

KOMITET METROLOGII I APARATURY NAUKOWEJ
POLSKIEJ AKADEMII NAUK

PROBLEMY METROLOGII ELEKTRONICZNEJ I FOTONICZNEJ 2

Praca zbiorowa
pod redakcją Janusza Mroczi



Oficyna Wydawnicza Politechniki Wroc³awskiej

PROBLEMY METROLOGII
ELEKTRONICZNEJ I FOTONICZNEJ

KOMITET METROLOGII I APARATURY NAUKOWEJ
POLSKIEJ AKADEMII NAUK

**PROBLEMY METROLOGII
ELEKTRONICZNEJ
I FOTONICZNEJ
2**

*Praca zbiorowa
pod redakcją Janusza Mroczki*



Oficina Wydawnicza Politechniki Wrocławskiej
Wrocław 2009

Opracowanie redakcyjne i korekta

Hanna JUREK

Projekt okładki

Janusz MROCZKA

Wszelkie prawa zastrzeżone. Żadna część niniejszej książki, zarówno w całości, jak i we fragmentach, nie może być reprodukowana w sposób elektroniczny, fotograficzny i inny bez zgody wydawcy i właściciela praw autorskich.

© Copyright by Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2009

OFICYNĄ WYDAWNICZĄ POLITECHNIKI WROCŁAWSKIEJ

Wybrzeże Wyspiańskiego 27, 50-370 Wrocław

<http://www.oficyna.pwr.wroc.pl>

e-mail: oficwyd@pwr.wroc.pl

ISBN 978-83-7493-468-8

Drukarnia Oficyny Wydawniczej Politechniki Wrocławskiej. Zam. nr 503/2009.

SPIS RZECZY

Przedmowa	7
Damian Konrad SZCZUCZYŃSKI, Janusz MROCZKA, Matematyczno-fizyczne uwarunkowania zagadnienia odwrotnego w analizie wielkości cząstek układów dyspersyjnych z wykorzystaniem rozpraszania światła	11
Waldemar WÓJCIK, Sławomir CIĘSZCZYK, Wyznaczanie informacji na podstawie widm absorpcyjnych oraz emisyjnych z wykorzystaniem metod modelowania matematycznego	93
Czesław ŁUKIANOWICZ, Skaterometria powierzchni nierównych	165
Marek ZIELIŃSKI, Anna ZAWADZKA, Dariusz CHABERSKI, Pomiary składu izotopowego wiązki atomowej z wykorzystaniem spektrometru masowego czasu przelotu	223
Ewa FRĄCZEK, Wojciech FRĄCZEK, Janusz MROCZKA, Wiry optyczne jako znaczniki fazy w polu optycznym	323
Piotr MASŁOWSKI, Katarzyna BIELSKA, Agata CYGAN, Roman CIURYŁO, Ryszard S. TRAWIŃSKI, Rola pomiaru częstotliwości w pomiarach i analizie kształtu linii widmowych	387

Gdy badamy wszechświat w obu kierunkach, napotykamy na struktury o wymiarach kwintylion razy mniejszych lub większych od naszej skali. Ponieważ zakres ten jest skończony, możemy mieć nadzieję, że któregoś dnia w pełni zrozumiemy dostępne nam struktury istniejące we wszechświecie, od najmniejszych do największych.

Stephen Hawking, Cambridge, 3 lutego 1984

PRZEDMOWA

W każdej dziedzinie nauki należącej do nauk przyrodniczych można wyróżnić trzy grupy zagadnień, które są rozwijane przez zespoły ludzkie uprawiające daną dziedzinę. Należą do nich problemy: teoretyczne, empiryczne i aplikacyjne. Powszechnie znany jest trwający od wieków spór, co jest pierwsze (ważniejsze) czy teoria, czy eksperyment. Wypowiedzi różnych uczonych z różnych okresów ukazują wagę danego problemu. Arystoteles, który wypowiadając się na temat roli doświadczenia w poznaniu, pisał: „(...) byt działa jako bodziec i wywołuje wrażenie; do wrażenia dołącza się myślenie i rozwija swoją działalność”. Einstein twierdził: „(...) doświadczenie może potwierdzić lub obalić teorię, ale nawet setki doświadczeń nie prowadzą do teorii”.

Heller pisał: „(...) doświadczenie na zawsze pozostanie punktem wyjścia i ostatecznym sprawdzianem naukowych teorii, ale świat naszych badań jest bogatszy od ciasnej empirii”. Odnosząc to do metrologii można powiedzieć, że problemy teoretyczne i empiryczne wzajemnie się przeplatają, tworząc – mimo indywidualnej odrębności każdej z nich – harmonijną jedność. To ich wzajemne interakcje tworzą bogactwo i piękno poznawcze metrologii.

Wątek aplikacyjny, którym zajmuje się miernictwo, jest najlepszym przykładem ukazującym sposób wdrażania osiągnięć metodologicznych metrologii. Przeważającą część wszystkich zagadnień empirycznych oraz technicznych stanowią pomiary. Stąd też zawsze aktualne jest pytanie: jakie możliwości poznania daje pomiar oraz jakie są granice jego stosowalności?

Zmierzyć znaczy porównać daną wielkość fizyczną z jednostką umowną, tj. ściśle określoną porcją danej wielkości. Wynik pomiaru danej wielkości fizycznej to iloczyn liczbowej wartości badanej wielkości oraz jej jednostki miary.

Aby dany pomiar był możliwy do wykonania, muszą być spełnione następujące warunki:

- badana wielkość fizyczna musi być ściśle zdefiniowana i mieć stałą wartość w trakcie pomiaru i w warunkach jego dokonywania,
- jednostka miary danej wielkości musi być zdefiniowana, niezmienna, odtwarzalna i umożliwiająca wykonanie porównania,
- porównanie winno być możliwe i to ze znaną dokładnością,
- winniśmy dysponować odpowiednimi środkami do dokonania pomiaru.

Granice pomiaru wynikają z niemożliwości spełnienia jednego z podanych postulatów. Granice te mogą ulegać zmianie wraz z postępem, jaki następuje w przedmiotach realnego poznania, bądź też ostatecznie, o ile wynikają one z możliwości poznawczych podmiotu realnego poznania. Podstawowym postulatem metrologii jest fakt, że wynik pomiaru nie jest identyczny z wartością rzeczywistą wielkości mierzonej. Zasadniczą tego przyczyną jest ograniczona zdolność naszych zmysłów oraz narzędzi realnego poznania, które to są niczym innym niż tworem naszych możliwości intelektualnych czy też predyspozycji lub niedyspozycji naszej świadomości jako najistotniejszego podmiotu w procesie poznania. Dzisiaj można bez cienia wątpliwości stwierdzić, że pomiar należy do jednej z podstawowych czynności naukowych. Obserwując osiągnięcia człowieka w poznaniu, zarówno w skali mikro, jak i makro, zauważamy pojawiające się konsekwencje „natury” samego pomiaru. Zmusza nas to do akceptacji i przyjmowania takich pojęć jak prawdopodobieństwo, niepewność, przedział ufności, rozmycie, ziarnistość itp. Utwierdza nas to w przekonaniu, że powszechnie przyjęty model w odniesieniu do otaczającej nas rzeczywistości o charakterze statycznym czy dynamicznym winien być przez nas bardziej wnikliwie analizowany, a szczególnie fakt, że pomiar w swojej istocie zakłada w chwili komparacji statyczny charakter porównywanych wielkości.

Istota dokonywanych przez nas pomiarów określona jest w dziedzinie i zakresie stosowalności metod i pojęć fizyki klasycznej, traktując zjawiska jako ciągłe w czasie i zdeterminowane. Coraz częściej zauważamy, że to nasze „codzienne doświadczenie” ukształtowane przez determinizm, w którym dane nam było wyrastać, nie może być dosłownie przenoszone na prezentowaną i rozwijaną przez nas metodologię w procesie poznania, a w przypadku opracowywania coraz to nowszych narzędzi realnego poznania winno być wykorzystywane świadomie i z dużą pokorą. To nasze „codzienne doświadczenie” niekiedy utwierdza nas w przekonaniu, że otaczająca nas rzeczywistość jest ciągła i zdeterminowana, a skonstruowany przez nas przyrząd pomiarowy tylko w „niewielkim” stopniu może wpływać na badane przez nas zjawisko. W rzeczywistości jest inaczej i żadne z podanych założeń nie jest ściśle spełnione. O tym informuje nas zasada nieoznaczoności, która w swoim pierwotnym zapisie dotyczyła ograniczenia w możliwości opisu toru, po którym porusza się cząstka w zależności od jej pędu i jakże dobrze znanemu nam pomiarowi, w którym rejestratorem toru poruszającej się cząstki była klisza fotograficzna o określonym liniowym rozmiarze ziarna

emulsji fotograficznej ($\Delta x \approx 10^{-6}$ m), co pozwoliło określić niedokładność pomiaru prędkości na poziomie $\Delta v \approx 10^2$ m/s.

Zasada nieoznaczoności dotyczy nie tylko położenia i pędu. W formalizmie mechaniki kwantowej występują również inne tzw. wielkości kanonicznie sprzężone. Przykładem może być tu czas i energia. W praktyce oznacza to, że na wykonanie jakiegokolwiek pomiaru w stanie nietrwałym mamy ograniczony czas – rzędu średniego czasu życia układu w tym stanie. Chcąc mieć pewność, iż układ znajduje się w stanie o ustalonej energii, winniśmy obserwować go nieskończenie długo. Podobnie chcąc przekonać się, że fala jest falą płaską, winniśmy ją obserwować w całej nieskończonej przestrzeni. Stany nietrwałe muszą więc charakteryzować się pewnym rozmyciem energii, które w ich układzie spoczynkowym odpowiada rozmyciu ich masy. Widzimy więc, że tylko stan podstawowy będzie stanem nierozmytym, natomiast wszystkie stany wzbudzone nie będą miały dokładnie określonej wartości energii. Stąd też linie emitowane przez atom mają pewną naturalną szerokość, co oznacza, że zwiększając dokładność pomiaru, nie uczynimy tej linii węższą. Mówiąc o pomiarze w mikroświecie, uzmysławiamy sobie dwa fakty. Pierwszy, wynikający z zasady nieoznaczoności, który ostrzega nas o bezkrytycznym stosowaniu pojęć fizyki klasycznej do opisu stanu makroobiektów. Drugi, że przeprowadzane doświadczenia informujące nas o stanie mikroobiektów mają charakter makroskopowy (klisza fotograficzna), czyli aparat pojęciowy użyty do ich opisu powinien odpowiadać teoriom, które go stworzyły, a wywodzą się z fizyki klasycznej. Równania mechaniki kwantowej opisują stan łączny układu; w pomiarze jest to badane zjawisko i przyrząd pomiarowy. Niemożliwe jest zatem wydzielenie z wyniku pomiaru stanu samego zjawiska, jaki istniał przed wprowadzeniem przyrządu. Dobre wytłumaczenie tego podaje Heisenberg, mówiąc o poruszających się fotonach mogących ze sobą interferować. Chcąc to zjawisko zaobserwować, musielibyśmy zużyć przynajmniej jeden z nich, co uniemożliwi zajście zjawiska interferencji. To umożliwia nam zauważenie, że możliwość przeprowadzenia obserwacji daje nam aparat pojęciowy fizyki klasycznej. Postęp, w jakim rozwijają się współczesne narzędzia realnego poznania dzięki coraz to doskonalszym zdobyczom elektroniki, informatyki oraz nowym sposobom detekcji wielkości fizycznych, pozwala sądzić, że tylko czas dzieli nas od kolejnych ciekawych osiągnięć w tej dziedzinie. Aby urealnić nasze oczekiwania, zauważmy jakie ograniczenia muszą być minimalizowane. Są to: ograniczenia wymiarowe określone w zasadzie nieoznaczoności; ograniczenia ilościowe wynikające z możliwości poznania z zadawalającą dokładnością tylko wybranych cech badanego zjawiska oraz ograniczenia wynikające z niedoskonałego poznania uwarunkowanego świadomością obserwatora będącego ostatecznym odbiorcą rezultatów eksperymentu poznania. Nietrudno wskazać, które z tych ograniczeń będzie wymagało długiego czasu, a może nawet kilku pokoleń.

Kontynuując idee publikowania corocznie osiągnięć naukowych metrologów pod auspicjami Komitetu Metrologii i Aparatury Naukowej Polskiej Akademii Nauk pragniemy przekazać kolejny tom stanowiący zbiór opracowań różnych ośrodków, gdzie

zagadnienia metrologiczne są rozwijane, a prezentowane materiały stanowiły w większości prace na stopień. Tą formą pragniemy dowieść kolejne tezy:

- metrologia stanowi metodologiczną podstawę wszystkich nauk empirycznych i odkrywa jedność nauk przyrodniczych,

- techniką metrologii jest miernictwo, za pomocą którego dokonujemy wartościowania naszych spostrzeżeń oraz które rozwija praktyki inżynierskie innych nauk, jak medycyna, biologia, chemia, biotechnologia i wiele nauk im pokrewnych,

- to rozwój coraz to doskonalszych narzędzi pomiarowych decyduje o jakości i rozwoju produkcji przemysłowej, a w niektórych kategoriach może być przedmiotem narodowej specjalności w międzynarodowym podziale rynku *high-tech*,

- autoidentyfikacja środowiska metrologów nie zawsze pokazuje jej rzeczywistą liczebność, przeważająca bowiem większość inżynierów i badaczy identyfikuje się z dziedziną zastosowań pomiarów, a nie samymi pomiarami, co można zauważyć w osiągnięciach nauk pokrewnych, np. bioniki, fotoniki, a szczególnie, że jako środowisko metrologiczne wypracowujemy swoje trwałe miejsce w intelektualnym dialogu interdyscyplinarnym, prezentując jedność poczynąń metrologicznych, mimo spontanicznej specjalizacji.

Janusz Mroczka

*zagadnienie odwrotne, pomiary pośrednie,
równanie całkowe Fredholma pierwszego rodzaju,
pomiary nefelometryczne, pomiary turbidymetryczne*

Damian Konrad SZCZUCZYŃSKI,
Janusz MROCZKA*

MATEMATYCZNO-FIZYCZNE UWARUNKOWANIA ZAGADNIENIA ODWROTNEGO W ANALIZIE WIELKOŚCI CZĄSTEK UKŁADÓW DISPERSYJNYCH Z WYKORZYSTANIEM ROZPRASZANIA ŚWIATŁA

Opracowanie przedstawia wybrane aspekty zagadnienia odwrotnego w wyznaczaniu rozkładu wielkości cząstek fazy zdyspergowanej układów dyspersyjnych na podstawie wyników pomiarów charakterystyk rozpraszania światła w układzie dyspersyjnym. Problem ten reprezentuje niezwykle istotne w metrologii zagadnienie odwrotne w pomiarach pośrednich. Przez analizę fizycznych podstaw pomiarów nefelometrycznych i turbidymetrycznych w pracy wykazano, że przy odpowiednich założeniach upraszczających modele matematyczne obu rodzajów pomiarów przyjmują postać równania całkowego Fredholma pierwszego rodzaju. Dzięki temu zagadnienie odwrotne w pomiarach nefelometrycznych i turbidymetrycznych rozważać można na podstawie ogólnej teorii problemów odwrotnych wyrażonych równaniem całkowym Fredholma pierwszego rodzaju, czemu poświęcono dalsze części opracowania. Wykazano złe postawienie w sensie Hadamarda tej klasy zagadnień odwrotnych. Na podstawie aparatu pojęciowego analizy funkcjonalnej, a szczególnie teorii funkcji osobliwych, ukazano przyczyny słabego uwarunkowania tego typu problemów odwrotnych. Następnie omówiono wybrane techniki rozwiązywania rozpatrywanego zagadnienia odwrotnego, koncentrując się na metodach numerycznych. Szczególną uwagę poświęcono roli wiedzy apriorycznej o poszukiwanym rozwiązaniu, którą procedury inwersyjne posługują się w różnych postaciach w celu minimalizacji efektów słabego uwarunkowania. W ostatniej części opracowania zaprezentowano badania symulacyjne prowadzone przez autorów. Wśród omawianych przedsięwzięć badawczych przedstawiono badania dotyczące metody regularyzacji Tichonowa z ograniczeniami. Przytaczane wyniki wykazują istotną przewagę tej techniki nad klasyczną regularyzacją.

1. WSTĘP

Jednym z kluczowych problemów metrologii jest zagadnienie odwrotne w pomiarach pośrednich. Polega ono na wyznaczeniu określonej charakterystyki badanego

* Politechnika Wroclawska, Wydział Elektroniki, Katedra Metrologii Elektronicznej i Fotonicznej, ul. B. Prusa 53/55, 50-317 Wrocław.

obiekty reprezentowanej przez funkcję $f(a)$ na podstawie innej funkcji $g(y)$ stanowiącej przedmiot bezpośrednich pomiarów. Zadanie opiera się na zależności wiążącej mierzoną funkcję $g(y)$ z szukaną funkcją $f(a)$:

$$g(y) = F[f(a)], \quad (1.1)$$

gdzie F – operator. Relacja (1) stanowi model matematyczny pomiarów pośrednich.

Problem wyznaczenia funkcji $f(a)$ na podstawie funkcji $g(y)$ według zależności (1.1) określa się jako zagadnienie odwrotne dla odróżnienia go od zagadnienia wprost polegającego na wyznaczeniu wyjścia modelu matematycznego pomiarów, czyli funkcji $g(y)$ dla zadanej funkcji $f(a)$. Zadanie wprost odpowiada więc symulacji pomiarów z zastosowaniem modelu matematycznego (1.1).

Relację (1.1) interpretować można w kategoriach związku przyczynowo-skutkowego. Rozważaną charakterystykę $f(a)$ obiektu traktować można jako przyczynę, natomiast odpowiadającą jej funkcję $g(y)$, będącą rezultatem pomiaru, jako skutek. W związku z tym zagadnienie wprost rozumieć można jako przewidywanie skutków wynikających ze znanej przyczyny, natomiast zagadnienie odwrotne jako odtwarzanie przyczyny obserwowanych skutków.

Zagadnienie odwrotne opisane związkiem (1.1) ma jednoznaczne rozwiązanie tylko wówczas, gdy odwzorowanie (1.1) jest różnowartościowe (jednojednoznaczne), czyli różnym funkcjom $f(a)$ przyporządkowuje różne funkcje $g(y)$. W takim przypadku rozwiązanie (1.1) formalnie można wyrazić przez odwzorowanie odwrotne:

$$f(a) = F^{-1}[g(y)]. \quad (1.2)$$

Uzyskana w wyniku rzeczywistych pomiarów funkcja $\tilde{g}(y)$ dotknięta jest wpływem zakłóceń pomiarowych. W związku z tym może ona nie należeć do obrazu operatora F , a co za tym idzie równanie (1.1), które dla funkcji $\tilde{g}(y)$ przyjmuje postać:

$$\tilde{g}(y) = F[f(a)], \quad (1.3)$$

może nie mieć rozwiązania. W takim przypadku poszukuje się rozwiązania przybliżonego $\hat{f}(a)$, które uzyskuje się z zastosowaniem określonej aproksymacji operatora odwrotnego \hat{F}^{-1}

$$\hat{f}(a) = \hat{F}^{-1}[\tilde{g}(y)]. \quad (1.4)$$

Wiele zagadnień odwrotnych wykazuje tzw. złe postawienie (ang. *ill-posedness*), tzn. nie spełnia warunków dobrego postawienia (ang. *well-posedness*) Hadamarda [1, 2]:

- rozwiązanie $f(a)$ istnieje dla każdej funkcji $g(y)$,
- rozwiązanie $f(a)$ jest jednoznaczne,
- rozwiązanie $f(a)$ jest stabilne, tzn. zależy w sposób ciągły od funkcji $g(y)$.

Warunek stabilności rozwiązania sformułować można również jako postulat, by dla każdego dążącego do zera ciągu zaburzeń funkcji $g(y)$ odpowiadający mu ciąg zaburzeń rozwiązania $f(a)$ także dążył do zera [2].

Z niewłaściwym postawieniem zagadnienia wiąże się również jego słabe uwarunkowanie, które polega na tym, że niewielka nawet niepewność danych pomiarowych $\tilde{g}(y)$ prowadzi do olbrzymiej niepewności rozwiązania $f(a)$ uzyskiwanego na podstawie tych danych [2, 3]. Ze względu na jakość, oznacza to, że wyrażony równaniem (1.3) warunek zgodności rozwiązania $f(a)$ z danymi pomiarowymi $\tilde{g}(y)$ nie wystarcza do dokładnego i jednoznacznego określenia rozwiązania, ponieważ bardzo szeroki przedział funkcji $f(a)$ jest zgodny z danymi pomiarowymi $\tilde{g}(y)$ z dokładnością do nawet bardzo małej niepewności pomiarowej $\tilde{g}(y)$ [2].

Wynikająca ze słabego uwarunkowania quasi-niedookreśloność zagadnienia powoduje, że uzyskanie rozwiązania o niewielkim poziomie niepewności wymaga zastosowania dodatkowej wiedzy o rozwiązaniu posiadanej *a priori*. Informacja aprioryczna w postaci dodatkowych warunków nakładanych na poszukiwane rozwiązanie umożliwia wybranie wąskiego przedziału rozwiązań spośród bardzo szerokiego przedziału rozwiązań spełniających wyjściowe zagadnienie odwrotne [3].

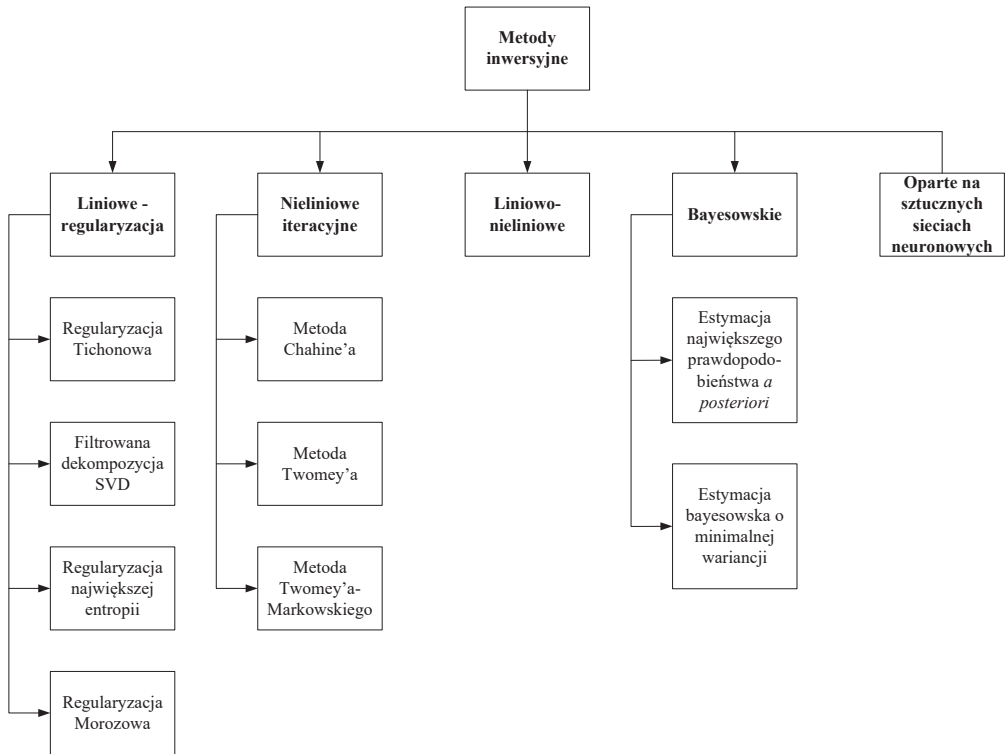
Obszerną klasę problemów odwrotnych stanowią zagadnienia, w których operator F stanowi operator całkowy Fredholma [1, 4]:

$$g(y) = K[f(a)] = (Kf)(y) = \int_{a_1}^{a_2} K(y, a) f(a) da. \quad y_1 \leq y \leq y_2. \quad (1.5)$$

Równanie (1.5) jest równaniem całkowym Fredholma pierwszego rodzaju. Zagadnienie odwrotne wyrażone równaniem (5) jest ogólnie źle postawione i słabo uwarunkowane [2, 4].

Analityczne rozwiązanie równania całkowego Fredholma pierwszego rodzaju znalezione zostały tylko dla nielicznych, szczególnych postaci funkcji jądra całkowego $K(y, a)$ [5]. W związku z tym w większości przypadków zagadnienie odwrotne rozwiązuje się z zastosowaniem metod numerycznych. W tym celu równanie (1.5) poddaje się dyskretyzacji, której rezultatem jest układ algebraicznych równań liniowych [1, 3, 4, 5]. W następstwie złego postawienia i słabego uwarunkowania zagadnienia (1.5) otrzymywany układ równań wykazuje bardzo słabe uwarunkowanie numeryczne [1, 3, 4, 5].

Opracowanych zostało wiele numerycznych technik rozwiązywania problemów odwrotnych, wyrażonych równaniem całkowym Fredholma pierwszego rodzaju. Każda z tych metod zwanych algorytmami lub procedurami inwersyjnymi wykorzystuje określoną formę wiedzy apriorycznej o poszukiwanym rozwiązaniu, w celu minimalizacji efektów słabego uwarunkowania zagadnienia. Ogólną klasyfikację metod inwersyjnych zaprezentowano na rysunku 1.1 [1, 3, 4].



Rys. 1.1. Ogólna klasyfikacja metod inwersyjnych wraz z przykładami technik należących do poszczególnych grup
 Fig. 1.1. General classification of the inverse methods with examples of techniques belonging to particular groups

Wyróżnić można kilka grup technik inwersyjnych. Metody liniowe poszukują najlepszej aproksymacji operatora odwrotnego do operatora odwzorowującego dyskretną reprezentację funkcji $f(a)$ w dyskretną reprezentację funkcji $g(y)$. Techniki te znajdują uzasadnienie teoretyczne na gruncie analizy funkcjonalnej i algebry liniowej. Są one reprezentowane przez różne odmiany regularyzacji [1, 3, 4].

Iteracyjne metody nieliniowe znajdują rozwiązanie w toku iteracyjnych poprawek założonego a priori rozwiązania początkowego. Korekty te mają charakter nieliniowy. Ich kierunek i wielkość dobierana jest tak, by rozwiązania uzyskiwane w kolejnych iteracjach wykazywały coraz lepszą zgodność z danymi pomiarowymi. Dzięki temu proces poprawek jest zbieżny z prawdziwym rozwiązaniem problemu [3].

Do rozwiązywania zagadnienia stosuje się również sztuczne sieci neuronowe. Sieć uczone jest (przez korekcję współczynników wagowych) odwzorowania dyskretniej reprezentacji funkcji $f(a)$ w dyskretną reprezentację funkcji $g(y)$, czyli modelu pomiarów (1.5) lub ogólniej (1.1). Nauczona sieć działa jako model odwrotny, tzn. po poda-

niu na jej wejście dyskretnej reprezentacji funkcji $g(y)$, na jej wyjściu otrzymuje się dyskretną reprezentację funkcji $f(a)$ [6].

Osobną grupę metod inwersyjnych stanowią techniki bayesowskie. W odróżnieniu od wcześniej omówionych procedur nie dokonuje się tu jawnego rozwiązywania zagadnienia odwrotnego. W metodach tych zakłada się pewien rozkład prawdopodobieństwa *a priori* rozwiązania, odzwierciedlający niepewność wiedzy apriorycznej o rozwiązaniu. Następnie rzeczywiste dane pomiarowe wykorzystuje się do skorygowania wiedzy apriorycznej o rozwiązaniu. Odbywa się to przez wyznaczenie według twierdzenia Bayesa warunkowego rozkładu prawdopodobieństwa rozwiązania pod warunkiem, że w rezultacie pomiarów uzyskane zostały określone dane pomiarowe. Ten wynikowy rozkład prawdopodobieństwa rozwiązania, zwany rozkładem *a posteriori*, stanowi probabilistyczną informację o rozwiązaniu. Na jego podstawie możliwe jest wyznaczenie różnych estymatorów rozwiązania [3]. Estymator największego prawdopodobieństwa *a posteriori* oznacza rozwiązanie najbardziej prawdopodobne, pod warunkiem zarejestrowania określonych wyników pomiarów [3]. Estymator bayesowski o minimalnej wariancji stanowi natomiast wartość oczekiwaną rozwiązania, pod warunkiem zarejestrowania określonych danych pomiarowych [3]. Techniki bayesowskie w rzeczywistości realizowane są z wykorzystaniem symulacji Monte Carlo. W tym przypadku w celu uzyskania rozkładu prawdopodobieństwa *a posteriori* rozwiązania dokonuje się analizy statystycznej wielkiej liczby rozwiązań zagadnienia wprost otrzymywanych dla dyskretnej reprezentacji funkcji $f(a)$ losowanych z założonego rozkładu prawdopodobieństwa *a priori* [3].

Przy odpowiednich założeniach upraszczających za pomocą równania całkowego Fredholma pierwszego rodzaju wyrazić można rozważane w niniejszej pracy zagadnienie odwrotne, polegające na wyznaczeniu rozkładu wielkości cząstek fazy zdyspergowanej układu dyspersyjnego na podstawie mierzonych charakterystyk rozpraszania światła w układzie dyspersyjnym [5]. Do charakterystyk tych należą: zależność objętościowej funkcji rozpraszania od kąta rozpraszania $\beta(\theta)$, będąca przedmiotem pomiarów nefelometrycznych oraz zależność całkowitego współczynnika osłabienia światła od długości fali rozpraszanego światła $c(\lambda)$, która jest przedmiotem pomiarów turbidymetrycznych [5]. Założenia upraszczające postulują, że w układzie dyspersyjnym zachodzi jednokrotne i niekoherentne rozpraszanie światła będącego niespolaryzowaną falą płaską. Założenia te umożliwiają łatwe wyprowadzenie modeli matematycznych pomiarów nefelometrycznych i turbidymetrycznych z teorii rozpraszania światła na pojedynczej cząstce fazy zdyspergowanej. Spośród wielu teorii tego typu w pracy rozważane są dwa modele: teoria rozpraszania Mie oraz teoria dyfrakcji Fraunhofera stanowiąca przybliżenie teorii Mie [5, 7, 8].

Przedmiotem zainteresowania w niniejszym opracowaniu są: modele matematyczne pomiarów nefelometrycznych i turbidymetrycznych, złe postawienie oraz słabe uwarunkowanie zagadnienia odwrotnego w analizie wielkości cząstek układów dyspersyjnych według pomiarów nefelometrycznych i turbidymetrycznych oraz nume-

rycznych metod rozwiązywania tego problemu wykorzystujących różne formy wiedzy apriorycznej o rozwiązaniu.

2. ANALIZA WIELKOŚCI CZĄSTEK UKŁADÓW DYSPERSYJNYCH

2.1. UKŁADY DYSPERSYJNE

Układy dyspersyjne są to układy dwu- lub wielofazowe złożone z fazy ciągłej, zwanej fazą rozpraszającą lub ośrodkiem dyspersyjnym, i zawieszonych w niej drobnych cząstek stanowiących fazę rozproszoną, zwaną też fazą zdyspergowaną [9].

2.2. WIELKOŚĆ CZĄSTEK FAZY ZDYSPERGOWANEJ UKŁADU DYSPERSYJNEGO

Jedną z najważniejszych właściwości fizycznych układów dyspersyjnych jest wielkość cząstek fazy zdyspergowanej, ponieważ ma ona zasadniczy wpływ na pozostałe właściwości układu [5, 9].

Cząstki fazy zdyspergowanej mogą mieć różny, często nieregularny, kształt. Wymaga to sprecyzowania miar wielkości (rozmiaru) cząstek, gdyż w tych warunkach średnica w znaczeniu dosłownym nie może być stosowana jako parametr charakteryzujący wielkość cząstek [5, 9].

W wielu zjawiskach fizycznych wykorzystywanych w analizie wielkości cząstek fazy zdyspergowanej zachowanie rzeczywistych cząstek opisać można w przybliżeniu jako zachowanie cząstek kulistych o określonej średnicy, zwanej średnicą zastępczą [5, 9]. W związku z tym średnice zastępcze, definiowane odpowiednio do przyjętej metody pomiaru, stosuje się jako miary wielkości cząstek. Istnieje wiele rodzajów średnic zastępczych [5, 9]. Przybliżenie kształtu rzeczywistej cząstki kulą o odpowiedniej średnicy zastępczej redukuje liczbę parametrów koniecznych do scharakteryzowania cech geometrycznych cząstki do jednego parametru – średnicy zastępczej, co stanowi istotne uproszczenie opisu matematycznego zjawisk fizycznych zachodzących z udziałem cząstek fazy zdyspergowanej [5].

W przypadku analizy wielkości cząstek fazy zdyspergowanej z zastosowaniem rozpraszania światła jako średnicę zastępczą przyjmuje się średnicę objętości d_v , która stanowi średnicę kuli o objętości takiej samej jak badana cząstka [5, 9]. W analogiczny sposób określić można promień objętości a_v . Przybliżenie rzeczywistych cząstek za pomocą cząstek kulistych o tej samej objętości uzasadniają wyniki wielu prac doświadczalnych i teoretycznych, które wykazały, że w warunkach naturalnych cząstki fazy zdyspergowanej, charakteryzujące się dużą różnorodnością kształtów i chaotyczną orientacją w przestrzeni, rozpraszają światło tak, jak jednorodne kule o identycznej objętości [5].

W dalszej części opracowania przez promień cząstki a rozumiany będzie zawsze jej promień objętości a_v .

Fazę zdyspergowaną, tworzoną przez cząstki o różnej wielkości, opisuje się za pomocą funkcji rozkładu wielkości cząstek. Rozkład wielkości cząstek jest funkcją promienia cząstek a , określającą liczbę cząstek dN_v w jednostce objętości układu dyspersyjnego, których promienie leżą w nieskończenie małym przedziale $(a, a + da)$, według wzoru [5]:

$$dN_v = N_v f(a) da, \quad (2.2.1)$$

gdzie N_v – całkowita liczba cząstek w jednostce objętości układu dyspersyjnego. Całkowitą liczbę cząstek o promieniu $a_1 \leq a \leq a_2$ w jednostce objętości układu określa wzór [5]:

$$N_v(a_1, a_2) = N_v \int_{a_1}^{a_2} f(a) da. \quad (2.2.2)$$

Rozkład wielkości spełnia warunek unormowania do jedności:

$$\int_0^{\infty} f(a) da = 1. \quad (2.2.3)$$

2.3. ZNACZENIE ANALIZY WIELKOŚCI CZĄSTEK UKŁADÓW DISPERSYJNYCH

Analiza wielkości cząstek fazy zdyspergowanej układów dyspersyjnych ma bardzo duże znaczenie w wielu dziedzinach działalności naukowo-technicznej i przemysłowej takich, jak [5]:

- chemia fizyczna,
- inżynieria materiałowa,
- nauki o ziemi,
- biochemia, biofizyka i biologia molekularna,
- medycyna,
- farmacja i przemysł kosmetyczny,
- przemysł spożywczy,
- ochrona środowiska,
- optyka morza.

Zainteresowanie analizą wielkości cząstek wynika z powszechności układów dyspersyjnych w przyrodzie. Ich przykłady stanowią: wody naturalne z zawieszonym w nich planktonem oraz cząstkami nieorganicznymi, krew i kolonie bakteryjne, unoszące się w powietrzu pyły, dymy i mgły, masa perłowa oraz minerały takie, jak opal.

Układami dyspersyjnymi są również liczne produkty przemysłowe – od artykułów spożywczych takich, jak mleko i majonez przez leki i kosmetyki w postaci past, kremów, żeli i pianek po materiały budowlane takie, jak beton, farby emulsyjne i stałe pianki [5]. Do układów dyspersyjnych zalicza się również tworzywa kompozytowe.

2.4. TECHNIKI ANALIZY WIELKOŚCI CZĄSTEK UKŁADÓW DYSPERSYJNYCH

Istnieje wiele metod pomiaru wielkości cząstek fazy zdyspergowanej. Do klasycznych technik pomiarowych zalicza się metody [5]:

- przesiewania,
- pomiarów mikroskopowych,
- grawimetryczną,
- sedymentacji,
- centryfugi,
- konduktometryczną.

Zasadniczą wadą tych technik jest ich inwazyjność – konieczność wstępnego przygotowania badanej próbki, co często prowadzi do zmiany jej pierwotnych właściwości [5].

Wady tej pozbawione są optyczne metody pomiarowe oparte na zjawisku rozpraszania światła w układach dyspersyjnych. Przewaga technik optycznych nad pozostałymi metodami pomiaru wielkości cząstek wynika z ich bezinwazyjności, związanej z tym szybkości oraz możliwości prowadzenia pomiarów *in situ* [5]. Optycznym metodom analizy wielkości cząstek fazy zdyspergowanej układów dyspersyjnych poświęcona jest dalsza część niniejszego opracowania.

3. PODSTAWY FIZYCZNE ORAZ IDEA OPTYCZNYCH TECHNIK ANALIZY WIELKOŚCI CZĄSTEK UKŁADÓW DYSPERSYJNYCH

Optyczne techniki analizy wielkości cząstek układów dyspersyjnych należą do klasy pomiarów pośrednich. Bezpośrednim pomiarom poddawane są wielkości fizyczne opisujące rozpraszanie światła na cząstkach fazy zdyspergowanej układu dyspersyjnego, a następnie na podstawie wyników tych pomiarów wyznaczany jest rozkład wielkości cząstek fazy zdyspergowanej z wykorzystaniem związku matematycznego pomiędzy mierzonymi wielkościami a szukaną funkcją rozkładu. Wniosekowanie o rozkładzie wielkości cząstek na podstawie wyników pomiarów wielkości fizycznych, opisujących rozpraszanie światła w układzie dyspersyjnym, stanowi zagadnienie odwrotne (ang. *inverse problem*).

Istnieje wiele optycznych technik analizy wielkości cząstek układów dyspersyjnych różniących się rodzajem wielkości fizycznych podlegających bezpośrednio

pomiarowi [8, 9]. Niniejsza praca poświęcona jest dwóm technikom: nefelometrii i turbidymetrii. W dalszej części rozdziału przedstawione zostały podstawy fizyczne oraz idee obu tych metod pomiarowych.

3.1. ROZPRASZANIE ŚWIATŁA W UKŁADACH DYSERSYJNYCH

Rozpraszanie światła jest zjawiskiem fizycznym polegającym na zmianie przestrzennego rozkładu wiązki światła na skutek oddziaływania z ośrodkiem, w którym to światło propaguje [10].

Mechanizm rozpraszania fal elektromagnetycznych polega na pochłanianiu promieniowania padającego przez atomy lub cząsteczki ośrodka, co prowadzi do ich wzbudzenia, następstwem którego jest wtórne promieniowanie fal elektromagnetycznych przez te atomy lub cząsteczki przy ich powrocie do niższych poziomów energetycznych. Promieniowanie wtórne stanowi promieniowanie rozproszone w ośrodku [5, 7]. Część pochłoniętej energii elektromagnetycznej może zostać przekształcona w inne formy energii takie, jak energia cieplna. Zjawisko to nosi nazwę absorpcji [7]. Ponieważ rozpraszanie i absorpcja są ściśle ze sobą związane, terminu rozpraszanie używa się często do określenia obu tych zjawisk zachodzących jednocześnie [7].

Promieniowanie rozproszone przez cały układ zawierający olbrzymią liczbę atomów lub cząsteczek jest sumarycznym efektem wtórnego promieniowania elektromagnetycznego emitowanego przez wszystkie atomy lub cząsteczki wzbudzone przez falę padającą [5, 7]. W ośrodkach, w których odległości między atomami lub cząsteczkami są znacznie mniejsze od długości fali światła padającego, każdy atom lub cząsteczka poddana jest oddziaływaniu nie tylko fali padającej, lecz również wypadkowej wtórnych fal wypromieniowanych przez wszystkie pozostałe atomy lub cząsteczki w układzie. Jednocześnie jednak wtórne promieniowanie każdej molekuly zależne jest od całkowitego działającego na nią promieniowania. W efekcie mamy do czynienia z wzajemnym sprzężeniem molekuł, co prowadzi do elektromagnetycznego problemu wielu ciał [7]. Ścisły opis tego zjawiska, określanego jako rozproszenie wielokrotne, jest niezwykle skomplikowany [7, 11].

Jeżeli odległości między atomami lub cząsteczkami w ośrodku są większe, można zaniedbać ich wzajemne sprzężenie, przyjmując przybliżenie rozproszenia jednokrotnego. W tej sytuacji wypadkowe pole elektryczne wewnątrz i na zewnątrz ośrodka obliczane jest jako superpozycja pól elektrycznych wtórnych fal elektromagnetycznych emitowanych przez wszystkie wzbudzone molekuly oraz pola elektrycznego fali padającej. Sumowanie pól uwzględnia różnice faz poszczególnych fal. Zakłada się przy tym, że różnice te są stałe w czasie, czyli fale są ze sobą spójne [5, 7].

Można wykazać, że w idealnie jednorodnym ośrodku wynikiem superpozycji fal wtórnych jest pojedyncza fala propagująca w ośrodku w kierunku zgodnym z kierunkiem propagacji fali padającej z prędkością c/n , gdzie c jest prędkością światła w próżni, a n współczynnikiem charakteryzującym rozważany ośrodek, zwanym

współczynnikiem załamania światła. Fala padająca zostaje w ośrodku całkowicie wygaszona. W przypadku ośrodka jednorodnego, mimo występowania molekularnego mechanizmu rozpraszania, nie zachodzi rozpraszanie w skali makroskopowej, przejawiające się zmianą kierunku propagacji światła. Ścisły opis oddziaływań światła z układami jednorodnymi prowadzi do zależności współczynnika załamania od właściwości ośrodka – liczby molekuł w jednostce objętości oraz ich polaryzowalności, czyli skłonności do zmiany rozkładu ładunku elektrycznego pod wpływem zewnętrznego pola elektrycznego [7].

Rozpraszanie światła rozumiane jako zmiana przestrzennego rozkładu wiązki występuje wyłącznie w ośrodkach niejednorodnych, tj. takich, w których współczynnik załamania nie jest jednakowy we wszystkich punktach, lecz zmienia się od punktu do punktu [5, 7, 10]. W rzeczywistości nawet układy jednorodne makroskopowo są w określonej skali niejednorodne, ponieważ można w nich wyróżnić pojedyncze atomy lub cząsteczki. Rozważania teoretyczne prowadzą jednak do wniosku, że ośrodki, w których odległości między molekułami są znacznie mniejsze od długości fali światła, można uznać za optycznie jednorodne [10].

Przyczyną niejednorodności optycznej ośrodka może być obecność w nim niewielkich cząstek innej substancji [5, 7, 10]. Przez cząstkę rozumie się agregat wielkiej liczby atomów lub cząsteczek jednego rodzaju. W przypadku gdy liczba ta jest wystarczająco duża, właściwości optyczne materiału, z którego cząstka jest zbudowana, w pełni opisuje parametr makroskopowy, jakim jest współczynnik załamania światła [7]. Niejednorodność optyczna tego typu występuje w przypadku układów dyspersyjnych, gdyż faza zdyspergowana tych układów stanowi zbiór cząstek o współczynniku załamania różniącym się od współczynnika załamania ośrodka dyspersyjnego, w którym te cząstki są zawieszane [5].

Podobnie jak w omawianym wcześniej przypadku atomów lub cząsteczek, jeżeli średnie odległości między cząstkami fazy zdyspergowanej są znacznie mniejsze od długości fali światła padającego, następuje wzajemne sprzężenie cząstek za pośrednictwem pola elektromagnetycznego, co prowadzi do zjawiska rozproszenia wielokrotnego [5]. Jak poprzednio, jeżeli średnie odległości cząstek są większe, można zastosować przybliżenie rozproszenia jednokrotnego [5]. W tej sytuacji pole elektryczne fali rozproszonej przez zbiór cząstek fazy zdyspergowanej stanowi superpozycję pól fal rozproszonych przez poszczególne cząstki (z uwzględnieniem faz), z których każda poddana jest oddziaływaniu wyłącznie pola elektrycznego fali padającej tak, jakby stanowiła pojedynczą cząstkę rozpraszającą w całym ośrodku dyspersyjnym.

Jeżeli dodatkowo założy się, że liczba cząstek w układzie dyspersyjnym jest bardzo duża, a ich położenie i orientacja przestrzenna – losowa i zmienna w czasie, fazy fal rozproszonych przez poszczególne cząstki w dowolnym punkcie obserwacji są również losowe (nie związane żadną systematyczną zależnością) oraz zmienne w czasie. Dzięki temu cząstki fazy zdyspergowanej można traktować jako niespójne względem siebie źródła promieniowania rozproszonego. W związku z tym natężenia fal

rozproszonych przez poszczególne cząstki mogą być bezpośrednio sumowane. Rozpraszanie światła przez cały zbiór cząstek fazy zdyspergowanej określa się w tym przypadku jako rozproszenie niespójne lub niekoherentne [5].

Dzięki założeniom o rozpraszaniu jednokrotnym i niekoherentnym możliwe jest łatwe sformułowanie opisu rozpraszania światła w układzie dyspersyjnym na podstawie wyników analizy rozpraszania światła na pojedynczej cząstce fazy zdyspergowanej.

3.2. MODELOWANIE ROZPRASZANIA ŚWIATŁA W UKŁADACH DYSPERSYJNYCH

Jak wynika z rozważań przedstawionych w punkcie 3.1, ścisły opis rozpraszania światła w układach dyspersyjnych jest niezwykle skomplikowany, co powoduje, że jego zastosowanie do analizy wielkości cząstek napotyka poważne trudności lub jest wręcz niemożliwe. Na potrzeby rozwiązywania rozważanego zagadnienia odwrotnego celowe jest wykorzystanie prostszych modeli przybliżonych, przyjmujących wiele założeń upraszczających [5].

W badaniach prowadzonych przez autorów przyjęte zostały następujące założenia upraszczające:

- w układzie zachodzi jednokrotne i niekoherentne rozpraszanie światła,
- rozpraszanie światła ma charakter sprężysty, czyli zachodzi bez zmiany długości fali,
- światło padające stanowi płaska, monochromatyczna i niespolaryzowana fala elektromagnetyczna,
- ośrodek dyspersyjny jest jednorodny i izotropowy,
- fazę zdyspergowaną tworzą jednorodne i izotropowe cząstki o kształcie kulistym,
- właściwości optyczne ośrodka dyspersyjnego i fazy zdyspergowanej w pełni opisują ich zespolone współczynniki załamania światła.

Podane założenia są uzasadnione w wielu rzeczywistych przypadkach nefelometrycznych i turbidymetrycznych pomiarów wielkości cząstek układów dyspersyjnych.

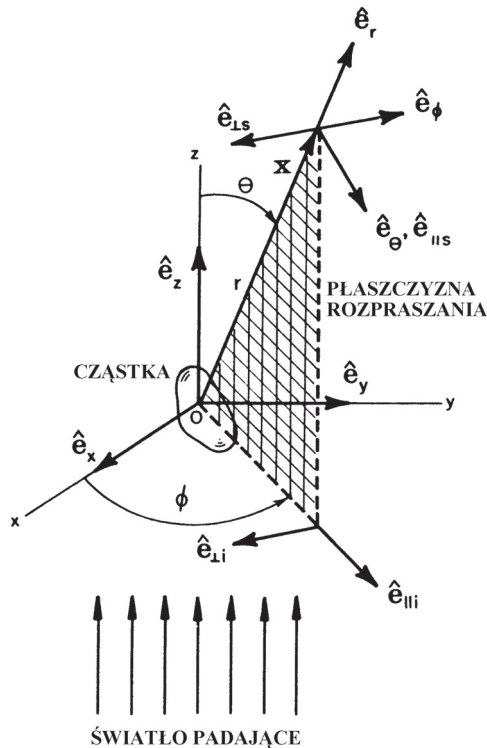
Konsekwencją założenia o rozpraszaniu jednokrotnym i niekoherentnym jest możliwość obliczenia natężenia światła rozproszonego w danym kierunku przez cały układ dyspersyjny, jako sumy natężeń światła rozproszonego w tym kierunku przez poszczególne cząstki fazy zdyspergowanej, z których każda poddana jest oddziaływaniu wyłącznie fali padającej tak, jakby stanowiła pojedynczą cząstkę rozpraszającą w całym ośrodku dyspersyjnym [5]. Dzięki temu właściwości optyczne układu dyspersyjnego w prosty sposób wynikają z właściwości pojedynczych cząstek rozpraszających.

Założenia o rozpraszaniu sprężystym, jednorodności i izotropowości ośrodka dyspersyjnego oraz jednorodności, izotropowości i kulistym kształcie cząstek fazy zdy-

spergowanej umożliwiają zastosowanie do opisu właściwości optycznych pojedynczej cząstki stosunkowo proste teorie fizyczne omówione w dalszej części pracy.

3.2.1. MODELOWANIE ROZPRASZANIA ŚWIATŁA NA POJEDYNCZEJ CZĄSTCE FAZY ZDYSPERGOWANEJ

Układ współrzędnych stosowany do opisu rozpraszania światła na pojedynczej cząstce fazy zdyspergowanej zaprezentowano na rysunku 3.2.1.1.



Rys. 3.2.1.1. Sferyczny układ współrzędnych (r, θ, ϕ) stosowany do opisu rozpraszania światła na pojedynczej cząstce [7]

Fig. 3.2.1.1. Spherical coordinate system (r, θ, ϕ) used for describing light scattering by single particle [7]

Fala padająca rozchodzi się w dodatnim kierunku osi z . Kierunek obserwacji fali rozproszonej przez cząstkę, zwany kierunkiem rozpraszania, określa wektor \mathbf{r} . Płaszczyzna wyznaczona przez kierunek rozpraszania i oś z nazywa się płaszczyzną rozpraszania. Kąt ϕ zwany jest kątem azymutalnym, natomiast kąt θ – kątem rozpraszania [7].

Przyjęto, że rozważana cząstka kulista ma promień a , materiał, z którego cząstka jest wykonana, charakteryzuje się zespolonym współczynnikiem załamania światła $N_1 = n_1 + ik_1$, gdzie i oznacza jednostkę urojoną, $i = \sqrt{-1}$, natomiast ośrodek dyspersyjny, w którym cząstka jest umieszczona, charakteryzuje się zespolonym współczynnikiem załamania światła $N_0 = n_0 + ik_0$.

Liczba falowa światła padającego i rozpraszanego przez cząstkę jest dana wzorem [7]

$$k = \frac{2\pi N_0}{\lambda_0}, \quad (3.2.1.1)$$

gdzie λ_0 – długość fali światła padającego i rozpraszanego w próżni.

Względny współczynnik załamania cząstki względem ośrodka wyraża się wzorem [7]

$$m = \frac{N_1}{N_0}. \quad (3.2.1.2)$$

Parametr wielkościowy cząstki, zwany inaczej parametrem Mie, zdefiniowany jest wzorem [7]

$$x = ka = \frac{2\pi a N_0}{\lambda_0}. \quad (3.2.1.3)$$

Właściwości optyczne jednorodnej i izotropowej cząstki kulistej umieszczonej w jednorodnym i izotropowym ośrodku dyspersyjnym, są w pełni zdeterminowane przez dwie wielkości: m i x [5, 7].

Jeśli idzie o nefelometryczną technikę analizy wielkości cząstek układów dyspersyjnych, istotne jest natężenie promieniowania rozpraszanego przez cząstkę w określonym kierunku rozpraszania \mathbf{r} wyznaczonym przez kąty θ i ϕ . Natężenie to zdefiniowane jest jako stosunek strumienia energii (czyli ilości energii przepływającej w jednostce czasu) promieniowania rozpraszanego przez cząstkę w mały kąt bryłowy wokół kierunku $\mathbf{r} - d\phi_E(\mathbf{r})$ do wielkości tego kąta bryłowego $d\Omega(\mathbf{r})$:

$$\mathfrak{I}_s(\mathbf{r}) = \frac{d\phi_E(\mathbf{r})}{d\Omega(\mathbf{r})}. \quad (3.2.1.4)$$

Szczegółowe rozważania prowadzą do wniosku, że w przyjętych założeniach upraszczających natężenie promieniowania rozpraszanego przez cząstkę w określonym kierunku spośród parametrów określających ten kierunek zależne jest wyłącznie od kąta rozpraszania θ i dane jest zależnością [7]

$$\mathfrak{I}_s(\theta, x, m) = \frac{1}{k^2} S_{11}(\theta, x, m) I_0, \quad (3.2.1.5)$$

gdzie: $S_{11}(\theta, x, m)$ – element tzw. macierzy Muellera, I_0 – energetyczne natężenie oświetlenia cząstki przez falę padającą, zdefiniowane jako stosunek strumienia energii promieniowania padającego na cząstkę do pola powierzchni rzutu tej cząstki na płaszczyznę prostopadłą do kierunku propagacji fali padającej [7].

Jeśli idzie o turbidymetryczną technikę analizy wielkości cząstek układów dyspersyjnych, istotny jest przekrój czynny cząstki na osłabienie (ekstynkcję) światła. Osłabienie oznacza ubytek strumienia energii promieniowania padającego, będący łącznym efektem rozpraszania promieniowania padającego w kierunkach innych niż kierunek promieniowania padającego oraz absorpcji promieniowania padającego przez cząstkę. Przekrój czynny cząstki na osłabienie określa się jako stosunek ubytku strumienia energii promieniowania padającego w wyniku osłabienia (ekstynkcji) $\phi_{E,ext}$ do energetycznego natężenia oświetlenia cząstki przez falę padającą I_0 [7]

$$C_{ext} = \frac{\phi_{E,ext}}{I_0}. \quad (3.2.1.6)$$

Wielkość C_{ext} zależy od x i m [7].

Postacie zależności $S_{11}(\theta, x, m)$ oraz $C_{ext}(x, m)$ wynikają ze szczegółowych teorii rozpraszania światła na pojedynczej cząstce. W badaniach prowadzonych przez autorów stosowane były dwie teorie tego typu:

- teoria rozpraszania Mie,
- teoria dyfrakcji Fraunhofera.

3.2.1.1. TEORIA ROZPRASZANIA MIE

Teoria Mie jest teorią rozpraszania światła na pojedynczej jednorodnej i izotropowej cząstce kulistej o dowolnej wielkości umieszczonej w jednorodnym i izotropowym ośrodku dyspersyjnym [7]. Wyznaczone przez teorię zależności są wynikiem ścisłego rozwiązania zagadnienia brzegowego elektromagnetyzmu [5, 7].

Rozkłady przestrzenne pola elektrycznego i magnetycznego fali rozproszonej na cząstce wyrażają się nieskończonymi szeregami funkcji specjalnych. W pomiarach nefelometrycznych zależność $S_{11}(\theta, x, m)$ zgodnie z teorią Mie [7, 8] określają następujące zależności:

$$S_{11}(\theta, x, m) = \frac{1}{2} \left[|S_1(\theta, x, m)|^2 + |S_2(\theta, x, m)|^2 \right], \quad (3.2.1.1.1)$$

gdzie [7, 8]:

$$S_1(\theta, x, m) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} (a_n \pi_n + b_n \tau_n), \quad (3.2.1.1.2)$$

$$S_2(\theta, x, m) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} (a_n \tau_n + b_n \pi_n). \quad (3.2.1.1.3)$$

Równania (3.2.1.1.2) i (3.2.1.1.3) stanowią rozwinięcia funkcji odpowiednio $S_1(\theta, x, m)$ i $S_2(\theta, x, m)$ w nieskończone szeregi funkcji specjalnych π_n i τ_n , przy czym wielkości a_n i b_n są współczynnikami tych rozwinięć.

Funkcje specjalne π_n i τ_n zależne są wyłącznie od kąta rozpraszania θ i zdefiniowane są przez zależności [7, 8]

$$\pi_n(\cos\theta) = \frac{P_n^1(\cos\theta)}{\sin\theta}, \quad (3.2.1.1.4)$$

$$\tau_n(\cos\theta) = \frac{dP_n^1(\cos\theta)}{d\theta} = -\sin\theta \frac{dP_n^1(\cos\theta)}{d(\cos\theta)}, \quad (3.2.1.1.5)$$

gdzie $P_n^1(\cos\theta)$ – stowarzyszona funkcja Legendre’a pierwszego rodzaju stopnia n pierwszego rzędu.

Współczynniki a_n i b_n , zwane współczynnikami rozpraszania, zależne są wyłącznie od parametru wielkościowego cząstki x oraz jej względnego współczynnika załamania m i określone są przez wzory [7]:

$$a_n = \frac{m\psi_n(mx)\psi_n'(x) - \psi_n(x)\psi_n'(mx)}{m\psi_n(mx)\xi_n'(x) - \xi_n(x)\psi_n'(mx)}, \quad (3.2.1.1.6)$$

$$b_n = \frac{\psi_n(mx)\psi_n'(x) - m\psi_n(x)\psi_n'(mx)}{\psi_n(mx)\xi_n'(x) - m\xi_n(x)\psi_n'(mx)}, \quad (3.2.1.1.7)$$

w których funkcje $\psi_n(\rho)$ oraz $\xi_n(\rho)$ oznaczają funkcje Riccatiego–Bessela. Funkcje te zdefiniowane są przez wyrażenia [7]:

$$\psi_n(\rho) = \rho j_n(\rho), \quad (3.2.1.1.8)$$

$$\xi_n(\rho) = \rho h_n^{(1)}(\rho), \quad (3.2.1.1.9)$$

gdzie: $j_n(\rho)$ – sferyczna funkcja Bessela pierwszego rodzaju rzędu n , $h_n^{(1)}(\rho)$ – sferyczna funkcja Hankela pierwszego rodzaju rzędu n (zwana też sferyczną funkcją Bessela trzeciego rodzaju rzędu n).

Istotna ze względu na pomiary turbidymetryczne zależność $C_{\text{ext}}(x, m)$ wyznaczona przez teorię Mie również ma postać szeregu nieskończonego [7]:

$$C_{\text{ext}}(x, m) = \frac{2\pi}{k^2} \operatorname{Re} \left\{ \sum_{n=1}^{\infty} (2n+1)(a_n + b_n) \right\}. \quad (3.2.1.1.10)$$

W badaniach prowadzonych przez autorów obliczenia numeryczne funkcji $S_{11}(\theta, x, m)$, $C_{\text{ext}}(x, m)$ oraz wszystkich wielkości pośrednich realizowane były według algorytmu opisanego przez Bohrena i Huffmana [7]. Algorytm ten wykorzystuje proste zależności rekurencyjne pomiędzy funkcjami specjalnymi kolejnych rzędów.

3.2.1.2. TEORIA DYFRAKCJI FRAUNHOFERA

Duża złożoność omawianej wcześniej teorii Mie stanowi poważne utrudnienie w jej zastosowaniu do analizy wielkości cząstek układów dyspersyjnych. Powoduje to, że teoria Mie często zastępowana jest przybliżonymi modelami zjawiska rozpraszania światła, w których przyjmowane są określone założenia upraszczające. Założenia te ograniczają stosowalność teorii przybliżonych do pewnych szczególnych przypadków spełniających stawiane postulaty [5, 8].

W badaniach prowadzonych przez autorów w charakterze przybliżonego modelu rozpraszania światła zastosowano teorię dyfrakcji Fraunhofera.

W teorii dyfrakcji Fraunhofera przyjmuje się następujące założenia upraszczające [8, 9]:

- cząstka rozpraszająca jest całkowicie nieprzezroczysta,
- cząstka rozpraszająca ma wymiar znacznie większy od długości fali światła.

Postulaty te można wyrazić formalnie jako [8]

$$|m-1| \gg 1 \quad \text{oraz} \quad x \gg 1. \quad (3.2.1.2.1)$$

Model rozproszenia Fraunhofera światła na pojedynczej cząstce kulistej o promieniu a opisuje wyłącznie efekt dyfrakcji na nieprzezroczystym dysku o promieniu a stanowiącym rzut tej cząstki na płaszczyznę prostopadłą do kierunku propagacji światła padającego. Zjawiska odbicia i załamania są pomijane [5, 9]. W związku z tym wyznaczone przez teorię Fraunhofera charakterystyki rozpraszania światła zależą tylko od parametru wielkościowego cząstki x , są natomiast niezależne od względnego współczynnika załamania cząstki względem otaczającego ją ośrodka m [5, 8, 9].

Teoria dyfrakcji Fraunhofera opiera się na zasadzie Huygensa–Fresnela, która stwierdza, że każdy punkt, do którego dotarło czoło fali padającej, jest źródłem wtórnej fali kulistej. Interferencja fal wtórnych daje w rezultacie czoło fali propagującej w rozważanym ośrodku w chwili późniejszej [9]. Jeżeli część czoła fali padającej zostanie wycięta przez nieprzezroczystą cząstkę, efektem interferencji fal wtórnych pochodzących od wszystkich punktów leżących na pozostałej części czoła fali jest zmiana dotychczasowego kształtu czoła fali, co wiąże się ze zmianą kierunku propagacji fali, zwaną ugięciem lub dyfrakcją [9].

Teoria Fraunhofera jest skalarną teorią dyfrakcji, co oznacza, że odnosi się do fal stanowiących zaburzenia skalarnych wielkości fizycznych. Fale świetlne stanowią zaburzenia pola elektrycznego i magnetycznego, czyli pól wektorowych. Zastosowanie teorii do przybliżonego opisu rozpraszania światła jest możliwe wyłącznie za sprawą założenia, że światło padające jest niespolaryzowane [8].

Istotną ze względu na pomiary nefelometryczne zależność $S_{11}(\theta, x)$ w teorii dyfrakcji Fraunhofera określa następujący wzór [7]

$$S_{11}(\theta, x) = \left| x \frac{(1 + \cos \theta)}{2} \frac{J_1(x \sin \theta)}{\sin \theta} \right|^2, \quad (3.2.1.2.2)$$

gdzie $J_1(\rho)$ – funkcja Bessela pierwszego rodzaju rzędu pierwszego. Dla małego kąta rozpraszania zastosować można przybliżenie $\sin \theta \approx \theta$ oraz $\cos \theta \approx 1$, co prowadzi do przybliżonego wzoru

$$S_{11}(\theta, x) = \left| x^2 \frac{J_1(x\theta)}{x\theta} \right|^2. \quad (3.2.1.2.3)$$

Istotna ze względu na pomiary turbidymetryczne wielkość C_{ext} , wyznaczona przez teorię dyfrakcji Fraunhofera, zależy wyłącznie od promienia cząstki a i dana jest wyrażeniem [7]

$$C_{\text{ext}} = 2\pi a^2. \quad (3.2.1.2.4)$$

Najważniejsze zalety teorii dyfrakcji Fraunhofera w stosunku do teorii rozpraszania Mie stanowią jej prostota oraz brak zależności wyznaczonych przez teorię charakterystyk rozpraszania światła od współczynnika załamania cząstki względem otaczającego ją ośrodka m . Empiryczne wartości współczynnika m charakteryzują się zwykle dość znaczną niepewnością [5].

3.2.2. MODELOWANIE ROZPRASZANIA ŚWIATŁA NA FAZIE ZDYSPERGOWANEJ

Modelowanie rozpraszania światła na fazie zdyspergowanej, stanowiącej zbiór złożony z bardzo wielkiej liczby cząstek, opiera się na teorii przenoszenia energii promienistej stworzonej przez Chandrasekhara i Preisendorfera [5].

Zgodnie z przyjętymi w punkcie 3.2 założeniami upraszczającymi rozmieszczenie cząstek fazy zdyspergowanej w ośrodku dyspersyjnym jest całkowicie losowe, ale średnia liczba cząstek, przypadająca na jednostkę objętości N_v zwana koncentracją cząstek, jest jednakowa w całym układzie. Podobnie, jednakowa w całym układzie jest przypadająca na jednostkę objętości średnia liczba cząstek o promieniach leżących w nieskończenie małym przedziale $(a, a + da)$, która, zgodnie ze wzorem (2.2.1), równa jest $N_v f(a) da$. Wynika stąd, że każda elementarna objętość układu dysper-

syjnego dV rozprasza światło w identyczny sposób. W związku z tym rozpraszanie światła w układzie dyspersyjnym scharakteryzować można za pomocą wielkości opisujących rozpraszanie światła przez dowolnie wybraną elementarną objętość dV .

W nefelometrycznej metodzie pomiaru wielkości cząstek układu dyspersyjnego, natężenie promieniowania rozproszonego przez elementarną objętość dV zależy wyłącznie od kąta rozpraszania θ . Wobec tego objętościowa funkcja rozpraszania zależna jest wyłącznie od kąta rozpraszania. Jest ona zdefiniowana następującym wzorem [5]:

$$\beta(\theta) = \frac{d\mathfrak{I}_s(\theta)}{I_0 dV}, \quad (3.2.2.1)$$

gdzie: dV – elementarna objętość układu, $d\mathfrak{I}_s(\theta)$ – natężenie promieniowania rozproszonego pod kątem θ przez elementarną objętość układu dV , I_0 – energetyczne natężenie oświetlenia elementarnej objętości układu przez falę padającą.

Dla ustalonych długości fali światła rozpraszanego λ oraz względnego współczynnika załamania cząstki względem ośrodka m , traktowanych jako parametry (3.2.1.5) określające natężenie światła rozproszonego pod kątem θ przez pojedynczą cząstkę, przedstawić można w postaci uwidaczniającej zależność od promienia cząstki:

$$\mathfrak{I}_{s,\text{poj.cz}}(\theta, a) = \frac{1}{k^2} S_{11}(\theta, a) I_0. \quad (3.2.2.2)$$

Przyjęte założenia upraszczające o rozpraszaniu jednokrotnym i niekoherentnym umożliwiają obliczenie natężenia światła rozproszonego w danym kierunku przez elementarną objętość dV układu dyspersyjnego jako sumy natężeń światła rozproszonego w tym kierunku przez poszczególne cząstki fazy zdyspergowanej w objętości dV traktowane niezależnie. W związku z tym natężenie światła rozproszonego pod kątem θ przez wszystkie cząstki, o promieniach leżących w infinytezymalnie małym przedziale $(a, a + da)$ zawarte w objętości dV , uzyskuje się przez pomnożenie natężenia $\mathfrak{I}_{s,\text{poj.cz}}(\theta, a)$ danego wzorem (3.2.2.2) przez liczbę cząstek $N_v f(a) da dV$:

$$d\mathfrak{I}_s(\theta, a) = \frac{1}{k^2} S_{11}(\theta, a) I_0 N_v f(a) da dV. \quad (3.2.2.3)$$

Całkowite natężenie promieniowania rozproszonego pod kątem θ przez elementarną objętość układu dV – $d\mathfrak{I}_s(\theta)$ obliczyć można przez scałkowanie danych wzorem (3.2.2.3) natężeń $d\mathfrak{I}_s(\theta, a)$ pochodzących od cząstek o różnych promieniach:

$$d\mathfrak{I}_s(\theta) = I_0 dV \int_0^{\infty} N_v \frac{1}{k^2} S_{11}(\theta, a) f(a) da. \quad (3.2.2.4)$$

Podstawiając wyrażenie (3.2.2.4) do wzoru (3.2.2.1), uzyskuje się:

$$\beta(\theta) = \int_0^{\infty} N_v \frac{1}{k^2} S_{11}(\theta, a) f(a) da . \quad (3.2.2.5)$$

W zależności od teorii rozpraszania światła na pojedynczej cząstce fazy zdyspergowanej, zastosowanej w modelowaniu matematycznym rozpraszania światła na fazie zdyspergowanej, występująca w równaniu (3.2.2.5) funkcja $S_{11}(\theta, a)$ określona jest wzorami:

- (3.2.1.1)–(3.2.1.3) oraz (3.2.1.1.1)–(3.2.1.1.9) w przypadku teorii rozpraszania Mie,
- (3.2.1.1), (3.2.1.3) oraz (3.2.1.2.2) lub przybliżonym wzorem (3.2.1.2.3) w przypadku teorii dyfrakcji Fraunhofera.

Ze względu na turbidymetryczną technikę analizy wielkości cząstek układów dyspersyjnych istotny jest całkowity współczynnik osłabienia (ekstynkcji) światła, który definiuje się jako stosunek całkowitego ubytku strumienia energii promieniowania padającego na elementarną objętość dV układu dyspersyjnego w wyniku osłabienia (ekstynkcji) $\phi_{E, \text{ext}, dV}$ do energetycznego natężenia oświetlenia elementarnej objętości układu przez falę padającą I_0 oraz do elementarnej objętości układu dV [5, 8]:

$$c = \frac{\phi_{E, \text{ext}, dV}}{I_0 dV}, \quad (3.2.2.6)$$

Jeżeli cylindryczną wiązkę światła o polu przekroju poprzecznego S biegnącego przez układ dyspersyjny w określonym kierunku, osłoniłoby się od promieniowania rozproszonego dochodzącego z innych kierunków oraz założyłoby się, że w układzie dyspersyjnym nie występują pierwotne źródła światła, to na infinytezymalnym odcinku dz drogi światła zjawisko osłabienia spowodowałoby ubytek strumienia energii wiązki równy

$$\phi_{E, \text{ext}, dV} = SI(z) - SI(z + dz) = SI(z) - S[I(z) + dI(z)] = -SdI(z), \quad (3.2.2.7)$$

gdzie $I(z)$ oznacza energetyczne natężenie oświetlenia w wiązce światła po przebyciu drogi z w układzie dyspersyjnym. Na drodze dz wiązki światła znajduje się objętość:

$$dV = Sdz \quad (3.2.2.8)$$

układu dyspersyjnego. Podstawiając związki (3.2.2.7) i (3.2.2.8) oraz $I_0 = I(z)$ do wzoru (3.2.2.6), uzyskuje się

$$c = -\frac{1}{I(z)} \frac{dI(z)}{dz}, \quad (3.2.2.9)$$

co stanowi różniczkową postać prawa Bougera–Lamberta–Beera [5, 8].

Scałkowanie równania (3.2.2.9) prowadzi do zależności [5, 8]:

$$I(z) = I(0) \exp(-cz). \quad (3.2.2.10)$$

Według zależności (3.2.1.6) całkowity ubytek strumienia energii promieniowania o długości fali λ , padającego na pojedynczą cząstkę fazy zdyspergowanej, spowodowany osłabieniem (ekstynkcją) światła przez tę cząstkę wyrazić można następującym wzorem, w którym uwidoczniła jest zależność od długości fali światła rozpraszanego oraz promienia cząstki:

$$\phi_{E,\text{ext,poj.cz.}}(\lambda, a) = C_{\text{ext,poj.cz.}}(\lambda, a) I_0. \quad (3.2.2.11)$$

Przyjęte w punkcie 3.2 założenia upraszczające o rozpraszaniu jednokrotnym i niekoherentnym umożliwiają obliczenie spowodowanego osłabieniem całkowitego ubytku strumienia energii promieniowania padającego na elementarną objętość dV układu dyspersyjnego jako sumy ubytków spowodowanych przez poszczególne cząstki fazy zdyspergowanej w objętości dV traktowane niezależnie. W związku z tym ubytek strumienia energii promieniowania o długości fali λ , wywołany przez wszystkie cząstki o promieniach leżących w infimezymalnie małym przedziale $(a, a + da)$ zawarte w objętości dV , uzyskuje się przez pomnożenie ubytku strumienia $\phi_{E,\text{ext,poj.cz.}}(\lambda, a)$ według (3.2.2.11) przez liczbę cząstek $N_v f(a) da dV$

$$\phi_{E,\text{ext,dV}}(\lambda, a) = C_{\text{ext,poj.cz.}}(\lambda, a) I_0 N_v f(a) da dV. \quad (3.2.2.12)$$

Całkowity ubytek strumienia energii promieniowania o długości fali λ w elementarnej objętości układu dV obliczyć można przez scałkowanie danych wzorem (3.2.2.12) strumieni $\phi_{E,\text{ext,dV}}(\lambda, a)$, pochodzących od cząstek o różnych promieniach:

$$\phi_{E,\text{ext,dV}}(\lambda) = I_0 dV \int_0^{\infty} N_v C_{\text{ext,poj.cz.}}(\lambda, a) f(a) da. \quad (3.2.2.13)$$

Podstawiając wyrażenie (3.2.2.13) do wzoru (3.2.2.6), uzyskuje się:

$$c(\lambda) = \int_0^{\infty} N_v C_{\text{ext,poj.cz.}}(\lambda, a) f(a) da. \quad (3.2.2.14)$$

W zależności od teorii rozpraszania światła na pojedynczej cząstce fazy zdyspergowanej zastosowanej w modelowaniu matematycznym rozpraszania światła na fazie zdyspergowanej, występująca w równaniu (3.2.2.14) funkcja $C_{\text{ext,poj.cz.}}(\lambda, a)$ określona jest wzorami:

- (3.2.1.1)–(3.2.1.3) oraz (3.2.1.1.6)–(3.2.1.1.10) w przypadku teorii rozpraszania Mie,
- (3.2.1.2.4) w przypadku teorii dyfrakcji Fraunhofera.

Można zauważyć, że zależności (3.2.2.5) oraz (3.2.2.14) kluczowe ze względu na odpowiednio nefelometryczną i turbidymetryczną analizę wielkości cząstek układów dyspersyjnych mają postać równania całkowego Fredholma pierwszego rodzaju [5, 8]:

$$g(y) = \int_0^{\infty} K(y, a) f(a) da, \quad y_{\min} \leq y \leq y_{\max}, \quad (3.2.2.15)$$

gdzie: $g(y)$ – zależność stanowiąca charakterystykę rozpraszania światła w układzie dyspersyjnym będąca przedmiotem zainteresowania w określonej technice analizy wielkości cząstek: $\beta(\theta)$ w technice nefelometrycznej, $c(\lambda)$ w technice turbidymetrycznej, $K(y, a)$ – jądro równania całkowego o szczegółowej postaci zależnej od rozważanej charakterystyki rozpraszania światła w układzie dyspersyjnym oraz teorii rozpraszania światła na pojedynczej cząstce fazy zdyspergowanej, zastosowanej w modelowaniu matematycznym rozpraszania światła na fazie zdyspergowanej.

3.3. IDEA OPTYCZNYCH TECHNIK ANALIZY WIELKOŚCI CZĄSTEK UKŁADÓW DISPERSYJNYCH

Optyczne techniki analizy wielkości cząstek układów dyspersyjnych opierają się na bezpośrednim pomiarze określonych zależności $g(y)$, stanowiących charakterystyki rozpraszania światła w układzie dyspersyjnym, a następnie na wyznaczeniu szukanego rozkładu wielkości cząstek fazy zdyspergowanej układu $f(a)$ na podstawie wyników pomiarów z wykorzystaniem związku o ogólnej postaci (3.2.2.15).

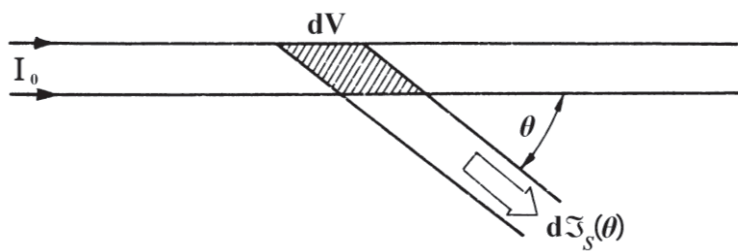
Przedmiotem rozważań w niniejszej pracy są dwie techniki pomiarowe:

- nefelometria,
- turbidymetria.

3.3.1. POMIARY NEFELOMETRYCZNE

W nefelometrycznej technice analizy wielkości cząstek układów dyspersyjnych bezpośredniemu pomiarowi poddawana jest zależność objętościowej funkcji rozpraszania od kąta rozpraszania $\beta(\theta)$ [5].

Zasada pomiaru funkcji $\beta(\theta)$ wynika z zależności (3.2.2.1) definiującej tę wielkość fizyczną. W celu określenia objętościowej funkcji rozpraszania β dla danego kąta rozpraszania θ dokonuje się pomiaru natężenia promieniowania rozproszonego pod kątem θ przez elementarną objętość układu $dV - d\mathfrak{V}_s(\theta)$ [5]. Zasadę pomiaru objętościowej funkcji rozpraszania $\beta(\theta)$ w nefelometrii zilustrowano na rysunku 3.3.1.1.



Rys. 3.3.1.1. Zasada pomiaru objętościowej funkcji rozpraszania $\beta(\theta)$ w nefelometrii [5]
 Fig. 3.3.1.1. Idea of volumetric scattering function $\beta(\theta)$ measurement in nephelometry [5]

Wyznaczenie rozkładu wielkości cząstek fazy zdyspergowanej $f(a)$, na podstawie zależności całkowitej wiążącej funkcję $\beta(\theta)$ z funkcją $f(a)$, wymaga znajomości przebiegu funkcji $\beta(\theta)$ w pewnym przedziale kąta rozpraszania θ . Dlatego w ramach pomiarów nefelometrycznych wyznaczane są wartości funkcji $\beta(\theta)$ dla wielu kątów θ .

Pomiar objętościowej funkcji rozpraszania omówioną metodą nastęrcza wielu problemów technicznych, z których najistotniejszymi są [5]:

- dokładne wyznaczenie kąta rozpraszania θ ,
- zmiana rozważanej objętości ośrodka dV przy zmianie kąta θ (rys. 3.3.1.1).

3.3.2. POMIARY TURBIDYMETRYCZNE

W turbidymetrycznej technice analizy wielkości cząstek układów dyspersyjnych bezpośredniemu pomiarowi poddawana jest zależność całkowitego współczynnika osłabienia (ekstynkcji) od długości fali rozpraszanego światła $c(\lambda)$ [5].

Zasada pomiaru całkowitego współczynnika osłabienia c wynika z zależności (3.2.2.9) lub (3.2.2.10) będących sformułowaniami prawa Bougera–Lamberta–Beera. W celu wyznaczenia $c(\lambda)$ mierzony jest spadek energetycznego natężenia oświetlenia w wiązce światła monochromatycznego o długości fali λ , po przebyciu przez nią określonego odcinka drogi w układzie dyspersyjnym. Konieczne jest przy tym spełnienie podanych w punkcie 3.2.2 warunków stosowalności zależności (3.2.2.9) i (3.2.2.10), tzn. [5]:

- osłonięcie wiązki światła od promieniowania rozproszonego docierającego z kierunków innych od kierunku propagacji wiązki, co wymusza zastosowanie wiązki ograniczonej cylindrycznie; warunek ten spełniony jest zawsze jedynie w przybliżeniu tym dokładniej, im mniejszy jest stosunek szerokości wiązki do długości jej drogi, czyli im mniejszy jest kąt bryłowy, z którego detektor rejestruje światło. Najkorzystniejsze jest w tym przypadku zastosowanie wiązki promieniowania laserowego ze względu na jej małą rozbieżność [5],

- zapewnienie, by w układzie dyspersyjnym nie występowały pierwotne źródła światła.

Spełnienie tych wymogów gwarantuje, że mierzony spadek energetycznego natężenia oświetlenia w wiązce światła po przebyciu przez nią określonego odcinka drogi w układzie dyspersyjnym spowodowany jest wyłącznie przez zjawisko osłabienia, co zapewnia poprawną wartość całkowitego współczynnika osłabienia wyznaczonego na podstawie mierzonego spadku.

Doświadczalne wyznaczenie współczynnika $c(\lambda)$, według zależności (3.2.2.9), polega na:

- pomiarze energetycznego natężenia oświetlenia $I(\lambda)$ w wiązce światła monochromatycznego o długości fali λ w określonym punkcie układu dyspersyjnego,
- pomiarze spadku energetycznego natężenia oświetlenia $\Delta I(\lambda)$ w wiązce światła po przebyciu przez nią możliwie jak najkrótszej drogi Δz w układzie dyspersyjnym,
- obliczeniu współczynnika $c(\lambda)$ na podstawie następującego wzoru, stanowiącego przybliżenie wzoru (3.2.2.9) przez zastąpienie różniczek różnicami skończonymi

$$c(\lambda) = -\frac{1}{I(\lambda)} \frac{\Delta I(\lambda)}{\Delta z}. \quad (3.3.2.1)$$

Doświadczalne wyznaczenie współczynnika $c(\lambda)$, na podstawie zależności (3.2.2.10), polega na:

- pomiarze energetycznego natężenia oświetlenia $I_0(\lambda)$ w wiązce światła monochromatycznego o długości fali λ w określonym punkcie układu dyspersyjnego,
- pomiarze energetycznego natężenia oświetlenia $I_z(\lambda)$ w wiązce światła po przebyciu przez nią drogi z w układzie dyspersyjnym,
- obliczeniu współczynnika $c(\lambda)$, na podstawie wzoru (3.2.2.10):

$$I_z(\lambda) = I_0(\lambda) \exp(-cz), \quad (3.3.2.2)$$

skąd wynika, że:

$$c(\lambda) = -\frac{1}{z} \ln \left[\frac{I_z(\lambda)}{I_0(\lambda)} \right]. \quad (3.3.2.3)$$

Wyznaczenie rozkładu wielkości cząstek fazy zdyspergowanej $f(a)$ według zależności całkowitej wiążącej funkcję $c(\lambda)$ z funkcją $f(a)$ wymaga znajomości przebiegu funkcji $c(\lambda)$ w pewnym przedziale długości fali światła rozpraszanego λ . Dlatego w ramach pomiarów turbidymetrycznych wyznaczane są wartości funkcji $c(\lambda)$ dla wielu długości fali λ .

Ze wzoru (3.2.1.2.4) wynika, że funkcja $C_{\text{ext,poj.cz.}}(\lambda, a)$ w teorii dyfrakcji Fraunhofera nie zależy od długości fali λ . Zgodnie ze wzorem (3.2.2.14) oznacza to, że całkowity współczynnik osłabienia c również nie zależy od długości fali λ . Ponieważ wyznaczenie rozkładu wielkości cząstek fazy zdyspergowanej $f(a)$ w turbidymetrii opiera się na badaniu zmienności współczynnika c w funkcji długości fali λ , wynika stąd, że

w tym celu nie może zostać wykorzystany model matematyczny pomiarów turbidymetrycznych oparty na teorii dyfrakcji Fraunhofera. W badaniach prowadzonych przez autorów stosowany był model matematyczny wynikający z teorii rozpraszania Mie.

4. ZAGADNIENIE ODWROTNE W ANALIZIE WIELKOŚCI CZĄSTEK UKŁADÓW DYSPERSYJNYCH Z WYKORZYSTANIEM TECHNIK OPTYCZNYCH

Zagadnienie odwrotne w analizie wielkości cząstek układów dyspersyjnych z wykorzystaniem technik optycznych polega na wnioskowaniu o rozkładzie wielkości cząstek fazy zdyspergowanej $f(a)$ na podstawie wyników pomiarów zależności $g(y)$ stanowiącej charakterystykę rozpraszania światła w układzie dyspersyjnym. Wnioskowanie to opiera się na wykorzystaniu związku matematycznego pomiędzy funkcją mierzoną $g(y)$ a szukaną funkcją rozkładu $f(a)$, który w przypadku obu rozważanych technik pomiarowych – nefelometrii oraz turbidymetrii wyraża się równaniem całkowym Fredholma pierwszego rodzaju o ogólnej postaci wyrażonej wzorem (3.2.2.15). Zależność wiążąca funkcję $g(y)$ z funkcją $f(a)$ stanowi model matematyczny pomiarów danego typu [5].

Matematycznie rozwiązanie rozpatrywanego problemu odwrotnego sprowadza się do rozwiązania równania (3.2.2.15) z funkcją jądra $K(y, a)$ właściwą dla rozważanej charakterystyki rozpraszania światła w układzie dyspersyjnym oraz teorią rozpraszania światła na pojedynczej cząstce fazy zdyspergowanej, zastosowanej w modelowaniu matematycznym rozpraszania światła na fazie zdyspergowanej. Niestety ze względu na złożoną postać jądra całkowego $K(y, a)$ rozwiązania analityczne znane są wyłącznie dla prostych przypadków szczególnych.

4.1. NIEWŁAŚCIWE POSTAWIENIE ZAGADNIENIA ODWROTNEGO

Olbrzymie trudności, których następcza rozwiązywanie rozważanego zagadnienia odwrotnego (3.2.2.15), są konsekwencją tzw. niewłaściwego postawienia (ang. *ill-posedness*). Oznacza to, że problem nie spełnia przynajmniej jednego z trzech warunków dobrego postawienia (ang. *well-posedness*) Hadamarda [1, 2]:

- rozwiązanie $f(a)$ jest dla każdej funkcji $g(y)$,
- rozwiązanie $f(a)$ jest jednoznaczne,
- rozwiązanie $f(a)$ jest stabilne, tzn. zależy w sposób ciągły od funkcji $g(y)$.

Warunek stabilności rozwiązania można również sformułować jako żądanie, by dla każdego dążącego do zera ciągu zaburzeń funkcji $g(y)$ odpowiadający mu ciąg zaburzeń rozwiązania $f(a)$ także dążył do zera [2].

Zakładając, że jądro całkowe $K(y, a)$ jest całkowne oraz funkcja $f(a)$ jest rozwiązaniem równania (3.2.2.15), czyli:

$$g(y) = \int_0^{\infty} K(y, a) f(a) da, \quad (4.1.1)$$

funkcja $f(a) + \delta f_n(a)$, gdzie $\delta f_n(a) = A \sin na$, A – stała, spełnia równość [2, 4]:

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_0^{\infty} K(y, a) [f(a) + \delta f_n(a)] da &= \lim_{n \rightarrow \infty} \int_0^{\infty} K(y, a) [f(a) + A \sin(na)] da \\ &= g(y) + \lim_{n \rightarrow \infty} \int_0^{\infty} K(y, a) A \sin(na) da. \end{aligned} \quad (4.1.2)$$

Oznaczając:

$$\delta g_n(y) = \int_0^{\infty} K(y, a) A \sin(na) da, \quad (4.1.3)$$

oraz uwzględniając związek [2, 4]:

$$\lim_{n \rightarrow \infty} \int_0^{\infty} K(y, a) \sin(na) da = 0, \quad (4.1.4)$$

uzyskuje się:

$$\lim_{n \rightarrow \infty} \delta g_n(y) = 0. \quad (4.1.5)$$

Oznacza to, że ciąg zaburzeń $\delta g_n(y)$ funkcji $g(y)$ taki, że $\delta g_n(y) \rightarrow 0$ dla $n \rightarrow \infty$ powoduje ciąg zaburzeń $\delta f_n(a)$ rozwiązania $f(a)$ problemu nie dążący do zera dla $n \rightarrow \infty$. Wykazuje to, że rozwiązanie problemu odwrotnego (3.2.2.15) jest niestabilne, a sam problem źle postawiony (ang. *ill-posed*) [2, 4].

4.2. SŁABE UWARUNKOWANIE ZAGADNIENIA ODWROTNEGO

Z niewłaściwym postawieniem rozważanego zagadnienia odwrotnego (3.2.2.15) wiąże się jego słabe uwarunkowanie (ang. *ill-conditioning*) polegające na tym, że niewielkie nawet zakłócenia danych pomiarowych $g(y)$ powodują niezwykle duże zaburzenia rozwiązania $f(a)$ [1, 2, 3, 4, 5].

Wgląd w przyczyny słabego uwarunkowania zagadnienia odwrotnego daje jego sformułowanie w kategoriach pojęć analizy funkcjonalnej. Przekształcenie całkowe wyrażone równaniem (3.2.2.15) zapisać można w postaci równania operatorowego [1]

$$g(y) = (Kf)(y) = \int_0^{\infty} K(y, a) f(a) da, \quad y_{\min} \leq y \leq y_{\max}, \quad (4.2.1)$$

gdzie K – operator przekształcający funkcję $f(a)$ w funkcję $g(y)$.

Niech funkcja $f(a)$ należy do przestrzeni Hilberta H_1 , którą stanowi przestrzeń $L_2[0, \infty)$ wszystkich funkcji całkowalnych z kwadratem w przedziale $[0, \infty)$, z iloczynem skalarnym określonym wzorem

$$\langle t, u \rangle_{H_1} = \int_0^{\infty} t(a) u(a) da. \quad (4.2.2)$$

Niech funkcja $g(y)$ należy do przestrzeni Hilberta H_2 , którą stanowi przestrzeń $L_2[y_{\min}, y_{\max}]$ wszystkich funkcji całkowalnych z kwadratem w przedziale $[y_{\min}, y_{\max}]$, z iloczynem skalarnym określonym wzorem

$$\langle t, u \rangle_{H_2} = \int_{y_{\min}}^{y_{\max}} t(y) u(y) dy. \quad (4.2.3)$$

Niech dodatkowo jądro $K(y, a)$ jest całkowalne z kwadratem, tzn.

$$\int_{y_{\min}}^{y_{\max}} dy \int_0^{\infty} K(y, a) da < \infty \quad (4.2.4)$$

oraz operator $K : H_1 \rightarrow H_2$ jest operatorem zwartym [1]. Wówczas są funkcje $u_n(y)$ oraz $v_n(a)$, zwane funkcjami osobliwymi, i rzeczywiste liczby dodatnie, zwane wartościami osobliwymi σ_n takie, że $\sigma_1 \geq \sigma_2 \geq \dots > 0$, $\sigma_n \rightarrow 0$, gdy $n \rightarrow \infty$ oraz [1, 12]:

$$(Kv_n)(y) = \int_0^{\infty} K(y, a) v_n(a) da = \sigma_n u_n(y), \quad (4.2.5)$$

$$(K^* u_n)(a) = \int_{y_{\min}}^{y_{\max}} K(y, a) u_n(y) dy = \sigma_n v_n(a), \quad (4.2.6)$$

gdzie K^* – operator sprzężony w stosunku do operatora K . Funkcje $v_n(a)$ tworzą bazę ortonormalną w dziedzinie operatora K – przestrzeni H_1 , natomiast funkcje $u_n(y)$ tworzą bazę ortonormalną w obrazie operatora K – przestrzeni H_2 . W związku z tym funkcję $f(a)$ z przestrzeni H_1 oraz funkcję $g(y)$ z przestrzeni H_2 przedstawić można odpowiednio w postaci następujących rozwinięć:

$$f(a) = \sum_{n=1}^{\infty} f_n v_n(a), \quad (4.2.7)$$

$$g(y) = \sum_{n=1}^{\infty} g_n u_n(y), \quad (4.2.8)$$

w których współczynniki, zwane współczynnikami Fouriera, określone są przez wzory:

$$f_n = \langle f, v_n \rangle_{H_1} = \int_0^{\infty} f(a) v_n(a) da, \quad (4.2.9)$$

$$g_n = \langle g, u_n \rangle_{H_2} = \int_{y_{\min}}^{y_{\max}} g(y) u_n(y) dy. \quad (4.2.10)$$

Podstawiając wyrażenia (4.2.7) i (4.2.8) do równania (4.2.1), oraz stosując związek (4.2.5), uzyskuje się:

$$g_n = \sigma_n f_n. \quad (4.2.11)$$

Wynika stąd, że [1, 12]:

$$g(y) = (Kf)(y) = \int_0^{\infty} K(y, a) f(a) da = \sum_{n=1}^{\infty} \sigma_n f_n v_n(y) \quad (4.2.12)$$

oraz

$$f(a) = \sum_{n=1}^{\infty} \frac{g_n}{\sigma_n} v_n(a) = \sum_{n=1}^{\infty} \frac{\langle g(y), u_n(y) \rangle_{H_2}}{\sigma_n} v_n(a). \quad (4.2.13)$$

Rozwinięcie (4.2.13), zwane szeregiem Picarda, określa rozwiązanie problemu odwrotnego sformułowanego za pomocą równania (4.2.1). Warunkiem istnienia tego rozwiązania jest zbieżność szeregu, co ściśle wyrazić można jako [1]:

$$\sum_{n=1}^{\infty} \left| \frac{\langle g(y), u_n(y) \rangle_{H_2}}{\sigma_n} \right|^2 < \infty. \quad (4.2.14)$$

Warunek (4.2.14) określa się jako kryterium Picarda [1].

W przypadku rozważanego zagadnienia odwrotnego jądro $K(y, a)$ ma charakter funkcji „gładkiej”, tzn. zmieniającej się stosunkowo wolno wraz ze zmienną a . W związku z tym jakościowo operator całkowy K powoduje „wygładzenie” funkcji $f(a)$, tzn. ma właściwość filtrowania z niej wielkoczęstotliwościowych oscylacji oraz gwałtownych zmian wartości [4]. W sposób bardziej ścisły wyrazić to można przez dwie ogólne właściwości, często obserwowane w przypadku wygładzających operatorów całkowych [4]:

• wartości osobliwe operatora σ_n bardzo szybko maleją ze wzrostem indeksu n , zbliżając się do zera,

• funkcje osobliwe $u_n(y)$ oraz $v_n(a)$ wraz ze wzrostem indeksu n prezentują coraz bardziej oscylacyjny przebieg, tzn. wykazują coraz więcej zmian znaków.

Jakościowo rozwinięcia fourierowskie (4.2.7) oraz (4.2.8) odpowiednio funkcji $f(a)$ i $g(y)$ stanowią ich dekompozycje na ortogonalne składowe oscylacyjne o rosnących częstotliwościach. Zgodnie ze wzorem (4.2.11), operator całkowy K traktować można jako filtr, który tłumi wielkoczęstotliwościowe oscylacje funkcji $f(a)$, odpowiadające współczynnikom fourierowskim f_n o dużych indeksach n , przez ich wymnożenie przez bardzo małe wartości osobliwe σ_n , co powoduje, że w wynikowej funkcji $g(y)$ istotne są jedynie wolnozmiennie składowe o małych indeksach g_n . Innymi słowy dane pomiarowe są małowzmacnione na wielkoczęstotliwościowe oscylacje funkcji $f(a)$, nawet gdy mają one znaczną amplitudę [3, 4]. Z kolei rozwiązywanie zagadnienia odwrotnego działa w przeciwną stronę – wiąże się ze znacznym wzmocnieniem wielkoczęstotliwościowych oscylacji funkcji $g(y)$, odpowiadających współczynnikom fourierowskim g_n o dużych indeksach n , przez ich dzielenie przez bardzo małe wartości osobliwe σ_n , co powoduje, że osiągają one duże amplitudy w rozwiązaniu $f(a)$ [3, 4].

Dotychczasowa analiza nie uwzględniała tego, że funkcja mierzona podlega zakłóceniom przez losowe błędy pomiarowe oraz błędy zaokrągleń. W rezultacie uzyskana w wyniku pomiarów funkcja $\tilde{g}(y)$ różni się od funkcji $g(y)$, spełniającej równanie (4.2.1) wyrażające model matematyczny procesu pomiarowego. Zakładając model addytywnych błędów losowych, wpływ zakłóceń z ich strony uwzględnić można przez dodanie członu zakłóceń $e(y)$ do wyjścia idealnego modelu $g(y)$ w równaniu (4.2.1):

$$\tilde{g}(y) = g(y) + e(y) = (Kf)(y) + e(y) = \int_0^{\infty} K(y, a) f(a) da + e(y). \quad (4.2.15)$$

Funkcja $\hat{f}(a)$ wyznaczona przez rozwiązanie równania (4.2.1) dla funkcji $\tilde{g}(y)$ różni się od rzeczywistej funkcji $f(a)$ i stanowi jedynie pewne jej przybliżenie. Zastępując w wyrażeniu (4.2.13) funkcję $g(y)$ przez funkcję $\tilde{g}(y)$ oraz uwzględniając związek (4.2.15), uzyskuje się [1]:

$$\begin{aligned} \hat{f}(a) &= \sum_{n=1}^{\infty} \frac{\langle \tilde{g}(y), u_n(y) \rangle_{H_2}}{\sigma_n} v_n(a) = \sum_{n=1}^{\infty} \frac{\langle g(y) + e(y), u_n(y) \rangle_{H_2}}{\sigma_n} v_n(a) \\ &= \sum_{n=1}^{\infty} \frac{\langle g(y), u_n(y) \rangle_{H_2}}{\sigma_n} v_n(a) + \sum_{n=1}^{\infty} \frac{\langle e(y), u_n(y) \rangle_{H_2}}{\sigma_n} v_n(a). \end{aligned} \quad (4.2.16)$$

Wynika stąd, że błąd rozwiązania $\hat{f}(a)$ w stosunku do prawdziwego rozwiązania $f(a)$ stanowi składnik:

$$r(y) = \sum_{n=1}^{\infty} \frac{\langle e(y), u_n(y) \rangle_{H_2}}{\sigma_n} v_n(a) = \sum_{n=1}^{\infty} \frac{e_n}{\sigma_n} v_n(a) = \sum_{n=1}^{\infty} r_n v_n(a), \quad (4.2.17)$$

gdzie

$$e_n = \langle e, u_n \rangle_{H_2} = \int_{y_{\min}}^{y_{\max}} e(y) u_n(y) dy, \quad (4.2.18)$$

$$r_n = \frac{e_n}{\sigma_n} = \langle r, v_n \rangle_{H_1} = \int_0^{\infty} r(a) v_n(a) da. \quad (4.2.19)$$

Wyrażenie (4.2.16) zapisać można więc w postaci

$$\hat{f}(a) = \sum_{n=1}^{\infty} \frac{\langle \tilde{g}(y), u_n(y) \rangle_{H_2}}{\sigma_n} v_n(a) = f(a) + \sum_{n=1}^{\infty} \frac{\langle e(y), u_n(y) \rangle_{H_2}}{\sigma_n} v_n(a). \quad (4.2.20)$$

Wprowadzając oznaczenie

$$\tilde{g}_n = \langle \tilde{g}, u_n \rangle_{H_2} = \int_{y_{\min}}^{y_{\max}} \tilde{g}(y) u_n(y) dy, \quad (4.2.21)$$

wzorowi (4.2.16) nadać można postać

$$\begin{aligned} \hat{f}(a) &= \sum_{n=1}^{\infty} \frac{\tilde{g}_n}{\sigma_n} v_n(a) \\ &= \sum_{n=1}^{\infty} \frac{g_n}{\sigma_n} v_n(a) + \sum_{n=1}^{\infty} \frac{e_n}{\sigma_n} v_n(a) = \sum_{n=1}^{\infty} \frac{g_n + e_n}{\sigma_n} v_n(a) \\ &= \sum_{n=1}^{\infty} \hat{f}_n v_n(a), \end{aligned} \quad (4.2.22)$$

gdzie

$$\hat{f}_n = \frac{\tilde{g}_n}{\sigma_n} = \frac{g_n + e_n}{\sigma_n} = f_n + \frac{e_n}{\sigma_n} = f_n + r_n. \quad (4.2.23)$$

Jak wcześniej wykazano rozwinięcia fourierowskie funkcji $f(a)$ i $r(a)$ oraz $g(y)$ i $e(y)$ na odpowiednie funkcje osobliwe operatora K stanowią ich dekompozycje na ortogonalne składowe oscylacyjne o rosnących częstotliwościach. Szum reprezento-

wany przez funkcję $e(y)$ charakteryzuje się przebiegiem szybkozmiennym, co powoduje, że współczynniki e_n jego rozwinięcia fourierowskiego (4.2.18) przyjmują istotne wartości nawet dla znacznych n , bardzo wolno maleją one ze wzrostem n , a często pozostają stałe lub nawet rosną przy zwiększającym się n . Jednocześnie wolna od zakłóceń funkcja $g(y)$ jest funkcją gładką, co sprawia, że współczynniki g_n jej rozwinięcia fourierowskiego (4.2.8) szybko maleją ze wzrostem n . W związku z tym głównie wielokrotnościowe oscylacje szumu $e(y)$ zostają silnie wzmocnione w procesie rozwiązywania zagadnienia odwrotnego za sprawą dzielenia we wzorze (4.2.19) współczynników e_n przez znikomo małe wartości osobliwe σ_n odpowiadające dużym indeksom n . Powoduje to, że jedynie dla małych n wartości współczynników \hat{f}_n , określone przez wyrażenie (4.2.23), są zbliżone do wartości f_n odpowiadających prawdziwemu rozwiązaniu. Dla dużych n wartości współczynników \hat{f}_n w coraz większym stopniu zdominowane są przez wielokrotnościowe składowe błędy r_n , ponieważ współczynniki fourierowskie f_n prawdziwego rozwiązania $f(a)$ szybko maleją ze wzrostem n , gdyż rozwiązanie to jest funkcją gładką. Zjawisko to prowadzi ostatecznie do olbrzymiego wzrostu wielkości błędu rozwiązania $\|r(a)\|$ dla niewielkiej nawet wielkości błędów pomiarowych $\|e(y)\|$, co oznacza słabe uwarunkowanie [1].

Reasumując, przyczyną złego uwarunkowania jest wygładzające działanie operatora całkowego K przejawiające się bardzo szybkim zmniejszaniem się jego wartości osobliwych σ_n przy wzroście n . Powoduje to, że w rozwiązaniu $\hat{f}(a)$ największemu wzmocnieniu podlegają wielokrotnościowe składowe szumu $e(y)$ [1, 3, 4].

Formalnym warunkiem istnienia rozwiązania $\hat{f}(a)$ jest kryterium Picarda [1]:

$$\sum_{n=1}^{\infty} \left| \frac{\langle \tilde{g}(y), u_n(y) \rangle_{H_2}}{\sigma_n} \right|^2 < \infty. \quad (4.2.24)$$

Zgodnie z kryterium, współczynniki Fouriera funkcji $\tilde{g}(y)$ przy wzroście n muszą zanikać szybciej niż wartości osobliwe σ_n operatora całkowego K [1]. Niespełnienie warunku (4.2.24) oznacza złe postawienie, natomiast im większa wartość wyrażen pod sumą dla dużych n , tym gorsze jest uwarunkowanie zagadnienia odwrotnego [1].

Stopień niewłaściwego postawienia oraz słabego uwarunkowania problemu narasta, gdy zmniejsza się szybkość zmniejszania się wartości osobliwych σ_n przy wzroście n w stosunku do szybkości zmniejszania się współczynników Fouriera funkcji $\tilde{g}(y)$ [1, 4]. Przyjmuje się często, że stopień niewłaściwego postawienia oraz słabego uwarunkowania zagadnienia jest [1]:

- umiarkowany, jeżeli $\sigma_n = o(n^{-\alpha})$ dla pewnego $\alpha > 0$,
- poważny w innych przypadkach, jeżeli na przykład $\sigma_n = o(e^{-n})$.

5. ANALITYCZNE ROZWIĄZANIA ZAGADNIENIA ODWROTNEGO W ANALIZIE WIELKOŚCI CZĄSTEK UKŁADÓW DYSPERSYJNYCH Z WYKORZYSTANIEM TECHNIK OPTYCZNYCH

Jak wynika z wcześniejszych rozważań, jądro $K(y, a)$ w równaniu całkowym (3.2.2.15) ma niezwykle skomplikowaną postać w przypadku obu rozważanych technik pomiarowych – nefelometrii oraz turbidymetrii zarówno w sytuacji, gdy modele matematyczne pomiarów oparte są na teorii rozpraszania Mie, jak i na jej przybliżeniu – teorii dyfrakcji Fraunhofera. Powoduje to, że rozwiązania analityczne problemu odwrotnego znane są wyłącznie dla prostszych przypadków szczególnych [5].

W niniejszym rozdziale prezentowane jest rozwiązanie analityczne znalezione przez Šifrina dla przypadku pomiarów nefelometrycznych, których model matematyczny oparty jest na teorii dyfrakcji Fraunhofera.

Podstawiając wzór (3.2.1.2.3) do równania (3.2.2.5) oraz uwzględniając związek (3.2.1.3), uzyskuje się następującą szczególną postać równania całkowego (3.2.2.15) dla przypadku pomiarów nefelometrycznych, których model matematyczny oparty jest na teorii dyfrakcji Fraunhofera [5]:

$$\beta(\theta) = \frac{N_v}{k^2 \theta^2} \int_0^\infty x^2 J_1^2(x\theta) f(x) dx. \quad (5.1)$$

W równaniu (5.1) rozkład wielkości cząstek oraz inne wielkości został wyrażony jako funkcja parametru wielkościowego cząstek x .

Metoda Šifrina rozwiązywania równania całkowego (5.1) bazuje na zastosowaniu funkcji specjalnej [5]

$$\varphi_n(\mu) = \int_0^\infty z^n e^{-\mu z} J_1^2(z) dz, \quad (5.2)$$

gdzie n – liczba całkowita. Funkcja ta ma następującą własność [5]:

$$\varphi_n(\mu) = -\frac{d\varphi_{n-1}(\mu)}{d\mu} = (-1)^n \frac{d^n \varphi_0(\mu)}{d\mu^n}. \quad (5.3)$$

Można zauważyć, że funkcja $\varphi_n(\mu)$ dana równaniem (5.2) stanowi transformatę Laplace'a funkcji $z^n J_1^2(z)$. Na tej podstawie rozwiązanie równania (5.1) ma postać [5]

$$f(x) = \frac{k^2}{N_v} \int_0^\infty D(\theta) h_D(x\theta) d\theta, \quad (5.4)$$

gdzie [5]

$$D(\theta) = \frac{d}{d\theta}[\beta(\theta)\theta^3], \quad (5.5)$$

$$h_D(x\theta) = 2\pi x \theta J_1(x\theta) Y_1(x\theta), \quad (5.6)$$

przy czym $Y_1(z)$ – funkcja Bessela drugiego rodzaju rzędu pierwszego, zwana też funkcją Neumana rzędu pierwszego.

Funkcja $f(x)$ określona przez wzory (5.4), (5.5) oraz (5.6) stanowi analityczne rozwiązanie Šifrina zagadnienia odwrotnego sformułowanego za pomocą równania całkowego (5.1). W rzeczywistości w numerycznym obliczaniu funkcji $f(x)$ we wzorze (5.4) całkowanie w przedziale $[0, \infty)$ przybliżone zostaje całkowaniem w skończonym przedziale $[\theta_{\min}, \theta_{\max})$, co wprowadza pewien błąd [5]. Wartość θ_{\min} zdeterminowana jest przez szerokość stosowanej w pomiarach wiązki światła. Wartość θ_{\max} wynika natomiast z faktu, że dla cząstek fazy zdyspergowanej o rozmiarach $x \sim 1$, dla których adekwatna jest stosowana tu teoria dyfrakcji Fraunhofera, mierzona objętościowa funkcja rozpraszania $\beta(\theta)$ szybko maleje ze wzrostem kąta rozpraszania θ i dla dużych kątów mierzone natężenia światła rozpraszanego są porównywalne z szumami źródła światła i detektora [5].

Metoda Šifrina rozwiązywania zagadnienia odwrotnego nie minimalizuje niekorzystnego wpływu słabego uwarunkowania problemu na uzyskiwaną funkcję $f(x)$. W związku z tym procedura numerycznego obliczania rozwiązania $\hat{f}(x)$ na podstawie zakłóconej błędami losowymi funkcji $\tilde{\beta}(\theta)$, przebiegająca zgodnie ze wzorami (5.4)–(5.6), prowadzi do znacznego wzmocnienia błędów w rozwiązaniu $\hat{f}(x)$. Otrzymywany w ten sposób rozkład $\hat{f}(x)$ charakteryzuje się dużą niepewnością przy niewielkim nawet poziomie niepewności danych pomiarowych $\tilde{\beta}(\theta)$.

6. NUMERYCZNE METODY ROZWIĄZYWANIA ZAGADNIENIA ODWROTNEGO W ANALIZIE WIELKOŚCI CZĄSTEK UKŁADÓW DYSPERSYJNYCH Z WYKORZYSTANIEM TECHNIK OPTYCZNYCH

Fakt, że rozwiązania analityczne rozważanego zagadnienia odwrotnego znane są wyłącznie dla prostszych przypadków szczególnych oraz że rozwiązania te dotknięte są niekorzystnym wpływem słabego uwarunkowania, powoduje, że do rozwiązywania problemu stosuje się techniki numeryczne. Metody te, zwane technikami, procedurami lub algorytmami inwersyjnymi, przeznaczone są dla zdyskretyzowanej postaci zagadnienia odwrotnego.

6.1. DYSKRETYZACJA ZAGADNIENIA ODWROTNEGO

Istnieje wiele metod dyskretyzacji równania całkowego (3.2.2.15) [1, 3]. W badaniach prowadzonych przez autorów wykorzystywana była technika oparta na zastosowaniu kolokacji oraz aproksymacji całkowania przez kwadraturę numeryczną [1, 13].

Zakładając, że szukany rozkład wielkości cząstek $f(a)$ przyjmuje znaczne wartości jedynie w pewnym ograniczonym przedziale $[a_{\min}, a_{\max}]$, a poza nim jego wartości są zaniedbywane małe, występującą w równaniu (3.2.2.15) całkę w przedziale $[0, \infty)$ przybliżyć można za pomocą całki w skończonym przedziale $[a_{\min}, a_{\max}]$:

$$\int_0^{\infty} K(y, a) f(a) da \approx \int_{a_{\min}}^{a_{\max}} K(y, a) f(a) da . \quad (6.1.1)$$

Następnie całkę w przedziale $[a_{\min}, a_{\max}]$ aproksymować można za pomocą kwadratury numerycznej, która całkę przybliży za pomocą ważonej sumy wartości funkcji podcałkowej [1]

$$\int_{a_{\min}}^{a_{\max}} K(y, a) f(a) da \approx \sum_{j=1}^q w_j K(y, a_j) f(a_j) , \quad (6.1.2)$$

gdzie w_j – współczynniki wagowe. Punkty a_j oraz odpowiadające im wagi wybrać można na różne sposoby.

W metodzie prostokątów (ang. midpoint rule) przedział $[a_{\min}, a_{\max}]$ dzielony jest na q równych podprzedziałów, każdy o szerokości [1, 13]

$$\Delta a = \frac{a_{\max} - a_{\min}}{q} . \quad (6.1.3)$$

Podprzedziałem j -tym jest [1, 13]

$$[a_{\min} + (j-1)\Delta a, a_{\min} + j\Delta a) \quad (6.1.4)$$

Każdy z punktów a_j stanowi środek odpowiedniego j -tego podprzedziału [1, 13]:

$$a_j = a_{\min} + \left(j - \frac{1}{2} \right) \Delta a . \quad (6.1.5)$$

Funkcja podcałkowa aproksymowana jest za pomocą funkcji odcinkami stałej, tzn. przyjmującej w kolejnych podprzedziałach stałe wartości równe wartości funkcji w środkach tych podprzedziałów [1, 13]:

$$\int_{a_{\min}}^{a_{\max}} K(y, a) f(a) da \approx \sum_{j=1}^q K(y, a_j) f(a_j) \Delta a . \quad (6.1.6)$$

Wynika stąd, że dla metody prostokątów [1, 13]:

$$w_j = \Delta a. \quad (6.1.7)$$

W metodzie trapezów (ang. trapezoid rule) funkcja podcałkowa aproksymowana jest za pomocą funkcji odcinkami liniowej. Odpowiadający temu rozkład punktów a_j określa wzór [1]

$$a_j = a_{\min} + (j-1)\Delta a, \quad (6.1.8)$$

gdzie [1]

$$\Delta a = \frac{a_{\max} - a_{\min}}{q-1}. \quad (6.1.9)$$

Współczynniki wagowe w metodzie trapezów dane są wzorem [1]

$$w_j = \begin{cases} \Delta a, & 2 \leq j \leq p-1 \\ \frac{1}{2}\Delta a, & j = 1, p. \end{cases} \quad (6.1.10)$$

Dokładność aproksymacji całki przez kwadraturę numeryczną wzrasta wraz ze wzrostem q .

W rzeczywistych pomiarach nie jest możliwe wyznaczenie ciągłej zależności wielkości mierzonej od zmiennej kontrolowanej $g(y)$, a jedynie dyskretnego zbioru wartości tej wielkości g_i dla określonych wartości zmiennej kontrolowanej y_i [1, 2, 3]:

$$g_i = g(y_i), \quad y_{\min} \leq y_i \leq y_{\max}, \quad i = 1, \dots, p. \quad (6.1.11)$$

Punkty pomiarowe y_i zwane są punktami kolokacji [1]. Jeżeli punkty te rozmieszczone są odpowiednio gęsto w przedziale $[y_{\min}, y_{\max}]$, możliwe jest odtworzenie przebiegu zależności $g(y)$ na podstawie wartości g_i .

Ogólnie dokładność aproksymacji funkcji $g(y)$ przez zbiór dyskretnych wartości $g_i = g(y_i)$ wzrasta wraz ze wzrostem p .

Wyrażając każdą z wartości g_i za pomocą równania (3.2.2.15) oraz wykorzystując przybliżenie (6.1.1) i przybliżenie (6.1.2) polegające na zastosowaniu kwadratury numerycznej, otrzymuje się następujący układ równań [1]:

$$g_i = g(y_i) = \sum_{j=1}^q w_j K(y_i, a_j) f(a_j), \quad i = 1, \dots, p. \quad (6.1.12)$$

Wprowadzając wektory kolumnowe:

$$\mathbf{g} = [g_1 \quad g_2 \quad \dots \quad g_p]^T = [g(y_1) \quad g(y_2) \quad \dots \quad g(y_p)]^T, \quad (6.1.13)$$

$$\mathbf{f} = [f_1 \ f_2 \ \dots \ f_q]^T = [f(a_1) \ f(a_2) \ \dots \ f(a_q)]^T, \quad (6.1.14)$$

gdzie \mathbf{A}^T – transpozycja macierzy \mathbf{A} oraz macierz

$$\mathbf{K} = [K_{ij}], \quad (6.1.15)$$

której elementy określa wzór

$$K_{ij} = w_j K(y_i, a_j), \quad i = 1, \dots, p, \quad j = 1, \dots, q, \quad (6.1.16)$$

układ równań (6.1.12) zapisać można w postaci [1, 3, 8, 13]

$$\mathbf{g} = \mathbf{Kf}. \quad (6.1.17)$$

Dyskretyzacja zatem równania całkowego Fredholma pierwszego rodzaju (3.2.2.15) prowadzi do układu liniowych równań algebraicznych (6.1.17).

6.2. ROZWIĄZANIE NAJMNIEJSZYCH KWADRATÓW Z MINIMALNĄ NORMĄ ZDYSKRETYZOWANEGO ZAGADNIENIA ODWROTNEGO

Dyskretnym sformułowaniem rozpatrywanego zagadnienia odwrotnego jest układ p liniowych równań algebraicznych z q niewiadomymi (6.1.17). Ponieważ w ogólnym przypadku $p \neq q$, układ ten może nie mieć rozwiązania otrzymywanego przez odwrócenie macierzy \mathbf{K} . W związku z tym w dalszych rozważaniach przez rozwiązanie układu (6.1.17) rozumiane będzie rozwiązanie najmniejszych kwadratów z minimalną normą, które dla układu (6.1.17) istnieje zawsze [1, 3, 4, 14, 15].

Rozwiązanie najmniejszych kwadratów z minimalną normą wyznaczyć można według rozkładu macierzy \mathbf{K} na wartości osobliwe (ang. *SVD – Singular Value Decomposition*), który dany jest wzorem [1, 3, 4, 14, 15]

$$\mathbf{K} = \mathbf{USV}^T, \quad (6.2.1)$$

gdzie: \mathbf{U} – macierz ortogonalna o wymiarach $p \times p$, \mathbf{V} – macierz ortogonalna o wymiarach $q \times q$, \mathbf{S} – macierz o wymiarach $p \times q$ mająca postać [1, 3, 4, 14, 15]

$$\mathbf{S} = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0}_{r, q-r} \\ \mathbf{0}_{p-r, r} & \mathbf{0}_{p-r, q-r} \end{bmatrix}, \quad (6.2.2)$$

przy czym $\mathbf{\Sigma}$ – macierz diagonalna o wymiarach $r \times r$, gdzie r – rząd macierzy \mathbf{K} , czyli maksymalna liczba liniowo niezależnych kolumn lub wierszy macierzy \mathbf{K} , $\mathbf{0}_{m,n}$ – macierz o wymiarach $m \times n$ wypełniona samymi zerami. Elementy diagonalne macierzy $\mathbf{\Sigma}$ są dodatnie i ustawione w porządku malejącym: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$. Zwane są one wartościami osobliwymi macierzy \mathbf{K} . Odpowiadające im kolumny macierzy \mathbf{U} oraz \mathbf{V} stanowią wektory osobliwe macierzy \mathbf{K} [1, 3, 4, 14, 15].

Podstawiając wyrażenia (6.2.1) wraz z (6.2.2) do równania (6.1.17), uzyskuje się

$$\mathbf{g} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0}_{r,q-r} \\ \mathbf{0}_{p-r,r} & \mathbf{0}_{p-r,q-r} \end{bmatrix} \mathbf{V}^T \mathbf{f}, \quad (6.2.3)$$

skąd wynika, że

$$\mathbf{U}^T \mathbf{g} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0}_{r,q-r} \\ \mathbf{0}_{p-r,r} & \mathbf{0}_{p-r,q-r} \end{bmatrix} \mathbf{V}^T \mathbf{f}. \quad (6.2.4)$$

Oznaczając

$$\mathbf{g}' = \mathbf{U}^T \mathbf{g} = \begin{bmatrix} \mathbf{g}'_1 \\ \mathbf{g}'_2 \end{bmatrix}, \quad (6.2.5)$$

gdzie \mathbf{g}'_1 – wektor kolumnowy utworzony z pierwszych r elementów wektora \mathbf{g}' , \mathbf{g}'_2 – wektor kolumnowy utworzony z pozostałych $(p-r)$ elementów wektora \mathbf{g}'

$$\mathbf{f}' = \mathbf{V}^T \mathbf{f} = \begin{bmatrix} \mathbf{f}'_1 \\ \mathbf{f}'_2 \end{bmatrix}, \quad (6.2.6)$$

gdzie \mathbf{f}'_1 – wektor kolumnowy utworzony z pierwszych r elementów wektora \mathbf{f}' , \mathbf{f}'_2 – wektor kolumnowy utworzony z pozostałych $(q-r)$ elementów wektora \mathbf{f}' , równanie (6.2.4) zapisać można w postaci układu [14]

$$\begin{cases} \mathbf{g}'_1 = \boldsymbol{\Sigma} \mathbf{f}'_1 + \mathbf{0}_{r,q-r} \mathbf{f}'_2 \\ \mathbf{g}'_2 = \mathbf{0}_{p-r,r} \mathbf{f}'_1 + \mathbf{0}_{p-r,q-r} \mathbf{f}'_2 \end{cases} \quad (6.2.7)$$

Ponieważ macierz $\boldsymbol{\Sigma}$ jest nieosobliwa [14]

$$\mathbf{f}'_1 = \boldsymbol{\Sigma}^{-1} \mathbf{g}'_1, \quad (6.2.8)$$

co w tym przypadku przyjmuje szczególnie prostą postać

$$f'_{1,n} = \sigma_n^{-1} g'_{1,n}, \quad n=1, \dots, r. \quad (6.2.9)$$

Wektor \mathbf{f}'_2 jest dowolny, co oznacza, że rozwiązanie układu (6.1.17) jest niejednoznaczne [14]. Natomiast dla każdego wektora $\mathbf{g}'_2 \neq \mathbf{0}$ drugie z równań w układzie (6.2.7) jest sprzeczne, co oznacza sprzeczność układu (6.1.17) [14].

Zgodnie ze wzorem (6.2.5) wektor \mathbf{g}' stanowi reprezentację wektora \mathbf{g} w bazie ortonormalnej utworzonej przez kolumny macierzy \mathbf{U} , zgodnie zaś ze wzorem (6.2.6) wektor \mathbf{f}' stanowi reprezentację wektora \mathbf{f} w bazie ortonormalnej utworzonej przez

kolumny macierzy \mathbf{V} . Niech:

$$\mathbf{g}'^* = \begin{bmatrix} \mathbf{g}'_1 \\ \mathbf{0}_{p-r} \end{bmatrix}, \quad (6.2.10)$$

$$\mathbf{g}'_0 = \begin{bmatrix} \mathbf{0}_r \\ \mathbf{g}'_2 \end{bmatrix}, \quad (6.2.11)$$

$$\mathbf{f}'^* = \begin{bmatrix} \mathbf{f}'_1 \\ \mathbf{0}_{q-r} \end{bmatrix}, \quad (6.2.12)$$

$$\mathbf{f}'_0 = \begin{bmatrix} \mathbf{0}_r \\ \mathbf{f}'_2 \end{bmatrix}, \quad (6.2.13)$$

gdzie $\mathbf{0}_n$ – n -elementowy wektor kolumnowy. Na podstawie wcześniejszych rozważań można stwierdzić, że wektor \mathbf{g}'^* stanowi rzut ortogonalny wektora \mathbf{g} na r -wymiarową przestrzeń kolumnową macierzy \mathbf{K} rozpiętą przez r pierwszych kolumn macierzy \mathbf{U} , wektor \mathbf{g}'_0 stanowi rzut ortogonalny wektora \mathbf{g} na $(p-r)$ -wymiarowe dopełnienie ortogonalne przestrzeni kolumnowej macierzy \mathbf{K} w przestrzeni rozpiętej przez kolumny macierzy \mathbf{U} oraz że wektor \mathbf{f}'^* stanowi rzut ortogonalny wektora \mathbf{f} na r -wymiarową przestrzeń wierszową macierzy \mathbf{K} utworzoną przez r pierwszych kolumn macierzy \mathbf{V} , wektor \mathbf{f}'_0 stanowi rzut ortogonalny wektora \mathbf{f} na $(q-r)$ -wymiarowe dopełnienie ortogonalne przestrzeni wierszowej macierzy \mathbf{K} w przestrzeni rozpiętej przez kolumny macierzy \mathbf{V} . Z właściwości rzutu ortogonalnego wynika, że dla danego wektora \mathbf{g}' należącego do p -wymiarowej przestrzeni euklidesowej spośród wszystkich wektorów w należących do r -wymiarowej podprzestrzeni najmniejszą normę euklidesową wektora $\mathbf{g}'_0 = \mathbf{g}' - \mathbf{w}$ zapewnia wektor $\mathbf{w} = \mathbf{g}'^*$. Każdy zatem wektor:

$$\begin{aligned} \mathbf{f}' &= \mathbf{f}'^* + \mathbf{f}'_0 = \begin{bmatrix} \Sigma^{-1} \mathbf{g}'_1 \\ \mathbf{0}_{q-r} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_r \\ \mathbf{f}'_2 \end{bmatrix} = \begin{bmatrix} \Sigma^{-1} & \mathbf{0}_{r,p-r} \\ \mathbf{0}_{q-r,r} & \mathbf{0}_{q-r,p-r} \end{bmatrix} \begin{bmatrix} \mathbf{g}'_1 \\ \mathbf{g}'_2 \end{bmatrix} + \begin{bmatrix} \mathbf{0}_r \\ \mathbf{f}'_2 \end{bmatrix} \\ &= \begin{bmatrix} \Sigma^{-1} & \mathbf{0}_{r,p-r} \\ \mathbf{0}_{q-r,r} & \mathbf{0}_{q-r,p-r} \end{bmatrix} \mathbf{g}' + \begin{bmatrix} \mathbf{0}_r \\ \mathbf{f}'_2 \end{bmatrix} \end{aligned} \quad (6.2.14)$$

minimalizuje wyrażenie

$$\|\mathbf{g}'_0\|_2 = \|\mathbf{g}' - \mathbf{g}'^*\|_2 = \|\mathbf{U}\mathbf{g}' - \mathbf{U}\mathbf{g}'^*\|_2 = \|\mathbf{g} - \mathbf{U}\mathbf{g}'^*\|_2. \quad (6.2.15)$$

Równość (6.2.15) wynika z ortogonalności macierzy \mathbf{U} oraz ze związku (6.2.5). Stosując w wyrażeniu (6.2.15) kolejno zależności: (6.2.10), (6.2.7), (6.2.6) oraz (6.2.3), (6.2.1) i (6.2.2), uzyskuje się

$$\begin{aligned} \|\mathbf{g}'_0\|_2 &= \|\mathbf{g} - \mathbf{U}\mathbf{g}^{*'}\|_2 = \left\| \mathbf{g} - \mathbf{U} \begin{bmatrix} \mathbf{g}'_1 \\ \mathbf{0}_{p-r} \end{bmatrix} \right\|_2 = \left\| \mathbf{g} - \mathbf{U} \begin{bmatrix} \Sigma \mathbf{f}'_1 + \mathbf{0}_{r,q-r} \mathbf{f}'_2 \\ \mathbf{0}_{p-r} \end{bmatrix} \right\|_2 \\ &= \left\| \mathbf{g} - \mathbf{U} \begin{bmatrix} \Sigma & \mathbf{0}_{r,q-r} \\ \mathbf{0}_{p-r,r} & \mathbf{0}_{p-r,q-r} \end{bmatrix} \begin{bmatrix} \mathbf{f}'_1 \\ \mathbf{f}'_2 \end{bmatrix} \right\|_2 = \left\| \mathbf{g} - \mathbf{U} \begin{bmatrix} \Sigma & \mathbf{0}_{r,q-r} \\ \mathbf{0}_{p-r,r} & \mathbf{0}_{p-r,q-r} \end{bmatrix} \mathbf{V}^T \mathbf{f} \right\|_2 \\ &= \|\mathbf{g} - \mathbf{K}\mathbf{f}\|_2 = \|\boldsymbol{\varepsilon}\|_2. \end{aligned} \quad (6.2.16)$$

Uwzględniając wzory (6.2.5) i (6.2.6) w wyrażeniu (6.2.14), można stwierdzić, że każdy wektor

$$\mathbf{f} = \mathbf{V}\mathbf{f}' = \mathbf{V} \begin{bmatrix} \Sigma^{-1} & \mathbf{0}_{r,p-r} \\ \mathbf{0}_{q-r,r} & \mathbf{0}_{q-r,p-r} \end{bmatrix} \mathbf{U}^T \mathbf{g} + \mathbf{V} \begin{bmatrix} \mathbf{0}_r \\ \mathbf{f}'_2 \end{bmatrix} \quad (6.2.17)$$

minimalizuje normę euklidesową wektora rezyduów, czyli: $\|\boldsymbol{\varepsilon}\|_2 = \|\mathbf{g} - \mathbf{K}\mathbf{f}\|_2$. Wobec tego każdy z tych wektorów stanowi rozwiązanie najmniejszych kwadratów układu równań (6.1.17) [14]. Rozwiązań takich jest nieskończenie wiele, ponieważ wektor \mathbf{f}'_2 jest dowolny. Jak wynika ze wzoru (6.2.17), przyjmując $\mathbf{f}'_2 = \mathbf{0}_{q-r}$, otrzymuje się rozwiązanie najmniejszych kwadratów z minimalną normą euklidesową dane wyrażeniem [15]:

$$\mathbf{f}_{LSMN} = \mathbf{V} \begin{bmatrix} \Sigma^{-1} & \mathbf{0}_{r,p-r} \\ \mathbf{0}_{q-r,r} & \mathbf{0}_{q-r,p-r} \end{bmatrix} \mathbf{U}^T \mathbf{g}. \quad (6.2.18)$$

Przyjmując oznaczenie [15]:

$$\mathbf{K}^+ = \mathbf{V} \begin{bmatrix} \Sigma^{-1} & \mathbf{0}_{r,p-r} \\ \mathbf{0}_{q-r,r} & \mathbf{0}_{q-r,p-r} \end{bmatrix} \mathbf{U}^T, \quad (6.2.19)$$

można zapisać [15]:

$$\mathbf{f}_{LSMN} = \mathbf{K}^+ \mathbf{g}. \quad (6.2.20)$$

Macierz \mathbf{K}^+ zwana jest pseudoodwrotnością macierzy \mathbf{K} [15].

Oznaczając kolejne kolumny macierzy \mathbf{U} i macierzy \mathbf{V} odpowiednio \mathbf{u}_n i \mathbf{v}_n oraz zapisując iloczyn skalarny wektorów \mathbf{t} i \mathbf{u} w przestrzeni euklidesowej jako

$$\langle \mathbf{t}, \mathbf{u} \rangle = \mathbf{t}^T \mathbf{u} = \mathbf{t} \mathbf{u}^T, \quad (6.2.21)$$

rozwiązanie (6.2.18) wyrazić można wzorem [1, 3, 4]

$$\mathbf{f}_{\text{LSMN}} = \sum_{n=1}^r \frac{\mathbf{u}_n^T \mathbf{g}}{\sigma_n} \mathbf{v}_n = \sum_{n=1}^r \frac{\langle \mathbf{g}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n. \quad (6.2.22)$$

W rzeczywistości rząd układu równań (6.1.17) zwykle wynosi

$$r = \min(p, q), \quad (6.2.23)$$

czyli układ charakteryzuje się pełnym rzędem wierszowym $r = p$ lub kolumnowym $r = q$.

Należy zwrócić uwagę na podobieństwo wzoru (6.2.22) oraz (4.2.13), jak również podobieństwa par wzorów: (4.2.7) i (6.2.6), (4.2.8) i (6.2.5) oraz (4.2.11) i (6.2.9). Podobieństwa te wynikają z faktu, że macierz \mathbf{K} reprezentuje operator liniowy K_{dyskr} przekształcający wektor \mathbf{f} należący do q -wymiarowej przestrzeni euklidesowej w wektor \mathbf{g} należący do p -wymiarowej przestrzeni euklidesowej zgodnie ze wzorem (6.1.17). Operator K_{dyskr} oraz operator całkowy K rozważany w punkcie 4.2 różnią się wymiarem oraz rodzajem przestrzeni będących ich dziedzinami i obrazami – dla operatora K dziedzinę i obraz stanowią nieskończeniowymiarowe przestrzenie funkcyjne Hilberta L_2 z iloczynem skalarnym określonym wzorem (4.2.2) dla dziedziny i (4.2.3) dla obrazu, natomiast dla operatora K_{dyskr} dziedzinę i obraz stanowią skończeniowymiarowe przestrzenie euklidesowe l_2 z iloczynem skalarnym określonym ogólnie wzorem (6.2.21). Wartości osobliwe i wektory osobliwe macierzy \mathbf{K} są wartościami osobliwymi i wektorami osobliwymi operatora K_{dyskr} reprezentowanego przez tę macierz. Ze wzoru (6.2.1) wynika bowiem

$$\mathbf{KV} = \mathbf{US}, \quad (6.2.24)$$

$$\mathbf{K}^T \mathbf{U} = \mathbf{VS}, \quad (6.2.25)$$

co zapisać można w postaci odpowiednio

$$\mathbf{Kv}_n = \sigma_n \mathbf{u}_n, \quad n = 1, \dots, r \quad (6.2.26)$$

$$\mathbf{K}^T \mathbf{u}_n = \sigma_n \mathbf{v}_n, \quad n = 1, \dots, r \quad (6.2.27)$$

Na podstawie wzorów (6.2.1) i (6.2.2) wyprowadzić można następujące związki [15]:

$$\mathbf{K}^T \mathbf{K} = \mathbf{V} \begin{bmatrix} \Sigma^2 & \mathbf{0}_{r, q-r} \\ \mathbf{0}_{q-r, r} & \mathbf{0}_{q-r, q-r} \end{bmatrix} \mathbf{V}^T, \quad (6.2.28)$$

$$\mathbf{K} \mathbf{K}^T = \mathbf{U} \begin{bmatrix} \Sigma^2 & \mathbf{0}_{r, p-r} \\ \mathbf{0}_{p-r, r} & \mathbf{0}_{p-r, p-r} \end{bmatrix} \mathbf{U}^T. \quad (6.2.29)$$

Z zależności (6.2.28) wynika, że jeżeli układ równań (6.1.17) ma pełen rząd kolumnowy, czyli $q = r$, macierz $\mathbf{K}^T\mathbf{K}$ jest nieosobliwa, a jej wartości własne i wektory własne stanowią odpowiednio: σ_i^2 oraz \mathbf{v}_i , $i = 1, \dots, q$. Mnożąc (6.1.17) lewostronnie przez \mathbf{K}^T , uzyskuje się

$$\mathbf{K}^T\mathbf{g} = \mathbf{K}^T\mathbf{K}\mathbf{f}, \quad (6.2.30)$$

skąd

$$\mathbf{f}_{LS} = (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\mathbf{g}. \quad (6.2.31)$$

Jest to zwykle rozwiązanie najmniejszych kwadratów [1, 3, 14, 15].

Z zależności (6.2.29) wynika, że jeżeli układ równań (6.1.17) ma pełen rząd wierszowy, czyli $p = r$, macierz $\mathbf{K}\mathbf{K}^T$ jest nieosobliwa, a jej wartości własne i wektory własne stanowią odpowiednio: σ_i^2 oraz \mathbf{u}_i , $i = 1, \dots, p$. Przez podstawienie można sprawdzić, że rozwiązanie (6.2.18) przybiera w tym przypadku postać

$$\mathbf{f}_{MN} = \mathbf{K}^T(\mathbf{K}\mathbf{K}^T)^{-1}\mathbf{g}. \quad (6.2.32)$$

Jest to zwykle rozwiązanie z minimalną normą.

6.3. SŁABE UWARUNKOWANIE ZDYSKRETYZOWANEGO ZAGADNIENIA ODWROTNEGO

Następstwem słabego uwarunkowania zagadnienia odwrotnego, sformułowanego za pomocą równania całkowego (3.2.2.15), jest słabe uwarunkowanie numeryczne układu równań (6.1.17) stanowiącego zdyskretyzowaną postać równania (3.2.2.15) [1, 3, 4, 5].

Macierz \mathbf{K} reprezentuje operator liniowy K_{dyskr} będący dyskretną aproksymacją operatora całkowego K zdefiniowanego równaniem (4.2.1). W związku z tym operator K_{dyskr} wykazuje właściwości analogiczne do własności operatora K . Operator K_{dyskr} powoduje wygładzanie wektora \mathbf{f} , tzn. ma właściwość filtrowania z niego wielkoczęstotliwościowych oscylacji oraz gwałtownych zmian wartości. Ścisłej sformułować to można w postaci dwóch ogólnych właściwości [4]:

- wartości osobliwe σ_n operatora K_{dyskr} bardzo szybko maleją ze wzrostem indeksu n , zbliżając się do zera,
- wektory osobliwe \mathbf{u}_n oraz \mathbf{v}_n wraz ze wzrostem indeksu n prezentują coraz bardziej oscylacyjny przebieg, tzn. wykazują coraz więcej zmian znaków.

Analogicznie do przypadku operatora K , dane pomiarowe są małowzruszone na wielkoczęstotliwościowe oscylacje w wektorze \mathbf{f} nawet, gdy mają one znaczną amplitudę. Z kolei rozwiązywanie zagadnienia odwrotnego działa w przeciwną stronę – wiąże się ze znacznym wzmocnieniem wielkoczęstotliwościowych oscylacji występujących

w wektorze \mathbf{g} [3, 4]. Ponieważ $f'_{1,n}$ i $g'_{1,n}$ stanowią współczynniki rozwinięcia wektorów odpowiednio \mathbf{f} i \mathbf{g} na ortogonalne składowe oscylacyjne o rosnących częstotliwościach zgodnie ze wzorami odpowiednio (6.2.6) i (6.2.5), wzmocnienie to wynika z tego, że współczynniki $f'_{1,n}$ o dużych indeksach n są rezultatem dzielenia współczynników $g'_{1,n}$ przez znikomo małe wartości osobliwe σ_n we wzorze (6.2.9) [3, 4].

Uwzględnienie w analizie addytywnych losowych błędów pomiarowych oraz błędów zaokrągleń zakłócających dane pomiarowe \mathbf{g} następuje przez zastąpienie wektora \mathbf{g} , będącego wyjściem idealnego modelu (6.1.17) przez wektor rzeczywistych danych pomiarowych [3, 4]

$$\tilde{\mathbf{g}} = \mathbf{g} + \mathbf{e}, \quad (6.3.1)$$

gdzie \mathbf{e} – wektor reprezentujący zakłócenia addytywne (proces stochastyczny). Równanie (6.1.17) przyjmuje w tej sytuacji postać [3, 4, 8]

$$\tilde{\mathbf{g}} = \mathbf{K}\mathbf{f} + \mathbf{e}. \quad (6.3.2)$$

Wektor $\hat{\mathbf{f}}_{\text{LSMN}}$ wyznaczony na podstawie rzeczywistych danych pomiarowych $\tilde{\mathbf{g}}$ stanowi jedynie pewne przybliżenie wektora prawdziwego rozwiązania \mathbf{f}_{LSMN} [3]. Zastępując \mathbf{g} przez $\tilde{\mathbf{g}}$ we wzorze (6.2.22) oraz uwzględniając zależność (6.3.1), uzyskuje się

$$\begin{aligned} \hat{\mathbf{f}}_{\text{LSMN}} &= \sum_{n=1}^r \frac{\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n = \sum_{n=1}^r \frac{\langle \mathbf{g} + \mathbf{e}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n \\ &= \sum_{n=1}^r \frac{\langle \mathbf{g}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n + \sum_{n=1}^r \frac{\langle \mathbf{e}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n. \end{aligned} \quad (133)(6.3.3)$$

Wynika stąd, że błąd rozwiązania $\hat{\mathbf{f}}_{\text{LSMN}}$ w stosunku do prawdziwego rozwiązania \mathbf{f}_{LSMN} stanowi składnik:

$$\mathbf{r} = \sum_{n=1}^r \frac{\langle \mathbf{e}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n. \quad (6.3.4)$$

Wyrażenie (6.3.3) zapisać można więc w postaci

$$\hat{\mathbf{f}}_{\text{LSMN}} = \mathbf{f}_{\text{LSMN}} + \mathbf{r}. \quad (6.3.5)$$

Analogicznie do przypadku operatora całkowego K , szum reprezentowany przez wektor \mathbf{e} charakteryzuje się przebiegiem szybkozmiennym, co powoduje, że współczynniki jego rozwinięcia na wektory osobliwe $\mathbf{u}_n - \langle \mathbf{e}, \mathbf{u}_n \rangle$ przyjmują istotne wartości nawet dla znacznych n , bardzo wolno maleją one ze wzrostem n , a często pozostają

stają stałe lub nawet rosną ze zwiększającym się n . Jednocześnie wolny od zakłóceń wektor \mathbf{g} reprezentuje funkcję gładką, co sprawia, że współczynniki jego rozwinięcia na wektory osobliwe $\mathbf{u}_n - \langle \mathbf{g}, \mathbf{u}_n \rangle$ szybko maleją ze wzrostem n . W związku z tym głównie wielkoczęstotliwościowe oscylacje szumu \mathbf{e} zostają silnie wzmocnione w procesie rozwiązywania zagadnienia odwrotnego za sprawą dzielenia we wzorze (6.3.4) współczynników $\langle \mathbf{e}, \mathbf{u}_n \rangle$ przez znikomo małe wartości osobliwe σ_n odpowiadające dużym indeksom n . Powoduje to, że jedynie dla małych n wartości współczynników rozwinięcia wektora $\hat{\mathbf{f}}_{\text{LSMN}}$ na wektory osobliwe $\mathbf{v}_n - \langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle / \sigma_n$ są zbliżone do wartości $\langle \mathbf{g}, \mathbf{u}_n \rangle / \sigma_n$ odpowiadających prawdziwemu rozwiązaniu \mathbf{f}_{LSMN} . Dla dużych n wartości współczynników $\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle / \sigma_n$ w coraz większym stopniu zdominowane są przez wielkoczęstotliwościowe składowe błędu $\langle \mathbf{e}, \mathbf{u}_n \rangle / \sigma_n$, ponieważ współczynniki $\langle \mathbf{g}, \mathbf{u}_n \rangle / \sigma_n$ prawdziwego rozwiązania \mathbf{f}_{LSMN} szybko maleją ze wzrostem n . Zjawisko to prowadzi ostatecznie do olbrzymiego wzrostu wielkości błędu rozwiązania $\|\mathbf{r}\|$ dla niewielkiej nawet wielkości błędów pomiarowych $\|\mathbf{e}\|$, co oznacza słabe uwarunkowanie [1, 3, 4].

Reasumując, za słabe uwarunkowanie dyskretnej postaci zagadnienia odwrotnego (6.1.17) odpowiedzialne jest wygładzające działanie operatora liniowego reprezentowanego przez macierz \mathbf{K} , przejawiające się bardzo szybkim zmniejszaniem się jego wartości osobliwych σ_n przy wzroście n . Sprawia to, że w rozwiązaniu $\hat{\mathbf{f}}_{\text{LSMN}}$ największemu wzmocnieniu podlegają wielkoczęstotliwościowe składowe szumu \mathbf{e} [1, 3, 4].

Stopień słabego uwarunkowania problemu narasta, gdy zwiększa się szybkość zmniejszania się wartości osobliwych σ_n przy wzroście n w stosunku do szybkości zmniejszania się współczynników $\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle$ [1, 4]. Uwarunkowanie zagadnienia pogarsza się wraz ze zwiększaniem się wartości wyrażenia [1, 4]

$$C_{\text{Picard}} = \sum_{n=1}^r \left| \frac{\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle}{\sigma_n} \right|^2. \quad (6.3.6)$$

Za formalny warunek dobrego uwarunkowania przyjmuje się dyskretne kryterium Picarda, które stwierdza, że aby zagadnienie wykazywało dobre uwarunkowanie, współczynniki fourierowskie $\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle$ średnio muszą zanikać szybciej niż wartości osobliwe σ_n macierzy \mathbf{K} [1, 4]. Zdyskretyzowany problem odwrotny (6.1.17) spełnia dyskretne kryterium Picarda, jeżeli odpowiadający mu problem ciągły (4.2.1) spełnia ciągłe kryterium Picarda (4.2.24) [4, 16].

Liczbowa miarą uwarunkowania numerycznego układu równań (6.1.17) jest liczba warunkowa macierzy \mathbf{K} zdefiniowana jako stosunek największej do najmniejszej wartości osobliwej macierzy \mathbf{K} [3, 17]:

$$\kappa_2(\mathbf{K}) = \frac{\sigma_1}{\sigma_r}. \quad (6.3.7)$$

W związku z bardzo szybkim zmniejszaniem się wartości osobliwych σ_n przy wzroście n liczba warunkowa osiąga olbrzymie wartości, co jest oznaką bardzo słabego uwarunkowania układu równań (6.1.17) [3].

Małe wartości osobliwe macierzy \mathbf{K} świadczą o tym, że macierz ta jest bliska macierzy osobliwej, co oznacza, że jej wiersze są wzajemnie niemal liniowo zależne oraz jej kolumny są wzajemnie niemal liniowo zależne [3, 4].

Małe wartości osobliwe σ_n dla dużych n powodują, że układ równań (6.1.17) jest bliski układowi niedookreśloności [3]. Znaczna niepewność rozwiązania $\hat{\mathbf{f}}_{\text{LSMN}} - \Delta\hat{\mathbf{f}}_{\text{LSMN}}$ przy niewielkiej nawet niepewności danych pomiarowych $\tilde{\mathbf{g}} - \Delta\tilde{\mathbf{g}}$ jakościowo traktowana może być jako efekt tego, że warunek spełniania przez wektor $\hat{\mathbf{f}}_{\text{LSMN}}$ układu równań (6.1.17), czyli warunek jego zgodności z danymi pomiarowymi $\tilde{\mathbf{g}}$, nie określa w sposób dokładny i jednoznaczny tego wektora. Istnieje bowiem cały szeroki przedział wektorów $[\hat{\mathbf{f}}_{\text{LSMN}} - \Delta\hat{\mathbf{f}}_{\text{LSMN}}, \hat{\mathbf{f}}_{\text{LSMN}} + \Delta\hat{\mathbf{f}}_{\text{LSMN}}]$ spełniających układ równań (6.1.17) z zadaną dobrą dokładnością $\Delta\tilde{\mathbf{g}}$ [3]. Uzyskanie jednoznacznie określonego rozwiązania zagadnienia odwrotnego wymaga uzupełnienia układu równań (6.1.17), wyrażającego warunek zgodności rozwiązania z danymi pomiarowymi przez dodatkowy warunek nałożony *a priori* na poszukiwane rozwiązanie. W procesie rozwiązywania problemu konieczne jest więc zastosowanie dodatkowej informacji apriorycznej o rozwiązaniu [3, 18].

6.4. REGULARYZACJA ZDYSKRETYZOWANEGO ZAGADNIENIA ODWROTNEGO

Jak wykazano w punkcie 6.3, słabe uwarunkowanie zagadnienia odwrotnego wyrażonego układem równań (6.1.17) ma swoje źródło w wygładzających właściwościach operatora liniowego reprezentowanego przez macierz \mathbf{K} , czego przejawem jest bardzo szybkie zmniejszanie się wartości osobliwych σ_n tego operatora przy wzroście n . Powoduje to, że wnoszone przez szum wielkoczęstotliwościowe oscylacje w danych pomiarowych $\tilde{\mathbf{g}}$ podlegają niezwykle silnemu wzmocnieniu w rozwiązaniu $\hat{\mathbf{f}}_{\text{LSMN}}$, w efekcie czego rozwiązanie te mają nierealistycznie oscylacyjny przebieg. W związku z tym naturalnym warunkiem, który należy nałożyć *a priori* na poszukiwane rozwiązanie, w celu jego dookreślenia i przez to poprawy uwarunkowania za-

gładkości jest warunek gładkości funkcji $\hat{f}(a)$ reprezentowanej przez wektor $\hat{\mathbf{f}}$ [1, 3, 4]. Warunek ten wynika z apriorycznej wiedzy, że funkcje rozkładu wielkości cząstek fazy zdyspergowanej układów dyspersyjnych $f(a)$ są funkcjami gładkimi [3]. Zastosować można różne ilościowe miary gładkości funkcji $f(a)$. Uwzględnienie warunku gładkości rozwiązania $\hat{f}(a)$ powoduje, że zamiast rozwiązania najmniejszych kwadratów z minimalną normą $\hat{\mathbf{f}}_{\text{LSMN}}$ oryginalnego słabo uwarunkowanego problemu (6.1.17) znajdowane jest rozwiązanie najmniejszych kwadratów z minimalną normą innego, lecz bliskiego mu, problemu charakteryzującego się lepszym uwarunkowaniem. Zagadnienie to zwane jest zagadnieniem zregularyzowanym, a jego rozwiązanie – rozwiązaniem zregularyzowanym $\hat{\mathbf{f}}_{\text{LSMNReg}}$ [1, 3, 4, 18–21].

Niezależnie od przyjętej miary gładkości funkcji $f(a)$, warunek gładkości uwzględnić można ogólnie przez wprowadzenie odpowiednich współczynników wagowych t_n , $n = 1, \dots, r$ do rozwiązania danego wzorem (6.2.18) lub równoważnym mu wzorem (6.2.22), co daje następujące rozwiązanie zregularyzowane [1, 3, 4, 22–25]:

$$\hat{\mathbf{f}}_{\text{LSMNReg}} = \sum_{n=1}^r t_n \frac{\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n. \quad (6.4.1)$$

Współczynniki wagowe t_n zwane są również współczynnikami filtrującymi [1, 3, 4, 22–24]. Definiując macierz diagonalną

$$\mathbf{T} = \begin{bmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & t_r \end{bmatrix}, \quad (6.4.2)$$

rozwiązanie zregularyzowane można zapisać w postaci [1]:

$$\hat{\mathbf{f}}_{\text{LSMNReg}} = \mathbf{V} \begin{bmatrix} \mathbf{T}\boldsymbol{\Sigma}^{-1} & \mathbf{0}_{r,p-r} \\ \mathbf{0}_{q-r,r} & \mathbf{0}_{q-r,p-r} \end{bmatrix} \mathbf{U}^T \tilde{\mathbf{g}}. \quad (6.4.3)$$

Przyjmując oznaczenie [1]:

$$\mathbf{K}^\# = \mathbf{V} \begin{bmatrix} \mathbf{T}\boldsymbol{\Sigma}^{-1} & \mathbf{0}_{r,p-r} \\ \mathbf{0}_{q-r,r} & \mathbf{0}_{q-r,p-r} \end{bmatrix} \mathbf{U}^T, \quad (6.4.4)$$

można zapisać [1]:

$$\hat{\mathbf{f}}_{\text{LSMNReg}} = \mathbf{K}^\# \tilde{\mathbf{g}}. \quad (6.4.5)$$

Macierz $\mathbf{K}^\#$ zwana jest macierzą regularyzacji [1]. Współczynniki filtrujące często, choć nie zawsze, spełniają warunek [1]:

$$0 \leq t_n \leq 1, \quad n = 1, \dots, r, \quad (6.4.6)$$

Macierz \mathbf{T} w pełni charakteryzuje daną procedurę regularyzacji.

Operację wygładzania przy wyznaczaniu rozwiązania $\hat{\mathbf{f}}_{\text{LSMNRreg}}$, według wzoru (6.4.1) realizuje się przez odpowiedni dobór współczynników filtrujących t_n . W związku z apriorycznym założeniem o gładkości funkcji $f(a)$ współczynniki rozwinięcia (6.2.22) prawdziwego rozwiązania \mathbf{f}_{LSMN} na wektory osobliwe \mathbf{v}_n , stojące przy wielkoczęstotliwościowych wektorach \mathbf{v}_n o większych indeksach $n - \langle \mathbf{g}, \mathbf{u}_n \rangle / \sigma_n$, powinny mieć małe wartości. Jak wynika z zależności (6.3.3), obserwowane w rzeczywistości olbrzymie wartości współczynników rozwinięcia rozwiązania $\hat{\mathbf{f}}_{\text{LSMN}}$ na wektory osobliwe \mathbf{v}_n o większych indeksach $n - \langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle / \sigma_n$ są zdominowane przez składową $\langle \mathbf{e}, \mathbf{u}_n \rangle / \sigma_n$ pochodzącą od szumu, która dla dużych n jest ciągle znaczna w związku z oscylacyjnym przebiegiem szumu i istotnie przewyższa składową $\langle \mathbf{g}, \mathbf{u}_n \rangle / \sigma_n$ pochodzącą od gładkiej nie zakłóconej funkcji $g(y)$. Współczynniki rozwinięcia $\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle / \sigma_n$ dla dużych n niosą więc znikomą informację o prawdziwym rozwiązaniu \mathbf{f}_{LSMN} i reprezentują błąd tego rozwiązania \mathbf{r} pochodzący od szumu \mathbf{e} . Stąd odpowiednie wytłumienie – „odfiltrowanie” składników $\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle \mathbf{v}_n / \sigma_n$ dla dużych n istotnie zmniejsza niepewność rozwiązania $\hat{\mathbf{f}}_{\text{LSMN}}$, powodując nieznaczną tylko utratę informacji niesionej przez dane pomiarowe $\tilde{\mathbf{g}}$ [3]. Wynika stąd, iż współczynniki filtrujące powinny być dobrane tak, by w największym stopniu tłumione były wielkoczęstotliwościowe wyrazy rozwinięcia (6.3.3) rozwiązania $\hat{\mathbf{f}}_{\text{LSMN}}$ na wektory osobliwe \mathbf{v}_n odpowiadające większym indeksom n lub równoważnie mniejszym wartościom osobliwym σ_n [3]. Filtracja powinna minimalizować utratę informacji niesionej przez dane pomiarowe $\tilde{\mathbf{g}}$, jednocześnie w jak największym stopniu eliminując z rozwiązania oscylacyjne artefakty, a tym samym, zmniejszając niepewność tego rozwiązania [3]. Zaproponowano wiele sposobów doboru współczynników filtracji [1, 3, 4, 22–25].

6.4.1. REGULARYZACJA TICHONOWA

W regularyzacji Tichonowa współczynniki filtrujące określone są przez wyrażenie [1, 3, 4, 19–21]:

$$t_n = \frac{\sigma_n^2}{\sigma_n^2 + \lambda^2}, \quad n = 1, \dots, r, \quad (6.4.1.1)$$

gdzie $\lambda > 0$ – parametr regularyzacji.

Można wykazać, że w przypadku, gdy macierz \mathbf{K} ma pełen rząd kolumnowy, czyli $r = q < p$, rozwiązanie zregularyzowane dane wzorem (6.4.3) przy współczynnikach filtrujących określonych wyrażeniem (6.4.1.1) stanowi wektor $\hat{\mathbf{f}}_{\text{LSTich}}$ taki, że [1, 3, 4, 19–21]:

$$\hat{\mathbf{f}}_{\text{LSTich}} = \operatorname{argmin}_{\mathbf{f}} \{ \|\tilde{\mathbf{g}} - \mathbf{K}\mathbf{f}\|_2^2 + \lambda^2 \|\mathbf{f}\|_2^2 \}. \quad (6.4.1.2)$$

Najbardziej ogólną postać zregularyzowanego rozwiązania Tichonowa określa wzór [1, 3, 4, 19–21, 26, 27]:

$$\hat{\mathbf{f}}_{\text{LSTich}} = \operatorname{argmin}_{\mathbf{f}} \{ \|\tilde{\mathbf{g}} - \mathbf{K}\mathbf{f}\|_2^2 + \lambda^2 \|\mathbf{L}(\mathbf{f} - \mathbf{f}^*)\|_2^2 \}, \quad (6.4.1.3)$$

gdzie: \mathbf{L} – macierz o liczbie kolumn q , \mathbf{f}^* – rozwiązanie zagadnienia odwrotnego założone *a priori*.

Dla $\mathbf{L} = \mathbf{I}_q$, gdzie \mathbf{I}_q – macierz jednostkowa o wymiarze $q \times q$ oraz $\mathbf{f}^* = \mathbf{0}_q$ rozwiązanie (6.4.1.3) przyjmuje szczególną postać (6.4.1.2). Stosując odpowiednie transformacje, możliwe jest przekształcenie postaci ogólnej zagadnienia (6.4.1.3) w tzw. postać standardową daną wyrażeniem [4]:

$$\hat{\mathbf{f}}_{\text{LSTich,std}} = \operatorname{argmin}_{\mathbf{f}} \{ \|\tilde{\mathbf{g}} - \overline{\mathbf{K}}\mathbf{f}\|_2^2 + \lambda^2 \|\bar{\mathbf{f}} - \bar{\mathbf{f}}^*\|_2^2 \}, \quad (6.4.1.4)$$

gdzie pozioma kreska nad symbolami wektorów i macierzy oznacza ich postacie po transformacji.

Zagadnienie minimalizacyjne (6.4.1.3) przedstawić można jako problem minimalizacji funkcji [3, 26, 27]:

$$J(\mathbf{f}) = \|\mathbf{L}(\mathbf{f} - \mathbf{f}^*)\|_2^2 \quad (6.4.1.5)$$

z ograniczeniem nierównościowym [26, 27]:

$$\|\tilde{\mathbf{g}} - \mathbf{K}\mathbf{f}\|_2^2 < \varepsilon, \quad (6.4.1.6)$$

gdzie $\varepsilon > 0$ – zadana niewielka wartość.

$$\hat{\mathbf{f}}_{\text{LSTich}} = \min_{\mathbf{f}: \|\tilde{\mathbf{g}} - \mathbf{K}\mathbf{f}\|_2^2 < \varepsilon} \{ J(\mathbf{f}) \}. \quad (6.4.1.7)$$

Wyrażenie (6.4.1.3) stanowi rezultat przekształcenia zagadnienia minimalizacji z ograniczeniem (6.4.1.7) w zagadnienie minimalizacji bez ograniczeń z zastosowaniem metody mnożników Lagrange’a. Rolę mnożnika Lagrange’a spełnia w tym przypadku parametr regularyzacji [26, 27].

Ze wzoru (6.4.1.7) wynika, że zregulowanym rozwiązaniem zagadnienia odwrotnego (6.1.17) jest wektor $\hat{\mathbf{f}}_{\text{LSTich}}$, który minimalizuje funkcję $J(\mathbf{f})$, zapewniając jednocześnie ustaloną niewielką wartość odchylenia wyjścia modelu dla rozwiązania $\hat{\mathbf{f}}_{\text{LSTich}} - \mathbf{K}\hat{\mathbf{f}}_{\text{LSTich}}$ od rzeczywistych danych pomiarowych $\tilde{\mathbf{g}}$, zgodnie z warunkiem (6.4.1.6). Wyrażenie $\|\tilde{\mathbf{g}} - \mathbf{K}\hat{\mathbf{f}}\|_2^2$ stanowi wielkość minimalizowaną przez zwykłe rozwiązanie najmniejszych kwadratów (6.2.31) i stanowi miarę zgodności rozwiązania $\hat{\mathbf{f}}_{\text{LS}}$ z danymi pomiarowymi $\tilde{\mathbf{g}}$. W rezultacie słabego uwarunkowania rozwiązanie $\hat{\mathbf{f}}_{\text{LS}}$ charakteryzuje się olbrzymią niepewnością $\Delta\hat{\mathbf{f}}_{\text{LS}}$, co jakościowo traktować można jako niejednoznaczność rozwiązania $\hat{\mathbf{f}}_{\text{LS}}$, ponieważ istnieje cały szeroki przedział wektorów $[\hat{\mathbf{f}}_{\text{LS}} - \Delta\hat{\mathbf{f}}_{\text{LS}}, \hat{\mathbf{f}}_{\text{LS}} + \Delta\hat{\mathbf{f}}_{\text{LS}}]$ spełniających warunek (6.4.1.6), czyli zgodnych z danymi pomiarowymi z dokładnością do ε [3]. W procedurze regularyzacji z całego szerokiego zakresu wektorów $\hat{\mathbf{f}}_{\text{LS}}$ wybiera się ten, który spełnia nałożony *a priori* warunek minimalizacji funkcji $J(\mathbf{f})$. Wyrażenie $J(\mathbf{f})$ stanowi miarę odstępstwa wektora \mathbf{f} od apriorycznie założonego charakteru. Warunek zatem minimalizacji przez poszukiwane rozwiązanie odstępstwa od apriorycznie ustalonej charakterystyki dookreśla to rozwiązanie, zmniejszając jego niepewność. Funkcja $J(\mathbf{f})$ nazywana jest funkcją kary [1, 19–21].

Jak uzasadniono we wcześniejszych rozważaniach, założoną *a priori* charakterystyką poszukiwanego rozwiązania $\hat{\mathbf{f}}_{\text{LSTich}}$ powinna być gładkość funkcji $\hat{f}_{\text{LSTich}}(a)$. Miarę braku gładkości funkcji $f(a)$ ogólnie wyrazić można wzorem (6.4.1.5) dla $\mathbf{f}^* = \mathbf{0}_q$, czyli [1, 3, 4, 26, 27]:

$$J(\mathbf{f}) = \|\mathbf{L}\mathbf{f}\|_2^2 \quad (6.4.1.8)$$

Najczęściej stosowanymi miarami są [1, 3, 4, 26, 27]:

- norma euklidesowa wektora \mathbf{f} , której odpowiada macierz

$$\mathbf{L} = \mathbf{I}_q, \quad (6.4.1.9)$$

- norma euklidesowa wektora różnic s -tego rzędu wektora \mathbf{f} , której odpowiada macierz pasmowa o wymiarze $(q-s) \times q - \mathbf{L}_{s,(q-s) \times q}$. Dla kilku początkowych s macierz ta przyjmuje następujące postacie [1, 26]:

– dla $s = 1$:

$$\mathbf{L}_{1,(q-1) \times q} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix}, \quad (6.4.1.10)$$

– dla $s = 2$:

$$\mathbf{L}_{2,(q-2)\times q} = \mathbf{L}_{1,(q-2)\times(q-1)} \mathbf{L}_{1,(q-1)\times q} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}, \quad (6.4.1.11)$$

– dla $s = 3$:

$$\mathbf{L}_{3,(q-3)\times q} = \mathbf{L}_{1,(q-3)\times(q-2)} \mathbf{L}_{2,(q-2)\times q} = \begin{bmatrix} 1 & -3 & 3 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -3 & 3 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 1 & -3 & 3 & 3 & -1 \end{bmatrix}. \quad (6.4.1.12)$$

Zachodzi związek

$$\mathbf{L}_{s,(q-s)\times q} = \mathbf{L}_{1,(q-s)\times(q-s+1)} \mathbf{L}_{s-1,(q-s+1)\times q}. \quad (6.4.1.13)$$

Dla $\mathbf{f}^* \neq \mathbf{0}_q$ funkcja kary $J(\mathbf{f})$ stanowi miarę odstępstwa wektora \mathbf{f} od założonego *a priori* rozwiązania \mathbf{f}^* . Jeżeli $\mathbf{L} = \mathbf{I}_q$, odstępstwo mierzone jest normą euklidesową wektora różnicy $(\mathbf{f} - \mathbf{f}^*)$.

Można wykazać, że rozwiązaniem problemu (6.4.1.3) jest [26]

$$\hat{\mathbf{f}}_{\text{LSTich}} = (\mathbf{K}^T \mathbf{K} + \lambda \mathbf{L})^{-1} (\mathbf{K}^T \tilde{\mathbf{g}} + \lambda \mathbf{L} \mathbf{f}^*). \quad (6.4.1.14)$$

Należy zauważyć, że dla $\lambda = 0$ zregularyzowane zagadnienie odwrotne (6.4.1.3) przechodzi w wyjściowe słabo uwarunkowane zagadnienie wyznaczenia zwykłego rozwiązania najmniejszych kwadratów (6.2.31) układu równań (6.1.17). Fakt ten uzasadnia określanie rozwiązania zregularyzowanego jako rozwiązania najmniejszych kwadratów z karą. Wraz ze wzrostem parametru regularyzacji λ problem zregularyzowany coraz bardziej różni się od problemu oryginalnego, co sprawia, że rozwiązanie zregularyzowane $\hat{\mathbf{f}}_{\text{LSTich}}$ wykazuje coraz większe odstępstwo od prawdziwego rozwiązania najmniejszych kwadratów (6.2.31). Jednocześnie jednak zwiększanie parametru λ powoduje poprawę uwarunkowania numerycznego zagadnienia zregularyzowanego [1, 3, 4]. Wynika to bezpośrednio z następującego rozumowania. Zapisując wzór (6.4.1.1) w postaci

$$t_n = \frac{1}{1 + \left(\frac{\lambda}{\sigma_n} \right)^2}, \quad n=1, \dots, r \quad (6.4.1.15)$$

zauważa się, że współczynniki filtrujące t_n maleją wraz ze wzrostem λ w stopniu tym większym, im mniejsza jest odpowiadająca temu współczynnikowi wartość osobliwa σ_n . Oznacza to silniejsze tłumienie wyrazów rozwinięcia (6.4.1) o większych n , a co za tym idzie minimalizację wielkości błędu rozwiązania $\|\mathbf{r}\|^2$ przy zadanej wielkości szumu $\|\mathbf{e}\|^2$ w danych pomiarowych.

Wynika stąd, że parametr regularyzacji λ powinien być dobierany przez kompromis między zgodnością rozwiązania z danymi pomiarowymi w sensie wielkości $\|\tilde{\mathbf{g}} - \mathbf{Kf}\|_2^2$ a uwarunkowaniem zregularyzowanego zagadnienia odwrotnego [3].

6.4.2. REGULARYZACJA TICHONOWA Z OGRANICZENIAMI

Klasyczna regularyzacja Tichonowa nie uwzględnia dwóch ograniczeń fizycznych, które powinna spełniać szukana funkcja rozkładu wielkości cząstek fazy zdyspergowanej układu dyspersyjnego $f(a)$

- warunku nieujemności funkcji $f(a)$

$$f(a) \geq 0 \text{ dla każdego } a \in (0, \infty), \quad (6.4.2.1)$$

- warunku unormowania funkcji $f(a)$ do jedności

$$\int_0^{\infty} f(a) da = 1. \quad (6.4.2.2)$$

Dyskretnymi odpowiednikami tych ograniczeń są odpowiednio:

- dyskretny warunek nieujemności

$$\mathbf{f} \geq \mathbf{0}_q, \quad (6.4.2.3)$$

• dyskretny warunek unormowania do jedności, który uzyskuje się przez założenie, że funkcja $f(a)$ przyjmuje znaczne wartości jedynie w pewnym ograniczonym przedziale $[a_{\min}, a_{\max}]$, a poza nim jej wartości są zaniedbywalnie małe, co umożliwia zastąpienie występującej w równaniu (6.4.2.2) całki w przedziale $[0, \infty)$ za pomocą całki w skończonym przedziale $[a_{\min}, a_{\max}]$

$$\int_0^{\infty} f(a) da \approx \int_{a_{\min}}^{a_{\max}} f(a) da, \quad (6.4.2.4)$$

którą następnie można aproksymować za pomocą kwadratury numerycznej

$$\int_{a_{\min}}^{a_{\max}} f(a) da \approx \sum_{j=1}^q w_j f(a_j). \quad (6.4.2.5)$$

Po uwzględnieniu wyrażeń (6.4.2.4) i (6.4.2.5), dyskretny warunek unormowania do jedności przyjmuje postać

$$\mathbf{w}^T \mathbf{f} = 1, \quad (6.4.2.6)$$

gdzie

$$\mathbf{w} = [w_1, w_2, \dots, w_q]^T. \quad (6.4.2.7)$$

Nieuwzględnienie podanych warunków powoduje, że klasyczna regularyzacja Tichonowa nie zapewnia uzyskania rozwiązań mających sens fizyczny – nieujemnych i unormowanych do jedności [3].

Zregularyzowane rozwiązanie Tichonowa stanowi rozwiązanie problemu minimalizacji bez ograniczeń (6.4.1.3). Uwzględnienie w tym problemie ograniczeń (6.4.2.6) oraz (6.4.2.7) prowadzi do zagadnienia minimalizacji z ograniczeniami. Dla $\mathbf{f}^* = \mathbf{0}_q$ otrzymuje się następujące wyrażenie określające zregularyzowane rozwiązanie Tichonowa z ograniczeniami, zwane również rozwiązaniem najmniejszych kwadratów z karą i ograniczeniami:

$$\begin{aligned} \hat{\mathbf{f}}_{\text{LSTichOgr}} &= \arg \min_{\mathbf{f}} \{(\tilde{\mathbf{g}} - \mathbf{Kf})^T (\tilde{\mathbf{g}} - \mathbf{Kf}) + \lambda^2 (\mathbf{L}\mathbf{f})^T (\mathbf{L}\mathbf{f})\} \\ &= \arg \min_{\mathbf{f}} \{\tilde{\mathbf{g}}^T \tilde{\mathbf{g}} - 2\tilde{\mathbf{g}}^T \mathbf{Kf} + \mathbf{f}^T \mathbf{K}^T \mathbf{Kf} + \lambda^2 \mathbf{f}^T \mathbf{L}^T \mathbf{L}\mathbf{f}\} \\ &= \arg \min_{\mathbf{f}} \{\mathbf{f}^T (\mathbf{K}^T \mathbf{K} + \lambda^2 \mathbf{L}^T \mathbf{L})\mathbf{f} - 2\tilde{\mathbf{g}}^T \mathbf{Kf}\}, \end{aligned} \quad (6.4.2.8)$$

przy ograniczeniach (6.4.2.3) i (6.4.2.6). Oznaczając

$$\mathbf{A} = 2(\mathbf{K}^T \mathbf{K} + \lambda^2 \mathbf{L}^T \mathbf{L}), \quad (6.4.2.9)$$

$$\mathbf{b} = -2\mathbf{K}^T \tilde{\mathbf{g}}, \quad (6.4.2.10)$$

wyrażeniu (6.4.2.8) można nadać postać

$$\hat{\mathbf{f}}_{\text{LSTichOgr}} = \arg \min_{\mathbf{f}} \left\{ \frac{1}{2} \mathbf{f}^T \mathbf{A} \mathbf{f} + \mathbf{b}^T \mathbf{f} \right\}. \quad (6.4.2.11)$$

Z postaci wzoru (6.4.2.11) oraz ograniczeń (6.4.2.3) i (6.4.2.6) wynika, że rozważany problem minimalizacji z ograniczeni stanowi zagadnienie programowania kwadratowego [28, 29]. W związku z tym do wyznaczenia zregularyzowanego rozwiązania Tichonowa z ograniczeniami posłużyć się można standardowymi technikami programowania kwadratowego. W badaniach prowadzonych przez autorów zastosowano metodę aktywnego zbioru.

6.4.3. OBCIĘTY ROZKŁAD NA WARTOŚCI OSOBLIWE

Rozwiązanie oparte na obciętym rozkładzie na wartości osobliwe (ang. TSVD – *Truncated Singular Value Decomposition*) uzyskuje się przez zastosowanie współczynników filtrujących danych wzorem [1, 3, 4, 24, 30]

$$t_n = \begin{cases} 1, & n \leq l, \\ 0, & n > l, \end{cases} \quad (6.4.3.1)$$

gdzie $l \leq r$ jest ustaloną liczbą *a priori*.

Uwzględniając wzór (6.4.3.1) w wyrażeniu (6.4.1), uzyskuje się zregularyzowane rozwiązanie TSVD [1, 3, 4]

$$\hat{\mathbf{f}}_{TSVD} = \sum_{n=1}^l \frac{\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n. \quad (6.4.3.2)$$

Jak wynika ze wzoru (6.4.3.2), rozwiązanie TSVD powstaje przez obcięcie rozwinięcia (6.3.3), określającego rozwiązanie $\hat{\mathbf{f}}_{LSMN}$, na l -tym wyrazie. W ten sposób l pierwszych wyrazów o indeksach $n = 1, \dots, l$ pozostaje nie zmienionych, natomiast następne $(r - l)$ o indeksach $n = l + 1, \dots, r$ zostaje zaniebanych w rozwiązaniu $\hat{\mathbf{f}}_{TSVD}$ [1, 3, 4]. Zgodnie z wcześniejszymi rozważaniami powoduje to poprawę uwarunkowania rozwiązania kosztem niewielkiej utraty informacji zawartej w pominiętych wyrazach rozwinięcia (6.3.3) [3].

Obcięcie rozwinięcia (6.3.3) wiąże się z wygładzeniem zregularyzowanego rozwiązania $\hat{\mathbf{f}}_{TSVD}$ w stopniu tym większym, im mniejsza jest liczba l uwzględnianych w tym rozwinięciu wyrazów. Procedura obcięcia stanowi więc formę wiedzy apriorycznej dotyczącej gładkości rozwiązania, a liczba l ilościową postać tej informacji.

Wielkość l stanowi parametr regularyzacji dla rozwiązania TSVD. Podobnie, jak parametr λ w regularyzacji Tichonowa, wartość l powinna być dobierana za pomocą kompromisu pomiędzy zgodnością rozwiązania z danymi pomiarowymi w sensie wielkości $\|\tilde{\mathbf{g}} - \mathbf{Kf}\|_2^2$ a uwarunkowaniem zregularyzowanego zagadnienia odwrotnego. Zmniejszanie parametru l powoduje poprawę uwarunkowania kosztem pogorszenia zgodności rozwiązania z danymi pomiarowymi [3, 31].

6.4.4. DOBÓR PARAMETRU REGULARYZACJI

Istnieje wiele procedur doboru parametrów regularyzacji podejmujących próbę ustalenia kompromisu pomiędzy zgodnością rozwiązania z danymi pomiarowymi a uwarunkowaniem zregularyzowanego zagadnienia odwrotnego [1, 3, 4, 31, 32, 33].

Metody te stosowane są do poszukiwania optymalnych wartości zarówno parametrów zmieniających się w sposób ciągły, jak parametr λ w regularyzacji Tichonowa, jak i w sposób skokowy, jak będący liczbą naturalną parametr l w rozwiązaniu TSVD. W dalszych rozważaniach parametr regularyzacji oznaczany będzie ogólnie γ .

6.4.4.1. ZASADA ROZBIEŻNOŚCI

Zasada rozbieżności (ang. *discrepancy principle*) zakłada, że znana jest estymata błędu pomiarowego [1, 3, 4, 34]

$$\hat{\varepsilon} = \|\hat{\mathbf{e}}\|_2 = \|\tilde{\mathbf{g}} - \mathbf{g}\|_2, \quad (6.4.4.1.1)$$

Metoda polega na doborze takiego parametru regularyzacji γ , w którym euklidesowa norma wektora rezyduów odpowiadającego rozwiązaniu zregularyzowanemu uzyskanemu dla parametru $\gamma - \hat{\mathbf{f}}_{\text{Reg},\gamma}$ spełnia warunek [1, 4]

$$\|\tilde{\mathbf{g}} - \mathbf{K}\hat{\mathbf{f}}_{\text{Reg},\gamma}\|_2 = \hat{\varepsilon}. \quad (6.4.4.1.2)$$

Jeżeli kryterium (6.4.4.1.2) spełniane jest przez więcej niż jedną wartość γ , wybierana jest wartość γ odpowiadająca większemu stopniowi regularyzacji, ponieważ zapewnia ona lepsze uwarunkowanie zagadnienia przez silniejsze wygładzanie rozwiązania [1, 4].

W sytuacji gdy warunek (6.4.4.1.2) nie jest spełniony w sposób dokładny przez żadną z wartości γ , co typowe jest w przypadku γ będącego liczbą naturalną (np. w metodzie TSVD), wybierana jest wartość γ minimalizująca różnicę między lewą i prawą stroną równania (6.4.4.1.2) [1, 4].

6.4.4.2. KRYTERIUM KRZYWEJ L

Krzywą L (ang. *L-curve*) nazywa się wykres utworzony przez punkty o współrzędnych [1, 3, 4, 32]

$$(\log \|\tilde{\mathbf{g}} - \mathbf{K}\hat{\mathbf{f}}_{\text{Reg},\gamma}\|_2, \log \|\mathbf{L}\hat{\mathbf{f}}_{\text{Reg},\gamma}\|_2) \quad (6.4.4.2.1)$$

dla parametru regularyzacji γ zmieniającego się w sposób ciągły lub skokowy w określonym zakresie.

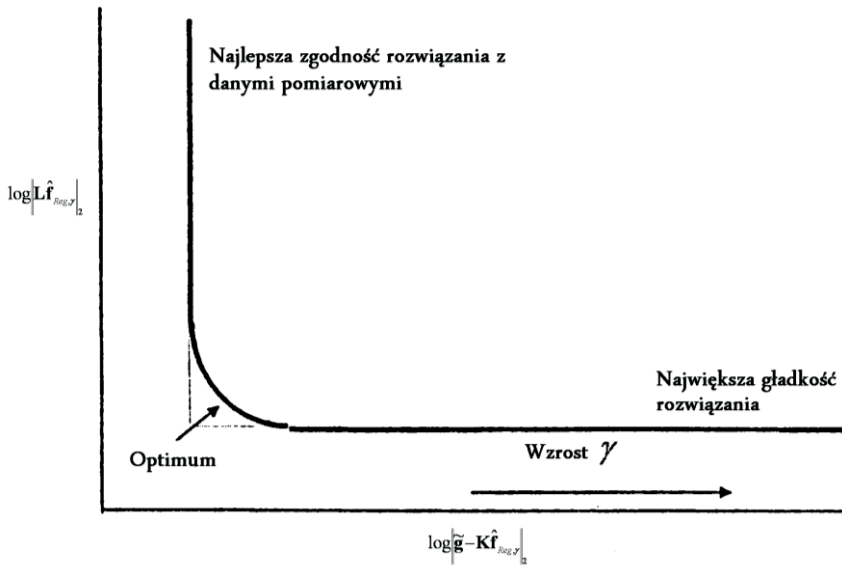
Typowy przebieg krzywej L zilustrowano na rysunku 6.4.4.2.1.

Nazwa wykresu wynika z tego, że typowo przybiera on kształt litery „L”. Fakt ten wyjaśnić można, opierając się na następującym rozumowaniu. Na podstawie wzoru (6.4.1) dla regularyzacji w postaci standardowej, czyli dla $\mathbf{L} = \mathbf{I}_q$, zachodzi związek [1]

$$\|\mathbf{L}\hat{\mathbf{f}}_{\text{Reg},\gamma}\|_2^2 = \|\hat{\mathbf{f}}_{\text{Reg},\gamma}\|_2^2 = \sum_{n=1}^r t_n^2 \frac{\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle^2}{\sigma_n^2}. \quad (6.4.4.2.2)$$

Ponadto

$$\mathbf{K}\hat{\mathbf{f}}_{\text{Reg},\gamma} = \mathbf{K} \sum_{n=1}^r t_n \frac{\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{v}_n = \sum_{n=1}^r t_n \frac{\langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle}{\sigma_n} \mathbf{K}\mathbf{v}_n. \quad (6.4.4.2.3)$$



Rys. 6.4.4.2.1. Typowa postać krzywej L [3]
 Fig. 6.4.4.2.1. Typical form of L -curve [3]

Uwzględniając zależność (6.2.26), otrzymuje się

$$\mathbf{K}\hat{\mathbf{f}}_{\text{Reg},\gamma} = \sum_{n=1}^r t_n \langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle \mathbf{u}_n. \quad (6.4.4.2.4)$$

Dla wektora $\tilde{\mathbf{g}}$ można natomiast przedstawić rozwinięcie w bazie ortonormalnej utworzonej przez kolumny macierzy \mathbf{U} pochodzącej z rozkładu macierzy \mathbf{K} na wartości osobliwe:

$$\tilde{\mathbf{g}} = \sum_{n=1}^p \langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle \mathbf{u}_n. \quad (6.4.4.2.5)$$

Biorąc pod uwagę wyrażenia (6.4.4.2.4) i (6.4.4.2.5), uzyskuje się:

$$\tilde{\mathbf{g}} - \mathbf{K}\hat{\mathbf{f}}_{\text{Reg},\gamma} = \sum_{n=1}^r (1-t_n) \langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle \mathbf{u}_n + \sum_{n=r+1}^p \langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle \mathbf{u}_n, \quad (6.4.4.2.6)$$

co ostatecznie daje:

$$\|\tilde{\mathbf{g}} - \mathbf{K}\hat{\mathbf{f}}_{\text{Reg},\gamma}\|_2^2 = \sum_{n=1}^r (1-t_n)^2 \langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle^2 + \sum_{n=r+1}^p \langle \tilde{\mathbf{g}}, \mathbf{u}_n \rangle^2. \quad (6.4.4.2.7)$$

W przypadku wartości parametru γ odpowiadających niewielkiemu stopniowi regularyzacji współczynniki filtrujące [1]:

$$t_n \approx 1, \quad n = 1, \dots, r. \quad (6.4.4.2.8)$$

W tym zakresie wartości współczynników t_n wyrażenia t_n^2 zmieniają się znacznie szybciej przy niewielkich zmianach t_n niż wyrażenia $(1-t_n)^2$. Zgodnie ze wzorami (6.4.4.2.2) i (6.4.4.2.7) niewielkie zmiany współczynników t_n wywołane zmianą parametru γ powodują zatem znacznie szybszą zmianę wartości $\log \|\tilde{\mathbf{g}} - \mathbf{K}\hat{\mathbf{f}}_{\text{Reg},\gamma}\|_2$ niż wartości $\log \|\mathbf{L}\hat{\mathbf{f}}_{\text{Reg},\gamma}\|_2$. Odpowiada to niemal pionowemu przebiegowi krzywej \mathbf{L} w zakresie wartości parametru γ wiążących się z niewielkim stopniem regularyzacji [1].

W przypadku wartości parametru γ odpowiadających znacznemu stopniowi regularyzacji współczynniki filtrujące [1]

$$t_n \approx 0, \quad n = 1, \dots, r. \quad (6.4.4.2.9)$$

W tym zakresie wartości współczynników t_n wyrażenia t_n^2 zmieniają się znacznie wolniej przy niewielkich zmianach t_n niż wyrażenia $(1-t_n)^2$. Zgodnie ze wzorami (6.4.4.2.2) i (6.4.4.2.7) zatem niewielkie zmiany współczynników t_n wywołane zmianą parametru γ powodują znacznie wolniejszą zmianę wartości $\log \|\tilde{\mathbf{g}} - \mathbf{K}\hat{\mathbf{f}}_{\text{Reg},\gamma}\|_2$ niż wartości $\log \|\mathbf{L}\hat{\mathbf{f}}_{\text{Reg},\gamma}\|_2$. Odpowiada to niemal poziomemu przebiegowi krzywej \mathbf{L} w zakresie wartości parametru γ wiążących się ze znacznym stopniem regularyzacji [1].

Metoda krzywej \mathbf{L} poszukuje optymalnej wartości parametru regularyzacji γ jako wartości γ odpowiadającej punktowi leżącemu w rogu krzywej \mathbf{L} , tzn. punktowi o maksymalnej krzywiznie [1, 3, 4, 32].

W przypadku parametru γ , będącego liczbą naturalną (np. w metodzie TSVD), otrzymuje się wykres złożony z izolowanych punktów, czyli tzw. dyskretną krzywą \mathbf{L} . W celu wyznaczenia optymalnej wartości parametru γ , znajduje się ciągłą krzywą \mathbf{L} dopasowaną do punktów dyskretnej krzywej \mathbf{L} , a następnie wyznacza róg krzywej ciągłej. Punkt dyskretnej krzywej \mathbf{L} najbliższy wyznaczonemu rogowi krzywej ciągłej uznaje się za róg dyskretnej krzywej \mathbf{L} , a odpowiadającą mu wartość γ za optymalną wartość parametru regularyzacji [4].

6.4.4.3. UOGÓLNIONA WALIDACJA WZAJEMNA

Metoda uogólnionej walidacji wzajemnej (ang. GCV – *Generalized Cross-Validation*) opiera się na spostrzeżeniu, że jeżeli usunięty zostałyby dowolnie wybrany element $\tilde{\mathbf{g}}_n$, $n = 1, \dots, p$, wektora danych pomiarowych $\tilde{\mathbf{g}}$, rozwiązanie zregularyzo-

wane $\hat{\mathbf{f}}_{\text{Reg},\gamma}$ powinno umożliwiać poprawne przewidzenie tego elementu [3, 4, 33]. Dodatkowo w podejściu GCV postuluje się, że wybór parametru regularyzacji ma być niezależny od transformacji ortogonalnych wektora $\tilde{\mathbf{g}}$ [4]. Prowadzi to do uznania za optymalną wartość parametru regularyzacji takiej wartości γ , która minimalizuje funkcję [1, 4]:

$$G(\gamma) = \frac{\|\tilde{\mathbf{g}} - \mathbf{K}\hat{\mathbf{f}}_{\text{Reg},\gamma}\|_2^2}{(\text{tr}(\mathbf{I}_p - \mathbf{K}\mathbf{K}_\gamma^\#))^2}, \quad (6.4.4.3.1)$$

gdzie: $\text{tr}(\mathbf{A})$ – operator wyznaczania śladu macierzy kwadratowej \mathbf{A} , $\mathbf{K}_\gamma^\#$ – dana wzorem (6.4.4) macierz regularyzacji dla parametru regularyzacji γ stosowana do obliczania rozwiązania zregularyzowanego $\hat{\mathbf{f}}_{\text{Reg},\gamma}$ na podstawie danych pomiarowych $\tilde{\mathbf{g}}$ zgodnie ze wzorem [1, 4]:

$$\hat{\mathbf{f}}_{\text{Reg},\gamma} = \mathbf{K}_\gamma^\# \tilde{\mathbf{g}}. \quad (6.4.4.3.2)$$

Wzór (6.4.4.3.1) używany może być dla parametrów regularyzacji zmieniających się zarówno w sposób ciągły, jak i skokowy [1, 4].

W kategoriach jakościowych uzasadnieniem przyjętej postaci (6.4.4.3.1) minimalizowanej funkcji $G(\gamma)$ jest fakt, że licznik stanowi kwadrat normy euklidesowej wektora rezyduów, będący miarą braku zgodności rozwiązania $\hat{\mathbf{f}}_{\text{Reg},\gamma}$ z danymi pomiarowymi $\tilde{\mathbf{g}}$, natomiast mianownik stanowi miarę stopnia regularyzacji zagadnienia – zwiększa się wraz ze wzrostem stopnia regularyzacji [1]. Minimalizacja funkcji $G(\gamma)$ odpowiada więc osiągnięciu jak najlepszej zgodności rozwiązania $\hat{\mathbf{f}}_{\text{Reg},\gamma}$ z danymi pomiarowymi $\tilde{\mathbf{g}}$ (minimalizacja licznika) z jednoczesnym jak najwyższym stopniem regularyzacji (maksymalizacja mianownika) [1]. Wyznaczana w ten sposób wartość parametru regularyzacji stanowi kompromis między zapewnieniem zgodnością rozwiązania z danymi pomiarowymi a poprawą uwarunkowania zregularyzowanego zagadnienia odwrotnego.

6.5. NIELINIOWE ITERACYJNE METODY INWERSYJNE

Omawiane dotychczas różne odmiany regularyzacji zagadnienia odwrotnego należą do liniowych technik inwersyjnych, ponieważ rozwiązanie w sposób liniowy zależy od wektora danych pomiarowych $\tilde{\mathbf{g}}$ [3, 35–38].

W nieliniowych iteracyjnych metodach inwersyjnych przyjmowane jest pewne rozwiązanie początkowe $\mathbf{f}^{(0)}$ oparte zwykle na posiadanej wiedzy apriorycznej o rozwiązanym problemie. Następnie rozwiązanie początkowe poddawane jest sukcesyw-

nym udoskonaleniom polegającym na mnożeniu tego rozwiązania w kolejnych iteracjach przez czynniki korekcyjne $c^{(1)}, c^{(2)}, \dots, c^{(k_{\text{iter}})}$, gdzie k_{iter} – liczba wszystkich iteracji [3, 35–38]. Zatem rozwiązanie w k -tej iteracji $\mathbf{f}^{(k)}$ uzyskiwane jest przez pomnożenie rozwiązania z iteracji $(k-1)$ -szej $\mathbf{f}^{(k-1)}$ przez obliczony w iteracji k -tej czynnik korekcyjny $c^{(k)}$ zgodnie z zależnością [3, 35–38]:

$$\mathbf{f}^{(k)} = c^{(k)} \mathbf{f}^{(k-1)}. \quad (6.5.1)$$

Czynniki korekcyjne $c^{(k)}$ wyznaczane są w poszczególnych iteracjach na podstawie danych pomiarowych $\tilde{\mathbf{g}}$, bieżącego rozwiązania $\mathbf{f}^{(k-1)}$ oraz stosowanej informacji apriorycznej. Procedura iteracyjna powinna być zbieżna, czyli rozwiązania otrzymywane w kolejnych iteracjach powinny cechować się coraz lepszą zgodnością z danymi pomiarowymi $\tilde{\mathbf{g}}$ oraz z założoną *a priori* charakterystyką. Rozwiązania $\mathbf{f}^{(k)}$ powinny dążyć do prawdziwego rozwiązania \mathbf{f} [3, 35–38].

Ze względu na multiplikatywny charakter iteracyjnych poprawek, przebiegających zgodnie ze wzorem (6.5.1) oraz zależnością czynników korekcyjnych $c^{(k)}$ od bieżącego rozwiązania, metody te określane są jako nieliniowe [3, 35–38].

W badaniach prowadzonych przez autorów stosowane były dwie nieliniowe procedury iteracyjne:

- metoda Twomeya,
- metoda Twomeya–Markowskiego.

6.5.1. METODA TWOMEYA

W metodzie Twomeya poprawki iteracyjne przebiegają według wzoru [3, 36–38]

$$f_j^{(k+1)} = (1 + r_i^{(k)} K(y_i, a_j)) f_j^{(k)} = (1 + r_i^{(k)} w_j^{-1} K_{ij}) f_j^{(k)}, \quad (6.5.1.1)$$

$$i = 1, \dots, p, \quad j = 1, \dots, q,$$

gdzie [3, 36–38]

$$r_i^{(k)} = \frac{\tilde{g}_i}{\sum_{j=1}^q K_{ij} f_j^{(k)}} - 1 = \frac{\tilde{g}_i}{\mathbf{k}_i \mathbf{f}^{(k)}} - 1, \quad (6.5.1.2)$$

przy czym \mathbf{k}_i oznacza i -ty wiersz macierzy \mathbf{K} .

Jak wynika z zależności (6.5.1.2), współczynniki $r_i^{(k)}$ są równe pomniejszonym o jeden stosunkom rzeczywistych wyników pomiarów \tilde{g}_i do odpowiadających im wartości teoretycznych g_i obliczanych jako wyjście modelu matematycznego (6.1.17) dla bieżącego rozwiązania $\mathbf{f}^{(k)}$ [3, 36–38]

$$\mathbf{g}_i = \mathbf{k}_i \mathbf{f}^{(k)}. \quad (6.5.1.3)$$

Na początku każdej iteracji na podstawie bieżącego rozwiązania $\mathbf{f}^{(k)}$ obliczane są współczynniki $r_i^{(k)}$ na podstawie wzoru (6.5.1.2) dla $i = 1, \dots, p$. Następnie wyznaczone współczynniki wykorzystywane są w danej iteracji do poprawek wartości $f_j^{(k)}$ przebiegających według wzoru (6.5.1.1) dla $i = 1, \dots, p, j = 1, \dots, q$ [36, 38].

Stosowane są różne kryteria zatrzymania dla procedury iteracyjnych poprawek w metodzie Twomeya. Najczęściej wykorzystuje się jeden z trzech warunków [3, 37, 38]:

1. Osiągnięcie zadowalającej zgodności rzeczywistych danych pomiarowych \tilde{g}_i z odpowiednimi wartościami teoretycznymi g_i obliczonymi dla bieżącego rozwiązania $\mathbf{f}^{(k)}$ w oparciu o wzór (6.5.1.3). Przez zadowalającą zgodność rozumie się sytuację, gdy różnice $(\tilde{g}_i - g_i)$ średnio są bliskie niepewnościom pomiarowym wyników \tilde{g}_i . Przyjmując za miarę tych niepewności odchylenia standardowe błędów pomiarowych wielkości \tilde{g}_i oznaczone σ_i , rozważany warunek można ściśle sformułować z zastosowaniem wartości statystyki chi-kwadrat [3, 37, 38]:

$$\hat{\chi}_{(k)}^2 = \frac{1}{p} \sum_{i=1}^p \left(\frac{\tilde{g}_i - \mathbf{k}_i \mathbf{f}^{(k)}}{\sigma_i} \right)^2. \quad (6.5.1.4)$$

Estymata $\hat{\chi}_{(k)}^2$ jest miarą średniej rozbieżności pomiędzy rzeczywistymi danymi pomiarowymi \tilde{g}_i a odpowiednimi wartościami teoretycznymi g_i odniesionej do niepewności pomiarowych. W związku z tym warunkiem zatrzymania procedury iteracyjnych poprawek jest osiągnięcie stanu, gdy [38]

$$\hat{\chi}_{(k)}^2 < 1. \quad (6.5.1.5)$$

2. Osiągnięcie stanu, gdy zmniejszenie wartości $\hat{\chi}_{(k)}^2$ w stosunku do wartości z poprzedniej iteracji $\hat{\chi}_{(k-1)}^2$ jest mniejsze od założonego poziomu [37]. Oznacza to, że kontynuowanie procedury iteracyjnych korekt rozwiązania nie powoduje już znaczącej poprawy zgodności rzeczywistych danych pomiarowych \tilde{g}_i z odpowiednimi wartościami teoretycznymi g_i . Inaczej mówiąc, algorytm osiągnął już rozwiązanie dostatecznie bliskie rozwiązaniu prawdziwemu [37]. Zwykle za warunek zatrzymania algorytmu przyjmuje się stan, gdy procentowe zmniejszenie wartości $\hat{\chi}_{(k)}^2$ w stosunku do wartości $\hat{\chi}_{(k-1)}^2$ jest mniejsze od ustalonej wielkości α , wynoszącej np. 5% [37]:

$$\frac{\hat{\chi}_{(k)}^2 - \hat{\chi}_{(k-1)}^2}{\hat{\chi}_{(k-1)}^2} \cdot 100 < \alpha. \quad (6.5.1.6)$$

3. Osiągnięcie stanu, gdy zmiana rozwiązania $\mathbf{f}^{(k)}$ w stosunku do rozwiązania z poprzedniej iteracji $\mathbf{f}^{(k-1)}$ jest mniejsza od założonego poziomu [2]. Oznacza to, że kolejne iteracje praktycznie nie spowodują już dalszej poprawy rozwiązania, gdyż algorytm osiągnął już rozwiązanie bliskie rozwiązaniu prawdziwemu [2]. Zwykle za warunek zatrzymania algorytmu przyjmuje się stan, gdy [2]:

$$\|\mathbf{f}^{(k)} - \mathbf{f}^{(k-1)}\|_2 < \delta, \quad (6.5.1.7)$$

gdzie δ – ustalona mała wartość dodatnia, np. 10^{-6} .

Słabe uwarunkowanie zagadnienia odwrotnego (6.1.17) powoduje znaczne spowolnienie lub nawet zanik zbieżności algorytmu Twomeya [2, 38]. Aby zapobiec wpadnięciu algorytmu w nieskończoną pętlę, często wprowadza się dodatkowy warunek stopu – procedura jest przerywana, gdy liczba iteracji przekroczy z góry ustaloną liczbę k_{\max} [38].

Złe uwarunkowanie problemu jest również przyczyną silnej zależności rozwiązania końcowego od apriorycznego rozwiązania początkowego $\mathbf{f}^{(0)}$ [2, 3]. Rozwiązanie ostateczne często ma przebieg bardzo zbliżony do przyjętego rozwiązania wejściowego $\mathbf{f}^{(0)}$. Od rozwiązania $\mathbf{f}^{(0)}$ zależy też szybkość zbieżności algorytmu Twomeya. Rozwiązanie wejściowe $\mathbf{f}^{(0)}$ stanowi więc formę wiedzy apriorycznej, umożliwiającej minimalizację efektów słabego uwarunkowania zagadnienia odwrotnego (6.1.17) [3].

Najważniejszym skutkiem złego uwarunkowania problemu są jednak tzw. artefakty w rozwiązaniu. Są to zaburzenia zwykle o charakterze gwałtownych wielkoczęstotliwościowych oscylacji, nie występujące w prawdziwym rozwiązaniu [2, 3, 38]. Stwierdzono, że przeprowadzenie zbyt dużej liczby iteracyjnych poprawek rozwiązania nie wpływa już praktycznie na poprawę zgodności rozwiązania z danymi pomiarowymi, natomiast znacząco wznaga oscylacje w rozwiązaniu [38]. W związku z tym warunek zatrzymania algorytmu powinien zapewniać kompromis pomiędzy zgodnością rozwiązania z danymi pomiarowymi a jego gładkością [38]. Warunek zatem zakończenia procedury stanowi drugą obok rozwiązania wejściowego $\mathbf{f}^{(0)}$ formę informacji apriorycznej [38].

6.5.2. METODA TWOMEYA–MARKOWSKIEGO

Środki, którymi posługuje się omawiana w poprzednim punkcie procedura Twomeya w celu uwzględnienia informacji apriorycznej o poszukiwanym rozwiązaniu, okazują się niewystarczające, by zapobiec skutkom słabego uwarunkowania zagadnienia odwrotnego. W rezultacie słabe uwarunkowanie problemu (6.1.17) prowadzi do znacznego spowolnienia lub zaniku zbieżności algorytmu, a uzyskiwane rozwiązania zakłócone są artefaktami w postaci silnych oscylacji wielkoczęstotliwościowych. W związku z tym opracowano wiele modyfikacji metody Twomeya [2, 3, 37, 38].

Udoskonalenie dokonane w ramach algorytmu Twomeya–Markowskiego polega na wprowadzeniu do metody Twomeya procedury wygładzania rozwiązań $\mathbf{f}^{(k)}$, co stanowi jawne uwzględnienie apriorycznego warunku gładkości rozwiązania [38].

W technice Twomeya–Markowskiego na początku aprioryczne rozwiązanie początkowe $\mathbf{f}^{(0)}$ przekazywane jest na wejście zwykłego algorytmu Twomeya w celu uzyskania rozwiązania, które trafia następnie na wejście pętli zwanej pętlą wygładzania i obliczeń Twomeya [38]. W pętli tej przeprowadzane jest iteracyjne udoskonalanie rozwiązania przez [38]:

- iteracyjną procedurę wygładzania bieżącego rozwiązania $\mathbf{f}^{(k)}$,
- realizację zwykłego algorytmu Twomeya, dla którego rozwiązanie wejściowe stanowi wygładzone rozwiązanie otrzymane w poprzednim kroku.

Wymienione dwa działania składają się na pojedynczy przebieg pętli wygładzania i obliczeń Twomeya.

Celem stosowania algorytmu Twomeya w ramach procedury Twomeya–Markowskiego jest poprawa zgodności rozwiązania z danymi pomiarowymi. Odbywa się ona kosztem zmniejszenia gładkości rozwiązania, czyli zwiększenia jego odstępstwa od założonego *a priori* charakteru [38]. Gładkość rozwiązania przywracana jest przez iteracyjną procedurę wygładzania rozwiązania. Wygładzanie prowadzi jednak do pogorszenia zgodności rozwiązania z danymi pomiarowymi, co ponownie wymaga użycia algorytmu Twomeya. Iteracyjne stosowanie naprzemiennie wygładzania oraz algorytmu Twomeya ma więc na celu uzyskanie rozwiązania o zadowalającej zgodności z danymi pomiarowymi oraz jednocześnie możliwie jak najbardziej gładkiego, czyli zgodnego z założoną apriorycznie charakterystyką [38]. Osiągnięcie kompromisu pomiędzy tymi dwiema właściwościami odbywa się przez odpowiedni dobór parametrów działania procedury wygładzania i algorytmu Twomeya [38].

Procedura wygładzania polega na iteracyjnym wykonywaniu operacji opisanej następującym wzorem [38]:

$$f_j := \begin{cases} \frac{3}{4}f_1 + \frac{1}{4}f_2 & \text{dla } j = 1, \\ \frac{1}{4}f_{j-1} + \frac{1}{2}f_j + \frac{1}{4}f_{j+1} & \text{dla } j = 2, \dots, q-1, \\ \frac{1}{4}f_{q-1} + \frac{3}{4}f_q & \text{dla } j = q, \end{cases} \quad (6.5.2.1)$$

gdzie $=$ – operator przypisania, czyli nadania nowej wartości. Operacja (6.5.2.1) polega na wyznaczeniu średniej ruchomej dla wektora \mathbf{f} .

Ponieważ każdy kolejny przebieg procedury wygładzania powoduje pogarszanie zgodności rozwiązania z danymi pomiarowymi, iteracyjne wygładzanie rozwiązania przerywane jest w momencie, gdy średnie różnice pomiędzy rzeczywistymi danymi pomiarowymi \tilde{g}_i oraz odpowiadającymi im teoretycznymi wartościami g_i obliczonymi dla bieżącego rozwiązania \mathbf{f} w oparciu o wzór (6.5.1.3), odniesione do odchyłek standardowych σ_i będą większe od założonego poziomu. W związku z tym za warunek przerwania iteracyjnej procedury wygładzania przyjęto [38]

$$\hat{\chi}^2 > \hat{\chi}_{\text{limit}}^2, \quad (6.5.2.2)$$

gdzie $\hat{\chi}_{\text{limit}}^2$ – założona *a priori* wartość graniczna. Często przyjmuje się $\hat{\chi}_{\text{limit}}^2 = 1$ [38].

Aby zapobiec możliwości wpadnięcia procedury wygładzania w nieskończoną pętlę, dodatkowo wprowadza się awaryjny warunek zatrzymania tej procedury. Jest ona przerywana po wykonaniu założonej liczby s_{max} przebiegów [38].

Iteracyjne udoskonalanie rozwiązania w pętli wygładzania i obliczeń Twomeya powoduje, że w kolejnych przebiegach pętli charakteryzuje się ono coraz większą gładkością przy zachowanej zgodności z rzeczywistymi danymi pomiarowymi [38]. Algorytm Twomeya–Markowskiego wykazuje więc zbieżność do rozwiązania zagadnienia odwrotnego (6.1.17), które jest najbardziej gładkie w sensie przyjętej miary.

Za miarę braku gładkości rozwiązania \mathbf{f} przyjęć można średnią wartość bezwzględną różnicy drugiego rzędu wektora \mathbf{f} [38]:

$$R = \frac{1}{q-2} \sum_{j=2}^{q-1} |f_{j+1} - 2f_j + f_{j-1}|. \quad (6.5.2.3)$$

Po dostatecznie dużej liczbie przebiegów pętli wygładzania i obliczeń Twomeya wartość R ulega już jedynie bardzo nieznacznemu zmniejszeniu w kolejnych przebiegach, a niekiedy może pozostawać stała lub nawet rosnać. W związku z tym za warunek przerwania pętli wygładzania i obliczeń Twomeya przyjmuje się najczęściej osiągnięcie momentu, gdy [38]

$$\frac{R^{(k)}}{R^{(k-2)}} > \mathcal{G}, \quad (6.5.2.4)$$

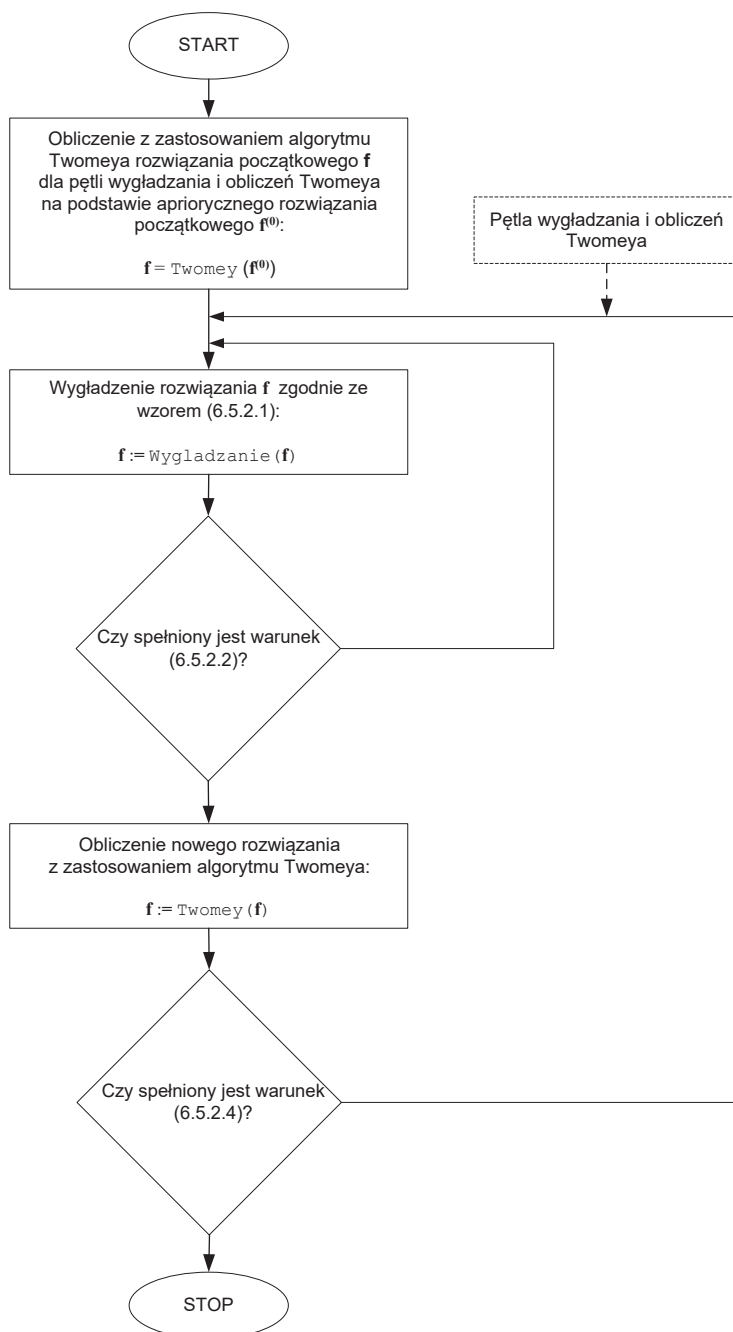
gdzie $R^{(k)}$ – wartość R po k -tej iteracji, \mathcal{G} – wartość graniczna ustalona *a priori* na podstawie wiedzy o rozwiązaniu. Zwykle przyjmuje się $\mathcal{G} = 0,95$ lub $\mathcal{G} = 0,975$.

Dla pętli wygładzania i obliczeń Twomeya wprowadza się dodatkowy awaryjny warunek zatrzymania. Pętla jest przerywana po wykonaniu założonej liczby k_{max} przebiegów [38].

Przebieg działań w algorytmie Twomeya–Markowskiego zaprezentowano na rysunku 6.5.2.1.

Dzięki zastosowaniu iteracyjnej procedury wygładzania algorytm Twomeya–Markowskiego wykazuje znacznie większą odporność na skutki słabego uwarunkowania zagadnienia odwrotnego niż algorytm Twomeya. Przejawia się to:

- znacznie szybszą zbieżnością procedury Twomeya–Markowskiego [38],
- mniejszym wpływem apriorycznie przyjętego rozwiązania początkowego $\mathbf{f}^{(0)}$ na ostateczne rozwiązanie [38],
- znacznie mniejszym nasileniem występowania artefaktów w rozwiązaniu w postaci wielkoczęstotliwościowych oscylacji [37, 38] w porównaniu z klasyczną procedurą Twomeya.



Rys. 6.5.2.1. Przebieg działań w algorytmie Twomeya–Markowskiego
Fig. 6.5.2.1. Workflow of the Twomey–Markowski algorithm

Poprawę odporności na efekty słabego uwarunkowania zagadnienia odwrotnego zawdzięcza się temu, że iteracyjna procedura wygładzania wprowadza do algorytmu aprioryczny warunek gładkości rozwiązania.

6.6. BAYESOWSKA METODA ROZWIĄZYWANIA
ZDYSKRETYZOWANEGO ZAGADNIENIA ODWROTNEGO
OPARTA NA PROCEDURZE MONTE CARLO

Wszystkie omawiane dotąd metody inwersyjne usiłowały rozwiązać zdyskretyzowane zagadnienie odwrotne (6.1.17) w sposób bezpośredni – przez poszukiwanie odpowiedniego operatora odwrotnego lub jego przybliżenia. Zupełnie inne podejście do rozwiązywania problemu odwrotnego prezentują metody bayesowskie [3, 39]. Rozwiązanie zagadnienia odwrotnego znajdowane jest przez analizę statystyczną powtarzanych wielokrotnie rozwiązań zagadnienia wprost, czyli zadania wyznaczenia wyjścia modelu (6.1.17) dla danego wektora \mathbf{f} . Analiza ta przeprowadzana jest na podstawie tzw. statystyki bayesowskiej [3, 39].

W podejściu bayesowskim dokonuje się jawnego założenia probabilistycznego *a priori* na temat szukanego wektora \mathbf{f} . Założenie to ma postać łącznej funkcji gęstości prawdopodobieństwa *a priori* wektora \mathbf{f} – $p_0(\mathbf{f})$ [3, 39]. Rzeczywiste dane pomiarowe $\tilde{\mathbf{g}}$ służą do skorygowania apriorycznej wiedzy o szukanym wektorze \mathbf{f} przez modyfikację łącznej funkcji gęstości prawdopodobieństwa wektora \mathbf{f} z zastosowaniem twierdzenia Bayesa [3, 39]. Zgodnie z tym twierdzeniem łączna warunkowa funkcja gęstości prawdopodobieństwa wektora \mathbf{f} z założeniem, że w rezultacie pomiaru zarejestrowane zostały dane pomiarowe $\tilde{\mathbf{g}}$, zwana łączną funkcją gęstości prawdopodobieństwa *a posteriori* wektora \mathbf{f} , dana jest wzorem [3, 39]

$$p(\mathbf{f}|\tilde{\mathbf{g}}) = \frac{p_0(\mathbf{f})p_L(\tilde{\mathbf{g}}|\mathbf{f})}{p(\tilde{\mathbf{g}})}, \quad (6.6.1)$$

gdzie: $p(\tilde{\mathbf{g}})$ – łączna funkcja gęstości prawdopodobieństwa wektora danych pomiarowych $\tilde{\mathbf{g}}$ (funkcja gęstości prawdopodobieństwa tego, że w toku pomiarów uzyskane zostaną wyniki $\tilde{\mathbf{g}}$), $p_L(\tilde{\mathbf{g}}|\mathbf{f})$ – łączna warunkowa funkcja gęstości prawdopodobieństwa wektora danych pomiarowych $\tilde{\mathbf{g}}$, przy założeniu, że prawdziwe rozwiązanie problemu odwrotnego stanowi wektor \mathbf{f} (funkcja gęstości prawdopodobieństwa tego, że w toku pomiarów uzyskane zostaną wyniki $\tilde{\mathbf{g}}$, przy założeniu, że prawdziwe rozwiązanie problemu odwrotnego stanowi wektor \mathbf{f}). Funkcja $p_L(\tilde{\mathbf{g}}|\mathbf{f})$ zwana jest również funkcją wiarygodności danych pomiarowych $\tilde{\mathbf{g}}$. Funkcja $p(\mathbf{f}|\tilde{\mathbf{g}})$ stanowi gęstość prawdopodobieństwa tego, że rozwiązaniem zagadnienia odwrotnego jest wektor \mathbf{f} z założeniem, że w rezultacie pomiarów uzyskano wyniki $\tilde{\mathbf{g}}$ [15]. Jest to więc proba-

bilistyczna wiedza o rozwiązaniu problemu \mathbf{f} wynikająca z wiedzy *a priori* o tym rozwiązaniu skorygowanej według wyników pomiarów $\tilde{\mathbf{g}}$ [3, 39].

W rzeczywistości podejście bayesowskie do rozwiązywania problemu odwrotnego realizowane jest z wykorzystaniem techniki Monte Carlo [3]. Metoda polega na wylosowaniu m wektorów \mathbf{f} oznaczonych \mathbf{f}_k , $k = 1, \dots, m$ z wielowymiarowego rozkładu prawdopodobieństwa opisanego łączną funkcją gęstości prawdopodobieństwa *a priori* $p_0(\mathbf{f})$. Dla każdego z wylosowanych wektorów \mathbf{f}_k , $k = 1, \dots, m$, obliczane jest wyjście modelu (6.1.17), czyli odpowiadający wektorowi \mathbf{f}_k wektor teoretycznych danych pomiarowych \mathbf{g}_k . Obliczenia te często określane są jako symulacja pomiarów. Każdemu wektorowi \mathbf{f}_k oraz odpowiadającemu mu wektorowi \mathbf{g}_k przypisać można prawdopodobieństwo masowe [3]

$$P(\mathbf{f}_k) = P(\mathbf{g}_k) = \frac{1}{m}, \quad k = 1, \dots, m \quad (6.6.2)$$

Ze wzoru (6.6.2) wynika, że

$$\sum_{k=1}^m P(\mathbf{f}_k) = \sum_{k=1}^m P(\mathbf{g}_k) = 1. \quad (6.6.3)$$

Ponadto zdarzenia polegające na uzyskiwaniu w danym przebiegu losowania określonego wektora \mathbf{f}_k i odpowiadającego mu wektora \mathbf{g}_k są zdarzeniami wzajemnie wykluczającymi się.

Dla każdego wektora \mathbf{g}_k , $k = 1, \dots, m$, obliczana jest łączna warunkowa funkcja gęstości prawdopodobieństwa uzyskania w rezultacie rzeczywistych pomiarów wektora danych pomiarowych $\tilde{\mathbf{g}}$, zakładając, że w efekcie symulacji pomiaru otrzymany został wektor $\mathbf{g}_k - p(\tilde{\mathbf{g}}|\mathbf{g}_k)$ [3]. Niech wektor błędów pomiarowych

$$\mathbf{e} = \tilde{\mathbf{g}} - \mathbf{g} \quad (6.6.4)$$

ma p -wymiarowy rozkład normalny

$$p_e(\mathbf{e}) = \left(\frac{1}{\sigma_e \sqrt{2\pi}} \right)^p \exp \left(-\frac{1}{2\sigma_e^2} \mathbf{e}^T \mathbf{e} \right), \quad (6.6.5)$$

odpowiadający założeniu, że elementy wektora \mathbf{e} stanowią wzajemnie nieskorelowane zmienne losowe o rozkładzie normalnym z zerową wartością oczekiwaną i jednakowym odchyleniem standardowym σ_e [3]. Wówczas [3]

$$p(\tilde{\mathbf{g}}|\mathbf{g}_k) = p_e(\tilde{\mathbf{g}} - \mathbf{g}_k) = \left(\frac{1}{\sigma_e \sqrt{2\pi}} \right)^p \exp \left(-\frac{1}{2\sigma_e^2} (\tilde{\mathbf{g}} - \mathbf{g}_k)^T (\tilde{\mathbf{g}} - \mathbf{g}_k) \right). \quad (6.6.6)$$

Łączna warunkowa funkcja gęstości prawdopodobieństwa wektora symulowanych danych pomiarowych \mathbf{g}_k z założeniem, że rzeczywiste dane pomiarowe stanowi wektor $\tilde{\mathbf{g}}$, czyli łączną funkcją gęstości prawdopodobieństwa *a posteriori* wektora \mathbf{g}_k , określona jest wzorem Bayesa [3]

$$p(\mathbf{g}_k|\tilde{\mathbf{g}}) = \frac{P(\mathbf{g}_k)p(\tilde{\mathbf{g}}|\mathbf{g}_k)}{P(\tilde{\mathbf{g}})}, \quad (6.6.7)$$

gdzie $P(\tilde{\mathbf{g}})$ – prawdopodobieństwo uzyskania w wyniku rzeczywistych pomiarów wektora danych pomiarowych $\tilde{\mathbf{g}}$. Ponieważ prawdopodobieństwa $P(\mathbf{g}_k)$ spełniają związek (6.6.3) oraz zdarzenia polegające na uzyskiwaniu w danym przebiegu losowania określonego wektora \mathbf{g}_k są zdarzeniami wzajemnie wykluczającymi się, prawdopodobieństwo $P(\tilde{\mathbf{g}})$ obliczyć można ze wzoru na prawdopodobieństwo całkowite [3]

$$P(\tilde{\mathbf{g}}) = \sum_{k=1}^m p(\tilde{\mathbf{g}}|\mathbf{g}_k)P(\mathbf{g}_k). \quad (6.6.8)$$

Po podstawieniu wyrażenia (6.6.8) do wzoru (6.6.7) otrzymuje się ostatecznie zależność [3]

$$p(\mathbf{g}_k|\tilde{\mathbf{g}}) = \frac{P(\mathbf{g}_k)p(\tilde{\mathbf{g}}|\mathbf{g}_k)}{\sum_{k=1}^m p(\tilde{\mathbf{g}}|\mathbf{g}_k)P(\mathbf{g}_k)}. \quad (6.6.9)$$

Uwzględniając związek (6.6.2), wzór (6.6.9) uprościć można do postaci [3]

$$p(\mathbf{g}_k|\tilde{\mathbf{g}}) = \frac{p(\tilde{\mathbf{g}}|\mathbf{g}_k)}{\sum_{k=1}^m p(\tilde{\mathbf{g}}|\mathbf{g}_k)}. \quad (6.6.10)$$

Pamiętając, że wektor \mathbf{g}_k odpowiada wektorowi \mathbf{f}_k , można zapisać [3]:

$$p(\mathbf{f}_k|\tilde{\mathbf{g}}) = p(\mathbf{g}_k|\tilde{\mathbf{g}}). \quad (6.6.11)$$

Ostatecznie więc [3]

$$p(\mathbf{f}_k|\tilde{\mathbf{g}}) = \frac{p(\tilde{\mathbf{g}}|\mathbf{g}_k)}{\sum_{k=1}^m p(\tilde{\mathbf{g}}|\mathbf{g}_k)}. \quad (6.6.12)$$

Funkcja $p(\mathbf{f}_k|\tilde{\mathbf{g}})$ stanowi łączną warunkową funkcję gęstości prawdopodobieństwa tego, że na wejście modelu przekazany został wektor \mathbf{f}_k , z założeniem, że w rezultacie

rzeczywistych pomiarów uzyskany został wektor danych pomiarowych $\tilde{\mathbf{g}}$, czyli łączną funkcją gęstości prawdopodobieństwa *a posteriori* wektora \mathbf{f}_k . Funkcja $p(\mathbf{f}_k|\tilde{\mathbf{g}})$ niesie probabilistyczną wiedzę o rozwiązaniu problemu \mathbf{f} , wynikającą z wiedzy *a priori* o tym rozwiązaniu skorygowanej według wyników pomiarów $\tilde{\mathbf{g}}$ [3].

Liczba symulacji m przeprowadzanych w praktyce jest bardzo wielka.

Znajomość funkcji $p(\mathbf{f}_k|\tilde{\mathbf{g}})$ umożliwia wyznaczenie różnych estymatorów szukanego wektora \mathbf{f} . Najczęściej stosowane są [3]:

- estymator największego prawdopodobieństwa *a posteriori*,
- estymator bayesowski o minimalnej wariancji.

Estymator największego prawdopodobieństwa *a posteriori* $\hat{\mathbf{f}}_{MAP}$ zdefiniowany jest jako moda łącznej funkcji gęstości prawdopodobieństwa *a posteriori* $p(\mathbf{f}_k|\tilde{\mathbf{g}})$, czyli [3]

$$\hat{\mathbf{f}}_{MAP} = \arg \max_{\mathbf{f}_k} \{p(\mathbf{f}_k|\tilde{\mathbf{g}})\}. \quad (6.6.13)$$

Estymator bayesowski o minimalnej wariancji $\hat{\mathbf{f}}_{MVBE}$ zdefiniowany jest jako warunkowa wartość oczekiwana wektora \mathbf{f} z założeniem, że w wyniku pomiarów uzyskany został wektor danych pomiarowych $\tilde{\mathbf{g}}$ [3]

$$\hat{\mathbf{f}}_{MVBE} = E(\mathbf{f}|\tilde{\mathbf{g}}) = \sum_{k=1}^m \mathbf{f}_k p(\mathbf{f}_k|\tilde{\mathbf{g}}). \quad (6.6.14)$$

Podane estymatory traktuje się jako rozwiązania problemu odwrotnego.

7. BADANIA SYMULACYJNE

W Katedrze Metrologii Elektronicznej i Fotonicznej Politechniki Wrocławskiej prowadzone są badania nad matematyczno-fizycznymi uwarunkowaniami zagadnienia odwrotnego w analizie wielkości cząstek fazy zdyspergowanej układów dyspersyjnych z wykorzystaniem rozpraszania światła. Przedmiotem badań są następujące problemy:

- wpływ na uwarunkowanie zagadnienia odwrotnego czynników takich, jak liczba i rozmieszczenie punktów pomiarowych, dokładność dyskretnej aproksymacji szukanego rozkładu wielkości cząstek, właściwości fizyczne układu dyspersyjnego, parametry fizyczne pomiaru oraz model matematyczny pomiarów,

- projektowanie procesu pomiarowego zapewniające maksymalną poprawę uwarunkowania zagadnienia,

- wpływ wspomnianych wcześniej czynników na jakość rozwiązania problemu odwrotnego,

- dobór optymalnych dla danych warunków technik inwersyjnych.

Rezultaty badań prezentują prace takie, jak [11] dotycząca zagadnienia odwrotnego w analizie wielkości cząstek układów dyspersyjnych w warunkach rozproszenia wielokrotnego, [40] przedstawiająca metodę momentów w analizie wielkości cząstek układów dyspersyjnych, [41] omawiająca zastosowanie transformat całkowitych w tej dziedzinie oraz [42] poruszająca problemy analizy wielkości cząstek z zastosowaniem rozpraszania światła pod niewielkim kątem.

W niniejszym punkcie omówione zostały wybrane wyniki badań symulacyjnych przeprowadzonych przez autorów.

7.1. CEL, ZAKRES I METODYKA BADAŃ

Prezentowane w pracy badania symulacyjne obejmowały zagadnienie odwrotne w pomiarach nefelometrycznych i turbidymetrycznych. W przypadku pomiarów nefelometrycznych rozważano dwa omawiane wcześniej modele matematyczne: model oparty na teorii rozpraszania Mie oraz model oparty na teorii dyfrakcji Fraunhofera. W przypadku pomiarów turbidymetrycznych wykorzystywany był wyłącznie model oparty na teorii rozpraszania Mie. W celu rozwiązywania problemu odwrotnego stosowano różne numeryczne techniki inwersyjne. W pracy omówiono trzy przedsięwzięcia badawcze:

- badanie porównawcze jakości rekonstrukcji funkcji rozkładu wielkości cząstek fazy zdyspergowanej układu dyspersyjnego na podstawie wyników pomiarów nefelometrycznych, przeprowadzonej na podstawie modelu matematycznego wynikającego z teorii rozpraszania Mie oraz na modelu matematycznym wynikającym z teorii dyfrakcji Fraunhofera z zastosowaniem różnych technik inwersyjnych,

- badanie porównawcze jakości rekonstrukcji funkcji rozkładu wielkości cząstek fazy zdyspergowanej układu dyspersyjnego na podstawie wyników pomiarów nefelometrycznych oraz wyników pomiarów turbidymetrycznych przeprowadzonej według modeli matematycznych wynikających z teorii rozpraszania Mie z zastosowaniem różnych technik inwersyjnych,

- badanie jakości rekonstrukcji funkcji rozkładu wielkości cząstek fazy zdyspergowanej układu dyspersyjnego na podstawie wyników pomiarów turbidymetrycznych, przeprowadzonej z zastosowaniem regularyzacji Tichonowa z ograniczeniami na tle innych metod inwersyjnych.

Wszystkie trzy przedsięwzięcia badawcze zrealizowano na podstawie podobnej metodyki. Badania przebiegały w dwóch etapach. W fazie pierwszej w toku symulacji pomiarów dla zadanej testowej funkcji rozkładu wielkości cząstek generowane były dane pomiarowe określonego rodzaju według wybranego modelu matematycznego. Dane te poddawano następnie wpływowi addytywnego, stacjonarnego, nieskorelowanego szumu gaussowskiego o zerowej wartości oczekiwanej. W drugiej fazie badań dla wygenerowanych danych pomiarowych z zastosowaniem wybranych technik inwersyjnych rozwiązywano zagadnienie odwrotne. Wykorzystywano w tym

celu ten sam model matematyczny, który użyty został w fazie pierwszej do generacji danych.

7.1.1. BADANIE WPŁYWU MODELU MATEMATYCZNEGO POMIARÓW
NA JAKOŚĆ REKONSTRUKCJI FUNKCJI ROZKŁADU WIELKOŚCI CZĄSTEK
WEDŁUG WYNIKÓW POMIARÓW NEFELOMETRYCZNYCH

Celem badań było określenie i porównanie jakości rekonstrukcji funkcji rozkładu wielkości cząstek fazy zdyspergowanej dokonanej na podstawie dwóch modeli matematycznych pomiarów nefelometrycznych – model wynikający z teorii rozpraszania Mie oraz model wynikający z teorii dyfrakcji Fraunhofera z zastosowaniem wybranych technik inwersyjnych [43].

Stanowiące pierwszą fazę badań symulacje pomiarów nefelometrycznych realizowane były dla następujących warunków:

- ciągu wartości promienia cząstek fazy zdyspergowanej

$$\mathbf{a} = [a_1 \quad a_2 \quad \dots \quad a_q]^T, \quad (7.1.1.1)$$

gdzie: $q = 200$, a a_j określają wzory (6.1.3) i (6.1.4),

- ciągu $p = 181$ wartości kąta rozpraszania:

$$\theta_i = (i-1) \cdot 1^\circ, \quad i = 1, \dots, p, \quad (7.1.1.2)$$

- liczby cząstek fazy zdyspergowanej w jednostce objętości układu dyspersyjnego $N_v = 10^{15} \text{ m}^{-3}$,

- długości fali światła rozpraszanego $\lambda = 0,6328 \mu\text{m}$,

- zespolonego współczynnika załamania cząstek fazy zdyspergowanej $N_1 = 1,55 + i \cdot 0,00$,

- zespolonego współczynnika załamania ośrodka dyspersyjnego $N_0 = 1,00 + i \cdot 0,00$,

- testowej funkcji rozkładu wielkości cząstek fazy zdyspergowanej danej wzorem

$$\begin{aligned} f_{\text{test}}(a) = & 0,33 f_{\text{norm}}(a, 0,45 \mu\text{m}, 0,15 \mu\text{m}) \\ & + 0,17 f_{\text{norm}}(a, 0,6 \mu\text{m}, 0,1 \mu\text{m}) \\ & + 0,5 f_{\text{lognorm}}(a, 0,1, 0,5), \end{aligned} \quad (7.1.1.3)$$

gdzie: $f_{\text{norm}}(a, \mu_{\text{norm}}, \sigma_{\text{norm}})$ – funkcja gęstości prawdopodobieństwa rozkładu normalnego zmiennej a o wartości oczekiwanej μ_{norm} i odchyleniu standardowym σ_{norm} , $f_{\text{lognorm}}(a, \mu_{\text{lognorm}}, \sigma_{\text{lognorm}})$ – funkcja gęstości prawdopodobieństwa rozkładu lognormalnego zmiennej a o parametrach μ_{lognorm} i σ_{lognorm} . Dyskretną reprezentację funkcji $f_{\text{test}}(a)$ stanowił wektor

$$\mathbf{f}_{\text{test}} = f_{\text{test}}(\mathbf{a}) = [f_{\text{test}}(a_1) \quad f_{\text{test}}(a_2) \quad \dots \quad f_{\text{test}}(a_q)]^T. \quad (7.1.1.4)$$

Symulacje przeprowadzone zostały na podstawie dwóch rozważanych modeli matematycznych:

- modelu opartego na teorii rozpraszania Mie,
- modelu opartego na teorii dyfrakcji Fraunhofera.

Rezultatem symulacji były dwa wektory symulowanych danych pomiarowych:

- $\boldsymbol{\beta}_{\text{Mie}}$ odpowiadający zastosowaniu modelu opartego na teorii rozpraszania Mie,
- $\boldsymbol{\beta}_{\text{Fraunhofer}}$ odpowiadający zastosowaniu modelu opartego na teorii dyfrakcji Fraunhofera.

Fraunhofera.

Wektory $\boldsymbol{\beta}_{\text{Mie}}$ i $\boldsymbol{\beta}_{\text{Fraunhofer}}$ poddano następnie zakłóceniu przez addytywny stacjonarny i nieskorelowany szum gaussowski o zerowej wartości oczekiwanej, reprezentowany przez wektor losowy $\boldsymbol{\varepsilon}$, którego elementy stanowią nieskorelowane zmienne losowe o rozkładzie normalnym z zerową wartością oczekiwaną i jednakowym odchyleniem standardowym $\sigma_{\varepsilon} = 1\% \max(\boldsymbol{\beta})$, gdzie $\boldsymbol{\beta}$ oznacza $\boldsymbol{\beta}_{\text{Mie}}$ lub $\boldsymbol{\beta}_{\text{Fraunhofer}}$. W ten sposób uzyskano odpowiednie wektory zakłócone szumem $\tilde{\boldsymbol{\beta}}_{\text{Mie}}$ oraz $\tilde{\boldsymbol{\beta}}_{\text{Fraunhofer}}$.

Ostatecznie symulacja oraz zaszumianie dostarczyły czterech wektorów danych pomiarowych $\boldsymbol{\beta}_{\text{Mie}}$, $\boldsymbol{\beta}_{\text{Fraunhofer}}$ oraz $\tilde{\boldsymbol{\beta}}_{\text{Mie}}$ i $\tilde{\boldsymbol{\beta}}_{\text{Fraunhofer}}$.

Druga faza badań obejmowała rozwiązywanie zagadnienia odwrotnego w pomiarach nefelometrycznych. Dla każdego z wektorów: $\boldsymbol{\beta}_{\text{Mie}}$, $\boldsymbol{\beta}_{\text{Fraunhofer}}$ oraz $\tilde{\boldsymbol{\beta}}_{\text{Mie}}$ i $\tilde{\boldsymbol{\beta}}_{\text{Fraunhofer}}$ wyznaczany był z zastosowaniem wybranych technik inwersyjnych wektor $\hat{\mathbf{f}}$ reprezentujący wartości rekonstruowanej funkcji rozkładu wielkości cząstek fazy zdyspergowanej $\hat{f}(a)$ w dokładnie tych samych punktach, w których wektor \mathbf{f}_{test} reprezentował wartości funkcji $f_{\text{test}}(a)$, czyli:

$$\hat{\mathbf{f}} = \hat{f}(\mathbf{a}) = [\hat{f}(a_1) \quad \hat{f}(a_2) \quad \dots \quad \hat{f}(a_q)]^T. \quad (7.1.1.5)$$

Podczas rozwiązywania zagadnienia odwrotnego dla każdego wektora danych pomiarowych stosowany był ten sam model matematyczny, który był używany w czasie generacji tego wektora w ramach symulacji.

Wykorzystane zostały następujące metody inwersyjne:

• regularyzacja Tichonowa z minimalizacją pięciu różnych miar braku gładkości poszukiwanego rozwiązania:

- normy euklidesowej wektora $\hat{\mathbf{f}}$,
- normy euklidesowej wektora różnic pierwszego, drugiego i trzeciego rzędu wektora $\hat{\mathbf{f}}$,
- normy euklidesowej różnicy wektora $\hat{\mathbf{f}}$ i wektora przyjętego *a priori* rozwiązania próbnego:

$$\mathbf{f}^* = [1 \ 1 \ \dots \ 1]^T \mu\text{m}^{-1}. \quad (7.1.1.6)$$

- metoda obciętego rozkładu na wartości osobliwe (TSVD).

Wartości parametru regularyzacji dla każdego procesu rozwiązywania zagadnienia odwrotnego dobierane były indywidualnie tak, by zapewnić jak najlepszą dokładność rekonstrukcji w danym przypadku.

Następnie dla każdego z uzyskanych wektorów $\hat{\mathbf{f}}$ określano miarę jego odstępstwa od prawdziwego rozwiązania problemu odwrotnego – wektora \mathbf{f}_{test} . Za miarę tę przyjęto wielkość

$$\Delta = \sqrt{\frac{\|\hat{\mathbf{f}} - \mathbf{f}_{\text{test}}\|_2^2}{q}}. \quad (7.1.1.7)$$

7.1.2. BADANIE PORÓWNAWCZE JAKOŚCI REKONSTRUKCJI FUNKCJI ROZKŁADU WIELKOŚCI CZĄSTEK NA PODSTAWIE WYNIKÓW POMIARÓW NEFELOMETRYCZNYCH I TURBIDYMETRYCZNYCH

Celem badań było określenie i porównanie jakości rekonstrukcji funkcji rozkładu wielkości cząstek fazy zdyspergowanej dokonanej na podstawie wyników pomiarów nefelometrycznych oraz wyników pomiarów turbidymetrycznych na podstawie modeli matematycznych pomiarów wynikających z teorii rozpraszania Mie z zastosowaniem wybranych technik inwersyjnych [44].

Symulacje pomiarów nefelometrycznych i turbidymetrycznych w pierwszej fazie badań przeprowadzono dla tych samych warunków:

- ciągu wartości promienia cząstek fazy zdyspergowanej, jak w punkcie 7.1.1,
- liczby cząstek fazy zdyspergowanej w jednostce objętości układu dyspersyjnego $N_v = 10^{15} \text{ m}^{-3}$,
- zespolonego współczynnika załamania cząstek fazy zdyspergowanej $N_1 = 1,55 + i \times 0,00$,
- zespolonego współczynnika załamania ośrodka dyspersyjnego $N_0 = 1,00 + i \times 0,00$,
- testowej funkcji rozkładu wielkości cząstek fazy zdyspergowanej danej wzorem (7.1.1.3) i reprezentowanej przez wektor \mathbf{f}_{test} określony wzorem (7.1.1.4).

Symulacja pomiarów nefelometrycznych realizowana była na podstawie modelu matematycznego wynikającego z teorii rozpraszania Mie dla ciągu wartości kąta rozpraszania, jak w punkcie 7.1.1, przy długości fali światła rozpraszanego $\lambda = 0,6328 \mu\text{m}$. W rezultacie symulacji uzyskano wektor danych pomiarowych β .

Symulacja pomiarów turbidymetrycznych prowadzona była według modelu matematycznego wynikającego z teorii rozpraszania Mie dla ciągu $p = 161$ długości fali:

$$\lambda_i = [0,25 + (i-1)0,003125] \mu\text{m}, \quad i = 1, \dots, p, \quad (7.1.2.1)$$

obejmującego zakres od 0,25 do 0,75 μm . W rezultacie symulacji uzyskano wektor danych pomiarowych \mathbf{c} .

Wektory β i \mathbf{c} poddano następnie zakłóceniu przez addytywny stacjonarny i nieskorelowany szum gaussowski o zerowej wartości oczekiwanej, reprezentowany przez wektor losowy $\boldsymbol{\varepsilon}$, którego elementy stanowią nieskorelowane zmienne losowe o rozkładzie normalnym z zerową wartością oczekiwaną i jednakowym odchyleniem standardowym $\sigma_{\varepsilon} = 1\% \times \max(\mathbf{g})$, gdzie \mathbf{g} oznacza β lub \mathbf{c} . W ten sposób uzyskano odpowiednio wektory zakłócone szumem $\tilde{\beta}$ oraz $\tilde{\mathbf{c}}$.

Ostatecznie symulacja oraz zaszumianie dostarczyły czterech wektorów danych pomiarowych β , \mathbf{c} oraz $\tilde{\beta}$ i $\tilde{\mathbf{c}}$.

Druga faza badań obejmowała rozwiązywanie zagadnienia odwrotnego w pomiarach nefelometrycznych i turbidymetrycznych. Dla każdego z wektorów β , \mathbf{c} oraz $\tilde{\beta}$ i $\tilde{\mathbf{c}}$ wyznaczany był z zastosowaniem wybranych technik inwersyjnych wektor $\hat{\mathbf{f}}$ reprezentujący wartości rekonstruowanej funkcji rozkładu wielkości cząstek fazy zdyspergowanej $\hat{f}(a)$ w dokładnie tych samych punktach, w których wektor \mathbf{f}_{test} reprezentował wartości funkcji $f_{\text{test}}(a)$ zgodnie ze wzorem (7.1.1.5).

Podczas rozwiązywania zagadnień odwrotnych stosowane były te same metody inwersyjne, opisane w badaniach w punkcie 7.1.1. Jak poprzednio, wartości parametru regularyzacji dla każdego procesu rozwiązywania zagadnienia odwrotnego dobierane były indywidualnie tak, by zapewnić jak najlepszą dokładność rekonstrukcji w danym przypadku.

Następnie dla każdego z uzyskanych wektorów $\hat{\mathbf{f}}$ określano zdefiniowaną wzorem (7.1.1.7) miarę jego odstępstwa od prawdziwego rozwiązania problemu odwrotnego – wektora \mathbf{f}_{test} .

7.1.3. BADANIE JAKOŚCI REKONSTRUKCJI FUNKCJI ROZKŁADU WIELKOŚCI CZĄSTEK Z ZASTOSOWANIEM REGULARYZACJI TICHONOWA Z OGRANICZENIAMI

Cel badań stanowiło określenie jakości rekonstrukcji funkcji rozkładu wielkości cząstek fazy zdyspergowanej układu dyspersyjnego na podstawie wyników pomiarów turbidymetrycznych przeprowadzonej z zastosowaniem regularyzacji Tichonowa z ograniczeniami na podstawie modelu matematycznego wynikającego z teorii rozpraszania Mie. Dokładność rekonstrukcji oceniana była na tle innych metod inwersyjnych nie nakładających na szukaną funkcję rozkładu żadnych ograniczeń fizycznych.

W pierwszej fazie badań wygenerowane zostały testowe dane pomiarowe w toku symulacji pomiarów turbidymetrycznych dla następujących warunków:

- układu dyspersyjnego stanowiącego zawiesinę cząstek lateksu (faza zdyspergowana) o kształcie kulistym w wodzie (ośrodek dyspersyjny); zespolony współczynnik

załamania lateksu w funkcji długości fali światła obliczano na podstawie wzoru danego w [45]; zespolony współczynnik załamania wody w funkcji długości fali światła obliczano na podstawie wzoru podanego w [46] dla następujących warunków: temperatury $T = 293,15$ K (20 °C), ciśnienia $p = 101,325$ kPa,

- ciągu $q = 4000$ promieni cząstek fazy zdyspergowanej \mathbf{a} rozłożonych zgodnie z wzorami (6.1.3) i (6.1.4) w przedziale $[a_{\min}, a_{\max}] = [300 \text{ nm}, 700 \text{ nm}]$,

- ciągu $p = 17$ długości fali światła rozpraszane λ określonego wzorem

$$\lambda_i = [300 + (i - 1) \cdot 25] \text{ nm}, \quad i = 1, \dots, p, \quad (7.1.3.1)$$

obejmującego przedział $[300 \text{ nm}, 700 \text{ nm}]$,

- testowej funkcji rozkładu wielkości cząstek fazy zdyspergowanej danej wzorem

$$\begin{aligned} f_{\text{test}}(a) &= 0,57 f_{\text{norm}}(a, 500 \text{ nm}, 75 \text{ nm}) \\ &+ 0,03 f_{\text{norm}}(a, 650 \text{ nm}, 15 \text{ nm}) \\ &+ 0,4 f_{\text{lognorm}}(a, 6, 15, 0, 1). \end{aligned} \quad (7.1.3.2)$$

Dla uproszczenia w obu fazach badań – przy generacji testowych danych pomiarowych oraz podczas rozwiązywania zagadnienia odwrotnego dla danych testowych przyjęto założenie, że liczba cząstek w jednostce objętości układu dyspersyjnego $N_v = 1 \text{ m}^{-3}$. Dzięki temu, że założenie to zostało poczynione na obu etapach badań, arbitralnie wybrana wartość $N_v = 1 \text{ m}^{-3}$ nie miała wpływu na wynik – rekonstruowaną funkcję rozkładu wielkości cząstek.

W rezultacie symulacji uzyskano wektor danych pomiarowych \mathbf{c} , który poddano następnie zakłóceniu przez addytywny stacjonarny i nieskorelowany szum gaussowski o zerowej wartości oczekiwanej, reprezentowany przez wektor losowy $\boldsymbol{\varepsilon}_p$, którego elementy stanowią nieskorelowane zmienne losowe o rozkładzie normalnym z zerową wartością oczekiwaną i jednakowym odchyleniem danym wzorem

$$\sigma_{\varepsilon,p} = \frac{p}{100} \max(\mathbf{c}), \quad (7.1.3.3)$$

gdzie p – stała określająca wielkość sztucznych błędów pomiarowych.

Zakłócenie wektora \mathbf{c} przeprowadzono dla następujących wartości parametru p :

- $p = 0,0\%$,
- $p = 2,5\%$,
- $p = 5,0\%$,
- $p = 7,5\%$.

Dla każdej wartości parametru p generowanych było $N_{\text{symul}} = 100$ realizacji wektora błędów pomiarowych $\boldsymbol{\varepsilon}_p$. Ostatecznie symulacja pomiarów oraz zaszumianie dostarczyły $N_{\text{symul}} = 100$ realizacji czterech wektorów danych pomiarowych \mathbf{c}_p : $\mathbf{c}_{0,0\%}$, $\mathbf{c}_{2,5\%}$, $\mathbf{c}_{5,0\%}$ oraz $\mathbf{c}_{7,5\%}$ oznaczonych $\mathbf{c}_{p,k}$, $k = 1, \dots, N_{\text{symul}}$.

W drugiej fazie badań dla każdego z wektorów $\mathbf{c}_{p,k}$ obliczany był przez rozwiązanie zagadnienia odwrotnego wektor $\hat{\mathbf{f}}_{p,k}$ reprezentujący wartości rekonstruowanej funkcji rozkładu wielkości cząstek fazy zdyspergowanej $\hat{f}_{p,k}(a)$ w punktach $a_{\text{inwer},j}$ tworzących ciąg $q_{\text{inwer}} = 400$ promieni cząstek fazy zdyspergowanej $\mathbf{a}_{\text{inwer}}$ rozłożonych zgodnie ze wzorami (6.1.3) i (6.1.4) w przedziale $[a_{\text{min}}, a_{\text{max}}] = [330 \text{ nm}, 700 \text{ nm}]$. Następnie dla każdej wartości parametru p obliczane były dwa wektory:

- średnia arytmetyczna wektorów $\hat{\mathbf{f}}_{p,k}$ dana wzorem

$$\bar{\mathbf{f}}_p = \frac{1}{N_{\text{symul}}} \sum_{k=1}^{N_{\text{symul}}} \hat{\mathbf{f}}_{p,k}. \quad (7.1.3.4)$$

- wektor odchyłeń standardowych odpowiadających sobie elementów wektorów $\hat{\mathbf{f}}_{p,k}$ dany wzorem:

$$\boldsymbol{\sigma}_{\hat{\mathbf{f}},p} = [\sigma_{\hat{\mathbf{f}},p,1} \quad \sigma_{\hat{\mathbf{f}},p,2} \quad \dots \quad \sigma_{\hat{\mathbf{f}},p,q_{\text{inwer}}}]^T, \quad (7.1.3.5)$$

gdzie

$$\sigma_{\hat{\mathbf{f}},p,j} = \left[\frac{1}{N_{\text{symul}} - 1} \sum_{k=1}^{N_{\text{symul}}} (\hat{f}_{p,k,j} - \bar{f}_{p,j})^2 \right]^{1/2}, \quad j=1, \dots, q_{\text{inwer}}. \quad (7.1.3.6)$$

Na podstawie wektora $\boldsymbol{\sigma}_{\hat{\mathbf{f}},p}$ wyznaczano syntetyczną miarę niepewności rozwiązania $\bar{\mathbf{f}}_p$ zdefiniowaną wyrażeniem

$$\Sigma_p = \left[\frac{1}{q_{\text{inwer}}} \sum_{k=1}^{q_{\text{inwer}}} \sigma_{\hat{\mathbf{f}},p,j}^2 \right]^{1/2}. \quad (7.1.3.7)$$

Ponadto obliczano syntetyczną miarę odchylenia zrekonstruowanej funkcji rozkładu wielkości cząstek od testowej funkcji rozkładu wielkości cząstek

$$\Delta_p = \left\{ \frac{1}{q_{\text{inwer}}} \sum_{j=1}^{q_{\text{inwer}}} [\bar{f}_{p,j} - f_{\text{test}}(a_{\text{inwer},j})]^2 \right\}^{1/2}. \quad (7.1.3.8)$$

Dla celów porównawczych obok zregulowanego rozwiązania Tichonowa z ograniczeniami wyznaczane były zwykłe zregulowane rozwiązania Tichonowa. Oba rodzaje rozwiązań obliczane były w zastosowaniu różnych miar braku gładkości szukanego rozwiązania:

- normy euklidesowej wektora $\hat{\mathbf{f}}$,
- normy euklidesowej wektora różnic pierwszego rzędu wektora $\hat{\mathbf{f}}$,

- normy euklidesowej wektora różnic drugiego rzędu wektora $\hat{\mathbf{f}}$,
- normy euklidesowej wektora różnic trzeciego rzędu wektora $\hat{\mathbf{f}}$.

Dla każdego wektora $\hat{\mathbf{f}}_{p,k}$ oraz miary braku gładkości szukanego rozwiązania optymalna wartość parametru regularyzacji wyznaczana była jednokrotnie – wyłącznie dla zagadnienia zwykłej regularyzacji Tichonowa – z użyciem omawianej wcześniej metody krzywej **L**. Ta sama wartość parametru regularyzacji była następnie wykorzystywana w odpowiednim zagadnieniu regularyzacji Tichonowa z ograniczeniami.

7.2. WYNIKI BADAŃ

7.2.1. WYNIKI BADAŃ WPŁYWU MODELU MATEMATYCZNEGO POMIARÓW NA JAKOŚĆ REKONSTRUKCJI FUNKCJI ROZKŁADU WIELKOŚCI CZĄSTEK NA PODSTAWIE WYNIKÓW POMIARÓW NEFELOMETRYCZNYCH

Wyniki badań symulacyjnych jakości rekonstrukcji funkcji rozkładu wielkości cząstek fazy zdyspergowanej dokonanej na podstawie dwóch modeli matematycznych pomiarów nefelometrycznych – model wynikający z teorii rozpraszania Mie oraz model wynikający z teorii dyfrakcji Fraunhofera z zastosowaniem wybranych procedur inwersyjnych zaprezentowano w tabeli 7.2.1.1 [43].

Tabela 7.2.1.1. Zestawienie wartości parametru $\Delta[\mu\text{m}^{-1}]$ charakteryzującego dokładność rekonstrukcji funkcji rozkładu wielkości cząstek fazy zdyspergowanej według dwóch badanych modeli matematycznych – model wynikający z teorii rozpraszania Mie oraz model wynikający z teorii dyfrakcji Fraunhofera dla wyników symulowanych pomiarów nefelometrycznych zakłóconych i niezakłóconych szumem

Table 7.2.1.1. Comparison of the values of the parameter $\Delta[\mu\text{m}^{-1}]$ characterizing the accuracy of reconstruction of particle size distribution of dispersed phase on the basis of two mathematical models under studies – model following from Mie scattering theory and model following from Fraunhofer diffraction theory for the results of simulated nephelometric measurements corrupted and not corrupted by noise

Model matematyczny zaszumienie danych	Mie, niezaszumione	Mie, zaszumione	Fraunhofer, niezaszumione	Fraunhofer, zaszumione
Procedura inwersyjna				
Tichonow, norm.	0,068938	0,34742	0,011602	0,11868
Tichonow, I rz.	0,074651	0,18218	0,065687	0,14748
Tichonow, II rz.	0,010262	0,13346	0,021501	0,15706
Tichonow, III rz.	0,00043019	0,20925	0,0096068	0,21619
Tichonow, aprior.	0,065849	0,33624	0,0028791	0,11492
TSVD	0,046616	0,33921	0,0050756	0,10086
Średnia	0,044458	0,25796	0,019392	0,142532

Z danych zestawionych w tabeli wynika, że funkcje rozkładu wielkości cząstek fazy zdyspergowanej $\hat{f}(a)$ odtworzone według modelu matematycznego wynikającego z teorii dyfrakcji Fraunhofera charakteryzują się ogólnie mniejszym odchyleniem mierzonym wartością Δ od rzeczywistej – testowej funkcji rozkładu wielkości cząstek fazy zdyspergowanej $f_{\text{test}}(a)$ w porównaniu z funkcjami odtworzonymi na podstawie modelu matematycznego wynikającego z teorii rozpraszania Mie [43]. Uogólnienie to potwierdzają wartości błędu rekonstrukcji Δ uśrednionego dla poszczególnych stosowanych technik inwersyjnych. Dla danych niezakłóconych szumem średni błąd rekonstrukcji w przypadku zastosowania modelu Fraunhofera jest około 2,3 razy mniejszy niż w przypadku zastosowania modelu Mie. Dla danych zakłóconych szumem średni błąd rekonstrukcji na podstawie modelu Fraunhofera jest około 1,8 razy mniejszy od błędu rekonstrukcji na podstawie modelu Mie [43].

Uzyskane wyniki wskazują, że rekonstrukcja funkcji rozkładu wielkości cząstek fazy zdyspergowanej na podstawie modelu Fraunhofera charakteryzuje się mniejszą podatnością na niekorzystny wpływ słabego uwarunkowania zagadnienia odwrotnego w porównaniu z rekonstrukcją według modelu Mie, o ile spełnione są warunki stosowalności modelu Fraunhofera [43].

Otrzymane rezultaty uzasadniają stosowanie modelu Fraunhofera, którego istotnymi zaletami, jeśli idzie o metrologię są prostota oraz brak zależności od współczynników załamania [43].

7.2.2. WYNIKI BADAŃ PORÓWNAWCZYCH JAKOŚCI REKONSTRUKCJI FUNKCJI ROZKŁADU WIELKOŚCI CZĄSTEK NA PODSTAWIE WYNIKÓW POMIARÓW NEFELOMETRYCZNYCH I TURBIDYMETRYCZNYCH

Wyniki badań symulacyjnych jakości rekonstrukcji funkcji rozkładu wielkości cząstek fazy zdyspergowanej dokonanej na podstawie wyników pomiarów nefelometrycznych oraz wyników pomiarów turbidymetrycznych na podstawie modeli matematycznych pomiarów wynikających z teorii rozpraszania Mie z zastosowaniem wybranych technik inwersyjnych zaprezentowano w tabeli 7.2.2.1 [44].

Z danych zestawionych w tabeli wynika, że funkcje rozkładu wielkości cząstek fazy zdyspergowanej $\hat{f}(a)$ odtworzone na podstawie wyników symulowanych pomiarów turbidymetrycznych wykazują ogólnie mniejsze odchylenie mierzone wartością Δ od rzeczywistej – testowej funkcji rozkładu wielkości cząstek fazy zdyspergowanej $f_{\text{test}}(a)$ w porównaniu z funkcjami odtworzonymi na podstawie wyników symulowanych pomiarów nefelometrycznych [44]. Potwierdzeniem tego uogólnienia są wartości błędu rekonstrukcji Δ uśrednionego dla poszczególnych stosowanych technik inwersyjnych. Dla danych niezakłóconych szumem średni błąd rekonstrukcji na podstawie wyników symulowanych pomiarów turbidymetrycznych jest blisko 1,2 razy mniejszy niż średni błąd rekonstrukcji na podstawie wyników symulowanych pomiarów nefe-

lometrycznych. Dla danych zakłóconych szumem natomiast średni błąd rekonstrukcji na podstawie wyników symulowanych pomiarów turbidymetrycznych jest ponad 2,2 razy mniejszy od średniego błędu rekonstrukcji na podstawie wyników symulowanych pomiarów nefelometrycznych [44].

Tabela 7.2.2.1. Zestawienie wartości parametru $\Delta[\mu\text{m}^{-1}]$ charakteryzującego dokładność rekonstrukcji funkcji rozkładu wielkości cząstek fazy zdyspergowanej na podstawie wyników symulowanych pomiarów nefelometrycznych oraz turbidymetrycznych zakłóconych i niezakłóconych szumem

Table 7.2.2.1. Comparison of the values of the parameter $\Delta[\mu\text{m}^{-1}]$ characterizing the accuracy of reconstruction of particle size distribution of dispersed phase on the basis of results of simulated nephelometric and turbidimetric measurements corrupted and not corrupted by noise

Procedura inwersyjna \ Rodzaj pomiaru, zaszumienie danych	nefelometria, niezaszumione	nefelometria, zaszumione	turbidymetria, niezaszumione	turbidymetria, zaszumione
Tichonow, norm.	0,068938	0,34742	0,058207	0,20016
Tichonow, I rz.	0,074651	0,18218	0,096930	0,092585
Tichonow, II rz.	0,010262	0,13346	0,0028386	0,022230
Tichonow, III rz.	0,00043019	0,20925	0,00040669	0,039482
Tichonow, aprior.	0,065849	0,33624	0,039513	0,15754
TSVD	0,046616	0,33921	0,025040	0,18746
Średnia	0,044458	0,25796	0,037156	0,116576

7.2.3. WYNIKI BADANIA JAKOŚCI REKONSTRUKCJI FUNKCJI ROZKŁADU WIELKOŚCI CZĄSTEK Z ZASTOSOWANIEM REGULARYZACJI TICHONOWA Z OGRANICZENIAMI

Wyniki badania jakości rekonstrukcji funkcji rozkładu wielkości cząstek fazy zdyspergowanej układu dyspersyjnego na podstawie wyników pomiarów turbidymetrycznych, przeprowadzonej z zastosowaniem regularyzacji Tichonowa z ograniczeniami według modelu matematycznego wynikającego z teorii rozpraszania Mie zaprezentowano w tabelach 7.2.3.1 i 7.2.3.2.

Na podstawie danych zaprezentowanych w tabeli 7.2.3.1 stwierdzić można, że niepewność Σ_p rozwiązania uzyskanego z zastosowaniem regularyzacji Tichonowa z ograniczeniami w porównaniu z rozwiązaniem uzyskanym stosując zwykłą regularyzację Tichonowa, jest:

- nieco większa dla miary braku gładkości wyrażonej normą euklidesową wektora rozwiązania dla wszystkich wartości parametru p , jak również dla miary braku gładkości wyrażonej normą euklidesową wektora różnic pierwszego rzędu rozwiązania w przypadku p równego 0,0%, 2,5% oraz 5,0%,

- znacznie mniejsza dla wszystkich pozostałych miar braku gładkości oraz wartości parametru p .

Tabela 7.2.3.1. Porównanie niepewności $\Sigma_p[\text{nm}^{-1}]$ funkcji rozkładu wielkości cząstek fazy zdyspergowanej zrekonstruowanej z zastosowaniem zwykłej regularyzacji Tichonowa oraz regularyzacji Tichonowa z ograniczeniami dla różnych miar braku gładkości szukanego rozwiązania przy różnych wartościach parametru p

Table 7.2.3.1. Comparison of uncertainty $\Sigma_p[\text{nm}^{-1}]$ of particle size distribution of dispersed phase reconstructed with use of ordinary Tikhonov regularization and constrained Tikhonov regularization for various measures of lack of smoothness of sought solution at various values of parameter p

p (%)	Metoda inwersyjna							
	Norma		Różnica I rz.		Różnica II rz.		Różnica III rz.	
	bez ograni- czeń	z ograni- czeniami	bez ograni- czeń	z ograni- czeniami	bez ograni- czeń	z ograni- czeniami	bez ograni- czeń	z ograni- czeniami
0,0	2,72E-18	4,71E-18	4,92E-18	5,48E-18	2,31E-17	4,74E-18	4,83E-17	4,10E-18
2,5	0,000358	0,000404	0,000338	0,000417	0,001287	7,53E-05	0,008871	0,000738
5,0	0,000432	0,0005	0,000599	0,000612	0,002557	0,00022	0,01758	0,001023
7,5	0,000564	0,000633	0,28612	0,001303	0,003266	0,000251	0,025459	0,00124

Tabela 7.2.3.2. Porównanie odchylenia $\Delta_p[\text{nm}^{-1}]$ funkcji rozkładu wielkości cząstek fazy zdyspergowanej zrekonstruowanej z zastosowaniem zwykłej regularyzacji Tichonowa oraz regularyzacji Tichonowa z ograniczeniami od testowej funkcji rozkładu dla różnych miar braku gładkości szukanego rozwiązania przy różnych wartościach parametru p

Table 7.2.3.2. Comparison of deviation $\Delta_p[\text{nm}^{-1}]$ of particle size distribution of dispersed phase reconstructed with use of ordinary Tikhonov regularization and constrained Tikhonov regularization from test particle size distribution for various measures of lack of smoothness of sought solution at various values of parameter p

p (%)	Metoda inwersyjna							
	Norma		Różnica I rz.		Różnica II rz.		Różnica III rz.	
	bez ograni- czeń	z ograni- czeniami	bez ograni- czeń	z ograni- czeniami	bez ograni- czeń	z ograni- czeniami	bez ograni- czeń	z ograni- czeniami
0,0	0,001046	0,000678	0,002576	0,000496	0,016733	0,000891	0,052086	0,000511
2,5	0,001156	0,000784	0,00421	0,000525	0,016785	0,000897	0,054167	0,000454
5,0	0,001207	0,000876	0,004508	0,000578	0,016851	0,000929	0,056762	0,000568
7,5	0,001243	0,000922	0,02861	0,00057	0,016745	0,000963	0,055744	0,000704

Na podstawie danych zaprezentowanych w tabeli 7.2.3.2 stwierdzić można, że odchylenie Δ_p od testowej funkcji rozkładu rozwiązania uzyskanego z zastosowaniem regularyzacji Tichonowa z ograniczeniami jest mniejsze niż rozwiązania uzyskanego z zastosowaniem zwykłej regularyzacji Tichonowa dla wszystkich miar braku gładkości oraz wszystkich wartości parametru p . Przewaga regularyzacji Tichonowa z ograniczeniami nad zwykłą regularyzacją Tichonowa staje się bardziej widoczna w przypadku miar braku gładkości wyrażonych normami euklidesowymi różnic pierwszego, drugiego i trzeciego rzędu wektora rozwiązania. Przewaga ta rośnie znacząco wraz z wartością p – dla $p = 0,0\%$ Δ_p dla regularyzacji z ograniczeniami jest około 1,5 razy mniejsze niż dla regularyzacji bez ograniczeń, natomiast dla $p = 5,0\%$ Δ_p jest mniejsze o ponad jeden rząd wielkości, natomiast dla $p = 7,5$ – o około dwa rzędy wielkości.

Reasumując, przeprowadzone badania symulacyjne wykazały, że rezultaty rekonstrukcji funkcji rozkładu wielkości cząstek z zastosowaniem regularyzacji Tichonowa z ograniczeniami są znacznie lepsze od rezultatów uzyskiwanych z zastosowaniem zwykłej regularyzacji Tichonowa w sensie mniejszej ogólnej niepewności oraz mniejszego ogólnego odchylenia od rzeczywistej – testowej funkcji rozkładu. Metoda z ograniczeniami okazuje się lepsza dla wszystkich poziomów zaszumienia danych p oraz dla wszystkich miar braku gładkości szukanego rozwiązania, aczkolwiek jej przewaga uwidacznia się najbardziej przy większych wartościach p . Należy podkreślić, że regularyzacja z ograniczeniami z powodzeniem odtwarza rozkład wielkości cząstek w warunkach, w których regularyzacja bez ograniczeń całkowicie zawodzi.

8. PODSUMOWANIE

Opracowanie przedstawia wybrane aspekty zagadnienia odwrotnego w wyznaczaniu rozkładu wielkości cząstek fazy zdyspergowanej układów dyspersyjnych na podstawie wyników pomiarów nefelometrycznych i turbidymetrycznych. Zadanie to stanowi przykład zagadnienia odwrotnego w pomiarach pośrednich, które jest jednym z kluczowych problemów metrologii.

W pracy analizie poddano fizyczne podstawy pomiarów nefelometrycznych i turbidymetrycznych. Wykazano, że przyjmując odpowiednie założenia upraszczające takie, jak założenie o rozpraszaniu sprężystym, jednokrotnym i niekoherentnym, modele matematyczne obu rodzajów pomiarów wyrazić można za pomocą równania całkowitego Fredholma pierwszego rodzaju. Jądro równania wynika z przyjętej teorii rozpraszania światła na pojedynczej cząstce fazy zdyspergowanej. Rozważano dwie teorie – ścisłą teorię Mie rozpraszania światła oraz stanowiącą jej przybliżenie teorię dyfrakcji Fraunhofera.

Udowodniony fakt umożliwia zaliczenie zagadnienia odwrotnego w pomiarach nefelometrycznych i turbidymetrycznych do obszernej i ważnej praktycznie klasy problemów odwrotnych wyrażonych równaniem całkowym Fredholma pierwszego rodzaju. W związku z tym w dalszej części opracowania rozważane były właściwości matematyczne zagadnień odwrotnych należących do tej klasy. Wykazano, że problemy te wykazują tzw. złe postawienie w sensie Hadamarda, ponieważ ich rozwiązanie jest niestabilne, tzn. nie zależy w sposób ciągły od danych wejściowych (pomiarowych). Na podstawie aparatu pojęciowego analizy funkcjonalnej, a w szczególności teorii funkcji osobliwych, udowodnione również zostało słabe uwarunkowanie zagadnień odwrotnych wyrażonych równaniem całkowym Fredholma pierwszego rodzaju, polegające na tym, że niewielka nawet niepewność danych pomiarowych prowadzi do olbrzymiej niepewności rozwiązania uzyskiwanego według tych danych. Omówiono jakościową interpretację słabego uwarunkowania jako następstwa tego, że wyrażony przez model matematyczny warunek zgodności rozwiązania z danymi pomiarowymi

nie wystarcza do dokładnego i jednoznacznego określenia rozwiązania, ponieważ bardzo szeroki przedział szukanych funkcji jest zgodny z danymi pomiarowymi z dokładnością do nawet bardzo małej niepewności pomiarowej tych danych. Przedstawiono również niezwykle istotną rolę, jaką dodatkowa wiedza o szukanym rozwiązaniu ma *a priori* spełnia podczas wyznaczania rozwiązania o niewielkim poziomie niepewności, umożliwiając wybranie wąskiego przedziału rozwiązań spośród bardzo szerokiego przedziału rozwiązań spełniających wyjściowe zagadnienie odwrotne.

Na podstawie zaprezentowanej analizy właściwości matematycznych zagadnień odwrotnych, wyrażonych równaniem całkowym Fredholma pierwszego rodzaju, w kolejnej części opracowania przedstawiono metody rozwiązywania tego typu zagadnień. Jako przykład rozwiązania analitycznego omówione zostało rozwiązanie Śifrina dla przypadku pomiarów nefelometrycznych, których model matematyczny oparty jest na teorii dyfrakcji Fraunhofera. Ponieważ analityczne rozwiązania równania całkowego Fredholma pierwszego rodzaju znalezione zostały tylko dla nielicznych szczególnych postaci funkcji jądra całkowego, w pracy skoncentrowano się na numerycznych metodach rozwiązywania zagadnienia odwrotnego. Omówiono dyskretyzację problemu prowadzącą do słabo uwarunkowanego numerycznie układu algebraicznych równań liniowych. Następnie zaprezentowano wybrane numeryczne techniki inwersyjne, szczególną uwagę poświęcając wiedzy apriorycznej o poszukiwanym rozwiązaniu, którą posługują się poszczególne procedury w celu minimalizacji efektów słabego uwarunkowania zagadnienia.

Analizie poddano metody regularyzacji, stosując podejście oparte na przybliżonej wcześniej teorii rozkładu macierzy na wartości osobliwe. Wśród omawianych technik regularyzacji przedstawiono regularyzację Tichonowa z ograniczeniami. Metoda ta w przeciwieństwie do technik klasycznych uwzględnia dwa fizyczne ograniczenia nałożone na szukaną funkcję rozkładu wielkości cząstek – warunek nieujemności funkcji oraz warunek jej unormowania do jedności. Dzięki temu uzyskiwane rozwiązania są realistyczne fizycznie. W pracy wykazano, że uwzględnienie wspomnianych ograniczeń prowadzi do zagadnienia programowania kwadratowego, które może być rozwiązane z wykorzystaniem istniejących algorytmów takich, jak metoda aktywnego zbioru. Regularyzacja Tichonowa z ograniczeniami jest przedmiotem badań prowadzonych przez autorów.

Jako reprezentantów algorytmów nieliniowych przedstawiono metodę Twomeya oraz jej udoskonalenie – metodę Twomeya–Markowskiego.

Zaprezentowano też bayesowskie podejście do rozwiązywania zagadnienia odwrotnego, w praktyczny sposób realizowane z wykorzystaniem metody Monte Carlo.

Ostatnia część opracowania poświęcona została prowadzonym przez autorów badaniom symulacyjnym zagadnienia odwrotnego w pomiarach nefelometrycznych i turbidymetrycznych. W pracy zaprezentowano trzy przedsięwzięcia badawcze:

- badanie porównawcze jakości rekonstrukcji funkcji rozkładu wielkości cząstek układu dyspersyjnego na podstawie wyników pomiarów nefelometrycznych przepro-

wadzonej według modelu matematycznego wynikającego z teorii rozpraszania Mie oraz modelu matematycznego wynikającego z teorii dyfrakcji Fraunhofera,

- badanie porównawcze jakości rekonstrukcji funkcji rozkładu wielkości cząstek układu dyspersyjnego na podstawie wyników pomiarów nefelometrycznych oraz wyników pomiarów turbidymetrycznych przeprowadzonych według modelu matematycznego wynikającego z teorii rozpraszania Mie,

- badanie jakości rekonstrukcji funkcji rozkładu wielkości cząstek układu dyspersyjnego na podstawie wyników pomiarów turbidymetrycznych, przeprowadzonej z zastosowaniem regularyzacji Tichonowa z ograniczeniami na tle innych metod inwersyjnych.

Dzięki przedstawionym wynikom możemy stwierdzić, że:

- rekonstrukcja funkcji rozkładu wielkości cząstek układu dyspersyjnego według modelu Fraunhofera charakteryzuje się ogólnie lepszą dokładnością w porównaniu z rekonstrukcją na podstawie modelu Mie, o ile spełnione są warunki stosowalności modelu Fraunhofera,

- rekonstrukcja funkcji rozkładu wielkości cząstek układu dyspersyjnego na podstawie wyników symulowanych pomiarów turbidymetrycznych charakteryzuje się ogólnie lepszą dokładnością w porównaniu z rekonstrukcją na podstawie wyników symulowanych pomiarów nefelometrycznych,

- funkcje rozkładu wielkości cząstek układu dyspersyjnego, odtworzone z zastosowaniem regularyzacji Tichonowa z ograniczeniami, wykazują znacznie mniejszą ogólną niepewność oraz lepszą zgodność z rzeczywistą funkcją rozkładu w porównaniu z rezultatem rekonstrukcji z zastosowaniem zwykłej regularyzacji Tichonowa.

LITERATURA

- [1] LINGJEARDE O.C., CHRISTOPHERSEN N., *Regularization Principles, Solving Ill-Posed Inverse Problems*, 1998.
- [2] CRUMP J.G., SEINFELD J.H., *A New Algorithm for Inversion of Aerosol Size Distribution Data*, *Aerosol Sci. and Technol.*, 1982, 1, 15–34.
- [3] KANDLIKAR M., RAMACHANDRAN G., *Inverse Methods for Analysing Aerosol Spectrometer Measurements: A Critical Review*, *J. Aerosol Sci.*, 1999, 30 (4), 413–437.
- [4] HANSEN P.C., *Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems*, *Numerical Algorithms*, 6, 1–35.
- [5] MROCZKA J., *Metrologiczne problemy wykorzystania światła rozproszonego do badań rozkładu wielkości cząstek w roztworach dyspersyjnych*, *Metrologia i Systemy Pomiarowe*. Monografia 4, Warszawa 1990.
- [6] POLAK A.G., MROCZKA J., *Pośrednie pomiary właściwości obiektów złożonych*, [W:] *Problemy metrologii elektronicznej i fotonicznej*, J. Mroczka (red.), Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2008.
- [7] BOHREN C.F., HUFFMAN D.R., *Absorption and scattering of light by small particles*, Wiley-Interscience, New York 1983.
- [8] JONES A.R., *Light scattering for particle characterization*, *Progr. Energy Combust. Sci.*, 1992, 25, 1–53.

- [9] XU R., *Particle characterization: light scattering methods*, Academic Publishers Kluwer, Dordrecht, Boston, London 2000.
- [10] JAWORSKI B.M., DIETŁAF A.A., *Fizyka. Poradnik encyklopedyczny*, wyd. drugie, PWN, Warszawa 1996.
- [11] CZERWIŃSKI M., MROCZKA J., GIRASOLE T., GOUESBET G., GREAHN G., *Light Transmittance Predictions under Multiple Light Scattering Conditions. Part 2: Inverse Problem – Particle Size Determination*, *Appl. Opt.*, 2001, 40(9), 1525–1531.
- [12] VIERA G., BOX M.A., *Information content analysis of aerosol remote-sensing experiments using an analytic eigenfunction theory: anomalous diffraction approximation*, *Appl. Opt.*, 1985, 24(24), 4525–4533.
- [13] LESNIC D., ELLIOTT L., INGHAM D.B., *A Numerical Analysis of the Data Inversion of Particle Sizing Instruments*, *J. Aerosol Sci.*, 1996, 27(7), 1063–1082.
- [14] BRANDT S., *Analiza danych*, PWN, Warszawa 1998.
- [15] BJÖRCK Å., DAHLQUIST G., *Metody numeryczne*, PWN, Warszawa 1983.
- [16] HANSEN P.C., *Computation of the singular value expansion*, *Computing*, 1988, 40, 185–199.
- [17] <http://mathworld.wolfram.com>
- [18] VASIN V.V., AGEEV A.L., *Ill-Posed Problems with A Priori Information*, VSP, Utrecht 1995.
- [19] TIKHONOV A.N., *Solution of incorrectly formulated problems and the regularization method*, *Dokl. Akad. Nauk. SSSR*, 1963, 151, 501–504 = *Soviet Math. Dokl.*, 1963, 4, 1035–1038.
- [20] TIKHONOV A.N., ARSEININ V.Y., *Solutions of Ill-Posed Problems*, Winston & Sons, Washington, D.C, 1977.
- [21] TIKHONOV A.N., GONCHARSKY A.V., *Ill-Posed Problems in Natural Sciences*, MIR Publishers, Moscow 1987.
- [22] HANSEN P.C., *Regularization, GSVD and truncated GSVD*, *BIT*, 1989, 29, 491–504.
- [23] HANSEN P.C., *Relations between SVD and GSVD of discrete regularization problems in standard and general form*, *Lin. Alg. Appl.*, 1990, 141, 165–176.
- [24] HANSEN P.C., *The truncated SVD as a method for regularization*, *BIT*, 1987, 27, 543–553.
- [25] HANSEN P.C., SEKII T., SHIBAHASHI H., *The modified truncated SVD method for regularization in general form*, *SIAM J. Sci. Stat. Comput.*, 1992, 13, 1142–1150.
- [26] TWOMEY S., *On the Numerical Solution of Fredholm Integral Equations of the First Kind by the Inversion of the Linear System Produced by Quadrature*, *J. Assoc. Comp. Mach.*, 1963, 10 (1), 97–101.
- [27] PHILLIPS D.L., *A Technique for the Numerical Solution of Certain Integral Equations of the First Kind*, *J. Assoc. Comp. Mach.*, 1962, 9(1), 84–97.
- [28] MARTOS B., *Programowanie nieliniowe. Teoria i metody*, PWN, Warszawa 1983.
- [29] ZANGWILL W.I., *Programowanie nieliniowe*, WNT, Warszawa 1974.
- [30] HANSEN P.C., *Truncated SVD solutions to discrete ill-posed problems with ill-determined numerical rank*, *SIAM J. Sci. Stat. Comput.*, 1990, 11, 503–518.
- [31] VOGEL C.R., *Optimal choice of a truncation level for the truncated SVD solution of linear first kind integral equations when data are noisy*, *SIAM J. Numer. Anal.*, 1986, 23, 109–117.
- [32] HANSEN P.C., O’LEARY D.P., *The use of the L-curve in the regularization of discrete ill-posed problems*, *SIAM J. Sci. Comput.*, 1993, 14, 1487–1503.
- [33] GOLUB G.H., HEATH M., WAHBA H., *Generalized cross validation as a method for choosing a good ridge parameter*, *Technometrics*, 1979, 21, 215–224.
- [34] MOROZOV V.A., *Methods for Solving Incorrectly Posed Problems*, Springer Verlag, New York 1984.
- [35] CHAHINE M.T., *Determination of the temperature profile in an atmosphere from its outgoing radiance*, *J. Opt. Soc. Am.*, 1968, 58, 1634–1637.

- [36] TWOMEY S., *Comparison of constrained linear inversion and an iterative non-linear algorithm applied to the indirect estimation of particle size distributions*, J. Comput. Phys., 1975, 18, 188–200.
- [37] WINKLMAYR W., WANG H., JOHN W., *Adaptation of the Twomey Algorithm to the Inversion of Cascade Impactor Data*, J. Aerosol Sci. and Technol., 1990, 13, 322–331.
- [38] MARKOWSKI G.R., *Improving Twomey's Algorithm for Inversion of Aerosol Measurement Data*, Aerosol Sci. and Technol., 1987, 7, 127–141.
- [39] RAY SMITH C., GRANDY W.T., JR. (eds.), *Maximum-Entropy and Bayesian Methods in Inverse Problems*, Reidel, Boston 1985.
- [40] MROCZKA J., *Method of moments in light scattering data inversion in the particle size distribution*, Opt. Comm., 1993, 99, 147–151.
- [41] MROCZKA J., *Integral transform technique in particle sizing*, J. Aerosol Sci., 1989, 20, 1075–1077.
- [42] MROCZKA J., *Light scattering at a small angle with the use of Lorenz-Mie theory in particle sizing*, Part. Part. Syst. Charact., 1989, 6, 86–88.
- [43] MROCZKA J., SZCZUCZYŃSKI D.K., *Wpływ modelu rozproszenia światła na jakość rozwiązań zagadnienia odwrotnego w pomiarach nefelometrycznych*, Pomiary, Automatyka, Kontrola, 2007, 53 (9bis, t. 1), 242–245.
- [44] MROCZKA J., SZCZUCZYŃSKI D.K., *Problem odwrotny – jakość rekonstrukcji funkcji rozkładu wielkości cząstek w pomiarach nefelometrycznych i turbidymetrycznych*, Pomiary, Automatyka, Kontrola, 2007, 53 (9bis, t. 1), 246–249.
- [45] December 1 1996 *Index of Refraction, Technical Note-007* (Duke Scientific Corporation).
- [46] September 1997 *Release on the Refractive Index of Ordinary Water Substance as a Function of Wavelength, Temperature and Pressure* (Erlangen, Germany: The International Association for the Properties of Water and Steam).

MATHEMATICAL AND PHYSICAL ASPECTS OF INVERSE PROBLEM IN PARTICLE SIZING WITH USE OF LIGHT SCATTERING

The study presents chosen aspects of inverse problem in determining particle size distribution of dispersed phase of particulate system on the basis of measured characteristics of light scattering in particulate system. The problem represents extremely important for metrology inverse problem in indirect measurements. By analysis of physical basis of nephelometric and turbidimetric measurements it was shown in the work that under appropriate simplifying assumptions mathematical models of both kinds of measurements take the form of Fredholm integral equation of the first kind. Hence the inverse problem in nephelometric and turbidimetric measurements can be analysed on the basis of the general theory of inverse problems expressed by Fredholm integral equation of the first kind, what is the subject of further parts of the study. Ill-posedness in the Hadamard sense of this class of inverse problems is proved. Based on functional analysis and particularly singular functions theory reasons of ill-conditioning of this kind of inverse problems were revealed. Subsequently chosen techniques of solving considered inverse problem were discussed focusing on numerical methods. An attention was paid to the role of *a priori* knowledge about sought solution which inverse procedures employ in various forms for the sake of minimizing effects of ill-conditioning. In the last part of the study simulation research conducted by authors were presented. Among discussed research undertakings research concerning constrained Tikhonov regularization were presented. Mentioned results prove substantial advantage of this technique over classical regularization.

*analiza widm, spektroskopia,
modelowanie matematyczne,
pomiary pośrednie*

Waldemar WÓJCIK,
Sławomir CIĘSZCZYK*

WYZNACZANIE INFORMACJI NA PODSTAWIE WIDM ABSORPCYJNYCH ORAZ EMISYJNYCH Z WYKORZYSTANIEM METOD MODELOWANIA MATEMATYCZNEGO

W pracy zaprezentowana została metoda analizy widm w średniej podczerwieni oparta na modelowaniu matematycznym. Umożliwia ona wyznaczanie informacji z pomiarów *in situ* różnorodnych procesów, w których mierzone są widma gazów. W procesach takich jak spalanie czy zgazowywanie substancji, zmienia się nie tylko zawartość gazów, ale również ich temperatura. Modele wykorzystujące modelowanie matematyczne przekształcają zawartość gazów oraz ich temperaturę w widmo. Wyznaczanie parametrów na podstawie widm pomiarowych jest zagadnieniem odwrotnym.

1. WPROWADZENIE

Określenie zawartości oraz temperatury składników gazowych dla różnych procesów fizykochemicznych jest zadaniem trudnym. Pomiary wykonuje się z wykorzystaniem sond inwazyjnych oraz procedur analitycznych przeprowadzanych poza diagnozowanym obiektem. W niektórych procesach, gdzie zachodzi potrzeba analizy składu gazów i temperatury nie ma możliwości wykonania pomiarów bezpośrednich. Do takich procesów zaliczyć można: pomiar składu spalin silników samolotowych, pomiar składu gazów wulkanicznych, pomiar zawartości gazowych składników zanieczyszczeń środowiskowych, określanie miejsc wycieków substancji gazowych z instalacji przemysłowych oraz ich zawartości w powietrzu atmosferycznym, czy pomiary procesów fizykochemicznych w obiektach zamkniętych. W przypadku procesów zmieniających się w czasie, zachodzi często potrzeba określania zmian koncen-

* Katedra Elektroniki, Politechnika Lubelska.

tracji składników. Inwazyjne procedury pomiarowe nie zawsze można zastosować do ciągłego monitoringu procesów. W takich przypadkach pomiary wykonywane mogą być z pewnej odległości. Takie metody pomiarowe zaliczyć można do grupy pomiarów zdalnych (ang. *remote sensing*). Dwa główne powody wykorzystywania takich metod to brak możliwości dostępu do procesu z powodu odległości, np. pomiary atmosferyczne oraz brak dostępu z powodu trudnych warunków panujących wokół procesu, np. wysoka temperatura oraz zapylenie w pobliżu przemysłowych obiektów z procesem spalania. We wspomnianych obiektach przemysłowych stosuje się specjalne sondy chłodzone wodą lub powietrzem, które dodatkowo są zaprojektowane tak by czoło sondy nie podlegało zabrudzeniom. W obiektach przemysłowym na podstawie widma promieniowania odbywać się może określanie cech diagnostycznych procesu, a zatem służyć ono może jako sygnał wykorzystywany do diagnostyki czy też sterowania procesem. Przy takim podejściu widmo nie zawsze jest analizowane w celu określenia zawartości składników, lecz dostarcza ono informację o stanie procesu przez określenie wartości jakiegoś parametru (np. stosunku stechiometrycznego).

Obecnie obserwować można trend w zwiększaniu wykorzystania odnawialnych źródeł energii. Zwiększa się liczba procesów, w których spalaniu podlega biomasa. Procesy spalania biomasy lub też współspalania biomasy z innymi paliwami są intensywnie badane oraz rozwijane przez wiele ośrodków naukowych oraz przemysłowych [10], [36]. Badania te mają na celu prowadzenie takich procesów, tak aby ich efektywność była wysoka, a ilości emitowanych związków szkodliwych dla środowiska możliwie małe, zgodnie z obowiązującymi normami. Określenie zawartości gazów powstających podczas spalania substancji służyć może również do prognozowania rozprzestrzeniania się pożarów [10], [39]. Właściwie modelowanie ich rozprzestrzeniania się w terenach leśnych oraz nieużytkach możliwe jest dopiero przy pełnej znajomości ilości emitowanych gazów oraz wpływu temperatury degradacji biomasy na tę ilość.

Pomiar *in situ* widm absorpcji oraz promieniowania może być przeprowadzony z wykorzystaniem różnorodnych technik spektralnych. Techniki te można podzielić na aktywne i pasywne. W aktywnych, oprócz analizatora widma, wymagane jest dodatkowe źródło promieniowania, gdyż mierzą one widmo absorpcyjne. Do ich pracy niezbędna jest zamknięta ścieżka pomiarowa pomiędzy źródłem a detektorem przechodząca przez badane środowisko. W wyniku analizy cech absorpcji składników określić można skład gazów, koncentrację pyłów, wartość temperatury, które w przypadku procesów technicznych mogą pełnić rolę parametrów diagnostycznych. W obiektach przemysłowych część ścieżki pomiarowej zastąpić można światłowodami, oddalając w ten sposób źródło promieniowania oraz analizator od agresywnego środowiska. Pasywne techniki pomiaru widm polegają na pomiarze naturalnego promieniowania badanego środowiska. Układ pomiarowy ogranicza się tu jedynie do analizatora widma. Urządzenia do pomiaru widm emisyjnych nazywane są również analizatorami widma o otwartej ścieżce pomiarowej. Nazwa ta wywodzi się od kla-

sycznych spektrometrów absorpcyjnych, w których w jednym urządzeniu zawarty jest zarówno analizator widma, jak też źródło promieniowania. Ścieżka pomiarowa jest tu zamknięta w takim sensie, że biegnie wewnątrz urządzenia. Badane próbki pobiera się z badanego procesu do odpowiednich kuwet, które następnie umieszcza się w spektrometrze. O zamkniętej ścieżce pomiarowej można mówić również w absorpcyjnych pomiarach *in situ*, w których ścieżka biegnie przez badane środowisko, ale poza urządzeniem pomiarowym. Pomiar metodami pasywnymi dostarcza eksperymentatorowi widma pasywne. W części literatury widma z takich pomiarów nazywane są widmami emisyjnymi. Trudno się z tym nie zgodzić. Należy jednak uściślić pewne aspekty pomiarów pasywnych. Widma mierzone metodami pasywnymi mogą być widmami ściśle emisyjnymi, ale często są widmami emisyjno-absorpcyjnymi, ponieważ można w nich wyróżnić zarówno cechy emisyjne, jak i absorpcyjne. Jeśli analizator widma mierzy widmo promieniowania gorącego obszaru procesu, np. płomienia, to mierzone widmo będzie widmem emisyjnym. Jeżeli natomiast pomiędzy strefą gorącą płomienia a sondą pomiarową znajdować się będą składniki o niższej temperaturze, to będą one absorbować promieniowanie płomienia. Stąd też w widmie pasywnym pojawią się cechy absorpcji. Dlatego też widma mierzone technikami pasywnymi określać można jako widma pasywne, zostawiając termin widmo emisyjne dla pomiarów cech składników o wysokiej temperaturze.

Skupiając się na pomiarach w zakresie bliskiej i średniej podczerwieni, możemy mieć do czynienia z przyrządami spektralnymi do pomiaru widm o różnej liczbie kanałów spektralnych oraz różnym zakresie spektralnym. Najprostszymi urządzeniami mogą być filtry na jeden zakres, stosowane np. w pirometrii. Zwiększając ilość dostępnych informacji zawartych w widmie, zwiększać należy liczbę kanałów spektralnych. Możemy więc mieć do czynienia z dwu lub więcej kanałowymi pirometrami składającymi się z oddzielnych filtrów spektralnych, które zapewnić mogą jedynie ograniczoną rozdzielczość spektralną. Pomiar wielopasmowe z większą rozdzielczością wymagają zastosowania spektrometrów, które zbudowane są na bazie siatek dyfrakcyjnych lub interferometrów. Największą rozdzielczość jednak mają spektrometry oparte na interferometrze Michelsona, zwane ze względu na sposób przetwarzania sygnału pomiarowego spektrometrami fourierowskimi (FTIR – *Fourier Transform Infrared Spectrometer*).

Zawartości składników oraz wartość temperatury może być określana z wykorzystaniem widm mierzonych spektrometrem FTIR (*Fourier Transform Infrared Spectrometer*) o otwartej ścieżce pomiarowej (*open path*, OP-FTIR) [10], [11], [18], [36], [37], [39]. Jest to urządzenie rzadziej spotykane niż typowy laboratoryjny spektrometr FTIR, który jest zamkniętym układem składającym się ze źródła promieniowania, kuwety z badaną zawartością oraz interferometru Michelsona wraz z detektorem. Mieszanki zawarte w kuwecie pomiarowej umieszczane są wewnątrz spektrometru. Cechy absorpcji poszczególnych składników określają ich zawartość. Spektrometr OP-FTIR jest analizatorem widma docierającego do niego z zewnątrz. W zastosowaniach

praktycznych służyć on może jako analizator naturalnego widma promieniowania procesu (pomiar pasywny) lub też w wykorzystaniu dodatkowego źródła promieniowania w taki sposób, by diagnozowany proces znajdował się na ścieżce pomiarowej pomiędzy tym źródłem a spektrometrem (pomiar aktywny). Podstawowymi zaletami spektrometrów OP-FTIR są:

- 1) możliwość wykonania pomiarów *in situ*,
- 2) brak konieczności pobierania próbek do analizy oraz wykonywania manipulacji próbkami (chłodzenie, umieszczanie w zamkniętej kuwecie),
- 3) wykonywanie pomiaru jest szybkie w porównaniu do innych metod,
- 4) cały zakres spektralny widma jest wyznaczany w jednym akcie pomiarowym,
- 5) układ pomiarowy jest bardzo prosty,
- 6) na podstawie widma FTIR określać można jednocześnie zawartość wielu składników, a dla pomiarów *in situ* mogą to być również składniki aktywnie reagujące chemicznie.

Analiza ilościowa widm pasywnych jest zadaniem skomplikowanym. Dokładność procedury wyznaczania zawartości składników oraz ich temperatury zależy od użytego algorytmu numerycznego. Pasywne techniki FTIR znalazły zastosowanie m. in. do określania zawartości spalin silników samolotowych, zanieczyszczeń atmosferycznych, gazów wulkanicznych, gazów z kominów instalacji przemysłowych, wycieków z instalacji gazowych oraz rurociągów. Możliwości metod fourierowskich zależą od dwu parametrów: rozdzielczości pomiaru oraz czasu jego wykonania. Rozdzielczość determinuje możliwość rozróżniania cech absorpcji poszczególnych składników, wpływa znacząco na zmianę mierzonej wartości absorpcji od zawartości danego komponentu. Dla określonego spektrometru zmniejszenie rozdzielczości pomiaru zmniejsza czas akwizycji danych, ale również zwiększa stosunek sygnału do szumu. W warunkach laboratoryjnych w celu zwiększenia stosunku sygnału do szumu stosuje się wielokrotny pomiar widma, a następnie jego uśrednianie. Dynamiczne zmiany transmisyjności oraz emisji promieniowania dla pomiarów *in situ* sprawiają, że z tej możliwości skorzystać nie można.

W dalszej części przedstawiona zostanie metodologia wyznaczania informacji jakościowych i ilościowych na podstawie pomiarów *in-situ* widm absorpcyjnych oraz emisyjnych procesów o zmieniającej się temperaturze i zawartości składników gazowych czy też pyłowych. Pomiarów takie mogą być źródłem informacji bezpośrednio o miejscu zachodzenia procesu, a więc emisji gazów, pyłów oraz zmian temperatury. Ogólnie mogą to być informacje diagnostyczne o badanym obiekcie czy też procesie. W dalszej części omówione zostaną:

- a) klasyczna analiza ilościowa widm mierzonych spektrometrem FTIR o zamkniętej wewnętrznej ścieżce pomiarowej,
- b) modelowanie transmisyjności gazów w zależności od ich stężenia oraz temperatury,
- c) wpływ parametrów spektrometru FTIR na widmo pomiarowe, charakterystyka spektralna spektrometru,

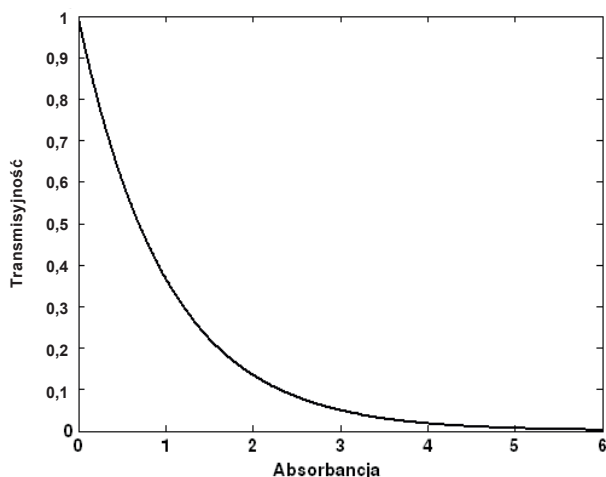
d) podstawowe parametry charakteryzujące promieniowanie spektralne w podczerwieni i w modelowaniu promieniowania jednorodnej oraz niejednorodnej ścieżki pomiarowej,

e) zagadnienie odwrotne w ilościowej analizie widm absorpcyjnych oraz emisyjnych mierzonych technikami *in situ*,

f) pomiary oraz analiza widm mierzonych spektrometrem FTIR o otwartej ścieżce pomiarowej.

2. KLASYCZNA ANALIZA ILOŚCIOWA POMIARU FTIR ORAZ JEJ OGRANICZENIA

Wyznaczanie informacji ilościowych na podstawie widm absorpcyjnych przeprowadza się z wykorzystaniem prawa Beera. W klasycznych technikach analizy zakłada się liniowość zależności zawartości składnika oraz absorpcji. W praktyce założenie takie często nie jest spełnione ze względu na nieliniowość zjawisk fizykochemicznych zachodzących w badanej mieszaninie oraz wpływie charakterystyki spektralnej spektrometru. Procedura automatycznego wyznaczenia krzywej kalibracyjnej, stosowana w większości nowoczesnych spektrometrów, umożliwia kalibrację w zakresie dwóch jednostek absorpcji, a rzadko i raczej w nietypowych układach, do czterech jednostek absorpcji. Jest to spowodowane małymi zmianami transmisyjności w zależności od absorpcji w zakresie 3–4 jednostek (rys. 1) i znacznym wzrostem szumów pomiarowych.



Rys. 1. Problem zależności transmisyjności od absorpcji
Fig. 1. Transmissivity versus absorbance

Przed przystąpieniem do kalibracji należy ustalić wszystkie parametry pomiarowe (rozdzielczość, apodyzację, uzupełnianie zerami interferogramu, korekcję fazy interferogramu). Mieszaniny stosowane do kalibracji powinny, jeśli to możliwe, zawierać pozostałe, aktualnie nie kalibrowane, składniki występujące w badanych próbkach o podobnej zawartości. Dobra znajomość składu próbek kalibracyjnych umożliwi poprawne przeprowadzenie procedury kalibracji. Próbkę oraz układ pomiarowy powinien znajdować się w takiej samej temperaturze podczas kalibracji oraz pomiarów właściwych. Na podstawie widm ze spektrometrów FTIR, które mogą być mierzone w dość szerokim zakresie długości fal, można wyznaczać wiele komponentów jednocześnie. Wartość absorbancji dla określonego składnika mieszaniny można zapisać następująco:

$$A_{\lambda} = a_{\lambda}lc, \quad (1)$$

gdzie: A_{λ} – absorbancja dla danej długości fali λ , a_{λ} – absorpcyjność składnika, l – długość ścieżki pomiarowej, c – koncentracja określonego składnika.

Z wykorzystaniem spektrometru mierzymy transmisyjność, tj.

$$\tau_{\lambda} = I_{\lambda}(0)/I_{\lambda}(l), \quad (2)$$

$I_{\lambda}(0)$ – natężenie promieniowania źródła, $I_{\lambda}(l)$ – natężenie promieniowania po przejściu przez badaną próbkę.

Na podstawie transmisyjności obliczyć możemy absorbancję badanego ośrodka według zależności:

$$A_{\lambda} = -\log(\tau_{\lambda}). \quad (3)$$

Stosowane są różnorodne procedury analizy ilościowej, ale najczęściej jest to regresja wielowymiarowa [19], [31], [39] (MLS – *multiple least-squares method*). W postaci macierzowej wartość absorbancji możemy zapisać następująco:

$$\mathbf{A} = \mathbf{K}\mathbf{C} + \mathbf{R}, \quad (4)$$

gdzie: \mathbf{A} – wartości absorbancji na kilku długościach fali, \mathbf{K} – współczynniki absorpcji, \mathbf{C} – zawartości poszczególnych składników mieszaniny, \mathbf{R} – wektor pozostałości dla poszczególnych długości fal.

Jeśli wyeliminowany zostanie wektor pozostałości, równanie (4) można zapisać w postaci

$$\mathbf{A} = \mathbf{K}\mathbf{C}. \quad (5)$$

Przed przystąpieniem do wyznaczania koncentracji musimy znać wartości macierzy \mathbf{K} z zależności:

$$\mathbf{K} = \mathbf{A}\mathbf{C}^{-1}. \quad (6)$$

Odwrotność macierzy C^{-1} obliczyć można jedynie, gdy jest ona macierzą kwadratową. Jeśli w celu zwiększenia poprawności rozwiązania stosujemy większą liczbę próbek do kalibracji, to macierz \mathbf{K} wyznaczamy następująco:

$$\mathbf{K} = \mathbf{A}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1} \quad (7)$$

Obliczanie nieznaney zawartości koncentracji na podstawie pomiaru A_{pom} przeprowadzamy za pomocą następującego wzoru:

$$\mathbf{C}_{\text{pom}} = (\mathbf{K}\mathbf{K}^T)^{-1}\mathbf{K}^T\mathbf{A}_{\text{pom}}, \quad (8)$$

gdzie: \mathbf{A}_{pom} – wartości absorbancji zmierzone dla kilku długości fal, \mathbf{C}_{pom} – wyznaczana wartość koncentracji poszczególnych składników mieszaniny.

Bardziej złożone metody regresji stosuje się w przypadkach nakładających się lub podobnych widm dla kilku składników. Są to [19], [29], np. regresja cząstkowa (PLS – *partial least square*), regresja głównych składowych (PCR – *principal component regression*), lokalnie ważona regresja (LWR – *locally weighted regression*) oraz sieci neuronowe. Stopień trudności analizy ilościowej zwiększa duża liczba składników mieszaniny, zarówno mierzonych, jak też zakłócających pomiar.

3. MODELOWANIE TRANSMISYJNOŚCI GAZÓW W ZALEŻNOŚCI OD STĘŻENIA ORAZ TEMPERATURY

W pomiarach *in situ* trudno jest zastosować bezpośrednio klasyczną metodę regresji wielowymiarowej. Głównymi przyczynami jest występująca zmienność temperatury badanego środowiska gazowego oraz możliwa duża dynamika zmian zawartości składników. Jeśli mierzymy transmisyjność zamkniętego obiektu, np. komory spalania lub reaktora chemicznego, to nie mamy wpływu na długość badanej ścieżki pomiarowej. Stąd też przeprowadzenie kalibracji w sensie klasycznym nie jest możliwe. W literaturze znaleźć można różne pośrednie metody kalibracji. Metody takie dostosowywane są do typowych parametrów procesu, np. dla średniej temperatury, dla kilku charakterystycznych wartości temperatury lub też dla typowych wartości temperatury oraz zawartości składników. Metoda ta (ang. *Synthetic Calibration/Classical Least Square Retrievals* – syntetyczna kalibracja/wyznaczanie niewiadomych klasyczną metodą najmniejszych kwadratów) stosowana była m.in. do badań dużej skali laboratoryjnych pożarów biomasy [18]. Ograniczenia, jakie napotykamy stosując te metody, to wyznaczanie niewiadomych przy uśrednionych wartościach innych parametrów, np. zawartości składnika gazowego przy zakładanej średniej temperaturze. Drugie ograniczenie to wpływ zawartości składników na transmisyjność. Pomiaru transmisyjności *in situ* przeprowadzane są dla określonej jej wartości wynikającej z charakteru

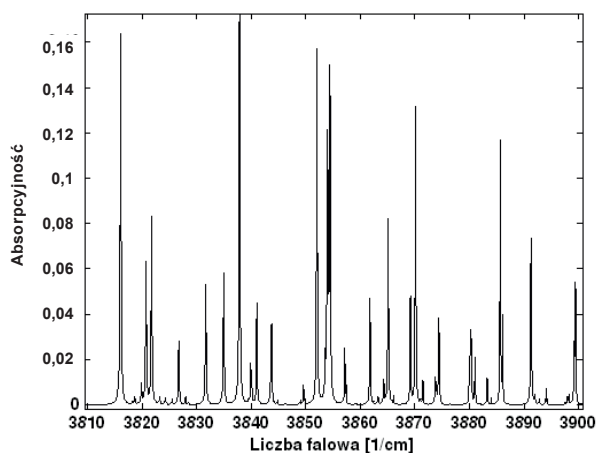
procesu. W klasycznych pomiarach spektrometrami FTIR eksperymentator może skracać lub wydłużać ścieżkę pomiarową przez wykorzystanie kuwet z wieloma odbiciami. Obiekt rzeczywisty cechują jego własne i określone wartości transmisyjności determinowane przez zawartość składników i temperaturę. Pomiar może charakteryzować silna nieliniowość zależności zawartości składnika i absorpcji. Ogólna relacja pomiędzy widmem a zawartością składników i temperaturą jest trudna do wyznaczenia. Stworzyć jednak można symulacyjne modele transmisyjności w zależności od zawartości składników oraz temperatury. Przy czym modele takie opisywać mogą zarówno homogeniczną ścieżkę pomiarową, jak też ścieżkę niejednorodną pod względem temperatury oraz koncentracji. Do celów obliczeniowych przyjmuje się często jednorodność ścieżki pomiarowej, dzięki czemu uproszczone zostają obliczenia numeryczne. W pomiarach diagnostycznych zainteresowani możemy być stosunkiem zawartości składników, np. CO/CO₂. Niejednorodność temperatury ścieżki pomiarowej w niewielkim stopniu wpływa na wyznaczenie takiego stosunku.

Rozwiązaniem problemu kalibracji dla pomiarów *in situ* są metody matematycznego modelowania cech spektralnych gazów. Pozwalają one obliczać transmisyjność ścieżki pomiarowej o dowolnych parametrach. Może to być również ścieżka niejednorodna, którą do obliczeń dzielimy na odpowiednią liczbę odcinków jednorodnych. Istnieją dwie główne grupy metod modelowania cech spektralnych gazów. Są to metody fizyczne oraz statystyczne. Modele fizyczne zwane są również modelami linia po linii, ponieważ za ich pomocą wyznacza się poszczególne linie rotacyjne gazów. Metody fizyczne wykorzystują prawa fizyczne opisujące ruch molekuł, które mogą poruszać się ruchem obrotowym oraz drgać zgodnie z zasadami teorii kwantowej. Energia fotonów absorbowanych lub emitowanych przez molekuły może przyjmować jedynie ściśle określone poziomy kwantowe. W widmie objawia się to prążkami (liniami) rotacyjnymi. Każda z tych linii ma określony kształt (intensywność, szerokość połówkowa), który jest zależny od temperatury, ciśnienia i składu mieszaniny. Linie rotacyjne są bardzo wąskie i leżą blisko siebie, tworząc tzw. pasma wibracyjne. Modele fizyczne składają się z kilku tabel parametrów dla dużej liczby linii rotacyjnych. Poszczególne parametry modeli wyznaczone zostają z wykorzystaniem metod laserowych oraz wysokorozdzielczych pomiarów FTIR. Parametry dla wysokich temperatur określone są za pomocą teorii z pomiarowym sprawdzeniem wybranych wartości. W typowym zakresie pomiarowym spektrometrów FTIR zawartych od kilkuset do nawet kilkuset tysięcy linii. Obliczanie transmisyjności dla tak dużej liczby linii jest zadaniem wymagającym dużych nakładów obliczeniowych. Dlatego też w niektórych obliczeniach lepiej jest zastosować statystyczne modele obliczeniowe absorpcji gazów. Powstały one głównie w celu obliczania transportu ciepła w obiektach przemysłowych (kotłach, piecach). Wyznaczenie ciepła oddawanego przez proces spalania do ścian kotła w dużych obiektach przemysłowych wymaga znacznej mocy obliczeniowej. Dodatkowe zastosowanie metod linia po linii do obliczania transmisyjności gazów znacznie wydłużyłoby czas obliczeń. Założeniem metod statystycznych jest su-

mowanie linii rotacyjnych w określonych pasmach, najczęściej o szerokości 25 cm^{-1} lub 50 cm^{-1} . Przy czym określane są pewne średnie, a więc statystyczne parametry linii takie jak: powierzchnia, intensywność, szerokość połówkowa. Parametry określa się tak, by zmiany transmisyjności pasma wraz z koncentracją oraz temperaturą odpowiadały zmianom rzeczywistym.

3.1. BAZA DANYCH HITRAN ORAZ HITEMP

Praktyczną realizacją modeli fizycznych są bazy danych zawierające poszczególne parametry linii rotacyjnych. Dla gazów atmosferycznych najbardziej znaną i często używaną bazą danych jest Hitran [31]. Ograniczeniem tej bazy jest temperatura dla jakiej można ją stosować. Wraz ze wzrostem wartości temperatury wzrasta błąd związany z modelowaniem linii gazowych. Dla temperatur, jakie panują w procesach spalania, stosować powinno się bazę Hitem. Hitran zawiera dane dotyczące środka linii, intensywności, współczynników poszerzenia linii w zależności od ciśnienia oraz poziomu najmniejszej energii przejścia. Baza ta zawiera parametry takich gazów, jak: CO_2 , H_2O , CO , NO , NO_2 , SO_2 , CH_4 , N_2O , HCl , HNO_3 . Do wyznaczenia intensywności linii dla temperatur innych niż temperatura odniesienia, stosuje się podstawowy poziom energii przejścia. Wyznaczenia szerokości linii dla zmieniających się temperatur polega na wykorzystaniu wykładnika potęgi współczynnika poszerzenia linii. Parametry bazy Hitran są wyznaczone z pomiarów. Jedynym parametrem wyznaczanym z teorii kwantowej jest podstawowa energii przejścia. W bazie danych Hitemp znajdują się dodatkowe linie aktywne dla temperatury powyżej 600 K. Parametry tych linii wyznaczane są z zależności teoretycznych.



Rys. 2. Absorpcyjność spektralna pary wodnej ($T = 400 \text{ K}$, $l = 1 \text{ cm}$, ułamek molowy zawartości = 0,01), na podstawie bazy Hitran

Fig. 2. Water vapour absorption ($T = 400 \text{ K}$, $l = 1 \text{ cm}$, mole fraction 0.01), Hitran database

Dla każdego gazu współczynnik absorpcji α zależy od intensywności przejścia absorpcji S [$\text{cm}^{-2}\text{atm}^{-1}$], funkcji kształtu linii ϕ [cm], całkowitego ciśnienia P [atm], procentowego stosunku molowego danego gazu x oraz długości ścieżki optycznej l , co można zapisać w postaci zależności [25]

$$\alpha = S\phi Pxl. \quad (9)$$

Intensywność linii przejścia absorpcji jest zależna od liczby cząstek o niższym stanie kwantowym. Liczba cząstek w poszczególnych stanach kwantowych opisana jest rozkładem Boltzmanna oraz odpowiednim prawdopodobieństwem przejścia dla konkretnych molekuł.

Temperatura wpływa na intensywność linii w następujący sposób [31], [40]:

$$S(T) = S(T_0) \frac{Q(T_0)}{Q(T)} e^{\frac{hcE''}{k} \left(\frac{1}{T} - \frac{1}{T_0} \right)} \frac{1 - e^{-\frac{hcv_0}{kT}}}{1 - e^{-\frac{hcv_0}{kT_0}}}, \quad (10)$$

gdzie: c – prędkość światła [cm/s], h – stała Planca [Js], k – stała Boltzmanna [J/K], E'' – energia stanu niższego [J], v_0 – częstotliwość przejścia [cm^{-1}], Q – sumaryczny stan energetyczny cząsteczki.

Na kształt linii ϕ wpływają procesy fizyczne, które powodują zmianę szerokości poszczególnych linii. Podstawowe mechanizmy zmiany kształtu linii dla temperatury poniżej 2000 K, to: poruszanie się cząstek ze względu na temperaturę oraz zderzenia cząstek pomiędzy sobą. Kształt linii opisany jest charakterystycznym profilem, który jest rozkładem natężenia danej linii widmowej wokół częstotliwości centralnej.

Wzrost szerokości linii pod wpływem temperatury związany jest ze zjawiskiem Dopplera, które polega na zmianie częstotliwości poruszającego się względem obserwatora źródła fali. Zmiana częstotliwości zależy od względnej prędkości poruszania się źródła. Pod wpływem wzrostu temperatury zwiększa się średnia prędkości poruszania się cząstek, a prędkości poruszania się cząstek opisuje rozkład Maxwella. Kształt poszerzenia linii związany z poszerzeniem temperaturowym opisany jest funkcją Gaussa [31]

$$\phi_D(\nu) = \frac{2}{\Delta\nu_D} \left(\frac{\ln 2}{\pi} \right)^{1/2} \exp \left(-4 \ln 2 \left(\frac{\nu - \nu_0}{\Delta\nu_D} \right)^2 \right), \quad (11)$$

gdzie $\Delta\nu_D$ – szerokość połówkowa poszerzenia dopplerowskiego, którą obliczyć można za pomocą następującego wzoru:

$$\Delta\nu_D = 7,16210^{-7} v_0 \left(\frac{T}{M} \right)^{1/2}, \quad (12)$$

gdzie: m jest masą cząsteczki, a M jest masą molową [g/mol].

Linie spektralne leżące w zakresach o większych długościach fal poszerzane są bardziej. Cząstki o mniejszej masie molowej podlegają większemu poszerzeniu niż cząstki o większej masie molowej.

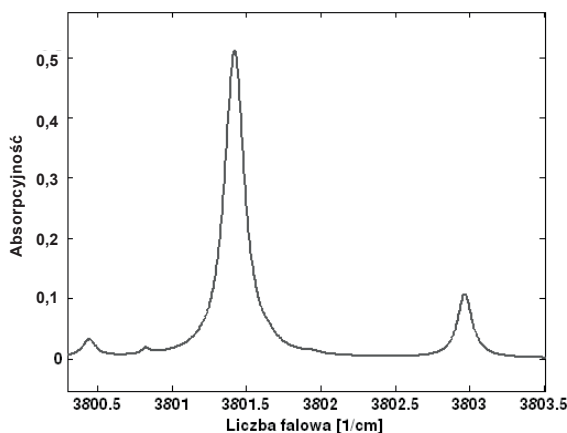
Poszerzenie linii w znacznym stopniu zależy również od oddziaływań pomiędzy molekułami. Jeżeli zderzenia zachodzą częściej, to zmiana szerokości linii jest odpowiednio większa. Zjawisko poszerzenia kolizyjnego zależy od częstości zderzeń pomiędzy molekułami, które zależą od ciśnienia oraz temperatury mieszaniny. Liczba zderzeń rośnie wraz ze wzrostem ciśnienia, stąd też ten rodzaj poszerzenia linii nazywany jest ciśnieniowym. Profil linii spektralnej związanej z tym poszerzeniem modeluje się za pomocą funkcji Lorentza [35]

$$\phi_K = \frac{1}{2\pi} \frac{\Delta\nu_C}{(\nu - \nu_0)^2 + \left(\frac{\Delta\nu_C}{2}\right)^2}, \quad (13)$$

gdzie $\Delta\nu_C$ – szerokość połówkowa linii, na którą wpływa ciśnienie całkowite mieszaniny oraz ułamki molowe zawartości poszczególnych cząstek wraz z ich szerokościami połówkowymi.

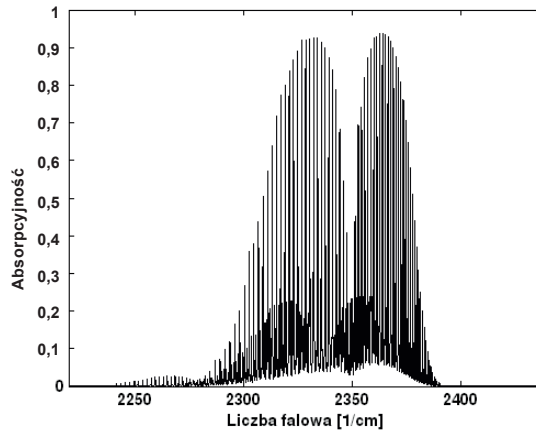
W modelowaniu linii muszą być uwzględnione obydwa rodzaje poszerzeń, tj. Dopplera oraz kolizyjne. Profil linii, jakim opisuje się wpływ obydwu zjawisk na kształt linii, nazywa się profilem Voigta, który jest splotem funkcji Gaussa oraz Lorentza [1], [35], tj.

$$\phi_V(\nu) = \int_{-\infty}^{+\infty} \phi_C(\nu') \phi(\nu - \nu') d\nu'. \quad (14)$$



Rys. 3. Kształt przykładowej linii spektralnej pary wodnej ($T = 500$ K, $l = 1$ cm, ułamek molowy zawartości = 0,01) na podstawie bazy Hitran

Fig. 3. Water vapour spectral line ($T = 500$ K, $l = 1$ cm, mole fraction 0.01), Hitran database



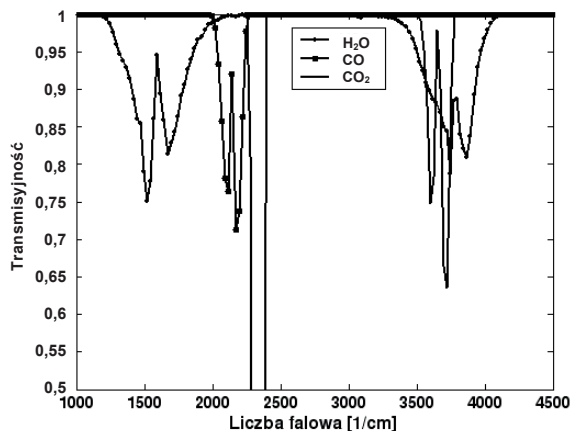
Rys. 4. Absorpcyjność spektralna dwutlenku węgla ($T = 400$ K, $l = 1$ cm, ułamek molowy zawartości = 0,01) na podstawie bazy Hitran

Fig. 4. Carbon monoxide spectral absorption ($T = 400$ K, $l = 1$ cm, mole fraction 0.01), Hitran database

3.2. STATYSTYCZNE MODELE OBLICZANIA PARAMETRÓW SPEKTRALNYCH GAZÓW

Jedne z bardziej popularnych statystycznych modeli do obliczania parametrów spektralnych gazów to statystyczny model wąskopasmowy SNB (ang. *Statistical Narrow Band*) oraz Radcal. Mogą one być używane do temperatury 2500 K. Dla tych modeli powierzchnie linii spektralnych zostają zsumowane w pasmach o szerokości 25 cm^{-1} . Aby poprawnie modelować zmianę transmisyjności poszczególnych kanałów spektralnych, należy znaleźć zmiany transmisyjności takiego kanału spektralnego od temperatury, zawartości analizowanego składnika oraz innych składników występujących w mieszaninie. W tym celu określa się statystyczne parametry linii zawartych w tym kanale. Każda linia rotacyjna opisana jest położeniem, intensywnością, szerokością, parametrami określającymi wpływ zmian szerokości połówkowej oraz intensywności wraz z temperaturą, ciśnieniem i zawartością. W paśmie o szerokości 25 cm^{-1} znajduje się wiele zachodzących na siebie linii. Współczynnik absorpcji danego pasma można obliczyć, sumując powierzchnię poszczególnych linii, uwzględniając ich zachodzenie na siebie. Dużą liczbę linii oraz ich różnorodność w zakresie modelowanego pasma spektralnego powodują trudności w takich obliczeniach. W literaturze zaproponowano wiele modeli różniących się między sobą sposobem podejścia do obliczania statystycznych parametrów w paśmie. Do celów pomiarowych wykorzystuje się często model statystyczny SNB. Obliczanie parametrów modelu opiera się na wielokrotnym wyznaczeniu transmisyjności dla różnych głębokości optycznych i dopasowywaniu parametrów modelu metodą najmniejszych kwadratów. W obliczeniach dobiera się parametry tak, by transmisyjność zmieniała się od 0,05 do

0,95 co 0,05. Dzięki temu osiąga się dobrą dokładność modelowania dla różnych głębokości optycznych. Przykładowy wykres absorpcyjności dla modelu SNB przedstawiono na rysunku 5.



Rys. 5. Porównanie transmisyjności głównych gazowych składników spalania ($T = 400\text{ K}$, $l = 1\text{ cm}$, ułamek molowy zawartości dla każdego gazu 0,01), model SNB

Fig. 5. Transmissivity comparison of main combustion product ($T = 400\text{ K}$, $l = 1\text{ cm}$, mol fraction of each gas 0.01), SNB model

Średnie poszerzenie linii gazów zależne od temperatury i zawartości gazów może być przedstawione dla mieszaniny CO, CO₂, H₂O następującymi wzorami [33]:

$$\bar{\gamma}_{\text{CO}} = \frac{p}{p_s} \left[0,075 X_{\text{CO}_2} \left(\frac{T_s}{T} \right)^{0,6} + 0,12 X_{\text{H}_2\text{O}} \left(\frac{T_s}{T} \right)^{0,82} + 0,06 \left(\frac{T_s}{T} \right)^{0,7} (1 - X_{\text{CO}_2} - X_{\text{H}_2\text{O}}) \right], \quad (15)$$

$$\bar{\gamma}_{\text{CO}_2} = \frac{p}{p_s} \left(\frac{T_s}{T} \right)^{0,7} [0,07 X_{\text{CO}_2} + 0,1 X_{\text{H}_2\text{O}} + 0,058 (1 - X_{\text{CO}_2} - X_{\text{H}_2\text{O}})], \quad (16)$$

$$\bar{\gamma}_{\text{H}_2\text{O}} = \frac{p}{p_s} \left\{ 0,462 X_{\text{CO}_2} \frac{T_s}{T} + \left(\frac{T_s}{T} \right)^{0,5} [0,079 (1 - X_{\text{CO}_2} - X_{\text{O}_2}) + 0,106 X_{\text{CO}_2} + 0,036 X_{\text{O}_2}] \right\}, \quad (17)$$

gdzie $p_s = 1\text{ atm}$ oraz $T_s = 296\text{ K}$.

3.3. WŁAŚCIWOŚCI SPEKTRALNE LOTNYCH CZĄSTEK STAŁYCH

Jeżeli w badanym środowisku znajdują się lotne cząstki stałe, to powodują one absorpcję promieniowania proporcjonalną do sumy pól przekroju poprzecznego cząstek.

Jeśli założymy, że cząstki są kulkami o promieniu r , to transmisyjność warstwy l można opisać wzorem

$$\tau_{cs} = \exp(-\pi r^2 n_g l). \quad (18)$$

Ilość ziaren n_g drobin oblicza się wykorzystując gęstości materiału ρ [kg/m³], z jakiego składa się pył oraz koncentracji cząstek w mieszaninie μ [kg/m³] [20]:

$$n_g = \mu / \left(\frac{4}{3} \pi r^3 \rho \right). \quad (19)$$

Współczynnik pochłaniania wynosi odpowiednio

$$K_p = \pi r^2 n_g = \frac{3\mu}{4r\rho}. \quad (20)$$

Podane wzory są słuszne, jeśli założymy, że emisyjność powierzchni drobin jest równa jedności, co nie jest spełnione w warunkach rzeczywistych. Dodatkowo drobiny różnią się rozmiarami oraz kształtami. Do celów obliczeniowych stosowane są współczynniki korekcyjne, wyznaczone doświadczalnie.

W przypadku pomiarów transmisyjności płomieni pojawia się widmo o charakterze ciągłym, pochodzące głównie od sadzy. Dla płomieni z pyłu węglowego dominują cechy spektralne popiołu i żużla. Cząstki te różnią się w znaczny sposób rozmiarami, a ich koncentracja jest duża i zmienia się w znacznym stopniu. Obszar płomienia można podzielić na 3 części. Pył węglowy odgrywa dużą rolę w początkowym obszarze płomienia pyłowego. Parametry optyczne takiego medium zależą głównie od koncentracji składników oraz ich rozmiarów (rozkładu cząstek). Następny obszar zawiera węgiel odbarwiony, przekształcający się w ostateczne produkty spalania oraz sadzę. W strefie trzeciej dominuje popiół. Parametry spektralne węgla, sadzy, popiołu i żużla zależą bardzo znacznie od ich składu oraz rozkładu wielkości cząstek. Pomiary absorpcji promieniowania ponad strefą spalania dla obiektu przemysłowego pokazują, że nawet dla dużego zapylenia w widmie widoczne są cechy spektralne gazów.

Transmisyjność warstwy o rozmiarze l zawierającą sadzę można zapisać w zależności od współczynnika tłumienia K_s

$$\tau_\lambda = \exp(-K_s l). \quad (21)$$

Spektralny współczynnik tłumienia K_s zależy od liczby cząstek (o średniej średnicy D) w jednostce objętości N_s , przekrojów powierzchni dla absorpcji oraz rozpraszania, którego wartość do celów obliczeniowych często się zaniedbuje. Współczynnik absorpcji można wyrazić w postaci [13]

$$K_{\text{abs}} = N_s C_{\text{abs}} = f_V \frac{36\pi F(\lambda)}{\lambda}, \quad (22)$$

gdzie względna objętość sadzy f_V opisana jest następującym wzorem:

$$f_V = \frac{\pi D^3 N_s}{6}, \quad (23)$$

a $F(\lambda)$, która jest funkcją zespolonego współczynnika refrakcji ($m = n - jk$) zależnością

$$F(\lambda) = \frac{1}{6} \operatorname{Im} \left| \frac{m^2 - 1}{m^2 + 2} \right| = \frac{nk}{(n^2 - k^2 + 2)^2 + 4n^2 k^2}. \quad (24)$$

Z elektromagnetycznej teorii Lorenza–Mie [28] wyznacza się tzw. efektywny spektralny współczynnik absorpcji, czyli stosunek spektralnego przekroju absorpcji do przekroju fizycznego cząstki. Inaczej mówiąc, stosunek energii zaabsorbowanej przez cząsteczkę do energii docierającej do cząstki.

Naturalna droga absorpcji l_{nda} określa grubość warstwy sadzy o względnej objętości $f_V = 1$, której transmisyjność wynosi $1/e$

$$l_{\text{nda}} = \frac{f_V}{K_{\text{abs}}} = \frac{\lambda}{36\pi F(\lambda)}. \quad (25)$$

Względną zawartość sadzy można obliczyć ze współczynnika refrakcji i pomiarów transmisyjności, wykorzystując zależność

$$f_V = \ln(\tau_\lambda) \frac{\lambda}{36\pi F(\lambda)L} = \ln(\tau_\lambda) \frac{l_{\text{nda}}}{L}. \quad (26)$$

Empiryczne współczynniki opisujące parametry podawane są w literaturze, np. dla zakresu 300–800 nm [13] przyjmują postać:

$$n = 1,811 + 0,1263 \ln(\lambda) + 0,027[\ln(\lambda)]^2 + 0,0417[\ln(\lambda)]^3, \quad (27)$$

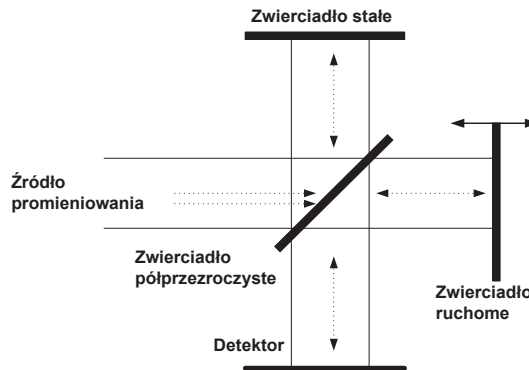
$$k = 0,5821 + 0,1213 \ln(\lambda) + 0,2309[\ln(\lambda)]^2 - 0,01[\ln(\lambda)]^3. \quad (28)$$

4. CHARAKTERYSTYKA SPEKTRALNA SPEKTROMETRU FTIR

Spektrometry FTIR cechuje dobra dokładność pomiarów oraz wysoka rozdzielczość. Zjawiskiem wykorzystywanym w spektrometrach FTIR jest interferencja dwóch promieni optycznych. Intensywność interferencji w funkcji różnicy długości

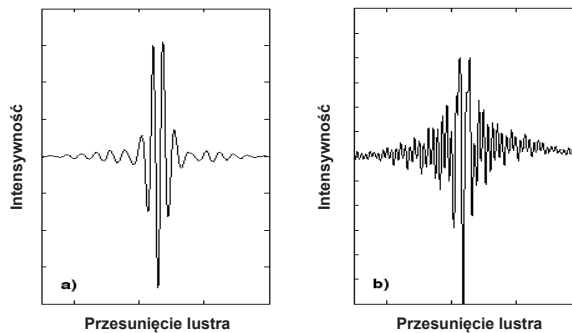
interferujących wiązek światła (przesunięcia ruchomego lustra interferometru) nosi nazwę interferogramu.

Główną częścią układu spektrometru jest interferometr Michelsona (rys. 6), który składa się z dzielnika światła, lustra stałego oraz lustra ruchomego. Po przejściu przez interferometr częstotliwości optyczne zostają zamienione na częstotliwości akustyczne. Promieniowanie ze źródła kierowane jest na półprzezroczyste lustro, gdzie w połowie zostaje odbite do lustra ruchomego, a w połowie przepuszczone do lustra stałego (lub odwrotnie). Odbite wiązki światła wracają na dzielnik, w którym zachodzi ich interferencja. Różnica długości dróg optycznych, powoduje różnicę faz pomiędzy wiązkami. W zależności od zmiany fazy następuje interferencja destruktywna bądź konstruktywna. Interferencja destruktywna zachodzi, gdy różnica dróg optycznych wynosi połowę długości fali. Interferogram zamieniany jest na widmo z wykorzystaniem dyskretnej transformaty Fouriera.



Rys. 6. Interferometr Michelsona

Fig. 6. Michelson interferometer



Rys. 7. Interferogram dla ciała czarnego (a)

oraz (b) interferogram dla dużej liczby linii gazowych (pomiar własny)

Fig. 7. Black body interferogram (a), and (b) interferogram for gas spectra lines

Maksymalna intensywność interferogramu występuje dla zerowego przesunięcia ruchomego zwierciadła (ZPD – *zero path difference*). Tylko w tym jednym miejscu wszystkie długości fali interferują konstruktywnie. W przypadku idealnego interferometru Michelsona, detektora i systemu próbkowania spektrum nie ma składowej urojonej. Jeżeli interferogram jest symetryczny względem środka, to obliczone na jego podstawie spektrum nie ma części urojonej. Rzeczywiste składniki urządzenia oraz ich niedoskonałości są przyczyną pojawienia się składowej urojonej. Główne powody zniekształceń fazowych interferogramu, to: przesunięcie fazy związane ze wzmacniaczem i filtrami elektronicznymi, błędy fazy powodowane przesunięciem próbek w stosunku do ich faktycznego położenia na ścieżce poruszania się ruchomego lustra interferometru, dyspersja podzielnika światła. Korekcja fazy umożliwia korekcję niesymetryczności interferogramu. Część optyczna oraz elektroniczna wprowadza systematyczne przesunięcie fazy, które można dosyć dokładnie określić. Szum w równy sposób wpływa na część rzeczywistą i urojoną widma.

Jeśli interferogram byłby rzeczywiście symetryczny, to wystarczyłoby zmierzyć jedynie jego połowę. W praktyce, aby dobrze określić zmiany fazy dla poszczególnych miejsc interferogramu, a więc również wyznaczyć charakterystyczne dla instrumentu przesunięcie fazowe, mierzy się interferogram dwustronny. Następnie oblicza się fazę dla specjalnie zmierzonego interferogramu o małej rozdzielczości, a więc dla małego przesunięcia zwierciadła ruchomego. Tak obliczony kąt fazowy Θ_λ służy do odtworzenia symetryczności interferogramów mierzonych dla wyższych rozdzielczości. Wyznaczany jest on z zależności

$$\Theta_\lambda = \arctg(\text{Im}(B(\lambda))/\text{Re}(B(\lambda))), \quad (29)$$

gdzie: (Re) – część rzeczywista widma, (Im) – część urojona widma, $B(\lambda)$ – widmo zespolone otrzymane z niesymetrycznego interferogramu.

W metodzie Formana korekcji fazy oblicza się odwrotną transformatę z obliczonego poprzednio kąta fazowego $\exp(j\Theta_\lambda)$, a następnie wykonuje się operację jej spłotu z interferogramem. Odtwarza to jego symetryczność.

Metoda Mertza korekcji fazy jest odpowiednikiem korekcji Formana, ale przeprowadzana w dziedzinie widma. Metoda ta jest bardziej efektywna obliczeniowo [19].

Minimalna częstotliwość próbkowania interferogramu zależy od szybkości poruszania się lustra interferometru oraz od maksymalnej liczby falowej, jaka jest mierzona. Według kryterium Nyquista, minimalna częstotliwość próbkowania musi być co najmniej dwa razy większa od maksymalnej częstotliwości sygnału. Jeśli oznaczymy największą mierzoną liczbę falową jako \mathcal{G}_{\max} , a prędkość poruszania się lustra jako v_l , to najwyższa częstotliwość fali interferogramu wynosi

$$f_{mi} = v_l \cdot \mathcal{G}_{\max} . \quad (30)$$

Interferogram jest funkcją położenia przesuwającego się zwierciadła. Zakładając, że zwierciadło ruchome porusza się ze stałą prędkością $v_l = \Delta y/\Delta t$, to fala o częstości

ω oraz wektorze falowym k_f podlega na ruchomym lustrze częstotliwościowemu przesunięciu Dopplera [14]

$$\Delta\omega = 2k_f v_l = (4\pi/\vartheta)v_l. \quad (31)$$

Uśrednioną intensywność promieniowania, padającą na detektor, można zapisać następującym wzorem [14]:

$$\bar{I} = (\bar{I}_0/2)[1 + \cos(\Delta\omega t)]. \quad (32)$$

Jeżeli prędkość poruszającego się zwierciadła jest znana, to na podstawie częstotliwości przesunięcia Dopplera można wyznaczyć częstotliwości fali padającej, tj.:

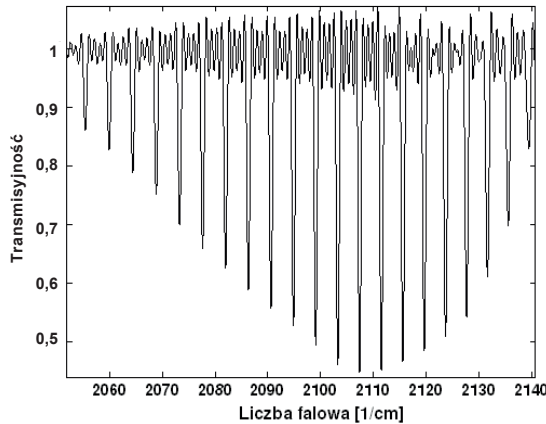
$$\omega = (c/V_l)\Delta\omega/2. \quad (33)$$

Jak widać ze wzoru (33) następuje przetworzenie wysokiej częstotliwości optycznej na częstotliwości z zakresu akustycznego. Warunkiem poprawnego przetwarzania jest jednostajne poruszanie się ruchomego lustra. Różnica dróg optycznych pomiędzy obydwoma wiązkami zapisana może być następująco:

$$\Delta s = \Delta s_0 + 2v_l t, \quad (34)$$

wynika z niej odpowiednia różnica fazy

$$\delta = (2\pi/\vartheta)\Delta s. \quad (35)$$



Rys. 8. Pojawienie się transmisyjności większej od jedności w wyniku zastosowania prostokątnego kształtu okna nałożonego na interferogram
Fig. 8. Rectangular window interferogram apodization effect on the transmissivity of gas

Dokładność wyznaczania różnicy długości dróg optycznych determinuje dokładność wyznaczania położenia widma względem liczb falowych. Najczęściej wykorzystuje się

laser He-Ne o długości fali 632,8 nm jako źródło odniesienia o dokładanie znanej długości fali. Promieniowanie lasera przechodzi przez interferometr i powoduje pojawienie się na detektorze sygnału o charakterze cosinusoidalnym. Różnica dróg optycznych, jaka występuje pomiędzy kolejnymi maksimami, odpowiada długości fali lasera.

Lustro ruchome porusza się po pewnej określonej drodze, stąd też sygnał na interferogramie jest ograniczony. Następstwem tego jest kształt linii spektrometru w postaci $\text{sinc}(x) = \sin(x)/x$ i znaczące listki boczne każdej linii spektralnej. Może to powodować w widmie wartość transmisyjności większą od jedności, bardzo trudną do interpretacji. Oczywiście fizycznie takie wartości nie występują, jednak pojawiają się one w mierzonym widmie jako cecha cyfrowego przetwarzania sygnałów.

Mnożenie interferogramu przez okno (funkcję apodyzującą) inne niż prostokątne powoduje osłabienie listków bocznych widma oraz niestety poszerzenie listka głównego. Z kształtem okna związana jest również rozdzielczość spektrometru. Można ją definiować na kilka sposobów w zależności od przyjętego kryterium. W spektroskopii często używa się kryterium Rayleigha, które mówi o tym, że dwa komponenty mogą być rozróżnione, jeżeli maksimum intensywności jednego nachodzi na pierwsze minimum intensywności drugiego. Dla spektrometrów FTIR rozdzielczość $\Delta\theta$ zależy od maksymalnej różnicy ścieżek optycznych Δs_{max} OPD (ang. *Optical Path Difference*). Przyjmuje się, że jest ona odwrotnie proporcjonalna do maksymalnej różnicy dróg optycznych interferometru

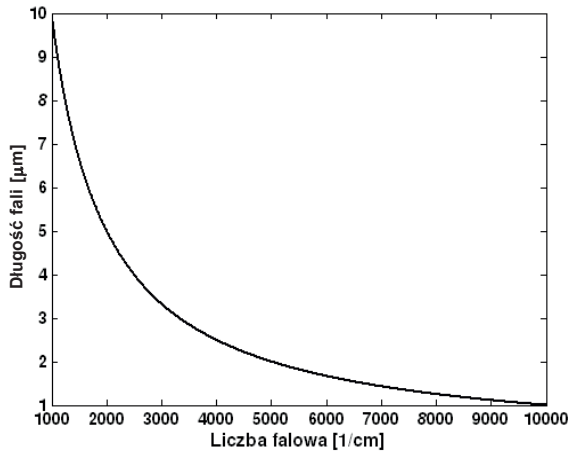
$$\Delta\theta = 1/\Delta s_{\text{max}} . \quad (36)$$

W oknie pomiarowym, innym niż prostokątnym nałożonym na interferogram następuje poszerzenie listka głównego widma okna, a w konsekwencji poszerzenie linii mierzonych w widmie. Takie dodatkowe poszerzenie związane ze zmianą kształtu okna powoduje, że poszczególne linie rotacyjne mogą się nakładać na siebie. Rozdzielczość spektrometru determinuje możliwości odróżniania dwu linii spektralnych leżących w widmie blisko siebie. Wysoka rozdzielczość polepsza możliwość lokalizacji linii spektralnych oraz ogranicza nakładanie się widm blisko położonych pasm lub też poszczególnych linii. Kosztem polepszenia rozdzielczości jest wzrost szumu i zmniejszenie precyzji analizy. Mniejsze rozdzielczości analizy dla tego samego spektrometru umożliwiają uzyskanie lepszego stosunku sygnału do szumu SNR oraz zwiększenie dynamiki pomiaru. Dodatkowo zmniejsza się również liczba analizowanych danych. W analizie ilościowej większe znaczenie od rozdzielczości ma SNR. Pomiar z wysoką rozdzielczością może ułatwić odróżnienie linii rotacyjnych poszczególnych gazów w nakładających się pasmach. Spektrometry FTIR mają pewne zalety w stosunku do spektrometrów dyspersyjnych. Główną przewagą FTIR jest, przy tej samej rozdzielczości pomiaru, wielokrotnie większy SNR. To samo widmo jest mierzone w wielokrotnie krótszym czasie w stosunku do pomiaru spektrometrem dyspersyjnym. Większa szybkość pomiaru związana jest z jednoczesnym pomiarem widma

dla wszystkich liczb falowych. Interferogram jest funkcją przemieszczenia ruchomego lustra wyrażaną w cm, stąd też widmo FTIR, będące transformatą Fouriera interferogramu, przedstawiane jest w funkcji liczbą falowej o jednostce 1/cm.

Tab. 1. Porównanie liczb falowych i długości fal
Tab. 1. Wavelength and wavenumber comparison

Liczba falowa (cm ⁻¹)	Długość fal (μm)	Rozdzielczość (cm ⁻¹)	Rozdzielczość (nm)
20000	0,5	1	0,025
10000	1	1	0,1
5000	2	1	0,4
3333,3	3	1	0,9
2500	4	1	1,6
2000	5	1	2,5
1000	10	1	10



Rys. 9. Zależność długości fali od liczby falowych
Fig. 9. Relationship between wavenumber and wavelength

Jeśli promieniowanie docierające do spektrometru oznaczymy jako $W(\vartheta)$, to interferogram $I_s(x)$ można zapisać jako

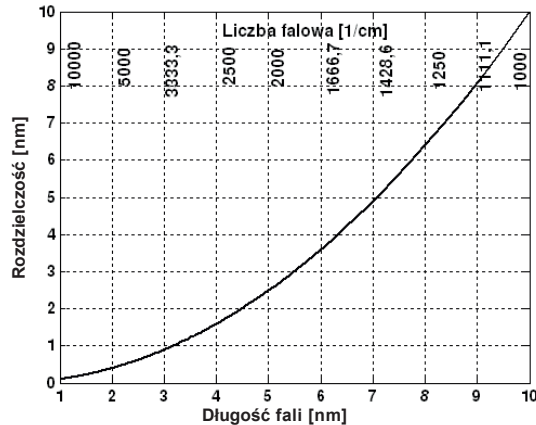
$$I_s(x) = k_s \int_0^{+\infty} W(\vartheta) \{1 + \cos(2\pi\vartheta x)\} d\vartheta, \quad (37)$$

gdzie k_s jest współczynnikiem charakterystycznym dla każdego interferometru, przepuszczalności dzielnika światła oraz współczynników odbicia lusterek. Informacje niesie jedynie część interferogramu oznaczona jako $I(x)$, która podlega zmianom wraz

z przesunięciem się ruchomego zwierciadła. Widmo otrzymane w wyniku pomiaru zapisać można następująco:

$$W_s(\nu) = \int_{-\infty}^{+\infty} I(x) \cdot o(x) \cdot \exp(-2\pi j\nu x) dx, \quad (38)$$

gdzie $o(x)$ jest kształtem okna.



Rys. 10. Rozdzielczość dla długości fali [nm] odpowiadająca rozdzielczości 1 [1/cm] przy liczbach falowych

Fig. 10. Resolution in wavenumber and wavelength

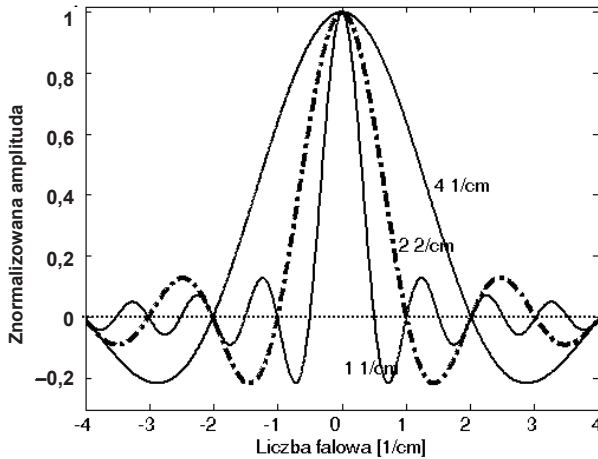
Widmo mierzone różni się od widma rzeczywistego ze względu na charakter spektralny interferometru $o(x)$. Widmo mierzone w ogólnym przypadku obliczane z wykorzystaniem transformaty Fouriera będzie zespolone ze względu na wspomniane wcześniej zjawisko dodatkowego przesunięcia fazowego. Po operacji korekcji fazy obliczać można jedynie cosinusową część transformaty Fouriera. Funkcja $o(x)$ jest związana z maksymalną różnicą dróg optycznych Δs_{\max} . W najprostszym przypadku jest to funkcja o stałej wartości dla różnic pomiędzy drogami optycznymi $s \in \langle -s_{\text{im}}, s_{\text{im}} \rangle$; określa ona charakter odpowiedzi widmowej spektrometru na wymuszenie monochromatyczne. Kształt linii spektrometru (ILS – ang. *instrumental line shape* [4], [5]) powoduje zniekształcenie widma mierzonego w stosunku do widma rzeczywistego. Zamiast funkcji stałej stosuje się wiele innych rodzajów funkcji symetrycznych. Funkcje te nazywane są również funkcjami apodyzującymi (ang. *apodization function*) oraz oknami. Z właściwości transformaty Fouriera wynika, że mnożenie w dziedzinie drogi (czasu) odpowiada operacji splotu w dziedzinie widma. Wpływ charakteru okna (funkcji apodyzującej) na widmo może zostać łatwo określony. Wid-

mo mierzone przez spektrometr $W(\nu)$ zależy od widma docierającego do spektrometru $W_s(\nu)$ oraz funkcji okna o widmie $O(\nu)$

$$W(\nu) = W_s(\nu) \otimes O(\nu). \quad (39)$$

Jest to operacja splotu dwóch funkcji w dziedzinie widma.

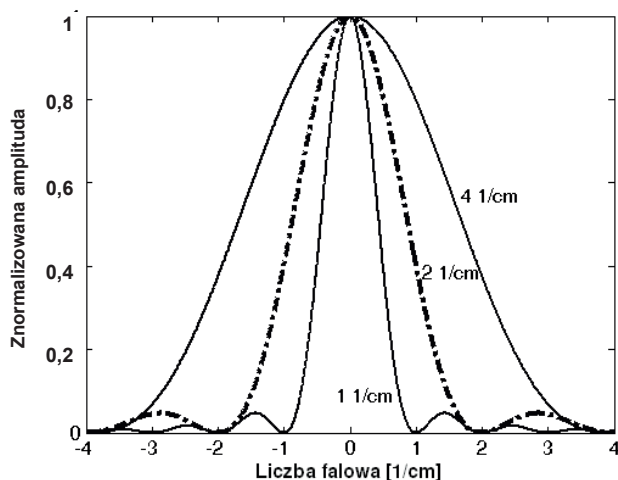
Przez wybór kształtu okna kształtować można charakterystykę odpowiedzi spektrometru na monochromatyczną linię spektralną. Zastosowanie okna prostokątnego daje najlepsze możliwości rozróżnienia dwóch linii leżących blisko siebie. Niestety dla tego okna pojawiają się znacznej wielkości listki boczne. Mogą one powodować występowanie wspomnianych wartości transmisyjności większych od jedności oraz utrudniać analizę linii spektralnych o mniejszej intensywności, znajdujących się w pobliżu linii o dużej intensywności. Znaczenie tego zjawiska rośnie szczególnie w obszarze dużej gęstości interpretowanych linii.



Rys. 11. Porównanie rozdzielczości spektrometru dla okien prostokątnych
Fig. 11. Spectrometer resolution comparison for rectangular apodization

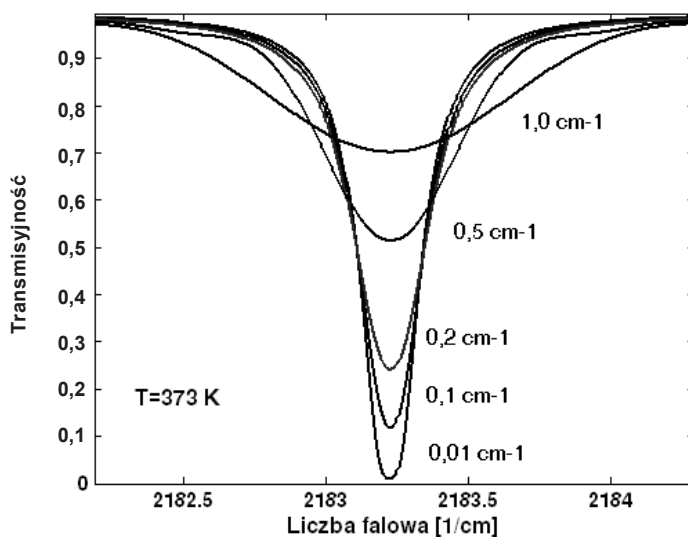
Niektórzy producenci spektrometrów definiują rozdzielczość jako szerokość półkową monochromatycznej linii spektralnej w przypadku zastosowania trójkątnej funkcji okna. Okno trójkątne ma 1,5-krotnie szersze widmo listki głównego w stosunku do okna prostokątnego. Jeżeli interferogram podlega jedynie naturalnemu ograniczeniu, czyli obcięciu do długości $s \in \langle -s_{im}, s_{im} \rangle$, to widmo monochromatycznej linii (rys. 12), mierzonej przez spektrometr, określone jest następującym wzorem:

$$O(\nu) = \int_{-s_{im}}^{s_{im}} e^{-j2\pi s} ds. \quad (40)$$

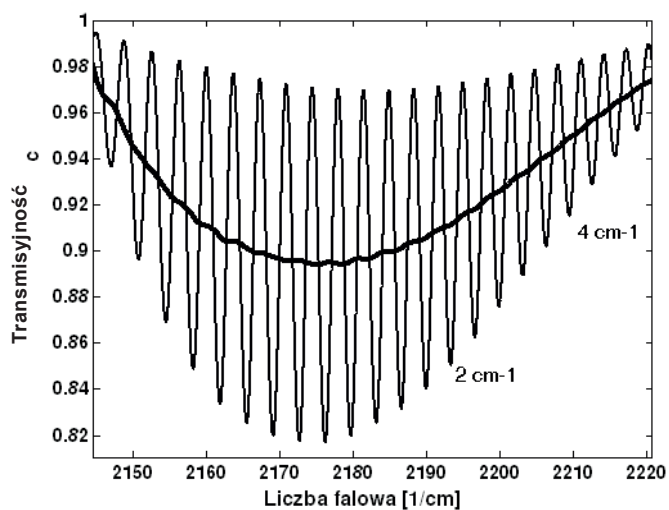


Rys. 12. Porównanie rozdzielczości spektrometru dla okien trójkątnych
Fig. 12. Spectrometer resolution comparison for triangle apodization

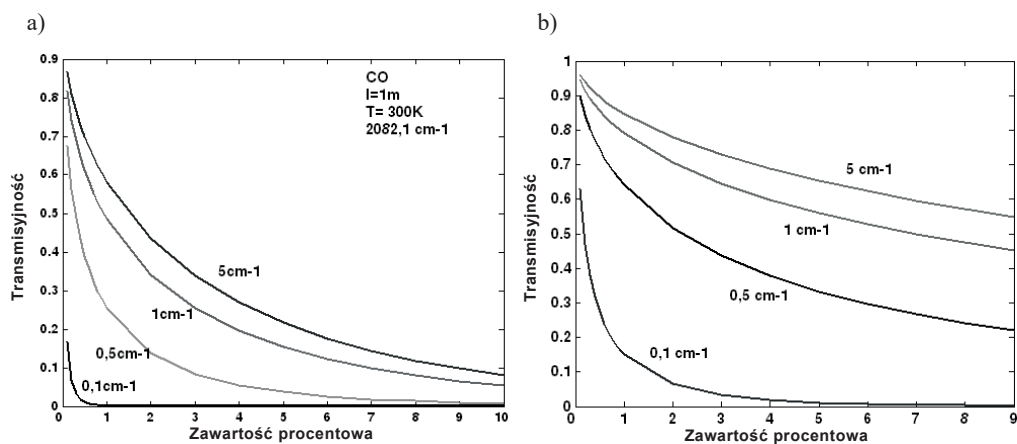
Dla pokazania wpływu rozdzielczości na kształt mierzonych linii, koncentracji i temperatury na transmisyjność i absorbcję przeprowadzono obliczenia dla CO, CO₂, H₂O oraz NO na rysunkach 13–31.



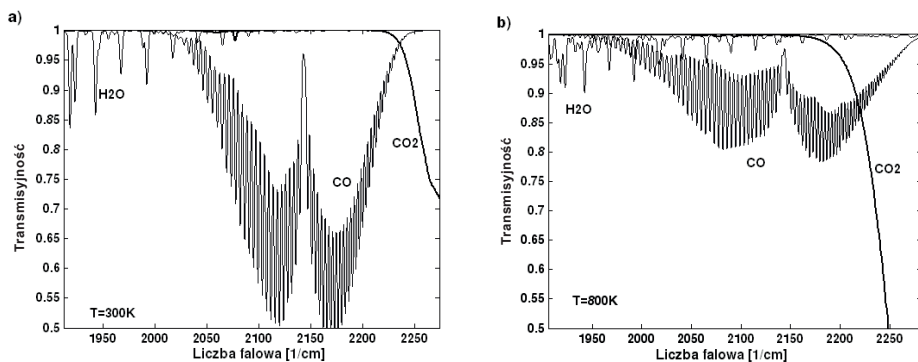
Rys. 13. Wpływ rozdzielczości na kształt mierzonej linii rotacyjnej dla tlenku węgla
Fig. 13. Resolution influence on the shape of measured carbon monoxide rotational line



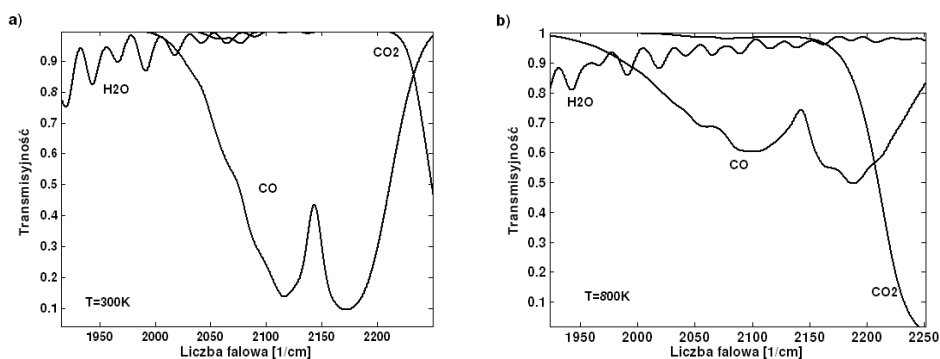
Rys. 14. Wpływ rozdzielczości na pomiar,
brak widoczności linii rotacyjnych dla tlenku węgla
Fig. 14. Resolution influence on the shape of measured rotational line,
vanishing of rotational shape



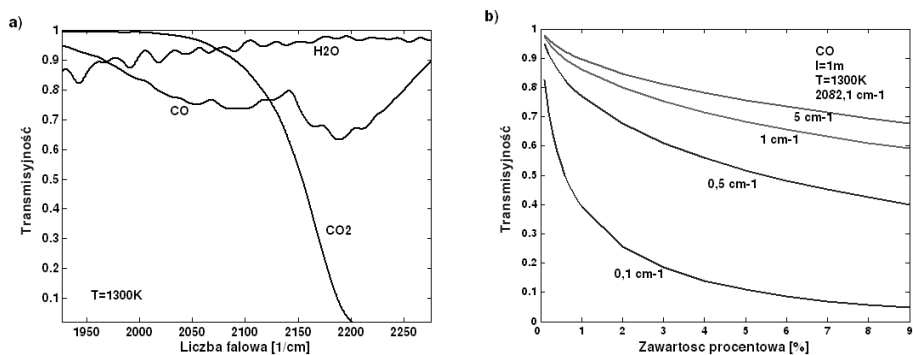
Rys. 15. Zależność transmisyjności CO na długości fali $2082,1 \text{ cm}^{-1}$ od zawartości procentowej dla różnych rozdzielczości pomiaru FTIR, przy temperaturze 300 K oraz 1300 K
Fig. 15. Transmissivity versus concentration of CO gas at $2082,1 \text{ cm}^{-1}$ wavenumber.
Four FTIR resolution, temperature of 300 K and 1300 K



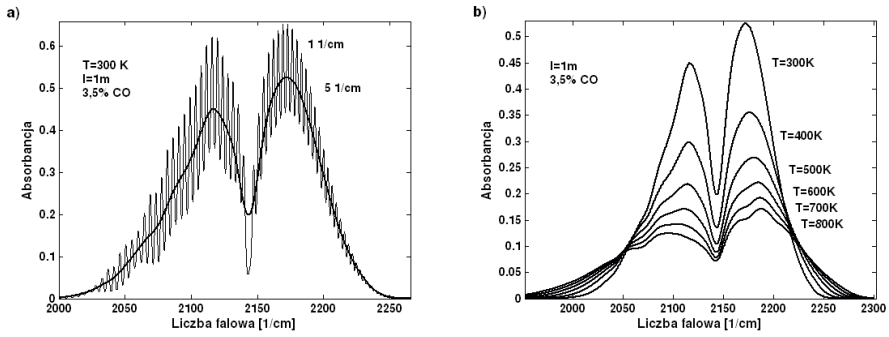
Rys. 16. Transmisyjność CO, CO₂ oraz H₂O dla dwu temperatur, rozdzielczość 1 cm⁻¹
 Fig. 16. Spectrum of CO, CO₂, H₂O for 300 K and 800 K, 1 cm⁻¹ resolution



Rys. 17. Transmisyjność CO, CO₂ oraz H₂O dla dwu temperatur, rozdzielczość 5 cm⁻¹
 Fig. 17. Spectrum of CO, CO₂, H₂O for 300 K and 800 K, 5 cm⁻¹ resolution

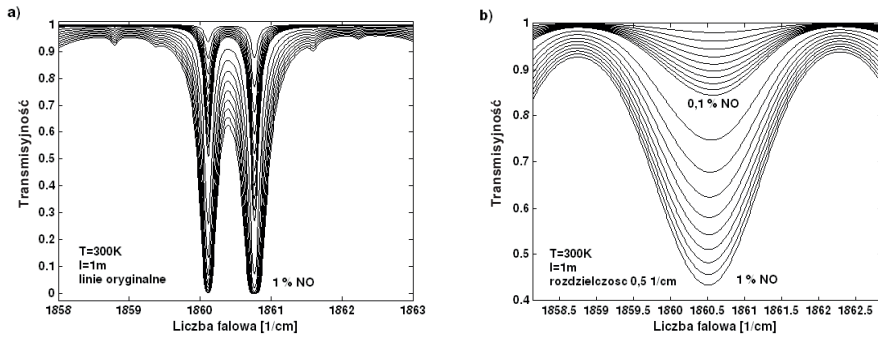


Rys. 18. Transmisyjność CO, CO₂ oraz H₂O dla $T = 1300$ K, wpływ rozdzielczości na transmisyjność
 Fig. 18. Spectrum of CO, CO₂, H₂O for 300 K and 800 K, 1 cm⁻¹ resolution. Transmissivity versus concentration of CO gas at 2082.1 cm⁻¹ wavenumber. Four FTIR resolution, temperature of 1300 K



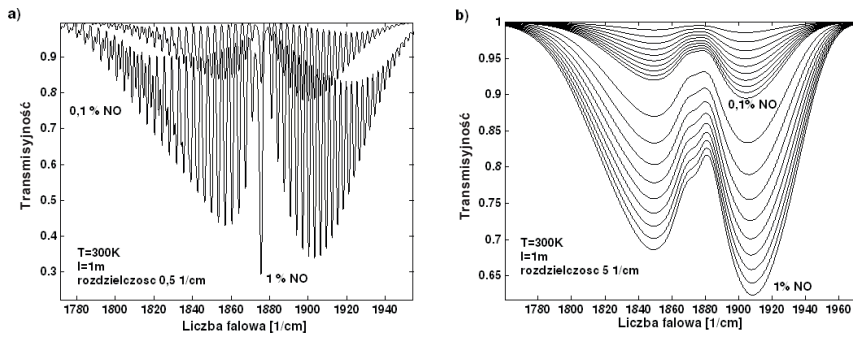
Rys. 19. Absorbancja CO $l = 300$ K dla dwu rozdzielczości, wpływ temperatury na widmo dla rozdzielczości 5 cm^{-1}

Fig. 19. a – Spectrum of CO for 300 K, 1 and 5 cm^{-1} resolution, b – influence of temperature on CO spectrum, 5 cm^{-1} resolution



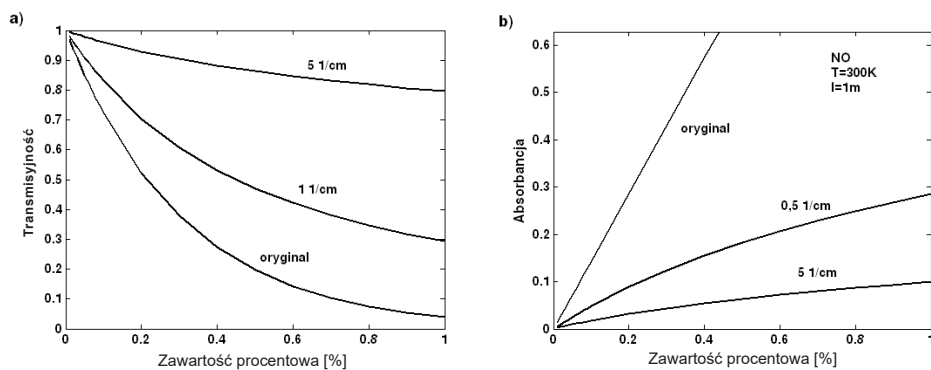
Rys. 20. Zależność transmisyjności NO $T = 300$ K dla zawartości 0,01–1%, linie oryginalne oraz rozdzielczość $0,5 \text{ cm}^{-1}$

Fig. 20. NO spectrum, $T = 300$ K, 0.01–1% volume fraction, original lines and 0.5 cm^{-1} resolution

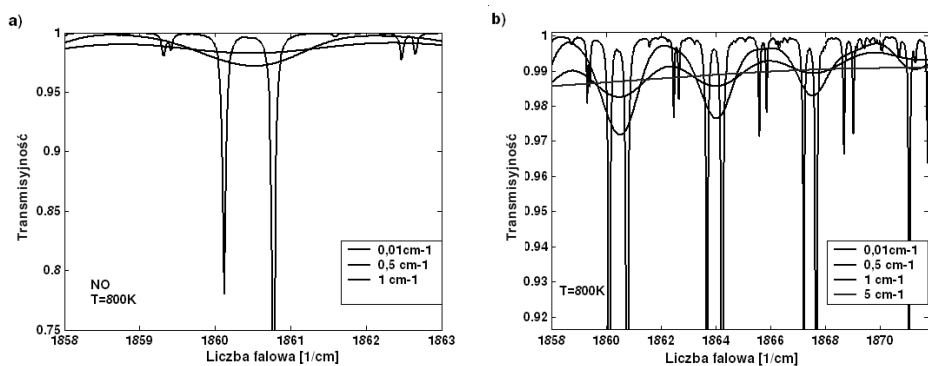


Rys. 21. Zależność transmisyjności NO $T = 300$ K dla zawartości 0,01–0,1–1%, dla rozdzielczość $0,5 \text{ cm}^{-1}$ oraz 5 cm^{-1}

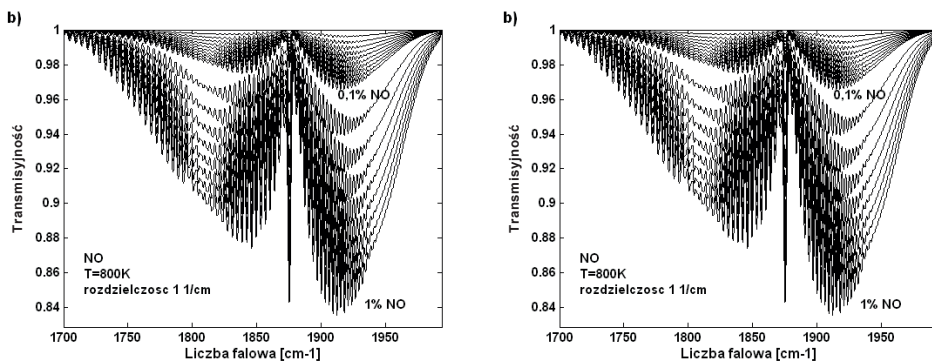
Fig. 21. NO spectrum, $T = 300$ K, 0.01–1% volume fraction, 0.5 cm^{-1} and 5 cm^{-1} resolution



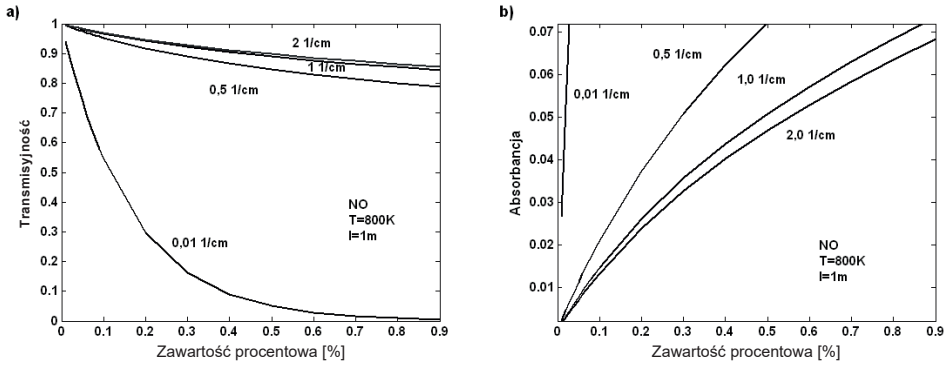
Rys. 22. Transmisyjność oraz absorbancja NO, $T = 300$ K, długość fali $1908,2 \text{ cm}^{-1}$
 Fig. 22. Transmissivity nad absorbance versus concentration of NO at $1908,2 \text{ cm}^{-1}$ wavenumber



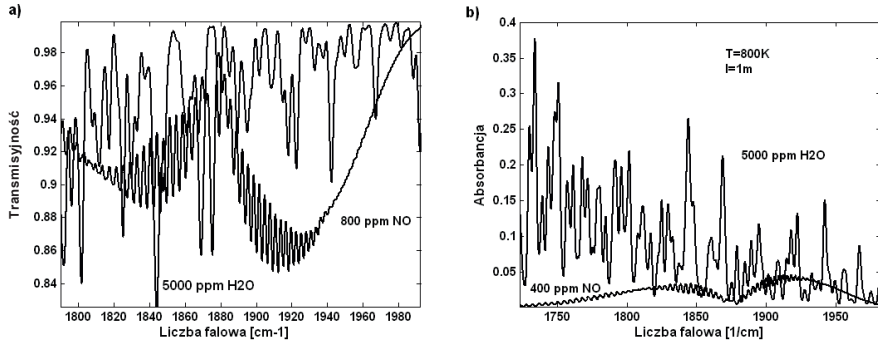
Rys. 23. Transmisyjność NO, $T = 800$ K dla różnych rozdzielności
 Fig. 23. Effect of resolution on NO transmissivity spectrum



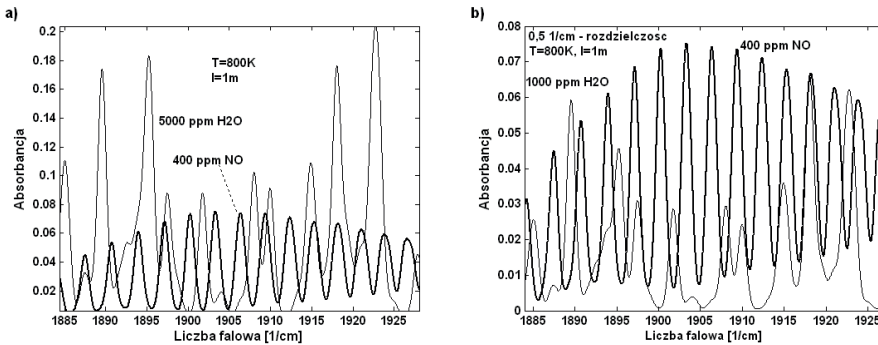
Rys. 24. Widmo NO, $T = 800$ K, linie oryginalne oraz dla rozdzielności 1 cm^{-1}
 Fig. 24. NO spectrum, $T = 800$ K, original lines and 1 cm^{-1} resolution



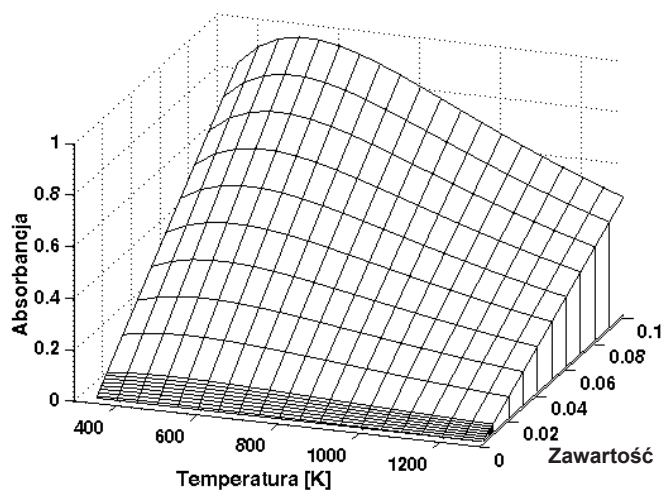
Rys. 25. Transmisyjność oraz absorbancja NO, $T = 800 \text{ K}$, długość fali 1915 cm^{-1}
 Fig. 25. Transmissivity nad absorbance versus concentration of NO at 1908.2 cm^{-1} wavenumber.
 Four FTIR resolution, $T = 800 \text{ K}$



Rys. 26. Porównanie widm transmisyjności i absorbancji NO oraz H₂O, $T = 800 \text{ K}$, rozdzielczość 1 cm^{-1}
 Fig. 26. Comparison of NO and H₂O spectrum, $T = 800 \text{ K}$, 1 cm^{-1} resolution

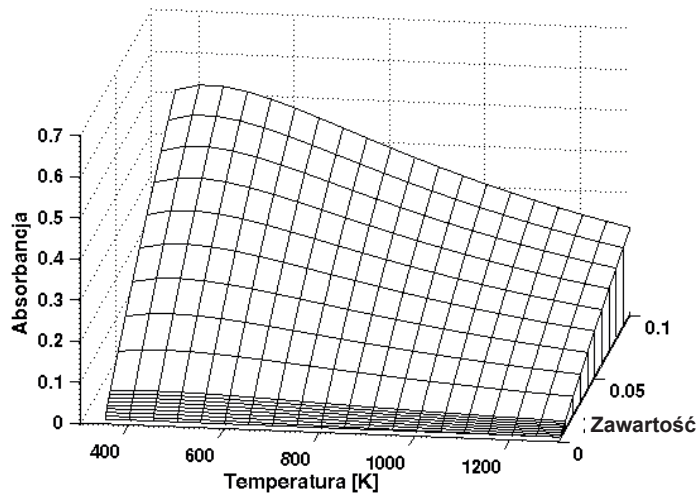


Rys. 27. Porównanie absorbancji NO oraz H₂O, $T = 800 \text{ K}$, rozdzielczość $0,5 \text{ cm}^{-1}$
 Fig. 27. Comparison of NO and H₂O spectrum, $T = 800 \text{ K}$, $0,5 \text{ cm}^{-1}$ resolution



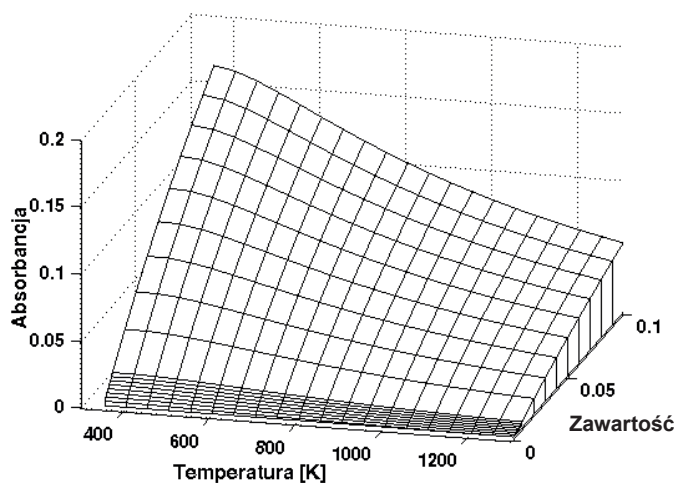
Rys. 28. Zależność absorbancji linii CO $2077,7 \text{ cm}^{-1}$ od temperatury oraz zawartości (ułamek objętościowy, linia oryginalna)

Fig. 28. Absorbance versus concentration and temperature of CO at 2077.7 cm^{-1} wavenumber (volume fraction, original line)

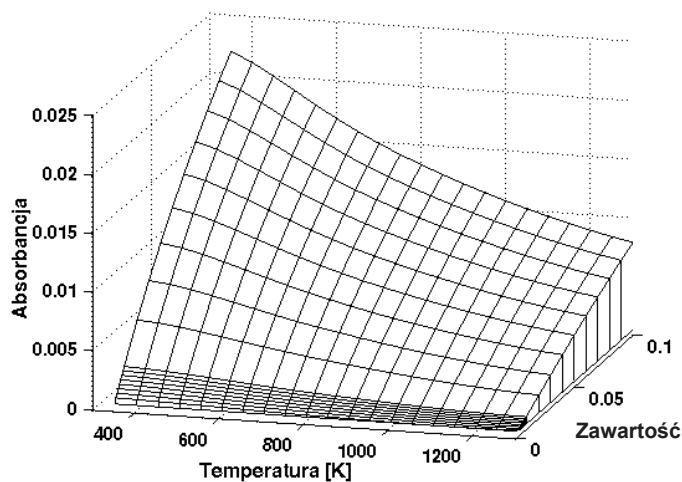


Rys. 29. Zależność absorbancji linii CO $2077,7 \text{ cm}^{-1}$ od temperatury oraz zawartości (ułamek objętościowy, rozdzielczość $0,1 \text{ cm}^{-1}$, apodyzacja trójkątna)

Fig. 29. Absorbance versus concentration and temperature of CO $2077,7 \text{ cm}^{-1}$ at wavenumber (volume fraction, 0.1 cm^{-1} resolution, triangle apodization)



Rys. 30. Zależność absorbancji linii CO $\nu = 2077,7 \text{ cm}^{-1}$ od temperatury oraz zawartości (ułamek objętościowy, rozdzielczość $0,5 \text{ cm}^{-1}$, apodyzacja trójkątna)
 Fig. 30. Absorbance versus concentration and temperature of CO at 2077.7 cm^{-1} wavenumber (volume fraction, 0.5 cm^{-1} resolution, triangle apodization)



Rys. 31. Zależność absorbancji linii CO $2077,7 \text{ cm}^{-1}$ od temperatury oraz zawartości (ułamek objętościowy, rozdzielczość 5 cm^{-1} , apodyzacja trójkątna)
 Fig. 31 Absorbance versus concentration and temperature of CO at 2077.7 cm^{-1} wavenumber (volume fraction, 5 cm^{-1} resolution, triangle apodization)

W spektrometrach FTIR mogą być zastosowane różnego rodzaju detektory. Do głównych zaliczyć można:

- detektory termiczne, czułe na zmianę temperatury detektora, powodowaną absorpcją promieniowania; są one niestety dość wolne,

- detektory kwantowe (np. MCT – *mercury cadmium tellurium*), które wykorzystują zjawisko przeniesienia elektronów na wyższe poziomy energii pod wpływem promieniowania; niestety do swojej pracy wymagają chłodzenia ciekłym azotem.

Spektrometr o otwartej ścieżce optycznej jest zbudowany nieco inaczej niż typowe spektrometry. Nie ma on szerokopasmowego źródła promieniowania, gdyż analizowane jest promieniowanie przychodzące z zewnątrz.

5. MODELOWANIE WIDM EMISYJNYCH I TRANSPORTU ENERGII PROMIENISTEJ

Promieniowanie cieplne jest promieniowaniem elektromagnetycznym, w zakresie od 0,1 do 1000 μm , emitowane przez ciała o temperaturze wyższej od zera bezwzględnego [21].

Strumień energii radiacyjnej to przepływ energii radiacyjnej Q w czasie t przez określoną powierzchnię F , co można opisać zależnością

$$\dot{Q} = \partial Q / \partial t. \quad (41)$$

Gęstość strumienia energii radiacyjnej określamy wzorem

$$\dot{q} = \partial \dot{Q} / \partial F. \quad (42)$$

Intensywnością promieniowania (w skrócie intensywność) nazywamy gęstość strumienia energii radiacyjnej w pewnym kącie bryłowym Ω , co można zapisać

$$I = \partial \dot{q} / \partial \Omega. \quad (43)$$

Monochromatyczną gęstość emisji powierzchni ciała doskonale czarnego (widmo lub rozkład spektralny promieniowania) określamy zależnością zwaną prawem Plancka [32]

$$e_{c\lambda}(T) = \frac{2\pi c_1 \lambda^{-5}}{e^{c_2/\lambda T} - 1}, \quad (44)$$

gdzie: λ – długość fali, T – temperatura, k – stała Boltzmanna, h – stała Plancka, c_1, c_2 – stałe, których wartości są następujące:

$$c_1 = hc^2 = 0,59544 \cdot 10^{-16}, \quad \text{Wm}^2, \quad (45)$$

$$c_2 = hc/k = 1,4388 \cdot 10^{-2}, \quad \text{mK}. \quad (46)$$

W radiometrii stosuje się wielkość zwaną energetyczną luminancją spektralną, która jest określona zależnością [6]

$$L_{\lambda}(T) = \frac{2c_1 \lambda^{-5}}{e^{c_2/\lambda T} - 1}, \quad \text{W}/(\text{cm}^2 \text{ sr } \mu\text{m}). \quad (47)$$

Maksimum spektralnego rozkładu emisji dla danej długości fali można obliczyć, wykorzystując prawo przesunięć Wiena

$$\lambda_{\text{max}} T = 2898, \quad \mu\text{m K}. \quad (48)$$

Prawo Stefana–Boltzmana opisuje całkowitą energię promieniowania ciała doskonale czarnego, a przyjmuje postać

$$\dot{e}_c = \sigma_c T^4, \quad (49)$$

gdzie: $\sigma_c = 5,67 \cdot 10^{-8}$, $\text{Wm}^{-2}\text{K}^{-4}$ – stała Stefana–Boltzmana.

Jeśli w ośrodku nie zachodzi rozpraszanie energii, zmianę intensywności promieniowania na określonej długości fali, wynikającą z absorpcji, zapisać można następująco

$$dI_{\lambda} = -\alpha_{\lambda} I_{\lambda} dl, \quad (50)$$

$$\alpha_{\lambda} = \alpha_{\lambda}(T, p, C_i). \quad (51)$$

gdzie: α_{λ} – lokalny monochromatyczny współczynnik pochłaniania zależy od temperatury, ciśnienia oraz koncentracji poszczególnych składników.

Intensywność promieniowania wyrazić można następująco:

$$I_{\lambda}(l) = I_{\lambda}(0) \exp \left[- \int_0^l \alpha_{\lambda}(l') dl' \right]. \quad (52)$$

Dla ośrodka jednorodnego prawo Bouguera–Lamberta przyjmuje postać

$$I_{\lambda}(l) = I_{\lambda}(0) \exp(-\alpha_{\lambda} l). \quad (53)$$

gdzie transmisyjność można zapisać jako

$$\tau_{\lambda} = \exp(-\alpha_{\lambda} l). \quad (54)$$

Absorpcyjność monochromatyczną jednorodnego ośrodka można wyznaczyć z zależności

$$a_{\lambda} = 1 - \exp(-\alpha_{\lambda} l), \quad (55)$$

Zdefiniować można również grubość optyczną

$$\kappa_{\lambda}(L) = \int_0^L \alpha_{\lambda}(l) dl, \quad (56)$$

która dla ośrodka jednorodnego przyjmuje postać

$$\kappa_{\lambda}(L) = \alpha_{\lambda}L. \quad (57)$$

Jeżeli jednorodny ośrodek ma temperaturę wyższą niż otoczenie to promieniuje energię. Emisyjność takiego ośrodka zapisać można następująco:

$$\varepsilon_{\lambda} = \exp(-\alpha_{\lambda}l). \quad (58)$$

Intensywność promieniowania, z uwzględnieniem prawa Plancka $L_{\lambda}(T)$, wyrazić można następująco:

$$I_{\lambda} = \varepsilon_{\lambda}L_{\lambda}(T). \quad (59)$$

Wyrażenie (59) stosowane jest do obliczania temperatury oraz zawartości sadzy w pomiarach pirometrycznych procesów spalania. Występują dwa podstawowe rodzaje pirometrów – pirometry całkowitego promieniowania oraz pirometry częściowego promieniowania. Pirometry częściowego promieniowania dzielimy na: monochromatyczne (jednobarwowe), wykorzystujące bardzo wąskie pasmo spektralne, dwubarwowe, wykorzystujące stosunek natężenia promieniowania na dwóch długościach fal, wielobarwne, wykorzystujące natężenie promieniowania na kilku długościach fal.

W procesach spalania najczęściej wykorzystuje się zakres długości fal (400–1000 nm). W tym zakresie pomiarowym intensywność promieniowania ciała doskonale czarnego można zastąpić prawem Wiena, wtedy, uwzględniając emisyjność ośrodka ε_{λ} , otrzymuje się

$$I_{\lambda,T} = \varepsilon_{\lambda}(c_1/\lambda^5)\exp(-c_2/\lambda T). \quad (60)$$

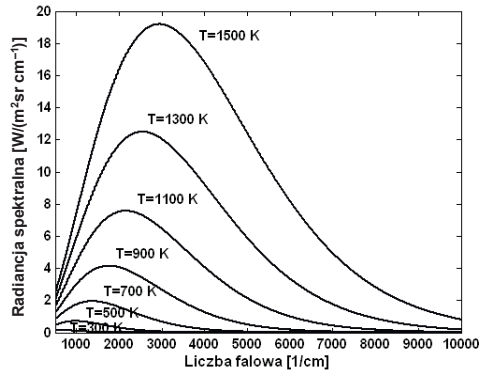
Dzieląc przez siebie intensywność promieniowania na dwu długościach fal λ_1 oraz λ_2 , otrzymuje się zależność

$$A = \left(\frac{I_1}{I_2} \right) = \left(\frac{\varepsilon_1}{\varepsilon_2} \right) \frac{\lambda_2^5 (\exp(c_2/\lambda_2 T))}{\lambda_1^5 (\exp(c_2/\lambda_1 T))}. \quad (61)$$

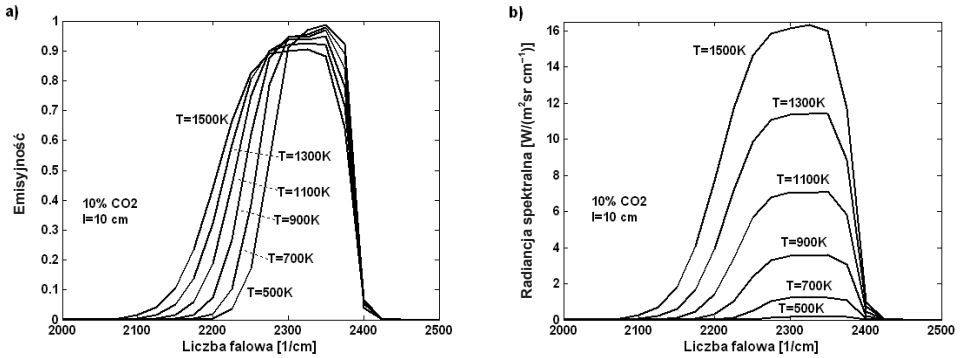
Jeżeli mierzone długości fali leżą blisko siebie, to przyjmuje się, że w danym zakresie emisyjność jest stała, tj. $\varepsilon_1 = \varepsilon_2$. Temperaturę można wtedy wyznaczyć z następującego wzoru:

$$T = \frac{c_2(1/\lambda_2 - 1/\lambda_1)}{(\ln(I_1/I_2) + 5 \ln(\lambda_1/\lambda_2))}. \quad (62)$$

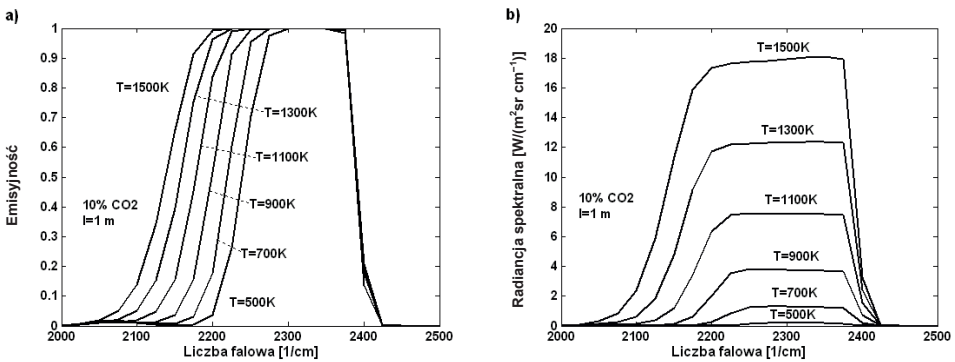
Dla zobrazowania obliczeń wyniki przedstawiono na rysunkach 32÷38.



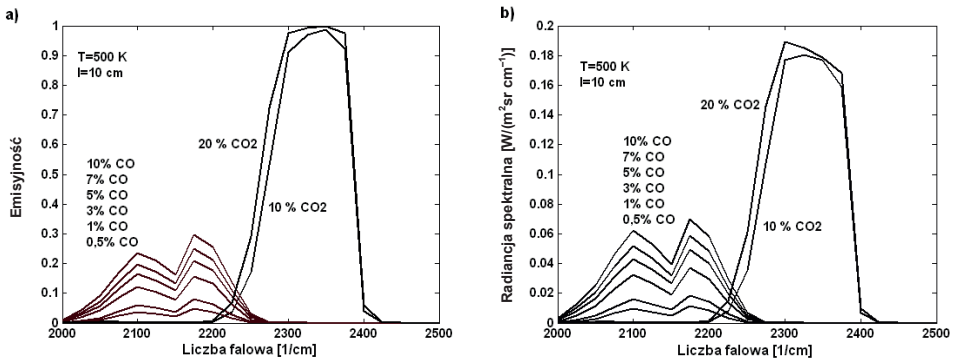
Rys. 32. Radiacja spektralna dla ciała doskonale czarnego
Fig. 32. Black body spectral radiance



Rys. 33. Emisyjność oraz widmo promieniowania CO₂ (10%, 10 cm) dla różnych temperatur
Fig. 33. Emissivity and emission spectrum CO₂ (10%, 10 cm) at different temperature

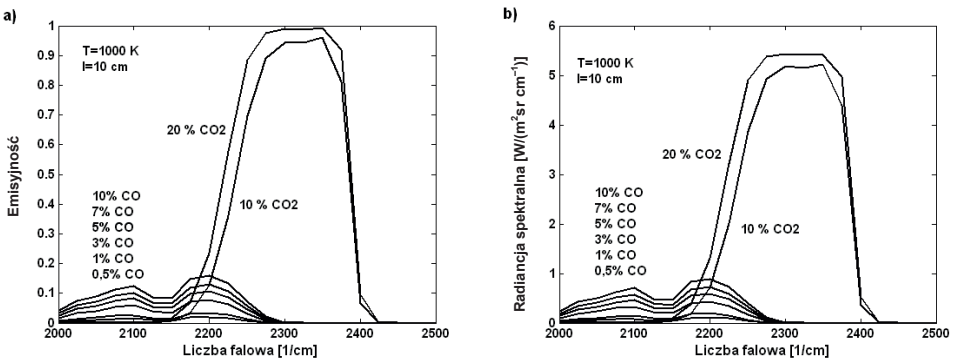


Rys. 34. Emisyjność oraz widmo promieniowania CO₂ (10%, 1 m) dla różnych temperatur
Fig. 34. Emissivity and emission spectrum CO₂ (10%, 1 m) at different temperature



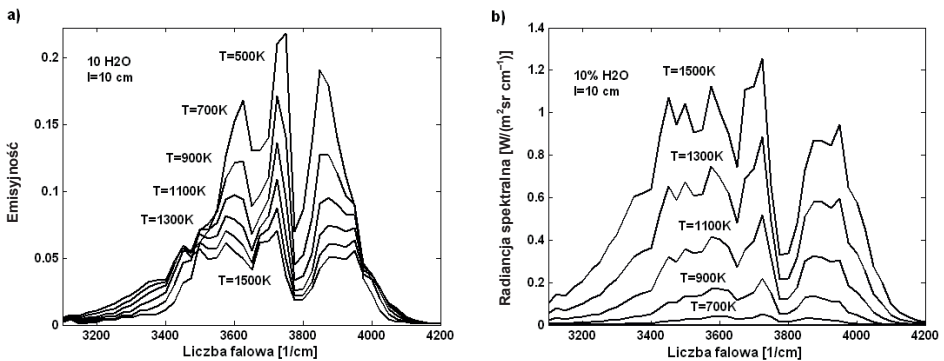
Rys. 35. Emisyjność oraz widmo promieniowania CO₂ i CO (10 cm, $T=500$ K)

Fig. 35. Emissivity and emission spectrum CO₂ (10 cm, $T=500$ K)



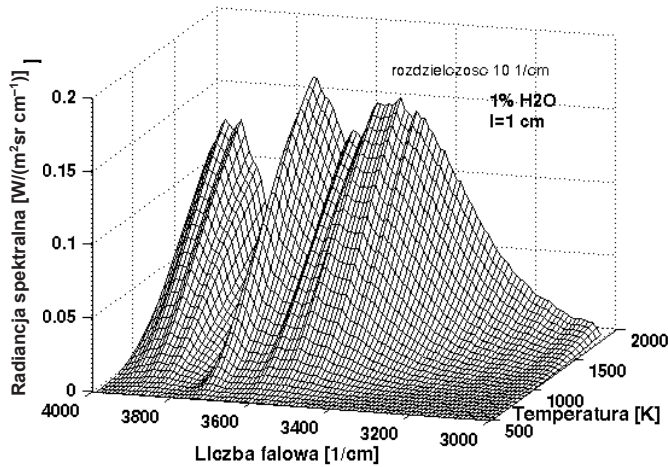
Rys. 36. Emisyjność oraz widmo promieniowania CO₂ i CO (10 cm, $T=1000$ K)

Fig. 36. Emissivity and emission spectrum CO₂ and CO (10 cm, $T=1000$ K)



Rys. 37. Emisyjność oraz widmo promieniowania H₂O (10%, 10 cm) dla różnych temperatur

Fig. 37. Emissivity and emission spectrum H₂O (10%, 10 cm) at different temperature



Rys. 38. Radiancja spektralna H₂O (1%, 1 cm) dla różnych temperatur
 Fig. 38. Emission spectrum, H₂O (1%, 1 cm) at different temperature

Do obliczania intensywności promieniowania niejednorodnych obszarów promieniujących używa się równania transportu energii radiacyjnej (transportu energii promienistej lub też transportu promieniowania). Równanie wyprowadzić można wykorzystując prawo Bouguera–Lamberta, w którym uwzględniamy tłumienie, jak też emisję promieniowania następującą w ośrodku.

W ośrodku bez rozpraszania intensywność promieniowania w jednym kierunku można przedstawić następująco:

$$I_{\lambda}(l + dl) = I_{\lambda}(l) + dI_{\lambda} . \quad (63)$$

Na zmianę intensywności promieniowania dI_{λ} wpływ ma zarówno pochłanianie, jak też proces emisji. Można więc te zmiany opisać zależnościami

$$dI_{\lambda a} = -\alpha_{\lambda}(l)I_{\lambda}(l)dl , \quad (64)$$

$$dI_{\lambda e} = \alpha_{\lambda}(l)B(T, l)_{\lambda} dl , \quad (65)$$

gdzie $B_{\lambda}(T, l)$ – intensywność promieniowania ciała doskonale czarnego w temperaturze T warstwy dl .

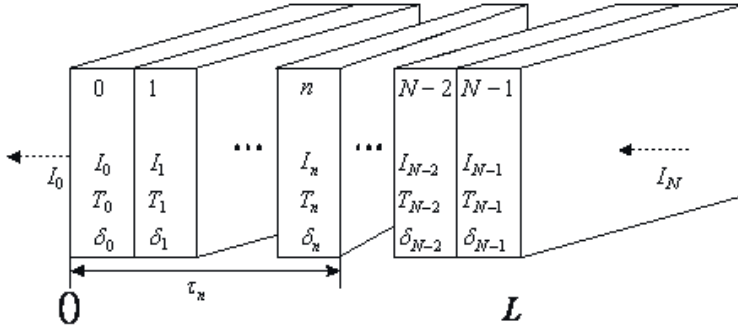
Zmiana intensywności promieniowania w warstwie dl określona jest następująco:

$$dI_{\lambda} = \alpha_{\lambda}(l)[I_{c\lambda} - I_{\lambda}(l)]dl . \quad (66)$$

Jeżeli ścieżkę obserwacji podzieli się na jednorodne warstwy, to monochromatyczna intensywność promieniowania docierającego do obserwatora z poprzedniej warstwy można zapisać jako:

$$I_0 = I_1\delta_0 + (1 - \delta_0)B_0, \quad (67)$$

gdzie: B_0 – promieniowanie ciała doskonale czarnego o temperaturze warstwy 0, δ_0 – transmisyjność warstwy 0.



Rys. 39. Podział ścieżki pomiarowej o długości L na jednorodne warstwy
Fig. 39. Measurement path dividing in layers with constant parameters

Stosując zależność (67) do następnych warstw, otrzymujemy:

$$I_1 = I_2\delta_1 + (1 - \delta_1)B_1, \quad (68)$$

$$I_2 = I_3\delta_2 + (1 - \delta_2)B_2, \quad (69)$$

$$I_3 = I_4\delta_3 + (1 - \delta_3)B_3, \quad (70)$$

i ogólnie

$$I_n = I_{n+1}\delta_n + (1 - \delta_n)B_n. \quad (71)$$

gdzie: B_n – promieniowanie ciała doskonale czarnego o temperaturze warstwy n , δ_n – transmisyjność warstwy n .

Intensywność promieniowania, które wychodzi z drugiej warstwy, z uwzględnieniem tłumienia warstwy pierwszej i zerowej, otrzyma następującą postać [8]:

$$I_0 = [I_2\delta_1 + (1 - \delta_1)B_1]\delta_0 + (1 - \delta_0)B_0. \quad (72)$$

Po wymnożeniu otrzymamy następujące wyrażenie:

$$I_0 = I_2\delta_1\delta_0 + (1 - \delta_1)B_1\delta_0 + (1 - \delta_0)B_0. \quad (73)$$

Jeśli uwzględnimy intensywność promieniowania wychodzącą z warstwy trzeciej, otrzymujemy zależność

$$I_0 = [I_3\delta_2 + (1 - \delta_2)B_2]\delta_1\delta_0 + (1 - \delta_1)B_1\delta_0 + (1 - \delta_0)B_0, \quad (74)$$

a po prostych przekształceniach otrzymamy

$$I_0 = I_3 \delta_2 \delta_1 \delta_0 + (1 - \delta_2) B_2 \delta_1 \delta_0 + (1 - \delta_1) B_1 \delta_0 + (1 - \delta_0) B_0. \quad (75)$$

Podane postępowanie można prowadzić aż do ostatniej warstwy, czy też w przypadku obiektów zamkniętych do promieniowania przeciwległej ściany.

Iloczyn transmisyjności poszczególnych warstw można zastąpić transmisyjnością całościową obliczaną od obserwatora do warstwy n , czyli

$$\tau_n = \prod_{n'=0}^n \delta_{n'}, \quad (76)$$

$$\tau_0 = \delta_0. \quad (77)$$

Uwzględniając monochromatyczną transmisyjność całościową τ_n , intensywność promieniowania przechodzącego przez trzy warstwy można zapisać

$$I_0 = I_4 \tau_3 + B_3(\tau_2 - \tau_3) + B_2(\tau_1 - \tau_2) + B_1(\tau_0 - \tau_1) + B_0(1 - \tau_0), \quad (78)$$

lub

$$I_0 = I_4 \tau_3 - B_3(\tau_3 - \tau_2) - B_2(\tau_2 - \tau_1) - B_1(\tau_1 - \tau_0) - B_0(\tau_0 - 1). \quad (79)$$

W poszczególnych nawiasach pojawiają się różnice transmisyjności, które odzwierciedlają jej zmianę na drodze obserwacji w każdej warstwie. Przy czym transmisyjność oblicza się od obserwatora w kierunku przeciwległej ściany. Dlatego też zmianę transmisyjności zapisać można następująco:

$$\Delta \tau_n = \tau_n - \tau_{n-1} \quad \text{dla } n \geq 1, \quad (80)$$

$$\Delta \tau_0 = \tau_0 - 1 \quad \text{dla } n = 0. \quad (81)$$

Ponieważ transmisyjność obliczana jest od obserwatora w kierunku przeciwległej ściany, jej zmiany będą ujemne, tzn.:

$$I_0 = I_4 \tau_{3-0} - B_3 \Delta \tau_3 - B_2 \Delta \tau_2 - B_1 \Delta \tau_1 - B_0 \Delta \tau_0. \quad (82)$$

Intensywność promieniowania docierającego do niejednorodnego ośrodka I_N , np. promieniowanie przeciwległej ściany względem sondy pomiarowej, po przejściu przez N warstw dociera do tej sondy. Druga część równania składa się z emisji promieniowania przez poszczególne warstwy. Intensywność promieniowania docierająca do obserwatora można zapisać następująco:

$$I_0 = I_N \tau_{N-1} - B_{N-1} \Delta \tau_{N-1} - B_{N-2} \Delta \tau_{N-2} - \dots - B_2 \Delta \tau_2 - B_1 \Delta \tau_1 - B_0 \Delta \tau_0, \quad (83)$$

lub w zwartej formie

$$I_0 = I_N \tau_{N-1} - \sum_{n=0}^{N-1} B_n \Delta \tau_n . \quad (84)$$

Zmiana transmisyjności następuje na pewnej długości drogi optycznej (warstwie) Δl , a jeśli wszystkie warstwy będą tej samej długości, to można ją wyrazić w postaci

$$I_0 = I_N \tau_{N-1} - \sum_{n=0}^{N-1} B_n \frac{\Delta \tau_n}{\Delta l_n} \Delta l_n . \quad (85)$$

Przechodząc od zmian dyskretnych do ciągłych, zależność można zapisać następująco:

$$I_0 = I_N \tau_L - \int_0^L B(l) \frac{\partial \tau(l)}{\partial l} dl , \quad (86)$$

gdzie

$$\tau(l) = \exp \left(- \int_0^l \alpha(l') dl' \right) , \quad (87)$$

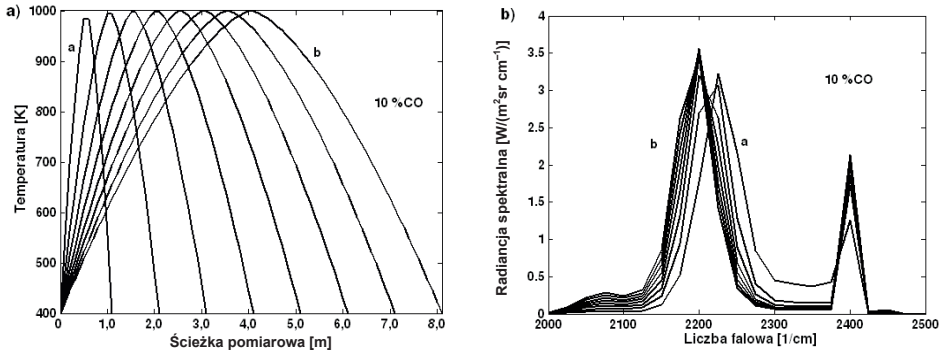
$$\tau_L = \exp \left(- \int_0^L \alpha(l') dl' \right) . \quad (88)$$

Jeśli promieniujące gazy nie znajdują się w zamkniętym obiekcie i promieniowanie przeciwległej ściany nie występuje, to intensywność promieniowania niejednorodnego ośrodka można zapisać następująco:

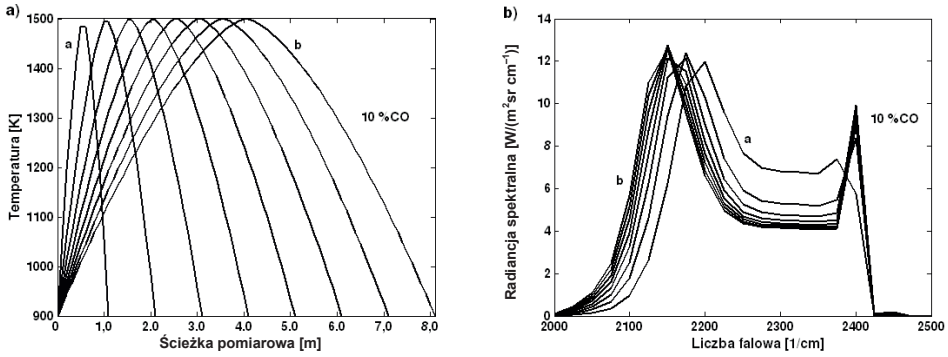
$$I_0 = - \int_0^L B(l) \frac{\partial \tau(l)}{\partial l} dl , \quad (89)$$

Taką postać równania transportu energii radiacyjnej, w której występuje transmisyjność można spotkać w pracach dotyczących zdalnych pomiarów spektralnych atmosfery. W literaturze można znaleźć kilka równoważnych postaci równania radiacyjnego transportu energii [7], [26], [32].

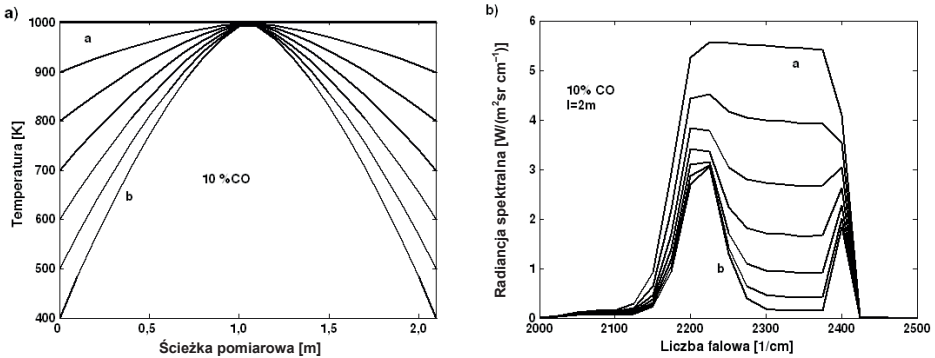
Wykorzystując model SNB oraz równanie radiacyjnego transferu energii, przeprowadzono symulację widma promieniowania dla różnego rodzaju gazowych ścieżek pomiarowych, dla zmieniających się kształtów rozkładu temperatury przy zawartości 10% dwutlenku węgla w rozpatrywanym gazie. Są to paraboliczne rozkłady temperatury o zmieniającej się temperaturze maksymalnej od 1000 do 1500 K. Okazuje się, że zmiany maksymalnej temperatury, kształtu rozkładu temperatury ścieżki obserwacji wpływają na widmo promieniowania w znaczący sposób.



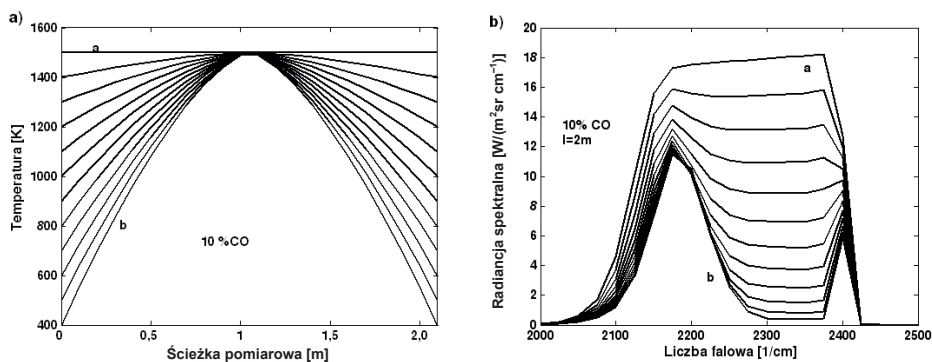
Rys. 40. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz jej widmo promieniowania
Fig. 40. Measurement path temperature profile and radiance spectrum



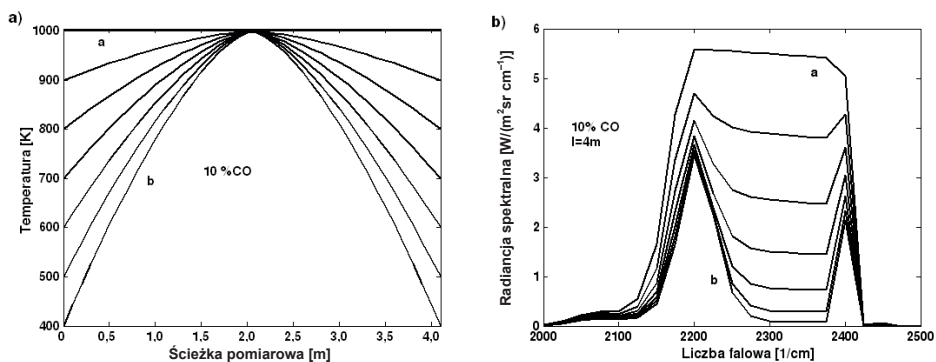
Rys. 41. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz jej widmo promieniowania
Fig. 41. Measurement path temperature profile and radiance spectrum



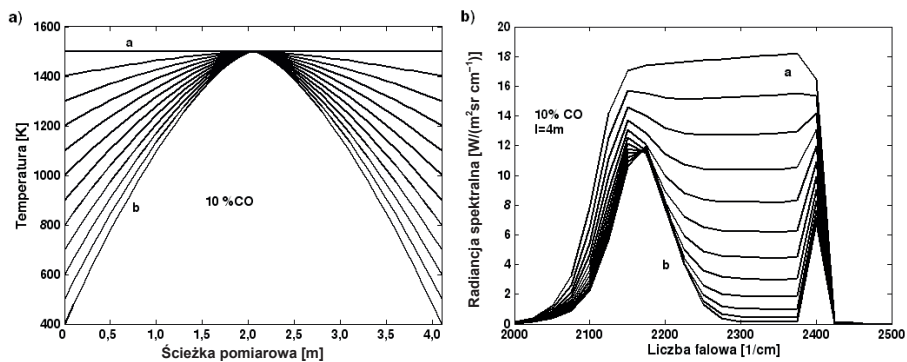
Rys. 42. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz jej widmo promieniowania
Fig. 42. Measurement path temperature profile and radiance spectrum



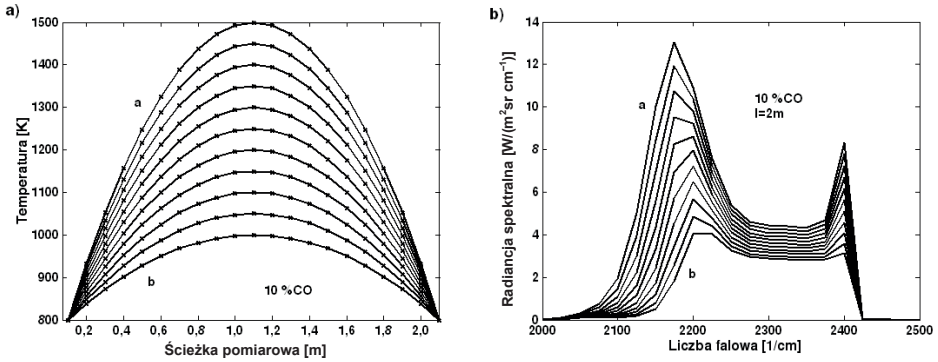
Rys. 43. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz jej widmo promieniowania
 Fig. 43. Measurement path temperature profile and radiance spectrum



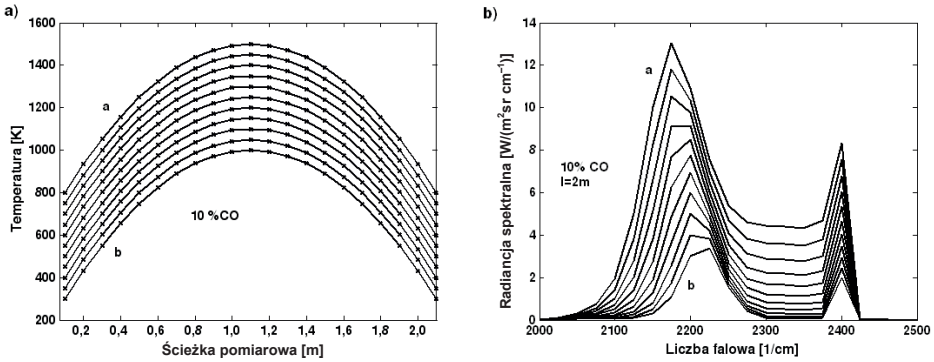
Rys. 44. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz jej widmo promieniowania
 Fig. 44. Measurement path temperature profile and radiance spectrum



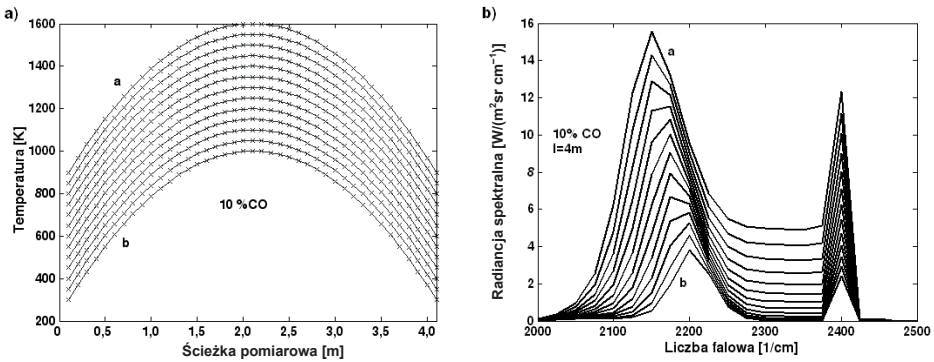
Rys. 45. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz jej widmo promieniowania
 Fig. 45. Measurement path temperature profile and radiance spectrum



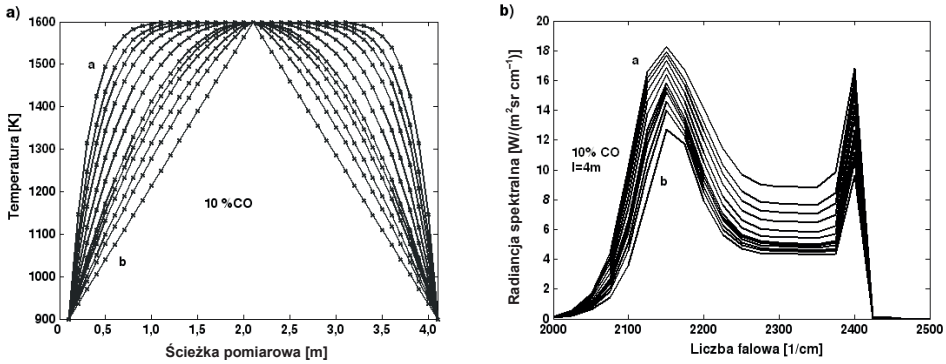
Rys. 46. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz jej widmo promieniowania
 Fig. 46. Measurement path temperature profile and radiance spectrum



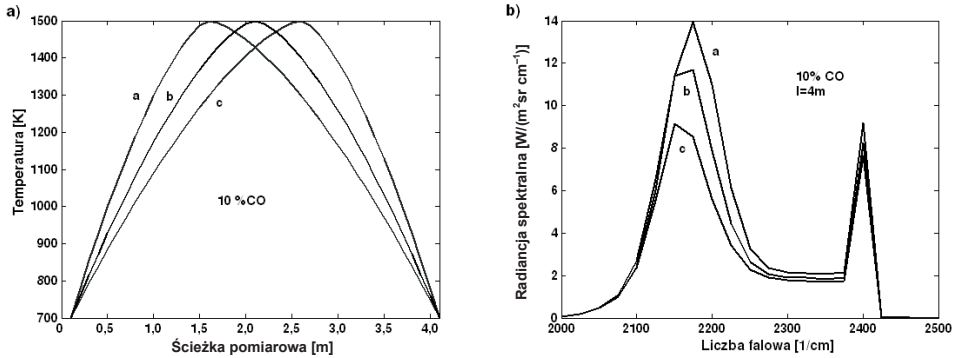
Rys. 47. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz jej widmo promieniowania
 Fig. 47. Measurement path temperature profile and radiance spectrum



Rys. 48. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz jej widmo promieniowania
 Fig. 48. Measurement path temperature profile and radiance spectrum



Rys. 49 Rozkład temperatury wzdłuż ścieżki pomiarowej oraz jej widmo promieniowania
Fig. 49. Measurement path temperature profile and radiance spectrum



Rys. 50. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz jej widmo promieniowania
Fig. 50. Measurement path temperature profile and radiance spectrum

Jeśli w ośrodku zachodzi zarówno absorpcja, jak też rozpraszanie, to współczynnik osłabienia promieniowania przyjmuje następującą postać:

$$K_{\lambda} = \alpha_{\lambda} + b_{\lambda}, \quad (90)$$

gdzie: α_{λ} – współczynnik absorpcji, b_{λ} – współczynnik rozpraszania.

Równanie transportu energii promienistej dla ośrodka, w którym zachodzi pochłanianie i rozpraszanie promieniowania, przybiera następującą postać różniczkową [20], [26], [32]

$$\frac{\partial I_{\lambda}(l)}{\partial l} = -\alpha_{\lambda}(l)I_{\lambda}(l) + \alpha_{\lambda}(l)B_{\lambda}(l) - b_{\lambda}(l)I_{\lambda}(l) + \frac{b_{\lambda}}{4\pi} \int_{\Omega=4\pi} I_{\lambda}(l, \Omega) p(\Omega) d\Omega, \quad (91)$$

gdzie: $p(\Omega)$ – funkcja fazowa charakteryzująca anizotropowość ośrodka.

Funkcja fazowa spełnia następujące równanie:

$$\frac{1}{4\pi} \int_{4\pi} p(\Omega) d\Omega = 1. \quad (92)$$

Jeżeli funkcja fazowa jest stałą dla każdego kąta, tj. $p(\Omega) = 1$, to ośrodek jest izotropowy, a energia jest rozpraszana równomiernie we wszystkich kierunkach.

W ośrodku z rozpraszaniem zmiana intensywności promieniowania wynika z czterech zjawisk:

- zmniejszenia intensywności promieniowania przez absorpcję,
- wzrostu intensywności promieniowania przez emisję własną ośrodka,
- zmniejszania intensywności promieniowania przez rozpraszanie promieniowania we wszystkich kierunkach,
- wzrostu intensywności promieniowania w wyniku rozpraszania w rozpatrywanym kierunku.

6. POŁĄCZENIE METODY EMISYJNEJ I ABSORPCYJNEJ

Wyznaczanie zawartości składników na podstawie pomiarów widm absorpcyjnych daje lepsze wyniki niż na podstawie widm emisyjnych. Nawet jeśli temperatura mieszaniny gazów jest wysoka, a ścieżka obserwacji w przybliżeniu jednorodna. Metody emisyjne (pasywne) znajdują zastosowanie w przypadkach, w których nie można użyć metod absorpcyjnych dla pomiarów *in situ*. Metody pasywne cechuje ogólnie wysoka czułość na temperaturę. Jeśli zaś wyznaczamy zawartość składników, lepsze wyniki otrzymamy, analizując widma absorpcyjne niż gdybyśmy analizowali widma emisyjne tego samego procesu. Stąd też w przypadkach, w których zastosować można zarówno pomiar transmisyjny, jak też emisyjny, zastosować należy metodę mieszaną, a więc wykorzystać zarówno widmo absorpcyjne, jak też emisyjne. Dla pomiarów *in situ*, w których ścieżka pomiarowa przebiega przez badany proces, łatwiej jest przygotować i przeprowadzić analizę widma absorpcyjnego. W połączonej metodzie emisyjno-absorpcyjnej zakłada się jednorodność obserwowanej ścieżki pomiarowej. Przy czym zakłada się w niej jednorodność parametrów monitorowanej ścieżki optycznej. Transmisyjność mierzonej ścieżki obserwacji o długości l wynosi

$$\tau_{\lambda}(T, c) = \exp(-\alpha_{\lambda}(T, c)l). \quad (93)$$

W przypadku ścieżki pomiarowej składającej się jedynie z gazów – bez substancji lotnych, np. sadza – transmisyjność w pasmach, w których nie występują cechy spektralne gazów będzie równa jedności. Gazy promieniują na tej samej długości fali, na której absorbują promieniowanie. Intensywność promieniowania zależy od zawartości gazów oraz ich temperatury, co można przedstawić następującą zależnością:

$$I_\lambda = [1 - \tau_\lambda(T, c)] B_\lambda(T) = \varepsilon_\lambda(T, c) B_\lambda(T), \quad (94)$$

gdzie $B_\lambda(T)$ – intensywność promieniowania ciała doskonale czarnego.

Jak widzimy w poprzednim wzorze intensywność promieniowania jest funkcją temperatury, ponieważ wraz z nią zmienia się intensywność promieniowania ciała doskonale czarnego oraz transmisyjność mieszaniny. Do wyznaczenia zawartości składników na podstawie pomiarów transmisyjności potrzebna jest znajomość temperatury. Wyznaczanie zawartości składnika na podstawie jego widma absorpcyjnego o znanej temperaturze jest zadaniem o wiele prostszym niż jednoczesne wyznaczenie jego zawartości oraz temperatury. Analizując w podobny sposób widma emisyjne określonego składnika, w celu wyznaczenia temperatury, znacznie łatwiej jest to zrobić dla znanych jego zawartości, niż jednocześnie wyznaczać zawartość i temperaturę. Wykorzystując zarówno widmo emisyjne, jak też absorpcyjne oraz obliczając na ich podstawie tzw. znormalizowane widmo promieniowania I_λ^N , wyeliminować można konieczność jednoczesnego wyznaczenia składu i temperatury. Obliczenie temperatury na podstawie I_λ^N jest operacją stosunkowo prostą. Znormalizowaną intensywność promieniowania obliczamy, wykorzystując zmierzoną transmisyjność τ_λ oraz zmierzoną intensywność promieniowania I_λ z zależności [2], [24]

$$I_\lambda^N = I_\lambda / (1 - \tau_\lambda). \quad (95)$$

Dla jednorodnej ścieżki pomiarowej zmierzone widmo emisyjne jest iloczynem emisyjności oraz intensywności promieniowania ciała doskonale czarnego, co można zapisać

$$I_\lambda^N = I_\lambda / (1 - \tau_\lambda) = (1 - \tau_\lambda) B_\lambda(T) / (1 - \tau_\lambda) = B_\lambda(T), \quad (96)$$

przy czym zależność jest poprawna dla $\tau_\lambda \neq 1$, a więc głównie dla pasm gazowych, w których występuje absorpcja promieniowania. Znormalizowana intensywność promieniowania jest równa intensywności promieniowania ciała doskonale czarnego.

Jeśli ścieżka pomiarowa zawiera jedynie substancje takie jak sadza, obserwowana będzie zarówno absorpcja, jak i rozpraszanie. Transmisję ścieżki optycznej o długości l zapisuje się wtedy następująco:

$$\tau_\lambda = \exp(-n_c A F_e l), \quad (97)$$

gdzie: n_c – gęstość cząstek, A – przekrój czynny cząsteczek, F_e – współczynnik ekstynkcji.

Znormalizowaną intensywność promieniowania można przedstawić następująco [24]:

$$I_\lambda^N = I_\lambda / (1 - \tau_\lambda) = F_a / F_e B_\lambda(T). \quad (98)$$

Przy jednorodnej ścieżce współczynnik ekstynkcji wyraża się wzorem

$$F_e = F_a + F_s, \quad (99)$$

gdzie: F_s – współczynnik rozpraszania, F_a – współczynnik absorpcji.

Na podstawie transmisyjności można wyznaczyć pośrednio gęstość cząstek. Jeżeli użyje się teorii Mie do wyznaczania współczynników F , to obliczyć można rozmiary cząstek. Wygodnie jest do tego wykorzystać pasma spektralne o charakterystycznych wartościach współczynników F , na przykład

$$F_a/F_e \approx 1. \quad (100)$$

Jeśli w mieszaninie występują cząstki stałe i gazy, transmisyjność zapisać możemy następująco:

$$\tau_\lambda = \exp[-(n_c A F_e + \alpha_\lambda) L]. \quad (101)$$

Znormalizowane natężenie promieniowania dla cząstek i gazów przybiera następującą postać [3]:

$$I_\lambda^N = I_\lambda / (1 - \tau_\lambda) = \{1 - \exp[-(n_c A F_a + \alpha_\lambda) L]\} B_\lambda(T) / \{1 - \exp[-(n_c A F_e + \alpha_\lambda) L]\}. \quad (102)$$

Metoda ta znalazła szerokie zastosowanie do badania płomieni, do których możliwy jest dostęp z dwóch stron [23]. Ciekawe są badania przeprowadzone w kotłach energetycznych powyżej strefy płomienia, gdzie długość ścieżki pomiarowej wynosiła 10 m [27]. W wyniku tych pomiarów okazało się, że pomimo braku widoczności przeciwległej ściany w zakresie widzialnym, to w zakresie średniej podczerwieni ścieżka optyczna jest przezroczysta. Transmisyjność ścieżki w obiekcie wynosiła zero jedynie w zakresach pochłaniania gazów CO_2 (2400–2150 cm^{-1}) oraz H_2O (4100–3100 cm^{-1}). Na podstawie pomiarów transmisyjności słabszych pasm wyznaczyć można koncentrację pary wodnej i dwutlenku węgla. Silne pasma na długości 10 m powodują całkowite pochłanianie promieniowania.

7. ANALIZA WIDM JAKO ZAGADNIENIE ODWROTNE

Analiza widm emisyjnych i absorpcyjnych procesów, w których następuje zarówno zmiana koncentracji, jak też temperatury nie może polegać na klasycznej metodzie kalibracji. Zbudowanie modelu kalibracyjnego polegającego na przekształceniu zmierzonego widma w wielkości poszukiwane nie zawsze jest możliwe. Symulacyjny model widma jest funkcją temperatury oraz składu. Jeśli mamy widma pomiarowe, to interesuje nas wyznaczenie na ich podstawie zawartości gazów oraz ich temperatura. Model widma określa zależność występującą pomiędzy wielkościami wyznaczanymi a mierzonymi, stąd też nazywany jest on modelem prostym lub też modelem bezpośrednim.

Poszukiwane informacje wyznacza się, wykorzystując model bezpośredni oraz dostępne dane pomiarowe, a zadanie takie określane jest jako zagadnienie odwrotne [8]. Jest to typowy przykład pomiarów pośrednich obiektów złożonych [30]. Model bezpośredni oprócz modelowania widm uwzględniać musi charakterystykę przetwarzania układu pomiarowego. Model prosty przedstawić można w następującej postaci:

$$\mathbf{y} = F(\mathbf{x}, \mathbf{c}_m), \quad (103)$$

gdzie: $\mathbf{y} = [y_1, y_2, \dots, y_m]^T$ – wektor pomiaru (mierzone widmo), $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ – wektor wyznaczanych zmiennych lub parametrów, takich jak temperatura oraz zawartość składników, \mathbf{c}_m – dodatkowe parametry modelu.

Operator F jest nieliniowy, dodatkowo, do wyznaczenia wartości wektora \mathbf{y} niezbędne są parametry cech spektralnych gazów, które dostępne są w postaci tabelarycznej. Ze względu na opisaną metodę wyznaczania transmisyjności gazów, jak też podział badanej ścieżki pomiarowej na warstwy, równania odwrotnego F^{-1} nie można rozwiązać analitycznie, a jedynie numerycznie. Wektor wyznaczanych parametrów podlega ograniczeniom związanym z naturą badanego procesu, czyli:

$$\mathbf{x}_{\min} \leq \mathbf{x} \leq \mathbf{x}_{\max}. \quad (104)$$

Jeśli analizujemy widma emisyjne, model bezpośredni opiera się na równaniu transportu energii radiacyjnej, które w wersji całkowitej jest równaniem Fredholma pierwszego rodzaju (stałe granice całkowania oraz niewiadoma pod znakiem całki). Do rozwiązania zagadnienia odwrotnego należy sformułować funkcję celu i wykorzystać do jej minimalizacji metody optymalizacji.

Zagadnienie odwrotne określić można jako wyznaczanie informacji o nieznanymi wielkościach na podstawie danych pomiarowych (widm) powiązanych modelem matematycznym z poszukiwanymi wartościami (temperatura, skład). Ponadto mamy dodatkowe informacje zmniejszające zbiór możliwych rozwiązań. Rozwiązanie problemu odwrotnego wymaga minimalizacji różnicy pomiędzy wektorem widma mierzonego a wektorem widma otrzymanego z modelu. Funkcja celu w najprostszym przypadku przyjmuje postać

$$F_C(\mathbf{x}) = \|\mathbf{y} - F(\mathbf{x})\|^2 = \sum_i r_i^e(\mathbf{x}), \quad (105)$$

gdzie

$$\mathbf{R}(\mathbf{x}) = (r_1(\mathbf{x}), r_2(\mathbf{x}), \dots, r_M(\mathbf{x}))^T. \quad (106)$$

Wektor reszt $\mathbf{R}(\mathbf{x})$ jest jedynym wskaźnikiem błędu rozwiązania zagadnienia odwrotnego. Inny typ błędu rozwiązania problemu odwrotnego to różnica pomiędzy wartością rzeczywistą x a wartością estymaty otrzymaną w wyniku rozwiązania zagadnienia odwrotnego. Niestety błąd taki wyznaczyć można tylko dla danych symulacyjnych lub też pomiaru wartości wektora \mathbf{x} metodami alternatywnymi.

Analiza widm *in situ* nie podlega ostrym rygorom, takim jak analiza składu mieszanin w laboratorium. Do celów diagnostycznych często wystarcza informacja o średnich wartościach parametrów czy kierunku zmian wartości. W większości problemów praktycznych i tak nie mamy wyboru, gdyż to, jaką informację możemy uzyskać, determinuje nam sam mierzony proces oraz układ pomiarowy.

W niektórych przypadkach, takich jak wyznaczanie profilu temperatury ścieżki obserwacji, informacji zawartych w widmie może być za mało do pełnego odtworzenia parametrów. Oznacza to, że zagadnienie odwrotne jest niedookreślone, istnieje więcej parametrów do wyznaczenia niż dostępnych danych. Przyczyną tego może być mała rozdzielczość spektrometru, zmniejszająca liczbę informacji pomiarowych. Jeśli w wyznaczanym rozkładzie temperatury ścieżki obserwacji znajdują się miejsca o szybkich zmianach temperatury, to niedookreślenie zagadnienia nie pozwoli ich wyznaczyć. Niedookreślone zagadnienie ma kilka rozwiązań, dla których wartość reszt funkcji celu jest mała. W takich przypadkach ogranicza się możliwe rozwiązania, do mających fizyczne uzasadnienie, np. przez nałożenie ograniczenia co do gładkości rozwiązania. Podczas rozwiązywania zadania odwrotnego pojawić się może również niestabilność oraz problem dużych zmian w rozwiązaniu przy małych zmianach wektora pomiarowego.

Dla rozkładu temperatury stabilność rozwiązania osiągnąć można przez przedstawienie parametrów wyznaczanego profilu w postaci pewnej funkcji lub zespołu funkcji bazowych. Wyboru funkcji dokonuje się na podstawie wcześniejszych danych statystycznych, dotyczących poszukiwanego profilu. Jeżeli mamy dodatkową wiedzę na temat badanego procesu, należy ją wykorzystać do ograniczenia zbioru możliwych rozwiązań. Informacje mogą pochodzić z dodatkowych pomiarów innymi metodami. Rozwiązanie problemu odwrotnego może wymagać regularyzacji, która modyfikuje funkcję celu w taki sposób, by zapewnić stabilność rozwiązania oraz zbliżyć go do wartości rzeczywistych. Najpopularniejszą metodą jest regularyzacja Tikhonova, w której do funkcji celu dodany zostaje dodatkowy człon wprowadzający dodatkowe informacje oraz ograniczający zbiór rozwiązań, np. do rozwiązań gładkich. W procesie minimalizacji funkcji celu z członem regularyzującym, rozważyć należy stopień wpływu tego członu na rozwiązanie. Zbyt duży wpływ regularyzacji powoduje ograniczenie wykorzystania informacji zawartych w danych pomiarowych. Regularyzacja ma za zadanie stabilizację, rozwiązanie oraz ograniczenie wpływu błędów pomiaru na rozwiązanie. Człon regularyzujący jest najczęściej funkcjonalem wygładzającym. Funkcja celu z regularyzacją przyjmuje następującą postać:

$$F_C(\mathbf{x}) = \|\mathbf{y} - F(\mathbf{x})\|^2 + \alpha_r \|\mathbf{L}_r(\mathbf{x} - \mathbf{x}_0)\|^2, \quad (107)$$

gdzie macierz \mathbf{L}_r wygląda następująco:

$$\mathbf{L}_r = \mathbf{I}, \quad (108)$$

gdzie \mathbf{I} – macierz jednostkowa,

$$\mathbf{L}_r = \begin{bmatrix} -1 & 1 & 0 & \dots \\ 0 & -1 & 1 & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & -1 & 1 \end{bmatrix}, \quad (109)$$

$$\mathbf{L}_r = \begin{bmatrix} -1 & 1 & 0 & \dots \\ 0 & -1 & 1 & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} -1 & 1 & 0 & \dots \\ 0 & -1 & 1 & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & -1 & 1 \end{bmatrix}. \quad (110)$$

Parametr regularyzacji α_r wpływa na zmniejszenie różnicy pomiędzy widmem symulacyjnym oraz widmem pomiarowym, a z drugiej strony gładkością rozwiązania. Wektor wartości poszukiwanych \mathbf{x}_0 jest wstępną wartością określającą obszar poszukiwań rozwiązania. Regularyzacja z członem kary ma kilka zalet, z których główna to dobra aproksymacja rzeczywistego rozwiązania nawet w przypadku słabego uwarunkowania nieliniowego operatora F . Oczywiście zaleta ta pojawia się dla dobrze dobranego parametru regularyzacji, którego wartość rozwiązuje problem pomiędzy wpływem członu regularyzacji na rozwiązanie a dopasowaniem poszukiwanego wektora do danych pomiarowych. Dostępne są różne metody wyznaczenia parametru α_r . Ogólną zasadą jest, że duża wartość członu regularyzacji powoduje dużą niedokładność dopasowania widma pomiarowego do widma modelowanego w wyniku rozwiązania zagadnienia odwrotnego. Mała wartość członu regularyzacji poprawia jakość dopasowania, niestety jednocześnie powoduje wzrost wpływu błędów pomiarowych oraz oscylacji w wielkościach wyznaczanych.

Do minimalizacji funkcji celu użyć należy metod optymalizacji, które podzielić można na lokalne oraz globalne. Metody globalne, takie jak metoda algorytmów genetycznych oraz symulowanego wyżarzania, umożliwiają znalezienie minimum funkcji spośród wielu minimów lokalnych. Niestety czas potrzebny do obliczeń, tj. liczba iteracji powoduje, że metody takie nie nadają się do obliczeń *on-line* w celu diagnozowania procesu. Z drugiej zaś strony parametry wyznaczone na podstawie widm podlegają fizycznym ograniczeniom. Ponadto w większości przypadków eksperymentator ma dodatkowe informacje, dzięki którym wektor startowy iteracji może być wybrany dość blisko obszaru wektora rzeczywistego. W lokalnych metodach optymalizacji krok iteracyjny przyjmuje postać:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k, \quad (111)$$

gdzie: \mathbf{d}_k – kierunek poszukiwań, α_k – długość kroku iteracji.

Istnieje wiele metod wyznaczania kierunku poszukiwań, np. metoda gradientów sprzężonych:

$$\mathbf{d}_k = -\mathbf{g}_k + \beta_k \mathbf{d}_{k-1}, \quad (112)$$

gdzie β_k jest skalarem dobieranym w zależności od konkretnej metody.

W innych metodach, np. największego spadku, kierunek poszukiwań określony jest zależnością:

$$\mathbf{d}_k = -\mathbf{B}_k^{-1} \mathbf{g}_k, \quad (113)$$

gdzie \mathbf{B}_k w metodzie największego spadku jest macierzą jednostkową.

Metoda Newtona polega na rozwiązaniu liniowej aproksymacji pierwszej części rozwinięcia funkcji w szereg Taylora, tj.:

$$F(\mathbf{x}_{k+1}) \approx F(\mathbf{x}_k) + \mathbf{J}(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = 0, \quad (114)$$

w wyniku czego otrzymujemy następującą postać iteracyjną:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{J}(\mathbf{x}_k)^{-1} (F(\mathbf{x}_k) - \mathbf{y}), \quad (115)$$

a Jakobian określony jest następująco:

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial F_1(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial F_1(\mathbf{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial F_m(\mathbf{x})}{\partial x_n} \end{bmatrix}. \quad (116)$$

Metoda Gaussa–Newtona (GN) wykorzystuje informacje o drugich pochodnych, w której Hesjan wyrazić można następująco:

$$\mathbf{H}(\mathbf{x}) = \mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}) + \mathbf{G}(\mathbf{x}). \quad (117)$$

Przyjmując, że $\mathbf{G}(\mathbf{x})$ jest małe, otrzymuje się aproksymację Hesjanu z wykorzystaniem Jakobianu, wtedy:

$$\mathbf{d}_k = -(\mathbf{J}(\mathbf{x}_k)^T \mathbf{J}(\mathbf{x}_k))^{-1} \nabla \mathbf{J}(\mathbf{x}_k), \quad (118)$$

$$\nabla \mathbf{J}(\mathbf{x}_k) = \mathbf{J}(\mathbf{x}_k)^T (F(\mathbf{x}_k) - \mathbf{y}). \quad (119)$$

Metoda Levenberga–Marquardta jest metodą GN z dodatkowym tłumieniem oscylacji. Parametr tłumienia λ'_k wpływa na kierunek i długość kroku. Podczas kolejnych iteracji parametr można zmniejszać, kontrolując przybliżanie się do metody GN lub metody największego spadku, tj.:

$$\mathbf{d}_k = -(\mathbf{J}(\mathbf{x}_k)^T \mathbf{J}(\mathbf{x}_k) + \lambda_k^r \mathbf{I})^{-1} \nabla \mathbf{J}(\mathbf{x}_k), \quad (120)$$

lub też

$$\mathbf{d}_k = -(\mathbf{J}(\mathbf{x}_k)^T \mathbf{J}(\mathbf{x}_k) + \lambda_k^r \text{diag}(\mathbf{H}))^{-1} \nabla \mathbf{J}(\mathbf{x}_k). \quad (121)$$

W kolejnych krokach iteracyjnych parametr regularyzacji λ_k^r może się zmniejszać, taką metodę nazywamy regularyzowaną metodę GN. Parametr regularyzacji spełniać powinien następujący warunek [17], [22]:

$$\lambda_k^r > 0, \quad 1 \leq \frac{\lambda_k^r}{\lambda_{k-1}^r} \leq r, \quad \lim_{k \rightarrow \infty} \lambda_k^r = 0 \quad (122)$$

Krok iteracyjny [16]

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\lambda_k^r \mathbf{L}^T \mathbf{L} + \mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}))^{-1} [\mathbf{J}(\mathbf{x}_k)^T (F(\mathbf{x}_k) - \mathbf{y}) + \lambda_k^r \mathbf{L}^T \mathbf{L}(\mathbf{x}_k - \mathbf{x}_a)]. \quad (123)$$

Jedną z możliwych procedur na wyznaczanie parametru regularyzacji w każdym kroku iteracyjnym jest wzór [15]:

$$\lambda_k^r = \beta_k \lambda_{\min}^r + (1 - \beta_k) \lambda_{k-1}^r, \quad (124)$$

$$\beta_k = r_k / r_{k-1}. \quad (125)$$

Ograniczenia co do wielkości tak dobieranego parametru określa się przed procedurą optymalizacji. Zmniejszająca się sekwencja parametrów określa wielkość reszt r_k . Jeśli wektor reszt zmniejsza się, to parametr w kolejnej iteracji jest bliski wartości poprzedniej. Jeśli wektor reszt zmniejsza się niewiele, parametr zostaje znacznie zwiększony w stosunku do wybranej jego wielkości minimalnej λ_{\min}^r . Dla kilku początkowych iteracji wprowadza się większą regularyzację, a stopniowo zmniejsza się ją wraz z przybliżaniem się rozwiązania do celu.

8. ANALIZA WRAŻLIWOŚCI ZAGADNIENIA ODWROTNEGO

Rozwiązując zagadnienie odwrotne w pomiarach *in situ* widm dla konkretnego procesu przeważnie jesteśmy zorientowani co do przewidywanych zakresów temperatury oraz zawartości poszczególnych składników. Szczególnie jeśli pomiary widm miałyby być wykonywane w celu diagnostyki *on-line* obiektów z procesami spalania. Można więc przeprowadzić analizę wpływu zmian wyznaczanych parametrów na widmo. Analiza wrażliwości powinna uwzględniać wpływ zmian temperatury oraz koncentracji wszystkich składników znajdujących się w mieszaninie gazowej na widmo absorpcyjne lub emisyjne. Umożliwi ona określenie możliwości wyznaczania po-

szczególnych składników czy też temperatury na ścieżce obserwacji. Dla dedykowanych systemów diagnostyczno-pomiarowych wybrać można najlepszy zbiór długości fal, zapewniając pożądane cechy układów monitorujących proces. Czułość i selektywność pomiaru determinowana jest składem mieszaniny. Analiza wrażliwości określa zdolność wyznaczenia konkretnego parametru przy zakłócającym wpływie innych parametrów. Zmiana absorpcji od temperatury dla jednorodnej ścieżki pomiarowej może być zapisana następująco:

$$\frac{\partial a}{\partial T} = \exp(-\alpha l) l \frac{\partial \alpha}{\partial T} = \tau \frac{\partial \chi}{\partial T}, \quad (126)$$

gdzie: $\chi = \alpha l$ oraz $\tau = \exp(-\alpha l)$, $\alpha(T)$ – monochromatyczny współczynnik pochłaniania.

Intensywność promieniowania pewnej długości fali zapisać można jako:

$$I = B(T)(1 - \exp(-\alpha l)) = B(T)a, \quad (127)$$

gdzie: $B(T)$ – intensywność promieniowania Plancka dla ciała doskonale czarnego, l – długości ścieżki pomiarowej, $a = a(T)$ – absorpcyjność monochromatyczna.

Wpływ zawartości danego składnika na transmisyjność określony jest następująco:

$$\frac{\partial \tau}{\partial c} = -\frac{\partial \chi}{\partial c} \tau. \quad (128)$$

Intensywność promieniowania przy zmianach temperatury zmienia się z dwóch powodów: zmiany intensywności promieniowania ciała doskonale czarnego oraz absorpcyjności składników. Promieniowanie Plancka dla danej długości fali można przedstawić następująco:

$$B(\lambda, T) = \frac{c_1 \lambda^3}{\exp(c_2 \lambda / T) - 1}, \quad (129)$$

Zmiana promieniowania Plancka dla zmiany temperatury określona jest następująco:

$$\frac{\partial B}{\partial T} = \frac{c_1 c_2 \lambda^4 \exp(c_2 \lambda / T)}{T^2 [\exp(c_2 \lambda / T) - 1]^2}, \quad (130)$$

Ostatecznie zależność zmiany promieniowania od temperatury wyrazić można następującym wzorem:

$$\frac{\partial I}{\partial T} = B(T) \tau \frac{\partial \chi}{\partial T} + a \frac{c_1 c_2 \lambda^4 \exp(c_2 \lambda / T)}{T^2 [\exp(c_2 \lambda / T) - 1]^2}. \quad (131)$$

W przypadku niejednorodności ścieżki pomiarowej do celów obliczeniowych oraz analizy, dzielimy ją na N warstw, w których zakładamy homogeniczność parametrów (temperatury i zawartości składników). Jeśli transmisyjność pojedynczej warstwy oznaczona zostanie przez τ_n^w , to transmisyjność całej badanej ścieżki jest iloczynem poszczególnych transmisyjności. W przypadku pomiarów pasywnych, intensywność promieniowania docierającego do obserwatora oblicza się, wykorzystując równanie transportu energii radiacyjnej. W modelu warstwowym parametry odnoszące się do całości analizowanej ścieżki pomiarowej definiowane są przez wykorzystanie parametrów charakteryzujących każdą z warstw. W celu określenia wpływu poszczególnych czynników w określonej warstwie na intensywność promieniowania należałoby obliczyć odpowiednie pochodne cząstkowe. Analiza taka pokazuje wpływ każdego parametru na promieniowanie w mierzonym zakresie długości fal oraz możliwość jego wyznaczenia przez analizę promieniowania [8].

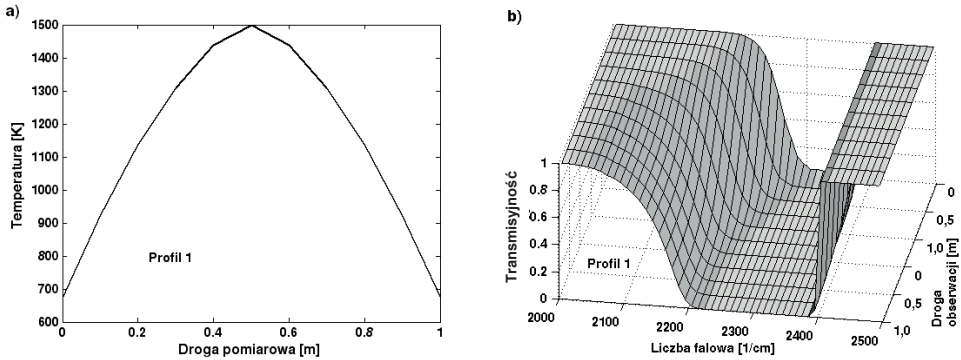
Niejednorodność ścieżek pomiarowych uwzględnia się najczęściej w przypadku pomiarów pasywnych intensywności promieniowania. Widmo emisyjne docierające do obserwatora zależy od temperatury oraz zmian transmisyjności ścieżki pomiarowej liczonej od obserwatora w przeciwnym kierunku. Stąd też analizę niehomogenicznych ścieżek pomiarowych przeprowadza się z wykorzystaniem funkcji wag określonych następującym wzorem [9], [34]:

$$W = \frac{\partial \tau(l)}{\partial l}. \quad (132)$$

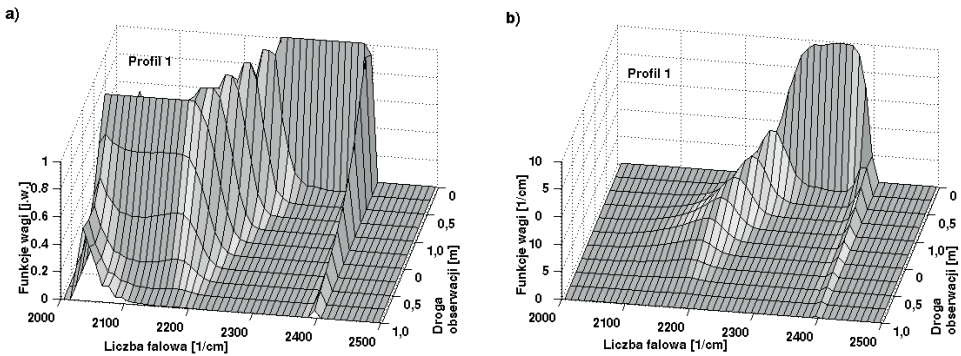
Funkcje wagi wskazują na określoną część całkowitej intensywności promieniowania, jaka dociera do obserwatora z danego miejsca ścieżki obserwacji. Intensywność promieniowania ścieżki pomiarowej docierającego do obserwatora jest ważoną sumą intensywności promieniowania w poszczególnych warstwach. A więc funkcje wagi określają warstwy o największych zmianach transmisyjności, a co za tym idzie wskazują region na ścieżce pomiarowej, w którym można wyznaczać poszukiwane parametry. Ponieważ funkcja wagi jest pochodną transmisyjności optycznej po drodze, będzie więc miała ona największą wartość tam, gdzie występuje silna absorpcja. Przy czym im dalej od obserwatora, tym transmisyjność mniejsza. Jeśli duża lokalna wartość absorpcji występuje dla miejsca, w którym transmisyjność liczona od obserwatora jest mała, to zmiana transmisyjności nie będzie tam miała dużych wartości. Jeśli zainteresowani jesteśmy intensywnością promieniowania poszczególnych warstw, wyznaczyć musimy, dla jakich liczb falowych docierające promieniowanie pochodzi właśnie z tej warstwy [8], [9].

Do analizy niejednorodnych ścieżek pomiarowych za pomocą funkcji wag wykorzystane zostanie modelowanie w średniej podczerwieni, ze średnią rozdzielczością spektralną ($10\text{--}25\text{ cm}^{-1}$). Zależności pomiędzy poszukiwanymi parametrami a promieniowaniem opisuje nieliniowy układ równań wykorzystujący równanie transportu energii radiacyjnej. Analityczny opis zależności transmisyjności od poszczególnych paramet-

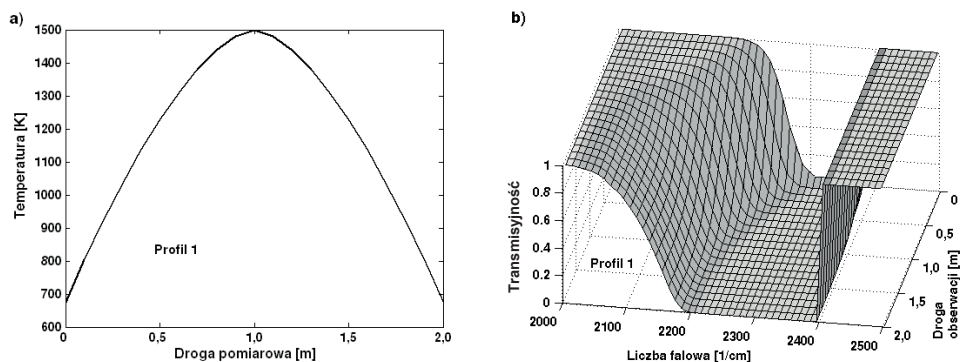
trów nie jest możliwy, gdyż zależy ona od koncentracji, ciśnienia i temperatury. Do zobrazowania funkcji wag wykorzystano transmisyjność pasma CO_2 około $4,3 \mu\text{m}$. Funkcje wag zostaną unormowane do jedności, co umożliwi zobrazowanie wpływu rozkładu temperatury na funkcje wagi. Dla każdej liczby falowej funkcja wagi ma charakterystyczny kształt. Przeanalizujemy typowy rozkład temperatury występujący w procesach spalania, który możemy określić jako osiowo-symetryczny z maksimum temperatury w środku. Jeśli szerokość połówkowa funkcji wagi zajmuje dużą część ścieżki optycznej, to za pomocą tej długości fali wyznaczyć możemy jedynie średnie właściwości kilku warstw. Oznacza to, że szerokie funkcje wagi nie pozwolą na wyznaczanie struktury ścieżki pomiarowej. Jeśli pomiar wykonywany jest z małą rozdzielczością, to funkcje wag na poszczególnych liczbach falowych będą na siebie zachodzić, co oznacza zmniejszenie liczby niezależnych informacji. Funkcje wagi rozpatrzone zostaną na podstawie kilku rozkładów temperatury oraz długości ścieżek obserwacji.



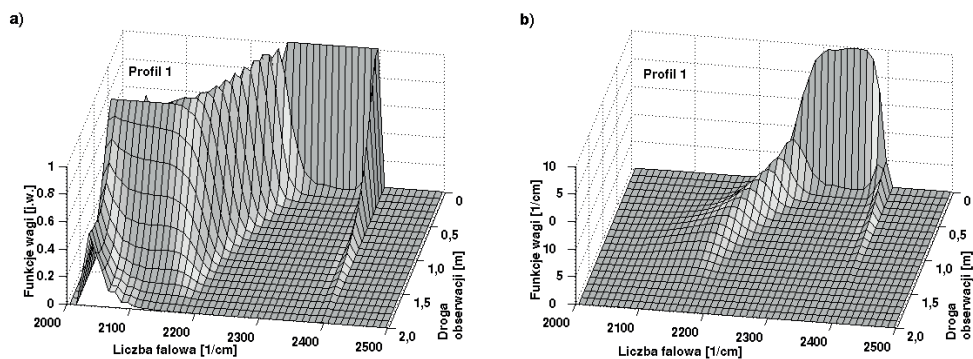
Rys. 51. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 1, 1 m)
 Fig. 51. Temperature profile and transmissivity of measurement path (profile 1, 1 m)



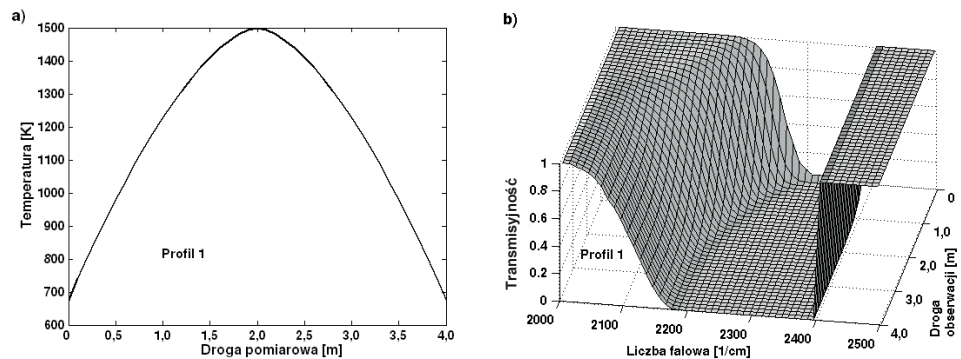
Rys. 52. Unormowane i zwykłe funkcje wagi (profil 1, 1 m)
 Fig. 52. The normalized and normal weighting function (profile 1, 1 m)



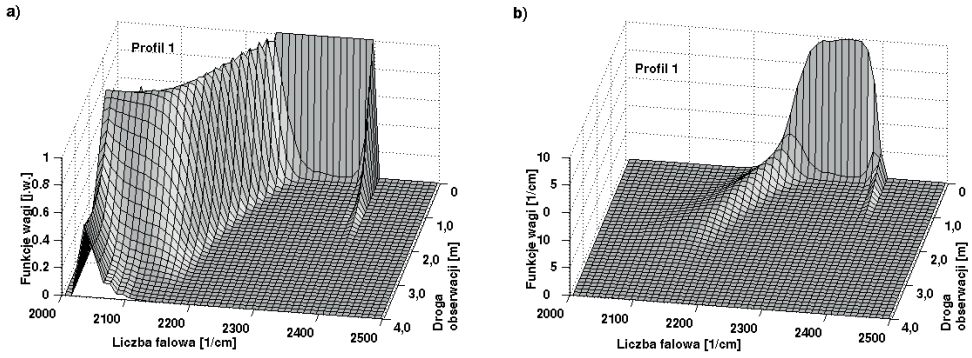
Rys. 53. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 1, 2 m)
 Fig. 53. Temperature profile and transmissivity of measurement path (profile 1, 2 m)



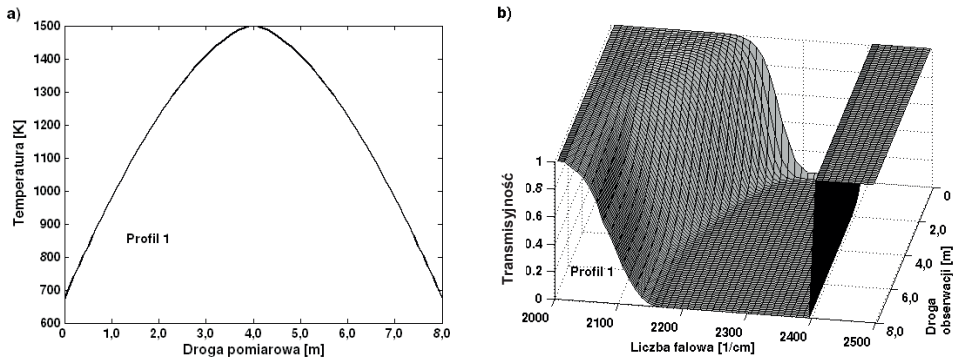
Rys. 54. Unormowane i zwykłe funkcje wagi (profil 1, 2 m)
 Fig. 54. The normalized and normal weighting function (profile 1, 2 m)



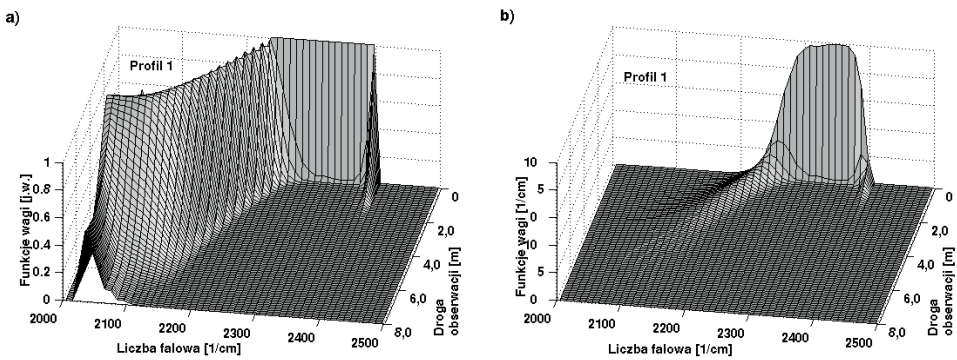
Rys. 55. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 1, 4 m)
 Fig. 55. Temperature profile and transmissivity of measurement path (profile 1, 4 m)



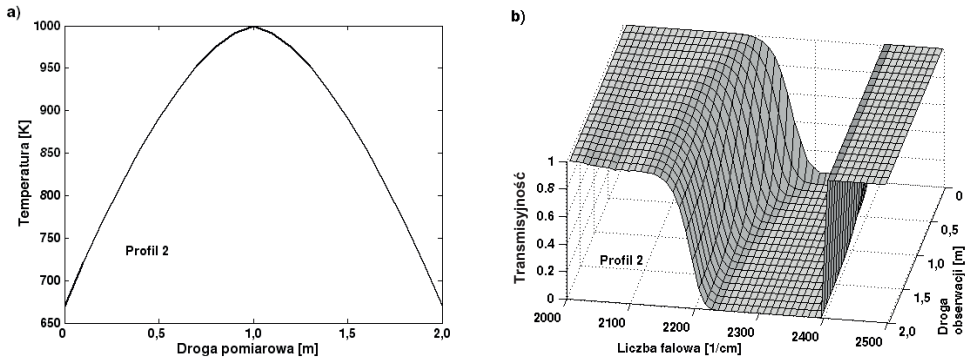
Rys. 56. Unormowane i zwykłe funkcje wagi (profil 1, 4 m)
 Fig. 56. The normalized and normal weighting function (profile 1, 4 m)



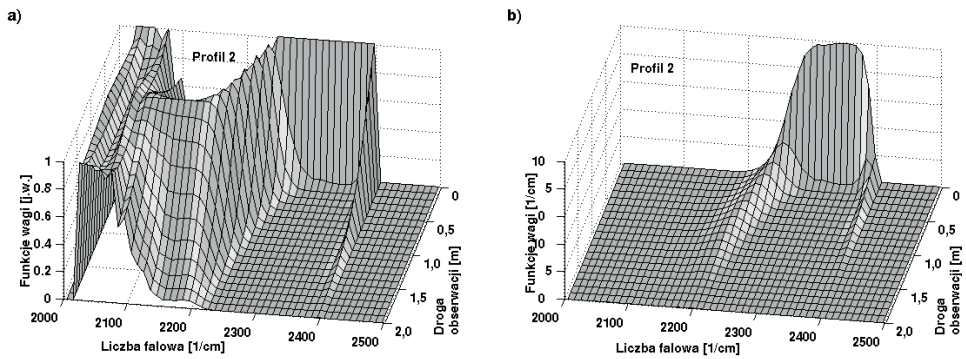
Rys. 57. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 1, 8 m)
 Fig. 57. Temperature profile and transmissivity of measurement path (profile 1, 8 m)



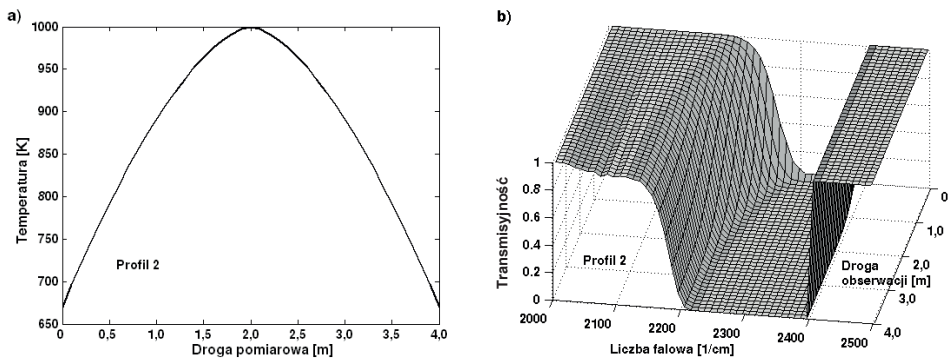
Rys. 58. Unormowane i zwykłe funkcje wagi (profil 1, 8 m)
 Fig. 58. The normalized and normal weighting function (profile 1, 8 m)



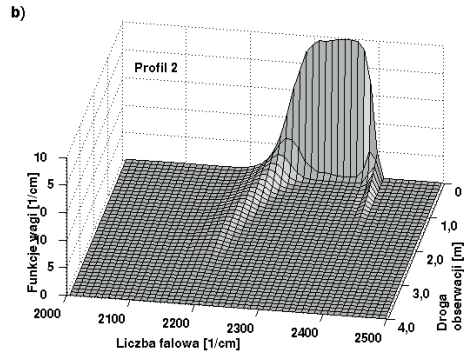
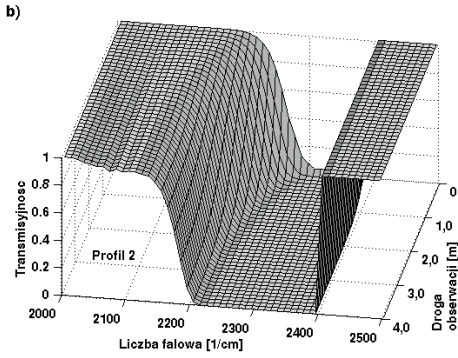
Rys. 59. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 2, 2 m)
 Fig. 59. Temperature profile and transmissivity of measurement path (profile 2, 2 m)



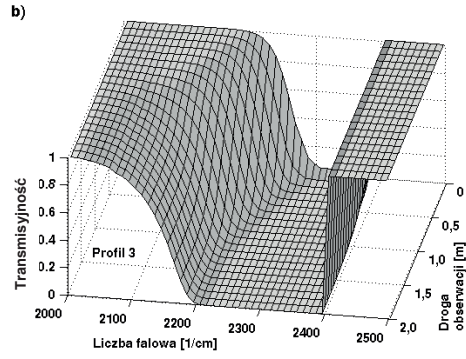
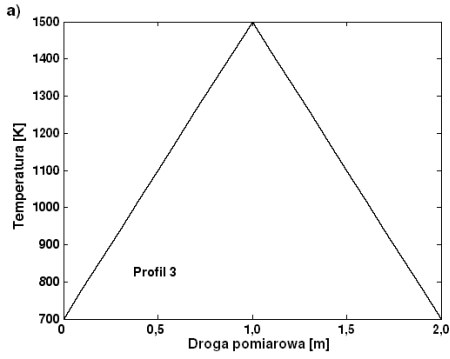
Rys. 60. Unormowane i zwykłe funkcje wagi (profil 2, 2m)
 Fig. 60. The normalized and normal weighting function (profile 2, 2 m)



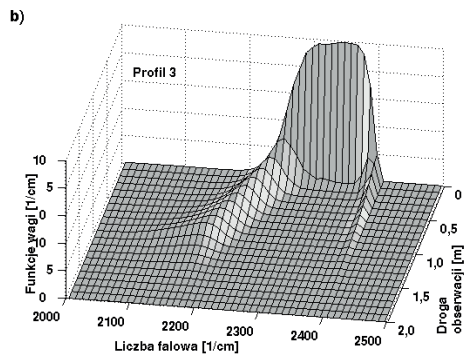
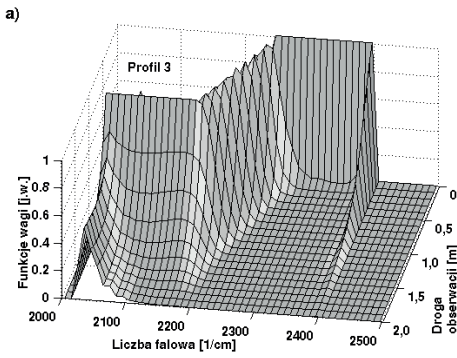
Rys. 61. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 2, 4 m)
 Fig. 61. Temperature profile and transmissivity of measurement path (profile 2, 4 m)



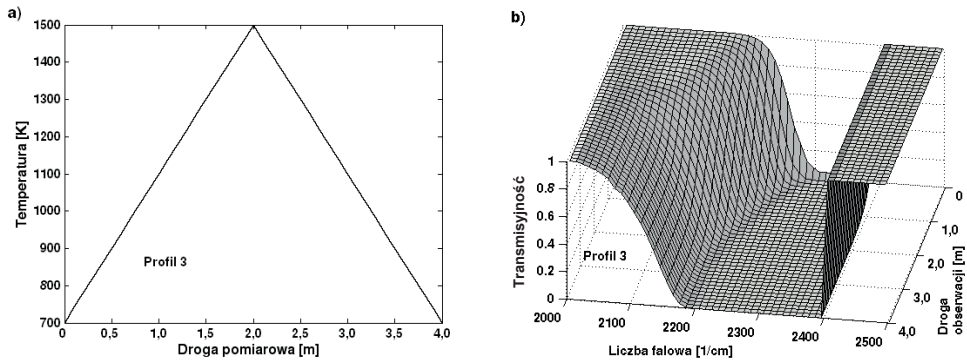
Rys. 62. Unormowane i zwykłe funkcje wagi (profil 2, 4 m)
 Fig. 62. The normalized and normal weighting function (profile 2, 4 m)



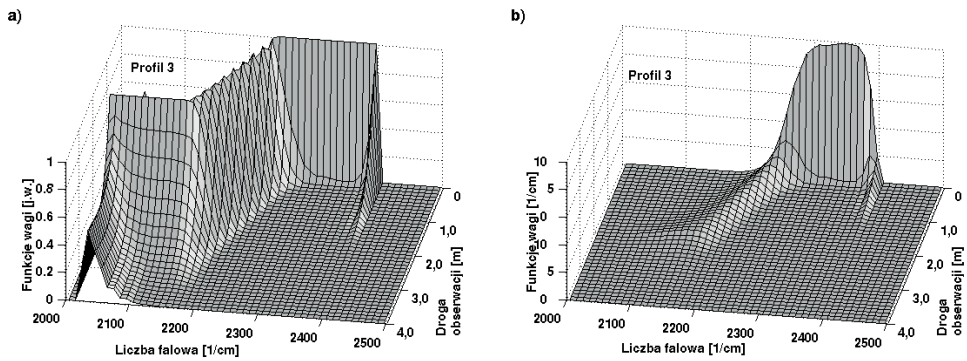
Rys. 63. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 3, 2 m)
 Fig. 63. Temperature profile and transmissivity of measurement path (profile 3, 2 m)



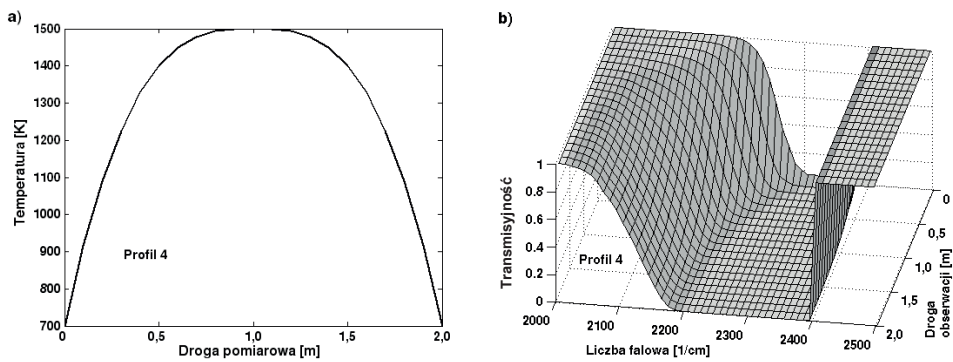
Rys. 64. Unormowane i zwykłe funkcje wagi (profil 3, 2 m)
 Fig. 64. The normalized and normal weighting function (profile 3, 2 m)



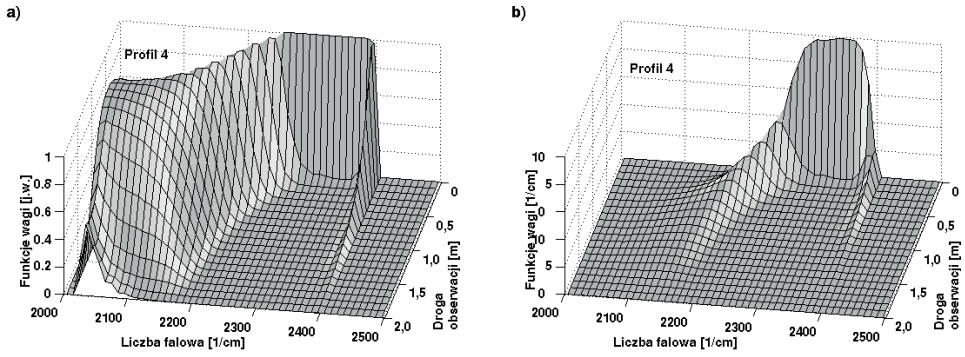
Rys. 65. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 3, 4 m)
 Fig. 65. Temperature profile and transmissivity of measurement path (profile 3, 4 m)



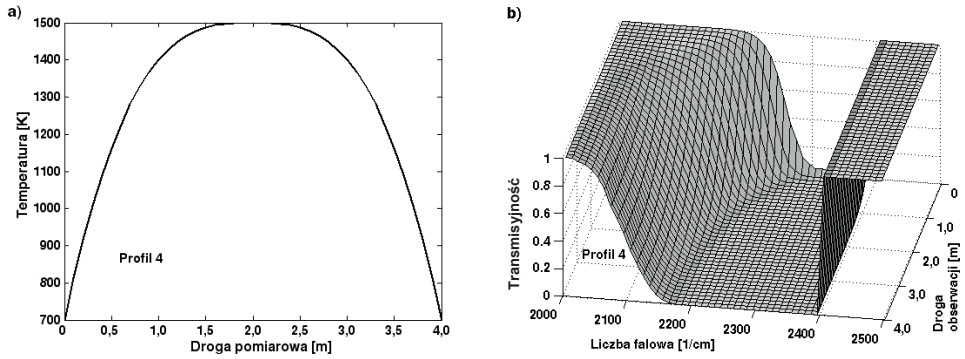
Rys. 66. Unormowane i zwykłe funkcje wagi (profil 3, 4m)
 Fig. 66. The normalized and normal weighting function (profile 3, 4 m)



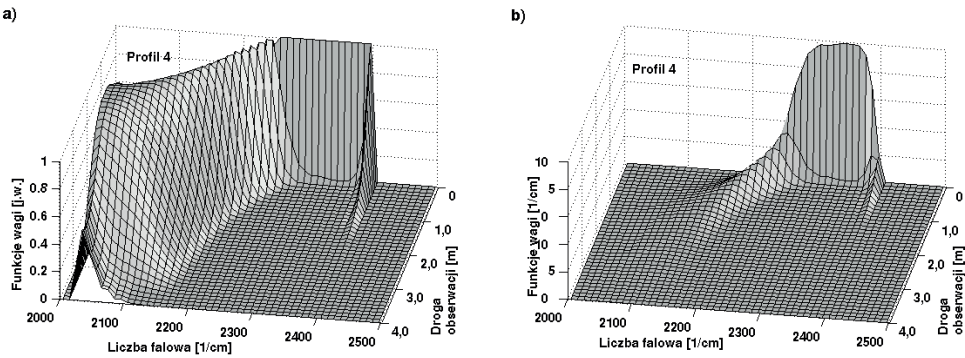
Rys. 67. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 4, 2 m)
 Fig. 67. Temperature profile and transmissivity of measurement path (profile 4, 2 m)



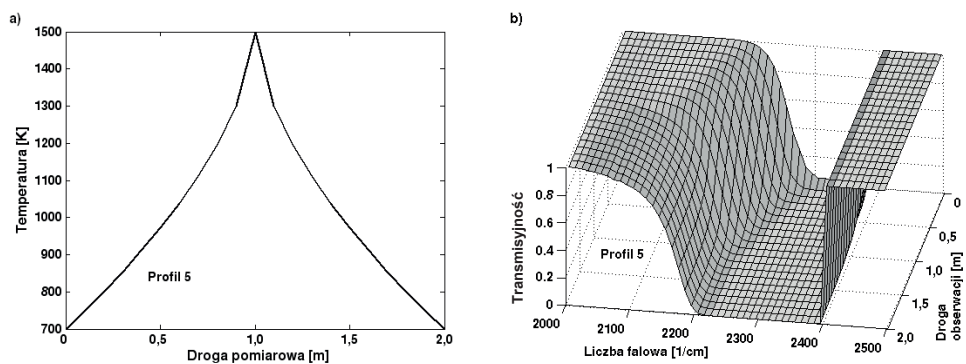
Rys. 68. Unormowane i zwykłe funkcje wagi (profil 4, 2 m)
 Fig. 68. The normalized and normal weighting function (profile 4, 2 m)



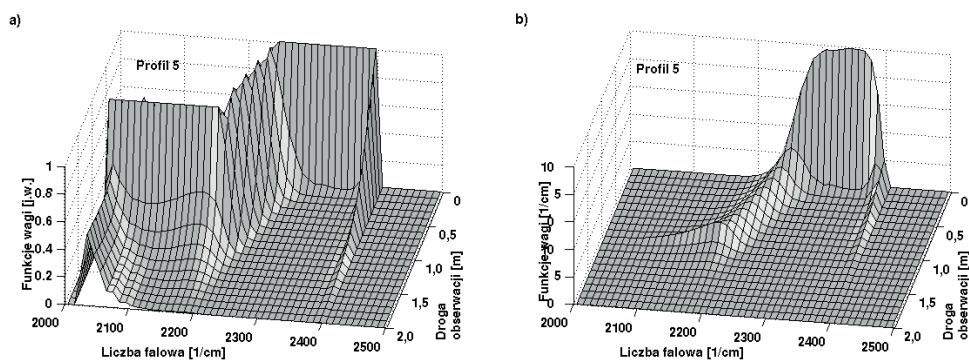
Rys. 69. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 4, 4 m)
 Fig. 69. Temperature profile and transmissivity of measurement path (profile 4, 4 m)



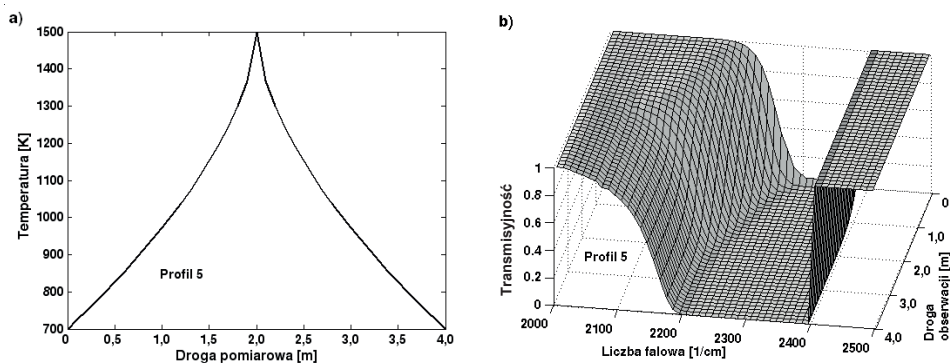
Rys. 70. Unormowane i zwykłe funkcje wagi (profil 4, 4 m)
 Fig. 70. The normalized and normal weighting function (profile 4, 4 m)



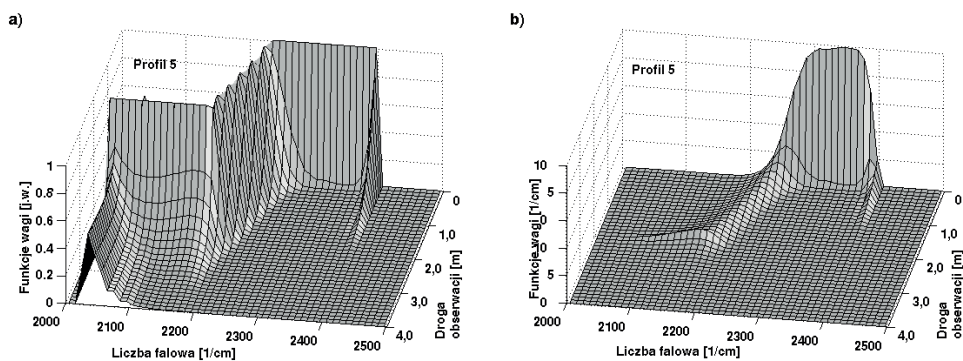
Rys. 71. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 5, 2 m)
 Fig. 71. Temperature profile and transmissivity of measurement path (profile 5, 2 m)



Rys. 72. Unormowane i zwykłe funkcje wagi (profil 5, 2 m)
 Fig. 72. The normalized and normal weighting function (profile 5, 2 m)



Rys. 73. Rozkład temperatury wzdłuż ścieżki pomiarowej oraz transmisyjność (profil 5, 4 m)
 Fig. 73. Temperature profile and transmissivity of measurement path (profile 5, 4 m)

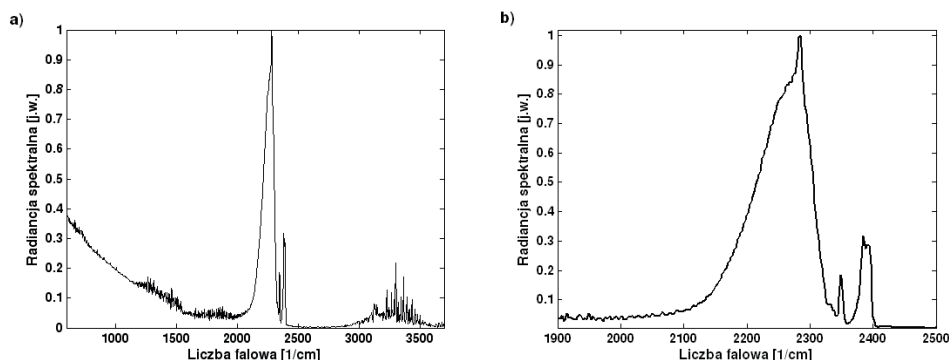


Rys. 74. Unormowane i zwykłe funkcje wagi (profil 5, 4 m)
 Fig. 74. The normalized and normal weighting function (profile 5, 4 m)

Z przeprowadzonych obliczeń funkcji wag, których wyniki zaprezentowano na rysunkach 51–74, wynika, że kształt funkcji wag zależy od: zawartości składników, długości ścieżki, temperatury oraz kształtu jej rozkładu. Dla unormowanej do jedności wagi obserwować można przesunięcie maksimum wraz ze zmianą kształtu rozkładu oraz średniego poziomu temperatury rozkładu. Zwykłe funkcje wagi pokazują natomiast wartości wag, z jakim intensywność promieniowania na danej długości fali wchodzi do intensywności obserwowanej przez sondę ($l = 0$). Wzrost długości drogi optycznej powoduje względne przesunięcie (w stosunku do długości drogi pomiarowej) maksimum funkcji wagi dla danego kanału w stronę punktu obserwacji (2015 cm^{-1}). Wzrastają wartości funkcji wag w kanałach spektralnych o słabszym tłumieniu, np. na końcach pasma wibracyjnego CO_2 ($2050\text{--}2150 \text{ cm}^{-1}$). Jeśli w danym paśmie spektralnym na obserwowany składnik nakłada się pasmo od innych gazów, to maksimum funkcji wagi przesunięte zostaje w kierunku obserwatora, dodatkowo zmniejsza się jego amplituda. Wzrost wartości temperatury konkretnej warstwy zmienia rozkład transmisyjności w kierunku granicy pasma wibracyjnego i pobudzone są bardziej jego brzegi ($2050\text{--}2250 \text{ cm}^{-1}$). Dla tych samych poziomów granicznych temperatury zmiana jej rozkładu wpływa na zmianę jej gradientu. W promieniowaniu bardziej widoczne będą miejsca o dużym gradientie temperatury z powodu silnych zmian transmisyjności tych warstw, co powoduje skupienie w tym obszarze większych wartości funkcji wagowych. Najbardziej charakterystycznym dla typowych procesów spalania jest rozkład osiowo-symetryczny lub prawie osiowo-symetryczny, mający maksimum temperatury w środku drogi optycznej. Dla tego rodzaju rozkładów maksymalne wartości funkcji wagi rozkładają się od punktu obserwacji do punktu temperatury maksymalnej. Przekraczając na ścieżce pomiarowej punkt temperatury maksymalnej, mamy do czynienia z obniżeniem efektywności estymacji parametrów rozkładu temperatury. W takim przypadku wykonywany jest pomiar promieniowania z dwu stron ścieżki obserwacji, lub jeśli to wskazane, założenie osiowej symetryczności rozkładu.

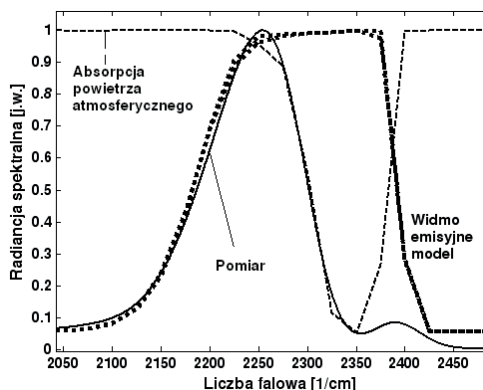
9. POMIARY I ICH ANALIZA

Wstępne pomiary widma promieniowania płomienia przeprowadzone zostały z wykorzystaniem palnika laboratoryjnego. Na rysunkach 75 i 76 widać charakterystyczne cechy spektralne dwutlenku węgla oraz pary wodnej. Na tle całości promieniowania w zakresie średniej podczerwieni zdecydowanie największą emisyjność ma CO_2 w paśmie spektralnym $2000\text{--}2500\text{ cm}^{-1}$. W zakresie $3000\text{--}3500\text{ cm}^{-1}$ oraz $1000\text{--}2000\text{ cm}^{-1}$ widoczna jest emisyjność pary wodnej. W przedstawionych pomiarach nie widać cech tlenku węgla, dlatego wnioskować można, że proces spalania przebiegał efektywnie. Wykorzystując model statystyczny SNB transmisyjności gazów, zaaprosymowano (rys. 76) widmo emisyjne dwutlenku węgla w zakresie $2000\text{--}2250\text{ cm}^{-1}$. Zakres spektralny $2250\text{--}2400\text{ cm}^{-1}$ odpowiada absorpcji promieniowania przez zimny dwutlenek węgla znajdujący się między mierzonym procesem a spektrometrem.



Rys. 75. Radiancja spektralna mierzona dla palnika laboratoryjnego

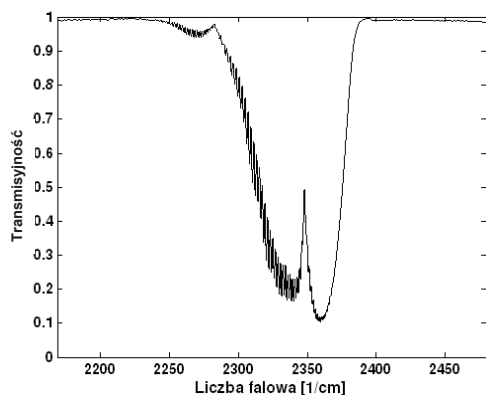
Fig. 75. Laboratory burner spectral radiance



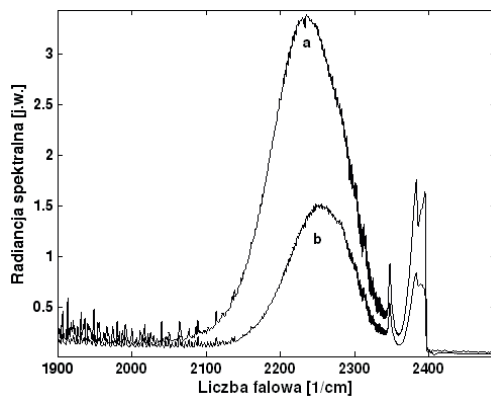
Rys. 76. Porównanie widma modelowanego z widmem mierzonym palnika laboratoryjnego

Fig. 76. Comparison of laboratory burner measured and calculated spectral radiance

Pomiary widm promieniowania przeprowadzone zostały na stanowisku do spalania gazów ze zgazowywarki biomasy o mocy 0,5 MW [8], [38]. Płomień był rozpalany z wykorzystaniem palnika gazowego. Spektrometr był oddalony od gorącej strefy, stąd we wszystkich widmach obserwować można absorpcję CO₂ znajdującego się w powietrzu. Aby pokazać wpływ części ścieżki pomiarowej, znajdującej się w powietrzu, wykonano pomiar absorpcji atmosferycznego dwutlenku węgla na tej samej odległości jak w przypadku pomiaru płomienia (rys. 77). Źródłem promieniowania były nagrzane elementy palnika. Zakres spektralny 2300–2400 cm⁻¹ jest zdominowany przez absorpcję zimnego CO₂ znajdującego się w powietrzu. Efekt ten można wyeliminować, stosując sondę światłowodową oraz światłowód do transmisji promieniowania z analizowanego procesu do spektrometru.



Rys. 77. Widmo absorpcyjne dwutlenku węgla zawartego w powietrzu
Fig. 77. Atmospheric carbon dioxide absorption spectrum

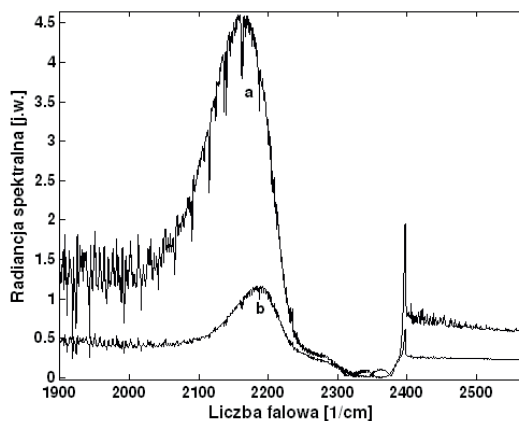


Rys. 78. Przykład widma emisyjnego:
a – podczas normalnej pracy palnika, pomiar prostopadle do osi płomienia, b – podczas rozpalania
Fig. 78. The emission from the flame showing CO₂ spectrum (measurement from the side of combustion chamber)

Długość obserwowanego płomienia wynosiła około 2 m. W pomiarach od strony wylotu spalin odległość pomiędzy palnikiem a otworem inspekcyjnym wynosiła 8 m. Na drodze tej spaliny ochładzają się w wyniku oddawania ciepła do chłodzonych ścian komory. Dlatego też promieniowanie płomienia jest absorbowane przez spaliny znajdujące się na ścieżce pomiarowej. Spaliny mają mniejszą temperaturę od warstw płomienia, dlatego na tle gorącej strefy płomienia powodują absorpcję. Emisja oraz absorpcja promieniowania następuje w określonych zakresach spektralnych (rys. 78, 79).

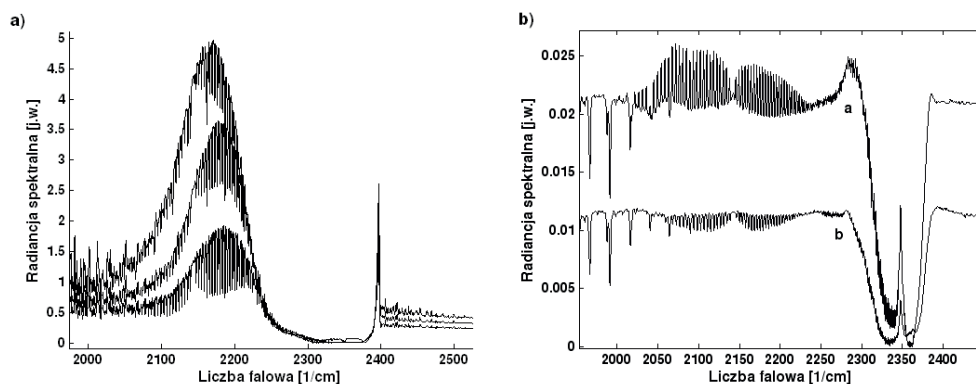
Pokazany na rysunkach 78 oraz 79 zakres spektralny podzielić można na kilka części. W zakresie 2300–2400 cm⁻¹ obserwować można silną absorpcję promieniowania przez chłodniejszy dwutlenek węgla, znajdujący się w kanale ze spalinami (rys. 78). W następnej części 2150–2300 cm⁻¹ zaobserwować można emisję promieniowa-

nia dwutlenku węgla zaraz po zapłonie oraz absorpcję podczas dalszej pracy palnika ze względu na nagrzewanie się kanału ze spalinami. Zaraz po zapłonie płomień ma stosunkowo niską temperaturę i promieniowanie CO_2 następuje w zakresie $2150\text{--}2300\text{ cm}^{-1}$ (rys. 79b). Zwiększanie temperatury płomienia powoduje zwiększenie emisji promieniowania dwutlenku węgla w trzecim zakresie $2100\text{--}2200\text{ cm}^{-1}$ (rys. 79a).



Rys. 79. Przykład widma emisyjnego mierzonego podczas normalnej pracy palnika, pomiar od strony wylotu spalin

Fig. 79. The emission from the flame showing CO_2 spectrum (measurement along combustion chamber)



Rys. 80. a – widmo emisyjne podczas wygaszania płomienia mierzone od strony wylotu spalin, b – przejście linii emisyjnych CO w linie absorpcyjne (pomiar prostopadle do osi płomienia)

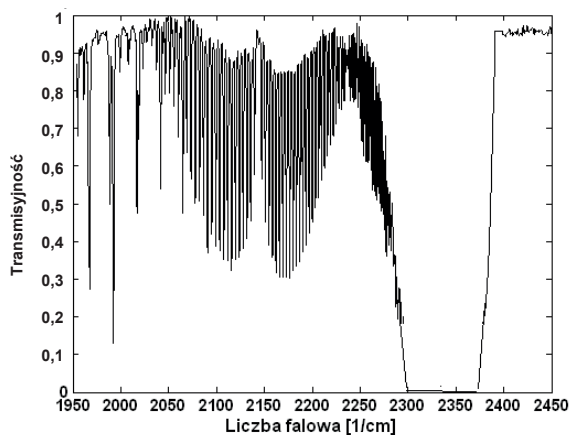
Fig. 80. a – emission spectra measured along combustion chamber, b – transmission from emission to absorption CO lines (measurement from the side of combustion chamber)

W widmie od strony wylotu spalin obserwuje się wyższy poziom promieniowania o charakterze ciągłym ($2500\text{--}2400\text{ cm}^{-1}$). W przypadku pomiarów z boku komory

spalania (rys. 78) odległość pomiędzy płomieniem a spektrometrem wynosiła 1 m. Absorpcja przez powietrze atmosferyczne była więc mniejsza niż dla pomiarów od strony wylotu spalin. Gdy spektrometr został ustawiony od strony wylotu spalin, promieniowanie pochodzące od płomienia pokonuje około 6 m w kanale ze spalinami oraz 1,5 m w powietrzu atmosferycznym. W zakresie $1900\text{--}2150\text{ cm}^{-1}$ widoczne są linie rotacyjne pary wodnej. Dla spektrometru ustawionego od strony wylotu spalin linie spektralne pary wodnej mają charakter absorpcyjny oraz emisyjny. Dla spektrometru ustawionego z boku palnika linie pary wodnej mają charakter emisyjny.

Na rysunku 80 widoczne są wyraźnie absorpcyjne linie rotacyjne tlenku węgla ($2050\text{--}2230\text{ cm}^{-1}$). Podczas powolnego procesu wygaszania linie te oznaczają, że w spalinach jest duża zawartość tlenku węgla. Linie rotacyjne CO widoczne są na tle emisyjnych cech CO_2 , pochodzących z płomienia ($2050\text{--}2250\text{ cm}^{-1}$). W trakcie procesu powolnego zmniejszania się wielkości płomienia oraz jego temperatury obserwować można zmniejszanie się intensywności promieniowania CO_2 . Zmniejszanie temperatury płomienia powoduje mniej efektywną konwersję CO w CO_2 .

Na rysunku 80b pokazane jest widmo z wyraźnymi cechami emisyjnymi CO. Pomiar wykonany był podczas znacznego zmniejszenia się temperatury płomienia. Powodowało to większą zawartość CO poza obszarem płomienia. W dalszym ciągu do palnika dostarczane były gazy ze zgazowywarki o dużej zawartości CO. Spektrometr ustawiony był z prostopadłe do osi płomienia w odległości ponad 2 m od wylotu palnika. Po obu stronach komory spalania, znajdują w tym miejscu otwory inspekcyjne. Promieniowanie nie pochodziło więc od nagranych ścian komory spalania, stąd możliwość uwidocznienia się pasma CO w postaci linii emisyjnych. Po pewnym czasie temperatura spalin zmniejszyła się, a linie emisyjne CO przeszły w linie absorpcyjne.

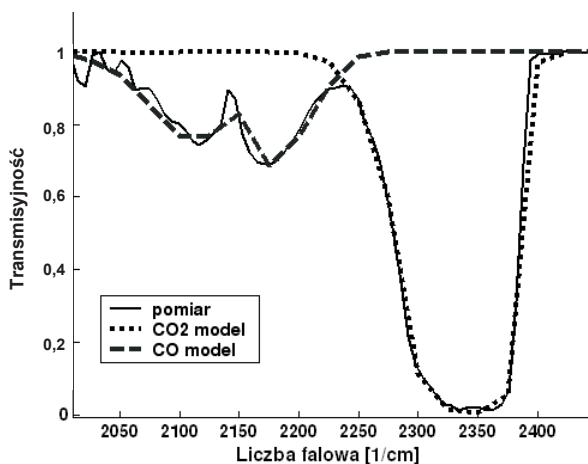


Rys. 81. Absorpcyjne widmo pomiarowe wysokiej rozdzielczości (1 cm^{-1} , apodyzacja trójkątna) zawierające linie spektralne CO oraz CO_2

Fig. 81. CO and CO_2 transmissivity spectrum measured with 1 cm^{-1} resolution

Na rysunku 81 przedstawiono widmo obserwowane od strony wylotu spalin w trakcie wygaszania płomienia. Widoczne są w nim cechy absorpcyjne tlenku węgla ($1950\text{--}2250\text{ cm}^{-1}$), ponieważ był on nadal dostarczany do palnika ze zgazowywarki. W zakresie $2250\text{--}2380\text{ cm}^{-1}$ widoczna jest absorpcja dwutlenku węgla. Jest to przykład, który można nazwać pasywnym widmem absorpcyjnym. Jest ono pasywne, ponieważ jako źródła promieniowania zostało wykorzystane naturalne promieniowanie gorących elementów palnika.

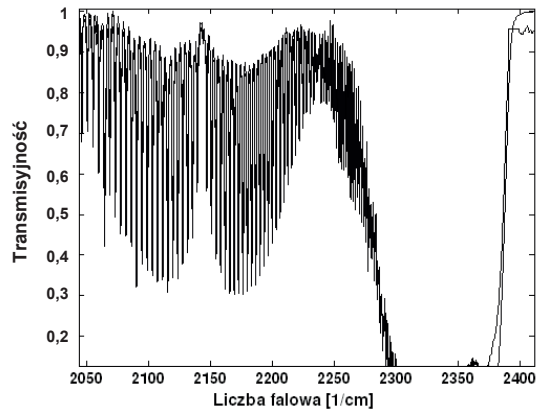
Widmo z rysunku 81 poddano wygładzeniu w kanałach o szerokości 10 cm^{-1} . Stąd też powstało widmo przedstawione na rysunku 82, które można analizować ilościowo z wykorzystaniem modeli SNB. Do dalszej analizy przyjęto jednorodność temperatury oraz koncentracji CO i CO₂ na ścieżce obserwacji przez płomień. Z wykorzystaniem modelu SNB dopasowano widmo pomiarowe do widm symulacyjnych CO oraz CO₂. Obliczona temperatura wyniosła 400 K, co jest bliskie pomiarom termoparami. Informacje, wyznaczone z wykorzystaniem modelu SNB, posłużyły jako wartości startowe dla dalszej analizy wykorzystującej dokładne modele gazów (Hitran, Hitemp oraz modelowanie charakterystyki spektralnej spektrometru).



Rys. 82. Porównanie widma pomiarowego niskiej rozdzielczości z widmami symulacyjnymi dla CO oraz CO₂

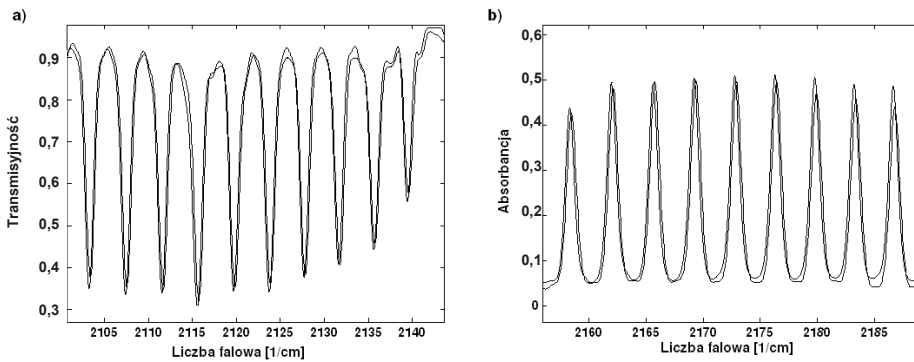
Fig. 82. Comparison of measured and simulated CO and CO₂ low resolution spectrum

Na rysunkach 83 i 84 pokazano dopasowanie widma pomiarowego do widma symulacyjnego w przypadku pomiarów spektrometrem FTIR o rozdzielczości 1 cm^{-1} i zastosowaniu apodyzacji trójkątnej. Otrzymano wynik $T = 390\text{ K}$, ułamek objętościowy zawartości CO 0,01, CO₂ 0,015, para wodna 0,1. Są to wielkości odniesione do ścieżki optycznej o długości jednego metra. Pomiar wykonany był na długości 8 m.



Rys. 83. Porównanie widma pomiarowego z widmami symulacyjnymi dla CO oraz CO₂ (rozdzielczość 1 cm⁻¹)

Fig. 83. Comparison of measured and simulated CO and CO₂ spectrum (1 cm⁻¹ resolution)



Rys. 84. Porównanie widma pomiarowego z widmem symulacyjnym dla CO (rozdzielczość 1 cm⁻¹): a – transmisyjność, b – absorbancja

Fig. 84. Comparison of measured and simulated CO spectrum (1 cm⁻¹ resolution), a – transmissivity, b – absorbance

9.1. PODSUMOWANIE I WNIOSKI

Jest potrzeba monitorowania różnorodnych procesów, w których zachodzi emisja gazów. Do procesów takich, odgrywających ostatnio coraz większą rolę, zaliczamy spalanie biomasy. Ponieważ spalaniu podlegają substancje lub też gazy powstałe podczas zgazowywania substancji o różnym składzie, proces taki wymaga metod pomiarowych analizujących zachodzące zjawisko w miejscu jego zachodzenia. Dostęp do

procesu umożliwiającą spektrometry o otwartej ścieżce pomiarowej, pozwalającą na wykonanie pomiarów zdalnych. Za pomocą takich urządzeń tworzyć można aktywne, jak i pasywne konfiguracje pomiarowe. W niektórych procesach, szczególnie zachodzących w zamkniętych obiektach, możliwy jest jedynie pasywny pomiar widm emisyjnych.

Mierzone widma, a szczególnie emisyjne, nie mogą być analizowane w klasyczny sposób znany ze spektroskopii laboratoryjnej. Główne powody, to: zmiana zarówno składników, jak i ich temperatury, występujące nieliniowości pomiędzy mierzonym widmem a zawartością składnika charakterystyczne dla badanych ścieżek pomiarowych. Do analizy takich widm wykorzystać należy modele transmisyjności gazów. Dla widm emisyjnych wykorzystać należy równanie transportu energii radiacyjnej. Modelowanie widma potraktować można jako zagadnienie bezpośrednie, w którym na podstawie wielkości takich, jak zawartość oraz temperatura składników wyznaczamy odpowiednie widmo. Zagadnienie odwrotne będzie polegało na wyznaczeniu zawartości składników i temperatury na podstawie zmierzonego widma. W procesie ilościowej analizy wyników pomiarów mamy do czynienia właśnie z zagadnieniem odwrotnym.

Aby określić możliwości wyznaczania poszczególnych parametrów należy przeprowadzić analizę wrażliwości. Oczywiście wykonuje się ją dla przewidywanych zakresów wyznaczanych wartości.

Praktyczne zastosowanie opisanej metody przedstawione zostało na przykładzie pomiarów widm ze zgazowywarki biomasy. Wyznaczono zawartość składników oraz temperatury na podstawie widm mierzonych pasywnie.

LITERATURA

- [1] ASFAW A., *A fast method of modeling spectral lines*, JoQS&RT70, 2001, pp. 129–137.
- [2] BAK J., CLAUSEN S., *FTIR transmission-emission spectrometry of gases at high temperatures: demonstration of Kirchhoff's law for a gas in an enclosure*, JoQS&RT, Vol. 61, No 5, 1999, pp. 687–694.
- [3] BATES S.C., MORRISON P.W., SOLOMON Jr. P.R., *Infrared Monitoring of Combustion*, Environmental Sensing and Combustion Diagnostics, SPIE, Vol. 1434, 1991, pp. 28–38.
- [4] BERNARDO C., GRIFFITH D., *Fourier transform spectrometer instrument lineshape (ILS) retrieval by Fourier deconvolution*, JoQS&RT, 95, 2005, pp. 141–150.
- [5] BIANCHINI G., RASPOLINI P., *Characterisation of instrumental line shape distortion due to path difference dependent phase errors in a Fourier transform spectrometer*, Infrared Physics & Technology, 41, 2000, pp. 287–292.
- [6] BIELECKI Z., ROGALSKI A., *Detekcja sygnałów optycznych*, Wydawnictwa Naukowo-Techniczne, Warszawa 2001.
- [7] BLOKH A.G., *Heat transfer in steam boiler furnaces*, Springer-Verlag, Berlin 1988.
- [8] CIĘSZCZYK S., *Pasywna diagnostyka procesu spalania*, Rozprawa doktorska, Lublin 2008.
- [9] CIĘSZCZYK S., *Zagadnienie odwrotne w pasywnych pomiarach optycznych*, materiały konferencyjne: VIII International Workshop for Candidates for a Doctor's Degree OWD'2006, Wisła, 21–24 October, 2006, pp. 271–276.

- [10] DE CASTRO A.J. et al., *Open-path Fourier transform infrared spectrometry characterization of low temperature combustion gases in biomass fuels*, Infrared Physics & Technology, 51, 2007, pp. 21–30.
- [11] CHRISTIAN T.J. et al., *Comprehensive laboratory measurements of biomass-burning emissions: 2. First intercomparison of open-path FTIR, PTR-MS, and GC-MS/FID/ECD*, Journal of Geophysical Research, Vol. 109, 2004, D02311.
- [12] CLAUSEN S., *Infrared combustion diagnostics in fluctuating flames*, Air Pollution and Visibility Measurements, Munich, Federal Republic of Germany, 19–23. 06. 95, Proc. SPIE, Vol. 2506, pp. 30–44.
- [13] DE IULIIS S., BARBINI M., BENECCHI S., CIGNOLI F., ZIZAK G., *Determination of the Soot Volume Fraction in an Ethylene Diffusion Flame by Multiwavelength Analysis of Soot Radiation*, Combustion and Flame, 115, 1998, pp. 253–261.
- [14] DENTRODER W., *Spektroskopia laserowa*, PWN, Warszawa 1993.
- [15] DOICU A., SCHREIER F., HESS M., *An iterative regularization method with B-spline approximation for atmospheric temperature and concentration retrievals*, Environmental Modeling & Software, 20, 2005, pp. 1101–1109.
- [16] DOICU A., SCHREIER F., HESS M., *Iterative regularization methods for atmospheric remote sensing*, JoQS&RT, 83, 2004, pp. 47–61.
- [17] DOICU A., SCHREIER F., HESS M., *Iteratively regularized Gauss-Newton method for atmospheric remote sensing*, Computer Physics Communication, 148, 2002, pp. 214–226.
- [18] FLORES-JARDINES E. et al., *Investigation of temperature and gas concentration distribution in hot exhausts (airplanes and burners) by scanning imaging FTIR spectrometry*, SPIE, Vol. 5979, 2005, pp. 365–376.
- [19] GRIFFITHS P.R., DE HASETH J.A., *Fourier Transform Infrared Spectrometry*, John Wiley & Sons, Inc., 2007, Hoboken, New Jersey.
- [20] KOSTOWSKI E., *Promieniowanie cieplne*, PWN, Warszawa 1993.
- [21] KOSTOWSKI E., *Przepływ ciepła*, Wydawnictwo Politechniki Śląskiej, Gliwice 2000.
- [22] KHAN T., SMIRNOVA A., *1D inverse problem In diffusion based optical tomography using iteratively regularized Gauss-Newton algorithm*, Applied Mathematics and Computation, 161, 2005, pp. 149–170.
- [23] MARRAN D., COSGROVE J., NEIRA J., MARKHAM J., RUTKA R., STRANGE R., *Turbine engine exhaust gas measurements using in-situ FT-IR emission/transmission spectroscopy*, Optical Methods for Industrial Processes, Proceedings of SPIE, Vol. 4201, 2001, pp. 118–127.
- [24] MARKHAM J.R., HANG Y.P., CARANGELO R.M., SOLOMON P.R., *FT-IR emission transmission tomography of coal flame Twenty-Third Symposium on Combustion*, The Combustion Institute, 1990, pp. 1869–1975.
- [25] MIHALCEA R.M., BAER D.S., HANSON R.K., *A diode-laser absorption sensor system for combustion emission measurements*, Meas. Sci. Technol., 9, 1998, pp. 327–338.
- [26] MILES J. et al., *Analysis of thermal radiation in coal-fired furnaces*, Thermosense XIX: An International Conference on Thermal Sensing and Imaging Diagnostic Applications, Apr. 1997, Proc. SPIE, Vol. 3056, pp. 20–32.
- [27] MORRISON P.W. Jr., COSGROVE J.E., CARANGELO R.M., CARANGELO M.D., SOLOMON P.R., LEROUEIL P., THORN P.A., *Fourier transform infrared (FTIR) instrumentation for monitoring recovery boilers*, Tappi Journal, Vol. 74, No. 12, December 1991, pp. 68–78.
- [28] AL-NIMIR M.A., ARPACI V.S., NAJJAR Y.S., *Temperature and soot concentration in a high soot density flame*, Int. J. Heat Mass Transfer., Vol. 39, No. 6, 1996, pp. 1235–1241.
- [29] PIOTROWSKI J., KOSTYRKO K., *Wzorcowanie aparatury pomiarowej*, PWN, Warszawa 2000.
- [30] POLAK A.G., MROCZKA J., *Pośrednie pomiary właściwości obiektów złożonych*, Problemy metrologii elektrycznej i fotonicznej (red. J. Mroczka), Oficyna Wydawnicza PWr, Wrocław 2008, pp. 15–78.

- [31] ROTHMAN L.S. et al., *The HITRAN 2004 molecular spectroscopic database*, Journal of Quantitative Spectroscopy and Radiative Transfer, 96, 2005, pp. 139–204.
- [32] RUDNICKI Z., *Radiacyjny przepływ ciepła w piecach przemysłowych*, Wydawnictwo Politechniki Śląskiej, Gliwice 1998.
- [33] SOUFIANI A., TAINÉ J., *High temperature gas radiative property parameters of statistical narrow-band model for H₂O, CO₂ and CO, and correlated-K model for H₂O and CO₂*, Int. J. Heat Mass Transfer, Vol. 40, No. 4, 1997, pp. 987–991.
- [34] USTINOV E., *Weighting function with separated maxima and kernel of inverse problems for ground-based passive remote sensing*, JoQSRT, 66, 2000, pp. 55–67.
- [35] WEBBER M.E., *Diode laser measurements of NH₃ and CO₂ for combustion and bioreactor applications*, PhD Thesis, January 2001, Stanford University.
- [36] WEISSINGER A., FLECKL T., OBERNBERGER I., *In situ FTIR spectroscopic investigations of species from biomass fuels in a laboratory-scale combustor: the release of nitrogenous species*, Combustion and Flame, 137, 2004, pp. 403–417.
- [37] WÓJCIK W., CIĘSZCZYK S., KOMADA P., KISAŁA P., *Pomiary widma procesów spalania do celów diagnostycznych w zakresie średniej podczerwieni*, Diagnostyka Procesów i Systemów, EXIT, pp. 271–276, Warszawa 2007.
- [38] WÓJCIK W., CIĘSZCZYK S., KOMADA P., KISAŁA P., *Pomiary widm procesów spalania z wykorzystaniem spektrometru FTIR*, Elektronika, 6, 2008, 230–232.
- [39] YOKELSON R.J., GRIFFITH D.W., WARD D.E., *Open-path Fourier transform infrared studies of large-scale laboratory biomass fires*, Journal of Geophysical Research, Vol. 101, 1996, pp. 21,067–21,080.
- [40] ZHANG H., MODEST M., *Evaluation of the Planck-mean absorption coefficients from HITRAN and HITEMP databases*, JoQS&RT, Vol. 73, No. 6, 2002, pp. 649–653.

INFORMATION RETRIEVAL FROM IN-SITU EMISSION AND
ABSORPTION SPECTRUM MEASUREMENTS OF
GASES BASED ON MATHEMATICAL MODELING

This article presents analysis method of mid-infrared spectrum, based on mathematical modeling. Such method allow, determining information from in-situ spectrum measurements of gases. Processes such as combustion or substance gasification are characterized by not only gas concentration variation but also temperature fluctuation. Direct models, based on mathematical modeling, transform temperature and gas concentration into emission or absorption spectrum. Parameter retrieval based on spectrum measurement is called an inverse problem.

Czesław ŁUKIANOWICZ*

SKATEROMETRIA POWIERZCHNI NIERÓWNYCH

W pracy przedstawiono podstawy pomiarów nierówności powierzchni z wykorzystaniem zjawiska rozpraszania światła. Pomiaru takie znajdują zastosowanie w wielu dziedzinach nauki i techniki, między innymi w takich, w których pojawia się potrzeba dokonania szybkiej i bezstykowej oceny nierówności precyzyjnie obrobionych powierzchni. Praca w syntetycznej formie opisuje podstawowe teorie wyjaśniające zjawisko rozpraszania światła przez powierzchnie nierówne oraz najważniejsze techniki i metody pomiaru nierówności powierzchni oparte na analizie światła rozproszonego przez powierzchnię. Większość z tych metod pozwala oceniać nierówności stosunkowo gładkich powierzchni, których wysokość nierówności jest porównywalna z długością fali światła. W końcowej części pracy sformułowano wnioski wynikające z przeprowadzonych rozważań i analiz. Wnioski te, podobnie jak cała praca, mogą być punktem wyjścia do dalszych badań w zakresie skaterometrii powierzchni nierównych.

1. WPROWADZENIE

1.1. NIERÓWNOŚCI POWIERZCHNI I ICH POMIAR

Powierzchnie rzeczywiste są na ogół nierówne. Pierwotną przyczyną istnienia nierówności powierzchni jest korpuskularna budowa materii. Jednak największy wpływ na postać, strukturę i wymiary nierówności powierzchni mają te procesy, w wyniku których kształtowana jest powierzchnia przedmiotu. Nierówności powierzchni przedmiotów tworzonych przez człowieka przyjmują najczęściej charakterystyczną postać, zależną w głównej mierze od zastosowanych narzędzi, materiałów i sposobów obróbki powierzchni [14], [13]. Mają one wpływ na wiele różnorodnych właściwości powierzchni. Dlatego też w wielu dziedzinach nauki i techniki występuje potrzeba ilościowej oceny nierówności powierzchni. Dokonanie takiej oceny wymaga przede wszystkim odpowiedniego – często znormalizowanego – opisu powierzchni nierównych. Opis ten jest na ogół uproszczony i dostosowany do aktualnego stanu techniki

* Politechnika Koszalińska, Wydział Mechaniczny, Katedra Inżynierii Produkcji, Zakład Metrologii i Jakości, ul. Raclawicka 15-17, 75-620 Koszalin.

pomiarowej. Zmienia się on wraz z rozwojem tej techniki i staje się coraz pełniejszy. Świadczą o tym między innymi zmiany w normach krajowych i międzynarodowych, dotyczących struktury geometrycznej powierzchni, zachodzące w ostatnich kilkudziesięciu latach. Normy te zawierają coraz więcej parametrów i funkcji charakteryzujących powierzchnie nierówne. W celu ilościowej oceny nierówności powierzchni, opracowano także kilka różnorodnych sposobów, pozwalających mierzyć wybrane geometryczne właściwości powierzchni nierównych. Ilościowemu opisowi nierówności powierzchni oraz różnym metodom ich pomiaru poświęcono wiele artykułów, rozpraw naukowych, a także liczne książki. Najbardziej fundamentalne opracowania książkowe w tym zakresie zostały przygotowane przez D.J. Whitehouse'a [71], [70], T.R. Thomasa [63], K.J. Stouta i L. Blunta [59]. Z książek polskich trzeba tu wymienić prace B. Nowickiego [48], M. Wieczorowskiego, A. Cellarego i J. Chajdy [72], K.E. Oczosia i V. Lubimova [49] oraz P. Pawlusa [53].

Pomiary nierówności powierzchni dokonywane są zarówno w laboratoriach badawczych, jak i w warunkach produkcyjnych. Ważną rolę w tych pomiarach odgrywają metody optyczne, a wśród nich metody oparte na zjawisku rozpraszania światła przez powierzchnie nierówne. Od kilkunastu lat obserwuje się wzrost zainteresowania metodami pomiaru nierówności powierzchni wykorzystującymi rozpraszanie światła. Zbiór takich metod oceny i pomiaru nierówności powierzchni nazywany jest niekiedy skaterometrią nierówności powierzchni [15] lub skaterometrią powierzchni obrobionych [4], [38]. Słowo skaterometria (ang. *scatterometry*) ma szersze znaczenie. Oznacza ono techniki pomiarowe oparte na wykorzystaniu zjawiska rozpraszania promieniowania. Dzięki takim metodom pomiarowym możliwa jest na przykład ocena chropowatości powierzchni podczas ruchu przedmiotu mierzonego [68], [69], [36], [37], co ma istotne znaczenie w zastosowaniach praktycznych. Mimo wielu zalet – do których można zaliczyć przede wszystkim bezstykowy sposób pomiaru, krótki czas pomiaru, a także wysoką czułość – istotną przeszkodą w rozwoju i szerszym wykorzystaniu tych metod jest złożoność zjawiska rozpraszania światła przez powierzchnie nierówne. Jednym z najważniejszych celów niniejszej pracy jest pełniejsze przedstawienie tego zjawiska, ze względu na możliwości jego wykorzystania w metrologii nierówności powierzchni.

1.2. ZJAWISKO ROZPRASZANIA ŚWIATŁA

Ze zjawiskiem rozpraszania światła przez nierówności powierzchni spotykamy się bardzo często. Wzrokowe postrzeganie większości otaczających nas obiektów jest możliwe dzięki temu, że nierówne powierzchnie tych obiektów odbijają i rozpraszają padające na nie światło. Oko ludzkie doskonale reaguje na efekt rozpraszania światła i łatwo odróżnia powierzchnie odbijające światło w sposób zwierciadlany od powierzchni nierównych, które rozpraszają światło w wielu kierunkach. Wizualna obserwacja powierzchni, ustawianej pod różnymi kątami w stosunku do obserwatora

i padającego na nią światła, umożliwia wykrycie drobnych wad, rys, zniekształceń i zanieczyszczeń powierzchni. Wizualna obserwacja powierzchni często jest wspomagana także przyrządami optycznymi, takimi np. jak lupa lub mikroskop. Może być ona podstawą jakościowej oceny stanu powierzchni. Ilościowa ocena tego stanu wymaga właściwej interpretacji zjawiska odbicia światła od powierzchni oraz przeprowadzenia odpowiednich pomiarów.

Oddziaływanie światła z materią odbijającego go obiektu rozpatruje się najczęściej jako skomplikowany proces elektrodynamiczny. Polega on na pobudzaniu do drgań naładowanych elektrycznie cząsteczek obiektu. Wymuszone drgania cząsteczek – wywołane przez elektromagnetyczną falę świetlną, padającą na powierzchnię obiektu – generują wtórne fale elektromagnetyczne. Te wtórne fale są spójne i interferują ze sobą. W tym wyraża się ogólna zasada Huygensa–Fresnela, stanowiąca podstawę optyki falowej [10]. Procesy generacji i interferencji fal wtórnych decydują o kierunku propagacji i natężeniu fali odbitej.

Jeśli powierzchnia odbijająca jest idealnie płaska i optycznie jednorodna, to kierunek propagacji płaskiej fali światła, odbitej od tej powierzchni, określony jest prawem odbicia światła i zależy od kierunku propagacji fali padającej. Fala świetlna odbita od takiej powierzchni nie ulegnie rozproszeniu. Można wykazać, że w tej sytuacji, zgodnie z zasadą Huygensa–Fresnela, wynik interferencji fal wtórnych będzie zerowy dla wszystkich kierunków, z wyjątkiem kierunku odbicia zwierciadlanego. Co się jednak stanie, gdy powierzchnia odbijająca światło będzie powierzchnią nierówną? Odpowiedź na to pytanie nie jest prosta. W przypadku pewnych powierzchni nierówności mogą występować w postaci oddzielnych wgłębień lub wzniesień, powodujących miejscowe, lokalne odstępstwa powierzchni od idealnej płaszczyzny. Znacznie częściej nierówności przyjmują postać okresowych lub nieokresowych odchyłeń od płaszczyzny, występujących w wielu punktach powierzchni. Na postawione wcześniej pytanie można udzielić jedynie bardzo ogólnej odpowiedzi. Pojawienie się na powierzchni odbijającej nawet niewielkich nierówności zmieni przebieg procesu interferencji fal wtórnych. W rezultacie tego fala świetlna, odbita od powierzchni nierównej, będzie propagować także w innych kierunkach niż kierunek odbicia zwierciadlanego. Proces ten, nazywany rozpraszaniem światła na powierzchni nierównej, opisywany jest na ogół za pomocą teorii dyfrakcji światła.

Zjawisko rozpraszania fal elektromagnetycznych przez powierzchnie nierówne było przedmiotem licznych badań. W ich wyniku uzyskano wiele danych doświadczalnych. Opracowano też wiele teorii rozpraszania światła – opartych głównie na teorii dyfrakcji fal – umożliwiających przewidywanie wyników eksperymentów. Teorie te są na ogół bardzo złożone, co utrudnia ich wykorzystanie w praktyce pomiarowej. Do prac, w których najpełniej zaprezentowano różne teorie rozpraszania fal przez powierzchnie nierówne, należy zaliczyć obszerne opracowania P. Beckmanna i A. Spizzichino [6], F.G. Bassa i I.M. Fuksa [3], J.A. Ogilvy [51], J. Stovera [60], a także pracę A.A. Maradudina [42]. Upraszczenie tych teorii prowadzi najczęściej do ograniczenia zakresu ich

stosowalności i zmniejsza dokładność opisu zjawiska. Wykorzystanie teorii rozpraszania światła w metrologii nierówności powierzchni komplikuje się dodatkowo, ze względu na konieczność spełnienia przez układ pomiarowy i powierzchnię wielu założeń, postulowanych przez te teorie. Analiza możliwości zastosowania niektórych z tych teorii w metrologii nierówności powierzchni została oparta na monografii [41] i stanowi główny temat niniejszej pracy.

2. TEORETYCZNE PODSTAWY ROZPRASZANIA ŚWIATŁA PRZEZ POWIERZCHNIE NIERÓWNE

Rozpraszanie światła jest zjawiskiem towarzyszącym propagacji światła w ośrodkach materialnych. Przejawia się ono w takim oddziaływaniu materii ośrodka i fali świetlnej, które powoduje, że część energii światła wysyłana jest w innych kierunkach niż kierunek propagacji fali pierwotnej. Na gruncie falowej teorii światła, oddziaływanie to wyjaśniane jest najczęściej pochłanianiem energii fali pierwotnej przez nieoświetlone elektrycznie cząstki ośrodka materialnego, np. elektrony. Proces absorpcji energii fali świetlnej wywołuje wymuszone drgania cząstek, co prowadzi do generowania i reemisji przez te cząstki fal wtórnych. Fale wtórne są falami sferycznymi. Jeśli ośrodek materialny jest optycznie jednorodny i izotropowy, to w wyniku interferencji spójnych fal wtórnych powstaje jedynie fala świetlna, której kierunek propagacji jest zgodny z kierunkiem fali pierwotnej. W tym przypadku światło nie jest rozpraszane. Wszelkie niejednorodności optyczne ośrodka materialnego powodują, że fale wtórne nie są spójne. W rezultacie ośrodek materialny rozprasza światło w różnych kierunkach. Jeśli częstotliwość wymuszonych drgań cząstek i częstotliwość fal wtórnych są równe częstotliwości pierwotnej fali świetlnej, to takie rozpraszanie nazywamy rozpraszaniem sprężystym. Spośród wielu rodzajów rozpraszania światła w tej pracy rozpatrzono tylko rozpraszanie sprężyste.

Idea interferencji fal wtórnych wyjaśnia zjawisko rozpraszania światła w ośrodku materialnym. Podobna myśl legła u podstaw teorii Fresnela, tłumaczącej prostoliniowe rozchodzenie się światła. Zgodnie z tą teorią naruszenie jednorodności ośrodka wywołuje zjawiska dyfrakcyjne. Dlatego teoria dyfrakcji jest podstawowym narzędziem optyki, stosowanym w analizie rozpraszania światła przez ośrodki niejednorodne, w tym przez nierówności powierzchni.

Kwantowa teoria światła objaśnia zjawisko rozpraszania światła zderzeniami fotonów z cząstkami materii. W wyniku tych zderzeń fotony i cząstki materii często zachowują swą pierwotną energię. Zderzenia takie nazywane są zderzeniami sprężystymi. Jeśli w ośrodku materialnym istnieją niejednorodności optyczne, to zderzenia sprężyste zmieniają kierunek propagacji niektórych fotonów. Prowadzi to do rozpraszania światła, przy czym częstość światła rozproszonego nie ulega zmianie. Takie rozpraszanie światła nazywane jest rozpraszaniem sprężystym. Rozpraszanie niesprężyste

żyste powstaje wtedy, gdy foton – podczas zderzenia z cząstką – utraci lub uzyska dodatkową energię. Zmiana energii fotonu powoduje zmianę częstości światła rozproszonego. Przykładem niesprężystego rozpraszania światła może być molekularne rozpraszanie Brillouina i rozpraszanie Ramana [31], [30]. Ten rodzaj rozpraszania światła nie jest rozpatrywany w tej pracy.

2.1. PROSTE MODELE ROZPRASZANIA ŚWIATŁA PRZEZ POWIERZCHNIE NIERÓWNE

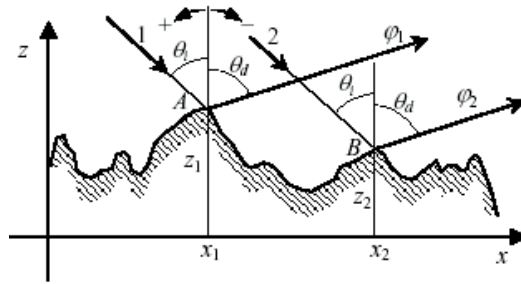
Jak już o tym wspomniano, rozpraszanie światła zachodzi wtedy, gdy na swej drodze napotyka ono niejednorodności ośrodka optycznego. W wyniku tego zmieniają się lokalnie kierunki propagacji światła. Niejednorodności ośrodka często występują w jego wnętrzu lub na granicy rozdzielającej dwa ośrodki optyczne. Wewnętrzna niejednorodność ośrodka może być spowodowana np. wtrąceniami oraz fluktuacjami gęstości i anizotropii materiału. Światło rozprzestrzeniając się napotyka niekiedy granicę rozdziału dwóch różnych, lecz wewnętrznie jednorodnych, ośrodków optycznych. Jeśli granica ta jest nierówna, to światło ulega na niej rozproszeniu, podobnie jak w ośrodku niejednorodnym. Stwierdzono, że zjawisko rozpraszania światła na granicy dwóch ośrodków optycznych ujawnia się tym wyraźniej, im bardziej nierówna jest ta granica. Na tym spostrzeżeniu opiera się podstawowa idea skaterometrii nierówności powierzchni.

Analiza zjawiska rozpraszania światła przez powierzchnie chropowate dokonywana jest najczęściej metodami optyki fizycznej. Światło rozpatrywane jest wtedy jako fala elektromagnetyczna, która, odbijając się od powierzchni nierównej, zostaje zmodulowana fazowo i amplitudowo. Zmienia się także stan polaryzacji tej fali. W wielu pracach ignoruje się jednak stan polaryzacji światła rozproszonego, analizując jedynie skalarną amplitudę pola elektrycznego.

2.1.1. KRYTERIUM RAYLEIGHA

W analizie rozpraszania światła ważne jest określenie warunków, dla których odbicie światła zmienia się z odbicia zwierciadlanego w odbicie dyfuzyjne. Choć warunki takie nie mogą być ściśle określone, warto jest rozpatrzyć to zagadnienie po to, aby ustalić, jakie czynniki wpływają na rozpraszanie światła przez powierzchnię.

O tym jak zostanie odbite światło od powierzchni decydują nie tylko jej nierówności, ale także długość fali oraz kąt padania światła na powierzchnię. Pierwsze rozważania dotyczące rozpraszania fal akustycznych przeprowadził Lord Rayleigh. Zostały one później uogólnione na przypadek odbicia fal elektromagnetycznych.



Rys. 1. Odbicie fali od powierzchni nierównej
Fig. 1. Reflection of the wave from rough surface

Niech płaska fala światła o długości λ pada pod kątem θ_i na powierzchnię nierówną (rys. 1). Różnica dróg optycznych Δr między falami odbitymi w punktach A i B wynosi

$$\Delta r = (x_2 - x_1)(\sin \theta_d + \sin \theta_i) + (z_1 - z_2)(\cos \theta_i + \cos \theta_d), \quad (1)$$

gdzie z_1 i z_2 są wysokościami nierówności w punktach o współrzędnych x_1 i x_2 , natomiast θ_i i θ_d oznaczają odpowiednio kąt padania i kąt rozproszenia. Znaki kątów pokazano na rysunku 1. W przypadku, gdy $\theta_d = -\theta_i$, otrzymamy

$$\Delta r = 2\Delta z \cos \theta_i, \quad (2)$$

gdzie $\Delta z = z_1 - z_2$ jest różnicą wysokości nierówności z_1 i z_2 . Różnica faz zatem $\Delta \varphi = \varphi_2 - \varphi_1$ między obiema częściami fali równa jest

$$\Delta \varphi = \frac{2\pi}{\lambda} \Delta r = 2k\Delta z \cos \theta_i, \quad (3)$$

przy czym symbol $k = 2\pi/\lambda$ oznacza liczbę falową.

Jeżeli różnica faz $\Delta \varphi$ wynosi zero, to odbicie fali od powierzchni nierównej jest odbiciem zwierciadlanym. Jeśli zaś bezwzględna wartość różnicy faz $|\Delta \varphi| = \pi$ radianów, to obie części fali odbitej są w przeciwnej fazie. W wyniku interferencji nastąpi zatem wygaszenie światła propagującego w kierunku odbicia zwierciadlanego. W tym przypadku, zgodnie z zasadą zachowania energii, światło odbite od powierzchni musi propagować w innych kierunkach niż kierunek odbicia zwierciadlanego. Energia światła odbitego od powierzchni ulega rozproszeniu w różnych kierunkach. Rayleigh przyjął, że powierzchnie wprowadzające bezwzględną różnicę faz $|\Delta \varphi| < \pi/2$ rad będą zaliczane do powierzchni gładkich. Po nałożeniu tego warunku na prawą stronę wyrażenia (3) otrzymuje się tzw. kryterium Rayleigha:

$$\Delta z_{\max} < \frac{\lambda}{8 \cos \theta_i}. \quad (4)$$

Określa ono, jaka jest największa, dopuszczalna według tego kryterium, wysokość nierówności Δz_{\max} powierzchni gładkiej. Oczywiście kryterium to jest umowne i niedoskonałe. Jednakże ważną zaletą tego kryterium jest jego prostota oraz to, że określone jest ono za pomocą trzech łatwych do zmierzenia parametrów geometrycznych.

2.1.2. MODEL OPARTY NA ZASADACH OPTYKI GEOMETRYCZNEJ

Niekiedy, w celu uproszczenia analizy, falowy model rozpraszania światła zastępowany bywa modelem promieniowym. Otrzymuje się go, zakładając, że długość fali światła dąży do zera. Powierzchnia nierówna traktowana jest wtedy jak zbiór miniaturowych zwierciadeł, od których światło odbija się zgodnie z prawem odbicia. Model ten, oparty na zasadach optyki geometrycznej, ze względu na swoją prostotę jest chętnie używany w praktyce [12], zwłaszcza w odniesieniu do powierzchni o dużych nierównościach. W języku angielskim określa się go terminem *mirror facet model*. Jego przydatność do wnikliwego opisu zjawiska rozpraszania światła jest niewielka. Jednak może on być pożyteczny w zastosowaniach praktycznych, gdzie dopuszcza się znaczne uproszczenia. Podobny sposób modelowania nierówności powierzchni losowych zaproponował P. Beckmann w pracy [6], stosując teorię łańcuchów Markowa.

Analiza zjawiska rozpraszania światła przez powierzchnie nierówne oparta na prostym modelu wykorzystującym zasady optyki geometrycznej może być przydatna np. podczas określania ogólnej postaci rozkładu natężenia światła rozproszonego [64], [41]. Postać takiego rozkładu zależy od wielu czynników. Wpływają na nią wymiary oraz ukształtowanie nierówności, a także nominalny kształt powierzchni i sposób jej oświetlenia. Przybliżony przestrzenny kształt tego rozkładu można określić, wykorzystując prosty, geometryczny model rozpraszania światła [41].

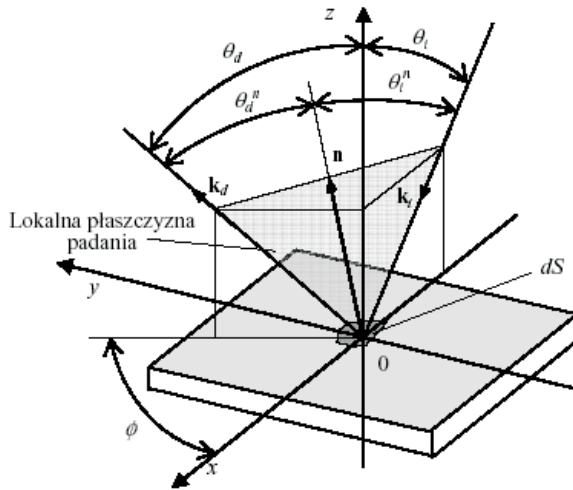
Niech płaska fala światła pada na zwierciadlane odbijający element dS powierzchni $z = f(x, y)$. Jednostkowy wektor normalnej \mathbf{n} jest prostopadły do elementu dS . Wektory falowe fali padającej \mathbf{k}_i i fali odbitej \mathbf{k}_d oraz wektor normalnej \mathbf{n} leżą w jednej płaszczyźnie, nazywanej lokalną płaszczyzną padania, co pokazano na rysunku 2. Długości rzutów n_x , n_y i n_z wektora normalnej \mathbf{n} , na osie kartezjańskiego układu współrzędnych $0xyz$, określone są zależnościami:

$$n_x = -n_z \operatorname{tg} \gamma_x, \quad (5)$$

$$n_y = -n_z \operatorname{tg} \gamma_y, \quad (6)$$

$$n_z = \frac{1}{\sqrt{1 + \operatorname{tg}^2 \gamma_x + \operatorname{tg}^2 \gamma_y}}, \quad (7)$$

gdzie: γ_x , γ_y – kąty między osiami $0x$ i $0y$ a stycznymi (w punkcie odbicia) do linii przecięcia powierzchni $z = f(x, y)$ płaszczyznami równoległymi do $0xz$ i $0yz$.



Rys. 2. Zwierciadlane odbicie światła od elementu powierzchni dS
 Fig. 2. Specular reflection of light from a surface element dS

Kąty γ_x , γ_y nachylenia stycznych do powierzchni $z = f(x, y)$ nie przekraczają zakresu od $-\pi/2$ do $\pi/2$ radiana. Można je wyznaczyć na podstawie pochodnych cząstkowych pierwszego rzędu funkcji $z = f(x, y)$, korzystając z równań

$$\gamma_x = \arctg \frac{\partial f(x, y)}{\partial x}, \quad (8)$$

$$\gamma_y = \arctg \frac{\partial f(x, y)}{\partial y}. \quad (9)$$

Wartości kątów γ_x i γ_y zależą od nominalnego kształtu powierzchni (np. powierzchnie płaskie, walcowe, stożkowe, sferyczne) oraz od kształtu i wymiarów nierówności.

Ten uproszczony model odbicia światła od zwierciadlanych elementów dS , tworzących powierzchnię $z = f(x, y)$, umożliwi w przybliżeniu określenie przestrzennego kształtu rozkładu natężenia światła rozproszonego. Na ogół w tym celu rozpatruje się położenie w przestrzeni zbioru wektorów falowych – określonych dla fal elementarnych, tworzących falę padającą i falę odbitą – tak jak uczyniono to w pracy [64]. Analiza taka, w zależności od przyjętego modelu nierówności, może mieć charakter deterministyczny lub statystyczny.

Kierunek światła odbitego, określony wektorem \mathbf{k}_d lub wersorem \mathbf{w}_d , zależy od kierunku światła padającego – określonego z kolei wektorem \mathbf{k}_i lub wersorem \mathbf{w}_i – oraz od kierunku normalnej \mathbf{n} . Zależność tę wyraża związek

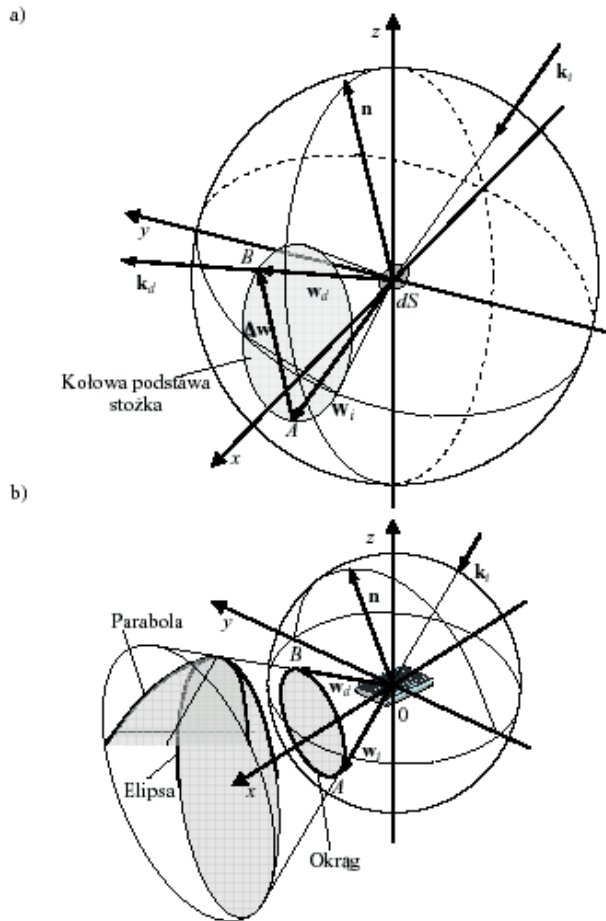
$$\mathbf{w}_d = \mathbf{w}_i + \Delta \mathbf{w} \quad (10)$$

przy czym

$$\Delta \mathbf{w} = -2(\mathbf{n} \cdot \mathbf{w}_i) \mathbf{n} = [-2|\mathbf{n}||\mathbf{w}_i| \cos(\pi - \theta_i^n)] \mathbf{n} \quad (11)$$

natomiast

$$|\Delta \mathbf{w}| = 2 \cos \theta_i^n \quad (12)$$



Rys. 3. Układ geometryczny rozpraszania: a) sfera do analizy geometrii rozpraszania, b) kształt rozkładu natężenia światła rozproszonego przez powierzchnię $z = f(y)$, w zależności od położenia płaszczyzny obserwacji

Fig. 3. Scattering geometry: a) sphere for scattering analysis, b) form of scattered light distribution obtained from $z = f(y)$ depending to position of observation plane

Wektor $\Delta \mathbf{w}$ jest różnicą wersorów \mathbf{w}_d oraz \mathbf{w}_i . Jest on równoległy do wektora \mathbf{n} . W przypadku niektórych powierzchni, np. powierzchni szlifowanych, można przyjąć,

że nierówności mają kształt rowków równoległych do jednej z osi układu współrzędnych, np. do osi $0x$. Opisane są one funkcją $z \approx f(y)$. Powierzchnia o takich nierównościach składa się z wielu elementów dS o wartościach kąta γ_x bliskich zeru, lecz o różnych wartościach kąta γ_y . Analizę rozpraszania światła przez takie powierzchnie przeprowadzono w pracy [41]. Pokazała ona, że ogólna postać rozkładu natężenia światła rozproszonego przez takie powierzchnie ma kształt fragmentu krzywej stożkowej. Zilustrowano to na rysunku 3.

Rodzaj krzywej zależy od kierunku wektora falowego \mathbf{k}_i fali padającej, a także od położenia płaszczyzny obserwacji względem stożka zakreślonego przez wektory \mathbf{k}_d . Najczęściej krzywą tą jest elipsa, okrąg lub parabola. W przypadku szczególnym, gdy wektor falowy \mathbf{k}_i leży w płaszczyźnie prostopadłej do kierunku nierówności (na rys. 3 jest to płaszczyzna $0yz$), rozkład natężenia światła rozproszonego przez powierzchnię $z \approx f(y)$ jest położony na odcinku prostej, będącej linią przecięcia płaszczyzny padania i płaszczyzny obserwacji. Wektor $\Delta\mathbf{w}$ jest wtedy położony w płaszczyźnie $0yz$ niezależnie od wartości kątów γ_y .

Dla powierzchni szlifowanych kąty γ_x zmieniają się w pewnym niewielkim zakresie. Z tego powodu pojawia się nieznaczące rozproszenie poprzeczne. Krzywe stożkowe opisujące rozkład natężenia światła rozproszonego ulegają pewnemu poprzecznemu rozmyciu. Będzie ono tym większe, im większy będzie zakres zmian kątów γ_x .

To, że rozkład natężenia światła rozproszonego jest tylko fragmentem krzywej stożkowej, wynika na ogół z ograniczonego zakresu zmiany kąta γ_y dla powierzchni rzeczywistych. Ale nawet wtedy, gdy kąt ten zmienia się w zakresie od $-\pi/2$ do $\pi/2$ radiana, rozkład natężenia światła rozproszonego przez powierzchnie nieprzezroczyste, nie może być obserwowany w tych punktach płaszczyzny obserwacji, w których współrzędna z jest ujemna. Powodem tego są odbicia wielokrotne, pojawiające się na pewno wtedy, gdy wartość bezwzględna kąta odbicia θ_d , mierzonego względem osi $0z$, przekroczy $\pi/2$ radiana.

2.2. DYFRAKCYJNE MODELE ROZPRASZANIA ŚWIATŁA

Optyka fizyczna traktuje światło jako fale elektromagnetyczne. Do matematycznego opisu takich fal wykorzystuje się wektorowe równania Maxwella [10], [54]. Uwzględnia to elektryczne i magnetyczne właściwości światła oraz jego polaryzację. W celu uproszczenia analizy rozpraszania światła przez powierzchnie nierówne analizuje się tylko zachowanie pola elektrycznego fal świetlnych. Często także pomija się polaryzacyjne właściwości tych fal. Rozpatruje się wtedy światło niespolaryzowane, albo przyjmuje się, że jest ono spolaryzowane liniowo, ograniczając analizę do jednej składowej wektora pola elektrycznego. Te założenia umożliwiają traktowanie światła jako fali skalarnej [10].

Przyjmując w dalszych rozważaniach podane założenia, dokonano analizy dyfrakcji płaskiej fali świetlnej na powierzchni nierównej. Celem tej analizy jest określenie

rozkładu zespolonej amplitudy lub natężenia światła rozproszonego przez powierzchnię nierówną w obszarze Fraunhofera.

2.2.1. TEORIA KIRCHHOFFA

Często wykorzystywanym podejściem do analizy rozpraszania fal na powierzchniach nierównych jest stosowanie teorii dyfrakcji Kirchhoffa. Teoria ta, opracowana przez G.R. Kirchhoffa do analizy dyfrakcji światła na otworach wykonanych w płaskich nieprzezroczystych ekranach, została następnie zaadaptowana do analizy pola dyfrakcyjnego fali odbitej od powierzchni chropowatych [11], [6], [3]. Stosowanie teorii Kirchhoffa opiera się na założeniu, że fala odbija się w każdym punkcie powierzchni tak samo jak odbijałaby się od nieskończonej płaszczyzny stycznej w danym punkcie. Dlatego niekiedy ten sposób analizy rozpraszania fali przez nierówności powierzchni nazywany jest w literaturze teorią Kirchhoffa, metodą Kirchhoffa lub metodą płaszczyzny stycznej [3], [51]. Jest on stosowany do analizy rozpraszania fal na stosunkowo dużych nierównościach, o wysokości większej od długości fali światła. W metodzie tej lokalne promienie krzywizny ρ w poszczególnych punktach powierzchni muszą spełniać warunek

$$\frac{1}{\sqrt[3]{k\rho}} \ll \cos\theta_i, \quad (13)$$

gdzie podobnie jak poprzednio k oznacza liczbę falową, a θ_i kąt padania światła na powierzchnię. Aby warunek (13) był spełniony, lokalne promienie krzywizny powinny być znacznie większe od długości fali światła λ .

Niech na jednorodną powierzchnię nierówną $z = f(x, y)$, której płaszczyzna średnia pokrywa się z płaszczyzną $z = 0$ kartezjańskiego układu współrzędnych, pada monochromatyczna fala płaska o długości λ . Niech kierunek padania określony jest w ogólnym przypadku wektorem \mathbf{r}_i . Załóżmy wstępnie, że kąt padania $\theta_i = 0$. Wiadomo, że w każdym punkcie – o współrzędnych danych wektorem \mathbf{r} – zespolona amplituda $E(\mathbf{r})$ monochromatycznej fali świetlnej, propagującej w ośrodku jednorodnym i izotropowym, spełnia równanie falowe Helmholtza [10]

$$\nabla^2 E(\mathbf{r}) + k^2 E(\mathbf{r}) = 0, \quad (14)$$

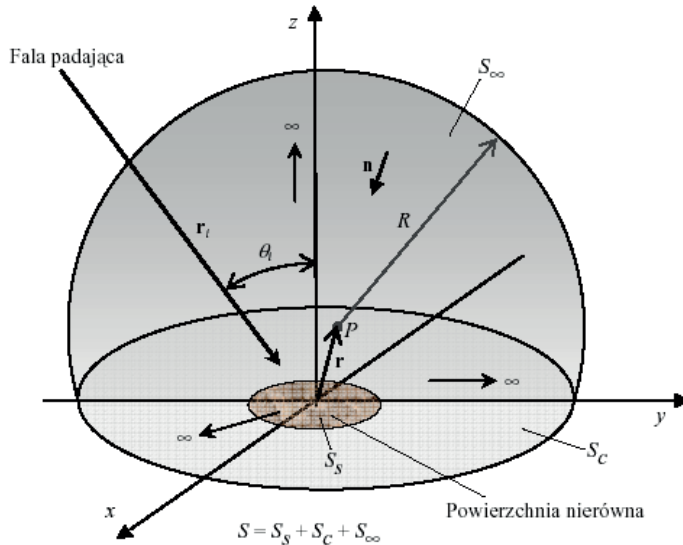
gdzie symbol ∇^2 oznacza operator Laplace'a, natomiast $k = 2\pi/\lambda$ jest liczbą falową. Równanie to jest spełnione także przez zespoloną amplitudę $E(\mathbf{r}_s)$ fali padającej, we wszystkich punktach powierzchni określonych wektorem \mathbf{r}_s . Jeśli znana jest zespolona amplituda $G(\mathbf{r}, \mathbf{r}_s)$ innej fali, spełniająca równanie Helmholtza, to korzystając ze wzoru Greena i stosując do obu fal równanie (14), otrzymuje się równanie (15), nazywane twierdzeniem całkowym Helmholtza–Kirchhoffa

$$E(\mathbf{r}) = \frac{1}{4\pi} \iint_S \left[E(\mathbf{r}_s) \frac{\partial G(\mathbf{r}, \mathbf{r}_s)}{\partial n} - G(\mathbf{r}, \mathbf{r}_s) \frac{\partial E(\mathbf{r}_s)}{\partial n} \right] dS, \quad (15)$$

gdzie: $E(\mathbf{r})$ – zespolona amplituda fali świetlnej w punkcie obserwacji P , którego położenie określa wektor \mathbf{r} , $E(\mathbf{r}_s)$ – zespolona amplituda fali światła na powierzchni S , w punktach określonym wektorem \mathbf{r}_s , $G(\mathbf{r}, \mathbf{r}_s)$ – funkcja Greena, S – powierzchnia całkowania złożona z powierzchni nierównej S_s , z części S_c płaszczyzny Oxy oraz z części sfery S_∞ , o środku w punkcie P i promieniu $R \rightarrow \infty$; ($S = S_s + S_c + S_\infty$). Pochodne cząstkowe w równaniu (15) brane są względem normalnej \mathbf{n} , skierowanej do wnętrza powierzchni (rys. 4). Funkcja Greena ma postać

$$G(\mathbf{r}, \mathbf{r}_s) = \frac{\exp(ik|\mathbf{r}, \mathbf{r}_s|)}{|\mathbf{r}, \mathbf{r}_s|}, \quad (16)$$

gdzie k jest liczbą falową, i – jednostką urojoną, natomiast $|\mathbf{r} - \mathbf{r}_s|$ oznacza długość wektora oddzielającego dwa punkty, z których jeden ma położenie określone wektorem \mathbf{r} , natomiast drugi znajduje się na powierzchni S . Funkcja ta opisuje falę kulistą, o amplitudzie równej jedności, emitowaną z punktu P , którego położenie określa wektor \mathbf{r} .



Rys. 4. Powierzchnia całkowania w równaniu (15)

Fig. 4. Surface of integration in equation (15)

Jeżeli zespolona amplituda $E(\mathbf{r}_s)$ fali świetlnej spełnia dwa warunki Sommerfelda, tzn. warunek skończoności

$$\lim_{|\mathbf{r}-\mathbf{r}_s| \rightarrow \infty} E(\mathbf{r}_s) = 0 \quad (17)$$

oraz warunek wypromieniowania

$$\lim_{|\mathbf{r}-\mathbf{r}_s| \rightarrow \infty} \left[\frac{\partial E(\mathbf{r}_s)}{\partial n} - ikE(\mathbf{r}_s) \right] |\mathbf{r}-\mathbf{r}_s| = 0 \quad (18)$$

to całkowanie po powierzchni S w równaniu (15) można zastąpić całkowaniem po powierzchni $S_s + S_c$, gdyż całka po powierzchni S_∞ jest równa zeru [25].

Rozwiązanie równań (14) i (15) wymaga wybrania odpowiednich warunków brzegowych. W metodzie Kirchhoffa przyjmuje się, że na powierzchni S_c spełnione są dwa warunki brzegowe, warunek Dirichleta

$$E(\mathbf{r}_s) = 0, \quad (19)$$

oraz warunek Neumanna

$$\frac{\partial E(\mathbf{r}_s)}{\partial n} = 0, \quad (20)$$

które łącznie nazywane są warunkami brzegowymi Kirchhoffa. Przyjmując warunki brzegowe Kirchhoffa, powierzchnię całkowania w równaniu (15) można ograniczyć do S_s , w wyniku czego otrzymuje się następującą zależność:

$$E(\mathbf{r}) = \frac{1}{4\pi} \iint_{S_s} \left[E(\mathbf{r}_s) \frac{\partial G(\mathbf{r}, \mathbf{r}_s)}{\partial n} - G(\mathbf{r}, \mathbf{r}_s) \frac{\partial E(\mathbf{r}_s)}{\partial n} \right] dS, \quad (21)$$

Ściśle rzecz biorąc, jednoczesne przyjęcie warunków brzegowych (19) i (20) jest niedopuszczalne, gdyż prowadzi do wniosku, że pole świetlne zanika także na całej powierzchni S_s , co jest sprzeczne z realnie istniejącą sytuacją. Na problem ten zwrócono uwagę w kilku pracach [25], [51]. Mimo tej sprzeczności teoria Kirchhoffa w wielu przypadkach daje wyniki zgodne z doświadczeniem.

Jak wykazano między innymi w pracach [25] i [60], przyjęcie w równaniu (21) funkcji Greena o postaci (16) i założenie, iż długość wektora \mathbf{r} jest znacznie większa od długości fali światła, prowadzi do otrzymania wzoru dyfrakcyjnego Fresnela–Kirchhoffa. Zakłada się przy tym, iż fala padająca jest falą kulistą. Jeżeli przyjąć, że na powierzchnię, z kierunku określonego wektorem \mathbf{r}_i , pada fala płaska o amplitudzie A , to zamiast wzoru Fresnela–Kirchhoffa, uzyskuje się zależność

$$E(\mathbf{r}) = \frac{1}{i\lambda} \iint_{S_s} E(\mathbf{r}_s) \frac{\exp(ik|\mathbf{r}-\mathbf{r}_s|)}{|\mathbf{r}-\mathbf{r}_s|} dS, \quad (22)$$

gdzie $E(\mathbf{r}_s)$ jest zespoloną amplitudą światła na powierzchni nierównej, której współczynnik odbicia jest równy jedności w każdym punkcie powierzchni

$$E(\mathbf{r}_s) = A \exp(ik|\mathbf{r}_i - \mathbf{r}_s|) \frac{\cos(\mathbf{n}, \mathbf{r} - \mathbf{r}_s) - \cos(\mathbf{n}, \mathbf{r}_i - \mathbf{r}_s)}{2}. \quad (23)$$

Czynnik $[\cos(\mathbf{n}, \mathbf{r} - \mathbf{r}_s) - \cos(\mathbf{n}, \mathbf{r}_i - \mathbf{r}_s)]/2$, występujący w wyrażeniu (23), nazywany jest współczynnikiem nachylenia. Jeżeli odległość punktu obserwacji P od powierzchni jest znacznie większa niż wymiary oświetlonego obszaru powierzchni, to – przy prostopadłym padaniu fali płaskiej na powierzchnię o niewielkim nachyleniu zboczy nierówności – można przyjąć, że współczynnik nachylenia jest bliski jedności. Stosowane jest przy tym tzw. przybliżenie przyosiowe. Polega ono na lokalizacji punktu obserwacji P w pobliżu kierunku odbicia zwierciadlanego. Przy prostopadłym padaniu na powierzchnię fali płaskiej, przybliżenie przyosiowe oznacza, że obserwacji światła rozproszonego dokonuje się w niewielkim obszarze pewnej płaszczyzny $z = \bar{r}$, w pobliżu punktu przecięcia tej płaszczyzny z osią Oz . W takim przypadku – dla trzech wyrażeń: $1/|\mathbf{r} - \mathbf{r}_s|$, $\cos(\mathbf{n}, \mathbf{r} - \mathbf{r}_s)$ oraz $\cos(\mathbf{n}, \mathbf{r}_i - \mathbf{r}_s)$, które występują w równaniach (22) i (23) – stosuje się następujące równości przybliżone

$$\frac{1}{|\mathbf{r} - \mathbf{r}_s|} \approx \frac{1}{\bar{r}}, \quad (24)$$

$$\cos(\mathbf{n}, \mathbf{r} - \mathbf{r}_s) \approx -\cos(\mathbf{n}, \mathbf{r}_i - \mathbf{r}_s) \approx 1, \quad (25)$$

gdzie \bar{r} jest stałą odległością płaszczyzny $z = 0$ od równoległej do niej płaszczyzny obserwacji, w której zlokalizowany jest punkt P . Przybliżenia (24) i (25) przyjmowane są dla wszystkich rozpatrywanych punktów powierzchni oraz punktów, w których analizowane jest pole dyfrakcyjne. Zależności (24) nie można niestety zastosować w odniesieniu do długości $|\mathbf{r} - \mathbf{r}_s|$, która występuje w szybko zmieniającym się czynniku fazowym równania (22).

Zgodnie z pracami [25], [60], [24] długość $|\mathbf{r} - \mathbf{r}_s|$, zawarta w wykładniku wyrażenia podcałkowego w równaniu (22), jest najczęściej określana z wykorzystaniem przybliżenia Fresnela

$$|\mathbf{r} - \mathbf{r}_s| \approx \bar{r} + \frac{x_p^2}{2\bar{r}} + \frac{y_p^2}{2\bar{r}} - \frac{x_s x_p}{\bar{r}} - \frac{y_s y_p}{\bar{r}} + \frac{x_s^2}{2\bar{r}} + \frac{y_s^2}{2\bar{r}}, \quad (26)$$

lub przybliżenia Fraunhofera

$$|\mathbf{r} - \mathbf{r}_s| \approx \bar{r} + \frac{x_p^2}{2\bar{r}} + \frac{y_p^2}{2\bar{r}} - \frac{x_s x_p}{\bar{r}} - \frac{y_s y_p}{\bar{r}}, \quad (27)$$

gdzie: \bar{r} – odległość płaszczyzny obserwacji od płaszczyzny $z = 0$, x_p, y_p – współrzędne punktu P w płaszczyźnie obserwacji, x_s, y_s – współrzędne punktu na po-

wierzchni nierównej. Podstawiając (27) do (22) i uwzględniając przybliżenie (24), po dokonaniu przekształceń, otrzymuje się:

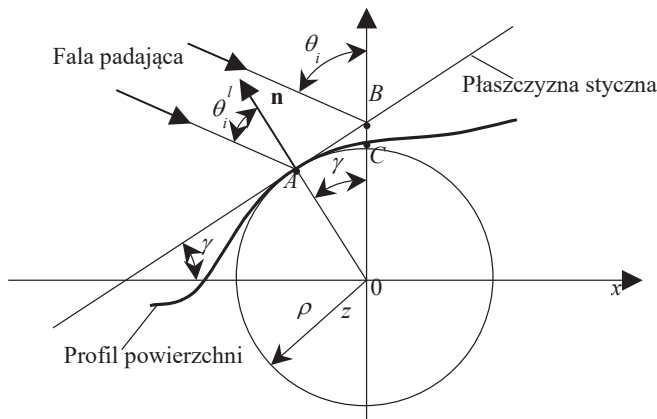
$$E(\mathbf{r}) = \frac{\exp\left[ik\bar{r} + \frac{ik}{2\bar{r}}(x_p^2 + y_p^2)\right]}{ik\bar{r}} \iint_{S_s} E(\mathbf{r}_s) \exp\left[\frac{-ik}{\bar{r}}(x_s x_p + y_s y_p)\right] dS. \quad (28)$$

W metodzie Kirchhoffa zakłada się, że powierzchnia rozpraszająca jest powierzchnią gładką [3], [51]. Założenie to wyraża zespoloną amplitudę światła $E(\mathbf{r}_s)$ w dowolnym punkcie powierzchni w postaci sumy amplitud fali padającej na powierzchnię oraz fali odbitej od płaszczyzny stycznej w danym punkcie. Wyrażenia dla zespolonej amplitudy światła i jej pochodnej mają postać

$$E(\mathbf{r}_s) = [1 - \rho_A(\mathbf{r}_s)] E_i(\mathbf{r}_s), \quad (29)$$

$$\frac{\partial E(\mathbf{r}_s)}{\partial n} = [1 - \rho_A(\mathbf{r}_s)] \frac{\partial E_i(\mathbf{r}_s)}{\partial n}, \quad (30)$$

gdzie $E_i(\mathbf{r}_s)$ oznacza zespoloną amplitudę światła padającego na powierzchnię, natomiast $\rho_A(\mathbf{r}_s)$ jest amplitudowym współczynnikiem odbicia. Współczynnik ten zależy od lokalnego kąta padania światła θ_i^l oraz optycznych właściwości materiału powierzchni. W ogólnym przypadku jest on wielkością zespoloną, przy czym dla fal wektorowych można go wyznaczyć, dla obu składowych wektora natężenia pola elektrycznego, stosując wzory Fresnela [6]. Takie określenie zespolonej amplitudy światła na powierzchni nierównej jest podobne do sposobu stosowanego w teorii dyfrakcji Kirchhoffa podczas analizy dyfrakcji światła na otworach wykonanych w płaskich, nieprzezroczystych ekranach.



Rys. 5. Konstrukcja płaszczyzny stycznej

Fig. 5. Construction of tangent plane

Odbicie światła w każdym punkcie powierzchni rozpatrywane jest w metodzie Kirchhoffa tak samo jak odbicie od nieskończonej płaszczyzny stycznej w tym punkcie, co pokazano na rysunku 5. Postępowanie takie jest poprawne pod warunkiem, że na płaszczyźnie stycznej da się określić obszar, którego wymiary liniowe są znacznie większe od długości fali światła λ , odległość zaś poszczególnych punktów tego obszaru od powierzchni nierównej jest niewielka. Warunki te wyrażają dwie nierówności

$$AB \gg \frac{\lambda}{2\pi \cos \theta_i^l}, \quad (31)$$

$$BC \ll \frac{\lambda \cos \theta_i^l}{2\pi}, \quad (32)$$

gdzie: λ – długość fali światła, θ_i^l – lokalny kąt padania, długości zaś odcinków AB oraz BC pokazano na rysunku 5. Przyjmując, że spełniona jest zależność

$$BC \approx AB\gamma \approx \frac{(AB)^2}{\rho} \quad (33)$$

co zachodzi wtedy, gdy pokazany na rysunku 5 kąt γ jest niewielki, po podstawieniu (33) do (32) i uwzględnieniu nierówności (23) otrzymuje się

$$k\rho \cos^3 \theta_i^l \gg 1, \quad (34)$$

gdzie $k = 2\pi/\lambda$ jest liczbą falową, ρ oznacza promień krzywizny powierzchni w danym punkcie, natomiast θ_i^l jest lokalnym kątem padania.

Warunek (34) określony jest dla lokalnego kąta padania θ_i^l . Jeśli powierzchnia nierówna charakteryzuje się niewielkimi wartościami kątów nachylenia zboczy nierówności γ , to lokalny kąt padania światła θ_i^l , w wyrażeniu (34), zastępuje się ogólnym kątem padania θ . Wtedy warunek (34) można przekształcić do postaci określonej wcześniej zależnością (13).

Jak pokazano w pracach [6], [3] i [51], ogólne wyrażenia dla zespolonej amplitudy światła rozproszonego przez powierzchnię nierówną w metodzie Kirchhoffa uzyskuje się przez podstawienie do równania (21) zależności (16), (29) i (30). Postać tych wyrażzeń zależy od wybranego układu odniesienia oraz od przyjętych założeń i zastosowanych aproksymacji. Jeśli założy się, że funkcja $z = f(x, y)$ i jej pierwsza pochodna są stacjonarnymi funkcjami losowymi z normalnym rozkładem rzędnych, to szczególnie można wykazać, iż w obszarze Fraunhofera stosunek statystycznie uśrednionej zespolonej amplitudy światła odbitego zwierciadlanie, do zespolonej amplitudy światła odbitego od płaszczyzny $z = 0$, jest proporcjonalny do funkcji charakterystycznej nierówności powierzchni

$$\frac{\bar{E}_s}{E_0} = \exp(-2k^2\sigma^2 \cos^2 \theta_i), \quad (35)$$

gdzie: \bar{E}_s – uśredniona zespolona amplituda światła odbitego od powierzchni nierównej w kierunku zwierciadlanym, E_0 – zespolona amplituda światła odbitego od płaszczyzny $z = 0$, σ – odchylenie standardowe wysokości nierówności powierzchni, k – liczba falowa, θ_i – kąt padania.

W pracy [3] w podobny sposób zastosowano metodę Kirchhoffa do analizy rozpraszania światła przez duże losowe nierówności powierzchni, znacznie większe od długości fali światła. Wykazano tam, że przestrzenny rozkład natężenia światła rozproszonego jest proporcjonalny do dwuwymiarowej funkcji gęstości prawdopodobieństwa pochodnej powierzchni

$$I(\theta_d, \phi, \theta_i) = |E(\theta_d, \phi, \theta_i)|^2 \propto (2\pi\pi^2 S_0 p(z')), \quad (36)$$

gdzie: $I(\theta_d, \phi, \theta_i)$ – natężenie światła rozproszonego przez powierzchnię w kierunku określonym kątami θ_d i ϕ przy padaniu światła na powierzchnię pod kątem θ_i , $E(\theta_d, \phi, \theta_i)$ – zespolona amplituda światła rozproszonego w tym samym kierunku, θ_d – biegunowy kąt rozproszenia, ϕ – azymutalny kąt rozproszenia, θ_i – kąt padania, S_0 – pole rzutu oświetlonego obszaru powierzchni na płaszczyznę $z = 0$, $p(z')$ – funkcja określająca gęstość prawdopodobieństwa pochodnej powierzchni.

2.2.2. TEORIA ZABURZEŃ

Równanie falowe Helmholtza (6) jest równaniem liniowym. Umożliwia to wykorzystanie zasady superpozycji i przedstawienie dowolnego rozwiązania tego równania w postaci sumy monochromatycznych fal płaskich o różnych amplitudach, fazach i kierunkach propagacji [24]. Jeżeli znany jest kształt nierówności i właściwości fizyczne powierzchni rozpraszającej, oświetlonej falą płaską, to można poszukiwać rozwiązania równania (14) w postaci kąтового widma fal płaskich, spełniającego warunki graniczne na powierzchni. Taką teorię rozpraszania fal skalarnych, padających prostopadłe na powierzchnię o zarysie sinusoidalnym, opracował Lord Rayleigh. Jej analizę przeprowadził P. Beckmann w pracy [6]. Teoria ta, rozwinięta potem przez S.O. Rice'a [56] dla powierzchni losowych, nazywana jest teorią zaburzeń (ang. *perturbation theory*), teorią Rayleigha lub teorią Rayleigha–Rice'a. Wektorową teorię zaburzeń zastosowali między innymi E.L. Church, H.A. Jenkinson i J.M. Zavada do analizy rozpraszania światła na powierzchniach optycznych [17], [18].

Istota teorii zaburzeń polega na poszukiwaniu fali rozproszonej w postaci takiej sumy fal płaskich, która spełnia warunki graniczne na powierzchni. Według tej teorii nierówności zaburzają falę odbitą od gładkiej powierzchni. W ogólnym przypadku pole fali rozproszonej jest tworzone przez nieskończoną sumę fal płaskich. Suma ta

jest szybko zbieżna tylko wtedy, gdy nierówności powierzchni są małe w porównaniu z długością fali światła λ . Teoria zaburzeń wymaga, aby funkcja $z = f(x, y)$, opisująca powierzchnię nierówną, spełniała następujące warunki:

$$\frac{2\pi}{\lambda} |f(x, y)| \cos \theta_i \ll 1, \quad (37)$$

$$|\nabla f(x, y)| \cos \theta_i \ll 1, \quad (38)$$

gdzie symbol ∇ (nabla) oznacza operator różniczkowy pierwszego rzędu, natomiast θ_i jest kątem padania światła na powierzchnię.

Najważniejsza idea teorii zaburzeń polega na rozwinięciu w szereg Taylora zespolonej amplitudy fali świetlnej $E(\mathbf{r}_s)$ na powierzchni rozpraszającej

$$E(\mathbf{r}_s) = E(\mathbf{r}_0) + f(x, y) \frac{\partial E(\mathbf{r}_0)}{\partial z} + \frac{f^2(x, y)}{2!} \frac{\partial^2 E(\mathbf{r}_0)}{\partial z^2} + \dots, \quad (39)$$

gdzie \mathbf{r}_s i \mathbf{r}_0 są wektorami określającymi odpowiednio położenie punktów na powierzchni rozpraszającej $z = f(x, y)$ oraz na płaszczyźnie $z = 0$, będącej jednocześnie płaszczyzną średnią powierzchni nierównej. Najczęściej stosuje się przybliżenie wykorzystujące dwa pierwsze składniki tego szeregu, stojące z prawej strony równania (39). Otrzymuje się wtedy

$$E(\mathbf{r}_s) \approx E(\mathbf{r}_0) + f(x, y) \frac{\partial E(\mathbf{r}_0)}{\partial z}. \quad (40)$$

Teoria zaburzeń wykorzystująca przybliżenie (40) nazywana jest, według [51], teorią zaburzeń pierwszego rzędu. Teoria zaburzeń drugiego rzędu jest nieco bardziej skomplikowana i wyraża zespoloną amplitudę fali świetlnej na powierzchni nierównej za pomocą trzech pierwszych składników występujących z prawej strony wyrażenia (39).

Aby uniknąć sprzeczności teorii Kirchhoffa, w metodzie zaburzeń przyjmuje się tylko jeden z warunków brzegowych Kirchhoffa (warunek Dirichleta lub warunek Neumanna) i dokonuje się odpowiedniego wyboru funkcji Greena. Jeśli zastosuje się przybliżenie (40), to warunek brzegowy Dirichleta (19) można zapisać w postaci

$$E(\mathbf{r}_s) \approx E(\mathbf{r}_0) + f(x, y) \frac{\partial E(\mathbf{r}_0)}{\partial z} = 0. \quad (41)$$

Funkcja Greena dobierana jest tak, że albo sama funkcja $G^*(\mathbf{r}, \mathbf{r}_0)$, albo jej pochodna $\partial G^*(\mathbf{r}, \mathbf{r}_0)/\partial n$ są równe zeru na płaszczyźnie $z = 0$. Przyjmując warunek brzegowy Dirichleta (41), funkcję Greena określa się najczęściej w postaci:

$$G^*(\mathbf{r}, \mathbf{r}_0) = \frac{\exp(ik|\mathbf{r} - \mathbf{r}_0|)}{|\mathbf{r} - \mathbf{r}_0|} - \frac{\exp(ik|\mathbf{r}^* - \mathbf{r}_0|)}{|\mathbf{r}^* - \mathbf{r}_0|}, \quad (42)$$

gdzie $G^*(\mathbf{r}, \mathbf{r}_0)$ jest funkcją Greena, alternatywną do funkcji danej wzorem (16), natomiast \mathbf{r}^* jest wektorem wyznaczającym położenie punktu P^* , będącego odbiciem zwierciadlanym punktu P względem płaszczyzny $z = 0$. Wektorowi \mathbf{r} , określającemu punkt P o współrzędnych x, y, z , odpowiada wektor \mathbf{r}^* , określający punkt P^* o współrzędnych $x, y, -z$. Funkcja Greena wyrażona równaniem (42) jest algebraiczną sumą dwóch fal kulistych wyemitowanych w przeciwnych fazach z dwóch punktów: z punktu P , określonego wektorem \mathbf{r} , oraz z punktu P^* , określonego wektorem \mathbf{r}^* . Przy wyborze warunku brzegowego Neumanna alternatywna funkcja Greena jest sumą dwóch składników, występujących z prawej strony równania (42). Funkcja ta opisuje dwie fale kuliste, o zgodnych fazach, wyemitowane z tych samych punktów co poprzednio, czyli z P i P^* .

Zależność określającą zespoloną amplitudę światła rozproszonego przez powierzchnię, w wybranym punkcie P , otrzymuje się w wyniku podstawienia zespolonej amplitudy światła na płaszczyźnie $z = 0$ do wzoru (15), który wyraża całkowite twierdzenia Helmholtza–Kirchhoffa. Zgodnie z teorią zaburzeń pierwszego rzędu, zespolona amplituda światła rozproszonego przez powierzchnię na płaszczyźnie $z = 0$ jest równa sumie dwóch składników

$$E(\mathbf{r}_0) \approx E_0(\mathbf{r}_0) + E_1(\mathbf{r}_0), \quad (43)$$

natomiast składnik zerowego rzędu $E_0(\mathbf{r}_0)$, nazywany polem koherentnym, jest równy

$$E_0(\mathbf{r}_0) = -\rho_A E_i(\mathbf{r}_0), \quad (44)$$

gdzie: ρ_A – amplitudowy współczynnik odbicia, $E_i(\mathbf{r}_0)$ – zespolona amplituda światła padającego w płaszczyźnie $z = 0$. Składnik zerowego rzędu $E_0(\mathbf{r}_0)$, dany wyrażeniem (44), jest rozwiązaniem problemu rozpraszania światła przez idealnie gładką płaszczyznę dla warunku brzegowego Dirichleta. Natomiast składnik pierwszego rzędu $E_1(\mathbf{r}_0)$, występujący w równaniu (43) – nazywany często polem niekoherentnym lub dyfuzyjnym – zaburza pole w płaszczyźnie $z = 0$. Uwzględniając zależności (29), (30) i (41), może on być określony następująco:

$$E_1(\mathbf{r}_0) = f(x, y) \left[\rho_A \frac{\partial E_i(\mathbf{r}_0)}{\partial z} - \frac{\partial E_i(\mathbf{r}_0)}{\partial z} \right], \quad (45)$$

gdzie $f(x, y)$ jest funkcją opisującą nierówności powierzchni.

Zgodnie z tym można przyjąć, że zespolona amplituda światła w punkcie P , określonym wektorem \mathbf{r} , jest sumą zespolonej amplitudy światła odbitego od gładkiej płaszczyzny $z = 0$ oraz zespolonej amplitudy światła zaburzonego przez nierówności powierzchni. Uwzględniając równanie (15) zależność (45) przyjmie postać

$$E_1(\mathbf{r}) = \frac{1}{4\pi} \iint_{S_0} \left\{ f(x, y) \left[\rho_A \frac{\partial E_i(\mathbf{r}_0)}{\partial z} - \frac{\partial E_i(\mathbf{r}_0)}{\partial z} \right] \right\} \frac{\partial G^*(\mathbf{r}, \mathbf{r}_0)}{\partial z} dS_0, \quad (46)$$

gdzie: $f(x, y)$ – funkcja opisująca nierówności powierzchni, ρ_A – amplitudowy współczynnik odbicia, $E_i(\mathbf{r}_0)$ – zespolona amplituda światła padającego w płaszczyźnie $z = 0$, $G^*(\mathbf{r}, \mathbf{r}_0)$ – alternatywna funkcja Greena, S_0 – pole rzutu oświetlonego obszaru powierzchni na płaszczyznę $z = 0$. W równaniach (45) i (46), w odróżnieniu od równań (15) i (30), pochodne cząstkowe mogą być brane po współrzędnej z , gdyż dla płaszczyzny $z = 0$ kierunek i zwrot wektora normalnej \mathbf{n} pokrywają się z kierunkiem i zwrotem osi Oz .

Uśrednione w czasie natężenie światła rozproszonego $I_1(\mathbf{r})$ w punkcie P , według teorii zaburzeń pierwszego rzędu, jest iloczynem zespolonej amplitudy światła, wyrażonej równaniem (38), oraz amplitudy z nią sprzężonej

$$\langle I_1(\mathbf{r}) \rangle = E_1(\mathbf{r}) E_1^*(\mathbf{r}), \quad (47)$$

gdzie $E_1^*(\mathbf{r})$ jest zespoloną amplitudą światła, sprzężoną z amplitudą $E_1(\mathbf{r})$, natomiast nawiasy $\langle \rangle$ oznaczają uśrednienie w czasie. Jak wynika z pracy J.A. Ogilvy [51], natężenie światła rozproszonego $I_1(\mathbf{r})$ określone jest wyrażeniem

$$\langle I_1(\mathbf{r}) \rangle = \frac{16 \pi^2 S_0 \cos^2(\theta_i) \cos^2(\theta_d)}{|\mathbf{r}|^2 \lambda^4} S(f_x, f_y), \quad (48)$$

przy czym

$$f_x = \frac{\sin(\theta_d) \cos \phi - \sin \theta_i}{\lambda}, \quad (49)$$

oraz

$$f_y = \frac{\sin(\theta_d) \sin \phi}{\lambda}, \quad (50)$$

gdzie: $\langle I_1(\mathbf{r}) \rangle$ – uśrednione w czasie względne natężenie światła rozproszonego w punkcie P , określonym wektorem \mathbf{r} , opisujące pole niekoherentne, S_0 – pole rzutu oświetlonego obszaru powierzchni na płaszczyznę $z = 0$, θ_i – kąt padania, θ_d – kąt rozproszenia, λ – długość fali światła, $S(f_x, f_y)$ – dwuwymiarowa funkcja gęstości widmowej mocy nierówności powierzchni, f_x, f_y – częstości przestrzenne nierówności powierzchni wyznaczone w kierunkach osi Ox i Oy , ϕ – azymutalny kąt rozproszenia. Dla ścisłości należy zaznaczyć, że azymutalny kąt rozproszenia ϕ , wyznaczany w płaszczyźnie Oxy , przyjmuje wartości dodatnie, gdy odmierzany jest od dodatniej części osi Ox w kierunku dodatniej części osi Oy .

Dotychczasowe rozważania przeprowadzono, przyjmując, że zespolona amplituda światła na powierzchni nierównej spełnia warunek brzegowy Dirichleta (19). Założenie, iż na powierzchni nierównej pole świetlne spełnia zamiast warunku Dirichleta warunek brzegowy Neumanna (20), umożliwi przeprowadzenie podobnej analizy [51]. Mimo iż jest ona nieco bardziej skomplikowana, jej wynikiem jest zależność podobna do równania (48).

Przedstawiona teoria zaburzeń, mimo wielu zalet, ma istotny niedostatek. Nie spełnia ona zasady zachowania energii. Wynika to z faktu, że składowa koherentna pola rozproszonego, dana wyrażeniem (44), nie zależy od nierówności powierzchni, podczas gdy pole dyfuzyjne jest od nierówności powierzchni zależne, co wynika z równania (45). Dlatego też, jak zaznaczono to na początku rozważań, stosowanie teorii zaburzeń może być uzasadnione jedynie dla powierzchni, których gradient wysokości nierówności jest niewielki, sama zaś wysokość nierówności jest mała w porównaniu z długością fali światła.

2.3. ROZPRASZANIE ŚWIATŁA PRZEZ POWIERZCHNIE OKRESOWE

Ujmując rzecz historycznie, rozpraszanie światła na powierzchniach okresowych było rozpatrywane wcześniej niż rozpraszanie na powierzchniach nieokresowych. Stało się tak zapewne ze względu na to, że analiza rozpraszania światła na powierzchniach okresowych jest prostsza niż na powierzchniach nieokresowych. Nie wymaga ona bowiem zastosowania skomplikowanych metod rachunku prawdopodobieństwa i statystyki matematycznej. Dogodnym sposobem analizy rozpraszania światła, szczególnie przez powierzchnie okresowe, jest teoria dyfrakcji światła.

Jak już o tym wspomniano, analiza rozpraszania fal przez powierzchnie nierówne dokonana została po raz pierwszy przez lorda Rayleigha. Rozpatrywał on rozpraszanie fal akustycznych, padających prostopadle na powierzchnię sinusoidalną. Metoda zastosowana przez Rayleigha oparta była na podobnych założeniach jak teoria zaburzeń. Została ona następnie rozwinięta, tak by można ją było stosować do rozpraszania fal elektromagnetycznych. Dokonano także uogólnienia tej metody w taki sposób, aby obejmowała ona przypadek powierzchni nieokresowych i przypadek nieprostopadłego padania fali na powierzchnię nierówną. Do badania rozpraszania fal przez powierzchnie okresowe stosowano także metodę Kirchhoffa [11]. Ogólne rozwiązanie otrzymane tą metodą daje się łatwo adaptować dla powierzchni okresowych.

Mimo że w praktyce powierzchnie z okresowymi nierównościami są spotykane znacznie rzadziej niż powierzchnie z nierównościami nieokresowymi lub losowymi, analiza rozpraszania światła przez powierzchnie okresowe jest ważna zarówno teoretycznie, jak i praktycznie. Znaczenie powierzchni okresowych w teorii rozpraszania światła wynika z tego, że analiza rozpraszania światła na niektórych prostych powierzchniach okresowych pozwala przewidzieć rezultaty rozpraszania na powierzchniach o bardziej skomplikowanej geometrii nierówności. Jak wiadomo, powierzchnie

okresowe o profilach sinusoidalnym i cosinusoidalnym są powierzchniami szczególnymi z tego względu, że dowolną powierzchnię nierówną, korzystając z szeregu Fouriera lub całki Fouriera, można traktować jako superpozycję zbioru ortogonalnych funkcji harmoniczych, z których każda ma odpowiednią amplitudę, przestrzenną długość fali i fazę. Wyniki analizy rozpraszania światła przez powierzchnie o profilu sinusoidalnym mogą być zatem uogólnione, dzięki zasadzie superpozycji, na powierzchnie o bardziej złożonym kształcie nierówności. Analiza dyfrakcji światła przez struktury i powierzchnie okresowe ma również duże znaczenie praktyczne, zwłaszcza w tych dziedzinach, gdzie stosowane są np. odbiciowe, fazowe siatki dyfrakcyjne, hologramy, wzorce inkrementalne itp.

Przykładem powierzchni okresowej, niezależnej od y , dla której zachodzi równość $z = f(x + \Lambda) = f(x)$, może być fazowa, odbiciowa siatka dyfrakcyjna

$$z = f(x, y) = a \sin\left(\frac{2\pi}{\Lambda}x + \alpha\right), \quad (51)$$

gdzie a jest amplitudą, α fazą nierówności, wielkość zaś Λ nazywana jest stałą siatki dyfrakcyjnej, okresem lub przestrzenną długością fali sinusoidalnej. Niech płaska fala światła, o długości λ , pada pod kątem θ_i na powierzchnię okresową $z = f(x)$, np. opisaną wzorem (51), a płaszczyzną padania niech będzie płaszczyzna $0xz$. Kątowe położenie maksimum dyfrakcyjnych określa równanie

$$\sin \theta_q = -\sin \theta_i + q \frac{\lambda}{\Lambda}, \quad \text{dla } q = 0, \pm 1, \pm 2, \dots, \quad (52)$$

gdzie θ_q oznacza kąt ugięcia światła, odpowiadający danemu maksimum, natomiast liczba q nazywana jest rzędem interferencji lub rzędem dyfrakcyjnym.

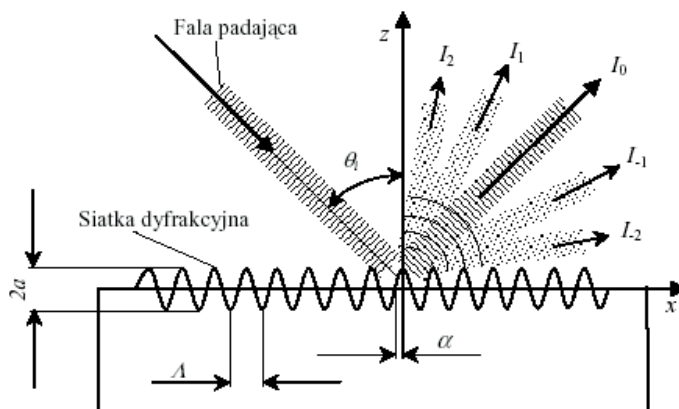
Równanie (52) nosi nazwę równania siatki dyfrakcyjnej. Otrzymuje się je z zależności (1), stosując ją do funkcji okresowej. Rozpatruje się przy tym dwa punkty powierzchni o jednakowych wartościach współrzędnej z , oddalone od siebie o Λ . Dla takich punktów różnice współrzędnych wynoszą odpowiednio: $z_1 - z_2 = 0$ i $x_2 - x_1 = \Lambda$. Zgodnie z równaniem (1) różnica drogi optycznej Δr , wyznaczona dla fal odbitych od powierzchni w tych punktach, jest równa

$$\Delta r = \Lambda(\sin \theta_d + \sin \theta_i), \quad (53)$$

gdzie symbolami θ_d i θ_i oznaczono odpowiednio kąt rozproszenia i kąt padania, Λ zaś oznacza stałą siatki dyfrakcyjnej. Maksima dyfrakcyjne powstaną wtedy, gdy różnica faz, wynikająca z różnicy drogi optycznej, będzie równa całkowitej wielokrotności 2π rad. Oznacza to konieczność spełnienia warunku

$$\frac{2\pi}{\lambda} \Lambda(\sin \theta_d + \sin \theta_i) = 2\pi q. \quad (54)$$

Po przekształceniu warunku (54), zakładając, iż $\theta_d = \theta_q$, otrzymuje się równanie (52). Kątowe położenie maksimów dyfrakcyjnych pokazano na rysunku 6.



Rys. 6. Dyfrakcja na siatce sinusoidalnej
Fig. 6. Diffraction from a sinusoidal grating

W pracy [10] rozpatrzono dyfrakcję płaskiej fali światła monochromatycznego na siatce dyfrakcyjnej, zawierającej N okresowych nierówności o dowolnym profilu. Natężenie światła ugiętego przez taką siatkę, w wybranym kierunku, jest iloczynem dwóch funkcji: funkcji interferencyjnej $H(p)$ i funkcji intensywności $I^0(p)$. Funkcja interferencyjna $H(p)$ jest funkcją okresową. Dla siatki odbiciowej zależy ona od sumy algebraicznej sinusów kąta rozproszenia θ_d i kąta padania θ_i . Znaki kątów θ_d i θ_i podano na rysunku 1. Funkcja interferencyjna zawiera wyraźne maksima, których natężenie jest równe N^2 , przy czym N jest liczbą okresów siatki w oświetlonym obszarze. Maksima te występują wtedy, gdy spełniony jest warunek (54). Funkcja intensywności $I^0(p)$ zależy od współczynnika odbicia światła i profilu powierzchni pojedynczej nierówności, fala świetlna bowiem odbijana w poszczególnych punktach powierzchni doznaje zmian amplitudy i fazy. Funkcja ta jest proporcjonalna do kwadratu modułu transformaty Fouriera funkcji odbicia, określonej w [10]. Mnożenie funkcji interferencyjnej $H(p)$ przez funkcję intensywności $I^0(p)$ powoduje zmianę natężenia kolejnych maksimów interferencyjnych. Rozkład natężenia światła ugiętego przez powierzchnię okresową jest rezultatem modulacji amplitudowej funkcji interferencyjnej $H(p)$ przez funkcję intensywności $I^0(p)$.

Korzystając z teorii Kirchhoffa, w pracach [6], [60], [25] pokazano, że w odbiciu płaskiej fali światła, której amplituda równa się jedności, od powierzchni fazowej sinusoidalnej siatki dyfrakcyjnej, względne natężenie światła $I(\mathbf{r})$ w punkcie P określonym wektorem \mathbf{r} , jest proporcjonalne do sumy kwadratów funkcji Bessela pierwszego rodzaju rzędu q

$$I(\mathbf{r}) \propto \sum_{q=-\infty}^{q=\infty} \left[J_q \left(\frac{\Delta\varphi_{\max}}{2} \right) \right]^2, \quad (55)$$

przy czym

$$\Delta\varphi_{\max} = \frac{4\pi}{\lambda} a (\cos\theta_i + \cos\theta_d), \quad (56)$$

gdzie: $J_q(\cdot)$ – funkcja Bessela pierwszego rodzaju rzędu q , $\Delta\varphi_{\max}$ – maksymalna różnica faz fali świetlnej, wprowadzana przez siatkę dyfrakcyjną, λ – długość fali światła, a – amplituda sinusoidalnej siatki dyfrakcyjnej, θ_i – kąt padania, θ_d – kąt rozproszenia. Przy prostopadłym padaniu płaskiej fali światła na powierzchnię, w sytuacji gdy amplituda sinusoidalnej siatki dyfrakcyjnej a jest znacznie mniejsza od długości fali światła λ , maksymalne względne natężenie światła $I(\theta_q)$ w q -tym rzędzie dyfrakcji jest w przybliżeniu równe funkcji Bessela pierwszego rodzaju rzędu q

$$I(\theta_q) \approx \left[J_q \left(\frac{4\pi a}{\lambda} \right) \right]^2, \quad (57)$$

Zakłada się przy tym, że kąty rozproszenia $\theta_i = \theta_q$, określające położenie kolejnych maksimum dyfrakcyjnych, są bliskie zeru. Dla większych wartości amplitudy a siatki sinusoidalnej wyznaczenie natężenia światła w q -tym maksimum dyfrakcyjnym staje się bardziej skomplikowane, gdyż wymaga sumowania funkcji Bessela odpowiadających kolejnym rzędom dyfrakcyjnym [25]. W przypadku gdy nierówności powierzchni opisane są inną funkcją okresową niż funkcja dana równaniem (51), pole dyfrakcyjne można wyznaczyć jako superpozycję fal ugiętych na sinusoidalnych i cosinusoidalnych siatkach dyfrakcyjnych, stanowiących rozwinięcie powierzchni nierównej w szereg Fouriera.

Do podobnych wniosków prowadzi analiza odbicia światła od powierzchni okresowych z wykorzystaniem teorii Rayleigha, przedstawiona w pracach [6], [51], [60]. Jeśli fala płaska pada na powierzchnię sinusoidalną, daną równaniem (51), której amplituda a jest znacznie mniejsza od długości fali λ , faza zaś $\alpha = \pi/2$, to dla światła ugiętego w rzędzie q otrzymuje się zależność

$$J_l \left(\frac{2\pi a}{\lambda} \cos\theta_i \right) = \sum_{q=-\infty}^{q=\infty} (-i)^q \sqrt{I(\theta_q)} J_{q+l} \left(\frac{2\pi a}{\lambda} \cos\theta_q \right), \quad (58)$$

gdzie: $J_m(\cdot)$ – funkcja Bessela pierwszego rodzaju rzędu m ; q, l – liczby całkowite, i – jednostka urojona, $I(\theta_q)$ – maksymalne względne natężenie światła w q -tym rzędzie dyfrakcji, θ_i – kąt padania, θ_q – kąt ugięcia dla q -tego rzędu.

Równanie (58) wynika z rozwinięcia w szereg Fouriera dwóch funkcji, które opisują falę padającą na powierzchnię i falę rozproszoną oraz z porównania odpowiednich

współczynników Fouriera w obu rozwinięciach. Szereg z prawej strony równania (58) jest szybkozbieżny, jeżeli $2\pi a \ll \lambda$. W takim przypadku można go ograniczyć do kilku składników. Umożliwia to utworzenie układu równań liniowych, z którego następnie wyznacza się względne natężenia światła $I(\theta_q)$ w kilku rzędach dyfrakcyjnych. Przedstawiona metoda analizy rozpraszania światła przez powierzchnie okresowe nazywana jest niekiedy metodą Rayleigha–Fouriera. Odbicie fal elektromagnetycznych od powierzchni okresowych rozpatrzono także w pracy [6], analizując to zjawisko kilkoma sposobami.

W pracy [60], korzystając z wektorowej teorii zaburzeń, wyprowadzono równania określające względne natężenia światła $I(\theta_{\pm 1})$ w maksimach dyfrakcyjnych ± 1 rzędu, przy odbiciu fali płaskiej od sinusoidalnej siatki dyfrakcyjnej. Analizy dokonano dla dwóch składowych światła spolaryzowanego: składowej położonej w płaszczyźnie padania oraz składowej leżącej w płaszczyźnie prostopadłej do płaszczyzny padania. Przy niewielkich kątach padania θ_i – dla siatki, której amplituda a jest mała w porównaniu z długością fali światła λ – oba równania przyjmują wspólną, przedstawioną postać:

$$I(\theta_{\pm 1}) \approx \left(\frac{2\pi a}{\lambda} \cos \theta_i \right)^2. \quad (59)$$

Teoria zaburzeń i teoria Kirchhoffa dają podobne rezultaty w przypadku, gdy amplituda nierówności jest mała w porównaniu z długością fali światła. Świadczy o tym porównanie zależności (57) i (59), jeśli uwzględni się, iż w równaniu (57) przyjęto $\theta = 0$ oraz że dla małych wartości argumentu funkcje Bessela pierwszego rodzaju $J_{\pm 1}(\cdot)$ aproksymuje się zależnościami

$$J_1(2ka) \approx \frac{2ka}{2} = ka, \quad (60)$$

$$J_{-1}(2ka) \approx -\frac{2ka}{2} = -ka, \quad (61)$$

gdzie $k = 2\pi/\lambda$ jest liczbą falową. Zaletą teorii zaburzeń jest to, że nie nakłada ona ograniczeń na przestrzenną długość fali nierówności powierzchni A , podczas gdy teoria Kirchhoffa wymaga, aby $A \gg \lambda$. Teoria zaburzeń może być jednak stosowana jedynie do nierówności o małych wysokościach. Natomiast teorię Kirchhoffa można stosować dla dowolnych wysokości nierówności. W tym przypadku ograniczenia nakładane są na przestrzenną długość fali lub na promienie krzywizny nierówności. W pracach [51] i [60] porównano obie teorie ze względu na ich dokładności.

2.4. ANALIZA DYFRAKCJI W SKATEROMETRII NIERÓWNOŚCI POWIERZCHNI

Mimo prostoty dyfrakcyjnego modelu rozpraszania światła przez powierzchnie okresowe, jego wykorzystanie do oceny powierzchni rzeczywistych nie jest łatwe

i wymaga niekiedy zastosowania specjalnych technik, takich np. jak sieci neuronowe [33]. W podstawowej postaci można go zastosować do oceny parametrów nierówności powierzchni jedynie w kilku prostych przypadkach. Występuje to np. wtedy, gdy powierzchnia opisana jest funkcją daną równaniem (51). W ogólnym przypadku powierzchni okresowych, z równania (52) otrzymuje się tylko wyrażenie dla przestrzennej długości fali Λ

$$\Lambda = \frac{\lambda q}{(\sin \theta_q + \sin \theta_i)}, \quad (62)$$

gdzie λ jest długością fali światła, q rzędem dyfrakcyjnym, natomiast θ_q i θ_i oznaczają kolejno kąt padania i kąt rozproszenia, odpowiadający q -temu maksimum dyfrakcyjnemu. Znając kąt padania θ_i oraz długość fali światła λ , można na podstawie równania (62) i wartości kąta rozproszenia θ_q w wybranym maksimum dyfrakcyjnym wyznaczyć przestrzenną długość fali Λ nierówności okresowych.

Jeśli rozpatrując równanie (62) przyjmie się, że kąt padania θ_i jest równy zeru i uwzględni, iż dla małych wartości kątów rozproszenia spełnione są równości

$$\sin \theta_q = \frac{\lambda q}{\Lambda} \approx \theta_q, \quad (63)$$

gdzie θ_q jest kątem rozproszenia określającym położenie q -tego maksimum dyfrakcyjnego. Równanie (62), przy podanych założeniach przyjmuje postać:

$$\Lambda \approx \frac{\lambda q}{\theta_q}, \quad (64)$$

Zastosowanie do równania (64) prawa sumowania błędów oraz zasady najniekorzystniejszego przypadku, umożliwi otrzymanie następującego wyrażenia dla błędu bezwzględnego $\Delta\Lambda$ przestrzennej długości fali Λ

$$\Delta\Lambda \approx \pm \left(\left| \frac{\Lambda}{\lambda} \Delta\lambda \right| + \left| -\frac{\Lambda^2}{\lambda q} \Delta\theta_q \right| \right), \quad (65)$$

gdzie: $\Delta\Lambda$ – błąd bezwzględny przestrzennej długości fali Λ , Λ – przestrzenna długość fali, λ – długość fali światła, q – rząd dyfrakcyjny, $\Delta\theta_q$ – błąd pomiaru kąta rozproszenia θ_q , odpowiadającego q -temu maksimum dyfrakcyjnemu. Błąd względny pomiaru przestrzennej długości fali wynosi:

$$\frac{\Delta\Lambda}{\Lambda} \approx \pm \left(\left| \frac{\Delta\lambda}{\lambda} \right| + \left| -\frac{\Lambda}{\lambda q} \Delta\theta_q \right| \right) \quad (66)$$

przy czym w równaniach (65) i (66) zastosowano takie same oznaczenia. Z równań (65) i (66) wynika, że wartości błędu bezwzględnego $\Delta\Lambda$ i względnego $\Delta\Lambda/\Lambda$ rosną

wraz ze wzrostem przestrzennej długości fali λ . Jeśli idzie o metrologię mikronierówności powierzchni jest to sytuacja dosyć niezwykła i szczególnie korzystna. Oznacza ona, że wykorzystując zjawisko dyfrakcji światła można dokładniej mierzyć nierówności, z krótszymi przestrzennymi długościami fali λ . Pomiarów kąta rozproszenia θ_i lepiej jest dokonywać w wyższych rzędach dyfrakcyjnych, gdyż wtedy błędy pomiaru przestrzennej długości fali λ będą mniejsze.

W przypadku powierzchni sinusoidalnej dyfrakcyjny model rozpraszania światła umożliwia także wyznaczenie amplitudy nierówności powierzchni. Dokonuje się tego na podstawie pomiaru względnego natężenia światła w jednym z maksimów dyfrakcyjnych ± 1 rzędu. W tym celu można skorzystać z zależności (59), która po przekształceniu przyjmuje postać

$$a \approx \frac{\lambda}{2\pi \cos \theta_i} \sqrt{I(\theta_{\pm 1})}, \quad (67)$$

przy czym zastosowano tu takie same oznaczenia jak poprzednio. Zależność (67) spełniona jest, jeśli założymy, iż amplituda a powierzchni sinusoidalnej jest znacznie mniejsza od długości fali światła λ .

Analizując efekty dyfrakcyjne w układach pomiarowych, przeznaczonych do oceny nierówności powierzchni, przyjmuje się na ogół, iż powierzchnia oświetlana jest światłem spójnym o znanym przestrzennym rozkładzie fazy. Wykorzystanie do tego celu światła spójnego wynika najczęściej ze świadomości, iż przy odbiciu od powierzchni nierównej fala świetlna jest przede wszystkim modulowana fazowo. Zastosowanie światła spójnego gwarantuje zatem, że informacja o fazie fali świetlnej nie zostanie utracona, chociaż odzyskanie i zinterpretowanie tej informacji może nie być łatwe. W pewnych sytuacjach zastosowanie światła spójnego okazuje się jednak niepożądane, ze względu na niekorzystne efekty, takie np. jak szумы koherentne powstające w wyniku zjawiska plamkowania. Szумы te, co prawda tylko niekiedy, można tłumić. Komplikuje to jednak układ pomiarowy oraz utrudnia odzyskanie informacji o fazie fali świetlnej odbitej od powierzchni nierównej.

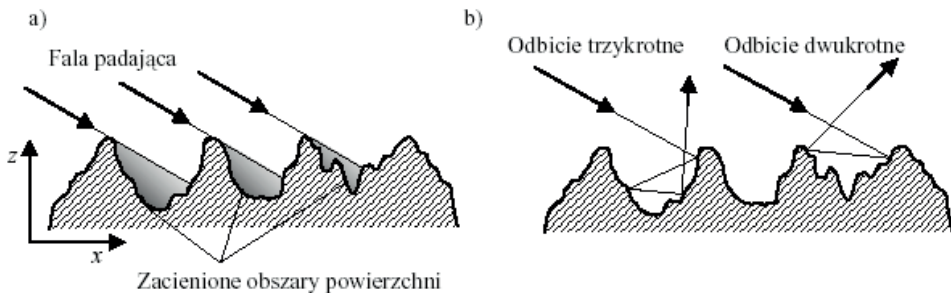
W przypadku stosowania do oświetlenia powierzchni mierzonej światła niespójnego, szczegółowa informacja zawarta w fazie odbitej fali świetlnej jest najczęściej bezpowrotnie tracona, ponieważ w procesie detekcji światła następuje proces uśrednienia fazy. Światło niespójne może być jednak stosowane wtedy, gdy przedmiotem zainteresowania jest uśredniona informacja o nierównościach powierzchni mierzonej. W takiej sytuacji analiza zjawiska dyfrakcji powinna uwzględniać zastosowanie oświetlenia niespójnego.

W dotychczasowych rozważaniach milcząco przyjmowano założenie, że padająca na powierzchnię nierówną fala świetlna dociera bez przeszkód do wszystkich punktów powierzchni. Ponadto zakładano też, że fala odbita od powierzchni również bez przeszkód dociera do punktu obserwacji. Jednakże w rzeczywistości wymienione założe-

nia nie zawsze udaje się spełnić, odbiciu światła od powierzchni nierównej mogą bowiem towarzyszyć różnorodne zjawiska, takie np. jak samozacienianie i samoprzesłanianie powierzchni oraz odbicia wielokrotne. Zjawiska te trudno jest uwzględnić w przedstawionym modelu dyfrakcyjnym. Ich analiza zostanie przedstawiona w kolejnym punkcie tego rozdziału.

2.5. ZJAWISKA TOWARZYSZĄCE ROZPRASZANIU ŚWIATŁA PRZEZ POWIERZCHNIE NIERÓWNE

Niekiedy, zwłaszcza przy dużych kątach padania światła na powierzchnię nierówną, ukształtowanie nierówności powierzchni może sprzyjać pojawianiu się pewnych dodatkowych zjawisk towarzyszących rozpraszaniu światła. Do zjawisk tych należą tak zwane samozacienianie powierzchni oraz odbicia wielokrotne. Samozacienianie, czyli zacienianie niektórych obszarów powierzchni przez nierówności, występuje przy dużym gradientzie funkcji opisującej powierzchnię, szczególnie wtedy, gdy kąty padania światła na powierzchnię są duże. Te same czynniki, tzn. mocno rozwinięta powierzchnia i duże wartości kąta padania, sprzyjają powstawaniu odbić wielokrotnych. Istotę zjawisk samozacieniania powierzchni i odbić wielokrotnych zilustrowano na rysunku 7.



Rys. 7. Zjawiska towarzyszące rozpraszaniu światła przez powierzchnię:

a) samozacienianie powierzchni, b) odbicia wielokrotne

Fig. 7. Attendant phenomena in light scattering from surface:

a) surface self-shadowing, b) multiple reflections

2.5.1. SAMOZACIENIANIE POWIERZCHNI

Teorie opisujące rozpraszanie światła przez powierzchnie nierówne zakładają na ogół, że padająca na powierzchnię fala oświetla bezpośrednio każdy punkt powierzchni. Jednakże założenie to nie zawsze jest spełnione. Jeśli na powierzchnię nierówną $z = f(x)$, pod kątem θ_i w płaszczyźnie Oxz , pada płaska fala światła, to – w przybliżeniu promieniowym – samozacienienie powierzchni wystąpi wtedy, gdy dla któregośkol-

wiek z punktów należących do oświetlanego obszaru powierzchni spełniona będzie odpowiednia z następujących nierówności:

$$-\frac{df(x)}{dx} > \operatorname{ctg} \theta_i, \quad (68)$$

$$\frac{df(x)}{dx} > \operatorname{ctg} \theta_i. \quad (69)$$

Nierówność (68) jest warunkiem wystąpienia samozacienień w sytuacji, którą przedstawiono na rysunku 7a. Jeśli zwrot osi $0x$, układu współrzędnych $0xy$ pokazanego na rysunku 7a, zostanie zmieniony na przeciwny, to wystąpienie samozacienień powierzchni określa nierówność (69). Spełnienie odpowiedniego z warunków (68) lub (69) oznacza, że na powierzchni istnieją takie punkty, które będą zacienione przez inne fragmenty powierzchni.

To czy dany punkt powierzchni będzie bezpośrednio oświetlony przez falę padającą, czy też nie, zależy nie tylko od kąta padania światła, ale także od wysokości nierówności i kątów nachylenia stycznych do powierzchni. Dla powierzchni zdeterminowanych, np. powierzchni okresowych, określenie wielkości obszarów oświetlonych nie nastęrcza większych trudności. W przypadku powierzchni opisanych stacjonarnymi i ergodycznymi funkcjami losowymi można obliczyć prawdopodobieństwo oświetlenia danego punktu powierzchni, niezależnie od współrzędnej z tego punktu oraz kąta γ nachylenia stycznej w tym punkcie. Prawdopodobieństwo $P(\theta)$ tego, że dany punkt powierzchni będzie oświetlony falą płaską, padającą pod kątem θ_i określone jest wg [3] następującym równaniem:

$$P(\theta_i) = \frac{\sum_{i=1}^{i=n} l_i}{l_c} = S(\theta_i, z') \int_{-\infty}^{\operatorname{ctg} \theta_i} p(z') dz', \quad (70)$$

gdzie: l_i – długości rzutów oświetlonych obszarów powierzchni $f(x)$ na oś $0x$, l_c – całkowita długość powierzchni oświetlonej wzdłuż osi $0x$, $S(\theta_i, z')$ – funkcja zacienienia powierzchni, $z'(x)$ – pochodna powierzchni wzdłuż osi $0x$, $p(z')$ – gęstość prawdopodobieństwa pochodnej powierzchni w kierunku x .

Funkcja zacienienia $S(\theta_i, z')$, występująca w równaniu (70), może być interpretowana jako prawdopodobieństwo tego, że fala płaska, padająca na powierzchnię pod kątem θ_i , oświetla bezpośrednio dany punkt powierzchni, w którym kąt nachylenia stycznej wynosi γ – bez względu na wartość rzędnej z w tym punkcie i wartości pochodnych wyższych rzędów. Funkcja ta zależy od kąta padania θ_i oraz od wartości pochodnej $df(x)/dx = z'(x) = \operatorname{tg} \gamma$.

Idea wykorzystania funkcji zacienienia oparta jest na założeniu, że średnie natężenie światła \bar{I}_{sh} , otrzymane w wyniku rozproszenia i towarzyszących mu samozacienie

nień, równe jest iloczynowi funkcji zacielenia i średniego natężenia światła rozproszonego \bar{I} , wyznaczonego bez uwzględnienia zacielen [51]:

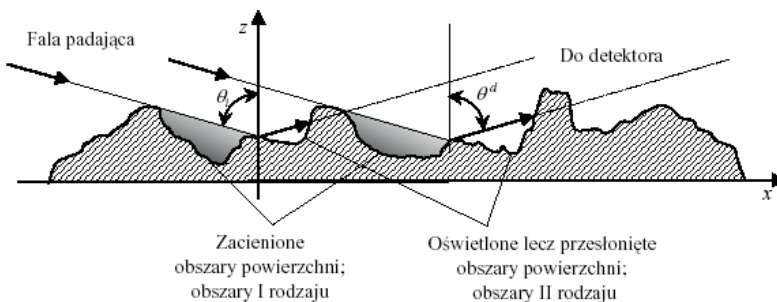
$$\bar{I}_{sh} \approx \bar{I} S(\theta_i, z'), \quad (71)$$

przy czym

$$0 \leq S(\theta_i, z') \leq 1. \quad (72)$$

Równanie (71) jest równaniem przybliżonym, gdyż zakłada całkowitą statystyczną niezależność średniego natężenia światła rozproszonego oraz funkcji zacielenia $S(\theta_i, z')$. Założenie to nie zawsze jest spełnione. Jednakże wyznaczając średnie natężenie światła rozproszonego metodą Kirchhoffa, przyjmuje się, iż jest ono niezależne od pochodnej powierzchni, co pozwala uznać równanie (71), w tym przypadku, za wystarczająco dokładne.

Funkcja zacielenia może mieć bardziej ogólny charakter, jeżeli uwzględni się fakt, że samozacielenia powierzchni dotyczą nie tylko światła padającego na powierzchnię, ale także światła odbitego od niej. Nierówności powierzchni mogą zacielen inne obszary powierzchni zarówno w odniesieniu do źródła światła, jak i w stosunku do punktu lub kierunku obserwacji, czy też miejsca, w którym umieszczono detektor światła rozproszonego. Ten ostatni przypadek niekiedy nazywany jest zjawiskiem samoprzesłania powierzchni. Oba rodzaje samozacielenia powierzchni pokazano schematycznie na rysunku 8. Obszary I rodzaju są zacielen w taki sposób, że nie dociera do nich bezpośrednio fala świetlna padająca na powierzchnię po kącie θ_i . Obszary II rodzaju są oświetlone falą padającą, lecz światło odbite od powierzchni w tych obszarach nie dociera bezpośrednio do detektora, jeśli detektor jest umieszczony pod odpowiednio dużym kątem detekcji θ^d . Obszary II rodzaju są więc przesłonięte przez sąsiadujące z nimi nierównościami powierzchni i nie są analizowane przez układ detekcyjny. Dla uproszczenia analizy często uznaje się, że obszary II rodzaju są zacielen – tzn. przesłonięte przez powierzchnię – w stosunku do detektora. Umożliwia to zastosowanie w podobny sposób analizy w odniesieniu do obszarów I i II rodzaju.



Rys. 8. Samozacielenie i samoprzesłanie powierzchni
Fig. 8. Surface self-shadowing and surface self-hiding

W pracy F.G. Bassa i I.M. Fuksa [3] oraz w pracy J.A. Ogilvy [51] opisano sposoby wyznaczania funkcji zacienienia i pokazano jej wykresy. Do wyznaczania tej funkcji wykorzystuje się teorię przewyższeń. Teoria ta, nazywana też teorią skoków poza-barierowych [50], określa relacje występujące między funkcją losową, np. funkcją $f(x)$ opisującą powierzchnię nierówną, a pewną zdeterminowaną funkcją $\varphi(x)$, którą może być np. prosta łącząca punkt zacinający i punkt zacieniany. Przewyższeniem funkcji losowej $f(x)$ nazywane jest przekroczenie przez tę funkcję wartości $\varphi(x)$. Jeżeli przekroczenie następuje w taki sposób, że w punkcie przecięcia się obu funkcji $df(x)/dx > d\varphi(x)/dx$, to przewyższenie jest dodatnie. Jeśli natomiast w punkcie przecięcia obu funkcji $df(x)/dx < d\varphi(x)/dx$, to przewyższenie jest ujemne.

Stosując teorię przewyższeń do wyznaczania funkcji zacienienia, uwzględniającej zarówno samozacienienia powierzchni w stosunku do źródła światła, jak i w stosunku do detektora światła rozproszonego, w pracy [3] otrzymano następującą zależność:

$$S(\theta_i, \theta^d, z') = \frac{1}{1 + L(a_i) + L(b_d)}, \quad (73)$$

przy czym parametry a_i i b_d dane są równaniami

$$a_i = \frac{\text{ctg } \theta_i}{\sigma_{z'}}, \quad (74)$$

$$b_d = \frac{\text{ctg } \theta^d}{\sigma_{z'}}, \quad (75)$$

natomiast funkcje $L(a_i)$ i $L(b_d)$ wynoszą odpowiednio:

$$L(a_i) = \text{tg } \theta_i \int_{\text{ctg } \theta_i}^{\infty} (z' - \text{ctg } \theta_i) p(z') dz', \quad (76)$$

$$L(b_d) = \text{tg } \theta^d \int_{\text{ctg } \theta^d}^{\infty} (z' - \text{ctg } \theta^d) p(z') dz', \quad (77)$$

gdzie $\sigma_{z'}$ jest odchyleniem standardowym pochodnej powierzchni $z' = df(x)/dx$, $p(z')$ zaś oznacza gęstość prawdopodobieństwa pochodnej powierzchni.

Jak wynika z przytoczonych równań, funkcja zacienienia powierzchni $S(\theta, \theta^d, z')$, przy danych kątach padania θ i detekcji θ^d , zależy tylko od statystycznych właściwości pochodnej powierzchni $z'(x)$, wyrażonych odchyleniem standardowym $\sigma_{z'}$ i gęstością prawdopodobieństwa $p(z')$. Również średnie natężenie światła, otrzymane w rezultacie rozproszenia światła i towarzyszących mu zacień, zgodnie z równaniem (71), zależy jedynie od gęstości prawdopodobieństwa pochodnej powierzchni. Podob-

ny wnioski można sformułować w odniesieniu do rozkładu natężenia światła rozproszonego. Rozważania te są oczywiście prawidłowe w przybliżeniu optyki geometrycznej, tzn. wtedy, gdy długość fali światła λ jest znacznie mniejsza od wymiarów charakteryzujących nierówność powierzchni.

Problem samozacieniania powierzchni był dotychczas rozpatrywany jedynie dla powierzchni dwuwymiarowych $z = f(x)$. Uogólnienie tych rozważań dla powierzchni trójwymiarowych $z = f(x, y)$ jest bardziej złożone. Wymaga ono analizy odpowiednich przekrojów powierzchni, zależnych od kierunku oświetlenia oraz właściwej składowej trójwymiarowej funkcji określającej pochodną powierzchni $z' = \nabla f(x, y)$.

W procesie pomiaru nierówności, samozacienianie powierzchni jest przeważnie traktowane jako zjawisko niepożądane, zakłócające i tak już skomplikowany proces rozpraszania światła. Dlatego najczęściej dąży się do eliminacji lub przynajmniej ograniczenia samozacieniania powierzchni, zmniejszając wartość kąta padania światła θ_i na powierzchnię rozpraszającą. Jeśli nie jest to możliwe, uwzględnia się wpływ samozacienień powierzchni na pole rozproszone, wykorzystując w tym celu funkcję zacienienia $S(\theta_i, \theta^d, z')$. Należy jednak pamiętać o tym, że wartości funkcji zacienienia mają istotne znaczenie przy dużych kątach padania θ_i , przy których, zgodnie z nierównością (13), stosowanie teorii Kirchhoffa jest ograniczone.

Funkcja zacienienia może być traktowana jako miara pewnej właściwości powierzchni. Właściwość, która umożliwia – przy odpowiednim oświetleniu i sposobie obserwacji – tworzenie charakterystycznego, zacienionego obrazu powierzchni. Obraz ten zależy od sposobu i kierunku oświetlenia powierzchni, a także od metody obserwacji światła odbitego. Analiza obrazu powierzchni tworzącej samozacienienia jest w wielu przypadkach bardzo pożyteczna. Na jej podstawie można ocenić niektóre charakterystyczne cechy nierówności powierzchni i wyznaczyć ich pewne parametry geometryczne. Badając zacienienia powierzchni, korzystnie jest zmieniać nie tylko kąt padania światła θ_i , ale także położenie powierzchni względem kierunku oświetlenia. Widoczne jest tu pewne podobieństwo z metodą ciemnego pola – niezwykle użytecznym i powszechnie stosowanym sposobem obserwacji w mikroskopii optycznej.

2.5.2. ODBICIA WIELOKROTNE

Analiza zjawiska rozpraszania światła przez powierzchnie nierówne dokonywana jest na ogół przy założeniu, że padająca na powierzchnię fala świetlna, odbija się od powierzchni tylko raz. Zjawisko odbicia światła od powierzchni płaskiej jest w tym sensie aktem jednorazowym. Fala świetlna po odbiciu się od powierzchni propaguje w przestrzeni otaczającej powierzchnię. Nie ma powtórnego oddziaływania powierzchni i odbitej fali świetlnej. Jeżeli na powierzchni pojawiają się niewielkie nierówności, kąt padania fali zaś nie będzie zbyt duży, to można przyjąć, że nierówna powierzchnia odbijająca światło – podobnie jak wspomniana poprzednio powierzch-

nia płaska – nie wpływa ponownie na falę odbitą. Sytuacja zmieni się, jeżeli nierówności powierzchni oraz kąty padania światła na powierzchnię będą większe. Mogą wtedy zachodzić odbicia wielokrotne. Odbicia te dotyczą tylko niektórych fragmentów frontu fali padającej na powierzchnię.

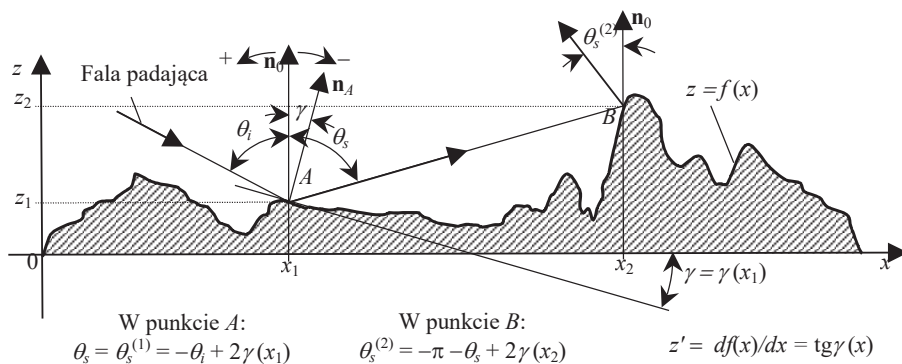
Niech na powierzchnię nierówną $z = f(x)$, pod kątem θ_i w płaszczyźnie $0xz$, pada płaska fala światła. Rozważmy, w przybliżeniu promieniowym, odbicie tej fali w dowolnie wybranym punkcie powierzchni o współrzędnych x_1, z_1 (rys. 9). Odbicia wielokrotne wystąpią wtedy, gdy układ równań

$$\left. \begin{aligned} z &= -x \operatorname{ctg} \theta_s + z_1 + x_1 \operatorname{ctg} \theta_s, \\ z &= f(x), \end{aligned} \right\} \quad (78)$$

gdzie

$$\theta_s = -\theta_i + 2\gamma(x_1), \quad (79)$$

będzie miał, oprócz punktu o współrzędnych x_1, z_1 , co najmniej jedno takie dodatkowe rozwiązanie x_2, z_2 , które wystąpi dla argumentu $x_2 > x_1$ przy $\theta_s < 0$, albo dla argumentu $x_2 < x_1$ przy $\theta_s > 0$. W równaniach (78) i (79) przyjęto następujące oznaczenia: θ_i – kąt padania światła na powierzchnię, x_1, z_1 – współrzędne punktu powierzchni, w którym rozpatrywane jest odbicie światła, θ_s – kąt odbicia światła w rozpatrywanym punkcie powierzchni, $\gamma(x_1)$ – kąt nachylenia stycznej do powierzchni w punkcie odbicia, $z = f(x)$ – funkcja opisująca powierzchnię.



Rys. 9. Odbicie wielokrotne od powierzchni nierównej

Fig. 9. Multiple reflection from a rough surface

Po stwierdzeniu, że dla danego fragmentu frontu fali świetlnej nastąpiło dwukrotne odbicie od powierzchni, trzeba rozpatrzyć w podobny sposób, czy nie nastąpią kolejne odbicia – najpierw trzecie, potem czwarte, a w ogólnym przypadku n -te. Kąt padania $\theta_i^{(n)}$, w każdym kolejnym punkcie odbicia, zależy od kąta odbicia $\theta_s^{(n-1)}$ w punkcie

poprzednim. Bezwzględna wartość kąta $\theta_i^{(n)}$, w odróżnieniu od kąta θ_i , może przekraczać $\pi/2$. Kąt $\theta_i^{(n)}$ dany jest równaniami:

$$\theta_i^{(n)} = \pi + \theta_s^{(n)}, \quad \text{jeśli } \theta_s^{(n-1)} < 0 \text{ dla } n = 2, 3, 4, \dots, \quad (80)$$

$$\theta_i^{(n)} = -\pi + \theta_s^{(n)}, \quad \text{jeśli } \theta_s^{(n-1)} > 0 \text{ dla } n = 2, 3, 4, \dots, \quad (81)$$

gdzie: $\theta_i^{(n)}$ – kąt padania światła w punkcie n -tego odbicia, $\theta_s^{(n-1)}$ – kąt odbicia światła w punkcie $n-1$, n – kolejny numer punktu, w którym zachodzą odbicia wielokrotne. Kąty θ_i , $\theta_i^{(n)}$, θ_s i $\theta_s^{(n-1)}$ – jak pokazano na rysunku 9 – wyznacza się względem normalnej \mathbf{n}_0 do płaszczyzny $z = 0$. Kąt jest dodatni, gdy wyznaczany jest od normalnej \mathbf{n}_0 odwrotnie do ruchu wskazówek zegara. Kąt odmierzany od normalnej \mathbf{n}_0 zgodnie z ruchem wskazówek zegara jest ujemny.

Warunkiem odbicia światła w kolejnym punkcie powierzchni, o współrzędnych x_n, z_n , jest wystąpienie co najmniej jednego dodatkowego rozwiązania układu równań:

$$\left. \begin{aligned} z &= -x \operatorname{ctg} \theta_s^{(n)} + z_n + x_n \operatorname{ctg} \theta_s^{(n)}, \\ z &= f(x), \end{aligned} \right\} \quad (82)$$

przy czym

$$\theta_s^{(n)} = -\theta_i^{(n)} + 2\gamma(x_n) \quad \text{dla } n = 2, 3, 4, \dots, \quad (83)$$

które to rozwiązanie spełniać będzie podobne postulaty, jak rozwiązanie układu (78). To dodatkowe rozwiązanie powinno zatem występować dla wartości $x_{n+1} > x_n$ przy $\theta_s^{(n)} < 0$ albo dla wartości $x_{n+1} < x_n$ przy $\theta_s^{(n)} > 0$, gdzie $\theta_s^{(n)}$ jest kątem odbicia w punkcie n .

W celu ilościowego określenia wymiarów obszarów powierzchni sprzyjających odbiciom wielokrotnym, należy rozwiązywać układ równań (78) kolejno we wszystkich punktach powierzchni. Następnie procedura ta powinna być stosowana rekurencyjnie, z wykorzystaniem układu równań (82), dla kolejnych punktów, w których nastąpiły odbicia wielokrotne. Jak widać, analiza odbić wielokrotnych jest zadaniem trudnym z kilku powodów. Po pierwsze, nieznana jest na ogół funkcja $z = f(x)$, opisująca powierzchnię. Po drugie, nieznane są wartości kątów odbicia w kolejnych punktach powierzchni. Po trzecie, liczba rozpatrywanych punktów powierzchni może być bardzo duża.

Niezależnie od tego, że analiza odbić wielokrotnych jest zadaniem trudnym, warto jest zwrócić uwagę na pewną charakterystyczną cechę tego zjawiska. Rozpatrując nominalnie płaską powierzchnię nierówną $z = f(x)$, łatwo zauważyć, że odbicia wielokrotne na pewno pojawiają się wtedy, gdy w rozpatrywanym punkcie powierzchni kąty odbicia światła θ_s lub $\theta_s^{(n)}$ będą spełniać odpowiedni z podanych warunków

$$|\theta_s| > \frac{\pi}{2}, \quad \text{albo} \quad |\theta_s^{(n)}| > \frac{\pi}{2}. \quad (84)$$

Odbicia wielokrotne mogą pojawić się także przy mniejszych wartościach kątów θ_s i $\theta_s^{(n)}$, ale nie jest to pewne. Ze wzoru (79), zmodyfikowanego dla dowolnego punktu powierzchni, i z pierwszej nierówności (84) otrzymuje się

$$|-\theta_i + 2\gamma(x)| > \frac{\pi}{2}, \quad (85)$$

Spełnienie nierówności (85) – przez kąt padania θ_i oraz przez kąt nachylenia stycznej $\gamma(x)$ w danym punkcie powierzchni – daje pewność, że fala odbita w tym punkcie jeszcze co najmniej raz odbije się od powierzchni.

Dla powierzchni opisanej stacjonarną i ergodyczną funkcją losową, o danej gęstości prawdopodobieństwa kątów nachylenia zboczy nierówności $p(\gamma)$, prawdopodobieństwo wystąpienia odbić wielokrotnych $P(\theta_i, \gamma)$, wynikające ze spełnienia nierówności (85), określone jest równaniem

$$P(\theta_i, \gamma) = P\left(|-\theta_i + 2\gamma(x)| > \frac{\pi}{2}\right) = 1 - \int_{\frac{\theta_i}{2} - \frac{\pi}{4}}^{\frac{\theta_i}{2} + \frac{\pi}{4}} p(\gamma) d\gamma. \quad (86)$$

Mimo wielu wysiłków, dotychczas nie opracowano dokładnych i praktycznie zadowalających, metod analizy odbić wielokrotnych. Zarówno teoria Kirchhoffa, jak i teoria zaburzeń pierwszego rzędu, pomijają zagadnienie odbić wielokrotnych. Jedną z koncepcji zmierzających do rozwiązania tego problemu polega na tym, aby rozpatrywać rozpraszanie lub dyfrakcję światła na powierzchni płaskiej, na której, według określonej zasady, rozmieszczone są zbiory pojedynczych, identycznych nierówności o zdefiniowanym kształcie [51]. Nierówności te mają małe wymiary, w porównaniu z długością fali światła, i mogą być wklęsłe lub wypukłe. Najczęściej przyjmowany jest półcylicyndryczny lub półkulisty, wypukły kształt pojedynczych nierówności. Analiza odbić wielokrotnych dla takiego uproszczonego modelu powierzchni wymaga najpierw ustalenia, w jaki sposób światło rozpraszane jest przez pojedynczą nierówność. Zazwyczaj pole rozproszone przez pojedynczy element wyznaczone jest jako suma pola uzyskanego w wyniku odbicia od idealnej powierzchni i pola rozproszonego przez pojedynczą kulę lub walec. Kolejne etapy analizy to przyjęcie modelu przestrzennego rozmieszczenia nierówności i dokonanie superpozycji fal odbitych z uwzględnieniem oddziaływania sąsiednich nierówności.

W pracy [71] przyjęto, że z każdym kulistym elementem rozpraszającym związany jest lokalny układ współrzędnych sferycznych. Pole rozproszone przez pojedynczy element kulisty ma symetrię osiową, to znaczy nie zależy od kąta azymutalnego określonego dla każdego elementu w płaszczyźnie Oxy . Pole to przedstawiono w postaci nieskończonej sumy iloczynów ortogonalnych wielomianów Legendre'a i sferycznych funkcji Hankela. Kolejne wielomiany Legendre'a reprezentują składowe harmoniczne

poła rozproszonego. Każda składowa spełnia warunki graniczne na powierzchni. Zakładając odpowiednie rozmieszczenie elementów rozpraszających na płaszczyźnie i dokonując superpozycji pól rozproszonych przez wszystkie elementy, w pracy [71] otrzymano równanie pozwalające, dzięki metodzie iteracyjnej, ocenić pole rozproszone z uwzględnieniem odbić wielokrotnych. Mimo dużej dokładności obliczeń uzyskiwanej w metodach iteracyjnych, o przydatności tego sposobu analizy odbić wielokrotnych decyduje stopień zgodności przyjętego modelu nierówności z rzeczywistością.

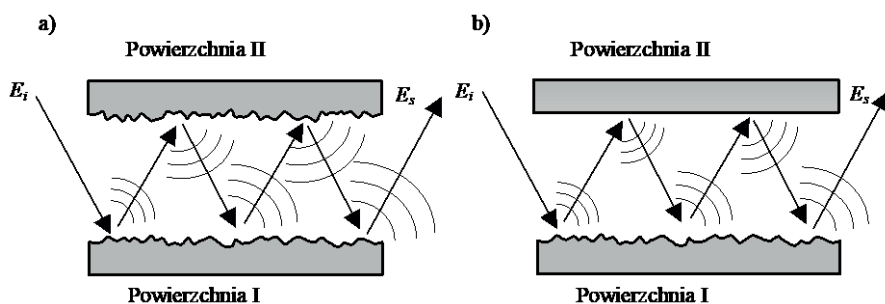
Znane są także inne metody analizy, które uwzględniają odbicia wielokrotne w zagadnieniach związanych z rozpraszaniem fal elektromagnetycznych przez powierzchnie nierówne. Najbardziej dokładne, ale jednocześnie skomplikowane i trudne w praktycznych zastosowaniach, są metody wykorzystujące równania całkowite Fredholma drugiego rodzaju. W Polsce problematyka zastosowania metod matematycznych, w tym równań całkowitych, do analizy rozpraszania światła, zwłaszcza przez ciecze i gazy, rozwijana jest w Katedrze Metrologii Elektronicznej i Fotonicznej Wydziału Elektroniki Politechniki Wrocławskiej przez J. Mroczkę i jego współpracowników [45], [73], [9], [46], [19], [20], [26]. Oddzielną grupę metod analizy rozpraszania światła na nierównościach powierzchni, z uwzględnieniem odbić wielokrotnych, stanowią metody rachunku wariacyjnego. Są to niestety też metody złożone i mało przydatne w zastosowaniach technicznych. W pracy J.A. Ogilvy [51] przedstawiono różne sposoby rozwiązywania równań całkowitych Fredholma i opisano podstawowe rodzaje metod wariacyjnych.

Odbicia wielokrotne, podobnie jak samozacienienia powierzchni, są często niepożądane. Występuje to wtedy na przykład, gdy zjawisko rozpraszania światła jest wykorzystywane do oceny nierówności powierzchni. Nierówności są w wielu przypadkach oceniane pośrednio, np. na podstawie pomiarów przestrzennego rozkładu natężenia światła rozproszonego przez powierzchnię. Odbicia wielokrotne sprzyjają rozpraszaniu światła w szerszym zakresie kątów, co w rezultacie może prowadzić do błędnej oceny nierówności powierzchni.

Rozpatrując tę sytuację na gruncie optyki falowej, można stwierdzić, że powierzchnia nierówna, na której zachodzą odbicia wielokrotne, wprowadza znacznie głębszą przestrzenną modulację fazową fali świetlnej niż powierzchnia, od której fala odbija się tylko raz. Modulacja fazowa, powodowana odbiciami wielokrotnymi, jest trudna do przewidzenia i oszacowania. Rezultatem odbić wielokrotnych jest także przestrzenna modulacja amplitudowa fali świetlnej, gdyż reflektancja całkowita powierzchni rzeczywistych jest mniejsza od jedności. Dlatego każde kolejne odbicie od powierzchni powoduje zmniejszenie amplitudy i natężenia światła odbitego. Jeżeli reflektancja jest stała na całym oświetlonym obszarze powierzchni, to natężenie fali odbitej n -krotnie jest proporcjonalne do n -tej potęgi reflektancji. Głębokość modulacji amplitudowej wzrasta, jeśli maleje reflektancja i rośnie liczba odbić n . Modulacja amplitudowa i modulacja fazowa, wywołane odbiciami wielokrotnymi, dotyczą tylko niektórych, na ogół trudnych do zlokalizowania, fragmentów czoła fali świetlnej.

Jak wynika z tego, liczne odbicia wielokrotne – w rezultacie wprowadzenia dodatkowej modulacji fazowej i amplitudowej fali odbitej – stanowią przeszkodę w skorelowaniu rozkładu natężenia światła rozproszonego przez powierzchnię z parametrami nierówności, które na niej występują. Może to stawiać pod znakiem zapytania przydatność metod oceny nierówności powierzchni, opartych na wykorzystaniu zjawiska rozpraszania światła. Na szczęście, jak pokazano wcześniej, prawdopodobieństwo wystąpienia odbić wielokrotnych dla większości precyzyjnie obrobionych powierzchni jest niewielkie, zwłaszcza przy kątach padania bliskich zeru.

Dotychczas rozważany był problem odbić wielokrotnych. Polega on na tym, że światło padające na powierzchnię odbija się od niej więcej niż jeden raz. Pomiedzy kolejnymi odbiciami światło propaguje swobodnie w przestrzeni otaczającej powierzchnię. Niekiedy zachodzi taka sytuacja, w której światło odbite od powierzchni nierównej zostanie na nią skierowane ponownie, na skutek odbicia od innej powierzchni. Tę sytuację – pokazaną schematycznie na rysunku 10 – nazywa się rozpraszaniem wielokrotnym padającej fali świetlnej E_i przez powierzchnię nierówną. Rozpraszanie wielokrotne, nawet na małych nierównościach, powoduje duże zaburzenia amplitudy i fazy pola wynikowego E_s .



Rys. 10. Rozpraszanie wielokrotne w wyniku odbicia od powierzchni: a) nierównej, b) gładkiej
Fig. 10. Multiple scattering due to reflection from: a) a rough surface, b) a smooth surface

Są one skutkiem kumulowania się zaburzeń fali świetlnej spowodowanych kolejnymi aktami rozpraszania. Zaburzenia te będą większe wtedy, gdy liczba aktów rozpraszania będzie wzrastać oraz wtedy, gdy większe będą nierówności dodatkowej powierzchni odbijającej. Rozpraszanie wielokrotne pojawia się np. podczas przejścia światła przez cienkie warstwy, na co wskazano w pracy [52].

W pracy [3] przedstawiono złożoną teorię, umożliwiającą przybliżone obliczanie pola wynikowego fali świetlnej, które powstaje na skutek wielokrotnego rozpraszania przez powierzchnie o niewielkich nierównościach. Teoria ta oparta jest na wykorzystaniu metody funkcji Greena, opracowanej wcześniej na potrzeby elektrodynamiki kwantowej i kwantowej teorii pola. Główna idea tej teorii polega na tym, że zaburzenie fali świetlnej spowodowane nierównościami zawarte jest nie w samym równaniu

falowym Helmholtza, lecz w warunkach granicznych tego równania. Mimo przydatności tej teorii w analizie zjawiska rozpraszania światła, dotychczas nie znalazła ona zastosowania w pomiarach nierówności powierzchni.

Rozpraszanie wielokrotne, podobnie jak odbicia wielokrotne, jest na ogół zjawiskiem niepożądanym w pomiarach nierówności metodami rozpraszania światła, gdyż komplikuje dodatkowo i tak już złożoną sytuację. Niekiedy jednak może ono być pożyteczne. Jeśli na przykład oceniamy nierówności bardzo gładkiej powierzchni, to jednokrotne rozpraszanie światła na tej powierzchni może powodować na tyle małe zaburzenia fali odbitej, że ich detekcja będzie bardzo trudna. Rozpraszanie wielokrotne spowoduje natomiast korzystny wzrost sygnału pomiarowego, proporcjonalny do liczby aktów rozpraszania.

3. METODY OCENY NIERÓWNOŚCI POWIERZCHNI OPARTE NA WYKORZYSTANIU ZJAWISKA ROZPRASZANIA ŚWIATŁA

3.1. WPROWADZENIE

Metody, które rozpatrzono w niniejszym rozdziale, umożliwiają dokonywanie oceny nierówności na podstawie zjawiska rozpraszania fal świetlnych oraz promieniowania podczerwonego i nadfioletowego, przez powierzchnię przedmiotu mierzonego. Jeżeli powierzchnia jest płaska i idealnie gładka, to światło odbijane jest od niej zgodnie z prawem odbicia światła. Energia światła odbitego od tej powierzchni emitowana jest w jednym kierunku, który nazywany jest kierunkiem odbicia zwierciadlanego. Odbicie takie nazywa się odbiciem zwierciadlanym. Jeśli na powierzchni występują nierówności, których wymiary porównywalne są z długością fali światła, to część energii zostanie odbita zgodnie z prawem odbicia, część zaś ulegnie rozproszeniu w wielu kierunkach. W przypadku gdy nierówności są znacznie większe od długości fali światła, przeważająca część energii światła odbitego ulega rozproszeniu. Odbicie takie nazywane jest odbiciem dyfuzyjnym.

Jednym z powodów zainteresowania metodami rozpraszania światła jest ich wysoka czułość, umożliwiająca mierzenie bardzo małych nierówności powierzchni, których wysokość wynosi od dziesiątych części nanometra do kilkuset nanometrów [21], [22]. Tak niewielkie nierówności powierzchni są trudne do zmierzenia. Inną ważną zaletą metod rozpraszania światła jest możliwość wykorzystania ich do szybkiej kontroli nierówności powierzchni przedmiotów obrobionych mechanicznie, np. powierzchni szlifowanych, polerowanych, docieranych itp., wykonywanych w wydajnych, zautomatyzowanych procesach produkcyjnych. W tym przypadku mierzone nierówności mogą być nieco większe.

Niezależnie od zakresu mierzonych nierówności powierzchni ich ocena metodami rozpraszania światła dokonywana jest na ogół na podstawie pomiarów następujących wielkości [40]:

- natężenia światła odbitego zwierciadlanie,
- natężenia światła rozproszonego w jednym lub w kilku kierunkach,
- całkowitego natężenia światła rozproszonego,
- rozkładu natężenia światła odbitego od powierzchni.

W języku polskim nie ma jak dotychczas powszechnie przyjętego nazewnictwa, dotyczącego metod oceny nierówności opartych na wykorzystaniu zjawiska rozpraszania światła. Jak już wspomniano w p. 1.1, coraz częściej ogół takich metod określa się terminem: skaterometria nierówności powierzchni. Słowo *skaterometria* utworzono od angielskiego terminu *scattering* – rozpraszanie.

Metody wykorzystujące pomiar natężenia światła odbitego zwierciadlanie nazywane są często metodami reflektometrycznymi lub rzadziej odbiciowymi. Nazwy te obejmują również takie metody, w których – oprócz natężenia światła odbitego zwierciadlanie – mierzone jest także natężenie światła rozproszonego w jednym lub w kilku wybranych kierunkach. Należy odnotować, iż nazwa *metody reflektometryczne* stosowana jest niekiedy w literaturze w szerszym sensie i obejmuje wszystkie te metody rozpraszania, w których następuje odbicie światła od powierzchni badanej.

Metody oparte na pomiarze całkowitego natężenia światła rozproszonego nie doczekały się dotychczas powszechnie przyjętej nazwy w języku polskim. W literaturze anglojęzycznej stosuje się w odniesieniu do nich termin *total integrated scattering* (lub *total integrated scatter*) i oznacza się je symbolem TIS. W języku polskim czasem nazywane one są metodami integracyjnymi [41].

Natomiast metody, w których dokonywany jest pomiar kąтового rozkładu natężenia światła odbitego od powierzchni, nazywane są niekiedy metodami goniofotometrycznymi. W języku angielskim metody te określa się jako *angle-resolved scattering* lub *differential scattering* i oznacza odpowiednio symbolami ARS oraz DS. Podobnie w języku rosyjskim, w odniesieniu do tych metod używa się terminu *дифференциальное рассеяние*. W języku polskim metody te nazywane bywają metodami różniczkowymi [41].

Niezależnie od podanych nazw różnych metod rozpraszania światła w stosunku do metod, których podstawy teoretyczne oparte są na teorii dyfrakcji, używa się niekiedy w języku polskim nazwy metody dyfrakcyjne, analogicznie jak w języku angielskim określenia *diffraction methods*.

W fizyce, w metodach rozpraszania światła nierówności powierzchni oceniane są na podstawie pomiarów natężenia światła – czyli kwadratu modułu zespolonej amplitudy fali świetlnej, odbitej w określonym kierunku – oraz na podstawie wartości kątów określających kierunki propagacji tej fali. Natężenie światła odbitego w danym kierunku zależy nie tylko od kąta padania, długości fali światła i nierówności powierzchni. Zależy ono także od natężenia światła padającego oraz od rodzaju i struktury materiału powierzchni. Aby uniknąć wpływu tych czynników na wyniki pomiarów nierówności powierzchni stosuje się często pomiary względne. Polegają one na wyznaczeniu stosunku dwóch tak dobranych wielkości, że stosunek ten zależy przede wszystkim od stanu nierówności badanej powierzchni.

3.2. METODY REFLEKTOMETRYCZNE

Pierwsze próby wykorzystania pomiarów natężenia światła odbitego do oceny nierówności powierzchni podjęto w Niemczech w latach 30. XX wieku. G. Schmaltz badał rozkład światła odbitego od powierzchni, porównując z nim krzywą Gaussa. B. Lange rozpoczął badania połysku powierzchni. Z czasem opracowano obiektywne metody oceny połysku powierzchni przedmiotów wykonanych ze stopów metali, papieru [32], drewna, tworzyw sztucznych itp.

W latach 60. XX w. – w rezultacie prac prowadzonych między innymi przez H.E. Bennetta i J.O. Porteusa [7] oraz P. Beckmanna [6], [5] – opracowano teorię, określającą zależność między odchyleniem standardowym wysokości nierówności a stosunkiem reflektancji zwierciadlanej i całkowitej. Zależność tę wyraża równanie

$$\frac{R_s}{R_0} = \exp \left[- \left(\frac{4\pi\sigma \cos\theta_i}{\lambda} \right)^2 \right], \quad (87)$$

gdzie: R_s – reflektancja zwierciadlana, R_0 – reflektancja całkowita, σ – odchylenie standardowe wysokości nierówności, θ_i – kąt padania, λ – długość fali światła. Równanie to otrzymano, zakładając spełnienie następujących warunków:

- odchylenie standardowe wysokości nierówności powierzchni jest znacznie mniejsze niż długości fali światła ($\sigma \ll \lambda$),

- odstępów nierówności S_i są znacznie większe od długości fali światła ($S_i \gg \lambda$), jednak na tyle małe, na ile kierunki propagacji światła rozproszonego nie są zbyt bliskie kierunkowi odbicia zwierciadlanego,

- nierówności powierzchni opisuje stacjonarna i ergodyczna funkcja losowa, natomiast rozkład wysokości nierówności powierzchni jest rozkładem normalnym.

Ostatni z wymienionych warunków – dotyczący normalności rozkładu wysokości nierówności – nie jest konieczny, jak wykazali to E.L. Church, H.A. Jenkinson i J.M. Zavada w swoich pracach [16], [17]. Niekiedy dodatkowo przyjmuje się założenie, że powierzchnia odbijająca ma nieskończenie wysoką przewodność elektryczną. Oznacza to, że powierzchnia taka nie pochłania światła padającego, odbijając je w całości.

Zależność (87), będąca podstawą reflektometrycznych, a także integracyjnych metod oceny nierówności powierzchni, bezpośrednio wynika z zależności (35). Należy tylko wziąć pod uwagę, że natężenie światła jest uśrednionym w czasie kwadratem modułu zespolonej amplitudy światła.

3.2. METODY INTEGRACYJNE

Z metodami reflektometrycznymi ściśle związane są metody integracyjne, w których podstawą oceny nierówności powierzchni jest pomiar reflektancji dyfuzyjnej [60], [8]. W celu wyznaczenia reflektancji dyfuzyjnej mierzy się sumaryczne (zinte-

growane) natężenie światła rozproszonego we wszystkich kierunkach poza zwierciadlanym. Ponieważ reflektancja całkowita R_0 jest sumą reflektancji zwierciadlanej R_s i reflektancji dyfuzyjnej R_d , to lewa strona zależności opisanej równaniem (87) może być przekształcona do postaci

$$\frac{R_s}{R_0} = \frac{R_0 - R_d}{R_0} = 1 - \frac{R_d}{R_0}, \quad (88)$$

Stosunek reflektancji dyfuzyjnej do reflektancji całkowitej, oznaczany symbolem TIS (*total integrated scattering*), jest zatem równy

$$\text{TIS} = \frac{R_d}{R_0} = 1 - \exp \left[- \left(\frac{4\pi\sigma \cos\theta_i}{\lambda} \right)^2 \right], \quad (89)$$

Wykorzystując równanie (89) do oceny powierzchni o bardzo małych nierównościach, stosuje się często dwa uproszczenia. Po pierwsze, dla takich powierzchni przyjmuje się, że

$$\text{TIS} = \frac{R_d}{R_0} = 1 - \exp \left[- \left(\frac{4\pi\sigma \cos\theta_i}{\lambda} \right)^2 \right] \approx \left(\frac{4\pi\sigma \cos\theta_i}{\lambda} \right)^2, \quad (90)$$

gdyż odchylenie standardowe σ wysokości nierówności powierzchni jest wielokrotnie mniejsze od długości fali światła λ . Wtedy wartość odchylenia standardowego wysokości nierówności określana jest na podstawie równania:

$$\sigma \approx \frac{\lambda}{4\pi \cos\theta_i} \sqrt{\text{TIS}}. \quad (91)$$

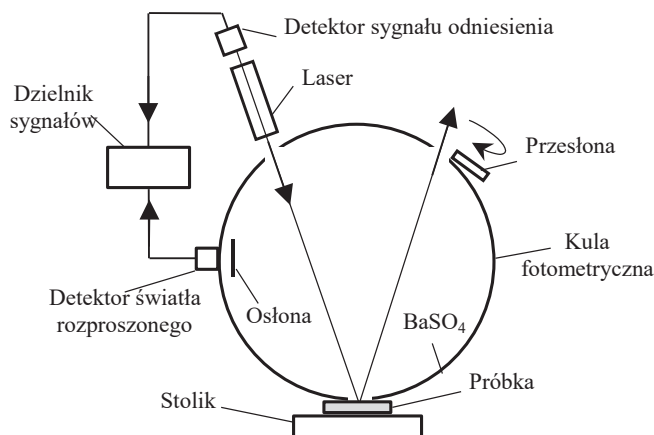
Po drugie, jeśli reflektancja dyfuzyjna R_d stanowi około 1% reflektancji całkowitej R_0 , jak na przykład w przypadku powierzchni optycznych, to TIS można określić w przybliżeniu na podstawie pomiarów natężenia światła odbitego zwierciadlanie i dyfuzyjnie:

$$\text{TIS} = \frac{R_d}{R_0} = \frac{R_d}{R_s + R_d} \approx \frac{R_d}{R_s}. \quad (92)$$

Najczęściej pomiary parametru *TIS* prowadzone są w specjalnych układach pomiarowych. Jeden z takich układów wykorzystuje półkuliste zwierciadło wklęsłe, zaproponowane na początku XX w. do pomiarów całkowitego natężenia światła rozproszonego przez amerykańskiego fizyka Williama W. Coblentza. Jest ono określane w literaturze anglojęzycznej terminem *Coblentz sphere*. Zwierciadło to umożliwia skupić praktycznie całe promieniowanie rozproszone przez badaną powierzchnię na powierzchni detektora fotoelektrycznego, umieszczonego w ognisku zwierciadła [58],

[23]. Układ ten jest bardzo czuły na usytuowanie próbki i detektora fotoelektrycznego względem zwierciadła półkulistego. W innych układach pomiarowych do pomiaru parametru TIS zamiast kuli Coblenta stosowana jest kula fotometryczna, nazywana również kulą całkującą lub kulą Ulbrichta. Wnętrze kuli pokryte jest materiałem rozpraszającym. Na ogół stosuje się w tym celu siarczan baru (BaSO_4), tlenek magnezu (MgO) lub różne odmiany sproszkowanego policzterofluoroetylenu, takie jak teflon, spectralon, halon. Detektor umieszczony wewnątrz kuli umożliwia wyznaczenie wartości proporcjonalnej do całkowitego natężenia światła rozproszonego. Zaletą takiego rozwiązania jest to, że wynik pomiaru w niewielkim stopniu zależy od pozycji detektora fotoelektrycznego. Różne konfiguracje takich przyrządów opisano w pracy [41], natomiast na rysunku 11 pokazano schemat takiego układu.

Pomiary nierówności metodami integracyjnymi dokonywane są często w wielu miejscach badanej powierzchni. W tym celu mierzona próbka umieszczana jest na przesuwym stoliku, dokonującym pomiarów w kolejnych miejscach. Szerszy opis metod integracyjnych, ich analizę, a także wyniki badań uzyskane za pomocą tych metod, przedstawiono między innymi w książkach J.M. Bennett i L. Mattssona [8] oraz J.C. Stovera [60] oraz w pracach [58], [23], [27], [61], [62].



Rys. 11. Układ do pomiaru parametru TIS z kulą fotometryczną
Fig. 11. System for measuring TIS with Ulbricht sphere

Pomiary całkowitego natężenia światła rozproszonego umożliwiają ocenę odchylenia standardowego nierówności powierzchni w zakresie od 0,1 nm do około 35 nm. Poszerzenie zakresu pomiarowego w stronę większych nierówności powierzchni wymaga zastosowania promieniowania o większej długości fali. Amerykańska organizacja ASTM (*American Society for Testing and Materials*) ustaliła sposób kontroli chropowatości powierzchni zwierciadeł metalowych metodą TIS [2]. Sposób ten polega na wykorzystaniu światła lasera He-Ne, o długości fali $\lambda = 632,8$ nm, zwierciadła półku-

listego oraz odpowiednich wzorców rozpraszających światło dyfuzyjnie. Wzorce takie opracowano w NIST (*National Institute of Standards and Technology*) w USA. Wspomniana norma [2] została wycofana w 2003 r. i dotychczas nie została zastąpiona inną.

3.3. METODY RÓŻNICZKOWE

Oddzielną grupę metod rozpraszania światła stanowią metody różniczkowe. Ich cechą szczególną jest to, iż oceny nierówności powierzchni tymi metodami dokonuje się na podstawie pomiarów kąтового rozkładu natężenia światła rozproszonego [204]. Rozkład ten, przy odpowiednim oświetleniu niektórych powierzchni, można mierzyć w wybranej płaszczyźnie, np. w płaszczyźnie padania. W tym przypadku natężenie światła rozproszonego będzie funkcją jednej zmiennej, której rolę spełnia na ogół kąt rozproszenia θ_d . Graficzny obraz tej funkcji nazywa się indyktrycą rozproszenia. Dają one pełniejszą informację o zjawisku rozpraszania światła przez powierzchnię oraz o samej powierzchni, w porównaniu z metodami opartymi na pomiarze kilku pojedynczych wartości natężenia światła rozproszonego lub odbitego zwierciadlanie.

Rozkład natężenia światła rozproszonego można też wyznaczać w przestrzeni. W tym przypadku, dla danego kąta padania θ_i , będzie on funkcją dwóch zmiennych: biegunowego kąta rozproszenia θ_d i azymutalnego kąta rozproszenia ϕ . Funkcja ta, opisująca przestrzenny rozkład natężenia światła rozproszonego, oznaczana jest często symbolami BRDF (*bidirectional reflectance distribution function*) lub BSDF (*bidirectional scatter distribution function*). W pracy [47] F.E. Nicodemus zdefiniował ją jako stosunek luminancji energetycznej, w określonym kierunku rozproszenia, do mocy promienistej padającej na jednostkę powierzchni. Luminancję energetyczną L_e , w określonym kierunku rozproszenia, wyznacza się wewnątrz elementarnego kąta bryłowego $d\Omega$. Funkcję BRDF wyraża zależność

$$BRDF = \frac{L_e}{P_i / S_0} = \frac{dP / d\Omega}{P_i \cos \theta_d} = \frac{16\pi^2}{\lambda^4} \cos(\theta_i) \cos(\theta_d) QS(f_x, f_y), \quad (93)$$

przy czym

$$L_e = \frac{dP}{S_0 d\Omega \cos \theta_d}, \quad (94)$$

gdzie: L_e – luminancja energetyczna powierzchni oświetlonej, w określonym kierunku rozproszenia, S_0 – pole powierzchni oświetlonej, dP – moc promienista rozproszona wewnątrz elementarnego kąta bryłowego, $d\Omega$ – elementarny kąt bryłowy, P_i – moc promienista padająca na powierzchnię, λ – długość fali światła, θ_i – kąt padania, θ_d – kąt rozproszenia, Q – bezwymiarowy czynnik, zależny od azymutalnego kąta rozproszenia ϕ , uwzględniający także stan polaryzacji światła padającego i optyczne właści-

wości powierzchni, $S(f_x, f_y)$ – dwuwymiarowa funkcja gęstości widmowej mocy nierówności powierzchni odbijającej, f_x, f_y – częstości przestrzenne nierówności.

Łatwo zauważyć podobieństwo równania (93) do wzoru (48). Podstawowa różnica między tymi wyrażeniami polega na występowaniu we wzorze (93) bezwymiarowego czynnika polaryzacyjnego Q . Różnica ta wynika stąd, że wzór (93) otrzymano, korzystając z wektorowej teorii zaburzeń, zależność zaś (48) z teorii skalarnej.

Zależność (93) – jak pokazano w pracach E.L. Churcha, H.A. Jenkinsona i J.M. Zavady [16], [17], J.C. Stovera [60] oraz J.M. Bennett i L. Mattsona [8] – może być zapisana w różny sposób. Końcowa część wyrażenia (93) wynika z wektorowej teorii rozpraszania fal na powierzchniach nierównych, nazywanej także wektorową teorią zaburzeń Rayleigha–Rice’a [60]. Wyrażenie to jest podstawą do oceny ważnej charakterystyki – gęstości widmowej mocy nierówności powierzchni $S(f_x, f_y)$. Może być ono stosowane do oceny powierzchni, których wysokości nierówności są znacznie mniejsze od długości fali światła. W pomiarach rozkładu natężenia światła odbitego od takich powierzchni, stosowane są układy pomiarowe o szerokim zakresie dynamicznym, na ogół nie mniejszym niż 120 dB, oraz precyzyjne układy mierzące kąt rozproszenia. Różnorodne zastosowania metod różniczkowych przedstawiono między innymi w pracach [55], [57], [67].

Opisane dotychczas metody różniczkowe przeznaczone są do oceny bardzo małych nierówności powierzchni. Stosuje się je w procesach kontroli powierzchni elementów optycznych, krzemowych podłoży układów scalonych itp. Zaletami tych metod są przede wszystkim bezstykowość, wysoka czułość oraz krótki czas pomiaru. Zakres wysokości nierówności mierzonych tymi metodami, z wykorzystaniem promieniowania widzialnego, zawiera się – dla średniego kwadratowego odchylenia wysokości nierówności powierzchni σ – od 0,1 do 10 nm. Jeżeli stosowane jest promieniowanie podczerwone, o długości fali $\lambda = 10,6 \mu\text{m}$, zakres ten wynosi od 2 do około 200 nm i może być zwiększony. W USA opracowano normę, określającą procedurę i warunki pomiarów nierówności powierzchni metodami różniczkowymi [1].

Idea metod różniczkowych – polegająca na ocenie nierówności powierzchni na podstawie pomiaru i analizy kąтового rozkładu natężenia światła rozproszonego – wykorzystana została także do pomiarów powierzchni o większych nierównościach. Ten szczególny rodzaj metod różniczkowych na ogół nie jest oparty na falowym modelu światła. Opiera się on najczęściej na przybliżeniu promieniowym oraz na zasadach optyki geometrycznej. W tym przypadku nierówna powierzchnia przedmiotu jest traktowana jako zbiór odpowiednio zorientowanych, płaskich miniaturowych zwierciadeł, tak jak przedstawiono to w p. 2.1.2. Indykatrysa rozproszenia może być interpretowana w tym modelu jako wykres proporcjonalny do gęstości prawdopodobieństwa kątów nachylenia mikrozwierciadeł tworzących powierzchnię. Nierówności powierzchni ocenia się na podstawie wybranych parametrów rozkładu natężenia światła rozproszonego, takich jak szerokość połówkowa indykatrysy rozproszenia, wariancja kątów rozproszenia itp. Rozkład natężenia światła rozproszonego wyzna-

czany jest na ogół za pomocą detektorów fotoelektrycznych ze sprzężeniem ładunkowym, czyli liniałów lub matrycowych kamer CCD, połączonych z komputerem. Zwiększa to szybkość pomiaru i nie wymaga ruchu detektora względem powierzchni mierzonej.

Takie metody znajdują zastosowanie w pomiarach aktywnych nierówności powierzchni części maszyn, np. w procesach szlifowania, polerowania itp. Głównym zadaniem tych metod jest stwierdzenie, czy powierzchnia została prawidłowo obrabiona. Często system kontrolno-pomiarowy umożliwia dokonywanie pomiaru podczas ruchu przedmiotu, np. w czasie obróbki lub tuż po jej zakończeniu. Zakres wysokości nierówności mierzonych tymi metodami wynosi od 10 do około 1000 nm. Przykładem takiego systemu jest układ opisany w pracy [36], a także przedstawiony w pracy [68] przyrząd Lasercheck, opracowany w firmie Optical Dimensions LLC (USA). Lasercheck przeznaczony jest do aktywnych pomiarów nierówności powierzchni obrabianych w zakresie parametru R_a od 5 do 500 nm. Kąt padania światła na powierzchnię wynosi 75° . Światło rozproszone mierzone jest liniałem detektorów fotoelektrycznych oraz dwoma oddzielnymi detektorami.

3.4. METODY WYKORZYSTUJĄCE OPTYCZNE PRZEKSZTAŁCENIE FOURIERA

Specyficzną grupę metod oceny nierówności powierzchni na podstawie kąowego rozkładu natężenia światła rozproszonego stanowią metody oparte na wykorzystaniu optycznego przekształcenia Fouriera. Metody te niekiedy zalicza się do metod dyfrakcyjnych, gdyż u ich podstaw leży zjawisko dyfrakcji fal świetlnych na nierównościach powierzchni. W teorii odwzorowania optycznego i w teorii dyfrakcji znaczącą rolę odgrywają przekształcenia całkowe, a zwłaszcza przekształcenie Fouriera [25], [24], [29]. Podstawy teoretyczne tego przekształcenia, szeroko stosowanego w wielu dziedzinach nauki i techniki, wynikają z teorii szeregów trygonometrycznych. Przekształcenie Fouriera dwuwymiarowych sygnałów optycznych może być realizowane za pomocą stosunkowo prostych układów, takich np. jak soczewki. Umożliwia ono częstotliwościową analizę sygnałów w liniowych i przestrzennie niezmienniczych układach optycznych.

Rozpatrywane dalej metody oceny nierówności powierzchni, wykorzystujące optyczne przekształcenie Fouriera, podobnie jak metody różniczkowe, umożliwiają wyznaczenie funkcji gęstości widmowej mocy, a niekiedy także funkcji autokorelacji nierówności powierzchni. Główna idea tych metod polega na oświetleniu powierzchni – lub jej przezroczystej repliki – falą płaską światła spójnego, o stałej amplitudzie, i zarejestrowaniu obrazu dyfrakcyjnego w obszarze Fraunhofera. Jeśli reflektancja całkowita powierzchni jest we wszystkich jej punktach jednakowa, to nierówności powierzchni powodują jedynie modulację fazową fali padającej. Fala świetlna, zmodulowana fazowo przez nierówności powierzchni, może być traktowana jako superpozycja fal płaskich o odpowiednich kierunkach propagacji. Zbiór takich fal nazywany

jest widmem kątowym fal płaskich [25]. Jeżeli głębokość modulacji fazowej jest niewielka – a występuje to wówczas, gdy wysokość nierówności powierzchni jest znacznie mniejsza od długości fali światła – to widmo kątowe fal płaskich daje w obszarze Fraunhofera obraz proporcjonalny do funkcji gęstości widmowej mocy nierówności powierzchni.

W metodach oceny nierówności powierzchni wykorzystujących optyczne przekształcenie Fouriera pomiarów, rejestracji i analizy natężenia światła w obrazie dyfrakcyjnym dokonuje się – w odróżnieniu od wielu metod różniczkowych – tylko w niewielkim zakresie kątów rozproszenia. Do pomiarów rozkładu natężenia światła w polu dyfrakcyjnym stosowane są często wieloelementowe, liniowe lub matrycowe detektory fotoelektryczne CCD. Unika się wtedy mechanicznego skanowania pola dyfrakcyjnego ruchomym detektorem fotoelektrycznym, co występuje w niektórych metodach różniczkowych.

Idea pomiarów nierówności powierzchni z wykorzystaniem optycznego przekształcenia Fouriera zrodziła się kilka lat po wynalezieniu lasera. Analiza zjawiska dyfrakcji światła spójnego, ugiętego na obiektach amplitudowo-fazowych, doprowadziła do spostrzeżenia, że zjawisko to może być opisane za pomocą teorii przekształcenia Fouriera. W pracy [25] pokazano, że w obszarze Fraunhofera, którego odległość \bar{r} od płaszczyzny obiektu musi spełniać warunek

$$\bar{r} \gg \frac{k(x_s^2 + y_s^2)_{\max}}{2}, \quad (95)$$

gdzie $k = 2\pi/\lambda$ oznacza liczbę falową, zespolona amplituda fali świetlnej jest proporcjonalna do dwuwymiarowej transformaty Fouriera zespolonej amplitudy światła, w płaszczyźnie P_s położonej bezpośrednio za obiektem uginającym. Zależność tę – w przypadku prostopadłego padania światła spójnego na obiekt uginający, co pokazano na rysunku 12 – wyraża równanie

$$E(x_p, y_p) = \frac{\exp\left[\frac{ik}{2\bar{r}}(x_p^2 + y_p^2)\right]}{i\lambda\bar{r}\exp(-ik\bar{r})} \iint_{-\infty}^{\infty} E(x_s, y_s) \left[\frac{-ik}{\bar{r}}(x_p x_s + y_p y_s)\right] dx_s dy_s, \quad (96)$$

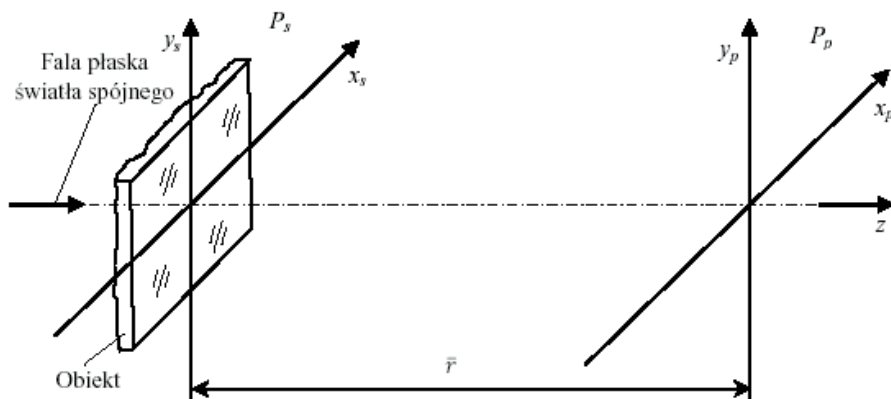
gdzie: $E(x_p, y_p)$ – zespolona amplituda światła w płaszczyźnie P_p , i – jednostka urojona, k – liczba falowa, \bar{r} – odległość płaszczyzny P_p od płaszczyzny P_s , x_p, y_p – współrzędne prostokątne w płaszczyźnie P_p , λ – długość fali światła, $E(x_s, y_s)$ – zespolona amplituda światła w płaszczyźnie P_s , x_s, y_s – kartezjańskie współrzędne prostokątne w płaszczyźnie P_s .

Wyrażenie (96) jest podobne do równania (28), ale zespolone amplitudy światła $E(x_s, y_s)$ i $E(\mathbf{r}_s)$ występujące w obu równaniach są nieco inne. Podobieństwo obu równań można zauważyć uwzględniając, że wektor \mathbf{r} określa położenie punktu P o współ-

rzędnych x_p, y_p, \bar{r} , natomiast wektor \mathbf{r}_s wyznacza punkt na powierzchni nierównej. Główna różnica między nimi polega na tym, że w równaniu (96) powierzchnia całkowania jest płaszczyzną, a w równaniu (28) jest to powierzchnia nierówna S_s . Przejście od całkowania po powierzchni S_s do całkowania po jej rzucie S_0 na płaszczyznę $z = 0$ następuje zgodnie z zależnością

$$dx dy = dS_0 = \frac{n_z}{|\mathbf{n}|} d(S_s) = \frac{n_z}{|\mathbf{n}|} dS, \quad (97)$$

gdzie n_z jest długością rzutu jednostkowego wektora \mathbf{n} , prostopadłego do powierzchni nierównej, na oś $0z$ układu współrzędnych $0xyz$. Ograniczenie obszaru całkowania w równaniu (28) do powierzchni S_s jest mniej istotne, gdyż funkcja $E(\mathbf{r}_s)$, jak wynika to z warunków brzegowych Kirchhoffa, jest na powierzchni S_c równa zero. Można więc obszar całkowania w tym równaniu rozciągnąć od $-\infty$ do $+\infty$.



Rys. 12. Prostopadłe padanie płaskiej fali światła na obiekt uginający
Fig. 12. Normal incidence of the plane light wave on the object

Bezsoczewkowy układ, pokazany na rysunku 12, jest rzadko stosowany do oceny nierówności powierzchni metodami optycznego przekształcenia Fouriera ze względu na to, że spełnienie warunku (95) wymaga, aby płaszczyzna obiektu P_s i płaszczyzna analizy pola dyfrakcyjnego P_p były oddalone od siebie nawet o kilkadziesiąt metrów. W celu zmniejszenia gabarytów stosuje się specjalne soczewki, których płaszczyzny ogniskowe, odpowiadające płaszczyznom P_s i P_p , są oddalone od siebie o kilkadziesiąt centymetrów. W układach tych rolę obiektu uginającego światło może spełniać nierówna powierzchnia przedmiotu lub przezroczysta replika powierzchni, odwzorowująca nierówności [41].

Jeżeli obiekt, oświetlony spójną falą płaską o jednostkowej amplitudzie, jest umieszczony w przedmiotowej płaszczyźnie ogniskowej soczewki o równych ogni-

skowych, to zespoloną amplitudę światła w obrazowej płaszczyźnie ogniskowej soczewki określa równanie

$$E(x_p, y_p) = \frac{1}{i\lambda f} \mathfrak{F}[E(x_s, y_s)], \quad (98)$$

natężenie światła zaś wynosi odpowiednio

$$I(x_p, y_p) = \frac{1}{(\lambda f)^2} |\mathfrak{F}[E(x_s, y_s)]|^2, \quad (99)$$

gdzie: $E(x_p, y_p)$ – zespolona amplituda światła w obrazowej płaszczyźnie ogniskowej soczewki, i – jednostka urojona, f – długość ogniskowej soczewki, λ – długość fali światła, x_p, y_p – współrzędne prostokątne w obrazowej płaszczyźnie ogniskowej soczewki, $\mathfrak{F}[\]$ – symbol oznaczający operację przekształcenia Fouriera, $E(x_s, y_s)$ – zespolona amplituda światła w przedmiotowej płaszczyźnie ogniskowej soczewki, x_s, y_s – współrzędne prostokątne w przedmiotowej płaszczyźnie ogniskowej soczewki, $I(x_p, y_p)$ – natężenie światła w obrazowej płaszczyźnie ogniskowej soczewki.

Z równania (99) wynika, że natężenie światła w obrazowej płaszczyźnie ogniskowej soczewki jest proporcjonalne do gęstości widmowej mocy zespolonej amplitudy światła w przedmiotowej płaszczyźnie ogniskowej. Jeżeli zatem zespoloną amplitudę światła zmodulowanego przez obiekt uczynić proporcjonalną do funkcji $z = f(x_s, y_s)$, opisującej nierówności powierzchni, to pomiar natężenia światła w obrazowej płaszczyźnie ogniskowej soczewki pozwoli na wyznaczenie ważnej charakterystyki – gęstości widmowej mocy nierówności powierzchni. Wskazano na to między innymi w pracach [65], [66].

Jeśli światło odbija się od powierzchni, to, pomijając występujący przy tym skok fazy, zmiany fazy fali światła są równe

$$\Delta\varphi(x_s, y_s) = \frac{4\pi}{\lambda} f(x_s, y_s), \quad (100)$$

gdzie: $\Delta\varphi(x_s, y_s)$ – funkcja opisująca zmiany fazy fali światła, λ – długość fali światła, $z = f(x_s, y_s)$ – funkcja opisująca nierówności powierzchni. Zakłada się przy tym jednorodność powierzchni odbijającej. Znaczy to, iż amplitudowy współczynnik odbicia ma stałą wartość we wszystkich punktach powierzchni. Odbicie światła od takiej powierzchni powoduje tylko fazową modulację fali świetlnej, nie modulując jej amplitudy. Zespolona amplituda światła $E(x_s, y_s)$, występująca w równaniach (97), (98) i (99), określona w płaszczyźnie przedmiotu oświetlonego falą o jednostkowej amplitudzie, jest zatem równa

$$E(x_s, y_s) = \rho_A \exp [i\Delta\varphi(x_s, y_s)], \quad (101)$$

przy czym

$$\rho_A = \rho_A(x_s, y_s) = \text{const}, \quad (102)$$

gdzie: $E(x_s, y_s)$ – zespolona amplituda fali światła w płaszczyźnie przedmiotu, ρ_A – amplitudowy współczynnik odbicia światła, $\Delta\varphi(x_s, y_s)$ – funkcja opisująca zmiany fazy fali świetlnej wywołane nierównościami powierzchni.

Zmiany fazy fali świetlnej, jak wynika z równania (100), są proporcjonalne do funkcji $z = f(x_s, y_s)$, opisującej nierówności powierzchni. Jednakże równanie (101) pokazuje, że zespolona amplituda $E(x_s, y_s)$ zmienia się nieproporcjonalnie do zmian fazy fali świetlnej. W aspekcie metrologicznym jest to niekorzystne, gdyż utrudnia uzyskanie jednoznacznej relacji między zespoloną amplitudą $E(x_s, y_s)$ i funkcją $z = f(x_s, y_s)$, opisującą nierówności powierzchni. Problem ten nasila się wraz ze wzrostem zakresu zmian wysokości nierówności.

Innymi słowy z równania (101) wynika, że w ogólnym przypadku zespolona amplituda światła $E(x_s, y_s)$ nie jest proporcjonalna do wysokości nierówności badanej powierzchni $z = f(x_s, y_s)$. Jest to prostą konsekwencją tego, że odbicie światła od powierzchni nierównej, wprowadza modulację fazową fali odbitej. Brak proporcjonalności zespolonej amplitudy światła $E(x_s, y_s)$ do wysokości nierówności $z = f(x_s, y_s)$ stawia pod znakiem zapytania możliwość pomiarów nierówności z wykorzystaniem optycznego przekształcenia Fouriera. Warto jednak zauważyć, iż w sytuacji, gdy zakres zmian wysokości nierówności powierzchni jest bardzo mały w porównaniu z długością fali światła, zależność między wysokością nierówności powierzchni $z = f(x_s, y_s)$ a zespoloną amplitudą światła $E(x_s, y_s)$ można uznać w przybliżeniu za liniową. W tym przypadku pomiary nierówności powierzchni z wykorzystaniem optycznego przekształcenia Fouriera mają pełne uzasadnienie.

Jak już o tym wspomniano, inną ważną zaletą pomiarów nierówności powierzchni, z wykorzystaniem optycznego przekształcenia Fouriera, jest nieznaczny wpływ ruchu powierzchni mierzonej na postać i położenie obrazu dyfrakcyjnego. Ruch powierzchni mierzonej w płaszczyźnie $0x_s y_s$ praktycznie nie wpływa na położenie obrazu dyfrakcyjnego w płaszczyźnie $0x_p y_p$. Jest to szczególnie korzystne wtedy, gdy optyczne przekształcenie Fouriera stosowane jest do szybkiej oceny nierówności podczas ruchu powierzchni, np. w procesie jej wytwarzania. W procesach obróbki powierzchni często jednak pojawiają się czynniki mogące zakłócać pomiary optyczne. Należą do nich wióry, ciecze chłodziwo-smarujące, pyły, ścierniwo itp. Wpływ tych czynników na przebieg pomiaru powinien być eliminowany lub ograniczany.

Metody różniczkowe i metody wykorzystujące optyczne przekształcenie Fouriera mają jeszcze jedną, dodatkową zaletę. Określają one, na podstawie obrazu dyfrakcyjnego, kierunkowość struktury geometrycznej powierzchni. W tym celu poddaje się analizie obraz dyfrakcyjny, otrzymany w wyniku rozproszenia światła przez powierzchnię nierówną. Podobny efekt uzyskuje się, analizując kwadrat modułu transformaty Fouriera mikroskopowego obrazu powierzchni.

Określenie kierunkowości struktury geometrycznej powierzchni ma duże znaczenie w zautomatyzowanych systemach produkcyjnych. W systemach tych stosuje się różnorodne układy, zwłaszcza optyczne, służące do rozpoznawania i identyfikacji położenia powierzchni przedmiotów na podstawie transformat Fouriera ich obrazów powierzchni lub obrazów dyfrakcyjnych.

W przypadku powierzchni bardzo gładkich analiza obrazów dyfrakcyjnych umożliwia badanie mikroskopijnych defektów powierzchni. Układ pomiarowy, wykorzystujący optyczne przekształcenie Fouriera, wykrywający oraz mierzący takie defekty i zanieczyszczenia powierzchni podłoży krzemowych układów scalonych, przedstawiono w pracach [43], [44]. Okresowe nierówności występujące na powierzchni badanej powodują pojawienie się wyraźnych ekstremów w transformatach Fouriera mikroskopowych obrazów powierzchni i w obrazach dyfrakcyjnych światła rozproszonego przez takie powierzchnie.

3.5. UWAGI KOŃCOWE

W poszukiwaniu szybkich i dokładnych metod pomiaru nierówności powierzchni dużo uwagi poświęca się pomiarom optycznym, w tym także metodom rozpraszania światła. Niestety zjawisko rozpraszania światła jest bardzo złożone. Co prawda opracowane zostały matematyczne modele zjawiska rozpraszania światła, lecz są to modele albo bardzo skomplikowane, które trudno zaadaptować do zastosowań praktycznych, albo na tyle przybliżone, że opisują rozpraszanie światła w dużym uproszczeniu. Metody pomiaru nierówności, które rozpatrzono w tej pracy, opierają się najczęściej na znacznie uproszczonych modelach zjawiska rozpraszania światła. Wyjątek stanowią tu niektóre metody różniczkowe i metody oparte na optycznym przekształceniu Fouriera.

Wielką zaletą optycznych pomiarów nierówności powierzchni, w tym także metod skaterometrycznych, jest bezstykowy sposób pomiaru. Umożliwia on nieinwazyjną, nieniszczącą ocenę powierzchni. Jest to ważne w przypadku powierzchni, które są mało odporne na zanieczyszczenia, podatne na odkształcenia i wrażliwe na różnorodne inne oddziaływania fizyczne bądź chemiczne, pojawiające się podczas styku z inną powierzchnią. W pomiarach nierówności metodami rozpraszania światła korzystne jest również to, iż pomiarem objęty jest pewien obszar powierzchni przedmiotu mierzonego, a nie pojedynczy profil. Umożliwia to bowiem ocenę przestrzennych właściwości i ukształtowania nierówności powierzchni.

Pomiary nierówności metodami rozpraszania światła mają także inne zalety, takie jak możliwość oceny powierzchni podczas ruchu przedmiotu mierzonego oraz krótki czas pomiaru. Sprzyja to zastosowaniu metod rozpraszania światła w pomiarach aktywnych nierówności, a także w kontroli i nadzorowaniu procesów obróbki powierzchni. Pewnym ograniczeniem omawianych metod w takich zastosowaniach jest ich wrażliwość na zakłócenia, w postaci pyłów i innych zanieczyszczeń występują-

cych w strefie obróbki. Wymagane jest wtedy, np. oczyszczenie powierzchni mierzonoj i stworzenie takich warunków pomiaru, które wyeliminują lub ograniczą wpływ czynników zakłócających.

Z przeprowadzonej analizy wynika, że pomiary nierówności powierzchni metodami rozpraszania światła opierają się często na uproszczonych modelach tego zjawiska. W literaturze brak jest wnikliwej analizy teoretycznej zjawiska rozpraszania przez powierzchnie nierówne, w aspekcie jego zastosowania w pomiarach nierówności powierzchni. Istniejące modele teoretyczne odnoszą się na ogół do tzw. zadania prostego, które polega na określeniu, jak rozproszy się światło na danej powierzchni nierównej. Modele te mają wiele ograniczeń i na ich podstawie trudno jest określić przestrzenną postać rozkładu natężenia światła rozproszonego. Podczas rozwiązywania zadania prostego ważnym zagadnieniem jest wybór właściwego modelu teoretycznego opisującego rozpraszanie światła. Wybór taki dokonywany jest w zależności od rodzaju i zakresu wysokości nierówności powierzchni oraz od sposobu jej oświetlenia. Dodatkowe komplikacje wynikają z powodu pojawiania się podczas odbicia światła od powierzchni nierównej tzw. samozacieniania powierzchni i odbić wielokrotnych. Warto jednak zauważyć, że coraz częściej pojawiają się propozycje wykorzystania zjawiska samozacieniania powierzchni w pomiarach [34], [28]. Okazuje się, że zjawisko to może być skutecznie zastosowane od oceny nierówności powierzchni o znacznych wysokościach, w tym takich, które nie mogą być zmierzone innymi metodami skaterometrycznymi. Przykładem tego są pomiary nierówności narzędzi ściernych, które nie tylko charakteryzują się znacznymi wysokościami nierówności, ale też dużą niejednorodnością materiału poszczególnych składników tworzących takie narzędzia.

Racjonalne stosowanie metod rozpraszania światła, oprócz analizy zadania prostego, wymaga rozwiązania bardziej złożonego problemu – tzw. zadania odwrotnego. Polega ono na wyznaczeniu charakterystyk nierówności powierzchni na podstawie pomiarów rozkładu natężenia światła rozproszonego. Zadanie to sprowadza się zatem do odzyskania, ze zmierzonych wartości natężenia światła rozproszonego, informacji o fazie fali świetlnej i wysokości nierówności w kolejnych punktach powierzchni. W ogólnym przypadku nie zostało ono dotychczas rozwiązane. Podstawowym utrudnieniem w rozwiązaniu zadania odwrotnego jest to, że relacje między wielkościami charakteryzującymi rozkład natężenia światła rozproszonego a funkcją opisującą nierówności są często złożone i niejednoznaczne. Jeden z wariantów zadania odwrotnego bardziej szczegółowo rozpatrzono w pracach [41], [35], [39].

4. PODSUMOWANIE I WNIOSKI

Rozważania przedstawione w niniejszej pracy pokazują, że od kilkudziesięciu lat zjawisko rozpraszania światła jest wykorzystywane do pomiarów i oceny nierówności powierzchni. W wielu ośrodkach naukowych prowadzone są badania zmierzające do

lepszego poznania i opisu tego zjawiska. Doskonalone są istniejące metody pomiaru nierówności, oparte na analizie światła rozproszonego. Opracowywane są nowe metody i przyrządy pomiarowe. Zainteresowanie tymi metodami pomiaru wynika z ich wielu zalet, są to bowiem metody bezstykowe, szybkie, a także bardzo czułe, co wykorzystuje się w pomiarach najbardziej gładkich powierzchni. Ponadto metodami rozpraszania światła można ocenić stan nierówności na pewnym obszarze, a nie tylko w wybranym przekroju badanej powierzchni.

Niestety metody te mają także wiele ograniczeń. Jednym z głównych czynników, które ograniczają wykorzystanie metod rozpraszania światła w pomiarach nierówności powierzchni, jest bardzo złożona postać sygnału pomiarowego. Fala świetlna, będąca nośnikiem tego sygnału, jest zmiennym polem elektromagnetycznym o skomplikowanej strukturze. W wyniku odbicia od powierzchni nierównej, może ona doznawać zmian amplitudy, fazy, częstotliwości, stanu polaryzacji i kierunku propagacji. Niektóre parametry fali mogą też zmieniać się w procesie jej rozprzestrzeniania. Dodatkowym utrudnieniem, często występującym w praktyce, jest konieczność analizy przestrzennego rozkładu parametrów fali odbitej od powierzchni. Zmierzenie tak złożonego sygnału pomiarowego i dokonanie, na jego podstawie, prawidłowej oceny nierówności powierzchni nie jest zadaniem łatwym, zwłaszcza dlatego, że relacje między sygnałem pomiarowym a nierównościami powierzchni ustalane są w wielu założeniach upraszczających. Dotychczasowy rozwój teorii rozpraszania światła, technik optycznych, elektronicznych i fotonicznych oraz skomputeryzowanych metod analizy sygnałów pomiarowych daje jednak nadzieję na stopniowe pokonywanie wymienionych trudności.

Na podstawie analiz przeprowadzonych w niniejszej pracy, dotyczących zastosowania w pomiarach nierówności powierzchni zjawiska rozpraszania światła, można sformułować następujące wnioski ogólne:

1. Podstawowym modelem teoretycznym, stosowanym do analizy zjawiska rozpraszania światła na powierzchniach nierównych, jest teoria dyfrakcji. Dyfrakcyjny model rozpraszania światła na nierównościach powierzchni ma najczęściej postać teorii zaburzeń lub teorii płaszczyzny stycznej, nazywanej także teorią Kirchhoffa. Wykorzystanie każdej z tych teorii wymaga spełnienia przez nierówności powierzchni odpowiednich założeń, szerzej przedyskutowanych w pracy. Teoria płaszczyzny stycznej dobrze opisuje rozpraszanie światła na powierzchniach, których wysokość nierówności i lokalne promienie krzywizny są większe od długości fali światła. Natomiast teoria zaburzeń ma zastosowanie do analizy rozpraszania światła na powierzchniach o wysokościach nierówności małych w porównaniu z długością fali światła, których gradient wysokości jest także niewielki.

2. Analiza zjawiska rozpraszania światła na powierzchniach, których wysokość nierówności jest znacznie mniejsza od długości fali światła, wskazuje, że w rezultacie pomiarów rozkładu natężenia światła rozproszonego w obszarze Fraunhofera można ocenić funkcję gęstości widmowej mocy nierówności powierzchni. W przy-

padku powierzchni okresowej dyfrakcyjny model rozpraszania światła umożliwia także wyznaczenie przestrzennej długości fali nierówności powierzchni, na podstawie pomiarów kąta rozproszenia, który odpowiada danemu rzędowi dyfrakcyjnemu. Pomiarów kąta rozproszenia warto jest dokonywać w wyższych rzędach dyfrakcyjnych, gdyż wtedy błędy pomiaru przestrzennej długości fali stają się mniejsze. Szczególną okolicznością jest to, iż dokładniej można mierzyć nierówności z krótszymi, przestrzennymi długościami fali. W przypadku powierzchni sinusoidalnej, dyfrakcyjny model rozpraszania światła umożliwia także wyznaczenie amplitudy nierówności powierzchni. W tym celu konieczny jest pomiar względnego natężenia światła w maksimum dyfrakcyjnym pierwszego rzędu. Wymienione właściwości metod rozpraszania światła sprzyjają ich stosowaniu w pomiarach nierówności powierzchni na ogół o niewielkich wysokościach i małych przestrzennych długościach fali. Istnieją jednak wyjątki od tej reguły. Na przykład, wykorzystanie zjawiska samozacieniania powierzchni pozwala w pewnych warunkach znacznie rozszerzyć zakres wysokości mierzonych nierówności powierzchni.

3. Rozpraszaniu światła na powierzchniach nierównych mogą towarzyszyć takie zjawiska jak samozacienianie powierzchni i odbicia wielokrotne. Jeśli idzie o pomiary nierówności powierzchni metodami rozpraszania światła, są to na ogół zjawiska niepożądane. Analiza przedstawiona w pracy wskazuje, że oświetlenie powierzchni w taki sposób, aby kąty padania były równe lub bliskie zeru, sprzyja ograniczeniu wpływu samozacienień powierzchni i odbić wielokrotnych na wynikowy rozkład natężenia światła rozproszonego. Wpływ niewielkich samozacienień powierzchni można uwzględnić, wprowadzając do analizy rozpraszania światła statystyczną funkcję zacienienia powierzchni.

4. Przy danym sposobie oświetlenia powierzchni, rozkład natężenia światła rozproszonego zależy od ukształtowania powierzchni nominalnej oraz od wymiarów, położenia i kształtu nierówności. Rozważania przeprowadzone w pracy pokazują, że zastosowanie zasad optyki geometrycznej umożliwia przybliżone określenie przestrzennego kształtu tego rozkładu. Przestrzenna postać rozkładu natężenia światła rozproszonego może zatem stanowić podstawę do ogólnego wnioskowania o ukształtowaniu powierzchni nominalnej, jej usytuowaniu względem fali oświetlającej powierzchnię, a także o położeniu, kształcie i wymiarach nierówności powierzchni.

Mimo iż przedstawione wnioski ogólne nie zawierają gotowych rozwiązań niektórych problemów występujących w pomiarach nierówności powierzchni metodami rozpraszania światła, to wraz z wieloma analizami zawartymi w niniejszej pracy mogą być one pomocne w teoretycznej analizie, praktycznym wykorzystaniu i dalszym rozwoju tych metod. Wiele zagadnień poruszonych w pracy – takich np. jak teoretyczne podstawy rozpraszania światła przez powierzchnie nierówne, samozacienianie powierzchni, odbicia wielokrotne – nie znajduje pełnego wyjaśnienia w obecnym stanie wiedzy i wymaga prowadzenia dalszych badań.

Przedstawione w pracy rozważania dotyczyły rozpraszania światła na powierzchniach nierównych. Jednak proponowane tu sposoby analizy, jak i otrzymane wnioski, mogą być z uwzględnieniem pewnych warunków przeniesione na inny zakres fal elektromagnetycznych. Wymaga to za każdym razem szczegółowej analizy i sprawdzenia, czy w innym zakresie fal elektromagnetycznych nie wystąpią specyficzne uwarunkowania i ograniczenia.

LITERATURA

- [1] ASTM E2387 – 05, *Standard Practice for Goniometric Optical Scatter Measurements*. DOI: 10.1520/E2387-05.
- [2] ASTM F1048-87(1999), *Standard Test Method for Measuring the Effective Surface Roughness of Optical Components by Total Integrated Scattering*. DOI: 10.1520/F1048-87R99.
- [3] BASS F.G., FUKS I.M., *Wave Scattering from Statistically Rough Surfaces*, Oxford, Pergamon Press, 1979.
- [4] BAUMGART J.W., TRUCKENBRODT H., *Scatterometry of honed surfaces*, *Optical Engineering*, Vol. 37, No. 5, 1998, 1435–1441.
- [5] BECKMANN P., *Scattering of light by rough surfaces*, *Progress in Optics*, Vol. VI, Wolf E. (ed.), Amsterdam, North-Holland, 1967, 53–69.
- [6] BECKMANN P., SPIZZICHINO A., *The Scattering of Electromagnetic Waves from Rough Surfaces*, Oxford, Pergamon Press, 1963.
- [7] BENNETT H.E., PORTEUS J.O., *Relation Between Surface Roughness and Specular Reflectance at Normal Incidence*, *Journal of the Optical Society of America*, Vol. 51, No. 2, 1961, 123–129.
- [8] BENNETT J.M., MATTSSON L., *Introduction to Surface Roughness and Scattering*, Washington, D.C., Optical Society of America, 2nd ed., 1999.
- [9] BORKOWSKI J., MROCZKA J., *Application of the discrete Fourier transform linear interpolation method in the measurement of volume scattering function at small angle*, *Optical Engineering*, Vol. 39, No. 6, 2000, 1576–1586.
- [10] BORN M., WOLF E., *Principles of Optics*, Oxford, Pergamon Press, 1968.
- [11] БРЕХОВСКИХ Л.М., *Дифракция электромагнитных волн на неровной поверхности*, Доклады АН СССР, Т. 81, No. 6, 1951, 1023–1026.
- [12] BRODMANN R., THURN G., *Roughness measurement of ground, turned and shot-peened surfaces by the light scattering method*, *Wear*, Vol. 109, No. 3, 1986, 1–13.
- [13] BURAKOWSKI T., *Areologia. Powstanie i rozwój*, Radom, Instytut Technologii Eksploatacji – PTB 2007.
- [14] BURAKOWSKI T., WIERZCHOŃ T., *Inżynieria powierzchni metali*, PWN, Warszawa 1995.
- [15] CHABROS W., STAROŃSKI L., *Skaterometria nierówności powierzchniowych*, Biuletyn Informacyjny – Optyka, nr 1–2, Instytut Optyki Stosowanej, Warszawa 1993, 9–24.
- [16] CHURCH E.L., JENKINSON H.A., ZAVADA J.M., *Measurement of the Finish of Diamond-Turned Metal Surfaces by Differential Light Scattering*, *Optical Engineering*, Vol. 16, No. 4, 1977, 360–374.
- [17] CHURCH E.L., JENKINSON H.A., ZAVADA J.M., *Relationship between Surface Scattering and Microtopographic Features*, *Optical Engineering*, Vol. 18, No. 2, 1979, 125–136.
- [18] CHURCH E.L., ZAVADA J.M., *Residual surface roughness of diamond-turned optics*, *Applied Optics*, Vol. 14, No. 8, 1975, 1788–1795.

- [19] CZERWIŃSKI M., MROCZKA J., GIRASOLE T., GOUESBET G., GREHAN G., *Light-transmittance predictions under multiple-light-scattering conditions. I. Direct problem: hybrid-method approximation*, Applied Optics, Vol. 40, No. 9, 2001, 1514–1524.
- [20] CZERWIŃSKI M., MROCZKA J., GIRASOLE T., GOUESBET G., GREHAN G., *Light-transmittance predictions under multiple-light-scattering conditions. II. Inverse problem: Particle size determination*, Applied Optics, Vol. 40, No. 9, 2001, 1525–1531.
- [21] DOBOSZ M., *Optyczne metody pomiaru chropowatości – metody rozpraszania światła*. Mechanik, nr 10, 1984, 529–532.
- [22] DOBOSZ M., RATAJCZYK E., *Optyczne metody pomiaru chropowatości. Metody kontrastu*. Mechanik, nr 3, 1985, 151–153.
- [23] DUPARRÉ A., *Light scattering techniques for the inspection of microcomponents and microstructures*. [W:] *Optical Inspection of Microsystems*, W. Osten (ed.), Optical Science and Engineering, Vol. 109, Boca Raton: Taylor & Francis, 2007, 103–119.
- [24] GNIADK K., *Optyczne przetwarzanie informacji*, PWN, Warszawa 1992.
- [25] GOODMAN J.W., *Introduction to Fourier Optics*, New York – London, McGraw-Hill Book Co., 1968.
- [26] GUSZKOWSKI T., MROCZKA J., *Rozwiązanie problemu odwrotnego w turbidymetrii z wykorzystaniem sieci neuronowej*, Pomiary, Automatyka, Kontrola, nr 9bis, 2007, s. 239–241.
- [27] HOU H., YI K., SHANG S., SHAO J., FAN Z., *Measurements of light scattering from glass substrates by total integrated scattering*, Applied Optics, Vol. 44, 2005, 6163–6166.
- [28] IAQUINTA J., ŁUKIANOWICZ C., *Inferring texture from shadows*, Measurement Science Review. Vol. 6, Sec. 3, No. 4, 2006, 21–24.
- [29] JÓZWICKI R., *Teoria odwzorowania optycznego*, PWN, Warszawa 1988.
- [30] KIELICH S., *Molekularna optyka nieliniowa*, PWN, Warszawa–Poznań 1977.
- [31] ЛАХИЦЕВ Г.С., *Онмука*, Наука, Москва, 1976.
- [32] LINDSTRAND M., *Gloss: measurement, characterization and visualization – in the light of visual evaluation*, Linköping Studies in Science and Technology, 2002, Thesis No. 975.
- [33] LONARDO P.M., BRUZZONE A.A., LONARDO A.M., *Analysis of Machined Surfaces through Diffraction Patterns and Neural Networks*, Annals of the CIRP, Vol. 44, No.1, 1995, 509–512.
- [34] ŁUKIANOWICZ C., IAQUINTA J., *Surface topography assessment of abrasive sheets by light scattering and self-shadowing methods*, Archives of Civil and Mechanical Engineering, Vol. 2, 2005, 71–84.
- [35] ŁUKIANOWICZ C., *Inverse problem in scatterometry of rough surfaces*, Optica Applicata, No. 2-3, 2003, 316–327.
- [36] ŁUKIANOWICZ C., KARPIŃSKI T., *Optical system for measurement of surface form and roughness*, Measurement Science Review, Vol. 1, No. 1, 2001, 151–154.
- [37] ŁUKIANOWICZ C., KARPIŃSKI T., *Scatterometry of Ground Surfaces*, Measurement Science Review, Vol. 3, No. 3, 2003, 21–24.
- [38] ŁUKIANOWICZ C., KARPIŃSKI T., *Scatterometry of machined surfaces*. [W:] Proceedings XIV IMEKO World Congress, Vol. VIII, Vienna 2000, 181–185.
- [39] ŁUKIANOWICZ C., *Modelling of direct and inverse problem in light scattering by rough surfaces*, Optica Applicata, nr 2–3, 2003, 329–336.
- [40] ŁUKIANOWICZ C., *Surface roughness assessment by light scattering*, Metrology and Measurement Systems, Vol. IX, No. 3, 2002, 247–260.
- [41] ŁUKIANOWICZ CZ., *Podstawy pomiarów nierówności powierzchni metodami rozpraszania światła*, Wydawnictwo Uczelniane Politechniki Koszalińskiej, Koszalin 2001.
- [42] MARADUDIN A.A., MÉNDEZ E.R., LESKOVA T.A., *Designer Surfaces*, Amsterdam, Elsevier, 2008.

- [43] MIYOSHI T., TAKAYA Y., SAITO K., *Micromachined Profile Measurement by Means of Optical Inverse Scattering Phase Method*, Annals of the CIRP, Vol. 45, No. 1, 1996, 497–500.
- [44] MIYOSHI T., TAKAYA Y., SAITO K., *Nanometer Measurement of Silicon Wafer Surface Texture Based on Fraunhofer Diffraction Pattern*, Annals of the CIRP, Vol. 44, No. 1, 1995, 489–492.
- [45] MROCZKA J., *Metrologiczne problemy wykorzystania światła rozproszonego do badania rozkładu wielkości cząstek w roztworach dyspersyjnych*, Metrologia i Systemy Pomiarowe, Monografia, 4, 1990.
- [46] MROCZKA J., WYSOCZAŃSKI D., *Plane-wave and Gaussian beam scattering on an infinite cylinder*, Optical Engineering, Vol. 39, No. 3, 2000, 763–770.
- [47] NICODEMUS F.E., *Reflectance nomenclature and directional reflectance and emissivity*, Applied Optics, Vol. 9, June, 1970, 1474–1475.
- [48] NOWICKI B., *Struktura geometryczna. Chropowatość i falistość powierzchni*, WNT, Warszawa 1991.
- [49] OCZOŚ K. E., LUBIMOV V., *Struktura geometryczna powierzchni. Podstawy klasyfikacji z atlasem charakterystycznych powierzchni kształtowanych*, Oficyna Wydawnicza Politechniki Rzeszowskiej, Rzeszów 2003.
- [50] OCZOŚ K.E., LUBIMOW W., *Nowe aspekty trójwymiarowej (3D) analizy chropowatości powierzchni obrobionej*, Mechanik, No. 8–9, 1998, 471–476.
- [51] OGILVY J.A., *Theory of Wave Scattering from Random Rough Surfaces*, Adam Hilger, Bristol, Philadelphia and New York 1991.
- [52] OHLÍDAL I., VIŽĎA F., OHLÍDAL M., *Optical analysis by means of spectroscopic reflectometry of single and double layers with correlated randomly rough boundaries*, Optical Engineering, Vol. 34, No. 6, 1995, 1761–1768.
- [53] PAWLUS P., *Topografia powierzchni – pomiary, analiza, oddziaływanie*, Oficyna Wydawnicza Politechniki Rzeszowskiej, Rzeszów 2005.
- [54] PETYKIEWICZ J., *Optyka falowa*, Wyd. Politechniki Warszawskiej, Warszawa 1980.
- [55] RAYMOND C.J., *Scatterometry for Semiconductor Metrology*, [W:] *Handbook of Silicon Semiconductor Metrology*, A.C. Diebold (ed.), Chapter 18, New York 2001, 477–513.
- [56] RICE S.O., *Reflection of Electromagnetic Waves from Slightly Rough Surfaces*, [W:] *Theory of Electromagnetic Waves*, Kline M. (ed.), Wiley-Interscience, New York 1951, 351–387.
- [57] RÖNNELID M., ADSTEN M., LINDSTRÖM T., NOSTELL P., WÄCKELDÅRD E., *Optical Scattering from Rough-Rolled Aluminum Surfaces*, Applied Optics, Vol. 40, 2001, 2148–2158.
- [58] SCHRÖDER S., GLEICH S., DUPARRÉ A., *Measurement system to determine the total and angle-resolved light scattering of optical components in the deep-ultraviolet and vacuum-ultraviolet spectral regions*, Applied Optics, Vol. 44, 2005, 6093–6107.
- [59] STOUT K.J., BLUNT L., *Three-Dimensional Surface Topography*, 2nd ed., Penton Press, London 2000.
- [60] STOVER J.C., *Optical Scattering: Measurement and Analysis*, 2nd ed., SPIE Optical Engineering Press, Bellingham, Wash., USA, 1995.
- [61] SYNAK R., PAWEŁCZAK M., LIPIŃSKI W., LIS T., *Laserowy przyrząd do oceny chropowatości powierzchni na podstawie pomiaru parametru TIS*, Materiały X Krajowej i Międzynarodowej Konferencji Metrologia w Technikach Wytwarzania, tom II, Kraków, 2003, 623–632.
- [62] SYNAK R., PAWEŁCZAK M., LIPIŃSKI W., LIS T., *Skaterometr laserowy SL 31*, II Konferencja Optoelektronika 2003, Komunikaty, Poznań, 2003, s. 116–118.
- [63] THOMAS T.R., *Rough Surfaces*, 2nd ed., Imperial College Press, London 1999.
- [64] THURN G., *Automatisierte Oberflächenprüfung durch rechnergestütztes Messen der Streulichtverteilung*, Technische Universität Berlin, Berlin 1984.
- [65] THWAITE E.G., *The direct measurement of the power spectrum of rough surfaces by optical Fourier transformation*, Wear, Vol. 57, No. 1, 1979, 71–80.

- [66] THWAITE E.G., *The Extension of Optical Angular Scattering Techniques to the Measurement of Intermediate Scale Roughness*, Annals of the CIRP, Vol. 31, No. 1, 1982, 463–465.
- [67] TIAN G.Y., LU R.-S., *Hybrid vision system for online measurement of surface roughness*, Journal of the Optical Society of America A, Vol. 23, 2006, 3072–3079.
- [68] VALLIANT J.G., FOLEY M.P., BENNETT J.M., *Instrument for on-line monitoring of surface roughness of machined surfaces*, Optical Engineering, Vol. 39, No. 12, 2000, 3247–3254.
- [69] WANG S., TIAN Y., TAY C.J., QUAN C., *Development of a Laser-Scattering-Based Probe for On-Line Measurement of Surface Roughness*, Applied Optics, Vol. 42, No. 7, 2003, 1318–1324.
- [70] WHITEHOUSE D.J., *Handbook of Surface and Nanometrology*, Institute of Physics Publishing, Bristol 2003.
- [71] WHITEHOUSE D.J., *Handbook of Surface Metrology*, Institute of Physics Publishing, Bristol and Philadelphia 1994.
- [72] WIECZOROWSKI M., CELLARY A., CHAJDA J., *Przewodnik po pomiarach nierówności powierzchni, czyli o chropowatości i nie tylko*, Zakład Poligraficzno-Wyd., M-Druk Poznań 2003.
- [73] WYSOCZAŃSKI D., *Światło rozproszone w pomiarach właściwości materiałów kompozytowych*, Elektronizacja, 10, 1997, 7–12.

SCATTEROMETRY OF ROUGH SURFACES

This work presents main problems covering the measuring of surface roughness by light scattering methods. These methods are applied in automated manufacturing processes, particularly for measuring precisely machined surfaces. Diffraction model of light scattering from rough surfaces is discussed. These considerations include principal light scattering theories: Kirchhoff theory and Rayleigh–Rice perturbation theory. Short analysis of light scattering from periodical rough surfaces is also given. Influences of surface self-shadowing and multiple reflections on resultant field scattered from rough surface are analysed. Next, the light scattering methods of measuring the surface roughness are presented. In the final part of the work the main general conclusions following from presented analyses and investigations are given. These conclusions and discussions presented in the work can be treated as a summation of hitherto author's works in the scatterometry of rough surfaces.

*spektrometria masowa czasu przelotu,
funkcja intensywności,
pomiar odcinka czasu*

Marek ZIELIŃSKI*, Anna ZAWADZKA*
Dariusz CHABERSKI*

POMIARY SKŁADU IZOTOPOWEGO WIĄZKI ATOMOWEJ Z WYKORZYSTANIEM SPEKTROMETRU MASOWEGO CZASU PRZELOTU**

Praca dotyczy problemu pomiaru składu izotopowego wiązki atomowej z wykorzystaniem spektrometru masowego czasu przelotu. Pierwsza część pracy zawiera wprowadzenie wyjaśniające, czym jest spektrometria masowa oraz opis sytemu pomiarowego służącego do badania składu izotopowego wiązki atomowej. W części kolejnej przedstawiono układy pomiarowe oraz przeprowadzono analizę ich dokładności. Wskazano na źródła błędów, sposoby ich ograniczania i redukcji. W części końcowej zaprezentowano wyniki przeprowadzonych eksperymentów oraz wnioski z nich wynikające.

1. SPEKTROMETRIA MASOWA

1.1. WPROWADZENIE

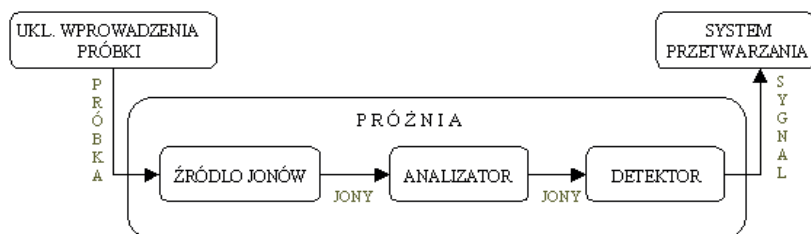
Spektrometria masowa jest metodą analizy, szeroko stosowaną m.in. w fizyce atomowej, reakcjach jon–cząstka, geochronologii, wszelkich analizach chemicznych, naukach biomedycznych itp. Cechuje ją przede wszystkim wyjątkowo wysoka czułość oraz nieporównywalnie niska granica wykrywalności. Uzyskiwane ze spektrometru widmo masowe jest statystycznym rozkładem mas badanych atomów i cząstek.

Zasada działania spektrometru masowego polega na rozróżnianiu cząstek na podstawie ich stosunku masy m do ładunku z . Każda badana substancja jest jonizowana, potem dzięki użyciu analizatora (separatora) cząstki o różnych masach są rozdzielane, a intensywność występowania w zależności od ich masy wykrywana na detektorze. Po przetworzeniu przez komputer zarejestrowanych danych otrzymuje się widmo masowe badanej substancji.

* Instytut Fizyki, Uniwersytet Mikołaja Kopernika w Toruniu.

** Praca wykonana w ramach projektu MNiSW Nr 3 T10C 015 30.

Od czasu zarejestrowania pierwszego widma masowego prostych cząstek (J.J. Thomson, 1912r.) skonstruowano wiele różnych urządzeń. Konstrukcja wszystkich spektrometrów masowych zgodna jest z ogólnym schematem budowy przedstawionym na rysunku 1.1.



Rys. 1.1. Schemat budowy spektrometru masowego
Fig. 1.1. Block diagram of mass spectrometer

Wprowadzenie próbki do układu może odbywać się w sposób bezpośredni lub za pomocą odpowiednich układów wlotowych (np. chromatografów). Wytwarzane jony analizowanej próbki (źródło jonów) są rozdzielane w zależności od ich mas (analyzer) i wykrywane w detektorze, gdzie przypisywane są im masy (ściślej – stosunek m/z) i intensywności występowania. Dane te są rejestrowane i przetwarzane (komputer). W niektórych przyrządach źródło jonów nie leży w obszarze próżni – jonizacja zachodzi pod ciśnieniem atmosferycznym.

1.2. ŹRÓDŁA JONÓW

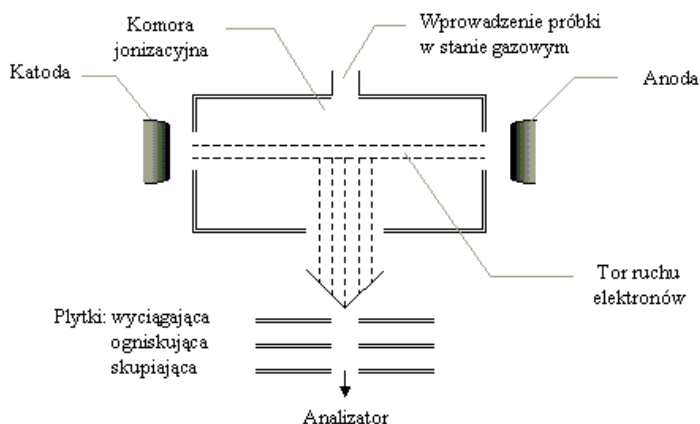
Obszar, w którym zachodzi jonizacja ogranicza się na ogół do małej komory próżniowej*, w której analizowana substancja przeprowadzana jest w jony w fazie gazowej. Istnieje wiele sposobów jonizacji związków chemicznych, z których najważniejsze przedstawiono w tabeli 1.1.

Do najczęściej stosowanych metod należy jonizacja elektronami (EI) [2]. Elektrony energetyczne emitowane przez rozgrzane włókno żarowe (katodę), przyspieszane są w kierunku anody (pułapki elektronowej) na skutek różnicy potencjałów między elektrodami. W strumieniu elektronów umieszcza się analizowaną próbkę w fazie gazowej (rys. 1.2). Elektrony o odpowiednio dużej energii, oddziałując z analizowanymi cząstkami (atomami), powodują oderwanie elektronu z cząstki i utworzenie jonu molekularnego: $M + e^- \rightarrow M^{+\bullet} + 2e^-$.

* Przeprowadzanie jonizacji w warunkach próżniowych nie dotyczy technik jonizacji pod ciśnieniem atmosferycznym (API). Powszechne stosowanie niskiego ciśnienia w obszarze jonizacji związane jest z trudnością skonstruowania połączenia źródła jonów pod ciśnieniem atmosferycznym z analizatorem, którego ciśnienie nie powinno przekraczać 10^{-5} Tr.

Tabela 1.1. Czynniki wywołujące jonizację
Table 1.1. Implicants of ionization

Czynnik	Przykład
Elektrony	Jonizacja elektronami (EI)
Reakcje jonów	Jonizacja chemiczna (CI), w tym pod ciś. atmosferycznym (APCI)
Pola elektryczne	Jonizacja polem (FI), desorpcja polem (FD), jonizacja elektrohydrodynamiczna (EHI)
Pola elektryczne + rozpylanie	Termosprej (TSP), elektrosprej (ESI)
Szybko poruszające się atomy	Bombardowanie szybkimi atomami (FAB)
Szybko poruszające się jony	Spektrometria mas jonów wtórnych (SIMS)
Fotony	Desorpcja laserowa (LD): w stałej matrycy (MALDI), jonizacja wielofotonowa (MPI)
Produkty rozszczepiania jądrowego	Desorpcja plazmowa (PD)
Wyładowanie jarzeniowe	Spektrometria mas przez wyładowanie żarowe (GDMS)
Ciepło	Jonizacja termiczna
Iskra elektryczna	Spektrometria mas ze źródłem iskrowym (SSMS)



Rys. 1.2. Schemat źródła jonów z jonizacją elektronami
Fig. 1.2. Diagram of ion source with electron ionization

Jony molekularne ulegają fragmentacji. W przypadku złożonych jonów molekularnych mogą tworzyć się całe ścieżki fragmentacyjne. Proces fragmentacji jonów molekularnych związany jest ze wszystkimi metodami jonizacji. Jest to duże utrudnienie w przypadku analiz złożonych związków chemicznych (głównie związków organicznych). W takich sytuacjach wykorzystuje się metodę jonizacji chemicznej, gdzie jony są wytwarzane na skutek zderzeń jonizowanej cząstki z obecnymi w źródle jonami pierwotnymi. W widmach otrzymanych tą metodą obserwuje się jony o masie związanej w określony sposób z masą cząsteczkową badanej substancji i stąd identyfikacja cząsteczek jest znacznie prostsza.

Skuteczną metodą otrzymywania jonów w fazie gazowej jest desorpcja laserowa. Impulsy laserowe ogniskowane na małej powierzchni próbki powodują wyrzucenie kilku pikomoli substancji w formie jonów i obojętnych cząstek, które reagują ze sobą w obszarze o ich zwiększonym stężeniu tuż nad powierzchnią próbki. Jest to metoda stosowana głównie do badań powierzchni i analizy lokalnego składu próbki, wymagająca użycia analizatorów o jednoczesnej detekcji lub analizatorów czasu przelotu [1].

1.3. ANALIZATORY

Najpowszechniejszą metodą separacji jonów pod względem stosunku ich masy do ładunku m/z jest użycie oddzielnych pól magnetycznych i elektrycznych lub też pól połączonych razem. Większość podstawowych różnic pomiędzy takimi spektrometrami leży w sposobie stosowania tych pól.

Tabela 1.2. Charakterystyka najważniejszych rodzajów analizatorów.

Zastosowane skróty: m – masa, z – ładunek, PE – pole elektryczne, PM – pole magnetyczne, DC – prąd stały, \mathbf{E} – wektor pola elektrycznego, \mathbf{B} – wektor indukcji magnetycznej, E_k – energia kinetyczna, v_0 – prędkość początkowa

Table 1.2. Characteristics of most important types of analyzers

Budowa	Opis działania	Charakterystyka
1	2	3
Analizator z sektorem magnetycznym		
W zależności od geometrii pola: magnesy z zakrzywioną rurą analizatora	Dla ustalonej wartości PM jony o tym samym momencie pędu i tym samym z poruszają się po tym samym, dobrze określonym torze, trafiając do detektora	<ul style="list-style-type: none"> • Pojedyncze ogniskowanie • Otrzymywanie widm o małej rozdzielczości • Separacja całkowitych wartości m/z dla $m \approx 5000$
Analizator z sektorem magnetoelektrycznym		
Sektor elektryczny: 2 gładkie, zakrzywione płyty metalowe, w poprzek, których utrzymywany jest potencjał elektryczny	Sektor elektryczny rozdziela jony zależnie od ich E_k (trajektoria jonu w PE nie zależy od masy) – funkcja ogniskująca. Zogniskowane jony trafiają do sektora magnetycznego.	<ul style="list-style-type: none"> • Podwójne ogniskowanie • Wysoka zdolność rozdzielcza • Najdokładniejsze pomiary mas • Wykrywalność do wart. m/z ok. 16000 • Potencjały przyspieszające rzędu 1–10 kV • Konieczność stosowania wysokiej próżni
Analizator kwadrupolowy		
4 równoległe metalowe pręty, do każdej pary przeciwległych prętów przykładana się napięcie DC oraz potencjał zmieniający się z częstością radiową	Ze wszystkich jonów oscylujących w zmiennym PE wytworzonym przez pręty, do detektora trafiają jony o stabilnej trajektorii (dla zadanego \mathbf{E} są to jony o ściśle określonej m), reszta jonów oscylując chaotycznie, zderza się z prętami nie docierając do detektora.	<ul style="list-style-type: none"> • Wykrywalność do wart. m/z ok. 4000 • Rozdzielczość ok. 4000 • Szybkie skanowanie wielu widm • W porównaniu z analizatorami z sektorem magnetycznym: łatwiejszy do zautomatyzowania, połączenia z różnymi zestawami pomiarowymi; dokładniejszy w kontroli; prostszy konstrukcyjnie, mniejszy; trwalszy

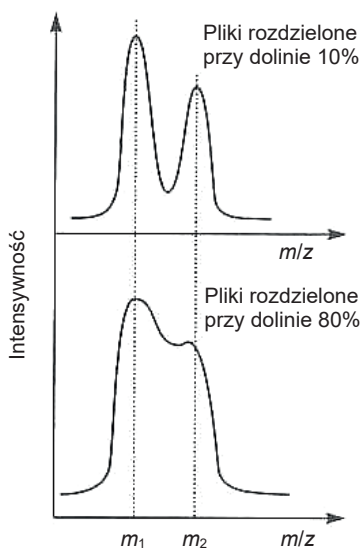
1	2	3
Analityzator czasu przelotu (TOF)		
Obszar swobodnego lotu: rura na stałym potencjale Całość pod próżnią	Obszar przyspieszający: jony o tym samym z uzyskują tę samą v_0 (niezależnie od m). W obszarze swobodnego lotu prędkość jonów zależy od m (nie od z). Jony o różnych m uzyskują różne prędkości i różny jest ich czas przelotu do detektora.	<ul style="list-style-type: none"> • W prostych TOF mała zdolność rozdzielcza • Zastosowanie ukł. korygujących i sprzężeń umożliwia uzyskanie rozdzielczości $>25\,000$ przy dużej czułości • Idealny przy jonizacji impulsowej, np. do źródeł typu MALDI • Nieograniczony zakres mas
Analityzatory cyklotronowego rezonansu jądrowego (ICR)		
Sześcienna komora pod wysoką próżnią umieszczona w PM magnesu nadprzewodzącego (ok. 7T). Przeciwnieległe ściany tworzą 3 pary płytek: pułapkowych, przekaźnika i odbiornika	Jony o niewielkiej prędkości w dużym PM mogą zostać złapane w pułpkę na torze kołowym. Prędkość kątowna jonu zależy od B oraz od m/z . Impulsy o częstości radiowej przykładane do płytek przekaźnika powodują wzbudzenie jonów do krążenia po większych orbitach. Powiększające się trajektorie jonów w końcu przetną się z płaszczyznami płytek odbiornika – w obwodzie indukuje się prąd proporcjonalny do liczby jonów o danej m .	<ul style="list-style-type: none"> • Duże możliwości zastosowania w badaniach dużych cząstek • Jednoczesna detekcja wszystkich jonów • Odpowiednio duża rozdzielczość wymaga bardzo wysokiej próżni • Ekstremalnie duża zdolność rozdzielcza dla najmniejszych mas szybko maleje wraz z ich wzrostem • Przy stosowaniu transformaty Fouriera wymagana znaczna pojemność komputera

Spektrometrem wykorzystującym tylko pole elektryczne jest m.in. kwadrupol. Może on pracować jako analityzator skaningowy, wykrywając jony kolejno w określonych przedziałach czasu. Podobnie może skanować analityzator z sektorem magnetycznym. Tego typu urządzenia wykorzystywane są głównie w analizie organicznej.

Analityzator czasu przelotu wymaga tego, aby jony były wytwarzane w następujących po sobie pakietach, jest zatem bardzo dobrze przystosowany do pracy z impulsowymi źródłami jonów. Coraz częściej stosuje się urządzenia, w których łączy się szeregowo wiele analityzatorów. Krótką charakterystykę podstawowych analityzatorów przedstawiono w tabeli 1.2.

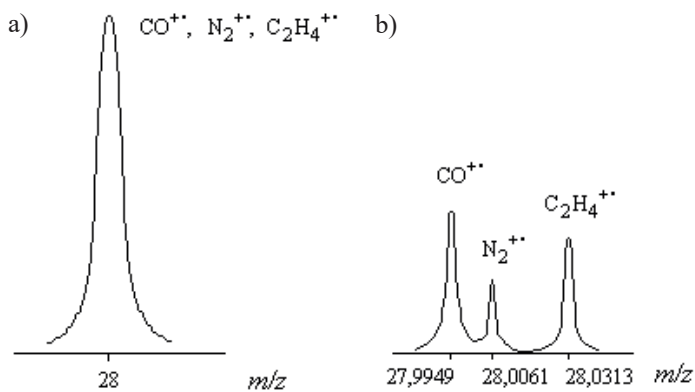
Niezależnie od budowy, wszystkie analityzatory charakteryzuje się za pomocą trzech parametrów: zakresu mas, zdolności rozdzielczej i czułości. Zakres mas określa górną granicę mas, czyli granicę, możliwe do zmierzenia wartości m/z . Rozdzielczość jest zdolnością rozróżniania sygnałów pochodzących od dwóch jonów o sąsiadujących wartościach m/z . Przyjmuje się, że dwa piki są rozdzielone, gdy intensywność doliny między nimi wynosi 10% intensywności słabszego piku, co ilustruje rysunek 1.3 [4]. Jest to tzw. rozdzielczość na poziomie 10%. Jeżeli założymy, że Δm jest najmniejszą różnicą mas między dwoma rozdzielonymi pikami o masach m i $m + \Delta m$, to rozdzielczość R definiuje się jako:

$$R = \frac{m}{\Delta m} . \quad (1.1)$$



Rys. 1.3. Schemat ilustrujący pojęcie doliny
Fig. 1.3. Idea of valley

Wymagania co do rozdzielności szybko rosną wraz ze zwiększającą się masą. Aby zaobserwować dwa jony różniące się masą o $\Delta m = 0,005$ dla jonów o $m \approx 20$, rozdzielność musi wynosić 4000, a dla jonów o $m \approx 100$ rozdzielność musiałaby już wynosić 20000. W warunkach szybkiego przemiatania na najlepszych spektrometrach z trudem uzyskuje się rozdzielność 10000.



Rys. 1.4. Obraz obszaru jonu molekularnego o $m/z = 28$ dla dwóch rozdzielności:

a) $R = 100$, b) $R = 3000$

Fig. 1.4. Image of the area of molecular ion $m/z = 28$ for two resolutions:

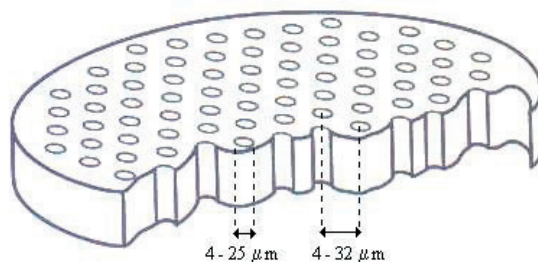
a) $R = 100$, b) $R = 3000$

Duża rozdzielczość jest szczególnie istotna w przypadku analiz organicznych, gdzie badane jony mogą mieć skomplikowaną budowę, a poszczególne produkty fragmentacji charakteryzujące daną cząstkę mogą różnić się o setne części masy atomowej. Na rysunku 1.4 przedstawiono możliwą interpretację jonu o $m/z = 28$ w zależności od rozdzielczości analizatora.

Czułość w spektroskopii mas definiowana jest w dwojaki sposób: w analizie jakościowej jest to minimalna ilość substancji koniecznej do otrzymania możliwego do interpretacji widma (sygnał charakterystycznych jonów musi być większy od poziomu szumu), a w analizie ilościowej jest to minimalna ilość analizowanego związku, która daje sygnał równy dziesięciokrotnej wartości poziomu szumów. Czułość ilościowa jest również nazywana granicą wykrywalności. Czułość zależy m.in. od sposobu zbierania danych. Najczulszą metodą jest wykrywanie wybranych reakcji, mniej czułym sposobem wykrywanie wybranych jonów i najmniej czułym rozwiązaniem jest skanowanie.

1.4. DETEKTORY

Wiązka jonów, ulegając detekcji, zostaje przekształcona w użyteczny sygnał. Słaby prąd jonowy można przekształcić w sposób bezpośredni (płyta fotograficzna, puszka Faradaya) lub pośredni (powielacze jonowe). Najczęściej stosowanymi detektorami są powielacze elektronowe, płytki mikrokanalikowe i fotopowielacze. Na wejściu każdego z tych detektorów umieszczone są dynody konwersyjne – dla jonów dodatnich o potencjale ujemnym, a dla jonów ujemnych o potencjale dodatnim. Jon na anodzie konwersyjnej powoduje emisję elektronów, które kierowane są do detektora*, gdzie powielane kaskadowo ulegają wielokrotnemu wzmocnieniu. Największym wzmocnieniem (do 10^8) charakteryzują się płytki mikrokanalikowe [5]. Detektor ten zbudowany jest z płytki z nawierconymi cylindrycznie, równoodległymi kanalikami. Różnica potencjałów między wejściem a wyjściem płytki jest rzędu 1 kV (rys. 1.5).



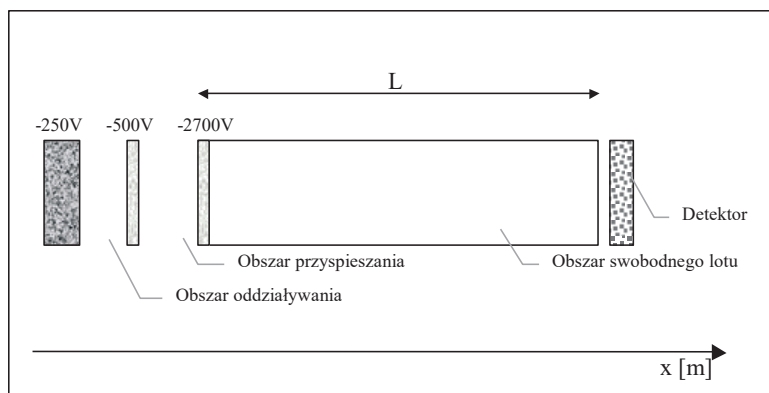
Rys. 1.5. Przekrój płytki z mikrokanalikami
Fig. 1.5. Cross-section of microchannel plate

* W przypadku fotopowielacza elektrony emitowane przez dynody konwersyjne przyspieszane są w kierunku ekranu fosforescencyjnego, gdzie są przekształcane w fotony.

Na powierzchni każdego kanalika naniesiony jest półprzewodnik, z którego w procesie powielania elektronów emitowane są elektrony. Pojedyncza płytka ma wzmocnienie około 10^5 , stosując wiele płytek można uzyskać wzmocnienie 10^8 . Na wyjściu każdego kanalika anoda zbiera wzmocnione elektrony, przekazując sygnał do miernika. Usytuowanie na płytce kanalików, do których trafiają elektrony, zależy od wartości m/z jonu wywołującego powielenie elektronowe.

1.5. SPEKTROMETRY CZASU PRZELOTU

Paczka jonów o różnym stosunku m/z , ale równej energii i momencie pędu wchodząc w obszar stałego pola elektrycznego są rozdzielane zgodnie z ich stosunkiem m/z . Dla jonów o jednostkowym ładunku ($z = 1$) separacja zachodzi tylko na podstawie masy jonu. Masę tę określa się na podstawie czasu, jakiego potrzebuje jon, aby przebyć daną odległość L .



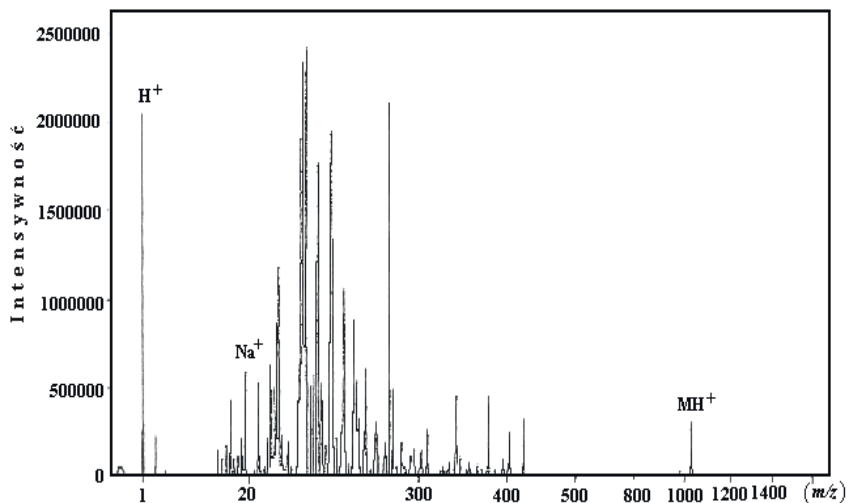
Rys. 1.6. Schemat spektrometru czasu przelotu
Fig. 1.6. Diagram of time-of-flight spectrometer

Pole elektryczne przyspiesza wszystkie jony do tej samej energii kinetycznej. Gdy jony przejdą przez siatkę, lecąc przez obszar dryfu (swobodnego lotu), osiągną prędkości zależne od ich masy. Typowa odległość obszaru źródłowego jest rzędu 0,5 cm, długość obszaru przelotu od 15 cm do 8 m. Napięcie przyspieszające na ogół jest z zakresu 3–30 kV, a czas przelotu rzędu 5–100 μs [6–10].

1.6. SKALOWANIE

Widmo czasu przelotu może być bezpośrednio przekonwertowane na widmo masowe, dzięki znajomości napięcia przyspieszającego i długości lotu. W praktyce mierzy się czas przelotu, co najmniej dwóch jonów o znanych masach, z których określa-

ne są stałe a , b w empirycznym równaniu $m/z = at^2 + b$. Widmo masy skalowane jest na ogół jonami Na^+ i H^+ . Na tej podstawie określa się pozostałe nieznanne masy.



Rys. 1.7. Widmo masowe z zaznaczonymi pikami kalibracyjnymi H^+ i Na^+

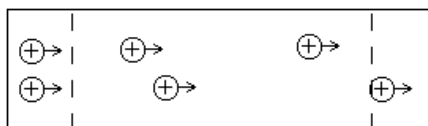
Fig. 1.7. Mass-spectrum with calibrations picks of H^+ and Na^+

1.7. ROZKŁADY POCZĄTKOWE

Największym problemem spektrometrów czasu przelotu jest zdolność rozdzielcza. Już w 1955 r. zauważono, że tak naprawdę nie wiemy, czy wszystkie jony powstają w tym samym miejscu, w tym samym czasie i czy ich energie początkowe są sobie równe. Rozróżniamy trzy zasadnicze rozproszenia początkowe.

1.7.1. ROZKŁAD CZASU

Gdy czas powstawania jonów nie jest identyczny, jon, który powstał wcześniej, pokona obszar przelotu, utrzymując stałą różnicę w czasie Δt i przestrzeni w stosunku do jonu powstałego później.



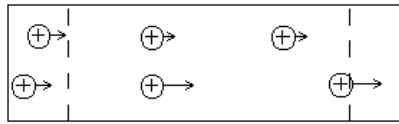
Rys. 1.8. Początkowy rozkład czasu powstania jonów: $t_1 \neq t_2$

Fig. 1.8. Initial distribution of ions creation time: $t_1 \neq t_2$

Rozdzielczość masową można polepszyć zwiększając czas przelotu. Aby to uzyskać, można zmniejszyć napięcie przyspieszające i/lub zwiększyć długość obszaru przelotu.

1.7.2. ROZKŁAD POŁOŻENIA

Jony powstałe bliżej źródła będą przyspieszone do większych energii kinetycznych niż te powstałe w pobliżu siatki wypychającej, wejdą one w obszar dryfowania później, ale mają większe prędkości, więc wyprzedzą te uformowane przy siatce i zostaną wychwycone przez detektor wcześniej.

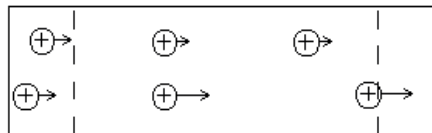


Rys. 1.9. Początkowy rozkład miejsca powstania jonów: $x_1 \neq x_2$
Fig. 1.9. Initial distribution of ions creation place: $x_1 \neq x_2$

Możliwe jest takie przygotowanie pola, aby jony o danej masie skupiały się w miejscu położonym w pewnej odległości od źródła (dwukrotna odległość płytki-siatka). Położenie tej płaszczyzny jest niezależne od masy, jakkolwiek jony o różnych masach dotrą do tego ogniska w różnych czasach. Ogniskowanie przestrzenne na ogół osiąga się zmniejszając napięcie wypychające, należy jednak pamiętać, że skupianie początkowej energii kinetycznej wymaga wysokiego napięcia. Stosuje się zatem dwuetapowe przyspieszanie: w mniejszym polu koryguje się rozkład przestrzenny, w większym rozkład energii. (Pole jest przyłożone w płaszczyźnie skupiającej, gdzie jej ognisko staje się wirtualnym źródłem jonów.) W technikach desorpcyjnych jony tworzone są na powierzchni, nie w fazie gazowej, rozkład przestrzenny zatem jest znacznie zminimalizowany.

1.7.3. ROZKŁAD ENERGII KINETYCZNEJ

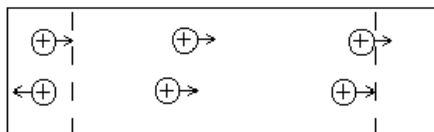
Jony uformowane z różną prędkością początkową będą miały po przyspieszeniu różne prędkości końcowe, osiągną zatem detektor po różnych czasach.



Rys. 1.10. Początkowy rozkład energii kinetycznej jonów: $E_1 \neq E_2$
Fig. 1.10. Initial distribution of kinetic energy of: $E_1 \neq E_2$

Jest to najtrudniejszy do korekcji warunek początkowy w przyrządach liniowych, gdyż wzrastające długości przelotu zwiększają szerokość piku wraz z ich rozdzielaniem względem różnych mas. W celu kompensacji różnic w początkowych energiach kinetycznych stosuje się w przyrządy ogniskujące energię, np. reflektrom.

Gdy jony są formowane w fazie gazowej, rozkład początkowy energii kinetycznych zawiera również jony o tej samej szybkości początkowej, ale różnych kierunkach prędkości.



Rys. 1.11. Początkowy rozkład kierunków poruszania się jonów: $v_1 \neq -v_2$
Fig. 1.11. Initial distribution of ions trajectories: $v_1 \neq -v_2$

Potrzebny jest pewien czas na to, aby jon obrócił się w odpowiednią stronę. Czas przelotu zależy więc również od początkowej pozycji jonu. Dłuższe obszary przelotu zmniejszają efekty obracania się jonów, dodatkowo duże napięcia minimalizują konsekwencje rozkładu początkowego energii kinetycznej.

1.8. ZDOLNOŚĆ ROZDZIELCZA I JEJ OGRANICZENIA

Lasery impulsowe minimalizują niepewność czasu formowania się jonów. Jeżeli jony powstają z gładkiej przewodzącej powierzchni równoległej do siatki przyspieszającej, zminimalizowany zostanie rozkład przestrzenny, początkowy zaś rozkład energii kinetycznej może zostać skorygowany wysokim napięciem i skompensowany reflektromem umieszczonym w analizatorze. Alternatywnie rozkład energii kinetycznej można zmniejszyć przez chłodzenie zderzeniami albo wyciąganie jonów ortogonalnie do kierunku desorpcji.

Większość współczesnych przyrządów minimalizuje rozkłady czasowe i przestrzenne przez desorpcję jonów z ekwipotencjalnej powierzchni, używając lasera impulsowego, cząstek o energii rzędu MeV lub impulsów wiązki elektronów (keV). Takie urządzenia są znacznie prostsze – obszar ekstrakcji jonów jest statyczny. Nie ma niepewności lokalizacji, wszystkie jony osiągają tę samą prędkość z przyspieszania. Jeżeli obszar przelotu jest długi w stosunku do obszaru źródłowego, przybliżone równanie sugeruje, że zdolność rozdzielcza jest niezależna zarówno od obu mas, jak i długości przelotu:

$$R = \frac{m}{\Delta m} \cong \frac{eV}{U_0}. \quad (1.2)$$

Jeżeli przyjąć, że jony desorbują z energią 1 eV, niech źródło będzie o napięciu $U_0 = 10$ keV okazuje się, że zdolność rozdzielcza R dochodzi do 10 000! Takie rezultaty nie są obserwowane w żadnych urządzeniach liniowych. Niemożność obserwowania zdolności rozdzielczej przewidzianej podanym równaniem przypisana jest częściowo technicznym ograniczeniom w konstrukcji przyrządów czasu przelotu. Wiele z tych ograniczeń wciąż może być minimalizowana ciągłym ulepszaniem detektorów i technologii rejestrowania jonów. Inne ograniczenia, takie jak rozkłady początkowe czy rozpady metastabilne wynikają z samego procesu jonizacji. Należy również uwzględnić skończoną grubość próbki – jony są desorbowane w nieco różnych miejscach obszaru źródłowego, stąd też są różnie przyspieszane. Ponieważ niepewności położenia początkowego są efektywnie przekształcane na rozkład końcowej energii kinetycznej, powinny one być skutecznie usuwane (reflektron).

Jony mogą powstawać na powierzchni dopiero po upływie pewnego czasu od impulsu jonizacyjnego. Jest to prawda, szczególnie dla jonów powstających w fazie gazowej we wtórnej reakcji i fakt ten wyjaśnia powszechnie znaną obserwację, że zdolność rozdzielcza jest lepsza dla nieorganicznych jonów niż dla organicznych.

Stabilność wysokiego napięcia używanego do przyspieszania jonów, odpowiedź detektora, jitter (przy uruchamianiu analizatora czas–amplituda) umożliwiają zawężenie szerokości pików, które mogą zostać rejestrowane. Z drugiej strony wszelkie nieregularności w gładkości powierzchni próbki, zaburzenia pola w obszarze drutów siatek przyspieszających dadzą w rezultacie różne długości torów jonów i czasów, ażeby mogły one opuścić obszar oddziaływań.

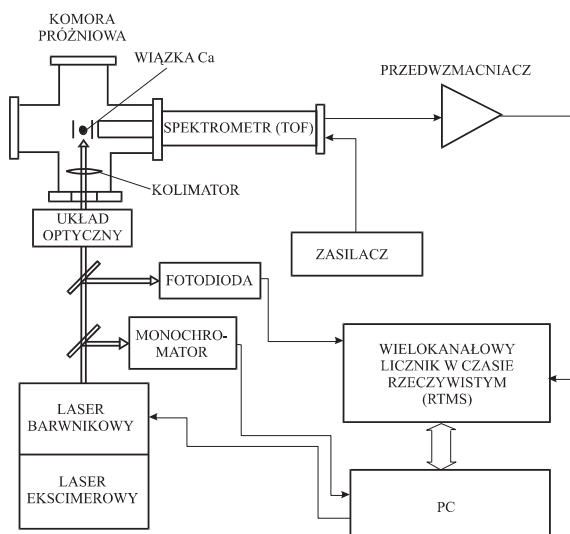
2. SYSTEM POMIAROWY DO BADANIA SKŁADU IZOTOPOWEGO WIĄZKI ATOMOWEJ

Podstawowymi elementami aparatury wykorzystywanej w badaniu składu izotopowego wiązki atomów wapnia są [6–13]:

- komora próżniowa wraz z układem próżniowym i automatycznym układem kontroli próżni,
- układ wytwarzający wiązkę atomów wapnia,
- układ laserów impulsowych,
- układ optyczny,
- spektrometr masowy czasu przelotu,
- układ detekcyjny.

Schemat układu doświadczalnego przedstawiono na rysunku 2.1.

W uproszczony sposób działanie tego układu można przedstawić następująco: każdy impuls światła wiązki laserowej z lasera barwnikowego, za pomocą układu optycznego, zostaje wprowadzony do komory próżniowej, w której jest ogniskowany w obszarze oddziaływania spektrometru czasu przelotu i oddziałuje z wiązką atomów



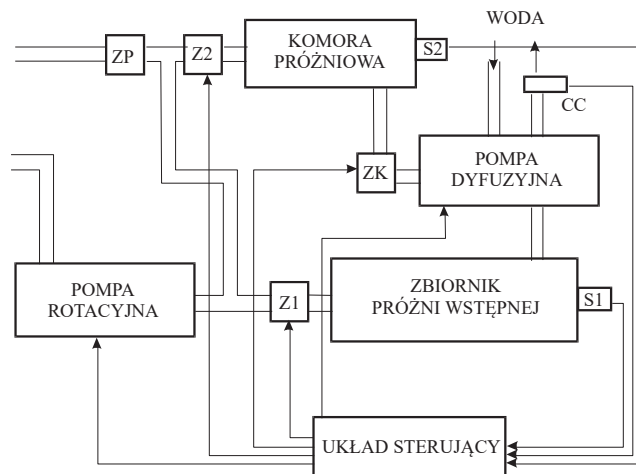
Rys. 2.1. Schemat blokowy aparatury pomiarowej
 Fig. 2.1. Block diagram of the measurement stand

wapnia. W wyniku oddziaływania tych wiązek następuje proces wielofotonowej jonizacji. Następnie, pod wpływem działania zewnętrznego pola elektrycznego, jony są wpychane do spektrometru masowego czasu przelotu, na którego końcu znajduje się detektor. Jako detektor została zastosowana para płytek mikrokanalikowych, w których sygnał jonowy jest wzmacniany. Ładunki, będące wielokrotnością ładunku jonu (wzmocnienie rzędu 10^5), zbierane są na anodzie i jako sygnał w postaci wielu impulsów elektrycznych podawane są na wejście przedwzmacniacza. Sygnał ten po każdym impulsie lasera barwnikowego jest wzmacniany i doprowadzany na wejście systemu pomiarowego współpracującego z komputerem PC. Sygnał wyzwalający licznik pochodzi od szybkiej fotodiody, na którą kierowana jest część energii impulsu lasera. Wychwytywanie niezjonizowanych atomów wapnia z wiązki atomowej zapewnia wymrażarka z ciekłym azotem.

2.1. UKŁAD PRÓŻNIOWY

Do badania składu izotopowego wiązki atomowej niezbędne jest utrzymywanie próżni rzędu 10^{-6} – 10^{-7} mbar, co umożliwia układ przedstawiony na rysunku 2.2. W układzie tym komora próżniowa połączona jest przez zawór klapowy ZK z pompą dyfuzyjną. Pompa ta służy do wytwarzania wysokiej próżni w komorze próżniowej, a jako czynnik pompujący zastosowano olej syntetyczny. Zawór klapowy służy do odcięcia pompy od komory w dwóch przypadkach: rozgrzewania pompy dyfuzyjnej oraz wzrostu ciśnienia w komorze. Od strony wylotu pompy dyfuzyjnej znajduje się

zbiornik próżni wstępnej, w którym próżnia wytwarzana jest za pomocą pompy rotacyjnej. Zastosowanie tego zbiornika umożliwia pompowanie przez cały czas pompą dyfuzyjną z wyłączoną czasowo pompą rotacyjną, a także zabezpiecza olej w pompie



Rys. 2.2. Schemat blokowy układu próżniowego wraz z komorą próżniową
Fig. 2.2. Block diagram of vacuum chamber with vacuum unit

dyfuzyjnej przed spalaniem w przypadku awarii pompy rotacyjnej. Zbiornik próżni wstępnej połączony jest z pompą rotacyjną przez zawór zabezpieczający Z1, zamykany podczas przerw w pracy pompy rotacyjnej. Zawór Z2 umożliwia wstępne odpompowanie komory próżniowej z pominięciem zbiornika próżni wstępnej. Zawory Z1, Z2 i ZK są otwierane elektropneumatycznie. Zawór ZP otwierany jest ręcznie i służy do zapowietrzania komory próżniowej.

Sondy S1 i S2 służą do kontroli ciśnień w obu zbiornikach próżniowych, przy czym sonda S2 umożliwia pomiar ciśnienia w komorze próżniowej (jest to sonda jonizacyjna umożliwiająca pomiar ciśnienia od 10^{-4} – 10^{-8} mbar), a sonda S1 (10^{-3} – 10^{-3} mbar) w zbiorniku próżni wstępnej. Układ zawiera również czujnik ciśnienia wody CC.

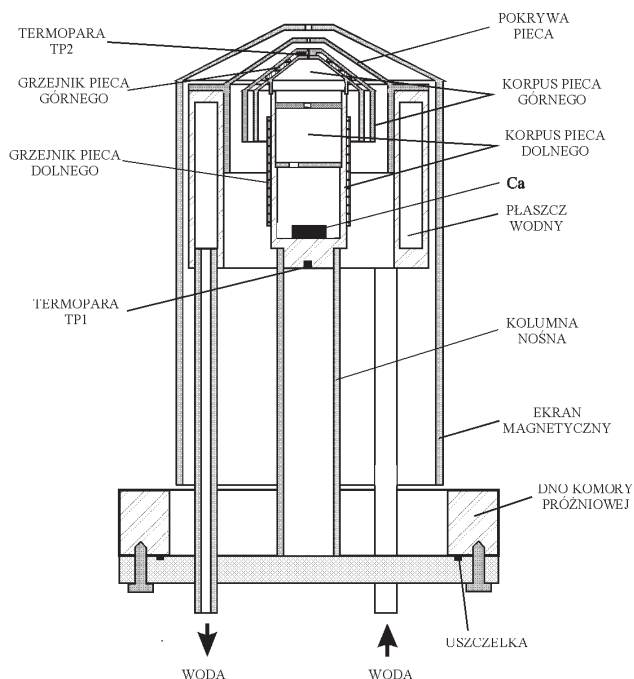
Całym układem próżniowym zarządza układ sterujący, który może pracować w dwóch trybach pracy: trybie sterowania ręcznego i trybie automatycznym. Sygnał z sondy S1 umożliwia automatyczne uruchomienie pompy rotacyjnej i zaworu Z1 (zawór Z1 otwierany jest po siedmiu minutach od włączeniu pompy rotacyjnej) – jeśli ciśnienie w zbiorniku próżni wstępnej wzrośnie do wartości 4×10^{-2} mbar oraz automatyczne wyłączenie pompy i zamknięcie zaworu Z1 – jeśli ciśnienie spadnie do wartości 2×10^{-3} mbar. W układzie sterującym znajduje się kilka układów zabezpieczających, które w przypadku: braku wody, braku fazy zasilania, zbyt wysokiego ciśnienia w komorze próżniowej lub zbiorniku próżni wstępnej powodują całkowite wyłączenie układu i uniemożliwiają ponowne samoczynne włączenie. Taki sposób

zabezpieczenia eliminuje możliwość wystąpienia wyładowania elektrycznego w komorze próżniowej w przypadku niekontrolowanego wzrostu ciśnienia w jej wnętrzu.

2.2. UKŁAD GENERACJI WIĄZKI

Do wytwarzania wiązek atomów lub cząstek pierwiastków czy związków chemicznych, będących w normalnych warunkach ciałami stałymi, wykorzystuje się procesy parowania lub sublimacji materiałów podgrzewanych do odpowiednio wysokich temperatur. W warunkach niskiego ciśnienia temperatury te są zwykle niższe niż temperatury wrzenia, topnienia lub sublimacji określone w warunkach normalnych.

Zasada działania źródła jest następująca: w pewnej objętości, połączonej z układem próżniowym jedynie kanałem o niewielkich wymiarach, wytwarza się stosunkowo wysokie ciśnienie pary danej substancji. Na skutek różnicy ciśnień między źródłem i układem próżniowym następuje wypływ atomów lub cząsteczek substancji do układu próżniowego. W ten sposób w obszarze o niższym ciśnieniu tworzy się strumień cząstek substancji o rozkładzie kątowym zależnym od kształtu i wymiarów kanału, wymiarów przesłon oraz od koncentracji cząstek gazów reszkowych. Źródłem wiązki atomowej stosowanym w eksperymencie jest piec przedstawiony na rysunku 2.3 [8–10].



Rys. 2.3. Schemat pieca wytwarzającego wiązkę atomową
Fig. 2.3. Diagram of the atomic-beam-oven

Składa się on: z pieca właściwego, ekranu magnetycznego, płaszczu wodnego oraz korpusu umożliwiającego zamocowanie w komorze próżniowej. Dolna komora dolnej części pieca spełnia rolę rezerwuaru, w którym umieszcza się substancję w postaci granulatu. Grzejnik umożliwia podgrzanie wapnia do temperatury 600–650 °C. Przesłona pełni funkcję wstępnego kolimatora wiązki atomowej. Pary powstałe na skutek podgrzewania przedostają się do górnej części przez otwory w przesłonach. Górna część pieca stanowi stożkowa pokrywa mająca położoną centralnie dyszę. Spełnia ona funkcję głównego kolimatora wiązki atomowej. Górna część pieca jest w takim układzie podgrzewana do temperatury wyższej względem dolnej o około 100–200 °C. Do uzyskania tej temperatury służy bifilarny grzejnik górny. Do pomiaru temperatury dyszy służy, umieszczona bezpośrednio przy niej, termopara TP2. Piec otoczony jest tzw. płaszczem wodnym z pokrywą pieca chłodzonym ciągłym przepływem wody. Natomiast płaszcz wodny osłonięty jest warstwą μ -metal, który spełnia rolę ekranu magnetycznego. Wiązka atomowa przechodzi do komory próżniowej przez niewielkiej średnicy otwór w płaszczu wodnym, który jest dodatkowym elementem kolimującym na jej drodze.

Atomy po przejściu przez obszar oddziaływania są osadzane na tzw. zimnym palcu wymrażarki chłodzonej ciekłym azotem. Zimny palec stanowi cylinder będący jednocześnie fragmentem górnej pokrywy komory próżniowej. Dzięki śladowi na palcu można określić szerokość wiązki atomowej w obszarze oddziaływania. Średnica śladu wynosi 1,5 cm i ma ostre granice, odległość obszaru oddziaływania od powierzchni pieca wynosi 4 cm, co daje średnicę wiązki w obszarze oddziaływania 0,4 cm. Do oszacowania gęstości wiązki w obszarze oddziaływania można posłużyć się wzorem (Ramsey – 1956):

$$I = 8,34 \cdot 10^{22} \frac{pA_S A_D}{l^2 \sqrt{MT}} \text{ molekuł/s,} \quad (2.1)$$

gdzie:

I – strumień cząstek wylatujących z pieca,

p – ciśnienie par w piecu [Pa],

A_S – powierzchnia źródła [cm²],

A_D – powierzchnia detektora (obszaru, w którym określamy gęstość) [cm²],

l – odległość od źródła [cm],

M – masa atomowa,

T – temperatura pieca [K].

Zależność ciśnienia par w piecu do temperatury pieca dana jest zależnością (Nesmeyanov – 1963):

$$\log p = A - \frac{B}{T} + CT + D \log T \quad (2.2)$$

gdzie:

T – temperatura [K],

A, B, C, D – stałe charakterystyczne dla danego materiału.

Do oszacowania gęstości wiązki jako źródło przyjmuje się otwór wewnętrznej przegrody pieca, natomiast jako detektor wylotową dyszę pieca. Atomy opuszczające dyszę wylotową pieca podlegają statystyce Maxwella i średnia prędkość atomów u wylotu pieca dana jest wzorem (Ramsey – 1956):

$$v = \frac{2}{\sqrt{\pi}} \sqrt{\frac{2kT}{m}}. \quad (2.3)$$

Podstawowe parametry wiązki atomowej obliczone z uwzględnieniem podanych założeń, dla dwóch temperatur pieca 600 i 650 °C (tab. 2.1).

Tabela 2.1. Podstawowe parametry wiązki atomowej dla dwóch temperatur pieca
Table 2.1. Basics parameters of the atomic beam for two temperatures of oven

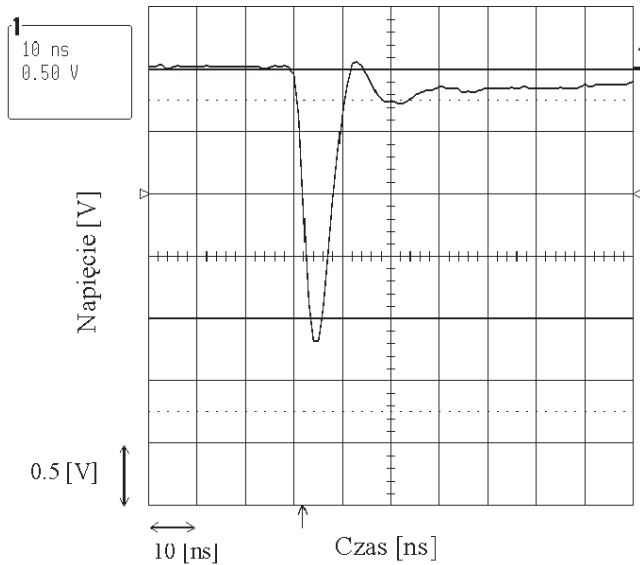
Temperatura pieca dolnego [°C]	600	650
Ciśnienie par w piecu [mbar]	$1,56 \cdot 10^{-2}$	$5,51 \cdot 10^{-2}$
Gęstość strumienia atomów [atomów/s]	$1,92 \cdot 10^{13}$	$6,59 \cdot 10^{13}$
Średnia prędkość atomów [m/s]	$6,8 \cdot 10^{-2}$	$7,0 \cdot 10^{-2}$
Gęstość wiązki w obszarze oddziaływania [atomów/cm ³]	$2,25 \cdot 10^{11}$	$7,7 \cdot 10^{11}$

2.3. UKŁAD LASEROWY

Na układ laserowy składają się dwa lasery: laser ekscimerowy XeCl, służący do wzbudzenia akcji laserowej lasera barwnikowego oraz laser barwnikowy [10]. W eksperymencie wykorzystany został laser ekscimerowy, którego ośrodkiem optycznie czynnym jest mieszanina Ne–Xe–HCl. Długość fali promieniowania wynosi $\lambda = 308$ nm. Przy tej długości fali oraz napięciu zasilania 25 kV energia pojedynczego impulsu lasera wynosi około 200 mJ.

Oscylogram typowego impulsu laserowego z lasera barwnikowego otrzymany dla diody fotoluminescencyjnej przedstawiony jest na rysunku 2.4. Szerokość połówkowa tego impulsu wynosi ~ 10 ns, a czas opóźnienia promieniowania odpowiadającego rozpoczęciu wyładowania ~ 30 ns. Dużą szybkość narastania napięcia w kanale udaje się uzyskać dzięki odpowiedniej konstrukcji samego kanału oraz odpowiedniemu systemowi wzbudzenia. Laser barwnikowy pompowany jest przez laser ekscimerowy XeCl. Schemat pracy lasera przedstawiono na rysunku 2.5 [11, 12].

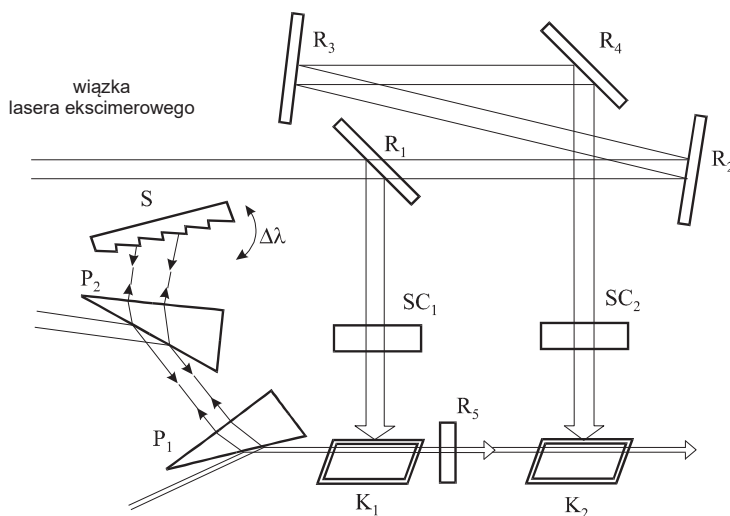
Wiązka pompująca z lasera ekscimerowego pada na zwierciadło R_1 o współczynniku odbicia równym około 0,1. Część odbita wiązki jest skupiana przez soczewkę cylindryczną SC_1 na kuwecie K_1 z barwnikiem. Akcja laserowa zachodzi między



Rys. 2.4. Oscylogram impulsu lasera barwnikowego
 Fig. 2.4. Pulse from the dye laser

zwierciadłem R_5 a dyfrakcyjną siatką odbiciową S w ustawieniu Littrowa. Rozszerzenie wiązki laserowej w celu oświetlenia jak największej liczby rys na siatce S zachodzi na pryzmatach P_1 i P_2 , stanowiących układ rozszerzenia wiązki laserowej. Kąty łamiące pryzmatów są tak dobrane (w zależności od materiału, z którego są wykonane), aby wiązka wyjściowa wychodziła prostopadle do powierzchni pryzmatu i do niego wracała. Następnie wyjściowa wiązka laserowa przez zwierciadło R_5 pada na kuetę z barwnikiem K_2 . Jednocześnie przepuszczona przez zwierciadło R_1 część wiązki lasera ekscimerowego, kolimowana na konfokalnym układzie lusterek R_2 i R_3 oraz kierowana przez lustro R_4 , dociera do tej kuety, powodując inwersję obsadzeń w barwniku. W kuetce K_2 następuje emisja wymuszona przez wyjściową (z oscylatora) wiązkę laserową i wzmocnienie końcowej liczby fotonów wyjściowych. Zmieniając kąt nachylenia siatki S względem kierunku padającej wiązki, można sterować długością emitowanego promieniowania. Szerokość spektralna wychodzącej wiązki wynosi około $0,5 \text{ cm}^{-1}$. Kuetki K_1 i K_2 podłączone są do pomp roztworu barwnika. Stężenie barwnika w roztworze we wzmacniaczu kwantowym K_2 jest trzykrotnie mniejsze niż w oscylatorze laserowym K_1 .

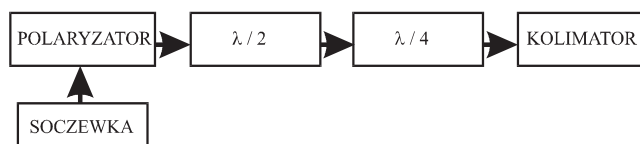
Powierzchnie elementów optycznych zostały doszlifowane z bardzo dużą precyzją do tzw. gładkości laserowej. W celu zmniejszenia strat przy odbiciu wiązek światła padających na powierzchnie optyczne zostały one pokryte warstwami antyodblaskowymi.



Rys. 2.5. Schemat lasera barwnikowego
Fig. 2.5. Diagram of the dye laser

2.4. UKŁAD OPTYCZNY

Stosowany układ optyczny można podzielić na dwie części [12]. Część pierwsza składa się z: soczewki o ogniskowej 1,5 m (służącej do kolimacji wiązki lasera na długości około 4 m), polaryzatora, półfalówki ($\lambda/2$), ćwierćfalówki ($\lambda/4$) – znajdujących się na zewnątrz komory próżniowej i służących do ustawiania polaryzacji: liniowej lub kołowej. Natomiast część drugą układu stanowi kolimator spolaryzowanej wiązki laserowej znajdujący się we wnętrzu komory. Rolę kolimatora spełnia achromatyczny obiektyw skupiający, ustawiony tak, że jego ognisko znajduje się w środku obszaru oddziaływania spektrometru masowego. Umożliwia to uzyskanie skupienia wiązki laserowej do średnicy 100 μm na wiązce atomowej. Schemat blokowy całego układu optycznego przedstawiono na rysunku 2.6.



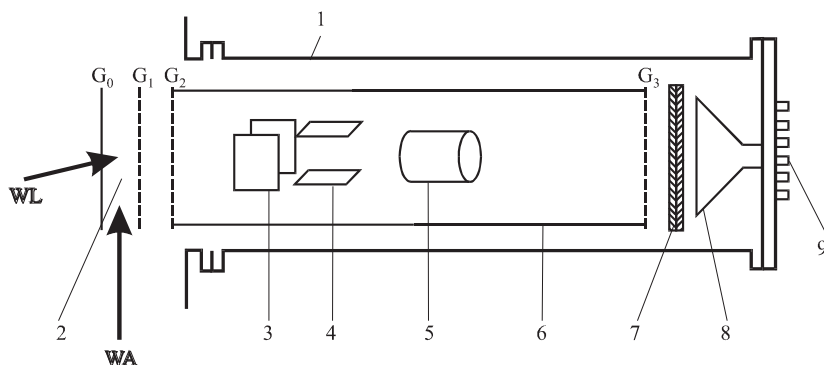
Rys. 2.6. Schemat blokowy układu optycznego
Fig. 2.6. Block diagram of the optical unit

W układzie tym użyty został pryzmat polaryzujący Glan–Laser z jednym oknem wyjściowym. Pryzmat ten złożony jest z dwóch jednakowych pryzmatów

prostokątnych, wykonanych ze szpatu islandzkiego, wyciętych tak, aby oś optyczna kryształu była równoległa do krawędzi dwuściennego kąta α . Używane w eksperymencie: półfalówka i ćwierćfalówka zbudowane są na achromatycznych rombach Fresnela.

2.5. SPEKTROMETR MASOWY CZASU PRZELOTU (TOF)

W przeprowadzonych eksperymentach wykorzystany był spektrometr masowy czasu przelotu model TOF-101 wyprodukowany przez firmę Comstock, który może być stosowany zarówno w jonizacji impulsem laserowym, jak i wiązką elektronową. Podstawowy układ spektrometru w jonizacji wiązką laserową przedstawiono na rysunku 2.7 [8–12].



Rys. 2.7. Schemat spektrometru czasu przelotu
Fig. 2.7. Diagram of the time-of-flight spectrometer

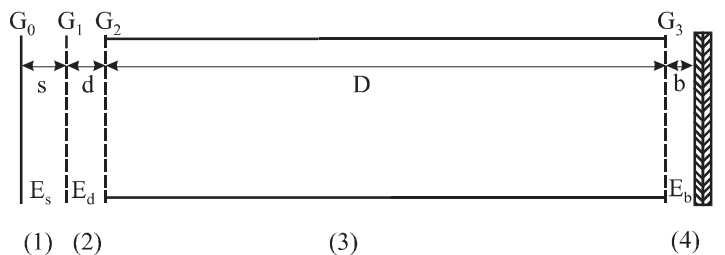
Wiązka laserowa WL przechodzi pomiędzy siatkami G_0 – G_1 (obszar fotojonizacji), powodując jonizację wiązki atomowej WA. Powstałe jony są wyrzucane do obszaru G_1 – G_2 , gdzie są przyspieszane.

W przypadku jonizacji impulsem laserowym czas trwania impulsu powinien być dostatecznie krótki (mniejszy niż kilkadziesiąt ns), wtedy wszystkie wytworzone jony są wypychane przez stałe pole z obszaru jonizacji do obszaru przyspieszającego i dalej do obszaru swobodnego przelotu. Jeśli czas trwania impulsu jest dłuższy, korzystnie jest wytwarzać jony w warunkach zerowego pola i po krótkim opóźnieniu wepchnąć impulsem elektrycznym.

W spektrometrze masowym czasu przelotu można wyróżnić cztery podstawowe obszary, co pokazano na rysunku 2.8 [11, 12]:

- (1) obszar tworzącego jednorodnego pola elektrycznego E_s stosowanego w obszarze jonizacji o długości s ,
- (2) obszar jednorodnego elektrycznego pola przyspieszającego E_d długości d ,

- (3) obszar swobodnego przelotu D ,
 (4) obszar hamowania b .



Rys. 2.8. Podstawowe obszary spektrometru masowego czasu przelotu
 Fig. 2.8. Basic areas of time-of-flight spectrometer

Obszar (1) ograniczony jest siatkami G_0 i G_1 . Potencjały zostały tak dobrane, aby uzyskać maksymalną wydajność zbierania i wypychania jonów z obszaru oddziaływania do obszaru przyspieszającego pomiędzy siatkami G_1 i G_2 jak i do obszaru swobodnego przelotu spektrometru, który jest ograniczony siatkami G_2 i G_3 . Wartości potencjałów są wynikiem przeprowadzenia wielu testów określających wpływ ich wartości (przy stałych pozostałych parametrach) na liczbę zliczanych jonów oraz ich rozkłady w TOF [11]. Siatki G_2 i G_3 są połączone elektrycznie z obszarem swobodnego przelotu i obie znajdują się na stałym potencjale -2700 V. Obszar swobodnego przelotu spektrometru masowego ma kształt cylindra o długości 1 m i średnicy 5 cm. Taka konstrukcja zapewnia brak pola elektrycznego w jego wnętrzu. Znajdująca się na końcu spektrometru para płytek mikrokanałowych stanowi detektor sygnału jonowego. Przednia powierzchnia płytek znajduje się na potencjale -2300 V, natomiast tylna powierzchnia na potencjale -300 V. Anoda jest na potencjale 0 V i zapewnia impedancję falową 50 Ω .

Każdy z czterech obszarów spektrometru wnosi swój wkład do czasu przelotu jonu. Całkowity zatem czas przelotu jest sumą czasów w poszczególnych obszarach (Cotter – 1994):

$$t_c = t_1 + t_2 + t_3 + t_4. \quad (2.4)$$

Całkowitą energię jonu o masie m i ładunku q , wchodzącego do rury po przejściu obszarów: jonizacji i przyspieszania można zapisać jako:

$$U = U_x + qsE_s + qdE_d, \quad (2.5)$$

gdzie:

U_x – początkowa energia kinetyczna jonu poruszającego się wzdłuż prostej równoległej do osi spektrometru,

s – długość obszaru jonizacji,

E_s – natężenie pola w obszarze jonizacji,

d – długość obszaru przyspieszania,
 E_d – natężenie pola w obszarze przyspieszania.

W obszarach (1) i (2) pod wpływem zewnętrznego pola jon porusza się ruchem jednostajnie przyspieszonym [12]. Korzystając z równania ruchu, czas przelotu w tych obszarach można ogólnie przedstawić jako:

$$t = \frac{\sqrt{v_0^2 - 2as'} - v_2}{a}. \quad (2.6)$$

Jeżeli podstawimy odpowiednie wyrażenia na prędkość początkową i przyspieszenie, otrzymamy dla obszaru (1)

$$v_0 = \sqrt{\frac{2U_x}{m}}, \quad a = \frac{qE_s}{m}, \quad (2.7)$$

natomiast

$$t_1 = \frac{\sqrt{2m}}{qE_s} \left(\sqrt{U_x + qsE_s} - \sqrt{U_x} \right), \quad (2.8)$$

dla obszaru (2)

$$v_0 = \sqrt{\frac{2(U_x + qsE_s)}{m}}, \quad a = \frac{qE_d}{m}, \quad (2.9)$$

$$t_2 = \frac{\sqrt{2m}}{qE_d} \left(\sqrt{U} - \sqrt{qsE_s + U_x} \right). \quad (2.10)$$

W obszarze (3) jon porusza się ruchem jednostajnym, czas przelotu zatem wynosi

$$t_3 = D \sqrt{\frac{m}{2U}}. \quad (2.11)$$

Odpowiednio czas przelotu w obszarze (4) wynosi

$$t_4 = \frac{\sqrt{2m}}{qE_b} \left(\sqrt{U - qbE_b} - \sqrt{U} \right). \quad (2.12)$$

Ponieważ obszary (1), (2) i (4) są na ogół dużo mniejsze niż obszar (3) czasy przelotu dla tych obszarów można zaniedbać. Wówczas wzór na całkowity czas przelotu jonu przyjmuje prostą postać:

$$T = D \sqrt{\frac{M}{2E}}, \quad (2.13)$$

gdzie:

M – masa jonu,

D – długość obszaru swobodnego przelotu,

E – całkowita energia jonu.

Przekształcając wzór (2.13), otrzymujemy:

$$M = 2E \frac{T^2}{D^2}. \quad (2.14)$$

Z podanego wzoru wynika, że masa jest proporcjonalna do kwadratu czasu przelotu, dlatego widmo masowe jest nieliniowe. Obliczając pierwszą pochodną ze wzoru (2.14), otrzymujemy

$$dM = \frac{4E}{D^2} T dT. \quad (2.15)$$

Jeśli $M/\Delta M$ zdefiniujemy jako rozdzielczość, to z (2.14) i (2.15) otrzymamy:

$$\frac{M}{\Delta M} = \frac{T}{2\Delta T}. \quad (2.16)$$

W ten sposób dla stałych wartości D i E dokładność pomiaru masy jest określona przez dokładność pomiaru T i ΔT . Jeśli zatem wszystkie jony byłyby początkowo nieruchome i tworzone w tym samym miejscu, to czas przelotu dla wszystkich jonów mających taki sam stosunek masy do ładunku powinien być taki sam. Rozdzielczość masy zależałaby wtedy tylko od rozdzielczości czasowej detektora i układu detekcyjnego. W rzeczywistości rozdzielczość masy jest również uzależniona od położenia początkowego i od rozkładu początkowej energii kinetycznej.

2.6. UKŁAD DETEKCYJNY

W układach pracujących z impulsowymi źródłami wzbudzającymi, takimi jak np. impulsowy laser barwnikowy, sygnał pochodzący z detektora (para płytek mikrokanałowych) ma charakter paczki impulsów, która może być stosunkowo szeroka i przesunięta w czasie (widmo masowe czasu przelotu) względem impulsu wyzwalającego (impuls lasera barwnikowego). Pojedyncze impulsy w paczce mają szerokość połówkową od 1 do 10 ns. W tego typu przypadkach wybór metody pomiarowej zależy od częstości występowania impulsów, ich liczby oraz wymaganej rozdzielczości czasowej [13]. W przypadku małej liczby impulsów sygnału wejściowego najczęściej stosowane są metody oparte na bramkowych systemach zliczających. W systemach tych ograniczenie liczby impulsów sygnału wejściowego spowodowane jest ograniczoną rozdzielczością liczników oraz szerokością pasm wzmacniaczy i przedwzmacniaczy.

Ciąg przypadkowych impulsów, pojawiający się na wyjściu detektora, który stanowią płytki mikrokanałowe może być modelowany jako proces przypadkowy punktowy będący matematycznym modelem procesu o rozkładzie Poissona. Proces ten jest szczególnie przydatny do opisu zjawisk fizycznych, których pojedyncze zdarzenia zachodzą w bardzo krótkim czasie, rozmieszczonych w sposób losowy [14]. Proces ten dobrze nadaje się do modelowania sygnału wyjściowego z detektora spektrometru masowego czasu przelotu. Intensywność występowania pojedynczych zdarzeń w wyniku realizacji takich procesów jest bardzo istotna zarówno dla opisu samego procesu, jak też systemu pomiarowego. W przypadku procesów o małej intensywności, dla których można przyjąć, że prawdopodobieństwo pojawienia się kolejnego impulsu w odstępie czasu równym czasowi trwania impulsu jest zawsze równe zero, dobrym modelem jest proces punktowy. W przypadku intensywności średniej prawdopodobieństwo to jest większe od zera, ale wciąż małe, a obserwowany proces można traktować jako proces filtrowany. W takich przypadkach można stosować cyfrowe metody analizy sygnału. Funkcję intensywności można wówczas zmierzyć dwoma sposobami:

- zliczając liczbę impulsów w wąskich kanałach czasowych dla wielu realizacji procesu,
- grupując według kanałów czasowych o założonej szerokości zarejestrowane stemple czasowe poszczególnych impulsów dla wielu realizacji procesu.

W przypadku bardzo dużych intensywności badanego procesu sygnał na wyjściu detektora nie przedstawia już ciągu pojedynczych impulsów, lecz jest proporcjonalny do ich średniej liczby w jednostce czasu. W takim przypadku stosuje się metody analogowe, polegające na pomiarze prądu średniego i analizie funkcji gęstości widmowej mocy uzyskanego sygnału. Cechą znamioną procesów niestacjonarnych jest możliwość występowania wszystkich wymienionych przypadków w różnych przedziałach czasu tego samego procesu.

Systemy pomiaru funkcji intensywności stosowane w spektrometrii masowej o nanosekundowej rozdzielczości działają na zasadzie zliczania impulsów, których momenty pojawiania mogą być traktowane jako proces punktowy o rozkładzie Poissona. Dla dowolnej chwili $t_0 \leq s < t$ przyrost zliczeń $N_{s,t} = N_t - N_s$, podlega rozkładowi Poissona, co można wyrazić w postaci zależności

$$P(N_{s,t} = n) = (n!)^{-1} \left(\int_s^t \rho(t) dt \right)^n \exp \left[- \left(\int_s^t \rho(t) dt \right) \right], \quad (2.17)$$

gdzie $\rho(t)$ jest funkcją intensywności procesu [13, 14].

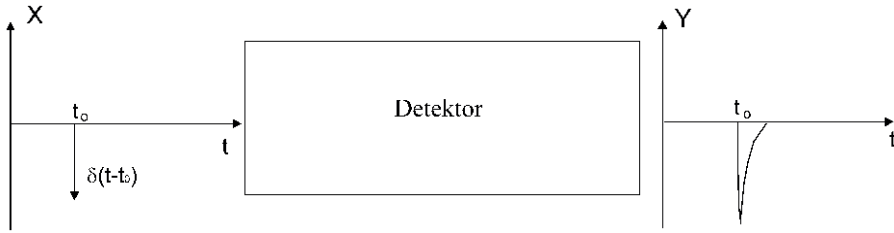
Nośnikiem informacji o rzeczywistym procesie fizycznym jest sygnał pomiarowy. Elementem generującym sygnał elektryczny jest detektor [15–18]. Sposób detekcji sygnału, rodzaj i parametry detektora oraz jego umiejscowienie w stosunku do

źródła procesu fizycznego wpływają na sygnał wyjściowy detektora. Nie zawsze sygnał wyjściowy z detektora można uważać za prosty proces stacjonarny lub nie-stacjonarny, czasami wykazuje on strukturę złożoną z kilku niezależnych procesów lub jest procesem podwójnie przypadkowym. Odpowiednie przeprowadzenie procesu fizycznego oraz odpowiednie umiejscowienie detektora względem obszaru oddziaływania, gdzie zachodzi proces fizyczny, umożliwia rozdzielenie w czasie składników procesu złożonego oraz ich analizę jako sumy oddzielnych procesów.

Efekt ten wykorzystuje się w spektrometrii masowej czasu przelotu [5, 6, 19, 20] oraz w systemach czasowo rozdzielczej detekcji fotonów [21–28]. Procesy te zwykle charakteryzuje duża dynamika zmian funkcji intensywności, w wyniku czego w krótkim przedziale czasu na wyjściu detektora może pojawić się zarówno mała, jak i duża liczba impulsów. W przypadku dużej intensywności sygnału model punktowy nie uwzględnia efektów pojawiających się w związku ze skończonym pasmem przenoszenia detektora. W takich przypadkach detektor można traktować jak filtr dolnoprzepustowy, a sygnał wyjściowy jest wówczas sygnałem filtrowanym [29].

Pomiar funkcji intensywności procesu może być realizowany zarówno w systemie analogowym, jak i cyfrowym. W przypadku bardzo dużych intensywności sygnał wyjściowy z detektora nie może być traktowany jako ciąg pojedynczych impulsów o przypadkowym rozmieszczeniu, ponieważ detektor stanowi układ filtrujący. Jest to jedyny przypadek, w którym analogowe metody integracyjne są bardziej odpowiednie do analizy sygnału niż metody cyfrowe [30]. W sytuacji kiedy intensywność procesu jest bardzo mała, metody analogowe nie są skuteczne, a metody cyfrowe są w tym zakresie bardzo efektywne [15, 30]. Szczególnie trudnym do analizy jest przypadek, kiedy funkcja intensywności wykazuje bardzo dużą dynamikę zmian. W takich przypadkach proces można badać różnymi metodami w określonych przedziałach czasu, co stanowi dość skomplikowaną i czasochłonną procedurę pomiarową lub prościej, stosować metody cyfrowe z korekcją uzyskanych rezultatów dla przedziałów czasu, w których intensywność procesu jest bardzo duża [31]. Zastosowanie impulsowego źródła wzbudzenia procesu fizycznego [32–36], umożliwia analizę czasowo rozdzielczą, a w przypadku procesów złożonych, ich dekompozycję na pojedyncze procesy. Za przykład zastosowań impulsowych źródeł wzbudzenia dla procesów fizycznych, poza spektrometrią czasu przelotu, można podać pomiary zaników populacji atomów ze stanów wzbudzonych do stanów o mniejszej energii oraz detekcję jonów i cząsteczek dla spektroskopii fourierowskiej [32–36].

Istotnym elementem w procesie badania rzeczywistych procesów fizycznych jest detektor, w którym w spektrometrii masowej czasu przelotu są płytki mikrokanałowe. Proces detekcji sygnału pomiarowego można przedstawić jako proces filtracji dolnoprzepustowej, co zilustrowano na rysunku 2.9 [9, 29].



Rys. 2.9. Proces detekcji sygnału pomiarowego
 Fig. 2.9. Detection process of the measurement signal

Przyjmując, że idealny sygnał wejściowy $X(t)$ może być wyrażony, jako suma losowo rozmieszczonych w czasie funkcji impulsowych $\delta(t - t_i)$, gdzie t_i jest zmienną losową, wówczas każdy pojedynczy impuls wejściowy $\delta(t - t_0)$ zostanie transformowany do postaci zależnej od transmitancji detektora. Jeżeli transmitancja detektora $K(s)$ wyraża się funkcją

$$K(s) = \frac{K_0}{(s - s_1)(s - s_2)}, \quad (2.18)$$

gdzie:

K_0 – stała zależna od konstrukcji detektora (czułość),

s_1, s_2 – bieguny transmitancji ($s_1 \neq s_2$), wówczas odpowiedź detektora na ten sygnał przyjmie postać [29]

$$Y(t) = A \left[\exp\left(-\frac{t-t_0}{\tau_1}\right) - \exp\left(-\frac{t-t_0}{\tau_2}\right) \right], \quad (2.19)$$

gdzie

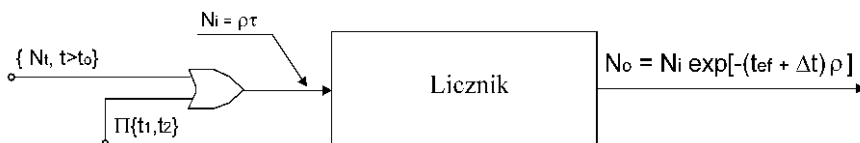
$$\tau_1 = \frac{1}{|s_1|}, \quad \tau_2 = \frac{1}{|s_2|}, \quad aA = \frac{K_0}{s_1 - s_2}.$$

Amplituda impulsów wyjściowych może zależeć także od współrzędnych przestrzennych, związanych z położeniem detektora.

Każdy system cyfrowy, do którego zostaną doprowadzone impulsy z detektora, ma określoną rozdzielczość Δt (minimalny czas, jaki musi upłynąć od impulsu poprzedzającego, aby kolejny impuls wejściowy został zaakceptowany jako oddzielny impuls). W przypadku jednorodnego, stacjonarnego procesu Poissona o intensywności ρ i detektora rzeczywistego, transformującego proces punktowy w ciąg impulsów o szerokości efektywnej t_{ef} , prawdopodobieństwo zaakceptowania impulsu wejściowego będzie równe prawdopodobieństwu zdarzenia, że w czasie $t = t_{ef} + \Delta t$ nie pojawi się żaden impuls, co można zapisać w postaci

$$P(N_t = 0) = \exp[-\rho(t_{ef} + \Delta t)]. \quad (2.20)$$

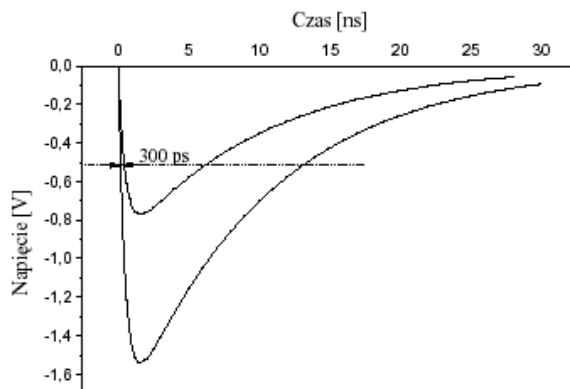
Jeżeli układem cyfrowym będzie licznik o rozdzielczości Δt , to liczba impulsów zliczonych N_0 może być różna od liczby impulsów N_i , które pojawiły się na wejściu licznika w tym samym czasie. Proces ten zilustrowano na rysunku 2.10.



Rys. 2.10. Proces zliczania impulsów o rozkładzie Poissona
Fig. 2.10. Counting process of the pulses with Poisson distribution

Jeżeli przyjmiemy, że transmitancję detektora opisuje zależność (2.18), wówczas impulsy będą miały kształt opisany zależnością (2.19). Jeżeli system pomiarowy będzie reagował na amplitudę impulsów, wówczas przy pewnych fluktuacjach ich wysokości pojawi się błąd określenia położenia impulsu.

Efekt ten zilustrowano na rysunku 2.11. Jak widać na rysunku, błąd ten osiąga wartość kilkuset pikosekund. Zmniejszenie wartości napięcia progowego zmniejsza wprawdzie ten błąd, ale zwiększa ryzyko pogorszenia stosunku sygnał–szum (SNR).

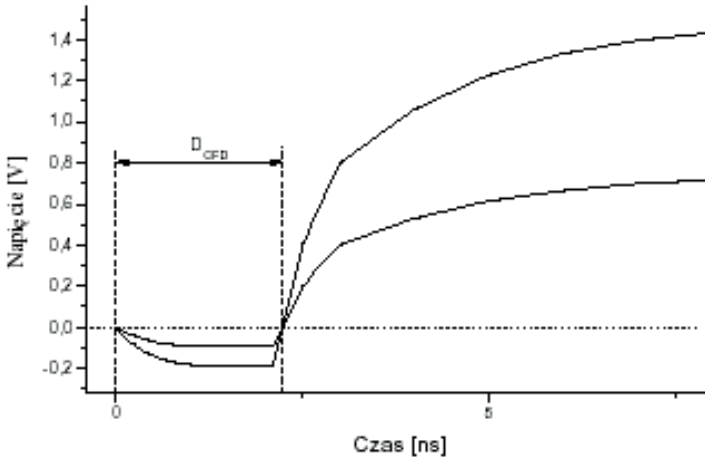


Rys. 2.11. Efekt powstawania błędu położenia impulsów na skutek zmian ich amplitudy
Fig. 2.11. Effect of pulse-position-error caused by amplitude fluctuations

Jak widać na rysunku, błąd ten osiąga wartość kilkuset pikosekund. Zmniejszenie wartości napięcia progowego zmniejsza wprawdzie ten błąd, ale zwiększa ryzyko pogorszenia stosunku sygnał–szum (SNR). Efekt ten jest możliwy do wyeliminowania przez zastosowanie dyskryminatora ze stałym współczynnikiem tłumienia (ang. *constant fraction discriminator*) co zilustrowano na rysunku 2.12 [13]. Pod-

czas takiej detekcji każdy impuls przechodzi przez dwa kanały. W pierwszym kanale jest opóźniany o czas równy D_{CFD} i odwracany (zmienia polaryzację), a w drugim jest tłumiony ze stałym współczynnikiem tłumienia W_{CFD} . Tak uformowane sygnały są sumowane i podawane na komparator. Sygnał z wyjścia komparatora podawany jest na wejście przerzutnika monostabilnego, który generuje impuls o standardowym kształcie.

Błąd określenia położenia impulsu jest w tym przypadku zwykle mniejszy od 100 ps. Na rysunku 2.12 pokazano takie same impulsy jak na rysunku 2.11, po przejściu przez układ dyskryminatora ze stałym współczynnikiem tłumienia.



Rys. 2.12. Efekt działania układu dyskryminatora ze stałym współczynnikiem tłumienia
Fig. 2.12. Effect of constant fraction discriminator application

Zarówno wartość współczynnika tłumienia W_{CFD} , jak i wielkość opóźnienia D_{CFD} powinny być optymalizowane względem wartości stałych czasowych τ_1 i τ_2 (2.19) [37].

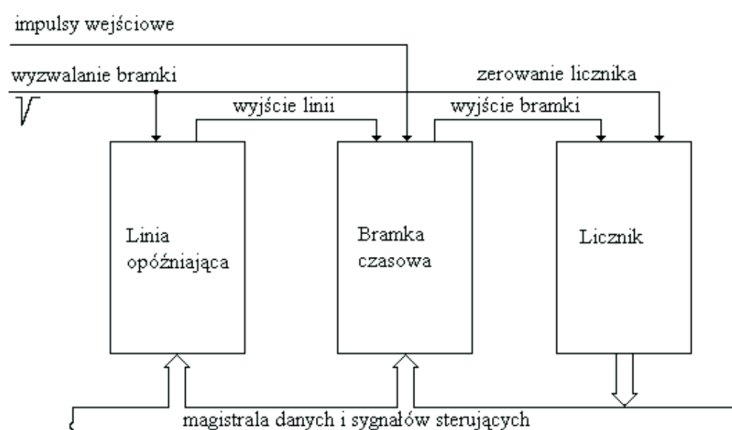
Dyskryminatory ze stałym współczynnikiem tłumienia produkowane komercyjnie mają na ogół ustalone parametry W_{CFD} i D_{CFD} , a optymalizacja tych parametrów może prowadzić do dalszej redukcji błędu określenia położenia impulsu. Należy podkreślić, że jakkolwiek czas opóźnienia D_{CFD} nie jest duży, ale czas trwania impulsu standardowego jest zwykle rzędu kilku ns, co ogranicza dopuszczalną intensywność procesu. Z tego też powodu ten sposób detekcji ma zastosowanie w systemach pomiarowych wysokorozdzielczych, w których intensywność procesu analizowanego jest zwykle ograniczona z innych powodów (np. czas interpolacji). W systemach pomiarowych o rozdzielczości nanosekundowej, w których dopuszcza się większą intensywność impulsów wejściowych, ten rodzaj detekcji zwykle nie jest stosowany.

3. POMIAROWY SYSTEM Z BRAMKĄ PRZESUWNĄ

Systemy zliczania impulsów w bramce czasowej stosowane w spektrometrii masy można podzielić na dwie grupy:

- systemy z pojedynczą, przesuwaną bramką czasową,
- systemy wielokanałowe, z wieloma bramkami czasowymi.

Najprostszymi systemami zliczania impulsów w bramce czasowej są systemy z bramką przesuwaną (ang. *Box-car*) [38–41]. W cyfrowym systemie z bramką przesuwaną opóźnienie otwarcia i szerokość bramki czasowej są realizowane cyfrowo [38, 39]. Schemat takiego systemu przedstawiono na rysunku 3.1 [6].



Rys. 3.1. Schemat cyfrowego systemu z bramką przesuwaną

Fig. 3.1. Diagram of the digital box-car system

Zasadniczymi elementami cyfrowego systemu są: programowana linia opóźniająca, programowana bramka czasowa oraz asynchroniczny licznik binarny. Istotnymi dla uzyskania wymaganych parametrów metrologicznych, aczkolwiek nie pokazanymi na schemacie, elementami systemu są wysokostabilny generator wewnętrzny oraz wejściowe układy analogowe, wpływające na kształt i poziom impulsów wejściowych i wyzwalających. Cyfrowy system z bramką przesuwaną może być zrealizowany bądź to jako autonomiczny system pomiarowy, bądź jako wirtualny system pomiarowy [38, 42–44].

W przypadku kiedy jest to system wirtualny, komputer pełni rolę graficznego interfejsu użytkownika, pamięci danych oraz steruje pracą poszczególnych modułów. Sygnał wyzwalający, oznaczający początek analizowanego procesu, otwiera bramkę po czasie równym ustalonemu programowo opóźnieniu. Po zamknięciu bramki stan licznika przepisywany jest do bufora danych, co oznacza również, że zakończył się kolejny cykl zliczania. Po odczytaniu danych następuje wykonanie kolejnego cyklu pomiarowego bądź wysłanie kolejnego sygnału wyzwalającego.

Jednym z najważniejszych elementów cyfrowego systemu jest cyfrowa, programowana linia opóźniająca. Zadaniem jej jest generacja impulsu otwierającego bramkę czasową po ustalonym czasie opóźnienia, liczonym zawsze od momentu pojawienia się impulsu wyzwalającego. Minimalna zmiana programowanego opóźnienia jest zależna od częstotliwości generatora wzorcowego. Zastosowanie w układzie linii opóźniającej licznika synchronicznego zmniejsza błąd nieliniowości układu.

W cyfrowym systemie z bramką przesuwną sygnał wyjściowy z programowanej linii opóźniającej steruje układem generacji bramki czasowej [42]. Impuls wejściowy otwierający bramkę pojawia się na wejściu układu sterującego bramki po czasie $\tau_D \pm \varepsilon_D$ (gdzie τ_D stanowi opóźnienie, a ε_D błąd graniczny linii opóźniającej) od momentu pojawienia się na wejściu linii opóźniającej impulsu wyzwalającego. Licznik układu bramkującego synchronizowany jest przez generator wzorcowy układu linii opóźniającej, co zapewnia całkowitą synchronizację układu linii opóźniającej i bramki czasowej.

Cyfrowe systemy z bramką przesuwą znajdują szerokie zastosowanie podczas analizy procesów niestacjonarnych o małej dynamice zmian funkcji intensywności [45–46]. Systemy te były stosowane w spektrometrii masowej czasu przelotu do połowy lat dziewięćdziesiątych ubiegłego wieku.

4. POMIAROWY SYSTEM WIELOKANAŁOWY

W przypadku kiedy badany proces analizować należy w stosunkowo długim czasie przy dużej dynamice zmian funkcji intensywności procesu, stosowanie systemu z pojedynczą bramką przesuwą staje się bardzo nieefektywne ze względu na małą wartość współczynnika wykorzystania cyklu pomiarowego W_{CP} . Jest on w przypadku systemu z bramką przesuwą równy

$$W_{CP} = \frac{\tau}{\tau_D + \tau + \tau_A}, \quad (4.1)$$

gdzie τ jest czasem otwarcia bramki, a τ_A jest czasem potrzebnym do przesłania i analizy danych w jednym cyklu pomiarowym.

Kiedy suma czasów $\tau_D + \tau_A \gg \tau$, wówczas współczynnik wykorzystania cyklu pomiarowego jest bardzo mały, a zgromadzenie określonej liczby danych wymaga dłuższego czasu. Z drugiej strony każdy cykl pomiarowy wymaga ponownego wzbudzenia badanego procesu. W wielu przypadkach dla zapewnienia stałych warunków prowadzenia eksperymentu badawczego liczba realizacji procesu musi być ograniczona. Zwiększenie współczynnika wykorzystania cyklu pomiarowego ma wówczas bardzo istotne znaczenie, ponieważ redukuje całkowity czas trwania eksperymentu i umożliwia przyjęcie założeń o stałych warunkach określających prowadzone badania.

Najprostszym systemem wielokanałowym z pojedynczym układem generacji bramek czasowych jest system z długim rejestrem przesuwным [47]. Szeregowy wyjście z rejestru ogranicza częstotliwość źródła wzbudzającego proces. Długość rejestru przesuwного jest ograniczona fluktuacjami fazowymi generatora zegarowego, który musi być doprowadzony do wszystkich elementów rejestru w ściśle określonym przedziale czasu. Podane trudności oraz koszt realizacji takich układów w postaci zintegrowanej stanowią poważną barierę dla ich stosowania.

Oprócz rejestru przesuwного i demultipleksa w układzie generacji bramek czasowych może być zastosowany generator pierścieniowy [48–52], pracujący w układzie z zamkniętą pętlą fazową. Układy generacji bramek czasowych, wykorzystujące generator pierścieniowy, stosują linię opóźniającą składającą się z inwerterów [49] lub wzmacniaczy sterowanych napięciem [48].

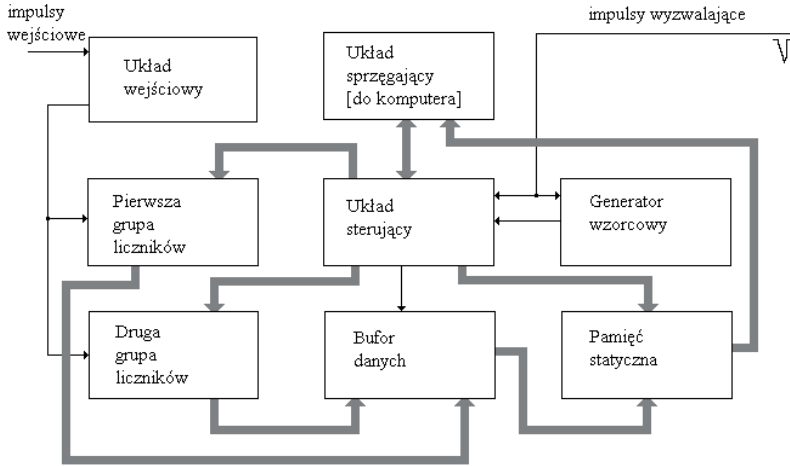
Zależność opóźnienia od wartości podawanego na wejście napięcia jest jednak silnie nieliniowa. Ogólnie trudno jest uzyskać wysokostabilne elementy linii opóźniającej o małych wartościach opóźnień. Opóźnienie pojedynczego elementu decyduje w tym przypadku o rozdzielczości systemu pomiarowego. Opóźnienia elementarne uzyskiwane w technologii CMOS są w granicach od 0,5 do 1,5 ns [50–52]. Stosowane w niektórych rozwiązaniach 32-elementowe lub dłuższe linie opóźniające wymagają dodatkowego dekodera lub stosowania specjalnej konfiguracji pamięci o wielobitowym dostępie. Poważnym problemem jest konieczność realizacji układów w wersji scalonej ASIC, co prowadzi do znacznych kosztów ich wytwarzania, biorąc pod uwagę niewielką skalę produkcji takich układów. Istnieje także możliwość zastosowania układów standardowych ECL, stanowiących programowane linie opóźniające z możliwością korekcyj opóźnienia. Rozdzielczość tych linii wynosi 20 ps, a zakres opóźnienia 2 ns. Istnieje możliwość zmiany tego opóźnienia w granicach 100 ps za pomocą zewnętrznego źródła napięciowego. Jednakże przy większej liczbie elementów opóźniających mogą pojawić się problemy z kompensacją przesunięć fazy sygnału zegarowego w poszczególnych miejscach układu.

4.1. SYSTEMY O ROZDZIELCZOŚCI NANOSEKUNDOWEJ

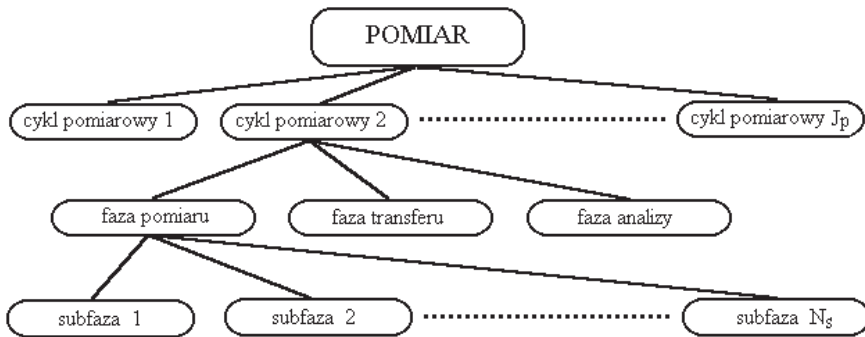
Wielokanałowy system zliczający z dwiema grupami liczników zapewnia bardzo dużą wartość współczynnika wykorzystania cyklu pomiarowego oraz dużą dopuszczalną intensywność procesu analizowanego [7, 9, 29, 38, 53–56].

Schemat systemu pomiarowego przedstawiono na rysunku 4.1 [7]. Układ wejściowy systemu zawiera wielostopniowy wzmacniacz–dyskryminator, którego zadaniem jest wzmocnienie impulsów wejściowych o amplitudzie większej od progu dyskryminatora. Wzmocnione impulsy są zliczane cyklicznie przez dwie grupy liczników. Proces zliczania w każdym cyklu pomiarowym rozpoczyna impuls wyzwalający, synchronizujący system pomiarowy z procesem badanym, a kończy zapełnienie ustalonego programowo obszaru pamięci, do której wpisywane są dane z po-

szczególnych liczników przez bufor danych. Układ sterujący generuje sygnały bramkujące, zeruje liczniki oraz przepisuje ich zawartość do pamięci statycznej. Algorytm pojedynczego cyklu pomiarowego zilustrowano na rysunku 4.2.



Rys. 4.1. Schemat systemu wielokanałowego o dwóch grupach liczników
Fig. 4.1. Block diagram of the multichannel system with double counters-unit

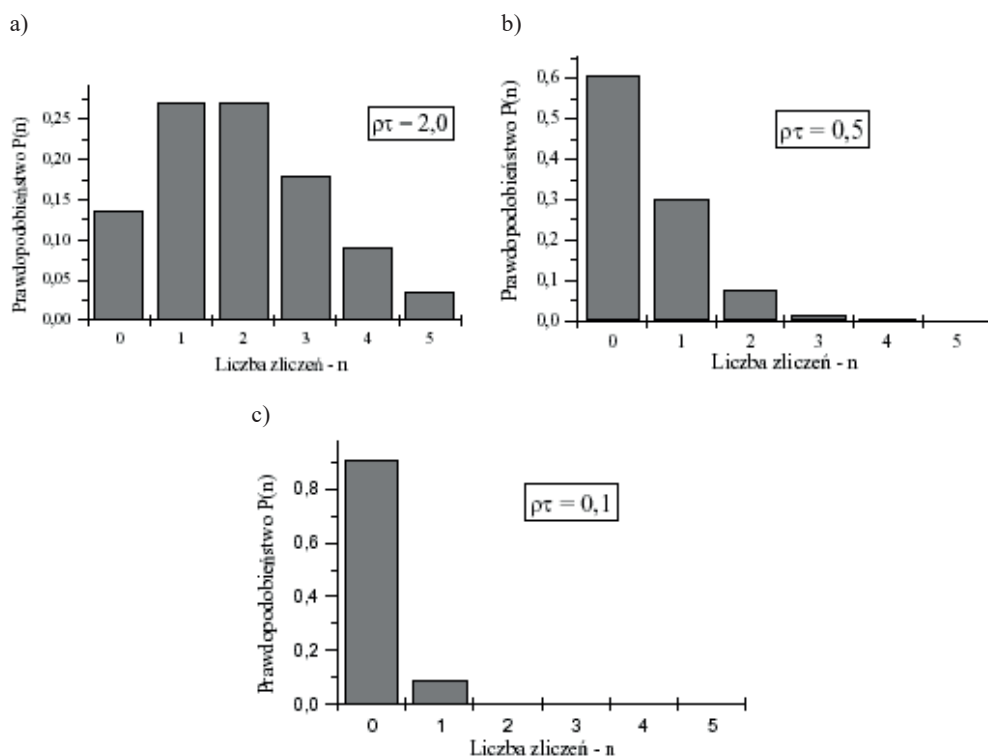


Rys. 4.2. Algorytm cyklu pomiarowego wielokanałowego systemu zliczającego
Fig. 4.2. Measurement cycle algorithm of multichannel counting system

Każdy pomiar składa się z założonej liczby cykli pomiarowych J_p , a każdy cykl pomiarowy składa się z faz: pomiaru, transferu i analizy, które system realizuje cyklicznie, aż do wykonania założonej liczby pełnych cykli pomiarowych. Liczba cykli pomiarowych J_p jest ustalana przez użytkownika w zależności od dynamiki badanego procesu i wymaganej dokładności określenia funkcji intensywności. W fazie transferu, dane są przesyłane przez łącze równoległe z pamięci statycznej do pamięci komputera, a w fazie analizy są one dekodowane i grupowane według kanałów. Liczba subfaz N_s pojedynczej

fazy pomiaru zależy od konstrukcji systemu pomiarowego. Czas transferu jednego 16-bitowego słowa z pamięci statycznej do pamięci komputera zależy od częstotliwości pracy magistrali komputera. Całkowity czas transferu podczas jednego cyklu pomiarowego zależy zatem zarówno od częstotliwości pracy magistrali komputera, jak też od liczby analizowanych kanałów czasowych. Czas analizy jest zależny od typu komputera, zastosowanej pamięci oraz liczby analizowanych kanałów czasowych. Ograniczenie liczby analizowanych kanałów czasowych umożliwia pracę systemu pomiarowego w trybie czasu rzeczywistego do częstotliwości źródła wzbudzenia równej 10 Hz.

W przypadku wielokanałowych systemów zliczających o dużej rozdzielczości szerokość generowanych bramek czasu jest rzędu kilku nanosekund, co sprawia, że oczekiwana liczba zliczeń dla jednego licznika, w jednym cyklu pomiarowym jest bardzo mała, co zilustrowano na rysunku 4.3.



Rys. 4.3. Rozkład prawdopodobieństwa liczby zliczeń
Fig. 4.3. Probability distribution of number of counts

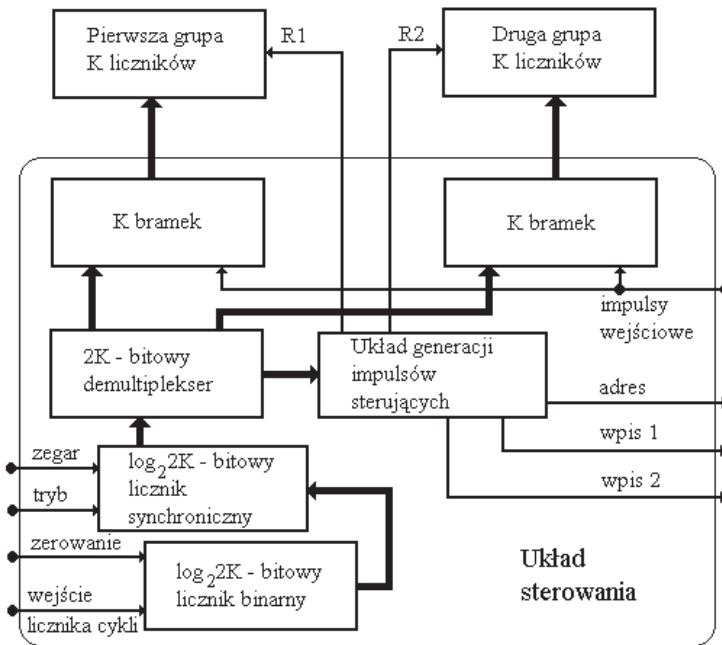
Przy szerokości bramki czasu równej $\tau = 5$ ns iloczyn $\rho\tau$ jest równy 0,1 dla intensywności $\rho = 2 \cdot 10^7 \text{ s}^{-1}$. Prawdopodobieństwo braku akceptacji impulsu przez licznik $P_{nz}(\rho, \Delta t')$ można obliczyć na podstawie zależności

$$P_{nz}(\rho; \Delta t') = 1 - \exp(-\rho \Delta t'). \quad (4.2)$$

Przyjmując $\Delta t' = 3$ ns, otrzymamy $P_{nz}(\rho; \Delta t') = 0,058$. Intensywność procesu ρ_{3dB} , przy której liczba impulsów zliczonych będzie o 3 dB mniejsza od impulsów na wejściu licznika, można wyznaczyć z zależności [57]

$$\rho_{3dB} = \frac{\ln \sqrt{2}}{t_{ef} + \Delta t}. \quad (4.3)$$

Próba zwiększenia intensywności badanego procesu w celu zwiększenia liczby zliczeń prowadzi do przekroczenia wartości ρ_{3dB} , równej w tym przypadku $\rho_{3dB} = 1,1 \cdot 10^8$ s⁻¹ i zmniejszenia liczby zliczeń na skutek wzrostu prawdopodobieństwa $P_{nz}(\rho; \Delta t')$. Jeżeli nie można zwiększyć liczby zliczeń w trakcie pojedynczego pomiaru w systemach z wąską bramką czasową, można ograniczyć pojemność licznika, co jest szczególnie istotne w systemie wielokanałowym.

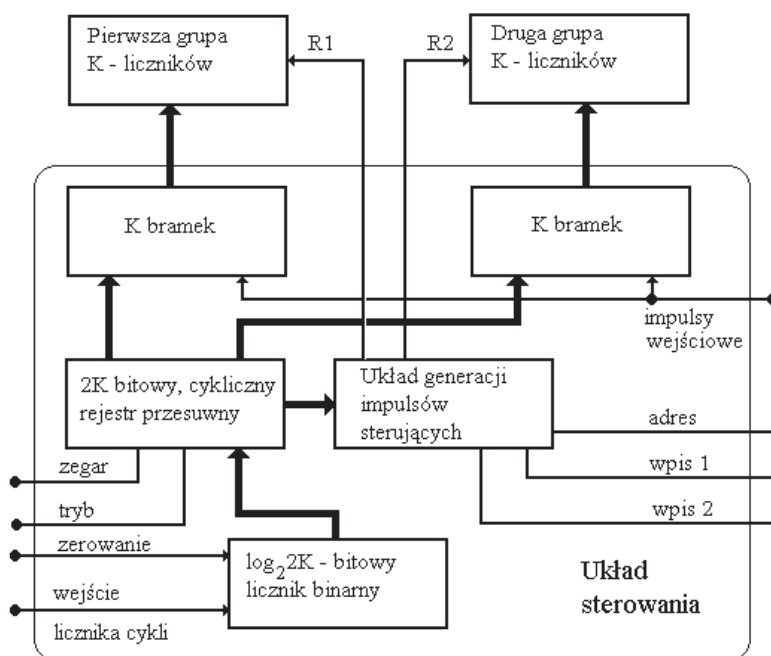


Rys. 4.4. Sposób generacji bramek czasowych z zastosowaniem demultipleksera
Fig. 4.4. Method of time-gates generation with demultiplexer

Z podanych względów w systemie wielokanałowym o rozdzielczości 5 ns zastosowano liczniki dwubitowe [7, 38], a w systemie o rozdzielczości 2,5 ns liczniki jednobitowe [54, 57]. Ograniczenie pojemności licznika wpływa oczywiście na do-

kładność pomiaru funkcji intensywności. W wielu przypadkach, kiedy maksymalna intensywność analizowanego procesu jest mała, błąd ten można pominąć [58]. Sposoby generowania sygnałów bramkujących w wielokanałowym systemie zliczającym o nanosekundowej rozdzielczości przedstawiono na rysunkach 4.4 i 4.5 [55, 57, 59].

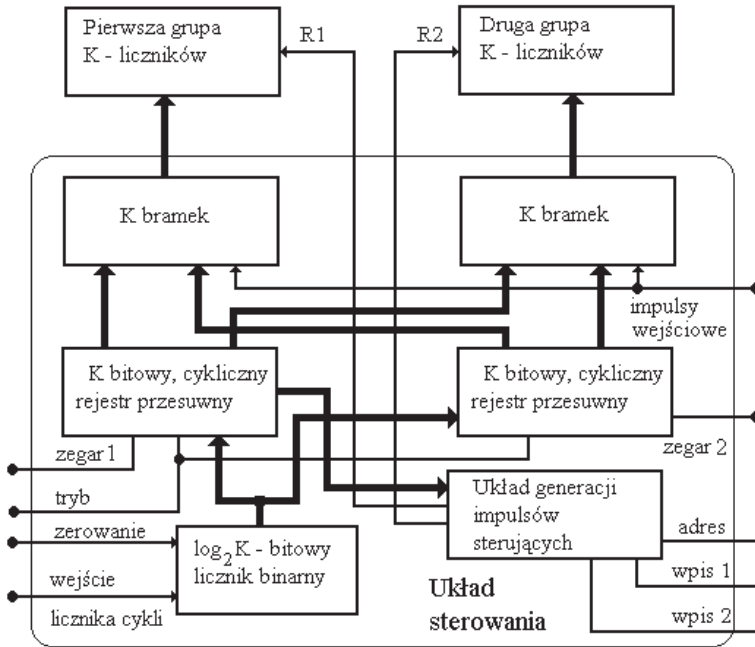
Dzięki zastosowaniu dodatkowego licznika binarnego oraz licznika synchronicznego lub rejestru przesunego z możliwością programowania stanu początkowego układy te zapewniają możliwość przydzielania różnych liczników i bramek do poszczególnych kanałów czasowych w kolejnych cyklach pomiarowych. Zastosowanie w układzie generacji bramek czasowych rejestru przesunego zmniejsza w znacznym stopniu różnice w szerokości poszczególnych bramek czasowych, zmniejszając tym samym prawdopodobieństwo zaliczenia impulsu wejściowego do niewłaściwego kanału czasowego.



Rys. 4.5. Sposób generacji bramek czasu z zastosowaniem rejestru przesunego

Fig. 4.5. Method of time-gates generation with shift register

W celu zwiększenia rozdzielczości systemu można zastosować wielofazowy układ generacji bramek czasu, którego schemat pokazano na rysunku 4.6 [57].

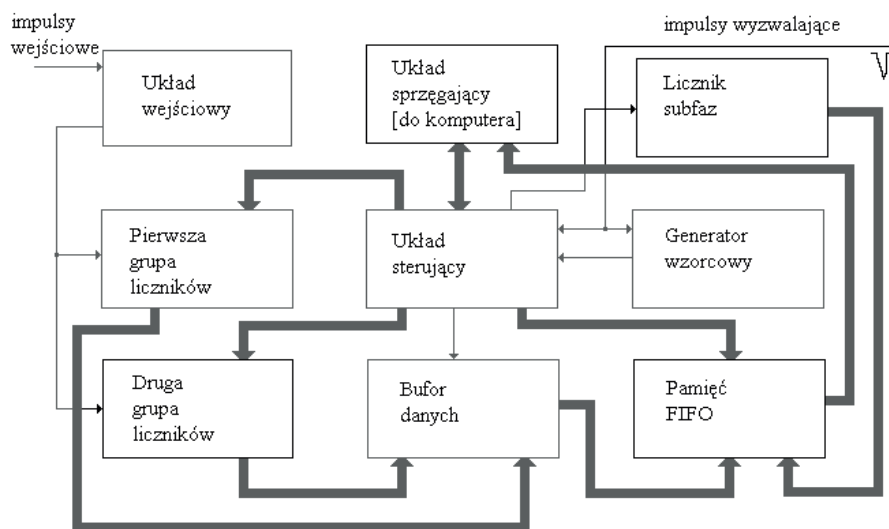


Rys. 4.6. Dwufazowy układ generacji bramek czasowych
 Fig. 4.6. Two-phase time-gates generation module

W systemie pomiarowym wykorzystującym dwufazowy układ generacji bramek czasowych, przy częstotliwości zegara równej 200 MHz, uzyskano rozdzielczość 2,5 ns.

Ograniczenie częstotliwości pracy źródła wzbudzenia, spowodowane czasem potrzebnym do wykonania operacji transferu i analizy danych można znacznie zredukować lub całkowicie wyeliminować. W przypadkach kiedy częstotliwość źródła wzbudzenia badanego procesu powinna być duża można ograniczyć czas transferu danych, rezygnując z zapamiętywania bajtów zerowych oraz dokonując transferu i analizy danych w tle aktualnie wykonywanych cykli pomiarowych [53]. Zapamiętywane są tylko stany liczników z tych subfaz cyklu pomiarowego, w których pojawił się chociaż jeden impuls na wejściu systemu pomiarowego.

Sposób ten wymaga jednakże zapamiętywania oprócz stanu liczników również numeru subfazy cyklu pomiarowego, co przedstawiono na rysunku 4.7. Ograniczenie liczby danych przechowywanych w pamięci buforowej systemu pomiarowego wpływa na zmniejszenie zarówno czasu transferu danych do pamięci komputera, jak i na zmniejszenie czasu analizy danych.



Rys. 4.7. Schemat wielokanałowego systemu zliczającego do analizy procesów wzbudzanych z dużą częstotliwością

Fig. 4.7. Block diagram of the multichannel system for high-frequency excited processes analysis

Ze względu na dużą dopuszczalną maksymalną intensywność badanego procesu, a tym samym możliwość zastosowania systemu pomiarowego do analizy procesów o dużej dynamice zmian funkcji intensywności, wielokanałowe systemy znalazły szczególnie duże zastosowanie w spektrometrii masowej [6, 9, 53].

4.2. SYSTEMY O ROZDZIELCZOŚCI PIKOSKUNDOWEJ

Rozdzielczość pikosekundowa wymagana jest w wielu przypadkach pomiarów funkcji intensywności procesów niestacjonarnych [49, 60–62, 63–66]. Do procesów takich zaliczyć można spektrometrię masową, niektóre procesy zaniku obsadzeń selektywnie wzbudzonych poziomów atomów, efekt sonoluminescencji oraz procesy zderzeń elektron–atom. Ze względu na trudności, jakie pojawiają się podczas stosowania generatorów wzorcowych o bardzo wysokiej częstotliwości [47] dla uzyskania pikosekundowej rozdzielczości systemów pomiarowych, stosuje się metody polegające na różnych sposobach interpolacji odcinka czasu. Stosowane w celu precyzyjnego określenia odcinka czasu metody polegają na zamianie czasu na inną wielkość fizyczną (napięcie) [63], na rozszerzeniu skali czasu [64–66] bądź też na wykorzystaniu metody noniusza elektronicznego. W tym ostatnim przypadku wykorzystuje się:

- generatory synchronizowane fazowo [67],
- linie opóźniające [68].

Wszystkie wymienione systemy pomiarowe zapewniają wysoką, pikosekundową rozdzielczość przy znacznym ograniczeniu dopuszczalnej intensywności impulsów. Ograniczenie to jest wynikiem konieczności przeznaczenia dość znacznych przedziałów czasu na interpolację mierzonych odcinków czasu.

Zamiana wielkości mierzonej, jaką jest czas, na napięcie prowadzi do możliwości pomiaru z pikosekundową rozdzielczością niewielkich odcinków czasu, rzędu dziesiątek i setek nanosekund. Czas uzyskania rezultatu jest w tym przypadku relatywnie mały, i zależy od szybkości zastosowanego wielokanałowego analizatora wysokości impulsów lub przetwornika A/C [63]. Zaletą systemów pomiarowych z przetwarzaniem czasu na napięcie jest ich relatywnie duża szybkość przy dużej, pikosekundowej rozdzielczości, a wadą – bardzo mały zakres pomiarowy.

Systemy pomiaru odcinka czasu z rozszerzeniem skali czasu są wykorzystywane od wielu lat [65, 66], przede wszystkim ze względu na prostą budowę układu pomiarowego. Ze względu na prostą konstrukcję, realizowane są chętnie w postaci układów scalonych [69]. Współczynnik rozszerzenia skali czasu w większości praktycznych zastosowań układu przyjmuje wartości od kilku do kilkuset razy [71]. Duża wartość tego współczynnika umożliwia uzyskanie dużej rozdzielczości, jednakże ogranicza dopuszczalną intensywność impulsów. Układ pomiaru odcinka czasu z rozszerzeniem skali czasu charakteryzuje mały zakres pomiarowy, rzędu dziesiątek nanosekund, który można zwiększyć tylko podczas wykorzystaniu wysokostabilnego wzorca czasu [70]. Zastosowanie dwóch generatorów w układzie umożliwia uzyskanie pikosekundowej rozdzielczości przy średnim czasie interpolacji równym 350 ns.

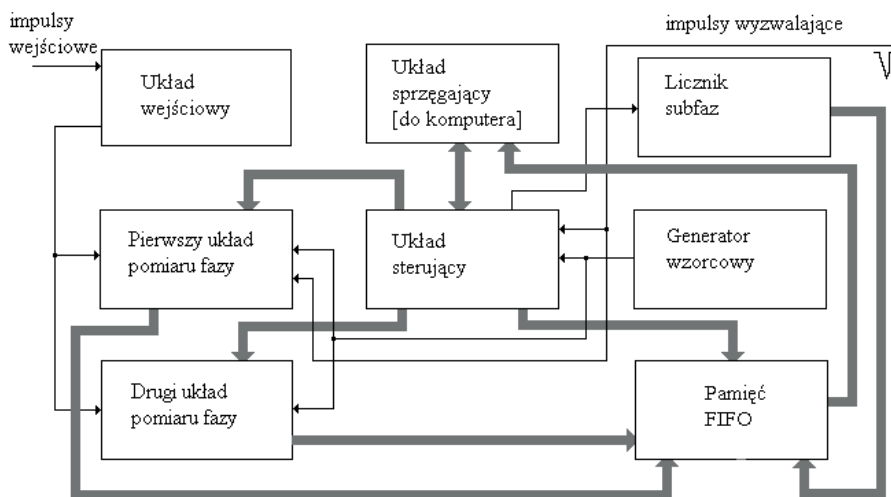
W pomiarach odcinka czasu również wykorzystywane są układy z noniuszem elektronicznym [72]. Realizacja systemu pomiarowego, w którym linie opóźniające wykonano w postaci połączonych szeregowo elementów o opóźnieniach różniących się o kilkadziesiąt do kilkuset pikosekund [65, 68, 73] umożliwiły znaczne zmniejszenie czasu interpolacji oraz uproszczenie układu pomiarowego. Podobne linie są wykorzystywane także w systemach generacji opóźnienia czasu [74]. Przy dużej rozdzielczości linii oraz relatywnie dużych wartościach opóźnień silnie rośnie czas interpolacji wraz ze wzrostem wartości mierzonego odcinka czasu, a układ pomiarowy wymaga dużej liczby elementów linii opóźniającej. Problem ten można częściowo wyeliminować, stosując układ dzielący mierzony odcinek czasu na trzy segmenty, z których tylko dwa, o ograniczonym (okresem zegara) zakresie, mierzone są metodą noniusza [75, 76].

Innym rozwiązaniem jest zastosowanie generatorów o niewielkiej różnicy częstotliwości synchronizowanych fazowo impulsami wejściowymi i wyzwalającymi [77, 78]. Opracowanie metody synchronizacji fazowej generatorów pracujących w układzie z zamkniętą pętlą fazową umożliwiło konstrukcję systemów pomiaru odcinka czasu o rozdzielczości lepszej niż 100 ps [67, 79]. Niestety dużej rozdzielczości interpolatorów towarzyszy ograniczenie dopuszczalnej intensywności impulsów wejściowych. Powielenie niezależnych kanałów interpolacji prowadzi wprawdzie do re-

dukcji tego ograniczenia, ale zwiększa jednocześnie stopień złożoności układu pomiarowego [80].

W przypadku systemów z bramką przesuwaną istnieje możliwość analizy procesów o dużej intensywności, lecz ograniczenie ich stosowania wynika głównie z małej efektywności wykorzystania cyklu pomiarowego oraz małej rozdzielczości.

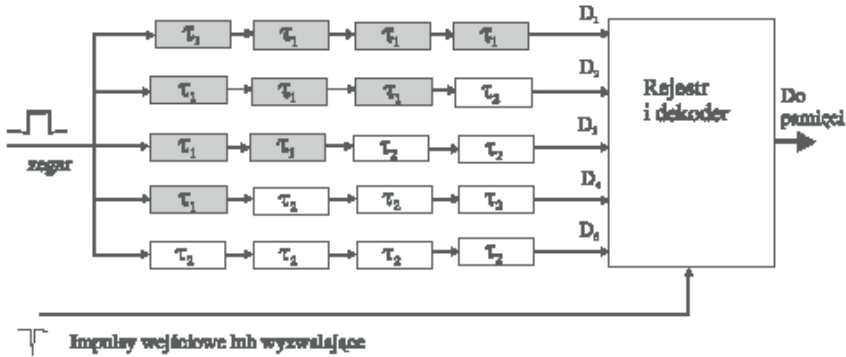
Zwiększanie rozdzielczości systemów wielokanałowych wykorzystujących zliczanie impulsów w bramkach czasowych, jest możliwe tylko w wyniku zastosowania jeszcze szybszych układów scalonych, zwiększenia stopnia integracji systemu pomiarowego oraz zwiększenia liczby faz zegara stosowanego w systemie pomiarowym. Zastąpienie pojedynczej grupy liczników układem pomiaru fazy generatora wzorcowego umożliwi zwiększenie rozdzielczości systemu pomiarowego z zachowaniem dużej dopuszczalnej intensywności analizowanego procesu [81]. Schemat takiego systemu pomiarowego przedstawiono na rysunku 4.8.



Rys. 4.8. Schemat systemu pomiarowego o pikosekundowej rozdzielczości

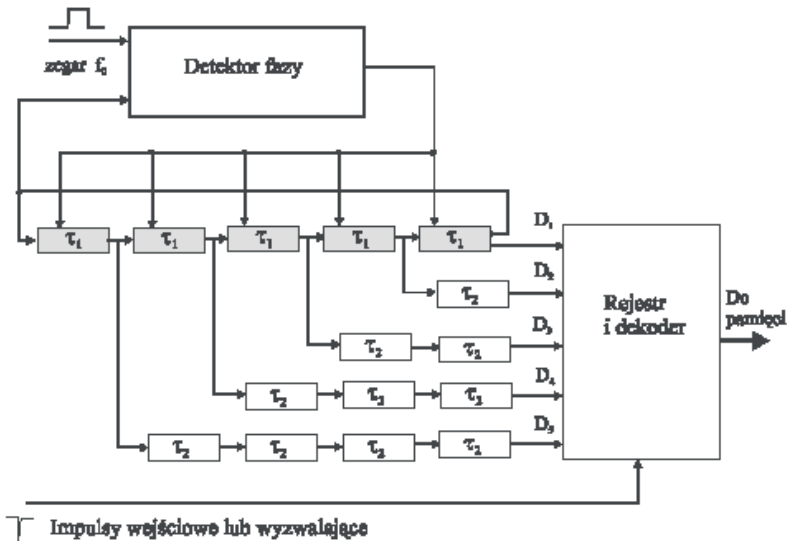
Fig. 4.8. Diagram of the measurement system of the resolution of order picoseconds

W przedstawianym systemie mierzona jest faza generatora wzorcowego w momencie pojawienia się impulsu wyzwalającego oraz impulsów wejściowych. W liczniku subfaz zliczana jest liczba pełnych cykli generatora od chwili pojawienia się impulsu wyzwalającego. Intensywność badanego procesu (impulsów na wejściu) nie jest w tym przypadku ograniczona przez całkowity czas potrzebny do zapisu danych do pamięci FIFO (kilka ns) ze względu na zastosowanie dwóch układów pomiaru fazy. System taki umożliwia analizę procesów o dużej dynamice funkcji intensywności, wzbudzanych z dużą częstotliwością. Pojedynczy układ pomiaru fazy, zawierający linie opóźniające o dwóch różnych wartościach opóźnień, przedstawia rysunek 4.9 [81].



Rys. 4.9. Schemat układu pomiaru fazy, wykorzystującego dwa rodzaje opóźnień
 Fig. 4.9. Diagram of phase measurement module with two different delay elements

W celu poprawy stabilności układu można stabilizować opóźnienia jednej linii, wykorzystując ją jako generator pierścieniowy pracujący w układzie zamkniętej pętli fazowej [48, 49, 82–85, 50–52]. Schemat układu pomiaru fazy z zastosowaniem sterowanej napięciem linii opóźniającej przedstawiono na rysunku 4.10.

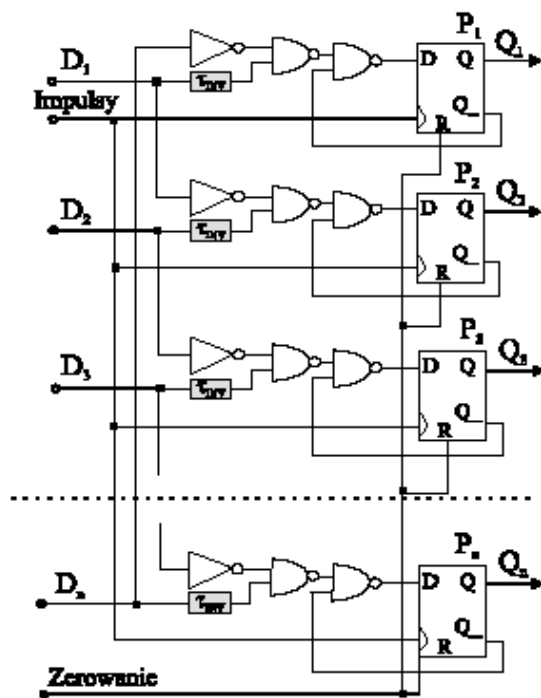


Rys. 4.10. Schemat układu pomiaru fazy ze stabilizowaną linią opóźniającą
 Fig. 4.10. Diagram of the phase measurement module with stabilized delay line

Zastosowanie takiej metody stabilizowania opóźnienia linii w układzie zamkniętej pętli fazowej wyklucza wykorzystanie standardowych układów PLD i FPGA do budowy linii opóźniających.

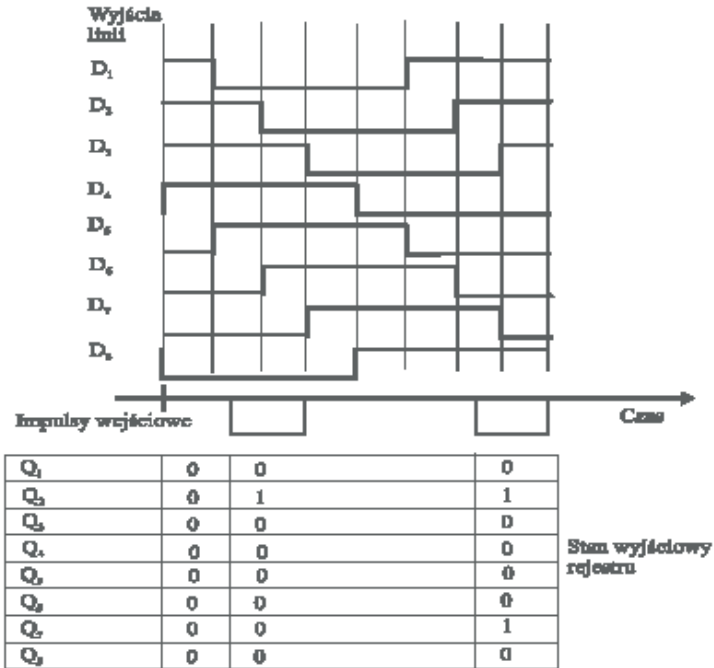
Istnieje jednakże możliwość zastosowania układów standardowych linii opóźniających wykonanych w technice ECL. Linie te umożliwiają zmianę opóźnienia zarówno przez programowanie (rozdzielczość wynosi 20 ps), jak też za pomocą zewnętrznego źródła napięciowego (w granicach 100 ps) [86].

Schemat rejestru, w którym przesunięcia fazowe zegara w momentach pojawiania się impulsów wejściowych są sumowane w czasie trwania pojedynczej subfazy cyklu pomiarowego przedstawiono na rysunku 4.11.



Rys. 4.11. Schemat rejestru układu pomiaru fazy
Fig. 4.11. Diagram of the phase measurement-module register

W celu wyeliminowania niepożądanego czasu martwego (jeden okres generatora), powstającego po impulsie wyzwalającym wskutek konieczności jednoznacznego określenia początku cyklu pomiarowego, można zastosować oddzielny rejestr do zapisu fazy generatora w chwili pojawienia się impulsu wyzwalającego lub wprowadzić dodatkowe opóźnienie (jeden okres generatora) dla impulsów wejściowych, wówczas impuls wyzwalający zostanie odseparowany od ciągu impulsów wejściowych, co umożliwi dokładne określenie początku cyklu pomiarowego. Sposób pracy rejestru zilustrowano na rysunku 4.12.



Rys. 4.12. Zależność stanu rejestru układu pomiaru fazy od momentów pojawiania się impulsów
 Fig. 4.12. States of the phase-measurement register versus input pulses positions

W przedstawionym systemie pomiarowym ograniczenie dopuszczalnej intensywności impulsów spowodowane jest głównie przez czas zapisu danych do rejestru oraz właściwości częstotliwościowe układu wejściowego, a zwłaszcza typu zastosowanych elementów i układów scalonych.

W układach programowanych trudno jest uzyskać elementy opóźniające o małych wartościach opóźnienia. Stosunkowo łatwo natomiast można uzyskać dwa elementy opóźniające, różniące się o kilkadziesiąt lub kilkaset pikosekund. W tym celu można wykorzystać właściwości podstawowych bloków logicznych układów CPLD [87].

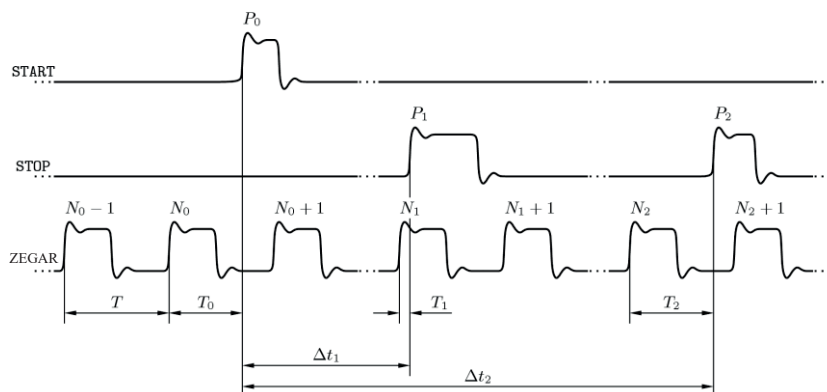
Wykorzystanie układów CPLD powoduje jednak dość duży błąd nieliniowości (ok. 200 ps) układu pomiarowego, wynikający z różnicy opóźnień poszczególnych elementów linii opóźniającej. Błąd ten, jako błąd systematyczny, może być korygowany [88, 89]. Jednym ze sposobów korekcji jest wykorzystanie sieci neuronowych [90].

Zastosowanie linii opóźniających zapewnia znaczny wzrost rozdzielczości systemu pomiarowego bez ograniczenia dopuszczalnej intensywności analizowanego sygnału. W układzie pomiaru fazy generatora można zastosować elementy opóźniające o bardzo małej różnicy opóźnień. W takim przypadku liczba elementów opóźniających musi być jednak duża, co prowadzi do wzrostu niepożądanych fluktuacji fazowych na wejściach rejestru.

5. SYSTEMY POMIAROWE IMPLEMENTOWANE W UKŁADACH FPGA

Zastosowanie układów FPGA nowej generacji o znacznie większym stopniu integracji umożliwia zwiększenie stopnia integracji całego systemu pomiarowego oraz dalszą poprawę parametrów metrologicznych [91–95]. Zastosowanie wielosegmentowej linii opóźniającej o regularnej strukturze wewnętrznej ma za zadanie zwiększyć rozdzielczość wyznaczania czasu przelotu [96–98]. O rozdzielczości układu decyduje opóźnienie pojedynczego segmentu wielosegmentowej linii opóźniającej (WLO). Informacja pochodząca z WLO zapisana jest najczęściej w pewnym dialekcie kodu termometrycznego i w celach optymalizacyjnych wymaga się jej kodowania do kodu naturalnego binarnego [95]. W ustalonym okresie zegara wzorcowego wzrost rozdzielczości będzie powodował wzrost liczby segmentów WLO. Im dłuższe słowo na wyjściu WLO, tym bardziej skomplikowana może się okazać konstrukcja konwertera kodu i tym samym czas jego konwersji będzie dłuższy [99].

Parametry metrologiczne układu pomiarowego, takie jak rozdzielczość, maksymalna intensywność rejestrowanych impulsów, nieliniowość pomiarowa decydują o dokładności pomiaru składu izotopowego wiązki atomowej [95, 100]. W układach wykorzystujących WLO w celu zmniejszenia wpływu zbyt dużego opóźnienia konwertera kodu na maksymalną intensywność rejestrowanych impulsów stosuje się jednocześnie proces przetwarzania danych pomiarowych. Jeżeli natomiast opóźnienia dostępnych elementów są niewystarczające (zbyt duże) do uzyskania żądanej rozdzielczości, można zastosować wielokrotne WLO (WWLO) [95]. Jeżeli charakterystyka uzyskanej WLO/WWLO jest niezadowalająca (nieliniowości są zbyt duże), można dokonać jej linearyzacji przez podłączenie do połączeń między segmentami pasożytniczych pojemności tak, aby zwiększyć stałą czasową obwodu [95].



Rys. 5.1. Wyznaczania odcinka czasu
Fig. 5.1. Time-interval determining

Na rysunku 5.1 przedstawiono sposób wyznaczania odcinka czasu pomiędzy rozpoczęciem a zakończeniem przelotu jonu. Zdarzenie P_0 reprezentuje jonizację wiązki atomowej i rozpoczęcie przelotu jonu/jonów, zdarzenia P_1 , P_2 i kolejne oznaczają moment zakończenia przelotu, czyli dotarcie jonu do detektora. Momenty wystąpienia zdarzeń nazwano stemplami czasowymi. Odcinek czasu wyznaczany jest jako różnica dwóch stempli czasowych [95, 101].

Wartość każdego stempla czasowego określana jest względem zegara wzorcowego ZEGAR. Liczbę pełnych okresów zegara wzorcowego, które pojawiły się od momentu inicjalizacji modułu, wyznacza się, stosując zliczanie impulsów zegarowych, natomiast przesunięcie T_i pomiędzy końcem okresu zegara wzorcowego a zboczem narastającym rejestrowanego impulsu (START lub STOP) wyznacza się stosując wielosegmentowe linie opóźniające (WLO).

Wartość stempla czasowego dla zdarzenia P_i wynosi

$$t_i = N_i T + T_i, \quad i \geq 0, \quad (5.1)$$

gdzie T jest okresem zegara wzorcowego.

Odcinek czasu Δt_i (czas przelotu) pomiędzy dwoma zdarzeniami P_0 (jonizacja) i P_i (zakończenie przelotu), gdzie $i > 0$ wynosi

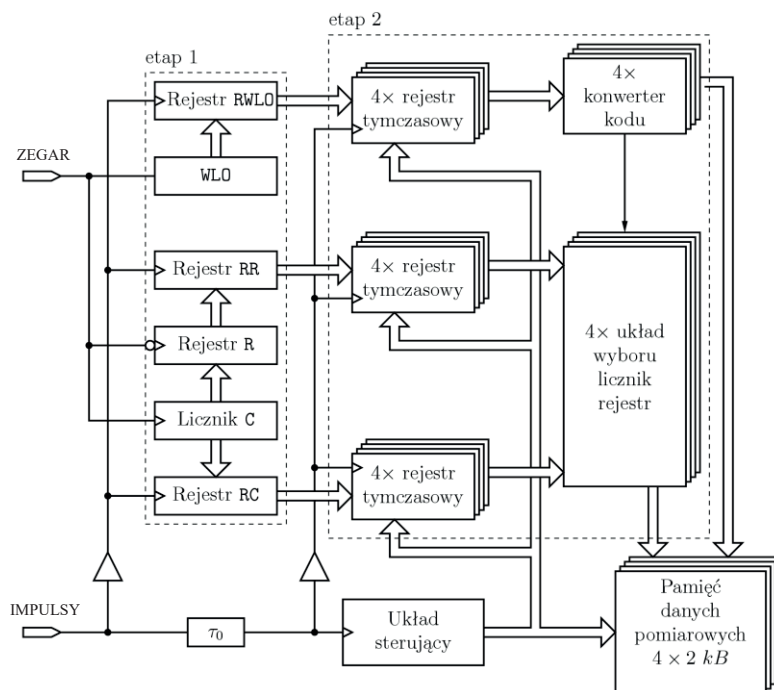
$$\Delta t_i = t_i - t_0 = (N_i - N_0)T + T_i - T_0. \quad (5.2)$$

Na rysunku 5.2 przedstawiono schemat blokowy modułu pomiaru odcinka czasu. Sygnał IMPULSY jest złożeniem sygnałów START i STOP w sposób, który zostanie zaprezentowany później. Zadaniem modułu jest rejestrowanie stempli czasowych dla zdarzeń reprezentowanych przez narastające zbocza sygnału IMPULSY względem wzorcowego sygnału zegarowego ZEGAR. WLO oraz rejestr RWLO służą do wyznaczenia przesunięcia T_i , natomiast licznik C i rejestr R wraz z rejestrami RR i RC służą do wyznaczenia wartości N_i .

Zastosowanie wzajemnie uzupełniającej się pary licznik–rejestr jest konieczne ze względu na asynchroniczność impulsów na wejściu IMPULSY względem sygnału zegarowego ZEGAR. Każde narastające zbocze sygnału ZEGAR powoduje inkrementację stanu licznika C, natomiast każde zbocze opadające tego sygnału powoduje przepisanie stanu licznika C do rejestru R. W przypadku pojawienia się zbocza narastającego na wejściu IMPULSY podczas inkrementacji licznika C, informacji o liczbie pełnych okresów zegara N_i może dostarczyć rejestr R.

Każde zbocze narastające sygnału IMPULSY powoduje zapisanie stanu WLO do rejestru RWLO, stanu licznika C do rejestru RC oraz stanu rejestru R do rejestru RR. Dla systemu o rozdzielczości 500 ps, który został zbudowany z wykorzystaniem 32-bitowej WLO i zakresie pomiarowym rozszerzonym przez 27-bitowy licznik powstaje wówczas 96-bitowy stempel czasowy, który cechuje się nadmiarowością informacji oraz nieoptymalnością jej kodowania. Informacja o przesunięciu T_i zapisana jest w kodzie pseudoter-

mometrycznym, natomiast informację o liczbie pełnych okresów zegara wzorcowego cechuje redundancja.



Rys. 5.2. Schemat blokowy modułu pomiaru odcinka czasu

Fig. 5.2. Block Diagram of time-interval measuring module

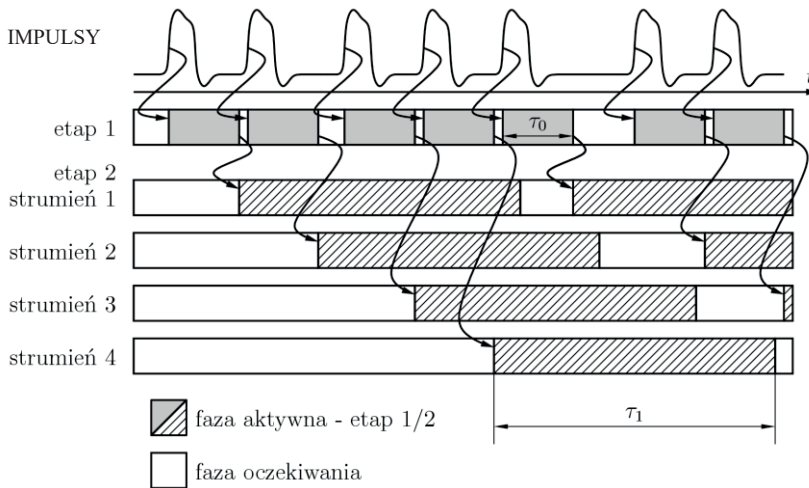
Trzy grupy rejestrów tymczasowych (cztery rejestry tymczasowe w grupie) zostały zaimplementowane w celu jak najszybszego odbierania informacji pochodzącej z rejestrów RC, RR i RWLO i następnym przekazywaniu tej informacji do bloków konwersji kodu i bloków wyboru licznik–rejestr. Rejestry tymczasowe są jednocześnie rejestrami wejściowymi bloków konwersji kodu i wyboru licznik–rejestr, stąd czas niezbędny na dokonanie konwersji kodu i czas niezbędny na dokonanie wyboru licznik–rejestr decydują bezpośrednio o maksymalnej intensywności rejestrowanych impulsów. Przez cały czas wykonywania konwersji lub wyboru stanu rejestru tymczasowego nie może zostać zmieniony, a więc rejestracja impulsu jest niemożliwa, co ogranicza maksymalną intensywność rejestrowanych impulsów. Zwielenokrotnienie liczby bloków konwertujących i bloków wyboru umożliwia znacząco zwiększyć maksymalną intensywność rejestrowanych impulsów. W zrealizowanym układzie czterokrotne zwiększenie liczby bloków konwertujących i wybierających umożliwiło uzyskanie ponad 2,5-krotnie większej maksymalnej intensywności rejestrowanych impulsów. Dla struktury programowalnej VIRTEX XCV300 [102] maksymalna inten-

sywność rejestrowanych impulsów została zwiększona ze $100 \cdot 10^6$ impulsów na sekundę do ponad $266 \cdot 10^6$ impulsów na sekundę.

Ideę zrównoleglenia procesu przetwarzania zilustrowano na rysunku 5.3. Każde zbocze narastające sygnału IMPULSY powoduje uaktywnienie na czas τ_0 elementów etapu 1. Po czasie τ_0 , który wyznaczany jest przez element o opóźnieniu τ_0 (rys. 5.2) następuje przekazanie zadania przetwarzania do jednego z czterech bloków etapu 2. Czas przetwarzania w etapie 2 wynosi τ_1 . W przypadku spełnienia kryterium

$$\frac{\tau_1}{n} \leq \tau_0, \quad n = 4, \quad (5.3)$$

gdzie n jest liczbą strumieni etapu 2, maksymalna intensywność rejestrowanych impulsów ograniczona jest przez czas przetwarzania etapu 1.

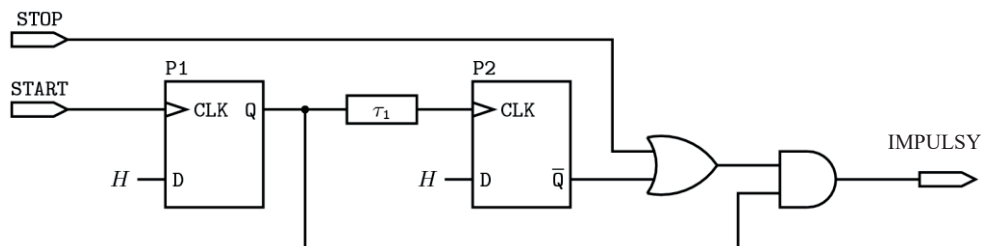


Rys. 5.3. Idea zrównoleglenia procesu przetwarzania danych pomiarowych
Fig. 5.3. Idea of parallelization data measuring process

Zadaniem bloku wejściowego jest połączenie sygnałów START i STOP w jeden wspólny sygnał IMPULSY. Połączenie tych sygnałów umożliwia rejestrację stempla czasowego dla START(u) i stempli czasowych dla STOP(u/ów) z wykorzystaniem jednego toru pomiarowego. Takie rozwiązanie minimalizuje ilość wykorzystanych zasobów struktury programowalnej oraz ułatwia proces kalibracji i usuwania błędów.

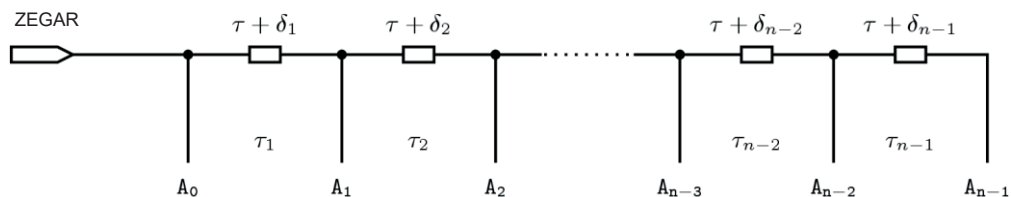
Pierwsze zbocze narastające na wejściu START powoduje ustawienie stanu przerzutnika P1, co skutkuje ustawieniem stanu wysokiego na wyjściu IMPULSY. Przez czas τ_1 stan wysoki na wyjściu utrzymuje się niezależnie od stanu na wejściu STOP. Po czasie τ_1 stan przerzutnika P2 zostaje ustawiony i o stanie wyjścia IMPULSY decyduje stan sygnału STOP.

Opóźnienie τ_1 zostało dobrane eksperymentalnie. Zbyt długi impuls wpłynie niekorzystnie na maksymalną intensywność rejestracji impulsów, które się pojawią tuż po wyzwoleniu, natomiast zbyt krótki impuls po przejściu przez elementy opóźniające może zostać rozmyty, ze względu na różny czas narastania i opadania sygnałów wewnątrz struktury FPGA.



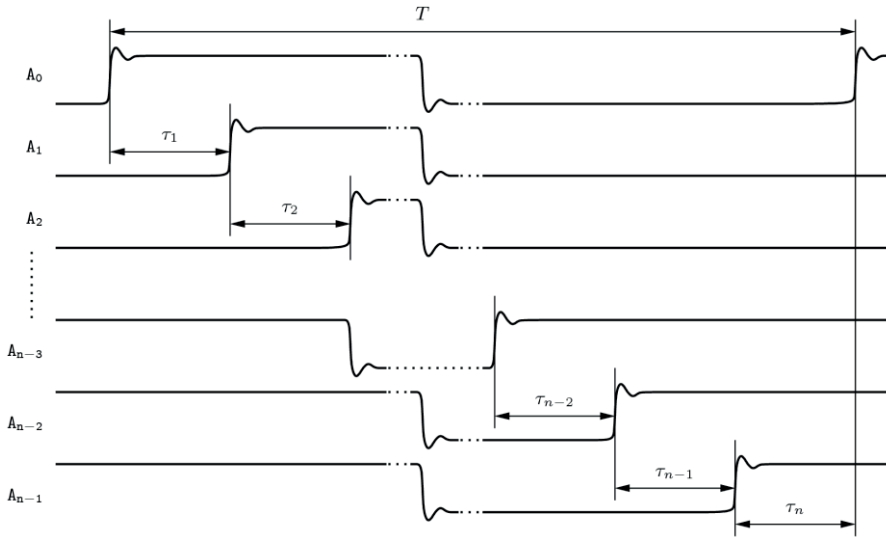
Rys. 5.4. Sposób dokonania połączenia sygnału wyzwalającego (START) z sygnałem impulsów (STOP)
Fig. 5.4. The method of connecting triggering-signal (START) and pulses-signal (STOP)

Licznikowa metoda wyznaczania wartości stempla czasowego pozwala uzyskać rozdzielczość pomiaru równą okresowi zegara wzorcowego T . Gdy wypełnienie sygnału zegarowego jest bliskie 0,5, wówczas możliwe jest wyznaczenie wartości stempla czasowego z rozdzielczością $T/2$. Zastosowanie wielosegmentowych linii opóźniających (WLO) pozwala uzyskać wielofazowy sygnał zegarowy, co daje możliwość zwielokrotnienia rozdzielczości. Zastosowanie n -segmentowej linii opóźniającej umożliwia n -krotne zwiększenie rozdzielczości. Na rysunku 5.5 przedstawiono schemat WLO o nominalnym opóźnieniu pojedynczego segmentu równym τ .



Rys. 5.5. Schemat wielosegmentowej linii opóźniającej
Fig. 5.5. Scheme of multi-tap-delay-line

Aby wykorzystanie WLO w systemie pomiaru odcinka czasu było optymalne, suma wszystkich $(n - 1)$ opóźnień powinna wynosić $T - \tau$. Na wyjściu WLO uzyskuje się zespół sygnałów (rys. 5.6), w których następny $(i + 1)$ jest opóźniony względem poprzedniego (i) o τ . Zatrzaśnięcie stanu wyjść WLO w momencie zajścia zdarzenia pozwala ocenić stopień propagacji zbrocza narastającego sygnału CLOCK poprzez WLO, co w efekcie umożliwia ustalenie odcinka czasu T_i (rys. 5.1).



Rys. 5.6. Sygnały wyjściowe wielosegmentowej linii opóźniającej
 Fig. 5.6. Output signals of multi-tap-delay-line

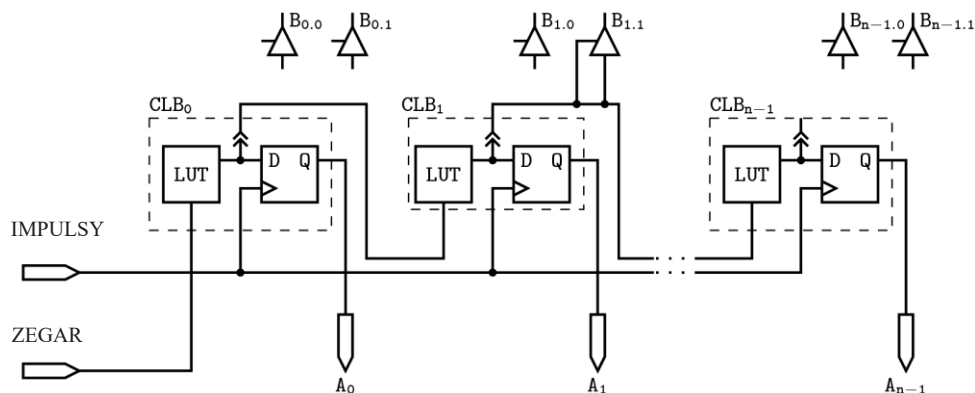
Charakterystyka rzeczywistej WLO jest nieliniowa, a odchylenia od wartości nominalnej δ_i , $i = [1, n - 1]$ mogą wynosić nawet połowę wartości opóźnienia nominalnego. W zależności od sposobu implementacji WLO stosuje się wiele różnych technik linearyzacji charakterystyki WLO.

Implementacja WLO do struktury programowalnej FPGA może polegać na wykorzystaniu tablic LUT (ang. *Look-up-table*) [102] w celu uzyskania pojedynczych opóźnień τ (rys. 5.7). Charakterystyka WLO zaimplementowanej do struktury FPGA może zostać zlinearyzowana przez podłączenia do wyjść bloków konfigurowalnych odpowiedniej liczby niewykorzystanych wejść komponentów struktury FPGA [102]. Bufory trójstanowe (rys. 5.7) podłączone do wyjść bloków CLB powodują wzrost stałej czasowej, co prowadzi do zwiększenia czasu narastania sygnału i w efekcie powoduje wzrost czasu propagacji zbocza narastającego. Podłączenie obu wejść (sterującego i danych) bufora do sygnału łączącego dwa bloki CLB powoduje średnie zwiększenie opóźnienia, w przypadku struktury XCV300 PQ240, o około 15 ps.

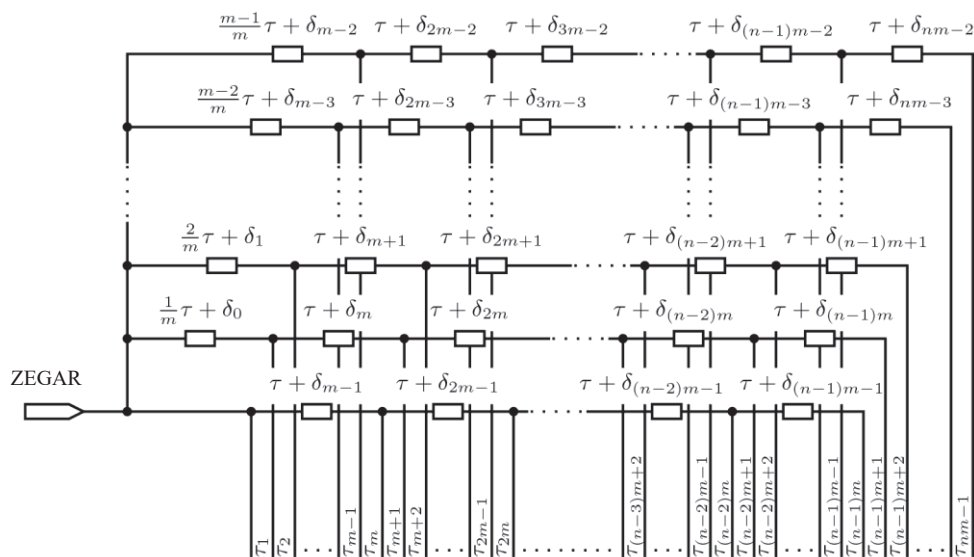
Stosowanie WLO w systemach pomiaru odcinka czasu umożliwia uzyskanie znaczącej poprawy rozdzielczości. Rozdzielczość ta jest jednak ograniczona przez minimalne regularnie dostępne w strukturze opóźnienia. Dalszy wzrost rozdzielczości można uzyskać, stosując wielokrotną WLO (rys. 5.8). Zastosowanie podwójnej WLO powoduje dwukrotne zwiększenie rozdzielczości, z zastosowaniem m -sztuk linii WLO następuje m -krotny wzrost rozdzielczości.

W projektowaniu WWLO odchylenia opóźnień pojedynczych segmentów składowej WLO mają znacznie większy wpływ na względne (względem rozdzielczości)

błędy nielineowości. Dodatkowo, jeżeli błąd nielineowości różnicowej INL dla i -tego wyjścia przekracza wartość $T/2$ mn, to może nastąpić wygenerowanie niewłaściwego kodu na wyjściu WWLO. Będzie to przypadek, w którym zbocze narastające na wyjściu i pojawi się później niż zbocze narastające na wyjściu $i + 1$. W projektach z użyciem WWLO stosowanie linearyzacji charakterystyki składowych WLO, w sposób przedstawiony na rysunku 5.7, jest już konieczne dla $m \geq 4$.



Rys. 5.7. Implementacja wielosegmentowych linii opóźniających w strukturze FPGA
Fig. 5.7. Implementation of multi-tap-delay-line into FPGA structure



Rys. 5.8. Model wielokrotnej wielosegmentowej linii opóźniającej
Fig. 5.8. Model of multiple-multi-tap-delay-line

Każda WWLO składa się z m sztuk, łatwych do implementacji WLO (linie poziome – rysunek 5.8) oraz jednej wysokorozdzielczej (pionowej) WLO o opóźnieniu pojedynczego segmentu równym τ/m . Niewątpliwą zaletą WWLO jest to, że większość składowych opóźnień uzyskiwana jest z łatwo dostępnych w strukturze FPGA bloków, a poważne problemy konstrukcyjne sprawia wyłącznie wysokorozdzielcza pionowa linia opóźniająca.

Opóźnienie τ_i pojedynczego segmentu m -krotnej WLO wynosi

$$\tau_i = \frac{\tau}{m} + \sum_{j=0}^{\left\lfloor \frac{i+m-2}{m} \right\rfloor + \left\lfloor \frac{i+m-1}{m} \right\rfloor - 1} (-1)^j \delta_{(i-1-j \bmod 2) \bmod m + \left\lfloor \frac{j}{2} \right\rfloor m}, \quad (5.4)$$

gdzie mod jest operacją zwracającą resztę z dzielenia, a operator $\lfloor \cdot \rfloor$ zwraca część całkowitą liczby rzeczywistej. W przypadku szczególnym przy implementacji podwójnej WLO opóźnienie pojedynczego segmentu wynosi

$$\tau_i = \frac{\tau}{2} + \sum_{j=0}^{i-1} \delta_j (-1)^{j+i+1}, \quad i \in 1, 2n-1. \quad (5.5)$$

Wysokorozdzielczą (pionową) WLO można uzyskać, wykorzystując elementy służące do implementacji przeniesień dla operacji arytmetycznych (CARRY) w układach FPGA. Opóźnienie pojedynczego elementu CARRY wynosi dla struktury XCV300 około 70 ps i może zostać wystarczająco precyzyjnie wyregulowane przez podłączenie określonej liczby buforów trójstanowych.

Implementacja WLO, której wszystkie segmenty zbudowane są w ten sposób, jest również możliwa, ale wymaga stosunkowo dużych struktur programowalnych oraz zegara wzorcowego o wysokiej częstotliwości. Na przykład 64-segmentowa WLO wymaga zegara wzorcowego o częstotliwości ponad 220 MHz i struktury programowalnej, która będzie miała przynajmniej 64 wiersze bloków konfigurowalnych.

Cały moduł pomiarowy można opisać w języku VHDL [103, 104]. Większość bloków modułu została opisana z wykorzystaniem standardowym konstrukcji tego języka. Warto skomentować jednak sposób implementacji WLO, która została uzyskana z wykorzystaniem tablic LUT zawartych w blokach CLB. Aby użyć elementów LUT do budowy segmentów WLO, zastosowano konstrukcję generic/port map języka. Opis 32-segmentowej WLO wygląda następująco:

```
t00: lut1 generic map("10") port map(clock, s);
p00: lut1 generic map("10") port map(s, sa(0));
p02: lut1 generic map("10") port map(sa(0), sa(1));
p04: lut1 generic map("10") port map(sa(1), sa(2));
...
p28: lut1 generic map("10") port map(sa(13), sa(14));
```

```

p30: lut1 generic map("10") port map(sa(14), sa(15));
p01: lut1 generic map("10") port map(s, sb(0));
p03: lut1 generic map("10") port map(sb(0), sb(1));
p05: lut1 generic map("10") port map(sb(1), sb(2));
...
p29: lut1 generic map("10") port map(sb(13), sb(14));
p31: lut1 generic map("10") port map(sb(14), sb(15));

```

Konstrukcja *generic map* umożliwia przekazanie pełnionej przez LUT funkcji logicznej (w tym przypadku bufor), natomiast przez port map przekazano informację o połączeniu sygnałów. Element t00 ma za zadanie zróżnicować opóźnienia dla wejść obu składowych WLO. Jego położenie zostało dobrane eksperymentalnie tak, aby różnica opóźnień pomiędzy wejściem górnej i wejściem dolnej WLO wynosiła 500 ps. W ten sposób uzyskano wysokorozdzielczą pionową WLO. Elementy pxx o numerach parzystych xx opisują górną, natomiast elementy o numerach nieparzystych dolną WLO.

Aby określić położenie w strukturze segmentów WLO, do projektu został dołączony plik UCF z ograniczeniami użytkownika (ang. *User Constraints File*), który wygląda następująco

```

INST "t00" LOC="CLB_R22C23.S0";
INST "p00" LOC="CLB_R24C26.S1";
INST "p02" LOC="CLB_R24C27.S1";
INST "p04" LOC="CLB_R24C28.S1";
...
INST "p28" LOC="CLB_R24C40.S1";
INST "p30" LOC="CLB_R24C41.S1";
INST "p01" LOC="CLB_R18C26.S1";
INST "p03" LOC="CLB_R18C27.S1";
INST "p05" LOC="CLB_R18C28.S1";
...
INST "p29" LOC="CLB_R18C40.S1";
INST "p31" LOC="CLB_R18C41.S1";.

```

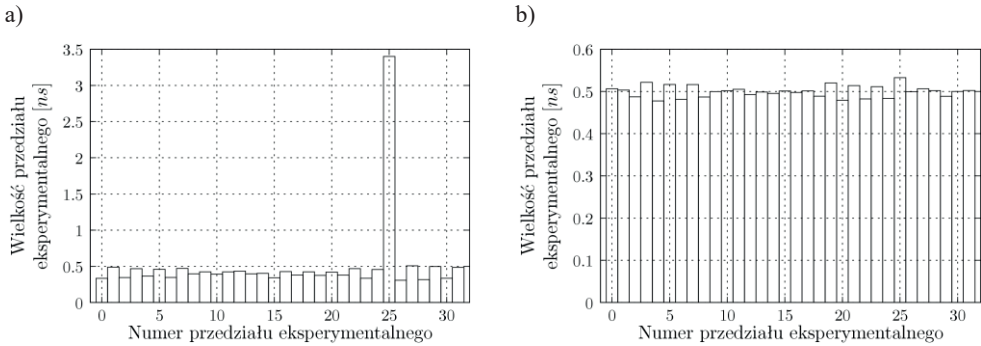
Na przykład notacja R22C23.S0 oznacza 22. wiersz, 23. kolumnę i zerowy blok konfiguracyjny w wybranej grupie (dla struktury XCV-300 występują dwa bloki CLB w grupie).

Charakterystykę modułu pomiarowego wyznacza się w sposób statystyczny. Wykorzystuje się tutaj test gęstości stanów. Na wejście STOP wprowadza się asynchroniczny w stosunku do zegara wzorcowego CLOCK sygnał zegarowy. Każde zbocze narastające tego sygnału powoduje rejestrację stempla czasowego, z którego w teście odzyskuje się wyłącznie informację o przesunięciu T_i (rys. 5.1). Im opóźnienie seg-

mentu jest większe, tym liczba zliczonych stempli czasowych z indeksujących ten segment jest większa. Oznaczając jako N liczbę wszystkich rejestracji, natomiast przez N_i liczbę stempli indeksujących i segment, opóźnienie τ_i może zostać opisane zależnością

$$\tau_i = T \frac{N_i}{N}, \quad (5.6)$$

gdzie T jest okresem zegara wzorcowego.



Rys. 5.9. Charakterystyki modułu pomiarowego: a – bez dokonania procesu linearyzacji; b – po dokonaniu linearyzacji wielosegmentowej linii opóźniającej

Fig. 5.9. Measuring module characteristics: a – without linearization process; b – after the linearization process has been applied

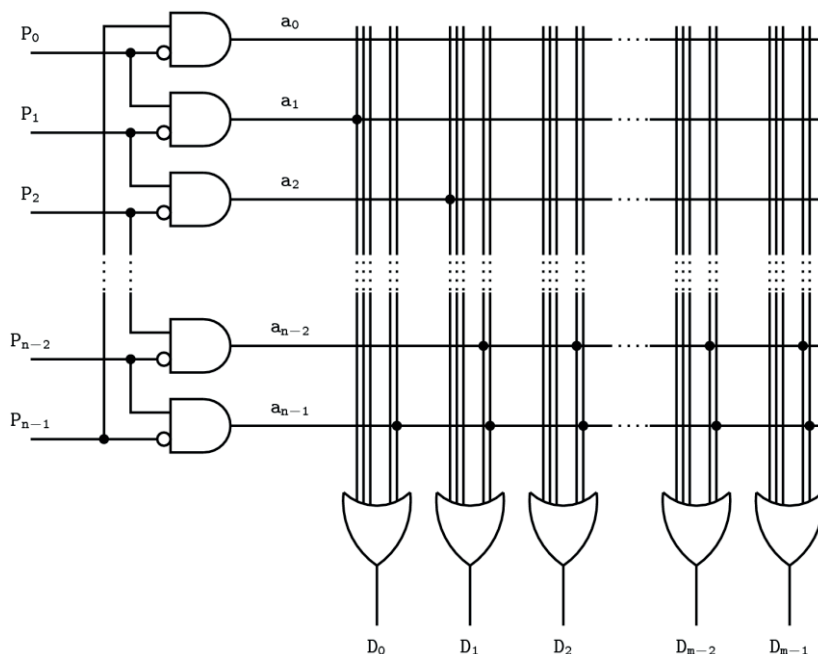
Na rysunku 5.9a przedstawiono charakterystykę modułu o rozdzielczości 500 ps, zbudowanego z wykorzystaniem 32-segmentowej podwójnej linii opóźniającej, dla której nie wykonano procesu linearyzacji. Po dokonaniu linearyzacji, która polegała na odpowiednim przyłączeniu około 1200 wejść buforów trójstanowych charakterystyka (rys. 5.9b) znacznie się poprawiła.

Innym sposobem dokonania linearyzacji mogłoby być zmniejszenie okresu zegara wzorcowego tak, aby zmniejszyć wartość τ_{25} (rys. 5.9a). Rozwiązanie to wymagałoby jednak zastosowania układu syntezującego, który spowodowałby zwiększenie fluktuacji fazy zegara, co niekorzystnie wpłynęłoby na parametry metrologiczne modułu. W module tym okres zegara T wynosi 62,5 ns, co przy 32 segmentach daje 500 ps opóźnienia na segment. W przypadku zmniejszenia okresu jednostka rozdzielczości uległaby pogorszeniu.

Standardowy opis w języku VHDL (wykorzystanie konstrukcji *when-else*) do opisu konwertera kodu pseudotermometrycznego na kod naturalny binarny doprowadził do powstania bloku o czasie propagacji równym około 32 ns (struktura XCV300). Zastosowanie go w module pomiarowym spowodowałoby znaczne ograniczenie maksymalnej intensywności rejestrowanych impulsów lub konieczne byłoby zastosowanie

ośmiu konwerterów kodu tak, aby spełnić kryterium (5.3) Implementacja kolejnych konwerterów doprowadziłaby do zwiększenia ilości wykorzystanych zasobów struktury FPGA.

Aby zapewnić dużą (ponad $250 \cdot 10^6$ impulsów/sekundę) intensywność rejestrowanych impulsów, dokonano optymalizacji konstrukcji konwertera kodu. W konwerterze tym zapewniono minimalną liczbę bloków CLB, przez które musi przejść sygnał z wejścia na wyjście (rys. 5.10).



Rys. 5.10. Schemat konwertera kodu pseudotermometrycznego na kod naturalny binarny
Fig. 5.10. Scheme of code converter from pseudo-thermometric code to natural binary code

Jest to konwerter dwuetapowy. W pierwszym etapie dokonywana jest konwersja z kodu pseudotermometrycznego (sygnały $P_0 - P_{n-1}$) na kod 1 z N (sygnały $a_0 - a_{n-1}$). Drugi etap konwersji stanowią $n/2$ wejściowe funkcje logiczne OR, które zostały podłączone do wyjść etapu pierwszego tak, aby dla kolejnych stanów w kodzie 1 z N na wyjściu konwertera uzyskiwać kolejne stany w kodzie naturalnym binarnym. Przykładowo do sygnału a_0 nie został podłączony żaden funkcja OR, do sygnału a_1 podłączono funkcję, którego wyjście określa stan D_0 , a do sygnału a_{31} zostało podłączonych pięć funkcji OR określających stan na pięciu najmniej znaczących wyjściach od D_4 do D_0 . Rozwiązanie to umożliwiło uzyskać czas konwersji równy około 16 ns, co pozwoliło, przy zastosowaniu 4 takich konwerterów, uzyskanie maksymalnej intensywności rejestrowanych impulsów na poziomie $266 \cdot 10^6$ impulsów na sekundę.

6. DOKŁADNOŚĆ SYSTEMU MONITOROWANIA SPEKTROMETRU MASOWEGO CZASU PRZELOTU

Każdy system pomiarowy, w tym także umożliwiający pomiar funkcji intensywności badanego procesu, ma skończoną dokładność określoną przez wiele czynników. Wśród tych czynników można wyodrębnić czynniki zewnętrzne, czynniki związane z właściwościami badanego procesu oraz czynniki wewnętrzne związane z systemem pomiarowym. Niektóre z nich, jak niestabilność generatora wzorcowego [105–118] czy odchyłki fazowe spowodowane szumami [105, 119, 110] oraz odbiciami sygnałów w liniach transmisyjnych [117, 120], stanowią źródła błędów we wszystkich systemach pomiaru funkcji intensywności. Inne zaś, jak rozdzielczość układu zliczającego [121, 29, 122, 57, 123, 124, 58, 125, 126] są wspólne dla bramkowanych systemów zliczających. Są też takie czynniki, które są charakterystyczne tylko dla określonego systemu pomiarowego, jak efekt modulacji szerokości bramki czasowej występujący w systemie wielokanałowym.

Systemy charakteryzujące się nano- i pikosekundową rozdzielczością coraz częściej realizowane są z wykorzystaniem układów VLSI [47, 51, 52, 81–83, 85]. Układy te umożliwiają zmniejszenie zakłóceń spowodowanych zniekształceniami fazowymi wprowadzanymi przez linie transmisyjne. W szczególności dotyczy to układów o zwiększonym zakresie pomiarowym, wykorzystujących generatory wzorcowe. Układy takie wymagają odpowiedniej topologii doprowadzeń, zapewniającej minimalne zniekształcenia fazowe sygnału, oraz kompensację opóźnień wynikających z planarnego rozmieszczenia poszczególnych bloków funkcjonalnych systemu. Problem zniekształceń fazowych wprowadzanych przez linie transmisyjne może być skutecznie ograniczany przez budowę układów i systemów pomiarowych o dużej skali integracji, co można osiągnąć, stosując szybkie układy CPLD i FPGA.

6.1. WPŁYW INTENSYWNOŚCI IMPULSÓW NA PROCES ZLICZANIA

Jeżeli maksymalna wartość intensywności impulsów wejściowych ρ_{\max} spełnia warunek [29, 57]

$$\rho_{\max} < \frac{0,35}{\Delta t + t_{ef}} \quad (6.1)$$

gdzie Δt – rozdzielczość, a t_{ef} – efektywny czas trwania impulsu, to można stosować technikę zliczania impulsów. W przypadku przeciwnym należy uwzględnić straty w liczbie zliczeń lub stosować metodę integracji analogowej. Stosując technikę zliczania impulsów w wąskim przedziale czasowym o szerokości τ , licznik n -bitowy zliczy w ciągu M cykli pomiarowych, przy intensywności procesu ρ oraz oczekiwanej liczbie impulsów N_{i_1} pojawiającej się na wejściu licznika w jednym cyklu pomiarowym, N_{o_M} impulsów, co można wyrazić w postaci zależności

$$N_{o_M} = MN_{i_1} \exp[-\rho(\Delta t + t_{ef})]. \quad (6.2)$$

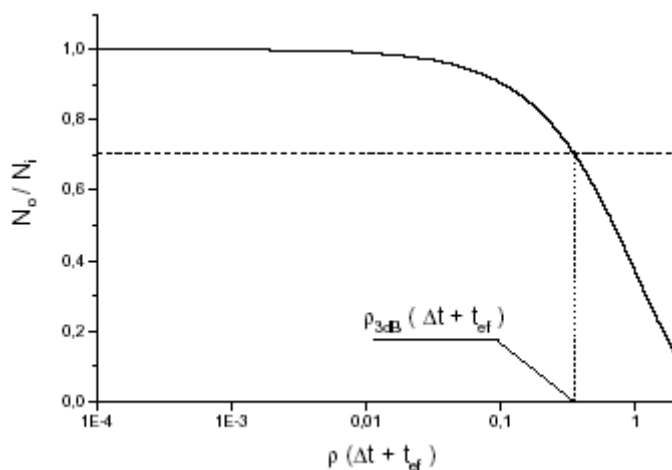
Intensywność procesu ρ wyrazić można w postaci

$$\rho = \frac{N_{i_1}}{\tau}. \quad (6.3)$$

Oznaczając $\Delta t' = \Delta t + t_{ef}$ oraz $N_{i_M} = MN_{i_1}$ zależność (6.2) przyjmie postać

$$N_{o_M} = N_{i_M} \exp\left[-\frac{N_{i_1} \Delta t'}{\tau}\right]. \quad (6.4)$$

Zazwyczaj dla uniknięcia efektu „gubienia impulsów” podczas procesu zliczania ogranicza się maksymalną intensywność procesu ρ_{3dB} , powodującą trzydecybelowe zmniejszenie liczby impulsów zliczonych w stosunku do liczby impulsów pojawiających się na wejściu licznika, co zilustrowano na rysunku 6.1.



Rys. 6.1. Wpływ rozdzielczości licznika na proces zliczania
Fig. 6.1. Counter resolution influence on counting process

Błąd zliczania ΔN wyrazić można w postaci

$$\Delta N = N_{i_M} \left(1 - \exp\left[-\frac{N_{i_1} \Delta t'}{\tau}\right]\right), \quad (6.5)$$

a błąd względny ΔN_{wz}

$$\Delta N_{wz} = \frac{\Delta N}{N_{i_M}} = 1 - \exp\left[-\frac{N_{i_1} \Delta t'}{\tau}\right]. \quad (6.6)$$

W przypadku kiedy maksymalna intensywność jest zbyt duża, podane zależności mogą być wykorzystane do korekcji liczby zliczeń we wszystkich typach systemów zliczających stosowanych w spektrometrii masowej czasu przelotu.

6.2. WPŁYW INTENSYWNOŚCI ZAKŁÓCEŃ NA PROCES ZLICZANIA

Istotny wpływ na proces zliczania wywiera obecność impulsów zakłócających (tła) na wejściu systemu pomiarowego. Prawdopodobieństwo pojawienia się na wejściu licznika n_i impulsów w przedziale czasu τ dla wartości oczekiwanej równej μ wynosi [127–135]

$$p(n_i, \mu) = \frac{\mu^{n_i} \exp(-\mu)}{n_i!}. \quad (6.7)$$

Zakładając, że ρ_s – intensywność sygnału i ρ_b – intensywność zakłóceń zewnętrznych przyjmują stałe wartości w wąskiej bramce czasowej τ , prawdopodobieństwo zliczenia n_k impulsów przez licznik wyraża rozkład Bernoulliego

$$P \frac{n_k}{n_i} = \binom{n_i}{n_k} p^{n_k} (1-p)^{n_i-n_k}, \quad (6.8)$$

gdzie n_i stanowi liczbę impulsów pojawiających się na wejściu licznika w czasie τ , a p jest prawdopodobieństwem uwzględnienia impulsu przez licznik. Prawdopodobieństwo uwzględnienia każdego kolejno pojawiającego się impulsu jest równe prawdopodobieństwu nie pojawienia się impulsu w czasie równym Δt (rozdzielczość licznika), i wyraża się zależnością

$$p = \exp[-(\rho_s + \rho_b)\Delta t]. \quad (6.9)$$

Można wykazać [13], że liczba zliczeń podlega także rozkładowi Poissona, a stosunek sygnału do szumu wyraża zależność [38]

$$\text{SNR} = \rho_s \sqrt{\frac{\tau p}{\rho_s + \rho_b}}. \quad (6.10)$$

Przyjmując następujące oznaczenia:

$N_s = \rho_s \tau$ – średnia liczba impulsów sygnału, pojawiających się na wejściu licznika w bramce czasowej o szerokości τ ,

$N_b = \rho_b \tau$ – średnia liczba impulsów tła, pojawiających się na wejściu licznika w bramce czasowej o szerokości τ ,

$N_t = (\rho_s + \rho_b) \Delta t$ – średnia liczba wszystkich impulsów na wejściu licznika w przedziale czasu o szerokości Δt ,

stosunek sygnał–szum można wyrazić w postaci [38]

$$\text{SNR} = N_s \sqrt{\frac{\exp(-N_t)}{N_s + N_b}}. \quad (6.11)$$

Należy podkreślić, że w przypadku kiedy intensywność pojawiania się impulsów na wejściu licznika jest mała, pomijalnie mała jest też średnia liczba impulsów N_t przypadająca na bardzo mały odcinek czasu Δt . Dla małych wartości średnich liczb impulsów sygnału N_s i tła N_b stosunek sygnał–szum wynosi

$$\text{SNR} \cong \frac{N_s}{\sqrt{N_s + N_b}}. \quad (6.12)$$

W przypadku kiedy w analizowanym zakresie $N_b \ll N_s$, stosunek ten można przybliżyć zależnością $\text{SNR} \cong \sqrt{N_s}$. Kiedy intensywność pojawiania się impulsów na wejściu licznika jest duża, wówczas stosunek sygnał–szum można w przybliżeniu wyrazić zależnością

$$\text{SNR} \cong \frac{1}{\sqrt{N_t}}, \quad (6.13)$$

lub zależnością [13]

$$\text{SNR} \cong \sqrt{N_s \exp(-N_t)}. \quad (6.14)$$

Z przedstawionych wykresów wynika, że zbyt duża intensywność impulsów sygnału lub sygnału i tła może powodować gwałtowne zmniejszenie liczby zliczeń. Z zależności (6.11) wynika, że proces ten rozpocznie się, gdy oczekiwana liczba impulsów na wejściu licznika, pojawiających się w czasie zliczania, czyli w bramce czasowej o szerokości τ przekroczy wartość $\tau/\Delta t$. Biorąc pod uwagę podane zależności i wykresy można stwierdzić, że w celu utrzymania wartości SNR na wysokim poziomie, w dużym zakresie intensywności impulsów wejściowych rozdzielczość systemu zliczającego powinna być jak największa ($\Delta t \rightarrow 0$), a liczba impulsów tła ograniczona ($N_b \rightarrow 0$), czemu służy np. stosowanie dyskryminatorów amplitudy impulsów wejściowych pochodzących z fotopowielacza.

Rozdzielczość systemu zliczającego jest zawsze ograniczona, co powoduje konieczność ograniczenia maksymalnej intensywności impulsów pojawiających się na wejściu licznika. Dla małych wartości ρ zależność (6.10) przyjmie postać

$$\text{SNR}_{\rho \ll \rho_{3\text{dB}}} \cong \rho_s \sqrt{\frac{\tau}{\rho_s + \rho_b}}. \quad (6.15)$$

Możliwość polepszenia zatem stosunku sygnał–szum przez zwiększenie intensywności sygnału jest ograniczona. Istnieje jednak możliwość poprawy tego parametru

przez wielokrotną analizę tego samego procesu. Przy wykonaniu M cykli pomiarowych oraz przy $N_i \ll 1$ stosunek sygnał–szum może być wyrażony zależnością

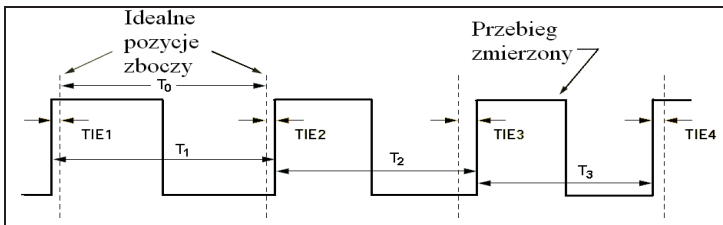
$$\text{SNR}(M) = \sqrt{\sum_{i=1}^M N_{s_i}} = \sqrt{M \bar{N}_s}, \quad (6.16)$$

gdzie \bar{N}_s jest wartością średnią liczby impulsów sygnału w bramce czasowej τ .

6.3. WPLYW NIESTABILNOŚCI GENERATORA WZORCOWEGO

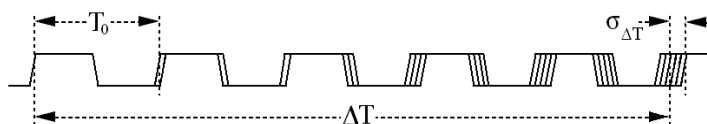
Głównym zadaniem generatora wzorcowego w systemach pomiaru funkcji intensywności jest odtworzenie skali czasu. Ponadto generator wzorcowy jest wykorzystywany w brankowanych systemach zliczania do generacji bramek czasowych. Szczególnie w systemach z branką przesuwną jest wykorzystywany także do generowania opóźnienia czasu otwarcia bramki czasowej. W systemach o rozdzielczości pikosekundowej generator wzorcowy wykorzystywany jest głównie w celu poszerzenia zakresu pomiarowego. W szczególnym przypadku w systemach wykorzystujących generatory synchronizowane fazowo stanowi on element noniusza.

Stabilność średniokresowa (mierzona w przedziale czasu: dzień–miesiąc) stosowanych w systemach pomiarowych kwarcowych generatorów wzorcowych zawiera się w granicach 10^{-6} – 10^{-10} i znacznie zależy od stopnia kompensacji wpływu temperatury. W przypadku kiedy wymagana jest większa stabilność wzorca, stosowane też są rubidowe wzorce atomowe o stabilności rzędu 10^{-11} [136]. Jednym z istotniejszych problemów związanych z badaniem stabilności generatora jest pomiar chwilowych, względnych odchyłek częstotliwości i fazy [105–118, 137–140]. Fluktuacje generatorów wzorcowych pokazane na rysunku 6.2 mogą być charakteryzowane zarówno w przedziale częstotliwości, jak i w przedziale czasu [105–107, 110, 137, 140, 141].



Rys. 6.2. Błąd odcinka czasowego spowodowany fluktuacjami generatora
Fig. 6.2. Time-interval-error caused by oscillator fluctuations

Jedną z metod pomiaru akumulowanych fluktuacji generatora (rys. 6.3) w przedziale czasu jest metoda wielokrotnego pomiaru odcinków czasu składających się z określonej liczby okresów generatora [142].



Rys. 6.3. Akumulowany błąd odcinka czasu spowodowany fluktuacjami generatora
Fig. 6.3. Long term accumulated jitter caused by oscillator fluctuations

Pomiaru odcinka czasu dokonuje się każdorazowo, wykorzystując wielosegmentową linię opóźniającą implementowaną w strukturze układu FPGA, zawierającą również wszystkie pozostałe bloki sprzętowe systemu pomiarowego, z wyjątkiem generatorów.

Precyzja pomiaru odcinka czasu zależy od rozdzielczości linii opóźniającej, która zależy od parametrów zastosowanego układu FPGA. Rozdzielczość ta w przypadku układu XC2VP4 (Virtex 2 Pro) wynosi 70 ps, a dla układu XC4VLX25 (Virtex 4) 50 ps.

Zmierzone odcinki czasu, stanowiące różnice stempli czasowych pobranych w wyniku wielu cykli pomiarowych, można rozważać jako zmienne losowe Δt_N , o normalnym rozkładzie prawdopodobieństwa. Oczywiście odchylenie standardowe dla zmiennych losowych Δt_N będzie różne od odchylenia standardowego charakteryzującego fluktuacje okresu sygnału generatora badanego ε

$$\Delta t_N = NT_0 + \sum_{n=1}^N \varepsilon_n . \quad (6.17)$$

Funkcja gęstości prawdopodobieństwa dla fluktuacji ε może być opisana zależnością

$$p(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{\varepsilon^2}{2\sigma_\varepsilon^2}} , \quad (6.18)$$

gdzie σ_ε stanowi odchylenie standardowe zmiennej ε .

Odchylenie standardowe odcinka Δt_N będzie zależęć od korelacji pomiędzy odchyłkami ε_i . W przypadku braku korelacji pomiędzy odchyłkami ε_i , wówczas wariancja odcinka czasu Δt_N będzie sumą wariancji poszczególnych zmiennych losowych. Przyjmując taki przypadek dla fluktuacji okresu generatora, wariancja odcinka czasu Δt_N może być wyrażona w postaci

$$\sigma_{\Delta t_N}^2 = N2\sigma_\varepsilon^2 . \quad (6.19)$$

W przypadku kiedy istnieje korelacja pomiędzy fluktuacjami, wariancja odcinka czasu jest większa [132]. Reasumując, można przyjąć, że w ogólności wariancja przyjmie postać

$$\sigma_{\Delta t_N} = N^a \sigma_\varepsilon^2, \quad (6.20)$$

gdzie $1 \leq a \leq 2$.

Jak widać z podanych zależności, wariancja odcinka czasu wyznaczonego przez N okresów zegara, zależy znacznie od liczby okresów. Analizując zależności (6.19) oraz (6.20) wartości maksymalne funkcji gęstości prawdopodobieństwa przyjmują postać

$$p(t = NT_0) = \frac{1}{\sqrt{2\pi}\sigma_{\Delta t_N}} = \frac{1}{\sqrt{2\pi}N^a\sigma_\varepsilon} = cN^{-a/2}, \quad (6.21)$$

gdzie

$$c = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon}, \quad (6.22)$$

oraz

$$\sigma_{\Delta t_N} = \frac{N^{a/2}}{c\sqrt{2\pi}} \quad (6.23)$$

Do pełnego opisu rozmycia czasowego generatora potrzebne są zatem tylko dwa parametry c oraz a , które mogą być wyznaczone z normalnych rozkładów wcześniej zgromadzonych stempli czasowych. Znając te parametry, można prognozować odchylenie standardowe dla dowolnego odcinka czasu z zależności (6.22) i (6.23).

6.4. BŁĘDY BRAMKOWANYCH SYSTEMÓW ZLICZANIA IMPULSÓW

Wymienione wcześniej źródła błędów związane z generatorem wzorcowym powodują w bramkowanych systemach zliczających zniekształcenie efektywnej bramki czasowej, co w konsekwencji prowadzi do błędów pomiaru funkcji intensywności. Przyjmując, że stosowany generator wzorcowy ma okres równy T_w oraz wariancję $\sigma_3^2(T_w)$, odcinek czasu t_m zmierzony za pomocą tego generatora można zapisać w postaci

$$t_m = N_w T_w + \sum_{i=1}^{N_w} x_i, \quad (6.24)$$

gdzie x_i stanowią zmienne losowe (odchyłki czasu od wartości T_w), a N_w jest liczbą cykli generatora wzorcowego.

Przyjmując, że zmienne te podlegają rozkładowi Gaussa z wariancją $\sigma_3^2(T_w)$ oraz zerową wartością średnią, zmienna losowa t_m będzie podlegała także rozkładowi Gaussa z wartością średnią równą $N_w T_w$ i wariancją

$$\sigma_x^2(t_m) = N_w^\alpha \sigma_x^2(T_w). \quad (6.25)$$

Funkcję gęstości prawdopodobieństwa zmiennej losowej t_m można przedstawić w postaci

$$p(t_m) = \frac{1}{\sigma_x(T_w) \sqrt{2\pi N_w^\alpha}} \exp\left\{-\frac{(t_m - N_w T_w)^2}{2N_w^\alpha \sigma_x^2(T_w)}\right\}. \quad (6.26)$$

Przyjmując że funkcję pojedynczej, idealnej bramki czasowej o szerokości τ można opisać zależnością

$$\Pi(t, T_w, N_w, \tau) = \mathbb{1}[t - N_w T_w] - \mathbb{1}(t - N_w T_w - \tau), \quad (6.27)$$

a rzeczywistą w postaci

$$\Pi_r(t, t_m, \tau) = \mathbb{1}[t - t_m] - \mathbb{1}(t - t_m - \tau), \quad (6.28)$$

efektywną funkcję bramki (dla wielu cykli pomiarowych) przedstawić można w postaci

$$\Pi_e(t, T_w, N_w, \tau) = \int_0^\infty p(t_m) \Pi_r(t, t_m, \tau) dt_m. \quad (6.29)$$

Można zatem wyrazić efektywną funkcję bramki w postaci splotu idealnej funkcji bramki

$$\Pi(t, \tau) = \mathbb{1}[t] - \mathbb{1}[t - \tau], \quad (6.30)$$

z funkcją gęstości prawdopodobieństwa zmiennej losowej t_m (6.26) [123]

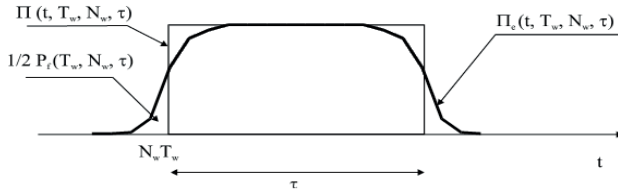
$$\Pi_e(t, T_w, N_w, \tau) = \Pi(t, \tau) \otimes \frac{1}{\sigma_x(T_w) \sqrt{2\pi N_w^\alpha}} \exp\left\{-\frac{(t - N_w T_w)^2}{2N_w^\alpha \sigma_x^2(T_w)}\right\}. \quad (6.31)$$

Efektywną i idealną bramkę czasową pokazano na rysunku 6.4. Bramka efektywna jest wynikiem uśrednienia funkcji bramek dla nieskończenie dużej liczby cykli pomiarowych [13]. Różnica między bramką idealną a efektywną jest tym większa, im gorsza jest stabilność generatora wzorcowego i im większy jest czas opóźnienia otwarcia bramki t_m , odmierzany za pomocą generatora.

W systemie z bramką przesuwą efektywna bramka jest modyfikowana w wyniku niestabilności generatora wzorcowego linii opóźniającej, a w systemie wielokanałowym bramki efektywne dla poszczególnych kanałów czasowych modyfikowane są przez niestabilność generatora wzorcowego systemu pomiarowego. W wyniku zmiany kształtu bramki czasowej istnieje skończone prawdopodobieństwo $P_f(T_w, N_w, \tau)$ zaliczenia do kanału czasowego o szerokości τ , leżącego w odległości czasowej $N_w T_w$ od

momentu rozpoczęcia cyklu pomiarowego impulsów wejściowych nienależących do tego kanału czasowego, które wyraża zależność

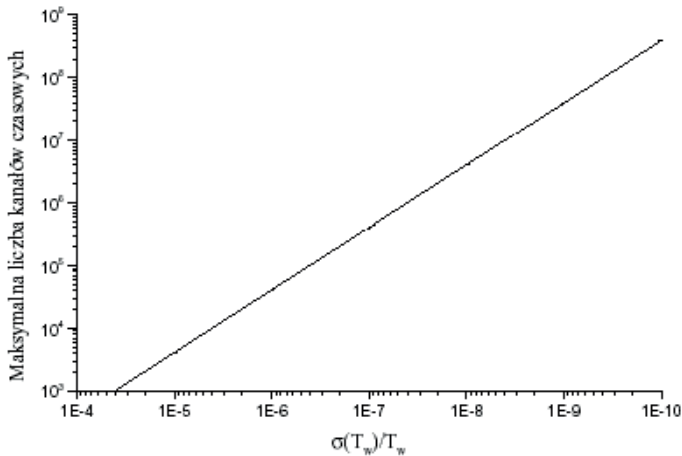
$$P_f(T_w, N_w, \tau) = \frac{1}{\tau} \left[\int_0^{N_w T_w} \Pi_e(t, T_w, N_w, \tau) dt + \int_{N_w T_w + \tau}^{\infty} \Pi_e(t, T_w, N_w, \tau) dt \right]. \quad (6.32)$$



Rys. 6.4. Idealna i efektywna bramka czasowa

Fig. 6.4. Ideal and effective time-gate

Jest to główna przyczyna ograniczenia zakresu pomiarowego wszystkich systemów pomiarowych o rozdzielczości nano- i pikosekundowej niezależnie od ich konstrukcji [13].



Rys. 6.5. Maksymalna liczba kanałów czasowych systemu wielokanałowego

Fig. 6.5. Maximal number of time-channel for multichannel scaler

Przyjmując [123], że w najgorszym przypadku dwukrotna wartość odchylenia standardowego dla odcinka czasu równego zakresowi pomiarowemu nie przekracza jednej dziesiątej szerokości bramki czasowej, obliczono maksymalną liczbę kanałów czasowych dla systemu wielokanałowego w funkcji względnego odchylenia standar-

dowego generatora wewnętrznego [123], co przedstawiono na rysunku 6.5. W podanych rozważaniach przyjęto stałą wartość szerokości bramki τ . W praktyce szerokość bramki także się zmienia i można przyjąć, że jest ona zmienną losową o rozkładzie Gaussa, niezależną od zmiennej t_m

$$\tau = \eta_\tau + \varepsilon_\tau, \quad (6.33)$$

gdzie η_τ jest wartością oczekiwaną, a ε_τ odchyłką zmiennej losowej τ .

W niektórych systemach z bramką przesuwaną [42, 143] bramka generowana jest bez udziału generatora wzorcowego. W takim przypadku efektywną funkcję bramki wyrazić można zależnością

$$\Pi_e(t, T_w, N_w, \tau) = \Pi(t, \tau) \otimes p(t_m) \otimes p(\tau), \quad (6.34)$$

w której $p(\tau)$ jest funkcją gęstości prawdopodobieństwa zmiennej losowej τ .

Ze względu na niezależność zmiennych losowych t_m i τ , zależność (6.34) można przedstawić w postaci

$$\Pi_e(t, T_w, N_w, \tau) = \Pi(t, \tau) \otimes \frac{\exp\left\{-\frac{(t - N_w T_w - \eta_\tau)^2}{2[N_w^\alpha \sigma_x^2(T_w) + \sigma_\varepsilon^2(\tau)]}\right\}}{\sqrt{2\pi[N_w^\alpha \sigma_x^2(T_w) + \sigma_\varepsilon^2(\tau)]}}. \quad (6.35)$$

W przypadku systemów wielokanałowych, jak i niektórych systemów z bramką przesuwaną szerokość bramki może być równa jednemu lub kilku okresom generatora wzorcowego, co można zapisać w postaci

$$\eta_\tau = N_g T_w, \quad (6.36)$$

gdzie N_g – liczba okresów generatora tworzących bramkę.

Zależność (6.35) można zatem przedstawić w prostszej postaci

$$\Pi_e(t, T_w, N_w, \tau) = \Pi(t, \tau) \otimes \frac{\exp\left\{-\frac{[t - T_w(N_w + N_g)]^2}{2\sigma_x^2(T_w)[N_w^\alpha + N_g^\alpha]}\right\}}{\sigma_x(T_w)\sqrt{2\pi[N_w^\alpha + N_g^\alpha]}}. \quad (6.37)$$

W systemach z bramkami czasowymi istotnym problemem jest synchronizacja układu generacji bramek czasowych z badanym procesem fizycznym [123, 144].

Istnieją dwie możliwości, pokazane na rysunku 6.6:

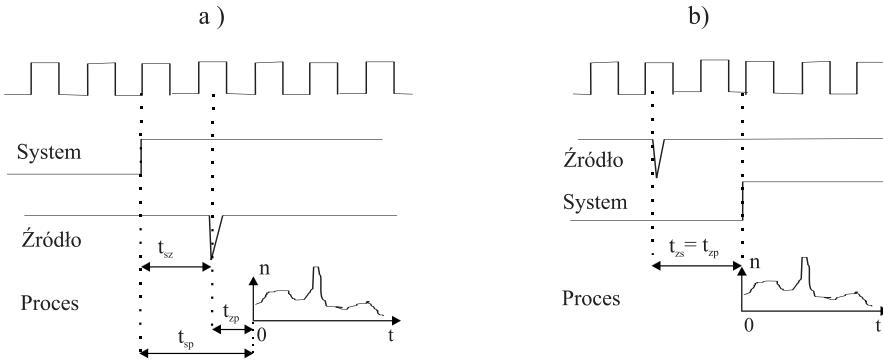
- a) system pomiarowy synchronizuje źródło wzbudzenia procesu badanego,
- b) źródło wzbudzenia procesu badanego synchronizuje system pomiarowy.

Odcinki czasu t_{sp} i t_{zp} można przedstawić jako zmienne losowe o rozkładzie normalnym:

$$t_{sp} = \eta_{sp} + \varepsilon_{sp}, \quad (6.38)$$

$$t_{zp} = \eta_{zp} + \varepsilon_{zp}, \quad (6.39)$$

gdzie η_{sp} i η_{zp} są wartościami oczekiwanymi reprezentującymi średnią wartość opóźnienia, a ε_{sp} i ε_{zp} stanowią odchyłki od wartości średnich tych opóźnień.



Rys. 6.6. Sposoby synchronizacji
Fig. 6.6. Synchronization methods

W przypadku kiedy system pomiarowy synchronizuje źródło wzbudzenia badanego procesu, odcinek czasu t_{sp} pomiędzy sygnałem synchronizującym a początkiem obserwowanego procesu składa się z dwóch odcinków t_{sz} i t_{zp} . Po kompensacji opóźnienia η_{sp} funkcję bramki można przedstawić w postaci splotu

$$\Pi_{e_{sz}}(t, T_w, N_w, \tau) = \Pi_e(t, T_w, N_w, \tau) \otimes \frac{\exp\left\{-\frac{t^2}{2\sigma_x^2(\varepsilon_{sp})}\right\}}{\sigma_x(\varepsilon_{sp})\sqrt{2\pi}}. \quad (6.40)$$

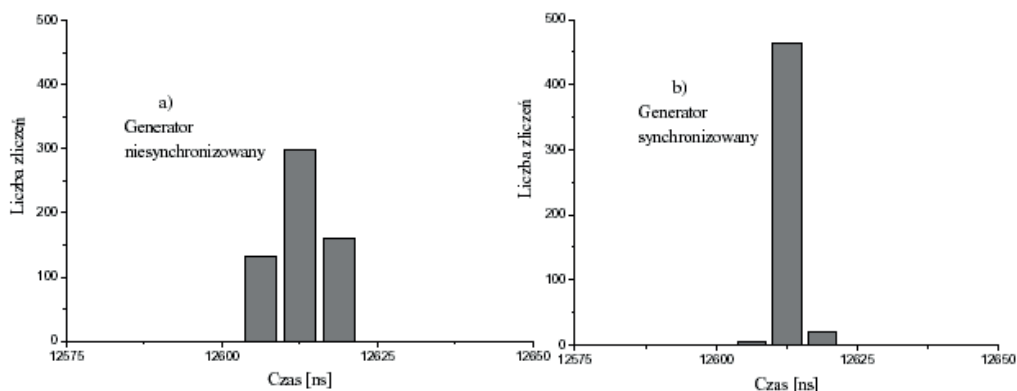
W przypadkach kiedy źródłem wzbudzenia procesu jest laser impulsowy zmienną losową t_{sz} charakteryzuje zwykle duża odchyłka standardowa, co prowadzi w konsekwencji do znacznego zniekształcenia bramki czasowej (rozmycie). Z tego powodu często (spektrometria masowa [9, 7, 53, 54]) stosuje się sposób synchronizacji pokazany na rysunku 6.6b. Metoda ta nie wykazuje już tej wady, ponieważ odchyłka standardowa zmiennej t_{zs} jest wielokrotnie mniejsza od odchyłki standardowej zmiennej t_{sz} . Wadą tej metody są trudności wynikające z konieczności dostosowania fazy wysokostabilnego wzorca częstotliwości do źródła wzbudzenia.

D. Chu [79] opracował sposób synchronizacji fazowej wysokostabilnych generatorów stosowanych w układzie noniusza systemu pomiaru odcinka czasu o pikosekundowej rozdzielczości. Generator działający na tej zasadzie [144, 145] został

zastosowany do wielokanałowego systemu zliczającego [146], współpracującego ze spektrometrem masowym czasu przelotu.

W trakcie testów zmieniano sposób synchronizacji. W przypadku pierwszym zastosowano generator wzorcowy autonomiczny w stosunku do zewnętrznego generatora impulsów, a w przypadku drugim generator synchronizowany fazowo impulsami z generatora zewnętrznego [144]. Efekt zastosowania generatora synchronizowanego fazowo zilustrowano na rysunku 6.7.

W przypadku systemów o dużej rozdzielczości (mała szerokość bramki τ) niestabilność generatora wzorcowego oraz błędy związane z synchronizacją powodują, że funkcja bramki ma kształt funkcji Gaussa, a prawdopodobieństwo właściwego przydziału impulsu wejściowego jest bardzo małe, w szczególności dla odległych od początku układu współrzędnych kanałów czasowych.



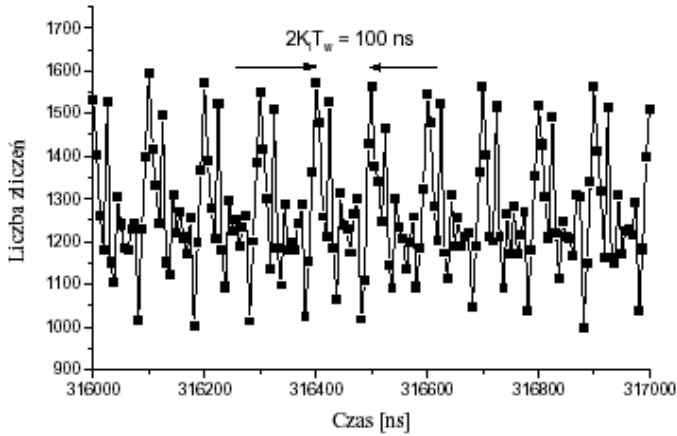
Rys. 6.7. Efekt zliczania impulsów

Fig. 6.7. Effect of pulse counts

W przypadku systemów wielokanałowych dodatkowym źródłem zakłóceń może być sam układ generacji bramek czasowych [57]. Niedokładna kompensacja opóźnień czasowych między układem generacji bramek a poszczególnymi licznikami powoduje efekt pasożytnej modulacji szerokości bramek czasowych, której częstotliwość f_{mb} jest związana z czasem trwania fazy pomiaru i w przypadku kiedy $\tau = T_w$, wyraża się zależnością

$$f_{mb} = \frac{1}{2K_1 T_w} \cdot \quad (6.41)$$

Efekt modulacji szerokości bramek czasowych dla systemu wielokanałowego o ośmiu licznikach w grupie ($K_1 = 8$) i szerokości bramek czasowych $\tau = T_w = 6,25$ ns zilustrowano na rysunku 6.8 [59]. Efekt ten powoduje okresowo zmienną liczbę zliczeń w poszczególnych kanałach czasowych.



Rys. 6.8. Szkodliwa modulacja szerokości bramek czasowych w wielokanałowym systemie zliczającym
Fig. 6.8. Parasitic time-gate width modulation in multichannel scaler

Pomijając wpływ wszystkich wymienionych czynników na funkcję bramki, poza efektem modulacji, funkcję tę można wyrazić w postaci [123]

$$\Pi(t, T_w, N_i, K_l) = \begin{cases} 1 & \text{dla } (N_i - 1)T_w + \Delta\tau_e \leq t \leq N_i T_w + \Delta\tau_e, \\ 0 & \text{dla } t < (N_i - 1)T_w + \Delta\tau_e \text{it } > N T_w + \Delta\tau_e, \end{cases} \quad (6.42)$$

gdzie szerokość pojedynczej bramki wyraża zależność $\tau_k = T_w + \Delta\tau_k$, a odchyłkę efektywną zależność

$$\Delta\tau_e = \sum_{k=1}^{(N_i-1) \bmod 2K_l} \Delta\tau_k. \quad (6.43)$$

Oprócz błędów kształtu bramki w wielokanałowych systemach zliczających mogą pojawić się dodatkowo błędy zliczania wynikające z niedoskonałej kompensacji różnicy opóźnienia sygnałów bramkujących wejścia poszczególnych liczników oraz ze skończonej szerokości impulsów wejściowych. Błędy te polegają:

- na braku akceptacji impulsu wejściowego przez jakikolwiek licznik,
- podwójnym zliczaniu tego samego impulsu w sąsiednich kanałach czasowych.

Błąd braku akceptacji impulsu wejściowego można wyeliminować w wyniku kompensacji opóźnień sygnałów bramkujących, co jest tym łatwiejsze, im większy jest stopień integracji systemu pomiarowego. W celu niedopuszczenia do wystąpienia efektu podwójnego czy też wielokrotnego zliczania tego samego impulsu, można stosować układ różniczkujący, ograniczający szerokość impulsów wejściowych oraz metodę blokowania kolejnego licznika [57, 59, 123, 124] w pojedynczym cyklu pomiarowym.

W systemach z przesuwną bramką czasową wymienione źródła błędów, takie jak niestabilność generatora wzorcowego czy błąd synchronizacji odnoszą się przede wszystkim do cyfrowej linii opóźniającej.

6.5. BŁĘDY SYSTEMÓW Z REJESTRACJĄ STEPLI CZASOWYCH WYKORZYSTUJĄCYCH LINIE OPÓŹNIAJĄCE

Najbardziej istotnymi błędami systemów z rejestracją stempli czasowych są błędy nieliniowości, których źródłem jest WLO [96, 97]. Niezależnie od implementacji opóźnienia poszczególnych segmentów WLO będą się różnić. Wszelkie sprzętowe metody linearyzacji mogą jedynie przybliżyć uzyskaną charakterystykę do idealnej [95]. Znajomość charakterystyki systemu umożliwia wykonanie korekcji zmierzonych wartości stempli czasowych lub rozkładów prawdopodobieństwa odcinków czasu [95, 147–150].

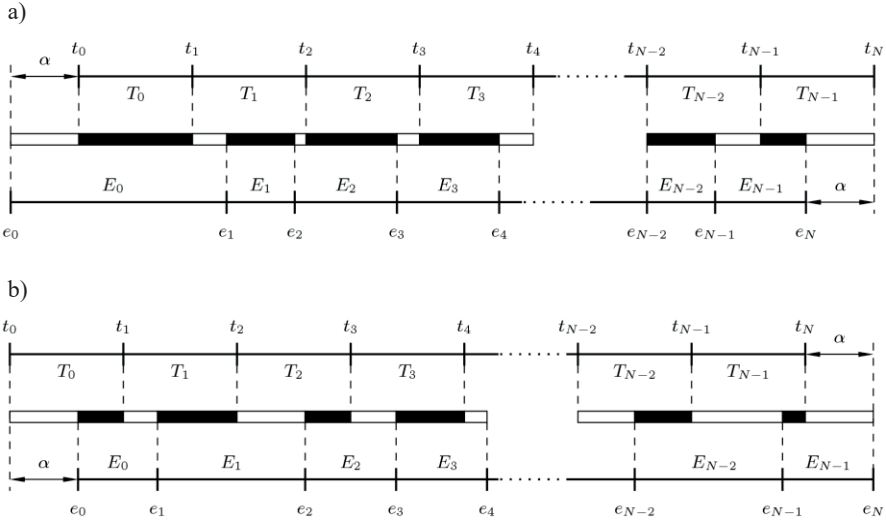
Wyznaczanie składu izotopowego wiązki atomowej polega na wyznaczeniu histogramu odcinków czasu, a więc metody korygujące rozkłady odcinków czasu, będą miały w tym przypadku zastosowanie [13, 95]. W rozdziale tym zostaną zaprezentowane dwie metody, z których pierwsza minimalizuje wpływ nieliniowości systemu na histogram odcinków czasu oraz druga, która dodatkowo minimalizuje wpływ kwantyzacji. Charakterystyki systemów są wyznaczone w sposób opisany w podrozdziale 5.5.

6.5.1. MINIMALIZACJA WPŁYWU NIELINIOWOŚCI

W metodzie tej każdemu opóźnieniu segmentu WLO E_i (przedziałowi eksperymentalnemu) przyporządkowuje się jeden lub więcej przedziałów teoretycznych T_j , tak jak zostało to pokazane na rysunku 6.9. W jej podstawowym wariacie szerokości wszystkich przedziałów teoretycznych T_j są takie same i równe wartości średniej wszystkich przedziałów eksperymentalnych. Można jednak przyjąć dowolne wartości przedziałów teoretycznych, które zostają przyporządkowane przedziałom eksperymentalnym. Szczególnie wielkości przedziałów teoretycznych mogą być mniejsze od wartości średniej wszystkich przedziałów eksperymentalnych, co umożliwi wykonanie histogramu z rozdzielczością większą od średniej rozdzielczości modułu pomiarowego.

Zgodnie z rysunkiem 6.9, początki i końce przedziałów eksperymentalnych i teoretycznych zostały oznaczone kolejno symbolami e_i oraz t_i , a ich szerokości odpowiednio przez E_i oraz T_i . Parametr α pozwala na przesunięcie przedziałów eksperymentalnych E_i tak, aby istniało maksymalne ich pokrycie z przedziałami teoretycznymi T_i . Różnica pomiędzy początkami/końcami par przedziałów teoretycznych i eksperymentalnych została oznaczona przez $r_i(\alpha)$, co dla $i = [0, N]$ można zapisać jako

$$r_i(\alpha) = e_i + \alpha - t_i \quad (6.44)$$



Rys. 6.9. Przyporządkowanie przedziałów teoretycznych do eksperymentalnych: a) $\alpha > 0$; b) $\alpha \leq 0$

Fig. 6.9. Assignment of theoretical partitions to experimental ones: a) $\alpha > 0$; b) $\alpha \leq 0$

Chcąc uzyskać maksymalne pokrycie par przedziałów, wystarczy zminimalizować sumę kwadratów wszystkich różnic $r_i(\alpha)$, oznaczoną jako $r(\alpha)$, co prowadzi do minimalizacji wyrażenia

$$r(\alpha) = \sum_{i=0}^N r_i^2 = \sum_{i=0}^N (e_i + \alpha - t_i)^2 \quad (6.45)$$

ze względu na parametr α .

Parametr α , dla którego występuje maksymalne pokrycie par T_i, E_i ma wartość

$$\alpha = \frac{1}{N+1} \sum_{i=0}^N (t_i - e_i). \quad (6.46)$$

Druga pochodna funkcji $r(\alpha)$ względem α jest dodatnia dla dowolnego α , stąd w punkcie określonym równaniem 6.46 $r(\alpha)$ przyjmuje wartość minimalną, czyli następuje maksymalne pokrycie przedziałów teoretycznych i eksperymentalnych.

W etapie drugim wyznacza się elementy $P_{i,j}$, z których każdy przedstawia udział i -tego przedziału teoretycznego w j -tym przedziale eksperymentalnym. W celu sprawnego przedstawienia algorytmu, który tego dokonuje, wprowadza się następujące oznaczenia

$$\underline{E}_i = [e_i, e_{i+1}], \quad (6.47)$$

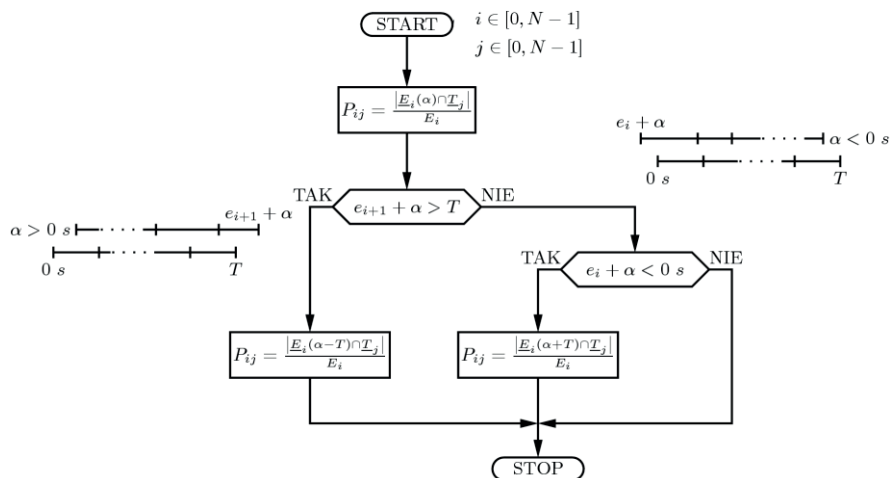
$$\underline{T}_i = [t_i, t_{i+1}], \quad (6.48)$$

$$|\underline{E}_i| = E_i = e_{i+1}, e, \quad (6.49)$$

$$|\underline{T}_i| = T_i = t_{i+1}, t, \quad (6.50)$$

$$\underline{E}_i(\alpha) = [e_{i+1} + \alpha, e_i + \alpha]. \quad (6.51)$$

$$\underline{E}_i \cap \underline{T}_j = \begin{cases} [\max(e_i, t_j), \min(e_{i+1}, t_{j+1})] & \text{gdy } \max(e_i, t_j) < \min(e_{i+1}, t_{j+1}) \\ 0 & \text{w pozostałych przypadkach} \end{cases}. \quad (6.52)$$



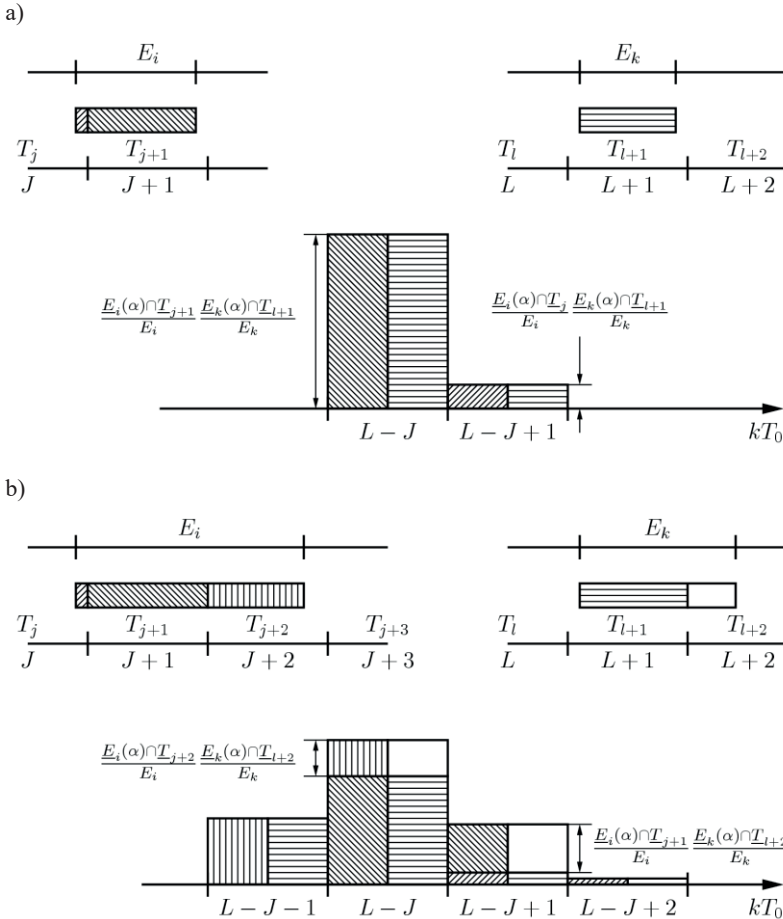
Rys. 6.10. Algorytm wyznaczania współczynników tablicy udziałów
Fig. 6.10. Algorithm of participation array coefficients calculation

Symbole \underline{T}_j i \underline{E}_i oznaczają parę współrzędnych określających początek i koniec i -tego przedziału kolejno teoretycznego i eksperymentalnego, T_i oraz E_i są natomiast szerokościami tych przedziałów, operator $()$ pozwala na przesuwanie przedziału, natomiast operator \cap zwraca przedział wspólny obu przedziałów.

Mając wprowadzone podane oznaczenia, algorytm metody minimalizującej wpływ nieliniowości modułu można przedstawić w następujący sposób (rys. 6.10). Jeżeli któryś z przedziałów eksperymentalnych E_i pokrywa się z którymś z przedziałów teoretycznych T_j , element P_{ij} tablicy udziałów otrzymuje wartość proporcjonalną do części wspólnej obu przedziałów, a odwrotnie proporcjonalna do szerokości przedziału eksperymentalnego E_i , w przeciwnym przypadku element P_{ij} otrzyma wartość 0. Wszystkie sumy elementów tablicy udziałów dla ustalonego przedziału eksperymentalnego są równe jedności, co można zapisać jako

$$\sum_{j=0}^{N-1} P_{ij} = 1, \quad i = [0, N - 1], \quad (6.53)$$

a więc elementy $P_{i,j}$ dla ustalonego przedziału eksperymentalnego mogą być traktowane jako prawdopodobieństwa udziału tego przedziału w j -tym przedziale teoretycznym.



Rys. 6.11. Przykłady obliczania przyczynku do histogramu różnic dla dwóch stempli czasowych:
 a – pierwszy stempel czasowy został zakwalifikowany do dwóch przedziałów teoretycznych,
 drugi do jednego, b – pierwszy stempel czasowy został zakwalifikowany
 do trzech przedziałów teoretycznych, drugi do dwóch

Fig. 6.11. Calculation examples of two-time-stamps difference histogram contribution:
 a – first time-stamp has been classified to two theoretical partitions, second one to one theoretical
 partition b – first time-stamp has been classified to three theoretical partitions,
 second one to two theoretical partitions

Dysponując tablicą $P_{i,j}$, istnieje możliwość obliczenia prawdopodobieństwa uzyskania określonej różnicy pomiędzy dwoma stemplami czasowymi, które zostały zaliczone kolejno do i -tego oraz j -tego przedziału eksperymentalnego. Wówczas histo-

gram odcinków czasowych na pozycji kT_0 , gdzie T_0 jest szerokością przedziału teoretycznego, a k jest liczbą całkowitą nieujemną, zostanie zwiększony o wartość

$$\rho_{ij}(kT_0) = \sum_{\substack{(N-1, N-1) \\ \{j=0, l=0\}, S_l - S_j = kT_0}} P_{ij} P_{kl}, \quad (6.54)$$

gdzie $S_{llk} = T_0(l/k + nN_{llk})$, a N_{llk} jest liczbą cykli zegara wzorcowego, które zostały zliczone przez licznik pełnych okresów od momentu inicjalizacji modułu pomiarowego. Sposób obliczania histogramu różnic kolejnych pomiarów przedstawiono na rysunku 6.11.

W przykładzie (a) dla pierwszego stempla czasowego przedział eksperymentalny E_i rozkłada się na dwa przedziały teoretyczne T_j i T_{j+1} , kolejno w proporcjach $E_i \cap T_j$ oraz $E_i \cap T_{j+1}$, a dla drugiego stempla czasowego przedział eksperymentalny E_k zostaje przypisany przedziałowi teoretycznemu T_{l+1} w całości. Histogram różnicy stempli czasowych, który standardowo dla różnicy dwóch stempli czasowych zostałby uzupełniony o czynnik równy 1 na pozycji $k-i$, jest w tym przypadku uzupełniony na pozycji $L - J + 1$ o czynnik ρ_{L-J+1} równy

$$\rho_{L-J+1} = P_{i,j} P_{k,l+1} = \frac{|E_i(\alpha) \cap T_j|}{E_i} \frac{|E_k(\alpha) \cap T_{l+1}|}{E_k} \quad (6.55)$$

oraz na pozycji $L - J$ o czynnik ρ_{L-J} równy

$$\rho_{L-J+1} = P_{i,j+1} P_{k,l+1} = \frac{|E_i(\alpha) \cap T_{j+1}|}{E_i} \frac{|E_k(\alpha) \cap T_{l+1}|}{E_k}. \quad (6.56)$$

W przykładzie (b) pierwszy stempel czasowy zostaje umieszczony w przedziale eksperymentalnym E_i , natomiast drugi stempel czasowy w przedziale eksperymentalnym E_k . Przedział eksperymentalny E_i pokrywa trzy przedziały teoretyczne T_j , T_{j+1} , T_{j+2} , a przedział eksperymentalny E_k pokrywa dwa przedziały teoretyczne T_{l+1} oraz T_{l+2} . Histogram różnic zostanie uzupełniony na pozycji $L - J - 1$ o wartość

$$\rho_{L-J-1} = P_{i,j+2} P_{k,l+1} = \frac{|E_i(\alpha) \cap T_{j+2}|}{E_i} \frac{|E_k(\alpha) \cap T_{l+1}|}{E_k} \quad (6.57)$$

na pozycji $L - J$ o wartość

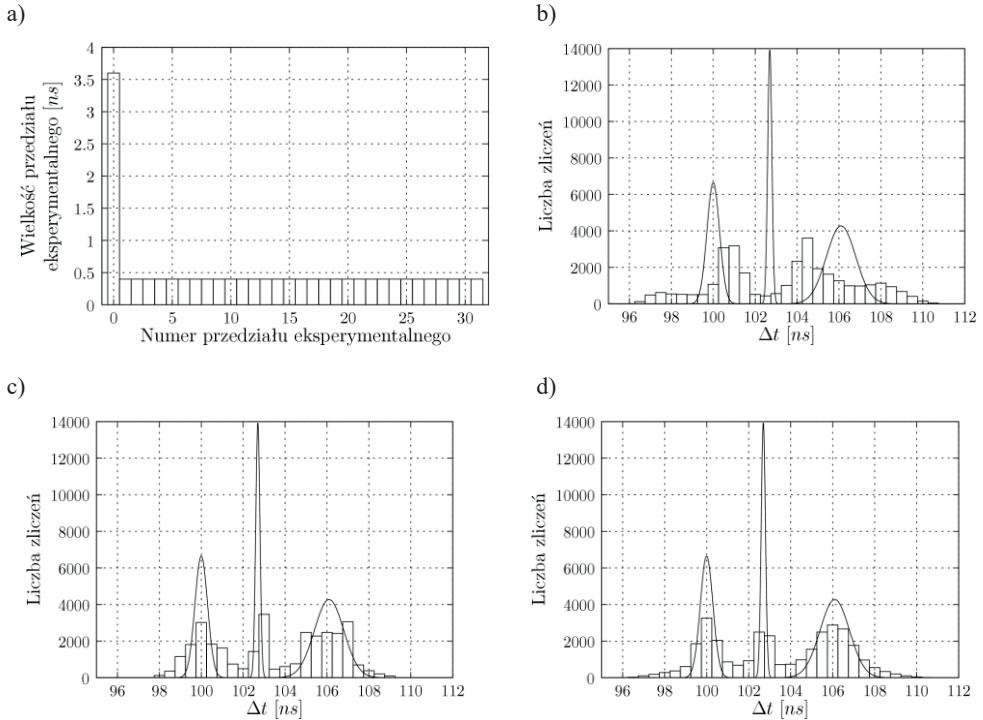
$$\begin{aligned} \rho_{L-J} &= P_{i,j+2} P_{k,l+2} + P_{i,j+1} P_{k,l+1} \\ &= \frac{|E_i(\alpha) \cap T_{j+2}|}{E_i} \frac{|E_k(\alpha) \cap T_{l+1}|}{E_k} + \frac{|E_i(\alpha) \cap T_{j+1}|}{E_i} \frac{|E_k(\alpha) \cap T_{l+1}|}{E_k}, \end{aligned} \quad (6.58)$$

na pozycji $L - J + 1$ o wartość

$$\begin{aligned} \rho_{L-J+1} &= P_{i,j+2}P_{k,l+2} + P_{i,j+1}P_{k,l+1} \\ &= \frac{|\underline{E}_i(\alpha) \cap \underline{T}_{j+1}|}{E_i} \frac{|\underline{E}_k(\alpha) \cap \underline{T}_{l+2}|}{E_k} + \frac{|\underline{E}_i(\alpha) \cap \underline{T}_j|}{E_i} \frac{|\underline{E}_k(\alpha) \cap \underline{T}_{l+1}|}{E_k}, \end{aligned} \quad (6.59)$$

natomiast pozycja $L - J + 2$ zostanie zwiększona o

$$\rho_{L-J+2} = P_{i,j}P_{k,l+2} = \frac{|\underline{E}_i(\alpha) \cap \underline{T}_j|}{E_i} \frac{|\underline{E}_k(\alpha) \cap \underline{T}_{l+2}|}{E_k}. \quad (6.60)$$



Rys. 6.12. Metoda minimalizacji wpływu nieliniowości modułu pomiarowego na wyznaczany histogram odcinków u : a – przykładowa charakterystyka modułu pomiarowego; b – histogram odcinków czasu uzyskany metodą standardową – odcinek wyznaczany jest jako różnica dwóch stempli (linia ciągła przedstawia rozkład odcinków czasu, które nie zostały poddane symulacji pomiaru);

c – metoda kompensująca błąd nieliniowości przez uwzględnienie wektora korekcyjnego (charakterystyki INL); d – metoda minimalizująca wpływ nieliniowości modułu opisana w tym podrozdziale
 Fig. 6.12. Measuring module nonlinearity influence minimizing method on time-interval histogram calculation: a – measuring module exemplary characteristic; b – time-interval histogram obtained by using standard method – the interval is being calculated as a difference of two time stamps (continuous line shows the distribution of time-intervals that has not been measured); c – nonlinearity error compensating method by using correcting vector (INL characteristic); d – measuring module nonlinearity influence minimizing method, described in this section

Działanie opisanej metody przedstawiono z wykorzystaniem trzech zbiorów odcinków czasu, charakteryzujących się wartością średnią, licznością oraz odchyleniem standardowym. Odcinki czasowe ze wszystkich trzech zbiorów zostały w losowej kolejności dodane do zbioru wynikowego i następnie poddane symulacji pomiaru wirtualnym modułem pomiarowym, którego charakterystyka została przedstawiona na rysunku 6.12a.

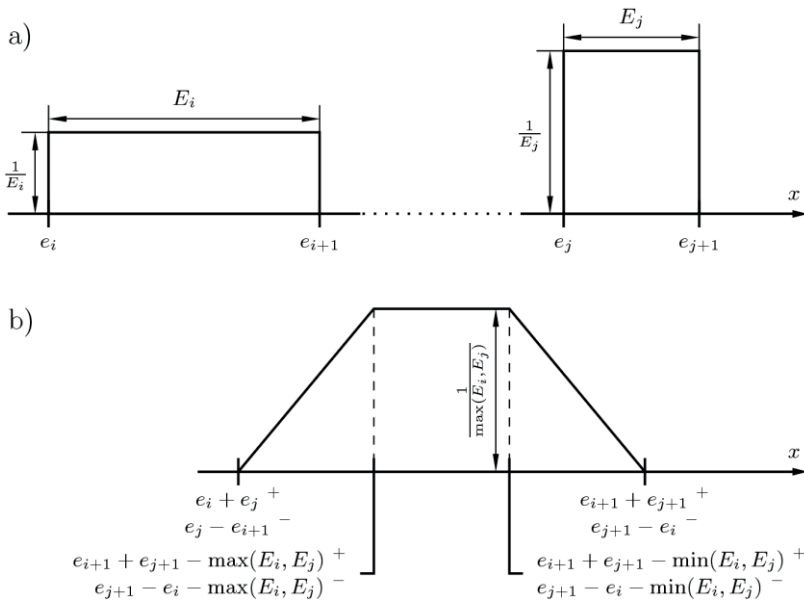
Wartość średnia pierwszego zbioru wynosi 100 ns, drugiego 102,7 ns, natomiast trzeciego 106,1 ns. Liczność pierwszego zbioru wynosi 10000, drugiego 7000, a trzeciego 15000. Odchylenie standardowe pierwszego zbioru odcinków czasu wynosi 0,3 ns, drugiego 0,1 ns i trzeciego 0,7 ns. Wyznaczony standardową metodą (różnice dwóch kolejnych stempli czasowych) histogram zmierzonych odcinków czasu został przedstawiony na rysunku 6.12b. Linia ciągła przedstawia rozkład odcinków czasu, które nie zostały poddane symulacji pomiaru. Rozkład ten został unormowany w taki sposób, aby pole powierzchni pod krzywą ciągłą było równe polu powierzchni histogramu i został on naniesiony na rysunki 6.12b,c,d w celach porównawczych. Jak widać, ze względu na bardzo dużą nieliniowość charakterystyki modułu pomiarowego histogram uzyskany standardową metodą odbiega w sposób znaczący od przewidywań. Występują dwa wyraźne zamiast trzech maksimów, dodatkowo są one nieprawidłowo umiejscowione. Na rysunku 6.12c przedstawiono histogram odcinków czasu uzyskany z wykorzystaniem kompensacji błędu nieliniowości, której idea polega na dodawaniu wektora korekcyjnego do wyznaczanych wartości stempli czasowych. Dogłębna analiza wykazała, że przedstawione za pomocą tego histogramu informacje o zawartościach odcinków czasu nie odbiegają w sposób znaczący od zadanych. Histogram sporządzony metodą opisaną w tym podrozdziale najlepiej przedstawia zadane do zmierzenia odcinki czasu (rys. 6.12d). Widać wyraźnie, że maksima dla wszystkich trzech zbiorów odcinków czasu są poprawnie umiejscowione, a szerokość każdego z rozkładów jest proporcjonalna do odchylenia standardowego zbioru odcinków czasu, któremu on odpowiada.

6.5.2. MINIMALIZACJA WPLYWU NIELINIOWOŚCI I KWANTYZACJI

Poprzednio opisana metoda zminimalizowała wpływ nieliniowości charakterystyki modułu pomiarowego na histogram odcinków czasu. Metoda, która zostanie opisana w tym podrozdziale, dodatkowo uwzględnia kwantyzację, jakiej poddawany jest mierzony odcinek czasu. Idea jej działania opiera się na tym, że rozkład prawdopodobieństwa sumy dwóch wielkości równy jest spłotowi rozkładów prawdopodobieństwa ich wystąpienia [150–153]. W przypadku odcinka czasu, który jest różnicą dwóch stempli czasowych, spłot zostanie wykonany w taki sposób, że argument funkcji opisującej rozkład pierwszego stempla czasowego będzie miał zmieniony znak.

Pomiaru odcinka czasu dokonujemy przez wyznaczenie dwóch stempli czasowych. Jeżeli stempel czasowy zostanie zakwalifikowany do i/j -tego przedziału ekspe-

rymentalnego o szerokości E_{ij} to rozkład prawdopodobieństwa jego wystąpienia jest równomierny i równy $1/E_{ij}$ w zakresie eksperymentu (rys. 6.13a). Poza przedziałem E_{ij} przyjmuje się, że prawdopodobieństwo jest zerowe. Rozkład prawdopodobieństwa odcinka czasu, określonego przez dwa stemple czasowe, które zakwalifikowały się kolejno do i -tego oraz j -tego eksperymentu jest splotem dwóch rozkładów równomiernych o wartościach $1/E_i$ oraz $1/E_j$ i szerokościach równych kolejno E_i oraz E_j , tak jak to zostało pokazane na rysunku 6.13b (-). Innymi słowy, uzyskany splot przedstawia rozkład prawdopodobieństwa odcinka czasu będącego wynikiem rejestracji dwóch stempli czasowych, które zakwalifikowały się odpowiednio do eksperymentu E_i oraz E_j .



Rys. 6.13. Wyznaczanie różnicy dwóch stempli czasowych przez splot ich rozkładów: a – rozkład prawdopodobieństwa dwóch stempli czasowych; b – rozkład prawdopodobieństwa uzyskanego odcinka czasu (-)

Fig. 6.13. Determining of two-time-stamp difference by using convolution of their probability density functions: a – probability density function of two time-stamps; b – probability density function of time-interval (-)

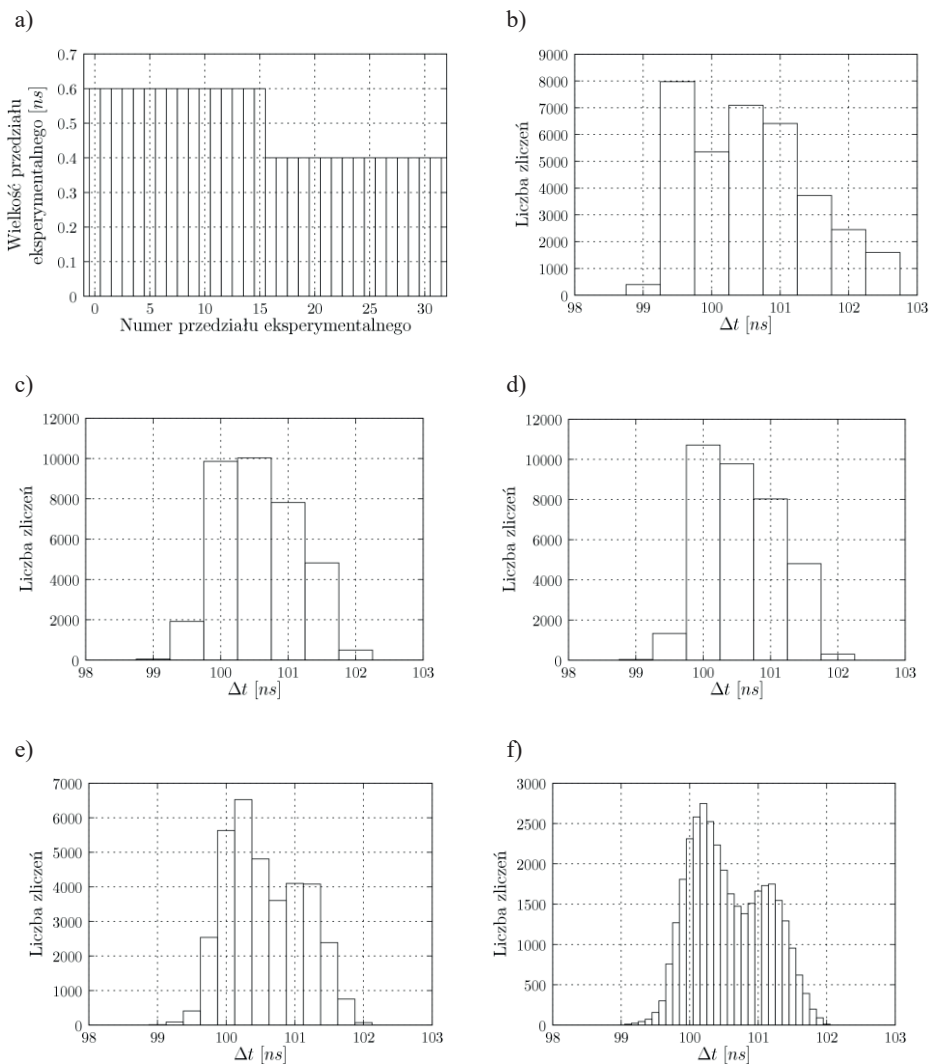
Skonstruowane tym sposobem rozkłady prawdopodobieństwa odcinków czasu są sukcesywnie dodawane do histogramu wynikowego, który może być w tym przypadku sporządzany z dowolną rozdzielczością, niezależnie od rozdzielczości aparatury pomiarowej. Niech $\sigma_{E_{ij}}(t)$ będzie funkcją opisującą rozkład prawdopodobieństwa stempla czasowego, który został zaliczony do przedziału eksperymentalnego E_{ij} ,

wówczas histogram odcinków czasu w przedziale $\Delta t = [t_1, t_2]$ zostanie zwiększony o wartość

$$\rho_{i,j}(t_1, t_2) = \int_{t_1}^{t_2} \sigma_{E_i}(-t) \otimes \sigma_{E_j}(t) dt \quad (6.61)$$

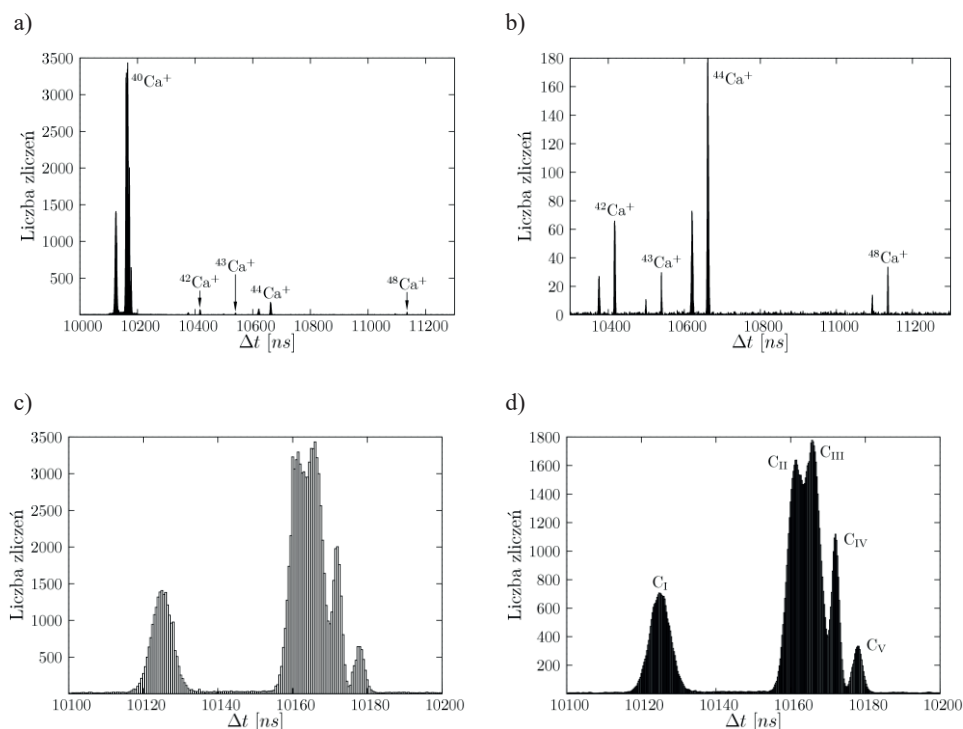
gdzie operator \otimes dokonuje splotu funkcji. Na potrzeby implementacji tej metody została wyznaczona postać analityczna funkcji $\rho_{E_{i,j}}(t_1, t_2)$ w takim celu, aby określanie przyczynków do histogramu odcinków czasu było precyzyjne i nie wymagało dużej mocy obliczeniowej. Metoda ta daje możliwość uzyskania dużo bardziej czytelnych histogramów od tych uzyskiwanych poprzednimi metodami, co dla spektrometrii masowej czasu przelotu, gdzie identyfikacji izotopu dokonuje się głównie w sposób wizualny, ma ogromne znaczenie.

W celu zademonstrowania zalet opisanej metody minimalizacji wpływu nieliniowości oraz kwantyzacji wygenerowane zostały dwa zbiory odcinków czasu, które nieznacznie różniły się długością. Pierwszy zbiór składał się z 20 000 odcinków czasu o średniej wartości równej 100,17 ns oraz odchyleniu standardowym równym 0,03 ns, drugi natomiast zbiór składał się z 15 000 odcinków czasu, jego wartość średnia wynosiła 101,11 ns, a odchylenie standardowe było równe 0,02 ns. Następnie odcinki czasu z obu zbiorów zostały umieszczone w zbiorze wynikowym w losowej kolejności i poddane symulacji pomiaru modułem pomiarowym, którego charakterystykę przedstawiono na rysunku 6.14a. Na rysunku 6.14b przedstawiono histogram odcinków czasu wyznaczonych ze zmierzonych danych w sposób standardowy, czyli przez różnice dwóch stempli czasowych. Jak łatwo zauważyć, uzyskany rezultat w sposób znaczący odbiega od oczekiwanego. Jeżeli do wyznaczenia histogramu użyta zostanie metoda minimalizacji wpływu nieliniowości charakterystyki modułu pomiarowego, to zadane odcinki czasu przedstawiane są w sposób prawidłowy (maksyma rozkładów znajdują się we właściwych miejscach), ale ze względu na niewielką różnicę ich długości (0,94 ns) histogram jest mało czytelny (rys. 6.14c). Podobny rezultat zostanie otrzymany, jeżeli do wyznaczenia histogramu zostanie użyta metoda minimalizacji wpływu nieliniowości oraz kwantyzacji, a histogram wynikowy zostanie wykonany ze średnią rozdzielczością aparatury pomiarowej (w tym przypadku 500 ps), co zostało przedstawione na rysunku 6.14d. Jeżeli jednak histogram zostanie sporządzony z rozdzielczością 250 ps (rozdzielczość dwa razy większa niż średnia rozdzielczość modułu), to obecność dwóch rodzajów odcinków czasu zaczyna być widoczna (rys. 6.14e). Wykonanie histogramu z rozdzielczością 100 ps spowoduje, że zawarte odcinki czasu są wyraźnie widoczne, a maksima ich wystąpienia mogą zostać wyznaczone z dużo większą dokładnością (rys. 6.14f).



Rys. 6.14. Metoda minimalizacji wpływu nieliniowości modułu oraz kwantyzacji: a – przykładowa charakterystyka modułu; b – histogram odcinków czasu uzyskany metodą standardową; c – metodą minimalizującą wpływ nieliniowości modułu; d – metodą minimalizującą wpływ nieliniowości modułu oraz kwantyzacji uzyskany z rozdzielczością 500 ps; e – z rozdzielczością 250 ps; f – z rozdzielczością 100 ps

Fig. 6.14. Method of influence minimization of measuring module nonlinearity and quantization: a – measuring module exemplary characteristic; b – time-interval histogram obtained by using standard method; c – method that minimizes only influence of nonlinearity; d – method that minimizes influence of nonlinearity and quantization applied with resolution equal to 500 ps; e – with resolution equal to 250 ps; f – with resolution equal to 100 ps



Rys. 6.15. Histogramy uzyskane podczas pomiaru składu izotopowego wiązki atomów wapnia: a – całe zarejestrowane widmo izotopowe wapnia, b – widmo izotopowe wapnia bez izotopu $^{40}\text{Ca}^+$, c – fragment widma izotopowego wapnia dla $^{40}\text{Ca}^+$ uzyskany standardową metodą, d – fragment widma izotopowego wapnia $^{40}\text{Ca}^+$ uzyskany z użyciem metody minimalizacji wpływu nieliniowości i kwantyzacji zastosowanej z rozdzielczością 250 ps (komponenty składowe CII oraz CIII stały się wyraźnie widoczne)

Fig. 6.15. Histograms obtained during measuring of isotope composition of calcium atomic beam: a – the whole registered calcium isotopic spectrum; b – calcium isotopic spectrum without $^{40}\text{Ca}^+$ isotope; c – chosen fragment of calcium isotopic spectrum for $^{40}\text{Ca}^+$ obtained using standard method; d – chosen fragment of calcium isotopic spectrum of $^{40}\text{Ca}^+$ obtained by using the method that minimizes the influence of nonlinearity and quantization applied with resolution equal to 250 ps (components CII and CIII have become clearly visible)

Drugim eksperymentem, który uwidacznia skuteczność opisanej metody jest analiza odcinków czasu uzyskanych podczas badania składu izotopowego wiązki atomowej wapnia. Zostało ono przeprowadzone z użyciem spektrometru masowego czasu przelotu firmy COMSTOCK model TOF-101 oraz modułu pomiaru odcinka czasu opisanego w rozdziale 5.

Na rysunku 6.15 przedstawiono histogramy uzyskane podczas eksperymentu. Wyznaczone z eksperymentu procentowe zawartości izotopów wapnia (pola powierzchni poszczególnych pików) są zgodne z danymi literaturowymi. Zgodność ta w głównej

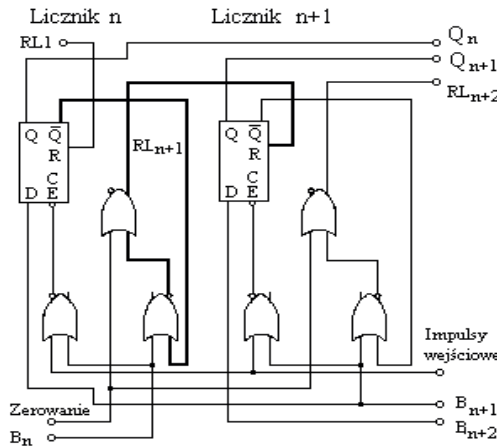
mierze zawdzięcza się optymalizacji modułu pomiaru odcinka czasu pod kątem maksymalnej intensywności rejestrowanych zdarzeń (głównie przez optymalizacje konwertera kodu oraz zrównoleglenie operacji przetwarzania i gromadzenia danych pomiarowych). Wynika z tego, że maksymalna intensywność jonów docierających do detektora ρ_{\max} była dużo mniejsza od maksymalnej (dopuszczalnej) intensywności zdarzeń rejestrowanych przez moduł pomiarowy (wyrażenie (6.1) było spełnione).

Na rysunku 6.15c przedstawiono histogram czasu przelotu dla izotopu wapnia $^{40}\text{Ca}^+$, natomiast na rysunku 6.15d przedstawiono histogram czasu przelotu, które zostały wyznaczone z wykorzystaniem opisanej metody. Komponenty są w tym przypadku bardzo dobrze widoczne i ich obecność nie budzi wątpliwości. Wyjaśnienie pochodzenia poszczególnych komponentów można znaleźć w pracach [154–157] oraz rozdziale 8.

7. SPOSOBY OGRANICZANIA I REDUKCJI BŁĘDÓW

Całkowita lub częściowa redukcja błędów systematycznych oraz ograniczanie błędów przypadkowych należą do podstawowych celów stawianych w trakcie opracowania jakiegokolwiek systemu pomiarowego. W przypadku systemów stosowanych w spektrometrii masowej czasu przelotu o nanosekundowej rozdzielczości można ograniczyć błędy zliczania impulsów wejściowych [57–59, 123, 124].

W systemach wielokanałowych konieczna jest eliminacja efektu podwójnego zliczania. W celu eliminacji tego efektu stosuje się metodę częściowego lub całkowitego blokowania kolejnego kanału czasowego [38, 54].

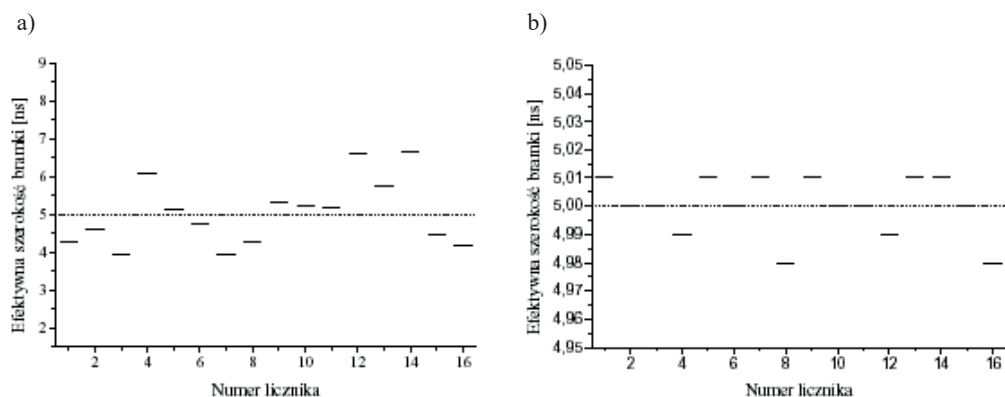


Rys. 7.1. Sposób zabezpieczenia przed podwójnym zliczaniem

Fig. 7.1. Double counts protection circuit

Wykorzystanie dwufazowego sposobu generacji bramek czasowych umożliwia, oprócz dwukrotnego zwiększenia rozdzielczości systemu pomiarowego, realizację prostego układu zabezpieczającego system pomiarowy przed podwójnymi zliczeniami. Zalety tej pozbawiony jest sposób polegający na programowym blokowaniu kolejnego kanału czasowego, aczkolwiek za cenę niewielkiego zwiększenia czasu analizy uzyskuje się znaczne uproszczenie konstrukcji układu. Działanie układu zabezpieczającego z całkowitą blokadą kolejnego kanału czasowego, wykorzystującego dwufazowy układ generacji bramek czasowych, pokazano na rysunku 7.1 [57].

Pasożytniczą modulację szerokości bramek czasowych, którą zilustrowano na rysunku 6.8, można zredukować, stosując metodę zmiennego przyporządkowania liczników do kanałów czasowych [55, 57, 59].



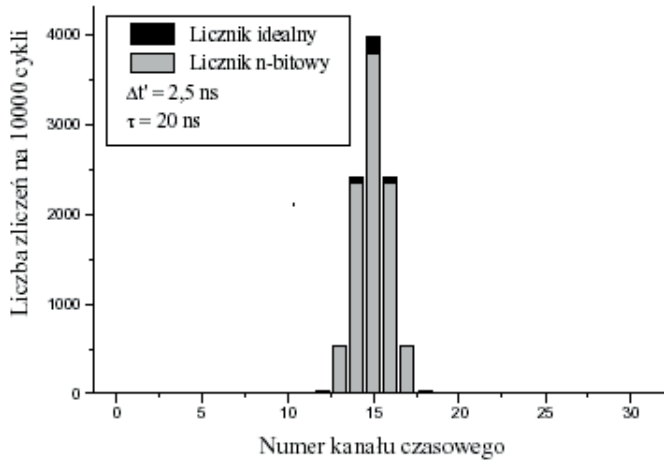
Rys. 7.2. Efektywne szerokości bramek czasowych: a – stałe przyporządkowanie liczników, b – zmienne przyporządkowanie liczników

Fig. 7.2. Effective time-gates width:

a – constant allocation of counters, b – variable allocation of counters

Metoda polega na cyklicznej zmianie przyporządkowania liczników do poszczególnych kanałów czasowych. W tym celu układ sterowania systemem wielokanałowego zawiera dodatkowy licznik binarny zliczający impulsy pochodzące ze źródła wzbudzenia badanego procesu. Stan tego licznika wskazuje na licznik przyporządkowany do kanału pierwszego w każdym cyklu pomiarowym. Zastosowanie tej metody znacznie zmniejsza różnice w efektywnej szerokości poszczególnych kanałów czasowych, co zilustrowano na rysunku 7.2 [57].

W praktyce zmniejszenie fluktuacji szerokości bramki czasowej oznacza zmniejszenie estymatora wariancji odchyłek dla uzyskanych rezultatów pomiarowych. W przypadku systemu z bramką przesuwną efekt redukcji liczby zliczeń, pojawiający się na skutek skończonej rozdzielczości licznika (n -bitowego) oraz skończonej szerokości impulsów wejściowych, zilustrowano na rysunku 7.3.



Rys. 7.3. Efekt redukcji liczby zliczeń w systemach z bramką przesuwaną
Fig. 7.3. Number of counts reduction in shifted gate systems

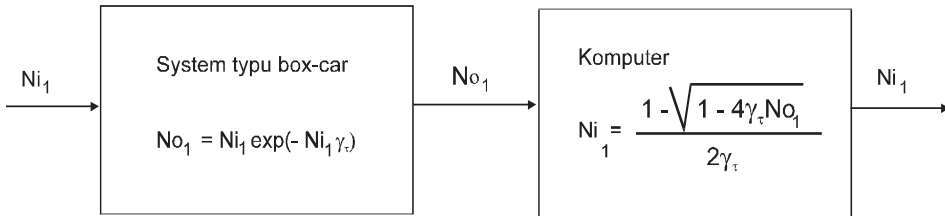
W takich przypadkach można dokonać korekty liczby zliczeń w poszczególnych kanałach czasowych. Odzyskanie „zgubionych impulsów” jest możliwe, gdy wykorzystamy równanie (6.4).

Liczbę impulsów wejściowych można określić, wiedząc, że jest spełniony warunek

$$\frac{N_{i_1} \Delta t'}{\tau} \ll 1, \quad (7.1)$$

co prowadzi do przybliżonej zależności [13]

$$N_{0_1} \cong N_{i_1} - \gamma_{\tau} N_{i_1}^2. \quad (7.2)$$



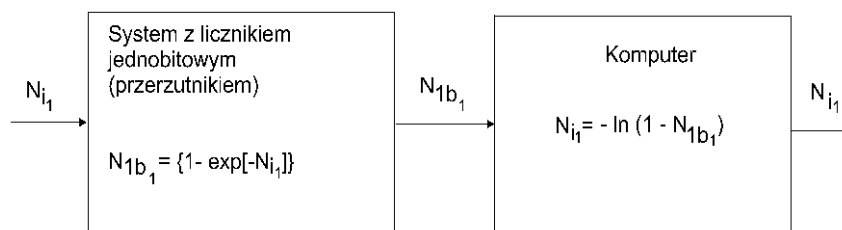
Rys. 7.4. Schemat algorytmu korekcyjnego dla systemu z bramką przesuwaną
Fig. 7.4. Diagram of the correction algorithm for shifted gate system

Liczbę impulsów wejściowych otrzymuje się, rozwiązując równanie (7.2) przy spełnionym warunku $4\gamma_{\tau} N_{0_1} < 1$.

$$N_{i_1} = \frac{1 - \sqrt{1 - 4\gamma_\tau N_{0_1}}}{2\gamma_\tau}. \quad (7.3)$$

Przeprowadzona analiza może być użyteczna w szczególności dla systemów z wąską przesuwą bramką czasową. Schemat algorytmu korekcyjnego przedstawiono na rysunku 7.4 [13, 58]. W systemach zliczających z wąską bramką czasową wartość oczekiwana liczby zliczeń na wyjściu wynosi $\mu = \tau \rho$, gdzie τ – szerokość bramki, a ρ – intensywność procesu. Wartość ta jest zwykle < 1 . W systemach wielokanałowych jest podobnie dla poszczególnych kanałów czasowych $\mu_i = \tau \rho_i$, gdzie i – numer kanału czasowego. W związku z tym prawdopodobieństwo zliczenia więcej niż jednego impulsu w trakcie pojedynczego cyklu pomiarowego jest małe $P(n > 1) \approx 0$ i prowadzi do uproszczenia systemu zliczającego do licznika jednobitowego.

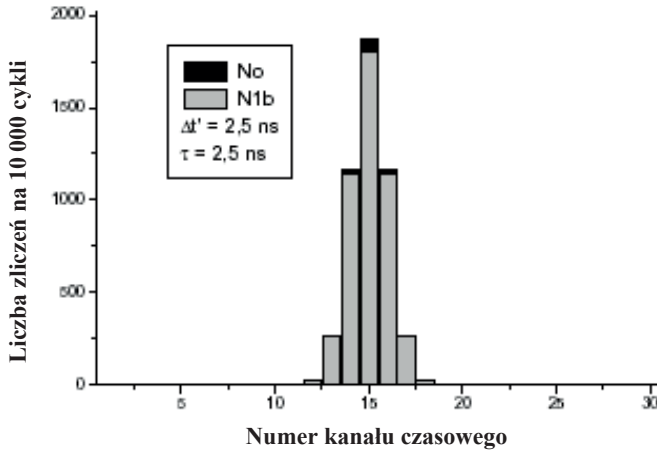
Stosowanie licznika jednobitowego uzasadnia rozkład prawdopodobieństwa dla małych wartości $\mu < 0,5$, dla których prawdopodobieństwo pojawienia się więcej niż jednego impulsu na wejściu licznika w czasie τ jest mniejsze niż 0,1. Nawet tak małe prawdopodobieństwo, z dużą liczbą cykli pomiarowych, może powodować istotne straty w liczbie zliczanych przez licznik impulsów. W takim przypadku istnieje także możliwość dokonania korekty liczby zliczonych impulsów, według algorytmu korekcyjnego, który przedstawiono na rysunku 7.5 [13].



Rys. 7.5. Schemat algorytmu korekcyjnego dla systemu z licznikiem jednobitowym
Fig. 7.5. Diagram of the correction algorithm for system with one-bit counter

Działanie liczników n -bitowego i jednobitowego w przypadku stosowania wąskiej bramki czasowej porównano na rysunku 7.6.

Ujemnym skutkiem stosowania układów zabezpieczających wielokanałowe systemy przed podwójnymi zliczeniami są dodatkowe straty w liczbie zliczeń, wynikające z wprowadzenia dodatkowych nie rozszerzających się stref czasu martwego. Straty te są szczególnie istotne w kanałach obejmowanych przez te przedziały czasu, w których intensywność procesu jest bardzo duża. Zakładając, że częstość pojawiania się impulsów w poszczególnych kanałach czasowych opisuje rozkład prawdopodobieństwa Poissona, straty te można oszacować. Można zatem dokonać korekty liczby zliczeń dla analizowanego procesu.



Rys. 7.6. Efekt zliczania impulsów przez licznik n -bitowy N_0 i jednobitowy N_{1b}
 Fig. 7.6. Effect of counting by n -bit counter N_0 , and one-bit counter N_{1b}

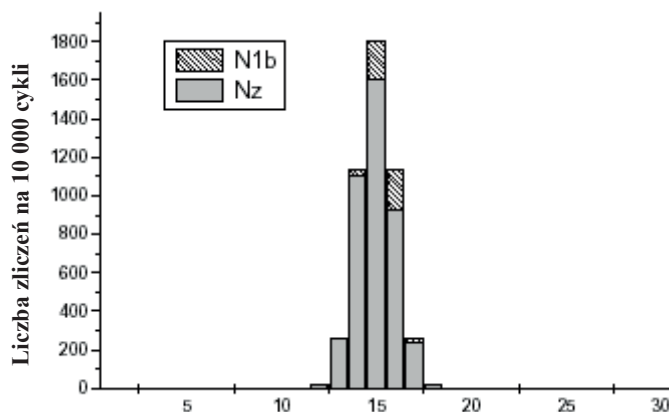
Skutkiem stosowania zabezpieczenia jest zmniejszenie liczby zliczeń $N_z(p)$ w kolejnym kanale czasowym p , na skutek dodatkowego blokowania bramki $G(p)$ sygnałem $V(p-1)$. Efekt ten nie jest istotny przy $\rho\tau \ll 1$ ze względu na niewielkie prawdopodobieństwo zablokowania kolejnego kanału czasowego równe

$$P(V_{(p-1)} = 0) = \frac{N_z(p-1)}{M}. \quad (7.4)$$

W przypadku kiedy warunek ten nie jest spełniony, należy wprowadzić poprawkę do liczby zliczeń we wszystkich kanałach czasowych z wyjątkiem pierwszego. Liczbę zliczeń $N_{zM}(p)$ dla systemu z zabezpieczeniem przed podwójnymi zliczeniami wyrazić można w postaci [13]

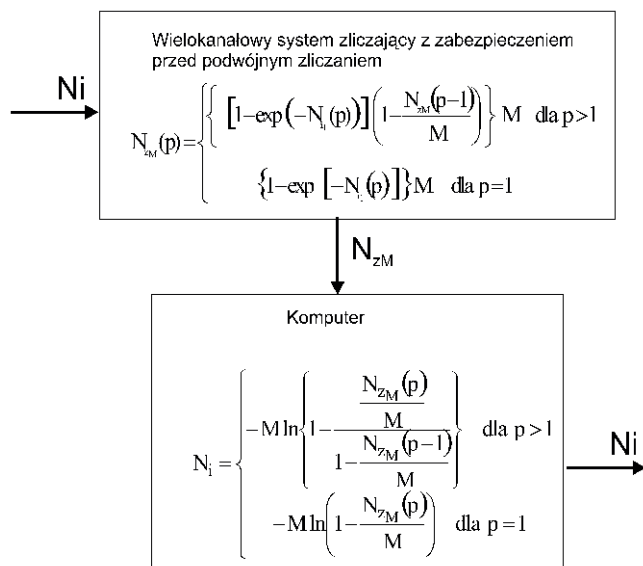
$$N_{zM}(p) = \begin{cases} \left\{ -\ln \left[\exp(-N_{i_1}(p)) \left(1 - \frac{N_{zM}(p-1)}{M} \right) \right] \right\} M & \text{dla } p > 1 \\ \{1 - \exp[-N_{i_1}(p)]\} M & \text{dla } p = 1 \end{cases}. \quad (7.5)$$

Efekt blokady kanałów czasowych w systemie wielokanałowym, wynikający ze stosowania układu zabezpieczającego przed podwójnym zliczaniem impulsów dla licznika jednobitowego, zilustrowano na rysunku 7.7. Schemat algorytmu korekcyjnego dla wielokanałowego systemu zliczającego z układem zabezpieczającym przed podwójnym zliczaniem przedstawiono na rysunku 7.8 [13].



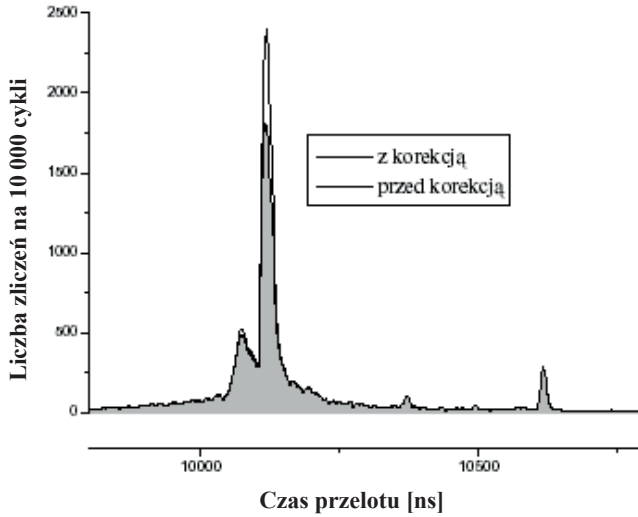
Rys. 7.7. Efekt zabezpieczenia systemu wielokanałowego, gdzie: N_{1b} – liczba zliczeń bez układu zabezpieczającego, N_z – liczba zliczeń przy stosowaniu układu zabezpieczającego
 Fig. 7.7. Effect of multichannel scaler protection: N_{1b} – number of counts without protection, N_z – number of counts with protection

Wszystkie straty w liczbie zliczeń w systemach wielokanałowych zależą od intensywności zliczanych impulsów. Straty spowodowane zastosowaniem liczników jednobitowych rosną szybko wraz ze wzrostem szerokości bramek czasowych.



Rys. 7.8. Algorytm korekcyjny dla wielokanałowego systemu zliczającego
 Fig. 7.8. Correction algorithm for multichannel scaler

Wynik zastosowania algorytmu korekcyjnego w spektrometrii masowej czasu przelotu w badaniach nad wielofotonową jonizacją atomów wapnia pokazano na rysunku 7.9 [58].



Rys. 7.9. Efekt zastosowania algorytmu korekcyjnego w spektrometrii masowej czasu przelotu
Fig. 7.9. Effect of correction algorithm application in time-of-flight mass spectrometry

W systemach o rozdzielczości pikosekundowej czas martwy jest zwykle wielokrotnie dłuższy od szerokości pojedynczego kanału czasowego [55]. Oznacza to, że w każdym cyklu pomiarowym pojedynczy impuls wejściowy blokuje j kolejnych kanałów czasowych. Liczba zliczonych impulsów w kanale p ma wpływ na liczbę zliczeń w kanałach od p do $p + j$.

Oznaczając przez:

$N_i(p)$ – liczbę impulsów pojawiających się na wejściu kanału p w ciągu M cykli pomiarowych,

$N_i(q)$ – liczbę impulsów pojawiających się na wejściu kanału q w ciągu M cykli pomiarowych ($p < q < p + j$),

$N_0(p)$ – liczbę zliczeń w kanale p w ciągu M cykli pomiarowych,

$N_0(q)$ – liczbę zliczeń w kanale p w ciągu M cykli pomiarowych,

można określić prawdopodobieństwo pojawienia się impulsu w kanale p blokującego kanał q

$$P_M(p \rightarrow q) = \frac{N_0(p)}{M}, \quad (7.6)$$

oraz prawdopodobieństwo pojawienia się impulsu na wejściu kanału q

$$P_M(q) = \frac{N_i(q)}{M}. \quad (7.7)$$

Prawdopodobieństwo utraty impulsu w kanale q , spowodowane pojawianiem się impulsów w kanale p , będzie zatem równe

$$P_{M_{\text{str}}}(p \rightarrow q) = P_M(q)P_M(p \rightarrow q). \quad (7.8)$$

Straty w kanale $p + j$ będą zatem wyrażone zależnością

$$P_{M_{\text{str}}}(p + j) = \sum_{q=p}^{p+j} P_M(p + j)P_M(q \rightarrow p + j). \quad (7.9)$$

Zależność (7.9) można przedstawić w postaci

$$\frac{N_i(p + j) - N_0(p + j)}{M} = \frac{N_i(p + j)}{M} \sum_{q=p}^{p+j} \frac{N_0(q)}{M}. \quad (7.10)$$

Liczba impulsów pojawiających się na wejściu kanału $p + j$ w ciągu M cykli pomiarowych może być obliczona z zależności

$$N_i(p + j) = \frac{N_0(p + j)}{1 - \sum_{q=p}^{p+j} \frac{N_0(q)}{M}}. \quad (7.11)$$

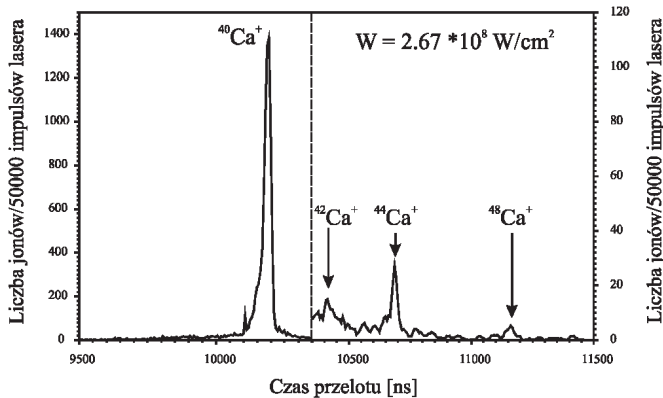
Korekcja liczby zliczeń może być dokonana według zależności (7.11) dopiero po wykonaniu wszystkich M cykli pomiarowych. Wpływ czasu martwego na liczbę zliczeń systemu pomiarowego wykorzystującego interpolatory odcinka czasu można ograniczyć, stosując system interpolacji wielokanałowej w stosunku do pojedynczego sygnału pomiarowego [62].

8. POMIARY SKŁADU IZOTOPOWEGO WIĄZKI ATOMÓW WAPNIA

Opisane wcześniej metody pomiaru umożliwiają otrzymanie widma czasu przelotu (TOF) w przedziale 1–40 μs z rozdzielczością 6,25 ns niezależnie od intensywności wiązki lasera i długości fali promieniowania. Przykładowe widmo TOF uzyskiwane w eksperymencie przy gęstości wiązki atomowej Ca wynoszącej $2 \cdot 10^{11}$ atom/ cm^3 przedstawiono na rysunku 8.1.

W celu uwidocznienia struktury izotopowej, widmo przedstawione na tym rysunku zostało podzielone na dwie części. Każda z nich ma odrębną oś pionową oznaczającą liczbę zarejestrowanych jonów powstałych podczas 50 000 impulsów lasera. Zawartość

różnych izotopów dla atomu wapnia wynosi: ^{40}Ca – 96,97%, ^{42}Ca – 0,64%, ^{43}Ca – 0,145%, ^{44}Ca – 2,06%, ^{46}Ca – 0,00033%, ^{48}Ca – 0,18%. Podczas analizy otrzymanego widma udało się oznaczyć tylko izotopy: ^{40}Ca , ^{42}Ca , ^{44}Ca oraz ^{48}Ca . Otrzymane czasy przelotu jonów Ca^+ dla poszczególnych izotopów obliczone ze wzoru (2.13) w przypadku zerowej energii początkowej wynoszą: ^{40}Ca – 10,2 μs , ^{42}Ca – 10,45 μs , ^{44}Ca – 10,7 μs , ^{48}Ca – 11,2 μs . Lokalne maksima odpowiadające tym izotopom uzyskane w pomiarze przypadają dla czasu nieznacznie krótszego. Niemniej jednak znajdują się w przedziale błędu spowodowanego: zaniedbaniem skrócenia czasu przelotu przez przyspieszanie, jakie uzyskują jony, przelatując przez obszar przyspieszania spektrometru oraz zaniedbaniem przez wytworzony jon pewnej składowej prędkości początkowej w kierunku detektora. Brak w widmie izotopu ^{46}Ca można łatwo wytłumaczyć niewielką zawartością procentową tego izotopu w całej populacji. Natomiast uzasadnienie niewystępowania izotopu ^{43}Ca nie jest już takie proste, gdyż jego zawartość procentowa jest tylko nieznacznie mniejsza niż izotopu ^{48}Ca – dobrze widocznego w widmie. Czas przelotu jonu izotopu ^{43}Ca wynosi 10,6 μs . Jeśli uwzględni się przesunięcie całego widma w kierunku krótszego czasu, to dla czasu około 10,57 μs można zauważyć wzmocnienie sygnału jonowego z liczbą zliczeń porównywalną do liczby jonów dla izotopu ^{48}Ca . Jednoznaczna interpretacja tego wyniku jest jednak niemożliwa. Głównym tego powodem jest poszerzenie każdego z pików odpowiadającym poszczególnym izotopom Ca, które pociąga za sobą rozmycie całego widma czasu przelotu.

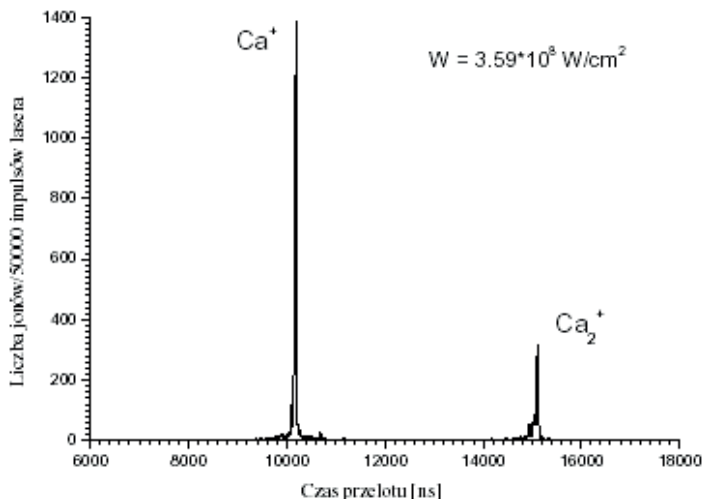


Rys. 8.1. Widmo określające skład izotopowy otrzymane z zastosowaniem systemu wielokanałowego
Fig. 8.1. Mass-spectrum obtained by using multichannel scaler

Położenie pików w widmie czasu przelotu silnie zależy od stabilności napięcia przyspieszającego. Niestabilność rzędu ± 1 V daje przesunięcie położenia maksimum o ± 5 ns [156]. Rozmycie całego widma może być również spowodowane występowaniem efektów plazmowych, tzn. efektów wywołanych wzajemnym oddziaływaniem jonów i elektronów powstałych w wyniku oddziaływania z impulsem laserowym [158, 159].

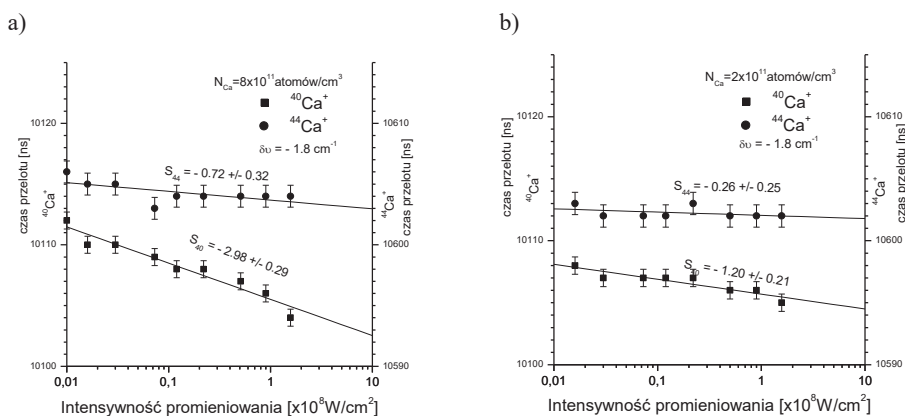
Pojawianie się tych efektów wynika z powodu istnienia różnych ruchliwości elektronów i jonów w statycznym polu elektrycznym. Na początku trwania impulsu laserowego rośnie liczba jonów, a pole przez nie wywołane przesuwają poziomy energii atomów w wiązce (statyczne przesunięcie Starka) w kierunku większej energii. Dodatkowe przesunięcie poziomów wprowadzone jest przez promieniowanie laserowe (dynamiczny efekt Starka). Podczas trwania zatem impulsu laserowego wydajność procesu w danej chwili zależna jest od aktualnej wartości przesunięcia poziomu rezonansowego w wyniku oddziaływania pola. Zasadniczo pułapkowanie elektronów, które może ekranować pole jonów, jest bez znaczenia przy gęstości jonów mniejszej niż 10^8 jon/cm³. Przy większych gęstościach jonów efekt ten staje się coraz bardziej znaczący [160]. Elektrony oscylują wokół chmury jonowej i w ten sposób ekranują pole wywołane przez jony, zachowując neutralność powstałej plazmy. Dopóki zachowana jest neutralność, dopóty jony nie są wypychane z obszaru oddziaływania. Krótko mówiąc, fotojonizacja jest modulowana przez oscylacje elektronów. A zatem można oczekiwać, że jony nie będą wypychane w sposób ciągły, lecz paczkami. Pierwsza chmura zaczyna przemieszczać się w zewnętrznym polu elektrycznym, następna chmura jest tworzona dopiero wtedy, gdy elektrony przywrócą możliwość fotojonizacji. Sytuacja ta powtarza się przez cały czas trwania impulsu laserowego. Pod koniec impulsu laserowego chmury są ustawione wzdłuż zewnętrznego pola elektrycznego w kierunku wejścia do spektrometru masowego TOF. Następnie chmury jonów, znajdujące się bliżej wejścia spektrometru, są przyspieszane przez „dalsze” chmury. W tym samym czasie chmury „dalsze” są opóźniane przez odpychanie kulombowskie chmur znajdujących się przed nimi. Dodatkowo na każdą wyprodukowaną chmurę działa pole pochodzące od elektronów. Zjawisko oddziaływania elektronów na jony jest bardzo krótkotrwałe ze względu na dużą ruchliwość elektronów. Wszystkie opisane zjawiska powodują zmianę kształtu powstających pików jonowych w zależności od liczby produkowanych jonów lub względnego przesunięcia pików (odpowiadających poszczególnym rodzajom jonów) w widmie masowym otrzymanym ze spektrometru TOF [156].

Oprócz izotopów Ca^+ w widmie czasu przelotu przede wszystkim można by się spodziewać pojawienia się pików: Ca^{2+} – odpowiadającego dwukrotnie zjonizowanemu atomowi wapnia i Ca_2^+ – będącego klasterem jonu Ca^+ i cząstki obojętnej. Niemniej jednak w większości serii pomiarowych nie udało się uzyskać wyników potwierdzających występowanie w widmie TOF wymienionych struktur. Najprostszy sposób oszacowania czasu przelotu Ca^{2+} wymaga podzielenia czasu odpowiadającego Ca^+ przez czynnik 1,414 i analogicznie do Ca_2^+ pomnożenie przez czynnik 1,414. Daje to odpowiednio wyniki: 7,3 μs dla Ca^{2+} i 14,7 μs dla Ca_2^+ . Podczas wykonywania eksperymentu okazało się, że występowanie tego typu struktur może zależeć od kilku czynników, spośród których najbardziej istotną rolę odgrywa intensywność promieniowania laserowego. Na rysunku 8.2 przedstawiono widmo masowe w przedziale 6–18 μs dla największej intensywności promieniowania laserowego uzyskanej podczas wszystkich pomiarów [12].



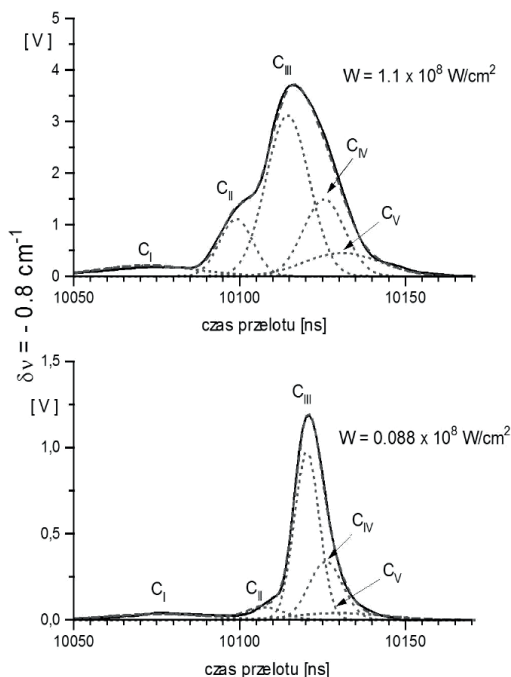
Rys. 8.2. Widmo masowe otrzymane z systemu wielokanałowego w przedziale 6–18 μ s
 Fig. 8.2. Mass-spectrum obtained by multichannel scaler in range 6–18 μ s

Na rysunku widoczne jest pojawienie się dodatkowego pików dla czasu przelotu około 15 μ s, dosyć dobrze odpowiadającego czasowi przelotu cząstki Ca_2^+ . Różnica pomiędzy wartością eksperymentalną a oszacowaną teoretycznie może wynikać z kilku czynników: niedoszacowania, występowania efektów plazmowych, ale także z możliwej różnicy czasu tworzenia się jonu i klastra (tzn. czas tworzenia klastra intuicyjnie wydaje się być dłuższy niż czas powstania jonu). Drugim czynnikiem warunkującym występowanie pików Ca_2^+ okazała się długość fali padającego promieniowania.



Rys. 8.3. Odległość pomiędzy pikami $^{40}\text{Ca}^+$ i $^{44}\text{Ca}^+$ w zależności od intensywności lasera
 Fig. 8.3. Distance between $^{40}\text{Ca}^+$ and $^{44}\text{Ca}^+$ peaks versus laser intensity

Analizując wyniki otrzymane dla trójfotonowej jonizacji wapnia, należy zwrócić uwagę na efekty, które wydają się być silnie związane z efektem kolektywnego oddziaływania cząstek naładowanych: jonów między sobą i jonów z elektronami. Zaobserwowane efekty to: zmiana kształtu (asymetria) rejestrowanego głównego pików jonowego i przesunięcia czasu przelotu jonów w widmie otrzymanym ze spektrometru. Wielkość tych efektów zależy głównie od gęstości wiązki atomowej, gęstości ładunku, intensywności promieniowania laserowego i stopnia dostrojenia do przejścia rezonansowego do stanu $4s6s^1S_0$ [154]. Pojawianie się efektu przesunięcia położenia pików $^{40}\text{Ca}^+$ i $^{44}\text{Ca}^+$ pomiędzy sobą w funkcji zmiany intensywności promieniowania lasera zależy głównie od gęstości wiązki atomowej w obszarze oddziaływania i w konsekwencji liczby otrzymywanych jonów (rys. 8.3). Informację o tym efekcie dają różne nachylenia S_{40} i S_{44} linii położenia pików izotopów $^{40}\text{Ca}^+$ i $^{44}\text{Ca}^+$ w funkcji intensywności promieniowania lasera. Nachylenia S_{40} i S_{44} pokazują ponadto drugi efekt związany z trójfotonową jonizacją atomów wapnia przez stan $4s6s^1S_0$. Im większy bowiem jest współczynnik nachylenia linii położenia pików izotopów wapnia w funkcji intensywności promieniowania, tym bardziej dany pik jest przesunięty w skali czasowej, względem impulsu wyzwalającego, w stronę krótszego czasu przelotu. Dla gęstości 2×10^{11} atomów/cm³ efekt jest bardzo mały, współczynniki S wynoszą odpowiednio:



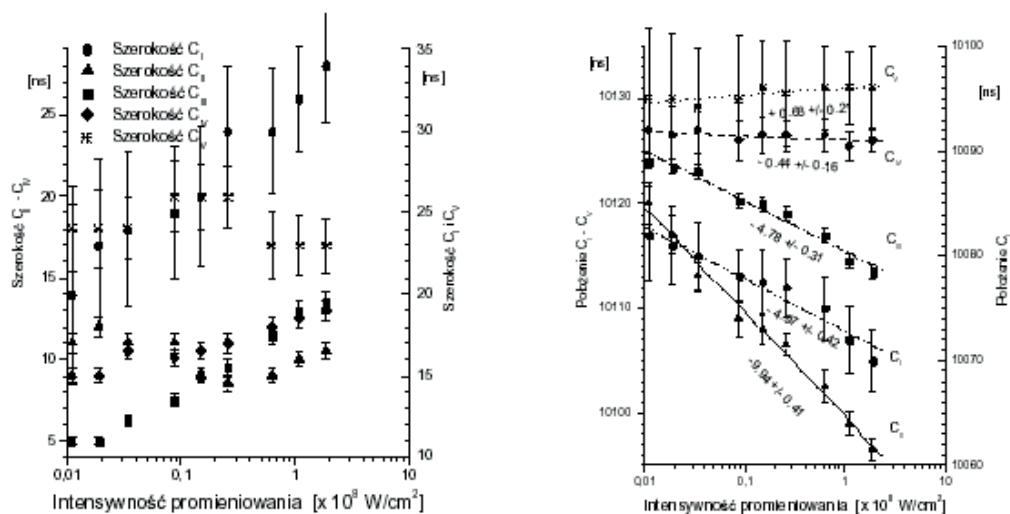
Rys. 8.4. Rozkład na składowe gaussowskie pików $^{40}\text{Ca}^+$
 Fig. 8.4. Gaussian decomposition of $^{40}\text{Ca}^+$ peak

$S_{40} = -1,20 \pm 0,24$ i $S_{44} = -0,26 \pm 0,24$. Natomiast dla gęstości 8×10^{11} atomów/cm³ efekt staje się bardzo widoczny: $S_{40} = -2,98 \pm 0,29$ i $S_{44} = -0,72 \pm 0,32$. Z powodu istnienia pewnej wartości progowej obserwowanych zjawisk wszystkie pozostałe wyniki omawiane w tym rozdziale odnoszą się do gęstości wiązki atomowej 8×10^{11} atomów/cm³. Na rysunku 8.4 przedstawiono zależność odległości między pikami i nachylenia linii S_{40} i S_{44} w funkcji gęstości promieniowania laserowego dla czterech różnych odstrojeń od rezonansu. W obszarze niewielkich odstrojeń od rezonansu zmiana położenia pików $^{40}\text{Ca}^+$ jest prawie jednakowa. Współczynnik nachylenia S dla tego pików oscyluje wokół wartości $-3,5$. Natomiast nachylenie linii położenia pików $^{44}\text{Ca}^+$ maleje w miarę zbliżania się do rezonansu. Podyktowane jest to wzrostem liczby jonów.

Wzrost liczby jonów powoduje wzrost siły odpychania się chmur izotopów podczas ich rozseparowywania. Dla dużych odstrojeń (rzędu 50 cm^{-1}), rysunek 4.4a, efekty wzajemnego odpychania się chmur jonowych zanikają. Następuje dość znaczne zmniejszenie liczby jonów pojawiających się w obszarze oddziaływania i zanikanie wszelkich efektów plazmowych (kolektywnych). Należy nadmienić, że występuje pełna zgodność pomiędzy wynikami: zmiany położenia pików w zależności od intensywności promieniowania laserowego dla różnych odstrojeń od rezonansu i zmiany położenia pików w funkcji odstrojenia od rezonansu dla różnych intensywności promieniowania laserowego. Położenie pików Ca^+ rejestrowanych z użyciem spektrometru masowego TOF jest bardzo czułe na zmiany napięcia zasilania spektrometru. Analiza wyników otrzymanych ze spektrometru masowego dla różnych odstrojeń i intensywności promieniowania laserowego wykazuje istnienie asymetrii pików $^{40}\text{Ca}^+$. Dla małej intensywności promieniowania asymetria spowodowana jest istnieniem wydłużonego zbocza opadającego profilu jonowego. Dla większych intensywności zbocze narastające przesuwa się. Występuje bardzo duży rozrzut wartości asymetrii w miarę zbliżania się do rezonansu.

Położenie maksimum pików $^{40}\text{Ca}^+$ nie znajduje się zawsze w tym samym kanale czasowym układu monitorującego i zależy od mocy lasera oraz odstrojenia od rezonansu. Można więc powiedzieć, że asymetria jest między innymi wynikiem przesuwania się maksimum pików $^{40}\text{Ca}^+$ na osi czasu. Można założyć, że przestrzenny rozkład gęstości promieniowania laserowego ma charakter funkcji Gaussa, więc i rozkład przestrzenny jonów w kierunku osi x powinien mieć podobny charakter. W wyniku zastosowania tej metody najlepsze dopasowanie do profilu doświadczalnego uzyskano dla pięciu składowych pików $^{40}\text{Ca}^+$.

Wyniki rozkładu na składowe gaussowskie przedstawiono na rysunku 8.4. Pierwszy komponent oznaczony jako C_I znajduje się około 50 ns przed głównym pikiem. Komponenty C_{II} – C_V zawarte są w głównym piku $^{40}\text{Ca}^+$. W wyniku tak przeprowadzonej analizy możliwe jest przedstawienie położenia na osi czasu szerokości i powierzchni każdego komponentu w funkcji intensywności promieniowania lub w funkcji odstrojenia od rezonansu.



Rys. 8.5. Szerokość i położenie składowych piku $^{40}\text{Ca}^+$
Fig. 8.5. Width and position of the $^{40}\text{Ca}^+$ peak

Na rysunku 8.5 przedstawiono pozycje poszczególnych komponentów w funkcji intensywności promieniowania laserowego. Położenie komponentu C_{III} , nazwanego głównym, odpowiada zmianom w czasie położenia maksimum głównego piku w funkcji intensywności promieniowania laserowego. Charakter tej zależności jest podobny do zależności położenia piku $^{40}\text{Ca}^+$ w funkcji intensywności promieniowania laserowego. Linie łączące czasowe położenia komponentów C_{II} i C_{IV} w funkcji intensywności promieniowania laserowego zbiegają się w jednym punkcie wraz z linią dla komponentu C_{III} . Natomiast komponenty C_I i C_V wykazują inną zależność przesunięcia w czasie od gęstości promieniowania laserowego. Sytuacja jest podobna w przypadku szerokości poszczególnych komponentów. Szerokość komponentów $C_{II} - C_{IV}$, a także zmiana ich szerokości jest podobna (jedynie szerokość komponentu C_{II} dla małych gęstości wykazuje lekkie odchylenie w stosunku do komponentów C_{III} i C_{IV}), natomiast dla komponentów C_I i C_V szerokość i dynamika jej zmian jest dużo większa.

Biorąc pod uwagę uzyskane wyniki, można stwierdzić, że tylko komponenty C_{II} , C_{III} i C_{IV} mają znaczenie fizyczne. Pozostałe komponenty są wynikiem występowania efektu zwanego ogólnie efektem aparaturowym. Pojawianie się komponentu C_V jest związane najprawdopodobniej z metodą detekcji sygnału. Istnienie komponentu C_I nie jest jeszcze w pełni zrozumiałe. Jest on widoczny zarówno przy małych gęstościach jonów, jak i przy dużych. Ponadto jest widoczny niezależnie od metody detekcji sygnału ze spektrometru. Jednakże oba komponenty C_I i C_V są bardzo małe w stosunku do pozostałych, szczególnie dla dużych intensywności promieniowania laserowego.

W procesie fotojonizacji podczas fazy produkcji jonów (tzn. w czasie trwania impulsu laserowego) może powstać relatywnie duża gęstość jonów (10^{10} jonów/ cm^3),

oprócz efektów nazywanych ogólnie oddziaływaniami atomowymi, pojawiają się zjawiska związane z powstawaniem plazmy, które mogą modyfikować otrzymane wyniki. O ile oddziaływania atomowe, głównie pojedyncze zderzenia, mają istotny wpływ dla zakresu małych odległości, o tyle efekty plazmowe, jako zjawiska makroskopowe, mogą zmienić średnie wielkości, np. rozkład prędkości w obszarze oddziaływania i obszarze przyspieszania spektrometru, co prowadzi do zmiany kształtu otrzymanego sygnału jonowego ze spektrometru masowego. W związku z tym do pełnego opisu zjawiska fotojonizacji należy dodać opis powstałej plazmy z uwzględnieniem efektów wynikających ze zbiorowego (kolektywnego) oddziaływania powstających w wyniku fotojonizacji ładunków. Efekty te ogólnie nazywane są efektami kolektywnymi.

Ze względu na miejsce produkcji jonów (obszar oddziaływania spektrometru masowego ze stałym polem elektrycznym), jak i ze względu na sposób detekcji sygnału (spektroskopia masowa czasu przelotu) pojawianie się oddziaływań kolektywnych może objawiać się zmianą wartości i kształtu rejestrowanego prądu elektronowego [158] lub zmianą kształtu widma masowego. Tak więc w czasie fazy produkcji jonów, jak i fazy zbierania sygnału może nastąpić zmiana kształtu powstających pików jonowych w zależności od liczby produkowanych jonów lub względne przesunięcia pików odpowiadających poszczególnym rodzajom jonów [3, 156].

Pojawianie się efektów kolektywnych wynika z istnienia różnych ruchliwości elektronów i jonów w polu elektrycznym. Na początku trwania impulsu laserowego liczba jonów rośnie i pole wywołane przez nie przesuwają poziomy energii atomów w wiązce (statyczne przesunięcie Starka). Dodatkowe przesunięcie poziomów wprowadzone jest przez promieniowanie laserowe (dynamiczny efekt Starka). W ten sposób podczas fazy produkcji jonów (podczas trwania impulsu laserowego) wydajność procesu w danej chwili zależy od aktualnej wartości przesunięcia poziomu rezonansowego w wyniku oddziaływania pola kolektywnego. Dodatkowo dla dużej gęstości jonów (więcej niż 10^8 jonów/cm³) efekt ekranowania pola jonowego przez elektrony staje się znaczący. Elektrony oscylują wokół chmury jonowej i w ten sposób ekranują pole wywołane przez jony, zachowując neutralność powstałej plazmy. Dopóki zachowana jest neutralność, dopóty jony nie są wypychane z obszaru oddziaływania.

W podanym modelu oddziaływań kolektywnych wynik jonizacji jest modulowany przez oscylację elektronową z częstotliwością plazmową. Dlatego jony nie są wypychane w sposób ciągły, lecz w paczkach. Po „wypchnięciu” elektronów przez zewnętrzne pole elektryczne pierwsza chmura jonów rozpoczyna ruch w zewnętrznym polu. Druga chmura powstaje wtedy, gdy elektrony „wezwoła” na ponowne jej powstanie. Sytuacja ta powtarza się aż do momentu zakończenia się impulsu laserowego. Pod koniec impulsu laserowego chmury są poukładane w szereg wzdłuż linii sił zewnętrznego pola elektrycznego i poruszają się w kierunku anody spektrometru masowego. Chmura znajdująca się bliżej siatki G₂ i G₃ spektrometru jest dodatkowo przyspieszana w wyniku kulombowskiego odpychania się chmur, natomiast chmura

znajdująca się na końcu szeregu będzie najbardziej opóźniana. Ponadto na każdą wyprodukowaną chmurę działa dodatkowo pole pochodzące od elektronów. Zjawisko oddziaływania elektronów na jony jest bardzo krótkotrwałe ze względu na dużą ruchliwość elektronów. Dodatkowo następuje kulombowskie i termiczne rozszerzanie się każdej chmury z osobna. W ten sposób następuje poszerzenie pików w widmie masowym otrzymanym ze spektrometru, modyfikacja wzajemnych relacji w czasie przelotu przez rurę swobodnego przelotu spektrometru, jak i zmiana kształtu otrzymywanego widma.

9. PODSUMOWANIE

W pracy dokonano opisu problematyki związanej z pomiarem składu izotopowego wiązki atomowej z wykorzystaniem spektrometru masowego czasu przelotu. Szczegółowo omówiono wszystkie składowe zestawy pomiarowe oraz metody minimalizujące wpływ niedoskonałości aparatury na wynik pomiaru. Dokonano przeglądu źródeł jonów, detektorów oraz analizatorów, przy czym skupiono się prawie wyłącznie na systemie ze spektrometrem masowym czasu przelotu. Typowe dla tego systemu elementy, takie jak układ próżniowy, układ generacji wiązki, układ laserowy i optyczny zostały przedstawione w rozdziale drugim. Rozdział trzeci zawiera opis układów pomiarowych z pojedynczą bramką przesuwaną, a rozdział czwarty opisuje systemy wielokanałowe o rozdzielczości piko- i nanosekundowej. Rozdział piąty przedstawia układy pomiarowe implementowane w strukturach programowalnych FPGA i opisane w języku VHDL, które wykorzystują wielosegmentowe linie opóźniające. Rozdział szósty przedstawia wybrane metody, które zwiększają dokładność pomiarową systemu ze spektrometrem masowym czasu przelotu. Dokonano tutaj, między innymi, opisu wpływu zbyt dużej intensywności rejestrowanych impulsów na liczbę zliczeń oraz opisu wpływu niestabilności generatora wzorcowego na dokładność wyznaczania odcinków czasu. Sporo miejsca w tym rozdziale poświęcono na przedstawienie metod minimalizujących wpływ nieliniowości charakterystyki modułu pomiarowego na histogram odcinków czasowych. W ostatnim rozdziale przedstawiono wyniki pomiarów uzyskanych podczas badania składu izotopowego wiązki atomów wapnia wraz z ich interpretacją.

LITERATURA

- [1] JOHNSTONE R.A.W., ROSE M.E., *Spektrometria mas*, PWN, Warszawa, 2001.
- [2] KIENITZ H., *Massenspektrometrie*, Verlag Chemie GmbH, Weinheim, 1968.
- [3] BOWE P., GIAMMANCO F., O'NEILL R.W., M. VAN DER BURGT P.J., SLEVIN J.A., *Observation of collective effects in the multiphoton ionization of atomic deuterium*, Phys. Rev. A, 1998, 58, 1389–1398.

- [4] HOFFMANN E., CHARETTE J., STROOBANT V., *Spektrometria mas*, PWN, Warszawa, 1998.
- [5] COTTER R.J., *Time-of-flight mass spectrometry*, American Chemical Society, Washington, 1994.
- [6] ZIELIŃSKI M., KARASEK K., PŁÓCIENNIK P., DYGDAŁA R.S., *Digital acquisition of the signals in atomic physics experiments*, 28 th EGAS, 1996 Graz, C4-82, pp. 382–383.
- [7] ZIELIŃSKI M., KARASEK K., DYGDAŁA R.S., *Fast, real-time multichannel scaler, construction and applications*, Rev. Sci. Instr., 67, 1996, pp. 3325–3331.
- [8] DYGDAŁA R.S., ZAWADZKA A., PŁÓCIENNIK P., ZIELIŃSKI M., *Three-Photon Ionization of Ca Through the Autoionizing States*, EGAS 1997 Berlin, pp. 587–588.
- [9] ZIELIŃSKI M., DYGDAŁA R.S., PŁÓCIENNIK P., *Spektrometria masowa czasu przelotu z wykorzystaniem wielokanałowego systemu zliczającego*, Metr. i Sys. Pom., t. 5, nr 1–2, 1998, pp. 117–125.
- [10] DYGDAŁA R.S., KARASEK K., GIAMMANCO F., KOBUS J., PABJANEK-ZAWADZKA A., RACZYŃSKI A., ZAREMBA J., ZIELIŃSKI M., *Three-photon Ionization of Ca*, J. Phys. B: At. Mol. Opt. Phys., 1998, 32, 2259–2278.
- [11] KARASEK K., *Badanie efektów fotojonizacji w atomach wapnia*, Praca doktorska, UMK, Toruń, 1999.
- [12] ZAWADZKA A., *Jonizacja wielofotonowa z dwuelektronowym wzbudzeniem w atomach wapnia*, Praca doktorska, UMK, Toruń, 2001.
- [13] ZIELIŃSKI M., *Wybrane aspekty pomiaru funkcji intensywności procesów niestacjonarnych w zakresie nano- i pikosekundowych rozdzielczości systemu pomiarowego*, praca habilitacyjna UMK/WAT, 2000.
- [14] SNYDER D.L., *Random point processes*, John Wiley & Sons, New York, 1975.
- [15] GHIONI M., COVA S., SAMORI C., ZAPPA F., *True constant fraction trigger circuit for picosecond photon-timing with ultrafast microchannel plate photomultipliers*, Rev. Sci. Instr., 1997, Vol. 68, No. 5, 2228–2236.
- [16] SPINELLI A., DAVIS L.M., *Actively quenched single-photon avalanche diode for high repetition rate one-gated photon counting*, Rev. Sci. Instr., 1996, Vol. 67, No. 1, 55–60.
- [17] WURZ P., GUBLER L., *Fast microchannel plate detector for particles*, Rev. Sci. Instr., 1996, Vol. 67, No. 5, pp. 1790–1793.
- [18] KEMNITZ K., PFEIFER L., AINBUND M.R., *Detector for multichannel spectroscopy and fluorescence lifetime imaging on the picosecond timescale*, Nucl. Instr. Meth. Phys. Res., 1997, A 387, 86–87.
- [19] BOKELMANN V., SPENGLER B., KAUFMANN R., *Dynamical Parameters of Ion Ejection and Ion Formation in Matrix-Assisted Laser Desorption/Ionization*, Eur. Mass Spectrom., 1, 1995, 81–93.
- [20] WILEY W.C., MCLAREN I.H., *Time-of-flight mass spectrometer with improved resolution*, Rev. Sci. Instr., 1955, Vol. 26, No. 12, 1150–1157.
- [21] STACEWICZ T., KRAIŃSKA-MISZCZAK M., *Time-resolved photon counting with digital oscilloscope*, Meas. Sci. Technol., 1997, No. 8, pp. 453–455.
- [22] VITEBINI M., ADRIANI A., DI DONFRANCESCO G., *Single photon detection and timing for a Lidar experiment*, Rev. Sci. Instr., 1987, Vol. 58, No. 10, pp. 1833–1839.
- [23] SCHWEMMER G., DOMBROWSKI M., KORB C.L., MILROD J., WALDEN H., KAGANN R.H., *A lidar system for measuring atmospheric pressure and temperature profiles*, Rev. Sci. Instr., 1987, Vol. 58, No. 12, pp. 2226–2237.
- [24] KELLY L.A., TRUNK J.G., SUTHERLAND J.C., *Time-resolved fluorescence polarization measurements for entire emission spectra with a resistive-anode, single-photon-counting detector: The fluorescence omnilyzer*, Rev. Sci. Instr., 1997, 68, (6), pp. 2279–2286.
- [25] IMASAKA T., OGAWA T., ISHIBASHI N., *Nanosecond time-resolved spectrometry with a tunable dye laser and a simple pulse-gated photon counter*, Analyt. Chem., 1979, Vol. 51, No. 4, pp. 502–504.

- [26] PALOJARVI P., MAATTA K., KOSTAMOVAARA J., *Integrated time-of-flight laser radar*, IEEE Trans. Instr. Meas., 1997, Vol. 46, No. 4, pp. 996–999.
- [27] CHANG B., HOETZLEIN R.C., MUELLER J.A., GEISER J.D., HOUSTON P.L., *Improved two dimensional product imaging: The real-time ion-counting method*, Rev. Sci. Instr., 1998, Vol. 69, No. 4, pp. 1665–1670.
- [28] BINKOWSKI M., SZYMAŃSKI M., BAKSALARY J., *Double-beam spectrometer of nanosecond resolution*, Opt. Appl., 1996, Vol. 26, No. 3, pp. 159–169.
- [29] ZIELIŃSKI M., *Systems of analysis of the random, point non-stationary processes with Poisson statistics*, KKM'98, Gdańsk, 1998, T. 3, pp. 37–45.
- [30] JENKINS A., GOULD R.W., GEDCKE D., *Quantitative X-Ray Spektrometry*, Marcel Dekker, New York, 1995.
- [31] WESTPHAL G.P., *Quantitative gamma spectroscopy at very high counting rates*, Nucl. Instr. and Meth., 1980, No. 178, pp. 609–611.
- [32] DEMTRODER W., *Spektroskopia laserowa*, PWN, Warszawa, 1993.
- [33] DEMAS J.N., *Excited state lifetime measurements*, Academic Press, New York, 1983.
- [34] IMHOF R.E., READ F.H., *Measurements of lifetimes of atoms, molecules and ions*, Rep. Prog. Phys., 1977, Vol. 40, pp. 1–104.
- [35] DYGDALA R.S., ZAWADZKA A., GIAMMANCO F., ZIELIŃSKI M., *Multi-photon ionization of calcium*, Int. Conf. on LASER 97, New Orleans, Soc. of Opt. & Quant. El. USA, STS Press, McLean, 1998, VA, pp. 311–319.
- [36] DYGDALA R.S., GIAMMANCO F., KARASEK K., PABJANEK A., ZIELIŃSKI M., *Three-Photon Ionization of Ca through $4s(2S) 6s 1 S0$ state*, 28 th EGAS, Graz, 1996, pp. 347–348.
- [37] ANFINRUD P.A., STRUVE W.S., *Minimization of discrimination jitter in time-correlated single photon counting*, Rev. Sci. Instr., 1989, 60, (4), p. 800.
- [38] ZIELIŃSKI M., KARASEK K., PŁÓCIENNIK P., DYGDALA R.S., *Digital gated single particle counting systems, design and applications*, Metrologia i Systemy Pomiarowe, T. III 3–4/ 1996, pp. 199–211.
- [39] DYGDALA R.S., FUSO F., ARIMONDO E., ZIELIŃSKI M., *Applications of a box-car in pulsed laser spectroscopy*, 26th EGAS Conf, Barcelona, 1994, pp. 305–306.
- [40] YANG P., MYERS D.P., LI G., HIEFTJE G.M., *Constant-fraction discrimination/boxcar integrator for plasma source time-of-flight mass spectrometry*, Appl. Spectrosc., 1995, Vol. 49, No. 5, pp. 660–664.
- [41] COLLIER J.L., GODDARD B.J., GOODE D.C., MARKA S., TELLE H.H., *A low-cost gated integrator boxcar averager*, Meas. Sci. Technol. (UK), 1996, Vol. 7, No. 9, pp. 1204–1211.
- [42] DYGDALA R.S., FUSO F., ARIMONDO E., ZIELIŃSKI M., *Modular digital box-car for applications in pulsed laser spectroscopy*, Rev. Sci. Instr., 1995, 66, pp. 3507–3512.
- [43] WINIECKI W., *Projektowanie przyrządów wirtualnych z wykorzystaniem zintegrowanych środowisk programowych*, MWK'97, Zegrze, 1997, pp. 91–136.
- [44] DES-JARDIN L., *Virtual instruments and the role of software*, Mc Graw-Hill Inc, 1993.
- [45] LIN S.F.R., ABBAS A., WIKSWO J.P., *High-resolution high-speed synchronous epifluorescence imaging of cardiac activation*, Rev. Sci. Instrum., 1997, Vol. 68, No. 1, pp. 213–217.
- [46] ŁUKASIAK Z., WYRZYKOWSKI Z., SYLWISTY J., BAŁA W., *Time-resolved study of luminescence properties of porous silicon in micro and nanosecond range*, Proc of SPIE, 1999, Vol. 3725, pp. 102–106.
- [47] SASAKI O., TANIGUCHI T., OHSKA T.K. et al, *1,2 GHz shift register IC for dead-time-less TDC application*, IEEE Trans. Nucl. Sci., 1989, Vol. 36, No. 1, pp. 512–516.
- [48] LOINAZ M.J., WOOLEY B.A., *A CMOS multichannel IC for pulse timing measurements with 1-mV sensitivity*, IEEE Jour. of Solid-state Circuits, 1995, Vol. 30, No. 12, pp. 1339–1349.

- [49] ARAI Y., IKENO M., *A time digitizer CMOS gate-array with a 250 ps time resolution*, IEEE Journal of Solid State Circuits, 1996, Vol. 31, No. 2, pp. 212–220.
- [50] ARAI Y., MATSUMURA T., ENDO K., *A CMOS 4 ch x 1 k time memory LSI with 1 ns/b resolution*, IEEE Solid-State Circuits, 1992, Vol. 27, pp. 359–364.
- [51] LJUSLIN C. et al, *An integrated 16-channel CMOS time to digital converter*, IEEE Trans. Nucl. Sci., 1994, Vol. 41, pp. 1104–1108.
- [52] WATANABE T. et al. *A CMOS time-to-digital converter LSI with half-nanosecond resolution using a ring gate delay line*, IEICE Trans. Electron., 1993, Vol. E76-C, pp. 1774–1779.
- [53] ZIELIŃSKI M., PŁÓCIENNIK P., KARASEK K., ZAWADZKA A., *System monitorowania dla spektrometru masowego czasu przelotu (time-of-flight)*, Konf. Automation 97, Warszawa, 1997, t. 2, pp. 428–435.
- [54] ZIELIŃSKI M., PŁÓCIENNIK P., ZAWADZKA A., DYGDALA R.S., *Wielokanałowy system zliczania fotonów i jonów, pracujący w czasie rzeczywistym*, Konf., MWK'97, Zegrze, 1997, t. 3, pp. 125–132.
- [55] ZIELIŃSKI M., DYGDALA R.S., KARASEK K., *Sposób zliczania impulsów w wielokanałowych systemach zliczających czasu rzeczywistego oraz układ do zliczania impulsów w wielokanałowych systemach zliczających czasu rzeczywistego*, Patent PL 179439 B1, 2000.
- [56] ZIELIŃSKI M., DYGDALA R.S., *Systemy pomiaru funkcji intensywności niestacjonarnych procesów o rozkładzie Poissona w czasie rzeczywistym*, XXXI MKM, 1999, Białystok, pp. 101–108.
- [57] ZIELIŃSKI M., PŁÓCIENNIK P., DYGDALA R.S., *Counting errors in a real-time multichannel scaler*, Rev. Sci. Instr., 1997, 69, (1), pp. 299–305.
- [58] ZIELIŃSKI M., *Correction of the counting losses of a real-time multichannel-scaler for a very high intensity of particle flux*, Measurement, 2000, Vol. 28, No. 4, pp. 225–233.
- [59] ZIELIŃSKI M., DYGDALA R.S., PŁÓCIENNIK P., KARASEK K., *Counting Errors Protection in the Real-time Multichannel Scaler*, Metrologia i Sys. Pom., 1997, No. 3–4, pp. 223–234.
- [60] RETTIG J. B., DOBOS L., *Picosecond time interval measurements*, IEEE Trans. Instr. Meas., 1995, Vol. 44, No. 2, pp. 284–287.
- [61] ANDRIA G., ATTIVISSIMO F., LANZOLLA A., *Interpolation techniques for non-stationary signal analysis*, Measurement, 1997, Vol. 21, No. 1–2, pp. 1–16.
- [62] PŁÓCIENNIK P., ZIELIŃSKI M., DYGDALA R.S., KARASEK K., *Wielokanałowy interpolator odcinka czasowego*, MWK'99, Zegrze, 1999, t. 3, pp. 81–86.
- [63] LAMPTON M., RAFANTI R., *A high speed wide dynamic range time-to-digital converter*, Rev. Sci. Instr., 1994, Vol. 65, No. 11, pp. 3577–3583.
- [64] MAATTA K.E., KASTAMOVAARA J.K., RISTO R.A., MYLLYLA A., *Time-to-digital converter for fast, accurate laser rangefinding*, SPIE Vol. 1010, Industrial Insp. pp. (60–67), 1988.
- [65] KALISZ J., SZPLET R., PASIERBIŃSKI J., PONIECKI A., *Field-programmable-gate-array-based time-to-digital converter with 200-ps resolution*, IEEE Trans. on Instr., 1997, Vol. 46, pp. 71–75.
- [66] KALISZ J., SZPLET R., *Time-to-digital converter with direct coding and 100 ps resolution*, Electronics letters, 1995, Vol. 31, No. 19, pp. 1658–1659.
- [67] CHU D.C., ALLEN M.S., FOSTER A.S., *Universal counter resolves picoseconds in time interval measurements*, Hewlett-Packard Journal, 1978, August, pp. 2–7.
- [68] KALISZ J., SZPLET R., PELKA R., PONIECKI A., *Single-chip interpolating time counter with 200 ps resolution and 43 s range*, IEEE Trans. Instr. & Meas., 1997, Vol. 46, No. 4, pp. 851–856.
- [69] SIMPSON M.L., BRITTON C.L., WINTENBERG A.L., YOUNG G.R., *An integrated CMOS time interval measurement system with subnanosecond resolution for the WA-98 calorimeter*, IEEE J. Solid-State Circ., 1997, Vol. 32, No. 2, pp. 198–205.
- [70] SUMNER R., BLANAR G., *New techniques for measuring time intervals with very high resolution*, LeCroy Research Systems, New York, 1996.

- [71] TURKO B.T., *Multichannel interval timer*, IEEE Trans. Nuc. Sci., 1984, Vol. NS-31, No. 1, pp. 167–171.
- [72] KINDLMANN P.J., SUNDERLAND J., *Phase stabilized vernier chronotron*, Rev. Sci. Instr., 1966, Vol. 37, No. 4, pp. 445–452.
- [73] RAHKONEN T., KOSTAMOVAARA J., *The use of stabilized CMOS delay lines in the digitization of short time intervals*, Proc. IEEE Symp. Circuits Syst., Singapore, 1991, Vol. 4, pp. 2252–2255.
- [74] MOYER G.C., CLEMENTS M. et al., *The delay vernier pattern generation technique*, IEEE Jour. of Solid-state Circ., 1997, Vol. 32, No. 4, pp. 551–561.
- [75] NUTT R., *Digital time intervals meter*, Rev. Sci. Instr., 1968, Vol. 39, pp. 1342–1345.
- [76] KALISZ J., PAWŁOWSKI M., PEŁKA R., *Error analysis and design of the Nutt time-interval digitizer with picosecond resolution*, J. Phys. E. Sci. Instr. 1987, Vol. 20, pp. 1330–1341.
- [77] TŁACZAŁA W., BASIŁADZE S.G., SMIRNOW W.A., *Przetwornik czas-kod cyfrowy*, Nukleonika, 1974, t. XIX, No. 9, s. 777–783.
- [78] BASIŁADZE S.G., TŁACZAŁA W., *Metoda przetwarzania przedziałów czasowych w kod cyfrowy w szerokim zakresie pomiarowym*, Nukleonika, 1974, t. XIX, No. 6, pp. 551–555.
- [79] CHU D.C., *The triggered phase-locked oscillator*, Hewlett-Packard Journal, August 1978, pp. 8–9.
- [80] PŁÓCIENNIK P., ZIELIŃSKI M., DYGDALA R.S., *Multichannel interpolator for measurement of the time resolved distribution*, 31 st EGAS, Marseille, 1999, pp. 316–317.
- [81] ZIELIŃSKI M., ZDUNIAK J., *Zastosowanie linii opóźniających zbudowanych na strukturach PLD do wielokanalowego systemu o dużej rozdzielczości*, MWK'99, Zegrze, 1999, t. 3, pp. 149–154.
- [82] CITTOLIN S., *CMOS data acquisition baseline design*, Nucl. Instr. Meth. A, 1994, Vol. 344, pp. 161–165.
- [83] SU D., LOINAZ M., MASUI C., WOOLEY B., *Experimental results and modeling techniques for substrate noise in mixed-signal integrated circuits*, IEEE Solid-State Circuits, 1993, Vol. 28, pp. 420–430.
- [84] LEO W., *Techniques for nuclear and particle physics experiments*, New York, Springer-Verlag, 1994.
- [85] LOINAZ M., WOOLEY B., *A BiCMOS time interval digitizer based on fully-differential, current-steering circuits*, IEEE Solid-State Circuits, 1994, Vol. 29, pp. 707–714.
- [86] *MECL integrated circuits*, Motorola Inc, 1989.
- [87] ŁUBA T., JASIŃSKI K., ZBIERZCHOWSKI B., *Specjalizowane układy cyfrowe w strukturach PLD i FPGA*, WKŁ, Warszawa 1997.
- [88] PEŁKA R., KALISZ J., SZPLET R., *Nonlinearity correction of the integrated time-to-digital converter with direct coding*, IEEE Trans. on Instr. and Meas., 1997, Vol. 46, No. 2, pp. 449–453.
- [89] PEŁKA R., *Linearizing transfer characteristics of precision time-interval digitizer with neural network*, Biuletyn WAT, 1995, t. XLIV, No. 10 (518), pp. 53–72.
- [90] TADEUSIEWICZ R., *Sieci neuronowe*, AOWRM, Warszawa 1993.
- [91] ZIELIŃSKI M., CHABERSKI D., GRZELAK S., FRANKOWSKI R., KOWALSKI M., *High resolution real-time multichannel scaler implemented in single FPGA device*, XVII IMEKO World Congress, Dubrovnik, Croatia, June 22–27, 2003, pp. 708–711.
- [92] ZIELIŃSKI M., CHABERSKI D., KOWALSKI M., FRANKOWSKI R., GRZELAK S., *High resolution time interval measuring system implemented in single FPGA device*, Measurement, 2004, Vol. 35, No. 3, pp. 311–317.
- [93] CHABERSKI D., GRZELAK S., *Remote measurements in time of flight mass spectroscopy*, Joint International IMEKO TC1+TC7 Symposium, Germany, Ilmenau, 2005, September 21–24, pp. 150–152.
- [94] GRZELAK S., CHABERSKI D., FRANKOWSKI R., ZIELIŃSKI M., KOWALSKI M., *System pomiaru odcinka czasowego implementowany w strukturze FPGA*, XXXIV Międzyuczelniana Konferencja Metrologów, Wrocław, 2002, 8–12 wrzesień, t. III, pp. 51–56.

- [95] CHABERSKI D., *System pomiaru funkcji intensywności strumienia jonów w spektrometrii masyowej czasu przelotu*, praca doktorska, Wojskowa Akademia Techniczna, 2006.
- [96] CHABERSKI D., FRANKOWSKI R., KOWALSKI M., ZIELIŃSKI M., GRZELAK S., *Wielosegmentowe linie opóźniające i ich zastosowanie w systemach pomiarowych*, XXXIV Międzyuczelniana Konferencja Metrologów, Wrocław, 2002, 8–12 września, t. III, pp. 7–14.
- [97] FRANKOWSKI R., GRZELAK S., KOWALSKI M., CHABERSKI D., ZIELIŃSKI M., *Time-interval measurements, using tapped delay lines implemented in programmable logical structures*, Proceedings of the 12th IMEKO TC4 International Symposium, Part 2, Electrical Measurements and Instrumentation, Zagreb, Croatia, 2002, September 25–27, pp. 333–336.
- [98] GRZELAK S., CHABERSKI D., ZIELIŃSKI M., *System pomiaru odcinka czasowego z wykorzystaniem metody noniusza implementowany w strukturze FPGA*, Metrologia Wspomagana Komputorowo, Waplewo, 2003, 26–29 maja, pp. 67–72.
- [99] ZIELIŃSKI M., CHABERSKI D., GRZELAK S., *Time-interval measuring modules with short deadtime*, Metrology and Measurement Systems, 2003, Vol. 10, No. 3, pp. 241–251.
- [100] ZIELIŃSKI M., GRZELAK S., CHABERSKI D., FRANKOWSKI R., KOWALSKI M., *High resolution time-interval measuring module with Vernier scale*, Elektronika, 2005, No. 2–3, pp. 68–70.
- [101] GRZELAK S., *Wysokorozdzielcze układy pomiaru odcinka czasowego dla zastosowań w spektrometrii fotonów i jonów*, praca doktorska, Wojskowa Akademia Techniczna, 2006.
- [102] *Virtex - 2.5 V Field Programmable Gate Arrays Product Specification*, 2001 April, Technical Documentation.
- [103] KALISZ J., *Język VHDL w praktyce*, Wydawnictwa Komunikacji i Łączności, Warszawa, 2002.
- [104] SKAHILL K., *VHDL for programmable logic*, Addison-Wesley Publishing Company, 1996.
- [105] AUDOIN C., DIMARCO N., *Stochastic models of stable frequency and time sources and their relationship*, IEEE Trans. Instr. Meas., 1993, Vol. 42, No. 3, pp. 682–688.
- [106] ALLAN D.W., *Statistics of atomic frequency standards*, Proc. IEEE, 1996, Vol. 54, No. 2, pp. 221–230.
- [107] BARNES J.A., ALLAN D.W., *An approach to the prediction of coordinated universal time*, Frequency, 1967, Vol. 5, Nov., pp. 15–20.
- [108] LESAGE P., AYI T., *Characterization of frequency stability: analysis of the modified Allan variance and properties of its estimate*, IEEE Trans. Instr. Meas., 1984, Vol. IM33, No. 4, pp. 332–336.
- [109] COSART L.D., PEREGRINO L., TAMBE A., *Time domain analysis and its practical application to the measurement of phase noise and jitter*, IEEE Trans. Instr. Meas., 1997, Vol. 46, No. 4, pp. 1016–1019.
- [110] VERNOTTE F., GROSLAMBERT J., GAGNEPAIN J.J., *A new method of measurement of the different types of noise altering the output signal of oscillators*, IEEE Trans. Instr. Meas., 1993, Vol. 42, No. 6, pp. 968–975.
- [111] CASTELLINI G., VIGNOLI V., GABBANI V., *125 MHz ECL digital oscillator synchronizable both in frequency and in phase*, IEEE Trans. Instr. Meas., 1996, Vol. 45, No. 1, pp. 318–320.
- [112] WALTER T., *Characterizing frequency stability: a continuous power-law model with discrete sampling*, IEEE Trans. Instr. Meas., 1994, Vol. 43, No. 1, pp. 69–77.
- [113] CRANINCK X., STEYAERT M.J., *A 1,8-GHz low-phase-noise voltage-controlled oscillator with prescaler*, IEEE J. Solid-State Circ., 1995, Vol. 30, No. 12, pp. 1474–1482.
- [114] NOVOTNY I.I., AUSTIN J., KELKAR R., STRAYER D., WYATT S., *Fully integrated CMOS phase-locked loop with 15 to 240 MHz locking range and ≤ 50 ps jitter*, IEEE J. Solid-State Circ., 1995, Vol. 30, No. 11, pp. 1259–1266.
- [115] MCNEILL J.A., *Jitter in ring oscillators*, IEEE J. Solid-State Circ., 1995, Vol. 32, No. 6, pp. 870–879.

- [116] BERKEL W., KOFFMANE G., EATOCK F.L., SCHMID P., HOPKE H., SNACKERS H.J., *500 MHz and 300 MHz programmable pulse generators*, Hewlett-Packard J., 1990, pp. 64–75.
- [117] GIOVANINI T., *Simple approach optimizes ECL clock networks*, *Microwaves & RF.*, 1992, pp. 87–93.
- [118] COSART L.D., PEREGRINO L., TAMBE A., *Time domain analysis and its practical application to the measurement of phase noise and jitter*, IEEE Instr. and Meas. Techn. Conf., Brussels, 1996.
- [119] HASSE L., SPIRALSKI L., *Szumy elementów i układów elektronicznych*, WNT, Warszawa, 1981.
- [120] OSIOWSKI J., SZABATIN J., *Podstawy teorii obwodów*, WNT, Warszawa, 1995, t. 3.
- [121] JENKINS A., GOULD R.W., GEDCKE D., *Quantitative X-Ray Spectrometry*, Marcel Dekker, New York, 1995.
- [122] APPLICATION NOTE 50, *Instrumentation for fluorescence lifetime spectrometry*, EG&G ORTEC, www.ortec.com.
- [123] ZIELIŃSKI M., *Digital gated single-particle counting system, the errors analysis*, *Instrumentation and Measurement*, IEEE Transactions on Volume 49, Issue 5, Oct. 2000, pp. 1069–1076.
- [124] ZIELIŃSKI M., DYGDAŁA R.S., PŁÓCIENNIK P., *The counting errors in real-time multichannel scaler*, 29 th EGAS, Berlin, 1997, pp. 440–441.
- [125] ZIELONKO R., *O optymalizacji algorytmu pomiaru czasu i częstotliwości w obecności zakłóceń o rozkładzie Poissona*, *Zeszyty Naukowe Politechniki Gdańskiej, Elektronika*, XXV, 1970, No. 153, pp. 69–80.
- [126] *AN51, Pulse-processing electronics for single-photon counting*, EG&G ORTEC.
- [127] GICHMAN I.I., SKOROCHOD A.W., *Wstęp do teorii procesów stochastycznych*, PWN, Warszawa, 1968.
- [128] PAPOULIS A., *Prawdopodobieństwo, zmienne losowe i procesy stochastyczne*, WNT, Warszawa, 1972.
- [129] KINGMAN J.F.C., *Poisson Processes*, Oxford University Press, New York, 1993.
- [130] BARTLETT M.S., *An introduction to stochastic processes*, Cambridge, 1978.
- [131] JAWORSKI J., *Matematyczne podstawy metrologii*, WNT, Warszawa, 1979.
- [132] SOBCZYK K., *Stochastyczne równania różniczkowe*, WNT, Warszawa, 1996.
- [133] SOBCZAK W., *Metody statystyczne w elektronice*, WNT, Warszawa, 1971.
- [134] SNYDER D.L., *Random point processes*, John Wiley & Sons, New York, 1975.
- [135] JENKINS A., GOULD R.W., GEDCKE D., *Quantitative X-Ray Spectrometry*, Marcel Dekker, New York, 1995.
- [136] *Rubidium frequency standards*, Tekelec - Temex, <http://www.tekelec-temex.com>, 1998.
- [137] BARNES J.A., CHI A.R., CUTLER L.S. et al., *Characterization of frequency stability*, IEEE Trans. Instr. Meas., 1971, Vol. IM-20, No. 2, pp. 105–20.
- [138] RUTMAN J., *Characterization of phase and frequency instabilities in precision frequency sources: fifteen years of progress*, Proc. IEEE, 1978, Vol. 66, p. 1048.
- [139] RUTMAN J., WALLS F.L., *Characterization of frequency stability in precision frequency sources*, Proc. IEEE, 1991, Vol. 79, p. 952.
- [140] BREGNI S., *Clock stability characterization and measurement in telecommunications*, IEEE Trans. Instr. Meas., 1997, Vol. 46, No. 6, pp. 1284–1294.
- [141] KROUPA V., *Frequency stability: Fundamentals and measurements*, IEEE Press, New York, 1983.
- [142] ZIELIŃSKI M., KOWALSKI M., CHABERSKI D., GRZELAK S., FRANKOWSKI R., *System do pomiaru rozmycia czasowego generatorów i jego zastosowania*, MWK2008, Waplewo, 2008, S4, p. 105.
- [143] DYGDAŁA R.S., FUSO F., ARIMONDO E., ZIELIŃSKI M., *Applications of a box-car in pulsed laser spectroscopy*, 26th EGAS Conf, Barcelona, 1994, pp. 305–306.

- [144] ZIELIŃSKI M., PŁÓCIENNIK P., KARASEK K., *Budowa i zastosowanie w systemach pomiarowych synchronizowanego generatora wzorcowego*, KKM'98, Gdańsk, 1998, t. 3, pp. 143–150.
- [145] ZIELIŃSKI M., KARASEK K., ZAWADZKA A., DYGDALA R.S., *Ocena stabilności generatora przy pomocy wielokanałowego systemu zliczającego*, MWK'99, Zegrze, 1999, t. 3, pp. 143–148.
- [146] ZIELIŃSKI M., *Sprawozdanie z realizacji projektu badawczego: Projektowanie i budowa systemów pomiarowych do badania zależności czasowych, pracujących w czasie rzeczywistym*, KBN, 1999.
- [147] PEŁKA R., KALISZ J., SZPLET R., *Nonlinearity correction of the integrated time-to digital converter with direct coding*, IEEE Trans. Instrum. Meas., 1997, Vol. 46, No. 2, pp. 449–452.
- [148] KALISZ J., *Review of methods for time interval measurements with picosecond resolution*, Metrologia, 2004, Vol. 41, No. 1, pp. 17–32.
- [149] BARONTI F., FANUCCI L., LUNARDINI D., RONCELLA R., SALETTI R., *On the differential nonlinearity of time-to-digital converters based on delay-locked-loop delay lines*, IEEE Transactions on Nuclear Science, 2001, Vol. 48, No. 6, pp. 2424–2431.
- [150] ZIELIŃSKI M., CHABERSKI D., KOWALSKI M. GRZELAK S., FRANKOWSKI R., PŁÓCIENNIK P., *Time of flight mass spectrometer monitoring system implemented in single FPGA device*, International Multi Conference on Systems, Signal & Devices, Summary Proceedings, 2007 March 19–22, Tunisia, p. 333.
- [151] BIELSKI A., CIURYŁO R., *Podstawy metod opracowania pomiarów*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń, 2001.
- [152] TAYLOR J.R., *Wstęp do analizy błęd pomiarowego*, PWN, Warszawa, 1995.
- [153] JAWORSKI J., *Matematyczne podstawy metrologii*, Wydawnictwa Naukowo-Techniczne, Warszawa, 1979.
- [154] DYGDALA R.S., KARASEK K., GIAMMANCO F., KOBUS J., PABJANEK-ZAWADZKA A., RACZYŃSKI A., ZAREMBA J., ZIELIŃSKI M., *Three-photon ionization of Ca*, J. Phys. B: At. Mol. Opt. Phys., 1998, Vol. 31, pp. 2259–2278.
- [155] DYGDALA R.S., ZAWADZKA A., LISAK D., PŁÓCIENNIK P., TRAWINSKI R.S., *Investigation of highly excited states of calcium by three-photon ionization*, Eur. Phys. J. D, 2004, Vol. 30, pp. 15–22.
- [156] DYGDALA R.S., KARASEK K., STEFANSKI K., ZAWADZKA A., RUMIANOWSKI R., ZIELIŃSKI M., *The plasma phenomena in three-photon ionization of Ca*, J. Phys. D: Appl. Phys., 2000, Vol. 33, No. 1, pp. 41–53.
- [157] ZIELIŃSKI M., DYGDALA R.S., PŁÓCIENNIK P., *Counting errors in a real-time multichannel scaler*, Rev. Sci. Instrum., 1998, Vol. 69, No. 1, pp. 299–305.
- [158] GIAMMANCO F., *Collective effects in laser multiphoton ionization, I Experimental results*, Phys. Rev. A, 1989, Vol. 40, pp. 5160–5170.
- [159] KHAN M.A., KHAWAJA E.E., AL-JUWAIR H.A., J. Phys. B: At. Mol. Phys., 1990, Vol. 23, pp. 533–539.
- [160] GIAMMANCO F., SPINELLI N., *Plasma collective effects in atomic physics*, 1996, pp. 108–126.

MEASUREMENTS OF THE ATOMIC BEAM COMPOSITION USING TIME-OF-FLIGHT SPECTROMETER

In this paper the problem of isotopic composition measurement of the atomic beam by time-of-flight spectrometer is discussed. The first part of the paper contains introduction, which explains the mass spectrometry and describes the measurement system for an investigation of the atomic beam isotopic composition. In the second part, the measurement units are presented and their precision is discussed, the sources of errors are shown and the methods of their limit and reduction are presented. The experimental results are shown and discussed at the end.

*wiry optyczne, interferometria,
metody lokalizacji wirów optycznych,
metody pomiarowe określenia znaku wirów optycznych*

Ewa FRĄCZEK,
Wojciech FRĄCZEK,
Janusz MROCZKA*

WIRY OPTYCZNE JAKO ZNACZNIKI FAZY W POLU OPTYCZNYM

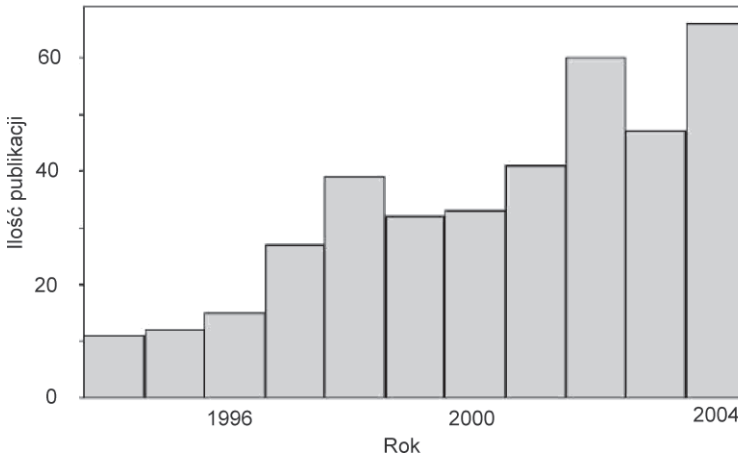
Niniejsza praca poświęcona jest interferometrii na wirach optycznych; zawiera opis rodzaju nieciągłości w polu optycznym, znanym jako wir optyczny. W publikacji tej są zaprezentowane metody generowania sieci wirów optycznych, ze szczególną uwagą zwróconą na metody interferencyjne. Przetawione fizyczne podstawy na temat wirów optycznych stanowią podstawę do technicznych realizacji systemów pomiarowych. Prezentowane sposoby lokalizacji centrów wirów optycznych opisane zostały szczegółowo ze względu na znaczenie lokalizacji punktów wirowych w zastosowaniu wirów optycznych jako znaczników fazy w polu optycznym. Uzyskane dokładności lokalizacji zostały przeanalizowane w połączeniu z przewidywanymi zakłóceniami. Wiedza na temat lokalizacji punktów wirowych może być użyteczna do obliczenia właściwości pola optycznego lub mierzonego obiektu. W niektórych przypadkach może być konieczna znajomość znaków ładunków topologicznych zlokalizowanych wirów optycznych. W pracy tej przedstawione są dwie metody określenia znaku wiru optycznego.

1. WPROWADZENIE

Zainteresowanie wirami optycznymi rozpoczęło się w drugiej połowie lat 70. XX w. Od prac opublikowanych przez Berryego i Nye'a o poświęconych rozpraszaniu mikrofal na antarktycznej pokrywie lodowej. Okazało się, że w wiązkach rozproszonych propagują się liczne defekty fazowe, w tym punktowe nieciągłości nazywane wówczas osobliwościami śrubowymi. Zapoczątkowana przez tych autorów teoria stosuje się do całego zakresu widma fal elektromagnetycznych, w tym również fal świetlnych. W latach 80. i 90. ubiegłego wieku pojawiło się wiele prac poświęconych teorii i zastosowaniu wirów optycznych. Obecnie mamy już z tej tematyki wydawnictwa książ-

* Katedra Metrologii Elektronicznej i Fotonicznej, Politechnika Wrocławska.

kowe [5], [63], [75]. Pod koniec lat 90. XX w. zaczęto mówić o nowej gałęzi współczesnej optyki, której nadano nazwę *singular optics*. Z każdym rokiem liczba publikacji dotyczących wirów optycznych wzrasta. Dla przykładu zamieszczony jest wykres sporządzony przez A.S. Desyatnikova, L. Tornera i Yu. S. Kivshara, przedstawiający ilość publikacji z dziedziny „nonlinear singular optics”, które ukazały się w czasopiśmie recenzowanych od 1994 roku [22].



Rys. 1. Liczba publikacji z dziedziny „nonlinear singular optics”, które ukazały się w czasopiśmie recenzowanych w kolejnych latach [22]

Fig. 1. The number of papers in the field of *nonlinear singular optics* cited in this review vs. the publication year [22]

Warto wspomnieć, że wiry optyczne mają również swoją prehistorię. Prapoczątków optyki nieciągłości można się doszukiwać w opublikowanej w 1830 roku teorii tęczy i odwzorowania punktu przez osiowo-symetryczny układ optyczny (plamka Aire’go) [10]. Wiry optyczne wpisują się w modny nurt badań nad nieciągłościami w szerokiej klasie pól skalarnych i wektorowych obecnych w różnych gałęziach fizyki (np. fizyka ciała stałego [48], teoria fazy nadciekłej [2], mechanika kwantowa [16]). Najbardziej zaawansowane prace z dziedziny wirów optycznych związane są z pułapkowaniem i manipulowaniem mikrocząstkami przez manipulatory optyczne (*optical tweezers*) [7], [35], [69]. Urządzenia te znalazły szerokie zastosowanie w mikrobiologii. Jednym z ciekawszych zastosowań wirów optycznych jest wykorzystanie punktowej nieciągłości fazy pola optycznego w teleskopie [46]. Teleskop ten ma w swoim wnętrzu maskę wytwarzającą wir optyczny, który wykorzystany jest do wygaszania światła obserwowanej gwiazdy. Osłabienie tak silnego źródła światła umożliwia obserwację planet. Prowadzone są również prace nad zastosowaniem wirów optycznych w technice światłowodowej [40], [70] oraz w kryptografii kwantowej [19]. Najbardziej reprezentatywną pracą przedstawiającą cechy fizyczne nieciągłości optycznych

jest praca J.F Nye'a *Natural focusing and fine structure of light* [63]. Obecnie wiry optyczne są obiektem zainteresowania wielu ośrodków naukowych na świecie, wśród nich znajduje się Katedra Metrologii Elektronicznej i Fotonicznej Politechniki Wrocławskiej, w której prowadzone są prace w zakresie interferometrii na wirach optycznych. Jest to stosunkowo niedawno powstała gałąź interferometrii, w której pożądana jest obecność nieciągłości fazowych w polu optycznym, a ich położenie jest wykorzystywane do pomiaru właściwości optycznych.

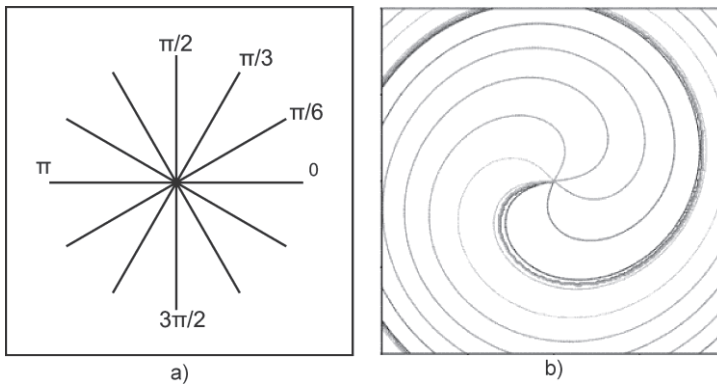
1.1. INTERFEROMETRIA NA WIRACH OPTYCZNYCH

Wir optyczny, jako punktowa nieciągłość w rozkładzie pola optycznego [31], może występować jako obiekt niepożądany lub może być sztucznie wprowadzony w pole interferencyjne. Istnieje wiele metod generacji wirów optycznych, a jedną z nich, umożliwiającą powstanie sieci wirów optycznych, jest metoda interferencyjna wykorzystywana w opisywanych eksperymentach [53]. Niniejsza praca poświęcona jest zastosowaniu wirów optycznych w *interferometrii na wirach optycznych* (IWO). Nieciągłości fazowe, włączając w to wiry optyczne, z powodu występowania nieokreślonej wartości fazy utrudniają analizę interferogramów. Dlatego wymagają bardziej złożonych metod analizy lub w standardowej analizie są pomijane. W technice IWO analizowane są właśnie nieciągłości fazowe i ich otoczenie. Wykorzystanie nieciągłości fazowych do analizy pola optycznego wymaga ich precyzyjnego zlokalizowania w obrazie interferencyjnym, a następnie analizy ich wzajemnego rozmieszczenia. Znajomość zmian geometrii sieci wirów umożliwia uzyskanie wartości wielkości mierzonych. Ponieważ wir optyczny jest punktem ciemnym, w którym poziom sygnału wynosi zero, pomiar położenia takiego punktu nie powinien odbywać się wprost, ale na podstawie rozkładu fazy pola w okolicy wiru optycznego. Taki pomiar wymaga opracowania metod identyfikacji charakterystycznych cech pola optycznego w sąsiedztwie wiru. Uzyskany metodą dedukcyjną model wiru optycznego nie może być bezpośrednio wykorzystany do lokalizacji wirów optycznych oraz do określenia ich znaków, ze względu na zespoloną amplitudę fali pola optycznego charakterystyczną dla wiru optycznego. Konieczne staje się połączenie istniejącego opisu z informacją uzyskaną drogą indukcyjną. Matematyczny i fizyczny opis wiru optycznego oraz jego propagacji jest dobrze znany i przedstawiany w literaturze [62], [66]. Istnieje analityczny opis wiru optycznego oraz wiązki, w której taka struktura jest propagowana. Wiele trudniejszy, ze względu na stopień skomplikowania, jest opis analityczny wiązek propagujących grupy wirów optycznych. Z tego samego powodu nie są wykorzystywane równania analityczne do obliczania lokalizacji punktów wirowych w polu optycznym. Do tego celu wykorzystuje się wyłącznie procedury numeryczne (punkt 2). Określenie położenia punktów wirowych w zależności od wartości wielkości oddziałującej możliwe jest po przeprowadzeniu eksperymentu w układzie rzeczywistym interferometru na wirach optycznych lub z wykorzystaniem jego modelu. Określenie wpływu wielkości oddziałującej na położenie punktów wirowych

wych w tym wypadku najlepiej jest określić w sposób indukcyjny. Rozwiązanie tego zagadnienia określanego w metrologii jako „problem wprost” [13] jest pierwszym etapem tej pracy. Następnie to zagadnienie jest odwracane i na podstawie analizy struktury sieci punktów wirowych wyznaczana jest wartość wielkości mierzonej (problem odwrotny). Połączenie modelu teoretycznego, obejmującego wir optyczny z eksperymentem, w którym wykorzystane są sieci wirów, umożliwi określenie cech podających lokalizację punktów wirowych w analizowanym obszarze pola optycznego oraz określenie znaków ich ładunków topologicznych. Lokalizacja węzłów sieci wirów optycznych powstałych z wykorzystaniem techniki IWO jest podstawą do dalszej części pracy nad wykorzystaniem nieciągłości fazowych w polu optycznym do celów pomiarowych. Określenie wartości wielkości mierzonej wymaga opracowania metod do analizy struktury sieci oraz metod interpretacji uzyskanych danych.

1.2. WŁAŚCIWOŚCI NIECIĄGŁOŚCI FAZY W POLU OPTYCZNYM

Pojedynczy wir optyczny jest związany z punktową izolowaną nieciągłością w rozkładzie fazy w polu świetlnym [31], [75]. Na rysunku 2 przedstawiono przekrój frontu falowego, który zawiera pojedynczy wir. Linie równej fazy zbiegają się w jednym punkcie, w którym wartość fazy jest nieokreślona (jest to izolowany punkt osobliwy w rozkładzie fazy pola optycznego), linie te mogą mieć wartości z przedziału $\langle 0, 2\pi \rangle$ lub $\langle -\pi, \pi \rangle$ (zależnie od przyjętej konwencji).



Rys. 2. Linie równej fazy wokół punktu wirowego:
a – dla fali opisanej wzorem (1); b – dla wiązki gaussowskiej z wirami

Fig. 2. Equiphase lines around the vortex point:

a – for wave described by formula (1); b – for gaussian beam with optical vortex

Najprostsze równanie spełniające równanie Laplace’a, które opisuje monochromatyczną, liniowo spolaryzowaną falę przenoszącą wir optyczny ma następującą postać [31]:

$$U_{\pm}^{(n)} = (a \pm ib)^n = A^{(n)} \exp(\pm in\theta) \quad (1)$$

gdzie:

- U – amplituda zespolona,
- A – amplitudę fali,
- n – liczba całkowita określa wartość topologicznego ładunku wiru optycznego,
- a, b – wartość rzeczywista i urojona funkcji,
- θ – wartość fazy.

Jeżeli wartość fazy wynosi $+n\theta$, to równanie (1) opisuje falę przenoszącą dodatni wir, jeżeli natomiast wartość fazy wynosi $-n\theta$, to mamy ujemny wir optyczny. Wyrażenie to jest rozbieżne przy odległości od centrum wiru optycznego dążącej do nieskończoności, powinno zatem być traktowane jako przybliżony opis rzeczywistych wiązek.

Topologiczny znak wiru optycznego (ładunek topologiczny [9], [36]) definiuje się przez wyrażenie [20], [71] z użyciem następującej zależności:

$$s = \frac{1}{2\pi} \oint_G d\theta = \frac{1}{2\pi} \oint_G \nabla\theta d\vec{R} \quad (2)$$

gdzie:

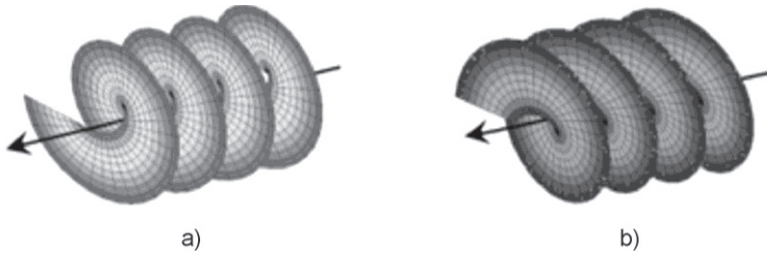
- G – nie przecinająca się krzywa zamknięta nie przechodząca przez nieciągłość fazową i nie zawierająca innej nieciągłości oprócz badanej,
- R – wektor położenia,
- θ – wartość fazy.

Wartością wyrażenia (2) są liczby całkowite, których znak określa znak ładunku topologicznego wiru, a wartość jest równa wartości tego ładunku.

Z warunku ciągłości, jaki musi spełniać amplituda zespolona U fali świetlnej wynika, że w punkcie, w którym jest wir optyczny amplituda (i natężenie) fali musi wynosić zero.

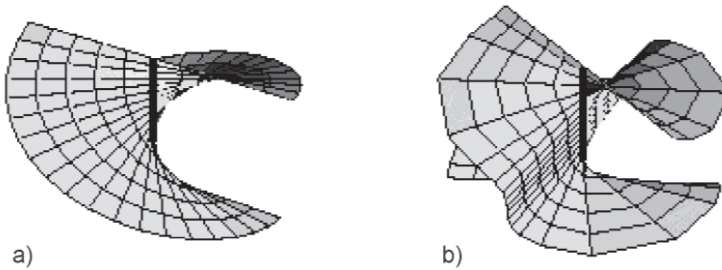
Kolejną charakterystyczną cechą wiązek przenoszących wir optyczny jest azymutalna faza określona wyrażeniem $\exp\{-in\theta\}$ (równ. (1)). Wyrażenie to jest odpowiedzialne za helikalny kształt frontu falowego tych wiązek oraz za stowarzyszony z nimi niezerowy moment pędu.

Oś helikoidy, przedstawionej na rysunku 3, leży wzdłuż kierunku rozchodzenia się fali i jest zbiorem punktów, w których faza jest nieokreślona. W przekroju płaszczyzną otrzymujemy obraz linii równej fazy zbiegającej się do punktu, czyli wir optyczny [14], [73] (rys. 2). Skręcony front falowy występujący w najbliższym otoczeniu wiru optycznego nie musi mieć kształtu regularnej helisy. Dalej znajduje się przykład niejednorodnego rozkładu fazy w pobliżu wiru optycznego zaczerpnięty z pracy M.R. Dennisa.



Rys. 3. Przykład heliakalnych fal. Przedstawiono helikalne fronty falowe dla ładunków topologicznych: a – $m = 1$, b) – $m = -1$ [51]

Fig. 3. Example of helical waves. Represented are the wave fronts of helical modes for the topological charge: a – $m = 1$, b – $m = -1$ [51]



Rys. 4. Dwie powierzchnie równej fazy w otoczeniu wiru optycznego: a – jednorodny helikalny front falowy, b – niejednorodny front falowy [21]

Fig. 4. Two surfaces of constant phase in the vicinity of a twisted wave dislocation: a – uniform twist, b – nonuniform twist [21]

Kierunek obrotu helikalnego frontu falowego determinuje znak ładunku topologicznego. Jeśli front falowy ulega skręceniu zgodnym z kierunkiem wskazówek zegara, analizując wzdłuż kierunku propagacji, to znak wiru będącego w jego centrum jest dodatni, w przypadku lewoskrętnej helikoidy – wir ma znak ujemny [8].

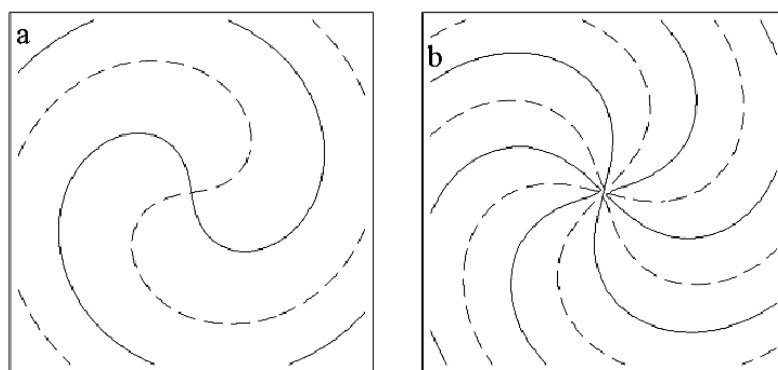
Helikalny kształt frontu falowego jest odpowiedzialny za niezerowy moment pędu fali świetlnej [6], który z kolei sprawia, że wiry są stabilną strukturą w ramach frontu falowego. W skład całkowitego momentu pędu fali świetlnej wchodzi również moment pędu związany z jej polaryzacją. Przez analogię do fizyki atomu, gdzie elektronowi krążącemu wokół jądra przypisuje się własny moment pędu – spin oraz orbitalny moment pędu, moment pędu fali świetlnej związany z wirem optycznym nazywamy orbitalnym momentem pędu. Z niezerowego momentu pędu wynika również fakt, iż wiry optyczne powstają lub zanikają parami (dodatni z ujemnym). Ponieważ wiry optyczne wskazują wiele podobieństw do cząstek naładowanych, czasem określa się je mianem quasi-cząstek.

Istnienie niezerowego momentu pędu w wiązkiach przenoszących wiry optyczne zostało potwierdzone doświadczalnie. Dobrym przykładem jest tu eksperyment wyko-

nany przez H. He, M. Friese, N. Heckenberga i H. Rubinsztein-Dunlopa [38]. Polegał on na tym, że zawieszinę drobnych cząstek, o dużym współczynniku absorpcji, oświetlono falą świetlną przenoszącą wiry optyczne. Absorpcji fali świetlnej przez cząsteczki zawiesziny towarzyszyło przekazanie momentu pędu z fali do cząsteczek. W efekcie czego zawiesziny wirowały w lewo lub w prawo zależnie od znaku oddziałującego z nimi wiru.

W punkcie, w którym występuje osobliwość fazy fali świetlnej o amplitudzie zespolonej U , wartość rzeczywista funkcji $U(\text{Re}(U))$ oraz wartość urojona funkcji $U(\text{Im}(U))$ wynoszą zero [30], [33], [75]:

$$\begin{cases} \text{Re}(U) = 0 \\ \text{Im}(U) = 0 \end{cases} \quad (3)$$



Rys. 5. Przykłady ilustrujące przecięcie linii zerowych $\text{Re}(U)$ (linia ciągła) i $\text{Im}(U)$ (linia przerywana):
a – dla wiru optycznego $m = 1$, b – dla wiru optycznego $m = -3$ [56]

Fig. 5. Plot of lines $\text{Re}(U) = 0$ (solid lines) and $\text{Im}(U) = 0$ (dashed lines):
a – for optical vortex $m = 1$, b – for optical vortex $m = -3$ [56]

Amplitudę zespoloną U fali świetlnej (gdy $z = \text{const}$) można przedstawić w postaci

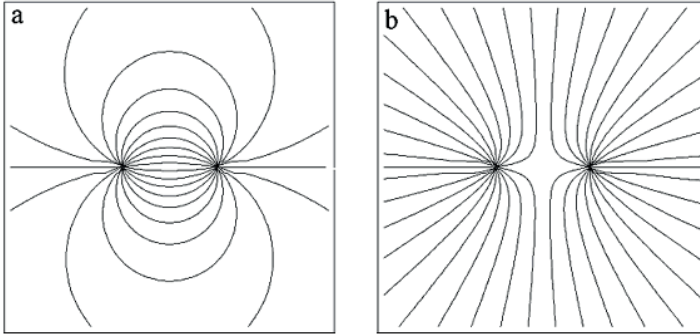
$$U(x, y) = U_{\text{Re}}(x, y) + i \cdot U_{\text{Im}}(x, y) \quad (4)$$

gdzie $U_{\text{Re}}(x, y)$ jest częścią rzeczywistą amplituda zespolona fali świetlnej $U(x, y)$, a $U_{\text{Im}}(x, y)$ – częścią urojoną.

W punkcie przecięcia linii zerowych funkcji U_{Re} i U_{Im} faza fali jest nieokreślona, a wokół tej osobliwości rozbudowuje się wir optyczny (rys. 5). Należy podkreślić, że wir optyczny to nie tylko punkt osobliwy – nazywany „punktem wiru”, ale otoczenie tego punktu. W trójwymiarowym obrazie wir optyczny oznacza tę część frontu falo-wego, która ma heliakalną geometrię.

W czasie propagacji w wolnej przestrzeni wiry mogą być tworzone lub niszczone, lecz całkowity topologiczny ładunek zostaje zachowany [31]. Ogólnie rzecz biorąc,

w miejscach, w których występuje nieokreśloność fazy, tworzone są bliźniacze ładunki topologiczne o tej samej wartości, lecz przeciwnych znaków.



Rys. 6. Linie równej fazy pomiędzy dwoma wirami optycznymi:
a – przeciwnych znaków (wir dodatni i ujemny); b – tych samych znaków (wiru dodatnie) [56]

Fig. 6. Equiphasing lines of two vortices:

a – one negative and one positive vortex; b – two positive vortices [56]

Na rysunku 6 przedstawione są dwa rozkłady fazy pola optycznego, dla dwóch par wirów. Łatwo zauważalne jest podobieństwo pomiędzy rozkładem fazy pola optycznego pomiędzy dwoma wirami a rozkładem linii sił pola elektrycznego wytworzonego przez dwa ładunki elektryczne o przeciwnych lub tych samych znakach. Rysunek 6b przedstawia dwa bliskie sobie wiry równoimienne, w których sąsiedztwie nie ma wirów przeciwnego znaku. Tego typu fronty falowe są trudne do uzyskania i utrzymania podczas propagacji fali. O możliwości praktycznej realizacji piszą I. Freund, N. Shvartsman oraz V. Freilikhen [29], opisując nieciągłości występujące w polach speklowych, np. układ wirów przedstawiony przez nazwany dyslokacją „combo”.

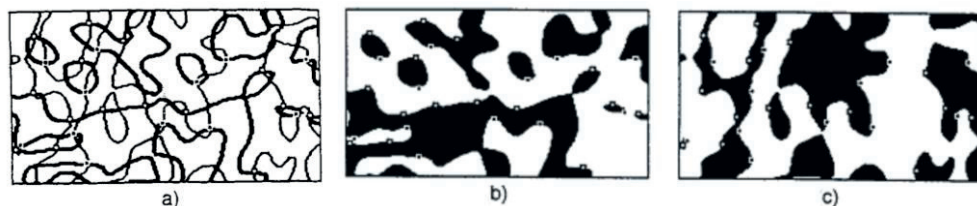
Dany front falowy może zawierać liczny zbiór wirów optycznych, dla których można sformułować następującą regułę znaków: *Dwa sąsiednie wiry, połączone linią równej fazy muszą być przeciwnych znaków* [31].

Aby przedstawić proste zastosowanie reguły znaków, rozważmy obszar, w którym znajdują się dwa sąsiednie wiry R i Q ze znakami odpowiednio $(+)$ i $(-)$. Jeśli mamy wir-bliźniak, stworzony pomiędzy R i Q , to reguła znaków wymaga następującej konfiguracji: $R(-)(+)Q$ i zakazuje konfiguracji $R(+)(-)Q$.

Trzy wnioski wynikające z reguły znaków:

- i. Znak każdego pojedynczego wiru w polu falowym automatycznie ustala znaki wszystkich wirów w tym polu.
- ii. Jeśli znak wiru znajdującego się w polu falowym zostanie zmieniony, znaki wszystkich pozostałych wirów muszą ulec zmianie.
- iii. Znak pierwszego wiru tworzonych podczas ewolucji pola falowego ustala znaki wszystkich przyszłych wirów [31].

Jedną z metod zaproponowaną przez I. FreundA i N. ShvartsmanA [30], [31], służącą do określania znaków zbioru wirów optycznych powstałych w optycznym polu speklowym, jest metoda czarno-białych map. Metoda ta wymaga wykreślenia dwóch map. Jedna z tych map przedstawia znak gradientu fazy wzdłuż osi x , a druga – znak gradientu fazy wzdłuż osi y . Po złożeniu ich razem powstaje wykres, na którym widoczne są punkty przecięcia $\varphi_x = \partial\varphi/\partial x$ oraz $\varphi_y = \partial\varphi/\partial y$. Punkty te wyznaczają położenie wirów optycznych.



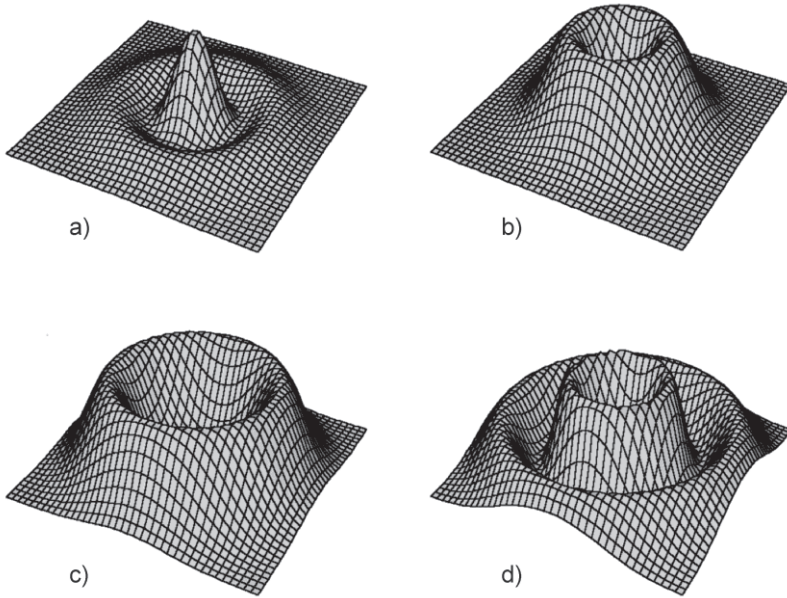
Rys. 7. Przykład czarno-białych map. Mapa: a – przedstawia powstałe kontury z nałożenia dwóch map (b) i (c). Mapa b – dla φ_x , c – dla φ_y . Kolor czarny oznacza obszar dodatni, kolor biały oznacza obszar ujemny [34]

Fig. 7. Example of black-white maps. Map: a – show outline arise from two maps (b) and (c). b – Map of φ_x , c – Map of φ_y . Black colour marked positive region, white colour marked negative region [34]

Na rysunku 7 kolor biały przypisany jest obszarowi, w którym rozważana funkcja ma wartości dodatnie, kolor czarny oznacza, iż funkcja ma wartości ujemne. Analiza gradientu rozkładu przedstawionego na dwóch mapach (rys. 7b i 7c) w okolicy punktu przecięcia umożliwia określenie znaku wirów optycznych.

Pierwsze prace z tematyki wirów optycznych dotyczyły pól optycznych wytworzonych przez fale monochromatyczne. W roku 2002 M.V. Berry stwierdził, że możliwe jest obserwowanie różnobarwnego rozkładu natężenia w pobliżu wiru optycznego [11], [12]. J. Leach oraz M.J. Padgett zaproponowali doświadczenie, w którym zaobserwowali ten efekt [45]. Fizycy z Uniwersytetu w Glasgow twierdzą, że światło białe zawierające wiry optyczne może być zastosowane w kwantowej komunikacji [5], [22] oraz do pułapkowania i manipulowania mikrocząstkami biologicznymi w urządzeniach nazywanych manipulatorami optycznymi (ang. *optical tweezers*) [5], [7], [35].

Typowe pole optyczne zawiera dużą liczbę wirów optycznych. Wytworzenie pola wolnego od takich struktur wymaga precyzyjnie kontrolowanych warunków, a każde zaburzenie wprowadzone do tak wygenerowanej wiązki powoduje powstanie wirów optycznych. Aby zdać sobie sprawę jak trudno jest uzyskać monochromatyczną wiązkę wolną od wirów optycznych, wystarczy zdać sobie sprawę, że jest to równoważne z uzyskaniem pola wolnego od spekli, z którymi wiry optyczne są ściśle związane [30], [32]. Ta właściwość wirów optycznych określana jest zwrotem *generic* [63], które nie ma jeszcze powszechnie przyjętego polskiego tłumaczenia.



Rys. 8. Rozkład natężenia światła w płaszczyźnie obserwacji modów LG_p^l dla następujących wartości parametrów p i l : (a) $p = 1; l = 0$; (b) $p = 0; l = 1$; (c) $p = 0; l = 3$; (d) $p = 1; l = 2$. Wiązki b, c, d zawierają wir optyczny. Skala osi z nie jest zachowana [56]

Fig. 8. Light intensity distribution in LG modes in the beam waist plane:
(a) $p = 1; l = 0$; (b) $p = 0; l = 1$; (c) $p = 0; l = 3$; (d) $p = 1; l = 2$. Waves b, c and d include optical vortices.
The plots have different z -scales

W typowym koherentnym polu optycznym wiry optyczne, podobnie jak spekle, występują w dużej ilości, ale rozpatrywane pojedynczo są drobnymi fragmentami całej wiązki. Łatwo jest również uzyskać wiązki, dla których pojedyncze wiry optyczne są dominującymi strukturami. Przykładem mogą tu być tak zwane mody Laguerra–Gausa (LG) promieniowania laserowego. Rozkład amplitudy w wiązce LG można przedstawić za pomocą następującego wyrażenia [66]:

$$u_{nm}^{LG} = E_0 A_{lm}^L \frac{w_0}{w} \left(\frac{\sqrt{2\rho}}{w} \right)^l \exp\left(-\frac{\rho^2}{w^2}\right) L_p^l\left(\frac{2\rho^2}{w^2}\right) \exp\left\{i\left[\frac{k\rho^2}{2R_z} + l\varphi - (2p+l+1)\arctan\left(\frac{z}{z_R}\right) + kz\right]\right\} \quad (5)$$

gdzie: E_0 – amplituda wiązki, A_{lm}^L – znormalizowana stała, w_0 – określa rozmiar „plamki” ($1/e$), $w_z = w_0 \sqrt{1+z^2/z_R^2}$ jest to poprzeczny rozmiar wiązki w płaszczyźnie z . Ta

wartość określa punkt, w którym amplituda o wartości maksymalnej $\sqrt{I_0}$ w centrum wiązki ($\rho=0$) osiąga wartość e razy mniejszą $\sqrt{I_0}/e$ ($\rho=w_z$); z_R – zakres Rayleigha ($z_R = k w_0^2/2$) – opisuje wartości, dla których natężenie osiąga połowę wartości; $\text{atan}(z/z_R)$ – faza Gouya, która zmienia się od 0 dla $z=0$ do $-\pi/2$ dla $z \rightarrow \infty$; R_z – promień krzywizny, dla $z=0$ mamy $R_z = \infty$ – front falowy jest płaski, dla $z \gg z_R$ front falowy staje się sferyczny; $L_p^l(x)$ – wielomian Laguerre’a, $l = m - n$ oraz $p = \min(n, m)$. Zamiennie z oznaczeniem u_{mn}^{LG} stosowane jest również oznaczenie u_{lp}^{LG} .

Dalej przedstawionych jest parę przykładów rozkładu natężenia światła w płaszczyźnie obserwacji modów LG_p^l .

2. METODY GENEROWANIA WIRÓW OPTYCZNYCH

Istnieje wiele metod generowania wirów optycznych powstałych na potrzeby różnych zastosowań. Dalej, w sposób skrótowy, opisanych jest kilka najpopularniejszych metod, takich jak: metoda holograficzna, metoda soczewek cylindrycznych, metoda spiralnej płytki fazowej, metoda podziału frontu falowego oraz metoda interferencyjna. Ze względu na wykorzystanie w technice Interferometrii na wirach optycznych najwięcej uwagi poświęcono metodzie Interferencyjnej. Metoda ta była wykorzystywana do generacji sieci wirów optycznych na potrzeby przeprowadzonych eksperymentów. Opisane metody nie stanowią wszystkich możliwych do wykorzystania sposobów otrzymywania helikalnych frontów falowych, można wymienić jeszcze takie metody, jak: separacja modów laserowych [1], [6], [39], [75] czy metodę opierającą się na wykorzystaniu zjawisk nieliniowych [3], [49].

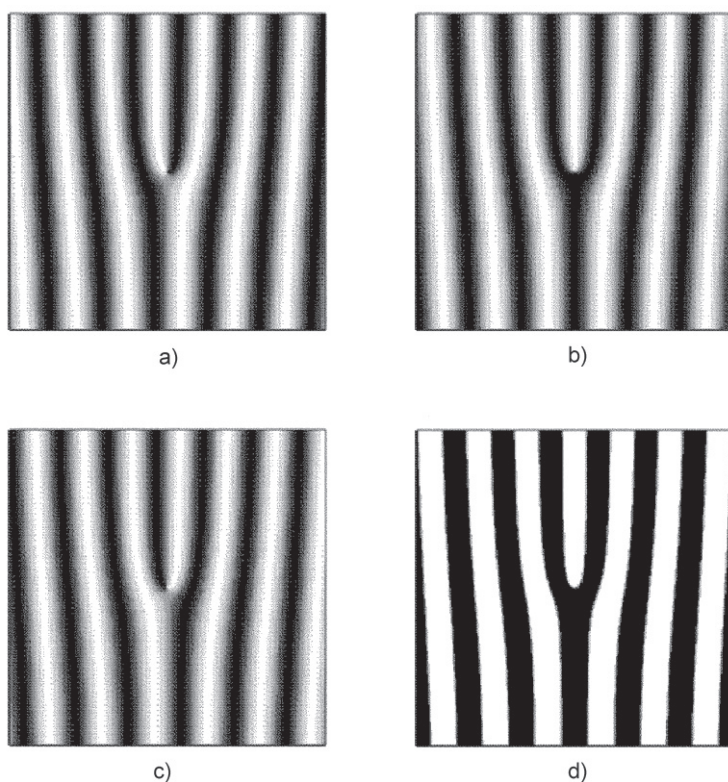
Opisywane metody generacji wirów optycznych zostały stworzone w celu sztucznego wprowadzenia wirów optycznych w pole optyczne jako obiektów dominujących, tworzących określoną strukturę. Samo uzyskanie wirów optycznych, nieregularnie rozmieszczonych, nie jest trudne. Typowe pole optyczne zawiera dużą liczbę wirów optycznych, których występowanie jest ściśle związane z występowaniem spekli [30], [32]. Ta właściwość wirów optycznych określana jest zwrotem *generic* [63].

2.1. METODA HOLOGRAFICZNA

Metoda holograficzna jest niewątpliwie najpopularniejsza i najczęściej opisywana w literaturze [39], [52], [68], [75]. Do generowania wiązek z wirami stosowane są specjalnie zaprojektowane hologramy syntetyczne. Zaletą metody holograficznej jest jej duża elastyczność, dzięki czemu w łatwy sposób można zaprojektować hologramy generujące nawet skomplikowane układy wirów optycznych. Ponadto przy niedużych

wymaganiach, co do parametrów hologramów, procedura ta nie wymaga użycia wyspecjalizowanego sprzętu.

W metodzie holograficznej wiązka ulega dyfrakcji na specjalnie wytworzonym hologramie syntetycznym. Zamiast wytwarzać hologram klasyczną metodą, zostaje on wygenerowany za pomocą komputera (obliczany jest układ prążków). Wydrukowanie tak powstałego układu prążków na folii, z użyciem drukarki laserowej o rozdzielczości 600 dpi, umożliwi otrzymanie hologramów o niskiej jakości nadających się na pokazy w pracowniach studenckich. Najwyższej jakości hologramy uzyskuje się dzięki zastosowaniu litografii elektronowej. Obecnie dynamicznie zmieniające się struktury układów prążków w doświadczeniach uzyskiwane są za pomocą sterowanej komputerowo matrycy ciekłych kryształów.

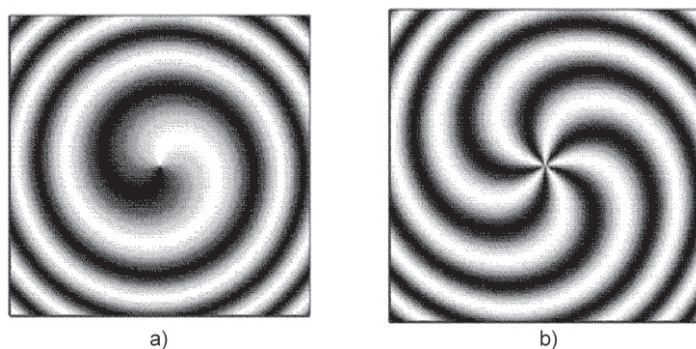


Rys. 9. Przykładowe hologramy służące do generacji wiru optycznego [56]
 Fig. 9. Example of holograms used to generation the optical vortex [56]

Na rysunku 9 przedstawiono charakterystyczną strukturę prążków. W punkcie, w którym znajduje się wir optyczny, obserwujemy rozwidlanie się prążka. Taki kształt prążków nazywa się również „widelcem”. Na rysunku 9 (w części a, b oraz c) poka-

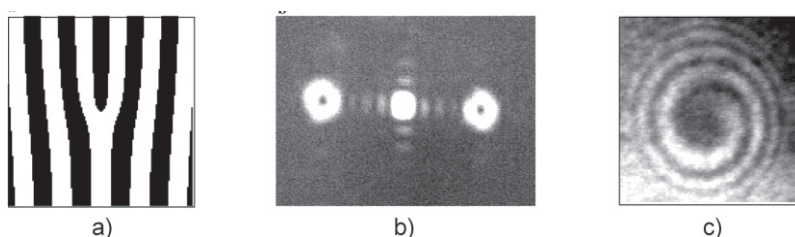
zono również, że w zależności od fazy początkowej rozwidlony prążek utworzony jest z minimalnego lub maksymalnego natężenia światła. Rozwidlający się prążek ma symetryczny kształt, gdy różnica faz pomiędzy wiązką odniesienia a przedmiotową wynosi $\pi/2$. Na rysunku 9d z kolei zilustrowano różnicę w wyglądzie prążków powstałą po binaryzacji obrazu. Obserwujemy tu wyraźną granicę pomiędzy prążkami jasnymi a ciemnymi.

Dalej (rys. 10) przedstawione są inne przykłady osiowych hologramów generujących wiry o ładunku 1 (rys. 10a) oraz wir o ładunku 5 (rys. 10b).



Rys. 10. Przykład osiowych hologramów generujących wiry o ładunku: a – $m = 1$ oraz b – $m = 5$ [56]
 Fig. 10. The one-axis holograms generation optical vortex with charge: a – $m = 1$ and b – $m = 5$ [56]

Jeżeli wygenerowany hologram oświetlimy wiązką rekonstruującą, to otrzymamy wiry optyczne w falach odpowiadających kolejnym rzędom ugięcia. Wir optyczny znajduje się w wiązkach rzędów 1, 2, 3 itd. (w wiązce zerowego rzędu nie obserwujemy wiru).



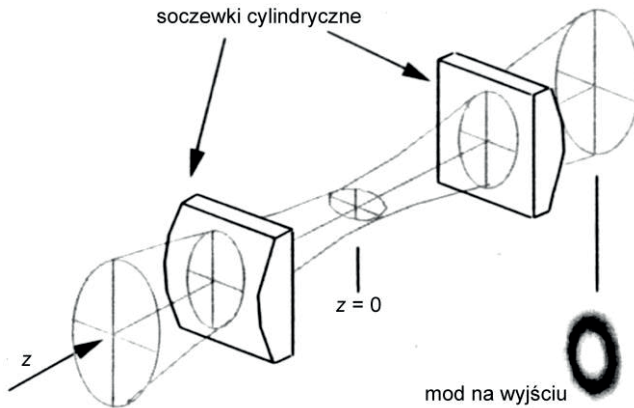
Rys. 11. Przykład hologramu wygenerowanego komputerowo: a – oraz rekonstrukcji tego hologramu
 b, c – przedstawia interferencję wiązki z wirem z płaską poosiową falą odniesienia
 Fig. 11. Example of numerical holograms: a – and reconstruction of the hologram
 b, c – the interferogram of the first order diffraction beam with coaxial plane wave

Na rysunku 11a przedstawiono przykładowy hologram wygenerowany numerycznie i wydrukowany na folii. Na rysunku 11b przedstawiono obraz tego hologramu, z oświetleniem wiązką laserową. W rzędzie +1 i -1 widoczne są nieciągłości w postaci czarnych

punktów świadczących o obecności wirów w tych wiązках. W wyższych rzędach, nie pokazanych na fotografii, generują się wiry o większym ładunku topologicznym, które w wyniku propagacji rozpadają się na grupy wirów o jednostkowym ładunku. Na rysunku 11c przedstawiono interferencję wiązki z wirem z płaską poosiową falą odniesienia. W wyniku tej interferencji otrzymujemy charakterystyczny spiralny kształt prążków.

2.2. METODA SOCZEWEK CYLINDRYCZNYCH

Wiązkę z wirami optycznymi można również otrzymać, używając dwóch jednakowych soczewek cylindrycznych [5], [6], [18], [75]. Wynik działania układu soczewek cylindrycznych można porównać do wyniku działania płytki ćwierćfalowej, która zamienia światło liniowo spolaryzowane na kołowo spolaryzowane. Na rysunku 12 przedstawiony jest układ dwóch soczewek cylindrycznych, które wytwarzają helikalny front falowy.

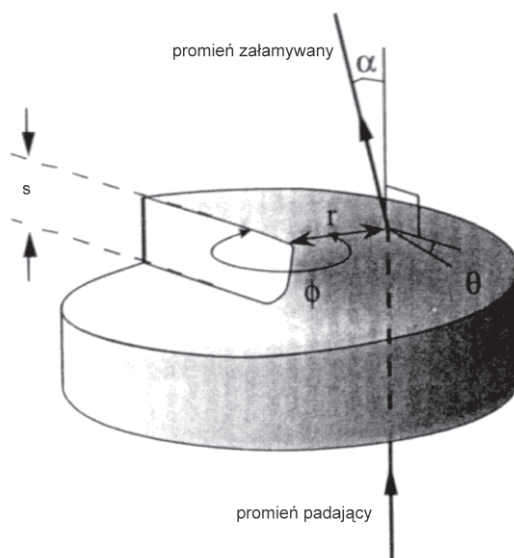


Rys. 12. Układ dwóch cylindrycznych soczewek [6]
Fig. 12. Setup of two cylindrical lenses [6]

Soczewki cylindryczne o ogniskowej f , przedstawione na rysunku 12, oddalone są od siebie o odległość równą $\sqrt{2}f$. W takim układzie z wejściowej wiązki Hermite–Gausa (HG) otrzymujemy wiązkę Laguerre–Gausa (LG_0^1). Jeżeli odległość między soczewkami zwiększymy do $2f$, to układ działa analogicznie do półfalówki, tzn. np. wejściową helisę lewoskrętną zamieni na prawoskrętną na wyjściu.

2.3. METODA SPIRALNEJ PŁYTKI FAZOWEJ

W następnej metodzie helikalny front falowy uzyskiwany jest z użyciem przezroczystej spiralnej płytki fazowej [5], [6], [15], [41] przedstawionej na rysunku 13.



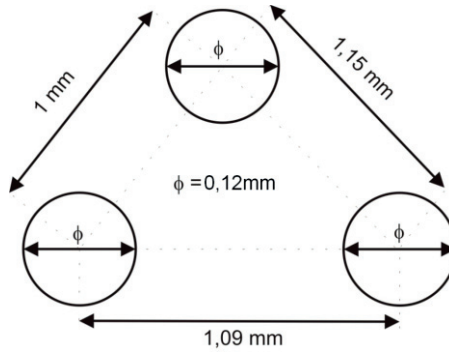
Rys. 13. Spiralna płytko fazowa [6]

Fig. 13. Spiral phase-plate [6]

Przedstawiona (rys. 13) płytko fazowa zaprojektowana jest tak, że skok oznaczony literką s jest rzędu całkowitej krotności długości fali oświetlającej płytkę λ . Na rysunku tym pokazany jest bieg przykładowego promienia świetlnego biegnącego wzdłuż osi optycznej. Kąt θ określa krzywiznę skręconej powierzchni płytki. Kąt α określa odchylenie przykładowego promienia świetlnego po przejściu przez spiralną płytkę fazową od osi optycznej. Wadą skręconych płytek fazowych jest to, że należy wykonywać je na określonej długości fali. Ponieważ do generacji wirów najczęściej wykorzystuje się światło z zakresu widzialnego, więc używane długości fali są rzędu kilkuset nanometrów. Tak małe długości używanych fal stanowią dodatkowe utrudnienia w wykonywaniu skręconych płytek fazowych. Obecnie do wytwarzania spiralnych płytek fazowych stosowana jest technika ciekłych kryształów lub polimerów [37].

2.4. METODA PODZIAŁU FRONTU FALOWEGO

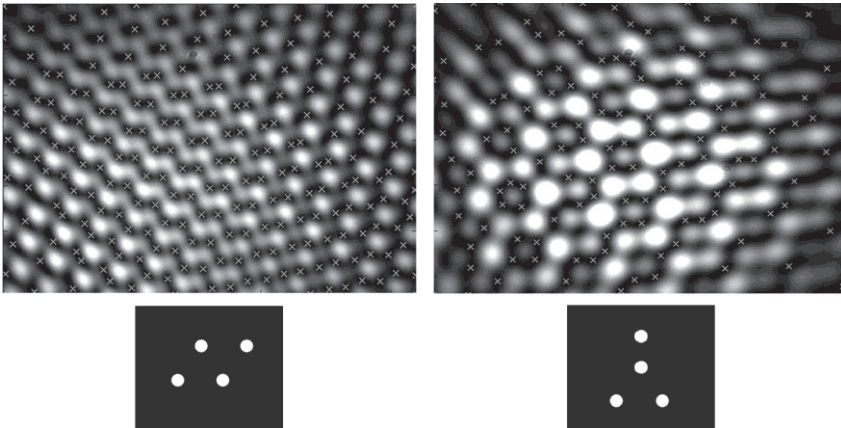
Następna metoda tworzenia wirów optycznych polega na podziale frontu falowego przez ustawienie na drodze wiązki laserowej płyty z otworkami. Przykładowe rozmieszczenie i wielkość otworków przedstawiane są na rysunku (rys. 14). W takim rozmieszczeniu otworów uzyskać można heksagonalną strukturę sieci wirów optycznych. Do generacji wirów optycznych wykorzystana została idea Younga [79].



Rys. 14. Przykład rozmieszczenia otworów i ich wielkości umożliwiające uzyskanie heksagonalnej struktury sieci wirów optycznych [58]

Fig. 14. Example of holes arrangement and size to obtain hexagonal lattice of vortices [58]

Na rysunku 15 przedstawiono inne przykładowe struktury sieci wirów w zależności od rozmieszczenia otworów w płycie na drodze wiązki światła.



Rys. 15. Przykładowe sieci wirów uzyskane przez podział frontu falowego za pomocą czterech otworów. Różne rozmieszczenie otworów i odpowiednie obrazy interferencyjne [58]

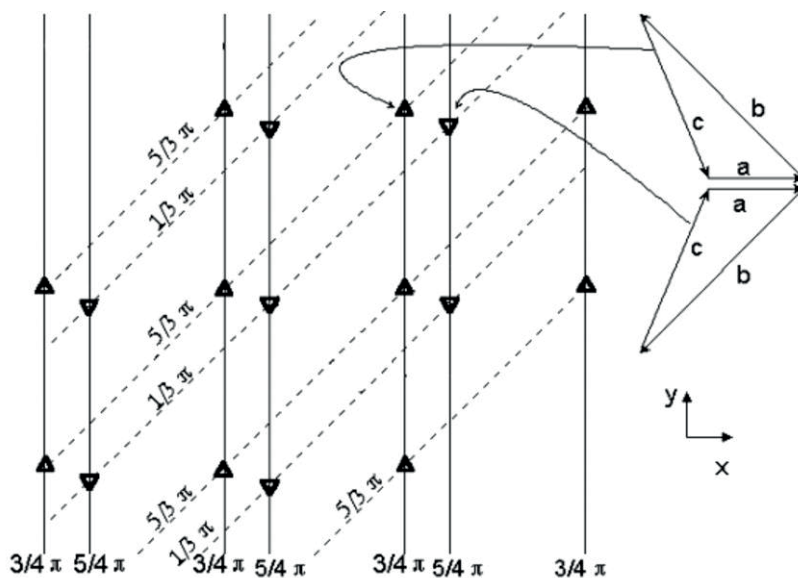
Fig. 15. Example of vortices lattice obtained by wave-front division with the use four holes.

Various of holes arrangements and adequate interference patterns [58]

2.5. METODA INTERFERENCYJNA

Interferencja dwóch lub większej liczby fal może dać efekt taki, że amplituda wypadkowa w izolowanych punktach będzie wynosić zero. W takich punktach powstają wiry optyczne. Ze względu na zastosowanie szczególnie ciekawy jest przypadek interferencji trzech fal płaskich, w wyniku czego powstaje regularna sieć wirów optycz-

nych [53], [61]. Dalej pokazany jest schematycznie przykład wyniku nakładania się trzech fal płaskich.

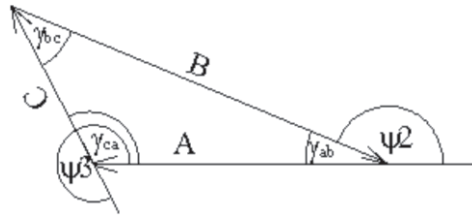


Rys. 16. Linie równej fazy trzech interferujących fal płaskich [54]

Fig. 16. Phase distribution of three plane waves [54]

Rysunek 16 został wykonany dla fal o amplitudach $A = 1$, $B = 3,3$, $C = 2,7$ oraz odpowiednio wektorach falowych $\mathbf{k}_a(0, 0, k)$, $\mathbf{k}_b(0,01k, 0, k)$, $\mathbf{k}_c(0,01k, -0,01k, k)$, z założeniem, że jednorodna fala płaska zadana jest przez amplitudę i kosinusy kierunkowe. O gęstości linii równej fazy będzie decydował kąt nachylenia wektora falowego \mathbf{k} do osi z . O orientacji tych linii zadecydują kąty między wektorem falowym a osiami x i y . Orientację linii równej fazy wyznaczy nam linia będąca przecięciem powierzchni falowej fali płaskiej i płaszczyzny obserwacji.

W zilustrowanym przypadku (rys. 16) zakładamy, że pierwsza fala pada prostopadle do płaszczyzny rejestracji. Wzdłuż kierunku propagacji tej fali orientujemy oś z . Druga fala pada pod pewnym, niewielkim kątem w stosunku do osi z . Oś x układu współrzędnych określamy tak, aby była prostopadła do linii równej fazy wyznaczonej dla drugiej fali w płaszczyźnie obserwacji (ciągłe linie na rysunku). Trzecia fala pada pod innym kątem w stosunku do osi z niż fala druga. Jej linie równej fazy, w płaszczyźnie obserwacji, wyznaczone są przez linie przerywane (rys. 16). W punktach, w których wektory amplitudy zespolonej fal tworzą trójkąt powstają wiry optyczne. Suma wektorów amplitud zespolonych \mathbf{A} , \mathbf{B} , \mathbf{C} w punkcie, w którym utworzony jest trójkąt wynosi zero. Aby powstał trójkąt z wartości amplitud trzech interferujących fal, suma dowolnych dwóch wartości amplitud musi być większa od wartości trzeciej amplitudy [53].

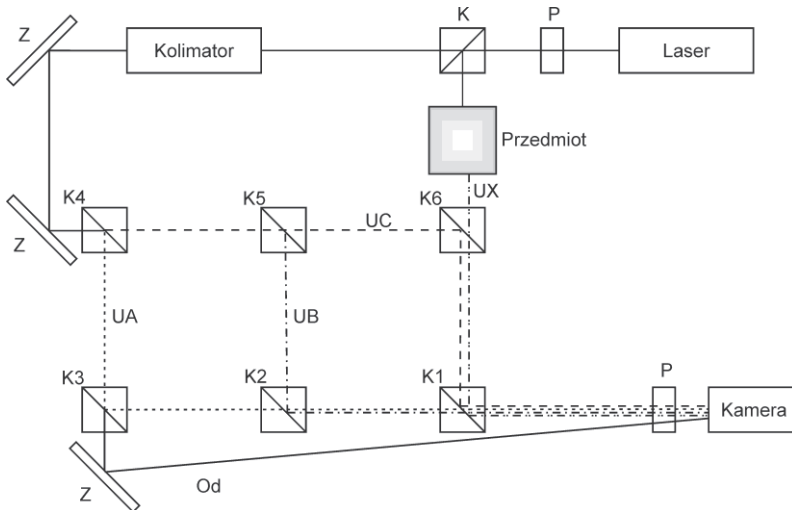


Rys. 17. Trójkąt utworzony z wektorów amplitud zespolonych **A**, **B**, **C** [56]
 Fig. 17. The triangle formed from complex amplitude vectors **A**, **B**, **C** [56]

Na rysunku 17 przedstawiono trójkąt utworzony z wektorów amplitudy zespolonej trzech interferujących fal. Przez symbole ψ_2 i ψ_3 oznaczona jest faza względna odpowiednio pomiędzy falami UA i UB oraz UA i UC . Jak widać z rysunku, pomimo iż faza jest nieokreślona w punkcie, w którym występuje zero amplitudy, to z geometrii trójkąta jesteśmy w stanie wyznaczyć fazę względną pomiędzy interferującymi falami.

2.5.1. SCHEMAT POMIAROWY INTERFEROMETRU NA WIRACH OPTYCZNYCH

Na przedstawionym dalej schemacie układu interferometru sieć wirów optycznych jest generowana w wyniku interferencji trzech fal płaskich (np. wiązki UA , UB , UC) [59]. Po przejściu przez kolimator rozszerzona wiązka laserowa pada na układ sześciu kostek światłodziących. Układ ten ma za zadanie podzielić wiązkę padającą na trzy fale płaskie, które pochyłone są pod różnymi kątami do osi optycznej. Odpowiednio



Rys. 18. Schemat ideowy układu interferometru na wirach optycznych
 Fig. 18. Experimental setup of the Optical Vortex Interferometer

skalibrowany układ umożliwiłaby otrzymanie regularnej sieci wirów optycznych. Kostki światłodzielnice w bloku interferometru rozmieszczone są w ten sposób, aby odległości między sąsiednimi kostkami były takie same. Takie ustawienie kostek umożliwia obserwację prążków o dużym kontraście, ponieważ różnica dróg optycznych fal UA , UB i UC jest mała (mniejsza od 3 cm). Oś optyczną interferometru wyznaczają dwa punkty leżące w środku kostek $K3$ i $K1$.

Legenda:

K – kostka światłodzielnica;

Z – zwierciadło;

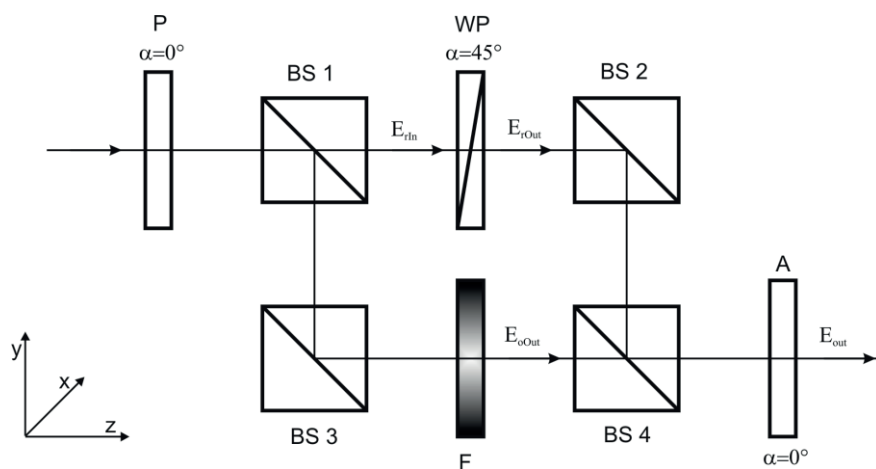
P – polaryzator;

UA, UB, UC – trzy fale płaskie;

Od – wiązka odniesienia;

UX – wiązka przedmiotowa.

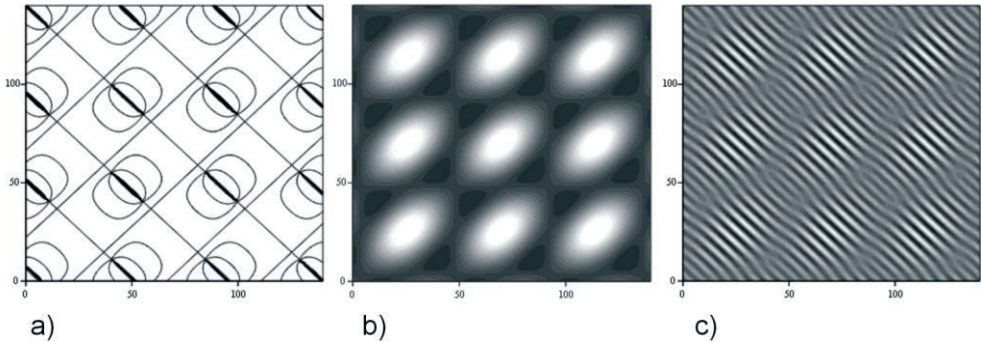
Na rysunku 19 przedstawiony jest inny układ interferometru umożliwiający generację regularnej struktury wirów optycznych [43]. W przeciwieństwie do zaprezentowanego jest on stabilniejszy, ponieważ występują w nim tylko dwie drogi optyczne interferujących wiązek. Wykorzystane są natomiast w tym układzie elementy polaryzacyjne. Nie jest jednak możliwe w układzie tym uzyskanie wszystkich czterech obrazów interferencyjnych wymaganych przez niektóre z opisywanych metod lokalizacji wirów optycznych (punkt 2). Rozkłady regularnych sieci nieciągłości optycznych można również otrzymać w interferometrze, w którym podział na interferujące fale odbywa się z użyciem dwóch pryzmatów Wollastona [44].



Rys. 19. Schemat interferometru wraz z układem współrzędnych: P – polaryzator, WP – pryzma Wollaston'a, A – analizator, F – filtr transmisyjny, BS1-4 – cztery kostki światłodzielnice [43]

Fig. 19. Setup of the interferometr setup and orientation of coordinate axes: P – polarizer; WP – Wollaston prism; A – analyzer; F – transmission filter; BS1-4 – four beam splitters [43]

W wyniku interferencji trzech fal płaskich otrzymujemy charakterystyczną regularną sieć wirów optycznych, która jest dokładniej opisana w punkcie 2.5.3. Na rysunku 20 przedstawiono przykładowe wyniki interferencji trzech fal płaskich, uzyskane w symulacji numerycznej.



Rys. 20. Linie równej fazy (a) i rozkład natężenia (b) powstały w wyniku interferencji trzech fal płaskich, (c) rozkład natężenia powstały w wyniku interferencji trzech fal płaskich z płaską wiązką odniesienia (na osi x i y znajdują się odległości podane w pikselach – rozmiar piksela wynosi $9 \mu\text{m}$)

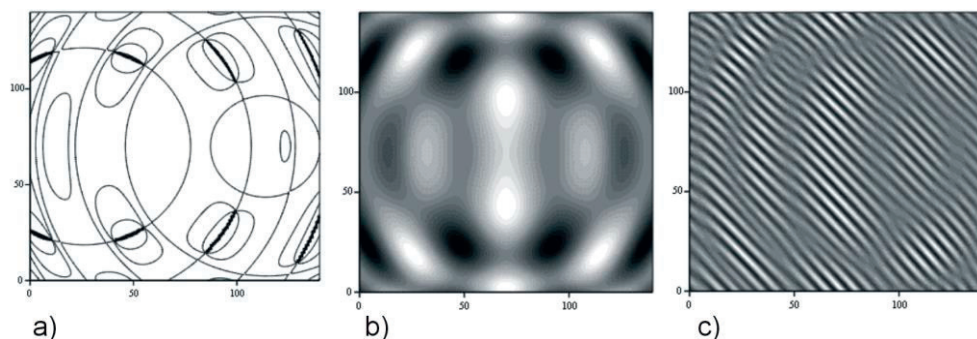
Fig. 20. The equiphase lines (a) and intensity distribution (b) generated by three plane waves interference, (c) intensity distribution generated by interference of three plane waves and reference plane wave (x and y axes in pixels – one pixel size is $9 \mu\text{m}$)

Na rysunku 20a i 20b przedstawiono rozkład fazy oraz natężenie powstałe w wyniku interferencji trzech fal płaskich o równych amplitudach. Na rysunku 20c widać charakterystyczną strukturę prążków rozwidlonych w miejscach występowania nieciągłości fazy, powstała w skutek dodania do trzech interferujących fal płaskich czwartej fali odniesienia o amplitudzie trzykrotnie większej od amplitudy jednej z trzech fal. Na interferogramie przedstawionym na rysunku 20b widać, że obszary o najmniejszym natężeniu, zawierające nieciągłości fazowe, są wyraźnie rozseparowane, co umożliwia zorientowanie się, gdzie znajdują się wiry optyczne.

Sieć wirów optycznych powstaje nie tylko w przypadku interferencji trzech fal płaskich. Dalej przedstawiony jest przykład symulacji numerycznej, w którym interferują ze sobą dwie fale płaskie oraz fala sferyczna.

Na rysunku 21, tak jak poprzednio, przedstawiono rozkład fazy (rys. 21a) i natężenia (rys. 21b) trzech interferujących fal (dwóch płaskich i jednej sferycznej) bez wiązki odniesienia oraz wynik interferencji trzech fal z płaską falą odniesienia (rys. 21c) o amplitudzie trzykrotnie większej od amplitudy jednej z fal płaskich. Położenia powstałych wirów optycznych nie tworzą, tak jak w przykładzie, regularnej sieci. Można jednak dopatrzeć się pewnej prawidłowości w położeniach kolejnych par wirów optycznych po przeanalizowaniu rozkładu fazy. Obserwacja samego rozkładu natężenia trzech fal nie jest wystarczająca do wskazania położenia wirów

optycznych. Dopiero położenia widelców (rys. 21c) zdradzają rozmieszczenie nieciągłości fazowych.



Rys. 21. Linie równej fazy (a) i rozkład natężenia (b) powstały w wyniku interferencji dwóch fal płaskich z falą sferyczną, (c) rozkład natężenia powstały w wyniku interferencji dwóch fal płaskich z falą sferyczną oraz płaskiej wiązki odniesienia (na osi x i y znajdują się odległości podane w pikselach – rozmiar piksela wynosi $9\ \mu\text{m}$)

Fig. 21. Equiphasic lines (a) and intensity distribution (b) generated by interference of two plane waves and a spherical wave, (c) intensity distribution generated by interference of two plane waves and one spherical wave and reference plane wave (x and y axes in pixels – one pixel size is $9\ \mu\text{m}$)

2.5.2. REGULARNE SIECI WIRÓW OPTYCZNYCH

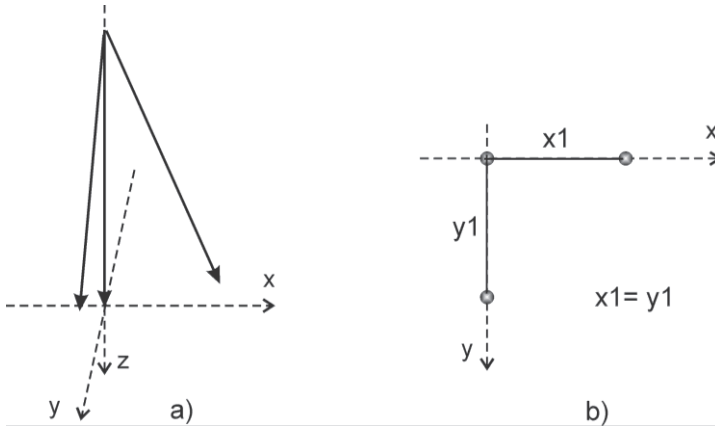
Występuje kilka sposobów generowania regularnych sieci wirów optycznych, np. przestrzenne sieci wirów optycznych można uzyskać za pomocą wygiętych płytek szklanych [42], w wyniku działania efektów nieliniowych generowanych w oparach rubidu [23], [24], [74] lub za pomocą rotującego kondensatu Bosego-Einsteina ^{87}Rb [47], jak również regularna sieć wirów może być wygenerowana w układzie interferometrycznym [53], [65] lub za pomocą odpowiednio przygotowanego hologramu [68], [75]. Regularne sieci wirów optycznych znalazły zastosowania np. do formowania dwuwymiarowych struktur fonicznych z użyciem trzech fal eliptycznie spolaryzowanych [60]. W niniejszej pracy zostaną opisane regularne sieci wirów optycznych powstałe w wyniku interferencji trzech fal płaskich.

2.5.3. INTERFERENCJA TRZECH FAL PŁASKICH

W wyniku interferencji trzech fal płaskich UA , UB i UC w płaszczyźnie obserwacji powstaje regularna sieć wirów optycznych. O odległościach między położeniami sąsiednich wirów decydują kierunki propagacji interferujących fal płaskich oraz wartości ich amplitud. Dalej przedstawiono kilka różnych przypadków sieci wirów powstałych na skutek interferencji trzech płaskich fal UA , UB i UC .

2.5.3.1. SIEĆ PAR

Przypadek 1 ilustruje sytuację, w której trzy fale płaskie rozchodzą się pod takimi kątami względem siebie, że w wyniku interferencji fal UA i UB powstają prążki pionowe; fal UA i UC – prążki poziome; fal UB i UC – prążki pochylone pod kątem 45° względem poprzednich prążków, przy czym prążki I_{AB} oraz I_{AC} mają taką samą gęstość. Przez „poziome” rozumiemy tu zorientowane równoległe do dłuższego boku matrycy CCD. Przykładowe kierunki rozchodzenia się fal pokazane są schematycznie na rysunku 22.

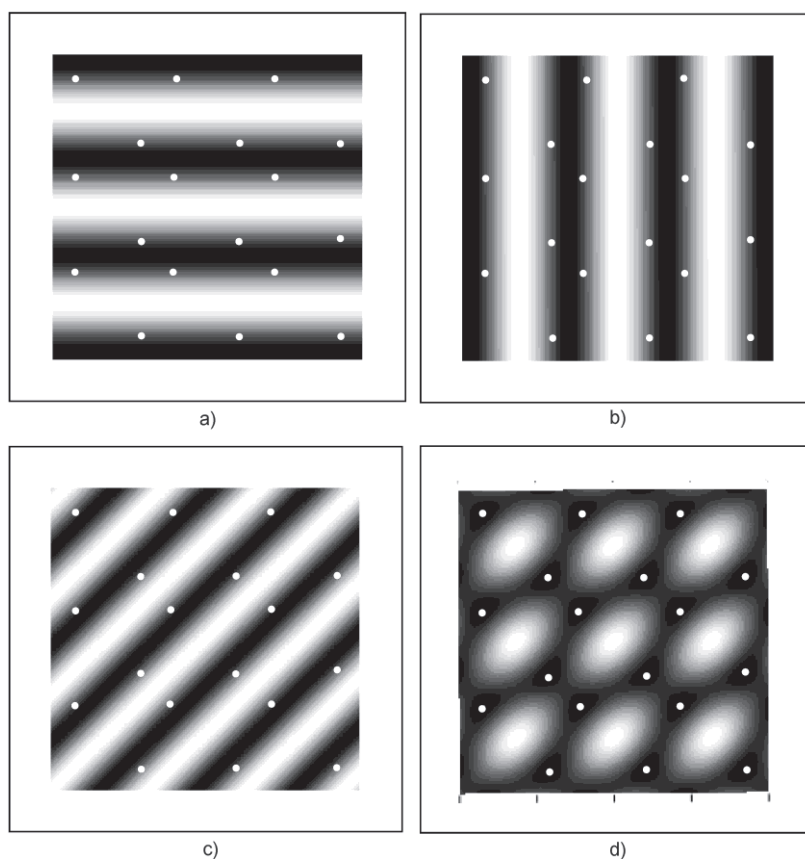


Rys. 22. Kierunki rozchodzenia się fal tworzących sieć par wirów optycznych. Wektory falowe trzech fal płaskich (a) oraz punkty oznaczające rzuty końców wektorów falowych na płaszczyźnie xy (b)
 Fig. 22. Propagation of waves which create lattice of optical vortices pair. Wave-vectors of three plane waves (a) and projection of wave-vectors endpoints on xy plane (b)

Na rysunku 23d przedstawiono wynik symulacji numerycznej interferencji trzech fal płaskich o takich samych wartościach amplitud. Na rysunku 23a, 23b i 23c przedstawiono interferogramy I_{AB} , I_{AC} oraz I_{BC} z zaznaczonymi, przez czerwone punkty, położeniami wirów optycznych. Wiry optyczne powstają tylko w przypadku interferencji trzech fal. Na interferogramach, przedstawiających interferencje pomiędzy każdą parą fal płaskich, naniesiono położenia wirów optycznych, ponieważ dane te wykorzystywane są w metodach lokalizacji oraz określania znaku wirów optycznych. Metody te opisane są w punkcie 5.

Powstała sieć została nazwana „siecią par”, ponieważ wyraźnie widać, które wiry optyczne tworzą parę wirów o przeciwnych znakach – są to wiry leżące w najmniejszej odległości od siebie.

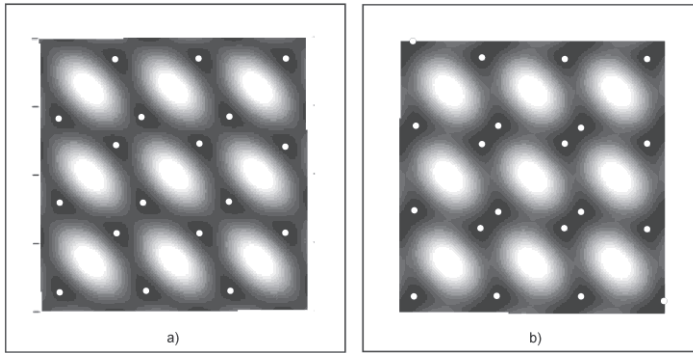
Na odległość między parą wirów mają wpływ wartości amplitud interferujących fal. Im większa jest różnica pomiędzy wartościami amplitud, tym bliżej siebie tworzą się sparowane wiry. Aby zilustrować to zjawisko, wygenerowane zostały dwa interfe-



Rys. 23. Wynik interferencji trzech fal płaskich UA , UB i UC – sieć par wirów optycznych:
 a – prążki interferencyjne utworzone przez fale UA i UB , b – prążki interferencyjne utworzone przez fale UA i UC , c – prążki interferencyjne utworzone przez fale UB i UC ; d – obraz interferencyjny utworzony przez fale UA , UB i UC . Białe kółka wskazują położenia punktów wirowych (na osiach x i y wartości odległości zostały pominięte)

Fig. 23. Result of interference of three plane waves UA , UB and UC – lattice of optical vortices pair:
 a – interference fringes generated by waves UA and UB , b – interference fringes generated by waves UA and UC , c – interference fringes generated by waves UB and UC , d – interference pattern generated by waves UA , UB and UC . Points indicate locations of vortex points (scales of axes x and y where omitted)

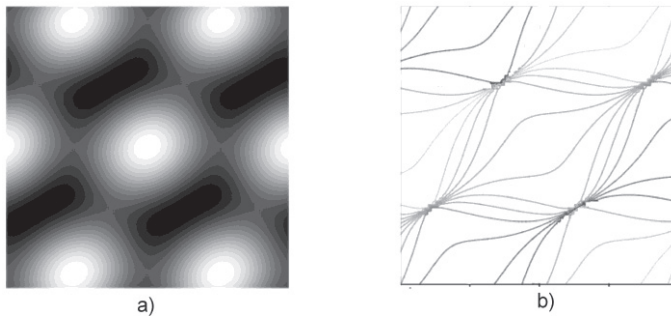
rogramy (rys. 24). Na pierwszym z nich (rys. 24a) widzimy wynik interferencji trzech fal płaskich o takich samych wartościach amplitud, natomiast na drugim (rys. 24b) przedstawiony jest interferogram, w przypadku gdy amplituda jednej z fal jest większa od pozostałych ($A = 1,7$, $B = C = 1$). Duże różnice w wartościach amplitud interferujących fal mogą spowodować, że rozróżnienie położenia dwóch wirów optycznych o przeciwnych znakach stanie się praktycznie niemożliwe.



Rys. 24. Rozkład natężenia trzech interferujących fal płaskich tworzących sieć par wirów optycznych. Wartości amplitud: a – $A = B = C = 1$, b – $A = 1,7$ $B = C = 1$. Białe kółka wskazują położenia punktów wirowych (na osiach x i y wartości odległości zostały pominięte)

Fig. 24. Intensity distribution of three interfering plane waves which generate lattice of optical vortices pair. Amplitude values: a – $A = B = C = 1$, b – $A = 1,7$ $B = C = 1$. Points indicate locations of vortex points (scales of axes x and y where omitted)

Jeżeli amplitudy interferujących trzech fal płaskich wynoszą $A = 2$, $B = C = 1$, to w miejscach, w których występują zera natężenia znajdują się nieciągłości fazowe, ale nie są to wiry optyczne. Przykładowy rozkład natężenia oraz fazy, dla wymienionego przypadku, pokazany jest na rysunku 25.



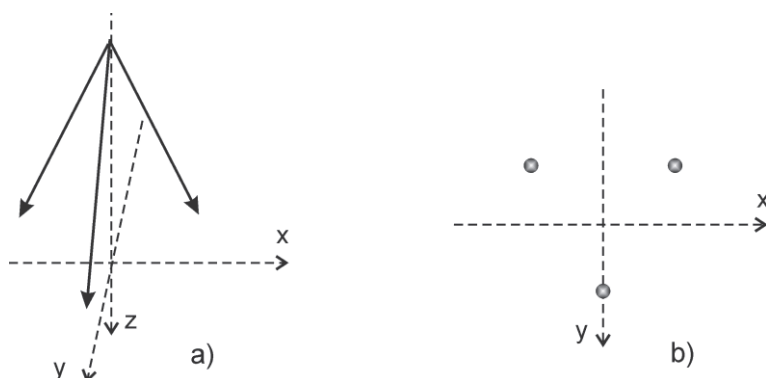
Rys. 25. Wynik interferencji trzech fal płaskich o amplitudach $A = 2$ oraz $B = C = 1$: a – rozkład natężenia, b – linie równej fazy (na osiach x i y wartości odległości zostały pominięte)

Fig. 25. Result of interference of three plane waves where amplitude $A = 2$ and $B = C = 1$: a – intensity distribution, b – equiphase lines (scales of axes x and y where omitted)

2.5.3.2. SIEĆ SZEŚCIOKĄTÓW

Przypadek 2 ilustruje sytuację, w której trzy fale płaskie rozchodzą się pod takimi względnymi kątami, że w wyniku interferencji fal UA i UB oraz UA i UC powstają prążki skośne, natomiast fal UB i UC – prążki pionowe, przy czym prążki I_{AB} oraz I_{AC}

mają taką samą gęstość. Przykładowe kierunki rozchodzenia się fal pokazane są schematycznie na rysunku 26.



Rys. 26. Kierunki rozchodzenia się fal tworzących sieć sześciokątną wirów optycznych. Wektory falowe trzech fal płaskich (a), oraz punkty oznaczające rzuty końców wektorów falowych na płaszczyźnie xy (b)

Fig. 26. Propagation of waves which create hexagonal lattice of optical vortices.

Wave-vectors of three plane waves (a), and projection of wave-vectors endpoints on xy plane (b)

Na rysunku 26b rzuty końców wektorów falowych tworzą trójkąt równoboczny.

Na rysunku 27d przedstawiono wynik interferencji trzech fal płaskich o takich samych wartościach amplitud. Na rysunku 27a, 27b i 27c przedstawiono interferogramy I_{AB} , I_{AC} oraz I_{BC} z zaznaczonymi jako czerwone kropki położeniami wirów optycznych. Powstała sieć została nazwana „siecią sześciokątów”, ponieważ położenia wirów optycznych tworzą sześciokąty foremne. Ten rodzaj sieci wirów optycznych został przedstawiony przez K.W. Nichollsa i J. F. Nye’a [61].

W podanych przykładach pokazano sieci wirów optycznych, których uzyskanie w układzie pomiarowym jest dość proste i powtarzalne. Istnieje bardzo wiele możliwości ustawień kątów nachylenia wektorów falowych trzech fal płaskich, w wyniku interferencji których powstają sieci wirów o nieco innej geometrii. Jednak we wszystkich tych przypadkach sieci wirów optycznych wykazane są pewne cechy wspólne. Wiry dodatnie (ujemne) tworzą się na przecięciach układu linii równoległych. Kierunek tych linii jest zgodny z kierunkiem prążków interferencyjnych fal $UA + UB$, $UA + UC$, $UB + UC$.

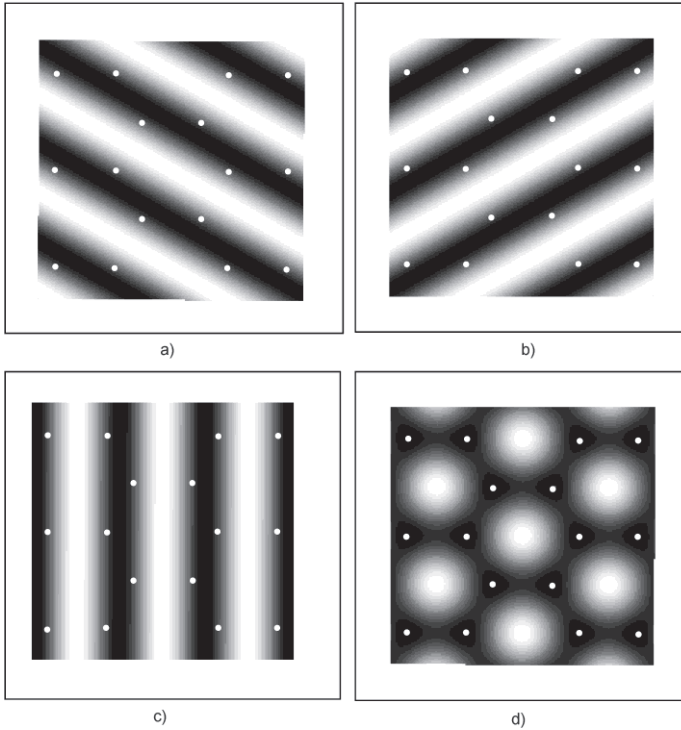
Sieć wirów wygenerowana przez trzy fale płaskie ma wiele szczególnych właściwości; to znaczy takich, które nie są obecne w sieciach wirów generowanych przez więcej niż trzy fale. Wyszczególniono te właściwości:

- Geometria dodatniej (ujemnej) podsieci wirów zależy tylko od współrzędnych wektorów falowych i nie zależy od amplitud i faz początkowych interferujących fal.

- Amplitudy trzech interferujących fal, w punktach, gdzie faza jest nieciągła, w sposób jednoznaczny definiują względne kąty fazowe między tymi falami.

- Geometria sieci wirów (obejmująca wiry dodatnie i ujemne) nie zależy od fazy interferujących fal. Gdy zmienia się faza początkowa jednej, dwu lub trzech interferujących fal, sieć przesuwa się jak ciało sztywne.

- Trzy końce wektorów falowych definiują płaszczyznę. Wiry optyczne propagują się prostopadle do tej płaszczyzny

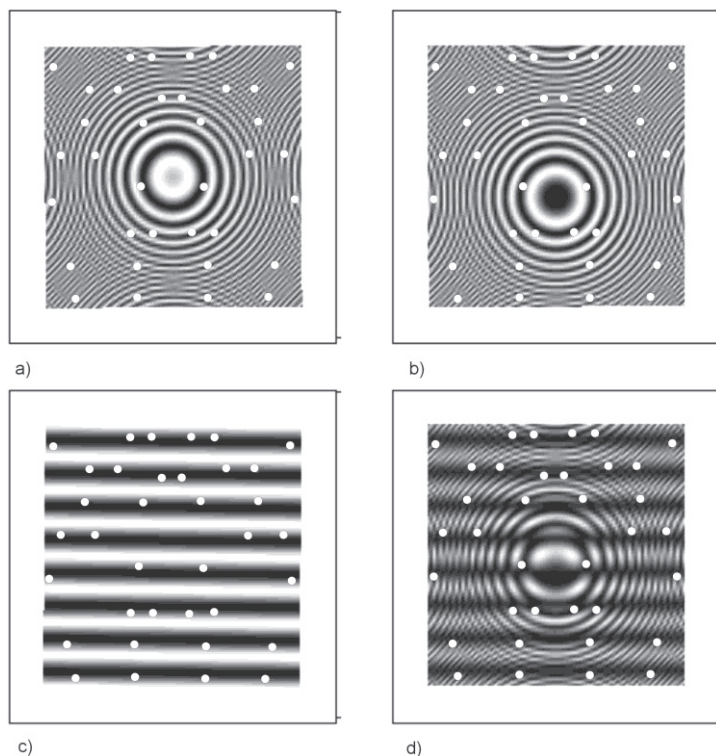


Rys. 27. Wynik interferencji trzech fal płaskich UA , UB i UC – sieć sześciokątów:
 a – prążki interferencyjne utworzone przez fale UA i UB , b – prążki interferencyjne utworzone przez fale UA i UC , c – prążki interferencyjne utworzone przez fale UB i UC ; d – obraz interferencyjny utworzony przez fale UA , UB i UC . Białe kółka wskazują położenia punktów wirowych (na osiach x i y wartości odległości zostały pominięte)

Fig. 27. Result of interference of three plane waves UA , UB and UC – hexagonal lattice:
 a – interference fringes generated by waves UA and UB , b – interference fringes generated by waves UA and UC , c – interference fringes generated by waves UB and UC , d – interference pattern generated by waves UA , UB and UC . Points indicate locations of vortex points (scales of axes x and y where omitted)

2.5.4. INTERFERENCJA DWÓCH FAL PŁASKICH Z FALĄ SFERYCZNĄ

W wyniku interferencji dwóch fal płaskich UA , UB z falą sferyczną US w płaszczyźnie obserwacji powstaje również sieć wirów optycznych. O odległościach między położeniami sąsiednich punktów wirowych decydują kierunki propagacji interferujących fal oraz wartości ich amplitud. Dalej przedstawiono przykład interferencji trzech fal UA , UB i US (dwóch fal płaskich i jednej sferycznej).



Rys. 28. Wynik interferencji dwóch fal płaskich UA , UB i fali sferycznej US – przykład I:
 a – prążki interferencyjne utworzone przez fale UA i US , b – prążki interferencyjne utworzone przez fale UB i US , c – prążki interferencyjne utworzone przez fale UA i UB , d – prążki interferencyjne utworzone przez fale UA , UB i US . Białe kółka wskazują położenia wirów optycznych
 Fig. 28. Result of interference of two plane waves UA , UB and spherical wave US – The I example:
 a – interference fringes generated by waves UA and US , b – interference fringes generated by waves UB and US , c – interference fringes generated by waves UA and UB , d – interference pattern generated by waves UA , UB and US . Points indicate locations of vortex points (scales of axes x and y where omitted)

Na rysunku 28 przedstawiono wyniki interferencji uzyskane z symulacji numerycznych, kolejno fali płaskiej z falą sferyczną I_{AS} (rys. 28a) następnej fali płaskiej z tą samą falą sferyczną I_{BS} (rys. 28b); dwóch fal płaskich I_{AB} (rys. 28c); trzech fal I_{ABS}

(rys. 28d). Wartości amplitud wszystkich trzech fal, w płaszczyźnie obserwacji, były zbliżone do siebie. Tak jak poprzednio czerwone punkty oznaczają położenia wirów optycznych. Na poprzednim rysunku widać, że wiry tworzą sieci różnych wzorów, ale nie tak prostych jak w przypadku interferencji trzech fal płaskich. W zasadzie trudno jest przewidzieć rozmieszczenie wirów optycznych bez wcześniejszych obliczeń, dlatego ten rodzaj sieci nie został wykorzystany w pomiarach testujących działanie interferometru na wirach optycznych.

3. METODY LOKALIZACJI WIRÓW OPTYCZNYCH

Interferometr na wirach optycznych wytwarza sieć wirów optycznych, które służą za punkty charakterystyczne w polu interferencyjnym. Aby można było je w tym celu wykorzystać, konieczne są precyzyjne metody lokalizacji położenia punktów wiru. Chcę podkreślić, że dokładność lokalizacji punktu wiru ma absolutnie kluczowe znaczenie dla rozdzielczości interferometru na wirach optycznych. W literaturze można znaleźć kilka teoretycznych metod lokalizacji wirów optycznych [4], [34], [75] oraz doświadczalne metody lokalizacji wirów optycznych [54], [55], [77], [78]. W punktach, w których występują wiry optyczne, obserwujemy minimum natężenia światła. Analiza obszarów minimalnej wartości natężenia światła jest trudna, ponieważ obszary te są najbardziej wrażliwe na szумы. Aby uniknąć przekłamań w oznaczeniu położenia wirów optycznych, omówione poniżej procedury lokalizacyjne bazują na właściwościach otoczenia punktów wiru.

Dalej zastało przedstawionych sześć metod lokalizacji wirów optycznych:

- metoda minimów,
- metoda amplitud,
- metoda trójkątów,
- metoda cosinusów,
- metoda wektorów,
- metoda widelców z przesunięciem fazy.

Metody amplitud, trójkątów, cosinusów i wektorów wymagają zarejestrowania czterech interferogramów (metody wieloekspozycyjne): I_{ABC} , I_{AB} , I_{AC} i I_{BC} oraz wstępnego oszacowania położenia wirów optycznych. Podczas szacowania położenia wirów optycznych wykorzystywana jest metoda minimów (metoda jednoekspozycyjna). Metoda widelców z przesunięciem fazy wymaga zarejestrowania kilkunastu interferogramów.

3.1. METODA MINIMÓW

Ponieważ na brzegach zarejestrowanych obrazów mogą znajdować się zniekształcenia obrazu, obszary te zostają pominięte w dalszej analizie. Pominięcie obszarów

przy krawędzi obrazu może być realizowane na etapie wstępnej obróbki lub w dalszych etapach przez pominięcie zlokalizowanych punktów w okolicach brzegów analizowanego obrazu. Kolejnym krokiem jest określenie i usunięcie tła. Z przeprowadzonych testów wynika, że poziom tła w zarejestrowanych obrazach interferencyjnych w wykorzystanym układzie zależy głównie od obszaru rejestrowanego względem apertury wiązek. Jednakże poziom tła nie jest jednorodny w obszarze analizowanym. W celu określenia poziomu tła badany obszar podzielony został na mniejsze obszary, w których określone było minimum natężenia odpowiadające najmniejszemu minimum lokalnemu w danym obszarze (dane z interferogramu I_{ABC}). Ponieważ w punktach wirowych natężenie powinno wynosić zero, dlatego różnica określana jest jako poziom tła. Obszar matrycy CCD, kamery wykorzystywanej w przeprowadzonych pomiarach, o wymiarach $7,2 \times 5,4$ mm, podzielony został na kilkanaście podobszarów potrzebnych do określenia poziomu tła. Tak uzyskanych kilkanaście wartości służyły jako dane wejściowe do interpolacji rozkładu natężenia promieniowania tła na cały obszar badany. Określony poziom tła następnie był odejmowany od wartości natężeń odczytanych ze wszystkich czterech zarejestrowanych interferogramów (I_{AB} , I_{BC} , I_{AC} , I_{ABC}). Tak przygotowane obrazy interferencyjne dwuwiązkowe stanowią dane wejściowe dla dokładniejszych metod lokalizacji opisanych dalej. Natomiast interferogram I_{ABC} poddawany zostaje dalszej analizie.

Właściwa procedura lokalizacji minimów podzielona jest na etapy. W pierwszym etapie następuje odseparowanie od siebie wzajemnie obszarów zawierających minima lokalne. W tym celu wykonywana jest binaryzacja obrazu z określonym poziomem granicznym, punktom o natężeniu wcześniej określonego poziomu przypisywane jest 0, a punktom o natężeniu dalej określonego poziomu wartość 1. Poziom graniczny jest każdorazowo dobierany automatycznie tak, aby ilość obszarów zawierających minimum nazwanych wyspami była jak największa i jednocześnie obszary te były jak najmniejsze (ze względu na szybkość wykonywania dalszych operacji). Ze względu na możliwe zakłócenia obrazu do określenia poziomu granicznego nie są uwzględniane wyspy o bardzo małej powierzchni. W przeprowadzonych pomiarach powierzchnia ta odpowiadała 10 pikselom matrycy. Poziom graniczny powinien być dobierany każdorazowo do zarejestrowanego obrazu, ponieważ ustalenie stałego poziomu może powodować, że obszary zawierające minima nie zostaną rozdzielone w zależności od natężeń interferujących wiązek lub poziomu tła. W przypadku wykonywania serii pomiarów w zbliżonych warunkach poziom graniczny nie zmieni się i może być określony jednorazowo. Wrażliwość na wartość poziomu granicznego wynika bezpośrednio z rozkładu natężenia w okolicy punktu wirowego.

W następnym etapie lokalizacji w obszarach interferogramu I_{ABC} określonych przez wyspy określany jest pojedynczy punkt (w każdym z obszarów), w którym natężenie jest minimalne.

W grupie tak wyznaczonych minimów mogą znajdować się minima nie odpowiadające położeniom punktów wirowych. Dodatkowo występujące minima natężeń

nia są pomijane. O tym, które minima wysp zostaną pominięte decyduje ich odległość od sąsiednich. Ponieważ struktura sieci wirów optycznych powstałych przy interferencji trzech fal płaskich jest określona, nie mogą wystąpić wiry bardzo blisko siebie. W celu określenia odległości, w której wiry w takiej sieci nie mogą wystąpić, określony został parametr:

$$odl = k \sqrt{\frac{P}{nw}}, \quad (6)$$

gdzie:

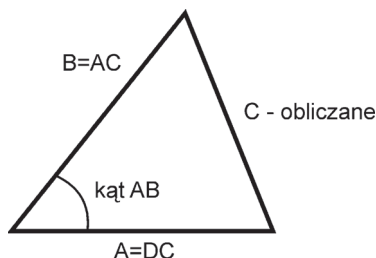
- P – powierzchnia analizowanego interferogramu,
- nw – liczba wszystkich wirów,
- k – współczynnik.

Parametr ten został wprowadzony ze względu na występujące zakłócenia w obrazie interferencyjnym powodujące w procesie analizy wydzielenie kilku wysp w bezpośrednim sąsiedztwie jednego punktu wirowego i jego wprowadzenie wyeliminowało problem interpretacji takiego obszaru jako zawierającego kilka punktów wirowych. Wartość tego parametru zależy od gęstości sieci wirów i dzięki temu jedna procedura może być wykorzystywana bez modyfikacji do analizy obrazów z różną liczbą wirów. Współczynnik k dobierany jest w zależności od stopnia deformacji sieci wirów oraz od jakości analizowanego obrazu. Większa dopuszczalna deformacja sieci wirów wymusza mniejszą wartość współczynnika k , co jest możliwe pod warunkiem wystarczającej jakości obrazu. Im więcej zakłóceń w obrazie, tym większy współczynnik k . W przeprowadzonych testach metod lokalizacji wynosił 0,3.

Testy numeryczne metody minimów przeprowadzane na teoretycznie wygenerowanych i w różny sposób zaburzonych falach płaskich wykazały, że zlokalizowane położenia wirów optycznych różnią się od rzeczywistych o mniej niż 3 piksele ($< 27 \mu\text{m}$) [55].

3.2. METODA AMPLITUD

W pierwszym kroku zastosowania metody amplitud następuje przeskalowanie wartości natężeń odczytanych z obrazów interwencyjnych I_{AB} , I_{AC} oraz I_{BC} w taki sposób, aby wartość minimalna wynosiła -1 , a maksymalna 1 . W ten sposób otrzymywane są wartości cosinusa fazy względnej pomiędzy falami U_{AB} , U_{AC} i U_{BC} . Następnie wydzielane są podmacierze, które w środku zawierają punkty wirowe, wyznaczone na podstawie lokalizacji metodą minimów. Kolejny krok to konstruowanie trójkątów z amplitud interferujących fal (z nieznormalizowanych interferogramów) i cosinusa fazy względnej (ze znormalizowanych interferogramów od -1 do 1). Aby wyjaśnić, w jaki sposób odbywa się konstruowanie trójkątów z amplitud i cosinusa fazy względnej, posłużono się przykładem:



Rys. 29. Przykład konstrukcji trójkąta z amplitud i cosinusa fazy względnej interferujących fal
 Fig. 29. Example of triangle formed from amplitudes and relative phase of interfering waves

Z interferogramów I_{AC} i I_{BC} odczytywane są długości boków B i A oraz cosinus kąta pomiędzy falami UA i UB (rys. 29). Z twierdzenia cosinusów obliczana jest długość boku C

$$C = \sqrt{I_{BC} + I_{AC} - 2\sqrt{I_{BC}I_{AC}} \cos(\angle AB)}. \quad (7)$$

Następnie wyznaczana jest różnica w wartościach pomiędzy amplitudą obliczoną C a wartością odczytaną z interferogramu I_{AB} , czyli $|C - AB|$ (rys. 29: $C = AB$). W taki sam sposób wyznaczane są różnice wartości amplitud $|B - AC|$ oraz $|A - BC|$. W kolejnym kroku obliczana jest suma s

$$s = \left| \frac{C - AB}{AB} \right| + \left| \frac{B - AC}{AC} \right| + \left| \frac{A - BC}{BC} \right|. \quad (8)$$

Suma przedstawiona równaniem (8) obliczana jest dla każdego punktu podmacierzy zawierającej punkt podejrzany w środku. Ostatni krok polega na znalezieniu najmniejszej wartości sumy (równ. (8)), która wskazuje położenie wiru optycznego.

Procedura jest powtarzana dla każdego punktu wyznaczonego wcześniej metodą minimów. W ten sposób lokalizowane są kolejno wszystkie punkty wirowe z całego obrazu interferencyjnego. Podobszary wokół punktu „podejrzanego” wydzielane są dla przyspieszenia wykonania wszystkich obliczeń. Możliwe jest wykorzystanie metody amplitud bez wcześniejszej lokalizacji punktów „podejrzanych” metodą minimów. Takie rozwiązanie jest jednak bardzo czasochłonne, ponieważ wykonanie wymaganych obliczeń dla całego obszaru interferogramów wymaga wielokrotnie więcej czasu niż zastosowanie wstępnej lokalizacji, a następnie lokalizacji metodą minimów już tylko w określonych podobszarach. Ma to szczególne znaczenie w zastosowaniu interpolacji, ponieważ wtedy liczba obliczeń znacznie wzrasta.

3.3. METODA TRÓJKĄTÓW

W pierwszym kroku metody trójkątów, podobnie jak w metodzie amplitud, następuje przeskalowanie wartości natężeń zarejestrowanych na interferogramach I_{AB} , I_{AC} oraz I_{BC} . Przeskalowanie odbywa się w taki sposób, aby wartości minimalne natężeń odpowiadały wartości 0, a wartości maksymalne odpowiadały wartości 255. Przeskalowanie to odbywa się dla każdego z wymienionych interferogramów oddzielnie. Następnie, tak jak w poprzednio opisanej metodzie, wydzielane są podobszary zawierające punkty wirowe na podstawie wyników lokalizacji wstępnej.

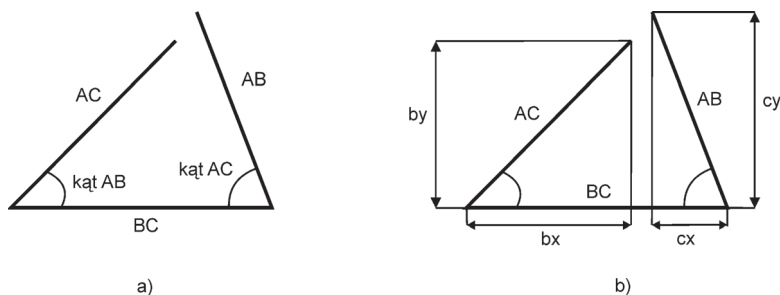
W kolejnym kroku obliczana jest różnica natężeń w każdym analizowanym punkcie podobszaru według zależności (9). Różnice te są wynikiem porównania długości boków trójkąta, który ze względu na przeskalowanie wykonane na początku powinien być trójkątem równobocznym w punkcie wirowym.

$$\begin{aligned} |AB - BC| &= T1, \\ |AB - AC| &= T2, \\ |AC - BC| &= T3. \end{aligned} \tag{9}$$

Powstałe w ten sposób macierze $T1$, $T2$, $T3$ stanowią dane wejściowe do następnego kroku opisywanej metody. Dla każdej współrzędnej z macierzy $T1$, $T2$, $T3$ wybierana jest wartość największa spośród wartości w tych trzech macierzach i zapisywana do nowej tablicy wynikowej W . Położenie minimalnej wartości w macierzy W wskazuje położenie punktu wirowego w analizowanym podobszarze. Podane operacje są powtarzane dla kolejnych podobszarów w celu określenia położenia kolejnych punktów wirowych.

3.4. METODA COSINUSÓW

W metodzie cosinusów również w pierwszym kroku następuje przeskalowanie wartości natężeń interferogramów I_{AB} , I_{AC} oraz I_{BC} . Tak jak w przypadku wcześniej opisanej metody amplitud przeskalowanie jest takie, aby wartość minimalna wynosiła -1 , a maksymalna 1 , w celu uzyskania wartości cosinusa fazy względnej pomiędzy falami odpowiednio UAB , UAC i UBC . Następnie wydzielane są podobszary zawierające punkty wirowe na podstawie wyników lokalizacji wstępnej w ten sam sposób co w poprzednio opisanych metodach. W kolejnym kroku wybierany jest jeden z boków trójkąta za podstawę rzutów, np. bok A (BC) (rys. 30). Odcinki AB , BC , AC odpowiadają odpowiednio wartościom natężeń interferogramów I_{AB} , I_{BC} oraz I_{CA} , a wartości cosinusów kątów AB , BC , AC wartościom z interferogramów I_{AB} , I_{BC} oraz I_{CA} po przeskalowaniu.



Rys. 30. Przykład konstrukcji trójkąta z amplitud i cosinusów fazy względnej interferujących fal (a) oraz rzutów dwóch boków trójkąta na trzeci bok tego trójkąta (b)

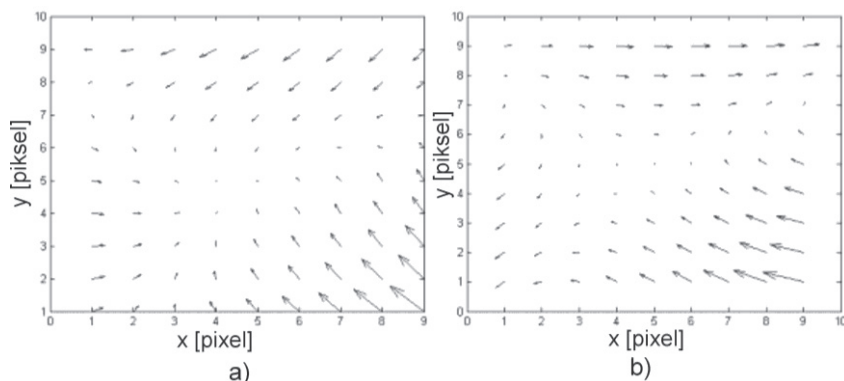
Fig. 30. Example of triangle formed from amplitudes and relative phase of interfering waves (a) and two sides projection on third side of the triangle (b)

Na wybrany bok rzutowane są długości pozostałych boków „trójkąta” (figura z trzech odcinków, niedomknięta), co daje w wyniku wartości bx oraz cx . Następnie wartości by oraz cy są wyznaczane przez rzut boków AB i AC na oś prostopadłą do wybranego boku (rys. 30b). Następnie wyznaczane są wartości Δxa i Δya z zależności

$$\Delta xa = BC - (bx + cx),$$

$$\Delta ya = by - cy. \quad (10)$$

Wartości Δxa i Δya obliczane są dla wszystkich punktów podmacierzy zawierających punkty podejrzane w środku. Wartości te są współrzędnymi wektora zwanego wektorem pseudoamplitudy (\mathbf{paA}). W kolejnych krokach należy obrać za podstawę rzutów kolejne dwa boki trójkąta i powtórzyć czynności doprowadzające do otrzymania kolejnych współrzędnych wektorów pseudoamplitudy (\mathbf{paB} oraz \mathbf{paC}). Na rysunku 31 pokazano przykład rozkładu wektorów pseudoamplitudy w otoczeniu wiru optycznego.

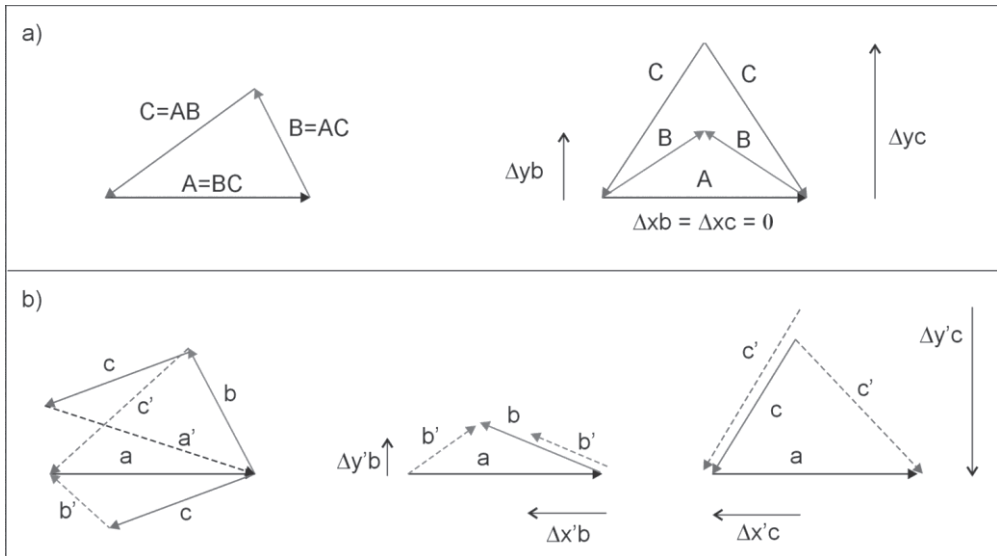


Rys. 31. Ilustracja przykładowych rozkładów wektorów pseudoamplitudy w otoczeniu wiru optycznego
Fig. 31. Example of pseudo-amplitude vector distributions in neighbourhood of optical vortex

Następnie należy przeanalizować zmiany kierunku wektora pseudoamplitudy w otoczeniu punktu podejrzanego osobno dla każdego bazowego boku trójkąta. Wir optyczny znajduje się w miejscu, wokół którego zmiany kierunku wektora pseudoamplitudy są największe. Dla każdego wiru otrzymujemy trzy współrzędne położenia (rzuty na trzy boki trójkąta), które, po uśrednieniu, wskazują położenie wiru optycznego.

3.5. METODA WEKTORÓW

W pierwszym kroku w metodzie wektorów wyznaczane są amplitudy fal UA , UB i UC na podstawie wartości natężeń interferogramów I_{AB} , I_{AC} oraz I_{BC} . Następnie wydzielane są podobszary zawierające punkty wirowe na podstawie wyników lokalizacji wstępnej. W kolejnym kroku należy wybrać jeden z boków trójkąta za podstawę nowego trójkąta, np. bok A (rys. 32a) i skonstruować dwa trójkąty o bokach B, A, B oraz C, A, C . Na rysunku 32a dużymi literami oznaczona jest wartość amplitudy odczytanej w punktach wirowych, natomiast małymi literami oznaczone są wartości amplitud odczytane w pobliżu punktów, w których znajduje się wir optyczny (rys. 32b). W punktach znajdujących się w sąsiedztwie punktów wirowych dwa trójkąty zbudowane są z boków b', a, b oraz c', a, c .



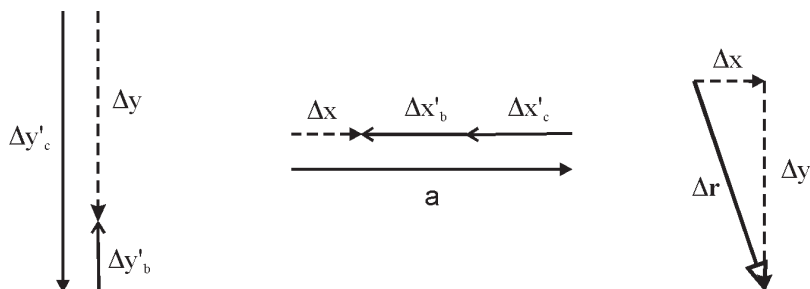
Rys. 32. Przykład konstrukcji trójkątów z amplitud interferujących fal oraz rzutów wektora pseudofazy:

a) amplitudy w punkcie wirowym, b) amplitudy w pobliżu punktu wirowego

Fig. 32. Example of the triangle formed from amplitudes and projections of the pseudo-phase vector:

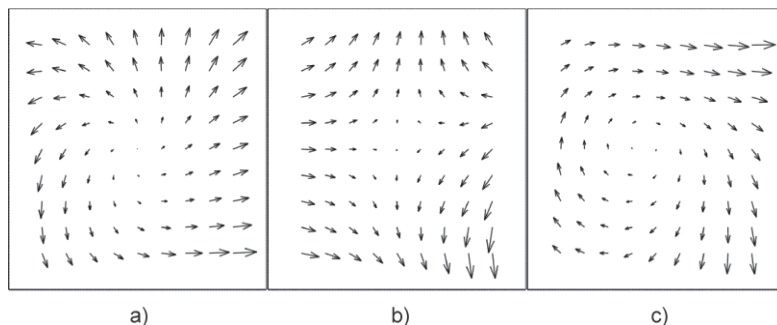
a) amplitudes in the vortex point, b) amplitudes near the vortex point

Na rysunku 32 przedstawione są przykładowe konstrukcje odcinków Δy_b , Δy_c , Δx_b , Δx_c w punkcie nieciągłości oraz $\Delta y'_b$, $\Delta y'_c$, $\Delta x'_b$, $\Delta x'_c$ w sąsiedztwie wiru optycznego. Odcinki te służą do wyznaczenia składowych wektora, który będzie nazywany na użytek tej pracy wektorem pseudofazy ($\Delta \mathbf{r}$). Na rysunku 33 przedstawiono zasadę konstrukcji wektora pseudofazy z odcinków $\Delta y'_b$, $\Delta y'_c$, $\Delta x'_b$, $\Delta x'_c$ wyznaczonych na bazie boku A .



Rys. 33. Przykład konstrukcji wektora pseudofazy wykorzystywanej w metodzie wektorów
Fig. 33. The example of the pseudo-phase vector building using in vectors method

W kolejnych krokach należy obrać za podstawę rzutów kolejne dwa boki trójkąta i powtórzyć czynności doprowadzające do otrzymania kolejnych współrzędnych wektorów pseudofazy. Na rysunku 34 pokazano przykładowe rozkłady wektorów pseudofazy w otoczeniu wiru optycznego.

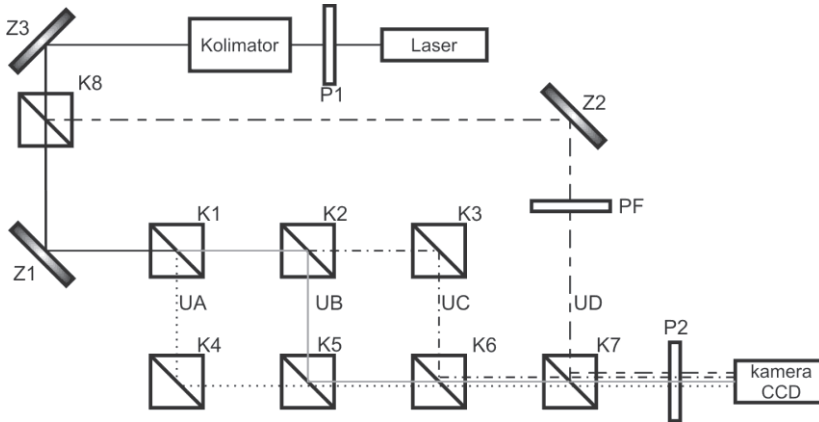


Rys. 34. Ilustracja przykładowych rozkładów wektorów pseudofazy w otoczeniu wiru optycznego
Fig. 34. Example of the pseudo-phase vector distributions in neighbourhood of the vortex point

Następnie należy przeanalizować zmiany kierunku wektora pseudofazy w otoczeniu punktu podejrzanego osobno dla każdego bazowego boku trójkąta. Wir optyczny znajduje się w miejscu, wokół którego zmiany kierunku wektora pseudofazy są największe. Dla każdego wiru otrzymujemy trzy współrzędne położenia (konstrukcje dla trzech boków trójkąta), które, po uśrednieniu, wskazują położenie punktu wirowego.

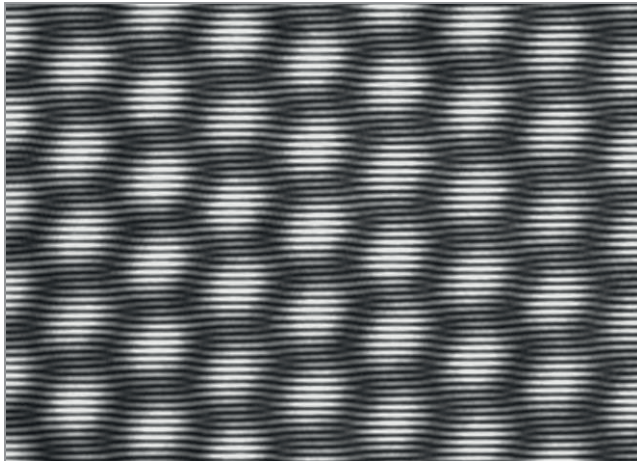
3.6. METODA WIDELCÓW Z PRZESUNIĘCIEM FAZY

Metoda widelców z przesunięciem fazy wymaga zarejestrowania serii obrazów interferencyjnych czterech fal, z czego trzy (UA , UB i UC) mają zbliżone wartości amplitud, czwarta (UD) natomiast ma amplitudę w przybliżeniu równą sumie trzech $UA + UB + UC$. Kąt pomiędzy tą wiązką a każdą z trzech pozostałych jest większy od kąta pomiędzy dowolną z UA , UB , UC . Układ pomiarowy przedstawiony jest na rysunku (rys. 35).



Rys. 35. Układ pomiarowy do rejestracji obrazów interferencji czterech wiązek z możliwością przesunięcia fazy jednej z nich

Fig. 35. Measurement setup to register interference pattern of four waves with possibility of phase shift one of them



Rys. 36. Przykładowy obraz interferencyjny, interferencji czterech wiązek (eksperyment)

Fig. 36. Example of interference pattern of four waves (experiment)

Fala UD w układzie tym przechodzi przez sterowany cyfrowo przesuwnik fazy (PF), co umożliwia rejestrację serii obrazów z krokowo zmieniającą się wartością fazy fali UD . Przykładowy obraz interferencyjny przedstawiony jest na rysunku 36.

Charakterystyczną strukturą prążków w przykładowym obrazie jest rozwidlenie zwane również „widelcem”. Rozwidlenie powstaje w wyniku interferencji fali przenoszącej wir optyczny z falą płaską. Powstanie rozwidlenie w wyniku takiej interferencji oraz cechy powstałej struktury opisane są dokładniej w rozdziale poświęconym określeniu znaku ładunku topologicznego wiru optycznego (punkt 5).

W pierwszym etapie metody lokalizacji punktów wirowych jest obliczany rozkład zmian natężenia na zarejestrowanym obszarze spowodowany przesunięciami fazy fali czwartej UD .

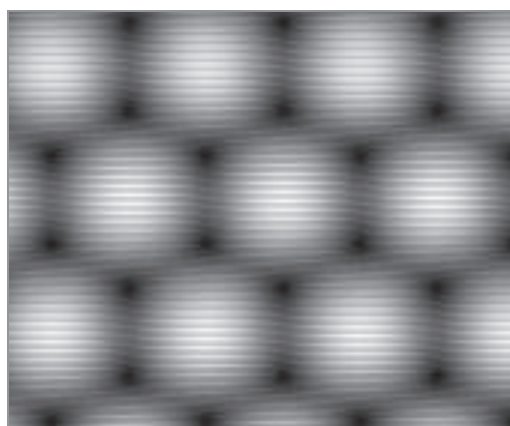
$$\Delta I = |I_2 - I_1| + |I_3 - I_2| + \dots + |I_n - I_{n-1}|, \quad (11)$$

gdzie:

I_k – k -ty rozkład natężenia,

n – liczba zarejestrowanych obrazów.

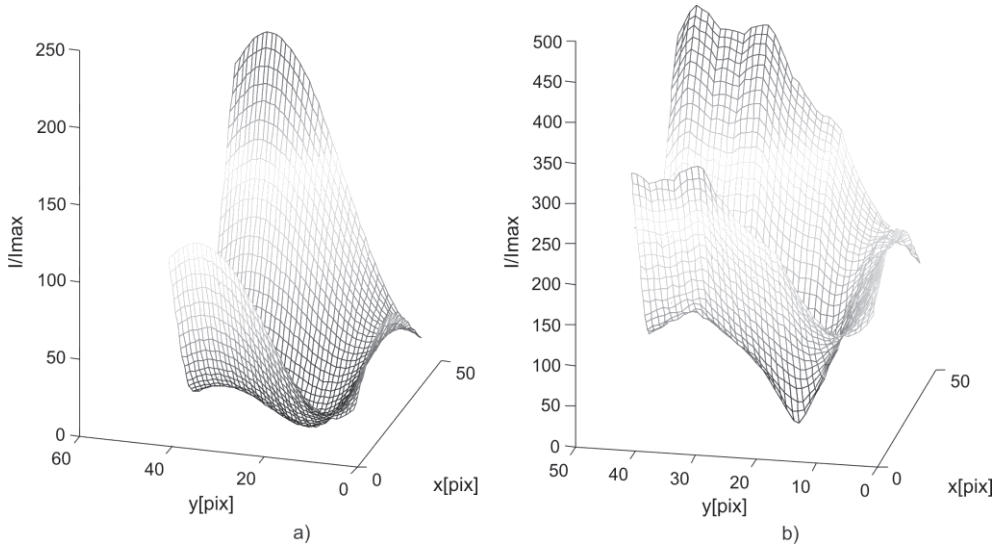
Ponieważ zmianie w kolejnych obrazach serii podlega faza fali UD , zaobserwować można przesunięcie prążków interferencyjnych w całym obszarze. Jednakże położenie „widelców” jest niezmiennie, ponieważ ich lokalizacja jest ściśle związana z położeniem punktów wirowych. Ponieważ w punktach wirowych natężenie światła rejestrowane przez detektor jest równe natężeniu fali UD i nie zależy od pozostałych trzech wiązek UA , UB , UC ($UA + UB + UC = 0$ w punkcie wirowym), w tych punktach zmiany natężenia nie występują. Wynikiem pierwszego etapu działania opisywanej metody jest obraz, w którym występują zaciemnienia odpowiadające położeniu punktów wirowych (rys. 37).



Rys. 37. Rozkład zmian natężenia przy zmianie fazy fali UD

Fig. 37. Distribution of intensity changing with the phase shifting of wave UD

Dalej działanie metody widelców z przesunięciem fazy jest analogiczne do działania metody minimów z tą różnicą, że minima lokalne są lepiej określone. Obszary ciemne są mniejsze i nachylenie powierzchni w okolicy punktów wirowych jest większe niż przy wykorzystaniu pojedynczego obrazu interferencyjnego trzech wiązek (rys. 38).



Rys. 38. a – rozkład natężenia w sąsiedztwie minimum lokalnego w obrazie interferencyjnym trzech wiązek, b – rozkład zmian natężenia policzony dla tego samego obszaru.

Przesunięcie fazy $\pi/9$, liczba obrazów analizowanych: 18 (rozkłady uzyskane w wyniku symulacji)

Fig. 38. a – distribution of the intensity in neighbourhood of local minimum of interference pattern of three waves, b – distribution of the intensity changing calculated for the same area. Phase shift step is $\pi/9$ and quantity of analysed images is 18 (distribution have been obtained by simulation)

4. ANALIZA DOKŁADNOŚCI LOKALIZACJI WIRÓW OPTYCZNYCH Z UŻYCIEM METOD CZTEROEKSPOZYCYJNYCH

W celu oszacowania błędów wynikających z algorytmów numerycznych procedur lokalizujących położenia wirów optycznych, zostały przeprowadzone symulacje numeryczne, a następnie wyniki symulacji zostały porównane z wynikami pomiarów.

4.1. OBLICZENIA TEORETYCZNE

W symulacjach numerycznych interferogramy I_{AB} , I_{AC} , I_{BC} oraz I_{ABC} były generowane z równań falowych fal płaskich, przy czym kierunki rozchodzenia się fal płaskich

były dobierane w taki sposób, aby w płaszczyźnie elementu CCD (576×768 pikseli o rozmiarach $9 \mu\text{m}$ na $9 \mu\text{m}$) otrzymać prążki interferencyjne o żądanej gęstości i nachyleniu. Tak wygenerowane interferogramy służyły do testowania metod lokalizacji.

Testy zostały przeprowadzone z użyciem następujących fal płaskich:

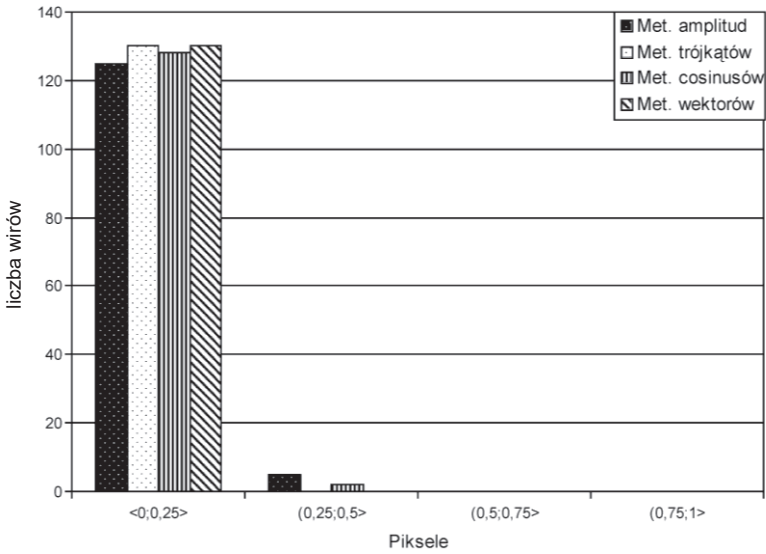
- a) Trzy fale płaskie o jednakowych amplitudach.
- b) Trzy fale płaskie o różnych amplitudach $A = 0,95$, $B = 1,05$, $C = 1,00$.
- c) Trzy fale płaskie o jednakowych amplitudach zaburzone innymi falami.
- d) Trzy fale płaskie o różnych amplitudach $A = 0,95$, $B = 1,05$, $C = 1,00$ zaburzone innymi falami.
- e) Dwie fale płaskie oraz jedna fala płaska zaburzona fazowo. Amplitudy wszystkich fal były jednakowe.

We wszystkich wymienionych kombinacjach było generowanych kilka przypadków z różnymi gęstościami prążków wytworzonych z interferujących fal UB i UC (od 8,5 do 18 prążków mieszczących się na matrycy CCD, co około jeden prążek), czyli grubości pojedynczych prążków były następujące: 90,4 piksela (8,5 prążków pionowych; $768/8,5 = 90,4$), 73,1 piksela, 66,8 piksela, 56,9 piksela, 53,0 piksela, 50,0 piksela, 46,5 piksela, 42,7 piksela (18 prążków pionowych). Wraz ze wzrostem gęstości obserwowanych prążków interferencyjnych rośnie liczba wirów optycznych. Im więcej wytworzonych jest wirów, tym dłużej działają procedury lokalizacji, natomiast zbyt mała liczba punktów nieciągłości fazy może okazać się niewystarczająca do przeprowadzenia części pomiarów (np. pomiar jakości płytek płaskorównoległych). Na etapie testowania nowych metod lokalizacji ustalone zostało, że 300–600 wirów optycznych jest wystarczającą ilością potrzebną do większości zastosowań interferometru na wirach optycznych, dlatego też metody lokalizacyjne były testowane na interferogramach, na których obserwowano od 88 wirów optycznych (8,5 prążków pionowych) do 414 wirów optycznych (18 prążków pionowych). Metoda trójkątów była dodatkowo przetestowana aż do około 1800 wirów optycznych (37 prążków pionowych).

Metody lokalizacji były testowane dla sieci sześciokątów oraz par wirów optycznych, a ponieważ nie zaobserwowano różnic w działaniu procedur w zależności od typu sieci, tylko od odległości między wirami, dalsze przykłady przedstawiono, wykorzystując sieć sześciokątów.

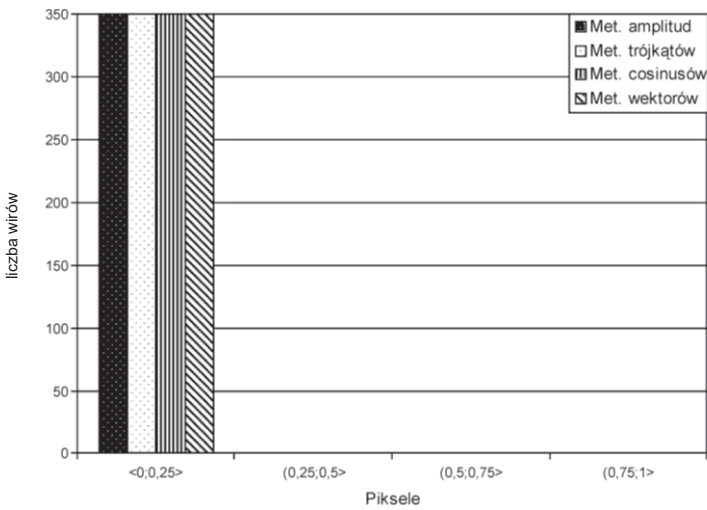
4.1.1. HISTOGRAMY – TRZY FALE PŁASKIE O JEDNAKOWYCH AMPLITUDACH

W trakcie tworzenia histogramów rzeczywiste położenia wirów optycznych wyznaczone były z użyciem metody opisanej w punkcie 1.2, polegającej na znalezieniu punktów przecięcia części rzeczywistej i urojonej funkcji U_{ABC} w zerze. Punkty te zostały wyznaczone z dokładnością $0,9 \mu\text{m}$ ($1/10$ piksela – interpolacja dziesięciokrotna), natomiast położenia wirów optycznych, które znajdowały metody lokalizacji, były obliczane z dokładnością $2,25 \mu\text{m}$ ($1/4$ piksela – czterokrotna interpolacja podmacierzy zawierających punkty zaobserwowane w środku). Podział piksela na mniejsze części miał na



Rys. 39. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy fale płaskie o równych amplitudach oraz gdy na interferogramie I_{BC} obserwujemy 10,5 prążków pionowych (130 wirów)

Fig. 39. Diagram of localisation methods accuracy in case where are generated three plane waves with equal amplitudes and where are 10,5 vertical fringes on interferogram I_{BC} (130 vortices)



Rys. 40. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy fale płaskie o równych amplitudach oraz gdy na interferogramie I_{BC} obserwujemy 16,5 prążków pionowych (350 wirów)

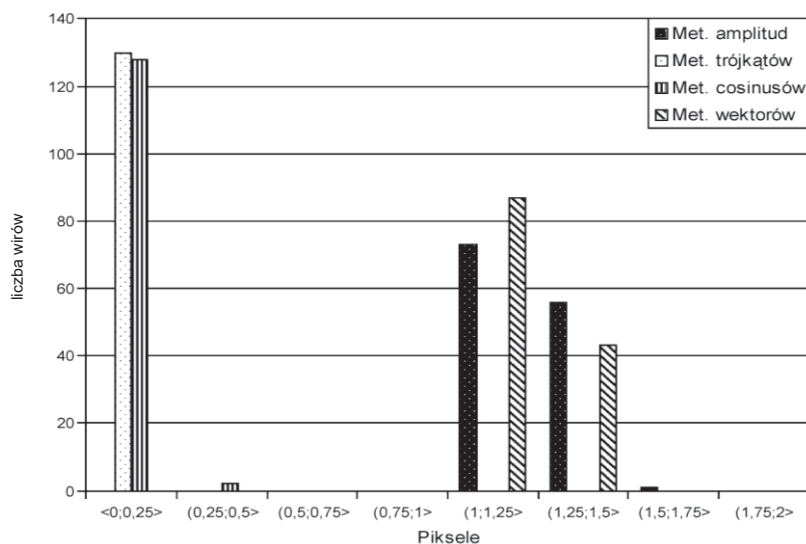
Fig. 40. Diagram of localisation methods accuracy in case where are generated three plane waves with equal amplitudes and where are 16,5 vertical fringes on interferogram I_{BC} (350 vortices)

celu ustalenie dokładności działania metod lokalizacji punktów wirowych, w przypadku użycia kamery cyfrowej o mniejszym rozmiarze piksela oraz potrzebny był do zaobserwowania niewielkich zmian w dokładności działania metod lokalizacji. Do wstępnego szacowania położenia punktów nieciągłych używana była metoda minimów, która określa położenia wirów z dokładnością rzędu $27\ \mu\text{m}$ (3 piksele). Dlatego też podmacierze, które wydzielane są w początkowych etapach działania metod lokalizacji, zawierające punkty zaobserwowane w środku, są rozmiaru 9 na 9 pikseli.

W przypadku kiedy trzy interferujące fale płaskie są idealne i mają takie same amplitudy, wszystkie metody (amplitud, trójkątów, cosinusów, wektorów) działają prawidłowo. Prawie wszystkie wiry optyczne znajdowane są z dokładnością poniżej $4,5\ \mu\text{m}$ ($\frac{1}{2}$ piksela) niezależnie od gęstości prążków. Dalej zostały przedstawione dwa przykładowe wykresy ilustrujące dokładność działania metod lokalizacji.

4.1.2. HISTOGRAMY – TRZY FALE PŁASKIE O RÓŻNYCH AMPLITUDACH

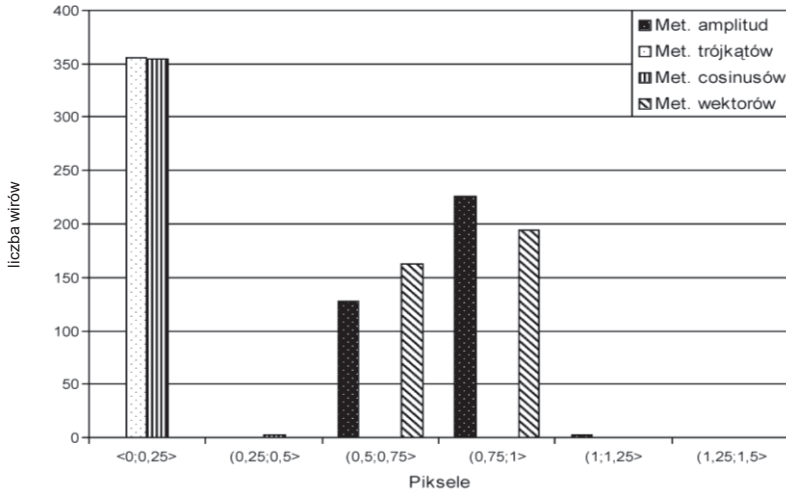
W przypadku kiedy trzy interferujące fale płaskie mają różne amplitudy, metody cosinusów i trójkątów działają prawidłowo. Prawie wszystkie wiry optyczne znajdowane są z dokładnością poniżej $4\ \mu\text{m}$, $5\ \mu\text{m}$, niezależnie od gęstości prążków. Natomiast metody amplitud i wektorów lokalizują wiry optyczne z błędem około $9\ \mu\text{m}$.



Rys. 41. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy fale płaskie o różnych amplitudach oraz gdy na interferogramie I_{BC} obserwujemy 10,5 prążków pionowych (130 wirów)

Fig. 41. Diagram of localisation methods accuracy in case where are generated three plane waves with different amplitudes and where are 10,5 vertical fringes on interferogram I_{BC} (130 vortices)

Metody te dają najlepsze wyniki, gdy ilość prążków pionowych I_{BC} jest większa od 14. Wartości amplitud dla poszczególnych fal wynoszące: $A = 0,95$, $B = 1,05$, $C = 1,00$ w przeliczeniu na natężenia mają następujące wartości: $I_a = 0,9$, $I_b = 1,1$ oraz $I_c = 1,0$. Dalej zostały przedstawione dwa przykładowe wykresy ilustrujący dokładność działania metod lokalizacji w tym przypadku.

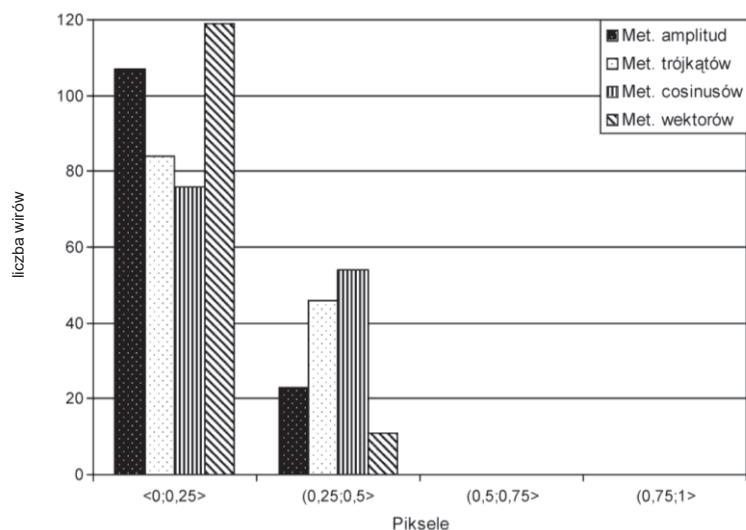


Rys. 42. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy fale płaskie o różnych amplitudach oraz gdy na interferogramie I_{BC} obserwujemy 16,5 prążków pionowych (356 wirów)

Fig. 42. Diagram of localisation methods accuracy in case where are generated three plane waves with different amplitudes and where are 16,5 vertical fringes on interferogram I_{BC} (356 vortices)

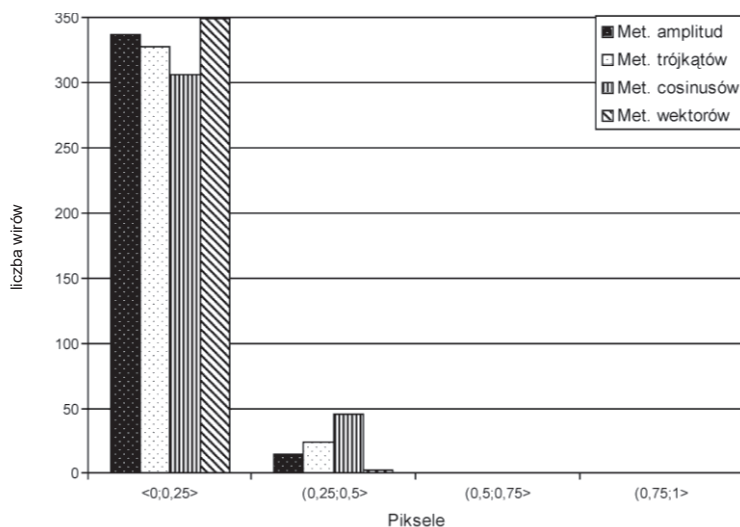
4.1.3. HISTOGRAMY – TRZY FALE PŁASKIE ZABURZONE O JEDNAKOWYCH AMPLITUDACH

W pomiarach eksperymentalnych nie mierzymy parametrów idealnych fal płaskich. W rzeczywistości zawsze występują zakłócenia. Aby interferogramy wygenerowane numerycznie były podobne do tych obserwowanych podczas pomiarów, do fal płaskich zostały dodane inne fale o amplitudach nie przekraczających 5% wartości amplitudy fali nośnej, czyli tym razem $UA = 1 \exp(i\mathbf{k}\mathbf{r} + \varphi) + Az \exp(i\mathbf{k}\mathbf{r}_2)$, $UB = 1 \exp(i\mathbf{k}\mathbf{r} + \varphi) + Bz \exp(i\mathbf{k}\mathbf{r}_2)$ oraz $UC = 1 \exp(i\mathbf{k}\mathbf{r} + \varphi) + Cz \exp(i\mathbf{k}\mathbf{r}_2)$ przy czym amplitudy Az , Bz , Cz były mniejsze niż 0,05. Zbyt duże wartości amplitudy fal zaburzających mogą spowodować deformację sieci wirów optycznych, co nie było obserwowane na interferogramach eksperymentalnych. Podczas prób różnych generacji numerycznych były dodawane małe zaburzenia losowe, ale nie zostały zaobserwowane zmiany w dokładności lokalizowania wirów optycznych. Dalej zostały przedstawione dwa przykładowe wykresy ilustrujące dokładność działania metod lokalizacji.



Rys. 43. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy zaburzone fale płaskie o jednakowych amplitudach oraz gdy na interferogramie I_{BC} obserwujemy 10,5 prążków pionowych (130 wirów)

Fig. 43. Diagram of localisation methods accuracy in case where are generated three disturbed plane waves with equal amplitudes and where are 10.5 vertical fringes on interferogram I_{BC} (130 vortices)



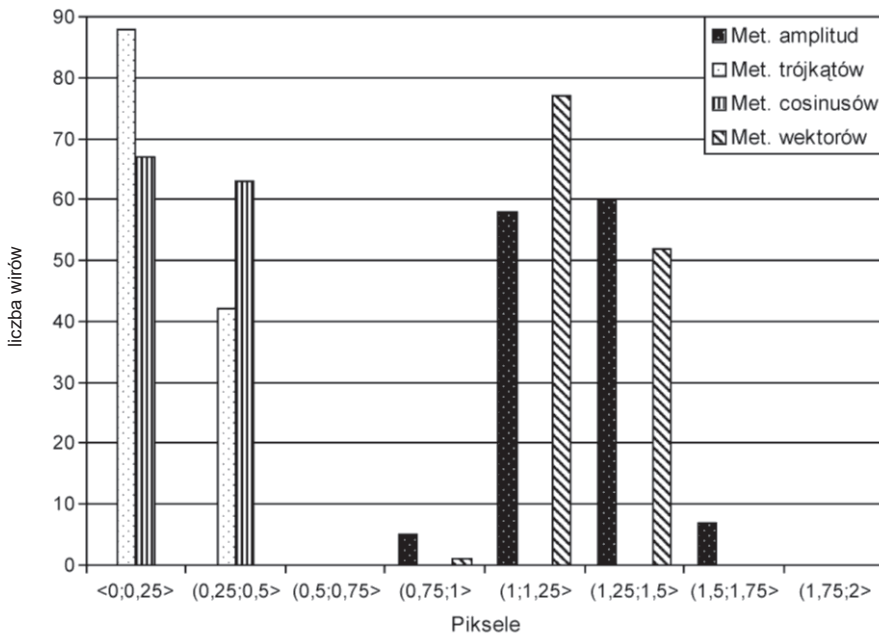
Rys. 44. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy zaburzone fale płaskie o jednakowych amplitudach oraz gdy na interferogramie I_{BC} obserwujemy 16,5 prążków pionowych (351 wirów)

Fig. 44. Diagram of localisation methods accuracy in case where are generated three disturbed plane waves with equal amplitudes and where are 16.5 vertical fringes on interferogram I_{BC} (351 vortices)

W przypadku kiedy trzy interferujące fale płaskie zaburzone innymi falami mają jednakowe amplitudy, wszystkie metody (amplitud, cosinusów, trójkątów, wektorów) działają prawidłowo. Wszystkie wiry optyczne znajdowane są z dokładnością poniżej $4,5 \mu\text{m}$ ($\frac{1}{2}$ piksela) z wyjątkiem przypadku, gdy liczba prążków pionowych I_{BC} jest mniejsza niż 9 prążków (grubość prążka wynosiła około 85 pikseli). W tym przypadku wiry optyczne znajdowane są w odległości do $6,75 \mu\text{m}$ ($\frac{3}{4}$ piksela) od położenia rzeczywistych.

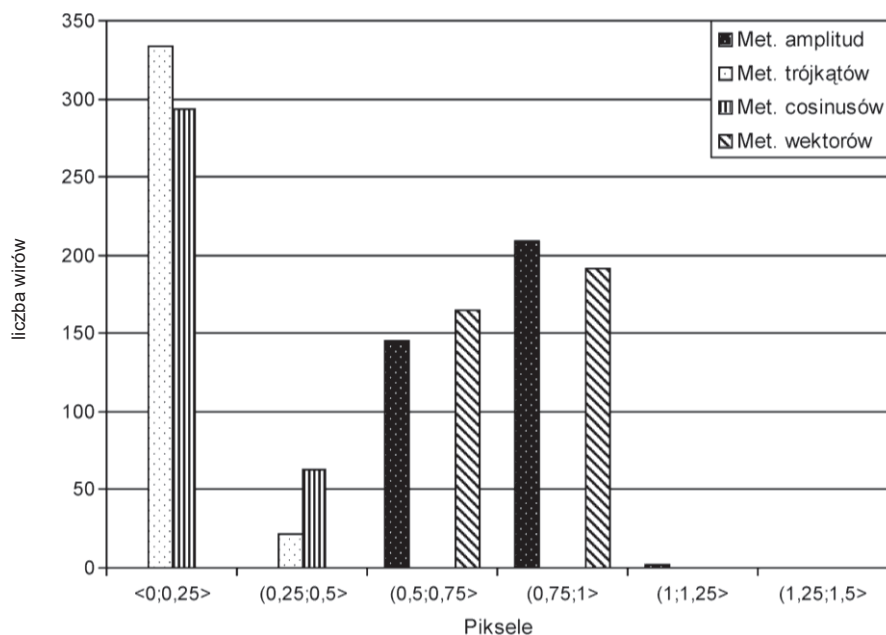
4.1.4. HISTOGRAMY – TRZY FALE PŁASKIE ZABURZONE O RÓŻNYCH AMPLITUDACH

W przypadku gdy trzy interferujące fale płaskie zaburzone innymi falami mają dodatkowo różne amplitudy, obserwujemy podobieństwa zarówno w działaniu metod amplitud i wektorów, jak i w działaniu metod trójkątów i cosinusów – występuje błąd systematyczny widoczny na histogramach przedstawionych na rysunkach 45, 46 i 47.



Rys. 45. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy zaburzone fale płaskie o różnych amplitudach oraz gdy na interferogramie I_{BC} obserwujemy 10,5 prążków pionowych (130 wirów)

Fig. 45. Diagram of localisation methods accuracy in case where are generated three disturbed plane waves with different amplitudes and where are 10,5 vertical fringes on interferogram I_{BC} (130 vortices)



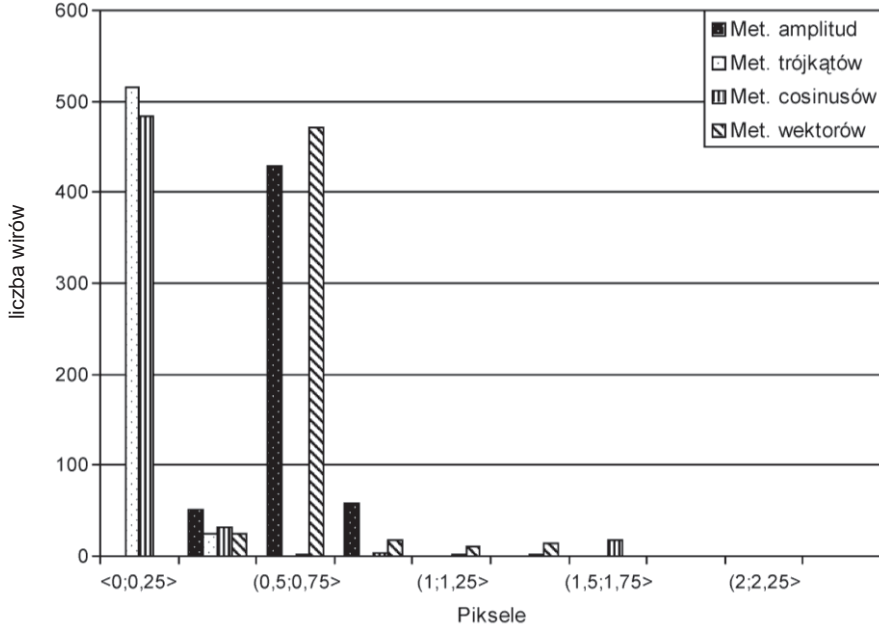
Rys. 46. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy zaburzone fale płaskie o różnych amplitudach oraz gdy na interferogramie I_{BC} obserwujemy 16,5 prążków pionowych (356 wirów)

Fig. 46. Diagram of localisation methods accuracy in case where are generated three disturbed plane waves with different amplitudes and where are 16,5 vertical fringes on interferogram I_{BC} (356 vortices)

Na rysunkach 45 i 46 widać, że wraz ze wzrostem gęstości prążków pionowych I_{BC} zmniejsza się błąd lokalizacji wirów optycznych. Metody trójkątów i cosinusów lokalizują punkty nieciągłości fazowych z dokładnością $4,5 \mu\text{m}$ ($\frac{1}{2}$ piksela), przy czym dla gęstszych prążków więcej wirów optycznych znajdujących się w odległości około $2,25 \mu\text{m}$ ($\frac{1}{4}$ piksela) od rzeczywistego położenia. Metody amplitud i wektorów działają niepoprawnie, w przypadku gdy prążki pionowe są rzadkie – praktycznie wszystkie wiry optyczne znalezione są o ponad $9 \mu\text{m}$ (1 piksel) dalej od rzeczywistego położenia. Wraz ze wzrostem gęstości prążków I_{BC} błąd lokalizacji tych dwóch metod maleje. Na rysunku 47 przedstawiono wyniki uzyskane dla przypadku, gdy na interferogramie I_{ABC} wygenerowano 540 wirów optycznych.

Na rysunku 47 przedstawiona jest dokładność lokalizacji metod, gdy liczba prążków pionowych zarejestrowanych na interferogramie I_{BC} wynosi 20. Jak widać, błąd systematyczny, z jakim metody amplitud i wektorów lokalizują punkty nieciągłości fazowych, występuje niezależnie od gęstości rejestrowanych prążków interferencyjnych. Wielkość tego błędu zależna jest jednak od gęstości prążków: wraz ze wzrostem liczby wirów optycznych zmniejsza się błąd lokalizacji. Na rysunku 47 widać, że je-

żeli sieć wirów optycznych będzie utworzona z 540 wirów, to w takim przypadku metoda cosinusów około 4% wirów lokalizuje dalej niż 9 μm (1 piksel). Metoda trójkątów pracuje bez zarzutów.

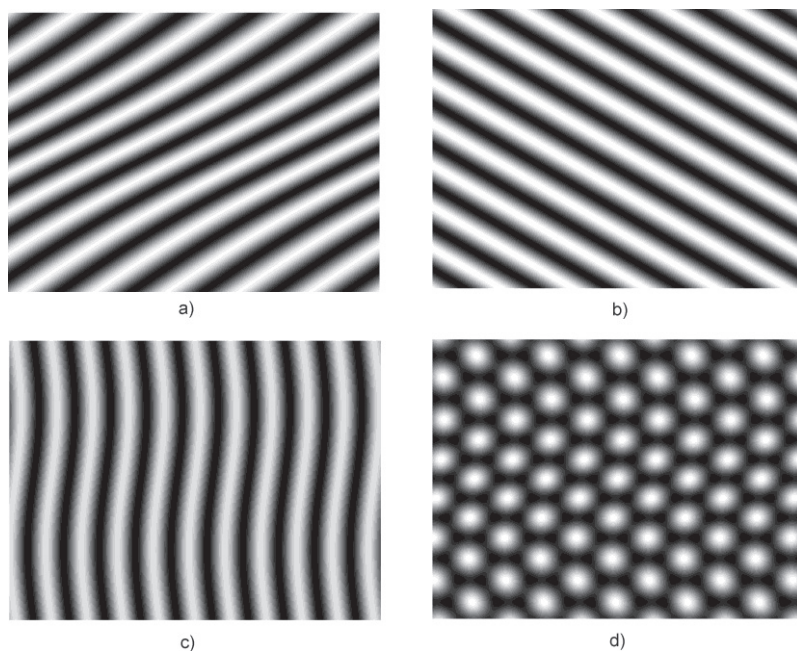


Rys. 47. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy zaburzone fale płaskie o różnych amplitudach oraz gdy na interferogramie I_{BC} obserwujemy 20 prążków pionowych (540 wirów)

Fig. 47. Diagram of localisation methods accuracy in case where are generated three disturbed plane waves with different amplitudes and where are 20 vertical fringes on interferogram I_{BC} (540 vortices)

4.1.5. HISTOGRAMY – DWIE FALE PŁASKIE ORAZ JEDNA FALA PŁASKA ZABURZONA FAZOWO

Ostatni rodzaj zaburzenia, jakim zostawały poddawane metody lokalizacyjne, to zaburzenie fazowe, przy czym sinusoidalnej zmianie fazy poddawana była tylko fala UB , a amplitudy trzech fal były różne. Wartości amplitud były takie same jak w poprzednich podrozdziałach ($A = 0,95$, $B = 1,05$, $C = 1,00$). Zaburzenie fazowe było testowane, ponieważ na interferogramach zarejestrowanych podczas eksperymentu obserwowane były drobne zakrzywienia prążków. Obserwowane deformacje prążków były mniejsze niż te wygenerowane w testach numerycznych. Na rysunku 50 przedstawione są interferogramy I_{AB} , I_{AC} , I_{BC} oraz I_{ABC} powstałe w wyniku interferencji dwóch fal płaskich i trzeciej fali płaskiej zaburzonej fazowo.



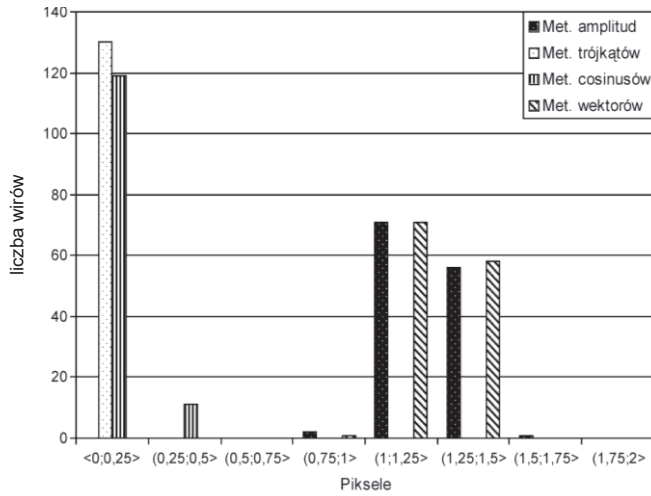
Rys. 48. Wynik interferencji dwóch fal płaskich o różnych amplitudach A i C z trzecią falą płaską zaburzoną fazowo – sieć sześciokątów: a – prążki interferencyjne utworzone przez fale UA i UB , b – prążki interferencyjne utworzone przez fale UA i UC , c – prążki interferencyjne utworzone przez fale UB i UC , d – prążki interferencyjne utworzone przez fale UA , UB i UC .

Na interferogramie I_{BC} obserwujemy 10,5 prążków pionowych (130 wirów)

Fig. 48. Result of interference of two plane waves with different amplitudes A and B and third plane wave with phase disturbed – hexagonal lattice: a – interference fringes generated by waves UA i UB , b – interference fringes generated by waves UA and UC ; c) interference fringes generated by waves UB and UC , d – interference pattern generated by waves UA , UB and UC .

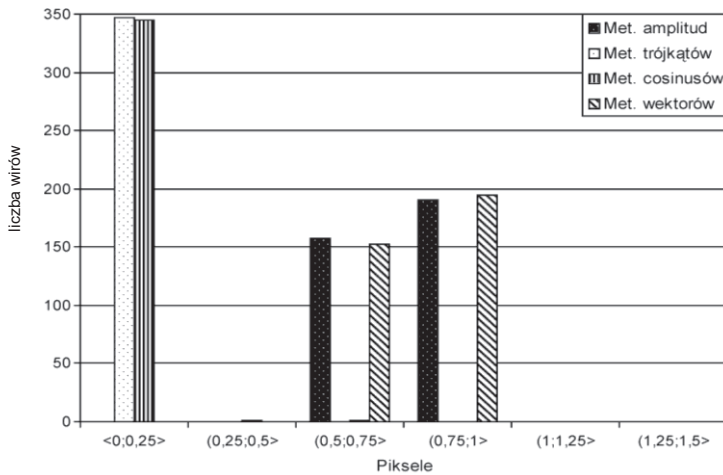
We observed 10,5 vertical fringes on interferogram I_{BC} (130 vortices)

W przypadku gdy spośród trzech fal płaskich, których amplitudy są różne, jedną zaburzymy fazowo, wówczas metody trójkątów i cosinusów lokalizują wiry optyczne z dokładnością $4,5 \mu\text{m}$ ($\frac{1}{2}$ piksela) dla rzadszej sieci wirów (rys. 49) oraz z dokładnością $2,25 \mu\text{m}$ ($\frac{1}{4}$ piksela) dla gęstszej sieci wirów (rys. 50), przy czym błąd metody cosinusów rośnie, gdy liczba prążków pionowych przekracza liczbę 19 (rys. 51). Natomiast metody amplitud i wektorów, tak jak w poprzednim punkcie (trzy fale płaskie o różnych amplitudach zaburzone innymi falami), obarczone są błędem systematycznym. W tym przypadku również te dwie metody radzą sobie lepiej, gdy analizują interferogramy zawierające gęstsze prążki. Błąd lokalizacji tych metod spada do $9 \mu\text{m}$ (1 piksela) dopiero, kiedy na interferogramie I_{BC} obserwujemy około 16 prążków. Na rysunku 49 zostały przedstawione przykładowe wykresy ilustrujące dokładność działania metod lokalizacji.



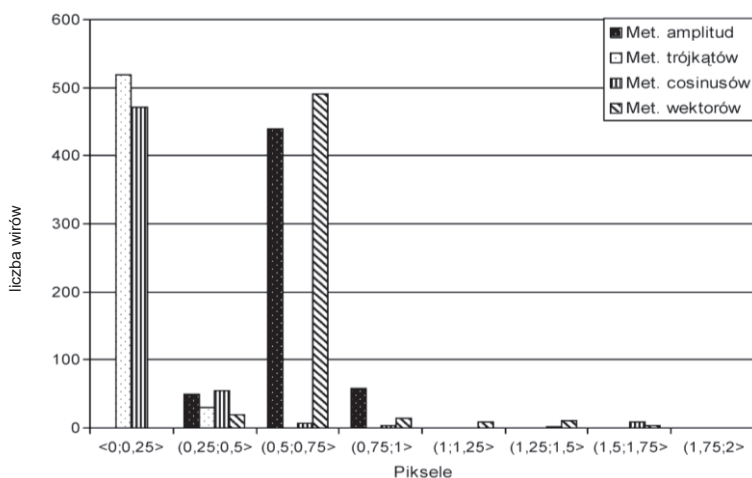
Rys. 49. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy fale płaskie o różnych amplitudach oraz gdy jedna z interferujących fal ma zaburzony front falowy (130 wirów w analizowanym obszarze)

Fig. 49. Diagram of localisation methods accuracy in case where are generated three plane waves with different amplitudes and where one of interfering wave have disturbed wave-front (130 vortices in analyzed area)



Rys. 50. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy fale płaskie o różnych amplitudach oraz gdy jedna z interferujących fal ma zaburzony front falowy (347 wirów w analizowanym obszarze)

Fig. 50. Diagram of localisation methods accuracy in case where are generated three plane waves with different amplitudes and where one of interfering wave have disturbed wave-front (347 vortices in analyzed area)



Rys. 51. Wykres ilustrujący dokładność działania metod lokalizacji w przypadku, gdy generowane są trzy fale płaskie o różnych amplitudach oraz gdy jedna z interferujących fal ma zaburzony front falowy (548 wirów w analizowanym obszarze)

Fig. 51. Diagram of localisation methods accuracy in case where are generated three plane waves with different amplitudes and where one of interfering wave have disturbed wave-front (548 vortices in analyzed area)

4.1.6. WYKRESY FAZOWE

Sposób porównywania metod lokalizacji, który był przedstawiony w punkcie 4 nie może być zastosowany do analizy danych eksperymentalnych, ponieważ nie możemy tu zastosować metody opisanej w punkcie 1.2, polegającej na znalezieniu punktów przecięcia części rzeczywistej i urojonej funkcji U_{ABC} w zerze (metoda teoretyczna). Zostanie więc przedstawiony drugi sposób oceny poprawności działania opracowanych metod lokalizacji. Polega on na wyliczeniu wartości fazy w zlokalizowanych punktach. Jeżeli punkty zostały poprawnie znalezione, to powinny one utworzyć dwie płaszczyzny fazowe odpowiadające wirom dodatnim i ujemnym. Rozrzut punktów w takiej płaszczyźnie fazowej jest wyznacznikiem poprawności działania analizowanej metody lokalizacji. Dokładność tej metody jest ograniczona przez nierównomierności amplitud interferujących fal. Jednak dla starannie przygotowanych wiązek błędy wprowadzone przez zmienność amplitudy nie są duże i nie podważają wartości użytej metody testowej. Warto podkreślić, że zaburzenia fazowe interferujących fal płaskich nie mają wpływu na działanie zaproponowanej metody testowej.

Analiza wielkości rozrzutu punktów w płaszczyznach o równej wartości fazy została przedstawiona na przykładzie 11,5 prążków zajmujących przestrzeń 576 pikseli. Przy takiej gęstości jeden prążek zajmuje około 50 pikseli ($11,5 \cdot 50 = 575$). Wraz z krokiem równym jednemu pikselowi faza zmienia się o $0,126$ radiana ($2\pi/50$)

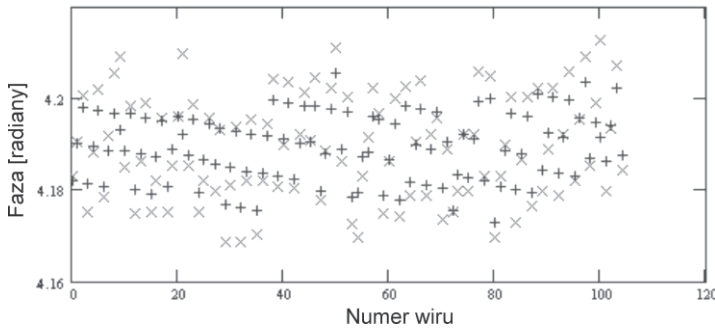
$\approx 0,126$ radiana). Ponieważ w metodzie teoretycznej stosowana jest interpolacja dziesięciokrotna, więc jednemu krokowi w obliczeniach numerycznych odpowiada zmiana fazy o $0,0126$ radiana. Faza względna pomiędzy falami UA i UC liczona jest z twierdzenia cosinusów:

$$\gamma_{ca} = \arccos\left(\frac{(Ia)^2 + (Ic)^2 - (Ib)^2}{2IaIc}\right) \quad (12)$$

gdzie wszystkie oznaczenia są zgodne z rysunkiem 18, a symbole Ia , Ib , Ic oznaczają natężenia fal czytane kolejno, w punktach wirowych, z odpowiednich interferogramów (np. $Ia = I_{BC}$). Faza względna wyznaczana jest z następującej zależności (rys. 18)

$$\psi_3 = \pi + \gamma_{ca}. \quad (13)$$

Błąd bezwzględny wyznaczenia kąta pomiędzy falami UA i UC (twierdzenie cosinusów) oszacowany metodą różniczki zupełnej wynosi $\pm 0,020$ radiana. Sumując błąd kwantyzacji fazy z błędem wynikającym z metody obliczeniowej, otrzymujemy: $0,0126 + 0,020 = 0,0326$ radiana. Na wykresie 52 przedstawione są wartości fazy względnej pomiędzy idealnymi falami UA i UC obliczone dla kolejnych wirów dodatnich. Plussem oznaczono punkty otrzymane metodą teoretyczną, a krzyżykiem – metodą trójkątów.



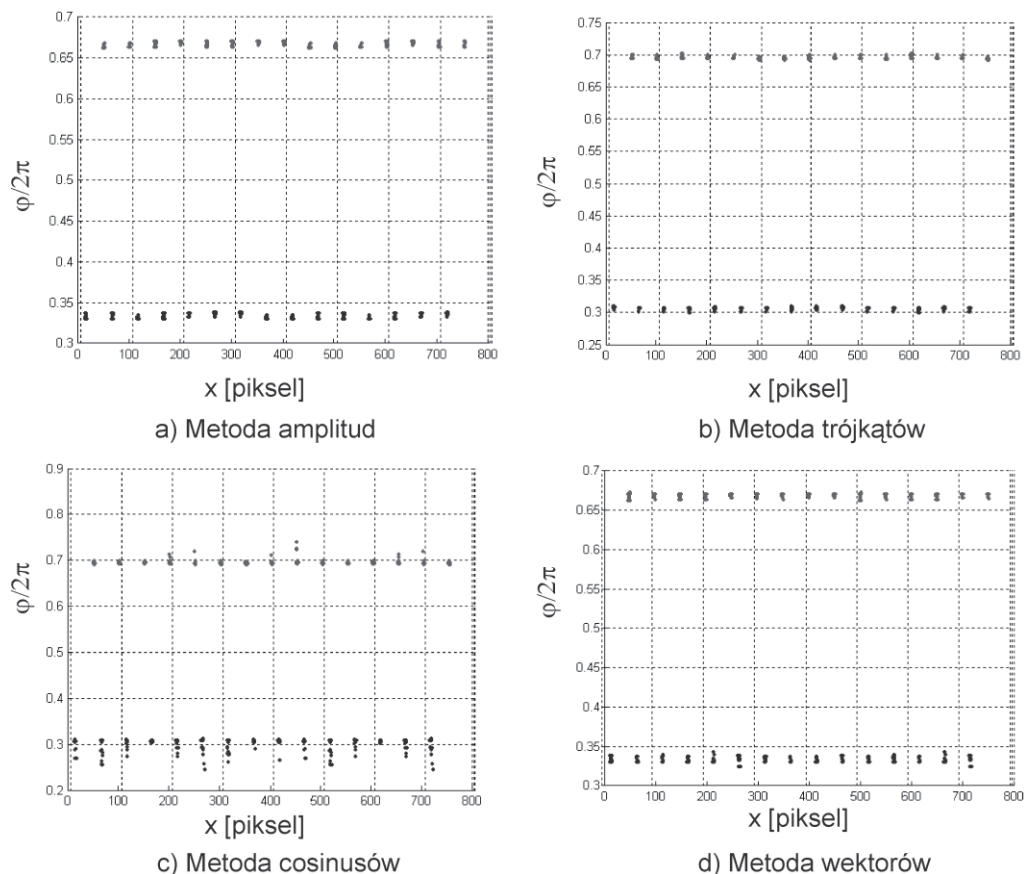
Rys. 52. Wykres przedstawiający wartości fazy względnej pomiędzy falami UA i UB obliczone dla kolejnych wirów o znaku dodatnim. '+' – metodą teoretyczną, 'x' – metodą trójkątów
Fig. 52. Diagram of relative phase value between waves UA and UB calculated for vortex with positive sign. '+' – theoretical method, 'x' – triangle method

Z rysunku 52 można odczytać, że rozrzut punktów przedstawionych na wykresie wynosi:

- dla metody teoretycznej: $0,033$ radiana;
- dla metody trójkątów: $0,044$ radiana.

Jak widać rozrzut wartości fazy wyznaczony metodą teoretyczną dla wirów tego samego znaku zgadza się z przewidywaniami. Przedział wartości faz wyznaczonych metodą trójkątów

tów jest większy od przedziału otrzymanego metodą teoretyczną – metoda trójkątów jest mniej dokładna. Rozrzut wartości fazy względnej w płaszczyznach fazowych, otrzymanych w wyniku pracy metodą lokalizacji, będzie większy niż w przypadku rozrzutu wartości fazy wyznaczonych metodą teoretyczną, ponieważ do błędów wynikających z kwantyzacji fazy i metody obliczeniowej fazy dochodzą błędy poszczególnych metod lokalizacji.



Rys. 53. Wykresy rozrzutu wartości fazy względnej, przedstawione dla czterech metod lokalizacyjnych, otrzymane z danych teoretycznych dla trzech fal płaskich o różnych amplitudach oraz gdy widocznych jest 15 prążków pionowych. Na osi x są współrzędne wirów, na osi y jest faza podzielona przez 2π

Fig. 53. Diagrams of the relative phase value spread for four localization methods obtained from theoretical data for three plane waves with different amplitudes and where 15 vertical fringes are visible. On the x axis are coordinate of vortex points and on y axis are phase divided by 2π

Rozrzut wartości fazy względnej pomiędzy dowolnie wybranymi dwoma falami w punktach wirowych może być wyznacznikiem poprawności działania metod lokalizacyjnych, w przypadku gdy nie mamy równań falowych, czyli w analizie danych

eksperymentalnych. Dalej przedstawione są przykładowe wykresy rozrzutu wartości fazy względnej otrzymanych z danych teoretycznych dla trzech fal płaskich o różnych amplitudach dla około 15 prążków pionowych. Na osi x są współrzędne wirów, na osi y jest faza przeskalowana na części długości fali ($2\pi \rightarrow \lambda$).

Na rysunku 53 widać, że najmniejszy rozrzut wartości fazy względnej występuje w płaszczyznach wirów zlokalizowanych metodą trójkątów. Rozrzut ten jest rzędu 0,04–0,05 radiana. Nieco większy rozrzut punktów obserwujemy w metodzie amplitud. Jest on większy o około 0,01 radiana. W metodzie tej obserwujemy przesunięcie wartości fazy. Płaszczyzny wirów dodatnich (górną grupę) oraz ujemnych (dolną grupę) są przesunięte o około 0,04 radiana ku sobie. Przesunięcie to spowodowane jest systematycznym błędem metody występującym przy lokalizowaniu punktów osobliwych. To samo przesunięcie w wyznaczeniu wartości fazy obserwujemy w metodzie wektorów, jednak w tym przypadku występuje większy rozrzut punktów w płaszczyźnie. Jest on dwukrotnie większy niż w przypadku metody trójkątów, czyli może wynosić nawet 0,1 radiana. Najgorzej w tym porównaniu wypadła metoda cosinusów. Rozrzut fazy sięga kilku dziesiątych części radiana.

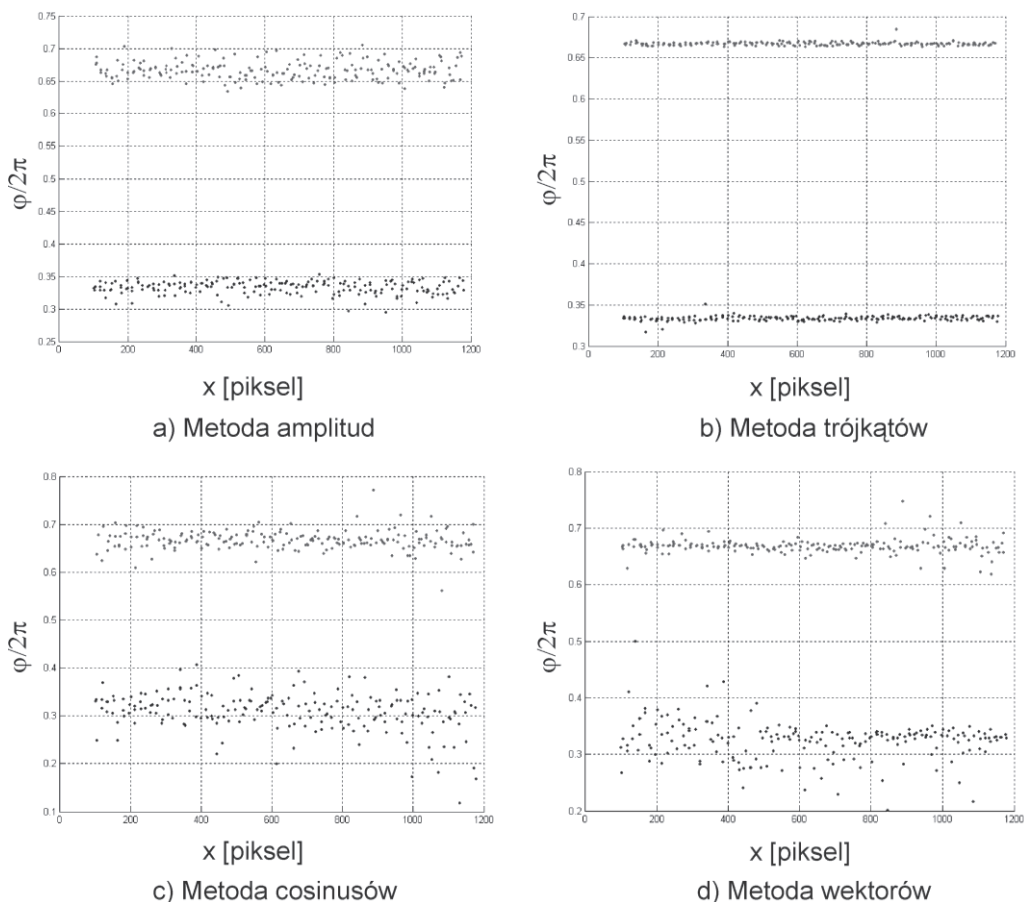
4.1.7. PODSUMOWANIE

Ze wszystkich metod lokalizacji przedstawionych w tym rozdziale najlepiej działa metoda trójkątów. Niezależnie od gęstości rejestrowanych prążków oraz różnicy natężeń interferujących fal (różnica pomiędzy dwoma – największą i najmniejszą – wartościami natężenia trzech interferujących fal jest mniejsza niż 30%) metoda trójkątów lokalizuje wiry optyczne z dokładnością 4,5 μm . Metody amplitud i wektorów działają bardzo podobnie. Obydwie są czułe na zmiany natężeń w trzech interferujących falach i obydwie są obciążone błędem systematycznym, którego pochodzenia nie udało się do tej pory ustalić. Błąd z jakim lokalizowane są wiry optyczne przez metody amplitud oraz wektorów jest duży (wynikający przede wszystkim z błędu systematycznego) – większy niż 9 μm . Wartość błędu wynosi 9 μm , gdy gęstość prążków pionowych na interferogramie I_{BC} jest większa niż 16 (jeden prążek powinien zajmować mniej niż 48 pikseli). Metody te mogą być jednak używane do analizy przesunięć wirów, np. do metody pomiaru małych kątów obrotu fali płaskiej [57], ponieważ w tym przypadku błąd systematyczny lokalizowania punktów wirowych nie ma znaczenia. Metoda cosinusów jest czuła na zaburzenia, które trudno jest wyeliminować w układzie pomiarowym (rys. 53). Ponieważ w symulacjach numerycznych, pomimo wielu prób, nie udało się wygenerować trzech fal zaburzonych w dokładnie taki sposób, w jaki był obserwowany w danych eksperymentalnych, więc błąd tej metody szacuję na większy od 9 μm . Z symulacji numerycznych wynika, że metoda cosinusów lokalizuje wiry optyczne z małym błędem 4,5 μm , jeżeli grubość rejestrowanych prążków zawiera się w przedziale $\langle 45,40 \rangle$ pikseli (od 17 do 19 prążków pionowych). Niewielki przedział poprawnego działania metody cosinusów jest jej wadą.

Wyniki testów numerycznych wskazują jednoznacznie na poprawność koncepcji zaproponowanych metod lokalizacji. Ich ostateczna weryfikacja musi być oczywiście przeprowadzona przez eksperyment, co będzie tematem następnego paragrafu.

4.2. WYNIKI EKSPERYMENTALNE

Na rysunku 54 pokazane są wykresy fazowe danych eksperymentalnych uzyskanych w układzie interferometru z automatycznymi przesłonami. Interferogramy były



Rys. 54. Wykresy rozrzutu wartości fazy względnej, przedstawione dla czterech metod lokalizacyjnych, otrzymane z danych eksperymentalnych, kiedy jest widocznych 21 prążków poziomych (grubość prążka około 50 pikseli). Na osi x są współrzędne wirów, na osi y jest faza podzielona przez 2π

Fig. 54. Diagrams of the relative phase value spread for four localization methods obtained from experimental data for three plane waves with different amplitudes and where 15 vertical fringes are visible (width of fringe are about 50 pixels). On the x axis are coordinate of vortex points and on y axis are phase divided by 2π

zarejestrowane 12-bitową kamerą pomiarową o rozmiarach elementu CCD 1344×1024 pikseli (rozmiar piksela wynosił: $6,45 \mu\text{m}$ na $6,45 \mu\text{m}$).

Jeżeli w metodzie trójkątów odrzucimy parę punktów, które wyraźnie odstają od pozostałych (powstałe z powodu pyłków na elementach optycznych), to rozrzut punktów w płaszczyznach fazowych, rysunek 54b, wynosi $0,05$ radiana. W pozostałych metodach obserwujemy duży rozrzut punktów. Jest to spowodowane zaburzeniami amplitudy, które wprowadza układ pomiarowy. Metoda trójkątów, jak widać, jest odporna na zaburzenia amplitudowe występujące w układzie pomiarowym. W metodzie cosinusów i wektorów występują dodatkowo błędy systematyczne.

Należy tu rozdzielić dwie sprawy. Jak było wspomniane zmienność amplitud interferujących fal zmniejsza dokładność metody amplitud, wektorów i cosinusów, czego można się było spodziewać na podstawie analizy numerycznej. Z drugiej strony zmienność amplitud powoduje, że w punktach wiru optycznego fazy względne interferujących fal nie są takie same. Obserwowany rozrzut kąta fazowego otrzymany metodą trójkątów nie może być zatem podstawą do precyzyjnego określenia dokładności tej metody. Małego rozrzutu należy oczekiwać, ze względu na istniejącą zmienność amplitud. W wynikach otrzymanych za pomocą metody trójkątów obserwujemy najmniejszy rozrzut kąta fazowego. Błędy wykazane przez pozostałe metody wykraczają poza zakres zmienności względnych faz, wynikający ze zmienności amplitudowych interferujących fal.

4.3. PODSUMOWANIE

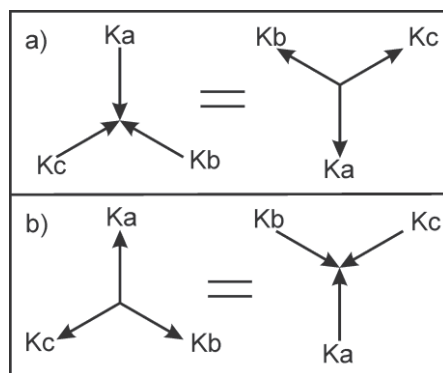
Metody lokalizacji wirów optycznych mają podstawowe znaczenie dla osiągnięć interferometru na wirach optycznych. Na podstawie przeprowadzonych testów numerycznych i doświadczalnych, których typowe wyniki zostały przedstawione wcześniej, można stwierdzić, że dysponujemy grupą metod, która pozwala na dokładne pomiary z użyciem interferometru na wirach optycznych, co było najważniejszym zadaniem na drodze do jego praktycznych zastosowań. Praktycznie dodatkową zaletą opracowanych metod jest to, że wykorzystują elementarne operacje na macierzach. W efekcie czego lokalizacja wirów, z użyciem komputera klasy PC przebiega szybko.

5. METODY ROZRÓŻNIANIA ZNAKÓW WIRÓW OPTYCZNYCH

Znak wiru optycznego jest zdeterminowany, jak było wspomniane w punkcie 1.2., przez kierunek obrotu helikalnego frontu falowego [8], [75], [76]: jeśli front falowy ulega skręceniu zgodnym z kierunkiem wskazówek zegara, to znak wiru będącego w jego centrum jest dodatni, w przypadku lewoskrętnej helikoidy – wir ma znak ujemny. Znajomość znaku wiru optycznego rozwiązuje problem sklejanía fazy [50], [67] oraz jest pomocna w rekonstrukcji frontu falowego. Analizy teoretyczne rozróż-

niania znaków wirów optycznych [5], [72], omówione w punkcie 1.2, nie mogą być zastosowane do danych doświadczalnych, ponieważ wymagają one znajomości funkcji amplitudy zespolonej. Jedną z doświadczalnych metod rozróżniania znaków wirów optycznych zaproponowała H.V. Bogatiyova [17]. Jest to metoda umożliwiająca rozpoznanie znaku wiru przenoszonego w wiązce Laguerre–Gausa. W niniejszym rozdziale zostaną przedstawione i omówione dwie eksperymentalne metody rozróżniania znaków wirów optycznych rozmieszczonych w regularnej sieci [25]–[28].

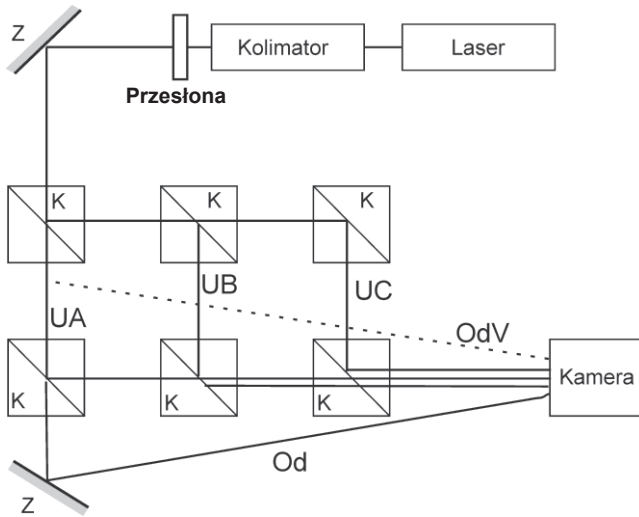
Informacje zapisane na interferogramach powstałych w wyniku interferencji trzech fal płaskich (I_{ABC} , I_{AB} , I_{AC} , I_{BC}) w interferometrze na wirach optycznych są niewystarczające do określenia znaków wirów optycznych, ponieważ dwa interferogramy przedstawiające dwie sieci wirów różniące się tylko znakami, są takie same. Dwa identyczne interferogramy przedstawiające sieci wirów o dwóch konfiguracjach znaków można uzyskać z interferencji trzech fal płaskich w dwóch układach różniących się kierunkami rozchodzenia się tych fal. Na rysunku 55 przedstawione są dwa przykładowe ułożenia rzutów wektorów falowych na płaszczyznę obserwacji, w których powstają identyczne interferogramy z wirami o przeciwnych znakach.



Rys. 55. Układy rzutów wektorów falowych A, B i C na płaszczyznę obrazu. Przypadki (a) oraz (b) odpowiadają otrzymaniu przeciwnych znaków wirów optycznych, ale jednakowych rozkładów natężenia

Fig. 55. Arrangement of projection of the wave normal vectors A, B and C on the screen. The cases (a) and (b) correspond to give opposite signs of optical vortices but the same intensity distribution

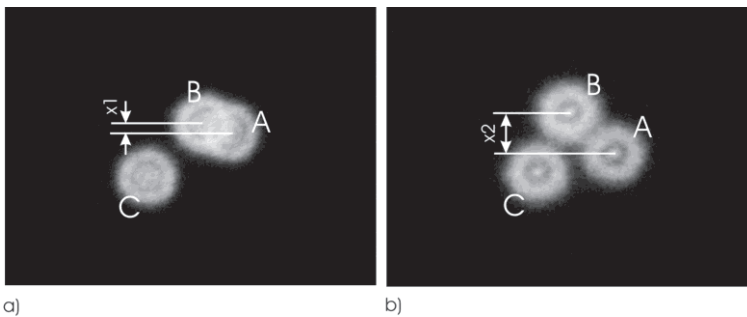
W wyniku interferencji trzech fal płaskich, których kierunki rozchodzenia się są takie, jak pokazano na rysunku 55a i 55b, powstają dwa identyczne interferogramy. Wiry optyczne wprawdzie w obu przypadkach mają takie samo położenie, ale różnią się znakami. W dalszej części tego rozdziału przedstawione są dwa eksperymentalne sposoby rozróżniania znaków wirów optycznych. Pomiar mający określić znak wirów był przeprowadzony za pomocą układu przedstawionego dalej (rys. 56).



Rys. 56. Układ pomiarowy stosowany do określania znaków wirów optycznych
 Fig. 56. Measurement setup to determinate optical vortex signs

5.1. METODA I

Analizując rysunek 55 można łatwo zauważyć, że w celu określenia ładunków topologicznych wirów optycznych należy sprawdzić kierunek propagacji, np. fali UA . Pomiar kierunku rozchodzenia się interferujących wiązek jest prosty – wystarczy zmniejszyć średnicę wiązki za kolimatorem, używając przesłony irysowej (rys. 56) do takiej wielkości, aby wiązki były rozróżnialne. Następnie należy zarejestrować natężenia trzech fal w dwóch położeniach ekranu (rys. 56).

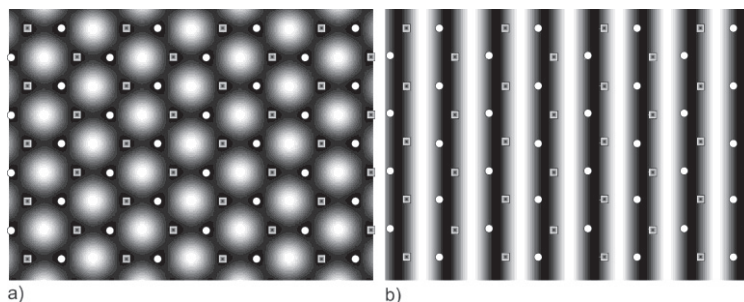


Rys. 57. Względne położenie trzech fal UA , UB i UC : a – blisko – w odległości około 10 cm, b – daleko – w odległości około 110 cm (średnica wiązki około 6 mm)

Fig. 57. Relative arrangement of waves UA , UB and UC :

a – near – about 10 cm, b – far – about 110 cm (diameter of beam is about 6 mm)

Na rysunku 57 pokazane są natężenia trzech fal UA , UB i UC zarejestrowane w odległości około 10 cm za interferometrem (rys. 57a) oraz w odległości około 110 cm za interferometrem (rys. 57b). Wektor falowy fali UA skierowany jest w dół.



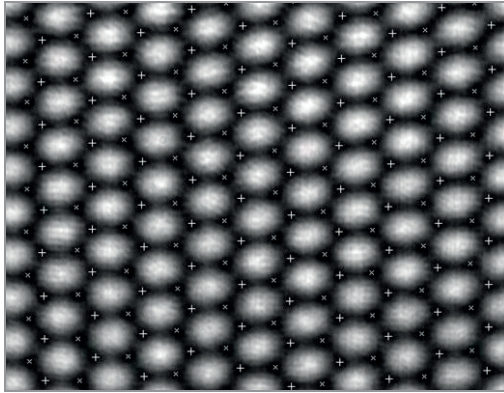
Rys. 58. Symulacja interferencji trzech wiązek płaskich z zaznaczonymi położeniami punktów wirowych ('+' – dodatnich, 'x' – ujemnych), a – punkty wirowe zaznaczone na interferogramie I_{ABC} , b – te same punkty wirowe jak w przypadku a) zaznaczone na interferogramie I_{BC}

Fig. 58. Simulation of three plane waves interference with positions and signs of vortex points marked ('+' – positive; 'x' – negative): a – vortex points are marked on interferogram I_{ABC} , b – the same vortex points as case a) are marked on interferogram I_{BC}

Do rozróżnienia dwóch grup wirów optycznych, różniących się ładunkiem topologicznym, używany jest interferogram fal UB i UC . W celu wyjaśnienia sposobu rozdziału dwóch grup wirów optycznych sporządzony został rysunek 58 przedstawiający interferogram I_{ABC} z zaznaczonymi położeniami wirów optycznych oraz interferogram I_{BC} z naniesionymi położeniami wirów optycznych.

Na rysunku 58b widać, że wiry o różnych znakach (kwadraty i kółka), naniesione na interferogram I_{BC} , leżą po dwóch przeciwnych stronach prążków interferencyjnych. Analiza rozkładu natężenia pola interferencyjnego dwóch fal płaskich UB i UC , w kierunku prostopadłym do prążków, wzdłuż osi x (z lewej do prawej), pozwala oddzielić wiry tworzące się, gdy wartość natężenia sumy tych dwóch fal wzrasta (kwadraty), od wirów tworzące się, gdy wartość natężenia sumy tych dwóch fal maleje (kółka). W ten sposób rozdzielone wiry optyczne tworzą dwie grupy o różnych znakach. Znak przypisywany jest, po sprawdzeniu kierunku rozchodzenia się wiązki UA , na podstawie wyników symulacji teoretycznych. Jeżeli kierunek rozchodzenia się fali płaskiej jest taki jak na rysunku 55a (wektor falowy fali UA skierowany jest w dół), to wiry oznaczone kwadratami (rys. 58) są dodatnie, w przeciwnym razie – rysunek 58b (wektor falowy fali UA skierowany jest do góry) – wiry oznaczone kwadratami (rys. 58) są ujemne.

Na rysunku 59 przedstawiony jest przykładowy interferogram I_{ABC} uzyskany doświadczalnie z zaznaczonymi położeniami i znakami („+” – wir dodatni, „x” – wir ujemny) wirów optycznych. Znaki wirów optycznych były wyznaczone metodą opisaną wcześniej.



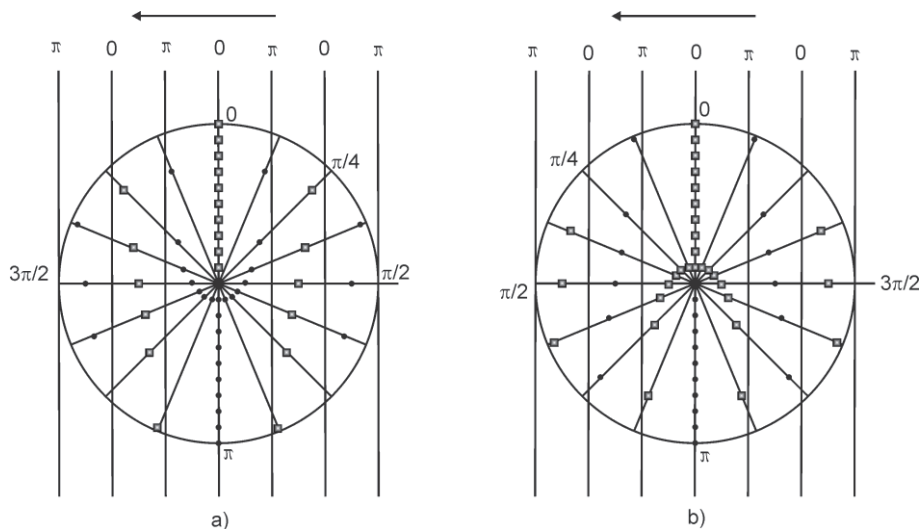
Rys. 59. Wynik zastosowania metody określania znaków wirów optycznych w rzeczywistym pomiarze, wir dodatni jest oznaczony przez znak „+”, ujemny przez „-”

Fig. 59. Result of apply the method to determine signs of optical vortices in experimental measurement, positive optical vortex is marked by “+” and negative by “-”

5.2. METODA II

Druga metoda rozróżniania znaków wirów optycznych wymaga zarejestrowania dodatkowego interferogramu z wiązką odniesienia (na rys. 56 oznaczona symbolem R), której wartość natężenia jest co najmniej równa sumie natężeń trzech interferujących fal. Warunek ten gwarantuje, że wiązka odniesienia nie wygeneruje dodatkowych wirów optycznych. W wyniku interferencji zaburzenia U_{ABC} z płaską falą odniesienia uzyskujemy charakterystyczny rozkład prążków z tak zwanymi „widelcami” [29], [38], [75]. Rozgałęzione prążki – widelce wskazują położenia wirów optycznych oraz ich znaki.

Na rysunku 60 przedstawiono przykładowy rozkład fazy fali o helikalnym kształcie frontu falowego z nałożonym rozkładem fazy płaskiej fali odniesienia. Czarnymi kółkami zaznaczono kilka punktów, w których występują minima interferencyjne, natomiast kwadratami oznaczono maksima interferencyjne. Na rysunku 60a widać kształt, utworzony z czarnych kółek, widelca skierowanego w górę, a na rysunku 60b widać kształt, utworzony z kwadratów, widelca skierowanego w dół. Na rysunku 60a znajduje się prawoskrętny helikalny front falowy więc wir optyczny znajdujący się w centrum wykreślonego okręgu ma znak dodatni [8] (patrzmy na interferogram, czyli w kierunku biegu interferujących wiązek), widelec zatem skierowany w górę oznacza wir o znaku dodatnim. Na rysunku 61b widać, że widelec skierowany w dół oznacza wir o znaku ujemnym. Na rysunku widać „czarny” widelec utworzony z minimów interferencyjnych (rys. 61a) oraz „biały” widelec utworzony z maksimów interferencyjnych (rys. 61b). To czy obserwowany widelec będzie miał czarną barwę lub białą zależy od różnicy faz pomiędzy falą przenoszącą wir optyczny a falą odniesienia.



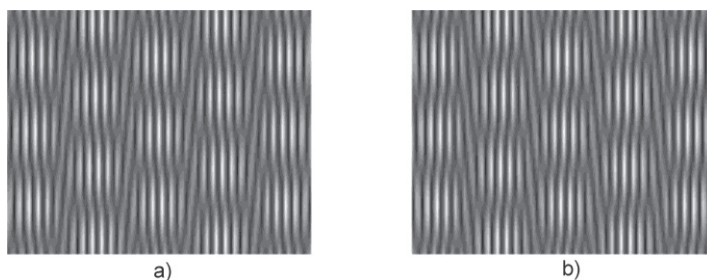
Rys. 60. Przykład interferencji fali przenoszącej wir optyczny z falą płaską (zaznaczony strzałką rzut wektora falowego płaskiej fali odniesienia) ilustrujący powstawanie widelców:

a – skierowany do góry – wir dodatni, b – widelec skierowany do dołu – wir ujemny.

Czarne kółka wskazują minima, kwadraty – maksima natężenia światła

Fig. 60. Example of interference wave with optical vortex include and plane wave (plane reference wave normal vector projection marked by black arrow) illustrate generation of fork-like fringes:

a – up – positive sign vortex, b – down – negative sign of vortex. The black circle mark minimum and square mak maximum intensity of light



Rys. 61. Wygenerowany rozkład natężenia dla interferencji trzech fal płaskich z płaską falą odniesienia, której kierunek propagacji jest taki jak pokazano na rysunku 58:

a – linią przerywaną *OdV*, b – linią ciągłą *Od*

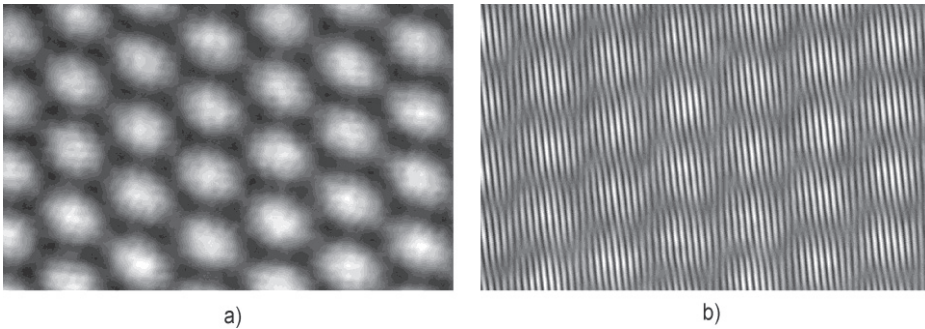
Fig. 61. Theoretically calculated distribution of intensity for interference three plane waves with fourth reference wave with direction of propagation is shown on figure 58:

a – by dashed line *OdV*, b – by solid line *Od*

Układ widelców określający parę wirów o przeciwnych znakach („góra – plus” oraz „dół – minus”) zmienia się, jeżeli kierunek propagacji fali odniesienia zmienia się

na przeciwny. Na rysunku 61 przedstawione są dwa interferogramy pokazujące sieć wirów optycznych powstałą w wyniku interferencji takich samych fal płaskich. Zmianie ulegał tylko kierunek propagacji fali odniesienia.

Na rysunku 61 widać zmiany orientacji widelców wskazujących wiry optyczne o tych samych znakach, która zależy tylko od kierunku propagacji fali odniesienia. Przypadek pokazany na rysunku 61a odpowiada sytuacji, gdy fala odniesienia OdV biegnie wzdłuż linii przerywanej zaznaczonej na rysunku 56; przypadek pokazany na rysunku 61b odpowiada zaś sytuacji, gdy fala odniesienia Od biegnie wzdłuż linii ciągłej zaznaczonej na rysunku 56. W układzie pomiarowym najprostszym rozwiązaniem jest skierowanie fali odniesienia w taki sposób jak pokazano na rysunku 56 – fala Od . W tym przypadku widelce skierowane do góry oznaczają wir o ładunku dodatnim, widelce zaś skierowane do dołu oznaczają wir o ładunku ujemnym. Na rysunku 62 pokazane są wyniki eksperymentalne.



Rys. 62. Zarejestrowany rozkład natężenia: a – sześciokątnej sieci wirów optycznych, b – ta sam sieć wirów optycznych po interferencji z falą odniesienia
 Fig. 62. Intensity distribution registered: a – hexagonal lattice of optical vortices, b – the same lattice of optical vortices after interference with reference wave

5.3. PODSUMOWANIE

Obydwie zaproponowane eksperymentalne metody rozpoznawania znaków wirów optycznych wymagają rejestracji interferogramu I_{ABC} – w celu lokalizacji położenia wirów optycznych. Metoda I wymaga rejestracji interferogramu I_{BC} oraz rozkładu natężenia trzech fal UA , UB i UC o zmniejszonych aperturach w dwóch położeniach. Metoda II wymaga rejestracji wyniku interferencji czterech fal I_{ABC0d} . Rezultaty obu metod są zgodne. Obydwie metody są proste w zastosowaniu i analizie. W tej sytuacji decyzja o zastosowaniu metody I lub drugiej zależy od konkretnego problemu pomiarowego; czy wygodniej jest dokonać pomiaru przesunięć zwężonych wiązek, w dwóch przesuniętych płaszczyznach, czy użyć dodatkowej fali odniesienia.

LITERATURA

- [1] ABRAMOCHKIN E.G., LOSEVSKY N., VOLOSTNIKOV V., *Generation of spiral-type laser beams*, Opt. Comm., 141, 1997, 59–64.
- [2] ABRİKOSOW A.A., *Nadprzewodniki drugiego rodzaju i sieci wirów*, Postępy Fizyki, Tom 55, Zeszyt 5, 2004, 199.
- [3] ACKEMANN T., KRIEGE E., LANGE W., *Phase singularities via nonlinear beam propagation in sodium vapor*, Opt. Comm. **155**, 1995, 339.
- [4] AKSENOV V.P., TIKHOMIROVA O., *Theory of singular-phase reconstruction for an optical speckle field in the turbulent atmosphere*, J. Opt. Soc. Am. A. Vol. 19, No. 2, 2002, 351–359.
- [5] ALLEN L., BARNETT S.M., PADGETT M.J., *Optical Angular Momentum*, Institute of Physics Publishing, London, 2003.
- [6] ALLEN L., PADGETT M.J., BABIKER M., *The orbital angular momentum of light*, Progress in Optics, E. Wolf (Ed.), Vol. XXXIX Chapter IV, 1999.
- [7] ASHKIN A., DZIEDZIC J.M., BJORKHOLM J. E., CHU S., *Observation of a single-beam gradient force optical trap for dielectric particles*, Optics Letters, 11, 1986, 288–290.
- [8] BASISTY I.V., SOSKIN M.S., VASNETSOV M.V., *Optical wavefront dislocation and their properties*, Opt. Comm., 199, 1995, 604–612.
- [9] BERRY M.V., *Much ado about nothing: optical dislocation lines (phase singularities, zeros, vortices...)*, International Conference on Singular Optics, SPIE, Vol. 3487, 1998, 1–15.
- [10] BERRY M.V., *Making waves in physics. Three wave singularities from the miraculous 1830s*, Nature, Vol. 403, 2000, 21.
- [11] BERRY M.V., *Coloured phase singularities*, New Journal of Physics, 4, 2002, 66.1–66.14.
- [12] BERRY M.V., *Exploring the colours of dark light*, New Journal of Physics, 4, 2002, 74.1–74.14.
- [13] BARZYKOWSKI J., *Współczesna metrologia*, WNT, Warszawa, 2004.
- [14] BAZHENOV V.Yu., SOSKIN M.S., VASNETSOV M.V., *Screw dislocations in light wavefronts*, J. Mod. Opt., 39, 1992, 985–990.
- [15] BEIJERSBERGEN M.W., COERWINKEL R.P.C., KRISTENSEN M., WOERDMAN J.P., *Helical-wavefront laser beam produced with a spiral phaseplate*, Opt. Comm., 112, 1994, 321–327.
- [16] BIALYNICKI-BIRULA I., BIALYNICKA-BIRULA Z., ŚLIWA C., *Motion of vortex lines in quantum mechanics*, Phys. Rev. A, Vol. 61, 2000, 032110.
- [17] BOGATIRYOVA H.V., FELDE C.V., POLYANSKII P.V., *Referenceless testing of vortex optical beams*, Optica Applicata, Vol. XXXIII, No. 4, 2003.
- [18] COURTIAL J., PADGETT M.J., *Performance of a cylindrical lens mode converter for producing Laguerre–Gaussian laser modes*, Opt. Comm., 159, 1999, 13–18.
- [19] COURTIAL J., O’HOLLERAN K., *Some of the mechanical and quantum-mechanical properties of optical vortices*, Eur. Phys. J., 145, 2007, 35–47.
- [20] DENNIS M.R., *Topological Singularities in Wave Fields*, University of Bristol, praca doktorska, November 2001.
- [21] DENNIS M.R., *Local phase structure of wave dislocation lines: twist and twirl*, J. Opt. A: Pure Appl. Opt., 6, 5, 2004, S202–S208.
- [22] DESYATNIKOV A.S., TORNER L., KIVSHAR Yu.S., *Optical Vortices and Vortex Solitons*, Progress in Optics, E. Wolf (Ed.), 2005, 47.
- [23] DREISCHUH A., PAULUS G.G., ZACHER F., *Quasi-two-dimensional dark spatial solutions and generation of mixed phase dislocations*, Appl. Physics B, 69, 1999, 107–11.
- [24] DREISCHUH A., CHERVENKOV S., NESHEV D., PAULUS G.G., WALTHER H., *Generation of lattice structures of optical vortices*, Opt. Soc. America, 19 (3), 2002, 550–6.

- [25] FRĄCZEK E., FRĄCZEK W., *The experimental method to recognize signs of optical vortices*, SPIE Proc., 5945, 2004.
- [26] FRĄCZEK E., FRĄCZEK W., MROCZKA J., *The experimental method for topological charge determination of optical vortices in a regular net*, Optical Engineering, 44 (2), 2005, 25601.
- [27] FRĄCZEK E., FRĄCZEK W., MASAJADA J., *The new method of topological charge determination of optical vortices in the interference field of the optical vortex interferometer*, Optik, Vol. 117, No. 9, 2006, 423–425.
- [28] FRĄCZEK E., FRĄCZEK W., *Two methods to determination topological charge in regular net of optical vortices*, SPIE Proc. 5858, 2005.
- [29] FREUND I., SHVARTSMAN N., FREILIKHEN V., *Optical dislocation networks in highly random media*, Opt. Comm., 101, 1993, 247–264.
- [30] FREUND I., *Optical vortices in Gaussian random wave fields: statistical probability densities*, J. Opt. Soc. Am. A, 11, 1994, 1644–1652.
- [31] FREUND I., SHVARTSMAN N., *Wave-field phase singularities: The sign principle*, Phys. Rev. Lett., 50, 1994, 5164.
- [32] FREUND I., *Saddles, singularities, and extrema in random phase fields*, Phys. Rev. E, 52, 1995, 2348–2360.
- [33] FREUND I., FREILIKHEN V., *Parametrization of anisotropic vortices*, J. Opt. Soc. Am. A., Vol. 14, No. 8, 1997, 351–359.
- [34] FREUND I., *Vortex derivatives*, Opt. Comm., 137, 1997, 118–126.
- [35] GRIER D.G., *A revolution in optical manipulation*, Nature, 424, 2003, 810–816.
- [36] HALPERIN B.I. *Statistical mechanics of topological defects*, Les Houches Session XXV – Physics of Defects, 813–857, North-Holland, Amsterdam, 1981.
- [37] HARKE B., KELLER J., ULLAL C.K., WESTPHAL V., SHONLE A., HELL S.W., *Resolution scaling in STED microscopy*, Opt. Express, Vol. 16, No. 6, 2008, 4154–4162.
- [38] He H., FRIESE M.E.J., HECKENBERG N.R., RUBINSZTEIN-DUNLOP H., *Direct observation of transfer of angular momentum to absorptive particles from a laser beam with a phase singularity*, Phys. Rev. Lett., 75, 1995, 826.
- [39] HECKENBERG N.R., McDUFF R., SMITH C.P., WHITE A.G., *Generation of optical phase singularities by computer-generated holograms*, Opt. Lett., 17, 1992, 221–223.
- [40] JOHNSON E.G., STACK J., KOEHLER CH., *Light Coupling by a Vortex Lens into Graded Index Fiber*, Journal of Lightwave Technology, Vol. 19, No. 5, 2001.
- [41] KHONINA S.N., SHINKARYEV V.V., SOIFER V.A., USPLENIEV G.V., *The phase rotor filter*, J. Mod. Opt., 39, 1992, 1147–1154.
- [42] KIM G.H., JEON J.H., NOH Y.C, KO K.H., MOON H.J., LEE J.H., CHANG J.S., *An array of phase singularities in a self-defocusing medium*, Opt. Comm., 147, 1998, 131–7.
- [43] KURZYNOWSKI P., WOŹNIAK W.A., FRĄCZEK E., *Optical vortices generation using Wollstone prism*, App. Optics, Appl. Optics, Vol. 45, No. 30, 2006, 7898–7903.
- [44] KURZYNOWSKI P., BORWIŃSKA M., *Generation of vortex-type markers in a one-way setup*, Appl. Opt., 46, 2007, 676–679.
- [45] LEACH J., PADGETT M.J., *Observation of chromatic effects near a white-light vortex*, New J. Phys., 5, 2003, 154.1–154.7
- [46] LEE J.H., FOO G., JOHNSON E.G., SWARTZLANDER G.A., *Experimental Verification of an Optical Vortex Coronagraph*, PRL, 97, 2006, 053901.
- [47] MADISON K.W., CHEVY F., WOHLLEBEN W., DALIBARD J., *Vortex formation in a stirred Bose-Einstein condensate*, Phys. Rev. Lett., 84, 2000, 806–809.
- [48] MAIOROV B., OSQUIGUIL E., *Vortex solid state in $YBa_2Cu_3O_{7-\delta}$ twinned crystals*, Phys. Rev. B, Vol. 64, 2001, 052511.

- [49] MAMAEV A.V., SAFFMAN M., ZAZULYA A.A., *Propagation of dark stripe beams in nonlinear media: snake instability and creation of optical vortices*, Phys. Rev. Lett., 76, 1996, 2262.
- [50] D. MALACARA, M. SERVIN, Z. MALACARA, *Interferogram Analysis for Optical Testing*, Marcel Dekker, New York, 1998.
- [51] MARRUCCI L., MANZO C., PAPARO D., *Optical Spin-to-Orbital Angular Momentum Conversion in Inhomogeneous Anisotropic Media*, Phys. Rev. Lett., 96, 2006, 163905.
- [52] MASAJADA J., *Synthetic holograms for optical vortices generation – Image evaluation*, Optik, 110, 1999, 554–558.
- [53] MASAJADA J., DUBIK B., *Optical vortex generation by three plane wave interference*, Opt. Comm., 198, 2001, 21–27.
- [54] MASAJADA J., POPIOŁEK-MASAJADA A., WIELICZKA D.M., *The interferometric system using optical vortices as phase marks*, Opt. Comm., 207, 2002, 85–93.
- [55] MASAJADA J., POPIOŁEK-MASAJADA A., FRĄCZEK E., FRĄCZEK W., *Vortex point localization problem in optical vortices interferometry*, Opt. Comm., 234, 2004, 23–28.
- [56] MASAJADA J., *Optical vortices and their application to interferometry*, OWPW Wrocław, 2004.
- [57] MASAJADA J., *Small-angle rotation measurement using optical vortex interferometer*, Opt. Comm., 239, 2004, 373–381.
- [58] MASAJADA J., POPIOŁEK-MASAJADA A., M. LENIEC, *Creation of vortex lattices by a wavefront division*, Opt. Express, Vol. 15, No. 8, 2007, 5196.
- [59] MASAJADA J., *Interferometr optyczny i sposób wytwarzania obrazu interferencyjnego*, zgłoszenie patentowe nr P368665.
- [60] MOA W., ZHONG Y., DONG J., WANG H., *Crystallography of two-dimensional photonic lattices formed by holography of three noncoplanar beams*, J. Opt. Soc. Am. B., Vol. 22, No. 5, 2005, 1085–1091.
- [61] NICHOLLS K.W., NYE J.F., *Three-beam model for studying dislocations in wave pulses*, J. Phys. A: Math. Gen., 20, 1987, 4673–4696.
- [62] NYE J.F., BERRY M.V., *Dislocations in wave trains*, Proc. R. Soc. Lond. A., 336, 1974, 165–190.
- [63] NYE J.F., *Natural focusing and fine structure of light*, IOP Bristol, 1999.
- [64] PATPRSKI K., KUJAWIŃSKA M., SAŁBUT L., *Interferometria laserowa*, OWPW, Warszawa, 2005.
- [65] PATR A.S., KHARE A., *Interferometric array generation*, Opt. Comm., 38, 2006, 37–45.
- [66] PEARSON J.E., MCGILL T.C., KURTIN S., YARIVT A., *Diffraction of Gaussian Laser Beams by a Semi-Infinite Plane*, J. Opt. Soc. Am., 59, 1969, 1440–1445.
- [67] ROBINSON D.W Edit., REID G.T., *Interferogram Analysis*, Institute of Physics Publishing, London, 1993.
- [68] SACKS Z.S., ROZAS D., SWARTZLANDER G.A., *Holographic formation of optical-vortex filaments*, J. Opt. Soc. Am. B, 15, No. 8, 1998, 2226–2234.
- [69] SCHONBRUN E., PIESTUN R., JORDAN P., COOPER J., WULFF K., COURTIAL J., PADGETT M., *3D interferometric optical tweezers using a single spatial light modulator*, Opt. Expr., 13, 2005, 3777–3786.
- [71] SALGUEIRO J.R., KIVSHAR Y.S., *Switching with vortex beams in nonlinear concentric couplers*, Opt. Expr., Vol. 15, No. 20, 2007, 12916–12921.
- [72] SENTHILKUMARAN P., *Optical phase singularities in detection of laser beam collimation*, Appl. Opt., 42, 2003, 6314–6320.
- [73] SOSKIN M.S., GORSHKOV V.N., VASNETSOV M.V., MALOS J.T., HECKENBERG N.R., *Topological charge and angular momentum of light beams carrying optical vortices*, Physical Review A, 56, 1997, 4064–75.
- [74] SOSKIN M.S., VASNETSOV M.V., *Singular optics*, Progress in Optics, 42, Chapter 4, 2001.

- [75] TIKHONENKO V., CHRISTOU J., LUTHER-DAVIES B., KIVSHAR Y.S., *Observation of vortex solitons created by the instability of dark soliton stripes*, Opt. Letters, 21, 1996, 1129–31.
- [76] VASNETSOV M.V., STALIUNAS K., *Optical vortices*, Nova Science Publishers, 1999.
- [77] VELZEL C., MASAJADA J., *Superresolution phase image microscope*, Opt. Appl., 39, 1999, 293–300.
- [78] WANG W., ISHII N., HANSON S.G., MIYAMOTO Y., TAKEDA M., *Phase singularities in analytic signal of white-light speckle pattern with application to micro-displacement measurement*, Opt. Comm., 248, 2005, 59–68.
- [79] WANG W., HANSON S.G., MIYAMOTO Y., TAKEDA M., *Experimental Investigation of Local Properties and Statistics of Optical Vortices in Random Wave Fields*, PRL 94, 2005, 103902.
- [80] YOUNG T., *Experimental demonstration of the General Law of the Interference of Light*, Philos. Trans. R. Soc. London 94, 1804.

OPTICAL VORTICES AS MARKERS IN THE OPTICAL FIELD

This work is devoted optical vortex interferometry, contain description of kind of singularities in the optical field known as optical vortex. In this publication are presented methods to generate a lattice of optical vortex with focused attention on interference method. Presented basis about optical vortices are grounds for technical realization of measurement systems. Submitted ways to obtain coordinates of centre of vortices are describe in details due to meaning vortex points localisation in applied optical vortices as phase markers in the optical field. The obtained accuracy of localisation have been analysed with connection to possible to occur trouble. The knowledge about localisation of vortex point can be useful to calculate measured object or optical field property. In some application may be required information about signs of topological charge of localised optical vortices. In this work are presented two methods to optical vortex sign discrimination.

*parametry kształtów linii widmowych,
pomiar częstotliwości promieniowania laserowego*

Piotr MASŁOWSKI*, Katarzyna BIELSKA*,
Agata CYGAN*, Roman CIURYŁO*,
Ryszard S. TRAWIŃSKI*

ROLA POMIARU CZĘSTOTLIWOŚCI W POMIARACH I ANALIZIE KSZTAŁTU LINII WIDMOWYCH**

W niniejszej pracy został opisany cyfrowy spektrograf laserowy wysokiej zdolności rozdzielczej służący do precyzyjnego badania kształtu linii widmowej 326,1 nm izotopów ^{113}Cd oraz ^{114}Cd zaburzonej przez gazy szlachetne, zbudowany w Instytucie Fizyki Uniwersytetu Mikołaja Kopernika. Szczególną uwagę zwrócono na układ pomiaru częstości wiązki generowanej przez laser barwnikowy. Przedstawiono również metody analizy danych otrzymywanych w pomiarze oraz sposoby weryfikacji uzyskiwanych wyników.

1. WPROWADZENIE

Obserwując widmo promieniowania elektromagnetycznego, emitowanego lub absorbowanego przez atomy lub cząsteczki w fazie gazowej, widzimy charakterystyczny układ ciemnych i jasnych prążków. Wzór ten jest niepowtarzalny i charakterystyczny dla określonej substancji i jako swoisty odcisk palca może być wykorzystany do jej identyfikacji. Jeżeli przyjrzymy się dokładniej pojedynczemu prążkowi reprezentującemu określoną linię widmową, zauważymy, że ma on skończoną szerokość spektralną – emitowane światło nie jest ściśle monochromatyczne oraz że obszar prążka nie ma ostro zaznaczonych granic. Widzimy pewien rozkład natężenia promieniowania nazywany kształtem linii widmowej. O ile obecność w badanym widmie określonej linii widmowej i jej natężenie daje nam informację o występowaniu określonego pierwiastka i jego koncentracji, o tyle kształt linii widmowej niesie informacje o stanie danego atomu (np. jego prędkości) i jego oddziaływaniu z otaczającymi go innymi atomami lub cząsteczkami, jak również zewnętrznymi polami elektrycznymi i magnetycznymi.

* Instytut Fizyki, Uniwersytet Mikołaja Kopernika w Toruniu.

** Praca wykonana w ramach projektu MNiSW nr 1 P03B 065 29.

Pomimo upływu ponad stu lat od pierwszych pionierskich prac poświęconych problematyce kształtów linii, tematykę tę trudno uznać za wyczerpaną. Wraz z rozwojem laserowych źródeł światła nastąpił dynamiczny rozwój nowoczesnych technik pomiarowych, takich jak laserowa spektroskopia absorpcyjna, laserowo indukowana fluorescencja (LIF), spektroskopia strat we wnęce (ang. *cavity ring-down spectroscopy* – CRDS) czy nieliniowe metody bezdopplerowskie. Metody te umożliwiły uzyskanie znaczącego wzrostu dokładności wyników przeprowadzanych badań spektroskopowych. Ułatwia to obserwację coraz subtelniejszych efektów wpływających na kształt linii widmowych, co umożliwia z jednej strony weryfikację z coraz większą dokładnością przewidywań teoretycznych dotyczących opisu emisji promieniowania w obecności gazu zaburzającego, a z drugiej strony coraz lepszy stosunek sygnału do szumu osiągniany w tych pomiarach tworzy zapotrzebowanie na coraz dokładniejsze profile teoretyczne.

Przykładem przełożenia się zwiększonej dokładności otrzymywanych wyników na możliwość dokładniejszej weryfikacji założeń teoretycznych może być użycie ich do sprawdzania wyników obliczeń numerycznych mających na celu wyznaczenie potencjałów oddziaływania międzycząsteczkowego. Dzięki wzajemnemu „oddziaływaniu” fizyków teoretycznych i doświadczalnych możliwy jest postęp w badaniach zarówno od strony obliczeniowej, jak i pomiarowej.

Wraz ze wzrostem stosunku sygnału do szumu w przeprowadzanych pomiarach pojawia się zapotrzebowanie na realistyczne profile kształtów linii, wychodzące poza stosowany dotychczas profil Voigta, które umożliwiłyby wyznaczenie z danych pomiarowych parametrów kształtu linii takich, jak szerokość ciśnieniowa linii, przesunięcie oraz ich natężenie z dokładnością rzędu ułamków procenta. Przykładem mogą być tutaj pomiary zawartości różnych substancji w próbkach gazu czy ciągłe monitorowanie poziomu stężenia wybranych substancji w określonych środowiskach, np. halach fabrycznych czy salach muzealnych. Zastosowanie w tym celu metod spektroskopii wysokiej zdolności rozdzielczej umożliwia uzyskanie wyników w czasie rzeczywistym (czas jednego pomiaru jest rzędu kilku sekund) z precyzją na poziomie jednej części na miliard (ppb), bez potrzeby uprzedniego przygotowywania próbek i wielokrotnej kalibracji urządzeń. Obecnie dostępne są już na rynku urządzenia komercyjne, bazujące na wspomnianej wcześniej metodzie CRDS, pozwalające mierzyć stężenia substancji śladowych, takich jak metan z precyzją 1 ppb [1], mierzyć poziom wilgotności [2] czy wyznaczać stosunek stężeń izotopów węgla z pomiaru absorpcji linii CO₂ z dokładnością 0,3% [1]. Rozwijane są także urządzenia, które służyłyby do celów medycznych, określając stężenie określonych związków w powietrzu wydychanym przez pacjenta, mających być symptomem schorzenia [3]. Ponieważ wszystkie wymienione urządzenia bazują na wyznaczeniu częstotliwości i natężenia linii absorpcyjnych określonych cząsteczek, ich precyzja jest ograniczona, poza innymi przyczynami, także przez zastosowany model profilu linii. W przypadku zastosowania profilu linii nie oddającego jej rzeczywistego kształtu, otrzymane wyniki mogą być obciążone

systematycznym błędem (np. [4]). Na potrzebę zastosowania realistycznych kształtów linii zwrócono uwagę także podczas badań atmosfery i zmian klimatu.

Prowadzenie precyzyjnych badań widm umożliwiających pomiar parametrów kształtów linii z dokładnością przewyższającą 1% wymaga uwzględniania wielu efektów fizycznych podczas analizy danych doświadczalnych. Większość analiz widm jest wykonywana z użyciem profilu Voigta, uwzględniającego dopplerowskie i zderzeniowe rozszerzenie linii widmowych. Jednak w przypadku gdy droga swobodna cząsteczek między zderzeniami jest mniejsza lub porównywalna z długością fali absorbowanego światła, dopplerowska składowa profilu linii może ulec zawężeniu, jest to tak zwane zwężenie Dickego. Ponadto w przypadku gdy atomy lub cząsteczki zaburzające (tzw. zaburzacze) badany gaz są cięższe lub mają masę porównywalną z masą atomów lub cząsteczek emitujących (bądź absorbujących) promieniowanie, zderzeniowa szerokość i przesunięcie linii nie mogą być traktowane jako niezależne od prędkości emitera (absorbera). Pojawia się tutaj korelacja pomiędzy ciśnieniowym (zderzeniowym) a dopplerowskim mechanizmem rozszerzenia linii, która może prowadzić do zwężenia profilu linii oraz powodująca jej asymetrię.

W ostatnim czasie ukazało się wiele prac poświęconych rozwojowi, analizom, testom i porównaniom istniejących modeli kształtu linii widmowych [5,6] wykorzystujących dane doświadczalne z różnych laboratoriów na świecie. Badania kształtów linii widm molekularnych mają istotne znaczenie nie tylko dla fizyki cząsteczkowej, ale są konieczne do prowadzenia badań ukierunkowanych na ochronę środowiska naturalnego i zrozumienie globalnych procesów zachodzących na naszej planecie. Prace te wpływają też na rozszerzenie naszej wiedzy astrofizycznej. Precyzyjne dane charakteryzujące słabe linie molekularne są szczególnie potrzebne podczas prowadzenia badań układów planetarnych oraz atmosfer planetarnych [7], w tym atmosfery Ziemi. Na przykład satelitarne systemy pomiarowe mogą służyć do globalnego monitoringu składu atmosfery z dużą rozdzielczością przestrzenną. Jednak tworzenie tego typu systemów wymaga precyzyjnych informacji na temat właściwości wykorzystywanych w nich przejść cząsteczkowych. Wymagania dotyczące spektrometrów i danych spektroskopowych wykorzystywanych w najnowszych satelitach monitorujących atmosferę zostały ostatnio scharakteryzowane w pracy [8]: *Most importantly, spectra must be recorded using an excellent spectrometer with well characterized instrumental line shape and also configured with detectors of minimal non-linearity. High signal to noise ratios (2000:1) at Doppler-limited resolution and long-term instrument stability are required.* Dalej autorzy [8] podają, że niepewność wyznaczanych natężeń linii nie powinna przekraczać 0,3%. Wymagania te mogą spełnić najnowsze metody spektroskopowe stosowane w badaniach kształtów linii widmowych, takie jak spektroskopia CRDS pod warunkiem odpowiednio starannego przeprowadzenia eksperymentu i analizy danych. Zostało to zademonstrowane przez współautora niniejszej pracy na przykładzie linii H_2O w podczerwieni [4,9]. Badania nad teoretycznym opisem precyzyjnych widm doświadczalnych przyczynią się do opracowania standardów obowiązujących w tworzeniu nowej

generacji baz danych spektroskopowych służących do badań atmosfery. Obecnie istniejące bazy danych (takie jak HITRAN) nie uwzględniają żadnych efektów kształtu linii poza ciśnieniowym rozszerzeniem i przesunięciem linii. Utrudnia to modelowanie doświadczalnych widm molekularnych, a zwłaszcza prowadzi do systematycznych, nawet kilkuprocentowych błędów w spektroskopowym wyznaczaniu koncentracji gazu.

Przykładem zastosowań praktycznych mogą być badania przeprowadzane w ramach programu SCIAMACHY prowadzonego przez Europejską Agencję Kosmiczną [10], czy planowane badania misji satelitarnej OCO (Orbiting Carbon Observatory) [11]. W przypadku tej ostatniej głównym zadaniem będzie monitorowanie stężenia CO₂, gazu mającego udział w powstawaniu efektu cieplarnianego, którego stężenie wzrasta na skutek działalności człowieka [12], tak aby stworzyć globalną mapę stężenia CO₂ i określić miejsca emisji i absorpcji tego gazu na świecie [11, 13]. W tym celu satelita będzie wyposażony w trzy spektrometry, pracujące w regionie 1,61 μm i 2,06 μm, co odpowiada wybranym pasmom absorpcyjnym CO₂, oraz 0,76 μm co pozwala na pomiary absorpcji pasma A cząsteczki O₂, służącego do kalibracji danych otrzymanych dla CO₂. Aby zarejestrować gradienty zmiany stężenia dwutlenku węgla, urządzenia powinny być w stanie wykrywać zmiany rzędu 1 ppm na tle typowej koncentracji 380 ppm. Aby osiągnąć taką zdolność rozdzielczą, potrzebna jest znajomość natężeń i parametrów ciśnieniowego rozszerzenia linii CO₂ i O₂ w wybranych przedziałach częstości z dokładnością do 0,3% lub lepszą [8,11]. Niestety dzisiejsze bazy parametrów linii cząsteczek O₂ i CO₂ [14] zawierają dane wyznaczone z dokładnością 2–3%. Za główne źródło błędów ograniczających dokładność, z jaką były wyznaczone te wartości, uważane są problemy z odpowiednim przygotowaniem próbek, dokładnym scharakteryzowaniem układu pomiarowego oraz modele kształtów linii użyte w analizie. Członkowie zespołu przygotowującego opisywaną misję wskazują na konieczność używania realistycznych profili linii [8], uwzględniających takie efekty, jak mieszanie linii, zwężenie Dickego czy efekty związane z korelacją pomiędzy dopplerowskim a ciśnieniowym rozszerzeniem linii (tzw. efekty speed-dependent). W celu uzyskania parametrów kształtów linii z wymaganą precyzją rozpoczęto wiele badań, których przykładem są badania pasma A cząsteczki tlenu, wykonywane za pomocą metody CRDS [15, 16].

Podobne problemy występują także w spektroskopii atmosfer innych planet. Istniejące metody badawcze umożliwiają pomiary takich czynników, jak skład atmosfery, rozkład ciśnienia, temperatury czy struktury chmur. Zainteresowanie astrofizyków skierowane jest także na dokładne wyznaczenie stosunków zawartości różnych izotopów określonych pierwiastków, co umożliwia np. wyciąganie wniosków na temat sposobu powstawania planet (przykładowo przez porównanie stosunków stężeń izotopów tlenu czy azotu [17]). Do prawidłowej analizy danych otrzymanych w wyniku badań satelitarnych czy prowadzonych z powierzchni Ziemi, na przykład w modelowaniu atmosfery planet za pomocą modeli transportu promieniowania konieczna jest bardzo dokładna znajomość takich parametrów, jak: długość fali badanej linii, jej na-

tężenie, parametry ciśnieniowego rozszerzenia, przesunięcia i asymetrii linii wynikające ze skończonego czasu trwania zderzenia czy efektu mieszania linii [7, 17].

Wybór linii interkombinacyjnej 326,1 nm kadmu, jako obiektu badań kształtu linii, wynikał przede wszystkim z tego, że linia ta w przypadku izotopów parzysto-parzystych (takich jak stosowany izotop ^{114}Cd) jest dobrze izolowana, pozbawiona wpływu sąsiednich linii widmowych.

Dodatkowo, stosunek mas atomów zaburzających do masy atomu kadmu w przypadku zaburzenia przez gazy szlachetne zmienia się od 0,035 dla układu Cd–He do 1,15 dla układu Cd–Xe, co umożliwia w dość szerokim zakresie stosunków mas zbadanie efektów związanych z korelacją pomiędzy ciśnieniowym a dopplerowskim rozszerzeniem linii widmowej, które jest zależne od podanego stosunku mas. Kolejną zaletą tej linii jest możliwość wykorzystania w badaniach parzysto–nieparzystego izotopu (np. ^{113}Cd), dla którego linia widmowa 326,1 nm składa się z dwóch składowych nadsubtelnych. W układzie takim można oczekiwać, że równoległe z efektami korelacyjnymi i asymetrią zderzeniową wynikającą ze skończonego czasu trwania zderzenia, mogłaby pojawić się tzw. asymetria mieszania linii, występująca w przypadku, gdy linie widmowe emitowane są z poziomów leżących blisko siebie. Efekt ten jest powszechny w widmach molekularnych, jednak w przypadku układów atomowych, jak wynika z wyliczeń teoretycznych, powinien mieć bardzo małą wartość i jeszcze nie był zaobserwowany [18]. Jak widać, linia 326,1 nm Cd jest doskonałym „poligonem doświadczalnym”, umożliwiającym badanie z dużą precyzją wielu efektów związanych z emisją promieniowania w obecności gazu zaburzającego i na podstawie tych wyników opracowanie metod pomiarowych i analizy danych mających zastosowanie dla innych układów atomowych i molekularnych.

W ostatnich latach znaczenie atomu kadmu istotnie wzrosło ze względu na możliwe jego wykorzystanie w nowoczesnych eksperymentach bazujących na chłodzeniu atomów. Atom kadmu ma podobną konfigurację elektronową jak inne używane w spektroskopii zimnych atomów pierwiastki: Ca, Mg, Sr, i Yb. W wyniku tego możemy rozważać możliwość wykorzystania atomów Cd np. do budowy zegara optycznego (ze względu na długożyciowy stan $^3\text{P}_0$ [19]) czy zastosowań w optyce kwantowej [20], natomiast wąski rezonans $^1\text{S}_0 \rightarrow ^3\text{P}_1$ umożliwi uzyskanie bardzo niskiej granicy chłodzenia [21]. Dotychczas udało się zbudować pułapkę magnetooptyczną (MOT) działającą na przejściu $^1\text{S}_0 \rightarrow ^1\text{P}_1$ (229 nm) [20]. Ponadto optyczne rezonanse Feshbacha w pobliżu interkombinacyjnej linii oferują możliwość efektywnej zmiany oddziaływań międzyatomowych [22]. Biorąc pod uwagę podane zastosowania atomów Cd, oczywista staje się potrzeba uzyskania bardzo dokładnych informacji o strukturze poziomów energetycznych atomu kadmu i o jego oddziaływaniu z innymi atomami.

Chronologicznie pierwsze pomiary kształtu interkombinacyjnej linii 326,1 nm kadmu wykonane w naszym zespole przeprowadzono za pomocą ciśnieniowo przestrajanego interferometru Fabry–Perota o przedziale dyspersji $0,33\text{ cm}^{-1}$. Uzyskano wartości parametrów ciśnieniowego rozszerzenia i przesunięcia linii 326,1 nm ^{114}Cd

zaburzonej pięcioma gazami szlachetnymi: He, Ne, Ar, Kr, Xe w zakresie ciśnień do około 100 Torr [23–26]. Otrzymane na tym etapie wyniki umożliwiły weryfikację dostępnych wówczas potencjałów oddziaływania w układach Cd–GS [27, 28]. Wskazywały jednocześnie na występowanie w układzie Cd–Xe efektów związanych z korelacją pomiędzy ciśnieniowym a dopplerowskim rozszerzeniem linii.

W wyniku przeprowadzonych pomiarów interferometrycznych stwierdzono, że w istniejących warunkach pomiarowych nie było możliwe zarejestrowanie asymetrii zderzeniowej linii, której wartość była zbyt mała i maskowana przez periodyczność funkcji aparaturowej interferometru Fabry–Perota. Następną próbę rejestracji asymetrii kształtu linii przeprowadzono za pomocą metody klasycznej spektroskopii absorpcyjnej, z siatkowym monochromatorem skanningowym jako elementem dyspersyjnym. Wykonane w ten sposób pomiary kształtu linii 326,1 nm ^{114}Cd zaburzonego Kr w zakresie do 600 Tr [29], zgodnie z przewidywaniami teoretycznymi, pokazały występowanie asymetrii kształtu linii, której wartość wzrastała wraz z ciśnieniem gazu zaburzającego. Pomiary te były jednak obarczone błędem systematycznym wynikającym z niepełnej znajomości funkcji aparaturowej wykorzystanego monochromatora, która – w przeciwieństwie do funkcji aparaturowej interferometru Fabry–Perota – nie jest dana w postaci analitycznej.

Aby umożliwić poprawną rejestrację efektu asymetrii zderzeniowej, podczas badania kształtu linii 326,1 nm ^{114}Cd , zdecydowano się skorzystać z metody laserowo indukowanej fluorescencji (LIF). W tym celu zbudowano cyfrowy spektrometr laserowy [30], w którym wiązka laserowa o szerokości widmowej 1 MHz była przestrajana w obrębie profilu badanej linii widmowej z jednoczesnym rejestrowaniem natężenia sygnału fluorescencji. W przypadku spektrometru tego typu możliwe było pominięcie w analizie jego funkcji aparaturowej. Przeprowadzone pomiary kształtu linii 326,1 nm ^{114}Cd zaburzonej przez hel i neon [31], argon [32], krypton [33], ksenon [34] (jak również [30] oraz prace w niej cytowane) oraz gazy molekularne ([35], oraz prace w niej cytowane) pozwoliły zarówno na obserwację asymetrii zderzeniowej, jak i efektów korelacyjnych między ciśnieniowym a dopplerowskim rozszerzeniem linii.

Kolejnym naturalnym krokiem była decyzja o zbadaniu wymienionych efektów w przypadku wykorzystania parzysto–nieparzystego izotopu ^{113}Cd w celu wyznaczenia wartości całkowitej asymetrii linii (nazywanej dalej asymetrią dyspersyjną) będącej sumą asymetrii zderzeniowej i ewentualnej asymetrii mieszania linii. Przeprowadzone pierwsze pomiary pilotażowe [36] pokazały, że pomiar asymetrii dyspersyjnej w tym układzie jest możliwy, jednak z powodu zbyt małej precyzji uzyskiwanych wyników, konieczna była modyfikacja spektrometru laserowego tak, aby uzyskać zarówno większą dokładność wyznaczenia częstotliwości wiązki laserowej, jak i większy stosunek sygnału do szumu.

Prezentowane tutaj zagadnienia badań kształtu linii 326,1 nm ^{113}Cd zaburzonej przez argon i ksenon zostały częściowo przedstawione we wcześniejszych pracach [36–40].

W niniejszej pracy szczególną uwagę zwrócono na:

- konieczność precyzyjnego pomiaru częstotliwości promieniowania laserowego – przedstawiono zastosowane rozwiązanie sprzętowe,
- uwzględnienie niepewności wyznaczenia częstotliwości promieniowania laserowego w procesie dopasowania profili teoretycznych do zmierzonych kształtów linii i analizie jakości dopasowania.

2. KSZTAŁTY CIŚNIENIOWO ROZSZERZONYCH LINII WIDMOWYCH

Badania ciśnieniowego rozszerzenia i przesunięcia linii widmowych prowadzone są od ponad stu lat. Zjawisko to próbowano wyjaśniać najpierw w ujęciu klasycznym, następnie półklasycznym i kwantowym. Na gruncie fizyki klasycznej powstały dwie grupy teorii ciśnieniowego rozszerzenia linii: teorie zderzeniowe i kwazistatyczne. Teorie zderzeniowe opierały się na założeniu, że oddziaływanie emitera z zaburzaczem odbywa się jedynie w czasie krótkotrwałego zderzenia. Emiter traktowany był jako klasyczny drgający oscylator. Zmiana kształtu linii widmowej była spowodowana zaburzeniem fazy drgań tego oscylatora w wyniku zderzenia atomu emitera z zaburzaczem. Podstawy tych teorii zostały opracowane przez Lorentza w 1904 roku i rozwijane przez wielu innych autorów. Korzystając z nich można podać wzór na rozkład natężeń ciśnieniowo rozszerzonej i przesuniętej izolowanej linii widmowej:

$$I(\omega) = \frac{\gamma_L}{2\pi} \frac{1}{(\omega - \omega_0 - \Delta)^2 + \left(\frac{\gamma_L}{2}\right)^2}. \quad (1)$$

Parametry otrzymanego profilu γ_L i Δ oznaczają odpowiednio szerokość połówkową profilu i przesunięcie względem częstości niezaburzonej ω_0 . Teorie zderzeniowe były rozwijane między innymi przez Lenza [41, 42], Weisskopfa [43], Lindholma [44], Foleya [45] i Andersona [46,47]. Teorie te zakładały bardzo krótki czas oddziaływania emitera z zaburzaczem, tak więc podejście to było słuszne dla małych koncentracji i dużych prędkości atomów zaburzaczy.

Przeciwstawną grupą teorii były tak zwane teorie kwazistatyczne, zakładające, że emitery nieustannie znajduje się w polu sił pochodzących od praktycznie nieruchomych zaburzaczy. Podejście to było rozwijane w pracach Holstmarka [48, 49], Jabłońskiego [50], Kuhna [51, 52] i Margenau [53, 54]. Podejście kwazistatyczne poprawnie opisują kształt skrzydeł linii i jest prawdziwe dla dużych stężeń i małych prędkości zaburzaczy.

Pierwsza kwantowo-mechaniczna teoria ciśnieniowego rozszerzenia linii widmowych została sformułowana w 1945 r. przez Jabłońskiego [55]. Rozwinięciem jej są prace Barangera [56–58] i Sobelmana [59], którzy rozbudowali teorię Jabłońskiego do

postaci uwzględniającej efekty nieadiabacyjne. Szudy i Baylis [60–63], opierając się na teorii Jabłońskiego, stworzyli ogólną adiabaticzną teorię kształtu linii, opisującą profil linii w całym zakresie częstości, w przeciwieństwie do wcześniejszych teorii. Należy zaznaczyć, że korzystając z teorii Barangera i zakładając adiabaticzność oddziaływań w układzie emiter–zaburzacz, otrzymamy wyrażenie na kształt linii widmowej identyczne z przewidywaniami adiabaticznej teorii Szudego i Baylisa.

Ograniczymy się tutaj do podania wybranych wyników teorii ciśnieniowego rozszerzenia linii widmowych, które opisują zjawiska zachodzące w badanych układach Cd–Ar i Cd–Xe. Przedstawiono je w ujęciu kwantowej teorii nieadiabaticznej kształtu linii, opartej na teorii Barangera. W przypadku omawianych układów są one zgodne z przewidywaniami teorii Szudego–Baylisa. Ogólne ujęcie problematyki teorii ciśnieniowego rozszerzenia linii można znaleźć w obszernych pracach przeglądowych [64, 65].

2.1. PRZYPADEK IZOLOWANEJ LINII WIDMOWEJ

Warunkiem koniecznym do kwantowo-mechanicznego opisu zjawisk zachodzących podczas emisji promieniowania przez układ atomu emitera i n atomów zaburzaczy jest zdefiniowanie hamiltonianu całego układu. Zakładając, że oddziaływanie między zaburzaczami nie wpływa na oddziaływanie między emiterym a zaburzaczami oraz że emitery nie oddziałują ze sobą, co jest spełnione dla małych stężeń emiterów, hamiltonian takiego układu możemy zapisać w postaci:

$$H = H_E + H_Z + V(t), \quad (2)$$

gdzie H_E jest hamiltonianem niezaburzonego emitera, H_Z hamiltonianem zaburzaczy, a $V(t)$ jest zależnym od czasu operatorem oddziaływania między emiterym a zaburzaczami. Hamiltonian dany wzorem (2) będzie spełniał równania:

$$H|i\rangle = E_i|i\rangle \quad \text{oraz} \quad H|f\rangle = E_f|f\rangle, \quad (3)$$

gdzie $|i\rangle$ oraz $|f\rangle$ oznaczają odpowiednio stan początkowy i stan końcowy dla wybranego przejścia, natomiast E_i oraz E_f odpowiednie energie tych stanów.

Kształt linii łatwo jest wyrazić przez odwrotną transformatę Fouriera funkcji $\Phi(s)$:

$$J(\omega) = \frac{1}{\pi} \operatorname{Re} \int_0^{\infty} ds \Phi(s) \exp(i\omega s). \quad (4)$$

Funkcję $\Phi(s)$ nazywamy funkcją korelacji. Wyrażenie to jest ogólnym wyrażeniem na kształt linii widmowych i stanowi punkt wyjścia dla dalszych rozważań. Jednak nie można go wykorzystać w praktyce bez dodatkowych przybliżeń i uproszczeń.

Przyjmując założenia, że zaburzacze traktowane są jako cząstki poruszające się niezależnie od siebie po torach klasycznych, a oddziaływanie emitera z n zaburzacza-
mi sprowadza się do oddziaływań emitera w danej chwili tylko z jednym zaburzaczem
(zderzenia binarne), możemy występującą w równaniu (4) funkcję korelacji wyrazić
w postaci:

$$\Phi(s) = \exp(-i\omega_{if}s - Ng_{if}(s)), \quad (5)$$

gdzie ω_{if} oznacza częstość przejścia równą $(E_i - E_f)/\hbar$, N oznacza stężenie zaburza-
czy oraz spełniony jest związek

$$Ng_{if}(s) = 2\pi N \int d^3\vec{v}_{EZ} f_{\mu}(\vec{v}_{EZ}) \bar{v}_{EZ} \int_0^{+\infty} d\rho \rho \times \int_{-\infty}^{+\infty} dt_0 \left\langle 1 - U_{ii}(t_0 + s, t_0) U_{ff}^{-1}(t_0 + s, t_0) \right\rangle_{\text{Ang. Av.}}. \quad (6)$$

Występujący w tym równaniu parametr ρ jest tak zwanym parametrem zderzenia
(odległością największego zbliżenia). Rozkład prędkości atomów zaburzacza wzglę-
dem atomów emitera \bar{v}_{EZ} dany jest rozkładem Maxwella $f_{\mu}(\bar{v}_{EZ})$

$$f_{\mu}(\bar{v}_{EZ}) = \left(\frac{\mu}{2\pi k_B T} \right)^{3/2} \exp\left(-\frac{\mu v_{EZ}^2}{2k_B T} \right), \quad (7)$$

gdzie k_B jest stałą Boltzmanna, T oznacza temperaturę, natomiast μ jest masą zredu-
kowaną układu emiter–zaburzaczy.

Występujący w równaniu (6) operator $U(t_0 + s, t_0)$ (gdzie indeksy dolne ii i ff
oznaczają odpowiednio stan początkowy i końcowy przejścia) jest unitarnym operato-
rem ewolucji w czasie

$$U(t_0 + s, t_0) = \mathfrak{S} \exp \left[-\frac{i}{\hbar} \int_{t_0}^{t_0+s} \tilde{V}(t) dt \right], \quad (8)$$

przeprowadzającym układ od stanu w momencie t_0 do stanu w chwili $t_0 + s$ i zawiera-
jącym operator chronologicznego uporządkowania w czasie \mathfrak{S} , oraz operator $\tilde{V}(t)$
zdefiniowany jako

$$\tilde{V}(t) = \exp\left(\frac{i}{\hbar} H_E t \right) V(t) \exp\left(-\frac{i}{\hbar} H_E t \right). \quad (9)$$

Symbolem $V(t)$ oznaczono oddziaływanie pojedynczego zaburzacza z emiternem
w przedziale czasu od chwili t_0 do $t_0 + s$, czyli jest on jednocząstkowym operatorem
ewolucji.

Wielkość $Ng_{if}(s)$, występującą w równaniu (5), można przedstawić jako liniową funkcję s [46]

$$Ng_{if}(s) = \zeta + iX + \left(\frac{\gamma_L}{2} + i\Delta \right) s, \quad (10)$$

gdzie γ_L – oznaczającą szerokość linii i Δ – oznaczającą jej przesunięcie można wyrazić zależnością [57]

$$\frac{\gamma_L}{2} + i\Delta = 2\pi N \int d^3\vec{v}_{EZ} f_\mu(\vec{v}_{EZ}) v_{EZ} \int_0^{+\infty} d\rho \rho \left\langle 1 - S_{ii} S_{ff}^{-1} \right\rangle_{\text{Ang.Av.}}. \quad (11)$$

Z kolei asymetrię linii X oraz współczynnik ζ można przedstawić jako:

$$\begin{aligned} \zeta + iX &= 2\pi N \int d^3\vec{v}_{EZ} f_\mu(\vec{v}_{EZ}) v_{EZ} \int_0^{+\infty} d\rho \rho \\ &\times \int_{-\infty}^{+\infty} dt_0 \left\langle 1 + S_{ii} S_{ff}^{-1} - U_{ii}(t_0, -\infty) U_{ff}^{-1}(t_0, -\infty) - U_{ii}(+\infty, t_0) U_{ff}^{-1}(+\infty, t_0) \right\rangle_{\text{Ang.Av.}}, \end{aligned} \quad (12)$$

co pokazano w pracy [66]. Operatory rozpraszania S , które znajdujemy w wyrażeniach (11) i (12), spełniają zależność $S = U(+\infty, -\infty)$. Po zastosowaniu równań (10), (11) oraz (12) otrzymamy ogólne wyrażenie na profil linii widmowej

$$I(\omega) = \frac{e^{-\zeta} \cos X}{\pi} \frac{\frac{\gamma_L}{2} [1 + \mathcal{G}(\omega - \omega_0 - \Delta)]}{\left(\frac{\gamma_L}{2} \right)^2 + (\omega - \omega_0 - \Delta)^2}, \quad (13)$$

w którym zmienna $\mathcal{G} = \frac{2\text{tg}X}{\gamma_L}$ dana jest wzorem

$$\mathcal{G} = \frac{2\text{tg}X}{\gamma_L}. \quad (14)$$

Profil ten różni się od symetrycznego profilu lorentzowskiego [67] asymetryczną poprawką typu dyspersyjnego zależną od parametru asymetrii \mathcal{G} . Współczynnik ζ powoduje z kolei swoiste spłaszczenie całego profilu linii. Kształt linii w postaci (13) jest równoważny profilom uzyskanym przez Andersona i Talmana [68] i Al-Saqabi i Peach [69] uzyskanym w ramach klasycznej teorii adiabatycznej oraz profilowi uzyskanemu przez Royer [70] rozwijającemu kwantowo-mechaniczne podejście Baran-gera [56].

Dla niskich ciśnień gazu zaburzającego spełniony jest warunek $|\zeta + iX| \ll 1$, wobec czego możemy zastosować przybliżenia

$$e^{-\zeta} \cos X \approx 1 \quad \text{oraz} \quad \vartheta \approx \frac{2X}{\gamma_L}. \quad (15)$$

Profil dany równaniem (13) przy tych założeniach przekształci się do postaci

$$I(\omega) = \frac{1}{\pi} \frac{\frac{\gamma_L}{2} + X(\omega - \omega_{if} - \Delta)}{(\omega - \omega_{if} - \Delta)^2 + \left(\frac{\gamma_L}{2}\right)^2}. \quad (16)$$

Funkcja kształtu linii w tej postaci jest równoważna funkcji wynikającej z kwantowej adiabaticznej teorii Szudego i Baylisa [60–65]. W przypadku układu j nienakładających się linii o częstości generowanego promieniowania ω_j , gdzie j numeruje pary poziomów początkowych i końcowych (czyli dokonujemy przejścia $\omega_{if} \rightarrow \omega_j$), rozkład natężenia tych linii będzie dany wzorem

$$I(\omega) = \frac{1}{\pi} \frac{\frac{\gamma_L}{2} + X(\omega - \omega_{if} - \Delta)}{(\omega - \omega_{if} - \Delta)^2 + \left(\frac{\gamma_L}{2}\right)^2}. \quad (17)$$

Jeśli w przypadku pojedynczej linii ograniczymy się do przybliżenia adiabaticznego (co odpowiada założeniu, że oddziaływanie emitera z zaburzaczem możemy opisać za pomocą pojedynczych krzywych potencjalnych), to

$$\left\langle U_{ii}(t_0 + s, t_0) U_{ff}^{-1}(t_0 + s, t_0) \right\rangle_{\text{Ang. Av.}} = \exp[-i\eta(t_0 + s, t_0)]. \quad (18)$$

gdzie

$$\eta(t_0 + s, t_0) = \frac{1}{\hbar} \int_{t_0}^{t_0 + s} dt [V_i(R(t)) - V_f(R(t))]. \quad (19)$$

Wówczas szerokość γ_L i przesunięcie linii Δ możemy wyznaczyć z zależności

$$\frac{\gamma_L}{2} + i\Delta = 2\pi N \int d^3 \vec{v}_{EZ} f_{\mu}(\vec{v}_{EZ}) \vec{v}_{EZ} \times \int_0^{+\infty} d\rho \rho \{1 - \exp[-i\eta(+\infty, -\infty)]\}. \quad (20)$$

Równanie to jest identyczne z zależnościami otrzymanymi przez Lindholma [44] i Foleya [45] w ramach klasycznej teorii zderzeniowej. W przypadku adiabaticznym współczynniki ζ i X dane są wzorem [66, 68]

$$\zeta + iX = 2\pi N \int d^3\vec{v}_{EZ} f_{\mu}(\vec{v}_{EZ}) v_{EZ} \int_0^{+\infty} d\rho \rho \times \int_{-\infty}^{+\infty} dt_0 \{1 + \exp[-i\eta(+\infty, -\infty)] - \exp[-i\eta(t_0, -\infty)] - \exp[-i\eta(+\infty, t_0)]\}. \quad (21)$$

2.2. PRZYPADEK NAKŁADAJĄCYCH SIĘ LINII WIDMOWYCH

Rozważany wcześniej przypadek linii izolowanej dotyczy zwykle układów atomowych. W przypadku widm gazów molekularnych linie widmowe są najczęściej położone bardzo blisko siebie.

Opis teoretyczny kształtu nakładających się linii widmowych został podany w podejściu kwantowo-mechanicznym przez Barangera [57] i był rozwijany w kolejnych pracach przez Kolba i Griema [71], Fano [72], Ben-Reuven [73] i Gordona [74]. Na podstawie teorii Barangera, z założeniem nieskończenie krótkiego czasu zderzeń między emitерem a zaburzaciami, otrzymuje się profil będący sumą pojedynczych profili linii oraz dodatkowych wyrazów opisujących efekty interferencyjne.

Przyjmijmy, że linia j odpowiada przejściu ze stanu początkowego i_j do stanu końcowego f_j . W tej notacji ten sam stan początkowy może być oznaczony różnymi symbolami i_j lub $i_{j'}$, w przypadku kiedy przejścia j i j' mają wspólny stan początkowy. Podobna sytuacja dotyczy oznaczeń stanu końcowego, kiedy dwa różne przejścia j i j' mają wspólny stan końcowy. W przypadku nakładania się blisko siebie położonych linii (j, j', j'', \dots), zgodnie z teorią przedstawioną przez Barangera [57], ogólne wyrażenie na kształt linii widmowej możemy zapisać jako

$$I(\omega) = \text{Re Tr}[\hat{D}\hat{F}(\omega)\hat{\rho}]. \quad (22)$$

Symbol Tr oznacza ślad po wszystkich elementach w przestrzeni linii, która jest zbudowana z elementów $|i_j f_j^*\rangle$, będących wektorami stanów w przestrzeni linii.

Operator \hat{D} jest zdefiniowany przez relację

$$\langle\langle i_{j'} f_{j'}^* | \hat{D} | i_j f_j^* \rangle\rangle = \langle i_{j'} | \vec{d} | f_{j'} \rangle \langle f_j | \vec{d} | i_j \rangle = D_{j'j}, \quad (23)$$

gdzie \vec{d} jest momentem dipolowym przejścia.

Operator $\hat{\rho}$ jest zdefiniowany przez relację

$$\langle\langle i_{j'} f_{j'}^* | \hat{\rho} | i_j f_j^* \rangle\rangle = \rho_{i_j} \delta_{j'j}, \quad (24)$$

gdzie ρ_{i_j} jest macierzą gęstości dla danego stanu początkowego.

W przypadku przyjęcia założenia o nieskończeniu krótkim czasie trwania zderzeń występujący w wyrażeniu (22), operator $\hat{F}(\omega)$ możemy wyrazić jako

$$F(\omega) = \frac{1}{\pi} [\hat{\kappa} - i(\omega - \hat{\omega}_0)]^{-1}, \quad (25)$$

gdzie $\hat{\omega}_0$ jest zdefiniowany przez relację

$$\langle\langle i_j f_j^* | \hat{\omega}_0 | i_j f_j^* \rangle\rangle = \omega_j \delta_{j'j}, \quad (26)$$

a ω_j oznacza częstość przejścia $(E_i - E_f)/\hbar$. Operator $\hat{\kappa}$ występujący w wyrażeniu (25) jest zdefiniowany jako operator w przestrzeni linii

$$\hat{\kappa} = N \{1 - \hat{S}^{(i)} \otimes \hat{S}^{(f)}\}_{AV}, \quad (27)$$

i działa on na elementy przestrzeni linii w taki sposób, że

$$\langle\langle i_j f_j^* | \hat{\kappa} | i_j f_j^* \rangle\rangle = N \{ \delta_{j'j} - \langle i_j | S | i_j \rangle \langle f_j | S | i_j \rangle \}. \quad (28)$$

Po podstawieniu operatora \hat{F} danego wzorem (25) do ogólnego wyrażenia na kształt linii widmowej (22) otrzymamy wyrażenie

$$I(\omega) = \text{Re Tr}[\hat{D}\hat{F}(\omega)\hat{\rho}], \quad (29)$$

które można przedstawić w równoważnej postaci

$$I(\omega) = \sum_j \frac{P_j}{\pi} \frac{\frac{\gamma_j}{2} + Y_j(\omega - \omega_j - \Delta_j)}{(\omega - \omega_j - \Delta_j)^2 + \left(\frac{\gamma_j}{2}\right)^2}, \quad (30)$$

gdzie ω_j , γ_j , ω_j i P_j oznaczają odpowiednio niezaburzoną częstość, szerokość półkową, przesunięcie i względne natężenie wybranej linii j . Wyraz Y_j jest parametrem asymetrii spowodowanej zjawiskiem mieszania linii, który znika dla pojedynczej linii. Jak widać, otrzymany profil jest sumą symetrycznych profili lorentzowskich pochodzących od poszczególnych linii oraz asymetrycznego profilu dyspersyjnego spowodowanego nakładaniem się – mieszaniem linii. Z tego powodu wypadkowy profil będzie asymetryczny i ten typ asymetrii zwyczajowo jest nazywany *asymetrią mieszania linii* (ang. *line-mixing asymmetry*).

Parametry P_j i Y_j spełniają równanie:

$$P_j(1 - iY_j) = \sum_{j'j''} A_{jj'}^{-1} \rho_{i_j} D_{j'j''} A_{j''j}, \quad (31)$$

gdzie elementy macierzy $A_{j'j} = \langle\langle i_{j'} f_{j'}^* | \hat{A} | i_j f_j^* \rangle\rangle$ są dane wyrażeniem

$$\langle\langle i_{j'} f_{j'}^* | \hat{A}^{-1} \hat{h} \hat{A} | i_j f_j^* \rangle\rangle = \left(\frac{\gamma_j}{2} + i\Delta_j + i\omega_j \right) \delta_{j'j}, \quad (32)$$

czyli zdefiniowana w ten sposób macierz \hat{A} przeprowadza macierz h do postaci diagonalnej. Pomimo to, że wyrażenia (30) i (17) mają formalnie identyczną postać, fizyczne znaczenie parametrów asymetrii X_j i Y_j jest kompletnie różne, co może prowadzić do wyciągnięcia błędnych wniosków co do fizycznych podstaw asymetrii występującej w obserwowanych kształtach linii. Warto zauważyć, że podczas gdy parametr asymetrii zderzeniowej X_j zazwyczaj przyjmuje taki sam znak dla danego zaburzacza, wartości Y_j mogą przyjąć zarówno znak dodatni, jak i ujemny. Ponadto powinny spełniać regułę sum podaną przez Pine'a [75]:

$$\sum_j P_j Y_j = 0. \quad (33)$$

Podstawą stosowanego wcześniej przybliżenia uderzeniowego jest założenie, że zderzenia między atomami emitera i zaburzacza następują natychmiastowo, tak że czas trwania jest do pominięcia w porównaniu z odstępem czasu pomiędzy kolejnymi zderzeniami. Jak pokazano dla przypadku izolowanej linii widmowej uwzględnienie skończonego czasu trwania zderzenia jest jednak konieczne do prawidłowego opisu kształtu linii widmowej. Kształt linii widmowej uwzględniający obydwie te efekty został podany w pracy Ciuryło i Szudego [76]. W przypadku uwzględnienia skończonego czasu trwania zderzenia operator \hat{F} przyjmie postać

$$\hat{F}(\omega) = \frac{1}{\pi} [\hat{\kappa} - i(\omega - \hat{\omega}_0)]^{-1} [1 + \hat{G}(\omega)], \quad (34)$$

gdzie $\hat{G}(\omega)$ jest operatorem zależnym od czasu trwania zderzenia. Po podstawieniu wyrażenia (25) do ogólnego wyrażenia (22) na rozkład natężenia promieniowania, po wprowadzeniu kilku przybliżeń, można pokazać, że [76]

$$I(\omega) = \sum_j \frac{P_j}{\pi} \frac{\frac{\gamma_j}{2} + \chi_j(\omega - \omega_j - \Delta_j)}{(\omega - \omega_j - \Delta_j)^2 + \left(\frac{\gamma_j}{2}\right)^2}, \quad (35)$$

a parametry P_j i χ_j spełniają zależność:

$$P_j(1 - i\chi_j) = \sum_{j''j'''} A_{jj''}^{-1} (\delta_{j''j'''} + G_{j''j'''}^{(0)}) \rho_{i_j''} D_{j''j'''} A_{j''j}. \quad (36)$$

Gdy czas trwania zderzenia można pominąć, wtedy $G_{j'j}^{(0)} = 0$ i parametr χ_j można traktować jako parametr mieszania linii Y_j występujący we wzorze (30). Podobnie dla linii izolowanych element macierzowy $A_{j'j}$ staje się równy $A_{j'j}$. Wówczas wyrażenie (35) staje się identyczne ze wzorem (16) i parametr χ_j staje się odpowiednikiem parametru X_j opisującym asymetrię spowodowaną przez skończony czas trwania zderzenia. W przypadku jednoczesnego występowania obydwu tych efektów, pomimo podobieństwa wyrażeń (30) i (17) znaczenie parametru χ_j jest różne od odpowiadających mu parametrów X_j i Y_j . Ogólnie parametr χ_j nie będzie więc równy zwyczajnej sumie wyrazów opisujących asymetrię zderzeniową X_j i asymetrię mieszania linii Y_j . Jednak, jak pokazano w pracy [76], dla słabego sprzężenia linii i małych stężeń zaburzaczy prawdziwe jest przybliżenie

$$\chi_j \approx Y_j + X_j. \quad (37)$$

W tym przybliżeniu parametr χ_j staje się parametrem asymetrii dyspersyjnej, który jest sumą parametrów asymetrii zderzeniowej X_j i asymetrii mieszania linii Y_j . W praktyce w analizie danych eksperymentalnych z uwagi na podobny charakter tych dwóch typów asymetrii wyznaczany jest tylko parametr χ_j , natomiast rozróżnienie wpływu poszczególnych składowych asymetrii na kształt rejestrowanej linii i prawidłowa interpretacja wymaga dużej ostrożności. Sytuację komplikuje możliwość występowania innych typów asymetrii. Asymetria linii może być również spowodowana zależnością przesunięcia ciśnieniowego linii widmowej od prędkości emitera (tzw. efekty korelacyjne), które zostaną szerzej omówione w rozdziale 2.3. Innym źródłem asymetrii może być korelacja pomiędzy zderzeniami zmieniającymi prędkość emitera i zderzeniami zmieniającymi fazę. Problem ten był przedyskutowany przez Rautiana i Sobelmana [77], natomiast eksperymentalnie efekt ten był badany np. przez Pine'a [78, 79] dla linii HF zaburzonej argonem. Jak pokazano w pracy [80], w przypadku badanych w niniejszej pracy układów Cd–Xe i Cd–Ar efekt ten jest niemierzalny. Nieuwzględnienie w pełni struktury badanej linii także może wprowadzać dodatkową asymetrię.

2.3. WPŁYW RUCHU ATOMU EMITERA NA KSZTAŁT LINII WIDMOWEJ

Dotychczasowe rozważania nie uwzględniały efektów związanych z ruchem emitera: analiza była przeprowadzona w układzie odniesienia związanym z emiterym. W przypadku rzeczywistego eksperymentu atomy emitera nie są nieruchome w układzie odniesienia związanym z laboratorium (w którym dokonujemy pomiaru). Jeśli założymy niezależność parametrów ciśnieniowego rozszerzenia i przesunięcia linii od prędkości emitera (czyli założymy, że zjawisko ciśnieniowego rozszerzenia linii

i rozszerzenie dopplerowskie są statystycznie niezależne od siebie), to kształt linii będziemy mogli przedstawić jako:

$$I(\omega) = \int d^3\vec{v}_E f_{m_E}(\vec{v}_E) I_{\text{zderz}}(\omega - \vec{k}\vec{v}_E), \quad (38)$$

gdzie $I_{\text{zderz}}(\omega - \vec{k}\vec{v}_E)$ oznacza profil zderzeniowy przesunięty dopplerowsko o częstotliwość $\vec{k}\vec{v}_E$, wektor \vec{k} jest wektorem falowym emitowanego promieniowania, \vec{v}_E jest prędkością emitera, natomiast $f_{m_E}(\vec{v}_E)$ jest maxwellowskim rozkładem prędkości emiterów

$$f_{m_E}(\vec{v}_E) = \left(\frac{m_E}{2\pi k_B T} \right)^{3/2} \exp\left(-\frac{m_E v_E^2}{2k_B T} \right), \quad (39)$$

gdzie k_B – stała Boltzmanna, m_E – masa emitera, T – temperatura gazu. Jeśli $I_{\text{zderz}}(\omega)$ nie zależy od v_E w inny sposób niż przez przesunięcie dopplerowskie $\vec{k}\vec{v}_E$, to profil dany wyrażeniem (38) jest splotem funkcji profilu $I_{\text{zderz}}(\omega)$ opisującej profil rozszerzony zderzeniowo i profilu dopplerowskiego $I_D(\omega)$

$$I(\omega) = I_{\text{zderz}}(\omega) \otimes I_D(\omega) = \int_{-\infty}^{+\infty} d\omega' I_{\text{zderz}}(\omega - \omega') I_D(\omega'). \quad (40)$$

Rozkład dopplerowski opisany jest zależnością

$$I_D(\omega) = \frac{2\sqrt{\ln 2}}{\gamma_D \sqrt{\pi}} \exp\left[-\frac{4 \ln 2}{\gamma_D^2} (\omega - \omega_0)^2 \right], \quad (41)$$

natomiast parametr γ_D jest szerokością dopplerowską linii

$$\gamma_D = \frac{2\omega_0}{c} \sqrt{\frac{2kT \ln 2}{m_E}}. \quad (42)$$

Jeśli profil zderzeniowy $I_{\text{zderz}}(\omega)$ jest opisany profilem Lorentza, to wypadkowy kształt linii opisanej wzorem (38) jest symetryczny i jest to dobrze znany profil Voigta (VP). Jeśli przyjmiemy, że rozszerzenie zderzeniowe jest opisywane przez profil uzyskany w ramach teorii nieadiabaticznej Barangera – wzór (16), uwzględniający dyspersyjną poprawkę wprowadzającą asymetrię, to wypadkowy kształt linii będzie splotem profilu Gaussa oraz profilu będącego sumą profilu lorentzowskiego i poprawki typu dyspersyjnego. Asymetryczny profil wypadkowy jest nazywany asymetrycznym profilem Voigta (AVP) [81].

Obydwa profile VP i AVP zakładają statystyczną niezależność między ciśnieniowym a dopplerowskim rozszerzeniem linii. Jak pokazano w [82] i [83], korelacje po-

między tymi zjawiskami mogą prowadzić do odstępstw pomiędzy profilami VP i AVP, a rzeczywistymi kształtami linii, wprowadzając dodatkowe przesunięcie i asymetrię rejestrowanego profilu.

W przypadku gdy emiter porusza się w układzie laboratoryjnym, faza funkcji korelacji promieniowania (6) ulegnie zmianie spowodowanej jego ruchem w tym układzie odniesienia. Funkcję korelacji dla linii izolowanej będzie można opisać wyrażeniem [83]

$$\Phi(s) = \left\langle \exp[-i\omega_j s - i\vec{k}\vec{r} - Ng(s)] \right\rangle_{\text{Emit.Av.}}, \quad (43)$$

gdzie $\langle \dots \rangle_{\text{Emit.Av.}}$ oznacza uśrednienie po zespole statystycznym emiterów, natomiast wektor \vec{r} jest wektorem zmiany położenia emitera w czasie od t_0 do $t_0 + s$ [83–85]. Iloczyn skalarny $\vec{k}\vec{r}$ wprowadza właśnie zmianę fazy funkcji korelacji spowodowaną przez ruch emitera. Po wprowadzeniu funkcji gęstości prawdopodobieństwa $W(\vec{r}, s; \vec{v}_E)$ tego, że emiter po czasie s znajdzie się w położeniu \vec{r} , jeśli w chwili początkowej miał prędkość \vec{v}_E i znajdował się w początku układu odniesienia, będzie możliwe wykonanie uśrednienia po zespole statystycznym emiterów, zgodnie ze wzorem (43). Wówczas funkcję korelacji będziemy mogli zapisać jako [86]

$$\Phi(s) = e^{-i\omega_j s} \int d^3\vec{v}_E f_{m_E}(\vec{v}_E) \int d^3\vec{r} W(\vec{r}, s; \vec{v}_E) \times \exp\{-i\vec{k}\vec{r} - Ng(s; \vec{v}_E)\}. \quad (44)$$

Jak widać, tak funkcja korelacji $\Phi(s)$, jak i określony przez nią kształt linii widmowej $I(\omega)$ będzie zależał od wyboru postaci funkcji $W(\vec{r}, s; \vec{v}_E)$. W przypadku kiedy założymy, że ruch emitera jest ruchem swobodnym po torach prostoliniowych, funkcja ta będzie opisana wyrażeniem

$$W(\vec{r}, s; \vec{v}_E) = \delta^3(\vec{r} - \vec{v}_E s), \quad (45)$$

gdzie $\delta^3(\vec{r} - \vec{v}_E s)$ jest trójwymiarową funkcją δ -Diraca. Jeśli podstawimy funkcję korelacji $\Phi(s)$ (44) z tak zdefiniowaną funkcją gęstości prawdopodobieństwa $W(\vec{r}, s; \vec{v}_E)$ do równania (4) otrzymamy wyrażenie na profil linii w postaci

$$I_{\text{SDAVP}}(\omega) = \text{Re} \frac{1}{\pi} \int d^3\vec{v}_E f_{m_E}(\vec{v}_E) \frac{\exp[-\zeta(\vec{v}_E) - iX(\vec{v}_E)]}{\frac{\gamma_L(\vec{v}_E)}{2} - i[\omega - \omega_j - \Delta(\vec{v}_E) - \vec{k}\vec{v}_E]}. \quad (46)$$

Parametry kształtu linii $\gamma_L(\vec{v}_E)$, $\Delta(\vec{v}_E)$, $\zeta(\vec{v}_E)$ i $X(\vec{v}_E)$ są w tym wypadku zależne od prędkości emitera $\Delta(\vec{v}_E)$ [82, 83, 87] i dane są równaniami [66, 83]

$$\frac{\gamma_L(\nu_E)}{2} + i\Delta(\nu_E) = 2\pi N \int d^3\vec{\nu}_{EZ} f_{m_z}(\vec{\nu}_E + \vec{\nu}_{EZ}) \nu_{EZ} \times \int_0^{+\infty} d\rho \rho \left\langle 1 - S_{ii} S_{ff}^{-1} \right\rangle_{\text{Ang.Av.}}, \quad (47)$$

$$\begin{aligned} \zeta(\nu_E) + iX(\nu_E) &= 2\pi N \int d^3\vec{\nu}_{EZ} f_{m_z}(\vec{\nu}_E + \vec{\nu}_{EZ}) \nu_{EZ} \int_0^{+\infty} d\rho \rho \\ &\times \int_{-\infty}^{+\infty} dt_0 \left\langle 1 + S_{ii} S_{ff}^{-1} - U_{ii}(t_0, -\infty) U_{ff}^{-1}(t_0, -\infty) - U_{ii}(+\infty, t_0) U_{ff}^{-1}(+\infty, t_0) \right\rangle_{\text{Ang.Av.}}. \end{aligned} \quad (48)$$

Biorąc pod uwagę, że prędkości ruchu emitera $\vec{\nu}_E$, ruchu zaburzacza względem emitera $\vec{\nu}_E$ oraz ruchu zaburzacza $\vec{\nu}_E$ są związane relacją $\vec{\nu}_E$, występująca w podanych równaniach funkcja $f_{m_z}(\vec{\nu}_E + \vec{\nu}_{EZ})$ jest maxwellowskim rozkładem prędkości zaburzaczy. Zależność parametrów kształtu linii od prędkości emitera wyraża się przez operatory U i S , które są zależne od energii oddziaływania emiter–zaburzacz w stanach początkowych i końcowych. Z tego powodu, aby wyznaczyć parametry γ_L , Δ , ζ i X w funkcji prędkości emitera, musimy znać potencjały oddziaływania dla badanego układu emiter–zaburzacz.

Aby wyznaczyć wartości parametrów kształtu linii widmowej, należy dokonać uśrednienia wielkości opisanych w równaniach (47) i (48) po maxwellowskim rozkładzie prędkości emiterów

$$\gamma_L = \int f_{m_e}(\vec{\nu}_E) \gamma_L(\nu_E) d^3\vec{\nu}_E, \quad (49)$$

$$\Delta = \int f_{m_e}(\vec{\nu}_E) \Delta(\nu_E) d^3\vec{\nu}_E, \quad (50)$$

$$\chi = \int f_{m_e}(\vec{\nu}_E) \chi(\nu_E) d^3\vec{\nu}_E, \quad (51)$$

$$\zeta = \int f_{m_e}(\vec{\nu}_E) \zeta(\nu_E) d^3\vec{\nu}_E. \quad (52)$$

Profil opisany wyrażeniem (46) uwzględnia zarówno asymetrię zderzeniową, zależność parametrów kształtu linii od prędkości emitera, jak i korelację pomiędzy ciśnieniowym a dopplerowskim rozszerzeniem linii. Profil ten określa się w literaturze jako *zależny od prędkości asymetryczny profil Voigta* i oznacza skrótem SDAVP (pochodzącym od angielskiej nazwy *speed-dependent asymmetric Voigt profile*) [81].

W przypadku gdy masa zaburzacza jest znacznie mniejsza od masy emitera (stosunek mas $\alpha = \frac{m_z}{m_e} \ll 1$), założenie o statystycznej niezależności ciśnieniowego i dopplerowskiego rozszerzenia linii jest spełnione. Odpowiada to sytuacji, w której atomy emitera mają prędkość znacznie mniejszą od prędkości atomów zaburzacza. W takim

wypadku możemy pominąć zależność parametrów kształtu linii od prędkości emitera i zależny od prędkości profil Voigta (SDAVP) dany wzorem (46) przejdzie wtedy w asymetryczny profil Voigta (AVP) (40). W innym przypadku, gdy będziemy mogli pominąć asymetrię zderzeniową linii (przyjmując wartość $X = 0$), w wyniku otrzymamy symetryczny, zależny od prędkości profil Voigta (SDVP). Jeśli pominiemy obydwa te efekty jednocześnie, przyjmując $X = 0$ i brak zależności parametrów γ_L , Δ i ζ od prędkości emitera, otrzymamy profil Voigta (VP).

Patrząc na postać profilu SDAVP danego równaniem (46) widzimy, że obliczenie tego profilu polega na uśrednieniu po maxwellowskim rozkładzie prędkości emitera przyczynków lorentzowskich o szerokościach $\gamma_L(v_E)$, przesunięciach $\Delta(v_E)$ i symetrii $X(v_E)$. Parametry te zależą od prędkości emitera, tak więc szerokość połowkowa profilu SDAVP może być mniejsza niż szerokość asymetrycznego profilu Voigta o identycznych parametrach γ_L , Δ i X . Zależność szerokości lorentzowskiej γ_L od prędkości prowadzi do zmniejszenia szerokości dopplerowskiej γ_D . Podobny efekt może być spowodowany przez zderzeniowe zwężenie linii, tzw. zwężenie Dickiego [84]. Wynika ono z tego, że gdy średnia droga swobodna emitera między następującymi po sobie zderzeniami jest mniejsza od długości fali emitowanego promieniowania, kolejne zderzenia emitera z zaburzaczami powodują zmianę kierunku poruszania się emitera. Tak więc zmiana położenia emitera w danym przedziale czasu jest mniejsza niż w przypadku ruchu emitera po torze prostoliniowym. Powoduje to również zmniejszenie zmiany fazy funkcji korelacji (43), co z kolei skutkuje zmniejszeniem przesunięcia dopplerowskiego emitowanego promieniowania, a więc także zmniejszeniem szerokości połowkowej obserwowanej linii. Profile uwzględniające zwężenie Dickiego zostały podane w pracach Galatry'ego [88] oraz Nelkina i Ghataka [89]. Jak zostało to pokazane w pracy [80], z powodu korelacji między zderzeniami zmieniającymi prędkość i zmieniającymi fazę efekt ten jest pomijalny w badanych przez nas układach. W wypadku jednoczesnego występowania efektów korelacyjnych i zwężenia Dickiego rozróżnienie mechanizmów leżących u podstawy obserwowanego zwężenia linii wymaga szczegółowej analizy. Zależność przesunięcia ciśnieniowego linii Δ od prędkości emitera prowadzi do dodatkowej asymetrii obserwowanego kształtu linii. Korelacje pomiędzy ciśnieniowym a dopplerowskim rozszerzeniem linii są więc źródłem kolejnej asymetrii, niezależnej od asymetrii zderzeniowej czy asymetrii mieszania linii.

Widzimy więc, że w przypadku występowania korelacji pomiędzy ciśnieniowym a dopplerowskim rozszerzeniem linii ani profil VP, ani profil AVP nie opisują poprawnie rzeczywistego kształtu linii. Należy jednak zaznaczyć, że profil SDAVP (lub SDVP) będzie poprawnie opisywał kształt linii tylko w przypadku, gdy dokładnie znamy energię oddziaływania w układzie emiter–zaburzacz. Jeśli zastosujemy wartości przybliżone energii oddziaływania, otrzymany profil teoretyczny może znacząco różnić się od zarejestrowanego profilu doświadczalnego.

3. APARATURA POMIAROWA

Pomiary opisane w niniejszej pracy zostały przeprowadzone metodą Laserowo Indukowanej Fluorescencji (LIF) z użyciem cyfrowego spektrometru laserowego zbudowanego w Zakładzie Fizyki Atomowej, Molekularnej i Optycznej Instytutu Fizyki UMK.

Badanie coraz subtelniejszych efektów wpływających na rejestrowany kształt linii 326,1 nm kadmu oraz rozpoczęcie badań kształtu tej linii dla parzysto-nieparzystego izotopu ^{113}Cd spowodowało potrzebę modernizacji układu, a zwłaszcza zwiększenie dokładności pomiaru częstotliwości światła wzbudzającego fluorescencję.

3.1. BUDOWA I ZASADA DZIAŁANIA SPEKTROMETRU LASEROWEGO

Schemat układu pomiarowego przedstawiono na rysunkach 1 oraz 2. Do generacji światła wykorzystano aktywnie stabilizowany pierścieniowy laser barwnikowy firmy Coherent CR899-21 (LB) pompowany laserem argonowym Innova 400-15 (LA). Zastosowano barwnik DCM-Special (4-dicyjanometylo-2-metyl-6-4-dimetylaminystryl-4H-pyran), który umożliwia generację światła w zakresie od 632 nm do 700 nm.

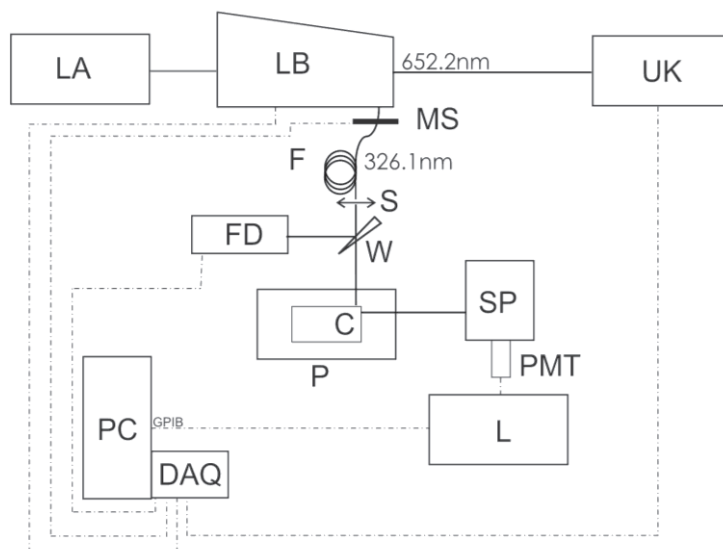
Aby otrzymać długość fali odpowiadającą badanej linii 326,1 nm ^{113}Cd ($5^3\text{P}_1-5^1\text{S}_0$), zastosowano wewnątrzwnętkowy podwójacz częstości (odpowiednio wycięty kryształ LiIO_3) otrzymujący po podwojeniu częstości długości fal w zakresie 320–350 nm. Laser ten ma możliwość przestrajania w zakresie 30 GHz (około 1 cm^{-1}) przez zmianę zewnętrznego napięcia od -5 V do $+5\text{ V}$. Laser pierścieniowy pracujący w opisanym układzie generuje dwie wiązki: podstawową o częstości ω i wiązkę podwojoną o częstości 2ω (UV).

Wiązka podwojona była skierowana do pieca, w którym umieszczono komórkę fluorescencyjną zawierającą badany izotop kadmu i gaz zaburzający (rys. 1), natomiast wiązka podstawowa była wykorzystywana do kontroli częstotliwości światła generowanego przez układ laserowy (rys. 2).

Komórka fluorescencyjna z ramieniem bocznym była umieszczona w piecu w stałej temperaturze 700 K, stabilizowanej z dokładnością do 1 K. Koniec ramienia bocznego był utrzymywany w temperaturze niższej niż reszta komórki. W warunkach pomiaru temperatura ramienia bocznego była ustalona na poziomie 411 K, co odpowiada stężeniu kadmu równemu $4 \cdot 10^{12}$ atomów w cm^3 [90].

Aby zapewnić maksymalny stosunek sygnału do szumu, odpowiednio dobrano położenie obszaru w komórce, w którym wzbudzana jest fluorescencja kadmu. Po między laserem a piecem z komórką fluorescencyjną umieszczony jest światłowód, zapewniający stosunkowo małe straty (ok. 20%) w interesującym nas obszarze widmowym. Wiązka laserowa po wyjściu ze światłowodu skierowana jest na klin odbijający stałą część wiązki w kierunku miernika mocy, gdyż podczas przestrajania lasera zmienia się moc wiązki podwojonej. Do pomiarów mocy wiązki podwojonej UV użyto fotodetektora firmy FEMTO typ DLPCA-S, odznaczającego się dużą szybko-

ścią działania, wysoką liniowością uzyskiwanego sygnału oraz odpornością na zakłócenia zewnętrzne.



Rys. 1. Schemat cyfrowego spektrometru laserowego:

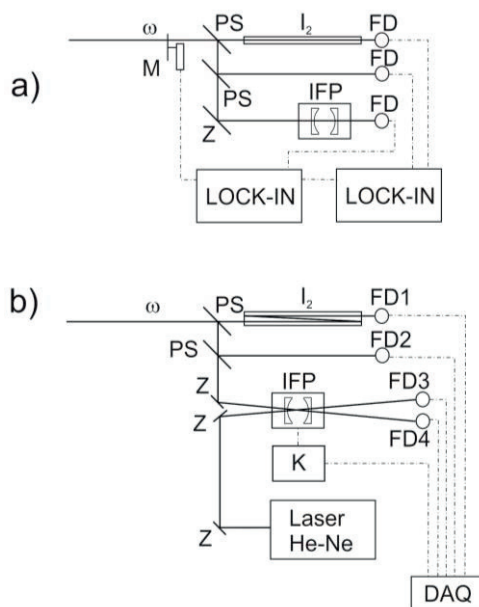
LA – laser argonowy, LB – laser barwnikowy pierścieniowy, UK – układ kalibracyjny, MS – migawka semaforowa, F – światłowód, S – soczewka, W – klin optyczny, FD – fotodetektor, C – komórka, P – piec, SP – spektrograf, PMT – fotopowielacz, L – licznik impulsów, DAQ – karta pomiarowa, PC – komputer

Fig. 1. Scheme of digital laser spectrometer:

LA – argon laser, LB – dye ring laser, UK – callibration setup, MS – shutter, F – optic fiber, S – lens, W – beam splitting wedge, FD – photodetector, C – cell, P – oven, SP – spectrograph, PMT – photomultiplier, L – pulse counter, DAQ – data aquisition system, PC – computer

Aby umożliwić pomiar promieniowania tła w układzie, zastosowano migawkę semaforową MS, umieszczoną tuż przed światłowodem. Sygnał fluorescencji kadmu był rejestrowany za pomocą fotopowielacza (FP) pracującego w trybie zliczania fotonów. Przed samym fotopowielaczem umieszczony był spektrograf siatkowy z siatką dyfrakcyjną 600 rys/mm, pracujący w niskim rzędzie ugięcia, którego zadaniem było wyeliminowanie światła rozproszonego.

Spektrometr pracuje w sposób cykliczny [30, 35]. W każdym kroku pomiarowym po przestrojeniu lasera barwnikowego na odpowiednią długość fali rejestrowane są w odpowiedniej sekwencji czasowej: natężenie sygnału fluorescencji kadmu wzbudzonej przez wiązkę podwojoną 2ω , prąd ciemny fotopowielacza, moc wiązki podwojonej 2ω oraz sygnały z układu kalibracyjnego. Następnie laser barwnikowy przestrajany jest na długość fali odpowiadającą następnemu krokowi pomiarowemu.



Rys. 2. Układ kalibracyjny: a) wersja pierwotna, b) wersja po modernizacji:

PS – płytkę światłdzielącą, Z – zwierciadło, M – modulator, I_2 – komórka absorpcyjna z parami jodu, IFP – interferometr Fabry–Perota, FD – fotodetektor, LOCK-IN – detektor fazoczuły, K – sterownik piezoelementu IFP, Laser He–Ne – laser stabilizowany częstotliwościowo, DAQ – karta pomiarowa, ω – wiązka lasera barwnikowego o częstotliwości podstawowej

Fig. 2. Calibration setup: a) previous configuration, b) after modernisation:

PS – beam splitter, Z – mirror, M – modulator, I_2 – absorption cell with molecular iodine vapor, IFP – Fabry–Perot interferometer, FD – photodetector, LOCK-IN – phase-sensitive detector, Laser He–Ne – frequency stabilised laser, DAQ – data acquisition system, ω – fundamental dye laser beam

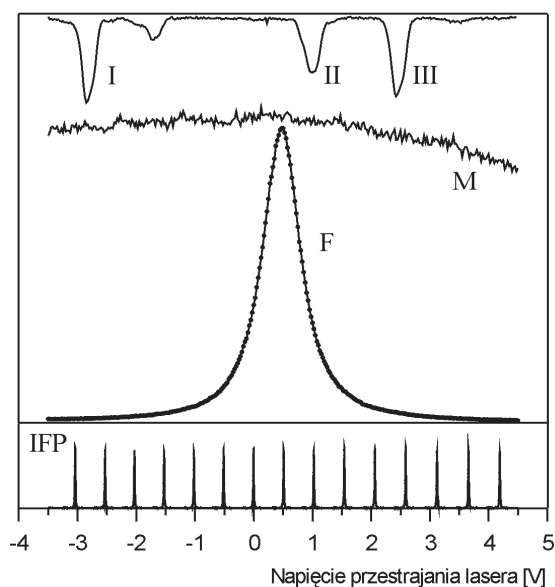
3.2. UKŁAD POMIARU CZĘSTOTLIWOŚCI ŚWIATŁA LASEROWEGO

Na rysunku 2a przedstawiono pierwotny układ kalibracyjny służący do pomiaru częstotliwości światła laserowego generowanego przez laser.

Wiązka światła o częstotliwości podstawowej wychodząca z lasera barwnikowego została podzielona za pomocą płytek światłdzielących PS na trzy wiązki. Pierwsze dwie były wykorzystane jako klasyczny układ dwuwiązkowy do pomiaru absorpcji par I_2 , a trzecia służyła do pomiaru częstotliwości względnej promieniowania lasera barwnikowego.

W trakcie każdego z kroków pomiarowych mierzone było natężenie wiązki przechodzącej przez komórkę z parami jodu oraz, natężenie wiązki referencyjnej, w efekcie czego otrzymywane było widmo absorpcyjne jodu. Wyznaczając położenie linii (przez ten termin rozumiemy napięcie lub liczbę falową odpowiadającą maksimum

absorpcji linii) jodowych w skali napięcia przestrajającego laser (patrz rys. 3) oraz porównując je z dostępnymi danymi doświadczalnymi otrzymywano bezwzględne znaczniki częstości. Liniom absorpcyjnym par I_2 widocznym na rysunku 3 według *Atlasu widma absorpcyjnego widma cząsteczki I_2* i jego erraty [91, 92] odpowiadają następujące liczby falowe: $\tilde{\nu}_I = 15327,6997 \text{ cm}^{-1}$, $\tilde{\nu}_{II} = 5328,0722 \text{ cm}^{-1}$, $\tilde{\nu}_{III} = 15328,2112 \text{ cm}^{-1}$. Trzecia część wiązki podstawowej wychodzącej z lasera była skierowana na interferometr Fabry–Perota (IFP) o długości 5 cm i przedziale dyspersji $FSR \approx 1,5 \text{ GHz}$. Interferometr podczas pomiaru nie był skanowany, odległość pomiędzy jego zwierciadłami pozostawała niezmienną. Maksyma transmisji interferometru pojawiające się podczas przestrajania lasera dawały w ten sposób znaczniki częstości, oddalone od siebie o przedział dyspersji interferometru.



Rys. 3. Zestaw danych pomiarowych otrzymywanych w wyniku pomiaru kształtu linii za pomocą cyfrowego spektrometru laserowego przed jego modernizacją: F – sygnał fluorescencji, M – profil mocy wiązki UV, IFP – sygnał transmisji nieprzestrajanego interferometru Fabry–Perota, I – III – maksima absorpcji par jodu wykorzystywane w analizie danych doświadczalnych. Rysunek zaczerpnięto z pracy [30]

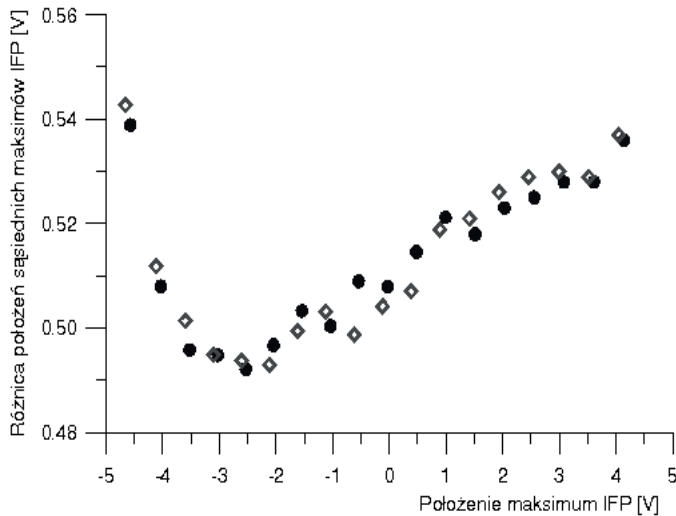
Fig. 3. Data set obtained in the line shape measurement with digital laser spectrometer before its modernization. F – fluorescence signal, M – profile of UV beam's power, IFP – transmission from not tuned Fabry–Perot interferometer, I–III – maxima of iodine vapor absorption used during data analysis. Figure taken from [30]

Detekcja wszystkich sygnałów wykonywana była z użyciem fotodiod. Sygnał transmisji był mierzony metodą detekcji fazoczułej. W tym celu przed całym układem

kalibracyjnym zainstalowany był modulator (M) typu EG&G 197, sygnały zaś rejestrowano za pomocą dwóch wzmacniaczy fazoczułych (LOCK-IN) model 7260 firmy EG&G. Sygnał po detekcji był przesyłany do komputera sterującego (PC) i tam zapisywany.

3.3. WPŁYW NIELINIOWOŚCI PRZESTRAJANIA LASERA BARWNIKOWEGO NA MIERZONY KSZTAŁT LINII WIDMOWEJ

Przestrojenie częstotliwości pracy lasera pierścieniowego CR 899-21 odbywa się za pomocą zmiany drogi optycznej we wnętrzu lasera przez obrót płytki Brewstera. Okazuje się jednak, że zmiana częstotliwości nie jest funkcją liniową kąta obrotu płytki. Również to, że płytką jest obracana przez galwoelement sprawia, że kąt obrotu płytki nie zmienia się liniowo z napięciem podawanym przez sterownik [93]. Laser Coherent CR 899-21 ma możliwość dołączenia tzw. modułu Autoscan, który daje możliwość ciągłego przestrajania się o więcej niż 30 GHz oraz ma wbudowany układ kompensujący nieliniowość przestrajania lasera. Niestety moduł Autoscan nie współpracuje z laserem wyposażonym w wewnątrzwnętkowy podwajacz częstości, co spowodowało konieczność skompensowania nieliniowości przestrajania lasera barwnikowego w inny sposób.



Rys. 4. Wykres różnicy położenia kolejnych maksimum transmisji interferometru Fabry–Perot w zależności od położenia maksimum transmisji na skali napięcia przestrajającego laser U_p .

Punktami różnych kształtów oznaczono dane uzyskane z dwóch niezależnych pomiarów.

Rysunek zaczerpnięto z pracy [30]

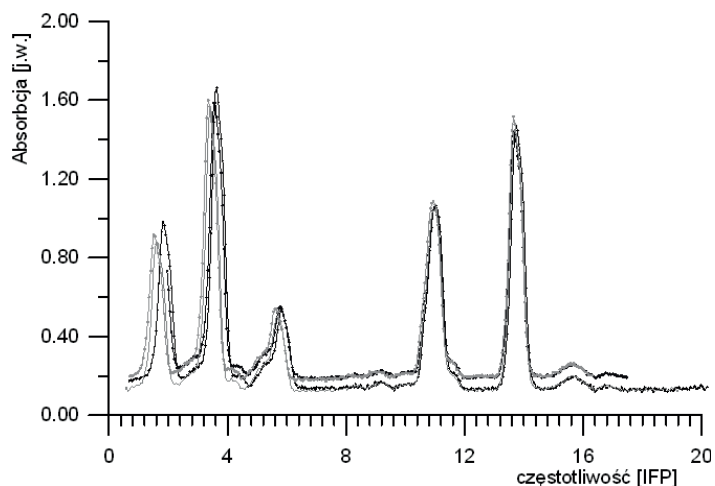
Fig. 4. Diagram of successive maxima of Fabry–Perot interferometer transmission positions differences depending on position of transmission's maximum in the scale of tuning voltage U_p .

Circles and diamonds mark data obtained from two independent measurements. Picture taken from [30]

Problem nieliniowości przestrajania laserów występuje we wszystkich laserach przestrajanych za pomocą piezoelementu. Przykładem mogą być wcześniejsze prace poświęcone kształtom linii, np. [75, 93]. Podobnie jak we wspomnianych pracach, w celu zmniejszenia wpływu nieliniowości przestrajania lasera na kształt rejestrowanej linii widmowej w zbudowanym układzie kalibracyjnym zastosowano interferometr Fabry–Perota i skierowano na niego za pomocą płytki światłodzielącej część wiązki podstawowej z lasera barwnikowego.

Na rysunku 4 przedstawiono różnice w położeniu między sąsiednimi maksimami transmisji interferometru Fabry–Perota w funkcji napięcia przestrajającego laser.

Jak widać częstotliwość pracy lasera jest nieliniową funkcją podawanego napięcia przestrajającego laser barwnikowy. W pierwszych wersjach układu kalibrującego spektrometru [95], korzystając z zarejestrowanego sygnału transmisji IFP dla każdego punktu pomiarowego, przypisywano odpowiednią liczbę falową w następujący sposób [35]: do każdego z maksimów transmisji IFP dopasowywano krzywą Gaussa, dzięki czemu wyznaczano położenie każdego z maksimów w skali napięcia przestrajającego U_p . Położenia te wyznaczają punkty równoodległe na skali częstości. Następnie dopasowywano funkcje Gaussa do maksimów linii absorpcyjnych jodu (punkty I, II i III na rys. 3), których liczby falowe są znane [91, 92]. Znajdując położenie tych maksimów można było wykreślić krzywą dyspersji.



Rys. 5. Widma cząsteczki I_2 zarejestrowane z użyciem spektrometru z pierwotnym układem kalibracji. Skala częstości wyrażona jest w jednostkach równych przedziałowi dyspersji użytego interferometru Fabry–Perota. Kolejne pomiary wykonane były w ciągu kilkunastu dni. Widoczna jest wyraźna zmiana odległości między maksimami absorpcji w zależności od pomiaru

Fig. 5. I_2 spectra obtained with spectrometer with previous calibration system. Frequency scale is given in units of free spectral range of Fabry–Perot interferometer used in the experiment.

Successing measurements were made during several days. One can notice clear change of distance between absorption maxima depending on the measurement

Trzeba tutaj zaznaczyć, że w tym miejscu zrobiono założenie o liniowej zmianie częstotliwości pomiędzy maksimami transmisji IFP, co jak pokazały dalsze badania, było niewystarczające w przypadku badania tak subtelnych efektów jak asymetria dyspersyjna układu dwóch linii położonych blisko siebie. Dobrą ilustracją tego problemu może być sytuacja przedstawiona na rysunku 5, gdzie pokazano kilka przykładów zarejestrowanego widma jodu, używanego do kalibracji osi częstotliwości. Pomiarzy wykonano w kilku przypadkowych dniach i przypadkowych porach. Zostały one nałożone na siebie, a skalę częstości przesunięto tak, aby jedna z linii jodu pokrywała się we wszystkich widmach. Jak wynika z rysunku, odległości pomiędzy liniami jodu nie były stałe. Powodem takiego wyniku były najprawdopodobniej nieskompensowana całkowicie nieliniowość przestrajania lasera barwnikowego oraz wpływ warunków zewnętrznych (zmiany ciśnienia powietrza w klimatyzowanym pomieszczeniu laboratoryjnym) na przedział dyspersji użytego interferometru Fabry–Perota.

Można więc stwierdzić, że wprowadzony sposób kalibracji skali częstotliwości zmniejszył wpływ efektu nieliniowego przestrajania lasera barwnikowego do poziomu umożliwiającego badanie takich efektów, jak rozszerzenie ciśnieniowe linii, asymetria linii spowodowana skończonym czasem zderzenia czy efekt korelacji pomiędzy ciśnieniowym a dopplerowskim rozszerzeniem linii w przypadku izolowanej linii 326,1 nm ^{114}Cd . Jednak przeprowadzone pilotażowe pomiary asymetrii dyspersyjnej w układzie ^{113}Cd –Ar [36], pomimo wykazania obecności tego typu asymetrii, zwróciły także uwagę na wpływ nieliniowości przestrajania lasera na precyzyjne wyznaczenie częstotliwości jego promieniowania, a co za tym idzie na jakość uzyskanych wyników. Niezbędna okazała się dalsza przebudowa istniejącego układu kalibracyjnego, służącego do precyzyjnego pomiaru częstotliwości względnej promieniowania emitowanego przez układ laserowy.

Schemat nowego układu kalibracyjnego przedstawiono na rysunku 2b. Idea jego działania bazuje na pomiarze w czasie rzeczywistym położenia maksimów transmisji IFP wiązki podstawowej lasera barwnikowego względem punktu nie zmieniającego swego położenia na skali częstotliwości [75]. Konstrukcja ta jest podobna do układów zastosowanych w zbudowanych częściowo przy naszym współdziałaniu na Uniwersytecie w Toronto spektrometrach laserowych ([96, 97] oraz prace w nich cytowane).

Przez konfokalny interferometr Fabry–Perota przechodzą dwie wiązki: wiązka podstawowa (o długości fali 652,2 nm) z lasera barwnikowego oraz wiązka ze stabilizowanego lasera helowo-neonowego (oznaczonego jako He–Ne na rys. 2b). Laser ten generuje promieniowanie o długości fali równej 632,8 nm i charakteryzuje się długoczasową stabilnością częstotliwości generowanego promieniowania – rzędu 3 MHz w okresie ośmiu godzin pracy. Za pomocą opisywanego układu mierzona jest częstotliwość względna wiązki podstawowej lasera barwnikowego. Aby uzyskać bezwzględną wartość częstotliwości wiązki podwojonej, czyli tej, która jest użyta do wzbudzenia par kadmu, w analizie uzyskanego widma wykorzystuje się rejestrowane

podczas pomiaru widmo absorpcyjne par I_2 oraz korzysta się z założenia o podwajaniu częstotliwości podstawowej przez kryształ $LiIO_3$ umieszczony we wnętrzu lasera barwnikowego.

Zastosowany w naszym układzie konfokalny interferometr Fabry–Perota składa się z dwóch zwierciadeł o promieniu krzywizny $R = 5$ cm, umieszczonych w odległości $d = R$. Przedział dyspersji (FSR) tego interferometru [98]:

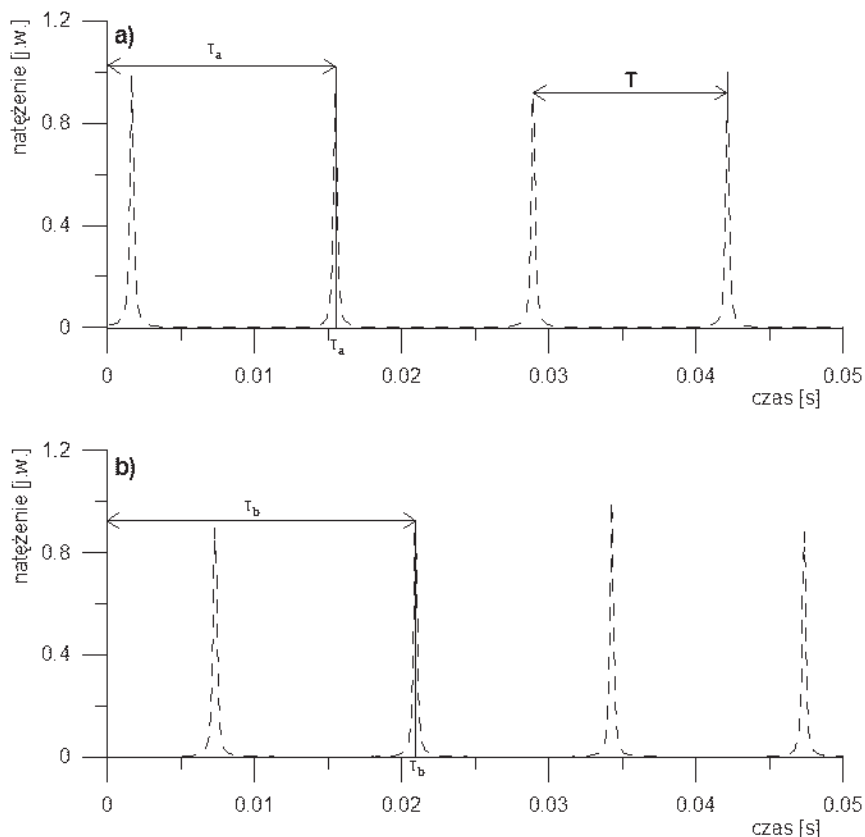
$$FSR = \frac{c}{4d}, \quad (53)$$

jest równy 1,5 GHz. Finesse rzeczywiste tego interferometru wynosi 150 [99]. Przechodzenie interferometru odbywa się przez zmianę odległości między jego zwierciadłami, a dokładniej przez zmianę położenia jednego ze zwierciadeł zamontowanych na elemencie piezoelektrycznym. Pod wpływem zmiany napięcia sterującego w zakresie 0–100 V zmienia on swoją długość o około 1 μm . Piłokształtny sygnał sterujący pochodzi ze sterownika interferometru (oznaczenie K na rys. 2b), który jednocześnie generuje synchroniczny znormalizowany sygnał TTL, wykorzystywany przez program nadzorujący proces pomiarowy. Częstotliwość przechodzenia wynosi około 16 Hz, co umożliwia dostatecznie dokładną detekcję sygnału transmisji obu wiązek przez interferometr za pomocą karty pomiarowej. Próbkowanie sygnałów odbywa się z częstotliwością 8 kHz. Wybór konfokalnego IFP podyktowany był o połowę mniejszym, niż w przypadku interferometru płaskiego, przedziałem dyspersji przy tej samej odległości zwierciadeł, ale przede wszystkim dużo mniejszą wrażliwością na ustawienie zwierciadeł względem padającej na interferometr wiązki laserowej [100, 101], co ułatwiło skierowanie wiązek z dwóch laserów do interferometru, a także ich późniejszą separację i detekcję.

Na początek rozważmy sytuację, gdy do IFP kierowana jest tylko wiązka lasera barwnikowego. Dodatkowo założymy, że przechodzenie piezoelementu jest liniowe. Na rysunku 6 przedstawiono dwa wykresy natężenia wiązki przechodzącej przez interferometr Fabry–Perota zarejestrowane w dwóch różnych chwilach. Jak widać z rysunku, zmiana położenia wybranego z maksimum interferencyjnych (na przykład drugie maksimum od lewej), wynosi $\tau_b - \tau_a$. Znając przedział dyspersji (FSR) użytego interferometru, możemy oszacować zmianę częstotliwości badanej wiązki laserowej, jaka nastąpiła pomiędzy tymi dwoma pomiarami w wyniku przestrojenia lasera:

$$\Delta\nu = \frac{\tau_b - \tau_a}{T} FSR, \quad (54)$$

gdzie T – oznacza czas, w którym interferometr Fabry–Perota zmieni swoją długość w taki sposób, aby przejść pomiędzy dwoma kolejnymi maksimumami jego transmisji. Tak więc w przypadku przedstawionym na rysunku 6 zmiana częstotliwości przechodzącej wiązki laserowej wynosi około 0,4 GHz.

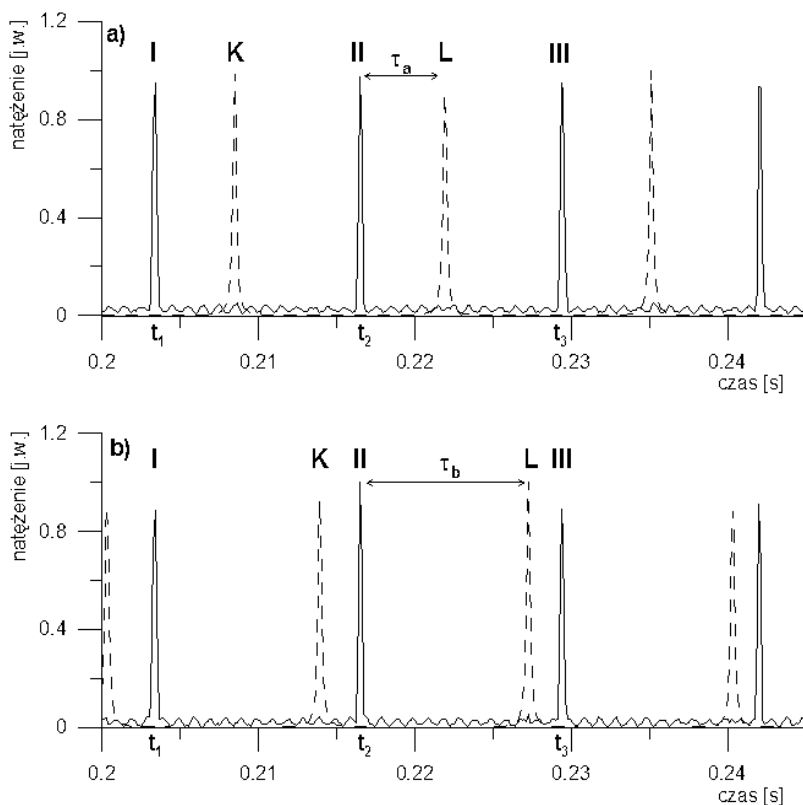


Rys. 6. Wykresy sygnału z konfokalnego interferometru Fabry–Perot po skierowaniu na niego wiązki lasera barwnikowego, otrzymane w wyniku przeprowadzenia dwóch pomiarów (a) i (b)
 Fig. 6. Diagrams of dye laser signal from confocal Fabry–Perot interferometer obtained for two measurements: (a) and (b)

Położenia poszczególnych maksimów interferencyjnych zostały określone za pomocą skali czasu, jednak nie są one powtarzalne ze względu na niestabilność odległości między zwierciadłami interferometru, zmiany temperatury i ciśnienia powietrza oraz nieregularności działania piezoelementu zmieniającego odległość między zwierciadłami interferometru. Aby określić dokładną zmianę częstotliwości promieniowania lasera barwnikowego na podstawie odczytów położenia danego maksimum interferencyjnego, potrzebny jest pewien punkt odniesienia o niezmiennym położeniu na skali częstotliwości, względem którego można byłoby określić zmianę położenia tych maksimów interferencyjnych. W opisywanym układzie rolę tę pełnią maksima interferencyjne wiązki pochodzącej ze stabilizowanego lasera helowo-neonowego (oznaczonego jako He–Ne na rys. 2b).

Na rysunku 7, analogicznie do rysunku 6, przedstawiono sygnał z konfokalnego interferometru Fabry–Perot, na który skierowano zarówno wiązkę podstawową z lase-

ra barwnikowego, jak i wiązkę pochodzącą z lasera He–Ne. Widoczne są teraz dwa zestawy maksimów interferencyjnych: oznaczone linią przerywaną interferencyjne wiązki lasera barwnikowego oraz oznaczone linią ciągłą maksima interferencyjne wiązki lasera He–Ne.

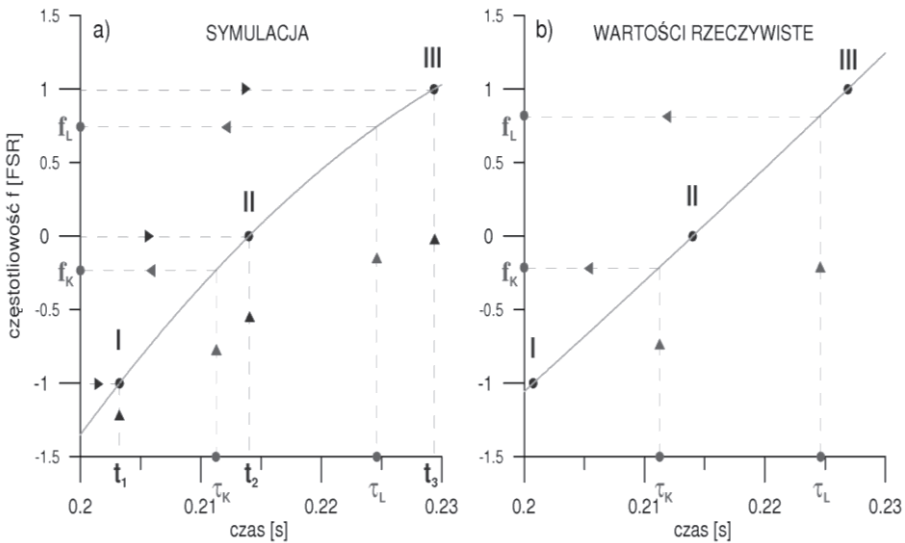


Rys. 7. Sygnały z konfokalnego interferometru Fabry–Perot dla wiązki z lasera barwnikowego (–) i dla wiązki z lasera He–Ne (—) otrzymane w wyniku przeprowadzenia dwóch pomiarów (a) i (b). t_1 , t_2 , t_3 – położenia maksimów transmisji wiązki lasera He–Ne na skali czasu

Fig. 7. Signals of dye laser (–) and He–Ne laser (—) beams from confocal Fabry–Perot interferometer obtained for two measurements (a) and (b). t_1 , t_2 , t_3 – He–Ne laser beam transmission maxima positions on the time scale

Jak widać z rysunku, zakres przestrajania interferometru obejmuje kilka rzędów interferencji, do dalszej analizy konieczne jest dokonanie wyboru zarówno jednego z maksimów interferencyjnych wiązki lasera He–Ne, który będzie służyć jako punkt odniesienia (możliwym wyborem jest na przykład maksimum oznaczone jako II na rys. 7), jak i jednego z maksimów interferencyjnych wiązki lasera barwnikowego, którego położenie będziemy śledzić (przykładowym wyborem może być maksimum L). Znajdując

odległości τ_a i τ_b pomiędzy maksimami II i L na rysunkach 7a i 7b oraz obliczając ich różnicę, jesteśmy w stanie obliczyć zmianę w położeniu maksimum L pomiędzy pomiarami przedstawionymi na rysunkach 7a i 7b. Porównując tę wartość z odległością na skali czasu odpowiadającą odległości pomiędzy kolejnymi maksimami interferencyjnymi dla wiązki z lasera barwnikowego (użyjemy tutaj zdefiniowanej wcześniej wartości T) będziemy w stanie, korzystając ze wzoru (54), wyznaczyć wartość, o jaką zmieniła się częstotliwość promieniowania generowanego przez laser barwnikowy. Otrzymana wartość będzie jednak wartością przybliżoną. Należy pamiętać, że podobnie jak w przypadku przestrajania lasera pierścieniowego, odpowiedź elementu piezoelektrycznego na przyłożone napięcie zewnętrzne nie jest liniowa. Oznacza to, że jeśli zmieniamy napięcie sterujące piezoelementem w taki sposób, aby w jednakowych odstępach czasu zmieniać napięcie sterujące o jednakowe wartości napięcia (sygnał piłokształtny), to zmiana odległości pomiędzy zwierciadłami konfokalnego IFP w funkcji czasu będzie niestety nieliniowa.



Rys. 8. Wykresy zależności częstotliwości od czasu przestrajania piezoelementu ilustrujące przejście od położenia maksimów interferencyjnych na skali czasu do skali częstości. I, II, III – położenia maksimów wiązki lasera He-Ne, τ_s, f_s – położenia maksimów fundamentalnej wiązki lasera barwnikowego, (—) – dopasowany do położenia maksimów lasera He-Ne wielomian stopnia drugiego. Na wykresie (a) wyolbrzymiono nieliniowość pracy piezoelementu, (b) – rzeczywiste dane pomiarowe
 Fig. 8. Dependence of frequency on time of piezo servo tuning showing change of interference maxima positions from time to frequency scale. I, II, III – positions of maxima of He-Ne laser beam, τ_s, f_s – positions of fundamental dye laser beam maxima, (—) – quadratic polynomial fitted to positions of He-Ne laser maxima. On diagram (a) nonlinearity of piezo servo is magnified, (b) – real data from measurement

Na rysunku 8a przedstawiono zastosowany sposób przejścia od odległości wyrażonych w skali czasu do odległości wyrażonych w skali częstotliwości.

Na osi odciętych oznaczono skalę czasu, natomiast oś rzędnych jest osią częstotliwości o jednostkach równych jednemu przedziałowi dyspersji użytego konfokalnego interferometru Fabry–Perota dla częstości promieniowania lasera He–Ne. W tak przygotowanym układzie współrzędnych zaznaczamy położenia maksimów interferencyjnych wiązki lasera He–Ne (punkty I–III). Ich współrzędne na osi czasu ($t_1 \div t_3$) wyznaczono z pomiaru sygnału z konfokalnego interferometru Fabry–Perota (rys. 7), natomiast na osi częstotliwości będą one równo oddalone od siebie o wartość równą jednemu przedziałowi dyspersji.

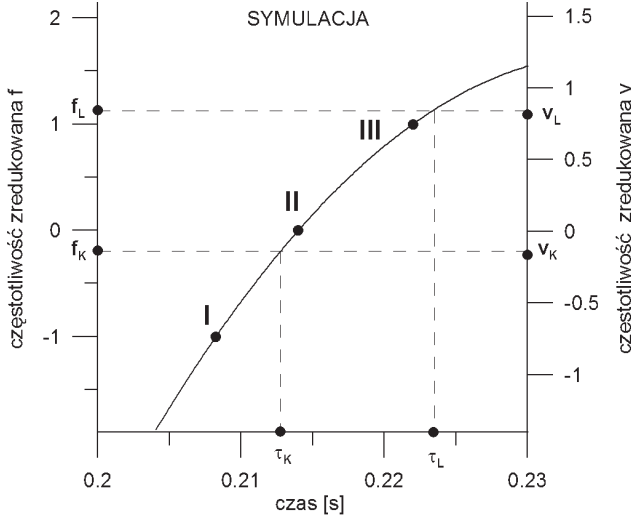
W miejscu przecięcia prostych rzutujących otrzymujemy punkty I–III określające w skali czasu i częstotliwości położenia trzech maksimów interferencyjnych z lasera He–Ne. Położenia tych trzech maksimów interferencyjnych możemy opisać za pomocą wielomianu drugiego stopnia charakteryzującego przedstawioną na rysunku 8a zależność. Otrzymamy funkcję, która pozwoli na przejście w prosty sposób od pomiarów odległości w skali czasu do pomiarów odległości w skali częstotliwości, uwzględniającą także nie w pełni liniowe przestrajanie piezoelementu interferometru. Jeżeli na osi czasu zaznaczymy teraz położenie zmierzonych maksimów transmisji lasera barwnikowego τ_K i τ_L , to za pomocą wspomnianej funkcji kwadratowej będziemy w stanie wyznaczyć ich położenia w skali częstotliwości f w jednostkach równych jednemu przedziałowi dyspersji IFP dla częstości promieniowania lasera He–Ne. Otrzymane wartości f nazwiemy dalej częstotliwością zredukowaną.

Otrzymana wartość będzie w pełni uwzględniała efekty związane z nieliniowością, jaka może wystąpić podczas przestrajania piezoelementu użytego w interferometrze. Na wykresie 8a dla czytelności rysunku zawyżono efekt nie w pełni liniowego przestrajania piezoelementu, natomiast na wykresie 8b przedstawiono rzeczywiste dane z interferogramów przedstawionych na rysunku 7. Wartości częstotliwości f_K i f_L otrzymano, korzystając z krzywej wyznaczonej na podstawie położenia maksimów transmisji wiązki lasera He–Ne na skali czasu. Warunek rezonansu we wnętrzu interferometru Fabry–Perota jest jednak zależny od długości fali. Tak więc rezonans dla wiązki lasera barwnikowego będzie występował przy zmianie odległości między zwierciadłami interferometru o inną wartość niż w przypadku lasera He–Ne. Stosunek zmiany odległości k , która powoduje przejście od jednego rezonansu do drugiego, a co za tym idzie, stosunek czasów, w jakim interferometr przestrajają się pomiędzy rezonansami dla każdej z wiązek, będzie dany wyrażeniem

$$k = \frac{\Delta d_B}{\Delta d_{\text{He-Ne}}} = \frac{\lambda_B}{\lambda_{\text{He-Ne}}} \approx \frac{T_B}{T_{\text{He-Ne}}}, \quad (55)$$

gdzie Δd_B i $\Delta d_{\text{He-Ne}}$ oznaczają zmianę odległości pomiędzy zwierciadłami interferometru, która jest wymagana, aby przejść pomiędzy kolejnymi rezonansami dla wiązki

odpowiednio lasera barwnikowego i helowo-neonowego, λ_B i $\lambda_{\text{He-Ne}}$ oznaczają długości fali obydwu laserów, natomiast T_B i $T_{\text{He-Ne}}$ oznaczają czas, w jakim interferometr przestrasza się pomiędzy kolejnymi rezonansami dla każdej z wiązek.



Rys. 9. Wykres zależności częstotliwości od czasu przestrajania piezoelementu ilustrujące przejście od położenia maksimum interferencyjnych na skali czasu do skali częstości zredukowanej. I, II, III – położenia maksimum wiązki lasera He-Ne, τ_x, f_x – położenia maksimum wiązki podstawowej lasera barwnikowego, (—) – wyznaczony z położenia maksimum lasera He-Ne wielomian stopnia drugiego. Na osi rzędnych po lewej stronie wykresu umieszczono skalę częstotliwości zredukowanej f , natomiast na osi rzędnych po prawej stronie wykresu umieszczono skalę częstotliwości zredukowanej ν , z wprowadzoną poprawką uwzględniającą różnicę w długościach fali obydwu laserów.

Dla przejrzystości rysunku stosunek długości fal obydwu wiązek ustalono na $k = 1,3$

Fig. 9. Dependence of frequency on time of piezo servo tuning showing change of interference maxima position scale from time to frequency scale. I, II, III – positions of maxima of He-Ne laser beam, τ_x, f_x – positions of fundamental dye laser beam maxima, (—) – quadratic polynomial fitted to positions of He-Ne laser maxima. On the left vertical axis there is scale of reduced frequency f , while on the right one there is scale of reduced frequency ν with correction including difference of lasers' wavelengths.

In order to make the picture clear proportion of wavelengths was assumed to be equal to 1.3

W opisanym eksperymencie parametr k jest równy 1,3. Aby otrzymać skorygowane ze względu na długość fali wartości częstotliwości zredukowanej ν_K i ν_L , należy przekształcić wartości f_K i f_L zgodnie z formułą:

$$\nu_K = f_K \frac{f_I - f_{II}}{f_K - f_L}, \quad (56)$$

gdzie wartości f_K i f_L są wartościami częstotliwości zredukowanej dla maksimum interferencyjnych wiązki lasera barwnikowego przed wprowadzeniem poprawki, a f_I i f_{II} są

położeniami dwóch kolejnych maksimum interferencyjnych wiązki lasera He–Ne, których różnica jest równa jedności.

Na rysunku 9 przedstawiono wykres analogiczny do wykresu 8, ale na osi rzędnych po prawej stronie wykresu umieszczono skalę częstotliwości zredukowanej ν , uwzględniająca różnicę w długościach fali promieniowania użytych laserów. Dla przejrzystości rysunku różnicę tę zwiększono dziesięciokrotnie (założono $k = 1, 3$).

Aby zmierzyć względną zmianę częstotliwości promieniowania emitowanego przez laser barwnikowy przestrajany w czasie $\Delta t = t_2 - t_1$, określamy położenie wybranego maksimum interferencyjnego ν_L w chwili czasu t_1 , które oznaczmy jako $\nu(1)$, i powtarzamy pomiar, aby otrzymać położenie ν_L w chwili czasu t_2 , które oznaczmy jako $\nu(2)$, a następnie korzystamy ze wzoru

$$\Delta \nu = (\nu(2) - \nu(1)) \cdot FSR. \quad (57)$$

Aby zwiększyć dokładność wyznaczania położenia maksimum interferencyjnych wiązki podstawowej lasera barwnikowego, zamiast śledzić zmiany położenia określonego maksimum transmisji interferometru (np. L) na skali częstotliwości zredukowanej ν w kolejnych punktach pomiarowych, możemy zmierzyć zmianę wartości średniej $\bar{\nu}$ z położenia dwóch sąsiednich maksimum interferencyjnych, na przykład maksimum oznaczonych jako K i jako L na rysunku 7

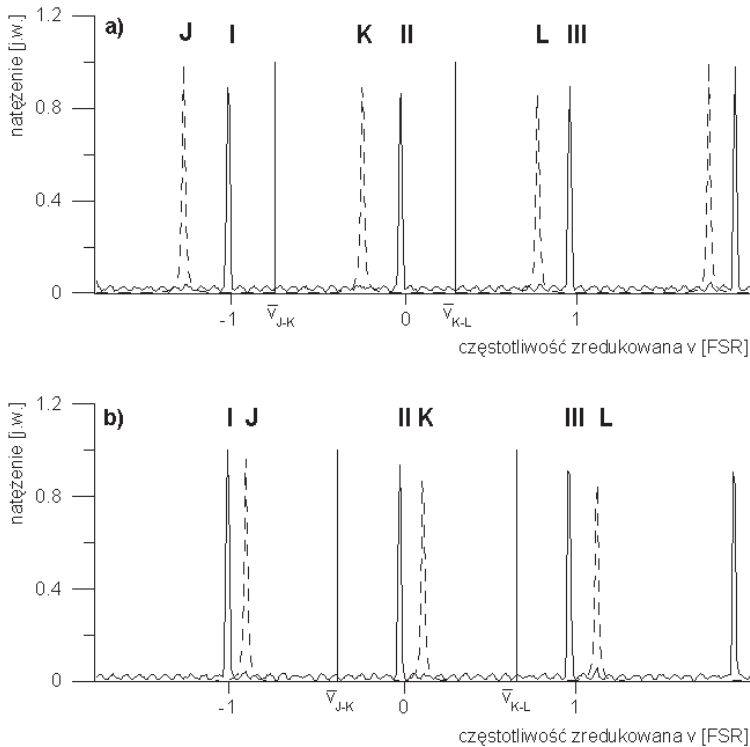
$$\bar{\nu} = \frac{\nu_L + \nu_K}{2}. \quad (58)$$

Sposób wyznaczenia względnej zmiany częstości w danym przedziale czasu nie ulega wtedy zmianie w stosunku do opisanej procedury i możemy nadal zastosować wzór (57), jeśli wartości $\nu(1)$ i $\nu(2)$ zamienimy na odpowiednie wartości średnie $\bar{\nu}(1)$ i $\bar{\nu}(2)$ dane wzorem (58).

Wynikiem działania programu sterującego dla określonego punktu pomiarowego jest wartość średnia częstotliwości zredukowanej $\bar{\nu}$ wraz z jej odchyleniem standardowym $S_{\bar{\nu}}$. Aby śledzić względną zmianę częstotliwości w przestrajaniu układu laserowego przez obszar linii widmowej, możemy to zrobić w sposób przedstawiony wcześniej, wyznaczając wartość $\bar{\nu}(i)$ w każdym z punktów pomiarowych widma numerowanych przez zmienną i . Wynik pomiaru w i -tym punkcie będzie aktualną częstotliwością zredukowaną w danym punkcie, obliczoną względem położenia na skali częstotliwości jednego z maksimum interferencyjnych wiązki lasera He–Ne, oznaczonego jako II na rysunku 7. Jeśli, zgodnie ze wzorem (57), porównamy wartość $\bar{\nu}(i)$ z wartością $\bar{\nu}(1)$ otrzymaną w pierwszym kroku pomiarowym, otrzymamy aktualną częstotliwość lasera względem częstotliwości początkowej w pierwszym punkcie pomiarowym:

$$v_i = (\bar{v}(i) - \bar{v}(1))FSR . \quad (59)$$

Jeśli jednak popatrzymy na rysunek 10a, zauważymy, że przedstawiony algorytm będzie działał dopóki nie zdarzy się sytuacja, w której jedno z maksimum lasera barwnikowego (K lub L) nie przekroczy wartości zerowej, wyznaczonej przez maksimum interferencyjne lasera He–Ne oznaczone jako II. Inaczej rzecz ujmując, nasz algorytm przestałby działać, jeżeli położenie któregośkolwiek z tych maksimum na skali częstotliwości zredukowanej na rysunku 8 przekroczyłoby zakres $(-1, 1)$. Tak więc w praktyce zastosowanie takiego układu można by było wykorzystać do pomiarów zmian częstotliwości maksymalnie o wartość równą jednemu przedziałowi dyspersji użytego interferometru, np. stabilizacji pracy lasera wokół określonej częstotliwości lub do wykonywania skanów częstotliwości lasera barwnikowego jedynie w zakresie jednego przedziału dyspersji IFP.



Rys. 10. Sygnały z konfokalnego IFP dla dwóch kolejnych punktów pomiarowych (a) i (b) w przypadku zmiany położenia jednego ze śledzonych maksimum interferencyjnych lasera barwnikowego tak, że znajdzie się ono poza przedziałem $(-1, 1)$ na skali częstotliwości
 Fig. 10. Signals from confocal IFP for two succeeding points of the scan (a) and (b) in case of change of position of one chosen interference maxima of dye laser in such a way that it goes out of range $(-1, 1)$ on frequency scale

Na rysunku 10 przedstawiono wykresy sygnałów z konfokalnego IFP właśnie w takiej sytuacji, gdy nastąpiła całkowita zmiana (np. zwiększenie) częstotliwości o więcej niż jeden przedział dyspersji. Na górnym wykresie widzimy sygnał otrzymany w wyniku pomiaru w punkcie i , skalę tworzą zaznaczone linią ciągłą maksima interferencyjne lasera He–Ne oznaczone jako I, II i III, natomiast maksima interferencyjne wiązki lasera barwnikowego oznaczone jako K i L służą nam do wyznaczenia wartości średniej $\bar{\nu}(i)$. Załóżmy, że pomiędzy pomiarami przedstawionymi na rysunkach 10a i 10b częstotliwość promieniowania zmieniła się tak, że jedno z maksimów interferencyjnych lasera barwnikowego, które było brane pod uwagę podczas wyznaczania $\bar{\nu}(i)$, znalazło się poza obszarem wyznaczonym przez maksima I–III. Widzimy, że przy większych zmianach częstotliwości nasz algorytm nie będzie działał prawidłowo. Problemem jest sposób lokalizacji maksimów interferencyjnych wiązki lasera barwnikowego, wyznaczających wartość średnią $\bar{\nu}(i)$ na skali częstotliwości. Jeśli algorytm byłby napisany w ten sposób, aby wyszukiwać maksima interferencyjne w zakresie $(-1, 1)$ na skali zredukowanych częstotliwości ν , to w wyniku działania algorytmu otrzymalibyśmy błędne wartości średniej częstotliwości zredukowanej $\bar{\nu}(i)$. Innym podejściem jest śledzenie przez program konkretnych dwóch maksimów interferencyjnych (np. numer K i L) wykorzystywanych do wyznaczenia wartości średniej $\bar{\nu}(i)$, ale wtedy byłibyśmy zmuszeni rozszerzyć naszą skalę częstotliwości opartą na maksimach interferencyjnych lasera He–Ne. Po rozszerzeniu jej tak, aby obejmowała przedział $(-2, 2)$, zakres przestrajania lasera, w którym moglibyśmy mierzyć częstotliwość wynosiłby tylko dwa przedziały dyspersji interferometru, co przy użytym w naszym eksperymencie interferometrze nie jest wartością wystarczającą (jego przedział dyspersji wynosi 1,5 GHz, natomiast zakres przestrajania lasera wynosi 30 GHz). Nie możemy także ciągle rozszerzać naszej skali częstotliwości, np. do zakresu $(-10, 10)$, ponieważ piezoelement przestrajający wnękę konfokalnego IFP zmienia swoją długość jedynie o wartość równą długości fali, co daje nam jedynie cztery maksima interferencyjne.

Możemy jednak wykorzystać fakt, że za każdym razem, gdy położenie jednego z maksimów interferencyjnych wiązki lasera barwnikowego wykroczy poza zakres $(-1, 1)$, na przykład znajdzie się w położeniu $\nu_L > 1$ w przypadku zwiększania częstotliwości pracy lasera, w zakresie $(-1, 1)$ możemy znaleźć drugie maksimum interferencyjne (maksimum J na rys. 10), którego położenie jest równe

$$\nu_J = \nu_L - 2. \quad (60)$$

Wartość średniej z położen maksimów interferencyjnych J i K zmieni się w takim przypadku na

$$\bar{\nu}_{J-K} = \bar{\nu}_{K-L} - 1. \quad (61)$$

Tak więc jeśli położenie jednego z maksimów transmisji na osi skali częstotliwości zredukowanej będzie większe od 1, to , co należałoby zrobić to zapamiętać ten fakt

i podczas obliczania całkowitej zmiany częstotliwości w danym punkcie pomiarowym do aktualnej wartości $\bar{\nu}_{J-K}$ należałoby dodać wartość jednego przedziału dyspersji.

Podobnie w przypadku przestrajania lasera barwnikowego w taki sposób, aby zmniejszała się częstotliwość promieniowania (możemy wtedy przyjąć dolny wykres na rysunku 10 jako uzyskany w pierwszym punkcie pomiarowym, a górny jako uzyskany w następnym punkcie pomiarowym). W takim przypadku położenie maksimum interferencyjnego J przyjmie wartość mniejszą od -1 , ale jednocześnie w przedziale $(0, 1)$ pojawi się nowe maksimum, oznaczone jako L . Analogicznie do poprzedniego przypadku będą spełnione zależności

$$\nu_L = \nu_J + 2, \quad (62)$$

oraz

$$\bar{\nu}_{K-L} = \bar{\nu}_{J-K} + 1. \quad (63)$$

Podobnie do poprzedniego przypadku, jeżeli położenie jednego z maksimów transmisji na osi skali częstotliwości zredukowanej będzie mniejsze od -1 , to to, co należałoby zrobić, to zapamiętać ten fakt i podczas obliczania całkowitej zmiany częstotliwości w danym punkcie pomiarowym od aktualnej wartości $\bar{\nu}_{K-L}$ należałoby odjąć wartość jednego przedziału dyspersji.

Jeśli jedna z takich sytuacji zdarzy się więcej niż raz, lub będą one zdarzać się naprzemiennie, to za każdym razem musimy odpowiednio zmniejszać lub zwiększać licznik takich przejść (oznaczymy liczbę przejść jako N). Ponieważ śledzimy położenie wartości średniej $\bar{\nu}(i)$ w kolejnych krokach pomiarowych, zdarzeniem zwiększającym lub zmniejszającym licznik N będzie przejście tej wartości przez punkt zero na skali częstotliwości oraz zmiana wartości $\bar{\nu}(i)$ o więcej niż $0,5$ (jeśli $\bar{\nu}(i)$ zmieni znak z ujemnego na dodatni, będzie to oznaczało zmniejszenie licznika o 1 , jeśli zmieni znak na ujemny, będzie to oznaczało zwiększenie licznika o 1). Będziemy wtedy w stanie śledzić względnie zmiany częstotliwości promieniowania lasera barwnikowego w dowolnie dużym zakresie, limitowanym jedynie przez jego zakres przestrajania. Ostatecznie, wyrażenie na całkowitą względną zmianę częstotliwości od pewnej częstotliwości określonej przez położenie jednego z maksimów interferencyjnych lasera He-Ne przybiera postać

$$\nu_i = (N + \bar{\nu}(i))FSR. \quad (64)$$

Jeśli w poprzednim wzorze od średniego położenia $\bar{\nu}(i)$ odejmiemy położenie $\bar{\nu}(1)$, zmierzone w pierwszym punkcie pomiarowym, otrzymamy wartości zmiany częstotliwości względem pierwszego pomiaru częstotliwości $\bar{\nu}(1)$ w naszym skanie

$$\nu_i = (N + \bar{\nu}(i) - \bar{\nu}(1))FSR. \quad (65)$$

Różnica pomiędzy użyciem wzoru (65) i (64) polega więc jedynie na różnym położeniu punktu zerowego na otrzymanej skali częstotliwości.

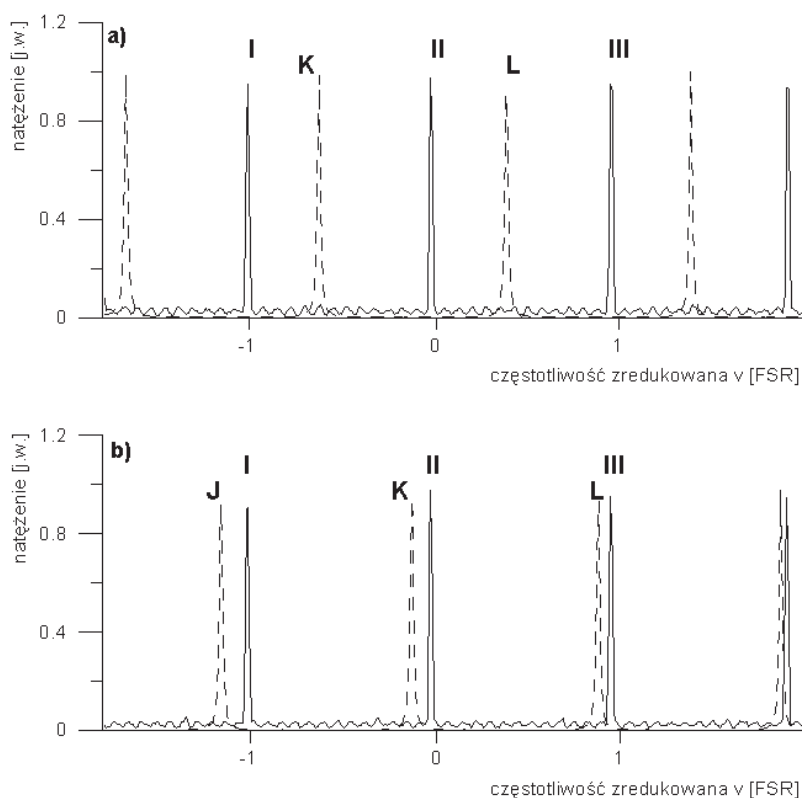
Algorytm realizujący podany sposób pomiaru częstotliwości względnej śledzi zmiany średniego położenia $\bar{v}(i)$ oraz rozpoznaje jego wielokrotne przejście przez punkt zero.

3.4. TESTY I DOBÓR WARUNKÓW POMIARU CZĘSTOTLIWOŚCI

Opisany sposób pomiaru częstotliwości względnej będzie działał w sposób prawidłowy, jeśli zostanie spełnionych kilka warunków.

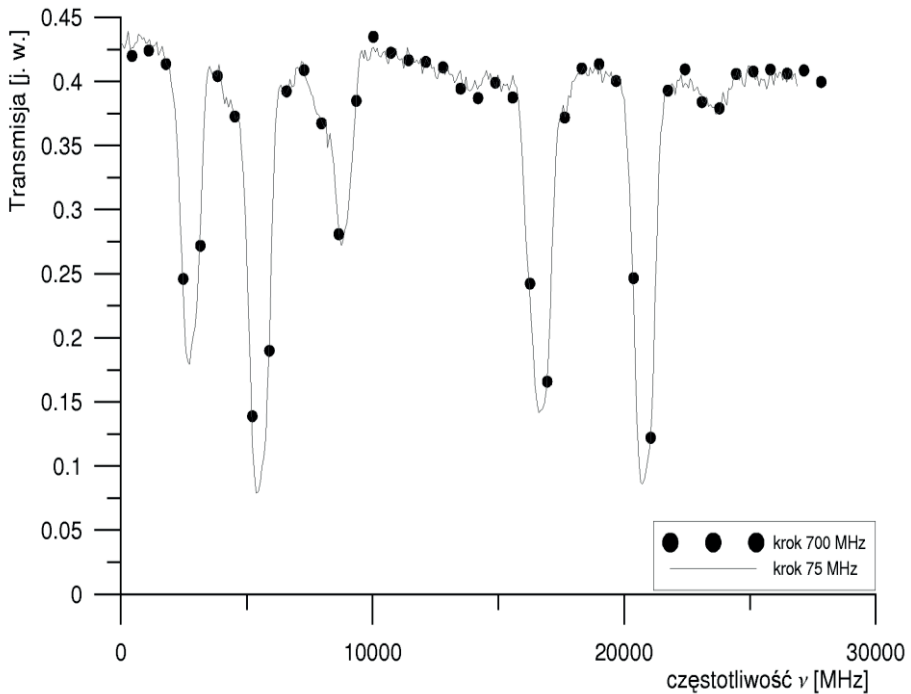
1. Problem doboru odpowiedniego kroku przestrajania lasera barwnikowego.

Rozpatrzmy sytuację przedstawioną na rysunku 11, gdzie pokazano sygnał z interferometru w dwóch kolejnych punktach pomiarowych odpowiadających przestrojeniu lasera barwnikowego o częstotliwość równą połowie przedziału dyspersji IFP. Problemem w tym wypadku jest stwierdzenie czy na dolnym wykresie oznaczone jako K



Rys. 11. Sygnał z konfokalnego IFP w przypadku ustalenia zbyt dużego kroku zmiany częstotliwości pracy lasera, równego połowie przedziału dyspersji
 Fig. 11. Signal from confocal IFP in case of too large step in change of frequency dye laser (equal to half of free spectral range)

zostało maksimum interferencyjne oznaczone jako K , czy jako L na górnym wykresie. Tak więc maksymalny pojedynczy krok, z jakim teoretycznie moglibyśmy przestrajać laser barwnikowy, wynosi $1/2 \text{ FSR} = 750 \text{ MHz}$. W naszym układzie pomiarowym, z uwagi na to, że w badaniach kształtów linii chcemy wykonać pomiar profilu badanej linii Cd z maksymalnie dużą rozdzielczością, przyjęta odległość pomiędzy kolejnymi punktami pomiarowymi na skali częstotliwości wynosi 75 MHz , co przy przedziale dyspersji interferometru równym $1,5 \text{ GHz}$ i finesse równym 150 umożliwia dokładne śledzenie położenia $\bar{\nu}$. Aby sprawdzić poprawność działania algorytmu, wykonano dwa pomiary: jeden z typowym krokiem równym 75 MHz , drugi z krokiem 700 MHz , co stanowi około 47% wartości przedziału dyspersji interferometru. Na rysunku 12 – analogicznym do rysunku 5, przedstawiono widmo absorpcji par I_2 uzyskane w każdym z pomiarów. Jak widać, wartości transmisji uzyskane w punktach pomiarowych należących do pomiaru z krokiem równym 700 MHz odpowiadają wartościom uzyskanym w pomiarze z krokiem równym 75 MHz , co potwierdza prawidłowe działanie programu.



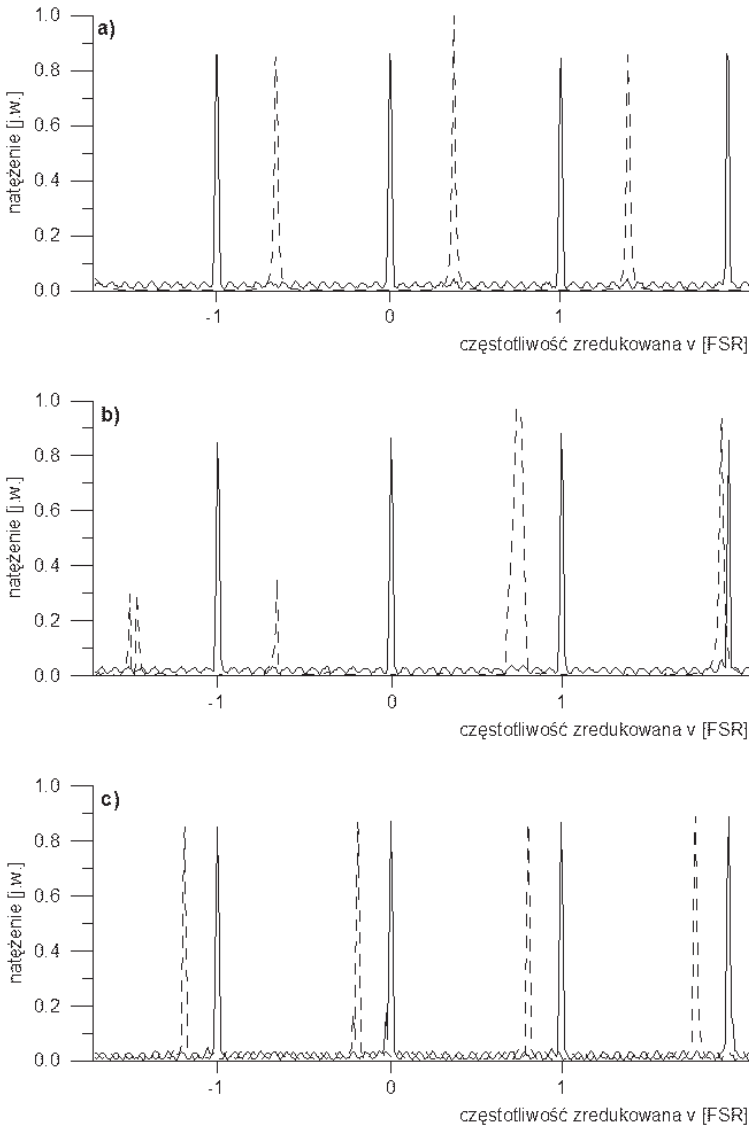
Rys. 12. Wykres sygnału transmisji wiązki podstawowej lasera barwnikowego przez pary I_2 , uzyskany w wyniku wykonania dwóch pomiarów z krokiem odpowiednio 75 MHz i 700 MHz
 Fig. 12. Diagram of transmission of the fundamental beam of dye laser through iodine vapor obtained in two measurements with frequency step 75 MHz and 700 MHz , respectively

2. Zapewnienie odpowiedniego położenia maksimów interferencyjnych wiązki lasera He–Ne.

Kolejną komplikacją mogącą wpłynąć na sposób działania układu mierzącego częstotliwość promieniowania jest możliwość otrzymania maksimów interferencyjnych wiązki lasera He–Ne w takim położeniu, w którym będą się one znajdować na początku lub końcu zakresu przestrajania konfokalnego interferometru FP. W tym wypadku zmierzone wartości położenia poszczególnych maksimów $\bar{\nu}(i)$ miałyby duże niepewności standardowe (rzędu dużej części przedziału dyspersji), co z kolei powodowałoby, że warunek na wielkość $\Delta\bar{\nu}$ w podprogramie wyznaczającym $\bar{\nu}$ nie zostałby spełniony i wymuszane byłoby powtórzenie pomiaru. Jednak łatwo jest zapobiec takiej sytuacji, sprawdzając przed każdym pomiarem położenie maksimów interferencyjnych wiązki lasera He–Ne na osi czasu. W przypadku kiedy są one położone na początku zakresu przestrajania IFP, możemy zmienić ich położenie, regulując odpowiednio składową stałą napięcia sterującego piezoelementem. Spowoduje to przesunięcie maksimów interferencyjnych na skali czasu.

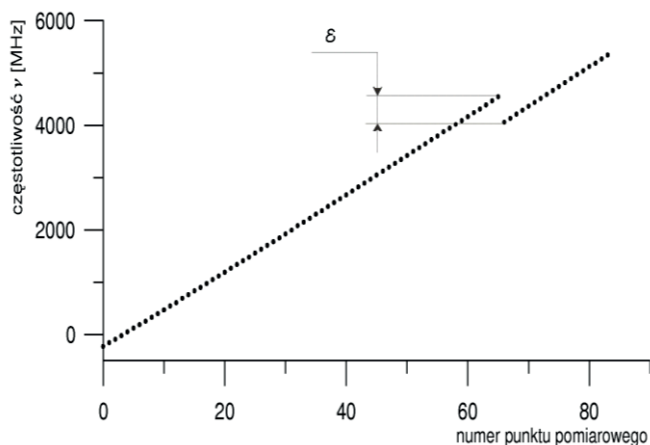
3. Zapewnienie jednomodowej pracy lasera barwnikowego.

Warunkiem koniecznym dla prawidłowego działania opisywanej procedury pomiaru częstotliwości względnej lasera barwnikowego jest także zapewnienie ciągłej zmiany jego częstotliwości. Laser typu Coherent CR-899-21 wyposażony w układ aktywnej stabilizacji pracuje jednomodowo na jednym z modów wężki rezonansowej [93, 102]. Na skutek zmiany ciśnienia powietrza w pokoju, wynikającej z pracy systemu klimatyzacji, zmian temperatury czy drgań budynku, laser może w sposób niekontrolowany przejść z pracy na jednym z modów wężki rezonansowej na pracę na innym modzie etalonu grubego [93], co w praktyce oznacza nagłą zmianę częstotliwości jego promieniowania o około 10 GHz. Sytuację taką przedstawiono na rysunkach 13 i 14. Na pierwszym rysunku przedstawiono sygnał z konfokalnego IFP. Podczas przestrajania interferometru nastąpiła niepożądana zmiana modu, na którym pracował laser barwnikowy. Spowodowało to powstanie obrazu interferencyjnego różniącego się znacząco od tego, który uzyskiwaliśmy w pracy jednomodowej. Zmierzona wartość $\bar{\nu}$ będzie charakteryzowała się dużym odchyleniem standardowym, co spowoduje wymuszenie przez program sterujący powtórzenia pomiaru w danym punkcie pomiarowym. Jeśli zmiana modu, na którym pracuje laser była jednorazowa, otrzymane wartości częstotliwości ν , pomimo że tworzą liniową skalę (patrz rys. 14), nie będą już prawidłowe, ponieważ nie ma możliwości określenia, o ile przedziałów dyspersji IFP zmieniła się częstotliwość promieniowania lasera barwnikowego. Otrzymana po skoku modu wartość częstotliwości $\bar{\nu}(i)$ jest poprzednio zmierzoną wartością $\bar{\nu}(i)$ sprzed skoku modu zwiększoną o wartością losową z przedziału $(-FSR, FSR)$ – oznaczoną na rysunku 14 jako ε . Powoduje to konieczność powtórzenia danego pomiaru kształtu linii widmowej od początku.



Rys. 13. Sygnały z konfokalnego IFP ((--)) – wiązka z lasera barwnikowego, (—) – wiązka z lasera He-Ne) w przypadku nagłej zmiany modu pracy lasera barwnikowego. Na wykresie (a) przedstawiono sygnały z IFP sprzed momentu zmiany modu, na którym pracuje laser barwnikowy, (b) – sygnały z IFP w momencie zmiany modu, (c) – sygnały z IFP po skoku modu. Przesunięcie maksimum interferencyjnych wiązki lasera barwnikowego pomiędzy rysunkami (a) i (c) jest przypadkowe

Fig. 13. Signals from confocal IFP ((--)) – dye laser beam, (—) – He-Ne laser beam) in case of sudden change of dye laser mode. On diagram (a) signal before dye laser mode change, (b) – at the moment of mode change, (c) – after mode change. Shift of interference maxima of dye laser beam between diagrams (a) and (c) is accidental



Rys. 14. Odczyty częstotliwości ν w kolejnych krokach pomiarowych, w przypadku wystąpienia zmiany modu pracy lasera barwnikowego. Symbol ϵ oznacza zmianę częstotliwości światła generowanego przez laser barwnikowy w wyniku skoku modu

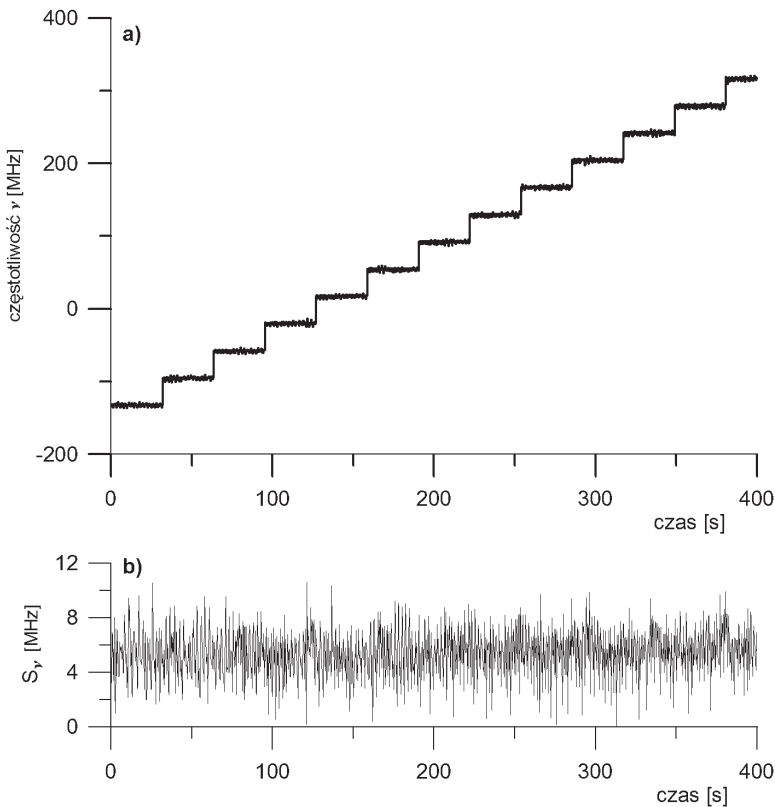
Fig. 14. Obtained values of frequency ν in successive measurement steps in case of change of dye laser mode. ϵ is frequency change of emitted light due to the change of the laser mode

Na rysunku 15 przedstawiono wykres częstotliwości względnej w czasie skokowego przestrajania lasera z krokiem 37,5 MHz (dla wiązki podstawowej), co odpowiada połowie kroku przyjętego podczas pomiarów kształtów linii. Dla większej przejrzystości rysunku 15 czas mierzenia częstotliwości w każdym z punktów pomiarowych ustalono na około 30 s, podczas gdy w przypadku rzeczywistego pomiaru wynosi on 9 s. Z rysunku wynika, że laser przestrajają się płynnie o zadaną wartość kroku i jest on stabilny na określonej częstotliwości. Na rysunku 15b przedstawiono odpowiadające poprzedniemu rysunkowi wartości niepewności standardowych poszczególnych pomiarów częstotliwości. Kształtują się one na poziomie około 2–9 MHz (wartość średnia odchylenia standardowych wynosi 5,3 MHz) i taka jest precyzja pomiarów częstotliwości wykonanych za pomocą opisanego układu. Warto zaznaczyć, że wartość ta jest kilkakrotnie mniejsza od dokładności komercyjnie dostępnych mierników długości fali. Przykładowo urządzenie WA-1500 firmy EXFO (dawniej Burleigh) dokonuje pomiaru bezwzględnego długości fali, charakteryzującego się odchyleniem standardowym pomiaru częstotliwości równym 60 MHz [103]. Czas przeprowadzenia pomiaru częstotliwości przez opisany system wynosi około 300 ms (dla porównania dla wspomnianego lambdametru WA-1500 czas ten wynosi 1 s), co jest czasem w zupełności wystarczającym dla naszych celów (w czasie testów wyznaczono optymalny czas zliczania fotonów, równy 9 sekundom dla każdego z kroków pomiarowych).

Przedstawiony sposób pomiaru częstotliwości względnej wiązki podstawowej generowanej przez laser barwnikowy umożliwia praktycznie całkowite wyeliminowanie błędów związanych z nieliniowością przestrajania tego lasera. Ponadto umożliwia on

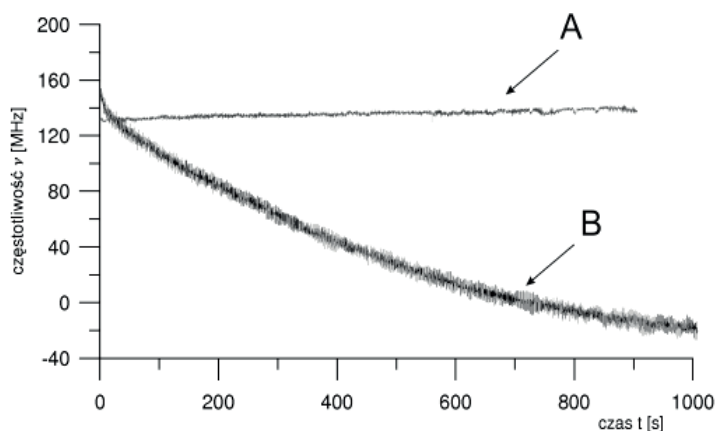
prawidłową ocenę niepewności pomiarowej otrzymywanego sygnału fluorescencji i, co za tym idzie, prawidłową interpretację wyników. Szerzej zostanie to omówione w następnym podrozdziale poświęconym analizie wykonanych pomiarów kształtu linii.

W celu uzyskania stabilnej pracy lasera barwnikowego, umożliwiającej jego przestrajanie w sposób ciągły, tak jak to przedstawiono na rysunku 15, konieczne jest jego odpowiednie przygotowanie. Pierwszym z warunków zapewniających optymalny tryb pracy jest uzyskanie dobrej jakości wiązki pompującej z lasera argonowego, tzn. wiązki o przekroju gaussowskim. Przeprowadzone testy wykazały, mimo iż laser Coherent Innova 450-15 jest wyposażony w układ *PowerTrack*, zapewniający aktywną stabilizację jego mocy wyjściowej, pełne ustabilizowanie jego pracy następuje około dwie godziny po włączeniu. Podobnie sam laser barwnikowy wymaga pewnego czasu na ustabilizowanie swojej pracy na wybranej częstotliwości w reżimie jednomodowym.

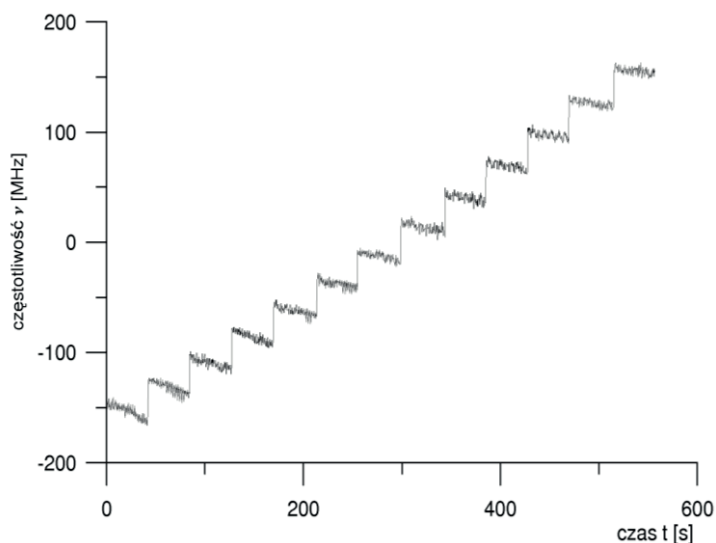


Rys. 15. Zależność częstotliwości względnej oraz jej niepewności standardowej $\Delta(\nu)$ od czasu podczas przestrajania lasera z krokiem 37,5 MHz (w przypadku wiązki 652 nm z lasera barwnikowego), co odpowiada połowie kroku stosowanego w pomiarach

Fig. 15. Dependence of relative frequency and its standard uncertainty $\Delta(\nu)$ on time while tuning the laser with step of 37.5 MHz (in case of 652 nm dye laser beam), which is half of measurement step



Rys. 16. Zależność częstotliwości od czasu przy nieprzestrzajonym laserze barwnikowym, A – pomiary częstotliwości lasera po odczekaniu dwóch godzin po włączeniu lasera barwnikowego, B – pomiary częstotliwości rozpoczęte chwilę po uruchomieniu lasera barwnikowego
 Fig. 16. Dependence of frequency on time for dye laser, A – frequency measurement started two hours after turning the dye laser on, B – frequency measurement started just after turning the dye laser on



Rys. 17. Zależność częstotliwości względnej od czasu podczas przestrajania lasera z krokiem 37,5 MHz, co odpowiada połowie kroku stosowanego w pomiarach. Rejestrację częstotliwości rozpoczęto natychmiast po włączeniu lasera barwnikowego. W przeciwieństwie do sytuacji przedstawionej na rysunku 15, częstotliwość generowanego promieniowania wykazuje dryf w każdym z kroków pomiarowych
 Fig. 17. Dependence of relative frequency on time during laser tuning with step of 37.5 MHz (half of measurement step). Data acquisition was started immediately after turning on the laser.
 In opposition to results shown in fig. 15, laser frequency drifts in every measurement step

Na rysunku 16 przedstawiono wykres częstotliwości zaraz po włączeniu lasera barwnikowego (B) oraz taki sam pomiar rozpoczęty po upływie dwóch godzin od odsłonięcia wiązki pompującej laser barwnikowy (A). Jeśli pomiar kształtu linii zostałby rozpoczęty w pierwszym okresie, jak wynika z rys. 16, laser zmieniałby częstość generowanego światła podczas każdego, trwającego dziewięć sekund kroku pomiarowego. Widać także, że w pierwszym okresie laser pracuje mniej stabilnie, o czym świadczą większe szумы częstotliwości pracy lasera na krzywej oznaczonej (B) na rysunku 16.

Przypadek ten zaprezentowano na rysunku 17, gdzie, analogicznie do rysunku 15, przedstawiono kolejne odczyty częstotliwości względnej podczas przestrajania lasera z krokiem równym 37,5 MHz. W porównaniu do poprzednio przedstawionego wykresu, pokazane wartości wyraźnie wskazują na dryf częstotliwości, na której pracuje laser w każdym z kroków pomiarowych.

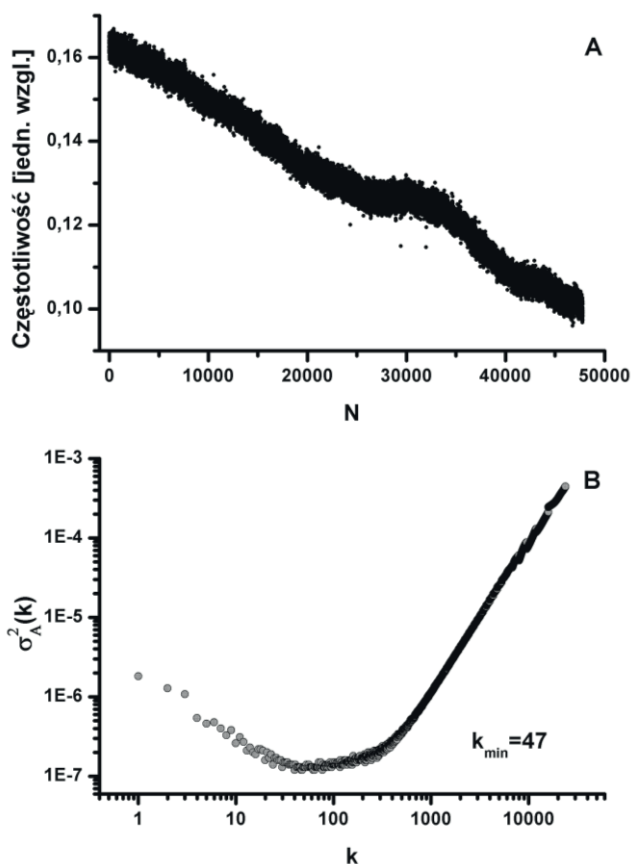
Jak zostało to opisane w poprzednim podrozdziale, laser barwnikowy jest przestrajany w zakresie 30 GHz przez podanie zewnętrznego napięcia U_p z przedziału (-5 V, 5 V). Jednak doboru kroku zmiany napięcia i co za tym idzie, rozdzielczości spektralnej, z jaką przeprowadzamy pomiar, należy dokonać z pewną ostrożnością. Z jednej strony zależy nam na maksymalnie dużej rozdzielczości dokonywanego pomiaru kształtu badanej linii widmowej, limitowanej przez szerokość spektralną światła lasera i błąd wyznaczenia jego częstotliwości, z drugiej strony zbyt duża liczba punktów pomiarowych wydłużyłaby znacząco pomiar. Byłoby to niekorzystne ze względu na trudności z utrzymaniem jednomodowej pracy lasera barwnikowego przez tak długi czas. Eksperymentalnie wyznaczono optymalną wartość kroku napięcia przestrajającego laser na 0,025 V, co odpowiada przestrojeniu wiązki podstawowej lasera barwnikowego o 75 MHz. Umożliwia to otrzymanie 400 punktów pomiarowych w pełnym zakresie przestrajania lasera. Jak pokazały liczne testy, dalsze zwiększanie rozdzielczości nie poprawiało znacząco dokładności pomiarów, natomiast przy kroku równym 75 MHz nie sprawiało problemów utrzymanie stabilnej, jednomodowej pracy lasera.

3.5. OPTYMALIZACJA POMIARU CZĘSTOTLIWOŚCI WZGLĘDNEJ LASERA BARWNIKOWEGO Z UŻYCIEM WARIANCJI ALLANA

W celu zapewnienia poprawnej pracy całego układu pomiarowego należy jeszcze dokonać wyboru optymalnego czasu pomiaru częstotliwości względnej lasera barwnikowego. Jak już wspomniano, w każdym kroku pomiarowym (przy danej częstotliwości generowanego promieniowania laserowego) wykonywany jest za pomocą układu i procedury opisany kilkukrotny pomiar częstotliwości względnej promieniowania laserowego. Należy mieć jednak na uwadze, że czas uśredniania nie może być dowolnie długi. Z rachunku błędów wiemy, że im dłużej prowadzimy pomiar danej wielkości X i im więcej wyników pomiarów wykorzystamy do obliczenia średniej arytmetycznej,

tycznej, tym dokładniej tę wielkość wyznaczymy, czyli że wariancja mierzonej wielkości będzie malała wraz ze wzrostem liczby pomiarów.

Jednak gdy mierzona wielkość ma skłonność do dryfowania, tzn. do stopniowego zwiększania (lub zmniejszania) swojej wartości podczas przejścia od jednego do drugiego kroku pomiarowego, zwiększanie liczby pomiarów nie umożliwia dokładniejszego wyznaczenia mierzonej wielkości.



Rys. 18. Wykres wariancji Allana dla serii pomiarów częstotliwości promieniowania lasera barwnikowego

Fig. 18. Allan variation plot for series of dye laser frequency measurements

Istnieje pewna skończona liczba pomiarów wartości X (skończony czas uśredniania), w której niepewność wartości mierzonej wielkości jest najmniejsza. Dalsze zwiększanie liczby pomiarów (czasu uśredniania) wpływa jedynie na pogorszenie otrzymywanych wyników. David W. Allan [104] podał sposób na znalezienie optymalnego czasu uśredniania, przy którym warto jeszcze uśredniać wyniki pomiaru.

W tym celu należy sporządzić wykres tzw. wariacji Allana $\sigma_A^2(k)$ od parametru k – liczby pomiarów uwzględnianych podczas wyznaczania wartości średniej

$$\sigma_A^2(k) = \frac{1}{2M} \sum_{n=0}^{M-1} [\delta_{n+1}(k) - \delta_n(k)]^2, \quad (66)$$

gdzie

$$\delta_n(k) = \frac{1}{k} \sum_{i=1}^k X_{nk+i}, \quad (67)$$

$n = 0, 1, \dots, M; M = (N/k) - 1; k = 1, \dots, N/2,$

gdzie N oznacza całkowitą dostępną liczbę pomiarów danej wielkości. Oznaczając jako τ odstęp czasu między kolejnymi pomiarami danej wielkości, a jako T_N całkowity czas pomiaru N próbek, optymalny czas uśredniania spełnia warunek $\tau < T_{\text{opt}} < T_N$ i dany jest zależnością $T_{\text{opt}} = k_{\text{min}}\tau$, gdzie k_{min} oznacza wartość parametru k , dla którego wariacja Allana osiąga minimum.

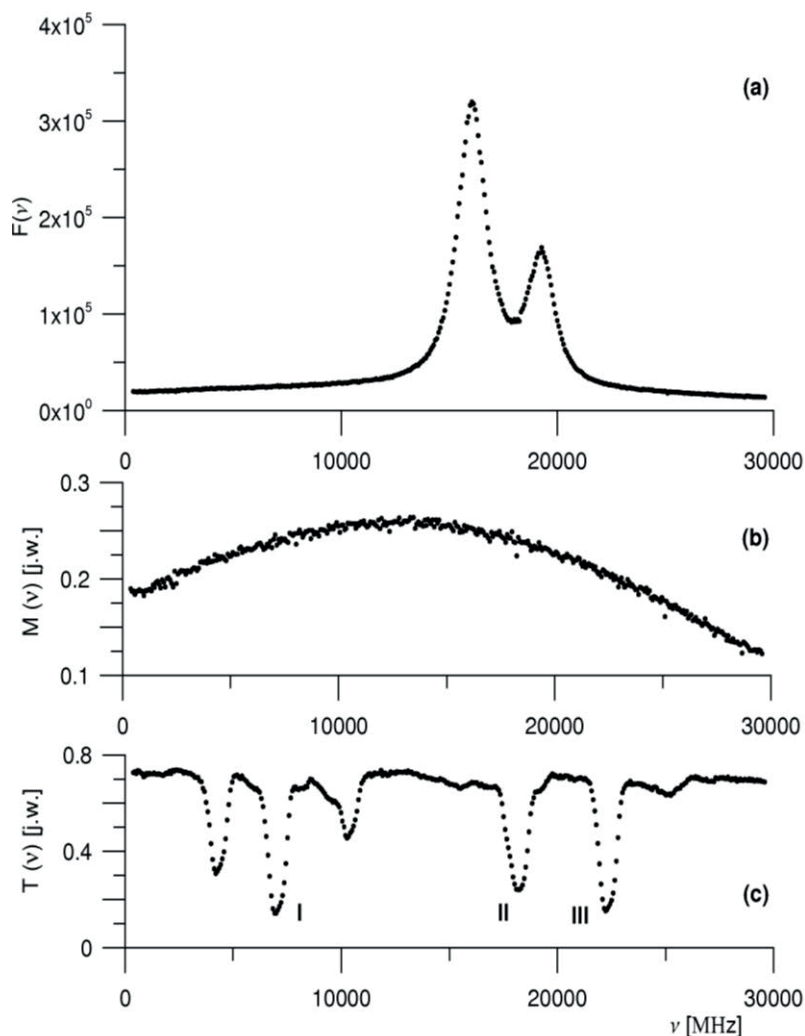
Na rysunku 18a przedstawiono przykładowy pomiar zmiany częstotliwości lasera barwnikowego w czasie. Widoczny jest ewidentny dryf częstotliwości lasera. Rysunek 18b ilustruje zmianę wariacji Allana w funkcji parametru k . Widoczne dla $k = 47$ minimum funkcji odpowiada optymalnemu czasowi uśredniania $T_{\text{opt}} \approx 6$ s, który jest zdecydowanie dłuższy od stosowanego w badaniach kształtu linii czasu pomiaru częstotliwości względnej lasera barwnikowego wynoszącego około 0,3 sekundy.

4. ANALIZA ZAREJESTROWANYCH KSZTAŁTÓW LINII WIDMOWYCH

4.1. WSTĘPNE PRZYGOTOWANIE DANYCH DOŚWIADCZALNYCH

Po przeprowadzeniu pomiaru kształtu linii za pomocą opisanego w poprzednim rozdziale spektrometru otrzymujemy następujący zestaw danych pomiarowych: liczbę zliczeń fotopowielacza przy odsłoniętej wiązce lasera – $F(\nu)$, moc wiązki padającej na fotodetektor, która jest proporcjonalna do mocy wiązki laserowej padającej na komórkę fluorescencyjną – $M(\nu)$, liczbę zliczeń fotopowielacza przy zasłoniętej wiązce laserowej – $F_s(\nu)$, na którą składa się prąd ciemny fotopowielacza oraz sygnał promieniowania tła w pomieszczeniu laboratoryjnym, sygnał z fotodetektora przy zasłoniętej wiązce laserowej, proporcjonalny do sumy szumu detektora i mocy promieniowania tła w pomieszczeniu laboratoryjnym – $M_s(\nu)$, sygnał transmisji wiązki podstawowej lasera barwnikowego przez pary jodu $I_2 - T(\nu)$.

Wszystkie wymienione wartości mierzymy w funkcji częstotliwości względnej, liczonej względem pierwszego punktu pomiarowego. Przykładowy zestaw najważniejszych sygnałów przedstawiono na rysunku 19.



Rys. 19. Zestaw części danych uzyskiwanych podczas pomiaru kształtu linii:

a) – sygnał fluorescencji $F(\nu)$, b) – krzywa mocy wiązki laserowej $M(\nu)$,
 c) – widmo transmisji dla cząsteczki jodu $T(\nu)$, I–III – linie absorpcyjne jodu

Fig. 19. Part of data set obtained during measurement of line shape:

a) – fluorescence signal $F(\nu)$, b) – laser beam power curve $M(\nu)$,
 c) – transmission spectrum of iodine molecule $T(\nu)$, I–III – iodine absorption lines

W celu przystąpienia do dalszej analizy zarejestrowanego profilu przechodzimy od względnej skali częstotliwości wyznaczonej względem pierwszego punktu pomiarowego do innej, względnej skali liczb falowych wyznaczonej względem położenia niezaburzonej linii $^{114}\text{Cd} - \tilde{\nu}_0$. W dalszej części pracy przez pojęcie „położenie” linii

będziemy rozumieli liczbę falową (lub częstotliwość) odpowiadającą maksimum składowej lorentzowskiej tej linii. Zmiana skali częstotliwości na skalę liczb falowych nie jest niezbędna, jednak umożliwia łatwiejsze porównywanie uzyskanych wyników z danymi dostępnymi w literaturze. Przeskalowania dokonujemy w dwóch krokach: w pierwszym etapie przechodzimy do bezwzględnej skali liczb falowych, wykorzystując w tym celu zarejestrowane wraz z sygnałem fluorescencji widmo cząsteczki I_2 . Przedstawione ono zostało na rysunku 19c, gdzie cyframi rzymskimi I, II i III zaznaczono trzy wybrane linie absorpcyjne wykorzystywane w analizie. Liniom tym odpowiadają następujące liczby falowe: $\tilde{\nu}_I = 15\,327,6997\text{ cm}^{-1}$, $\tilde{\nu}_{II} = 15\,328,0722\text{ cm}^{-1}$, $\tilde{\nu}_{III} = 15\,328,2112\text{ cm}^{-1}$ [91, 92]. Niestety linie te mają strukturę nadsubtelną: linie o liczbach falowych $\tilde{\nu}_I$ i $\tilde{\nu}_{III}$ mają po 15 składowych, natomiast linia $\tilde{\nu}_{II}$ ma 21 składowych [92]. Tak więc na wyznaczenie ich położenia mają wpływ niedokładności wyznaczenia liczb podanych w atlasie widma jodu (wg autorów dokładność wyznaczenia bezwzględnego położenia linii jodowej jest równa $0,002\text{ cm}^{-1}$) oraz niedokładności związanej z wyznaczeniem położenia danej linii w nierozseparowanym widmie I_2 . Oszacowanie ewentualnego maksymalnego błędu systematycznego związanego z wyznaczeniem położenia linii I_2 możemy otrzymać wyznaczając wartość resztkową parametru ciśnieniowego przesunięcia badanej linii widmowej.

W celu wykorzystania linii widma cząsteczki I_2 jako znaczników bezwzględnych liczb falowych dla wiązki podwojonej, w dalszej analizie wykorzystamy podwojone wartości liczb falowych $2 \cdot \tilde{\nu}_I = 30\,655,4106\text{ cm}^{-1}$, $2 \cdot \tilde{\nu}_{II} = 30\,656,1556\text{ cm}^{-1}$, $2 \cdot \tilde{\nu}_{III} = 30\,656,4336\text{ cm}^{-1}$. Po przejściu od skali częstotliwości ν do skali liczb falowych $\tilde{\nu}$ i jej przeskalowaniu względem trzech linii absorpcyjnych I_2 otrzymujemy bezwzględną skalę liczb falowych. Następnym krokiem jest przejście do skali względnej z zerem przyjętym na skali liczb falowych w położeniu liczby falowej odpowiadającej niezaburzonej linii izotopu ^{114}Cd . Do tego celu wykorzystujemy znaną liczbę falową dla tego przejścia $\tilde{\nu}_{\text{Cd}} = 30\,656,094\text{ cm}^{-1}$ [105]. W wyniku prostego przeskalowania otrzymujemy zestaw danych pomiarowych w zależności od różnicy liczby falowej wiązki ultrafioletowej lasera i liczby falowej linii $326,1\text{ nm }^{114}\text{Cd}$.

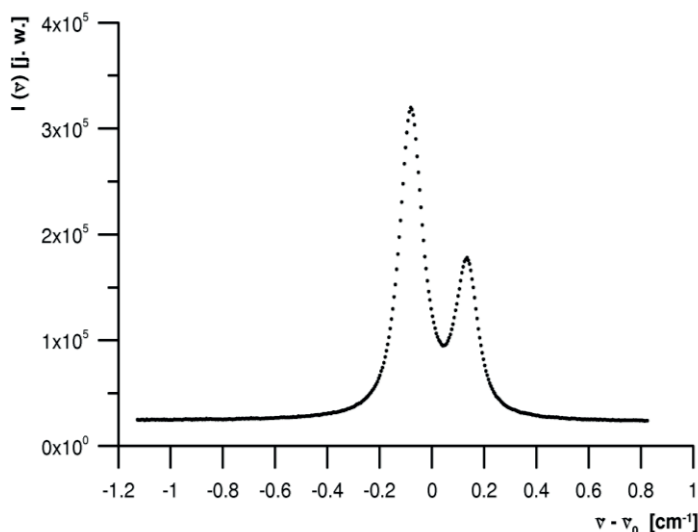
Podczas przestrajania lasera zmienia się moc wiązki laserowej, konieczna jest więc odpowiednia korekta zarejestrowanego profilu linii, biorąca pod uwagę zależność mocy wiązki podwojonej od liczby falowej generowanego promieniowania. Najłatwiej to zrobić, wykorzystując zmierzony profil mocy (wykres b na rysunku 19), przez który dzielimy zarejestrowany sygnał fluorescencji (wykres a na rysunku 19) i mnożymy wynik dzielenia przez moc wiązki ultrafioletowej dla liczby falowej odpowiadającej maksimum linii $M(\tilde{\nu}_m)$, tak aby zachować informację o liczbie zliczeń fotopowielacza w maksimum badanej linii

$$I(\tilde{\nu}) = \frac{F(\tilde{\nu})}{M(\tilde{\nu}_m)} M(\tilde{\nu}_m). \quad (68)$$

Aby wziąć pod uwagę także ewentualny wpływ resztkowego promieniowania w pomieszczeniu laboratoryjnym na kształt badanej linii (np. światło pochodzące od kontrolki urządzeń pomiarowych, świecenie rozgrzanego pieca itp.), wykorzystujemy sygnały $F_t(\tilde{\nu})$ i $M_t(\tilde{\nu})$ zarejestrowane przy zasłoniętej wiązce laserowej. Tak więc korekta zarejestrowanego kształtu linii ma ostatecznie formę:

$$I(\tilde{\nu}) = \frac{F(\tilde{\nu}) - F_t(\tilde{\nu})}{M(\tilde{\nu}) - M_t(\tilde{\nu})} M(\tilde{\nu}_m). \quad (69)$$

Kształt linii widmowej po zastosowaniu opisanej korekty, przedstawiony na rysunku 20, jest punktem wyjścia do dalszej analizy.



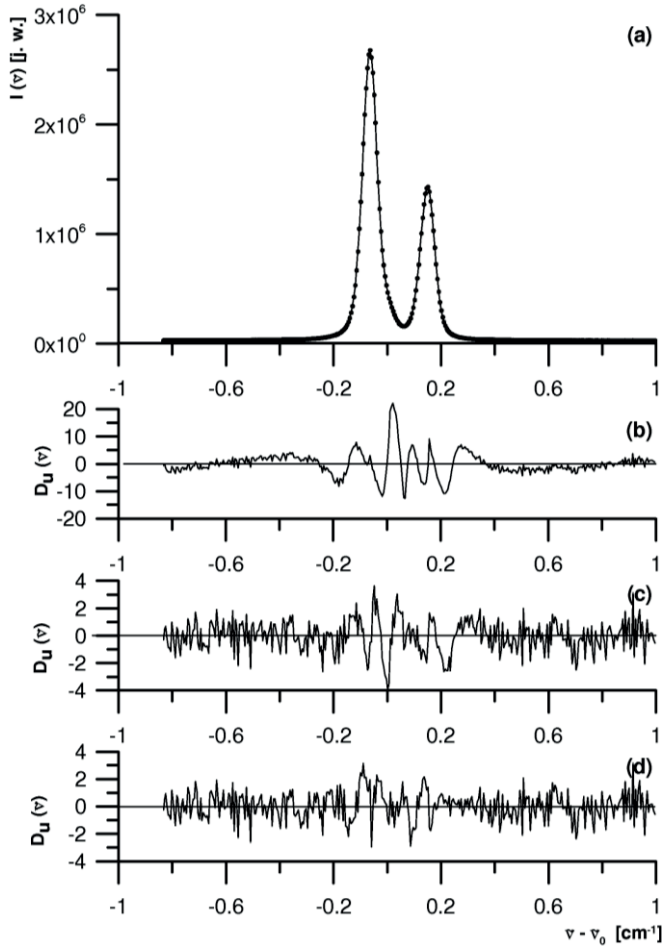
Rys. 20. Przykładowy profil zarejestrowanej linii 326,1 nm ^{113}Cd zaburzonej przez ksenon po przeprowadzeniu wstępnej analizy danych

Fig. 20. Example of recorded profile of 326.1 nm ^{113}Cd line perturbed by xenon after initial data analysis

4.2. ANALIZA KSZTAŁTU LINII

Stosowana w badaniach kształtu linii widmowych analiza zmierzonych profili doświadczalnych $I(\tilde{\nu})$ polega na dopasowaniu do nich różnych profili teoretycznych, biorących pod uwagę efekty fizyczne wpływające na kształt badanej linii i wyznaczeniu parametrów tych dopasowań. Otrzymane zależności wartości wyznaczonych parametrów od koncentracji gazu zaburzającego umożliwiają obserwację wpływu (lub jego braku) poszczególnych efektów na obserwowany kształt linii. W procedurze dopasowywania użyto powszechnie stosowanego algorytmu Marquardta [106, 107]. Pomiar

asymetrii dyspersyjnej przeprowadzono na przykładzie przejścia $5^3P_1 - 5^1S_0$ w parzysto-nieparzystym izotopie ^{113}Cd . Ze względu na niezerowy spin jądra ($I = 1/2$) występuje rozszczepienie nadsubtelne badanej linii 326,1 nm ^{113}Cd na dwie składowe odpowiadające przejściom $F' = 1/2 \rightarrow F'' = 1/2$ oraz $F' = 3/2 \rightarrow F'' = 1/2$.



Rys. 21. Profil linii 326,1 nm ^{113}Cd zaburzonej przez ksenon pod ciśnieniem 42 Tr, a – profil doświadczalny (\bullet) i dopasowany do niego profil teoretyczny biorący pod uwagę wszystkie składowe izotopowe (—), różnice ważone $D_u(\tilde{\nu})$ pomiędzy profilem doświadczalnym a profilem teoretycznym uwzględniającym: b – jedynie 2 składowe nadsubtelne izotopu ^{113}Cd , c – izotopy ^{113}Cd i ^{114}Cd , d – osiem składowych izotopowych o zawartości zgodnej z tabelą 1

Fig. 21. Profile of 326.1 nm ^{113}Cd line perturbed by xenon at pressure 42 Tr, a – experimental (\bullet) and fitted theoretical profile taking into consideration all isotope components (—), weighted residuals $D_u(\tilde{\nu})$ between experimental and theoretical profile taking into consideration: b – only 2 hyperfine components of ^{113}Cd isotope, c – isotopes ^{113}Cd and ^{114}Cd , d – eight isotope components of abundance consistent with data in table 1

Do zdefiniowanego wzorem (69) skorygowanego profilu doświadczalnego dopasowywano profil teoretyczny dany wzorem

$$I_{\text{dop}}(\tilde{\nu}) = I_r + \sum_{i=1}^m b_i I_i(\tilde{\nu} - d_i), \quad (70)$$

będący sumą dopasowanych odpowiednich profili teoretycznych $I_i(\tilde{\nu})$, gdzie d_i oznacza przesunięcie izotopowe lub nadsubtelne danej składowej linii, b_i – natężenie danej składowej linii (gdzie $\sum_i b_i = 1$), I_r – natężenie promieniowania światła rozproszonego. Okazało się jednak, że po przyjęciu liczby składowych $m = 2$ (dla obydwu składowych nadsubtelnych) i dopasowaniu profilu teoretycznego do danych doświadczalnych otrzymano wyraźne niedopasowanie w centrum linii (przedstawiono to na rys. 21b). Do porównania jakości dopasowań użyto tak zwanych różnic ważonych $D_u(\tilde{\nu})$, zdefiniowanych jako iloraz różnicy $D(\tilde{\nu}) = I_{\text{exp}}(\tilde{\nu}) - I_{\text{dop}}(\tilde{\nu})$ oraz niepewności standardowej sygnału w danym punkcie – $u_1(\tilde{\nu})$ [108].

Podobne zachowanie było widoczne także podczas pomiarów linii 326,1 nm ^{114}Cd [30, 35]. Okazało się więc konieczne uwzględnienie pozostałych, śladowych składowych izotopowych linii Cd. Na rysunku 21c pokazano różnice ważne $D_u(\tilde{\nu})$ otrzymane w przypadku dopasowania do danych doświadczalnych profilu $I(\tilde{\nu})$ uwzględniającego trzy linie składowe (dwie składowe nadsubtelne izotopu ^{113}Cd i jedną izotopu ^{114}Cd , których łączna zawartość w użytej próbce kadmu według producenta wynosiła 98,27%). W tym przypadku widać znaczącą poprawę w stosunku do wykresu 21b, jednak w centrum linii w dalszym ciągu widoczne jest niedopasowanie. Wykres 21d przedstawiający różnice ważne $D_u(\tilde{\nu})$, otrzymane z uwzględnieniem ośmiu składowych izotopowych, daje najlepsze dopasowanie. W każdym z opisywanych dopasowań za wartości względne natężeń każdej ze składowych przyjęto wartości składu procentowego użytej próbki Cd podane w atęście dostarczonym przez producenta (wartości te przedstawiono w tab. 1).

Tabela 1. Skład izotopowy kadmu użytego w pomiarach linii 326,1 nm ^{113}Cd według danych producenta [109]

Izotop	106	108	110	111	112	113	114	116
Zawartość [%]	<0,01	<0,01	0,12	0,12	1,28	95,8	2,47	0,21

W dalszym ciągu tej pracy przez pojęcie dopasowania profilu będziemy rozumieli dopasowanie do danych doświadczalnych funkcji danej wzorem (70), gdzie I_i oznacza konkretny profil teoretyczny, w którym uwzględniono wszystkie składowe izotopowe oraz rozszczepienie nadsubtelne linii 326,1 nm izotopów nieparzystych (czyli przyjęto $m = 10$).

W celu zbadania roli zderzeń, efektów związanych z korelacją pomiędzy ciśnieniowym a dopplerowskim rozszerzeniem linii oraz wpływu mieszania linii, do danych doświadczalnych dopasowywano cztery rodzaje profili:

- profil Voigta (VP) – najprostszy ze stosowanych profili. Jest on symetryczny, uwzględnia ciśnieniowe rozszerzenie i przesunięcie linii oraz zakłada brak korelacji pomiędzy ciśnieniowym a dopplerowskim rozszerzeniem linii. Umożliwia wyznaczenie szerokości składowej dopplerowskiej γ_D , szerokości składowej lorentzowskiej γ_L oraz ciśnieniowego przesunięcia linii Δ ,

- asymetryczny profil Voigta (AVP) – profil uwzględniający asymetrię dyspersyjną (opisującą zarówno zderzeniową jak i związaną z efektem mieszania linii), w którym oprócz parametrów γ_D , γ_L i Δ dopasowywany jest także parametr asymetrii χ ,

- zależny od prędkości profil Voigta (SDVP) – uwzględniający efekt korelacji pomiędzy ciśnieniowym a dopplerowskim rozszerzeniem linii,

- zależny od prędkości asymetryczny profil Voigta (SDAVP) – uwzględniający zarówno efekty związane z korelacją pomiędzy ciśnieniowym a dopplerowskim rozszerzeniem linii, jak i asymetrię profilu.

Ostatni z wymienionych profili – SDAVP – jest dany wzorem (46).

Numerycznie wygodniejsze jest jednak użycie profilu SDAVP w postaci podanej przez m.in. Ciuryło [110]

$$I_{\text{SDAVP}}(\omega) = \frac{2}{\pi^{3/2} \omega_D} \int_{-\infty}^{+\infty} e^{-x^2} x \times \left\{ \arctan[A(x)] + \frac{\chi B_A(\alpha; x)}{2} \ln[1 + A^2(x)] \right\} dx, \quad (71)$$

gdzie

$$A(x) = \frac{\omega - \omega_0 - \Delta B_S(\alpha; x) + x \omega_D}{\frac{\gamma_L}{2} B_W(\alpha; x)}. \quad (72)$$

Wielkość $\omega_D = \omega_0 v_{m_E} / c$ jest związana z szerokością składowej dopplerowskiej profilu linii γ_D zależnością:

$$\omega_D = \frac{\gamma_D}{2\sqrt{\ln 2}}. \quad (73)$$

Funkcje $B_W(\alpha; x)$, $B_S(\alpha; x)$, $B_A(\alpha; x)$ opisują zależność szerokości lorentzowskiej γ_L , przesunięcia ciśnieniowego Δ i asymetrii χ od prędkości emitera. Są one funkcjami stosunku masy zaburzacza do masy emitera α oraz prędkości zredukowanej emitera x , danej wzorem

$$x = \frac{v_E}{v_{m_E}}, \quad (74)$$

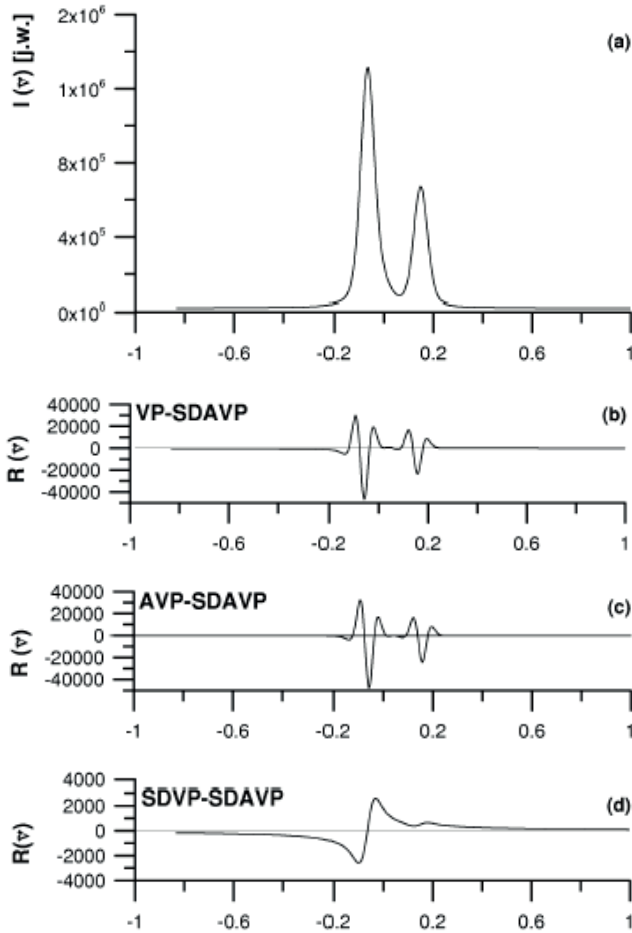
gdzie

$$v_{m_E} = \sqrt{2k_B T / m_{m_E}} .$$

W dalszej części pracy funkcje te będą nazywane funkcjami SD (ang. *speed-dependent*). Podobnie jak i w przypadku wzoru (46), po wprowadzeniu kilku uproszczeń ($B_W(\alpha; x) = B_S(\alpha; x) = B_A(\alpha; x) = 1$ i $\chi = 0$, oraz obydwu jednocześnie), ze wzoru (71) możemy otrzymać wyrażenia odpowiednio na profile AVP, SDVP i VP. Wyznaczenie wartości funkcji $B_W(\alpha; x)$, $B_W(\alpha; x)$, $B_W(\alpha; x)$ wymaga założenia określonej postaci energii oddziaływania pomiędzy emiterym i zaburzaczem.

Do pokazania różnic między wymienionymi profilami teoretycznymi przeprowadzono prostą symulację numeryczną. Wygenerowano kształty linii ^{113}Cd , używając wszystkich rozważanych profili VP, AVP, SDVP i SDAVP oraz zakładając funkcje $B_W(\alpha; x)$ i $B_W(\alpha; x)$ odpowiadające układowi Cd–Xe z założeniem potencjału van der Waalsa. Do uproszczenia obrazu ustalono liczbę składowych linii równą 2, tak aby modelowały jedynie obydwie składowe nadsubtelne linii 326,1 nm izotopu ^{113}Cd , co odpowiada przyjęciu $m = 2$ we wzorze (70) oraz uwzględnieniu odpowiednich natężeń z tabeli 1. Parametry γ_D , γ_L oraz χ wybrano tak, aby odpowiadały wartościom otrzymanym z pomiarów dla ciśnienia gazu zaburzającego równego 49,3 Torr (co odpowiada koncentracji $15,65 \times 10^{17}$ atom·cm⁻³). Następnie wykreślono różnice pomiędzy obliczonymi profilami VP, AVP i SDVP a obliczonym profilem SDAVP. Wyniki obliczeń przedstawiono na rysunku 22. Wykres 22b pokazuje wyraźne różnice w centrum pomiędzy profilami SDAVP i VP, z których pierwszy uwzględnia efekty spowodowane zależnościami parametrów kształtu linii od prędkości emitera oraz asymetrię, natomiast drugi zupełnie pomija te efekty.

Jeśli popatrzymy na wykres 22c, widzimy, że charakter różnic jest podobny do różnic między profilami VP i SDAVP. Ponieważ te różnice wykreślono dla profili AVP i SDAVP, możemy je zinterpretować jako wpływ efektów korelacji między dopplerowskim a ciśnieniowym rozszerzeniem linii na obserwowany profil linii. Widać, że objawiają się one głównie w centrum linii oraz, porównując ich wielkość z wielkością różnic na wykresie 22b, możemy stwierdzić, że ewentualny wpływ efektów korelacyjnych na kształt linii jest dużo większy od wpływu asymetrii dyspersyjnej. Potwierdzeniem tego są różnice przedstawione na wykresie 22d, gdzie wyraźnie widać asymetrię profilu SDAVP na skrzydłach linii, jednakże wielkość różnic jest około dziesięciokrotnie mniejsza. Nasuwającym się wnioskiem jest stwierdzenie, że aby prawidłowo opisać kształt linii, gdy występuje asymetria dyspersyjna, konieczne będzie użycie profilu dobrze oddającego korelacje pomiędzy dopplerowskim a ciśnieniowym rozszerzeniem linii, a to jest związane ze znajomością energii oddziaływania.



Rys. 22. Symulacja numeryczna profilu linii 326,1 nm ^{113}Cd zaburzonego ksenonem pod ciśnieniem 49,3 Tr: a) wygenerowany profil SDAVP, b) różnice $R(\tilde{\nu})$ między profilami

VP i SDAVP, c) różnice $R(\tilde{\nu})$ pomiędzy profilami AVP i SDAVP,

d) różnice $R(\tilde{\nu})$ pomiędzy profilami SDVP i SDAVP

Fig. 22. Numerical simulation of profile of 326.1 nm ^{113}Cd line perturbed by xenon under pressure 49.3 Tr: a) generated SDAVP profile, b) differences $R(\tilde{\nu})$ between VP and SDAVP profiles, c) differences $R(\tilde{\nu})$ between AVP and SDAVP profiles,

d) differences $R(\tilde{\nu})$ between SDVP and SDAVP profiles

4.3. OCENA JAKOŚCI DOPASOWANIA

Po wykonaniu dopasowania profilu teoretycznego $I_t(\tilde{\nu})$ do zmierzonego kształtu linii $I_{\text{exp}}(\tilde{\nu})$ zachodzi potrzeba oceny jakości dopasowania. Pod uwagę należy wziąć

zarówno poprawność założonego modelu i wynikającego z niego profilu teoretycznego, jak i takie zagadnienia, jak oszacowanie niepewności standardowej związanej z pomiarem czy kwestię niezależności od siebie poszczególnych wyników. Pominięcie jednego lub kilku z tych czynników może prowadzić do błędnej oceny i interpretacji otrzymanych dopasowań.

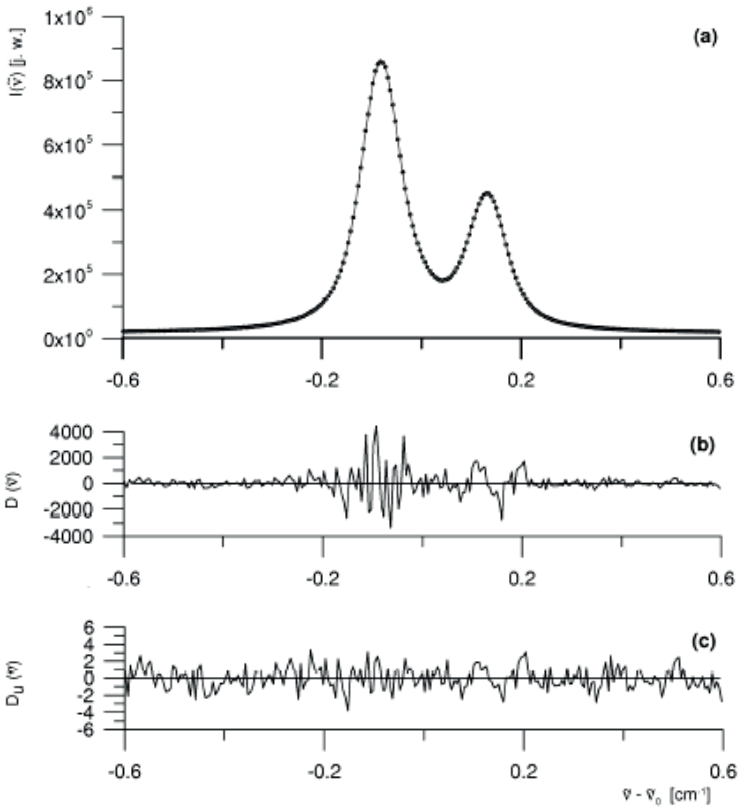
W prowadzonej analizie kształtu linii dopasowanie profili teoretycznych do danych doświadczalnych wykonano z użyciem metody najmniejszych kwadratów, korzystając z algorytmu Marquardta. W takim przypadku najbardziej popularną metodą oceny jakości dopasowania jest test χ^2 . Ponieważ symbol χ zwyczajowo jest używany do oznaczenia parametru asymetrii kształtu linii, w dalszej części pracy test ten będzie nazywany testem Chi^2 oraz wyznacznikiem jakości dopasowania będzie wartość Chi^2 . Test Chi^2 mówi nam, czy na danym poziomie istotności możemy przyjąć hipotezę, że dane doświadczalne są dobrze opisywane założonym profilem. Jest tak, jeśli spełniony jest warunek [111]

$$Chi_{exp}^2 < Chi_{\alpha}^2(n-r), \quad (75)$$

gdzie Chi_{exp}^2 jest wartością zmiennej Chi^2 otrzymaną z dopasowania, natomiast $Chi_{\alpha}^2(n-r)$ jest krytyczną wartością tej zmiennej dla poziomu istotności α i liczby stopni swobody $n-r$, gdzie n jest liczbą punktów pomiarowych, r liczbą niezależnie dopasowywanych parametrów, natomiast liczba Chi_{exp}^2 dana jest wzorem

$$Chi_{exp}^2 = \sum_i \frac{\varepsilon_i^2}{\sigma_i^2}. \quad (76)$$

Symbol ε_i został tutaj użyty do oznaczenia różnicy między profilem eksperymentalnym a teoretycznym dla każdego z punktów pomiarowych, a σ_i^2 oznacza wariancję rozkładu błędów pomiarowych w i -tym punkcie pomiarowym. Warto zauważyć, że test ten daje nam wiarygodne wyniki, jeśli wielkość Chi_{exp}^2 (76) jest rzeczywiście opisana rozkładem Chi^2 , a co za tym idzie gęstość prawdopodobieństwa rozkładu błędów pomiarowych ε_i jest dana rozkładem normalnym. Ponadto we wzorze (76) występuje wielkość σ_i , która w przypadku dopasowywania profilu teoretycznego do danych doświadczalnych jest estymowana przez niepewność standardową u_i , z jaką dokonano pomiaru w danym punkcie pomiarowym. W przypadku niepoprawnego oszacowania błędów eksperymentalnych uzyskana wartość Chi_{exp}^2 będzie także nieprawidłowa. Tak więc warunkami koniecznymi do zastosowania testu Chi^2 do oceny jakości dopasowania profilu teoretycznego do danych doświadczalnych są: występowanie tylko błędów przypadkowych oraz odpowiednie oszacowanie niepewności standardowej u_i .



Rys. 23. Profil linii 326,1 nm ^{113}Cd zaburzonej przez argon pod ciśnieniem 197 Tr:
 a – profil doświadczalny (\bullet) i dopasowany do niego profil teoretyczny SDAVP (—),
 b – różnice $D(\tilde{\nu})$ pomiędzy profilem doświadczalnym a dopasowanym profilem teoretycznym,
 (c) różnice ważone $D_u(\tilde{\nu})$ pomiędzy profilem doświadczalnym a dopasowanym profilem teoretycznym

Fig. 23. Profile of the line 326.1 nm ^{113}Cd perturbed by argon under pressure 197 Tr:
 a – experimental profile (\bullet) and fitted theoretical SDAVP profile (—),
 b – residuals $D(\tilde{\nu})$ between experimental profile and fitted theoretical profile,
 c – residuals $D_u(\tilde{\nu})$ between experimental profile and fitted theoretical profile

Należy jednak zauważyć, że ocena jakości dopasowania nie może się opierać jedynie na tym kryterium. Na przykład w przypadku wykonania pomiarów, które nie są niezależne od siebie i próby dopasowania do nich krzywej teoretycznej możemy otrzymać pozytywny wynik testu χ^2 , pomimo ewidentnego niedopasowania. Konsekwencją tego może być wyciągnięcie błędnych wniosków co do poprawności zastosowanego modelu.

W analizie kształtów linii widmowych pomocna jest wizualna ocena dopasowania poszczególnych profili teoretycznych do danych doświadczalnych. Najprostszą metodą może być wykorzystanie różnic między nimi

$$D(\tilde{\nu}) = I_{\text{exp}}(\tilde{\nu}) - I_t(\tilde{\nu}). \quad (77)$$

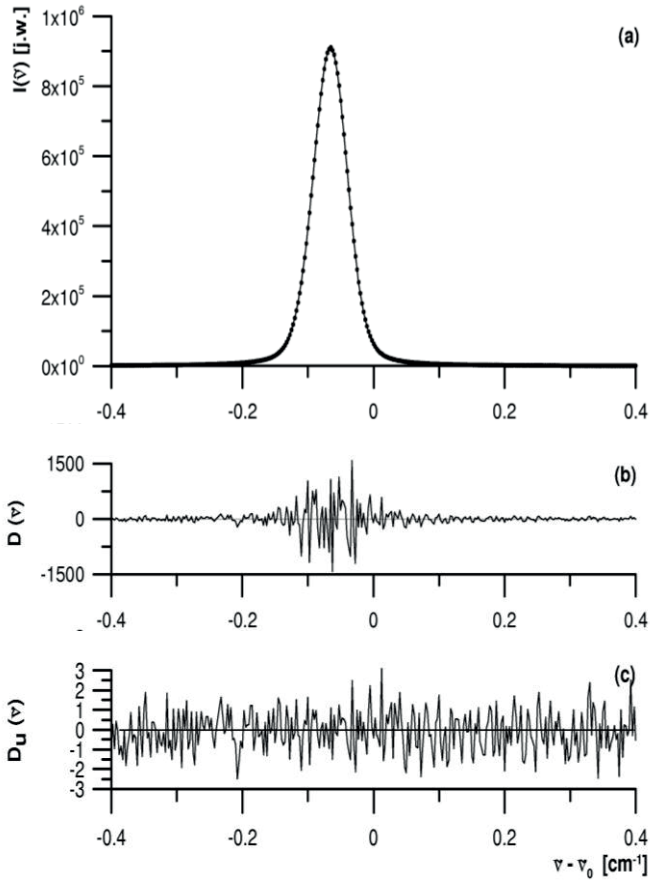
Na rysunku 23 przedstawiono przykładowy pomiar kształtu linii 326,1 nm ^{113}Cd zaburzonej argonem dla ciśnienia zaburzacza 197 Tr. Na rysunku 23b wykreślono różnice pomiędzy dopasowanym profilem a danymi doświadczalnymi. Użycie powszechnie stosowanej w badaniach kształtów linii różnicy $D(\tilde{\nu})$ uwypukla przede wszystkim niedopasowania w okolicach centrum linii, natomiast nie eksponuje ona odstępstw dopasowywanego profilu od danych doświadczalnych na skrzydłach linii. Jest to szczególnie niekorzystne w badaniach asymetrii linii, ponieważ objawia się ona głównie na skrzydłach badanej linii widmowej, jak pokazano np. na rysunku 22d. Z tego względu wskazane jest stosowanie tzw. różnic ważonych $D_u(\tilde{\nu})$, zdefiniowanych jako iloraz różnicy $D(\tilde{\nu})$ oraz niepewności standardowej danego sygnału w danym punkcie – $u(\tilde{\nu})$ [108]

$$D_u(\tilde{\nu}) = \frac{I_{\text{exp}}(\tilde{\nu}) - I_t(\tilde{\nu})}{u(\tilde{\nu})}. \quad (78)$$

Na rysunku 23c przedstawiono różnice ważone. Oddają one niedopasowania na dalekich skrzydłach linii widmowej w takim samym stopniu jak w centrum linii, co umożliwia dokładną analizę dopasowania w tych obszarach linii.

Aby pokazać słuszność takiego podejścia, wykonano prosty eksperyment numeryczny. Wygenerowano profil SDAVP linii 326,1 nm ^{114}Cd zaburzonej przez argon pod ciśnieniem 42 Tr. W tym przypadku dla uproszczenia sytuacji wybrano izotop ^{114}Cd – jest to pojedyncza linia bez struktury nadsubtelnej. Następnie do obliczonych wartości dodano szum gaussowski o wartości średniej równej zero i wariancji u_2 odpowiadającej niepewności standardowej pomiaru fluorescencji w naszym eksperymencie. Do tak otrzymanych danych pseudo-doświadczalnych dopasowano za pomocą algorytmu Marquardta ten sam profil teoretyczny. Wyniki przedstawiono na rysunku 24.

Podobnie jak w przypadku danych z pomiaru widać, że użycie różnic ważonych $D_u(\tilde{\nu})$ umożliwia lepszą ocenę jakości dopasowania poza centrum linii niż użycie zwykłych różnic $D(\tilde{\nu})$. Dodatkowo warto zauważyć, że przedstawione na rysunku 24c różnice ważone mieszczą się w granicach $\pm 3u$, co jest spowodowane tym, że do generacji tego profilu użyliśmy szumu gaussowskiego o wariancji równej u^2 . W przypadku dopasowania profilu teoretycznego do danych doświadczalnych, jeśli otrzymane wartości różnic ważonych są dużo większe lub dużo mniejsze niż $3u$, oznacza to niedoszacowanie lub przeszacowanie niepewności przypadkowych lub występowanie innych źródeł niepewności.



Rys. 24. Symulacja profilu linii 326,1 nm ^{114}Cd zaburzonej przez argon pod ciśnieniem 42 Tr: a – profil pseudodoświadczalny wygenerowany przy użyciu profilu SDAVP z dodanym szumem gaussowskim (\bullet) i dopasowany do niego profil teoretyczny SDAVP(—), b – różnice $D(\tilde{\nu})$ między profilem pseudodoświadczalnym a dopasowanym profilem teoretycznym, c – różnice ważone $D_u(\tilde{\nu})$ pomiędzy profilem pseudo-doświadczalnym a dopasowanym profilem teoretycznym

Fig. 24. Simulation of profile of the 326.1 nm ^{114}Cd line perturbed by argon under pressure of 42 Tr: a – pseudo-experimental profile generated with SDAVP profile with Gaussian noise added (\bullet) and fitted theoretical SDAVP profile SDAVP(—), b – residuals $D(\tilde{\nu})$ between pseudo-experimental and theoretical profile, c – weighted residuals $D_u(\tilde{\nu})$ between pseudo-experimental and theoretical profile

Wartości doświadczalne natężenia badanej linii w naszym eksperymencie wyznaczone są według wzoru (69). Tak więc niepewność standardową zarejestrowanego sygnału, niezbędną do obliczenia różnic ważonych $D_u(\tilde{\nu})$, w każdym z punktów możemy oszacować, korzystając ze wzoru

$$u(\tilde{\nu}) = \sqrt{\left(\frac{M(\tilde{\nu}_m)}{M(\tilde{\nu}) - M_t(\tilde{\nu})}\right)^2 [u_F^2(\tilde{\nu}) + u_{F_t}^2(\tilde{\nu})] + \left(\frac{[F(\tilde{\nu}) - F_t(\tilde{\nu})]M(\tilde{\nu}_m)}{(M(\tilde{\nu}) - M_t(\tilde{\nu}))^2}\right)^2 [u_{M_t}^2(\tilde{\nu}) + u_{M_t}^2(\tilde{\nu})]}, \quad (79)$$

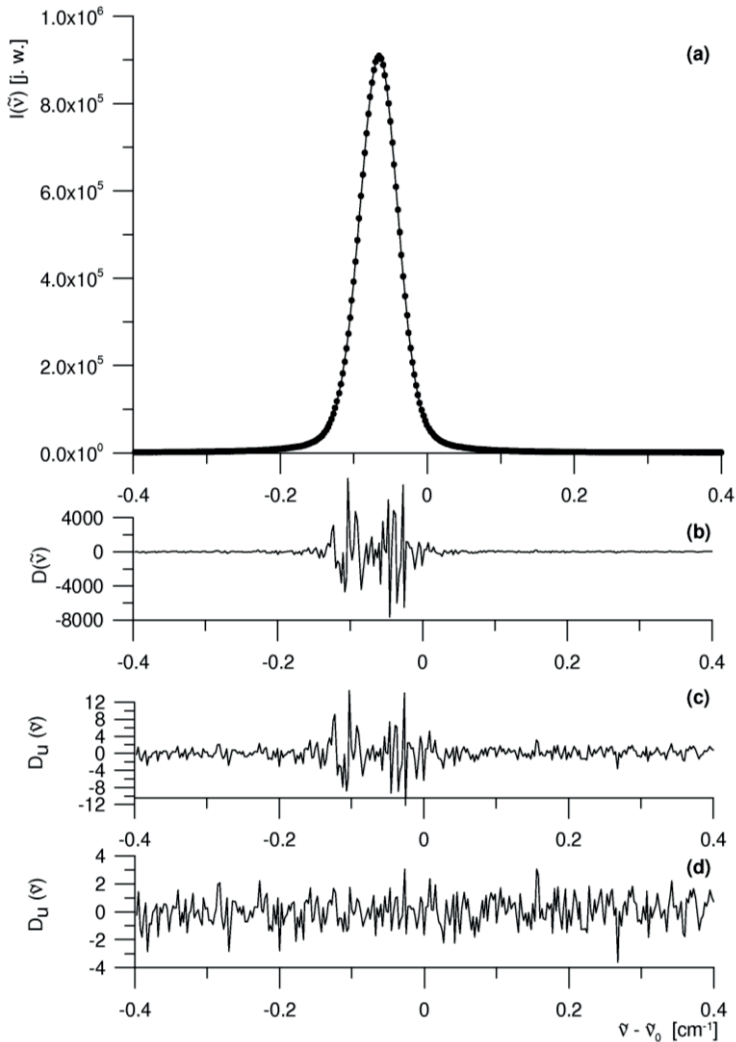
gdzie $u_{M_{UV}}(\tilde{\nu})$ jest niepewnością standardową $M_{UV}(\tilde{\nu})$. Sygnał $F(\tilde{\nu})$ i tło $F_t(\tilde{\nu})$ są rejestrowane za pomocą metody zliczeń fotonów i podlegają rozkładowi Poissona, więc ich niepewności standardowe $u_F(\tilde{\nu})$ i $u_{F_t}(\tilde{\nu})$ równe są pierwiastkom kwadratowym tych wielkości: $u_F(\tilde{\nu}) = \sqrt{F(\tilde{\nu})}$ oraz $u_{F_t}(\tilde{\nu}) = \sqrt{F_t(\tilde{\nu})}$. Niepewności $M(\tilde{\nu}_m)$ nie bierzemy pod uwagę, gdyż wszystkie punkty doświadczalne są mnożone przez tę samą wartość $M(\tilde{\nu}_m)$. We wzorze (79) można także pominąć wartości $M_t(\tilde{\nu})$ i $u_{M_t}^2(\tilde{\nu})$, ponieważ mają one pomijalnie mały wpływ na ostateczną wartość niepewności.

Warto jednak zwrócić uwagę, że poprzedni wzór byłby słuszny tylko w przypadku, gdyby wartości liczb falowych (lub częstotliwości) były wyznaczone dokładnie lub niepewność, z jaką byłyby wyznaczone, byłaby pomijalna. W naszym przypadku jednak wartości liczb falowych, jak to widać na rysunku 15, zostały wyznaczone z określoną niepewnością pomiarową.

Aby sprawdzić, jak wpływa to na wykres różnic ważonych, przeprowadzono symulacje w sposób podobny do tej przedstawionej na rysunku 24 z tą różnicą, że w generacji naszego pseudodoświadczalnego profilu oprócz szumu gaussowskiego dodanego do obliczonych wartości „sygnału”, do wartości liczb falowych dodano także szum gaussowski o wariancji odpowiadającej niepewności standardowej pomiaru częstotliwości lasera w naszym układzie pomiarowym podniesionej do kwadratu. Na rysunku 25 przedstawiono wyniki symulacji dla linii 326,1 nm ^{114}Cd zaburzonej argonem pod ciśnieniem 42 Tr. Na pierwszym z wykresów różnic przedstawiono zwykłe różnice $D(\tilde{\nu})$, na drugim różnice ważne $D_u(\tilde{\nu})$, biorąc pod uwagę jedynie błąd wyznaczenia wartości sygnału w danym punkcie pomiarowym (zgodnie ze wzorem (79)). Zaobserwować można odstępstwa od założonego profilu pojawiające się na bliskich skrzydłach linii. Spowodowane jest to właśnie błędami w wyznaczeniu wartości liczb falowych, które mają największy wpływ na rejestrowany sygnał w tych rejonach linii, w których nachylenie profilu linii jest największe. Aby uwzględnić w obliczeniach niepewności ten efekt, można do niepewności pomiaru sygnału w każdym punkcie dodać niepewność wynikającą z wyznaczenia liczby falowej w tym punkcie.

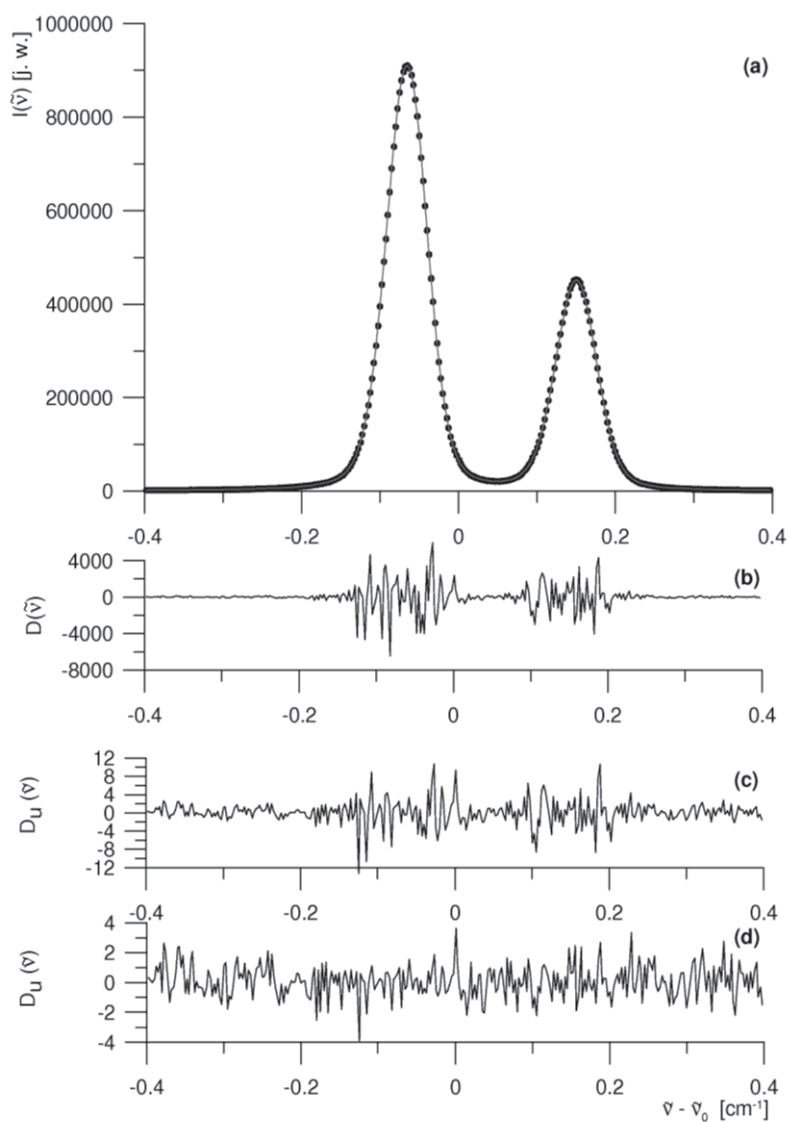
Przyczynę wnoszoną przez niepewność wyznaczenia liczb falowych do całkowitej niepewności pomiaru można przybliżyć za pomocą wzoru

$$u_{d\tilde{\nu}}^2(\tilde{\nu}) = \left(\frac{\partial I_{\text{exp}}(\tilde{\nu})}{\partial \tilde{\nu}}\right)^2 u_{\tilde{\nu}}^2(\tilde{\nu}), \quad (80)$$



Rys. 25. Symulacja profilu linii 326,1 nm ^{114}Cd zaburzonej przez argon pod ciśnieniem 42 Tr, a – profil pseudo-doswiadczalny wygenerowany z użyciem profilu SDAVP z dodanym szumem gaussowskim zarówno do natężenia sygnału, jak i do wartości liczb falowych (•) i dopasowany do niego profil teoretyczny (—), b – różnice $D(\tilde{\nu})$, c – różnice ważone $D_u(\tilde{\nu})$ obliczone z wykorzystaniem wag danych wzorem (79), d – różnice ważone $D_u(\tilde{\nu})$ obliczone z wykorzystaniem wag danych wzorem (81)

Fig. 25. Simulation of profile of the 326.1 nm ^{114}Cd line perturbed by argon under pressure of 42 Torr, a – pseudo-experimental profile generated with SDAVP profile with gaussian noise added both to intensity of the signal and to values of wavenumbers (•) and fitted theoretical SDAVP profile (—), b – residuals $D(\tilde{\nu})$, c – weighted residuals $D_u(\tilde{\nu})$ calculated using weights given by formula (79), d – weighted residuals $D_u(\tilde{\nu})$ calculated using weights given by formula (81)



Rys. 26. Symulacja profilu linii 326,1 nm ^{113}Cd zaburzonej przez argon pod ciśnieniem 42 Tr, oznaczenia jak na rysunku 25

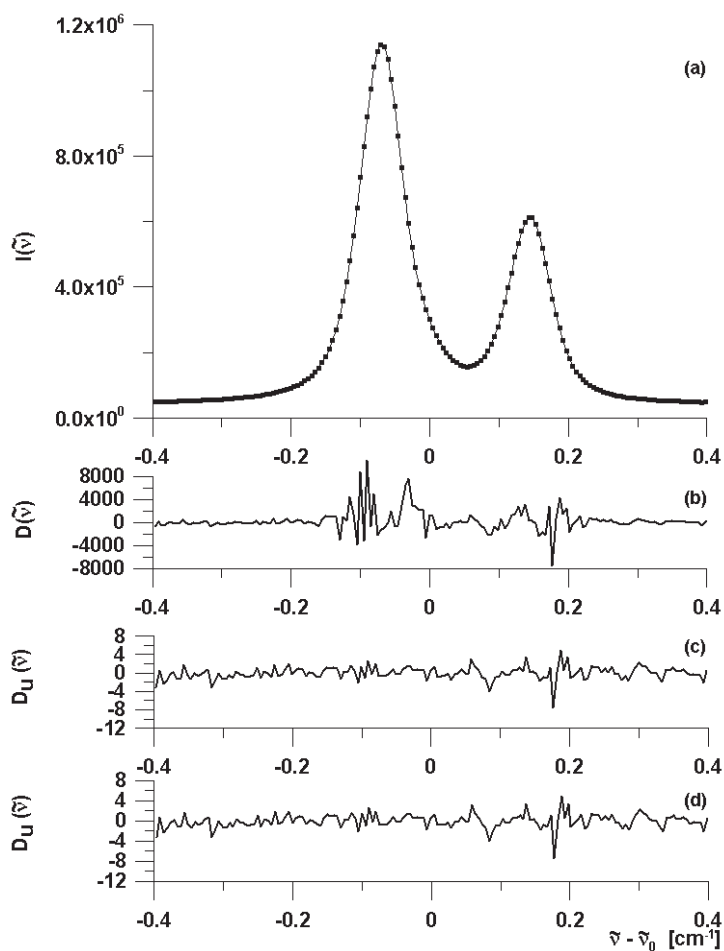
Fig. 26. Simulation of profile of the 326.1 nm ^{113}Cd line perturbed by argon under pressure of 42 Torr, symbols like in figure 25

a niepewność zarejestrowanego natężenia światła $I_{\text{exp}}(\tilde{\nu})$ będzie dana wzorem

$$u_c^2(\tilde{\nu}) = \sqrt{u_p^2(\tilde{\nu}) + \left(\frac{\partial I_{\text{exp}}(\tilde{\nu})}{\partial \tilde{\nu}} \right)^2 u_{\tilde{\nu}}^2(\tilde{\nu})}, \quad (81)$$

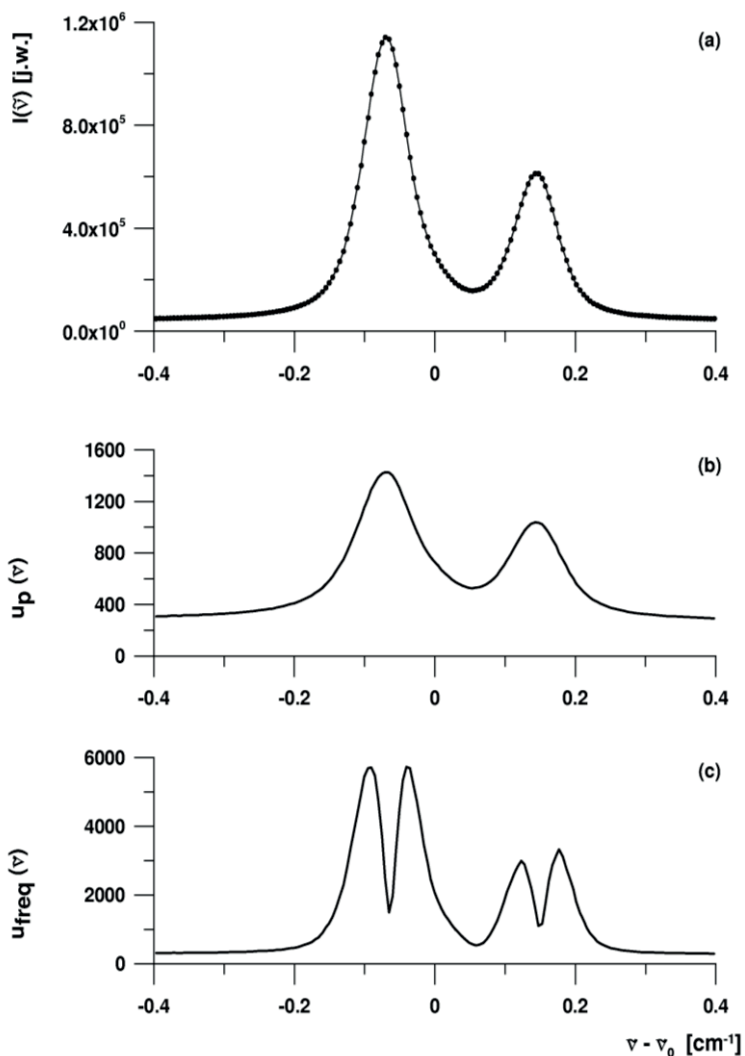
gdzie $u_p(\tilde{\nu})$ oznacza niepewność standardową związaną z pomiarem fluorescencji (daną wzorem (79)), a $u_{\tilde{\nu}}(\tilde{\nu})$ jest niepewnością standardową pomiaru liczby falowej w danym punkcie pomiarowym. Różnice ważone wyliczone w ten sposób wykorzystano podczas obliczania różnic pomiędzy profilem pseudodoświadczalnym i dopasowanym profilem teoretycznym przedstawionym na rysunku 25d. Widoczny jest tutaj brak oscylacji występujących w obszarze bliskich skrzydeł linii pojawiających się w przypadku nieuwzględnienia niepewności wyznaczenia liczby falowej $\tilde{\nu}$. Ponadto różnice między dopasowaniem a symulowanymi wynikami doświadczalnymi mieszczą się w zakresie od $-3u$ do $3u$, co jest zgodne z założeniem przyjętym w generowaniu tych danych.

Podobny eksperyment numeryczny możemy wykonać dla linii 326,1 nm izotopu ^{113}Cd o takim samym ciśnieniu (42 Tr) argonu jako gazu zaburzającego. Wyniki symulacji przedstawiono na rysunku 26. Także i tutaj są widoczne oscylacje różnic ważonych w obszarach o dużym pochyleniu zbocza linii w przypadku nieuwzględnienia niepewności pomiaru częstości. Po ich uwzględnieniu oscylacje przestają być widoczne. Na rysunku 27 przedstawiono przykładowy wynik pomiaru linii 326,1 nm ^{113}Cd zaburzonego ksenonem pod ciśnieniem 106 Tr. Widzimy, że przy nieuwzględnieniu błędu wyznaczenia częstotliwości odstępstwa od dopasowanego profilu mają podobny charakter do odstępstw otrzymanych z opisanej symulacji. Na rysunku 27d przedstawiono różnice ważone $D_u(\tilde{\nu})$ z wagami określonymi wzorem (81). Warto zauważyć, że wartości różnic ważonych ogólnie zawierają się w przedziale $(-3; 3)$. Oznacza to, że rozkład odstępstw profilu eksperymentalnego $I_{\text{exp}}(\tilde{\nu})$ od dopasowanego profilu teoretycznego $I(\tilde{\nu})$ jest dobrze opisywany rozkładem normalnym o wartości średniej równej zero i wariancji równej u^2 . Świadczy to o prawidłowym oszacowaniu całkowitej niepewności standardowej wyznaczenia natężenia światła dla każdego z punktów pomiarowych, a więc także o prawidłowym oszacowaniu przyczynków do całkowitej niepewności pomiarowej pochodzących od wszystkich czynników mających wpływ na wynik pomiaru. Zgodnie ze stwierdzeniem z początku rozdziału, w przypadku niedoszacowania lub przeszacowania niepewności standardowej pomiaru wyrażenie ε^2/σ^2 miałoby nieprawidłową wartość, podobnie jak wartość $\text{Chi}_{\text{exp}}^2$ dana wzorem (76) i co za tym idzie, wyniki testu Chi^2 nie byłyby wiarygodne.



Rys. 27. Profil linii 326,1 nm ^{113}Cd zaburzonej przez ksenon pod ciśnieniem 106 Tr:
 a – profil doświadczalny (\bullet) i dopasowany do niego profil teoretyczny (—), b – różnice $D(\tilde{\nu})$,
 c – różnice ważone $D_u(\tilde{\nu})$ z wykorzystaniem wag danych wzorem (79),
 d – różnice ważone $D_u(\tilde{\nu})$ z wykorzystaniem wag danych wzorem (81)

Fig. 27. Profile of the 326.1 nm ^{113}Cd line perturbed by xenon under pressure of 106 Tr:
 a – experimental profile (\bullet) and fitted theoretical profile (—), b – residuals $D(\tilde{\nu})$,
 c – weighted residuals $D_u(\tilde{\nu})$ calculated using weights given by formula (79),
 d – weighted residuals $D_u(\tilde{\nu})$ calculated using weights given by formula (81)



Rys. 28. Profil linii 326,1 nm ^{113}Cd zaburzonej przez ksenon pod ciśnieniem 42 Tr:
 a – profil doświadczalny (•) i dopasowany do niego profil teoretyczny (—),
 b – wkład do niepewności całkowitej pomiaru pochodzący z niepewności pomiaru natężenia fluorescencji, obliczony według wzoru (79), c – wkład do niepewności całkowitej pomiaru pochodzący z niepewności pomiaru częstotliwości lasera, obliczony z (80)

Fig. 28. Profile of the 326.1 nm ^{113}Cd line perturbed by xenon under pressure of 42 Tr:
 a – experimental profile (•) and fitted theoretical profile (—), b – contribution in total uncertainty of the measurement originating from uncertainty of fluorescence measurement, counted with formula (79), c – contribution in total uncertainty of the measurement originating from uncertainty of laser beam frequency measurement, calculated from formula (79)

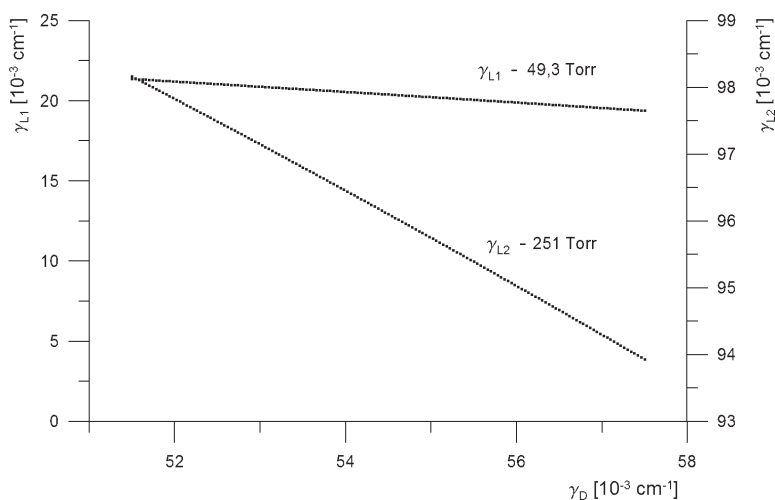
Możemy przyrzeć się wkładom poszczególnych niepewności standardowych związanych odpowiednio z pomiarem natężenia fluorescencji, jak i pomiarem liczby falowej generowanego przez układ laserowy promieniowania we wzorze (81).

Okazuje się, że dla niskich ciśnień, a co za tym idzie stromych profili rejestrowanego kształtu linii, niepewność standardowa pomiaru częstotliwości ma większy wpływ niż niepewność związana z pomiarem natężenia fluorescencji (patrz rys. 28). Wpływ ten oczywiście zmniejsza się wraz ze wzrostem ciśnienia gazu zaburzającego.

Podsumowując, należy podkreślić, że uwzględnienie niepewności standardowej pomiaru częstotliwości lasera nie poprawia samych wyników dopasowań. Umożliwia ono jednak prawidłową ocenę ich poprawności zarówno za pomocą testu χ^2 , jak i oceny wizualnej. Dzięki temu możliwa jest ocena poprawności oszacowania niepewności pomiarowej, co umożliwia także kontrolę poprawności pracy układu pomiarowego.

4.4. DOPASOWYWANIE PROFILU DO POŁOŻONYCH BLISKO SIEBIE SKŁADOWYCH LINII WIDMOWEJ

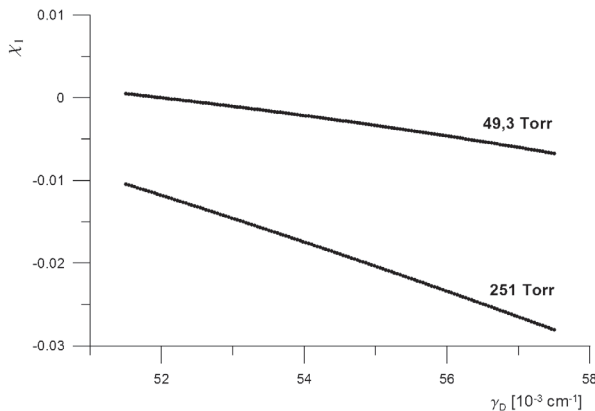
Metoda Marquardta [106, 107] znajdowania globalnego minimum w wielowymiarowej przestrzeni parametrów dopasowywanej funkcji jest metodą ogólnie znaną i była używana wielokrotnie do dopasowywania kształtów linii widmowych, także we wcześniejszych pomiarach układów Cd–gazy szlachetne oraz Cd–gazy molekularne.



Rys. 29. Wartości dopasowanej szerokości składowej lorentzowskiej γ_L kształtu linii ^{113}Cd zaburzonej Xe w zależności od ustalonej szerokości składowej dopplerowskiej γ_D dla ciśnienia ksenonu równego 49,3 Tr (γ_{L1}) oraz 251 Tr (γ_{L2})

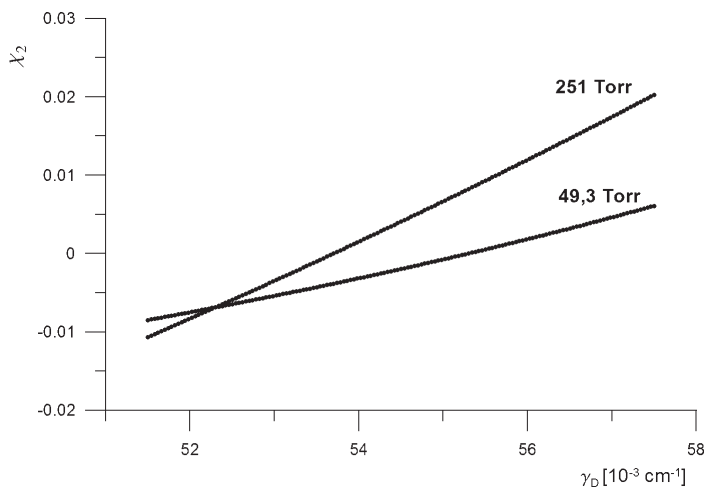
Fig. 29. Fitted values of Lorentzian width γ_L of shape of ^{113}Cd line perturbed by Xe depending on fixed Doppler width γ_D for xenon pressure equal to 49.3 Tr (γ_{L1}) and 251 Tr (γ_{L2})

W wartościach początkowych parametrów dopasowywanej funkcji bliskich wartościom rzeczywistym (jak bliskie wartości są wartościami wystarczająco bliskimi zależy m.in. od koncentracji gazu zaburzającego czy stopnia skomplikowania dopasowywanego profilu) metoda ta jest szybkozbieżna. Jednak nawet w przypadku analizy pojedynczej linii z użyciem profilu SDAVP pojawiają się niekorzystne ze względu na zbieżności i niezawodności dopasowania efekty: zachodzi korelacja pomiędzy dopasowywaną szerokością dopplerowską i asymetrią, a dodatkowo dla wysokich koncentracji gazu zaburzającego wyznaczana z dopasowania wartość szerokości składowej dopplerowskiej γ_D staje się coraz gorzej określona. Ten ostatni efekt spowodowany jest tym, że kształt linii przy dużych koncentracjach zaburzacza ma charakter coraz bardziej lorentzowski, ze zmniejszającą się rolą składowej dopplerowskiej. Przykładem mogą być badania kształtu linii 422,7 nm Ca zaburzonej przez gazy szlachetne wykonane przez grupę Lewisa [94, 112]. Podczas analizy bardziej skomplikowanych układów, opisywanych funkcjami o dużej liczbie parametrów, sytuacja się komplikuje jeszcze bardziej, czego przykładem może być opisany w niniejszej pracy układ dwóch linii (dwie składowe nadsztetne linii 326,1 nm ^{113}Cd). Dla większych koncentracji zaburzacza następuje nakładanie się obydwu linii, co powoduje silniejszą wzajemną korelację pomiędzy niektórymi dopasowywanymi parametrami. Na rysunkach 29–31 przedstawiono przykładowe wyniki dopasowań profilu SDAVP do wyników pomiaru kształtów linii 326,1 nm ^{113}Cd zaburzonej ksenonem pod ciśnieniem odpowiednio 49,3 i 251 Tr. Na pierwszym z rysunków przedstawiono otrzymane wartości szerokości składowej lorentzowskiej χ_1 w zależności od ustalonej (stałej podczas dopasowywania) wartości szerokości składowej dopplerowskiej γ_D .



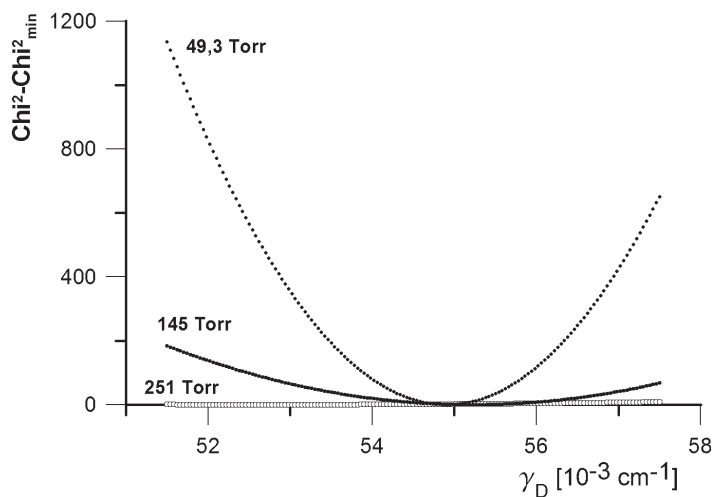
Rys. 30. Wartości dopasowanej wartości parametru asymetrii χ_1 kształtu linii ^{113}Cd zaburzonej Xe w zależności od ustalonej szerokości składowej dopplerowskiej γ_D dla ciśnienia ksenonu równego 49,3 Tr oraz 251 Tr

Fig. 30. Fitted values of asymmetry parameter χ_1 of shape of ^{113}Cd line perturbed by Xe depending on fixed Doppler width γ_D for xenon pressure equal to 49.3 Tr and 251 Tr



Rys. 31. Wartości dopasowanej wartości parametru asymetrii χ_2 kształtu linii ^{113}Cd zaburzonej Xe w zależności od ustalonej szerokości składowej dopplerowskiej dla ciśnienia ksenonu równego 49,3 Tr oraz 251 Tr

Fig. 31. Fitted values of asymmetry parameter χ_2 of shape of ^{113}Cd line perturbed by Xe depending on fixed Doppler width γ_D for xenon pressure equal to 49.3 Tr and 251 Tr



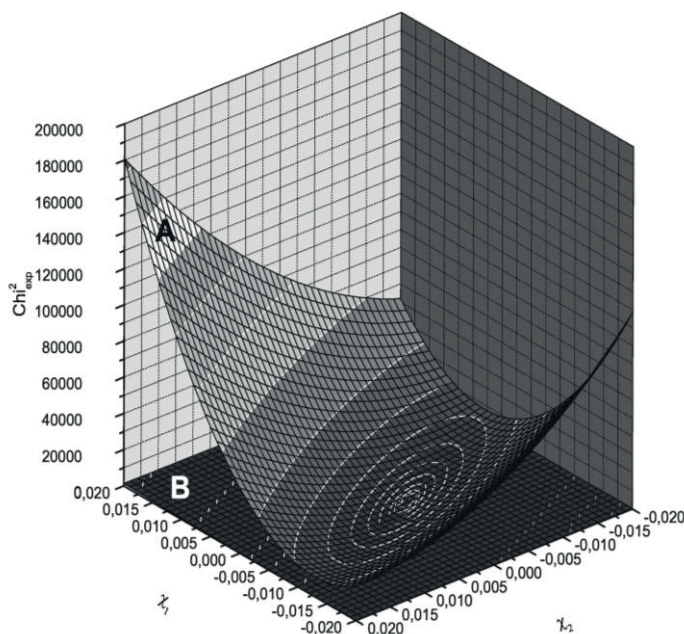
Rys. 32. Zależność różnicy pomiędzy parametrem χ^2 a wartością minimalną $\chi_{\text{exp min}}^2$ dopasowania od ustalonej wartości parametru składowej dopplerowskiej γ_D . Przedstawiono zależności dla trzech ciśnień gazu zaburzającego z zakresu 49,5–251 Tr

Fig. 32. Dependence of difference between χ^2 parameter and minimum value of fitted profile's $\chi_{\text{exp min}}^2$ on fixed value of Doppler width γ_D . Dependences for three pressures of perturbing gas from range of 49.5–251 Tr are presented

Wyniki wskazują wyraźną korelację dopasowanej szerokości składowej lorentzowskiej z szerokością składowej dopplerowskiej γ_D , jednak jest ona znacznie silniejsza dla wyższego ciśnienia (251 Tr). Wynik ten jest zgodny z wcześniej wspomnianymi wynikami Lewisa i innych dla układu Ca-Ar [112], jednak w tym przypadku zachowanie to staje się widoczne przy niższym ciśnieniu gazu zaburzającego. Na rysunkach 30 oraz 31 przedstawiono wartości dopasowanych parametrów asymetrii dyspersyjnej χ_1 i χ_2 kształtu linii 326,1 nm ^{113}Cd zaburzonej Xe w zależności od ustalonej szerokości składowej dopplerowskiej γ_D . Otrzymane wyniki wskazują na silną korelację parametrów obydwu asymetrii i szerokości dopplerowskiej γ_D , ponadto są one w podobnym stopniu skorelowane zarówno dla niskiego (49,3 Tr), jak i wysokiego ciśnienia (251 Tr). Ponieważ wartość $\text{Chi}_{\text{exp}}^2$ jest głównym kryterium poszukiwania minimum globalnego w metodzie Marquardta, zachowanie się tej wartości przy zmianie poszczególnych parametrów dopasowania determinuje szybkość zbiegania metody do globalnego minimum. Na rysunku 32 przedstawiono wykresy wzrostu wartości Chi^2 podczas ustalania kolejnych wartości szerokości dopplerowskiej z zakresu od 51,5 do $57,5 \cdot 10^{-3} \text{ cm}^{-1}$ z krokiem $0,2 \cdot 10^{-3} \text{ cm}^{-1}$ (jako wartość zerową przyjęto minimalną wartość $\text{Chi}_{\text{exp min}}^2$ otrzymaną po dopasowaniu profilu teoretycznego ze zmieniającym parametrem γ_D). W tym wypadku widoczny jest spadek wzrostu wartości $\text{Chi}_{\text{exp}}^2$ przy zmianie parametru γ_D wraz ze wzrostem ciśnienia gazu zaburzającego. W przypadku najniższego ciśnienia $\text{Chi}_{\text{exp}}^2$ wzrosło maksymalnie o około 1136, podczas gdy dla najwyższego ciśnienia maksymalny wzrost wartości $\text{Chi}_{\text{exp}}^2$ wyniósł 9,1.

W przypadku wyznaczania minimum lokalnego dla parametrów asymetrii obydwu składowych nadsubtelnych χ_1 i χ_2 , których wpływ na otrzymany kształt linii jest stosunkowo mały (asymetria zmienia kształt linii o mniej niż 1%), dwuwymiarowa przestrzeń parametrów χ_1 i χ_2 jest stosunkowo „płaska” względem $\text{Chi}_{\text{exp}}^2$ (co zostało pokazane na rysunkach 33 i 34). Podobnie jak dla poprzednich wykresów przedstawiono tutaj dwa skrajne przypadki: ciśnienie gazu zaburzającego, równe 49,3 Tr i 251 Tr. Dla każdego z tych ciśnień wykonano dwuwymiarową mapę wartości $\text{Chi}_{\text{exp}}^2$ w zależności od ustalonych parametrów asymetrii obydwu składowych nadsubtelnych. Parametry χ_1 oraz χ_2 zmieniano w zakresie od $-0,020$ do $0,020$ z krokiem $0,01$, co dało łącznie 1600 wartości $\text{Chi}_{\text{exp}}^2$. W przypadku niskiego ciśnienia przy zmianie wartości któregoś z parametrów asymetrii $\text{Chi}_{\text{exp}}^2$ stosunkowo szybko zwiększa swoją wartość, co przekłada się na dobre działanie algorytmu Marquardta. W drugim wypadku (rys. 33 i rys. 35) $\text{Chi}_{\text{exp}}^2$ zmienia swoją wartość w bardzo małym stopniu. Należy zwrócić uwagę na zmianę skali pomiędzy rysunkami 34 i 35. Maksymalna zmiana wartości $\text{Chi}_{\text{exp}}^2$ w przypadku niskiego

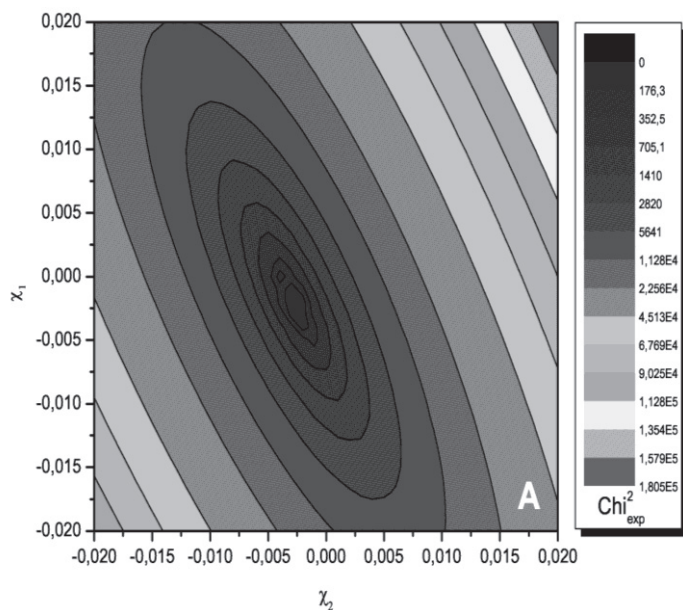
ciśnienia na całej powierzchni parametrów χ_1 i χ_2 odpowiada zmianie wartości Chi_{exp}^2 w obszarze wewnątrz czterech pierwszych poziomic zaznaczonych na płaszczyźnie parametrów χ_1 oraz χ_2 w przypadku ciśnienia 251 Tr (rys. 35).



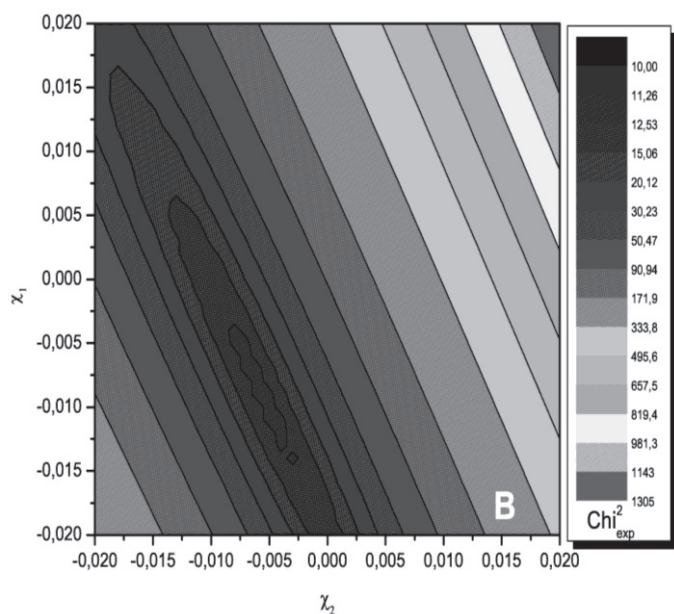
Rys. 33. Wykres zależności parametru $Chi_{\text{exp_min}}^2$ otrzymanego z dopasowania profilu SDAVP do danych doświadczalnych od ustalonych wartości parametrów asymetrii składowych nadsubtelnych izotopu ^{113}Cd χ_1 i χ_2 . Powierzchnia A odpowiada ciśnieniu gazu zaburzającego równemu 49,3 Tr, powierzchnia B odpowiada ciśnieniu 251 Tr

Fig. 33. Dependence of $Chi_{\text{exp_min}}^2$ parameter obtained from fit of SDAVP profile to experimental data on fixed values of asymmetry parameters χ_1 and χ_2 of hyperfine components ^{113}Cd isotope. Area A corresponds with perturbing gas pressure 49.3 Tr, area B corresponds with pressure 251 Tr

Dodatkowo kształt konturów ograniczających obszary o określonych wartościach Chi_{exp}^2 w tym drugim przypadku wskazuje na korelację parametrów asymetrii (rys. 35). Tak więc, podczas wyznaczania parametru asymetrii dla wyższych ciśnien gaz zaburzającego należy zwrócić uwagę na zachowanie dopasowywanych parametrów kształtu linii. Ponieważ wraz ze wzrostem ciśnienia kształt linii ma charakter coraz bardziej lorentzowski, występuje korelacja między szerokościami składowych dopplerowskich i lorentzowskich, co objawia się zmianą (np. zmniejszeniem) dopasowanej wartości szerokości składowej dopplerowskiej i odpowiadającej jej zmianą (np. zwiększeniem) dopasowanej wartości szerokości składowej lorentzowskiej.



Rys. 34. Rzut powierzchni A z rysunku 33 na płaszczyznę χ_1, χ_2 dla ciśnienia 49,3 Tr
 Fig. 34. Projection of surface A (fig. 33) on the plane χ_1, χ_2 for pressure of 49.3 Tr



Rys. 35. Rzut powierzchni B z rysunku 33 na płaszczyznę χ_1, χ_2 dla ciśnienia 251 Tr
 Fig. 35. Projection of surface B (fig. 33) on the plane χ_1, χ_2 for pressure of 251 Tr

Z powodu korelacji pomiędzy szerokością składowej dopplerowskiej a parametrami asymetrii obydwu składowych nadsubtelnych może to prowadzić do zmiany parametrów asymetrii. Z wymienionych powodów w przypadku naszej linii 326,1 nm ^{113}Cd . zaburzonej przez gazy szlachetne, wyznaczone wartości parametrów asymetrii dyspersyjnej będą obarczone dodatkową niepewnością. Precyzja tych wyników będzie ulegała zmniejszeniu wraz ze wzrostem ciśnienia gazu zaburzającego. Spadek precyzji wynika z nałożenia na siebie obydwu składowych nadsubtelnych linii 326,1 nm ^{113}Cd .

5. PODSUMOWANIE

Prowadzone w Zakładzie Fizyki Atomowej, Molekularnej i Optycznej Instytutu Fizyki UMK systematyczne badania kształtu interkombinacyjnej linii widmowej kadmu 326,1 nm wykazały, że w przypadku izotopu parzysto-nieparzystego, gdy linia ta ma dwie składowe nadsubtelne, konieczna jest bardzo precyzyjna znajomość częstotliwości promieniowania laserowego wzbudzającego fluorescencję kadmu. Dotychczas stosowany układ kalibracyjny wykorzystujący równoczesny z pomiarem sygnału fluorescencji pomiar widma absorpcyjnego par jodu oraz znaczniki (piki transmisji) z pasywnie stabilizowanego interferometru Fabry–Perota (IFP) okazał się niewystarczający i został zastąpiony układem składającym się z przestrajanego piezoelektrycznie interferometru Fabry–Perota oraz stabilizowanego częstotliwościowo lasera helowo-neonowego. Zbudowany układ kalibracyjny umożliwia pomiar częstotliwości względnej promieniowania pierścieniowego lasera barwnikowego z precyzją lepszą niż dostępne urządzenia komercyjne, co przekłada się na istotne zwiększenie precyzji uzyskiwanych wartości parametrów kształtu linii.

W procesie analizy kształtu zmierzonej linii widmowej, po dopasowaniu do niej odpowiedniego profilu teoretycznego stosowana jest procedura oceny jakości dopasowania, mająca na celu stwierdzenie, czy zastosowany profil teoretyczny poprawnie opisuje badany kształt linii widmowej.

Jak pokazano, stosowane powszechnie milczące założenie, że wartości zmiennej niezależnej – wartości częstotliwości promieniowania laserowego są znane z niepewnością, która może być pominięta, prowadzi do istotnych przekłamań w procedurze oceny jakości dopasowania profili. Przeprowadzone symulacje numeryczne i zmierzone doświadczalnie kształty linii wykazały, że w celu przeprowadzenia poprawnej analizy kształtu linii konieczne jest uwzględnienie niepewności wyznaczenia częstotliwości promieniowania laserowego. Jest to szczególnie istotne w przypadku linii widmowych mających strukturę nadsubtelną, tak jak w przypadku linii 326,1 nm parzysto-nieparzystego izotopu ^{113}Cd .

LITERATURA

- [1] Materiały reklamowe firmy Picarro dostępne na stronie internetowej www.picarro.com
- [2] Materiały reklamowe firmy Tiger Optics dostępne na stronie internetowej www.tigeroptics.com
- [3] THORPE M.J., BALSLEV-CLAUSEN D., KIRCHNER M. S., YE J., *Human breath analysis via cavity-enhanced optical frequency comb spectroscopy*, Optics Express, 2008, Vol. 16, No. 4 2387.
- [4] LISAK D., HODGES J., CIURYŁO R., *Comparison of semiclassical line-shape models to rovibrational H₂O spectra measured by frequency-stabilized cavity ring-down spectroscopy*, Phys. Rev. A, 2006, Vol. 73, 012507.
- [5] DUGGAN P., SINCLAIR P.M., BERMAN R., MAY A.D., DRUMMOND J.R., *Testing Lineshape Models: Measurements for $\nu = 1 - 0$ CO Broadened by He and Ar*, J. Mol. Spectrosc., 1997, Vol. 186, 90; CHAUSSARD F., SAINT-LOUP R., BERGER H., JOUBERT P., BRUET X., BONAMY J., ROBERT D., *Speed-dependent line profile: A test of a unified model from the Doppler to the collisional regime for molecule–molecule collisions*, J. Chem. Phys., 2000, Vol. 113, No. 12, 4951; BRAULT J.W., BROWN L.R., CHACKERIAN C. Jr., FREEDMAN R., PREDOI-CROSS A., PINE A.S., *Self-broadened ¹²C¹⁶O line shapes in the $\nu = 2 \leftarrow 0$ band*, J. Mol. Spectrosc., 2003, Vol. 222, 220.
- [6] CIURYŁO R., SZUDY J., *Speed-dependent pressure broadening and shift in the soft collision approximation*, J. Quant. Spectrosc. Radiat. Transfer, 1997, Vol. 57, No. 3, 411; CIURYŁO R., PINE A.S., SZUDY J., *A generalized speed-dependent line profile combining soft and hard partially correlated Dicke-narrowing collisions*, J. Quant. Spectrosc. Radiat. Transfer, 2001, No. 68, 257; CIURYŁO R., SHAPIRO D.A., DRUMMOND J.R., MAY A.D., *Solving the line-shape problem with speed-dependent broadening and shifting and with Dicke narrowing. II. Application*, Phys. Rev. A, 2002, Vol. 65, 012502.
- [7] STRONG K., TAYLOR F.W., CALCUTT S.B., REMEDIOS J.J., BALLARD J.J., *Spectral parameters of self- and hydrogen-broadened methane from 2000 to 9500 cm⁻¹ for remote sounding of the atmosphere of jupiter*, J. Quant. Spectrosc. Radiat. Transfer, 1993, Vol. 50, 363.
- [8] MILLER C.E., BROWN L.R., TOTH R.A., BENNER D.C., DEVI V.M., *Spectroscopic challenges for high accuracy retrievals of atmospheric CO₂ and the Orbiting Carbon Observatory (OCO) experiment*, C. R. Physique, 2005, Vol. 6, 876.
- [9] HODGES J.T., LISAK D., *Frequency-stabilized cavity ring-down spectrometer for high-sensitivity measurements of water vapor concentration*, Appl. Phys. B, 2006, Vol. 85, 375.
- [10] van DIEDENHOVEN B., HASEKAMP O.P., ABEN I., *Surface pressure retrieval from SCIAMACHY measurements in the O₂ A Band: validation of the measurements and sensitivity on aerosols*, Atmos. Chem. Phys., 2005, Vol. 5, 2109.
- [11] CRISP D., ATLAS R.M., BREON F.-M., BROWN L.R., BURROWS J.P., CIAIS P., CONNOR B.J., DONEY S.C., FUNG I.Y., JACOB D.J., MILLER C.E., O'BRIEN D., PAWSON S., RANDERSON J.T., RAYNER P., SALAWITCH R.J., SANDER S.P., SEN B., STEPHENS G.L., TANS P.P., TOON G.C., WENNBERG P.O., WOFSY S.C., YUNG Y.L., KUANG Z., CHUDASAMA B., SPRAGUE G., WEISS B., POLLOCK R., KENYON D., SCHROLL S., *The Orbiting Carbon Observatory (OCO) mission*, Advances in Space Research, 2004, Vol. 34, 700.
- [12] CICERONE R.J., BARREN E.J., DICKINSON R.E., FUNG I.Y. et al., *Climate Change Science: An Analysis of Some Key Questions*, National Research Council, Washington 2001.
- [13] HAAG A., *Earth Monitoring: The crucial measurement*, Nature, 2007, Vol. 450, 785.
- [14] ROTHMAN L.S., JACQUEMART D., BARBE A., BENNER D.C., BIRK M., BROWN L.R., CARLEER M.R., CHACKERIAN C., CHANCE K., COUDERT L.H., DANA V., DEVI V.M., FLAUD J.M., GAMACHE R.R., GOLDMAN A., HARTMANN J.M., JUCKS K.W., MAKI A.G., MANDIN J.Y., MASSIE S.T., ORPHAL J., PERRIN A., RINSLAND C.P., SMITH M.A.H., TENNYSON J., TOLCHENOV R.N., TOTH R.A., VANDER AUWERA J., VARANASI P.,

- WAGNER G., *The HITRAN 2004 molecular spectroscopic database*, J. Quant. Spec. Rad. Transfer, 2005, Vol. 96, 139.
- [15] ROBICHAUD D.J., HODGES J.T., BROWN L.R., LISAK D., MASŁOWSKI P., OKUMURA M., MILLER C.E., *Experimental intensity and lineshape parameters of the oxygen A-band using frequency-stabilized cavity ring-down spectroscopy*, J. Mol. Spectrosc., 2008, Vol. 248, 1.
- [16] ROBICHAUD D.J., HODGES J.T., MASŁOWSKI P., YEUNG L.Y., OKUMURA M., MILLER C.E., BROWN L.R., *High-accuracy transition frequencies for the O₂ A-band*, J. Mol. Spectrosc., 2008, Vol. 251, 27.
- [17] DROSSART P., *Infrared spectroscopy of planetary atmospheres*, C. R. Physique, 2005, Vol. 6, 817.
- [18] SANCHEZ-FORTÚN STOKER J., DICKINSON A.S., *Line mixing in H broadening of the Na 3P–3D lines*, J. Phys. B., 2003, Vol. 36, 1309.
- [19] HOYT C. W., BARBER Z.W., OATES C.W., FORTIER T.M., DIDDAMS S.A., HOLBERG L., *Observation and Absolute Frequency Measurements of the ¹S₀-³P₀ Optical Clock Transition in Neutral Ytterbium*, Phys. Rev. Lett., 2005, Vol. 95, 083003.
- [20] BRICKMAN K.-A., CHANG M.-S., ACTON M., CHEW A., MATSUKEVICH D., HALJAN P.C., BAGNATO V.S., MONROE C., *Magneto-optical trapping of cadmium*, Phys. Rev. A, 2007, Vol. 76, 043411.
- [21] XU X., LOFTUS T.H., SMITH M.J., HALL J.L., GALLAGHER A., YE J., *Dynamics in a two-level atom magneto-optical trap*, Phys. Rev. A, 2002, Vol. 66, 011401.
- [22] FEDICHEV P.O., KAGAN YU., SHLYAPNIKOV G.V., WALRAVEN J.T.M., *Influence of Nearly Resonant Light on the Scattering Length in Low-Temperature Atomic Gases*, Phys. Rev. Lett., 1996, Vol. 77, No. 14, 2913; THEIS M., THALHAMMER G., WINKLER K., HELMWIG M., RUFF G., GRIMM R., HECKER DENSCHLAG J., *Tuning the Scattering Length with an Optically Induced Feshbach Resonance*, Phys. Rev. Lett., 2004, Vol. 93, No. 12, 123001; ENOMOTO K., KASA K., KITAGAWA M., TAKAHASHI Y., *Optical Feshbach Resonance Using the Intercombination Transition* (praca wysłana do publikacji); CIURYŁO R., TIESINGA E., JULIENNE P.S., *Optical tuning of the scattering length of cold alkaline-earth-metal atoms*, Phys. Rev. A, 2005, Vol. 71, 030701.
- [23] BIELSKI A., BRYM S., CIURYŁO R., DOMYSŁAWSKA J., LISICKI E., TRAWIŃSKI R.S., *Low pressure broadening and shift of the cadmium intercombination line 326.1 nm (⁵1S₀-⁵3P₁) perturbed by He and Ne*, J. Phys. B: At. Mol. Phys., 1994, Vol. 27, 5863.
- [24] BRYM S., DOMYSŁAWSKA J., *Influence of Krypton on the Shape and Shift of the 326.1 nm (⁵1S₀-⁵3P₁) Cd Intercombination Line*, Phys. Scr., 1995, Vol. 52, 511.
- [25] BRYM S., CIURYŁO R., LISICKI E., TRAWIŃSKI R.S., *Pressure Broadening and Shift of the 326.1 nm Cd Line Perturbed by Argon*, Phys. Scr., 1996, Vol. 53, 541.
- [26] BRYM S., CIURYŁO R., TRAWIŃSKI R.S., BIELSKI A., *Speed-dependent collisional effects on the 326.1-nm Cd line perturbed by Xe*, Phys. Rev. A, 1997, Vol. 56, No. 6, 4501.
- [27] CZUCHAJ E., SIENKIEWICZ J., *Pseudopotential calculation of the adiabatic potentials and oscillator strengths of cadmium-rare-gas pairs*, J. Phys. B: At. Mol. Opt. Phys., 1984, Vol. 17, 2251.
- [28] CZUCHAJ E., STOLL H., PREUSS H., *Pseudopotential SCF/CI calculations for the potential energies of the MHe and MNe (M=Mg, Cd, Hg) systems*, J. Phys. B: At. Mol. Opt. Phys., 1987, Vol. 20, 1487.
- [29] DOMYSŁAWSKA J., BIELSKI A., TRAWIŃSKI R.S., *Collision-time asymmetry of the ¹¹⁴Cd 326.1 nm line perturbed by Krypton*, J. Quant. Spectr. Radiat. Transf., 1999, Vol. 61, No. 6, 735.
- [30] TRAWIŃSKI R.S., *Asymetria linii widmowej 326.1 nm kadmu zaburzonej przez gazy szlachetne*, praca habilitacyjna, Toruń 2002.
- [31] BIELSKI A., LISAK D., TRAWIŃSKI R.S., SZUDY J., *Effects of Low-Polarizability Perturbbers on the Cadmium Intercombination Line*, Acta. Phys. Pol. A, 2003, Vol. 103, No. 1, 23.

- [32] BIELSKI A., LISAK D., TRAWIŃSKI R.S., *Collision-time asymmetry of the ^{114}Cd 326.1 nm line perturbed by Ar*, Eur. Phys. J. D, 2001, Vol. 14, 27.
- [33] TRAWIŃSKI R.S., BIELSKI A., LISAK D., *Collision-time asymmetry and speed-dependent effects on the ^{114}Cd 326.1 nm line perturbed by Kr*, Acta Phys. Polon. A, 2001, Vol. 99, 243.
- [34] BIELSKI A., CIURYŁO R., DOMYSŁAWSKA J., LISAK D., TRAWIŃSKI R.S., SZUDY J., *Laser-induced fluorescence study of collision-time asymmetry and speed-dependent effects on the ^{114}Cd 326.1 nm line perturbed by Xe*, Phys. Rev. A, 2000, Vol. 62, 032511.
- [35] LISAK D., *Ciśnieniowe rozszerzenie linii kadmu 326.1 nm zaburzonej przez gazy szlachetne*, praca doktorska, Toruń 2002.
- [36] BIELSKI A., CIURYŁO R., DOMYSŁAWSKA J., LISAK D., MASŁOWSKI P., SZUDY J., TRAWIŃSKI R.S., *Observation of the Line-Mixing and Collision-Time Asymmetry of the 5^1S_0 - 5^3P_1 Line of the Even-Odd ^{113}Cd Isotope*, Acta Phys. Pol. A, 2004, Vol. 105, 329.
- [37] MASŁOWSKI P., DOMYSŁAWSKA J., BIELSKA K., CIURYŁO R., LISAK D., TRAWIŃSKI R.S., SZUDY J., *Asymmetry of hyperfine-structure components of the 5^1S_0 - 5^3P_1 ^{113}Cd line perturbed by argon*, Eur. Phys. J. Special Topics, 2007, Vol. 144, 239.
- [38] BIELSKA K. *Rozszerzenie ciśnieniowe składowych nadsubtelnych linii widmowej 326,1 nm ^{113}Cd zaburzonej argonem i ksenonem*, praca magisterska, Toruń 2007.
- [39] CYGAN (ŚLIWIŃSKA) A., *Badanie ciśnieniowego rozszerzenia linii widmowej 326.1 nm izotopu kadmu 113 zaburzonej przez neon*, praca magisterska, Toruń 2008.
- [40] MASŁOWSKI P., *Zderzenia optyczne kadmu z gazami szlachetnymi*, praca doktorska, Toruń 2008.
- [41] LENZ W., *Eigene korrespondenzäßige Betrachtungen*, Z. Phys, 1924, Vol. 25, 299.
- [42] LENZ W., *General theory of spectral lines broadening*, Z. Phys, 1933, Vol. 80, 423.
- [43] WEISSKOPF V., *The Width of Spectral Lines in Gases*, Phys. Zeits., 1933, Vol. 34, 1.
- [44] LINDHOLM E., *Pressure Broadening of Spectral Lines*, Arkiv. Mat. Astr. Fys., 1945, Vol. 32A, 17.
- [45] FOLEY H.M., *The Pressure Broadening of Spectral Lines*, Phys. Rev., 1946, Vol. 69, No. 11-12, 616.
- [46] ANDERSON P.W., *Pressure Broadening in the Microwave and Infra-Red Regions*, Phys. Rev., 1949, Vol. 76, No. 5, 647.
- [47] ANDERSON P.W., *A Method of Synthesis of the Statistical and Impact Theories of Pressure Broadening*, Phys. Rev. A, 1952, Vol. 86, 809.
- [48] HOLTSMARK J., *On the Broadening of Spectral Lines*, Z. Phys, 1919, Vol. 20, 162.
- [49] HOLTSMARK J., *On the Broadening of Spectral Lines II*, Z. Phys., 1924, Vol. 25, 73.
- [50] JABŁOŃSKI A., *On the Impact Broadening of Spectral Lines and the Energy Transfer at Collisions*, Z. Phys., 1931, Vol. 70, 723.
- [51] KUHN H.G., LONDON F., *Limitation of the Potential Theory of the Broadening of Spectral Lines*, Phil. Mag., 1934, Vol. 18, 938.
- [52] KUHN H.G., *Pressure Shift of Spectral Lines*, Phys. Rev., 1937, Vol. 52, 133.
- [53] MARGENAU H., *Theory of Pressure Effects of Foreign Gases on Spectral Lines*, Phys. Rev., 1935, Vol. 48, 755.
- [54] MARGENAU H., *Statistical Theory of Pressure Broadening*, Phys. Rev., 1951, Vol. 82, 156.
- [55] JABŁOŃSKI A., *General Theory of Pressure Broadening of Spectral Lines*, Phys. Rev., 1945, Vol. 68, 78.
- [56] BARANGER M., *Simplified Quantum-Mechanical Theory of Pressure Broadening*, Phys. Rev., 1958, Vol. 111, No. 2, 481.
- [57] BARANGER M., *Problem of Overlapping Lines in the Theory of Pressure Broadening*, Phys. Rev., 1958, Vol. 111, No. 2, 494.
- [58] BARANGER M., *General Impact Theory of Pressure Broadening*, Phys. Rev., 1958, Vol. 112, No. 3, 855.

- [59] SOBELMAN I.I., *Broadening of Spectral Lines due to Collisions with Electrons*, Opt. Spekr., 1956, Vol. 1, 617.
- [60] SZUDY J., BAYLIS W.E., *Unified Franck-Condon Treatment of Pressure Broadening of Spectral Lines*, J. Quant. Spectrosc. Radiat. Transf., 1975, Vol. 15, 641.
- [61] SZUDY J., BAYLIS W.E., J. Quant. Spectrosc. Radiat. Transf., 1977, Vol. 17, 269 (errata do pracy [73]).
- [62] SZUDY J., BAYLIS W.E., *Asymmetry in Pressure-Broadened Spectral Lines*, J. Quant. Spectrosc. Radiat. Transf., 1977, Vol. 17, 681.
- [63] SZUDY J., *Rola zasady Francka–Conдона w zagadnieniu ciśnieniowego rozszerzenia linii widmowych*, praca habilitacyjna, Toruń 1976.
- [64] ALLARD N., KIELKOPF J., *The effect of neutral nonresonant collisions on atomic spectral lines*, Rev. Mod. Phys., 1982, Vol. 54, No. 4, 1103.
- [65] SZUDY J., BAYLIS W.E., *Profiles of line wings and rainbow satellites associated with optical and radiative collisions*, Phys. Rep., 1996, Vol. 266, 127.
- [66] CIURYŁO R., SZUDY J., TRAWIŃSKI R.S., *Non-adiabatic approach to the symmetry of pressure broadened spectral lines*, J. Quant. Spectrosc. Radiat. Transf., 1997, Vol. 57, No. 4, 551.
- [67] LORENTZ H.A., *The absorption and emission lines of gaseous bodies*, Proc. R. Acad., 1906, Vol. 8, 591.
- [68] ANDERSON P.W., TALMAN J.D., Conference on Broadening of Spectral Lines, University of Pittsburg, unpublished (1955); *Pressure Broadening of Spectral Lines at General Pressures*, Bell. Teleph. Syst., Tech. Publ. No. 3117.
- [69] AL-SAQABI B.N.I., PEACH G., *Unified theories of the pressure broadening and shift of spectral lines: II. Van der Waals interactions*, J. Phys. B, 1987, Vol. 20, 1175.
- [70] ROYER A., *Low density approximations in the adiabatic theory of pressure broadening*, Acta Phys. Pol. A, 1978, Vol. 54, 805.
- [71] KOLB A.C., GRIEM H.R., *Theory of Line Broadening in Multiplet Spectra*, Phys. Rev., 1958, Vol. 111, No. 2, 514.
- [72] FANO U., *Pressure Broadening as a Prototype of Relaxation*, Phys. Rev., 1963, Vol. 131, No. 1, 259.
- [73] BEN-REUVEN A., *Symmetry Considerations in Pressure Broadening Theory*, Phys. Rev., 1966, Vol. 141, No. 1, 34; *Impact Broadening of Microwave Spectra*, 1966, Vol. 145, No. 1, 7; *Spectral Line Shapes in Gases in the Binary–Collision Approximation*, Adv. Chem. Phys., 1975, Vol. 33, 235.
- [74] GORDON R.G., *On the Pressure Broadening of Molecular Multiplet Spectra*, J. Chem. Phys., 1967, Vol. 46, 448; GORDON R.G., MCGIUNIS R.P., *Line Shapes in Molecular Spectra*, J. Chem. Phys., 1968, Vol. 49, 2455.
- [75] COULOMBE M.J., PINE A.S., *Linear Scan Control Of Tunable Lasers Using a Scanning Fabry-Perot*, Applied Optics, 1979, Vol. 18, No. 10, 1505.
- [76] CIURYŁO R., SZUDY J., *Line-mixing and collision-time asymmetry of spectral line shapes*, Phys. Rev. A, 2001, Vol. 63, 042714.
- [77] RAUTIAN S.G., SOBELMAN I.I., *The Effect of Collisions on the Doppler Broadening of Spectral Lines*, Usp. Fiz. Nauk, 1966, Vol. 90, 209 (1966) [Sov. Phys. Usp., 1967, Vol. 9, 701].
- [78] PINE A.S., *Line shape asymmetries in Ar-broadened HF ($\nu = 1-0$) in the Dicke-narrowing regime*, J. Chem. Phys., 1994, Vol. 101, 3444.
- [79] PINE A.S., *Asymmetries and correlations in speed-dependent Dicke-narrowed line shapes of argon-broadened HF*, J. Quant. Spectrosc. Radiat. Transf., 1999, Vol. 62, 397.
- [80] LISAK D., BIELSKI A., CIURYŁO R., DOMYŚLAWSKA J., TRAWIŃSKI R.S., SZUDY J., *On the role of Dicke narrowing in the formation of atomic line shapes in the optical domain*, J. Phys. B: At. Mol. Opt. Phys., 2003, Vol. 36, 3985.

- [81] HARRIS M., LEWIS E.L., MCHUGH D.R., SHANNON I., *The full Voigt profile and collision time asymmetry for profiles of calcium 442.7 nm perturbed by krypton*, J. Phys. B:At. Mol. Phys., 1984, Vol. 17, L661.
- [82] BERMAN P.B., *Speed-Dependent Collisional Width and Shift Parameters in Spectral Profiles*, J. Quant. Spectrosc. Radiat. Transf., 1972, Vol. 12, 1331.
- [83] WARD J., COOPER J., SMITH E.W., *Correlation Effects in the Theory of Combined Doppler and Pressure Broadening – I. Classical Theory*, J. Quant. Spectrosc. Radiat. Transf., 1974, Vol. 14, 555.
- [84] DICKE R.H., *The Effect of Collisions upon the Doppler Width and Spectral Lines*, Phys. Rev., 1953, Vol. 89, No. 2, 472.
- [85] WITTKÉ J.P., DICKE R.H., *Redetermination of the Hyperfine Splitting in the Ground State of Atomic Hydrogen*, Phys. Rev., 1956, Vol. 103, No. 3, 620.
- [86] CIURYŁO R., SZUDY J., *Speed-dependent pressure broadening and shift in the soft collision approximation*, J. Quant. Spectrosc. Radiat. Transf., 1997, Vol. 57, No. 3, 411.
- [87] FARROW R.L., RAHN L.A., SITZ G.O., ROSASCO G.J., *Observation of the Speed-Dependent Collisional Inhomogeneity in H₂ Vibrational Line Profiles*, Phys. Rev. Lett., 1989, Vol. 63, No. 7, 746.
- [88] GALATRY L., *Simultaneous Effect of Doppler and Foreign Gas Broadening on Spectral Lines*, Phys. Rev., 1961, Vol. 122, No. 4, 1218.
- [89] NELKIN M., GHATAK A., *Simple Binary Collision Model for Van Hove's G_s(r, t)*, Phys. Rev., 1964, Vol. 135, No. 1A, A4.
- [90] NESMEJANOV A.N., *Dawlenie para chemiczeskich elementow*, Akademia Nauk ZSRR, Moskwa 1961.
- [91] GERSTENKORN S., LUC P., *Atlas du Spectre d'Absorption de la Molecule d'Iode*, Editions du C.N.R.S., Paris 1978.
- [92] GERSTENKORN S., LUC P., *Absolute iodine (I₂) standards measured by means of Fourier transform spectroscopy*, Rev. Phys. Appl., 1979, Vol. 14, 791.
- [93] *Ring Dye Laser Systems. The Essence of Innovation*, Coherent, Laser Products Division, Palo Alto, USA 1988.
- [94] HARRIS M., LEWIS E.L., MCHUGH D., SHANNON I., *Measurement of the broadening and asymmetry of the calcium resonance line (422.7 nm)*, J. Phys. B: At. Mol. Phys., 1986, Vol. 19, 3207.
- [95] BIELSKI A., CIURYŁO R., DOMYSŁAWSKA J., LISAK D., TRAWIŃSKI R.S., WOLNIKOWSKI J., *Pressure broadening and shift of the 326.1 nm Cd line perturbed by H₂ and D₂*, Acta Phys. Pol. A, 2000, Vol. 97, No. 6, 1003.
- [96] VITCU A., CIURYŁO R., WEHR R., DRUMMOND J.R., MAY A.D., *High-resolution tunable mid-infrared spectrometer based on difference-frequency generation in AgGaS₂*, Applied Optics, 2004, Vol. 43, No. 25, 4965.
- [97] WEHR R., DRUMMOND J.R., MAY A.D., *Design of a difference-frequency infrared laser spectrometer for absorption line-shape studies*, Applied Optics, 2007, Vol. 46, 978.
- [98] DEMTRÖDER W., *Laser Spectroscopy*, 3rd Edition, Springer 2002.
- [99] *Instrukcja obsługi analizatora widma*, COBRABiD, Wrocław 1992.
- [100] BOYD G.D., GORDON J.P., *Confocal multimode resonator for millimeter through optical wavelength masers*, Applied Optics, 1961, Vol. 40, 489.
- [101] BOYD G.D., KOGELNIK H., *Generalized confocal resonator theory*, Bell Syst. Techn. J., 1962, Vol. 41, 1347.
- [102] MASŁOWSKI P., *Układ pomiarowy do badania długoczasowych zaników fluorescencji*, praca magisterska, Toruń 2003.

- [103] *Specyfikacja techniczna lambdametru WA-1500*, Burleigh.
- [104] ALLAN D.W., *Statistics of Atomic Frequency Standards*, Proceedings of the IEEE, 1966, Vol. 54, No. 2, 221-230.
- [105] MOORE C.E., *Atomic energy levels*, NSRDS-NBS, Vol. 3, U.S. Dep. of Commerce, National Bureau of Standards, Washington 1971.
- [106] MARQUARDT D.W., *An algorithm for least-squares estimation of non-linear parameters*, J. Soc. Industr. Appl. Math., 1963, Vol. 11, 431.
- [107] PRESS W., TEUKOLSKY S., VETTERLING W., FLANNERY B., *Numerical Recipes in Fortran*, Cambridge University Press, Cambridge 1992.
- [108] CIURYŁO R., BIELSKI A., DOMYSŁAWSKA J., SZUDY J., TRAWIŃSKI R.S., *Effect of dissociative recombination on spectral line profiles in neon glow discharge*, J. Phys. B: At. Mol. Opt. Phys., 1994, Vol. 27, 4181.
- [109] Certyfikat izotopu ^{113}Cd wystawiony przez firmę Techsnabexport.
- [110] CIURYŁO R., *Shapes of pressure- and Doppler-broadened spectral lines in the core and near wings*, Phys. Rev. A, 1998, Vol. 58, No. 2, 1029.
- [111] BRANDT S., *Analiza danych*, Wydawnictwo Naukowe PWN, Warszawa 1998.
- [112] SHANNON I., HARRIS M., MCHUGH D.R., LEWIS E.L., *Low-pressure spectral line profiles: an analysis in terms of asymmetric speed-dependent Voigt profiles*, J. Phys. B: At. Mol. Phys., 1986, Vol. 19, 1409.

ROLE OF FREQUENCY MEASUREMENT IN MEASUREMENTS AND ANALYSIS OF SPECTRAL LINE SHAPES

This paper describes high resolution digital laser spectrometer and precise measurements of shape of 326.1 nm ^{113}Cd and ^{114}Cd line performed in the Institute of Physics at Nicolaus Copernicus University. Special attention was paid to precision of the measurement of dye laser beam frequency. Methods of data analysis and revision of results are also presented.



**Wydawnictwa Politechniki Wrocławskiej
są do nabycia w księgarni „Tech”
plac Grunwaldzki 13, 50-377 Wrocław
budynek D-1 PWr., tel. 071 320 29 35
Prowadzimy sprzedaż wysyłkową**

ISBN 978-83-7493-468-8