ograniczony przez strukturę krystaliczną i pętle dyslokacyjne przyjmują postać zbliżoną do elips. Obrazy źródła otrzymane powyższą metodą pokrywają się z analogicznymi obrazami otrzymanymi dla krzemu przy pomocy mikroskopii podczerwonej i potwierdzają hipotezę Franka-Reada, zarówno dla germanu, jak i dla krzemu.



Rys. 21. Porównanie rozkładu dyslokacji ujawnianych kolejno trzema odczynnikami: a) mieszanką nr 5, b) mieszanką nr 3, c) mieszanką nr 2

5. OMÓWIENIE WYNIKÓW

W celu porównania oddziaływania poszczególnych trawień określano gęstość dyslokacji pierwszego rodzaju wzdłuż określonej średnicy pewnej próbki germanu trawiąc ją różnymi mieszankami. Wyniki przedstawiono na rys. 21. Porównanie to może nie jest dostatecznie precyzyjne ze względu na to, że pomiędzy poszczególnymi powierzchniami trawienia istnieje odległość 50 — 100μ spowodowana koniecznością obróbki mechanicznej i chemicznej w celu ujawnienia dyslokacji następną mieszanką. Jednak, jak widać na rys. 21, różnice gęstości dyslokacji w poszczególnych punktach nie przekraczają na ogół 20%, co ze względów wyżej podanych jak i z powodu możliwych błędów umiejscowienia badanego obszaru należy uważać za zupełnie dobrą jednoznaczność określania gęstości dyslokacji kawędziowych różnymi mieszankami.



Rys. 22. Trzy przykłady przebiegu gęstości dyslokacji krawędziowych wzdłuż średnicy monokryształu germanu

Monokryształ wyciągany z fazy ciekłej posiada na ogół niejednorodne warunki chłodzenia wskutek nierównomiernego rozkładu temperatury zarówno w kierunku osiowym, jak i promieniowym. W wyniku powyższych nierównomierności temperatury w krysztale powstają naprężenia. Zewnętrzna część kryształu, jako chłodzona szybciej niż środek, będzie ściskała wnętrze walca. W wyniku działania powstałych sił ścinających pojawi się poślizg kryształu i zatem dyslokacji. Jak podaje Billig [9], w pierwszym przybliżeniu gęstość dyslokacji wyznaczonej przez natężenia powstałe w wyniku niejednorodności temperatury wyraża się następująco:

$$N_D = \frac{a}{b} \frac{\delta T}{\delta r}, \tag{2}$$

gdzie:

N_D — gęstość dyslokacji,
a — współczynnik rozszerzalności termicznej,
b — moduł wektora Burgersa,
$\frac{\delta T}{\delta r}$ — temperaturowy gradient promieniowy.

Wobec tego zbadanie rozkładu dyslokacji wzdłuż średnicy monokryształu może dać pojęcie o rozkładzie temperatury w zakresie plastyczności germanu [15]. Typowe rozkłady dyslokacji pierwszego typu w krysztale germanu przedstawione są na rys. 22 abc. Długość kreski na wykresie odpowiada średnicy pola obserwowanego przez mikroskop. Nierównomierności przebiegu krzywych są spowodowane prawdopodobnie głównie niesymetrycznością wyciąganego monokryształu, gdyż otrzymano go w urządzeniu nie pozwalającym na obracanie.

Wyniki dotyczące dyslokacji drugiego typu są trudne do zinterpretowania ze względu na małą powtarzalność. Spowodowane jest to czułością tych wgłębień na odchylenie od płaszczyzny o minimalnej prędkości trawienia oraz trudnością utrzymania właściwych warunków trawienia. Poza tym nawet drobne nierówności powierzchni kryształu spowodowane polerowaniem chemicznym uniemożliwiają wykonanie zdjęcia wskutek małej głębi ostrości przy dużych powiększeniach. Dotychczas żadni autorzy nie stwierdzili charakterystycznych rozkładów dyslokacji drugiego typu w monokryształach germanu.

6. WNIOSKI

Uważa się, że wgłębienia pierwszego rodzaju odpowiadają dyslokacjom typu krawędziowego. Dyslokacje te były opisane w doświadczeniu Vogela i in. [6], którzy potwierdzili hipotezę, że granica ziarn o małym

kącie składa się z dyslokacji krawędziowych, tj. dużych wgłębień. Natomiast przyjmuje się, że małym wgłębieniom odpowiadają dyslokacje złożone, o orientacjach różnych od krawędziowej. Za hipotezą tą przemawia fakt, że gęstość ich odpowiada gęstości wyznaczonej z pomiarów rentgenowskich [8], które najprawdopodobniej wyznaczają bezwzględną zawartość dyslokacji w germanie.

Z doświadczeń uzyskanych w czasie przygotowywania próbek wynika, że do ujawniania dyslokacji krawędziowych najdogodniejsza jest mieszanka nr 2. Działa ona najpewniej i na ogół daje jednoznaczne obrazy trawień nawet w przypadku niezbyt dokładnie wypolerowanych powierzchni, w przeciwieństwie do pozostałych mieszanek. Pewność działania tej mieszanki wynika z samoczynnego termostatowania procesu trawienia, gdy przy pozostałych mieszankach istnieje znaczny wpływ temperatury otoczenia oraz zmiany temperatury w czasie samego ujawniania.

Badanie dyslokacji złożonych metodami chemicznymi przedstawia wiele trudności spowodowanych przyczynami omówionymi powyżej. Najlepsze wyniki otrzymano przez trawienie mieszanką nr 6. Jednak wydaje się, że do ilościowych badań dyslokacji złożonych okażą się konieczne inne metody wykrywania dyslokacji, np. analiza rentgenowska.

W zakończeniu pragnąłbym podziękować drowi J. Auleytnerowi z Instytutu Fizyki Uniwersytetu Warszawskiego za szereg cennych uwag dotyczących niniejszej pracy.

Instytut Podstawowych Problemów Techniki PAN
Zakład Elektroniki

WYKAZ LITERATURY

1. Brochocki A.: Technologia kryształów germanu. — Elektronika 3 (1957), Nr 10—11, s. 20.
2. Shockley W., Read W. T. jr: Quantitative Predictions from Dislocation Models of Crystal Grain Boundaries. — Phys. Rev. 75 (1949), s. 692.
3. Cottrell A. H.: Dislocation and Plastic Flow in Crystals. Oxford University Press, 1953.
4. Read W. T. jr: Dislocations in Crystals. Mc. Graw-Hill Co., New York 1953.
5. Kobus A.: Podstawy geometrycznej teorii dyslokacji w kryształach. — Rozprawy Elektrotechniczne, t. VI (1960) z. 3.
6. Vogel F. L., Pfann W. G., Corey H. E., Thomas E. E.: Observations of Dislocation in Lineage Boundaries in Germanium. — Phys. Rev. 90 (1953), s. 489.
7. Batterman B. W.: Hillocks, Pits and Etch Rate in Germanium Crystals. — Journ. Appl. Phys. 28 (1957), s. 1236.

8. Kurtz A. D., Kulin S. A., Averbach B. L.: Effect of Dislocations on the Minority Carrier Lifetime in Semiconductors. — Phys. Rev. 101 (1956), s. 1285.
9. Billig E.: Some Defects in Crystals Grown from the Melt. I. Defects Caused by Thermal Stressess. — Proc. Roy. Soc. 235A (1956), Nr 1200, s. 37.
10. Billig E.: Defects in Germanium Crystals Grown from the Melt. — Brit. J. Appl. Phys. 7 (1956), s. 375.
11. Ellis S. G.: Dislocations in Germanium. — Journ. Appl. Phys. 26 (1955), s. 1140.
12. van Bueren H. G.: Influence of Lattice Defects on the Electrical Properties of Cold-Worked Metals. — Philips Res. Rep. 12 (1957), s. 1—45 oraz 190—239.
13. Ellis S. G.: Microscopic Examination of Germanium Crystals and Transitors. W pracy zbiorowej: Transistors. I. Bell. Labs. Princeton, 1956.
14 Tyler W. W., Dash W. C.: Dislocation Arrays in Germanium. — Journ. Appl. Phys. 28 (1957), s. 1221.
15. Brochocki A., Kobus A.: Wpływ warunków wyciągania monokryształów germanu na rozkład dyslokacji krawędziowych. Komunikat wygłoszony na I Krajowej Naradzie Elektroniki — Warszawa — X — 1958 r.
16. Auleytner J., Godwood K., Krilov J.: Observations on Etching of Germanium Crystals. — Bull. l'Acad. Pol. d. Sci. Cl. 3. V (1957), Nr 6, s. 639.

А. КОБУС

ХИМИЧЕСКИЕ МЕТОДЫ ОБНАРУЖЕНИЯ ДИСЛОКАЦИИ В КРИСТАЛЛАХ ГЕРМАНИЯ

Резюме

В статье обсуждаются химические методы обнаружения дислокации в кристаллах германия. В результате соответственной химической обработки на определённых поверхностях кристалла возникают углубления соответствующие точке пересечения дислокационной линии и поверхности кристалла. Доказательство совпадения углублений с дислокациями в кристаллах германия предложено в 1953 г. Фогелём и др. Механизм возникновения углублений ещё точно не изучён, хотя известно, что углубление возникает из-за появления энергетической аномалии в кристаллической решётке в окружности дислокации. Описаны некоторые химические смеси применяемые для обнаружения дислокации и приведены микроскопические снимки представляющие фигуры полученные при воздействии разных реактивов на поверхность. Получено, что химическое травление обнаруживает две основные группы углублений, которых плотность в типовых кристаллах германия равна соответственно: 10^2—10^4 см$^{-2}$ и 10^5—10^7 см$^{-2}$. Углубления с менее значительной плотностью создают характеристические системы, являющиеся функцией способа вытягивания монокристалла; такие углубления относятся к краевым дислокациям. Углубления второго рода расположены беспорядочно; они относятся к дислокациям с ориентацией различной от краевой. В заключении обсуждена пригодность отдельных травлющих смесей применяемых для обнаружения дислокации.

A. KOBUS

CHEMICAL METHODS OF DISCLOSING DISLOCATIONS IN GERMANIUM CRYSTALS

Summary

The paper presents a review of chemical methods for disclosing dislocations in germanium crystals. After an appropriate chemical etching, on the surface of crystal there appear etch pits correspond to intersections of the dislocation line and the surface. The chemical methods, in spite of their drawbacks (one does not know whether all the dislocations are disclosed), distinguish themselves by simplicity and do not require a big equipment. A number of reagents applied to disclosing dislocations are given and their usefulness is discussed.

A. KOBUS

MÉTHODES CHIMIQUES DE LA DÉCOUVERTE DES DISLOCATIONS DANS LES CRISTAUX DE GERMANIUM

Résumé

L'article présente une révue des méthodes chimiques employées pour la découverte des dislocations dans les cristaux de germanium. Après un traitement chimique des figures de corrosion se forment à la surface du cristal, correspondant à l'intersection de la ligne de dislocation avec la surface. Les méthodes chimiques quoiqu'elles ne sont pas sans reproche (il n'est pas sûr, si toutes les dislocations ont été découvertes) se distinguent par leur simplicité et n'exigent pas de grand équipement. On a presenté un nombre de réactifs employés pour la découverte de dislocations et on a discuté leur utilité.

A. KOBUS

CHEMISCHE METHODEN DER ENTDECKUNG DER VERSETZUNGEN IN GERMANIUMKRISTALLEN

Zusammenfassung

Der Artikel gibt die Übersicht der chemischen Methoden bei Entdeckung der Versetzung in Germaniumkristallen. Nach entsprechender chemischer Bearbeitung entstehen auf der Oberfläche des Kristalls die Ätzpünktchen, die der Kreuzung der Versetzunglinie mit der Oberfläche entsprechen. Die chemischen Methoden, obwohl sie nicht vollkommen fehlerfrei sind (es ist nicht ganz sicher, ob alle Versetzungen bereits entdeckt worden sind) zeichnen sich mit grosser Einfachkeit aus und verlangen keine grosse Ausstattung. Verschiedene anwendbare Reagenten wurden angegeben und ihre Nützlichkeit erörtert.

621.314.7

WŁADYSŁAW MAJEWSKI JR

Parametry falowe i robocze tranzystora

Rękopis dostarczono 31.3.1959

Artykuł stanowi próbę zastosowania klasycznej teorii czwórników do układów tranzystorowych. Podano w nim podstawowe wzory z teorii czwórników a następnie zastosowano je do układów tranzystorowych. Wyprowadzono wzory na oporności falowe i tamowności falowe tranzystora w trzech podstawowych układach pracy. Podane przykłady liczbowe dotyczą pojedynczych tranzystorów i wzmacniacza o sprzężeniu transformatorowym.

Następnie podano przykłady obliczania parametrów roboczych tranzystorów i ich łańcuchów.

1. ELEMENTY TEORII CZWÓRNIKÓW

1.1. Zagadnienia ogólne

Klasyczna teoria czwórników powstała na gruncie układów biernych — przede wszystkim filtrów i korektorów. Do czasu wprowadzenia tranzystora nie znajdowała ona większego zastosowania do czwórników czynnych. Lampa elektronowa ze względu na wybitnie jednokierunkowe działanie nie jest typowym czwórnikiem czynnym i rzadko bywa rozpatrywana na gruncie teorii czwórników. Dopiero tranzystor mogący stanowić klasyczny przykład czwórnika czynnego wymaga stosowania teorii czwórników. Mimo to nie wszystkie narzędzia klasycznej teorii czwórników znalazły zastosowanie w teorii układów tranzystorowych. Powszechnie stosowane są równania zasadnicze czwórnika w różnych postaciach i związane z nimi parametry opornościowe, przewodnościowe, mieszane itp. Parametry te podawane są jako wielkości pierwotne charakteryzujące właściwości tranzystora. Parametry wtórne nie są oparte na teorii czwórników, lecz tworzone są w sposób analogiczny do stosowanego w teorii układów lampowych. Do tych parametrów należą współczynniki wzmocnienia napięcia [V/V], prądu [mA/mA], mocy [mW/mW]. Wprowadza się również

parametry analogiczne do wzmocności skutecznej, a mianowicie wzmocnienie mocy odniesienia i niekiedy jego odpowiednik napięciowy [3], oraz analogiczne do oporności falowych, a mianowicie optymalne oporności generatora i obciążenia.

Wielkości te są wprowadzane z reguły w sposób odmienny od stosowanego w teorii czwórników[1]).

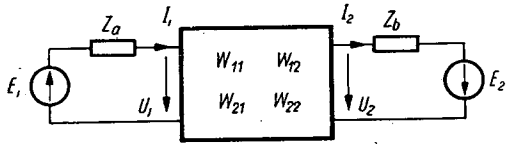
Artykuł niniejszy stanowi próbę konsekwentnego zastosowania klasycznej teorii czwórników do układów tranzystorowych.

Rozdział pierwszy artykułu zawiera podstawowe elementy klasycznej teorii czwórników w ujęciu stosowanym w teletransmisji przewodowej[2]) w zakresie koniecznym do zrozumienia dalszych rozważań.

Przyjmując, że czytelnik w zasadzie zna podstawy teorii czwórników w rozdziale tym ujęto zagadnienie teorii czwórników w sposób szkicowy z pominięciem dowodów. Oznaczenia stosowane w niniejszym rozdziale zgodne są z oznaczeniami używanymi w teorii czwórników [1]. Podawane są również oznaczenia stosowane w teorii układów tranzystorowych.

1. 2. Parametry oporowe czwórnika

Czwórnik pracujący między opornościami Z_a i Z_b[3]) jak na rys. 1 może być opisany układem następujących równań:



Rys. 1. Czwórnik między źródłem a odbiornikiem

$$U_1 = W_{11} I_1 + W_{12} I_2, \tag{1a}$$

$$U_2 = W_{21} I_1 + W_{22} I_2, \tag{1b}$$

$$U_1 = E - Z_a I_1, \tag{2a}$$

$$U_2 = - Z_b I_2. \tag{2b}$$

Równania (1) są to równania oporowe czwórnika. Równania (2) są to równania źródła i odbiornika przy kierunku przekazywania energii z lewa na prawo, tj. gdy $E_1 = E$ i $E_2 = 0$.

---

[1]) Wyjątek stanowi artykuł [4], w którym wprowadza się oporności optymalne przez pomiar oporności wejściowej w stanie jałowym i w stanie zwarcia [patrz wzór (10) niniejszego artykułu].

[2]) Rozdział ten oparty jest głównie na pracy [1].

[3]) Oporności Z_a i Z_b, podobnie jak wszystkie inne oporności, napięcia i prądy występujące w niniejszych rozważaniach, mogą być w ogólnym przypadku wielkościami zespolonymi. Jednak dla uproszczenia „daszki" nie będą stosowane.

Równania (2) przy kierunku przekazywania energii z prawa na lewo, tj. gdy $E_1 = 0$ i $E_2 = E$, przybiorą postać:

$$U_1 = -Z_a I_1, \tag{3a}$$

$$U_2 = E - Z_b I_2. \tag{3b}$$

Parametry W_{11}, W_{12}, W_{21}, W_{22} charakteryzują sam czwórnik i noszą nazwę parametrów oporowych. W teorii układów tranzystorowych oznaczane są one zwykle przez r_{11}, r_{12}, r_{21}, r_{22}.

Równania (1) mogą być przedstawione również w postaci macierzowej. Odpowiednia macierz czwórnika nosi nazwę macierzy opornościowej lub szeregowej.

Oporności źródła i odbiornika Z_a i Z_b oznaczane są zwykle w teorii układów tranzystorowych odpowiednio przez R_g i R_o.

Wyjaśnimy teraz sens fizyczny parametrów oporowych czwórnika i omówimy metodę ich wyznaczania.

Załóżmy najpierw, że energia przekazywana jest z lewa na prawo i ponadto, że $I_2 = 0$ (stan jałowy po stronie wtórnej). Wtedy z równania (1a) otrzymamy:

$$W_{11} = \left[\frac{U_1}{I_1}\right]_{I_2=0} \tag{4a}$$

Ze względu na to, że oporność wejściowa pierwotna stanu jałowego określona jest zależnością

$$W_{1o} = \left[\frac{U_1}{I_1}\right]_{I_2=0}, \tag{4b}$$

otrzymujemy ostatecznie

$$W_{11} = W_{1o}; \tag{4c}$$

W_{11} równe jest więc oporności wejściowej pierwotnej stanu jałowego.

Biorąc pod uwagę równanie (2a) otrzymujemy przy tych samych założeniach

$$W_{21} = \left[\frac{U_2}{I_1}\right]_{I_2=0}. \tag{5a}$$

Oznaczając ten stosunek przez M_1 i nazywając go opornością wzajemną pierwotną czwórnika otrzymujemy

$$W_{21} = M_1. \tag{5b}$$

Załóżmy teraz, że energia przekazywana jest z prawa na lewo i ponadto, że $I_1 = 0$ (stan jałowy po stronie pierwotnej czwórnika). Postępując w sposób analogiczny jak poprzednio otrzymamy teraz

$$W_{22} = \left[\frac{U_2}{I_2}\right]_{I_1 = 0} = -W_{2o}, \tag{6}$$

$$W_{12} = \left[\frac{U_1}{I_2}\right]_{I_1 = 0} = -M_2. \tag{7}$$

Znaki " — " w powyższych równaniach wynikają z przyjętych kierunków strzałkowania, a wielkości W_{2o} i M_2 oznaczają odpowiednio oporność wejściową wtórną stanu jałowego i oporność wzajemną wtórną.

Wzory (4) ÷ (7) dają metodę pomiaru parametrów oporowych. Pomiar W_{1o} i W_{2o} jest łatwy i często stosowany, pomiar M_1 i M_2 jest trudniejszy.

1.3. Parametry falowe czwórnika

Czwórnik scharakteryzować można także przy pomocy innych rodzin parametrów. Do wielu celów bardzo przydatna jest rodzina parametrów falowych. Należą do niej oporności falowe i tamowności falowe.

Zdefiniujemy najpierw oporności falowe. Opornościami falowymi pierwotną i wtórną nazywamy parę oporności Z_1, Z_2 o następujących własnościach:

a) jeżeli oporność obciążenia $Z_b = Z_2$, to oporność wejściowa pierwotna $W_1 = Z_1$ (kierunek przekazywania energii z lewa na prawo)

b) jeżeli oporność obciążenia $Z_a = Z_1$, to oporność wejściowa wtórna $W_2 = Z_2$ (kierunek przekazywania energii z prawa na lewo).

Oprócz oporności falowych pierwotnej Z_1 i wtórnej Z_2 wprowadza się także oporność falową średnią Z i przekładnię oporową p.

Wielkości te związane są wzorami

$$Z_1 = \frac{Z}{p}, \tag{8a}$$

$$Z_2 = pZ. \tag{8b}$$

Powyższe parametry falowe związane są z parametrami oporowymi następującymi wzorami:

$$Z = \sqrt{W_{11}W_{22} - W_{12}W_{21}}, \tag{9a}$$

$$p = \sqrt{\frac{W_{22}}{W_{11}}}. \tag{9b}$$

Oporności falowe mogą być również wyznaczone za pomocą pomiaru oporności wejściowych w stanie jałowym i w stanie zwarcia ze wzorów

$$Z_1 = \sqrt{W_{1o} W_{1z}}, \tag{10a}$$

$$Z_2 = \sqrt{W_{2o} W_{2z}}, \tag{10b}$$

gdzie W_{1z} i W_{2z} — oznaczają odpowiednią oporność wejściową pierwotną i wtórną stanu zwarcia.

W teorii układów tranzystorowych wielkość W_{1z} jest oznaczana przez h_{11} i stanowi jeden z parametrów mieszanych tranzystora.

Zdefiniujemy obecnie tamowności falowe. Tamownością falową pierwotną nazywamy wielkość charakteryzującą przekazywanie mocy zespolonej z lewa na prawo w przypadku, gdy $Z_b = Z_2$ (stan dopasowania falowego po stronie wtórnej).

$$\Gamma_1 = \frac{1}{2} \ln \frac{U_1 I_1}{U_2 I_2}. \tag{11}$$

Tamownością falową wtórną nazywamy wielkość charakteryzującą przekazywanie mocy zespolonej z prawa na lewo w przypadku, gdy $Z_a = Z_1$ (stan dopasowania falowego po stronie pierwotnej).

$$\Gamma_2 = \frac{1}{2} \ln \frac{U_2 I_2}{U_1 I_1}. \tag{12}$$

Wielkości Γ_1 i Γ_2 są w ogólnym przypadku wielkościami zespolonymi i mogą być przedstawione w postaci części rzeczywistej i urojonej. Rozważmy dla przykładu

$$\Gamma_1 = A_1 + jB_1, \tag{13}$$

gdzie wielkość

$$A_1 = \frac{1}{2} \ln \left| \frac{U_1 I_1}{U_2 I_2} \right| \text{ przy } Z_b = Z_2$$

charakteryzuje tłumienie mocy pozornej spowodowane przez czwórnik i nosi nazwę tłumienności falowej pierwotnej.

Wielkość $B_1 = \arg \frac{U_1 I_1}{U_2 I_2}$ przy $Z_b = Z_2$ charakteryzuje przesunięcie fazowe wprowadzane przez czwórnik i nosi nazwę przesuwności falowej pierwotnej.

Podobnie

$$\Gamma_2 = A_2 + jB_2. \tag{14}$$

Oprócz tamowności falowej pierwotnej i wtórnej wprowadza się także tamowność falową średnią i przekładnię energetyczną. Wielkości te związane są wzorami:

$$\Gamma_1 = \Gamma + \ln p_e, \tag{15a}$$

$$\Gamma_2 = \Gamma - \ln p_e. \tag{15b}$$

Powyższe parametry falowe związane są z parametrami oporowymi następującymi wzorami[1]):

$$\Gamma = \operatorname{artgh}\frac{Z_1}{W_{11}} = \operatorname{artgh}\frac{Z_2}{W_{22}}, \tag{16a}$$

$$p_e = \sqrt{\frac{W_{12}}{W_{21}}}. \tag{16b}$$

Zamiast tłumienności falowej pierwotnej (ew. wtórnej) wprowadza się niekiedy — szczególnie w przypadku czwórników czynnych — wzmocność falową pierwotną (ew. wtórną) określoną wzorem:

$$S_1 = -A_1 = \frac{1}{2}\ln\left|\frac{U_2 I_2}{U_1 I_1}\right|. \tag{17}$$

Wprowadzić można również wzmocność falową napięciową i prądową pierwotną (ew. wtórną)[2]):

$$S_{1u} = \ln\left|\frac{U_2}{U_1}\right| = S_1 + \frac{1}{2}\ln p, \tag{18a}$$

$$S_{1i} = \left|\frac{I_2}{I_1}\right| = S_1 - \frac{1}{2}\ln p. \tag{18b}$$

Z wzorów powyższych można wyznaczyć współczynniki wzmocnienia napięcia i prądu:

$$k_u = \left|\frac{U_2}{U_1}\right| = e^{S_{1u}}, \tag{19a}$$

$$k_i = \left|\frac{I_2}{I_1}\right| = e^{S_{1i}}. \tag{19b}$$

Wielkości A_1, A_2, S_1, S_2, $\ln p$, $\ln p_e$ itd. są zwykle wyrażone w neperach [N]. Przeliczenie na decybele [dB] opiera się na wykorzystaniu przybliżonej równości $1\text{N} \approx 8{,}68$ dB.

1.4. Metoda rachunkowa wyznaczania parametrów falowych na podstawie parametrów oporowych

Wyznaczanie oporności falowych Z_1 i Z_2 nie nastręcza żadnych trudności i może być przeprowadzane bezpośrednio wg wzrów (9) i (8).

---

[1]) Wzór ten wynika ze wzorów (4.34a) i (4.34b) pracy [1].

[2]) Wprowadzenie takich wielkości jest szczególnie wygodne w teorii czwórników czynnych.

Natomiast wyznaczanie Γ_1 i Γ_2, a ściśle mówiąc etap pośredni — wyznaczenie Γ — wymaga dodatkowych wyjaśnień. Zagadnienie to sprowadza się do rozwiązania następującego problemu matematycznego: z równania $\operatorname{tgh}(A + jB) = Te^{j\varphi_T}$ wyznaczyć A i B, gdy T i φ_T są dane.

Można wykazać, że rozwiązanie powyższego problemu prowadzi do następujących równań:

$$\operatorname{tgh} 2A = \frac{2T}{1+T^2}\cos\varphi_T, \tag{20a}$$

$$\operatorname{tg} 2B = \frac{2T}{1-T^2}\sin\varphi_T. \tag{20b}$$

Wzór (20b) określa kąt $2B$ z dokładnością do wielokrotności π. Ponadto można wykazać, że drugie ramię tego kąta położone jest:

a) w pierwszej ćwiartce, gdy $T < 1$ i $0 < \varphi_T < \frac{\pi}{2}$,

b) w drugiej ćwiartce, gdy $T > 1$ i $0 < \varphi_T < \frac{\pi}{2}$,

c) w trzeciej ćwiartce, gdy $T > 1$ i $-\frac{\pi}{2}, < \varphi_T < 0$,

d) w czwartej ćwiartce, gdy $T < 1$ i $-\frac{\pi}{2}, < \varphi_T < 0$,

Rozważamy obecnie ważny w praktyce przypadek, gdy T jest rzeczywiste.

Jeżeli $T < 1$, a $\varphi_T = 0$,

to

$$\operatorname{tgh} A = T, \tag{21a}$$

$$2B = 0. \tag{21b}$$

Jeżeli $T > 1$, a $\varphi_T = 0$, to $2B = \pi$ i A trzeba obliczyć ze wzoru (20a). Istnieją nomogramy i tablice funkcji tgh argumentu zespolonego umożliwiające szybkie, chociaż mniej dokładne rozwiązanie powyższego zagadnienia.

1.5. Łańcuchy czwórników zbudowane na zasadzie dopasowania falowego

Zajmiemy się z kolei wyznaczaniem parametrów falowych zastępczego czwórnika dla pewnego szczególnego rodzaju łańcuchów, a mianowicie łańcuchów zbudowanych na zasadzie dopasowania falowego. Łańcuchem zbudowanym na zasadzie dopasowania falowego nazywamy łańcuch, w którym dla dowolnej pary czwórników oporność falowa pier-

wotna czwórnika następnego równa jest oporności falowej wtórnej czwórnika poprzedniego.

Oporność falowa pierwotna czwórnika zastępczego dla tak utworzonego łańcucha równa jest oporności falowej pierwotnej pierwszego czwórnika. Oporność falowa wtórna równa jest oporności falowej wtórnej ostatniego czwórnika. Tamowność falowa pierwotna (wtórna) jest równa sumie tamowności falowych pierwotnych (wtórnych) wszystkich czwórników składowych.

1.6. Oporność wejściowa czwórnika

Oporności falowe pierwotna i wtórna są szczególnymi wartościami oporności wejściowej pierwotnej i wtórnej czwórnika. Natomiast w przypadku ogólnym oporności wejściowe określone są następującymi wzorami:

$$W_1 = Z_1 \frac{1 + q_2 e^{-2\Gamma}}{1 - q_2 e^{-2\Gamma}}, \tag{22a}$$

$$W_2 = Z_2 \frac{1 + q_1 e^{-2\Gamma}}{1 - q_1 e^{-2\Gamma}}, \tag{22b}$$

gdzie Γ jest tamownością średnią czwórnika, zaś q_1 i q_2 noszą nazwę współczynników niedopasowania i określone są wzorami:

$$q_1 = \frac{Z_a - Z_1}{Z_a + Z_1}, \tag{23a}$$

$$q_2 = \frac{Z_b - Z_2}{Z_b + Z_2}. \tag{23b}$$

Istnieją nomogramy ułatwiające wyznaczanie współczynników odbicia (np. [1] str. 169).

Ze wzorów (22) wynika oczywiście określenie oporności falowych: np. przy $Z_b = Z_2$ mamy $q_2 = 0$ i $W_1 = Z_1$.

1.7. Tłumienność skuteczna i wzmocność skuteczna czwórnika

Tłumienność falowa czwórnika charakteryzuje sam czwórnik, nie mówi natomiast nic o jakości współpracy czwórnika z danym źródłem i danym odbiornikiem. Informacji o tym dostarcza wielkość nazywana tłumiennością skuteczną.

Tłumienność skuteczna pierwotna czwórnika pracującego między opornościami Z_a i Z_b określona jest wzorem:

$$A_{s1} = \frac{1}{2} \ln \left| \frac{U_s I_s}{U_2 I_2} \right|; \tag{24}$$

$|U_2\ I_2|$ jest mocą pozorną wydzielającą się na oporności Z_b,

$|U_s\ I_s|$ jest mocą pozorną odniesienia, która wydzieliłaby się na oporności Z_a połączonej bezpośrednio ze źródłem o oporności Z_a. W przypadku, gdy Z_a jest rzeczywiste, jest to największa moc, jaką może dać źródło. Podobnie tłumienność skuteczna wtórna określona jest wzorem:

$$A_{s2} = \frac{1}{2} \ln \left| \frac{U_s I_s}{U_1 I_1} \right|. \tag{25}$$

Tłumienność skuteczna pierwotna (wtórna) może być obliczona z następującego wzoru:

$$A_{s1(2)} = A_{j(2)} + \ln \left| \frac{Z_a + Z_1}{2\sqrt{Z_a Z_1}} \right| + \ln \left| \frac{Z_b + Z_2}{2\sqrt{Z_b Z_2}} \right| + \ln |1 - q_1 q_2 e^{-2\Gamma}|. \tag{26}$$

Istnieją nomogramy ułatwiające wyznaczanie poszczególnych składników powyższego wzoru (np. [1] str. 179).

Ze wzoru (26) wynika, że w przypadku obustronnego dopasowania falowego tłumienność skuteczna równa jest tłumienności falowej. Jeżeli oporności źródła i odbiornika Z_a i Z_b są rzeczywiste, to stan dopasowania falowego jest stanem optymalnym ze względu na przekazywanie mocy.

Niedopasowania po stronie pierwotnej lub wtórnej zwykle pogarszają warunki przekazywania mocy. Pogorszenie to jest scharakteryzowane właśnie przyrostem tłumienności skutecznej.

Oprócz tłumienności skutecznej pierwotnej (ew. wtórnej) wprowadza się także wzmocność skuteczną pierwotną (ew. wtórną) określoną wzorem:

$$S_{s1} = -A_{s1} = \frac{1}{2} \ln \left| \frac{U_2 I_2}{U_s I_s} \right|. \tag{27}$$

Jest to wielkość analogiczna do stosowanego w teorii układów tranzystorowych wzmocnienia mocy odniesienia:

$$k_{P_o} = \left| \frac{U_2 I_2}{\frac{E^2}{4 Z_a}} \right| = \left| \frac{U_2 I_2}{U_s I_s} \right|. \tag{28}$$

Istnieje następujący związek między tymi wielkościami:

$$k_{P_o} = e^{2 S_{s1}} = e^{-2 A_{s1}}.$$

Oporności wejściowe i tłumienność (wzmocność) skuteczna noszą w odróżnieniu od parametrów falowych nazwę parametrów roboczych.

Wzmocność skuteczna pierwotna (wtórna) może być obliczona z następującego wzoru analogicznego do (26):

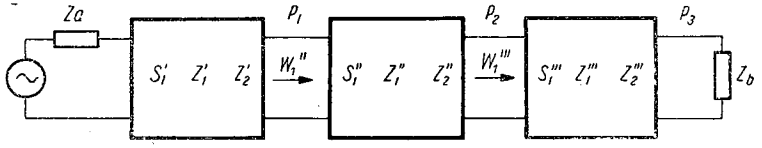
$$S_{s1(2)} = S_{1(2)} - \ln\left|\frac{Z_a + Z_1}{2\sqrt{Z_a Z_1}}\right| - \ln\left|\frac{Z_b + Z_2}{2\sqrt{Z_b Z_2}}\right| - \ln|1 - q_1 q_2 e^{-2\Gamma}|. \quad (29a)$$

Wzór ten zapisuje się często w następującej formie skróconej:

$$S_{s1} = S_{1(2)} - \Delta_1 - \Delta_2 - \Delta_{12}. \quad (29b)$$

1.8. Tłumienność skuteczna i wzmocność skuteczna łańcucha dowolnych czwórników

Rozważymy łańcuch trzech czwórników (rys. 2), co do których nie stawiamy żadnych ograniczeń pod względem ich wzajemnego dopasowania.



Rys. 2. Łańcuch trzech czwórników

Nasze rozumowanie może być bez żadnych trudności uogólnione na większą ilość czwórników.

Oznaczając moce wydzielone w kolejnych przekrojach przez P_1, P_2, P_3, a moc odniesienia przez P_s, możemy napisać:

$$S_{s1} = \frac{1}{2}\ln\frac{P_3}{P_s} = \frac{1}{2}\ln\frac{P_1}{P_s}\cdot\frac{P_2}{P_1}\cdot\frac{P_3}{P_2} = S'_{s1} + S''_{skr1} + S'''_{skr1}. \quad (30)$$

We wzorze powyższym zastosowano następujące oznaczenia:

S_{s1} — wzmocność skuteczna pierwotna całego łańcucha,
S'_{s1} — wzmocność skuteczna pierwotna pierwszego czwórnika,
S''_{skr1} — wzmocność skrośna[1]) pierwotna drugiego czwórnika,
S'''_{skr2} — wzmocność skrośna pierwotna trzeciego czwórnika.

[1]) Wzmocnością (tłumiennością) skrośną czwórnika nazywamy wielkość wyrażoną tym samym wzorem co wzmocność (tłumienność) falową, lecz bez stawiania warunku dotyczącego wartości oporności obciążenia czwórnika.

Pierwszy etap obliczenia polega na kolejnym wyznaczeniu oporności wejściowych pierwotnych trzeciego czwórnika (W_1''') i drugiego czwórnika (W_1''). Następnie obliczamy wzmocności skrośne wg wzorów:

$$S'''_{skr1} = S'''_1 - \ln\left|\frac{Z_b + Z'''_2}{2\sqrt{Z_b Z'''_2}}\right| - \frac{1}{2}\ln|1 - (q'''_2)^2 e^{-4\Gamma'''}|, \tag{31}$$

$$S''_{skr1} = S''_1 - \ln\left|\frac{W'''_1 + Z''_2}{2\sqrt{W'''_1 Z''_2}}\right| - \frac{1}{2}\ln|1 - (q''_2)^2 e^{-4\Gamma''}|, \tag{32}$$

oraz wzmocność skuteczna S'_{s1} wg wzoru (29).

2. PARAMETRY FALOWE TRANZYSTORA

2.1. Zagadnienia ogólne

Zastosujemy obecnie ogólną teorię czwórników w ujęciu klasycznym rozważaną w poprzednim rozdziale do układów tranzystorowych. Wyprowadzimy ogólne wzory wiążące parametry falowe tranzystora w trzech różnych układach pracy z jego parametrami własnymi (r_e, r_b, r_m, r_k) oraz podamy przykłady obliczenia parametrów falowych. Następnie wyznaczymy parametry falowe łańcucha utworzonego z tranzystorów dopasowanych do siebie falowo. Początkowo zakładać będziemy, że parametry własne tranzystora są rzeczywiste, potem rozpatrzymy również przypadek parametrów zespolonych.

2.2. Parametry falowe tranzystora w układzie o wspólnej bazie (WB)

Z teorii układów tranzystorowych znane są następujące związki między parametrami oporowymi i parametrami własnymi tranzystora w układzie o wspólnej bazie:

$$W_{11} = r_{11} = r_e + r_b, \tag{33a}$$

$$W_{12} = r_{12} = r_b, \tag{33b}$$

$$W_{21} = r_{21} = r_m + r_b, \tag{33c}$$

$$W_{22} = r_{22} = r_k + r_b. \tag{33d}$$

Wyznaczymy najpierw oporności falowe.

Podstawiając (33) do (9) otrzymujemy:

$$Z = \sqrt{r_{11} r_{22} - r_{12} r_{21}} = \sqrt{(r_e + r_b)(r_k + r_b) - r_b (r_m + r_b)},$$

$$p = \sqrt{\frac{r_{22}}{r_{11}}} = \sqrt{\frac{r_k + r_b}{r_e + r_b}}.$$

Zakładając $r_b \ll r_k$ i $r_b \ll r_m$, co jest zawsze spełnione, oraz biorąc pod uwagę, że $r_m = \alpha r_k$, otrzymujemy następujące wzory przybliżone:

$$Z = \sqrt{r_k [r_e + r_b (1 - \alpha)]}, \tag{34}$$

$$p = \sqrt{\frac{r_k}{r_e + r_b}}. \tag{35}$$

Podstawiając otrzymane wyrażenia do (8) otrzymujemy następujące wyrażenia na Z_1 i Z_2:

$$Z_1 = \sqrt{(r_e + r_b) [r_e + r_b (1 - \alpha)]}, \tag{36}$$

$$Z_2 = r_k \sqrt{\frac{r_e + r_b (1 - \alpha)}{r_e + r_b}}. \tag{37}$$

Otrzymane wyrażenia są identyczne ze znanymi z teorii układów tranzystorowych wyrażeniami na optymalne wartości oporności źródła i odbiornika. Wyrażenia te uzyskuje się zwykle przez poszukiwanie maksimum funkcji

$$k_{p_o} = f(R_g, R_o). \tag{38}$$

Sprowadza się to do obliczania pochodnych cząstkowych i przyrównywania ich do zera.

Oczywiście z definicji parametrów falowych i skutecznych czwórnika wynika, że stan dopasowania falowego jest stanem optymalnym[1]) ze względu na przekazywanie mocy ze źródła do odbiornika. Dlatego zgodność wzorów na oporności falowe z wzorami na optymalne oporności źródła i generatora nie jest przypadkowa.

Wyznaczymy teraz pozostałe parametry falowe.

Podstawiając (37) i (33d) do (16a) otrzymujemy:

$$\operatorname{tgh} \Gamma = \frac{Z_2}{r_{22}} \cong \frac{Z_2}{r_k} = \sqrt{\frac{r_e + r_b (1 - \alpha)}{r_e + r_b}}. \tag{39}$$

Zakładając, że r_e, r_b, r_k, α są rzeczywiste stwierdzamy, że

$$\frac{r_e + r_b (1 - \alpha)}{r_e + r_b} < 1.$$

Możemy więc stosować wzory (21), z których otrzymujemy:

$$A = \operatorname{artgh} \frac{Z_2}{r_k}, \tag{40}$$

$$B = 0. \tag{41}$$

[1]) W przypadku parametrów rzeczywistych.

Jeżeli r_e, r_b, r_k, α są zespolone, to A i B wyznaczyć możemy ze wzorów (20).

Podstawiając (33) do (16a) otrzymujemy

$$p_e = \sqrt{\frac{r_{12}}{r_{21}}} \cong \sqrt{\frac{r_b}{\alpha\, r_k}}. \tag{42}$$

W dalszym ciągu ze wzorów (15) otrzymujemy następujące wyrażenia[1]):

$$S_1 = -\operatorname{artgh}\frac{Z_2}{r_k} + \ln\sqrt{\frac{\alpha\, r_k}{r_b}}, \tag{43}$$

$$B_1 = 0 \tag{44}$$

$$A_2 = \operatorname{artgh}\frac{Z_2}{r_k} + \ln\sqrt{\frac{\alpha\, r_k}{r_b}}, \tag{45}$$

$$B_2 = 0. \tag{46}$$

Z otrzymanych wzorów wynikają następujące wnioski:

1) wzmocność falowa pierwotna tranzystora jest dodatnia, a więc tranzystor w układzie o wspólnej bazie wprowadza wzmocnienie przy przekazywaniu jej od wejścia do wyjścia,

2) tłumienność falowa wtórna tranzystora jest dodatnia, a więc tranzystor w układzie o wspólnej bazie wprowadza tłumiene mocy przy przekazywaniu jej od wyjścia do wejścia.

3) przesunięcia fazy wprowadzane przez tranzystor równe są zeru dla obu kierunków przekazywania mocy (oczywiście w założeniu rzeczywistych parametrów własnych).

Wyrażenie (43) stanowi odpowiednik wyrażenia na maksymalne wzmocnienie mocy odniesienia $k_{p_o\,\max}$, które może być otrzymane z (38).

Ze względu na skomplikowane postacie wzorów wykazanie równoważności tych wyrażeń nie jest łatwe. Równoważność ta wynika jednak z definicji parametrów falowych.

Przykład 2 — 1

Przyjmując przeciętnie spotykane dla tranzystora warstwowego wartości parametrów własnych $r_b = 500\,\Omega$; $r_e = 25\,\Omega$; $r_k = 1\,\mathrm{M}\Omega$; $\alpha = 0{,}96$ otrzymujemy ze wzorów (34) ÷ (37):

$$Z = \sqrt{r_k\,[r_e + r_b\,(1-\alpha)]} = 6{,}72\,\mathrm{k}\,\Omega$$

oraz

$$p = \sqrt{\frac{r_k}{r_e + r_b}} = 43{,}7;$$

[1]) Przy założeniu rzeczywistych parametrów własnych.

$$Z_1 = \frac{Z}{p} = 154\,\Omega$$

oraz

$$Z_2 = p\,Z = 293\,\mathrm{k}\Omega.$$

Następnie z (40) i (41) otrzymujemy

$$A = \operatorname{artgh}\frac{Z_2}{r_k} = \operatorname{artgh} 0{,}293 = 0{,}32\,\mathrm{N},$$

$$p_e = \sqrt{\frac{r_b}{a\,r_k}} = \frac{1}{43{,}8},$$

$$\ln\sqrt{\frac{a\,r_k}{r_b}} = -\ln p_e = 3{,}78\,\mathrm{N}.$$

Ostatecznie korzystając z (43) ÷ (46) mamy:

$$S_1 = -0{,}32 + 3{,}78 = 3{,}46\,\mathrm{N} = 30\,\mathrm{dB},$$

$$A_2 = 0{,}32 + 3{,}78 = 4{,}10\,\mathrm{N} = 35{,}5\,\mathrm{dB},$$

$$B_1 = B_2 = 0.$$

Przykład 2 — 2

Obliczmy parametry falowe tranzystora przy częstotliwości granicznej. Przyjmiemy następujące wartości parametrów własnych: $r_b = 500\,\Omega$, $r_e = 25\,\Omega$, $r_k = 1\,\mathrm{M}\Omega$ oraz

$$a = \frac{0{,}96}{\sqrt{2}}\,e^{-j\,45^\circ} = 0{,}68\,e^{-j\,45^\circ}.$$

Obliczamy najpierw oporności falowe.

Wprowadzając oznaczenia:

$$a = r_e + r_b(1 - a),$$
$$b = r_e + r_b$$

otrzymujemy:

$$Z_1 = \sqrt{ab},$$
$$Z_2 = r_k\sqrt{\frac{a}{b}}.$$

Obliczając a i b dostajemy:

$$a = 373\,e^{j\,43^\circ},$$
$$b = 525.$$

Mamy stąd:

$$Z_1 = 443\,e^{j\,21{,}5^\circ}\,\Omega,$$
$$Z_2 = 842\,e^{j\,21{,}5^\circ}\,\mathrm{k}\Omega.$$

Obliczamy następnie przekładnię energetyczną p_e i średnią tamowność falową Γ. Otrzymujemy:

$$p_e = \sqrt{\frac{r_b}{\alpha r_k}} = \frac{1}{36{,}9} e^{j\,22{,}5^\circ}, \quad \text{skąd } \ln p_e = -3{,}6\,\mathrm{N},$$

$$\operatorname{tgh} \Gamma = \frac{Z_2}{r_k} = 0{,}842\, e^{j\,21{,}5^\circ}.$$

Korzystamy teraz ze wzorów (20) i otrzymujemy

$$A = \frac{1}{2} \operatorname{artgh} \frac{2 \cdot 0{,}842}{1 + (0{,}842)^2} \cos 21{,}5^\circ = 0{,}77,$$

$$B = \frac{1}{2} \operatorname{arctg} \frac{2 \cdot 0{,}842}{1 - (0{,}842)^2} \sin 21{,}5^\circ = 32{,}5^\circ.$$

Ponadto stwierdzamy, że drugie ramię kąta 2B leży w pierwszej ćwiartce.

Teraz możemy wyznaczyć tamowności falowe pierwotną i wtórną oraz ich składowe.

$$\Gamma_1 = 0{,}77 - 3{,}6 + j\,22{,}5^\circ + j\,32{,}5^\circ = -2{,}83 + j\,55^\circ.$$

Otrzymujemy stąd $S_1 = 2{,}83$ N.

Dalej

$$\Gamma_2 = 0{,}77 + 3{,}6 - j\,22{,}5^\circ + j\,32{,}5^\circ = 4{,}37 + j\,10^\circ.$$

otrzymujemy stąd $A_2 = 4{,}37$ N.

2.3. Parametry falowe tranzystora w układzie o wspólnym emiterze (WE)

Z teorii układów tranzystorowych znane są następujące związki między parametrami oporowymi a parametrami własnymi tranzystora w układzie o wspólnym emiterze:

$$W_{11} = r_{11} = r_b + r_e, \tag{47a}$$

$$W_{12} = r_{12} = r_e, \tag{47b}$$

$$W_{21} = r_{21} = r_e - r_m, \tag{47c}$$

$$W_{22} = r_{22} = r_e + r_k - r_m. \tag{47d}$$

Postępując jak w poprzednim punkcie możemy otrzymać następujące wyrażenia na oporność falową średnią Z i przekładnię oporową p.

$$Z = \sqrt{(r_b + r_e)(r_e + r_k - r_m) - r_e (r_e - r_m)},$$

$$p = \sqrt{\frac{r_e + r_k - r_m}{r_b + r_e}}.$$

Przyjmując założenia upraszczające jak w punkcie 2.2 otrzymujemy następujące wyrażenia przybliżone

$$Z = \sqrt{r_k\,[r_e + r_b\,(1-\alpha)]}, \tag{48}$$

$$p = \sqrt{\frac{r_k\,(1-\alpha)}{r_b + r_e}}, \tag{49}$$

Jest rzeczą interesującą, że wyrażenie otrzymane na średnią oporność falową Z dla układu o wspólnym emiterze jest takie same jak dla układu o wspólnej bazie.

Wyrażenia na oporność falową pierwotną i wtórną mają teraz postać:

$$Z_1 = \sqrt{(r_b + r_e)\left(r_b + \frac{r_e}{1-\alpha}\right)}, \tag{50}$$

$$Z_2 = r_k(1-\alpha)\sqrt{\frac{r_b + \dfrac{r_e}{1-\alpha}}{r_b + r_e}}. \tag{51}$$

Powyższe wyrażenia są oczywiście identyczne z odpowiednimi wyrażeniami na optymalne wartości oporności źródła i odbiornika, podobnie jak to ma miejsce dla układu o wspólnej bazie.

Wyznaczymy teraz pozostałe parametry falowe.

Postępując jak w punkcie poprzednim otrzymujemy:

$$\operatorname{tgh}\Gamma = \frac{Z_2}{r_k(1-\alpha)} = \sqrt{\frac{r_b + \dfrac{r_e}{1-\alpha}}{r_b + r_e}}. \tag{52}$$

Zakładając, że parametry własne r_e, r_b, r_k, α są rzeczywiste, stwierdzamy, że

$$\frac{r_b + \dfrac{r_e}{1-\alpha}}{r_b + r_e} > 1.$$

Korzystając ze wzoru (20a) otrzymujemy wtedy

$$A = \frac{1}{2}\operatorname{artgh}\frac{2T}{1+T^2}, \tag{53}$$

gdzie

$$T = \frac{Z_2}{r_k(1-\alpha)} = \sqrt{\frac{r_b + \dfrac{r_e}{1-\alpha}}{r_b + r_e}} \tag{54}$$

oraz

$$B = \frac{\pi}{2}. \tag{55}$$

Wzorów (21) nie można tu stosować, gdyż $T > 1$. Jeżeli wielkości r_e, r_b, r_k, α są zespolone, to A wyznaczamy z (20a), a B z (20b).

W dalszym ciągu otrzymujemy:

$$p_e = \sqrt{\frac{r_e}{r_e - r_m}} \cong \sqrt{\frac{r_e}{-r_m}} = j\sqrt{\frac{r_e}{\alpha r_k}}, \tag{56}$$

stąd

$$\ln p_e = \ln \sqrt{\frac{r_e}{\alpha r_k}} + j\frac{\pi}{2}. \tag{57}$$

Korzystając z wzorów (15) otrzymujemy następujące wyrażenia[1]):

$$S_1 = -A + \ln \sqrt{\frac{\alpha r_k}{r_e}}, \tag{58}$$

$$B_1 = \pi, \tag{59}$$

$$A_2 = A + \ln \sqrt{\frac{\alpha r_k}{r_e}}, \tag{60}$$

$$B_2 = 0, \tag{61}$$

gdzie A określone jest wzorem (53).

Z otrzymanych wzorów wynikają następujące wnioski:

1) wzmocność falowa pierwotna jest dodatnia, gdyż $\ln \sqrt{\frac{\alpha r_k}{r_e}} > A$, a więc tranzystor w układzie o wspólnym emiterze wprowadza wzmocnienie mocy przy przekazywaniu jej od wejścia do wyjścia.

2) tłumienność falowa wtórna tranzystora jest dodatnia, a więc tranzystor w układzie o wspólnym emiterze wprowadza tłumienie mocy przy jej przekazywaniu od wyjścia do wejścia.

3) przesunięcie fazowe wprowadzane przez tranzystor w układzie o wspólnym emiterze przy kierunku przenoszenia mocy od wejścia do wyjścia jest równe π, a nie 0 jak w układzie o wspólnej bazie.

Wyrażenie (58) stanowi oczywiście odpowiednik wyrażenia na k_{pomax}.

[1]) Przy założeniu rzeczywistych parametrów własnych.

Przykład 2 — 3

Przyjmując wartości parametrów własnych tranzystora jak w przykładzie (2—1) otrzymujemy następujące wartości parametrów falowych:

$$Z = \sqrt{r_b [r_e + r_k (1 - \alpha)]} = 6{,}72\ \mathrm{k}\Omega,$$

$$p = \sqrt{\frac{r_k (1 - \alpha)}{r_b + r_e}} = 8{,}75,$$

$$Z_1 = \frac{Z}{p} = 770\ \Omega,$$

$$Z_2 = p\, Z = 59\ \mathrm{k}\Omega.$$

Dla obliczenia A wyznaczamy najpierw T

$$T = \frac{Z_2}{r_k (1 - \alpha)} = 1{,}47,$$

stąd

$$A = \frac{1}{2} \operatorname{artgh} \frac{2T}{1 + T^2} = \frac{1}{2} \operatorname{artgh} 0{,}933 = 0{,}84\ \mathrm{N}.$$

Dalej wyznaczamy

$$p_e = j \sqrt{\frac{r_e}{\alpha r_k}} = j \frac{1}{196},$$

stąd

$$\ln |p_e| = -5{,}28\ \mathrm{N}.$$

Ostatecznie otrzymujemy następujące wzmocności, tłumienności i przesuwności:

$$S_1 = -0{,}84 + 5{,}28 = 4{,}44\ \mathrm{N} = 38{,}5\ \mathrm{dB},$$
$$A_2 = 0{,}84 + 5{,}28 = 6{,}12\ \mathrm{N} = 53{,}2\ \mathrm{dB},$$
$$B_1 = \pi,$$
$$B_2 = 0.$$

Porównując otrzymane wyniki z wynikami przykładu (2—1) można wysnuć następujące, znane zresztą wnioski:

1) oporność falowa pierwotna jest większa dla układu o wspólnym emiterze niż dla układu o wspólnej bazie,

2) oporność falowa wtórna jest większa dla układu o wspólnej bazie niż dla układu o wspólnym emiterze,

3) przekładnia oporowa jest dla układu o wspólnej bazie znacznie większa niż dla układu o wspólnym emiterze.

4) wzmocność falowa pierwotna jest większa dla układu o wspólnym emiterze niż dla układu o wspólnej bazie.

5) przesuwność falowa pierwotna wynosi 0 dla układu o wspólnej bazie, a π dla układu o wspólnym emiterze.

6) tłumienność falowa wtórna jest większa dla układu o wspólnym emiterze niż dla układu o wspólnej bazie. Oznacza to, że oddziaływanie wsteczne jest mniejsze w układzie o wspólnym emiterze niż w układzie o wspólnej bazie.

2.4. Parametry falowe tranzystora w układzie o wspólnym kolektorze (WK)

Ze względu na to, że metoda wyprowadzania wzorów na parametry falowe układu o wspólnym kolektorze jest taka sama jak w poprzednich dwóch przypadkach, ograniczymy się jedynie do podania końcowych rezultatów.

$$Z_1 = \sqrt{r_k\left(r_b + \frac{r_e}{1-\alpha}\right)}, \tag{62}$$

$$Z_2 = r_k(1-\alpha)\sqrt{\frac{r_b}{r_k} + \frac{r_e}{r_k(1-\alpha)}}, \tag{63}$$

$$S_1 = -\frac{1}{2}\ln(1-\alpha), \tag{64}$$

$$A_2 = -\frac{1}{2}\ln(1-\alpha). \tag{65}$$

Ze względu na prostotę wzoru (64) łatwo można się przekonać, że jest on równoważny wzorowi na $k_{p_{o\,max}}$ dla układu o wspólnym kolektorze.

Przykład 2—4

Zakładając parametry własne tranzystora jak w przykładzie 2—1 otrzymamy następujące parametry falowe:

$$Z_1 = 33{,}6\,\mathrm{k\Omega},$$
$$Z_2 = 1{,}34\,\mathrm{k\Omega},$$
$$S_1 = 1{,}61\,\mathrm{N} = 14\,\mathrm{dB},$$
$$A_2 = 1{,}61\,\mathrm{N} = 14\,\mathrm{dB}.$$

Nie będziemy tu przeprowadzali szczegółowego porównania wzmacniaczy tranzystorowych pracujących w różnych układach, gdyż zagadnienie to omawiane jest w teorii układów tranzystorowych. Zwrócimy jedynie uwagę na fakt, że dla układu o wspólnym kolektorze zachodzi nierówność $Z_1 > Z_2$, podczas gdy dla układów poprzednich było $Z_1 < Z_2$.

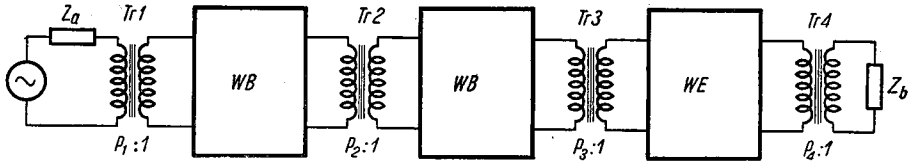
2.5. Łańcuch tranzystorów zbudowany na zasadzie dopasowania falowego

Z łańcuchem tranzystorów zbudowanym na zasadzie dopasowania falowego mamy do czynienia w przypadku wzmacniaczy wielostopniowych ze sprzężeniem transformatorowym. Stosowanie transformatorów międzystopniowych umożliwia zrealizowanie warunku dopasowania falowego między poszczególnymi tranzystorami.

Opierając się na rozważaniach z punktu 1.5 rozważmy następujący przykład.

Przykład 2—5

Rozpatrzymy trzystopniowy wzmacniacz tranzystorowy ze sprzężeniem transformatorowym (rys. 3) pracujący między opornościami $Z_a = Z_b = 600\Omega$. Pierw-



Rys. 3. Tranzystorowy wzmacniacz o sprzężeniu transformatorowym

sze dwa stopnie pracują w układzie o wspólnej bazie i mają następujące parametry falowe:

$$Z_1 = 154\,\Omega; \quad Z_2 = 293\,\mathrm{k}\,\Omega; \quad S_1 = 3{,}47\,\mathrm{N}; \quad A_2 = 4{,}09\,\mathrm{N}$$

Trzeci stopień pracuje w układzie o wspólnym emiterze i ma parametry falowe:

$$Z_1 = 770\,\Omega; \quad Z_2 = 59\,\mathrm{k}\,\Omega; \quad S_1 = 4{,}45\,\mathrm{N}, \quad A_2 = 6{,}11\,\mathrm{N}.$$

Należy wyznaczyć przekładnie transformatorów wynikające z warunków dopasowania oraz wyznaczyć parametry falowe wzmacniacza jako całości.

Obliczymy najpierw, jakie muszą być przekładnie transformatorów *Tr2* i *Tr3*, aby rozpatrywany łańcuch trzech tranzystorów był łańcuchem zbudowanym na zasadzie dopasowania falowego.

$$p_2 = \sqrt{\frac{293}{0{,}154}} = 43{,}7.$$

Oczywiście przekładnia transformatora *Tr2* równa jest przekładni oporowej tranzystora w układzie o wspólnej bazie.

$$p_3 = \sqrt{\frac{293}{0{,}770}} = 19{,}6.$$

Zgodnie z rozważaniami z punktu 1.5 rozpatrywany łańcuch może być zastąpiony czwórnikiem o następujących parametrach falowych:

$$Z_1 = 154\,\Omega,$$
$$Z_2 = 59\,\mathrm{k}\Omega,$$
$$S_1 = 3{,}47 + 3{,}47 + 4{,}45 = 11{,}39\,\mathrm{N} = 98{,}8\,\mathrm{dB},$$
$$A_2 = 4{,}09 + 4{,}09 + 6{,}11 = 14{,}29\,\mathrm{N} = 124\,\mathrm{dB}.$$

Transformatory *Tr1* i *Tr4* stosujemy dla uzyskania dopasowania falowego łańcucha do źródła i do odbiornika. Przekładnie tych transformatorów powinny wynosić:

$$p_1 = \sqrt{\frac{600}{154}} = 1{,}98,$$

$$p_4 = \sqrt{\frac{59}{0{,}6}} = 9{,}94.$$

Łańcuch uzyskany z poprzedniego przez uzupełnienie go transformatorami *Tr1* i *Tr4* może być zastąpiony czwórnikiem o następujących parametrach falowych:

$$Z_1 = Z_2 = 600\,\Omega,$$
$$S_1 = 11{,}39\,\mathrm{N},$$
$$A_2 = 14{,}29\,\mathrm{N}.$$

Ze względu na to, że $Z_1 = Z_2$, wzmocność falowa napięciowa S_{1u} i wzmocność falowa pierwotna prądowa S_{1i}[1]) równe są wzmocności falowej pierwotnej S_1.

W rozważaniach powyższych założono, że tłumienności transformatorów równe są zeru.

2.6. Znaczenie fizyczne parametrów falowych tranzystora

Przypomnimy obecnie znaczenie fizyczne parametrów falowych tranzystora. Oporności falowe Z_1 i Z_2 tranzystora tworzą parę oporności mających następujące właściwości:

a) jeżeli porność obciążenia $R_0 \equiv Z_b = Z_2$, to oporność wejściowa pierwotna $W_1 = Z_1$,

b) jeżeli oporność generatora $R_g \equiv Z_a = Z_1$, to oporność wejściowa wtórna $W_2 = Z_2$. Przy spełnieniu warunków a) i b) istnieją najlepsze warunki przekazywania mocy przez tranzystor[2]).

Oporności Z_1 i Z_2 są identyczne z wprowadzonymi w inny sposób opornościami $R_{g\mathrm{opt.}}$ i $R_{o\mathrm{opt.}}$

Wzmocność falowa pierwotna S_1 równa jest połowie logarytmu naturalnego stosunku mocy po stronie wtórnej do mocy po stronie pierwotnej przy spełnieniu warunku dopasowania falowego po stronie wtórnej

[1]) Zob. 1.3.

[2]) W założeniu, że parametry własne tranzystora są rzeczywiste.

i przy kierunku przekazywania mocy od strony pierwotnej do strony wtórnej. Wielkość S_1 jest więc maksymalnym wzmocnieniem mocy wyrażonym w mierze logarytmicznej [N]. W przypadku obustronnego dopasowania falowego wzmocność falowa S_1 jest równa wzmocności skutecznej S_{s1} i charakteryzuje przekazywanie mocy od źródła do odbiornika w warunkach optymalnych.

Tłumienność falowa wtórna A_2 równa jest połowie logarytmu naturalnego stosunku mocy po stronie wtórnej do mocy po stronie pierwotnej przy spełnieniu warunku dopasowania falowego po stronie pierwotnej i przy kierunku przekazywania mocy od strony wtórnej do strony pierwotnej. Wielkość A_2 charakteryzuje więc efekt oddziaływania wstecznego w tranzystorze.

Reasumując: zespół parametrów falowych Z_1, Z_2, S_1, A_2, B_1, B_2 nie tylko charakteryzuje sam tranzystor, jak zespół parametrów oporowych, ale również daje bezpośrednie informacje o pracy tranzystora jako wzmacniacza w pewnych określonych warunkach.

3. PRZYKŁADY WYZNACZANIA PARAMETRÓW ROBOCZYCH TRANZYSTORA

3.1. Zagadnienia ogólne

Zajmiemy się obecnie wyznaczaniem oporności wejściowych i wzmocności skutecznej. Nie będziemy wyprowadzać wzorów ogólnych, które byłyby nieprzejrzyste i mało przydatne, a ograniczymy się jedynie do rozważania kilku przykładów liczbowych.

3.2. Wyznaczanie oporności wejściowych

Odpowiednie wzory ogólne zaczerpnięte z teorii czwórników zostały podane w punkcie 1.6. Obecnie rozpatrzymy przykład liczbowy.

Przykład 3 — 1

Należy obliczyć oporność wejściową tranzystora w układzie o wspólnej bazie, dla stanu zwarcia, jeżeli parametry falowe tego tranzystora wynoszą: $Z_1 = 154\ \Omega$, $Z_2 = 293\ \mathrm{k}\Omega$, $S_1 = 3{,}46$ N, $A_2 = 4{,}1$ N, $B_1 = B_2 = 0$.

Obliczymy najpierw średnią tamowność falową określoną wzorem:

$$\Gamma = \frac{\Gamma_1 + \Gamma_2}{2} = \frac{-S_1 + A_2 + j(B_1 + B_2)}{2}. \tag{66}$$

W rozpatrywanym przypadku mamy:

$$\Gamma = \frac{-3{,}46 + 4{,}1}{2} = 0{,}32\ \mathrm{N}.$$

Następnie korzystając z (23b) i biorąc pod uwagę, że $Z_b = 0$ otrzymujemy:

$$q_2 = \frac{Z_b - Z_2}{Z_b + Z_2} = -1.$$

W dalszym ciągu ze wzoru (22a) otrzymujemy:

$$W_1 = Z_1 \frac{1 + q_2 e^{-2\Gamma}}{1 - q_2 e^{-2\Gamma}} = 154 \frac{1 - 0{,}57}{1 + 0{,}57} = 42\,\Omega,$$

Ze stosowanego wzoru (22a) wynika, że oporność wejściowa pierwotna jest tym bardziej zbliżona do oporności falowej pierwotnej, im mniejszy współczynnik odbicia q_2 i im większa jest tamowność średnia. Łatwo zauważyć, że tamowność średnia jest tym większa, im tłumienność falowa wtórna A_2 jest większa od wzmocności falowej pierwotnej S_1.

3.3. Wyznaczanie wzmocności skutecznej

Odpowiednie wzory ogólne zaczerpnięte z teorii czwórników zostały podane w punkcie 1.7. Obecnie rozpatrzymy przykład liczbowy.

Przykład 3 — 2

Jednostopniowy wzmacniacz tranzystorowy w układzie o wspólnym emiterze mający parametry falowe: $Z_1 = 770\,\Omega$; $Z_2 = 59\,\text{k}\Omega$; $S_1 = 4{,}44$ N; $A_2 = 6{,}12$ N pracuje między opornościami $Z_a = 1\,\text{k}\Omega$, $Z_b = 10\,\text{k}\Omega$. Należy wyznaczyć wzmocność skuteczną w tych warunkach pracy.

Korzystamy tu z wzoru (29b).

$$S_{s1} = S_1 - \Delta_1 - \Delta_2 - \Delta_{12}. \tag{29b}$$

Obliczając poszczególne składniki otrzymujemy:

$$\Delta_1 = \ln \left| \frac{Z_a + Z_1}{2\sqrt{Z_a Z_1}} \right| = 0{,}01\,\text{N},$$

$$\Delta_2 = \ln \left| \frac{Z_b + Z_2}{2\sqrt{Z_b Z_2}} \right| = 0{,}36\,\text{N},$$

$$\Delta_{12} = \ln \left| 1 - q_1 q_2 e^{-2\Gamma} \right| = 0{,}02\,\text{N}.$$

W ostatnim wzorze symbol Γ oznacza tamowność średnią określoną wzorem (66). Ostatecznie $S_{s1} = 4{,}05$ N.

Wzmocność skuteczna jest więc w rozważanym przypadku mniejsza od falowej o ok. 0,4 N. Spowodowane jest to przede wszystkim znacznym niedopasowaniem na wyjściu. Wpływ niedopasowania na wejściu i oddziaływanie wyjścia na wejście jest bardzo mały.

Powstaje pytanie, kiedy przy obliczaniu wzmocności skutecznej możemy pomijać poszczególne wyrazy we wzorze (29b).

Rozważając wyrażenie $\Delta = \ln \frac{A+B}{2\sqrt{AB}}$, gdzie A i B są wielkościami rzeczywistymi, otrzymamy następujące wyniki[1]).

$$\Delta \leqslant 0{,}1\ \mathrm{N} \text{ dla } \frac{A}{B} \leqslant 2{,}5,$$

$$\Delta \leqslant 0{,}2\ \mathrm{N} \text{ dla } \frac{A}{B} \leqslant 3{,}75,$$

$$\Delta \leqslant 0{,}3\ \mathrm{N} \text{ dla } \frac{A}{B} \leqslant 5{,}0,$$

Rozważając wyrażenie $\Delta' = \ln|1 - A|$, gdzie A jest wielkością rzeczywistą otrzymamy[1]):

$$-0{,}1\,\mathrm{N} \leqslant \Delta' \leqslant +0{,}1\,\mathrm{N} \text{ dla } -0{,}1 \leqslant A \leqslant 0{,}1.$$

Z powyższych danych widać, że w praktyce można często pominąć wpływy dość znacznych niedopasowań.

3.4. Wyznaczanie wzmocności skutecznej łańcucha

Odpowiednie wzory ogólne zaczerpnięte z teorii czwórników zostały podane w punkcie 1.8. Obecnie rozpatrzymy przykład liczbowy.

Przykład 3-3

Wzmacniacz tranzystorowy składa się z dwóch stopni w układzie o wspólnym emiterze, sprzęgniętych ze sobą układem RC. Parametry falowe tranzystora w układzie o wspólnym emiterze — jak w przykładzie 3-2. Należy wyznaczyć wzmocność skuteczną wzmacniacza przy założeniu, że źródło dopasowane jest do pierwszego tranzystora, a odbiornik do drugiego, tzn.

$$Z_a = Z_1' \quad \text{oraz} \quad Z_b = Z_2''$$

Korzystamy ze wzoru (30). Ze względu na dopasowanie na wejściu i wyjściu wzór upraszcza się do postaci:

$$S_{s1} = S_{s1}' + S_1'', \tag{67}$$

gdzie S'_{s1} oznacza wzmocność skuteczną pierwszego stopnia, a S''_1 wzmocność falową drugiego stopnia. Ponadto oporność wejściowa pierwotna drugiego stopnia równa jest (ze względu na dopasowanie) oporności falowej pierwotnej: $W''_1 = Z''_1$: Biorąc pod uwagę, że $Z'_1 = Z''_1 = Z_1$, $Z'_2 = Z''_2 = Z_2$ oraz $S'_1 = S''_1 = S_1$, możemy napisać ostatecznie:

[1]) Na podstawie nomogramów ze str. 179 pracy [1].

$$S_{s1} = S_1 - \ln\left|\frac{Z_1 + Z_2}{2\sqrt{Z_1 Z_2}}\right| + S_1 = 2\,S_1 - \ln\left|\frac{Z_1 + Z_2}{2\sqrt{Z_1 Z_2}}\right| =$$

$$= 2 \cdot 4{,}44 - \ln\frac{60 + 0{,}8}{2\sqrt{60 \cdot 0{,}8}} = 8{,}88 - 1{,}48 = 7{,}4\,\mathrm{N}.$$

4. WNIOSKI KOŃCOWE

Parametry falowe tranzystora charakteryzują całkowicie jego własności, podobnie zresztą jak np. parametry oporowe. Mają one jednak pewną wyższość nad parametrami oporowymi polegającą na tym, że określają one jednocześnie optymalne warunki pracy (oporności zamknięcia) i podają, jaka wzmocność może być w tych warunkach z tranzystora uzyskana. Zaletą stosowania parametrów falowych jest również łatwość obliczania wzmacniaczy ze sprzężeniem transformatorowym.

Jeżeli tranzystor z pewnych względów nie pracuje w warunkach optymalnych, to wówczas parametry falowe są niewystarczające i trzeba obliczać parametry robocze. Należy jednak zwrócić uwagę na to, że przy niewielkich odstępstwach od warunków optymalnych można posługiwać się nadal parametrami falowymi.

Stosowanie parametrów falowych i roboczych tranzystora umożliwia również stosunkowo proste obliczanie parametrów roboczych dowolnego łańcucha.

Wyznaczanie parametrów falowych i parametrów roboczych tranzystora wymaga rachunków skomplikowanych tylko pozornie. W rzeczywistości obliczenia nie są bardziej skomplikowane niż obliczenia analogicznych wielkości przy zastosowaniu zwykle stosowanej techniki rachunkowej. Korzystne jest przy tym to, że wzmocność tranzystora przy stosowaniu parametrów falowych jest wyrażona od razu w mierze logarytmicznej.

Autor składa serdeczne podziękowanie prof. drowi W. Nowickiemu za cenne uwagi i pomoc przy opracowywaniu niniejszej pracy.

Politechnika Warszawska
Katedra Teletransmisji Przewodowej

WYKAZ LITERATURY

1. Nowicki W.: Zasady teletransmisji przewodowej, t. 1, PWN, Warszawa 1953.
2. Shea R. F.: Principles of Transistor Circuits.
   D. Van Nostrand Co., New York 1953.
3. Coblenz A., Owens H.: Transistors — Theory and Applications. 1955.
4. Rosenberg W.: Eingenschaften und Anwendung des Flächentransistors. Nachrinchtentechnik Nr 11, 1956.

В. МАЕВСКИ (юнр)

СОБСТВЕННЫЕ И РАБОЧИЕ ПАРАМЕТРЫ ПОЛУПРОВОДНИКОВОГО ТРИОДА

Резюме

Статья является попыткой применения классической теории четырёхполюсников к транзисторным схемам. Приведены в ней основные формулы из теории четырёхполюсников; эти формулы применяются далее к транзисторным схемам. Выведены формулы на собственные сопротивления и постоянные передачи полупроводникового триода для трёх основных схем соединения. Приведены численные примеры касающиеся одиночных полупроводниковых триодов и усилителя с трансформаторной связью.

В дальнейшем приведены примеры расчёта рабочих параметров полупроводниковых триодов и их цепей.

Собственные параметры полупроводникового триода характеризуют полностью его свойства подобным образом как напр. резистивные параметры. Имеют они всё-таки некоторое превосходство над резистивными параметрами состоящее в том, что они определяют одновремённо оптимальные условия работы (сопротивление) источника и сопротивление нагрузки и сообщают, какое усиление можно в этих условиях получить с полупроводникового триода. Достоинством применения собственных параметров является также облегчение расчёта усилителей с трансформаторной связью.

Если полупроводниковый триод по каким-нибудь причинам не работает в оптимальных условиях, то, в таком случае, собственные параметры недостаточны и следует рассчитывать рабочие параметры. Следует всё-таки обратить внимание на то, что при небольших отклонениях от оптимальных условий можно и в дальнейшем применять собственные параметры.

W. MAJEWSKI

IMAGE AND WORKING PARAMETERS OF THE TRANSISTOR

Summary

The present paper constitutes a trial of applying the classical theory of four-poles to transistor circuits. Basic formulae of the theory of four-poles are presented and then applied to transistor circuits. Formulae for image impedances and image transfer coefficients in three basic circuits are deduced in the paper. The numerical examples presented concern single transistors and an amplifier with the transformer coupling.

Next, examples of the calculation of working parameters of transistors and their chains are presented.

Image parameters of the transistor characterize completely its properties, likewise e.g. the impedance parameters. The former, however, have a certain superiority over the latter consisting in the fact that at the same time they determine the optimum working conditions (terminating impedances) and inform what gain can in such conditions be obtained from the transistor. An advantage of the application of image parameters is also the facility of calculating amplifiers with the transformer coupling.

If a transistor does not work in optimum conditions, then image parameters do not suffice and it is necessary to calculate the working parameters. It should be mentioned, however, that in the case of small deviations from the optimum conditions, it is still possible to use the image parameters.

621.3.08:621.3.015.33:621.315.63

JANUSZ LECHOSŁAW MAKSIEJEWSKI

Obliczanie napięcia wierzchołka słupa trafionego przez piorun, metodami stosowanymi dla linii długich

Rękopis dostarczono 3.2.1960

W pracy poddano analizie obliczenia napięcia wierzchołka słupa trafionego przez piorun metodami stosowanymi dla linii długich z punktu widzenia przydatności tych metod do obliczeń praktycznych.

1. WSTĘP

Ostatnio ukazały się publikacje [1], [2] rozpatrujące stosowane w praktyce sposoby obliczania napięcia wierzchołka słupa linii energetycznej trafionego przez piorun. W pracy niniejszej zagadnienie to rozważymy przy pomocy metod ogólnych stosowanych dla linii długich w celu zbadania wartości praktycznej tych metod w danym przypadku.

Stan nieustalony w obwodzie o stałych rozłożonych możemy rozpatrzyć dwoma metodami klasycznymi, a mianowicie metodą d'Alemberta oraz metodą Fouriera. Zajmiemy się kolejno obydwoma metodami.

2. METODA d'ALEMBERTA

Początkowo bierzemy pod uwagę układ obliczeniowy jak na rys. 1, a więc przyjmujemy, że przęsła przewodu odgromowego są nieskończenie długie. Zakładamy, że słup stanowi linię bez strat. Wychodzimy z równań napięcia i prądu w słupie:

$$U_s(p) = A_1 e^{-\gamma x} + A_2 e^{\gamma x}, \tag{1}$$

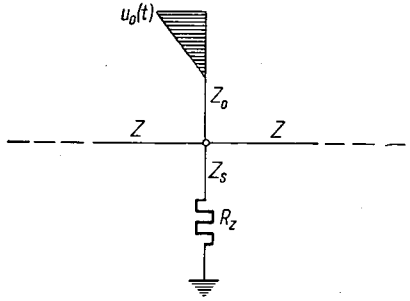
$$I_s(p) = \frac{1}{Z_s}\left[A_1 e^{-\gamma x} - A_2 e^{\gamma x}\right], \tag{2}$$

$$\gamma = \frac{p}{v_1}, \tag{3}$$

gdzie:

γ — współczynnik przenoszenia,
v_1 — szybkość rozchodzenia się fal,
Z_s — oporność falowa słupa.

Stałe całkowania A_1 i A_2 wyznaczamy z warunków brzegowych. Na podstawie zasady Thevenina falę napięciową w kanale pioruna u_o trafiającą



Rys. 1. Układ obliczeniowy przy założeniu, że przęsła przewodu odgromowego są nieskończenie długie

w punkt węzłowy możemy rozważać jako włączenie źródła napięcia $2u_o$, o oporności wewnętrznej równej oporności falowej kanału pioruna Z_o. Dla wierzchołka słupa ($x = 0$) mamy więc:

$$U_s(p) = 2\,U_o(p) - [I_s(p) + 2\,I_1(p)]\,Z_o, \tag{4}$$

gdzie:

$I_1(p) = \dfrac{U_s(p)}{Z}$ — prąd w przewodzie odgromowym,

Z — oporność falowa pojedynczego przewodu odgromowego.

Na końcu słupa o wysokości h, tj. dla $x = h$

$$U_s(p) = I_s(p)\,R_z, \tag{5}$$

gdzie R_z — oporność uziemienia słupa.

Z podanych warunków brzegowych (4) i (5) otrzymujemy:

$$A_1 = \frac{U_o(p)\,Z_s(R_z + Z_s)\,e^{\gamma h}}{a_1\,e^{\gamma h} - b_1\,e^{-\gamma h}},$$

$$A_2 = \frac{U_o(p)\,Z_s(R_z - Z_s)\,e^{-\gamma h}}{a_1\,e^{\gamma h} - b_1\,e^{-\gamma h}}.$$

gdzie:

$$a_1 = \frac{[Z Z_o + Z_s (Z + 2 Z_o)][R_z + Z_s]}{2 Z},$$

$$b_1 = \frac{[Z Z_o - Z_s (Z + 2 Z_o)][R_z - Z_s]}{2 Z}.$$

Podstawiając powyższe wyrażenia na stałe do (1) i (2) otrzymamy dla transformat napięcia i prądu równania:

$$U_s(p) = U_o(p) \frac{Z_s [(R_z + Z_s) e^{\gamma (h-x)} + (R_z - Z_s) e^{-\gamma (h-x)}]}{a_1 e^{\gamma h} - b_1 e^{-\gamma h}}, \tag{6}$$

$$I_s(p) = U_o(p) \frac{[(R_z + Z_s) e^{\gamma (h-x)} - (R_z - Z_s) e^{-\gamma (h-x)}]}{a_1 e^{\gamma h} - b_1 e^{-\gamma h}}. \tag{7}$$

Wyrażenia na $U_s(p)$ i $I_s(p)$ możemy przepisać w następujący sposób:

$$U_s(p) = U_o(p) \frac{Z_s [(R_z + Z_s) e^{p(\tau_o - \tau)} + (R_z - Z_s) e^{-p(\tau_o - \tau)}]}{a_1 e^{p\tau_o} - b_1 e^{-p\tau_o}}, \tag{8}$$

$$I_s(p) = U_o(p) \frac{[(R_z + Z_s) e^{p(\tau_o - \tau)} - (R_z - Z_s) e^{-p(\tau_o - \tau)}]}{a_1 e^{p\tau_o} - b_1 e^{-p\tau_o}}, \tag{9}$$

gdzie:

$\tau_o = \frac{h}{v_1}$ — czas rozchodzenia się fali o szybkości v_1 wzdłuż całego słupa,

$\tau = \frac{x}{v_1}$ — czas rozchodzenia się fali o szybkości v_1 wzdłuż odcinka słupa o długości x.

Czasy te wprowadzamy w celu ułatwienia interpretacji otrzymanych poniżej wyników.

W dalszym ciągu wszystkie obliczenia będziemy przeprowadzać dla napięcia.

Dzieląc licznik i mianownik (8) przez $e^{p\tau_o}$ otrzymamy

$$U_s(p) = \frac{U_o(p) Z_s [(R_z + Z_s) e^{-p\tau} + (R_z - Z_s) e^{-p(2\tau_o - \tau)}]}{a_1 \left(1 - \frac{b_1}{a_1} e^{-2p\tau_o}\right)}. \tag{10}$$

Ponieważ $Re\ p > 0$ [3], to moduł $e^{-2p\tau_o}$ jest mniejszy od jedności. Następnie mamy $\left|\frac{b_1}{a_1}\right| < 1$, a więc $\frac{1}{1 - \frac{b_1}{a_1} e^{-2p\tau_o}}$ można rozpatrywać jako

sumę szeregu geometrycznego. Wzór (10) można zatem przekształcić następująco:

$$U_s(p) = U_0(p)\,\alpha\left[e^{-p\tau} + \beta_2 e^{-p(2\tau_0-\tau)}\right]\left[1 + \beta e^{-2p\tau_0} + \right.$$
$$\left. + \beta^2 e^{-4p\tau_0} + \beta^3 e^{-6p\tau_0} + \ldots\right], \tag{11}$$

przy czym:

$$\alpha = \frac{Z_s(R_z + Z_s)}{a_1} = \frac{2\,Z\,Z_s}{Z\,Z_0 + Z_s(Z + 2\,Z_0)}, \tag{12}$$
$$\beta = \frac{b_1}{a_1} = \frac{[Z\,Z_0 - Z_s(Z + 2\,Z_0)](R_z - Z_s)}{[Z\,Z_0 + Z_s(Z + 2\,Z_0)](R_z + Z_s)}.$$

O ile wprowadzimy oznaczenia

$$\beta_1 = \frac{Z\,Z_0 - Z_s(Z + 2\,Z_0)}{Z\,Z_0 + Z_s(Z + 2\,Z_0)}, \tag{13}$$

$$\beta_2 = \frac{R_z - Z_s}{R_z + Z_s}, \tag{14}$$

mamy

$$\beta = \beta_1\,\beta_2$$

oraz

$$U_s(p) = U_0(p)\,\alpha\left[e^{-p\tau} + \beta_2 e^{-p(2\tau_0-\tau)} + \beta_1\beta_2 e^{-p(2\tau_0+\tau)} + \right.$$
$$\left. + \beta_1\beta_2^2 e^{-p(4\tau_0-\tau)} + \beta_1^2\beta_2^2 e^{-p(4\tau_0+\tau)} + \ldots\right]. \tag{15}$$

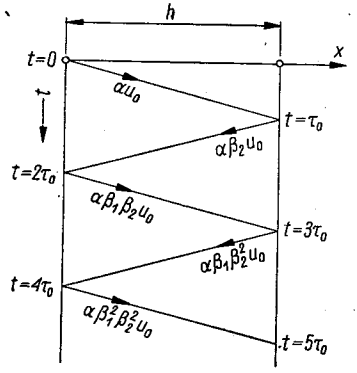
Korzystając z twierdzenia o przesunięciu, możemy napisać wyrażenie dla funkcji czasowej

$$u_s(t) = \alpha\{u_0(t-\tau)\,\mathbf{1}_\tau + \beta_2 u_0[t-(2\tau_0-\tau)]\,\mathbf{1}_{2\tau_0-\tau} +$$
$$+ \beta_1\beta_2 u_0[t-(2\tau_0+\tau)]\,\mathbf{1}_{2\tau_0+\tau} + \beta_1\beta_2^2 u_0[t-(4\tau_0-\tau)]\,\mathbf{1}_{4\tau_0-\tau} +$$
$$+ \beta_1^2\beta_2^2 u_0[t-(4\tau_0+\tau)]\,\mathbf{1}_{4\tau_0+\tau} + \ldots\}, \tag{16}$$

przy czym $\mathbf{1}_{t-t_1}$ oznacza funkcję jednostkową, która jest równa zeru dla $t<t_1$ i równa jedności dla $t>t_1$.

Otrzymujemy rozwiązanie w postaci sumy fal wędrownych. Pierwsza fala napięciowa pojawia się w rozpatrywanym punkcie w chwili $\tau = \frac{x}{v_1}$ licząc od momentu uderzenia pioruna. Druga fala przybywa po czasie $2(\tau_0-\tau)$ od chwili przyjścia pierwszej fali, tj. po czasie potrzebnym dla przebiegnięcia dwukrotnie odcinka $h-x$, stanowi więc falę odbitą od końca słupa. Każda następna fala powstaje w wyniku odbicia się od początku względnie od końca słupa. A więc β_1 stanowi współczynnik odbicia na początku słupa, β_2 — współczynnik odbicia na końcu słupa, natomiast α — współczynnik przejścia. Stąd wynika metoda praktyczna,

w której rozpatruje się przebiegi fal operując wyłącznie współczynnikami odbicia i przejścia bez zestawiania odpowiednich równań napięcia i prądu. W sposób przejrzysty przebiegi te przedstawia rozkład jazdy fal



Rys. 2. Rozkład jazdy fal według Bewleya dla przypadku z rys. 1

według Bewleya podany na rys. 2. Metoda d'Alemberta zwana jest również metodą fal wędrownych.

Dla wierzchołka słupa $\tau = 0$, a więc

$$u_s(t) = \alpha[u_0(t) + \beta_2(1+\beta_1)u_0(t-2\tau_0)\mathbf{1}_{2\tau_0} + \beta_1\beta_2^2(1+\beta_1)u_0(t-4\tau_0)\mathbf{1}_{4\tau_0} + \ldots]. \tag{17}$$

Analogiczne rozważania można przeprowadzić dla prądu.

O ile przyjąć skończoną długość sąsiednich przęseł, należy również rozważyć przebiegi w przewodach odgromowych. W tym celu oprócz równań dla napięcia i prądu w słupie (1) i (2) należy wziąć pod uwagę analogiczne równania napięcia i prądu w przewodzie

$$U_1(p) = A_3 e^{-\gamma y} + A_4 e^{\gamma y}, \tag{18}$$

$$I_1(p) = \frac{1}{Z}[A_3 e^{-\gamma y} - A_4 e^{\gamma y}]. \tag{19}$$

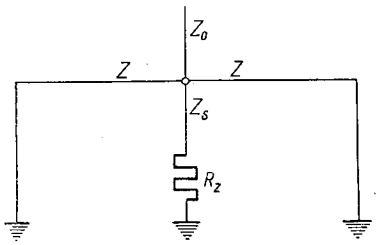
Stałe całkowania A_1, A_2, A_3 i A_4 wyznaczamy na podstawie warunków brzegowych. Dla wierzchołka słupa trafionego przez piorun ($x=0$, $y=0$)

$$U_1(p) = U_s(p). \tag{20}$$

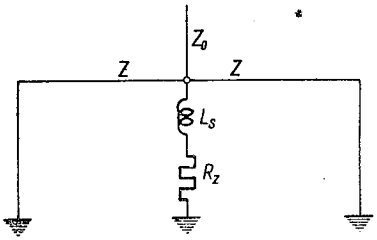
$U_s(p)$ dla wierzchołka jak poprzednio musi spełniać równanie (4), natomiast dla końca słupa ($x=h$) warunek (5). Podobnie jak w pracy [1] przyjmujemy, że sąsiednie słupy stanowią wprost zwarcie z ziemią (rys. 3), a więc dla $y = l$ (l — długość przęsła)

$$U_1(p) = 0. \tag{21}$$

Równania (4), (5), (20), (21) pozwalają na wyznaczenie stałych, jednakże odpowiednie wyrażenia są złożone, co uniemożliwia rozwiązanie w postaci przejrzystej i dogodnej do obliczeń. W sposób stosunkowo prosty



Rys. 3. Układ obliczeniowy przy uwzględnieniu skończonej długości przęseł i przy przyjęciu, że sąsiednie słupy stanowią zwarcie z ziemią



Rys. 4. Układ obliczeniowy przy zastąpieniu słupa trafionego przez indukcyjność skupioną i przy przyjęciu, że sąsiednie słupy stanowią zwarcie z ziemią

można wówczas wyznaczyć przebiegi wspomnianą wyżej metodą praktyczną, w której posługujemy się wyłącznie współczynnikami przejścia oraz odbicia [1].

Rozpatrzymy jeszcze uproszczenie obliczeń układu o skończonej długości sąsiednich przęseł dyskutowane w pracy [1] i polegające na zastąpieniu słupa trafionego przez indukcyjność skupioną L_s. Układ obliczeniowy dla tego przypadku przedstawia rys. 4. Należy wówczas wziąć pod uwagę tylko równania napięcia i prądu w przewodzie (18) i (19). Dla wierzchołka słupa ($y=0$) mamy warunek

$$U_1(p) = 2\,U_0(p) - [I_s(p) + 2\,I_1(p)]\,Z_0, \tag{22}$$

przy czym prąd w słupie w tym przypadku

$$I_s(p) = \frac{U_1(p)}{p\,L_s + R_z}. \tag{23}$$

Natomiast dla $y = l$ musi być spełniony warunek (21). Warunki te pozwalają na wyznaczenie stałych A_3 i A_4. Ostatecznie otrzymamy

$$U_1(p) = \frac{2\,U_0(p)\,Z\,(p\,L_s + R_z)\,[e^{\gamma(l-y)} - e^{-\gamma(l-y)}]}{[(Z + 2\,Z_0)(p\,L_s + R_z) + Z\,Z_0]\,e^{\gamma l} - [(Z - 2\,Z_0)(p\,L_s + R_z) + Z\,Z_0]\,e^{-\gamma l}}. \tag{24}$$

Wprowadzając $\tau_0' = \dfrac{l}{v_1}$ — czas rozchodzenia się fali wzdłuż całego przęsła i $\tau' = \dfrac{y}{v_1}$ — czas rozchodzenia się fali wzdłuż odcinka przęsła o długości y oraz stosując takie same jak poprzednio przekształcenia mamy:

$$U_1(p) = U_0(p)\,\alpha'\,[e^{-p\tau'} - e^{-p(2\tau_0' - \tau')}]\,[1 + \beta'\,e^{-2p\tau_0'} + \beta'^2\,e^{-4p\tau_0'} + \ldots]. \tag{25}$$

Otrzymujemy w ten sposób wzór o identycznej postaci jak (11), przy czym

$$\alpha' = \frac{2Z(pL_s + R_z)}{(Z + 2Z_o)(pL_s + R_z) + ZZ_o}, \tag{26}$$

$$\beta' = \beta'_1 \beta'_2,$$

$$\beta'_1 = -\frac{(Z - 2Z_o)(pL_s + R_z) + ZZ_o}{(Z + 2Z_o)(pL_s + R_z) + ZZ_o}; \qquad \beta'_2 = -1. \tag{27}$$

Zatem

$$U_1(p) = U_o(p)\,\alpha' [e^{-p\tau'} - e^{-p(2\tau'_o - \tau')} - \beta'_1 e^{-p(2\tau'_o + \tau')} + \\ + \beta'_1 e^{-p(4\tau'_o - \tau')} + \beta'^2_1 e^{-p(4\tau'_o + \tau')} \ldots]. \tag{28}$$

Dla wierzchołka słupa $\tau' = 0$ i wówczas

$$U_1(p) = U_o(p)\,\alpha' [1 - (1 + \beta'_1) e^{-p2\tau'_o} + \beta'_1 (1 + \beta'_1) e^{-p4\tau'_o} \ldots]. \tag{29}$$

Współczynniki α' oraz β'_1 stanowią funkcje operatorowe. Należy teraz wyznaczyć funkcje czasowe odpowiadające poszczególnym wyrazom wyrażenia (29).

Poszukujemy rozwiązania dla zadanego przebiegu fali pioruna. Przyjmujemy, że napięcie pioruna jest określone funkcją

$$u_o(t) = At,$$

gdzie A jest stałą. Funkcja operatorowa odpowiadająca tej funkcji czasowej

$$At \mathrel{\widehat{=}} \frac{A}{p}.$$

Pisząc pierwszy wyraz wyrażenia (29) w postaci rozwiniętej otrzymamy

$$U_o(p)\,\alpha' = \frac{2AZ(pL_s + R_z)}{p[(pL_s + R_z)(Z + 2Z_o) + ZZ_o]} = \frac{M(p)}{pN(p)}. \tag{30}$$

Dla funkcji operatorowej typu $F(p) = \dfrac{M(p)}{N(p)}$ możemy zastosować twierdzenie o rozkładzie

$$F(p) \mathrel{\widehat{=}} F(t) = \frac{M(0)}{N(0)} + \sum_{k=1}^{n} \frac{M(p_k)}{p_k N'(p_k)} e^{p_k t} \tag{31}$$

a następnie możemy wyznaczyć funkcję czasową odpowiadającą (30)

$$u_{w_1} = \frac{1}{p} F(t) = \int_o^t F(t)\,dt.$$

W danym przypadku równanie $N(p) = 0$ ma tylko jeden pierwiastek

$$p_1 = -\frac{(Z + 2Z_o)R_z + ZZ_o}{(Z + 2Z_o)L_s}. \tag{32}$$

Dalej obliczamy pochodną $N(p)$ względem p

$$N'(p) = (Z + 2Z_o)L_s.$$

Oznaczając przez B_1 wartość funkcji dla $p = 0$

$$B_1 = \frac{M(0)}{N(0)} = \frac{2AZR_z}{(Z + 2Z_o)R_z + ZZ_o}$$

oraz wprowadzając oznaczenie

$$B_2 = \frac{M(p_1)}{p_1 N'(p_1)} = \frac{2AZ^2Z_o}{(Z + 2Z_o)[(Z + 2Z_o)R_z + ZZ_o]}$$

otrzymujemy

$$F(t) = B_1 + B_2 e^{p_1 t} \tag{33}$$

i ostatecznie szukana funkcja czasowa

$$u_{w1} = B_1 t - \frac{B_2}{p_1}(1 - e^{p_1 t}). \tag{34}$$

Drugi wyraz wyrażenia (29) ma postać

$$-U_o(p)\alpha'(1 + \beta_1') = -\frac{8AZZ_o(pL_s + R_z)^2}{p[(Z + 2Z_o)(pL_s + R_z) + ZZ_o]^2} = \frac{M(p)}{pN(p)}. \tag{35}$$

Obecnie należy skorzystać ze wzoru dla przypadku, gdy równanie $N(p) = 0$ posiada jeden pierwiastek p_1 powtarzający się n razy

$$F(t) = \frac{M(0)}{N(0)} + \sum_{k=1}^{n} \frac{H_{k_n}}{(n-k)!} t^{n-k} e^{p_1 t}, \tag{36}$$

przy czym

$$H_{k_n} = \frac{1}{(k-1)!}\left\{\frac{d^{k-1}}{dp^{k-1}}\left[\frac{(p-p_1)^n}{p}\frac{M(p)}{N(p)}\right]\right\}_{p=p_1}. \tag{37}$$

W danym przypadku $n = 2$. Obliczamy wyrażenia pomocnicze

$$H_{1_2} = -\frac{8AZZ_o(p_1L_s + R_z)^2}{p_1L_s^2(Z + 2Z_o)^2},$$

$$H_{2_2} = -\frac{8AZZ_o(p_1^2L_s^2 - R_z^2)}{p_1^2L_s^2(Z + 2Z_o)^2}.$$

Wprowadzając oznaczenie

$$H_{0_2} = \frac{M(0)}{N(0)} = -\frac{8\,A\,Z\,Z_o\,R_z^2}{[(Z+2\,Z_o)R_z + Z\,Z_o]^2}$$

mamy wówczas

$$F(t) = H_{0_2} + H_{1_2}\,t\,e^{p_1 t} + H_{2_2}\,e^{p_1 t}, \tag{38}$$

a dalej całkując otrzymujemy wyrażenie dla funkcji czasowej odpowiadającej drugiemu wyrazowi

$$u_{w_2} = H_{0_2}\,t + \frac{H_{1_2}}{p_1}\,t\,e^{p_1 t} + \left(\frac{H_{1_2}}{p_1^2} - \frac{H_{2_2}}{p_1}\right)(1 - e^{p_1 t}). \tag{39}$$

Podobnie wyznaczamy funkcje czasowe odpowiadające dalszym wyrazom. Trzeciemu wyrazowi odpowiada funkcja

$$u_{w_3} = H_{0_3}\,t - \left[\frac{H_{1_3}}{p_1^3} - \frac{H_{2_3}}{p_1^2} + \frac{H_{3_3}}{p_1}\right](1 - e^{p_1 t}) +$$

$$- \left[\frac{H_{1_3}}{p_1^2} - \frac{H_{2_3}}{p_1}\right] t\,e^{p_1 t} + \frac{H_{1_3}}{2\,p_1}\,t^2\,e^{p_1 t},$$

gdzie:

$$H_{0_3} = \frac{K}{p_1^2}\,R_z^2[(Z - 2\,Z_o)R_z + Z\,Z_o],$$

$$H_{1_3} = -K(p_1 L_s + R_z)^2\,[(Z - 2\,Z_o)(p_1 L_s + R_z)\,Z\,Z_o],$$

$$H_{2_3} = -\frac{K}{p_1}(p_1 L_s + R_z)\{2\,p_1^2 L_s^2(Z - 2\,Z_o) + (p_1 L_s - R_z)[(Z - 2\,Z_o)R_z + Z\,Z_o]\},$$

$$H_{3_3} = -\frac{K}{p_1^2}\{p_1^3 L_s^3(Z - 2\,Z_o) + R_z^2[(Z - 2\,Z_o)R_z + Z\,Z_o]\},$$

natomiast

$$K = \frac{8\,A\,Z\,Z_o}{p_1 L_s^3 (Z + 2\,Z_o)^3}.$$

Całkowite napięcie wierzchołka słupa jest sumą fal składowych u_{w1} $u_{w2}, u_{w3} \ldots u_{wn}$, przy czym mnożniki $e^{-p2k\tau'_o}$ występujące w wyrażeniu (29) wskazują zgodnie z twierdzeniem o przesunięciu na początek działania poszczególnych fal. Mamy więc ostatecznie

$$u_1(t) = \left[B_1 t - \frac{B_2}{p_1}(1 - e^{p_1 t})\right]\mathbf{1}_0 +$$

$$+\left\{H_{0_2}(t - 2\tau_0') + \frac{H_{1_2}}{p_1}(t - 2\tau_0')\, e^{p_1(t-2\tau_0')} + \left[\frac{H_{1_2}}{p_1^2} - \frac{H_{2_2}}{p_1}\right][1 - e^{p_1(t-2\tau_0')}]\right\}\mathbf{1}_{2\tau_0'} +$$

$$+\left\{H_{0_3}(t - 4\tau_0') - \left[\frac{H_{1_3}}{p_1^3} - \frac{H_{2_3}}{p_1^2} + \frac{H_{3_3}}{p_1}\right]\left[1 - e^{p_1(t-4\tau_0')}\right] +\right.$$

$$\left. - \left[\frac{H_{1_3}}{p_1^2} - \frac{H_{2_3}}{p_1}\right](t - 4\tau_0')\, e^{p_1(t-4\tau_0')} + \frac{H_{1_3}}{2p_1}(t - 4\tau_0')^2\, e^{p_1(t-4\tau_0)}\right\}\mathbf{1}_{4\tau_0'} + \cdots \quad (40)$$

3. METODA FOURIERA

Rozważymy najpierw najprostszy układ obliczeniowy (rys. 1) odpowiadający założeniu, że przęsła są nieskończenie długie. Korzystamy z równania (6) dla transformaty napięcia, którą przekształcamy przechodząc od funkcji wykładniczych do funkcji hiperbolicznych:

$$U_s(p) = U_o(p)\frac{Z_s\,[R_z \cosh p(\tau_o - \tau) + Z_s \sinh p(\tau_o - \tau)]}{c_1 \sinh p\tau_o + d_1 \cosh p\tau_o}, \quad (41)$$

przy czym wprowadzamy następujące oznaczenia:

$$c_1 = \frac{Z Z_o R_z + Z_s^2 (Z + 2Z_o)}{2Z},$$

$$d_1 = \frac{Z Z_o Z_s + Z_s R_z (Z + 2Z_o)}{2Z}.$$

Przyjmujemy, jak poprzednio, liniowy przebieg fali pioruna, podstawiamy więc $U_o(p) = \frac{A}{p}$, otrzymując

$$U_s(p) = \frac{AZ_s}{p}\,\frac{[R_z \cosh p(\tau_o - \tau) + Z_s \sinh p(\tau_o - \tau)]}{[c_1 \cdot \sinh p\tau_o + d_1 \cdot \cosh p\tau_o]} = \frac{M(p)}{pN(p)}. \quad (42)$$

Dla wyrażenia $\frac{M(p)}{N(p)}$ stosujemy twierdzenie o rozkładzie (31), a następnie wyznaczamy $u_s(t)$ drogą całkowania.

Znajdujemy wartości p, dla których $N(p) = 0$, a więc wyznaczamy pierwiastki równania:

$$c_1 \sinh p\tau_o + d_1 \cosh p\tau_o = 0,$$

czyli

$$\operatorname{tgh} p\tau_o \parallel -\frac{d_1}{c_1}. \quad (43)$$

Ponieważ pierwiastki te mogą być zespolone, zakładamy

$$p\tau_o = \varphi + j\psi.$$

Podstawiając powyższe wyrażenie do równania (43), otrzymujemy po przekształceniu

$$\operatorname{tgh} p\tau_o = \frac{\sinh\varphi \cdot \cosh\varphi + j \sin\psi \cdot \cos\psi}{\cosh^2\varphi - \sin^2\psi} = -\frac{d_1}{c_1}. \tag{44}$$

Równanie (44) rozpada się na dwa równania:

$$\frac{\sinh\varphi \cdot \cosh\varphi}{\cosh^2\varphi - \sin^2\psi} = -\frac{d_1}{c_1}, \tag{45}$$

$$\sin\psi \cdot \cos\psi = 0. \tag{46}$$

Równanie (46) daje pierwiastki $\psi = k\frac{\pi}{2}$, gdzie k jest dowolną liczbą całkowitą. Podstawiając $\psi = k\frac{\pi}{2}$ do równania (45) otrzymujemy przy k parzystym

$$\operatorname{tgh}\varphi = -\frac{d_1}{c_1}. \tag{47}$$

Równanie to ma rozwiązanie, gdy $\frac{d_1}{c_1} < 1$. Otrzymujemy jeden ujemny pierwiastek, który oznaczamy przez $-\mu_1$. Natomiast przy k nieparzystym równanie (45) daje

$$\operatorname{tgh}\varphi = -\frac{c_1}{d_1}. \tag{48}$$

Równanie (48) jest rozwiązalne, gdy $\frac{c_1}{d_1} < 1$ — odpowiedni pierwiastek oznaczmy przez $-\mu_2$.

Ostatecznie, gdy $\frac{d_1}{c_1} < 1$, otrzymujemy

$$\tau_o p_k = -\mu_1 + jk\pi, \tag{49}$$

natomiast gdy $\frac{d_1}{c_1} > 1$

$$\tau_o p_k = -\mu_2 + j\frac{(2k+1)}{2}\pi. \tag{50}$$

W ten sposób otrzymujemy nieskończony ciąg pierwiastków. Następnie wyznaczamy pochodną

$$N'(p) = \tau_o [c_1 \cosh p\tau_o + d_1 \sinh p\tau_o] \tag{51}$$

oraz wartość funkcji dla $p = 0$

$$\frac{M(0)}{N(0)} = \frac{A Z_s R_z}{d_1}. \tag{52}$$

Szukamy napięcia wierzchołka słupa, a więc przyjmujemy $\tau = 0$. Podstawiając wyrażenia (49), (51) oraz (52) do (31) po przekształceniach otrzymujemy

$$F(t) = D_1 + D_2 \sum_{k=-\infty}^{+\infty} \frac{e^{p_k t}}{p_k \tau_o}, \tag{53}$$

gdzie:

$$D_1 = \frac{AZ_s R_z}{d_1},$$

$$D_2 = \frac{AZ_s (c_1 R_z - d_1 Z_s)}{c_1^2 - d_1^2}.$$

Sumowanie należy rozciągnąć na wszystkie pierwiastki p_k.

W przypadku gdy $\frac{d_1}{c_1} < 1$, należy uwzględnić, że pierwiastki są określone zależnością (49). Mamy wówczas

$$F(t) = D_1 + D_2 e^{-\frac{\mu_1 t}{\tau_o}} \sum_{k=-\infty}^{+\infty} \frac{e^{jk\pi\frac{t}{\tau_o}}}{-\mu_1 + jk\pi}. \tag{54}$$

Uwzględniając, że:

$$\frac{e^{jk\pi\frac{t}{\tau_o}}}{-\mu_1 + jk\pi} + \frac{e^{-jk\pi\frac{t}{\tau_o}}}{-\mu_1 - jk\pi} = \frac{2\left[k\pi \sin k\pi\frac{t}{\tau_o} - \mu_1 \cos k\pi\frac{t}{\tau_o}\right]}{\mu_1^2 + k^2\pi^2}, \tag{55}$$

mamy

$$F(t) = D_1 + D_2 e^{-\frac{\mu_1 t}{\tau_o}} \left\{ -\frac{1}{\mu_1} + 2 \sum_{k=1}^{+\infty} \frac{\left(k\pi \sin k\pi\frac{t}{\tau_o} - \mu_1 \cos k\pi\frac{t}{\tau_o}\right)}{(\mu_1^2 + k^2\pi^2)} \right\}. \tag{56}$$

a dalej całkując w granicach od 0 do t, otrzymujemy ostatecznie

$$u_s(t) = D_1 t - D_2 \tau_o \left\{ \frac{\left(1 - e^{-\frac{\mu_1 t}{\tau_o}}\right)}{\mu_1^2} + \right.$$

$$+ 2 \sum_{k=1}^{+\infty} \frac{1}{(\mu_1^2 + k^2\pi^2)^2} \cdot \left[e^{-\frac{\mu_1 t}{\tau_o}} \left(2k\pi\mu_1 \sin k\pi \frac{t}{\tau_o} + \right.\right.$$

$$\left.\left. - (\mu_1^2 - k^2\pi^2) \cos k\pi \frac{t}{\tau_o} \right) + (\mu_1^2 - k^2\pi^2) \right] \Bigg\}. \qquad (57)$$

W ten sposób otrzymane rozwiązanie jest w postaci nieskończonego szeregu Fouriera.

Gdy $\frac{d_1}{c_1} > 1$, pierwiastki są dane zależnością (50) i wyrażenie (53) przybiera postać

$$F(t) = D_1 + D_2 e^{-\frac{\mu_2 t}{\tau_o}} \sum_{k=-\infty}^{+\infty} \frac{e^{j\frac{(2k+1)}{2}\frac{\pi t}{\tau_o}}}{-\mu_2 + j\frac{(2k+1)}{2}\pi}.$$

Występującą w tym wzorze sumę możemy rozłożyć na dwa składniki, pierwszy stanowiący sumę w granicach od $k = 0$ do $+\infty$, a drugi — sumę w granicach od $k = -1$ do $-\infty$. Biorąc pod uwagę, że wartość czynnika $2k + 1$ przy $k = -1, -2, -3 \ldots$ różni się od wartości $2k + 1$ przy $k = 0, 1, 2 \ldots$ tylko znakiem, możemy napisać:

$$F(t) = D_1 + D_2 e^{-\frac{\mu_2 t}{\tau_o}} \sum_{k=0}^{+\infty} \left\{ \frac{e^{j\frac{(2k+1)}{2}\frac{\pi t}{\tau_o}}}{-\mu_2 + j\frac{(2k+1)}{2}\pi} + \frac{e^{-j\frac{(2k+1)}{2}\frac{\pi t}{\tau_o}}}{-\mu_2 - j\frac{(2k+1)}{2}\pi} \right\},$$

a następnie przekształcając mamy

$$F(t) = D_1 + 2 D_2 e^{-\frac{\mu_2 t}{\tau_o}} \sum_{k=0}^{+\infty} \frac{\frac{(2k+1)}{2}\pi \sin\frac{(2k+1)}{2}\frac{\pi t}{\tau_o} - \mu_2 \cos\frac{(2k+1)}{2}\frac{\pi t}{\tau_o}}{\mu_2^2 + \left(\frac{2k+1}{2}\pi\right)^2}. \qquad (58)$$

Całkując (58) otrzymujemy szukane rozwiązanie

$$u_s(t)=D_1 t-2D_2\tau_o\sum_{k=0}^{+\infty}\frac{1}{\left[\mu_2^2+\left(\frac{2k+1}{2}\pi\right)^2\right]^2}\left\{e^{-\frac{\mu_2 t}{\tau_o}}\left[(2k+1)\pi\mu_2\sin\frac{(2k+1)}{2}\frac{\pi t}{\tau_o}+\right.\right.$$

$$\left.\left.-\left[\mu_2^2-\left(\frac{2k+1}{2}\pi\right)^2\right]\cos\frac{(2k+1)}{2}\frac{\pi t}{\tau_o}\right]+\mu_2^2-\left(\frac{2k+1}{2}\pi\right)^2\right\}. \quad (59)$$

Przy pomocy metody Fouriera rozpatrzmy jeszcze układ obliczeniowy z rys. 4. W tym celu przekształcamy równanie (24) otrzymane dla tego układu wprowadzając funkcje hiperboliczne oraz podstawiamy dla liniowego przebiegu prądu pioruna $U_o(p)=\frac{A}{p}$. Dla napięcia wierzchołka słupa ($\tau'=0$) otrzymujemy

$$U_1(p)=\frac{2AZ(pL_s+R_z)\sinh p\tau_o'}{p\left[Z(pL_s+R_z+Z_o)\sinh p\tau_o'+2Z_o(pL_s+R_z)\cosh p\tau_o'\right]}=\frac{M(p)}{pN(p)}. \quad (60)$$

Stosując twierdzenie o rozkładzie obliczamy

$$\frac{M(0)}{N(0)}=0,$$

$$N'(p)=\left[2Z_o(pL_s+R_z)\tau_o'+ZL_s\right]\sinh p\tau_o'+$$
$$+\left[Z(pL_s+R_z+Z_o)\tau_o'+2Z_oL_s\right]\cosh p\tau_o'$$

oraz pierwiastki równania $N(p)=0$.

Równanie $N(p)=0$ prowadzi do równania transcendentalnego

$$\operatorname{tgh} p\tau_o'=-\frac{2Z_o(pL_s+R_z)}{Z\left[pL_s+R_z+Z_o\right]}. \quad (61)$$

Równanie (61) daje się rozwiązać jedynie metodami graficznymi lub numerycznymi. Rozwiążemy to równanie metodą graficzną stosując odwzorowanie konformiczne.

Funkcję $\operatorname{thp}\tau_o'$ wyrażamy przez funkcję wykładniczą i wówczas równanie (61) możemy sprowadzić do postaci:

$$e^{2p\tau_o'}=\frac{\left[pL_s(Z-2Z_o)+ZR_z+ZZ_o-2Z_oR_z\right]}{\left[pL_s(Z+2Z_o)+ZR_z+ZZ_o+2Z_oR_z\right]}. \quad (62)$$

Wprowadzając nową zmienną $z=2p\tau'_o$, przy czym $z=\xi+j\eta$, oraz nowe oznaczenia na odpowiednie wyrazy równania (62), otrzymujemy równanie

$$e^z=\frac{az+b}{cz+d}, \quad (63)$$

w wyniku podstawienia $z = \xi + j\eta$ do funkcji homograficznej

$$w = \frac{az + b}{cz + d},$$

po przekształceniu otrzymujemy

$$w = \frac{(a\xi + b)(c\xi + d) + ac\eta^2}{(c\xi + d)^2 + c^2\eta^2} + j\frac{\eta(ad - bc)}{(c\xi + d)^2 + c^2\eta^2}. \tag{64}$$

Część rzeczywistą funkcji w oznaczmy przez u, zaś część urojoną przez v, a więc

$$w = u + jv.$$

Odwzorowanie funkcji homograficznej dla $\xi = \text{const}$, $-\infty < \eta < +\infty$ w płaszczyźnie w przedstawia rodzinę kół o równaniu

$$u^2 + v^2 - u\left[\frac{a}{c} + \frac{a\xi + b}{c\xi + d}\right] + \frac{a}{c}\frac{(a\xi + b)}{(c\xi + d)} = 0. \tag{65}$$

Środki tych kół leżą na osi u (współrzędne środków

$$u = \frac{1}{2}\left[\frac{a}{c} + \frac{a\xi + b}{c\xi + d}\right],\ v = 0).$$

Promienie kół możemy obliczyć z wyrażenia

$$r_\xi = \frac{1}{2}\left[\frac{a\xi + b}{c\xi + d} - \frac{a}{c}\right].$$

Dla $\eta \to 0$ mamy

$$\lim_{\eta \to 0} w = \frac{a\xi + b}{c\xi + d} + j0,$$

natomiast dla $\eta \to \pm\infty$

$$\lim_{\eta \to \pm\infty} w = \frac{a}{c} + j0.$$

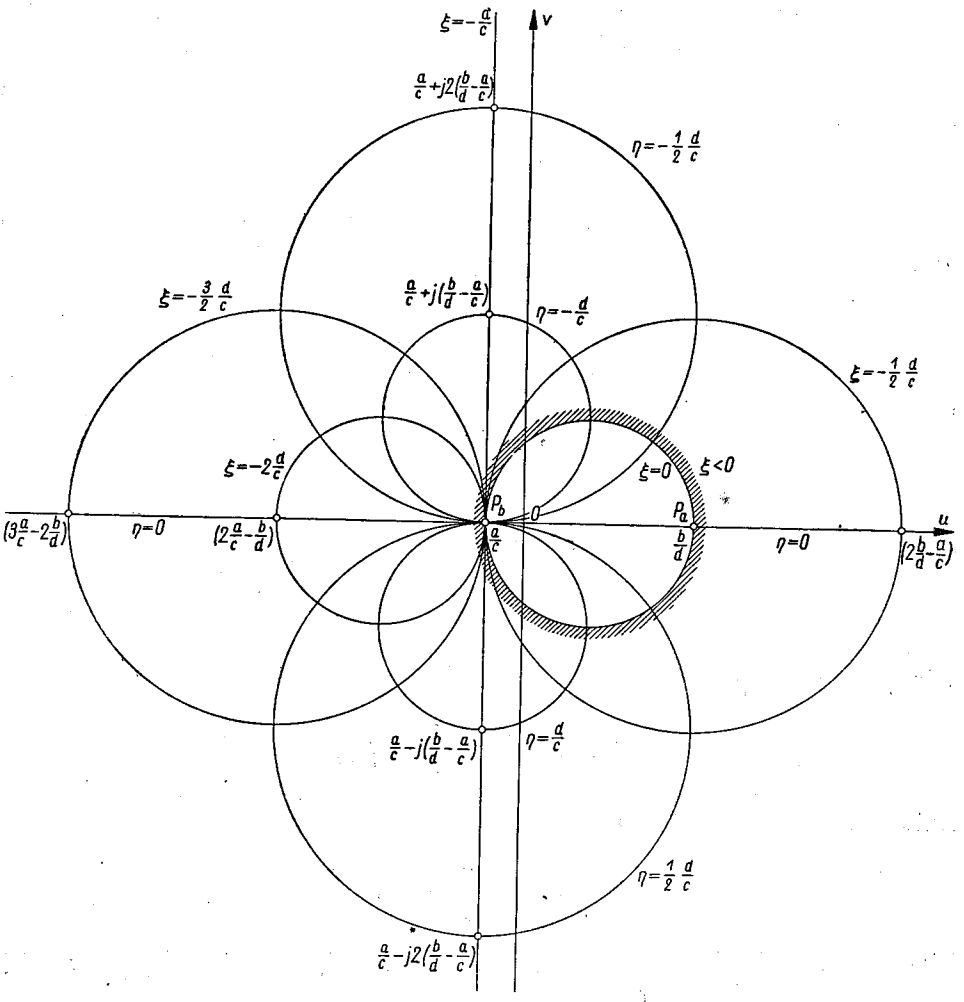
Stąd wynika, iż okręgi wszystkich kół $\xi = \text{const}$, przechodzą przez punkt o współrzędnych $u = \frac{a}{c}$, $v = 0$.

Odwzorowania funkcji homograficznej dla $\eta = \text{const}$, $-\infty < \xi < +\infty$ stanowią układ kół o równaniu

$$u^2 + v^2 - u\frac{2a}{c} - v\frac{(ad - bc)}{c^2\eta} + \frac{a^2}{c^2} = 0. \tag{66}$$

Środki tych kół leżą na prostej $u = \frac{a}{c}$ $\left(\text{współrzędne środków } u = \frac{a}{c},\right.$ $\left. v = \frac{ad - bc}{2c^2\eta}\right)$, natomiast promienie daje wyrażenie

$$r_\eta = \frac{ad - bc}{2c^2\eta}.$$



Rys. 5. Odwzorowanie konformiczne funkcji homograficznej $w = \frac{az+b}{cz+d}$

Gdy $\xi \to 0$,

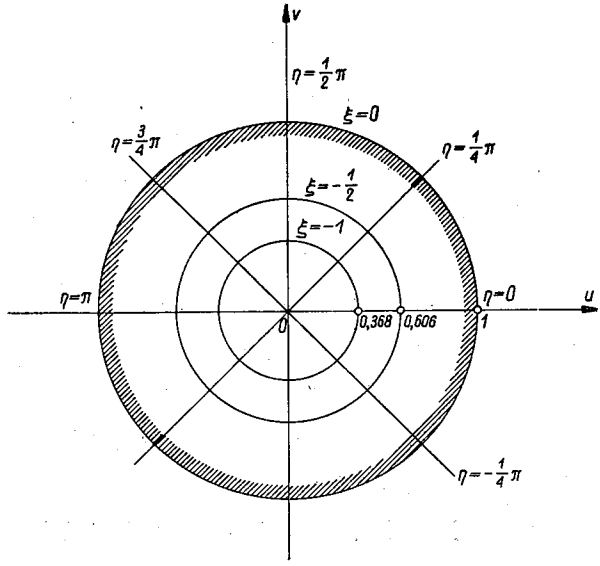
$$\lim_{\xi \to 0} w = \frac{bd + ac\eta^2}{d^2 + c^2\eta^2} + j\frac{\eta(ad - bc)}{d^2 + c^2\eta^2},$$

a dla $\xi \to \pm \infty$

$$\lim_{\xi \to \pm \infty} w = \frac{a}{c} + j\,0\,.$$

Widzimy stąd, że wszystkie okręgi dla $\eta = \text{const}$ przechodzą przez punkt $u = \frac{a}{c}, v = 0$. Koła te są ortogonalne do kół odpowiadających $\xi = \text{const}$.

Na rys. 5 podano przykładowo odwzorowanie funkcji homograficznej dla $\xi = \text{const}$ oraz dla $\eta = \text{const}$. Z uwagi na to, że Z_0 mało różni się od Z, natomiast $R_z < Z$, mamy $\frac{a}{c} < 0$ oraz $0 < \frac{b}{d} < 1$. Punkt P_a odpowiada $z = 0$, punkt 0 zaś $z = -\frac{b}{a}$. Przez punkt P_b przechodzą wszyst-



Rys. 6. Odwzorowanie konformiczne funkcji wykładniczej $w = e^z$

kie koła dla różnych $\xi = \text{const}$ oraz $\eta = \text{const}$, przy czym dla kół $\xi = \text{const}$ odpowiada on $\eta \to \pm \infty$, natomiast dla kół $\eta = \text{const}$ — $\xi \to \pm \infty$.

Dla zbadania odwzorowania funkcji wykładniczej $w = e^z$ podstawiamy $z = \xi + j\eta$

$$w = e^z = e^{\xi}(\cos\eta + j\sin\eta).$$

Przedstawiając w we współrzędnych biegunowych otrzymujemy

$$w = re^{j\theta}\,.$$

Mamy więc

$$r = e^{\xi}, \quad \theta = \eta.$$

W ten sposób dla $\xi = \text{const}$ odwzorowaniem funkcji e^z są koła koncentryczne o środku w początku układu, natomiast dla $\eta = \text{const}$ rodzina linii prostych wychodzących promieniowo z początku układu. Odwzorowania te przedstawiono na rys. 6.

Funkcja homograficzna przekształca półpłaszczyznę zmiennej zespolonej $-\infty < \xi < 0$, $-\infty < \eta < +\infty$ na obszar zewnątrz okręgu $\xi = 0 = \text{const}$, natomiast funkcja wykładnicza przekształca tę płaszczyznę na wnętrze okręgu $\xi = 0 = \text{const}$ — na rys. 5 i 6 odpowiednie obszary zaznaczono kreskowaniem. W obszarze wspólnym obydwu odwzorowaniom należy poszukiwać pierwiastków z równania (63). Z powyższych rozważań przewidujemy, że części rzeczywiste pierwiastków z będą ujemne.

Mając odwzorowania funkcji homograficznej oraz funkcji wykładniczej dla różnych wartości $\xi = \text{const}$, możemy wyznaczyć punkty płaszczyzny w odpowiadające tym samym wartościom ξ dla obu funkcji jako punkty przecięcia odpowiednich kół — punkty te utworzą krzywą, którą oznaczmy przez G. Podobnie wyznaczamy punkty płaszczyzny w odpowiadające tym samym wartościom η dla obu funkcji jako punkty przecięcia odpowiednich kół oraz półprostych. Ponieważ jednak każda półprosta odpowiada wielu wartościom η różniącym się o $2n\pi$, punkty przecięcia utworzą cały szereg krzywych odpowiadających kolejno $n = 0, 1, 2 \ldots$ — krzywe te oznaczmy przez F_n. Punkty przecięcia krzywej G z krzywymi F_n odpowiadają tym samym wartościom ξ oraz η dla funkcji homograficznej oraz wykładniczej, a więc odpowiadają rozwiązaniom równania (63).

Ponieważ oś Ou odpowiada części wspólnej odwzorowania obydwu funkcji dla $\eta = 0 = \text{const}$, więc otrzymamy jeden pierwiastek rzeczywisty o wartości $\ln\frac{b}{d} < \xi < 0$. Pozostałe pierwiastki będą zespolone.

Dla coraz to wyższych wartości n część rzeczywista pierwiastków zespolonych dąży do wartości $\ln\frac{a}{c}$, natomiast część urojona wzrasta nieograniczenie. Przechodząc od zmiennej z do zmiennej p otrzymamy szukane pierwiastki równania (61).

Obliczmy składowe $u_{1}(t)$ odpowiadające pierwiastkowi rzeczywistemu

$$p_o = -\zeta$$

oraz pierwiastkom sprzężonym

$$p_k = -\delta_k \pm j\,\omega_k.$$

Wprowadzając oznaczenia

$$\frac{M(p_o)}{p_o N'(p_o)} = S,$$

$$\frac{M(p_k)}{p_k N'(p_k)} = P_k \pm j Q_k,$$

otrzymamy

$$F(t) = Se^{-\zeta t} + \sum_{k=1}^{\infty} (P_k + j Q_k) e^{(-\delta_k + j\omega_k)t} + \sum_{k=1}^{\infty} (P_k - j Q_k) e^{(-\delta_k - j\omega_k)t}.$$

Wyrażenie to możemy przekształcić do postaci

$$F(t) = Se^{-\zeta t} + 2 \sum_{k=1}^{\infty} e^{-\delta_k t} (P_k \cos \omega_k t - Q_k \sin \omega_k t), \tag{67}$$

a dalej całkując otrzymujemy $u_1(t)$

$$u_1(t) = \frac{S}{\zeta}\left(1 - e^{-\zeta t}\right) + 2 \sum_{k=1}^{\infty} \frac{1}{\delta_k^2 + \omega_k^2} \left\{ e^{-\delta_k t} [(P_k \omega_k + Q_k \delta_k) \sin \omega_k t + \right.$$

$$\left. - (P_k \delta_k - Q_k \omega_k) \cos \omega_k t] + P_k \delta_k - Q_k \omega_k \right\}. \tag{68}$$

4. PRZYKŁADY OBLICZEŃ

Dla ilustracji naszych rozważań przeprowadzimy obliczenia obydwoma metodami dla układów rozpatrywanych przez J. L. Jakubowskiego [1]. Przyjmujemy więc następujące parametry:

$$Z_o = 300\,\Omega;\ Z_s = 200\,\Omega;\ Z = 400\,\Omega;\ R_z = 10\,\Omega;$$

$$h = 15\,\text{m};\ l = 150\,\text{m};\ A = 7500\,\text{kV}/\mu\text{s}.$$

Rozpatrzmy najpierw układ z rys. 1. Przeprowadzając obliczenia metodą d'Alemberta wyznaczamy współczynnik przejścia i współczynniki odbicia. Współczynniki te w danym przypadku będą następujące:

$$\alpha = 0{,}50000;\ \beta_1 = 0{,}25000;\ \beta_2 = 0{,}90476.$$

Czas potrzebny na przebycie przez fale długości słupa 15 m $\tau_0 = 0{,}05\ \mu$s. Na napięcie wierzchołka słupa w naszym przypadku otrzymujemy wyrażenie:

$$u_s(t) = 3750\,t - 2544{,}6\,(t - 0{,}1)\,\mathbf{1}_{0,1} - 575{,}57\,(t - 0{,}2)\,\mathbf{1}_{0,2} +$$

$$- 130{,}19\,(t - 0{,}3)\,\mathbf{1}_{0,3} - \cdots \tag{69}$$

Wartości u_s obliczone na podstawie wzoru (69) dla różnych czasów t podane są w tablicy 1. Przebieg tego napięcia daje krzywa *1* na rys. 7.

Przeprowadzając obliczenia dla układu z rys. 1 drugą metodą należy wyznaczyć pierwiastki równania (43).

W naszym przypadku $\frac{d_1}{c_1} = 0{,}63103 < 1$. Otrzymujemy

$$\tau_0 p_k = -0{,}74319 + jk\pi.$$

Rozwiązanie będzie więc przedstawione wzorem (57), przy czym

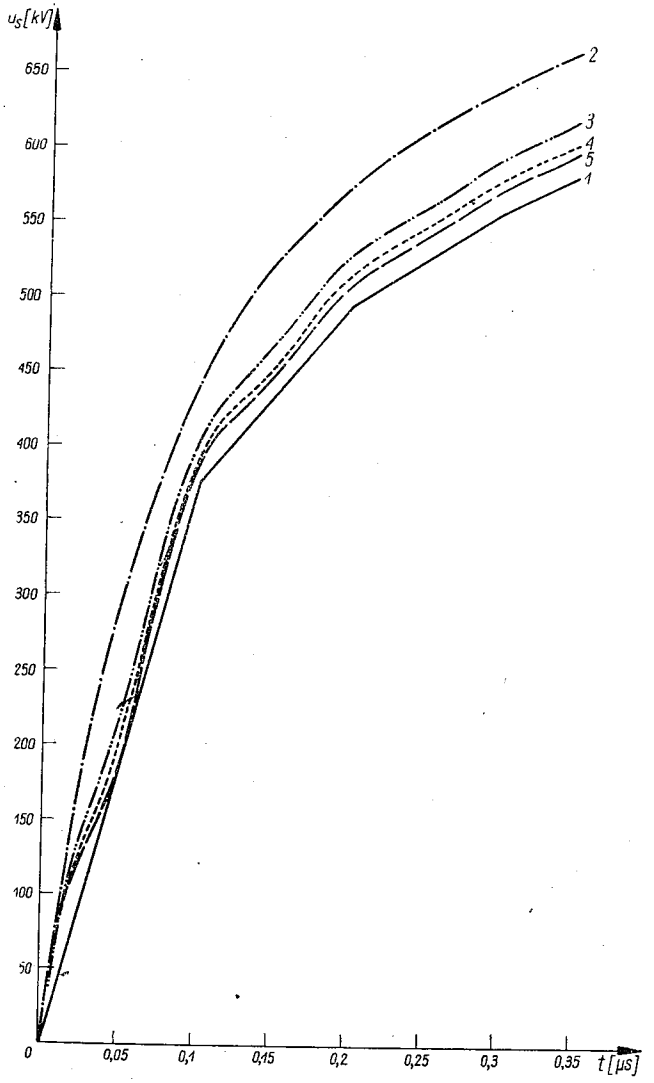
$$\mu_1 = 0{,}74319;\ D_1 = 461{,}54;\ D_2 = -281{,}25.$$

Tablica 1

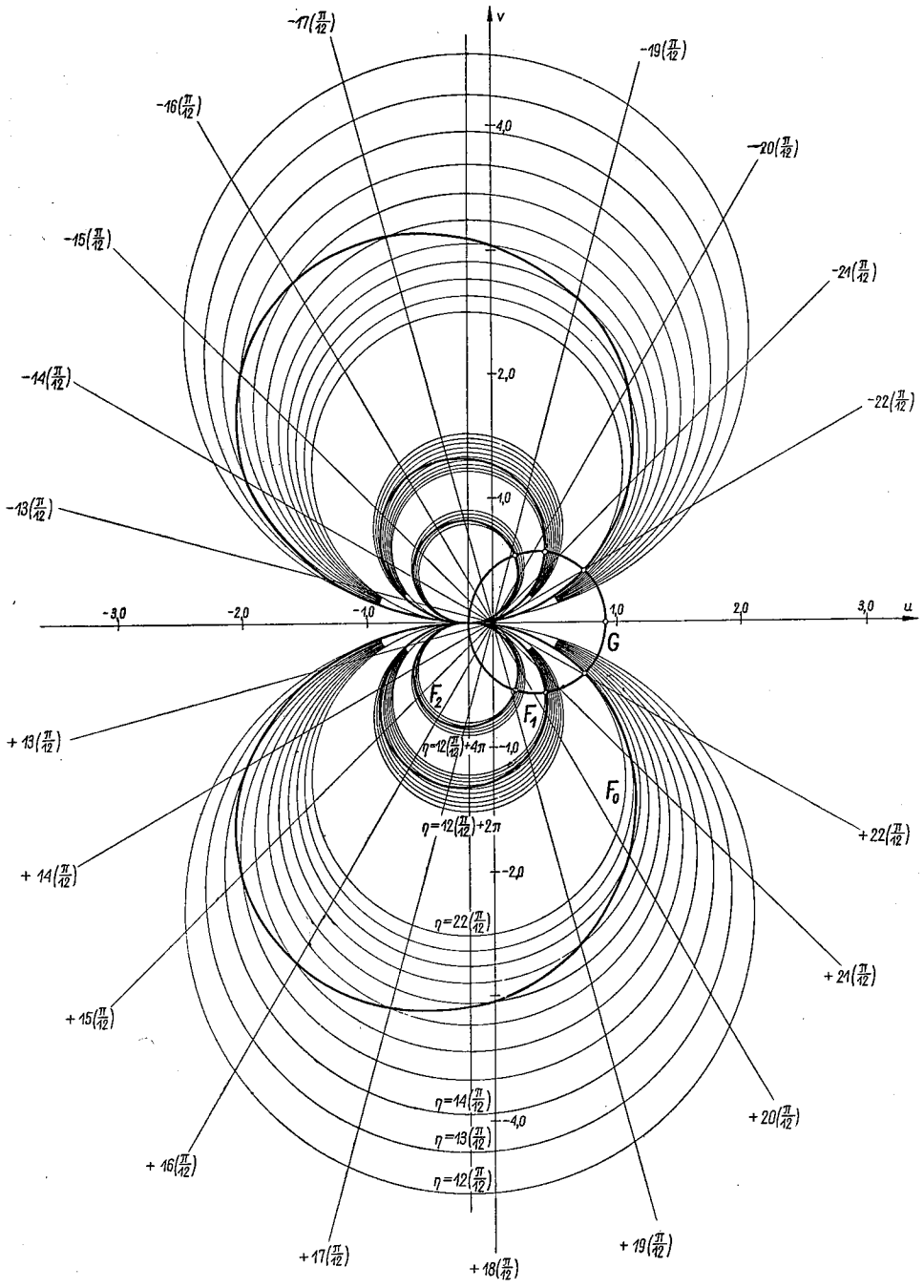
t	Napięcie wierzchołka słupa u_s dla układu z rys. 1				
	metoda d'Alemberta	metoda Fouriera: sumy składowych			
		1	1+2	1+2+3	1+2+3+4
μs	kV	kV	kV	kV	kV
0,0125	46,87	92,11	86,39	81,41	79,42
0,0250	93,75	169,6	138,0	126,6	121,1
0,0375	140,6	234,9	176,9	159,0	150,7
0,0500	187,5	290,1	218,9	198,7	189,6
0,0625	234,4	336,9	268,4	250,1	241,8
0,0750	281,2	376,8	320,6	305,9	299,3
0,0875	328,1	410,9	367,3	355,6	350,4
0,1000	375,0	440,2	402,8	392,3	387,5
0,1125	390,1	465,5	426,8	415,4	410,2
0,1250	405,1	487,5	443,0	429,8	423,8
0,1375	420,2	506,7	456,2	441,6	434,9
0,1500	435,3	523,7	470,2	455,1	448,2
0,1625	450,3	538,7	485,9	471,1	464,5
0,1750	465,4	552,2	502,1	488,2	481,9
0,1875	480,5	564,4	517,2	504,0	498,0
0,2000	495,5	575,5	529,7	516,7	510,8
0,2125	503,4	585,6	539,6	526,4	520,4
0,2250	511,3	595,1	547,7	534,1	528,0
0,2375	519,1	603,9	555,1	541,3	534,9
0,2500	527,0	612,2	562,8	548,8	542,4
0,2625	534,9	620,1	570,8	556,9	550,6
0,2750	542,8	627,6	578,9	565,2	559,0
0,2875	550,6	634,8	586,8	573,2	567,1
0,3000	558,5	641,8	594,1	580,6	574,4
0,3125	564,8	648,5	600,8	587,2	581,1
0,3250	571,0	655,1	607,1	593,4	587,2
0,3375	577,2	661,6	612,9	599,2	592,9
0,3500	583,5	667,9	619,4	605,7	599,4

Wartości u_s obliczone metodą Fouriera dla rozmaitych t przy uwzględnieniu kolejno coraz to większej liczby składowych od pierwszej do czwartej zestawiono w tablicy 1. Odpowiednie wykresy dla napięcia u_s podają krzywe *2* ÷ *5* z rys. 7.

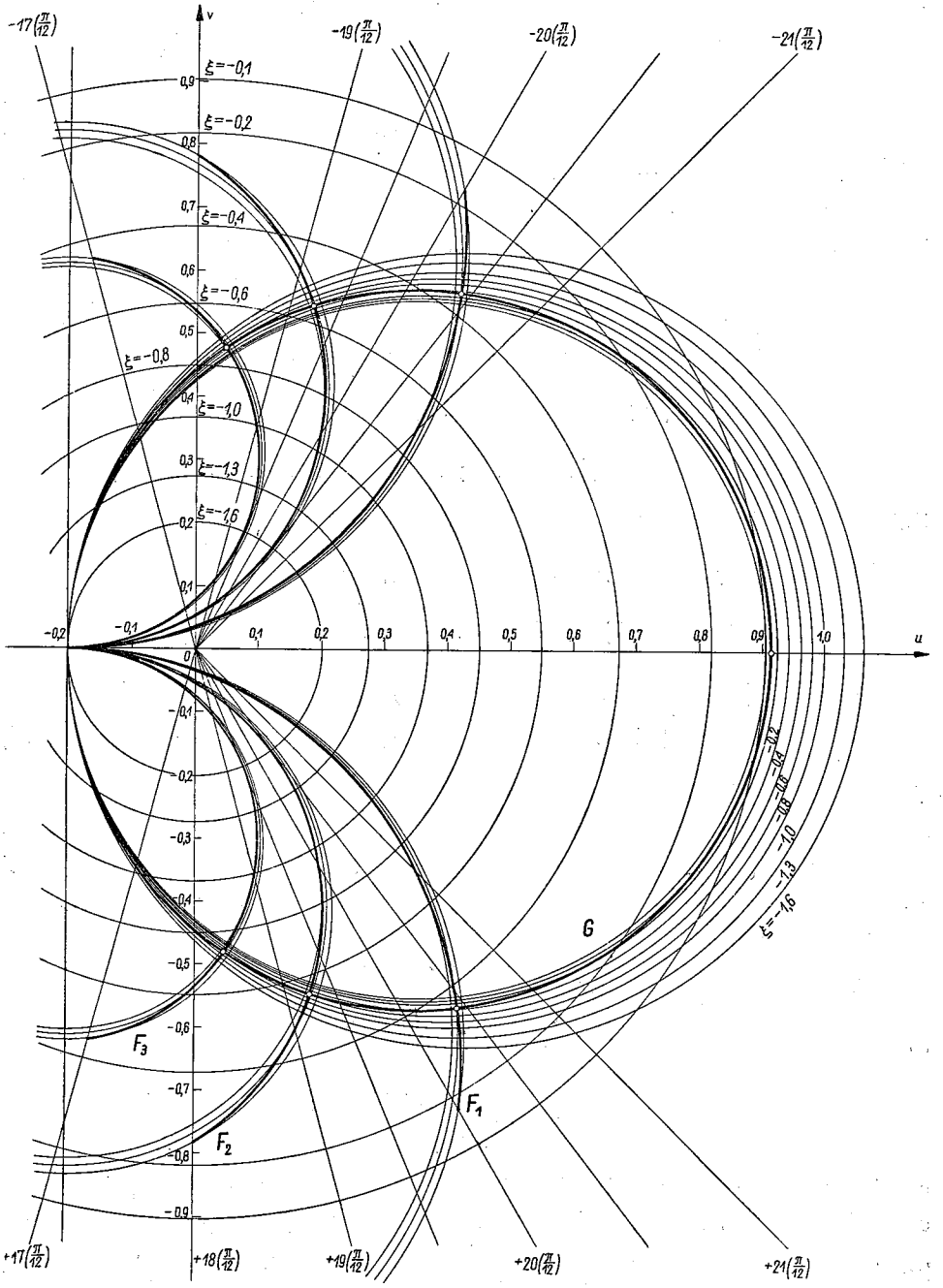
Z porównania tych krzywych z podanym na tym samym rysunku przebiegiem napięcia otrzymanym metodą d'Alemberta wynika, iż



Rys. 7. Przebieg czasowy napięcia wierzchołka słupa u_s określony dla układu z rys. 1: *1* — metoda d'Alemberta, *2* ÷ *5* — metoda Fouriera przy uwzględnieniu kolejno coraz większej liczby składowych, *2* — 1 składowa, *3* — 1 plus 2, *4* — 1 plus 2, plus 3, *5* — 1 plus 2, plus 3, plus 4



Rys. 8. Wyznaczenie pierwiastków równania transcendentalnego $e^z = \frac{az+b}{cz+d}$ metodą odwzorowania konformicznego. Wykres odpowiada rozpatrywanemu przypadkowi liczbowemu ($a = -1$, $b = 59$, $c = 5$, $d = 65$)



Rys. 9. Powiększony fragment rys. 8 przy pominięciu krzywej F_0 i podaniu dodatkowo części krzywej F_3

w miarę uwzględnienia większej liczby składowych krzywa coraz bardziej zbliża się do przebiegu rzeczywistego.

Z kolei przeprowadzimy obliczenia dla układu z rys. 4, w którym słup zastąpiono przez indukcyjność skupioną. Indukcyjność słupa określa się z zależności $L_s = \tau_o Z_s$. Podstawiając odpowiednie wartości otrzymujemy $L_s = 10\ \mu H$. Aby wyznaczyć napięcia pierwszą metodą, korzystamy ze wzoru (40) i w tym celu obliczamy wartości liczbowe wchodzących do tego wzoru stałych:

$$p_1 = -13;\ B_1 = 461{,}54;\ B_2 = 5538{,}5;$$

$$H_{0_2} = -42{,}603;\ H_{1_2} = 79754;\ H_{2_2} = -7157{,}4;$$

$$\tau_o' = 0{,}5\ \mu s;\ H_{0_3} = -38{,}671;\ H_{1_3} = 1{,}1484 \cdot 10^6;$$

$$H_{2_3} = -1{,}1902 \cdot 10^5;\ H_{3_3} = 1{,}4787 \cdot 10^3.$$

Ostatecznie otrzymujemy

$$u_1(t) = [461{,}54\,t + 426{,}03\,(1 - e^{-13t})]\,\mathbf{1}_0 - [42{,}603\,(t-1) +$$

$$+ 6134{,}9\,(t-1)\,e^{-13\,(t-1)} + 78{,}652\,(1 - e^{-13\,(t-1)})]\,\mathbf{1}_{1,0} +$$

$$- [38{,}671\,(t-2) + 67{,}764\,(1 - e^{-13\,(t-2)}) - 2359{,}6\,(t-2)\,e^{-13\,(t-2)} +$$

$$+ 44{,}71\,(t-2)^2\,e^{-13\,(t-2)}]\,\mathbf{1}_{2,0} + \cdots \qquad (70)$$

Wyniki obliczeń napięcia u_1 wg powyższego wzoru podano w tablicy 2, natomiast odpowiedni przebieg napięcia przedstawia krzywa *1* na rys. 10.

Przy obliczaniu napięcia dla układu z rys. 4 drugą metodą należy wyznaczyć pierwiastki równania transcendentalnego (61).

W naszym przypadku otrzymujemy równanie

$$\operatorname{tgh}\frac{p}{2} = -\frac{3\,(p+1)}{2\,(p+31)},$$

które sprowadzamy do postaci

$$e^z = \frac{-z+59}{5z+65},$$

a więc zgodnie z przyjętymi oznaczeniami mamy $a = -1$, $b = 59$, $c = 5$, $d = 65$. Rozwiązując te równanie przy pomocy odwzorowania konformicznego wykreślamy odwzorowania funkcji $w = \frac{-z+59}{5z+65}$ oraz funkcji $w = e^z$ dla różnych wartości $\xi = \text{const}$ i $\eta = \text{const}$. Rysunek 8 przedstawia krzywą G oraz krzywe F_0, F_1 i F_2. Na rys. 9 przedstawiającym

powiększony fragment rys. 8 podano krzywą G oraz interesujące nas części krzywych F_1, F_2 i F_3. Z wykresów tych otrzymujemy wartości pierwszych pięciu pierwiastków. W naszym przypadku $\tau'_o = 0{,}5$, zachodzi więc równość $z = 2p\tau'_o = p$ i otrzymujemy bezpośrednio pierwiastki

$$p_0 = -0{,}0885;\ p_1 = -0{,}170 \pm j \cdot 5{,}7639;\ p_2 = -0{,}353 \pm j \cdot 11{,}632;$$

$$p_3 = -0{,}554 \pm j \cdot 17{,}609;\ p_4 = 0{,}731 \pm j \cdot 23{,}758.$$

Składowe napięcia $u_1(t)$ odpowiadające tym pierwiastkom obliczamy ze wzoru (68). Stałe występujące w tym wzorze przyjmują w naszym przypadku następujące wartości:

$$S = 404{,}69;\ P_1 = -135{,}86;\ P_2 = -232{,}71;$$

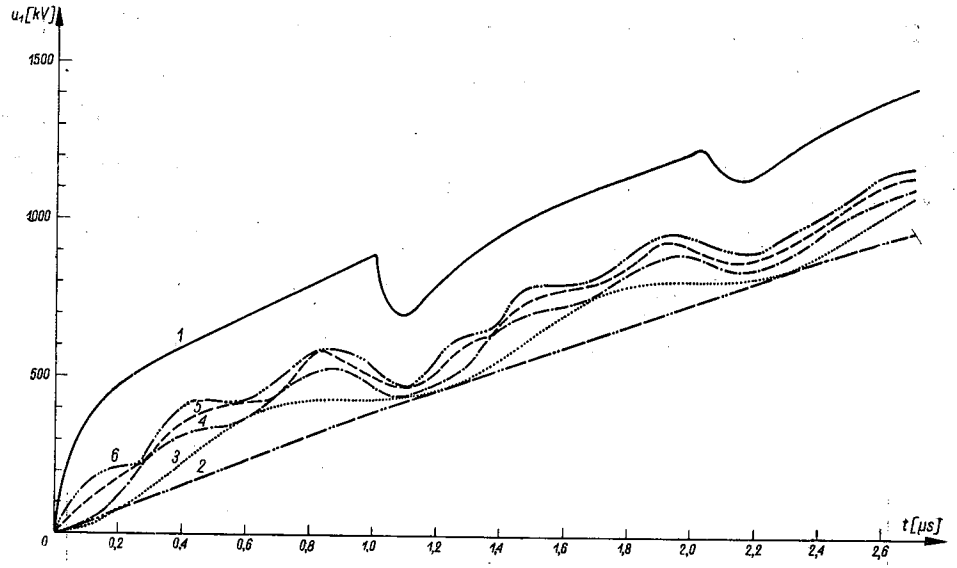
$$P_3 = -299{,}49;\ P_4 = -323{,}29;\ Q_1 = -190{,}85;$$

$$Q_2 = -300{,}79;\ Q_3 = -341{,}05;\ Q_4 = -329{,}14.$$

Tablica 2

t	Napięcie wierzchołka słupa u_1 dla układu z rys. 4					
	Metoda d'Alemberta	Metoda Fouriera — sumy składowych				
		1	1 + 2	1 + 2 + 3	1 + 2 + 3 + 4	1+2+3+4+5
μs	kV	kV	kV	kV	kV	kV
0,1363	416,4	54,81	40,94	52,49	157,9	191,3
0,2725	539,5	109,0	126,9	224,2	228,9	229,8
0,4088	612,6	162,5	237,6	322,1	363,7	415,5
0,5450	677,2	215,3	339,1	346,1	406,7	409,7
0,6813	740,4	267,6	404,0	424,7	427,6	476,8
0,8176	803,4	319,2	426,6	517,4	582,8	588,9
0,9538	866,3	370,2	425,5	501,9	531,4	577,4
1,090	699,7	420,5	431,5	444,6	469,4	478,8
1,226	834,7	470,3	469,7	497,6	559,7	602,4
1,363	941,6	519,4	545,3	630,3	645,5	658,2
1,499	1013	568,1	641,5	711,5	758,7	798,1
1,635	1074	616,1	729,9	748,5	792,3	808,0
1,771	1132	663,5	787,8	821,4	842,3	878,8
1,908	1189	710,3	810,6	890,5	945,6	964,1
2,044	1225	756,6	813,6	878,6	906,6	940,5
2,180	1138	802,4	822,5	845,8	881,5	902,4
2,316	1221	847,6	858,1	896,0	944,6	976,3
2,453	1310	892,2	924,7	1000	1024	1047
2,589	1377	936,4	1008	1069	1116	1146
2,725	1434	980,0	1085	1113	1150	1174
2,861	1487	1023	1137	1179	1209	1238

Tablica 2 podaje wartości u_1 odpowiadające pierwszej składowej a także kolejno sumie pierwszych kilku składowych od dwóch do pięciu, obliczonych dla różnych t. Krzywe *2* ÷ *6* na rys. 10 przedstawiają



Rys. 10. Przebieg czasowy napięcia wierzchołka słupa u_1 określony dla układu z rys. 4: *1* — metoda d'Alemberta, *2* ÷ *6* metoda Fouriera przy uwzględnieniu kolejno coraz to większej liczby składowych, *2* — 1 składowa, *3* — 1 plus 2, *4* — 1 plus 2, plus 3, *5* — 1 plus 2, plus 3, plus 4, *6* — 1 plus 2, plus 3, plus 4, plus 5

odpowiednie wykresy napięcia u_1. Z rysunku tego widać, że mimo uwzględnienia pięciu składowych krzywa *6* znacznie jeszcze odbiega od przebiegu napięcia otrzymanego przy pomocy metody d'Alemberta, który podaje krzywa *1*.

5. WNIOSKI

Na podstawie rozważonych przykładów widać, że metoda Fouriera jest bardzo uciążliwa, gdyż wymaga uwzględnienia znacznej liczby składowych, a poza tym prowadzi do równań transcendentalnych, które, ogólnie biorąc, można rozwiązać tylko w sposób przybliżony metodami wykreślnymi lub numerycznymi. W ten sposób metoda Fouriera nie jest przydatna przy rozwiązywaniu rozpatrywanych zagadnień.

Metoda d'Alemberta, zwana również metodą fal wędrownych, prowadzi do prostej i przejrzystej metody praktycznej, w której operuje

się wyłącznie współczynnikami przejścia i odbicia. Dlatego też metoda ta oddaje duże usługi przy rozważaniu stanów nieustalonych w liniach.

Politechnika Warszawska
Katedra Wysokich Napięć

WYKAZ LITERATURY

1. Jakubowski J. L.: Układy zastępcze stosowane przy obliczaniu napięcia wierzchołka słupa linii energetycznej trafionego przez piorun. — Rozprawy Elektrotechniczne. Tom III, 1957, z. 2, s. 231.
2. Jakubowski J. L.: Eliminacja oporności falowej przy uderzeniu pioruna w słup linii energetycznej. — Arch. El. Tom VI, 1957, z. 2, s. 279.
3. Kontorowicz M. I.: Operacjonnoje isczislenje i niestacjonarnyje jawlenja w elektriczeskich ciepjach. Tłumaczenie polskie, Warszawa 1956.
4. Whittaker E. T., Robinson G.: The Calculus of Observations. London and Glasgow 1940.

Я. Л. МАКСЕЕВСКИ

РАСЧЁТ НАПРЯЖЕНИЯ ВЕРШИНЫ ОПОРЫ, ПОРАЖЁННОЙ МОЛНИЕЙ ПО МЕТОДАМ ПРИМЕНЯЕМЫМ ДЛЯ ДЛИННЫХ ЛИНИЙ

Резюме

В последнее время появились публикации [1], [2] рассматривающие применяемые на практике способы расчёта напряжения вершины опоры энергетической линии поражённой молнией. Целью настоящей работы является обсуждение практического значения расчёта указанного напряжения по общим, применяемым для длинных линий, методам, то есть по методу Даламбера и по методу Фурьера.

Напряжение вершины опоры рассчитывается по обоим методам для двух самых важных эквивалентных схем, принимая неограниченный линейный рост волны в канале молнии. В первой схеме полагается, что соседние секции троса бесконечно длинные (рис. 1), а во второй схеме, что эти секции имеют некоторую определённую длину (рис. 3). Точное обсуждение этого последнего случая весьма сложно, и поэтому мы вводим упрощения обсуждённые в публикации [1], которые состоят в замещении поражённой опоры сосредоточенной индуктивностью. Полученная таким образом схема представлена на рис. 4.

По методу Даламбера получается решение в виде суммы блуждающих волн. Для первого случая решение представлено формулой (17), а для второго случая формулой (40). Метод Даламбера ведёт к практическому методу, в которой используются исключительно коэффициенты отражения и преломления. Поэтому указанный метод широко применяется для изучения переходных процессов в линиях.

При использовании метода Фурьера решение получается в виде бесконечного ряда. Решение для первого случая выражено формулой (57) или (59), а для второго случая бесконечной суммой членов уравнения (68).

Метод Фурьера представляет некоторые затруднения, так как он требует решения трансцендентального уравнения, которое в общем случае может быть решено только приближённо, используя численные или графические методы. Кроме того вышеуказанный метод является весьма затруднительным, так как для получения решения следует учесть значительное количество составляющих. Таким образом метод Фурьера не пригоден для решения изучаемых вопросов.

Для иллюстрации изучения приведены числовые примеры. Результаты расчёта напряжения вершины в зависимости от времени по формулам, полученным для первого случая, приведены на рис. 7, а для второго — на рис. 10. На обоих рисунках непрерывная линия соответствует методу Даламбера, а прерывистые линии — результатам полученным по методу Фурьера с учётом различного количества членов.

J. L. MAKSIEJEWSKI

CALCULATION OF VOLTAGE-TO-EARTH OF THE TOP OF A POWER-LINE TOWER STRICKEN BY LIGHTNING, BY METHODS USED FOR TRANSMISSION LINES

Summary

An analysis of methods used in practice for calculating voltage-to-earth of the top of a power-line tower stricken by lightning was subject to recent publications [1], [2]. The aim of the present paper is to examine the practical value of calculation of this voltage by means of general methods used for transmission lines i.e. d'Alembert method and Fourier method.

The tower top voltage is calculated by both methods for two equivalent circuits, being of the greatest importance, under the assumption of unlimited linear increase of the wave in the lightning channel. In the first circuit the adjacent ground wire sections are considered as infinitely long (Fig. 1). The second circuit takes into account a finite length of these ground wire sections (Fig. 3). The rigorous considerations of this last case is very involved and consequently certain simplification based on substitution of the tower struck by lightning for concentrated inductance is introduced as in [1]. The circuit thus obtained is shown in Fig. 4.

The d'Alembert method gives the solution in the form of the sum of travelling waves. For the first case the solution is represented by equation (17) and for the second case by (40). The d'Alembert method leads to a practical method in which we deal only with the coefficients of reflection and refraction. This last method is therefore generally used for the considerations of transients in lines.

Using the Fourier method the solution is given by infinite series. The solution for the first case is now expressed by (57) or (59) and for the second case by the infinite sum of terms of equation (68).

The Fourier method offers some difficulties as it requires the solution of transcendental equation which in general can be solved only approximately by numerical or graphical methods. Moreover, this method is too cumbersome as, in order to obtain a solution, a very large number of voltage components must be taken. In this way the Fourier method is not suitable for the solution of the problems under consideration.

As an illustration of the considerations, numerical examples are presented. The results of computations of top voltage, as a function of time from formulae obtained for the first case, are shown in Fig. 7 and for the second case — in Fig. 10. On both figures the continuous line corresponds to the d'Alembert method and dotted lines — to the results obtained according to the Fourier method with different numbers of the terms taken.

621.317.33:621.313.333.2

FELIKS ANDRZEJEWSKI, ZBIGNIEW ORZESZKOWSKI

Pomiarowa metoda wyznaczania współczynnika wypierania prądu w wirnikach maszyn indukcyjnych

Rękopis dostarczono 7.10.1959

Podano zasady metod matematycznych, wykreślnych i pomiarowych wyznaczania współczynnika wzrostu oporności czynnej wskutek wypierania prądu. Przedstawiono metodę autorów, omówiono jej dokładność i porównano z przytoczonymi innymi metodami. Przedstawiono wyniki pomiarów i obliczeń dla wykonanych maszyn dużych mocy, przeprowadzono analizę i wyciągnięto wnioski.

1. WSTĘP

Oporność rzeczywista uzwojenia wirnika silnika asynchronicznego zależy od częstotliwości i kształtu pręta; określenie tej oporności dla wirników klatkowych, przy pewnych wartościach częstotliwości, jest bardzo ważne, gdyż od tego zależą parametry rozruchu i pracy silnika.

Wzrost oporności wirnika ze wzrostem częstotliwości jest spowodowany zjawiskiem wypierania prądu i jest określany tzw. współczynnikiem wypierania. Wyznaczyć ten współczynnik można przy pomocy metod matematycznych [1], [2], [3], wykreślnych [5] lub też metody pomiarowej [4].

W metodach matematycznych jak i wykreślnych przyjmuje się, że w częściach prętów znajdujących się w promieniowych kanałach wentylacyjnych oraz w pierścieniach zwierających nie ma wypierania prądu. Tak samo we wzorach matematycznych nie można uwzględnić oporności lutu łączącego pręty z pierścieniami, która zależy od czynników przypadkowych, a może stanowić znaczną wartość w stosunku do oporności prętów i pierścienia. Istnieją jeszcze inne przyczyny zmieniające rozkład prądu w pręcie, jak np. bocznikujące działanie blach przylegających bezpośrednio do pręta, które na dość dużej głębokości są zawsze zwarte, szczególnie w dużych maszynach, oraz dające taki sam skutek kliny sta-

lowe, umieszczone pod prętami uzwojenia (stosowane często w przypadku prętów trapezowych). Wszystkie te czynniki sprawiają, że obliczone wartości współczynnika mogą się znacznie różnić od rzeczywiście istniejących.

Opracowane są dwie metody pomiarowe, Syromiatnikowa [4] i autorów niniejszego artykułu. Przy pomocy tych metod można doświadczalnie określić dla wykonanej maszyny współczynnik wypierania.

W artykule podane są zasadnicze wzory metod obliczeniowych [1], [2], [3], sposób wykreślny wyznaczania współczynnika wypierania [5], metoda pomiarowa [4] oraz metoda pomiarowa zaproponowana przez autorów. Przeprowadzono analizę dokładności metody, podano wyniki obliczeń i pomiarów oraz wyciągnięto wnioski odnośnie do wpływu czynników nie ujętych rachunkiem.

2. METODA RACHUNKOWA [1], [2], [3] WYZNACZANIA WSPÓŁCZYNNIKA WZROSTU OPORNOŚCI k_2

Oporność uzwojenia wirnika można traktować jako sumę oporności tej części uzwojenia, która nie podlega wypieraniu R_{2p} (części uzwojenia będące w powietrzu), oraz oporności części uzwojenia podlegającej temu zjawisku R_{2n} (część uzwojenia będąca w żelazie). Wzrost oporności określa się współczynnikiem k_2, który jest stosunkiem oporności części pręta podlegającej wypieraniu przy $f = 50$ oraz przy $f = 0$, a więc

$$k_2 = \frac{R_{2n(50)}}{R_{2n(0)}}.$$

Całkowite oporności wynoszą wtedy przy $f = 50$ Hz

$$R_{2(50)} = R_{2p} + R_{2n(50)} = R_{2p} + k_2 R_{2n(0)},$$

i przy $f = 0$ Hz

$$R_{2(0)} = R_{2p} + R_{2n(0)}.$$

Celem metody rachunkowej jest wyznaczenie współczynnika k_2 w zależności od kształtu i liczby prętów w żłobku.

2.1. Pręty o przekroju prostokątnym

Gdy w żłobku znajduje się wiele przewodów o przekroju prostokątnym (rys. 1), to współczynnik wypierania jest określony wzorem

$$k_2 = \varphi(\xi) + p(p-1)\psi(\xi), \tag{1}$$

zaś

$$\varphi(\xi) = \xi \frac{\operatorname{sh} 2\xi + \sin 2\xi}{\operatorname{ch} 2\xi - \cos 2\xi}, \tag{2}$$

$$\psi(\xi) = 2\xi \frac{\operatorname{sh} \xi - \sin \xi}{\operatorname{ch} \xi + \cos \xi}, \tag{3}$$

$$\xi = 2\pi h \sqrt{\frac{n b}{b_ż} \cdot \frac{f}{\varrho 10^5}} \tag{4}$$

we wzorach oznaczają:

p — numer warstwy, w której leży dany przewód,

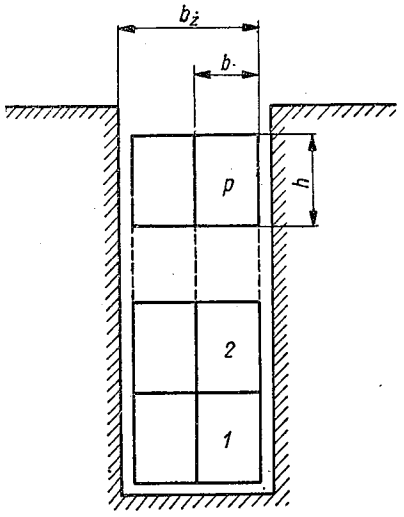
n — liczba przewodów w warstwie,

b, $b_ż$ — szerokość przewodu i żłobka [cm],

h — wysokość pręta [cm],

ϱ — oporność właściwa przewodu $\left[\frac{\Omega\,\mathrm{mm}^2}{\mathrm{m}}\right]$.

Dla $\xi \leqslant 1$ można przyjąć



Rys. 1. Układ przewodów w żłobku, p — liczba warstw

$$\psi(\xi) \approx 1 + \frac{4}{45}\xi^4 \quad \text{i} \quad \psi(\xi) = \frac{1}{3}\xi^4,$$

a dla $\xi > 2$

$$\varphi(\xi) \approx \xi \quad \text{i} \quad \psi(\xi) = 2\xi.$$

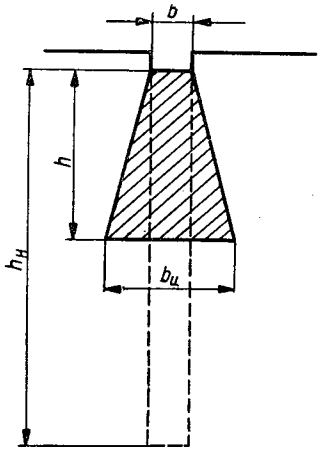
Gdy w żłobku znajduje się jeden pręt, to wtedy $n = 1$ oraz $p = 1$ i wzory (1) oraz (4) upraszczają się; dla klatki odlewanej $b = b_ż$.

W przypadku prętów o przekroju kołowym współczynnik wypierania jest mniejszy niż dla przekroju prostokątnego i oblicza się go na podsta-

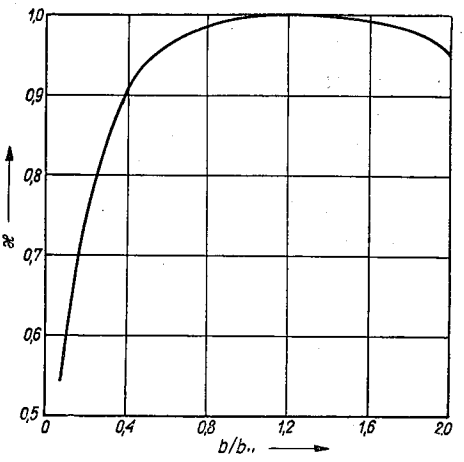
wie poprzednio podanych wzorów, uwzględniając zależność $b = h = d$, gdzie d jest średnicą przewodu.

2.2. Pręty trapezowe

W prętach o przekroju trapezowym (rys. 2) współczynnik wypierania, oprócz poprzednio wymienionych parametrów, zależy jeszcze od stosunku $u = \frac{b}{b_u}$ i rośnie ze zmniejszeniem się tego stosunku.



Rys. 2. Pręt o kształcie trapezowym i równoważny mu pręt prostokątny



Rys. 3. Krzywa zależności przewodności żłobkowej dla kształtu trapezowego pręta w funkcji stosunku

$$\frac{b}{b_u}\left[\varkappa = f\left(\frac{b}{b_u}\right)\right]$$

Dla wyznaczenia współczynnika k_2 kształt trapezowy zamienia się na prostokątny, posiadający ten sam przekrój $q = b\ h_H = (b+b_u)\frac{h}{2}$. Otrzymamy następujące wyrażenie dla k_2 jako funkcję $u = \frac{b}{b_u}$

$$k_2 = \frac{1+u}{1+u\,[2\,\psi(\xi)-1]}\,[\varphi(\xi)]^2. \tag{5}$$

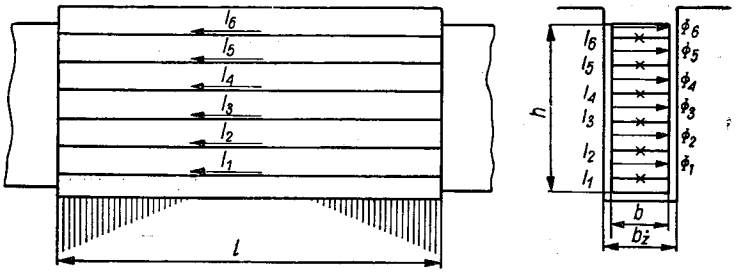
Przy pomocy podanych niżej wzorów można obliczyć wymiary pręta trapezowego wyrażonego przez wymiary równoważnego pręta prostokątnego

$$h = h_H\sqrt{\frac{2u}{\varkappa(1+u)}}, \qquad b = b_z\sqrt{\frac{2\varkappa u}{1+u}}, \qquad b_u = b_H\sqrt{\frac{2\varkappa}{u(1+u)}}, \tag{6}$$

gdzie zależność $\varkappa = \eta\left(\frac{b}{b_u}\right)$ jest przedstawiona graficznie na rys. 3. We wzorze (4) należy jako wartość $\frac{b}{b_ż}$ podstawić średnią dla wysokości h [1], [3].

3. METODA WYKREŚLNA H. NACKE [5]

Wzory matematyczne określające współczynnik k_2 nie dają poglądu na fizykalny przebieg prądów w poszczególnych elementach pręta. Uzu-



Rys. 4. Podział pręta na części elementarne

pełnieniem metod rachunkowych jest metoda wykreślna [5], obrazująca rozkład prądów w pręcie.

Część pręta znajdującą się w żelazie dzielimy wzdłuż wysokości h na pewną liczbę równych pasków, np. $n = 6$ rys. 4. W każdym z tych elementarnych pasków płynie prąd I_m, tak, że wytworzone przez te elementarne prądy strumienie są sprzężone z poszczególnymi elementami pręta. Paski te posiadają indukcyjność $L_t = \mu_0 \, h \, \frac{l}{n} \, b_ż$, której odpowiada oporność indukcyjna $X_t = 2\pi f \, L_t$ oraz oporność czynna $r_t = \varrho \frac{l\,n}{b\,h}$ (oznaczenia na rys. 4).

Pierwszy element pręta (licząc od dna żłobka) stanowi szeregowe połączenie obu oporności $(r_t + X_t)$, dając spadek napięcia $\hat{I}_1 \, (r_t + j\,X_t)$. Dla drugiego elementu mamy następujące równanie spadków napięć

$$\hat{I}_2 \, r_t = \hat{I}_1 (r_t + j\,X_t),$$

skąd

$$\hat{I}_2 = \hat{I}_1 \frac{r_t + j\,X_t}{r_t} = \hat{I}_1 + \hat{I}_1 \, j \frac{X_t}{r_t}. \qquad (7)$$

Paski 1 i 2 są sprzężone elementarnym strumieniem Φ_2, wytworzonym przez sumę prądów $(\hat{I}_1 + \hat{I}_2)$. Ten wspólny strumień wytwarza *SEM* równoważoną przez spadek napięcia $(\hat{I}_1 + \hat{I}_2)X_t$, który jest włączony szeregowo do poprzedniego układu równoległego.

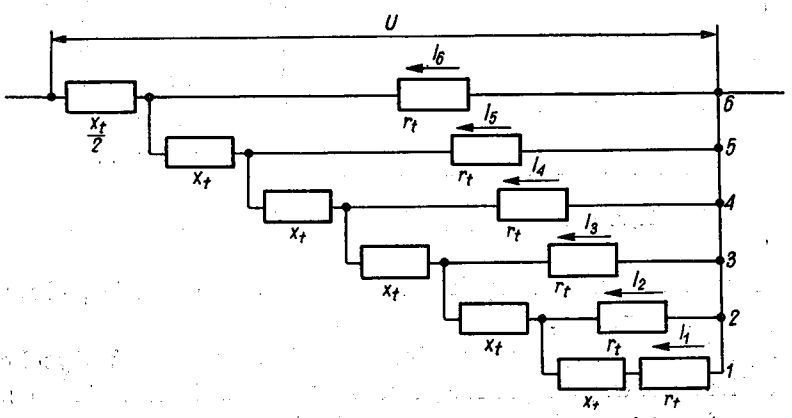
Element 3 jest objęty strumieniami Φ_1 i Φ_2, mamy więc następujące równanie

$$\hat{I}_3 r_t = \hat{I}_2 r_t + j(\hat{I}_1 + \hat{I}_2) X_t,$$

skąd

$$\hat{I}_3 = \hat{I}_2 + j(\hat{I}_1 + \hat{I}_2)\frac{X_t}{r_t}. \tag{8}$$

Postępując w ten sposób można utworzyć układ zastępczy przedstawiony na rys. 5. Dla ostatniego elementu n oporność indukcyjna wynosi $\frac{x_t}{2}$ dlatego, że posiada on tylko jedną przynależną mu przestrzeń



Rys. 5. Układ zastępczy dla elementarnych części pręta

Na podstawie otrzymanego układu zastępczego można wykonać wykres wektorowy prądów, podany na rys. 6; otrzymany wykres jest w pewnej skali również wykresem napięć, rys. 7. Do budowy obu tych wykresów potrzebne są tylko znajomości X_t i r_t, zaś wartość prądu I_1 przyjmuje się dowolną.

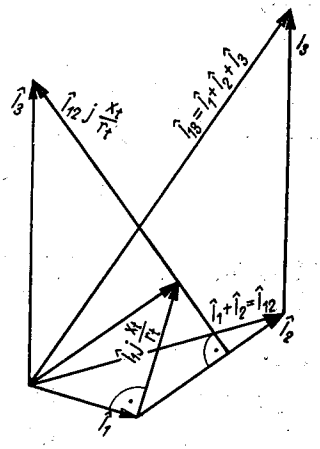
Otrzymany z wykresu (rys. 7) wektor napięcia rozkłada się na składową czynną (zgodną z kierunkiem prądu całkowitego) i bierną — prostopadłą do kierunku I. Składowa czynna wynosi $U_{\sim w} = Ir_w$, zaś bierna $U_{\sim B} = j\,I\,X_w$, gdzie r_w jest opornością czynną pręta dla prądu zmiennego, a X_w — opornością rozproszenia przestrzeni żłobka.

Do wyznaczenia współczynnika k_2 potrzebne są wartości napięcia przy prądzie zmiennym $U_{\sim w}$ i stałym $U_{=w} = \frac{I r_w}{n}$

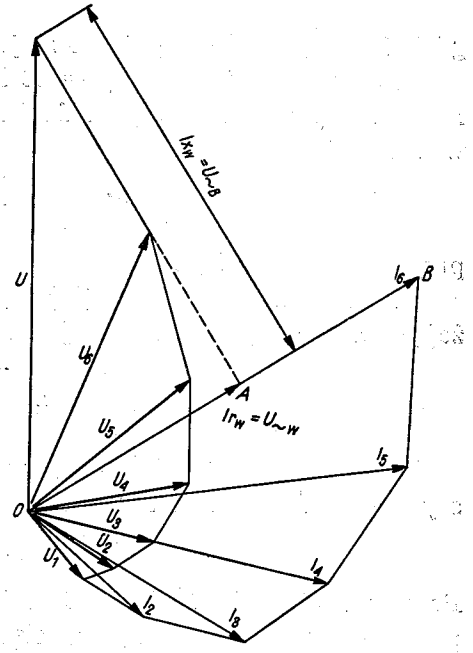
i wówczas

$$k_2 = \frac{U_{\sim w}}{U_{=w}} = \frac{OA}{\frac{OB}{n}}. \tag{9}$$

Jeżeli mamy pręty o przekroju trapezowym, to dzieli się je na elementy odpowiednio małe, aby można było przyjąć, że każdy z nich jest



Rys. 6. Wykres wektorowy prądów wg układu zastępczego



Rys. 7. Wykres wektorowy prądów i napięć dla poszczególnych elementów pręta

prostokątem o tej samej wysokości, a różnej szerokości. W schemacie zastępczym należy dla każdego elementu uwzględnić odpowiednią oporność czynną i bierną.

4. METODA POMIAROWA PODANA PRZEZ SYROMIATNIKOWA [4]

Wyznaczenie oporności przy prądzie stałym $R'_{2(0)}$ przeprowadza się z pomiaru obciążenia, przyjmując, że przy małych poślizgach oporność wirnika $R'_{2(s)}$ jest praktycznie równa $R'_{2(0)}$, a więc można napisać:

$$R'_{2(0)} \approx R'_{2(s)} = \frac{P_{12}\, s}{3\, I_2^{12}} \tag{10}$$

gdzie P_{12} — moc pola wirującego.

Prąd I'_2 oblicza się ze wzoru

$$I'_2 = I'_{2n}\frac{I_1}{I_{1n}} \approx I'_{2n}\frac{P_1}{P_{1n}}$$

lub też

$$I'_2 \approx I_1 \cos\varphi_1 = \frac{P_1}{3U_1},$$

gdzie indeksy 1 odnoszą się do stojana, a indeksy 2 do wirnika.

Oporność wirnika przy $f = 50$ Hz jest wyznaczana na podstawie pomiaru mocy zwarcia P_z jako funkcji prądu stojana przy $f = 50$ Hz ze wzoru

$$R'_{2(50)} = \frac{P_z - 3I_{1z}^2 R_1}{3I_{2z}^2}, \tag{11}$$

przy czym $I_{1z} \approx I'_{2z}$.

Na podstawie zależności (10) i (11) określa się współczynnik a wyrażający wzrost oporności uzwojenia wirnika na skutek wypierania

$$a = \frac{R'_{2(50)}}{R'_{2(0)}}. \tag{12}$$

Wyznaczenie wartości $R'_{2(0)}$ ze wzoru (10) wymaga obciążenia maszyny oraz znajomości strat w żelazie, ponieważ

$$P_{12} = P_1 - P_{\dot{z}} - P_{Cu1}.$$

Pomiar obciążenia średnich i dużych maszyn jest jednak trudny do przeprowadzenia, a w wytwórniach maszyn, zazwyczaj niewykonalny. Inną niedogodnością jest jeszcze i to, że przy obciążeniu maszyna nagrzewa się i nie można określić wówczas temperatury, przy jakiej mierzone są oporności R_1, $R_2(0)$ oraz $R_2(50)$, co pociąga za sobą błędne ich wyznaczanie. Obciążenie zaś mocą częściową daje bardzo mały poślizg, powstają trudności dokładnego określenia P_{12}, a oprócz tego występuje wtedy większy wpływ prądu magnesującego, przez co trudniej jest wyznaczyć dokładnie prąd I'_2.

Z przyczyn wyżej podanych metoda ta nadaje się do badania wypierania w maszynach mniejszych mocy.

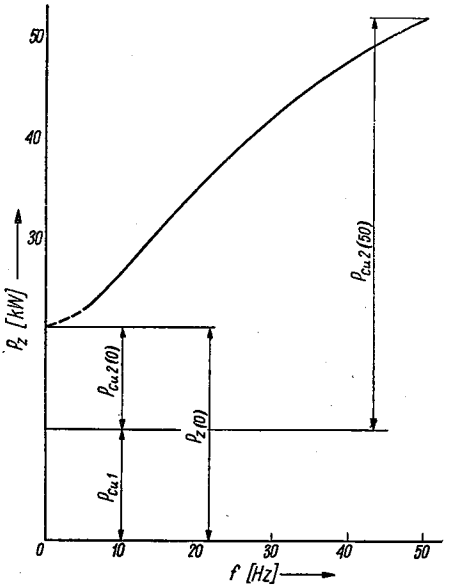
5. METODA POMIAROWA ZAPROPONOWANA PRZEZ AUTORÓW

5.1. Wyznaczenie współczynników a oraz k_2

Podczas pomiarów badana maszyna jest zahamowana. Dokonuje się pomiaru mocy zwarcia P_z w funkcji prądu stojana I_{1z} dla różnych war-

tości częstotliwości w granicach $f = 5 \div 50$ Hz. Pomiar przeprowadza się prądem małym o dowolnej wartości. W ten sposób eliminuje się z układu pomiarowego transformatory miernicze, które dają przy niskich wartościach f znaczne błędy pomiarowe. Ponadto unika się nagrzania maszyny, co jest również bardzo ważne ze względu na dokładność otrzymanych wyników.

Uzyskane z pomiarów zależności przedstawia się graficznie jako rodzinę krzywych $P_z = \varphi(I_{1z})$, dla różnych wartości f=const, i na pod-



Rys. 8. Zależność mocy zwarcia od częstotliwości $P_z = \psi(f)$ przy $I_{zn} = \text{const}$ dla silnika 1000 kW, $n = 1000$ obr/min

stawie tych charakterystyk sporządza się krzywą $P_z = \Psi(f)$ dla dowolnego $I_{1z} = \text{const}$, rys. 8. Ekstrapolując tę krzywą do $f = 0$ (linia przerywana) otrzymamy odcinek wyrażający straty $P_z(0)$ odpowiadające prądowi stałemu [6].

Z wykresu $P_z = \Psi(f)$ dla $I_{1z} = \text{const}$, a więc i dla $I'_{2z} = \text{const}$ można wyznaczyć straty miedzi wirnika z następujących zależności:

$$P_{Cu\,2\,(0)} = P_{z\,(0)} - P_{Cu\,1\,(0)} = P_{z\,(0)} - 3\,I_{1z}^2 R_1, \tag{13}$$

$$P_{Cu\,2\,(50)} = P_{z\,(50)} - P_{Cu\,1\,(50)} = P_{z\,(50)} - 3\,I_{1z}^2 \cdot k_1 \cdot R_1, \tag{14}$$

gdzie:

P_{Cu1}, P_{Cu2} — straty w miedzi stojana i wirnika przy odpowiednich częstotliwościach,

R_1 — oporność uzwojenia stojana pomierzona prądem stałym,

k_1 — współczynnik wypierania prądu w uzwojeniu stojana.

Ze względu na to, że pomiar jest przeprowadzany małym prądem, maszyna ma temperaturę otoczenia i dla niej wyznacza się oporność stojana R_1.

Zazwyczaj przyjmuje się, że wypierania prądu w stojanie nie ma, wtedy $k_1 = 1$. Nie jest to jednak całkowicie słuszne i w dalszym ciągu będzie omówiony wpływ tego uproszczenia na wynik.

Z uwagi na to, że dla $P_z = \Psi(f)$ prąd wirnika jest stały $I'_{2z} \approx \text{const}$, stosunek więc pomierzonych mocy wirnika jest równy stosunkowi oporności

$$a = \frac{P_{Cu\,2\,(50)}}{P_{Cu\,2\,(0)}} = \frac{R'_{2\,(50)}}{R'_{2\,(0)}}, \tag{15}$$

a zatem współczynnik wzrostu oporności wirnika a wyznacza się ze stosunku mocy.

Na podstawie tych samych pomiarów można również wyznaczyć współczynniki k_2 dla części uzwojenia umieszczonych w żelazie, tak jak się to wyznacza przy pomocy metod rachunkowych. W tym celu należy jedynie dodatkowo obliczyć oporność R'_{2p} i wyliczyć stosunek

$$c = \frac{R'_{2p}}{R'_{2(0)}} = \frac{R'_{2p} \cdot 3 I'^2_{2z}}{P_{Cu2\,(0)}} \approx \frac{R'_{2p} \cdot 3 I^2_{1z}}{P_{Cu2\,(0)}}, \tag{16}$$

a ponieważ

$$R'_{2\,(0)} = R'_{2p} + R'_{2n\,(0)},$$

$$R'_{2\,(50)} = R'_{2p} + R'_{2n\,(50)} = R'_{2p} + k_2 R'_{2n\,(0)},$$

po przekształceniach otrzymuje się

$$a = k_2(1 - c) + c$$

i następnie

$$k_2 = \frac{a - c}{1 - c}. \tag{17}$$

W podobny sposób wyznacza się współczynnik k_2 w punkcie 4.

5.2. Dokładność wyznaczenia współczynnika a

5.2.1. Wpływ wypierania prądu w uzwojeniu stojana

Wypieranie prądu pociąga za sobą wzrost strat dodatkowych i dlatego jest ono szkodliwe w stojanie, gdzie częstotliwość ma stałą wartość. W obliczeniach przyjmuje się na ogół, że w uzwojeniu stojana wypierania nie ma. W rzeczywistości jednak ono istnieje i w dobrze zaprojektowanym stojanie straty dodatkowe tym spowodowane mogą dochodzić do 33% strat w miedzi liczonych dla prądu stałego [1]. Te straty dodatkowe występują przede wszystkim w maszynach dużych i szybkobieżnych,

gdzie ilość prętów w żłobku jest mała, tak że przy zastosowaniu nawet możliwie dużej ilości gałęzi równoległych wysokość pręta przekracza wartość krytyczną.

Tablica 1

Moc kW	Obroty synchron.	$\frac{P_{z(50)}}{P_{z(0)}}$	Uwagi
1250	3000	2,94	pręty trapezowe
1000	1000	2,45	„ prostokątne
800	375	1,28	wirnik pierścieniowy, pręty prostokątne
570	750	2,88	dwuklatkowy
800/400	1000/500	2,72/2,52	dwuklatkowy

Ze względu na wartości współczynnika k_1 zawarte w granicach $1 \div 1{,}33$ otrzymamy wyrażenia na straty w miedzi

$$P_{Cu1(50)} = (1 \div 1{,}33) P_{Cu1(0)},$$
$$P_{Cu2(50)} = P_{z(50)} - (1 \div 1{,}33) P_{Cu1(0)}.$$

Przyjmując następnie $P_{Cu2(0)} \approx P_{Cu1(0)} = 0{,}5 P_{z(0)}$ oraz, że $\frac{P_{z(50)}}{P_{z(0)}} \approx 3$ (wartość średnia wyników uzyskanych z pomiarów dużych maszyn, tabl. 1), otrzymuje się

$$P_{Cu2(50)} \approx 3 P_{z(0)} - (1 \div 1{,}33) P_{Cu1} = 3 P_{z(0)} - 0{,}5 (1 \div 1{,}33) P_{z(0)} \approx$$
$$\approx (2{,}5 \div 2{,}34) P_{z(0)}.$$

Z powyższego wynika, że przy założeniu, iż w stojanie nie ma wypierania, można w granicznym przypadku popełnić błąd osiągający wartość

$$\frac{2{,}5 - 2{,}34}{2{,}34} 100 \approx + 7\ \%,$$

a więc jest tendencja wyznaczenia za dużej wartości $P_{Cu2(50)}$.

5.2.2. *Wpływ strat w żelazie*

Przy pomiarze prądem znamionowym (silnik zahamowany) napięcie zwarcia wynosi $U_z = 20 \div 30\%\ U_n$, zaś *SEM* stojana $E_z \approx 0{,}5\ U_z$, wobec tego straty w żelazie stojana przy zwarciu mają wartość

$$P_{\dot{z}1(50)} \approx \left(\frac{1}{100} \div \frac{1}{45}\right) P_{\dot{z}1n};$$

w rzeczywistości będą one jeszcze mniejsze, bo nie występują straty dodatkowe w żelazie.

Przy zahamowanej maszynie straty w żelazie wirnika są mniej więcej równe stratom w stojanie

$$P_{ż(50)} = P_{ż1(50)} + P_{ż2(50)} \approx \left(\frac{1}{50} \div \frac{1}{22}\right) P_{ż1n},$$

a ponieważ

$$P_{ż1n} \approx 3/5\,(P_{Cu1(0)} + P_{Cu2(0)}) \approx 3/5\,P_{z(0)},$$

oraz

$$P_{z(50)} \approx 3\,P_{z(0)},$$

straty w żelazie przy zwarciu i $f = 50$ Hz wynoszą

$$P_{ż(50)} \approx (0{,}5 \div 1{,}0)\%\, P_{z(50)}.$$

Nie uwzględniając strat w żelazie wyznacza się straty przy zwarciu $P_{z(50)}$, a tym samym i straty w miedzi wirnika $P_{Cu2(50)}$ o (0,5 ÷ 1,0)% za duże.

5.2.3. Wpływ indukcyjności X_5' i X_0 (rozproszenia wirnika i pola głównego)

Wartość pomierzona oporności wirnika jest wyznaczona wzorem

$$R'_{2pm} = R_{pm} - R_{1pm} = \frac{R'_2}{\left(\frac{R'_2}{X_0}\right)^2 + \left(1 + \frac{X'_2}{X_0}\right)^2},$$

gdzie R'_2 jest wartością rzeczywistą oporności czynnej wirnika, a indeksy *pm* odnoszą się do wartości pomierzonych.

Z powyższego równania widać, że wartość pomierzona R'_{pm} jest mniejsza od wartości rzeczywistej R'_2. Jak wykazały pomiary [6], błąd ten przy $f = 50$ Hz jest ujemny i leży w granicach (5÷10)%, a co do wartości jest mniej więcej równy sumie uchybów spowodowanych nieuwzględnieniem wypierania w stojanie (p. 5.2.1) i strat w żelazie (p. 5.2.2.); uchyby te kompensują się więc i w pomiarach można je pomijać. Wobec powyższego, błąd przy wyznaczaniu współczynnika *a*, sprowadza się praktycznie do błędów przyrządów pomiarowych przy wyznaczaniu mocy P_z i wynosi przy starannym pomiarze $\Delta a \approx \pm 3\%$.

5.3 Dokładność wyznaczenia współczynnika k_2

5.3.1. Wpływ nieuwzględniania zmian współczynnika c

Założenie, że współczynnik c = const, tzn. że wypieranie nie występuje w częściach prętów znajdujących się w wentylacyjnych kanałach

promieniowych i pierścieniach zwierających, prowadzi do błędu, ponieważ w tych elementach również to zjawisko zachodzi. Według [1] wzrost oporności w tych częściach przy $f = 50$ Hz może dojść do 20% i wobec tego dla oporności $R_{2p(50)}$ będzie słuszna zależność

$$R_{2p(50)} = yR_{2p(0)} \approx (1 \div 1{,}2) R_{2p(0)}.$$

Uwzględniając powyższe, wyrażenie (17) osiągnie wartość

$$k_{2y} = \frac{a - yc}{1 - c},$$

a więc otrzymamy wynik za duży ($k_2 > k_{2y}$) i błąd względny stąd wynikający wyniesie

$$\Delta' k_2\% = \frac{k_2 - k_{2y}}{k_{2y}} 100 = \left(\frac{a - c}{a - yc} - 1\right) 100\%.$$

Z danych konstrukcyjnych otrzymano średnio dla szybkobieżnych maszyn z prętami trapezowymi $c \approx 0{,}51$, a dla maszyn z prętami prostokątnymi $c \approx 0{,}37$.

Należy tu jeszcze zaznaczyć, że Syromiatnikow [4] przyjmuje do swoich obliczeń średnio $R_{2p(0)} \approx 30\%\ R_{2(0)}$, a $R_{2n(0)} \approx 70\%\ R_{2(0)}$, tak że wtedy $c \approx 0{,}3$.

Przyjmując średnią wartość $a \approx 4{,}0$ oraz $y = 1{,}2$ otrzymamy graniczne wartości uchybów $\Delta' k_2$ rzędu $(2 \div 3)\%$. Jak widać, uchyb ten jest nieznaczny i maleje ze wzrostem a oraz zmniejszaniem się c.

5.3.2. Wpływ błędnego określenia wielkości c

Wartość c określa się częściowo z pomiaru, a częściowo z obliczeń (R'_{2p}). Chociaż obliczenie wartości R'_{2p} jest stosunkowo proste dla $f = 0$, to jednak występują tu takie czynniki, jak oporność lutu (łączącego pręty z pierścieniami zwierającymi), dokładność zachowania wymiarów, oporności właściwej materiałów itp., które powodują, że wartości obliczone mogą się znacznie różnić od rzeczywistych.

Uchyb względny k_2 spowodowany błędnym określeniem wielkości $R'_{2p(0)}$ i związanej z tym wielkości c wynosi

$$\Delta'' k_2\% = \pm \frac{\Delta c\,(a - 1)}{(a - c)(1 - c)} 100\%,$$

gdzie Δc — uchyb wartości bezwzględnej.

Przyjmując, że uchyb Δc może osiągnąć wartość $\Delta c \leqslant \pm 20\%c$, uchyb graniczny względny wyniesie dla $a = 4$, $c = 0{,}5$, $\Delta'' k_2 \approx \pm 17\%$,

$$\text{dla } a = 4, \quad c = 0{,}3, \quad \Delta'' k_2 \approx \pm 7\,\%$$

Jak z powyższego przeliczenia widać, uchyb ten zależy przede wszystkim od wartości c.

5.3.3. *Wnioski*

Reasumując, na dokładność wyznaczenia współczynnika k_2 wpływają uchyby Δa, $\Delta' k_2$ oraz $\Delta'' k_2$; dwa pierwsze, jak to wykazano poprzednio, są stosunkowo niewielkie w porównaniu z $\Delta'' k_2$, spowodowanym niedokładnym wyznaczeniem $R_{2p(0)}$. Błąd ten jest nie do uniknięcia i to zarówno w metodzie rachunkowej, jak też wykreślnej i pomiarowej.

6. ANALIZA WYNIKÓW OBLICZEŃ I POMIARÓW

6.1. *Wyniki obliczeń i pomiarów*

W celu przeprowadzenia analizy i wyciągnięcia wniosków przedstawione zostały w tabl. 2 wyniki obliczeń i pomiarów wykonanych w D.Z.W.M.El. M-5 na różnego typu dużych maszynach asynchronicznych.

Tablica 2

Moc kW	Obroty synchr.	Współczynnik wzrostu oporności a				Rodzaj uzwojenia wirnika
		Field	Nacke	Syromiat-nikow	Autorzy	
		I	II	III	IV	
1250	3000	3,98	3,34	4,36	4,38	pręty trapezowe $u = \frac{3}{9,5}$, $h = 55$ mm
1000	1000	3,66	3,21	3.94	4,04	pręty prostokątne 4×55 mm
800	375	1,38	1,27	1,44	1,47	pierścieniowy wym. pręta 3×12 mm, na szer. żłobka 2 pręty na wys. żłobka 2 pręty
dwubiegowy 800/400	1000/500	3,61/4,55	—	4,10/5,03	4,25/5,11	dwuklatkowy
570	750	2,87	—	3,24	3,30	dwuklatkowy

Metodami rachunkową i wykreślną zostały wyznaczone współczynniki wypierania a, aby można było je porównać z wynikami uzyskanymi z pomiarów.

6.2. Porównanie wyników

Jak z zestawienia (tabl. 2) widać, wyniki uzyskane metodą pomiarową autorów są większe od wyników uzyskanych metodą Syromiatnikowa od 0,5 ÷ 3,5%, różnice są więc niewielkie.

Wyniki uzyskane metodą wykreślną są niższe od wyników otrzymanych z obliczeń od 6 ÷ 16% w stosunku do wyników obliczonych. Metodą wykreślną uzyskuje się tym dokładniejsze wyniki, im na większą liczbę elementów podzieli się pręt. Przytoczone wyniki otrzymano przy podziale na $n = 6$ i stąd przy dużych zredukowanych wysokościach pręta trapezowego różnica wynosi 16%, przy pręcie prostokątnym, gdy wysokość jest znacznie mniejsza 12%, zaś przy prętach uzwojenia fazowego już tylko 6,5%.

Porównując wyniki obliczone (kolumna I) z pomiarowymi (kolumna IV) widać, że wartości obliczone dla prętów trapezowych i prostokątnych są niższe o około 9% (w stosunku do wartości pomierzonych), dla uzwojenia fazowego o około 6%. Największe różnice bo 11 ÷ 15% występują dla uzwojeń dwuklatkowych, wynika to z trudności samych obliczeń.

Różnice w porównywanych wynikach spowodowane przez czynniki technologiczno-montażowe (zwiększenie oporności na skutek lutowań, zwarcie blach w sąsiedztwie prętów i stosowanie dodatkowych elementów mocujących umieszczonych pod prętami trapezowymi) i założenia upraszczające (przyjmuje się, że w częściach prętów w kanałach promieniowych i pierścieniach zwierających nie ma wypierania prądu) są stosunkowo niewielkie i ujemne; jest to o tyle korzystne, że daje pewien zapas momentu rozruchowego.

7. CHARAKTERYSTYKA METOD

Rachunkowe metody wyznaczania współczynników a lub k_2 są rozszerzeniem i uzupełnieniem metody opracowanej przez Fielde'a[1]) [1], [2], [3]. Wprawdzie przy wyprowadzaniu matematycznych wyrażeń przyjęte są pewne uproszczenia omówione już poprzednio, to jak widać z wyników pomiarów, różnice te są niewielkie. Metoda rachunkowa jest niezastąpiona ze względu na stosunkową prostotę obliczeń i możliwości obliczenia wypierania prądu dla różnych kształtów przekrojów prętów. Pewnym mankamentem jest to, że obliczenia ujmują tylko zjawiska od strony ilościowej, nie dając obrazu rozkładu gęstości prądu w pręcie, co utrudnia zrozumienie fizykalnego sensu zjawiska.

[1]) Trans. Amer. Inst. Electr. Engrs., 1905.

Metoda wykreślna [5] daje doskonały pogląd na zjawiska zachodzące w pręcie, jednak posługiwanie się nią w normalnej pracy obliczeniowej jest trudne ze względu na dużą pracochłonność. Przy dzieleniu pręta na pewną skończoną liczbę elementów otrzymuje się wyniki niższe niż metodę Fielde'a. Jest to skutek założenia stałości strumienia na wysokości elementarnego pręta. Im na większą liczbę elementów dzieli się wysokość, tym wynik jest dokładniejszy, lecz również wzrasta trudność wykonania wykresu, szczególnie występuje to przy prętach trapezowych.

Metody pomiarowe są cennym narzędziem dla sprawdzenia wyników obliczeń i określenia tych wszystkich czynników, których w metodach matematyczno-wykreślnych uwzględnić nie można. Taka konfrontacja obliczeń z pomiarami może doprowadzić do statystycznego ujęcia całokształtu czynników ubocznych. Ma to szczególne znaczenie w tych wszystkich przypadkach, gdy kształty prętów są skomplikowane, oraz w silnikach dwuklatkowych. Na podstawie wyników można wprowadzić uproszczenia we wzorach obliczeniowych.

Metoda Syromiatnikowa [4], jak to już było podane poprzednio, ze względu na potrzebę obciążenia maszyny nadaje się raczej dla maszyn mniejszych mocy. Można również potrzebne dane uzyskać bez obciążenia, z wykresu kołowego, wówczas jednak metoda nabiera cech metody obliczeniowej. Określenie współczynnika wzrostu oporności jest tu stosunkowo proste dlatego, że wszystkie dane potrzebne do określenia a, są wyznaczone w ramach normalnych badań objętych normami.

Metoda autorów jest prosta, nadaje się do pomiaru maszyn wszelkich mocy i daje dużą dokładność. Możliwość stosowania do pomiaru małych prądów ogranicza moc źródła do niewielkiej wartości. Zasadniczym warunkiem uzyskania dużej dokładności wyników jest dokładne zdjęcie charakterystyki $P_z = \psi(f)$ przy $I_1 = I_2' = \text{const}$ oraz określenie oporności R_1, $R_{2(0)}$, $R_{2(50)}$ w tej samej temperaturze, co nie przedstawia tu żadnych trudności. Na podstawie otrzymanych wyników można stwierdzić, że metoda [4] i metoda autorów mają prawie ten sam stopień dokładności wyznaczenia współczynnika a.

Politechnika Wrocławska
Katedra Maszyn Elektrycznych, Katedra Pomiarów Elektrycznych

WYKAZ LITERATURY

1. Schuisky W.: Induktionsmaschinen. Springer-Verlag, Wien 1957.
2. Richter R.: Elektrische Maschinen, B. II, Springer-Verlag, Berlin 1930.
3. Schuisky W.: Stromverdrängungsmotoren. — Archiv f. Elektrotechnik, 27, 1933.

4. Syromiatnikow I. A.: Praca silników asynschronicznych. PWT, Warszawa 1953.
5. Nacke H.: Ersatzschaltung zur Ermittlung der Stromverdrängung in den Stäben von Wirbelstromläufermotoren. — ETZ, Nr 22, 1954.
6. Andrzejewski F., Orzeszkowski Z.: Nowa metoda wyznaczania strat w uzwojeniach wirników maszyn asynchronicznych. Rozprawy Elektrotechniczne, Nr 4, 1960.

Ф. АНДРЖЕЕВСКИ, З. ОРЖЕШКОВСКИ

ИЗМЕРИТЕЛЬНЫЙ МЕТОД ОПРЕДЕЛЕНИЯ КОЭФФИЦИЕНТА ВЫТЕСНЕНИЯ В РОТОРАХ АСИНХРОННЫХ ДВИГАТЕЛЕЙ

Резюме

В статье описаны расчётные, графические а также измерительные методы определения коэффициента приращения активного сопротивления, вызванного вытеснением.

Расчётный метод основан на работах Фильда, развитых также другими авторами [1], [2], [3]. Указаныний метод разработан для стержней с найболее часто встречающимися в производстве крупных электрических машин, профилями, т. е. с круглым, прямоугольным или же трапецоидальным сечением [формулы, (1), (2), (3), (4), (5), (6)].

Графический метод дан Наке [5]. Этот метод имеет прежде всего большое показательное значение и он даёт определённый взгляд на физические явления, возникающие в стержне. Формулы (7) и (8) определяют токи в отдельных элементах стержня (рис. 4). На рисунке 5 представлена эквивалентная схема для отдельных элементов, на которое разделен стержень. Из формул разработанных для отдельных частей схемы можно составить векторную диаграмму токов (рис. 6) и напряжений (рис. 7). Формула (9) определяет нужный коэффициент k_2, который, так как и в расчетном методе, выражает отношение сопротивлений при $f = 50$ гц и $f = 0$ гц, для тех частей стержней у которых выступает явление вытеснения.

Метод предложенный Сыромятниковым [4] является измерительным методом и здесь коэффициент вытеснения представляет отношение всего сопротивления ротора при $f = 50$ гц и $f = 0$ гц.

Сопротивление при постоянном токе $R'_{2(0)}$ ($f = 0$) можно определить измеряя нагрузку [форм. (10)], а сопротивление при $f = 50$ гц на основании измерения мощности короткого замыкания [форм. (11)]. Коэффициент увеличения сопротивления в виде отношения обоих вышеуказанных нами сопротивлений, представлен формулой (12).

Метод измерения коэффициента вытеснения разработанный авторами основан на измерении мощности короткого замыкания машины в зависимости от частоты при постоянном значении тока ротора; значение тока является произвольной величиной и поэтому может быть на столько мало, что не вызовет нагрева обмотки. Из графика $P_z = \Psi(f)$ для $I_2 = \text{const}$ (рис. 8) экстраполяцией определяется мощность $P_{z(0)}$ а для $f = 50$ гц мощность $P_{z(50)}$.

Из форм. (13) и (14) получаем коэффициент вытеснения a [форм. (15)]. Можно также подобным образом как по расчётному или графическому методу, определить коэффициент k_2 из выражения (17).

В пунктах 5.2 и 5.3 рассмотрены факторы имеющие влияние на точность определения коэффициентов a и k_2.

Из анализа и учёта всех факторов следует, что погрешность составляет $\Delta a = \pm 3\%$, что по видимому является незначительной величиной.

В пункте 6 дан анализ результатов полученных на основании расчётов и измерений, для крупных асинхронных двигателей.

Метод предложенный авторами отличается простотой выполнения и большой точностью, он пригоден при измерениях двигателей любых мощностей. Возможность произведения измерений с незначительными токами позволяет применить источники к небольшой мощностью.

Основным условием получения высокой точности измерения является снятие характеристики $P_z = \Psi(f)$ при $I_1 = I_2 = \text{const}$, а также определение сопротивления $R_1, R_{2(0)}, R_{2(50)}$.

В таблице 2 сведены результаты по расчётному [1], [2], [3], графическому [5] и измерительному методу [4] и по методу авторов. Отклонение результатов обоих измерительных методов мало и не превышает 4%.

Результаты по графическому методу (колонка II, таблица 2) ниже на около $6 \div 16\%$ по сравнению с расчетными данными (колонка I, таблица 2). Отклонения результатов будут тем меньше, чем больше число элементарных полос на какие разделен стержень.

Расчётные данные (колонка I) ниже чем результаты измерений (колонка IV) в среднем на $9 \div 15\%$.

При сравнении результатов видно, что технологически-монтажные факторы, включая упрощающие положения, дают небольшую погрешность с отрицательным знаком.

F. ANDRZEJEWSKI, Z. ORZESZKOWSKI

DIE MESSMETHODE ZUR BESTIMMUNG DES WIDERSTANDSFAKTORS IN DEN LÄUFERN VON DREHSTROMMOTOREN

Zusammenfassung

Im Artikel sind die mathematischen, graphischen und messtechnischen Verfahren zur Bestimmung der Erhöhung des Wirkwiderstandes infolge der Stromverdrängung angegeben.

Die von Field angegebenen und von anderen Verfassern erweiterten Formeln [1], [2], [3] sind für die in der Praxis meistens verwendeten Rund-, Hoch- und Trapezstäbe kurz erfasst. Gl. (1), (2), (3), (4), (5), (6).

Das von Nacke angegebenes graphisches Verfahren gibt einen sicheren Einblick in die phisikalischen Vorgänge bei der Stromverdrängung. Die Gleichungen (7), und (8) bestimmen die Teilströme. Abb. 5 stellt die Ersatzschaltung für den Ersatzleiter dar. Gemäss der Gleichungen für die Teilströme lässt sich das Zeigerdiagramm der Ströme (Abb. 6) und Spannungen (Abb. 7) aufstellen.

Auf Grund der Gleichung (9) bestimmt man, für den im Flussdurchtrittsraum liegenden Leiter, den gesuchten Wirkfaktor k_2, der ähnlich wie in den Rechnungsverfahren das Verhältnis des Wirkwiderstandes bei $f = 50$ Hz und $f = 0$ Hz darstellt.

In der Messmethode von Syromiatnikow [4] entspricht der Faktor a Gl. (12) dem Verhältnis des ganzen Wirkwiderstandes des Rotors bei $f = 50$ Hz und bei $f = 0$ Hz.

Der Wirkwiderstand ohne Stromverdrängung d.h. bei Gleichstrom ist aus der Belastungsprobe Gl. (10) und der Wirkwiderstand bei $f = 50$ Hz aus der Kurzschlussprobe Gl. (11) bestimmt.

Die von Verfassern angegebene Methode beruht auf der Messung der beim Kurzschlussversuch aufgenommenen Leistung als Funktion der Frequenz beim konstanten Läuferstrom. Da die Messung auch bei kleinem Strom durchgeführt werden kann, erwärmt sich die Maschine während des Versuches nicht.

Aus der Kurve $P_z = \psi(f)$ bei $I_2 = \text{konst}$ (Abb. 8) bestimmt man die Leistung $P_{z(50)}$ für $f = 50$ Hz und durch die Extrapolation $P_{z(0)}$ für $f = 0$ Hz. Mit Hilfe von Gleichungen (13), (14) berechnet man den Faktor a. Man kann auch, so wie in dem mathematischen und graphischen Verfahren den Wirkfaktor k_2 bestimmen. Gl. (17).

In Abschnitten 5.2 und 5.3 ist der Einfluss von verschiedenen Parametern auf die Genauigkeit der Messung der Faktoren a und k_2 besprochen. Die Analyse der Messgenauigkeit ergab, dass der wahrscheinliche Fehler Δa ungefähr $\pm$ 3% beträgt.

Im Abschnitt 6 sind die Mess- und Berechnungsergebnisse für grosse Drehstrommotoren angegeben.

Die von Verfassern vorgeschlagene Messmethode ist einfach und genau. Da die Kurzschlussströme während des Versuches klein sind, braucht man keinen grossen Hilfsgenerator.

Die Genauigkeit der Methode ist im ersten Falle von der Genauigkeit der Messung der Funktion $P_z = \psi(f)$ bei $I_1 = I_2 = \text{konst.}$ und der Widerstände R_1, $R_{2(0)}$, $R_{2(50)}$ abhänging.

In der Tabelle 2 sind die Engebnisse aus dem mathematischen Verfahren [1]. [2], [3], graphischen Verfahren von Nacke [5], Messverfahren von Syromiatnikow [4] und der von Verfassern angegebenen Messmethode zusammengestellt.

Die Ergebnisse der beiden Messmethoden sind sehr ähnlich und unterscheiden sich nicht mehr als 4%.

Die Ergebnisse des graphischen Verfahrens (Spalte II, Tabelle 2) sind durch schnittlich 6 ÷ 16% kleiner, als die Ergebnisse aus dem mathematischen Verfahren (Spalte I, Tabelle 2). Bei grösserer Zahl von Teilleiter sind die Unterschiede kleiner.

Die Ergebnisse des mathematischen Verfahrens (Spalte I) sind durchschnittlich 9 ÷ 15% kleiner als die Messergebnisse (Spalte IV).

534.6:534.7

WIKTOR JASSEM

Wstępna analiza spektrograficzna głosek polskich

Rękopis dostarczono 1.3.1958

Praca ma na celu dokonanie pierwszej próby scharakteryzowania sygnałów akustycznych odpowiadających głoskom polskiego języka przy pomocy spektrografu. Rozróżniono trzy typy sygnałów informacyjnych: quasiperiodyczne, szumy oraz sygnały mieszane stanowiące superpozycję szumów na sygnały quasiperiodyczne. W identyfikacji poszczególnych głosek szczególnie ważną rolę odgrywają zakresy wzmocnienia w widmie. Analiza trójwymiarowa, dokonana w czasie, częstotliwości i intensywności pozwala na segmentację, tj. oddzielenie w sygnale zmiennym w czasie pod względem rozkładu energii w widmie charakterystycznych odcinków odpowiadających poszczególnym głoskom. Samogłoski charakteryzują się dwoma względnie trzema formantami w tej części przebiegu, gdzie zmiany w widmie energetycznym są najwolniejsze. Stanowią one przebiegi quasiperiodyczne. Podobne przebiegi odpowiadają spółgłoskom płynnym, z tym jednak, że dla ich identyfikacji ważne są nie tylko te części przebiegu, które są względnie ustalone w swym widmie energetycznym, lecz także te, w których to widmo ulega wyraźnej zmianie w czasie. Te nieustalone przebiegi, zwane transientami, są również istotne dla identyfikacji spółgłosek zwartych oraz tzw. samogłosek niesylabicznych. Spółgłoski trące są bądź szumami, bądź superpozycją szumu na przebieg quasiperiodyczny, przy czym przebiegi akustyczne odpowiadające im są względnie ustalone. Spółgłoski zwarto-trące stanowią bezpośrednio po sobie następujące kombinacje przebiegów analogicznych do tych, które charakteryzują spółgłoski zwarte i trące.

Przedstawiona analiza odnosi się do głosu jednej osoby.

Zrozumienie żywej mowy zależne jest od spełnienia przez centralny układ nerwowy osoby słuchającej dwóch podstawowych funkcji: a) zidentyfikowania zasłyszanych sygnałów akustycznych oraz b) powiązania w ośrodku mowy zidentyfikowanych sygnałów z określonymi treściami (znaczeniami) według istniejącej w danej społeczności językowej konwencji.

Ostatnie badania w zakresie elektroakustycznej analizy mowy (prowadzone szczególnie intensywnie w Stanach Zjednoczonych) pozwalają przypuszczać, iż istnieje możliwość skonstruowania aparatury, która naślado-

wałaby pierwszą z wymienionych wyżej czynności mózgowych[1]). Elektronowe maszyny do tłumaczenia w pewnej mierze spełniają funkcję zbliżoną do drugiej z wymienionych czynności.[2])

Automatyczna identyfikacja sygnałów akustycznych mowy miałaby doniosłe znaczenie dla techniki łączności. Pozwalałaby ona zamienić bardzo skomplikowane drgania powstające w procesie mówienia na proste sygnały, na przykład sygnały o stałej częstotliwości zróżnicowane tylko długością trwania. Telefonia wielokrotna uzyskałaby możliwość przeprowadzania na torze przenoszącym niewielkie pasmo znacznej ilości rozmów. Przy tym częstotliwości sygnałów mogłyby być dość niskie (rzędu kilkuset Hz), albowiem elementy znaczące (którymi w mowie są głoski, tak jak w piśmie litery) następują po sobie w przeciętnie szybkiej mowie z częstotliwością rzędu kilkunastu na sekundę. W urządzeniu odbiorczym sygnały przenoszone mogłyby być bądź zamienione z powrotem na mowę, bądź na łatwo i szybko czytelne znaki.

Innym cennym zastosowaniem aparatury identyfikującej sygnały akustyczne mowy byłyby maszyny samopiszące (tj. piszące pod dyktando bez pośrednictwa stenotypisty). Konstrukcje takich maszyn są w toku[3]).

Identyfikacja akustycznych sygnałów mowy opiera się na ich analizie w trzech wymiarach: czasie, amplitudzie i częstotliwości[4]). Analizę taką wykonać można przy pomocy spektrografu akustycznego, przy czym najlepszym typem do tego rodzaju badań okazał się tzw. sonagraf produkowany przez Kay Electric Co. Zasady działania i szczegóły konstrukcyjne dotyczące sonagrafu znaleźć można w cyklu artykułów opublikowanych w Journal of the Acoustical Society of America vol. 22, 1946. Sonagraf, przy pomocy którego wykonano referowane w niniejszym artykule analizy, jest jednym z najnowszych modeli analizującym w sposób ciągły pasmo od 85 do 8000 Hz filtrem o szerokości 300 Hz lub 45 Hz.

---

[1]) J. Wiren & H. L. Stubbs, Electronic Binary Selection System for Phoneme Classification, JASA vol. 25, str. 1082 i nast., 1956.

[2]) W. N. Locke & A. D. Booth (red.) Machine Translation of Languages, New York 1955 (dzieło zbiorowe); P. L. Garvin, Machine Translation, Reports of the 8th International Congress of Linguists, str. 103 i nast., Oslo 1957; N. D. Andriejew: Maszinnyj pierewod i problema jazyka-posriednika, Woprosy jazykoznanija 5, 1957, str. 117 i nast.

[3]) D. B. Fry i P. Denes: Mechanical Speech Recognition, Communication Theory, str. 426 i nast., London 1953; D. B. Fry i P. Denes: Experiments in Mechanical Speech Recognition, Information Theory, 3rd London Symposium, str. 206 i nast., London 1956; H. F. Olson i H. Belar: Phonetic Typerwriter, JASA, vol. 28, 1956, str. 1072 i nast.

[4]) Ogólne zasady takiej analizy przedstawiłem z mgrem inż. J. Suwalskim w artykule „Analiza akustyczna polskich głosek szumiących", Przegląd Telekomunikacyjny nr 1, 1957, str. 5 i nast.

Zasadniczą trudność w konstrukcji aparatury identyfikującej sygnały akustyczne żywej mowy jest nie tyle ich złożoność, ile fakt, iż określony element funkcjonalny mowy, jakim jest dana głoska, jest realizowany akustycznie bardzo różnie. Jeżeli obierzemy sobie jakąś głoskę danego języka — przypuśćmy polskie **a** — i dokonamy analizy znacznej ilości przykładów występowania tej głoski w mowie, to przekonamy się, że w określonych parametrach, którymi trzeba operować, występuje dla tej głoski w jej licznych realizacjach poważny rozrzut. Zdarza się, że dany sygnał akustyczny określony wielkościami odnoszącymi się do odpowiednich parametrów w skali bezwzględnej w jednym wypadku musi być zidentyfikowany jako jedna głoska, a w innym jako inna głoska. Pierwszą próbę wyjaśnienia tego zjawiska dał M. Joos w pracy Acoustic Phonetics, Suppl. to Language vol. 24 no. 2, Baltimore 1948. Dalsze badania potwierdziły słuszność zasadniczych jego tez[1]). Okazuje się, że poszczególne wypadki, które identyfikuje się językowo jako określoną głoskę, nie stanowią zespołu statystycznie jednorodnego. Zróżnicowania pomiędzy nimi nie mają charakteru czysto przypadkowego.

Następujące czynniki różnicują w pierwszej linii sygnały akustyczne, które są identyfikowane językowo jako ta sama głoska:

1. Indywidualne cechy wymowy. Pierwszorzędne znaczenie ma tu skala głosu (bas, tenor, sopran itd.).
2. Sąsiedztwo innych sygnałów.

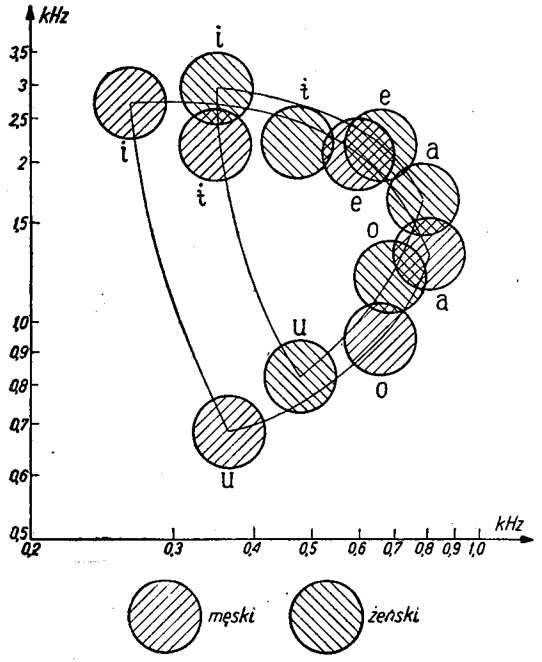
Jak indywidualne cechy głosu, w szczególności jego skala, wpływają na identyfikację głosek, wyjaśnimy posługując się tymczasem pewnym uproszczeniem, które istoty sprawy nie zniekształci, a pozwoli możliwie prosto je przedstawić.

W polskim języku posiadamy 6 samogłosek tzw. „ustnych" (w odróżnieniu od „nosowych", jak **ą**, **ę**). Są to: **i**, **y**, **e**, **a**, **o**, **u**. Przyjmijmy, iż możemy w sposób wystarczający opisać akustycznie te samogłoski posługując się dwiema wielkościami, które umieścimy na dwóch osiach prostokątnego płaskiego układu współrzędnych. Niechaj, zgodnie z istotnym stanem, wielkości te odnoszą się do częstotliwości. Jedna z tych wielkości odnosi się do tzw. formantu niższego (F_1), a druga do formantu wyższego (F_2). Wyznaczmy na naszym układzie punkty odpowiadające poszczególnym samogłoskom ustnym głosu męskiego oraz punkty odpowiadające tymże

[1]) Wymienić tu należy przede wszystkim: R. K. Potter i J. C. Steinberg: Toward the Specification of Speech, JASA vol. 22, 1950, str. 807 i nast.; G. E. Peterson: The Phonetic Value of Vowels, Language vol. 27, 1951, str. 541 i nast. (przed.: Bell Telephone System Monogr. nr 1982); G. E. Peterson i H. L. Barney: Control methods Used in a Study of Vowels, JASA, vol. 24, 1952, str. 175 i nast. (przedruk jw.); P. Ladefoged: Information Conveyed by Vowels, JASA, vol. 29, 1957, str. 98 i nast.

samogłoskom dla głosu żeńskiego wraz z polem rozrzutu. Otrzymamy obraz jak na rys. 1.

Jak widać z rysunku, może się zdarzyć, że pewnym wielkościom (F_1, F_2) odpowiada w głosie męskim **a**, a w żeńskim **o**, albo też w mę-



Rys. 1. Zależność formantów samogłoskowych od skali głosu

skim **e**, a w żeńskim **y**. Jest jednak rzeczą pierwszorzędnej wagi, że **system wzajemnych stosunków** pomiędzy głoskami pozostaje dla obu głosów taki sam. System samogłoskowy głosu żeńskiego jest **przesunięty** w obu współrzędnych względem głosu męskiego w kierunku wyższych częstotliwości.

Jest rzeczą ciekawą, że decyduje tu **nie wysokość tonu podstawowego**, lecz właśnie **skala głosu**. Jeżeli na przykład głos barytonowy na tonie 200 Hz (mniej więcej *as* w oktawie małej) wymówi lub zaśpiewa określoną samogłoskę i na tym samym tonie tę samą samogłoskę wymówi lub zaśpiewa mezzosopran, to formanty okażą się różne. Fakt ten tłumaczy się następująco: Każda samogłoska ma określoną artykulację. Język przyjmuje dla każdej z nich określone położenie niezależnie od częstotliwości drgań wiązadeł głosowych, tworząc komory rezonansowe o określonych częstoliwościach własnych. Przy regularnej budowie anatomicznej organów mowy na ogół osoby o większej budowie czaszki posiadają dłuższe wiązadła głosowe. Dłuższe wiązadła decydują o niższej skali głosowej,

a zarazem większe komory rezonansowe — o niższej częstotliwości rezonansu.

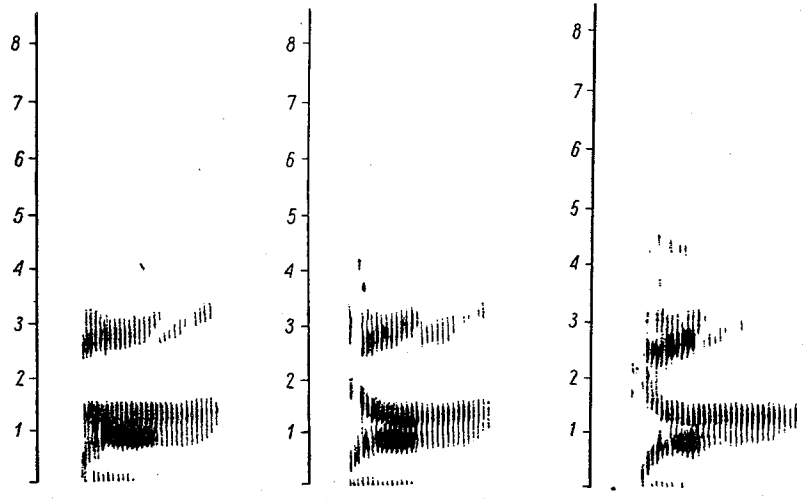
Wypływają stąd dwa ważne wnioski. Jeden natury technicznej, drugi — metodycznej:

1. Aparatura identyfikująca głoski musi reagować na skalę głosu mówiącego.

2. Przy analizie głosek jako sygnałów akustycznych nie można traktować każdej z nich oddzielnie, lecz każdą na tle systemu głosek w danym języku z jednej strony, a w związku z typem głosu — z drugiej. A więc analizę należy rozpocząć od badania możliwie całego systemu głoskowego jednej osoby (ewentualnie jednego typu głosu), a następnie stwierdzić przesunięcia znalezionych wielkości w innych głosach.

Trzeba dodać, że prawdopodobnie indywidualne zróżnicowania, poza skalą głosu, zależą jeszcze od innych czynników, których jednak dotychczas nie ustalono z dostateczną pewnością.

Z kolei wyjaśnimy, na czym polegają zróżnicowania zależne od **otoczenia** głoskowego, czyli od tego, jaki sygnał po danym sygnale następuje, względnie jaki go poprzedza. Przypuśćmy, że głoskę **a** analizujemy w takim wyrazie jak *bal*, a potem w takich wyrazach jak *dal* lub *garb*. Okaże się, że w pierwszej części (na osi czasowej) formanty samogłoski **a** będą różne w każdym z tych trzech wyrazów. W szczególności tzw. formant drugi F_2 będzie w początkowej fazie głoski **a** wyższy w wyrazie *dal* niż w wyrazie *bal*, a jeszcze wyższy w wyrazie *garb*; widać to na rys. 2 ukazującym spektrogramy sylab **ba, da, ga.**



Rys. 2. Spektrogramy sylab *ba, da, ga*

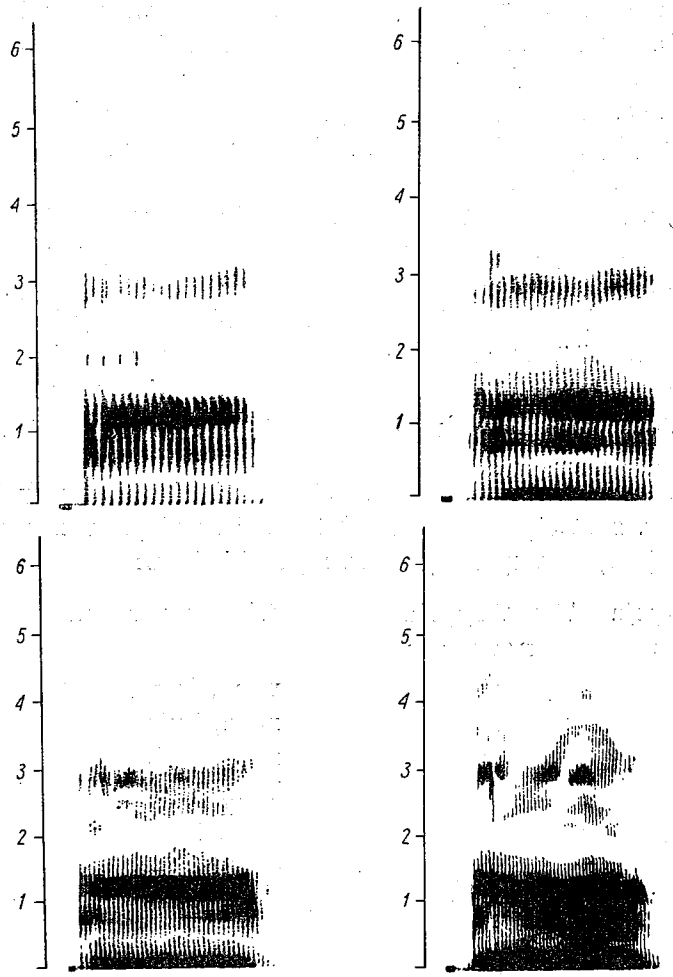
Niniejsza praca oparta jest na materiale ok. 400 obrazów spektrogra-

ficznych, na które składają się przeważnie całkowite wyrazy, a prócz tego pewna ilość odrębnych sylab i izolowanych głosek oraz krótkich wypowiedzi dwu lub trzywyrazowych. W obliczu faktu, iż materiał musiał być z konieczności dość ograniczony ilościowo ze względu na krótki okres czasu, w którym można było dysponować sonagrafem, trzeba było podjąć decyzję co do wyboru. Aby otrzymać po kilka do kilkunastu (a w trudniejszych przypadkach ponad 20) przykładów każdej głoski, trzeba było — zgodnie z powyższymi wywodami — ograniczyć się w zasadzie do wymowy jednej osoby, czyniąc pierwszy krok na drodze ustalenia cech polskich głosek w analizie trójwymiarowej. Ze względów natury zewnętrznej trzeba było oprzeć się na własnej wymowie, przedstawiającej przeciętny typ polszczyzny kulturalnej. Skala mojego głosu (bas-baryton) mieści się mniej więcej w granicach 70÷200 Hz. Dla orientacji kilkadziesiąt analiz wykonałem dla kilku dalszych osób. Jest rzeczą oczywistą, że badania należy kontynuować analizując różne głosy.

Przy pomocy sonagrafu można dokonać dwojakiego typu analizy widmowej. W obu przypadkach przebieg trwający nie dłużej niż 2,4 sekundy utrwalony jest magnetycznie i wielokrotnie odczytywany. Za każdym odczytem układ heterodynowy wybiera z sygnału wstęgę o szerokości 45 Hz lub 300 Hz, przy czym częstotliwość środkowa filtru przy każdym odczycie przesuwa się w skali częstotliwości o odstęp mniejszy niż szerokość wstęgi. Skala częstotliwości obejmuje zakres 85÷8000 Hz. W pierwszym przypadku analizie podlega cały utrwalony przebieg. Wynik takiej analizy, zapisany przy pomocy sonagrafu, posiada na osi odciętych czas, a na osi rzędnych częstotliwość. Napięcie wyższe od krytycznego powoduje powstanie iskry zaczerniającej. Wielkości amplitudy nie można z takiej analizy określić ilościowo. Można natomiast na zapisie prześledzić miejsca określone współrzędnymi czasu i częstotliwości, w których występują maksima amplitudy. W drugim przypadku wybiera się na utrwalonym sygnale określone miejsce w czasie. Podczas kolejnych odczytów zapisanego sygnału w oznaczonym punkcie czasowym włączony zostaje układ, który dokonuje analizy według całki fourierowskiej, której granice określa moment włączenia i stała czasu układu filtrującego. Otrzymujemy wykres, który wskazuje poziom spektralny amplitudy w funkcji częstotliwości.

Jedna grupa polskich głosek, do której należą samogłoski, przedstawia sygnały w przybliżeniu periodyczne. Występują one na spektrogramach wykonywanych przy pomocy filtru szerokowstęgowego (300 Hz) w formie pionowych prążków równoległych do osi częstotliwości (gdy częstotliwość tonu podstawowego jest mniejsza niż 300 Hz). Odstęp między prążkami jest wprost proporcjonalny do okresu tonu podstawowego. Na rys. 3 widoczne są spektrogramy samogłoski **a** wymówionej na czterech różnych

tonach, od niskiego do wysokiego. Warto zwrócić uwagę na to, że bez względu na wysokość tonu maksima energii występują w tych samych częstotliwościach, mianowicie w okolicy tonu podstawowego, a następnie



Rys. 3. Samogłoska a na czterech tonach

ok. 800, 1200 oraz 2800 Hz. Występowanie lub brak składowych harmonicznych stanowi podstawę zasadniczego podziału wszystkich polskich głosek na dwie grupy. Do pierwszej należą te, które wykazują składowe harmoniczne. Są to głoski **dźwięczne**. Głoski drugiej grupy — **bezdźwięczne** — takich składowych nie posiadają[1]).

[1]) W szepcie, którego jednak nie uważa się za normalną mowę, wszystkie głoski są bezdźwięczne, przy czym zróżnicowania pomiędzy niektórymi głoskami ulegają zatarciu.

Zanim przejdziemy do szczegółowego omówienia poszczególnych głosek i ich typów, przedstawimy **transkrypcję** fonetyczną, którą się w dalszym ciągu będziemy posługiwać. Zwyczajna ortografia polska (podobnie jak pisownia większości języków) nie jest z punktu widzenia fonetycznego całkiem konsekwentna. Na przykład w wyrazie *róża* litery *ó* oraz *ż* oznaczają te same głoski, które w wyrazie *burza* zapisujemy odpowiednio przez *u* oraz *rz*. A zatem określona głoska w zwyczajnej ortografii może być różnie zapisywana. Na odwrót, określonej literze mogą w różnych wyrazach odpowiadać różne głoski. Na przykład *kładka* doskonale rymuje się z *klatka; d* w pierwszym wyrazie wymawia się całkiem tak samo jak *t* w drugim. Poza tym, mimo ustalonej pisowni, określony wyraz może wymawiać się różnie w zależności od brzmienia wyrazu sąsiedniego. Na przykład w swobodnej, naturalnej wymowie na końcu wyrazu *brat* występuje głoska bezdźwięczna w zestawieniu *brat Stefana*, ale dźwięczna (taka jak *d* w *buda*) w zestawieniu *brat Zygmunta*. Otóż transkrypcja jest to ściśle fonetyczna pisownia reprezentująca wymowę (brzmienie) w sposób całkowicie konsekwentny.

W polskich pracach fonetycznych stosuje się dwa systemy transkrypcji, jeden tzw. slawistyczny (starszy) i drugi międzynarodowy (nowszy). Poniżej podajemy oba systemy wraz z przykładami:

Transkrypcja starsza	Transkrypcja nowsza	jak w wyrazach
i	i*	wino, igła
y	i*	ryba, byłem
e	e*	rzeka, chleb
a	a*	brat, staw
o	o*	skok, płot
u	u*	kruk, wróg
ǫ	ɔ̃*	wąs, wąż
ę	ɛ̃*	kęs, węch
i̯	j*	jodła, wiosło
u̯	w*	dłoń, pył [a]
p	p	pole, babka
t	t	tor, budka
k	k	kot, próg [b]
ḱ	c	kino, kiełbasa
b	b*	burza, kup drożdży [c]
d	d*	dom, brat Zygmunta [c]
g	g*	góra, brak wody [c]
ǵ	ɟ*	gil, giełda
f	f	fajka, krew [b]

s	s	sarna, wóz [b]
š	ʃ	koszyk, ryż [b]
ś	ɕ	sieć, świat, więź [b], siła
x	x	**ch**ata, herbata [d]
v	v*	**w**oda, wino
z	z*	ząb, sto**s** drewna [e]
ž	ʒ	rzeka, żaba, ma**sz** dwa złote [c]
ź	ʑ	źrebię, zima, **z**iemia, powieś go [c]
c	t͡s	**c**ena, schad**z**ka
č	t͡ʃ	**cz**ytać, brydż [b]
ć	t͡ɕ	sierść, **ci**chy, **ci**asny, żerdź [b]
ʒ	d͡z*	**dz**won, nic więcej [c]
ǯ	d͡ʒ*	drożdże, nie le**cz** go [c]
ʑ	d͡ʑ*	**dź**wignia, **dzi**siaj, **dzi**ura, choćby [c]
m	m*	**m**atka, gęba [e]
n	n*	**n**ora, są**d** [e]
ń	ɲ*	ko**ń**, **ni**ski, sa**ni**e, pięć [e]
ŋ	ŋ*	ban**k**, ręka [e]
r	r*	**r**yba, ka**r**a
ł	ɫ*	**ł**ydka, ko**ł**o

a) w wymowie potocznej, wyjąwszy północno-wschodnią część Polski

b) w osobno wymówionym wyrazie

c) w swobodnej, szybkiej wymowie

d) w wymowie okolic północno-wschodnich wyrazy pisane przez **h** mają inną głoskę

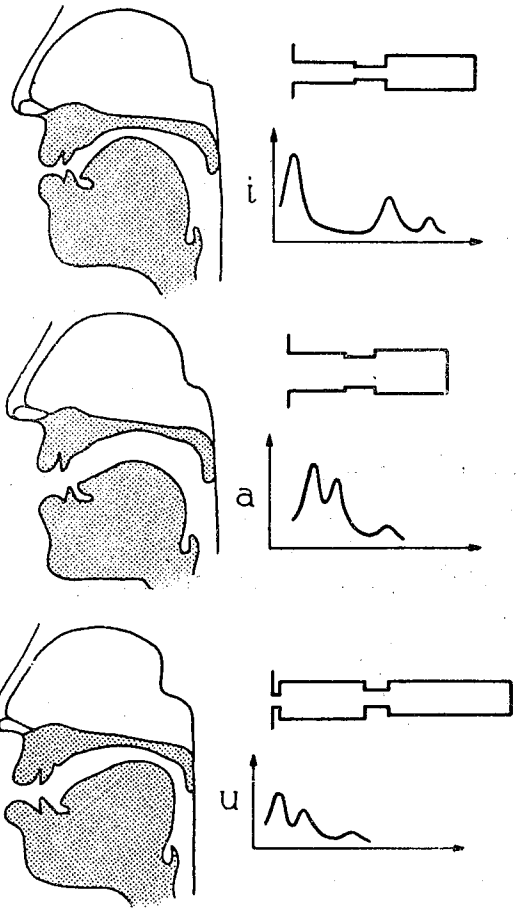
e) litery **ę** i **ą** przed *p*, *t*, *k*, *b*, *d*, *g* wymawia się jako zespoły dwóch głosek, z których pierwszą jest **e** lub **o**, a drugą **m**, **n**, **ɲ** lub **ŋ**, np. *ląd*, wymawia się tak jak *lont*.

W niniejszym artykule stosować będziemy transkrypcję międzynarodową m. in. dlatego, że takiej właśnie transkrypcji używa się w prawie wszystkich pracach z zakresu fonetyki akustycznej.

W zestawieniu powyższym obok znaku transkrypcji międzynarodowej zaznaczono gwiazdką głoski **dźwięczne.** Pozostałe głoski są **bezdźwięczne.**

W grupie dźwięcznych wyróżnić trzeba głoski, które składają się **wyłącznie** z tonów **harmonicznych**, czyli stanowią przebiegi w przybliżeniu **periodyczne.** Tutaj należą wszystkie samogłoski, a dalej **m, n, ɲ, ŋ, l, ł** oraz **j** i **w**. Zgodnie z tym, co powiedziano wyżej (str. 338), na spektrogramach wykonanych przy pomocy filtru szerokowstęgowego, te głoski występują w postaci prążków wzdłuż skali częstotliwości. Ujmiemy ją wspólną nazwą **płynnych.**

Rozpatrując dalej i klasyfikując głoski polskie będziemy musieli zwrócić uwagę na **skupienia energii** w pewnych częstotliwościach, pojawiające się w postaci zaciemnień wzgl. zgrubień prążków, na obrazach spektrograficznych wykonanych przy użyciu filtru szerokowstęgowego. Trzeba w tym miejscu zaznaczyć, że w mowie występują znaczne zróżnicowania tak całkowitego poziomu, jak poziomu poszczególnych składowych. Otóż ujemną stroną sonagrafu jest niedostateczna dynamika



Rys. 4. Samogłoski **i, a, u**
(przekroje artykulacyjne, uproszczone schematy układów akustycznych, charakterystyki częstotliwościowe)

przy trójwymiarowej analizie mowy. Stąd zdarza się nierzadko, że jeśli dany wyraz zawiera głoski o dość znacznej różnicy poziomów (mniej więcej powyżej 20 dB), wówczas skupienia energii charakteryzujące stosunkowo ciche głoski nie wystąpią na spektrogramie (chyba że się przesteruje mocniejsze głoski, co oczywiście jest równie niepożądane). Trze-

ba się więc zastrzec, że nie na wszystkich załączonych tu spektrogramach dadzą się zaobserwować wzmocnienia typowe dla wszystkich głosek danego wyrazu. Większą dynamikę można uzyskać na widmach opisanych wyżej (str. 338), ale i tu nie przekracza ona 36 dB.

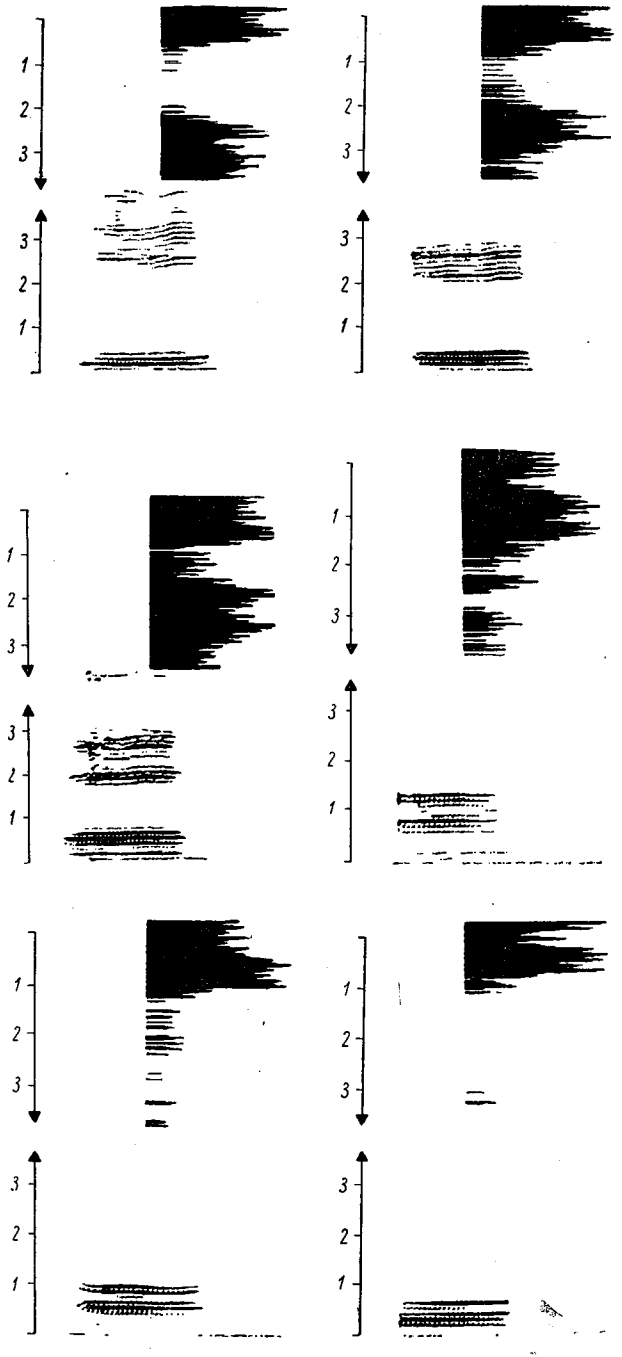
Samogłoski charakteryzują się bardzo wyraźnymi **formantami**, czyli częstotliwościami własnymi komór rezonansowych. Dla unaocznienia zależności pomiędzy położeniem organów mowy a formantami podajemy na rys. 4 przekroje uwidaczniające artykulację trzech samogłosek polski **i, a, u** wraz ze schematami odpowiadających im uproszczonych układów akustycznych oraz charakterystyką częstotliwościową tych układów.

Jak widać z rys. 4, wnętrze jamy ustnej przy artykulacji samogłosek można porównać do układu dwóch sprzężonych komór rezonansowych, które pobudzone drganiami wiązadeł głosowych o licznych harmonicznych daje widmo z wyraźnymi wzmocnieniami w trzech częstotliwościach. Dalsze wzmocnienia mają już z reguły stosunkowo niski poziom i rzadziej są widoczne na spektrogramach.

Rysunek 5 przedstawia 6 polskich **samogłosek ustnych i, ɨ, e, a, o, u** wymówionych **oddzielnie** w postaci spektrogramów oraz widm chwilowych wykonanych przy użyciu filtru wąskowstęgowego (45 Hz). Ponieważ częstotliwość tonu podstawowego wynosi tu około 125 Hz, tak na spektrogramach, jak i na widmach widoczne są poszczególne harmoniczne. Na spektrogramach **i, ɨ, e** widoczne są trzy zakresy wzmocnienia. Na spektrogramach **a, o, u** trzeci formant nie jest widoczny z powodu zbyt niskiego poziomu. Natomiast trzy formanty można odczytać na wszystkich widmach wyjąwszy **u**, przy którym widać tylko dwa niższe formanty. Samogłoski uwidocznione na rys. 5 mają formanty w następujących częstotliwościach (w hercach)[1]).

samogłoska	F_1	F_2	F_3
i	270	2600	3300
ɨ	320	2250	2800
e	620	2000	2750
a	820	1250	2400
o	650	1000	2100
u	380	700	

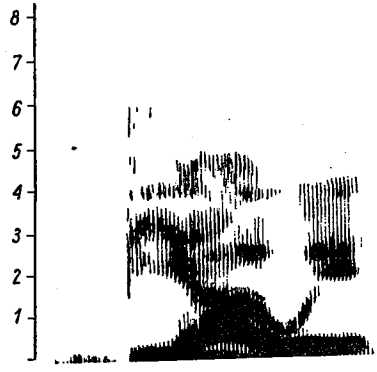
[1]) Por. W. Jassem; The Formants of Sustained Polish Vowels, The Study of Sounds, str. 335 i nast., Tokyo 1957. Gruntowne studium, w którym rozpatrzono w wyczerpujący sposób zależność formantów samogłoskowych od układu rezonatorów jamy ustnej, stanowi praca T. Chiba i M. Kajiyama: The Vowels, its Nature and Structure, Tokyo 1941.



Rys. 5. Spektrogramy 6 polskich samogłosek ustnych **i, ɨ, e, a, o, u**

Jeżeli samogłoska występuje w sąsiedztwie innych głosek, tj. z reguły spółgłosek, nie stanowi ona już w całości przebiegu tak ustalonego, jak to widoczne jest na rys. 5. W początkowej części przebiegu samogłoskowego formanty uzależnione są od głoski poprzedzającej, a w końcowej — od następującej. Występują tu przebiegi **nieustalone** ze wzmocnieniami przesuwającymi się w skali częstotliwości, tak jak to widzieliśmy na rys. 2.

Przy przebiegach ustalonych formanty występują na spektrogramie jako ciemniejsze pasma równoległe do skali czasu. Przy przebiegach nieustalonych zaś otrzymujemy pasma nachylone względem skali czasu pod różnymi kątami, przy czym kąt nachylenia zazwyczaj ulega zmianie o charakterze ciągłym w funkcji czasu. Ze względu na ich wygląd na spektrogramie formanty przebiegów nieustalonych przyjęto potocznie nazywać „wygięciami". Przy jednych połączeniach sąsiadujących głosek wygięcia są znaczniejsze, przy innych mniejsze tak pod względem wielkości kąta, jak i czasu trwania. W pewnym stopniu sąsiednie głoski uzależniają po-



Rys. 6. Spektrogram wyrazu *biały* (**bjawi**)

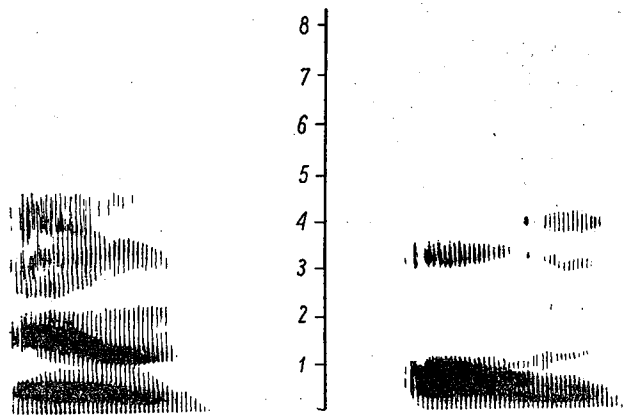
zycję formantów także i w ustalonej — środkowej — części samogłoski. Dla ogólnej orientacji rozważymy rys. 6 przedstawiający spektrogram wyrazu *biały* (wymowa w transkrypcji: **bjawi**).

W pierwszej części samogłoski **a** formant niższy F_1 wygięty jest w dół a wyższy F_2 — w górę. W centralnej części, na mniej więcej $^1/_2$ przebiegu samogłoski F_1 ustalony jest na częstotliwości ok. 700 Hz[1]), aby pod koniec znów opaść. F_2 od początku do końca samogłoski opada, ale w centralnej jej części ma on wyraźne przegięcie i przez krótki odcinek czasu rzędu kilku centisekund jest niemal ustalony na częstotliwości ok.

[1]) Jest to częstotliwość nieco niższa niż w samogłosce **a** wymówionej oddzielnie. Ale też tak poprzedzająca, jak i następująca głoska ma F_1 w częstotliwości znacznie niższej niż 800 Hz.

1200 Hz. Formant trzeci F_3 jest wyraźny dopiero w drugiej części samogłoski i utrzymuje się ok. 2500 Hz. Końcowe i w wyrazie *biały* jest na mniej więcej $^3/_4$ przebiegu ustalone. W pierwszej części, nie przekraczającej $^1/_4$ przebiegu, F_2 przechodzi z niższej do wyższej częstotliwości, po czym ustala się na 2100 Hz. F_1 i F_3 utrzymują się bez wyraźniejszych zmian w okolicy 300 i 2600 Hz.

Samogłoski nosowe, nawet wymówione oddzielnie, nie stanowią przebiegu ustalonego, lecz przebieg tego typu, jaki otrzymujemy z kombinacji dwóch głosek, jak to widać na rys. 7.



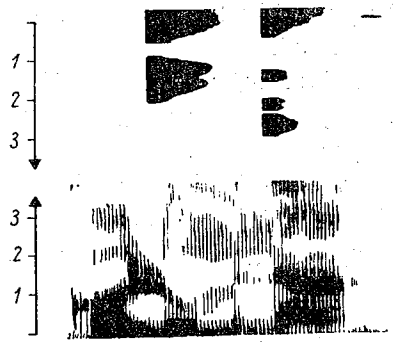
Rys. 7. Samogłoski ẽ i õ

W pierwszej części ẽ ma dwa formanty niższe około 500 i 1800 Hz, w drugiej ok. 300 i 1300. Dwa niższe formanty õ leżą w pierwszej części ok. 400 i 800 Hz, w drugiej — niemal się zlewają wokół 500 Hz. Dla samogłosek nosowych ważny jest dalszy formant ok. 2000 Hz, szczególnie w drugiej części (słabo widoczny przy õ na rys. 7). Samogłoski składające się z dwóch elementów w skali czasu, czyli z dwóch **segmentów,** nazywają się **dyftongami.** Takimi są właśnie polskie ẽ i õ[1]).

Głoski **j** i **w** noszą nazwę **samogłosek niesylabicznych.** Akustycznie i artykulacyjnie są one bardzo zbliżone do samogłosek **i** względnie **i** oraz **u** (tj. samogłosek o najniższym F_1). Jak widać z rys. 6, 16, 17 i 18 formanty **j** wypadają w częstotliwościach 300, 2000 ÷ 2400 oraz 2600 ÷ 3000 Hz, a formanty **w** ok. 300 oraz 700 Hz.

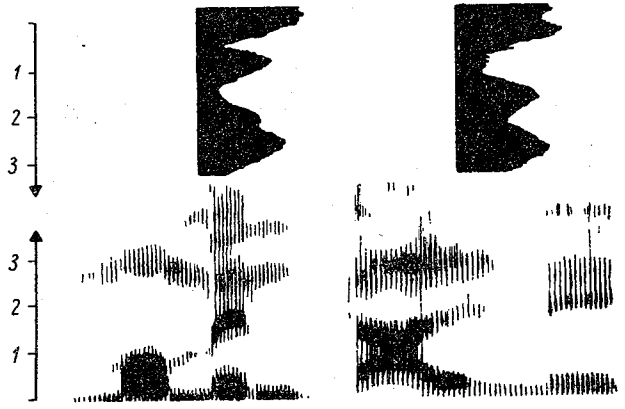
[1]) W języku francuskim samogłoski nosowe są jednorodne. Opracowań spektrograficznych głosek francuskich jednak dotychczas brak. O dyftongicznym charakterze polskich samogłosek nosowych zob. T. Benni: Samogłoski polskie, analiza fizjologiczna i systematyka, Prace Towarzystwa Naukowego Warszawskiego nr 8, 1912, str. 56 i nast. Por. W. Jassem: Polish, Le Maître Phonétique ser. III, nr 88, 1947, str. 29. Zob. też w niniejszym artykule rys. 29 i 20.

Spółgłoski płynne podzielić trzeba na dwie grupy: nosowe i boczne. Spółgłoski **nosowe** mają dwa formaty charakterystyczne dla całej tej grupy. Pierwszy F_1 w okolicy tonu podstawowego, a drugi, słabszy, F_3 nieco powyżej 2 kHz. Około 2,5 ÷ 2,8 kHz występuje jeszcze zwykle formant dalszy, F_4. F_3 około 2 kHz charakteryzuje tak spółgłoski, jak i samogłoski nosowe (por. wyżej). Jest to więc tzw. „formant nosowy". Na razie brak



Rys. 8. Spektrogram wyrazu *kolumna* (**kolumna**) z widmami głosek **l** i **n**

dostatecznie pewnych danych, które pozwoliłyby w sposób niewątpliwy wyjaśnić zależność artykulacji, polegającej na przepuszczeniu powietrza przez kanały nosowe, ze wspomnianym formantem. Na rys. 8 i 9 widocz-



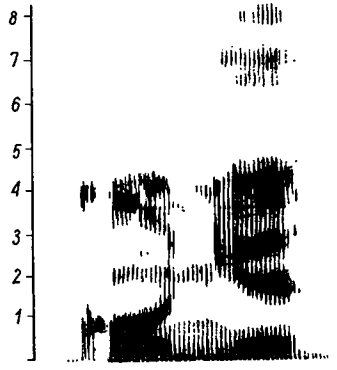
Rys. 9. Spektrogram wyrazu *momentalny* (**momentalni**) z widmami głosek **m** i **l**

ne są formanty głosek **m** i **n** w okolicy tonu podstawowego, dalej około 2 kHz oraz ok. 2,5 kHz.

Na widmie głoski **n** na rys. 8 F_3 przy 2 300 Hz oraz F_4 przy 2800 Hz są wyraźnie widoczne.

Do grupy spółgłosek nosowych należą cztery: **m**, **n**, **ɲ**, **ŋ**. Przejście pomiędzy F_1 spółgłoski nosowej a F_1 sąsiedniej samogłoski jest raptowne, tzn. nie obserwuje się tu „wygięć". Dobitnie widoczne to jest na rys. 9 przedstawiającym spektrogram wyrazu *momentalny* (2 razy **m**, 2 razy **n**).

Spółgłoski nosowe różnią się pomiędzy sobą położeniem formantu drugiego F_2 oraz wygięciem formantu drugiego sąsiadującej samogłoski. Formant drugi **m** leży przy 700 ÷ 900 Hz, a F_2 **n** przy 1200 ÷ 1400 Hz. Cztery formanty **m** i **n** widoczne są na widmach na rysunkach 8 i 9 (filtr szerokowstęgowy). Bardziej wyraziste niż formanty F_2 spółgłosek nosowych są jednak wygięcia formantu F_2 sąsiednich samogłosek. W dalszym ciągu takie wygięcia będziemy nazywać **transientami samogłoskowymi**[1]). Na rysunku 9 widać obustronne lekkie wygięcie w dół F_2 samogłoski **o**, który w środkowej jej części leży ok. 800 Hz (a więc nieco niżej niż przy izolowanej samogłosce); F_2 **m** jest zresztą w pewnej mie-



Rys. 10. Spektrogram wyrazu *konie* (**koɲe**)

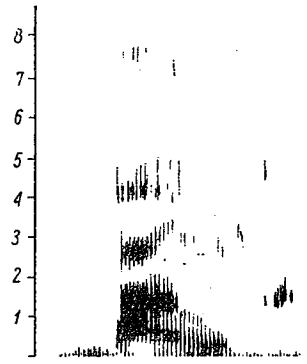
rze też uzależniony od sąsiedztwa samogłoskowego. Początkowe **m** w **momentalni** ma F_2 przy 700 Hz, a F_2 **m** pomiędzy **o** a **e** przesuwa się w czasie trwania spółgłoski nieco w górę. W pierwszej części **e** widoczny jest bardzo wyraźnie transient wiążący F_2 **m** z F_2 **e**, który zresztą przy krótkim trwaniu samogłoski (ok. 0,07 sec) i obustronnym sąsiedztwie spółgłosek o niższym F_2 osiąga tylko 1700 Hz. Wreszcie rys. 8 ukazuje transient samogłoskowy końcowego **a** wiążący F_2 **n** przy 1400 Hz z F_2 **a** przy 1100 Hz.

Oba transienty samogłosek otaczających **ɲ** na rys. 10 wyginają się ku 2 kHz. Przyjmujemy zatem, zgodnie z tym, co zauważono dla **m** i **n**,

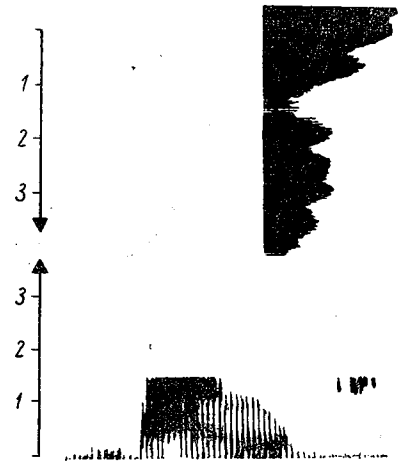
[1]) W pracy A. Malécot: Acoustic Cues for Nasal Consonants, Language (USA) vol. 32, 1956, str. 294 i nast., wykazano, że transienty samogłoskowe są ważniejsze dla identyfikacji słuchowej spółgłosek nosowych niż położenie formantu drugiego w samych spółgłoskach.

że formant charakterystyczny dla **ɲ** leży w tej samej częstotliwości co F_3 wszystkich spółgłosek nosowych mimo wzmocnienia ok. 750 Hz, które określimy jako formant dodatkowy, ponieważ nie wiąże się z formantami samogłoskowymi.

Spółgłoska **ŋ** posiada oba formanty charakterystyczne dla spółgłosek nosowych z tym, że F_1 leży nieco wyżej niż przy pozostałych **nosowych**, mianowicie ok. 400 Hz. Około 1000 Hz przypada formant do-



Rys. 11. Spektrogram wyrazu *bank* (**baŋk**)



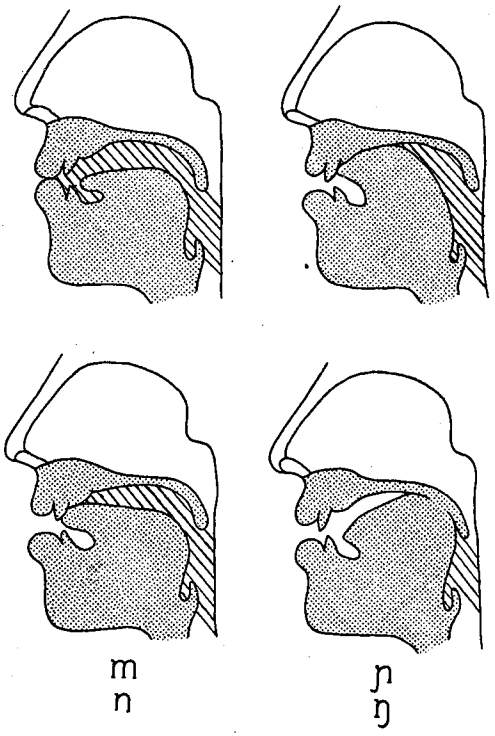
Rys. 12. Widmo głoski **ŋ** w wyrazie *bank* (**baŋk**)

datkowy, podobnie jak przy **ɲ**. Formant charakterystyczny dla **ŋ** leży **powyżej** formantu „nosowego", a mianowicie ok. $3 \div 3{,}2$ kHz. F_2 **a** na rys. 11 nie wykazuje transientu, natomiast wyraźny jest transient F_3 **a**, który od 2600 Hz podnosi się do 3200 Hz. Rysunek 12 ukazuje widmo głoski **ŋ** w wyrazie *bank* (filtr szerokowstęgowy) z widocznymi wszystkimi formantami w częstotliwościach: 400 Hz, 1000 Hz, 2000 Hz i 3250 Hz.

Można zatem spółgłoski nosowe ułożyć w szereg: **m**, **n**, **ɲ**, **ŋ**, według wysokości charakterystycznego formantu — od najniższej do najwyższej częstotliwości. To uszeregowanie zgodne jest z klasyfikacją artykulacyjną — od najbardziej przedniego (tzw. wargowego) **m** do najbardziej tylnego (miękkopodniebiennego) **ŋ**.

Na rysunku 13 pokazano położenie języka przy artykulacji spółgłosek nosowych. Nasuwa się przypuszczenie, że formant charakterystyczny F_2 jest związany z komorą rezonansową zaznaczoną kreskami na rysunku.

Do grupy spółgłosek **bocznych** należą dwie spółgłoski **l** i **ɫ**. Głoska **l** charakteryzuje się przede wszystkim dwoma formantami, z których pierwszy



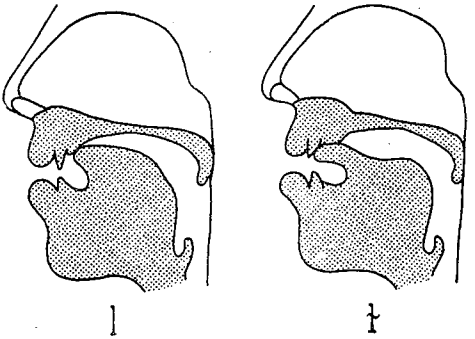
Rys. 13. Przekroje artykulacyjne spółgłosek nosowych

F_1 leży w pobliżu 300 Hz, drugi waha się pomiędzy 1300 ÷ 1900 Hz w zależności od sąsiednich samogłosek. Tak na przykład w wyrazie **kolumna** (rys. 8) F_2 **l** nie przekracza częstotliwości 1300 Hz, ponieważ F_2 tak poprzedzającego **o**, jak i następującego **u** leżą poniżej 1 kHz. Natomiast w **momentalni** F_2 **l** ma częstotliwość 1700 Hz. Silnie uzależniony od sąsiedztwa samogłoskowego jest także trzeci formant **l**. W sąsiedztwie samogłosek o stosunkowo niskich formantach F_2 i F_3 formant trzeci **l** obniża się aż do 1700 Hz, podczas gdy w innych przypadkach dochodzi on do 3 kHz (por. rysunki 8 i 9). Spółgłoska **ł** (używana obecnie przez aktorów na scenie, w szczególnie starannej dykcji, a potocznie w północno-wschodniej Polsce) ma F_1 w tej samej częstotliwości co **i**, natomiast drugi formant **ł** leży ok. 700 Hz, a więc znacznie niżej niż przy **l**. Trzeci formant leży ok. 2,6 kHz. Istotniejszym wahaniom pozycje formantów głoski **l** nie ulegają. Na rys. 14 widoczne jest oprócz spektrogramu wyrazu *łydka* (**łitka**) widmo **ł** ukazujące trzy formanty. Artykulacyjnie **l** i **ł** powstają w ten sposób, że przód języka zwiera się z dziąsłem ponad przednimi siekaczami, podczas gdy boki języka nie przylegają na całej długości do sklepienia jamy ustnej, tak iż po bokach pozostaje wolna szczelina, przez

którą przechodzi wydechane powietrze. Stąd też nazwa spółgłosek bocznych. Częstotliwość formantu drugiego, a być może też i trzeciego, prawdopodobnie stoi w związku z położeniem języka wewnątrz jamy ustnej, jak to wydaje się sugerować rys. 15.

Rys. 14. Spektrogram wyrazu *łydka* (**łątka**) z widmem głoski **ł**

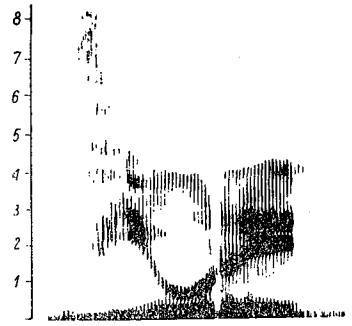
Warto nawiasowo zwrócić uwagę na fakt, że dwa niższe formanty są takie same w przypadku głosek **ł** i **w**, których jak to wyżej wskazano — używa się wymiennie w polskim języku.



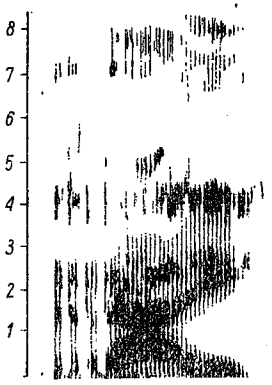
Rys. 15. Przekroje artykulacyjne głosek l i ł

Następną grupę stanowią spółgłoski **szumiące,** zwane także **trącymi** lub niezbyt trafnie **szczelinowymi.** Grupę tę charakteryzuje **szum** w szerszym lub węższym zakresie powyżej 2 kHz. Grupa dzieli się na dwa typy. Pierwszy nie posiada tonu podstawowego ani składowych harmonicznych; są to szumiące **bezdźwięczne.** Drugi — to przebiegi stanowiące superpozycję szumu na przebieg w przybliżeniu periodyczny, o tonie podstawowym z kilkoma do kilkunastu harmonicznych. Spółgłoski szumiące układają się w pary, przy czym każda para ma określoną cechę artykulacyjną i akustyczną.

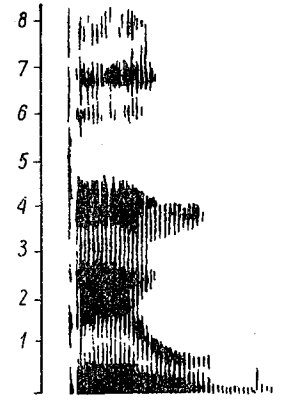
Pierwszą parę stanowią głoski **f** i **v**. Ich szum ma bardzo niski poziom i rozkłada się niemal na cały zakres analizowany. Widoczne są jednak wzmocnienia w wyższych częstotliwościach około 6 ÷ 7 kHz. Poza tym



Rys. 16. Spektrogram wyrazu *wióry* (**vjurɨ**)

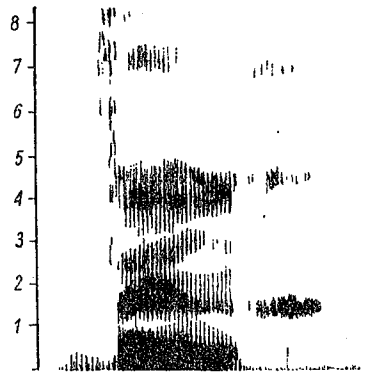


Rys. 17. Spektrogram wyrazu *raj* (**raj**)

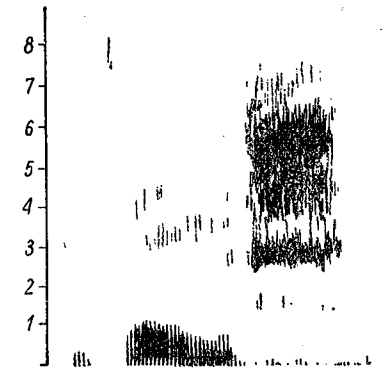


Rys. 18. Spektrogram wyrazu *pył* (**pɨw**)

występują tu lekkie wzmocnienia przy 1500 Hz i 2500 Hz. Na spektrogramach wyrazów *wióry* (**vjurɨ**) i *gniew* (**gɲef**) (rysunki 16 i 21) dadzą się one łatwo zaobserwować, natomiast na rysunkach 19 i 20 z powodu



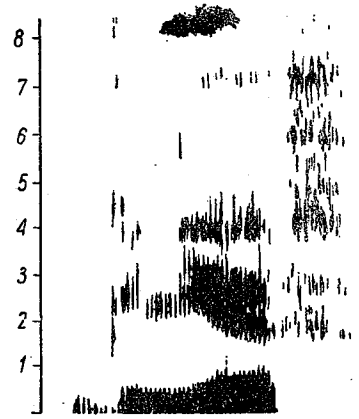
Rys. 19. Spektrogram wyrazu *węch* (**vẽx**)



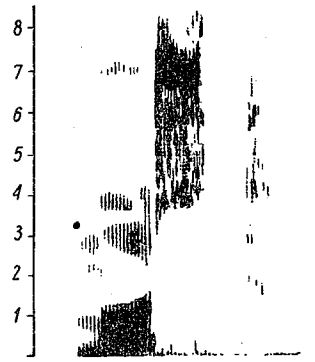
Rys. 20. Spektrogram wyrazu *wąż* (**võʃ**)

zbyt niskiego poziomu w stosunku do samogłosek głoska **v** jest prawie niewidoczna, podobnie jak na rys. 34. Transienty samogłoskowe w sąsiedztwie **f** i **v** dążą do częstotliwości 1,5 i 2,5 kHz charakterystycznych dla tych spółgłosek (zob. szczególnie rys. 21).

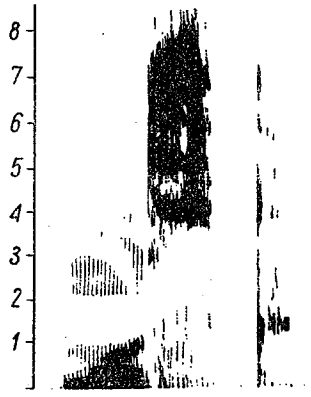
Głoski **s** i **z** odznaczają się wysokim poziomem szumu, koncentrującym się powyżej 4 kHz. Pomiędzy 6 ÷ 7 kHz zaznacza się zazwyczaj dodatkowe wzmocnienie. Cechy te widoczne są na rys. 22, 23 i 24.



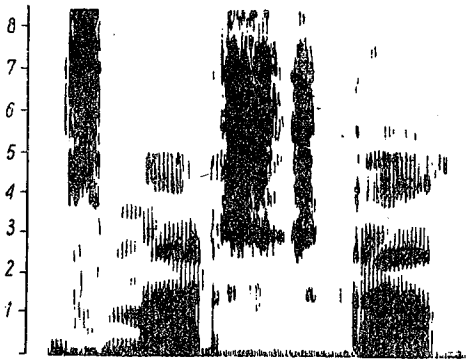
Rys. 21. Spektrogram wyrazu *gniew* (gɲef)



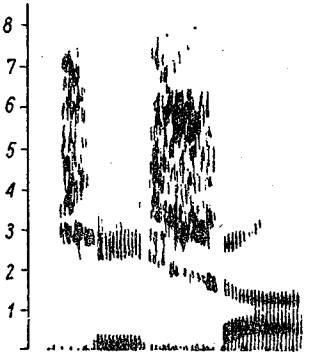
Rys. 22. Spektrogram wyrazu *most* (most)



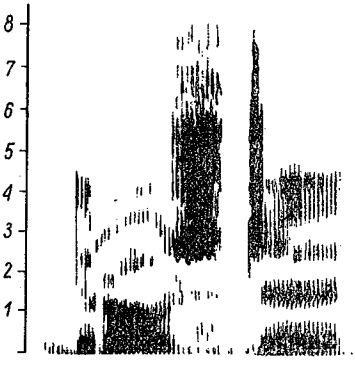
Rys. 23. Spektrogram wyrazu *mózg* (musk)



Rys. 24. Spektrogram wyrazu *zmarszczka* (zmarʃt͡ʃka)



Rys. 25. Spektrogram wyrazu *cisza* (t͡ɕiʃa)



Rys. 26. Spektrogram wyrazu *drożdże* (drozd͡ʒe)

Znamienne dla s i z są transienty sąsiednich samogłosek. Dążą one do częstotliwości ok. 1,2 kHz. W przebiegu samych spółgłosek s, z wyraźniejszego wzmocnienia w tej częstotliwości jednak nie stwierdzono.

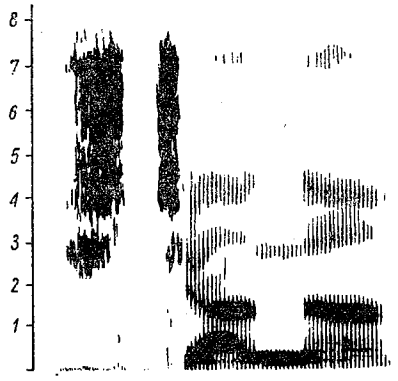
Przy głoskach ʃ i ʒ zakres szumów jest nieco szerszy niż przy s i z i rozpoczyna się już w okolicy 3 kHz, a nawet niżej. Na samym początku zakresu szumów, ok. 3 kHz, występuje przy ʃ i ʒ poziom szczytowy w widmie. Poza tym spółgłoski te charakteryzują się dodatkowym, bardzo wąskim zakresem szumu, którego częstotliwość jest w znacznym stopniu zależna od sąsiednich głosek. W sąsiedztwie spółgłoskowym (np. w wyrazie *zmarszczka*, rys. 24) ten odrębny zakres szumów przypada w okolicy częstotliwości 1,4 ÷ 1,5 kHz. Zależność tego osobnego zakresu szumów od F_2 sąsiednich samogłosek ilustruje dobitnie rys. 25. W czasie trwania głoski zmienia się on od ok. 2,3 kHz do ok. 1,5 kHz, przy czym transient F_2 a kontynuuje jeszcze to „opadanie", aż do ustalonego przebiegu a. Większe wzmocnienie szumu przy 3 kHz i słabsze przy 1,4 kHz widoczne jest też przy głosce ʒ na rys. 26.

Zwarty zakres szumu przy głoskach ɕ i ʑ rozciąga się pomiędzy 3,5 kHz a 7 kHz. Prócz tego w częstotliwości ok. 2,5 kHz występuje odrębny zakres szumów o dość znacznym poziomie. Te cechy widoczne są na rysunkach 27 i 28. Jeśli F_2 sąsiedniej samogłoski leży poniżej 2,5 kHz, występuje transient jak na załączonych spektrogramach wyrazów *ściana* i *ziarno*.

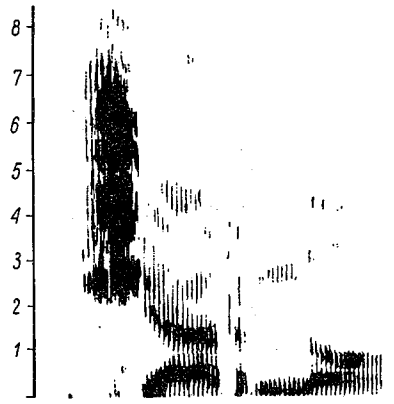
Spółgłoska x jest bezdźwięczna, a jej dźwięczny odpowiednik jest używany bardzo rzadko [1]). Ma ona dwa wąskie zakresy szumów: Jeden występuje niezmiennie w częstotliwości 4 kHz, a drugi jest silnie uzależniony od sąsiedztwa. Ten niższy zakres szumów ma niemal tę samą częstotliwość co F_2 sąsiedniej samogłoski. Widać to na rysunkach 19, 29, 30 i 31. Jednakże powyżej 1,8 kHz niższy zakres szumów nie występuje; taka samogłoska jak i, ze stosunkowo wysokim F_2, ma ten formant w sąsiedztwie x w nieco niższej częstotliwości niż przy wymówieniu w izolacji (por. rys. 30). Około 6÷7 kHz występuje często na spektrogramach głoski x jeszcze trzeci zakres szumów.

---

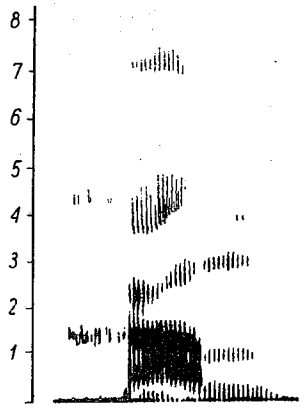
[1]) W niniejszym artykule świadomie pominięto problemy fonematyki jako zasadniczo raczej językoznawcze niż akustyczne. Nie jest jeszcze rzeczą jasną, jak dalece pojęcie fonemu byłoby istotne przy automatycznym rozpoznawaniu dźwięków mowy. Wspominam o tym w tym miejscu, gdyż kwestię dźwięcznego odpowiednika głoski x można objaśnić jedynie fonematycznie. Chodzi tu mianowicie o tzw. wariant kombinatoryczny. Pojęcie to należy ściśle do fonematyki. Por. W. Jassem: Fonetyka języka angielskiego, str. 37 i nast., Warszawa, 1954, oraz W. Jassem: Węzłowe zagadnienia fonematyki, Biuletyn Pol. Tow. Językoznawczego zeszyt XV, 1956, str. 24 i nast.



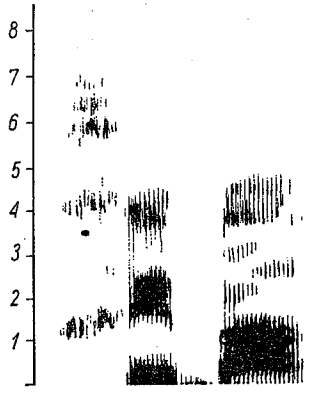
Rys. 27. Spektrogram wyrazu *ściana* (**ɕt͡ɕana**)



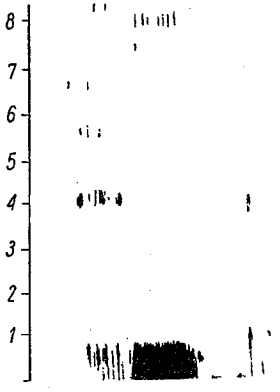
Rys. 28. Spektrogram wyrazu *ziarno* (**ʒarno**)



Rys. 29. Spektrogram wyrazu *cham* (**xam**)

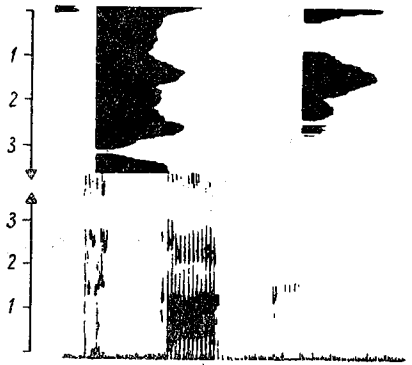


Rys. 30. Spektrogram wyrazu *chyba* (**xʲba**)

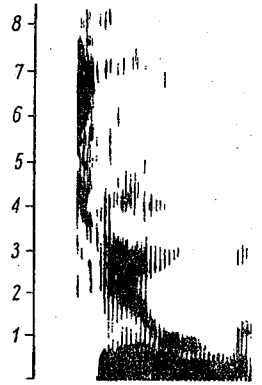


Rys. 31. Spektrogram wyrazu *huk* (**xuk**)

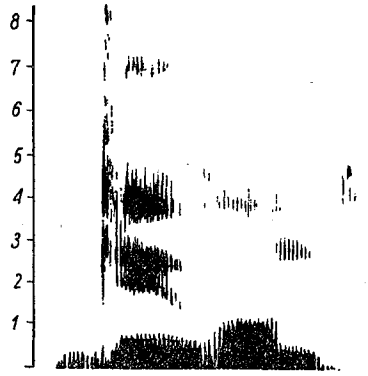
Zależności między artykulacją a cechami akustycznymi głosek szumiących nie zostały jeszcze bliżej zbadane[1]). W każdym razie mamy tu zapewne do czynienia z wirującymi ruchami powietrza na skutek przeciskania się przez wąską szczelinę pod wzmożonym ciśnieniem.



Rys. 32. Spektogram wyrazu *ptak* **(ptak)** z widmami segmentów szumiących **p** i **k**



Rys. 33. Spektrogram wyrazu *kieł* **(cel)**



Rys. 34. Spektrogram wyrazu *Giewont* (**jevont**)

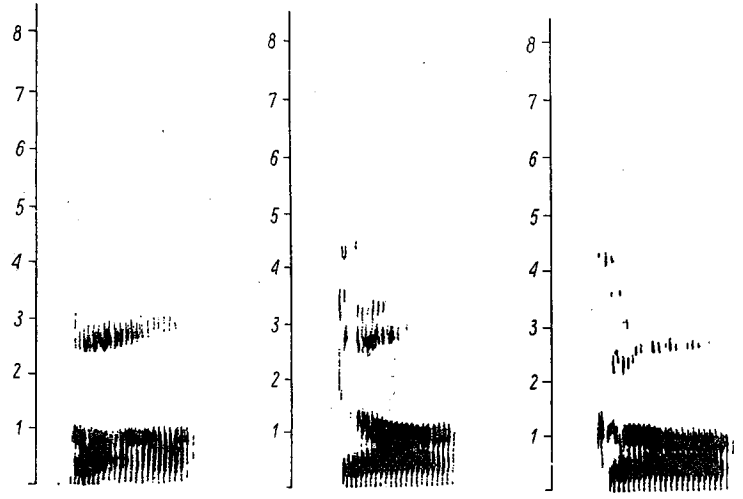
Dalszą grupę głosek obejmuje się nazwą **zwartych** albo **wybuchowych.** Są to głoski niejednorodne w tym sensie, że składają się akustycznie z dwóch lub trzech części w skali czasu, czyli z dwóch wzgl. trzech **segmentów.** Pierwszy segment nazwiemy — używając terminologii zaczerpniętej z fonetyki artykulacyjnej — **zwarciem.** W przypadku spółgłosek

[1]) Cechy spektralne spółgłosek trących omawiane są w następujących pracach: T. Tarnóczy: Die akustische Struktur der stimmlosen Engelaute, Acta Linguistica Academiae Scientiarum Hungaricae, str. 313 i nast., Budapest 1954; K. S. Harris: Cues for the Identification of the Fricatives of American English, JASA, vol. 26, 1954, str. 952 i nast.; G. W. Hughes i M. Halle: Spectral Properties of Fricative Consonants, JASA, vol. 28, 1956, str. 303 i nast.

bezdźwięcznych zwarcie jest równoważne akustycznie z brakiem sygnału, czyli przerwą. Czas trwania tej przerwy jest rzędu 0,05 sec. Jeśli spółgłoska zwarta bezdźwięczna występuje po krótszej lub dłuższej przerwie w nadawaniu (mówieniu), wówczas zwarcie zlewa się z tą przerwą, a zatem „informacyjnie" nie istnieje — nie stanowi odrębnego segmentu. W spółgłoskach dźwięcznych zwarcie wykazuje strukturę harmoniczną ze skupieniem energii w okolicy tonu podstawowego. Poziom ogólny tego segmentu jest bardzo niski, tak że przy słabym wysterowaniu czasem na spektrogramach także i dźwięczne głoski zwarte mogą w pierwszej części mieć wygląd krótkiej przerwy. Drugi segment jest to „wybuch", czyli (stosując terminologię fonetyczną) **plozja**, stanowiąca bardzo krótki przebieg zbliżony charakterem do pojedynczego impulsu. Plozja występuje na spektrogramach jako wąski prążek pionowy w szerokim zakresie częstotliwości, nieraz na całym prawie zakresie analizowanym. W pewnych częstotliwościach charakterystycznych, o których mowa niżej, ten prążek ma zgrubienia. Trzecia część ma charakter szumów superponowanych w przypadku głosek dźwięcznych na przebieg w przybliżeniu harmoniczny, skupionych wokół określonych częstotliwości. Po względem długości trwania ta trzecia część rozciąga się od wartości zerowej (to znaczy może jej być brak) do ok. 0,05 sec. Niższe wartości występują przy dźwięcznych, wyższe przy bezdźwięcznych. Jeśli po spółgłosce zwartej bezdźwięcznej następuje dalsza spółgłoska albo jeśli się ona znajduje przed przerwą w nadawaniu (pauzą), część szumiąca często osiąga maksymalne wartości czasowe. Jest ona wtedy wyraźnie słyszalna i nosi nazwę **aspiracji.** Jeżeli po zwartej dźwięcznej następuje dalsza spółgłoska (przed pauzą taka spółgłoska w polskim języku nie występuje), wówczas po segmencie szumiącym pojawia się jeszcze dalszy, dodatkowy segment, zbliżony akustycznie do samogłoski **i**. Taka samogłoska jest przez językoznawców określana mianem „szwa". Głoski zwarte są bardzo częste w polskim i załączone spektrogramy dają liczne ich przykłady. I tak rysunki 8, 9, 10, 14, 18, 32 i 33 pokazują zwarte bezdźwięczne **k, t, p** i **c** przed samogłoskami; rysunki 14, 22, 23, 31, 32 zwarte bezdźwięczne **p, t, k** przed spółgłoską lub w pozycji końcowej. Rysunki 6, 11, 30 i 34 dają przykłady na zwarte dźwięczne **b** i **g** przed samogłoską; rysunki 21 i 26 ilustrują zwarte dźwięczne **g** i **d** przed spółgłoską.

Spółgłoski zwarte układają się w pary **dźwięczna-bezdźwięczna.** Pary te różnią się między sobą poziomem oraz częstotliwością skupień energii w segmencie plozji i segmencie szumiącym jako też transientami samogłosek sąsiednich. Wszystkie zwarte mają w części następującej po zwarciu słabsze lub silniejsze skupienie energii w częstotliwości 4 kHz (jeśli jest ono szczególnie słabe, może być na spektrogramach niewidoczne).

Przed spółgłoską i przed pauzą przy spółgłoskach zwartych bezdźwięcznych stwierdza się następujące wzmocnienia: **p** — przy 1,5 i 2,5 kHz; słabsze przy 4 kHz; **k** — silniejsze przy 1,5 kHz, słabsze przy 2,5 kHz i jeszcze słabsze przy 4 kHz; **t** — najsilniejsze przy 4 kHz, słabsze przy 1,8 i 2,5 kHz. Transienty samogłoskowe F_2 skierowane są ku następującym częstotliwościom: przy **p-b** 700÷1300 Hz w zależności od samogłoski: niższe F_2 wyginają się ku niższym wartościom, wyższe — ku wyższym.



Rys. 35. Spektrogramy sylab **bo, do, go**

Przy **k-g** F_2 kierują się w przypadku samogłoski **e** ku 2000 Hz, w wypadku **o** i **u** ku 800÷900 Hz, a F_2 **a** kieruje się bądź ku 1,5, bądź 2,0 kHz [1]). Przy **t-d** F_2 sąsiednich samogłosek skierowany jest ku częstotliwości 1,8 kHz. Jeżeli po zwartej następuje samogłoska, wówczas skupienia energii widoczne w segmencie plozji i szumu przypadają w częstotliwościach, ku którym skierowany jest F_2 samogłoski. Para **c-ɟ** jest odrębna od pozostałych. W polskim języku spółgłoski te występują tylko przed **i** oraz **e**. Zakres szumów jest znacznie szerszy niż przy innych zwartych — od 2,5 kHz niemal do końca zakresu analizowanego — i tą cechą **c-ɟ** zbliżają się do typu zwarto-trących, o których będzie mowa w dalszym ciągu. Transienty F_2 dążą w tym przypadku do częstotliwości 2,5 kHz.

F_1 sąsiedniej samogłoski (zwłaszcza następującej) w przypadku zwartych dźwięcznych wygięty jest łagodnie w dół ku częstotliwości tonu

[1]) Od jakich dodatkowych okoliczności zależy ta alternatywa, tego na razie nie udało mi się stwierdzić.

podstawowego; wygięcie to jest mniej wyraźne w sąsiedztwie zwartych bezdźwięcznych.

Rysunek 35 (por. też rys. 2) wskazuje porównawczo transienty F_2 i F_1 w sylabach **bo, do, go** [1]).

Ostatnia grupa spółgłosek to tzw. **zwarto-trące.** Wykazuje ona ogólne cechy podobne do grupy poprzedniej (zwartych) z następującymi różnicami: 1) Plozja jest bardzo słaba i dlatego rzadko widoczna na spektrogramach. 2) Segment trący (szumiący) jest zawsze obecny, przy czym zakres szumów jest szeroki. 3) Względna długość trwania segmentu trącego wydaje się być istotną cechą rozróżniającą grupy zwartych, zwarto-trących i trących. Z zastrzeżeniem, że dźwięczne mają z zasady krótszy segment trący (szumiący) niż analogiczne bezdźwięczne, można ustalić porządek: zwarte — zwarto-trące — trące według rosnącego czasu trwania segmentu szumiącego, najdłuższego zatem przy bezdźwięcznych trących, a najkrótszego przy dźwięcznych zwartych[2]).

Plozja (o ile jest zauważalna) występuje przy zwarto-trących jako prążek równoległy do skali częstotliwości, o niezbyt szerokim zakresie, z centrum w okolicy 3 ÷ 3,5 kHz.

Poszczególne pary zwarto-trących różnią się pomiędzy sobą skupieniami energii w części szumiącej. Z kolei zaś segment szumiący jest tu (poza długością trwania) analogiczny jak przy odpowiednich głoskach szumiących, tak jak na to wskazuje transkrypcja: **t͡s-d͡z, t͡ʃ-d͡ʒ, t͡ɕ-d͡ʑ.** Również transienty samogłosek po zwarto-trących są takie, jak przy odpowiednich trących. Transienty samogłosek poprzedzających są podobne jak w sąsiedztwie **t-d.**

Wskazane wyżej cechy zwarto-trących ilustrują rysunki 24, 26 i 27.

Odrębny typ stanowi spółgłoska **r** zwana **drżącą.** Jest ona podobna do głoski zwartej dźwięcznej z tym, że segment zwarcia jest szczególnie krótki (rzędu 0,02 sec). Plozja występuje na spektrogramach jako prążek w szerokim pasmie częstotliwości. Po plozji występuje zawsze element

---

[1]) Analizie głosek zwartych poświęcono szereg wnikliwych prac, z których najważniejsze są następujące: P. C. Delattre, A. M. Liberman i F. S. Cooper: Acoustic Loci & Transitional Cues for Consonants, JASA, vol. 27, 1955, str. 769 i nast.; E. Fischer-Jorgensen: Acoustic Analysis of Stop Consonants, Miscellanea Phonetica II, str. 42 i nast., London 1954; L. N. Stevens i A. S. House: Studies of Formant Transitions Using a Vocal Tract Analog, JASA, vol. 28, 1956, str. 576 i nast.

[2]) L. J. Gerstman stwierdza w notatce: Noise Durations as a Cue for Distinguishing among Fricative, Affricate and Stop Consonants, JASA vol. 28, 1956, że przez skrócenie czasu trwania segmentu szumiącego w sylabie **ʃa** (np. przez demagnetyzację zapisu magnetycznego) otrzymuje się sylabę słyszaną jako **t͡ʃa**, a po dalszym skróceniu — **ka** wzgl. **ta.** Stosunki czasowe mogą w tym wypadku okazać się istotne przy automatycznej identyfikacji mowy.

„szwa" (por. wyżej), to jest krótka samogłoska z formantami ok. 400, 1300 oraz 2500 Hz. Długość tej samogłoski zawiera się w granicach 0,02 ÷ ÷ 0,04 sec. Sąsiednie samogłoski — tak poprzedzająca jak i następująca — wykazują transienty skierowane do wskazanych częstotliwości. Opisana głoska **r** występuje na rysunkach 16, 24, 26 i 28. Przy dobitnym wymawianiu wskazany przebieg powtarza się kilkakrotnie, jak na rys. 17.

*

Gruntowna analiza spektrograficzna, która mogłaby być podstawą do automatycznej identyfikacji polskich głosek wymagać będzie nie tylko przebadania wymowy znacznej liczby osób, ale także bliższego określenia wzajemnych zależności sąsiadujących głosek, niż to było możliwe w powyższym wstępnym studium. Wydaje się, że przedstawione wyżej badania wraz z odnośną literaturą dotyczącą innych języków wskażą właściwą drogę dla dalszych dociekań.

Analizy zreferowane w niniejszym artykule wykonano w czasie kilkutygodniowego pobytu w Londynie za skierowaniem Instytutu Podstawowych Problemów Techniki PAN. Profesorowi D. Fry, kierownikowi Department of Phonetics University College London, za łaskawe udostępnienie mi laboratorium, kol. P. Denesowi, kierownikowi laboratorium, za cenne wskazówki techniczne, a wreszcie laborantom pracującym w tymże Department za pomoc techniczną składam serdeczne podziękowanie. Departament of Phonetics UCL poniósł również koszt cennego papieru do zapisów spektrograficznych.

Instytut Podstawowych Problemów Techniki
Polskiej Akademii Nauk

В. ЯССЭМ

ВСТУПИТЕЛЬНЫЙ СПЕКТРОГРАФИЧЕСКИЙ АНАЛИЗ ПОЛЬСКИХ ЗВУКОВ РЕЧИ

Резюме

Целью работы является проведение впервые определения акустических сигналов соответствующих звукам польского языка с помощью спектрографа. Выделены три типа информационных сигналов: квазипериодические, шумы а также смешанные сигналы, составляющие суперпозицию шумов на квазипериодические сигналы. В процессе идентификации отдельных звуков особенно важную роль играют диапазоны усиления в спектре. Трехразмерный анализ произведенный по времени, по частоте и по интенсивности даёт возможность сегментации, то есть выделения в сигнале, изменяющемся во вре-

мени по распределению энергии в спектре, характеристических отрезков, соответствующих отдельным звукам. Гласные характеризируются двумя или же тремя формантами в этой части процесса, в которой изменения в энергетическом спектре являются самыми медленными. Они являются квазипериодическими процессами. Похожие процессы соответствуют плавным согласным (м, н, л, р) с тем однако, что для идентификации их имеют значение не только те части процесса, которые являются относительно установившимся в своём энергетическом спектре, но также и те, у которых этот спектр подвергается отчётливому изменению во времени. Эти неустановившиеся процессы, называемые транзиентами, являются тоже существенными для идентификации смычных согласных, а также так называемых неслоговых гласных. Щелевые согласные являются или же шумами, или же суперпозицией шума на квазипериодический процесс, при чём акустические процессы соответствующие им являются относительно установившимися. Смычно-щелевые согласные являются непосредственно одна за другой следующими комбинациями процессов аналогических к тем, которые характеризуют смычные и щелевые согласные. Представленный анализ относится к голосу одного человека.

W. JASSEM

A PRELIMINARY SPECTROGRAPHIC ANALYSIS OF POLISH SPEECH SOUNDS

Summary

This is a first attempt to find the characteristics of the acoustic signals corresponding to the sounds of Polish speech made with the aid of a sound spectrograph, Three kinds of information-bearing elements have been distinguished: near-periodic, noise-like, and such as form a superposition of a noise-like signal on a near-periodic. In the identification of the individual speech sounds, a particularly important role is played by the energy concentrations in the spectrum. A three-dimensional, energy—frequency—time analysis reveals a segmentation of the time variable speech signal into elements corresponding to the individual sounds. Vowels are characterized by two or three formants at the moments when the rate of change of the energy-spectrum is minimum. They are near-periodic. Similar signals correspond to liquid consonants, but here not only the relatively steady segments, but also the neighbouring transitions are characteristic. Such transitions are also essential to the identification of the stops and the non-syllabic vowels. The fricatives are noise-like signals alone or superposed on near-periodic signals. They have a relatively steady distribution of energy with frequency. The affricates are essentially close combinations of stops and fricatives.

The analysis is based on data relating to one male voice.

ERRATA

do artykułu J. L. Maksiejewskiego „Metody obliczania prądu w odgromniku przy uderzeniu pioruna w linię w niedużej odległości od odgromnika"
zeszyt 1, tom V, 1959

Strona	Miejsce	Jest	Powinno być
55	wzór (6)	u_2'	u_2''
57	wzór (22)	$p_1 = \frac{1}{T}$	$p_1 = -\frac{1}{T}$
61, rys. 5	Pierwszy skok prądu i_o w przypadku przebiegu b powinien być również wykonany linią przerywaną.		
64	wiersz 2 g.	этих теорий	обих концепций
64	„ 15 g.	то есть цепи	то есть рассмотривая цепь
64	„ 23 g.	i	i_o
64	„ 26 g.	приближенные результаты	приближенно те же самые результаты
64	„ 29 g.	верхней медианы	медианы верхней
65	„ 2 g.	opproach	approach

POLSKA AKADEMIA NAUK
INSTYTUT PODSTAWOWYCH PROBLEMÓW TECHNIKI

ROZPRAWY ELEKTROTECHNICZNE

TOM VI · ZESZYT 4

KWARTALNIK

PAŃSTWOWE WYDAWNICTWO NAUKOWE
WARSZAWA 1960

ADRES REDAKCJI

Warszawa, Politechnika, Plac Jedności Robotniczej 1
Katedra Teleransmisji Przewodowej, Gmach Mechaniki

PAŃSTWOWE WYDAWNICTWA NAUKOWE
Warszawa, Miodowa 10

Nakład 600 (469 + 131)	Oddano do składania 9.VIII.1960
Ark. wyd. 17,25, ark. druk. 16.	Podpisano do druku w grudniu 1960
Papier druk. sat. kl. III, 80 g, 70×100	Druk ukończono w grudniu 1960
Zamówienie nr 1510/60	Cena zł. 25.— C-36

Drukarnia im. Rewolucji Październikowej, Warszawa

621.3.01/.02

JERZY OSIOWSKI

Znaczenie i zastosowania dystrybucji w teorii obwodów[1)]

Rękopis dostarczono 17.2.1960

Praca składa się z dwóch części. Część pierwsza poświęcona jest wprowadzeniu w podstawy matematyczne teorii dystrybucji. Omówiono najważniejsze pojęcia i własności dotyczące dystrybucji Schwartza i Mikusińskiego ze szczególnym uwzględnieniem przekształcenia Fouriera i Laplace'a dla dystrybucji.

Część druga zawiera przegląd wybranych zagadnień teorii obwodów, traktowanych i rozwiązywanych na gruncie dystrybucji zarówno w dziedzinie czasowej (równania różniczkowe), jak i częstotliwościowej (funkcje przenoszenia).

WSTĘP

Od dawna już znany jest powszechnie fakt, iż w wielu bardzo różnorodnych zagadnieniach matematycznych, fizycznych czy nawet technicznych klasyczne pojęcie funkcji i cały aparat analizy matematycznej na tym pojęciu oparty — są w jakimś stopniu niewystarczające do opisu i badania tych zagadnień, w związku z czym odczuwa się potrzebę rozszerzenia pojęcia funkcji, jakiegoś wyjścia poza krąg możliwości klasycznej analizy.

Nie wnikając w szczegóły i nie chcąc zajmować uwagi czytelnika znanymi przykładami, warto jednak podkreślić zarysowujące się niejako dwa kierunki „krytyki" pojęcia funkcji i rozwoju jego uogólnień.

Pierwszy — można by scharakteryzować jako kierunek matematyczno-teoretyczny. Już w najprostszych zagadnieniach analizy matematycznej spotykamy co krok sytuację polegającą na tym, że dane działanie jest wykonalne tylko przy takich czy innych założeniach ograniczających. Nie każda funkcja ciągła ma pochodną, nie każdy szereg

[1)] Praca ta została oparta na referacie problemowym wygłoszonym na III Symposium Teorii Łączności w Szklarskiej Porębie w dniu 18.9.1959 r.

funkcyjny (zbieżny) można całkować lub różniczkować wyraz po wyrazie, nie każda funkcja całkowalna ma transformatę Fouriera, nie każda funkcja okresowa całkowalna z kwadratem ma szereg Fouriera do niej samej zbieżny, nie zawsze można przestawiać kolejność takich działań, jak przejście do granicy, całkowanie, różniczkowanie itp. itp. Przykładów takich, lub im podobnych, można by mnożyć wiele, sięgając głębiej do poszczególnych dziedzin matematyki. Zagadnienia takie — nazwijmy je zagadnieniami regularnościowymi — obok zasadniczego znaczenia teoretycznego mają również i wyraźny aspekt praktyczny: konieczność stałego kontrolowania przy rozwiązywaniu każdego problemu, czy są spełnione założenia stosowalności użytego działania, metody, wzoru, — czy wielkości, którymi chcemy operować, rzeczywiście istnieją itp. itp.

Wszystkie te problemy „niewykonalności" w klasycznej analizie rodzą niewątpliwie koncepcję, żeby tak uogólnić pojęcie funkcji, aby w tym zbiorze funkcji uogólnionych było np. zawsze wykonalne działanie różniczkowania. Wydaje się, że jest to jeden nurt, z którego wyłoniła się teoria dystrybucji.

Drugi kierunek można by nazwać f i z y c z n o-p r a k t y c z n y m. W wielu problemach fizycznych aparat funkcyjny okazał się niewystarczający do opisu samego zjawiska, a tym samym nie nadający się do jego badania. Znów nie wnikając w znane szczegóły i przykłady przypominamy tylko choćby problem gęstości. O ile przy rozkładzie objętościowym, powierzchniowym czy liniowym masy, gęstość wyraża się funkcją całkowalną, o tyle przy następnym etapie idealizacji geometrycznej — przejściu do masy punktowej — gęstość $\mu(M)$ przestaje być funkcją, bowiem warunki:

$$\int_V \mu(M)\, d\, x_M = m,$$

gdy $M_0 \epsilon V$
oraz

$$\int_V \mu(M)\, d\, x_M = 0,$$

gdy M_0 nie należy do V (M_0 — punkt, w którym znajduje się masa punktowa m), nie mogą być spełnione przez żadną funkcję. Stało się to punktem wyjścia dla wprowadzonej już od dawna a ugruntowanej przez Diraca [6] i związanej zwykle z jego imieniem — dziś już szeroko rozpowszechnionej w wielu rozmaitych problemach (np. [19], [26]) — teorii tzw. funkcji impulsowych (δ—funkcji). Jednakże cała ta teoria, która w fizyce odegrała i odgrywa nadal ogromną rolę, nie może być poprawnie wprowadzona i uzasadniona na gruncie klasycznej analizy i stanowi z matematycznego punktu widzenia jedynie zbiór formalnych reguł postępowania,

prowadzących przy pewnej dozie intuicji i wyczucia do poprawnych rezultatów (analogia do rachunku operatorów Heaviside'a). Dlatego też znaczenie teorii dystrybucji w tym zakresie wydaje się ogromne — daje ona zupełnie ścisłe, jasne podstawy matematyczne oraz wypełnia konkretną treścią pojęciową formalnie ekstrapolowane z analizy reguły działań.

Warto jeszcze kilka słów poświęcić sprawie stosunku rachunku operatorowego (przekształcenia Laplace'a) do teorii dystrybucji — szczególnie w zakresie nas interesującym. Wiadomo, że stosując rachunek operatorowy w zagadnieniach równań różniczkowych, można w wielu przypadkach, w pewnym sensie automatycznie a jednocześnie w sposób poprawny, uniknąć trudności, jakie przy metodach klasycznych wynikają np. przy występowaniu funkcji nieciągłych, ich różniczkowaniu itp., a więc w przypadkach, które — ściśle biorąc — wymagałyby dystrybucyjnego traktowania równania. Wprowadzenie przez van der Pola [25] transformat wielomianowych pozwoliło na formalne rozszerzenie stosowalności rachunku operatorowego poza klasyczny zakres, jednakże ścisłe podstawy matematyczne postępowanie to uzyskało dopiero w oparciu o dystrybucje (por. [16], [30], [8]). W tym sensie w zakresie równań różniczkowych, a więc i w teorii obwodów występują dość ważne powiązania dystrybucji z rachunkiem operatorowym.

Osobne miejsce w tych powiązaniach zajmuje rachunek operatorów Mikusińskiego [22]. Jak wiadomo, dzięki nieistnieniu w dziedzinie funkcji modułu mnożenia splotowego, pojęcie operatora w sensie Mikusińskiego jest istotnym uogólnieniem pojęcia funkcji i obejmuje m. in. tzw. operatory liczbowe, będące odpowiednikami δ-dystrybucji (por. [23]). W tym znaczeniu można więc mówić o „równoległości" (w pewnym zakresie) rachunku operatorów Mikusińskiego i dystrybucji.

W części I pracy rozpatrzone są w dużym skrócie najważniejsze pojęcia i własności dystrybucji w dwóch podstawowych ujęciach: ujęciu funkcjonałowym, pochodzącym od Sobolewa [31] i Schwartza [27], [28], [29], a ostatnio szczegółowo opracowanym przez Gelfanda i Sziłowa [11], [12], [13] (dystrybucje Schwartza), oraz w ujęciu ciągowym opartym na pracach Mikusińskiego [20], [21], [24], a także Korevaara [16] (dystrybucje Mikusińskiego). Aby nie powiększać i tak dość pokaźnej objętościowo tej części, nie zostały już omówione inne koncepcje dystrybucji (Bochner, König, Sikorski i in.) oraz ograniczono się tylko do rozpatrzenia dystrybucji jednej zmiennej.

W części II podane zostały przykłady ujęcia dystrybucyjnego niektórych zagadnień z teorii obwodów, rozwiązywania równań obwodów w dziedzinie dystrybucji, związków pomiędzy charakterystykami układu elektrycznego itp.

CZĘŚĆ I — DYSTRYBUCJE

1. DYSTRYBUCJE JAKO FUNKCJONAŁY LINIOWE

W rozdziale tym rozpatrzymy w skrócie najważniejsze pojęcia i własności dotyczące dystrybucji Schwartza[1]). Konieczne wydaje się przy tym owówienie także kilku pojęć wstępnych.

1.1. Pojęcie funkcjonału

Jednym z podstawowych pojęć analizy funkcjonalnej jest pojęcie operatora, tj. funkcji abstrakcyjnej przyporządkowującej każdemu elementowi x pewnego zbioru X jednoznacznie określony element y innego zbioru Y. Jeśli zbiór Y jest zbiorem liczb, to operator nazywa się funkcjonałem. Szczególne znaczenie mają funkcjonały w przestrzeniach funkcyjnych, tzn. w przypadku, gdy zbiór X jest zbiorem funkcji φ określonej klasy (tzn. $x \equiv \varphi$). Na przykład jeśli jako zbiór X przyjmiemy zbiór wszystkich funkcji $\varphi(t)$ ciągłych w przedziale $[a, b]$, tzn. $X \equiv C\,[a, b]$, to przykładami funkcjonałów w przestrzeni X, które będziemy oznaczać przez f, $f(\varphi)$ lub (f, φ), będą:

$$(f_1, \varphi) = \varphi^2(c) \qquad (a \leqslant c \leqslant b),$$

$$(f_2, \varphi) = \int_a^b \varphi(t)\,dt,$$

$$(f_3, \varphi) = \min_{t \in [a, b]} \varphi(t) \quad \text{itp.}$$

1.2. Dystrybucje jako funkcjonały liniowe w przestrzeni K funkcji podstawowych

Rozpatrywać będziemy funkcje rzeczywiste $\varphi(t)$ spełniające następujące warunki:

1) $\varphi(t)$ jest określona w przedziale $(-\infty, +\infty)$ i posiada w tym przedziale wszystkie pochodne, tzn. $\varphi(t) \in C^\infty$ $(-\infty, +\infty)$;

2) istnieje taki przedział ograniczony $[a, b]$, na zewnątrz którego: $\varphi(t) \equiv 0$. Przedział ten może być oczywiście różny dla różnych funkcji $\varphi \in K$. Funkcje $\varphi(t)$ spełniające te warunki będziemy nazywać funk-

[1]) Poza wymienioną już we wstępie literaturą źródłową, dystrybucje Schwartza były rozpatrywane pod kątem widzenia potrzeb i ewentualnych zastosowań technicznych również w pracach: [5], [8], [9], [32].

cjami podstawowymi, a ich zbiór — przestrzenią podstawową lub przestrzenią K (por. [11]).

Przykładem funkcji podstawowych są np. funkcje:

$$\varphi(t) = \begin{cases} e^{-\left(\frac{1}{t-a}+\frac{1}{b-t}\right)}; & a<t<b \\ 0 \qquad\qquad\quad ; & t \leqslant a, \quad t \geqslant b \end{cases} \tag{1}$$

Zauważamy, że przestrzeń K jest liniowa, tzn. jeśli $\varphi_1 \in K$, $\varphi_2 \in K$, to dla dowolnych liczb rzeczywitych k_1 i k_2 : $k_1 \varphi_1 + k_2 \varphi_2 \in K$. Z podanego określenia wynika także wprost, że jeśli $\varphi \in K$, to przy dowolnym n także $\varphi^{(n)} = \frac{d^n \varphi}{d t^n} \in K$, ale np. do K nie zawsze już należą funkcje pierwotne $\int \varphi(t)\, dt$.

Zbieżność w przestrzeni podstawowej K określa się następująco [11]. Mówimy, że ciąg $\{\varphi_n\}$, $\varphi_n \in K$ jest zbieżny do zera w przestrzeni K, jeśli istnieje wspólny dla wszystkich φ_n przedział $[a, b]$ taki, że:

$$\begin{aligned} &1)\ \varphi_n(t) \equiv 0, \quad \text{gdy} \quad t<a \quad \text{lub} \quad t>b \\ &2)\ \varphi_n^{(k)}(t) \Rightarrow 0, \quad n \to \infty, \quad t \in [a,b] \end{aligned} \tag{2}$$

dla $k = 0, 1, 2 \ldots$ Drugi warunek oznacza, że funkcje φ_n oraz ich pochodne $\varphi_n^{(k)}$ dowolnego k-tego rzędu dążą w przedziale $[a, b]$ jednostajnie do zera.

W przestrzeni K funkcji podstawowych φ można wprowadzić funkcjonał liniowy $f \equiv (f, \varphi)$, tzn. taką zależność, która każdej funkcji $\varphi \in K$ przyporządkowuje liczbę (f, φ) i spełnia warunki:

1) jeśli $\varphi_1 \in K$, $\varphi_2 \in K$ oraz k_1 i k_2 są dowolnymi liczbami rzeczywistymi, to:

$$(f, k_1 \varphi_1 + k_2 \varphi_2) = k_1 (f, \varphi_1) + k_2 (f, \varphi_2), \tag{3}$$

2) jeśli ciąg $\{\varphi_n\}$, $\varphi_n \in K$ dąży do zera w przestrzeni K, to: $(f, \varphi_n) \to 0$ (warunek ciągłości funkcjonału).

Dystrybucją (w sensie Schwartza) będziemy nazywać każdy funkcjonał liniowy w przestrzeni podstawowej K.

Dystrybucje (w sensie Schwartza) definiuje się także jako funkcjonały liniowe w innych przestrzeniach funkcyjnych, które mogą być rozmaicie definiowane [12]. Ma to duże znaczenie i teoretyczne i praktyczne. W naszych rozważaniach ograniczymy się tylko do najprostszego przypadku dystrybucji w przestrzeni K (pewien wyjątek stanowi rozdz. 1.9, dotyczący przekształcenia Fouriera).

1.3. Przykłady

Rozpatrzymy teraz szczegółowiej kilka ważnych przykładów.

a) Niech $f(t)$ oznacza dowolną funkcję określoną w przedziale $(-\infty, +\infty)$ i lokalnie całkowalną. Utwórzmy funkcjonał:

$$(f, \varphi) = \int_{-\infty}^{+\infty} f(t)\,\varphi(t)\,dt, \tag{4}$$

który jak łatwo sprawdzić, jest funkcjonałem liniowym w przestrzeni K, a więc — dystrybucją. Wobec własności funkcji φ, całkowanie odbywa się przy tym zawsze w pewnym przedziale ograniczonym.

Dystrybucje o postaci (4) nazywamy dystrybucjami regularnymi lub dystrybucjami typu funkcji.

Wynika stąd, że każda funkcja lokalnie całkowalna określa jednoznacznie dystrybucję regularną (4). Jest i odwrotnie — jeśli dany jest funkcjonał typu (4), to z dokładnością do wartości funkcji w zbiorze miary zero jest tym samym jednoznacznie wyznaczona funkcja $f(t)$. Inaczej, jeśli:

$$\int_{-\infty}^{+\infty} f_1(t)\,\varphi(t)\,dt = \int_{-\infty}^{+\infty} f_2(t)\,\varphi(t)\,dt$$

dla każdej funkcji $\varphi \in K$, to $f_1(t) = f_2(t)$ prawie wszędzie, tzn. co najwyżej z wyjątkiem zbioru miary zero.

Jeśli utożsamimy więc wszystkie funkcje równe prawie wszędzie i będziemy je traktować jako jedną funkcję, to każdej funkcji lokalnie całkowalnej $f(t)$ odpowiada jednoznacznie dystrybucja regularna (4) i odwrotnie — każda dystrybucja typu (4) wyznacza jednoznacznie pewną funkcję lokalnie całkowalną.

Można więc mówić o identyfikacji dystrybucji regularnych i funkcji lokalnie całkowalnych.

b) Rozpatrzymy funkcjonał:

$$(f, \varphi) = \varphi(a), \tag{5}$$

gdzie a jest ustalonym punktem przedziału $(-\infty, +\infty)$. Jest to funkcjonał liniowy, a więc dystrybucja, bowiem:

$$(f, k_1\varphi_1 + k_2\varphi_2) = k_1\varphi_1(a) + k_2\varphi_2(a) = k_1(f, \varphi_1) + k_2(f, \varphi_2)$$

oraz

$$(f, \varphi_n) = \varphi_n(a) \to 0,$$

gdy $\{\varphi_n\}$ dąży do zera w przestrzeni K.

Dytrybucja ta nie jest jednak regularna, bowiem funkcjonał (5) można przedstawić w postaci całki Stieltjesa:

$$\int_{-\infty}^{+\infty} \varphi(t)\, d\,\mathbf{1}(t-a) = \varphi(a), \tag{6}$$

która nie da się sprowadzić do postaci (4). Dystrybucje, które nie są regularne, nazywać będziemy dystrybucjami s i n g u l a r n y m i. Istnienie dystrybucji singularnych właśnie sprawia, iż dystrybucje stanowią istotne uogólnienie funkcji (lokalnie całkowalnych).

Dystrybucja singularna (5), zwana δ - d y s t r y b u c j ą lub d y s t r y b u c j ą D i r a c a jest pełnym odpowiednikiem pseudofunkcji Diraca $\delta(t-a)$, a związek definicyjny (5) jest, porównując do (4), odpowiednikiem tzw. własności filtrującej $\delta(t-a)$, tj. własności wyrażonej wzorem:

$$\int_{-\infty}^{+\infty} \delta(t-a)\,\varphi(t)\,dt = \varphi(a). \tag{7}$$

Dystrybucję Diraca zapisuje się zwykle w postaci

$$(\delta_a, \varphi) = \varphi(a). \tag{8}$$

lub

$$\big(\delta(t-a), \varphi\big) = \varphi(a).$$

Przez analogię do zapisu (4) dystrybucji regularnej często — szczególnie w zastosowaniach — pisze się u m o w n i e:

$$(f, \varphi) = \int_{-\infty}^{+\infty} f(t)\,\varphi(t)\,dt,$$

niezależnie od tego, czy (f, φ) jest dystrybucją regularną, czy singularną. Pozwala to np. na zapisywanie δ-dystrybucji w postaci:

$$(\delta_a, \varphi) = \int_{-\infty}^{+\infty} \delta_a(t)\,\varphi(t)\,dt = \int_{-\infty}^{+\infty} \delta(t-a)\,\varphi(t)\,dt, \tag{9}$$

a więc na zachowaniu pewnego wygodnego i przyjętego w praktyce formalizmu Dirakowskiej teorii funkcji impulsowych przy nadaniu mu jednak zupełnie ściśle określonej treści dystrybucyjnej.

c) Pozostawiamy czytelnikowi do sprawdzenia, że funkcjonały:

$$\int_{-\infty}^{+\infty} \varphi^2(t)\,dt; \quad \int_{-\infty}^{+\infty} f(t)\,\varphi(t)\,dt + 1; \quad \max_{t\,\epsilon\,(-\infty\,+\infty)} |\varphi(t)|; \quad \sqrt[3]{\varphi(0)}$$

nie są dystrybucjami, bowiem nie są funkcjonałami liniowymi w przestrzeni K.

1.4. Podstawowe działania na dystrybucjach

a) Równość dystrybucji. Mówimy, że dwie dystrybucje f_1 i f_2 są identyczne, jeśli dla każdej funkcji podstawowej φ zachodzi równość:

$$(f_1, \varphi) = (f_2, \varphi). \tag{10}$$

Określenie to pociąga za sobą omówioną już w poprzednim punkcie identyfikację na gruncie dystrybucji funkcji lokalnie całkowalnych równych prawie wszędzie, gdyż dla takich funkcji $f_1(t)$ i $f_2(t)$ mamy

$$(f_1, \varphi) = \int_{-\infty}^{+\infty} f_1(t)\,\varphi(t)\,dt = \int_{-\infty}^{+\infty} f_2(t)\,\varphi(t)\,dt = (f_2, \varphi).$$

b) Suma dystrybucji. Jeśli f_1 i f_2 są dystrybucjami, to przez ich sumę $f_1 + f_2$ rozumiemy dystrybucję:

$$(f_1 + f_2, \varphi) \overset{df}{=} (f_1, \varphi) + (f_2, \varphi). \tag{11}$$

Łatwo sprawdzić, że jest to istotnie funkcjonał liniowy w przestrzeni K, a więc dystrybucja. Jeśli f_1 i f_2 są dystrybucjami regularnymi, to ich suma jest oczywiście też dystrybucją regularną, bowiem wówczas:

$$(f_1 + f_2, \varphi) = (f_1, \varphi) + (f_2, \varphi) = \int_{-\infty}^{+\infty} f_1(t)\,\varphi(t)\,dt +$$

$$+ \int_{-\infty}^{+\infty} f_2(t)\,\varphi(t)\,dt = \int_{-\infty}^{+\infty} [f_1(t) + f_2(t)]\,\varphi(t)\,dt. \tag{12}$$

c) Mnożenie dystrybucji przez liczbę. Jeśli f jest dystrybucją oraz k — dowolną liczbą rzeczywistą, to dystrybucję $k \cdot f$ definiujemy jako:

$$(kf, \varphi) \equiv k(f, \varphi) \overset{df}{=} (f, k\varphi), \tag{13}$$

przy czym znów jest łatwo sprawdzić, że jest to istotnie funkcjonał liniowy w przestrzeni K, a więc dystrybucja. Jeśli f jest dystrybucją regularną, to kf też jest oczywiście dystrybucją regularną, bowiem:

$$(kf, \varphi) = (f, k\varphi) = \int_{-\infty}^{+\infty} f(t)\,[k\varphi(t)]\,dt = \int_{-\infty}^{+\infty} [kf(t)]\,\varphi(t)\,dt. \tag{14}$$

Własności b) i c) świadczą o tym, że zbiór dystrybucji określonych w przestrzeni K jest przestrzenią liniową. Jest to tzw. przestrzeń sprzężona z K.

d) Mnożenie dystrybucji przez funkcję klasy C^∞

$(-\infty, +\infty)$. Załóżmy, że $a(t)$ jest funkcją klasy C^∞ $(-\infty, +\infty)$, tzn. posiada wszystkie pochodne w przedziale $(-\infty, +\infty)$. Iloczyn af funkcji $a(t)$ i dytrybucji f można wówczas zdefiniować analogicznie jak w przypadku mnożenia przez liczbę, tzn.

$$(af, \varphi) \overset{df}{=} (f, a\varphi), \tag{15}$$

przy czym (sprawdzamy!) jest to istotnie funkcjonał liniowy w K, a więc dystrybucja.

Jako przykład rozpatrzmy iloczyn $a\delta_a$. Mamy:

$$(a\delta_a, \varphi) = (\delta_a, a\varphi) = a(a)\varphi(a) = a(a)(\delta_a, \varphi) \tag{16}$$

lub w tradycyjnym zapisie

$$a(t)\delta(t-a) = a(a)\delta(t-a). \tag{17}$$

Jest to znana zależność (często stosowane są jej przypadki szczególne, np. $a(t)\,\delta(t) = a(0)\,\delta(t)$; $t^n\delta(t) = 0$, n — naturalne itp. która tutaj uzyskała pełne uzasadnienie.

Zwracamy uwagę, że założenia o istnieniu w $(-\infty, +\infty)$ wszystkich pochodnych funkcji $a(t)$ nie można przy stosowaniu definicji (15) osłabić, gdyż tylko przy tym założeniu $a\varphi \in K$ [tzn. $a(t)$ jest multiplikatorem w przestrzeni K, a więc $(f, a\varphi)$ jest funkcjonałem w przestrzeni K (liniowość łatwo sprawdzić).

Jeśli f jest dystrybucją regularną, to af jest też oczywiście dystrybucją regularną, bowiem

$$(af, \varphi) = (f, a\varphi) = \int_{-\infty}^{+\infty} f(t)\,[a(t)\varphi(t)]\,dt = \int_{-\infty}^{+\infty} [a(t)f(t)]\,\varphi(t)\,dt. \tag{18}$$

1.5 Różniczkowanie i całkowanie dystrybucji

Załóżmy, że $f(t)$ jest funkcją klasy $C^1(-\infty, +\infty)$. Można więc rozpatrywać dystrybucje regularne (f, φ) i (f', φ), przy czym otrzymujemy przez całkowanie przez części prosty związek:

$$(f', \varphi) = \int_{-\infty}^{+\infty} f'(t)\varphi(t)\,dt = f(t)\varphi(t)\Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} f(t)\varphi'(t)\,dt =$$

$$= -\int_{-\infty}^{+\infty} f(t)\varphi'(t)\,dt = (f, -\varphi'). \tag{19}$$

Jeśli (f, φ) jest dowolną dystrybucją (niekoniecznie regularną), to łatwo sprawdzić, że $(f, -\varphi')$ jest też dystrybucją, wobec tego można

w ogólnym przypadku zdefiniować pochodną f' dystrybucji f w postaci funkcjonału:

$$(f', \varphi) \overset{df}{=} (f, -\varphi') = -(f, \varphi'), \tag{20}$$

przy czym, jak z tego wynika, pochodna (20) istnieje dla każdej dystrybucji. Ponieważ jest ona także dystrybucją, a więc można ją dalej według tej samej reguły różniczkować. Mamy więc o fundamentalnym znaczeniu fakt — w dziedzinie dystrybucji różniczkowanie jest zawsze wykonalne i każda dystrybucja ma pochodną dowolnie wysokiego rzędu, będącą też dystrybucją i wyrażającą się funkcjonałem:

$$(f^{(n)}, \varphi) \overset{df}{=} (f, (-1)^n \varphi^{(n)}). \tag{21}$$

W szczególności więc każda funkcja lokalnie całkowalna traktowana jako dystrybucja regularna posiada w dziedzinie dystrybucji pochodne wszystkich rzędów, które nie są już, ogólnie biorąc, dystrybucjami regularnymi. Taką pochodną funkcji lokalnie całkowalnej traktowanej jako dystrybucja nazywa się często pochodną dystrybucyjną tej funkcji. Jeśli funkcja taka posiada prawie wszędzie pochodną lokalnie całkowalną, to — jak zobaczymy na przykładach — pochodna dystrybucyjna może być z nią identyczna, ale może również tak nie być! Ponadto pochodna dystrybucyjna funkcji lokalnie całkowalnej, zgodnie z definicją (20), istnieje zawsze, podczas gdy funkcja może zwykłej pochodnej w ogóle nie posiadać.

W dziedzinie dystrybucji zachowują się reguły różniczkowania; jeśli f i g są dystrybucjami, k — stałą oraz $\alpha(t) \in C^\infty(-\infty, +\infty)$, to słuszne są wzory:

$$(f+g)' = f' + g'; (kf)' = kf'; (\alpha f)' = \alpha' f + \alpha f'. \tag{22}$$

Rozpatrzymy teraz kilka prostych przykładów.

a) Jeśli $f(t) \in C^1(-\infty, +\infty)$, to zgodnie z (19) i (20) widzimy, że pochodną dystrybucyjną (f', φ) można utożsamić ze zwykłą pochodną $f'(t)$, tzn.

$$(f', \varphi) = \int_{-\infty}^{+\infty} f'(t)\,\varphi(t)\,dt. \tag{23}$$

Można wykazać, że zachodzi to dla szerszej klasy funkcji $f(t)$; ogólnie pochodną (f', φ) można wtedy i tylko wtedy przedstawić w postaci (23), tzn. można pochodną dystrybucyjną f' utożsamić z (istniejącą prawie wszędzie) zwykłą pochodną $f'(t)$, gdy funkcja $f(t)$ jest absolutnie ciągła (por. [10]).

b) Niech $f(t) = \mathbf{1}(t)$. Pochodna w zwykłym sensie istnieje prawie wszędzie z wyjątkiem punktu $t = 0$ i jest równa $f'(t) = 0$ (a więc jest funkcją lokalnie całkowalną). Nie jest ona jednak identyczna z pochodną dystrybucyjną, bowiem mamy:

$$(\mathbf{1}', \varphi) = -(\mathbf{1}, \varphi') = -\int_{-\infty}^{+\infty} \mathbf{1}(t)\,\varphi'(t)\,dt = -\int_{0}^{\infty} \varphi'(t)\,dt = \varphi(0) = (\delta_0, \varphi),$$

gdzie $\delta_0 \equiv \delta(t)$,
a więc

$$\mathbf{1}'(t) = \delta(t), \tag{24}$$

tzn. pochodną dystrybucyjną funkcji jednostkowej jest δ-dystrybucja.

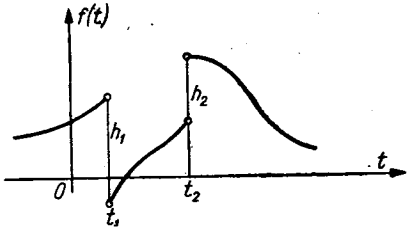
Dystrybucji $(\mathbf{1}', \varphi) = (\delta, \varphi)$ nie można przedstawić w postaci (4), gdyż jest ona dystrybucją singularną. Można ją natomiast przedstawić w postaci całki Stieltjesa:

$$(\delta, \varphi) = (\mathbf{1}', \varphi) = \int_{-\infty}^{+\infty} \varphi(t)\,d\,\mathbf{1}(t).$$

Ogólnie, do tego, aby dystrybucję $(f'\,\varphi)$ można było przedstawić w postaci całki Stieltjesa:

$$(f', \varphi) = \int_{-\infty}^{+\infty} \varphi(t)\,d\,f(t), \tag{25}$$

trzeba i wystarcza, aby $f(t)$ była funkcją z ograniczoną zmiennością. Dystrybucje typu (25) można nazwać dystrybucjami typu miary i można je utożsamiać z miarami Stieltjesa.



Rys. 1. Wykres funkcji przedziałami regularnej

c) Funkcja $f(t)$ przedziałami regularna (rys. 1) posiada prawie wszędzie (z wyjątkiem punktów t_k, w których funkcja ma skoki) pochodną $f'(t)$, która jest funkcją lokalnie całkowalną. Nie jest ona identyczna jednak z pochodną dystrybucyjną, gdyż łatwo można wykazać, że:

$$f'_{dystr} = f'(t) + \sum_k h_k\,\delta(t - t_k), \tag{26}$$

gdzie: $h_k = f(t_k + 0) - f(t_k - 0)$ oznaczają wartości skoków funkcji $f(t)$ w w punktach t_k. Symbol: f'_{dystr} należy rozumieć jako pochodną dystrybucji (f, φ).

d) Mamy:

$$(\delta'_a, \varphi) = -(\delta_a, \varphi') = -\varphi'(a) \tag{27}$$

lub ogólniej (n naturalne):

$$(\delta_a^{(n)}, \varphi) = (-1)^n (\delta_a, \varphi^{(n)}) = (-1)^n \varphi^{(n)}(a). \tag{28}$$

Związki te określają dystrybucje pochodne względem δ-dystrybucji i w zapisie tradycyjnym przybierają znaną postać:

$$\int_{-\infty}^{+\infty} \delta'(t-a)\,\varphi(t)\,dt = -\varphi'(a),$$

$$\int_{-\infty}^{+\infty} \delta^{(n)}(t-a)\,\varphi(t)\,dt = (-1)^n \varphi^{(n)}(a).$$

Funkcjonały (27) i (28) nie dadzą się już przedstawić w postaci całek Stieltjesa.

e) Istnieją, jak wiadomo, funkcje wszędzie ciągłe i nie posiadające nigdzie pochodnej. Na przykład (Weierstrass) funkcja:

$$f(t) = \sum_{k=1}^{\infty} \frac{1}{2^k} \sin 20^k t.$$

W dziedzinie dystrybucji posiada ona jednak, jak każda dystrybucja, pochodną określoną funkcjonałem:

$$(f', \varphi) = -(f, \varphi') = -\int_{-\infty}^{+\infty} f(t)\,\varphi'(t)\,dt.$$

Jeśli dystrybucja f jest pochodną dystrybucji g, tzn. $g' = f$, to dystrybucję g nazywamy c a ł k ą (dystrybucją pierwotną) d y s t r y b u c j i f i piszemy:

$$g = \int f\,dt. \tag{29}$$

Ponieważ

$$(g', \varphi) = (g, -\varphi'),$$

wobec tego funkcjonał g jest określony równością:

$$(g, -\varphi') = (f, \varphi). \tag{30}$$

Można wykazać, że dla każdej dystrybucji f istnieje dystrybucja pierwotna g.

1.6. Niektóre zagadnienia zbieżności w dziedzinie dystrybucji

Omówimy pokrótce najważniejsze pojęcia dotyczące zbieżności w dziedzinie dystrybucji.

Będziemy mówić, że ciąg dystrybucji f_1, f_2 ... f_n ... jest zbieżny do dystrybucji f, jeśli dla każdej funkcji podstawowej φ zachodzi równość:

$$\lim_{n\to\infty}(f_n, \varphi) = (f, \varphi). \tag{31}$$

Piszemy wówczas, że: $f_n \to f$ lub: $\lim_{n\to\infty} f_n = f$. Analogicznie, jeśli f_x jest dystrybucją zależną od parametru x w pewnym przedziale, to mówimy, że dystrybucja f jest granicą f_x przy $x \to x_0$, jeśli dla każdej funkcji podstawowej φ zachodzi związek:

$$\lim_{x\to x_0}(f_x, \varphi) = (f, \varphi). \tag{32}$$

Piszemy wówczas, że:

$$f_x \to f \text{ przy } x \to x_0$$

lub

$$\lim_{x\to x_0} f_x = f,$$

Wykażemy teraz, że jeśli: $\lim_{n\to\infty} f_n = f$, to:

$$\lim_{n\to\infty} f_n' = f', \tag{33}$$

tzn., że ze zbieżności ciągu dystrybucji f_n do dytrybucji f wynika zbieżność ciągu pochodnych do pochodnej f'. Istotnie, bowiem dla każdej funkcji podstawowej φ mamy:

$$(f_n', \varphi) = (f_n, -\varphi') \to (f, -\varphi') = (f', \varphi)$$

przy $n \to \infty$.

Rozpatrzmy kilka przykładów.

a) Załóżmy, że ciąg funkcji $f_n(t)$ lokalnie całkowalnych jest w przedziale $(-\infty, +\infty)$ zbieżny do funkcji $f(t)$ lokalnie całkowalnej. Powstaje pytanie, czy ciąg dystrybucji regularnych f_n jest zbieżny do dystrybucji regularnej f, tzn. czy ciąg $f_n(t)$ jest także dystrybucyjnie zbieżny do $f(t)$. Postaramy się sprawę tę wyjaśnić.

Zgodnie z definicją ciąg $\{f_n(t)\}$ będzie dystrybucyjnie zbieżny do $f(t)$, jeśli dla każdej funkcji podstawowej:

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} f_n(t)\,\varphi(t)\,dt = \int_{-\infty}^{+\infty} f(t)\,\varphi(t)\,dt, \tag{34}$$

do czego nie wystarcza jednak zwykła zbieżność funkcji $f_n(t)$ do

$f(t)$, a więc ogólnie biorąc, odpowiedź na postawione pytanie jest n e g a t y w n a [por. przykład podany w punkcie c)].

Jednakże równość (34) będzie spełniona przy pewnych dodatkowych założeniach. Na przykład równość (34) zachodzi dla każdej funkcji $\varphi \in K$ (a więc $f_n \to f$ dystrybucyjnie), jeśli ciąg funkcji $f_n(t)$ dąży j e d n o s t a j n i e do $f(t)$ w każdym przedziale ograniczonym. Równość ta zachodzi także (por. [11]), gdy $f_n(t)$ dąży do $f(t)$ p r a w i e w s z ę d z i e oraz ponadto:

$$|f_n(t)| < K \tag{35}$$

dla wszystkich n i t, tzn. wszystkie wyrazy ciągu $\{f_n(t)\}$ są wspólnie ograniczone.

Dla przykładu niech:

$$f_n(t) = \frac{1}{n} \sin(n^2 t).$$

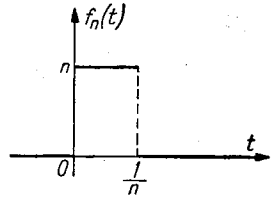
Ciąg ten jest jednostajnie zbieżny do zera w każdym przedziale ograniczonym (bowiem $|f_n(t)| \leqslant \frac{1}{n} \to 0$), a więc

$$\lim_{n \to \infty} f_n(t) = 0$$

zarówno w zwykłym, jak i dystrybucyjnym sensie. Jednakże ciąg pochodnych:

$$\psi_n(t) = f_n'(t) = n \cos(n^2 t),$$

który jest rozbieżny w zwykłym sensie, jest — zgodnie z (33) — d y-



Rys. 2. Wykres funkcji $f_n(t)$ z przykładu b

s t r y b u c y j n i e z b i e ż n y d o z e r a! To samo dotyczyłoby ciągów utworzonych z wyższych pochodnych funkcji $f_n(t)$.

b) Niech (rys. 2):

$$f_n(t) = \begin{cases} n; & 0 < t < \frac{1}{n} \\ 0; & t < 0,\ t > \frac{1}{n}. \end{cases} \tag{36}$$

Rozpatrzmy granicę ciągu dystrybucji regularnych (f_n, φ) przy $n \to \infty$. Mamy:

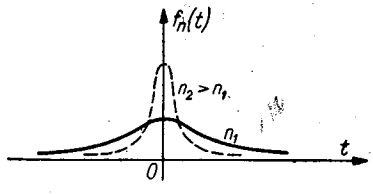
$$(f_n, \varphi) = \int_{-\infty}^{+\infty} f_n(t)\,\varphi(t)\,dt = n\int_{0}^{\frac{1}{n}} \varphi(t)\,dt = \varphi(t_n),$$

gdzie $0 < t_n < \frac{1}{n}$. W granicy otrzymujemy:

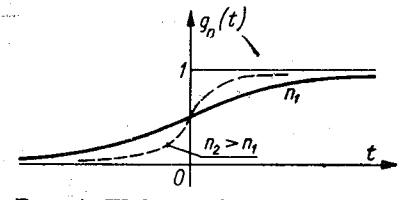
$$\lim_{n\to\infty}(f_n, \varphi) = \lim_{n\to\infty}\varphi(t_n) = \varphi(0) = (\delta, \varphi),$$

a więc granicą dystrybucyjną ciągu funkcji $f_n(t)$ jest δ-dystrybucja, tzn.

$$\lim_{n\to\infty} f_n(t) = \delta(t). \tag{37}$$



Rys. 3. Wykresy funkcji $f_n(t)$ z przykładu c



Rys. 4. Wykresy funkcji $g_n(t)$ z przykładu c

c) Niech (rys. 3):

$$f_n(t) = \frac{1}{\pi}\cdot\frac{n}{n^2t^2+1}. \tag{38}$$

Oznaczamy:

$$g_n(t) = \int_{-\infty}^{t} f_n(\tau)\,d\tau = \frac{n}{\pi}\int_{-\infty}^{t}\frac{d\tau}{n^2\tau^2+1} = \frac{1}{\pi}\operatorname{arctg} nt + \frac{1}{2}.$$

Zauważamy, że (rys. 4):

$$\lim_{n\to\infty} g_n(t) = \lim_{n\to\infty}\left(\frac{1}{\pi}\operatorname{arctg} nt + \frac{1}{2}\right) = \begin{cases} 1; & t>0 \\ \frac{1}{2}; & t=0 \\ 0; & t<0 \end{cases}$$

a więc:

$$\lim_{n\to\infty} g_n(t) = \mathbf{1}(t),$$

przy czym równość ta zachodzi zarówno w zwykłym, jak i dystrybucyjnym sensie (wynika to stąd, że $|g_n(t)| < 1$ oraz $g_n(t) \to \mathbf{1}(t)$ przy $t \neq 0$, a więc prawie wszędzie [por. pkt 1.6a)]. Wobec tego zgodnie z (33):

$$\lim_{n\to\infty} f_n = \lim_{n\to\infty} g_n' = \delta(t),$$

tzn. ciąg (38) jest dystrybucyjnie zbieżny do δ-dystrybucji. Różniczkując (38):

$$\psi_n(t) = f_n'(t) = -\frac{2}{\pi} \cdot \frac{n^3 t}{(n^2 t^2 + 1)^2} \tag{39}$$

dostaniemy zgodnie z (33):

$$\lim_{n \to \infty} \psi_n = \delta'(t), \tag{40}$$

a więc ciąg (39) jest dystrybucyjnie zbieżny do dystrybucji δ'. Można oczywiście takie postępowane przedłużać dalej.

Zauważmy także, że w zwykłym sense ciąg (39) jest przy $n \to \infty$ zbieżny do zera, bowiem

$$\lim_{n \to \infty} \psi_n(t) = -\frac{2}{\pi} \lim_{\varepsilon \to 0} \frac{n^3 t}{(n^2 t^2 + 1)^2} = 0 \tag{41}$$

dla każdego t (!).

W podanych przykładach występowały ciągi funkcji lokalnie całkowalnych zbieżne dystrybucyjne do dystrybucji δ lub jej pochodnych. Ciągi takie, tzw. ciągi „funkcji aproksymujących", stanowią, jak wiadomo, podstawę całej tzw. „aproksymacyjnej" teorii δ-funkcji, traktowanych jako ich „granice" por. [26]. Jednakże wzory (39, (40) i (41) dobitnie świadczą o sprzecznościach, na jakie są skazane z góry próby wprowadzenia δ-funkcji przy pomocy tylko aparatu klasycznej analizy.

Nadmieniamy także, iż zachodzi ogólna własność polegająca na tym, że [11] każdą dystrybucję można przedstawić jako granicę dystrybucji regularnych.

Kilka słów o szeregach w dziedzinie dystrybucji. Mówimy, że szereg $f_1 + f_2 + f_n + \ldots$, którego wyrazami są dystrybucje, jest zbieżny do dystrybucji s, jeśli $s_n \to s$, gdzie:

$$s_n = \sum_{k=1}^{n} f_k$$

jest n-tą sumą cząstkową tego szergu, i piszemy:

$$s = \sum_{k=1}^{\infty} f_k.$$

Z twierdzenia o różniczkowaniu ciągu — wzór (33) — wynika wprost, że jeśli:

$$s = \sum_{k=1}^{\infty} f_k,$$

to

$$s' = \sum_{k=1}^{\infty} f'_k,$$

a więc w dziedzinie dystrybucji każdy szereg zbieżny można różniczkować wyraz po wyrazie.

1.7. Dystrybucje rzędu skończonego

Ważne teoretycznie i praktycznie jest pojęcie rzędu dystrybucji.

Mówimy [15], że dystrybucja f jest rzędu skończonego, jeśli istnieje taka funkcja $F(t)$ lokalnie całkowalna w przedziale $(-\infty, +\infty)$ i taka liczba całkowita nieujemna r, że:

$$F^{(r)} = f. \tag{42}$$

Symbol $F^{(r)}$ oznacza tu r-tą pochodną dystrybucyjną funkcji $F(t)$, a równość jest rozumiana w sensie równości dystrybucji — por. (10).

Najmniejszą z liczb r, dla których można dobrać funkcję $F(t)$ spełniającą (42), nazywa się rzędem dystrybucji f.

Wobec zależności (24) otrzymujemy np., że δ-dystrybucja jest dystrybucją rzędu pierwszego i, analogicznie, $\delta^{(n)}$ — dystrybucją rzędu n-tego.

Z równości (42) wynika od razu, że każda dystrybucja rzędu skończonego jest pochodną dystrybucyjną określonego rzędu funkcji ciągłej w przedziale $(-\infty, +\infty)$.

Dystrybucje rzędu skończonego odgrywają bardzo ważną rolę zarówno w teorii, jak i w zastosowaniach, jednakże nie wszystkie dystrybucje są rzędu skończonego. Jako przykład dystrybucji rzędu nieskończonego może służyć dystrybucja określona jako suma szeregu dystrybucyjnego:

$$s = \sum_{k=0}^{\infty} \delta_k^{(k)}, \tag{43}$$

Pokażemy, że wzór (43) istotnie określa dystrybucję. Oznaczając:

$$s_n = \sum_{k=0}^{n} \delta_k^{(k)}$$

mamy dla każdej funkcji podstawowej φ zgodnie z (28):

$$(s_n, \varphi) = \sum_{k=0}^{n} (\delta_k^{(k)}, \varphi) = \sum_{k=0}^{n} (-1)^k \varphi^{(k)}(k)$$

oraz przy $n \to \infty$:

$$\lim_{n\to\infty}(s_n, \varphi) = \sum_{k=0}^{\infty} (-1)^k \varphi^{(k)}(k).$$

Ponieważ zgodnie z własnością definicyjną każda funkcja podstawowa φ zeruje się wraz ze wszystkimi swoimi pochodnymi na zewnątrz pewnego przedziału ograniczonego $[a, b]$, więc dla każdej funkcji φ występujący w ostatnim wzorze szereg redukuje się do sumy skończonej. Zgodnie z definicją zbieżności szeregu dystrybucji otrzymujemy zatem, że szereg (43) jest dystrybucyjnie zbieżny i określa dystrybucję:

$$(s, \varphi) = \sum_{k=0}^{N_\varphi} (-1)^k \varphi^{(k)}(k), \tag{44}$$

gdzie N_φ jest największą liczbą naturalną leżącą we wspomnianym już przedziale $[a, b]$, ogólnie biorąc, różnym dla różnych funkcji podstawowych φ [Sprawdzenie liniowości i ciągłości funkcjonału (44) pozostawiamy Czytelnikowi].

1.8. Splot w dziedzinie dystrybucji

Rozpatrzmy najpierw przypadek funkcji. Niech $f(t)$ i $g(t)$ oznaczają funkcje bezwzględnie całkowalne w przedziale $(-\infty, +\infty)$. Wówczas ich splot:

$$f \star g = \int_{-\infty}^{+\infty} f(\tau)\, g(t-\tau)\, d\tau = \int_{-\infty}^{+\infty} f(t-\tau)\, g(\tau)\, d\tau \tag{45}$$

jest też funkcją bezwzględnie całkowalną. Dystrybucja $(f*g, \varphi)$ jest więc regularna i może być przedstawiona w postaci:

$$(f \star g, \varphi) = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} f(\tau)\, g(t-\tau)\, d\tau\right] \varphi(t)\, dt =$$

$$= \int_{-\infty}^{+\infty} f(\tau) \left[\int_{-\infty}^{+\infty} g(t-\tau)\, \varphi(t)\, dt\right] d\tau = \int_{-\infty}^{+\infty} f(\tau) \left[\int_{-\infty}^{+\infty} g(t)\, \varphi(t+\tau)\, dt\right] d\tau. \tag{46}$$

Wprowadzimy teraz następujące określenie. Dystrybucję g nazywamy multiplikatorem splotowym w przestrzeni K, jeśli dla każdej (ustalonej) funkcji $\varphi \in K$ funkcjonał:

$$\big(g, \varphi(t+\tau)\big) = \Psi(\tau) \tag{47}$$

jest funkcją podstawową zmiennej τ, tzn. $\Psi \in K$.

Jeśli g jest przy tym funkcjonałem regularnym, to:

$$\Psi(\tau) = \big((g, \varphi(t+\tau)\big) = \int_{-\infty}^{+\infty} g(t)\,\varphi(t+\tau)\,dt,$$

a więc zależność (46) można wówczas przepisać w postaci:

$$(f \star g, \varphi) = \int_{-\infty}^{+\infty} f(\tau)\,\Psi(\tau)\,d\tau = (f, \Psi), \tag{48}$$

gdzie funkcja Ψ jest określona wzorem (47).

Zależność (48) może być teraz użyta do określenia splotu dwóch dystrybucji [33].

Niech f, g — dystrybucje, przy czym g jest multiplikatorem splotowym w przestrzeni K. Wówczas s p l o t e m $f*g$ dystrybucji f i g nazywamy dystrybucję:

$$(f \star g, \varphi) \overset{df}{=} (f, \Psi), \tag{49}$$

gdzie Ψ jest funkcją podstawową określoną przez (47).

Zarówno δ-dystrybucja, jak wszytkie jej pochodne są multiplikatorami splotowymi w przestrzeni K, bowiem (dla ustalonego u):

$$\big(\delta_u, \varphi(t+\tau)\big) = \varphi(u+\tau),$$
$$\big(\delta'_u, \varphi(t+\tau)\big) = -\varphi'(u+\tau),$$
$$\big(\delta_u^{(n)}, \varphi(t+\tau)\big) = (-1)^n \varphi^{(n)}(u+\tau),$$

przy czym, prawe strony tych równości są funkcjami podstawowymi zmiennej τ. Ogólniej, multiplikatorami splotowymi są dystrybucje o tzw. ograniczonym nośniku [11], [33].

Jeśli więc f jest dowolną dystrybucją, to na podstawie (49) dostajemy np.

$$(f \star \delta, \varphi) = (f, \varphi),$$
$$(f \star \delta', \varphi) = (f, -\varphi') = (f', \varphi) = (f' \star \delta, \varphi),$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$(f \star \delta^{(n)}, \varphi) = \big(f, (-1)^n \varphi^{(n)}\big) = (f^{(n)}, \varphi) = (f^{(n)} \star \delta, \varphi)$$

lub w zapisie tradycyjnym:

$$f \star \delta = f; \quad f \star \delta' = f' \star \delta = f'; \quad f \star \delta^{(n)} = f^{(n)} \star \delta = f^{(n)}. \tag{50}$$

Jeśli $P(x)$ jest dowolnym wielomianem, to mamy ogólnie:

$$f \star P\left(\frac{d}{dt}\right)\delta = P\left(\frac{d}{dt}\right) f \star \delta = P\left(\frac{d}{dt}\right) f. \tag{51}$$

Analogiczne zależności można otrzymać dla splotów z dystrybucją $\delta_a = \delta(t - a)$.

Można wykazać, że przy pewnych założeniach (które są spełnione na przykład, gdy jeden z czynników jest δ-dystrybucją lub jej pochodną) określony przez (49) splot zachowuje podstawowe własności, występujące w dziedzinie funkcyjnej: przemienność[1]), łączność, rozdzielność względem dodawania itp. Słuszna jest wówczas także zależność:

$$D(f \overset{\star}{,} g) = Df \star g = f \star Dg, \tag{52}$$

gdzie D — dowolny jednorodny operator różniczkowy. Jeśli: $D \equiv \frac{d^n}{dt^n}$, dostajemy uogólnioną „formułę Duhamela" dla dystrybucji:

$$(f \star g)^{(n)} = f^{(n)} \star g = f \star g^{(n)}, \tag{53}$$

której przypadkami szczególnymi były dwie spośród zależności (50).

1.9. Przekształcenie Fouriera

Wprowadzenie przekształcenia Fouriera dla dystrybucji wymaga pewnego zmodyfikowania przestrzeni funkcji podstawowych. Obecnie pod pojęciem funkcji podstawowej będziemy rozumieć funkcje zespolone $\varphi(t)$ zmiennej rzeczywistej t spełniające bez zmian warunki 1) i 2) podane w rozdz. 1.2. Przestrzeń tych funkcji będziemy nadal oznaczać przez K.

Wprowadzenie funkcji podstawowych zespolonych nie wnosi żadnych istotnych zmian ani pojęciowych ani rachunkowych. Jedynie, jeśli $f(t)$ jest funkcją zespoloną lokalnie całkowalną, to dystrybucję regularną (f, φ) rozumiemy teraz jako funkcjonał:

$$(f, \varphi) = \int_{-\infty}^{+\infty} \overline{f(t)}\, \varphi(t)\, dt, \tag{54}$$

gdzie $\overline{f(t)}$ jest funkcją zespoloną sprzężoną z $f(t)$. Podobnie jeśli a jest liczbą zespoloną lub funkcją zespoloną, klasy C^∞ $(-\infty, +\infty)$, to:

$$(af, \varphi) = \overline{a}\,(f, \varphi) = (f, \overline{a}\,\varphi). \tag{55}$$

---

[1]) W definicji splotu dystrybucji $f*g$ zakładaliśmy, że druga z nich, tzn. g, jest multiplikatorem splotowym w przestrzeni K. Przy zmianie kolejności na $g*f$ może się więc zdarzyć, iż założenie to nie będzie spełnione. Mimo to, jak można wykazać [11], przy podanych założeniach zmieniając we wzorze (49) miejscami f i g i uwzględniając tę zmianę w (47) otrzymuje się również dystrybucję, i to równą splotowi $f*g$, a więc $g*f = f*g$.

Każda funkcja podstawowa $\varphi(t)$ ma $\mathcal{F}$-transformatę równą:

$$\tilde{\varphi} = \mathcal{F}\varphi = \int_{-\infty}^{+\infty} e^{-i\omega t}\varphi(t)\,dt = (e^{i\omega t}, \varphi). \tag{56}$$

Zbiór $\mathcal{F}$-transformat $\tilde{\varphi}(\omega)$ funkcji podstawowych $\varphi(t)$ będziemy oznaczać przez Z (por. [10], [11], [12]).

Jeśli $f(t)$ jest dowolną funkcją bezwzględnie całkowalną oraz $\tilde{f} = \mathcal{F}f$, to (równość Parsevala):

$$(f, \varphi) = \int_{-\infty}^{+\infty} \overline{f(t)}\,\varphi(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \overline{f(t)}\left[\int_{-\infty}^{+\infty} \tilde{\varphi}(\omega)\,e^{i\omega t}\,d\omega\right] dt =$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tilde{\varphi}(\omega)\left[\int_{-\infty}^{+\infty} \overline{f(t)}\,e^{i\omega t}\,dt\right] d\omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \overline{\tilde{f}(\omega)}\,\tilde{\varphi}(\omega)\,d\omega = \frac{1}{2\pi}(\tilde{f}, \tilde{\varphi}) \tag{57}$$

Widzimy więc, że funkcjonał (f, φ) może być w tym przypadku wyrażony przez inny funkcjonał $(\tilde{f}, \tilde{\varphi})$ określony jednak już nie w przestrzeni K, lecz w przestrzeni Z (bowiem $\tilde{\varphi} \in Z$).

Związek (57) pozwala na zdefiniowanie przekształcenia Fouriera $\tilde{f} = \mathcal{F}f$ dla dowolnej dystrybucji f jako funkcjonału liniowego $(\tilde{f}, \tilde{\varphi})$ w przestrzeni Z określonego wzorem [11].

$$(\tilde{f}, \tilde{\varphi}) \overset{df}{=} 2\pi(f, \varphi). \tag{58}$$

Przy tej definicji zachowują się podstawowe własności przekształcenia Fouriera. Na przykład jeśli $\tilde{f} = \mathcal{F}f$, to:

$$(\tilde{f}', \tilde{\varphi}) = 2\pi(f', \varphi) = 2\pi(f, -\varphi') = 2\pi\frac{1}{2\pi}(\tilde{f}, -i\omega\tilde{\varphi}) =$$

$$= -i\omega(\tilde{f}, \tilde{\varphi}) = (i\omega\tilde{f}, \tilde{\varphi}),$$

czyli:

$$\mathcal{F}f' = i\omega\mathcal{F}f \tag{59}$$

lub ogólnie:

$$\mathcal{F}\left(P\left(\frac{d}{dt}\right)f\right) = P(i\omega)\,\mathcal{F}f, \tag{60}$$

a także:

$$P\left(\frac{d}{d\omega}\right)\mathcal{F}f = \mathcal{F}\left(P(-it)f\right), \tag{61}$$

gdzie $P(x($ jest dowolnym wielomianem.

Jeśli: $\tilde{f} = \mathcal{F} f$, to wówczas także:

$$f = \mathcal{F}^{-1} \tilde{f},$$

tzn. dystrybucja f jest wynikiem odwrotnego przekształcenia Fouriera wykonanego nad $\tilde{f}$.

Przykłady:

a) Proste przekształcenia prowadzą do następujących relacji:

$$(\tilde{\delta}_a, \tilde{\varphi}) = 2\pi (\delta_a, \varphi) = 2\pi \varphi(a) = 2\pi \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\omega a} \tilde{\varphi}(\omega)\, d\omega = (e^{-ia\omega}, \tilde{\varphi}),$$

$$(e^{\tilde{iat}}, \tilde{\varphi}) = 2\pi (e^{iat}, \varphi) = 2\pi \int_{-\infty}^{+\infty} e^{-iat} \varphi(t)\, dt = 2\pi \tilde{\varphi}(a) = 2\pi (\delta_a, \tilde{\varphi})$$

równoważnych związkom:

$$\begin{aligned} &\mathcal{F}\delta(t-a) = e^{-ia\omega}; \quad && \mathcal{F}^{-1} e^{-ia\omega} = \delta(t-a); \\ &\mathcal{F} e^{iat} = 2\pi \delta(\omega - a); \quad && \mathcal{F}^{-1} \delta(\omega - a) = \frac{1}{2\pi} e^{iat}. \end{aligned} \tag{62}$$

Przy $a = 0$ otrzymujemy:

$$\mathcal{F}\delta(t) = 1; \qquad \mathcal{F} 1 = 2\pi \delta(\omega). \tag{63}$$

Ponieważ:

$$\cos at = \frac{e^{iat} + e^{-iat}}{2}, \qquad \sin at = \frac{e^{iat} - e^{-iat}}{2i},$$

więc zgodnie z (62) dostaniemy:

$$\begin{aligned} &\mathcal{F} \cos at = \pi [\delta(\omega + a) + \delta(\omega - a)], \\ &\mathcal{F} \sin at = i\pi [\delta(\omega + a) - \delta(\omega - a)]. \end{aligned} \tag{64}$$

b) Jeśli $P(x)$ jest dowolnym wielomianem, to wobec (61) otrzymamy:

$$\mathcal{F} P(t) = P\left(i \frac{d}{d\omega}\right) \mathcal{F} 1 = 2\pi P\left(i \frac{d}{d\omega}\right) \delta(\omega). \tag{65}$$

W szczególności, gdy $P(t) = t^n$, dostajemy:

$$\mathcal{F} t^n = 2\pi (i)^n \delta^{(n)}(\omega). \tag{66}$$

Analogicznie w oparciu o (60) otrzymujemy:

$$\mathcal{F}\left(P\left(\frac{d}{dt}\right)\delta(t)\right) = P(i\omega), \tag{67}$$

a więc w szczególności, gdy $P(t) = t^n$:

$$\mathcal{F}\delta^{(n)}(t) = (i\omega)^n. \tag{68}$$

Podamy jeszcze twierdzenie będące uogólnieniem twierdzenia Borela o transformacie splotu.

Jeśli f jest dowolną dystrybucją, a g jest dystrybucją spełniającą pewne warunki, to:

$$\mathcal{F}[f * g] = \mathcal{F}f \cdot \mathcal{F}g. \tag{69}$$

Twierdzenie (69) np. bardzo łatwo jest udowodnić w ważnym praktycznie przypadku, gdy $g \equiv P\left(\frac{d}{dt}\right)\delta$. Mamy wówczas bowiem zgodnie z (51), (60) i (67):

$$\mathcal{F}\left[f * P\left(\frac{d}{dt}\right)\delta\right] = \mathcal{F}\left[P\left(\frac{d}{dt}\right)f\right] = \mathcal{F}f \cdot P(i\omega) = \mathcal{F}f \cdot \mathcal{F}g.$$

Ogólnie, twierdzenie (69) jest słuszne, jeśli g jest dystrybucją w tzw. ograniczonym nośniku (por. [11]).

2. DYSTRYBUCJE JAKO KLASY RÓWNOWAŻNE CIĄGÓW

W rozdziale tym omówimy krótko koncepcję, podstawowe pojęcia i własności dystrybucji Mikusińskiego [24] (por. także [4] rozpatrywanych w przedziale $(-\infty, +\infty)$ oraz — z pewnymi modyfikacjami Korevaara [16] — w przedziale $[0, +\infty)$.

2.1. Ciągi podstawowe i ich równoważność

Niech $f_n(t)$ będą funkcjami ciągłymi w dowolnym przedziale $T_1 < t < T_2$ $(-\infty \leqslant T_1 < T_2 \leqslant +\infty)$. Ciąg $\{f_n(t)\}$ będziemy nazywać ciągiem podstawowym, jeśli istnieje taki ciąg funkcyjny $\{F_n(t)\}$ oraz liczba całkowita $k \geqslant 0$, że

1) zachodzi związek

$$F_n^{(k)}(t) \equiv \frac{d^k F_n(t)}{dt^k} = f_n(t), \tag{70}$$

2) ciąg $\{F_n(t)\}$ jest w (T_1, T_2) niemal jednostajnie zbieżny[1]). Notu-

[1]) To znaczy jest zbieżny jednostajnie w każdym przedziale domkniętym leżącym w przedziale (T_1, T_2).

jemy to symbolicznie w postaci: $F_n(t) \rightrightarrows$, lub $F_n(t) \rightrightarrows F(t)$, gdzie $F(t)$ jest funkcją graniczną ciągu $\{F_n(t)\}$.

Rozpatrzmy proste przykłady.

a) Każdy ciąg $\{f_n(t)\}$ funkcji ciągłych niemal jednostajnie zbieżny[1]) jest ciągiem podstawowym [wówczas w definicji mamy $k = 0$ i $F_n(t) \equiv f_n(t)$]. W szczególności ciągiem podstawowym jest ciąg stały $\{f(t)\}$, tj. ciąg, którego każdy wyraz jest równy tej samej funkcji ciągłej $f(t)$. W ten sposób każdej funkcji ciągłej w przedziale odpowiada ciąg podstawowy $\{f(t)\}$.

b) Ciąg — por. (38) —

$$f_n(t) = \frac{1}{\pi} \cdot \frac{n}{n^2 t^2 + 1}, \tag{71}$$

którego wyrazy mają przebieg podany na rys. 3 jest ciągiem podstawowym, bowiem można wykazać (rachunki pomijamy), że ciąg:

$$F_n(t) = \frac{t}{\pi} \operatorname{arctg} nt + \frac{t}{2} - \frac{1}{2\pi n} \ln(1 + n^2 t^2)$$

jest w każdym przedziale ograniczonym jednostajnie zbieżny do funkcji $t \cdot \mathbf{1}(t)$, a przy tym:

$$F_n''(t) = f_n(t).$$

Sam ciąg (71) jest w każdym przedziale (T_1, T_2) przy $T_1 \cdot T_2 \geqslant 0$ niemal jednostajnie zbieżny do zera, natomiast w żadnym przedziale (T_1, T_2), zawierającym zero (tj. przy $T_1 \cdot T_2 < 0$) nie jest w ogóle zbieżny (bowiem jest rozbieżny przy $t = 0$).

Analogicznie ciąg $g_n(t)$ o wyrazach:

$$g_n(t) = \begin{cases} 0 & ; \ -\infty < t \leqslant 0 \\ n^2 t & ; \ 0 \leqslant t \leqslant \dfrac{1}{n} \\ -n^2 t + 2n & ; \ \dfrac{1}{n} \leqslant t \leqslant \dfrac{2}{n} \\ 0 & ; \ \dfrac{2}{n} \leqslant t < \infty \end{cases} \tag{72}$$

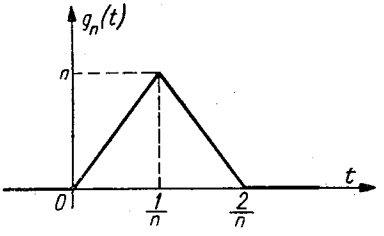
(por. rys. 5) jest ciągiem podstawowym, łatwo jest bowiem sprawdzić, że

[1]) Jeśli nie będziemy wyraźnie wymieniać przedziału, to umawiamy się, iż chodzić będzie zawsze o przedział $(-\infty, +\infty)$.

ciąg:

$$G_n(t) = \begin{cases} 0 & ; \quad -\infty < t \leqslant 0 \\ \frac{1}{6} n^2 t^3 & ; \quad 0 \leqslant t \leqslant \frac{1}{n} \\ -\frac{1}{6} n^2 t^3 + n t^2 - t + \frac{1}{3n} & ; \quad \frac{1}{n} \leqslant t \leqslant \frac{2}{n} \\ t & ; \quad \frac{2}{n} \leqslant t < +\infty \end{cases} \tag{73}$$

jest w każdym przedziale ograniczonym jednostajnie zbieżny do funkcji



Rys. 5. Wykres funkcji $g_n(t)$ określonej wzorem (72)

$t \cdot \mathbf{1}(t)$, a ponadto:

$$G_n''(t) = g_n(t),$$

c) Ciąg funkcji:

$$f_n(t) = n \cos(n^2 t)$$

jest w każdym przedziale (T_1, T_2) ciągiem podstawowym, bowiem ciąg:

$$F_n(t) = \frac{1}{n} \sin(n^2 t)$$

jest w każdym przedziale ograniczonym zbieżny jednostajnie do zera por. [rozdz. 1.6a)], a przy tym:

$$F_n'(t) = f_n(t).$$

Sam ciąg $\{f_n(t)\}$ przy tym nie jest zbieżny w żadnym przedziale (T_1, T_2).

Zwracamy uwagę, że jeśli $\{f_n(t)\}$ jest ciągiem podstawowym, to ciąg $\{F_n(t)\}$ i liczba k spełniające warunki definicyjne 1) i 2) nie są bynajmniej jedyne. Jeśli bowiem warunki te są spełnione przez dany ciąg $\{F_n(t)\}$ i daną liczbę k, to — jak łatwo sprawdzić — są one spełnione także przez każdy ciąg funkcji

$$F_n^{\star}(t) = \int_{t_0}^{t} \int_{t_0}^{\tau_m} \dots \int_{t_0}^{\tau_2} F_n(\tau_1)\, d\tau_1\, d\tau_2 \dots d\tau_m$$

i liczbę $k + m$.

2.2. Równoważność ciągów podstawowych, określenie dystrybucji

Dwa ciągi podstawowe $\{f_n(t)\}$ i $\{g_n(t)\}$ nazywa się równoważnymi [24], jeżeli istnieje taka liczba całkowita $k \geqslant 0$ i takie dwa ciągi $\{F_n(t)\}$ i $\{G_n(t)\}$, że:

$$F_n^{(k)}(t) = f_n(t); \quad G_n^{(k)}(t) = g_n(t) \tag{74}$$

oraz

$$F_n(t) \rightrightarrows \Phi(t); \quad G_n(t) \rightrightarrows \Phi(t). \tag{75}$$

Na przykład ciągi $\{f_n(t)\}$ i $\{g_n(t)\}$ rozpatrywane w przykładzie 2.1b) są równoważne, ponieważ ciągi $\{F_n(t)\}$ i $\{G_n(t)\}$ są zbieżne jednostajnie do tej samej funkcji $t \cdot \mathbf{1}(t)$, a przy tym: $f_n(t) = F_n''(t)$ oraz $g_n(t) = G_n''(t)$.

Fakt równoważności ciągów $\{f_n(t)\}$ i $\{g_n(t)\}$ zapisujemy symbolicznie w postaci:

$$\{f_n(t)\} \sim \{g_n(t)\}.$$

Wprowadzone pojęcie równoważności pozwala podzielić zbiór wszystkich ciągów podstawowych w danym przedziale na rozłączne klasy ciągów równoważnych, tzn. każdy ciąg podstawowy należy do dokładnie jednej takiej klasy.

Każdą klasę ciągów równoważnych w danym przedziale (T_1, T_2) nazywa się dystrybucją (w sensie Mikusińskiego) w tym przedziale.

Wszystkie ciągi równoważne danemu ciągowi podstawowemu $\{f_n(t)\}$ określają więc tę samą dystrybucję, można zatem na gruncie tej definicji mówić o utożsamianiu wszystkich ciągów równoważnych danemu ciągowi podstawowemu (tzn. w ramach tej samej klasy).

Dystrybucję wyznaczoną przez klasę ciągów równoważnych ciągowi $\{f_n(t)\}$ będziemy oznaczać przez $[f_n(t)]$.

Zauważamy, że każda funkcja ciągła $f(t)$ w danym przedziale wyznacza dystrybucję $[f(t)]$, tj. klasę ciągów podstawowych równoważnych ciągowi stałemu $\{f(t)\}$.

Jest oczywiście i odwrotnie — jeśli dystrybucja ma postać $[f(t)]$, tzn. jest klasą ciągów równoważnych ciągowi stałemu $\{f(t)\}$, to tym samym jest określona funkcja ciągła $f(t)$. W tym sensie więc można utożsa-

mieć dystrybucje typu $[f(t)]$ z funkcjami ciągłymi $f(t)$ [por. 1.3a)] i pisać po prostu:

$$f(t) = [f(t)]. \tag{76}$$

Ogólnie, używa się często zapisu

$$f(t) = [f_n(t)] \tag{77}$$

niezależnie od tego, czy dystrybucję $[f_n(t)]$ da się utożsamić z funkcją ciągłą, czy nie. Przykładem dystrybucji, której nie można utożsamić z funkcją, jest np. dystrybucja [por. 2.1b)] $\left[\frac{1}{\pi}\frac{n}{n^2t^2+1}\right]$, którą nazywa się δ-dystrybucją [por. 1.6c)]:

$$\delta(t) = \left[\frac{1}{\pi}\frac{n}{n^2t^2+1}\right]. \tag{78}$$

Omówimy pokrótce, jak się definiuje najważniejsze działania na dystrybucjach w ujęciu ciągowym.

Będziemy zakładać, że $f(t) = [f_n(t)]$ i $g(t) = [g_n(t)]$ są dowolnymi dystrybucjami określonymi w tym samym przedziale (T_1, T_2).

a. Równość dystrybucji. Mamy:

$$[f_n(t)] = [g_n(t)], \tag{79}$$

wtedy i tylko wtedy, gdy: $\{f_n(t)\} \sim \{g_n(t)\}$.

b. Suma dystrybucji. Mamy:

$$[f_n(t)] + [g_n(t)] = [f_n(t) + g_n(t)]. \tag{80}$$

c. Iloczyn dystrybucji przez liczbę λ. Mamy:

$$\lambda[f_n(t)] = [\lambda f_n(t)]. \tag{81}$$

Można sprawdzić (szczegóły pomijamy), że związki (80) i (81) istotnie określają dystrybucje, tzn. ciągi $\{f_n(t) + g_n(t)\}$ oraz $\{\lambda f_n(t)\}$ są ciągami podstawowymi, oraz jeśli $\{f_n^*(t)\} \sim \{f_n(t)\}$ oraz $\{g^*_n(t)\} \sim \{g_n(t)\}$, to także $\{f_n(t) + g_n(t)\} \sim \{f^*_n(t) + g^*_n(t)\}$ oraz $\{\lambda f_n(t)\} \sim \{\lambda f^*_n(t)\}$.

Analogicznie do (81) definiuje się iloczyn dystrybucji $[f_n(t)]$ przez funkcję $a(t)$ klasy C^∞ [por. 1.4d)].

2.3. Pochodna dystrybucji

Zauważymy przede wszystkim, że jeśli $\{f_n(t)\}$ jest ciągiem podstawowym i $f_n(t) \in C^m$, to ciąg $\{f_n^{(m)}(t)\}$ jest też ciągiem podstawowym. Ponadto, jeśli $\{f_n(t)\} \sim \{g_n(t)\}$ oraz $f_n(t) \in C^m$, $g_n(t) \in C^m$, to także $\{f_n^{(m)}(t)\} \sim \{g_n^{(m)}(t)\}$. Rozważmy teraz dystrybucję $[f_n(t)]$ i załóżmy, że $f_n(t) \in C^1$,

Ciąg $\{f'_n(t)\}$ jest zatem ciągiem podstawowym i pochodną dystrybucji $[f_n(t)]$ definiujemy jako:

$$[f_n(t)]' \overset{df}{=} [f'_n(t)]. \tag{82}$$

tj. jako klasę ciągów równoważnych ciągowi $\{f'_n(t)\}$.

Jeśli $[f_n(t)] = [f(t)]$ i przy tym $f(t) \in C^1$, to zgodnie z (82):

$$[f(t)]' = [f'(t)], \tag{83}$$

a więc pochodna dystrybucyjna funkcji $f(t)$ (tj. pochodna dystrybucji $[f(t)]$) jest identyczna ze zwykłą pochodną $f'(t)$.

Powstaje jednak pytanie, czy każdą dystrybucję można przedstawić jako $[f_n(t)]$, gdzie $f_n(t) \in C^1$ [1]), a co za tym idzie, czy każda dystrybucja ma pochodną określoną przez (82).

Odpowiedź na to pytanie jest pozytywna ([24], [16]). Na podstawie twierdzenia aproksymacyjnego Weierstrassa łatwo jest np. wykazać [24], że każdą dystrybucję można przedstawić w postaci $[p_n(t)]$, gdzie $p_n(t)$ — wielomiany, a więc funkcje klasy C^1.

Wynika stąd (por. 1.5), że każda dystrybucja ma pochodną dowolnie wysokiego rzędu, określoną związkiem:

$$[f_n(t)]^{(m)} = [f_n^{(m)}(t)], \tag{84}$$

przy czym $f_n(t) \in C^m$.

Weźmy teraz dowolną dystrybucję $f(t) = [f_n(t)]$. Z określenia wynika, że $f_n(t) = F_n^{(k)}(t)$ i przy tym $F_n(t) \rightrightarrows F(t)$.

Ponieważ $\{F_n(t)\} \sim \{F(t)\}$, więc: $[F_n(t)] = [F(t)] = F(t)$, a zatem zgodnie z (84):

$$f(t) = [f_n(t)] = [F_n^{(k)}(t)] = [F_n(t)]^{(k)} = F^{(k)}(t), \tag{85}$$

tzn. każda dystrybucja (w sensie Mikusińskiego) jest pochodną dystrybucyjną pewnego rzędu funkcji ciągłej.

Własność ta pozwala na zorientowanie się we wzajemnym stosunku między dystrybucjami w ujęciu funkcjonałowym (Schwartza) i w ujęciu ciągowym (Mikusińskiego).

Przy definicji funkcjonałowej powyższą własność posiadały jedynie dystrybucje rzędu skończonego — por. rozdz. 1.7. Nie jest to przypadkowe, można bowiem wykazać [24], iż zachodzi równoważność pojęć dystrybucji w sensie Mikusińskiego i dystrybucji rzędu skończonego w sensie Schwartza przy podanych poprzednio definicjach. Warto jednak pod-

[1]) Lub — co jest równoważne — czy wśród ciągów równoważnych ciągowi $\{f_n(t)\}$ istnieje zawsze taki ciąg $\{\varphi_n(t)\}$, że $\varphi_n(t) \in C^1$.

kreślić, że można tak zmodyfikować definicję dystrybucji w sensie Mikusińskiego (por. [24]), aby obejmowała ona również dystrybucje rzędu nieskończonego — por. (43).

Widzieliśmy, że każdą funkcję ciągłą $f(t)$ można utożsamiać z dystrybucją $[f(t)]$ i odwrotnie. Ważny ten fakt można uogólnić także na funkcje lokalnie całkowalne. Niech $f(t)$ będzie funkcją lokalnie całkowalną; wówczas $\int_{t_0}^{t} f(\tau)d\tau$ jest funkcją ciągłą i prawie wszędzie

$$\left(\int_{t_0}^{t} f(\tau)\,d\tau\right)' = f(t). \tag{86}$$

Funkcja $\int_{t_0}^{t} f(\tau)d\tau$, jako ciągła, może być traktowana jako dystrybucja $[\int_{t_0}^{t} f(\tau)d\tau]$, a wobec (85) jej pochodna dystrybucyjna $[\int_{t_0}^{t} f(\tau)d\tau]'$ może być utożsamiona z funkcją $f(t)$. Wymaga to oczywiście utożsamiania takich funkcji $f(t)$ i $g(t)$ lokalnie całkowalnych, dla których:

$$\int_{t_0}^{t} f(\tau)\,d\tau \equiv \int_{t_0}^{t} g(\tau)\,d\tau,$$

tzn. równych prawie wszędzie [por. 1.3a) i 1.4a)].

W ten sposób przypadkami szczególnymi dystrybucji są zarówno funkcje ciągłe, jak i, ogólnie, funkcje lokalnie całkowalne.

P r z y k ł a d. Wróćmy do przykładu 2.1b). Dystrybucja:

$$t \cdot \mathbf{1}(t) = \left[\frac{t}{\pi}\operatorname{arctg} n t + \frac{t}{2} - \frac{1}{2\pi n}\ln(1+n^2t^2)\right] \tag{87}$$

jest funkcją ciągłą, a wyrazy ciągu $\left\{\frac{t}{\pi}\operatorname{arctg} nt + \frac{t}{2} - \frac{1}{2\pi n}\ln(1+n^2t^2)\right\}$ są różniczkowalne. Funkcję tę można przedstawić w postaci:

$$t \cdot \mathbf{1}(t) = \int_{-\infty}^{t} \mathbf{1}(\tau)\,d\tau,$$

a więc prawie wszędzie (z wyjątkiem $t = 0$):

$$\left(t \cdot \mathbf{1}(t)\right)' = \mathbf{1}(t)$$

i wobec tego zgodnie z poprzednią uwagą o funkcjach lokalnie całkowalnych pochodna dystrybucyjna (87) jest równa:

$$\mathbf{1}(t) = \left[\frac{1}{2} + \frac{1}{\pi}\operatorname{arctg} n t\right]. \tag{88}$$

Ponieważ wyrazy ciągu $\left\{\frac{1}{2}+\frac{1}{\pi}\operatorname{arctg} n t\right\}$ są różniczkowalne, wobec tego pochodna (88), która już nie jest funkcją lokalnie całkowalną, jest równa — por. (78) —

$$\delta(t)=\left[\frac{1}{\pi}\,\frac{n}{n^2 t^2+1}\right];$$

oznaczając pochodną δ-dystrybucji przez $\delta'(t)$ otrzymujemy stąd od razu — por. (39) i (40) —

$$\delta'(t)=\left[-\frac{1}{\pi}\,\frac{2n^3 t}{(n^2 t^2+1)^2}\right], \tag{89}$$

przy czym różniczkowanie to można przedłużać dalej.

2.4. Niektóre zagadnienia zbieżności

Zbieżność ciągu dystrybucji w sensie Mikusińskiego określa się następująco [24]. Mówimy, że ciąg dystrybucji $\{f_n(t)\}$ jest zbieżny do dystrybucji $f(t)$, i piszemy:

$$f_n(t)\to f(t) \quad \text{lub} \quad \lim_{n\to\infty} f_n(t)=f(t),$$

jeżeli istnieje taki ciąg $\{F_n(t)\}$ funkcji ciągłych, funkcja ciągła $F(t)$ oraz taka liczba $k\geqslant 0$, że

$$F_n(t)\rightrightarrows F(t)$$

oraz:

$$F_n^{(k)}(t)=f_n(t); \qquad F^{(k)}(t)=f(t),$$

przy czym różniczkowanie w tych związkach jest dystrybucyjne.

Z definicji tej bezpośrednio wynika, że jeśli $f_n(t)\to f(t)$, to także

$$f_n^{(m)}(t)\to f^{(m)}(t), \tag{90}$$

a więc każdy ciąg zbieżny dystrybucji można (dystrybucyjnie!) różniczkować wyraz po wyrazie (por. 1.6).

Ważne jest w szczególności następujące twierdzenie [24]. Ciąg $\{f_n(t)\}$ funkcji ciągłych jest dystrybucyjnie zbieżny do dystrybucji $f(t)$ wtedy i tylko wtedy, gdy:

$$f(t)=[f_n(t)],$$

tzn., gdy jest ciągiem podstawowym dystrybucji $f(t)$.

Wynika stąd np., że ciągi $\left\{\frac{1}{\pi}\,\frac{n}{n^2t^2+1}\right\}$ i $\left\{-\frac{1}{\pi}\,\frac{2n^3t.}{(n^2t^2+1)^2}\right\}$ są — por. (78) i (89) — dystrybucyjnie zbieżne odpowiednio do $\delta(t)$ i $\delta'(t)$ [por. 1.6c)], co można zapisać w postaci:

$$\delta(t) = \lim_{n\to\infty} \frac{1}{\pi}\,\frac{n}{n^2t^2+1}; \qquad \delta'(t) = \lim_{n\to\infty}\left[-\frac{1}{\pi}\,\frac{2n^3t}{(n^2t^2+1)^2}\right]. \tag{91}$$

Zwracamy uwagę, że granice we wzorach (91) są brane w znaczeniu dystrybucyjnym. W „tradycyjnym" ujęciu δ-funkcji granice we wzorach tych były, jak wiadomo, rozumiane w zwykłym sensie, co prowadziło do podstawowych sprzeczności takiego ujęcia [por. 1.6c)][1]).

Zbieżność szeregu dystrybucji oraz twierdzenie o różniczkowaniu szeregu zbieżnego wyraz po wyrazie formułuje się dokładnie tak samo, jak w ujęciu funkcjonałowym (por. 1.6).

2.5. Dystrybucje w przedziale $[0, +\infty)$

We wszystkich zagadnieniach elektrycznych rozpatrywanych w czasie, w których wyróżnia się chwilę początkową, od której interesuje nas zachowanie się badanego układu, wystarcza znajomość rozwiązań w przedziale $[0, +\infty)$. Jak wiadomo, do tego rodzaju zagadnień głównie dostosowany jest rachunek operatorowy. Z tego względu wydaje się, iż szczególnie duże znaczenie w zastosowaniach mogą mieć dystrybucje rozpatrywane w przedziale $[0, +\infty)$.

Przy rozpatrywaniu dystrybucji w przedziale $[0, \infty)$ dogodne jest nieco zmodyfikować podstawowe określenia [16]. Omówimy je pokrótce zwracając głównie uwagę na te momenty, które wykazują pewne różnice w stosunku do omawianych w rozdziałach 2.1÷2.4 dystrybucji na całej osi. Cała symbolika i system oznaczeń pozostają bez zmian.

Będziemy teraz zakładać, że rozpatrywane funkcje $f_n(t)$ są funkcjami ciągłymi w przedziale $(0, +\infty)$.

a) Ciągiem podstawowym będziemy nazywać ciąg $\{f_n(t)\}$,

[1]) Aby nie dopuścić do ew. nieporozumień, jakie mogłyby powstać przy interpretacji wzorów typu (91), podobnie jak i analogicznych wzorów w rozdz. 1.6c) — można by wprowadzić jakiś umowny symbol świadczący o tym, że zbieżność jest rozumiana dystrybucyjnie np. [16]:

$$\delta(t) = {}^{*}\lim_{n\to\infty} \frac{1}{\pi}\,\frac{n}{n^2t^2+1}.$$

$f_n(t) \in C(0, \infty)$[1]), jeżeli istnieje taka liczba całkowita $k \geqslant 0$, że ciąg $\{F_n(t)\}$, gdzie[2]):

$$F_n(t) = \int_0^t \int_0^{\tau_k} \dots \int_0^{\tau_2} f_n(\tau_1)\, d\tau_1\, d\tau_2 \dots d\tau_k = \int_0^t f_n(\tau) \frac{(t-\tau)^{k-1}}{(k-1)!}\, d\tau, \qquad (92)$$

jest niemal jednostajnie zbieżny w przedziale $(0, +\infty)$.

Widzimy więc, że (92) spełnia także podstawowy związek (70), tj.

$$F_n^{(k)}(t) = f_n(t), \qquad (93)$$

jednakże w definicji rozdziału 2.1 $F_n(t)$ były d o w o l n y m i funkcjami spełniającymi (93), podczas gdy funkcje (92) są jednoznacznie określone[3]).

b) R ó w n o w a ż n o ś ć ciągów podstawowych rozumiemy dokładnie tak samo, jak poprzednio. Mówimy, że ciągi podstawowe $\{f_n(t)\}$ i $\{g_n(t)\}$ są równoważne, jeśli istnieje taka liczba $k \geqslant 0$, że ciągi $\{F_n(t)\}$ i $\{G_n(t)\}$, gdzie

$$F_n(t) = \int_0^t f_n(\tau) \frac{(t-\tau)^{k-1}}{(k-1)!}\, d\tau; \qquad G_n(t) = \int_0^t g_n(\tau) \frac{(t-\tau)^{k-1}}{(k-1)!}\, d\tau, \qquad (94)$$

są niemal jednostajnie zbieżne do tej samej funkcji, tj.

$$F_n(t) \rightrightarrows \Phi(t); \qquad G_n(t) \rightrightarrows \Phi(t).$$

c) D e f i n i c j a d y s t r y b u c j i, reguły działań na dystrybucjach, zasada utożsamiania dystrybucji $[f(t)]$ z funkcją ciągłą $f(t)$ oraz cała symbolika — pozostają bez żadnych zmian.

---

[1]) W oryginalnym ujęciu Korevaara [16], w definicji tej zakłada się, że $f_n(t)$ są jedynie funkcjami lokalnie całkowalnymi (wówczas: $k \geqslant 1$). W naszym opracowaniu zmieniono więc nieco tę definicję, starając się o to, aby możliwie mało odbiegała ona od definicji z rozdz. 2.1. Zwracamy ponadto uwagę na to, że zmiana założenia ciągłości funkcji $f_n(t)$ na ich lokalną całkowalność lub odwrotnie nie zmienia zupełnie istoty samych dystrybucji; można by np. definicję ciągu podstawowego — rozdz. 2.1 — oprzeć również tylko na założeniu lokalnej całkowalności, co spowodowałoby jedynie pewne drobne zmiany formalne, związane z wprowadzeniem jako ciągów podstawowych także pewnych ciągów funkcji nieciągłych.

[2]) Przy $k = 0$ mamy: $F_n(t) = f_n(t)$.

[3]) Wiąże się to z tym, że można by także rozszerzyć obszar określoności funkcji $f_n(t)$ i $F_n(t)$ na całą oś przyjmując: $f_n(t) \equiv 0$ i $F_n(t) \equiv 0$ przy $-\infty < t < 0$ oraz $f_n(0) = 0$ (dla zachowania ciągłości) i w tym sensie traktować rozpatrywany przypadek jako przypadek szczególny dystrybucji na całej osi.

d) Pochodną dystrybucji wprowadzamy tak samo, jak w 2.3 z pewnym założeniem dodatkowym związanym z punktem $t = 0$. Rozważmy dystrybucję $[f_n(t)]$ i załóżmy, że $f_n(t) \in C^1\ (0, +\infty)$ oraz $f_n(0) = 0$[1]. Wówczas $\{f'_n(t)\}$ jest też ciągiem podstawowym. Ponadto, jeśli $\{f_n(t)\}$ i $\{g_n(t)\}$ są ciągami różniczkowalnymi oraz $\{f_n(t)\} \sim \{g_n(t)\}$, to także (łatwo sprawdzić: $\{f'_n(t)\} \sim \{g'_n(t)\}$ (por. 2.3) i wobec tego pochodną dystrybucji $[f_n(t)]$ definiujemy jako — por. (82) —

$$[f_n(t)]' = [f'_n(t)], \tag{95}$$

tj. jako klasę ciągów równoważnych ciągowi $\{f'_n(t)\}$.

Ponieważ można wykazać [16] (por. także 2.3), że każdą dystrybucję można przedstawić w postaci $[f_n(t)]$, gdzie $\{f_n(t)\}$ jest ciągiem różniczkowalnym, wobec tego każda dystrybucja posiada pochodną określoną przez (95), a tym samym i wszystkie pochodne.

Identycznie jak poprzednio, każda funkcja $f(t)$ lokalnie całkowalna może być traktowana jako istniejąca prawie wszędzie zwykła pochodna funkcji ciągłej $\int_0^t f(\tau)d\tau$, tzn. — por. (86) —

$$\left(\int_0^t f(\tau)\,d\tau\right)' = f(t) \tag{96}$$

i wobec tego może być utożsamiana z pochodną dystrybucyjną

$$\left[\int_0^t f(\tau)\,d\tau\right]' \text{ dystrybucji } \left[\int_0^t f(\tau)\,d\tau\right].$$

e) Funkcja jednostkowa a funkcja stała. W przedziale $[0, +\infty)$ jest sens rozpatrywać dystrybucje $\mathbf{1}_a(t) = \mathbf{1}(t-a)$ oraz $\delta_a(t) = \delta(t-a)$ i dalsze pochodne przy $a \geqslant 0$, przy czym oczywiście

$$\delta_a(t) = \frac{d}{dt}\mathbf{1}_a(t) \tag{97}$$

(różniczkowanie dystrybucyjne) dla każdego $a \geqslant 0$. Ponieważ jednak w przedziale $[0, \infty)$ funkcja $\mathbf{1}(t)$ jest identyczna z funkcją stałą równą 1 [por. przypisek 3) na str. 398], wobec tego mamy (różniczkowanie dystrybucyjne!):

$$\frac{d}{dt}c\,\mathbf{1}(t) = \frac{dc}{dt} = c\,\delta(t), \tag{98}$$

c — stała.

[1]) Ciąg podstawowy spełniający te dwa warunki będziemy nazywać ciągiem różniczkowalnym.

Zwracamy uwagę, że pochodną (97) można „skonstruować" następująco. Niech $\{f_n(t)\}$ oznacza ciąg funkcji $f_n(t) \in C^1(0, \infty)$ spełniający warunki: $f_n(0) = 0$; $\int_0^t f_n(\tau)\, d\tau \rightrightarrows t\,\mathbf{1}(t-a)$, $a \geqslant 0$. Wówczas:

$$\delta_a(t) = [f_n'(\tau)].$$

Konstrukcję takiego ciągu podano przykładowo w rozdz. 2.1b) ($f_n' = g_n$, $a = 0$).

f) Pochodna dystrybucyjna funkcji w przedziale $[0, +\infty)$. Niech $f(t) \in C^1(0, \infty)$ oraz istnieje $f(0)$. Rozpatrzmy dystrybucję $[f(t) - f(0)\cdot\mathbf{1}(t)]$. Ciąg $\{f(t) - f(0)\cdot\mathbf{1}(t)\}$ jest wobec założeń różniczkowalny i wobec tego:

$$[f(t) - f(0)\cdot\mathbf{1}(t)]' = [f'(t)].$$

Z drugiej strony [różniczkowanie sumy oraz uwaga w punkcie e) i wzór (98)]:

$$[f(t) - f(0)\cdot\mathbf{1}(t)]' = [f(t)]' - f(0)\,\delta(t)$$

i ostatecznie:

$$[f(t)]' = [f'(t)] + f(0)\,\delta(t) \tag{99}$$

lub:

$$f'_{dystr.} = f'(t) + f(0)\,\delta(t). \tag{100}$$

Jest to przypadek szczególny wzoru (26) przy $t_0 = 0$, $h_0 = f(0)$. Wzór (100) można by bez trudu uogólnić na przypadek funkcji przedziałami regularnej. Otrzymalibyśmy (26) przy $t_k \geqslant 0$.

Analogicznie postępuje się przy dalszych pochodnych. Na przykład jeśli $f(t) \in C^n(0, \infty)$ oraz istnieją wartości: $f(0)$, $f'(0)$ ..., $f^{(n-1)}(0)$, to:

$$[f(t)]^{(n)} = [f^{(n)}(t)] + f^{(n-1)}(0)\,\delta(t) + \ldots + f(0)\,\delta^{(n-1)}(t). \tag{101}$$

2.6. Przekształcenie Laplace'a dystrybucji w przedziale $[0, \infty)$

Wprowadzenie dystrybucji w przedziale $[0, \infty)$ pozwala na proste określenie przekształcenia Laplace'a dla dystrybucji, a tym samym rozszerzenie zakresu stosowalności rachunku operatorowego [16][1]).

[1]) Można również wprowadzić przekształcenie Laplace'a dystrybucji w sensie Schwartza — por. np. [30]. Własności i zasady stosowania przekształcenia Laplace'a dla dystrybucji funkcjonałowych zostały ostatnio szerzej omówione w pracy [8].