

1. Podstawowe wielkości fizyczne

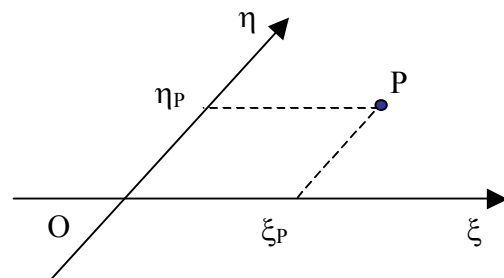
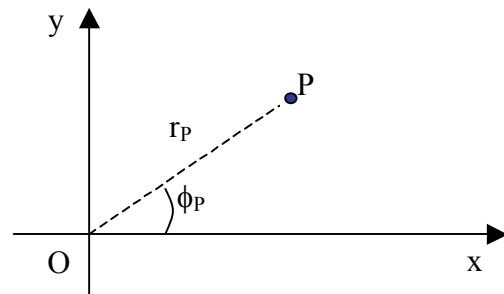
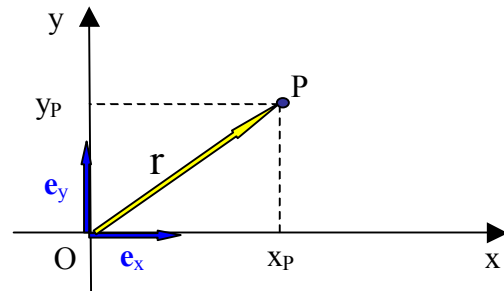
Położenie \mathbf{r}

Układ współrzędnych wymaga określenia jego początku O , trzech kierunków w przestrzeni oraz wzorca odległości. Wzorec metra jest definiowany jako odległość, na którą dotrze światło w próżni w czasie $1/3 \cdot 10^{-8}$ sekundy.

W kartezjańskim układzie współrzędnych położenie określone jako odległość w trzech (gdy rozpatruje się trzy wymiary) ortogonalnych kierunkach. Wektorem położenia punktu P jest $\mathbf{r} = x_P \mathbf{e}_x + y_P \mathbf{e}_y + z_P \mathbf{e}_z$, gdzie \mathbf{e}_x , \mathbf{e}_y , \mathbf{e}_z oznaczają wersory kierunkowe (symbole wielkości wektorowych są pogrubione).

W układzie biegunowym (stosowanym w dwóch wymiarach, dla trzech wymiarów stosuje się odpowiednio układ sferyczny lub cylindryczny) położenie określone przez wartość kąta ϕ i odległość r wzdłuż promienia łączącego z początkiem układu.

Układy mogą być nieortogonalne, gdy osie współrzędnych ξ i η nie są względem siebie prostopadłe.



Czas t

Wzorec czasu zdefiniowany jest poprzez proces, którego zależność od czasu jest znana (zakłada się, że jest znana). Wzorcem może być np. obiekt poruszający się ze stałą prędkością lub ruchem okresowym. Wzorec czasu był powiązany min. z ruchem Ziemi wokół Słońca. Obecnie definicja sekundy wiąże się z czasem trwania odpowiedniej liczby okresów drgań fali świetlnej emitowanej przez atomy cezu.

Prędkość \mathbf{v}

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \mathbf{r}}{\Delta t}$$

$$\mathbf{v} = v_x \mathbf{e}_x + v_y \mathbf{e}_y + v_z \mathbf{e}_z$$

Przyspieszenie \mathbf{a}

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2 \mathbf{r}}{dt^2}$$

Przy wyznaczaniu prędkości lub przyspieszenia Δt nie jest nieskończenie małe, ale znacząco mniejsze od wielkości typowych dla rozważanego problemu (np. mierząc prędkość samochodu $\Delta t \sim$ ułamki sekundy, przesuwanie się kontynentów $\Delta t \sim$ lata).

Wielkości fizyczne nie są nieskończone. Wielkość traktuje się jako nieskończenie wielką lub małą, gdy wykracza poza rozważany przedział typowych dla danego zagadnienia wartości.

Przykładowo:

1 rok $\approx \pi \cdot 10^7$ s, godzina wykładu $\approx 10^{-4}$ roku, wiek studenta $\approx 10^9$ s.

wiek Wszechświata $\approx 13,8$ mld lat $\approx 4 \cdot 10^{17}$ s,

czas życia rezonansów (rodzaj cząstek) $\approx 10^{-23}$ s

czasy połowicznego rozpadu niektórych pierwiastków $\approx 10^{24}$ s

Ograniczenia dotyczą także prędkości: graniczną prędkością przesyłania energii (informacji) jest prędkość światła w próżni: $c \approx 3 \cdot 10^8$ m/s.

Budowa materii

Cząstkami elementarnymi są kwarki, leptoni i bozony oddziaływania. Otaczająca nas materia składa się w zasadzie z dwóch rodzajów kwarków: górnego u (od *ang. up*) i dolnego d (*ang. down*), dwóch leptonów: elektronu e i neutrina ν oraz bozonów oddziaływania.

Istnieją jeszcze kwarki i leptoni związane z materią występującą w warunkach wysokich energii (takich jak np. panowały na początku istnienia Wszechświata). Są to kwarki: s (dziwny - *ang. strange*), c (powabny - *ang. charm*), b (niski - *ang. bottom*, lub piękny - *ang. beauty*), t (wysoki - *ang. top*, lub prawdziwy - *ang. true*), oraz odpowiadające im leptoni: mion, neutrino mionowe, taon, neutrino taonowe.

Kwarki i elektron obdarzone są ładunkiem elektrycznym: jeśli ładunek elektronu przyjąć jako równy -1 , to ładunek kwarka u wynosi $+2/3$ a kwarka d wynosi $-1/3$. Kwarki nie występują jako odosobnione cząstki: złożenia dwóch kwarków są cząstkami nazywanymi mezonami, a trzech to bariony. Proton jest cząstką składającą się z trzech kwarków 'uud' o ładunku sumarycznym $+1$ a neutron jest złożeniem kwarków 'udd' o sumarycznym zerowym ładunku. Rozmiary kwarków i leptonów są zbyt małe, aby można je zmierzyć, natomiast rozmiary protonu i neutronu są rzędu 10^{-15} m, zaś ich masy rzędu 10^{-27} kg. Protony i neutrony łączą się w układy stanowiące jądra atomowe. Po dołączeniu do jądra elektronów powstaje atom, którego rozmiary są rzędu 10^{-10} m. Atomy łączą się w związki chemiczne, kryształy itd.

Dla porównania:

promień Ziemi $\approx 6 \cdot 10^6$ m, masa Ziemi $\approx 10^{24}$ kg

odległość Ziemi od środka naszej Galaktyki $\approx 2 \cdot 10^{20}$ m

rozmiar obserwowanego Wszechświata $\approx 10^{26}$ m

Dokładność pomiarów

Przykładowa niedokładność pomiarów:

masy elektronu $\pm 0,0005\%$

masy Ziemi $\pm 0,1\%$

częstotliwości promieniowania przy przejściu między dwoma poziomami struktury nadsubtelnej wodoru $\pm 10^{-10}\%$

Dokładność pomiarów ograniczona jest niedoskonałością układów pomiarowych (mierników i wzorców), ale także stosowanymi teoriami oraz efektami kwantowymi. Np. zgodnie z kwantową zasadą nieoznaczoności, niedokładność pomiaru pędu δp (tzn. gdy pęd jest określony jako $p \pm \delta p$) i niedokładność pomiaru położenia δx przy jednoczesnym pomiarze spełnia nierówność: $\delta p \delta x > \hbar$, gdzie $\hbar = 10^{-34}$ kg m²/s jest stałą Plancka. Oznacza to, że np. określenie położenia protonu z dokładnością większą niż jego średnica ($\delta x \sim 10^{-15}$ m) praktycznie wyklucza jednoczesne określenie jego prędkości v (bo wtedy nieoznaczoność

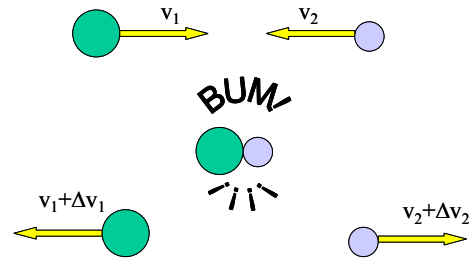
pomiaru prędkości: $\delta v = \frac{\delta p}{m_{\text{protonu}}} > \frac{\hbar}{m_{\text{protonu}} \delta x} \approx 10^8 \frac{\text{m}}{\text{s}}$, czyli określenie jego prędkości

obarczone jest niedokładnością porównywalną z prędkością światła).

Masa (bezwładna) m

W wyniku zderzenia sprężystego dwóch ciał zmieniają ich prędkości:

$\Delta v = v_{\text{po zderzeniu}} - v_{\text{przed zderzeniem}}$. Zmiany te są różne w zależności od prędkości początkowych, jednakże ich stosunek jest stały: $|\Delta v_1/\Delta v_2| = \text{const.}$ Może to posłużyć jako definicja masy $m_2/m_1 = |\Delta v_1/\Delta v_2|$ (pozwala na porównanie masy ciała m_2 z masą wzorca m_1). Masa zgodnie z taką definicją jest miarą bezwładności ciał (im większa masa tym mniejsza zmiana prędkości).

Pęd $p = mv$

Definicja masy bezwładnej (podana powyżej) jest tożsama zasadzie zachowania pędu: pęd sumaryczny przed zderzeniem = pęd po zderzeniu, czyli w wyniku zderzenia wzrost pędu pierwszego ciała = zmniejszenie pędu drugiego ciała: $m_1|\Delta v_1| = m_2|\Delta v_2|$. Zasada zachowania pędu wynika z założenia jednorodności przestrzeni, tzn. tego, że własności przestrzeni nie zmieniają się w wyniku translacji (zmiany położenia). I tak np. analizując dwuwymiarowy ruch kulki po płaskiej poziomej powierzchni, gdy powierzchnia ta jest gładka (jednorodna) pęd powinien być zachowany. Dla powierzchni chropowatej (niejednorodnej) występuje siła tarcia i pęd nie jest zachowany.

Jednostki wielokrotne: $10^{\pm n}$ gdzie $n=0,3,6,\dots$

- $10^{18}=1\text{E}$ (eksa)
- $10^{15}=1\text{P}$ (peta)
- $10^{12}=1\text{T}$ (tera)
- $10^9=1\text{G}$ (giga)
- $10^6=1\text{M}$ (mega)
- $10^3=1\text{k}$ (kilo)
- $10^0=1$
- $10^{-3}=1\text{m}$ (mili)
- $10^{-6}=1\mu$ (mikro)
- $10^{-9}=1\text{n}$ (nano)
- $10^{-12}=1\text{p}$ (piko)
- $10^{-15}=1\text{f}$ (femto)
- $10^{-18}=1\text{a}$ (atto)

2. Siły i ich źródła

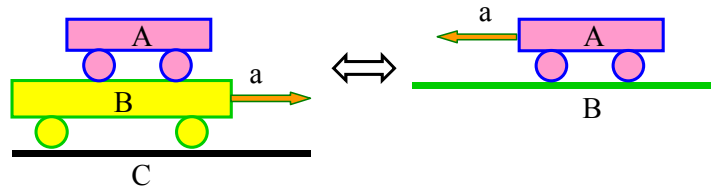
Siła \mathbf{F}

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d(m\mathbf{v})}{dt} = \{\text{gdy masa się nie zmienia w czasie}\} = m\mathbf{a}$$

Siła jest tym, co zmienia pęd a w konsekwencji prędkość ciał (zmiana prędkości jest tym łatwiejsza im jest mniejsza masa bezwładna ciała). Określenie siły działającej na obiekt w każdym miejscu rozpatrywanej przestrzeni oznacza zdefiniowanie pola siły (ogólnie termin "pole czegoś" oznacza przyporządkowanie wartości "czegoś" w każdym miejscu w przestrzeni, np. pole temperatury określa zależność temperatury od położenia).

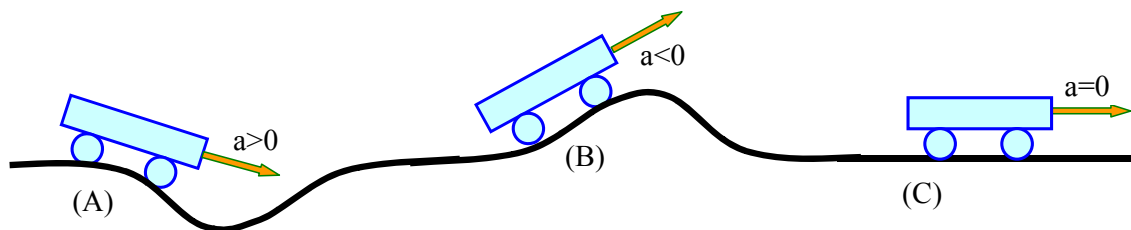
Źródłami sił są cztery podstawowe oddziaływania: grawitacyjne, elektromagnetyczne, jądrowe słabe oraz jądrowe silne.

Siły pojawiają się również w nieinercjalnych układach odniesienia, tzn. takich, które przyspieszają względem innych układów. Przykładem jest sytuacja przedstawiona na rysunku obok: na



wózek A nie działa żadna siła, więc jest nieruchomy względem podłoża C, natomiast względem przyspieszającego wózka B porusza się z przyspieszeniem $-\mathbf{a}$. Zatem analizując ruch wózka A względem B, wózek ten zachowuje się tak, jakby na niego działała siła $\mathbf{F} = -m\mathbf{a}$, gdzie m jest jego masą. Siła ta występuje tylko w układzie przyspieszającym (układzie związanym z wózkiem B) i dlatego nazywana jest pozorną lub siłą bezwładności, bo jej wartość zależy od masy bezwładnej ciała. Podobny mechanizm dotyczy występowania w układzie wirującym przyspieszenia odśrodkowego (siły odśrodkowej) oraz przyspieszenia (siły) Coriolisa działającego min. na ciała spadające i poruszające się na Ziemi.

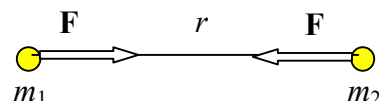
Występowanie układów nieinercjalnych z siłami pozornymi jest utożsamiane z zakrzywieniem przestrzeni. W układzie dwuwymiarowym zakrzywienie takie można wyobrazić sobie jako "dołek", w pobliżu którego pojawia się siła wciągająca do dołka (wózek (A) na rysunku poniżej). Sytuacja przeciwna występuje w pobliżu wzniesienia, wokół którego występuje odpychanie (wózek (B)). W płaskiej przestrzeni (układzie inercjalnym) nie ma sił pozornych (wózek (C)).



Siła grawitacyjna:

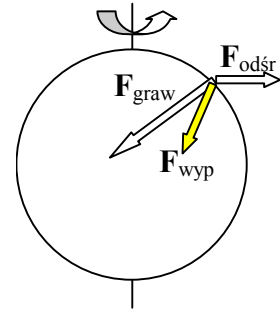
$$F = G \frac{m_1 m_2}{r^2}$$

gdzie $G = 6,6 \cdot 10^{-11} \text{ Nm}^2/\text{kg}^2$ jest stałą grawitacji, r odległością pomiędzy dwoma ciałami a m masą



grawitacyjną ("ładunkiem" siły grawitacyjnej).

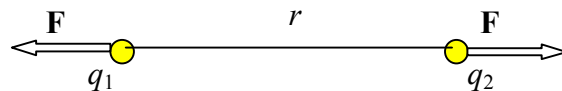
Z doświadczenia wynika, że masa bezwładna i grawitacyjna są równoważne. Siła wypadkowa F_{wyp} działająca na ciała na powierzchni Ziemi jest złożeniem siły przyciągania grawitacyjnego do środka Ziemi F_{grav} oraz siły odśrodkowej $F_{\text{odśr}}$ związanej z ruchem obrotowym Ziemi, prostopadłej do osi obrotu Ziemi. Siła grawitacyjnego przyciągania zależy od masy grawitacyjnej ciała zaś siła odśrodkowa jest siłą pozorną równą iloczynowi przyspieszenia odśrodkowego i masy bezwładnej. Gdyby dla różnych ciał stosunek masy bezwładnej i masy grawitacyjnej był różny, to wypadkowa siła miałaby różny kierunek. Doświadczenie wykazuje, że pion jest taki sam dla wszystkich ciał, czyli masa bezwładna i grawitacyjna są tą samą wielkością fizyczną. W konsekwencji pole grawitacyjne jest równoważne istnieniu nieinercyjnego układu odniesienia, czyli polu sił pozornych. Ogólna teoria względności Einsteina łączy geometrię (kształt i zakrzywienie) przestrzeni z rozkładem masy.



Siły elektryczne (elektromagnetyczne)

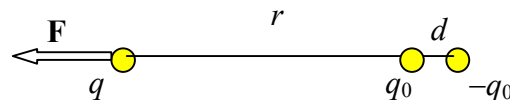
Pomiędzy dwoma nieruchomymi ładunkami elektrycznymi:

$$F = -\frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2}$$



gdzie $1/(4\pi\epsilon_0) = 9 \cdot 10^9 \text{ Nm}^2/\text{C}^2$ jest współczynnikiem proporcjonalności, r jest odległością zaś q wartością ładunku elektrycznego (całkowitą wielokrotnością ładunku elementarnego elektronu $e = -1,6 \cdot 10^{-19} \text{ C}$). Ładunki mogą być dodatnie bądź ujemne a siła odpowiednio przyciągająca (dla ładunków o różnych znakach) bądź odpychająca (dla ładunków o takich samych znakach).

Dwa ładunki q_1 i q_2 o tej samej wartości, ale przeciwnych znakach $q_1 = -q_2 = q_0$ tworzą tzw. dipol elektryczny. Siła działająca na ładunek q w polu elektrycznym



wytworzonym przez dipol elektryczny, w przypadku gdy wszystkie ładunki znajdują się na jednej prostej, wynosi:

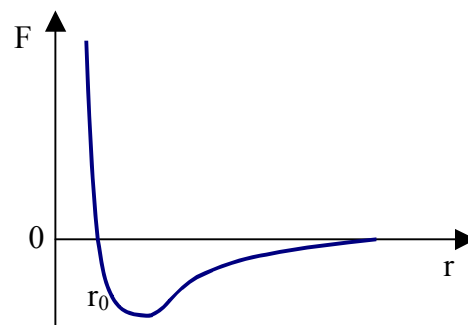
$$F = -\frac{1}{4\pi\epsilon_0} \frac{q_0 q}{r^2} + \frac{1}{4\pi\epsilon_0} \frac{q_0 q}{(r+d)^2} = -\frac{q_0 q}{4\pi\epsilon_0} \left(\frac{1}{r^2} - \frac{1}{(r+d)^2} \right) = -\frac{q_0 q}{4\pi\epsilon_0} \frac{(2r+d)d}{r^2(r+d)^2},$$

czyli dla $d \ll r$:

$$F \approx -\frac{q_0 q}{4\pi\epsilon_0} \frac{2d}{r^3}.$$

Ogólnie w polu dipola: $F \propto 1/r^3$, podczas gdy w polu pojedynczego ładunku (monopola) $F \propto 1/r^2$.

Siła działająca na ładunek w polu wytworzonym przez ładunek i dipol elektryczny: $F \propto \alpha/r^2 + \beta/r^3$, gdzie pierwszy człon odpowiada oddziaływaniu z pojedynczym ładunkiem a drugi z dipolem. Gdy stałe α i β mają różne znaki siła ta na bliskich



odległościach $r \ll r_0$ ma kierunek zgodny ze znakiem β , dla $r = r_0 = -\beta/\alpha$ jest równa zero, a dla $r > r_0$ ma kierunek zgodny ze znakiem α . Zatem siła ta w zależności od odległości będzie przyciągająca lub odpychająca, a na dużych odległościach maleje do zera.

Podobny charakter jak powyżej ma siła oddziaływania elektrycznego pomiędzy atomami (lub cząsteczkami), które mają złożony rozkład ładunku elektrycznego. Dwa atomy będące blisko siebie odpychają się, w odległości r_0 od siebie nie działają na siebie żadną siłą, na większych odległościach przyciągają się a na bardzo dużych siła znowu się zeruje. Odległość r_0 , poniżej której występuje siła odpychająca wyznacza rozmiar atomu i określa średnią odległość pomiędzy atomami w kryształach. Zmniejszanie tej odległości (ściskanie) wywołuje reakcję w postaci siły odpychającej zaś zwiększanie tej odległości (rozciąganie) powoduje wystąpienie siły przyciągającej. Jest to źródło sił: chemicznych, krystalicznych, sprężystości, tarcia, etc.

Układ ładunków wytwarza pole elektryczne, które może być sumarycznie opisywane wektorem natężenia pola elektrycznego \mathbf{E} określającego siłę działającą na ładunek q w polu innych ładunków: $\mathbf{F} = q\mathbf{E}$. Analogicznie można opisywać pole magnetyczne poprzez wektor indukcji pola magnetycznego \mathbf{B} . Tym razem jednak pole wytwarzane jest przez poruszające się ładunki (w magnesach stałych są to ruchy elektronów wewnątrz atomów a w elektromagnesach prąd elektryczny) i pole to oddziałuje na poruszające się ładunki. Siła działająca na ładunek q w polu \mathbf{B} jest również proporcjonalna do prędkości \mathbf{v} , z jaką się ten ładunek porusza: $F \sim qvB$. Dokładnie: $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$, gdzie \times oznacza iloczyn wektorowy, którego wartość jest równa iloczynowi wartości vB oraz sinusa kąta pomiędzy tymi wektorami a kierunek jest prostopadły do wektorów \mathbf{v} i \mathbf{B} .

Ogólnie pola elektryczne i magnetyczne są ze sobą powiązane i dlatego mówi się o zjawiskach elektromagnetycznych.

Siły jądrowe

Siły te występują na bardzo małych odległościach, charakterystycznych dla jąder atomowych. Dlatego dotyczą one tylko oddziaływań pomiędzy cząstkami elementarnymi. Siły jądrowe silne między innymi łączą kwarki w protony i neutrony oraz łączą neutrony i protony w jądra atomowe. Siły jądrowe słabe między innymi powodują rozpad β jąder i neutronu.

W podejściu kwantowym oddziaływanie pomiędzy cząstkami odbywa się poprzez wymianę wirtualnej cząstki będącej nośnikiem tego oddziaływania. Cząstkami przenoszącymi oddziaływania są następujące bozony:

- gluon (jądrowe silne),
- foton (elektromagnetyczne),
- bozony pośrednie W^- , W^+ oraz Z (jądrowe słabe),
- grawiton (grawitacyjne).

Dodatkowo postuluje się istnienie bozonu Higgsa, który odpowiada za tzw. oddziaływanie skalarnie (nie dotyczące oddziaływań pomiędzy dwoma cząstkami) powodujące to, że cząstki mają różne masy. Eksperymentalne potwierdzenie istnienia (lub nie) bozonu Higgsa przewidywane jest w najbliższych latach.

W modelu standardowym oddziaływania jądrowe słabe i elektromagnetyczne są formą jednego oddziaływania elektroslabego. Są również prowadzone próby stworzenia teorii, w której zunifikowane byłyby wszystkie oddziaływania. Oddziaływaniem sprawiającym największą trudność dla jednolitej teorii jest oddziaływanie grawitacyjne, które jest o wiele rzędów wielkości słabsze niż pozostałe oddziaływania.

Oddziaływania fundamentalne:	Natężenie względne	Zasięg
Grawitacyjne	$6 \cdot 10^{-39}$	∞
Elektromagnetyczne	$7 \cdot 10^{-3}$	∞
Jądrowe słabe	10^{-5}	10^{-18} m
Jądrowe silne	1	10^{-15} m

Zasady dynamiki Newtona

Zasady te podane w 1687 roku, obecnie formułuje się następująco:

1. Jeśli na ciało nie działa żadna siła wypadkowa, to ciało jest nieruchome lub porusza się ruchem jednostajnym ($\mathbf{F} = 0 \Rightarrow \mathbf{v} = 0$ lub $\mathbf{v} = \text{const.}$).
2. Szybkość zmiany pędu ciała jest równa sile działającej na to ciało ($\mathbf{F} = d\mathbf{p}/dt$).
3. Siły wzajemnego oddziaływania dwóch ciał są sobie równe, lecz przeciwnie skierowane ($\mathbf{F}_{1 \rightarrow 2} = -\mathbf{F}_{2 \rightarrow 1}$).

Ad 1. Tzw. zasada bezwładności jest w rzeczywistości definicją inercjalnego układu odniesienia i stwierdzeniem, że istnieją układy inercjalne (tzn. takie, w których ciała są nieruchome lub poruszają się ze stałą prędkością). Pozostawanie w spoczynku nie jest tylko szczególnym przypadkiem ruchu jednostajnego, bo np. wszystkie fotony poruszają się ze stałą prędkością a nie istnieją fotony nieruchome.

Ad. 2. Jest to definicja siły, lecz również stwierdzenie, że przyczyną zmiany pędu jest siła.

Ad. 3. Jest to tzw. prawo akcji i reakcji. Nie jest ono jednak ściśle spełnione (szczególnie dla ciał odległych poruszających się z dużymi przyspieszeniami), bowiem zakłada, że oddziaływania rozchodzą się z nieskończoną prędkością. W rzeczywistości oddziaływania rozprzestrzeniają się ze skończoną prędkością i np. zmiana położenia jednego ciała wywoła zmianę kierunku działania siły na drugie ciało po czasie związanym z dotarciem odpowiedniego bozonu.

3. Podstawy teorii względności

Względność praw

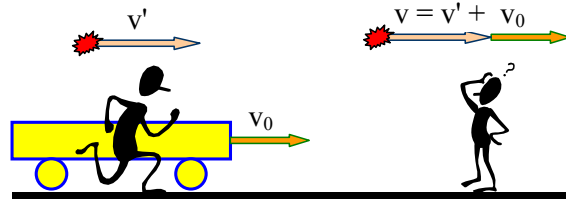
Prawa przyrody mają jednakową postać we wszystkich układach tzn. nie ma wyróżnionego punktu odniesienia we Wszechświecie (Ziemia porusza się w Układzie Słonecznym, który porusza się wokół centrum naszej Galaktyki, która oddala się od innych galaktyk itd.).

Wyróżnienie pewnych układów wiąże się jedynie np. z tym, że pojawiają się w nich siły lub, że pewne obiekty są w nich nieruchome (np. z punktu widzenia ludzi wyróżnionym układem jest układ związany ze środkiem Ziemi, lecz prawa fizyczne obowiązujące na Ziemi są takie same na każdej innej planecie o takim samym przyspieszeniu grawitacyjnym na jej powierzchni).

Światło rozchodzi się w próżni \Rightarrow nie wyróżnia żadnego układu odniesienia \Rightarrow wszystkie układy dla światła są równoważne \Rightarrow **prędkość światła w próżni względem każdego układu odniesienia jest taka sama**. Zostało to potwierdzone doświadczalnie: Michelson (w 1881) wykazał, że nie ma ośrodka (nazywanego eterem) koniecznego do rozchodzenia się światła i w przeciwieństwie np. do fal dźwiękowych światło propaguje się w próżni. W laboratorium CERN (w 1964) dokonano pomiaru prędkości fotonów γ emitowanych przez mezony π^0 poruszające się z prędkością $0,99975c$ i pokazano, że bez względu na prędkość źródła światła jego prędkość jest zawsze taka sama.

Dodawanie prędkości

Skoro prędkość światła ma taką samą wartość względem każdego obserwatora to nie obowiązuje klasyczna reguła dodawania prędkości, mówiąca, że prędkość v' mierzona przez obserwatora poruszającego się z prędkością v_0 różni się od prędkości v mierzonej przez nieruchomego obserwatora o wartość v_0 : $v = v' + v_0$.



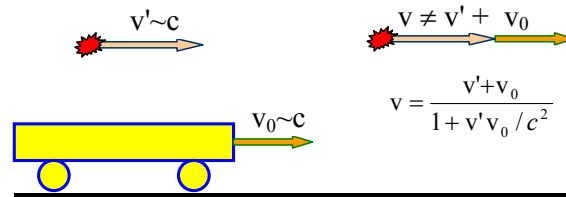
Aby prędkość światła była identyczna dla wszystkich obserwatorów prędkość względna wynosi:

$$v = \frac{v' + v_0}{1 + v'v_0/c^2}$$

i wówczas, gdy $v' = c$ to również $v = c$.

Natomiast, jeśli $v' < c$ to zawsze $v < c$. Dla

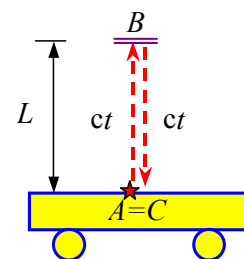
prędkości małych w porównaniu z prędkością światła $v'v_0 \ll c^2$ wzór ten staje się identyczny z klasycznym. W ogólności nie jest możliwe, aby obiekt o skończonej masie mógł osiągnąć prędkość światła w próżni c .



Dylatacja czasu

Rozpatrzmy następujący proces: światło od punktu A biegnie do punktu B na zwierciadle oddalonym o odległość L , odbija się i wraca do początkowego punktu $C = A$. Czas przelotu w obie strony wynosi $2t = 2L/c$.

Ten sam proces obserwowany z innego układu odniesienia wygląda inaczej. Jeśli źródło światła i zwierciadło odbijające znajdują się na wózku poruszającym się z prędkością v_0 to droga, którą przebywa



światło od punktu początkowego do zwierciadła i z powrotem jest dłuższa, gdyż zanim światło dotarło do punktu początkowego wózek przejechał pewną odległość. Jednak prędkość światła jest taka sama, stąd czas na to potrzebny musi być dłuższy.

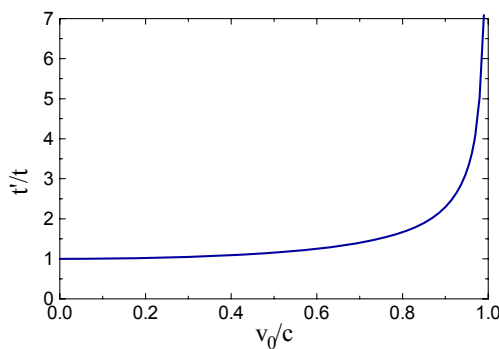
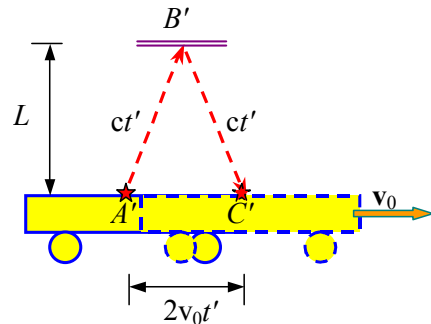
Z trójkąta prostokątnego uzyskuje się: $(ct')^2 = (v_0 t')^2 + L^2$,
czyli $(c^2 - v_0^2)t'^2 = L^2$. Czas, jaki trwał ten proces:

$$2t' = \frac{2L}{c\sqrt{1 - v_0^2/c^2}} > 2t = \frac{2L}{c}.$$

Porównując czas trwania tego samego procesu mierzony względem dwóch różnych punktów odniesienia uzyskuje się, zatem:

$$t' = \frac{t}{\sqrt{1 - v_0^2/c^2}},$$

gdzie czas t mierzony w układzie własnym (w którym zachodzi proces), jest krótszy niż czas t' w układzie względem którego układ własny się porusza. Wydłużenie czasu spowodowane poruszaniem się układu nazywa się dylatacją.



Zjawisko dylatacji staje się znaczące dopiero przy prędkościach porównywalnych z prędkością światła. Nawet, gdy $v_0 = c/2$ to wydłużenie wynosi jedynie kilkanaście procentów: $t'/t \approx 1,15$. Aby wydłużenie czasu było dwukrotne należy poruszać się z prędkością równą 87% prędkości światła.

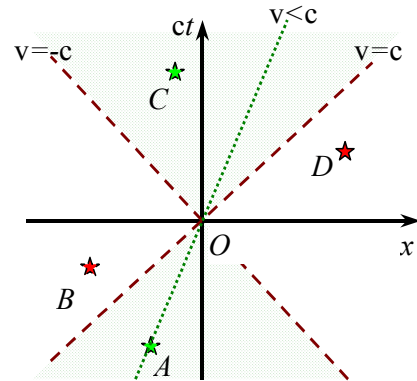
Dylatacja czasu jest dobrze potwierdzona w licznych eksperymentach i tłumaczy wiele zjawisk. Min. potwierdza wydłużenie średniego

czasu życia szybko poruszających się cząstek, (bo w układzie własnym cząstki, czyli takim, w którym ona jest nieruchoma, czas upływa wolniej niż w układzie laboratoryjnym, względem którego porusza się ona z dużą prędkością). Przykładowo miony mają średni czas życia $2,2 \cdot 10^{-6}$ s a docierają do powierzchni Ziemi, mimo, że są wytwarzane w górnych warstwach atmosfery (przebywają odległość rzędu 100km, czyli gdyby czas ich życia się nie wydłużył musiałyby poruszać się z prędkością przeszło 100-krotnie przewyższającą prędkość światła). Wspomniane na początku mezony π^0 wykorzystane do pomiaru prędkości światła, poruszające się z prędkością $0,99975c$, ulegały rozpadowi po czasie prawie 45 razy dłuższym niż mezony nieruchome.

Dylatację czasu można zmierzyć również dla małych prędkości. W 1972 roku w USA umieszczono w samolotach pasażerskich okrążających Ziemię po 4 zegary cezowe i porównano ich wskazania z zegarami pozostawionymi w laboratorium. W rzeczywistości z powodu ruchu obrotowego Ziemi, względem nieruchomego inercyjnego układu odniesienia najwolniej poruszały się zegary w samolotach lecących na zachód. Po powrocie zegary lecące na wschód były opóźnione o 60ns a lecące na zachód spieszyły się o 270ns w stosunku do zegarów laboratoryjnych. Oprócz wpływu efekt kinematyczny rzędu ± 100 ns na pomiar miał wpływ efekt grawitacyjny rzędu +150ns.

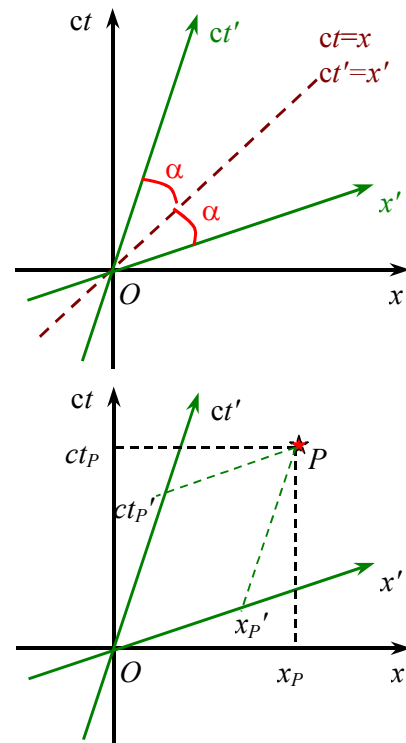
Czasoprzestrzeń

Czas w teorii względności nie jest wielkością niezależną od układu odniesienia i traktuje się go podobnie jak współrzędne położenia. Przestrzeń (trójwymiarowa) z czasem tworzy zatem czterowymiarową czasoprzestrzeń. Aby wymiar współrzędnej czasowej był taki sam jak wymiary współrzędnych przestrzennych, czas na swojej osi pomnożony jest przez prędkość światła. Wykres przedstawiający dwa wymiary (czasowy i jeden przestrzenny) przedstawiony jest obok. Punkty w czasoprzestrzeni określają zdarzenia, którym przyporządkowane jest miejsce i czas. Względem punktu O (obserwator umieszczony w początku układu współrzędnych: $x = 0$, $t = 0$) zdarzenia A i B miały miejsce w przeszłości ($t < 0$) zaś zdarzenia C i D dopiero się odbędą ($t > 0$). Prędkość światła jest prędkością graniczną również dla przesyłania informacji. Dlatego o zaistnieniu zdarzenia można się dowiedzieć dopiero po czasie nie krótszym niż potrzebny na przebycie przez światło odległości dzielącej od zdarzenia. Informacja o zdarzeniu A mogła dotrzeć do obserwatora O nawet z prędkością mniejszą niż c (kropkowana zielona linia $v = x/t < c$). Jednak informacja o zdarzeniu B dotrze do miejsca $x = 0$ dopiero w przyszłości. Dlatego do przeszłości obserwatora O należą tylko zdarzenia wewnątrz obszaru ograniczonego liniami światła $v = c$ i $v = -c$ (obszar zaznaczony). Podobnie jest ze zdarzeniami w przyszłości: na zdarzenie C można mieć jeszcze wpływ będąc w miejscu O, podczas gdy zdarzenie D odbędzie się niezależnie od zdarzenia O. Stąd do przyszłości zdarzenia O należy tylko obszar zaznaczony (zawierający też zdarzenie C), podczas gdy pozostałe zdarzenia są niezależne od zdarzenia O.



Układ poruszającego się obserwatora

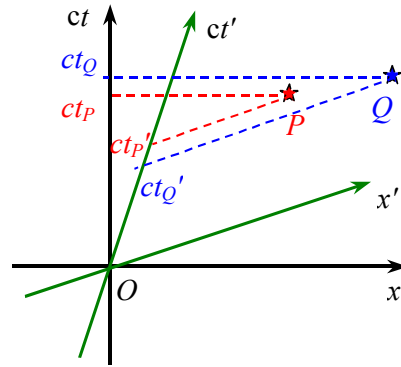
Początek układu współrzędnych przestrzennych wraz z upływem czasu przesuwa się wzdłuż osi współrzędnej czasowej (zdarzenie O określa położenie początku układu przestrzennego $x = 0$ w chwili $t = 0$). Dlatego oś czasu można traktować jako położenie w czasoprzestrzeni obserwatora znajdującego się w początku układu przestrzennego. Jeśli drugi obserwator oddala się z prędkością $v_0 < c$, to jego położenie wyznacza prosta nachylona do osi ct . Prosta ta wyznacza jednocześnie oś czasu ct' dla tego drugiego obserwatora, (który znajduje się w początku swojego układu współrzędnych przestrzennych). W klasycznym podejściu oś współrzędnej przestrzennej x' drugiego obserwatora powinna pokrywać się z osią x . Jednak z warunku, że prędkość światła jest jednakowa dla wszystkich obserwatorów wynika, że linia wyznaczająca promień świetlny (określona przez $ct = x$ a dla drugiego układu $ct' = x'$) jest dwusieczną kąta pomiędzy osią czasową ct (ct') i przestrzenną x (x') w każdym układzie. W rezultacie oś x' jest nachylona symetrycznie do osi x tak jak oś ct' do osi ct .



Współrzędne zdarzenia P wyznacza się w układzie poruszającym się (x', ct') tak jak w układach z nieortogonalnymi osiami.

Jednoczesność zdarzeń

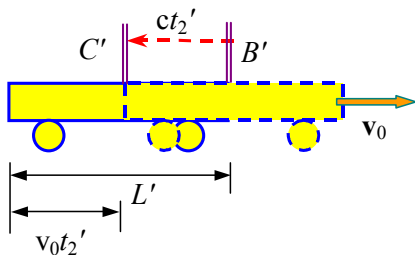
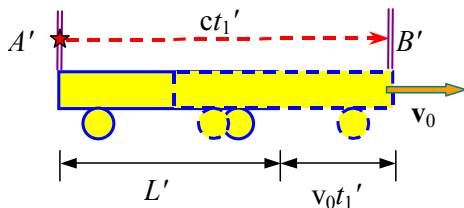
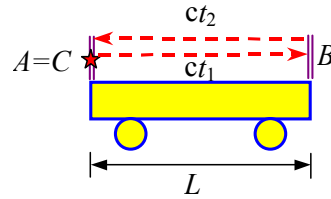
Na wykresie czasoprzestrzennym obok zaznaczone są dwa zdarzenia P i Q oraz odpowiadające im współrzędne czasowe w dwóch układach odniesienia (nieruchomym i poruszającym się). W układzie nieruchomym zdarzenie P zaszło przed zdarzeniem Q ($t_P < t_Q$) natomiast w układzie poruszającym kolejność zdarzeń była odwrotna ($t'_P > t'_Q$). Skoro kolejność tych zdarzeń jest różna dla różnych obserwatorów nie można powiedzieć w sposób bezwzględny, które z nich zaszło wcześniej. Taka niejednoznaczność dotyczy tylko zjawisk niezależnych od siebie, tzn. takich, dla których informacja pomiędzy nimi musiałaby się poruszać szybciej niż prędkość światła (odcinek je łączący leży na prostej odpowiadającej prędkości nadświetlnej). Oznacza to, że ani zdarzenie P nie może wpłynąć na przebieg zdarzenia Q ani na odwrót. Zdarzenia te są od siebie odseparowane przestrzennie.



Kolejność zdarzeń nie budzi wątpliwości w momencie, gdy jedno ze zdarzeń znajduje się w obszarze przeszłości (lub przyszłości) drugiego zdarzenia (tak jak np. zdarzenia A i O lub O i C na pierwszym z wykresów czasoprzestrzeni).

Kontrakcja długości

Rozpatrzmy teraz proces, gdy światło biegnie wzdłuż wózka o długości L . Czas t_1 przelotu do przodu (od A do B) jest równy czasu powrotu t_2 (od B do C) a czas przelotu w obie strony wynosi $t = t_1 + t_2 = 2L/c$. Dla tego samego procesu obserwowanego w układzie, względem którego wózek się porusza z prędkością v_0 , światło biegnąc do przodu (od A' do B') przebywa dłuższą drogę niż biegnąc z powrotem (od B' do C'). Odległości te wynoszą odpowiednio: $ct_1' = L' + v_0 t_1'$, oraz $ct_2' = L' - v_0 t_2'$, a związane z tym czasy:



$$t_1' = \frac{L'}{c - v_0}, \text{ oraz } t_2' = \frac{L'}{c + v_0}.$$

Całkowity czas przelotu w obie strony wynosi w rezultacie:

$$t' = t_1' + t_2' = \frac{2L'c}{(c - v_0)(c + v_0)} = \frac{2L'}{c(1 - v_0^2/c^2)}.$$

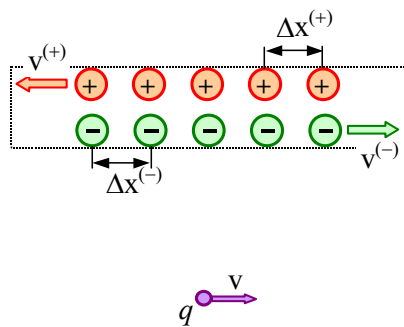
Jednocześnie zgodnie ze zjawiskiem dylatacji pomiędzy czasami t i t' zachodzi związek

$t = t' \sqrt{1 - v_0^2/c^2}$, który może być spełniony jedynie wtedy, gdy długość wózka w układzie poruszającym się L' jest mniejsza niż w układzie nieruchomym (własnym) L :

$$L' = L \sqrt{1 - v_0^2/c^2} \leq L.$$

Skrócenie wymiarów podłużnych nazywane jest kontrakcją długości.

Relatywistyczna kontrakcja długości między innymi tłumaczy powstawanie pola magnetycznego. W układzie nieruchomych ładunków istnieje tylko pole elektryczne a pole magnetyczne wytwarzane jest przez poruszające się ładunki. Ruch ładunków tworzy prąd



elektryczny, który obrazowo przedstawiony jest na rysunku obok jako strumień ładunków dodatnich i ujemnych. Jeśli ładunki dodatnie i ujemne się równoważą (jest ich tyle samo i poruszają się z jednakowymi prędkościami w przeciwnych kierunkach), to nie ma pola elektrycznego. Dlatego, gdy ładunek q jest nieruchomy ($v = 0$), to nie działa na niego żadna siła. Jednak, gdy ładunek q zaczyna się poruszać równoległe do kierunku prądu z prędkością v , to w jego układzie własnym prędkość ładunków dodatnich jest inna niż prędkość ładunków ujemnych. W rezultacie

relatywistycznego skrócenia odległości między ładunkami $\Delta x^{(+)}$ i $\Delta x^{(-)}$ stają się różne (bo różne są prędkości), co powoduje że liczba ładunków dodatnich na jednostkę długości jest inna niż ładunków ujemnych. Oznacza to, że wypadkowa różnica gęstości ładunków jest różna od zera (ładunki przestają się równoważyć) i istnieje pole elektryczne oddziałujące na ładunek q . Pojawienie się siły działającej na poruszający się ładunek wiąże się z istnieniem pola magnetycznego, które jak widać można sprowadzić do relatywistycznej transformacji pola elektrycznego. Mimo, że relatywistyczny efekt od pojedynczego ładunku jest bardzo mały, to duża liczba poruszających się ładunków daje znaczącą wartość pola magnetycznego. Osobne traktowanie pól elektrycznego i magnetycznego wynika jedynie z prostoty matematycznej.

Pęd i siła

Zasada zachowania pędu powinna być spełniona w każdym inercjalnym układzie odniesienia. Ponieważ nie obowiązuje klasyczne dodawanie prędkości, pęd zdefiniowany jako $\mathbf{p} = m\mathbf{v}$ nie spełnia tego warunku. Dlatego relatywistycznie pęd ma inną postać:

$$\mathbf{p} = \frac{m\mathbf{v}}{\sqrt{1 - v^2/c^2}}$$

i w rezultacie jego wartość rośnie do nieskończoności wraz ze wzrostem prędkości do prędkości światła. Czasami wielkość $m/\sqrt{1 - v^2/c^2}$ nazywa się masą relatywistyczną. Definicja siły ($\mathbf{F} = d\mathbf{p}/dt$) nie zmienia się, lecz teraz siła nie jest już iloczynem masy i przyspieszenia:

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d}{dt} \left(\frac{m\mathbf{v}}{\sqrt{1 - v^2/c^2}} \right) = \frac{m\mathbf{a}}{\sqrt{1 - v^2/c^2}} + m\mathbf{v} \frac{d}{dt} \left(\frac{1}{\sqrt{1 - v^2/c^2}} \right).$$

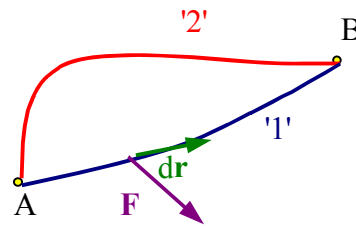
4. Praca i energia

Praca W

$$dW = \mathbf{F} \cdot d\mathbf{r}, \text{ czyli } W = \int_A^B \mathbf{F} \cdot d\mathbf{r},$$

gdzie A, B są odpowiednio początkiem i końcem drogi wzdłuż której wykonywana jest praca. Jeśli praca ta nie zależy od drogi (tzn. praca wzdłuż drogi '1' jest taka sama jak wzdłuż dowolnej innej drogi '2') to wtedy

$W = E_B - E_A$, gdzie E_A jest energią w punkcie A a E_B energią w punkcie B. Warto zauważyć, że praca wiąże się z procesem (zmianą) a energia jest przyporządkowana stanowi (miejscu). Nie zawsze jednak można posługiwać się energią, bo gdyby rozważać przykładowo pracę, jaką należy wykonać, aby przesunąć obiekt w obecności sił tarcia (np. ciężką szafę po podłodze) to istotną staje się droga, po której odbywa się przesunięcie.



Energia kinetyczna

wiąże się z pracą wykonaną na zmianę prędkości $v = dr/dt$:

$$E_{KIN} = \int_{v=0}^v \mathbf{F} \cdot d\mathbf{r} = \int_{v=0}^v \frac{d\mathbf{p}}{dt} \cdot d\mathbf{r} = \int_{v=0}^v d(m\mathbf{v}) \cdot \frac{d\mathbf{r}}{dt} = m \int_{v=0}^v d\mathbf{v} \cdot \mathbf{v} = m \int_{v=0}^v v dv = \frac{1}{2} mv^2$$

W układzie o dużej liczbie poruszających się obiektów wartość średnia ich energii kinetycznej określa temperaturę układu. Np. temperatura powietrza wiąże się ze średnią energią kinetyczną poruszających się cząsteczek gazu. Pomiar temperatury za pomocą klasycznego termometru polega na przekazywaniu energii w wyniku zderzeń cząsteczek gazu ze ściankami termometru, co prowadzi do wzrostu prędkości drgań atomów ośrodka (czyli zwiększeniu ich energii kinetycznej) a zatem wzrostu temperatury.

Energia potencjalna

nie jest związana z prędkością ciał a dotyczy oddziaływań między nimi. Zbliżenie lub oddalenie dwóch ciał, które oddziałują ze sobą (w wyniku np. przyciągania grawitacyjnego lub odpychania elektrycznego) wymaga wykonania pracy, czyli zmiany energii. Praca wykonana przeciw działającemu siłom zwiększa energię potencjalną ciał. Dlatego układy dążą do stanu, w którym energia potencjalna jest najmniejsza z możliwych. Przykładowo spadanie jabłek z drzew można tłumaczyć dążnością do zmniejszenia ich energii potencjalnej w polu grawitacyjnym Ziemi.

Moc P

określa pracę jaka jest wykonana w jednostce czasu: $P = \frac{dW}{dt}$.

Przykładowo podstawowe tempo metabolizmu u człowieka wymaga zużywania ok. 80 J (dżuli) energii w ciągu 1 s, czyli mocy 80 W (watów). Dodatkowo praca mózgu to ok. 40W a serca w spoczynku ok. 15 W. Sumarycznie student, który siedzi spokojnie na wykładzie i nie śpi "pracuje" z mocą ok. 150 W. Jednak już jazda na rowerze wymaga mocy ok. 500 W a w przypadku bardzo dużego wysiłku fizycznego ok. 1000 W. Z tego jednak wykonywana praca użyteczna nie przekracza 100 W. Koń jest zdolny do wykonywania pracy użytecznej z mocą ok. 500 W a przez krótkie okresy nawet 700÷800 W. Stąd jednostka mocy 1 KM (koń mechaniczny) = 746 W. Średnio zwierzęta są w stanie ok. 25% zużywanej energii zamienić na pracę mechaniczną. Dla porównania spalanie 1 litra benzyny daje ok. $3.1 \cdot 10^7$ J. Zjedzenie 100 g czekolady daje energię ok. 500 kcal, czyli $500 \times 4180 \text{ J} \approx 2.1 \cdot 10^6 \text{ J}$.

Zasada zachowania energii

jest wynikiem jednorodności przestrzeni ze względu na upływ czasu. Gdyby np. wraz z upływem czasu zmieniała się wartość stałej grawitacyjnej, zmieniałaby się siła przyciągania grawitacyjnego ciał i związana z tym energia nie byłaby zachowana. Gdy własności przestrzeni nie zmieniają się z upływem to sumaryczna energia nie ulega zmianie. W przypadku występowania sił tarcia także jest zachowana energia: praca wykonana, aby pokonać siłę tarcia zwiększa temperaturę trących powierzchni, zatem zwiększa energię cząstek i atomów. Sumaryczna energia makroskopowych obiektów i energia wewnętrzna ciał, związana z energią atomów i cząsteczek jest wówczas zachowana.

Relatywistyczna postać energii

W teorii względności pęd $p = \frac{mv}{\sqrt{1-v^2/c^2}}$, zatem w wyprowadzeniu energii kinetycznej:

$$E_{KIN} = \int_{v=0}^v \mathbf{F} \cdot d\mathbf{r} = \dots = m \int_{v=0}^v v d\left(\frac{v}{\sqrt{1-v^2/c^2}}\right) = \dots = \frac{mc^2}{\sqrt{1-v^2/c^2}} - mc^2$$

Energia kinetyczna jest różnicą czynnika zależnego od prędkości i stałego czynnika mc^2 . Zwiększając energię kinetyczną o ten stały czynnik uzyskuje się energię:

$$E = E_{KIN} + mc^2 = \frac{mc^2}{\sqrt{1-v^2/c^2}}, \text{ która dla ciał w spoczynku (dla } v=0) \text{ wynosi } E = mc^2.$$

Zgodnie z relatywistyczną postacią energii, gdy prędkość ciała dąży do prędkości światła w próżni $v \rightarrow c$ to energia dąży do nieskończoności $E \rightarrow \infty$. W konsekwencji żaden obiekt o niezerowej masie nie może osiągnąć prędkości światła (bo potrzeba by było wykonać nieskończenie wielką pracę). Jedynie cząstki o zerowej masie spoczynkowej ($m=0$), takie jak fotony, mogą się poruszać z prędkością światła (wówczas energia określana jest innym wzorem). Przekazywanie energii, czyli również informacji nie może się odbywać z prędkością większą niż c . Teoria względności dopuszcza istnienie cząstek (zwanymi tachionami), których prędkość jest większa od prędkości światła, które jednak nie mogą być spowolnione do prędkości c (co oznacza, że nie można ich zatrzymać a jedynie np. anihilować).

Konsekwencją wzoru $E = mc^2$ jest równoważność masy i energii. Zasada zachowania energii zawiera w sobie oprócz energii kinetycznej i potencjalnej również masę układu (pomnożoną przez c^2). Możliwe są zarówno procesy, w których zwiększa się masa układu kosztem zmniejszenia energii (np. mogą powstać nowe cząstki) jak również procesy, w których masa układu zamieniana jest na energię.

Proton (p) i neutron (n) są barionami i możliwe są procesy przemiany jednego w drugi. Ponieważ jednak proton ma dodatni ładunek elektryczny, zaś neutron jest obojętny elektrycznie, w procesie przemiany jednej cząstki w drugą musi być wyemitowana cząstka o takim ładunku, aby sumaryczny ładunek był zachowany. Cząstką taką może być elektron (e^-) o ładunku ujemnym lub antyelektron (e^+), czyli pozyton (o ładunku dodatnim). Dodatkowo jednak w przemianach cząstek musi być zachowana liczba leptonów, więc wytworzeniu elektronu towarzyszyć musi wytworzenie antyneutrino ($\bar{\nu}$) zaś wytworzeniu pozytonu towarzyszyć musi wytworzenie neutrino (ν). Stąd możliwe są następujące rozpady:

$$n \rightarrow p + e^- + \bar{\nu}$$

$$p \rightarrow n + e^+ + \nu$$

Jednakże tylko neutron ulega samoistnemu rozpadowi (średnio po ok. 15 minutach) zaś proton uważany jest za cząstkę stabilną. Wynika to z faktu, że masa neutronu jest nieznacznie większa od masy protonu i dzięki temu w rozpadzie swobodnego neutronu jest zachowywana energia. Aby mogło dojść do rozpadu protonu, należy dostarczyć dodatkowej energii (np. zderzyć proton z inną cząstką).

Przykładami zamiany masy na energię są procesy syntezy i rozpadu jąder atomowych. Masa jąder atomowych zazwyczaj jest mniejsza niż suma mas swobodnych nukleonów, z których składa się jądro. Dzięki temu nie dochodzi do samoistnego rozpadu jąder (gdyby masy te były identyczne, to nie byłoby powodu, dla którego nukleony grupowałyby się w jądra). Trzeba dostarczyć energii, aby jądro atomowe uległo rozpadowi (dostarczona energia musi przynajmniej równoważyć deficyt masy jądra w stosunku do sumarycznej masy swobodnych składników). Występujący w jądrach atomowych deficyt masy jest jednak różny dla różnych atomów. I tak deficyt masy przypadający na jeden nukleon jest najmniejszy dla jąder lekkich i bardzo ciężkich zaś największy dla jąder składających się z około 60 nukleonów. Dlatego łączenie dwóch lekkich jąder w jedno cięższe zwiększa deficyt masy i powoduje uwolnienie energii (tak jak to się dzieje na Słońcu i innych gwiazdach jak również w bombach wodorowych). Podobnie wyzwala się energia przy rozpadzie jąder ciężkich (co wykorzystuje się w elektrowniach atomowych jak również w bombach atomowych).

Zaskakujące problemy związane z masą i energią dostarcza kosmologia. Ogólna teoria względności Einsteina wiąże kształt i geometrię przestrzeni z rozkładem masy. Idealizując Wszechświat jako obszar o jednorodnym rozkładzie masy, której gęstość wynosi ρ można z ogólnej teorii względności uzyskać równanie opisujące zależność od czasu odległości R pomiędzy dwoma punktami w przestrzeni. Ponieważ może to być odległość pomiędzy dwoma dowolnymi punktami, nazwano ją promieniem Wszechświata. Równanie to ma postać:

$$\left(\frac{1}{R} \frac{dR}{dt}\right)^2 = \frac{8\pi G}{3} \rho + \frac{1}{3} \Lambda,$$

gdzie G jest stałą grawitacji zaś Λ jest tzw. stałą kosmologiczną. Stała kosmologiczna dodana została przez Einsteina do swoich równań po to, aby istniało rozwiązanie, w którym Wszechświat nie zmienia rozmiarów (tzn. $dR/dt = 0$) a jednocześnie ma niezerową gęstość masy (tzn. $\rho > 0$). Obecnie wiadomo, że Wszechświat się rozszerza ($dR/dt > 0$) a ok. 13,8 miliardów lat temu jego rozmiar był bliski zeru. Rozszerzanie nie odbywało się z jednakową prędkością i był okres zwany inflacją, gdy prędkość rozszerzania Wszechświata była nawet większa od prędkości światła. Obecnie przyjmuje się, że $dR/dt = H_0 R$ gdzie $H_0 \approx 66 \text{ km/s Mpc}$ jest stałą Hubble'a ($1 \text{ Mpc} \approx 3 \cdot 10^{22} \text{ m}$). Wykorzystując stałą Hubble'a można przepisać powyższe równanie do postaci:

$$\left(\frac{1}{R} \frac{dR}{dt}\right)^2 = H_0^2 (\Omega_M + \Omega_\Lambda),$$

gdzie $\Omega_M \propto \rho$ zaś $\Omega_\Lambda \propto \Lambda$. Stała kosmologiczna Λ pełni taką samą funkcję, co gęstość masy (czyli energii). Jednak wraz z rozszerzaniem Wszechświata maleje gęstość jego masy ($\rho R^3 c^2 = E = \text{const}$) natomiast stała kosmologiczna nie powinna zmieniać swej wartości. Odpowiadająca stałej kosmologicznej energia rośnie proporcjonalnie wraz z objętością Wszechświata $\propto R^3$. Według najnowszych szacunków $\Omega_M \approx 0,3$ zaś $\Omega_\Lambda \approx 0,7$, co oznacza, że decydujący wpływ na nasz Wszechświat ma energia, której gęstość nie zależy od objętości Wszechświata. Na marginesie, dla "normalnej" masy astrofizycy mają problemy w określeniu gdzie jest zawarta znaczna jej część i jak na razie znane są źródła masy dające tylko wartość $\Omega_M \approx 0,05$ (pozostała część nazywana jest ciemną materią).

5. Ruch falowy

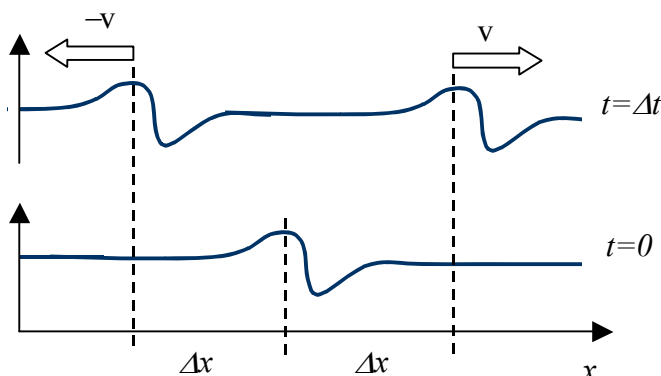
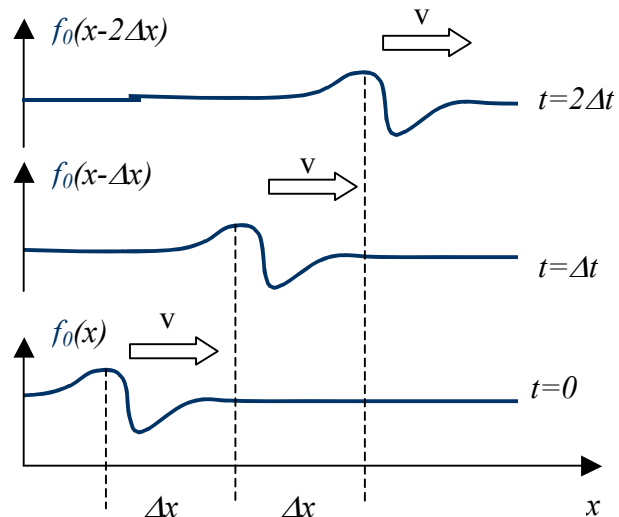
Fale

Poruszać mogą się nie tylko obiekty materialne, ale także rozkłady wartości różnych wielkości fizycznych. Przemieszczające się zaburzenie (odstępstwa od wartości średniej) nazywane jest falą. Przykładem fali jest fala na powierzchni wody, gdzie w stanie równowagi granica pomiędzy powietrzem a wodą jest nieruchoma. Po wrzuceniu do wody kamienia, w miejscu tym poziom wody się lokalnie obniży a tuż obok podwyższy. Takie zaburzenie nie pozostaje w miejscu, ale przesuwa się po powierzchni (w postaci fali kołowej). Cząsteczki wody nie podążają jednak za falą a tylko przesuwają się w górę i dół (dokładniej po torach eliptycznych) powodując podwyższenie lub obniżenie słupa wody. Zatem w kierunku poziomym nie przesuwają się cząstki a zmieniona wysokość powierzchni wody.

Falę tworzą zaburzenia różnych wielkości. Falą jest przesuwaające się wygięcie wzdłuż węża gumowego. Zmiana ciśnienia w gazie lub naprężenia w ciałach stałych tworzy falę akustyczną (dźwiękową). Są wreszcie fale elektromagnetyczne, w których zmiany dotyczą wartości pól elektrycznego i magnetycznego.

Równanie falowe

Rozkład zaburzenia pewnej wielkości f tworzącego falę przesuwa się w przestrzeni i dlatego jest opisywany funkcją położenia x (w przypadku jednego wymiaru) oraz czasu t : $f(x, t)$. Jeśli w chwili $t = 0$ zaburzenie ma postać $f(x, 0) = f_0(x)$, to po upływie czasu Δt funkcja ta jest przesunięta na odległość Δx , czyli $f(x, \Delta t) = f_0(x - \Delta x)$. Ponieważ przesunięcie Δx wiąże się z prędkością przesuwania się fali v jak $\Delta x = v\Delta t$, stąd w dowolnej chwili czasu $f(x, t) = f_0(x - vt)$. Zatem falę opisywać trzeba nie jako funkcję niezależnych zmiennych x i t , ale jako funkcję zmiennej $\phi = k(x - vt)$.



Stała k dodana została, aby ϕ zwane fazą fali było bezwymiarowe (k ma wymiar odwrotności długości).

Opisywana fala przemieszcza się wzdłuż osi x z prędkością v . Jednakże identyczne zaburzenie powinno się przesuwać również w przeciwną stronę (po wrzuceniu kamienia do wody wytworzona fala rozchodzi się po powierzchni we wszystkich kierunkach). Zatem oprócz fali

poruszającej się z prędkością $+v$ opisywanej fazą $\phi = k(x - vt)$ istnieje także fala poruszająca się z prędkością $-v$, dla której faza ma postać $\phi = k(x + vt)$. Obie fale opisywane są jednym równaniem różniczkowym:

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 f}{\partial t^2},$$

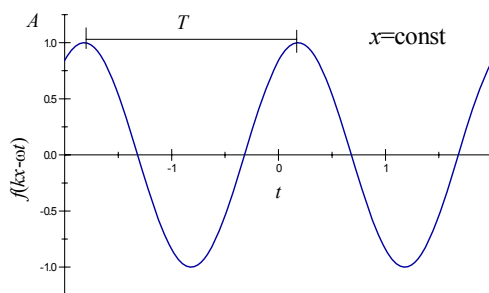
zwanym równaniem falowym. Równanie to opisuje jednocześnie dwie przeciwbieżne fale w układzie odniesienia, względem którego obie te fale mają taką samą wartość prędkości v . Prędkość ta jest prędkością rozchodzenia się fali względem nieruchomego ośrodka. Gdyby obserwować fale z układu poruszającego się, prędkości w przeciwnie strony byłyby inne (a postać równania falowego byłaby bardziej skomplikowana). Tak jednak nie jest dla fal elektromagnetycznych (w szczególności dla światła). W przeciwieństwie np. do temperatury czy ciśnienia pola elektryczne i magnetyczne istnieją również w obszarach, w których nie ma atomów. Dlatego fale elektromagnetyczne mogą rozchodzić się w próżni nie wyróżniając żadnego ośrodka. Prędkość fali na wodzie jest określona względem środka ciężkości wody, natomiast prędkości światła nie można odnieść do żadnego układu. Dlatego zgodnie z teorią relatywistyczną fala elektromagnetyczna we wszystkich układach ma jednakową prędkość równą c , a równanie falowe dla światła we wszystkich układach ma identyczną postać.

Fale harmoniczne (monochromatyczne)

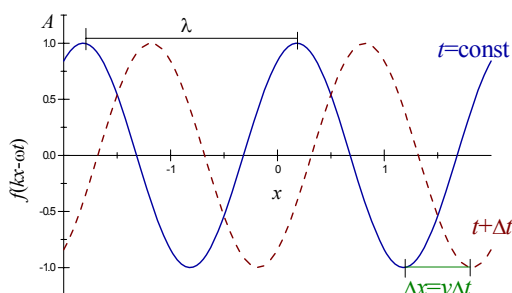
Fala może być opisywana przez różne funkcje $f_0(x \pm vt)$. Najczęściej jednak rozważana jest funkcja cosinus:

$$f = A \cos(kx \pm \omega t),$$

gdzie A jest amplitudą a $\omega = kv$ częstością. Fala o jednej częstości dla fal dźwiękowych jest falą harmoniczną a dla fal świetlnych opisuje światło monochromatyczne (o jednej barwie).



W fali harmonicznnej (monochromatycznej) mierząc wartość zaburzenia w określonym miejscu $x = \text{const}$ w funkcji czasu uzyskuje się przebieg oscylacyjny z okresem T . Okres jest czasem, po którym funkcja ma znowu taką samą wartość: $A \cos[kx - \omega(t + T)] = A \cos(kx - \omega t)$, czyli faza różni się o: $\omega T = 2\pi$. Okres wiąże się także z częstotliwością drgań $\nu = 1/T$ i w rezultacie pomiędzy częstością ω oraz częstotliwością fali ν jest zależność $\omega = 2\pi\nu$.

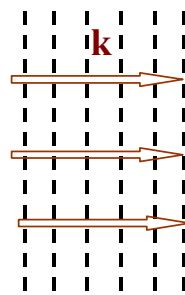


Mierząc wartość zaburzenia w określonej chwili czasu $t = \text{const}$ w funkcji odległości uzyskuje się identyczny jak poprzednio wykres. Okres przestrzenny nazywany jest długością fali λ i wiąże się ze stałą k zależnością: $k = 2\pi/\lambda$. Wielkość k nazywana jest liczbą falową. Po upływie czasu Δt rozkład zaburzenia przesunie się o odległość Δx z prędkością $v = \omega/k = \lambda\nu$. Prędkość ta określa przemieszczanie się punktów o stałej wartości zaburzenia (np.

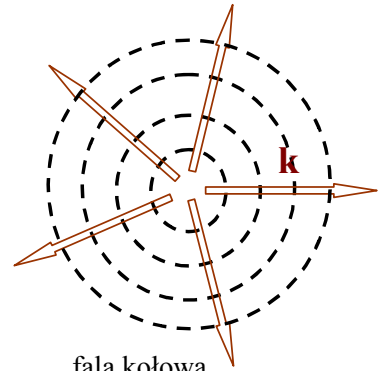
maksimum), czyli określa prędkość przemieszczania się stałej wartości fazy (argumentu funkcji). Stąd prędkość v nazywa się prędkością fazową fali.

Powyższe rozważania ograniczały się do jednego wymiaru przestrzennego (fala rozchodząca się wzdłuż osi x). Dla fal propagujących się na powierzchni (np. fala na powierzchni wody) czy w objętości (np. fala dźwiękowa) należy określić jak się zmienia zaburzenie w pozostałych wymiarach. Powierzchnia, na której faza fali ma jednakową wartość, (czyli

zaburzenie też ma jednakową wartość) nazywana jest powierzchnią stałej fazy. Jeśli powierzchniami tymi są płaszczyzny, fala nazywana jest płaską. Analogicznie są definiowane fale sferyczne, cylindryczne, a wśród fal powierzchniowych fale liniowe, kołowe itp.. Kierunek, wzdłuż którego fala się przesuwa określony jest przez wektor falowy \mathbf{k} o długości równej liczbie falowej. Lokalnie wzdłuż wektora falowego fala może być opisywana tak, jak fala jednowymiarowa.



fala liniowa

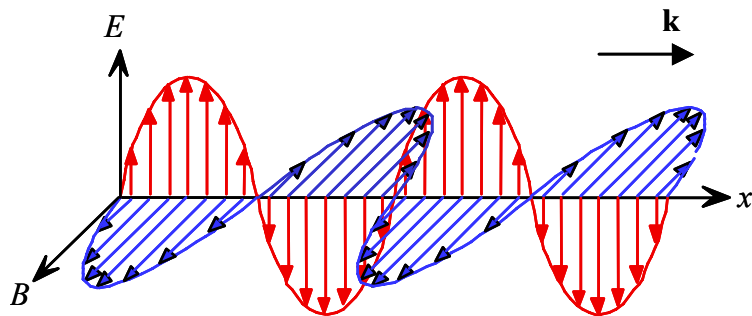


fala kołowa

Fale elektromagnetyczne

Zmieniające się w czasie pole elektryczne wytwarza pole magnetyczne. Podobnie zmiany pola magnetycznego wytwarzają pole elektryczne. Wzajemne generowanie pól elektrycznego i magnetycznego przesuwa się w przestrzeni jako fala elektromagnetyczna. W fali

elektromagnetycznej pole elektryczne i magnetyczne są względem siebie prostopadłe i prostopadłe do kierunku prędkości fali wyznaczanego przez wektor falowy. Rysunek ilustruje wzajemne położenie wektorów pola elektrycznego \mathbf{E} i magnetycznego \mathbf{B} oraz wektora falowego \mathbf{k} dla fali monochromatycznej.



Istnieje wiele rodzajów fal elektromagnetycznych. Zmienne pole elektryczne o częstotliwości $\nu = 50$ Hz również wytwarza falę, lecz jej długość $\lambda = c/\nu = 6 \cdot 10^6$ m = 6 000 km jest tak duża, że nie ma sensu stosować formalizmu falowego do prądu zmiennego o tej częstotliwości. Z drugiej strony fale o częstotliwościach tak dużych, że odpowiadająca im długość fali jest porównywalna z rozmiarami atomów, wygodniej jest opisywać jak cząstki (fotony). W szczególności ze względu na sposób wytwarzania i właściwości rozróżnia się następujące rodzaje fal elektromagnetycznych, przy czym granice pomiędzy nimi nie są ściśle określone.

fale radiowe: długości fal $\lambda \sim \text{m} \div \text{km}$, a odpowiadający im zakres częstotliwości

$\nu \sim 10^3 \div 10^8$ Hz; wytwarzane przez zmienny prąd elektryczny w antenach; na potrzeby łączności radiowej dzielone na fale długie, średnie, krótkie, UKF itp.

mikrofale: $\lambda \sim \text{cm}$ ($\nu \sim 10^9$ Hz), zaliczane również do fal radiowych, wytwarzane przez lampy elektronowe i układy półprzewodnikowe; mniej niż inne fale radiowe ulegają ugięciom i słabiej są pochłaniane przez jonosferę (stąd wykorzystywane do łączności dalekozasięgowej i w radiolokacji); silnie pochłaniane przez organizmy żywe (stosowane w piecykach mikrofalowych).

podczerwień (promieniowanie cieplne): $\lambda \sim 0,76 \mu\text{m} \div \text{mm}$ ($\nu \sim 10^{10} \div 10^{14}$ Hz), źródłem podczerwieni są ciała o temperaturze $T > 3\text{K}$ (stąd nazwa promieniowanie cieplne) lub pobudzone do świecenia gazy; zakres $\lambda \sim 0,76 \mu\text{m} \div 20 \mu\text{m}$ nazywany bliską

podczerwienią ma właściwości fizyczne podobne do światła widzialnego (odpowiednio pozostał zakres jest tzw. podczerwienią daleką).

światło widzialne: $\lambda \sim 0,38 \mu\text{m} \div 0,76 \mu\text{m}$ ($\nu \sim 10^{15}$ Hz), zakres fal wyróżniony przez zdolność detekcyjną oka ludzkiego; mechanizm wytwarzania światła widzialnego podobnie jak podczerwieni i nadfioletu wiąże się ze stanami energetycznymi atomów.

ultrafiolet: $\lambda \sim 10 \text{ nm} \div 0,38 \mu\text{m}$ ($\nu \sim 10^{16}$ Hz), silnie pochłaniany przez tlen i ozon oraz przez ludzką skórę.

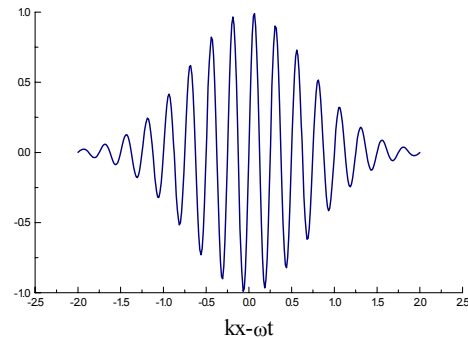
promieniowanie X (rentgenowskie): $\lambda \sim 0,01 \text{ nm} \div 10 \text{ nm}$ ($\nu \sim 10^{19}$ Hz), wytwarzane min. przez szybkie elektrony uderzające w materię stałą (w wyniku ich wyhamowywania), co jest wykorzystywane w lampach rentgenowskich; słabo pochłaniane przez materię (stąd ich wykorzystanie do prześwietlania przedmiotów).

promienie gamma (γ): $\lambda < 0,1 \text{ nm}$ ($\nu > 10^{19}$ Hz), wytwarzane w obrębie jądra atomowego i procesów jądrowych; traktowane jako strumień cząstek (fotonów).

Ponieważ własności światła widzialnego, bliskiej podczerwieni i bliskiego nadfioletu są podobne, łączny ich zakres nazywa światłem.

Fale niemonochromatyczne

Fala monochromatyczna (harmoniczna) jest pewną idealizacją zakładającą, że czas jej trwania i odległość występowania jest nieskończona. W rzeczywistych sytuacjach spotyka się fale w postaci impulsów (takie jak np. obok na rysunku), które jednak lokalnie bardzo dobrze przypominają falę monochromatyczną. Prędkość przemieszczania się takiego impulsu nazywana jest prędkością grupową i często ma inną wartość niż prędkość fazowa fali.



Natężenie pola elektrycznego \mathbf{E} płaskiej monochromatycznej fali elektromagnetycznej opisuje funkcja: $\mathbf{E} = \mathbf{A} \cos(\omega t - kx + \varphi)$. Fale niemonochromatyczne zapisuje się jako złożenie (superpozycja) fal monochromatycznych o różnych częstościach w postaci całki:

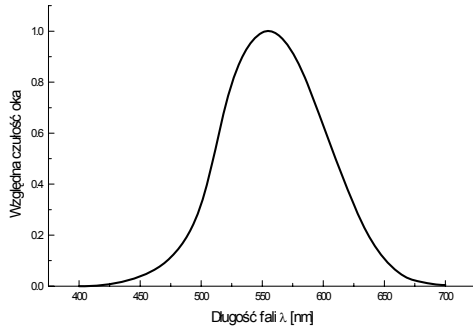
$$\mathbf{E} = \int \mathbf{A}_{\omega} \cos(\omega t - kx + \varphi_{\omega}) d\omega,$$

gdzie amplitudy A_{ω} i fazy φ_{ω} są różne dla różnych częstości. Rozkład amplitud A_{ω} względem częstości nazywa się widmem promieniowania. Ponieważ pomiędzy częstością ω a długością fali istnieje bezpośrednia zależność $\lambda = 2\pi c/\omega$, widmo promieniowania przedstawia się również jako zależność analogicznych amplitud A_{λ} od długości fali.

Rzeczywiste źródła światła nie wytwarzają idealnych fal monochromatycznych. Przykładowo Słońce wysyła jednocześnie fale elektromagnetyczne ze wszystkich zakresów, tzn. widmo jego promieniowania obejmuje wszystkie długości fal. Maksimum widma słonecznego przypada na zakres światła widzialnego. Jest to jednocześnie promieniowanie słabo pochłaniane w atmosferze natomiast rozpraszane, co powoduje, że Ziemia jest oświetlana z całej powierzchni nieba. Nic dziwnego, że ten zakres promieniowania jest widziany przez oko ludzkie i większości zwierząt. Zarówno wytwarzanie jak i detekcja fal elektromagnetycznych z różnych zakresów jest dokonywana z wykorzystaniem różnych mechanizmów. Każdy detektor reaguje na fale tylko z pewnego zakresu a jego czułość jest zależna od długości fali. Zatem sygnał dochodzący z detektora promieniowania jest zależny od przekrywania się widma mierzonych fal elektromagnetycznych z widmem czułości detektora.

Widzenie światła

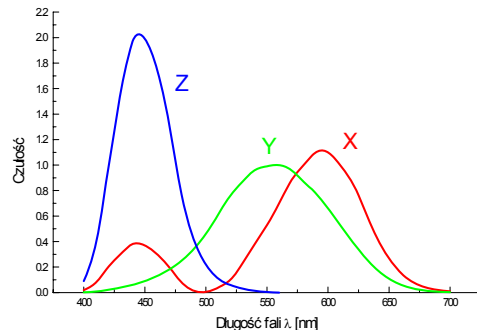
Naturalnym detektorem fal elektromagnetycznych jest oko ludzkie, które postrzega fale z zakresu od długości $\lambda \approx 3,8 \cdot 10^{-7} \text{ m} = 380 \text{ nm}$, czyli światła fioletowego poprzez światło niebieskie, zielone, żółte, pomarańczowe do światła czerwonego i ciemnoczerwonego o



długości do $\lambda \approx 760 \text{ nm}$. Widmo czułości oka przedstawione jest w przybliżeniu na wykresie obok. Czułość oka jest największa dla fal $\lambda \approx 550 \text{ nm}$, co oznacza, że fala monochromatyczna o tej długości daje wrażenie znacznie jaśniejszej niż fale monochromatyczne o innych długościach niosących taką samą energię. Granice zakresu widzialnego są przybliżone, gdyż o postrzeganiu światła decyduje energia fali (jasność), ale również wpływ ma budowa osobnicza oka, jego ewentualne wady, wiek, zmęczenie etc. Należy

także pamiętać, że wrażenie widzenia obrazu powstaje w mózgu na podstawie sygnałów dochodzących z obu oczu.

Gdyby na dnie oka występował tylko jeden rodzaj receptorów, widzielibyśmy obrazy czarno-białe (w różnych odcieniach szarości) bez wrażenia koloru. Tak się dzieje przy słabym oświetleniu (widzenie nocne), gdzie światło jest odbierane przez bardzo czuły układ znajdujący się w pręcikach. Przy silnym oświetleniu pręciki są zasłanianie (aby się nie uszkodziły) a widzenie odbywa się za pośrednictwem receptorów w czopkach. W czopkach znajdują się trzy różne detektory o maksimach czułości w obszarach błękitu, oranżu i czerwieni (oznaczonych odpowiednio symbolami Z, Y i X na przybliżonym wykresie ich widm czułości przedstawionym obok). Sumaryczny sygnał docierający do mózgu od tych detektorów daje wrażenie jasności, zaś stosunek sygnałów (X:Y:Z) daje wrażenie barwy. Oznacza to, że do wywołania wrażenia dowolnej barwy postrzeganej przez ludzi wystarczą trzy źródła światła niezależnie pobudzające detektory X, Y i Z (barwa biała wiąże się z jednakowym pobudzeniem wszystkich detektorów a czarna z brakiem sygnału). Wykorzystuje się to min. w telewizji (system trzech kolorów RGB - red, green, blue, czyli czerwony, zielony i niebieski) a także druku (CMYK - cyan, magenta, yellow, black, czyli turkusowy, różowy, żółty i czarny).



Mechanizm widzenia barw dopuszcza, że źródła światła o różnym widmie dają takie same wrażenie barwy (są to tzw. metamery). Oświetlone nimi obiekty mogą mieć jednak różne kolory, gdyż barwa światła odbitego (rozproszonego) wynika z pochłaniania fal o pewnych długościach, (które mogą akurat nie występować w jednym z metamerów). Brak jednego lub dwóch detektorów w czopkach skutkuje mniejszą rozróżnialnością barw (barwy różne przy trzech detektorach mogą być metamerami dla dwóch detektorów) a w przypadku braku wszystkich trzech detektorów widzenie staje się czarno-białe (z wykorzystaniem pręcików). Wady takie nazywa się daltonizmem. Z kolei brak lub uszkodzenie pręcików uniemożliwia widzenie przy słabym oświetleniu (w nocy), co nazywane jest kurzą ślepotą.

6. Interferencja światła

Natężenie światła:

Częstotliwość fali świetlnej jest tak duża, że nie istnieje detektor zdolny do podążania za zmianami pola elektrycznego. Dlatego detektory światła mierzą nie wartość pola a uśredniony po czasie jego kwadrat. Wielkością mierzona proporcjonalną do energii fali jest

$$I = \langle E^2 \rangle / Z_0,$$

nazywane natężeniem światła (dokładniej: jest to energia padająca na element powierzchni w jednostce czasu). Z_0 jest stałą zaś $\langle \rangle$ oznacza uśrednienie po czasie, w którego wyniku

między innymi uzyskuje się $\langle \cos(\omega t + \varphi) \rangle = 0$, $\langle \cos^2(\omega t + \varphi) \rangle = \frac{1}{2}$. W rezultacie, jeśli fala

jest monochromatyczna, czyli $E = A \cos(\omega t - kx + \varphi)$, to jej natężenie wynosi $I = A^2 / 2Z_0$.

Nakładanie się dwóch fal

Jeśli dwie fale monochromatyczne, opisywane wzorami: $E_1 = A_1 \cos(\phi_1)$, $E_2 = A_2 \cos(\phi_2)$, (gdzie $\phi_1 = \omega_1 t - k_1 x + \varphi$, $\phi_2 = \omega_2 t - k_2 x + \varphi$), o natężeniach odpowiednio I_1 i I_2 , nałożą się w pewnej płaszczyźnie $x = \text{const}$, to:

wypadkowe pole elektryczne E jest sumą pól składowych: $E = E_1 + E_2$;

wypadkowe natężenie światła I nie jest prostą sumą natężeń, gdyż:

$$I = \langle E^2 \rangle / Z_0 = \langle (E_1 + E_2)^2 \rangle / Z_0 = \langle (E_1)^2 + (E_2)^2 + 2E_1 E_2 \rangle / Z_0 = I_1 + I_2 + 2\langle E_1 E_2 \rangle / Z_0$$

Zatem pojawia się dodatkowy czynnik $2\langle E_1 E_2 \rangle$, który da się zapisać w postaci:

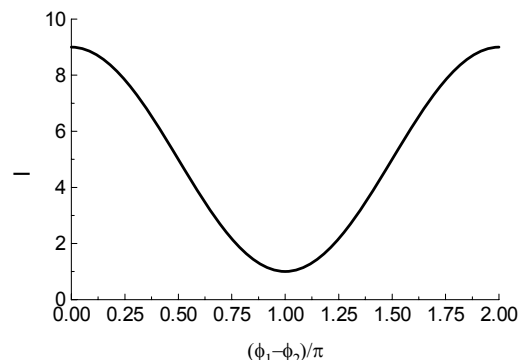
$$\langle 2E_1 E_2 \rangle = \langle 2A_1 A_2 \cos \phi_1 \cos \phi_2 \rangle = A_1 A_2 \langle \cos(\phi_1 + \phi_2) + \cos(\phi_1 - \phi_2) \rangle$$

Obie fazy ϕ_1 i ϕ_2 są liniowymi funkcjami czasu, więc wartości średnie z cosinusa ich sumy i różnicy powinny się zerować i wówczas natężenie wypadkowe jest sumą natężeń: $I = I_1 + I_2$.

Jednakże, gdy częstości obu fal składowych są takie same ($\omega_1 = \omega_2$), to różnica faz jest niezależna od czasu: $\phi_1 - \phi_2 = k_1 x - k_2 x$ i wówczas wypadkowe natężenie światła wynosi:

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\phi_1 - \phi_2).$$

Natężenie wypadkowe I może być większe lub mniejsze od sumy natężeń fal składowych w zależności od znaku funkcji cosinus. Dla przypadku jednakowych natężeń fal składowych $I_1 = I_2$ wypadkowe natężenie może być równe zero, gdy $\phi_1 - \phi_2 = \pi$ (fale się wygaszą), lub być dwa razy większe niż suma natężeń, gdy $\phi_1 - \phi_2 = 0$ (fale się wzmacniają). Wykres obok przedstawia wartość wypadkową natężenia dla fal składowych o różnych natężeniach $I_1 = 4$ i $I_2 = 1$ w zależności od niezmienniczej w czasie różnicy faz między nimi.

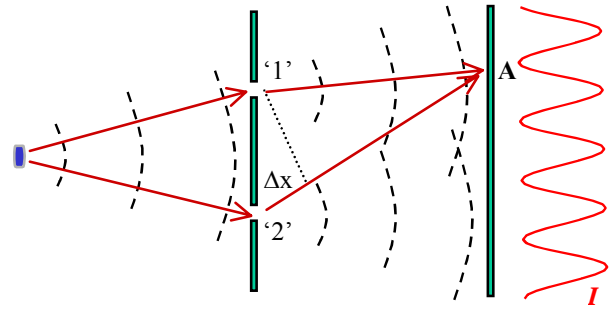


Interferencja

Nałożenie się dwóch fal, w którego wyniku wypadkowe natężenie nie jest sumą natężeń nazywa się interferencją a człon $2\sqrt{I_1 I_2} \cos(\phi_1 - \phi_2)$ nazywa się członem interferencyjnym.

Interferencję można obserwować w układzie przedstawionym na rysunku. Źródła kulistych fal monochromatycznych '1' i '2' (o jednakowych fazach) oświetlają płaski ekran. Droga, jaką musi przebyć fala do punktu A na ekranie od źródła '1' jest krótsza o Δx od drogi fali ze źródła '2'. Dlatego w punkcie A pojawi się między obiema falami różnica faz $\phi_1 - \phi_2 = -k\Delta x$.

Różnica faz jest inna w każdym punkcie i dlatego w wyniku interferencji obserwuje się lokalnie maksima i minima wypadkowego natężenia światła, czyli tzw. prążki interferencyjne. Należy podkreślić, że całkowita energia jest zachowana a skoro w jednych miejscach fale się wygaszają, to w innych muszą się wzmacnić. Zatem interferencja powoduje redystrybucję rozkładu energii (natężenia) fali.



Spójność światła

W rzeczywistości nie istnieją dwa różne źródła, których fazy zawsze byłyby identyczne, czyli różnica faz $\phi_1 - \phi_2$ nie zawsze jest stała w czasie. Gdy zmiany różnicy faz są na tyle duże, że obserwacja interferencji jest niemożliwa, źródła obu fal składowych nazywa się niespójnymi. Źródła są spójne wtedy, gdy różnica faz jest na tyle stabilna, że obserwuje się ich interferencję. Czyli do wystąpienia interferencji potrzeba spójnych fal o tej samej częstotliwości. Dlatego min. na stole oświetlonym przez dwie lampy nie widać prążków interferencyjnych.

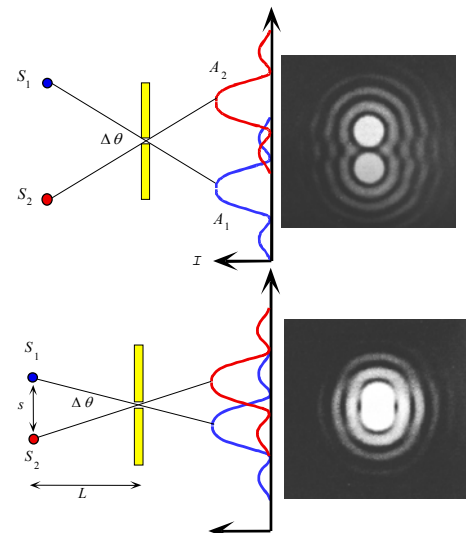
Dyfrakcja

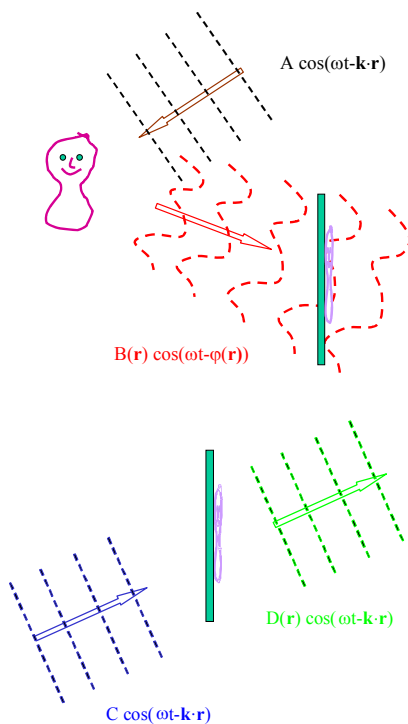
Czoło fali można traktować jako zbiór punktowych źródeł fal. Każde z takich źródeł wytwarza falę kulistą, która się nakłada na fale wytworzone przez pozostałe źródła. Kierunek rozchodzenia się czoła fali jest zdeterminowany przez interferencję takich fal składowych i dlatego czoło fali wypadkowej jest prostopadłe do kierunku przesuwania się fali. Powierzchnie wyznaczone przez kolejne położenia czoła fali są powierzchniami stałej fazy.

Powyższy model propagacji fal nazywany jest modelem Huygensa (czyt.: *hojhensa*). Pozwala on zrozumieć, co się dzieje, gdy na drodze fali znajduje się przeszkoda, np. w postaci nieprzeźroczystej przesłony, jak na rysunkach obok. Składowe fale kuliste promieniujące w obszar za przesłoną nie są wówczas

wygaszane interferencyjnie, w przeciwieństwie do przypadku, gdy nie ma przesłony. Dlatego fala wypadkowa ulega ugięciu na krawędzi przesłony. Zjawisko to nazywane jest dyfrakcją fali.

Dyfrakcja powoduje między innymi, że obraz punktu w rzeczywistym układzie optycznym jest rozmyty (każda soczewka ma skończoną szerokość, czyli należy ją traktować jak umieszczoną w otworze nieprzeźroczystej przesłony). Dlatego obrazy dwóch punktów będą widoczne jako niezależne tylko, gdy są odpowiednio od siebie odległe (górny rysunek obok), natomiast gdy są blisko siebie stają się nierozróżnialne (dolny rysunek obok). Ogranicza to rozdzielczość układów optycznych takich jak mikroskopy czy teleskopy.





Holografia

Płaska fala monochromatyczna $A \cos(\omega t - \mathbf{k} \cdot \mathbf{r})$ ma niezależną od położenia amplitudę A (i natężenie) oraz prostą zależność fazy od położenia. W ogólniejszym przypadku fali monochromatycznej, która nie jest falą płaską:

$$B(\mathbf{r}) \cos(\omega t - \varphi(\mathbf{r})),$$

zarówno amplituda $B(\mathbf{r})$ jak i faza $\varphi(\mathbf{r})$ może być złożoną funkcją położenia \mathbf{r} . Przykładowo w fali odbitej od obiektu oświetlonego falą płaską (rysunek) amplituda niesie informację o przestrzennej zdolności odbijającej obiektu zaś faza określając odległość, jaką przebywa fala, niesie informację o kształcie przestrzennym tego obiektu.

Zwykle zdjęcia zapisują natężenie światła tworząc dwuwymiarowy obraz rozkładu jasności proporcjonalny do natężenia światła odbitego od obiektu $\mathcal{I} = \zeta |B(\mathbf{r})|^2$. Po oświetleniu wywołanego zdjęcia fala przechodząca (jak również odbita) ma rozkład natężenia proporcjonalny do jasności obrazu $D(\mathbf{r}) = \zeta |B(\mathbf{r})|^2 C$ lecz fazę fali oświetlającej zdjęcie. W ten sposób gubiona jest informacja o kształcie obiektu w trzecim wymiarze.

Zapis fazy jest możliwy przez wykorzystanie zjawiska interferencji a ten sposób zapisu trójwymiarowości obiektów stosuje się w holografii. Hologram jest zdjęciem obrazu interferencyjnego światła o znanej fazie (zwanej falą odniesienia np. fali płaskiej $C \cos(\omega t - \mathbf{k} \cdot \mathbf{r})$) ze światłem odbitym od obiektu fotografowanego (falą obrazową). Po wywołaniu hologramu otrzymujemy rozkład jasności proporcjonalny do obrazu interferencyjnego fali odniesienia z falą obrazową:

$$\mathcal{I} = \zeta |B(\mathbf{r})|^2 + \zeta |C|^2 + 2\zeta B(\mathbf{r})C \cos(\varphi(\mathbf{r}) - \mathbf{k} \cdot \mathbf{r}).$$

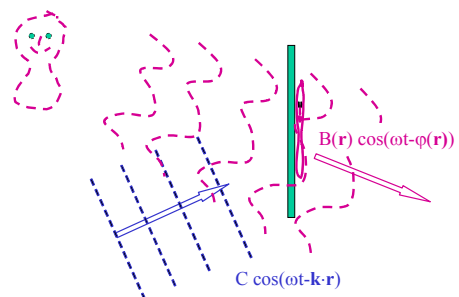
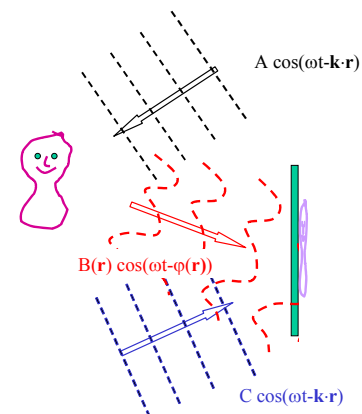
Oświetlając hologram taką samą falą, jak stosowana do jego zapisu fala odniesienia, fala przechodząca ma postać:

$$\mathcal{I} C \cos(\omega t - \mathbf{k} \cdot \mathbf{r}) = \{...\} + 2\zeta B(\mathbf{r})C^2 \cos(\varphi(\mathbf{r}) - \mathbf{k} \cdot \mathbf{r}) \cos(\omega t - \mathbf{k} \cdot \mathbf{r}) = \{...\} + \zeta B(\mathbf{r})C^2 \cos(\omega t + \varphi(\mathbf{r}) - 2\mathbf{k} \cdot \mathbf{r}) + \underline{\zeta B(\mathbf{r})C^2 \cos(\omega t - \varphi(\mathbf{r}))},$$

gdzie $\{...\} = [\zeta |B(\mathbf{r})|^2 + \zeta |C|^2] C \cos(\omega t - \mathbf{k} \cdot \mathbf{r})$ oraz skorzystano z wzoru na iloczyn cosinusów.

Podkreślony człon opisujący falę przechodzącą przez hologram jest z dokładnością do ζC^2 identyczny z falą obrazową. Oznacza to, że w świetle przechodzącym przez hologram odtworzeniu uległa fala zawierająca również fazę fali odbitej od obiektu fotografowanego niosąca informacje o trójwymiarowych własnościach obiektu.

Zapisując i odtwarzając hologram trzeba korzystać ze spójnych monochromatycznych źródeł światła (laserów). Są jednak specjalne rodzaje hologramów, które można oglądać w świetle białym. W takich hologramach obiekt odtwarzany jest przez falę o jednej długości i dlatego są one jednobarwne. Aby stworzyć wrażenie dowolnej barwy można jednak zapisać na jednej fotografii trzy hologramy odtwarzane przez fale o różnych długościach. Dla prostych obiektów można tworzyć hologramy tzw. syntetyczne, w których obraz interferencyjny jest wyliczany komputerowo.



7. Światło w ośrodkach materialnych

Współczynnik załamania

W ośrodkach materialnych dochodzi do oddziaływania pomiędzy falą elektromagnetyczną a ładunkami elektrycznymi, z których zbudowany jest ośrodek. Oddziaływanie to polega na tym, że pod wpływem pola elektrycznego fali elektromagnetycznej dochodzi do przemieszczania elektronów wewnątrz atomów i cząsteczek jak również jonów w strukturach jonowych. Przesuwanie ładunków względem ich położenia równowagi odbywa się z częstotliwością zmian pola elektrycznego. Drgające ładunki powracając do stanu początkowego oddają energię drgań do ośrodka (np. zwiększając jego temperaturę) lub wypromieniowują wtórną falę elektromagnetyczną. Przekazanie energii do ośrodka wiąże się z absorpcją fali elektromagnetycznej a ośrodek taki jest nieprzeźroczysty dla danej fali.

W przypadku, gdy drgające ładunki generują wtórną falę elektromagnetyczną, ośrodek taki może być postrzegany jako przeźroczysty, gdyż wtórna fala ma taką samą częstotliwość jak fala pierwotna. Obie fale: pierwotna i wtórna interferują ze sobą tworząc wypadkową falę elektromagnetyczną. Wypadkowa fala ma częstotliwość fali pierwotnej i efektywną prędkość n razy mniejszą od prędkości światła w próżni: $v = c/n$, gdzie n nazywane jest współczynnikiem załamania. Dla fali monochromatycznej:

$$\mathbf{E} = \mathbf{A} \cos(\omega t \pm kx + \varphi)$$

liczba falowa $k = 2\pi/\lambda$ wiąże się z prędkością i częstością fali: $k = \omega/v = (\omega/c)n$. Odpowiednio w próżni liczba falowa $k_0 = 2\pi/\lambda_0 = \omega/c$, czyli w ośrodkach materialnych długość fali λ jest krótsza niż długość fali w próżni $\lambda = \lambda_0/n$, natomiast liczba falowa jest większa: $k = k_0 n$.

Załamanie światła

Przejście fali do ośrodka, w którym porusza się z inną prędkością wiąże się ze zmianą kierunku rozchodzenia się tej fali. Jeśli płaska fala monochromatyczna w ośrodku o współczynniku załamania n_1 porusza się pod kątem α do osi y (rysunek) to pole tej fali można zapisać jako:

$$E_1 = A_1 \cos(\omega t - k_{1x}x - k_{1y}y + \varphi_1),$$

gdzie k_{1x} i k_{1y} są składowymi wektora falowego k_1 :

$$k_{1x} = k_1 \sin\alpha = k_0 n_1 \sin\alpha, \text{ oraz } k_{1y} = k_1 \cos\alpha = k_0 n_1 \cos\alpha.$$

Analogicznie po przejściu do ośrodka o współczynniku załamania n_2 fala ma postać:

$$E_2 = A_2 \cos(\omega t - k_{2x}x - k_{2y}y + \varphi_2),$$

gdzie $k_{2x} = k_2 \sin\beta = k_0 n_2 \sin\beta$, oraz $k_{2y} = k_2 \cos\beta = k_0 n_2 \cos\beta$.

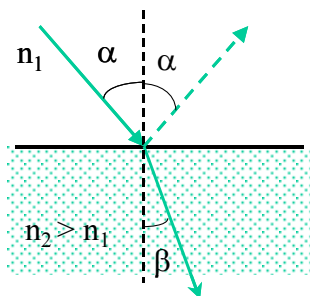
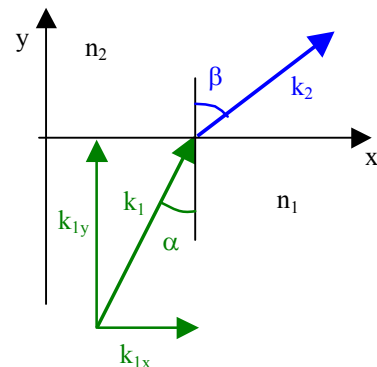
Zmiany w czasie pola po obu stronach granicy pomiędzy ośrodkami (dla $y = 0$) powinny odbywać się w identyczny sposób, czyli w każdej chwili czasu t stosunek pól

$$E_1(y=0)/E_2(y=0) = \text{const.}$$

Jest to możliwe tylko wtedy, gdy $k_{1x} = k_{2x}$, czyli gdy:

$$n_1 \sin\alpha = n_2 \sin\beta.$$

Jest to tzw. prawo załamania (Sneliusa). Kąt α jest kątem padania w ośrodku o współczynniku załamania n_1 a β jest kątem załamania w ośrodku o współczynniku n_2 . Przy padaniu światła na granicę dwóch ośrodków pojawia się też fala odbita. Kąt odbicia takiej fali jest równy kątowi padania, co można pokazać przeprowadzając identyczne rozumowanie jak powyżej.



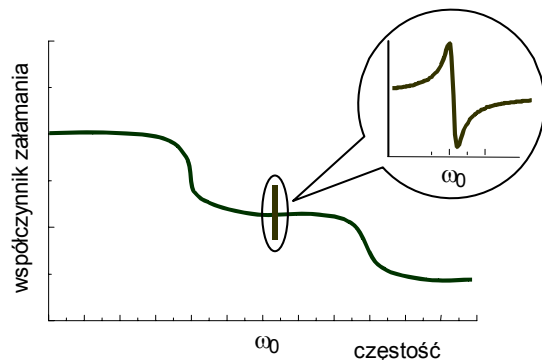
Rozpraszanie światła

Gdy granica pomiędzy ośrodkami o różnych współczynnikach załamania nie jest idealnie płaska, fale odbite i załamane z różnych miejsc na granicy rozchodzą się w różnych kierunkach. Dlatego światło odbijane i przechodzące przez chropowatą granicę dwóch ośrodków (np. zarysowane szkło) ulega rozproszeniu. Podobnie jest z ośrodkami niejednorodnymi składającymi się z obszarów o różnym współczynniku załamania (np. mgła składająca się z kropelek wody w powietrzu, czy mleko składające się z kropelek tłuszczu w wodzie): w wyniku wielokrotnych odbić i załamania dochodzi do rozproszenia światła i ośrodek może być całkowicie nieprzeźroczysty, mimo że nie pochłania światła (każdy jednorodny obszar jest przeźroczysty).

Charakter rozproszenia zależy od rozmiarów niejednorodności (np. średnicy kropelek wody w powietrzu czy szerokości rys na powierzchni szkła). Dla niejednorodności porównywalnych z długością fali światła, rozproszenie zależy od długości fali. Tak jest dla cieczy i gazów, w których termiczne fluktuacje gęstości przenoszą się na przestrzenne fluktuacje wartości współczynnika załamania. To z kolei powoduje rozpraszanie światła odwrotnie proporcjonalne do czwartej potęgi długości fali świetlnej ($\propto \lambda^{-4}$). Zatem rozproszenie to, zwane rozproszeniem Rayleigha, jest większe dla światła o krótszej fali, czyli w gazach i cieczach jest silniej rozpraszane światło niebieskie niż czerwone. Tym tłumaczy się niebieski kolor nieba i czerwień zachodzącego lub wschodzącego Słońca.

Dyspersja

Indukowanie wtórnej fali elektromagnetycznej w ośrodkach materialnych zależy od częstotliwości fali. Dla wolno zmieniającego się pola elektrycznego (fali o małej częstotliwości) przesunięcie ładunków jest większe niż dla pól szybkozmiennych (o dużych częstotliwościach). Wynika to z faktu, że czas potrzebny na przemieszczenie ładunku jest skończony i dla pól szybko zmieniających się ładunki nie zdążą osiągnąć położenia maksymalnych. Dlatego też wraz ze wzrostem częstotliwości wartość współczynnika załamania maleje i dla promieniowania rentgenowskiego prędkość fali w ośrodku jest taka sama jak w próżni tzn. $n = 1$. Dla każdego ośrodka istnieją charakterystyczne częstotliwości rezonansowe, o których to częstotliwościach fala elektromagnetyczna jest całkowicie absorbowana. Obecność absorpcji zmienia charakter oddziaływania z ośrodkiem powodując, że wraz ze wzrostem częstotliwości rośnie współczynnik załamania (tak jak wokół częstości ω_0 na rysunku obok). Dla światła współczynnik załamania zazwyczaj rośnie ze wzrostem częstotliwości i zawiera się w przedziale $1 < n < 3$.



Zależność współczynnika załamania od częstotliwości nazywana jest dyspersją. Dyspersja jest min. przyczyną rozszczepienia światła białego w pryzmacie i powstawania tęczy (fale o różnych częstotliwościach mają inny kąt załamania)

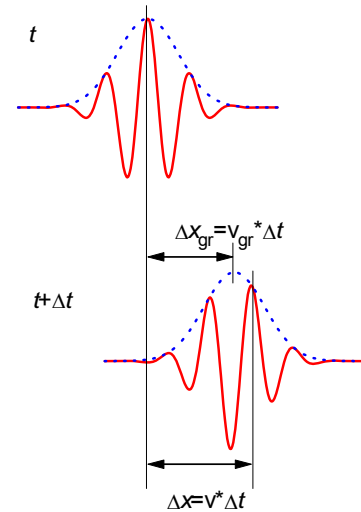
Prędkości fali elektromagnetycznej

W pobliżu częstości absorbowanych przez ośrodek współczynnik załamania może przyjmować zarówno bardzo duże jak i małe wartości, w tym może być nawet mniejszy od jedności. Współczynnik załamania o wartości mniejsza od jedności oznacza, że prędkość fali jest większa od prędkości światła w próżni: jeśli $0 < n < 1$, to $v = c/n > c$. Wbrew pozorom nie jest to sprzeczne z tym, że prędkość przesyłania energii nie może przekroczyć prędkości

światła w próżni. Prędkość $v = \omega/k = c/n$ jest prędkością fazową (prędkością przesuwania się powierzchni stałej fazy) i nie określa prędkości, z jaką przesyłana jest energia. Można się o tym przekonać analizując propagację impulsu światła.

W impulsie prędkość fazowa nie musi być taka sama jak prędkość grupowa, z jaką przesuwa się impuls.

Powierzchnia stałej fazy może się przesuwać wewnątrz impulsu. Na rysunku zilustrowano sytuację, w której prędkość fazowa v jest większa od prędkości grupowej v_{gr} impulsu. Oczywistym jest fakt, że prędkość, z jaką przesyłana jest energia (czy też informacja zapisana w impulsie) jest prędkością całego impulsu a nie prędkością, z jaką przesuwa się powierzchnia stałej fazy. W skrajnych sytuacjach może się nawet zdarzyć, że prędkość fazowa przyjmie przeciwny znak do prędkości grupowej (tzn. powierzchnia stałej fazy będzie się przesuwać do tyłu), co odpowiada ujemnej wartości współczynnika załamania.

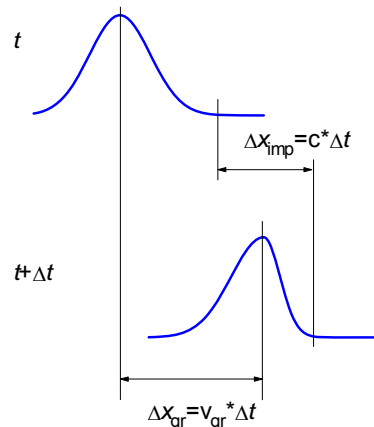


Prędkość grupowa

Prędkość grupową określa się przez pochodną liczby falowej po częstotliwości:

$$\frac{1}{v_{gr}} = \frac{\partial k}{\partial \omega} = \frac{\partial}{\partial \omega} \left(\frac{\omega}{c} n \right) = \frac{n}{c} + \frac{\omega}{c} \frac{\partial n}{\partial \omega} = \frac{1}{c} \left(n + \omega \frac{\partial n}{\partial \omega} \right)$$

Należy jednak mieć świadomość, że prędkość grupowa w pewnych specyficznych sytuacjach również nie opisuje prędkości impulsu. Jeśli kształt impulsu zmienia się wraz z propagacją, to prędkość początku impulsu jest różna od prędkości środka impulsu, opisywanej prędkością grupową (rysunek). Prędkość maksimum może być wtedy nawet większa od prędkości światła w próżni. Jednakże czoło impulsu porusza się zawsze z prędkością światła w próżni, niezależnie od ośrodka.



Sytuacja, w której maksimum impulsu porusza się szybciej niż światło w próżni może mieć miejsce w ośrodkach silnie absorbujących światło, które normalnie nie są przezroczyste. Jeśli jednak ośrodek ten będzie pompowany optycznie (tak jak ośrodki laserujące) może stać się przezroczysty. Współczynnik załamania n takiego ośrodka jak i jego pochodna po częstotliwości fali $\partial n / \partial \omega$ może przyjmować wówczas skrajnie duże wartości. Jeśli pochodna $\partial n / \partial \omega$ będzie ujemna to prędkość grupowa (prędkość maksimum impulsu) może okazać się większa od c albo nawet może być ujemna. W 2000 roku w NEC Research Institute w Princeton przepuszczono impuls światła przez odpowiednio spreparowane pary cezu. Maksimum impulsu światła przechodzącego przez sześciocentymetrową warstwę takiego ośrodka wyprzedzało maksimum impulsu biegnącego w próżni o 62 nanosekundy, czyli o około 20 metrów. Inną skrajną sytuację uzyskuje się, gdy pochodna $\partial n / \partial \omega$ jest dodatnia i ma dużą wartość. Impuls światła porusza się wtedy bardzo wolno. W 1999 roku w Rowland Institute for Science w Cambridge uzyskano w spreparowanych parach sodu impuls światła poruszający się z prędkością 17m/s (ok. 60 kilometrów na godzinę). W 2001 roku udało się taki impuls zatrzymać na niemal milisekundę, czyli czas, w którym światło w próżni przebywa drogę 300 kilometrów.

Absorpcja i emisja światła

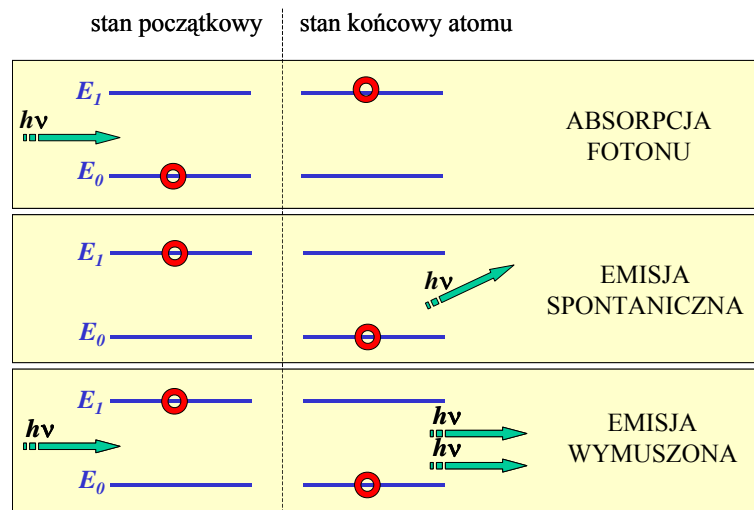
Fale mają prędkość, energię i pęd, o czym można się przekonać obserwując np. zniszczenia brzegu morza wywołane pracą fal. Mimo, że w klasycznym ujęciu fale i cząstki są różnymi pojęciami (np. trudno przyporządkować falam masę) to w opisie kwantowym fale i cząstki mają takie same właściwości (każda cząstka jest jednocześnie falą i na odwrót). Dlatego fale utożsamia się ze strumieniem poruszających się cząstek. W szczególności światło jest jednocześnie falą elektromagnetyczną jak i strumieniem cząstek nazywanych fotonami. Masa fotonu jest zerowa, gdyż fotony poruszające się z prędkością c przy niezerowej masie miałyby zgodnie ze wzorami relatywistycznymi między innymi nieskończony pęd i energię. Energia fotonu E_{fot} jest proporcjonalna do częstotliwości ν fali świetlnej: $E_{\text{fot}} = h\nu$ (h jest stałą Plancka) zaś pęd jest proporcjonalny do liczby falowej $p = hk$. Wytworzenie i pochłonięcie fotonu wiąże się z procesami wewnątrz atomów. Pochłonięcie fotonu przez atom, nazywane absorpcją, powoduje, że energia potencjalna atomu wzrasta o energię pochłoniętego fotonu. Jeśli atom w stanie podstawowym ma energię E_0 , to po absorpcji fotonu $h\nu$ jego energia wynosi $E_1 = E_0 + h\nu$ i atom jest w tzw. stanie wzbudzonym. Stany wzbudzone nie mogą mieć dowolnej energii i dlatego nie każdy foton może być pochłonięty przez dany atom. Rozkład częstości fotonów, które mogą być pochłonięte przez dany atom nazywany jest widmem absorpcji. Widmo absorpcji jest charakterystyczne dla każdego atomu.

Atom w stanie wzbudzonym nie przebywa długo i po pewnym czasie dążąc do minimalizacji własnej energii przechodzi z powrotem do stanu podstawowego (średnia wartość tego czasu jest tzw. czasem życia w stanie wzbudzonym). Różnica energii pomiędzy stanem wzbudzonym a podstawowym $E_1 - E_0$ może być przekazana na wytworzenie fotonu o częstotliwości $\nu = (E_1 - E_0)/h$, może też np. zwiększyć energię kinetyczną, czyli temperaturę ośrodka. Wytwarzanie fotonów nazywa się emisją, a wyemitowane fotony mają energię i częstotliwość taką samą jak fotony pochłonięte. Stąd widmo emisyjne jest identyczne z absorpcyjnym.

Emisja fotonu zachodząca po czasie w przybliżeniu równym czasowi życia stanu wzbudzonego nazywana jest spontaniczną.

Emisję fotonu można jednak wymusić wcześniej. Czynnikiem wymuszającym emisję fotonu jest inny foton o takiej samej częstotliwości. Proces emisji wymuszany przez inny foton nazywany jest emisją wymuszoną lub stymulowaną. Wyemitowany

w ten sposób foton ma nie tylko identyczną częstotliwość, co foton wymuszający emisję, ale także kierunek rozchodzenia się i fazę. W rezultacie światło wytworzone w procesie emisji wymuszonej składa się z uporządkowanego strumienia fotonów, podczas gdy emisja spontaniczna wytwarza światło rozchodzące się w różnych kierunkach i niespójne.



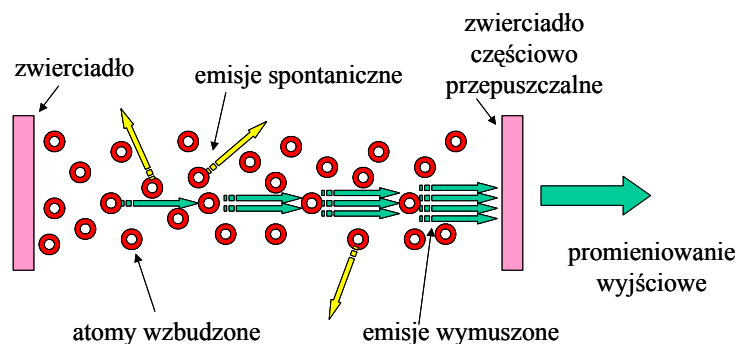
Lasery

Światło z większości źródeł (w tym np. z żarówki) powstaje w wyniku emisji spontanicznej. Lasery są natomiast źródłami światła, które zostało wytworzone w wyniku emisji wymuszonej (termin laser jest skrótem, w którym ostatnie trzy litery powstały z ang. *stimulated emission of radiation*, czyli wymuszonej emisji promieniowania). Dzięki temu

promieniowanie laserowe jest monochromatyczne i spójne, a także rozchodzące się w jednym kierunku. Cechy te pozwalają min. na zogniskowanie światła na bardzo małej powierzchni i dzięki temu uzyskaniu gigantycznych gęstości mocy fali świetlnej. Wykorzystuje się to min. w precyzyjnej mikroobróbkę materiałów (np. do cięcia materiałów, zgrzewania itp.) i medycynie (skalpele, mikrochirurgia dna oka itp.). Monochromatyczność i spójność laserów jest powodem ich powszechnego wykorzystywania min. w telekomunikacji światłowodowej, metrologii i holografii.

W celu uzyskania promieniowania laserowego należy atomy doprowadzić do stanu wzbudzonego i spowodować, aby emisja zachodziła w sposób wymuszony. Proces wzbudzania atomów (tzw. pompowania, czyli zwiększania ich energii) może być wywołany zarówno przez oświetlanie ich ze źródeł klasycznych lub innych laserów, jak również przez przepływający prąd elektryczny czy np. w wyniku reakcji chemicznych. Pompowanie powinno wzbudzić ponad połowę wszystkich atomów biorących udział w akcji laserowej. W przeciwnym przypadku, wyemitowane fotony zostaną pochłonięte przez atomy będące w stanie podstawowym. Innymi słowy, aby ośrodek był przezroczysty dla fotonów, to więcej atomów musi być gotowa do ich emisji niż do ich absorpcji. Stan, w którym więcej atomów jest w stanie wzbudzone niż w stanie podstawowym nazywany jest inwersją obsadzeń. Inwersja obsadzeń nie jest stanem normalnym, w którym większość stanowią cząstki o mniejszej energii. Odwrócenie, czyli inwersja normalnego rozkładu energii cząstek interpretowane jest jako stan o ujemnej temperaturze.

Jeżeli już doprowadzi się do inwersji obsadzeń, należy zadbać, aby emisja odbywała się w odpowiednim kierunku i była procesem wymuszonym. W tym celu ośrodek laserujący umieszcza się pomiędzy dwoma zwierciadłami, czyli w tzw. rezonatorze optycznym (rysunek). Rezonator taki uprzywilejowuje jeden kierunek rozchodzenia się fotonów. Odbijające się od obu zwierciadeł fotony wymuszają emisję kolejnych fotonów powodując lawinowy wzrost natężenia światła. Fotony emitowane spontanicznie pod kątem opuszczają rezonator i nie wpływają na proces emisji w kierunku wyjściowym. Rezonator zapewnia wzmocnienie promieniowania zarówno w wyróżnionym kierunku jak i o określonej częstotliwości.

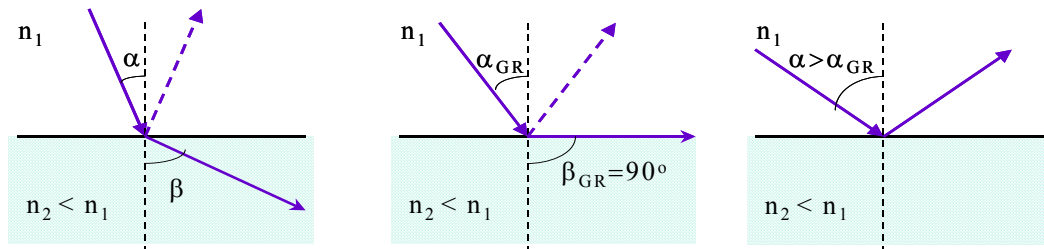


Moc promieniowania laserów może wydawać się nieduża w porównaniu z klasycznymi źródłami światła, takimi jak żarówki i lampy. Promieniowanie laserowe jest jednak wysoce monochromatyczne i zazwyczaj w postaci wąskiej wiązki o małej rozbieżności. Stwarza to duże zagrożenie szczególnie dla oczu. Promieniowanie laserowe z zakresu widzialnego i bliskiej podczerwieni (o długości fali $0,4\ \mu\text{m} < \lambda < 1,4\ \mu\text{m}$) jest silnie ogniskowane na dnie oka, gdzie tworzy plamkę o małej średnicy (rzędu setnych części milimetra). Jest to wyjątkowo niebezpieczne, gdyż o uszkodzeniu tkanki decyduje gęstość energii, która w wyniku zogniskowania na tak małych powierzchniach wzrasta około 100 000 razy. Dlatego w praktyce za bezpieczne uważa się lasery o mocy mniejszej niż pojedyncze miliwaty (zależy to oczywiście od długości fali, szerokości wiązki itp.). Promieniowanie laserowe z zakresu nadfioletu i dalszej podczerwieni jest absorbowane przez rogówkę oka (mogąc spowodować zapalenie rogówki) lub przez soczewkę (może wywołać pewne formy katarakty).

8. Światłowodowy

Całkowite wewnętrzne odbicie

Przechodzenie światła do drugiego ośrodka i obserwowany przy tym efekt załamania może się odbywać tylko w pewnym zakresie kątów padania. Granicznym kątem padania jest kąt α_{GR} , dla którego fala załamana porusza się wzdłuż granicy rozdzielającej oba ośrodki (kąt załamania $\beta_{GR} = 90^\circ$), co po wstawieniu do wzoru na prawo załamania prowadzi do wyniku: $\sin \alpha_{GR} = n_2/n_1$. Dla kątów padania $\alpha > \alpha_{GR}$ wartość $n_1/n_2 \sin \alpha$ jest większa od jedności, prawo załamania przestaje obowiązywać i światło ulega całkowitemu wewnętrznemu odbiciu.

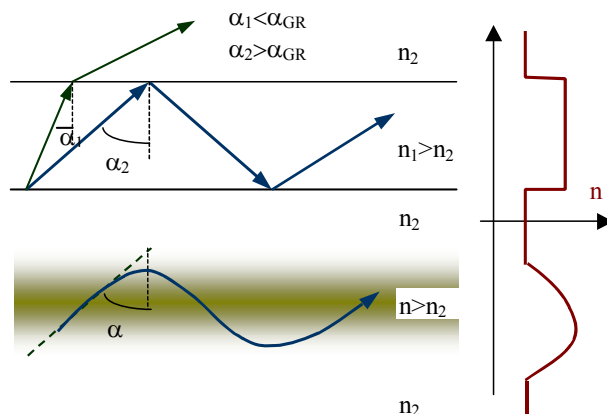


Zjawisko całkowitego odbicia występuje, gdy $n_1/n_2 > 1$, czyli gdy fala pada z ośrodka o większym współczynniku załamania na ośrodek o mniejszym współczynniku załamania np. od strony szkła o współczynniku załamania $n_1 > 1$ na granicę z powietrzem, którego współczynnik załamania $n_2 \approx 1$.

Zjawisko całkowitego wewnętrznego odbicia wykorzystuje się w układach optycznych m.in. w elementach pełniących rolę zwierciadeł. Wykonuje się wtedy pryzmaty o powierzchniach polerowanych pod odpowiednim kątem, aby światło wchodząc do pryzmatu ulegało całkowitemu wewnętrznemu odbiciu na tych powierzchniach. W szczególności dla pryzmatu powstałego ze ściętego narożnika sześcianu światło po kilkukrotnym całkowitym odbiciu wewnętrznym wraca dokładnie w kierunku, z którego padało (dla zwierciadeł płaskich ma to miejsce tylko, gdy światło pada prostopadłe na zwierciadło). Zestawy takich pryzmatów zostały m.in. umieszczone na Księżycu jako zwierciadła odbijające impulsy laserowe wysyłane z Ziemi (dzięki temu wykonano pomiary odległości Księżyca od Ziemi z dokładnością do kilkunastu centymetrów). Zjawisko całkowitego wewnętrznego odbicia wykorzystywane jest także w światłowodach do transmisji światła z wyjątkowo małymi stratami na duże odległości.

Struktura światłowodów

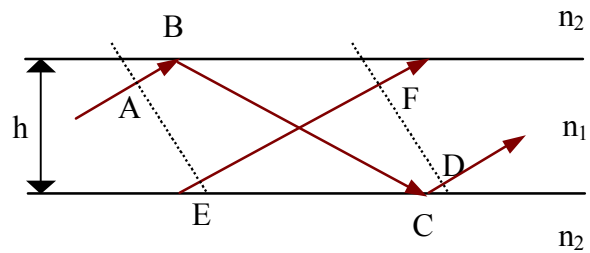
Umieszczony w środku rdzeń światłowodu ma większy współczynnik załamania niż otoczenie. Promień świetlny w światłowodzie padający pod kątem większym od kąta granicznego ulega całkowitemu wewnętrznemu odbiciu i zostaje uwięziony w obszarze rdzenia. Do uzyskania takiego efektu nie jest jednak niezbędna skokowa granica, na której występuje odbicie. Jeżeli współczynnik załamania będzie się zmieniał w sposób ciągły, to promienie będą zakrzywiane w stronę obszaru o większym współczynniku załamania. W identyczny sposób tłumaczyć można działanie soczewek skupiających: w



środku soczewka jest grubsza, co jest równoznaczne z większym przyrostem fazy fali, czyli średnia wartość współczynnika załamania jest większa niż na brzegu soczewki.

Mody światłowodowe

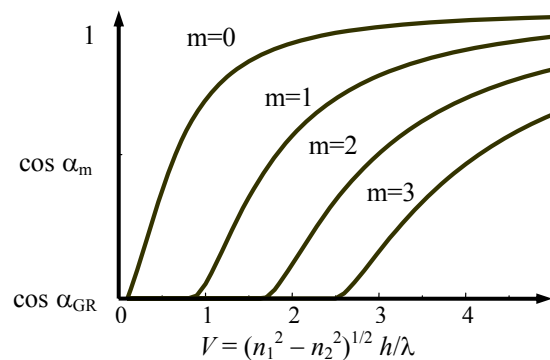
Prowadzenie fali świetlnej wewnątrz światłowodu wymaga, aby kąt padania był większy od kąta granicznego. Nie jest to jednak warunek wystarczający. Jeśli fala kolejno odbita na dwu granicach (tor promienia ABCD na rysunku) nie będzie miała tej samej fazy, co fala, która przebyła inną drogę (tor promienia EF) to obie fale będą się wygaszać i energia fali będzie wypływać na zewnątrz światłowodu. Dla propagacji światła wewnątrz światłowodu konieczna jest pozytywna interferencja fal płaskich odbijanych na granicach, czyli różnica faz pomiędzy nimi musi być całkowitą wielokrotnością 2π . Dla przypadku jak na rysunku warunek ten wynosi:



$$\left(\overline{AB} + \overline{BC} + \overline{CD} - \overline{EF} \right) \frac{2\pi}{\lambda} n_1 + \Delta\phi_B + \Delta\phi_C = 2\pi m, \quad m = 0, 1, 2, \dots$$

gdzie oprócz różnicy dróg optycznych (będących iloczynem odległości i liczby falowej) uwzględniona jest zmiana fazy $\Delta\phi$ powstająca przy odbiciu w punktach B i C. Analogiczny warunek można wprowadzić dla promieni zakrzywianych w falowodzie o nieskokowej zmianie współczynnika załamania światła. Uwzględnienie warunku interferencji powoduje, że dla każdej wartości liczby całkowitej m uzyskuje się co najwyżej jeden kąt α_m i w rezultacie liczba kątów α_m dla których propaguje się światło w światłowodzie jest skończona. Postać pola związana z jednym kątem α_m nazywana jest modem światłowodowym.

Liczba modów i odpowiadająca im wartość kąta α_m zależy od wielkości zwanej unormowaną częstotliwością $V = (n_1^2 - n_2^2)^{1/2} h / \lambda$, gdzie h jest grubością światłowodu. Wraz ze wzrostem wartości V rośnie liczba modów prowadzonych w światłowodzie (rysunek obok). W światłowodzie jednomodowym grubość h musi być mała w porównaniu z długością fali λ i mała musi być różnica współczynników załamania pomiędzy rdzeniem a otoczeniem $n_1^2 - n_2^2$. Dla światła widzialnego średnice rdzenia muszą być wówczas rzędu pojedynczych mikrometrów.



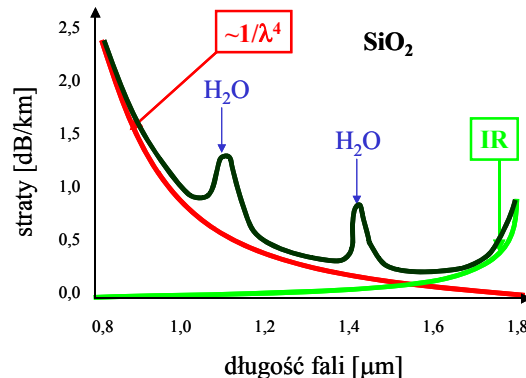
Wytwarzanie światłowodów jednomodowych wymaga bardziej zaawansowanych technologii. Światłowody takie mają jednak zalety w porównaniu ze światłowodami wielomodowymi. W światłowodzie wielomodowym każdemu z modów odpowiada inna prędkość fazowa i inna prędkość grupowa. Zatem światło wprowadzone do światłowodu wielomodowego jest przesyłane z różnymi prędkościami. W szczególności, gdy przesyłany jest impuls świetlny, to dociera on na końcu światłowodu wielomodowego wydłużony i zniekształcony. Również rozkład poprzeczny pola odbiega od rozkładu wejściowego, gdyż jest wynikiem interferencji wielu modów poprzesuowanych w fazie a ze względu na niejednorodność światłowodu różnice faz są przypadkowe. W światłowodach jednomodowych wydłużenie impulsów jest nieznaczne i wynika z dyspersji ośrodka. Odbywa się to jednak kosztem całkowitej energii

przenoszonej przez światłowód gdyż światło wewnątrz światłowodu jednomodowego jest zlokalizowane w obszarze o szerokości rzędu najwyżej kilku mikrometrów (w światłowodach wielomodowych nawet setek mikrometrów), co utrudnia wprowadzenie do światłowodu światła o dużej mocy.

Światłowody telekomunikacyjne

W zależności od przeznaczenia wytwarza się różne typy światłowodów. Najbardziej znanym są światłowody włókniste wykorzystywane w telekomunikacji światłowodowej. Mają one strukturę cylindryczną, w której współczynnik załamania zmienia się wzdłuż promienia. Część wewnętrzna nazywa się rdzeniem a część zewnętrzna płaszczem. Są to włókna szklane (szkło kwarcowe SiO_2) o średnicy 125 mikrometrów i z kilkumikrometrowym rdzeniem o wyższym współczynniku załamania, uzyskanym przez domieszkowanie szkła.

Prowadzenie światła w cylindrycznych dielektrykach analizowane było teoretycznie już na początku XX wieku, lecz drogę do zastosowań otworzyło wytworzenie w firmie Corning Glass Works (USA) w 1970 roku światłowodów ze szkła kwarcowego o niskim tłumieniu. Straty w wyprodukowanych wówczas światłowodach wynosiły ok. 20 dB/km, co oznacza, że po przejściu przez 1 km światłowodu światło ma 1% mocy wejściowej. Produkowane obecnie światłowody jednomodowe stosowane w telekomunikacji mają straty poniżej 0,2 dB/km dla światła o długości fali $\lambda = 1,55 \mu\text{m}$ (tzn. dla światła niewidocznego dla oka ludzkiego). Umożliwia to przesyłanie sygnałów świetlnych na odległości setek kilometrów bez konieczności wzmacniania (obecnie w telekomunikacji światłowodowej stosuje się odcinki rzędu 100 km). Dla światła o większych i mniejszych długościach fali rośnie tłumienność szkła. Dla fal dłuższych związane jest to z absorpcją szkła w podczerwieni (IR), a dla krótszych z rozpraszaniem $\propto 1/\lambda^4$. Dodatkowym źródłem strat jest absorpcja światła przez cząsteczki wody (H_2O), których wyeliminowanie jest bardzo trudne. Początkowo w telekomunikacji wykorzystywano światło o długości fali $\lambda = 1,3 \mu\text{m}$ ze względu na trudności z wytworzeniem źródeł światła (laserów półprzewodnikowych) dla światła $\lambda = 1,55 \mu\text{m}$.



W jednym włóknie światłowodowym można przysyłać niezależnie impulsy o różnych długościach fali, które dają się łatwo rozseparować na wyjściu (np. przy pomocy siatki dyfrakcyjnej). Dlatego pojemność linii transmisyjnej zwiększa się przez zwiększenie częstotliwości modulacji oraz przez zastosowanie światła o różnych długościach fali. Przyjmuje się, że w niedługiej przyszłości będzie wykorzystywane w pełni całe pasmo niskiej tłumienności w zakresie długości fal od ok. 1,28 μm do 1,63 μm , co pozwoli przysyłać informację z częstotliwością do 100 THz (10^{14} Hz).

Ograniczeniem przy zwiększaniu częstotliwości modulacji sygnału i liczby różnych długości fal są elektroniczne układy wzmacniające, które w klasycznych systemach składają się z detektora natężenia światła połączonego z elektronicznym wzmacniaczem zasilającym laser półprzewodnikowy, z którego zmodulowane światło wprowadzane jest do następnego odcinka światłowodu. Okazuje się jednak, że w światłowodach jest możliwe również wzmacnianie światła. Jony pierwiastków ziem rzadkich, na przykład erb i neodym, dodawane do szkła są odpowiednio stosowane do zbudowania optycznych wzmacniaczy i laserów

pompowanych optycznie. Pierwszy laser światłowodowy zbudowano w 1961 roku, wykorzystując rdzeń szklany domieszkowany neodymem o średnicy ok. 300 μm . W 1986 roku zbudowano laser w światłowodzie włóknistym domieszkowanym erbem, który może służyć jako źródło światła lub element wzmacniający sygnał przesyłany w światłowodach telekomunikacyjnych. Zastąpienie elektronicznych układów wzmacniających przez taki wzmacniacz optyczny pozwala na powiększenie pojemności linii transmisyjnej.

Inne rodzaje światłowodów

Wytwarzana są również włókniste światłowody wielomodowe o dużych średnicach rdzenia zarówno ze szkła jak i np. z polimerów. Wykorzystywane są one między innymi do oświetlania bądź przesyłania obrazów w wiązkach światłowodowych. Duże średnice rdzenia pozwalają na przesyłanie światła o dużej mocy.

Oprócz światłowodów włóknistych wytwarzane są światłowody o prostokątnym przekroju poprzecznym - światłowody paskowe lub światłowody płaskie. Są one zazwyczaj paskami lub warstwami na powierzchni płytek dielektrycznych (szklanych, kryształów elektrooptycznych, półprzewodników i in.). Ich tłumienność jest znacznie większa niż światłowodów włóknistych, ale i długości mniejsze (rzędu centymetrów), ograniczone rozmiarami podłoża. Światłowodami paskowymi są m.in. obszary aktywne laserów półprzewodnikowych. Wytwarzane są struktury światłowodów paskowych będących rezonatorami, przełącznikami optycznymi, modulatorami, głowicami pomiarowymi czujników różnych wielkości i in. Ze względu na miniaturyzację tych układów nazywa się je układami optyki scalonej. Szczególnie atrakcyjna wydaje się możliwość budowy scalonych układów na podłożu krzemowym, gdzie na jednej płytce połączone mogą być układy optyczne, mikromechaniczne i elektroniczne.