TEMAT

Х

OPTYKA FALOWA

1. Interferencja fal płaskich 🔶

Światło podobnie jak dźwięk możemy opisać jako falę, lecz w przeciwieństwie do dźwięku nie jest to fala podłużna (rys. 1.1) tylko poprzeczna



Rysunek 1.1. W fali dźwiękowej przemieszczają się obszary zagęszczonego i rozrzedzonego ośrodka (np. powietrza). Pasek u góry ilustruje, poprzez odcienie szarości, gęstość ośrodka. Dolna część rysunku ilustruje zmiany ciśnienia fali harmonicznej dźwiękowej. Taką falę nazywamy falą podłużną.

Fala świetlna jest szczególnym przypadkiem fali elektromagnetycznej.



Rysunek 1.2. Fala świetlna jest falą poprzeczną. Wektor pola elektrycznego **E** i magnetycznego **H** drga w kierunku prostopadłym do kierunku rozchodzenia się światła. Przez drgania wektora rozumiem tu oscylacje jego długości. Na rysunku światło rozchodzi się w kierunku wskazywanym przez długą strzałkę.

Definicja 1.1: Fala podłużna

Fala, której kierunek drgań jest zgodny z kierunkiem jej rozchodzenia się nazywamy falą podłużną

Definicja 1.2: Fala poprzeczna

Fala, której kierunek drgań jest prostopadły do kierunku jej rozchodzenia się nazywamy falą poprzeczną

Rysunek (1.3) pokazuje spektrum fal elektromagnetycznych. Spektrum to dzieli się na trzy duże klasy fal: fale radiowe o długości z przedziału [~1mln km–1mm], optyczne o długości z przedziału [1mm-0.01µm] i gamma o długości z przedziału [0.01µm-~1pm]. Fale optyczne dzielą się na trzy zakresy: podczerwień - $\lambda \in [1mm - 0,7\mum]$, światło widzialne - $\lambda \in [0,75\mu$ m-0,4µm] i nadfiolet - $\lambda \in [0,4\mu$ m-0,01µm]. Przedmiotem optyki falowej są własności fal

z zakresu optycznego. Niemniej wiele z tego co zostanie tu powiedziane dotyczy całego zakresu fal elektromagnetycznych. Podstawową teorią opisującą zjawiska elektromagnetyczne jest klasyczna elektrodynamika oparta o układ równań Maxwella (TXX) i zasadę zachowania ładunku elektrycznego (TXX). Z równań Maxwella wynika równanie falowe, którego rozwiązania, w ramach elektrodynamiki, interpretujemy jako fale elektromagnetyczne. Jednak optyka falowa zaczęła się rozwijać przed sformułowaniem układ równań Maxwella. W swej pierwszej wersji optyka falowa opierała się o zasadę Huygensa (§TIX 2)



Rysunek 1.3. Widmo (spektrum fal elektromagnetycznych). Zwróć uwagę jak wąskie jest pasmo światła widzialnego; źródło Wikipedia

Obok optyki falowej mamy również optykę geometryczną. Zwykle wykład z optyki zaczyna się od optyki geometrycznej, która uważana jest za teorię prostszą i powstała przed teorią falową. Skoro jednak dopiero co zakończyłem omawiać fale, to warto na gorąco przejść do optyki falowej, a dopiero w następnym kroku omówić optykę geometryczną.

Granice między zakresem fal radiowych, optycznych i gamma są dość umowne. Podstawą tego podziału jest to, że oddziaływanie fal z materią zależy od ich długości. Od fal optycznych oczekujemy między innymi, że będzie je można kształtować odwołując się do zjawiska załamania. Jednak już w dalekim nadfiolecie nie ma materiałów, które efektywnie załamuję fale świetlne. Z drugiej strony nie powinniśmy mieć możliwości emisji fal optycznych technikami radiowymi. Granica tej niemożności przesuwa się jednak, wraz z rozwojem techniki, w kierunku fal o mniejszej długości fali, co niekończenie znajduje odzwierciedlenie w umownych granicach między poszczególnymi rodzajami fal. W praktyce więc korelacja między podanymi zakresami fal a ich sposobem oddziaływania z materią jest dość luźna.

Na dobry początek powrócę do zagadnienia interferencji fal płaskich w kontekście fal świetlnych. Obliczę obraz interferencyjny dwóch fal płaskich.

Zadanie 1.1:

Na kliszy fotograficznej, leżącej w płaszczyźnie *z*=0 rejestrujemy interferencję dwóch fal płaskich, o tej samej długości fali $\lambda = 632,8$ nm i tej samej amplitudzie. Wektor falowy pierwszej fali (rysowanej na niebiesko) ma współrzędne $\mathbf{k}_{N}(0,0,k)$, gdzie $k=2\pi/\lambda$. Wektor falowy drugiej fali (rysowanej na czerwono) ma współrzędne $\mathbf{k}_{C}(0,0.01k,?)$. Załóż, że w początku układu współrzędnych fazy obu fal są zgodne. Oblicz wynik nakładania się tych fal.

Zwracam waszą uwagę na fakt, że fale świetlne są falami trójwymiarowymi. Zatem powierzchnia falowa świetlnej fali płaskiej tworzy zbiór płaszczyzny oddalonych o długość fali λ (rys. TIX 1.1.8).



Rysunek 1.4. Interferencję dwóch fal płaskich obserwujemy w płaszczyźnie ekranu E, która pokrywa się z osią y układu współrzędnych. Front falowv fali narysowanej na niebiesko jest równoległy do ekranu. W początku układu współrzędnych fazy obu fal są takie same, a suma obu fazorów (czarna strzałka) ma największą możliwą wartość. W innym punkcie ekranu P₁ faza fali niebieskiej jest taka sama, gdyż punkty ekranu leżą na powierzchni falowej fali niebieskiej. Faza fali czerwonej jest inna, gdyż powierzchnia falowa fali czerwonej nie pokrywa się z powierzchnią ekranu. Przez punkt P1 przechodzi inna powierzchnia równej fazy fali czerwonej (narysowana linią czerwoną przerywaną) reprezentująca inną wartość fazy. Odcinek d pokazuje różnice dróg między tymi obiema powierzchniami falowymi.

Przy rozwiązaniu tego zadania posłużę się modelem fazorowym (§TVIII 2.3). Zorientuję oś y układu współrzędnych zgodnie z płaszczyzną kliszy fotograficznej (ekranu) (rys. 1.4). W początku układu współrzędnych kąt między fazorami obu fal jest równy zeru. Obliczam sumę obu fazorów (rys. 1.4) oraz kwadrat tej sumy, czyli wartość natężenia światła. Przejdę teraz do innego punktu kliszy – P₁. W tym innym punkcie kliszy wektor falowy pierwszej fali płaskiej jest taki sam jak w początku układu współrzędnych – w końcu powierzchnia falowa to powierzchnia, na której fazory mają tą samą fazę. Ale faza drugiej fali jest inna, gdyż punkt P₁ leży na powierzchni falowej, która odpowiada innemu kątowi fazowemu. Musimy obliczyć wartość tego kąta.

W tym celu poprowadzę odcinek prostopadły do powierzchni falowych i przechodzący przez punkt P_1 . Odcinek *d* wyznacza odległość między

powierzchnią falową przechodzącą przez początek układu współrzędnych, a powierzchnią przechodzącą przez punkt P₁. Wiemy, że gdy przesuniemy się wzdłuż prostopadłej, o odcinek równy długości fali λ , to trafimy ponownie na powierzchnie, gdzie fazory mają tą samą fazę (rys. TIX 1.5). Ale gdy przesuniemy się o część λ , to wtedy faza zmieni się proporcjonalnie do wielkości tej części. Na przykład jeżeli odcinek *d* będzie miał długość równą pół długości fali to faza zmieni się o połowę całego obrotu, jeżeli będzie to jedna trzecia długości fali, to o jedną trzecią obrotu, jeżeli jedna dziesiąta długości fali to o jedną dziesiątą pełnego obrotu i tak dalej. Słowem faza fali zmieni się o

$$\delta \varphi = 2\pi \frac{d}{\lambda} = kd \tag{1.1}$$

Zobacz jak pracuje wzór (1.1). Wielkość d/λ mówi ile długości fal zmieści się w odcinku o długości d, a 2π , to wartość kąta pełnego wyrażonego w radianach. Gdy na przykład $d = \frac{1}{2}\lambda$, to $\delta \varphi = \pi$, czyli pół obrotu. A co będzie gdy na przykład $d = \frac{3}{2}\lambda$? Wtedy $\delta \varphi = 7\pi$, co daje trzy pełne obroty i pół kolejnego. Pełne obroty nas nie interesują, zatem ponownie zostaje pół obrotu. Aby zatem obliczyć różnicę faz między obiema powierzchniami falowymi musimy odległość między nimi przemnożyć przez liczbę falową k. Trzeba tylko obliczyć długość odcinka d. Aby to zrobić musimy wiedzieć jaki jest kąt nachylenia drugiej fali do osi układu współrzędnych. Informację tą skrywa w sobie wektor falowy.

Wektor falowy pierwszej fali ma niezerową tylko *z*-tową współrzędną. Oznacza to, że wektor ten jest równoległy do osi *z*, a powierzchnia falowa jest do tej osi prostopadła (rys. 1.5a). Wyznaczę *z*-tową współrzędną wektora falowego \mathbf{k}_2 drugiej fali. Skorzystam z faktu, że suma kwadratów współrzędnych wektora daje nam kwadrat długości tegoż wektora.

$$k_{Cz} = \pm \sqrt{k^2 - k_{Cy}^2} = \pm \sqrt{k^2 - 0.01k^2} = \pm \sqrt{0.99k}$$
 1.2

Pozostaje jeszcze kwestia znaku. Powiedzmy, że obie fale rozchodzą się zgodnie z kierunkiem osi *z*, wtedy współrzędna *z*-towa wektora falowego drugiej fali jest dodatnia.

$$k_{Cz} = \sqrt{0,99}k$$

Współrzędna *x*-owa wektora falowego drugiej fali jest równa zeru, co oznacza, że wektor ten jest nachylony tylko względem osi *y* i *z*; w efekcie powierzchnia falowa jest również nachylona tylko względem tych osi (rys. 1.5a). Obliczę kąty nachylenia wektora falowego drugiej fali do poszczególnych osi. Skorzystam z ogólnej zależności (rys. 1.5b).

$$\cos(\alpha_q) = \frac{k_q}{k}$$
 1.3



Rysunek 1.5. Interferencja dwóch fal z zadania (1.1). Dla czytelności rysunku powierzchnia falowa fali rysowanej na niebiesko została nieco odsunięta od osi *y*. Należy ją sobie wyobrażać jako nałożoną na oś *y*. Druga fala (rysowana na czerwono) również została odsunięta o ten sam mały odcinek. Fala narysowana na czerwono jest pochylona względem osi *y*. Rysunek **b** pokazuje kąty wektora falowego fali pochylonej. Rysunek **c** pokazuje trójkąt DEF równoważny trójkątowi ABC. Na różowo zaznaczone są obszary, w których powierzchnia falowa fali pochylonej wyprzedza, na ekranie, (pod względem fazy), falę równoległą do ekranu. Aby uzyskać fazę tej fali na ekranie należy do fazy w początku układu współrzędnych dodać fazę równą iloczynowi *kd*. W obszarze zaznaczonym na żółto fala pochylona spóźnia się. Tu należy wielkość *k d* dodać ze znakiem minus.

Kąt α_q jest kątem pomiędzy wektorem falowym a daną osią $q \in \{x, y, z\}$. Kosinusy tych kątów nazywają się kosinusami kierunkowymi wektora (§DA 2.6). Trzy takie kosinusy są współrzędnymi wektora o jednostkowej długości, równoległego do wektora wyjściowego. Jednostkowy wektor w kierunku wektora falowego będę oznaczał przez **k** z daszkiem $\hat{\mathbf{k}}(\cos(\alpha_x), \cos(\alpha_y), \cos(\alpha_z))$.

Z treści zadania wynika, że w początku układu współrzędnych fazy obu fal są takie same, czyli kąt fazowy między fazorami jest równy zeru. Znajdę wartość tego kąta w innych punktach ekranu. Będę poruszał się wzdłuż osi y. Zgodnie z rysunkiem (1.4) widać, że długość odcinka d, dla punktu P₁ ekranu wyraża się wzorem

$$d = y \cos(\alpha_y) = y \frac{k}{k_y}$$
 1.4

Skorzystałem z faktu, że relacje kątowe w trójkącie ABC z rysunku (1.5b) są takie same jak w trójkącie DEF zbudowanym na bazie wektora falowego (rys. 1.5c). Skoro mam obliczoną długość odcinka *d*, w zależności od położenia

punktu P_1 na osi y to korzystając z wyrażenia (1.1) mogę obliczyć zmianę wartości fazy drugiej fali, gdy przesuniemy się wzdłuż odcinka d.

$$\delta \varphi = k_C d = k_C \frac{k_{Cy}}{k_C} y = y k_{Cy}$$
 1.5

Ten prosty wynik wymaga komentarza. Po pierwsze widać jak ładnie "pracuje" wektor falowy. Jeżeli chcemy policzyć zmianę fazy przy przesunięciu wzdłuż jakiejś linii (u nas jest to linia będącą osią y) to wystarczy wielkość tego przesunięcia pomnożyć przez wartość rzutu wektora falowego na tą linię. Po drugie dla dodatnich y znak przy $\delta \varphi$ będzie przeciwny do znaku otrzymanego dla ujemnych y. Co oznacza ta dodatnia (ujemna) zmiana fazy wyjaśnia rysunek (1.5a).

Jeszcze jedna ważna rzecz. Wzór (1.5) pozwala obliczyć przyrost różnicy fazy pomiędzy fazorami obu fal, kiedy poruszamy się wzdłuż osi *y* zaczynając od punktu w początku układu współrzędnych. Nie uwzględnia natomiast faktu, że w początku układu współrzędnych różnica faz nie musi być równa zeru: $\delta \varphi_0 \neq 0$. Aby obliczyć fazę należy do obliczonego przyrostu dodać wartość początkową tej różnicy, czyli $\delta \varphi_0$. W naszym zadaniu ta początkowa różnica faz jest równa zeru, więc nie ma czego dodawać, jednak nie zawsze tak jest. Ostatecznie wzór na fazę fali czerwonej na ekranie, w danej ustalonej chwili czasu, przyjmie postać:

$$\delta \varphi = y k_{Cy} + \delta \varphi_0$$

Zanim przejdę dalej, pokażę ważny trik. Jeżeli obliczamy obraz interferencyjny (natężeniowy) jaki dają dwie interferujące fale (niekoniecznie płaskie), to możemy uznać, że fazory reprezentujące jedną z tych fal mają wszędzie kąt fazowy równy zeru. I to nawet wtedy kiedy ta wybrana fala nie jest równoległa do ekranu (rys. 1.6). Jest to równoważne obróceniu wszystkich fazorów, zaczepionych w danym punkcie o taki kąt, że fazor reprezentujący wybraną falę ma fazę równą zeru. W przypadku fali niebieskiej, równoległej do ekranu, możemy przyjąć, że faza fali niebieskiej jest równa zero wszędzie na ekranie. Fazory \mathbf{v}_{N} , które reprezentują tą falę mają współrzędne

$$v_{Nx} = A \text{ oraz } v_{Ny} = 0$$

1.7

1.6

Faza fazorów reprezentujących falę czerwoną różni się od fazy fazorów reprezentujących falę niebieską o $\delta \varphi$ dane wzorem (1.6). Wyznaczając kąt fazorów dla fali czerwonej względem odnośnych kątów fazowych fali niebieskiej możemy zapisać.

$$v_{Cx} = A\cos(\delta\varphi) \text{ oraz } v_{Cy} = A\sin(\delta\varphi)$$
 1.8



Rysunek. 1.6. a) W czterech punktach ekranu (P₁, ..., P₄) wyznaczone są fazory dwóch interferujących fal (niebieskie i czerwone strzałki). Fazory wypadkowe narysowane są na czarno. Wypadkowe fazory mają różne kąty fazowe, ale z punktu widzenia pomiaru nie ma to znaczenia, gdyż rejestrujemy natężenie światła (b), którego miarą jest kwadrat długości fazora. Operacja obliczenia kwadratu długości fazora gubi informację o jego fazie; c) dlatego przy obliczaniu wyniku pomiaru natężenia światła możemy fazory, w każdym punkcie obrócić tak, żeby fazory niebieskie miały kąt fazowy równy zero. W niczym nie zmieni to obliczonego natężenia światła.

Mogę teraz dodać, w każdym punkcie ekranu, fazory obu fal, w efekcie czego otrzymam nowy fazor w o współrzędnych

$$w_x = A + A\cos(\delta\varphi) = A(1 + \cos(\delta\varphi)) \text{ oraz } w_y = A\sin(\delta\varphi)$$
1.9

Natężenie światła to kwadrat długości fazora w

$$I = w_x^2 + w_y^2 = A^2 \left[\left(1 + \cos(\delta\varphi) \right)^2 + \sin^2(\delta\varphi) \right]$$

= 2A² (1 + cos(\delta\varphi)) 1.10

Korzystając z (1.6) mam:

$$I = 2A^2 \left(1 + \cos(k_{Cy}y + \delta\varphi_0) \right)$$
 1.11

W ten sposób rozwiązaliśmy zadanie dla punktów rozmieszczonych wzdłuż osi y. Co się stanie, gdy dla danego y zaczniemy poruszać się wzdłuż osi x? Nic się nie stanie, gdyż interferujące fale, nie są nachylone względem siebie wzdłuż tej osi. Zatem jeżeli poruszamy się wzdłuż osi x to, przy tym samym y odcinek d ma tą samą długość. Zmiany natężenia będą zachodziły tylko wzdłuż osi y, tak jak to pokazane jest na rysunku (1.7). Ten układ ciemnych i jasnych linii nazywa się interferogramem. Obliczony układ prążków jest periodyczny. Możemy obliczyć jego okres, czyli odległość pomiędzy środkami sąsiednich prążków jasnych lub ciemnych. Umówmy się, że przez jeden prążek będziemy rozumieli dwa sąsiednie paski jeden jasny drugi ciemny. Zgodnie z rysunkiem (1.4) wartość natężenie powtórzy się tam, gdzie odcinek d będzie równy

dokładnie λ , lub faza pod funkcją cosinus we wzorze (1.5) zmieni się o 2π , stąd mamy warunek

$$\delta y k_{Cy} = 2\pi \Longrightarrow \delta y = \frac{2\pi}{k_{Cy}}$$
1.12
$$\bigcup_{\substack{\text{UA=UB=1}\\ \text{oxA=cyA=0\\ \text{oxB=0; cyB=0.001}}} \bigcup_{\substack{\text{UA=UC=1}\\ \text{oxA=cyA=0\\ \text{oxC=0; cyC=0.002}}} \bigcup_{\substack{\text{UA=1; UD=1/3\\ \text{oxA=cyA=0\\ \text{oxD=0; cyD=0.001}}} \bigcup_{\substack{\text{oxD=0; cyD=0.001}}} \bigcup_{$$

Rysunek 1.7. Górna część rysunku przedstawia wygenerowane numerycznie interferogramy dwóch fal. *UA*, *UB*, *UC*, *UD* to amplitudy fal, cxA = cxB = cxC = cxD = 0 to cosinusy kierunkowe fal względem osi *x*. Z faktu, że są one równe zeru wynika, że powierzchnie falowe (wektory falowe) nie są nachylone w tym kierunku. *cyA*, *cyB*, *cyC*, *cyD* to cosinusy kierunkowe fal względem osi *y*. Dolna część rysunku przedstawia zmianę natężenia światła wzdłuż wybranej linii (oś y na pierwszym interferogramie) prostopadłej do kierunku prążków interferencyjnych.

Podstawiając wartość $k_{Cy} = 99,2918$ 1/mm otrzymujemy $\delta_y=0.063$ mm. No cóż, te prążki są gęste, ale wystarczy zmniejszyć k_{Cy} i prążki będą rzadsze, a przez to lepiej widoczne. Przy odpowiednio małym k_{Cy} mogą być nawet widoczne gołym okiem. Przykład zmiany grubości prążków interferencyjnych, przy zmianie względnego nachylenia fal możesz zobaczyć porównując lewy i środkowy interferogram z rysunku (1.6).

Zbadam przypadki graniczne. Niech wektory falowe obu fal będą równoległe. W tym przypadku oba wektory falowe mają takie same współrzędne; w szczególności $k_{Cy}=k_{Cy}=0$. Ze wzoru (1.12) wynika, że okres

prążków jest nieskończony, czyli że nie ma żadnych prążków. Ten sam wniosek możemy wyciągnąć z rysunku (1.4). Jeżeli oba fronty falowe są wzajemnie równoległe, to odległości *d* są wszędzie takie same. Zatem w płaszczyźnie całego ekranu mamy taką samą różnicę faz $\delta \varphi = \delta \varphi_0$ pomiędzy oboma falami i taką samą wartość sumy obu fal. Powiedzmy, że takie dwie równoległe powierzchnie falowe spotykają się dokładnie w fazie, to znaczy że kąty fazowe na ekranie są zgodne ($\delta \varphi = 0$). Amplitudy tych fal dodają się w prosty sposób; jeżeli obie fale mają te same amplitudy to po ich dodaniu mamy 2*A*, *a* natężenie w każdym punkcie ekranu wynosi 4*A*². Rozważmy teraz inną możliwość, obie fale spotykają się w przeciwnej fazie. Ich amplitudy nawzajem zniosą się i amplituda wypadkowa, tak samo jak natężenie na całym ekranie będzie równa zeru. A jeżeli kąt fazowy między falami będzie jakiś inny? Odwołam się do wzoru (1.11), kładąc $k_{Cv}=0$.

$$I = 2A^2 (1 + \cos(\delta \varphi_0))$$

1.13

Ponieważ faza początkowa φ_0 , w rozważanym przypadku szczególnym, jest w każdym punkcie ekranu stała wzór (1.13) daje, w każdym punkcie ekranu, tą samą liczbę. Kładąc $\delta\varphi_0=0$ lub $\delta\varphi_0=\pi$ otrzymamy dwa szczególne przypadki omówione wyżej. Dla innych wartości $\delta\varphi_0$ otrzymujemy wartości pośrednie, ale w całej płaszczyźnie ekranu takie same.

Jeżeli jesteś wyczulony na podstawowe zasady fizyki, to masz pewnie pytanie. Jeżeli dwie fale płaskie nałożymy na siebie tak aby były wzajemnie do siebie równoległe i nakładały się w przeciwfazie, to fale te ulegają tzw. interferencji destruktywnej i wypadkowe natężenie światła wynosi zero. Ale pierwsza fala niesie energię wyrażoną poprzez kwadrat jej amplitudy A^2 , podobnie druga. Razem energia niesiona przez te fale wyraża się przez $2A^2$. Jeżeli fale wygaszą się to energia znika i złamana zostaje zasada zachowania energii. A jeżeli fale interferują w fazie (interferencja konstruktywna), to zgodnie ze wzorem (1.13) ich energia wynosi $4A^2$, czyli dwa razy więcej niż wnoszą obie fale. Czyżby perpetuum mobile? Do sprawy tej jeszcze wrócę; na razie wstrzymaj się jeszcze na trochę od sprintu do urzędu patentowego¹.

Przy okazji przyglądania się rysunkowi (1.7) warto zrobić następującą uwagę. Falę charakteryzujemy używając dwa wektory: wektor falowy i fazor. Od strony fizycznej interpretacji to są dwa bardzo różne wektory. Wektor pokazuje kierunek rozchodzenia sie energii fali i "żvie" falowy w trójwymiarowej przestrzeni. Fazor nie pokazuje żadnego kierunku w przestrzeni. Fazor "żyje" we własnej dwuwymiarowej przestrzeni (rys. 1.8) i informuje nas o fazie i amplitudzie fali w danym punkcie. Obliczając zadania z optyki falowej nie można w żadnym razie mylić ze sobą tych dwóch "gatunków" wektorów

¹ Zresztą, jak pisałem w (§TIII_IV), urzędy patentowe nie przyjmują zgłoszeń dotyczących Perpetuum Mobile.

Do tej pory zakładałem, że dwie interferujące fale mają takie same amplitudy. Teraz założę, że amplitudy są różne: pierwsza fala ma amplitudę A_1 a druga A_2 ; przy czym niech $A_2 < A_1$. Efekty nakładania się fal o różnych amplitudach pokazane są na rysunku (1.7) (prawy interferogram). Obliczmy dokładnie jak wygląda obraz interferencyjny. Wzory (1.8) i (1.10) przyjmą teraz postać:

$$v_{Nx} = A_1 \text{ oraz } v_{Ny} = 0$$
 1.142

$$v_{Cr} = A_2 \cos(\delta \varphi) \operatorname{oraz} v_{Cr} = A_2 \sin(\delta \varphi)$$
 1.14b

Wzór (1.9) na współrzędne wypadkowego wirującego wektora ma teraz postać

$$w_x = A_1 + A_2 \cos(\delta \varphi)$$
 oraz $w_y = A_2 \sin(\delta \varphi)$ 1.15
Wzór (1.10) na długość wektora wypadkowego, czyli natężenie ma postać
 $I = w_x^2 + w_y^2 = A_1^2 + A_2^2 \cos^2(\delta \varphi) + 2A_1 A_2 \cos(\delta \varphi) + A_2^2 \sin^2(\delta \varphi)$ 1.16
Wzór ten często zapisuje się za pomocą natężeń
 $I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\delta \varphi)$ 1.17



Rysunek 1.8. Na ekranie (szara płaszczyzna) rejestrujemy interferencje dwóch fal płaskich. Na rysunku pokazana jest powierzchnia falowa fali czerwonej. Fala niebieska (równoległa do ekranu) nie została narysowana. Wektor k (czarna wektorem strzałka) jest falowym czerwonej fali. Współrzędne wektora falowego jak również punkty na ekranie i frontv falowe interferujacych fal wyznaczone są w układzie współrzędnych narysowanym w górnym, prawym rogu. W każdym punkcie ekranu dodajemy do siebie fazory fali czerwonej i niebieskiej. Fazory "żyją" przestrzeni te w dwuwymiarowej. Tak długo jak interesuje nas natężenie interferujących fal (czyli to co mierzymy w eksperymencie) tak długo każdym punkcie ekranu w możemv zaczepiać w dowolny sposób (nie patrząc na inne punkty ekranu) lokalne układy współrzędnych, w których liczymy sumy fazorów interferujacych fal.

gdzie I_1 i I_2 to natężenie pierwszej i drugiej fali płaskiej. Trzeci przykład prążków na rysunku (1.7) ma mniejszy kontrast. W tym trzecim przykładzie

11

jedna fala jest słabsza od drugiej, co powoduje, że minima natężenia są różne od zera a maksima mniejsze niż w przypadku, gdy obie amplitudy byłyby równe wartości większej (rys. 1.9). Widać to również w dolnej części rysunku (1.7). Trzeciemu interferogramowi odpowiada linia zielona o mniejszej różnicy pomiędzy maksymalną i minimalna różnicą natężeń. Mniejsza różnica natężeń to właśnie mniejszy kontrast, który widzimy na trzecim interferogramie. Wprowadzę wielkość, która nazywa się współczynnikiem kontrastu prążków interferencyjnych (interferogramu).

$$\gamma = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$
 1.18

Gdzie I_{max} to największe natężenie światła na interferogramie a I_{min} to najmniejsze natężenie światła na interferogramie. Tak zdefiniowana wielkość jest równa jedności, gdy I_{min} jest równe zeru. Jest to zarazem największa możliwa wartość tej wielkości. Gdy $I_{min}=I_{max}$ kontrast jest równy swej najmniejszej możliwej wartość czyli zeru; wtedy oczywiście żadne prążki nie są widoczne. Korzystając ze wzorów (1.18) i (1.17) obliczę kontrast dla interferogramu dwóch fal płaskich. Największe natężenie mamy gdy $\cos(\delta \varphi)=1$ a najmniejsze gdy $\cos(\delta \varphi)=-1$. Kontrast wynosi

$$\gamma = \frac{I_1 + I_2 + 2\sqrt{I_1I_2} - I_1 - I_2 + 2\sqrt{I_1I_2}}{I_1 + I_2 + 2\sqrt{I_1I_2} + I_1 + I_2 - 2\sqrt{I_1I_2}} = 2\frac{\sqrt{I_1I_2}}{I_1 + I_2}$$
 1.19



Rysunek 1.9. a) Gdy dwa fazory o jednakowej długości *A* sumujemy ze sobą, to długość fazora wypadkowego jest równa zeru gdy są one przeciwnie skierowane. Gdy fazory te są zgodnie skierowane, wtedy fazor wypadkowy ma największą długość równą 4A²; b) gdy jeden z fazorów skrócimy, na przykład o połowę, to przy ich przeciwnym skierowaniu również otrzymamy najmniejszą długość fazora wypadkowego. Jednak długość ta będzie większa od zera i równa ¹⁄₄ A². Gdy fazory te są w fazie długość fazora wypadkowego wynosi 9/4 A² i jest prawie dwukrotnie mniejsza od odpowiedniej wartości otrzymanej gdy długość obu fazorów wynosiła A.

Jak można było się spodziewać, kontrast zależy tylko od wzajemnych stosunków natężeń dwóch interferujących fal i jest największy gdy natężenia te są sobie równe.

Zadanie (1.1) opiera się na szczególnej konfiguracji interferujących fal. Po nauce związanej z jego rozwiązaniem, rozwiązanie bardziej ogólnego przykładu nie powinno sprawić nam kłopotu: Niech obie fale będą pochylone względem ekranu. Nasza zadanie brzmi tak:

Zadanie 1.2.

Na kliszy fotograficznej, leżącej w płaszczyźnie *z*=0 rejestrujemy interferencje dwóch fal płaskich, o tej samej długości fali $\lambda = 632,8$ nm. Wektor falowy fali "niebieskiej" ma współrzędne $\mathbf{k}_{N}(k_{Nx},k_{Ny},?)$, gdzie $k=2\pi/\lambda$. Wektor falowy fali "czerwonej" ma współrzędne $\mathbf{k}_{C}(k_{Cx},k_{Cy},?)$. Amplitudy tych fal wynoszą A_{C} i A_{N} . Załóżmy, że w początku układu współrzędnych fazy obu fal wynoszą: φ_{pN} dla fali niebieskiej i φ_{pC} dla fali czerwonej. Oblicz wynik nakładania się tych fal.

Możesz sam rozwiązać to zadanie krok po kroku, byłoby to nawet zalecane. Ja ograniczę się do podania rozwiązania

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos\left(x\delta k_x + y\delta k_y + \delta\varphi_0\right)$$
1.20

$$\delta k_x = \pm (k_{Cx} - k_{Nx}) \tag{1.20a}$$

$$\delta k_y = \pm \left(k_{Cy} - k_{Ny} \right) \tag{1.20b}$$

$$\delta\varphi_0 = \pm(\varphi_{0C} - \varphi_{0N}) \tag{1.20c}$$

Ponieważ funkcja cosinus jest parzysta nie ma znacznie czy we wzorach (1.20 ac) wybierzemy znak plus czy minus. Ponownie otrzymaliśmy wzór na prążki interferencyjne. Wzór ten pokazuje, że gęstość i orientacja prążków zależy od względnego nachylenia fal a położenie jasnych lub ciemnych centrów prążków zależy od różnicy faz interferujących fal w początku układu współrzędnych (jest to warunek początkowy).

Zauważ jeszcze jedną rzecz. Gdy obserwujemy obraz interferencyjny dwóch fal płaskich to jest on stabilny w czasie, pomimo, że fazory wirują z olbrzymią częstością. Rysunek (1.10) pokazuje dlaczego tak się dzieje. Wszystkie fazory wirują z taką samą częstością. Oznacza to, że względny kąt pomiędzy fazorami jest taki sam, a co za tym idzie długość fazora wypadkowego nie zmienia się w czasie, a co za tym idzie obraz interferencyjny jest niezmienny w czasie. Tak jest wtedy gdy fazory wirują z tą samą częstością. A jak częstość wirowania fazorów obu fal jest inna? To wtedy obraz nie jest stały i długość fazora wypadkowego zmienia się w czasie; w efekcie obraz interferencyjny oscyluje, tyle że te oscylacje zachodzą z mniejszą częstością niż częstość wirowania fazorów. Znamy już ten efekt z analizy dudnień (§TVIII 2.1).



Rysunek 1.10. Dwie fale płaskie, których powierzchnie falowe narysowane są na niebiesko i czerwono padają na ekran pod pewnym niezerowym kątem. W początku układu współrzędnych, w wybranej chwili t=0, fazy tych fal są odpowiednio $\pi/3$ i π . Z prawej strony rysunku mamy przedstawione: sumy fazorów (na czarno), fazor fali pierwszej (na niebiesko) i fali drugiej (na czerwono), w trzech różnych punktach na ekranie, dla trzech różnych czasów obserwacji. Tam, gdzie fala niebieska wyprzedza oś y, trzeba się cofnąć od powierzchni falowej do ekranu. Zatem idac w góre, wzdłuż ekranu, wirujace wektory tej fali obracaja sie zgodnie z ruchem wskazówek zegara (ujemne katy obrotu) a wirujące wektory fali czerwonej obracają się przeciwnie. Dla tego samego przesunięcia wzdłuż osi y wektor fali niebieskiej obraca się wolniej od wektora fali czerwonej. Jest tak dlatego, że fala czerwona jest bardziej pochylona niż fala niebieska. Widać to również przez porównanie długości odcinków d1 i d2. Dokonując obserwacji w kolejnych chwilach widzimy, że wirujące wektory obu fal obracają się z tą samą prędkością i w tę samą stronę. Zatem w każdej chwili czasu ich suma jest wektorem o tej samej długości, choć innym kącie fazowym. Mierzymy jednak energię fali, która jest kwadratem długości wektora wypadkowego. Zatem w każdej chwili czasu, dla danego punktu ekranu pomiar daje ten sam wynik.

1.1. Interferometr Michelsona

Do tej pory teoretyzowaliśmy. Pozostaje odpowiedzieć na pytanie: jak uzyskać obraz interferencyjny w układzie doświadczalnym? W dobie laserów jest to proste. Mając element światłodzielący, kolimator, laser, dwa zwierciadła możemy dowoli napatrzeć się na interferencję dwóch fal płaskich. Element światłodzielący częściowo przepuszcza światło a częściowo odbija. Wystarczy całość zmontować w układ interferometru Michelsona; najprostszego

i najszerzej stosowanego rodzaju interferometru, czyli urządzenia do "robienia" interferencji (rys. 1.1.1).



Rysunek 1.1.1. a) układ optyczny interferometru Michelsona składa się ze źródła światła; Kostka światłodzieląca *KS* dzieli padającą falę płaską na dwie fale płaskie. Jedna jest oznaczona na czerwono, a druga na zielono. Zwierciadła *Z1, Z2* zawracają obie wiązki do kostki; b) pokazuje tor wiązki rysowanej na czerwono od momentu jej wejścia na kostkę *KS*. Wiązka odbija się od warstwy granicznej kostki w kierunku zwierciadła *Z2,* od którego odbija się z powrotem w kierunku kostki światłodzielącej *KS* (linia przerywana). Na kostce ponownie się rozszczepia (linia przerywana z kropkami) na dwie wiązki; jedna kieruje się do kamery (wiązka *I*), druga z powrotem w kierunku lasera (wiązka II); c) bieg wiązki rysowanej linią zieloną. Tutaj przebieg jest podobny jak w przypadku pokazanym na rysunku (b), z tym, że pracuje zwierciadło *Z1*. W płaszczyźnie kamery interferują wiązki oznaczone *I* dla fali rysowanej na czerwono i zielono; d) Aby uzyskać obraz prążków interferencyjnych należy pochylić jedno ze zwierciadeł. W efekcie powierzchnie interferujących fal nie będą do siebie równoległe

Jak działa interferometr Michelsona? Źródłem światła jest zwykle laser (dlaczego laser wyjaśnię później). Wiązka po wyjściu z lasera trafia do kolimatora, który jest urządzeniem do wytwarzania fali płaskiej. Następnie pada na element światłodzielący; w naszym przypadku jest to kostka światłodzieląca. Element światłodzielący część wiązki przepuszcza a część odbija. Powiedzmy, że nasza kostka światłodzieląca robi to w stosunku 1:1, to znaczy połowa mocy wiązki jest odbijana a połowa przepuszczana. Wiązka odbita biegnie do zwierciadła, gdzie w wyniku odbicia zawraca i ponownie wchodzi do kostki światłodzielącej; tam ponownie ulega odbiciu i przejściu w stosunku 1:1. Wiązka światła, która przeszła odbija się od drugiego zwierciadła i również wraca do kostki, gdzie jej połowa przechodzi, a połowa się odbija. W płaszczyźnie detektora spotykają się dwie fale; spotykają i interferują tworząc obrazy takie, jak te pokazane na rysunku (1.1.2). Rysunek (1.1.2) nie przedstawia symulacji numerycznej, tylko wyniki doświadczenia. Jednak aby otrzymać obraz prążków jedno zwierciadło musi być pochylone rysunek (1.1.1d). Im większe pochylenie zwierciadła, tym większy kąt pomiędzy kierunkami rozchodzenia się fal spotykających się w płaszczyźnie detektora.

Zatrzymam się nad przypadkiem, gdy zwierciadła *Z1* i *Z2* są idealnie prostopadłe do kierunku padających fal. Wtedy fale, które spotykają się w płaszczyźnie kamery są wzajemnie do siebie równoległe. We wzorze (1.1.20) kładziemy

$$\delta k_x = 0 \text{ i } \delta k_y = 0 \qquad \qquad 1.1.1$$

Przy czym oś z zorientowana jest prostopadle do płaszczyzny detektora. Jeżeli fale spotykają się w przeciwfazie to rejestrowane natężenie światła musi być równe zeru. Jeżeli w fazie, to otrzymamy maksymalne natężenie światła na detektorze. W pośrednich przypadkach natężenie to będzie pośrednie. To jaki przypadek otrzymamy zależy od różnicy dróg jaką przebyły obie fale po podziale na kostce światłodzielącej. Jeżeli obie drogi są idealnie równe, lub różnią się o wielokrotność λ , to fale spotykają się w fazie jeżeli różnią się o nieparzystą wielokrotność $\lambda/2$, to spotykają się w przeciwfazie i w wyniku destruktywnej interferencji wygaszają się. Proste prawda? Ale wymaga precyzyjnego sterowania zwierciadłem. Ponieważ światło biegnie tam i z powrotem to dla uzyskania dodatkowego przesunięcia o $\lambda/2$ zwierciadło należy przesunąć o $\lambda/4$. Dla światła czerwonego o długości fali 600nm daje to zaledwie 150nm. Obserwując ruch prążków możemy z bardzo wysoką precyzją mierzyć przesuw zwierciadła. Precyzja interferometru otworzyła mu drogę do wielu zastosowań w nauce i technice.

A co z zasadą zachowania energii? Jak wiązki się wygaszą, to zniknie niesiona przez nie energia! Nic takiego się nie dzieje. Zapomnieliśmy o tym, że część światła przechodzi z powrotem w kierunku lasera. Zasada zachowania energii jest spełniona, gdyż jak światło w płaszczyźnie detektora interferuje w przeciwfazie, to światło w drugiej gałęzi interferuje w fazie. Skąd to wiemy? Wiemy to po pierwsze z doświadczenia, po drugie na mocy zasady zachowania energii (coś takiego musi się zdarzyć inaczej zasada ta zostaje złamana), a po trzecie z głębszej teorii – elektrodynamiki klasycznej, a po czwarte z jeszcze głębszej teorii elektrodynamiki kwantowej. Zanim dotrzemy do tych głębszych teorii musi nam wystarczyć wnioskowanie z zasady zachowania energii i doświadczenia. Natura jest tak urządzona, że jak uda nam się doprowadzić do interferencji destruktywnej, to gdzieś musi być druga gałąź układu, w której następuje interferencja konstruktywna, tak że bilans energii domyka się.

Aby uzyskać obraz prążkowy jedno ze zwierciadeł musimy pochylić (rys. 1.1.1d). Wtedy, w płaszczyźnie kamery fale nakładają się pod kątem. Sprawę ilustruje rysunek (1.1.2).



Rvsunek 1.1.2. Obraz prażków interferencyjnych dwóch fal płaskich. Jak widać prążki nie są idealnie proste, a rozkład natężeń jest również zaburzony. W układzie doświadczalnym nie sposób wytworzyć dwóch idealnych fal płaskich stałej 0 amplitudzie.

Porozmawiajmy jeszcze o kilku trudnościach technicznych. Co prawda koncentrujemy się na podstawowych zasadach optyki falowej, ale od czasu do czasu możemy kilka słów poświęcić technice, tak abyś wiedział, że życie proste nie jest. Jak to zwykle bywa przy analizie teoretycznej wzieliśmy pod uwage tylko kilka najważniejszych dla nas czynników. Na przykład takie fale płaskie. Skad je wziąć? Układ laser + kolimator nigdy nie wytworzy idealnych fal płaskich. Zawsze będą to nieco popsute fale płaskie. Nie bierzemy pod uwagę tego popsucia w analizie teoretycznej. Ba, nawet nie możemy tego zrobić, gdyż musielibyśmy znać dokładne charakterystyki lasera, użytych w kolimatorze soczewek, dokładność montażu całości instrumentarium, wady szkieł, z których wykonane są soczewki, własności powietrza, które wypełnia kolimator, etc. Również kostka światłodzieląca i zwierciadła psują nieco nasze fale. Powierzchnie kostek i zwierciadeł nie są idealnie płaskie, kostka nie ma idealnie równoległych ścian, itd. Kamera również dorzuca swoje trzy grosze do psucia wyniku doświadczenia. Krążą w niej ciemne prądy (szumy termiczne), matryca CCD nie jest idealnie jednorodna, charakterystyki poszczególnych komórek elementu CCD nie są idealnie liniowe, itd. W efekcie końcowym zamiast prażków takich jak są pokazane na rysunku (1.7) mamy takie jak na rysunku (1.1.2), który przedstawia obraz uzyskany w rzeczywistym interferometrze. W praktyce trzeba się sporo namęczyć, żeby zmniejszyć niechciane efekty do dopuszczalnego poziomu, lub wyeliminować ich wpływ inteligentnymi procedurami pomiarowymi i inteligentną analizą danych pomiarowych. Z drugiej strony takie wyniki jak te pokazane na rysunku (1.1.2) potwierdzają, że sformułowane przez nas reguły do opisu procesu nakładania się fal są bardzo dobre. Niedokładny jest model, bo nie uwzględnia tych wszystkich błędów układu pomiarowego, które opisałem wyżej.

W interferometrze spotykać się mogą różne fale, nie tylko płaskie. Rysunek (1.1.3) pokazuje przykładowe interferogramy idealne, czyli wygenerowane numerycznie dla fali płaskiej i kulistej. Analiza takich interferogramów rządzi się tymi samymi podstawowymi regułami co interferogramów dla fal płaskich. Tyle, że od strony technicznej analiza ta jest trudniejsza ze względu na bardziej skomplikowaną geometrią powierzchni falowych interferujących fal.



Rysunek 1.1.3. Wynik interferencji fal: płaskiej *A* i kulistej *B* (*A*+*B*), płaskiej *A* kulistej *C* (*A*+*C*) i kulistej *B* z kulistą *C* (*B*+*C*). Fala kulista *B* ma w płaszczyźnie interferencji mniejszy promień krzywizny niż fala kulista *C*; promienie krzywizn obu fal kulistych pokazuje górna część rysunku. Amplitudy obu fal kulistych w płaszczyźnie interferencji są takie same.

Na temat interferometrii napisano setki książek i podręczników oraz tysiące artykułów. W użyciu jest wiele różnego rodzaju interferometrów. Rysunek (1.1.4) pokazuje drugi pod względem popularności układ interferometru. Pod względem ilości zastosowań interferometr prawdopodobnie przegrywa tylko, z tak powszechnie stosowanymi przyrządami pomiarowymi jak linijka, waga, termometr, barometr, kątomierz, woltomierz czy amperomierz. Wielu użytkowników nie wie nawet, że używa do pomiarów interferometru. W dobie komputerów i mikro-optyki interferometr można ukryć w małym pudełku, a schowany w tym samym pudełku układ cyfrowy przeliczy uzyskane wyniki i w wygodnej formie wyświetli go na dołączonym do pudełka wyświetlaczu. Z interferometrami różnych rodzajów będziemy się jeszcze spotykać nie tylko na gruncie optyki ale również w tematach związanych z mechaniką kwantową.



Rysunek 1.1.4. Interferometr w układzie Macha-Zendera. Padająca fala płaska (choć może to być również fala o innej geometrii powierzchni falowej) dzielona jest na pierwszej kostce światłodzielącej KS na dwie wiązki. Wiązki te są następnie kierowane przez zwierciadła ZW na drugą kostkę KS, gdzie dochodzi do ich nałożenia. Interferencja obu fal rejestrowane jest na kamerze lub kliszy lub obserwowana na ekranie. Najważniejszą zaletą interferometru Macha-Zendera jest to, że badana próbka prześwietlona jest przez jedną z fal jednokrotnie (w interferometrze Michelsona światło przechodzi przez próbkę dwukrotnie). Największą wadą jest jego bardziej skomplikowana budowa i wiążąca się z tym mniejsza stabilności całej konstrukcji.

1.1.1. Skoki fazy

Wróćmy do interferencji dwóch fal płaskich o takich samych amplitudach. Jeżeli wydaje wam się, że nic ciekawego nie da się już na ten temat powiedzieć to jesteście w błędzie. Rysunek (1.1.5) obrazuje interferencję dwóch fal z użyciem fazorów i linii równej fazy. Z rysunku wynika, że wzdłuż prążka interferencyjnego nie zmienia się wartość ani amplitudy fazora wypadkowego ani względnej fazy między fazorami interferujących fal. Środki ciemnych prążków pola interferencyjnego fal o równej amplitudzie skrywają mroczną tajemnicę. Światło na tą tajemnicę rzuca rysunek (1.1.6). Gdy oba fazory mają równą wartość i dodają się w przeciwfazie, to gdy przekraczamy linię ciemnego prążka faza wypadkowej fali doznaje skoku fazy o wartość π . Co oznacza skok fazy o π , pokazuje rysunek (1.1.7). Jak widać zbliżając się do skoku natrafiamy na punkty o niejednoznacznie określonej wartości fazy. Jeżeli na dolnej płaszczyźnie na krawędzi faza fali wynosi α , to na górnej jest ona równa $\alpha+\pi$. Gdy zero nie występuje, tak jak przy interferencji dwóch fal o nierównych amplitudach, faza określona jest jednoznacznie.



Rysunek 1.1.5. a) Linie równej jednej z dwóch interferujących fal płaskich wyznaczone w płaszczyźnie detekcji. Różne kolory odpowiadają różnym wartościom fazy. Druga fala jest równoległa do płaszczyzny detekcji i jej faza jest wszędzie taka sama. Tam gdzie powtarzają się linie o tej samej wartości (narysowane tym samym kolorem), kąt i długość fazora wypadkowego mają tą samą wartość. b) obraz interferencyjny tych fal. Wzdłuż każdej linii równoległej do prążka fazory wypadkowe (narysowane na zielono) mają tą samą długość. Fazor niebieski odpowiada fali równoległej do ekranu, a czerwony fali nachylonej.

Sytuacja kiedy układ fizyczny "nie wie, jednoznacznie w jakim jest stanie jest dla nas trudna do zaakceptowania. Pomyśl, że masz kolegę, który ma jednocześnie numer buta 34 i 44. Nie chodzi tu o to, że numer buta kolegi znamy niedokładnie. Znamy go dokładnie, ale tak się składa, że kolega ma dwa rozmiary buta naraz (na obu stopach). Jak dobrać mu buty? Czy zatem podwójna wartość fazy nie powoduje jakichś kosmicznych problemów? Otóż nie, gdyż natura przesłoniła ten wstydliwy fakt zasłoną zerowego natężenia światła. Tak się zawsze składa, że tam gdzie faza "wariuje" natężenie światła spada do zera.

Uzyskanie miejsc skoku fazy dla dwóch fal nie jest proste. Amplitudy fal muszą być dokładnie takie same. W praktyce nie da się uzyskać ani dwóch fal o identycznych amplitudach, ani nawet jednej fali o dokładnie takiej samej amplitudzie na całej jej powierzchni falowej. Istnieje jednak prosta metoda na otrzymanie punktów, gdzie faza wariuje i to całkowicie. Rysunek (1.1.8) pokazuje rozkład wybranych linii równej fazy dla trzech interferujących fal *a,b,c*. Fala *a* jest równoległa do płaszczyzny obserwacji i ma wszędzie stałą fazę. Wyrysowane są ponadto linie równej fazy dla dwóch pozostałych fal. Istnieją takie punkty, że fazory tych trzech fal dodają się tworząc trójkąt. Mamy tam punkty zerowego natężenia fazy, a przy okazji są to punkty, w których faza nie jest dobrze określona.



Rysunek 1.1.6. Rysunek z lewej strony pokazuje sytuację w której fazor czerwony i niebieski mają takie same długości. Fazor niebieski obraca się względem czerwonego. Gdy jego faza różni się o π od fazy fazora czerwonego fazor wypadkowy jest wektorem zerowym. Tuż przed tym momentem fazor wypadkowy (zielony) ustawia się w kącie $\pi/2$. Gdy różnica faz wektora niebieskiego i czerwonego nieznacznie przekroczy π wektor zielony pojawia się z fazą $-3/2\pi$. Oznacza to, że w punkcie, w którym przekraczamy linię zerowej amplitudy faza doznaje skoku wartości o π . Obszar w pomarańczowym kole jest narysowany w powiększeniu u góry rysunku. Rysunek z prawej strony pokazuje tą samą sytuację, gdy wektor niebieski ma mniejszą długość od czerwonego. Wtedy kąt wektora zielonego zmienia się od wartości zero, gdy wektory niebieski i czerwony są równoległe do pewnej wartości maksymalnej, a następnie ponownie spada do zera i w sposób ciągły (bez nagłych skoków) przechodzi pod oś wyznaczoną przez wektor czerwony.



Rysunek 1.1.7. Poruszając się wzdłuż osi *x*-ów widzimy zmniejszającą się fazę φ fali płaskiej. W pewnym punkcie wartość fazy skacze o π , a potem ponownie opada w sposób ciągły. Taką nieciągłość nazywamy nieciągłością krawędziową.

Jak widać z rysunku (1.1.8) fazory dodają się do zera na dwa różne sposoby, tworząc regularną siatkę ciemnych punktów (rys. 1.1.9). Rysunek (1.1.10) pokazuje co się dzieje wokół punktu zerowego. Zbiegają się do niego wszystkie możliwe linie równej fazy. Centralny punkt kompletnie nie wie jaką ma fazę. Na szczęście jest całkowicie ciemny więc nie sposób dostrzec jak bardzo jest zakłopotany.



Rysunek 1.1.8. Wybrane linie równej fazy fali *b* i *c* – fala *a* jest równoległa do powierzchni rysunku. Ale *a*, *b*, *c* mają różne amplitudy, tak że dla kątów fazowych pokazanych na rysunku, w punktach przecięcia linii niebieskich i czerwonych fazory dodają się do zera tworząc regularną siatkę ciemnych punktów.



Rysunek 1.1.9. Z lewej przykład zarejestrowanej (w doświadczeniu) interferencji trzech fal płaskich; jasne punkty pokazują położenie wirów optycznych; z prawej wykreślona mapa linii równej fazy dla trzech interferujących fal (symulacja komputerowa). Widać, że linie równej fazy zbiegają się do punktów, przez w punktach tych faza nie jest jednoznacznie określona. W punktach tych natężenie fali jest równe zeru.

Takie rozkłady fazy nazywamy wirami optycznymi. Spowodowane jest to bardzo ciekawą geometrią frontu falowego w pobliżu punktu osobliwego, którą pokazuje rysunek (1.1.11). Opisane tu efekty nazywane są różnie. Mówimy o fazowych defektach, o nieciągłościach, singularnościach. Defekty grają w fizyce współczesnej pierwszoplanową rolę. Dość wspomnieć, że bez defektów nie byłoby półprzewodnikowej elektroniki. Od strony matematycznej defekty opisywane są metodami topologicznymi. W tej sytuacji musiałem skorzystać z okazji jaką daje temat interferencji fal do pierwszego spotkania z defektami.



Rysunek 1.1.10. a) suma fazorów reprezentujących trzy fale płaskie w dziesięciu wybranych punktach na ekranie. W środku (czerwone kółko) fazory dodają się do zera; b) wypadkowe fazory trzech fal dodanych na rysunku (a). W środku fazor wypadkowy jest wektorem zerowym. Zauważ, że gdy przekraczamy środek po liniach przerywanych to po obu stronach punktu zerowego mamy fazory obrócone o π .



Rysunek 1.1.11. a) geometria frontu falowego modelowego wiru optycznego. Wzdłuż osi z położone są punkty w których faza jest nieokreślona. Front falowy ma kształt helisy i nie ma w nim skoków między kolejnymi powierzchniami tak jak ma to miejsce dla fali płaskie (rys. TIX 1.1.9) czy sferycznej (rys. TIX 1.2.3); b) gdy określimy wartości fazy heliakalnego frontu falowego z rysunku (a), na płaszczyźnie prostopadłej do osi z, to do punktu na osi z będą się zbiegały linie równej fazy o wartości od 0 do 2π . Faza w tym punkcie będzie nieokreślona. Biała linia wyznacza punkt skoku fazy $0\rightarrow 2\pi$, charakterystyczny przy mapowaniu wartości kąta na linię prostą. W rzeczywistości żadnej tego typu nieciągłości nie ma (dla kąta $0=2\pi$). Nieciągłość jest tylko w punkcie centralnym. Wartości kąta podane są w radianach.

1.2. Płytka płasko-równoległa

No dobrze, jak już chcemy aby w naszym interferometrze pracowały fale płaskie, to jak możemy sprawdzić, czy fala wychodząca z kolimatora jest płaska? Podam prosty sposób na sprawdzenie czy fale sa płaskie czy nie. Weź płytkę płasko-równoległa i odbij od niej badaną falę. Fala odbita będzie złożeniem fali dobitej od górnej i dolnej powierzchni. Płytkę można ustawić tak, aby obie odbite wiązki nałożyły się. Jeżeli padająca fala jest płaska, to obrazem nałożonych wiazek będzie jednolicie oświetlona powierzchnia. Dlaczego tak się dzieje? Zgodnie z rysunkiem (1.2.1) płaska fala, która pada na powierzchnię płytki pod katem α ulega załamaniu i odbiciu. Kierunek fali załamanej wyznaczamy korzystając z prawa załamania (rys. TX 2.6), a fali odbitej na podstawie prawa odbicia (TX 2.3). Fala załamana ponownie trafia na granice między dwoma ośrodkami; tym razem jest to granica szkło-ośrodek zewnętrzny (ośrodek zewnętrzny to najczęściej powietrze). Na tej granicy ponownie dochodzi do odbicia i załamania fali. Części odbita musi pokonać granicę dwóch ośrodków i ponownie jest to granica szkło-ośrodek zewnętrzny. W efekcie otrzymujemy dwie odbite fale płaskie, których powierzchnie falowe są do siebie równoległe. Jeżeli fale te nałożą się na siebie to powstaje obraz interferencyjny. Obraz interferencyjny dwóch fal płaskich o wzajemnie do siebie równoległych powierzchniach falowych opisany jest wzorem (1.13). Pozostaje pytanie: Jaka

jest różnica faz pomiędzy tymi dwoma równoległymi do siebie falami płaskimi? Różnicę tą możemy szybko wyznaczyć, korzystając z rysunku (1.2.1). Niech Fbędzie wybranym punktem na ekranie. W punkcie tym spotykają się dwie fale o wzajemnie równoległych powierzchniach falowych.



Rysunek 1.2.1. Płaska fala padajaca płytkę płaskona równoległą ulega odbiciu od górnei cześci płytki (kolor czerwony) i od dolnej części (kolor niebieski). Inne części podzielonej fali np., tę przechodzaca przez dolna powierzchnię płytki pomijamy. Obie odbite fale interferują ze sobą.

Z rysunku widać, że odcinek EF ma taką samą długość jak odcinek CG. Zatem całą analizę możemy zakończyć na punktach *C* i *E*, niezależnie od położenia ekranu, gdyż relacje fazowe w punktach C i E pomiędzy obiema falami są takie same jak w punktach F i G. Fala wnikająca w płytkę musi przebyć drogę, od punktu A do punkt C, równą sumie odcinków AB+BC. Jeżeli chcemy uzyskać zmianę kąta fazora na tej drodze to musimy ją przemnożyć przez współczynnik załamania płytki. Dlaczego tak jest wyjaśniam zaraz na początku sekcji 2 tego tematu, teraz po prostu to przyjmijmy.

Jeżeli poprowadzimy prostopadłe do promienia niebieskiego w punkcie C to prostopadła ta będzie również prostopadła do promienia czerwonego i przetnie go w punkcie E. Zatem przez punkty E i C przechodzą powierzchnie falowe obu fal. Są to pierwsze takie punkty po wyjściu obu rozważanych promieni z płytki i w tych punktach będziemy porównywać różnicę faz między interferującymi falami. Do punktu E fala odbita w punkcie A doszła poprzez drogę AE pomnożoną przez współczynnik załamania ośrodka. Różnica przemnożonych przez współczynniki załamania dróg obu fal wynosi:

$$\delta = n_{\rm s} (\overline{\rm AB} + \overline{\rm BC}) - n_0 \overline{\rm AE}$$
 1.2.1

Z rysunku (1.2.1) widać, że

$$\overline{AB} = \overline{BC} = \frac{d}{\cos(\beta)}$$
 1.2.2a

$$\tan(\beta) = \frac{\overline{\text{AD}}}{d}$$
 1.2.2b

$$\sin(\alpha) = \frac{\overline{AE}}{2\overline{AD}}$$
 1.2.2c

Korzystając z tych wyrażeń wzór na różnicę dróg δ możemy zapisać

$$\delta = n_s \overline{AB} - n_o 2\overline{AD} \sin(\alpha) = 2n_s \frac{d}{\cos(\beta)} - 2n_o 2d \tan(\beta)\sin(\alpha) \qquad 1.2.3$$

W dalszych obliczeniach skorzystamy z prawa Snella $n_o \sin \alpha = n_s \sin \beta$ 1.2.4

Prawo Snella pozwala nam wyeliminować kąt α , z równania (1.2.3).

$$\delta = \frac{2n_s d}{\cos\beta} - n_o 2d \tan\beta \frac{n_s}{n_o} \sin\beta$$
 1.2.5

Po uporządkowaniu powyższego wyrażenia mamy

$$\delta = 2n_s d\left(\frac{1}{\cos\beta} - \tan\beta\sin\beta\right) = 2n_s d\left(\frac{1 - \sin^2\beta}{\cos\beta}\right)$$
 1.2.6

Ostatecznie otrzymujemy $\delta = 2n_s d \cos \beta$

Powyższy wzór wyraża różnicę, przemnożonych przez współczynnik załamania, dróg optycznych dla obu fal. Aby z niego skorzystać musimy znać różnicę faz pomiędzy obiema falami. W celu wyznaczenia różnicy faz trzeba przemnożyć obliczoną różnicę dróg przez wartość wektora falowego k; ale to nie wszystko. Na granicy ośrodek gęsty-ośrodek rzadki, a założymy teraz, że $n_o < n_s$, promień odbity ulega dodatkowemu skokowi fazy o π , który oczywiście należy uwzględnić. Skok fazy o π oznacza, że fala odbita w punkcie A jakby dodatkowo przebyła drogę $\lambda/2$ równoważną przesunięciu fazy o π . Ponieważ, we wzorze (1.2.1) wkład drogi dla tej fali odejmujemy, odejmiemy też efekt tego skoku. Stąd różnica fazy wynosi:

$$\delta\phi = k\,\delta - \pi = \frac{2\pi}{\lambda} 2n_s d\cos\beta - \pi = \frac{4\pi}{\lambda} n_s d\cos\beta - \pi \qquad 1.2.8$$

Gdy $\delta \varphi = m2\pi$ obie fale interferują konstruktywnie i w obrazie obserwujemy jasną, jednorodną plamę a gdy $\delta \varphi = (m + \frac{1}{2})2\pi$, obie fale wygaszają się i w obrazie obserwujemy minimum natężenia światła. Przy pośrednich wartościach różnicy fazy mamy pośrednie wartości natężenia światła.

Podsumowując stwierdzamy, że jeżeli na płytkę płasko-równoległą pada fala płaska, to w odbiciu obserwujemy interferencję dwóch fal: pierwsza powstaje w wyniku odbicia części fali padającej od górnej powierzchni płytki, druga od dolnej. Obie fale są do siebie równoległe i w efekcie obraz interferencyjny jest jednorodny. Natężenie światła w obrazie interferencyjnym zależy od różnicy faz pomiędzy obiema falami. Natężenie to otrzymujemy ze wzoru (1.13), przyjmując we wzorze (1.13), wyrażenie φ_0 , obliczoną ze wzoru (1.2.8) różnicę faz $\delta\varphi$.

Jeżeli natężenie światła w obrazie interferencyjnym uzyskanym po odbiciu dwóch fal od płytki płasko-równoległe nie jest jednorodne, to badana fala nie jest płaska. Na rysunku (1.2.2a) widać dwie przesunięte względem siebie fale płaskie. Weź dwie kartki papieru ustaw na płaskiej powierzchni jedną nad drugą i wzajemnie względem siebie przesuń. W obszarze nakładania się tych kartek nic się nie zmieni. Jeżeli fala nie jest płaska, to przesunięcie jednej fali względem drugiej zmienia wzajemne relacje fazowe. To oznacza, że obraz interferencyjny przestaje być jednorodny i pojawiają się prążki (rys. 1.2.2b).



Rysunek 1.2.2. a) dwie fale płaskie równoległe do siebie dodają się w taki sam sposób niezależnie od tego jak są przesunięte względem siebie w płaszczyźnie powierzchni falowej; b) dwie niepłaskie fale i nieprzesunięte dodają się tak jak fale płaskie, na ekranie zobaczymy jednolicie naświetloną płaszczyznę. Gdy fale te przesuniemy względem siebie ich powierzchnie falowe rozejdą się. Różnica odległości między punktami powierzchni falowych będzie się zmieniała od punktu do punktu. W efekcie zobaczymy prążki interferencyjne.

Zatem dysponując płytką płasko-równoległą możemy szybko sprawdzić jak płaska jest fala. Musimy być jednak pewni, że nasza płytka jest płasko równoległa! Rysunek (1.2.3) pokazuje co się dzieje gdy tak nie jest. Fala płaska padająca na "pogiętą" płytkę daje obraz prążkowy



Rysunek **1.2.3**. Górna powierzchnia płytki nie jest płaska. W efekcie dwa promienie (czarne linie) padające na płytkę w dwóch różnych punktach odbijaja górnej się od (czerwony) powierzchni dolnej i (niebieski) pod różnymi katami. Wychodzące z płytki promienie _ czerwone i niebieskie nie są już wzaiemnie równoległe, zatem reprezentowane przez nie fale odbite przestają bvć płaskie i obraz interferencyjny bardziej staje sie skomplikowany.

Aby test z użyciem płytki płasko-równoległej miał sens trzeba kupić odpowiednio dobrą płytkę. A to niestety sporo kosztuje. Oczywiście im większa płytka tym trudniej utrzymać wymaganą dokładność wykonania i tym większa

jest jej cena (przy płaskości lepszej niż 50nm i średnicy 5cm może to być powyżej 4000Euro).

To jeszcze nie koniec problemów. Jeżeli nawet obie powierzchnie płytki są płaskie, ale nie są wzajemnie równoległe to zamiast jednolicie oświetlonego pola zobaczymy na ekranie układ prążków. Dzieje się tak, ponieważ wzajemnie nierównoległe powierzchnie płaskie odbijają padającą falę płaską w różnych kierunkach. Chociaż fale odbite pozostają płaskie to ich kierunki rozchodzenia się nie są równoległe, co jak wiemy skutkuje w powstaniu układu prostych prążków. Dwie pochylone powierzchnie płytki tworzą klin optyczny (rys. 1.2.5), którym się teraz zajmę. Fale po odbiciu się od tych powierzchni są względem siebie nachylone i w obrazie interferencyjnym pojawiają się prążki.



Rysunek 1.2.5. Fala płaska pada na klin szklany o kącie γ . Fale odbita od pierwszej i od drugiej powierzchni są dalej falami płaskimi, ale ze względu na klinowatość powierzchnie płytki falowe tych fal sa wzajemnie pochylone. W efekcie na ekranie zaobserwujemy prążki interferencyjne, tak jak w przypadku interferometru Michelsona ze wzajemnie nachylonymi zwierciadłami.

W gruncie rzeczy płytka płasko równoległa działa jak interferometr Michelsona. Padająca fala ulega na jej powierzchni podziałowi na dwie, które następnie interferują. Można też uznać, że sytuacja jest odwrotna. To interferometr Michelsona działa jak płytka płasko-równoległa, która została podzielona. Jej pierwsza powierzchnia stała się zwierciadłem *Z1* a druga powierzchnia zwierciadłem *Z2*. Kostka światłodzieląca dba o podział wiązki świetlnej. Różnica faz wyrażona wzorem (1.2.8) zostaje wyrażona przez różnicą dróg dla obu gałęzi interferometru. Taka "podzielona" płytka płaskorównoległa" jest wygodna w użyciu gdyż, możemy łatwo umieścić w jednej z jej gałęzi jakiś przedmiot, lub zmienić różnicę faz δ (przesuwając jedno ze zwierciadeł), lub zmienić kąt między powierzchniami płytki pochylając jedno ze zwierciadeł; wtedy interferometr staje się odpowiednikiem klina optycznego.

1.2.1. Barwy interferencyjne

Wyniki przedstawione dla płytki płaskorównoległej pokazują, że dla danego kąta padania fale odbite, których długość spełnia warunek (1.2.8), przy $\delta \varphi = m2\pi$, ulegają wygaszeniu. Z efektami tego wygaszania każdy z nas się spotkał. Dobrym przykładem są barwy widziane na powierzchni bańki (rys. 1.2.6). Bańka mydlana może być traktowana jako cienka płytka (utworzona z warstwy mydlin) o zakrzywionej powierzchni. Zakrzywiona powierzchnia powoduje, zmianę orientacji wektora normalnego. Idąc wzdłuż powierzchni bańki natrafiamy na punkty, dla których warunek wygaszenia, przy danym kącie obserwacji, spełniony jest dla coraz to innej długości fali. Wygaszone fale nie docierają do obserwatora i zamiast światła białego widzimy układ wielobarwnych plam. Układ ten zmienia się wraz z kątem, pod którym patrzymy na dany fragment bańki. Podobny efekt widać na kałuży na której rozlana jest warstwa oleju lub benzyny (rys. 1.2.7). Barwy, które powstają na skutek interferencji destrukcyjnej w wiązce światła białego nazywamy barwami interferencyjnymi.



Rysunek 1.2.6. W świetle odbitym powierzchnia bańki wygląda jak pomalowana w barwne plamy. Barwy efektem wygaszania te sa interferencyjnego zachodzącego przy odbiciu światła od zewnętrznej warstwy błony i wewnętrznej mydlanej. Tam gdzie nie ma barw światło odbija się pod takim kątem, że nie dociera do aparatu W efekcie fotograficznego. przez bańkę wyraźnie widzimy znajdujące się w tle okno budynku.



Rysunek 1.2.7. Cienka warstwa oleju silnikowego rozlana na kałuży. Od górnej i dolnej powierzchni tej warstwy odbija się światło. Pod danym kątem obserwacji, dla pewnych barw następuje wygaszenia a dla innych wzmocnienie. W efekcie widzimy na powierzchni warstwy układ barwnych pasów.

1.3. Zastosowania techniczne

Płaska płytka może zostać napylona na szkło jako cienka warstwa o grubości mniejszej od jednego mikrometra. Przy odpowiednim doborze grubości takiej warstwy możemy uzyskać efekt wygaszania odbić, dla danej długości fali. Zwykle taką warstwę wykonuje się przy założeniu, że światło pada prostopadle do powierzchni, co ma miejsce w większości układów optycznych. Oczywiście warunek ten nigdy nie jest spełniony idealnie. Rysunek (1.3.1) pokazuje obraz płytki przed i po pokryciu cienką warstwą. Gdy układ optyczny musi pracować w świetle białym zamiast jednej warstwy kładzie się ich wiele. Każda następna warstwa pozwala na redukcję odbicia dla kolejnej długości fali. Gdy warstw jest kilka możemy zredukować odbicia dla kilku długości fali rozłożonych równomiernie w granicach widma widzialnego. Dla fal o długościach leżących pomiędzy tymi wybranymi redukcja odbić będzie gorsza. Niemniej ilość energii odbijanej dla tych pośrednich długości fal jest również wyraźnie obniżona.



Rysunek 1.3.1. Z lewej strony płytka przed pokryciem warstwą antyodblaskową; wyraźnie widać światło odbite. Z prawej ta sama płytka po naniesieniu warstwy antyodblaskowej.

Wydaje się, że mając do dyspozycji odpowiednie wzory w stosunkowo prosty sposób można zaprojektować nawet złożony układ cienkich warstw tłumiących odbicie w całym zakresie fal widzialnych. Praktyka nigdy nie jest taka prosta. Pozostaje kwestia znalezienia materiałów o współczynnikach załamania zgodnych z obliczonymi. Materiały te muszą spełniać kilka dodatkowych warunków. Muszą trwale przylegać do podłoża, lub do poprzedniej warstwy, muszą dać się nakładać, jako jednorodna warstwa przy użyciu znanych technik, muszą być odporne na warunki pracy; na przykład: na warunki atmosferyczne, na ścieranie. Stosowanie zbyt wielu różnych materiałów również nie jest wskazane. Każdy z nich nakłada się w osobnej komorze próżniowej o wysokiej czystości. Zastosowanie jednej komory jest praktycznie niemożliwe, gdyż oczyszczenie jej z poprzednio użytego materiału jest kosztowne i czasochłonne. Zwykle cienkie warstwy wykonuje sie naprzemiennie z dwóch rodzajów materiałów.

1.3.1. Fotografia Lippmanna

Pod koniec XIX wieku Gabriel Lippmann zaproponował układ w którym zjawisko interferencji fal wykorzystano do rejestracji barwnych fotografii na

monochromatycznych emulsjach fotograficznych². Schemat układu pokazany jest na rysunku (1.3.2). Obraz tworzony jest w płaszczyźnie płyty fotograficznej przez klasyczny układ optyczny. Cała tajemnica leży po stronie konstrukcji płyty fotograficznej. Składa się ona z grubej warstwy emulsji (kilka µm), pod która znajduje się rtęciowe zwierciadło. Padająca fala świetlna odbija się od zwierciadła a fala odbita tworzy z fala padającą falę stojącą (§TIX 3.1). Jak wiemy fala stojąca ma punkty (węzły fali stojącej), w których natężenie światła jest minimalne (wynosi zero lub jest bliskie zeru) oraz miejsca gdzie amplituda oscylacji natężenia światła przyjmuje wartość maksymalną (strzałki fali stojacej). W punktach gdzie sa strzałki fali stojacych i w ich sasiedztwie dochodzi do naświetlenia emulsji, wyniku czego wytrąca się metaliczne srebro tworząc cienkie częściowo odbijające warstwy. Odległość między kolejnymi warstwami zależy od długości fali padającego światła. W procesie odtworzenia obrazu układ warstw działa jak filtr interferencyjny, który w kierunku oka przepuszcza światło o długości fali dla jakiej układ w danym obszarze został zarejestrowany.



Rysunek 1.3.2. Schemat układu do fotografii barwnej Lippmanna. Zagęszczenia kropek pokazują miejsca gdzie w objętości emulsji tworzą się strzałki fali stojącej. W tym miejscu naświetlona błona tworzy częściowo odbijające warstwy srebra. Układ takich warstw, poprzez zjawisko destruktywnej interferencji działa jak wysokiej jakości filtr barwny.

Proces fotografii Lippmanowskiej jest trudny technicznie. Wymaga przygotowania drobno ziarnistej emulsji (średnica ziarna jest rzędu 10nm) o wysokiej przezroczystości. Płyta fotograficzna musi być uzupełniona o rtęciowe zwierciadło. W późniejszym okresie rtęć została zastąpiona warstwą powietrza, dalej jednak płyta miała złożoną budowę. Samo naświetlanie grubej emulsji trwa stosunkowo długo (rzędu minuty przy jasnym oświetleniu).

² P. Ranson, R Ouillon, J.P Pinan-Lucarre, W stulecie Nagrody Nobla z fizyki za rok 1908 przyznanej Gabrielowi Lippmannowi za fotografię barwną, Postępy Fizyki, **60**, 121-124 (2009)



Rysunek 1.3.3. Gabriel Jonas Lippmann 16.08.1845 Bonnevoie (ur. w w Luksemburgu, zm. 13.07.1921) - fizyk francuski znany głównie z badań nad zjawiskiem piezoelektrycznym oraz reiestracia obrazów barwnych. Za techniki uzyskiwania opracowanie barwnej fotografii otrzymał w w 1908 roku nagrodę Nobla z dziedziny fizyki; źródło Wikipedia.

Przy oglądaniu zdjęć płyta musi być oświetlona pod odpowiednim kątem, inaczej odtwarzane kolory nie odpowiadają rzeczywistym. Przyczyna tego jest taka sama jak przyczyna zmian układy kolorów na plamie oleju, przy zmianie kąta patrzenia lub oświetlenia. Skutkiem tego barwna fotografia Lipmmanna została wyparta przez inne rozwiązania bazujące na trójchromatycznych materiałach światłoczułych. Metoda Lipmmanna "ożyła" dzięki holografii. Modyfikacja techniki Lippmannowskiej pozwoliła Denisyukowi na opracowanie metody rejestracji hologramów barwnych (tzw. hologramy objętościowe).

Z drugiej strony fotografie wykonane techniką zaproponowaną przez Lippmanna są bardzo trwałe i dają wrażenie żywych barw, co jest charakterystyczne dla barw uzyskanych metodą interferencyjną. Rysunek (1.3.4) pokazuje przykłady fotografii wykonanych techniką Lippmanna.





Rysunek 1.3.4. Z lewej – zdjęcie wypchanej papugi zrobione przez Lippmanna w 1899 roku; z prawej inny przykład fotografii wykonanej przez Lippmanna; źródło Wikipedia

1.4. Detektor story

Zakończę ten rozdział ważną uwagą. W temacie poświęconym drganiom i falom używałem zapisu zespolonego, który ułatwia obliczenia. Na końcu cyklu obliczeniowego używałem funkcji Re, by przejść do wielkości rzeczywistych. W optyce nie jest to konieczne. Związane jest to z ogromną częstością fal świetlnych, rzędu stu tysięcy miliardów cykli na sekundę. Detektory, którymi dysponujemy nie potrafią działać tak szybko.

Rysunek (1.4.1) przedstawia model prostego detektora. Gdy nadlatująca kulka trafia w wahadło, wahadło wychyla się. Gdy wahadło przecina bramkę następuje zliczenie pojedynczego impulsu. Następnie wahadło wraca do położenia równowagi. Hamulec powoduje, że wahadło nie waha się tam i z powrotem ale szybko zatrzymuje się położeniu równowagi. Po uderzeniu następnej kulki cykl powtarza się. Wszystko działa dobrze dopóki kulki nie nadlatują zbyt często. Gdy następna kulka uderza w momencie, gdy wahadło jest w obszarze bramki nie jest ono w stanie wrócić do położenia równowagi (rys. 1.4.2).



Rysunek 1.4.1. Schemat detektora do liczenia nadlatujących kulek.



Rysunek 1.4.2. Gdy czas między uderzeniem kolejnych kulek jest zbyt mały detektor nie zdąży powrócić do stanu, podstawowego. Pomiar ilości kulek przestanie być możliwy

Następna kulka nie powoduje następnego klik i sprawy przybierają zły obrót. Pomiary są zafałszowane. W zależności od konstrukcji detektor może w takie sytuacji milczeć, lub zacząć ciągle buczeć. W tej drugiej sytuacji wiemy przynajmniej, że detektor jest przeładowany.

Ten prosty model ilustruje problem wszystkich detektorów, rejestrujących zjawiska w czasie. Każdy z nich ma swój charakterystyczny czas, nazywanym czasem relaksacji, po którym może dokonać pomiaru w następnym okienku czasowym. Jeżeli proces jest za szybki detektor nie jest w stanie go śledzić i zaczyna działać wadliwie. Fala świetlna oscyluje tak szybko, że nie ma detektorów zdolnych do rejestracji jej przebiegu w czasie. Jedyne co możemy zrobić to mierzyć średnią energię fali. Dlatego w optyce zwyczajowo stosujemy zapis zespolony i gdy fale są monochromatyczne zapominamy o Re. Jednak w radiotechnice, gdzie fale oscylują dużo wolniej nasze detektory nadążają za przebiegiem fali i możemy te przebiegi mierzyć. Wtedy po obliczeniu interferencji fal, w zapisie zespolonym, musimy użyć funkcji Re – nie wolno obliczać kwadratu modułu, gdyż niszczy to informację o przebiegu czasowym, chyba że nam na nim nie zależy. Dopiero po użyciu funkcji Re, otrzymany wynik, podnosimy do kwadratu.

Pokażę teraz, że przy drganiach szybszych od czasu relaksacji detektora wolno nam opuszczać funkcję Re, przy obliczaniu interferencji fal świetlnych. W tym celu dodam, w wybranym punkcie dwie fale

$$u_1(t) = A_1 \cos(\omega t + \theta_1)$$
 1.4.1a

$$u_2(t) = A_2 \cos(\omega t + \theta_2)$$
 1.4.1b

$$u(t) = u_1(t) + u_2(t) = A_1 \cos(\omega t + \theta_1) + A_2 \cos(\omega t + \theta_2)$$
1.4.1c

Wyrazy θ_1 i θ_2 reprezentują część przestrzenną rozkładu fali. Na przykład dla fali płaskiej (TIX 1.1.2) mamy

$$\theta_1 = -\mathbf{k} \cdot \mathbf{r_1} + \delta_1 \tag{1.4.2a}$$

$$\theta_2 = -\mathbf{k} \cdot \mathbf{r_2} + \delta_2 \tag{1.4.2b}$$

Natężenie wyrazi się wzorem

$$i = u^{2}(t) = (A_{1} \cos(\omega t + \theta_{1}) + A_{2} \cos(\omega t + \theta_{2}))^{2} = A_{1}^{2} \cos^{2}(\omega t + \theta_{1}) + A_{2}^{2} \cos^{2}(\omega t + \theta_{2}) + 2A_{1}A_{2} 2\cos(\omega t + \theta_{1}) + A_{2}^{2} \cos(\omega t + \theta_{2}) + 2A_{1}A_{2} 2\cos(\omega t + \theta_{2})$$

$$1.4.3$$

Detektor nie rejestruje tak szybkich przebiegów. Rejestruje natomiast średnią energię po czasie. Obliczę więc średnią po czasie z wyrażenia (1.4.3).

$$\langle i^{2} \rangle_{T} = \frac{1}{T} \int_{0}^{T} \left(A_{1}^{2} \cos^{2}(\omega t + \theta_{1}) + A_{2}^{2} \cos^{2}(\omega t + \theta_{2}) + 2A_{1}A_{2} 2\cos(\omega t + \theta_{1})\cos(\omega t + \theta_{2}) \right)^{2} dt$$
1.4.4

. . .

Przez nawias ostry > oznaczyłem wartość średnią, literka *T* wskazuje, że średnią liczmy po czasie jednego okresu. Dzięki liniowości całki możemy osobno liczyć średnie po wszystkich składnikach sumy. Wiemy już, że (TVIII_5.2.15)

$$\langle \cos^2(\omega t + \varphi) \rangle_T = \frac{1}{T} \int_0^T \cos^2(\omega t + \varphi) dt = \frac{1}{2T}$$
 1.4.5

Iloczyn cosinusów w (1.4.4) mogę rozpisać do postaci

 $\cos(\omega t + \theta_1)\cos(\omega t + \theta_2) =$ $(\cos(\omega t)\sin(\theta_1) + \cos(\theta_1)\sin(\omega t))(\cos(\omega t)\sin(\theta_2) +$ $\cos(\theta_2)\sin(\omega t)) = \cos^2(\omega t)\sin(\theta_1)\sin(\theta_2) +$ $\sin^2(\omega t)\cos(\theta_1)\cos(\theta_2) - \cos(\omega t)\sin(\omega t)\cos(\theta_1)\cos(\theta_2) +$ $\cos(\omega t)\sin(\omega t)\sin(\theta_1)\sin(\theta_1)$ 1.4.6

Człony mieszane uśredniają się do zera

$$\langle \sin(\omega t)\cos(\omega t)\rangle_T = \frac{1}{T} \int_0^T \sin(\omega t)\cos(\omega t) dt = 0$$
 1.4.7

Zbierając wszystkie wyrażenia mamy

$$\langle i^2 \rangle_T = \frac{1}{2T} [A_1^2 + A_2^2 + 2A_1 \sin(\theta_1) \sin(\theta_2) + 2A_2 \cos(\theta_1) \cos(\theta_2)]$$
 1.4.8

Korzystając ze wzorów redukcyjnych dla funkcji trygonometrycznych mam

$$\langle i^2 \rangle_T = \frac{1}{2T} (A_1^2 + A_2^2) + 2A_1 A_2 \cos(\theta_1 - \theta_2)$$
 1.4.9

Policzę to samo zapisując fale płaskie w postaci zespolonej. Wzory na fale przyjmą postać

$$u_1(t) = A_1 \exp\{i(\omega t + \theta_1)\}$$
 1.4.10a

$$u_2(t) = A_2 \exp\{i(\omega t + \theta_2)\}$$
 1.4.10b

$$\begin{split} \mathbf{i} &= |\mathbf{u}(t)|^2 = A_1^2 + A_2^2 + A_1 A_2 \exp\{i(\omega t + \theta_1)\}\exp\{-i(\omega t + \theta_2)\} + A_1 A_2 \exp\{-i(\omega t + \theta_1)\}\exp\{i(\omega t + \theta_2)\} = A_1^2 + A_2^2 + \\ A_1 A_2 \exp\{i(\theta_1 - \theta_2)\} + A_1 A_2 \exp\{-i(\theta_1 - \theta_2)\} \end{split}$$
 1.4.11

Korzystając ze wzoru Eulera (DD 1.1.3) mamy

$$i = A_1^2 + A_2^2 + 2A_1A_2\cos(\theta_1 - \theta_2)$$
 1.4.12

Porównując wzory (1.4.9) i (1.4.12) widać, że otrzymaliśmy to samo, z dokładnością do stałego czynnika mnożącego 1/2T. Musisz przyznać, że zapis zespolony był przy tym bardziej efektywny (obliczenia były krótsze, co już

wiemy z (§TVIII 2.3)). W przypadku optyki zapis zespolony zwalnia nas nadto z kłopotliwego liczenia średnich po czasie. Powyższe można uogólnić, to znaczy, jeżeli chcemy obliczyć natężenie wypadkowe *N* nałożonych na siebie fal o tej samej częstości, to możemy to robić bezpośrednio w zapisie zespolonym, a im więcej fal tym prostsze są obliczenia w reprezentacji zespolonej w porównaniu z reprezentacją w dziedzinie rzeczywistej. Jaka jest cena tej prostoty? Cóż, obliczamy natężenie światła z dokładności do czynnika mnożącego. Ale w doświadczeniu zwykle nie mierzymy wartości natężenia światła, tylko względny rozkład tego natężenia. W tej sytuacji czynnik mnożący nie ma znaczenia.

1.5. Amplituda zespolona

Na zakończenie tej sekcji garść dodatkowych wyjaśnień. Napisany poniżej wzór jest najbardziej ogólnym wyrażeniem opisującym fale, jakim do tej pory dysponujemy.

 $\mathbf{u}(\mathbf{r},t) = a(\mathbf{r},t)e^{i\phi(r,t)}$

1.5.1

1.5.2

W optyce wielkość opisaną wzorem nazywamy amplitudą zespoloną

Definicja 1.5.1: Amplituda zespolona

Wyrażenie w postaci (1.5.1) nazywamy amplitudą zespoloną, przy czym funkcja a opisuje amplitudę fali a funkcja φ , fazę fali

Kwadrat modułu amplitudy zespolone fali reprezentuje energię tej fali w danym punkcie, które w optyce nazywamy natężeniem światła

 $i(\mathbf{r}, t) = |u(\mathbf{r}, t)|^2 = a^2(\mathbf{r}, t)$

W tym stwierdzeniu jest poważne uproszczenie. Poza punktami, gdzie natężenie światła jest równe zeru, wyrażenie (1.5.2) ma skończoną wartość. Jeżeli danemu punktowi przypiszemy skończoną wartość natężenia (energii), to gęstość tej energii będzie nieskończona i mamy czarną dziurę (rys. 1.5.1a). Jest oczywiste, że tak być nie może. Aby tego uniknąć uważamy, że kwadrat amplitudy zespolonej fali reprezentuje gęstość energii na małym kawałku powierzchni; w granicy jest to nieskończenie mały kawałek (rys. 1.5.1b). Wtedy energia na powierzchni całego kawałka jest równa

$$i = \int_{S} u^2(\mathbf{r}, t) \mathrm{d}s$$
 1.5.3

Gdzie S oznacza, że należy całkować po całym wybranym fragmencie powierzchni, a d*s* jest powierzchnią nieskończenie małego fragmentu powierzchni. Zatem, gdy mówimy natężenie światła w punkcie, to mamy na myśli energię fali w małym otoczeniu tego punktu, a wyrażenie u² reprezentuje
powierzchniową gęstość energii fali świetlnej. Aby uzyskać energię (natężenie) fali świetlnej, musimy tą gęstość pomnożyć przez powierzchnię.



Rysunek 1.5.1. Przypisanie każdemu punktowi powierzchni energii równej u² prowadzi do nieskończonych gęstości energii i nieskończonych energii zgromadzonych na skończonej powierzchni; b) dlatego przez u² rozumiemy gęstość energii obliczonej dla nieskończenie bardzo małego (w granicy nieskończenie małego) elementu powierzchniowego.

2. Zasada Huygensa w optyce 🔶

Z zasadą Huygensa (okr. TIX_2.1) spotkaliśmy się w temacie poświęconym falom. Huygens zastosował ją również do światła, postulując, że jest ono falą świetlną. Jak już wiemy, z poprzedniego tematu, jego propozycja okazała się przedwczesna. Falowa teoria światła zaczęła zyskiwać uznanie w XIXw, by w jego połowie stać się teorią dominującą. Okazało się przy tym, że w przypadku fal świetlnych zasada Huygensa musi zostać zmodyfikowana, by uwzględnić efekty interferencyjne. Omówienie optyki falowej z perspektywy zmodyfikowanej zasady Huygensa będzie przedmiotem tego rozdziału

W środowisku jednorodnym promienie świetlne są prostopadłe do powierzchni falowych. Choć optykę geometryczną zajmę się w następnym temacie, to pojęcie promienia jest również użyteczne w teorii falowej. Promienie świetlne wskazują kierunki propagacji energii świetlnej. Rysunek (2.1) pokazuje bieg promieni dla fali kulistej.



Rysunek 2.1. Dwie powierzchnie falowe fali kulistej wraz z przykładem promieni świetlnych

Co to jest środowisko izotropowe. Izotropia oznacza w tym wypadku tyle, że fala elektromagnetyczna propaguje się tak samo we wszystkich kierunkach.

Będziemy się obecnie zajmować wyłącznie rozchodzeniem się światła w środowisku optycznie izotropowym. Próżnia, powietrze, woda, nienaprężone szkło są optycznie izotropowe. Będziemy również analizować przejście światła przez granicę dwóch różnych izotropowych optycznie ośrodków. Rysunek (2.2) pokazuje cztery powierzchnie falowe fali płaskiej wyznaczające punkty, gdzie w ustalonej chwili czasu faza fali jest równa zeru. Czerwone linie pokazują trzy przykładowe promienie tej fali, które muszą być prostopadłe do powierzchni falowych.



Rysunek 2.2. Powierzchnie równej fazy fali płaskiej są zaznaczone liniami przerywanymi. Przykładowe promienie zaznaczone są liniami czerwonymi. Pokazane są również przykładowe fazory na promieniach.

W próżni światło rozchodzi się z prędkością oznaczaną literą c, równą około trzystu tysiącom kilometrów na sekundę. W ośrodkach materialnych światło zwalnia. Stosunek prędkości światła w próżni c, do prędkości światła w ośrodku materialnym v_m nazywamy współczynnikiem załamania tego ośrodka.

Definicja 2.1: współczynnik załamania

Współczynnik załamania n ośrodka materialnego określony jest przez stosunek prędkości światła w próżni c do prędkości światła v_m w tym ośrodku

$$n = \frac{c}{v_m}$$
2.1

Fakt 2.1

Współczynnik załamania jest wielkością bezwymiarową

Wiązka światła z rysunku (2.2) przechodzi przez ośrodek o współczynniku załamania *n* (na przykład blok szklany). Zakładamy, że po obu stronach tego ośrodka jest powietrze, gdzie możemy przyjąć współczynnik załamania jako równy jeden. Długość fali w ośrodku o współczynniku załamania *n* zmniejsza się *n* razy, a częstość pozostaje taka sama (w przyszłości to uzasadnię). Oznacza to, że wewnątrz ośrodka powierzchnie falowe są w odległości λ/n , a liczba falowa (wartość wektora falowego) wynosi k' = nk

Zasada Huygensa tłumaczy fakt uginania się wiązki światła na przeszkodzie. Rysunek (2.3) przedstawia efekt zastosowania zasady Huygensa dla fali przechodzącej przez dwa niewielkie otwory



Rysunek 2.3. a) przez dwa bardzo małe otwory przechodzą fale prawie kuliste (TIX 1.21). Różnica dróg od jednego z otworów do punktu P i do punktu Q jest niewielka. Stąd natężenie światła od każdej z fal na ekranie jest praktycznie stałe, a co za tym idzie stałe jest również natężenie światła od sumy fal (przy zaniedbaniu zjawiska interferencji); b) tak wygląda naświetlony ekran. W przykładzie źródła fal kulistych są od siebie oddalone o 2mm, odległości do ekranu wynosi *z*=200mm, rozmiary ekranu to 10x10mm.

Wiemy jednak, że dwie nakładające się fale interferują ze sobą, przez co obserwowany obraz, dla dwóch otworów jest znacznie bardziej złożony (rys. 2.4),



Rysunek 2.4. a) przy oświetleniu obu otworów tą samą falą (np. falą płaską), natężenie światła na ekranie jest modulowane przez zjawisko interferencji. Na rysunku przyjęto parametry takie jak na rysunku (2.3b), długości fali wynosi 630nm. Prążki są drobne więc wyrysowane są na małym fragmencie ekranu z rysunku (2.3b); b) rozkład natężenia światła na tym samym drobnym fragmencie ekranu pokazuj, że brak prążków na rysunku (2.3b) nie jest kwestią tego, że są one zbyt drobne.

Interferencja fal świetlnych zaczęła być na dobre badana w pierwszej połowie XIX wieku³. W efekcie tych badań francuski fizyk Augustin Fresnel

³ Wcześniej badał ją między innymi Izaak Newton, stąd mamy dziś pojęcie prążków Newtona. Newton objaśnił powstawanie prążków na bazie teorii promienia. Jednak to w wieku XIX interferencja stała się przedmiotem systematycznych badań co doprowadziło do przyjęcia falowej teorii w optyce. Teoria falowa okazała się zgrabniejsza i skuteczniejsza w opisie zjawiska dyfrakcji i interferencji.

poprawił zasadę Huygensa, tak aby można było w jej ramach ująć zjawisko interferencji. Obecnie poprawioną zasadę Huygensa nazywamy zasadą Fresnela-Huygensa.

Określenie 2.1: Zasada Huygensa-Fresnela

Każdy punkt ośrodka, do którego dotarła fala świetlna staje się źródłem fali wtórnej. Fale wtórne wygenerowane przez punkty do, którego dotarło czoło fali interferują ze sobą.

Na rysunku (2.5) fala jest emitowana z punktu źródłowego Z. W pewnej odległości od punktu źródłowego front falowy jest sferą. Dwuwymiarowy przekrój przez jedną powierzchnię równej fazy reprezentuje, na rysunku, czerwony okrąg. Aby wyznaczyć parametry fali (to znaczy jej amplitudę i fazę) w punkcie, na przykład P, musimy dodać do siebie wszystkie fale wtórne wyemitowane z tego frontu falowego. Fale wtórne muszą być "dociągnięte" do punktu obserwacji. Takie dociągnięcie zostało narysowane dla trzech wybranych punktów będących źródłem fali wtórnej. Odpowiednie fronty falowe (a raczej ich części) są narysowane na zielono. Ponieważ mamy do czynienia z interferencją, po dodaniu fal wtórnych, w każdym punkcie obserwacji, należy obliczyć kwadrat tej sumy. Kwadrat tej sumy jest równy natężeniu fali w punkcie obserwacji. Przy rysowaniu tego rysunku założyłem, że środowisko jest jednorodne, to znaczy, że nie ma zmian współczynnika załamania. Wtedy, na mocy (fakt TIX 2.1) możemy rysować fale wtórne jako fale kuliste. Pojawienie się zmian rozkładu współczynnika załamania komplikuje całą konstrukcję.



2.5. Ze źródła punktowego Rysunek Z rozchodzi się fala. Znamy geometrię frontu falowego (czerwona linia) w pewnei odległości od źródła. Aby na tej podstawie wyznaczyć amplitudę światła w punkcie P, z każdego punktu znanego frontu falowego rysujemy fala wtórna, która siega do punktu P. Na rysunku czarne okręgi symbolizują fale wtórne z wybranych punktów. Ze względu na przejrzystość rysunku tylko trzy spośród wtórnych wyrysowanych fal zostałv "dociągnięte" do punktu P (te trzy fale narysowane sa na zielono). Następnie wszystkie te fale sumujemy. Aby policzyć natężenie wyniki podnosimy do kwadratu.

Mając za podstawę zasadę Huygensa-Fresnela zmierzymy się z następującym zadaniem. Na otwór kołowy pada fala świetlna wyemitowana z punktu źródłowego. Jakie jest natężenie światła w wybranym punkcie ekranu? Rysunek (2.7) przynosi odpowiedź na to pytanie, zgodnie z zasadą HuygensaFresnela. Dopóki nie dotrzemy do przeszkody dopóty nie dzieje się nic ciekawego. Fala rozchodzi się jako fal kulista o rosnącym promieniu. Przy dojściu do przeszkody musimy uważać, dlatego narysowałem front falowy, który dotarł do przeszkody i z tego frontu falowego wyrysowałem fale wtórne. Z faktu, że część frontu falowego zostanie zatrzymana i nie da swoich fal wtórnych wynikają ciekawe efekty.



Rysunek 2.6. Augustin Jean Fresnel (ur. 10.05.1788, zm. 14.07.1827). Francuski fizyk i inżynier. Jeden z głównych twórców falowej teorii światła. W 1818 napisał traktat o dyfrakcji. Odkrył i wyjaśnił stan polaryzacji kołowej światła oraz opracował teorię dwójłomności kryształów. Jako inżynier zajmował się budową i remontami dróg i mostów.



Rysunek 2.7. Z lewej strony na otwór pada fala z punktowego źródła światła. Taka fala ma sferyczny front falowy (fala kulista). Powierzchnię frontu falowego dzielimy na gęstą sieć punktów. Z każdego punktu wypuszczamy wtórną falę kulistą (wzór na amplitudę zespoloną fali kulistej jest na rysunku, zobacz też (TIX. 1.2.1). W wybranym punkcie ekranu P fale wtórne sumujemy, a moduł sumy podnosimy do kwadratu. Podobnie postępujemy dla wszystkich innych punktów ekranu, dla których chcemy policzyć natężenie światła.

Możemy wykorzystać do obliczeń fazory. Każdy punkt frontu falowego fali padającej stanowi punktowe źródło fali wtórnej. Kulistą falę wtórną możemy reprezentować jako zbiór promieni wychodzących we wszystkie strony (rys. 2.1). Któryś z tych promieni trafi w punkt P. Na rysunku (2.7) oprócz wtórnych fal kulistych pokazane są promienie trafiające w punktu P. Musimy teraz policzyć jak, wzdłuż każdego z tych promieni, w ustalonej chwili czasu, zmieni się faza fali (lub faza odpowiedniego fazora). W punkcie P dodajmy fale wtórne z odpowiednimi fazami lub odpowiadające im fazory. Kwadrat sumy fal lub długości fazorów reprezentuje natężenie fali w punkcie P.

Rysunek (2.8) przedstawia zmodyfikowany schemat obliczeń z użyciem fazorów. Na otwór pada fala kulista, a my bierzemy pod uwagę front falowy, który jest styczny do powierzchni otworu, w jego centrum (bo jest to wygodne).



Rysunek 2.8. a) Front falowy narysowany na czarno jest styczny do powierzchni otworu w jego środku. W ustalonej chwili czasu amplitudę i fazę tej fali reprezentuje fazor czerwony. Przez zielony punkt, leżący w płaszczyźnie otworu, przechodzi front falowy narysowany na zielono (przerywana lina). Fazory na tym froncie (zielona strzałka rysowana linią przerywaną)mają inną fazę niż na froncie czarnym (zielona strzałka). Zielona linia kreska-kropka zaczynająca się w środku sfery czarnej (i zielonej) reprezentuje promień świetlny. Promień ten przebija sferę czarną, i zieloną. Odległość między punktami przebicia, liczona wzdłuż promienia, i przemnożona przez liczbę falową k daje nam różnicę kątów fazora na sferze czarnej i zielonej. Ta sama konstrukcja pokazana jest również na przykładzie punktu pomarańczowego, który przechodzi powierzchnia równej fazy narysowana przez na pomarańczowo; b) konstrukcja z części (a) pozwala wyznaczyć fazę fali padającej na otwór, w każdym punkcie tegoż otworu. Możemy teraz uznać, że każdy punkt otworu emituje wtórną falę kulistą, z tym że fazy początkowe tych fal wtórnych są różne dla różnych punktów (bo powierzchnia otworu nie pokrywa się z powierzchnią falową fali padającej). Przy każdym z wyróżnionych punktów napisane jest równanie wtórnej fali kulistej. Zauważ, że równania te różnią się początkowym przesunięciem fazowym δ . Z każdego źródła fali wtórnej wybieramy promień, który trafia w punkt P na ekranie. Dalej postępujemy tak jak w przykładzie pokazanym na rysunku (2.7).

W punkcie styczności fazor fali jest jednocześnie fazorem tej fali wyznaczonym w płaszczyźnie otworu. Niestety inne punkty otworu nie są styczne do powierzchni falowej. Możemy jednak obliczyć fazę fazora tej fali w innych punktach otworu prowadząc promienie prostopadłe do powierzchni falowej.

Każdy taki promień przebije jakiś punkt w płaszczyźnie otworu. Długość odcinka łączącego punkt na froncie falowym z punktem w otworze po przemnożeniu przez *k* pozwala na wyliczenie zmiany kąta fazora przy przejściu fali padającej do danego punktu otworu. Dalej postępujemy tak jak w przypadku pokazanym na rysunku (2.8). Z każdego wybranego punktu otworu ciągniemy promień trafiający w wybrany punkt P na ekranie. Obliczamy kąt fazora w punkcie P i sumujemy wszystkie fazory. Powtarzamy tą procedurę dla wszystkich interesujących nas punktów w płaszczyźnie ekranu. Zastosowanie tej procedury ma znaczenie techniczne. Łatwiej jest sumować rozkłady fal od powierzchni płaskiej niż sferycznej. Ponadto możemy naszą procedurę uogólnić na falę padającą o dowolnym froncie falowym.

Uogólnienie przedstawionej procedury na dowolną falę padającą na otwór nie jest trudne. Powiedzmy, że mamy znaną falę padającą, której amplituda zespolona (def. 1.5.1) dana jest funkcją $u_{pad}(\mathbf{r},t)$. Nie musi to być fala ani płaska ani kulista. Fala pada na otwór, a naszym zadaniem jest obliczenie rozkładu natężenia światła na ekranie. Jeżeli padająca fala jest znana, to możemy obliczyć rozkład jej fazy i amplitudy w płaszczyźnie otworu. Każdy punkt otworu stanie się źródłem kulistej fali wtórnej, która będzie dziedziczyć fazę i amplitudę po fali padającej. Na rysunku (2.9) w trzech wybranych punktach fala padająca reprezentowana jest przez fazory. Dalej postępujemy tak jak na rysunku (2.8b).



Rysunek 2.9. Niech na otwór pada znana fala opisana amplitudą zespoloną u_{pad}, tak że w każdym punkcie otworu możemy obliczyć fazę i długość fazora tej fali. Na rysunku wyrysowane są trzy przykładowe fazory w trzech punktach otworu. Wyznaczając odcinki łączące kolejne punkty otworu z punktem obserwacji P, możemy wyliczyć ułożenie fazorów w punkcie P. Następnie fazory te sumujemy, a z sumy obliczamy kwadrat modułu.

Sumowanie fazorów w punkcie *P* możemy wykonać w dowolnie wybranym układzie współrzędnych. Fazor wypadkowy wyraża się wzorem

$$\mathbf{v}_p = \left(\sum_{i=1}^N v_{xi}; \sum_{i=1}^N v_{yi}\right)$$
 2.1

Gdzie v_{xi} , v_{yi} to współrzędna x i y fazora dla *i*-tego promienia dochodzącego do punktu P. Poszczególne współrzędne wyrażą się wzorem

$$v_{xi} = A_i \cos(\varphi_i) \tag{2.1a}$$

$$v_{vi} = A_i \sin(\varphi_i) \tag{2.1.b}$$

W tym wzorze A_i oznacza długość (wartość) *i*-tego fazora, a φ_i fazę *i*-tego fazora. Kąt fazowy fazora jest równy sumie kąta początkowego φ_0 wyznaczonego w płaszczyźnie otworu i przyrostu kąta $\delta \varphi$ związanego z przesunięcia się z danego punktu otworu do punktu obserwacji *P*.

$$\varphi_i = \varphi_0 + \delta \varphi_i \tag{2.2}$$

Przyrost kąta fazora można wyznaczyć licząc ile razy w danym odcinku r_i mieści się długość fali światła λ ; wynik trzeba przeliczyć na radiany, czyli przemnożyć przez 2π .

$$\delta \varphi_i = \frac{2\pi}{\lambda} r_i = k r_i$$
2.3

gdzie

$$r_{i} = \sqrt{\left(x_{f} - x_{p}\right)^{2} + \left(y_{f} - y_{p}\right)^{2} + \left(z_{f} - z_{p}\right)^{2}}$$
 2.4

W tym miejscu widać, że warto było przejść z kulistego frontu falowego na płaski ekran (rys. 2.8). Spróbuj policzyć r_i dla punktów znajdujących się na sferze; oczywiście da się ale jest trudniej. Zbierając te wszystkie wzory możemy zapisać

$$\mathbf{v}_{p} = \left(\sum_{\substack{i=1\\ v_{x}}}^{N} A_{i} \cos(k r_{i} + \delta \varphi_{i}); \sum_{\substack{i=1\\ v_{y}}}^{N} A_{i} \sin(k r_{i} + \delta \varphi_{i})\right)$$
2.5

Natężenie światła liczmy jako kwadrat długości fazora wypadkowego

$$I_p = v_x^2 + v_y^2 \tag{2.6}$$

Ze wzoru (TIX 1.2.1) na falę kulistą wiemy, że jej amplituda zmienia się wraz z odległością. Efekt ten zaniedbujemy. W optyce zwykle interesuje nas względny rozkład natężeń światła na detektorze. Względny rozkład amplitud fazorów na froncie falowym i w punkcie P jest praktycznie taki sam, gdyż różnice dróg liczonych wzdłuż poszczególnych promienie są na tyle małe, że względne różnice amplitud tym spowodowane są zaniedbywalnie małe. Zauważ, że te same różnice długość są istotne z punktu widzenia zmiany fazy. Faza jest dużo, dużo czulsza na zmianę długości drogi niż amplituda.

2.1. Strefy Fresnela

Przedstawiona wyżej technika rachunkowa, dla twórcy całej koncepcji Fresnela, była mało użyteczna. Osiągnięcie sensownej dokładności wymaga przeliczenia tysięcy promieni, co bez komputera jest zadaniem zbyt czasochłonnym. Fresnelowi pozostało zatem przetłumaczyć swoją zasadę na język analitycznych wyrażeń matematycznych i obliczać pojawiającej się złożone wyrażenia całkowe, lub wspomóc się trikami pozwalającymi wywnioskować wyniki. Przedstawię jeden z takich trików, który pozwolił na wyciągnięcie zaskakujących, a przez to cennych wniosków.

Rysunek (2.1.1) przestawia schemat rozumowania Fresnela. Otwór kołowy oświetlony jest poosiową falą płaską. Wszystkie fazory reprezentujące tą falę w otworze są w tej same fazie. Chcemy obliczyć natężenie światła w punkcie P na osi symetrii układu. Tak się przy tym stało (oczywiście nie przypadkiem), że różnica w długości drogi między punktem P i środkiem otworu, a punktem P i brzegiem otworu wynosi $\lambda/2$. Można z tego wywnioskować, że fazor w punkcie P, obliczony wzdłuż promienia centralnego będzie przesunięty w fazie w stosunku do fazora brzegowego o π .



Rysunek 2.1.1. Na otwór pada fala płaska poosiowa o długości fali λ . W płaszczyźnie otworu wszystkie fazory mają tą samą fazę. Punkt *P* na osi *z* wybieramy tak, że różnica długości odcinka od punktu *P* do ekranu i od ekranu do krawędzi otworu wynosi $\lambda/2$.

Fazory z pośrednich punktów przesunięte będą o pośrednie kąty, a suma fazorów rozwinie się wzdłuż półokręgu (rys. 2.1.1). Zastanówmy się co się stanie gdy w sytuacji pokazanej na rysunku powyżej wstawimy większy otwór,

tak że różnica odległość punkt P - otwór oraz punkt P - krawędzi otworu wyniesie λ ?

$$r_0 - r_2 = 2\frac{\lambda}{2}$$
 2.1.1

Z rysunku (2.1.2) widać, że w takim przypadku suma fazorów rozwinie się w okrąg, a amplituda fali w punkcie P będzie równa zeru. Jest to ostatnia rzecz, jakiej byśmy się spodziewali, na osi układu oświetlonego, poosiową falą płaską natężenie światła jest równe zeru!



Rysunek 2.1.2. Na otwór pada fala płaska poosiowa o długości fali λ . W płaszczyźnie otworu wszystkie wirujące wektory mają tą samą fazę. Punkt *P* na osi *z* wybieramy tak, że różnica długości odcinka od *P* do ekranu i od ekranu do krawędzi otworu wynosi λ .

gdy znów powiększymy otwór tak, że (rys. 2.1.3)

$$r_0 - r_3 = 3\frac{\lambda}{2}$$
 2.1.3

Wysumowane fazory "przebiegną" półtora okręgu i natężenie światła w punkcie P będzie takie jak w przypadku pokazanym na rysunku (2.1.1).

Teraz trochę nazewnictwa. Wytyczyliśmy rozmiary otworu, w którym różnice długości między punktem P i środkiem otworu oraz punktem P i brzegiem otworu wynosiły $n\lambda/2$. Możemy pomyśleć o dużym otworze podzielonym na strefy, które dla danego punktu obserwacji spełniają taką zależność. Strefy te nazywamy strefami Fresnela (rys. 2.1.4). Promień *n*-tej strefy Fresnela wynosi

$$R_{n} = \sqrt{\left(z_{0} + n\frac{\lambda}{2}\right)^{2} - z_{o}^{2}} = \sqrt{z_{0}n\lambda + n^{2}\frac{\lambda^{2}}{4}}$$
2.1.4

Ponieważ, w zakresie optycznym, długości fal są bardzo, bardzo małe (rzędu dziesięciotysięcznych części milimetra) człon zawierający λ^2 jest również bardzo, bardzo mały w porównaniu z członem $z_0 n \lambda$ i zwykle się go pomija.

$$R_n \approx \sqrt{z_0 n \lambda}$$
 2.1.5



Rysunek 2.1.3. Gdy otwieramy otwór tak, że różnica odległość między punktem P i środkiem otworu oraz punktem P i brzegiem otworu wynosi $r_0 - r_3 = 3\frac{\lambda}{2}$, to natężenie światła urośnie do wartości takiej jak w przypadku gdy różnica ta wynosiła $\frac{1}{2} \lambda$.



Rysunek 2.1.4. Strefy Fresnela dla otworu kołowego oświetlonego falą płaską dla punktu obserwacji *P*. Czarnobiały pasek to widok z boku, kolorowa tarcza to widok od przodu.

Na podstawie dotychczasowych spostrzeżeń możemy stwierdzić, że gdy w otworze mieści się, dla danego punktu obserwacji, parzysta liczba stref Fresnela, to natężenie światła w tym punkcie jest bliskie zera, gdy stref jest nieparzysta liczba, natężenie światła osiąga pewną wartość maksymalną, gdy mieści się pełna liczba stref Fresnela oraz kawałek następnej strefy natężenie światła ma wartości pośrednie.

Gdy chcemy zwiększyć natężenie światła w punkcie P możemy zamalować wszystkie pierścienie, których górne granice wyznaczają strefy Fresnela o parzystym numerze (rys. 2.1.5). Wtedy światło tych stref nie interferuje destruktywnie ze światłem z pierścieni wyznaczonych przez nieparzyste strefy. Wkład dają tylko pierścienie nieparzyste. Tak zamalowaną płytkę nazywamy płytką strefową Feresnela.



Rysunek 2.1.5. Zamalowujemy co drugą strefę Fresnela. W efekcie pierścienie o parzystych numerach nie wygaszają światła z pierścieni o nieparzystych numerach i w punkcie *P* otrzymujemy wyraźnie większe natężenie światła. W naszym przykładzie mamy sumę fazorów z czterech nieparzystych pierścieni, co oznacza, że natężenie światła jest szesnaście razy większe, niż w przypadku otworu o rozmiarach jednej strefy Fresnela (rys. 2.1.1).

Zamalowanie połowy stref Fresnela oznacza dużą stratę światła – połowa otworu nie przepuszcza światła. Można temu zaradzić w sposób pokazany na rysunku (2.1.6). Wykorzystując fakt, że długość fali jest mniejsza *n* razy (*n*-współczynnik załamania szkła) w powietrzu niż w szkle (rys. 2.2) możemy podciąć płytkę, w każdym nieparzystym pierścieniu, w taki sposób, aby wszystkie fazory przez to podcięcie nie wykonały połowy obrotu. Wtedy fazory z parzystych pierścieni będą sumowały się tak jak fazory z nieparzystych pierścieni. Głębokość podcięcia możemy wyliczyć korzystając ze wzoru

d a	h (d h) m l	
$a \underline{n}$	$= \underbrace{n + (a - n)}_{n + \frac{1}{2}} n + \frac{1}{2}$	
droga	droga	
optyczna	optyczna	216
przez	przez	2.1.0
niepodciętą	podciętą	
<i>cz</i> ęść	<i>cz</i> ęść	
płytki	płytki	

Stąd możemy obliczyć głębokość podcięcia h.

$$h = \frac{\lambda}{2(n-1)} \tag{2.1.7}$$

Jeżeli długość fali jest rzędu λ =0,5µm, a współczynnik załamania wynosi około n=1,5, to h wynosi około $h \approx 0.5$ µm $\approx \lambda$. Płytka pokazana na rysunku (2.1.6) to jest również płytka strefowa Fresnela.



Rysunek 2.1.6. Wykorzystamy fakt, że w powietrzu długość fali jest większa niż w szkle. Jeżeli teraz odpowiednio podetniemy nieparzyste strefy Fresnela to wirujące wektory z tych podciętych stref zgubią pół obrotu. W efekcie będą wchodziły do sumy wektorów tak jak wirujące wektory ze stref nieparzystych. Amplituda światła w punkcie *P* wzrośnie siedmiokrotnie w porównaniu z otworem o tym samym promieniu, a natężenie światła wzrośnie czterdzieści dziewięć razy. Jaką głębokość *h* musi mieć to podcięcie? Taką, aby droga optyczna przez podciętą część płytki była $\lambda/2$ razy mniejsza od drogi przez niepodciętą część płytki.

Jeszcze lepiej byłoby, gdyby wszystkie wirujące wektory sumowały się z tą samą fazą. Wymagałoby to innej głębokości podcięcia w każdym punkcie płytki. Prosty sposób na zaprojektowanie takiego elementu przedstawię w następnym temacie.

Łatwo można pokazać, że przy przyjęciu przybliżenia (2.1.5) powierzchnie wszystkich stref Fresnela są takie same. Pole powierzchni koła o promieniu R_n wynosi

$$S_n = \pi R_n^2 = \pi \lambda z n \tag{2.1.8a}$$

Pole koła o promieniu R_{n+1} wynosi

$$S_{n+1} = \pi R_{n+1}^2 = \pi \lambda z (n+1)$$
 2.1.8b

Pole pierścienia zawartego między strefą R_n i R_{n+1} wynosi

$$S = S_{n+1} - S_n = \pi \lambda z (n+1) - \pi \lambda z n = \pi \lambda z$$

$$2.1.8$$

Jak widać pole to nie zależy od numery stref n. Oznacza to, że szerokości stref Fresnela maleje wraz odchodzenie od środka otworu. Rysunek (2.1.7) przedstawia przykładowe wykresy.



Rysunek 2.1.7. a) granica pierwszych czterdziestu stref Fresenela dla trzech różnych przypadków; b) rozmiar pierwszych czterdziestu stref Fresenela dla trzech różnych przypadków

Element optyczny, który koncentruje energię świetlną fali płaskiej w zadanym punkcie nazywamy soczewką. Opisana wyżej dwa rodzaje płytek strefowych Fresnela (rys. 2.1.5 i 2.1.6) pełnią rolę elementu ogniskującego, tak jak klasyczna soczewka. Płytka strefowa Fresnela należy do tzw. dyfrakcyjnych elementów optycznych. Jej podstawową zaletą jest płaskość, za którą idzie niska waga w porównaniu do klasycznych soczewek. Wykres z rysunku (1.1.7) pokazuje, że wykonanie wysokiej jakości soczewek Fresnela może być trudne technologicznie ze względu na bardzo małe rozmiary zewnętrznych stref. Pomimo trudności technologicznych, i wad w stosunku do soczewek klasycznych takie elementy są dziś wykonywane. Wymaga to jednak zaawansowanej technologii. Do sprawy powrócę w następnym temacie.

No dobrze, a skąd wiadomo, że to wszystko prawda? Teoria Fresnela, jak wszystkie nowe zaskakujące koncepcje nie miała łatwego dzieciństwa. Została uznana dzięki temu, że miała potężnego sojusznika – eksperyment. W czasach Fresnela przeprowadzenie odpowiedniego eksperymentu łatwe nie było. Do doświadczenia nie możemy użyć światła białego, gdyż położenie stref Fresnela zależy od długości fali (2.1.4). Zatem gdy dla jednej fali wypadnie minimum natężenia, w danym punkcie obserwacji, dla innej będzie to maksimum a dla jeszcze innych będziemy mieli wartości pośrednie. Fresnel musiał więc wyizolować ze światła słonecznego, składową o możliwie wąskim spektrum długości fal. Dla nas doświadczenia ze strefami Fresnela są proste, przez dostęp do tanich kamer i laserów.

W 1818 roku Fresnel zgłosił swoją teorię do konkursu na pracę nad naturą światła. Było to świeżo po pracach Thomasa Younga nad interferencją fal przechodzących przez dwie szczeliny, co otworzyło debatę pomiędzy zwolennikami teorii korpuskularnej a zwolennikami teorii falowej. Jeden z członków komitetu Simeon Poisson dokładnie przestudiował pracę Fresnela próbując dowieść, że jest błędna. Doszedł do wniosku, że jeżeli teoria Fresenla jest prawdziwa, to musi zachodzić następujący efekt. Wstawmy w bieg fali płaskiej krążek, tak jak to pokazuje rysunek (2.1.8a)



Rysunek 2.1.8. a) na nieprzezroczysty krążek pada równoległa wiązka światła. Zgodnie z optyką geometryczną za krążkiem powinien rozciągać się obszar cienia o średnicy równej średnicy krążka. Jednak w centrum tego obszaru obserwujemy jasną plamkę. Obserwacja tej plamki była experimentum crucis dla teorii falowej Fresnela; b) Zdjęcie pokazujące obraz przesłony kołowej oświetlonej poosiową wiązką równoległą (poosiową falą płaską). W centrum widać słabą plamkę Poissona (Arago). Na brzegach obszaru cienia widać charakterystyczne dla dyfrakcji prążki. Omówimy je poniżej (zobacz rys. 3.1.2-3); c) to samo zdjęcie co (b) ale z podniesionym kontrastem części centralnej.

Według obliczeń Poissona za krążkiem, w centrum cienia powinna pojawić się słaba jasna plamka, co zdawało się być absurdalne. Przewodniczący komisji konkursowej Dominique Arago, podjął próbę sprawdzenia eksperymentalnego przewidzianego efektu. Eksperyment pokazał, że taka słaba plamka rzeczywiście się pojawia (rys. 2.1.8b-c), co ostatecznie przesądziło o zwycięstwie w konkursie Fresnela i szybkim rozpowszechnieniu się teorii falowej. Od tych zdarzeń plamka pojawiająca się za krążkiem nazywana jest plamką Arago lub plamką Poissona.

Eksperyment Arago jest pięknym przykładem czegoś co w nauce nazywamy eksperymentem krzyżowym lub eksperyment rozstrzygającym (łac. experimentum crucis). Experimentum crucis, to eksperyment, który w mocny sposób rozstrzyga na rzecz jednej z konkurencyjnych teorii.

3. Siatki dyfrakcyjne 🔶

Siatki dyfrakcyjne to żelazny temat każdego kursu fizyki. Wiąże się to z tym, że z jednej strony są to proste w analizie struktury dyfrakcyjne, z drugiej strony mają istotne znaczenie dla współczesnej nauki. Najprostsza siatka dyfrakcyjna składa się z rzędu naprzemiennie ułożonych przezroczystych i nieprzezroczystych pasków (rys. 3.1)



Rysunek 3.1. Najprostsza siatka dyfrakcyjna złożona jest z *N*, równomiernie rozłożonych powietrznych szczelin w nieprzezroczystym materiale. Odległość *d* między środkami szczelin nazywamy stałą siatki.

Analizę obrazu uzyskiwanego przez siatkę, przy oświetleniu falą płaską zaczniemy od prostego modelu. Potraktujemy siatkę dyfrakcyjną jako układ n świecących punktów (inaczej mówiąc zaniedbamy fakt, że szczeliny mają niezerową szerokość i wysokość (rys. 3.2). Jest to nasz model kulistej krowy dla siatki. Szukamy natężenia światła w punkcie P, takim, że odległość P od siatki jest dużo większa od stałej siatki. Wtedy różnica dróg od środka sąsiednich szczelin do punktu P, z dobrym przybliżeniem, wyraża się wzorem⁴.

$$\delta = d\sin(\varphi)$$

3.1a

3.1

Maksima natężenia światła będą widziane pod kątem, dla którego fazory liczone z sąsiednich szczelin ułożą się w punkcie P pod tym samym kątem. Słowem różnica sąsiednich dróg musi być wielokrotnością długości fali λ .

$d\sin(\varphi) = m\lambda$

Wzór ten mówi nam pod jakimi kątami będziemy widzieli maksima natężenia światła. Liczba *m* przyjmuje wartości $m=0,\pm1,\pm2,\ldots$ i nazywana jest rzędem ugięcia.

⁴ Zastosowane tu przybliżenie zostało już przestudiowane, zobacz rysunek (TIII 5.3.8)



Rysunek 3.2. Siatka dyfrakcyjna złożona z *N*-szczelin i oświetlona falą płaską poosiową. Natężenie światła wyznaczamy w punkcie *P*, który znajduje się w dużej odległości w porównaniu z rozmiarami siatki, tak, że kąt φ jest mały. Przy małym kącie φ różnice dróg optycznych jakie światło ma do punktu P od dwóch sąsiednich szczelin wynosi, z dobrym przybliżeniem, δ = $dsin(\varphi)$. Fazory od sąsiednich szczelin ułożą się w fazie, gdy różnica dróg optycznych między sąsiednimi szczelinami będzie równa całkowitej wielokrotności długości fali λ światła padającego.

Gdzie pojawią się minima? Minima pojawią się tam, gdzie fazory od poszczególnych punktów dodadzą się do zera, tak jak to było w przypadku stref Fresnela (rys. 3.1.2). Stąd mamy warunek

$$\Delta = n\delta = m'\lambda \Rightarrow n\,d\,\sin(\varphi) = m'\lambda \qquad 3.2a$$

$$d\sin(\varphi) = \frac{m'}{n}\lambda$$
 3.2b

W tym wzorze *n* jest liczbą szczelin. Bez trudu możesz zauważyć, że gdy *m* jest podzielne przez *n* to otrzymujemy wzór (3.1) wyrażający kąty widzenia maksimów, a nie minimów. Będzie tak dla przykładu, gdy m'=20 i n=10. Wtedy iloraz m'n=2 i mamy warunek na maksimum drugiego rzędu. W pozostałych przypadkach mamy warunek na minima. Teraz już wiesz dlaczego we wzorze (3.2a) jest wyrażenie *m'*, a nie *m*. Obie liczby dotyczą czegoś innego. Liczba *m* jest rzędem dyfrakcji, a *m'*, to liczba oznaczająca kolejne minima. Dla danego rzędu dyfrakcji, liczba m', zmienia się w zakresie od 1 do *n*-1, czyli między dwoma maksimami jest *n*-1 minimów. Oczywiście między dwoma minimami

Fakt 3.2.1:

Dla siatki o liczbie szczelin n, między dwoma sąsiednimi maksimami głównymi mamy n-1 minimów i n-2 maksimów bocznych.

Sprawę ilustruje rysunek (3.3). Z rysunku widać, że dla trzech szczelin i danego rzędu dyfrakcji m, oraz liczby szczelin n, warunek na minima ma postać

$$d\sin(\varphi) = \left(m + \frac{k}{n}\right)\lambda$$
3.3

k zmienia się od 1 do *n*-1. Wzór ten można uogólnić dla większej liczby szczelin. Jest on oczywiści równoważny wyrażeniu (3.2b) (pomyśl dlaczego).



Rysunek 3.3. Rozpatrzmy przypadek siatki o trzech szczelinach. Niech w punkcie P₁ ekranu fazory układają się w fazie (a). Suma tych fazorów pokazana jest również w części (b) rysunku. Idąc w prawo natrafimy na punkt P₂, w którym fazory dodadzą się do zera (utworzą trójkąt). Gdy przechodzimy od punktu P₁ do punktu P₂ wszystkie fazory kręcą się, ale z różnymi prędkościami kątowymi. W punkcie P₂ fazor czerwony wyprzedzi fazor czarny o $2\pi/3$, a fazor niebieski wyprzedzi fazor czarny o $4\pi/3$. Idąc dalej w prawo natrafimy na punkt P₂, w którym fazor czerwony ustawi się przeciwnie do czarnego a fazor niebieski ponownie ustawi się zgodnie. W tym punkcie dwa fazory będą zgodnie współpracować przeciw trzeciemu i uzyskamy boczne maksimum (zobacz też konstrukcje na rysunkach (4.1.2 i 4.1.7)). W punkcie P₄ fazory ponownie utworzą trójkąt, a w punkcie P₅ zejdą się w fazie. Widać z tego, że maksimum boczne ma trzy razy mniejszą amplitudę niż maksimum główne, co oznacza dziewięć razy mniejsze natężenie.

Mamy określone kątowe położenie maksimów i minimów. A co pomiędzy nimi? Każdy kolejny punkt świecący reprezentujący kolejne szczeliny siatki jest dalej od punktu P o δ w stosunku do swojego sąsiada. Oznacza to, że w punkcie P każdy kolejny fazor, reprezentujący światło z kolejnej szczeliny będzie obrócony, w stosunku do poprzedniego, o kąt

 $\theta = k\delta$

3.4

Musimy teraz obliczyć sumę tak poobracanych fazorów w punkcie P. Męczyć się z tym nie będę, tylko skorzystam z usług programu Mathematica. Musimy znaleźć sumę *n* fazorów (wektorów), o kątach różniących się od siebie o $jk\delta$, przy *j* rosnącym od zera do *n*; gdzie *n* to liczba szczelin mniej jeden, *n*=*N*-1. Przyjmując, że długość fazorów jest jednostkowa, a faza pierwszego z nich (*j*=0) jest równa zero, współrzędne *j*-tego fazora wyrażają się wzorem

$(\cos(j\theta), \sin(j\theta))$

3.5

Obliczę, z pomocą programu Mathematica współrzędne wektora wypadkowego

Krok pierwszy: wpisać sumę cosinusów i nakazać jego wyliczenie	$\sum_{j=0}^{N-1} \cos[j\theta]$	
Wynik	$\frac{1}{2}(1 - \cos[N\theta] + \cot[\frac{\theta}{2}]\operatorname{Sin}[N\theta])$	
Krok drugi: to samo uczynić dla ciągu sinusów	$\sum_{m=0}^{N-1} \operatorname{Sin}[j\theta]$	
Wynik	$-\operatorname{Csc}\left[\frac{\theta}{2}\right]\operatorname{Sin}\left[\frac{N\theta}{2}\right]\operatorname{Sin}\left[\frac{1}{2}(\theta-N\theta)\right]$	
Krok trzeci: teraz wpisać sumę kwadratów obu wyników i nakazać ich uproszczenie	$\left(\frac{1}{2}(1 - \cos[N\theta] + \cot[\frac{\theta}{2}]\operatorname{Sin}[N\theta])\right)^{2} + \left(-\operatorname{Csc}[\frac{\theta}{2}]\operatorname{Sin}[\frac{n\theta}{2}]\operatorname{Sin}[\frac{1}{2}(\theta - N\theta)]\right)^{2} / /\operatorname{Simplify}$	
Wynik	$\operatorname{Csc}\left[\frac{\theta}{2}\right]^{2}\operatorname{Sin}\left[\frac{n\theta}{2}\right]^{2}$	
M. 3.1. Cykl obliczeń dla siatki dyfrakcyjnej		

Na dobrą sprawę całość mógłbym zmieści tylko w jednym kroku (M.2.2).

Instrukcja do wykonania	$\left(\left(\sum_{j=0}^{N-1} \operatorname{Cos}[j\theta]\right)^2 + \left(\sum_{j=0}^{N-1} \operatorname{Sin}[j\theta]\right)^2\right)$ //Simplify
wynik	$\operatorname{Csc}\left[\frac{\theta}{2}\right]^2 \operatorname{Sin}\left[\frac{n\theta}{2}\right]^2$

M 3.2. Obliczenia dla siatki dyfrakcyjnej w jednym kroku

Przez Csc program Mathematica oznacza funkcję cosecant czyli odwrotność funkcji sinus

$$\operatorname{Csc}(x) = \frac{1}{\sin(x)}$$
 3.6a

Stąd mamy wzór

$$\underbrace{\operatorname{Csc}\left[\frac{\theta}{2}\right]^{2}\operatorname{Sin}\left[\frac{N}{2}\theta\right]^{2}}_{\text{tako rzecze Mathematica}} = \underbrace{\frac{\sin^{2}\left(\frac{N}{2}\theta\right)}{\frac{\sin^{2}\left(\frac{\theta}{2}\right)}{\frac{1}{\sqrt{2}}}}_{\substack{\text{thumaczenie} \\ \text{na nasz zapis}}} 3.7$$

Widać, że ze wzorem (3.7) będziemy mieli łopot dla θ =0 – mamy wyrażenie typu 0/0. Korzystając z technik obliczania granic tego typu wyrażeń można pokazać, że

$$\lim_{x \to 0} \frac{\sin(N x)}{\sin(x)} = N$$
3.8

Wystarczy tu skorzystać ze znanego nam faktu, że dla małych x, $sin(x) \approx x$ (DB 3.3). Zatem możemy zapisać

$$\lim_{x \to 0} \frac{\sin^2\left(\frac{N}{2}\right)}{\sin^2\left(\frac{x}{2}\right)} = N^2$$
3.9

Funkcja sinus przyjmie zera jeszcze w innych punktach, wszędzie tam gdzie $x=\pm m'\pi$. Co się wtedy stanie z naszym wyrażeniem? Zobaczmy jak to będzie z wyrażeniem

$$\lim_{x \to \pi} \frac{\sin(\pm N \ m'\pi)}{\sin(\pm m'\pi)} = N$$
3.10

Czyli dokładnie jak w przypadku (3.8). Widać również, że wyrażenie to będzie równe zeru tam, gdzie spełniony jest warunek

$$\sin\left(\frac{N\theta}{2}\right) = 0 \Rightarrow \frac{N\theta}{2} = m'\pi \Rightarrow \frac{\theta}{2} = \frac{m'}{N}\pi \qquad 3.11$$

Wzór ten zgadza się z poprzednio wyprowadzonym wzorem na minima (3.2b). Możemy się o tym przekonać wstawiając do wzoru (3.11) wyrażenie (3.3) a następnie, za δ wyrażenie (3.1a).

$$\frac{\theta}{2} = \frac{m'}{N}\pi \stackrel{3.4}{\Longrightarrow} \frac{k\delta}{2} = \frac{m}{N}\pi \stackrel{3.1a}{\Longrightarrow} \frac{kd\sin(\varphi)}{2} = \frac{m}{N}\pi \stackrel{k=\frac{2\pi}{\lambda}}{\Longrightarrow} d\sin(\varphi) = \frac{m}{N}\lambda \quad 3.12$$

Biorąc pod uwagę (3.7) i (3.11) wzór na rozkład intensywności światła za siatką dyfrakcyjną przyjmuje postać

$$i(\varphi) = \frac{\sin^2\left(\frac{N \ k \ d \ \sin(\varphi)}{2}\right)}{\sin^2\left(\frac{k \ d \ \sin(\varphi)}{2}\right)}$$

$$3.13$$

Rysunek (3.4) przedstawia przykładowy rozkład natężenia światła wyrysowany ze wzoru (3.13)



Rysunek 3.4. Unormowany do jedności wykres natężenie światła, obliczony na podstawie wzoru na obraz siatki, dla wybranych przykładów: a) N=2 i d=1mm; b) N=4 i d=4mm. Czerwone punkty wskazują minima natężenie z prawej strony centralnego maksimum. Widać, że między maksimami mamy N-1 minimów, co jest zgodne ze wzorem (3.2b i 3.3).

Ponieważ wzór na rozkład natężenia światła obowiązuje dla małych kątów φ możemy sinusy zastąpić wartościami kąta wyrażonymi w radianach.

$$i(\varphi) = \frac{\sin^2\left(\frac{N \ k \ d \ \varphi}{2}\right)}{\sin^2\left(\frac{k \ d \ \varphi}{2}\right)}$$

$$3.14$$

Przypominam, że podobną w swym duchu analizę prowadziłem (§TIX 3). Analizowałem tam rozkład natężenia światła wokół dwóch punktów (np. rysunki TIX 3.11 i 3.12). W naszym modelu odpowiada to siatce z dwoma szczelinami.

3.1. Analiza widmowa

Ze wzoru (3.1) widać, że położenie maksimów (za wyjątkiem zerowego rzędu) zależy od długości fali. Oznacza to, że dla światła białego poszczególne składowe widmowe będą wzmacnianie, dla tego samego rzędu dyfrakcji m, pod nieco innym kątem. W efekcie światło białe ulegnie rozszczepieniu, z wyjątkiem rzędu zerowego (m=0). Efekt ten zilustrowany jest na rysunku (3.1.1). Mówimy, że siatka dyfrakcyjna ma chromatyzm.

Gdy pobudzamy do świecenia gaz złożony z jednego rodzaju molekuł, to światło emitowane przez ten gaz składa się z fal o ściśle określonych długościach. Każdy atom czy chemiczna molekuła ma charakterystyczny rozkład długości emitowanych fal nazywany widmem (rys. 3.1.2). Siatka dyfrakcyjna pozwala na rozseparowanie poszczególnych składowych widma, co jest podstawą działania spektrometrów.



Rysunek 3.1.1. Siatka dyfrakcyjna rozszczepia światło białe. Na rysunku widać miejsca występowania kolejnych maksimów dla poszczególnych barw światła białego.

Przydatność siatki do budowy spektrometrów określa parametr nazywany rozdzielczością siatki. Rysunki (3.1.3-5) ilustrują problem rozdzielczości. Pokazane przebiegi natężenia światła zostały wykreślone ze wzoru (3.12).



Rysunek 3.1.2. Widmo atomów, od góry: wodoru, azotu, żelaza. Widmo na zdjęciach obejmuje zakres widzialny. Wymienione atomy emitują również fale o długości z zakresu podczerwonego i nadfioletowego.

Jeżeli fala padająca na siatkę składa się z składowych o różnych długościach, to składowe te nakładają się na siebie. Gdy ich obrazy są zbyt blisko wzajemne nakładanie może spowodować, że przestaną być dla nas rozróżnialne.



Rysunek 3.1.3. Na górze pokazany jest unormowany rozkład intensywność dla dwóch fal o długości λ_1 =500nm oraz λ_2 =610nm; $\Delta\lambda$ =110nm. Niebieska linia wskazuje położenie pierwszego lewego minimum dla fali λ_1 . Rysunek na dole pokazuje sumę obu rozkładów. Piki dla obu fal w pierwszym i minus pierwszym rzędzie dyfrakcji są dobrze widoczne.

W celu oceny rozdzielczości siatki dyfrakcyjnej stosuje się kryterium Rayleigha.

Określenie 3.1: Kryterium Rayleigha

Jeżeli maximum pierwszego obrazu leży dokładnie nad pierwszym minimum drugiego obrazu, to jesteśmy na granicy rozdzielczości, jeżeli leży bliżej, to obrazy nie są rozróżnialne.

Ostatni przykład rysunkowy (3.1.5) jest na granicy kryterium Rayleigha (przebieg niebieskiej kreski pokazuje, że główne maksimum jednej fali wypada dokładnie nad pierwszym minimum drugiej fali). W pierwszym przykładzie rysunkowym (3.1.3) kryterium Rayleigha spełnione jest z nawiązką, w drugim przykładzie (3.1.4) kryterium Rayleigha nie jest spełnione.



Rysunek 3.1.4. Na górze pokazany jest unormowany rozkład intensywność dla dwóch fal o długości λ_1 =500nm oraz λ_2 =540nm; $\Delta\lambda$ =40nm. Niebieska linia wskazuje położenie pierwszego lewego minimum dla fali λ_1 . Rysunek na dole pokazuje sumę obu rozkładów. Piki dla obu fal w pierwszym i minus pierwszym rzędzie dyfrakcji nakładają się na siebie tak, że praktycznie nie wiadomo czy pochodzą od fali o jednej czy dwóch długościach. Identyfikacja długości obu fal składowych może być niemożliwa.

Możemy wyprowadzić wzór na warunek Rayleigha na rozdzielczość siatki. Rozważmy falę o długości λ . W *m*-tym rzędzie dyfrakcji pierwsze minimum widziane jest pod kątem φ spełniającym warunek (3.3)

$$d\sin(\varphi) = \frac{(mn+1)\lambda}{n}$$
 3.1.1

Chcemy, żeby dla tego samego kąta, rzędu dyfrakcji oraz dla fali o długości $\lambda + \Delta \lambda$ widziane było maksimum główne. Stąd mamy warunek

$$d\sin(\varphi) = m(\lambda + \Delta\lambda)$$
3.1.2



Rysunek 3.1.5. Na górze pokazany jest unormowany rozkład intensywność dla dwóch fal o długości λ_1 =500nm oraz λ_2 =570nm; $\Delta\lambda$ =70nm. Niebieska linia wskazuje położenie pierwszego lewego minimum dla fali λ_1 . Rysunek na dole pokazuje sumę obu rozkładów. Piki dla obu fal w pierwszym i minus pierwszym rzędzie dyfrakcji są delikatnie widoczne, identyfikacja długości obu fal składowych powinna być możliwa.

Porównując te dwa warunki mamy

$$\frac{\lambda}{\Delta\lambda} = mn$$

Jest to bardzo szczęśliwy wynik. Rozdzielczość siatki nie zależy od stałej siatki ale od rzędu dyfrakcji m i ilości szczelin n. Należy jednak pamiętać, że wynik ten obowiązuje dla małych kątów obserwacji φ .

Policzmy ile wynosi rozdzielczość siatki o siedmiu szczelinach w pierwszym rzędzie dyfrakcji przy długości fali λ =500nm.

$$\Delta \lambda = \frac{\lambda}{mn} = \frac{500nm}{7} = 71,4nm$$
3.1.4

Odpowiada to przykładowi rysunkowemu (3.1.5). Zwiększyć rozdzielczość siatki możemy zwiększając ilość szczelin. Dla 21 szczelin otrzymamy w pierwszym rzędzie $\Delta\lambda$ =23,8nm (rys. 3.1.6)



Rysunek 3.1.6. Zwiększenie ilość szczelin z 7 (rys. 3.1.4) do 21 skutkuje zwiększeniem rozdzielczości siatki. λ_1 =500nm oraz λ_2 =540nm; $\Delta\lambda$ =40nm

Również w drugim rzędzie dyfrakcji sprawy przedstawiają się lepiej (rys. 3.1.7).



Rysunek 3.1.7. W drugim rzędzie dyfrakcji analizowana tu siatka przy siedmiu szczelinach ma wystarczającą zdolność rozdzielczą do odróżnienia fal o długości λ_1 =500nm oraz λ_2 =540nm. Zdolność rozdzielcza siatki wynosi teraz $\Delta\lambda$ =35,7nm

Współczesna technologia pozwala wytwarzać siatki o wielu tysiącach szczelin. Dla tysiąca szczelin w pierwszym rzędzie dyfrakcji dla światła o długości λ =500nm rozdzielczość wynosi $\Delta\lambda$ =0,5nm – a nie jest to robiący wrażenie wynik.

Jeszcze kilka uwag w kwestii kryterium Rayleigha. Po pierwsze należy pamiętać, że źródłem ograniczenia rozdzielczości siatek dyfrakcyjnych jest szum (rys. 3.1.8). Przy braku szumu stosując dostępne procedury matematyczne moglibyśmy określić długość obu fal ze znacznie większą dokładnością niż to wynika z kryterium Rayleigha. Po drugie kryterium to nie ma charakteru absolutnego. Należy je rozumieć tak: siatka o przyzwoitej jakości przy podstawowej analizie obrazu prowadzonej przez przyzwoicie wyszkolonego specjalistę powinna pozwolić na rozdzielenie dwóch linii widmowych, gdy długości odpowiednich fal spełniają kryterium Rayleigha. Jednakże, gdy siatka jest bardzo dobrej jakości i/lub rejestrujemy sygnał specjalnym detektorem i mamy do dyspozycji wysokiej klasy specjalistę i/lub specjalne algorytmy obróbki sygnału możemy uzyskać większą rozdzielczość niż by to wynikało z kryterium Rayleigha. Jeszcze inaczej rzecz ujmując kryterium Rayleigha jest punktem orientacyjnym dla dobrego inżyniera (jest to kryterium inżynierskie). Kiedy do akcji przystępuje doskonale panujący nad wysokiej jakości sprzętem i świetnie znający teorią inżynier-artysta wtedy kryterium Rayleigha jest zbyt ostrożnym podejściem do tematu. Kiedy za sprawę bierze się partacz lub sprzęt jest zły żadne kryterium nie pomoże.



Rysunek 3.1.8. Jeszcze raz przykład z rysunku (3.1.5). Teraz jednak za pomocą funkcji generacji liczb losowych dodałem do wykresu szum na poziomie 5% maksymalnej wartości sygnału czystego. Od razu widać, że uzyskane wykresy stały się mniej jednoznaczne. I choć zastosowany tu numeryczny model szumu jest prymitywny daje dobrą intuicję problemów jakie mamy w rzeczywistych pomiarach.

4. Całki dyfrakcyjne 🐥

Czeka nas zadanie przetłumaczenia zasady Fresnela-Huygensa na język matematyki. Patrząc na rysunek (4.1) możemy zapisać następujące wyrażenie

u(P)

$$= \exp\{i\omega t\} \iint_{\Sigma} u_{\text{pad}}(x; y; z) \underbrace{\frac{1}{|\mathbf{r}(x, y, z)|} \exp\{i|\mathbf{k}(x, y, z)||\mathbf{r}(x, y, z)|\}}_{\text{jednostkowa fala kulista}} dxdy \quad 4.1$$

Przypatrz się temu wyrażeniu: w każdym punkcie przeszkody o współrzędnych x,y,z=const, mamy jednostkową falę kulistą (TIX 1.2.1a), o fazie początkowej $\varphi=0$, która jest falą wtórną z zasady Fresnela-Huygensa (okr. 2.1). W ogólnym przypadku amplitudę zespoloną fali padającej u_{pad} na otwór, oznaczony tu przez Σ , możemy zapisać w postaci (1.5.1)

$$u_{pad}(x, y, z) = a_{pad}(x, y, z) \exp\{i\varphi_{pad}(x, y)\}\exp\{i\omega t\}$$
4.2

Wtedy wzór (4.1) przejedzie w (przyjmę dodatkowo, że otwór znajduje się w płaszczyźnie z=0)

$$u(P) = \exp\{i\omega t\} \iint_{\Sigma} \frac{a_{pad}(x, y)}{|\mathbf{r}(x, y)|} \exp\{i|\mathbf{k}(x, y)||\mathbf{r}(x, y)| + \phi_{pad}(x, y)\} dxdy$$

$$4.3a$$

Pod całką dalej mamy wzór na falę kulistą, tyle że teraz w każdym punkcie otworu wtórna fala kulista dziedziczy, po fali padającej, amplitudę a_{pad} i fazę początkową φ_{pad} . Aby obliczyć natężenie światła trzeba z powyższego wyrażenia policzyć moduł i podnieść go do kwadratu

$$i(P) = \left| \iint_{\Sigma} \frac{a_{pad}(x, y)}{|\mathbf{r}(x, y)|} \exp\{i|\mathbf{k}(x, y)||\mathbf{r}(x, y)| + \phi_{pad}(x, y)\} dxdy \right|^{2}$$

$$4.3b$$

Pod modułem mamy sumę (całka to suma nieskończenie wielu nieskończenie małych wkładów) amplitud fal wtórnych. Jak wiemy z (§TIX 3) taka suma podniesiona do kwadratu opisuje zjawisko interferencji. Zatem wzór (4.3) jest matematycznym zapisem zasady Huygensa-Fresnela.

Obliczanie ze wzorów (4.1 i 4.3) sprawia spore kłopoty nawet w prostych przypadkach. Aby uczynić je bardziej życiowymi konieczne są uproszczenia.

Przyjmę, że odległość do punktu obserwacji P jest duża, tak że z dobrym przybliżeniem mogę zastąpić \mathbf{r} w liczniku wyrażenia (4.1) przez odległość z_o (rys. 4.1)

$$|\mathbf{r}| \approx z_o$$



Rysunek 4.1. Schemat układu z otworem do wyprowadzenia przybliżenia Fresnela

Wzór (4.1) przyjmie postać

$$u(P) = \frac{\exp\{i\omega t\}}{z_o} \iint_{\Sigma} u_{pad}(x, y) \exp\{i|\mathbf{k}(x, y)||\mathbf{r}(x, y)|\} dxdy$$

$$4.4$$

Niestety tego samego przybliżenia nie możemy zastosować w wyrazie $\exp\{i|\mathbf{k}||\mathbf{r}|\}$. Powód jest następujący:

4.0.1. Wyrażenie $\exp\{i|\mathbf{k}||\mathbf{r}|\}$

Rysunek (4.2) przedstawia zachowanie się wyrażenia $\exp\{i|\mathbf{k}||\mathbf{r}|\}$, przy przyjęciu przybliżenia (4.4.a). Ponieważ wartość funkcji $exp{i coś}$ jest liczbą zespoloną o module równym jeden wszystkie wartości, przy zmieniającym się coś ułożą się na jednostkowym okręgu (rys. DD 1.1.5). Fakt, że badana funkcja jest cykliczna jest jednym z powodów, dla którego nie będzie można zastosować przybliżenia (4.4a). Na rysunku (4.2b) przyjąłem, że odległość między otworem a ekranem wynosi $z_0=100$ mm. Już przy kącie nachylenia promienia względem osi z równym ok. $0,2^{\circ}$ funkcja exp $\{i|\mathbf{k}||\mathbf{r}|\}$ powraca do punktu wyjścia. Czarny punkt pokazuje wartość przybliżoną $\exp\{ikz_n\}$. Ponieważ z_n jest stałe dla wszystkich promieni, wartość przybliżona jest stała. Gdy wartości dokładne przejda cały okrąg nasze przybliżenie traci wszelką użyteczność, w sumie może reprezentować dowolny punkt okręgu. W analizowanym przykładzie przybliżenie (4.4a) ma sens dla wartości kata mniejszej od około 0,02°. Dla obliczania praktycznych zagadnień dyfrakcyjnych katy te są zbyt małe. Warto zwrócić uwagę, że szybkie zmiany wrażenia $\exp\{i|\mathbf{k}||\mathbf{r}|\}$ związane są z dużą wartością liczby falowej k, która dla światła widzialnego ma wartość rzędu

4.4a

10 000 1/mm. Oznacza to, że z wyrażeniem $\exp\{i|\mathbf{k}||\mathbf{r}|\}$ musimy poradzić sobie inaczej.



Rysunek 4.2. a) otwór na który pada fala płaska dzielimy na punkty (tu podzielona jest górna część otworu. Następnie obliczamy długości promieni od kolejnych punktów w otworze do wybranego punktu ekranu (tu został wybrany punkt na osi); b) wyrażenie exp{*ikr*} obliczone zostało dla tego zbioru punktów. Punkty te układają się, na płaszczyźnie Arganda (def. DD 1.1.1), wzdłuż jednostkowego okręgu. Odległość otwór-ekran wynosi z_0 =100mm, najbardziej odchylony od osi *z* promień tworzy z nią kąt około 0,2°. Dla tego kąta punkt wraca do położenia obliczonego dla kąta zero. Czarny punkt pokazuje wyrażenie przybliżone (2.1.5) exp{*ikz*₀}. Ponieważ z_0 jest stałe, dla wszystkich promieni otrzymujemy tą samą wartość

. .

Długość wektora **r** możemy wyrazić wzorem (rys. 4.1)

$$\mathbf{r}| = \sqrt{(x - x_o)^2 + (y - y_o)^2 + z_o^2}$$
4.5

Inaczej możemy zapisać

. .

_ -

$$|\mathbf{r}| = z_o \sqrt{1 + \frac{(x - x_o)^2}{z_o^2} + \frac{(y - y_o)^2}{z_o^2}}$$
 4.6

Zajmę się teraz wyrażeniem typu $\sqrt{1+t}$, bo za takim wyrażeniem mamy do czynienia powyżej, jeżeli przyjmiemy, że *t* ma postać

$$t = \frac{(x - x_o)^2}{z_o^2} + \frac{(y - y_o)^2}{z_o^2}$$
4.7

Rozwiniemy to wyrażenie w szereg Taylora (§DB 3) wokół punktu *t*=0 ma postać

$$\sqrt{1+t} = 1 + \frac{1}{2}t - \frac{1}{8}t^2 + \frac{1}{16}t^3 - \frac{5}{128}t^4 + \cdots$$
4.8

Postąpimy teraz zgodnie z metodologią drogi krowy, czyli nieco skomplikujemy model. To nieco oznacza, że przyjmiemy teraz tzw. przybliżenie pierwszego rzędu i w szeregu (4.8) odrzucimy wyrazy w potędze dwa i większej

$$\sqrt{1+t} \approx 1 + \frac{1}{2}t \tag{4.9}$$

Wstawiając za t wyrażenie (4.7) otrzymujemy

$$\sqrt{1 + \frac{(x - x_o)^2}{z_o^2} + \frac{(y - y_o)^2}{z_o^2}} \approx 1 + \frac{1}{2} \left[\frac{(x - x_o)^2}{z_o^2} + \frac{(y - y_o)^2}{z_o^2} \right]$$
 4.10

Wstawiając to wyrażenie do wzoru (4.6) mamy

$$|\mathbf{r}| \approx z_o + \frac{1}{2} z_o \left[\frac{(x - x_o)^2}{z_o^2} + \frac{(y - y_o)^2}{z_o^2} \right]$$

$$4.11$$

Stąd

$$|\mathbf{r}| \approx z_o + \frac{1}{2} \left[\frac{(x - x_o)^2}{z_o} + \frac{(y - y_o)^2}{z_o} \right]$$
 4.12

Co jest szukanym przez nas przybliżeniem wartości |r|.

Rysunek (4.3) ilustruje zachowanie się przybliżenia (4.12) w wyrażeniu $\exp\{i|\mathbf{k}||\mathbf{r}|\}$. Rysunek (4.3a) pokazuje, że dla wartości kąta, dla których obliczony był rysunek (4.3b) różnice między oboma wyrażeniami $\exp\{i|\mathbf{k}||\mathbf{r}|\}$ i $\exp\{ikr_p\}$ są zaniedbywalnie małe. Nawet dla kątów bliskich 2° przybliżenie zachowuje się bardzo dobrze (rys. 4.3b).

Wyrażenie dokładne (4.5) Może być interpretowane jako równanie sfery o środku w punkcie (x_o , y_o), leżącym w płaszczyźnie otworu i promieniu $|\mathbf{r}_o|$, co jest zgodne z zasadą Huygensa (analizujemy wtórne fale kuliste dochodzące do danego punktu obserwacji *P*). Wyrażenie (4.11), które możemy zapisać w postaci,

$$2|\mathbf{r}|z_o \approx (x - x_o)^2 + (y - y_o)^2 + 2z_o^2$$
4.13

przedstawia parabolę. Ponieważ taką parabolę dobieramy dla każdej fali wtórnej, z punktu widzenia geometrii, nasze przybliżenie polega na zastąpieniu, dla każdego punktu ekranu, wtórnych fal kulistych najbardziej do nich podobnymi wtórnymi falami parabolicznymi. Dlatego to przybliżenie nazywa się również przybliżeniem parabolicznym. Inna nazwa to przybliżenie paraksjalne (przyosiowe). Nazwa wzięła się stąd, że przybliżenie ma wartość dla promieni biegnących pod niewielkim kątem w stosunku do osi układu optycznego.

Fakt 4.1

Przybliżenie (4.12) nosi w optyce nazwę przybliżenia parabolicznego, lub paraksjalnego.



Rysunek 4.4. a) przedstawia ten sam przykład co rysunek (4.2b), tyle że punkty czarne reprezentują teraz wyrażenie przybliżone $\exp\{ikr_p\}$. Widać dużą zgodność między wyrażeniem dokładnym i przybliżonym; b) położenie punktów dla kątów rosnących od 1,7° do 1.73°. Niebieskie kółko obejmuje parę punktów obliczoną dla tego samego kąta. Dla tych wartości kąta przybliżenie ciągle dobrze się spisuje.

4.1. Całka dyfrakcyjna Fresnela

Wyrażenie (4.12) podstawię do całki (4.4)

$$u(P) = \frac{\exp\{i\omega t\}}{z_o} \iint_{\Sigma} u_{pad}(x, y) \exp\{ik(z_o + \frac{1}{2}\left[\frac{(x - x_o)^2}{z_o} + \frac{(y - y_o)^2}{z_o}\right])\} dxdy$$
4.1.1

Długość wektora falowego $|\mathbf{k}|$ będę oznaczał przez liczbę falową k. Po przekształceniach mam

$$u(P) = \underbrace{\frac{\exp\{i\omega t\}\exp\{ikz_o\}}{z_o}}_{\Xi} \iint_{\Sigma} u_{pad}(x, y)\exp\left\{i\frac{k}{2z_o}([(x-x_o)^2 + (y-y_o)^2])\right\} dxdy$$

$$(P) = \underbrace{\frac{\exp\{i\omega t\}\exp\{ikz_o\}}{z_o}}_{\Xi} \iint_{\Sigma} u_{pad}(x, y)\exp\left\{i\frac{k}{2z_o}([(x-x_o)^2 + (y-y_o)^2])\right\} dxdy$$

$$(P) = \underbrace{\frac{\exp\{i\omega t\}\exp\{ikz_o\}}{z_o}}_{\Xi} \iint_{\Sigma} u_{pad}(x, y)\exp\left\{i\frac{k}{2z_o}([(x-x_o)^2 + (y-y_o)^2])\right\} dxdy$$

Jest to pierwsza postać całki dyfrakcyjnej Fresnela (całka dyfrakcyjna w przybliżeniu bliskiego pola, lub jeszcze inaczej w przybliżeniu parabolicznym, lub paraksjalnym). Drugą, w pełni równoważną postać otrzymamy rozpisując kwadratowe wyrazy pod funkcją exp i wyłączając przed znak całki te czynniki, które nie zależą od zmiennych całkowania.

$$u(P) = \frac{\exp\{i\omega t\}\exp\{ikz_o\}\exp\left\{i\frac{k}{2z_o}(x_o^2 + y_o^2)\right\}}{\sum_{\Sigma} u_{\text{pad}}(x, y)\exp\left\{-i\frac{k}{2z_o}(2xx_o + 2yy_o)\right\}\exp\left\{i\frac{k}{2z_o}(x^2 + y^2)\right\}}$$

$$(1.13)$$

Przedstawię obie wersje całki Fresnela w postaci w jakiej będę ją dalej zapisywał

$$u(P) = \Xi \iint_{\Sigma} u_{pad}(x, y) \exp\left\{i\frac{k}{2z_o}([(x - x_o)^2 + (y - y_o)^2])\right\} dxdy$$
4.1.4a

$$u(P) = \Omega \iint_{\Sigma} u_{pad}(x, y) \exp\left\{-i\frac{k}{z_o}(xx_o + yy_o)\right\} \exp\left\{i\frac{k}{2z_o}(x^2 + y^2)\right\} dxdy$$

$$(P) = \Omega \iint_{\Sigma} u_{pad}(x, y) \exp\left\{-i\frac{k}{z_o}(xx_o + yy_o)\right\} \exp\left\{i\frac{k}{2z_o}(x^2 + y^2)\right\} dxdy$$

$$(P) = \Omega \iint_{\Sigma} u_{pad}(x, y) \exp\left\{-i\frac{k}{z_o}(xx_o + yy_o)\right\} \exp\left\{i\frac{k}{2z_o}(x^2 + y^2)\right\} dxdy$$

Definicja 4.1.1: Całka dyfrkacyjna Fresnela (bliskiego pola)

Całka dyfrakcyjna Fresnela (bliskiego pola) wyraża się wzorami postaci (4.1.4) lub równoważnej.

Podsumowując: całkę dyfrakcyjną bliskiego pola możemy interpretować tak jak całkę dyfrakcyjną (4.4), z tym że wtórne fale kuliste zamieniamy na wtórne fale paraboliczne. Każda fala paraboliczna dziedziczy fazę i amplitudę po fali padającej. (komentarz nad 4.3b).

Wypadałoby policzyć przynajmniej jednej przykład z wykorzystaniem całki dyfrakcyjnej Fresnela. Względnie prostym ćwiczeniem będzie obliczenie dyfrakcji fali płaskiej na półpłaszczyźnie.

4.1.1. Przykład

Przedmiotem jest nieprzepuszczająca światła, nieograniczona półpłaszczyzna, którą umieścimy w początku układu współrzędnych prostopadle do osi z. Na tą półpłaszczyznę pada fala płaska propagująca się wzdłuż osi z (rys. 4.1.1). Chcemy obliczyć rozkład amplitudy zespolonej w płaszczyźnie ekranu znajdującego się w odległości z_o od półpłaszczyzny. Skorzystam ze wzoru (4.1.4a) pamiętając, że fala wychodząca z przedmiotu jest falą płaską obciętą na półpłaszczyźnie. Obcięcie to możemy zrealizować ustawiając granice całkowania po y na przedziale $[0; \infty]$. W ten sposób mamy

$$u(x_{o}, y_{o}) = \Xi \int_{0}^{\infty} \int_{-\infty}^{\infty} \exp\{ikz\} \exp\{\frac{ik}{2z_{o}}((x - x_{o})^{2} + (y - y_{o})^{2})\} dxdy$$

$$4.1.5$$



Rysunek 4.1.1. Fala płaska poosiowa oświetla nieprzezroczystą półpłaszczyznę.

W powyższej całce możemy wyłączyć funkcję $\exp\{ikz\}$ oraz dokonać jej faktoryzacji na całki obliczane po zmiennej x i y.

$$u(x_o, y_o) = \operatorname{Eexp}\{ikz\} T_x T_y$$

$$4.1.6a$$

$$T_{x} = \int_{-\infty}^{\infty} \exp\left\{\frac{ik}{2z_{o}}(x - x_{o})^{2}\right\} dx$$
 4.1.6b
$$T_{y} = \int_{0}^{\infty} \exp\left\{\frac{ik}{2z_{o}}(y - y_{o})^{2}\right\} dy$$
 4.1.6c

Przekształcimy te całki z pomocą podstawienia

$$t_x = \sqrt{\frac{2}{\lambda z_o}} (x - x_o) \Longrightarrow dt_x = \sqrt{\frac{2}{\lambda z_o}} dx$$
 4.1.7a

$$t_y = \sqrt{\frac{2}{\lambda z_o}} (y - y_o) \Longrightarrow dt_y = \sqrt{\frac{2}{\lambda z_o}} dy$$
 4.1.7b

$$t_{0y} = \sqrt{\frac{2}{\lambda z_o}} \left((y=0) - y_o \right) = -y_o \sqrt{\frac{2}{\lambda z_o}}$$

$$4.1.7c$$

Równanie (4.1.7c) określa dolną granicę całkowania dla całki (4.1.6c) wyrażonej przez zmienną t_y . Całki (4.1.2 b i c) przyjmą postać

$$T_x = \sqrt{\frac{\lambda z_o}{2}} \int_{-\infty}^{\infty} \exp\left\{\frac{i\pi}{2}t_x^2\right\} dt_x$$
4.1.8a

$$T_{y} = \sqrt{\frac{\lambda z_{o}}{2}} \int_{t_{0y}}^{\infty} \exp\left\{\frac{i\pi}{2}t_{y}^{2}\right\} dt_{y}$$

$$4.1.8b$$

Całki te są problematyczne i ich rozwiązania nie dadzą się wyrazić przez funkcje elementarne. Ponieważ jednak całki tej postaci pojawiają się przy wielu zagadnieniach, istotnych z punktu widzenia techniki i nauki, poradzono sobie wprowadzając nowe funkcje tzw. cosinusy i sinusy całkowe nazywane też całkami Fresnela (DX xx). Wzór (4.1.9a) podaje definicje sinusa całkowego, a wzór (4.1.9b) podaje definicje cosinusa całkowego.

$$S(x) = \int_{0}^{x} \sin\left(\frac{1}{2}\pi t^{2}\right) dt$$

$$4.1.9a$$

$$C(x) = \int_{0}^{x} \cos\left(\frac{1}{2}\pi t^{2}\right) dt$$

$$4.1.9b$$

Z pomocą tych funkcji specjalnych mogę obliczyć całki (4.1.8). Będę przy tym korzystał z własności sinusa i cosinusa całkowego (DX xx, xx). Zacznę od całki (4.1.8a). Mogę ją zapisać w postaci

$$T_x = \sqrt{\frac{\lambda z_o}{2}} \left[\int_{-\infty}^{\infty} \cos\left\{\frac{\pi}{2} t_x^2\right\} dt_x + i \int_{-\infty}^{\infty} \sin\left\{\frac{\pi}{2} t_x^2\right\} dt_x \right]$$

$$4.1.10$$

Wezmę się za całkę z cosinusem

$$\int_{-\infty}^{\infty} \cos\left\{\frac{\pi}{2}t_x^2\right\} dt_x = -\int_{0}^{-\infty} \cos\left\{\frac{\pi}{2}t_x^2\right\} dt_x + \int_{0}^{\infty} \cos\left\{\frac{\pi}{2}t_x^2\right\} dt_x$$

$$= -\frac{-1}{2} + \frac{1}{2} = \frac{1}{4}$$
4.1.11

Całka zawierająca sinus daje dokładnie taki sam wynik. Po dodaniu obu składników T_x przyjmie postać

$$T_x = \frac{1}{2} \sqrt{\frac{\lambda z_o}{2}}$$

$$4.1.12$$

Całka (4.1.8b) jest nieco bardziej skomplikowana

$$T_{y} = \sqrt{\frac{\lambda z_{o}}{2}} \left[\int_{t_{oy}}^{\infty} \cos\left\{\frac{\pi}{2}t_{x}^{2}\right\} dt_{x} + i \int_{t_{oy}}^{\infty} \sin\left\{\frac{\pi}{2}t_{x}^{2}\right\} dt_{x} \right]$$

$$4.1.13$$

Wezmę się za całkę z cosinusem

$$\int_{t_{0y}}^{\infty} \cos\left\{\frac{\pi}{2}t_{x}^{2}\right\} dt_{x} = \int_{0}^{\infty} \cos\left\{\frac{\pi}{2}t_{x}^{2}\right\} dt_{x} - \int_{0}^{t_{0y}} \cos\left\{\frac{\pi}{2}t_{x}^{2}\right\} dt_{x}$$

$$= \frac{1}{2} - C(t_{0y})$$
4.1.14

Podobnie możemy zapisać całkę z sinusem

$$\int_{t_{0y}}^{\infty} \sin\left\{\frac{\pi}{2}t_{x}^{2}\right\} dt_{x} = \int_{0}^{\infty} \sin\left\{\frac{\pi}{2}t_{x}^{2}\right\} dt_{x} - \int_{0}^{t_{0y}} \sin\left\{\frac{\pi}{2}t_{x}^{2}\right\} dt_{x}$$

$$= \frac{1}{2} - S(t_{0y})$$
4.1.15

Mamy zatem

$$T_{y} = \sqrt{\frac{\lambda z_{o}}{2}} \left(\mathsf{C}(\infty) - \mathsf{C}(t_{0y}) + i \,\mathsf{S}(\infty) - i\mathsf{S}(t_{0y}) \right)$$

$$4.1.16$$

Wyrażenie (4.1.5) przyjmie postać

Jan Masajada © - 45 tematów z fizyki

$$u(x_o, y_o) = \Xi \exp\{ikz\} \frac{\lambda z_o}{4} \left(\frac{1}{2} - C(t_{0y}) + \frac{i}{2} - iS(t_{0y})\right)$$
 4.1.17

Natężenie światła jest równe

$$i(x_o, y_o) = |\Xi|^2 \frac{(\lambda z_o)^2}{16} \left[\left(\frac{1}{2} - C(t_{0y}) \right)^2 - \left(\frac{1}{2} - S(t_{0y}) \right)^2 \right]$$

$$4.1.18$$

Wzór (4.1.18) opisuje szukane natężenie światła. Wykres natężenia przedstawiony jest na rysunku (4.1.2). Powinniśmy jeszcze wrócić do pierwotnych zmiennych. Wstawiając do (4.1.8a, b) wyrażenie (4.1.18) mamy

$$i(x_o, y_o) \sim \left[\left(\frac{1}{2} - C \left(-y_o \sqrt{\frac{2}{\lambda z_o}} \right) \right)^2 + \left(\frac{1}{2} - S \left(-y_o \sqrt{\frac{2}{\lambda z_o}} \right) \right)^2 \right]$$
 4.1.19



Rysunek 4.1.2. Wykres kwadratu natężenia światła wzdłuż osi *y* dla światła czerwonego o długości fali 633nm. Odległość półpłaszczyzny od ekranu wynosi z_o =100mm dla niebieskiego wykresu i z_o =300mm dla fioletowego wykresu. Oś y_o zorientowana jest tak jak na rysunku (4.1.2).

Na rysunku (4.1.2) krawędź przesłony jest na wysokości linii $y_o=0$, zatem zgodnie z zasadami optyki geometrycznej promienie świetlne nie powinny docierać w obszar $y_o<0$ (obszar cienia), a obszar $y_o\geq0$ powinien być równo naświetlony. Jednak zjawisko dyfrakcji powoduje, że w obszar cienia dostaje się światło. W obszarze jasnym, blisko linii krawędzi półpłaszczyzny natężenie światła oscyluje, przy czym oscylacje te zanikają wraz z oddalaniem się od tej granicy. Na fotografii wygląda to tak jak na rysunku (4.1.3).



Rysnek 4.1.3. Obraz dyfrakcyjny szczeliny oświetlonej poosiową falą płaską . Zamiast wyraźnie zaznaczonej granicy cienia widzimy układ prążków.

4.2. Całka dyfrakcyjna Fraunhofera

Jeżeli możemy przyjąć, że w całce dyfrakcyjnej Fresnela (4.1.3) człon

$$\frac{k}{2z_o}(x^2 + y^2) \approx 0 \tag{4.2.1}$$

to wtedy wyrażenie

$$\exp\left\{\frac{ik}{2z_o}(x^2+y^2)\right\} \approx 1$$
4.2.2

i całkę dyfrakcyjną Fresnela (4.1.3) możemy zapisać w postaci

$$u(P_o) = \Omega \iint_{\Sigma} u_{pad}(x, y, z) \exp\left\{-i\frac{k}{z_o}(xx_o + yy_o)\right\} dxdy$$

$$4.2.3$$

Jest to całka dyfrakcyjna w przybliżeniu dalekiego pola (przybliżeniu Fraunhofera).

Definicja 4.2.1: Całka dyfrakcyjna Fraunhofera (dalekiego pola)

Całka dyfrakcyjna Fraunhofera (dalekiego pola) wyraża się wzorami postaci (4.2.3) lub równoważnej.

Warunek (4.2.1) jest warunkiem silnym, to znaczy, że kąty rozchodzenia się promieni w układzie muszą być bardzo małe. Z drugiej strony wyrażenie (4.2.1) ma bardzo atrakcyjną formę. Pokażę zaraz, że w dalekim polu całka dyfrakcyjna sprowadza się do transformaty Fouriera (STIX 4.3) z funkcji u_{pad} opisującej falę padającą na otwór. Wprowadzę nowe wielkości (rys. 4.2.1)

$$f_x = \frac{\sin(\alpha_x)}{\lambda}; \ f_y = \frac{\sin(\alpha_y)}{\lambda}$$
 4.2.4

Wielkości te nazywane są częstościami przestrzennymi i mają wymiar odwrotności długości (def. TIX 4.3.3), co tłumaczy użycie nazwy "częstość"; przez analogię do częstości czasowej mającej wymiar jeden nad czas.

$$[częstość przestrzenna] = \frac{1}{[długość]}$$
4.2.5

Dla małych kątów α mamy

$$f_x = \frac{\sin(\alpha_x)}{\lambda} \approx \frac{\tan(\alpha_x)}{\lambda} \approx \frac{\alpha_x}{\lambda}; \quad f_y = \frac{\sin(\alpha_y)}{\lambda} \approx \frac{\tan(\alpha_y)}{\lambda} \approx \frac{\alpha_y}{\lambda}$$
 4.2.6

Ze względu na użyteczność, częstości przestrzenne, dla małych kątów, są definiowane również tak

$$f_x \approx \frac{\tan(\alpha_x)}{\lambda} = \frac{x_o}{\lambda z_o}; \ f_y \approx \frac{\tan(\alpha_y)}{\lambda} = \frac{y_o}{\lambda z_o}$$
 4.2.7



Rysunek 4.2.1. Układ kątów, dla których zdefiniowane są częstości przestrzenne.

Zapiszę całkę dalekiego pola (4.2.3) z użyciem częstości przestrzennych

$$u(P_{o}) = \Omega \iint_{\Sigma} u_{pad}(x, y) \exp\{-2\pi i \left(xf_{x} + yf_{y}\right)\} dxdy$$

$$4.2.8$$

Wzór powyższy ma postać transformaty Fouriera, pod warunkiem jednak, że rozciągniemy granicę całkowania na całą płaszczyznę

$$U(f_x, f_y) = \Omega \iint_{-\infty}^{\infty} u_{\text{pad}}(x, y) \exp\{-2\pi i (xf_x + yf_y)\} dx dy$$

$$= \Omega \mathcal{F} \left(u_{\text{pad}}(x, y) \right)$$

$$4.2.9$$

Aby to było możliwe granic całkowania musi "pilnować" funkcja u_{pad} . Jak zobaczymy poniżej nie stanowi to żadnego problemu. Zauważ również, że funkcja amplitudy zespolonej u stała się funkcją częstości przestrzennych U.

Inaczej wyrażenie (4.2.9) nie byłoby transformatą Fouriera. Wzory (4.2.7) pozwalają na prosty powrót do współrzędnych punktu obserwacjo P_0 .

Dlaczego tak usilnie dążyłem do postaci transformaty Fouriera (4.2.9). Być może już się domyślasz, że chodzi o metodę wiadra (§TI 4). Sprowadzenie wyrażenie na całkę dyfrakcyjną do bardzo dobrze znanego wyrażenia matematycznego jest prawdziwym darem niebios i byłoby głupotą nieskorzystanie z tej sposobności. Mając postać (4.2.9) możemy od ręki wykorzystać całą bogatą teorię transformat Fouriera oraz mnogość procedur numerycznych do ich obliczania. Na podstawie tej wiedzy możemy również dużo powiedzieć o samej fizyce dyfrakcji światła. Nasz zapał może zostać schłodzony faktem, że przybliżenie (4.2.1) prowadzące do transformaty Fouriera działa tylko dla małych kątów. Ale tu niebiosa okazały się również łaskawe. Zakres zastosowań przybliżenia dalekiego pola jest zaskakująco duży.

Pokażę, że transformatę Fouriera można traktować jako rozkład zaburzenia na fale płaskie. Oznacza to, że w przybliżeniu dalekiego pola zastępujemy kuliste fale wtórne, stycznymi do nich falami płaskimi. Zacznę od pokazania, że czynnik transformujący (wzór (4.2.9)) można interpretować jako opisujący falę płaską rozchodzącą się w kierunku wskazywanym przez częstości przestrzenne f_x i f_y .

$$\exp\{-2\pi i \left(xf_x + yf_y\right)\}$$
4.2.10

W tym celu przekształcę ten czynnik do postaci

$$\exp\{-2\pi i (xf_x + yf_y)\} \longrightarrow \exp\{-i\frac{2\pi}{\lambda} \left(x\frac{x_o}{z_o} + y\frac{y_o}{z_o}\right)\}$$

$$\rightarrow \exp\{-ik \left(x\frac{x_o}{r_o} + y\frac{y_o}{r_o}\right)\}$$

$$4.2.11$$

W ostatnim kroku wycofałem się z przybliżenia $r_o \rightarrow z_o$ (4.4a), Korzystając z rysunku (4.2.1) mamy

$$\exp\left\{-ik\left(x\sin(\alpha_x) + y\sin(\alpha_y)\right)\right\}$$

= $\exp\left\{-i\left(xk\cos(\beta_x) + yk\cos(\beta_y)\right)\right\}$
= $\exp\left\{-i\left(xk_x + yk_y\right)\right\}$
4.2.12

Ostatnie wyrażenie wygląda już jak wzór na falę płaską, tyle że brakuje nam składowej z k_z . Możemy ją jednak dopisać

$$\exp\{-i(x k_{x} + y k_{y} + zk_{z})\} = \exp\{-i(x k_{x} + y k_{y})\}\exp\{-i zk_{z}\}$$
4.2.13

I mamy wzór na falę płaską o jednostkowej amplitudzie.

Czy jednak wolno tak beztrosko dopisywać brakujący czynnik? Zwykle nie, ale w tym wypadku jest to uzasadnione. Dopisany czynnik już w całce dalekiego pola siedzi w przedcałkowy czynniku Ω (zobacz wzór (4.1.3)).

Widać teraz jak możemy intepretować wzór (4.2.9). Wybieramy punkt, dla którego liczymy wartość funkcji $U(f_x, f_y)$. W tym przypadku wybranie punktu oznacza wybranie konkretnej wartości częstości przestrzennych f_x , f_y . Zgodnie ze wzorem (4.2.9) wartość funkcji $U(f_x, f_y)$ jest sumą fal płaskich rozchodzących się z każdego punktu przedmiotu (całkowanie po x i y, to jest po całej płaszczyźnie przedmiotu), pod kątem określonym przez dane częstości f_x , f_y , z amplitudą zespoloną równą u_{pad}(x,y). Zatem każda fala płaska dziedziczy fazę i amplitudę po fali padającej. Przypomnij sobie, że w przybliżeniu bliskiego pola zamienialiśmy fale kuliste na paraboliczne, a każda fala paraboliczna dziedziczyła amplitudę i fazę początkową po fali padającej (komentarz pod def. 4.1.1).

4.3. Funkcja transmitancji

Do tej pory zajmowaliśmy się zagadnieniem dyfrakcji na płaskich otworach. To wąska klasa obiektów. Jednak prosta metoda pozwoli nam na użycie całek dyfrakcyjnych dla znacznie szerszej klasy układów. Rozważmy jakiś dowolny element płaski. Może to być na przykład czarno-białe przezrocze. Przezrocze zwykle zamknięte jest w ramce, której brzegi nie przepuszczają światła. Samo przezrocze w różnych punktach przepuszcza światło w różnym stopniu. Możemy opisać je funkcją przepuszczalności t(x, y) (nazywaną funkcją transmitancji). W punktach gdzie przezrocze jest całkowicie przezroczyste funkcja ta przyjmuje wartość jeden. W punktach gdzie światło jest całkowicie tłumione funkcja transmitancji przyjmuje wartość zero. W pozostałych punktach przezrocza natężenie światła jest tłumione tylko częściowo i tam funkcja transmitancji przyjmuje wartość między zero a jeden. Dla otworu funkcja transmitancji ma prostą konstrukcję. W punktach należących do otworu przyjmuje wartość jeden, a w każdym innym punkcie przyjmuje wartość zero (rys. 4.3.1).

Funkcja transmitancji pozwala opisać działanie przezrocza na falę padającą u_{pad}. Fala padająca po przejściu przez przezrocze ulega modyfikacji w swej części amplitudowej. Oznacza to, że geometria powierzchni falowej fali padającej nie ulega zmianie. Jednak rozkład energii na powierzchni falowej, na wyjściu z przezrocza jest inny niż przed wejściem na przezrocze. Za pomocą funkcji transmitancji t zapisujemy to tak

$$u_{\text{wych}}(x, y, z = z_p) = t(x, y)u_{\text{pad}}(x, y, z_p)$$

$$4.3.1$$

Funkcja opisująca falę padającą zależy oczywiście od zmiennych x,y,z. We wzorze powyższym przyjęliśmy, że ekran znajduje się w położeniu z_p . Zwykle przyjmuje się, że $z_p=0$, gdyż taki wybór początków układu współrzędnych

upraszcza obliczenia. Ja również będę przyjmował, że $z_p=0$ i wzór powyższy będę pisał w postaci

 $u_{\text{wych}}(x, y) = t(x, y)u_{\text{pad}}(x, y)$



Rysunek 4.3.1. Dla otworu funkcja transmitancji jest równa jeden dla punktów leżących wewnątrz otworu a zero dla punktów znajdujących się na zewnątrz otworu.

Jeżeli wykorzystamy technikę bielenia⁵, to nasze przezrocze stanie się praktycznie przezroczyste. Zmiany stopnia zaczernienia zostają zamienione na zmiany wartości współczynnik załamania. Światło przechodząc przez takie przezrocze nie zmienia swojej amplitudy ale zmienia swoją fazę. Ale żeby światło zmieniło swoją fazę przezrocze nie może być nieskończenie cienkie (dwuwymiarowe). W praktyce żadne przezrocze nie jest nieskończenie cienkie, a jego grubość choć niewielka, zwykle wielokrotnie przekracza wartość długości fali świetlnej. Na rysunku (4.3.2a) odbielone przezrocze pokazane jest w postaci szeregu kolorowych prostokątów. Każdy kolor oznacza inną wartość współczynnika załamania. Ponieważ przezrocze ma niezerową grubość w każdym obszarze fala przechodząca przez przezrocze dozna innego przesunięcie fazowego równego grubości przezrocza przemnożonego przez współczynnik załamania w danym obszarze razy wartość wektora falowego. W efekcie zmieni się geometria powierzchni falowej.

W wielu wypadkach, z dobrym przybliżeniem, możemy opisać odbielone przezrocze jako cienkie. Dzieje się tak wtedy, kiedy możemy uznać, że promień padający na przezrocze znajduje się na tej same wysokości co promień wychodzący z przezrocza. Warunek ten będzie spełniony, wtedy gdy promienie będą padały na przezrocze pod małymi kątami i wtedy gdy samo przezrocze będzie cienkie a takie właśnie są przezrocza (kierunki promieni wyznaczone są przez normlane do powierzchni falowej). Na rysunku (4.3.2b) wybrany promień

⁵ Technika ta pozwala na zmianę stopnia zaczernienia pewnych rodzajów klisz na zmiany współczynnika załamania światła. Klisza staje się przezroczysta, a światło przechodzące przez różne jej punkty doznaje różnych przesunięć fazowych.

padający na przezrocze wychodzi prawie na tej samej wysokości (prawie to samo y) co pada, takie przezrocze możemy traktować jako dwuwymiarowe.



Rysunek 4.2.2. a) przy przejściu przez przezroczystą płytkę o zmiennej grubość lub zmiennym współczynniku załamania zmienia się faza fali padającej (w każdym punkcie inaczej) a przez to również kształt frontu falowego. Na rysunku różne kolory odpowiadają różnym przesunięciom fazowym przy przejściu promieni po liniach prostopadłych do powierzchni płytki; b) układ jest cienki (dla danej fali) kiedy możemy przyjąć, że promień pada na pierwszą powierzchnie na takiej samej wysokości jak wychodzi z drugiej powierzchni (niebieski odcinek).

Uogólnimy tą definicję na dowolny układ optyczny. Zacznę od tego, że teraz wszystko to co przekształca falę świetlną będę nazywał układem optycznym. Przezrocze będzie dla mnie układem optycznym tak samo jak otwór czy soczewka.

Definicja 4.3.1: Cienki układ optyczny

Układ optyczny nazywamy cienkim, wtedy gdy możemy uznać, że promienie padają na ten układ na tej samej wysokości na której z niego wychodzą

Uwaga 4.3.1.

Należy pamiętać, że o tym czy układ optyczny możemy traktować jako cienki czy nie decyduje nie tylko jego fizyczna grubość ale największe wartości kątów promieni padających na ten układ.

Teraz możemy napisać funkcję opisującą przesunięcie fazy fali padających na nasze odbielone przezrocze.

 $u_{\text{wych}}(x, y) = u_{\text{pad}}(x, y) \exp\{i\psi(x, y)\}$ 4.3.2

 $\psi(x, y)$ jest tu oczywiście przesunięciem fazowym promieni padających na punkt o współrzędnych x,y i wychodzących w punkcie o tych samych współrzędnych (bo stosujemy przybliżenie cienkiego układu). Ponieważ przezrocze jest cienkie więc możemy uznać, że dla każdego promienia przesunięcie fazowe ψ równe jest grubości przezrocza *h* razy współczynnik załamania *n* przezrocze w punkcie padania promienia (rys. 4.3.3) (zobacz też rys. 2.2).



Rysunek 4.3.3. Dla cienkiego przedmiotu zmiana fazy jest proporcjonalna do grubości tego przedmiotu i jego współczynnika załamania.

Zatem, w ogólnym przypadku przezrocze może zmieniać fazę i amplitudę padającej fali. W tym ogólnym wypadku do opisania działania przezrocza na falę padającą wykorzystamy funkcję transmitancji przezrocza, którą złożymy ze zdefiniowanych wyżej części.

Definicja 4.3.2: Funkcja transmitancji układu optycznego

Funkcja transmitancji t(x,y) cienkiego układu optycznego opisuje zmianę amplitudy i fazy fali padającej po przejściu przez ten układ i ma postać

 $t(x, y) = a_t(x, y) \exp\{i\psi(x, y)\}$ 4.3.3

Popatrzmy jak działa funkcja transmitancji t(x, y) na falę padającą u_{pad}

 $u_{pad}(x, y) = a(x, y) \exp\{i\varphi(x, y)\}$

$$u_{\text{wych}}(x, y) = u_{\text{pad}}(x, y)t(x, y)$$

= $a(x, y)a_t(x, y)\exp\{i(\varphi(x, y) + \psi(x, y))\}$ 4.3.4

Całość możemy przedstawić na rysunku (4.2.4). Zauważ również, że funkcja transmitancji przedmiotu "pilnuje" granic całkowania. Przechodząc ze wzoru (4.2.8) na całkę dalekiego pola do postaci (4.2.9) tej całki rozszerzyłem granice całkowania na całą płaszczyznę, dodając zastrzeżenie, że granic całkowania musi "strzec" wyrażenie podcałkowe. Gdy opisujemy przedmiot przez funkcją transmitancji to mamy takiego strażnika. Tam gdzie przedmiot się kończy (na przykład mamy ramkę przeźrocza) funkcja transmitancji przyjmuje wartość zero i wyrażenie podcałkowe staje się równe zeru.

4.3.4.a



Rysunek 4.3.4. Transformacja fali padającej na przedmiot opisany dwuwymiarową funkcją transmitancji, w reprezentacji fazorowej. Działanie przedmiotu na fazory reprezentujące falę padająca pokazane jest w dwóch wybranych punktach P₁ i P₂. Działanie to można rozbić na dwa niezależne poddziałania: pierwsze z nich to zmiana długości fazora (czynnik $a_t(x,y)$), drugie z nich to zmiana fazy fazora (czynnik ($exp{i\psi(x,y)}$)) reprezentujących falę padającą w dwóch punktach przedmiotu: P₁ i P₂.

Poniżej podam krótki schemat postępowania jaki wyłania się z dotychczasowych rozważań. Zakładamy, że cienki układ optyczny leży w początku układu współrzędnych w płaszczyźnie x,y. Aby obliczyć efekt przejścia fali świetlnej przez ten układ musimy

Schemat 4.3.1: Obliczanie zagadnień dyfrakcyjnych

- Musimy znać amplitudę zespoloną dla fali padającej u_{pad}(x,y), w płaszczyźnie układu optycznego.
- Musimy znać funkcję transmitancji układu optycznego t(x,y)
- Określamy falę wychodzącą z układu uwych wzorem uwych(x,y) = upad(x,y) t(x,y)
- Aby obliczyć rozkład amplitudy fali na ekranie u_e(x_e, y_e) musimy odwołać się do jednej z całek dyfrakcyjnych (bliskiego lub dalekiego pola). Całkujemy wielkość u_{wych}.
- Aby obliczyć natężenie światła na ekranie obliczamy kwadrat modułu u_e : i = $|u_e|^2$.

Tu potrzebne jest kilka dodatkowych uwag. Poznane przez nas całki dyfrakcyjne mają postać:

$$u(x_o, y_o) = \iint_{-\infty}^{\infty} u_{\text{wych}}(x, y) K(x, y) \, dx \, dy$$

$$4.3.5$$

To czym różnią się całki dyfrakcyjne między sobą to czynnik transformujący K. Na przykład dla całki Fresnela czynnik transformujący ma postać

$$K(x, y) = \exp\left\{i\frac{k}{2z_o}([(x - x_o)^2 + (y - y_o)^2])\right\}$$
4.3.6

Znamy dwie całki dyfrakcyjne: Fresnela i Fraunhofera. Niemniej wyrażeń całkowych stosowanych do opisu dyfrakcji światła jest więcej. Każda z nich ma swoje wady i zalety i może być skutecznie stosowana w ściśle określonych okolicznościach. Wszystkie one są przybliżonym rozwiązaniem równań Maxwella, które wyznaczają ogólne ramy teorii dyfrakcji.

4.3.1. Zasada Babineta

Przy rozwiązywaniu problemów dyfrakcyjnych użyteczne jest zasada Babineta. Do jej sformułowania potrzebujemy pojęcia przesłon komplementarnych. Są to takie dwie przesłony, których fizyczne złożenie daje całkowicie nieprzezroczysty ekran.

Definicja 4.3.3: Przesłony komplementarne

Dwie przesłony nazywamy komplementarnymi, kiedy przepuszczająca części jednej z nich odpowiada dokładnie przesłaniającej części drugiej przesłony.

Przykładem przesłon komplementarnych są otwór kołowy i krążek o średnicy równej średnicy otworu. Widać z tego, że jeżeli złożymy otwartą część przesłon komplementarnych to dostaniemy wolną przestrzeń. Niech u₁ i u₂ będą obrazami dyfrakcyjnymi przesłony pierwszej i drugiej. Amplitudy zespolone u₁ i u₂ obliczmy ze wzorów

$$u_1 = \Im\{u_{pad}t_1\}$$

$$4.3.7a$$

$$4.3.7b$$

$$u_2 = \Im\left\{u_{pad}t_2\right\}$$
 4.3.7b

gdzie symbol \Im oznacza całkę dyfrakcyjną bliskiego lub dalekiego pola, t₁ i t₂ to funkcje transmitancji przesłon komplementarnych, u_{pad} to fala padająca. Z drugiej strony mamy

$$u_{1} + u_{2} = \Im\{u_{pad}t_{1}\} + \Im\{u_{pad}t_{2}\} = \Im\{u_{pad}(t_{1} + t_{2})\} = \Im\{u_{pad}\}$$
4.3.8

Gdzie skorzystałem z liniowości całek. Suma komplementarnych otworów daje całą przestrzeń, stąd ostatnie przejście w (4.3.8). Zauważ, że dla przesłon komplementarnych mamy

$$t_c = t_1 - (1 - t_2) \tag{4.3.9}$$

Gdzie t_c jest funkcją transmitancji przesłony całkowitej (to jest takiej, które nie przepuszcza światła w całej płaszczyźnie). Dla tej przesłony mamy

$$\Im\{t_c\} = \Im\{t_1\} - \Im\{1\} + \Im\{t_2\} = 0$$

$$4.3.10$$

Wyraz $\Im\{1\}$ reprezentuje wiązkę nieugiętą. Widać z tego, że poza obszarem, w którym wiązka nieugięta wnosi znaczący wkład mamy

$$\Im\{t_1\} + \Im\{t_2\} = 0 \Longrightarrow \Im\{t_1\} = -\Im\{t_2\} = \Im\{t_2\} e^{i\pi}$$

$$4.3.11$$

Oznacza to, że w tym obszarze obrazy przesłon dopełniających różnią się tylko przesunięcie fazowym o π . Wynik ten nosi nazwę zasady Babineta. Zasada Babineta uwidacznia się szczególnie dobrze w obrazach dalekiego pola, kiedy to

mamy szeroką część dyfrakcyjną i stosunkowo wąską wiązkę nieugiętą. W obszarze zdominowanym przez dyfrakcję obrazy przesłon złożonej z otworków są takie same jak obraz tej samej przesłony, dla której otworki zamieniono na układ pasujących do nich zatyczek, a pozostała część przesłony jest wolna. Przykładem takiego układy jest szczelina i pasujący do nie pasek nieprzezroczystego materiału. W obszarze dyfrakcji obrazy obu tych przesłon są takie same z dokładności do czynnika fazowego, którego jednak w obrazie natężeniowym nie widzimy.

4.4. Otwór prostokątny oświetlony falą płaską

Jako przykład zastosowania całki dyfrakcyjnej Fraunhofera obliczę obraz szczeliny prostokątnej oświetlonej poosiową falą płaską o amplitudzie *a* (rys. 4.4.1). Aby to zrobić postąpię według schematu (4.3.1). Fala padająca, to poosiowa fala płaska, która jest opisana wzorem (TIX 1.1.2)

$$u_{pad} = a \exp\{ikz\}$$

Funkcja transmitancji prostokątnej szczeliny ma postać

$$t(x,y) = \Pi\left(\frac{x}{a}\right) \,\Pi\left(\frac{y}{b}\right) \tag{4.4.2}$$

Funkcje prostokątne Π zdefiniowane są w (Dx xx). Fala wychodząca opisana jest wzorem

$$u_{\text{wych}} = u_{\text{pad}}t = a \exp\{ikz\} \prod\left(\frac{x}{2a}\right) \prod\left(\frac{y}{2b}\right)$$
 4.4.3

Wyrażenie to podstawiamy do całki dyfrakcyjnej dalekiego pola (4.2.3)

$$U(f_x, f_y) = \Omega \iint_{-\infty}^{\infty} \frac{\exp\{ikz\} \prod \left(\frac{x}{2a}\right) \prod \left(\frac{y}{2b}\right)}{\exp\{-2\pi i \left(x f_x + y f_y\right)\} dx dy}$$

$$4.4.4$$

Wyłączamy czynniki nie zawierające zmiennych całkowania

$$U(f_x, f_y) = \Omega \exp\{ikz\}$$

$$\iint_{-\infty} \Pi\left(\frac{x}{2a}\right) \Pi\left(\frac{y}{2b}\right) \exp\{-2\pi i \left(x f_x + y f_y\right)\} dx dy$$
4.4.5

Powyższa całka da się przedstawić w postaci iloczynu całek (da się sfaktoryzować)

$$\int_{-\infty}^{\infty} \Pi\left(\frac{x}{2a}\right) \exp\{-2\pi i (x f_x)\} dx \int_{-\infty}^{\infty} \Pi\left(\frac{y}{2b}\right) \exp\{-2\pi i (y f_y)\} dy \qquad 4.4.6$$

Obliczymy pierwszy składnik tego iloczynu. Funkcja prostokątna Π jedynie ogranicza nam obszar całkowania, możemy więc wprowadzić ją do granic całki.

4.4.1



Rysunek 4.4.1. Szczelina prostokątna o bokach 2a i 2b oświetlona jest monochromatyczną, poosiową falą płaską. Szukamy jej obrazu na ekranie umieszczonym w odległości z_0 , przy czym zakładamy, że odległość szczelina ekran jest dostatecznie duża aby można było zastosować przybliżenie dalekiego pola.

Obliczamy całki z poszczególnych składników tej całki

$$\int_{-a}^{a} \cos(2\pi x f_{x}) dx = \frac{1}{\pi f_{x}} \sin(2\pi a f_{x})$$

$$\int_{-a}^{a} \sin(2\pi x f_{x}) dx = 0$$
4.4.8

Stad mamy

$$\int_{-\infty}^{\infty} \Pi\left(\frac{x}{2a}\right) \exp\{-2\pi i x f_x\} dx = \frac{1}{\pi f_x} \sin(2\pi a f_x)$$

$$4.4.9$$

Wynik ten zapiszemy wykorzystując funkcję sinc (Dx xx)

$$\int_{-\infty}^{\infty} \Pi\left(\frac{x}{2a}\right) \exp\{-2\pi i x f_x\} dx = 2a \operatorname{sinc}(2af_x)$$

$$4.4.10$$

W podobny sposób obliczamy drugą całkę iloczynu

$$\int_{-\infty}^{\infty} \Pi\left(\frac{y}{2b}\right) \exp\{-2\pi i x f_y\} dy = 2b \operatorname{sinc}(2bf_y)$$

$$4.4.11$$

Po złożeniu całek mamy wynik

$$u(f_x, f_y) = \Omega \exp\{ikz\} 2a \ 2b \ \operatorname{sinc}(2af_x) \ \operatorname{sinc}(2bf_y)$$

$$4.4.12$$

Zwykle pomijamy stałe stojące przed całką

$u(f_x, f_y) \sim \operatorname{sinc}(2af_x) \operatorname{sinc}(2bf_y)$

Kilka uwag dotyczących otrzymanego rozwiązania. Po pierwsze, pomijając stałe stojącymi przed całkami, otrzymane wyrażenie jest wyrażeniem rzeczywistym, co jest pochodną faktu, że otwór nie zmienia fazy fali padającej. Po drugie otrzymaliśmy zależności od częstości przestrzennych (transformatę Fouriera) i jeżeli interesuje nas wartość dla konkretnego punkt na konkretnym ekranie to musimy, korzystając ze wzoru na częstości przestrzenne (4.2.7), przejść do współrzędnych przestrzennych. Musimy znać jeszcze odległość między ekranem a otworem. Po trzecie obliczyliśmy amplitudę zespoloną obrazu. Aby obliczyć rozkład natężenia musimy obliczyć kwadrat modułu tego wyrażenia

$$I(f_x, f_y) \sim \text{sinc}^2(2af_x) \text{sinc}^2(2bf_y)$$
 4.4.14

Rysunek (4.4.2) pokazuje przykładowy rozkład natężenia światła.



Rysunek **4.4.2.** (a) przedstawia rozkład obliczonego natężenia światła. Zwracam uwagę, że na osiach odłożone są częstości przestrzenne w jednostkach [1/mm]. Widać, że w centrum obrazu mamy silne maksimum, a po jego bokach występuję drobne pofałdowania. Aby zobaczyć strukturę tych pofałdowań, na rysunku (b) zmniejszyłem zakres wyrysowanych wartości. Wszystkie wartości I>0,01 zostały obcięte. Dzięki temu dobrze widoczne są boczne maksima. Szczelina jest kwadratem o boku 1x1mm.

4.4.13

Rysunek (4.4.3) Przedstawia otrzymany wynik w przekroju. Rysunek (4.4.4) przedstawia rozkład natężenia światła dla szczeliny prostokątnej. Widać, że natężenie światła ulega większemu rozmyciu dyfrakcyjnemu w kierunku osi f_y , w którym szczelina jest węższa. Jest to charakterystyczne dla obrazów w dalekim polu i od strony formalnej wynika z twierdzenia o podobieństwie dla transformat Fouriera (tw. TIX 4.3.3)



Rysunek 4.4.3. Przekrój wzdłuż osi f_y =0 wykresu z rysunku (4.3.2). Dokładnie widać różnicę natężenia pomiędzy maksimum głównym a maksimami bocznymi.



Rysunek 4.4.4. Obraz dyfrakcyjny na szczelinie prostokątnej o wymiarach a=2mm i b=0,5mm. Zauważ, że wzdłuż osi f_y odległości między maksimami bocznymi są większe niż wzdłuż osi f_x . Ale szczelina jest szersza wzdłuż osi x.

4. 5. Otwór kołowy oświetlony falą płaską

W układach optycznych najczęściej spotykamy się z kołowymi otworami. I z takim przypadkiem musimy się zmierzyć. Funkcję transmitancji otworu kołowego opiszemy funkcją circ (Dx xx)

$$t(r) = \operatorname{circ}\left(\frac{r}{a}\right)$$

$$4.5.1$$

Przy oświetleniu falą płaską poosiową obraz w dalekim polu będzie opisany wzorem

$$u(f_x, f_y) = \Omega \exp\{ikz\} \iint_{-\infty}^{\infty} \operatorname{circ}\left(\frac{r}{a}\right) \exp\{-2\pi i (x f_x + y f_y)\} dx dy \quad 4.5.2$$

Przy symetrii kołowej wygodnie jest przejść do układu biegunowego (§TIV 5.1.1).

$$x = \rho \cos(\varphi)$$
 4.5.2a

$$y = \rho \sin(\varphi) \tag{4.5.2b}$$

$$dx \, dy = \rho d\varphi dr \tag{4.5.2c}$$

$$f_x = \rho_f \cos(\theta) \tag{4.5.2d}$$

$$f_y = \rho_f \sin(\theta) \tag{4.5.2e}$$

Wyrażenie na obraz w dalekim polu przyjmie postać

$$u(f_x, f_y) = \overline{\Omega} \int_{0}^{\infty} \int_{0}^{2\pi} \operatorname{circ}\left(\frac{r}{a}\right) \rho \exp\{-2\pi i \left(\rho \rho_f \cos(\varphi)\cos(\theta) + \rho_f \sin(\varphi)\sin(\theta)\right)\} d\varphi dr$$

$$4.5.3$$

Korzystając ze wzorów redukcyjnych dla funkcji trygonometrycznych wyrażenie pod funkcją exp możemy zwinąć do postaci

$$u(f_x, f_y) = \overline{\Omega} \int_{0}^{\infty} \int_{0}^{2\pi} \operatorname{circ}\left(\frac{r}{a}\right) \rho \exp\left\{\frac{-2\pi i\rho \rho_f}{\cos(\varphi - \theta)}\right\} d\varphi \, dr \qquad 4.5.4$$

Otrzymane całki są trudne. Ale spróbujmy ją rozwiązać. Krok pierwszy polega na pozbyciu się funkcji circ. Jedyne co robi ta funkcja to obcięcie granic całkowania do obszaru otworu. Zatem możemy się jej pozbyć pod warunkiem, że zmienimy odpowiednio granice całkowania. Promień naszego otworu wynosi *a*, zatem mamy.

$$u(\rho_f, \theta) = \overline{\Omega} \int_{0}^{a} \int_{0}^{2\pi} \rho \exp\{-2\pi i\rho \rho_f \cos(\varphi - \theta)\} d\varphi dr$$

$$4.5.5$$

Policzmy najpierw całkę po φ .

$$\int_{0}^{2\pi} \exp\{-2\pi i\rho \,\rho_f \cos(\varphi - \theta)\} \mathrm{d}\varphi$$

$$4.5.6$$

Najpierw mała zamiana zmiennych

$$\psi = \varphi - \theta \rightarrow \varphi = \psi + \theta \rightarrow d\varphi = d\psi$$

Całka przyjmie postać

$$\int_{0}^{2\pi} \exp\{-2\pi i\rho \rho_{f} \cos(\psi)\} d\psi$$

$$4.5.7$$

Dziś zanim ktokolwiek powie, że nie daje rady obliczyć całki musi wpierw zobaczyć czy jej nie znajdzie w tablicach całek, lub czy nie da się jej obliczyć programem całkującym. Ja ją policzę pod programem Mathematica. Wynik zaznaczyłem na żółto (M.4.5.1). Reszta to informacja o tym, że wynik jest prawidłowy tylko wtedy gdy iloczyn $\rho \rho_f$ jest liczbą rzeczywistą (zaznaczone na niebiesko) oraz o tym, że gdy tak nie jest to program nie potrafi obliczyć całki (zaznaczenie na zielono). Ponieważ niebieski warunek jest w naszym przypadku spełniony zajmijmy się wynikiem, czyli częścią żółtą.

Integrate[Exp[$-2\pi I\rho \rho f Cos[\psi]$], { ψ , 0,2 π }]	Instrukcja całowania
If $[\rho \ \rho f] \in \text{Reals}, 2\pi \text{BesselJ}[0, 2\pi\rho\rho f], \text{Integrate}[e^{-2i\pi\rho\rho f \text{Cos}[\psi]}],$	Wynik
$\{\psi, 0, 2\pi\}$, Assumptions $\rightarrow \rho \rho f \notin \text{Reals}$	
M. 4.5.1. Obliczenie całki (4.5.7) w pakiecie Mathematica	

Widzimy w nim funkcję specjalną, tzw. funkcję Bessla zerowego rzędu (Dx xx). Zwykle funkcje Bessla oznaczamy symbolem $J_n(\rho)$, gdzie *n* oznacza tzw. rząd funkcji Bessla. Czyli w takich zwykłych oznaczeniach uzyskany w programie Mathematica wynik zapiszemy w postaci

 $2\pi \text{BesselJ}[0, 2\pi\rho\rho_f] = 2\pi J_0(2\pi\rho\rho_f)$

4.5.8

Wstawiamy go do wyjściowego wyrażenia i mamy

$$u(\rho_f, \theta) = \int_0^a \rho \ 2\pi J_0(2\pi\rho\rho_f) d\rho \qquad 4.5.9$$

Porządkując nieco sprawy mamy

$$u(\rho_f, \theta) = 2\pi \int_0^a \rho J_0(2\pi\rho\rho_f) d\rho \qquad 4.5.10$$

No właśnie, mamy całkowanie funkcji Bessla. Nie wygląda to ładnie. Ale zanim się załamiemy zapytamy o zdanie program Mathematica

Integrate[$2\pi\rho$ BesselJ[$0,2\pi\rho\rho$ f],{ $\rho,0,a$ }]	Instrukcja całkowania	
aBesselJ[1,2aπρf]	wynik	
ρf		
M. 4.5.2. Obliczenie całki (4.4.13) w pakiecie Mathematica		

Tłumacząc ten wynik na nasze mamy

$$\frac{a\text{BesselJ}[1,2a\pi\rho f]}{\rho f} = a \frac{J_1(2\pi a\rho_f)}{\rho_f}$$

$$4.5.11$$

Gdzie J_1 jest funkcją Bessla pierwszego rzędu. Czyż liczenie całek nie jest banalnie proste? Uzyskany wynik zapiszemy jeszcze w postaci

$$a\frac{J_{1}(2\pi a\rho_{f})}{\rho_{f}} = 2\pi a^{2} \frac{J_{1}(2\pi a\rho_{f})}{2\pi a\rho_{f}}$$

$$4.5.12$$

Przez analogię do funkcji sinc zdefiniuję funkcję somb (od angielskiego sombrero)

somb(r) =
$$2 \frac{J_1(\pi r)}{\pi r}$$
 4.5.13

Mnożenie przez 2 powoduje, że funkcja somb jest unormowana. Korzystając z funkcji somb mamy

$$u(\rho_f, \theta) = \underbrace{\pi a^2}_{S} \operatorname{somb}(2a\rho_f) = S \operatorname{somb}(2a\rho_f)$$

$$4.5.14$$

S oznacza oczywiście powierzchnię otworu. Wracając do współrzędnych prostokątnych mamy

$$u(f_x, f_y) = Ssomb\left(2a\sqrt{f_x^2 + f_y^2}\right)$$

$$4.5.15$$

Czas na zobrazowanie uzyskanego wyniku. Funkcja somb niewiele swym kształtem odbiega od funkcji Bessla (rys. 4.5.1)



Rysunek 4.5.1. Przebieg funkcji somb(x) i somb²(x) dla $x \in [-5,5]$. Widać, że podnoszenie do kwadratu znacznie redukuje wartość bocznych maksimów i zwęża maksimum główne.

Na rysunku (4.5.2) przedstawiony jest rozkład natężenia dla tego samego przykładu. Podnoszenie do kwadratu zwiększyło różnicę w wysokości między maksimum głównym a maksimami bocznymi.



Rysunek 4.5.2. Ten sam przykład co na poprzednim rysunku, tylko teraz obliczamy natężenie światła u². Maksimum główne jest szczuplejsze, a maksima boczne są na rysunku (a) ledwo widoczne. Lepiej pokazuje je rysunek (b), na którym maksimum główne zostało obcięte dla tej same wartości co na poprzednim rysunku.

Rysunek (4.5.3) pokazuje obraz dyfrakcyjny dla otworów o różnych średnicach



Rysunek 4.5.3. Natężenie u² dla otworów o promieniu kolejno r=1mm, r=0,5mm i r=0,25mm. Skale są takie jak na poprzednim rysunku. Widać, że zamykanie otworu powoduje większe rozmycie obrazu oraz spadek maksymalnego natężenia światła. To ostatnie jest zrozumiałe. Mniejszy otwór przepuszcza mniej światła. Większe rozmycie energii przy mniejszej średnicy otworu jest charakterystyczne dla obrazów w dalekim polu. Na rysunku (4.4.4) obraz jest również bardziej rozmyty w kierunku, w którym szczelina jest węższa. Większe rozmycie energii wpływa również na zmniejszenie wartości maksimum głównego.

4.6. Krótkie podsumowanie

W tym rozdziale dużo pracowałem na wyrażeniach przybliżonych. W fizyce stosowanej jest to chleb powszedni. Wyrażenia podstawowe dla danej teorii sa zwykle trudno obliczalne. Wystartowaliśmy od matematycznego zapisu zasady Huygensa-Fresnela (4.1). Pierwsza przybliżona wersja otrzymanego wzoru jest całka dyfrakcyjna Fresnela (4.1.4), a drugą całka dyfrakcyjna Fraunhofera (4.2.3). Całka dyfrakcyjna Fraunhofera ma wygodna postać transformaty Fouriera z fali przechodzącej przez przedmiot. Niestety, swoją rolę spełnia tylko w dużej odległości od przedmiotu. Istnieje jednak wystarczająco dużo praktycznych zagadnień, w których całka ta jest użyteczna. Zapis całki dyfrakcyjne Fraunhofera (dalekiego pola) w postaci transformaty Fouriera oznacza, że obraz obliczamy jako zależny od współrzędnych kątowych (4.2.7). Jeżeli coś zależy od kąta to znaczy, że skaluje się liniowo wraz z odległością. Inaczej mówiąc, jeżeli obserwujemy obraz w odległości z_o a potem przemieszczamy ekran do odległość m razy większej, to obraz powinien ulec mkrotnemu zwiększeniu, zachowując swoją wewnętrzną geometrię. Już z tego widać, gdzie całka dalekiego pola zawodzi. Podrozdział (2.1) dotyczący stref Fresnela pokazuje gdzie pojawią się kłopoty. Gdy liczba stref Fresnela, przy przesuwaniu się ekranu, zmienia się o jeden, to na osi maksimum natężenia światła zmienia się na minimum lub na odwrót. Zmiana położenia ekranu prowadzi do jakościowych zmian w obrazie; to nie jest obraz, który może być obliczony w przybliżeniu dalekiego pola. Aby to było możliwe, w obszarze przedmiotu musi się mieścić wyraźnie mniej niż jedna strefa Fresnela.

Inną cechą obrazu fourierowskiego jest odwrotne skalowanie rozmiarów poprzecznych przedmiotów. Wynika to z twierdzenia o podobieństwie

(tw. TIX 4.3.3). Gdy zmniejszamy rozmiary przedmiotu, to jego obraz rozmiary zwiększa (rys. 4.4.4). Nie jest to jednak cecha obrazów w bliskim polu.

Warto jeszcze wskazać na ciekawą cechę fali płaskiej poosiowej opisanej wzorem

 $u_{\rm p}(x, y, z) = A \exp\{ikz\}$ 4.6.1

W całkach dyfrakcyjnych bliskiego i dalekiego pola może ona zostać wyprowadzona przed znak całki, gdyż nie zależy od zmiennych całkowania (x,y). Co więcej, gdy obliczamy natężenie fali czynnik związany z falą płaską mnoży obliczoną całkę dyfrakcyjną przez liczbę A^2 . Oznacza to, że płaska fala poosiowa praktycznie nic od siebie nie dodaje do obserwowanego obrazu. To co widzimy jest efektem działania przedmiotu, opisanego funkcją transmitancji t(x,y). Słowem fala płaska poosiowa, to idealna sonda optyczna, która jest często w optyce wykorzystywana. Już fala płaska pochylona nie ma tej cechy, gdyż w opisującym ją wzorze mamy zależności od zmiennych całkowania. Zatem wynik całkowania zawiera spleconą informację i o przedmiocie i o fali. Aby mieć czystą informację o przedmiocie, trzeba umieć odpleść z wyniku informację o fali go oświetlającej. Jedynym wyjątkiem jest płaska fala poosiowa.

4.7. Siatka dyfrakcyjna

Wróćmy do kwestii siatki dyfrakcyjnej. Obliczyliśmy jej obraz, za pomocą fazorów, ograniczając się do reprezentowania poszczególnych szczelin przez ich punkty środkowe. Policzymy obraz siatki korzystając z przybliżenia dalekiego pola. Siatkę oraz układ współrzędnych, w którym to zrobimy pokazuje rysunek (4.7.1)



Rysunek 4.7.1. Analizujemy siatkę o stałej siatki *d*, szerokości szczelin *a*, wysokości szczelin *b*, liczbie szczelin *N*, opisaną w układzie współrzędnych jak na rysunku.

Funkcja transmitacji n-tej szczeliny ma postać

$$t_{n}(x,y) = \Pi\left(\frac{x-nd}{a}\right) \Pi\left(\frac{y}{b}\right)$$

$$4.7.1$$

Cała siatka to suma takich szczelin

$$t(x,y) = \sum_{n=0}^{N-1} \prod\left(\frac{x-nd}{a}\right) \prod\left(\frac{y}{b}\right) = \prod\left(\frac{y}{b}\right) \sum_{n=0}^{N-1} \prod\left(\frac{x-nd}{a}\right)$$

$$4.7.2$$

Zaczęliśmy indeksowanie szczelin od n=0, bo wtedy pierwsza szczelina, czyli ta dla której n=0, wypadnie w początku układu współrzędnych. Skoro pierwsza szczelina ma indeks n=0 to ostatnia ma indeks n=N-1.Przy oświetleniu falą płaską poosiową obraz w dalekim polu ma postać.

$$U(f_x, f_y) = \Omega \exp\{-i \, zk_z\} \exp\{ikz\} \iint_{-\infty}^{\infty} \Pi\left(\frac{y}{b}\right) \sum_{n=0}^{N-1} \Pi\left(\frac{x-nd}{a}\right)$$

$$4.7.3$$

Wyrażenie całkowe (4.7.3) możemy łatwo sfaktoryzować (przedstawić jako iloczyn całek pox i poy)

$$\iint_{-\infty}^{\infty} \Pi\left(\frac{y}{b}\right) \sum_{n=0}^{N-1} \Pi\left(\frac{x-nd}{a}\right) = \exp\{-2\pi i \left(x f_x + y f_y\right)\} dx dy$$

$$\int_{-\infty}^{\infty} \Pi\left(\frac{y}{b}\right) \exp\{-2\pi i y f_y\} dy \int_{-\infty}^{\infty} \sum_{n=0}^{N-1} \Pi\left(\frac{x-nd}{a}\right)$$

$$-\infty \exp\{-2\pi i x f_x\} dx$$
4.7.4

Pierwszą całkę już obliczaliśmy rozwiązując zagadnienie otworu prostokątnego (4.4.11)

$$\int_{-\infty}^{\infty} \Pi\left(\frac{y}{b}\right) \exp\{-2\pi i \, y f_y\} \mathrm{d}y = b \, \operatorname{sinc}(bf_y)$$

$$4.7.5$$

Korzystając z twierdzenia o liniowości transformaty Fouriera (tw. TIX 4.3.1) drugą całkę możemy zapisać w postaci sumy całek.

$$\sum_{n=0}^{N-1} \int_{-\infty}^{\infty} \Pi\left(\frac{x-nd}{a}\right) \exp\{-2\pi i x f_x\} dx$$

$$4.7.6$$

Wystarczy zatem, że policzymy całkę dla n=0, a potem dla kolejnych skorzystamy z twierdzenia o przesunięciu (tw. TIX 4.3.2). Dla n=0 mamy

$$\int_{-\infty}^{\infty} \Pi\left(\frac{x}{a}\right) \exp\{-2\pi i y f_y\} dx = a \operatorname{sinc}(a f_y)$$

$$4.7.7$$

Na mocy twierdzenia o przesunięciu mamy

$$\int_{-\infty}^{\infty} \Pi\left(\frac{x-nd}{a}\right) \exp\{-2\pi i x f_x\} dx$$

$$= \exp\{-2\pi i n df_x\} \int_{-\infty}^{\infty} \Pi\left(\frac{x}{a}\right) \exp\{-2\pi i x f_x\} dx$$

$$\underbrace{4.7.8}_{\text{sinc}(af_x)}$$

Zatem całe wyrażenie ma rozwiązanie

$$U(f_x, f_y) = \Omega \ ab \ \operatorname{sinc}(bf_y) \ \operatorname{sinc}(af_x) \sum_{i=0}^{N-1} \exp\{-2\pi i n df_x\}$$

$$4.7.9$$

Suma w tym wyrażeniu tworzy ciąg geometryczny o parametrach

wyraz pierwszy =
$$a_P = 1$$
 4./.10a

$$q = \exp\{-2\pi i df_x\}$$

$$4.7.10b$$

Suma *N* kolejnych wyrazów takiego ciągu wyraża się wzorem (choć suma we wzorze (3.7.9) kończy się na *N*-1, to wyrazów jest *N*, bo zaczynamy od *i*=0)

$$S_N = a_P \frac{1 - q^N}{1 - q} \tag{4.7.11}$$

Stąd mamy

$$U(f_x, f_y) = \Omega ab \operatorname{sinc}(bf_y) \operatorname{sinc}(af_x) \frac{1 - \exp\{-2\pi i N df_x\}}{1 - \exp\{-2\pi i df_x\}}$$

$$4.7.12$$

Ostatni ułamek możemy przekształcić stosując różne triki. Na przykład pomnożenie licznika i mianownika przez wyraz

$$\exp\{i\pi f_x Nd\}\exp\{i\pi f_x d\}$$

$$4.7.13$$

A potem rozwinięcie wyrazów $\exp\{i\varphi\}$ do postaci trygonometrycznej upraszcza całe wyrażenie. A co jeżeli nie masz wprawy w takich działaniach? Wtedy

1 7 10

możesz użyć dedykowanych programów komputerowych. Ja jak wiesz gonię do pracy pakiet Mathematica. Najpierw polecenie przekształć eksponenty na postać trygonometryczną

$\frac{1 - \exp[-2\pi I \text{fx}Nd]}{1 - \exp[-2\pi I \text{fx}d]} / / \text{ExpToTrig}$	Polecenia przekształcenia (4.7.12) do postaci
	trygonometrycznej
$\frac{1 - \cos[2dfxN\pi] + I \sin[2dfxN\pi]}{1 - \cos[2dfx\pi] + I \sin[2dfx\pi]}$	Wynik
$\frac{1 - \cos[2dfxN\pi] + ISin[2dfxN\pi]}{1 - \cos[2dfx\pi] + ISin[2dfx\pi]} //Simplify$	Polecenie uproszczenia
$Csc[dfx\pi](Cos[dfx(-1+N)\pi] - ISin[dfx(-1 + N)\pi])Sin[dfxN\pi]$	Wynik
$Cos[dfx(-1+N)\pi] - ISin[dfx(-1 + N)\pi]//TrigToExp$	Powrótdopostaciekspotencjalnejdlawyrażeniawokrągłymnawiasie
$e^{-idfx(-1+N)\pi}$	wynik
$e^{-idfx(-1+N)\pi}Csc[dfx\pi]Sin[dfxN\pi]$	Całość
M. 7.1. Przekształcenia wyrażenia (4.7.12)	

Co w tłumaczeniu na nasze oznaczenia daje wyrażenie

$$\frac{1 - \exp\{-2\pi i N df_x\}}{1 - \exp\{-2\pi i df_x\}} = e^{-i df_x(-1+N)\pi} \frac{\sin[\pi df_x N]}{\sin[\pi df_x]}$$
4.7.14

Całe wyrażenie (4.7.12) przyjmie postać

$$U(f_x, f_y) = \Omega e^{-idfx(-1+N)\pi}ab \operatorname{sinc}(bf_y) \operatorname{sinc}(af_x) \frac{\sin[\pi df_x N]}{\sin[\pi df_x]}$$
 4.7.15

A natężenie światła

$$I(f_x, f_y) = \Omega^2 a^2 b^2 \operatorname{sinc}^2(bf_y) \operatorname{sinc}^2(af_x) \frac{\sin^2[\pi df_x N]}{\sin^2[\pi df_x]}$$
 4.7.16

Jak widać w natężeniu światła nie ma wyrazu $e^{i\varphi}$. Z dokładnością do stałych wyrażenie (4.7.16) możemy zapisać tak

$$I(f_x, f_y) \propto \underbrace{\operatorname{sinc}^2(bf_y)\operatorname{sinc}^2(af_x)}_{\operatorname{szczelina}} \underbrace{\frac{\operatorname{sin}^2[\pi df_x N]}{\operatorname{sin}^2[\pi df_x]}}_{\operatorname{siatka}}$$

$$4.7.17$$

Wynik, który właśnie uzyskaliśmy składa się z dwóch znanych już nam wyrażeń. Pierwsze z nich to obraz dyfrakcyjny szczeliny (4.4.14), a drugi to obraz siatki obliczony przy założeniu, że siatkę można reprezentować przez zbiór punktów w środku każdej szczeliny (3.14). Przypomnę ten wzór

$$i = \frac{\sin^2\left(\frac{N \ k \ d \ \varphi}{2}\right)}{\sin^2\left(\frac{k \ d \ \varphi}{2}\right)} = \frac{\sin^2(N \ \pi \ d \ f_x)}{\sin^2(\pi d \ f_x)}$$

$$4.7.18$$

Skorzystałem tu z faktu, że $f_x = \alpha_x / \lambda = \varphi / \lambda$. Pierwszy człon nazywamy członem dyfrakcyjnym, bo opisuje dyfrakcję na pojedynczej szczelinie. Drugi człon to człon interferencyjny, bo opisuje interferencję fal wyemitowanych z *N* punktów. Rysunki (4.7.2 i 4.7.3) pokazują efekt działania tych dwóch członów.



Rysunek 4.7.2. Obraz dyfrakcyjny, dalekiego pola, dla siatki o parametrach: szerokość szczeliny a=1mm, stała siatki d=2mm, wysokość szczeliny d=2mm, ilość szczelin N=10. Z lewej strony wyrysowany jest rozkład natężenia światła dla pojedynczej szczeliny (czerwona linia), obraz interferencyjny (zielona linia) oraz obraz siatki (4.7.17) (czarna linia). Różnica w natężeniach światła powoduje, że rozkład od pojedynczej szczeliny jest praktycznie nie widoczny. Rysunek z prawej pokazuje obraz szczeliny we właściwej skali.



Rysunek 4.7.3. Aby sprawy lepiej uwidocznić możemy uzyskane obrazy częściowe unormować (normowanie nie zaburza kształtu krzywych). Rysunek z lewej pokazuje unormowany obraz siatki. Na rysunku z prawej nałożone są te same obrazy co na rysunku (4.7.2). Po unormowaniu widać jak rozkład natężenia światła od pojedynczej szczeliny (czerwony wykres) wpływa na natężenie światła w bocznych maksimach głównych obrazu interferencyjnego (zielony wykres). Należy pamiętać, że choć przebiegi natężenia we wszystkich trzech obrazach pozostają takie same ich rzeczywiste natężenia różnią się (rys. 4.7.2).

Rysunek (4.7.4) pokazuje obraz siatki dyfrakcyjnej dla innych jej parametrów.



Rysunek 4.7.4. Unormowany obraz członu interferencyjnego, dyfrakcyjnego i wzoru (4.7.18) dla siatki o parametrach: szerokość szczeliny a=0,5 stała siatki wynosi d=2, a ilość szczelin N=10. Mniejsza szerokość szczeliny, w porównaniu z poprzednim przykładem, powoduje większe rozmycie jej obrazu (czerwony wykres opada wolniej). W efekcie natężenie głównych maksimów bocznych maleje wolniej. Widać również, że drugie maksimum obrazu interferencyjnego jest całkowicie skasowane przez pierwsze minimum obrazu dyfrakcyjnego. Siatkę można tak sprofilować, że obraz dyfrakcyjny wygasi wszystkie maksima interferencyjne za wyjątkiem jednego. Taką siatkę nazywamy siatką profilowaną. Ma ona tą zaletę, że cała energia zostaje skierowana w jeden rząd dyfrakcji.

Uzyskaliśmy ciekawy wynik. Przejście światła przez siatkę daje się opisać jako iloczyn części interferencyjnej charakteryzującej rozkład szczelin, oraz części dyfrakcyjnej charakteryzującej obraz dyfrakcyjny pojedynczej szczeliny. Nasuwa się pytanie czy ten wynik można uogólnić. Powiedzmy, że mamy regularny rozkład otworów o tym samym kształcie (na przykład otwory trójkątne). Czy obraz dyfrakcyjny w dalekim polu można przedstawić jako iloczyn obrazu interferencyjnego charakteryzującego rozkład otworów i członu dyfrakcyjnego charakteryzującego obraz dyfrakcyjny pojedynczego otworu? Odpowiedź na to pytanie jest twierdząca. Nietrudno jest to pokazać, gdy się zna operację splotu, a to jeszcze przed nami.

4.8. Związek z koherencją

W (§TX IV) opisałem paczki falowe. Jak się okazuje nie możemy wygenerować czystego przebiegu harmonicznego, gdyż wymagałoby to nieskończenie długiego czasu pracy generatora fali. Każdy generator w pewnym momencie zaczyna emisję a w jakimś następnym ją kończy. W efekcie zamiast fali harmonicznej mamy paczkę falową, która jak wiemy z poprzednich rozważań oscyluje z określoną częstością, ale ma zanikającą amplitudę. Ma to swoje daleko idące konsekwencje. Rysunek (4.8.1a) pokazuje układ dwóch szczelin oświetlonych falą płaską. Wiemy teraz, że światło przychodzi w paczkach (w mechanice kwantowej te paczki staną się fotonami).



Rysunek 4.8.1. a) Dwie szczeliny oświetlone są falą płaską, która składa się z paczek falowych. Paczki ulegają podziałowi i interferują same z sobą. Dlaczego paczki interferują same z sobą a nie z innymi paczkami jest kwestią do której będę wracał. Teraz przyjmijmy, że tak jest. Różne drogi jakie mają do przebycia kopie każdej paczki do tego samego punktu na ekranie, zmieniają kontrast obserwowanych prążków, co lustruje część (b) rysunku.

Każda paczka zostaje rozdzielona na dwie ścieżki wyznaczone przez szczeliny. W płaszczyźnie rejestracji obie części paczki nakładają się. Widać jednak, że warunki nakładania się obu części paczki są różne w różnych częściach ekranu. Na środku paczki nakładają się bez żadnych względnych opóźnień. W górnej części paczka narysowana na czerwono wychodzi z górnej szczeliny i w płaszczyźnie ekranu wyprzedza paczkę narysowaną na niebiesko, która wychodzi z dolnej szczeliny. W dolnej części jest na odwrót. Efekty tych opóźnień mają wpływ na obraz interferencyjny, co wyjaśnia rysunek (4.8.1b). Gdy paczki nie opóźniają się jedna względem drugiej w każdym punkcie, gdzie się nakładają fazory reprezentujące te paczki mają tą samą długość. Gdy mamy opóźnienia jeden fazor jest krótszy od drugiego co skutkuje spadkiem kontrastu obrazu interferencyjnego aż do jego zaniku (zobacz też rys. (1.8)). To jak szybko spada kontrast zależy od długości paczki. Rysunek (4.8.2) ilustruje ten efekt.





Rysunek 4.8.2. Obliczony obraz prążków dla następujących parametrów: częstość paczki v_0 =4,74 10¹⁴Hz odległość między otworami *d*=3mm, odległość ekran-płyta z otworami *z*=1000mm, rozmiar obrazu - w pionie 2mm, w poziomie 6mm, obraz z lewej - długość koherencji światła 6,3µm, obraz z prawej – długość koherencji światła 25µm. Większa długość koherencji zapewnia lepszy kontrast prążków na brzegu rysunku.

Tylko co to jest ta długość paczki? Paczka ciągnie się do nieskończoności, tyle że im dalej od maksimum amplitudy tym mniejsza amplituda paczki. W pewnym momencie oscylacje paczki spadają do tak małej wartości, że nie ma to znaczenia (rys. TIX 4.2.2). Podam teraz użyteczną definicję długości paczki, uzasadnieniem tej definicji zajmę się w innym temacie.

Definicja 4.8.1. Długość koherencji (spójności) fal elektromagnetycznej

Długość koherencji paczki fal elektromagnetycznych przyjmuje się za równą c Δt , gdzie Δt to czas emisji impulsu fali, a c to prędkość fali.

Zagadnienie koherencji światła ma znaczenie w interferometrii. Ramiona interferometrów (np. rys. 1.1.1 i 1.1.4) powinny mieć jak najbardziej równą długość, tak aby dwie kopie paczki falowe nakładały się bez opóźnień. Opóźnienia będą skutkowały spadkiem kontrastu prążków interferencyjnych (1.18). Czułość interferometru na różnicę dróg obu ramion zależy od długości koherencji użytego światła. Dla przykładu źródła termiczne mają bardzo małą

drogę koherencji (kilka mikrometrów). Lampy gazowe mają długość koherencji od kilku centymetrów do nawet 50cm. Były one źródłem światła dla interferometrii przed wynalezieniem lasera. Tani laser gazowy helowo-neonowy (He-Ne) ma długość koherencji na poziomie kilkunastu centymetrów. Laser He-Ne jednomodowy, przy wysokiej jakości, może mieć długość koherencji przekraczającą 100m. Długość koherencji laserów specjalnych może przekroczyć 100km. Popularne lasery półprzewodnikowe mają długość koherencji rzędu 1mm, ale ich droższe modele, projektowane na wysoką koherencję, mogą osiągnąć długość rzędu 100m.