

W oku cyklonu rewolucji

Jakiś czas temu na Zjeździe Fizyków w Białymstoku wystąpiłam z referatem „Czy grozi nam rewolucja w nauczaniu?”. Uważam, co podziela też ze mną część kolegów, że spostrzeżenia i uwagi o stanie nauczania i propozycje na przyszłość były wtedy i nadal pozostają słuszne. Tymczasem w ciągu ostatnich dwóch lat sytuacja się radykalnie zmieniła. Znaleźliśmy się, nawet tego nie zauważając, w samym oku cyklonu edukacyjnej rewolucji. Wszystko się zmienia i to w zawrotnym tempie. I to nie tylko szkoła i jej otoczenie: rodzice, nauczyciele, edukatorzy, ale przede wszystkim sami uczniowie. Jest to pokolenie całkiem różne od swoich nawet tylko o 5 lat starszych kolegów. I to nie tylko dlatego, że obecnie uczniowie mają inne doświadczenia z życia codziennego, ale zmienił się (i stale dynamicznie się zmienia) sposób komunikacji między uczniami, język. Uczniowie mają inne priorytety, inaczej przeżywają emocje i inaczej je wyrażają. Pojawiły się bariery we wzajemnym rozumieniu ze starszym pokoleniem. Zwiększyły się różnice pomiędzy grupami uczniów.

W czasie trwania pandemii okazało się, że rola rodziców i nauczycieli jest kluczowa dla wykształcenia uczniów, przyszłych mieszkańców naszego kraju, naszej Ziemi. Dobrą lekcją z tego płynącą jest to, że rodzice musieli się zaangażować w proces edukacyjny i że bez ich istotnego wsparcia młody człowiek może nie mieć nawet szansy na prawidłowy rozwój. Smutny jest natomiast fakt, że nie wszyscy uczniowie mają takie możliwości. Tu będzie tkwiło źródło pogłębiających się nierówności społecznych.

Innym ważnym czynnikiem, niezbędnym dla prawidłowego rozwoju młodego pokolenia, są nauczyciele. Czy będą oni mieli czas i warunki na spełnienie swojej roli? Niestety to zależy od czynników politycznych, a te są raczej destrukcyjne. I znowu ten fakt będzie przyczyną nierówności szans edukacyjnych młodzieży, a co za tym idzie - również ekonomicznych.

Powtarzam, w udziale rodziców i nauczycieli chodzi o zapewnienie młodym ludziom szerokiej oferty możliwych dróg do wyboru, chodzi o szansę.

Czy potrafimy sprostać wyzwaniu i znaleźć wspólny język z młodymi ludźmi i zaproponować adekwatne metody nauczania? Arcytrudne zadanie, bo nie można zapominać, że fizyka i ogólnie nauka jako taka też doznała ogromnej akceleracji. Żyjemy w fascynującym czasie. Niewątpliwie osoby zajmujące się edukacją będą miały ręce pełne roboty, choć ostatnie lata nie zostały całkowicie zmarnowane, częściowo przygotowywano się do nowych wyzwań.

Z G.-M.



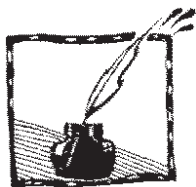
Contents

In the eye of the storm of revolution <i>Zofia Gołqb-Meyer</i>	1
A few words about epidemic modeling: SI, SIS, SIR and \mathcal{R}_0 <i>Ewa Gudowska-Nowak, Katarzyna Oleś, Bartłomiej Dybiec</i>	4
Letter to the Editor <i>Edward Malec i Andrzej Staruszkiewicz</i>	18
Editorial note.....	19
Light interference – when can we observe it? Part. I <i>Tomasz Kawalec</i>	20
Beauty captured in diamonds <i>Mariusz Mrózek</i>	30
What to read	46
3rd World Conference on Physics Education <i>Dagmara Sokołowska</i>	48



Spis treści

W oku cyklonu rewolucji <i>Zofia Gołąb-Meyer</i>	1
To i owo o modelowaniu epidemii: SI, SIS, SIR oraz \mathcal{R}_0 <i>Ewa Gudowska-Nowak, Katarzyna Oleś, Bartłomiej Dybiec</i>	4
List do redakcji <i>Edward Malec i Andrzej Staruszkiewicz</i>	18
Nota od redakcji	19
Interferencja światła – kiedy możemy ją zaobserwować? Część I <i>Tomasz Kawalec</i>	20
Piękno zamknięte w diamentach <i>Mariusz Mrózek</i>	30
Co czytać	46
III Światowa Konferencja Edukacji Fizyki <i>Dagmara Sokołowska</i>	48



To i owo o modelowaniu epidemii: SI, SIS, SIR oraz \mathcal{R}_0

Ewa Gudowska-Nowak,
Katarzyna Oleś, Bartłomiej Dybiec
Instytut Fizyki Teoretycznej
Uniwersytet Jagielloński

Modelowanie epidemiologiczne jest dobrze rozwiniętą dziedziną badań. Dzięki niemu możliwe jest lepsze zrozumienie mechanizmów rozprzestrzeniania się chorób, identyfikacja kluczowych elementów oraz ocena efektywności różnorodnych strategii interwencji. W najprostszych wariantach modele epidemiologiczne są modelami różniczkowymi, zakładającymi, że osobniki mogą znajdować się w dyskretnej (zależnej od modelu) liczbie stanów oraz są doskonale przemieszane. Oznacza to, że prawdopodobieństwo oddziaływania dwóch dowolnie wybranych osobników jest takie samo. Z jednej strony, w realnych sytuacjach, takie założenie nie zawsze jest spełnione – sieć kontaktów jest o wiele bardziej złożona. Zaś z drugiej strony pozwala ono na łatwiejsze konstruowanie licznych modeli epidemiologicznych i ich rozszerzeń.

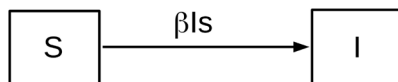
Podstawowe modele epidemiologiczne swoje nazwy zawdzięczają angielskim nazwom stanów, w jakich mogą znaleźć się osobnicy. W najprostszych przypadkach są to klasy S – zdrowy/podatny (*susceptible*), I – zarażony (*infected*) oraz R – ozdrowiały/uodporniony (*recovered*). Dodatkowo modele określają możliwe przejścia między stanami. Najprostsze i najpopularniejsze modele to modele: SI, SIS oraz SIR [4, 6, 7], które zostaną omówione w niniejszym artykule. W modelach tych osobnik zdrowy może zostać zakażony na skutek kontaktu z osobnikiem chorym, a pozostałe procesy (np. zdrowienie) mogą być spontaniczne. Zwykle zakłada się także, że proces epidemiczny zachodzi na tyle szybko, że można pominąć zjawiska demograficzne. Na podstawie modeli SI, SIS oraz SIR poprzez zwiększanie liczby możliwych stanów oraz przejść między nimi buduje się bardziej skomplikowane modele. Prowadzi to między innymi do modeli SIRS [4], SEIR [7] oraz wielu innych. Model SIRS odpowiada sytuacji, kiedy nie jest możliwe uzyskanie trwałej odporności, zaś w modelu SEIR choroba w początkowym okresie rozwija się w sposób utajony.

W niniejszym artykule prezentujemy wybrane własności najprostszych modeli epidemiologicznych SI, SIS oraz SIR. Przedstawiamy podstawowe informacje na temat odporności stadnej (populacyjnej) i powiązanego z nią bazowego współczynnika reprodukcji oraz omawiamy statystyczne własności testów diagnostycznych.

Model SI

Model SI jest odpowiednikiem modelu Verhulsta (logistycznego) w ekologii [7]. W modelu SI doskonale przemieszana populacja generalna o ustalonym rozmiarze N podzielona jest na osobników zdrowych (S) i zakażonych (I) ($N=S+I$).

Nie jest możliwe wyzdrowienie, zatem stan końcowy jest stanem, w którym wszyscy osobnicy są zarażeni. O szybkości zarażania decyduje liczba kontaktów między osobnikami zarażonymi i podatnymi, która zależy od średniej liczby kontaktów na jednostkę czasu, na jednego osobnika, które mogą prowadzić do zarażenia. Wielkość tę oznaczamy przez β . Jeśli w populacji jest I zarażonych osobników, to na jednostkę czasu dochodzi do $\beta \times I$ kontaktów, spośród których $\beta \times I \times s$ (gdzie $s = S/N$ jest podatną frakcją populacji) zachodzi z osobnikami podatnymi.



Rys. 1. Diagram dla modelu SI: możliwe stany wraz z szybkościami przejścia między nimi

Liczbę osobników zarażonych w chwili $t + \Delta$ można wyznaczyć na podstawie liczby zakażonych w chwili t

$$I(t + \Delta) = I(t) + I_{new}, \quad (1)$$

gdzie I_{new} jest liczbą nowych zakażeń w czasie Δ . W czasie Δ średnia liczba kontaktów zarażonych osobników, które prowadzą do zarażenia jest równa $\Delta \times \beta \times I$. Spośród $\Delta \times \beta \times I$ kontaktów $\Delta \times \beta \times I \times s$ zachodzi z osobnikami podatnymi i to właśnie one prowadzą do nowych zakażeń. Liczba nowych zakażeń jest równa

$$I_{new} = \Delta \times \beta \times I \times s. \quad (2)$$

Dzięki temu dostajemy

$$I(t + \Delta) = I(t) + \Delta \times \beta \times I \times s, \quad (3)$$

oraz

$$I(t + \Delta) - I(t) = \Delta \beta I s. \quad (4)$$

Dzieląc obustronnie przez Δ

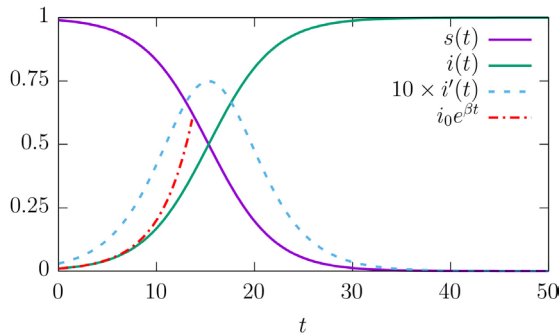
$$\frac{I(t + \Delta) - I(t)}{\Delta} = \beta I s, \quad (5)$$

oraz przechodząc z $\Delta \rightarrow 0$, otrzymujemy równanie opisujące model SI

$$\frac{dI}{dt} = \beta I s. \quad (6)$$

Prawa strona powyższego równania jest nieujemna, dlatego $I(t)$ jest niemalejąca – oznacza to, że liczba osobników zarażonych nie może maleć. Dzieląc powyższe równanie przez N oraz korzystając ze związku $s = 1 - i$, otrzymujemy równanie logistyczne [9, 10]

$$\frac{di}{dt} = \beta i(1 - i). \quad (7)$$



Rys. 2. Model SI: zależność zdrowej $s(t)$ oraz zarażonej $i(t)$ części populacji (linie ciągłe), zapadalności $i'(t)$ (linia przerywana) od czasu t dla $i_0=0,01$ oraz $\beta=0,3$. Dodatkowa linia (przerywano-kropkowana) przedstawia przybliżony rozmiar zarażonej części populacji od czasu w początkowym okresie epidemii: $i(t) \approx i_0 e^{\beta t}$

Rozwiązaniem równania (7) (dla warunku początkowego $i(0)=i_0$) jest funkcja logistyczna

$$i(t) = \frac{i_0 e^{\beta t}}{1 + i_0 (e^{\beta t} - 1)} = \frac{1}{1 + \frac{1 - i_0}{i_0} e^{-\beta t}}. \quad (8)$$

W granicy $t \rightarrow \infty$

$$\lim_{t \rightarrow \infty} i(t) = 1, \quad (9)$$

co oznacza, że wszyscy zostaną zarażeni (dla każdego i_0 oraz β). Początkowo (dla małych t) liczba chorych osobników rośnie wykładniczo

$$i(t) = \frac{i_0 e^{\beta t}}{1 + i_0 (e^{\beta t} - 1)} \approx i_0 e^{\beta t}. \quad (10)$$

Dzieje się tak dlatego, że początkowo w stanie I jest niewielka część populacji, dlatego $s(t) \approx 1$. W późniejszym okresie rośnie liczba zarażonych S , $s(t)$ staje się mniejsze od 1, znaczenie mają efekty nasyceniowe, a wzrost liczby zarażonych osobników przestaje być wykładniczy. W modelu SI liczba zarażonych osobników jest równa

$$I(t) = N \times i(t) = \frac{N_0 e^{\beta t}}{1 + \frac{N_0}{N} (e^{\beta t} - 1)}, \quad (11)$$

gdzie N_0 jest liczbą początkowo zarażonych osobników. Dla $t \approx 0$ obserwujemy wykładniczy wzrost $N(t) \approx N_0 e^{\beta t}$.

Zapadalność (zachorowalność) – liczba nowych przypadków w jednostce czasu Δ jest równa

$$I_{new} = I(t + \Delta) - I(t) \approx \frac{dI}{dt} \times \Delta = \frac{\Delta}{N} \frac{di}{dt}, \quad (12)$$

gdzie

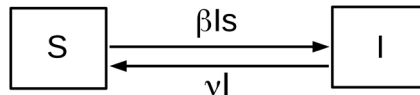
$$\frac{di}{dt} = \beta \frac{i_0 - 1}{i_0} \frac{e^{-\beta t}}{\left[1 + \frac{1 - i_0}{i_0} e^{-\beta t}\right]^2}. \quad (13)$$

Największy przyrost liczby zarażonych przypadków ma miejsce wtedy, gdy zapadalność jest maksymalna.

Na rysunku 2 przedstawiono przykładową krzywą logistyczną $i(t)$, krzywą zapadalności $i'(t) = \frac{di}{dt}$ oraz krzywą wykładniczą $i_0 e^{\beta t}$ dla $i_0 = 0,01$ oraz $\beta = 0,3$. Krzywe te pokazują, jaka część populacji jest zarażona $i(t)$, kiedy ma miejsce największy przyrost liczby zakażeń $i'(t)$ oraz w jaki sposób początkowo wzrasta rozmiar zarażonej części populacji $i(t) \approx i_0 e^{\beta t}$. Pokazano także, jaka część osobników pozostaje zdrowa $s(t) = 1 - i(t)$. W modelu SI liczba osobników zdrowych maleje w czasie, a zarażonych rośnie, aż do momentu zakażenia wszystkich. Największy wzrost liczby zarażonych osobników ma miejsce, gdy zapadalność $i'(t)$ jest maksymalna.

Model SIS

Naturalnym rozszerzeniem modelu SI jest założenie, że osobniki chore mogą wyzdrowieć. Szybkość zdrowienia opisuje parametr ν , który jest średnią liczbą ozdowień na jednostkę czasu. Na jego podstawie można wyznaczyć średni czas potrzebny do wyzdrowienia τ ($\tau = \frac{1}{\nu}$), który jest średnim czasem, przez jaki osobnik chory może zarażać osobniki zdrowe.



Rys. 3. Diagram dla modelu SIS: możliwe stany wraz z szybkościami przejścia między nimi

W modelu SIS proces zarażania wygląda tak samo, jak dla modelu SI: w jednostce czasu zakażeniu ulega $\beta \times I \times s$ osobników. Liczba osobników zarażonych zmniejsza się na skutek zdrowienia. W jednostce czasu (spontanicznie) zdrowieje $\nu \times I$ osobników. Dlatego zmiana liczby osobników zarażonych w czasie opisana jest równaniem

$$\frac{dI}{dt} = \beta I s - \nu I. \quad (14)$$

Dla osobników podatnych S mamy

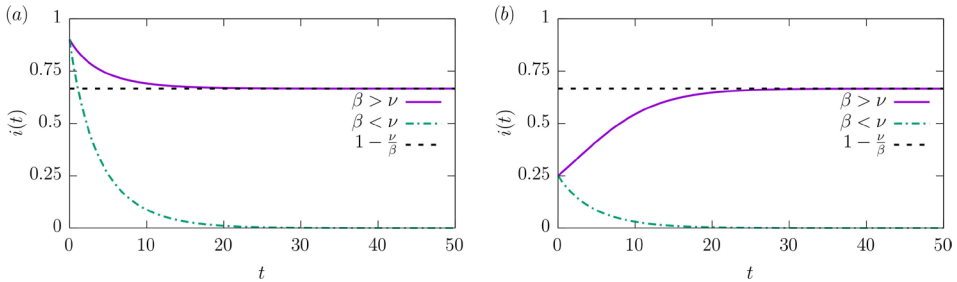
$$\frac{dS}{dt} = -\beta I s + \nu I. \quad (15)$$

Równanie (15) można także uzyskać, różniczkując po czasie związek $S = N - I$, z którego wynika

$$\frac{dS}{dt} = -\frac{dI}{dt}. \quad (16)$$

Łącznie równania opisujące zmianę liczby osobników w poszczególnych stanach mają postać

$$\begin{cases} \frac{dS}{dt} = -\beta Is + \nu I \\ \frac{dI}{dt} = \beta Is - \nu I \end{cases} \quad (17)$$



Rys. 4. Zależność zarażonej części populacji $i(t)$ od czasu t dla $\beta = 0,1$, $\nu = 0,3$ (linia ciągła) oraz $\beta = 0,3$, $\nu = 0,1$ (linia przerywano-kropkowana) z (a): $i(0) = 0,9$ oraz (b): $i(0) = 0,25$. Linie ciągłe odpowiadają sytuacji endemicznej ($\beta > \nu$), w której stała część populacji pozostaje zarażona.

Zdrowa część populacji jest równa $s(t) = 1 - i(t)$

Dzieląc stronami równania (14) oraz (15) przez całkowity rozmiar populacji, otrzymujemy układ równań na frakcje osobników w każdej z klas

$$\begin{cases} \frac{ds}{dt} = -\beta is + \nu i \\ \frac{di}{dt} = \beta is - \nu i \end{cases} \quad (18)$$

Ze względu na warunek $s(t) + i(t) = 1$ jedno z równań powyższego układu jest nadmiarowe. Z równania (18) oraz związku $s + i = 1$ dostajemy

$$\frac{ds}{dt} = -\beta(1-s)s + \nu(1-s) \quad (19)$$

lub

$$\frac{di}{dt} = \beta i(1-i) - \nu i, \quad (20)$$

które dla $\beta \neq \nu$ można przepisać w następującej formie

$$\frac{di}{dt} = i[(\beta - \nu) - \beta i] = (\beta - \nu)i \left[1 - \frac{\beta}{\beta - \nu} i \right]. \quad (21)$$

Równanie (21) jest bardzo podobne do równania (7), a jego rozwiązaniem jest

$$i(t) = \frac{i_0(\beta - \nu)e^{(\beta - \nu)t}}{(\beta - \nu) + i_0\beta(e^{(\beta - \nu)t} - 1)} = \frac{i_0(\beta - \nu)}{i_0\beta + (\beta - \nu - i_0\beta)e^{-(\beta - \nu)t}}. \quad (22)$$

Przypadek $\nu = \beta$ należy rozpatrzeć osobno. W takiej sytuacji

$$i(t) = \frac{i_0}{1 + i_0 \beta t}. \quad (23)$$

Rozmiar zdrowej części populacji jest równy

$$s(t) = 1 - i(t). \quad (24)$$

W granicy $t \rightarrow \infty$ zarażona jest następująca część populacji

$$\lim_{t \rightarrow \infty} i(t) = \begin{cases} 1 - \frac{\nu}{\beta} & \text{dla } \beta > \nu \\ 0 & \text{dla } \beta \leq \nu \end{cases}. \quad (25)$$

Dlatego jeśli $\beta < \nu$, choroba zostaje wyeliminowana, w przeciwnym razie (asymptotycznie) stała część populacji jest zarażona. Z faktu, że zarażona jest stała część populacji nie wynika, iż są to cały czas te same osobniki. Ze względu na ciągle zachodzące procesy zdrowienia i zarażenia osobniki zarażone i zdrowe zmieniają się, natomiast ich frakcja jest stała. Zachowanie asymptotyczne modelu SIS można przewidzieć dzięki analizie jakościowej. Dla $\beta \neq \nu$, równanie (20) ma dwa punkty stałe: $i^* = 0$ oraz $i^* = 1 - \frac{\nu}{\beta}$. Stabilnym punktem stałym jest $\max\left(0, 1 - \frac{\nu}{\beta}\right)$. Dlatego jeśli $1 - \frac{\nu}{\beta} < 0$, choroba zostaje wyeliminowana, a punkt stały $i^* = 1 - \frac{\nu}{\beta}$ staje się нефизyczny.

Rysunek 4 przedstawia zależność od czasu zarażonej części populacji $i(t)$ dla modelu SIS z $\beta = 0,1$, $\nu = 0,3$ (linia ciągła) oraz zamienionych wartości parametrów β i ν ($\beta = 0,3$, $\nu = 0,1$) z różnymi warunkami początkowymi: (a): $i(0) = 0,9$ oraz $i(0) = 0,25$. Dla $\beta > \nu$ (linie ciągłe) obserwujemy stan endemiczny – choroba nie zostaje wyeliminowana – $i(t) \rightarrow 1 - \nu/\beta$, bez względu na warunek początkowy. Natomiast dla $\beta < \nu$ choroba zostaje asymptotycznie w pełni wyeliminowana (linie przerywane). W celu podniesienia czytelności rysunku nie przedstawiono rozmiaru zdrowej części populacji, który jest równy $s(t) = 1 - i(t)$.

Model SIR

Model SIR jest modelem, w którym możliwe jest wyzdrowienie i uzyskanie odporności. Oprócz poznanych dotychczas klas osobników podatnych S i zarażonych I, istnieje klasa osobników ozdrowiałych R. Osobnicy ozdrowiali są osobnikami, które nie mogą zostać ponownie zarażeni – uzyskali oni trwałą odporność. Dlatego, w ogólności, mogą być także osobnikami, którzy zmarli lub zostali trwale usunięci bądź wyizolowani z populacji. W modelu SIR liczba osobników podatnych maleje na skutek zakażeń. Natomiast liczba osobników zarażonych rośnie ze względu na nowe zakażenia, a maleje ze względu na proces spontanicznych ozdowień. Liczba osobników ozdrowiałych tylko rośnie. Model SIR jest uważany za podstawowy model epidemiologiczny [4, 7].



Rys. 5. Diagram dla modelu SIR: możliwe stany wraz z szybkościami przejścia między nimi

Dla liczby osobników podatnych dostajemy

$$\frac{dS}{dt} = -\beta I s, \quad (26)$$

ponieważ

$$S(t + \Delta) = S(t) - I_{new}, \quad (27)$$

gdzie $I_{new} = \Delta\beta I s$. Dokonując analogicznych przekształceń dla równania (27) jak w równaniach (3) – (6), otrzymujemy równanie (26). Liczba osobników zarażonych wzrasta ze względu na nowe zakażenia oraz maleje ze względu na zdrowienie

$$\frac{dI}{dt} = \beta I s - \nu I, \quad (28)$$

ponieważ $\nu \times I$ jest liczbą ozdowień na jednostkę czasu. Liczba osobników ozdowiających rośnie ze względu na proces zdrowienia

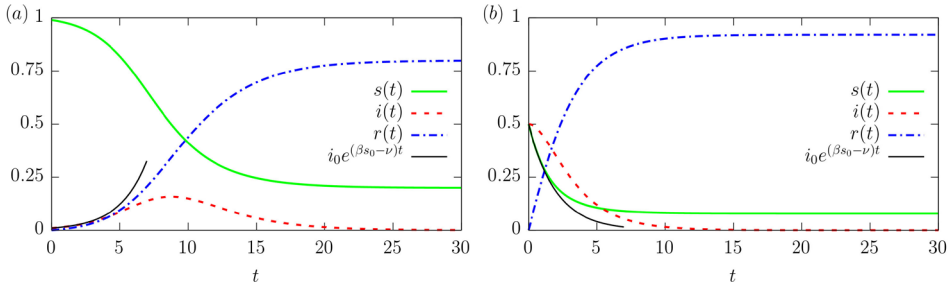
$$\frac{dR}{dt} = \nu I. \quad (29)$$

Dla modelu SIR równania opisujące zmianę liczby osobników w poszczególnych stanach mają następującą finalną postać

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta}{N} IS \\ \frac{dI}{dt} = \frac{\beta}{N} IS - \nu I. \\ \frac{dR}{dt} = \nu I \end{cases} \quad (30)$$

Układ równań (30) można rozwiązać tylko numerycznie. Dzieląc stronami równania (30) przez całkowity rozmiar populacji, otrzymujemy układ równań na frakcje osobników w każdej z klas

$$\begin{cases} \frac{ds}{dt} = -\beta i s \\ \frac{di}{dt} = \beta i s - \nu i. \\ \frac{dr}{dt} = \nu i \end{cases} \quad (31)$$



Rys. 6. Zależność zdrowej $s(t)$, zarażonej $i(t)$ oraz ozdrowiałej $r(t)$ części populacji dla $\beta = 1$, $\nu = 0,5$ z (a): $i(0) = 0,01$ oraz (b): $i(0) = 0,5$. Dodatkowa linia pokazuje krzywą wykładniczą opisującą zmianę zarażonej części populacji w początkowym okresie epidemii

Model SIR może zostać użyty do zdefiniowania bardzo ważnej wielkości: bazowego współczynnika reprodukcji \mathcal{R}_0 . W tym celu równanie (28) zapisujemy w postaci

$$\frac{dI}{dt} = \nu I \left[\frac{\beta}{\nu} s - 1 \right]. \tag{32}$$

O tym, czy liczba zarażonych rośnie, czy spada decyduje znak wyrażenia w nawiasie kwadratowym. Wprowadźmy oznaczenie $\mathcal{R}_0 = \frac{\beta}{\nu} s(0) \approx \frac{\beta}{\nu}$, ponieważ $s(0) \approx 1$, gdzie \mathcal{R}_0 jest bazowym współczynnikiem reprodukcji. Jest to liczba wtórnych zarażeń przypadających na jedno zakażenie pierwotne w populacji, w której wszystkie osobniki są podatne. We wczesnym stadium epidemii

$$\frac{dI}{dt} \approx \nu I \left[\frac{\beta}{\nu} s(0) - 1 \right] = \nu I \left[\frac{\beta}{\nu} - 1 \right] = \nu I [\mathcal{R}_0 - 1]. \tag{33}$$

Z powyższego równania można znaleźć zależność liczby zarażonych od czasu w początkowym ($s(0) \approx 1$) okresie epidemii ($t \approx 0$)

$$i(t) = i_0 e^{\nu(\mathcal{R}_0 - 1)t} = i_0 e^{(\beta - \nu)t} \tag{34}$$

oraz

$$I(t) = N_0 i(t) = N_0 e^{\nu(\mathcal{R}_0 - 1)t} = N_0 e^{(\beta - \nu)t}. \tag{35}$$

Dla $t \gg 0$ takie przybliżenie przestaje działać, ponieważ $s(t) \ll 1$.

Rysunek 6 pokazuje numerycznie znaną zależność zdrowej $s(t)$, zarażonej $i(t)$ oraz ozdrowiałej $r(t)$ części populacji wraz z początkową (wykładniczą) zależnością $i(t)$ dla (a): $i(0) = 0,01$ oraz (b): $i(0) = 0,5$ z $\beta = 1$ oraz $\nu = 0,5$. Rysunek 6(b) odpowiada takiemu warunkowi początkowemu, że $i'(t)|_{t=0} < 0$, dlatego $i(t)$ jest funkcją malejącą – liczba zarażonych tylko maleje. Takie samo zachowanie obserwujemy także dla niewielkiej liczby początkowo zarażonych osobników dla odpowiednio dobranych β i ν , dla których zachodzi $\beta/\nu < 1$. W modelu SIR z $\beta/\nu > 1$, rys. 6(a), krzywa $i(t)$ ma typowy dzwonowy kształt. Epidemia kończy się, gdy wszyscy zarażeni wyzdrowieją, dlatego w stanie końcowym są tylko zdrowe i ozdrowiałe osobniki.

Model SIR można zmodyfikować dodając możliwość powrotu z klasy R do klasy S. Takie rozszerzenie odzwierciedla sytuację, w której nie wszyscy osobnicy uzyskują odporność lub uzyskana odporność nie jest trwała. Dzięki możliwości takiego przejścia zwiększa się klasa możliwych zachowań. W szczególności, podobnie jak w modelu SIS, choroba może być endemiczna, tzn. stała część populacji jest zarażona.

Bazowy \mathcal{R}_0 i efektywny \mathcal{R}_t współczynnik reprodukcji

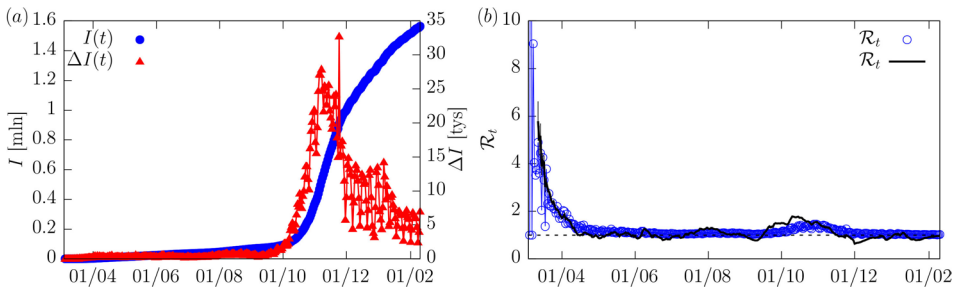
Jeśli β jest średnią liczbą kontaktów na jednostkę czasu, na jednego osobnika, które mogą prowadzić do zarażenia, a τ jest średnim czasem, przez jaki osobnik zarażony może zakażać ($\tau = \frac{1}{\nu}$), to liczba zarażeń przypadających na jednego zarażonego osobnika w pełni podatnej populacji jest równa bazowemu współczynnikowi reprodukcji

$$\mathcal{R}_0 = \beta\tau = \frac{\beta}{\nu}. \quad (36)$$

β zależy od prawdopodobieństwa zarażenia podczas pojedynczego kontaktu \tilde{p} oraz średniej całkowitej liczby kontaktów na jednostkę czasu \bar{c} . W najprostszym przypadku $\beta = \bar{c} \times \tilde{p}$. Wtedy bazowy współczynnik reprodukcji jest równy

$$\mathcal{R}_0 = \bar{c} \times \tilde{p} \times \tau. \quad (37)$$

Powyższy wzór sugeruje, że bazowy współczynnik reprodukcji może zostać ograniczony poprzez: zmniejszanie liczby kontaktów (izolacja, kwarantanna), obniżenie prawdopodobieństwa zarażenia (środki ochrony osobistej, dystansowanie społeczne, szczepienia) lub skracanie czasu zarażania (izolacja, działania terapeutyczne). Większość tych działań może zostać podsumowana przez #DDM: Dezynfekcja, Dystans, Maseczki.



Rys. 7. Zależności dla Polski: (a) liczba zarażonych $I(t)$ w milionach (lewa oś) oraz dzienna liczba wykrytych przypadków $\Delta I(t)$ w tysiącach (prawa oś) oraz (b) efektywny współczynnik reprodukcji \mathcal{R}_t obliczony z równania (41) (punkty) lub przy pomocy EpiEstim (linia ciągła)

Jeśli populacja jest tylko w części podatna, to efektywny współczynnik reprodukcji jest równy

$$\mathcal{R}_t = \mathcal{R}_0 s(t) = \mathcal{R}_0 \frac{S(t)}{N} = \frac{\beta}{\nu} \frac{S(t)}{N}, \quad (38)$$

ponieważ $\mathcal{R}_0 s(t)$ jest liczbą kontaktów z osobnikami podatnymi, które prowadzą do nowych zakażeń. Początkowo z dobrym przybliżeniem cała populacja składa się z osobników podatnych $s(0) \approx 1$ ($S(0) \approx N$). Jeśli β i ν są stałe, to efektywny współczynnik reprodukcji jest nierosnącą funkcją czasu. Zachodzi także $\mathcal{R}_t \leq \mathcal{R}_0$, dlatego \mathcal{R}_0 jest najbardziej pesymistycznym oszacowaniem \mathcal{R}_t .

W modelu SIR zmiana liczby osobników w czasie opisana jest układem równań (30). Równanie opisujące zmianę w czasie liczby zarażonych osobników może zostać przepisane jako

$$\frac{dI}{dt} = \nu I \left[\frac{\beta}{\nu} \frac{S}{N} - 1 \right] = \nu I [\mathcal{R}_t - 1], \quad (39)$$

gdzie $\mathcal{R}_t = \frac{\beta}{\nu} \frac{S(t)}{N}$ jest efektywnym współczynnikiem reprodukcji. W chwili $t = 0$: $S(0) \approx N$ i wtedy efektywny współczynnik reprodukcji jest równy bazowemu współczynnikowi reprodukcji \mathcal{R}_0 . Pochodną $I'(t)$ można przybliżyć przez

$$\frac{dI}{dt} \approx \frac{I(t + \Delta) - I(t)}{\Delta} = I_{t+1} - I_t, \quad (40)$$

ponieważ $\Delta = 1$ dzień jest częstością ogłaszania liczby zarażonych. Z równania (39) można obliczyć \mathcal{R}_t [2, 11]

$$\mathcal{R}_t = \frac{1}{\nu} \frac{I_{t+1} - I_t}{I_t} + 1 = \tau \frac{I_{t+1} - I_t}{I_t} + 1, \quad (41)$$

gdzie $\tau = \frac{1}{\nu}$ jest średnim czasem zarażania. Bardziej precyzyjne wyznaczenie efektywnego współczynnika reprodukcji oraz oszacowanie błędów jego wyznaczenia można przeprowadzić w środowisku R [8] przy pomocy pakietu EpiEstim [1, 2, 11].

Rysunek 7 przedstawia dane dla Polski [5] (IV 2019 – II 2020): (a) – ogólna liczba zarażeń $I(t)$ w milionach (lewa oś) oraz dzienna liczba zarażeń $\Delta I(t)$ w tysiącach (prawa oś), (b) – efektywny współczynnik reprodukcji \mathcal{R}_t obliczony z równania (41) (punkty), przy założeniu $\tau = 6,7$ oraz pakietu EpiEstim z $\tau = 6,7 \pm 5,2$.

Odporność stadna

Osobnik zarażony może zakażać tylko nieuodpornionych. W pełni podatnej populacji ($s \approx 1$) każdy osobnik zarażony zakaża średnio \mathcal{R}_0 osobników podatnych. W populacji, w której są osobniki uodpornione, osobnik zarażony zakaża średnio $\mathcal{R}_0 \times s$ osobników. Dlatego mamy

$$\mathcal{R}^* \leq \mathcal{R}_0 s = \mathcal{R}_0 (1 - p), \quad (42)$$

\mathcal{R}^* jest współczynnikiem reprodukcji dla częściowo uodpornionej populacji, a p zaszczepioną częścią populacji. Epidemia nie wybucha (liczba zarażonych maleje), jeśli $\mathcal{R}^* < 1$ czyli

$$\mathcal{R}_0 (1 - p) \leq 1. \quad (43)$$

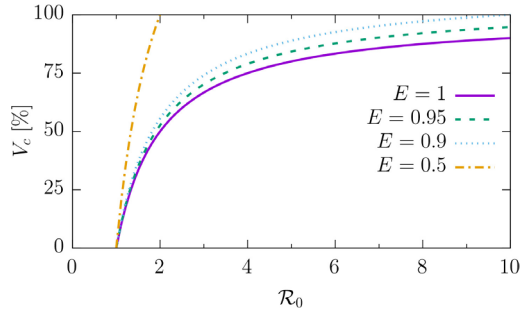
Stąd

$$p \geq 1 - \frac{1}{\mathcal{R}_0}. \quad (44)$$

Krytyczny poziom zaszczepień (uodpornień) [6] w idealnie przemieszanej populacji jest równy

$$p_c = 1 - \frac{1}{\mathcal{R}_0}. \quad (45)$$

Dla wzrastających \mathcal{R}_0 trzeba zaszczepić większą część populacji. Dla Covid-19 $2 \leq \mathcal{R}_0 \leq 6$, co oznacza, że w celu uzyskania odporności stadnej 50% – 83,3% populacji powinno być zaszczepione/odporne.



Rys. 8. Zależność krytycznego poziomu zaszczepień od bazowego współczynnika reprodukcji \mathcal{R}_0 dla różnych efektywności szczepienia E

Dla idealnej szczepionki [3] krytyczny poziom zaszczepień wynosi

$$V_c = p_c \quad (46)$$

Jeśli szczepionka ma skuteczność (efektywność E), to aby efektywnie uzyskać krytyczny poziom zaszczepień, należy wykonać większą liczbę szczepień

$$V_c \times E = p_c \quad (47)$$

Stąd krytyczny poziom zaszczepień jest równy

$$V_c = \frac{p_c}{E} = \frac{1 - \frac{1}{\mathcal{R}_0}}{E}. \quad (48)$$

Jeśli $E < 1 - \frac{1}{\mathcal{R}_0}$, nie da się uzyskać odporności stadnej, nawet szczepiąc całą populację. Rysunek 8 przedstawia zależność krytycznego poziomu zaszczepień w funkcji bazowego współczynnika reprodukcji dla kilku wybranych poziomów skuteczności szczepionki. Zakładając $2 \leq \mathcal{R}_0 \leq 6$, w porównaniu do idealnej szczepionki większa część populacji powinna być zaszczepiona/odporna dla $E = 0,9$: 55,5% – 92,6%, a dla $E = 0,95$: 52,6% – 87,7%.

Testy diagnostyczne i ich interpretacja

Stan epidemiczny danego osobnika ustala się na podstawie odpowiednich testów diagnostycznych. Każdy test scharakteryzowany jest przez dwa parametry: czułość (zdolność testu do wykrycia badanej choroby) oraz swoistość (zdolność wykrycia braku choroby). Czułość testu jest równa stosunkowi wyników prawdziwie dodatnich do sumy wyników prawdziwie dodatnich i fałszywie ujemnych. Natomiast swoistość testu to stosunek wyników prawdziwie ujemnych do sumy wyników prawdziwie ujemnych i fałszywie dodatnich. Czułość testu mówi, jaka część testów wykonanych na osobnikach zarażonych dała wyniki pozytywne, natomiast swoistość pokazuje, jaka część osobników zdrowych uzyskała ujemny wynik testu. Najlepsze testy to takie, których czułość i swoistość są bliskie jedności (w idealnym przypadku równe 1). Przykładowo dla czułości równej 0,99 (średnio) jeden na 100 przebadanych zarażonych osobników nie uzyskuje dodatniego wyniku testu. W przypadku testów wykonanych na 10000 zarażonych daje to (średnio) 100 osób.

Znając czułość ($P(+|I)$) oraz swoistość ($P(-|S)$) testu oraz wiedząc, jaka część populacji jest zarażona, korzystając ze wzoru Bayesa i wzoru na prawdopodobieństwo całkowite, można znaleźć prawdopodobieństwo, że osobnik który uzyskał dodatni wynik testu, jest istotnie chory

$$\begin{aligned} P(I|+) &= \frac{P(I \cap +)}{P(+)} = \frac{P(+|I)P(I)}{P(+|I)P(I) + P(+|S)P(S)} \\ &= \frac{P(+|I)P(I)}{P(+|I)P(I) + [1 - P(-|S)]P(S)}, \end{aligned} \quad (49)$$

oraz prawdopodobieństwo, że osobnik który uzyskał ujemny wynik testu, jest istotnie zdrowy

$$\begin{aligned} P(S|-) &= \frac{P(S \cap -)}{P(-)} = \frac{P(-|S)P(S)}{P(-|S)P(S) + P(-|I)P(I)} \\ &= \frac{P(-|S)P(S)}{P(-|S)P(S) + [1 - P(+|I)]P(I)}. \end{aligned} \quad (50)$$

Jeśli swoistość testu jest równa 1, to $P(I|+)=1$, ponieważ nie obserwujemy przypadków fałszywie pozytywnych (skutecznie wykluczamy chorobę). W przypadku, gdy czułość testu jest równa 1, to $P(S|-)=1$, ponieważ nie obserwujemy przypadków fałszywie ujemnych (skutecznie potwierdzamy chorobę). W innych przypadkach, aby wyznaczyć powyższe prawdopodobieństwa, konieczna jest wiedza, jaka część populacji choruje na daną chorobę. W idealnej sytuacji możliwe jest uzyskanie testu o czułości 1 i dużej swoistości. W rzeczywistych przypadkach czułość i swoistość testów diagnostycznych są mniejsze od 1 i, co więcej, mogą zależeć od sposobu wykonywania testu i pobierania próbek.

Z punktu widzenia interpretacji wyników testów kluczowe jest zrozumienie, czym jest czułości i swoistości testów. Jeśli swoistość testu wynosi 1, to nie dochodzi do sytuacji, gdy osoby zdrowe uzyskują dodatni wynik testu (*false positive*). W takiej sytuacji dodatni wynik testu oznacza bycie zarażonym (z prawdopodobieństwem 1). Analogicznie jeśli czułość testu jest równa 1, to nie mamy wyników fałszywie ujemnych (*false negative*). Dlatego ujemny wynik testu oznacza bycie zdrowym (z prawdopodobieństwem 1). Dla nieidealnych testów nie wiadomo, jakie jest prawdopodobieństwo bycia zarażonym w przypadku pozytywnego wyniku testu lub bycia zdrowym w przypadku ujemnego wyniku testu. W celu wyznaczenia tych prawdopodobieństw należy zastosować wzory (49) oraz (50). Prawdopodobieństwa te mogą znacznie różnić się od czułości i swoistości testu. W zależności od sytuacji mogą one być mniejsze lub większe od nich!

Czułość równa 1 oznacza, że zawsze wykrywamy zarażenie, jeśli dany osobnik jest zakażony. Swoistość równa 1 oznacza, że nie ma przyczyn innych, niż diagnozowana choroba, mogących prowadzić do pozytywnego wyniku testu. Dlatego łatwiejsze jest uzyskanie maksymalnej możliwej czułości niż swoistości. Jakość testów wyznacza się na podstawie badania testów. Testy też się testuje!

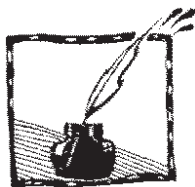
Podsumowanie

Modele epidemiologiczne pozwalają na lepsze zrozumienie mechanizmów rządzących rozprzestrzenianiem się epidemii. Przedstawione modele SI, SIS oraz SIR są podstawowymi modelami epidemiologicznymi, które stanowią punkt wyjścia dla konstrukcji bardziej złożonych modeli epidemiologicznych. Poza modelami epidemiologicznymi przedstawiliśmy matematyczne aspekty możliwości uzyskania odporności stadnej, które powiązane są z bazowym współczynnikiem reprodukcji oraz skutecznością szczepień. Bazowy współczynnik reprodukcji jest podstawową wielkością używaną do określenia poziomu zaszczepień gwarantującego uzyskanie odporności stadnej. Jego minimalizacja jest podstawowym celem strategii powstrzymywania epidemii. Stan epidemiologiczny osobników ocenia się na podstawie odpowiednich testów, dlatego przedstawiliśmy także kilka uwag dotyczących interpretacji wyników testów diagnostycznych, które odwołują się do czułości i swoistości.

Wybrana literatura

- [1] A. Cori. Package 'EpiEstim'. Imperial College, London, UK, 2019.
- [2] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez. *A new framework and software to estimate time-varying reproduction numbers during epidemics*. Am. J. Epidemiol., 178(9):1505-1512, 09 2013.
- [3] P. Fine, K. Eames, and D. L. Heymann. "*Herd immunity*": A rough guide. Clin. Infect. Dis., 52(7):911-916, 2011.
- [4] H. W. Hethcote. *The mathematics of infectious diseases*. SIAM Rev., 42(4):599-653, 2000.
- [5] <https://github.com/CSSEGISandData/COVID-19>, (2020).
- [6] S. A. Levin, T. G. Hallam, and L. J. Gross. *Applied mathematical ecology*. Springer Verlag, Berlin, 2012.
- [7] J. D. Murray. *Mathematical biology I: An introduction*. Springer, New York, 2002.

-
- [8] R Core Team. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
 - [9] H. G. Schuster and W. Just. *Deterministic chaos: An introduction*. Wiley-VCH, Weinheim, 2006.
 - [10] S. H. Strogatz. *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Westvie Press, 2015.
 - [11] J. Wallinga and P. Teunis. *Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures*. Am. J. Epidemiol., 160(6):509-516, 2004.



List do redakcji Fotonu

Szanowna Redakcjo,

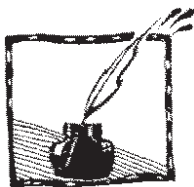
ze zdziwieniem odnotowaliśmy pojawienie się w najnowszym numerze Fotonu (Lato 2020), czasopisma sponsorowanego przez Wydział Fizyki, Astronomii i Informatyki UJ i przeznaczonego m. in. dla nauczycieli fizyki, artykułu *Uniwersalny układ odniesienia dla propagacji światła – nierozwiązany problem fizyki* napisanego przez Karola i Romana Szostków. Autorzy przywracają dawno porzuconą Teorię Eteru i kwestionują Szczególną Teorię Względności Einsteina, sformułowaną w ostatecznej postaci przez Hermanna Minkowskiego w 1909 roku. Niżej podpisani – Edward Malec i Andrzej Staruszkiewicz – pełnili w swoim czasie funkcje Kierownika Zakładu Teorii Względności i Astrofizyki UJ, czują się zatem zobligowani do skomentowania tego niewątpliwie skandalicznego (ich zdaniem) faktu.

Musimy przypomnieć, że fizyka teoretyczna polega na tworzeniu *modeli matematycznych* wybranych aspektów rzeczywistości fizycznej. Wielkie teorie fizyki – Mechanika Newtona, Elektrodynamika Maxwella, Ogólna Teoria Względności Einsteina, są właśnie takimi modelami. O ich wartości decyduje zakres zjawisk, które pozwalają opisać oraz ilościowa doskonałość opisu, którą pozwalają osiągnąć. Weźmy np. model zwany Elektrodynamiką Kwantową. Model ten opiera się na dwu filarach: Elektrodynamice Maxwella i teorii elektronu Diraca. Oba te filary z kolei opierają się na założeniu, że czasoprzestrzeń ma geometrię Minkowskiego, tak pochopnie odrzucaną przez Panów Szostków. W ramach Elektrodynamiki Kwantowej można obliczyć tzw. anomalny moment magnetyczny elektronu, a obserwacje potwierdzają 11 miejsc znaczących wyniku teoretycznego.

Zachęcamy Czytelników żeby rozważyli jaka jest szansa, że tych 11 miejsc znaczących zgadza się z obserwacjami przez jakiś złośliwy przypadek i to pomimo błędów w założeniach, jakim niewątpliwie byłoby przyjęcie fałszywej geometrii czasoprzestrzeni jako podstawy całego rachunku.

Jako Appendix do niniejszego listu umieszczamy link <https://radynaukowe.uw.edu.pl/wp-content/uploads/sites/17/2021/01/RNDNF-NR-13.pdf> pod którym można znaleźć – na stronach 3-12 streszczenia recenzji prac Pana Romana Szostka przedstawionych w celu uzyskania stopnia naukowego doktora habilitowanego. Jest wśród nich omówienie recenzji, którą jeden z nas (A.S.) przedstawił Radzie Wydziału Fizyki Uniwersytetu Warszawskiego.

Edward Malec i Andrzej Staruszkiewicz



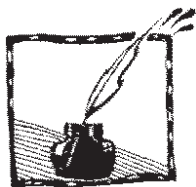
Nota od redakcji

Redakcja *Fotonu* dziękuje Panu prof. dr. hab. Edwardowi Malcowi oraz Panu prof. dr. hab. Andrzejowi Staruszkiewiczowi za list w sprawie publikacji w *Fotonie* 149 (Lato 2020) artykułu *Uniwersalny układ odniesienia dla propagacji światła – nierozwiązany problem fizyki*, autorstwa Panów Karola Szostka i Romana Szostka. Redakcja ubolewa nad niedołożeniem wystarczającej staranności w doborze recenzentów tego artykułu i zbyt pospieszną jego publikację. Artykuł, który miał być zachętą do dyskusji, stał się kontrowersyjny i nieakceptowalny dla środowiska – nie tylko krakowskich – fizyków zajmujących się relatywistyką.

Czytelników *Fotonu* zachęcamy do zapoznania się z poglądowymi artykułami publikowanymi w przeszłości w *Fotonie* na temat Szczególnej Teorii Względności, zaproponowanej przez Alberta Einsteina i, jak dotąd, potwierdzonej przez wszystkie obserwacje i badania doświadczalne:

- *O matematyce, fizyce i paradoksie bliźniąt*, Leszek Sokołowski, *Foton* 61 (1999), s. 63-65,
- *O prędkościach nadświatlnych*, Leszek Sokołowski, *Foton* 94 (Jesień 2006), s. 17-25,
- *O paradoksie bliźniąt nieco inaczej. Paradoks i jego kontekst – cz. I*, Leszek Sokołowski, *Foton* 126, s. 23-30,
- *O paradoksie bliźniąt nieco inaczej – cz. II. Geometria paradoksu*, Leszek Sokołowski, *Foton* 127, s. 36-44.

w imieniu redakcji *Fotonu*
Dagmara Sokołowska



Interferencja światła – kiedy możemy ją zaobserwować? część I

*Tomasz Kawalec, Instytut Fizyki
im. Mariana Smoluchowskiego UJ*

O interferencji światła możemy przeczytać w wielu podręcznikach szkolnych, akademickich oraz artykułach naukowych. Interferencja jest tam omawiana zarówno w kontekście opisu klasycznego, jak i kwantowego. W *Fotonie* temu tematowi był na przykład poświęcony dwuczęściowy artykuł w numerach 94 i 95 w 2006 roku. Kiedy zagadnienie interferencji staje się nam już bliższe, dowiadujemy się, że interferować może nie tylko światło (czy też w ogóle – fale), ale też cząstki, atomy czy nawet cząsteczki! Mam tu na myśli zakres od elektronów aż po fulereny. Istnieją nawet doświadczenia, w których rola materii i światła została niejako odwrócona. Zamiast interferencji światła, wynikającej z jego propagacji przez szczeliny (czyli struktury materialne), bada się interferencję atomów, poruszających się przez siatkę dyfrakcyjną „zbudowaną” ze światła!¹ Czy w tej sytuacji mogę napisać o interferencji coś ciekawego? Czy mogę dołożyć chociaż małą cegiełkę do już istniejącej, stabilnej budowli? Od strony samej fizyki – zapewne nie. Ale myślę, że mogę uczniom, ale też i studentom, pomóc zrozumieć niektóre aspekty tych fascynujących zjawisk.

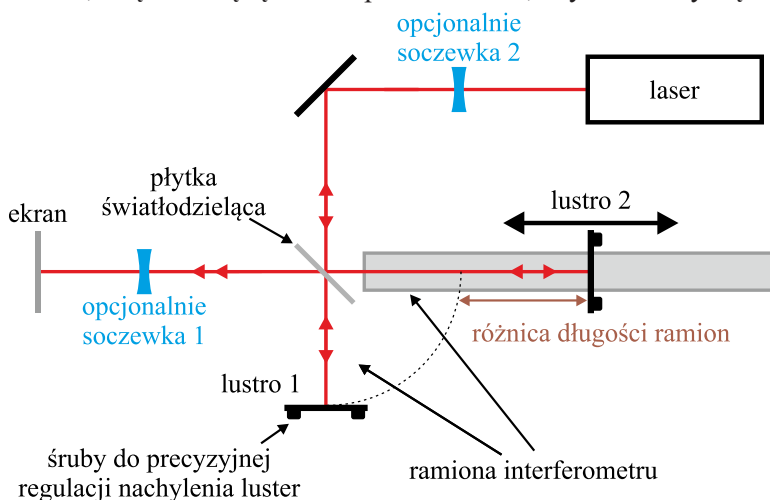
Kiedy jest mowa o interferencji światła, przychodzą nam na myśl głównie statyczne prążki interferencyjne – ciekawe układy obszarów jasnych i ciemnych, obserwowanych na ekranie czy na kamerze. Ale dlaczego czasem te prążki to zestaw równoległych linii, a czasem koncentryczne okręgi? Czy gdy na ekranie widzimy po prostu jasną, jednorodną plamę, to na pewno interferencja nie zachodzi? A może cała ta plama to jeden wielki, pojedynczy prążek? A może prążki istnieją, ale są zbyt gęste, żeby je zobaczyć gołym okiem? Albo z jakiegoś powodu przesuwają się tak szybko, że dla naszego oka struktura interferencyjna po prostu się uśrednia? Pójdźmy nawet dalej. Zgodnie z powszechnym przekonaniem, obserwacja interferencji jest możliwa tylko wtedy, gdy interferujące wiązki światła są spójne, czyli drgania jednej i drugiej fali są względem siebie „uporządkowane”. Dlatego spotykamy się z opinią, że interferujące wiązki powinny pochodzić z lasera – i to jednego lasera. A gdyby źródłami tych wiązek były dwa zupełnie niezależne lasery? Albo gdyby źródłem światła była lampa? Czy wtedy mamy szansę zaobserwować coś ciekawego? Z drugiej strony – czy użycie jednego lasera na pewno gwarantuje, że prążki się pojawiają? Na te pytania odpowiem w drugiej części artykułu. W pierwszej części zajmiemy się geometrią statycznych prążków – takich, które można łatwo obserwować gołym okiem lub

¹ Pojęcia „interferencja atomów”, ale też i „interferencja fotonów” są tak naprawdę nieprecyzyjne i mogą być mylące. Wyjaśnię to w drugiej części artykułu.

przy pomocy zwykłego aparatu lub kamery, bez stosowania szczególnych, dodatkowych zabiegów.

Przy omawianiu interferencji często posługujemy się pojęciem różnicy faz interferujących fal. Różnica faz mówi nam, jaka jest zależność pomiędzy „stanem” fal w danym punkcie – na przykład na ekranie. Dla światła mówi nam ona, czy jeśli dla jednej fali akurat wektor pola elektrycznego jest maksymalny, to dla drugiej też jest maksymalny, czy może maksymalny, ale o przeciwnym zwrocie, czy może jest w jakimś stanie pośrednim. Matematycznie różnica faz jest wyrażana w jednostkach takich jak kąty – najczęściej w radianach. Jasne prążki interferencyjne powstają w miejscach, w których różnica faz nakładających się fal jest parzystą wielokrotnością π (czyli w szczególności wynosi zero), a ciemne – gdy jest nieparzystą wielokrotnością π . W pierwszym przypadku fale dodają się tak, że się wzmacniają (wektory pola elektrycznego obu fal drgają zgodnie), a w drugim tak, że się wygaszają (wektory pola elektrycznego mają zawsze przeciwne zwroty).

Do obserwacji interferencji światła często wykorzystuje się dwuwiaźkowe interferometry Michelsona lub Macha-Zehndera. Każdy z nich można łatwo zbudować w laboratorium optycznym lub nawet kupić w postaci gotowego zestawu. Wiązki światła, które interferują, pochodzą z lasera (lub nawet wskaźnika laserowego), którego światło jest najpierw dzielone na płytce światłodzielącej. Następnie dzięki układowi optycznemu, złożonemu w najprostszym przypadku z samych lusterek, wiązki te są łączone z powrotem tak, aby nakładały się na siebie.



Rys. 1. Schemat interferometru Michelsona. Różnica dróg optycznych pomiędzy ramionami interferometru jest równa podwojonej różnicy ich długości. Czynnikiem 2 wynika z faktu, że światło w ramionach biegnie „tam i z powrotem”

Schemat interferometru Michelsona, wraz z uproszczonym biegiem wiązek, jest przedstawiony na rys. 1. Światło z lasera pada na płytkę światłodzielącą, która rozdziela je do dwóch ramion interferometru. Często używa się też terminu „płytką półprzepuszczalną”, jednak unikam tej nazwy, ponieważ rzadko kiedy ta

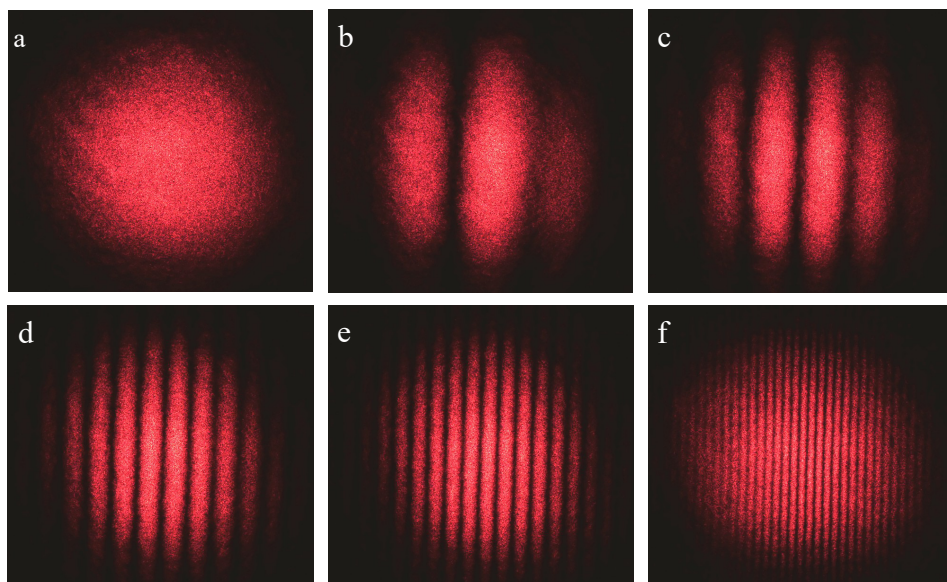
plytka faktycznie dzieli wiązkę światła „na pół” pod względem natężenia. Żeby nie odbiegać od głównego wątku, nie zajmujemy się tu celowo szczegółami toru wiązek wewnątrz płytki światłodzielącej. Światło odbite od lustra 1 i 2 trafia ponownie na płytkę światłodzielącą, dzięki czemu obie wiązki możemy na siebie nałożyć. To zadanie wymaga dużej dokładności – dlatego oba lusterka, a często też i płytka światłodzieląca, są zamontowane w uchwytych, których kąt nachylenia możemy precyzyjnie regulować specjalnymi śrubami. Ostatecznie część światła biegnie w stronę ekranu, a część – z powrotem w stronę lasera. O tym drugim przypadku często się zapomina – omówimy go później.

Na początek rozważmy przypadek, gdy wiązki światła padające na ekran biegną dokładnie po tym samym torze. Ponadto nie stosujemy ani soczewki 1, ani 2. Wtedy wzór na natężenie światła I na wyjściu interferometru Michelsona dla światła monochromatycznego przyjmuje postać:

$$I(\Delta x) = I_0 \left(1 + \cos \left(2\pi \frac{\Delta s}{\lambda} \right) \right), \quad (1)$$

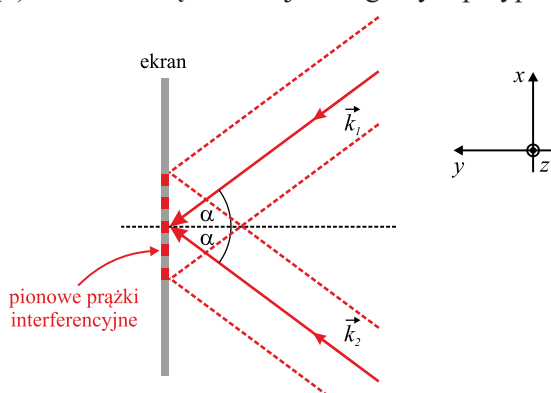
gdzie I_0 to natężenie światła laserowego na wejściu interferometru, Δs to różnica dróg optycznych w ramionach interferometru, a λ to długość fali światła². Wzór (1) tak naprawdę dotyczy sytuacji, gdy światło jest tak zwaną falą płaską, czyli nie skupia się ani nie rozbiega oraz wiązka nie jest przestrzennie ograniczona – czyli rozciąga się na boki w nieskończoność. Światło laserowe (ani żadne inne) w rzeczywistości falą płaską nie jest – nie istnieją w zwykłych ośrodkach naprawdę równoległe (skolimowane) wiązki światła. Jeśli jednak wiązka jest wystarczająco szeroka – powiedzmy na kilka milimetrów – opis poprzez fale płaskie jest zupełnie wystarczający. Co nam mówi wzór (1)? Po pierwsze, że natężenie światła na ekranie jest wszędzie takie samo – cały ekran jest oświetlony równomiernie. Nie ma w tym wzorze żadnych zmiennych związanych z położeniem na ekranie. Ponieważ mamy tu do czynienia tylko z imitacją fali płaskiej, to mowa jest oczywiście o oświetleniu ekranu na obszarze wyznaczonym przekrojem wiązek laserowych. Po drugie – natężenie światła zależy od różnicy dróg optycznych Δs w ramionach interferometru. Natężenie może przyjmować wartości od 0 do $2I_0$. Gdy delikatnie (w zakresie setek nanometrów) zmienimy położenie lustro 1 lub 2, cały oświetlony ekran będzie jaśniejszy lub ciemniejszy. W takim razie – gdzie są ci główni bohaterowie – prążki?! Pierwsza odpowiedź brzmi – to, co widać na ekranie, to właśnie po prostu taki jeden, duży prążek. Geometria (i w tym – symetria) układu sprawia, że oczekiwane przez nas ciekawsze prążki nie mają się tu jak pojawić. W doświadczeniach faktycznie da się to zaobserwować. Jeśli wiązki są bardzo precyzyjnie nałożone, to dostajemy obraz taki, jak na rys. 2 a. Innymi słowy – różnica faz między obydwoma wiązkami jest na ekranie w każdym punkcie taka sama. Jak zatem dostać „prawdziwe” prążki, takie jak te na rys. 2 b-f?

² Czytelnicy, którzy chcą znać wyprowadzenie tego wzoru, mogą skorzystać z rozumowania zaczynającego się od wzoru (2) poniżej i wziąć od razu $\alpha = 0$.



Rys. 2. Prążki interferencyjne liniowe o różnej gęstości. Kolejne zdjęcia zostały zrobione dla coraz większego kąta między wiązkami. Rozmiar plamy światła wynosił około 40 cm, dzięki zastosowaniu soczewki 2 (opis w tekście)

Musimy sprawić, żeby różnica faz między wiązkami była różna w różnych miejscach ekranu. Okazuje się, że wystarczy poprowadzić obie wiązki pod pewnym kątem względem siebie, tak jak to jest naszkicowane na rys. 3. Jak się zaraz okaże, do samych celów obserwacyjnych ten kąt musi być mały – poniżej 1 stopnia, ale w pewnych sytuacjach używa się i znacznie większych kątów. Jak znaleźć geometrię prążków w takiej sytuacji? Poniżej przedstawiony jest szkic jednego z rozwiązań, który bazuje na podejściu ogólniejszym niż to, które prowadziło do wzoru (1) i wzór ten będzie tutaj szczególnym przypadkiem.



Rys. 3. Geometria układu do otrzymywania liniowych prążków. Orientacja ekranu jest dokładnie taka sama, jak na rys. 1. Linie kreskowane umownie pokazują szerokość padających wiązek światła. Wektory falowe \vec{k}_1 i \vec{k}_2 wskazują kierunek rozchodzenia się fal. Kąt między wiązkami jest tu celowo duży, żeby rysunek był przejrzysty, chociaż bywają sytuacje, w których i takie się stosuje

Natężenie światła I jest proporcjonalne do średniej po czasie (zaznaczonej poziomą kreską) z kwadratu natężenia pola elektrycznego E :

$$I(\vec{r}) = \varepsilon_0 c \overline{E(\vec{r}, t)^2}. \quad (2)$$

Co prawda natężenie pola elektrycznego jest wektorem, ale tutaj wystarczy opis skalarny (liczbowy), o czym wspomnę później. W powyższym wzorze ε_0 to przenikalność elektryczna próżni, c to prędkość światła, a t to czas. Dlaczego pojawia się ta średnia po czasie? Chodzi o to, że pole elektryczne (i magnetyczne) światła drga tak szybko, że ani nasze oko, ani przyrządy pomiarowe nie są w stanie tych drgań bezpośrednio zarejestrować – są na to zbyt powolne. Ponadto, ponieważ zagadnienie robi nam się bardziej skomplikowane geometrycznie niż poprzednio, konieczne było wprowadzenie wektora położenia \vec{r} . Zwróćmy też uwagę na nazewnictwo, które – jak zaobserwowałem – bywa mylące nawet dla studentów. Mamy bowiem do czynienia z *natężeniem światła* i z *natężeniem pola elektrycznego*. To są różne wielkości fizyczne, z różnymi jednostkami.

Zgodnie z zaobserwowanym prawem przyrody (tak zwana zasada superpozycji), natężenia pola elektrycznego obydwu wiązek musimy do siebie dodać, a wynik dodawania wstawić do wzoru (3):

$$I(\vec{r}) = \varepsilon_0 c \left(\frac{1}{\sqrt{2}} E_0 \cos(\vec{k}_1 \cdot \vec{r} - \omega t) + \frac{1}{\sqrt{2}} E_0 \cos(\vec{k}_2 \cdot \vec{r} - \omega t + k_2 \Delta s) \right)^2. \quad (3)$$

W powyższym wzorze $\frac{1}{\sqrt{2}} E_0$ to amplituda pola elektrycznego światła w ramieniu interferometru – już po przejściu przez płytkę światłodziącą. Czynniki $\frac{1}{\sqrt{2}}$ może na pierwszy rzut oka wyglądać podejrzanie. Wynika on z tego, że płytka dzieli u nas natężenie światła na pół, a zależność między natężeniem światła i pola elektrycznego jest kwadratowa (wzór 2). ω to częstość naszych fal, a \vec{k} to tak zwany wektor falowy światła, który po prostu wskazuje kierunek rozchodzenia się fali. Kropki oznaczają iloczyn skalarny wektora falowego i wektora położenia. Przyjąłem, że dodatkową drogę optyczną Δs przebywa fala 2. Wiązki pochodzą z tego samego lasera, mają zatem tą samą częstotliwość i długość k wektora falowego:

$$|\vec{k}_1| = |\vec{k}_2| = k = \frac{2\pi}{\lambda}. \quad (4)$$

Na podstawie geometrii układu (zwróćmy uwagę na układ współrzędnych na rys. 3) wnioskujemy, że:

$$\vec{k}_1 \cdot \vec{r} = -k_x x + k_y y \quad \vec{k}_2 \cdot \vec{r} = k_x x + k_y y. \quad (5)$$

Po wstawieniu (5) do (3) przychodzi pora na wykonanie kilku przekształceń. Na początek skorzystajmy ze wzoru skróconego mnożenia:

$$I(\vec{r}) = \frac{1}{2} \varepsilon_0 c E_0^2 \times \left(\overline{\cos^2(\vec{k}_1 \cdot \vec{r} - \omega t)} + 2 \overline{\cos(\vec{k}_1 \cdot \vec{r} - \omega t) \cos(\vec{k}_2 \cdot \vec{r} - \omega t + k \Delta s)} + \overline{\cos^2(\vec{k}_2 \cdot \vec{r} - \omega t + k \Delta s)} \right). \quad (6)$$

Średnią sumy zamieniliśmy na sumę średnich, ponieważ wszędzie ta średnia dotyczy tego samego „odcinka” czasu. Teraz w członie mieszanym korzystamy z tożsamości trygonometrycznych i zapisujemy iloczyn cosinusów za pomocą sumy cosinusów z odpowiednimi argumentami. Ponadto wykonujemy niektóre średnie po czasie i otrzymujemy³:

$$I(\vec{r}) = \frac{1}{2} \varepsilon_0 c E_0^2 \left(\frac{1}{2} + 2 \frac{1}{2} \left(\overline{\cos(-2k_x x - k \Delta s)} + \overline{\cos(2k_y y - 2\omega t + k \Delta s)} \right) + \frac{1}{2} \right). \quad (7)$$

Pierwsza średnia po czasie nie wpływa na funkcję cosinus, ponieważ zależność od czasu tam zniknęła. Druga średnia po czasie powoduje, że jej człon się zeruje. Ponadto korzystamy z zależności $I_0 = \frac{1}{2} \varepsilon_0 c E_0^2$ i dostajemy:

$$I(x) = I_0 \left(1 + \cos \left(x \frac{4\pi}{\lambda} \sin \alpha + \frac{2\pi}{\lambda} \Delta s \right) \right). \quad (8)$$

Zgodnie z oczekiwaniami, ze względu na geometrię układu, prążki interferencyjne tworzą się w płaszczyźnie ekranu (nie ma zależności od zmiennej y) i są prostopadłe do płaszczyzny rysunku (nie ma zależności od zmiennej z). Z powyższego wzoru na natężeniowy profil prążków interferencyjnych wzdłuż osi x , możemy wyliczyć ich okres d :

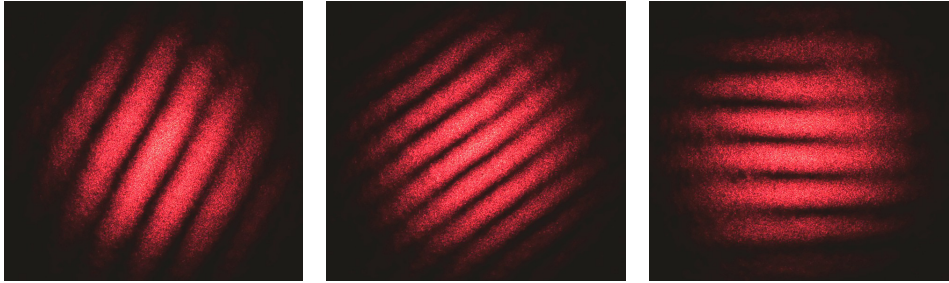
$$d = \frac{\lambda}{2 \sin \alpha}. \quad (9)$$

Okres ten zależy od długości fali oraz – co dla nas ważniejsze – od kąta α . Gdy kąt ten zmierza do zera, okres prążków zmierza do nieskończoności, a to jest zgodne z naszą wcześniejszą analizą. Gdy kąt rośnie – gęstość prążków również rośnie, tak jak to jest pokazane na rys. 2 b-f. Przy kącie 2α równym 30° , odległość między środkami prążków wynosi tylko około $1,2 \mu\text{m}$ dla popularnego lasera helowo-neonowego o długości fali 633 nm . Takie prążki nie są widoczne bezpośrednio gołym okiem. Dla celów demonstracyjnych możemy powiększyć obraz interferencyjny, wstawiając soczewkę rozpraszającą 2. Soczewka ta nie wpływa zasadniczo na samą strukturę ciemnych i jasnych pól, ale ułatwia obserwację. Manipulując płaszczyzną, w której jest zawarty kąt 2α między wiązkami, poprzez delikatną zmianę kąta ustawienia lusterek interferometru, możemy oczywiście obracać prążki interferencyjne – przykłady są pokazane na rys. 4.

Żeby nie poruszać kilku tematów jednocześnie, celowo pominąłem tutaj zagadnienie polaryzacji światła, czyli kierunku (lub kierunków) drgań wektora natężenia pola elektrycznego. Głębsza analiza oraz nieco uogólnione równanie (6) prowadzą do wniosku, że aby interferencja zaszła, polaryzacje wiązek z obydwu ramion nie mogą być względem siebie prostopadłe, a najlepiej – gdy są równoległe. W standardowych sytuacjach na szczęście ten warunek jest spełniony (lub prawie spełniony) bez naszych dodatkowych działań. Tutaj założyliśmy po cichu,

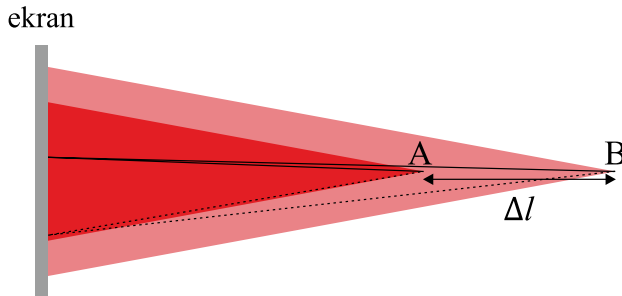
³ $\overline{\cos^2 \omega t} = \overline{\sin^2 \omega t} = \frac{1}{2}$. Wynik ten można dostać z całkowania, ale też z prostych rozważań geometrycznych – zachęcam do narysowania wykresu funkcji $\cos^2 x$ i skorzystania z jego symetrii w pionie. Natomiast $\overline{\cos \omega t} = \overline{\sin \omega t} = 0$. Wszystkie te przypadki dotyczą sytuacji, gdy średnią liczymy po okresie lub jego wielokrotności.

że obie wiązki są spolaryzowane prostopadle do płaszczyzny rysunku i dlatego pole elektryczne traktowaliśmy skalarnie, zamiast wektorowo we wzorze (3).



Rys. 4. Prążki interferencyjne dla różnych orientacji kąta między wiązkami

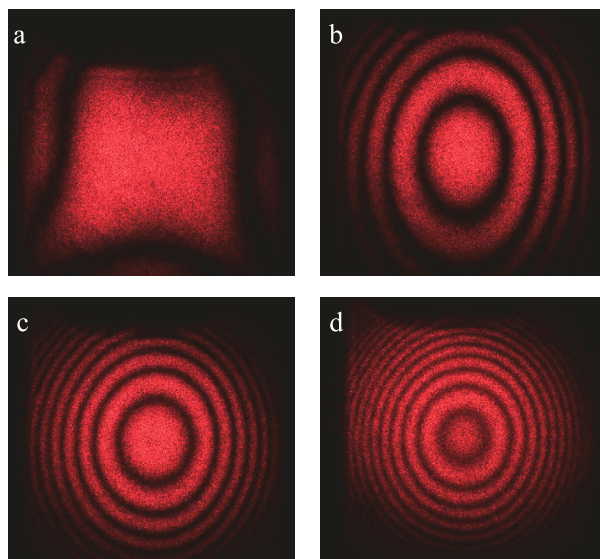
Pozostała nam jeszcze kwestia okrągłych prążków interferencyjnych – jak je uzyskać? Już wiemy, że geometria układu musi być inna niż poprzednio – tutaj bowiem potrzebujemy symetrii cylindrycznej. Najłatwiej ją uzyskać, wprowadzając soczewkę 1 na wejściu interferometru, zgodnie z rys. 1. Interferujące wiązki są wtedy rozbieżne, ale w przeciwieństwie do przypadku użycia tylko soczewki 2, teraz mamy możliwość kontrolowania różnicy położenia miejsc, z których obie wiązki się rozbiegają – spójrzmy na szkic na rys. 5. Dla uproszczenia, sam interferometr został tu pominięty, a pozostały jedynie interferujące wiązki.



Rys. 5. Szkic przedstawiający nałożenie rozbieżnych wiązek światła. Względna odległość Δl (pozornych) miejsc A i B rozbiegania się wiązek jest regulowana w rzeczywistości położeniem lustra w jednym z ramion interferometru. Czarne linie symbolizują, w ramach optyki geometrycznej, promienie światła wybiegające z punktów A i B. W zależności od miejsca padania na ekran, różnica faz między obydwojma promieniami jest różna – na ekranie powstają prążki interferencyjne. Ze względu na symetrię obrotową wzdłuż osi łączącej punkty A i B, prążki przyjmują kształt pierścieni

Gdyby względna odległość Δl była zbliżona do zera, w zasadzie powinniśmy zaobserwować na ekranie jeden duży, jednorodny prążek. Czynniki geometryczne i optyczne powodują jednak, że w naszym układzie ten prążek jest zniekształcony – popatrzmy na rys. 6 a. Natomiast w miarę zwiększania odległości Δl , po pierwsze pojawiają się koncentryczne prążki, a po drugie – prążki stopniowo się zagęszczają. Przykłady są pokazane na rys. 6 b-d, gdy różnica dróg optycznych była wydłużana o około 2 cm do każdego zdjęcia.

Zwróćmy tutaj jeszcze uwagę na to, że taka duża „zgrubna” zmiana różnicy dróg optycznych w zakresie centymetrów powoduje przede wszystkim wspomnianą zmianę generalnej struktury prążków (zagęszczanie lub rozrzedzanie). Natomiast to, czy akurat w centrum obrazu będzie jasny, czy ciemny prążek, czy też pośredni, zależy już niemal mikroskopowo od wartości różnicy dróg optycznych. Wystarczy delikatnie dotknąć któregoś z lusterek interferometru, a zmiana jego położenia o setki nanometrów spowoduje zamianę jasnego i ciemnego prążka (ale praktycznie nie zmieni gęstości prążków).

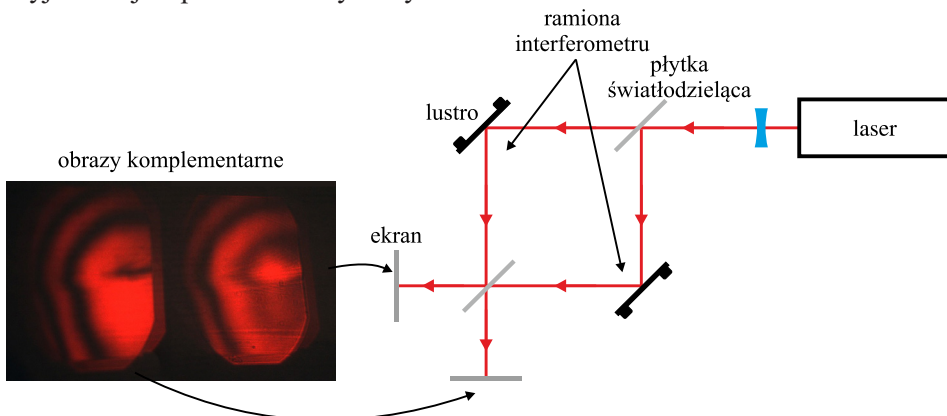


Rys. 6. Okrągłe prążki interferencyjne, uzyskane po zastosowaniu soczewki 1, a także pomocniczo soczewki 2. Kolejne zdjęcia zostały wykonane dla coraz większej różnicy dróg optycznych między ramionami interferometru (do kilku cm). Rozmiar obrazu interferencyjnego to około 15 cm

Można to bardzo ładnie zilustrować, ponieważ w interferometrze Michelsona częstym zabiegiem jest nie ręczne, ale automatyczne przesuwanie jednego z lusterek w zakresie kilku-kilkunastu mikrometrów, dzięki zamocowaniu go na elemencie piezoelektrycznym. Ruch lustra powoduje niewielką, płynną zmianę różnicy dróg optycznych w ramionach interferometru, a to z kolei prowadzi do przesuwania się prążków liniowych wzdłuż osi x z rys. 3. (zmiana Δs we wzorze 8) lub do koncentrycznego malenia lub rośnięcia prążków okrągłych. Oba przypadki są zilustrowane na krótkich filmach dołączonych do artykułu.

Na koniec głównej części artykułu chcę zwrócić uwagę na jeszcze jedną rzecz. Kiedy analizujemy działanie interferometrów, a szczególnie interferometru Michelsona, możemy poczuć pewien niepokój. Wyobraźmy sobie, że tak dobrze wyregulowaliśmy układ optyczny, że na ekranie obserwujemy tylko jeden wielki prążek. Na dodatek, dzięki odpowiedniemu doborowi różnicy dróg optycznych, jest to prążek zupełnie ciemny. Do ekranu nie dociera zatem żadne światło. Ale przecież my do interferometru świecimy laserem! Co się dzieje ze

światłem, z energią? Wszystko jest w porządku – interferometr tak naprawdę ma dwa wyjścia. Analiza układu prowadzi do wniosku, że kiedy w jednym z wyjść mamy prążek ciemny, to w drugim – jasny. Oba wyjścia są komplementarne. Problem w tym, że w interferometrze Michelsona to drugie wyjście idealnie pokrywa się z wejściem i często go nie zauważamy, ponieważ „wydobycie” światła z tego wyjścia jest możliwe, ale nieco kłopotliwe. Interferometr Macha-Zehndera natomiast ma te dwa wyjścia geometrycznie rozdzielone i bardzo dobrze dostępne. Szkic tego interferometru wraz z prążkami obserwowanymi na jego obydwu wyjściach jest przedstawiony na rys. 7.



Rys. 7. Schemat interferometru Macha-Zehndera wraz z przykładowymi obrazami interferencyjnymi z obu wyjść. Autorką zdjęcia prążków jest Aleksandra Sierant

Albert Michelson był fizykiem amerykańskim pochodzenia polsko-żydowskiego, żyjącym na przełomie XIX i XX wieku. Dokonał precyzyjnych pomiarów prędkości światła, a wraz z Edwardem Morley’em przeprowadził słynny eksperyment interferometryczny, który przyczynił się do powstania szczególnej teorii względności⁴. W 1907 roku został laureatem Nagrody Nobla za „skonstruowanie precyzyjnych instrumentów optycznych i wykonanie z ich wykorzystaniem pomiarów spektroskopowych i metrologicznych”. Skonstruował również interferometr gwiazdowy do zastosowań astronomicznych⁵.

Warto zauważyć, że interferometr Michelsona jest używany do dzisiaj, zarówno w badaniach naukowych, jak i w dydaktyce, niejednokrotnie w wersji praktycznie takiej samej, jak ta sprzed 140 lat. Zmodyfikowane konstrukcje są również bardzo popularne. Wariant z automatycznie przesuwającym lustrem jednego z ramion stanowi serce spektroskopii fourierowskiej, a wielkich rozmiarów interferometr Michelsona jest obecnie główną częścią optycznego detektora fal grawitacyjnych. Podobnie, do dzisiaj używa się interferometru gwiazdowego,

⁴ Bezpośredni wpływ wyników eksperymentu na sformułowanie przez Einsteina szczególnej teorii względności jest przedmiotem dyskusji, jednak wpływ co najmniej pośredni wydaje się pewny.

⁵ Biografia Alberta Michelsona jest przedstawiona w artykule w *Fotonie* 84 z 2004 roku.

przy pomocy którego Michelson i jego współpracownik 100 lat temu zmierzyli po raz pierwszy średnicę gwiazdy innej niż Słońce.

Przeglądając w Internecie materiały dotyczące działalności naukowej Alberta Michelsona, odniosłem wrażenie, że często informacje dotyczące pomiarów bezwzględnej prędkości światła i pomiarów z użyciem interferometru są wymieszane i wprowadzają czytelników w błąd. Dlatego poniżej przedstawiam w bardzo syntetycznej formie informacje o niektórych eksperymentach Michelsona.

1877-1879, 1924-1926, 1929-1933 (już częściowo po śmierci Michelsona) – pomiary bezwzględnej prędkości światła z użyciem aparatury bazującej na ulepszonej metodzie Léona Foucaulta, w której mierzy się czas przelotu światła (tam i z powrotem) na odcinku o znanej długości, z pomocą wirującego lustra. Pomiary te w żaden sposób nie bazowały na zjawisku interferencji światła.

1881 – pierwszy eksperyment z użyciem interferometru, który miał pomóc zrozumieć zachowanie hipotetycznego eteru w pobliżu Ziemi poprzez wykazanie zmian w różnicy czasu przelotu światła w prostopadłych ramionach interferometru, objawiającej się zmianami w położeniu prążków interferencyjnych. Eksperyment okazał się niewystarczająco dokładny.

1887 – kolejny eksperyment interferometryczny, przeprowadzony razem z Edwardem Morley’em. Ogólna idea pomiaru nie zmieniła się, ale interferometr został zmodyfikowany przez wprowadzenie wielokrotnych odbić światła w ramionach. Zwiększona czułość pozwoliła stwierdzić, że hipotetyczny eter musiałby być „całkowicie wleczony” przez Ziemię. Ani ten eksperyment, ani poprzedni z 1881 roku nie miały na celu pomiaru bezwzględnej prędkości światła.

1892-1893 – pomiar interferencyjny, dzięki któremu mechaniczny wzorzec metra, przechowywany w Paryżu, został powiązany z długością fali światła kadmu.

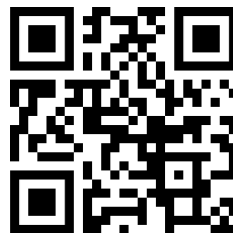
1920 – pierwszy pomiar średnicy odległej gwiazdy (Betelgezy), wykonany z użyciem interferometru gwiazdowego Michelsona.

Prążki interferencyjne otrzymane w interferometrze Michelsona dla lasera helowo-neonowego. Jedno z ramion interferometru ma płynnie zmienianą długość w zakresie około 500 nanometrów dzięki użyciu elementu piezoelektrycznego. W zależności od geometrii układu prążki mogą mieć różny kształt.

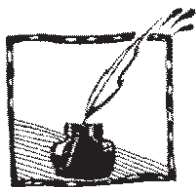
Zeskanuj i zobacz:



Prążki liniowe. <https://youtu.be/XsWBEsIzRMA>



Prążki koncentryczne. <https://youtu.be/4rtcpLYk8rM>



Piękno zamknięte w diamentach

Mariusz Mrózek

Wydział Fizyki, Astronomii i Informatyki Stosowanej,
Uniwersytet Jagielloński, Kraków

Diamenty są wieczne to tytuł filmu opartego na przygodach brytyjskiego agenta Jamesa Bonda z 1971 roku na podstawie książki Iana Fleminga pod tym samym tytułem. Diamenty fascynują jednak nie tylko filmowców i pisarzy, ale są obecne także wokół nas – mają bardzo szerokie zastosowanie: od przemysłu przez branżę jubilerską aż po współczesną naukę.

W diamencie mogą występować pewne niedoskonałości. Defekty te są niezwykłym polem doświadczalnym współczesnej nauki. Zmiany w strukturze krytalograficznej oraz w składzie diamentu powodują, że mogą one służyć jako czujniki magnetyczne, elektryczne, ciśnieniowe i temperaturowe. Znalazły one także zastosowanie w biologii i medycynie.

Czym jest diament?

Diament jest rzadkim minerałem występującym w przyrodzie. Jego nazwa wywodzi się z języka starogreckiego *adamas*, w łacinie określanego jako *diamentum* czyli w tłumaczeniu *niezniszczalny, niepokonany*. Jest to nawiązanie do jednej z jego wyjątkowych własności, jaką jest nadzwyczajna twardość. Do tej pory jest najtwardszą znaną substancją występującą w przyrodzie. Odpowiedzmy zatem na pytanie z nagłówka tego akapitu. Diament to po prostu węgiel, więc warto przybliżyć najpierw czym jest węgiel, by zrozumieć własności diamentu. Liczba atomowa węgla to 6, jest niemetalem z bloku p układu okresowego. Należy w nim do 14 grupy i ma 4 elektrony walencyjne. Najbardziej rozpowszechniony stabilny izotop to ^{12}C (98,9% – zawartość danego pierwiastka w pierwiastku naturalnie występującym, inaczej *abundancja*) ma taką samą liczbę protonów co neutronów. Kolejnym stabilnym izotopem jest ^{13}C (1,1 %) oraz radioizotop ^{14}C (śladowe ilości w przyrodzie). Ten ostatni jest radioaktywny i ma czas połowicznego rozpadu wynoszący 5730 lat (jest to czas, po którym zawartość pierwiastka w badanym materiale zmniejsza się o połowę), co powoduje, że ma zastosowanie w datowaniu materii organicznej. Maksymalny wiek próbek, dla których można stosować metodę datowania radiowęglowego to około 60 tysięcy lat¹. Węgiel jest czwartym najbardziej rozpowszechnionym pierwiastkiem we Wszechświecie², oraz drugim pod względem masy w ciele człowieka.

¹ Bagnicka M. „Węgiel – źródło informacji o wieku znalezisk”, *Analit*, 2018 (5), 17

² https://pl.wikipedia.org/wiki/Cz%C4%99sto%C5%9B%C4%87_wyst%C4%99powania_pierwiastk%C3%B3w_we_Wszech%C5%9Bwiecie

Niektóre pierwiastki chemiczne, głównie niemetale, między innymi węgiel, występują w różnych postaciach, które mają odmienne właściwości fizyczne i różną aktywność chemiczną. Postaci te zwane są odmianami alotropowymi i różnią się budową sieci krystalicznej. Główne odmiany alotropowe węgla to:

- grafit,
- diament,
- grafen,
- fulereny.

Grafit jest najbardziej rozpowszechnioną stabilną strukturą węgla. Jego sieć krystaliczna zbudowana jest z płaskich warstw ułożonych jedna nad drugą, z których każda warstwa przypomina strukturę plastra miodu. Atomy węgla są ułożone w regularne sześciokąty o wspólnych bokach. W obrębie każdej warstwy atomy są połączone silnymi wiązaniami kowalencyjnymi z trzema sąsiednimi atomami tego pierwiastka. Natomiast między warstwami występują tylko słabe oddziaływania (van der Waalsa), dlatego kryształy grafitu są miękkie i łatwo dają się łupać. Odległości między płaszczyznami są niemal 2,5 razy większe niż długości wiązań między atomami węgla w pierścieniach, stąd siła wiązań między warstwami jest mała. Dlatego poszczególne warstwy grafitu stosunkowo łatwo się rozdzielają. Wykorzystuje się to na przykład podczas pisania ołówkiem, którego rdzeń (rysik) jest z grafitu i podczas pisania na papierze odrywane są kolejne warstwy z rysika i nanoszone na papier. Grafit ma doskonałe właściwości smarne, jest tłusty w dotyku i wytrzymały na wysokie temperatury. Wykazuje wysoką odporność mechaniczną na ściskanie i niewielką na rozciąganie. Ponadto grafit i wyroby grafitowe nie rozpuszczają się w wodzie, charakteryzują się małą aktywnością chemiczną, łatwością składowania i utylizacji oraz brakiem negatywnego wpływu na środowisko.

Fulereny³ – są to cząsteczki zbudowane z parzystej liczby atomów węgla (od 28 do 1500), które tworzą zamkniętą strukturę. Najbardziej popularny – fuleren C_{60} – wyglądem przypomina piłkę futbolową i jest zbudowany z 60 atomów (oczywiście węgla), które tworzą 12 pierścieni pięciocłonowych i 20 pierścieni sześciocłonowych. Fullereny stosuje się w medycynie (nośniki leków)⁴; w życiu codziennym (wydłużanie trwałości żywności, produkcja materiałów fotoprzewodzących stosowanych w drukarkach); w elektrotechnice i katalizie (diody, katalizatory); w nanotechnologii (układy mikroelektroniczne); w elektronice polimerowej (fotoogniwa polimerowe)⁵.

Grafen – jest to warstwa grafitu o grubości jednego atomu⁶. Warstwa grafenu jest niezmiernie cienka, więc zazwyczaj nanosi się go na inne materiały, któ-

³ H.W. Kroto, et al „ C_{60} : Buckminsterfullerene”. Nature, 1985, 318 (6042), pp. 162-163.

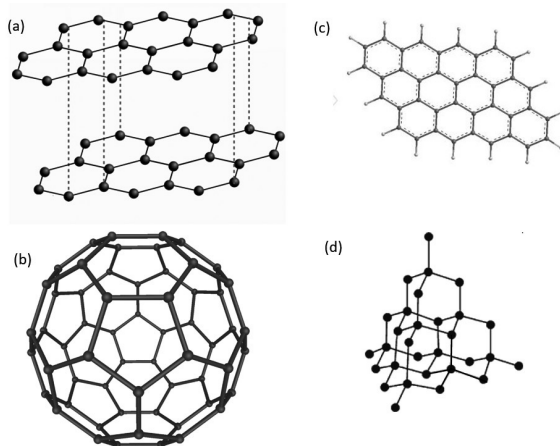
⁴ S. Goodarzi, et al, “Fullerene: biomedical engineers get to revisit an old friend”, Materials Today, 2017, pp. 460-480.

⁵ J. Nelson, “Polymer:fullerene bulk heterojunction solar cells”, 2011, Materials Today, 14 (10), pp. 462-470,

⁶ Novoselov, K. S et al, . “Electric Field Effect in Atomically Thin Carbon Films”. Science, 2004, 306 (5696) pp. 666–669.

rym nadaje on niezwykle właściwości. To znakomity przewodnik ciepła – dziesięciokrotnie lepszy od srebra. Podobnie jak grafit jest również bardzo dobrym przewodnikiem prądu elektrycznego. W porównaniu ze stalą ten nowy materiał ma od 5-6 razy mniejszą gęstość, jest dwukrotnie twardszy, 13-krotnie bardziej elastyczny i 100-krotnie bardziej wytrzymały na rozciąganie. Jego wiązania atomowe są tak ściśle, że przy pewnej modyfikacji może służyć jako filtr do uzdatniania słonecznej wody⁷. Obecnie to bardzo drogi materiał, ale ze względu na swoje doskonałe właściwości z pewnością zastąpi w przyszłości krzem w urządzeniach elektronicznych.

Teraz dochodzimy do głównego tematu, którym jest diament. Diament zbudowany jest z atomów tworzących regularną sieć przestrzenną o kształcie czworoscianu foremnego (tetraedru), w której każdy atom łączy się z czterema innymi atomami węgla. Równomiernie rozłożone, krótkie, a zarazem mocne wiązania kowalencyjne wpływają na bardzo dużą twardość tej odmiany alotropowej (w skali Mohsa twardość diamentu wynosi 10)⁸.



Rys. 1. Odmiany alotropowe węgla: (a) grafit, (b) fulleren, (c) grafen, (d) diament⁹

W zakresie widzialnym diament jest przezroczysty. Ta transparentność sięga tak naprawdę od światła ultrafioletowego (długość fali od 225 nm) do dalekiej podczerwieni (długość fali ponad 20 μm). Własności fizyczne diamentów to m.in. gęstość wynosząca 3,53 g/cm³, współczynnik załamania światła $n = 2,41$ (dla długości fali światła 532 nm), przerwa energetyczna – 5,4 eV, co powoduje, że jest dobrym izolatorem elektrycznym (chyba, że jest domieszkowany borem, co zmienia jego właściwości elektryczne), współczynnik przewodnictwa cieplnego = $2 \cdot 10^3$ W/mK (bardzo dobrze odprowadza ciepło), wytrzymałość na ściskanie wynosi 110 GPa, a wytrzymałość na rozciąganie 2,2 GPa. Jest odporny chemicznie na rozpuszczalniki, kwasy i zasady.

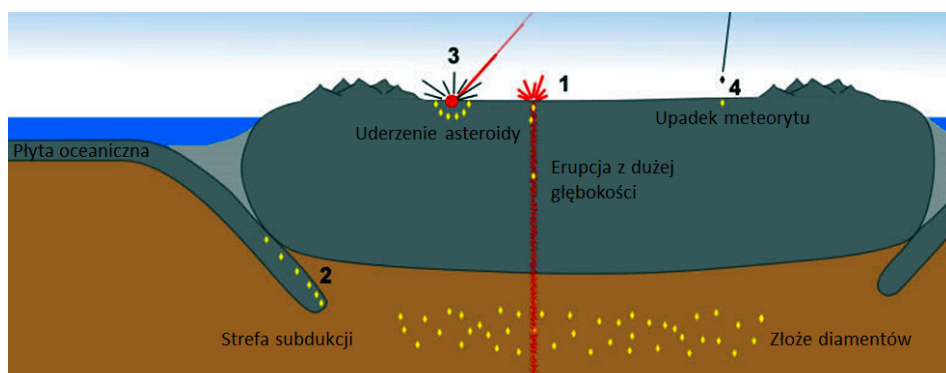
⁷ <https://www.graphene-info.com/graphene-water-treatment>.

⁸ <https://geology.com/minerals/mohs-hardness-scale.shtml>

⁹ [https://pl.wikipedia.org/wiki/W%C4%99giel_\(pierwiastek\)](https://pl.wikipedia.org/wiki/W%C4%99giel_(pierwiastek))

Powstawanie diamentów

Naturalnie diamenty występują bardzo rzadko na powierzchni naszej planety. Węgiel, aby mógł zamienić się w diament (skryształizować), potrzebuje odpowiednio wysokiego ciśnienia około 5 GPa, (dla porównania ciśnienie atmosferyczne jest 50 000 razy mniejsze) oraz wysokiej temperatury (1000–1600°C). Takie warunki występują w płaszczu Ziemi na głębokości 150–200 km. Czyli muszą istnieć jakieś procesy, które powodują wydostanie się diamentów na powierzchnię planety. Diamenty powstałe w skorupie ziemskiej mogą zostać wyrzucone na powierzchnię Ziemi przez erupcję wulkaniczną (1). Druga metoda to tak zwana subdukcja, czyli proces wpychania jednej płyty litosferycznej pod drugą (np. oceaniczną pod kontynentalną) (2). Diamenty mogły powstać również w wyniku uderzenia asteroidy w powierzchnię Ziemi, co spowodowało krystalizację węgla w glebie (3) lub mogły już być przyniesione z kosmosu przez meteoryt (4).



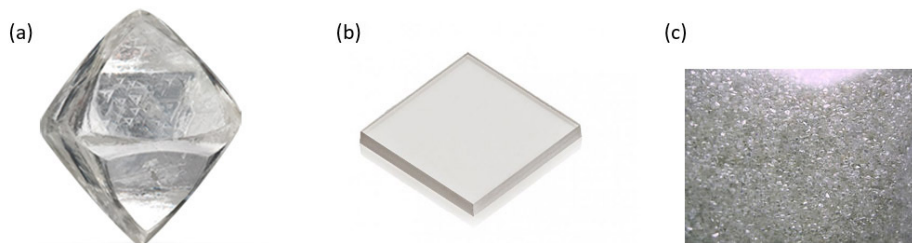
Rys. 2. Metody, dzięki którym diamenty znajdują się na powierzchni planety¹⁰

Z powodu trudności w wydobywaniu diamentów i ich ciągłego zapotrzebowania w branży przemysłowej i jubilerskiej naukowcy starali się znaleźć metody tworzenia sztucznych diamentów. Pierwszy raz otrzymano syntetyczne diamenty w 1954 roku w laboratorium firmy General Electric. Zastosowana metoda była bardzo podobna do sposobu, w jaki diamenty powstają w naturze w płaszczu Ziemi. Metoda ta nosi nazwę HPHT (ang. *High Pressure High Temperature*), czyli wysokie ciśnienie i wysoka temperatura. Prace nad zbudowaniem maszyny, która wytrzyma ciśnienia rzędu 10 GPa i temperaturę 1600°C trwały 4 lata. Udało się otrzymać diament, w którym nastąpiła pełna krystalizacja, czyli zamiana węgla w diament. Bardzo podobne prace prowadzili także w owym czasie naukowcy z ZSRR, lecz nie udało im się to do końca i otrzymywali oni tzw. czarne diamenty, czyli takie, w których nie do końca została przeprowadzona krystalizacja. Dzięki metodzie HPHT udaje się produkować próbki monokrystaliczne o objętości kilku milimetrów sześciennych. To wydarzenie zrewolucjonizowało przemysł diamentowy i uniezależniło branżę diamentów od tradycyjnego

¹⁰ <https://geology.com/articles/diamonds-from-coal/>.

wydobycia. Drugą stosowaną metodą jest chemiczne osadzanie z fazy gazowej CVD (ang. *Chemical Vapour Deposition*). Ta metoda polega na wprowadzeniu do komory reakcyjnej prekursorów H_2 , CH_4 oraz dostarczeniu energii w formie mikrofal (moc rzędu kW). Powoduje to powstawianie plazmy, która to rozkłada prekursory na pojedyncze składniki, a one osiadają na specjalnie przygotowanym podłożu. W ten sposób nakładane są kolejne warstwy diamentu. Dzięki temu, w większości przypadków, dostajemy próbki polikrystaliczne. Niestety proces ten jest bardzo powolny, gdy chcemy otrzymać grube próbki diamentowe.

Ostatnią metodą, dzięki której otrzymujemy diamenty, a tak naprawdę mikrodiamenty, jest metoda detonacyjna. Polega ona na wywołaniu eksplozji w silnym materiale wybuchowym zawierającym dodatek związków $C_aH_bN_cO_d$ z niedostatkim tlenu. Ciśnienie i temperatura niesione razem z falą uderzeniową powstającą w eksplozji przekraczają 10 GPa i 3000°C. Węgiel w tym procesie przechodzi w diament (ok. 75%) w postaci kryształitów o wielkościach od 4 nm w górę. Ta metoda ma szczególne zastosowanie w przemyśle. Tak powstałe mikrodiamenty stosuje się najczęściej w materiałach ściernych (tarcza szlifierska, piły diamentowe).



Rys. 3. Zdjęcie (a) naturalnego diamentu, (b) syntetycznego otrzymanego techniką HPHT, (c) proszku diamentowego otrzymanego techniką detonacyjną

Klasyfikacja diamentów

Diamenty, które mają zostać wykorzystane w branży jubilerskiej, muszą zostać scharakteryzowane (oszlifowany diament nazywa się brylantem). W tym celu stosuje się regułę 4C. Jest to skrót od 4 słów w języku angielskim zaczynających się na literę C: *carat* (masa w karatach), *cut* (cięcie, szlif), *color* (kolor) i *clarity* (przejrzystość, klarowność). Każdy sprzedawany przez jubilera diament powinien być opatrzony etykietą (certyfikatem), na której znajduje się jego masa, a także pozostałe parametry wg oceny eksperta.

Zacznijmy od masy, która podawana jest w jednostkach nazywanych **karatami**, oznaczanych symbolem ct. Nazwa „karat” pochodzi od nazwy drzewa szarańczyn strąkowy (łac. *ceratonia siliqua*), którego nasiona – nazywane chlebem świętojańskim – mają po wysuszeniu jednakową masę ok 0,2 g, stąd jeden karat odpowiada 0,2 grama. Karat można podzielić jeszcze na 100 punktów. Największy diament *Cullinan*, znaleziony w RPA, miał masę 3106 ct (621 g). Wykonano

z niego 105 brylantów, z czego największy *Cullinan I* miał 530 ct i zdobi teraz brytyjskie berło królewskie.

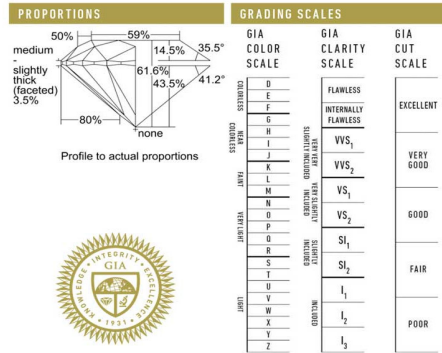
FACSIMILE
This is a digital representation of the original GIA Report. This representation might not be accepted in lieu of the original GIA Report in certain circumstances. The original GIA Report includes certain security features which are not reproducible on this facsimile.

GIA®

GIA REPORT
2326328623

Verify this report at GIA.edu

GIA NATURAL DIAMOND DOSSIER®	
March 25, 2019	
GIA Report Number	2326328623
Shape and Cutting Style	Round Brilliant
Measurements	5.10 - 5.12 x 3.15 mm
GRADING RESULTS	
Carat Weight	0.50 carat
Color Grade	G
Clarity Grade	VS2
Cut Grade	Excellent
ADDITIONAL GRADING INFORMATION	
Polish	Excellent
Symmetry	Excellent
Fluorescence	None
Clarity Characteristics.....	Cloud, Feather
Inscription(s): GIA 2326328623	



The results documented in this report refer only to the diamond described, and were obtained using the techniques and equipment available to GIA at the time of examination. This report is not a guarantee or valuation. For additional information and important limitations and disclaimers, please see GIA.edu/terms or call +1 800 421 7250 or +1 760 680 4500. ©2017 Gemological Institute of America, Inc.

reportcheck.gia.edu

GIA.edu

Rys. 4. Przykładowy certyfikat diamentu¹¹



Rys. 5. Różne typy szlifów diamentów¹²

Rys. 6. Zwieńczenie brytyjskiego berła królewskiego z widocznym brylantem¹³

Cut (rodzaj szlif). Ma bardzo istotny wpływ na jakość diamentu. Najbardziej popularny jest szlif brylantowy (okrągły, około 75 % wszystkich szlifów). Poza nim jest wiele innych. Wybrane szlify zostały przedstawione na rysunku 5.

¹¹ <https://www.i-diamants.com/en/diamond-certificates-gia-igi-hrd.html,00026>

¹² <https://blog.michelson.pl/2015/08/01/typy-szlifow-diamentow/>

¹³ <http://www.spbursztyn.slcd.pl/comenius/klejnoty.pdf>

Tylko idealny szlif gwarantuje wydobycie ukrytego piękna z kamienia i idealne odbijanie światła od faset. Ocena wykończenia i proporcji szlifów obejmuje następującą skalę:

- **Excellent** – doskonała,
- **Very Good** – bardzo dobra,
- **Good** – dobra,
- **Fair** – zadowalająca,
- **Poor** – słaba.

Według Amerykańskiej skali GIA (*Gemological Institute of America*), najczęściej używanej w Europie, skala barw diamentów oznaczana jest literami od D (najczystsze) do Z (intensywnie zabarwione), w następujący sposób:

- D, E, F – bezbarwne,
- G, H, I, J – bliskie bezbarwnych,
- K, L, M – lekko żółte,
- N, O, P, Q, R – jasne żółte,
- S, T, U, V, W, X, Y, Z – żółte.



Rys. 7. Wybrane kolory według klasyfikacji GIA¹⁴

Clarity (przejrzystość). Ocenę czystości diamentu wykonuje się przez obserwację za pomocą oka, używając przy tym lupy o min. 10-krotnym powiększeniu. Większość diamentów posiada znamiona wewnętrzne, tzw. inkluzje (bezpостaciowe wrostki) oraz znamiona zewnętrzne, tzw. skazy. Na podstawie ich zagęszczenia, rozmieszczenia oraz wielkości, określa się stopień czystości:

LC/FL – (Loup Clean, Free of Inclusions/Flawless) najczystszy diament, wolny od inkluzji oraz skaz,

VVS1, VVS2 – (Very, Very Small Inclusions) bardzo, bardzo małe zanieczyszczenia (inkluzje), widoczne tylko pod mikroskopem przy bardzo dokładnym badaniu. Niemal niewidoczne przez lupę o 10-krotnym powiększeniu,

VS1, VS2 – (Very Small Inclusions) bardzo małe zanieczyszczenia widoczne tylko przez lupę w tafli i podstawie diamentu,

SI – (Small Inclusions) nieznaczne inkluzje widoczne tylko przez lupę przy 10-krotnym powiększeniu, jednak gołym okiem są niedostrzegalne,

P1, P2, P3 – (1st, 2st, 3st Pique) są to kamienie najniższej klasy. Zanieczyszczenia są widoczne gołym okiem. Stopniowanie od 1 do 3 oznacza, że przy 1 inkluzje są widoczne gołym okiem, ale trudne do rozpoznania, natomiast przy 3 są widoczne gołym okiem oraz pokrywają ponad 50 % kamienia.

¹⁴ <https://www.gia.edu/diamond>

Poniżej przedstawione są ogólne zastosowania diamentów w różnych dziedzinach naszego życia.

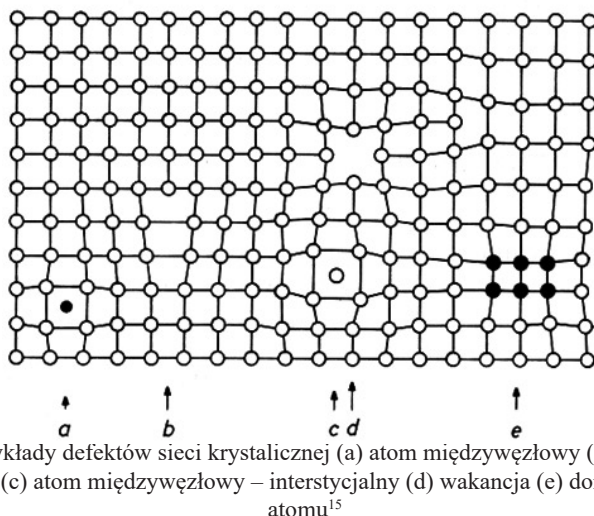
- W jubilerstwie do wyrobu biżuterii.
- Do produkcji materiałów ściernych, narzędzi tnących i skrawających, takich jak koronki diamentowe na końcówki wiertel, w tym też wiertła dentystyczne.
- Narzędzia do cięcia szkła (diamentem nacina się podziałki na najdokładniejszych przyrządach optycznych i siatki dyfrakcyjne spektrometrów).
- Diamentowe ostrza wykorzystywane w narzędziach chirurgicznych, takich jak diamentowy skalpel.
- Noże diamentowe służące do wygładzania tarcz ściernych, ostrzenia przyrządów skrawających i obrabiania powierzchni z dokładnością do pół mikrometra.
- Proszki i pasty ścierne wykorzystywane do szlifowania i polerowania nawet najtwardszych materiałów (diament szlifuje się diamentem). Około 80% wszystkich diamentów przemysłowych wykorzystuje się w stanie rozdrobnionym (jako proszek).
- Do wyrobu specjalistycznych urządzeń, takich jak końcówki do twardościomierzy, w instrumentach naukowych i precyzyjnych instrumentach pomiarowych.
- Jako materiał na półprzewodnik. Odpowiednio domieszkowany diament wykazuje własności półprzewodnikowe. Układy scalone na diamencie mogą pracować w wysokich temperaturach, jednocześnie tranzystor diamentowy ma wyższe napięcie przebicia i pracuje szybciej niż jego odpowiedniki krzemowe.

Diamenty wykorzystywane dla celów jubilerskich to ok. 20% wydobywanych. Reszta to diamenty przemysłowe.

Defekty występujące w kryształach

W naturze nigdy nie występują kryształy idealne. Kryształy naturalne zawierają różne defekty, które powodują pewne zniekształcenie sieci krystalicznej i są ośrodkami nagromadzenia energii (odkształcenia) wynikającej stąd, że każdy defekt powoduje przesunięcie atomów z ich położeń równowagi. Wpływ defektów sieci na własności kryształów jest bardzo istotny. Niedoskonałości te możemy podzielić na kilka kategorii:

- punktowe (wakancja – wolne miejsca w sieci krystalicznej, mogą to być także atomy międzywęzłowe),
- liniowe (m.in. krawędziowe, które powstają poprzez wprowadzenie dodatkowej płaszczyzny między nieco rozsunięte płaszczyzny sieciowe, oraz śrubowe, powstające w wyniku przesunięcia płaszczyzn atomowych),
- powierzchniowe (tzw. granice międzyziarnowe – to wąska strefa materiału, gdzie atomy ułożone są w sposób chaotyczny),
- przestrzenne (szczeliny i pory).



Rys. 8. Przykłady defektów sieci krystalicznej (a) atom międzywęzłowy (b) dyslokacja krawędziowa (c) atom międzywęzłowy – interstycjalny (d) wakancja (e) domieszka innego atomu¹⁵

Tak np. wakancje ułatwiają dyfuzję atomów, a wiadomo, że dyfuzja jest podstawą prawie wszystkich procesów i przemian zachodzących w materiałach. Z kolei defekty liniowe (dyslokacje) ułatwiają odkształcanie metali i bez nich przeróbka plastyczna niektórych materiałów byłaby niemożliwa. Bardzo ważne są także granice międzyziarnowe jako defekty powierzchniowe, które mogą być czynnikiem wzmacniającym materiał.

Idealny diament powinien składać się tylko z atomów węgla ^{12}C . Szczegółowe badania pokazują, że sztucznie wytworzone diamenty są bardziej zbliżone do ideału niż naturalne, a ich struktura jest niemal perfekcyjna. Wynika to z tego, że natura nie jest idealna, a całe jej piękno kryje się w defektach. Naturalnie w każdym diamencie występuje jeszcze izotop węgla (taka sama liczba protonów, inna neutronów) o liczbie masowej 13. Prócz tego występuje także azot ^{14}N , który zastępuje węgiel. W przypadku diamentów wyprodukowanych syntetycznie możemy zawartość domieszek kontrolować (zmniejszać lub zwiększać). Z tego względu możemy podzielić diamenty na:

- Typ Ia – stanowi 98% całkowitego wydobycia. Zawiera znaczne ilości azotu 14 (około 0,1%).
- Typ Ib – stanowi 1% całkowitego wydobycia. Zawiera do 0,01% azotu. Do tej grupy należą prawie wszystkie diamenty syntetyczne.
- Typ IIa – prawie nie zawiera azotu. Rzadko występuje w przyrodzie.
- Typ IIb – zawiera domieszkę boru – posiada błękitną barwę.

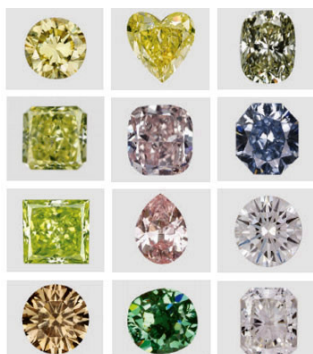
Centra barwne w diamentach

Najpopularniejszym defektem w diamencie jest centrum barwne. Jest to defekt w sieci krystalicznej zdolny do absorpcji i emisji światła. W większości polega na zastąpieniu atomu węgla atomem innego pierwiastka w różnych konfiguracjach.

¹⁵ https://www.tf.uni-kiel.de/matwis/amat/def_en/index.html

Obecnie znanych jest ponad 500 różnych typów centrów barwnych w diamentach. Na rysunku 9 przedstawione są zdjęcia diamentów zawierające różne centra barwne. Idąc od lewej górnej strony wierszami, mamy następujące:

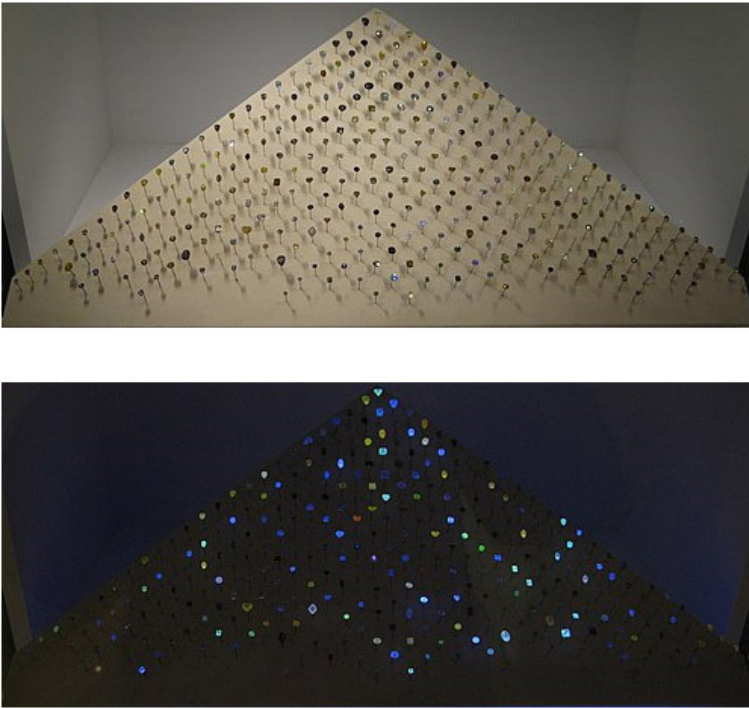
- **N3** – składa się on z 3 atomów azotu i wakancji między nimi (azot nadaje mu żółty kolor),
- **C** – jest to pojedynczy atom azotu otoczony węglem (azot nadaje mu żółty kolor),
- **H-rich** – zawiera dużą ilość wodoru (może dawać żółty lub fioletowy kolor w zależności od koncentracji),
- **H4** – mamy tu 4 atomy azotu i dwie wakancje połączone ze sobą (azot nadaje kolor żółty),
- **NV0** – połączenie azotu i wakancji bez wolnych elektronów (kolor różowy),
- **BV-** – duża zawartość boru połączonego z wakancją (niebieski kolor),
- **H3** – dwa atomy azotu połączone z wakancją (od żółtego po jasnozielony),
- **NV-** – połączenie azotu i wakancji z wolnym elektronem (kolor różowy) – zostanie omówiony szerzej.,
- **A** – dwa atomy azotu połączone ze sobą (brak koloru),
- **V** – wakancje połączone w klastry w strukturze diamentu (różne odcienie w zależności od liczby wakancji),
- **GR1** – pojedyncza wakancja o ładunku elektrycznym 0 (zielony odcień),
- **B** – składa się z 4 atomów i jednej wakancji (brak koloru).



Rys. 9. Zdjęcia diamentów z różnymi centrami barwnymi¹⁶

Każda konfiguracja domieszkowanego pierwiastka i wakancji może tworzyć różnego rodzaju zabarwienia diamentu. Bardzo interesującym dziełem sztuki wykorzystującym centra barwne w diamentach jest *Aurora Piramida Nadziei* – kolekcja 296 naturalnych diamentów w szerokiej gamie kolorów znajdująca się w Muzeum Historii Naturalnej w Londynie. Na rysunku 10 możemy zobaczyć, jak kolekcja wygląda w świetle białym (górne zdjęcie) oraz jej fluorescencję pod wpływem światła ultrafioletowego.

¹⁶ <https://www.withclarity.com/blog/2019/09/25/diamond-types-guide/>



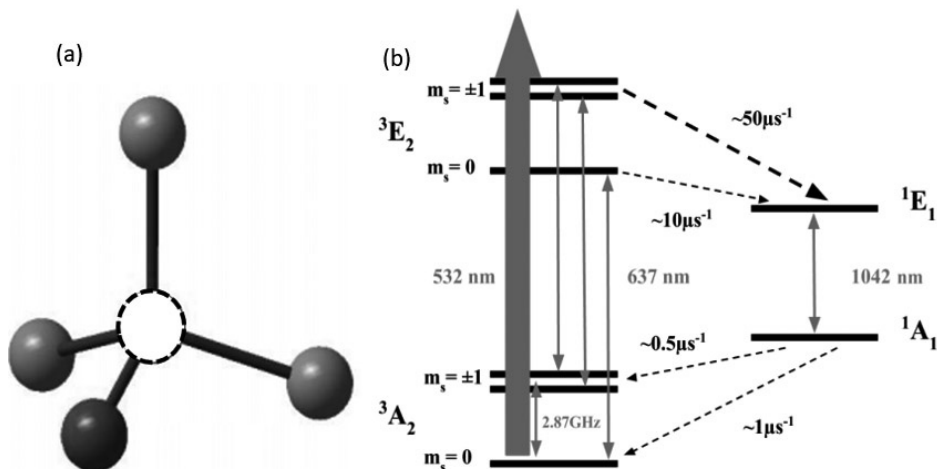
Rys. 10. *Aurora Piramida Nadziei*. Na górze widoczna w świetle białym, na dole fluorescencja pod wpływem światła ultrafioletowego¹⁷

Centrum barwne azot-wakancja

Centrum barwne azot-wakancja (NV) w diamencie składa się z azotu (najczęściej izotopu ^{14}N (99,634% naturalnego składu), rzadziej ^{15}N (0,366%)), w którego najbliższym sąsiedztwie znajduje się wakancja V. Atom azotu ^{14}N posiada pięć elektronów walencyjnych. Trzy z nich związane są z atomami węgla znajdującymi się wokół niego. Pozostałe dwa skierowane są do wakancji. Układ taki nazywany jest neutralnym centrum barwnym NV^0 . Centrum NV^0 wydaje się obiecującym kandydatem do zastosowań w systemach informacji kwantowej ze względu na swój paramagnetyzm (spin $S = 1/2$), lecz jest jeszcze stosunkowo słabo zbadany. Jeśli do wakancji przyłączony zostanie kolejny elektron, to ten defekt zamienia się na centrum barwne NV^- , którym będziemy się głównie zajmować. Rzadziej występuje sytuacja, gdy są tylko cztery elektrony i mamy wtedy centrum barwne NV^+ .

¹⁷ https://en.wikipedia.org/wiki/Aurora_Pyramid_of_Hope

Ponieważ defekt NV^- w diamencie słabo oddziałuje z siecią krystaliczną nawet w temperaturze pokojowej, to może być traktowany jako „sztuczny atom” uwieczony w sieci krystalicznej. Centrum barwne NV^- posiada symetrię C_{3v} .

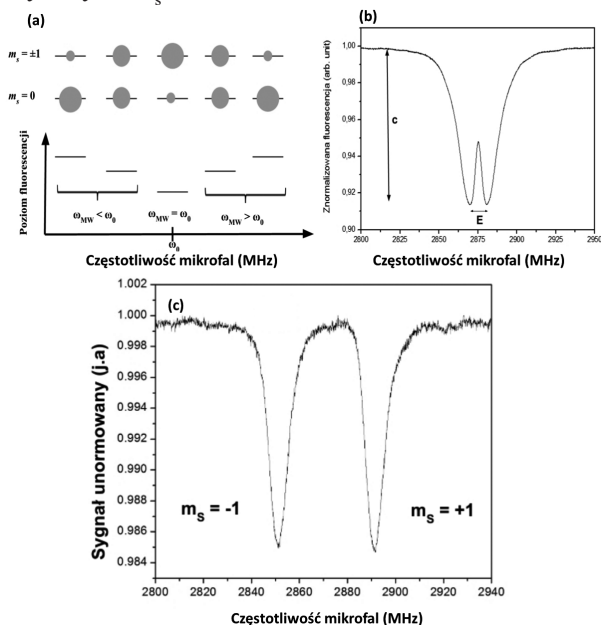


Rys. 11. (a) Budowa komórki elementarnej centrum barwnego NV^- . Czarna kulka oznacza azot, kulka oznaczona przerywaną linią oznacza wakancję, a szara węgiel. (b) Schemat poziomów elektronowych centrum. Linie ciągłe oznaczają przejścia promienne, a linie przerywane przejścia bezpromienne. Grubsza linia przerywana oznacza większe prawdopodobieństwo przejścia. Wartości liczbowe przy liniach przerywanych przedstawiają typowe raty relaksacji

Centrum barwne NV^- ma spin $S = 1$, a zatem posiada poziomy elektrony trypletowe i singletowe. Poziomy podstawowy (rys. 4b) jest trypletem. W zerowym polu magnetycznym jest on rozszczepiony o 2,87 GHz pod wpływem oddziaływania spin-spin na podpoziom $m_s = 0$ i dwa podpoziomy $m_s = \pm 1$. Stan wzbudzony centrum NV^- także jest trypletem. Oznaczony jest ³E₂ i podobnie jak stan podstawowy jest rozszczepiony o wartość 1,42 GHz. Przejście pomiędzy stanem podstawowym a wzbudzonym dla trypletu ma energię 1,945 eV, co odpowiada długości fali 637 nm. Przejście to odbywa się bez udziału fononów i nazywane jest przejściem zerofononowym (*ZPL – Zero-Phonon Line*). Stan singletowy składa się także ze stanu podstawowego ¹A₁ i wzbudzonego ¹E₁. Przejście ¹E₁ → ¹A₁ ma linię zerofononową o energii 1,19 eV (1042 nm). Wzbudzenie optyczne centrum NV^- jest możliwe dzięki istnieniu licznych poziomów oscylacyjnych (fononowych). Poziomy te są wzbudzane zielonym światłem a następnie szybko bezpromieniście kaskadują do stanu ³E₂, skąd relaksują do stanu podstawowego ³A₂, zarówno przez przejścia bezpośrednie, jak i przez przejścia międzysystemowe (za pośrednictwem stanów singletowych). Przez takie pompowanie optyczne z konwersją międzysystemową w układzie wytwarzana jest polaryzacja spinowa mogąca osiągnąć nawet 80%. Konwersja pomiędzy trypletem a singletem odbywa się niepromieniście (tzw. nieradiacyjne przejścia międzysystemowe *ISC – ang. Intersystem Crossing*), co przedstawione jest na rysunku 11b za pomocą przerywanych strzałek. Przejścia z trypletowego stanu wzbudzonego ³E₂ do sin-

gletu 1E_1 odbywają się z większym prawdopodobieństwem ze stanów $m_s = \pm 1$ niż z $m_s = 0$. Ze stanu 1E_1 następuje bardzo szybkie przejście do stanu 1A_1 (głównie promieniste, choć także możliwe jest niepromieniste). Następnie z dolnego stanu singletowego 1A_1 do trypletowego stanu podstawowego 3A_2 następuje niepromieniste przerzucenie populacji.

Dla centrów barwnych NV^- wykorzystywana jest podstawowa metoda badawcza nazwana optycznie wykrywanym rezonansem magnetycznym (*ODMR – ang. Optically Detected Magnetic Resonance*). Polega ona na obserwacji zmiany natężenia fluorescencji próbki wzbudzonej zielonym laserem podczas skanowania częstości pola mikrofalowego wokół rezonansu magnetycznego. Gdy centrum azot-wakancja zostanie wzbudzone ze stanu $m_s = 0$ poziomu podstawowego 3A_2 , to fluorescencja będzie silniejsza niż w przypadku wzbudzenia stanów $m_s = \pm 1$ 3A_2 (rys. 5a) (spowodowane jest to dostępnością innego kanału deekscytacji). Po wytworzeniu polaryzacji spinowej stanu podstawowego możliwa jest obserwacja przejść pomiędzy podpoziomami stanu podstawowego o różnych liczbach magnetycznych m_s .

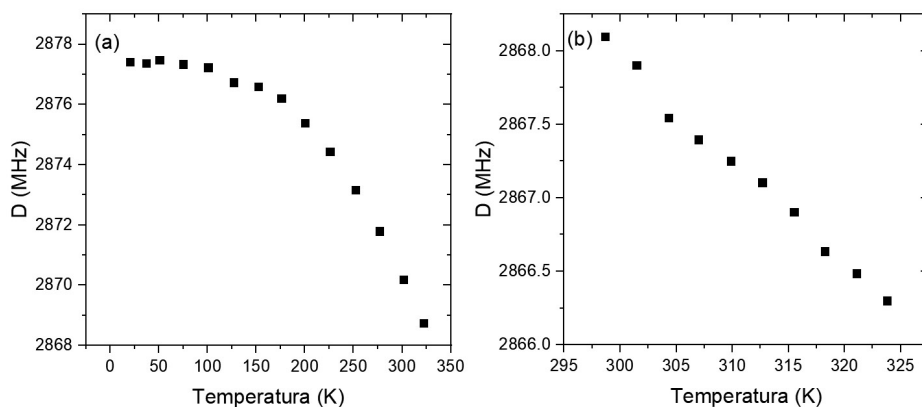


Rys. 12. (a) Rysunek schematycznie przedstawia populacje stanów $m_s = 0$ i $m_s = \pm 1$ oraz związane z nimi natężenie fluorescencji w funkcji częstotliwości mikrofal ω_{MW} względem rezonansu ω_0 .

Rozmiar szarej kropki symbolizuje populację danego stanu (założono tutaj, że pole mikrofalowe ma formę impulsu i może wywołać inwersję populacji stanów m_s , tzw. impuls π). (b) Przykładowe widmo ODMR dla próbki diamentowej NV^- zarejestrowane w temperaturze 77 K, w zerowym polu magnetycznym, (c oznacza głębokość rezonansu, E – wartość rozszczepienia zeropolowego). (c) Widmo ODMR w niezerowym polu magnetycznym. Widoczne jest rozszczepienie rezonansów magnetycznych

Na rysunku 12b widoczne jest widmo rezonansu ODMR dla centrum barwnego azot-wakancja w zerowym polu magnetycznym. Względna głębokość rezonansu (wyrażona w procentach) oznaczona jest literą c i nazwana jest kontrastem. Typowy kontrast dla zespołów centrów barwnych wynosi od kilku do kilkunastu procent dla zerowego pola magnetycznego, a dla pojedynczych centrów dochodzi maksymalnie do 40%.

W zerowym polu magnetycznym mamy rozszczepienie między stanami spinowymi równe 2,87 GHz. Stany $m_s = \pm 1$ są rozszczepione w zerowym polu magnetycznym z powodu występowania naprężenia sieci E . W niezerowym polu magnetycznym stany $m_s = \pm 1$ rozszczepiają się 5,6 MHz/G. Dzięki takiej analizie widm jesteśmy w stanie nie tylko określić wielkość przyłożonego pola magnetycznego, ale także podać jego dokładny kierunek. Pozwala to na magnetyczne mapowanie (wykreślanie rozkładu pola magnetycznego) danej powierzchni za pomocą centrum barwnego azot-wakancja. Na rysunku 12c widoczne jest widmo ODMR w niezerowym polu magnetycznym. Na podstawie różnicy położenia środków rezonansów możemy policzyć wartość pola magnetycznego przyłożonego do układu. Dla tych pomiarów wartość pola wynosiła 7 G.

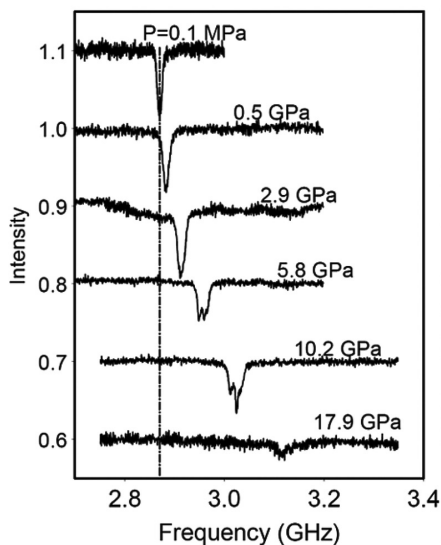


Rys. 13. Zależność położenia rezonansu ODMR dla temperatury (a) w zakresie od 10 K do 320 K, (b) wokół temperatury pokojowej

Położenie rezonansu ODMR w zerowym polu magnetycznym pochodzącego od centrum barwnego azot-wakancja (D) w diamencie zależy od temperatury. Efekt powstaje z powodu rozszerzalności cieplnej oraz oddziaływania elektronów z fononami, ale jego dokładny mechanizm nie jest jeszcze w pełni wyjaśniony. Na rysunku 13 przedstawione zostały wyniki pomiarów położenia rezonansu w zerowym polu magnetycznym w zależności od temperatury. Dla temperatur poniżej 80 K przesunięcie jest praktycznie niemierzalne (Rys. 13a). Powyżej 80 K widoczne jest przesunięcie rezonansu ku niższym częstotliwościom, coraz większe ze wzrostem temperatury. Ilustruje to przejście od dominacji oddziaływania spin-spin do przypadku zdominowanego przez oddziaływanie fononowe. Dla dostatecznie małego zakresu temperatur zmiana ta może być w przybliżeniu

traktowana jako liniowa. Dzięki temu centrum barwne NV^- może być miernikiem temperatury, nawet w organizmach żywych w temperaturze bliskiej pokojowej. W temperaturze bliskiej pokojowej przesunięcie to wynosi około -74 kHz/K (Rys. 13b). Gdy umieścimy w badanym materiale większą liczbę nanodiamentów z centrami barwnymi azot-wakancja, to mierząc sygnały z różnych miejsc próbki, można także mierzyć gradient temperatury.

Centrum barwne azot-wakancja może także służyć jako miernik ciśnienia wywieranego na badany układ. Także za pomocą widma ODMR można mierzyć wartość wywieranego ciśnienia. Na rysunku 14 przedstawione są przykładowe wyniki na podstawie referencji². Przesunięcie położenia rezonansu w zerowym polu magnetycznym (D) wynosi $14,58$ MHz/GPa. Oznacza to, że obserwując zmianę położenia rezonansu, odczytujemy wartość ciśnienia wywieranego na próbkę. Metoda ta może służyć jako alternatywa do stosowanych pomiarów wytrzymałościowych materiałów przemysłowych.

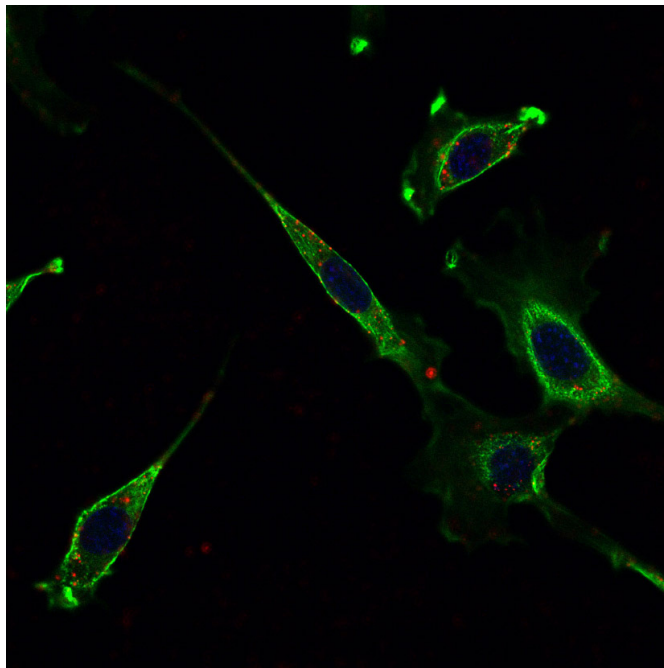


Rys. 14. Przykładowe widma ODMR w quasihydrostatycznym ośrodku ciśnieniowym NaCl. Linia przerywana wskazuje położenie rezonansu przy ciśnieniu atmosferycznym (Intensity – unormowany poziom fluorescencji, Frequency – częstotliwość mikrofal)¹⁸

Ponieważ nanodiament zawierający centrum barwne azot-wakancja składa się z węgla i azotu, nadaje się jako idealne medium do badań biologicznych. Taki nanodiament możemy umieścić w komórce biologicznej i wykorzystując możliwości centrów barwnych, badać jej własności. Najczęściej wykorzystuje się je do bioobrazowania. Na rysunku 15 przedstawione jest zdjęcie komórek biologicznych (wykonane za pomocą mikroskopu konfokalnego) z nanodiamentami

¹⁸ Marcus W. Doherty, Viktor V. Struzhkin, David A. Simpson, Liam P. McGuinness, Yufei Meng, Alastair Stacey, Timothy J. Karle, Russell J. Hemley, Neil B. Manson, Lloyd C. L. Hollenberg, and Steven Prawer, “Electronic Properties and Metrology Applications of the Diamond NV^- Center under Pressure”, Phys. Rev. Lett. 112, 047601 (2014).

zawierającymi centra barwne azot-wakancja. Komórka została oświetlona zielonym laserem (stąd zielona poświata komórek), co spowodowało, że nanodiamenty zaczęły świecić na czerwono. Dzięki nim możemy badać zmiany temperatury w komórkach pod wpływem czynników chemicznych, stanów zapalnych, zmian nowotworowych itd., co pozwoliłoby na szybką diagnozę pacjenta.



Rys. 15. Zdjęcie komórek biologicznych (zielone) z nanodiamentami zawierającymi centra barwne azot-wakancja (czerwone punkty)¹⁹

Zakończenie

Od wieków diamenty zachwycają nas swoim doskonałym pięknem i niezwyklejmi właściwościami, pokazują nam, że oprócz oczywistych zastosowań, z którymi je kojarzymy, mają cały wachlarz tajemnic.

Badania wykonywane z wykorzystaniem nanodiamentów z centrami barwnymi azot-wakancja dają możliwość praktycznego zastosowania w wielu naukach i dziedzinach życia, takich jak biologia, medycyna czy fizyka. Zastosowanie ich w biomedycynie i biotechnologii pozwala na bioobrazowanie z zachowaniem wyraźnej i długotrwałej fluorescencji. Z tego powodu nanodiamenty z centrami barwnymi wyróżniają się nie tylko na tle innych nanocząsteczek. Ciągłe badania pozwalają na odkrywanie coraz to nowych zastosowań dla centrów barwnych w diamentach.

Chciałbym podziękować dr. Adamowi Wojciechowskiemu za wspólne dyskusje na temat tych jakże szlachetnych kamieni.

¹⁹ Podziękowanie dla Armina Ebrahimi za udostępnienie zdjęcia.



Co czytać

Fizyka z serii The Manga Guide (PWN)

Hideo Nitta, Keita Takatsu, TREND-PRO Co., Ltd

Megumi to gwiazda sportu, jednak na lekcjach z fizyki jest słabym zawodnikiem. Na szczęście przyjaźni się z Ryotą – entuzjastą, który pomaga jej zrozumieć wszelkie zawiłości tej dziedziny nauki na przykładach wziętych z życia!

Czytając *The Manga Guide. Fizyka*, wraz z Megumi, poznasz zasady fizyki rządzące przedmiotami, takimi jak rolki czy proca, a także zachowaniem samochodu podczas hamowania oraz serwisami tenisowymi.

Błyskawicznie opanujesz trudne pojęcia, takie jak: pęd i ruch paraboliczny oraz zależność między siłą, masą i przyspieszeniem.

Nauczysz się również, jak:

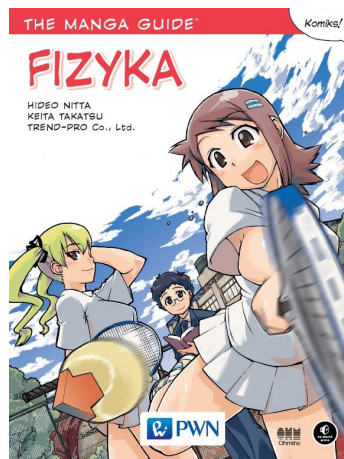
- w praktyce stosować trzy zasady dynamiki Newtona,
- określać, jak obiekty będą się poruszać po wzajemnym zderzeniu,
- rysować diagramy wektorowe i upraszczać trudne zagadnienia za pomocą trygonometrii,
- obliczać, w jaki sposób zmienia się energia kinetyczna obiektu, gdy rośnie jego energia potencjalna.

Seria *The Manga Guides* łączy formę ciekawych i zabawnych japońskich komiksów z praktyczną wiedzą z zakresu tak popularnych ostatnio dziedzin jak: fizyka, informatyka, czy biochemia. To kompilacja graficznych prac najważniejszych japońskich artystów zajmujących się mangą i wiedzy specjalistów z najbardziej prestiżowych uczelni.

Wkrótce w serii ukażą się także komiksy na temat:

- mikroprocesorów,
- baz danych,
- Wszechświata,
- biologii molekularnej,
- biochemii.

Podstawą zrozumienia fizyki jest właściwe „widzenie” tego, co chcemy zbadać.



W szczególności w mechanice klasycznej trzeba rozumieć, jak prawa fizyczne odnoszą się obiektów będących w ruchu. Niestety tradycyjne podręczniki rzadko zawierają odpowiednie obrazy takiego ruchu.

„The Manga Guide. Fizyka” jest próbą pokonania ograniczeń tradycyjnych podręczników przez użycie komiksu. Komiksy nie są jedynie zwykłymi ilustracjami – są ekspresyjnym i dynamicznym nośnikiem, który może przedstawiać upływ czasu. Umożliwiają obrazowe wyrażenie zmian w ruchu. Mogą także zmienić prawa, które wydają się nieciekawe, oraz nierealne scenariusze w rzeczy znane, przyjazne i łatwe do zrozumienia.

Główną postacią książki jest uczennica liceum, Megumi Ninomiya, dla której fizyka jest trudna. Moim szczerym pragnieniem jest to, aby tę książkę przeczytało możliwie jak najwięcej czytelników, którzy myślą, że „fizyka jest trudna” i którzy „nie lubią fizyki”, i aby znaleźli – choćby najmniejszą – przyjemność w nauce fizyki, tak jak stało się z Megumi.

Hideo Nitta

Nasze miejsce w kosmosie (PWN)

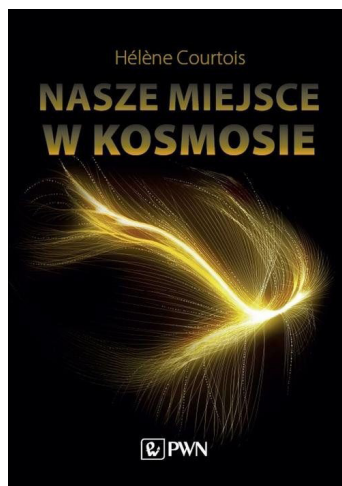
Hélène Courtois

Nasza galaktyka i jej sąsiedzi mkną przez Wszechświat w zdumiewającym tempie kilkuset kilometrów na sekundę. Fakt ten jest znany od początku lat 60. ubiegłego wieku. W tym czasie astrofizycy nie byli jeszcze w stanie w pełni wyjaśnić przyczyn tego zjawiska. W latach 90. amerykańscy naukowcy wskazali, że ruch ten może być związany z oddziaływaniem ogromnej masy tzw. Wielkiego Atraktora.

Poszukując Wielkiego Atraktora, zespół badawczy, którego członkiem była nasza Autorka, odkrył supergromadę galaktyk, do których należy nasza Droga Mleczna. Nazwali ją Laniakea (z hawajskiego niezmiernie niebiosy).

Ta książka ma na celu podzielenie się historią tego ogromnego odkrycia. Autorka w jasny i prosty sposób dokonuje przeglądu Wszechświata i rządzących nim praw fizyki.

Hélène Courtois - francuska astrofizyczka specjalizująca się w kosmografii. Jest profesorem na Uniwersytecie w Lyonie.



III Światowa Konferencja Edukacji Fizyki

Dagmara Sokołowska

W dniach 13-17 grudnia 2021 r. odbędzie się w trybie zdalnym III Światowa Konferencja Edukacji Fizyki. Pierwotnie miała ona zostać zorganizowana w lipcu 2020 r. w Hanoi w Wietnamie, jednakże ze względu na pandemię COVID-19 najpierw została przeniesiona na lato 2021, a następnie – na obecny termin w trybie zdalnym.



Podczas konferencji można będzie wziąć udział w wykładach plenarnych:

- Yaron Lehari (Izrael), *Crosscutting concepts in science education: Illusions, challenges and hopes*,
- Prathiba Jolly (Indie), *Innovation, Enterprise and Physics Education: Weaving Paradigms for the World of Work*,
- Paula Heron (USA), *Improving student learning in physics: The dual roles of conceptual understanding and reasoning ability*,
- Tomasz Greczyło (Polska), *Video clips in physics education*,
- Alexander P. Mazzolini (Australia), *Lessons learned and unlearned: A lifelong journey with 'Active Learning' as a constant companion*.

Ponadto organizatorzy zapraszają na kilkadziesiąt sesji prezentacji konferencyjnych – krótkich wystąpień ustnych, seminariów, warsztatów i sesji plakatowych.

Szczegółowy program znajduje się pod adresem:

<http://wcpe2020.hnue.edu.vn/Conference/Program/p/tentative-program-adapted-to-new-normal-78>

Organizacja tego dużego przedsięwzięcia w trybie on-line otwiera szereg nowych możliwości. Po pierwsze prezentacje konferencyjne zostaną podzielone na trzy strefy czasowe, tak aby potencjalny uczestnik z dowolnego miejsca na świecie mógł uczestniczyć w konferencji w dogodnym dla siebie czasie. Po drugie – opłata konferencyjna jest jak na warunki konferencji międzynarodowej bardzo niska i waha się, w zależności od daty wniesienia opłaty, w granicach 70-200 euro dla pracowników akademickich i 30-70 euro dla nauczycieli, studentów i doktorantów. Rejestracja trwać będzie do 5 grudnia 2021 r. – szczegóły: <http://wcpe2020.hnue.edu.vn/Registration>

W imieniu GIREP (*International Research Group on Physics Teaching*), jednego z trzech głównych organizatorów WCPE, serdecznie zapraszam do udziału w konferencji polskich dydaktyków fizyki wszystkich szczebli, nauczycieli szkolnych, doktorantów i studentów!