

Spójność światła

Doc. dr Andrzej KUJAWSKI, dr Jan MOSTOWSKI

W ostatnich latach często słyszy się o spójności światła. Na przykład, mówiąc o laserze zawsze podaje się, że światło lasera jest „spójne”, w przeciwieństwie do światła pochodzącego z innych źródeł, które jest niespójne. W rzeczywistości pojęcie spójności (lub koherencji) dotyczy wszystkich rodzajów fal, jednakże w przypadku fal świetlnych jest ono niezbędne do wyjaśnienia wielu zjawisk optycznych.

Istotę pojęcia spójności można zrozumieć badając interferencję fal na wodzie, wytworzonych przez końce dwóch drgających prętów. Każdy drgający pręt wytwarza falę powierzchniową kolistą rozchodzącą się od swojego źródła. Jeśli pręty drgają równomiernie z jednakową częstością, powstające fale dają trwały obraz interferencyjny; maksima i minima wychyleń występują stale w tych samych miejscach. Takie fale nazywamy spójnymi. Jeśli jednak pręty drgają nieregularnie, to znaczy ich częstości lub amplitudy wychyleń zmieniają się niezależnie w przypadkowych chwilach czasu, maksima interferencyjne powstają w coraz to innych miejscach. O takich falach mówimy, że są niespójne.

Ten poglądowy obraz nie daje jednak precyzyjnej odpowiedzi na pytanie, kiedy dwie fale możemy uważać za spójne? Zanim dokładniej odpowiemy na to pytanie w przypadku fal świetlnych, przypomnijmy, że w konwencjonalnych źródłach, zwanych też źródłami termicznymi (na przykład żarówka, płomień, ale nie laser!), światło powstaje w wyniku pobudzenia do świecenia niezależnie od siebie znacznej liczby atomów. Powstająca fala świetlna jest wynikiem nałożenia się olbrzymiej liczby fal pochodzących od każdego świecącego atomu.

Aby opisać fakt, że światło nie jest falą monochromatyczną, posługujemy się następującym wyrażeniem dla wartości natężenia pola elektrycznego fali świetlnej w punkcie określonym wektorem położenia \mathbf{r} i w chwili t

$$E(\mathbf{r}, t) = \varepsilon(\mathbf{r}, t) \cos(\omega t - \mathbf{k}\mathbf{r} + \alpha(t)). \quad (1)$$

Funkcje $\varepsilon(\mathbf{r}, t)$ i $\alpha(t)$ są nazywane amplitudą i fazą fali; $\omega = 2\pi\nu$, gdzie ν jest częstością fali, a \mathbf{k} wektorem falowym. Na przykład, jeśli $E(\mathbf{r}, t)$ opisuje światło emitowane przez pojedynczy atom, to amplituda $\varepsilon(\mathbf{r}_0, t)$ w ustalonym punkcie \mathbf{r}_0 jest funkcją malejącą w czasie, opisującą zanik świecenia po wzbudzeniu atomu. W tym przypadku $\varepsilon(\mathbf{r}_0, t) = \varepsilon_0 \exp[-(t-t_0)/t_a]$, gdzie t_0 opisuje czas, w którym atom został pobudzony do świecenia, a ε_0 jest stałą.

Charakterystyczny czas zaniku świecenia dla większości atomów wynosi $t_a \approx 10^{-8}$ s. Fazę $\alpha(t)$ w tym przypadku można przyjąć za stałą. Przypomnijmy, że w dziedzinie optycznej ω jest rzędu 10^{15} s $^{-1}$.

W przypadku źródła termicznego pełne pole elektryczne wytworzone przez dużą liczbę niezależnie świecących atomów jest sumą pól emitowanych przez poszczególne atomy. Na skutek chaotycznego ruchu atomów i ich zderzeń każdy z nich emituje światło w nieco inny sposób.

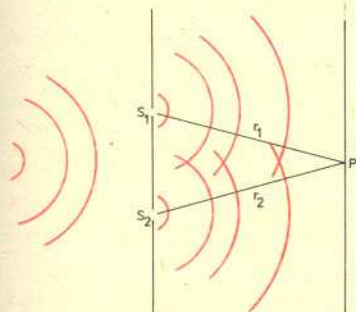
Całkowite pole, pochodzące od układu atomów, może być także scharakteryzowane wzorem (1). Nie można jednak w tym przypadku podać ogólnej dokładnej postaci $\varepsilon(\mathbf{r}, t)$ i $\alpha(t)$, gdyż zmiany ich mają charakter przypadkowy. Możemy jedynie stwierdzić, że istotnie zmieniają się one w czasie rzędu $t_s \approx 10^{-12}$ s. Czas ten nazywa się czasem spójności.

Rozważmy prosty przykład doświadczenia ukazującego interferencję światła, jakim jest doświadczenie Younga. Mamy tutaj (rys. 1) dwa źródła światła S_1 i S_2 , którymi są szczeliny lub otwory w nieprzezroczystym ekranie. W wyniku oświetlenia pola elektryczne w otworach S_1 i S_2 wynoszą odpowiednio $E(t, S_1)$ i $E(t, S_2)$. Każde z nich można opisać wzorem (1). Pole elektryczne w punkcie P powstaje w wyniku propagacji fali z otworów S_1 i S_2 do ekranu. Jeśli założymy, że odległości r_1 i r_2 punktu obserwacji P od źródeł S_1 i S_2 są znacznie większe od długości fali świetlnej, to pole dochodzące do ekranu możemy traktować jako superpozycję fal kulistych wychodzących z S_1 i S_2 . Mamy więc

$$E(t, P) = E_1 + E_2 = \frac{A_1}{r_1} E\left(t - \frac{r_1}{c}, S_1\right) + \frac{A_2}{r_2} E\left(t - \frac{r_2}{c}, S_2\right).$$

Stałe A_1 i A_2 zależne są od kształtu otworów, a ich dokładna postać nie będzie nam potrzebna; c jest prędkością światła. Każdy ze składników w powyższym wzorze określa stan pola w punkcie P pochodzącego od pól w S_1 i S_2 opóźnionych odpowiednio o czasy r_1/c i r_2/c . Zakładamy także dla uproszczenia, że obydwie fale są tak samo liniowo spolaryzowane.

W doświadczeniach rejestrujemy obraz interferencyjny na kliszy fotograficznej. Klisza zaczernia się proporcjonalnie do natężenia światła. W czasie naświetlania kliszy natężenie światła zmienia się, tak że klisza rejestruje jedynie pewną średnią wartość natężenia światła $I(P) = c \varepsilon_0 \langle E^2(P, t) \rangle$, gdzie $E(P, t)$ jest natężeniem pola w punkcie P , ε_0 przenikalnością dielektryczną próżni, a nawiasy ostre oznaczają uśrednianie. Oznaczając natężenia każdej z wiązek przez I_1 i I_2



Rys. 1

mamy $I_1(P) = c \epsilon_0 \langle E_1^2(P, t) \rangle$, $I_2(P) = c \epsilon_0 \langle E_2^2(P, t) \rangle$ i po wstawieniu do wzoru $I = c \epsilon_0 \langle (E_1 + E_2)^2 \rangle$ otrzymujemy

$$I(P) = I_1(P) + I_2(P) + \frac{2c\epsilon_0 A_1 A_2}{r_1 r_2} \langle \epsilon_1(t_1) \epsilon_2(t_2) \cos[\omega t_1 + \alpha_1(t_1)] \cos[\omega t_2 + \alpha_2(t_2)] \rangle \quad (2)$$

Aby obliczyć średnią wartość występującą we wzorze (2) musimy uwzględnić trzy charakterystyczne czasy: 1) okres drgań fali świetlnej $T = 2\pi/\omega \approx 10^{-15}$ s, 2) czas, w którym istotnie zmieniają się ϵ i α , czyli czas spójności $t_s \approx 10^{-12}$ s, 3) czas obserwacji obrazu interferencyjnego $t_0 \approx 10^{-2}$ s. Za czas obserwacji przyjęliśmy typowy czas naświetlania kliszy w aparacie fotograficznym. Może być on jednak krótszy przy innych metodach pomiaru natężenia. Po skorzystaniu ze wzoru na iloczyn cosinusów wyrażenie uśrednione we wzorze (2) przyjmuje postać

$$\frac{1}{2} \langle \epsilon_1 \epsilon_2 \cos[2\omega t + (r_1 + r_2)\omega/c + \alpha_1 + \alpha_2] \rangle + \frac{1}{2} \langle \epsilon_1 \epsilon_2 \cos[(r_1 - r_2)\omega/c + \alpha_1 - \alpha_2] \rangle,$$

gdzie α_1 i α_2 zależą odpowiednio od $t_1 = t - r_1/c$ i $t_2 = t - r_2/c$. Pierwszy uśredniany składnik jest funkcją szybko oscylującą w czasie, bowiem funkcja $\cos[2\omega t + (r_1 + r_2)\omega/c + \alpha_1 + \alpha_2]$ przyjmuje wielokrotnie wartości dodatnie i ujemne w czasie obserwacji ($\omega \approx 10^{15}$ s⁻¹, $t_0 \approx 10^{-2}$ s!), a zatem jej średnia czasowa jest równa zero. Wzór (2) przybiera więc postać

$$I = I_1 + I_2 + \frac{A_1 A_2}{r_1 r_2} c \epsilon_0 \left\langle \epsilon_1 \left(t - \frac{r_1}{c} \right) \epsilon_2 \left(t - \frac{r_2}{c} \right) \cos \left[(r_1 - r_2) \frac{\omega}{c} + \alpha_1 \left(t - \frac{r_1}{c} \right) - \alpha_2 \left(t - \frac{r_2}{c} \right) \right] \right\rangle \quad (3)$$

Wykazaliśmy w ten sposób, że wypadkowe natężenie obserwowane w typowym eksperymencie Younga jest sumą natężeń I_1 i I_2 pochodzących od poszczególnych źródeł oraz członu interferencyjnego.

Wartość członu interferencyjnego zależy od sposobu zrealizowania doświadczenia. Jeśli źródła S_1 i S_2 świecą zupełnie niezależnie, to funkcja $\alpha_1(t_1)$ przyjmuje wartości nie związane z wartościami $\alpha_2(t_2)$. W rezultacie różnica $\alpha_1(t_1) - \alpha_2(t_2)$ zmienia się istotnie w czasie rzędu 10^{-12} s. Czas obserwacji $t_0 \approx 10^{-2}$ s obejmuje znaczną liczbę „istotnych zmian”, a zatem na skutek obecności funkcji cosinus wyrażenie uśredniane wiele, wiele razy zmienia znak i wartość średnia wynosi zero. W tym przypadku znika człon interferencyjny i $I = I_1 + I_2$. O falach wychodzących z S_1 i S_2 mówimy, że są niespójne. Z powyższej analizy możemy wyciągnąć wniosek, że znikanie obrazu interferencyjnego nie jest jakąś wewnętrzną cechą światła. Jest to wynik określony możliwościami aparatury pomiarowej, która uśrednia szybkie fluktuacje natężenia światła. Stwierdzenie, że fale świetlne są niespójne, czyli że nie mogą interferować, w istocie oznacza, że aparatura pomiarowa nie rejestruje obrazów interferencyjnych w czasach odpowiednio krótkich.

Aby w omawianym doświadczeniu możliwa była obserwacja interferencji, należy tak skonstruować aparaturę, aby różnica $\alpha_1 - \alpha_2$ nie zmieniała się w czasie obserwacji. Można to uzyskać w ten sposób, że S_1 i S_2 są szczelinami lub otworami oświetlonymi przez punktowe (bardzo mały otwór) pojedyncze źródło. Wówczas fazy $\alpha_1(t_1)$ i $\alpha_2(t_2)$ fal w otworach S_1 i S_2 zmieniają się przypadkowo, jednak $\alpha_1(t) = \alpha_2(t)$. Obraz interferencyjny będzie widoczny, jeśli $\alpha_1 \left(t - \frac{r_1}{c} \right) - \alpha_2 \left(t - \frac{r_2}{c} \right)$ będzie stałe w czasie. Można to uzyskać, jeśli różnica dróg optycznych

między punktem obserwacji a dwoma źródłami $\left(\frac{r_1}{c} - \frac{r_2}{c} \right)$ jest co najwyżej rzędu 1 mm. Wówczas

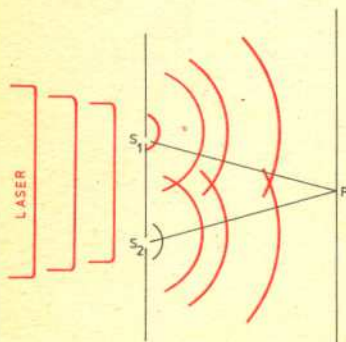
bowiem różnica między czasami $\left(t - \frac{r_1}{c} \right) - \left(t - \frac{r_2}{c} \right)$ jest mniejsza niż czas spójności

i $\alpha_1 \left(t - \frac{r_1}{c} \right) - \alpha_2 \left(t - \frac{r_2}{c} \right)$ jest wielkością prawie stałą w czasie. Gdyby światło było

monochromatyczne, to funkcje $\alpha(t)$ i $\epsilon(t)$ byłyby stałe. W tym przypadku nie mielibyśmy żadnego ograniczenia na rozmiar obrazu interferencyjnego, to znaczy prążki interferencyjne widoczne byłyby w całej płaszczyźnie ekranu. W rzeczywistości skończony czas spójności ogranicza rozmiar obszaru, w którym obserwujemy obraz interferencyjny.

Inną możliwość daje użycie do oświetlenia szczelin S_1 i S_2 światła odpowiedniego lasera (rys. 2). Mechanizm wytwarzania światła w laserze jest zupełnie inny niż w źródłach termicznych, bowiem atomy nie promieniują pojedynczo i niezależnie od siebie. Wprost przeciwnie, istota świecenia lasera polega na zgodnym w fazie promieniowaniu wszystkich atomów. W tym przypadku można teoretycznie uzasadnić, i jest to eksperymentalnie potwierdzone, że amplituda $\epsilon(t)$ i faza $\alpha(t)$ zmieniają się bardzo wolno, bowiem czas, w ciągu którego następują ich istotne zmiany, może osiągać wartość około $t_s \approx 0,1$ s lub nawet większą w przypadku specjalnych konstrukcji lasera. Wzór (3) określający wypadkowe natężenie światła pozostaje słuszny, ale uśrednianie należy przeprowadzić inaczej. Mianowicie w czasie obserwacji $t_0 \approx 10^{-2}$ s wielkości $\epsilon(t)$ i $\alpha(t)$ można uważać za stałe i wzór można przekształcić do postaci

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos[(r_1 - r_2)\omega/c + \alpha_1 - \alpha_2]. \quad (4)$$



Rys. 2

Tak więc przy użyciu światła lasera można łatwo oglądać obrazy interferencyjne w doświadczeniu Younga; nie jest konieczne używanie pojedynczego źródła punktowego. Okazuje się również, że zakres odległości szczelin S_1 i S_2 , dla których obserwuje się obraz interferencyjny, jest znacznie większy niż w przypadku termicznych źródeł punktowych. Przypadek opisany wzorem (4) odpowiada pełnej spójności fal wychodzących ze szczelin S_1 i S_2 . Maksymalna i minimalna wartość natężenia I wynoszą: $I_{\max} = (\sqrt{I_1} + \sqrt{I_2})^2$, $I_{\min} = (\sqrt{I_1} - \sqrt{I_2})^2$. Wzór (3) natomiast w sposób zupełnie ogólny opisuje spójność częściową światła, charakteryzując w granicznych sytuacjach zupełny brak spójności (niespójność) i pełną spójność.

W naszych rozważaniach charakteryzowaliśmy światło czasem spójności t_s , to znaczy czasem, w ciągu którego amplituda i faza fali pozostawały wielkościami stałymi. Czas ten zarówno dla światła termicznego jak i laserowego był dłuższy niż okres drgań fali świetlnej. Przy wyjaśnianiu powstawania prążków interferencyjnych w doświadczeniu Younga skorzystaliśmy z faktu, że czas spójności światła laserowego może być znacznie dłuższy niż czas spójności światła termicznego. Analiza teoretyczna doświadczenia Younga, jaką przeprowadziliśmy, w istocie dotyczyła zachowania się średniej wartości iloczynu dwóch pól elektrycznych $\langle E_1(t)E_2(t') \rangle$. Ta wielkość nazywana jest funkcją korelacji pierwszego rzędu. Istnieją doświadczenia, w których mierzy się uśrednione iloczyny większej niż dwa liczby pól, na przykład $\langle E_1(t_1)E_2(t_2)E_3(t_3)E_4(t_4) \rangle$ — są to funkcje korelacji wyższego rzędu. Na podstawie tych doświadczeń można wykazać, że światło lasera i światło termiczne różnią się nie tylko czasem i obszarem spójności.

W artykule tym przy omawianiu pojęcia spójności pierwszego rzędu wprowadziliśmy uśrednienie, bowiem amplituda i faza fali świetlnej zmieniają się w sposób przypadkowy. W ten sposób uwzględniliśmy statystyczne cechy fali świetlnej. W doświadczeniach, w których mierzy się funkcje korelacji wyższych rzędów wykazuje się, że statystyczne właściwości światła termicznego i laserowego istotnie się różnią.

Na zakończenie podkreśliśmy jeszcze, że cechy statystyczne pola świetlnego w inny i istotnie głębszy sposób są powiązane z kwantowymi właściwościami światła. Prowadzi to do takiej interpretacji wprowadzonego wyżej pojęcia spójności światła, w której podstawową rolę grają szybkości zliczeń fotonów.



Rozwiązanie zadania F 92. Kolejne ładowanie kondensatorów i połączenie w baterię pozwala uzyskać napięcie $2\mathcal{E}$, zaś ich równoczesne ładowanie tylko \mathcal{E} . Pozostaje więc manipulowanie źródłem i połączonymi kondensatorami. Rozważmy przypadek przedstawiony na rysunku. Napięcie na każdym z kondensatorów wynosi \mathcal{E} , a różnica potencjałów na otwartym wyłączniku równa jest $q_A - q_B = \mathcal{E}$. Długo też zamknięcie wyłącznika spowoduje przegrupowanie ładunku na kondensatorach. Kondensator 2 doładuje się i napięcie na nim przekroczy \mathcal{E} , natomiast kondensator 1 ulegnie częściowemu rozładowaniu. Wystarczy teraz ponownie naładować kondensator 1 do napięcia \mathcal{E} i powtórzyć tę operację aż do chwili, gdy napięcie na kondensatorze 2 będzie dostatecznie bliskie wartości $2\mathcal{E}$. Łącząc wtedy kondensatory w szereg różnoinniennymi okładkami otrzymuje się baterię o pożądanym napięciu. Dokonajmy teraz rozważań ilościowych. Niech kondensatory 1 i 2 mają odpowiednio pojemności c_1 i c_2 , napięcie przed i -tym doładowywaniem $u_1^{(i-1)} = \mathcal{E}$, $u_2^{(i-1)}$ oraz ładunki $q_1^{(i-1)} = c_1\mathcal{E}$, $q_2^{(i-1)} = c_2u_2^{(i-1)}$.

Z zasady zachowania ładunku dla zakresowanego obszaru wynika związek

$$q_1^{(i)} + q_2^{(i)} = q_1^{(i-1)} + q_2^{(i-1)} = c_1\mathcal{E} + c_2u_2^{(i-1)},$$

który po wyrażeniu ładunków przez napięcia i pojemności ma postać

$$c_1u_1^{(i)} + c_2u_2^{(i)} = c_1\mathcal{E} + c_2u_2^{(i-1)}. \quad (1)$$

Z rozkładu potencjału wzdłuż obwodu wynika:

$$\mathcal{E} + u_1^{(i)} - u_2^{(i)} = 0. \quad (2)$$

Rozwiązując układ równań (1), (2) uzyskuje się rekurencyjny wzór na napięcie kondensatora 2:

$$\begin{cases} u_2^{(i)} = 2 \frac{c_1}{c_1 + c_2} \mathcal{E} + \frac{c_2}{c_1 + c_2} u_2^{(i-1)} \\ u_2^{(0)} = \mathcal{E}. \end{cases}$$

Ze wzoru tego można wyznaczyć przyrost napięcia podczas i -tego doładowywania

$$\begin{cases} \Delta u_2^{(i)} = u_2^{(i)} - u_2^{(i-1)} = \left(\frac{c_2}{c_1 + c_2} \right) \Delta u_2^{(i-1)} \\ \Delta u_2^{(1)} = \frac{c_1}{c_1 + c_2} \mathcal{E}. \end{cases}$$

Widać stąd, że przyrosty napięć tworzą ciąg geometryczny o ilorazie $q = \frac{c_2}{c_1 + c_2}$.

Napięcie po N operacjach doładowywania otrzymamy sumując N wyrazów tego ciągu

$$u_2^{(N)} = \mathcal{E} + \sum_{i=1}^N \Delta u_2^{(i)} = \left[2 - \left(\frac{c_2}{c_1 + c_2} \right)^N \right] \mathcal{E}.$$

Z zależności tej wynika, że:

1°. $\lim_{N \rightarrow \infty} u_2^{(N)} = 2\mathcal{E}$, co potwierdza rozważania jakościowe.

2°. szybkość narastania jest tym większa, im większa jest wartość ilorazu c_1/c_2 .

W praktyce rzadko zależy nam na wysokim napięciu w baterii kondensatorów. Częściej chodzi o to, żeby energia bądź ładunek wyzwolane podczas rozładowywania były jak największe. Czy wtedy również najlepszy efekt otrzymamy przy dużym stosunku c_1/c_2 ?

