

$$(1 + \delta) \cos(\varphi_1 + \varepsilon) = \cos \varphi_1 + \sum_{k=2} g_{2k-1} \cos \varphi_{2k-1} \quad (7a)$$

$$(1 + \delta) \sin(\varphi_1 + \varepsilon) = \sin \varphi_1 + \sum_{k=2} -1^{k+1} g_{2k-1} \sin \varphi_{2k-1}. \quad (7b)$$

3. GRANICE BŁĘDU AMPLITUDY

Na podstawie wyżej wyprowadzonych wzorów (7a) i (7b) można napisać następującą zależność:

$$\begin{aligned} (1 + \delta)^2 &= 1 + 2 \sum_{k=2} g_{2k-1} \cos \varphi_{2k-1} \cos \varphi_1 + 2 \sum_{k=2} -1^{k+1} g_{2k-1} \sin \varphi_{2k-1} \sin \varphi_1 + \\ &+ \left[\sum_{k=2} g_{2k-1} \cos \varphi_{2k-1} \right]^2 + \left[\sum_{k=2} -1^{k+1} g_{2k-1} \sin \varphi_{2k-1} \right]^2. \end{aligned} \quad (8)$$

W tym miejscu należy zauważyć, iż

$$\begin{aligned} g_{2k-1} [\cos \varphi_{2k-1} \cos \varphi_1 + (-1)^{k+1} \sin \varphi_{2k-1} \sin \varphi_1] \\ = g_{2k-1} \cos[\varphi_{2k-1} + (-1)^k \varphi_1]. \end{aligned} \quad (9)$$

Dalsze przekształcenia wymagają rozpisania na składniki obu tych wyrażeń sumowych, które podnoszone są do kwadratu. Dostaje się tutaj:

$$\begin{aligned} &\left[\sum_{k=2} g_{2k-1} \cos \varphi_{2k-1} \right]^2 + \left[\sum_{k=2} (-1)^{k+1} g_{2k-1} \sin \varphi_{2k-1} \right]^2 = \\ &= g_3^2 + g_5^2 + g_7^2 + g_9^2 + \dots \\ &+ 2 g_3 g_5 \cos \varphi_5 \cos \varphi_3 + 2 g_3 g_7 \cos \varphi_7 \cos \varphi_3 + + \dots \\ &- 2 g_3 g_5 \sin \varphi_5 \sin \varphi_3 + 2 g_3 g_7 \sin \varphi_7 \sin \varphi_3 - + \dots \\ &+ 2 g_5 g_7 \cos \varphi_7 \cos \varphi_5 + 2 g_5 g_9 \cos \varphi_9 \cos \varphi_5 + + \dots \\ &- 2 g_5 g_7 \sin \varphi_7 \sin \varphi_5 + 2 g_5 g_9 \sin \varphi_9 \sin \varphi_5 - + \dots \\ &+ \dots - \dots + \dots = \\ &= \sum_{k=2} g_{2k-1}^2 + 2 \sum_{k=2} g_{2k-1} g_{2k+1} \cos(\varphi_{2k+1} + \varphi_{2k-1}) + \\ &+ 2 \sum_{k=2} g_{2k-1} g_{2k+3} \cos(\varphi_{2k+3} - \varphi_{2k-1}) + \\ &+ 2 \sum_{k=2} g_{2k-1} g_{2k+5} \cos(\varphi_{2k+5} + \varphi_{2k-1}) + \\ &+ 2 \sum_{k=2} g_{2k-1} g_{2k+7} \cos(\varphi_{2k+7} - \varphi_{2k-1}) + \dots \end{aligned} \quad (10)$$

Wyrażenie (9) przyjmuje wartość maksymalną wówczas, gdy jest

$$g_{2k-1} \cos[\varphi_{2k-1} + (-1)^k \varphi_1] = + |g_{2k-1}| \quad (11a)$$

zaś swoje minimum osiąga dla

$$g_{2k-1} \cos [\varphi_{2k-1} + (-1)^k \varphi_1] = -|g_{2k-1}|. \quad (11b)$$

Jeżeli warunek (11a) jest zachowany dla wszystkich wartości k , którym odpowiada $g \neq 0$, to wówczas wyrażenia (9) i (10) przyjmują swoje wartości największe. Przy spełnionym warunku (11b) występują natomiast wartości najmniejsze. Teza ta wymaga bliższego uzasadnienia.

Początkowo zakłada się, że warunek (11a) jest zachowany dla wszystkich wyższych harmonicznych, a więc i dla $(2n-1)$ -tej. Tym samym zachodzi:

$$\begin{aligned} \text{albo } g_{2n-1} > 0 \text{ i jednocześnie } \varphi_{2n-1} &= (-1)^{n+1} \varphi_1, \\ \text{albo } g_{2n-1} < 0 \text{ i jednocześnie } \varphi_{2n-1} &= (-1)^{n+1} \varphi_1 + 180. \end{aligned}$$

Wynika stąd, że:

$$\begin{aligned} \text{gdy } g_{2k-1} > 0 \text{ oraz } g_{2k+1} > 0 & \quad \text{to} \quad \varphi_{2k+1} + \varphi_{2k-1} = 0 \\ \text{gdy } g_{2k-1} < 0 \text{ oraz } g_{2k+1} < 0 & \quad \text{to} \quad \varphi_{2k+1} + \varphi_{2k-1} = 360 \\ \text{gdy } g_{2k-1} > 0 \text{ oraz } g_{2k+1} < 0 & \quad \text{to} \quad \varphi_{2k+1} + \varphi_{2k-1} = +180 \\ \text{gdy } g_{2k-1} < 0 \text{ oraz } g_{2k+1} > 0 & \quad \text{to} \quad \varphi_{2k+1} + \varphi_{2k-1} = -180 \end{aligned}$$

Podobne rozważania prowadzą do wniosku, iż

$$\begin{aligned} \text{gdy } g_{2k-1} g_{2k+3} > 0 & \quad \text{to} \quad \varphi_{2k+3} - \varphi_{2k-1} = 0, \\ \text{gdy } g_{2k-1} g_{2k+3} < 0 & \quad \text{to} \quad \varphi_{2k+3} - \varphi_{2k-1} = 180, \end{aligned}$$

Analogicznie dostaje się też:

$$\begin{aligned} \text{gdy } g_{2k-1} g_{2k+5} > 0 & \quad \text{to} \quad \varphi_{2k+5} + \varphi_{2k-1} = 0, \\ \text{gdy } g_{2k-1} g_{2k+5} < 0 & \quad \text{to} \quad \varphi_{2k+5} + \varphi_{2k-1} = 180, \end{aligned}$$

jak również

$$\begin{aligned} \text{gdy } g_{2k-1} g_{2k+7} > 0 & \quad \text{to} \quad \varphi_{2k+7} - \varphi_{2k-1} = 0, \\ \text{gdy } g_{2k-1} g_{2k+7} < 0 & \quad \text{to} \quad \varphi_{2k+7} - \varphi_{2k-1} = 180, \text{ itd.} \end{aligned}$$

Z powyższego widać, że tak spełniony warunek (11a) odpowiada pojawieniu się górnej granicy błędu amplitudy. Wzór (8) przybiera wówczas następującą postać:

$$\begin{aligned} [1 + \delta_g]^2 = 1 + \sum_{k=2} |g_{2k-1}|^2 + 2 \sum_{k=2} |g_{2k-1}| + 2 \sum_{k=2} |g_{2k-1}| * |g_{2k+1}| + \\ + 2 \sum_{k=2} |g_{2k-1}| * |g_{2k+3}| + 2 \sum_{k=2} |g_{2k-1}| * |g_{2k+5}| + \dots \end{aligned} \quad (12a)$$

Ten sam sposób postępowania może być zastosowany również w przypadku, gdy dla wszystkich wyższych harmonicznych zachowany jest warunek (11b). Stwierdza się wówczas, że

$$\begin{aligned} \text{gdy } g_{2k-1} g_{2k+1} > 0 & \quad \text{to} \quad \varphi_{2k+1} + \varphi_{2k-1} = 180, \\ \text{gdy } g_{2k-1} g_{2k+1} < 0 & \quad \text{to} \quad \varphi_{2k+1} + \varphi_{2k-1} = 0, \text{ oraz} \end{aligned}$$

gdy $g_{2k-1} g_{2k+3} > 0$ to $\varphi_{2k+3} - \varphi_{2k-1} = 180$,
 gdy $g_{2k-1} g_{2k+3} < 0$ to $\varphi_{2k+3} - \varphi_{2k-1} = 0$, itd.

Powstający błąd amplitudy osiąga teraz swoją dolną granicę, przy czym wyrażenie (8) może być obecnie zapisane jako:

$$\begin{aligned}
 [1 + \delta_d]^2 = & 1 + \sum_{k=2} |g_{2k-1}|^2 - 2 \sum_{k=2} |g_{2k-1}| - 2 \sum_{k=2} |g_{2k-1}| * |g_{2k+1}| - \\
 & - 2 \sum_{k=2} |g_{2k-1}| * |g_{2k+3}| + -2 \sum_{k=2} |g_{2k-1}| * |g_{2k+5}| - \dots
 \end{aligned} \quad (12b)$$

Z powyższych rozważań wynikają zatem związki

$$\left. \begin{aligned}
 [1 + \delta_d]^2 &= \left[1 - \sum_{k=2} |g_{2k-1}| \right]^2 \\
 [1 + \delta_g]^2 &= \left[1 + \sum_{k=2} |g_{2k-1}| \right]^2
 \end{aligned} \right\} \quad (12c)$$

prowadzące ostatecznie do zależności

$$|\delta_G| = \sum_{k=2} |g_{2k-1}| \quad (13)$$

określającej graniczną wartość błędu amplitudy. Granica ta nie zostanie przekroczona dla żadnego momentu początkowego serii próbkowań. Należy tu zauważyć, iż wiele typów przebiegów odkształconych nie pozwala na spełnienie warunku (11a), (11b) dla wszystkich występujących harmonicznich. W takich przypadkach rzeczywiście powstające ekstremalne błędy δ_m , δ_M mogą być znacznie mniejsze od wartości granicznej. Równość $\delta_M = \delta_G$ względnie $\delta_m = -\delta_G$ może bowiem zachodzić jedynie wówczas, jeżeli istnieje taka wartość momentu początkowego x , przy której dany warunek (11) jest spełniony dla wszystkich wyższych harmonicznich.

Z definicji (6) oznaczenia φ wynika, iż pojawienie się takiego stanu jest zależne od kątów fazowych α harmonicznich niewyeliminowanych. Możliwość spełnienia – dla różnych x – obu warunków (11a), (11b) występuje praktycznie wówczas, gdy istnieje tylko jedna harmoniczna niewyeliminowana. Spostrzeżenia te najlepiej objaśnią stosowne przykłady liczbowe. Celem uzyskania możliwie dobrej poglądowości założono, iż zastosowana realizacja metody zapewnia eliminację tylko trzeciej i piątej harmonicznej. Takiej realizacji odpowiadają następujące wartości stałych:

$$C(1) = 6,589113 \quad \text{oraz} \quad \beta = 138,000000 \text{ [deg]}.$$

Założono dalej, że dla badanego przebiegu amplitudy wyższych harmonicznich odpowiadają zależności

$$A_{2k-1} = A_1 / (2k-1)^2. \quad (14)$$

Wynikające stąd wartości wielkości pomocniczej g_{2k-1} zostały zestawione w poniższej tablicy.

T a b l i c a 1.

Przykładowe wartości g_{2k-1}

k	$2k-1$	g_{2k-1}
2	3	0
3	5	0
4	7	-0,012613
5	9	0
6	11	+0,008264
7	13	-0,003657
8	15	0
9	17	-0,002139
10	19	+0,002770

PRZYKŁAD „A”

Przebieg badany jest określony za pomocą równania

$$y = \sum_{k=1}^4 \frac{\sin (2k-1) x}{(2k-1)^2}, \quad (\text{A})$$

gdzie $x = \omega t$. Kształt tej krzywej jest podobny do przedstawionej na rysunku 3. Wartości kątów fazowych wynoszą tu $\alpha_1 = \alpha_3 = \alpha_5 = \alpha_7 = 0$. Błąd amplitudy wynika jedynie z występowania siódmej harmonicznej, której odpowiada ujemna wartość g . Na podstawie warunków (11) i przy uwzględnieniu definicji (6) dochodzi się do poniższych równań, z których wyznacza się momenty początkowe x , odpowiadające granicznemu błędowi amplitudy $\pm \delta_G$. Wartość dodatnia pojawia się, gdy $\cos(\varphi_7 + \varphi_1) = -1$. Wynika stąd, że ma zachodzić $(\alpha_7 + \alpha_1) + 8(x' + \beta) = n' * 360 + 180$. Przy uwzględnieniu przyjętych wartości α_1 , α_7 , β dostaje się zależność

$$x' = n' * 45 - 115,500. \quad (\text{A1})$$

W odniesieniu do minimum otrzymuje się analogicznie, że ma być $(\alpha_7 + \alpha_1) + 8(x'' + \beta) = n'' * 360$. Prowadzi to do związku

$$x'' = n'' * 45 - 138,000. \quad (\text{A2})$$

Stosownie do wzoru (13), graniczna wartość błędu wynosi tu $\delta_G = 1,261\%$. Zgodnie z równaniami (A1, A2) ekstrema winny występować dla niżej zestawionych momentów początkowych:

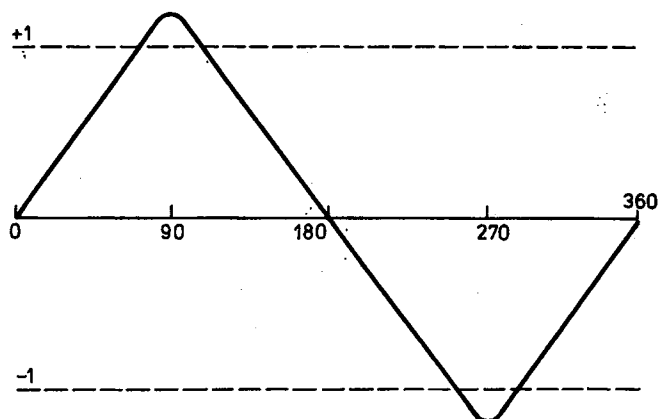
$$\begin{aligned} +\delta_G & \text{ przy } x = 19,500; \quad 64,500; \quad 109,500; \quad 154,500; \quad \dots \\ -\delta_G & \text{ przy } x = 42,000; \quad 87,000; \quad 132,000; \quad 177,000; \quad \dots \end{aligned}$$

Numeryczne badanie tego przykładu, przedstawione w [1], dało dokładnie takie same rezultaty.

PRZYKŁAD „B”

Badana krzywa jest teraz opisana równaniem

$$y = \sum_{k=1}^7 \frac{(-1)^{k+1} \sin (2k-1) x}{(2k-1)^2} \quad (\text{B})$$



Rys. 1 Przebieg odkształcony do przykładu „B”

Kształt tego przebiegu pokazuje rys. 1. Kąty fazowe mają tu następujące wartości: $\alpha_1 = \alpha_5 = \alpha_9 = \alpha_{13} = 0$, $\alpha_3 = \alpha_7 = \alpha_{11} = 180$. Rolę harmonicznym niewyeliminowanych grają 7, 11 oraz 13. Postępując podobnie jak w przykładzie „A”, ustala się równania, służące do wyznaczenia wartości momentu początkowego x . Liczba tych równań jest taka sama jak liczba harmonicznym niewyeliminowanych. W ten sposób otrzymuje się tu poniższe układy trzech równań:

— dla maksimum

$$\left. \begin{array}{ll} \text{przy } k = 4 \text{ ma być} & x = n' * 45 - 138,000 \\ \text{przy } k = 6 \text{ ma być} & x = n'' * 30 - 153,000 \\ \text{przy } k = 7 \text{ ma być} & x = n''' * 30 - 123,000 \end{array} \right\} \quad (\text{B1})$$

— dla minimum

$$\left. \begin{array}{ll} \text{przy } k = 4 \text{ ma być} & x = n' * 45 - 115,500 \\ \text{przy } k = 6 \text{ ma być} & x = n'' * 30 - 138,000 \\ \text{przy } k = 7 \text{ ma być} & x = n''' * 30 - 138,000 \end{array} \right\} \quad (\text{B2})$$

Z układu równań (B1) otrzymuje się, że:

przy $k = 4$ jest $x = 42; 87; 132; 177; 222; 267; 312;$
 przy $k = 6$ jest $x = 27; 57; 87; 117; 147; 177; 207;$
 przy $k = 7$ jest $x = 27; 57; 87; 117; 147; 177; 207;$

Wspólne wartości x , występujące dla wszystkich harmonicznnych niewyeliminowanych, są jak widać następujące:

$$x = 87; \quad 177; \quad 267; \quad 357 \quad (B3)$$

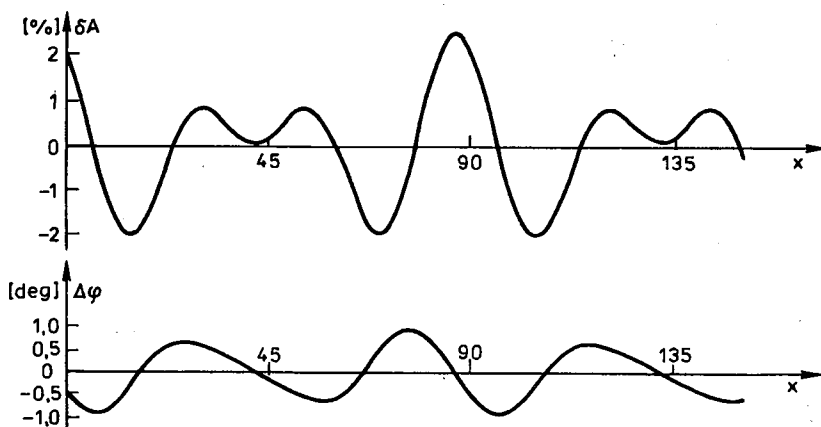
Z układu równań (B2) dostaje się natomiast:

przy $k = 4$ jest $x = 19,5; 64,5; 109,5; 154,5; 199,5;$

przy $k = 6$ jest $x = 12,0; 42,0; 72,0; 102,0; 132,0;$

przy $k = 7$ jest $x = 12,0; 42,0; 72,0; 102,0; 132,0.$

Jak łatwo zauważyć, nie istnieją tutaj wspólne wartości x .



Rys. 2. Błąd amplitudy (u góry) i błąd fazy (u dołu) dla przykładu „B”

Z powyższych rozważań wynika, że dla rozpatrywanego przykładu należy spodziewać się $\delta_M = +\delta_G$ oraz $\delta_m > -\delta_G$. Stosownie do wzoru (13) jest tutaj $\delta_G = 2,452\%$. Przykład ten zbadano numerycznie. Rysunek 2 ilustruje zależność wartości obu błędów od momentu początkowego serii próbkowań x . Przy zastosowaniu mniejszego kroku zmian x wyznaczono ekstrema błędu amplitudy, zestawione w tabelicy 2.

T a b l i c a 2.

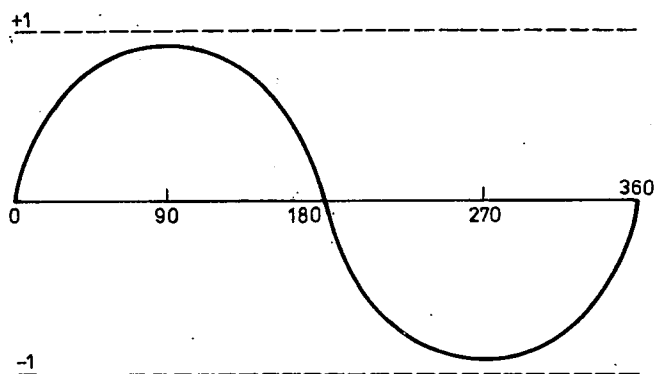
Ekstrema błędu amplitudy dla przykładu „B”

Lp.	x [deg]	δ [deg]	Uwagi:
1.	14,34	-1,998	abs. min.
2.	30,72	+0,847	lok. maks.
3.	42,00	+0,069	lok. min.
4.	53,28	+0,847	lok. maks.
5.	69,66	-1,998	abs. min.
6.	87,00	+2,453	abs. maks.

PRZYKŁAD „C”

Przebieg badany jest tutaj podobny do zastosowanego w przykładzie „A”, lecz zawiera trzy harmoniczne niewyeliminowane 7, 11 oraz 13. Określa go równanie

$$y = \sum_{k=1}^7 \frac{\sin (2k-1)x}{(2k-1)^2}. \quad (C)$$



Rys. 3. Przebieg odkształcony do przykładu „C”

Kształt tej krzywej przedstawiony jest na rys. 3. Przez postępowanie podobne do zastosowanego w przykładzie „B” dostaje się następujące zespoły wartości:

– dla maksimum

przy $k = 4$ jest $x = 19,5; 64,5; 109,5; 154,5; 199,5;$

przy $k = 6$ jest $x = 12,0; 42,0; 72,0; 102,0; 132,0;$

przy $k = 7$ jest $x = 27,0; 57,0; 87,0; 117,0; 147,0;$

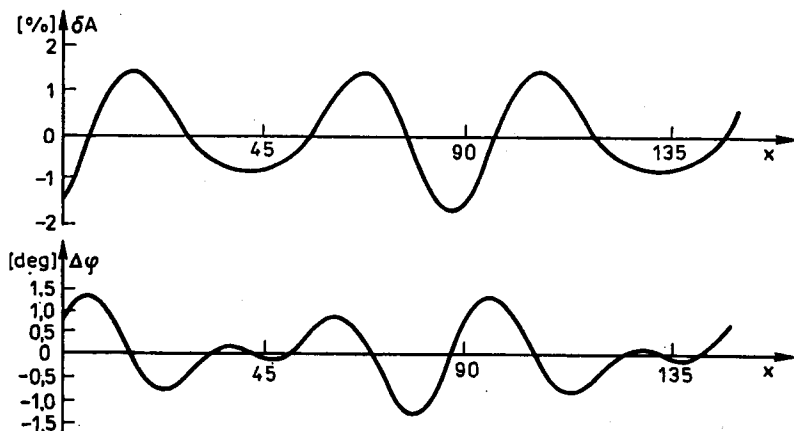
– dla minimum

przy $k = 4$ jest $x = 42,0; 87,0; 132,0; 177,0; 222,0;$

przy $k = 6$ jest $x = 27,0; 57,0; 117,0; 147,0; 177,0;$

przy $k = 7$ jest $x = 12,0; 42,0; 72,0; 102,0; 132,0;$

Zarówno jedna jak i druga seria nie wykazuje istnienia wartości wspólnych. Wynika stąd, że w rozpatrywanym przykładzie należy oczekiwać $\delta_M < +\delta_G$ oraz $\delta_m > -\delta_G$. Błąd graniczny jest taki sam jak w przykładzie „B” i wynosi 2,453%. I ten przykład został zbadany numerycznie. Zależność obu błędów od momentu początkowego x ilustruje rys. 4. Dodatkowo wyznaczone ekstrema błędu amplitudy zostały zestawione w następnej tablicy.



Rys. 4. Błąd amplitudy (u góry) i błąd fazy (u dołu) dla przykładu „C”

T a b l i c a 3.

Ekstrema błędów amplitudy dla przykładu „C”

Lp.	x [deg]	δ [deg]	Uwagi:
1.	16,35	+1,424	abs. maks.
2.	42,00	-0,801	lok. min.
3.	67,65	+1,424	abs. maks.
4.	87,00	-1,722	abs. min.

4. GRANICZNE WARTOŚCI BŁĘDU FAZY

Na podstawie zależności (7) znajduje się łatwo, że

$$\operatorname{tg}(\varphi_1 + \varepsilon) = \frac{\sin \varphi_1 + \sum_{k=2}^{\infty} (-1)^{k+1} g_{2k-1} \sin \varphi_{2k-1}}{\cos \varphi_1 + \sum_{k=2}^{\infty} g_{2k-1} \cos \varphi_{2k-1}}. \quad (15a)$$

Dalsze przekształcenia opierają się na zastosowaniu wzorów:

$$\operatorname{tg} \varphi_1 = \frac{\sin \varphi_1}{\cos \varphi_1} \quad \operatorname{tg}(\kappa - \lambda) = \frac{\operatorname{tg} \kappa - \operatorname{tg} \lambda}{1 + \operatorname{tg} \kappa \operatorname{tg} \lambda}. \quad (15b,c)$$

Celem uproszczenia zapisu wprowadza się przejściowo następujące oznaczenie pomocnicze

$$M = \left[\cos \varphi_1 + \sum_{k=2}^{\infty} g_{2k-1} \cos \varphi_{2k-1} \right] \cos \varphi_1. \quad (15d)$$

W pierwszej kolejności wyznacza się różnicę

$$\begin{aligned} \operatorname{tg}(\varphi_1 + \varepsilon) - \operatorname{tg} \varphi_1 &= \frac{1}{M} \left[\sum_{k=2} (-1)^{k+1} g_{2k-1} \sin \varphi_{2k-1} \cos \varphi_1 - \sum_{k=2} g_{2k-1} \cos \varphi_{2k-1} \sin \varphi_1 \right] = \\ &= -\frac{1}{M} \left[g_3 \sin \varphi_3 \cos \varphi_1 + g_3 \cos \varphi_3 \sin \varphi_1 - g_5 \sin \varphi_5 \cos \varphi_1 + g_5 \cos \varphi_5 \sin \varphi_1 + \right. \\ &\quad \left. + g_7 \sin \varphi_7 \cos \varphi_1 + g_7 \cos \varphi_7 \sin \varphi_1 - g_9 \cos \varphi_9 \sin \varphi_1 + g_9 \sin \varphi_9 \cos \varphi_1 + \dots = \right. \\ &= \frac{1}{M} \left(\sum_{k=2} (-1)^{k+1} g_{2k-1} \sin [\varphi_{2k-1} + (-1)^k \varphi_1] \right). \end{aligned} \quad (16a)$$

Inne przekształcenia zostają wykonane w wyrażeniu, stanowiącym mianownik wzoru (15c):

$$\begin{aligned} 1 + \operatorname{tg} \varphi_1 \operatorname{tg}(\varphi_1 + \varepsilon) &= \frac{1}{M} \left[1 + \sum_{k=2} g_{2k-1} \cos \varphi_{2k-1} \cos \varphi_1 + \sum_{k=2} (-1)^{k+1} g_{2k-1} \sin \varphi_{2k-1} \sin \varphi_1 \right] = \\ &= \frac{1}{M} \left[1 + g_3 \cos \varphi_3 \cos \varphi_1 - g_3 \sin \varphi_3 \sin \varphi_1 + g_5 \cos \varphi_5 \cos \varphi_1 + g_5 \sin \varphi_5 \sin \varphi_1 + \right. \\ &\quad \left. + g_7 \cos \varphi_7 \cos \varphi_1 - g_7 \sin \varphi_7 \sin \varphi_1 + g_9 \cos \varphi_9 \cos \varphi_1 + g_9 \sin \varphi_9 \sin \varphi_1 + \dots = \right. \\ &= \frac{1}{M} \left(1 + \sum_{k=2} g_{2k-1} \cos [\varphi_{2k-1} + (-1)^k \varphi_1] \right). \end{aligned} \quad (16b)$$

W ten sposób otrzymuje się zależność

$$\operatorname{tg} \varepsilon = \frac{\sum_{k=2} (-1)^{k+1} g_{2k-1} \sin [\varphi_{2k-1} + (-1)^k \varphi_1]}{1 + \sum_{k=2} g_{2k-1} \cos [\varphi_{2k-1} + (-1)^k \varphi_1]}. \quad (17)$$

Dość złożona postać powyższego wzoru nastręcza pewne trudności przy poszukiwaniu granicznych wartości błędu fazy ε . Dokładne rozwiązanie teoretyczne tego zagadnienia jest możliwe jedynie w przypadkach najprostszych, gdzie występują najwyżej dwie dowolne harmoniczne niewyeliminowane. Z uwagi na znaczną objętość tych rozważań, podaje się tutaj jedynie otrzymane wnioski.

4.1. BŁĄD FAZOWY DLA PRZYPADKÓW NAJPROSTSZYCH

A) Przy jednej harmonicznej niewyeliminowanej obowiązuje zależność

$$\operatorname{tg} \varepsilon_G = \frac{|g_{2k-1}|}{(1 - g_{2k-1})^2} \quad (18a)$$

a błąd o wartości ε_G pojawia się przy spełnieniu warunku

$$\cos [\varphi_{2k-1} + (-1)^k \varphi_1] = -g_{2k-1}. \quad (18b)$$

Dla przykładu „A”, przy zachowaniu poprzednio przyjętych założeń, otrzymuje się wartość $\varepsilon_G = 0,72269$ [deg], która powinna występować dla następujących momentów początkowych:

- jako maksimum przy $x = 8,160; 53,160; 98,160; \dots$ [deg]
- jako minimum przy $x = 30,840; 75,840; 120,840; \dots$ [deg]

B) Przypadek dwu harmonicznym niewyeliminowanych. Jedna z nich ma liczbę porządkową $(2k-1)$, a druga — $(2n-1)$. Wobec potrzeby nadania zależnościom możliwie prostej postaci, zostają wprowadzone następujące oznaczenia skrótowe:

$$t = 2k-1 \quad u = 2n-1. \quad (19)$$

Mogą tutaj wystąpić różne podprzypadki.

B') liczby k, n są obie parzyste lub obie nieparzyste. Istnieją wówczas dwa rozwiązania, z których podaje się tu jedynie odpowiedź o większej wartości:

$$\operatorname{tg} \varepsilon_G = \frac{|(-1)^k g_u - (-1)^n g_t|}{[1 - (g_u - g_t)^2]^{1/2}}. \quad (19a)$$

Wartość ε_G pojawia się przy jednoczesnym spełnieniu warunków

$$\cos[\varphi_u + (-1)^k \varphi_1] = -(g_u - g_t) \quad (19b)$$

$$\cos[\varphi_t + (-1)^n \varphi_1] = -(g_u - g_t). \quad (19c)$$

Obydwa kąty, określone lewą stroną zależności (19b, c), muszą różnić się znakiem.

Tytułem przykładu zbadano tutaj przebieg o równaniu

$$y = \sum_{k=1}^6 \frac{\sin[(2k-1)x + \alpha_{2k-1}]}{(2k-1)^2}, \quad (D)$$

gdzie przyjęto kąty fazowe o wartościach: $\alpha_1 = \alpha_3 = \alpha_5 = \alpha_9 = 0$, $\alpha_7 = +88,804$, $\alpha_{11} = -91,196$. Wspólne rozwiązania równań (19b, c), a mianowicie $x = 42,000; 132,000; 222,000; 312,000$ odpowiadają maksymalnemu błędowi fazy. Dla minimum wspólne rozwiązania nie istnieją. Graniczny błąd fazy wynosi tu $\varepsilon_G = 1,1963$ [deg]. Numeryczne badanie tego przykładu dało $\delta_M = +1,1963$ [deg] przy $x = 42,00$ [deg].

B'') Jedna z liczb k, n jest parzysta a druga nieparzysta. Tu również występują dwa rozwiązania, z których przytacza się tylko to, które daje większy rezultat:

$$\operatorname{tg} \varepsilon_G = \frac{|(-1)^k g_t(1 - g_t^2 + g_u^2) - (-1)^n g_u(1 + g_t^2 - g_u^2)|}{[1 - 2(g_t^2 + g_u^2) + (g_t^2 - g_u^2)^2]^{1/2}}. \quad (20a)$$

Wartość ε_G pojawia się przy jednoczesnym spełnieniu warunków

$$\cos[\varphi_t + (-1)^k \varphi_1] = \frac{2g_t}{1 + g_t^2 - g_u^2}, \quad (20b)$$

$$\cos[\varphi_u + (-1)^n \varphi_1] = \frac{2g_u}{1 + g_t^2 - g_u^2}. \quad (20c)$$

Znaki odnośnych kątów także tutaj muszą być przeciwne.

Jako przykład został tu zbadany przebieg o równaniu

$$y = \sum_{k=1}^5 \frac{\sin[(2k-1)x + \alpha_{2k-1}]}{(2k-1)^2} + \frac{\sin[13x + \alpha_{13}]}{13^2}, \quad (E)$$

w którym założono: $\alpha_4 = \alpha_3 = \alpha_5 = \alpha_9 = 0$, $\alpha_7 = +88,555$ oraz $\alpha_{13} = -89,581$. Wspólne wartości momentu początkowego, jakie odpowiadają występowaniu maksymalnego błędu fazy, wyznaczone z warunków (20b,c), są następujące: $x = 42,000$; $132,000$; $222,000$; $312,000$. Na podstawie wzoru (20a) dostaje się $\varepsilon_G = 0,9322$ [deg]. Podobnie jak w przykładzie D, tu także nie ma wspólnych x , odpowiadających minimum błędu fazowego. Numeryczne badanie tego przebiegu dało w wyniku $\delta_M = +0,9322$ [deg] przy $x = 42,06$ [deg]. Występująca tu niewielka różnica obu wartości x wynika przypuszczalnie z wpływu zaokrągleń w trakcie obliczeń.

4.2. BŁĄD FAZOWY DLA PRZYPADKU OGÓLNEGO

Rozważania nad przypadkami najprostszymi dają pewną podstawę do opracowania zależności przybliżonych, ważnych dla przypadków ogólnych, charakteryzujących się większą liczbą harmonicznym niewyeliminowanych. Zakłada się tutaj, że wszystkie niezerowe iloczyny

$$g_{2k-1} \cos[\varphi_{2k-1} + (-1)^k \varphi_1]$$

mają wartość bezwzględną pomijalnie małą wobec 1. Założenie to jest uzasadnione faktem, że czynnik g jest małym ułamkiem i oprócz tego argument funkcji trygonometrycznej leży w pobliżu ± 90 [deg]. Z dokładnego wzoru (17) można więc uzyskać zależność przybliżoną

$$\operatorname{tg} \varepsilon_G = \sum_{k=2} |g_{2k-1}|. \quad (21)$$

Graniczny błąd fazy pojawia się wówczas, gdy dla wszystkich niewyeliminowanych harmonicznym spełnione są jednocześnie niżej przytoczone warunki:

– dla maksimum ma być

$$(-1)^{k+1} g_{2k-1} \sin[\varphi_{2k-1} + (-1)^k \varphi_1] = +|g_{2k-1}| \quad (22a)$$

– dla minimum ma zachodzić

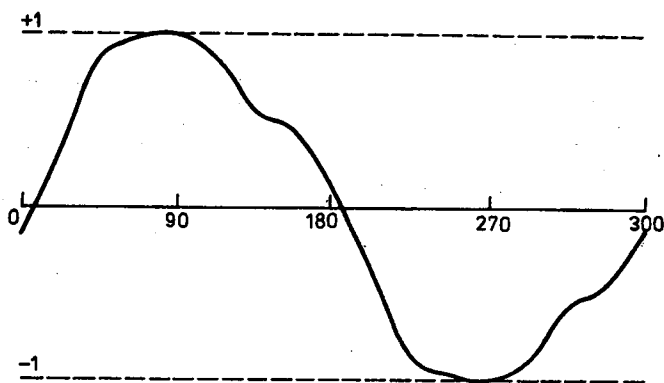
$$(-1)^{k+1} g_{2k-1} \sin[\varphi_{2k-1} + (-1)^k \varphi_1] = -|g_{2k-1}|. \quad (22b)$$

Zestaw rozwiązań układu równań (22a) musi zawierać minimum jedną wspólną odpowiedź x . To samo dotyczy warunków (22b).

Celem sprawdzenia praktycznej użyteczności przybliżonego wzoru (21) przeprowadzono przykładowe badania numeryczne. Przebieg odkształcony został określony równaniem o postaci

$$y = \sum_{k=1}^K \frac{\sin[(2k-1)x + \alpha_{2k-1}]}{(2k-1)^2}. \quad (F)$$

Przy $K=7$ założono: $\alpha=0$, $\alpha_3=\alpha_5=\alpha_{11}=\alpha_{13}=-90$, $\alpha_7=\alpha_9=+90$. Odpowiednia krzywa jest przedstawiona na rys. 5. Realizacja metody pozostaje taka sama, jak w przykładach poprzednich. Spełniony jest tutaj zespół warunków (22a), zaś



Rys. 5. Przebieg odkształcony dla przykładu „F”

oczekiwana wartość graniczna błędu fazy, wyliczona ze wzoru (21), wynosi $\varepsilon_G = 1,4054$ [deg]. Powinna ona wystąpić w pobliżu $x=42$ [deg]. Numeryczne badanie tego przebiegu dało wyniki zestawione w następnej tablicy.

Tablica 4.

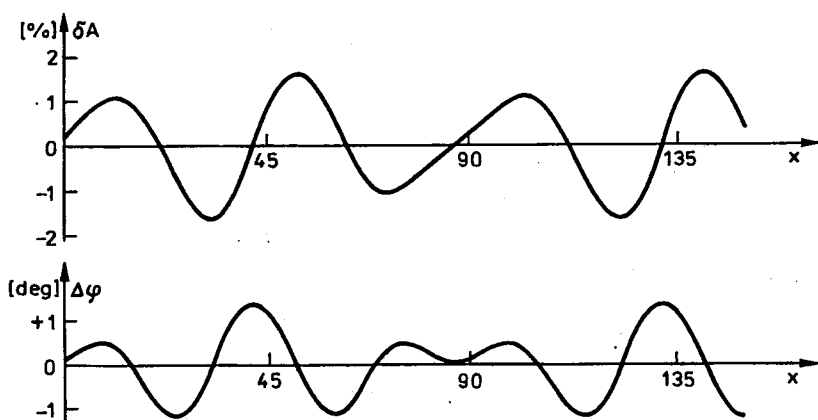
Ekstremalne błędy fazy dla przykładu „F”

Lp.	x [deg]	ε [deg]	Uwagi:
1.	8,25	+0,4785	lok. maks.
2.	24,75	-1,1528	abs. min.
3.	41,91	+1,4056	abs. maks.
4.	59,39	-1,1386	lok. min.
5.	75,72	+0,4875	lok. maks.
6.	87,00	+0,0396	lok. min.

Ten sam typ przebiegu był następnie badany przy założeniu różnych wartości kątów fazowych harmonicznym niewyeliminowanym. Wyznaczano jedynie błąd fazowy, występujący przy $x=42$ [deg], gdyż jest to punkt, w którym należy oczekiwać pojawienia się ekstremum (rys. 6)

Spodziewane wartości $|\varepsilon| = \varepsilon_G$ odpowiadają odmianom 2 i 7, bowiem tylko tu i to przy $x=42$ [deg] warunki (22a,b) są spełnione dla wszystkich niewyeliminowanych harmonicznym.

Dodatkowo badano błąd amplitudy i fazy dla 5. odmiany przebiegu „F”. Odpowiednie krzywe pokazane są na rys. 7, natomiast wyznaczone ekstrema błędu fazy zestawiono w tablicy 6.

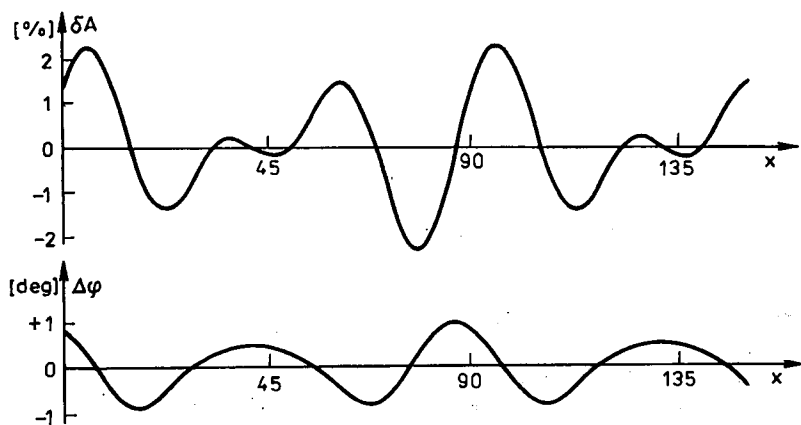


Rys. 6. Błąd amplitudy (u góry) i błąd fazy (u dołu) dla przykładu „F”

T a b l i c a 5.

Błąd fazy przy $x = 42$ [deg] dla różnych odmian przebiegu „F”

Lp.	α [deg]	α_{11} [deg]	α_{13} [deg]	ε [deg]
1.	+90	+90	+90	+0,0396
2.	-90	+90	+90	-1,4054
3.	+90	-90	+90	+0,9866
4.	-90	-90	+90	-0,4587
5.	+90	+90	-90	+0,4587
6.	-90	+90	-90	-0,9866
7.	+90	-90	-90	+1,4054
8.	-90	-90	-90	-0,0396

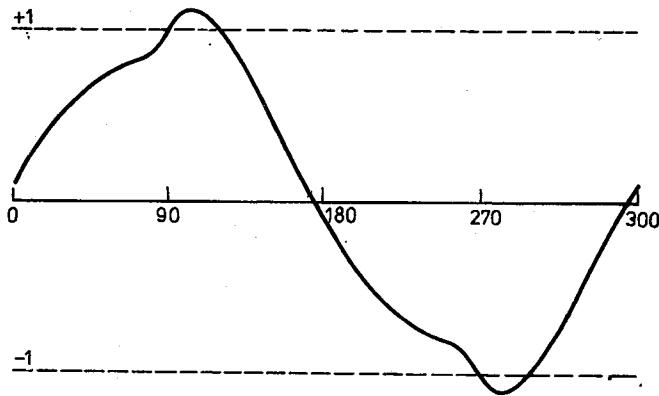


Rys. 7. Błąd amplitudy (u góry) i błąd fazy (u dołu) dla przykładu „F5”

T a b l i c a 6.

Ekstremalne błędy fazy dla przykładu „F5”

Lp.	x [deg]	ε [deg]	Uwagi:
1.	16,44	-0,8190	abs. min.
2.	42,14	+0,4587	lok. maks.
3.	67,82	-0,8126	lok. min.
4.	86,84	+0,9869	abs. maks.



Rys. 8. Przebieg odkształcony typu „F” przy $K=7$ oraz $\alpha=0$, $\alpha_3=\alpha_7=\alpha_{11}=+90$ [deg], $\alpha_5=\alpha_9=\alpha_{13}=-90$ [deg]

Na zakończenie zbadano jeszcze, jak zmienia się graniczny błąd fazy w zależności od widma harmonicznych niewyeliminowanych. Do tego zadania wykorzystano przebieg o równaniu „F” z następującymi wartościami kątów fazowych: $\alpha_1=\alpha_{15}=0$, $\alpha_3=\alpha_5=\alpha_{11}=\alpha_{13}=\alpha_{17}=\alpha_{19}=-90$, $\alpha_7=\alpha_9=+90$. Zakładano tutaj, że najwyższa harmoniczna niewyeliminowana ma liczbę porządkową $(2K-1)$. Dla wszystkich zastosowanych wartości K warunki (22a) są spełnione w pobliżu wartości $x=42$ [deg]. Błędy fazy, wyznaczone przez numeryczne badanie odnośnego przebiegu, są zestawione w tabelicy 9. Dodatkowa kolumna uwidacznia błąd graniczny, obliczony ze wzoru (21). Okazuje się, iż wspomniana zależność daje zupełnie zadowalającą dokładność wyznaczenia ε_G .

T a b l i c a 7.

Maksymalny błąd fazy w zależności od widma przebiegu odkształconego

Lp.	K	x [deg]	ε_M [deg]	ε_G [deg]	Wyst. harm.
1.	6	41,88	1,1963	1,1960	1; 11
2.	7	41,91	1,4056	1,4054	7; 11; 13
3.	8	41,91	1,4056	1,4054	7; 11; 13
4.	9	41,94	1,5280	1,5279	7; 11; 13; 17
5.	10	41,93	1,6867	1,6865	7; 11; 13; 17; 19

5. GRANICZNE WARTOŚCI BŁĘDU DLA RÓŻNYCH REALIZACJI METODY

Wyżej przedstawione rozważania wskazują, że graniczny błąd zarówno amplitudy jak i fazy zostaje określony ze wzoru (13) względnie (21) na podstawie wartości sumy

$$G = \sum_{k=2} |g_{2k-1}| = \sum_{k=2} \frac{A_{2k-1}}{A_1} * \left| \frac{C(2k-1)}{C(1)} \right|. \quad (23)$$

Suma powyższa jest zależna od dwu głównych czynników, bowiem

- wybrana realizacja metody wyznacza wartości stosunków $C(2k-1)/C(1)$ w zależności od k , tj. numeru harmonicznej,
- postać przebiegu badanego określa stosunkowe wartości amplitud poszczególnych harmonicznych A_{2k-1}/A_1 .

Określenie wartości sumy G , z jaką należy liczyć się w praktyce, wymaga więc przyjęcia stosownych założeń. Właściwości przebiegu badanego grają tu bardzo istotną rolę.

Regularne przebiegi okształcone, których opis w postaci szeregu Fouriera łatwo można znaleźć w literaturze [2], odpowiadają najczęściej ogólniejszym związkom typu

$$A_h = A_1 * h^{-r} \quad \text{gdzie } r = 1, 2 \quad \text{lub } 3 \quad (24a)$$

$$\alpha_h = \alpha_1 + n_h * 90 \quad \text{gdzie } n_h = 0, 1, 2, \text{ lub } 3. \quad (24b)$$

Zależności te zostały przyjęte do dalszych rozważań, gdzie wprowadzono także poniższe nazwy, określające typ badanego przebiegu okształconego:

- „wolno ubywający” przy $r = 1$,
- „zwyczajnie ubywający” przy $r = 2$,
- „szybko ubywający” przy $r = 3$.

Innym zagadnieniem jest określenie najwyższej harmonicznej, uwzględnianej jeszcze przy obliczaniu sumy wg wzoru (23). Wyższe harmoniczne są wywoływane lub tłumione przez rozmaite elementy, wchodzące w skład układu sieciowego. Wydaje się, że można tu przyjąć $h_{\max} = 37$ (czyli $k_{\max} = 19$). Liczne publikacje [3–7] wskazują bowiem, że udział 25. harmonicznej bywa zaledwie rzędu 5%.

Przyjmując wyżej przedstawione założenia wyliczono wartości sum G dla realizacji metody, różniących się liczbą eliminowanych harmonicznych. Dodatkowo podano wartość wskaźnika V , która stosownie do rozważań [5] określa potrzebną częstotliwość taktowania członu sterującego procesem próbkowania. Musi być ona V razy większa od częstotliwości przebiegu badanego. Ze względów praktycznych, częstotliwość taktowania nie powinna przekraczać paru MHz. Oznacza to, że przy badaniu przebiegów o częstotliwości technicznej, stosowane mogą być te realizacje metody, dla których V wynosi do 100 000.

Tablica 8.

Suma G dla różnych realizacji metody

Eliminowane harmoniczne	G [%] dla $r =$			V
	1	2	3	
3;	75,939	8,567	1,265	12
3;5;	42,604	3,329	0,324	60
3;5;7	25,277	1,598	0,115	420
3;5;7;11	9,953	0,500	0,027	4620
3;5;7;11;13	6,087	0,270	0,012	60060
3;5;7;11;13;17	3,140	0,116	0,004	1021020
3;5;7;11;13;17;19	1,961	0,068	0,002	19399380
3;5;7;11;13;17;19;23	0,806	0,027	0,001	446185740

ZAKOŃCZENIE

Dobłą dokładność pomiaru rokuje zastosowanie takiej realizacji, która daje wartość sumy G około 0,3%. Widać więc, że przebiegi szybko ubywające można badać według drugiej, zaś zwyczajnie ubywające — według piątej realizacji, ujętych w tablicy 8. Okazuje się także, iż otrzymanie $G \approx 2\%$ dla przebiegów wolno ubywających (co odpowiada siódmej realizacji z tabl. 8) nie jest technicznie możliwe z uwagi na zbyt dużą potrzebną częstotliwość taktowania.

Stwierdzenie to wskazuje na potrzebę opracowania takiej modyfikacji metody, która pozwoli na techniczną realizację wystarczająco dokładnego badania również przebiegów wolno ubywających. Jest to jednak osobne zagadnienie, wymagające odrębnego opracowania.

BIBLIOGRAFIA

1. Bretschneider G., Waldman E.: Zulaessige Oberschwingungen in Stromversorgungsnetzen. ETZ-a, 97 (1976), zesz. 2, str. 90—96
2. Bronštejn I.N., Semiendiajev K.A.: Spravočnik po matematikie. Gos. Izdat. Tekhniko-Teoret. Literatury, Moskwa, 1954
3. Handke A.: Teoretyczna i doświadczalna analiza harmoniczných prądu pobieranych przez wybrane układy tyrystorowe z elektroenergetycznej sieci zasilającej. Rozprawy Elektrot. 20 (1974), zesz. 2, str. 409—422
4. Handke A.: Analiza harmoniczných prądu generowanych w elektroenergetycznej sieci zasilającej przez trójfazowy tyrystorowy prostownik mostkowy. Rozprawy Elektrol. 23 (1977), zesz. 3, str. 710—739
5. Sawicki J.: Theoretische Grundlagen der speziellen Samplings-Methode zur Messung der komplexen Form elner Grundschwingung. Archiv fuer Elektrot. 1988, str. 337—348
6. Sawicki J.: The Finding of th Vector of the Fundamental Harmonic by the Method of a Particular Sampling. 11th IMEKO World Congress, Houston 1988, vol. Applications, str. 581—589
7. Wołoszyk M.: Rationalised Method of Particular Sampling for the Measurement of the Fundamental Harmonic. 12th IMEKO World Congress, Beijing 1991, Digest vol. I, str. 112—113

8. Žeželenko I.W., Ševcov K.K.: *Wysšyje garmoniki, generirujemye ustanovkami dugovoj elektrosvarki*. Električestvo, 94 (1974), zes. 2, str. 84—87
9. Žeželenko I.W., Sorokin W.M.: *Vysšyje garmoniki v električeskich sietjach*. Električestvo, 94 (1974), zes. 11, str. 23—28

J. SAWICKI

BASIC ERROR OF THE METHOD OF THE PARTICULAR SAMPLING

S u m m a r y

The method of the particular sampling is destined for measuring of the vector of the fundamental harmonic component of distorted waves. The theoretical principle of the method has been presented in [5]. The error of the measurement consists of two parts. The fact, that not all higher harmonic components are strictly eliminated is the cause of one of them, and another is the non-ideal features of A/D convertors. The present work gives the expressions for the limits of the first error. The numerical values of the errors for the various technical realisations of the method and for various regular types of the distorted waves are given.

Nowa metoda odbioru koherentnych sygnałów światłowodowych z binarną modulacją amplitudy

JERZY SIUZDAK

Instytut Telekomunikacji, Politechnika Warszawska

Otrzymano 1992.06.19

Autoryzowano 1992.10.20

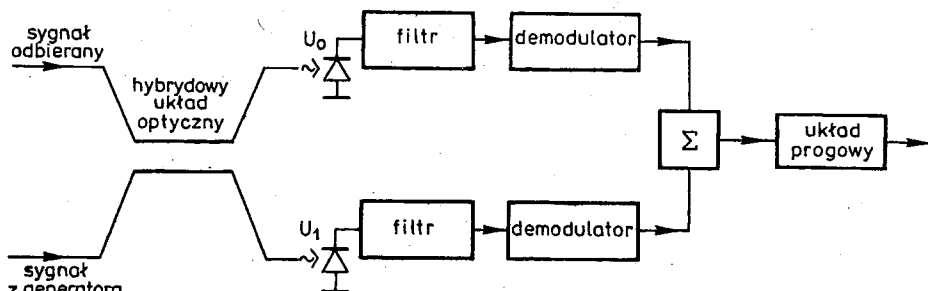
Zaprezentowano nową metodę detekcji sygnałów z binarną modulacją amplitudy. Metoda może znaleźć zastosowanie przy odbiorze koherentnych sygnałów światłowodowych przy szerokościach linii widmowych laserów porównywalnych z szybkością transmisji. Za pomocą modelowania komputerowego stwierdzono, że pogorszenie jakości odbioru spowodowane skończoną szerokością linii widmowych jest nieznaczne przy zastosowaniu opisanej metody.

1. WSTĘP

Zasadniczym problemem przy realizacji homodynowej detekcji koherentnej sygnałów optycznych są trudności z synchronizacją faz fal optycznych pochodzących od dwóch niezależnych laserów półprzewodnikowych. Jedno z rozwiązań tego problemu wykorzystuje specjalne sprzęgacze optyczne o wielu wejściach i wyjściach, co daje możliwość detekcji amplitudy i fazy sygnału optycznego bez uciekania się do synchronizacji fazy [1,2]. Sprzęgacz taki (ang. multiport optical network) daje na wyjściu sygnały optyczne pozostające z sobą w określonych zależnościach fazowych [1], co wykorzystywane jest w dalszej obróbce. Najważniejszym zjawiskiem pogarszającym jakość odbioru z wykorzystaniem tych sprzęgaczy jest niezerowa szerokość linii widmowych laserów półprzewodnikowych używanych jako nadajnik i generator lokalny. Zwiększa to znacznie stopień błędów (BER) jeżeli iloraz szerokości linii widmowych laserów B_L i szybkości transmisji $1/T$ przekracza określoną wartość. W przypadku binarnej modulacji amplitudy (ASK) konieczne jest stosowanie filtracji poddetekcyjnej [3] dla minimalizacji wpływu tego zjawiska. W bieżącej pracy zaprezentowana zostanie inna technika odbioru sygnałów z binarną modulacją amplitudy dla dużych wartości $B_L T$.

2. OPIS METODY

Metoda zostanie przedstawiona na przykładzie odbiornika wielofazowego (ang. phase diversity receiver) z układem specjalnego sprzęgacza optycznego o dwóch wejściach i dwóch wyjściach. Realizacja dla innych rodzajów sprzęgaczy jest analogiczna. Należy tutaj zaznaczyć, że układ tego sprzęgacza jest układem hybrydowym (dającym na wyjściu sygnały ortogonalne) a nie zwykłym sprzęgaczem światłowodowym. Schemat blokowy omawianego odbiornika przedstawiony jest na rys. 1.



Rys. 1. Schemat blokowy proponowanego odbiornika

Składa się on z (2×2) hybrydowego układu optycznego, fotodetektorów, filtrów predetekcyjnych, demodulatorów, sumatora i układu progowego. Sygnały na wyjściu fotodetektorów dane są przez [1,3]

$$u_k = R \sqrt{P_s P_L} b \cos[\Theta(t) + k\pi/2], \quad k = 0, 1. \quad (1)$$

W powyższym wzorze R jest czułością fotodetektora, P_s – mocą sygnału odbieranego, P_L – mocą sygnału z generatora lokalnego, $b=0$ lub 1 w zależności od wartości sygnału binarnego, $\Theta(t)$ jest szumem fazowym wywołanym skończonymi szerokościami linii widmowych laserów nadawczego i generatora lokalnego.

Szumy śrutowe na wyjściach fotodetektorów są niezależne i gaussowskie. Szumy termiczne zostaną pominięte gdyż przyjmujemy, że $P_L \gg P_s$. Gęstość widmowa szumów śrutowych wynosi

$$N = qRP_L/2, \quad -\infty < f < \infty \quad (2)$$

gdzie q jest ładunkiem elektronu. Z wzorów (1), (2) nietrudno wyliczyć, że maksymalna wartość stosunku mocy sygnału do mocy szumu na wyjściu filtru predetekcyjnego wynosi

$$\text{SNR} = \frac{RP_s}{qB}, \quad (3)$$

gdzie B jest pasmem szumowym ($0 < f < \infty$) filtru predetekcyjnego. Ze względu na to, że sygnał użyteczny zawiera nieliniową funkcję szumu fazowego $\Theta(t)$ trudno jest

podać układ optymalnej (tj. zapewniającej minimalną stopę błędów) demodulacji tego sygnału. W przypadku wąskich linii widmowych obydwu laserów spełniających warunek

$$B_L T \ll 1 \quad (4)$$

(gdzie B_L jest sumaryczną szerokością widmową obydwu laserów, zaś $1/T$ szybkością transmisji) w demodulatorze występuje jedynie układ podnoszenia do kwadratu, zaś filtr predetekcyjny jest filtrem dopasowanym do sygnału binarnego. W przypadku wzrostu wartości $B_L T$ zwiększa się szerokość pasma filtru predetekcyjnego zaś po kwadratorze występuje dodatkowo filtr podetekcyjny [4]. Zaprezentowany poniżej układ demodulacji jest uogólnieniem wspomnianych metod i polega na zastosowaniu filtracji nieliniowej.

Niech sygnał na wyjściu filtru predetekcyjnego dany będzie przez $s_{wek}(t)$, $k=0,1$. wówczas sygnał wyjściowy układu filtracji nieliniowej $s_{wyk}(t)$, $k=0,1$, można wyrazić przez

$$s_{wyk}(t) = \iint_{-\infty}^{+\infty} s_{wek}(x) s_{wek}(y) g(t-x, t-y) dx dy, \quad k = 0,1, \quad (5)$$

gdzie funkcja $g(x,y)$ jest odpowiedzią impulsową układu filtracji nieliniowej. Analiza układu odbiornika z filtrem (5) jest trudna ze względu na jego nieliniowość. W związku z tym dla znalezienia optymalnej odpowiedzi impulsowej $g(x,y)$ oraz określenia stosunku mocy sygnału i szumu na wyjściu zastosowano metodę modelowania komputerowego opisaną w następnym rozdziale.

3. SYMULACJA KOMPUTEROWA

Dla umożliwienia symulacji działania układu opisanego równaniem (5) konieczne było zastąpienie go układem dyskretnym. W tym celu przyjęto, że filtr predetekcyjny jest idealnym filtrem prostokątnym o transmitancji danej przez

$$H(f) = \begin{cases} 1 & \text{dla } -2/T < f < 2/T \\ 0 & \text{dla pozostałych } f. \end{cases} \quad (6)$$

Wskutek tego, że filtr o powyższej transmitancji przepuszcza zasadniczą część widma impulsu prostokątnego o czasie trwania T (czyli jednego bitu informacyjnego) można przyjąć, że sygnał na jego wyjściu jest praktycznie nieznieskształcony. Przyjęto następnie, że sygnał wyjściowy z filtru predetekcyjnego podlega idealnemu próbkowaniu, przy czym okres między sąsiednimi próbkami wynosi $T/4$. Otrzymuje się w ten sposób dwa ciągi próbek dla dwóch gałęzi odbiornika z rys. 1

$$A_i = \sin \Theta_i + n_i \quad (7)$$

$$B_i = \cos \Theta_i + m_i,$$

gdzie $\Theta_i = \Theta_{i-1} + \Delta_i$ jest szumem fazowym, przy czym Δ_i jest przypadkowym dryftem fazowym, zaś n_i , m_i są szumami śrutowymi. Wielkości Δ_i , n_i , m_i są niezależnymi zmiennymi gaussowskimi o zerowych wartościach oczekiwanych, których wariancje zależą od parametrów SNR, $B_L T$, przy czym zachodzą następujące relacje

$$E\{n_i n_j\} = 0, \quad E\{m_i m_j\} = 0, \quad E\{\Delta_i \Delta_j\} = 0 \quad \text{dla } i \neq j. \quad (8)$$

Dwie pierwsze są wynikiem przebiegu funkcji korelacji szumu śrutowego na wyjściu filtru predetekcyjnego, która zeruje się dla $t = lT/4$, l – całkowite, trzecia tym, że szum fazowy jest procesem o przyrostach niezależnych.

Ponieważ zgodnie z założeniem pojedynczy bit zajmuje okres czasu o długości T , zatem i funkcja $g(x, y)$ musi być ograniczona do takiego okresu czasowego:

$$g(x, y) = g(x, y) \text{ rect}(x) \text{ rect}(y) \quad (9)$$

gdzie

$$\text{rect}(x) = \begin{cases} 1 & \text{dla } 0 < x < T \\ 0 & \text{dla pozostałych } x \end{cases} \quad (10)$$

Z powyższych wzorów oraz z symetrii zagadnienia wynika, że sygnał wyjściowy filtru nieliniowego powinien być modelowany jako

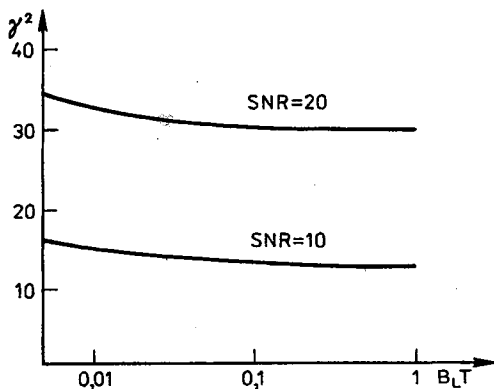
$$s_{wy0}(i) = \sum_{j=0}^3 A_j^2 - j + 2\beta_1 \sum_{j=0}^2 A_{i-j} A_{i-j-1} + 2\beta_2 \sum_{j=0}^1 A_{i-j} A_{i-j-2} + 2\beta_3 A_i A_{i-3} \quad (11)$$

gdzie β_1 , β_2 , β_3 są współczynnikami podlegającymi optymalizacji. Dla dolnej gałęzi w układzie z rys. 1 sygnał wyjściowy otrzymywany jest analogicznie.

W trakcie pracy układu sygnał wyjściowy z demodulatora ma charakter pseudo-przypadkowy i dla określenia jakości pracy urządzenia należy wprowadzić pewną jej miarę. Najbardziej oczywistą byłaby elementowa stopa błędów (BER). Niestety zbyt mała moc obliczeniowa komputera, którym dysponowano wykluczyła tego rodzaju rozwiązanie. W związku z tym jako miarę jakości demodulacji wybrano parametr γ^2 , gdzie

$$\gamma = \frac{E\{s_{wy}|1\} - E\{s_{wy}|0\}}{\sigma\{1\} + \sigma\{0\}} \quad (12)$$

przy czym $E\{s_{wy}|1\}$ jest wartością oczekiwaną sygnału dla momentu podejmowania decyzji przy nadanym bicie „1”, $E\{s_{wy}|0\}$ jest tą wartością dla nadanego sygnału „0”, $\sigma\{1\}$ jest odchyleniem standardowym sygnału wyjściowego w momencie podejmowania decyzji przy nadanej „1”, zaś $\sigma\{0\}$ jest tą samą wielkością dla nadanego „0”. Moment podejmowania decyzji pokrywa się w naszym przypadku z 4 próbką odpowiadającą danemu bitowi. Wielkość γ^2 ma sens zbliżony do stosunku sygnału do szumu na wyjściu i bywa często stosowana do analizy jakości nieliniowych układów detekcyjnych [4, 5].



Rys. 2. Wyniki symulacji

Wyniki symulacji pokazane są na rys. 2. Widać, że wielkość γ^2 zmniejsza się jedynie nieznacznie nawet dla dużych wartości $B_L T$, co świadczy o zalecie proponowanej metody. Współczynniki β_1 , β_2 , β_3 były zmieniane tak, aby otrzymać maksymalną wartość wielkości γ^2 . Warto dodać, że dla $B_L T \approx 0$ optymalne wartości współczynników $\beta_1 \approx \beta_2 \approx \beta_3 \approx 1$, co świadczy o tym, że demodulator można zrealizować jako połączenie filtru dopasowanego i kwadratora, zaś dla dużych $B_L T$ optymalne wartości spełniają zależność $\beta_1 \approx \beta_2 \approx \beta_3 \approx 0$.

PODSUMOWANIE

Zaprezentowano metodę odbioru sygnałów z binarną modulacją amplitudy, stanowiącą uogólnienie znanych metod. Jej zaletą jest to, że daje dobre rezultaty również przy szerokościach linii widmowych laserów półprzewodnikowych porównywalnych z szybkością transmisji. Zasadniczą jej wadą jest większa komplikacja układowa przy realizacji filtrów nieliniowych. W przypadku wąskich linii widmowych laserów półprzewodnikowych realizacja metody pokrywa się ze znanymi metodami [1].

BIBLIOGRAFIA

1. A.W. Davis et al.: *Phase diversity techniques for coherent optical receivers*. Lightwave Techn., 1987, vol. LT-5, str. 561–572
2. J. Siuzdak: *Wybrane problemy koherentnej komunikacji światłowodowej*. KST Bydgoszcz 1989, t. E, str. 112–116
3. J. Siuzdak, W. van Etten: *Heterodyne ASK multiport optical receivers using post detection filtering*. J. Lightwave Techn., 1990 vol. LT-8, str. 71–77
4. L.G. Kazovsky et al.: *ASK multiport optical homodyne receivers*. J. Lightwave Techn., 1987, vol. LT-5, str. 770–790
5. S.D. Personick: *Receiver design for digital fiber-optic communication systems*. Bell Syst. Techn. Journ., 1973, vol. 52, str. 843–874

J. SIUZDAK

A NOVEL METHOD OF RECEPTION OF COHERENT FIBER OPTIC SIGNALS WITH BINARY ASK

S u m m a r y

A novel method of detection of binary ASK modulated signals has been presented. The method may be applied in the detection of coherent fiber optic signals when the laser linewidths are comparable to the bit rate. It has been proved via computer simulation that the deterioration of the reception quality due to non-negligible laser linewidths is small when using the presented method.

On the gaussian approximation of the instantaneous frequency of the sources with non-zero spectral linewidth

JERZY SIUZDAK

Instytut Telekomunikacji, Politechnika Warszawska

Received 1992.06.19

Authorized 1992.10.20

The hypothesis on gaussian probability density function of the instantaneous frequency of the oscillator with a non-zero spectral linewidth is examined by means of numerical simulation. The results show that this is a good approximation only for a small linewidth, whereas it leads to errors when the linewidth increases.

1. INTRODUCTION

The analysis of communication systems using oscillators with substantial spectral linewidths, such as coherent optical communications systems which employ semiconductor lasers is generally difficult. The effect of non-negligible linewidth (phase noise) is often modelled [2,3,5] as producing a random measured intermediate frequency. Then it is possible to treat the receiver as measuring a single realisation of this frequency. This approximation is shown to be correct [4] if the linewidth is small compared to the bandwidth of the filters of the receiver. Nevertheless, it is used when this condition is violated [3,5]. The main objective of this paper is to examine whether it may be applied in the latter case.

2. MODEL DESCRIPTION

A block scheme of the investigated model is depicted in Fig. 1. The incoming signal is mixed with two reference signals being in quadrature. The frequency offset between the signal and the reference is assumed to be zero. The low pass filters reject double frequency terms so the inputs to the integrators are $\sin \Phi$ and $\cos \Phi$, where Φ is the phase noise of both the transmitter and receiver oscillators (lasers). The integrators take averages over time T . Their outputs are squared and added.

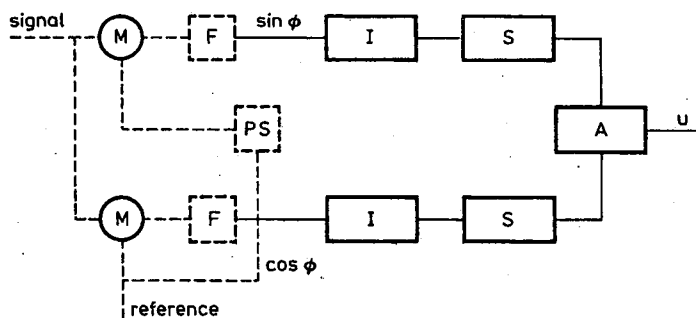


Fig. 1. Block diagram of the modelled scheme: M — multiplier, F — low pass filter, PS — 90° phase shifter, I — integrator, S — squarer, A — adder

This is in fact a well known detection scheme, and it is fully equivalent to the coherent optical ASK phase diversity receiver with a $\{2 \times 2\}$ optical hybrid [1]. Since our primary concern is the phase noise we will neglect any additive interference.

3. GAUSSIAN APPROXIMATION

To find the probability function of the output u , we follow the procedure developed in [3]. According to it, the phase noise is modelled as a random frequency deflection f which is approximately constant over the time T . Given f the value of u is

$$u = \left\{ \frac{1}{T} \int_t^{t+T} \sin(2\pi fx + \beta) dx \right\}^2 + \left\{ \frac{1}{T} \int_t^{t+T} \cos(2\pi fx + \beta) dx \right\}^2$$

$$= 2[1 - \cos(2\pi\alpha)] / (2\pi\alpha)^2. \quad (1)$$

Here $\alpha = fT$ and β is an arbitrary phase angle. The variable α has the following gaussian probability density function (pdf) [3]

$$p(\alpha) = \exp(-\pi\alpha^2/b) / \sqrt{b}. \quad (2)$$

Here b is the normalised to the bit rate width of the (lasers) spectral lines given by

$$b = (B_{TR} + B_{LO})T \quad (3)$$

where B_{TR} and B_{LO} are the spectral widths of the transmitter and the local oscillator, respectively. For the sake of simulation we are concerned with the probability $(P(u < u_0))$ that u is less than a fixed value u_0 . Since the function $u(|\alpha|)$ (1) decreases monotonically when $|\alpha|$ increases (except for very small values of u which are not considered here), this probability may be expressed as

$$P(u < u_0) = P(|\alpha| > \alpha_0) = 1 - \text{erf}(\alpha_0 \sqrt{\pi/b}) \quad (4)$$

where α_0 is the solution of eqn. (1) for $u = u_0$. The function (4) is depicted in Fig. 2 for various b .

4. SIMULATION

The operation of the scheme of Fig. 1 was simulated on a digital computer. The integrator was inevitably replaced with an adder. The number of points N , on which the integration time was divided, was large enough so that this number did not have any influence on the results. The phase noise was modelled by the recurrence relation

$$\Phi_n = \Phi_{n-1} + \sqrt{2\pi b/N} g_n \quad (5)$$

where g_n are independent gaussian variables with zero means and unit variances ($E\{g_i g_j\} = 0$ for $i \neq j$). The results of the simulation are marked as squares and circles in Fig. 2. Satisfactory agreement with the theory occurs only for the smallest value of b .

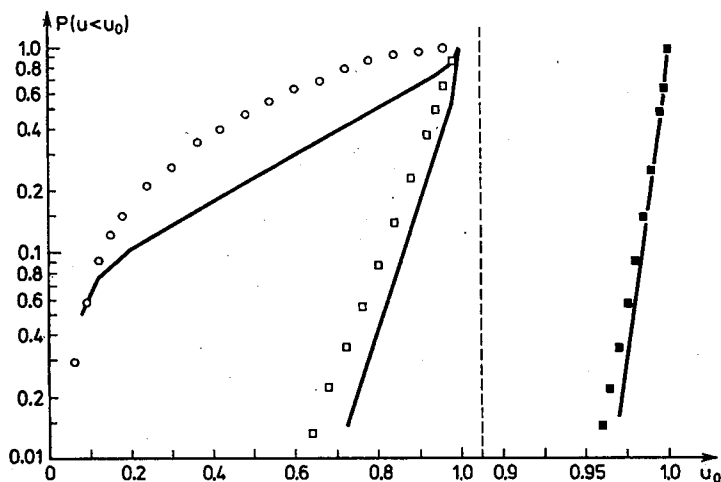


Fig. 2. Comparison between the theoretical curves (solid lines) and the results of simulation marked by circles and squares. Here $P(u < u_0)$ is the probability that the output is less than the fixed value u_0 . Note a change of the horizontal scale at the right part of the figure

CONCLUSIONS

The results show that the method used in [3, 5] to analyse communication receivers with oscillators of substantial linewidths give precise results only for small normalised linewidths b . When the value of b increases the errors of approximation also increase. This is contrary to the results obtained in [6]. One reason of this discrepancy may be the fact that the statistics of the actual DFB lasers used in the experiment [6] does not follow the gaussian model used here.

REFERENCES

1. A.W. Davis et al.: *Phase diversity techniques for coherent optical receivers*, Journ. Lightw. Techn., 1987, LT-5, pp. 561–572
2. R. Gagliardi: *Introduction to communication engineering*, John Willey & Sons, 1978
3. I. Garrett, G. Jacobsen: *Theoretical analysis of heterodyne optical receivers for transmission systems using (semiconductor) lasers with non-negligible linewidth*, Journ. Lightw. Techn., LT-5, 1986, pp. 323–334
4. I. Garrett, G. Jacobsen: *Statistics of laser frequency fluctuations in coherent optical receivers*, Electron. Lett., 1986, 22, pp. 168–170
5. G. Jacobsen, I. Garrett: *Theory for heterodyne optical ASK receivers using square law detection and postdetection filtering*, IEE Proc., Pt. J, 134 (1987), pp. 303–312
6. R.J.S. Pedersen et al.: *Measurement of the statistics of DFB laser frequency fluctuations*, Electr. Lett., 1988, 24, pp. 585–586

J. SIUZDAK

O GAUSSOWSKIEJ APROKSYMACJI CZĘSTOTLIWOŚCI CHWILOWEJ ŹRÓDEŁ
O NIEZEROWEJ SZEROKOŚCI LINII WIDMOWEJ

Streszczenie

Hipoteza o gaussowskiej funkcji gęstości prawdopodobieństwa częstotliwości chwilowej oscylatora z niezerową szerokością linii widmowej została zbadana za pomocą symulacji numerycznej. Rezultaty wykazują, że jest ona dobrym przybliżeniem jedynie dla małych szerokości linii, podczas gdy prowadzi do błędów gdy szerokość linii zwiększa się.

SPIS TREŚCI

H. Kudrewicz: Dyfrakcja na asymetrycznej półpłaszczyźnie impedancyjnej. Struktura rozwiązania w otoczeniu krawędzi	5
A. Zabłudowski, B. Dubalski: Generacja liczbowych ciągów pseudolosowych nad ciałem Galois $GF\ 2^k$ za pomocą rejestru akumulacyjnego	19
P. Dymarski, A. Chmielewski, S. Kula, E. Świercz: Algorytmy predykcji dla koderów ADPCM	35
M. Żółtowski: O szybkich algorytmach metody najmniejszych kwadratów w liniowej filtracji adaptacyjnej	55
T. Adamski: Zastosowanie filtracji kwantylowej do korekcji chwilowej niestałości momentu próbkowania w cyfrowych systemach pomiarowych	91
K. Bieńkowski, M. Ghabally: Wymiany danych i ruch komunikatów bezprzewodowych w mikroprocesorowych sieciach komputerowych klasy DELTA	119
W. Niemiec: O konstytutywnym rozłożonym parametrycznie modelowaniu jednoskładnikowych rzeczywistych procesów. Cz. I. Równania różniczkowe cząstkowe konstytutywne stanu opisujące jednoskładnikowe rzeczywiste procesy	131
W. Niemiec: O konstytutywnym rozłożonym parametrycznie modelowaniu jednoskładnikowych rzeczywistych procesów. Cz. II. Rozwiązywanie równań różniczkowych cząstkowych konstytutywnych stanu dla jednoskładnikowych rzeczywistych procesów	155
E. Danicki, D. Bogucki: Pojemność wzajemna dwóch przetworników międzypalcasowych w obudowie metalowej	187
J. Sawicki: Błąd podstawowy w metodzie szczególnego próbkowania	199
J. Siuzdak: Nowa metoda odbioru koherentnych sygnałów światłowodowych z binarną modulacją amplitudy	219
J. Siuzdak: O gaussowskiej aproksymacji częstotliwości chwilowej źródeł o niezerowej szerokości linii widmowej	225

CONTENTS

H. Kudrewicz: Diffraction by asymmetric half-plane. The structure of the solution near the edge	5
A. Zabłudowski, B. Dubalski: Generation of multilevel pseudorandom sequences over Galois field $GF\ 2^k$ obtained by the use accumulative register	19
P. Dymarski, A. Chmielewski, S. Kula, E. Świercz: Prediction algorithms for ADPCM coders	35
M. Żółtowski: On the last algorithms of linear least square method in adaptive filtering	55
T. Adamski: Applications of quantile filtering to sampling time jitter correction in digital signal measurement system	91
K. Bieńkowski, M. Ghabally: The data exchanges and the message flow in the wireless robotics microprocessor DELTA computer networks	119
W. Niemiec: On the constitutive distributed parameter modelling of the singlecomponent real processes. Part I. The partial differential constitutive state equations describing the singlecomponent real processes	131

W. N i e m i e c: On the constitutive distributed parameter modelling of the singlecomponent real processes. Part II. The solution of partial differential constitutive state equations for the singlecomponent real processes	155
E. D a n i c k i, D. B o g u c k i: Mutual capacitance of two interdigital transducers in the metal case	187
J. S a w i c k i: Basic error of the method of the particular sampling	199
J. S i u z d a k: A novel method of reception of coherent fiber optic signals with binary ASK	219
J. S i u z d a k: On the gaussian approximation of the instantaneous frequency of the sources with non-zero spectral linewidth	225

ERRATA

W Kwartalniku Elektroniki i Telekomunikacji 1992, t. 38 z. 4. na stronie 713 tytuł artykułu p. Prof. Józefa Wojasa powinien mieć brzmienie:

„Analiza promieniowania elektromagnetycznego w fotoelektronowych
badaniach powierzchniowych”

W artykule p. Prof. Józefa Wojasa p.t. „Dyskusja teorii fotoemisji” w podpisie pod rysunkiem 6 wydrukowano: eksploatacje zamiast: ekstrapolacje. Na str. 575 wydrukowano: „stożek uciezki” zamiast „stożek ucieczki”.

Na str. 573, 15 wiersz od góry we wzorze nr 30: przed ilorazem $\frac{a}{b}$ nie potrzebny minus

Za błędy w artykułach Autora i Czytelników przeprasza

Redakcja

Subscription for external subscribers:

The promotional subscription price in 1993 is \$ 80 including postage for institutions and \$ 75 for individuals, including postage. Subscriptions should be sent too the publisher, Polish Scientific Publishers PWN Ltd, Journal Division, Miodowa 10, 00-251 Warsaw, POLAND, fax (48) (22) 26 09 50, (48) (22) 26 71 63, with a copy by fast to Electronics and Telecommunications Quarterly 00-665 Warsaw, ul. Nowowiejska 15/19, p. 470. Subscription is occupied after showing a checque or transfer documents. Our bank account is as follows:

Bank account: PBK VIII O/Warszawa nr 370028-1052-139-112. At subscriber's request this journal will be air mailed at additional postage to 50% of gross price to European countries and 65% overseas.

Kwartalnik Elektroniki i Telekomunikacji

WARUNKI PRENUMERATY

Wpłaty na prenumeratę przyjmowane są na okresy kwartalne:

na teren kraju

- jednostki kolportażowe „RUCH” S.A. i urzędy pocztowe oddawcze, właściwe dla miejsca zamieszkania lub siedziby prenumeratora, oraz doręczyciele w miejscowościach, gdzie dostęp do urzędu jest utrudniony,
- od osób lub instytucji, zamieszkających lub mieszczących się w miejscowościach, w których nie ma jednostek kolportażowych „RUCH”, wpłaty należy wносить do „RUCHU” S.A. Oddział Warszawa, 00-958 Warszawa, ul. Towarowa 28. Konto: PBK XIII Oddział Warszawa nr 370044-1195-139-11. „RUCH” S.A. zapewnia dostawę pod wskazanym adresem pocztą zwykłą w ramach opłaconej prenumeraty.

na zagranicę

- „RUCH” S.A. Oddział Warszawa, 00-958 Warszawa, ul. Towarowa 28. Konto: PBK XIII Oddział Warszawa nr 370044-1195-139-11. Dostawa odbywa się pocztą zwykłą w ramach opłaconej prenumeraty, z wyjątkiem zlecenia dostawy pocztą lotniczą, której koszt w pełni pokrywa zleceniodawca.

Prenumerata ze zleceniem dostawy za granicę jest o 100 % wyższa od krajowej.

Dostawa zamówionej prasy następuje:

- przez jednostki kolportażowe „RUCH” S.A. — w sposób uzgodniony z zamawiającym,
- prenumerata pocztowa — pod wskazanym adresem, w ramach opłaconej prenumeraty.

Terminy przyjmowania przez „RUCH” S.A. wpłat na prenumeratę krajową i zagraniczną oraz przez Poczta Polską (tylko prenumerata krajowa):

<u>„RUCH” S.A.</u>		<u>Poczta Polska</u>	
— do 20 XI	na I kw. roku następnego	— do 25 XI	na I kw. roku następnego
— do 20 II	na II kw.	— do 25 II	na II kw.
— do 20 V	na III kw.	— do 25 V	na III kw.
— do 20 VIII	na IV kw.	— do 25 VIII	na IV kw.

Bieżące numery można nabyć w Księgarni Wydawnictwa Naukowego PWN, ul. Miodowa 10, 00-251 Warszawa. Również można je nabyć, a także zamówić (przesyła za zaliczeniem pocztowym) we Wzorcowni Ośrodka Rozpowszechniania Wydawnictw Naukowych PAN, Pałac Kultury i Nauki, 00-901 Warszawa.

UWAGA!

Poczynając od 1993 r. prenumeratę naszego czasopisma prowadzi również Zakład Kolportażu Wydawnictwa SIGMA-NOT. Zamówienia na prenumeratę należy składać w terminach kwartalnych. Wpłata na I półr. wynosi 60 000 zł

Warunkiem przyjęcia prenumeraty jest wpłata na konto wartości zamówionych egzemplarzy. Wpłat na prenumeratę należy dokonywać na blankietach do wpłat na rachunki bankowe na konto:

Zakład Kolportażu Wydawnictwa SIGMA-NOT Sp. z o.o.
00-950 Warszawa nr PBK III O/Warszawa 370015-1573-139-11

Zakład Kolportażu oferuje również możliwość zamawiania prenumeraty ze zleceniem wysyłki za granicę. Jej cena jest dwukrotnie wyższa od ceny prenumeraty normalnej, a zlecający powinien podać dokładny adres odbiorcy za granicą. Dodatkowych informacji udziela:

Zakład Kolportażu Wydawnictwa SIGMA-NOT Sp. z o.o.
00-716 Warszawa, skr. poczt. 1004, ul. Bartycka 20.
Telefony: 40-30-86, 40-35-89, 40-00-21 w. 249, 295, 299

POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

Nr Indeksu 363189
PL ISSN 0035-9386

KWARTALNIK
ELEKTRONIKI I TELEKOMUNIKACJI
ELECTRONICS AND
TELECOMMUNICATIONS
QUARTERLY

TOM XXXIX — ZESZYT 2

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1993

**POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI**

**KWARTALNIK
ELEKTRONIKI I TELEKOMUNIKACJI
ELECTRONICS AND
TELECOMMUNICATIONS
QUARTERLY**

TOM XXXIX — ZESZYT 2

**WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1993**

RADA REDAKCYJNA

Przewodniczący

prof. dr inż. ADAM SMOLIŃSKI
członek rzeczywisty PAN

Członkowie

prof. dr hab. inż. DANIEL JÓZEF BEM, prof. dr hab. inż. MICHAŁ BIAŁKO – czł. koresp. PAN,
prof. dr hab. inż. STEFAN HAHN – czł. koresp. PAN, prof. dr inż. ANDRZEJ HAŁAS, prof. dr inż.
ZDZISŁAW KACHLICKI, prof. dr hab. inż. BOHDAN MROZIEWICZ, prof. dr inż. JERZY
OSIOWSKI, prof. dr inż. WITOLD ROSIŃSKI – czł. rzecz. PAN, prof. dr hab. inż. STEFAN
WĘGRZYN – czł. rzecz. PAN, prof. dr hab. inż. WIESŁAW WOLIŃSKI – czł. koresp. PAN,
prof. dr inż. ANDRZEJ ZIELIŃSKI, prof. dr inż. MARIAN ZIENTALSKI

REDAKCJA

Redaktor Naczelny

prof. dr hab. inż. WIESŁAW WOLIŃSKI

Zastępca Redaktora Naczelnego

doc. dr inż. KRYSTYN PLEWKO

Sekretarz Odpowiedzialny

mgr KRYSTYNA LELAKOWSKA

ADRES REDAKCJI

00-665 Warszawa, ul. Nowowiejska 15/19 Politechnika, pok. 470
Instytut Telekomunikacji, Gmach im. prof. JANUSZA GROSZKOWSKIEGO

Dyżury Redakcji: środy i piątki, godz. 14–16
tel. 628 89 81 21007-737

Telefony domowe: Redaktora Naczelnego: 12 17 65

Zast. Red. Naczelnego: 26 83 41

Sekretarza Odpowiedzialnego: 25 29 18

WYDAWNICTWO NAUKOWE PWN

Warszawa, ul. Miodowa 10

Ark. druk. 13,25	Podpisano do druku w listopadzie 1993 r.
Papier offsetowy kl. III 80 g. B-1	Druk ukończono w grudniu 1993 r.

Skład: Wyd. **GP-BIT** W-wa ul. Marymoncka 34 Druk i oprawa: Drukarnia Braci Grodzickich,
Żabienice ul. Przelotowa 7

Probability distribution of sampling time jitter in equivalent time digital oscilloscopes

TOMASZ ADAMSKI

Instytut Podstaw Elektroniki, Politechnika Warszawska

Received 1992.10.08

Authorized 1993.01.08

The presence of time jitter between the trigger signal and the sampling pulse in equivalent-time digital oscilloscopes can cause appreciable distortion of the recorded waveform. If the probability distribution of the sampling time jitter is known it is possible to assess and correct this distortion by appropriate algorithms. The paper deals with methods for calculation of the probability distribution of the random variable describing the sampling time jitter.

1. INTRODUCTION

In electronic measurement systems with sampling (like digital oscilloscopes, fast data acquisition systems or picosecond waveform recorders) user cannot neglect random unstability of sampling, called sampling time jitter. This unstability is due to inherent noise phenomena and to some extent also to the method of sampling. In equivalent time digital oscilloscopes the jitter causes troublesome noisy scattering of the measured signal image and appreciable distortions of the recorded waveform. On the other hand the fast progress in sampling head design (see [9]) and measurements of signals in picosecond region create need of exact analysis and assessment of the sampling time jitter (abbreviated in the sequel to the STJ). The paper deals with the STJ present in equivalent time waveform recorders (ETWR) or digital oscilloscopes with repetitive sampling. The STJ is caused by the time jitter (or in other words phase noise) between the trigger signal and the sampling pulse during the triggering process. The simplest mathematical model of the triggering process is a real random variable τ called the first upcrossing time of the linear barrier (see Fig.1.). We would like to compute the probability distribution P_τ of the random variable τ , but we will consider a little more general problem: how can be computed the probability distribution P_τ of the random variable τ being the first upcrossing time of a moving continuous barrier $Br \in C(\mathbb{R}^+)$ by a real continuous stochastic process $(X(t))_{t \in \mathbb{R}^+}$. Some simple analytical methods (both exact and approximate) are

proposed in the paper to calculate the first upcrossing time distribution P_τ . These methods, called in the paper fast moving barrier methods or shortly FMB methods can be considered as a modification of the Rice series method.

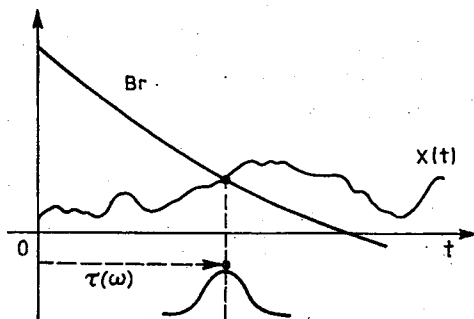


Fig. 1. The first upcrossing time of the moving barrier Br by the stochastic process $(X(t))_{t \in \mathbb{R}^+}$.

The essence of the FMB methods consists in replacing the P_τ computation problem by easier one: the computation of the mean value of upcrossings (of a moving continuous barrier $Br \in C(\mathbb{R}^+)$ by a continuous stochastic process $(X(t))_{t \in \mathbb{R}^+}$ on the fixed interval $I \subseteq \mathbb{R}^+$ and factorial moments. The mean value of upcrossings is given in the case of Gaussian stationary process $(X(t))_{t \in \mathbb{R}^+}$ and the barrier $Br(t) \equiv \text{const.}$ (the barrier is a level) by the well known Rice formula. In the paper, this formula is generalized on the nonstationary, nongaussian case with the barrier $Br \in C^{(1)}(\mathbb{R}^+)$. Sample paths of the process $(X(t))_{t \in \mathbb{R}^+}$ are not assumed to be differentiable. In the section 5 we prove that for a sufficiently fast linear barrier Br and a Gaussian process $(X(t))_{t \in \mathbb{R}^+}$, under not specially strict assumptions the probability distribution P_τ can be considered as Gaussian.

The moving barrier crossings problems are useful in noise and reliability assessment of many electronic systems (for example in PCM systems, phase meters, sampling oscilloscopes, trigger circuits or pulse generators).

All proofs of given in the paper theorems can be found in [21]. Basic facts and definitions concerning discussed problems are given in a systematic way in [1], [2], [3], [4] and [21].

2. THE FIRST UPCROSSING TIME

In a variety of practical problems involving random processes it is necessary to have statistical information on their level crossing (or more general, moving barrier crossing) properties. One of the most useful random variable giving such information is the first upcrossing time. Let $(X(t))_{t \in \mathbb{R}^+}$ be an arbitrary stochastic process with values in n -dimensional Euclidean space $\mathbb{R}^n$ and B a Borel subset of $\mathbb{R}^n$. The first entrance time τ (of the process $(X(t))_{t \in \mathbb{R}^+}$) to the set B is the function

$\tau: \Omega \rightarrow \mathbb{R}^+ \cup \{+\infty\}$ defined for each $\omega \in \Omega$ (where $(\Omega, \mathfrak{M}, P)$ is a probabilistic space), by the formula

$$\tau(\omega) = \begin{cases} \inf A & \text{if } A \neq \phi \\ +\infty & \text{if } A = \phi \end{cases} \quad (2.1)$$

where $A = \{t \in \mathbb{R}^+; X(t)(\omega) \in B\}$

Assume Br is a real continuous function over $\mathbb{R}^+$ i.e. $Br \in C(\mathbb{R}^+)$, which will be called in the sequel the barrier or moving barrier. The first upcrossing time of a continuous moving barrier $Br \in C(\mathbb{R}^+)$ by a real stochastic process $X(t)_{t \in \mathbb{R}^+}$ can be defined as the first entrance time of the two-dimensional process $(t, X(t))_{t \in \mathbb{R}^+}$ to the set $\{(t, x) \in \mathbb{R}^2; t \in \mathbb{R}^+, x > Br(t)\}$ or equivalently by the following formula

$$\tau(\omega) = \begin{cases} \inf \{t \in \mathbb{R}^+; X(t)(\omega) > Br(t)\} & \text{if } \{t \in \mathbb{R}^+; X(t)(\omega) > Br(t)\} \neq \phi \\ +\infty & \text{if } \{t \in \mathbb{R}^+; X(t)(\omega) > Br(t)\} = \phi. \end{cases} \quad (2.2)$$

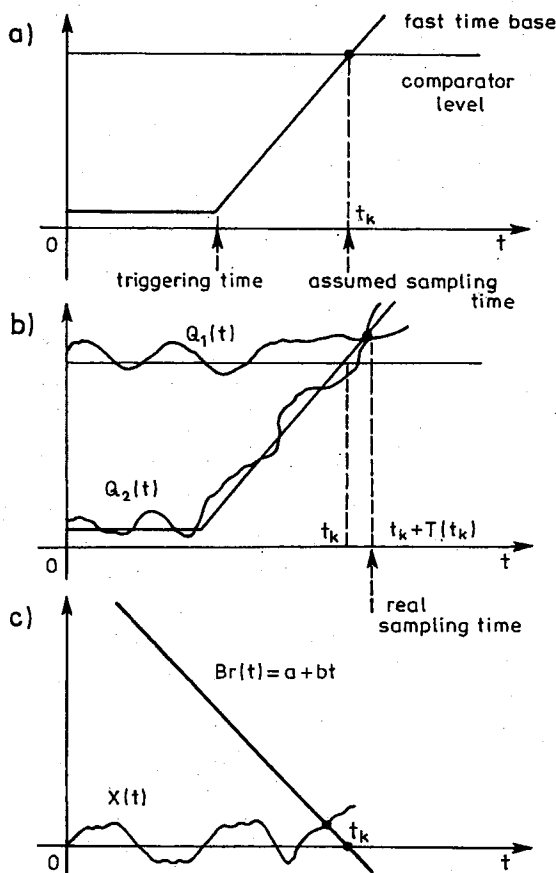


Fig. 2. Sampling time jitter in equivalent time waveform recorders as a linear barrier upcrossing time
 a) deterministic model without noise: b) probabilistic model with additive noise $(Q_1(t))_{t \in \mathbb{R}^+}$ and $(Q_2(t))_{t \in \mathbb{R}^+}$
 c) simplified probabilistic model, $X(t) = Q_1(t) - Q_2(t)$

In particular, if the moving barrier Br is a line parallel to the time axis then the first upcrossing time is called the first level upcrossing time.

If the process $(X(t))_{t \in \mathbb{R}^+}$ is continuous and $Br \in C(\mathbb{R}^+)$ then τ is a random variable i.e. τ is a $(\mathfrak{M}, B(\mathbb{R}^+))$ measurable function.

This section does not deal with theoretical aspects and properties of the random variable τ for example problem when $E(\tau) < +\infty$ or $D^2(\tau) < +\infty$. Solutions of such problems can be found in [1], [2] or [21].

Many noise phenomena in electronic circuits can be described with the random variable τ , for instance sampling time jitter in digital oscilloscopes or unstability of triggering time. The barrier crossing model describing the mechanism of the STJ is shown in the Fig.2

In the next section we describe exact and approximate methods for distribution calculation of the random variable τ . The problem of P_τ assesment is not new and partially solved (see [1]–[7], [21]) but in general formulation (even for Gaussian processes $(X(t))_{t \in \mathbb{R}^+}$) is difficult. For Markov processes it is possible to solve this problem with the Fokker–Planck equation¹ or Pontriagin equations (see [3]) but in general (i.e. nonmarkovian case) an only known method is the method of Rice series and Longuet–Higgins series, valuable under very specific assumptions.

3. THE MEAN NUMBER OF MOVING BARRIER UPCROSSINGS

Formulae for the mean number of level crossings for real, continuous, stationary, Gaussian stochastic processes were obtained by S.O.Rice many years ago. The stationary nongaussian case was analyzed by M.R.Leadbetter in [1]. Nonstationary, nongaussian cases (but for differentiable sample paths) were discussed in [3]. The section deals with formulae for the mean number of moving barrier crossings in nonstationary, nongaussian case, when sample paths are not assumed to be differentiable.

Let $N^+(Br, I)$ denotes the number of the barrier $Br \in C(\mathbb{R}^+)$ upcrossings by a stochastic process $(X(t))_{t \in \mathbb{R}^+}$ on the time interval $I \subset \mathbb{R}^+$. If $(X(t))_{t \in \mathbb{R}^+}$ is a continuous, real process and Br is a continuous function over $\mathbb{R}^+$ then $N^+(Br, I)$ is a random variable (see [1]). Under not specially strict assumptions, finding the probability distribution of τ can be considered as the easier problem of $E(N^+(Br, I))$ computation.

Two following theorems are basic for developed in the sequel, method of P_τ calculation (called in the sequel Fast Moving Barrier method or shortly FMB method). The FBM method is particularly well suited to STJ or phase noise analysis in pulse and digital circuits.

Theorem 3.1. (on a mean number of moving barrier upcrossings by stochastic process). Assume, $(X(t))_{t \in \mathbb{R}^+}$ is a real, continuous stochastic process defined on the probabilistic space $(\Omega, \mathfrak{M}, P)$, with the complete probabilistic measure P , $Br \in C^{(1)}(\mathbb{R}^+)$ is a function describing moving barrier, $I \subset \mathbb{R}^+$ is a fixed bounded

¹ This method is used widely in noise properties assesment of PLL (phase locked loop) systems.

interval and there is such $q_0 \in \mathbb{R}^+ \setminus \{0\}$ that for each $t \in \mathbb{R}^+$ and each $q \in (0, q_0]$ the random variable $(X(t), Y_q(t))$ (where $Y_q(t) \stackrel{\text{df}}{=} X(t+q)/q - X(t)/q$) have the joint density $g_{q,t} : \mathbb{R}^2 \ni (u, z) \rightarrow g_{q,t}(u, z) \in \mathbb{R}^+$ with the following properties:

1) the function $\mathbb{R} \ni u \rightarrow g_{q,t}(u, z) \in \mathbb{R}^+$ is uniformly continuous independently of $(q, t, z) \in (0, q_0] \times \mathbb{I} \times \mathbb{R}$ (i.e. for each $\varepsilon > 0$ there is such a $\delta > 0$ that for each $(q, t, z) \in (0, q_0] \times \mathbb{I} \times \mathbb{R}$ and each $u_1, u_2 \in \mathbb{R}$ the following implication is true $|u_1 - u_2| < \delta \Rightarrow |g_{q,t}(u_1, z) - g_{q,t}(u_2, z)| < \varepsilon$)

2) the function $\mathbb{R} \ni z \rightarrow g_{q,t}(u, z) \in \mathbb{R}^+$ is uniformly continuous independently of $(q, t, u) \in (0, q_0] \times \mathbb{I} \times \mathbb{R}$

3) for each $q \in (0, q_0]$ the function $\mathbb{R}^+ \times \mathbb{R}^2 \ni (t, u, z) \rightarrow g_{q,t}(u, z) \in \mathbb{R}^+$ is continuous

4) there is such a function $p : \mathbb{R}^+ \times \mathbb{R}^2 \rightarrow \mathbb{R}^+$ that the convergence

$$g_{q,t}(u, z) \xrightarrow{q \rightarrow 0} p(t, u, z) \text{ is uniform for } (t, u, z) \in \mathbb{I} \times \mathbb{R}^2$$

5) there is a measurable function $h : \mathbb{R}^+$ such that $\int_{\mathbb{R}^+} z h(z) l_1(dz) < +\infty$

and $g_{q,t}(u, z) \leq h(z)$ for each $q \in (0, q_0]$, $u \in \mathbb{R}$, $z, t \in \mathbb{R}^+$

$$\text{then } E(N^+(\text{Br}, \mathbb{I})) = \int_{\mathbb{I}} \left(\int_{\mathbb{R}^+} z p(t, \text{Br}(t), z + \text{Br}'(t)) l_1(dz) \right) l_1(dt), \quad (3.1)$$

where l_1 is the Lebesgue measure on $\mathbb{R}$.

Proof see [21]. In this theorem, we do not assume that the real stochastic process $X(t)_{t \in \mathbb{R}^+}$ is stationary nor that all sample paths are differentiable. If the process $X(t)_{t \in \mathbb{R}^+}$ is stationary in the strict sense then the above theorem can be formulated in the following way.

Theorem 3.2. Assume, $(X(t))_{t \in \mathbb{R}^+}$ is a real, continuous, stationary in the strict sense stochastic process defined on the probabilistic space $(\Omega, \mathfrak{M}, P)$, with the complete probabilistic measure P , $\text{Br} \in C^{(1)}(\mathbb{R}^+)$ is a function describing moving barrier, $\mathbb{I} \subset \mathbb{R}^+$ is a fixed bounded interval and there is such $q_0 \in \mathbb{R}^+ \setminus \{0\}$ that for each $t \in \mathbb{R}^+$ and each $q \in (0, q_0]$ the random variable $(X(0), Y_q(0))$ (where $Y_q(0) \stackrel{\text{df}}{=} X(q) - X(0)/q$) have the joint density $g_q : \mathbb{R}^2 \ni (u, z) \rightarrow g_q(u, z) \in \mathbb{R}^+$ with the following properties:

1) the function $\mathbb{R} \ni u \rightarrow g_{q,t}(u, z) \in \mathbb{R}^+$ is uniformly continuous in u independently of $(q, z) \in (0, q_0] \times \mathbb{R}$ and the function $\mathbb{R} \ni z \rightarrow g_q(u, z) \in \mathbb{R}^+$ is uniformly continuous independently of $(q, u) \in (0, q_0] \times \mathbb{R}$

2) for each $q \in (0, q_0]$ the function $\mathbb{R}^+ \ni (u, z) \rightarrow g_q(u, z) \in \mathbb{R}^+$ is continuous

3) there is such a function $p : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ that the convergence

$$g_q(u, z) \xrightarrow{q \rightarrow 0} p(u, z) \text{ is uniform for } (u, z) \in \mathbb{R}^2$$

4) there is a measurable function $h: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that $\int_{\mathbb{R}^+} z h(z) l_1(dz) < +\infty$ and $g_q(u, z) \leq h(z)$ for each $q \in (0, q_0]$, $u \in \mathbb{R}$, $z \in \mathbb{R}^+$

$$\text{then } E(N^+(\text{Br}, I)) = \int_I \left(\int_{\mathbb{R}^+} z p(\text{Br}(t), z + \text{Br}'(t)) l_1(dz) \right) l_1(dt). \quad (3.2)$$

Proof see [21].

Theorem 3.3. Assume the stochastic process $(X(t))_{t \in \mathbb{R}^+}$ is real, continuous, Gaussian and stationary with the mean value 0 and the correlation function $R(t)$, (with $R(0) > 0$). If exists $R''(0)$ and $\text{Br} \in C^{(1)}(\mathbb{R}^+)$ then the random variable $N^+(\text{Br}, I)$ has for each bounded interval $I \subset \mathbb{R}^+$ the mean value $E(N^+(\text{Br}, I)) < +\infty$ and

$$E(N^+(\text{Br}, I)) = \frac{1}{2\pi\sqrt{R(0)}} \int_I \exp\left(-\left(\frac{\text{Br}(t)^2}{2R(0)}\right)\right) (\sqrt{-R''(0)}) \cdot \exp\left(-\frac{(\text{Br}'(t))^2}{2(-R''(0))}\right) - \text{Br}'(t)\sqrt{2\pi} \Phi\left(-\frac{\text{Br}'(t)}{\sqrt{-R''(0)}}\right) l_1(dt), \quad (3.3)$$

where Φ is the probability integral $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{y^2}{2}\right) dy$.

Proof see [21]. From the above theorem we obtain the classical, well known result, so called the Rice formula (see [1]–[6]).

Corollary 3.4. (the classical Rice formula). Assume the stochastic process $(X(t))_{t \in \mathbb{R}^+}$ is real, continuous, Gaussian and stationary with the mean value 0 and the correlation function $R(t)$, (with $R(0) > 0$). If exists $R''(0)$ then the random variable $N^+(\text{Br}, I)$ for $\text{Br}(t) \equiv u$ (where u is arbitrary fixed real number) has for each bounded interval $I \subset \mathbb{R}^+$ the mean value $E(N^+(\text{Br}, I)) < +\infty$ and

$$E(N^+(\text{Br}, I)) = \frac{l_1(I)}{2\pi} \sqrt{\frac{-R''(0)}{R(0)}} \exp\left(\frac{-u^2}{2 \cdot R(0)}\right). \quad (3.4)$$

4. THE FAST MOVING BARRIER METHOD OF P_t CALCULATION

The fast moving barrier method of P_t calculation is based on the following theorem.

Theorem 4.1. Let $(X(t))_{t \in \mathbb{R}^+}$ be a real stochastic process defined on the probabilistic space $(\Omega, \mathfrak{M}, P)$ with a complete probabilistic measure P . Assume one dimensional probability distributions of the process $(X(t))_{t \in \mathbb{R}^+}$ are continuous and the moving barrier $Br \in C(\mathbb{R}^+)$. If there is an interval $I = [0, x_0] \subseteq \mathbb{R}$ such that $P(N^+(Br, I) \in \{0, 1\}) = 1$ then for each interval $I_1 \subseteq I \setminus \{0\}$ we have

$$P(\tau \in I_1) = E(N^+(Br, I_1)) \quad (4.1)$$

and for each interval $I_1 \subseteq I$ such that $0 \in I_1$

$$P(\tau \in I_1) = P(X(0) > Br(0)) + E(N^+(Br, I_1)), \quad (4.2)$$

where τ is the first upcrossing time of the moving barrier Br by the process $(X(t))_{t \in \mathbb{R}^+}$. In particular for each $x \in [0, x_0]$, we obtain

$$F_\tau(x) = E(N^+(Br, [0, x])) + P(X(0) > Br(0)) \quad (4.3)$$

where F_τ is a distribution function of the random variable τ .

Proof see [21]. The probability $P(X(0) > Br(0))$ in (4.2) is in practical situations very small and can be neglected. It can also be proved that if condition $P(N^+(Br, I) \in \{0, 1\}) = 1$ is not fulfilled the proposed method under weak assumptions becomes approximate with very small errors which can be easily computed (for more information see [21]).

From theorems 3.1 and 4.1 we obtain directly the following result.

Theorem 4.2. If real, continuous, stochastic process $(X(t))_{t \in \mathbb{R}^+}$ fulfills assumptions of the theorem 3.1 and τ is the first upcrossing time of a moving barrier $Br \in C^{(1)}(\mathbb{R}^+)$ by the stochastic process $(X(t))_{t \in \mathbb{R}^+}$ and there is such real number x_0 , that $P(N^+(Br, [0, x_0]) = 1) = 1$ then using notation from theorem 3.1 and 4.1 we have for each $x \in \mathbb{R}$

$$F_\tau(x) = \begin{cases} P(X(0) > Br(0)) + \int_{[0, x]} \int_{\mathbb{R}^+} z p(t, Br(t), z + Br'(t)) l_1(dz) l_1(dt) & \text{for } x \in [0, x_0] \\ 1 & \text{for } x > x_0 \\ 0 & \text{for } x < 0 \end{cases} \quad (4.4)$$

The formula (4.4) implies that the probability measure P_τ is a sum of Dirac measure concentrated in the point $t = 0$ (practically negligible in most cases) and the measure having the density f_τ given by the formula:

$$f_\tau(t) = \begin{cases} \int_{\mathbb{R}^+} z p(t, Br(t), z + Br'(t)) l_1(dt) & \text{for } t \in [0, x_0] \\ 0 & \text{for } t \notin [0, x_0]. \end{cases} \quad (4.5)$$

The n -th factorial moment of a random variable X with values in $N \cup \{0\}$ is the mean value $E(X \cdot (X-1) \cdot (X-2) \cdot \dots \cdot (X-n+1))$. In the sequel the n -th factorial moment of the random variable $N^+(\text{Br}, [0, t])$ is denoted by $\text{FM}(n, t)$.

Theorem 4.3. Let a real, continuous stochastic process $(X(t))_{t \in R^+}$ (defined on the probabilistic space with a complete probabilistic measure P) has continuous one dimensional distribution functions. Assume, the moving barrier is continuous i.e. $\text{Br} \in C(R^+)$ and for each $t \in R^+$ there is such a $k(t) \in N$ that $N^+(\text{Br}, [0, t]) \leq k(t) = 1$ (i.e. on the bounded interval $[0, t]$ P almost each sample path has the finite number of upcrossings) then

$$\sum_{i=1}^{k(t)} \frac{(-1)^{i+1}}{i!} \text{FM}(i, t) \leq P(\tau < t) \leq \sum_{i=1}^{k(t)} \frac{(-1)^{i+1}}{i!} \text{FM}(i, t) + P(X(0) > \text{Br}(0)). \quad (4.6)$$

If additionally $P(X(0) > \text{Br}(0)) = 0$ then

$$P(\tau < t) = 1 - P(N^+(\text{Br}, [0, t]) = 0) = \sum_{i=1}^{k(t)} \frac{(-1)^{i+1}}{i!} \text{FM}(i, t). \quad (4.7)$$

Proof see [21]. The above theorem gives both exact and approximate method for $P(\tau < t)$ calculation. The method is based on factorial moments, which are sometimes difficult for analytical assesment. It is of course disadvantage of the method but on the other hand we can easily estimate these moments from measurement (number of crossings measurement is much simpler than direct time measurement, particularly for wide band signals). Hence formulas (4.1), (4.6) and (4.7) can be also useful in experimental approach.

Assume, that in the theorem 4.1 the condition: "there is such a $x_0 \in R^+ \setminus \{0\}$ that $P(N^+(\text{Br}, [0, x])_0 \in \{0, 1\}) = 1$ " is not fulfilled and instead we have in a point $t \in R^+ \setminus \{0\}$, $P(N^+(\text{Br}, [0, t]) > 1) > 0$. In this case we easily asses the error $\varepsilon(t) \triangleq |F_\tau(t) - E(N^+(\text{Br}, [0, t])) - P(X(0) > \text{Br}(0))|$ introduced by the formula (4.3) in the point t with the inequality $\varepsilon(t) \leq \text{FM}(2, t)$. Hence the FBM method exact in formulation given in the theorem 4.1 can be also approximate one.

The mathematical basis of the sketched above FMB method for P_τ calculation is systematically developed in [21]. Other techniques for STJ assesment based on computer simulation and numerical methods were discussed for example in [20] and [21].

5. PROBABILITY DISTRIBUTION OF THE RANDOM VARIABLE τ FOR THE LINEAR BARRIER,

In this section we will show using a simple barrier crossing model, that under certain assumptions, the probability distribution of the random variables $T(t_1), T(t_2), \dots, T(t_M)$ describing the STJ in equivalent time points $t_1, t_2, \dots, t_M$ can be considered as Gaussian.

Assume that the real stochastic process $(X(t))_{t \in \mathbb{R}^+}$ is continuous, stationary and Gaussian with a correlation function $R(t)$, the barrier Br is linear and the assumptions of the theorem 3.1 are fulfilled. Applying described above FBM method and the formula (3.3) to the linear moving barrier given as $Br(t) = a + bt$, $a > 0$, $b < 0$, we obtain for arbitrary bounded interval $I \subset \mathbb{R}^+$

$$E(N^+(Br, I)) = \frac{1}{2\pi\sqrt{R(0)}} \int_I \exp\left(-\frac{(a+bt)^2}{2R(0)}\right) \left(\sqrt{-R''(0)} \exp\left(-\frac{b^2}{2(-R''(0))}\right) - b\sqrt{2\pi} \Phi\left(-\frac{b}{\sqrt{-R''(0)}}\right) \right) l_1(dt), \quad (5.1)$$

where Φ is the probability integral (i.e. the distribution function of the probability distribution $N(0,1)$ which gives according to (4.5) $f_t(t) = g_t(t)$ for $t \geq 0$ and $f_t(t) = 0$ for $t < 0$, where

$$g_t(t) = \frac{1}{2\pi\sqrt{R(0)}} \exp\left(-\frac{(a+bt)^2}{2R(0)}\right) \left(\sqrt{-R''(0)} \exp\left(-\frac{b^2}{2(-R''(0))}\right) - b\sqrt{2\pi} \Phi\left(-\frac{b}{\sqrt{-R''(0)}}\right) \right). \quad (5.2)$$

Let t_0 denotes the crossing point of linear moving barrier with the time axis i.e. $t_0 = -a/b$. From the formula (5.2) we obtain that if $a/b = \text{const.}$ and $b \rightarrow -\infty$ (i.e. the slope of the linear barrier becomes greater with the constant upcrossing point t_0) then for each $t \in \mathbb{R}$

$$\frac{g_t(t)}{\frac{-b}{\sqrt{2\pi R(0)}} \exp\left(-\frac{b^2(t+t_0)^2}{2R(0)}\right)} = \frac{\sqrt{-R''(0)} \exp\left(\frac{b^2}{2R''(0)}\right)}{-b\sqrt{2\pi}} + \Phi\left(\frac{-b}{\sqrt{-R''(0)}}\right) \rightarrow 1 \quad (5.3)$$

and $P(X(0) > Br(0)) \rightarrow 0$ when $b \rightarrow -\infty$. Hence for a sufficiently fast moving barrier (for example for $b/\sqrt{-R''(0)} > 3$) we can replace without significant error, the function $g_t(t)$ by the probability density function of the Gaussian distribution $N(0, -\sqrt{R(0)}/b)$. For typical noise parameters in ETWR the condition $b/\sqrt{-R''(0)} \gg 3$ seems to be fulfilled, which leads to the conclusion that STJ distribution in ETWR (i.e. distributions of random variables $T(t_1), T(t_2), \dots, T(t_M)$) can be considered as Gaussian. This effect can be also observed experimentally (see [12], [17]).

6. ERRORS INTRODUCED BY SAMPLING TIME JITTER

The STJ limits significantly accuracy of wide band sampling measurement systems.

This section deals with measurement errors caused by the STJ in such systems, particularly in ETWR. Assume the analog input signal $f: \mathbb{R} \rightarrow \mathbb{R}$ is sampled in equivalent time points $t_1, t_2, \dots, t_M$ (in particular we can have $t_k = k \cdot \Delta t$ but it is not important in the sequel).

Sampling time points $t_k, k \in \{1, 2, \dots, M\}$ are not exactly real sampling points. Because of noise phenomena present in sampling circuits real sampling points are randomly shifted on the time axis. This shifting is called the sampling time jitter. In relatively slow systems these shifts are not critical and errors caused by jitter can be neglected. In fast systems, like wide band equivalent-time waveform recorders or sampling oscilloscopes random shifting is in the range 10–200 ps and causes troublesome distortion of the measured signal.

Let $T(t_k)$ be the random variable describing the random shift of the real sampling point $t_k + T(t_k)$ related to assumed, deterministic one t_k . According to previous sections the random variables $T(t_1), T(t_2), \dots, T(t_M)$ can be considered as the first upcrossing times of linear moving barrier by the stochastic process $(X(t))_{t \in \mathbb{R}^+}$ describing noise. Because, in ETWR we repeat the sampling at the each point t_k (where $k = 1, 2, \dots, M$ and M is number of sampling points) in independent manner, we can describe the all sampling process by M random sequences $(f(t_k + T_n(t_k)))_{n=1}^{\infty}$ for $k = 1, 2, \dots, M$, where $(T_n(t_k))_{n=1}^{\infty}$ is a sequence of independent random variables with distributions equal to the distribution of $T(t_k)$. In practice, the number of samples for each k is limited to N_0 , then the input information of the ETWR consists of M random vectors of N_0 dimensions or in other words, of M finite random sequences $(f(t_k + T_n(t_k)))_{n=1}^{N_0}$ for $k = 1, 2, \dots, M$.

In the sequel we assume that the input signal is a $(B(\mathbb{R}), B(\mathbb{R}))$ measurable function $f: \mathbb{R} \rightarrow \mathbb{R}$ and is sampled in time points $t_1 + T(t_1)$, where $t_1, t_2, \dots, t_M \in [a, b]$ are deterministic sampling points. We would like to assess measurement errors introduced by sampling time jitter. In other words, we want to assess the value $er(t) \stackrel{\text{def}}{=} f(t + T(t)) - f(t)$ for a fixed sampling point t particularly for $t_1, t_2, \dots, t_M \in [a, b]$.

If the signal f is a $(B(\mathbb{R}), B(\mathbb{R}))$ measurable function (for instance f is continuous) then $er(t)$ is for each sampling point t a random variable but its probability distribution depends on f and is in most cases difficult to exact calculations.

Assume that jitter is limited (i.e. there is such a number $\varepsilon > 0$ that for each sampling point t , we have $|T(t)| \leq \varepsilon$). If the signal $f \in C^{(n)}(\mathbb{R})$ then from Taylor's formula we obtain

$$|er(t)| \leq \sum_{k=1}^{n-1} \frac{|f^{(k)}(t)|}{k!} \varepsilon^k + \frac{\varepsilon^n}{n!} \sup_{h \in [-\varepsilon, \varepsilon]} |f^{(n)}(t+h)|. \quad (6.1)$$

If $f^{(n+2)} \in C(\mathbb{R})$, $f^{(k)} \in L^1(\mathbb{R}, \mathcal{L}, l_1)$ for $k = 0, 1, \dots, n+2$ and $\mathcal{F}$ denotes the Fourier

transform then $\mathcal{F}(f^{(n)})(\omega) = (j\omega)^n \mathcal{F}(f)(\omega)$. Hence $|f^{(n)}(t)| \leq \frac{1}{2\pi} \int_{\mathbb{R}} |\omega|^n |F(\omega)| l_1(d\omega)$,

where F is a spectrum of the signal f . Then from (6.1), we have the following inequality

$$\begin{aligned}
 |er(t)| &\leq \sum_{k=1}^{n-1} \frac{|f^{(k)}(t)|}{k!} \varepsilon^k + \frac{\varepsilon^n}{2\pi n!} \int_{\mathbb{R}} |\omega|^n |F(\omega)| l_1(d\omega) \leq \\
 &\leq \frac{1}{2\pi} \sum_{k=1}^n \frac{\int_{\mathbb{R}} |\omega|^k |F(\omega)| l_1(d\omega) \varepsilon^k}{k!} \leq \frac{1}{\pi} \int_{\mathbb{R}^+} (e^{\varepsilon|\omega|} - 1) |F(\omega)| l_1(d\omega).
 \end{aligned} \quad (6.2)$$

If the spectrum F of the input signal f is limited i.e. there is such a ω_g , that $\text{supp } F \subset [-\omega_g, \omega_g]$ then the integral on the right side of (6.2) is finite and we obtain

$$|er(t)| \leq \frac{1}{\pi} \int_{[0, \omega_g]} (e^{\varepsilon\omega} - 1) |F(\omega)| l_1(d\omega). \quad (6.3)$$

The simplest but rather pessimistic assessment of $er(t)$, can be also obtained, when limited by ω_g spectrum of the signal f and limited jitter is assumed. In this case, specifically for $f(t) = A \cdot \sin(\omega_g t)$, where $A > 0$, we have $|er(t)| \leq A \cdot \omega_g \varepsilon$. In general, we do not assume, that the sampling time jitter is limited. In such a case, if the signal $f \in C^{(1)}(\mathbb{R})$, $T(t) \in L^2(\Omega, \mathcal{M}, P)$ and $\sup_{t \in \mathbb{R}} |f'(t)| < +\infty$ then from the Lagrange theorem about the mean value we have

$$D^2(er(t)) \leq \left(\sup_{t \in \mathbb{R}} |f'(t)| \right)^2 \cdot D^2(T(t)) \quad (6.4)$$

and for limited jitter

$$D^2(er(t)) \leq \varepsilon^2 \left(\sup_{t \in [a-\varepsilon, b+\varepsilon]} |f'(t)| \right)^2. \quad (6.5)$$

If $f \in C^{(3)}(\mathbb{R})$ and $f, f^{(k)} \in L^1(\mathbb{R}, \mathcal{L}, l_1)$ for $k = 1, 2, 3$ then (denoting spectrum of the signal f by F) the formulas (6.4) and (6.5) can be rewritten as (6.6) and (6.7).

$$D^2(er(t)) \leq \frac{2}{\pi} \left(\int_{\mathbb{R}^+} \omega |F(\omega)| d\omega \right) \cdot D^2(T(t)) \quad (6.6)$$

$$D^2(er(t)) \leq \frac{2\varepsilon^2}{\pi} \left(\int_{\mathbb{R}^+} \omega |F(\omega)| d\omega \right)^2. \quad (6.7)$$

If $f(t) = A \sin(\omega_g t)$ then from (6.4) we have $D^2(er(t)) \leq A^2 \omega_g^2 D^2(T(t))$.

The above error assessment deal with arbitrary sampling system where the STJ is significant. Assume that analysed in the sequel sampling system is an equivalent time waveform recorder. In this case for each $k = 1, 2, \dots, M$ a sequence of samples $(f(t_k + T_n(t_k)))_{n=1}^{N_k}$ can be gathered creating input data. This data directly displayed on the screen gives scattered (we say jittered) image of the measured signal. In general the statistics $(f(t_k + T_n(t_k)))_{n=1}^{N_k}$ taken for every $k = 1, 2, \dots, M$ allow to correct

distortions caused by jitter, but in principle the restoration needs deconvolution algorithm or in the case when jitter statistics depends on k , solution of the integral Fredholm equation of the first kind. Hence correction leads to the ill posed problem very sensitive to input data and needs complicated numerical regularization techniques (see [8]). In this context, simple methods, like mean value or quantile statistics computations, working quite correctly only under certain assumptions become attractive.

The simplest method to correct scattered signal image (mostly used by digital oscilloscopes producers) is the mentioned mean value computation i.e. the averaging method. In this method for each sampling point t the number

$$MN(t, N_0) \stackrel{\text{def}}{=} \frac{1}{N_0} \sum_{i=1}^{N_0} f(t+T_i(t)) \quad (6.8)$$

is computed. Let $h(t) \stackrel{\text{def}}{=} E(f(t+T_n(t)))$ and $g(\cdot, t)$ denotes the probability density function of the random variable $T(t)$ describing the STJ in the time point t . From the law of great numbers we have $MN(t, N_0) \rightarrow h(t)$ (where $N_0 \rightarrow +\infty$) with probability 1. Hence for sufficiently large N_0 we can admit

$$h(t) = \int_{\mathbb{R}} g(x, t) f(x) I_1(dx). \quad (6.9)$$

The function $t \rightarrow h(t)$ is obtained from measurement. Hence, when the STJ is present, arithmetic averaging of the set of samples related to fixed equivalent time sampling moment limits the band of the system. Below, we will assess measurement errors, when averaging is applied.

The averaging method is quite correct for linear signal $f(t)=At+B$, (where $A, B \in \mathbb{R}$) i.e. $MN(t, N_0) \rightarrow f(t)$ with probability 1 but in general (for example for pulse signals) it introduces some errors. These errors can be easily assessed in the following way.

We would like to find $|h(t)-f(t)|$ and $\sup_{t \in \mathbb{R}} |h(t)-f(t)|$ for fixed sampling time t . Assume that the input signal $f \in C^{(n)}(\mathbb{R})$ and for each $t \in \mathbb{R}$, $T(t) \in L^n(\Omega, \mathfrak{M}, P)$, $f(t+T(t)) \in L^1(\Omega, \mathfrak{M}, P)$. Using Taylor's formula we obtain

$$|h(t)-f(t)| \leq \sum_{k=1}^{n-1} \frac{|f^{(k)}(t)|}{k!} |E((T(t))^k)| + \frac{1}{n!} \sup_{x \in A} |f^{(n)}(t+x)| |E|(T(t))^n|, \quad (6.10)$$

where $A=[-\varepsilon, \varepsilon]$ if the sampling time jitter is limited (i.e. exists such a number a number $\varepsilon > 0$, that for each sampling time t , $|T_n(t)| < \varepsilon$) or $A=\mathbb{R}$ if the STJ is not limited. If $f^{(n+2)} \in L^1(\mathbb{R}, \mathcal{L}, 1_1)$ for $k=0, 1, \dots, n+2$ then denoting by F the Fourier transform of the signal f , we can (6.10) rewrite as

$$|h(t)-f(t)| \leq \sum_{k=1}^{n-1} \frac{|f^{(k)}(t)|}{k!} |E((T(t))^k)| + \frac{|E|((T(t))^n)|}{2\pi n!} \cdot \int_{\mathbb{R}} |\omega|^n |F(\omega)| I_1(d\omega) \leq$$

$$\leq \frac{1}{2\pi} \sum_{k=1}^{n-1} \frac{E|(T(t))^k| \int_{\mathbb{R}} |\omega|^n |F(\omega)| L_1(d\omega)}{k!} + \frac{E|(T(t))^n|}{2\pi n!} \int_{\mathbb{R}} |\omega|^n |F(\omega)| L_1(d\omega). \quad (6.11)$$

In particular from the formula (6.10) for $f \in C^{(1)}(\mathbb{R})$ and $f \in C^{(2)}(\mathbb{R})$ we have

$$|h(t) - f(t)| \leq \sup_{x \in A} |f'(t+x)| \cdot E|T(t)| \quad (6.12)$$

$$|h(t) - f(t)| \leq |f'(t)| \cdot E|T(t)| + \frac{1}{2} \sup_{x \in A} |f''(t+x)| \cdot E|T(t)|^2. \quad (6.13)$$

If additionally the STJ is limited then two inequalities (6.14), (6.15) follow from the formulas (6.12) and (6.13)

$$|h(t) - f(t)| \leq \varepsilon \sup_{x \in [-\varepsilon, \varepsilon]} |f'(t+x)| \text{ for } f \in C^{(1)}(\mathbb{R}) \quad (6.14)$$

$$|h(t) - f(t)| \leq \varepsilon |f'(t)| + \frac{\varepsilon^2}{2} \sup_{x \in [-\varepsilon, \varepsilon]} |f''(t+x)| \text{ for } f \in C^{(2)}(\mathbb{R}). \quad (6.15)$$

If the probability distribution of the random variables $T(t)$ is for each $t \in \mathbb{R}$ symmetric (i.e. jitter is symmetric), then formulas (6.10), (6.11) and (6.13) become simpler because odd moments $E((T(t))^k)$ vanish. In particular for $f \in C^{(2)}(\mathbb{R})$ from (6.13) we obtain

$$|h(t) - f(t)| \leq \frac{D^2(T(t))}{2\pi} \int_{\mathbb{R}^+} \omega^2 |F(\omega)| \cdot L_1(d\omega). \quad (6.17)$$

If sampling time jitter is limited, the signal f is analytical on $\mathbb{R}$ (i.e. $f \in C^{(\infty)}(\mathbb{R})$) and $\sup_{k \in \mathbb{N}} |f^{(k)}(t)| = G(t) < +\infty$ then because $E((T(t))^k) \leq \varepsilon^k$ we obtain

$$|h(t) - f(t)| \leq \sum_{k=1}^{\infty} \frac{|f^{(k)}(t)|}{k!} \leq \sum_{k=1}^{\infty} \frac{G(t)}{k!} \varepsilon^k = G(t)(e^\varepsilon - 1). \quad (6.18)$$

If additionally the random variable $T(t)$ has a symmetric probability distribution then $E((T(t))^k) = 0$ for k odd and (6.18) can be rewritten as (6.19)

$$|h(t) - f(t)| \leq \sum_{k=1}^{\infty} \frac{f^{(2k)}(t)}{(2k)!} |E((T(t))^{2k})| \leq G(t)(\text{ch}(\varepsilon) - 1). \quad (6.19)$$

From assessments (6.10)–(6.19) we can easily obtain error assessments (of the signal f measurement) in the supremum norm. For example, if $f \in C^{(4)}(\mathbb{R})$, $f^{(k)} \in L^1(\mathbb{R}, \mathcal{L}, L_1)$ for $k=1, \dots, 4$ and $E(T(t)) = 0$ for each $t \in \mathbb{R}$ then for ε -limited sampling time jitter we obtain

$$\sup_{t \in \mathbb{R}} |h(t) - f(t)| \leq \frac{\varepsilon^2}{2} \sup_{t \in \mathbb{R}} |f''(t)| \leq \frac{\varepsilon^2}{2\pi} \int_{\mathbb{R}^+} \omega^2 |F(\omega)| L_1(d\omega). \quad (6.20)$$

In the sequel we give several examples of distortions introduced by averaging when sampling time jitter is present.

Example 1. If for each $t \in [a, b]$ sampling time jitter is limited i.e. $|T(t)| \leq \varepsilon$ and a probability distribution $P(t, \cdot)$ of the random variable $T(t)$ is symmetric related to zero or $E(T(t)) = 0$, then arithmetic averaging for arbitrary fixed $t \in [a, b]$ does not introduce errors in the case of linear signals, i.e. for signals defined by the formula $f: [a-\varepsilon, b+\varepsilon] \ni t \rightarrow At+B \in \mathbb{R}$, where $A, B \in \mathbb{R}$. Indeed, $E(A(T(t)+t)+B) = At+B$ for $t \in [a, b]$.

Example 2. Let $f(t) = A \sin(\omega_g t)$; where $A, \omega_g \in \mathbb{R}^+$ and assume that $P(t, \cdot)$, the probability distribution of the random variable $T(t)$ is symmetric. Hence $h(t) = E(A \cdot \sin(\omega_g(t+T(t)))) = A(\sin(\omega_g t) E(\cos(\omega_g T(t))) + \cos(\omega_g t) \cdot E(\sin(\omega_g T(t)))) = A \sin(\omega_g t) \cdot E(\cos(\omega_g T(t))) + \cos(\omega_g t) \cdot E(\sin(\omega_g T(t)))$, where $\varphi(t, \cdot): \mathbb{R} \ni \omega \rightarrow E(e^{j\omega T(t)}) \in \mathbb{C}$ is a characteristic function of the probability distribution $P(t, \cdot)$. Similarly for a cosinusoidal signal $f(t) = A \cos(\omega_g t)$, we have $h(t) = A \cdot \cos(\omega_g t) \cdot \operatorname{Re} \varphi(t, \omega_g)$. For each $t, \omega \in \mathbb{R}$, $|\operatorname{Re} \varphi(t, \omega)| \leq |\varphi(t, \omega)| \leq |\varphi(t, 0)| \leq 1$.

Therefore, if the probability distribution of the random variable $T(t)$ does not depend on t , then "symmetric jitter" modifies an amplitude of the averaged sinusoid and introduces a phase shift but the shape of h remains sinusoidal (in particular may be $h(t) \equiv 0$ or the phase shift equal to π may occur). In the case where the random variable $T(t)$ has for each $t \in \mathbb{R}$ the uniform probability distribution on the interval $[-\varepsilon, \varepsilon]$ we obtain

$$h(t) = A \sin(\omega_g t) E(\cos(\omega_g T(t))) = \frac{A}{2\varepsilon} \sin(\omega_g t) \int_{-\varepsilon}^{\varepsilon} \cos(\omega_g x) dx = f(t) \cdot Sa(\omega_g \varepsilon), \quad (6.21)$$

$$\text{where } Sa: \mathbb{R} \ni x \rightarrow Sa(x) = \begin{cases} \sin x / x & \text{for } x \in \mathbb{R} \setminus \{0\} \\ 1 & \text{for } x = 0 \end{cases}$$

In the case, when the random variable $T(t)$ have for each $t \in \mathbb{R}$ the Gaussian probability distribution $N(0, \sigma)$, we obtain $\varphi(t, \omega) = \exp(-\sigma^2 \omega_g^2 / 2)$ then

$$h(t) = A \cdot \sin(\omega_g t) \operatorname{Re} \varphi(t, \omega_g) = A \cdot \sin(\omega_g t) \cdot \exp(-\sigma^2 \omega_g^2 / 2).$$

If the probability distribution of the random variable $T(t)$ does not depends on t but is not symmetric then the jitter can change both amplitude and introduces a phase shift φ (h remains a sinusoid) according to the formula

$$E(A \cdot \sin(\omega_g(T(t)+t))) = A_1 \sin(\omega_g t) + A_2 \cos(\omega_g t) = \tilde{A} \cdot \sin(\omega_g t + \varphi)$$

where $A_1 = A \cdot E(\cos(\omega_g T(t)))$, $A_2 = A \cdot E(\sin(\omega_g T(t)))$, $\tilde{A}^2 = A_1^2 + A_2^2$, and φ is solution of equations $\sin \varphi = A_2 / \tilde{A}$, $\cos \varphi = A_1 / \tilde{A}$.

If the signal $f \in C^{(1)}(\mathbb{R})$ is periodic with a period $T = 2\pi/\omega_g$ then its Fourier series

$$\sum_{n=-\infty}^{\infty} c_n e^{j\omega_g n t}$$

is uniform convergent for $t \in \mathbb{R}$ to $f(t)$. Hence

$$h(t) = E(f(t+T(t))) = \sum_{n=-\infty}^{\infty} c_n E(e^{j\omega_n(t+T(t))}) = \sum_{n=-\infty}^{\infty} c_n e^{j\omega_n t} \cdot \varphi(t, \omega_n) \quad (6.22)$$

Example 3. Assume the random variable $T(t)$ have for every $t \in \mathbb{R}$ the same uniform probability distribution on $[-\varepsilon, \varepsilon]$ and the input signal f is the step function $f(t) = 1(t)$. Then we have

$$h(t) = E(1(t+T(t))) = \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} 1(t+x) dx = \begin{cases} 1 & \text{for } t \geq \varepsilon \\ 1/2 + \frac{1}{2\varepsilon} t & \text{for } -\varepsilon < t < \varepsilon \\ 0 & \text{for } t \leq -\varepsilon \end{cases} \quad (6.24)$$

If $T(t)$ have for each $t \in \mathbb{R}$ the Gaussian probability distribution $N(0, \sigma)$ then we obtain

$$h(t) = E(1(t+T(t))) = \frac{1}{\sigma\sqrt{2}} \int_{\mathbb{R}} 1(t+x) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\sqrt{2}}{2\sigma} t\right) \quad (6.25)$$

where $\operatorname{erf}(x) \stackrel{\text{df}}{=} \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$ is the error function.

For both probability distributions (uniform and Gaussian) obtained rise time of measured step functions is significantly augmented by the sampling time jitter (for example up to 1.6ε in the case of uniform probability distribution).

If we correct errors introduced by the STJ in ETWR systems the accuracy of input A/D converters should be higher than implied by inequalities (6.1)–(6.7) and the practical rule: "the variance of the quantization noise should be even to the variance of the noise caused by STJ". More information on STJ errors in sampling systems can be found in [11]–[18].

8. CONCLUSIONS

The first upcrossing time of a moving barrier by a stochastic process is a good, general way of STJ problem modelling. The presented FBM methods of P_r calculation are simple and working under not specially restrictive assumptions. These methods are also useful in noise properties and reliability assesment of many electronic circuits, in particular, of trigger and sampling circuits used in wide band digital oscilloscopes. It was proved that for sufficiently fast linear barrier the probability distribution P_r can be considered as Gaussian. Some inequalities were given to assess errors caused by the STJ. They give useful information for designers choosing accuracy of A/D converters applied in wide band data acquisition systems. If we have information about STJ probability distributions we can try remove or correct errors caused by the STJ via for example deconvolution or median filtering.

The following dependence holds. Theoretical description of the STJ $\rightarrow$ measurement identification or computation of STJ distribution $\rightarrow$ STJ errors assessment and correction.

LIST OF SYMBOLS

- N — set of natural numbers
 R — set of real numbers
 R^+ — set of nonnegative real numbers, $R^+ = \{x \in R; x \geq 0\}$
 $B(R)$ — σ -field of Borel subsets of R
 $B(R^+)$ — σ -field of Borel subsets of R^+
 C — set of complex numbers
 R^n — n -dimensional Euclidean space
 $(\Omega, \mathfrak{M}, P)$ — probabilistic space
 I_1 — Lebesgue measure on R
 $L^p(\Omega, \mathfrak{M}, P)$ — $p \geq 1$, set of all random variables having the p -th moment
 $L^p(R, \mathcal{L}, I_1)$ — $p \geq 1$, set of p -th power integrable real functions over R
 $E(X)$ — mean value of the random variable X
 $D^2(X)$ — variance of the random variable X
 $C(R^+)$ — set of continuous functions over R^+
 $C^{(n)}(R)$ — set of real functions having continuous n -th order derivatives
 $N^+(Br, I)$ — random variable, the number of the barrier Br upcrossings by a stochastic process $(X(t))_{t \in R^+}$ on the time interval $I \subset R^+$
 $N(m, \sigma)$ — normal distribution with the mean value m and the variance σ^2

REFERENCES

1. M.R. Leadbetter, G. Lindgren, H. Rootzen: *Extremes and Related Properties of Random Sequences and Processes*. Springer Verlag, New York, 1983
2. H. Cramer, M.R. Leadbetter: *Stationary and Related Stochastic Processes, Sample Function Properties and Their Applications*. John Wiley & Sons, New-York, 1967
3. P. Kree, Ch. Soize: *Mathematics of Random Phenomena, Random Vibrations of mechanical Structures*. P.Reidel, Boston 1986
4. I.F. Blake: W.C. Lindsey: *Level-crossing problems for random processes*. IEEE Trans. on Inf. Theory, vol. IT-19, No.3, May 1973
5. W.I. Tichonov, W.I. Chimienko: *Stochastic processes Level Crossings, (in Russian)*. Moscow, 1987
6. S.O. Rice: *Mathematical Analysis of Random Noise*. Bell System Technical Journal, vol. 23, 1944, vol. 24, 1945
7. G.L. Cariolaro: *Time Jitter Caused by Noise in Trigger Circuits*. IEEE Trans on Inf. Theory, vol. IT-14 No 4, July 1968
8. W.L. Gans: *The Measurement and Deconvolution of Time Jitter in Equivalent-Time Waveform Samplers*. IEEE Trans. on Instr. and Measurement, No.1, March 1983
9. R. Y u: M. C a s e: *275 GHz 3-mask Integrated GaAs Sampling Circuit*. Electronics Letters; vol.26 No 13, June 1990
10. K. Fehert: *Telecommunications Measurements, Analysis and Instrumentation*. Prentice Hall, 1987
11. J.J. O'Reilly: *Temporal Quantization Noise*. Electronics Letters; March 82, vol.18, No 5
12. A.V. Balakrishnan: *On the problem of time jitter in sampling*. IRE Trans Inf. Theory, April IT-8, 1962
13. J.B. Szuhbator: *Power Spectrum of Sampled Signals Distorted by Jitter*. Electronic Letters, vol. 19, No.19, 1983

14. G. Ver kroost: *Noise Caused by Sampling Time Jitter with Applications to Sampling Frequency Conversion*. Signal Processing, Theory and Applications, EURASIP 1986
15. T.M. Souders, D.R. Flach, C. Hagwood, G.L. Yang: *The Effects of Timing Jitter in Sampling Systems*. IEEE Trans. Instrum. Measur. vol.39, February 1990
16. M.F. W ag d y, S.A. A w a d: *Effect of Sampling Jitter on Some Sine Wave Measurements*. IEEE Trans. Instrum. Measur., vol.39, February 1990
17. T. A d a m s k i: *Methods of Sampling Time Jitter Identification in Equivalent Time Waveform Recorders*. Proceedings of the Polish–Czech–Hungarian Workshop on Circuit Theory and Applications, Budapest–God 1991
18. T. A d a m s k i: *Jitter Noise in Sampling Frequency Converters*. Proceeding of the Hungarian–Polish–Czech Workshop on Circuit Theory and Applications, Cernice 1989
19. T. A d a m s k i: *The Mean Number of moving Barrier Crossing for Nonstationary Stochastic Process*. Proceedings of Hungarian–Polish–Czech Workshop on Circuit Theory and Applications, Warsaw–Kiry 1988
20. T. A d a m s k i: *Simple Method for the First Passage Time Assessment*. Acta Politechnika, Prace CVUT v Praze, No 1, Prague 1990
21. T. A d a m s k i: *Jitter in Electronic Circuits and Systems*. Technical Report, Institute of Electronics Fundamentals, Warsaw Univ. of Technology, Warsaw 1991

T. ADAMSKI

ROZKŁAD PRAWDOPODOBIENSTWA CHWILOWEJ NIESTAŁOŚCI MOMENTU PRÓBKOWANIA W OSCYLOSKOPACH CYFROWYCH ZE STROBOSKOPOWYM SPÓSOMBEM PRÓBKOWANIA

Streszczenie

W szerokopasmowych oscyloskopach cyfrowych ze stroboskopowym sposobem próbkowania, chwilowa niestałość momentu próbkowania (tzw. sampling time jitter) powoduje znaczne zniekształcenia rejestrowanego przebiegu. Praca dotyczy sposobów obliczania rozkładu prawdopodobieństwa zmiennej losowej τ opisującej tę niestałość oraz podaje warunki, przy których rozkład τ można uważać za normalny. Znajomość rozkładu zmiennej losowej τ umożliwia odtworzenie metodą rozplatania rzeczywistego kształtu sygnału mierzonego.

Effective bits of analog to digital conversion and sampling time jitter

TOMASZ ADAMSKI

Instytut Podstaw Elektroniki, Politechnika Warszawska

Received 1992.10.20

Authorized 1992.12.10

Sampling time jitter limits significantly accuracy of wide band measurement systems with sampling. The paper deals with influence of sampling time jitter on effective bits when repetitive (equivalent time) sampling and averaging algorithm are applied. It is shown that in the presence of sampling time jitter the method of effective bits obtaining remains (with some modifications) correct. The worst case assessments of number of effective bits for band limited signals (as a function of a jitter variance and a number of averaged samples) are computed in the paper.

1. INTRODUCTION

In wide band measurement systems like digital oscilloscopes or digital waveform recorders we cannot neglect random unstability of sampling time. Because of noise phenomena present in sampling circuits real sampling points are randomly shifted on the axis. This shifting is called the sampling time jitter (STJ). The STJ introduces errors to measurements and can cause troublesome scattering of the input signal image.

In relatively slow systems the mentioned shifts are not critical and errors caused by the STJ can be neglected. Additionally, in such systems, when repetitive (equivalent time) sampling mode is used we can decrease quantization effects by adding dither to the input signal and consecutive averaging. The above procedure is well known and leads to the notion of effective bits.

In fast systems, like wide band equivalent-time waveform recorders and sampling oscilloscopes, random shifting is in the range 10—200 ps and significant distortions of the measured signal occur. We show in the sequel that the method of effective bits remains correct (with some modifications) also when the STJ is present.

2. REPETITIVE SAMPLING WITH TIME JITTER

Below we describe shortly the mathematical model of sampling with time jitter. Assume the analog input signal $f: \mathbb{R} \rightarrow \mathbb{R}$ is a periodic, $(B(\mathbb{R}), B(\mathbb{R}))$ measurable function sampled in equivalent time points $t_1, t_2, \dots, t_M \in [0, b]$, where $b > 0$ is a period of f (in particular we can have $t_k = k \cdot \Delta t$). Let $T(t_k)$ be the random variable describing the random shift of the real sampling point $t_k + T(t_k)$ related to assumed, deterministic one t_k . In equivalent time waveform recorders (ETWR) we repeat the sampling at each point t_k (where $k=1, 2, \dots, M$ and M is a number of sampling points) in independent manner. Therefore we can describe the all sampling process by M random sequences $(f(t_k + T_n(t_k)))_{n=1}^{\infty}$, for $k=1, 2, \dots, M$, where $(T_n(t_k))_{n=1}^{\infty}$ is a sequence of independent random variables with probability distributions equal to the distribution of $T(t_k)$. In practice, the number of samples for each k is limited to $N_0 \in \mathbb{N}$, then the input information of the ETWR consists of M random vectors of N_0 dimensions or in other words, of M finite random sequences $(f(t_k + T_n(t_k)))_{n=1}^{N_0}$ for $k=1, 2, \dots, M$.

To assess measurement errors introduced by the STJ we have first to assess the value $er(t) \triangleq f(t + T(t)) - f(t)$ for a fixed sampling point t particularly for $t_1, t_2, \dots, t_M \in [0, b]$. If the signal f is a $(B(\mathbb{R}), B(\mathbb{R}))$ measurable function (for instance f is continuous) then $er(t)$ is a random variable for each sampling point t but its probability distribution depends on f and is in most cases difficult to calculate exactly. Denote by $\mathcal{K}(A, \omega_g)$ a set of real band limited (by ω_g) functions over $\mathbb{R}$ with values in the interval $[-A, A]$. The worst case function (i.e. most sensitive to the STJ) from $\mathcal{K}(A, \omega_g)$ is a sinusoid $f(t) = A \cdot \sin(\omega_g t)$. As a result we have for $f \in \mathcal{K}(A, \omega_g)$

$$D^2(er(t)) = D^2(f(t + T(t))) \leq A^2 \omega_g^2 D^2(T(t)). \quad (2.1)$$

The above error assessment deals with arbitrary sampling system where the STJ is significant. Assume $(Q(t))_{t \in \mathbb{R}}$ is a real, $\mathcal{F} \otimes B(\mathbb{R})$ measurable (where $\mathcal{F} = \sigma(Q(t); t \in \mathbb{R})$) stationary (in the strict and wide sense) stochastic process describing an additive noise corrupting the input deterministic signal f . If for each $k=1, 2, \dots, M$ the random sequence $(T_n(t_k))_{n=1}^{\infty}$ is the sequence of independent random variables with the same probability distribution and σ -fields $\sigma(T_n(t_k); n \in \mathbb{N})$, $\sigma(Q(t); t \in \mathbb{R})$ are independent then for each $k=1, 2, \dots, M$ a process $(Q_n(t_k))_{n=1}^{\infty}$ defined as $Q_n(t_k) = Q(nb + t_k + T_n(t_k))$ is well defined stationary one in the strict and wide sense (see theorem A1 from the appendix).

Let $(D_n)_{n=1}^{\infty}$ denotes a sequence of independent random variables with the uniform probability distribution on the interval $[-\Delta x/2, \Delta x/2]$, where Δx is a distance between (uniformly distributed) quantization levels. The sequence $(D_n)_{n=1}^{\infty}$ describes dither purposely added to the input signal. Finally we can describe for each $k=1, 2, \dots, M$ the input signal with the following random sequence.

$$(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n)_{n=1}^{\infty}. \quad (2.2)$$

We assume in the sequel that for each $k=1, 2, \dots, M$ σ -fields $\sigma(T_n(t_k); n \in \mathbb{N})$, $\sigma(Q(t); t \in \mathbb{R})$ and $\sigma(D_n; n \in \mathbb{N})$ are independent (i.e. dither, jitter and additive noise are

independent) but it does not imply that for fixed k , random variables $Q_n(t_k)$ and $f(t_k + T_n(t_k))$ are independent too. Note that for a bounded measurable function f , using the theorem A2 from the appendix we obtain

$$\begin{aligned} E(f(t_k + T_n(t_k))Q_n(t_k)) &= E(E(f(t_k + T_n(t_k))Q_n(t_k) | \sigma(T_n(t_k)))) = \\ &= E(E(f(t_k + x)Q(nb + x + t_k)) |_{x=T_n(t_k)}). \end{aligned}$$

Therefore, if $E(Q(t)) = 0$ for each $t \in \mathbb{R}$ then $E(E(f(t_k + x)Q(nb + x + t_k)) |_{x=T_n(t_k)}) = 0$ and finally

$$E(f(t_k + T_n(t_k))Q_n(t_k)) = 0. \quad (2.3)$$

3. SYSTEMATIC ERRORS OF AVERAGING ALGORITHM

Assume that analysed in the sequel sampling system is the ETWR. In this case for each $k=1, 2, \dots, M$ a sequence of samples $(f(t_k + T_n(t_k)) + Q_n(t_k))_{n=1}^{N_0}$ can be gathered creating input data. This data directly displayed on the screen gives scattered (we say jittered) image of the measured signal (even if the additive noise is not present). In general the statistics $(f(t_k + T_n(t_k)) + Q_n(t_k))_{n=1}^{N_0}$ taken for every $k=1, 2, \dots, M$ allow to correct distortions caused by jitter, but in principle the restoration needs deconvolution algorithm or in the case when jitter statistics depends on k , solution of the integral Fredholm equation of the first kind. Hence correction leads to the ill posed problem very sensitive to input data and needs complicated numerical regularization techniques (see [7], [8]). In this context, simple methods like mean value or quantile statistics computations (see [9]), working quite correctly only under certain assumptions become attractive. The simplest method to correct scattered signal image (mostly used by digital oscilloscopes producers) is the mentioned mean value computation i.e. the averaging method. In this method for each sampling point t the number

$$MN(t, N_0) \stackrel{\text{df}}{=} \frac{1}{N_0} \sum_{n=1}^{N_0} (f(t + T_n(t)) + Q_n(t_k)) \quad (3.1)$$

is computed. Let $h(t) \stackrel{\text{df}}{=} E(f(t + T_n(t)))$ and $g(\cdot, t)$ denotes the probability density function of the random variable $T(t)$ describing the STJ in the time point t . From the law of great numbers (see theorem A3 from the appendix) we have $MN(t, N_0) \rightarrow h(t)$ (where $N_0 \rightarrow \infty$) with probability 1. Hence for sufficiently large N_0 we admit

$$h(t) = \int_{\mathbb{R}} g(x, t) f(x) l_1(dx). \quad (3.2)$$

The function $t \rightarrow h(t)$ is obtained from measurement. Hence, when the STJ is present, arithmetic averaging of the set of samples (related to fixed equivalent time sampling moment) limits the band of the system. Below, we will assess measurement errors, when averaging is applied.

The averaging method is quite correct for linear signal $f(t) = At + B$, where $A, B \in \mathbb{R}$ i.e. $MN(t, N_0) \rightarrow f(t)$ with probability 1, but in general (for example for pulse signals) it introduces some systematic errors. These errors can be easily assessed in the following way.

We would like to find $|h(t) - f(t)|$ for fixed sampling time t and $\sup_{t \in \mathbb{R}} |h(t) - f(t)|$.

Assume that the input signal $f \in C^{(n)}(\mathbb{R})$ and for each $t \in \mathbb{R}$, $T(t) \in L^n(\Omega, \mathfrak{M}, P)$, $f(t + T(t)) \in L^1(\Omega, \mathfrak{M}, P)$. Using Taylor's formula we obtain

$$|h(t) - f(t)| \leq \sum_{k=1}^{n-1} \frac{|f^{(k)}(t)|}{k!} |E((T(t))^k)| + \frac{1}{n!} \left(\sup_{x \in A} |f^{(n)}(t+x)| E|(T(t))^n| \right), \quad (3.10)$$

where $A = [-\varepsilon, \varepsilon]$ if the sampling time jitter is limited (i.e. exists such a number $\varepsilon > 0$, that for each sampling time t , $|T_n(t)| < \varepsilon$) or $A = \mathbb{R}$ if the sampling time jitter is not limited.

If $f^{(n+2)} \in C(\mathbb{R})$, $f^{(k)} \in L^1(\mathbb{R}, \mathcal{L}, l_1)$ for $k=0, 1, \dots, n+2$ then denoting by F the Fourier transform of the signal f , we can (3.10) rewrite as

$$\begin{aligned} |h(t) - f(t)| &\leq \sum_{k=1}^{n-1} \frac{|f^{(k)}(t)|}{k!} |E((T(t))^k)| + \frac{E|((T(t))^n)|}{2\pi n!} \cdot \int_{\mathbb{R}} |\omega|^n |F(\omega)| l_1(d\omega) \leq \\ &\leq \frac{1}{2\pi} \sum_{k=1}^{n-1} \frac{|f^{(k)}(t)|}{k!} \int_{\mathbb{R}} |\omega|^n |F(\omega)| l_1(d\omega) + \frac{1}{2\pi} \frac{E|(T(t))^n|}{n!} \int_{\mathbb{R}} |\omega|^n |F(\omega)| l_1(d\omega) \end{aligned}$$

In particular from the formula (3.10) for $f \in C^{(1)}(\mathbb{R})$ and $f \in C^{(2)}(\mathbb{R})$ we have

$$|h(t) - f(t)| \leq \sup_{x \in A} |f'(t+x)| \cdot E|T(t)| \quad (3.12)$$

$$|h(t) - f(t)| \leq |f'(t)| E|T(t)| + \frac{1}{2} \sup_{x \in A} |f''(t+x)| \cdot E|T(t)|^2. \quad (3.13)$$

If additionally sampling time jitter is limited then two inequalities (3.14), (3.15) follow from the formulas (3.12) and (3.13)

$$|h(t) - f(t)| \leq \varepsilon \sup_{x \in [-\varepsilon, \varepsilon]} |f'(t+x)| \text{ for } f \in C^{(1)}(\mathbb{R}) \quad (3.14)$$

$$|h(t) - f(t)| \leq \varepsilon |f'(t)| + \frac{\varepsilon^2}{2} \sup_{x \in [-\varepsilon, \varepsilon]} |f''(t+x)| \text{ for } f \in C^{(2)}(\mathbb{R}) \quad (3.15)$$

If the probability distribution of the random variables $T(t)$ is for each $t \in \mathbb{R}$ symmetric (i.e. jitter is symmetric), then formulas (3.10), (3.11) and (3.13) become simpler because odd moments $E((T(t))^k)$ vanish. In particular for $f \in C^{(2)}(\mathbb{R})$ from (3.13) we obtain

$$|h(t) - f(t)| \leq \frac{1}{2} \sup_{x \in A} |f''(t+x)| \cdot D^2(T(t)). \quad (3.16)$$

If $f \in C^{(4)}(\mathbb{R})$ and $f^{(k)} \in L^1(\mathbb{R}, \mathcal{L}, l_1)$ for $k=1, 2, 3, 4$ then from (3.11) we obtain

$$|h(t) - f(t)| \leq \frac{D^2(T(t))}{2\pi} \int_{\mathbb{R}^+} \omega^2 |F(\omega)| l_1(d\omega). \quad (3.17)$$

If sampling time jitter is limited, the signal f is analytical on $\mathbb{R}$ (i.e. $f \in C^{(\infty)}(\mathbb{R})$) and $\sup_{k \in \mathbb{N}} |f^{(k)}(t)| = G(t) < +\infty$ then because $E((T(t))^k) \leq \varepsilon^k$ we obtain

$$|h(t) - f(t)| \leq \sum_{k=1}^{\infty} \frac{|f^{(k)}(t)|}{k!} \leq \sum_{k=1}^{\infty} \frac{G(t)}{k!} \varepsilon^k = G(t)(e^\varepsilon - 1). \quad (3.18)$$

If additionally the random variable $T(t)$ has a symmetric probability distribution then $E((T(t))^k) = 0$ for k odd and (3.18) can be rewritten as (3.19)

$$|h(t) - f(t)| \leq \sum_{k=1}^{\infty} \frac{|f^{(2k)}(t)|}{(2k)!} |E((T(t))^{2k})| \leq G(t)(\text{ch}(\varepsilon) - 1). \quad (3.19)$$

From assessments (6.10)–(6.19) we can easily obtain error assessments (of the signal f measurement) in the supremum norm. For example, if $f \in C^{(4)}(\mathbb{R})$, $f^{(k)} \in L^1(\mathbb{R}, \mathcal{L}, l_1)$ for $k=1, 2, 3, 4$ and $E(T(t))=0$ for each $t \in \mathbb{R}$ then for ε -limited sampling time jitter we obtain

$$\sup_{t \in \mathbb{R}} |h(t) - f(t)| \leq \frac{\varepsilon^2}{2} \sup_{t \in \mathbb{R}} |f''(t)| \leq \frac{\varepsilon^2}{2\pi} \int_{\mathbb{R}^+} \omega^2 |F(\omega)| l_1(d\omega). \quad (3.20)$$

In the sequel we give several examples of distortions introduced by averaging when sampling time jitter is present.

Example 1. If for each $t \in [a, b]$ sampling time jitter is limited i.e. $|T(t)| \leq \varepsilon$ and a probability distribution $P(t, \cdot)$ of the random variable $T(t)$ is symmetric related to zero or $E(T(t))=0$, then arithmetic averaging for arbitrary fixed $t \in [a, b]$ does not introduce errors in the case of linear signals, i.e. for signals defined by the formula $f: [a-\varepsilon, b+\varepsilon] \ni t \rightarrow At+B \in \mathbb{R}$, where $A, B \in \mathbb{R}$. Indeed, $E(A(T(t)+t)+B) = At+B$ for $t \in [a, b]$.

Example 2. Let $f(t) = A \sin(\omega_g t)$; where $A, \omega_g \in \mathbb{R}^+$ and assume that $P(t, \cdot)$ the probability distribution of the random variable $T(t)$ is symmetric. Hence $h(t) = E(A \cdot \sin(\omega_g(t+T(t)))) = A(\sin(\omega_g t) \cdot E(\cos(\omega_g T(t))) + \cos(\omega_g t) \cdot E(\sin(\omega_g T(t)))) = A \sin(\omega_g t) \cdot E(\cos(\omega_g T(t))) = A \cdot \sin(\omega_g t) \cdot \text{Re } \varphi(t, \omega_g)$, where $\varphi(t, \cdot): \mathbb{R} \ni \omega \rightarrow E(e^{j\omega T(t)}) \in \mathbb{C}$ is a characteristic function of the probability distribution $P(t, \cdot)$. Similarly for a consinusoidal signal $f(t) = A \cos(\omega_g t)$, we have $h(t) = A \cdot \cos(\omega_g t) \cdot \text{Re } \varphi(t, \omega_g)$. For each $t, \omega \in \mathbb{R}$, $|\text{Re } \varphi(t, \omega)| \leq |\varphi(t, \omega)| \leq |\varphi(t, 0)| \leq 1$. Therefore, if probability distribution of the random variable $T(t)$ does not depend on t , then "symmetric jitter" modifies an amplitude of the averaged sinusoid and introduces a phase shift but the shape of h remains sinusoidal (in particular may be

$h(t) \equiv 0$ or the phase shift equal to π may occur). In the case where the random variable $T(t)$ has for each $t \in \mathbb{R}$ the uniform probability distribution on the interval $[-\varepsilon, \varepsilon]$ we obtain

$$h(t) = A \sin(\omega_g t) E(\cos(\omega_g T(t))) = \frac{A}{2\varepsilon} \sin(\omega_g t) \int_{-\varepsilon}^{\varepsilon} \cos(\omega_g x) dx = f(t) \cdot \text{Sa}(\omega_g \varepsilon), \quad (3.21)$$

$$\text{where } \text{Sa} : \mathbb{R} \ni x \rightarrow \text{Sa}(x) = \begin{cases} \sin x / x & \text{for } x \in \mathbb{R} \setminus \{0\} \\ 1 & \text{for } x = 0. \end{cases}$$

In the case, when the random variable $T(t)$ have for each $t \in \mathbb{R}$ the Gaussian probability distribution $N(0, \sigma)$, we obtain $\varphi(t, \omega) = \exp(-\sigma^2 \omega_g^2 / 2)$ then

$$h(t) = A \cdot \sin(\omega_g t) \text{Re } \varphi(t, \omega_g) = A \cdot \sin(\omega_g t) \cdot \exp(-\sigma^2 \omega_g^2 / 2).$$

If the probability distribution of the random variable $T(t)$ does not depends on t but is not symmetric then the jitter can change both amplitude and phase and introduces a phase shift φ (h remains a sinusoidal) according to the formula

$$E(A \cdot \sin(\omega_g(T(t) + t))) = A_1 \sin(\omega_g t) + A_2 \cos(\omega_g t) = \tilde{A} \cdot \sin(\omega_g t + \varphi)$$

where $A_1 = A \cdot E(\cos(\omega_g T(t)))$, $A_2 = A \cdot E(\sin(\omega_g T(t)))$, $\tilde{A}^2 = A_1^2 + A_2^2$, and φ is a solution of equations $\sin \varphi = A_2 / \tilde{A}$, $\cos \varphi = A_1 / \tilde{A}$.

If the signal $f \in C^{(1)}(\mathbb{R})$ is periodic with a period $T = 2\pi / \omega_g$ then its Fourier

series $\sum_{n=-\infty}^{\infty} c_n e^{j\omega_g n t}$ is uniform convergent for $t \in \mathbb{R}$ to $f(t)$. Hence

$$h(t) = E(f(t + T(t))) = \sum_{n=-\infty}^{\infty} c_n E(e^{j\omega_g n(t + T(t))}) = \sum_{n=-\infty}^{\infty} c_n e^{j\omega_g n t} \cdot \varphi(t, \omega_g n) \quad (3.22)$$

Example 3. Assume the random variable $T(t)$ have for each $t \in \mathbb{R}$ the same uniform probability distribution on $[-\varepsilon, \varepsilon]$ and the input signal f is the step function $f(t) = 1(t)$. Then we have

$$h(t) = E(1(t + T(t))) = \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} 1(t + x) dx = \begin{cases} 1 & \text{for } t \geq \varepsilon \\ 1/2 + \frac{1}{2\varepsilon} t & \text{for } -\varepsilon < t < \varepsilon \\ 0 & \text{for } t \leq -\varepsilon. \end{cases} \quad (3.24)$$

If $T(t)$ have for each $t \in \mathbb{R}$ the Gaussian probability distribution $N(0, \sigma)$ then we obtain

$$h(t) = E(1(t + T(t))) = \frac{1}{\sigma\sqrt{2}} \int_{\mathbb{R}} 1(t + x) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \frac{1}{2} + \frac{1}{2} \text{erf}\left(\frac{\sqrt{2}}{2\sigma} t\right), \quad (3.25)$$

where $\text{erf}(x) \stackrel{\text{df}}{=} \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$ is the error function.

For both probability distributions (uniform and Gaussian) obtained rise time of the measured step function is significantly augmented by the STJ (for example up to $1.6 \cdot \varepsilon$ in the case of the uniform probability distribution). More information on STJ errors in sampling systems can be found in [1]–[8].

4. AVERAGING OF QUANTIZED SIGNALS IN JITTERED AND NOISY ENVIRONMENT

Let $\Delta x > 0$ be a distance between uniform distributed quantization levels and $QN: \mathbb{R} \rightarrow L$ (where $L = \{k \cdot \Delta x; k \in \mathbb{Z}\}$) denotes a quantization function defined by the formula: $(QN(x) = k \cdot \Delta x) \equiv (k \cdot \Delta x - \Delta x/2 < x \leq k \cdot \Delta x + \Delta x/2)$. Of course we have $\sup_{x \in \mathbb{R}} \text{dist}(x, L) = \Delta x/2$.

If a real stochastic process $(X(t))_{t \in \mathbb{R}}$ is stationary (in the strict sense) then a stochastic process $(QN(X(t)))_{t \in \mathbb{R}}$ is stationary too.

Fact. Assume: 1) for each $k=1,2,\dots, M$ a sequence of real, independent, random variables $(T_n(t_k))_{n=1}^{\infty}$ with the same probability distribution describes the STJ in the sampling point t_k .

2) $(D_n)_{n=1}^{\infty}$ is sequence of independent real random variables (describing dither) with the uniform probability distribution on the interval $[-\Delta x, \Delta x]$

3) the input signal $f: \mathbb{R} \rightarrow \mathbb{R}$ is $(B(\mathbb{R}), B(\mathbb{R}))$ measurable, periodic (with period b) function and for each $k=1, 2, \dots, M$ and $n \in \mathbb{N}$ $f(t_k + T_n(t_k)) \in L^1(\Omega, \mathfrak{M}, P)$.

4) $(Q(t))_{t \in \mathbb{R}}$ is a real, measurable, stationary (in the wide and strict sense) stochastic process (describing additive noise)

If for each $k=1,2,\dots, M$ σ -fields $\sigma(T_n(t_k); n \in \mathbb{N})$, $\sigma(Q(t); t \in \mathbb{R})$ and $\sigma(D_n; n \in \mathbb{N})$ are independent (i.e. dither, jitter and additive noise are independent) and $Q_n(t_k) \stackrel{\text{df}}{=} Q(bn + t_k + T(t_k))$ then for each sampling point t_k we have P almost everywhere

$$\left(\sum_{n=1}^{N_0} QN(f(t_k) + Q_n(t_k) + D_n) \right) / N_0 \rightarrow E(f(t_k + T_n(t_k))), \quad (4.1)$$

where $N_0 \rightarrow +\infty$.

Proof. The stochastic process $(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n))_{n=1}^{\infty}$ is stationary (in the strict sense). Hence the "quantized process" $(QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n))_{n=1}^{\infty}$ is stationary too. Using theorem A3 from the appendix we obtain

$$\left(\sum_{n=1}^{N_0} QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n) \right) / N_0 \rightarrow E(QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n)). \quad (4.2)$$

From the theorem A2 we have $E(QN(f(t_k) + Q_n(t_k) + D_n)) =$

$$\begin{aligned} &= E(E(QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n) | \sigma(T_n(t_k), Q_n(t_k)))) = \\ &= E(E(QN(f(t_k + y_1) + y_2 + D_n)) |_{(y_1, y_2) = (T_n(t_k), Q_n(t_k))}), \end{aligned} \quad (4.3)$$

but for each $a \in \mathbb{R}$ we have $E(QN(a + D_n)) = a$ (which is the essence of the effective bits method). Additionally from the theorem A4 from the appendix we have $E(Q_n(t_k)) = E(Q(t_k + T_n(t_k))) = 0$. Therefore we can rewrite the above equation in the form

$$\begin{aligned} E(E(QN(f(t_k + y_1) + y_2 + D_n)) |_{(y_1, y_2) = (T_n(t_k), Q_n(t_k))}) &= \\ &= E(E(f(t_k + y_1) + y_2) |_{(y_1, y_2) = (T_n(t_k), Q_n(t_k))}) = \\ &= E(f(t_k + T_n(t_k)) + Q_n(t_k)) = E(f(t_k + T_n(t_k))) \text{ Q.E.D.} \end{aligned} \quad (4.4)$$

The above theorem states theoretical basis for use the effective bits method when the STJ is present.

5. VARIANCE ASSESSMENT OF THE SUM OF QUANTIZED SAMPLES

We would like to obtain the number of effective bits as a function of N_0 -number of averaged samples, α_k — standard deviation of the STJ, β_k — standard deviation of the additive noise, B — number of bits of the input A/D converter and provided that the input signal $f \in \mathcal{X}(A, f_g)$. In order to do it we have to compute or assess the variance (5.1):

$$D^2 \left(\sum_{n=1}^{N_0} QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n) \right) / N_0. \quad (5.1)$$

Additionally the variance, as a function of N_0 gives some information on the convergence speed. Exact computation of (5.1) is difficult because the variance (5.1) depends on the input signal f and the quantization function QN . On the other hand assessment of (5.1) (independent of f) can be done in the following simple way.

If $X \in L^2(\Omega, \mathfrak{M}, P)$ then $QN(X) \in L^2(\Omega, \mathfrak{M}, P)$ and

$$(\Delta x)^2/4 + E(X)^2 - \Delta x \cdot E(|X|) \leq E((QN(X))^2) \leq (\Delta x)^2/4 + E(X)^2 + \Delta x \cdot E(|X|). \quad (5.2)$$

It follows immediately from the inequality: $E((X + \Delta x/2)^2 \cdot \chi_{\{X \geq 0\}}) +$

$$+ E((X - \Delta x/2)^2 \cdot \chi_{\{X \leq 0\}}) \leq E((QN(X))^2) \leq E((X + \Delta x/2)^2 \cdot \chi_{\{X \geq 0\}}) +$$

$+ E((X - \Delta x/2)^2 \cdot \chi_{\{X \leq 0\}})$. If $X \in L^1(\Omega, \mathfrak{M}, P)$ then $X - \Delta x/2 \leq QN(X) \leq X + \Delta x/2$ and we have

$$E(X - \Delta x/2) \leq E(QN(X)) \leq E(X + \Delta x/2). \quad (5.3)$$

From (5.2) and (5.3) we obtain

$$\begin{aligned} D^2(X) - \Delta x(E(|X|) + E(X)) &\leq D^2(QN(X)) = E((QN(X))^2) - (E(QN(X)))^2 \leq \\ &\leq D^2(X) + \Delta x(E(|X|) + E(X)) \end{aligned} \quad (5.4)$$

Assume for each $t_k, k=1, 2, \dots, M$ the random sequence $(Q_n(t_k))_{n=1}^\infty$ is the sequence of independent random variables (it follows for example from assumptions: $|T_n(t_k)| < b/4$ for $n \in \mathbb{N}$, the process $(Q(t))_{t \in \mathbb{R}}$ is Gaussian and its correlation function K fulfills condition: $\text{supp } K \subseteq (b/4, b/4)$). Therefore random sequences

$(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n)_{n=1}^{\infty}$ and $(QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n))_{n=1}^{\infty}$ are the sequences of independent random variables. Using (5.4) the variance (5.1) can be assessed as follows:

$$\begin{aligned} D^2\left(\sum_{n=1}^{N_0} QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n)/N_0\right) &= \\ &= \sum_{n=1}^{N_0} D^2(QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n))/N_0^2. \end{aligned} \quad (5.5)$$

Assume $f \in \mathcal{K}(A, \omega_g)$ then from (5.4), (2.1), (4.3), (4.4), the equation $E(QN(a + D_n)) = a$ for each $a \in \mathbb{R}$ and the equation $E(f(t_k + T_n(t_k)) \cdot Q_n(t_k)) = 0$ we have

$$\begin{aligned} D^2(QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n)) &= E((QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n))^2) - \\ &- (E(QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n)))^2 \leq (\Delta x)^2/4 + E((f(t_k + T_n(t_k)) + Q_n(t_k) + \\ &+ D_n)^2) + \Delta x \cdot E(|f(t_k + T_n(t_k)) + Q_n(t_k) + D_n|) - (E(f(t_k + T_n(t_k))))^2 \leq (\Delta x)^2/4 + \\ &+ E((f(t_k + T_n(t_k)) + Q_n(t_k))^2) + E((D_n)^2) + \Delta x \cdot (A + E(|Q_n(t_k)|) + E(|D_n|)) - \\ &- (E(f(t_k + T_n(t_k))))^2 \leq (\Delta x)^2/4 + (\Delta x)^2/12 + \Delta x \cdot (A + E(|Q_n(t_k)|) + \Delta x/4) + \\ &+ D^2(f(t_k + T_n(t_k))) + E(Q_n(t_k))^2 \leq A^2 \omega_g^2 D^2(T_n(t_k)) + E((Q_n(t_k))^2) + \\ &+ \Delta x \cdot (A + E(|Q_n(t_k)|)) + 7(\Delta x)^2/12 \end{aligned}$$

If we denote $E(|Q_n(t_k)|) = \gamma_k$, $D^2(T_n(t_k)) = \alpha_k^2$ and $D^2(Q_n(t_k)) = \beta_k^2$ then $D^2(QN(f(t_k) + Q_n(t_k))) \leq A^2 \omega_g^2 \alpha_k^2 + \beta_k^2 + \Delta x \cdot A + \Delta x \cdot A + \Delta x \cdot \gamma_k + 7(\Delta x)^2/12$ and $D^2\left(\sum_{n=1}^{N_0} QN(f(t_k + T_n(t_k)) + Q_n(t_k) + D_n)/N_0\right) \leq (A^2 \omega_g^2 \alpha_k^2 + \beta_k^2 + \Delta x \cdot (A + \gamma_k) + 7(\Delta x)^2/12) N_0$.

If the input signal f is not corrupted by the additive noise $(Q(t))_{t \in \mathbb{R}}$ then

$$D^2\left(\sum_{n=1}^{N_0} QN(f(t_k + T_n(t_k)) + D_n)/N_0\right) \leq (A^2 \omega_g^2 \alpha_k^2 + \Delta x \cdot A + 7(\Delta x)^2/12)/N_0.$$

Because $\Delta x = 2A/2^B$, where B is a number of bits of the input A/D converter, we obtain

$$\begin{aligned} D^2\left(\sum_{n=1}^{N_0} (QN(f(t_k + T_n(t_k)) + D_n)/N_0)\right) &\leq (7(2A/2^B)^2/12 + A^2(2\pi \cdot f_g)^2 \alpha_k^2 + 2A^2/2^B)/N_0 \stackrel{\text{df}}{=} \\ &\stackrel{\text{df}}{=} F(A, N_0, f_g, \alpha_k) \end{aligned}$$

6. EFFECTIVE BITS IN JITTERED ENVIRONMENT

Assume the input signal $f \in K(A, \omega_g)$ (where $\omega_g = 2\pi f_g$) is not corrupted by the additive noise. Hence a low bound of the number of effective bits $B_e(B, N_0, \alpha_k, f_g)$ is given by the formula

$$(B_e(B, N_0, \alpha_k, f_g) = d) \equiv (F(A, N_0, f_g, \alpha_k) = (2A/2^d)^2/12). \quad (6.1)$$

Therefore d can be computed from the equation

$$(7(2A/2^B)^2/12 + A^2(2\pi \cdot f_g)^2 \alpha_k^2 + 2A^2/2^B)/N_0 = (2A/2^d)^2/12 \quad (6.2)$$

which finally gives

$$d = -(\log_2((7(1/2^B)^2 + 12\pi^2 f_g^2 \alpha_k^2 + 6/2^B))/2 + \log_2(N_0)/2). \quad (6.3)$$

Similar results can be obtained when the additive noise is present. Low bounds of effective bits $B_e(B, N_0, \alpha_k, f_g)$ as a function of the input signal band f_g (appropriately for $N_0=16$, $N_0=512$ and $N_0=4096$) are shown in the Fig.1.

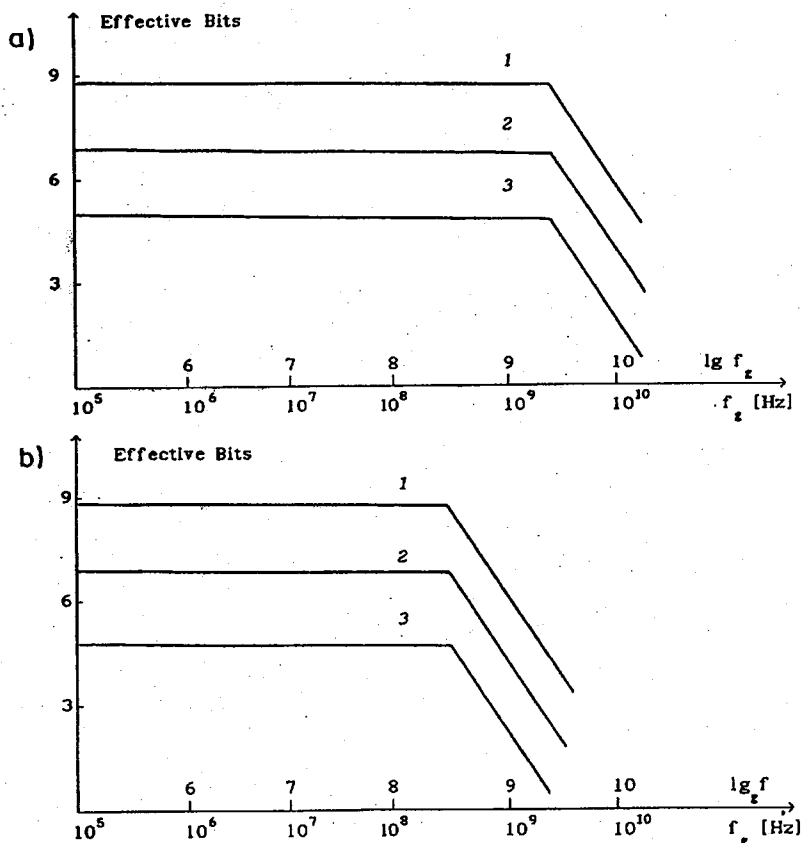


Fig.1. Bode plots of effective bits lower bounds. Effective bits lower bounds are shown as a function of the input signal band f_g , for $N_0 = 4096$ plot 1), for $N_0 = 512$ plot 2) and for $N_0 = 16$ plot 3). In the fig. a) the standard deviation of the sampling time jitter is even $\alpha_k = (D^2(T(t_k)))^{1/2} = 10$ ps, in the fig. b) $\alpha_k = (D^2(T(t_k)))^{1/2} = 100$ ps, in all cases $B = 8$.

CONCLUSION

In the presence of the sampling time jitter the method of effective bits is correct however the limit of the averaged sum is even to $E(f(t_k + T(t_k)))$ (instead of $f(t_k)$). Dither (uniformly distributed on the interval $[-\Delta x/2, \Delta x/2]$) introduced purposely on the input cannot be replaced by the noise caused by the STJ.

APPENDIX

Theorem A.1 Assume $(Q(t))_{t \in \mathbb{R}}$ is a real, stationary in the strict sense, $\mathcal{F} \otimes B(\mathbb{R})$ measurable stochastic process (where $\mathcal{F} = \sigma(Q(t); t \in \mathbb{R})$). If a real, stationary in the strict sense stochastic process $(T(t))_{t \in \mathbb{R}}$ describing sampling time jitter) and the stochastic process $(Q(t))_{t \in \mathbb{R}}$ are independent then the stochastic process $(Q(t + T(t)))_{t \in \mathbb{R}}$ is stationary in the strict sense. If additionally the stochastic process $(Q(t))_{t \in \mathbb{R}}$ is stationary in the wide sense then the stochastic process $(Q(t + T(t)))_{t \in \mathbb{R}}$ is stationary in the wide sense too.

Theorem A.2 Let $\xi: \Omega \rightarrow X$, $\eta: \Omega \rightarrow Y$ be random variables (defined on a probabilistic space $(\Omega, \mathfrak{M}, P)$) with values in measurable spaces respectively $(X, \mathcal{B}_1)$ and $(Y, \mathcal{B}_2)$. If a function $h: X \times Y \rightarrow \mathbb{R}$ is $(\mathcal{B}_1 \otimes \mathcal{B}_2, B(\mathbb{R}))$ measurable and $h(\xi, \eta) \in L^1(\Omega, \mathfrak{M}, P)$ then for each σ -field $\mathcal{F}$ such that $\mathcal{F} \subset \mathfrak{M}$, $\sigma(\eta) \subset \mathcal{F}$ we have

- 1) $E(h(\xi, \eta) | \mathcal{F}) = E(h(\xi, y) | \mathcal{F})|_{y=\eta}$.
- 2) If additionally σ -fields $\sigma(\xi)$ and $\mathcal{F}$ are independent then $E(h(\xi, \eta) | \mathcal{F}) = E(h(\xi, y) |_{y=\eta})$.

Theorem A.3 (Strong law of great numbers for strictly stationary stochastic processes). If $(X(n))_{n \in \mathbb{Z}}$ is a real strictly stationary stochastic process defined on a probabilistic space $(\Omega, \mathfrak{M}, P)$ and $X(n) \in L^1(\Omega, \mathfrak{M}, P)$ for each $n \in \mathbb{Z}$ then P almost everywhere

$$\left(\sum_{i=1}^{N_0} X(n) \right) / N_0 \xrightarrow{N_0 \rightarrow \infty} E(X(0)) \quad \text{and} \quad \sum_{i=N_0}^{N_0} X(n) / (2N_0 + 1) \xrightarrow{N_0 \rightarrow \infty} E(X(0)). \quad (\text{A.3})$$

Theorem A.4 Assume $(Q(t))_{t \in \mathbb{R}}$ is a real, second order, $\mathcal{F} \otimes B(\mathbb{R})$ measurable stochastic process (where $\mathcal{F} = \sigma(Q(t); t \in \mathbb{R})$). Denote $m(t) = E(Q(t))$. If a real stochastic process $(T(t))_{t \in \mathbb{R}}$ (describing sampling time jitter) and $(Q(t))_{t \in \mathbb{R}}$ are independent and for each $t \in \mathbb{R}$, $m(t + T(t)) \in L^1(\Omega, \mathfrak{M}, P)$ then for each $t \in \mathbb{R}$, $E(Q(t + T(t))) = E(m(t + T(t)))$. In particular, if $m(t) \equiv 0$ then $E(Q(t + T(t))) = 0$.

LIST OF SYMBOLS

$(\Omega, \mathfrak{M}, P)$ — probabilistic space

$\text{dist}(A, x)$ — distance between a set A and a point x in a metric space

N — set of natural numbers

- Z — set of integers
 Q — set of rational numbers
 R — set of real numbers
 R^+ — set of nonnegative real numbers, $R^+ = \{x \in R; x \geq 0\}$
 C — set of complex numbers
 $B(X)$ — σ -field of Borel sets of a metric space X
 I_1 — Lebesgue measure on R
 $L^p(\Omega, \mathcal{M}, P)$ — $p \geq 1$, set of all random variables having the p -th moment
 $L^p(R, \mathcal{L}, I_1)$ — $p \geq 1$, set of p -th power interable real functions over R
 $D^2(X)$ — variance of the random variable X
 $C(R^+)$ — set of continuous functions over R^+
 $C^{(n)}(R)$ — set of real functions having continuous n -th order derivatives

REFERENCES

1. L. E. L a r s o n: *High-Speed Analog to Digital Conversion with GaAs Technology*. Prospects, Trends and Obstacles; ISCAS 1988
2. B. L i u: *Effect of Timing Jitter on Performance of MTI Filters*. IEEE Transactions on Aerospace and Electronic Systems, may 1989
3. M. S h i n a g a w a, Y. A k a z a w a, T. W a k i m o t o: *Jitter Analysis of High Speed Sampling Systems*. IEEE Journal of Solid-State Circuits, Febr. 1990
4. T. M. S o u l d e r s: *The Effects of Timing Jitter in Sampling Systems*. IEEE Transaction on Instrumentation and Measurement, February 1990
5. M. F. W a g d y, S. S. A w a d: *Effect of sampling Jitter on Some Wave Measurement*. IEEE Transactions on Instrumentation and Measurement, February 1990
6. G. V e r k r o o s t: *Noise by Sampling Time Jitter with Applications to Sampling Frequency Conversion*. Signal Processing, Theory and Applications, EURASIP 1986
7. W. G a n s: *The Measurements and Deconvolution of Time Jitter in Equivalent Time Waveform Samplers*. IEEE Proceedings; March 1983
8. C. G r o e t s c h: *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman 1984
9. T. A d a m s k i: *Sampling Time Jitter Correction based on Quantile Filtering*. Measurement, elsevier, 1993, No 11

T. ADAMSKI

BITY EFEKTYWNE KONWERSJI A/C W OBECNOŚCI CHWILOWEJ NIESTAŁOŚCI MOMENTU PRÓBKOWANIA

Streszczenie

Chwilowa niestałość momentu próbkowania (sampling time jitter) znacznie ogranicza dokładność szerokopasmowych systemów próbkujących. W sytuacji gdy stosowane jest próbkowanie powtarzalne można dla podniesienia dokładności próbkowania w takich sytatemach zastosować metodę bitów efektywnych a ściślej zmodyfikowaną jej wersję opisaną w artykule. W pracy zostało podane również ograniczenie od dołu na liczbę bitów efektywnych w funkcji pasma sygnału wejściowego przy założeniu, że parametrami są liczba uśrednianych próbek i wariancja jitteru.

681.3.05
621.3.062

The basic criteria for reciprocal accomodation of the main elements of the analog-to-digital conversion system in PC-measurement instruments

ANNA DOMAŃSKA

Instytut Elektroniki i Telekomunikacji, Politechnika Poznańska

Received 1992.10.10

Authorized 1993.02.15

The analog-to-digital conversion system is the basic element of the PC-measurement instrument hardware. The issue of reciprocal accomodation of the elements of this system is examined, namely, the accomodation of the antialiasing filter, the sample-hold amplifier and the analog-to-digital converter. The criteria are the fixed level of the permissible error and the dynamics of the processed signal.

1. INTRODUCTION

PC-measurement instruments (PCMI) have been called third-generation measuring devices. The three basic elements of these devices are: PC, input hardware and corresponding software.

PC controls the data acquisition process, processes the acquired data according to required algorithms, presents the results and compiles the documentation.

The input hardware, called the data acquisition system (DAS), is composed of sensors, input conditioning systems, analog-to-digital conversion system (ADCS) and multiplexers (MUX). The input signal conditioning system accomodates the signal from the sensors or measurement converters to the ADCS. Accomodation may consist in amplification, compression, change of standard, filtering, ensuring the galvanic separation between the examined object and the PCMI etc. The ADCS changes the analog signal to the digital one. This process consists of the following operations: prefiltration (antialiasing filtration), sampling, quantifying. The anti-aliasing filter (AAF), the sample-hold amplifier (SHA), the analog-to-digital converter (ADC) are included in the conversion system.

The properties of the PCMI hardware are characterized by the parameters of the ADCS. The shaping of these parameters result from the previous establishing of time-accuracy conditions of data acquisition. These conditions are the starting point

for solving the accommodation problems, i.e. accommodation of ADCS to the input signal and accommodation of elements within the system. The parameters of ADCS are at the same time the criteria, according to which the properties of the remaining hardware are shaped.

The first step is to establish the required resolution (the bit format of data) and the speed of processing (the conversion time, the conversion frequency). The type of ADC is selected on this basis. The second step is to establish the structure of ADCS and to determine the parameters of the remaining elements (AAF, SHA) on the basis of the ADC parameters as well as the properties of the signal. Further considerations refer to the problems solved in the second step.

From the point of view of metrology it is essential not only for the device to perform the given measurement, but also for the obtained results to be reliable. This paper has presented the conditions for obtaining reliable measurement results. The conditions result from the cooperation of the elements in the ADCS. The employment of the elements with "nominally good" parameters is the necessary condition for obtaining suitably reliable measurement results. However, the nominal values of parameters are not the sufficient basis for formulating a quantitative evaluation of result if the conditions for reciprocal accommodation of elements in the ADCS and the conditions of accommodation of the ADCS and the output signal are not fulfilled.

From the point of view of the "measuring potential", the capabilities of the PCMI are huge. However, from the point of view of quality criteria, these capabilities are limited by resolution, data acquisition speed and the power of being independent of the influence of disturbances.

2. THE PLACE OF ADCS IN THE PCMI STRUCTURE

Originally, the structure of PCMI was as shown in Fig. 1. This configuration is used when the required resolution is not greater than 12 bits, or the dynamic range is less than 70 dB. In devices where the required resolution is higher, plug-in boards have been replaced by measurement modules, which are placed outside the computer and communicate with it through an interface. This structure is shown in Fig. 2.

As a result of this modification, the interference noise introduced by the computer itself (i.e. existing within the framework of the computer) has been

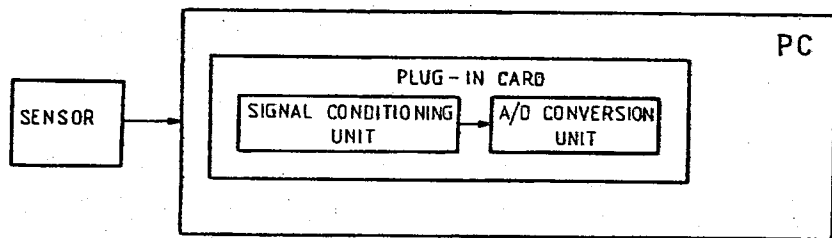


Fig. 1. The block scheme of the initial structure of PCMI

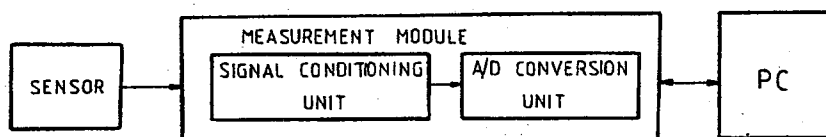


Fig. 2. The block scheme of the structure of PCMI with greater resolution

eliminated. It was possible to achieve better parameters of measurement processing. The analog signal circuits have been separated from the fast digital circuits of the computer which emit the noise and limit the dynamic range. The ADCS has been separated from the sources of noise, whose level exceeds the value of voltage corresponding to LSB (least significant bit) in the case of higher resolution processing.

3. THE SHAPING OF PARAMETERS OF THE PREFILTRATION CIRCUIT

The basic task of prefiltration is to prevent the distortion of the measurement signal, which might result from the sampling operation. Since the filter collaborates with the rest of elements in the ADCS, the quantities which influence the shaping of filter characteristic are the sampling rate and resolution of ADC.

The frequency of sampling of analog signal, which enables it to be reconstructed, results from the sampling theorem. If the boundary frequency corresponding to the highest frequency component of the signal is denoted f_{B0} , then the sampling rate f_s must fulfil the relationship $f_s \geq 2f_{B0}$. Therefore $2f_{B0} = f_{s0}$ defines the lowest possible value of f_s .

The spectrum of the signal sampled with the rate f_s (assuming that the sampling is uniform) is a periodical function and consist of replica of the spectrum of the analog signal (before the sampling has taken place) displaced by $k \cdot f_s$, $k = 0, \pm 1, \pm 2, \dots$. This can be expressed as follows:

$$X_s(\omega) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} X(\omega - k \cdot \omega_s), \quad T_s = \frac{1}{f_s}. \quad (1)$$

If the signal is sampled with the sampling rate $f_s \neq f_{s0}$, then the spectrum of signal after sampling depends on the relation between f_s and $2f_{B0}$. This relation is shown in Fig.3. The case shown in Fig.3c is valid when the sampling rate f_s is not fixed according to the sampling theorem. In this case in the neighbourhood of the frequency $\pm(2k+1) \cdot \frac{1}{2} f_s$ there appears the folding of the replica of the spectrum (aliasing). The spectrum distortions originating in this place are a source of errors in the representation of the signal.

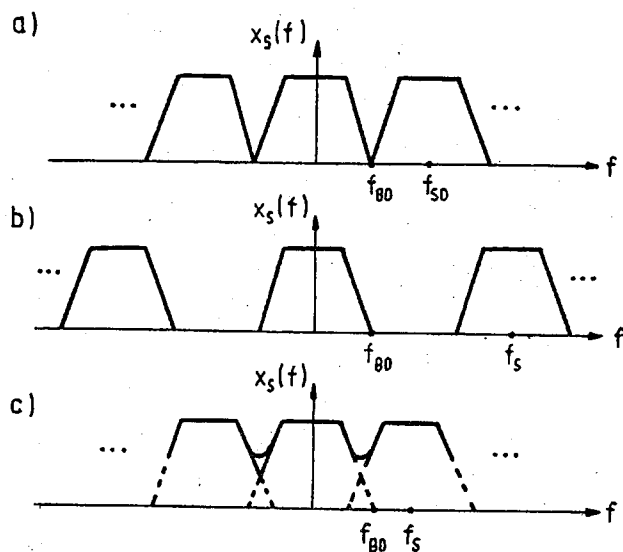


Fig. 3. The spectrum of the sampled signal: a) $f_s = 2f_{B0}$, b) $f_s > 2f_{B0}$, c) $f_s < 2f_{B0}$

It is possible to establish the proper value of f_s , if f_{B0} is definable unequivocally. To assure this, the sampling operation is preceded by low-pass prefiltration, called antialiasing filtration.

The typical measurement signals are in the lower band and the spectrum is fuzzy. As the result of antialiasing filtration, the signal band is finally established ($f_{B0} \rightarrow f_B$). Sometimes it is connected with the cutting off the highest components of the spectrum. These components may be cut off which have suitably low amplitudes, so that elimination of these components does not introduce any substantial distortions of the measurement signal within the limits of fixed error.

The ideal AAF has a flat characteristics in the signal band, infinitely large attenuation outside the transmission boundaries, constant group delay. The real AAF is an approximation of an ideal filter. Apart from the principles according to which its characteristics as low-pass filter is shaped, there are conditions resulting from the work of the filter with the analog-to-digital conversion module. The latter are the criteria connected with the resolution of the ADC and the intended sampling rate ($f_s = k \cdot 2f_B$, $k \geq 1$).

Let the normalized spectrum characteristics of AAF be $H(\omega)$. It is shaped so as to fulfil the following set of conditions:

C1: $H(\omega_B) = 1$, $\omega_B = 2\pi f_B$, f_B — is the frequency denoting the range of the flat part of the filter characteristics,

C2: $H(\omega_C) = \frac{1}{\sqrt{2}}$, $\omega_C = 2\pi f_C$, f_C — is the frequency denoting the 3-dB filter band, $f_C > f_B$,

- C3: $H(\omega_N) = \frac{1}{2} \cdot \frac{1}{2^N}$, $\omega_N = 2\pi f_N$ — is the frequency denoting the band outside which the imperfection of attenuation of the filter introduces an error not greater than the error corresponding to $\frac{1}{2}$ LSB of the applied ADC of the N-bits resolution,
- C4: $-\frac{d\varphi(\omega)}{d\omega} = \text{const}$, $-\omega_B \leq \omega \leq \omega_B$, this is the linearity condition of the phase characteristics of the filter — $\varphi(\omega)$, (i.e. constant group delay).

There are rigorous consequences for AAF of the assumption that the sampling rate f_s is close to the minimal frequency f_{s0} . However, most often f_s is the multiple of the f_{s0} (oversampling), since in the former case there are very high requirements regarding the steepness of the trailing edge and the attenuation efficiency of the filter on the reject band, and the costs of a filter with good parameters are high too. In the second case, the increase of the sampling rate causes the requirements for AAF to be softened. The filter may be simpler and therefore cheaper. A lower price is paid for the fact that the signal band is totally included in the flat part of the filter characteristics and the trailing edge will begin outside this band. The steepness of the edge may be less acute. It is illustrated in Fig.4.

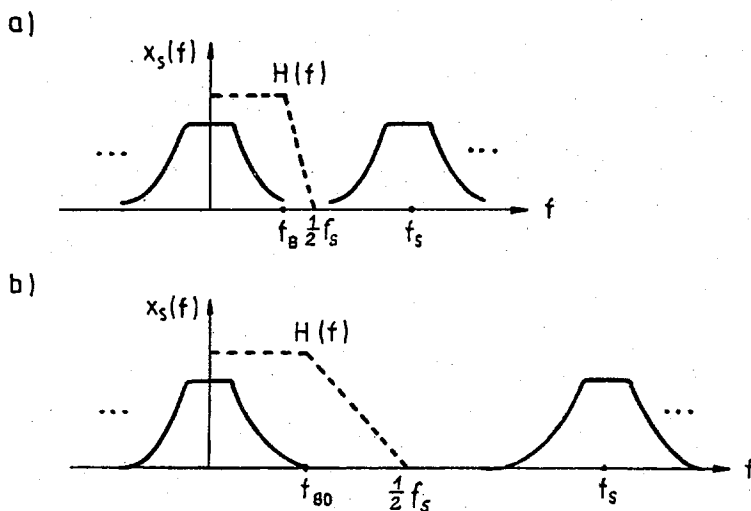


Fig. 4. Antialiasing filtering, $f_s = k \cdot f_{s0}$: a) small k , b) big k

The second advantage of sampling with the sampling rate being a multiple of f_{s0} is the reduction of power of quantization noise in the signal band. Assuming the uniform probability distribution of quantization error, the power density spectrum of the quantization noise is as follows (Otnes, Enochson, 1978):

$$G(f) = \frac{q^2}{12} \cdot \frac{1}{f_s}, \quad q = \frac{U_{FS}}{2^N} \quad (2)$$

q — resolution of ADC,

N — number of bits in the digital code word,

U_{FS} — full-scale voltage range for the analog signal.

Therefore the noise power in $[-f_B, f_B]$ band is:

$$G_0(f) = \int_{-f_B}^{f_B} G(f) df = \frac{q^2}{12} \cdot \frac{2f_B}{f_s} \quad (3)$$

If the level of quantization noise does not change (i.e. the number of bits remains the same), then the value of $G_0(f)$ is:

$$G_0(f) = G'_0(f) = \frac{q^2}{12} \quad \text{when } f_s = 2f_B \quad (4a)$$

or

$$G_0(f) = G''_0(f) = \frac{q^2}{12} \cdot \frac{2f_B}{f_s} \quad \text{when } f_s > 2f_B \quad (4b)$$

depending on the relation between f_s and f_B . If $f_s \gg 2f_B$ then $G''_0(f) \ll G'_0(f)$. The power spectrum of quantization noise takes up a wider band, so the power falling on the signal band is considerably reduced.

Sampling with the rate $f_s > f_{s0}$ appears also in connection with the issue of processing the signal with the adequately high resolution. The increase of the number of bits (N) of the ADC and resulting decrease of q causes the power density spectrum $G(f)$, as well as power $G_0(f)$, to decrease. The condition C3 is fulfilled for respectively greater frequency f_N (the increase of N). The frequency f_s is getting greater, too. On the other hand, the greater the difference between f_s and f_{s0} , the simpler the problem of antialiasing filtration. In the extreme case the AAF may become superfluous.

4. THE CONDITIONS SHAPING THE PARAMETERS OF SIGNAL SAMPLING OPERATION

In the course of the analog-digital processing the input signal U_{WE} should not undergo changes. Otherwise the assigning of the processing result to a specific actual value of the signal is laden with error. It is possible to define the processed signal band, for which this error does not exceed the width of the quantization interval. From Fig.5 it can be seen that:

$$\frac{dU_{WE}}{dt} \approx \frac{\Delta U_c}{T} \quad (6)$$

and the following is the goal:

$$\Delta U_{c \max} \leq \frac{U_{FS}}{2^N} \quad (7)$$

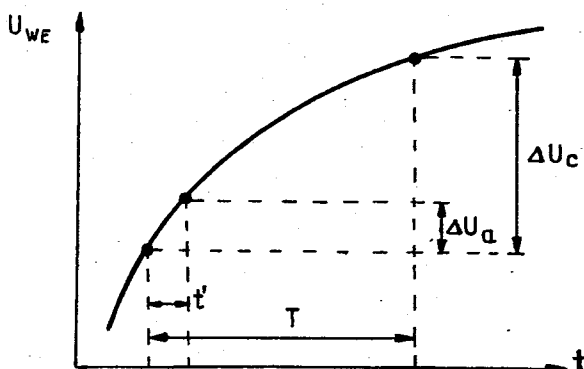


Fig. 5. The formation of the error resulting from the signal change: ΔU_c — in conversion time T_c in ADC, ΔU_a — in aperture time t'_a in SHA

The limiting form of the relationship between the maximal rate speed of input signal change and the admissible error is as follows:

$$\left(\frac{dU_{WE}}{dt}\right)_{l \max} = \frac{\Delta U_{c \max}}{T_c} = \frac{U_{FS}}{2^N} \cdot \frac{1}{T_c}. \quad (8)$$

Thus the required conversion time of ADC is:

$$T_c = \frac{U_{FS}}{2^N} \cdot \frac{1}{\left(\frac{dU_{WE}}{dt}\right)_{l \max}}. \quad (9)$$

If the frequency of the highest harmonic of the input signal (resulting from the antialiasing filtration) is f_B , $U_B = U_{B \max} \sin 2\pi f_B t$, and $2U_{B \max} \leq U_{FS}$, then the acceptable conversion time of ADC is:

$$T_c = \frac{1}{\pi 2^N \cdot f_B}. \quad (10)$$

For a specific ADC with a determined T_c , the maximal frequency of the processed signal f_B can be obtained from dependence (10). Also, it is possible to determine the frequency of taking samples f_S :

$$f_{S \max} = \frac{1}{T_c} = \pi \cdot 2^{N-1} \cdot f_{S \min}. \quad (11)$$

where $f_{S \min} = 2f_B$ (the value of sampling rate resulting from the sampling theorem).

If the value T_c is such that it does not assure the fulfilment of condition (7), then it is necessary to apply SHA. SHA samples the signal and remembers its value, keeping it constant for the time of conversion of ADC. In real SHA, the time of transition from the sampling phase to the storage phase (aperture time t'_a) does not equal zero. During this time the input signal undergoes a change. However, time t'_a is

shorter than time T_c . Assuming that the error should not exceed the width of quantization interval, the acceptable rate of signal change is greater. It can be written (see Fig.5) that:

$$\frac{dU_{WE}}{dt} \approx \frac{\Delta U_a}{t'} \quad (12)$$

Since the required version is:

$$\Delta U_{a \max} \leq \frac{U_{FS}}{2^N} \quad (13)$$

the limiting form of the relationship between the maximal rate of input signal change and the admissible error is as follows:

$$\left(\frac{dU_{WE}}{dt} \right)_{2 \max} = \frac{\Delta U_{a \max}}{t_a} = \frac{U_{FS}}{2^N} \cdot \frac{1}{t_a} \quad (14)$$

Knowing the value of t_a , it is possible to determine the admissible speed of input signal change. Considering the fact that t_a is much shorter than T_c ($T_c = k t_a$, $k \gg 1$), on the basis of (8) and (14) it can be written:

$$\left(\frac{dU_{WE}}{dt} \right)_{2 \max} = k \cdot \left(\frac{dU_{WE}}{dt} \right)_{1 \max} \quad (15)$$

The use of SHA before ADC allows to process the input signal of greater change rate. The acceptable speed of input signal change increases in such a degree in which t_a is shorter than T_c .

In case of high requirements regarding the accuracy and the speed of ADCS, SHA works with time compression t_a , achieved as a result of adequate choice of control signal. The error connected with the sampling operation is reduced to the boundary resulting from aperture indeterminacy t_{aj} (time jitter). The dependence (14) is then as follows:

$$\left(\frac{dU_{WE}}{dt} \right)_{3 \max} = \frac{U_{FS}}{2^N} \cdot \frac{1}{t_{aj}}, \quad t_{aj} < t_a \quad (16)$$

Since $T_c > t_a > t_{aj}$, then $\left(\frac{dU_{WE}}{dt} \right)_{3 \max} > \left(\frac{dU_{WE}}{dt} \right)_{2 \max} > \left(\frac{dU_{WE}}{dt} \right)_{1 \max}$.

The sampling frequency in the SHA + ADC depends on the acquisition time t_{aq} and aperture time t_a in SHA, and the conversion time T_c in ADC (Devias, 1985):

$$f_{S \max} = \frac{1}{t_{aq} + t_a + T_c} \quad (17)$$

It can be shown that $f_{S \max} \gg f_{S \min}$.

5. CONCLUSION

The consistently improving parameters of PCMI result from constant changes in two areas: the increase in scale integration of hardware elements and the application of circuits which support different analog-to-digital conversion methods currently used (Angello, 1990).

The integration of SHA and ADC in one chip was conducive to improving parameters, among others to shortening several times the processing time. For 12-bits sampling ADC, the conversion time is of the order of single microseconds, the aperture time is of the order a fractions of a microsecond. The nonlinearity error is $\pm 0,5$ LSB.

The present integration begins to include the whole hardware. One integrated circuit (IC) can hold DAS with a multichannel input. The parameters of inner circuits are determined by software. The conversion time of a 12 bit DAS does not exceed 10 microseconds. The differential nonlinearity error is $\pm 0,5$ LSB.

The second trend of changes is represented by those PCMI whose DAS uses oversampling delta-sigma data converters. The main advantage is reaching high resolution (20 bits and more). Also in this case the efficient use of delta-sigma conversion method has been made possible by progress in the integration technology. In one IC, which realizes the delta-sigma conversion, are included the quantizer with noise-shaping circuit and the digital filter. Processing according to this method allows to achieve a series of parameters which would be very difficult to obtain in the former types of ADCS. The qualitative change has also been obtained thanks to elimination of such elements as sample-hold circuitry, precision comparators and voltage dividers. In addition, the single-bit quantizer eliminates the differential nonlinearity associated with multilevel quantizer.

It may be taken for granted that the progress in the field of achieving improving parameters of the analog-to-digital conversion, and thus better accuracy properties of PCMI will be connected with wider and wider application of digital signal processing techniques.

Irrespective of the changes occurring in the configuration of PCMI, the issue of the evaluation of the reliability of measurement results remains essential in every case. The reciprocal accommodation of elements in the ADCS is one of the conditions for obtaining suitable reliability. The principles for accommodation presented in the paper must be fulfilled irrespectively of the ADCS configuration employed.

REFERENCES

1. A. Agnello: *16-bit conversion paves the way to high-quality audio for PCs*. Electronic Design, July 26, 1990, p.51—55
2. E. Davies: *Sample and hold — the key to fast A to D conversion*. Electronic Engineering, March, 1985, p.67—69
3. F. Goodenough: *ISSCC — Analog technology*, Electronic Design, February 20, 1992, p.64—75
4. F. Goodenough: *Geophysicists get 24-bit, 0.1—W ADC to sense seismic signals*. Electronic Design, February 20, 1992, p.34—38

5. Z. Kulka, A. Libura, M. Nadachowski: *Przetworniki analogowo-cyfrowe i cyfrowo-analogowe*. WKŁ, Warszawa, 1987
6. M. Łakomy, J. Zabrodzki: *Scalone przetworniki analogowo-cyfrowe i cyfrowo-analogowe*. PWN, Warszawa 1985
7. R. Otnes, L. Enochson: *Analiza numeryczna szeregów czasowych*. WNT, Warszawa 1978

A. DOMAŃSKA

PODSTAWOWE KRYTERIA WZAJEMNEGO DOPASOWANIA GŁÓWNYCH ELEMENTÓW
UKŁADU KONWERSJI ANALOGOWO-CYFROWEJ
W URZĄDZENIACH POMIAROWYCH Z KOMPUTERAMI OSOBISTYMI

Streszczenie

Układ konwersji analogowo-cyfrowej jest najistotniejszym elementem hardwarowej części urządzeń pomiarowych z komputerami osobistymi (PC-measurement instruments). Rozpatrywane jest zagadnienie wzajemnego dopasowania elementów tego układu: filtru antyaliasingowego, układu próbkująco-pamiętającego, przetwornika analogowo-cyfrowego. Wielkościami kryterialnymi są: ustalony poziom dopuszczalnego błędu oraz dynamika przetwarzanego sygnału.

Analiza struktury energetycznej i przejść międzypasmowych w arsenku galu

JÓZEF WOJAS

Profesor emerytowany, Politechnika Warszawska

Otrzymano 1992.04.15

Autoryzowano do druku 1992.12.12

Początkową część pracy traktuje o metodach zaburzeniowych symetrii kryształu stosowanych do wyprowadzenia pasm energetycznych arsenku galu z pasm energetycznych germanu. Dyskutowano o symetrii stanów w pasmach przewodnictwa i walencyjnym, o rozszczepieniu stanów i ich łączeniu się w zaburzeniu. Podano matematyczne równania na energię pasma przewodnictwa i dyskutowano kształty pasm energetycznych arsenku galu oraz zasady obliczania jego struktury pasm. Dokładnie przeanalizowano struktury: energetyczną pasma walencyjnego, przewodnictwa, przerwy wzbronionej i stanów domieszkowych. W podsumowaniu wyników wyznaczania struktury dość szczegółowo podano energie (powstałych) pików i minimów w widmie odbicia przypisywanych odpowiednim przejściom elektronów.

1. WPROWADZENIE

Chociaż stany elektronowe w pasmach scharakteryzowane są skończoną liczbą wektorów falowych $\mathbf{k}$, to liczba samych stanów jest nieskończona. Wynika stąd, że na każdy wektor falowy przypada nieskończona liczba stanów elektronowych. Aby zatem określić jednoznacznie stan elektronu, podaje się nie tylko jego wektor falowy, ale numeruje się także odpowiadające mu stany w kierunku wzrastającej energii. Stan elektronu określa się więc jednoznacznie przez podanie wektora falowego i numeru pasma. Każdy ze stanów ma zatem jednoznacznie określoną energię. Energię elektronu w n -tym pasmie opisuje się funkcją $E_n(\mathbf{k})$. Okazuje się, że $E_n(\mathbf{k})$ jest dla każdego pasma quasi-ciągłą funkcją wektora falowego. Dla kryształów o dużych objętościach, dla których dozwolone przez periodyczne warunki brzegowe wektory falowe leżą blisko siebie, funkcja ta jest prawie ciągłą.

Zespół funkcji $E_n(\mathbf{k})$ nazywamy strukturą pasmową danego kryształu. Funkcję tę wykorzystuje się w obliczeniach fizycznych właściwości ciał stałych.

Jednoznacznego wyboru wektora $\mathbf{k}$ zazwyczaj dokonuje się przez wybranie dla każdej reprezentacji¹ najkrótszego wektora falowego, który tę reprezentację generu-

¹ Reprezentacją grupy (zespołu elementów na przykład pewnych operacji symetrii) nazywamy zespół macierzy, dla których tablica iloczynów pokrywa się z tablicą grupową.

je; powstały w ten sposób obszar wektorów falowych $\mathbf{k}$ stanowi trójwymiarową strefę Brillouina, a wektory w tej strefie nazywane są często zredukowanymi wektorami falowymi. Zespół wektorów falowych na przykład $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$ tworzy w przestrzeni wektorów falowych pewną sieć; sieć tę nazywamy siecią wektorów falowych, zaś o punktach tworzących sieć mówimy jako sieciowych wektorach falowych. W teorii dyfrakcji operuje się pojęciem sieci odwrotnej. Wektory tej sieci są $(2\pi)^{-1}$ razy większe od sieciowych wektorów falowych [1] [2].

W fizyce ciała stałego sieć wektorów falowych nazywana jest zwyczajowo siecią odwrotną. Struktura pasmowa półprzewodników budowana jest na zależnościach E od $\mathbf{k}$.

Obecność przerw energetycznych, całkowite wypełnienie pasm najniższych czyli pasm walencyjnych, oraz niezapełnienie pasm wyżej leżących, czyli pasm przewodnictwa jest charakterystyczną cechą półprzewodników (a więc i dyskutowanego arsenku galu). Herman [3] przewidział na drodze teoretycznej, że układ pasm energetycznych kryształów o strukturze blendy cynkowej powinien być powiązany z układem pasm kryształów o strukturze diamentu.

2. ZABURZENIA SYMETRII W PÓLPRZEWODNIKACH

Callaway [4] rozszerzył idee Hermana i stosując jego teorię zaburzeń do wyrowadzenia pasm energetycznych obliczył w jaki sposób struktura pasmowa GaAs jest powiązana ze strukturą Ge. Przeprowadził on pierwsze, teoretyczne próby określenia energetycznej struktury pasmowej. Ponieważ Ga i As sąsiadują z Ge w periodycznym układzie pierwiastków, różnica potencjału krystalicznego między GaAs a Ge nie powinna być duża, a zatem użycie metod zaburzeniowych winno być prawidłowe.

Kryształ GaAs składa się z atomów Ga tworzących jedną sieć kubiczną powierzchniowo scentrowaną, a atomy As tworzą drugą sieć przesuniętą wzdłuż głównej przekątnej o $1/4$ jej długości. Zatem widzimy, że regularna struktura GaAs (tak jak diamentu i siarczku cynku) jest utworzona z dwu sieci przestrzennych F tylko przesuniętych względem siebie. Na komórkę elementarną przypadają tu 4 cząsteczki GaAs. Zatem potencjał zaburzający, „zmieniający” Ge w GaAs, można napisać następująco:

$$\Delta V = \sum_{\text{siatka 1}} (V_{Ga} - V_{Ge}) + \sum_{\text{siatka 2}} (V_{As} - V_{Ge})$$

gdzie wielkości V_{Ga} itd. są wyliczone z samouzgodnionych pól dla swobodnych atomów. Potencjał zaburzający może być wyrażony jako składający się z dwóch części:

$$\Delta V = \Delta V_{\text{sym.}} + \Delta V_{\text{antysym.}}$$

Callaway obliczył, że wielkość części antysymetrycznej $\Delta V_{\text{antysym.}}$ strefy jest znacznie większa niż wielkość części symetrycznej $\Delta V_{\text{sym.}}$, chociaż część symetryczna nie jest zaniedbywalna.

Znaleziono, że zaburzenie zwiększa przerwę energetyczną, przy czym większy przyczynik do tego wzrostu pochodzi z części antysymetrycznej. Część symetryczna zaburzenia może być tłumaczona teorią zaburzeń pierwszego rzędu. Energia wszystkich stanów ulega wtedy zwiększeniu, ponieważ zaburzenie części symetrycznej jest odpychające (na atomach galu i arsenu).

Wierzchołek pasma walencyjnego w punktach symetrii jest tworzony ze stanów p i jego energia ulega mniejszemu zwiększeniu, niż dno pasma przewodnictwa, które aczkolwiek mając charakter s , jest pod większym wpływem zaburzenia, ponieważ potencjał jest większy w pobliżu jądra. Część antysymetryczna potencjału zaburzającego ma zerową wartość oczekiwania dla wszystkich stanów w strefie, a więc należy tu użyć teorii zaburzeń drugiego rzędu, aby znaleźć zmiany energii.

W teorii zaburzeń drugiego rzędu zmiana energii stanu mającego funkcję falową Ψ_k^i (i -ta nieprzywiedlna reprezentacja wektora falowego k)

$$\Delta E_k^i = \sum_j \frac{|V_{ij}(\vec{k})|^2}{E_k^{i(0)} - E_k^{j(0)}} \quad (1)$$

gdzie $E_k^{i(0)}$ jest energią niezaburzoną stanu określonego funkcją Ψ_k^i , a $V_{ij}(\vec{k})$ jest elementem macierzowym równym:

$$V_{ij}(\vec{k}) = \int \Psi_k^{i*} \Delta V_{\text{antysym.}} \Psi_k^i d\tau.$$

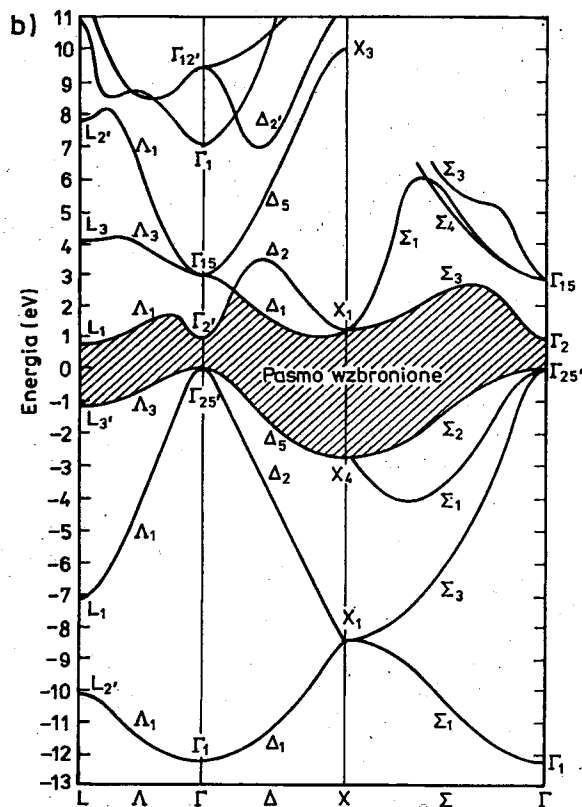
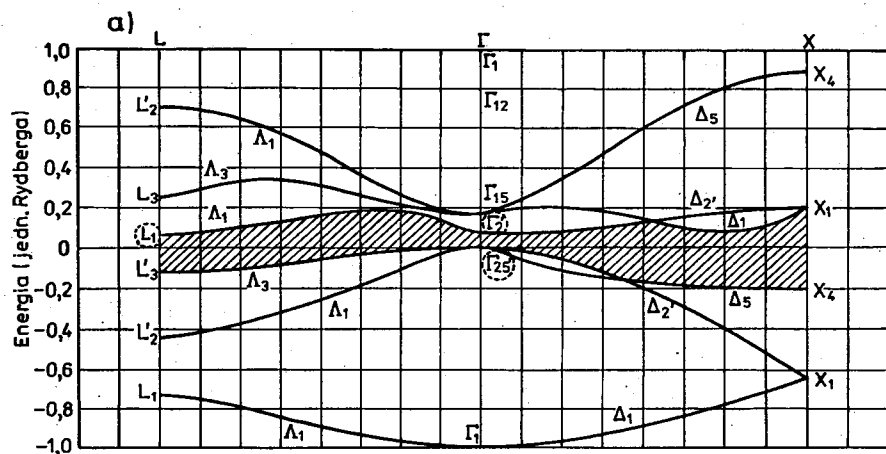
Z teorii grup można określić, które elementy macierzowe potencjału zaburzającego między stanami w punktach symetrii strefy w Ge nie są zerowe, a stąd mogą być określone zależności jakościowe między poziomem w GaAs a odpowiadającym stanem w Ge.

Ogólny obraz uzyskany przez Callaway'a przedstawia się następująco: dno pasma przewodnictwa GaAs ma nieznaczną tendencję do wzrostu względem Ge za wyjątkiem osi [100]. Jest całkiem możliwym, że na kierunkach [111] minima będą się przesuwaly od krawędzi strefy do środka, chociaż mogłyby leżeć na osiach [100]. Stany pasma walencyjnego w ogólności są obniżone przyczyniając się do zwiększenia przerwy energetycznej. Oddziaływanie spin — orbita pomijamy, gdyż nie zmienia ono w sposób zasadniczy tego obrazu.

Callaway [4] próbował ocenić wielkość wzrostu przerwy energetycznej przez użycie przybliżonych funkcji falowych w celu określenia wartości elementów macierzowych. Uzyskał on energie występujące w mianowniku równania (1) z obrazu struktury energetycznej Ge opartego na danych doświadczalnych i obliczeniach Hermana.

Wiemy już, że w celu określenia zmiany energii stanów musimy uzyskać elementy macierzowe $V_{ij}(k)$ i tzw. indeksy energii. Dla indeksów energii zastosujemy wartości pokazane na rys. 1a, b. Określimy teraz [4], na podstawie teorii grup, które elementy macierzowe tu występują.

Rozpatrzmy najpierw środek strefy Brillouina Γ [5]. Antysymetryczna część potencjału zaburzającego ma pełną symetrię grupy czworościennnej na dowolnej stronie siatki, ale ma znak przeciwny na dwu przenikających się sieciach sześciennych. Transformuje się ona zatem tak jak Γ'_2 . W celu określenia, czy element



Rys.1. a) Pasma energetyczne w germanie. Oddziaływanie spinowo—orbitalne zaniedbano. Wykresy oparto na danych doświadczalnych, obliczeniach Hermana i dobrej intuicji. Stany, w których położenia są wiadome — są cykliczne, b) Pasma energetyczne w germanie. Oddziaływanie sprzężenia spinowo—orbitalnego zaniedbano [15]

macierzowy V_{ij} jest różny od zera, rozpatrzmy iloczyn reprezentacji: $\Gamma_1 \times \Gamma'_2$. Jeżeli iloczyn ten zawiera reprezentację Γ_j , element macierzowy może być różny od zera.

Tablica 1

Prosty iloczyn reprezentacji. Antysymetryczna część zaburzenia transformuje się jak Γ'_2, L'_2, X'_2 [sieć kubiczna (regularna) powierzchniowo centrowana — (f.c.c)], $A'_2 W'_2$ (f.c.c.) i A_1 .

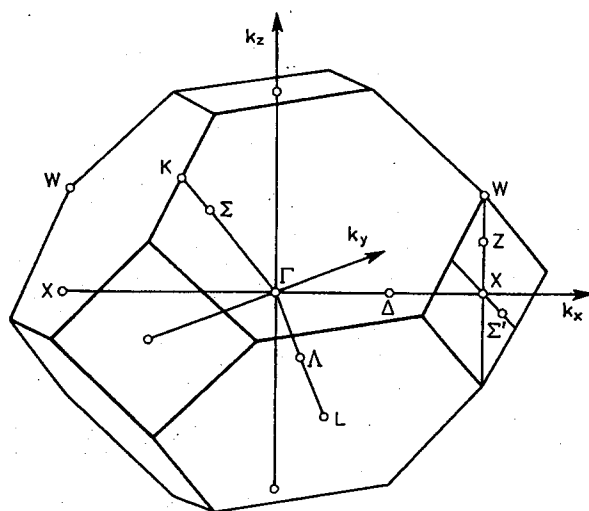
Γ_1 $\Gamma_1 \times \Gamma'_2$	$\Gamma_1 \Gamma_2 \Gamma_{12} \Gamma_{15} \Gamma_{25} \Gamma'_1 \Gamma'_2 \Gamma'_{12} \Gamma'_{15} \Gamma'_{25} \Gamma_6 \pm \Gamma_7 \pm \Gamma_8 \pm$ $\Gamma'_2 \Gamma'_1 \Gamma'_{12} \Gamma'_{25} \Gamma_{15} \Gamma_2 \Gamma_1 \Gamma_{12} \Gamma_{25} \Gamma_{15} \Gamma_7 \pm \Gamma_6 \pm \Gamma_8 \pm$
L_1 $L_1 \times L'_2$	$L_1 L_2 L_3 L'_1 L'_2 L_3 L_4 \pm L_4 \pm L_6 \pm$ $L_2 L_1 L_3 L_2 L_1 L_3 L_5 \pm L_4 \pm L_6 \pm$
X_1 $X_1 \times X'_2$	$X_1 X_2 X_3 X_4 X_5$ $X_1 X_2 X_4 X_3 X_5$
A_1 $A_1 \times A'_2$	$A_1 A_2 A'_1 A'_2 A_5 A_6 A_7$ $A'_2 A_1 A_2 A_1 A_5 A_7 A_6$
W_1 $W_1 \times W_2$	$W_1 W_2$ $W_1 W_2$
A_1 $A_1 \times A_1$	$A_1 A_2 A_3 A_4 A_5 A_6$ $A_1 A_2 A_3 A_4 A_5 A_6$

Tablica 1 daje niektóre iloczyny reprezentacji w pewnych punktach strefy. Widzimy, że stan pasma walencyjnego Γ'_{25} łączy się ze stanem pasma przewodnictwa Γ_{15} tak, że Γ'_{25} będzie obniżone a Γ_{15} podwyższone. Podobnie stan pasma przewodnictwa Γ'_2 łączy się ze stanami Γ_1 powyżej i poniżej. W związku z tym dno pasma walencyjnego Γ_1 jest również obniżone. Stan Γ'_2 podlega efektom kompensacyjnym. W punkcie L, który jest środkiem heksagonalnej powierzchni, antysymetryczna część zaburzenia transformuje się jak Γ'_2 . Najniższym poziomem pasma przewodnictwa jest L_1 . Poziom ten może nakładać się ze stanami L'_2 w pasmie walencyjnym i w pasmie przewodnictwa. Odpowiednie indeksy energetyczne są mniejsze niż w przypadku stanów pasma przewodnictwa Γ'_2 w środku strefy, a co za tym idzie ruch L'_2 może być bardziej gwałtowny niż Γ'_2 . Najwyższym stanem pasma walencyjnego jest L'_3 ; przesunie się on w dół ze względu na element macierzowy ze stanem pasma przewodnictwa L_3 (rys.1a).

Wzdłuż osi [111], pomiędzy Γ i L, antysymetryczna część potencjału transformuje się jak A_1 . A zatem mogą się łączyć tylko stany o tej samej symetrii. A_3 na szczycie pasma walencyjnego przesunie się w dół pod wpływem A_3 w pasmie przewodnictwa, A_1 na dnie pasma przewodnictwa odczuwa wpływ stanów A_1 w pasmie walencyjnym i przewodnictwa.

W punkcie X w środku płaszczyzny kwadratowej, zaburzenie ma symetrię X'_2 w sieci sześcienniej. A zatem stan X_4 w pasmie walencyjnym łączy się z X_3 w pasmie przewodnictwa i obniża się. Stany X_1 są rozszczepiane w zaburzeniu pierwszego rzędu; one również łączą się w zaburzeniu drugiego rzędu, tak że środek ciężkości tych dwu stanów na dnie pasma przewodnictwa będzie mieć tendencję do wznoszenia się.

Wzdłuż osi [100] antysymetryczna część zaburzenia ma symetrię Δ'_2 . Rozważmy najpierw symetrię w pobliżu środka strefy. Δ_1 na dnie pasma walencyjnego będzie zepchnięta w dół w wyniku oddziaływania z Δ'_2 . Δ'_2 przy wierzchołku pasma walencyjnego będzie podniesiona przez Δ_1 , lecz będzie mieć tendencję do spychania w dół przez Δ_1 z pasma przewodnictwa. Ponieważ to oddziaływanie będzie mieć znacznie mniejszy indeks energetyczny, ruch wypadkowy powinien być skierowany w dół. Podobnie Δ'_2 w pasmie przewodnictwa powinno być obniżone. Na drugim końcu osi [100], stan Δ_5 jest na szczycie pasma walencyjnego. Stan ten będzie obniżony przez nałożenie się z Δ_5 z pasma przewodnictwa. Stan Δ_1 pasma przewodnictwa przesunę się pod wpływem stanu Δ'_2 z pasma walencyjnego i stanu Δ'_2 z pasma przewodnictwa. Ten ostatni efekt będzie silny i popchnie Δ_1 niżej, tak aż dojdzie do dolnego rozszczepionego od podwójnie zdegenerowanego poziomu X_1 (rys.1a).



Rys. 2. Strefa Brillouina dla germanu i arsenku galu

W punkcie narożnym W (rys.2), rozszczepienie zdegenerowanych stanów W_2 na szczycie pasma walencyjnego jest istotne, ponieważ za wyjątkiem przekątnych płaszczyzny kwadratowej, jest to jedyne miejsce w strefie gdzie poziom płaszczyzny pasma walencyjnego wzniesie się. Istnieje zatem możliwość, że maksimum pasma walencyjnego przesunie się od Γ do W. Antysymetryczna część różnicy potencjału transformuje się tutaj jak W_2 w siatce sześcienniej. Środek ciężkości stanów utworzonych z W_1 przy dnie pasma walencyjnego obniży się, pozostałe zaś stany będą miały tendencje przeciwdziałające.

Wzdłuż przekątnej Z powierzchni kwadratowej istnieje tylko jedna nieprzywiedlna reprezentacja. A zatem te wszystkie stany będą się nakładać. Występuje tutaj również rozszczepienie pierwszego rzędu. Można oczekiwać, że rozszczepienie najniższych stanów Z_1 będzie większe niż rozszczepienie stanów na wierzchołku pasma

walencyjnego, ponieważ rozszczepienie stanów na wierzchołku pasma walencyjnego musi wchodzić do stanu X_4 , który nie jest rozszczepiony. W drugim rzędzie środków ciężkości rozszczepionych stanów przy wierzchołku pasma walencyjnego będzie podlegał wpływom kompensującym się.

Kombinacja tych informacji postuluje następujący ogólny obraz: przeważnie (za wyjątkiem punktów symetrii W i linii Z), stany przy wierzchołku i dnie pasma walencyjnego są obniżone. Prowadzi to do wzrostu przerwy energetycznej pomiędzy pasmami walencyjnym i przewodnictwa i do zwiększonej energii kohezyjnej GaAs w porównaniu z Ge. W zależności od względnych wielkości elementów macierzowych, możliwe jest że albo (a) maksimum pasma walencyjnego pozostaje w środku strefy, albo (b) przesuwa się do narożnego punktu W. Nieprawdopodobna jest jakkolwiek inna zmiana. Pasma przewodnictwa podlega sprzecznym tendencjom wznoszenia się i obniżania; tendencja do wznoszenia się jest prawdopodobnie najbardziej wyraźna za wyjątkiem osi [100]. Możliwe jest, że minimum pasma przesuwa się od punktu L do środka strefy, chociaż prawdopodobieństwo otrzymania minimów wzdłuż osi [100] nie może być całkowicie wykluczone. Istnieją pewne dane eksperymentalne, otrzymane przez Barrie [6], wykazujące małą efektywną masę elektronu w GaAs znacznie mniejszą od efektywnej masy elektronów w Ge [4] i tego samego rzędu co masa przewidywana dla Ge w środku strefy na drodze pośredniej z danych rezonansu cyklotronowego dla dziur. Można to traktować jako dowód, że minimum pasma przewodnictwa w GaAs znajduje się w środku strefy Γ (rys.1b i 2).

3. KSZTAŁT PASM GaAs

Według Callawaya [4] schemat energetycznych GaAs podobny jest do układu pasm w Ge z tą tylko różnicą, że w arsenku galu najniższe minimum pasma przewodnictwa leży przy $\vec{k} = 0$, czyli w środku strefy Brillouin'a (rys.3) [5], co potwierdza doświadczenie. Kształt pasma przewodnictwa w GaAs można znaleźć posługując się teorią Kane [7] którą opracował on dla InSb. Wykazano tu, że te dwa związki mają podobne struktury pasmowe i dlatego teoria Kane może być zaadaptowana bezpośrednio do GaAs.

Oddziaływanie między minimum pasma przewodnictwa (000), a maksimum pasma walencyjnego prowadzi do trzech pasm, które są rozwiązaniem następującego równania:

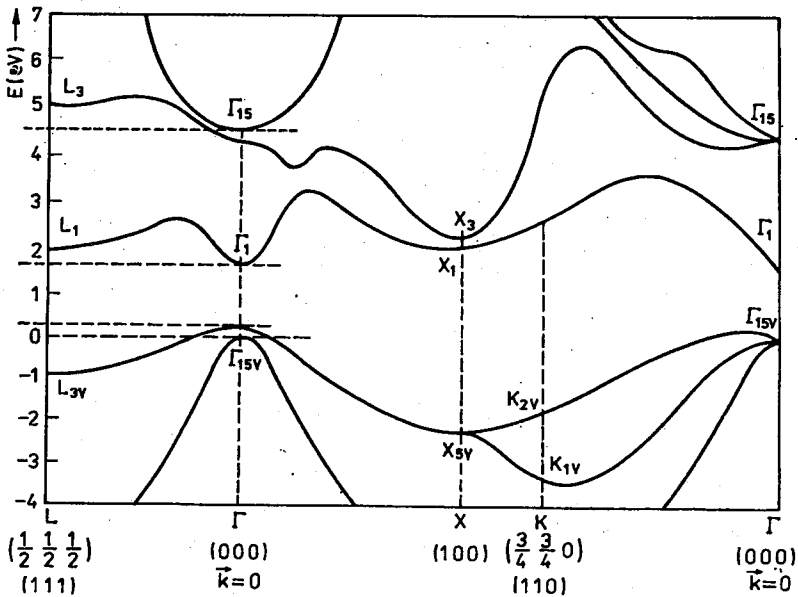
$$E_1(E_1 - E_G)(E_1 + \Delta) = \vec{k}^2 P^2 \left(E_1 + \frac{2\Delta}{3} \right), \quad (2)$$

gdzie $E_1 = E - \frac{\hbar^2 \vec{k}^2}{2m_0}$;

E — energia pasma przewodnictwa,

E_G — energia przerwy energetycznej dla przejść prostych,

Δ — rozszczepienie spin — orbita pasma walencyjnego w środku strefy,
 P — element macierzy pędu, łączący te trzy pasma będące rozwiązaniami równania (2).



Rys. 3. Energetyczny model pasm walencyjnych i przewodnictwa w GaAs obliczony metodą pseudo-potencjału (M.L.Cohen [5]). Poziome energetyczne na powierzchni czystej i pokrytej czem (G.W. Gobeli, F.G. Allen [32])

Dla małych wartości $\vec{k}$ energię pasma przewodnictwa podaje wzór:

$$E = E_G + \frac{\hbar^2 \vec{k}^2}{2m_0} + \frac{P^2 \vec{k}^2}{3} \left(\frac{2}{E_G} + \frac{1}{E_G + \Delta} \right) \quad (3)$$

Masę efektywną elektronów na dnie pasma przewodnictwa można wyznaczyć z wyrażenia:

$$E - E_G = \frac{\hbar^2 \vec{k}^2}{2m_b^*}$$

stąd

$$m_b^* = \frac{\hbar^2 \vec{k}^2}{2(E - E_G)},$$

gdzie m_b^* to masa efektywna na dnie pasma przewodnictwa, więc

$$P^2 = \frac{1,5 E_G \cdot E_G + \Delta}{3 E_G + 2 \Delta} \left(\frac{1}{m_b^*} - \frac{1}{m_0} \right), \quad (4)$$

Posługując się jednostkami atomowymi bierzemy:

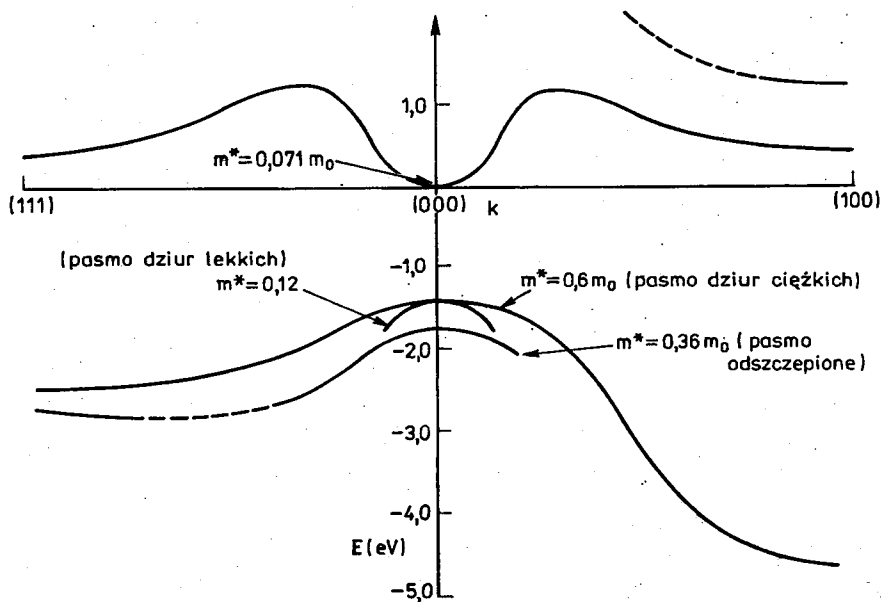
$E_G = 0,0561$; $\Delta = 0,0122$ i $m_b^* = 0,072$.

Wtedy $P^2 = 0,385$.

Podstawiając te wartości do równania (2) uzyskuje się więc rozwiązania przedstawiające kształt (rys.4):

- 1) pasmo przewodnictwa,
- 2) pasmo dziur lekkich,
- 3) pasmo odszczepione.

Korzystając więc z równania (2) wykazano, że pasmo dziur lekkich odstępkuje od paraboliczności mniej więcej o tyle samo co i pasmo przewodnictwa, a pasmo odszczepione jest nawet bliższe paraboli.



Rys. 4. Struktura pasmowa GaAs wg Hilsuma [6]

Prosta w zasadzie teoria Kane nie daje jednak żadnych wskazówek odnośnie kształtu pasma dziur ciężkich. Energię pasma przewodnictwa można w przybliżeniu zapisać w postaci:

$$E = 0,0561 + 6,95 \vec{k}^2 - 600 \vec{k}^4 \quad (5)$$

Jego odstępstwo od paraboliczności jest małe i wynosi tylko 10%, gdy $\vec{k}^2$ osiąga 10^{-3} jednostek atomowych.

4. OBLICZANIE PASM ENERGETYCZNYCH

Metoda ciasnego wiązania jest przypuszczalnie najprostszym podejściem stosowanym do opisu pasm energetycznych. Obliczenia kształtu pasm energetycznych dla dowolnego kryształu aczkolwiek są bardzo skomplikowane, są jednak dość bezpośrednie, o ile tylko zdecydujemy się na wybór konkretnego przybliżenia w opisie

oddziaływań wymiany między elektronami swobodnymi. W tym celu najpierw konstruuje się potencjał krystaliczny, a następnie rozwiązuje równanie wiekowe, z którego otrzymujemy własności własne energii i funkcje falowe.

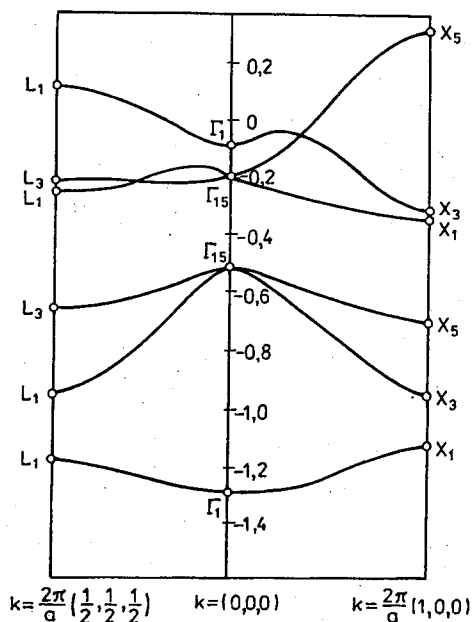
Dla przeprowadzenia tych obliczeń rozwinięto cały szereg technik obliczeniowych jak: a) metoda komórek (przybliżenie Wignera i Seitza), b) metoda fal płaskich, c) metoda ortogonalizowanych fal płaskich (OPW), d) metoda uzupełnionych fal płaskich (APW), e) metoda funkcji Greena i inne. Niektóre z tych metod zostały zastosowane do obliczeń struktury pasmowej GaAs. Pewne własności pasm energetycznych są konsekwencją istnienia symetrii kryształu, stąd z samej symetrii wiele wnioskujemy o pasmach energetycznych w rozpatrywanym kryształ. Dla otrzymania jednak samych pasm, musimy przeprowadzić szczegółowe obliczenia.

Pasma energetyczne elektronów w ciele stałym (w półprzewodniku) reprezentują jego podstawową strukturę elektronową. Ażeby otrzymać jakikolwiek rozsądny opis pasm energetycznych, należy przeprowadzać obliczenia całkowitej struktury pasmowej. Zrozumienie wyników tych obliczeń jest niezmiernie ważne dla właściwego rozumienia fizyki ciała stałego. Wyniki obliczeń struktur pasmowych przedstawiane są zazwyczaj dla wektorów falowych leżących wzdłuż linii symetrii.

4.1. PIERWSZE ZASADY OBLICZANIA STRUKTUR PASMOWYCH GaAs

Bassani i Yoshimine [8] pierwsi przeprowadzili badania związków III – V metodą ortogonalizowanych fal płaskich (OPW). Zaniedbano przy tym oddziaływanie spin – orbita.

Niektóre energie przejść dla GaAs (rys.5) wraz z ich ostatnimi wynikami doświadczalnymi zostały zebrane w tablicy 2. Z analizy tej tabeli wynika, że mimo



Rys. 5. Struktura GaAs wg [8]

Tablica 2

Porównanie obliczonych energii przejścia i piki odbicia dla GaAs

Przejście	Obliczona energia* przejścia [eV]	Pik odbicia** [eV]
$L_{3v} \rightarrow L_{1c}''$	10,4	9,6
$L_{3v} \rightarrow L_{3c}$	5,9	6,63; 6,89
$\Gamma_{15v} \rightarrow \Gamma_{15c}$	4,2	4,45
$L_{3v} \rightarrow L_{1c}$	5,4	2,90; 3,13
$\Gamma_{15v} \rightarrow \Gamma_{1c}$	5,7	1,54

* Dane, których obliczone wartości nie zwierają oddziaływania spin – orbita, lub innych (względnych) efektów.

** Dane wg Thompsona i in. Canad. J. Phys. 1966, t.44, p.2927

ogólnej zgodności, poziomy Γ_{1c} i Γ_{15c} są odwrotne. Poziomy Γ_{1c} i L_{1c} są również wysokie; są one szczególnie wygodne do wyboru potencjału krystalicznego, który określił Bassani [9] analizując elementy macierzowe hamiltonianu H. Potencjał ten można również oszacować doświadczalnie obserwując przesunięcie brzegów pasm energetycznych w przypadku wprowadzenia domieszek [10].

W większości obliczeń Bassani i Yoshimine konstruowali potencjał krystaliczny superponując potencjały atomowe; w których przyczynek coulombowski był obliczany z dokładnych funkcji falowych Hartree – Fock'a stanów walencyjnych, a przyczynek wymiany przy pomocy przybliżenia Slatera wolnych elektronów. W następnych obliczeniach wykorzystano potencjały atomowe Hartree – Focka – Slatera – Hermana i Skillmana [11] czyniąc wymianę slaterowską zwartą w użyciu. Te ostatnie obliczenia dają lepszą zgodność z doświadczeniem, niż ich pierwsze obliczenia dla Ge, w których używano zhybrydizowany potencjał atomowy.

Zgodnie z przybliżeniem Slatera wymiana krystaliczna jest proporcjonalna do pierwiastka sześciennego z krystalicznej gęstości ładunku, podczas gdy w powyższym przybliżeniu dodano pierwiastki sześciennie z atomowych gęstości ładunku. Dopasowywano więc $V(000)$ (średnia z potencjału krystalicznego w jednej komórce) do wartości $V_s(0)$, tak by najlepiej zbliżyć się do faktów doświadczalnych.

Dla GaAs użyto $V_s(0) = -2,0$ ryd, wtedy wartość potencjału $V(000)$ jest przynajmniej dwa razy większa. Wada innych transformacji fourierowskich potencjału jest spowodowana nieodpowiednim potencjałem w obszarach międzyjądrowych. Woodruff i Herman [12] zasugerowali poprawki do modelu potencjału pozwalające na zbliżenie $V(000)$ do $v_s(0)$.

Bassani i Yoshimine w swoich następnych obliczeniach dla Ge opierając się na procedurze Woodruffa otrzymali $V(000) = -2,17$ ryd. Przez proste uśrednienie potencjału komórki, używając stabelaryzowanych funkcji atomowych Hermana i Skillmana [13] dla Ga i dla As otrzymali oni $V(000) = -5,31$ ryd dla GaAs [8].

Herman i inni [14] stosowali wymianę Kohn – Sham do Ge i utworzyli schemat pośredni dający lepszą zgodność z doświadczeniem. Badano również L — zależność

(nielokalną) biorąc pod uwagę, że w metodzie Hartree–Focka dokładny potencjał jest różny dla różnych orbitali [11] [15].

Różne wyrażenia wymiany badali Brinkman i Goodman [16] i do obliczania struktury pasmowej Si stosowali metodę $\vec{k} \cdot \vec{p}$. Korelacyjne ekranowanie dyskutowali Robinson, Bassani, Knox i Schrieffer [17].

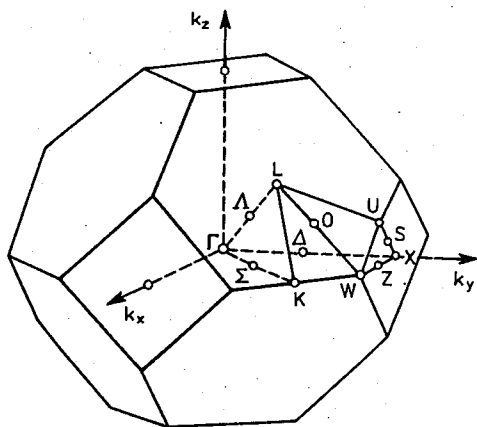
Żadna z prezentowanych metod konstrukcji potencjału krystalicznego nie daje jednak kompletnej struktury pasmowej zgodnej z doświadczeniem. To skierowało Hermana i innych badaczy [14] [18] na rozwój programów komputerowych, w których wprowadzono małe doświadczalne poprawki do pierwszych zasad samouzgodnionych. Złe współczynniki fourierowskie poprawia się aż do momentu, gdy wynikające z nich rezultaty struktury pasmowej będą się zgadzać z pewnymi poznanymi przejściami energetycznymi. Jest wtedy nadzieja, że te ewentualne poprawki uzyskają prawidłową interpretację fizyczną [19].

Na bazie otrzymanych rezultatów konkludowano, że główny odbiciowy pik nie odpowiada przejściu $X_{4v} \rightarrow X_{1c}$, jak poprzednio zakładano.

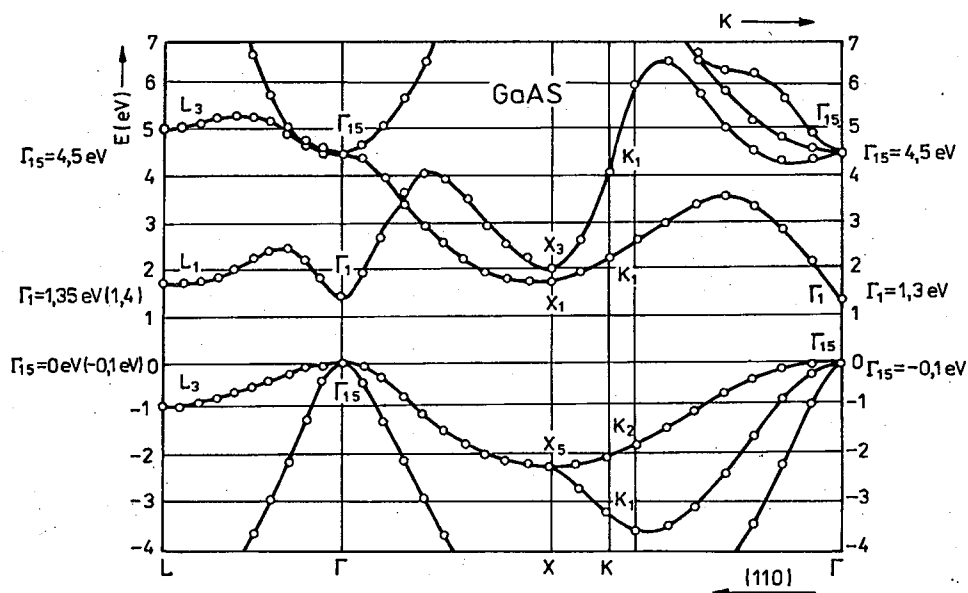
4.2. OBLICZENIA DOŚWIADCZALNE STRUKTUR PASMOWYCH Z PRZEJŚĆ OPTYCZNYCH

Przedyskutujemy teraz wyniki obliczeń uzyskane przy pomocy pseudopotencjału i metody $\vec{k} \cdot \vec{p}$. Metoda pseudopotencjału [20] opiera się na fakcie, iż termi pochodzące z ortogonalizacji do funkcji rdzenia kompensują składowe wielkiej częstotliwości potencjału krystalicznego w równaniu Schrödingera. Pozostałe termi są wyrażone jako potencjał lokalny, czyli tzw. pseudopotencjał. Metoda ta nie jest najlepsza dla wszystkich kryształów, ponieważ ostateczny term „potencjał” nie zawsze może być opisany przez potencjał lokalny.

Cohen i Bergstresser [21] wyprowadzili przy pomocy tej metody struktury pasmowe wielu półprzewodników. Pseudopotencjały dla związków III–V tworzą z pseudopotencjałów dla pierwiastków grupy IV przez dodanie niewielkiego antysymetrycznego potencjału. Pasma obliczano wzdłuż symetrycznych kierunków Δ , Δ , Z i linii łączącej X z U (rys.6). Ich rezultaty dla GaAs są reprodukowane na rys7.



Rys. 6. Strefa Brillouina diamentu lub struktur sfalerytowych



Rys. 7. Struktura energetyczna arsenku galu

Obliczone tu przejścia $X_{5v} \rightarrow X_{1c}$ dla kierunków GaAs wynoszą raczej 4 eV niż 5 eV jak to było niekiedy wyznaczane. W późniejszych obliczeniach [22] stałą dielektryczną $\epsilon_2(\omega)$ dla GaAs wyprowadzono zgodnie z [21] używając czynników struktury pseudopotencjału; otrzymana krzywa jest reprodukowana na rys. 8a. Dla porównania na rys. 8b przedstawiono krzywą $\epsilon_2(\omega)$ otrzymaną z optycznego odbicia [23]. Krzywa $\epsilon_2(\omega)$ według [22] jest bardziej podobna do krzywej odbiciowej R, niż do krzywej ϵ_2 według [23].

Z obliczeń wynikają następujące wnioski: Próg ϵ_2 przy 1,4 eV znaleziono dla przejścia $\Gamma_{15v} \rightarrow \Gamma_{1c}$, który poprzednio był przypisywany rozszczepieniu 1,5 eV. Znalaziono, że wzrost i płaskowzgórce w przedziale od 2,6 do 2,9 eV powstało z przejścia $L_{3v} \rightarrow L_{1c}$ przy 2,6 eV i $\Lambda_{3v} \rightarrow \Lambda_{1c}$ przy 2,85 eV. Na silny pik przy 4,4, (4,6) eV składają się przyczynki związane z otoczeniem punktu X, a w szczególności następujące przejścia pasmowe 4→5: $X_{5v} \rightarrow X_{1c}$ o 4,5 eV, $\Delta_{5v} \rightarrow \Delta_{1c}$ o 4,2 eV (gdzie Δ_{5v} oznacza zwyrodniałe pasma Δ_{3v} , Δ_{4v}) i przejście $\Sigma_{2v} \rightarrow \Sigma_{2v} \rightarrow \Sigma_{1c}$ uwydatnione silnym wychyleniem przy 4,4 eV. Przejścia pasmowe 4→6 są słabsze: $X_{5v} \rightarrow X_{3c}$ o 4,4 eV i $\Gamma_{15v} \rightarrow \Gamma_{15c}$ o 4,53 eV i obydwa zamaskowane są przejściami 4→5. Pik przy 5,9 (6,3) eV powstaje z przejścia $L_{3v} \rightarrow L_{3c}$.

Interpolacyjna metoda $\vec{p} \cdot \vec{k}$ oryginalnie rozwinięta przez Kane [24] była rozszerzona przez Cardonę i Pollaka [25] na całą strefę Brillouina. Zaobserwowano, że zawsze, gdy przejścia pomiędzy dwoma pasmami są zakazane przez symetrię środka albo brzegu strefy Brillouina to nieprawdopodobne są również przejścia dla bardzo małych obszarów z wnętrza strefy Brillouina, nawet jeśli nie są zabronione. To wyjaśnia dlaczego nie są obserwowane przejścia związane z najniższymi pasmami walencyjnymi. W badaniach tych również brano pod uwagę sprzężenia spin – orbita.

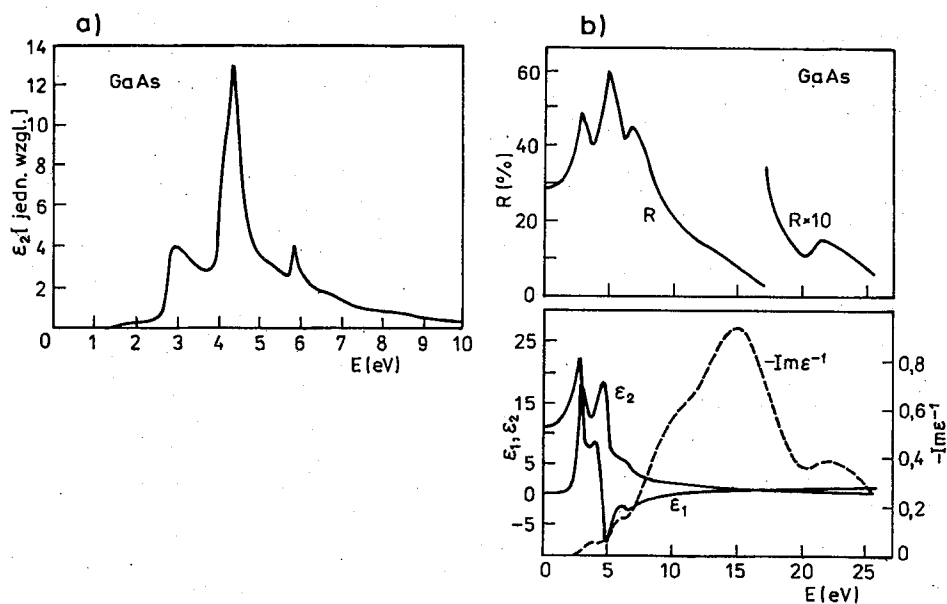
Przejścia w GaAs: dane doświadczalne i teoretyczne

Pik odbicia	Oznaczenie	Energia [eV]	Metoda	Nr publik.	Uwagi
E_0	$\Gamma_{1c} \rightarrow X_{1c}$	0.25 0.38	$\vec{k} \cdot \vec{p}$ zawierający spin oporności elektrycznej	24 53	
	$\Gamma_{15c} \rightarrow X_{1c}$	1.7	wydajność fotoelektryczna	36	
	$L_{1c} \rightarrow X_{1c}$	0.09 ± 0.02	wydajność fotoelektryczna	54	
	$*\Gamma_{15c} \rightarrow \Gamma_{1c}$	1.4 1.4 1.54	wydajność fotoelektryczna pseudopotencjał $\varepsilon(\omega)$ elektroodbitcie	36 18 55	
	$\Gamma_{15c}^{3/2} \rightarrow \Gamma_{1c}$	1.472 1.35	elektro-odbitcie wydajność fotoelektryczna	48 33	
E_1	$\Gamma_{15c}^{1/2} \rightarrow \Gamma_{1c}$	1.761 1.4	elektro-odbitcie wydajność fotoelektryczna	48 33	
	$L_{3c} \rightarrow L_{1c}$	2.6 2.9	pseudopotencjał $\varepsilon(\omega)$ EOPW	18 53	
	$A_{3c} \rightarrow A_{1c}$	2.898; 3.125 2.90; 3.13 2.9 2.85	elektroodbitcie odbitcie ε_2 z odbicia pseudopotencjał $\varepsilon(\omega)$	48 45 19 18	Wychodząc z GaAs zmienia się nachylenia wysokości pików w funkcji koncentracji.
	$X_{5c} \rightarrow X_{1c}$	4.07 4.36 4.05 4.2 4.6	odbitcie $\vec{k} \cdot \vec{p}$ zawierający spin pseudopotencjał $\varepsilon(\omega)$ EOPW OPW	45 24 18 53 1	niewytlumaczony

* Spis tych publikacji znajduje się w: phys. stat. sol., vol. 37, s. 27 (1970)

Tablica 3c.d.

E'_0	$X_{3p} \rightarrow X_{3p}$	5.1 4.8 4.4	OPW BOPW pseudopotencjał $\varepsilon(\omega)$	1 53 18	1 słaby, maskujący przejście
	$X_{1c} \rightarrow X_{3c}$	0.43 0.58 ± 0.04 0.6	wydajność fotoelektryczna wydajność fotoelektryczna EOPW	33, 34 54 53	
	$\Gamma_{15p} \rightarrow \Gamma_{15c}$	4.33	pseudopotencjał $\varepsilon(\omega)$	18	słaby, maskujący przejście
	$\Delta_{3p} \rightarrow \Delta_{1c}$	4.2 4.45 4.46; 4.64 4.42; 4.60	pseudopotencjał $\varepsilon(\omega)$ odbicie elektro- odbicie $\vec{k} \cdot \vec{p}$ zawierający spin	18 45 48 24	słaby, maskujący przejście zamykające się $\Gamma_{15p} \rightarrow \Gamma_{15c}$
	$\Sigma_{2p} \rightarrow \Sigma_{1c}$	4.4	pseudopotencjał $\varepsilon(\omega)$	18	silne przejście
E_2	$\Sigma_{2p} \rightarrow \Sigma_{1c}$	5.04 4.6 4.99; 5.345 5.2 5.0	odbicie ε_2 z odbicia elektroodbicie $\vec{k} \cdot \vec{p}$ $\vec{k} \cdot \vec{p}$ zawierające spin	45 19 48 22 24	
	$L_{3p} \rightarrow L_{3c}$	6.0 6.63; 6.89 6.3 5.9 5.9 6.2	pseudopotencjał odbicie ε_2 z odbicia pseudopotencjał $\varepsilon(\omega)$ OPW BOPW	17 45 19 18 1 53	
	$L_{3p} \rightarrow L''_{1c}$	9.6 10.4 9.0	odbicie OPW BOPW	45 1 53	
	d -- stan $\rightarrow$ pasmo przewodnictwa	22	odbicie	45	



Rys. 8. a) Urojona część funkcji dielektrycznej jako funkcja energii dla GaAs [22], b) Spektralna zależność odbicia R , rzeczywistej i urojonej części stałej dielektrycznej ϵ_1 i ϵ_2 oraz funkcja strat energii $\text{Im}\epsilon^{-1}$ dla GaAs

4.3. PODSUMOWANIE WYNIKÓW WYZNACZANIA STRUKTURY PASMOWEJ GaAs

We wcześniejszych obliczeniach struktury pasmowej, wzorowanych na pierwszych zasadach obliczeniowych, rozpatrywano tylko punkty symetryczne. Gdy zorientowano się, że wyniki doświadczalne mogły być skorelowane z teoretycznie wyznaczonymi przerwami energetycznymi, wyprowadzono dokładniejsze struktury pasmowe za pomocą interpolacyjnych technik — takich jak metoda pseudopotencjału czy $\vec{k} \cdot \vec{p}$. Korzystając z niedawno wyprowadzonych struktur pasmowych można było zinterpretować dodatkowo fakty doświadczalne i poprawić niektóre błędne przekazy. Na przykład pewne przejścia optyczne, które były przyporządkowane punktom symetrycznym odpowiadają punktom siodłowym wewnątrz strefy Brillouina.

W tablicy 3 są wyszczególnione i opisane energie i przyporządkowane im przejścia razem z faktami doświadczalnymi. Najwięcej informacji (faktów) doświadczalnych odnośnie struktury zawiera w sobie widmo optyczne. Struktura uzyskana z widm optycznych nie odpowiada bezpośrednio energiom przejść.

Obserwowany w GaAs pik odbiciowy⁽²⁾ dla 22 eV [26] przyporządkowano przejściom między nisko leżącymi stanami d a pasmem przewodnictwa. Pik odbiciowy przy 9,6 eV w GaAs przyporządkowano przejściu $L_{3v} \rightarrow L_{1c}^u$ o 10,4 eV na

⁽²⁾ Ważnym zagadnieniem badań widm odbicia „R” jest właśnie przyporządkowanie obserwowanych maksimów (w widmie odbicia) odpowiednim przejściem w strukturze pasmowej.

podstawie obliczeń metodą ortogonalizowanych fal płaskich (OPW) [9] i obliczeń Hermana [14] (EOPW — odpowiednik doświadczalny OPW określa to przejście o 9,0 eV). Dla GaAs pik odbiciowy z maksimami przy 6,63 i 6,89 eV jest przyporządkowany przejściu $L_{3v} \rightarrow L_{3c}$ z rozszczepieniem spinowo-orbitalnym przy L_{3v} [26]. OPW [14] i obliczenia pseudopotencjalne [21] umiejscowiają te przejścia w 5,9 i 6,0 eV, a według EOPW w 6,2 eV. Pik ε_2 powstały przy 6,1 eV jest zbliżony do energii obliczonego przejścia. Wyraźne piki w widmie odbicia GaAs w pobliżu 5 eV są traktowane jako bardzo słabe przejścia $X_{5v} \rightarrow X_{1c}$. Istnieją dowody poparte podobną identyfikacją w Ge i Si [15, 11, 27], że pik ten pochodzi od przejścia punktu siodłowego od $\Sigma_{2v} \rightarrow \Sigma_{1c}$. Struktura w widmie elektrycznego odbicia pojawiającego się przy 5,345 eV i 4,99 eV wykazuje zatem rozszczepienie 0,36 eV. Obliczenie [27] $\vec{k} \cdot \vec{p}$ umieszcza przejście $\Sigma_{2v} \rightarrow \Sigma_{1c}$ przy 5,2 eV ze spinowo-orbitalnym rozszczepieniem przy 5,0 eV [28]. Wczesne obliczenia OPW umiejscowiają zaś [8] przejście $X_{5v} \rightarrow X_{1c}$ przy 4,6 eV, a przejście $X_{5v} \rightarrow X_{3c}$ przy 5,1 eV; wg. metody EOPW miejsca te odpowiadają energiom 4,2 eV i 4,8 eV. Pik w widmie odbicia GaAs przy 4,45 eV przypisuje się przejściu $\Delta_{5v} \rightarrow \Delta_{3c}$ ściśle dochodzącemu do $\Gamma_{15v} \rightarrow \Gamma_{15c}$. Podobne przyporządkowanie zrobiono [29] dla struktury elektrycznych odbić przy 4,46 eV i 4,64 eV wykazujących rozszczepienie 0,18 eV.

Najnowszy sposób liczenia stałej dielektrycznej $\varepsilon_2(\omega)$ [22] oparty na czynnikach kształtu pseudopotencjału wykazuje, że pik przy 4,4 eV pochodzi od całego szeregu serii przejść: $\Gamma_{15v} \rightarrow \Gamma_{15c}$, $X_{5v} \rightarrow X_{3c}$ i $\Delta_{5v} \rightarrow \Delta_{3c}$ (Δ_{5c} dotyczy nie zredukowanej postaci grupy podwójnej); uważa się, że wszystkie przejścia są słabe z wyjątkiem $\Sigma_{2v} \rightarrow \Sigma_{1c}$.

Zwróćmy uwagę na fakt, że przyporządkowania te są w sprzeczności z przyporządkowaniami odbiciowymi i elektro-odbiciowymi dla $X_{5v} \rightarrow X_{3c}$ i $\Sigma_{2v} \rightarrow \Sigma_{1c}$ dyskutowanymi w poprzednim rozdziale. Obliczenia $\vec{k} \cdot \vec{p}$ [28] dają 4,42 i 4,60 eV dla przejścia $\Delta_{5v} \rightarrow \Delta_{3c}$ i rozszczepienie 0,18 eV zgadza się z maksymalnym rozszczepieniem na szczycie pasma walencyjnego wzdłuż Δ , obliczonym w pracy [30].

Jeśli chodzi o przejście $\Sigma_{2v} \rightarrow \Sigma_{1c}$, to ani mierzone przejście spinowo-orbitalne E'_0 (0,18 eV) ani pik E_2 (0,36 eV) nie zgadza się z obliczoną wartością [30] 0,06 eV dla maksymalnego rozszczepienia Σ . W metodzie pseudopotencjału obliczenia $\varepsilon_2(\omega)$ z GaAs [22] na przejście $X_{5v} \rightarrow X_{1c}$ przyjmuje się wartość 4,05 eV. Warto wspomnieć, że niewytłumaczony pik w widmie odbicia [26] pojawia się przy 4,07 eV.

Przy pomocy pomiarów odbicia od GaAs stwierdzono, że pik 3 eV nie zmienia swej energii równomiernie z koncentracją, jak to ma miejsce w przypadku piku o wartości 4,4 eV; zmiana w nachyleniu obserwowana jest w pobliżu 50 procentowej koncentracji. Sądzi się, że pik ten jest spowodowany przejściem $\Lambda_{3v} \rightarrow \Lambda_{3c}$, umieszczonym około wartości k_{111} . Uważa się, że zmiana nachylenia pochodzi od zmiany wartości k_{111} . Pasma, z którymi ma się tu do czynienia są prawie równoległe wzdłuż kierunków $\langle 111 \rangle$ pomiędzy przejściem $\Lambda_{3v} \rightarrow \Lambda_{1c}$ i L , tak, że przejście $L_{3v} \rightarrow L_{1c}$ jest prawie zdegenerowane z przejściem Λ .

W GaAs piki [26] uzyskane przez odbicie, widoczne są przy 2,90 i 3,13 eV. Obliczanie $\varepsilon_2(\omega)$ z GaAs z czynników [22] kształtu pseudopotencjału potwierdza, że piki te powstają z przejść $L_{3v} \rightarrow L_{1c}$ i $\Lambda_{3v} \rightarrow \Lambda_{1c}$.

Pomiary elektroodbicia [29] zdecydowanie rozszczepiają L_{3v} , na 2,898 i 3,125 eV. Oceniając je z kierunku przesunięć w widmie optycznym (z przyłożonym polem elektrycznym [31]), obie te wartości pochodzą od krawędzi punktu siodłowego. A zatem rozszczepienie A_{3v} powinno wynosić 0,23 eV. Z EOPW dla GaAs $L_{3v} \rightarrow L_{1c}$ jest 2,9 eV.

Minima w elektrycznym-odbiciowym widmie GaAs przy 1,427 i 1,761 eV przypisano przejściom odpowiednio [29] $\Gamma_{15v}^{3/2} \rightarrow \Gamma_{1c}$ i $\Gamma_{15v}^{1/2} \rightarrow \Gamma_{1c}$. W fotoelektrycznej wydajności te przejścia są widoczne przy 1,35 i 1,4 eV [32]. Obliczona wartość pseudopotencjału dla $\Gamma_{15v} \rightarrow \Gamma_{1c}$ wynosi 1,4 eV [22].

Następna dodatkowa identyfikacja była wykonana dla GaAs. Z wydajności fotoelektrycznej przejście $\Gamma_{15v} \rightarrow \Gamma_{1c}$ zostało ocenione [33] na 1,7 eV i rozszczepienie pasma przewodnictwa $X_{1c} \rightarrow X_{3c}$ na 0,43 eV [32; 34]. Dodatkowo, pomiary fotoelektrycznej wydajności umiejscawiają $X_{1c} \rightarrow X_{3c}$ na $0,58 \pm 0,04$ eV i $L_{1c} \rightarrow X_{1c}$ na $0,09 \pm 0,02$ eV. Z obliczeń EOPW wynika, że $X_{1c} \rightarrow X_{3c}$ wynosi 0,6 eV.

Marcus Gray [35] na podstawie zebranych danych dochodzi do następujących wniosków. Wyniki empirycznego wyznaczania struktury pasmowej zostały całkowicie skorelowane z informacjami uzyskanymi na drodze doświadczalnej. Częściowo dyskusyjne pozostaje położenie $A_{2v} \rightarrow \Sigma_{1c}$, a mianowicie nie ma całkowitej pewności, czy to przejście jest głównym przyczynkiem do piku refleksyjnego E_2 , jak sugerują obliczenia $\vec{k} \cdot \vec{p}$, czy może odpowiada pikowi E'_0 odbiciowemu (refleksyjnemu), jak sugerują to obliczenia $\varepsilon_2(\omega)$ przeprowadzone przy pomocy czynników kształtu pseudopotencjału?

Najbardziej logiczny sposób porównywania widma odbicia z obliczoną strukturą pasmową polega na zredukowanej postaci $\varepsilon_2(\omega)$. Natomiast najbardziej obiecującym podejściem wydaje się być pół-empiryczna metoda Hermana EOPW; bez wątpienia dużo lepszy wgląd do wartości tej metody będzie można otrzymać po obliczeniach punktów pośrednich w strukturze pasmowej, jak również $\varepsilon(\omega)$. Na podstawie pierwszych doświadczeń nie można jeszcze uzyskać pełnej całkowitej struktury pasmowej zgodnej z danymi eksperymentalnymi. Teoretyczne wyniki mogą być wprowadzone do dalszych rozważań, jeśli zamiana potencjału i efekty korelacyjne zostaną wyrażone w zadowalający sposób.

Nawet dla stosunkowo lekkich związków takich jak GaAs i GaP, istnieje możliwość dużych poprawek relatywistycznych, wykluczająca stosowanie techniki perturbacyjnej. Zachodzi wtedy konieczność wprowadzenia relatywistycznych równań falowych. Pozostaje jeszcze interesujące, czy nie relatywistyczne formułowanie jedno-elektronowe jest wystarczające dla dokładnego ilościowego przewidzenia wyniku eksperymentalnego zjawisk w półprzewodnikach [36] [37].

5. PASMA WALENCYJNE GaAs

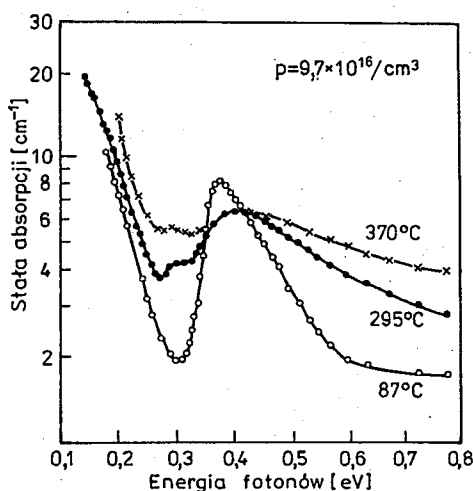
Niemal wszystkie związki grupy $A^{III} B^V$ (w tym i GaAs) winny być interpretowane na gruncie pasm walencyjnych. Struktura pasma walencyjnego GaAs jest bardzo podobna do struktury obserwowanej w Ge typu p , to znaczy z pasmem:

a) dziur ciężkich,
 b) dziur lekkich,
 c) i pasmem z dziurami o masie efektywnej pośredniej, odszczępionym od pozostałych dwu pasm oddziaływaniem spin—orbita⁽³⁾. Strukturę pasma walencyjnego GaAs określono z pomiarów absorpcji promieniowania podczerwonego na swobodnych dziurach kryształu typu *p*.

Z danych uzyskanych z pomiarów absorpcji w podczerwieni (powodowanej przez swobodne dziury w GaAs typu *p*) wydedukowano, że trzy piki absorpcyjne występujące po stronie niskoenergetycznej krawędzi absorpcji samoistnej [38] należy przypisać przejściom między trzema pasmami walencyjnymi, podobnie jak w Ge typu *p*. Tę samą strukturę stwierdzono we wszystkich próbkach typu *p*, co w konsekwencji jest dowodem, że te pasma nie są spowodowane obecnością nieznanych domieszek lub defektów sieci. Warto zaznaczyć, że struktury tej nie dają próbki typu *n*.

Powstałe widma (podobne do pików występujących w Ge typu *p*) można wytłumaczyć przejściami dziur między różnymi gałęziami — pasmami walencyjnymi arsenku galu.

Na rys.9 pokazano widma absorpcyjne GaAs w funkcji temperatury. Badana próbka była domieszkowana cynkiem. Widzimy, że przy obniżaniu temperatury pasma wyostrzają się i przesuwają nieco z położenia jakie miały w temperaturze pokojowej.

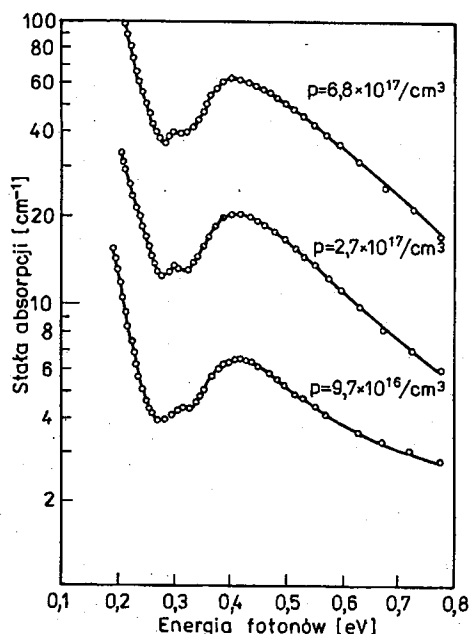


Rys. 9. Widma absorpcji GaAs typu *p* dla trzech temperatur

Próbki domieszkowane kobaltom pokazują struktury identyczne i podobną zależność temperaturową. Na rys.10 zaś pokazano widma absorpcyjne GaAs typu *p* w 295K dla pewnego przedziału koncentracji nośników.

Absorpcyjny przekrój czynny tj. stosunek współczynnika absorpcji do gęstości nośników swobodnych przy danej energii jest w przybliżeniu stały.

⁽³⁾ Oddziaływanie spinowo-orbitalne jest oddziaływaniem pomiędzy momentem magnetycznym spinowym — a orbitalnym polem magnetycznym.



Rys. 10. Widma absorpcji GaAs przy 295 K dla trzech zakresów koncentracji nośników [37]

Jeśli dla prostoty przyjmuje się strukturę pasma walencyjnego GaAs podobną do struktury Ge typu *p* [39; 40] to szczegóły widma na rys.9 mogą być wytłumaczone przejściami pionowymi między różnymi pasmami walencyjnymi GaAs. Pik absorpcyjny przy 0,42 eV przypisano przejściom między pasmem odszczepionym powstałym w wyniku sprzężenia (oddziaływania) spin – orbita, a pasmem dziur ciężkich; pik przy 0,31 eV przypisano przejściom między pasmem odszczepionym a pasmem dziur lekkich (absorpcja na swobodnych dziurach). Absorpcja zaczynająca się przy około 0,25 eV odpowiada z kolei przejściom między pasmami dziur lekkich i ciężkich, oczywiście idąc za interpretacją podobnego widma absorpcji Ge typu *p*.

W GaAs obserwuje się więc trzy przejścia pasmowe międzywalencyjne. Zasadniczym czynnikiem powstania możliwości obserwowania tych przejść jest to, aby rozszczepienie spin – orbita było mniejsze od przerwy energetycznej i aby absorpcja na swobodnych nośnikach nie poszerzała lub nie przesłaniała struktur; na przykład w krzemie nie obserwowano przejść międzywalencyjnych, ponieważ struktura jest prawdopodobnie przesłaniana przez absorpcję na swobodnych nośnikach [41].

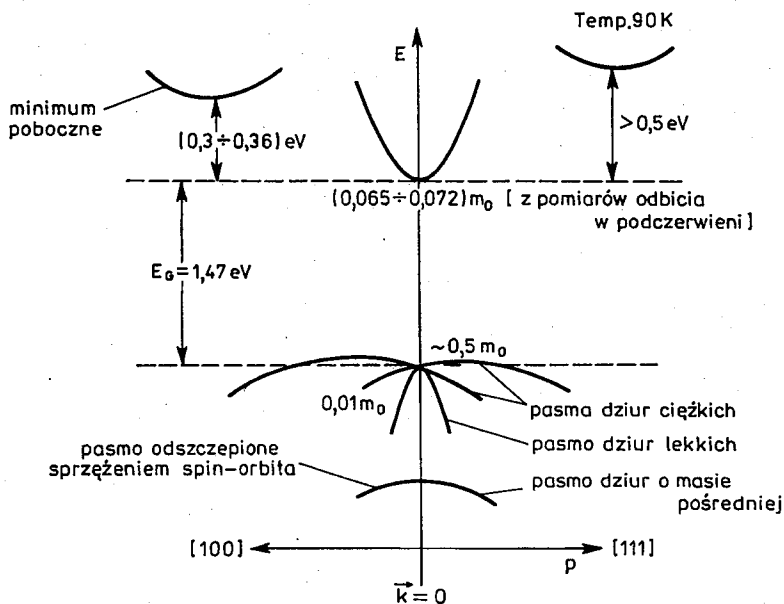
Z tych obserwacji Braunstein wydedukował, że rozszczepienie spin-orbita jest około 0,33 eV i że masa efektywna dziury lekkiej wynosi około 0,17 masy dziury ciężkiej. Zatem stosunki mas efektywnych dla dziur wynoszą:

$$\frac{m_1^*}{m_3^*} = 3,38 \quad \text{i} \quad \frac{m_3^*}{m_2^*} = 1,70^{(4)}.$$

(4) Oznaczenia indeksów przy masach efektywnych: 1 — pasmo dziur ciężkich, 2 — pasmo dziur lekkich, 3 — pasmo odszczepienie.

Ten dowód na istnienie dziur lekkich i ciężkich należy brać pod uwagę w teoriach transportu dla GaAs.

Z powodu braku płaszczyzny inwersji w sieci kryształów GaAs, oczekuje się, że degeneracja pasm dziur ciężkich będzie rozszczepiona [42; 43]; ich wierzchołki są względem siebie nieco rozsunęte. W konsekwencji maksymalne energetyczne pasma walencyjne nie leżą dokładnie w $\vec{k} = 0$ (rys.11), ale w bardzo małej odległości i o bardzo małą energię wyżej od maksimum pasma dziur lekkich. Uzyskanie dowodów na człon liniowy wymagało ilościowego zbadania kształtu piku pochodzącego od przejść między pasmami dziur lekkich i ciężkich. Wartość 0,33 eV uzyskana na rozszczepienie spin-orbita w GaAs pozwala na wykonanie oszacowania czasu jaki elektron walencyjny spędza w $\vec{k} = 0$ na atomie galu i arsenu. Rozszczepienia spin-orbita atomowych funkcji P: Ga, As i Ge oszacowane z widm atomowych wynoszą odpowiednio: 0,10; 0,30; 0,20 eV⁽⁵⁾.



Rys. 11. Uproszczony model pasmowy GaAs (m_{ef} dziur lekkich = 0,17 m_{ef} dziur ciężkich). Z powodu braku płaszczyzny (symetrii) inwersji maksima pasm dziur ciężkich nie leżą w środku strefy

Przyjmując ten sam współczynnik normalizacji dla GaAs jakiego wydaje się być odpowiednim dla Ge przy szacowaniu rozszczepienia spin-orbita w kryształach z wartości dla atomów swobodnych mianowicie z $\left(\frac{29}{20}\right)$ stwierdzono, że elektron na wierzchołku pasma walencyjnego przebywa 65% swego czasu na arsenie, a 35% na galu.

⁽⁵⁾ Oszacowania te zostały wykonane przez E. Kane'a.

6. PASMO PRZEWODNICTWA GaAs

Teoretyczne rozważania Callaway'a zasugerowały, że najniższe minimum pasma przewodnictwa GaAs powinno leżeć w środku strefy, co później zostało potwierdzone przez doświadczenia.

Z pomiarów magneto-oporu podłużnego i poprzecznego wynika, że najniższe minimum pasma przewodnictwa GaAs winno być symetryczne sferycznie i dlatego musi ono znajdować się w środku strefy. Glicksman [44] stwierdził, że w GaAs typu n o koncentracji elektronów $n = 4 \cdot 10^{17} \text{cm}^{-3}$ w temperaturze pokojowej występował izotropowy magnetoopór poprzeczny i tylko mały jego efekt podłużny. Również pomiary piezooporu przeprowadzone przez Sagar'a [45] na GaAs typu n wykazały sferyczne minimum pasma przewodnictwa w środku strefy.

Kształt krawędzi absorpcji podstawowej w GaAs jest w zgodności z przejściami prostymi z pasma walencyjnego do pasma przewodnictwa, a ponieważ maksimum pasma walencyjnego — jeśli pominiemy lekkie rozsuniecie — znajduje się w $\vec{k} = 0$, to najniższe główne minimum pasma przewodnictwa musi także być w środku pierwszej strefy⁽⁶⁾ Brillouina. Pasma przewodnictwa posiada również minima poboczne leżące daleko od środka strefy.

Callaway'owski obraz teoretyczny pasma przewodnictwa GaAs wskazuje, że w jego strukturze występują minima leżące na kierunkach [111] w przestrzeni $\vec{k}$ o energii nieco wyższej od energii minimum przy $\vec{k} = 0$. Callaway nie rozpatrywał energii minimów leżących na kierunkach [100].

Te minima poboczne zostały zaobserwowane doświadczalnie. Na przykład Spitzer i Whelan [44] obserwowali w GaAs typu n dodatkową absorpcję promieniowania podczerwonego w przedziale energii fotonów od 0,25 do 1,25 eV. Tę dodatkową absorpcję proporcjonalną do koncentracji elektronów n zinterpretowano jako wynik przejść z minimum pasma przewodnictwa w środku strefy do innych minimów o większej energii.

Dalszy dowód na minimum poboczne pasma przewodnictwa pochodzi z doświadczeń elektrycznych. Jeśli bowiem GaAs typu n jest grzany powyżej 500 K, współczynnik Halla R zwiększa się i przechodzi przez maksimum, gdy dochodzi i ustala się przewodnictwo samoistne σ_0 przy wystarczająco wysokiej temperaturze. ten wzrost R jest prawdopodobnie wynikiem zapełniania wyższych minimów, które występują wzdłuż kierunku [100].

Według innego autora — Hilsuma [46], wpływ minimów wyższych na przejścia elektronów pokazują pomiary współczynnika Halla R w temperaturach > 600 K. Bowiem w bezdomieszkowych próbkach GaAs typu n , w miarę jak temperatura rośnie — współczynnik Halla R przechodzi przez maksimum zanim rozpocznie się przewodnictwo samoistne. Należy tego oczekiwać jeśli elektrony są przenoszone wzbudzeniami termicznymi do wyższych minimów pasma przewodnictwa. Małeje

⁽⁶⁾ Prosta komórka sieci odwrotnej (sieci w przestrzeni Fourier'a), czyli najmniejsza objętość ograniczona płaszczyznami prostopadłymi do każdego — krótszego wektora (sieci odwrotnej) G w jego punkcie środkowym, została nazwana pierwszą strefą Brillouin'a.

wtedy koncentracja n , bo elektrony zapełniają wtedy tylko wyższe minima i jeszcze nie przewodzą. Tak będzie, jak powiedziano wyżej, jeśli elektrony termiczne wzbudzone są przenoszone tylko do minimów wyższych pasma przewodnictwa. Takiego przebiegu współczynnika Halla R można oczekiwać, jeśli przyjmuje się, że minimum wyższe jest o 0,36 eV.

Edwards, Slykhouse i Drickamer [47] wywnioskowali z przeprowadzonych pomiarów przesunięcia krawędzi absorpcji optycznej ze zmianą ciśnienia, że minima [100] leżą więcej niż 0,5 eV powyżej dna pasma przewodnictwa.

Jednakże Ehrenreich [49] przeanalizował wyniki doświadczalne rzucające światło na strukturę pasmową GaAs i doszedł do wniosku, że najniższe minimum pasma przewodnictwa leży w centrum strefy, z minimami wzdłuż kierunku [100]. Zostało później ustalone, że minima wzdłuż kierunków [100] są w istocie najniższymi minimami pobocznymi w pasmie przewodnictwa GaAs i że leżą one około 0,3 eV powyżej minimum przy $\vec{k} = 0$. Ponieważ jednym z parametrów użytych w rozwiązywaniu równania (2) jest m_b^* na dnie pasma przewodnictwa, więc jedyną dodatkową informacją jaką możemy uzyskać z tego rozwiązania (równanie 5) jest zmiana m^* ze zmianą $\vec{k}$.

Dla próbek zdegenerowanych zależność m^* od koncentracji nośników n może być obliczona bezpośrednio, ponieważ $\vec{k}$ jest powiązane z n zależnością:

$$3 \pi^2 n = \vec{k}^3$$

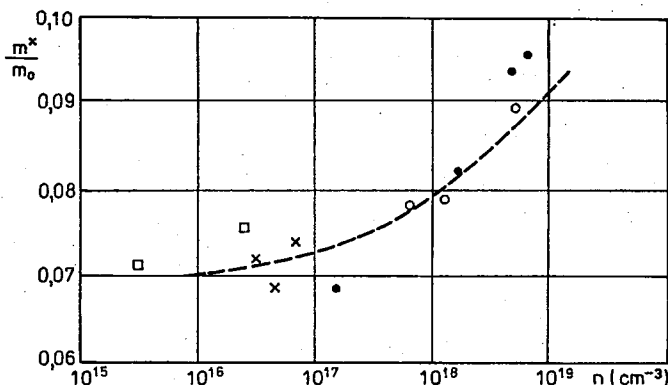
Można podać również przybliżone rozwiązanie równania (2); podaje ono też kształt pasma przewodnictwa (czyli zależność E od $\vec{k}$):

$$E = \frac{\hbar^2}{2m_b^*} \vec{k}^2 + \frac{\left(\frac{\hbar^2}{2m_b^*}\right)^2}{E_g} \vec{k}^4. \quad (6)$$

Równanie (6) odstępuje od paraboli więcej niż dokładniejsze rozwiązanie, które przedstawia sobą, równanie 5 i dlatego wskazuje ono na nieco większą zależność m^* od n . Zgodność tej zależności z doświadczeniem jest mimo wszystko dość dobra, jak to przedstawia właśnie rys.12. Widzimy więc, że masa efektywna elektronu m_b^* zależy od koncentracji elektronów n i od kształtu pasma przewodnictwa

Kształt pasma przewodnictwa wpływa również na położenie poziomu Fermiego $E_F^{(7)}$. Jego położenie zależy więc od degeneracji próbki, czyli kształtu pasma przewodnictwa. Położenie E_F zależy głównie od: domieszkowania, temperatury, masy efektywnej elektronów i dziur.

⁽⁷⁾ W OK potencjał chemiczny μ równy jest energii Fermiego, nazywanej w fizyce ciała stałego poziomem Fermiego.



Rys. 12. Zależność masy efektywnej od koncentracji elektronów: ○ Spitzer i Whelan [44], ● Cardona [30], □ Moss i Walton [49], × Palik, Stevenson i Wallis [50], — — — krzywa teoretyczna

7. PRZERWA ENERGETYCZNA W GaAs

Aby przenieść elektron ze stanu podstawowego do stanu wzbudzonego potrzebna jest pewna, skończona energia na przejście przerwy energetycznej. Określenie przerwy E_G dla GaAs z danych elektrycznych nie jest łatwe, ponieważ potrzebuje do jej wyznaczenia, pełne przewodnictwo samoistne występuje dopiero > 900 K (przy takich temperaturach często występują zmiany nieodwracalne).

Już wczesne eksperymenty na próbkach typu n wykazały ponadto niezwykle wzrost stałej Halla R_H z temperaturą T powyżej 600 K z maksimum występującym przy 870 K. Whelan i Wheatley [51] zasugerowali, że efekt ten jest spowodowany zmianą czynnika rozproszeniowego nośników, jak również może uzasadniać przewodnictwo w minimach [100], które jak już wiemy występują około 0,3 eV powyżej głównego minimum. Callaway doszedł do wniosku, że przerwa pasmowa w GaAs w centrum strefy Brillouin'a będzie się zwiększać od wartości przerwy w Ge równej 0,8 eV do wartości 1,2 eV.

Wartość przerwy pasmowej (w centrum strefy) obliczona z właściwości elektrycznych w OK wynosi 1,4 eV. Wartość ta jest z pewnością za mała, a pomiary na GaAs o wysokiej oporności dają wartość 1,58 eV, która jest w lepszej zgodności z przerwą określoną drogą optyczną wynoszącą 1,55 eV. E_G uzyskane z doświadczenia wyniosło 1,47 eV w 90 K i 1,40 eV w temperaturze pokojowej.

Wykazano, że E_G dla przejść prostych może być wydedukowana z absorpcji w podczerwieni, czyli z wyznaczenia krawędzi absorpcji podstawowej i wynosi ona 1,52 eV w OK, a około 1,4 eV w temperaturze pokojowej. Ponieważ maksima pasma dziur ciężkich są nieco przesunięte, to można powiedzieć, że w ogólności stany pasma walencyjnego są obniżone przyczyniając się tym do zwiększenia przerwy energetycznej. Ciśnienie wywierane na kryształ ma wpływ na szerokość przerwy energetycznej i po wpływie tego ciśnienia na szerokość przerwy wnioskuje się wzdłuż

jakiego kierunku występuje dane minimum pasma przewodnictwa. Jeżeli bowiem E_G nie zmienia się szybko z ciśnieniem, to prawdopodobnie wchodzące w grę minimum pasma przewodnictwa znajduje się wzdłuż kierunku [100].

Wraz z ciśnieniem przesuwa się krawędź absorpcji optycznej czyli zmienia się również szerokość przerwy energetycznej [52]. Szerokość przerwy energetycznej E_G maleje z ciśnieniem w stosunku $9,4 \cdot 10^{-6}$ eV/atm.

Przejścia proste z pasma walencyjnego do minimów pasma przewodnictwa przy krawędzi strefy są określone z odbicia w ultrafiolecie, na przykład przejścia proste na krawędzi (100) wynoszą 5,12 eV do najniższych minimów (100) pasma przewodnictwa, a 5,55 eV do minimów wyższych. Przy krawędzi (111) przerwa pasmowa dla przejść prostych wynosi 2,99 eV; odbicie w ultrafiolecie także ujawnia przejścia na tej krawędzi strefy z pasma walencyjnego do gałęzi pasma przewodnictwa położonej 6,6 eV wyżej.

8. PASMA DOMIESZKOWE

Kwantowa teoria stanów domieszkowych w półprzewodnikach podaje, że odpowiednia funkcja falowa zanika eksponencjalnie ze wzrostem odległości od domieszki. Do poszukiwania takich właśnie zanikających funkcji używa się metody $\mathbf{k} \cdot \mathbf{p}$, gdzie $\mathbf{k}$ jest wektorem falowym, a $\mathbf{p}$ — pędem elektronu. Uogólnienie formuły $\mathbf{k} \cdot \mathbf{p}$ na pasmo wzbronione jest równoważne analitycznemu przedłużeniu parabolicznego pasma energetycznego na przestrzeń zespolonych wektorów falowych.

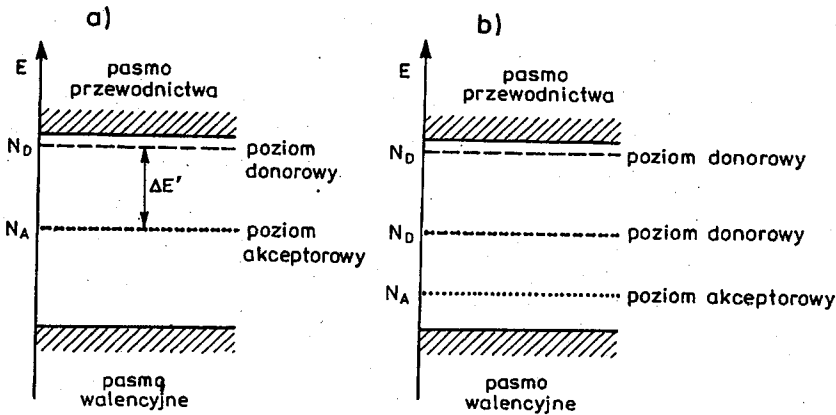
Zauważmy, że używając przedłużenia analitycznego pasma, możemy rozpatrywać zarówno stany leżące głęboko w obszarze pasma wzbronionego, jak również i te, które leżą blisko krawędzi pasma. Funkcja falowa, opisująca stan powierzchniowy leżący blisko krawędzi pasma, będzie zanikać w miarę oddalania się od powierzchni bardzo słabo. Gdy wnikamy głębiej w obszar energii wzbronionych, wówczas zanik ten staje się coraz ostrzejszy, a stany energetyczne, odpowiednio, coraz bardziej zlokalizowane.

Rozpatrując wysokooporową próbkę GaAs o donorach N_D i akceptorach N_A przy poziomie donorowym będącym bardzo blisko dna pasma przewodnictwa i poziomie akceptorowym o energii $\Delta E'$ poniżej niego, znaleziono średnią wartość $\Delta E'$ równą $\sim 0,6$ eV (rys.13). Jeżeli więc w GaAs $N_D \approx 10^{17} \text{ cm}^{-3}$, a $N_A \approx 10^{16} \text{ cm}^{-3}$, to próbka będzie typu n o koncentracji nośników w pobliżu 10^{17} cm^{-3} i oporności około $0,01 \Omega \text{ cm}$. Jeżeli zaś $N_A = 10^{17} \text{ cm}^{-3}$ i $N_D = 10^{16} \text{ cm}^{-3}$, to próbka będzie typu p , a koncentracja dziur będzie wynosiła około 10^{11} cm^{-3} i oporność będzie rzędu $10^6 \Omega \text{ cm}$.

Koncentrację dziur p obliczono ze wzoru:

$$p^2 = N_V N_A \exp \left[\frac{(\Delta E' - E_G)}{kT} \right] \quad (7)$$

gdzie N_V — gęstość stanów w pasmie walencyjnym,



Rys. 13. Wykresy poziomów energetycznych dla wysokooporowego GaAs, a) wg Allena [52], b) wg Blanca i Weisberga [52]

$$(N_V = 10^{19} \text{ cm}^{-3}, \text{ jeżeli } m_h^* \text{ wynosi } 0,5 m_0),$$

$\Delta E'$ — odległość poziomu akceptorowego od dna pasma przewodnictwa.

Chociaż wykonano próbki o wyżej podanych właściwościach elektrycznych, to znaczna ich część o dużej oporności, to jest rzędu $10^7 \Omega \text{ cm}$, może być próbkami typu n i to o koncentracji elektronów między 10^8 a 10^9 cm^{-3} . Względna wielkość N_A i N_D w takich próbkach możemy obliczyć ze wzoru:

$$n = \frac{N_D}{N_A - N_D} N_C \exp\left(-\frac{\Delta E'}{kT}\right) \quad (8)$$

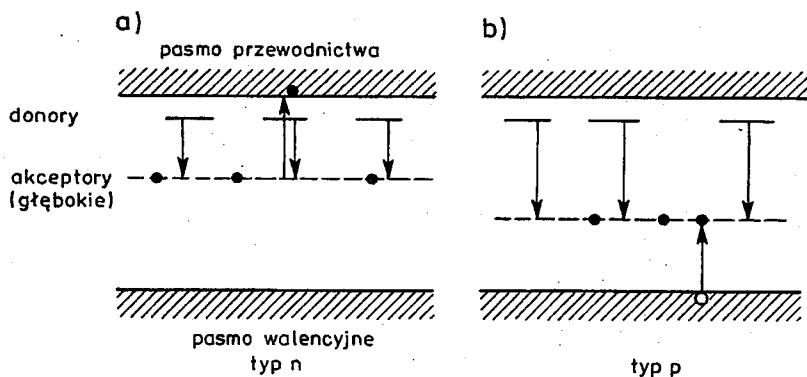
gdzie N_C — gęstość stanów w pasmie przewodnictwa,

$$(N_C = 4,7 \cdot 10^{17} \text{ cm}^{-3}, \text{ jeśli } m_e^* = 0,072 m_0).$$

Różnica między N_A i N_D wynosi 10% N_D dla niższych koncentracji elektronów i 1% N_D dla wyższych.

W wysokooporowych próbkach GaAs typu n poziom akceptorowy leży od 0,6 do 0,65 eV poniżej dna pasma przewodnictwa, a więc jest to poziom bardzo głęboki (daleko oddalony od pasma walencyjnego). W GaAs typu n akceptorem głębokim jest pasemko położone daleko w przerwie — bliżej poziomów donorowych (bliżej pasma przewodnictwa) rys.14.

Jeżeli w GaAs typu n domieszki donorowe stanowią płytki poziom domieszkowy, to położony jest on w pobliżu dna pasma przewodnictwa. Na przykład poziom donorowy wytwarzany przez domieszki krzemowe leży około 0,005 eV poniżej głównego dna pasma przewodnictwa. Ge jako domieszka zachowuje się w sposób podobny do Si. GaAs silniej domieszkowany wykazuje zwykle cechy przewodnictwa pasma domieszkowego. Najczystszy GaAs posiada całkowitą zawartość domieszek nieco większą niż 10^{16} cm^{-3} . W przypadku wysokooporowego GaAs typu n koncentracja nośników n jest o kilka rzędów wielkości mniejsza od N_D i N_A .



Rys. 14. Kompensacja domieszek donorowych przez głębokie akceptory w GaAs (tworzenie materiału typu *n* i typu *p*)

Występowanie dużego oporu próbek GaAs jest w zgodności z modelem, w którym domieszki donorowe z płytkim poziomem domieszkowym są kompensowane przez głęboki akceptor posiadający poziom w pobliżu środka przerwy pasmowej. Chociaż wystąpił nadmiar tych akceptorów, to materiał pozostanie typu *n*, jeżeli prawdopodobieństwo ponownego wzbudzenia pochwycionych elektronów do pasma przewodnictwa będzie większe od prawdopodobieństwa pochwycenia dodatkowych elektronów z pasma walencyjnego (rys.14). [GaAs intensywnie strefowo czyszczony wykazuje czasami bardzo wysoki opór i ma współczynnik Halla R wskazujący, że w temperaturze pokojowej koncentracja elektronów wynosi tylko 10^8 cm^{-3}].

Autor uważa, że ta wysoka oporność jest wynikiem domieszki akceptorowej (o głęboko położonym poziomie akceptorowym) kompensującej domieszki donorowe [53]. Po zredukowaniu ilości donorów leżących poniżej tej domieszki akceptorów, wszystkie donory tracą swoje elektrony na rzecz głębokich poziomów akceptorowych chociaż bardzo mała część tych elektronów jest wzbudzona z powrotem do pasma przewodnictwa. Wysoko oprowy GaAs może być wytwarzany przez domieszkowanie materiału typu *n* tlenem, kobaltem, lub niklem i sądzi się że zwykle tlen jest domieszką odpowiedzialną za ten efekt. Stwierdzono dalej, że Fe jest domieszką akceptorową i przez pomiar zmiany oporności z temperaturą oceniono energię jonizacji na 0,37 eV.

GaAs wytworzony bez rozmyślnego domieszkowania zwykle ma w temperaturze pokojowej oporność mniejszą niż $10 \Omega \text{ cm}$, a przypadkowa sztabka ma czasem bardzo wysoką oporność, większą niż $10^6 \Omega \text{ cm}$. Wydaje się, że izolowane donory w GaAs posiadają określoną energię jonizacji, ale że przy koncentracjach większych niż $\sim 5 \cdot 10^{17} \text{ cm}^{-3}$ pasmo domieszkowe utworzone oddziaływaniem między donorami zachodzi na pasmo przewodnictwa.

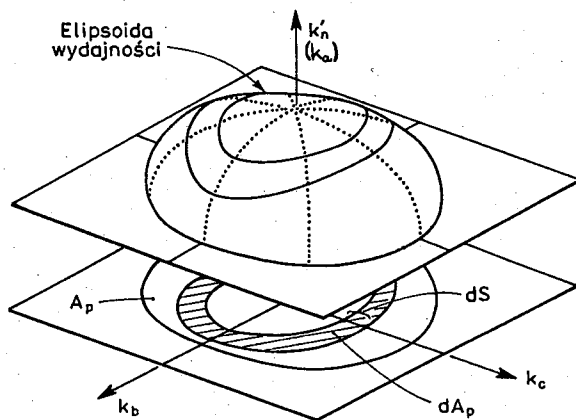
Pierwiastki grupy IV, jak: Si, Ge, Sn i Pb działają jako donory w GaAs, ale są one także w stanie wchodzić do sieci w postaci elektrycznie obojętnej, aż do wysokiej koncentracji. W tym przypadku prawdopodobnie wchodzi do sieci parami zastępując sąsiadujące atomy Ga i As.

Najefektywniejszymi donorami w GaAs są pierwiastki VI grupy: selen i tellur. Można je wprowadzać w ilości około 10^{19} cm^{-3} . Pierwiastki te w sieci GaAs zajmują pozycje arsenu.

Edmond [52] znalazł z pomiarów współczynnika Halla, że (Si i O_2 są głównymi domieszkami GaAs przedostającymi się do niego w czasie krystalizacji) maksymalna ilość donorów jaką można wprowadzić przez domieszkowanie GaAs krzemem wynosi około $4 \cdot 10^{18} \text{ cm}^{-3}$. Z pomiarów promieniami X wydedukowano, że Si może wejść do sieci GaAs w postaci elektrycznie naturalnej w dużej koncentracji do około 10^{20} cm^{-3} . Wyniki te można wyjaśnić w następujący sposób. Atomy Si mogą wchodzić do sieci na dwa sposoby: Przy niskich koncentracjach zastępują one atomy Ga i dają poziom donorowy, zaś przy koncentracji n większej niż $2 \cdot 10^{18} \text{ cm}^{-3}$ wchodzi one parami i dają centra neutralne.

Pomiary Rhoderick'a nad jądrowym rezonansem magnetycznym Ga w próbkach GaAs domieszkowanych Si potwierdziły, że Si wchodzi do sieci krystalicznej na te dwa sposoby.

Broom, Barrie i Ross [53] stwierdzili w GaAs typu n o koncentracji elektronów $n_e = 8 \cdot 10^{16} \text{ cm}^{-3}$ zachowanie typowe dla przewodnictwa pasma domieszkowego. Współczynnik Halla R przechodzi przez maksimum przy $\sim 70 \text{ K}$, a poniżej $\sim 10 \text{ K}$ współczynnik ten i przewodność σ były zgrubsza niezależne od temperatury. Z tych wyników przerwa energetyczna między pasmem domieszkowym, a pasmem przewodnictwa wynosi około $0,003 \text{ eV}$.



Rys. 15. Elipsoida wydajności dla trój-wymiarowych pasm, ilustrująca konstrukcję obliczonych krzywych rozkładu energetycznego fotoelektronów. Eliptyczny obszar dA_p jest rzutem na płaszczyznę $k_y k_z$ elementu elipsoidy wydajności wyciętego przez dwie koncentryczne kule, które są powierzchniami stałych energii elektronów w próżni. Przyjmuje się, że normalna do powierzchni $\hat{k}'_n$ leży wzdłuż $\hat{k}_n$.

Dla próbek typu n o koncentracji elektronów większej niż $5 \cdot 10^{17} \text{ cm}^{-3}$ stwierdzono, że współczynnik Halla R i przewodność σ pozostają praktycznie stałe od 300 K do $1,5 \text{ K}$, wykazując, że przy tej n_e pasmo domieszek donorowych zachodzi na pasmo przewodnictwa.

Z pośród atomów domieszkowych grupy II działających jako akceptory w GaAs, Zn jest najefektywniejszy, ponieważ jest rozpuszczalny do około 10^{20} cm^{-3} . Stwierdzono, że poziom akceptorowy Zn leży około 0,08 eV powyżej szczytu pasma walencyjnego. Cu zaś wprowadza poziom akceptorowy leżący 0,15 eV powyżej pasma walencyjnego (oprócz pewnych poziomów płytszych). Miedź z I grupy działająca w GaAs jako akceptor jest ważną domieszką akceptorową z powodu szybkiej dyfuzji poprzez kryształ. Cu dyfunduje procesem międzywęzłowym, ale jest także wprowadzana do sieci w pozycje węzłowe zastępując Ga, który jest większym z tych dwu atomów. Fuller i Whelan [56] znaleźli, że po dyfuzji Cu do GaAs wzrost koncentracji dziur p był dwu- lub trzykrotnie większy od koncentracji Cu. To oznacza, że Cu posiada co najmniej dwa płytke poziomy akceptorowe w GaAs.

Mg zachowuje się jak akceptor w GaAs przy koncentracjach wyższych od $5 \cdot 10^{18} \text{ cm}^{-3}$, natomiast przy koncentracjach niższych jak donor. Działanie donorowe mogło wystąpić jeżeli Mg wszedł do sieci międzywęzłowo [57].

Niektóre z domieszek, na przykład tlen, kobalt, nikiel tworzą, jak wiemy już, tak zwane głębokie poziomy akceptorowe położone w pobliżu środka przerwy energetycznej. W przypadku, gdy oprócz występowania głębokich poziomów akceptorowych jednocześnie i liczba donorów jest bardzo mała, na przykład rzędu 10^{16} cm^{-3} , to może jak wiemy, nastąpić kompensacja donorów przez akceptory, w wyniku czego rezystancja materiału wzrasta do wartości megomów; powstaje więc wysokooporowy GaAs, który nazywany jest często materiałem półizolującym [56].

Najefektywniejszymi akceptorami GaAs są więc pierwiastki II grupy: Zn i Cd. Można je wprowadzić do próbki w ilości do $\sim 10^{20} \text{ cm}^{-3}$. Wchodzą one do sieci GaAs na miejsce atomów Ga.

BIBLOGRAFIA

1. C. Kittel: *Wstęp do fizyki ciała stałego*. PWN. Warszawa 1974
2. Z. Trzaska, Durski: *Sieć odwrotna i interpretacja retigramów*. Wydawnictwa Politechniki Warszawskiej, Warszawa 1980
3. F. Herman: J.Elektron. 1, 103 (1955)
4. J. Callaway: *Energy bands in gallium arsenide*. J.Electron. 2, 330 (1957)
5. D. Brust, M.L. Cohen, J.C. Phillips: Phys. Rev. Lett. 9, 389 (1962);
W.E. Spitzer, R.E. Simon, Phys. Rev. Lett. 9, 387 (1963)
6. C. Hilsum, R. Barrie: Proc. Phys. Soc., 71, 676 (1958)
7. E.O. Kane: J.Phys.Chem.Solids 1, 249 (1957) *Band structure of indium antimonide*.
J.Phys.Chem.Solids 1, 249–261, Pergamon Press, New York 1957; J.Phys.Chem.Solids 1/2, 82 (1956)
8. F. Bassani, M. Yoshimine: Phys.Rev. 130, 20 (1963)
9. F. Bassani: *Physics of III–V Compounds*, I, Academic Press, Inc., New York 1966
10. J.C. Phillips: Solid State Phys. 18, 78 (1966)
11. W.A. Harrison: *Teoria ciała stałego*. Warszawa 1976
12. T.O. Woodruff: Phys.Rev. 103, 1159 (1966)
F. Herman: Phys.Rev. 93, 1214 (1955)
13. F. Herman, S. Skillman: *Atomic Structures Calculations*, Prentice–Hall, Inc., Englewood Cliffs, New York 1963
14. F. Herman, R.L. Kortum, C.D. Kuglin, J.P. van Dyke, S. Skillman: *Methods in Computational Physics*. Ed. B. Alden, S. Fernbach,

15. M. Rotenberg, 8, Academic Press, Inc., New York 1968
16. W. Brinkman, B. Goodman: Phys.Rev. 149, 597 (1966)
17. J. Robinson, F. Bassani, R.S. Knox, J.R. Schrieffer: Phys.Rev. Letters 9, 215 (1962)
F. Bassani, J. Robinson, B. Goodman, J.R. Schrieffer: Phys.Rev. 127, 1969 (1962)
18. F. Herman, C.D. Kuglin, R.L. Kortum: Proc.Symp.Energy Bands in Metals and Alloys, AIME, September 1967
F. Herman, R.L. Kortum, C.D. Kuglin, R.A. Short: J.Phys.Soc.Japan 21S, 7 (1966)
19. Gallium Aesnide and Related Compounds, Ed. C.M. Wolfe, Conference Series Number 45, Bristol and London 1978
20. W.A. Harrison: *Teoria ciala stalego*, PWN, Warszawa 1976
21. M.L. Cohen, T.K. Bergstresser: Phys.Rev. 141, 789 (1966)
22. W. Saslow, T.K. Bergstresser, C.K. Fong, M.L. Cohen, D. Brust: *Solid State Commun.* 5, 667 (1967)
23. H.R. Philipp, H. Ehrenreich: Phys. Rev. 129, 1550 (1963)
24. E.O. Kane: J.Phys.Chem.Solids 1, 83 (1956)
25. M. Cardona, F.H. Pollak: Phys.Rev.142, 530 (1966)
26. A.G. Thompson, J.C. Wooley, M. Rubenstein: Canad.J.Phys. 44, 2927 (1966)
27. F.H. Pollak, M. Cardona: J.Phys.Chem.Solids 27,423 (1966);
V.L. Berkovits, L.F. Ivantsov, I.V. Makarenko, T.A. Minashvili, V.I. Safarov: Sov.Phys.—Semicond. (USA), 21, 270 (1987)
28. F.H. Pollak, C.W. Higginbottam, M. Cardona, Proc.Internat.Conf. Phys. Semicond., Kyoto, 1966, p.20
29. A.G. Thompson, M. Cardona, K. Shaklee, J.C. Woolley: Phys.Rev.146, 601 (1966)
30. M. Cardona, F.H. Pollak, J.G. Broerman: Phys. Letters (Netherlands) 19, 276 (1965)
31. B.O. Seraphin: Proc.Phys.Soc. 87,239 (1966)
32. G.W. Gobeli, F.G. Allen: Phys.Rev. 127, 131 (1962); Phys.Rev. 137, A245 (1965)
33. R.C. Eden, J.C. Moll, W.E. Spicer: Phys.Rev. Letters 18, 597 (1967)
34. M.L. Cohen, J.C. Phillips: Phys.Rev. 139, 912 (1965)
35. A. Marcus Gray: Phys.Stat.Sol., 37, 11 (1970)
36. L.F. Eastman: *Gallium Arsenide and Related Compounds*, (St.Louis), Bristol and London 1976
37. R. Braunstein: Intervalence band transitions in gallium arsenide, J.Phys.Chem.Solids, 8, 280—282 (1959)
38. R. Braunstein, L. Magid: Phys.Rev. 111, 480 (1958)
39. A.H. Kahn: Phys.Rev. 97, 1947 (1965)
40. E.O. Kane: J.Phys. Chem. Solids 1, 83 (1956)
41. W.E. Spitzer, R.C. Eden: *Photoemission investigations of the band structure of semiconductors*. Proc. 9th Conference on physics of Semiconductors, Moscow, 1968, Leningrad, USSR: nauka
42. R.H. Parmenter: Phys.Rev. 100, 573 (1965)
43. N.E. Christensen, M. Cardona: Solid State Commun. (USA), 51, 491 (1984)
44. W.G. Spitzer, J.W. Whelan: Phys.Rev. 114, 59 (1959)
45. A. Sagar: *Westinghouse Research report*, 6—40602—3—R1
46. C. Hilsun: Proc.Phys.Soc. London 76, 414 (1960)
47. A.L. Edwards, T.E. Slykhouse, H. Drickamer: J.Phys.Chem. Solids 11, 140 (1959)
48. H. Ehrenreich: *Band and structure and electron transport of GaAs*. Phys.Rev. 120, 1951 (1960)
49. T.S. Moss, A.K. Walton: Proc. Phys. Soc., London, 74, 131, (1959)
50. E.D. Palik a. oth.: Phys. Rev. 124, 701, (1961)
51. J.M. Whelan, G.H. Wheatley: J.Phys.Chem.Solids 6, 169 (1968)
52. Y. Bor — Yam, J.D. Joannopoulos: Phys.Rev.B (USA), 30 1984 (1984);
J.W. Allen: Nature, London, 187, 403 (1960);
J. Blanc, L.R. Weisberg: Nature, Lond. 192, 155 (1961)

53. J. W o j a s: *Badania fotoelektrycznych i powierzchniowych własności monokryształów GaAs o termicznie czyszczonych powierzchniach*. Wyższa Szkoła Pedagogiczna, Olsztyn 1976
54. J.T. E d m o n d, R.F. B r o o m, F.A. C u n n e l l: *Report on Meeting on Semiconductors at Rugby*. The Physical Society, London, 1956, p.109
55. R.F. B r o o m: *Physics Lett.* 4, 330 (1963)
56. C.S. F u l l e r i n n i: *J. Electrochem. Soc.* 109, 880 (9162)
57. D. B o l m o n t, P. C h e n, C.A. S e b e n n e: *Electronic properties of Ga/GaAs (110) upon interface formation*. *Surface Science*, 117, 417 (1982)
58. *Handbook on Semiconductors*, Ed.Series T.S. Moss, Vol.1. *Band Theory and Transport Properties*. Volume Editor W. Paul, North—Holland, 1982

J. WOJAS

ANALYSIS OF ENERGETIC STRUCTURE AND INTERBAND TRANSITIONS OF GaAs

S u m m a r y

First part of the work treats of disturbance methods of crystalline symmetry which are used to lead energy bands of GaAs out of energy bands of Ge. Symmetry of states in conduction bands and valency bands, splitting of states and their joint during disturbance were discussed. Equations of conduction band energy were described and forms of energetic bands of GaAs were discussed. Principles of calculation of GaAs bands structure were discussed. In recapitulations of results of structure calculation energies of peaks and minimums in reflection spectrum answering proper electron transitions, were shown in detail. Energetic structures of valency band, conduction band, forbidden gap and impurity states were precisely analysed.

Modelling of photoacoustic Helmholtz resonators by means of acousto-electrical analogies

TOMASZ STARECKI

Instytut Podstaw Elektroniki, Politechnika Warszawska

Received 1992.11.06

Authorized 1993.01.22

Photoacoustic cells consisting of two chambers connected by a narrow tube are studied. Theoretical explanation of the resonance curves of the cells is based on the electrical analogies. Amplitude and phase angle of the photoacoustic signal are expressed as a function of chopping frequency, cell dimensions and buffer gas parameters. Theoretical results are compared with experimental values.

1. INTRODUCTION

One of the photoacoustic cells, which is suitable especially for spectroscopy of solids, is a Helmholtz resonator. For the first time it was applied in 1976 by McClelland and Kniseley [1]. Since then a lot of applications of Helmholtz resonator in photoacoustics have been reported [2-7]. However in most of these works calculated parameters of the cells were different from experimental values. One of the methods that can be used for theoretical calculations is modelling by means of electrical analogs. The method was applied by a few authors, but only Nordhaus and Peltz obtained good agreement with experiments [6, 7]. However they analyzed cases in which channel length was $200 \div 600$ mm, while in most practical solutions the channel is $10 \div 100$ times shorter.

In this paper it is considered if the method of acousto-electrical analogies is useful for precise determination of the resonance curves of a Helmholtz resonator with a short channel.

2. LUMPED ELEMENTS MODEL

Helmholtz resonator consists of two volumes connected by a narrow tube (Fig.1a). In a simple model of a such resonator the cavities and interconnecting duct are represented by lumped constant elements such as capacitors, resistor and

inductor (Fig.1b). Intensity of the light beam is described by the current source J . Pressure values correspond to the voltages across the capacitors. The only difficulty which occurs while using the method is that values of the elements of the equivalent circuit are defined in different way by different authors. In this paper three different cases are considered. The first definition is given by Morse [8] and Malecki [9, 10]:

$$R = \delta w^2 / 2\pi v; \quad L = \delta(1 + 1.6a)/S; \quad C_i = V_i / (\delta v^2). \quad (1)$$

Nolle [11] defines the elements as follows:

$$R = 8\pi\mu l / S^2; \quad L = 4\delta l / 35; \quad C_i = V_i / (\delta v^2). \quad (2)$$

Peltz and Nordhaus [6, 7] describe the elements as:

$$R = 8\pi\mu l / S^2; \quad L = \delta l / S; \quad C_i = V_i / (\delta v^2). \quad (3)$$

In the above equations channel length, radius and crosssectional area are denoted respectively by symbols l , a and S . Volumes of the cavities are described as V_i . Properties of the transmitting gas are taken into consideration by the viscosity μ , mass density δ and free-space sound velocity v . Symbol w denotes pulsation of the light beam modulation.

3. TRANSMISSION LINE MODEL

When the sound wavelength is considerably greater than the channel length the lumped elements model can be used. But when the length of the interconnecting duct increases, sound propagation in the tube becomes important. Therefore the elements in the model can be considered as homogeneously distributed. In such model the duct is represented by a transmission line, which characteristic impedance Z_w and propagation constant Γ_w are defined as follows [6, 7, 12]:

$$Z_w = (Z/Y)^{1/2}; \quad \Gamma_w = (Z \cdot Y)^{1/2}, \quad \text{where:} \quad (4)$$

$$Z = jw\delta / (S(1 - F_v \exp(j\Phi_v))); \quad (5)$$

$$Y = jwS(1 + (\sigma - 1)F_t \exp(j\Phi_t)) / \delta v^2. \quad (6)$$

Symbol σ , ($\sigma = C_p/C_v$) denotes an adiabate coefficient of the gas. The parameters F and Φ appearing in (5) and (6) are given in equation (7):

$$F \cdot \exp(j\Phi) = 2 \cdot J_1((-j)^{1/2}r) / ((-j)^{1/2}r \cdot J_0((-j)^{1/2}r)), \quad (7)$$

where J_0 and J_1 are Bessel functions of zero and first order. Indexes v and t in equations (5) and (6) suggest which variable r_v or r_t should be used instead of symbol r in expression (7). Both variables r_v and r_t are described below:

$$r_v = a(w\delta/\mu)^{1/2}; \quad r_t = a(w\delta c_p/a)^{1/2}. \quad (8)$$

The variable r_v determines viscous properties of the gas. Thermal properties of the gas, described by r_t , are taken into account by the mass density δ , heat capacity c_p and thermal conductivity a .

The whole circuit in which the transmission line is replaced by the equivalent T-section [13] is shown in Fig.1c.

4. EXPERIMENTS

Theoretical resonance curves obtained from the discussed models were verified by experimental data. A sketch of the cell used in the experiments is shown in Fig.1a. The cavity in which a sample (carbon black) was placed had a volume of 2.0 cm^3 .

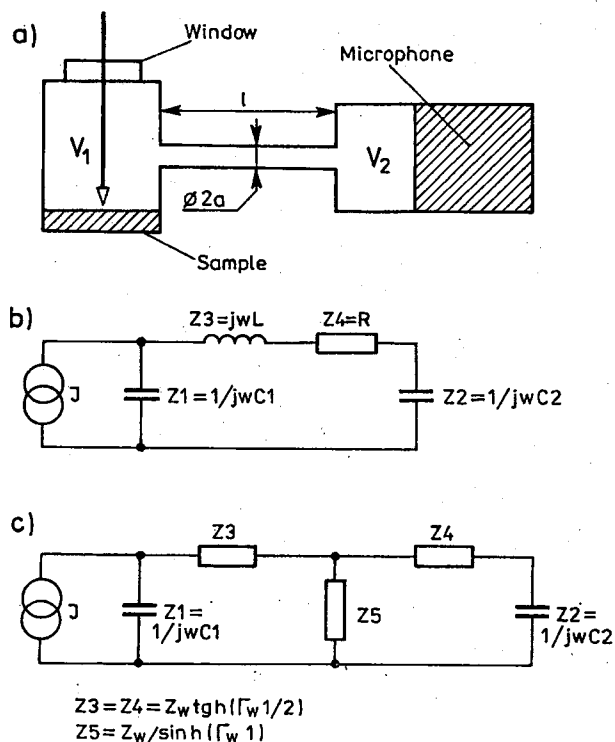


Fig. 1. Helmholtz resonator: a. Sketch of the cell, b. Lumped elements model of the cell, c. Transmission line model of the cell

A volume of the second chamber was being changed in the range of $1.13 \div 2.26 \text{ cm}^3$ by means of distance sleeves. The cavities were connected by a tube 15 mm long and 1.1 mm in diameter. All the measurements were performed in an automatic measurement system [14]. The resonance curves obtained in the measurements and the theoretical results calculated from the discussed models are shown in Fig.2÷4.

It can be seen that the experimental values are in good agreement with the results obtained from the transmission line model. Other models give worse

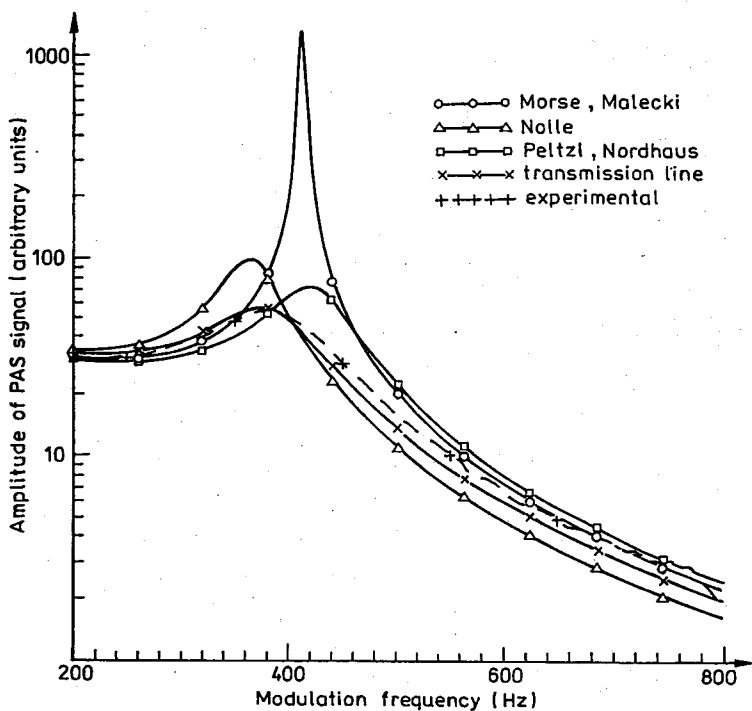


Fig. 2. Comparison of theoretical and experimental Helmholtz resonance curves. Microphone cavity volume $V_2 = 2.26 \text{ cm}^3$

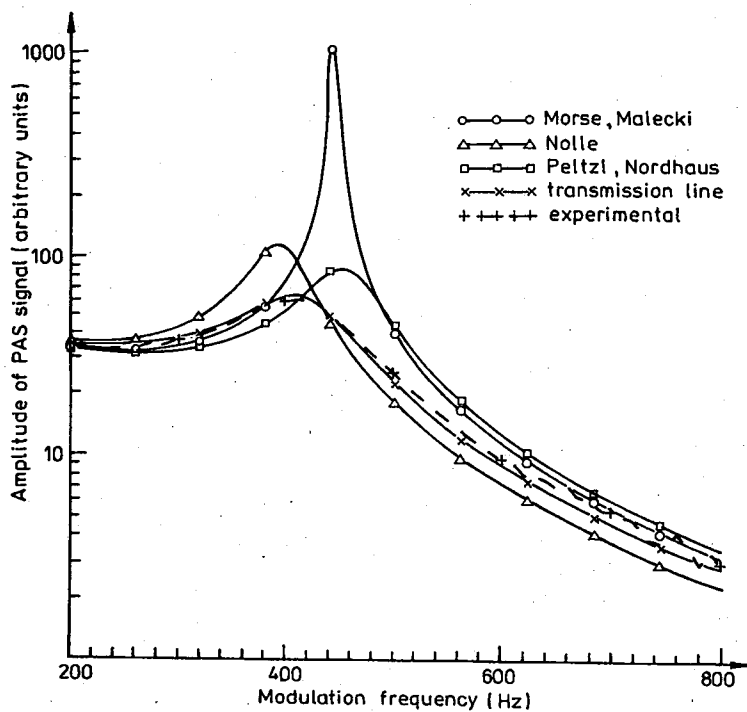


Fig. 3. Comparison of theoretical and experimental Helmholtz resonance curves. Microphone cavity volume $V_2 = 1.70 \text{ cm}^3$

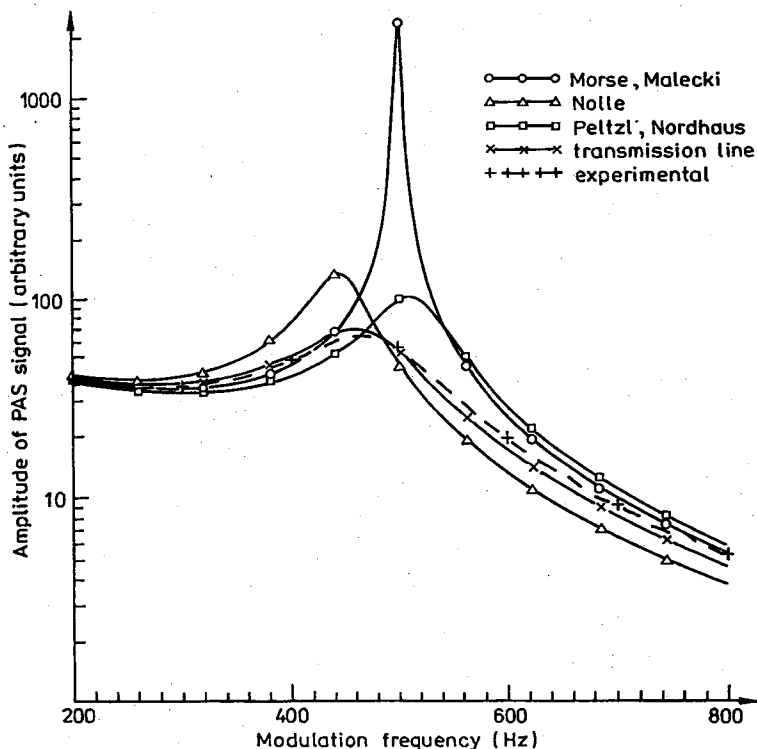


Fig. 4. Comparison of theoretical and experimental Helmholtz resonance curves. Microphone cavity volume $V_2 = 1,13 \text{ cm}^3$

approximation. The worst evaluation is obtained from the equations given by Morse.

CONCLUSIONS

Preliminary results show that acousto-electrical analogies can be used in modelling of photoacoustic Helmholtz resonators. The best fit to the experimental data was obtained from the transmission line model. However in the experiments carried out only one factor (the volume of the microphone chamber) was being changed, so that no general conclusions should be drawn yet. Further work in the area is in progress.

REFERENCES

1. J.F. McClelland, R.N. Kniseley: *Scattered light effects in photoacoustic spectroscopy* Appl. Opt., 1976, vol.15, p.2967÷2968
2. R.S. Quimby, P.M. Selzer, W.M. Yen: *Photoacoustic cell design: resonant enhancement and background signals*. Appl. Opt., 1977, vol.16, p.2630÷2632

3. N.C. Fernelius: *Helmholtz resonance effect in photoacoustic cells*. Appl. Opt. 1979, vol.18, p.1784÷1787
4. H. Coufal, U. Möller, S. Schneider: *Photoacoustic cells for measurement at various temperatures and pressures*. Proceedings of the First International Conference on the Photoacoustic Effect in Germany, 1981
5. W.A. McClenny, C.A. Bennet, G.M. Russwurm, R. Richmond: *Helmholtz resonator enhancement of photoacoustic signals*. Appl. Opt., 1981, vol.20, p.650÷653
6. O. Nordhaus, J. Pelzl: *Frequency dependence of resonant photoacoustic cells*. Appl. Phys., 1981, vol.25, p.221÷229
7. J. Pelzl, K. Klein, O. Nordhaus: *Extended Helmholtz resonator in low-temperature photoacoustic spectroscopy*. Appl. Opt., 1982, vol.21, p.94÷99
8. P.M. Morse: *Vibration and Sound*. McGraw-Hill, New York 1948, p.234÷235
9. I. Malecki: *Teoria fal i układów akustycznych*. PWN, Warszawa 1964, p.367÷375
10. I. Malecki: *Analogie elektro-mechano-akustyczne*. WPW, Warszawa 1981, p.19÷23
11. A.W. Nolle: *Small-signal impedance of short tubes*. J. Acoust. Soc. Am., 1953, vol.25, p.32÷39
12. A.H. Benade: *On the propagation of sound waves in a cylindrical conduit*. J. Acoust. Soc. Am., 1968, vol.44, p.616÷623
13. F.B. Daniels: *On the propagation of sound waves in a cylindrical conduit*. J. Acoust. Soc. Am., 1950, vol.22, p.563÷564
14. T. Starecki: *Concept of virtual instruments applied in photoacoustic measurements*, submitted to Rev. Sci. Instrum.

T. STARECKI

MODELOWANIE REZONANSOWYCH KOMÓR FOTOAKUSTYCZNYCH W UKŁADZIE HELMHOLTZA ZA POMOCĄ ANALOGII AKUSTO-ELEKTRYCZNYCH

Streszczenie

W artykule analizowane są komory fotoakustyczne powstałe z połączenia dwóch wnęk kanałem niewielkiej średnicy. Analiza teoretyczna krzywych rezonansowych komór prowadzona jest w oparciu o analogie akusto-elektryczne. Amplituda i faza sygnału fotoakustycznego przedstawiona jest w funkcji częstotliwości modulacji, wymiarów komory oraz parametrów wypełniającego ją gazu. Obliczenia teoretyczne porównane są z rezultatami pomiarów.

Praca układu impulsowego do zmniejszania zniekształceń prądu sieciowego ze zmienną częstotliwością kluczkowania

MAREK GOTFRYD

Wydział Elektryczny, Politechnika Rzeszowska

Otrzymano 1992.05.23

Autoryzowano do druku 1993.10.25

W artykule przedstawiono analizę układu impulsowego, pracującego ze zmienną częstotliwością kluczkowania, przeznaczonego do zmniejszania zniekształceń prądu sieciowego. Rozważany sposób pracy układu odpowiada uzyskiwanemu przy sterowaniu go z pomocą układu scalonego TDA 4814. W pracy wyprowadzono zależność na wartość napięcia wyjściowego układu oraz obliczono straty mocy w kluczu tranzystorowym i cewce indukcyjnej. Na tej podstawie dokonano porównania właściwości układu uzyskiwanych przy jego pracy ze zmienną i stałą częstotliwością kluczkowania.

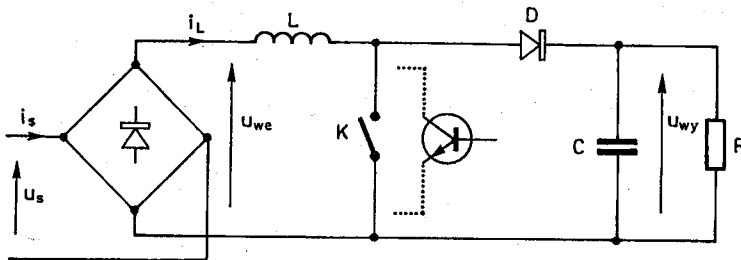
1. WSTĘP

Urządzenia elektryczne zawierające klasyczne, beztransformatorowe prostowniki sieciowe z obciążeniem pojemnościowym pobierają z sieci prąd o znacznie odkształconym przebiegu. Pogarsza to współczynnik mocy ($\cos \varphi$) tych urządzeń. W celu jego poprawy, tj w celu eliminacji harmonicznych prądu sieciowego stosuje się m.in. układy impulsowe włączane pomiędzy prostownik i kondensator obciążenia. Szczegółowa analiza pracy dwóch takich układów przedstawiona została wcześniej [1,2]. Pokazano wtedy, że jeden z nich, oparty na układzie impulsowego regulatora podwyższającego napięcie wejściowe, odznacza się wyraźnie większą sprawnością energetyczną od drugiego. Fakt ten znalazł potwierdzenie w praktyce aplikacyjnej — produkowany specjalizowany układ scalony TDA 4814 (SIEMENS) przeznaczony jest do sterowania pracą właśnie tego układu. Niestety, sterowanie to przebiega inaczej niż założono w [1]. Założono tam mianowicie, że częstotliwość kluczkowania układu pozostaje stała, a zmienia się czas załączenia klucza (tranzystorowego) w poszczególnych okresach kluczkowania, natomiast układ TDA 4814 realizuje funkcję przeciwną — czas załączenia klucza w poszczególnych okresach pozostaje stały, a zmienna jest częstotliwość kluczkowania. Z tego względu wyniki analizy zawartej w [1, 2] nie mogą znaleźć ścisłego zastosowania do oceny właściwości układu eliminacji zniekształceń prądu sieciowego z ww układem skalonym. Dlatego celem niniejszej pracy jest przeprowadzenie analizy działania tego układu z uwzględ-

nieniem odmiennego sposobu jego sterowania oraz — na jej podstawie — porównanie jego właściwości z właściwościami układu zanalizowanego wcześniej. Ze względu na — mimo wszystko — podobne sposoby działania układu dla obydwu sposobów sterowania, w toku poniższej analizy będą częściowo wykorzystywane rezultaty analizy poprzedniej [1, 2].

2. ANALIZA OGÓLNA

Schemat rozważanego układu w postaci uproszczonej przedstawiono na rys.1. Jak wiadomo, działanie takiego układu polega na pobieraniu energii z prostownika i gromadzeniu jej w cewce, gdy klucz K pozostaje zwarty oraz na oddawaniu jej do



Rys. 1. Schemat ideowy układu

obciążenia po rozwarciu klucza. Aby układ mógł działać poprawnie, konieczne jest aby napięcie wyjściowe zawsze przewyższało napięcie wejściowe oraz aby w końcu każdego okresu kluczkowania — tuż przed kolejnym zwarciem klucza — prąd w cewce był równy zero. Załóżmy wstępnie bezstratność układu tj. zerowe spadki napięć na przewodzących diodach i zwartym kluczu oraz zerową rezystancję cewki. Jako wejściowe można przyjąć napięcie wyjściowe prostownika tj.

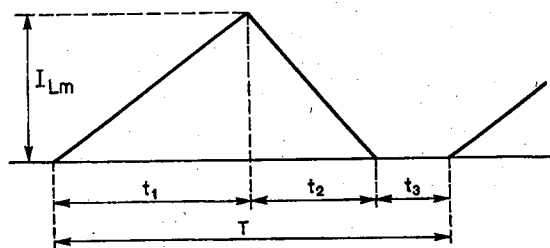
$$u_{we}(t) = U_{wcn} |\sin(\omega_s t)|, \quad (1)$$

gdzie U_{wcn} jest amplitudą napięcia sieciowego, ω_s jest jego pulsacją. Ograniczając rozważania do pierwszego półokresu napięcia sieciowego, znak wartości bezwzględnej w (1) można oczywiście pominąć. Dla łatwej realizacji układu konieczne jest, aby częstotliwość kluczkowania była wielokrotnie większa od częstotliwości sieci zasilającej. W tym przypadku napięcia $u_{we}(t)$ i $u_{wy}(t)$ można potraktować jako wolnozmiennne funkcje czasu w przeciągu każdego z okresów kluczkowania i przyjąć je wtedy jako (prawie) stałe. Wartości tych napięć oznaczane będą wtedy przez U_{wcn} i U_{wyn} , gdzie n oznacza n -ty okres kluczkowania w półokresie sieci. Powtarzając rozważania podane w [1] można stwierdzić, że prąd i_L będzie miał postać ciągu trójkątnych impulsów, tak jak to pokazano na rys.2. Ich parametry będą następujące [1]:

$$I_{Lmn} = \frac{U_{wcn}}{L} \cdot t_{in}, \quad (2)$$

$$t_{2n} = t_{1n} \cdot \frac{U_{wen}}{U_{wyn} - U_{wen}}, \quad (3)$$

$$T_n = t_{1n} + t_{2n} + t_{3n}. \quad (4)$$



Rys. 2. Przebieg prądu $i_L(t)$ w jednym z okresów kluczowania

W powyższych zależnościach t_1 jest czasem zwarcia klucza, t_2 — czasem zanikania prądu, $t_2 + t_3$ stanowi czas rozwarcia klucza; wszystkie te parametry dotyczą n -tego okresu kluczowania o długości T_n .

Przebieg wolnozmiennnej składowej prądu i_L będzie przebiegiem wartości średniej tych impulsów, obliczanej w każdym okresie T_n . Łatwo pokazać, że przebieg ten będzie się wyrażał zależnością [1]:

$$i_L(t) = \frac{u_{we}(t)}{2 \cdot L \cdot T} \cdot \frac{t_1^2}{1 - \frac{u_{we}(t)}{u_{wy}(t)}}. \quad (5)$$

Prąd sieciowy $i_s(t)$ będzie miał przebieg sinusoidalny, gdy prąd $i_L(t)$ będzie proporcjonalny do napięcia wejściowego. Otrzymamy stąd warunek

$$\frac{1}{T} \cdot \frac{t_1^2}{1 - \frac{u_{we}(t)}{u_{wy}(t)}} = const. \quad (6)$$

Można tu zauważyć dwie możliwości pracy układu:

$$1) T = const. \quad i \quad \frac{t_1^2}{1 - \frac{u_{we}(t)}{u_{wy}(t)}} = const.$$

$$\text{czyli} \quad t_1 \sim \sqrt{1 - \frac{u_{we}(t)}{u_{wy}(t)}} \quad (7)$$

$$2) t_1 = const. \quad i \quad \frac{1}{T \left(1 - \frac{u_{we}(t)}{u_{wy}(t)} \right)} = const.$$

$$\text{czyli } T \sim \frac{1}{1 - \frac{u_{we}(t)}{u_{wy}(t)}}.$$

Przypadek pierwszy został zanalizowany w pracach [1, 2], natomiast drugi — odpowiadający sposobowi pracy układu TDA 4814 — zostanie rozważony poniżej.

Na podstawie (4) i (8) można napisać

$$T_{\min} = T(u_{we} = 0) = t_1 + t_2(u_{we} = 0) + t_3(u_{we} = 0). \quad (9)$$

Zauważmy, że $t_2(u_{we} = 0) = 0$ (zal.(2)). W związku z tym zależność (9) można przedstawić w postaci

$$T_{\min} = t_1 + t_{30}, \quad (10)$$

gdzie $t_{30} = t_3(u_{we} = 0)$. Podstawiając (10) do (8) otrzymamy

$$T = \frac{T_{\min}}{1 - \frac{u_{we}(t)}{u_{wy}(t)}} = \frac{t_1 + t_{30}}{1 - \frac{u_{we}(t)}{u_{wy}(t)}}. \quad (11)$$

Porównując zależności (11) i (4), z uwzględnieniem zależności (3), uzyskamy

$$t_3 = \frac{t_{30}}{1 - \frac{u_{we}(t)}{u_{wy}(t)}}. \quad (12)$$

Podsumowując, aby — przy rozważanym sposobie sterowania pracą układu — prąd sieciowy był sinusoidalny potrzeba, by czasy t_1 , t_2 , t_3 w kolejnych okresach kluczkowania były równe:

$$t_1 = \text{const}, \quad t_2 = t_1 \cdot \frac{1}{\frac{u_{wy}(t)}{u_{we}(t)} - 1}, \quad t_3 = t_{30} \cdot \frac{1}{1 - \frac{u_{we}(t)}{u_{wy}(t)}}. \quad (13)$$

Spełnienie dwóch pierwszych zależności (13) w praktycznym układzie jest łatwe (druga jest spełniana automatycznie; wynika to z konfiguracji układu), natomiast spełnienie trzeciej może być trudniejsze, szczególnie w układzie samooscylicującym. Wymaga to bowiem uzależnienia przerw między impulsami (rys.2) od chwilowych wartości napięć U_{we} i U_{wy} . Ułatwieniem praktycznej realizacji układu będzie więc przyjęcie po prostu $t_{30} = 0$, co da również $t_3 = 0$. Wtedy impulsy prądu i_L będą następować bezpośrednio po sobie, bez przerw między nimi; w ten właśnie sposób pracuje układ TDA 4814 w układzie samooscylicującym, co wynika z jego schematu blokowego i aplikacyjnego [3]. Dla zachowania ogólności rozważań, dalsza analiza pracy i właściwości układu będzie prowadzona przy założeniu, że $t_{30} \geq 0$.

Aby otrzymać wyrażenie określające wartość napięcia wyjściowego układu, należy rozważyć wartość ładunku doprowadzanego i odprowadzanego z kondensatora C. Ładunek doprowadzany do kondensatora w n -tym okresie kluczkowania jest równy

$$Q_{dn} = \frac{I_{Lmn} \cdot t_{2n}}{2} \quad (14)$$

Po podstawieniu (2) i (3) otrzyma się

$$Q_{dn} = \frac{t_1^2}{2L} \cdot \frac{U_{wen}^2}{U_{wyn} - U_{wen}} \quad (15)$$

Ładunek odprowadzony z kondensatora C w tym samym okresie kluczkowania jest równy

$$Q_{odn} = \frac{U_{wyn}}{R} \cdot T_n \quad (16)$$

Wypadkowa zmiana ładunku jest oczywiście różnicą wyrażeń (15) i (16), natomiast odpowiadająca temu zmiana napięcia wyjściowego będzie równa

$$\Delta U_{wyn} = \frac{1}{C} \cdot \left(\frac{t_1^2}{2L} \cdot \frac{U_{wen}^2}{U_{wyn} - U_{wen}} - \frac{U_{wyn}}{R} \cdot T_n \right) \quad (17)$$

Pierwszy składnik wyrażenia w nawiasie można przekształcić następująco:

$$\begin{aligned} \frac{t_1^2}{2L} \cdot \frac{U_{wen}^2}{U_{wyn} - U_{wen}} &= \frac{t_1}{2L} \cdot \frac{U_{wen}^2}{U_{wyn}} \cdot t_1 \cdot \frac{U_{wyn}}{U_{wyn} - U_{wen}} = \frac{t_1}{2L} \cdot \frac{U_{wen}^2}{U_{wyn}} \cdot (t_1 + t_{2n}) = \\ &= \frac{t_1}{2L} \cdot \frac{U_{wen}^2}{U_{wyn}} \cdot \frac{t_1 + t_{2n}}{T_n} \cdot T_n \end{aligned} \quad (18)$$

Wyrażenie $\frac{(t_1 + t_{2n})}{T_n}$ ma stałą wartość, niezależną od okresu kluczkowania; na podstawie (3), (4) i (12) mamy bowiem

$$\frac{t_1 + t_{2n}}{T_n} = \frac{t_1}{t_1 + t_{30}} \quad (19)$$

Podstawiając (19) do (18), a (18) do (17) otrzymamy

$$\Delta U_{wyn} = \frac{1}{C} \cdot \left(\frac{t_1}{2L} \cdot \frac{U_{wen}^2}{U_{wyn}} \cdot \frac{t_1}{t_1 + t_{30}} - \frac{U_{wyn}}{R} \right) T_n \quad (20)$$

Wyrażenie powyższe można także zapisać inaczej

$$\frac{\Delta U_{wyn}}{T_n} = \frac{1}{C} \cdot \left(\frac{t_1}{2L} \cdot \frac{U_{wen}^2}{U_{wyn}} \cdot \frac{t_1}{t_1 + t_{30}} - \frac{U_{wyn}}{R} \right) \quad (21)$$

Lewa strona tego wyrażenia przedstawia średnią wartość pochodnej napięcia wyjściowego w okresie kluczkowania, tj, w przedziale czasu bardzo małym w stosunku do okresu napięcia sieciowego. Dlatego wyrażenie (21) można potraktować jako równanie różniczkowe dla napięcia wyjściowego uśrednianego w każdym okresie kluczkowania, czyli jako równanie dla wolnozmiennnej składowej tego napięcia. Można zatem napisać

$$\frac{du_{wy}(t)}{dt} = \frac{1}{C} \cdot \left(\frac{t_1}{2L} \cdot \frac{u_{we}^2(t)}{u_{wy}(t)} \cdot \frac{t_1}{t_1 + t_{30}} - \frac{u_{wy}(t)}{R} \right). \quad (22)$$

Jest to równanie o następującej postaci ogólnej:

$$\frac{du_{wy}(t)}{dt} = f(t) \cdot \frac{1}{u_{wy}(t)} - a \cdot u_{wy}(t), \quad (23)$$

gdzie:

$$f(t) = \frac{t_1}{2LC} \cdot \frac{t_1}{t_1 + t_{30}} \cdot u_{we}^2(t) = \frac{t_1}{2LC} \cdot \frac{t_1}{t_1 + t_{30}} \cdot U_{wem}^2 \sin^2(\omega_s t); \quad (24)$$

$$a = \frac{1}{RC}.$$

Procedura rozwiązywania równania o identycznej postaci ogólnej została przedstawiona w [1]. Uwzględniając tylko nieco inną postać funkcji $f(t)$ w tym przypadku można podać, że rozwiązaniem równania (22) jest funkcja

$$u_{wy}(t) = \frac{U_{wem}}{2} \cdot \sqrt{\frac{R}{L} \cdot \frac{t_1}{1 + \frac{t_{30}}{t_1}}} \sqrt{1 - \frac{1}{\sqrt{1 + (\omega_s RC)^2}}} \cdot \sin\left(2\omega_s t + \arctg \frac{1}{\omega_s RC}\right). \quad (25)$$

Czasowa zmienność napięcia $u_{wy}(t)$ jest identyczna jak w przypadku układu analizowanego w [1] — dla funkcji sterowania kluczem zgodnej z (7). Można to uzasadnić następująco. Obydwa układy, z założenia, pobierają w jednakowy sposób (sinusoidalny) prąd z sieci, obydwie też — przy założeniu ich bezstratności — muszą w jednakowy sposób przekazywać pobieraną z sieci energię do obciążenia. Dlatego przebiegi czasowe napięcia wyjściowego w obydwu przypadkach muszą być jednakowe. Pierwsze dwa czynniki w wyrażeniu (25) przedstawiają wartość średnią (długookresową) napięcia wyjściowego

$$U_{wys} = \frac{U_{wem}}{2} \cdot \sqrt{\frac{R}{L} \cdot \frac{t_1}{1 + \frac{t_{30}}{t_1}}}. \quad (26)$$

Ze względu na monotoniczną zależność $U_{wys}(t_1)$, możliwa jest regulacja tego napięcia przez zmiany czasu t_1 . Przekształcając (26) można uzyskać też zależność odwrotną tj. $t_1(U_{wys})$. Jest ona nieco złożona, bowiem jest dana przez równanie kwadratowe

$$t_1^2 \cdot \frac{R}{L} - t_1 \cdot \frac{4U_{wys}^2}{U_{wem}^2} - \frac{4U_{wys}^2}{U_{wem}^2} \cdot t_{30} = 0. \quad (27)$$

3. ANALIZA ENERGETYCZNA

W celu pełniejszej oceny właściwości użytkowych układu konieczne jest rozważenie występujących w nim strat mocy i oszacowanie możliwej do uzyskania sprawności energetycznej. Tak jak poprzednio [2], przyjęto rozważenie strat

- w tranzystorze (jako kluczu), składających się ze strat na nasycenie i strat w czasie przełączania;
- w cewce, składających się ze strat w rdzeniu i w uzwojeniu.

3.1. STRATY W TRANZYSTORZE

Energia strat w czasie nasycenia tranzystora w jednym (n -tym) okresie kluczowania jest równa:

$$W_{1n} = \int_0^{t_1} i_{Ln}(t) \cdot U_{ces} dt, \quad (28)$$

gdzie $i_{Ln}(t)$ jest przebiegiem prądu $i_L(t)$ w n -tym okresie kluczowania, U_{ces} jest napięciem nasycenia tranzystora. Dla uproszczenia przyjęto tu, że jest ono stałe. Na podstawie zależności (2) można obliczyć, że

$$W_{1n} = \frac{U_{ces} U_{wen}}{2L} \cdot t_1^2. \quad (29)$$

Energia tracona na nasycenie tranzystora w czasie wszystkich okresów kluczowania w jednym półokresie napięcia sieciowego będzie odpowiednią sumą wyrażeń (29)

$$W_1 = \sum_{n=1}^N \frac{U_{ces}}{2L} \cdot U_{wen} t_1^2 = \frac{U_{ces}}{2L} \cdot t_1 \cdot \sum_{n=1}^N U_{wen} t_1, \quad (30)$$

gdzie N jest liczbą okresów kluczowania w półokresie sieci.

Wyrażenia pod zankiem sumy można przekształcić następująco:

$$U_{wen} t_1 = U_{wen} \cdot \frac{t_1}{T_n} \cdot T_n = U_{wen} \cdot \frac{t_1}{t_1 + t_{2n} + t_{3n}} \cdot T_n. \quad (31)$$

Podstawiając zależności (3) i (12) do (31) otrzymamy

$$U_{wen} t_1 = U_{wen} \cdot \frac{t_1}{t_1 + t_{30}} \cdot \left(1 - \frac{U_{wen}}{U_{wyn}}\right) \cdot T_n. \quad (32)$$

Zależność (30) po podstawieniu (32) przyjmie postać

$$W_1 = \frac{U_{ces}}{2L} \cdot \frac{t_1}{t_1 + t_{30}} \cdot \sum_{n=1}^N U_{wen} \cdot \left(1 - \frac{U_{wen}}{U_{wyn}}\right) \cdot T_n. \quad (33)$$

Suma w powyższym wyrażeniu przedstawia sumę pól prostokątów o podstawach T_n i wysokościach $U_{wen} \left(1 - \frac{U_{wen}}{U_{wyn}}\right)$. Podstawy tych prostokątów zapełniają odcinek o długości połowy okresu sieci ($1/2f_s$). Ze względu na to, że podstawy te są bardzo małe w porównaniu z półokresem sieci, wartość sumy można przybliżyć całką oznaczoną

$$\sum_{n=1}^N \left(1 - \frac{U_{wen}}{U_{wyn}}\right) \cdot T_n \cong \int_0^{1/2f_s} u_{we}(t) \left(1 - \frac{u_{we}(t)}{u_{wy}(t)}\right) dt. \quad (34)$$

Przybliżenie to jest tym lepsze, im częstotliwości kluczowania jest wyższa; w typowych warunkach błąd przybliżenia jest jest mniejszy od 1%, jest więc praktycznie pomijalny. Aby obliczyć wartość całki (34) należy podstawić do niej wyrażenia (1) i (25) na $u_{we}(t)$ i $u_{wy}(t)$. Otrzymamy wówczas

$$\int_0^{1/2f} u_{we}(t) \left(1 - \frac{u_{we}(t)}{u_{wy}(t)}\right) \cdot dt = \frac{U_{wem}}{\omega_s} \cdot \left(2 - \frac{U_{wem}}{U_{wys}} \cdot II\right). \quad (35)$$

gdzie:

$$II = \int_0^{\pi} \frac{\sin x}{g(x)} dx, \quad (36)$$

natomiast

$$g(x) = \sqrt{1 - \frac{1}{\sqrt{1 + (\omega_s RC)^2}} \sin\left(2x + \arctg \frac{1}{\omega_s RC}\right)}. \quad (37)$$

Obliczone numeryczne wartości całki II dla różnych wartości $\omega_s RC$ zostały podane w [2]; są one zbliżone do $\pi/2$: np. dla $\omega_s RC = 5$ $II = 1,567 \dots$. Po podstawieniu (35) do (34) i następnie (34) do (33) otrzymamy ostatecznie

$$W_1 = \frac{U_{ces} U_{wem}}{2\omega_s L} \cdot \frac{t_1}{1 + \frac{t_{30}}{t_1}} \cdot \left(2 - \frac{U_{wem}}{U_{wys}} \cdot II\right). \quad (38)$$

Średnia w półokresie sieci moc strat na nasycenie będzie równa

$$P_1 = W_1 \cdot 2f_s = \frac{U_{ces} U_{wem}}{\pi L} \cdot \frac{t_1}{1 + \frac{t_{30}}{t_1}} \left(1 - \frac{1}{2} \frac{U_{wem}}{U_{wys}} II\right). \quad (39)$$

Spośród strat na przełączenie tranzystora, w rozważanym układzie należy uwzględnić tylko straty przy jego wyłączaniu. Przy jego załączaniu, ze względu na indukcyjne obciążenie, napięcie na tranzystorze bardzo szybko spada do wartości równej U_{ces} , a straty w tych warunkach zostały już uwzględnione w zależności (39).

Powtarzając rozważania zamieszczone w [2] można pokazać, że wyłączenie tranzystora w każdym z okresów kluczowania odbywa się przy napięciu na nim równym $\cong U_{wyn}$. Analogicznie jak to zrobiono w [2] założymy, że opadanie prądu tranzystora przy jego wyłączaniu odbywa się liniowo według zależności

$$i_{Tn}(t) = I_{Lmn} \cdot \left(1 - \frac{t}{t_{op}}\right), \quad 0 < t < t_{op}. \quad (40)$$

(czas t liczony jest tu od momentu rozpoczęcia opadania prądu). Energia strat na jedno wyłączenie tranzystora (tj. w jednym okresie kluczowania) będzie wtedy równa

$$W_{2n} = \int_0^{t_{op}} i_{Tn}(t) \cdot u_{ce}(t) dt = U_{wyn} \int_0^{t_{op}} I_{Lmn} \left(1 - \frac{t}{t_{op}}\right) dt. \quad (41)$$

Podstawiając (2) i całkując otrzymamy

$$W_{2n} = \frac{U_{\text{wyn}} U_{\text{wen}}}{2L} \cdot t_1 t_{\text{op}}. \quad (42)$$

Energia strat tranzystora na przełączanie tranzystora w całym półokresie sieci będzie sumą wyrażeń (42)

$$W_2 = \sum_{n=1}^N \frac{U_{\text{wyn}} U_{\text{wen}}}{2L} \cdot t_1 t_{\text{op}} = \frac{t_{\text{op}}}{2L} \cdot \sum_{n=1}^N U_{\text{wyn}} U_{\text{wen}} t_1. \quad (43)$$

Podobnie jak to zrobiono wcześniej ((31), (32)), wyrażenie pod znakiem sumy można przekształcić następująco

$$U_{\text{wyn}} U_{\text{wen}1} = U_{\text{wyn}} U_{\text{wen}} \cdot \frac{t_1}{T_n} \cdot T_n = U_{\text{wyn}} U_{\text{wen}} \cdot \frac{t_1}{t_1 + t_{30}} \left(1 - \frac{U_{\text{wen}}}{U_{\text{wyn}}}\right) \cdot T_n. \quad (44)$$

Podstawiając (44) do (43) otrzymamy

$$W_2 = \frac{t_{\text{op}}}{2L} \cdot \frac{t_1}{t_1 + t_{30}} \cdot \sum_{n=1}^N (U_{\text{wyn}} U_{\text{wen}} - U_{\text{wen}}^2) \cdot T_n. \quad (45)$$

Tak jak jak to było przy obliczaniu strat w nasyceniu, sumę w powyższej zależności można przybliżyć całką oznaczoną

$$W_2 = \frac{t_{\text{op}}}{2L} \cdot \frac{t_1}{t_1 + t_{30}} \int_0^{1/2f} (u_{\text{wy}}(t) u_{\text{we}}(t) - u_{\text{we}}^2(t)) \cdot dt. \quad (46)$$

Podstawiając zależności (1) i (25) na $u_{\text{we}}(t)$ i $u_{\text{wy}}(t)$ oraz wykonując całkowanie otrzyma się

$$W_2 = \frac{t_{\text{op}}}{2L} \cdot \frac{t_1}{t_1 + t_{30}} \cdot \frac{U_{\text{wen}} U_{\text{wys}}}{\omega_s} \cdot \left(I_3 - \frac{U_{\text{wen}}}{U_{\text{wys}}} \frac{\pi}{2}\right), \quad (47)$$

gdzie

$$I_3 = \int_0^{\pi} g(x) \sin(x) dx; \quad (48)$$

(zastosowano tu oznaczenie „I3” zgodne z użytym w [2]).

Wartości całki I3 podano w [2]; są one nieznacznie większe od 2 np. dla $\omega_s RC = 5$ I3 2,008..., dlatego dalej przyjęto I3 = 2. W tym przypadku średnią moc traconą na przełączanie tranzystora można wyrazić następująco:

$$P_2 = W_2 \cdot 2f_s = \frac{U_{\text{wen}} U_{\text{wys}}}{\pi L} \cdot \frac{t_{\text{op}}}{1 + \frac{t_{30}}{t_1}} \cdot \left(1 - \frac{U_{\text{wen}}}{U_{\text{wys}}} \frac{\pi}{4}\right). \quad (49)$$

W powyższej zależności interesująca jest „słaba” zależność mocy P_2 od czasu t_1 . Można to zrozumieć, gdyż np. ze wzrostem czasu t_1 wzrasta energia strat W_{2n} w każdym okresie kluczkowania, ale jednocześnie maleje ilość tych okresów w półokresie sieci (czyli maleje częstotliwość kluczkowania), co przeciwdziała silnemu wzrostowi mocy P_2 .

3.2. STRATY W CEWCE

Procedura obliczeń strat w cewce w rozważanym układzie może być analogiczna do przedstawionej w [2], z uwzględnieniem różnic, jakie wynikają z innego sposobu sterowania układu. Procedura ta zostanie przedstawiona w skrócie poniżej.

Straty w cewce powstają w jej rdzeniu (zwykle ferrytowym) i w jej uzwojeniu. Moc strat w rdzeniu w przypadku ogólnym wyraża się następującą przybliżoną zależnością [2]:

$$P_{fe} = V_e(\alpha B + \beta B^2) \left(\frac{f}{f_0} \right)^w, \quad (50)$$

gdzie:

- P_{fe} — moc strat w rdzeniu,
- V_e — objętość rdzenia,
- B — wartość skuteczna indukcji ($B \leq 0,25$ T),
- f — częstotliwość sinusoidalnych zmian indukcji,
- f_0 — 15000 Hz (częstotliwość odniesienia),
- α, β, w — stałe materiałowe; dla krajowego ferrytu F-807):
 $\alpha = 3 \cdot 10^4 \text{ W/m}^3 \text{ T}, \beta = 1,8 \cdot 10^6 \text{ W/m}^3 \text{ T}^2, w = 1,15$.

Indukcję w rdzeniu można powiązać z prądem w uzwojeniu cewki następującą zależnością:

$$B = \frac{L \cdot I}{A_z \cdot z}, \quad (51)$$

gdzie:

- L — indukcyjność cewki,
- I — natężenie prądu w uzwojeniu,
- A_z — powierzchnia przekroju poprzecznego rdzenia,
- z — liczba zwojów uzwojenia.

Trudność obliczania mocy strat w cewce polega na tym, że zmiany indukcji w rdzeniu nie są sinusoidalne, a dodatkowo zmiany te w kolejnych okresach kluczkowania są nieco inne (zmienia się kształt impulsów prądu i_L). Dlatego obliczenia strat w rdzeniu należy dokonać następująco. Przebieg prądu $i_{Ln}(t)$, w n -tym okresie kluczkowania (tzn. jeden impuls prądu i_L), należy potraktować jako fragment o długości T_n odpowiedniego przebiegu okresowego $i_{Ln}(t)$. Przebieg ten należy rozłożyć w szereg Fouriera i znaleźć amplitudy jego harmonicznych. Jak można policzyć, dla przebiegu jak na rys.2 harmoniczne te będą się wyrażać następująco:

$$I_{n0} = \frac{I_{Lmn}(t_1 + t_{2n})}{2T_n} \quad (k = 0), \quad (52)$$

$$I_{nk} = \sqrt{a_{nk}^2 + b_{nk}^2}, \quad \text{natomiast} \quad (53)$$

$$a_{nk} = \frac{I_{Lmn}}{\Pi \omega_n k^2} \left(\frac{\cos k \omega_n t_1 - 1}{t_1} + \frac{\cos k \omega_n t_1 - \cos k \omega_n (t_1 + t_{2n})}{t_{2n}} \right),$$

$$b_{nk} = \frac{I_{Lmn}}{\Pi \omega_n k^2} \left(\frac{\sin k \omega_n t_1}{t_1} + \frac{\sin k \omega_n t_1 - \sin k \omega_n (t_1 + t_{2n})}{t_{2n}} \right),$$

gdzie $\omega_n = \frac{2\pi}{T_n}$ i $k > 0$.

Dla każdej harmonicznej (o rzędzie nie przekraczającym przyjętej wartości maksymalnej) należy obliczyć odpowiadającą jej wartość skuteczną indukcji w rdzeniu

$$B_{nk} = \frac{L \cdot I_{nk}}{\sqrt{2} \cdot A_z \cdot z}. \quad (54)$$

Dla każdej tak obliczonej wartości indukcji trzeba obliczyć cząstkową energię strat w danym okresie kluczkowania

$$W_{fenk} = V_c (\alpha B_{nk} + \beta B_{nk}^2) \left(\frac{k}{f_0 T_n} \right)^w \cdot T_n. \quad (55)$$

Energia strat w rdzeniu w danym okresie kluczkowania będzie sumą energii strat dla poszczególnych harmonicznych

$$W_{fen} = \sum_{k=1}^h W_{fenk}. \quad (56)$$

Średnia moc strat w rdzeniu będzie ostatecznie równa

$$P_{fe} = W_{fe} \cdot 2f_s. \quad (58)$$

Obliczenie strat w uzwojeniu jest prostsze. Najpierw należy obliczyć jego rezystancję

$$r_{cu} = \frac{\rho \cdot l_z \cdot z^2}{k_{cu} \cdot S_0}, \quad (59)$$

gdzie:

ρ — rezystywność materiału uzwojenia (przyjęto miedź),

l_z — średnia długość jednego zwoju,

S_0 — powierzchnia okna wypełniana uzwojeniem,

k_{cu} — współczynnik wypełnienia okna (przyjęto $k_{cu} = 0.5$).

Energia strat w uzwojeniu w jednym okresie kluczkowania jest równa

$$W_{cun} = r_{cu} \left(I_{n0}^2 + \sum_{k=1}^h \frac{I_{nk}^2}{2} \right) \cdot T_n. \quad (60)$$

Łączna energia strat w uzwojeniu w półokresie sieci będzie sumą energii (60)

$$W_{cu} = \sum_{k=1}^N W_{cun}, \quad (61)$$

zaś średnia moc strat

$$P_{cu} = W_{cu} \cdot 2f_s. \quad (62)$$

Wszystkie przedstawione obliczenia ujęto w stosowanym programie komputerowym. Dzięki temu możliwe było wielokrotne powtarzanie tych obliczeń przy zmianach niektórych parametrów wyjściowych takich jak indukcyjność cewki, liczba zwojów, rodzaj rdzenia, wartość napięcia wyjściowego itp.

4. OBLICZENIA, WNIOSKI

Na podstawie powyższych rozważań dokonano szeregu obliczeń strat mocy przy różnych parametrach układu. Jako niezmienną przyjęto jego moc wyjściową $P_{wy} = 50$ W. Na tej podstawie obliczano rezystancję $R = U_{wys}^2/P_{wy}$ i następnie t_1 z równania (27). Wyniki tych obliczeń zestawiono w tablicy 1, w której, dla porównania, zamieszczano także wartości strat mocy dla identycznego układu, ale pracującego ze stałą częstotliwością kluczowania. Te ostatnie dane zaczerpnięto z [2]. W tablicy 1 zwraca uwagę fakt, że — przy przyjętych parametrach — roz-

Tablica 1

Wyniki obliczeń strat mocy w układzie dla $F = var$ i $F = const$

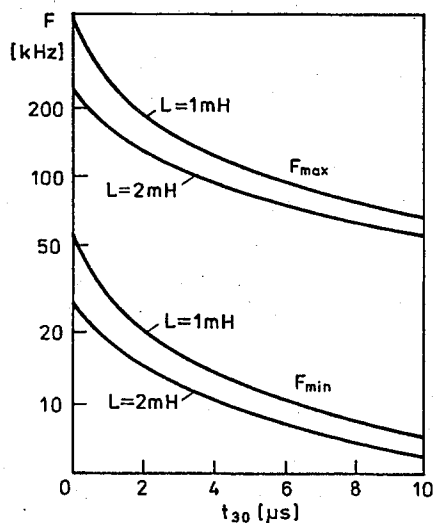
$P_{wy}=50$ W, $U_{wem}=311$ V, $U_{ces}=2$ V, $f_s=50$ Hz, $t_{op}=0,5$ μ s, $\omega_s RC=5$ $h=40$ rdzeń EE-30 F-807, $z=250$					
Moc strat, W		F=var $t_{30}=0$		F=const=15 kHz	
		$U_{wyz}=350$ V	$U_{wyz}=380$ V	$U_{wyz}=350$ V	$U_{wyz}=380$ V
na nasycenie tranzystora	$L=1$ mH	0,25	0,29	0,12	0,15
	$L=2$ mH	0,25	0,29	0,12	0,15
na przełączanie tranzystora	$L=1$ mH	5,30	6,79	1,61	1,94
	$L=2$ mH	2,65	3,39	1,14	1,37
łączna w tranzystorze	$L=1$ mH	5,55	7,08	1,73	2,09
	$L=2$ mH	2,90	3,69	1,26	1,52
w rdzeniu cewki	$L=1$ mH	0,08	0,09	0,05	0,05
	$L=2$ mH	0,08	0,09	0,06	0,06
w uzwojeniu cewki	$L=1$ mH	0,08	0,08	0,23	0,26
	$L=2$ mH	0,08	0,08	0,16	0,19
łączna w cewce	$L=1$ mH	0,16	0,17	0,28	0,31
	$L=2$ mH	0,16	0,17	0,22	0,25
łączna w układzie	$L=1$ mH	5,71	7,26	2,01	2,40
	$L=2$ mH	3,06	3,86	1,48	1,77

Tablica 2

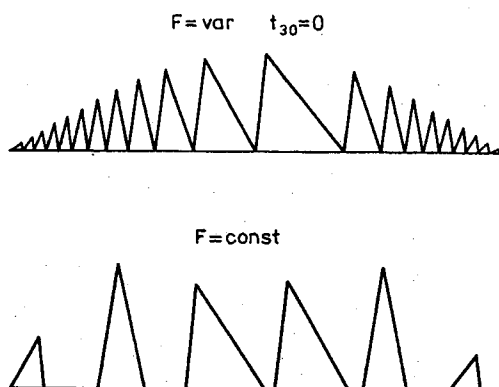
Zależność strat w układzie od czasu t_{30}

$P_{wy}=50$ W, $U_{wyz}=350$ V, $U_{wem}=311$ V, $U_{ces}=2$ V, $f_s=50$ Hz, $L=1$ mH $t_{op}=0,5$ μ s, $\omega_s RC=5$, $h=40$ rdzeń EE-30 F-807, $z=250$						
Moc strat, W	t_{30} , μ s	0	1	2	5	10
na nasycenie tranzystora		0.25	0.25	0.25	0.25	0.25
na przełączanie tranzystora		5.30	3.91	3.31	2.49	1.92
łączna w tranzystorze		5.55	4.16	3.56	2.74	2.17
w rdzeniu cewki		0.08	0.07	0.06	0.06	0.05
w uzwojeniu cewki		0.08	0.12	0.14	0.18	0.23
łączna w cewce		0.16	0.18	0.20	0.24	0.28
łączna w układzie		5.71	4.34	3.76	2.98	2.45

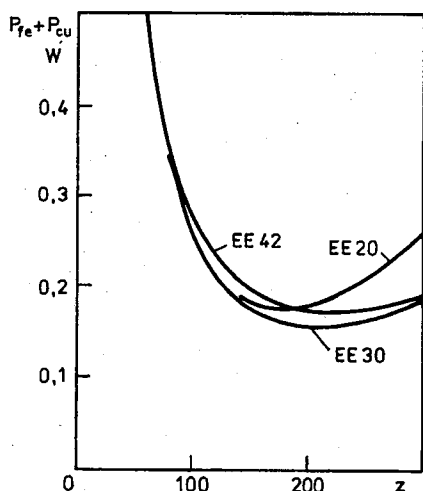
Rys. 3. Zależność granicznych wartości częstotliwości klucowania od czasu t_{30}



ważany układ jest lepszy od przedstawionego w [2] jedynie pod względem strat w uzwojeniu cewki, natomiast pod względem strat w tranzystorze i strat łącznych jest wyraźnie od niego gorszy. Straty w tranzystorze wyraźnie maleją przy wzroście indukcyjności, maleją też przy wzroście czasu t_{30} (tablica 2). Na tej podstawie można stwierdzić, że przyczyną dużych strat mocy w tranzystorze jest wysoka częstotliwość klucowania, z jaką pracuje układ. Jej wartość średnia jest rzędu 100 kHz, a wartość chwilowa zmienia się od setek kHz (dla $U_{we} = 0$) do rzędu 10 kHz (dla $U_{we} = U_{wem}$), co pokazano na rys.3. Zmienność częstotliwości klucowania w półokresie napięcia sieciowego jest dobrze widoczna na rys.4, na którym w jednakowej skali i w sposób podglądowy, pokazano przebiegi prądu $i_L(t)$ dla obydwu porównywanych układów. Widoczne też jest, że większej częstotliwości klucowania w rozważanym układzie towarzyszą nieco mniejsze amplitudy impulsów prądu $i_L(t)$. Dzięki temu, pomimo większej częstotliwości, straty w rdzeniu cewki nie są zbyt duże. Reasumując powyższe rozważania można stwierdzić, że możliwość zmniej-



Rys. 4. Przebieg prądu $i_L(t)$ w półokresie dla obydwu układów (rysunek poglądowy)



Rys. 5. zależność mocy strat w cewce od liczby jej zwojów

szenia strat w tranzystorze przez wzrost indukcyjności cewki i czasu t_{30} wiąże się z występującym wtedy spadkiem częstotliwości kluczkowania. Praktycznie możliwy jest tu tylko dobór indukcyjności, gdyż — jak wspomniano wcześniej — takie sterowanie układu, aby $t_{30} > 0$, może być dość trudne do realizacji (układ TDA 4814 tego nie umożliwia). Przy stosowaniu cewki o większej indukcyjności należy pamiętać, że indukcja w rdzeniu $B \sim L$, co grozi jego nasyceniem się. Można tego uniknąć przez zastosowanie większego rdzenia i/lub dobór liczby zwojów uzwojenia. Może to dodatkowo przyczynić się do zmniejszenia strat w układzie ze względu na występowanie wyraźnego minimum w zależności $P_{fe} + P_{cu} = f(z)$, co pokazano na rys.5.

Szacując sprawność energetyczną układu można ocenić, że dla $L = 1\text{mH}$ będzie ona mniejsza niż 85–88%, dla $L = 2\text{mH}$ może być zbliżona do 90%. Sprawność ta będzie więc o kilka % mniejsza niż układu pracującego z $F = \text{const}$ [2]. Oczywiście, że w obliczeniach zawartych w pracy poczyniono uproszczenia, które mają wpływ na uzyskane wyniki liczbowe — np. przy obliczaniu mocy P_2 założono stałość czasu t_{op} , co w pewnym stopniu zawyżyło uzyskane wartości tej mocy. Dodatkowe obliczenia pokazały, że to zawyżenie może być równe ok. 30–35%, a więc nie tak dużo.

Porównując ostatecznie właściwości rozważanego układu przy jego pracy ze zmienną lub stałą częstotliwością kluczkowania można stwierdzić, że pierwszy przypadek ($F = \text{var}$) cechuje się:

- mniejszą o kilka % sprawnością energetyczną;
- prostszym sterowaniem ($t_1 = \text{const}$);
- niewystępowaniem ograniczeń na wartość częstotliwości kluczkowania, jak to jest dla $F = \text{const}$.

Gdyby istniały specjalizowane układy sterujące dla przypadku $F = \text{const}$, wówczas taki sposób pracy układu eliminacji harmonicznego prądu sieciowego powinien być preferowany, ze względu na mniejsze straty mocy. Byłoby to istotne zwłaszcza dla układów o większej mocy wyjściowej.

BIBLIOGRAFIA

1. M. G o t f r y d: *Analiza układów impulsowych do eliminacji zniekształceń prądu sieciowego — zależności podstawowe*. Kwartalnik Elektroniki i Telekomunikacji, 1992, t.38, z.4
2. M. G o t f r y d: *Analiza układów impulsowych do eliminacji zniekształceń prądu sieciowego — zależności energetyczne*. Kwartalnik Elektroniki i Telekomunikacji, 1992, t.38, z.4
3. SIEMENS: *ICs für Industrielle Anwendungen*. Datenbuch 1989/90

M. GOTFRYD

S u m m a r y

THE ACTION OF PULSE CIRCUIT WORKING WITH VARIABLE SWITCHING FREQUENCY
USED FOR DISTORTIONS DECREASING OF MAINS CURRENT

In the paper there has been considered a pulse circuit used for distortions decreasing of input current of mains powered devices. The circuit operates with variable switching frequency; the way of its action is equal to the one obtained when the TDA 4814 integrated circuit is used. There has been derived, considered and calculated losses in the switching transistor and the coil. Basing on those results there has formula for output voltage of the circuit and there have been made a comparison between the properties of the circuit and the same one but working with constant switching frequency.

Równania konstytutywne ośrodka ciągłego dla procesu ciągłej krystalizacji masowej

Część I. Konstrukcja równań konstytutywnych ośrodka ciągłego dla procesu ciągłej krystalizacji masowej

Praca doktorska — Część I

WACŁAW NIEMIEC

*Niezależny Instytut — Centrum Edukacyjne „Wiedza Nauka”
dawniej — Politechnika Śląska w Gliwicach*

*Otrzymano 1993.01.10
Aytoryzowano do druku 1993.11.10*

Przedmiotem tej części artykułu jest opracowanie konstytutywnego rozłożonego parametrycznie modelu dla całego procesu ciągłej krystalizacji masowej ujmującego wszystkie jego procesy składowe:

- proces zarodkowania pierwotnego,
- proces wzrostu kryształów,
- proces zarodkowania wtórnego,

rozłożone w przestrzeni roboczej krystalizatora i czasie we wzajemnych fizykochemicznych relacjach. W oparciu o analizę literaturowych eksperymentów w zakresie kinetyk:

- reakcji zarodkowania pierwotnego,
- reakcji wzrostu kryształów zakłóconej w kinetyce procesem zarodkowania wtórnego,

oraz towarzyszące im zjawiska fizyczne podano oryginalną koncepcję modelowania całego procesu ciągłej krystalizacji masowej zwaną dalej „*koncepcją podwójnie miejscowo rozłożonych parametrów*”. Konsekwencją takiego całościowego podejścia było wyprowadzenie w pracy układów równań różniczkowych cząstkowych konstytutywnych stanu bilansów masy, energii i pędu dla przesyconego roztworu — proces zarodkowania pierwotnego oraz dla kryształu wewnątrz przesyconego roztworu — proces wzrostu kryształów, zaś zakłócenie kinetyki reakcji wzrostu kryształów poprzez proces zarodkowania wtórnego dane jest równaniem bilansu energii mechanicznej na powierzchni kryształu. Dla tego oryginalnego konstytutywnego rozłożonego parametrycznie modelu matematycznego całego procesu ciągłej krystalizacji masowej w pracy przeprowadzono:

- analizę zagadnień inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami przestrzenną i czasową,
- klasyfikację matematyczną układów równań różniczkowych cząstkowych konstytutywnych stanu bilansów masy, energii i pędu,
- dyskusję zagadnień sterowania opracowanym modelem krystalizacji.

1. WSTĘP

Proces krystalizacji jest naturalnym sposobem otrzymywania fazy krystalizacyjnej wewnątrz przesyconego roztworu. Z tego powodu proces ten znalazł szerokie pole zastosowań do:

- otrzymywania materiałów krystalicznych o zadanych właściwościach,
- oczyszczania substancji,

w procesach metalurgicznych oraz w procesach technologii chemicznej.

W naukowej literaturze o krystalizacji podkreśla się szczególne znaczenie następujących elementów tego procesu [1 – 20]:

- kinetyk procesów składowych,
- zjawisk fizykalnych procesu krystalizacji,

które są decydujące o wydajnościowych i jakościowych zagadnieniach produktów tego procesu i konsekwentnie powodują, że albo powstaje duża masa małych kryształów o charakterystycznych dla siebie własnościach fizykalnych lub przeciwnie pojedyncze kryształy o dużych rozmiarach i znacznie różnych właściwościach fizykalnych. Z tego punktu widzenia technologia krystalizacji rozróżnia następujące jego procesy składowe:

- zarodkowanie pierwotne — pojawianie się zarodków krystalicznych wewnątrz przesyconego roztworu,
- wzrost kryształów (wzrost fazy krystalicznej wewnątrz przesyconego roztworu),
- zarodkowanie wtórne (pojawienie się dodatkowych zarodków krystalicznych z obłamanych dendrytów i krawędzi kryształów podczas ich ruchu wewnątrz przesyconego roztworu).

Problemy kompletnego i konstytutywnego opisu wymienionych 3 (trzech) procesów nie zostały zrealizowane w światowej literaturze i dlatego proponowany konstytutywny sposób opisu całości fizykalnych aspektów tych procesów stanowi istotną nowość w podejściu do tak złożonego procesu w światowej literaturze.

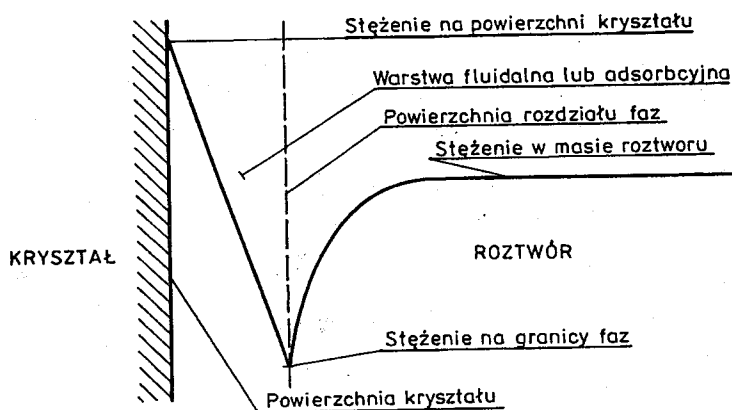
2. OMÓWIENIE TEMATU

Klasyczna literatura opisująca fizykalne aspekty krystalizacji dzieli ten proces na trzy procesy składowe [1 – 20]:

— zarodkowanie pierwotne: Przejawia się tym, że w przesyconym roztworze na skutek określonych warunków fizykalnych powstaje taki rozkład potencjału elektrochemicznego, który decyduje o układaniu się jonów, cząsteczek i elektronów w podstawowe elementy sieci krystalicznej. Powstaje sieć dendrytyczna, która decyduje o strukturze i właściwościach rosnącego kryształu. Im większa jest intensywność procesu zarodkowania pierwotnego, tym więcej małych kryształów jest wynikiem procesu krystalizacji [1 – 7, 11, 13, 15 – 16, 18, 21, 29].

— wzrost kryształów: Jest to kontynuacja procesu zarodkowania pierwotnego i dotyczy kryształu znajdującego się wewnątrz warstwy fluidalnej lub adsorbcyjnej, rozpiętej na układzie dendrytycznym wokół rosnącego kryształu.

Taka idea jest wiodącą wśród różnych teorii przedstawiających mechanizm reakcji wzrostu kryształów. Jednym z najbardziej istotnych czynników decydujących o zagadnieniach miejscowych reakcji wzrostu kryształów jest rozkład stężenia wokół rosnącego kryształu. Jest on określony poprzez rozkład temperatury w układzie



Rys. 1. Rozkład stężenia wokół rosnącego kryształu

roztwór – kryształ, a także pola wektora prędkości roztwór – kryształ. Geometrię zagadnienia lokalnego wokół rosnącego kryształu przedstawiono na rysunku 1. Istotne jest tutaj stężenie nasycenia dla generacji przesyconia. Proces ten jest bowiem związany z rozkładem przesyconia krystalizującego roztworu w układzie: roztwór przesycony – kryształ. Im większa jest intensywność procesu wzrostu kryształów tym większe pojedyncze kryształy są wynikiem procesu krystalizacji [1 – 11, 14 – 15, 17, 21 – 23, 27 – 29].

— zarodkowanie wtórne: zachodzi jako proces mechanicznego zakłócenia kinetyki wzrostu kryształów poprzez zderzenia: kryształ – kryształ, kryształ – mieszadło, kryształ – ścianki krystalizatora. Wykruszone w wyniku tych zderzeń dendryty lub naroża kryształów stanowią dodatkowe zarodki, które podlegają tak samo jak zarodki pierwotne procesowi wzrostu [18 – 25, 28 – 29].

Autorzy cytowanej literatury postulują następujący sposób opisu krystalizacji [1 – 20]:

- kompleksowy opis jej wszystkich procesów składowych,
- oparcie opisu procesu zarodkowania pierwotnego i wzrostu kryształów na bilansach masy, energii i pędu,
- uznają proces zarodkowania wtórnego jako zakłócenie kinetyki procesu wzrostu kryształów.

Chociaż Autorzy sami bardzo krytycznie oceniają swoje opracowania oraz ogólny stan poziomu literatury dotyczącej procesu krystalizacji zdając sobie sprawę, że „otrzymane wyniki są raczej konsekwencją przyjętych założeń niż rzeczywistym obrazem procesu” [6] to jednak nie są w stanie opisać w sposób pełny całości fizykochemii nawet pojedynczych procesów składowych.

3. CEL I TEZY PRACY

W światowej literaturze o krystalizacji twierdzi się, że proces ten nie został nawet częściowo opisany stąd też podjęcie tego tematu jest aktualne i celowe.

Celem rozprawy jest opis procesu krystalizacji przy pomocy równań konstytutywnych fizykalnych aspektów wszystkich jego procesów składowych: — zarodkowania pierwotnego, — wzrostu kryształów zakłóconego w kinetyce zarodkowaniem wtórnym z punktu widzenia ich wzajemnych zależności rozłożonych w czasie i przestrzeni roboczej krystalizatora Ω_{Rt} przy pomocy bilansów masy, energii i pędu. Uzasadnienie przyjętych założeń i metody rozwiązania uzasadniają udowodnienia następujących tez:

Teza I. Warunki lokalne objętość — czas $\bar{\Omega}_{(a,b,c)t}$ oraz $\bar{\Omega}_{K(a,b,c)t}$ decydują o intensywności reakcji zarodkowania pierwotnego wokół ekwiwalentnego punktu $Z(x, y, z, t)$ oraz reakcji wzrostu kryształów i zarodkowania wtórnego wokół ekwiwalentnego punktu $K(x, y, z, t)$.

Teza II. Punkty $Z(x, y, z, t)$ a także $K(x, y, z, t)$ oraz ich lokalne otoczenia $Z(x, y, z, t) \in \bar{\Omega}_{(a,b,c)t}$ oraz $K(x, y, z, t) \in \bar{\Omega}_{K(a,b,c)t}$ posiadają swoje wzajemnościowe obrazy wewnątrz przestrzeni roboczej — czas dla krystalizatora Ω_{Rt} — warunki globalne.

Teza III. Bilanse masy, energii i pędu opisują proces zarodkowania pierwotnego i wzrostu kryształów (zakłóconego w kinetyce zarodkowaniem wtórnym) ze względu na ich kinetyczne i zjawiskowe aspekty.

Teza IV. Równania konstytutywne ośrodka ciągłego wyprowadzone z bilansów masy, energii i pędu dla zarodkowania pierwotnego i wzrostu kryształów, zakłóconego w kinetyce zarodkowaniem wtórnym są inwariantne (niezmiennicze) względem swoich współczynników fizykalnych na podstawie odpowiednio sformułowanej zasady inwariancji konstytutywnej

Teza V. Zasada inwariancji konstytutywnej dla TEZY IV jest dana w formie układów równań różniczkowych cząstkowych ciągłości.

Teza VI. Równania różniczkowe cząstkowe konstytutywne stanu są transformalne do:

— funkcji źródłowych wewnątrz przestrzeni roboczej — czas Ω_{Rt} dla zarodkowania pierwotnego w punkcie $Z(x, y, z, t) \in \bar{\Omega}_{(a,b,c)t}$ i dla wzrostu kryształów zakłóconego w kinetyce zarodkowaniem wtórnym w punkcie $K(x, y, z, t) \in \bar{\Omega}_{K(a,b,c)t}$,

— funkcji źródłowych na brzegu F_{Rt} z Ω_{Rt} urządzenia krystalizującego przy czym; dla zarodkowania pierwotnego $R(x, y, z, t) \in F_{Rt}$ oraz dla wzrostu kryształów zakłóconego w kinetyce zarodkowaniem wtórnym $P(x, y, z, t) \in F_{Rt}$,

— zaś postacię jednorodną tych równań podlegają dekompozycji dla zjawisk opisywanych przez jednorodne równania różniczkowe cząstkowe pól potencjalnych i wirowych.

Teza VII. Wspólna unormowana amplituda funkcji własnych daje możliwość otrzymania funkcji Greena dla zjawisk procesów zarodkowania pierwotnego i wzrostu kryształów w zależności od wymiarów elementów lokalnych $\bar{\Omega}_{(a,b,c)t}$ oraz $\bar{\Omega}_{K(a,b,c)t}$. Elementy lokalne $\bar{\Omega}_{(a,b,c)t}$ oraz $\bar{\Omega}_{K(a,b,c)t}$ są zjawiskowo domknięte tzn. wpływ poszczególnych zjawisk fizykalnych na warunki lokalne bilansów masy, energii i pędu może być traktowany oddzielnie.

Teza VIII. Kompletne rozwiązania dla procesów składowych krystalizacji tj. zarodkowania pierwotnego i wzrostu kryształów zakłóconego w kinetyce zarodkowaniem wtórnym bazują na sumowaniu wektorów stanu rozwiązań analitycznych poszczególnych zjawisk o współrzędnych: masa, energia i pęd.

4. FIZYKO – CHEMICZNE PROCESY SKŁADOWE KRYSTALIZACJI PRZESYCONYCH ROZTWORÓW

Ciała stałe znajdują się w stanie krystalicznym lub amorficznym. Proces krystalizacji jest formowaniem się fazy stałej w przesyconym roztworze i jest podstawowym zjawiskiem w przyrodzie, a jako taki jest wykorzystywany w technice. Kryształy są nieodpartym dowodem istnienia ładu w przyrodzie. Miejsce każdej cząsteczki w kryształcie wyznacza dokładnie położenie jej sąsiadów. W kryształcie idealnym podstawowe układy elementarne są powtarzalne i powstaje w każdym miejscu sieć w sposób pod każdym względem identyczny. Ze względu na istnienie takiego powtarzalnego wzorca kryształ może rosnąć w nieskończoność poprzez zwykłe dodawanie dalszych cząstek, co jest wystarczające do jego wzrostu.

Gdy ciecz zawierająca jakąś obcą substancję rozpuszczoną zostanie poddana krzepnięciu, to pierwsze kryształy wytworzone w ten sposób są zwykle znacznie czystsze niż sama ciecz, roztwór. Dzieje się tak dlatego że ciecz, której struktura jest chaotyczna, może zawierać obce atomy lub cząsteczki o wymiarach nietypowych pojawiające się z mniejszym trudem niż kryształ mający ściśle określone wymiary sieciowe. W konsekwencji obce cząstki wykazują tendencję do unikania, miałyby z nim bowiem wyższą energię. Tendencji tej przeciwstawia się entropia, która zmierza do rozproszenia substancji rozpuszczonej w całym układzie, zarówno ciekłym jak i stałym. W danej temperaturze pomiędzy rozcieńczonym roztworem stałym, a bardziej stężonym roztworem ciekłym zostaje osiągnięta równowaga. Odrzucona przez kryształy substancja rozpuszczona koncentruje się w pozostałej cieczy, której skład staje się wskutek tego jeszcze mniej korzystny dla krystalizacji. Następuje przeto obniżenie temperatury krzepnięcia cieczy. Ściśle określona temperatura krzepnięcia substancji czystej przestaje się w obecności substancji rozpuszczonej w całym zakresie temperatur krzepnięcia. Przy dalszym oziębianiu, pojawia się warstwa kryształów ciała bogatszego w substancję rozpuszczoną, lecz jeszcze czystszych niż ciecz. Proces ten trwa dopóty, dopóki w odpowiednio niskiej temperaturze ostatnie ślady cieczy, która nie wzięła jeszcze udziału w danych fazach krystalizacji ale, już znacznie zgęszczonej substancji rozpuszczonej, nie zestaliły się w strefach zbliżonych do granic ziarn. Mamy więc do czynienia ze stopniowym wzrostem stężenia substancji rozpuszczonej od środków do granic ziarn. Literatura krystalizacji rozróżnia w procesie krystalizacji następujące procesy składowe: zarodkowanie pierwotne, wzrost kryształów oraz zarodkowanie wtórne, które są wzajemnie powiązane zjawiskami fizykalnymi w przestrzeni roboczej kry-

stalizatora i czasie. Procesy te oddziałują na siebie i produkt jest wynikiem tych oddziaływań. W oparciu o klasyczne pozycje literatury składowe procesy krystalizacji zawierają następujące fizyko – chemiczne zagadnienia [1 – 20].

4.1. PROCES ZARODKOWANIA PIERWOTNEGO

Rozkład potencjału elektro – chemicznego wokół punktu $Z(x, y, z, t)$ wywołuje elementy sieci krystalicznej zarodka „in statu nascendi”. Ten rozkład jest związany z wektorem stanu o współrzędnych: stężenie, temperatura, wektor prędkości [1 – 7, 11 – 13, 15 – 16, 18, 21 – 28]. Szywność sieci krystalicznej zarodka pozwala na zdefiniowanie lokalnych warunków zarodkowania pierwotnego w lokalnym elemencie objętość – czas $\bar{\Omega}_{(a_1, b_1, c_1)_t}$. Wektor stanu warunków fizykalnych wokół punktu $Z(x, y, z, t)$ związany jest z następującymi zjawiskami fizycznymi wewnątrz przesyconego roztworu:

- dyfuzyjny transport stężenia roztwór – roztwór,
- dyfuzyjny transport entalpii roztwór – roztwór,
- przewodnictwo cieplne roztwór – roztwór,
- wektor prędkości roztwór – roztwór.

Ta analiza zarodkowania pierwotnego prowadzi do następujących założeń:

- Założenie ZZZP1: Reakcja zarodkowania pierwotnego w przestrzeni roboczej — czas Ω_{Rt} da się sprowadzić do ekwiwalentnej reakcji zastępczej w punkcie $Z(x, y, z, t) \in \bar{\Omega}_{(a_1, b_1, c_1)_t}$.
- Założenie ZZZP2: Wymiary lokalnego prostopadłościennego otoczenia, w którym zachodzi reakcja zarodkowania pierwotnego, $(a_1, b_1, c_1)_t$ zawiera ją całą sieć zbudowaną z dendrytów.
- Założenie ZZZP3: Lokalny element objętość – czas $\bar{\Omega}_{(a_1, b_1, c_1)_t}$ uważamy ze względu na zachodzące zjawiska za domknięty pod względem ich bilansowych udziałów tzn. efekty wpływu poszczególnych zjawisk na wektor stanu w punkcie $Z(x, y, z, t) \in \bar{\Omega}_{(a_1, b_1, c_1)_t}$ mogą być rozpatrywane osobno.
- Założenie ZZZP4: Obowiązuje zasada wzajemności dla ośrodków anizotropowych i niejednorodnych w interpretacji Profesora Witolda Nowackiego [31 – 32].
- Założenie ZZZP5: Element $\bar{\Omega}_{(a_1, b_1, c_1)_t}$ w stanie statycznym jest w rozważanym procesie nieodkształcalny.

4.2. ANALIZA WZROSTU KRYSZTAŁÓW

Analiza warunków wzrostu kryształów na podstawie eksperymentów krystalizacji omawianych w literaturze przedmiotu prowadzi do sformułowania następujących poglądów [1 – 11, 14 – 15, 17, 21 – 23, 27 – 29]:

- kryształ wewnątrz warstwy fluoidalnej lub adsorbcyjnej rozpiętej na układzie dendrycznym wokół rosnącego kryształu rośnie według zadanej kinetyki reakcji wzrostu,

— kryształ wewnątrz warstwy fluidalnej lub adsorbcyjnej zajmuje ułamek objętościowy fazy krystalicznej,

— wektor stanu warunków fizykalnych wzrostu kryształów zależy od zmiennych stanu: ułamka objętościowego fazy krystalicznej, temperatury kryształu, prędkości kryształu, rozmiaru liniowego kryształu,

— literatura krystalizacji przyjmuje, że kryształ po zderzeniach zachowuje swój kształt i ma przykładowo strukturę regularną,

— kryształ pływa wewnątrz przesyconego roztworu z określoną prędkością,

— układ dendrytyczny jest sztywny o kształcie prostopadłościennym co pozwala założyć, że element lokalny zajmuje przestrzeń objętościowo—czasową $\bar{\Omega}_{K(a_2, b_2, c_2)t}$ oraz posiada wymiary „ a_2, b_2, c_2 ” dla czasu „ t ”,

— Przyjmujemy, że istnieje taki środek geometryczny $K(x, y, z, t) \in \bar{\Omega}_{K(a_2, b_2, c_2)t}$ do którego sprowadzam zastępczą reakcję wzrostu kryształów.

Z procesem wzrostu kryształów związane są ciągle zjawiska fizykalne zachodzące wewnątrz przesyconego roztworu:

— dyfuzyjny transport stężenia roztwór—kryształ,

— dyfuzyjny transport entalpii roztwór—kryształ,

— przewodnictwo cieplne roztwór—kryształ,

— wektor prędkości kryształu i warunek pływania roztwór—kryształ.

Stąd wynikają następujące założenia:

Założenie ZWK1: Reakcja wzrostu kryształów da się zastąpić ekwiwalentną reakcją sprowadzoną do punktu $K(x, y, z, t) \in \bar{\Omega}_{K(a_2, b_2, c_2)t}$.

Założenie ZWK2: Wymiary lokalnego prostopadłościennego otoczenia reakcji wzrostu kryształów $(a_2, b_2, c_2)t$ zawierają układ dendrytyczny łącznie z warstwą fluidalną lub adsorbcyjną.

Założenie ZWK3: W lokalnym elemencie domkniętym $\bar{\Omega}_{K(a_2, b_2, c_2)t}$ objętość—czas efekty wpływów poszczególnych zjawisk na wektor stanu w punkcie $K(x, y, z, t)$ mogą być rozpatrywane osobno.

Założenie ZWK4: Kryształ rośnie w układzie regularnym o osiach „ x, y, z, t ”.

Założenie ZWK5: Kryształ po zderzeniach zachowuje swój kształt.

Założenie ZWK6: Dla elementu $\bar{\Omega}_{K(a_2, b_2, c_2)t}$ obowiązuje zasada wzajemności dla ośrodków anizotropowych i niejednorodnych [31–32].

Założenie ZWK7: Element lokalny w stanie statycznym $\bar{\Omega}_{K(a_2, b_2, c_2)t}$ nie ulega odkształceniu.

4.3. ZARODKOWANIE WTÓRNE

Jest to proces zakłócający w sposób mechaniczny kinetykę wzrostu kryształów. Zakłócenie polega na obłamywaniu elementów sieci dendrycznej od rosnącego kryształu przez zderzenia: kryształ—kryształ, kryształ—mieszađło, kryształ—ścianki krystalizatora podczas ruchu kryształu wewnątrz krystalizującej zawiesiny: kryształu i przesycony roztwór [18–23, 28–29]. Proces zarodkowania wtórnego jest związany z charakterystycznymi rodzajami energii kryształu: E_z — energia kinetyczna zderzeń, E_w — energia wibracji, E_s — energia statyczna. Warunek wzrostu kryształów

ZWK5 mówiący iż kryształ po zderzeniach zachowuje swój kształt prowadzi do równania [18–23, 28–29]:

$$[E_z] + [E_w] = [E_s] \quad (4.3.1)$$

które jest związane z wektorem rozmiaru liniowego kryształu $L = L(L_x, L_y, L_z, t)$. W energii zderzeń kryształów $[E_z]$ mamy funkcję $f(L)$ zależną od następujących parametrów:

- współczynnika objętościowego kształtu kryształu,
- statystycznej liczbie zderzeń kryształu,
- gęstości kryształów,
- wektora rozmiaru liniowego kryształu.

Energia wibracji kryształów $[E_w]$ jest zależna od:

- tarcia kryształów w roztworze, która jest zależna od wartości przyjętych zmiennych stanu ($C, T, v, \eta_w, E, T_K, v_K$),
- drugiej pochodnej względem czasu rozmiaru liniowego kryształu.

Energia statyczna $[E_s]$ zależy od: energii ściskania kryształu i stałej Poissona $\bar{k}, h$,

- drugich pochodnych wektora liniowego kryształu względem zmiennych przestrzennych odpowiadających tensorowi naprężeń kryształu.

Powyższa analiza pozwala sformułować następujące założenia dla zarodkowania wtórnego:

- Założenie ZZW1:** Współczynnikiem objętościowym kształtu kryształu nazywamy odchylenie kształtu rosnącego kryształu od prostopadłościanu. Wartość tą przyjmujemy jako stałą w procesie.
- Założenie ZZW2:** Ilość zderzeń kryształów przyjmujemy jako średnią statystyczną.
- Założenie ZZW3:** Przyjmujemy uśrednioną wartość współczynnika tarcia kryształów w zawiesinie.
- Założenie ZZW4:** Przyjmujemy uśrednioną wartość modułu ściskania i stałej Poissona.

5. ZAGADNIENIA LOKALNE I GLOBALNE PRZY OPISIE KRYSTALIZACJI Z ROZTWORU

Rozważania rozdziału 4 pozwalają sformułować zagadnienie lokalne dla procesów składowych krystalizacji. Zagadnienia te wynikają z zasady wzajemności w ośrodkach anizotropowych i niejednorodnych, w których zjawiska krystalizacji zachodzą w sposób ciągły zgodnie z interpretacją prof. Witolda Nowackiego [31–32]. Warunki lokalne oraz globalne są ze sobą powiązane poprzez cyrkulacje wektorów normalnych powierzchni lokalnych i globalnej. Dla interpretacji zasady wzajemności wprowadzamy zmienny punkt $Q^R(\xi^R, \eta^R, \beta^R, \tau) \in \Omega_{Rt}$ przebiegający całą objętość przestrzennoczasową kryształizatora.

5.1. ZAGADNIENIA LOKALNE DLA PROCESÓW SKŁADOWYCH KRYSTALIZACJI

Proces krystalizacji powinien znaleźć swoją całkowitą interpretację poprzez zasadę wzajemności i ze względu na powyższe twierdzenie rozpatrujemy wzajemnościową interpretację poszczególnych składowych procesów. Przeprowadzony dowód wykazuje, że zasada wzajemności obejmuje procesy składowe krystalizacji i globalnie cały proces krystalizacji.

5.1.1. Zagadnienia lokalne i globalne dla zarodkowania pierwotnego

Dla zarodkowania pierwotnego definiuje się z punktu widzenia zasady wzajemności dla ośrodków anizotropowych i niejednorodnych warunki [21 — 26, 29]:

lokalne

$\bar{\Omega}_{(a_1, b_1, c_1)t}$ — element objętościowo — czasowy $m^3 - s$,
 $F_{(a_1, b_1, c_1)t}$ — zewnętrznie zorientowana powierzchnia dla $\bar{\Omega}_{(a_1, b_1, c_1)t}$ $m^2 - s$,
 $\mathbf{n}^{(z)}$ — wektor normalny do $F_{(a_1, b_1, c_1)t}$, — posiadający odpowiednią cyrkulację, $Z(x, y, z, t)$ — punkt ekwiwalentnej reakcji zarodkowania pierwotnego w elemencie $\bar{\Omega}_{(a_1, b_1, c_1)t}$.

Relacja według
 $\longleftrightarrow$
 zasady wzajemności
 [31 — 32].

globalne

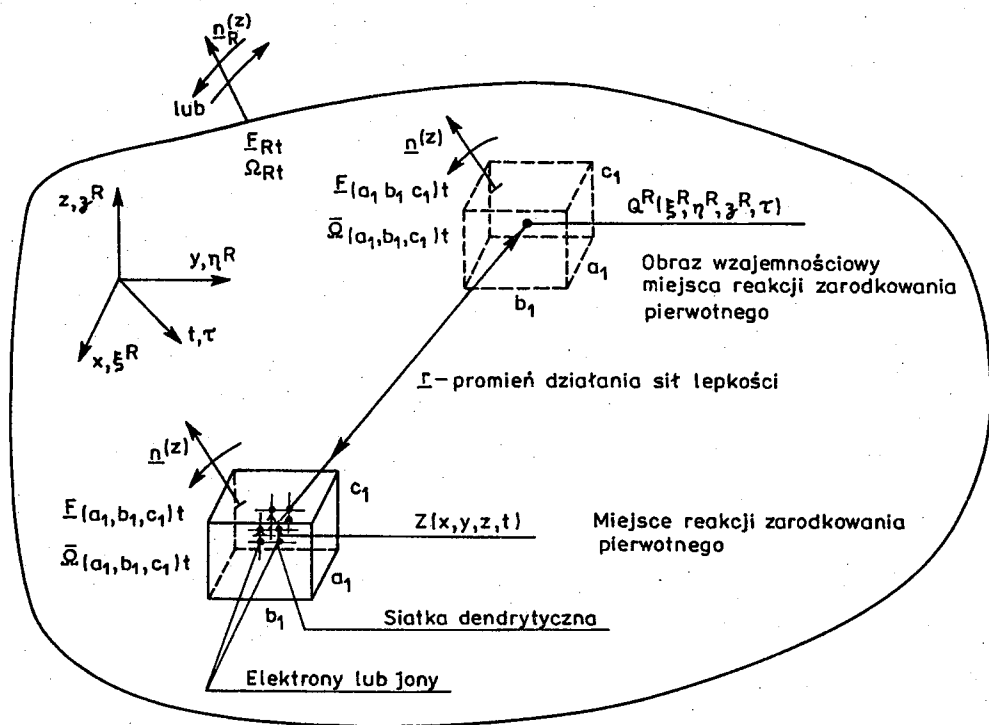
Ω_{Rt} — przestrzeń robocza — czas $m^3 - s$,
 F_{Rt} — zewnętrznie zorientowana powierzchnia dla Ω_{Rt} , $m^2 - s$,
 $\mathbf{n}_K^{(z)}$ — wektor orientacyjny do powierzchni F_{Rt} , z odpowiednią cyrkulacją,
 $Q^R(\xi^R, \eta^R, \beta^R, \tau)$ — obraz punktu $Z(x, y, z, t)$ w Ω_{Rt} wg. zasady wzajemności.

Są one związane ze zjawiskami fizycznymi będącymi odpowiedzią ośrodka ciągłego jakim jest przesycony roztwór na lokalną zmianę parametrów od reakcji zarodkowania pierwotnego. Wymiary lokalne $(a_1, b_1, c_1)t$ traktuje się jako największe możliwe wymiary siatki dendrycznej kryształu „in statu nascendi”. Cyrkulacja wektora normalnego $\mathbf{n}^{(z)}$ jest przeciwna do odpowiedniego wektora cyrkulacji dla procesu wzrostu kryształów zakłóconego w kinetyce zarodkowaniem wtórnym ze względu na to, że procesy zarodkowania pierwotnego i wzrostu kryształów są silnie przeciwne w sensie zapotrzebowania na produkt krystalizacji; duża ilość małych kryształów lub mała ilość dużych kryształów [38].

Zagadnienia lokalne dla zarodkowania pierwotnego oraz ich obraz spełniający zasadę wzajemności przedstawiono na rys.2.

W ośrodkach lepkich obraz wzajemnościowy jest związany z lepkością i promieniem działania sił wiskozyjnych. Lokalny element $\bar{\Omega}_{(a_1, b_1, c_1)t}$ jest związany z intensywnością reakcji zarodkowania pierwotnego V_B wzorem:

$$V_B = \bar{a} (C - C_s)^i \left[\frac{\text{mol}}{s} \right] \quad \text{przy } C \gg C_s \text{ oraz } i \geq 1 \quad (5.1.1.1)$$



Rys. 2. Koncepcja miejscowo rozłożonych parametrów dla reakcji zarodkowania pierwotnego

Reakcja zarodkowania pierwotnego może być określana w stosunku do właściwej masy molowej $G \left[\frac{\text{kg}}{\text{mol m}^3} \right]$ oraz daje ubytek stężenia od reakcji zarodkowania pierwotnego poprzez związek

$$V_B G = \bar{a} (C - C_s)^i G \left[\frac{\text{kg}}{\text{m}^3 \text{s}} \right] \quad (5.1.1.2)$$

gdzie: $\bar{a}$ — współczynnik temperaturowy dla reakcji zarodkowania pierwotnego $\left[\frac{\text{mol}}{\text{s}} \right] \left[\frac{\text{m}^3}{\text{kg}} \right]^i$, C — aktualne stężenie $\left[\frac{\text{kg}}{\text{m}^3} \right]$, C_s — stężenie nasycenia $\left[\frac{\text{kg}}{\text{m}^3} \right]$, i — rząd przesylenia dla reakcji zarodkowania pierwotnego, charakterystyczny dla danej substancji krystalizującej i procesu [1–7, 11–13, 15–16, 18, 21–26, 29].

5.1.2. Zagadnienia lokalne dla wzrostu kryształów zakłóconego w kinetyce zarodkowaniem wtórnym

Proces wzrostu kryształów zakłócony jest w kinetyce zarodkowaniem wtórnym ponieważ związany jest z zagadnieniami definiowanymi jako [21–23, 27–29]:

lokalne

$\bar{\Omega}_{K(a_2, b_2, c_2)t}$ — lokalnie rozłożona objętość—czas m^3-s ,
 $F_{K(a_2, b_2, c_2)t}$ — zewnętrznie zorientowana powierzchnia dla $\bar{\Omega}_{K(a_2, b_2, c_2)t}$ m^2-s ,
 $\mathbf{n}_K^{(z)}$ — wektor orientacyjny dla $F_{K(a_2, b_2, c_2)t}$ posiadający odpowiednią cyrkulację,
 $K(x, y, z, t)$ — punkt ekwiwalentnej reakcji wzrostu kryształów.

Relacja według

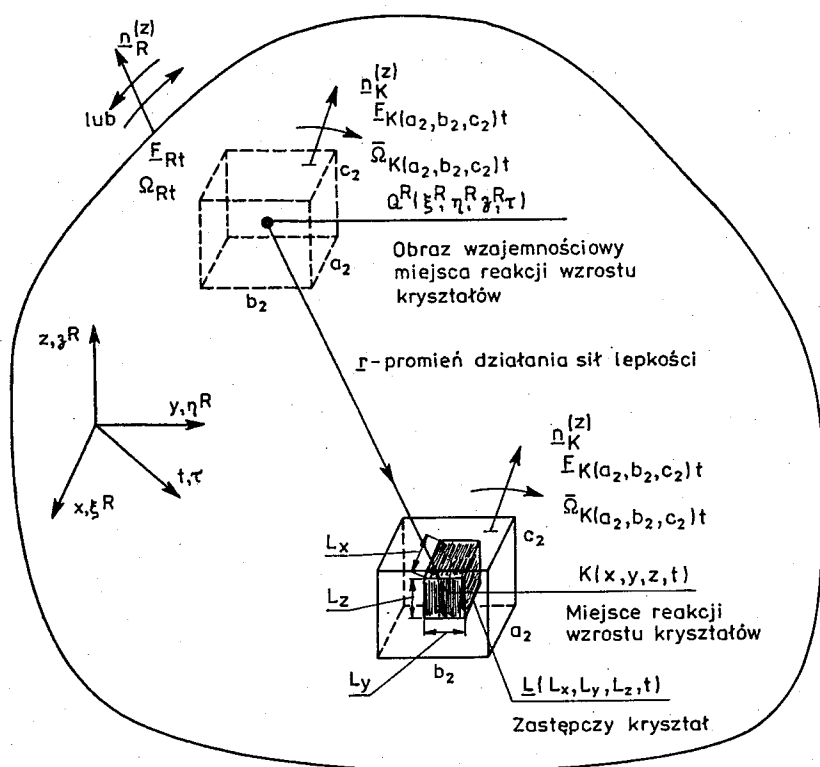


zasady wzajemności
 [31 – 32]

globalne

Ω_{Rt} — przestrzeń robocza—czas m^3-s ,
 F_{Rt} — zewnętrznie zorientowana powierzchnia dla Ω_{Rt} m^2-s ,
 $\mathbf{n}_R^{(z)}$ — wektor orientacyjny dla F_{Rt} z odpowiednią cyrkulacją,
 $Q^R(\xi^R, \eta^R, \beta^R, \tau)$ — obraz punktu $K(x, y, z, t)$ wg. zasady wzajemności.

Są one powiązane zjawiskami fizycznymi jako odpowiedzią ośrodka ciągłego jakim jest przesycony roztwór na zmiany parametrów lokalnych od reakcji wzrostu kryształów zakłóconej w kinetyce zarodkowaniem wtórnym. Wymiary lokalne $(a_2, b_2, c_2)t$ ujmują wymiary kryształu łącznie z jego warstwą fluidalną lub adsorbcyjną.



Rys. 3. Koncepcja miejscowo rozłożonych parametrów dla reakcji wzrostu kryształów

Postulujemy, że [38] cyrkulacja wektora $\mathbf{n}_K^{(z)}$ jest przeciwna do cyrkulacji $\mathbf{n}^{(z)}$ dla zarodkowania pierwotnego. Zagadnienia lokalne dla reakcji wzrostu kryształów oraz zarodkowania wtórnego i ich obraz wzajemnościowy przedstawione są na rys.3.

Podobnie jak dla zarodkowania pierwotnego także dla kryształu wewnątrz warstwy fluoidalnej lub adsorbcyjnej za obraz wzajemnościowy odpowiedzialna jest wiskozyjność roztworu. Intensywność reakcji wzrostu kryształów zakłócona w kinetyce zarodkowaniem wtórnym V_{GK} opisywana jest wzorem:

$$V_{GK} = \bar{b}(C - C_s)^j f(L) \left[\frac{\text{mol}}{\text{s}} \right] \quad \text{przy czym } C \gg C_s \text{ oraz } j \geq 1 \quad (5.1.2.1)$$

Jeżeli uwzględnimy gęstość własną kryształu $G_K \left[\frac{\text{kg}}{\text{mol m}^3} \right]$ oraz ułamek fazy krystalicznej E [objętość kryształu / $\bar{\Omega}_{K(a_1, b_1, c_1)t}$] otrzymujemy przyrost masy kryształów wewnątrz przesyconego roztworu wyrażony wzorem

$$V_{GK} G_K E = \bar{b}(C - C_s)^j f(L) G_K E \left[\frac{\text{kg}}{\text{m}^3 \text{s}} \right] \quad (5.1.2.2)$$

gdzie: $\bar{b}$ — współczynnik temperaturowy $\left[\frac{\text{mol}}{\text{sm}^2} \right] \left[\frac{\text{m}^3}{\text{kg}} \right]^j$, $f(L)$ — funkcja wpływu warunków powierzchniowych od zarodkowania wtórnego na szybkość wzrostu kryształów $[\text{m}^2]$, j — rząd przesyconienia dla reakcji wzrostu kryształów zależnie od substancji krystalizującej, L — wektor rozmiaru liniowego kryształu $[\text{m}]$, [1–11, 14–15, 17, 21–23, 27–29].

5.2. ZAGADNIENIA LOKALNE DLA KOMPLETNEGO PROCESU KRYSTALIZACJI

W sekcjach 4.1. oraz 4.2. podano, że istnieją wspólne i ciągłe zjawiska decydujące o procesie zarodkowania pierwotnego z jednej strony i o procesie wzrostu kryształów z zarodkowaniem wtórnym z drugiej strony. Literatura wymienia następujące zjawiska fizyczne wspólne dla procesu zarodkowania pierwotnego i wzrostu kryształów:

dyfuzyjny transport stężenia w układach: roztwór — roztwór i roztwór — kryształ,

dyfuzyjny transport entalpii w układach: roztwór — roztwór i roztwór — kryształ,

przewodnictwo cieplne w układach: roztwór — roztwór i roztwór — kryształ,

pole wektora prędkości w układach: roztwór — roztwór i roztwór — kryształ oraz warunek pływania kryształów w roztworze.

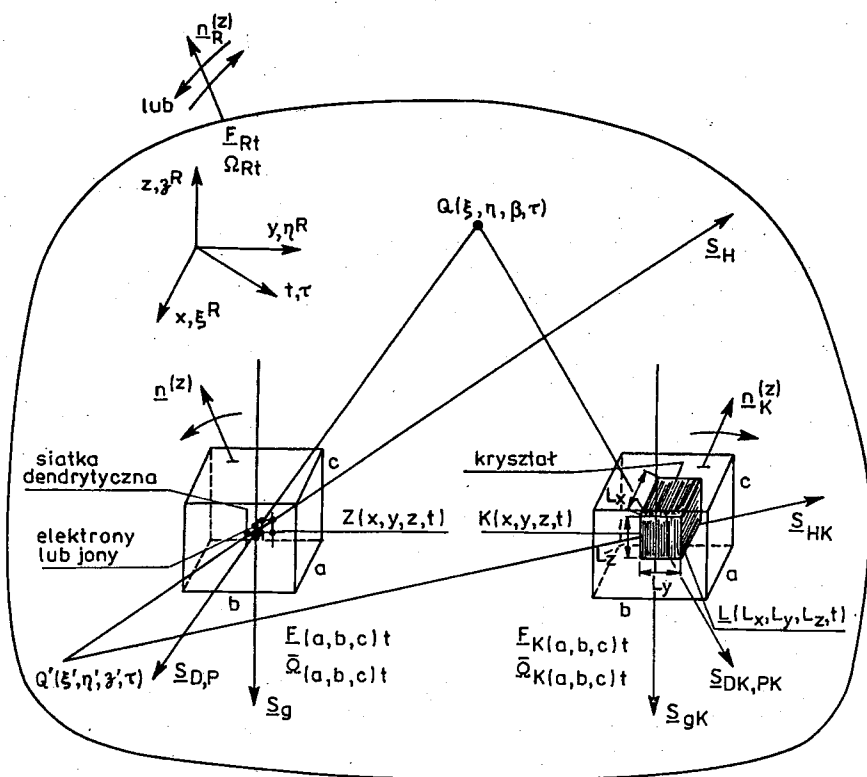
Reakcja zarodkowania pierwotnego oraz reakcja wzrostu kryształów zakłócona w kinetyce zarodkowaniem wtórnym odbywają się jednocześnie w przestrzeni roboczej i czasie Ω_{Rt} . Ich lokalne objętości $\bar{\Omega}_{(a_1, b_1, c_1)t}$ oraz $\bar{\Omega}_{K(a_2, b_2, c_2)t}$ poprzez wspólne zjawiska muszą posiadać ten sam obraz wynikający z zasady wzajemności dla ośrodków anizotropowych i niejednorodnych [31–32, 37, 21–29]. Wynika stąd ogólne założenie przyjęte w niniejszej pracy:

Założenie ZG1: Dla jednego układu wzajemnościowego przyjmuje się te same wymiary elementów $\bar{\Omega}_{(a,b,c)t}$ oraz $\bar{\Omega}_{K(a,b,c)t}$ ($a_1=a_2=a$, $b_1=b_2=b$, $c_1=c_2=c$) t i jest to związane z tym, że przyjęto punktową reprezentację równoważnych reakcji zarodkowania pierwotnego i wzrostu kryształów.

Ponieważ dyfuzja i przewodnictwo cieplne należą do pól potencjalnych — zaś pole wektora prędkości ma charakter wirowy (oddziaływanie sił grawitacyjnych jest uwzględnione osobno) wprowadzono dodatkowo założenie:

Założenie ZG2: Punkt zmienny $Q^R(\xi^R, \eta^R, \beta^R, \tau) \in \Omega_{Rt}$ posiada swoją interpretację dla pól potencjalnych

$$Q^R(\xi^R, \eta^R, \beta^R, \tau) \equiv Q(\xi, \eta, \beta, \tau) \quad (5.2.1)$$



Rys. 4. Koncepcja podwójnie miejscowo rozłożonych parametrów jako nowa metoda modelowania matematycznego procesu ciągłej krystalizacji masowej.

- $Q(\xi, \eta, \beta, \tau)$ — wzajemnościowy zmienny punkt dla zjawisk fizycznych pól potencjalnych,
 $S_{D,P}$ — strumień dyfuzja/przewodnictwo cieplne roztwór — roztwór,
 $S_{DK,PK}$ — strumień dyfuzja/przewodnictwo cieplne roztwór — kryształ,
 $Q'(\xi', \eta', \beta', \tau)$ — wzajemnościowy zmienny punkt dla zjawisk fizycznych pól wirowych,
 S_H — strumień turbulencji roztwór — roztwór,
 S_{HK} — strumień turbulencji roztwór — kryształ,
 S_g — strumień sił grawitacyjnych roztworu,
 S_{gK} — strumień sił grawitacyjnych kryształu

oraz dla pól wirowych

$$Q^R(\xi^R, \eta^R, \beta^R, \tau) \equiv Q'(\xi', \eta', \beta', \tau) \quad (5.2.2)$$

zagadnieniem istotnym do omówienia jest istnienie i ciągłe działanie pola sił ciężkości, które działają tak samo na zarodek „*in statu nascendi*” jak i na kryształ wewnątrz warstwy fluoidalnej lub adsorbcyjnej. Wynika stąd ogólne założenie

Założenie ZG3: Siły grawitacyjne działają ciągle na zarodek „*in statu nascendi*” oraz na rosnący kryształ.

Kompleksowe zagadnienia wzajemnościowe dla całego procesu krystalizacji są przedstawione na rys. 4. [21 – 26, 29].

Rysunek 4 zawiera tzw. koncepcję „podwójnie miejscowo rozłożonych parametrów” jako nową metodę całościowej analizy krystalizacji. Stanowi ona punkt wyjścia do konstytutywnego opisu całego procesu krystalizacji przy pomocy bilansów masy, energii i pędu. Na rys. 4 pokazano, że cyrkulacje wektorów $\mathbf{n}^{(z)}$ oraz $\mathbf{n}_K^{(z)}$ są wzajemnie przeciwne, zaś wybór cyrkulacji wektora $\mathbf{n}_K^{(z)}$ [38] zależy od zadań sterujących wydajnością i jakością procesu krystalizacji przy pomocy warunków brzegowych na powierzchni F_{Rt} . Rysunek 4 posiada interpretację strumieni pól potencjalnych i wirowych oraz sił grawitacyjnych.

6. FORMALIZM MATEMATYCZNY PRZYJĘTY DO OPISU PROCESU KRYSTALIZACJI

Rozdział ten poświęcony jest formalizacji matematycznej zjawisk mających wpływ na wydajność i jakość produktów procesu krystalizacji poprzez jego składowe procesy. Uwzględniamy przy tym lokalne i globalne aspekty oddziaływań kinetycznych i zjawiskowych zachodzących procesów składowych krystalizacji.

6.1. ZARODKOWANIE PIERWOTNE

Lokalna funkcja intensywności zarodkowania pierwotnego związana jest zmiennymi stanu oraz ich gradientami

$$V_B = V_B [C, \nabla C, T, \nabla T, \mathbf{v}, \nabla \mathbf{v}] \quad (6.1.1)$$

W Tablicy 1 umieszczono opis matematyczny kinetyki i zjawisk towarzyszących, decydujących o bilansach masy, energii i pędu wg. procedury bilansowej dla wektorów stanu o współrzędnych masa, energia i pęd:

$$\begin{aligned} \left[\begin{array}{c} \text{Przyrost} \\ \text{(faza płynna)} \end{array} \right] &= \left[\begin{array}{c} \text{Zjawiska pól potencjalnych} \\ \text{(faza płynna)} \end{array} \right] + \left[\begin{array}{c} \text{Zjawiska pola wirowego} \\ \text{(faza płynna)} \end{array} \right] + \\ &+ \left[\begin{array}{c} \text{Źródło} \\ \text{(faza płynna)} \end{array} \right] \end{aligned} \quad (6.1.2)$$

Tablica 1 zawiera wyrażenia definiujące strumienie zjawisk odpowiadające bilansom i zmiennym stanu. Stosuje się do nich tw. Gaussa oraz sumuje się z efektami objętościowymi odpowiednio do bilansu.

Tablica 1

Matematyczne formuły fizyko—chemicznych aspektów procesu zarodkowania pierwotnego

Bilanse Zjawiska	Masa	Energia	Pęd
Dyфуzja	$j_{Dc} = -D_c \text{ grad } C$ $\left[\frac{\text{kg}}{\text{m}^2 \text{s}} \right]$	$I_{Dc} = H_{Dc} \iiint_{\Omega_{(x,y,z)} t} \text{div}(D_c \text{ grad } C) \cdot d\Omega_{(x,y,z)} t$ [W]	
Konwekcja	$i_c = C v$ $\left[\frac{\text{kg}}{\text{m}^2 \text{s}} \right]$	$i_c = CH v$ $\left[\frac{\text{W}}{\text{m}^2} \right]$	$N_c^i = - \iiint_{\Omega_{(x,y,z)} t} [v \text{div}(Cv) + (Cv \text{grad}) v] d\Omega_{(x,y,z)}$ [N]
Przewodnictwo cieplne	$i_{1c} = -\lambda_c \text{ grad } T$ $\left[\frac{\text{W}}{\text{m}^2} \right]$		
Lepkość			$N_c^e = \iiint_{\Omega_{(x,y,z)} t} [\eta_w \nabla^2 v + \frac{\eta_w}{3} \text{grad div}] d\Omega_{(x,y,z)}$ [N]
Źródło	$J_\Omega = - \iiint_{\Omega_{(x,y,z)} t} V_B G d\Omega_{(x,y,z)}$ $\left[\frac{\text{kg}}{\text{s}} \right]$	$I_\Omega = - \iiint_{\Omega_{(x,y,z)} t} H_B V_B G d\Omega_{(x,y,z)}$ [W]	$N_\Omega = \iiint_{\Omega_{(x,y,z)} t} C_G \Omega_{(x,y,z)}$ [N]
Przyrost	$\frac{\partial}{\partial t} \iiint_{\Omega_{(x,y,z)} t} C d\Omega_{(x,y,z)}$ $\left[\frac{\text{kg}}{\text{s}} \right]$	$\frac{\partial}{\partial t} \iiint_{\Omega_{(x,y,z)} t} (CH) d\Omega_{(x,y,z)}$ [W]	$\frac{\partial}{\partial t} \iiint_{\Omega_{(x,y,z)} t} (Cv) d\Omega_{(x,y,z)}$ [N]

Tabela 2

Matematyczne formuły fizyko – chemii procesu wzrostu kryształów

Bilanse Zjawiska	Masa	Energia	Peł
Dyфуzja	$j_D = -D_K \text{grad}[C(1-E)]$ $\left[\frac{\text{kg}}{\text{m}^2 \text{s}} \right]$	$I_{DK} = H_c \iiint_{\Omega_K(x,y,z)} \text{div}(-D_K \text{grad} \cdot [C(1-E)]) d\Omega_{K(x,y,z)}$ [W]	
Konwekcja	$j_K = \rho_K E v_K$ $\left[\frac{\text{kg}}{\text{m}^2 \text{s}} \right]$	$i_K = \rho_K E H_K v_K$ $\left[\frac{\text{W}}{\text{m}^2} \right]$	$N_K^E = \iiint_{\Omega_K(x,y,z)} \rho_K [v_K \text{div}(E v_K) + (E v_K \text{grad}) v_K] d\Omega_{K(x,y,z)}$ [N]
Przewodnictwo cieplne	$i_{LK} = -\lambda_K \text{grad}[T_K(1-E)]$ $\left[\frac{\text{W}}{\text{m}^2} \right]$		
Lepkość			$N_O^K = \iiint_{\Omega_K(x,y,z)} \eta_w E v^2 v_K + \frac{\eta_w}{3} E g \text{grad div } v_K] d\Omega_{K(x,y,z)}$ [N]
Źródło	$J_O^K = \iiint_{\Omega_K(x,y,z)} V_{GK} G_K E d\Omega_{K(x,y,z)}$ $\left[\frac{\text{kg}}{\text{s}} \right]$	$I_O^K = \iiint_{\Omega_K(x,y,z)} H_{GK} V_{GK} G_K E \cdot d\Omega_{K(x,y,z)}$ [W]	$N_O^K = \iiint_{\Omega_K(x,y,z)} Q_K E + C(1-E)] g \cdot d\Omega_{K(x,y,z)}$ [N]
Przyrost	$\frac{\partial}{\partial t} \iiint_{\Omega_K(x,y,z)} (\rho_K E) d\Omega_{K(x,y,z)}$ $\left[\frac{\text{kg}}{\text{s}} \right]$	$\frac{\partial}{\partial t} \iiint_{\Omega_K(x,y,z)} (\rho_K E H_K) d\Omega_{K(x,y,z)}$ [W]	$\frac{\partial}{\partial t} \iiint_{\Omega_K(x,y,z)} (\rho_K E v_K) d\Omega_{K(x,y,z)}$ [N]

6.2. WZROST KRYSTAŁÓW

Lokalna funkcja intensywności wzrostu kryształów określona jest poprzez zmienne stanu oraz ich gradienty

$$V_{GK} = V_{GK}[C, VC, E, VE, T_K, VT_K, v_K, Vv_K] \quad (6.2.1)$$

Tablica 2 zawiera kinetyczne i zjawiskowe elementy decydujące o bilansach masy, energii i pędu wg. procedury sumowania wektorów stanu, masa kryształu, energia kryształu i pęd kryształu:

$$\begin{aligned} \left[\begin{array}{c} \text{Przyrost} \\ \text{(faza krystaliczna)} \end{array} \right] &= \left[\begin{array}{c} \text{Zjawiska pól potencjalnych} \\ \text{(faza krystaliczna)} \end{array} \right] + \left[\begin{array}{c} \text{Zjawiska pola wirowego} \\ \text{(faza krystaliczna)} \end{array} \right] + \\ &+ \left[\begin{array}{c} \text{Źródło} \\ \text{(faza krystaliczna)} \end{array} \right] \end{aligned} \quad (6.2.2)$$

W tablicy 2 umieszczono strumienie zjawisk z punktu widzenia użycia tw. Gaussa oraz ich efekty objętościowe odpowiednio do bilansu [1–11], 14–15, 17, 21–23, 27–29].

6.3. ZARODKOWANIE WTÓRNE

Funkcja $f(L)$ opisująca zagadnienie zarodkowania wtórnego w kinetyce wzrostu kryształów jest związana ze zmiennymi stanu i jej pochodnymi

$$f = f\left[L, \frac{\partial L}{\partial t}, \nabla L\right]. \quad (6.3.1)$$

Z założenia ZWK5, że kryształ po zderzeniach zachowuje swój kształt otrzymujemy równanie, które wyraża zależność (6.3.1) między energiami: E_z — energia kinetyczna zderzeń, E_w — energia wibracji, E_s — energia statyczna dla rosnącego kryształu.

Tablica 3

Matematyczne formuły mechanicznych energii dla procesu zarodkowania wtórnego

Energie mechaniczne kryształu	Energia kinetyczna zderzeń kryształu	Energia wibracji kryształu	Statyczna energia kryształu
Formuły			
Równania	$E_z = n\rho_K K_V L v_K^2$	$E_w = \frac{B}{2\rho_K} \frac{\partial^2 L}{\partial t^2}$	$E_s = h \left[\nabla^2 L + \frac{k}{k-2} \text{graddiv} L \right]$
	$\left[\frac{N}{m^3} \right]$	$\left[\frac{N}{m^3} \right]$	$\left[\frac{N}{m^3} \right]$

W tablicy 3 umieszczono dokładne postacie wyrażeń matematycznych dla energii E_z , E_w i E_s [18–23, 38, 28–29].

7. KONSTRUKCJA RÓWNAŃ KONSTITUTYWNYCH

Równania różniczkowe cząstkowe konstytutywne stanu dla ośrodka ciągłego powinny posiadać następujące cechy:

- powinny opierać się na zmiennych stanu opisujących dany proces,
- powinny uwzględniać fizyczne aspekty ośrodka,
- powinny spełniać zasadę inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami przestrzenną i czasową.

Założenie ZG4. Ciśnienie nie ma wpływu na reakcje zarodkowania pierwotnego i wzrostu kryształów zakłócone w kinetyce zarodkowaniem wtórnym.

Założenie ZG5. Stężenie roztworu krystalizującego definiuje się jako [kg substancji rozpuszczonej/m³ roztworu]. Gęstość rozpuszczalnika jako nośnika stężenia traktuje się jako układ odniesienia w rozważaniach bilansowych.

7.1. RÓWNANIA KONSTITUTYWNE ZARODKOWANIA PIERWOTNEGO

7.1.1. Bilans stężenia roztworu

Rozważamy współrzędne „masa” zawarte w Tablicy 1, otrzymując wyjściowe równanie całkowe bilansu stężenia w postaci:

$$\frac{\partial}{\partial t} \iiint_{\Omega_{(x,y,z)t}} C d\Omega_{(x,y,z)} = \iint_{F_{(x,y,z)t}} [-D_c \text{grad} C + C_v] dF_{(x,y,z)} - \iiint_{\Omega_{(x,y,z)t}} V_B G d\Omega_{(x,y,z)} \quad (7.1.1.1)$$

Uwzględniając wpływ cyrkulacji wektora $\mathbf{n}^{(z)}$ na sumowanie strumieni pól potencjalnych i wirowych oraz tw. Gaussa, równanie (7.1.1.1) otrzymuje postać [38]:

$$\frac{\partial}{\partial t} \iiint_{\Omega_{(x,y,z)t}} C d\Omega_{(x,y,z)} = \iiint_{\Omega_{(x,y,z)t}} [\text{div}(D_c \text{grad} C) - \text{div}(C_v)] d\Omega_{(x,y,z)} - \iiint_{\Omega_{(x,y,z)t}} V_B G d\Omega_{(x,y,z)} \quad (7.1.1.2)$$

Porównując operacje pod całkami potrójnymi mamy

$$\frac{\partial C}{\partial t} = \text{div}(D_c \text{grad} C) - \text{div}(C_v) - V_B G \quad (7.1.1.3)$$

a po przekształceniach

$$\frac{\partial C}{\partial t} - \text{grad} D_c \text{grad} C + v \text{grad} C = D_c \nabla^2 C - C \text{div} v - V_B G. \quad (7.1.1.4)$$

Wprowadzając definicje pochodnej śledczej [33, 21 – 26, 29]

$$\frac{DC}{Dt} = \frac{\partial C}{\partial t} - \text{grad} D_c \text{grad} C + v \text{grad} C \quad (7.1.1.5)$$

równanie (7.1.1.4) przyjmuje ostateczną postać

$$\frac{DC}{Dt} = D_c \nabla^2 C - C \text{div} v - V_B G. \quad (7.1.1.6)$$

Stosując warunek nieściśliwości dla roztworu lub też ze stałości obrazu wzajemnościowego dla elementu $\bar{\Omega}_{(a,b,c)t}$ danego warunkiem $\frac{DC}{Dt} = 0$ [33, 34, otrzymujemy równania ciągłości stężenia jako inwariancyjną grupę ciągłości [21 – 26, 29]:

— dla stałego współczynnika $D_c \neq D_c(x, y, z, t)$

$$\frac{\partial C}{\partial t} = -v \text{grad} C \quad (7.1.1.7)$$

— dla zmiennego współczynnika $D_c = D_c(x, y, z, t)$

$$\frac{\partial C}{\partial t} = \text{grad} D_c \text{grad} C - v \text{grad} C. \quad (7.1.1.8)$$

7.1.2. Bilans energii termicznej roztworu

Biorąc pod uwagę założenia ogólne ZG1 oraz ZG2 a także współrzędne „energia” z Tablicy 1, wyjściowa postać bilansu energii ma postać

$$\begin{aligned} \frac{\partial}{\partial t} \iiint_{\Omega_{(x,y,z)t}} (CH) d\Omega_{(x,y,z)} = & \iint_{F_{(x,y,z)t}} [-\lambda_c \text{grad} T + CHv] dF_{(x,y,z)} + H_c \iiint_{\Omega_{(x,y,z)t}} \\ & \text{div}(D_c \text{grad} C) d\Omega_{(x,y,z)} - \iiint_{\Omega_{(x,y,z)t}} H_B V_B G d\Omega_{(x,y,z)} \end{aligned} \quad (7.1.2.1)$$

Uwzględniając cyrkulację wektora $n^{(z)}$ oraz tw. Gaussa otrzymujemy [38]

$$\begin{aligned} \frac{\partial}{\partial t} \iiint_{\Omega_{(x,y,z)t}} (CH) d\Omega_{(x,y,z)} = & \iiint_{\Omega_{(x,y,z)t}} [\text{div}(\lambda_c \text{grad} T) - \text{div}(CHv)] d\Omega_{(x,y,z)} + \\ & + H_c \iiint_{\Omega_{(x,y,z)t}} \text{div}(D_c \text{grad} C) \cdot d\Omega_{(x,y,z)} - \iiint_{\Omega_{(x,y,z)t}} H_B V_B G d\Omega_{(x,y,z)} \end{aligned} \quad (7.1.2.2)$$

a porównując wyrażenia podcałkowe w równaniu (7.1.2.2) mamy

$$\frac{\partial}{\partial t} (CH) = \text{div}(\lambda_c \text{grad} T) - \text{div}(CHv) + H_c \text{div}(D_c \text{grad} C) - H_B V_B G. \quad (7.1.2.3)$$

Równanie (7.1.2.3) przekształcamy do postaci

$$\begin{aligned} H \frac{\partial C}{\partial t} + C \frac{\partial H}{\partial t} = & \text{grad} \lambda_c \text{grad} T + \lambda_c \nabla^2 T - CH \text{div} v - v C \text{grad} H - v H \text{grad} C - \\ & - H_B V_B G + H_c \text{grad} D_c \text{grad} C + H_c D_c \nabla^2 C. \end{aligned} \quad (7.1.2.4)$$

Do równania (7.1.2.4) wprowadzamy obecnie równanie ciągłości dla stałego współczynnika (7.1.1.7) pomnożone obustronnie przez entalpię „H”

$$H \frac{\partial C}{\partial t} = -v H \text{grad} C. \quad (7.1.2.5)$$

Odejmując obecnie od równania (7.1.2.4) równanie (7.1.2.5) otrzymuje się wynik

$$\begin{aligned} C \frac{\partial H}{\partial t} = & \text{grad} \lambda_c \text{grad} T + \lambda_c \nabla^2 T - CH \text{div} v - v C \text{grad} H - H_B V_B G + \\ & + H_c \text{grad} D_c \text{grad} C + H_c D_c \nabla^2 C. \end{aligned} \quad (7.1.2.6)$$

Następnym krokiem naszych rozważań jest wprowadzenie różniczki zupełnej elementu „CH” o składowych [37, 21–26, 29]:

— czasowej

$$C \frac{\partial H}{\partial t} = C c_p \frac{\partial T}{\partial t} + H_c \frac{\partial C}{\partial t} \quad (7.1.2.7)$$

przy czym

$$c_p = \left. \frac{\partial H}{\partial T} \right|_c \left[\frac{\text{J}}{\text{kg}^\circ\text{K}} \right] \text{ oraz } H_c = \left. \frac{\partial(CH)}{\partial C} \right|_T \left[\frac{\text{J}}{\text{kg}} \right],$$

— przestrzennej

$$C \text{grad} H = C c_p \text{grad} T + H_c \text{grad} C. \quad (7.1.2.8)$$

Równanie (7.1.2.8) mnożymy obecnie obustronnie przez „-v” do postaci

$$-v C \text{grad} H = -v C c_p \text{grad} T - v H_c \text{grad} C. \quad (7.1.2.9)$$

Stosując obecnie równania (7.1.2.7) oraz (7.1.2.9) w równaniu (7.1.2.6) równanie to przyjmuje postać

$$\begin{aligned} C c_p \frac{\partial T}{\partial t} + H_c \frac{\partial C}{\partial t} = & \text{grad} \lambda_c \text{grad} T + \lambda_c \nabla^2 T - CH \text{div} v - v C c_p \text{grad} T - \\ & - v H_c \text{grad} C - H_B V_B G + H_c \text{grad} D_c \text{grad} C + H_c D_c \nabla^2 C. \end{aligned} \quad (7.1.2.10)$$

Do równania (7.1.2.10) używamy teraz postać równania ciągłości stężenia dla zmiennego współczynnika (7.1.1.8) przemnożoną obustronnie przez entalpię „ H_c ” do postaci

$$H_c \frac{\partial C}{\partial t} = H_c \text{grad} D_c \text{grad} C - v H_c \text{grad} C. \quad (7.1.2.11)$$

Po odjęciu równania (7.1.2.11) od równania (7.1.2.10) i po przekształceniu otrzymujemy

$$C \frac{\partial T}{\partial t} - \frac{1}{c_p} \text{grad} \lambda_c \text{grad} T + v C \text{grad} T = \frac{\lambda_c}{c_p} \nabla^2 T - C T \text{div} v - \frac{H_B}{c_p} v_B G + \frac{H_c}{c_p} D_c \nabla^2 C. \quad (7.1.2.12)$$

Wprowadzając konsekwentnie definicję pochodnej śledczej [33, 21 – 26, 29]

$$C \frac{DT}{Dt} = C \frac{\partial T}{\partial t} - \frac{1}{c_p} \text{grad} \lambda_c \text{grad} T + v C \text{grad} T \quad (7.1.2.13)$$

do równania (7.1.2.12) mamy

$$C \frac{DT}{Dt} = \frac{\lambda_c}{c_p} \nabla^2 T - C T \text{div} v - \frac{H_B}{c_p} v_B G + \frac{H_c}{c_p} D_c \nabla^2 C. \quad (7.1.2.14)$$

Stosując warunek nieściśliwości lub stałości obrazu wzajemnościowego dla roztworu jako ośrodka ciągłego $C \frac{DT}{Dt} = 0$ [32, 33, 34,] otrzymujemy równania różniczkowe cząstkowe ciągłości temperatury jako inwariancyjną grupę ciągłości w przypadkach [21 – 26, 29]:

— dla stałych współczynników $D_c \neq D_c(x, y, z, t)$ oraz $\lambda_c \neq \lambda_c(x, y, z, t)$

$$\frac{\partial T}{\partial t} = -v \text{grad} T \quad (7.1.2.15)$$

— dla zmiennych współczynników $D_c = D_c(x, y, z, t)$ oraz $\lambda_c \neq \lambda_c(x, y, z, t)$

$$C c_p \frac{\partial T}{\partial t} = \text{grad} \lambda_c \text{grad} T - C c_p v \text{grad} T [+ H_c \text{grad} D_c \text{grad} C]^* \quad (7.1.2.16)$$

[]* — element dołączony dla pokazania ciągłości energii termicznej w strumieniu dyfuzji roztwór – roztwór.

7.1.3. Bilans pędu roztworu

Stosując elementy kolumny „pęd” z Tablicy 1 możemy napisać równanie wyjściowe dla bilansu pędu z uwzględnieniem wpływu przyjętej cyrkulacji wektora orientacyjnego $n^{(2)}$ powierzchni $F_{(x,y,z)t}$ [38]

$$\begin{aligned} \frac{\partial}{\partial t} \iiint_{\Omega_{(x,y,z)t}} (Cv) d\Omega_{(x,y,z)} = & \iiint_{\Omega_{(x,y,z)t}} \left[\eta_w \nabla^2 v + \frac{\eta_w}{3} \text{grad div } v \right] d\Omega_{(x,y,z)} - \\ & - \iiint_{\Omega_{(x,y,z)t}} \left[v \text{div}(Cv) + (Cv \text{grad})v \right] d\Omega_{(x,y,z)} + \iiint_{\Omega_{(x,y,z)t}} Cg d\Omega_{(x,y,z)}. \end{aligned} \quad (7.1.3.1)$$

porównując pod całkami potrójnymi elementy obu stron powyższego równania po prostych przekształceniach otrzymujemy

$$v \frac{\partial C}{\partial t} + C \frac{\partial v}{\partial t} = \eta_w \nabla^2 v + \frac{\eta_w}{3} \text{grad div } v - v \text{div}(Cv) - (Cv \text{grad})v + Cg. \quad (7.1.3.2)$$

Do rozważań wprowadzamy obecnie równanie ciągłości stężenia dla stałego współczynnika $D_e \neq D_e(x, y, z, t)$ o numerze (7.1.1.7) pomnożone obustronnie przez wektor prędkości „v” do postaci

$$v \frac{\partial C}{\partial t} = - v(v \text{grad } C). \quad (7.1.3.3)$$

Odejmując równanie (7.1.3.3) od równania (7.1.3.2) otrzymujemy

$$C \frac{\partial v}{\partial t} = \eta_w \nabla^2 v + \frac{\eta_w}{3} \text{grad div } v - Cv \text{div } v - (Cv \text{grad})v + Cg. \quad (7.1.3.4)$$

Równanie (7.1.3.4) przekształcamy do postaci

$$C \frac{\partial v}{\partial t} + (Cv \text{grad})v = \eta_w \nabla^2 v + \frac{\eta_w}{3} \text{grad div } v - Cv \text{div } v + Cg. \quad (7.1.3.5)$$

Wprowadzając definicję pochodnej śledczej [33, 21–26, 29]

$$C \frac{Dv}{Dt} = C \frac{\partial v}{\partial t} + (Cv \text{grad})v \quad (7.1.3.6)$$

równanie (7.1.3.5) przyjmuje postać śledczą

$$C \frac{Dv}{Dt} = \eta_w \nabla^2 v - Cv \text{div } v + \frac{\eta_w}{3} \text{grad div } v + Cg, \quad (7.1.3.7)$$

którą przekształcamy do

$$C \frac{Dv}{Dt} = \eta_w \nabla^2 v - Cv \text{div } v + Cg + \frac{\eta_w}{3} \text{grad div } v. \quad (7.1.3.8)$$

Stosując warunek nieściśliwości lub stałości obrazu wynikającego z zasady wzajemności dla ośrodków anizotropowych i niejednorodnych $C \frac{Dv}{Dt} = 0$ [33, 34, 32] mamy grupę ciągłości inwariancji konstytutywnej dla przypadków [21–26, 29]:

— dla stałego współczynnika $\eta_w \neq \eta_w(x, y, z, t)$

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \text{grad}) \mathbf{v} \quad (7.1.3.9)$$

— dla zmiennego współczynnika $\eta_w = \eta_w(x, y, z, t)$

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \text{grad}) \mathbf{v}. \quad (7.1.3.10)$$

7.1.4. Dyskusja kompletnej postaci równań konstytutywnych ośrodka ciągłego dla procesu zarodkowania pierwotnego i ich równań ciągłości

Równania konstytutywne bilansów masy, energii i pędu zbiorczo przedstawiają układ równań dla procesu zarodkowania pierwotnego w postaci:

$$\frac{DC}{Dt} = D_c \nabla^2 C - C \text{div} \mathbf{v} - V_B G \quad (1)$$

$$(I) \quad C \frac{DT}{Dt} = \frac{\lambda_c}{c_p} \nabla^2 T - CT \text{div} \mathbf{v} - \frac{H_B}{c_p} V_B G + \frac{H_c}{c_p} D_c \nabla^2 C \quad (2)$$

$$C \frac{D\mathbf{v}}{Dt} = \eta_w \nabla^2 \mathbf{v} - C \text{div} \mathbf{v} + C \mathbf{g} + \frac{\eta_w}{3} \text{grad} \text{div} \mathbf{v} \quad (3)$$

gdzie:

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \text{grad} \text{ — (suma odpowiednich operacji gradientowych) [33, 21–26, 29]}$$

Układ równań różniczkowych cząstkowych konstytutywnych stanu (I) jest związany z układami równań różniczkowych cząstkowych ciągłości z warunku $\frac{D}{Dt} = 0$ [33, 34,

32] dla przypadków [21–26, 29]:

— stałych współczynników $D_c \neq D_c(x, y, z, t)$ oraz $\lambda_c \neq \lambda_c(x, y, z, t)$, a także $\eta_w \neq \eta_w(x, y, z, t)$

$$\frac{\partial C}{\partial t} = -\mathbf{v} \text{grad} C \quad (1')$$

$$(I') \quad \frac{\partial T}{\partial t} = -\mathbf{v} \text{grad} T \quad (2')$$

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \text{grad}) \mathbf{v} \quad (3')$$

lub konsekwentnie,

— zmiennych współczynników $D_c = D_c(x, y, z, t)$ oraz $\lambda_c = \lambda_c(x, y, z, t)$, a także $\eta_w = \eta_w(x, y, z, t)$

$$\frac{\partial C}{\partial t} = \text{grad} D_c \text{grad} C - \mathbf{v} \text{grad} C \quad (1'')$$

$$(I'') \quad C c_p \frac{\partial T}{\partial t} = \text{grad} \lambda_c \text{grad} T - C c_p \mathbf{v} \text{grad} T [+ H_c \text{grad} D_c \text{grad} C]^x$$

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \text{grad}) \mathbf{v} \quad (3'')$$

$[]^x$ — element dołączony tylko dla prezentacji ciągłości energii cieplnej w strumieniu dyfuzji roztwór—roztwór.

Układ równań różniczkowych cząstkowych konstytutywnych stanu w postaciach pochodnych śledczych (I) jest niezależny od tego czy współczynniki fizyczne są stałe czy też zmienne. Ta niezależność wynika z faktów wyprowadzania układów równań różniczkowych cząstkowych ciągłości (I') — współczynniki stałe lub (I'') — współczynniki zmienne, jako grup inwariantnych dla udowodnienia zasady inwariancji konstytutywnej układu (I). Matematycznie układ równań różniczkowych cząstkowych konstytutywnych stanu (I) może być zakwalifikowany jako quasiliniowy ponieważ wszystkie jego najwyższe operacje cząstkowego różniczkowania są w pierwszej potęgce [36, 38, 35]:

$(\nabla^2 C)^{p=1}$, $(\nabla^2 T)^{p=1}$, $(\nabla^2 \mathbf{v})^{p=1}$, $(\text{grad div} \mathbf{v})^{p=1}$ oraz podlegają dekompozycji do liniowych równań różniczkowych cząstkowych zjawisk fizycznych.

7.2. RÓWNANIA KONSTITUTYWNE WZROSTU KRYSTAŁÓW

7.2.1. Bilans ułamka fazy krystalicznej

Biorąc pod uwagę współrzędne „masa” z Tablicy 2 wyjściowe równanie całkowite bilansu fazy krystalicznej jest określone przez:

$$\begin{aligned} \frac{\partial}{\partial t} \iiint_{\Omega_{K(x,y,z)t}} (\rho_K E) d\Omega_{K(x,y,z)} &= \iint_{F_{K(x,y,z)t}} \{ D_K \text{grad}[C(1-E)] - \rho_K E \mathbf{v}_K \} dF_{K(x,y,z)} + \\ &+ \iiint_{\Omega_{K(x,y,z)t}} V_{GK} G_K E d\Omega_{K(x,y,z)}. \end{aligned} \quad (7.2.1.1)$$

Uwzględniając wpływ cyrkulacji wektora $\mathbf{n}_K^{(2)}$ na sumowanie strumieni pól potencjalnych i wirowych oraz twierdzenie Gaussa, równanie (7.2.1.1) przekształca się do postaci [38]

$$\begin{aligned} \frac{\partial}{\partial t} \iiint_{\Omega_{K(x,y,z)t}} (\rho_K E) d\Omega_{K(x,y,z)} &= \iiint_{\Omega_{K(x,y,z)t}} \{ \text{div} [-D_K \text{grad}[C(1-E)]] + \\ &+ \text{div}(\rho_K E \mathbf{v}_K) \} d\Omega_{K(x,y,z)} + \iiint_{\Omega_{K(x,y,z)t}} V_{GK} G_K E d\Omega_{K(x,y,z)}. \end{aligned} \quad (7.2.1.2)$$

Porównując operacje pod całkami potrójnymi otrzymamy

$$\frac{\partial}{\partial t} (\rho_K E) = \operatorname{div}[-D_K \operatorname{grad}[C(1-E)]] + \operatorname{div}(\rho_K E \mathbf{v}_K) + V_{GK} G_K E. \quad (7.2.1.3)$$

Ponieważ stężenie „C” jest narzucone od roztworu i zmian „in statu nascendi” zarodkowania pierwotnego przy:

E — ułamek objętościowy fazy krystalicznej [objętość kryształu/ $\bar{\Omega}_{K(x,y,z)t}$],

(1-E) — ułamek objętościowy roztworu w $\bar{\Omega}_{K(x,y,z)t}$

operacje „gradC” zaniedbujemy w dalszych rozważaniach.

W związku z powyższym równanie (7.2.1.3) przekształca się do postaci

$$\frac{\partial E}{\partial t} = \frac{D_K}{\rho_K} CV^2 E + \operatorname{grad}\left(\frac{D_K}{\rho_K} C\right) \operatorname{grad} E + \mathbf{v}_K \operatorname{grad} E + E \operatorname{div} \mathbf{v}_K + \frac{V_{GK}}{\rho_K} G_K E. \quad (7.2.1.4)$$

Równanie (7.2.1.4.) przekształcamy następnie do:

$$\frac{\partial E}{\partial t} - \operatorname{grad}\left(\frac{D_K}{\rho_K} C\right) \operatorname{grad} E - \mathbf{v}_K \operatorname{grad} E = \frac{D_K}{\rho_K} CV^2 E + E \operatorname{div} \mathbf{v}_K + \frac{V_{GK}}{\rho_K} G_K E. \quad (7.2.1.5)$$

Do równania (7.2.1.5) wprowadzamy definicję pochodnej śledczej [33, 21–26, 29]

$$\frac{DE}{Dt} = \frac{\partial E}{\partial t} - \operatorname{grad}\left(\frac{D_K}{\rho_K} C\right) \operatorname{grad} E - \mathbf{v}_K \operatorname{grad} E \quad (7.2.1.6)$$

i otrzymujemy w ten sposób postać konstytutywną równania (7.2.1.5)

$$\frac{DE}{Dt} = \frac{D_K}{\rho_K} CV^2 E + E \operatorname{div} \mathbf{v}_K + \frac{V_{GK}}{\rho_K} G_K E. \quad (7.2.1.6)$$

Z warunku nieściśliwości kryształu wewnątrz warstwy fluidalnej lub adsorbcyjnej

$\frac{DE}{Dt} = 0$ [33, 34, 32] otrzymuje się równania różniczkowe cząstkowe ciągłości jako

grupę inwariancyjną [21–23, 29]:

— dla stałego współczynnika $D_K C \neq D_K C(x, y, z, t)$

$$\frac{\partial E}{\partial t} = \mathbf{v}_K \operatorname{grad} E \quad (7.2.1.8)$$

— dla zmiennego współczynnika $D_K C = D_K C(x, y, z, t)$

$$\frac{\partial E}{\partial t} = \operatorname{grad}\left(\frac{D_K}{\rho_K} C\right) \operatorname{grad} E + \mathbf{v}_K \operatorname{grad} E \quad (7.2.1.9)$$

7.2.2. Bilans energii termicznej fazy krystalicznej

Przy założeniach ogólnych ZG1 i ZG2 biorąc pod uwagę współrzędne „energia” z Tablicy 2, postać wyjściowa bilansu energii dla procesu wzrostu kryształów jest sformułowana następująco:

$$\frac{\partial}{\partial t} \iiint_{\Omega_{K(x,y,z)}t} (\rho_K E H_K) d\Omega_{K(x,y,z)} = \iint_{F_{K(x,y,z)}t} \{ \lambda_K \text{grad}[T_K(1-E)] - \rho_K E H_K \mathbf{v}_K \} dF_{K(x,y,z)} +$$

$$+ H_c \iiint_{\Omega_{K(x,y,z)}t} \text{div}(\mathbf{D}_K \mathbf{C} \text{grad} E) d\Omega_{K(x,y,z)} + \iiint_{\Omega_{K(x,y,z)}t} H_{GK} V_{GK} G_K E d\Omega_{K(x,y,z)}. \quad (7.2.2.1)$$

Uwzględnienie w równaniu wpływu cyrkulacji wektora $\mathbf{n}_K^{(z)}$ na znaki sumowania pól potencjalnych i wirowych oraz twierdzenie Gaussa prowadzi do [38]

$$\frac{\partial}{\partial t} \iiint_{\Omega_{K(x,y,z)}t} (\rho_K E H_K) d\Omega_{K(x,y,z)} = \iiint_{\Omega_{K(x,y,z)}t} \{ \text{div}[-\lambda_K \text{grad}[T_K(1-E)]] + \text{div}(\rho_K E H_K \mathbf{v}_K) \} \cdot$$

$$\cdot d\Omega_{K(x,y,z)} + H_c \iiint_{\Omega_{K(x,y,z)}t} \text{div}(\mathbf{D}_K \mathbf{C} \text{grad} E) d\Omega_{K(x,y,z)} + \iiint_{\Omega_{K(x,y,z)}t} H_{GK} V_{GK} G_K E d\Omega_{K(x,y,z)}. \quad (7.2.2.2)$$

Porównując wyrażenia podcałkowe w powyższym równaniu otrzymujemy:

$$\frac{\partial}{\partial t} (\rho_K E H_K) = \text{div}[-\lambda_K \text{grad}[T_K(1-E)]] + \text{div}(\rho_K E H_K \mathbf{v}_K) + H_c \text{div}(\mathbf{D}_K \mathbf{C} \text{grad} E) +$$

$$+ H_{GK} V_{GK} G_K E, \quad (7.2.2.3)$$

a po przekształceniach równania (7.2.2.3) mamy:

$$H_K \rho_K \frac{\partial E}{\partial t} + \rho_K E \frac{\partial H_K}{\partial t} = \text{grad}[-\lambda_K(1-E)] \text{grad} T_K - \lambda_K(1-E) \nabla^2 T_K + \rho_K H_K \mathbf{v}_K \cdot$$

$$\cdot \text{grad} E + \rho_K E \mathbf{v}_K \text{grad} H_K + \rho_K E H_K \text{div} \mathbf{v}_K + H_c \text{grad}(\mathbf{D}_K \mathbf{C}) \text{grad} E +$$

$$+ H_c \mathbf{D}_K \mathbf{C} \nabla^2 E + H_{GK} V_{GK} G_K E. \quad (7.2.2.4)$$

Do rozważań wprowadzamy obecnie równanie (7.2.1.8), ciągłości ułamka fazy krystalicznej w postaci przemnożonej przez „ $\rho_K H_K$ ”:

$$\rho_K H_K \frac{\partial E}{\partial t} = \mathbf{v}_K \rho_K H_K \text{grad} E, \quad (7.2.2.5)$$

które następnie odejmujemy od równania (7.2.2.4)

$$\rho_K E \frac{\partial H_K}{\partial t} = \text{grad}[-\lambda_K(1-E)] \text{grad} T_K - \lambda_K(1-E) \nabla^2 T_K + \rho_K E \mathbf{v}_K \text{grad} H_K +$$

$$+ \rho_K E H_K \text{div} \mathbf{v}_K + H_c \text{grad}(\mathbf{D}_K \mathbf{C}) \text{grad} E + H_c \mathbf{D}_K \mathbf{C} \nabla^2 E + H_{GK} V_{GK} G_K E. \quad (7.2.2.6)$$

Wprowadzając teraz do rozważań różniczkę zupełną elementu „ $\rho_K E H_K$ ” [37, 21–26, 29]:

— z pochodną czasową

$$\rho_K E \frac{\partial H_K}{\partial t} = \rho_K E c_{pK} \frac{\partial T_K}{\partial t} + H_c \rho_K \frac{\partial E}{\partial t} \quad (7.2.2.7)$$

gdzie

$$c_{pK} = \frac{\partial H_K}{\partial T_K} \Big|_{(\rho_K E)} \left[\frac{J}{kg^\circ K} \right] \text{ oraz } H_c = \frac{\partial (CH)}{\partial C} \Big|_T = \frac{\partial [(1-E)H_K]}{\partial (1-E)} \Big|_{T_K} \cdot \left[\frac{J}{kg} \right]$$

a także

— z pochodną przestrzenną

$$\rho_K E \text{grad} H_K = \rho_K E c_{pK} \text{grad} T_K + H_c \rho_K \text{grad} E. \quad (7.2.2.8)$$

Równanie (7.2.2.8) mnożymy teraz przez „ v_K ”

$$v_K \rho_K E \text{grad} H_K = v_K \rho_K E c_{pK} \text{grad} T_K + v_K H_c \rho_K \text{grad} E. \quad (7.2.2.9)$$

Kontynuując rozważania stosujemy równania (7.2.2.7) oraz (7.2.2.9) w równaniu (7.2.2.6) otrzymując

$$\begin{aligned} \rho_K E c_{pK} \frac{\partial T_K}{\partial t} + H_c \rho_K \frac{\partial E}{\partial t} = & \text{grad}[-\lambda_K(1-E)] \text{grad} T_K - \lambda_K(1-E) \nabla^2 T_K + v_K \rho_K E c_{pK} \cdot \\ & \cdot \text{grad} T_K + v_K H_c \rho_K \text{grad} E + \rho_K E H_K \text{div} v_K + H_c \text{grad}(D_K C) \text{grad} E + H_c D_K C \nabla^2 E + \\ & + H_{GK} V_{GK} G_K E. \end{aligned} \quad (7.2.2.10)$$

Niezbędne jest teraz użycie równania ciągłości (7.2.1.9) dla zmiennego współczynnika $D_K = D_K(x, y, z, t)$ przemnażając je obustronnie przez entalpię własną „ H_c ”

$$H_c \rho_K \frac{\partial E}{\partial t} = H_c \rho_K \text{grad}(D_K C) \text{grad} E + \rho_K H_c v_K \text{grad} E, \quad (7.2.2.11)$$

które następnie odejmujemy od równania (7.2.2.10) otrzymując

$$\begin{aligned} \rho_K E c_{pK} \frac{\partial T_K}{\partial t} + \text{grad}[\lambda_K(1-E)] \text{grad} T_K - v_K \rho_K E c_{pK} \text{grad} T_K = & -\lambda_K(1-E) \nabla^2 T_K + \\ & + \rho_K E H_K \text{div} v_K + H_c D_K C \nabla^2 E + H_{GK} V_{GK} G_K E. \end{aligned} \quad (7.2.2.12)$$

Przyjmując w równaniu (7.2.2.12) definicję pochodnej śledczej [33, 21–26, 29]

$$\rho_K E c_{pK} \frac{DT_K}{Dt} = \rho_K E c_{pK} \frac{\partial T_K}{\partial t} + \text{grad}[\lambda_K(1-E)] \text{grad} T_K - v_K \rho_K E c_{pK} \text{grad} T_K \quad (7.2.2.13)$$

równanie (7.2.2.12) przyjmuje ostateczną postać

$$E \frac{DT_K}{Dt} = \frac{-\lambda_K}{c_{pK} \rho_K} (1-E) \nabla^2 T_K + E T_K \text{div} v_K + \frac{H_{GK} V_{GK}}{c_{pK} \rho_K} V_{GK} G_K E + \frac{H_c D_K}{c_{pK} \rho_K} C \nabla^2 E. \quad (7.2.2.14)$$

Z warunku nieściśliwości $E \frac{DT_K}{Dt} = 0$ [33, 34, 32] otrzymuje się równania różniczkowe cząstkowe ciągłości dla następujących przypadków [21–23, 29]:

— dla stałych współczynników $[\lambda_K(1-E)] \neq [\lambda_K(1-E)(x, y, z, t)]$ oraz $[D_K C] \neq [D_K C(x, y, z, t)]$

$$\frac{\partial T_K}{\partial t} = v_K \text{grad} T_K \quad (7.2.2.15)$$

— dla zmiennych współczynników $[\lambda_K(1-E)] = [\lambda_K(1-E)(x, y, z, t)]$ oraz $[D_K C] = [D_K C(x, y, z, t)]$

$$\rho_K E c_{pK} \frac{DT_K}{Dt} = -\text{grad}[-\lambda_K(1-E)] \text{grad} T_K + v_K \rho_K E c_{pK} \text{grad} T_K [+ H_c \text{grad}(D_K C) \cdot \text{grad} E]^* \quad (7.2.2.16)$$

[*] — element dołączony dla prezentacji ciągłości energii termicznej w strumieniu dyfuzji roztwór – kryształ.

7.2.3. Bilans pędu fazy krystalicznej

Do rozważań bierzemy elementy „pęd” z Tablicy 2 jako podstawę do tworzenia bilansu pędu. Biorąc pod uwagę cyrkulację wektora $n_K^{(2)}$ wyjściowe równanie bilansowe dla pędu ma postać [38]

$$\begin{aligned} \frac{\partial}{\partial t} \iiint_{\Omega_{K(x,y,z)t}} (\rho_K E v_K) d\Omega_{(x,y,z)} &= \iiint_{\Omega_{K(x,y,z)t}} \left[\eta_w E \nabla^2 v_K + \frac{\eta_w}{3} E \text{grad} \text{div} v_K \right] d\Omega_{K(x,y,z)} + \\ &+ \iiint_{\Omega_{K(x,y,z)t}} \left[v_K \text{div}(E v_K) + (E v_K \text{grad}) v_K \right] d\Omega_{K(x,y,z)} + \iiint_{\Omega_{K(x,y,z)t}} \left[\rho_K E + \right. \\ &\left. + C(1-E) \right] g d\Omega_{K(x,y,z)}. \end{aligned} \quad (7.2.3.1)$$

Porównując w równaniu (7.2.3.1) operacje podcałkowe po przekształceniach otrzymujemy

$$\begin{aligned} v_K \rho_K \frac{\partial E}{\partial t} + \rho_K E \frac{\partial v_K}{\partial t} &= \eta_w E \nabla^2 v_K + \frac{\eta_w}{3} E \text{grad} \text{div} v_K + v_K \rho_K (v_K \text{grad} E) + \\ &+ \rho_K (E v_K \text{grad}) v_K + [\rho_K E + C(1-E)] g. \end{aligned} \quad (7.2.3.2)$$

Do powyższego równania stosujemy równanie ciągłości (7.2.1.8) dla stałego współczynnika $D_K C \neq D_K(x, y, z, t)$ przemnożone obustronnie przez wektor „ v_K ”

$$v_K \rho_K \frac{\partial E}{\partial t} = v_K \rho_K (v_K \text{grad} E). \quad (7.2.3.3)$$

Odejmując równanie (7.2.3.3) od równania (7.2.3.2) otrzymujemy

$$\rho_K E \frac{\partial \mathbf{v}_K}{\partial t} = \eta_w E \nabla^2 \mathbf{v}_K + \frac{\eta_w}{3} E \text{grad} \text{div} \mathbf{v}_K + \rho_K E \mathbf{v}_K \text{div} \mathbf{v}_K + \rho_K (E \mathbf{v}_K \text{grad}) \mathbf{v}_K + \\ + [\rho_K E + C(1-E)] \mathbf{g}. \quad (7.2.3.4)$$

Równanie (7.2.3.4) przekształcamy do postaci

$$\rho_K E \frac{\partial \mathbf{v}_K}{\partial t} - \rho_K (E \mathbf{v}_K \text{grad}) \mathbf{v}_K = \eta_w E \nabla^2 \mathbf{v}_K + \rho_K E \mathbf{v}_K \text{div} \mathbf{v}_K + [\rho_K E + C(1-E)] \mathbf{g} + \\ + \frac{\eta_w}{3} E \text{grad} \text{div} \mathbf{v}_K. \quad (7.2.3.5)$$

Wprowadzając definicję pochodnej śledczej [33, 21–26, 29]

$$\rho_K E \frac{D \mathbf{v}_K}{Dt} = \rho_K E \frac{\partial \mathbf{v}_K}{\partial t} - \rho_K (E \mathbf{v}_K \text{grad}) \mathbf{v}_K \quad (7.2.3.6)$$

równanie (7.2.3.5) przyjmuje postać konstytutywną

$$\rho_K E \frac{D \mathbf{v}_K}{Dt} = \eta_w E \nabla^2 \mathbf{v}_K + \rho_K E \mathbf{v}_K \text{div} \mathbf{v}_K + [\rho_K E + C(1-E)] \mathbf{g} + \\ + \frac{\eta_w}{3} E \text{grad} \text{div} \mathbf{v}_K. \quad (7.2.3.7)$$

Z warunku nieściśliwości $\rho_K E \frac{D \mathbf{v}_K}{Dt} = 0$ [33, 34, 32] otrzymujemy postać równania ciągłości obowiązującego dla następujących przypadków [21–23, 29]

— dla stałego współczynnika $[\eta_w E] \neq [\eta_w E](x, y, z, t)$

$$\frac{\partial \mathbf{v}_K}{\partial t} = (\mathbf{v}_K \text{grad}) \mathbf{v}_K \quad (7.2.3.8)$$

— dla zmiennego współczynnika $[\eta_w E] = [\eta_w E](x, y, z, t)$

$$\frac{\partial \mathbf{v}_K}{\partial t} = (\mathbf{v}_K \text{grad}) \mathbf{v}_K. \quad (7.2.3.9)$$

7.2.4. Dyskusja kompletnej postaci równań konstytutywnych ośrodka ciągłego dla procesu wzrostu kryształów i ich równań ciągłości

Zbiorcza postać równań konstytutywnych ośrodka ciągłego dla procesu wzrostu kryształów może być zapisana w postaci układu równań różniczkowych cząstkowych konstytutywnych stanu następująco:

$$\frac{DE}{Dt} = \frac{D_K}{\rho_K} C \nabla^2 E + E \text{div} \mathbf{v}_K + \frac{V_{GK}}{\rho_K} G_K E \quad (1)$$

$$(II) \quad E \frac{DT_K}{Dt} = - \frac{\lambda_K}{c_{pK} \rho_K} (1-E) \nabla^2 T_K + E T_K \operatorname{div} \mathbf{v}_K + \frac{H_{gK} V_{gK}}{c_{pK} \rho_K} G_K E + \\ + \frac{H_c D_K}{c_{pK} \rho_K} C \nabla^2 E \quad (2)$$

$$\rho_K E \frac{D\mathbf{v}_K}{Dt} = \eta_w E \nabla^2 \mathbf{v}_K + \rho_K E \mathbf{v}_K \operatorname{div} \mathbf{v}_K + [\rho_K E + C(1-E)] \mathbf{g} + \\ + \frac{\eta_w}{3} E \operatorname{grad} \operatorname{div} \mathbf{v}_K \quad (3)$$

przy warunku pływania kryształów wewnątrz przesyconego roztworu

$$\rho_K E \frac{D\mathbf{v}_K}{Dt} = C(1-E) \frac{D\mathbf{v}}{Dt} \quad (4)$$

oraz ogólnej definicji pochodnej śledczej:

$$\frac{D}{Dt} = \frac{\partial}{\partial t} - \mathbf{v}_K \operatorname{grad} + (\text{suma odpowiednich operacji gradientowych [33, 21–26, 29]}).$$

Układowi równań różniczkowych cząstkowych konstytutywnych stanu odpowiadają układy równań różniczkowych cząstkowych ciągłości z warunku dla pochodnych śledczych $\frac{D}{Dt} = 0$ [33, 34, 32] oraz przypadków [21–23, 29]:

— stałych współczynników $[D_K C] \neq [D_K C(x, y, z, t)]$, $[\lambda_K(1-E)] \neq [\lambda_K(1-E)(x, y, z, t)]$ oraz $[\eta_w E] \neq [\eta_w E(x, y, z, t)]$

$$\frac{\partial E}{\partial t} = \mathbf{v}_K \operatorname{grad} E \quad (1')$$

$$(II') \quad \frac{\partial T_K}{\partial t} = \mathbf{v}_K \operatorname{grad} T_K \quad (2')$$

$$\frac{\partial \mathbf{v}_K}{\partial t} = (\mathbf{v}_K \operatorname{grad}) \mathbf{v}_K, \quad (3')$$

— zmiennych współczynników $[D_K C] = [D_K C(x, y, z, t)]$, $[\lambda_K(1-E)] = [\lambda_K(1-E)(x, y, z, t)]$ oraz $[\eta_w E] = [\eta_w E(x, y, z, t)]$

$$\rho_K \frac{\partial E}{\partial t} = \operatorname{grad}(D_K C) \operatorname{grad} E + \rho_K \mathbf{v}_K \operatorname{grad} E \quad (1'')$$

$$\rho_K E c_{pK} \frac{\partial T_K}{\partial t} = - \operatorname{grad} [\lambda_K(1-E)] \operatorname{grad} T_K + \mathbf{v}_K \rho_K E c_{pK} \operatorname{grad} T_K [+ \\$$

$$(II'') \quad + H_c \operatorname{grad}(D_K C) \operatorname{grad} E]^* \quad (2'')$$

$$\frac{\partial v_K}{\partial t} = (v_K \text{grad}) v_K, \quad (3'')$$

[]* — element dołączony dla pokazania ciągłości energii termicznej w strumieniu dyfuzji roztwór – kryształ.

Quasiliniowy układ równań różniczkowych cząstkowych konstytutywnych stanu (II) jest niezmienniczy względem swoich współczynników fizykalnych na mocy układów równań ciągłości (II') lub (II'').

7.3 RÓWNANIE RÓŻNICZKOWE CZĄSTKOWE ZARODKOWANIA WTÓRNEGO

Założenie ZWK5 o postaci matematycznej

$$E_z + E_w = E_s \quad (7.3.1)$$

że kryształ po zderzeniach zachowuje swój kształt, w konsekwencji uwzględnienia postaci strumieni energii kryształu zawartych w Tablicy 3 przyjmuje postać matematyczną [18–23, 38, 28–29]:

$$n\rho_K K_V L v_K^2 + \frac{B}{2\rho_K} \frac{\partial^2 L}{\partial t^2} = h \left[\nabla^2 L + \frac{\bar{k}}{\bar{k}-2} \text{graddiv} L \right]. \quad (7.3.2)$$

8. ANALIZA JAKOŚCIOWA RÓWNAŃ KONSTITUTYWNYCH KRYSTALIZACJI

8.1. RÓWNANIA KONSTITUTYWNE KRYSTALIZACJI

Równania konstytutywne ośrodka ciągłego dla procesu krystalizacji mają odpowiednio do procesów składowych zbiorczą postać:

— dla procesu zarodkowania pierwotnego

$$\frac{DC}{Dt} = D_c \nabla^2 C - C \text{div} v - V_B G \quad (1)$$

$$(I) \quad C \frac{DT}{Dt} = \frac{\lambda_c}{c_p} \nabla^2 T - C T \text{div} v - \frac{H_B}{c_p} V_B G + \frac{H_c}{c_p} D_c \nabla^2 C \quad (2)$$

$$C \frac{Dv}{Dt} = \eta_w \nabla^2 v - C v + C g + \frac{\eta_w}{3} \text{graddiv} v, \quad (3)$$

gdzie:

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + v \text{grad} - (\text{suma odpowiednich operacji gradientowych}) [33, 21–26, 29].$$

i konsekwentnie

— dla procesu wzrostu kryształów

$$\frac{DE}{Dt} = \frac{D_K}{\rho_K} CV^2 E + E \operatorname{div} v_K + \frac{V_{GK}}{\rho_K} G_K E \quad (1)$$

$$(II) \quad E \frac{DT_K}{Dt} = - \frac{\lambda_K}{c_{pK} \rho_K} (1-E) \nabla^2 T_K + E T_K \operatorname{div} v_K + \frac{H_{GK} V_{GK}}{c_{pK} \rho_K} G_K E + \\ + \frac{H_c D_K}{c_{pK} \rho_K} CV^2 E \quad (2)$$

$$\rho_K E \frac{Dv_K}{Dt} = \eta_w E \nabla^2 v_K + \rho_K E v_K \operatorname{div} v_K + [\rho_K E + C(1-E)] g + \\ + \frac{\eta_w}{3} E \operatorname{grad} \operatorname{div} v_K \quad (3)$$

oraz warunek pływania kryształów wewnątrz przesyconego roztworu

$$\rho_K E \frac{Dv_K}{Dt} = C(1-E) \frac{Dv}{Dt}, \quad (4)$$

gdzie:

$$\frac{D}{Dt} = \frac{\partial}{\partial t} - v_K \operatorname{grad} + (\text{suma odpowiednich operacji gradientowych}) [33, 21-26, 29].$$

— zarodkowania wtórne

$$(III) \quad n \rho_K K_V L v_K^2 + \frac{B}{2 \rho_K} \frac{\partial^2 L}{\partial t^2} = h \left[\nabla^2 L + \frac{\bar{k}}{\bar{k}-2} \operatorname{grad} \operatorname{div} L \right]. \quad (1)$$

Układy równań różniczkowych cząstkowych konstytutywnych stanu (I) oraz (II) są quasiliniowe w sensie matematycznej teorii równań różniczkowych cząstkowych ponieważ wszystkie ich najwyższe pochodne cząstkowe są w pierwszej potęgce [36, 38, 35]

$$(\nabla^2 C)^{p-1}, (\nabla^2 E)^{p-1}, (\nabla^2 T)^{p-1}, (\nabla^2 T_K)^{p-1}, (\nabla^2 v)^{p-1}, (\nabla^2 v_K)^{p-1},$$

$$(\operatorname{grad} \operatorname{div} v)^{p-1}, (\operatorname{grad} \operatorname{div} v_K)^{p-1}$$

zaś zgodnie z teorią równań konstytutywnych [34] oraz [32] są traktowane jako liniowe, ponieważ:

- nieliniowości przechodzą do równań ciągłości jako grup inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami czasową i przestrzenną,
- ulegają dekompozycji do liniowych równań różniczkowych cząstkowych opisujących zjawiska fizyczne.

Układy równań różniczkowych cząstkowych konstytutywnych stanu (I) oraz (II) są niezmiennicze względem ich współczynników fizykalnych:

$D_c, \lambda_c, c_p, \eta_w$ — dla procesu zarodkowania pierwotnego, oraz

$D_K C, H_c D_K C, \lambda_K (1-E), c_{pK}, \eta_w E$ — dla procesu wzrostu kryształów, poprzez ich układy równań różniczkowych cząstkowych ciągłości.

Dlatego też współczynniki te traktowane są jako STAŁE przy analitycznym rozwiązaniu układów równań różniczkowych cząstkowych konstytutywnych stanu (I) oraz (II) z równaniem (III.1) jako zakłóceniem.

8.2. INWARIANCJA KONSTITUTYWNA

Zasada inwariancji konstytutywnej w postaci układów równań różniczkowych cząstkowych ciągłości z warunku $\frac{D}{Dt} = 0$ [33, 34, 32] dla procesu zarodkowania pierwotnego ma postać dla przypadków [21–26], 29].

— stałych współczynników fizykalnych $D_c \neq D_c(x, y, z, t)$, $\lambda_c \neq \lambda_c(x, y, z, t)$ oraz $\eta_w \neq \eta_w(x, y, z, t)$

$$\frac{\partial C}{\partial t} = -v \text{grad} C \quad (1')$$

$$(I') \quad \frac{\partial T}{\partial t} = -v \text{grad} T \quad (2')$$

$$\frac{\partial v}{\partial t} = -(v \text{grad})v \quad (3')$$

— zmiennych współczynników fizykalnych $D_c = D_c(x, y, z, t)$, $\lambda_c = \lambda_c(x, y, z, t)$ oraz $\eta_w = \eta_w(x, y, z, t)$

$$\frac{\partial C}{\partial t} = \text{grad} D_c \text{ grad} C - v \text{grad} C \quad (1'')$$

$$(I'') \quad C c_p \frac{\partial T}{\partial t} = \text{grad} \lambda_c \text{ grad} T - C c_p v \text{grad} T [+ H_c \text{grad} D \text{ grad} C]^* \quad (2'')$$

$$\frac{\partial v}{\partial t} = -(v \text{grad})v \quad (3'')$$

[]* — element dołączony tylko dla prezentacji ciągłości energii cieplnej w strumieniu dyfuzji roztwór—roztwór.

Odpowiednio do powyższych rozważań dla procesu wzrostu kryształów z warunku dla pochodnych śledczych $\frac{D}{Dt} = 0$ [33, 34, 32] otrzymuje się układy równań różniczkowych cząstkowych ciągłości jako grupy inwariancji konstytutywnej dla następujących przypadków [21–23, 29]:

— przy stałych współczynnikach fizykalnych $[D_K C] \neq [D_K C(x, y, z, t)]$, $[H_c D_K C] \neq [H_c D_K C](x, y, z, t)$, $[\lambda_K(1-E)] \neq [\lambda_K(1-E)(x, y, z, t)]$ oraz $[\eta_w E] \neq [\eta_w E(x, y, z, t)]$

$$\frac{\partial E}{\partial t} = \mathbf{v}_K \text{grad} E \quad (1')$$

$$(II') \quad \frac{\partial T_K}{\partial t} = \mathbf{v}_K \text{grad} T_K \quad (2')$$

$$\frac{\partial \mathbf{v}_K}{\partial t} = (\mathbf{v}_K \text{grad}) \mathbf{v}_K \quad (3')$$

— przy zmiennych współczynnikach fizycznych $[D_K C] = [D_K C(x, y, z, t)]$, $[H_c D_K C] = [H_c D_K C(x, y, z, t)]$, $[\lambda_K (1 - E)] = [\lambda_K (1 - E)](x, y, z, t)$ oraz $[\eta_w E] = [\eta_w E(x, y, z, t)]$,

$$\rho_K \frac{\partial E}{\partial t} = \text{grad}(D_K C) \text{grad} E + \rho_K \mathbf{v}_K \text{grad} E \quad (1'')$$

$$\rho_K E c_{pK} \frac{\partial T_K}{\partial t} = -\text{grad}[\lambda_K (1 - E)] \text{grad} T_K + \rho_K E c_{pK} \mathbf{v}_K \text{grad} T_K$$

$$(II'') \quad [+H_c \text{grad}(D_K C) \text{grad} E]^* \quad (2'')$$

$$\frac{\partial \mathbf{v}_K}{\partial t} = (\mathbf{v}_K \text{grad}) \mathbf{v}_K \quad (3'')$$

[]* — element dołączony dla pokazania ciągłości energii termicznej w strumieniu dyfuzji roztwór — kryształ.

Interpretację zasady inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami przestrzenną i czasową przeprowadzono dla procesów składowych: zarodkowania pierwotnego oraz wzrostu kryształów przy punktowym zakłóceniu zarodkowaniem wtórnym osobno a następnie dla całego procesu ciągłej krystalizacji masowej. Przy realizacji tego zagadnienia wprowadza się wektory promienie wodzące dla aspektów przestrzennych i czasowych równań różniczkowych cząstkowych ciągłości dla procesu zarodkowania pierwotnego i procesu wzrostu kryształów zakłóconego punktowno w kinetyce poprzez proces zarodkowania wtórnego. Proces zarodkowania pierwotnego posiada przypadki (przy $n^{(2)}$):

— dla stałych współczynników, układ równań różniczkowych cząstkowych ciągłości (I'); S_t — wektor promień wodzący, $[]$ — wektor o module skalara,

— operacje względem czasu „t”

$$S_{t,v} = \left[\frac{\partial C}{\partial t} \right]; \quad S_{t,T} = \left[\frac{\partial T}{\partial t} \right]; \quad S_{t,v} = \left[\frac{\partial \mathbf{v}}{\partial t} \right] \quad (8.2.1)$$

— skalar „C” w polu wektora prędkości „v”

$$S_{v,C} = [-\mathbf{v} \text{grad} C] \quad (8.2.2)$$

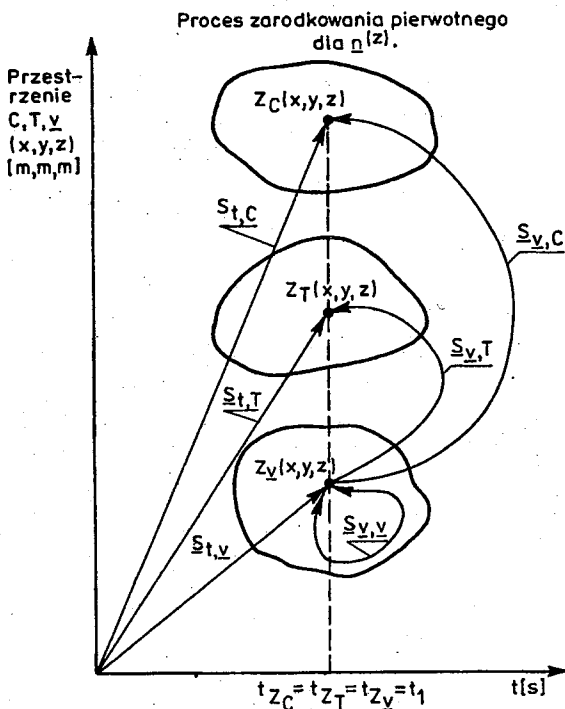
— skalar „T” w polu wektora prędkości „v”

$$S_{v,T} = [-\mathbf{v} \text{grad} T] \quad (8.2.3)$$

— wektor „ v ” w polu wektora prędkości „ v ”

$$S_{v,v} = [-(v \text{ grad})v]. \quad (8.2.4)$$

Na rysunku 5, przedstawiono zagadnienia interpretacji zasady inwariancji konstytutywnej dla powyższych elementów grup inwariancyjnych masy, energii i pędu danych układem równań (I')



Rys. 5. Interpretacja przestrzenno—czasowa zasady inwariancji konstytutywnej dla procesu zarodkowania pierwotnego przy stałych współczynnikach fizycznych

Z rysunku powyższego widać, że spełniony jest warunek

$$t_{Z_C} = t_{Z_T} = t_{Z_v} = t_1 \quad (8.2.5)$$

gdyż:

- brak jest elementów pól potencjalnych o określonych stałych czasowych,
- brak jest znaków „ $\pm$ ” lub „ $\mp$ ” dających zwłokę czasową od zmiany znaków,
- dla zmiennych współczynników, układ równań różniczkowych ciągłości (I'')
- operacje względem czasu „ t ”

$$S_{t,C} = \left[\frac{\partial C}{\partial t} \right]; \quad S_{t,T,C} = \left[C c_p \frac{\partial T}{\partial t} \right]; \quad S_{t,v} = \left[\frac{\partial v}{\partial t} \right] \quad (8.2.6)$$

— skalarna część związana z „ D_c , C ” — zjawisko dyfuzji w bilansie masy

$$S_{D_c, C} = [\text{grad} D_c, \text{grad} C] \quad (8.2.7)$$

— skalarna część związana z „ D_c , C ” — zjawisko dyfuzji w bilansie energii

$$S_{D_c, C, T} = [H_c \text{grad} D_c, \text{grad} C]^* \quad (8.2.8)$$

— skalarna część odpowiadająca „ λ_c , T ” — zjawisko przewodnictwa cieplnego w bilansie energii

$$S_{\lambda_c, T} = [\text{grad} \lambda_c, \text{grad} T] \quad (8.2.9)$$

— skalarna część odpowiadająca „ C ” — w polu wektora prędkości „ v ”

$$S_{v, C} = [-v \text{grad} C] \quad (8.2.10)$$

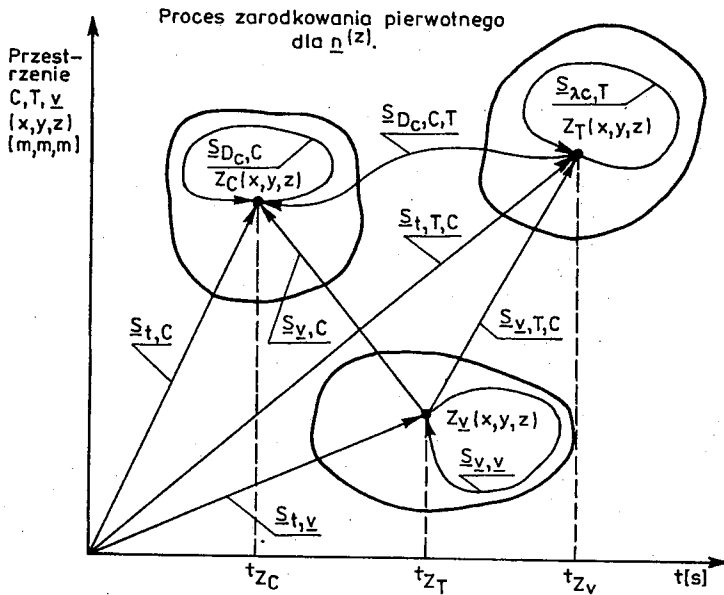
— skalarna część odpowiadająca „ T ” — w polu wektora prędkości „ v ”

$$S_{v, T, C} = [-C c_p v \text{grad} T] \quad (8.2.11)$$

— wektorowa część odpowiadająca „ v ” — w polu wektora prędkości „ v ”

$$S_{v, v} = [-(v \text{grad})v]. \quad (8.2.12)$$

Geometryczna interpretacja zasady inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami przestrzenną i czasową dla równań różniczkowych cząstkowych ciągłości (I'') jest dana na rysunku 6.



Rys. 6. Interpretacja przestrzenno—czasowa zasady inwariancji konstytutywnej dla procesu zarodkowania pierwotnego przy zmiennych współczynnikach fizycznych

Z rysunku 6 widać, że istnieją różne czasy odpowiadające współrzędnym masa, energia i pęd a związane z operacjami różniczkowania „ $\text{grad} D_c$ ” — zjawisko dyfuzji

oraz „grad λ_c ” — zjawisko przewodnictwa cieplnego. Tak więc dla procesu zarodkowania pierwotnego czasowy element pamięci ma postać:

$$t_{Z_C} \neq t_{Z_V} \neq t_{Z_T} \quad (8.2.13)$$

Przedstawione powyżej rozważania dotyczące zagadnień pamięci przestrzenno—czasowej dla procesu zarodkowania pierwotnego posiadają swój odpowiednik dla procesu wzrostu kryształów, przy $n_K^{(2)}$. Tak więc rozważmy dla procesu wzrostu kryształów przypadek stałych współczynników, któremu odpowiada układ równań różniczkowych ciągłości (II'), posiadający następującą wektorową interpretację; S_i — wektor promień wodzący, $[\]$ — wektor o module skalara. Tak więc z układu (II') można wyszczególnić operacje:

— operacje względem czasu „t”

$$S_{i,E} = \left[\frac{\partial E}{\partial t} \right]; \quad S_{i,T_K} = \left[\frac{\partial T_K}{\partial t} \right]; \quad S_{i,v_K} = \left[\frac{\partial v_K}{\partial t} \right] \quad (8.2.14)$$

— skalar „E” w polu wektora prędkości „v_K”

$$S_{v_K,E} = [v_K \text{grad} E] \quad (8.2.15)$$

— skalar „T_K” w polu wektora prędkości „v_K”

$$S_{v_K,T_K} = [v_K \text{grad} T_K] \quad (8.2.16)$$

— wektor „v_K” w polu wektora prędkości „v_K”

$$S_{v_K,v_K} = [(v_K \text{grad}) v_K] \quad (8.2.17)$$

Rozważania powyższe pozwalają sformułować geometryczną interpretację zasady inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami przestrzenną i czasową dla równań różniczkowych cząstkowych ciągłości stanowiących układ (II'), pokazaną na rysunku 7. Punkty $K_E(x, y, z)$, $K_{T_K}(x, y, z)$, $K_{v_K}(x, y, z)$ — miejsca zakłóceń procesem zarodkowania wtórnego.

Rysunek 7 daje możliwość stwierdzenia, że spełniony jest tutaj warunek

$$t_{K_E} = t_{K_{T_K}} = t_{K_{v_K}} = t_2 \quad (8.2.18)$$

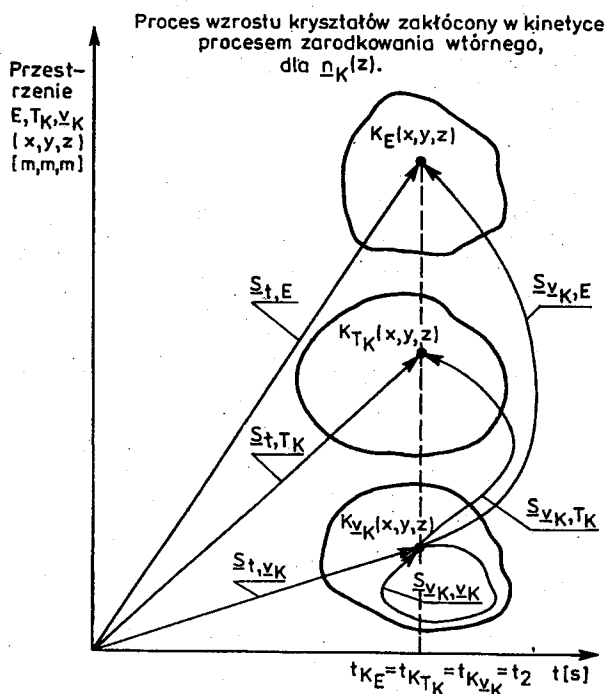
wynikający z:

- braku elementów pól o określonych stałych czasowych,
- braku przemiennych znaków „±” lub „∓” dających zwłokę czasową od tej przemienności znaków.

Przypadek zmiennych współczynników dla procesu wzrostu kryształów bazuje na układzie równań różniczkowych cząstkowych ciągłości (II'') zawierając operacje różniczkowania:

— operacje względem czasu „t”

$$S_{i,E} = \left[\rho_K \frac{\partial E}{\partial t} \right]; \quad S_{i,T_K,E} = \left[\rho_K E c_{pK} \frac{\partial T_K}{\partial t} \right]; \quad S_{i,v_K} = \left[\frac{\partial v_K}{\partial t} \right] \quad (8.2.19)$$



Rys. 7. Interpretacja przestrzenno—czasowa zasady inwariacji konstytutywnej dla procesu wzrostu kryształów zakłóconego w kinetyce procesem zarodkowania wtórnego przy stałych współczynnikach fizycznych

— skalarna część związana z „ D_K , E ” — zjawisko dyfuzji w bilansie masy kryształów

$$S_{D_K, E} = [\text{grad}(D_K C) \text{grad} E] \quad (8.2.20)$$

— skalarna część związana z „ D_K , $E T_K$ ” — zjawisko dyfuzji w bilansie energii kryształów

$$S_{D_K, E, T_K} = [H_c \text{grad}(D_K C) \text{grad} E]^* \quad (8.2.21)$$

— skalarna część odpowiadająca „ λ_K , T_K ” — zjawisko przewodnictwa cieplnego w bilansie energii kryształów

$$S_{\lambda_K, T_K} = [-\text{grad} \lambda_K (1 - E) \text{grad} T_K] \quad (8.2.22)$$

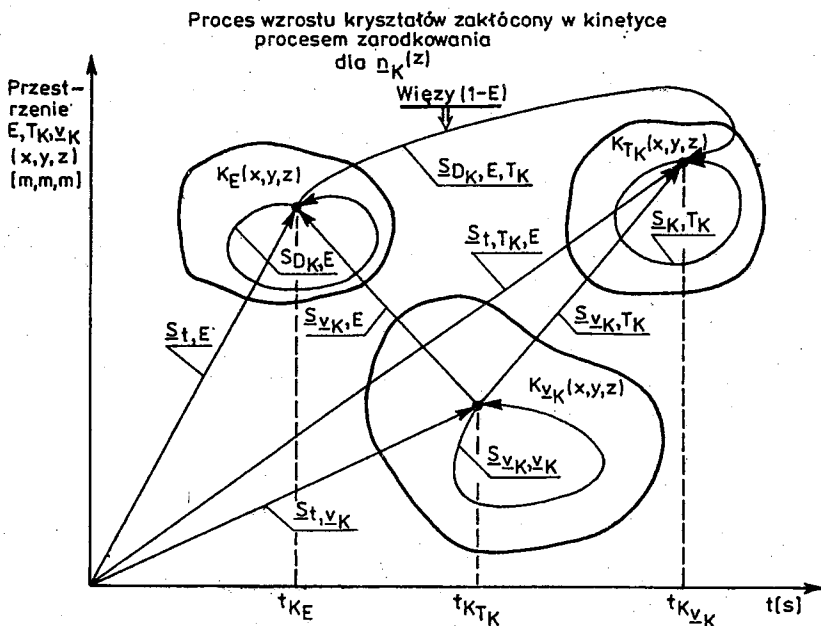
— skalarna część odpowiadająca „ E ” — w polu wektora prędkości „ $\underline{v}_K$ ”

$$S_{\underline{v}_K, E} = [\rho_K \underline{v}_K \text{grad} E] \quad (8.2.23)$$

— skalarna część odpowiadająca „ T_K ” — w polu wektora prędkości „ $\underline{v}_K$ ”

$$S_{\underline{v}_K, T_K, E} = [\rho_K E c_{pK} \underline{v}_K \text{grad} T_K] \quad (8.2.24)$$

— wektorowa część odpowiadająca „ $\underline{v}_K$ ” — w polu wektora prędkości „ $\underline{v}_K$ ”



Rys. 8. Interpretacja przestrzenno–czasowa zasady inwariancji konstytutywnej dla procesu wzrostu kryształów zakłóconego w kinetyce procesem zarodkowania wtórnego przy zmiennych współczynnikach fizycznych

$$S_{v_K, v_K} = [(v_K \text{ grad}) v_K] \quad (8.2.25)$$

Z rysunku 8 widać, że spełniony jest warunek:

$$t_{K_E} \neq t_{K_{v_K}} \neq t_{K_{T_K}} \quad (8.2.26)$$

W oparciu o rysunki 5 i 7 można (dla stałych współczynników) postawić zasadę inwariancji konstytutywnej w ujęciu przestrzenno–czasowym dla całego procesu ciągłej krystalizacji masowej, przy czym dla równoległe przebiegających procesów zarodkowania pierwotnego i wzrostu kryształów:

$$t_1 - t_2 \rightarrow 0 \quad \text{lub} \quad t_2 - t_1 \rightarrow 0 \quad (8.2.27)$$

a stąd

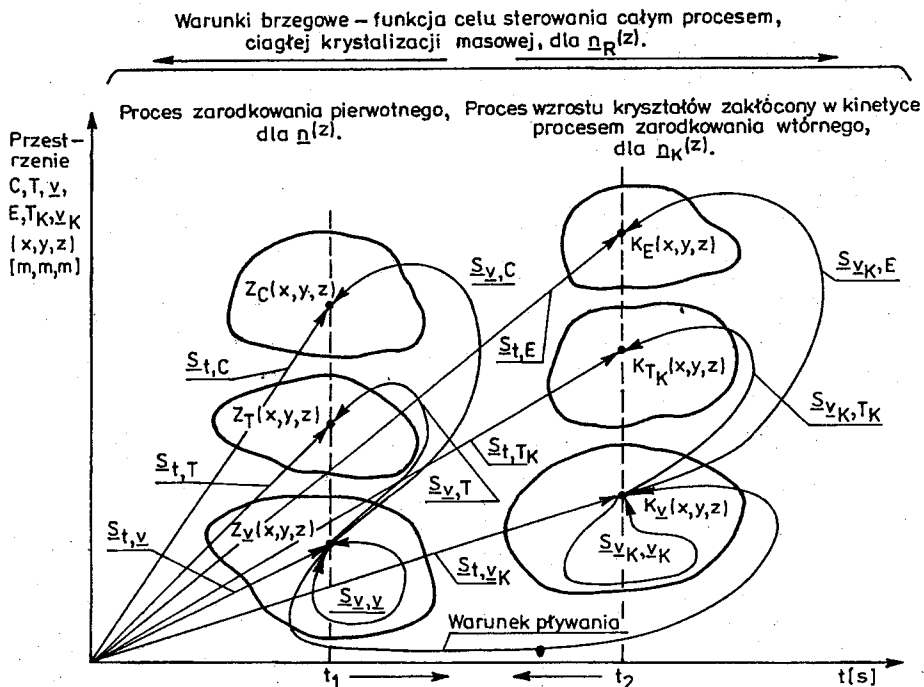
$$t_1 = t_2 = t \quad (8.2.28)$$

co jest przedstawione na rysunku 9.

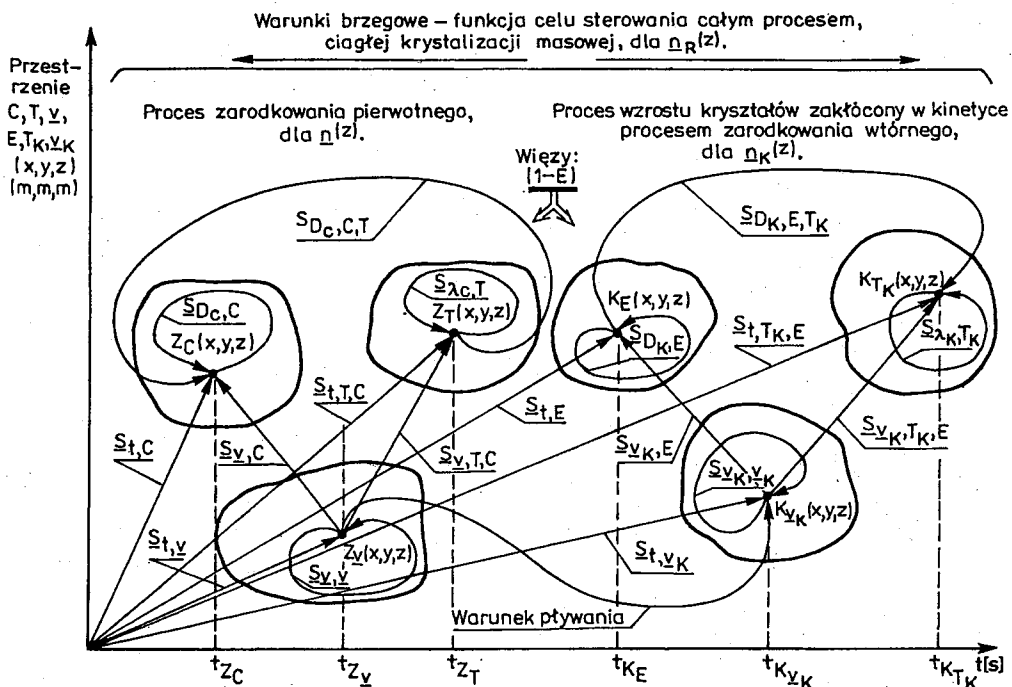
Konsekwentnie do tego całościowego podejścia zasada inwariancji konstytutywnej dla zmiennych współczynników w ujęciu przestrzenno–czasowym dla całego procesu ciągłej krystalizacji masowej, przy równocześnie zachodzących procesach zarodkowania pierwotnego i wzrostu kryształów przedstawiona jest na rysunku 10. Przy czym z rysunku 10 widać zwłoki czasowe dla współrzędnych:

$$\text{masa:} \quad t_{z_c} - t_{K_E} = t_{M1} \quad \text{lub} \quad t_{K_E} - t_{z_c} = t_{M2} \quad (8.2.29)$$

$$\text{energia:} \quad t_{z_r} - t_{K_{T_K}} = t_{H1} \quad \text{lub} \quad t_{K_{T_K}} - t_{z_r} = t_{H2} \quad (8.2.30)$$



Rys. 9. Interpretacja przestrzenno – czasowa zasady inwariancji konstytutywnej dla całego procesu ciągłej krystalizacji masowej przy jego stałych współczynnikach fizykalnych



Rys. 10. Interpretacja przestrzenno – czasowa zasady inwariancji konstytutywnej dla całego procesu ciągłej krystalizacji masowej przy jego zmiennych współczynnikach fizykalnych

pęd: $t_{z_v} - t_{K_v} = t_{P1}$ lub $t_{K_v} - t_{z_v} = t_{P2}$ (8.2.31)

Uwaga U1. Alternatywa „lub” we wzorach (8.2.27) oraz (8.2.29 – 30) związana jest z kolejnością występowania procesów: najpierw „zarodkowanie pierwotne” potem „wzrost kryształów” lub odwrotnie, najpierw „wzrost kryształów” a potem „zarodkowanie pierwotne”, zależnie od rozwiązań technologicznych.

9. ZASADA ADDYTYWNOŚCI W BILANSOWANIU KONSTITUTYWNYM

Dla rozważania addytywności w sensie konstytutywnego bilansowania przy konstytutywnym rozłożonym parametrycznie modelowaniu matematycznym niezbędne jest wprowadzenie pewnych postulatów stanowiących wyjściowe pewniki do formułowania potrzebnych bilansów dla pełnego opisu rozważanego procesu technologicznego z punktu widzenia wydajnościowych i jakościowych aspektów jego produktów. Postulaty te można przedstawić według następującej kolejności:

POSTULAT P1. Istnieją postacie wektorowe zjawisk fizykalnych tworzących konstruowany bilans.

POSTULAT P2. Postacie wektorowe zjawisk fizykalnych spełniają warunki sumowania odpowiednio do kryteriów znaków bilansowych konstruowanego bilansu.

POSTULAT P3. Postacie wektorowe zjawisk fizykalnych tworzących konstruowany bilans dają w stosunku do zorientowanej zewnętrznie powierzchni strumienie tych zjawisk fizykalnych przez rozważaną powierzchnię.

POSTULAT P4. Postacie wektorowe zjawisk fizykalnych tworzących konstruowany bilans oraz ich strumienie spełniają reguły i operacje ogólnej teorii pola.

POSTULAT P5. Rozpuszczalnik jest nośnikiem — a więc jest pasywny w stosunku do reakcji chemicznej lub innego typu, nie jest katalizatorem dla reakcji, a reaguje tylko stężenie [kg substancji rozpuszczonej/m³ roztworu].

POSTULAT P6. Do rozpuszczalnika odnoszą się te same zjawiska fizykalne co do roztworu o stężeniu „C”, [kg/m³].

Biorąc do rozważań postawione powyżej POSTULATY P1 – 6 można dać dla procesu ciągłej krystalizacji masowej następującą sugestię:

Założenie ZG6. Proces zarodkowania pierwotnego ma swój układ odniesienia w postaci rozpuszczalnika natomiast proces wzrostu kryształów zakłócony w kinetyce zarodkowaniem wtórnym odnoszony jest już do roztworu.

Z założenia ZG6 wynika następujący wniosek:

Wniosek W1. Dla rozpuszczalnika definiuje się następujące zmienne stanu: ρ , — gęstość materii rozpuszczalnika [kg/m³], T , — temperatura rozpuszczalnika [°K], v , — prędkość rozpuszczalnika [m/s].

Całość powyższych rozważań pozwala na postawienie następującego twierdzenia:

TWIERDZENIE TW1. Jeżeli proces zarodkowania pierwotnego opisany jest przy pomocy układu równań różniczkowych cząstkowych konstytutywnych stanu (I) to rozpuszczalnik opisuje układ równań różniczkowych cząstkowych konstytutywnych stanu

$$\frac{D\rho_r}{Dt} = D_r \nabla^2 \rho_r - \rho_r \operatorname{div} \mathbf{v}_r$$

$$(I_R) \quad \rho_r \frac{DT_r}{dt} = \frac{\lambda_r}{c_{pr}} \nabla^2 T_r - \rho_r T_r \operatorname{div} \mathbf{v}_r + \frac{H_{cr}}{c_{pr}} D_r \nabla^2 \rho_r$$

$$\rho_r \frac{D\mathbf{v}_r}{Dt} = \eta_{wr} \nabla^2 \mathbf{v}_r - \rho_r \mathbf{v}_r \operatorname{div} \mathbf{v}_r + \frac{\eta_{wr}}{3} \operatorname{grad} \operatorname{div} \mathbf{v}_r$$

$$\left| \begin{array}{c} \rho_{r0} \\ T_{r0} \\ \mathbf{v}_{r0} \end{array} \right| \quad (1)$$

$$+ \quad (2)$$

$$(3)$$

Warunki
początkowe.

gdzie:

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \operatorname{grad} \text{ — (suma odpowiednich operacji gradientowych) [33, 21–26], 29]$$

Dowód: Dowód twierdzenia TW1 polega na wprowadzeniu w konstrukcji układu równań różniczkowych cząstkowych konstytutywnych stanu (I) procesu zarodkowania pierwotnego, następujących zmian

$$C \rightarrow \rho_r + C \quad (4)$$

$$D_c \operatorname{grad} C \rightarrow D_r \operatorname{grad} \rho_r + D_c \operatorname{grad} C \quad (5)$$

$$T \rightarrow D_r + T \quad (6)$$

$$\lambda_c \operatorname{grad} T \rightarrow \lambda_r \operatorname{grad} T_r + \lambda_c \operatorname{grad} T \quad (7)$$

$$Cv \rightarrow \rho_r \mathbf{v}_r + Cv \quad (8)$$

i inne elementy konsekwentnie do zjawisk fizycznych.

Koniec dowodu

Z twierdzenia TW1 wynika następujący istotny wniosek:

WNIOSEK W2. Układ równań różniczkowych cząstkowych konstytutywnych stanu dla rozpuszczalnika (I_R) posiada wszystkie cechy charakterystyczne dla opisu procesu zarodkowania pierwotnego w sposób konstytutywny przy pomocy układu (I).

Układ równań różniczkowych cząstkowych konstytutywnych stanu dla rozpuszczalnika (I_R) jest określany jako oscylacje wokół położenia równowagi podawanej poprzez warunki początkowe: ρ_{r0} , T_{r0} , $\mathbf{v}_{r0}$. niemniej jednak oczywistym wydaje się być następujący wniosek;

WNIOSEK W3. Układ równań różniczkowych cząstkowych konstytutywnych stanu dla rozpuszczalnika (I_R) nie musi być rozwiązywany analitycznie, ponieważ stanowi on tylko układ odniesienia dla roztworu opisywanego w sposób konstytutywny poprzez układ (I). Niemniej jednak układ (I_R) posiada wszystkie cechy jego analitycznej rozwiązalności w stosunku do swoich warunków początkowych.

Należy podkreślić, że dla układów roztworów i ośrodków o wielu aktywnych składnikach zasada addytywności w stosunku do poszczególnych aktywnych składników podana jest w książce L. Schwartz'a dyskutującej zagadnienia metod matematycznych w fizyce [39].

10. ZAGADNIENIA PAMIĘCI PRZESTRZENNO – CZASOWEJ

Zagadnienia związane z zasadą inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami przestrzenną i czasową omówione w rozdziale 8 dają podstawę do przedyskutowania problemów przestrzenno – czasowej pamięci dla:

- poszczególnych bilansów przy opisie procesów składowych,
- poszczególnych procesów składowych,
- całego procesu ciągłej krystalizacji masowej.

W każdym z tych przypadków mamy do czynienia z zagadnieniami lokalnymi: — od reakcji miejscowej zarodkowania pierwotnego lub też reakcji wzrostu kryształów zakłóconej w kinetyce zarodkowaniem wtórnym i globalnymi — związanymi ze zjawiskami fizykalnymi będącymi odpowiedzią ośrodka ciągłego na zmiany współrzędnych źródłowego wektora stanu w punkcie $Z(x, y, z, t)$, lub też w punkcie $K(x, y, z, t)$, wewnątrz przestrzeni roboczej i czasie Ω_{Rt} , przy czym obowiązuje tutaj zasada wzajemności dla anizotropowych i izotropowych niejednorodnych ośrodków z pamięciami przestrzenną i czasową dla zjawisk fizykalnych pól potencjalnych — zmienny punkt $Q(\xi, \eta, \beta, \tau)$ oraz pól wirowych — zmienny punkt $Q'(\xi', \eta', \beta', \tau)$. Poglądy te przedstawiono poniżej

$$\begin{array}{ccccc}
 & & Q(\xi, \eta, \beta, \tau) & & \\
 & \swarrow & & \searrow & \\
 Z(x, y, z, t) & & & & K(x, y, z, t) \\
 & \nwarrow & & \nearrow & \\
 & & Q'(\xi', \eta', \beta', \tau) & &
 \end{array} \quad (10.1)$$

jako formę relacji zagadnień lokalnych i globalnych pamięci przestrzenno – czasowej [21 – 26, 29, 31, 37].

Pamięć lokalna dla $\bar{\Omega}_{(a,b,c)t}$ oraz $\bar{\Omega}_{K(a,b,c)t}$ związana jest ze stałymi czasowymi od reakcji w punktach $Z(x, y, z, t) \in \bar{\Omega}_{(a,b,c)t}$ — reakcji zarodkowania pierwotnego oraz $K(x, y, z, t) \in \bar{\Omega}_{K(a,b,c)t}$ — reakcji wzrostu kryształów zakłócona w kinetyce procesem zarodkowania wtórnego [34].

Pamięć globalna dla $\Omega_{Rt} - \bar{\Omega}_{(a,b,c)t}$ lub też $\Omega_{Rt} - \bar{\Omega}_{K(a,b,c)t}$ związana jest ze stałymi czasowymi zjawisk fizykalnych dla pól potencjalnych i wirowych, odpowiednio do formuły (10.1) [34]. Dla kontynuacji naszych rozważań niezbędne jest wprowadzenie następującego twierdzenia:

TWIERDZENIE TW2. Pamięci lokalna i globalna dla procesów: zarodkowanie pierwotne i wzrost kryształów zakłócony w kinetyce zarodkowaniem wtórnym zależą od warunków fizykalnych otoczeń punktów $Z(x, y, z, t) \in \bar{\Omega}_{(a,b,c)t}$ oraz $K(x, y, z, t) \in \bar{\Omega}_{K(a,b,c)t}$ o wymiarach $(a, b, c)t$.

Dowód

Dowód dotyczy zagadnień lokalnych. Niech istnieje punkt $A(x, y, z, t)$

$$\text{dla którego} \quad A(x, y, z, t) \equiv Z(x, y, z, t) \equiv K(x, y, z, t) \quad (10.2)$$

$$\text{spełniający} \quad A(x, y, z, t) \in \bar{\Omega}_{A(a,b,c)t} \quad (10.3)$$

$$\text{przy czym} \quad \bar{\Omega}_{A(a,b,c)t} \equiv \bar{\Omega}_{(a,b,c)t} \equiv \bar{\Omega}_{K(a,b,c)t} \quad (10.4)$$

Niech ponadto będą zdefiniowane funkcje:

$$\text{— skalarna } \Phi(A) = \Phi(x, y, z, t), \text{ oraz — wektorowa } \Psi(A) = \Psi(x, y, z, t) \quad (10.5)$$

które na brzegu $F_{A(a,b,c)t} \in \bar{\Omega}_{A(a,b,c)t}$ przyjmują postacie:

$$\Phi(x=a, y=b, z=c, t) = \Phi(a, b, c) \Phi(t) \text{ oraz } \Psi(x=a, y=b, z=c, t) = \Psi(a, b, c) \Psi(t) \quad (10.6)$$

i sprowadzają się do:

$$\Phi(a, b, c) = \text{const}, \quad \text{oraz}, \quad \Psi(a, b, c) = \text{const} \quad (10.7)$$

przy		przy
$\Phi(t)$ — źródło skalarne		$\Psi(t)$ — źródło wektorowe
wewnątrz $\bar{\Omega}_{A(a,b,c)t}$		wewnątrz $\bar{\Omega}_{A(a,b,c)t}$

zaś odpowiednie operacje różniczkowania cząstkowego

$$\nabla \Phi(a, b, c) = 0 \quad \text{oraz} \quad \nabla \Psi(a, b, c) = 0 \quad (10.8)$$

Konsekwencją tego podejścia są źródłowe wektory stanu:

z układu (I) dla
punktu $Z(x=a, y=b, z=c, t)$

$$\frac{dC}{dt} = -V_B G \quad (1)$$

$$(I^S) \quad C \frac{dT}{dt} = -\frac{H_B}{c_p} V_B G \quad (2)$$

$$\frac{dv}{dt} = g \quad (3)$$

z układu (II) dla
punktu $K(x=a, y=b, z=c, t)$

$$\frac{dE}{dt} = \frac{V_{GK}}{\rho_K} G_K E \quad (1)$$

$$(II^S) \quad E \frac{dT_K}{dt} = \frac{H_{GK} V_{GK}}{c_{pK} \rho_K} G_K E \quad (2)$$

$$\rho_K E \frac{dv_K}{dt} = [\rho_K E + C(1-E)]g \quad (3)$$

Uwaga U2. Przy całkowaniu równania (I^S2) stosujemy moduł równania (I^S1) — możliwość reakcji odwrotnej do zarodkowania pierwotnego — rozpuszczanie zarodków krystalicznych.

Uwaga U3. Jeżeli zagadnienia źródłowe są ważne dla $Z(a, b, c, t) \in F_{(a,b,c)t}$ oraz $K(a, b, c, t) \in F_{K(a,b,c)t}$ to również ważne wewnątrz $\bar{\Omega}_{(a,b,c)t}$ oraz $\bar{\Omega}_{K(a,b,c)t}$. Identyczny dowód dla zagadnień globalnych polega na poszukiwaniu funkcji brzegowych $\Phi(a, b, c)$ oraz $\Psi(a, b, c)$ od zjawisk fizykalnych w $\Omega_{Rt} - \bar{\Omega}_{A(a,b,c)t}$.

Koniec dowodu

Kontynuując nasze rozważania jako konsekwencję twierdzenia TW2 wprowadza się bardzo istotne następne twierdzenie:

TWIERDZENIE TW3. Funkcje brzegowe skalarna $\Phi(a, b, c) = 0$ oraz wektorowa $\Psi(a, b, c) = 0$ dla brzegów $F_{(a,b,c)t} \in \bar{\Omega}_{(a,b,c)t}$ — proces zarodkowania pierwotnego oraz $F_{K(a,b,c)t} \in \bar{\Omega}_{K(a,b,c)t}$ — proces wzrostu kryształów i proces zarodkowania wtórnego. Jedynymi aktywnymi źródłami działającymi wewnątrz $\bar{\Omega}_{(a,b,c)t}$ oraz $\bar{\Omega}_{K(a,b,c)t}$ są funkcje źródłowe ze scałkowania układów równań różniczkowych zwyczajnych (I^s) oraz (II^s) w przedziale czasu $[0, t)$.

Dowód

Wprowadzamy zewnętrzną i jedyną orientację [38]

$$\mathbf{n}^{(z)} \in F_{(a,b,c)t} \in \bar{\Omega}_{(a,b,c)t} \quad (10.9)$$

oraz

$$\mathbf{n}_K^{(z)} \in F_{K(a,b,c)t} \in \bar{\Omega}_{K(a,b,c)t} \quad (10.10)$$

a także zastępczo

$$A(x, y, z, t) \equiv Z(x, y, z, t) \equiv K(x, y, z, t) \quad (10.11)$$

przy czym

$$\mathbf{n}_A^{(z)} \in F_{A(a,b,c)t} \in \bar{\Omega}_{A(a,b,c)t} \quad (10.13)$$

Dla punktu $A(x=a, y=b, z=c, t)$ definiuje się funkcje brzegowe:

— od źródła wewnątrz $\bar{\Omega}_{A(a,b,c)t}$

$$\Phi^s(x, y, z) = \Phi(a, b, c) \quad \text{oraz} \quad \Psi^s(x, y, z) = \Psi(a, b, c) \quad (10.14)$$

— od zjawiska fizykalnego wewnątrz $\Omega_{Rt} - \bar{\Omega}_{A(a,b,c)t}$

$$\Phi^{zj}(x, y, z) = \Phi(a, b, c) \quad \text{oraz} \quad \Psi^{zj}(x, y, z) = \Psi(a, b, c) \quad (10.15)$$

Przyjmujemy znaki dla powierzchni $F_{A(a,b,c)t}$: znak „+” — strona źródła „s” oraz „-” — strona zjawiska „zj”. Odpowiednio do tego podejścia otrzymujemy wypadkowe wartości funkcji tematowych dla modułu powierzchni $F_{A(a,b,c)t}$ — traktowanej obustronnie [38]:

$$\Phi^s(x=a, y=b, z=c) - \Phi^{zj}(x=a, y=b, z=c) = 0 \quad (10.16)$$

oraz

$$\Psi^s(x=a, y=b, z=c) - \Psi^{sj}(x=a, y=b, z=c) = 0 \quad (10.17)$$

co należało udowodnić.

Koniec dowodu.

11. WNIOSKI KOŃCOWE DOTYCZĄCE KONSTRUKCJI RÓWNAŃ RÓŻNICZKOWYCH CZĄSTKOWYCH KONSTYTUTYWNYCH STANU DLA PROCESU CIĄGŁEJ KRYSTALIZACJI MASOWEJ

W niniejszej pracy wprowadzono kompletny, konstytutywny model matematyczny o parametrach rozłożonych procesu ciągłej krystalizacji masowej. Wszystkie operacje modelowania konstytutywnego poparto odpowiednimi założeniami technologicznymi, założeniami konstrukcji równań różniczkowych cząstkowych konstytutywnych stanu, interpretacją przestrzenno—czasową inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami przestrzenno—czasowymi, a także dyskusją zasady addytywności w bilansowaniu konstytutywnym oraz zagadnieniami pamięci przestrzenno—czasowej.

Dla poparcia poszczególnych tez wprowadzono odpowiednie twierdzenia matematyczne wraz z dowodami umożliwiającymi ich praktyczne użycie przy innym niż proces ciągłej krystalizacji masowej modelowaniu zagadnień technologicznych.

Należy podkreślić, że proces ciągłej krystalizacji masowej jest procesem skomplikowanym przy konstytutywnym rozłożonym parametrycznie modelowaniu ponieważ:

- składa się z trzech konkurencyjnych procesów składowych: zarodkowania pierwotnego oraz wzrostu kryształów zakłóconego w kinetyce zarodkowaniem wtórnym,
- istnieją więzy poprzez zjawiska fizyczne i zmienne stanu między procesami składowymi, a także zmiennymi stanu,
- istnieje układ odniesienia procesu zarodkowania pierwotnego do rozpuszczalnika,
- istnieje skomplikowana kinetyka procesów składowych dająca lokalne źródła dla zjawisk tych procesów,
- inne trudne do opisu konstytutywnego fizyko—chemiczne zagadnienie ujęte w wprowadzonym konstytutywnym modelu.

Inne rzeczywiste procesy technologiczne mają wielokrotnie o wiele bardziej prostą specyfikę co ułatwia użycie wprowadzonej w artykule konstytutywnej metody modelowania matematycznego ich fizycznych aspektów [30].

Reasumując należy stwierdzić, że w rozprawie podano kompletną teorię konstytutywnego rozłożonego parametrycznie modelowania, która powinna znaleźć zastosowanie do opisu każdego procesu technologicznego posiadającego określone aspekty fizyczne. Może ona służyć do doskonalenia już istniejących procesów technologicznych lub służyć do prognozowania przyszłych procesów technologicznych ponieważ ujmuje tak ilościowe jak i jakościowe aspekty produktów tych procesów technologicznych dla ich rozłożonego parametrycznie sterowania [40—41].

SYMBOLIKA

C	— stężenie krystalizującego roztworu	kg/m^3
C_s	— stężenie nasycenia dla krystalizującego roztworu	kg/m^3
C_0	— stężenie początkowe dla krystalizującego roztworu proces zaczyna się dla $C_0 = C_s$	kg/m^3
i	— rząd przesyconia krystalizującego roztworu dla reakcji zarodkowania pierwotnego	
$\bar{a}$	— współczynnik reakcji zarodkowania pierwotnego	$(\text{mol/s})/(\text{m}^3/\text{kg})^i$
G	— własna masa molowa dla reakcji zarodkowania pierwotnego	$\text{kg}/(\text{mol m}^3)$
H_B	— entalpia własna reakcji zarodkowania pierwotnego	J/kg
T	— temperatura krystalizującego roztworu	$^\circ\text{K}$
T_0	— temperatura początkowa krystalizującego roztworu	$^\circ\text{K}$
D_c	— współczynnik dyfuzji roztwór – roztwór	m^2/s
λ_c	— współczynnik przewodnictwa cieplnego roztwór – roztwór	$\text{W}/(\text{m}^\circ\text{K})$
H_c	— entalpia własna dla stężenia krystalizującego roztworu	J/kg
c_p	— ciepło właściwe krystalizującego roztworu	$\text{J}/(\text{kg}^\circ\text{K})$
v	— pole wektora prędkości krystalizującego roztworu	m/s
v_0	— początkowe pole wektora prędkości krystalizującego roztworu (od ciągłego mieszania)	m/s
η_w	— współczynnik lepkości dynamicznej	Ns/m^2
g	— stała grawitacyjna	m/s^2
j	— rząd przesyconia dla reakcji wzrostu kryształów	
b	— współczynnik reakcji wzrostu kryształów	$(\text{mol}/\text{m}^2\text{s})/(\text{m}^3/\text{kg})^j$
$f(L)$	— funkcja liniowego rozmiaru kryształu w kinetyce wzrostu kryształów	m^2
E	— ułamek objętościowy fazy krystalicznej wewnątrz przesyconego roztworu [objętość kryształu/ $\Omega_{K(a,b,c)t}$]	
$(1-E)$	— ułamek objętościowy roztworu	
E_0	— początkowa wartość ułamka objętościowego fazy krystalicznej wewnątrz przesyconego roztworu	
ρ_K	— gęstość właściwa kryształu	kg/m^3
L	— wektor liniowy rozmiaru kryształu	m
L_0	— wartość początkowa wektora liniowego rozmiaru kryształu	m
G_K	— własna masa molowa dla reakcji wzrostu kryształów	$\text{kg}/(\text{mol m}^3)$
H_{GK}	— entalpia własna reakcji wzrostu kryształów	J/kg

T_K	— temperatura kryształów	$^{\circ}\text{K}$
T_{K0}	— temperatura początkowa kryształów	$^{\circ}\text{K}$
D_K	— współczynnik dyfuzji roztwór—kryształ	m^2/s
λ_K	— współczynnik przewodnictwa cieplnego roztwór—kryształ	$\text{W}/(\text{m } ^{\circ}\text{K})$
c_{pK}	— ciepło właściwe rosnących kryształów	$\text{J}/(\text{kg } ^{\circ}\text{K})$
v_K	— pole wektora prędkości rosnących kryształów	m/s
v_{K0}	— początkowe pole wektora prędkości rosnących kryształów	m/s
n	— statystyczna liczba zderzeń kryształów na powierzchni kryształu	liczba/m^2
K_V	— objętościowy współczynnik kształtu kryształu	
B	— funkcja tarcia kryształów	$[\text{kg}/\text{m}^3]^2$
$\bar{k}$	— stała Poissona dla kryształów	
h	— moduł ściskania kryształu	N/m^2
$\bar{\Omega}_{(a,b,c)t}$	— lokalnie wybrany, rozłożony parametrycznie element objętość—czas dla reakcji zarodkowania pierwotnego, $x=a$, $y=b$, $z=c$ dla czasu „ t ” [m, m, m, s], zjawiskowo domknięty	$\text{m}^3 - \text{s}$,
$F_{(a,b,c)t}$	— zewnętrznie zorientowana powierzchnia dla elementu $\bar{\Omega}_{(a,b,c)t}$	$\text{m}^2 - \text{s}$
$\mathbf{n}^{(z)}$	— wektor orientacyjny, zewnętrzny, normalny do powierzchni $F_{(a,b,c)t}$ mający cyrkulację decydującą o znakach zjawisk fizycznych dla pól potencjalnych i wirowych przy sumowaniu ich wpływów na określone współrzędne wektorów stanu w bilansach masy, energii i pędu	
$\bar{\Omega}_{K(a,b,c)t}$	— lokalnie wybrany, rozłożony parametrycznie element objętość—czas dla reakcji wzrostu kryształów zakłóconej w kinetyce procesem zarodkowania wtórnego, mający wymiary [m, m, m, s], zjawiskowo domknięty	$\text{m}^3 - \text{s}$
$F_{K(a,b,c)t}$	— zewnętrznie zorientowana powierzchnia dla elementu $\bar{\Omega}_{K(a,b,c)t}$	$\text{m}^2 - \text{s}$
$\mathbf{n}_K^{(z)}$	— wektor orientacyjny, zewnętrzny, normalny do powierzchni $F_{K(a,b,c)t}$ mający cyrkulację decydującą o znakach zjawisk fizycznych dla pól potencjalnych i wirowych przy sumowaniu ich wpływów na określone współrzędne wektorów stanu w bilansach masy, energii i pędu. Cyrkulacja wektora $\mathbf{n}^{(z)}$ jest przeciwna do cyrkulacji wektora $\mathbf{n}_K^{(z)}$.	
Ω_{Rt}	— przestrzeń robocza objętość—czas dla krystalizatorów	$\text{m}^3 - \text{s}$
F_{Rt}	— zewnętrznie zorientowana powierzchnia dla Ω_{Rt}	$\text{m}^2 - \text{s}$

$\mathbf{n}_R^{(z)}$ — wektor orientacyjny, zewnętrzny, normalny do powierzchni F_{Rt} , którego cyrkulacja zależy od zagadnień sterowania na brzegu danym powierzchnią F_{Rt} , będący podstawą do określenia relacji między cyrkulacjami:

$$\mathbf{n}_R^{(z)} \xleftarrow{(\tilde{A})} \mathbf{n}_R^{(z)} \xrightarrow{(\tilde{B})} \mathbf{n}_K^{(z)}$$

($\tilde{A}$) — ważną rolę gra proces zarodkowania pierwotnego,

($\tilde{B}$) — ważny jest proces wzrostu kryształów zakłócony w kinetyce procesem zarodkowania wtórnego.

Wybór wariantu ($\tilde{A}$) lub ($\tilde{B}$) zależy od wymaganych własności produktu procesu ciągłej krystalizacji masowej.

Uwaga U.4: Elementy $\bar{\Omega}_{(a,b,c)t}$ oraz $\bar{\Omega}_{K(a,b,c)t}$ są zjawiskowo domknięte, tzn. wpływy poszczególnych zjawisk fizycznych na wektory stanu odpowiadające tym elementom mogą być rozpatrywane w sposób oddzielny.

Układ odniesienia.

C_r	— stężenie masy rozpuszczalnika	kg/m ³
C_{r0}	— stężenie początkowe rozpuszczalnika	kg/m ³
T_r	— temperatura rozpuszczalnika	°K
T_{r0}	— temperatura początkowa rozpuszczalnika	°K
$\mathbf{v}_r$	— pole wektora prędkości rozpuszczalnika	m/s
$\mathbf{v}_{r0}$	— początkowe pole wektora prędkości rozpuszczalnika	m/s
D_r	— współczynnik dyfuzji rozpuszczalnika	m ² /s
λ_r	— współczynnik przewodnictwa cieplnego rozpuszczalnika	W/(m°K)
c_{pr}	— ciepło właściwe rozpuszczalnika	J/(kg°K)
η_{wr}	— współczynnik lepkości dynamicznej rozpuszczalnika	Ns/m ²

BIBLIOGRAFIA

1. H.E.C. Powers: *Nucleation and early crystal growth*. Journal Ind. Chemists, 1963, pp. 351
2. J. Frawley, W. Childs: *Dynamic nucleation of supercooling metals*. Transactions of the Metallurgical Society of AIME, 1968, v. 242, s. 256
3. J.W. Mullin: *Crystallization*. Butterworths, London 1972
4. A. Van Hook: *Crystallization*. Reinhold Publishing Corporation, New York 1961
5. J. Nyvlt: *Industrial crystallization from solutions*. Butterworths, London 1971
6. J. Synowiec: *Wpływ niektórych czynników na średni wymiar kryształów soli nieorganicznych wytwarzanych w procesie krystalizacji masowej*. PWN, Wrocław 1971
7. J. Nyvlt: *Crystallization from solutions*. SVTL, Bratislava 1967
8. R.L. Parker: *Crystal growth mechanisms: energetics, kinetics and transport*. Academic Press, New York 1970
9. W. Ohara, R. Reid: *Modelling crystal growth rate from solution..* Prentice Hall Inc., London 1973
10. R. Triverdi: *Effects of interface kinetics on the growth rate of dendritic*. Journal of Crystal Growth, 1979, v. 47, s. 389

11. E. Fraś: *Teoretyczne podstawy krystalizacji. Część I.* AGH im. St. Staszica. Skrypty uczelniane nr 930, Kraków 1984
12. A. Yue, J. Youe: *The effect of concentration and temperature gradients on solute redistribution.* Journal of Crystal Growth, 1972, v. 13/14, s. 797
13. M. Flemings: *Solidification processing.* Mc Graw-Hill Book Company, New York 1974
14. W. Kurz, D. Fisher: *Dendrite growth at the limit of stability; Tip radius and spacing.* Acta Metallurgica, 1981, v. 29, s. 11
15. A. Ohno: *The solidification of metals.* Chijin Shoken, Tokyo 1980
16. G. Cole: *Transport processes and fluid flow in solidification.* Solidification ASM, Metals Park, Ohio 1971, s. 201
17. W. Wilcox, M. Fullmer: *Turbulent free convection in Czochralski crystal growth.* Journal Applied Physics 1965, v. 36, s. 2202
18. T. Bower, R. Brody, M. Flemings: *Measurements of solute redistribution in dendritic solidification.* Transactions of the Metallurgical Society of AIME 1966, v. 236, s. 624
19. G.R. Younquist, A.D. Randolph: *Secondary nucleation in a class II system: Ammonium Sulphate Water.* AIChE Journal 1972, v. 18, s. 421
20. T.P. Melia, W.P. Moffit: *Secondary nucleation from aqueous solutions.* Ind. and Eng. Chem. Fund. 1964, v. 3/4, s. 313
21. W. Niemiec: *Model matematyczny o parametrach rozłożonych procesu ciągłej krystalizacji masowej do celów sterowania adaptacyjnego.* Praca doktorska. Politechnika Śląska w Gliwicach 1979
22. W. Niemiec: *A mathematical model of the distributed parameters of the continuous mass crystallization process for the adaptive control. Part I. The construction of partial differential constitutive state equations for the continuous mass crystallization process.* Poznańskie Towarzystwo Przyjaciół Nauk, Wydział Nauk Technicznych, Prace Komisji Automatyki i Informatyki, Tom XV—1989, s. 81—118
23. W. Niemiec: *On the constitutive theory of modelling and information for distributed parameter control of the continuous mass crystallization process: Part I: Partial differential constitutive state equations describing phenomena of the continuous mass crystallization process.* Third Int'l Conference on Liquid Metal Engineering and Technology in Energy Production, Oxford, England, 9—13 April 1984, Paper No. 191
24. W. Niemiec: *On modelling and information construction for control of distributed parameter chemical processes in fluid phase.* 3rd IFAC Symposium "Control of Distributed Parameter Systems", Toulouse, France, 29 VI—2 VII, 1992, Session 23
25. W. Niemiec: *The constitutive theory of modelling and information for phenomenal distributed parameter control of multicomponent chemical processes in gas, fluid and solid phase. Part I. The constitutive distributed parameter model of multicomponent chemical processes in gas, fluid and solid phase.* 7th Miami Int'l Conference on Alternative Energy Sources, Miami Beach, Florida, USA, 9—11 December, 1985, Session "Hydrocarbons/Energy Transfer" pp. 579—588, Hem Publ. Corp. Berlin, Heilderberg, New York, London Paris, Tokyo
26. W. Niemiec: *On modelling, information construction and optimization for distributed parameter control of multicomponent chemical processes in fluid— and gas phase. Part I. Partial differential constitutive state equations of multicomponent chemical processes in fluid— and gas phase.* 21st Annual Meeting of the Society of Engineering Science, Virginia Polytechnic Institute and State University, Blacksburg, USA, 15—17 October, 1984, 16AM6—9
27. W. Niemiec: *The mathematical analysis of the kinetics of the crystal growth reaction during the continuous mass crystallization process.* 171st Meeting, The Electrochemical Society, Philadelphia, Pennsylvania, USA, 10—15 May, 1987, paper No. 68
28. W. Niemiec: *The application of methods of structural and solid mechanics to the considerations of kinetics problems of crystal growth and secondary nucleation in the suspension during the continuous mass crystallization process.* 8th European Crystallographic Meeting, Liege, Belgium, 4—6 May, 1986
29. W. Niemiec: *A mathematical dynamical model of distributed parameters of the continuous mass crystallization process. Part I. The construction of partial differential constitutive state equations for the*

- continuous mass crystallization process. 4th SMIRT Conference, Paris, France, 7–11 August, 1984, Session/paper L 1/7*
30. W. Niemiec: *Zagadnienia sterowania adaptacyjnego procesów przemysłowych. Zeszyty Naukowe Politechniki Śląskiej Automatyka, Zeszyt 50, 1981*
 31. W. Nowacki: *Dynamiczne zagadnienia termosprężystości. PWN, Warszawa 1966*
 32. W. Nowacki, Z. Olesiak: *Termodyfuzja w ciałach stałych. PWN, Warszawa 1992*
 33. B. Średniawa: *Hydrodynamika i teoria sprężystości. PWN, Warszawa 1977*
 34. A.J.A. Morgan: *On the construction of constitutive equations for continuous media. Archiwum Mechaniki Stosowanej, 1965, v. 17, no. 1, pp. 145–174*
 35. T. Trajdos: *Matematyka dla inżynierów. PWN, Warszawa 1974*
 36. C.A. Coulson, A. Jeffrey: *Fale—modele matematyczne. WNT, Warszawa 1982*
 37. A.N. Tichonov, A.A. Samarski: *Równania fizyki matematycznej. PWN, Warszawa 1963*
 38. W. Smirnow: *Matematyka wyższa. PWN, Warszawa 1960*
 39. L. Schwartz: *Metody matematyczne w fizyce. PWN, Warszawa 1984*
 40. Wł. Findeisen: *Wielopoziomowe układy sterowania. WNT, Warszawa 1974*
 41. S. Węgrzyn: *Podstawy automatyki. PWN, Warszawa 1974*

W. NIEMIEC

THE CONSTRUCTION OF CONSTITUTIVE EVALUATIONS OF THE CONTINUOUS MEDIUM FOR THE CONTINUOUS MASS CRYSTALLIZATION PROCESS

PART I. THE CONSTRUCTION OF CONSTITUTIVE EQUATIONS OF THE CONTINUOUS MEDIUM FOR THE CONTINUOUS MASS CRYSTALLIZATION PROCESS DOCTORAL DISSERTATION — PART I

Summary

The subject of this part of the article is preparation of the constitutive distributed parameter model for the complete continuous mass crystallization process conceiving all its composite processes:

- the primary nucleation process,
- the crystal growth process,
- the secondary nucleation process,

being distributed in the working space of the crystallizer and time in the mutual physicochemical relations.

On the basis of the analysis of the literature experiments in the area on the kinetics of the reactions:

- the primary nucleation reaction,
- the crystal growth reaction disturbed in the kinetics by the secondary nucleation process,

and accompanying them physical phenomena an original concept of modelling of complete continuous mass crystallization process called in farther considerations "double locally distributed parameter concept" has been given. As a consequence of the complete approach were deduced in the work the systems of partial differential constitutive state equations of mass, energy and momentum balances of the supersaturated solution — the primary nucleation process and for the crystal inside the supersaturated solution — the crystal growth process but the disturbance of the kinetics of the crystal growth reaction by the secondary nucleation process by the mechanic energy balance equation on the surface of the crystal is given. For this original constitutive distributed parameter mathematical model of the complete continuous mass crystallization process in the work:

- analysis of the constitutive invariance problems for the continuous media with the space and time memories,
- mathematical classification of the systems of partial differential constitutive state equations of mass, energy and momentum balances,
- discussion of the control problems given by the elaborated model of the crystallization have been considered.

Równania konstytutywne ośrodka ciągłego dla procesu ciągłej krystalizacji masowej

Część II. Analityczne rozwiązanie równań konstytutywnych ośrodka ciągłego dla procesu ciągłej krystalizacji masowej

Praca doktorska – Część II

WACŁAW NIEMIEC

*Niezależny Instytut – Centrum Edukacyjne „Wiedza Nauka”
dawniej Politechnika Śląska w Gliwicach*

*Otrzymano 1993.01.10
Autoryzowano do druku 1993.11.10*

W części II artykułu rozwiązano analitycznie oryginalny konstytutywny rozłożony parametrycznie model matematyczny dla całego procesu ciągłej krystalizacji masowej, obejmujący wszystkie procesy składowe:

- proces zarodkowania pierwotnego,
- proces wzrostu kryształów,
- proces zarodkowania wtórnego został rozwiązany w oryginalny i analityczny sposób.

Rozważono dwa przypadki tego analitycznego rozwiązania przy:

A^o — istnieniu warunków początkowych,

B^o — istnieniu warunków początkowych i brzegowych,
dla rozwiązań zjawisk fizycznych pól potencjalnych i wirowych w układzie roz-
twór–kryształ oraz energii mechanicznych kryształu. Przedyskutowano szeroki zakres
zastosowań otrzymanych analitycznych rozwiązań dla poszczególnych procesów składo-
wych oraz dla całego procesu ciągłej krystalizacji masowej.

1. WSTĘP

W tej części pracy skoncentrowano się nad rozwiązaniem analitycznym równań różniczkowych cząstkowych konstytutywnych stanu dla procesów składowych procesu ciągłej krystalizacji masowej jak [1–8]:

— zarodkowanie pierwotne

$$\frac{DC}{Dt} = D_c \nabla^2 C - C \operatorname{div} \mathbf{v} - V_B G \quad (1)$$

$$(I) \quad C \frac{DT}{Dt} = \frac{\lambda_c}{c_p} \nabla^2 T - CT \operatorname{div} \mathbf{v} - \frac{H_B}{c_p} V_B G + \frac{H_c}{c_p} D_c \nabla^2 C \quad (2)$$

$$C \frac{D\mathbf{v}}{Dt} = \eta_w \nabla^2 \mathbf{v} - C \mathbf{v} \operatorname{div} \mathbf{v} + C \mathbf{g} + \frac{\eta_w}{3} \operatorname{grad} \operatorname{div} \mathbf{v} \quad (3)$$

gdzie

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \text{grad} - (\text{suma odpowiednich operacji gradientowych}) [13], [1-8],$$

— wzrost kryształów

$$\frac{DE}{Dt} = \frac{D_K}{\rho_K} C \nabla^2 E + E \text{div} \mathbf{v}_K + \frac{V_{GK}}{\rho_K} G_K E \quad (1)$$

$$\begin{aligned} E \frac{DT_K}{Dt} = & - \frac{\lambda_K}{c_{pK} \rho_K} (1-E) \nabla^2 T_K + E T_K \text{div} \mathbf{v}_K + \frac{H_{GK} V_{GK}}{c_{pK} \rho_K} G_K E + \\ & + \frac{H_c D_K}{c_{pK} \rho_K} C \nabla^2 E \end{aligned} \quad (2)$$

$$\begin{aligned} \rho_K E \frac{D\mathbf{v}_K}{Dt} = & \eta_w E \nabla^2 \mathbf{v}_K + \rho_K E \mathbf{v}_K \text{div} \mathbf{v}_K + [\rho_K E + C(1-E)] \mathbf{g} + \\ & + \frac{\eta_w}{3} E \text{grad} \text{div} \mathbf{v}_K. \end{aligned} \quad (3)$$

Warunek pływania kryształów wewnątrz przesyconego roztworu

$$\rho_K E \frac{D\mathbf{v}_K}{Dt} = C(1-E) \frac{D\mathbf{v}}{Dt} \quad (4)$$

gdzie

$$\frac{D}{Dt} = \frac{\partial}{\partial t} - \mathbf{v}_K \text{grad} + (\text{suma odpowiednich operacji gradientowych}) [13], [1-8]$$

— zarodkowanie wtórne, jako zakłócenie kinetyki reakcji wzrostu kryształów

$$(III) \quad n \rho_K K_V L \mathbf{v}_K^2 + \frac{B}{2 \rho_K} \frac{\partial^2 L}{\partial t^2} = h \left[\nabla^2 L + \frac{\bar{k}}{k-2} \text{grad} \text{div} L \right]. \quad (1)$$

Przy czym rozwiązanie równania różniczkowego cząstkowego (III.1) dla zarodkowania wtórnego jest poprzez funkcję $f(L)$ zakłóceniem kinetyki V_{GK} reakcji wzrostu kryształów.

Dla układu równań różniczkowych cząstkowych konstytutywnych stanu opisujących proces zarodkowania pierwotnego (I), w Części I [1] tej pracy wyprowadzono zasadę inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami przestrzenną i czasową:

— dla stałych współczynników fizycznych, układ równań różniczkowych cząstkowych (I'), i odpowiednio

— dla zmiennych współczynników fizycznych, układ równań różniczkowych cząstkowych (II").

Tak postawiona zasada inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami przestrzenną i czasową dla układu równań różniczkowych cząstkowych konstytutywnych stanu (I) pozwala stwierdzić, że układ ten jest niezmienniczy w stosunku do jego fizycznych współczynników. Z tego względu współczynniki fizyczne układu (I) są przyjmowane jako STAŁE przy jego analitycznym rozwiązaniu: $D_c \neq D_c(x, y, z, t)$, $\lambda_c \neq \lambda_c(x, y, z, t)$ oraz $\eta_w \neq \eta_w(x, y, z, t)$, w dalszej części artykułu. Rozważania te obowiązują także dla procesu wzrostu kryształów.

Układowi równań różniczkowych cząstkowych konstytutywnych stanu (II) dla procesu wzrostu kryształów odpowiadają wprowadzone w Części I [1] niniejszej pracy układy równań różniczkowych cząstkowych ciągłości jako zasada inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami przestrzenną i czasową:

— dla stałych współczynników fizycznych, układ równań różniczkowych cząstkowych (II'), a także

— dla zmiennych współczynników fizycznych, układ równań różniczkowych cząstkowych (II'').

Układy równań różniczkowych cząstkowych ciągłości (II') oraz (II'') stanowią formalizację matematyczną zasady inwariancji konstytutywnej dla ośrodków ciągłych z pamięciami przestrzenną i czasową i świadczą o tym, że układ równań różniczkowych cząstkowych konstytutywnych stanu (II) jest niezmienniczy w stosunku do swoich fizycznych współczynników. Dlatego współczynniki fizyczne układu (II) zostały przyjęte jako stałe przy jego analitycznym rozwiązaniu: $[D_K C] \neq [D_K C(x, y, z, t)]$, $[H_c D_K C(x, y, z, t), [\lambda_K(1-E)]] \neq [\lambda_K(1-E)(x, y, z, t)]$ oraz $[\eta_w E] = [\eta_w E(x, y, z, t)]$.

Ponadto układem odniesienia dla procesu zarodkowania pierwotnego bazującego na przesyconym roztworze jest ROZPUSZCZALNIK, zaś układem odniesienia dla procesu wzrostu kryształów jest proces zarodkowania pierwotnego. Dlatego rozpuszczalnik jest układem odniesienia dla całego procesu ciągłej krystalizacji masowej ponieważ proces zarodkowania wtórnego jest jedynie zakłóceniem kinetyki reakcji wzrostu kryształów. Dlatego też układ równań różniczkowych cząstkowych konstytutywnych stanu dla rozpuszczalnika (I_R) ma wszystkie cechy równań różniczkowych cząstkowych konstytutywnych stanu i może być rozwiązany analitycznie. Pomimo to, że reakcje zarodkowania pierwotnego i wzrostu kryształów bazują na przesyconym roztworze NIE WYSTĘPUJE POTRZEBA ANALITYCZNEGO ROZWIĄZANIA UKŁADU (I_R), ponieważ przyjmujemy iż własności rozpuszczalnika są stałe w stosunku do substancji rozpuszczonej [13], [11–12], [1–8].

2. ANALIZA JAKOŚCIOWA RÓWNAŃ KONSTYTUTYWNYCH DLA ICH ANALITYCZNEGO ROZWIĄZANIA

2.1. PROCES ZARODKOWANIA PIERWOTNEGO

2.1.1. Zagadnienia źródłowe zjawisk fizycznych roztwór-roztwór

Do rozważań bierzemy układ równań różniczkowych cząstkowych konstytutywnych stanu (I) opisujących proces zarodkowania pierwotnego. Na bazie udowod-

nionego w Części I [1], tej pracy TWIERDZENIA TW2 dla reakcji zarodkowania pierwotnego otrzymujemy z układu (I) źródłowy układ równań różniczkowych zwyczajnych względem czasu odpowiadający punktowi równoważonej reakcji zarodkowania pierwotnego $Z(x, y, z, t)$ przy $x=a, y=b, z=c$ dla czasu „ t ” [2–8]:

$$(I^S) \quad \begin{aligned} \frac{dC}{dt} &= -V_B G & (1) \\ C \frac{dT}{dt} &= -\frac{H_B}{c_p} V_B G & (2) \\ \frac{dv}{dt} &= g & (3) \end{aligned} \quad \int_0^t$$

Funkcje źródłowe: $C^Z(t), T^Z(t), v^Z(t)$.
 Uwaga U1. Przy całkowaniu równania (I^S 2) stosujemy moduł równania (I^S 1), żeby uwzględnić możliwość odwrotnej reakcji do reakcji zarodkowania pierwotnego – rozpuszczanie zarodków.

Warunki początkowe: C_0, T_0, v_0 (v_0 – od ciągłego mieszania).

2.2. PROCES WZROSTU KRYSTAŁÓW

2.2.1. Zagadnienia źródłowe zjawisk fizykalnych roztwór-kryształ

Z układu równań różniczkowych cząstkowych konstytutywnych stanu dla procesu wzrostu kryształów (II) na bazie udowodnionego w Części I [1] TWIERDZENIA TW2 dla reakcji wzrostu kryształów otrzymuje się z układu (II) źródłowy układ równań różniczkowych zwyczajnych względem czasu dla równoważnej reakcji wzrostu kryształów w punkcie $K(x, y, z, t)$. Zagadnienie to odpowiada ustaleniu zmiennych $x=a, y=b, z=c$ dla czasu „ t ” [2–8]:

$$(II^S) \quad \begin{aligned} \frac{dE}{dt} &= \frac{V_{GK}}{\rho_K} G_K E = SE & (1) \\ E \frac{dT_K}{dt} &= \frac{H_{GK}}{c_{pK}} SE & (2) \\ \rho_K E \frac{dv_K}{dt} &= [\rho_K E + C(1-E)]g & (3) \end{aligned} \quad \int_0^t$$

Funkcje źródłowe: $E^K(t), T_K^K(t), v_K^K(t)$.
 Uwaga U2. Warunki Uwagi U1 są spełnione dla układu (II^S).
 W układzie (II^S): $S = \frac{V_{GK}}{\rho_K} G_K$
 Warunki początkowe: E_0, T_{K0}, v_{K0} .
 (v_0 – od ciągłego mieszania).

Warunek pływania: $\rho_K E \frac{dv_K}{dt} = C(1-E) \frac{dv}{dt}$ (4)

2.3. ANALIZA JEDNORODNEJ POSTACI UKŁADU RÓWNAŃ (I) — PROCES ZARODKOWANIA PIERWOTNEGO

Jednorodna postać układu równań różniczkowych cząstkowych konstytutywnych stanu (I) może być zapisana jako [2–8]:

$$\frac{DC}{Dt} = D_c \nabla^2 C - C \operatorname{div} \mathbf{v} \quad (1)$$

$$(I^h)C \quad \frac{DT}{Dt} = \frac{\lambda_c}{c_p} \nabla^2 T - CT \operatorname{div} \mathbf{v} + \frac{H_c}{c_p} D_c \nabla^2 C \quad (2)$$

$$C \frac{D\mathbf{v}}{Dt} = \eta_w \nabla^2 \mathbf{v} - C \mathbf{v} \operatorname{div} \mathbf{v} + \frac{\eta_w}{3} \operatorname{grad} \operatorname{div} \mathbf{v} \quad (3)$$

gdzie

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \operatorname{grad} - (\text{suma odpowiednich operacji gradientowych}) [13], [2-8].$$

Do rozważań wprowadzamy teraz definicję pól wirowych [14], [16], [2-8].

DEFINICJA D1. Istnieje wektor „B”, wektor potencjałów pola przepływu roztworu taki, że $\mathbf{v} = \operatorname{rot} \mathbf{B}$ oraz $\operatorname{div} \operatorname{rot} \mathbf{B} \equiv 0$

i stosując konsekwentnie DEFINICJĘ D1 do układu (I^h) otrzymujemy:

— dla pól potencjalnych

$$\frac{\partial C}{\partial t} = D_c \nabla^2 C \quad (1)$$

(I_1^{hp})

$$C \frac{\partial T}{\partial t} = \frac{\lambda_c}{c_p} \nabla^2 T + \frac{H_c}{c_p} D_c \nabla^2 C, \quad (2)$$

przy czym do układu (I_1^{hp}) następująca uwaga jest obowiązująca

Uwaga U3. Operacje teorii pola $\operatorname{grad} D_c \operatorname{grad} C = 0$ oraz $\operatorname{grad} \lambda_c \operatorname{grad} T = 0$ co oznacza, że wektor $\operatorname{grad} D_c \perp \operatorname{grad} C$ a także $\operatorname{grad} \lambda_c \perp \operatorname{grad} T$, co odpowiada zagadnieniom pól potencjalnych [14-16].

Następnie do równania $(I_1^{hp}2)$ stosujemy reguły matematyczne [1]

$$C \frac{\partial H}{\partial t} = C c_p \frac{\partial T}{\partial t} + H_c \frac{\partial C}{\partial t}, \quad (3)$$

zaś przy nieobecności źródeł jest spełnione dodatkowo [1]

$$H = c_p T. \quad (4)$$

Odpowiednio do tego podejścia układ (I_1^{hp}) przechodzi do postaci

$$\frac{H_c}{c_p} \frac{\partial C}{\partial t} = \frac{H_c}{c_p} D_c \nabla^2 C \quad (1)$$

(I_2^{hp})

$$C \frac{\partial T}{\partial t} + \frac{H_c}{c_p} \frac{\partial C}{\partial t} = \frac{\lambda_c}{c_p} \nabla^2 T + \frac{H_c}{c_p} D_c \nabla^2 C. \quad (2)$$

Odjęcie równania (I_2^{hp}) od równania (I_2^{hp}) pozwala otrzymać

$$\frac{\partial C}{\partial t} = D_c \nabla^2 C \quad (1)$$

$$(I_3^{hp}) \quad C \frac{\partial T}{\partial t} = \frac{H_c}{c_p} D_c C \quad (2)$$

$$C \frac{\partial T}{\partial t} = \frac{\lambda_c}{c_p} \nabla^2 T \quad (3)$$

— dla pola wirowego

$$\frac{DC}{Dt} = -C \operatorname{div} v \quad (1)$$

$$(I_1^{hr}) \quad C \frac{DT}{Dt} = -CT \operatorname{div} v \quad (2)$$

$$C \frac{Dv}{Dt} = \eta_w \nabla^2 v - C \operatorname{div} v + \frac{\eta_w}{3} \operatorname{grad} \operatorname{div} v \quad (3)$$

gdzie

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + v \operatorname{grad} [13], [2-8].$$

2.4. ANALIZA JEDNORODNEJ POSTACI UKŁADU (II) — PROCES WZROSTU KRYSTAŁÓW

Punktem wyjścia do rozważań tego rozdziału jest postać jednorodna układu równań różniczkowych cząstkowych konstytutywnych stanu, która w oparciu o układ (II) może być zapisana jako [2-8]:

$$\frac{DE}{Dt} = \frac{D_K}{\rho_K} C V^2 E + E \operatorname{div} v_K \quad (1)$$

$$(II^h) \quad E \frac{DT_K}{Dt} = \frac{-\lambda_K}{c_{pK} \rho_K} (1-E) \nabla^2 T_K + E T_K \operatorname{div} v_K + \frac{H_c D_K}{c_{pK} \rho_K} C V^2 E \quad (2)$$

$$\rho_K E \frac{Dv_K}{Dt} = \eta_w E \nabla^2 v_K + \rho_K E v_K \operatorname{div} v_K + \frac{\eta_w}{3} E \operatorname{grad} \operatorname{div} v_K \quad (3)$$

przy warunku pływania kryształów wewnątrz przesyconego roztworu

$$\rho_K E \frac{Dv_K}{Dt} = C(1-E) \frac{Dv}{Dt} \quad (4)$$

gdzie

$$\frac{D}{Dt} = \frac{\partial}{\partial t} - v_K \operatorname{grad} + (\text{suma odpowiednich operacji gradientowych}) [13], [2-8].$$

Podobnie jak w rozdziale 2.3. istotne jest oparcie rozważań na definicji pól wirowych [14–16], [2–8]:

DEFINICJA D2. Istnieje taki wektor $\mathbf{B}_K$, wektor potencjał pola przepływu fazy krystalicznej, taki że $\mathbf{v}_K = \text{rot}\mathbf{B}_K$ oraz $\text{div}\text{rot}\mathbf{B}_K \equiv 0$

i jako konsekwencję użycia DEFINICJI D2 dla układu (II^h) mamy:

— dla pól potencjalnych

$$\frac{\partial E}{\partial t} = \frac{D_K}{\rho_K} \text{CV}^2 E \quad (1)$$

(II_1^{hp})

$$E \frac{\partial T_K}{\partial t} = - \frac{\lambda_K}{c_{pK} \rho_K} (1-E) \nabla^2 T_K + \frac{H_c D_K}{c_{pK} \rho_K} \text{CV}^2 E. \quad (2)$$

Układu (II_1^{hp}) dotyczy następująca uwaga:

Uwaga U4. Operacje teorii pola $\text{grad}\hat{\mathbf{M}}\text{grad}E = 0$ oraz $\text{grad}\hat{\mathbf{V}}\text{grad}T_K = 0$ co oznacza, że $\text{grad}\hat{\mathbf{M}} \perp \text{grad}E$ a także $\text{grad}\hat{\mathbf{V}} \perp \text{grad}T_K$ zgodnie z teorią pól potencjalnych, przy czym [14–16]:

$$\hat{\mathbf{M}} = \frac{D_K}{\rho_K} \quad \text{oraz} \quad \hat{\mathbf{V}} = - \frac{\lambda_K}{c_{pK} \rho_K} (1-E).$$

Do równania $(II_1^{hp}2)$ stosujemy teraz reguły matematyczne [1]

$$\rho_K E \frac{\partial H_K}{\partial t} = \rho_K E c_{pK} \frac{\partial T_K}{\partial t} + H_c \rho_K \frac{\partial E}{\partial t} \quad (3)$$

oraz dla warunku braku źródeł zewnętrznych [1]

$$H_K = c_{pK} T_K. \quad (4)$$

Na bazie tych reguł matematycznych układ (II_1^{hp}) przyjmuje nową postać:

$$\frac{H_c}{c_{pK}} \frac{\partial E}{\partial t} = \frac{H_c}{c_{pK}} \frac{D_K}{\rho_K} \text{CV}^2 E \quad (1)$$

(II_2^{hp})

$$E \frac{\partial T_K}{\partial t} + \frac{H_c}{c_{pK}} \frac{\partial E}{\partial t} = - \frac{\lambda_K}{c_{pK} \rho_K} (1-E) \nabla^2 T_K + \frac{H_c D_K}{c_{pK} \rho_K} \text{CV}^2 E. \quad (2)$$

Następnie odejmujemy równanie $(II_2^{hp}1)$ od równania $(II_2^{hp}2)$ daje wynik

$$\frac{\partial E}{\partial t} = \frac{D_K}{\rho_K} \text{CV}^2 E \quad (1)$$

(II_3^{hp})

$$E \frac{\partial T_K}{\partial t} = \frac{H_c D_K}{c_{pK} \rho_K} \text{CV}^2 E \quad (2)$$

$$E \frac{\partial T_K}{\partial t} = \frac{-\lambda_K}{c_{pK} \rho_K} (1-E) \nabla^2 T_K, \quad (3)$$

— dla pola wirowego

$$\frac{DE}{Dt} = + E \operatorname{div} \mathbf{v}_K \quad (1)$$

$$(II_1^{hr}) \quad E \frac{DT_K}{Dt} = + E T_K \operatorname{div} \mathbf{v}_K \quad (2)$$

$$\rho_K E \frac{D\mathbf{v}_K}{Dt} = \eta_w E \nabla^2 \mathbf{v}_K + \rho_K E \mathbf{v}_K \operatorname{div} \mathbf{v}_K + \frac{\eta_w}{3} E \operatorname{grad} \operatorname{div} \mathbf{v}_K \quad (3)$$

przy warunku pływania kryształów wewnątrz przesyconego roztworu

$$\rho_K E \frac{D\mathbf{v}_K}{Dt} = C(1-E) \frac{D\mathbf{v}}{Dt} \quad (4)$$

gdzie

$$\frac{D}{Dt} = \frac{\partial}{\partial t} - \mathbf{v}_K \operatorname{grad} \quad [13], [2-8].$$

3. ANALITYCZNE ROZWIĄZANIE KONSTITUTYWNEGO ROZŁOŻONEGO PARAMETRYCZNIE MATEMATYCZNEGO MODELU DLA PROCESU CIĄGŁEJ KRYSTALIZACJI MASOWEJ

A⁰. ISTNIENIE WARUNKÓW POCZĄTKOWYCH

1⁰. Proces zarodkowania pierwotnego.

1⁰.1. Funkcje źródłowe dla zjawiskowych rozwiązań.

Całkowanie układu równań różniczkowych zwyczajnych względem czasu (t^s) w przedziale czasowym $[0, t]$ po uwzględnieniu Uwagi U1 daje możliwość otrzymania funkcji źródłowych [2-8]:

— z równania (I^s1) dla stężenia

$$C^z(t) = C_0 - \int_0^t V_B G \, dt \quad (1)$$

— z równania (I^s2) w oparciu o Uwagę U1 dla temperatury

$$T^z(t) = T_0 - \frac{H_B}{c_p} \ln \left| \frac{C^z(t)}{C_0} \right| \text{ zaś } C^z(t) \text{ jest definiowane wzorem (1)} \quad (2)$$

— z równania (I^s3) dla pola wektora prędkości roztworu

$$\mathbf{v}^z(t) = \mathbf{v}_0 + \mathbf{g}t \quad (3)$$

Otrzymane funkcje źródłowe $[C^z(t), T^z(t), v^z(t)]$ są powiązane z odpowiednimi zjawiskami wewnątrz Ω_{Rt} , pól potencjalnych i pola wirowego i tworzą źródłowy wektor stanu w punkcie $Z(x, y, z, t)$.

1^o.2. Zjawiskowe rozwiązania dla pól potencjalnych

1^o.2.1. Dyfuzyjny transport stężenia roztworu (zmienny punkt $Q(\xi, \eta, \beta, \tau)$).

Rozważmy dyfuzyjny transport stężenia roztworu o postaci $(I_3^{hp} \cdot 1)$ przepisanej tutaj jako

$$\frac{\partial C}{\partial t} - D_c \nabla^2 C = 0 \quad (1)$$

Funkcja stężenia może być przedstawiona następująco

$$C(x, y, z, t) = U(x, y, z) \bar{C}(t) \quad (2)$$

i postać równania (2) przekształca postać równania (1) do układu równań

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} + \lambda U = 0 \quad (3)$$

oraz

$$\frac{d\bar{C}}{dt} = -\lambda D_c \bar{C} \quad (4)$$

W oparciu o TWIERDZENIE TW3 w Części I [1] warunki brzegowe na brzegu $F_{(a,b,c)t}$ elementu objętość – czas $\bar{\Omega}_{(a,b,c)t}$ są równe zero

$$\begin{bmatrix} U(0, y, 0) \\ U(x, 0, 0) \\ U(0, 0, z) \end{bmatrix} = 0 \quad \text{oraz} \quad \begin{bmatrix} U(a, y, 0) \\ U(x, b, 0) \\ U(0, y, c) \end{bmatrix} = 0. \quad (5)$$

Wprowadzamy teraz formę postaci funkcji $U(x, y, z)$ napisaną poniżej

$$U(x, y, z) = X(x) \cdot Y(y) \cdot Z(z) \quad (6)$$

do równania (3) przekształcając to równanie do postaci

$$\begin{aligned} X'' + 9X &= 0, & \begin{bmatrix} X(0) \\ Y(0) \\ Z(0) \end{bmatrix} &= 0 & \text{and} & \begin{bmatrix} X(a) \\ Y(b) \\ Z(c) \end{bmatrix} &= 0 \\ Y'' + \mu Y &= 0, & & & & & \\ Z'' + \gamma Z &= 0, & & & & & \end{aligned} \quad (7)$$

Równania różniczkowe zwyczajne (7) względem współrzędnych przestrzennych mają rozwiązania

$$\begin{aligned}
 X(x) &= \sin \frac{n\pi x}{a}, & \vartheta &= \left(\frac{n\pi}{a}\right)^2 \\
 Y(y) &= \sin \frac{m\pi y}{b}, & \mu &= \left(\frac{m\pi}{b}\right)^2 \\
 Z(z) &= \sin \frac{k\pi z}{c}, & \gamma &= \left(\frac{k\pi}{c}\right)^2,
 \end{aligned} \tag{8}$$

przy czym osiągają wartości własne

$$\lambda_{n,m,k} = \left(\frac{n\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2 + \left(\frac{k\pi}{c}\right)^2 \tag{9}$$

oraz funkcje własne

$$U_{n,m,k} = A_{n,m,k} \sin \frac{n\pi x}{a} \sin \frac{m\pi y}{b} \sin \frac{k\pi z}{c}. \tag{10}$$

W rozważaniach naszych niezbędne jest uwzględnienie następującej uwagi [15], [2–8]:

Uwaga U5. Amplitudę $A_{n,m,k}$ można wybrać z punktu widzenia wymiarów miejscowych „ a , b , c ” dla czasu „ t ”, tak że norma funkcji własnych $U_{n,m,k}$ z wagą równą 1 (jeden) jest równa 1 (jeden).

Równanie normy ma postać [15, [2–8]:

$$N_e = \int_0^a \int_0^b \int_0^c U_{n,m,k}^2 dx dy dz = A_{n,m,k}^2 \int_0^a \sin^2 \frac{n\pi x}{a} dx \int_0^b \sin^2 \frac{m\pi y}{b} dy \int_0^c \sin^2 \frac{k\pi z}{c} dz = 1 \tag{11}$$

Ze scałkowania równania (11) można otrzymać wartość amplitudy $A_{n,m,k}$

$$A_{n,m,k} = \sqrt{\frac{8}{abc}}. \tag{12}$$

Funkcje własne przynależne do stałego punktu $Z(x, y, z, t)$ są więc

$$U_{n,m,k}(x, y, z) = \sqrt{\frac{8}{abc}} W(x, y, z) \tag{13}$$

gdzie

$$W(x, y, z) = \sin \frac{n\pi x}{a} \sin \frac{m\pi y}{b} \sin \frac{k\pi z}{c}$$

zaś dla punktu zmiennego $Q(\xi, \eta, \beta, \tau)$ funkcje własne mają postać

$$U_{n,m,k}(\xi, \eta, \beta) = \sqrt{\frac{8}{abc}} W(\xi, \eta, \beta) \quad (14)$$

przy

$$W(\xi, \eta, \beta) = \sin \frac{n\pi\xi}{a} \sin \frac{m\pi\eta}{b} \sin \frac{k\pi\beta}{c}.$$

Po scałkowaniu konsekwentnie do procedury rozwiązywania równania

$$C(t) = D_{n,m,k}^c e^{-\lambda_{n,m,k} D_c t} \quad (15)$$

gdzie

$$D_{n,m,k}^c = \iiint_{\Omega_x} \bar{C}_0 \sqrt{\frac{8}{abc}} W(\xi, \eta, \beta) d\xi d\eta d\beta \quad (16)$$

oraz

$$W(\xi, \eta, \beta) = \sin \frac{n\pi\xi}{a} \sin \frac{m\pi\eta}{b} \sin \frac{k\pi\beta}{c}.$$

Dla pełności rozważań przepisujemy ponownie funkcję źródłową stężenia, wzór (1^o.1.1):

$$C^z(t) = C_0 - \int_0^t V_B G dt. \quad (17)$$

Całość powyższych rozważań pozwala sformułować pełne rozwiązanie zagadnienia transportu stężenia krystalizującego roztworu w strumieniu dyfuzji roztwór—roztwór:

$$C(Z, t) = \int_0^t \iiint_{\Omega_x} \left\{ \sum_{\substack{n=1 \\ m=1 \\ k=1}}^{\infty} D_{n,m,k}^c e^{-\lambda_{n,m,k} D_c(t-\tau)} \frac{8}{abc} W(x, \xi; y, \eta; z, \beta) \right\} \left((C_0 - \int_0^{\tau} V_B G dt) d\xi d\eta d\beta d\tau, \right. \quad (18)$$

gdzie $W(x, \xi; y, \eta; z, \beta) = W(x, y, z) W(\xi, \eta, \beta)$.

Część rozwiązania (18) między nawiasami $\{ \}$ jest zjawiskową funkcją Greena dla zjawiska dyfuzyjnego transportu stężenia roztwór—roztwór:

$$G_{Fh}^c(Z, t; Q, \tau) = \left\{ \sum_{\substack{n=1 \\ m=1 \\ k=1}}^{\infty} D_{n,m,k}^c e^{-\lambda_{n,m,k} D_c(t-\tau)} \frac{8}{abc} W(x, \xi; y, \eta; z, \beta) \right\}. \quad (19)$$

Własności funkcji Greena (19) odpowiadają stałemu obrazowi wzajemnościowemu dla punktów [11–12], [15]:

$$\begin{array}{ccc} Z(x, y, z, t) & \longleftrightarrow & Q(\xi, \eta, \beta, \tau) \\ \in \bar{\Omega}_{(a,b,c)t} & \text{Stały wzajemnościowy obraz.} & \in \Omega_{Rt} \end{array} \quad (20)$$

1^o.2.2. Transport entalpii w strumieniu dyfuzji roztwór – roztwór
(zmienny punkt $Q(\xi, \eta, \beta, \tau)$)

Postać wyjściowa zagadnienia tematowego z przepisywania równania ($I_3^{p,2}$) jest

$$C \frac{\partial T}{\partial t} = \frac{H_c}{c_p} D_c \nabla^2 C. \quad (1)$$

Równanie źródłowe dla równania tematowego (1) ma postać

$$C \frac{dT}{dt} = \frac{H_c}{c_p} \frac{dC}{dt} \quad (2)$$

i to równanie źródłowe pochodzi z kombinacji równania (1) oraz następującej postaci równania dyfuzji stężenia roztworu

$$\frac{H_c}{c_p} \frac{\partial C}{\partial t} = \frac{H_c}{c_p} D_c \nabla^2 C. \quad (3)$$

Na bazie równania (1) oraz równania (3) można napisać

$$\frac{H_c}{c_p} \frac{\partial C}{\partial t} = C \frac{\partial T}{\partial t} = \frac{H_c}{c_p} D_c \nabla^2 C, \quad (4)$$

które może być transformowane do postaci

$$C \frac{\partial T}{\partial t} = \frac{H_c}{c_p} \left(\frac{\partial C}{\partial t} - D_c \nabla^2 C \right) \quad (5)$$

oraz dodatkowo

$$\frac{H_c}{c_p} \frac{\partial C}{\partial t} = \partial T. \quad (6)$$

Odpowiednio do rozważań rozdziału 1^o.2.1. rozwiązanie problemu dyfuzyjnego transportu entalpii roztwór – roztwór jest dane wzorem:

$$\begin{aligned} T(Z, t) = T_0(Z, t) + \frac{H_c}{c_p} \left[\ln \left| \int_0^t \iiint_{\Omega_x} \left\{ \sum_{\substack{n=1 \\ m=1 \\ k=1}}^{\infty} D_{n,m,k}^c e^{-\lambda_{n,m,k} D_c(t-\tau)} \frac{8}{abc} W(x, \xi; y, \eta; z, \beta) \right\} \right. \right. \\ \left. \left. \left(C_0 - \int_0^{t \rightarrow \tau} V_B G dt \right) d\xi d\eta d\beta d\tau \right| - \ln |C_0| \right]. \end{aligned} \quad (7)$$

W rozwiązaniu (7) obowiązuje taka sama funkcja Greena jak dla dyfuzyjnego transportu stężenia w roztworze — wzór (1^o.2.1.19).

1^o.2.3. Zmiany temperatury od przewodnictwa cieplnego roztwór — roztwór (zmienny punkt $Q(\xi, \eta, \beta, \tau)$)

Rozważania zaczynamy od przywołania pamięcią równania przewodnictwa cieplnego w postaci jednorodnej (I₃^{hp}3)

$$C \frac{\partial T}{\partial t} - \frac{\lambda_c}{c_p} \nabla^2 T = 0. \quad (1)$$

Wprowadzając funkcję temperatury

$$T(x, y, z, t) = V(x, y, z, t) \bar{T}(t) \quad (2)$$

transformujemy równanie (1) do postaci układu równań

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} + \lambda V = 0 \quad (3)$$

oraz

$$\frac{d\bar{T}}{dt} = -\lambda \frac{\lambda_c}{c_p C(t)} \bar{T}. \quad (4)$$

W oparciu o TWIERDZENIE TW3 w Części I [1] warunki brzegowe na brzegu $F_{(a,b,c)t}$ elementu objętość — czas $\bar{\Omega}_{(a,b,c)t}$ są zerowe — porównaj rozdział 1^o.2.1. W konsekwencji powtórzenia rozważań jak dla rozdziału 1^o.2.1. wartości własne mają postać

$$\lambda_{n,m,k} = \left(\frac{n\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2 + \left(\frac{k\pi}{c}\right)^2. \quad (5)$$

Odpowiednio też do rozważań rozdziału 1^o.2.1 otrzymujemy funkcje własne:

— dla stałego punktu $Z(x, y, z, t)$

$$V \lambda_{n,m,k}(x, y, z) = \sqrt{\frac{8}{abc}} W(x, y, z) \quad (6)$$

przy czym $W(x, y, z)$ jest określone w rozdziale 1^o.2.1.

— dla zmiennego punktu $Q(\xi, \eta, \beta, \tau)$

$$V_{n,m,k}(\xi, \eta, \beta) = \sqrt{\frac{8}{abc}} W(\xi, \eta, \beta) \quad (7)$$

przy $W(\xi, \eta, \beta)$ podanym w rozdziale 1^o.2.1.

Po scałkowaniu równania (4) otrzymujemy

$$\bar{T}(t) = D_{n,m,k}^T e^{-\lambda_{n,m,k} \frac{\lambda_c}{c_p} [t(t)]} \quad (8)$$

gdzie funkcja

$$[J(t)] = \int \frac{dt}{C_0 - \int_0^t V_B G dt},$$

zaś stała całkowania w równaniu (8) ma postać

$$D_{n,m,k}^T = \iiint_{\Omega_r} T_0 \sqrt{\frac{8}{abc}} W(\xi, \eta, \beta) d\xi d\eta d\beta. \quad (9)$$

Z rozdziału 1^o.1. przepisujemy wzór na funkcję źródłową temperatury reakcji zarodkowania pierwotnego jako — wzór (1^o.1.2)

$$T^z(t) = T_0 - \frac{H_B}{c_p} \ln \left| \frac{C^z(t)}{C_0} \right|. \quad (10)$$

Rozważania powyższe pozwalają sformułować pełne rozwiązanie zagadnienia zmian temperatury od zjawiska przewodnictwa cieplnego roztwór—roztwór

$$T(Z, t) = \int_0^t \iiint_{\Omega_r} \left\{ \sum_{\substack{n=1 \\ m=1 \\ k=1}}^{\infty} D_{n,m,k}^T e^{-\lambda_{n,m,k} \frac{\lambda_c}{c_p} [t-(\tau)]} \frac{8}{abc} W(x, \xi; y, \eta; z, \beta) \right\} \left[T_0 - \frac{H_B}{c_p} \ln \left| \frac{C^z(\tau)}{C_0} \right| \right] d\xi d\eta d\beta d\tau \quad (11)$$

gdzie $W(x, \xi; y, \eta; z, \beta)$ — patrz rozdział 1^o.2.1.

Rozwiązanie dane równaniem (11) bazuje na funkcji Greena dla zjawiska przewodnictwa cieplnego roztwór—roztwór w postaci

$$G_{Ph}^T(Z, t; Q, \tau) = \left\{ \sum_{\substack{n=1 \\ m=1 \\ k=1}}^{\infty} D_{n,m,k}^T e^{-\lambda_{n,m,k} \frac{\lambda_c}{c_p} [t-(\tau)]} \frac{8}{abc} W(x, \xi; y, \eta; z, \beta) \right\} \quad (12)$$

odpowiadającej stałemu zjawiskowemu obrazowi dla punktów [11–12], [15]

$$\begin{array}{ccc} Z(x, y, z, t) & \longleftrightarrow & Q(\xi, \eta, \beta, \tau) \\ \in \bar{\Omega}_{(a,b,c)t} & \text{Stały wzajemnościowy obraz.} & \in \Omega_{Rt} \end{array} \quad (13)$$

1^o.3. Zjawiskowe rozwiązanie dla pola wirowego

1^o.3.1. Pole wektora prędkości roztwór—roztwór (zmienny punkt $Q'(\xi', \eta', \beta', \tau)$).

Rozważania nasze zapoczątkujemy od przywołania do pamięci układu równań różniczkowych cząstkowych w formach pochodnych śledczych dla pola wirowego (I_1^{hr}) w postaci

$$\frac{DC}{Dt} = -C \operatorname{div} \mathbf{v} \quad (1)$$

$$C \frac{DT}{Dt} = -CT \operatorname{div} \mathbf{v} \quad (2)$$

$$C \frac{D\mathbf{v}}{Dt} = \eta_w \nabla^2 \mathbf{v} - C \operatorname{div} \mathbf{v} + \frac{\eta_w}{3} \operatorname{grad} \operatorname{div} \mathbf{v} \quad (3)$$

gdzie

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \operatorname{grad} \quad [13, [2-8].$$

Równanie (1) przemnażamy obustronnie przez wektor „ $\mathbf{v}$ ” i stosujemy w równaniu (3), które następnie przekształcamy do postaci

$$C \frac{D\mathbf{v}}{Dt} - \mathbf{v} \frac{DC}{Dt} = \eta_w \nabla^2 \mathbf{v} + \frac{\eta_w}{3} \operatorname{grad} \operatorname{div} \mathbf{v}. \quad (4)$$

Na bazie stałości obrazu wzajemnościowego — stałe wymiary „ a, b, c ” wynikłego z zasady wzajemności dla izotropowych i anizotropowych niejednorodnych ośrodków [11–12], równanie (4) rozkładamy na część statyczną i dynamiczną. Należy przy tym zwrócić uwagę:

Uwaga U6. Części statycznej w zapisie odpowiada $\bar{\Omega}_{(a,b,c)t=0} \equiv \bar{\Omega}_{(a,b,c)}$, zaś części dynamicznej $\bar{\Omega}_{(a,b,c)t}$. Dla równania (4) oba zapisy są równoważne [11–12].

Kontynuując dalej rozważania:

$$\eta_w \nabla^2 \mathbf{v} + \frac{\eta_w}{3} \operatorname{grad} \operatorname{div} \mathbf{v} = 0 \quad (5)$$

oraz

$$C \frac{D\mathbf{v}}{Dt} - \mathbf{v} \frac{DC}{Dt} = 0 \quad (6)$$

Konsekwencja użycia matematycznej reguły [14], [16]

$$\operatorname{rot} \operatorname{rot} \mathbf{v} = \operatorname{grad} \operatorname{div} \mathbf{v} - \nabla^2 \mathbf{v} \quad (7)$$

do równania (5) jest modyfikacja tego równania do postaci

$$\left(\eta_w + \frac{\eta_w}{3} \right) \operatorname{grad} \operatorname{div} \mathbf{v} - \eta_w \operatorname{rot} \operatorname{rot} \mathbf{v} = 0 \quad (8)$$

Następnie transformujemy równanie (8) do postaci

$$\bar{\nabla}^2 \mathbf{v} = 0 = \alpha \operatorname{grad} \operatorname{div} \mathbf{v} - \beta \operatorname{rot} \operatorname{rot} \mathbf{v} \quad (9)$$

gdzie

$$\alpha = \eta_w + \frac{\eta_w}{3} \quad \text{oraz} \quad \beta = \eta_w$$

W stanie statycznym, dla $t = 0$ każdy punkt o stałych współrzędnych $Z(x, y, z, t) = Z(x, y, z, t=0) = Z(x, y, z, 0)$ posiada pole wektora prędkości $v_0(x, y, z, 0) = v_0$ zaś rozkład statyczny pola wektora prędkości jest dany wzorem

$$v_0(Z, 0) = v_0(x, y, z, 0) = \iiint_{\Omega_x} \Gamma_G^W(Z, Q') v_0(Q') d\xi' d\eta' d\beta' \quad (10)$$

Uwaga U7. Przy zjawiskowych funkcjach Greena zapis punktu dla stanu statycznego może być $Z(x, y, z, t; t=0) = Z(x, y, z, 0)$ a stąd można napisać $v_0(Z; t=0) = v_0(x, y, z, t; t=0) = v_0(x, y, z, 0) = v_0$, lub też konsekwentnie dla $t=0$ zachodzi $Z(x, y, z, 0)$ a stąd $v_0(x, y, z, 0) = v_0$.

Równanie (10) jest ważne dla następujących przypadków tensora Greena dla elementu statycznego $\bar{\Omega}_{(a,b,c)}$ roztworu [16–17], [2–8]:

— przypadek izotropowy, Dodatek 1

$$\Gamma_G^W(Z, Q') = \Gamma_G(Z, Q'). \quad (11)$$

— przypadek anizotropowy, Dodatek 2

$$\Gamma_G^W(Z, Q') = \Gamma_{GA}(Z, Q'). \quad (12)$$

Kontynuując nasze roważania bierzemy do dyskusji dynamiczną część pola wektora prędkości „ v ” związaną ze współrzędną czasową „ t ” dla $\bar{\Omega}_{(a,b,c)t}$, tak więc

$$C \frac{Dv}{Dt} - v \frac{DC}{Dt} = 0. \quad (13)$$

Równanie (13) jest następnie scałkowane do postaci

$$\ln|N| - |C| = \ln|v| \quad (14)$$

i ostatecznie

$$v(t) = \frac{C^z(t)}{N_1} \quad (15)$$

gdzie: $C^z(t)$ — funkcja źródłowa dla stężenia w punkcie $Z(x, y, z, t)$ dla $x=a, y=b, z=c$, porównawczo do wzoru (1⁰.1.1)

$$C^z(t) = C_0 - \int_0^t v_B G dt, \quad (16)$$

zaś N_1 — stała całkowania przyjęta tak, że $N_1 = v(Z, 0)$ dla postaci równania (10). Należy uwzględnić ponadto, że działa jeszcze funkcja źródłowa pola wektora prędkości roztworu dana równaniem (1⁰.1.3) posiadającym postać

$$v^z(t) = v_0 + gt. \quad (17)$$

Całość rozważań tego rozdziału pozwala sformułować pełną postać rozwiązania dla pola wektora prędkości roztworu w postaci

$$v(Z, t) = \int_0^t \iiint_{\Omega_x} \left\{ \frac{1}{N_1} \left[C_0 - \int_0^{t \rightarrow (t-\tau)} V_B G dt \right] \Gamma_G^W(Z, Q') \right\} (v_0 + g\tau) d\xi' d\eta' d\beta' d\tau. \quad (18)$$

Dla równania (18) funkcja Greena pola wektora prędkości ma postać

$$G_{Ph}^v(Z, t; Q', \tau) = \left\{ \frac{1}{N_1} \left[C_0 - \int_0^{t \rightarrow (t-\tau)} V_B G dt \right] \Gamma_G^W(Z, Q') \right\}. \quad (19)$$

Wzajemnościowa interpretacja punktów o stałych i zmiennych współrzędnych może być zapisana w następujący sposób [11 – 12], [15]

$$\begin{array}{ccc} Z(x, y, z, t) & \longleftrightarrow & Q'(\xi', \eta', \beta', \tau) \\ \in \bar{\Omega}_{(a,b,c)t} & \text{Stały wzajemnościowy obraz.} & \in \Omega_{Rt} \end{array} \quad (20)$$

1^o.4. Kompletne rozwiązanie dla procesu zarodkowania pierwotnego przy obecności warunków początkowych roztworu

Rozwiązanie kompletne dla procesu zarodkowania pierwotnego przy obecności warunków początkowych przedstawia się przy pomocy wektora stanu o współrzędnych masa, energia i pęd, w stosunku do których sumuje się wpływy poszczególnych zjawisk pól potencjalnych i wirowych. Konsekwencją użycia wyników rozdziałów A^o.1.^o.1. – A^o.1^o.3. jest następująca formuła kompletnego rozwiązania [2 – 8]:

Kompletna postać rozwiązania	Funkcje źródłowe w punkcie Z(x, y, z, t)
$\begin{bmatrix} \text{Stężenie} \\ \text{Temperatura} \\ \text{Pole wektora} \\ \text{prędkości} \end{bmatrix}$	$\begin{bmatrix} \text{Obniżenie stężenia} \\ \text{(formuła A}^o.1^o.1.1\text{)} \\ \text{Obniżenie temperatury} \\ \text{(formuła A}^o.1^o.1.2\text{)} \\ \text{Przyrost pola wektora prędkości} \\ \text{(formuła A}^o.1^o.1.3\text{)} \end{bmatrix} +$
Dyfuzja Q → Z	Przewodnictwo cieplne Q → Z
$+ \begin{bmatrix} \text{Zmiany stężenia} \\ \text{(formuła A}^o.1^o.2.1.18\text{)} \\ \text{Zmiany temperatury} \\ \text{(formuła A}^o.1^o.2.2.7\text{)} \\ * \end{bmatrix}$	$+ \begin{bmatrix} * \\ \text{Zmiany temperatury} \\ \text{(formuła A}^o.1^o.2.3.11\text{)} \\ * \end{bmatrix} +$
Pole wektora pedkości Q' → Z	

$$+ \begin{bmatrix} \text{Ciągłość stężenia} \\ \text{(formuła A}^0.1^0.3.1\text{)} \\ \text{Ciągłość temperatury} \\ \text{formuła A}^0.1^0.3.2\text{)} \\ \text{Zmiany pola wektora prędkości} \\ \text{(formuła A}^0.1^0.3.18\text{)} \end{bmatrix} \quad (1) \quad * - \text{współrzędne, w których nie ma związków między zjawiskami fizykalnymi.}$$

2^o. Proces wzrostu kryształów

2^o.1. Funkcje źródłowe dla zjawiskowych rozwiązań

Dla otrzymania funkcji źródłowych odpowiadających ekwiwalentnej reakcji wzrostu kryształów sprowadzonej do punktu $K(x, y, z, t)$ całkujemy układ równań różniczkowych zwyczajnych względem czasu (II^S) w przedziale czasu $[0, t)$ odpowiedni do Uwagi U2:

— z równania ($II^S.1$) dla ułamka objętościowego fazy krystalicznej

$$E^k(t) = |E_0| e^{St} \quad (1)$$

gdzie $S = \frac{V_{GK}}{\rho_K} G_K$ oraz S — wielkość stała lub parametrycznie zmienna odpowiednio do procesu zarodkowania pierwotnego, stężenie roztworu.

— z równania ($II^S.2$) dla temperatury kryształów

$$T_K^k(t) = T_{K0} + \frac{H_{GK}}{c_{pK}} St \quad (2)$$

— z równania ($II^S.3$) dla pola wektora prędkości kryształów

$$v_K^k(t) = v_{K0} + \left\{ 1 + \frac{C^z(t) [1 - E^k(t)]}{\rho_K E^k(t)} \right\} g t, \quad (3)$$

gdzie: $C^z(t)$ oraz $E^k(t)$ — działają niezależnie od źródła pola wektora prędkości kryształów wewnątrz przesyconego roztworu.

Uwaga U8. Obecność funkcji źródłowej $C^z(t)$ we wzorze (3) odpowiada teorii wzrostu kryształów, która twierdzi iż wzrost kryształów polega na powstawaniu zarodków pierwotnych na powierzchni kryształów. We wzorze (3) elementy $C^z(t)$ oraz $E^k(t)$ są narzucone jako niezależne od $v_K^k(t)$ w zagadnieniach lokalnych, natomiast podlegają całkowaniu względem czasu w rozwiązaniach globalnych zjawisk fizykalnych prędkości roztwór — kryształ [1].

Warunek pływania kryształów wewnątrz przesyconego roztworu uzupełnia równanie (3):

$$\rho_K E \frac{dv_K}{dt} = C(1 - E) \frac{dv}{dt} \quad (4)$$

Funkcje źródłowe $[E^k(t), T_K^k(t), v_K^k(t)]$ tworzą źródłowy wektor stanu w punkcie $K(x, y, z, t)$ oraz są związane ze zjawiskami fizycznymi dla pól potencjalnych i wirowego wewnątrz Ω_{Rt} .

2^o.2. Zjawiskowe rozwiązania dla pól potencjalnych

2^o.2.1. Dyfuzyjny transport stężenia roztwór – kryształ (zmienny punkt $Q(\xi, \eta, \beta, \tau)$)

Rozważanie rozpoczynamy przypomnieniem równania dyfuzyjnego transportu stężenia roztwór – kryształ (II₃^{hp}.1) jak następuje

$$\frac{\partial E}{\partial t} - \frac{D_K}{\rho_K} \nabla^2 E = 0. \quad (1)$$

Całość rozważań jest tutaj powtórzeniem procedury rozwiązywania z rozdziału 1^o.2.1. dla dyfuzji roztwór – roztwór w procesie zarodkowania pierwotnego. Wprowadzamy więc funkcję ułamka objętościowego fazy kryształicznej

$$E(x, y, z, t) = \bar{U}(x, y, z) \bar{E}(t) \quad (2)$$

który modyfikuje równanie (1) do układu równań

$$\frac{\partial^2 \bar{U}}{\partial x^2} + \frac{\partial^2 \bar{U}}{\partial y^2} + \frac{\partial^2 \bar{U}}{\partial z^2} + \lambda \bar{U} = 0 \quad (3)$$

oraz

$$\frac{d\bar{E}}{dt} = -\lambda \frac{D_K}{\rho_K} C \bar{E}. \quad (4)$$

W oparciu o TWIERDZENIE TW3 w Części I [1] warunki brzegowe na brzegu $F_{K(a,b,c)t}$ elementu objętość – czas $\bar{\Omega}_{K(a,b,c)t}$ są równe zero. Otrzymujemy więc odpowiednio do rozdziału 1^o.2.1.

— wartości własne

$$\lambda_{n,m,k} = \left(\frac{n\pi}{a} \right)^2 + \left(\frac{m\pi}{b} \right)^2 + \left(\frac{k\pi}{c} \right)^2. \quad (5)$$

— funkcje własne dla punktu o stałych współrzędnych $K(x, y, z, t)$

$$\bar{U}_{n,m,k}(x, y, z) = \sqrt{\frac{8}{abc}} W(x, y, z) \quad (6)$$

— funkcje własne dla punktu o zmiennych współrzędnych $Q(\xi, \eta, \beta, \tau)$

$$\bar{U}_{n,m,k}(\xi, \eta, \beta) = \sqrt{\frac{8}{abc}} W(\xi, \eta, \beta) \quad (7)$$

Całkowanie równania (4) pozwala otrzymać

$$\bar{E}(t) = D_{n,m,k}^E e^{-\lambda_{n,m,k} \frac{D_K}{\rho_K} C t}. \quad (8)$$

Uwaga U9. Przy całkowaniu równania (4) stężenie $C = C(x, y, z, t)$ traktuje się jako stałe, narzucone przez proces zarodkowania pierwotnego, $C = \text{const}$.

W równaniu (8) stała całkowania ma postać

$$D_{n,m,k}^E = \iiint_{\Omega_K} E_0 \sqrt{\frac{8}{abc}} W(\xi, \eta, \beta) d\xi d\eta d\beta. \quad (9)$$

W tym miejscu rozważań przywołujemy do pamięci funkcję źródłową ułamka fazy krystalicznej z równania (2^o.1.1)

$$E^k(t) = |E_0| e^{St}. \quad (10)$$

Na bazie powyższych informacji pełne rozwiązanie równania dyfuzji stężenia w układzie roztwór–kryształ ma postać

$$E(K, t) = \int_0^t \iiint_{\Omega_K} \left\{ \sum_{\substack{n=1 \\ m=1 \\ k=1}}^{\infty} D_{n,m,k}^E e^{-\lambda_{n,m,k} \frac{D_K}{\rho_K} C(t-\tau)} \frac{8}{abc} W(x, \xi; y, \eta; z, \beta) \right\} |E_0| e^{S\tau} d\xi d\eta d\beta d\tau \quad (11)$$

Część rozwiązania (11) między nawiasami { } jest zjawiskową funkcją Greena dla dyfuzji roztwór–kryształ

$$G_{Ph}^E(K, t; Q, \tau) = \left\{ \sum_{\substack{n=1 \\ m=1 \\ k=1}}^{\infty} D_{n,m,k}^E e^{-\lambda_{n,m,k} \frac{D_K}{\rho_K} C(t-\tau)} \frac{8}{abc} W(x, \xi; y, \eta; z, \beta) \right\}. \quad (12)$$

Własności funkcji Greena (12) spełniają zasadę stałego obrazu wzajemnościowego dla punktów [11–12], [15]:

$$\begin{array}{ccc} K(x, y, z, t) & \longleftrightarrow & Q(\xi, \eta, \beta, \tau) \\ \in \bar{\Omega}_{K(a,b,c)t} & \text{Stały wzajemnościowy obraz.} & \in \Omega_{Rt} \end{array} \quad (13)$$

2^o.2.2. Transport entalpii w strumieniu dyfuzji roztwór–kryształ (zmienny punkt $Q(\xi, \eta, \beta, \tau)$)

Rozwiązanie zagadnienia tematowego rozpoczynamy od przypomnienia równania ($\Pi_{3.2}^{hp}$) mającego postać

$$E \frac{\partial T_K}{\partial t} = \frac{H_c D_K}{c_{pK} \rho_K} CV^2 E. \quad (1)$$

Równanie źródłowe dla równania tematowego (1) ma postać

$$E \frac{dT_K}{dt} = \frac{H_c}{c_{pK}} \frac{dE}{dt} \quad (2)$$

a wynika ono z kombinacji równania (1) oraz z następującej postaci równania dyfuzji stężenia roztwór–kryształ

$$\frac{H_c}{c_{pK}} \frac{\partial E}{\partial t} = \frac{H_c D_K}{c_{pK} \rho_K} CV^2 E. \quad (3)$$