

Metody symulacji w nanotechnologii

Skrócony opis:

Wykład zawiera wprowadzenie do technik stosowanych w fizyce i biofizyce do komputerowego modelowania nanostruktur i symulacji zachodzących w nich procesów. Wybrane przykłady to np. tunelowanie przez bariery potencjału, modelowanie struktury energetycznej układów grafenowych, optymalizacja naprężeń materiałów w skali nano czy dynamika molekularna nanocząstek o znaczeniu biologicznym. Omówienie metod poprzedzone jest krótkim wprowadzeniem do mechaniki kwantowej. Prezentowane metody mają odniesienie do metod numerycznych poznanych na II roku na zajęciach z metod numerycznych. Studenci piszą własne programy realizujące rozwiązanie danego problemu dla nanostruktur lub uczą się posługiwać gotowym pakietem numerycznym.

Pełny opis:

1. Wstęp do mechaniki kwantowej. Równanie Schroedingera, funkcja falowa, przestrzeń modelowa, macierzowa reprezentacja hamiltonianu, zagadnienie własne dla hamiltonianu.
2. Układy modelowe, atomy i molekuly. Wiązanie chemiczne, hybrydyzacja orbitali.
3. Ośrodek periodyczny, twierdzenie Blocha, wektor falowy, pasmowa struktura widma energii, grafen, nanorurki i nanowstęgi.
4. Metoda ciasnego wiązania, metody diagonalizacji.
5. Wprowadzenie do problemu naprężeń na przykładzie samorosnących kropek kwantowych.
6. Rodzaje i źródła odkształceń:
 - Odkształcenia a naprężenia. Opis tensorowy.
 - Przybliżenie ośrodka ciągłego.
 - Podejście atomistyczne.
7. Równanie Schroedingera zależne od czasu:
 - Ewolucja cząstki swobodnej, fale płaskie, gaussowski pakiet falowy.
8. Ruch w polu potencjału ograniczającego.
9. Rozpraszanie na barierze potencjału, tunelowanie, współczynniki odbicia i transmisji.
10. Tunelowanie przez podwójną bariera potencjału, efekty rezonansowe.
11. Metody numeryczne: Cranka-Nicholson i time splitting.
12. Dozwolone energie układu kwantowego: dyskretne, z zakresu ciągłego i rezonansowe (dyskretne rozmyte w widmie ciągłym).
13. Rozwiązanie równania własnego hamiltonianu dla cząstki w jednowymiarowym potencjale studnia-bariera.
14. Własności stanów rezonansowych.
15. Wprowadzenie do nanotechnologii układów biologicznych. Centralny dogmat biologii molekularnej.
16. Przegląd metod poszukiwania minimum funkcji wielu zmiennych na przykładzie pola siłowego.
17. Metody symulacji dynamiki molekularnej.
18. Modelowanie wytrzymałości biomateriałów (włókna).

Wykład 1.
3 godziny

Elementy mechaniki kwantowej

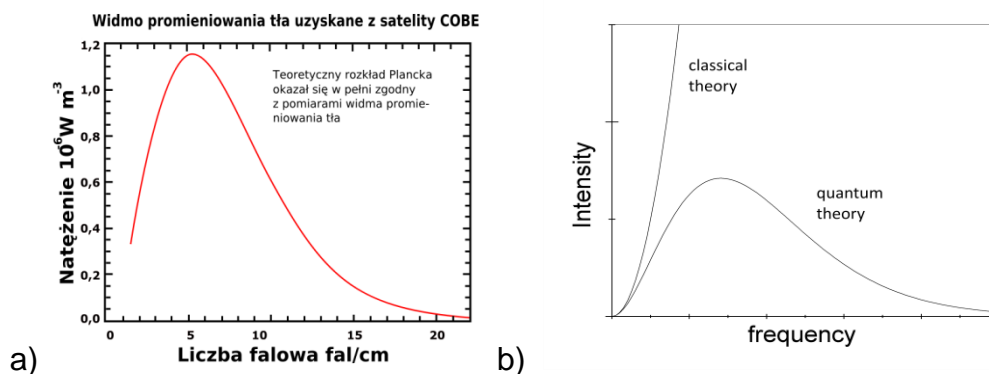
Geneza mechaniki kwantowej

Fizyka klasyczna opisuje układy makroskopowe, przy czym termin „makro” dotyczy układów o rozmiarach powyżej jednego mikrometra. Mogą to być pojedyncze obiekty o makroskopowych rozmiarach i masach (bryły sztywne lub obiekty modelowane punktami materialnymi) albo zbiory takich obiektów (np. układy planetarne). Jednym z zadań teorii fizycznej jest opis *stanu układu fizycznego*. W fizyce klasycznej stan układu złożonego z pewnej liczby cząstek materialnych jest jednoznacznie określony przez podanie położeń i pędów każdego elementu takiego układu w dowolnej chwili czasu. Wielkości te znajduje się rozwiązując równania mechaniki Newtona, które są równaniami różniczkowymi cząstkowymi II-go rzędu. Stan klasycznych układów złożonych z bardzo dużej liczby słabo-oddziałujących cząstek (gazów) opisuje się w ramach fizyki statystycznej, nie wnikając w ich mikroskopową strukturę, za pomocą makroskopowych parametrów takich jak temperatura, objętość i ciśnienie.

Przełom XIX i XX wieku przyniósł rozwój technik eksperymentalnych szczególnie spektroskopowych. Wykonano wówczas szereg doświadczeń, których rezultatów nie dało się wyjaśnić na gruncie fizyki klasycznej bez wprowadzenia nowej teorii opisującej rzeczywistość fizyczną na poziomie „mikro”, tj. pojedynczych atomów i molekuł. Należą do nich *promieniowanie ciała doskonale czarnego*, *efekt fotoelektryczny* oraz *doświadczenie Francka-Hertza*.

Promieniowanie ciała doskonale czarnego.

Ciało doskonale czarne to obiekt fizyczny całkowicie pochłaniający padające nie niego promieniowanie elektromagnetyczne. W ramach fizyki klasycznej zostało w XIX wieku sformułowane prawo promieniowania ciała doskonale czarnego, zwane prawem Rayleigha-Jeansa, według którego ilość wypromieniowanej energii powinna być proporcjonalna do czwartej potęgi częstotliwości promieniowania i to niezależnie od temperatury. Tymczasem doświadczenia wykazywały, że istnieje maksimum natężenia w widmie emitowanym i jest ono funkcją temperatury (prawo Wiena, Rys.1). Przykładowo, dla promieniowania powierzchni Słońca (6000 K) przypada dla długości fali 480 nm. Co więcej, ze sformułowania prawa Rayleigha-Jeansa wynika, że ciało powinno promieniować znacznie więcej energii w zakresie ultrafioletowym niż w zakresie widzialnym. Rozbieżność ta, nazywana *katastrofą w nadfiolecie* była głównym powodem poszukiwania nowej fizyki dotyczącej mikroświata.



Rys.1 a) Widmo promieniowania tła (wszechświata), jako pozostałość po Wielkim Wybuchu. Wikipedia http://pl.wikipedia.org/wiki/Cia%C5%82o_doskonale_czarne . b) Porównanie widma

promieniowania ciała doskonale czarnego (krzywa z maksimum) dobrze opisywanego przez kwantową teorię Plancka z przewidywaniami klasycznymi (krzywa monotonicznie rosnąca).

Źródłem promieniowania ciała są w mechanice klasycznej drgające ładunki elektryczne, tzw. oscylatory (np. elektrony w atomach) mogące emitować promieniowanie o dowolnej częstotliwości a tym samym o dowolnej energii. Żeby wyjaśnić rozbieżności z klasyczną krzywą (zob. Rys.1) Max Planck w 1900 roku przyjął, że takie oscylatory materii mogą emitować promieniowanie tylko w porcjach o energiach $E=h\nu$, gdzie ν – częstotliwość promieniowania ($\nu=c/\lambda$, λ - długość fali, c – prędkość światła), a h jest pewną stałą o wartości $4,135\ 667\ 516 \cdot 10^{-15}$ eVs. Od tego czasu porcje te nazywane są *kwantami*. Zrewolucjonizowało to całkowicie pogląd na strukturę materii, bo oznaczało m.in. to, że struktura energetyczna atomów, z których zbudowana jest cała materia, jest *dyskretna* czyli kwantowa (tzn., że tylko pewne wybrane wartości energii są osiągalne, dostępne).

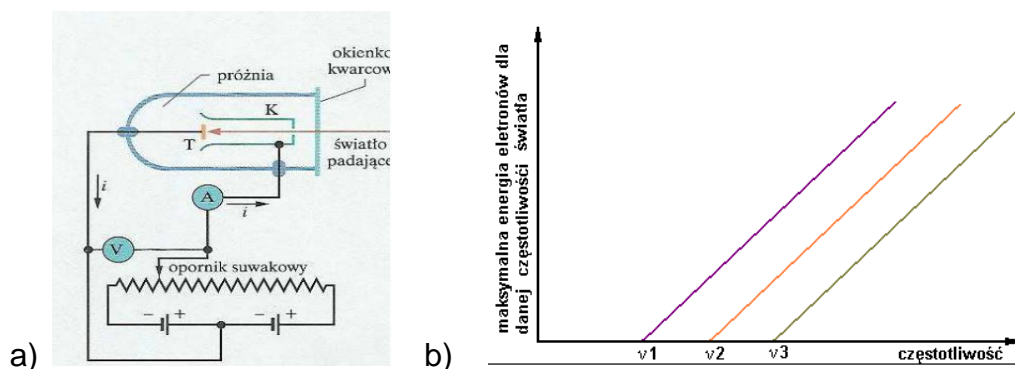
Zjawisko fotoelektryczne.

Efekt fotoelektryczny został odkryty w 1887 r. przez Heinricha Hertza, bardziej szczegółowo badany później przez Philippa von Lenarda a wyjaśniony przez Alberta Einsteina w 1905 roku w oparciu o teorię kwantów zaproponowaną przez Plancka.

W próżniowej lampie wpiętej w obwód elektryczny (Rys. 2a) światło pada na tarczę i wybija z niej elektrony. Przyłożony potencjał może przyspieszać elektrony do kolektora (K) lub hamować (w zależności od polaryzacji przyłożonego potencjału) i zwracać do emitera (T). Energia elektronów zwracanych $E=eVh$ (V – potencjał hamowania) NIE ZALEŻY od natężenia światła (Rys. 2b). Dla dowolnej częstotliwości padającego promieniowania zależność E – energii elektronów (czyli Vh) od częstości ν jest liniowa z progiem. Wniosek Einsteina:

$$h\nu = E + \Phi,$$

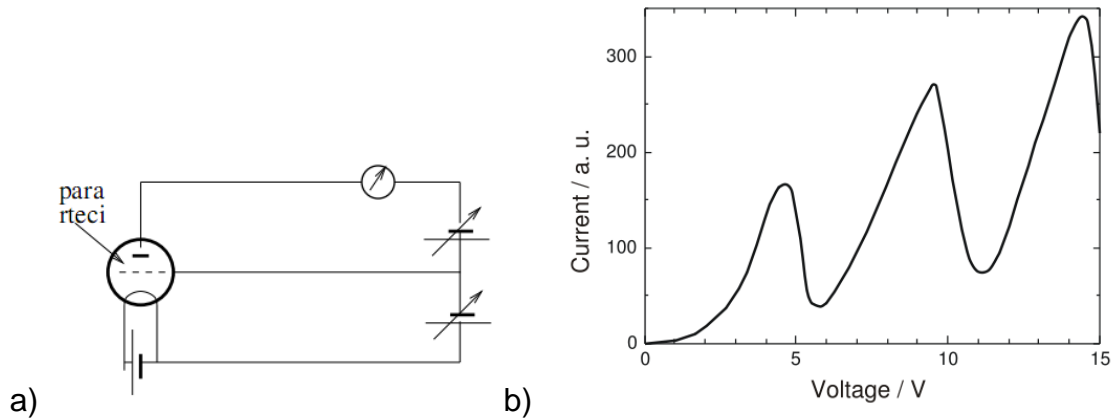
gdzie Φ to praca wyjścia z metalu (energia potrzebna na wybicie z metalu elektronów) Warto zauważyć, że nachylenie prostych energii na Rys. 2b jest stałe i równe dokładnie stałej Plancka.



Rys.2 a) Schemat układu do obserwacji zjawiska fotoelektrycznego. "Podstawy Fizyki", D. Halliday, R. Resnick, and J. Walker, PWN Warszawa 2007. b) Wykresy zależności energii emitowanych elektronów od częstotliwości padającego promieniowania, dla różnych metali, z których zbudowana jest tarcza (T). http://library.thinkquest.org/28383/nowe_teksty/html/2_8.html

Doświadczenie Francka-Hertza.

W doświadczeniu tym elektrony w nieelastycznych zderzeniach z atomami rtęci (Rys. 3) przekazują im energię w porcjach o wartości ok. 4.88 eV – spada wówczas natężenie prądu anodowego. Wniosek: stany energetyczne atomu rtęci są dyskretne; różnica pomiędzy energią stanu podstawowego a energią pierwszego stanu wzbudzonego w atomach rtęci wynosi ok. 4.88 eV.



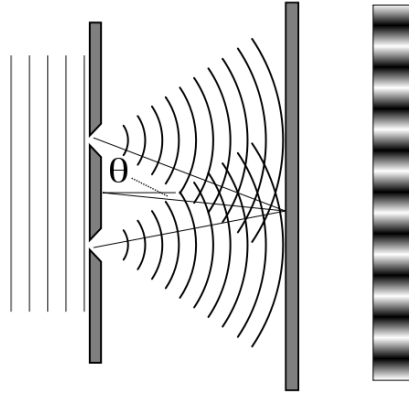
Rys.3 a) Schemat układu do obserwacji doświadczenia Francka-Hertza. b) Wykres prądu anodowego w funkcji napięcia przyspieszającego siatki.

http://en.wikipedia.org/wiki/Franck%E2%80%93Hertz_experiment

Fale materii, dyfrakcja promieni X i dyfrakcja elektronów.

Powyższe doświadczenia pokazują ewidentnie dyskretny czyli kwantowy charakter materii, przynajmniej jeśli mówimy o strukturze energetycznej elementarnych składników materii czyli atomów oraz gdy mamy na myśli wartości energii emitowanego (a także absorbowanego) promieniowania. Inne doświadczenia pokazują inną zaskakującą właściwość materii: jej elementarne składniki, z których jest zbudowana, np. elektrony zachowują się niekiedy jak fale.

Najbardziej charakterystyczną właściwością fal jest ich dyfrakcja (ugięcie na przeszkodach, przesłonach, itp.) oraz interferencja. Typowy obraz przejścia płaskiej fali elektromagnetycznej (światła) przez dwie szczeliny (o szerokości porównywalnej z długością fali), obserwowany na ekranie wygląda tak jak na Rys. 4. W wyniku konstruktywnego i destruktywnego sumowania się grzbietów kulistych fal powstających na szczelinach, na ekranie obserwujemy szereg jasnych i ciemnych prążków. Jeśli szczeliny są punktowe to prążki mają kształt koncentrycznych okręgów. Obraz prążków otrzymuje się też dla siatek dyfrakcyjnych zbudowanych z szeregu bardzo blisko siebie położonych szczelin. Gdy długość fali padającej jest bardzo mała trudno jest mechanicznie wytworzyć siatkę o bardzo gęstych szczelinach, dlatego dla promieni X (Roentgena) jako siatek używa się kryształów, a rolę szczelin (rys) odgrywają w nich położone płaszczyzny krystalograficzne złożone z regularnej sieci atomów.

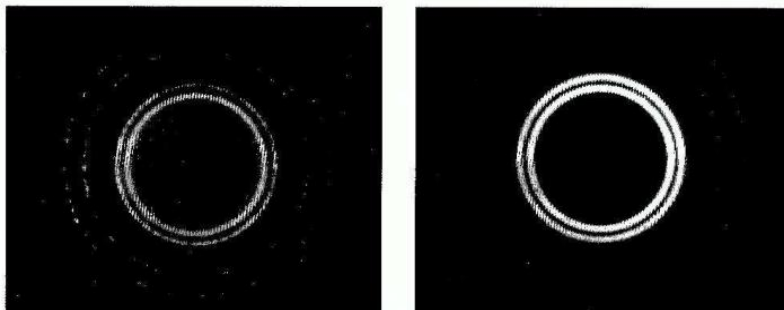
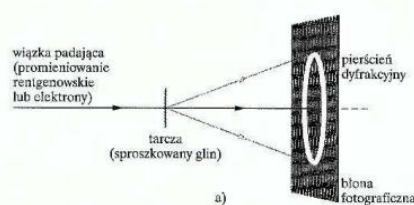


Rys.4 Dyfrakcja fali płaskiej na dwóch szczelinach i powstawanie prążków interferencyjnych na ekranie. http://en.wikipedia.org/wiki/Double-slit_experiment

Zaskakujące jest to, że bardzo podobny obraz otrzymujemy gdy zamiast fal (np. promieni X) przepuścimy przez układ elektrony. Jest to pokazane na Rys. 5.

DYFRAKCJA ELEKTRONÓW I PROMIENI X

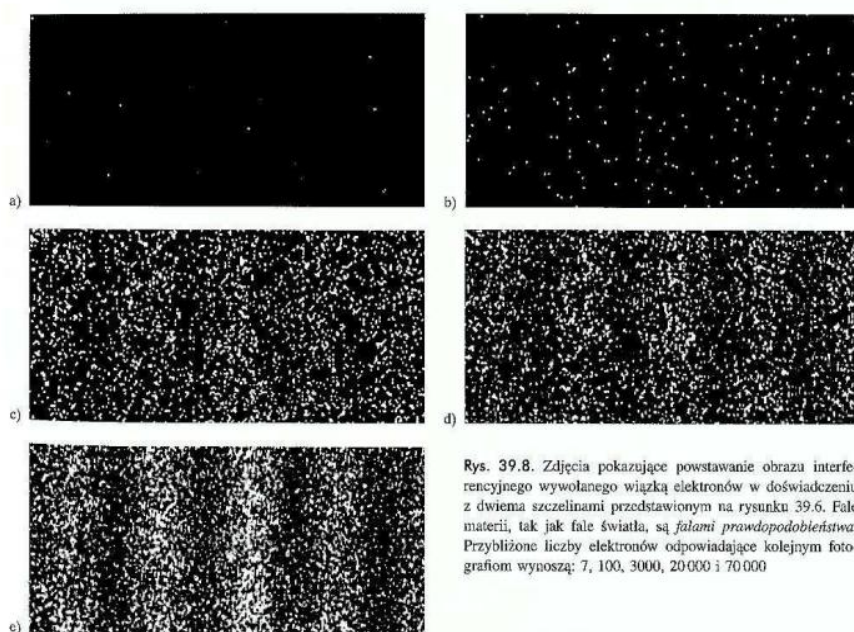
Rys. 39.9. a) Układ doświadczalny wykorzystywany do prezentacji falowego charakteru padającej wiązki metodami dyfrakcyjnymi. Obrazy dyfrakcyjne otrzymane dla b) wiązki promieniowania rentgenowskiego i c) wiązki elektronów (fali materii). Zwróć uwagę, że zasadnicza geometria obu obrazów jest identyczna.



Rys.5 Porównanie dyfrakcji promieni X (fal) i elektronów (cząstek) na kryształach glinu. "Podstawy Fizyki", D. Halliday, R. Resnick, and J. Walker, PWN Warszawa 2007.

W doświadczeniu z elektronami klasyczna fizyka podpowiada nam, że elektrony jako cząstki materialne muszą przejść albo przez jedną albo przez drugą szczelinę. W związku z tym dla gęstego strumienia elektronów padających na podwójną szczelinę, na luminescencyjnym (scyntylicyjnym) ekranie, który rozjaśnionym pikselem monitoruje zderzenie każdego elektronu z ekranem, spodziewamy się dwóch rozjaśnionych obszarów, bezpośrednio za szczelinami. Wynik doświadczenia jest jednak zupełnie inny – Rys.6. Choć każdy z elektronów „gdzieś” padnie, to obraz uzyskany w wyniku bardzo dużej liczby padających elektronów jest obrazem takim samym jak uzyskany dla fal!

FALE MATERII



Rys. 39.8. Zdjęcia pokazujące powstawanie obrazu interferencyjnego wywołanego wiązką elektronów w doświadczeniu z dwiema szczelinami przedstawionym na rysunku 39.6. Fale materii, tak jak fale światła, są *falami prawdopodobieństwa*. Przybliżone liczby elektronów odpowiadające kolejnym fotografiom wynoszą: 7, 100, 3000, 20 000 i 70 000

Rys.6 Obraz interferencyjny otrzymany na ekranie luminescencyjnym, w którym elektrony padają na ekran po przejściu przez podwójną szczelinę. Kolejne fazy odpowiadają coraz większej liczbie elektronów. "Podstawy Fizyki", D. Halliday, R. Resnick, and J. Walker, PWN Warszawa 2007.

Oznacza to, że nie jesteśmy w stanie śledzić trajektorii elektronu, tzn. nie możemy jednocześnie (w tej samej chwili) znać jego położenia i pędu. Nie jest to nasza ułomność, a właściwość materii, właściwość świata w którym żyjemy i jawi się on nam jako niedeterministyczny: znając stan cząstki w danej chwili czasu, jej stany w chwilach następnych możemy przewidywać tylko z pewnym prawdopodobieństwem. W mechanice kwantowej ta właściwość odzwierciedlona jest przez tzw. *zasadę nieoznaczoności* Heisenberga. Jasne i ciemne prążki na Rys. 6. odzwierciedlają fakt, że tylko z pewnym prawdopodobieństwem jesteśmy w stanie określić którądy (przez którą szczelinę) przeszedł dany elektron. Nowa fizyka, fizyka kwantowa musi mieć zatem charakter *probabilistyczny*.

Wybrane, podstawowe postulaty mechaniki kwantowej

Postulat1.

Stan układu fizyczny w mechanice kwantowej (układu kwantowego) opisywany jest w pełni poprzez pewną funkcję zwaną funkcją falowa albo funkcją stanu $\Psi(x, t)$, gdzie x oznacz zespół współrzędnych położeniowych wszystkich cząstek należących do układu, a t oznacza czas. Interpretację fizyczną posiada kwadrat modułu funkcji falowej. Wyrażenie

$$\Psi^* \Psi dx = |\Psi|^2 dx$$

jest prawdopodobieństwem znalezienia układu w objętości dx .

Zauważmy, że całka po całej przestrzeni z $|\Psi|^2$ musi być równa „1”, gdyż prawdopodobieństwo znalezienia układu fizycznego „gdziekolwiek” jest równe pewności. Tym samym, możliwe funkcje

falowe układów fizycznych muszą być, mówimy, „kwadratowo całkowalne” – całki z kwadratów ich modułów nie mogą być rozbieżne.

Postulat2.

Każdej mierzonej (mierzalnej) wielkości fizycznej A danego układu fizycznego (której mierzona wartość może zależeć od x oraz od t), odpowiada pewien operator \mathbf{A} . Operatory działają na funkcje. W wyniku działania $\mathbf{A}\Psi=\Phi$ otrzymujemy również funkcję (inną, w szczególnych przypadkach tę samą). Przykładowym operatorem może być branie pochodnej z funkcji po jakiejś zmiennej, lub mnożenie funkcji przez jakąś liczbę. Jeśli $\mathbf{A}\Psi=a\Psi$, (a – liczba) to Ψ jest tzw. funkcją własną operatora \mathbf{A} .

Postulat3.

Wynikami pomiaru wielkości fizycznej A mogą być tylko wartości własne odpowiadającego jej operatora \mathbf{A} .

Uwaga.

Wszystkie funkcje własne dowolnego operatora (nazwijmy je φ_i , oraz ponumerujmy $i=1,2,\dots$) tworzą bazę (funkcyjną) w przestrzeni funkcji kwadratowo całkowalnych (danego typu). Odpowiadające im wartości własne to a_i . Zatem, dowolną funkcję opisującą stan układu fizycznego można przedstawić w postaci kombinacji liniowej (rozwińcie w bazie)

$$\Psi = \sum_i c_i \varphi_i$$

Zauważmy analogię z przestrzenią wektorową. Podobnie jak w przestrzeni wektorowej definiuje się iloczyn skalarny dwóch funkcji φ_i oraz φ_j jako

$$\langle \varphi_i | \varphi_j \rangle = \int \varphi_i^* \varphi_j d\tau,$$

gdzie τ (wcześniej oznaczane także przez x) to zespół wszystkich współrzędnych przestrzennych. Symbol $\langle | \rangle$ jest tzw. notacją Diraca.

Funkcję nazywamy unormowaną jeśli $\langle \varphi_i | \varphi_i \rangle \geq 1$, a dwie funkcje są ortogonalne jeśli ich iloczyn skalarny znika.

W iloczynie skalarnym pod całką występują dwie funkcje. Jle jedna z nich może być wynikiem działania jakiegoś operatora na inną funkcję:

$$\langle \varphi | \Phi \rangle = \langle \varphi | H\psi \rangle = \langle \varphi | H | \psi \rangle.$$

Postulat4.

Jeśli układ znajduje się w stanie opisywanym funkcją falową Ψ to prawdopodobieństwo, że w wyniku pomiaru wielkości fizycznej A otrzymamy wielkość a_i dane jest przez:

$$P_i = |\langle \varphi_i | \Psi \rangle|^2 = \left| \int \varphi_i^* \Psi d\tau \right|^2 = |c_i|^2.$$

Dodatkowo:

A) Jeśli układ fizyczny znajduje się w stanie opisywanym funkcją falową Ψ , to wartości średniej (uzyskanej z wielu pomiarów) wielkości fizycznej A (dla układu ciągle w tym samym stanie) odpowiada tzw. wartość oczekiwana odpowiadającego jej operatora A :

$$\bar{A} = \langle \Psi | A \Psi \rangle = \int \Psi^* A \Psi d\tau.$$

B) Pęd to wektorowa wielkość fizyczna $\vec{p} = (p_x, p_y, p_z)$

operator pędu też jest wielkością wektorową, składa się z trzech składowych operatorów

np. $p_x = -i\hbar \frac{\partial}{\partial x}.$

Tym samym operator energii kinetycznej (przez analogię klasyczną)

$$T = \frac{mv^2}{2} = \frac{p^2}{2m} = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \quad (1)$$

Postulat5.

Funkcja falowa spełnia równanie Schrödingera

a) zależne od czasu t

$$i\hbar \frac{\partial \Psi(\mathbf{x}, t)}{\partial t} = H \Psi(\mathbf{x}, t),$$

gdzie dla układu złożonego z n cząstek $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ a każde $\mathbf{x}_i = (x_i, y_i, z_i)$, oraz

b) niezależne od czasu,

$$H \Psi(\mathbf{x}, t) = E \Psi(\mathbf{x}, t) \quad (2)$$

gdzie H – operator energii całkowitej układu (operator Hamiltona, w skrócie hamiltonian). $H = T + V$, a V to operator energii potencjalnej.

Jeśli V – energia potencjalna układu, nie zmienia się w czasie, tzn. $V=V(x)$, to stany własne H czyli stany układu fizycznego to stany stacjonarne i $\Psi(\mathbf{x}, t) = \Psi(\mathbf{x})$.

Uwaga, równanie Schrödingera (2) jest *de facto* równaniem różniczkowym. Jego rozwiązanie wymaga nałożenia na Ψ warunków brzegowych, to one powodują, że rozwiązaniem równania (2) jest na ogół dyskretny zbiór wartości i funkcji własnych

$$E_i, \Psi_i.$$

Rozwiązanie równania Schroedingera (2) jest podstawowym celem każdego kwantowego problemu fizycznego. Przez jego rozwiązanie znajdujemy bowiem funkcje falowe reprezentujące możliwe stany danego układu, a tym samym otrzymujemy pełną informację o danym układzie fizycznym.

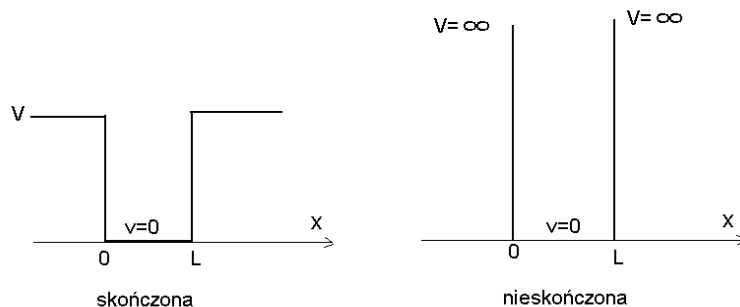
Rozwiązanie równania zależnego od czasu daje nam dodatkowo wgląd w dynamikę układu fizycznego, czyli daje nam możliwość poznania (i przewidzenia) jego zachowania się w czasie.

Najprostsze przykłady kwantowych układów fizycznych

Istnieje tylko kilka modelowych układów fizycznych, które dają się ściśle rozwiązać. Przedstawimy tu niektóre z nich. We wszystkich innych przypadkach reprezentujących złożone układy fizyczne równanie (2) musimy rozwiązywać w sposób przybliżony.

Cząstka w jednowymiarowej studni potencjału.

Wyobraźmy sobie cząstkę o masie m , która swobodnie może poruszać się tylko w jednym kierunku (w jednym wymiarze) np. x , na odcinku o długości L . Jeśli niezależnie od energii cząstki nie może ona opuścić odcinka L , to w mechanice kwantowej jej energia potencjalna reprezentowana jest przez dwie nieskończenie wysokie bariery umieszczone na końcach odcinka L . Jeśli cząstka mogłaby się wydostać z obszaru L przy energii równej np. V , to wówczas mówimy o studni potencjału o szerokości L i barierach potencjału o skończonej wysokości V . Oba przypadki reprezentowane są na Rys. 7.



Rys.7 Skończona (lewa) i nieskończona (prawa) jednowymiarowa studnia potencjału o szerokości L .

W obu przypadkach równanie Schroedingera (2) jest równaniem zależnym tylko od jednej zmiennej x . Jako równanie różniczkowe wymaga podania dwóch warunków brzegowych w celu uzyskania jego jednoznacznego rozwiązania. W przypadku nieskończonej studni potencjału takimi warunkami są znikanie funkcji falowej $\Psi(x)$ dla $x=0$ oraz $x=L$, $\Psi(0) = \Psi(L) = 0$, czyli na brzegach i poza granicami studni (cząstka nie może się wydostać ze studni, tzn. nie może istnieć poza studnią, a zatem prawdopodobieństwo znalezienia jej poza odcinkiem L jest równe zero). Warunki te powodują, że (2) ma tylko takie możliwe energie własne:

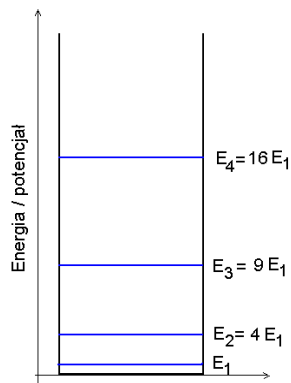
$$E_n = \frac{h^2 n^2}{8mL^2}, \quad (3)$$

gdzie n to liczba kwantowa numerująca kolejne rozwiązania, a odpowiadające im funkcje falowe

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi}{L}x\right), \quad n = 1, 2, 3, \dots$$

Schematycznie przedstawia to Rys. 8. Warto przyjrzeć się efektowi skali. Dla dużych, makroskopowych rozmiarów studni L oraz makroskopowych cząstek (o dużej masie) wzór (3) pokazuje, że możliwe poziomy energetyczne są bardzo gęsto położone na skali energii. Dla $L=1\text{m}$ oraz $m=1\text{kg}$, różnica E_2-E_1 wynosi ok. 10^{-66} Dżuła (10^{-47} eV) i jest absolutnie niemierzalna (za mała!), podczas gdy w warunkach mikroświata, tzn. np. dla elektronu w pułapce o rozmiarach 1 nm , ta różnica energii jest rzędu 1 eV i jest mierzalna. Ten przykład podpowiada nam, że otaczająca nas

makroskopowa rzeczywistość też poddaje się opisowi mechaniki kwantowej, jednak efekty kwantowe są tu na ogół znikome (niemierzalne).



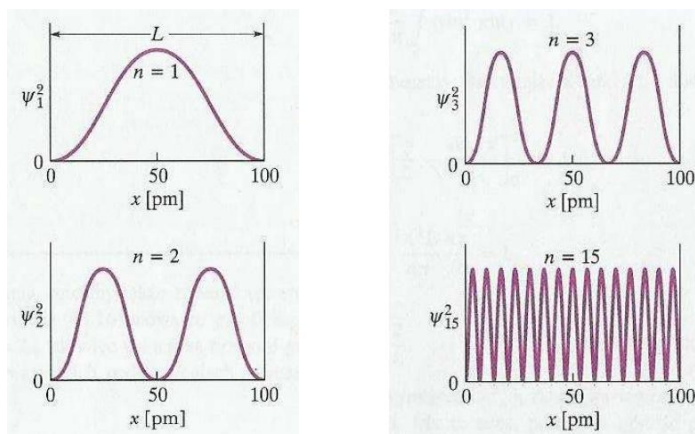
Rys.8 Schemat poziomów energetycznych cząstki w jednowymiarowej, nieskończonej studni potencjału.

Zwróćmy uwagę, że przejścia kwantowe pomiędzy poziomami w tym układzie, czyli wzbudzenia cząstki do wyższej energii np. za pomocą światła (fotonów) możliwe są tylko dla fotonów o energiach równych

$$h\nu = E_n - E_m$$

Proces taki nazywa się absorpcją. Proces odwrotny (deekscytacje) będzie źródłem fotonów (emisja). Oba widma (przez widmo rozumiemy zależność intensywności światła absorbowanego lub emitowanego od energii padających fotonów) mieć zatem będą dyskretny charakter (patrz doświadczenie Francka-Hertza, Rys. 3b).

Kwadraty modułów Ψ_n (gęstości prawdopodobieństwa) pokazane są na Rys. 9. Te prawdopodobieństwa przypominają „fale”.



Rys.9 Schemat poziomów energetycznych cząstki w jednowymiarowej, nieskończonej studni potencjału. „Podstawy Fizyki”, D. Halliday, R. Resnick, and J. Walker, PWN Warszawa 2007.

Symetria.

Układ fizyczny może być niezmienniczy ze względu na pewne operacje symetrii. Np. molekula dwuatomowa homojądrowa nie zmienia się jeśli poddamy ją obrotom wokół osi wiązania lub

odbiciom w płaszczyźnie prostopadłej do osi wiązania umieszczonej w połowie wiązania. Nieskończony kryształ będący obiektem periodycznym jest niezmienniczy względem symetrii *translacji*, tzn. przesunięcia o wektor sieci. Jeśli układ jest niezmienniczy ze względu na jakież operacje symetrii, to opisujący go hamiltonian \mathbf{H} też jest niezmienniczy a każdej operacji symetrii odpowiada pewien operator \mathbf{S} . Niezmienniczość \mathbf{H} względem operacji symetrii prowadzi do tego, że operatory, te komutują: $[\mathbf{H}, \mathbf{S}] = \mathbf{0}$, tzn. kolejność ich działania na funkcje falowe $\mathbf{H}\mathbf{S}\psi$ lub $\mathbf{S}\mathbf{H}\psi$ nie ma znaczenia – zawsze otrzymamy ten sam wynik. Łatwo pokazać, że takie dwa operatory mają wspólne funkcje własne, a zatem funkcje falowe opisujące stany układu będą musiały: (a) odzwierciedlać symetrie związane z operatorami \mathbf{S} , (b) być indeksowane dodatkowymi liczbami kwantowymi związanymi z wartościami własnymi \mathbf{S} .

Wykład 2.
3 godziny

Atomy, molekuly i elementy teorii ciała stałego

Atom wodoru.

Atom wodoru to najprostszy układ złożony z dwóch oddziałujących elektrostatycznie cząstek: protonu i elektronu. Ponieważ proton jest ok. 1000 razy cięższy od elektronu, możemy spokojnie założyć, że w obliczu ruchu elektronu jest on nieruchomy (bardzo, bardzo powolny) i przyjąć początek układu współrzędnych na nieruchomym protonie. Wówczas wodór jawi się nam jako ujemnie naładowany elektron poruszający się w kulombowskim polu pochodzącym od dodatnio naładowanego protonu. Zauważmy, że w klasycznej fizyce elektron poruszający się wokół jądra (tzn. z przyspieszeniem) emitowałby, czyli tracił energię i po prostu spadłby na proton; atom wodoru przestałby istnieć...

Energia potencjalna elektronu w kulombowskim polu elektrostatycznym protonu wynosi

$$V(r) = -\frac{e}{r} \quad (4)$$

gdzie e to ładunek elektronu a r to odległość elektronu od jądra (protonu). Zauważmy, że energia ta jest ujemna (bo cząstki się przyciągają) i może ona równoważyć kinetyczną energię elektronu (która jest zawsze dodatnia). Hamiltonian składa się z operatora energii kinetycznej \mathbf{T} , danego wyrażeniem (1) oraz energii potencjalnej (powyżej). Rozwiązanie równania

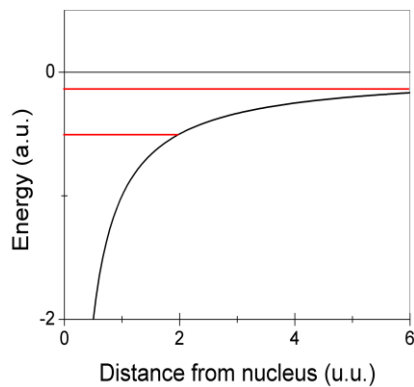
$$\mathbf{H}\Psi(\vec{r}, t) = E\Psi(\vec{r}, t),$$

(gdzie w konkretnym przypadku jednego elektronu użyto tu wektora położenia elektronu $\vec{r} = (x, y, z)$ w miejsce poprzednio używanych oznaczeń \mathbf{x} lub τ) znów daje dyskretne *widmo* energii własnych, czyli możliwych energii jakie może posiadać elektron w atomie wodoru. Wynoszą one:

$$E_n = -\frac{me^4}{8h^2\epsilon_0^2} \frac{1}{n^2}, \quad (5a)$$

gdzie n to znów liczba kwantowa numerująca kolejne poziomy energetyczne atomu wodoru. Energie te są ujemne, co oznacza, że elektron jest w atomie wodoru „związany”. Żeby go uwolnić, czyli zjonizować atom - trzeba dostarczyć dodatniej energii równoważącej energię wiązania. Dla małej masy elektronu (w kg jest to wielkość rzędu 10^{-30}) różnice pomiędzy poziomami energetycznymi są już całkiem spore i dobrze mierzalne. Energia stanu podstawowego, E_1 , w wodorze to -13,6 eV.

Zauważmy, że jest to również problem cząstki w studni potencjału, z tym, że teraz studnia (4) jest trójwymiarowa, o sferycznej symetrii, ma kształt kulombowskiego „lejka” i jest nieskończenie głęboka. Wraz z poziomami energetycznymi zobrazowana jest ona schematycznie na Rys. 10.

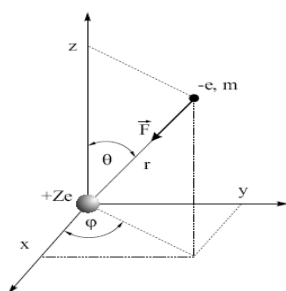


Rys.10 Schematy energii potencjalnej (czarna krzywa) i poziomów energetycznych (czerwona) elektronu w atomie wodoru w jednostkach atomowych ($m=h=e=1$).

Kiedy studnia kwantowa jest więcej niż jednowymiarowa, występuje wówczas na ogół degeneracja poziomów energetycznych. Z degeneracją mamy do czynienia wtedy, gdy jednej energii odpowiada więcej różnych funkcji (stanów kwantowych). Mamy z nią do czynienia również w przypadku atomu wodoru. Najniższy poziom energetyczny jest niezdegenerowany. Drugi poziom jest zdegenerowany czterokrotnie, a trzeci aż dziewięć-krotnie. Występowanie degeneracji poziomów energetycznych wiąże się na ogół z istnieniem jakiejś symetrii, której podlega dany układ. W przypadku wodoru mamy do czynienia z bardzo wysoką symetrią. Atom wodoru (i tym samym opisujący go hamiltonian) jest niezmienniczy ze względu na dowolne obroty wokół dowolnych osi przechodzących przez proton. Jest to tzw. symetria sferyczna, a wynikająca z niej degeneracja nazywa się *orbitalną*. Tę degenerację „numeruje” się tzw. orbitalnymi (l) i magnetycznymi (m) liczbami kwantowymi. Dla poziomu energetycznego n , nazywanego także *powłoką*, mamy n różnych orbitalnych typów funkcji falowych (tzw. orbitali), każdy z nich dodatkowo zdegenerowany $1x, 3x, 5x, \dots$ Typy te opisujemy literkami s, p, d, \dots i nazywamy *podpowłokami*. Funkcje falowe zależą od x, y, z albo przechodząc do sferycznego układu współrzędnych od r i dwóch kątów θ i ϕ . W sferycznym układzie współrzędnych (Rys. 10a) funkcje falowe wyglądają tak

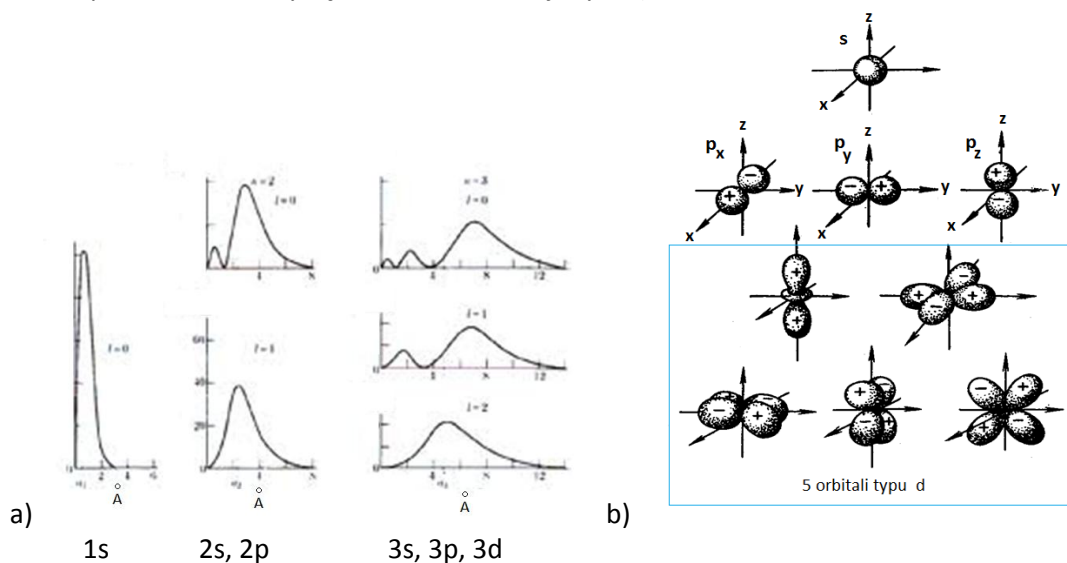
$$\Psi(x, y, z) = \Psi(r, \theta, \varphi) = R_n(r)Y_{l,m}(\theta, \varphi) = \Psi_{n,l,m}, \quad (5b)$$

gdzie $Y_{l,m}$ to tzw. funkcje kuliste, n nazywa się główną liczbą kwantową, l – liczbą orbitalną, a m magnetyczną liczbą kwantową. Energie w wodorze zależą tylko od n (przy nieobecności zewnętrznych pól).



Rys.10b Sferyczny układ współrzędnych

Zależność radialna (od r) funkcji falowych w każdym typie n jest taka sama i charakteryzuje się pewną liczbą węzłów (tzn. takich odległości od jądra dla których funkcja całkowicie się zeruje). Moduły kwadratów tych funkcji scałkowane po kątach dają tzw. radialne rozkłady gęstości elektronowej. Maksima tych gęstości pokazują te odległości od jądra, na których prawdopodobieństwo znalezienia elektronu jest największe. Można je nazwać najbardziej prawdopodobnymi orbitami, po których poruszałby się klasyczny elektron. Gęstości te są pokazane na Rys. 11a. Powierzchnie, na których całkowita gęstość ładunku ma stałą wartość nazywają się „warstwicami” Na Rys. 11b pokazane są kątowe „kształty” tych warstwic dla różnych orbitali, ukazujące rozkład gęstości elektronowej w poszczególnych stanach. (Uwaga: dla danego n ze względu na degenerację orbitalną każda kombinacja funkcji o różnym l i m też jest dobrą funkcją falową, zatem pokazane kształty są „możliwe” a nie „jedyne”).



Rys.11 Schematyczne radialne (a) i kątowe (b) rozkłady gęstości ładunku dla różnych orbitali wodorowych. Gęstości te można znaleźć w dowolnej książka z chemii kwantowej lub mechaniki kwantowej.

W atomach wieloelektronowych ze względu na kulombowskie oddziaływanie między elektronami, degeneracja orbitalna zostaje zniesiona (zaburzenie symetrii). Energie poziomów energetycznych zależą ciągle od n i teraz od l . Nie zależą jednak o magnetycznej liczby kwantowej m : każdy poziom o określonym (n, l) jest $2l+1$ krotnie zdegenerowany. Kolejność energetyczna poziomów jest następująca: 1s, 2s, 2p, 3s, 3p, 3d,, ale dla wyższych n odstępstwa spowodowane oddziaływaniem między elektronami powodują pewne zmiany. Każdy orbital atomowy może być ze

względu na zasadę Pauliego obsadzony tylko przez maksymalnie dwa elektrony (znajdujące się w różnych tzw. stanach spinowych). Jeśli kilka podpowłok elektronowych jest w pełni wypełnionych a inne są nieobsadzone, to taki atom nazywamy zamkniętopowłokowym. Niechętnie wchodzi on wówczas w reakcje z innymi atomami. Takie są np. atomy gazów szlachetnych. W przypadku atomów otwartopowłokowych „chętnie oddają” one lub „przyjmują” elektrony (tworząc ujemne lub dodatnie jony) dążąc do zamykania powłok elektronowych, gdyż takie układy są najbardziej korzystne energetycznie. Takie atomy łatwo wchodzi w reakcje z innymi tworząc molekuly i kryształy.

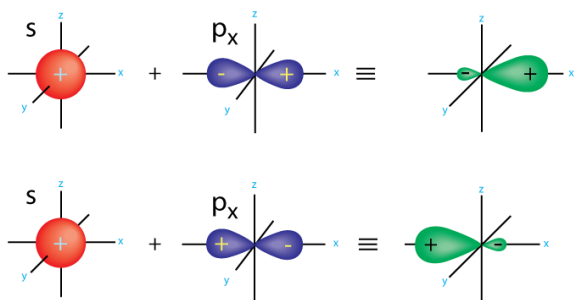
Orbitale zhybrydizowane i wiązania chemiczne

Wiązania chemiczne.

Atomy, choć elektrycznie obojętne mogą łączyć się i tworzyć molekuly oraz kryształy. Połączenie dwóch sąsiednich atomów nazywa się *wiązaniem chemicznym*. Istnieją różne typy wiązań. Gdy jednemu atomowi brakuje elektronu do zamknięcia zewnętrznej powłoki lub podpowłoki, np. Cl ($1s^2 2s^2 2p^2 3s^2 3p^5$) a inny posiada jeden elektron na zewnętrznej powłoce, np. Na ($1s^2 2s^2 2p^2 3s^1$) to sód oddaje jeden elektron a chlor go przyjmie. Powstaną wówczas dwa jony (każdy z nich zamkniętopowłokowy) naładowane jeden dodatnio drugi ujemnie, które będą się przyciągać elektrostatycznie. Powstanie wówczas wiązanie jonowe. Inny typ wiązań występuje w metalach: prawie swobodne elektrony tworzą ujemnie naładowaną chmurę, która utrzymuje kulombowsko „w ryzach” sieć dodatnio naładowanych (i odpychających się) jonów. Wiązanie *molekularne* albo van der Waals występuje pomiędzy atomami gazów szlachetnych lub pomiędzy dużymi molekulami (np. białka) ale także pomiędzy warstwami grafenu w graficie. W interesujących nas strukturach grafenowych, wiązania są *kowalencyjne*. Powstają przez nakładanie się orbitali zhybrydizowanych z sąsiednich atomów.

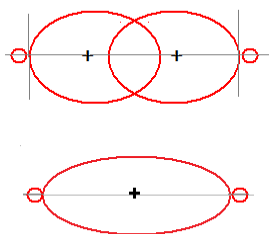
Orbitale zhybrydizowane.

Jak wspomniano wcześniej, w przypadku degeneracji poziomów energetycznych w atomie, z funkcji falowych (orbitali) należących do tej samej wartości energii można tworzyć kombinacje liniowe; każda z nich jest tak samo dobrą funkcją falową, orbitalem o innym przestrzennym rozkładzie gęstości elektronowej. W przybliżeniu, takie kombinacje można też tworzyć dla orbitali odpowiadających różnym, ale blisko energetycznie położonym poziomom. Taka sytuacja ma miejsce np. w węglu C ($1s^2 2s^2 2p^4$), gdzie energie poziomów 2s i 2p są bardzo bliskie (przynajmniej w porównaniu z odległością energetyczną do 1s). Z orbitali 2s i $2p_x$ można utworzyć dwa nowe orbitale tzw. zhybrydizowane, sp. Jeden jako sumę 2s i $2p_x$ a drugi jako ich różnicę. Ponieważ orbital $2p_x$ ma gęstość zlokalizowaną osiowo wzdłuż osi OX, ale dla $x > 0$ jest dodatni a dla $x < 0$ jest ujemny, powstałe orbitale zhybrydizowane będą „kierunkowe”, jeden z nich będzie miał „więcej” gęstości elektronowej dla $x > 0$ a drugi wręcz przeciwnie. Ilustruje to Rys. 12.



Rys.12. Tworzenie orbitali hybrydowanych typu sp . <http://www.grandinetti.org/comment/18>

Takie orbitale wyjaśniają trwałość wiązania w cząsteczce F_2 , co pokazuje Rys. 13. Nakładające się w obszarze między jądrami fluoru gęstości (chmury ładunku) elektronowej pochodzące od orbitali sp z obu atomów, równoważą (czyli wiążą) kulombowsko odpychające się dodatnie jony fluoru F^{1+} . Z połączenia (zsumowania) takich dwóch orbitali sp od różnych atomów zbudowany jest tzw. molekularny orbital wiążący (w tym przypadku tzw. orbital typu σ).

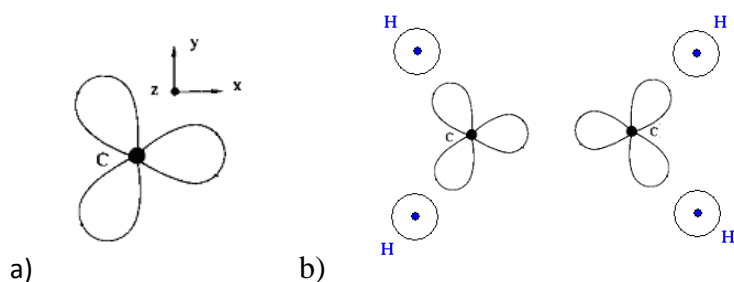


Rys.13. Tworzenie wiązania kowalencyjnego w molekułe F_2 . Suma dwóch orbitali zhybrydowanych sp daje wiążący orbital typu σ .

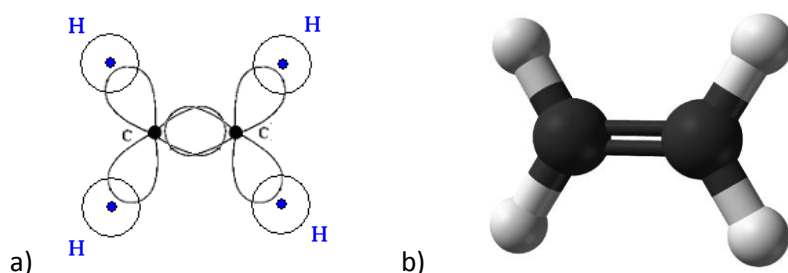
Hybrydyzacja sp^2 prowadzi do wyjaśnienia trwałości wiązań np. w benzenie, grafenie (i częściowo w etylenie C_2H_4). Powstaje wówczas gdy z jednego orbitala typu s i dwóch typu p (np. p_x i p_y) tworzy się trzy nowe kombinacje, a orbital p_z pozostawia bez zmian.

$$\begin{aligned}\phi_1 &= \frac{1}{\sqrt{3}}(s + \sqrt{2}p_x) \\ \phi_2 &= \frac{1}{\sqrt{6}}(\sqrt{2}s - p_x + \sqrt{3}p_y) \\ \phi_3 &= \frac{1}{\sqrt{6}}(\sqrt{2}s + p_x - \sqrt{3}p_y) \\ \phi_4 &= p_z\end{aligned}$$

Orbitale sp^2 też są kierunkowe i „wskazują” naroża trójkąta równobocznego. Ilustruje to Rys. 14. Natomiast Rys. 15 przedstawia przykład tworzenia wiązania kowalencyjnego w molekułe C_2H_4 .



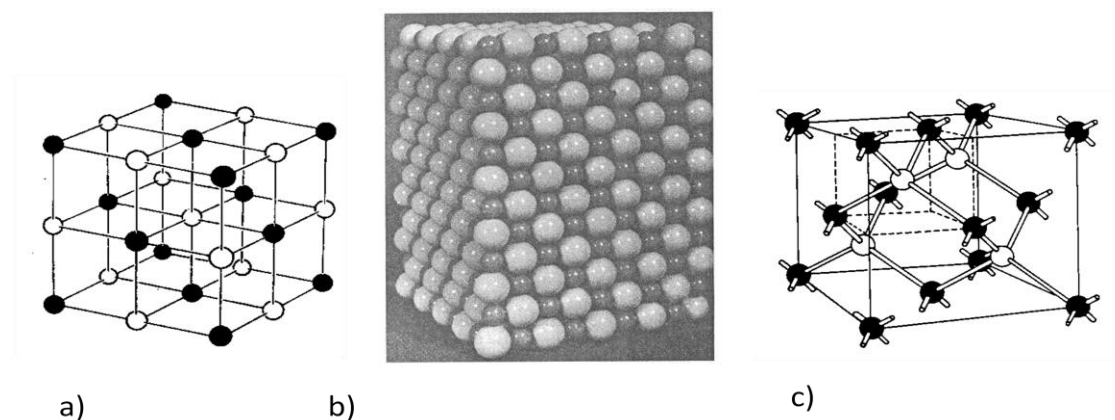
Rys.14. a) Schemat orbitali zhybrydyzowanych sp^2 . b) Orbitale $1s$ wodoru i sp^2 węgla przed utworzeniem wiązania w molekułe C_2H_4 .



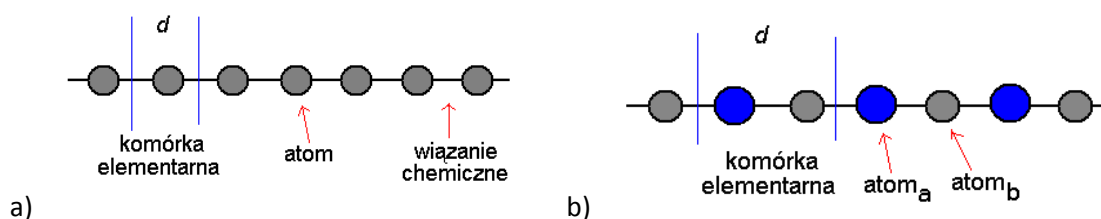
Rys.15. a) Tworzenie kowalencyjnego wiązania molekularnego w cząsteczce C_2H_4 . b) Model cząsteczki C_2H_4 . W rzeczywistości wiązanie pomiędzy węglami nie jest typu σ , ale tworzy się wiązanie typu π z dwóch orbitali hybrydyzowanych sp^3 węgla. <http://en.wikipedia.org/wiki/File:Ethylene-CRC-MW-3D-balls.png>

Potencjały periodyczne

Kryształy są przykładami układów fizycznych, w których mamy do czynienia z periodycznie powtarzającą się strukturą przestrzenną. Tę strukturę tworzą atomy połączone wzajemnie za pomocą wiązań chemicznych. Każdy kryształ można otrzymać poprzez powielanie tzw. *komórki elementarnej* zawierającej jeden lub więcej atomów tworzących strukturę krystaliczną sieci. Rys. 16 ukazuje przykłady sieci trójwymiarowych, a model sieci jednowymiarowej (łańcucha) pokazany jest na Rys. 17. W obu przypadkach kule obrazują pojedyncze atomy (różne oznaczone różnymi kolorami) a kreski symbolizują wiązania chemiczne między nimi.



Rys.16. Modele sieci krystalicznych, a) i b) NaCl, c) GaAs. "Wstęp do fizyki ciała stałego", C.W. Kittel, PWN, Warszawa 1999.



Rys.17. Model sieci jednowymiarowej. a) złożonej z takich samych atomów, b) złożonej z dwóch różnych atomów w komórce elementarnej. Taki przypadek nazywa się „siecią z bazą” w komórce elementarnej występuje więcej niż jeden atom.

Makroskopowy kryształ jest bardzo złożonym układem fizycznym. Zawiera miliony atomów, elektronów i protonów, silnie ze sobą oddziałujących. Ścisły kwantowo-mechaniczny opis takiego układu jest niemożliwy nie tylko ze względu na trudności obliczeniowe, ale także interpretacyjne (trudno interpretować własności funkcji falowej będącej funkcją tak wielu zmiennych). Dlatego też dokonuje się szeregu przybliżeń. Pierwszym z nich jest rozdzielenie ruchu jąder i elektronów (wspominaliśmy już o tym omawiając krótko atom wodoru). Następnym krokiem jest redukcja ciągle bardzo złożonego hamiltonianu opisującego stan bardzo wielu elektronów w kryształ (kryształ to bardzo duża molekula) do hamiltonianu jednoelektronowego. Nazywa się to *przybliżeniem jednoelektronowym*. Zakłada ono (rozsądnie), że możemy opisywać stany pojedynczych elektronów w molekuale lub kryształ (tak jak to ma miejsce np. w atomie wodoru) a funkcję falową całego układu zbuduje się w odpowiedni sposób z funkcji jednoelektronowych. Te jednoelektronowe funkcje nazywają się *orbitalami atomowymi*, *orbitalami molekularnymi* lub *orbitalami krystalicznymi*, odpowiednio w przypadku atomu, molekuly lub kryształu. Orbitale są rozwiązaniami jednoelektronowych równań typu (2) z tym, że teraz każdy z elektronów oddziałuje kulombowsko z periodyczną siecią dodatnio naładowanych jonów oraz z wszystkimi pozostałymi elektronami. W kolejnym przybliżeniu zarówno energię potencjalnego oddziaływania elektronu z jonami jak i z pozostałymi elektronami (które przecież nie tworzą sztywnej, periodycznej sieci) przybliża się periodycznym wyrażeniem. Np. dla jednowymiarowego kryształu takiego jak na Rys. 17 $V(x+d) = V(x)$, gdzie d - okres periodyczności, równy w tym przypadku rozmiarowi komórki elementarnej.

Twierdzenie Blocha i pasmowa struktura energii

Nietrudno się domyśleć, że skoro kryształ jest periodyczny to w zasadzie cała informacja o jego właściwościach będzie zgromadzona w strukturze pojedynczej komórki elementarnej z dodatkowo nałożonym warunkiem periodyczności. Ten warunek periodyczności to nic innego jak *symetria translacji*. Translacja to nic innego jak przesunięcie obiektu o pewien wektor. W przypadku kryształu jest to tzw. wektor sieci prostej, który w jednowymiarowym przypadku sprowadza się do odcinka d (Rys. 17) , który w ogólności nazywa się „stałą sieci”. Symetria translacji mówi, że po przesunięciu obiektu o wektor translacji dany obiekt „przechodzi w siebie” – w zasadzie się nie zmienia. Oczywiście może ona dotyczyć tylko układów nieskończonych, jednak dla układów zawierających bardzo dużą liczbę komórek elementarnych, przesunięcie kryształu o jedną komórkę przeprowadzi go w zasadzie „w siebie”, z bardzo drobnymi zmianami na brzegach kryształu. Jeśli interesują nas właściwości kryształu „zapisane” w jego strukturze krystalicznej a nie własności brzegowe, to z bardzo dobrym przybliżeniem możemy bardzo duży (makroskopowy) kryształ traktować jak kryształ nieskończony. To przybliżenie będzie miało pewne konsekwencje na możliwe wartości tzw. wektora falowego, ale o tym będzie mowa później. Tymczasem zajmijmy się najważniejszym twierdzeniem dotyczącym układów periodycznych. Jest nim twierdzenie Blocha. Mówi ono, że funkcja falowa opisująca stan elektronu w periodycznym potencjale spełnia warunek (tu przedstawiony tylko w jednym wymiarze x)

$$\Psi(x + d) = e^{ikd} \Psi(x) \tag{6a}$$

gdzie k to tzw. wektor falowy. W innej, ale równoważnej postaci twierdzenie Blocha wygląda tak

$$\Psi(x) = e^{ikx} u(x), \quad (6b)$$

gdzie $u(x)$ jest funkcją o periodyczności sieci, tzn. $u(x+d)=u(x)$.

Twierdzenie Blocha wynika z faktu, że jeśli dany układ jest *translacyjnie niezmienniczy* to reprezentujący go hamiltonian \mathbf{H} komutuje z operatorem translacji \mathbf{T} . A jeśli tak, to oba operatory mają wspólne funkcje własne, a zatem funkcje własne \mathbf{H} można ponumerować tak, jak funkcje własne \mathbf{T} . Tą liczbą kwantową numerującą wartości własne operatora translacji jest właśnie wektor falowy k , który w jednym wymiarze sprowadza się do skalaru k . Rzeczywiście, jeśli w (6a i b) mamy z prawej strony zależność od k , to lewa strona też musi być indeksowana liczbą k . Pozostaje nam jeszcze dowiedzieć się czym jest operator translacji. Jego działanie na funkcję falową polega na przesunięciu argumentu funkcji o wektor translacji

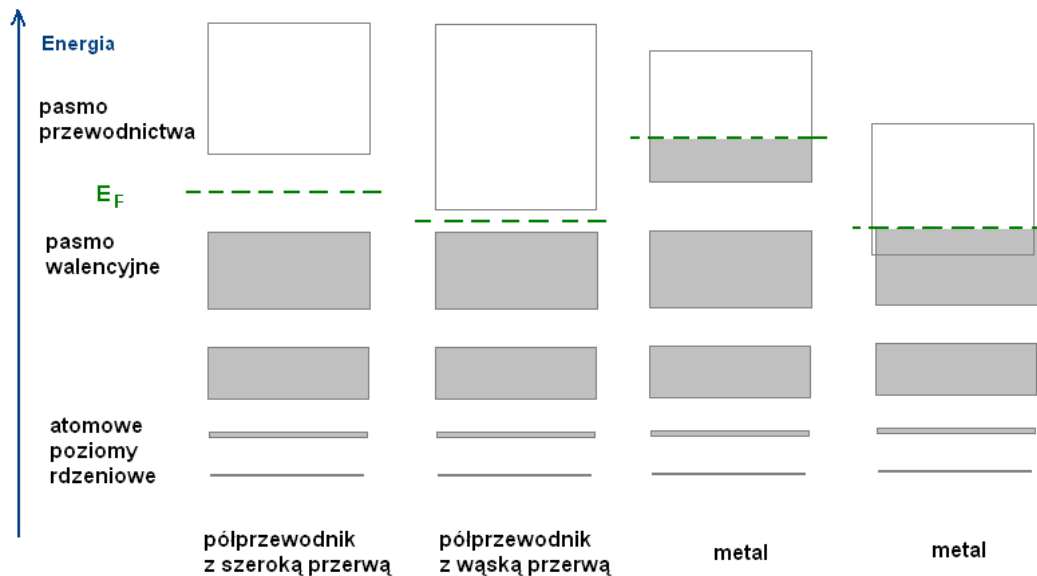
$$T_d \Psi(x) = \Psi(x+d) = e^{ikd} \Psi(x)$$

Ze względu na fakt, że e^{ikd} jest wielkością niemianowaną, k ma wymiar odwrotności długości. Periodyczny czynnik e^{ikd} zmienia się z okresem 2π , a zatem wystarczy, żeby k zmieniało się w zakresie od $(0, 2\pi/d)$ albo $(-\pi/d, +\pi/d)$. Ten zakres nazywa się *strefą Brillouina*. Dla nieskończonego kryształu k przyjmuje w tym zakresie wartości ciągłe, natomiast gdy kryształ jest skończony i zawiera (w kierunku x) N komórek elementarnych, k się dyskretyzuje i przyjmuje wartości $k = 0, \dots, 2\pi n/(dN), \dots, n=1,2,\dots,N$. Warto zauważyć, że dla dużego N , mamy bardzo dużo wartości k , bardzo blisko siebie (gęsto) położonych – czyli prawie jak ciągły indeks.

Niezależnie od k każda funkcja falowa jest rozwiązaniem równania (2) z periodycznym potencjałem (jakimś potencjałem w ramach komórki elementarnej). Te rozwiązania numerujemy „zwykłą” liczbą kwantową, np. n . Dla periodycznego potencjału będziemy mieli zatem wartości energii oznaczone jako $E_n(k)$, przy czym dla każdego n , k zmienia w przedziale $(-\pi/d, +\pi/d)$. Daje to obraz struktury energetycznej kryształów w postaci *pasem*. Pasma numerowane są przez n , a w ramach pasma energia zmienia się według $E_n(k)$, co nazywamy *relacją dyspersyjną* danego pasma. Liczba k – numeruje stany elektronowe w ramach każdego pasma.

Metale i półprzewodniki

Dla skończonego kryształu, pasma to „zbiory” energetycznie gęsto położonych poziomów jednoelektronowych, których jest jednak w paśmie skończona ilość. W skończonym kryształcie mamy też skończoną liczbę elektronów. Te wszystkie poziomy (stany) jednoelektrodowe musimy teraz obsadzić wszystkimi elektronami jakich dostarczą nam atomy tworzące kryształ. W rzeczywistości stany obsadzamy elektronami z powłok walencyjnych atomów tworzących kryształ, gdyż periodyczny potencjał w równaniu (2) dotyczy w zasadzie tylko walencyjnych elektronów (tych z zewnętrznymi powłokami elektronowymi atomów). Możliwe scenariusze obrazuje Rys. 18.



Rys.18. Schematy energetyczne pasm izolatorów, półprzewodników i metali.

O tym czy dany kryształ jest metalem, półprzewodnikiem czy izolatorem decyduje obsadzenie pasm i przerwy energetyczne pomiędzy pasmami. Jeśli każde dwa pasma są od siebie energetycznie oddzielone przerwą to o elektrycznym charakterze kryształu decyduje liczba elektronów walencyjnych (w zasadzie ich liczba w komórce elementarnej). Ponieważ każdy poziom jednoelektronowy może być ze względu na spin obsadzony maksymalnie dwoma elektronami, to przy parzystej liczbie elektronów kilka pasm jest całkowicie zapełnionych a pozostałe są nieobsadzone. Najwyższe energetycznie obsadzone pasmo nazywa się pasmem walencyjnym a najniższe nieobsadzone pasmem przewodnictwa. Jeśli przerwa energetyczna pomiędzy pasmem walencyjnym a pasmem przewodnictwa jest duża (kilka eV) to mamy do czynienia z izolatorem, jeśli jest mała (rzędu od 0.5 do 3 eV) mówimy o półprzewodniku samoistnym. Jeśli pasma się nakładają lub jeśli występuje przerwa energetyczna ale liczba elektronów jest nieparzysta (co skutkuje częściowym obsadzeniem ostatniego pasma), to mamy do czynienia z metalem. Gdy pasmo walencyjne zaledwie styka się z pasmem przewodnictwa mówimy wówczas o *semimetalu* lub *półmetal*. Taka sytuacja występuje w grafenie.

Ostatni obsadzony poziom elektronowy w metalu wyznacza tzw. poziom Fermiego, E_F (dla półprzewodników leży on w przerwie energii wzbronionych). Metal przewodzi prąd elektryczny dlatego, że bezpośrednio powyżej ostatniego obsadzonego poziomu elektronowego (powyżej E_F) istnieje *continuum* stanów nieobsadzonych, do których elektron może być wzbudzany (może sukcesywnie zwiększać swoją energię) poprzez powolne przyspieszanie w polu elektrycznym przyłożonym do metalu. W izolatorach i półprzewodnikach samoistnych jest to niemożliwe – wymagane jest wzbudzenie o kilka eV, które nie może być zrealizowane poprzez dodanie energii od zewnętrznego pola elektrycznego (gdyż jest ona zbyt mała dla typowych pól nierozrywających kryształu). Takie wzbudzenia realizuje się na ogół optycznie.

Metody obliczeniowe i wybrane struktury grafenowe

Przestrzenie modelowe.

Zamiast rozwiązywać równanie różniczkowe (2) można przybliżyć funkcję falową w skończonej bazie N funkcji φ (np. wzajemnie ortogonalnych i unormowanych - ortonormalnych)

$$\Psi = \sum_{i=1}^N c_i \varphi_i \quad (7)$$

i przybliżyć poszukiwaną energię układu jako

$$E = \langle \psi | H | \psi \rangle \quad (8)$$

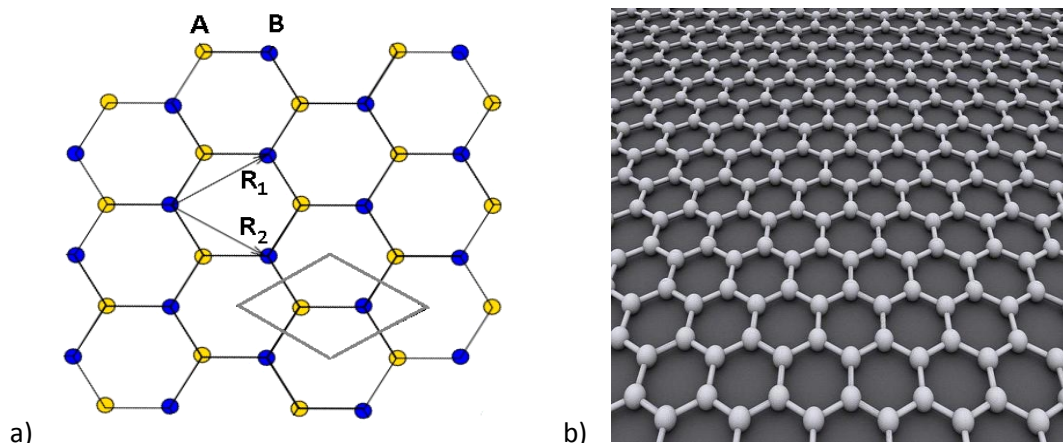
W bazie funkcji φ_i wyrażenie $\langle \psi | H | \psi \rangle$, a właściwie zbiór liczb $H_{ij} = \langle \varphi_i | H | \varphi_j \rangle$, przyjmuje postać macierzy, a zagadnienie $H\Psi(\mathbf{x}) = E\Psi(\mathbf{x})$ (równanie 1) przyjmuje postać zagadnienia własnego dla macierzy hamiltonianu H

$$\begin{bmatrix} H_{11} & \dots & H_{1N} \\ \dots & \dots & \dots \\ H_{N1} & \dots & H_{NN} \end{bmatrix} \begin{bmatrix} c_1 \\ \dots \\ c_N \end{bmatrix} = E \begin{bmatrix} c_1 \\ \dots \\ c_N \end{bmatrix} \quad (9)$$

Rozwiązanie polega na ZDIAGONALIZOWANIU macierzy H . Diagonalizacji dokonuje się na ogół za pomocą transformacji ortogonalnej (lub unitarnej gdy H jest zespolona) S , takiej, że $SHS^T = D$, gdzie D jest diagonalna, a wartości na diagonalu są wartościami własnymi macierzy H . Otrzymujemy N rozwiązań E_i (niektóre mogą się powtarzać – degeneracja). Te rozwiązania mają swoją interpretację zgodnie z *zasadą wariacyjną*, która stoi u podstaw przedstawionej wyżej metody przybliżonej zwanej *metodą wariacyjną*. Kolejne wartości E_i są górnymi przybliżeniami do kolejnych wartości własnych hamiltonianu, czyli przybliżają kolejne ściste energie układu fizycznego reprezentowanego przez hamiltonian H . Dla układów periodycznych H_{ij} - zależą od k , gdyż funkcje falowe (nawet te przybliżone) muszą zależeć od wektora falowego. Rozwiązując zatem zagadnienie (9) dla periodycznego układu otrzymujemy pasma $E_i = E_i(k)$. Żeby rzeczywiście otrzymać pasma, zagadnienie (9) trzeba rozwiązać dla wielu wartości k ze strefy Brillouina.

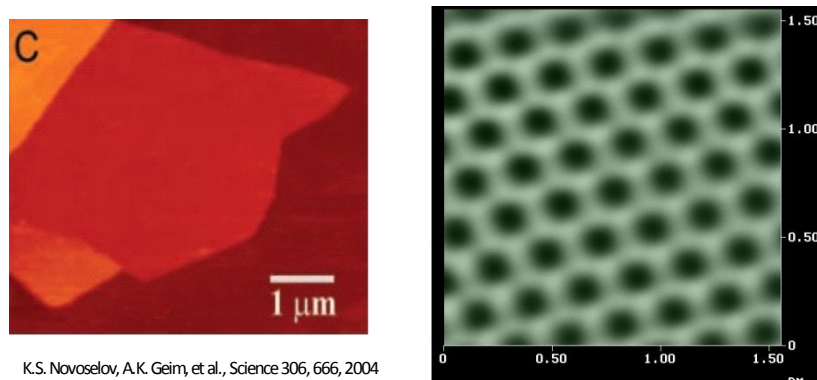
Grafen.

Grafen to dwuwymiarowa sieć rombowa węgla z bazą dwuatomową i wygląda jak sieć heksagonalna Rys. 19.



Rys.19. a) Heksagonalna sieć grafenu. Komórka elementarna zawiera dwa atomy węgla. R_1 i R_2 to dwa wektory bazowe sieci (z ich kombinacji powstają wektory translacji wskazujące i powielające kolejne komórki elementarne grafenu. b) Model grafenu. <http://en.wikipedia.org/wiki/Graphene>

Grafen został po raz pierwszy otrzymany w 2005 roku przez Geima i Novoselova, którzy w 2010 roku zostali za to uhonorowani nagrodą Nobla. Pojedyncze warstwy grafenu otrzymali oni odrywając za pomocą taśmy Scotch kolejne warstwy z grafitu pozostawionego na podłożu przez „ołówkę”. Bardzo łatwo jest rozpoznać pojedynczą warstwę, gdyż jest ona przezroczysta, podczas gdy już warstwa podwójna znacząco tłumi światło. Obecnie grafen otrzymuje się bardziej wyrafinowanymi metodami, np. przez odparowanie krzemu z warstw SiC lub przez osadzanie metodą CVD. Przykładowe obrazy pojedynczych płatków grafenu otrzymane za pomocą mikroskopii skaningowej i tunelowej pokazane są na Rys. 20.



K.S. Novoselov, A.K. Geim, et al., Science 306, 666, 2004

Rys.20. Mikroskopowe obrazy płatków grafenowych.
http://www.polymermicroscopy.com/eng_afm_graphit.htm
 and <http://www.graphene.manchester.ac.uk/story/image-gallery/>

Trwałość grafenu wyjaśniają wiązania kowalencyjne typu σ w płaszczyźnie, tworzone pomiędzy każdymi dwoma sąsiadami poprzez nakrywanie się orbitali hybrydowanych sp^2 . Na każdym węglu pozostaje jednak po jednym elektronie obsadzającym orbital p_z (prostopadły do płaszczyzny). Te orbitale też tworzą wiązania poprzez tworzenie orbitali molekularnych typu π , wiążących gdy dodajemy do siebie dwa orbitale p_z , lub gdy je odejmujemy. Ponieważ grafen jest periodyczny w 2 wymiarach, a każde dwa sąsiednie orbitale π są „takie same”, tworzy się z nich (ze względu na powielanie) orbital krystaliczny π . Temu orbitalowi, z względu na symetrię translacji odpowiadają pasma energetyczne, przez analogię nazywane pasmami π . Obsadzone jest tylko pasmo wiążące i

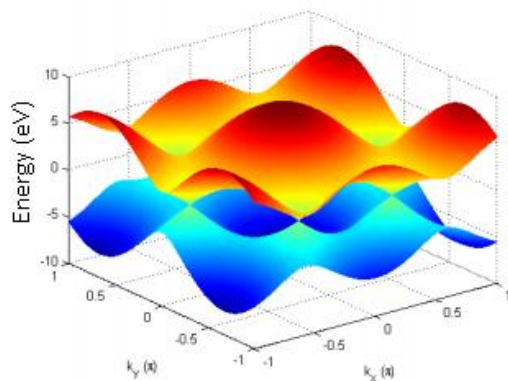
nazywa się pasmem walencyjnym; niewiążące pasmo π^* (antywiążące) to pasmo przewodnictwa. Oba pasma $E_i(\mathbf{k})$, $i=1,2$ (π oraz π^*) zależą od wektora falowego \mathbf{k} , ale teraz jest to rzeczywisty wektor o dwóch składowych k_x i k_y . Poziom Fermiego leży między nimi. To dlatego tylko te orbitale (a nie orbitale wiążące σ) decydują o własnościach elektrycznych grafenu i tylko im poświęcimy dalszą uwagę.

Metoda ciasnego wiązania.

Metoda ciasnego wiązania jest praktyczną realizacją metody przestrzeni modelowej. Jednocząstkowe funkcje (7) mające opisywać stany elektronów walencyjnych (elektronów π) w grafenie rozwijamy w bazie orbitali p_z . W komórce elementarnej są dwa takie orbitale, ale z powodu periodyczności kryształu, musimy najpierw utworzyć z nich orbitale zdelokalizowane (tzw. sumy Blocha), spełniające twierdzenie Blocha, a zatem zależne od k . Takie dwa orbitale bazowe bardzo upraszczają zagadnienie: w tej bazie macierz hamiltonianu ma wymiar 2 (tylko 2, ale kosztem tego że wszystko zależy od k). Tak prostą macierz (9) można zdiagonalizować analitycznie. Dostaniemy wówczas dwa rozwiązania $E(k)$, nazwane $E_+(k)$ – wiążące pasmo π oraz $E_-(k)$ – pasmo antywiążące.

$$E(k_x, k_y) = \pm t \left\{ 1 + 4 \cos\left(\frac{\sqrt{3}k_x a}{2}\right) \cos\left(\frac{k_y a}{2}\right) + 4 \cos^2\left(\frac{k_y a}{2}\right) \right\}^{1/2} \quad (10)$$

Strefa Brillouina (strefa zmienności dwóch składowych k_x i k_y wektora \mathbf{k}) jest teraz dwuwymiarowa i wygląda jak sześciokąt. Energie tych pasm pokazane są na Rys. 21.



Rys.21. Pasma π w grafenie. Praca doktorska M. Pelc, Toruń, 2012.

Macierz hamiltonianu ma 4 elementy. Te diagonalne odpowiadają sytuacji gdy w elemencie

$H_{ij} = \langle \varphi_i | H | \varphi_j \rangle$ spotyka się orbital p_z z orbitalem p_z na tym samym atomie. Jest to tzw. całka nawęzłowa. Ponieważ oba elementy diagonalne są takie same (takie same dwa atomy węgla w komórce elementarnej) zatem można przypisać im dowolną wartość, która *de facto* oznaczać będzie tylko identyczne przesunięcie wartości obu pasm na skali energii. Przyjęto się przypisywać im wartość zero, i tym samym jest to wartość energii na poziomie Fermiego. Elementy pozadiagonalne zależą od całki $H_{ij} = \langle \varphi_i | H | \varphi_j \rangle$ pomiędzy sąsiednimi orbitalami p_z oraz od k_x i k_y . Tę całkę również zamiast obliczać traktujemy jako parametr empiryczny, oznaczamy t i przypisujemy mu wartość 2.7

eV (wartość w jednostkach energii bowiem jest to element macierzowy operatora energii całkowitej). Po diagonalizacji otrzymamy (10).

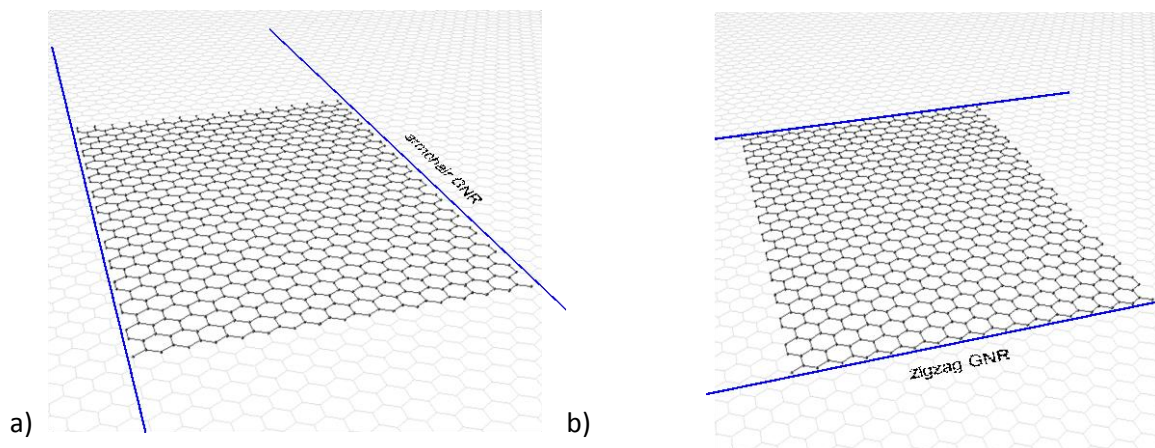
Skończone struktury grafenowe.

W przypadku skończonych struktur grafenowych (nieperiodycznych), które są w zasadzie dużymi molekułami zbudowanymi z atomów węgla, nie można mówić już o komórce elementarnej i wówczas baza funkcji w (7) musi zawierać wszystkie orbitale p_z zlokalizowane na wszystkich atomach. Nie ma już co prawda zależności od k , ale kosztem tego musimy zbudować i zdiagnozować bardzo duże macierze hamiltonianu. Dużym pocieszeniem jest fakt, że na ogół pracuje się w tzw. *przybliżeniu najbliższych sąsiadów*, w którym zaniedbuje się wszystkie elementy macierzowe łączące orbitale nie należące do sąsiednich węzłów węglowych. Jest to bardzo racjonalne przybliżenie. W efekcie macierz hamiltonianu jest prawie cała „pusta”, niezerowe elementy pojawiają się w zasadzie tylko na obrzeżach diagonalii (elementy której równają się zero). Po diagonalizacji otrzymujemy zbiór wartości własnych, które reprezentują poziomy energetyczne tej „molekuły”.

Periodyczne struktury grafenowe.

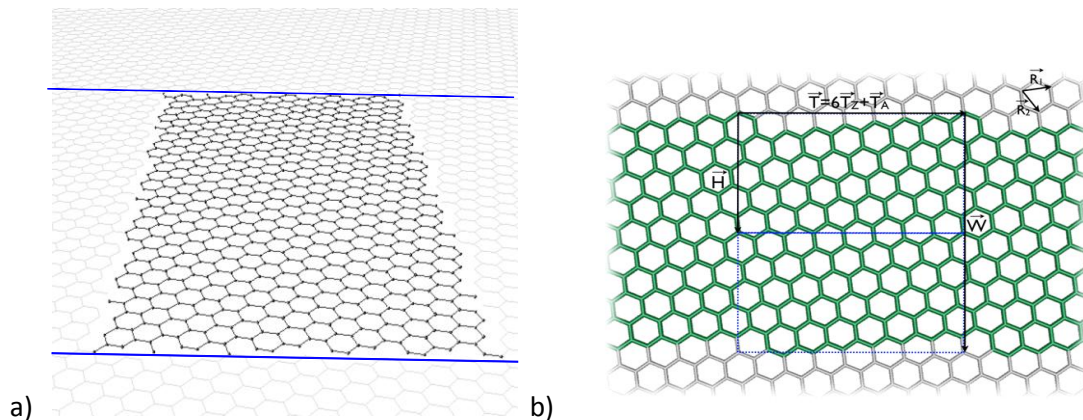
Nanowstęgi.

Periodyczne struktury grafenowe to nanowstęgi i nanorurki węglowe. Wstęgi powstają przez wycięcie z nieskończonego arkusza grafenu paska o zadanej szerokości i dowolnym kierunku. Gdy brzeg wstęgi jest wycięty wzdłuż wiązań pomiędzy atomami węgla otrzymujemy wówczas wstęgę o brzegu fotelowym (*armchair*). Natomiast gdy cięcie jest w kierunku prostopadłym, otrzymana wstęga ma brzeg typu zygzak (*zigzag*). Oba przypadki zilustrowane są na Rys. 22.



Rys.22. Wstęgi grafenowe. a) armchair oraz b) zigzag.

Wycinając wstęgę w dowolnym innym kierunku otrzymujemy tzw. wstęgi *chiralne*, przykład pokazany jest na Rys. 23. Taka wstęga może mieć nieregularne, choć periodyczne brzegi.



Rys.23. Przykłady wstęg chiralnych. W b) zaznaczono komórkę elementarną wstęgi. Jest to obszar ograniczony z zewnątrz prostokątem o długości T i szerokości W .

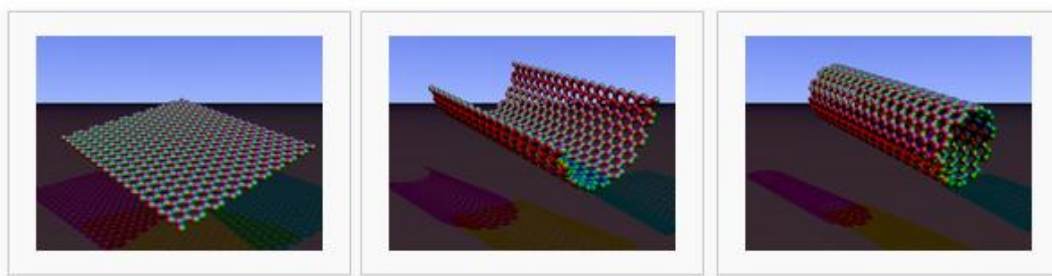
Każdą wstęgę można jednoznacznie zdefiniować przez podanie wektora translacji wstęgi T , jej szerokości W oraz szczegółów brzegu (jeśli chcemy modyfikować brzeg po wycięciu poprzez dodawanie lub odejmowanie węzłów brzegowych). Komórki elementarne periodycznych wstęg grafenowych zawierają wszystkie atomy mieszczące się w prostokącie zdefiniowanym przez wektor translacji wstęgi T (Rys. 23 b) i wektor definiujący szerokość wstęgi W .

Łatwo jest zbudować macierz hamiltonianu dla takiego układu. Procedura jest następująca:

1. numerujemy wszystkie węzły w danej komórce elementarnej i w obu sąsiednich (numeracja przebiega tylko do liczby atomów w komórce),
2. wyzerowujemy wszystkie elementy diagonalne (w zasadzie całą macierz),
3. w elementach poza-diagonalnych łączących węzły należące do tej samej komórki elementarnej wpisujemy wartość parametru t ,
4. w elementach diagonalnych łączących i -ty węzeł z komórki z j -tym węzłem z komórki po prawej stronie wstawiamy wartość $t \times \exp(ikd)$, a w elementach łączących k -ty węzeł komórki z l -tym węzłem komórki leżącej po lewej stronie wstawiamy $t \times \exp(-ikd)$
5. jeśli dany węzeł i -ty połączony jest z węzłem j -tym z tej samej komórki i komórki sąsiedniej (np. prawej) to wstawiamy wartość $t(1 + \exp(ikd))$, gdyż te węzły łączą się na dwa sposoby,
6. sprawdzamy czy macierz jest hermitowska (tzn. $H_{ij} = H_{ji}^*$)
7. ponieważ k zmienia się w granicach $0-2\pi/d$, zatem czynnik $\exp(\pm ikd)$, można zastąpić czynnikiem $\exp(\pm ik)$, i zmieniać k od $0-2\pi$

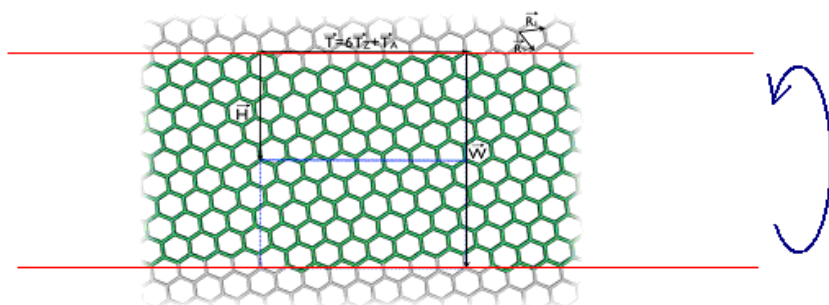
Nanorurki.

Wstęgi grafenowe można też zwinąć i tworzyć z nich nanorurki węglowe, tak jak to pokazano na Rys. 24.



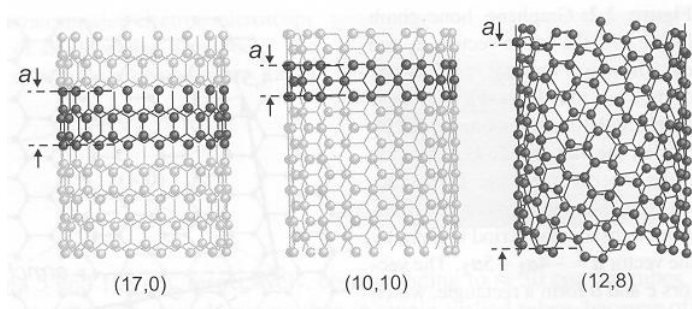
Rys.24. Tworzenie nanorurki ze wstęgi grafenowej. http://en.wikipedia.org/wiki/Carbon_nanotube

W przypadku nanorurek atomy węgla z jednego brzegu wstęgi są połączone z atomami z drugiego brzegu. Ilustruje to Rys. 25. Należy o tym pamiętać budując macierz hamiltonianu.



Rys.25. Tworzenie nanorurki ze wstęgi grafenowej poprzez zwinięcie i połączenie atomów z przeciwnych brzegów.

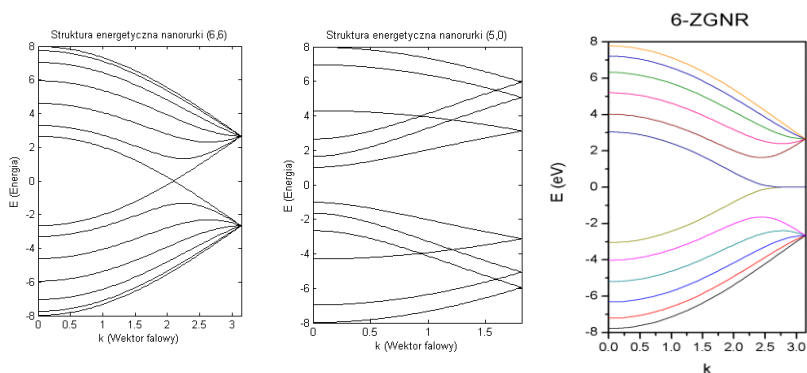
Mogą powstać 3 różne typy nanorurek pokazane na Rys. 26, fotelowe, zygzakowate i chiralne. Tym razem rodzaj nanorurki jest określony przez kształt brzegu na cięciu prostopadłym do osi nanorurki.



Rys.26. Trzy typy nanorurek węglowych: od lewej –zigzag, armchair, chiralna.

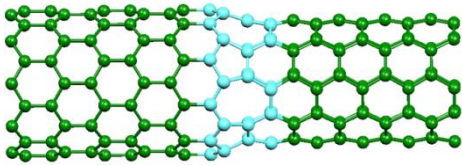
“Physical Properties of Carbon Nanotubes”, M.S. Dresselhouse, G. Dresselhouse, and R. Saito, Imperial College Press, London 1998.

Każdą nanorurkę węglową można jednoznacznie zdefiniować przez jej tzw. *wektor chiralny* zwany również *wektorem obwodu*. Na Rys. 25 pokrywa się on z wektorem W (w literaturze oznaczany jest przez C_h). Współrzędne wektora chiralnego w bazie wektorów definiujących sieć grafenu (wektory R_1 i R_2 na Rys. 19a) to dwie liczby naturalne (n,m) . Definiują one jednoznacznie każdą nanorurkę.



Rys.27. Pasmowa struktura energetyczna różnych grafenowych układów periodycznych. Od lewej: nanorurka fotelowa (6,6), nanorurka zygzak (5,0), wstęga zygzak o szerokości 6-ciu komórek elementarnych grafenu.

Ponieważ zarówno wstęgi grafenowe jak i nanorurki są obiektami jednowymiarowymi, ich struktura energetyczna to seria pasm $E_i(k)$. Trzy przykłady pokazane są na Rys. 27. W zależności od średnicy i skrętności nanorurek, a w przypadku nanowstęg w zależności od ich szerokości i kształtu brzegu, powstają struktury, które mogą być albo metaliczne albo półprzewodnikowe (na Rys. 27 widać, że nanorurka (6,6) oraz wstęga zygzak są metalami – brak przerwy energetycznej, natomiast nanorurka (5,0) jest półprzewodnikiem o przerwie energetycznej $\sim 2\text{eV}$). Ponadto, dwie różne nanorurki można łączyć za pomocą tzw. *defektów topologicznych* 5/7 (czasami 8). Pokazane jest to na Rys. 28. Z takich złącz można tworzyć *supersieci* oraz *kropki kwantowe*.



Rys.28. Złącze dwóch nanorurek zygzakowatej $(2n,0)$ oraz fotełowej (n,n) . Złącze jest zdefiniowane przez n par pięciokąt/siedmiokąt.

Zarówno z grafenem jak i z nanorurkami oraz nanowstęgami wiąże się ogromne nadzieje na przyszłą elektronikę węglową. Wykonano już prototypowe tranzystory zbudowane z nanorurek i wstęg grafenowych. Ich rozmiary to zaledwie kilka-kilkadziesiąt nanometrów.

Stany rezonansowe cząstek w nanostrukturach

Mirosław Bylicki

Podstawowe informacje o własnościach nanostruktury, a właściwie, o zachowaniu się uwięzionych w niej cząstek, można uzyskać modelując oddziaływanie takich cząstek z nanostrukturą za pomocą stosunkowo prostych potencjałów. Jak zaproponować taki potencjał? Zajmijmy się krótko przypadkiem kiedy nanostruktura jest zbudowana z kilku obszarów materiałów krystalicznych półprzewodnikowych (możliwe także izolatory i przewodniki) i rozważamy poruszający się w takiej strukturze elektron. Dla każdego z tych materiałów mamy charakterystyczny rozkład pasm dozwolonych energii, w szczególności pasmo przewodnictwa. Jeśli elektron ma energię w zakresie pasma przewodnictwa, to „czuje się on swobodny”. Jednak obszar zbudowany z danego materiału jest ograniczony (o rozmiarach rzędu kilku, kilkudziesięciu, kilkuset nanometrów). Elektron musi trafić na granicę z materiałem, w którym np. energia rozważanego elektronu przypada na przerwę energetyczną a pasmo przewodnictwa jest powyżej. Wtedy elektron ma trudności z wnikięciem do tego sąsiedniego obszaru. W ogólności takie wnikięcie nie jest jednak niemożliwe. Prawdopodobieństwo jego zajścia zależy od energii elektronu i skoku energii przewodnictwa na granicy materiałów. Całkowicie niemożliwe jest w przypadku idealnego izolatora. Wtedy ten skok jest nieskończony. Modelowy potencjał takiego układu będzie przyjmował w poszczególnych obszarach wartości energii odpowiadające dnom pasm przewodnictwa materiałów wypełniających te obszary.

Wbrew pozorom, sytuacja może być bardzo skomplikowana w zależności od przestrzennego rozkładu różnych materiałów tworzących nanostrukturę. Najprostszym przypadkiem jest rozkład kilku równoległych warstw, które są na tyle rozległe, że w przybliżeniu można je rozważać jako nieskończone. Wtedy potencjał nie zmienia się w płaszczyznach poszczególnych warstw. Zmienia się tylko w kierunku zmian materiału, przechodzenia jednej warstwy w inną. Model sprowadza się do jednego wymiaru odpowiadającego temu właśnie kierunkowi. Zagadnienie badania struktury energetycznej sprowadza się do rozwiązania równania Schroedingera (równania własnego dla operatora energii) jednowymiarowego układu prostokątnych studni i barier. Określenie, co jest studnią, co barierą zależy od względnej wartości potencjału.

Jaką dostajemy informację? Formalnie, matematycznie rzecz biorąc, otrzymujemy wartości własne operatora energii. Fizycznie – dowiadujemy się jakie energie elektronu w nanostrukturze są dozwolone. Mogą to być ściśle określone energie odpowiadające stanom związanym elektronu, tzn. elektronowi uwięzionemu w nanostrukturze tak, że nie może się z niej wydostać. Ich znajomość daje możliwość poznania własności optycznych nanostruktury, tzn. pozwala dowiedzieć się jakie światło nanostruktura może wysyłać lub pochłaniać oraz jaka musiałaby być częstotliwość fali elektromagnetycznej, która mogłaby uwolnić elektron z nanostruktury. Oprócz tych stanów, ściśle związanych, możliwe są też stany kwazizwiązane (albo pseudozwiązane). To takie, w których elektron wydaje się być związany w nanostrukturze, jednak może się z niej w skończonym czasie uwolnić bez otrzymania dodatkowej energii. Wyżej zasugerowano, że elektron może wnikać w obszar o pasmie przewodnictwa powyżej jego energii, to oznacza, że w sprzyjających okolicznościach może wydostać się na zewnątrz nanostruktury. Takim stanom też odpowiadają charakterystyczne energie. Nie są jednak ściśle określone (dyskretne), są nieco „rozmyte”. Znajomość takich stanów jest bardzo ważna,

np. ze względu na własności przewodzenia prądu przez nanostruktury, przedostawania się elektronów do wewnątrz i na zewnątrz nanostruktur.

Ta część zajęć poświęcona jest stanom pseudozwiązanym, zwanym także rezonansowymi. Zaczynamy od rozważenia prostego jednowymiarowego modelu, układu studnia-bariera. Na przykładzie analitycznych rozwiązań dla tego przypadku poznajemy podstawowe własności stanów rezonansowych. Następnie, przedstawiamy metody przeznaczone specjalnie do poszukiwania stanów rezonansowych: metodę stabilizacji i metodę obrotu zespolonego.

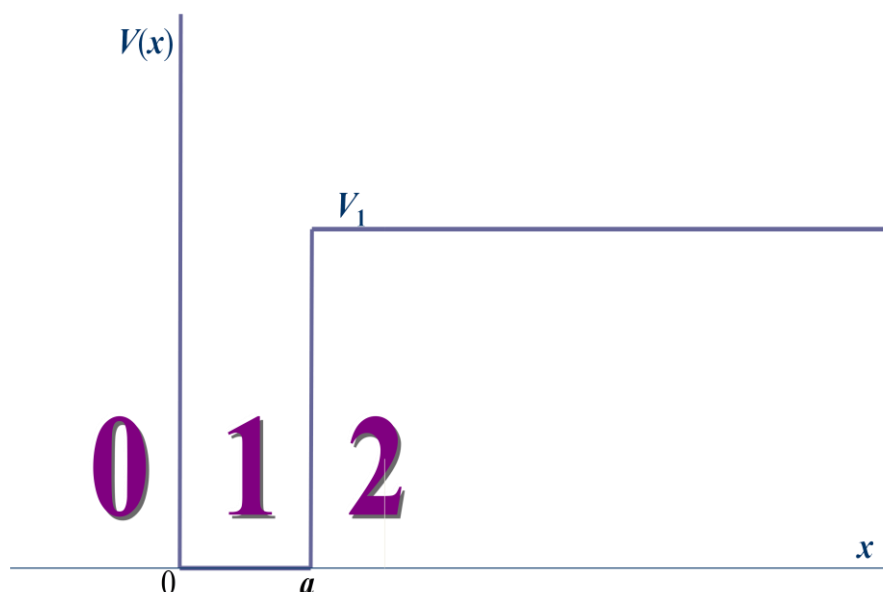
Układ studnia-bariera

Zajmujemy się tutaj jednowymiarowym układem studnia-bariera, który wydaje się być najprostszym przykładem ilustrującym fakt istnienia i właściwości stanów rezonansowych. Rozważamy najpierw tylko studnię potencjału, dla której znajdujemy stany związane. Później dodajemy do niej barierę potencjału, by pokazać, jak stany związane studni przeistaczają się w rezonansowe, jak zmieniają się ich własności.

Stany związane w studni potencjału

Cząstka, którą rozważamy (elektron) porusza się po prostej, jej położenie określa jedna współrzędna – x . Z nanostrukturą (resztą świata) oddziałuje tak, że zyskuje energię potencjalną $V(x)$. Nasza nanostruktura jest nanowarstwą półprzewodnika o grubości a . Dno pasma przewodnictwa w tej warstwie przyjmujemy za 0 na skali energii. Z jednej strony warstwa przylega do nieskończonego półprzewodnika o pasmie przewodnictwa położonym o V_1 wyżej. Z drugiej strony jest ograniczona idealnym izolatorem, w którym potencjał jest nieskończony (pasma przewodnictwa „nieskończenie wysoko”). Potencjał wygląda więc tak, jak na rys. 29 i zapisuje się wzorem

$$V(x) = \begin{cases} \infty, & \text{dla } x < 0, \\ 0, & \text{dla } 0 \leq x \leq a, \\ V_1, & \text{dla } x > a. \end{cases}$$



Rys. 29. Studnia potencjału. Może modelować układ nanowarstwy półprzewodnikowej (obszar o numerze 1) pomiędzy nieskończonym półprzewodnikiem (obszar 2), o dnie pasma przewodnictwa położonym o V_1 powyżej dna pasma przewodnictwa w nanowarstwie, a nieskończonym izolatorem (obszar 0). W zakresie energii pomiędzy 0 a V_1 spodziewamy się jedynie stanów związanych.

Naszym zadaniem jest rozwiązanie równania Schroedingera. Umówmy się najpierw, że interesuje nas energia z zakresu od 0 do V_1 . Poniżej zera nie możemy spodziewać się sensownych, fizycznych rozwiązań, ponieważ oznaczałyby one ujemną energię kinetyczną cząstki. Podobnie byłoby, gdyby cząstka o energii $0 \leq E \leq V_1$ znajdowała się bardzo daleko, próbowałaby uciekać do nieskończoności. Ta ucieczka nie byłaby możliwa. Zatem, w tak ograniczonym zakresie energii spodziewamy się tylko stanów związanych.

Ponieważ potencjał jest określony w szczególny sposób, w trzech obszarach, równanie Schroedingera rozwiązujemy dla każdego obszaru oddzielnie, a następnie otrzymane rozwiązania zszywamy w jedną całość. W obszarze „0” mamy

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \Phi_0(x) + (\infty - E)\Phi_0(x) = 0.$$

Wzajemne redukowanie się obu wyrazów dla dowolnej wartości x (w obszarze „0”) wymaga albo nieskończonej drugiej pochodnej Φ_0 albo $\Phi_0(x) = 0$. Pierwsza możliwość jest absurdalna. Jednak druga jest niesłychanie prosta matematycznie i oczywista fizycznie – funkcja falowa równa zero oznacza zerową gęstość prawdopodobieństwa cząstki w obszarze „0”. Tak właśnie miało być, elektron nie może wnikać do obszaru izolatora. Mamy więc rozwiązanie w obszarze „0”,

$$\Phi_0(x) = 0.$$

Na odcinku $[0, a]$ – w studni – równanie Schroedingera ma postać

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \Phi_1(x) + E\Phi_1(x) = 0$$

lub inaczej

$$\frac{d^2}{dx^2} \Phi_1(x) + k_1^2 \Phi_1(x) = 0, \text{ gdzie } k_1^2 \equiv \frac{2mE}{\hbar^2}, k_1 > 0.$$

Jest identyczne z równaniem oscylatora harmonicznego. Jego ogólne rozwiązanie można zapisać w postaci

$$\Phi_1(x) = A \sin k_1 x + B \cos k_1 x.$$

Pozostaje do znalezienia rozwiązanie w obszarze „2”. Równanie

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \Phi_2(x) + (V_1 - E)\Phi_2(x) = 0$$

jest równoważne

$$\frac{d^2}{dx^2} \Phi_2(x) - k_2^2 \Phi_2(x) = 0, \text{ gdzie } k_2^2 \equiv \frac{2m(V_1 - E)}{\hbar^2}, k_2 > 0.$$

Różni się od równania oscylatora harmonicznego znakiem przy stałej k_2^2 . Ogólne rozwiązanie,

$$\Phi_2(x) = C e^{-k_2 x} + D e^{k_2 x},$$

jest dowolną kombinacją liniową funkcji wykładniczych, zanikającej i „wybuchającej” w nieskończoności. Ta ostatnia jest niedopuszczalna ze względów fizycznych. Prawdopodobieństwo znalezienia cząstki w nieskończoności (bardzo daleko od studni) byłoby bardzo duże, a z powodów energetycznych cząstka nie może uciekać nieskończenie daleko od studni. Dlatego spośród wszystkich rozwiązań matematycznych, wybieramy tylko te dopuszczalne fizycznie, przyjmując $D = 0$. (Za chwilę zobaczymy, że z tego właśnie fizycznego warunku wynika dyskretny charakter dopuszczalnych energii). Ostatecznie rozwiązaniem w obszarze „2” jest

$$\Phi_2(x) = C e^{-k_2 x}.$$

Teraz rozwiązania z poszczególnych obszarów trzeba zszyć w jedną gładką całość. Zapewnimy w ten sposób, że gęstość prawdopodobieństwa znalezienia cząstki w położeniu x będzie zmieniać się w sposób ciągły. Żądamy więc by funkcja i jej pochodna (jeśli możliwe) była ciągła w punktach zszycia, tzn. w $x = 0$ i $x = a$, czyli aby

$$\begin{aligned} \Phi_0(0) &= \Phi_1(0), \\ \Phi_1(a) &= \Phi_2(a), \\ \frac{d\Phi_1}{dx} \Big|_{x=a} &= \frac{d\Phi_2}{dx} \Big|_{x=a}. \end{aligned}$$

(W naszym przypadku nie można zapewnić ciągłości pochodnej w $x = 0$.) Te formalne warunki prowadzą do układu szczegółowych równań liniowych na współczynniki A, B i C :

$$\begin{aligned} 0 &= A \sin 0 + B \cos 0 \\ A \sin k_1 a + B \cos k_1 a &= C e^{-k_2 a} \\ k_1 A \cos k_1 a - k_1 B \sin k_1 a &= -k_2 C e^{-k_2 a}. \end{aligned}$$

Z pierwszego wyniku natychmiast $B = 0$. Dwa pozostałe

$$\begin{aligned} A \sin k_1 a - C e^{-k_2 a} &= 0 \\ k_1 A \cos k_1 a + k_2 C e^{-k_2 a} &= 0 \end{aligned}$$

pozwalają wyznaczyć A i C . Jeśli te dwa równania są liniowo niezależne, to istnieje jedno jedyne rozwiązanie. Jest nim $A = 0$ i $C = 0$. Takie rozwiązanie jest нефизyczne – oznacza, że funkcja falowa jest równa 0 wszędzie, a zatem, prawdopodobieństwo znalezienia cząstki gdziekolwiek jest zerowe – tej cząstki nie ma (!).

Czy istnieją jakieś fizyczne rozwiązania? Mogą istnieć pod warunkiem, że powyższe dwa równania są liniowo zależne. Żeby takie były, macierz główna współczynników musi być osobiłwa, jej wyznacznik musi zniknąć. Żądamy zatem, aby

$$\begin{vmatrix} \sin k_1 a & -e^{-k_2 a} \\ k_1 \cos k_1 a & k_2 e^{-k_2 a} \end{vmatrix} = 0.$$

Właściwie jest to warunek na energię E , uwikłaną w k_1 i k_2 . Może się zdarzyć, że dla pewnych wartości energii z przedziału powyżej zera a poniżej V_1 , warunek powyższy jest spełniony. Wtedy istnieją stany cząstki związanej w studni dla takich właśnie energii.

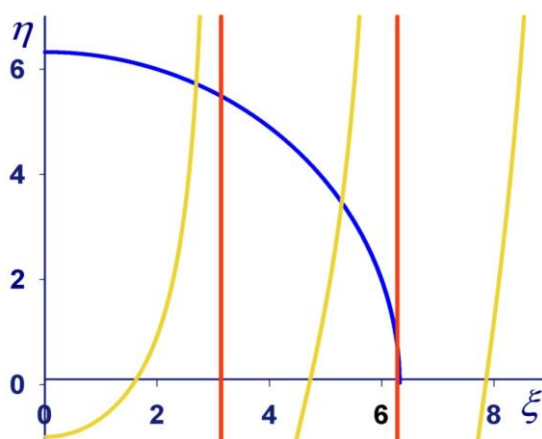
Równanie powyższe jest nieliniowe a jego rozwiązanie nie jest trywialne. Po rozwinięciu wyznacznika i przyjęciu (dla uproszczenia) oznaczeń $\xi \equiv k_1 a$ i $\eta \equiv k_2 a$, otrzymujemy

$$\eta = -\xi \cot \xi.$$

Z definicji przyjętych dla k_1 i k_2 mamy również tożsamościowy związek

$$\xi^2 + \eta^2 = \frac{2m}{\hbar^2} V_1 a^2 \equiv R^2.$$

Rysunek 30 przedstawia wykresy dwóch ostatnich zależności między ξ i η , wykonane dla konkretnych wartości parametrów $m = 1$ j.u., $a = 2$ j.u., $V_1 = 5$ j.u. i $\hbar = 1$ j.u. (Używamy tu jednostek umownych, specyficznych dla rozważanego układu. Umawiamy się, że masa rozważanej cząstki stanowi jednostkę masy. Jednostki energii są takie, że stała Plancka podzielona przez 2π jest równa 1.) Widać, że w tym przypadku mamy dwa punkty przecięcia wykresów, dwa rozwiązania. Łatwo uzasadnić, że w ogólnym przypadku liczba rozwiązań n jest określona przez warunek $(2n - 1)\frac{\pi}{2} \leq R < (2n + 1)\frac{\pi}{2}$. Jest zawsze skończona. Zależy od szerokości studni a i jej głębokości V_1 . Dla studni wąskiej lub płytkiej może nie istnieć ani jedno rozwiązanie.



Rys. 30. Wykresy zależności $\eta = -\xi \cot \xi$ (żółty) i $\xi^2 + \eta^2 = R^2$ (niebieski). Punkty ich przecięcia odpowiadają istnieniu niezerowych rozwiązań równania Schroedingera. Kolorem czerwonym wykreślono asymptoty pionowe w $\xi = m\pi$.

Wartości energii odpowiadających tym rozwiązaniom, wartości własne hamiltonianu, można uzyskać odczytując z wykresu wartości ξ i η dla punktów przecięcia wykresów. Zastosujemy nieco inny, choć także graficzny, wariant znajdowania tych energii. Ma być spełniony związek $\eta + \xi \cot \xi = 0$. Wykreślmy więc wartości funkcji

$$f(E) = \eta(E) + \xi(E) \cot \xi(E),$$

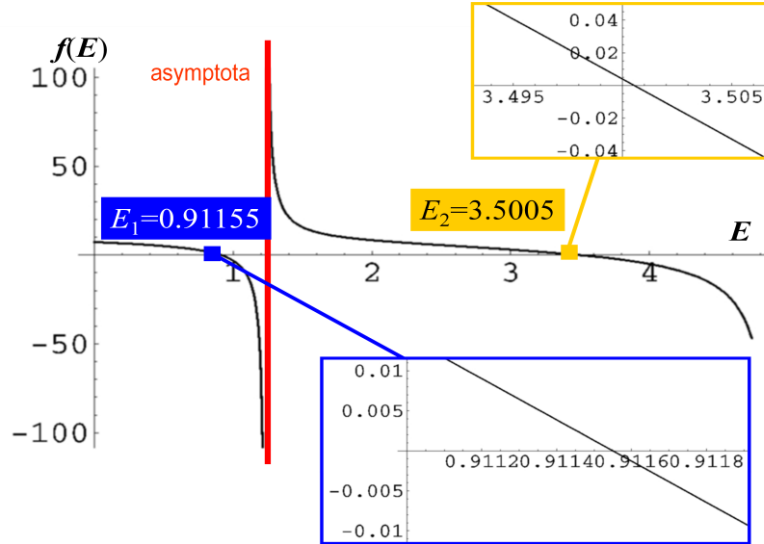
w zakresie energii, w którym problem rozważaliśmy, $0 < E < V_1$, i znajdziemy jej miejsca zerowe. Rysunek 31 przedstawia taki wykres dla przyjętych wcześniej wartości parametrów. Mamy dwa miejsca zerowe, to są dwie energie własne naszego układu: $E_1 \cong 0.91155$ i $E_2 \cong 3.5005$. Wstawiając je po kolei do jednego z układu dwóch – liniowo zależnych w takim przypadku – równań na A i C dostajemy związek między współczynnikami A i C :

$$C_i = A_i e^{k_2(E_i)a} \sin[k_1(E_i)a], \text{ dla } i = 1, 2.$$

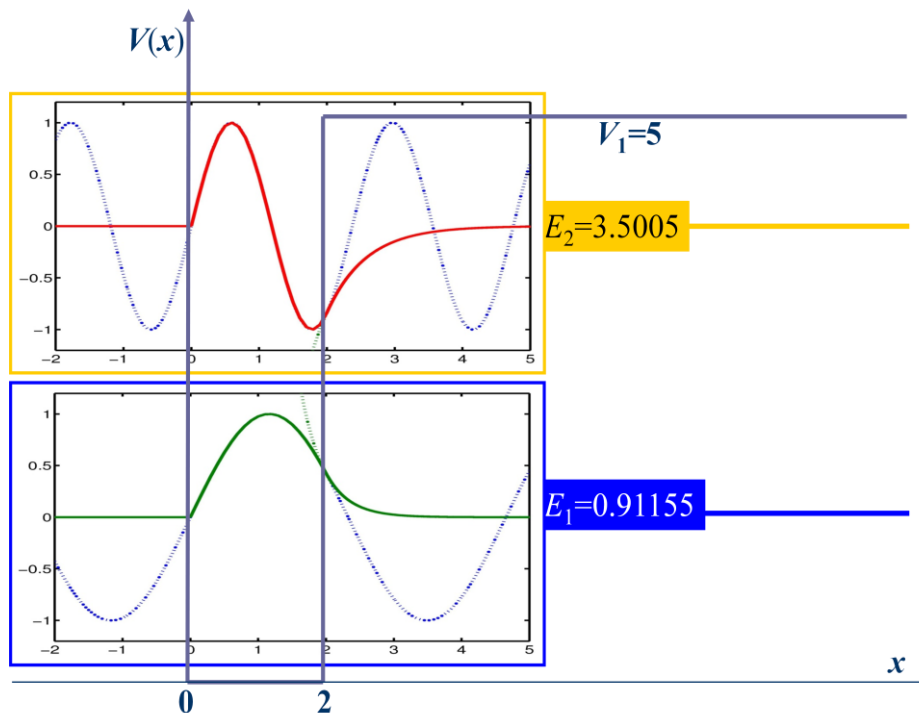
W rezultacie, funkcje falowe odpowiadające wartościom E_i daje się zapisać wzorem

$$\Psi_i(x) = A_i \begin{cases} 0, & \text{dla } x < 0, \\ \sin[k_1(E_i)x], & \text{dla } 0 \leq x \leq a, \\ e^{-k_2(E_i)(x-a)} \sin[k_1(E_i)a], & \text{dla } x > a. \end{cases}$$

Ich wykresy znajdują się na rysunku 32.



Rys. 31. Wykres funkcji $f(E) = \eta(E) + \xi(E) \cot \xi(E)$. Jej miejsca zerowe odpowiadają wartościom własnym energii (dozwołonym energiom badanego układu). Pionowa czerwona linia jest asymptotą.



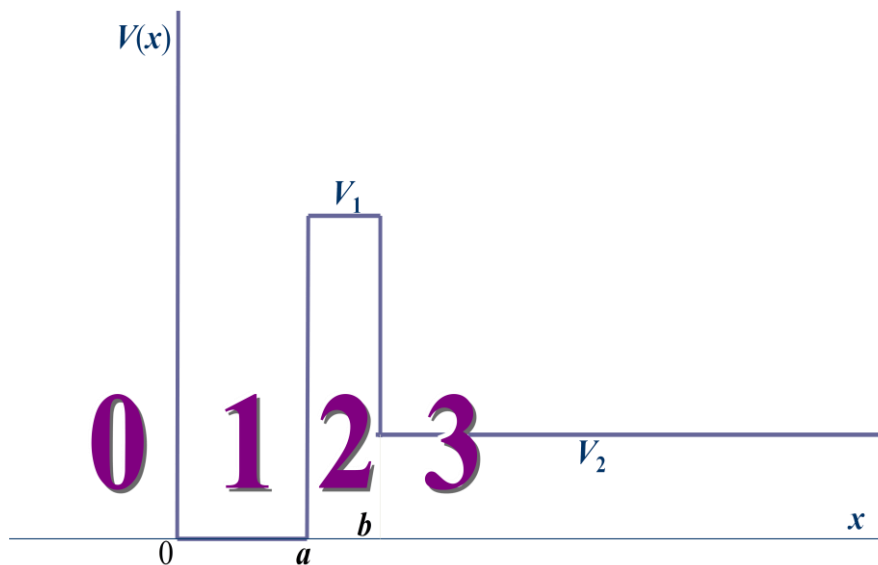
Rys. 32. Wykres funkcji falowych jedynych dwóch stanów związanych cząstki w rozważanej studni potencjału. Umieszczono je tradycyjnie na tle wykresu potencjału, każdą na wysokości odpowiadającej jej energii własnej. Ma to na celu wyłącznie ułatwienie skojarzeń. Jediną wspólną oś dla wszystkich wykresów na tym rysunku jest oś współrzędnych cząstki. Każdy z wykresów funkcji ma swoją, niezależną oś wartości funkcji (niezależną też od osi energii potencjalnej).

Stany rezonansowe układu studnia-bariera

Modyfikujemy teraz nasz układ. Obszar „2” jest teraz skończony, sięga do $x = b$. Stanowi barierę. Dalej mamy nieskończenie grubą warstwę materiału o dnie pasma przewodnictwa $V_2 < V_1$. Potencjał odczuwany przez rozważany przez nas elektron

$$V(x) = \begin{cases} \infty, & \text{dla } x \leq 0, \\ 0, & \text{dla } 0 \leq x \leq a, \\ V_1, & \text{dla } a < x \leq b, \\ V_2, & \text{dla } x > b \end{cases}$$

przedstawia rys. 33 .



Rys. 33. Potencjał studnia-bariera. Nieskończoną warstwę 2 z rys. 29 ograniczono do odcinka $[a, b]$. Stanowi ona teraz skończoną barierę potencjału, bo dalej (obszar 3) znajduje się nieskończony półprzewodnik o dnie pasma przewodnictwa V_2 powyżej dna pasma przewodnictwa w warstwie studni (obszar 1). Jeśli cząstka o energii powyżej V_2 znajduje się w studni (1), to może wydostać się z niej do obszaru 3 tunelując przez barierę 2. Jej stan nie będzie związany lecz rezonansowy.

Umówmy się, że interesują nas energie tego układu w zakresie $V_2 < E < V_1$. Podobnie jak poprzednio, rozwiązujemy równanie Schroedingera w poszczególnych obszarach. Rozwiązania dla obszarów „0” i „1” są takie jak w poprzednim przypadku. Obszar „2” jest teraz skończony, więc rosnąca funkcja wykładnicza nie stanowi „fizycznego zagrożenia”. Dopuszczalne rozwiązanie dla tego obszaru ma postać

$$\Phi_2(x) = C e^{-k_2 x} + D e^{k_2 x}.$$

Sytuacja w zewnętrznym obszarze, „3”, jest teraz inna niż w poprzednim przypadku – rozważane energie cząstki są powyżej wartości energii potencjalnej w tym obszarze. Nie ma powodu, żeby cząstka nie mogła oddalać się dowolnie daleko. Równanie Schroedingera ma tu postać

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \Phi_2(x) + (V_1 - E) \Phi_2(x) = 0$$

lub inaczej

$$\frac{d^2}{dx^2} \Phi_3(x) + k_3^2 \Phi_3(x) = 0, \text{ gdzie } k_3^2 \equiv \frac{2m(E - V_2)}{\hbar^2}, k_3 > 0.$$

Znów jest to równanie identyczne z równaniem oscylatora harmonicznego i można przyjąć ogólne rozwiązanie w postaci kombinacji liniowej sinusa i cosinusa o częstości kołowej k_3 lub w równoważnej postaci

$$\Phi_3(x) = F e^{-ik_3 x} + G e^{ik_3 x} = F(e^{-ik_3 x} + S e^{ik_3 x}).$$

Ta postać jest bardziej przyjazna, bo ma klarowną interpretację fizyczną. Funkcje $e^{ik_3 x}$ i $e^{-ik_3 x}$ są funkcjami własnymi operatora pędu ($\hat{p} = -i\hbar \frac{d}{dx}$) do wartości własnych, odpowiednio, $k_3 \hbar$ i $-k_3 \hbar$. Zatem pierwsza opisuje cząstkę o dodatnim pędzie, czyli uciekającą od układu studnia-bariera, a druga – cząstkę o ujemnym pędzie, zbliżającą się do bariery. Dlatego $e^{ik_3 x}$ nazywa się falą wychodzącą, a $e^{-ik_3 x}$ – falą wchodzącą. Gdyby rozważać sytuację fizyczną jako strumień cząstek padających na barierę, opisanych falą wchodzącą, rozpraszanych na układzie studnia-bariera, to fala wychodząca opisuje strumień cząstek odbitych od układu. Miarą „natężenia” tych strumieni są odpowiednio $|F|^2$ i $|G|^2$. Ich stosunek $|S|^2$ jest swego rodzaju współczynnikiem odbicia. Dlatego w rozwiązaniu szczególnie interesujący będzie współczynnik S , zwany macierzą rozpraszania.

Zszywanie rozwiązań dla poszczególnych obszarów w ciągłą i gładką globalną funkcję falową wymaga spełnienia formalnych warunków

$$\begin{aligned} \Phi_0(0) &= \Phi_1(0), \\ \Phi_1(a) &= \Phi_2(a), \\ \frac{d\Phi_1}{dx} \Big|_{x=a} &= \frac{d\Phi_2}{dx} \Big|_{x=a}, \\ \Phi_2(b) &= \Phi_3(b), \\ \frac{d\Phi_2}{dx} \Big|_{x=b} &= \frac{d\Phi_3}{dx} \Big|_{x=b}. \end{aligned}$$

Ich jawna postać jest następująca

$$\begin{aligned} 0 &= A \sin 0 + B \cos 0, \\ A \sin k_1 a + B \cos k_1 a &= C e^{-k_2 a} + D e^{k_2 a}, \\ k_1 A \cos k_1 a - k_1 B \sin k_1 a &= -k_2 C e^{-k_2 a} + k_2 D e^{k_2 a}, \\ C e^{-k_2 b} + D e^{k_2 b} &= F(e^{-ik_3 b} + S e^{ik_3 b}), \\ -k_2 C e^{-k_2 b} + k_2 D e^{k_2 b} &= -ik_3 F(e^{-ik_3 b} - S e^{ik_3 b}). \end{aligned}$$

Jest to układ pięciu równań liniowych jednorodnych na sześć niewiadomych: A, B, C, D, F, S . Oprócz oczywistego zerowego rozwiązania, istnieją również inne, których charakter jest jak najbardziej fizyczny. Oznacza to, że nie potrzeba spełniać dodatkowych warunków, czyli dla dowolnej energii E przyjętego do rozważań zakresu istnieją odpowiednie funkcje falowe spełniające równanie Schroedingera. Każda wartość energii $V_2 < E < V_1$ jest dozwolona.

Jednak rozwiązania te mogą zmieniać się w sposób bardzo istotny ze zmianą energii. Dlatego w dalszej części zajmujemy się badaniem zależności rozwiązania od energii, $\Psi(x; E)$, w zakresie $V_2 < E < V_1$. Zamiast rozwiązywać w pełni układ równań, zastanówmy się przez chwilę, co nas najbardziej w rozwiązaniu interesuje. Potem postaramy się znaleźć tylko odpowiedzi na interesujące pytania, pomijając mniej ważne a zajmujące dużo czasu i miejsca szczegóły. Myśląc o zagadnieniu

jako o strumieniu cząstek padających na układ studnia-bariera, odbijających się od bariery lub tunelujących przez nią do wnętrza studni, za najmniej interesującą część rozwiązania trzeba uznać część wewnątrz bariery (obszar 2). Nie będziemy więc poszukiwać współczynników C i D . Rozwiązanie wewnątrz studni zdeterminowane jest przez współczynnik A (bo $B = 0$). Od jego wartości zależy prawdopodobieństwo znalezienia cząstki wewnątrz studni. Wyznaczenie tego prawdopodobieństwa nie jest wcale prostą sprawą, bo funkcje nasze nie dają się unormować do jedynki (nie są „całkowalne z kwadratem”; nie znikają w nieskończoności). Gdyby funkcja stanu była unormowana, to prawdopodobieństwo znalezienia w studni cząstki o energii E obliczalibyśmy według przepisu $P = \int_0^a |\Psi(x; E)|^2 dx$. Przyjmując, że „nieunormowanie jest takie samo” niezależnie od energii, możemy uznać stosunek

$$\frac{\int_0^a |\Psi(x; E_1)|^2 dx}{\int_0^a |\Psi(x; E_2)|^2 dx} \cong \frac{|A(E_1)|^2}{|A(E_2)|^2}$$

za względne prawdopodobieństwo znalezienia cząstki w studni mówiące ile razy bardziej prawdopodobne jest znalezienie cząstki w studni, kiedy ma ona energię E_1 , niż wtedy, kiedy posiada energię o wartości E_2 . (Dalsze przybliżenie w powyższym równaniu, sprowadzenie całości do stosunku kwadratów modułu współczynnika A , wynika z przyjęcia, że zależność $k_1(E)$ jest znacznie mniej gwałtowna niż zależność $A(E)$.)

Na zewnątrz bariery istotną informację o rozwiązaniu zawiera macierz rozpraszania S . Współczynnik F możemy uznać za niezależny, który można zmieniać w sposób dowolny. (Mamy pięć równań na sześć niewiadomych. Rozwiązanie jest niejednoznaczne – można pięć niewiadomych wyznaczyć za pomocą szóstej. Wybór F do tej roli jest dość naturalny, to F określa natężenie strumienia padających cząstek).

Rozwiązujemy więc układ tylko ze względu na A i S . Oto rozwiązanie:

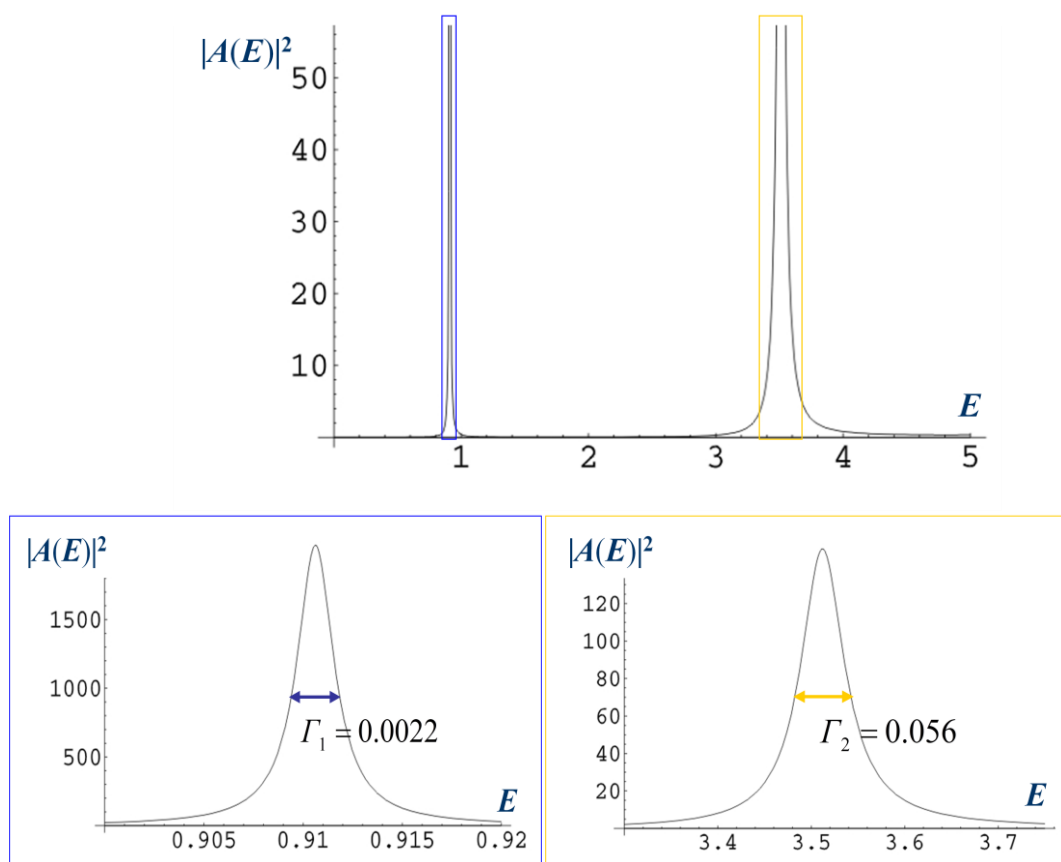
$$A = -F \frac{4ik_2 k_3 e^{2k_2(a+b)} e^{-ik_3 b}}{k_2 \sin k_1 a [e^{2k_2 b} (k_2 - ik_3) - e^{2k_2 a} (k_2 + ik_3)] + k_1 \cos k_1 a [e^{2k_2 b} (k_2 - ik_3) + e^{2k_2 a} (k_2 + ik_3)]}$$

$$S = \frac{k_2 \sin k_1 a [e^{2k_2 a} (k_2 - ik_3) - e^{2k_2 b} (k_2 + ik_3)] - k_1 \cos k_1 a [e^{2k_2 a} (k_2 - ik_3) + e^{2k_2 b} (k_2 + ik_3)]}{k_2 \sin k_1 a [e^{2k_2 b} (k_2 - ik_3) - e^{2k_2 a} (k_2 + ik_3)] + k_1 \cos k_1 a [e^{2k_2 b} (k_2 - ik_3) + e^{2k_2 a} (k_2 + ik_3)]}$$

Współczynniki są zespolone i zapisane tak złożonymi formułami, że trudno jest z nich samych sądzić cokolwiek o zależności od energii. Zobaczymy ją na wykresach. Rozważamy przykład dla konkretnych wartości parametrów układu, takich jak poprzednio, $m = 1$, $a = 2$, $V_1 = 5$ i $\hbar = 1$ oraz dodatkowo $V_2 = 0$ i $b = 3$.

Zaczynamy od wnętrza studni. Ponieważ nie samo A , lecz $|A(E)|^2$ określa względne prawdopodobieństwo znalezienia cząstki w studni, na rys. 34 wykreślamy zależność tej wielkości od energii. W przyjętym do rozważań zakresie energii od $V_2 = 0$ do $V_1 = 5$ mają miejsce dwa wybitne maksima w okolicach wartości energii będących jedynymi dozwolonymi energiami cząstki w studni. Wtedy cząstka mogła znaleźć się w studni mając jedynie energię E_1 lub E_2 . Teraz, w obecności

skończonej bariery, może znaleźć się w studni mając dowolną energię. Jednak prawdopodobieństwo tego jest niezwykle małe prawie dla wszystkich wartości energii. Tylko w sąsiedztwie E_1 i E_2 wrasta wielokrotnie, o kilka rzędów wielkości. Na dolnej części rys. 34 pokazano obydwie maksima bardziej szczegółowo. Ich profile mają identyczny charakter. Ich kształt jest opisany krzywą Lorentza. Profile te różnią się jednak znacznie szerokością mierzoną w połowie wysokości. Jeden, ten w pobliżu niższej energii E_1 , jest bardzo wąski, ma szerokość połówkową $\Gamma_1 = 0.0022$ j. u., a drugi, ten w pobliżu E_2 , ma szerokość znacznie większą, $\Gamma_2 = 0.056$ j. u. Szerokości mają istotne znaczenie fizyczne – są prawdopodobieństwami ucieczki cząstki ze studni w jednostce czasu. (Same wartości nie mają znaczenia, ponieważ przyjęte wartości parametrów układu nie mają szczególnego fizycznego zastosowania. Zwracamy jednak uwagę na wartości, bo poniżej odnajdziemy te same charakterystyki w zgoła innym miejscu).

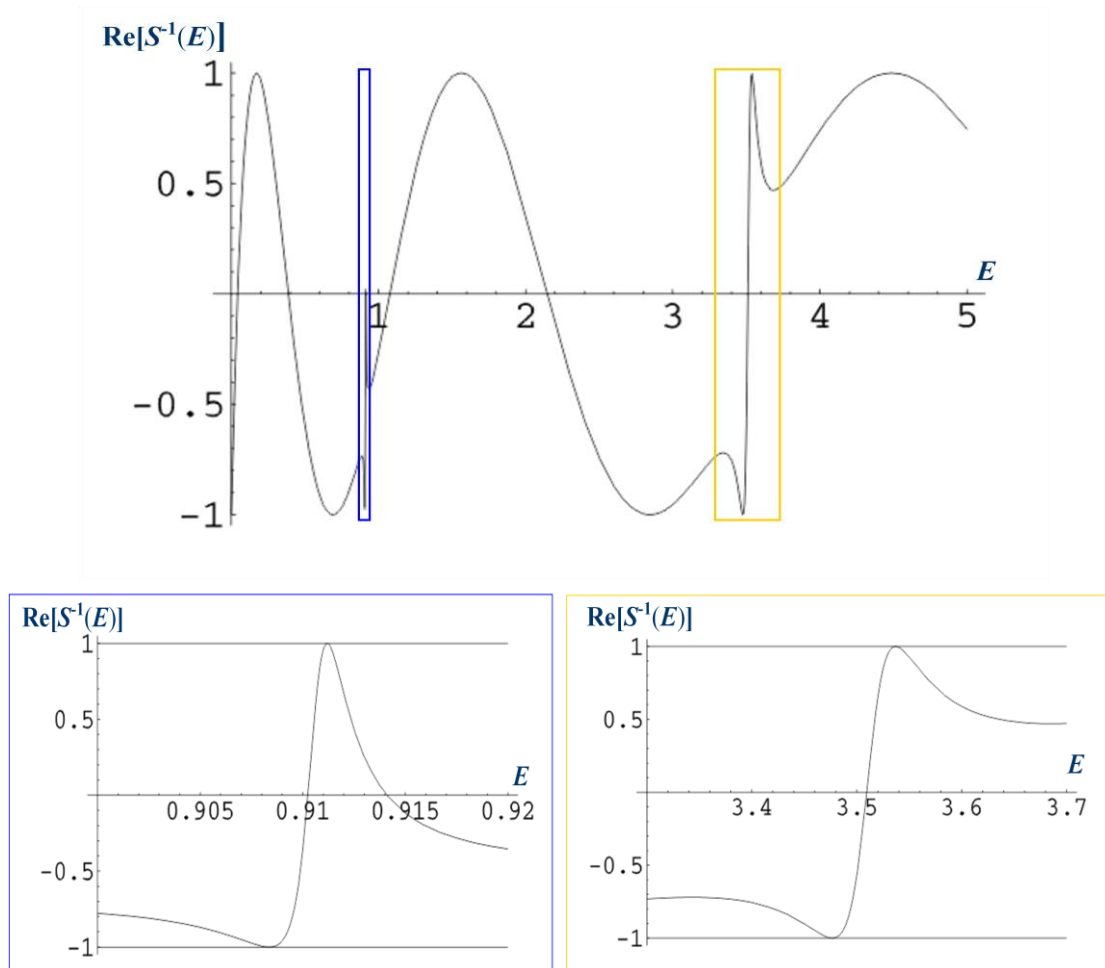


Rys. 34. Wielkość $|A(E)|^2$, do której proporcjonalne jest prawdopodobieństwo znalezienia cząstki w studni. Górny wykres obejmuje cały badany zakres energii. Obszary zaznaczone kolorowymi prostokątami są powiększone w dolnej części.

Najwyraźniej widać, że możliwość uwięzienia cząstki w studni ma charakter rezonansowy. Jego prawdopodobieństwo jest duże tylko w pobliżu energii własnych studni, w skończonych przedziałach o szerokościach określonych przez Γ – analogicznie do drgających układów, które łatwo pobudzić do drgań tylko z częstotliwościami w pobliżu częstotliwości drgań własnych układu.

Jeśli wpadanie cząstki do studni ma charakter rezonansowy, to również odbicie od bariery musi przejawiać gwałtowne rezonansowe zachowanie w pobliżu energii E_1 i E_2 . Rzeczywiście, widać

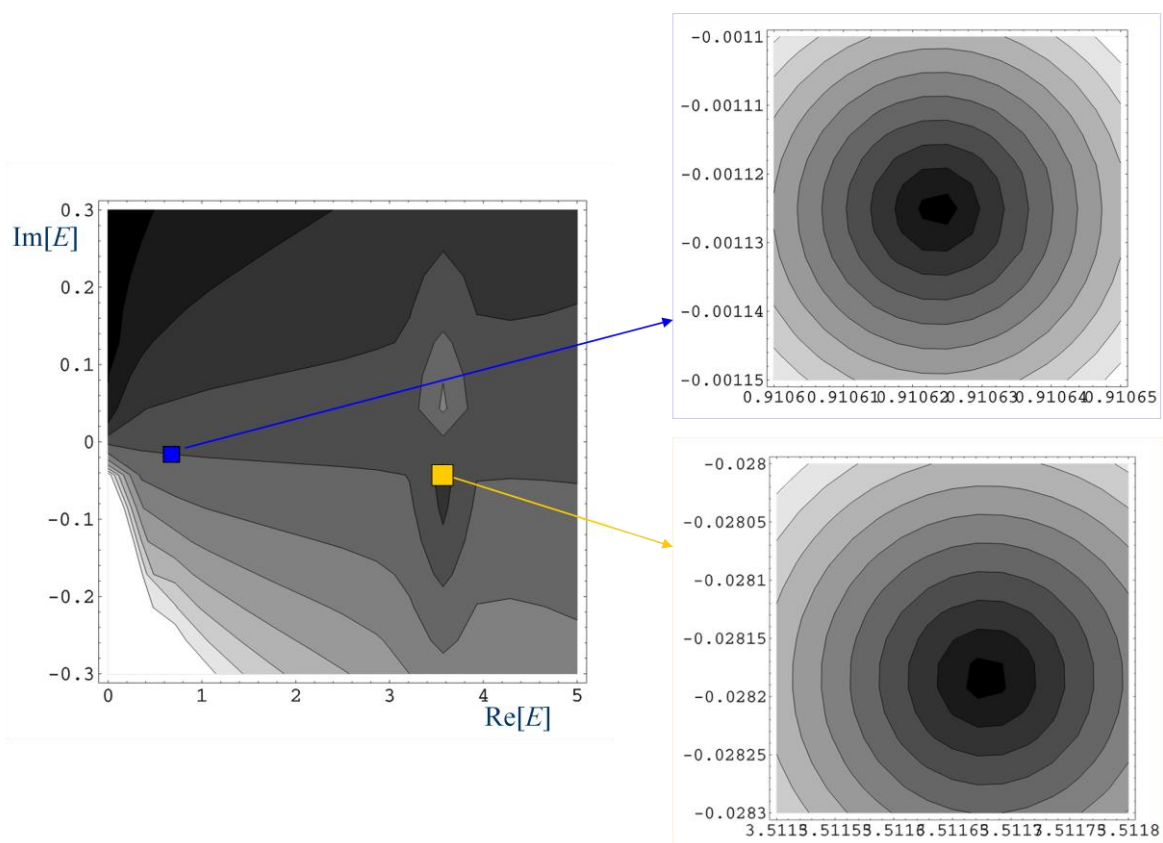
to na rys. 35, gdzie wykreślono zależność od energii części rzeczywistej odwrotności S . Na tle gładkiej, oscylacyjnej zależności pojawiają się dwie gwałtowne struktury w okolicach E_1 i E_2 . Ich powiększenia pokazują, że mają taki sam charakter, ale jedna jest wąska a druga szeroka. Podobne rezonansowe zachowanie będzie ujawniać się w części urojonej, argumente, module S . Zajmiemy się zbadaniem modułu. Zrobimy to w sposób nieco zaskakujący.



Rys. 35. Część rzeczywista odwrotności macierzy rozpraszania, współczynnika S . Górny wykres obejmuje cały badany zakres energii. Obszary zaznaczone kolorowymi prostokątami są powiększone w dolnej części.

Puśćmy wodze fantazji i przyjmijmy, że argument E jest zespolony. Dla matematyka nie stanowi to żadnego problemu. Dla fizyka, przyjęcie, że energia może mieć wartości zespolone jest trudne do zaakceptowania. Tym bardziej, że w kursie podstawowym fizyki kwantowej dowodzi się, iż wartości własne operatorów hermitowskich (jakimi mają być operatory wszystkich obserwowalnych wielkości fizycznych) są rzeczywiste. Tak, ale dotyczy to przestrzeni stanów unormowanych, związanych, nie dotyczy stanów nieograniczonych przestrzennie, jakimi są stany rezonansowe. Nie ma więc przyczyn zasadniczych, żeby wzbraniać się przed zespolonym charakterem E . Są psychologiczne. Ciągłe zaniepokojonych zapewniam, że i dla części rzeczywistej i dla urojonej znajdzie się właściwa interpretacja fizyczna.

Ze względów technicznych, które staną się oczywiste za chwilę, zajmiemy się $|S|^{-1}$. Ponieważ mamy badać zależność od dwuwymiarowego, zespolonego argumentu, posłużymy się wykresem konturowym. Przedstawiamy go na rys. 36. Widać wyraźnie dwa małe obszary, w których wartość $|S|^{-1}$ gwałtownie spada. Leżą one w pobliżu $\text{Re}(E) = E_1$ i $\text{Re}(E) = E_2$, ale wcale nie na osi $\text{Im}(E) = 0$, lecz poniżej. Jedno z minimów ma miejsce dla $\check{E}_1 \cong 0.91062 - i0.001125$, a drugie dla $\check{E}_2 \cong 3.51168 - i0.02818$. Jest dla nas oczywiste, że części rzeczywiste tych szczególnych liczb są bliskie wartościom energii własnych dla studni, $E_1 = 0.91155$ i $E_2 = 3.5005$. Zauważmy, że nie są jednak takie same. Różnice nie wynikają z jakiejś niedokładności, są fizycznym przejawem modyfikacji układu. Mówi się często, że są przesunięciem pochodzącym od kontinuum energii (w układzie studnia-bariera wszystkie energie powyżej $V_2 = 0$ są dozwolone).

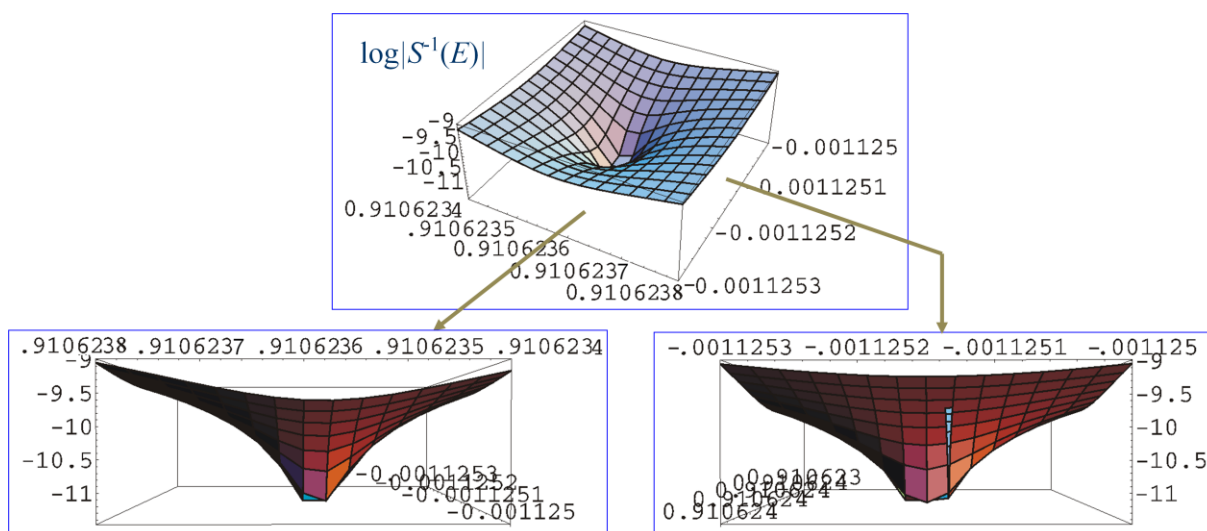


Rys. 36. Odwrotność modułu macierzy rozpraszania $|S|^{-1}$ w zależności od zespolonej zmiennej E . Im ciemniejszy odcień tym mniejsza wartość $|S|^{-1}$. Wykres po lewej stronie obejmuje cały badany zakres energii, $V_2 = 0 < \text{Re}(E) < V_1 = 5$. Obszary zaznaczone kolorowymi prostokątami są powiększone po prawej stronie.

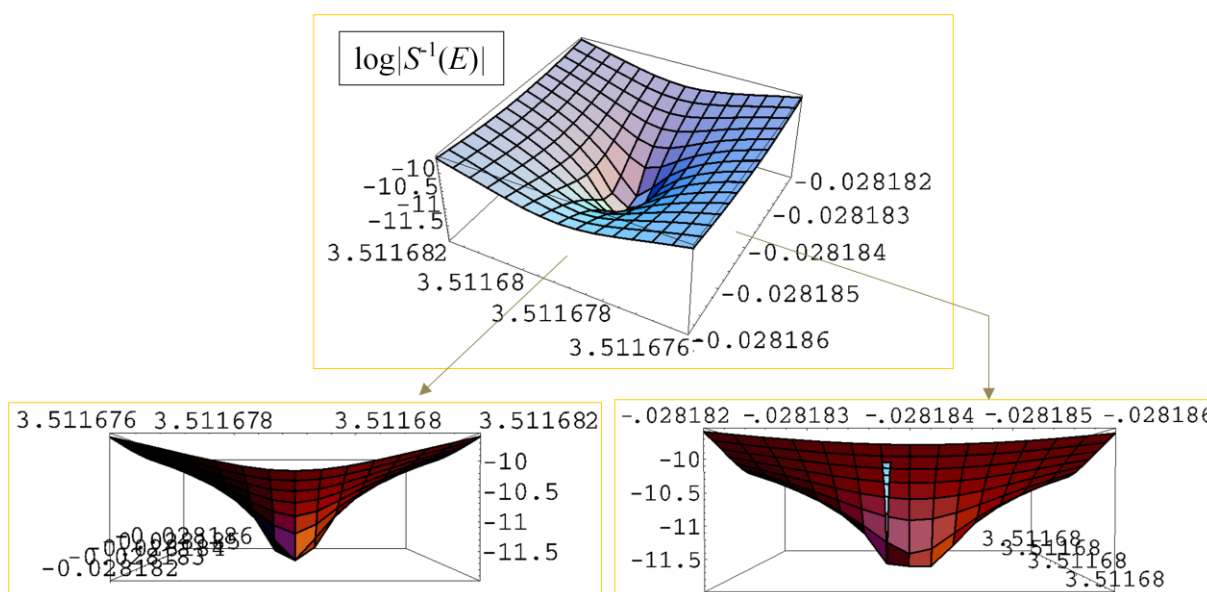
A co z częściami urojonymi? Zauważmy, że jeśli je podwoić, to otrzymujemy wartości szerokości półwkowych otrzymanych dla pików $|A(E)|^2$. Znaleźliśmy więc pewne szczególne zespolone wartości własne równania Schroedingera, których część rzeczywista odpowiada energetycznemu położeniu rezonansu, E_r , a część urojona – połowie jego szerokości półwkowej, Γ : $\check{E} = E_r - i\frac{\Gamma}{2}$.

Jak głębokie są owe minima widoczne na rys. 36? Czy sięgają wartości 0? Żeby to sprawdzić posługujemy się wykresami 3D na rys. 37 i 38. Przedstawiają one $\log|S^{-1}(E)|$ odpowiednio dla

niższego i wyższego energetycznie rezonansu. Podgląd z różnych stron pozwala lepiej odczytać położenie minimum. Na tych rysunkach minima osiągają wartość rzędu 10^{-11} . Zejście dowolnie blisko zera jest kwestią cierpliwości badacza. Rozważane przez nas minima są punktami, w których $S^{-1} = 0$. (Dlatego łatwiej było wykreślać S^{-1} zamiast S). Oznacza to, że w tych szczególnych punktach funkcja falowa nie zawiera fali wchodzącej. Jest tylko fala wychodząca. Są to szczególne warunki asymptotyczne, które można by przyjąć za definicję stanów rezonansowych. Są one stanami pseudozwiązanymi, które mogą się rozpaść. Zawierają więc tylko falę wychodzącą, która taki rozpad opisuje – opisuje uciekającą z układu cząstkę.



Rys. 37. $\log|S^{-1}|$ w zależności od zespolonej zmiennej E w pobliżu niższego energetycznie rezonansu rozważanego układu studnia-bariera; na dole oddzielnie widok od strony osi rzeczywistej (po lewej) i od strony osi urojonej (po prawej)



Rys. 38. $\log|S^{-1}|$ w zależności od zespolonej zmiennej E w pobliżu wyższego energetycznie rezonansu rozważanego układu studnia-bariera; na dole oddzielnie widok od strony osi rzeczywistej (po lewej) i od strony osi urojonej (po prawej)

Połączenie dwóch parametrów rezonansu, energetycznego położenia i szerokości, w zespolonej „energii” $\tilde{E} = E_r - i\frac{\Gamma}{2}$ ma szczególny sens w kontekście ewolucji stanu, czyli – w naszym przypadku – rozpadu układu rezonansowego. W pierwszym rozdziale podaliśmy postulaty mechaniki kwantowej. Według jednego z nich, ewolucja każdego stanu odbywa się zgodnie z równaniem Schroedingera zależnym od czasu. Szczególnym przypadkiem są stany własne hamiltonianu niezależnego od czasu. Ewoluuja one zgodnie z

$$\psi(x, t) = e^{-\frac{i}{\hbar}Et}\Psi(x), \text{ gdzie } \hat{H}\Psi(x) = E\Psi(x).$$

Wtedy gęstość prawdopodobieństwa $|\psi(x, t)|^2$ dla stanu rezonansowego zmienia się w czasie wykładniczo

$$|\psi(x, t)|^2 = \left| e^{-\frac{i}{\hbar}(E_r - i\Gamma/2)t}\Psi(x) \right|^2 = e^{-\frac{\Gamma}{\hbar}t} |\Psi(x)|^2.$$

Zatem w stanie rezonansowym prawdopodobieństwo znalezienia cząstki w studni maleje wykładniczo. Po czasie $\tau = \frac{\hbar}{\Gamma}$, zwanym średnim czasem życia, maleje e -krotnie.

Przybliżone metody znajdowania położenia i szerokości stanów rezonansowych

Proponujemy tu dwie metody o charakterze wariacyjnym polegające na rozwinięciu poszukiwanej przybliżonej funkcji falowej stanu rezonansowego w bazie zadanych funkcji unormowanych i optymalizacji względem dodatkowych (nieliniowych), specyficznych dla metody parametrów. Z punktu widzenia praktycznych obliczeń, największym problemem jest występowanie rozmytego poziomu rezonansowego na tle ciągłego widma energii odpowiadających stanom niezwiązanym. W związku z tym obliczenia wariacyjne polegające na rozwinięciu funkcji rezonansowej w bazie dają zestaw energii, wśród których trudno zidentyfikować wartość odpowiadającą stanowi rezonansowemu. Drugim problemem jest nienormowalność ściślej funkcji rezonansowej, co czyni jej reprezentację w bazie funkcji unormowanych istotnym przybliżeniem.

Metoda stabilizacji pozwala na drodze szczególnej optymalizacji wyłonić rezonansowy poziom energetyczny spośród energii odpowiadających stanom niezwiązanym. Jest jednak obciążona błędem ograniczenia poszukiwań funkcji rezonansowej do rozwinięcia w bazie funkcji całkownych z kwadratem. Metoda obrotu zespolonego pozwala znaleźć obydwie parametry rezonansu, położenie energetyczne i szerokość, jako rzeczywistą i urojoną część szczególnej zespolonej wartości własnej. Posługuje się transformacją przeprowadzającą funkcję rezonansową w funkcję całkowną z kwadratem (dającą się unormować). Dzięki temu stosowanie baz funkcji całkownych z kwadratem jest w pełni uzasadnione.

Metoda stabilizacji

Zakładamy, że dana jest baza N funkcji $\{\varphi_i(\vec{r})\}_{i=1}^N$. Są one unormowane i określone jednoznacznie. Jeśli zastosować tę bazę do rozwinięcia poszukiwanej funkcji rezonansowej, to

jedynymi parametrami wariacyjnymi – umożliwiającymi dopasowanie rozwinięcia tak, aby równanie Schroedingera było spełnione z najlepszą możliwą dokładnością – byłyby współczynniki rozwinięcia (parametry liniowe). Wprowadzimy dodatkowo jeden parametr nieliniowy, wspólny dla wszystkich funkcji bazowych, manipulowanie którym (wariowanie go) umożliwi nam identyfikację rozwiązań dla stanów rezonansowych. Będzie to parametr skalowania współrzędnych, $\vec{r} \rightarrow \alpha\vec{r}$. Wprowadzimy te przeskalowane współrzędne do funkcji bazowych otrzymując bazę funkcji w dodatkowym parametrem α , $\{\varphi_i^\alpha(\vec{r}) \stackrel{\text{def}}{=} \varphi_i(\alpha\vec{r})\}_{i=1}^N$. Dopiero w takiej bazie rozwijamy poszukiwaną funkcję rezonansową :

$$\psi(\vec{r}; c_1, \dots, c_N, \alpha) = \sum_{i=1}^N c_i \varphi_i^\alpha(\vec{r}).$$

Energia odpowiadająca funkcji ψ znajdujemy jako wartość oczekiwaną hamiltonianu:

$$E(c_1, \dots, c_N, \alpha) = \frac{\langle \psi(\vec{r}; c_1, \dots, c_N, \alpha) | \hat{H} | \psi(\vec{r}; c_1, \dots, c_N, \alpha) \rangle}{\langle \psi(\vec{r}; c_1, \dots, c_N, \alpha) | \psi(\vec{r}; c_1, \dots, c_N, \alpha) \rangle}.$$

Optymalna z punktu widzenia spełnienia równania Schroedingera jest ta funkcja $\psi(\vec{r}; c_1, \dots, c_N, \alpha)$, dla której

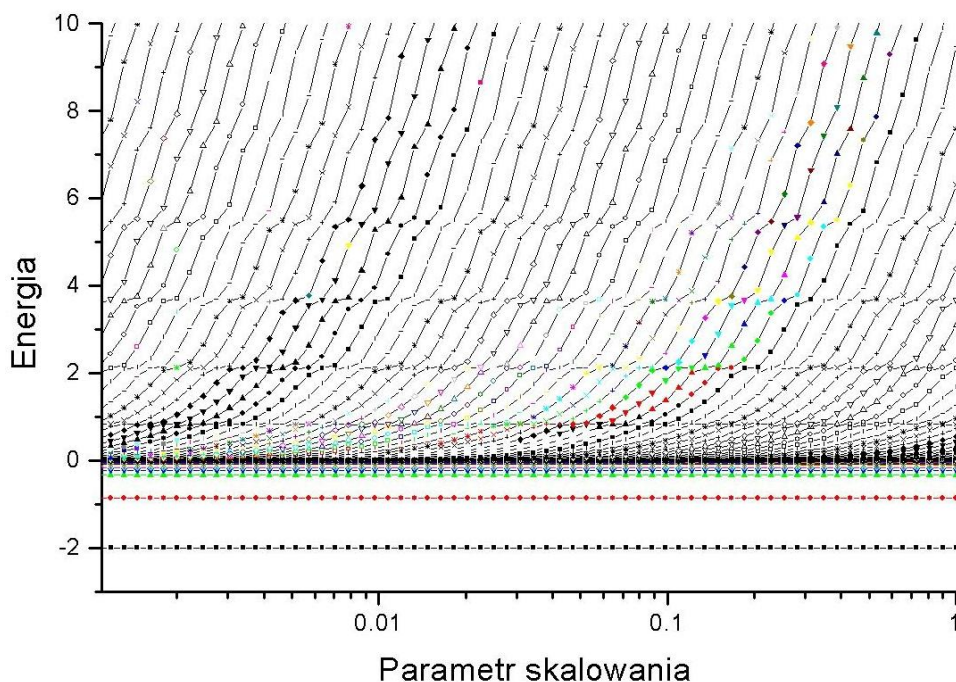
$$\left. \frac{\partial E(c_1, \dots, c_N, \alpha)}{\partial c_i} \right|_{c_i^{opt}} = 0 \quad \text{i} \quad \left. \frac{\partial E(c_1, \dots, c_N, \alpha)}{\partial \alpha} \right|_{\alpha^{opt}} = 0.$$

Ustalmy na razie wartość α i zajmijmy się pierwszym warunkiem. Ponieważ zależność od współczynników c_i jest jednoznacznie określona w sposób analityczny, warunek ten prowadzi po krótkich obliczeniach do równania macierzowego

$$\mathbf{H}(\alpha) \mathbf{C}^{opt}(\alpha) = E(\alpha) \mathbf{S}(\alpha) \mathbf{C}^{opt}(\alpha),$$

gdzie $\mathbf{C}^{opt}(\alpha)$ jest kolumnowym wektorem, którego składowymi są c_i^{opt} , $\mathbf{H}(\alpha)$ jest tzw. macierzą hamiltonianu zbudowaną z elementów $H_{ij}(\alpha) = \langle \varphi_i^\alpha | \hat{H} | \varphi_j^\alpha \rangle$ a $\mathbf{S}(\alpha)$ – macierzą iloczynów skalarnych funkcji bazowych $S_{ij}(\alpha) = \langle \varphi_i^\alpha | \varphi_j^\alpha \rangle$. Rozwiązujemy więc algebraiczny problem (uogólnione równanie własne) dla macierzy kwadratowych o wymiarze $N \times N$. Otrzymamy N rozwiązań – N rzeczywistych wartości własnych $E_k(\alpha)$. Wśród nich, niektóre mogą odpowiadać stanom związanym, liczne stanom niezwiązanym i pewne stanom rezonansowym. Nas interesują te ostatnie, ale nie potrafimy odróżnić ich od energii stanów niezwiązanym.

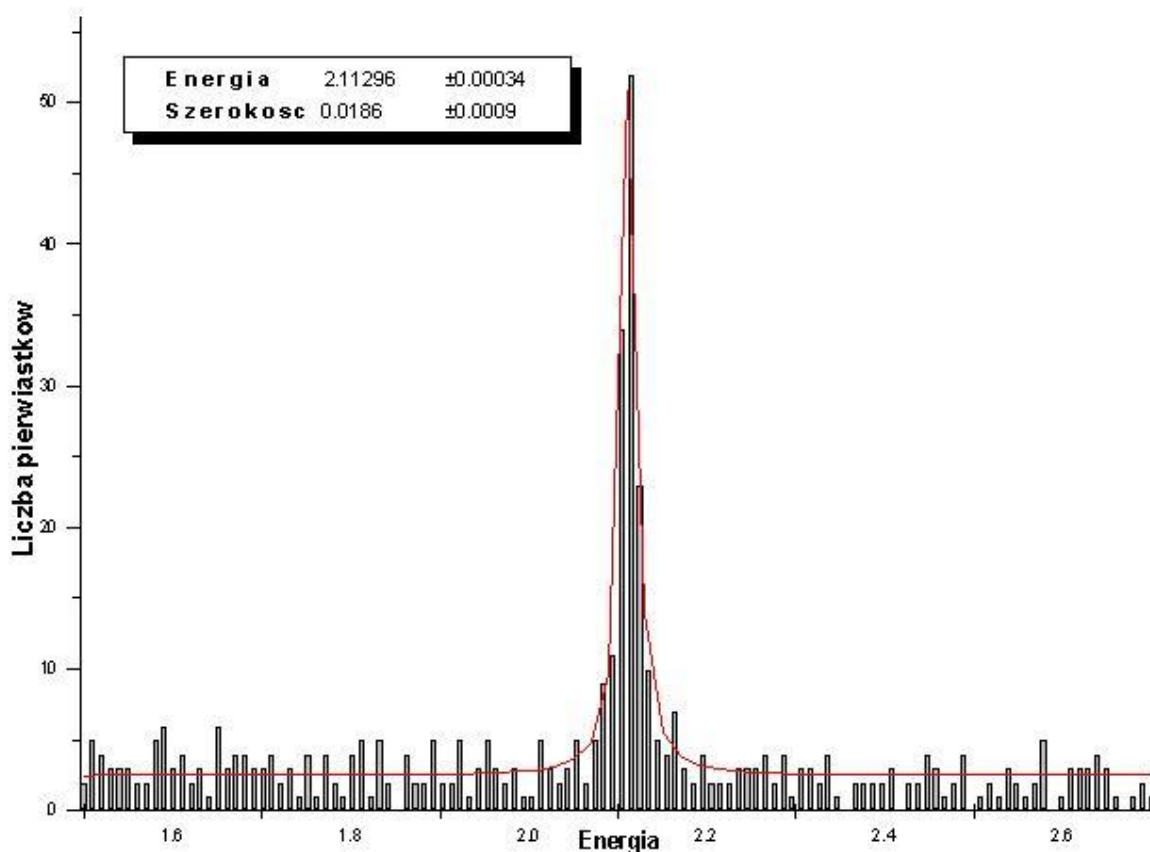
Zajmujemy się teraz poszukiwaniem optymalnej wartości parametru skalowania współrzędnych, α . Nie jesteśmy w stanie określić analitycznie zależności rozwiązań $E_k(\alpha)$ od α . Ustalamy więc punkt optymalny, dla którego $\frac{\partial E_k(\alpha)}{\partial \alpha} = 0$, na drodze numerycznej: 1) wykonujemy serię obliczeń dla wielu wartości α , które zmieniamy w sposób systematyczny, 2) wartości własne $E_k(\alpha)$ otrzymane dla danej wartości α porządkujemy od najmniejszej do największej, 3) wykreślamy zależność kolejnych pierwiastków E_k od α . Przykładowy obraz, otrzymany w taki sposób, pokazany jest na rysunku 39. W zasadzie, w miarę rozszerzania się funkcji na skutek skalowania $\alpha\vec{r}$,



Rys. 39. Stabilizacja energii pewnego układu w zależności od parametru skalowania współrzędnych, α . Poziome linie dla energii poniżej zera odpowiadają stanom związanym. (Są poziome, bo baza doskonale je odtwarza, niezależnie od parametru α). Obszar widma ciągłego energii rozciąga się powyżej zera. W tym zakresie, każdy z pierwiastków rośnie wraz ze wzrostem α . Jednak stabilizuje się na pewnych poziomach (odpowiadających rezonansom) w niewielkim przedziale wartości α , do momentu aż następny pierwiastek dotrze do tego poziomu i przejmie rolę reprezentowania energii stanu rezonansowego.

wszystkie E_k maleją monotonicznie zmierzając ku progowi, w którym zaczyna się widmo ciągłe energii. Jednak na pewnych poziomach stabilizują się. Te poziomy energetyczne przypisujemy stanom rezonansowym. Ich wartość można z pewną (ograniczoną) dokładnością odczytać z wykresu lub otrzymać numerycznie z równania $\frac{dE_k(\alpha)}{d\alpha} = 0$. Można też zastosować opisaną niżej procedurę, która pozwala oprócz położenia poziomu energetycznego określić także jego szerokość.

Systematyczne skalowanie współrzędnych pociąga za sobą systematyczne zmiany wartości własnych. Jest to swego rodzaju przeglądanie kontinuum dozwolonych energii. Jeśli liczba wartości parametru α i liczba funkcji bazowych, N , są dość duże, to otrzymujemy liczny zbiór wartości własnych, których gęstość względem energii reprezentuje dość dobrze gęstość stanów badanego układu. W praktyce reprezentujemy ją za pomocą histogramu – dzielimy rozpatrywany przedział energii (w którym poszukujemy stanów rezonansowych) na małe przedziały o jednakowej szerokości, zliczamy ile wartości własnych przypada na dany przedział (dla wszystkich wartości α). Jest dość oczywiste, że w tak przygotowanym histogramie, pojawią się maksima w pobliżu poziomów stabilizacji wartości własnych, tj. w pobliżu poziomów rezonansowych. Na rys. 40 przedstawiamy przykład takiego maksimum odpowiadającego poziomowi około 2.1 umownych jednostek na Rys.39.



Rys. 40. Gęstość stanów w zakresie widma ciągłego energii w okolicy jednego z rezonansów (Rys.39), reprezentowana przez rozkład pierwiastków pochodzących z obliczeń metodą stabilizacji. Czerwona krzywa jest dopasowaną do rozkładu pierwiastków krzywą Lorentza. Parametrami dopasowania są położenie i szerokość rezonansu podane na rysunku.

Takie maksima odpowiadają w istocie rzeczy maksimum pokazanym na Rys. 34. Mają rezonansowy charakter, który można opisać krzywą Lorentza. Dlatego, do rozkładu pierwiastków w pobliżu wybranego stanu rezonansowego dopasowujemy krzywą Lorentza

$$\rho(E) = \rho_0 + \frac{D}{(E - E_r)^2 + \frac{1}{4}\Gamma^2}.$$

Spśród czterech parametrów dopasowania (ρ_0, D, E_r, Γ), dwa mają istotne znaczenie dla określenia rezonansu. Są to położenie poziomu rezonansowego, E_r , i jego szerokość Γ . Ostatecznie wartości tych parametrów otrzymane w drodze dopasowania krzywej Lorentza do rezonansowej struktury w gęstości stanów (w rozkładzie pierwiastków) uznajemy za najbardziej wiarygodne wartości parametrów stanu rezonansowego jakie można otrzymać metodą stabilizacji.

Metoda obrotu zespolonego współrzędnych

Metoda obrotu zespolonego współrzędnych jest podobna do metody stabilizacji. Stosuje się w niej skalowanie współrzędnych za pomocą zespolonego czynnika $\alpha = e^{i\theta}$, gdzie $\theta > 0$ jest parametrem rzeczywistym. Transformacja $\vec{r} \rightarrow e^{i\theta} \vec{r}$ nazywana jest obrotem zespolonym

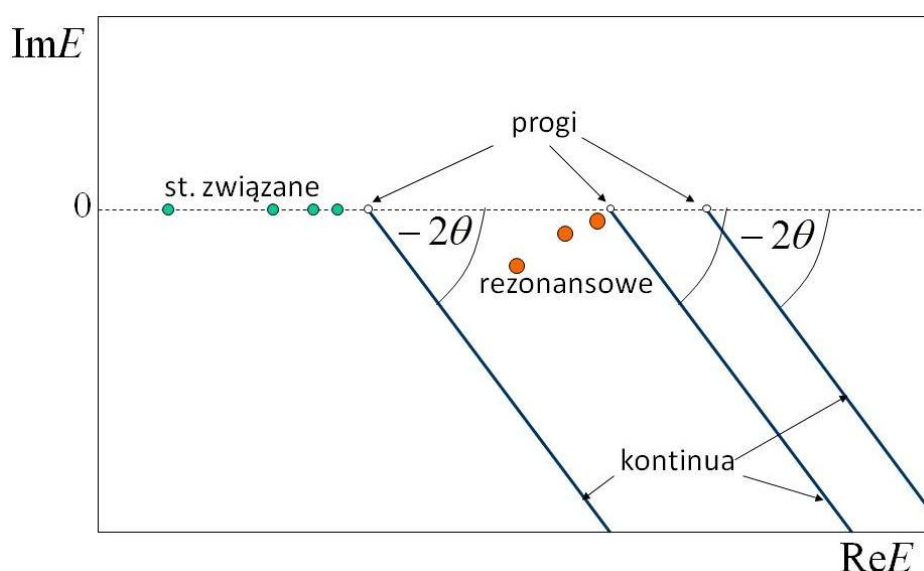
współrzędnych lub skalowaniem zespolonym. Stąd też metoda nazywana jest czasami skalowaniem zespolonym współrzędnych.

Metoda występuje w kilku odmianach. W najprostszej wersji, transformacja obrotu jest stosowana do współrzędnych w hamiltonianie, a nie w funkcjach. Udowodniono matematycznie, że widmo wartości własnych przetransformowanego operatora Hamiltona, $\hat{H}(e^{i\theta}\vec{r})$ jest w ścisłych relacjach z widmem oryginalnego operatora i można jego składowe interpretować fizycznie. W szczególności:

- 1) Dyskretne wartości własne oryginalnego operatora $\hat{H}(\vec{r})$, odpowiadające stanom związanym, nie zmieniają się na skutek obrotu zespolonego współrzędnych.
- 2) Widmo ciągłe energii odpowiadających stanom niezwiązanym obraca się względem progu widma ciągłego na płaszczyznę zespoloną o kąt -2θ .
- 3) Energie stanów rezonansowych są zespolone, odizolowane od widma ciągłego i w znacznym zakresie niezależne od kąta obrotu θ . Mają ujemną część urojoną.

Własności te zilustrowano na Rys. 41.

Oczywistą zaletą tej metody jest znajdowanie obu parametrów rezonansowego poziomu energetycznego w jednej liczbie zespolonej, $\check{E} = E_r - i\frac{\Gamma}{2}$, którą stosunkowo łatwo można odróżnić od zespolonych energii odpowiadających stanom niezwiązanym. Drugą – bardzo poważną – zaletą metody obrotu zespolonego współrzędnych jest fakt, że transformacja obrotu czyni rezonansową funkcję falową całkowalną z kwadratem. Ograniczenie się do stosowania bazy funkcji całkowalnych z kwadratem (wygodnych z powodów praktycznych) nie stanowi w tym przypadku przybliżenia (w przeciwieństwie do metody stabilizacji).



Rys. 41. Schemat przedstawiający widmo wartości własnych obróconego na sposób zespolony hamiltonianu. Energie odpowiadające stanom związanym i progom (początkom) widma ciągłego pozostają niezmiennic przez obrót. Półproste widma ciągłego są obrócone wokół progu widma o kąt 2θ na dolną półpłaszczyznę zespoloną. Wartości własne odpowiadające rezonansom pojawiają się na dolnej półpłaszczyźnie i są odseparowane od widma ciągłego.

W praktyce, z powodu skończonej przestrzeni (skończona baza), w której poszukujemy funkcji stanu rezonansowego, przybliżona energia rezonansu zależy w nieznaczny sposób od θ . Dlatego kąt obrotu traktujemy również jako parametr wariacyjny. Przepis na realizację obliczeń metodą obrotu zespolonego jest następujący:

- 1) Dokonujemy obrotu zespolonego współrzędnych w Hamiltonianie – otrzymujemy $\hat{H}(e^{i\theta}\vec{r})$.
- 2) Poszukujemy rezonansowego rozwiązania równania własnego operatora $\hat{H}(e^{i\theta}\vec{r})$ w postaci rozwinięcia $\psi(\vec{r}; c_1, \dots, c_N) = \sum_{i=1}^N c_i \varphi_i(\vec{r})$, gdzie φ_i są zadanymi funkcjami bazowymi.
- 3) Energię odpowiadającą takiemu rozwinięciu wyraża wartość oczekiwana

$$E(c_1, \dots, c_N, \theta) = \frac{\langle \psi(\vec{r}; c_1, \dots, c_N) | \hat{H}(e^{i\theta}\vec{r}) | \psi(\vec{r}; c_1, \dots, c_N) \rangle}{\langle \psi(\vec{r}; c_1, \dots, c_N) | \psi(\vec{r}; c_1, \dots, c_N) \rangle}.$$

- 4) Optymalne zestawy współczynników c_i i odpowiadające tym zestawom energie znajdujemy rozwiązując równanie macierzowe

$$\mathbf{H}(\theta)\mathbf{C}^{\text{opt}}(\theta) = E(\theta)\mathbf{S}\mathbf{C}^{\text{opt}}(\theta),$$

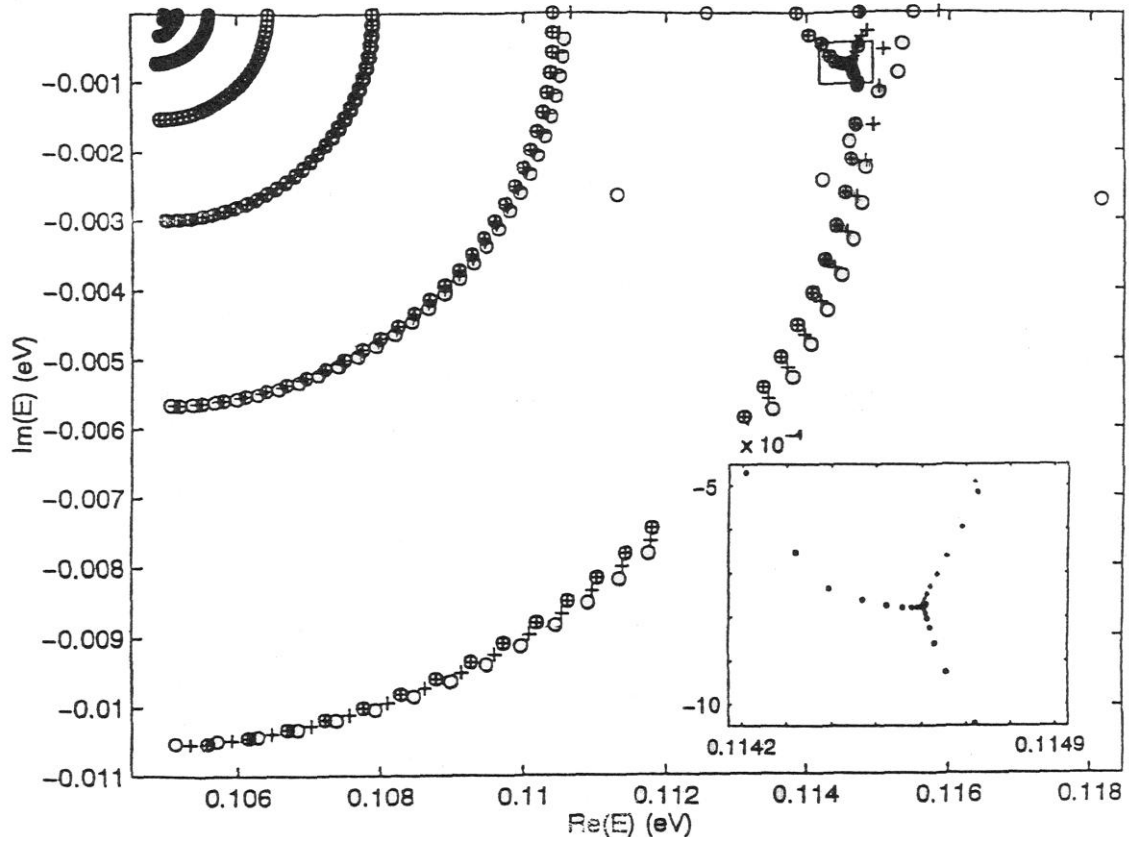
gdzie $\mathbf{C}^{\text{opt}}(\theta)$ jest kolumnowym wektorem, którego składowymi są $c_i^{\text{opt}}(\theta)$,

$\mathbf{H}(\theta)$ jest tzw. macierzą obróconego hamiltonianu zbudowaną z elementów

$H_{ij}(\theta) = \langle \varphi_i | \hat{H}(e^{i\theta}\vec{r}) | \varphi_j \rangle$, a \mathbf{S} – macierzą iloczynów skalarnych funkcji bazowych $S_{ij} = \langle \varphi_i | \varphi_j \rangle$. Ponieważ macierz $\mathbf{H}(\theta)$ nie jest hermitowska, otrzymujemy zespolone rozwiązania $E(\theta)$.

- 5) Część z nich powinna zależeć od θ w sposób właściwy dla energii stanów niezwiązanych. Tylko wybrane zespolone wartości własne, te odpowiadające stanom rezonansowym powinny być niezależne od θ . Powtarzamy rozwiązywanie równania z punktu 4) dla różnych wartości θ , żeby znaleźć optymalne energie rezonansowe z warunku $\left. \frac{dE(\theta)}{d\theta} \right|_{\theta^{\text{opt}}} = 0$. Oznacza to, że spośród wszystkich wartości własnych otrzymywanych w punkcie 4) wybieramy tę, która jest stabilna przy zmianie θ .
- 6) Interpretujemy część rzeczywistą takiej wartości własnej jako położenie energetyczne rezonansu, a część urojoną jako połowę jego szerokości energetycznej.

Przykładowe wyniki otrzymane metodą obrotu zespolonego przedstawia Rys. 42. Wszystkie wartości własne o części rzeczywistej w badanym zakresie energii umieszczono na płaszczyźnie zespolonej. Wyniki odpowiadają trzem nieco różnym bazom i trzydziestu wartościom θ zmieniającym się w równych odstępach w zakresie od 0 do $\frac{\pi}{4}$. Ujawniają się wyraźnie dwa typy θ -trajektorii (torów po których „porusza” się wybrana wartość własna ze zmianą kąta obrotu). Trajektorie w postaci charakterystycznych łuków odpowiadają widmu ciągłemu obracającemu się regularnie w miarę wzrostu kąta obrotu współrzędnych. Wyraźnie od nich różnią się trzy θ -trajektorie (po jednej dla każdej bazy) wykazujące znacznie wolniejszą zależność od θ . Te trzy θ -trajektorie, choć startują dla $\theta = 0$ w różnych punktach na osi rzeczywistej, zbiegają się we wspólnym punkcie, w którym stabilizuje się wartość własna (najwolniej zależy od kąta obrotu). Ten punkt należy uznać za optymalną wartość zespolonej energii rezonansu.



Rys. 42. θ -trajektorie otrzymane dla nanostruktury w postaci symetrycznego układu warstw półprzewodnikowych, dwóch barier i studni [M. Bylicki, R. Oszwałdowski, W. Jaskólski, *J. Phys.: Condensed Matter* 8 (1996) 6393]. Obszar zbieżności trajektorii rezonansowych, zaznaczony prostokątem na głównym rysunku, jest powiększony we wstawce.

Wprowadzenie do problemu odkształceń na przykładzie samorosnących kropek kwantowych

Michał Zieliński

Samorosnące kropki kwantowe

Jedną z najważniejszych metod wytwarzania nanostruktur jest tzw. epitaksja z wiązek molekularnych (MBE, ang. molecular beam epitaxy). W metodzie tej, w warunkach ultra-wysokiej próżni, na odpowiednio przygotowane podłoże krystaliczne, nanoszone są kolejne warstwy atomów lub molekuł (Rys.43). Źródłem strumienia atomów różnych pierwiastków lub molekuł (stąd nazwa MBE) są specjalne piecyki o stabilizowanej temperaturze, tzw. komórki efuzyjne (zwane też komórkami Knudsen). Proces MBE w połączeniu z, na przykład, spektroskopią RHEED (dyfrakcja wysokoenergetycznych elektronów, ang. Reflection High-Energy Electron Diffraction) umożliwia kontrolę wzrostu kolejnych warstw z dokładności do pojedynczych odległości międzyatomowych.

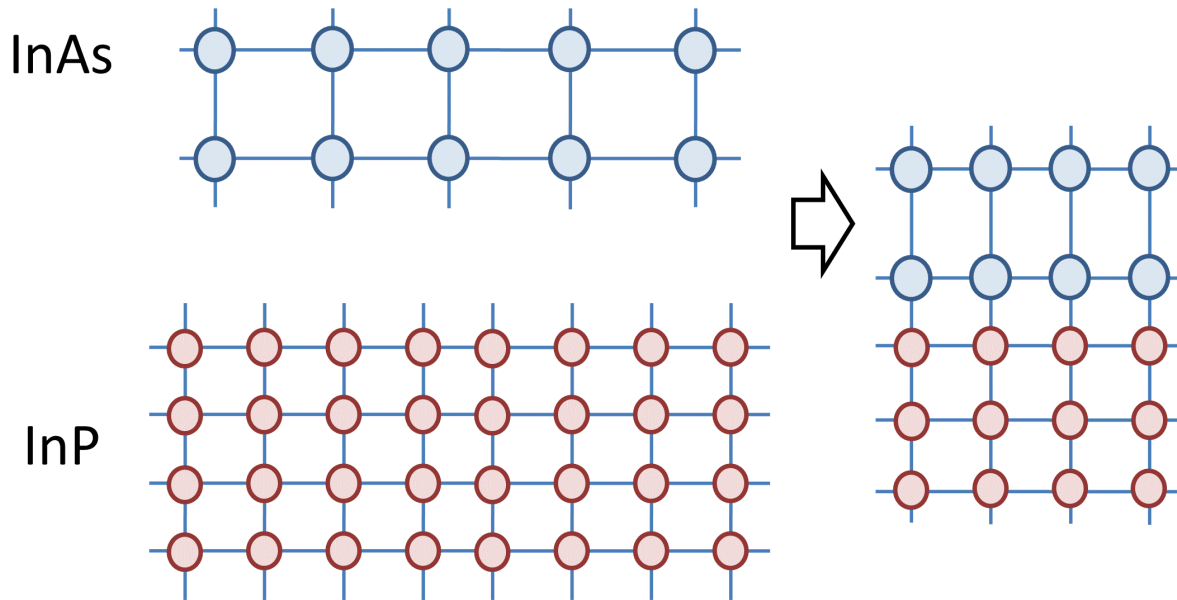


Rys.43. Wygląd zewnętrzny komory MBE (źródło: Internet).

W procesie wzrostu można zmienić skład chemiczny reagentów, np. poprzez zamknięcie i otwarcie odpowiednich komórek efuzyjnych. I tak, przykładowo, po osadzeniu kilkunastu monowarstw (ang. monolayers) arsenku galu (GaAs) można nanieść kolejne warstwy innego materiału, np. arsenku aluminium (AlAs). Jako, że materiały osadzone oraz podłoża różnią się właściwościami elektronowymi, w szczególności szerokością przerwy energii wzbronionej, to opisany proces zmiany składników umożliwia wytwarzanie studni kwantowych oraz tzw. supersieci.

Nie wszystkie materiały półprzewodnikowe można ze sobą łączyć, tak aby obszar złącza wolny był od dyslokacji oraz innych defektów struktury krystalicznej. W szczególności stałe sieci krystalicznych (związane z długością wiązań międzyatomowych) materiałów podłoża oraz osadzanego nie mogą zbyt różnić się od siebie. Dobrym przykładem są tutaj GaAs oraz AlAs, których stałe sieci różnią się o mniej niż 0.2%.

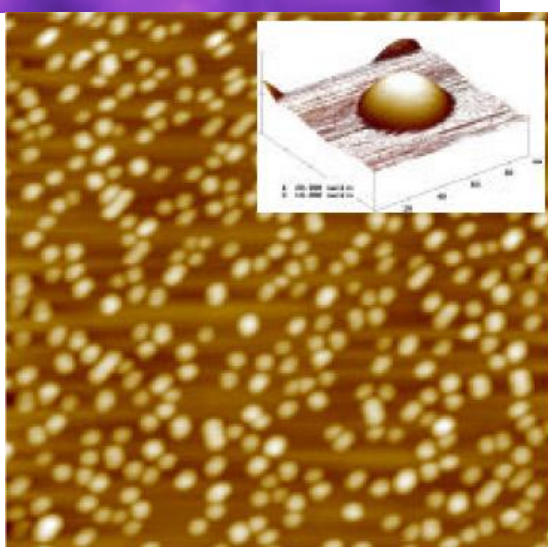
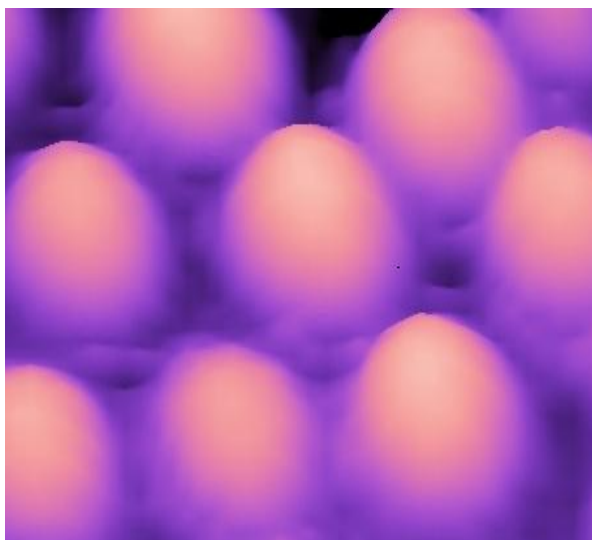
Bardzo ciekawa sytuacja ma miejsce w przypadku często stosowanych kombinacji materiałów półprzewodnikowych np. dla InAs/InP różnica stałych sieci wynosi około 3%, natomiast dla InAs/GaAs jest to aż 7%. Ze względu na wspomnianą różnicę stałych sieci, nanoszona warstwa oraz podłoże muszą się w jakiś sposób „dopasować” się (rys. 44).



Rys.44 Schemat ilustrujący dopasowanie stałej sieci epitaksjalnie nanoszonej warstwy do stałej sieci podłoża.

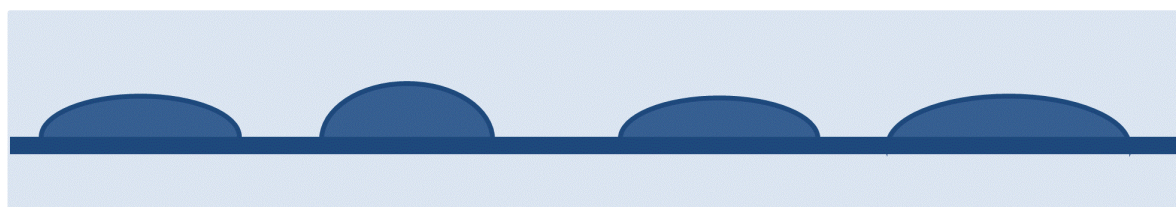
Dla materiałów różniących się istotnie stałą sieci, w procesie tzw. wzrostu Stranskiego-Krastanowa, pierwsza z osadzanych warstw (tzw. warstwa zwilżająca) dopasowuje swoją stałą sieci w płaszczyźnie złącza do stałej materiału podłoża (jest to tzw. pseudomorfizm). Dla materiałów widocznych na rysunku schematycznym oznacza to, że materiał warstwy (InAs) o większej stałej sieci krystalicznej, jest ściśnięty w płaszczyźnie złącza, a towarzyszy temu jednocześnie zwiększenie grubości warstwy (rozciągnięcie) w kierunku wzrostu, podczas gdy podłoże pozostaje praktycznie nie odkształcone. W takiej sytuacji mówimy, że warstwa ulega silnej deformacji lub odkształceniu, gdyż zmianie ulegają odległości międzyatomowe w stosunku do odległości charakterystycznych dla nieodkształconego kryształu objętościowego. Bardziej ilościowy opis tego zjawiska znajdziemy w dalszej części wykładu.

Kontynuacja procesu jednorodnego wzrostu materiałów niedopasowanych sieciowo doprowadziłaby do silnego odkształcenia nowo rosnącej warstwy. W rzeczywistości bardziej energetycznie korzystna jest sytuacja w której energia odkształceń minimalizowana jest w procesie wzrostu poprzez powstanie niejednorodnych zgrubień „wysp” nanoszonego materiału (Rys 45). Powstałe „wyspy” mają rozmiary rzędu nanometrów (średnica 10 -30 nanometrów; wysokość 1-5 nanometrów) oraz rozmieszczone są w nieregularny, przypadkowy sposób na powierzchni podłoża. Charakteryzują się jednak dość podobnym kształtem przypominającym soczewkę, a ich parametry takie jak skład chemiczny, rozmiar, czy gęstość rozmieszczenia na podłożu mogą być w dużym stopniu kontrolowalne poprzez dobór warunków wzrostu.



Rys.45 Obrazy z mikroskopu sił atomowych ilustrujące powstawanie „wysp” (kropek kwantowych) na podłożu o innej stałej sieci (<http://cqd.eecs.northwestern.edu>).

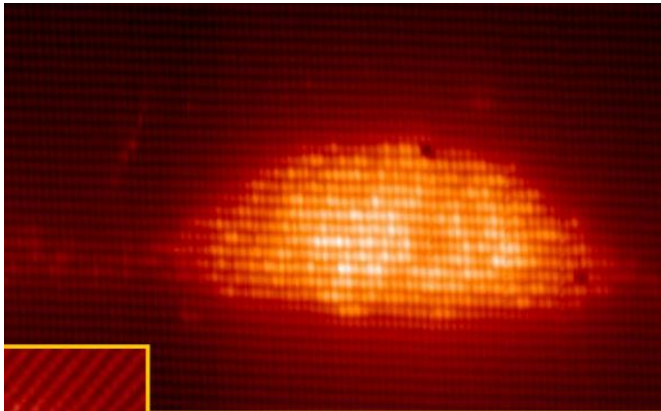
Zwykle proces wzrostu „wysp” jest przerywany, natomiast ponownie nanoszony jest materiał podłoża. W rezultacie otrzymuje się strukturę zbudowaną z „wtrąceń” materiału półprzewodnikowego otoczonego przez materiał podłoża (Rys. 46 oraz Rys. 47).



Rys.46. Schematyczny przekrój przez warstwę samorosnących kropek kwantowych.

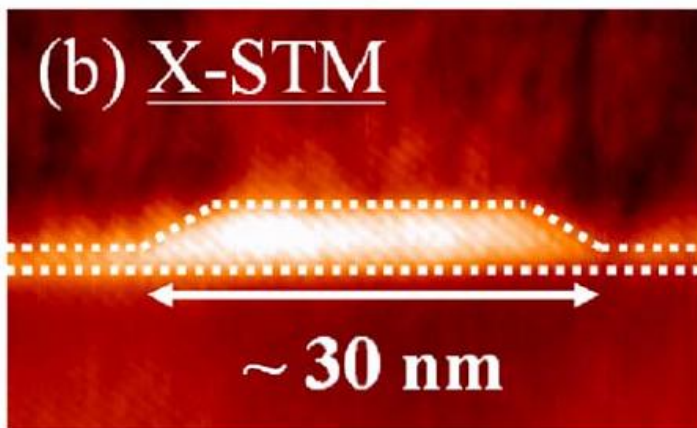
Odpowiedni dobór składu chemicznego, np. otaczającej matrycy o szerokiej przerwie energetycznej (GaAs lub InP) oraz wąsko-przerwowego materiału „wysp” (InAs) prowadzi do przestrzennego ograniczenia ruchu nośników ładunku (elektronów i dziur) do obszaru danej „wyspy”. Tego rodzaju nanostruktury nazywamy samorosnącymi kropkami kwantowymi (ang. self-assembled quantum dot), gdzie „samorosnący” lub „samoorganizujący” ilustruje charakter procesu wzrostu, a słowo „kropka

kwantowa” opisuje niewielki obszar przestrzeni ograniczony w trzech wymiarach barierami potencjału.



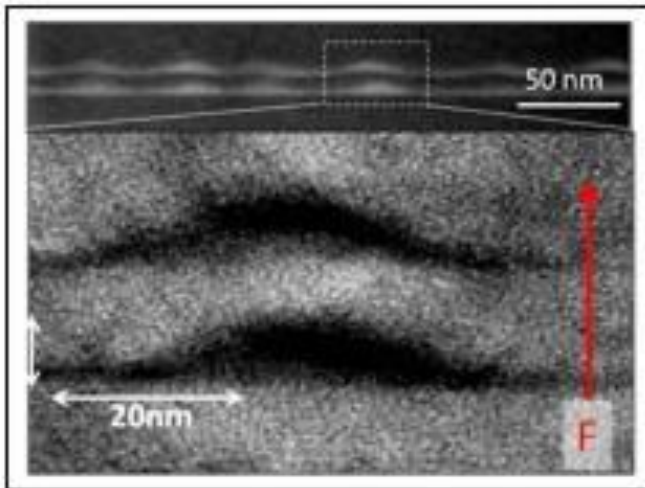
Rys.47 Obraz ze skaningowego mikroskopu tunelowego, przekrój (X-STM, ang. cross-section scanning tunneling microscope) przez pojedynczą samorosnącą kropkę kwantową InAs/GaAs o kształcie soczewki (źródło: prace P.M. Koenraad et al).

Kontrola parametrów wzrostu umożliwia także uzyskiwanie nanostruktur o innych niż soczewka kształtach. W szczególności technika „indium-flush” pozwala na tworzenie samorosnących kropek kwantowych o kształcie dysku (Rys. 48).

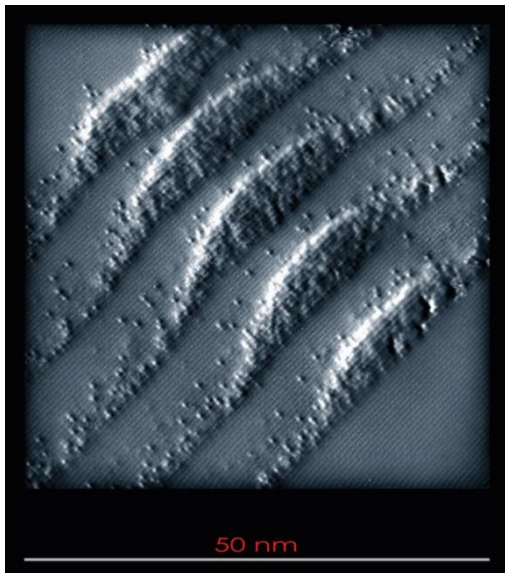


Rys.48. Obraz ze skaningowego mikroskopu tunelowego, przekrój (X-STM) przez pojedynczą samorosnącą kropkę kwantową InAs/GaAs o kształcie dysku (źródło: prace P.M. Koenraad et al).

Obok pojedynczych kropek kwantowych możliwe jest również stworzenie układów podwójnych kropek kwantowych (Rys. 49) lub nawet stosu kropek kwantowych (Rys. 50).



Rys.49. Obraz ze transmisyjnego mikroskopu elektronowego (TEM, ang. transmission electron microscopy), przekrój przez układ dwóch samorosnących kropek kwantowych InAs/GaAs (źródło: Internet).



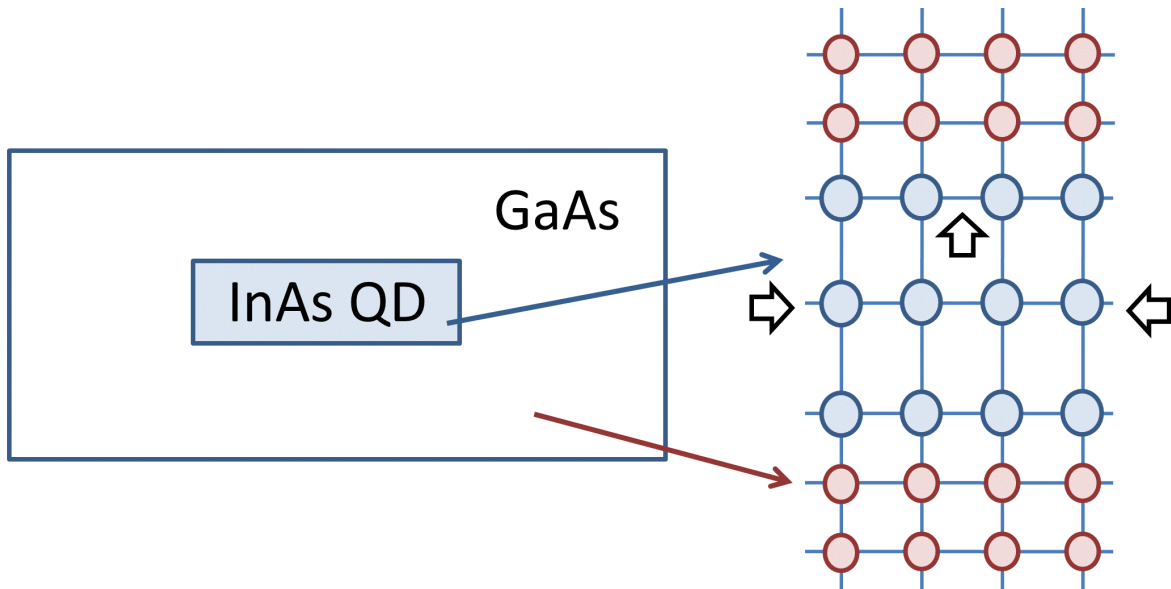
Rys.50. Obraz ze skaningowego mikroskopu tunelowego, przekrój (X-STM) przez pionowy stos samorosnących kropek kwantowych InAs/GaAs (źródło: Internet).

Własności widmowe pojedynczych kropek kwantowych przypominają własności atomów (dyskretne widmo energii), stąd czasem stosowana jest nazwa „sztuczne atomy”. Analogicznie, układy dwóch i więcej kropek kwantowych nazywa się często mianem „sztucznych molekuł”.

6. Rodzaje i źródła odkształceń

Tensorowy opis odkształceń

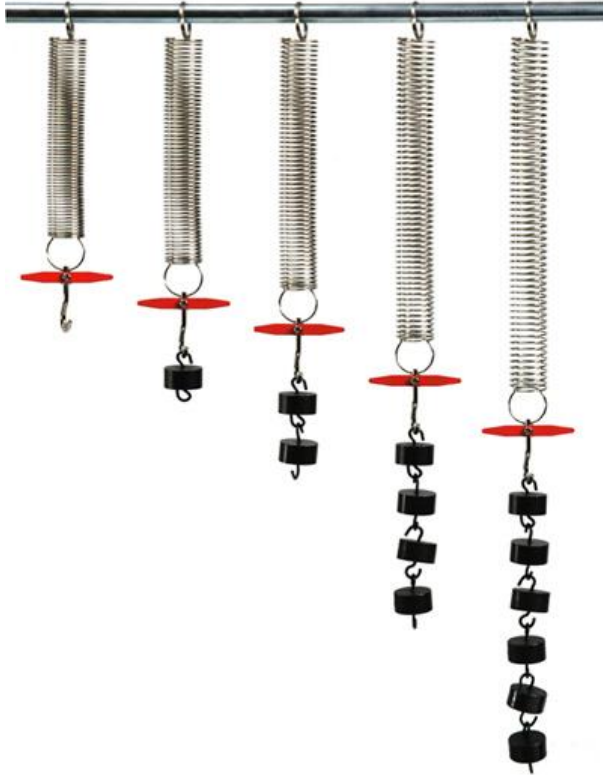
Omówione do tej pory samorosnące kropki kwantowe zawdzięczają swoje istnienie niedopasowaniu sieci krystalicznych materiału kropki i podłoża i związanym z tym efektom odkształceń. Z drugiej strony materiał kropki jest silnie odkształcony (Rys. 51), i jak należy przypuszczać efekty odkształceń mogą mieć istotny wpływ na własności widmowe danej kropki kwantowej.



Rys.51. Schemat deformacji odległości międzyatomowych dla kropki kwantowej InAs w matrycy GaAs.

Samorosnąca kropka kwantowa stanowi dość skomplikowany, trójwymiarowy układ, w pierwszej kolejności zastanówmy się więc jak ilościowo opisać deformacje (odkształcenia) dla zdecydowanie prostszego zagadnienia jednowymiarowego.

Zacznijmy od przykładu metalowej sprężyny na której zawieszamy coraz większe obciążenie (Rys. 52). Zgodnie z prawem Hooke'a oraz naszą intuicją stopień rozciągnięcia sprężyny zależy od jej rodzaju (np. grubości oraz materiału drutu z którego jej wykonana) oraz zawieszanej masy.



Rys.52. Deformacja (rozciąganie) sprężyny w funkcji rosnącego obciążenia.

Na użytek aktualnych rozważań nie interesują nas jednak ani siły związane z tym procesem, ani naprężenia występujące w drucie sprężyny, a jedynie jakaś miara jej deformacji. Wprowadzimy więc definicję odkształcenia sprężyny poprzez porównanie długości sprężyny zdeformowanej i nieodkształconej (Rys. 53) oraz podzielenie tej różnicy przez długość nieodkształconą.

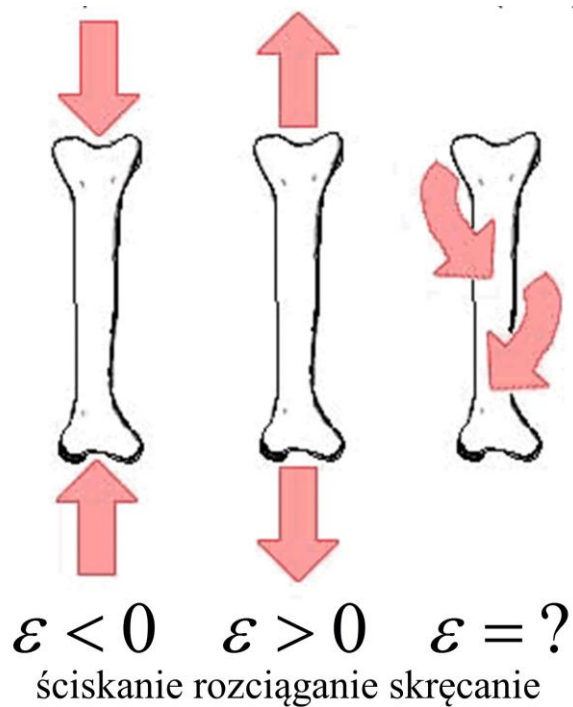


$$\epsilon = \frac{L - L_0}{L_0} = \frac{\Delta L}{L_0}$$

Rys.53. Ilustracja i definicja odkształcenia dla przypadku jednowymiarowego (źródło: Internet).

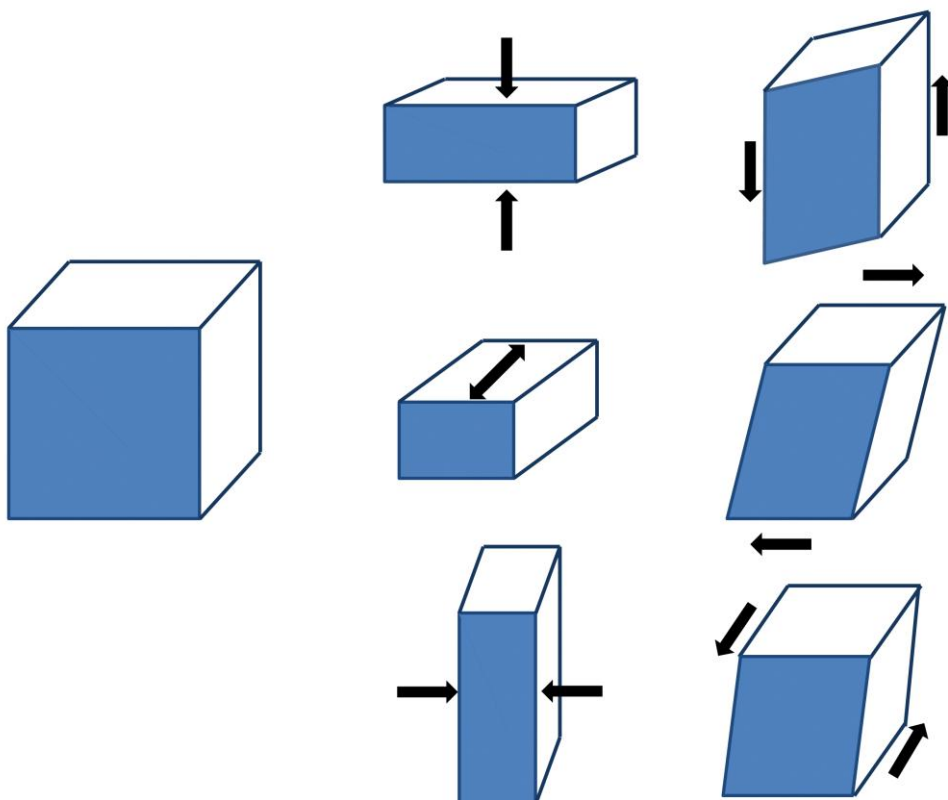
Definicja ta ma taką zaletę, że dla struktury nieodkształconej ϵ przyjmuje wartość równą zero. Znak ϵ opisuje charakter odkształceń: $\epsilon > 0$ odpowiada odkształceniom o charakterze rozciągającym, natomiast $\epsilon < 0$ odpowiada odkształceniom o charakterze ściskającym. Dodatkowo dzielenie przez długość sprężyny nieodkształconej pozwala wprowadzić jednakową miarę odkształceń dla obiektów różnej długości. Przykładowo metrowy drut rozciągnięty o 1 centymetr jest tak samo silnie odkształcony (1%), jak kilometrowy drut rozciągnięty o dziesięć metrów.

Definicja ϵ jest w pełni wystarczająca dla deformacji o charakterze jednowymiarowym. Problem pojawia się jednak dla bardziej skomplikowanych deformacji przestrzennych (Rys. 53b).



Rys.53b.. Ilustracja znaku odkształcenia dla przypadku jednowymiarowego oraz „kłopotu” z prostym opisem jednowymiarowym dla bardziej skomplikowanych deformacji o charakterze skręcającym.

W ogólnym przypadku w opisie należy uwzględnić istnienie pozostałych dwóch wymiarów. Ilustruje to rysunek Rys. 54, który pokazuje wiele (często skomplikowanych) możliwości w jakie można zdeformować jednostkowy sześcian. Możliwe jest więc rozciąganie i ściskanie wzdłuż określonych krawędzi sześcianu, a co więcej deformacja o charakterze skręcającym, bądź ścinającym.



Rys.54. Schemat różnych odkształceń dla trójwymiarowego sześcianu.

Wszystkie możliwe deformacje sześcianu można opisać za pomocą jednej wielkości macierzowej, tzw. tensora odkształceń:

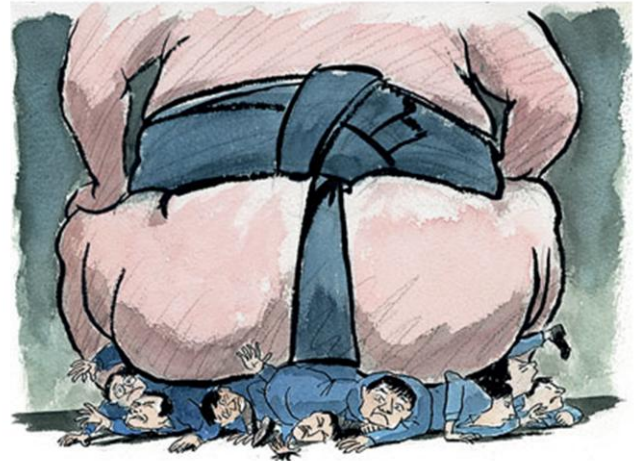
$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix}$$

Tensor ten jest symetryczny ($\varepsilon_{ij}=\varepsilon_{ji}$) oraz, co ważne, opisuje deformacje kształtu jednostkowego sześcianu, a nie jego obrót (wyrażony za pomocą macierzy obrotów) czy translacje (przesunięcia) w przestrzeni.

Rodzaje odkształceń

Bez wchodzenia w szczegóły matematyczne zilustrujemy (Rys. 55) główne rodzaje odkształceń oraz odpowiadające im postaci tensora odkształceń. W szczególności deformacja o charakterze hydrostatycznym (której doświadcza np. kabina łodzi podwodnej) oznacza jednorodne ściskanie ($\varepsilon < 0$), zatem żaden z kierunków przestrzennych nie jest wyróżniony ($\varepsilon_{xx}=\varepsilon_{yy}=\varepsilon_{zz}$). Ponieważ kształt deformowanej struktury jest zachowany to elementy niediagonalne tensora są dokładnie równe zero.

Z kolei dla odkształceń o charakterze osiowym, ściskaniu (rozciąganiu) w jednym z kierunków (np. $\epsilon_{zz} < 0$) towarzyszy rozciąganie (ściskanie) w dwóch pozostałych kierunkach ($\epsilon_{xx} = \epsilon_{yy} > 0$). Wybierając osie układu współrzędnych wzdłuż kierunku deformacji ponownie otrzymujemy diagonalną postać tensora odkształceń, przy czym elementy na diagonalu są różne (np. $\epsilon_{xx} = \epsilon_{yy} \neq \epsilon_{zz}$).



Hydrostatyczne

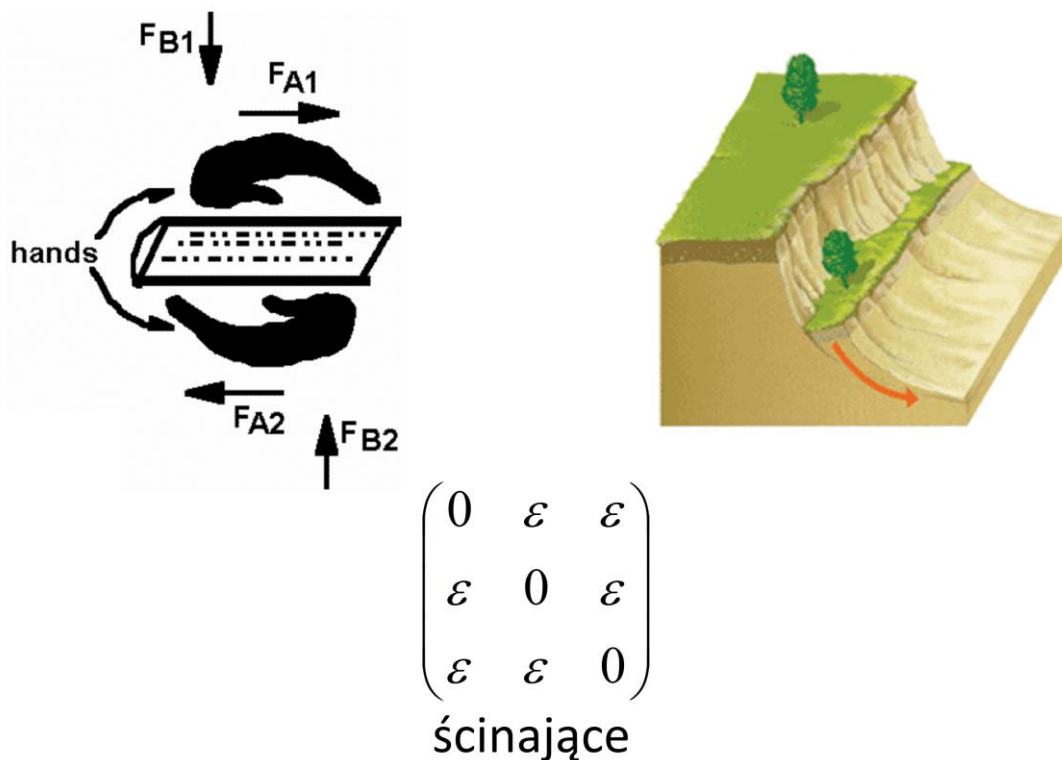
$$\begin{pmatrix} \epsilon & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & \epsilon \end{pmatrix}$$

Osiowe

$$\begin{pmatrix} \epsilon_{\leftrightarrow} & 0 & 0 \\ 0 & \epsilon_{\leftrightarrow} & 0 \\ 0 & 0 & \epsilon_{\updownarrow} \end{pmatrix}$$

Rys.55 Przykład odkształceń o charakterze hydrostatycznym i osiowym oraz odpowiadających im postaci tensor odkształceń (Źródło: Internet).

Jeszcze innym rodzajem odkształceń są te o charakterze ścinającym (ang. shear strain), zilustrowane na Rys. 56, w takim przypadku zmianie nie ulegają liniowe wymiary struktury, a odkształcenie ciała spowodowane przyłożeniem siły stycznej do jej powierzchni. Należy w tym miejscu dodać, że trzy wspomniane wyżej rodzaje deformacji mogą występować jednocześnie. Na przykład w przypadku „typowych” samorosnących kropek kwantowych, dominująca rolę odgrywają odkształcenia o zarówno hydrostatycznym, jak i osiowym, natomiast te o charakterze ścinającym są zdecydowanie mniej istotne.



Rys.56. Przykład odkształceń o ścinającym (Źródło: Internet).

Z elementów macierzy tensora odkształceń można skonstruować użyteczne wielkości pomocnicze będące miarą określonych typów deformacji. W szczególności ślad tensora odkształceń (suma elementów diagonalnych macierzy) jest ilościową miarą odkształceń hydrostatycznych:

$$Tr(\varepsilon) = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}$$

Dla niewielkich deformacji $Tr(\varepsilon)$ opisuje zmianę objętości jednostkowego sześcianu:

$$V \approx V_0 [1 + Tr(\varepsilon)]$$

Ujemne wartości $Tr(\varepsilon)$ opisują zatem odkształcenia o charakterze ściskającym, a dodatnie o charakterze rozciągającym. Użyteczności $Tr(\varepsilon)$ jest tym większa, że jest on niezmiennikiem, a więc wielkością niezależną od wyboru układu współrzędnych.

Miarą odkształceń osiowych może być np. wielkość zdefiniowana następująco:

$$B(\varepsilon) = \sqrt{(\varepsilon_{xx} - \varepsilon_{yy})^2 + (\varepsilon_{yy} - \varepsilon_{zz})^2 + (\varepsilon_{zz} - \varepsilon_{xx})^2}$$

Dla oddziaływań o charakterze czysto hydrostatycznym ($\epsilon_{xx}=\epsilon_{yy}=\epsilon_{zz}$) mamy zatem $B(\epsilon)=0$, gdyż wszystkie wymiary liniowe odkształcone się w ten sam sposób. Z kolei dla odkształceń o charakterze czysto osiowym jest $\text{Tr}(\epsilon)=0$, przy czym $B(\epsilon)\neq 0$, co odpowiada deformacji kształtu zachowującej objętość. Przypomnijmy raz jeszcze, że w ogólności możliwe są odkształcenia będą kombinacją wszystkich wspomnianych podstawowych typów odkształceń.

Odształcenia a naprężenia.

W naszych rozważaniach dotyczących kropek kwantowych opisaliśmy już jeden przykład: odkształcenia pochodzące od niedopasowania stałych sieci krystalicznych dwóch materiałów półprzewodnikowych. Odształcenia występują jednak w wielu innych układach, w szczególności makroskopowych m.in. konstrukcjach budowlanych, urządzeniach technicznych itp. Należy też doprecyzować, że mówiąc od odkształceniach odnosimy się do odkształceń sprężystych, a więc takich, które ustępuje po usunięciu siły, która je spowodowała. W tym wykładzie nie interesują nas odkształcenia plastyczne charakterystyczne dla takich materiałów jak np. plastelina czy glina.

Należy w tym miejscu rozróżnić także „odkształcenia” od „naprężeń”. Te pierwsze opisują deformację danego ciała w stosunku do przypadku niedokształconego, przy czym ϵ jest wielkością bezwymiarową. Naprężenia (oznaczenie σ) są z kolei pewną miarą sił związanych z tą deformacją, a ich jednostką jest paskal (N/m^2 , jednostka siły na jednostkę powierzchni). Przykładowo intuicja podpowiada nam, że 1% odkształcenie stalowego drutu o przekroju 1mm^2 wymaga przyłożenia zdecydowanie większej siły niż gumowego paska o identycznym przekroju.

Relacja pomiędzy dwoma wspomnianym wyżej wielkościami jest niczym innym jak prawem Hooke’a:

$$\sigma = E\epsilon,$$

gdzie współczynnik proporcjonalności E jest parametrem materiałowym zwanym modułem sprężystości (sztywności) podłużnej E lub krócej modułem Younga. Wielkość ta przyjmuje wartości np. 210 GPa dla stali i 0.1 GPa dla gumy, co dobrze ilustruje fundamentalnie różną sprężystość obu materiałów.

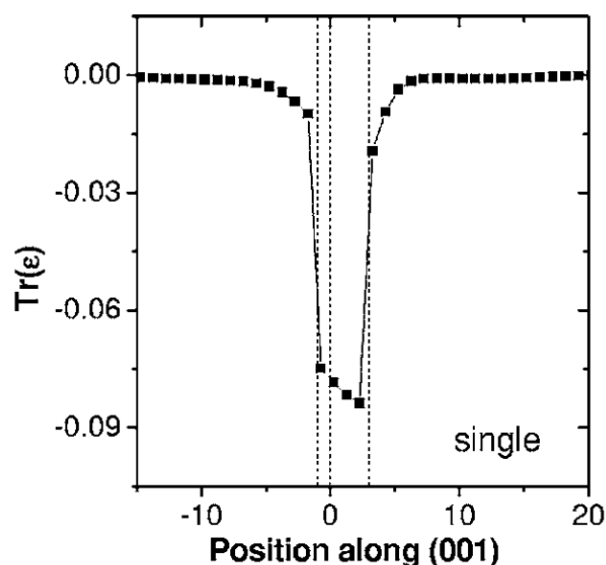
Podobnie jak odkształcenia, naprężenia są w ogólnym przypadku (dla materiałów anizotropowych) opisane wielkością tensorową (σ_{ij}), a relacja między obiema wielkościami zwana jest uogólnionym prawem Hooke’a:

$$\sigma_{kl} = \sum_{ij} c_{klij} \epsilon_{ij}$$

Ponieważ naprężenia oraz odkształcenia są tensorami drugiego rzędu, związek pomiędzy nimi jest zadany tensorem czwartego rzędu, c_{ijkl} , zwanym tensorem sztywności, którego składowe zwane są modułami sprężystości lub współczynnikami sztywności sprężyste. Tensor sztywności ma w ogólności $3^4=81$ składowych, jednak symetria sieci krystalicznej istotnie redukuje liczbę niezależnych składowych i tak dla struktury o symetrii regularnej (kubicznej) mamy jedynie 3 niezerowe składowe c_{ijkl} , zwyczajowo oznaczane jako C_{11} , C_{12} oraz C_{44} , przy czym zastosowano tutaj notację w której indeksy 1, 2 oraz 4 odpowiadają parom wskaźników: 11, 22, 23.

Przybliżenie ośrodka ciągłego oraz podejście atomistyczne.

Do tej pory omawialiśmy deformacje obiektu (sieci krystalicznej) jako całości. Na wstępie opisaliśmy jednak przykład samorosnącej kropki kwantowej oraz zasugerowaliśmy, że kropka kwantowa jest silnie odkształcona, podczas, gdy podłoże krystaliczne jest zdeformowane w mniejszym stopniu. Oznacza to, że dla w ogólnym przypadku deformacja, a więc i składowe tensora odkształceń, jest funkcją położenia. Jest to widoczne na Rys. 57, gdzie wewnątrz kropki kwantowej doznaje silnych odkształceń o charakterze ściskającym ($\text{Tr}(\epsilon) \approx -0.08$), podczas, gdy obszar w odległości 10 stałych sieci od kropki kwantowej pozostaje praktycznie niezdeformowany.



Rys.57. Ślad tensora odkształceń $\text{Tr}(\epsilon)$ w funkcji położenia atomów; przekrój wzdłuż kierunku wzrostu przez geometryczny środek kropki kwantowej InAs/GaAs o kształcie soczewki (Źródło: *Phys. Rev. B* **74**, 195339 (2006)).

Jednym ze sposobów wyznaczenia przestrzennego rozkładu odkształceń jest wykonanie obliczeń w tzw. przybliżeniu ośrodka ciągłego (ang. continuum elasticity), w którym zaniedbujemy atomową budowę materii, to znaczy ośrodek dzielimy na bardzo małe części, ale o takich samych własnościach jak cała badana objętość.

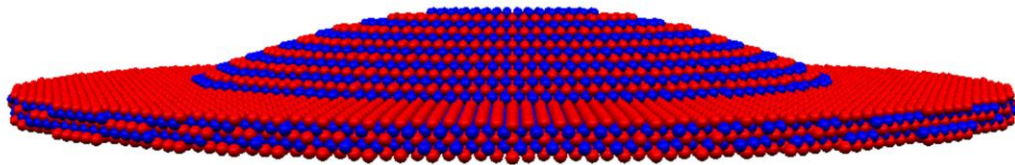
W interesującym nas przypadku definiujemy wartości energii sprężystej (odkształceń) w każdym punkcie dyskretnej, przestrzennej siatki. Przykładowo dla struktury o symetrii regularnej:

$$E_{CE}(x, y, z) = \frac{V}{2} \left[C_{11}(x, y, z)(\epsilon_{xx}^2 + \epsilon_{yy}^2 + \epsilon_{zz}^2) + C_{44}(x, y, z)(\epsilon_{yz}^2 + \epsilon_{zx}^2 + \epsilon_{xy}^2) + 2C_{12}(x, y, z)(\epsilon_{xx}\epsilon_{yy} + \epsilon_{zz}\epsilon_{xx} + \epsilon_{xx}\epsilon_{yy}) \right],$$

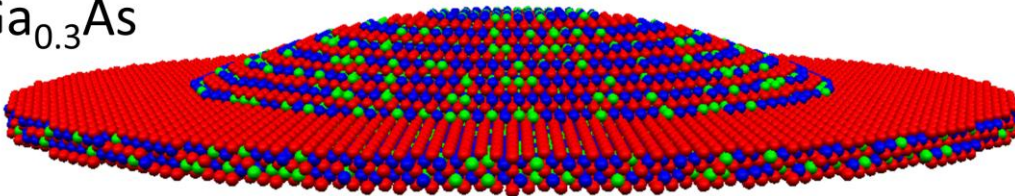
gdzie C_{11} , C_{12} oraz C_{44} to zdefiniowane wcześniej zależne od położenia (zatem inne dla kropki kwantowej oraz inne dla otaczającej matrycy) współczynniki sztywności sprężystej, a V to objętość związana z dyskretyzacją siatki (objętość „oczka” siatki). Wartości ϵ w każdym węźle siatki są znajdowane w ten sposób, aby energia całkowita układu, tj. suma E_{CE} po wszystkich punktach siatki, przyjmowała minimum.

W podejściu tym widoczny jest jednak pewien fundamentalny problem. Po pierwsze krok siatki (a więc i objętość V) musi być tak dobrany, aby zminimalizować błędy dyskretyzacji. Z drugiej strony, nawet jeśli krok siatki byłby bardzo mały, np. porównywalny ze stałą sieci krystalicznej, to i tak symetria dyskretnej siatki obliczeniowej jest w ogólności inna od symetrii sieci krystalicznej. W szczególności, dla materiałów takich jak InAs czy InP, może to być całkiem skomplikowana symetria typu blendy cynkowej lub wurcytu. Co więcej, jak widzieliśmy wcześniej na obrazach X-STM kropek kwantowych, atomy w kropce kwantowej i na jej brzegu mogą być umieszczone w sposób losowy. Fluktuacje składu chemicznego np. dla kropek kwantowych o mieszanym składzie chemicznym ($\text{In}_x\text{Ga}_{1-x}\text{As}$, Rys. 58) odgrywają istotną rolę i nie mogą być w sposób prosty modelowane w przybliżeniu ośrodka ciągłego.

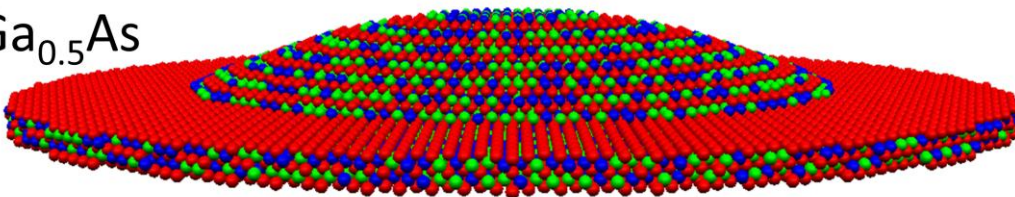
InAs



$\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$



$\text{In}_{0.5}\text{Ga}_{0.5}\text{As}$



Rys.58. Położenia atomów arsenu (kolor czerwony), indu (niebieski) oraz galu (zielony) dla warstwy zwilżającej oraz kropki kwantowej $\text{In}_x\text{Ga}_{1-x}\text{As}$ o kształcie soczewki oraz różnych wartości składu chemicznego. Otaczająca kropkę kwantową matryca GaAs została na rysunku pominięta.

Sposobem na opisane wyżej problemy jest uwzględnienie w obliczeniach atomowego, „ziarnistego” charakteru materii. W odróżnieniu od metod wykorzystujących przybliżenie ośrodka ciągłego tego rodzaju metody noszą ogólne miano „atomistycznych”.

Przykładem takiej właśnie metody jest VFF (ang. Valence Force Field), w której energia sprężystości zdefiniowana jest jako funkcja położenia atomów, a w szczególności długości wiązań międzyatomowych oraz kątów między wiązaniami. W interesującym nas przypadku kryształów o mikroskopowej symetrii tetraedralnej mamy:

$$E_{VFF} = \frac{1}{2} \sum_i^N \sum_{j=1}^4 A_{ij} \left((\mathbf{R}_i - \mathbf{R}_j)^2 - (d_{ij}^0)^2 \right)^2 + \sum_i^N \sum_{j=1}^3 \sum_{k=j+1}^4 B_{ijk} \left((\mathbf{R}_j - \mathbf{R}_i) \cdot (\mathbf{R}_k - \mathbf{R}_i) - \cos \theta_{ij}^0 d_{ik}^0 \right)^2,$$

W wyrażeniu tym znajdują się dwie sumy. W przypadku pierwszej sumy indeks i przebiega po wszystkich (N) atomach w układzie, a indeks j po (4) najbliższych sąsiadach danego atom i , A_{ij} oznacza z kolei stałą siłową związaną ze zmianą długości wiązania (ściskanie/rozciąganie) pomiędzy atomami i oraz j . Druga suma przebiega po wszystkich atomach oraz parach sąsiadów (ij), a stała siłowa B_{ijk} związana jest ze zmianą kąta między wiązaniami $ij-ki$ w stosunku do idealnej wartości dla kryształu tetraedrycznego ($\cos(\theta)=1/3$). W obu sumach \mathbf{R}_i oznacza położenia danego atomu.

Rozwiązanie problemu odkształceń w metodzie VFF polega na znalezieniu takich pozycji atomów (\mathbf{R}_i) dla których energia E_{vff} osiąga minimum. Dla samorosnących kropek kwantowych liczba atomów w domenie obliczeniowej może sięgać nawet 100 milionów atomów, co prowadzi do złożonego problemu obliczeniowego.

Znajdowanie minimum funkcji – metoda gradientów prostych

Istnieje wiele metod znajdowania ekstremum funkcji, jedną z nich jest metoda gradientu prostego (ang. *gradient descent*). Jest to metoda iteracyjna w której ekstremum osiągane jest w serii kolejnych kroków wykonywanych w kierunku największego wzrostu lub spadku, w zależności od tego czy celem optymalizacji jest minimum funkcji czy jej maksimum (Rys. 59).

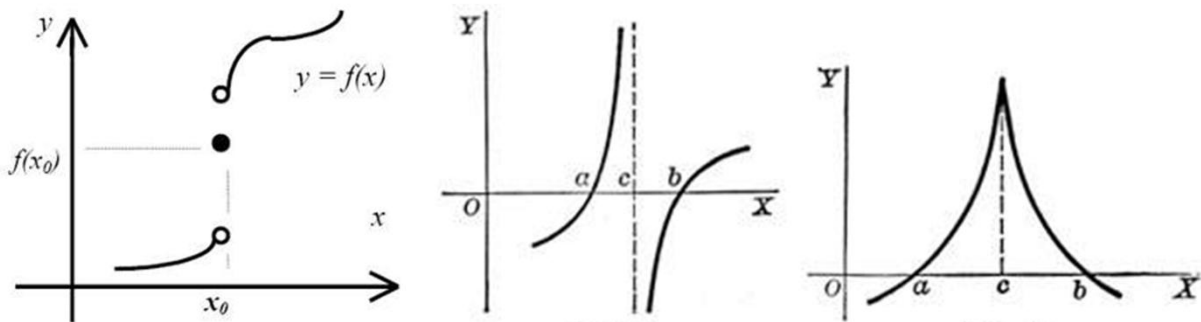


Rys.59. Najkrótsza droga na szczyt to zwykle ta najbardziej stroma.

Utożsamiając kierunek największego spadku z (minus) gradientem funkcji, bardziej sformalizowany algorytm wygląda następująco:

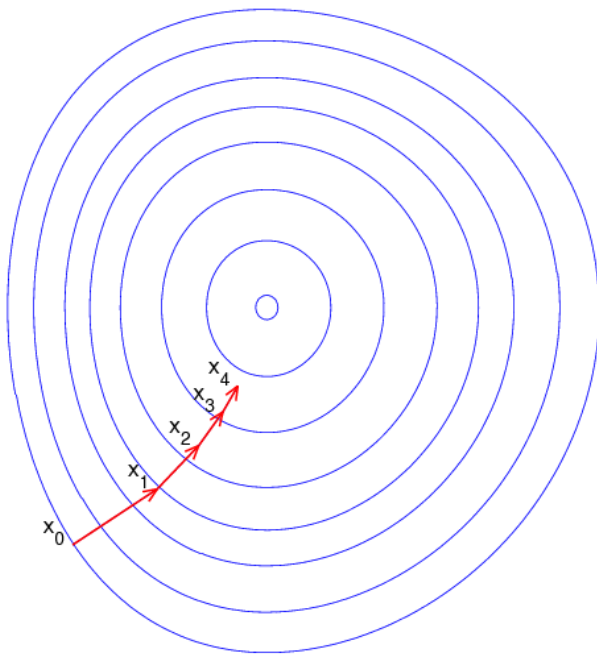
$$x_{k+1} = x_k - \alpha_k \nabla f_k$$

Przy czym $f : \mathbf{R}^N \mapsto \mathbf{R}$ jest różniczkowalną i ciągłą (Rys. 60) funkcją N zmiennych.



Rys.60. Przykłady funkcji nieciągłych i nieróżniczkowalnych.

Proces iteracyjny (Rys. 61) rozpoczyna się od punktu startowego x_0 , którego wybór jest arbitralny i zależy od konkretnego problemu. Długość kroku (α_k) powinna być z kolei tak dobrana, aby wartość funkcji w kolejnych punktach nie rosla ($f(x_0) > \dots > f(x_k) > f(x_{k+1})$). Zbyt duży krok sprawi, że metoda nie będzie numerycznie stabilna, z kolei zbyt mały krok spowolni proces znajdowania minimum. Podobnie jak x_0 , wybór α_k jest w dużym stopniu arbitralny i zależy od konkretnego zagadnienia.



Rys.61. Kilka kolejnych kroków w metodzie gradientu prostego.

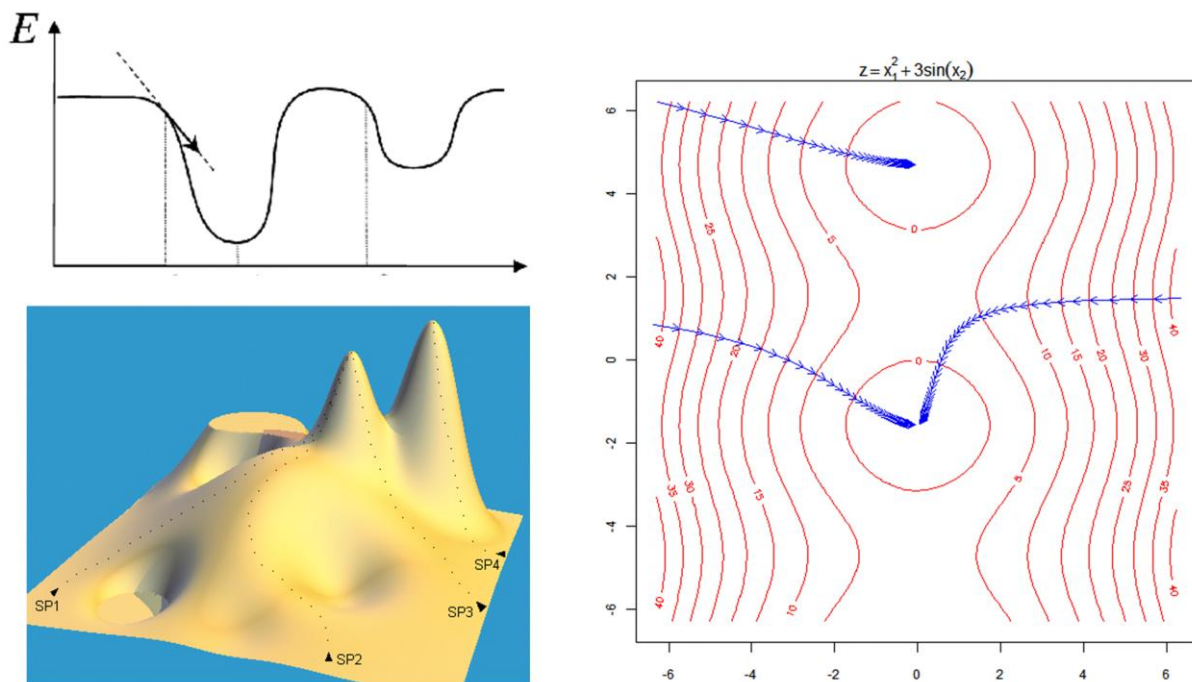
Jednym ze sposobów na uniknięcie powyższego problemu jest wariant metody gradientu prostego, tzw. metoda najszybszego spadku (ang. steepest descent), w której długość kroku nie jest parametrem, ale jest wyznaczana w każdym kroku iteracji, tak by funkcja osiągnęła minimum w na kierunku gradientu:

$$f(x_k - \alpha_k \nabla f(x_k)) = \min_{\alpha > 0} f(x_k - \alpha \nabla f(x_k))$$

Zaletą tego podejścia jest uniknięcie związanych kłopotów z wyborem α_k , z drugiej strony jednak strony metoda ta jest bardziej skomplikowana od metody gradientu prostego oraz przede wszystkim zwiększona jest ilość (czas) wykonywanych obliczeń numerycznych.

Niezależnie od sposobu wyboru α_k proces iteracyjny kontynuowany jest do momentu osiągnięcia minimum. W celu określenia, czy dany punkt dostatecznie dobrze przybliży minimum funkcji można użyć kilku kryteriów, np. sprawdzić czy długość wektora gradientu jest mniejsza od pewnej zadanej precyzji ($\|\nabla f(x_k)\| < \varepsilon$) lub odległość między kolejnymi punktami jest mniejsza od pewnej odległości minimalnej ($\|x_{k+1} - x_k\| < \delta$). Oba kryteria mogą być też stosowane jednocześnie.

Na zakończenie tej części wykładu należy zwrócić uwagę, że w metodzie gradientu prostego, jak i metodzie największego spadku, znajdowane jest lokalne, a nie globalne minimum funkcji (Rys.62). Dla funkcji posiadających wiele minimów lokalnych, to które z nich zostanie znalezione zależy m.in. od wyboru punktu startowego.



Rys.62 Jedno- i dwuwymiarowe funkcje posiadające lokalne minima.

Modelowanie właściwości elektrycznych i mechanicznych bionanoukładów

Wprowadzenie do nanotechnologii układów biologicznych. Centralny dogmat biologii molekularnej

Zastanówmy się nad fenomenem życia na Ziemi. Trwa ono nieprzerwanie od kilku miliardów lat. Obejmuje miliony różnorodnych organizmów. Biolodzy grupują je w odpowiednie podzbiory, np. gatunki. Organizmy te podlegają ewolucji, czyli zmianom. Zmiany mogą być korzystne (lepsze dopasowanie do środowiska) lub niekorzystne. Często gatunki żywych organizmów giną, istnieją jednak takie, które przetrwały miliony lat. Istotą utrzymania procesu życia jest rozmnażanie się – jedne osobniki giną, ale wydają „potomstwo”. Czasem obserwujemy okres hibernacji, zanim organizm przejawia swoje pełne cechy (rośliny, nasiona), czasem życie nowego pokolenia krytycznie zależy od opieki pokolenia rodziców (ssaki). W jaki sposób przekazywana jest informacja z jednego organizmu do organizmu potomnego o „planie życiowym”? Jakie są mechanizmy chemiczne pozwalają na lepszą adaptację do otoczenia? Dlaczego wielkie dinozaury wyginęły (ok. 60 mln lat temu), a małe ssaki przetrwały (kataklizm dużego meteorytu??). Dzięki postępom fizyki, chemii i biologii znamy odpowiedzi na te fundamentalne pytania. Znając struktury i mechanizmy biologiczne możemy pewne rozwiązania „wynalezione” przez Naturę w toku milionów lat ewolucji przenosić do techniki, w tym na poziomie nanoobiektów. W ten sposób w ostatnich kilkunastu latach narodziła się nowa gałąź wiedzy – bionanotechnologia [1, Ehud Gazit, Plenty of room for biology at the bottom: An introduction to bionanotechnology. Imperial College Press, 2007, zob. darmowy rozdział:

http://www.worldscientific.com/doi/suppl/10.1142/p465/suppl_file/p465_chap01.pdf .

Duży wysiłek też się wkłada w poprawianie natury i tutaj, przy projektowaniu nowych układów. przydaje się modelowanie komputerowe.

Organizmy żywe składają się (przeważnie) z tkanek, a te z komórek. W komórkach bieżą podstawowe procesy życiowe (metabolizm) – pobieranie składników energetycznych i budulcowych z otoczenia, tworzenie niezbędnych elementów komórki czy nowych komórek. Komórki są zwykle małe, mają rozmiary rzędu mikrometrów. Zbudowane są z biopolimerów, ich wnętrza wypełnione jest licznymi błonami, białkami i cząsteczkami sygnałowymi czy energetycznymi. Białka strukturalne, białka enzymatyczne i fosfolipidy (składniki błon) odgrywają najważniejszą rolę w funkcjonowaniu komórki. Są one syntetyzowane na bieżąco w odpowiednich rejonach komórki. Na przykład białka powstają w skomplikowanych fabrykach biochemicznych zwanych rybosomami. Każdy człowiek ma ok. 1,000,000,000,000,000 rybosomów. Strukturę przestrzenną takiej fabryki poznano niedawno a odkrycie to zostało uhonorowane Nagrodą Nobla (zob. Rys. 63).

**The Nobel Prize in Chemistry 2009
Venkatraman Ramakrishnan, Thomas
A. Steitz, Ada E. Yonath**



Rys. 63. Prof. Ada Jonath z Izraela podczas wizyty na UMK.

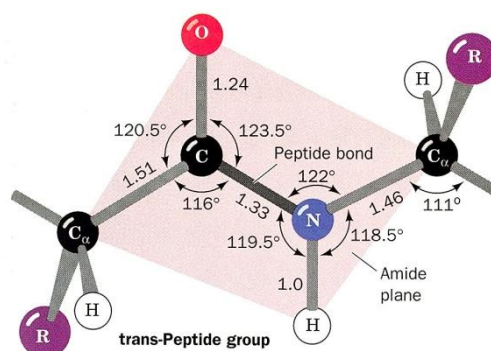
Strukturę nanofabryki – rybosomu można podziwiać np. w <http://www.rcsb.org/pdb/101/motm.do?momID=121>

oraz:

<http://www.dnatube.com/video/2425/Translation-Atomic-View>).

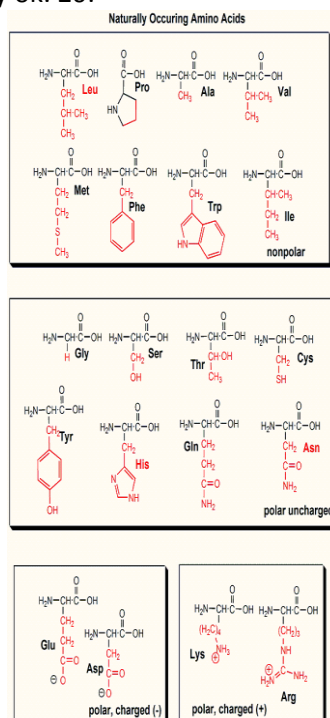
Jak działa taka maszyna biochemiczna? Jak to się dzieje, że z prostych składników – aminokwasów, powstają tak złożone białka jak hemoglobina (transport tlenu we krwi) czy aktyna (włókna mięśniowe)? Na te pytania odpowiadają symulacje komputerowe realizowane na największych komputerach na świecie (zob. np. modelowanie działania rybosomu: <http://www.dnatube.com/video/1228/tRNARibosome-Molecular-Dynamics-Simulation>, (proszę zwrócić uwagę na stosunkowo „słabą” rozdzielczość grafiki w tej dość leciwej animacji).

Białka są to biopolimery złożone z połączonych wiązaniami peptydowymi aminokwasów:



Rys. 64. Wiązanie peptydowe (źródło: D.Voet, J. Voet, Biochemistry, 2nd Ed, 1995)

Podstawowych aminokwasów mamy ok. 20:



Rys. 65. Struktury chemiczne podstawowych aminokwasów (Internet: <http://employees.csbsju.edu/hjakubowski/classes/ch112/proteins/aminoacidprot1.gif>).

Z nich zbudowane są praktycznie wszystkie białka obecne w organizmach żywych. Chemicy oczywiście znają więcej związków chemicznych tego typu (tysiące) ale w naturze wykorzystywanych jest tylko ok. 20.

Kolejność połączenia aminokwasów łańcuch jest zwykle bardzo ważna i determinuje właściwości fizyczne, chemiczne i biologiczne białka. Wynika to z tego, że aminokwasy mają łańcuchy boczne różnej wielkości, różniące się polarnością, ładunkiem czy hydrofobowością. Są to zatem elementy „konstrukcyjne” z których może wybierać nanobiotechnolog tworząc nowe systemy. Sama sekwencja (tzw. struktura pierwszorzędowa) jest ważna, ale jeszcze ważniejsze jest spontaniczne ułożenie się (ang. *folding*) tego łańcucha (ok. 100-1000 aminokwasów) w przestrzeni. To struktura 3D w pełni określa funkcję białka w komórce. Błędne zwinięcie jest zwykle wykrywane i takie egzemplarze białek są „tagowane” i przeznaczane do degradacji i recyklingu. Warto wiedzieć, że nie wszystkie białka potrzebne do życia mają określoną strukturę 3D. Zagadnienie to jest ostatnio modnym tematem badań.

Białka mogą pełnić rolę strukturalną albo być enzymami. Z białek zbudowane są m.in. jedwab, wełna i nici pajęczce.

Skąd rybosom ma „wiedzieć” jakie białko ma syntetyzować i w jakiej kolejności podłączać aminokwasy do tworzonego aktualnie łańcucha białkowego? Okazuje się, że fabryka ta jest precyzyjnie sterowana m.in. informacją zawartą w kwasie dezoksyrybonukleinowym (DNA). Komórki zawierają DNA (często jest on skoncentrowany w jądrze komórkowym, ale nie zawsze), który też jest biopolimerem. DNA zbudowany jest z jeszcze mniejszej liczby elementarnych składników - są to tzw. nukleotydy.

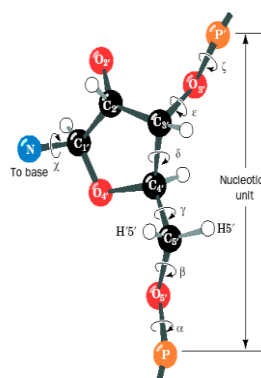
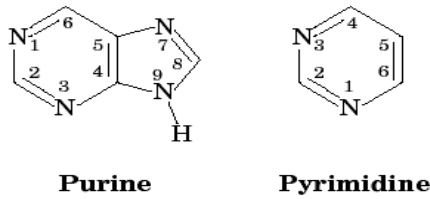


FIGURE 29-7 The conformation of a nucleotide unit is determined by the seven indicated torsion angles.

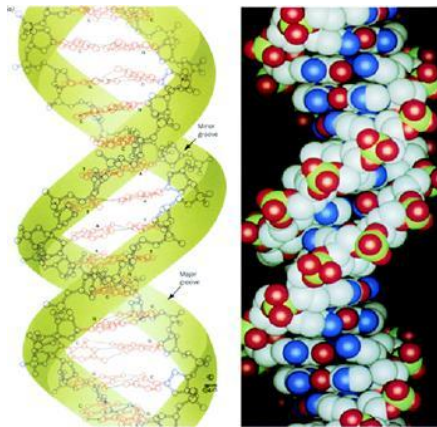
Rys. 66. Schemat budowy nukleotydu (źródło: D.Voet, J. Voet, *Biochemistry*, 2n Ed, 1995).

Nukleotyd składa się z reszty kwasu fosforowego, cukru opartego na pierścieniu pięciocłonowym i z zasady (nazwa chemiczna) – purynowej lub pirymidynowej (Rys. 66).



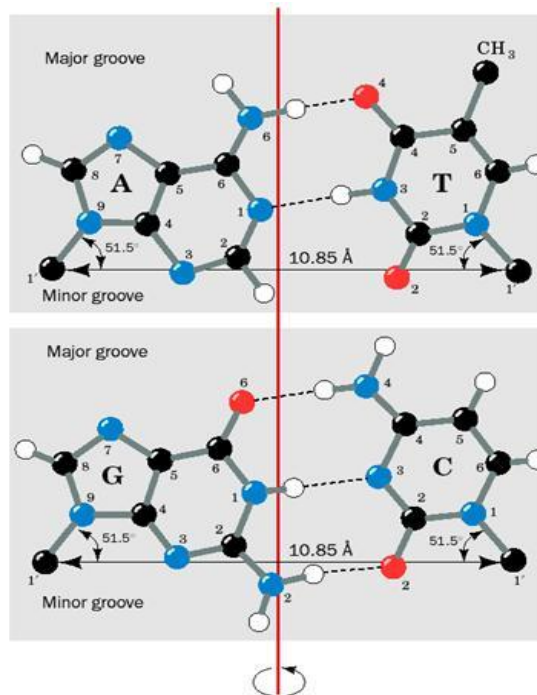
Rys. 67. Schemat ogólny pierścieni purynowych i pirymidynowych, kodujących informacje w DNA i RNA (<http://tonga.usp.edu/gmoyna/biochem341/lecture36.html>).

Strukturę przestrzenną DNA odkryli w latach 50 XX w. T. Watson i J. Crick. Struktura typowego B-DNA jest regularna, mówimy o podwójnej helisie, która przypomina schody:



Rys. 68. Budowa przestrzenna fragmentu B-DNA (źródło: D.Voet, J. Voet, Biochemistry, 2nd Ed, 199).

DNA zwykle występuje w jednej kopii w komórce. Jest to bardzo długa cząsteczka. Ludzkie DNA ma ok. 3 mld nukleotydów w każdym łańcuchu. Co ciekawe jeden z łańcuchów DNA jest komplementarny do drugiego, ponieważ zasady DNA występują w parach (A-T i G-C).



Rys. 69. Standardowe pary zasad Watsona-Cricka kodujące informację (źródło: D.Voet, J. Voet, Biochemistry, 2nd Ed, 1995).

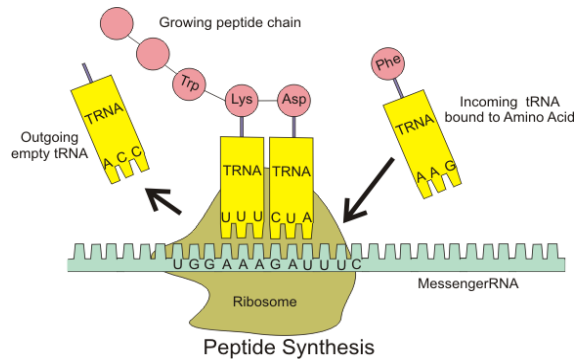
Określone fragmenty DNA zawierają w swojej sekwencji informacje o tym, jak ma być syntetyzowane białko. Nie muszą to być ciągłe odcinki biopolimeru, czasem jeden taki gen kodujący białko jest rozrzucony w kilku miejscach (eksony) oddzielonych miejscami niekodującymi (introny). Mediana długości eksonu wynosi 122 bp. Średnia liczba eksonów w genie to 8 (może być ich nawet 350) średnia długość sekwencji kodującej to 447 aminokwasów. W 3 mld liter ludzkiego DNA jest znacznie więcej obszarów niekodujących, niż kodujących. W roku 2000 odczytano sekwencję ludzkiego DNA. Okazało się, że genów mamy mniej niż wcześniej sądzono: ok. 20 687.

Kiedy białko ma być syntetyzowane, bo jest np. potrzebne do budowy potomnej komórki, fragment silnie skręconego i upakowanego w jądrze komórkowym DNA się rozplata. Rozplatanie to indukowane jest przez odpowiednie enzymy. Potem na odsłoniętej pojedynczej nici DNA następuje synteza stosunkowo krótkiego fragmentu komplementarnego (np. 3000 liter) złożonego z RNA (tzw. mRNA). RNA- kwas rybonukleinowy, jest b. podobny do DNA, nie występuje w nim tylko jedna zasada T, a zamiast niej mamy uracyl U; pierścień cukrowy jest nieco inny niż w DNA. mRNA opuszcza jądro komórkowe i podłącza się do rybosomu. Rybosom działa tak jak głowica magnetofonu i „odczytuje” informacje zawarte w sekwencji RNA. Okazuje się, że w przyrodzie obowiązuje uniwersalny kod genetyczny: określona trójka zasad DNA (czy RNA) koduje określony aminokwas. Łatwo zauważyć, że kod genetyczny musi być redundantny: mamy 64 różne kombinacje trojek 4 zasad, zaś tylko 20 aminokwasów.

1st Position (5' end)	2nd Position				3rd Position (3' end)
	U	C	A	G	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP TRP	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

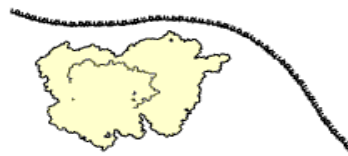
Rys. 70. Kod genetyczny (Internet: http://www.gendnademk.republika.pl/str%205%20kod%20dobre_pliki/kod.htm)

Redundatność chroni informację przed niszczącymi mutacjami w DNA (np. powodowanymi przez palenie papierosów!). Rybosom w oparciu o 3 litery np. AUU zezwala na włączenie do rosnącego łańcucha takiego właśnie aminokwasu (tu: Ile – Isoleucyna) , jaki jest zakodowany przez trójkę kodującą w mRNA. Proces ten jest b. szybki, wszystko jest kontrolowane przez odpowiednie enzymy.



Rys. 71. Synteza białka w nanomaszynie jaką jest rybosom (Internet: <http://en.wikipedia.org/wiki/Ribosome>).

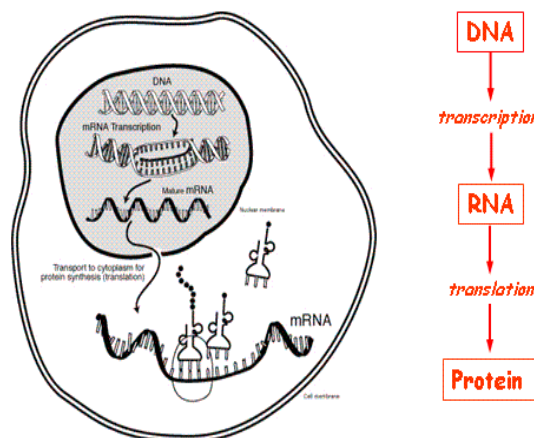
Można obejrzeć to animacji komputerowej (modelowanie procesu bionanotechnologicznego):



http://en.wikipedia.org/wiki/Translation_%28biology%29

Rys. 72. Animacja prezentująca schematycznie syntezę białka. To nie jest trajektoria MD tylko rodzaj naukowej animacji komputerowej.

Ważne: Przepływ informacji genetycznej od DNA → RNA → białko nazywa się centralnym dogmatem biologii molekularnej.



Warto też wiedzieć, że po syntezie łańcucha białkowego może on niekiedy podlegać dalszej obróbce, np. niektóre aminokwasy mogą być fosforyzowane.

Omówienie Centralnego Dogmatu Biologii Molekularnej było przedmiotem naszego 1-go wykładu. Rozumiemy teraz, że zaburzenia w genach (sekwencji DNA) mogą się odbić na zaburzeniach w białkach, a to z kolei może prowadzić do rozmaitych patologii czy chorób. Zmiany letalne powodują śmierć komórki, ale są też takie zmiany, które prowadzą do nieśmiertelności komórki w postaci niekończących się podziałów. W ten sposób wygląda mechanizm powstawania niektórych nowotworów. Powstaje cały czas otwarte pytanie jak temu zapobiegać, czyli jak zatrzymać niekontrolowany wzrost, jak chronić się przez „falszowaniem” informacji genetycznej przez czynniki zewnętrzne, np. promienie UV (opalenie - czerniak) czy dym papierosowy (-> rak płuc).

Zdrowie jest niezwykle pożądanym dobrem, zatem bogate społeczeństwa lożą ogromne środki na to, by poznać molekularne przyczyny chorób i by szukać leków na te patologie. Ponieważ cząsteczki biologiczne (m.in. białka, DNA) są małe - mają rozmiary rzędu nanometrów - powstałą całą dziedziną ich badań praktycznych nazywa się **bionanotechnologią**. Wykorzystuje ona nie tylko metody fizyczne, takie jak dyfrakcja promieni X, NMR czy AFM, ale również symulacje komputerowe.

Warto wiedzieć, że współczesne metody biologii molekularnej pozwalają zmusić niektóre organizmy żywe (bakterie, drożdże) do produkcji niemal dowolnego białka. Potrafimy też przeprowadzać metodami enzymatycznymi w zasadzie dowolną „edycję” nici DNA. Zrobotyzowane i skomputeryzowane maszyny pozwalają określić skład biopolimerów, zaś analiza bioinformatyczna pomaga optymalizować właściwości fizykochemiczne. W ramach bionanotechnologii próbuje się mieszać biopolimery z innymi ciekawymi materiałami lub obiektami, np. kropkami kwantowymi, szukając nowych, niezwykle obiektów, np. układów zdolnych efektywnie przetwarzać światło na prąd elektryczny. Przedmiotem dalszej części wykładów będzie prezentacja metod komputerowej (obliczeniowej) analizy właściwości strukturalnych, elektronowych i mechanicznych takich nanomateriałów.

Podstawy formalne metod modelowania.

Wprowadzenie

Idea symulacji jest dość prosta: Tworzymy model matematyczny biopolimeru (np. białka, kwasu nukleinowego, fosfolipidów czy błony, hormonów, etc.). Model ten musi być oparty na doświadczalnych danych uzyskanych metodami fizycznymi: musi uwzględniać rozmiary i masy atomów, ładunki elektryczne, efekty otoczenia wodnego – np. obecność jonów czy pH środowiska, temperaturę, naturę oddziaływań międzymolekularnych, ciśnienie.

Celem modelowania może być poszukiwanie optymalnych struktur tych biopolimerów, a można też badać jak szybko ich przestrzenne struktury (konformacje) się zmieniają. Można śledzić jak jedno białko oddziałuje z innym. Testuje się przy pomocy modeli co się stanie z polem elektrycznym wytwarzanym przez enzym, jeśli w jednym miejscu nastąpi zamiana jednego, danego aminokwasu na inny (mutacja) – czy efekt będzie duży, mały czy żaden. Można do takiego modelu białka

dopasowywać (dokować) małe cząsteczki chemiczne i wyszukiwać te, które łączą się szczególnie mocno. Mogą to być potencjalne leki. Można projektować białka o określonych (pożądanych) właściwościach fizykochemicznych. Można próbować nadać znanym białkom enzymatycznym inne funkcje biologiczne. Można też komputerowo testować mechanizmy katalityczne różnych enzymów, np. tych, które służą do wytwarzania piwa, czy substancji organicznych (akrylamid) przydatnych powszechnie w przemyśle i gospodarce.

Modelowanie komputerowe biopolimerów jest powszechnie stosowaną na całym świecie metodą badawczą pozwalającą uzyskać stosunkowo szybko nowe i normalnie niedostępne informacje o tych ważnych związkach chemicznych. Podejście to przyczyniło się wydatnie do wynalezienia kilku nowych leków, m.in. przeciwgrypowych, ulepszenia testów diagnostycznych, czy zaproponowania nowych procesów biotechnologicznych (zielona chemia). Dzięki modelowaniu nanobioukładów rozumiemy znacznie lepiej biologię molekularną komórki, procesy rozmnażania, ewolucji, metabolizm, fotosyntezę, procesy chorobowe. Poszukuje się też nowych materiałów, np. hybrydowych, podpatrując i wykorzystując rozwiązania wynalezione w toku milionów lat ewolucji (biomimetyka).

Obliczenia wymagają dużej wiedzy fizycznej, informatycznej i sporej wiedzy matematycznej. Komercyjne programy do modelowania komputerowego cząsteczek kosztują nawet setki tysięcy dolarów. Firmy piszące takie kody w okresie maksymalnego rozwoju zatrudniały nawet po 200 osób z doktoratami w tym sporo informatyków.

Obecnie na rynku dominuje kilka firm, jest też b. dużo łatwo dostępnego oprogramowania akademickiego, można z niego korzystać do celów badawczych czy edukacyjnych za darmo. Poznamy podstawy koncepcyjne głównych metod obliczeniowych i zapoznamy się na ćwiczeniach z dwoma popularnymi programami **ArgusLab** (obliczenia kwantowe i mechanika molekularna) oraz **NAMD+VMD** (dynamika molekularna).

„Mapa” metod modelowania molekularnego – metody kwantowe a klasyczne

W modelowaniu komputerowym układów chemicznych (cząsteczek lub ich kompleksów) czy biopolimerów interesują nas głównie dwa problemy:

- (a) jaka jest struktura przestrzenna cząsteczki ?
- (b) jak oddziałuje ona z innymi cząsteczkami lub otoczeniem ?

Chcemy znać odpowiedź na te pytania w oparciu o obliczenia, nie pomiary, które są zwykle bardzo kosztowne i trudne do wykonania. Teoria fizyczna właściwa do udzielenia tych odpowiedzi jest mechanika kwantowa.

Biocząsteczki są jednak zbyt duże – zawierają zbyt wiele elektronów by można było zajmować się każdym elektronem z osobna. Zatem stosuje się przybliżenia, w których najmniejszym obiektem analizowanym dynamicznie jest atom (reprezentowany punktem materialnym – nieskończeni małą kulką o określonej masie) lub nawet grupa atomów.

Metody kwantowe opisuje się w literaturze pewnymi akronimami odnoszącymi się do stosowanych przybliżeń. W molekularnej mechanice kwantowej (chemii kwantowej) spotykamy często podział podawany poniżej:

Metody molekularnej mechaniki kwantowej:

- „ab initio” – oparte na zasadzie wariacyjnej lub rachunku zaburzeń [np. metoda Hartree-Focka (HF), metoda oddziaływania konfiguracyjnego (CI), metoda wielociałowego rachunku

zaburzeń (MBPT), metoda sprzężonych klasterów (CC, CC SD), metoda rachunku zaburzeń MP2)

- DFT – Density Functional Theory (np. DFT B3LYP, LSDA, TD DFT i inne)
- Metody półempiryczne (metoda Huckla, metoda PPP, metoda TB, metoda AM1, PM3, PM6, metoda ZINDO).

Metody różnią się mocno stopniem „ściśłości” i przyjmowanymi przybliżeniami ale każda z nich ma swoje zalety i wady. Zaletą metod półempirycznych (AM1, PM3, MNDO) jakie poznamy w ramach pakietu ArgusLab jest szybkość obliczeń, niezła jakość wyników jeśli chodzi o optymalizację geometrii średniej wielkości molekuł, stosowalność do stosunkowo dużych układów (kilkaset atomów).

Główne pakiety do obliczeń kwantowych to Gaussian09, Gamess, NWChem, VASP, Jaguar, Hyperchem. Koszt niektórych licencji liczony jest w tysiącach dolarów rocznie.

Strukturę przestrzenną cząsteczki, istotna do oceny jej funkcji biologicznej czy technologicznej, wyznacza takie ułożenie atomów (odległości między nimi, kąty), które odpowiada minimum tzw. energii potencjalnej tego układu. Poniżej przedstawimy wyjaśnienie skąd bierze się koncept powierzchni energii potencjalnej i jak można wyznaczyć minima na tej powierzchni, Jest to bardzo fundamentalne pojęcie w dziedzinie modelowania komputerowego bio- i nano-struktur.

Idea powierzchni energii potencjalnej (PES)

Całkowita informacja o cząsteczce zawarta jest w jej funkcji falowej, którą tutaj w notacji Diraca oznaczymy symbolem $|\Phi\rangle$. Funkcję tę można uzyskać rozwiązując równanie Schroedingera:

$$\hat{H}|\Phi\rangle = \varepsilon|\Phi\rangle$$

Hamiltonian \hat{H} (operator energii) zależy tutaj zarówno od elektronowych (i,j) jak i jądrowych (A,B) stopni swobody (stosujemy jednostki atomowe):

$$\hat{H} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{j=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{i>j}^M \frac{1}{r_{ij}} + \sum_A^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}}$$

Jednostki atomowe (a.u.):

$$\begin{aligned} \hbar &= 1 \\ "c = 1" & \quad (137.) \\ m &= 1 \\ e &= 1 \end{aligned}$$

W rozważaniach dalszych przyjmuje się zwykle tzw. przybliżenie Borna-Oppenheimera, w którym zakłada się, że jądra atomów są nieskończenie masywne zatem pozostają nieruchome. Wówczas możliwe jest rozseparowanie elektronowych i jądrowych stopni swobody:

$$\hat{H}_{el} \Phi_{el} = \varepsilon_{el} \Phi_{el}$$

Tutaj Hamiltonian elektronowy ma postać:

$$\hat{H}_{el} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{j=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{i>j}^M \frac{1}{r_{ij}}$$

natomiast elektronowa funkcja falowa Φ_{el} zależy od położenia atomów $\{\vec{R}_A\}$ w sposób parametryczny, co symbolizuje średnik w zapisie wzoru:

$$\Phi_{el} = \Phi_{el}(\{\vec{r}_i\}; \{\vec{R}_A\})$$

$$\varepsilon_{el} = \varepsilon_{el}(\{\vec{R}_A\})$$

Widać, że energia elektronowa badanej biomolekuły ε_{el} zależy od położenia jąder atomów. Dla każdego indywidualnego ustawienia położenia jąder mamy inną funkcję falową Φ_{el} (zależną od współrzędnych elektronowych \vec{r}_i). Przy ustalonych położeniach atomów (odległościach R_{AB}) energia całkowita cząsteczki ma następującą postać:

$$\varepsilon_{tot}^{fix N} = \varepsilon_{el} + \sum_A \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}}$$

Hamiltonian jądrowy z kolei opiszemy wzorem \hat{H}_{nucl} :

$$\hat{H}_{nucl} = -\sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 + \langle \hat{H}_{nucl} \rangle_{\Phi_{el}} + \sum_A \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} =$$

$$-\sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 + \underbrace{\varepsilon_{el}(\{R_A\}) + \sum_A \sum_B^M \frac{Z_A Z_B}{R_{AB}}}_{\varepsilon_{tot}^{fix N}(\{R_A\})} =$$

$$-\sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 + \varepsilon_{tot}^{fix N}(\{R_A\}).$$

Widzimy w tym wyrażeniu człon $\varepsilon_{tot}^{fix N}$ (dla ustalonych położenia atomów – fix) stanowi potencjał, w którym poruszają się jądra atomowe. Tę właśnie wielkość nazywa się powierzchnią energii potencjalnej (PES ma wymiar energii). Energia potencjalna zależy od współrzędnych wszystkich atomów. W metodach klasycznych, np. omawianych dalej metodach dynamiki molekularnej, przybliżana jest wzorami klasycznymi (pola siłowe). Pola siłowe opracowywane są dla różnych nanoukładów, inne są dla kryształów (ciało stałe) inne dla białek i inne dla DNA.

Zauważmy, że rozwiązując metodami mechaniki kwantowej jądrowe równanie Schroedingera:

$$\hat{H}_{nucl} \Phi_{nucl} = \varepsilon_{nucl} \Phi_{nucl}$$

uzyskuje się informacje o drganiach, obrotach i translacjach cząsteczki. Te same informacje można uzyskać też drogą klasyczną, bez odwoływania się do mechaniki kwantowej, o ile znane są "rozsądne" wzory na powierzchnie energii potencjalnej. Wyrażenia takie zostały opracowane w oparciu o dane doświadczalne i obliczenia kwantowe i są dostępne albo „za darmo” albo w programach komercyjnych.

Pola siłowe

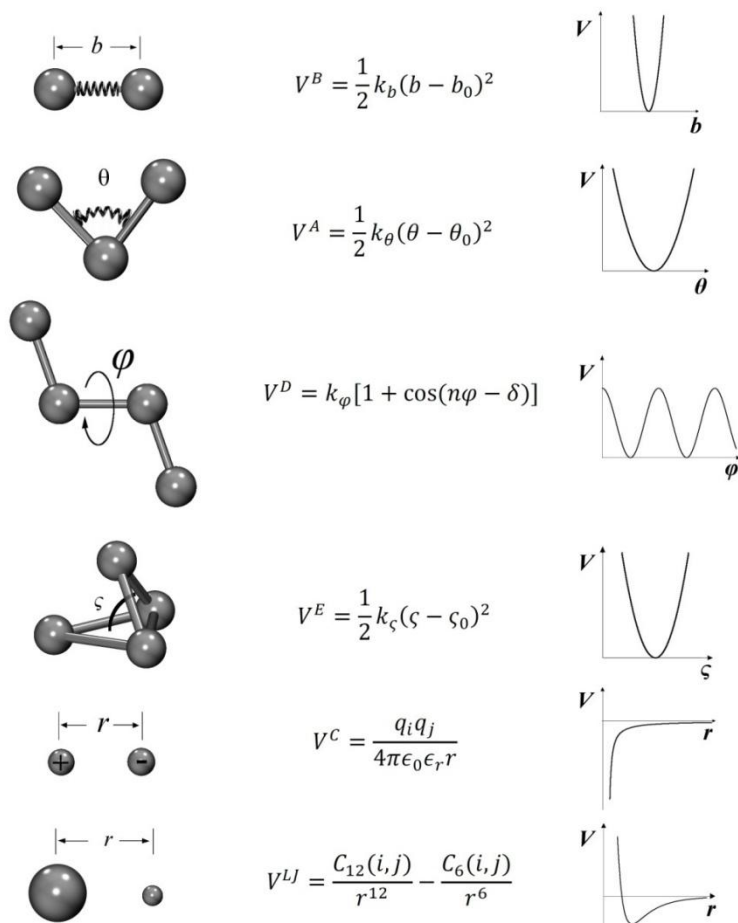
Podstawą wykorzystania pól siłowych w modelowaniu nanoukładów jest założenie, że klasyczne pola siłowe w postaci sparametryzowanych wzorów analitycznych są dobrymi przybliżeniami do prawdziwych powierzchni energii potencjalnej. Bardzo ważne jest też kolejne założenie: przyjmuje się że parametry opisujące poszczególne ugrupowania chemiczne, np. dany aminokwas, są przenaszalne z jednej biocząsteczki do innej, też zawierającej ten aminokwas.

Ogólne wyrażenie na pole siłowe jest następujące (uwaga są też bardziej złożone wzory!) :

$$V = \sum_{\text{bonds } i} V_i^B + \sum_{\text{bond angles } j} V_j^A + \sum_{\text{torsional angles}} V_k^D + \sum_{\text{improper angles } l} V_l^E + \sum_{\text{atom pairs } (r,s)} (V_{r,s}^C + V_{r,s}^P + V_{r,s}^{vdW})$$

Znaczenie fizyczne poszczególnych członów podano w podpisie do rysunku poniżej. Wyrazy te odpowiadają energiom wiązań chemicznych (V_i^B), energiom związanym z deformacjami kątów między wiązaniami (V_j^A , V_k^D , V_l^E) oraz dwuciałowym oddziaływaniami pomiędzy wszystkimi parami atomów obecnymi w badanym systemie. Te oddziaływania niewiążące składają się na energię kulombowską, związaną z występowaniem ładunków cząstkowych na atomach (V^C) oraz tzw. oddziaływaniami Van der Waalsa ($V^P + V^{vdW}$).

Warto też wiedzieć, że w modelowaniu stosuje się pola siłowe tzw. „*all atom*” czyli takie, w których występują indywidualne atomy, ale także, dla oszczędności czasu komputerowego, bardziej przybliżone modele, gdzie grupy atomów (np. cały aminokwas czy zasada DNA) stanowią centra potencjałów. Są to tzw. modele gruboziarniste i ich znaczenie ostatnio się zwiększa.



Rys. 73. (wg: H.Grubmueller, "Proteins as Molecular Machines: Force Probe Simulations", published in Computational Soft Matter: From Synthetic Polymers to Proteins, Lecture Notes, Norbert Attig, Kurt Binder, Helmut Grubmueller, Kurt Kremer (Eds.), John von Neumann Institute for Computing, Julich, Germany, NIC Series, Vol. 23, ISBN 3-00-012641-4, pp. 401-422, 2004. © 2004 by John von Neumann Institute for Computing), Oddziaływania składające się na pole siłowe. Drgania atomów wzdłuż wiązań opisane są potencjałem harmonicznym V^B , minimum potencjał osiąga na odległości równowagowej b_0 pomiędzy atomami połączonymi wiązaniem chemicznym. Kąty płaskie i tzw. kąty niewłaściwe (wychylenia poza płaszczyznę) są też opisywane potencjałami harmonicznymi, V^A i V^E , gdzie θ_0 oraz ζ_0 oznaczają odpowiednie kąty równowagowe. Kąty dwuścienne (torsyjne) opisywane są potencjałami periodycznymi V^D ; stałe siłowe oznaczone są symbolami k z odpowiednimi indeksami. V^C oznacza oddziaływanie kulombowskie, zaś potencjały Lennarda-Jonsa $V^{LJ} = V^P + V^{vdW}$, składają się z odpychania Pauliego, $V^P \sim r^{-12}$ oraz przyciągania van der Waalsa, $V^{vdW} \sim -r^{-6}$.

Popularne w modelowaniu białek i DNA pola siłowe to CHARMM, AMBER, GROMOS, OPLS. Odnośniki do ich pełnego opisu można znaleźć w Wikipedii. My będziemy na ćwiczeniach stosować pole CHARMM v.27.

Optymalizacja geometrii (=minimalizacja energii)

Wg tzw. hipotezy Anfinsena białka w stanach natywnych (fizjologicznych i stabilnych) znajdują się w minimum swojej powierzchni energii swobodnej F . Energia swobodna $F=E-TS$. Jeżeli pominiemy czynniki entropowe (S), co w wielu sytuacjach jest uprawnione, to znalezienie natywnego stanu białka (biopolimeru) sprowadza się do znalezienia optymalnej (minimalnej) energii. Odpowiadamy zatem na pytanie: przy jakim wzajemnym ustawieniu atomów energia E biomolekuły jest najmniejsza. Ponieważ E jest funkcją położeń w przestrzeni wszystkich atomów, zależy też od oddziaływania molekuły z otoczeniem, pytanie nie jest trywialne: jest to problem minimalizacji funkcji skalarnej naprawdę wielu zmiennych. Na przykład dla hemoglobiny możemy mieć łatwo w systemie (modelu) 20 tys. atomów co daje ok. 60 tys. zmiennych (parametrów) do optymalizowania! Sprawa nie jest jednak beznadziejna, i chociaż nie ma metody gwarantującej znalezienie globalnego minimum takiej funkcji, to znalezienie przy pomocy algorytmów minimum lokalnego jest w praktyce możliwe.

Minimalizacja „brute force”

Przy niewielkiej liczbie stopni swobody skanowanie całej przestrzeni konformacyjnej (np. co 10 stopni w kątach i co 0.1 Ang. w odległościach) jest wykonalne, ale dla interesujących nas modeli jest to błędna ścieżka, bowiem liczba struktur do obliczenia energii rośnie niemal wykładniczo.

Metody najszybszego spadku

Przedstawimy metodę najszybszego spadku (SD) na przykładzie funkcji jednej zmiennej $f(x)$.

1. Wybieramy punkt startowy x_0
2. W punkcie $x_k = x_0$ obliczamy gradient funkcji i „opatrujemy go” znakiem minus:

$$-\nabla f(x_k)$$

3. Kolejny punkt obliczamy wg wzoru:

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k)$$

4. Wielkość kroku w wybranym kierunku ustalamy dokonując minimalizacji kierunkowej (liniowej), czyli tak zmienimy czynnik γ_k , aby spełnić warunek:

$$f(x_k - \gamma_k \nabla f(x_k)) = \min_{\gamma > 0} f(x_k - \gamma \nabla f(x_k))$$

5. Iterujemy punkty 2-4 tak długo, aż spełnione będzie kryterium stopu:

$$\nabla f(x_k) \leq \varepsilon \text{ oraz } f(x_k) - f(x_{k-1}) \leq \varepsilon'$$

Schemat ten można uogólnić łatwo na wektory w przestrzeni wielowymiarowej.

Lepszą efektywność niż w SD uzyskuje się stosując **metodę sprzężonych gradientów** (kierunków, CG). W tym podejściu kierunek minimalizacji określamy nie bezpośrednio z gradientu, ale z „mieszanki” kierunku aktualnego kierunku gradientu i poprzedniego kierunku

1. Znajdź największą pochyłość
2. Jeśli pochyłość $< \epsilon$, zakończ poszukiwanie
3. Jeżeli jesteś w kroku nr 1 – zapamiętaj kierunek schodzenia (będzie to kierunek gradientu dla kroku nr 1)
4. Dla numeru kroku >1 wyznacz kierunek spadku będący kombinacją poprzedniego kierunku i kierunku najszybszego spadku w danym punkcie (sprzężone kierunki).
5. Poszukaj minimum w tym kierunku. Przejdź do tego minimum.
6. Idź do pkt. 1.

Są różne warianty sprzęgania tych kierunków i stąd różne nazwy metod, np. Fletcher-Reeves, Polak-Ribiere, Hestenes-Steifel (więcej zob. np. :

http://www.mymathlib.com/optimization/nonlinear/unconstrained/fletcher_reeves.html

Algorytm CG nie wymaga specjalnie dużej pamięci, jest stosunkowo szybko zbieżny, ale nie dla struktur dalekich od minimum. Jeśli wiemy, że struktura jest daleko od minimum, lepiej jest zastosować na początku algorytm SD.

Wadą opisanego podejścia w metodach mechaniki molekularnej jest konieczność liczenia gradientu funkcji energii, która ma na ogół złożony charakter i wymaga różniczkowania numerycznego. Opracowano zatem w 1976 roku inny algorytm, który ma cechy algorytmu CG, a nie wymaga liczenia prawdziwego gradientu – algorytm Powella.

(Zob. Press, WH; Teukolsky, SA; Vetterling, WT; Flannery, BP (2007). "[Section 10.7. Direction Set \(Powell's\) Methods in Multidimensions](#)". *Numerical Recipes: The Art of Scientific Computing* (3rd ed.). New York: Cambridge University Press.

lub <http://people.equars.com/marco/poli/phd/node55.html>)

Metody II rzędu (Newtona-Raphsona, N-R)

Szybszą zbieżność (w jednym kroku dla powierzchni drugiego stopnia) zapewniają metody, w których wykorzystuje się wiedzę o drugich pochodnych PES.

Przyjmijmy oznaczenia:

Niech \mathbf{x}_k oznacza punkt w przestrzeni konfiguracyjnej w k-tym kroku

\mathbf{d}_k – kierunek minimalizacji w k-tym kroku, -gradient:

$$-\nabla f(\mathbf{x}_k) = -\mathbf{g}_k$$

ω, λ – współczynniki skalarne

H – macierz drugich pochodnych (tzw. Hessian)

V – macierz przybliżona do odwrotności Hessianu

Wówczas wg metody N-R możemy wyznaczyć ekstremum (minimum) z wzoru:

$$x_{min} = x_0 - H_0^{-1} \nabla f(x_0)$$

Uzasadnienie wzoru wykracza poza zakres tego wykładu. Poniżej podajemy jedynie prosty, jednowymiarowy przykład że taki wzór faktycznie działa:

Przykład (jednowymiarowy, „prim” i „bis” oznaczają kolejne pochodne)

$$f(x) = a + bx + cx^2$$

$$f'(x) = b + 2cx$$

$$f''(x) = 2c$$

W minimum mamy $f'(x_{min}) = 0$ zatem $b + 2cx_{min} = 0$ czyli $x_{min} = -b/2c$

Stąd

$$x_{min} = (-f'(x) + 2cx) / 2c = x - f'(x) / f''(x).$$

W praktyce obliczanie drugich pochodnych, zwłaszcza dla tak dużych systemów jak biopolimery, jest niepraktyczne, bo wymaga bardzo dużej pamięci RAM. Opracowano zatem przybliżone metody, np. tzw. quasi Newtona (Davidona-Fletcher-Powella czy Broydena-Fletcher-Goldfarb-Fano), są one stosowane np. w programie ArgusLab do optymalizacji geometrii metodami półempirycznymi mechaniki kwantowej.

Metody całkowania równań ruchu (Newtona, Langevina)

W metodzie symulacji dynamiki bioukładu przy pomocy mechaniki klasycznej (dynamika molekularna) zakłada się, że ewolucję w czasie systemu da się opisać przy pomocy równań Newtona:

$$a_i = F_i / m_i$$

Chcąc określić położenie atomu i w chwili czasu t musimy znać jego masę m_i , siłę działającą na ten atom $F_i(t)$ oraz scałkować dwukrotnie powyższe równanie po czasie. Operację tę należy powtórzyć kolejno dla każdego atomu w układzie, czasem 20 tys., a czasem nawet 200 tys. razy, bo aż tak duże biomolekuły się obecnie symuluje. Znając wyrażenie na pole siłowe (przypomnijmy: energia E , lub czasem oznaczana V , zależy od aktualnych położeń $r(t)$ lub $x(t)$ wszystkich atomów badanego układu)

możemy poprzez obliczenie gradientu z tej funkcji (w punkcie przestrzeni konfiguracyjnej określonym aktualną „potrzebą” algorytmu i odpowiadającym danej chwili czasowej) znaleźć potrzebne siły F_i :

$$F_i = -\nabla E(\mathbf{x})$$

Znalezienie tzw. trajektorii atomu czyli funkcji $x(t)$ wymaga scałkowania dwukrotnego po czasie podanego wyrażenia na przyspieszenie a_i . W metodach symulacji komputerowej poszukuje się jak najefektywniejszego algorytmu realizacji tego całkowania. Oczywiście metod numerycznych jest b. dużo, ale w tej dziedzinie, ze względu na swoją prostotę, stabilność numeryczną i niewielkie wymagania pamięci największą popularność zdobył algorytm Verleta. Omówimy kilka jego wariantów. Czas w tym podejściu ulega dyskretyzacji. Wybór kroku czasowego całkowania musi być podyktowany znajomością badanego układu: zbyt duży krok czasowy da nam fałszywą trajektorię i w końcu „eksplozję” modelowanego układu, zbyt mały będzie generował zbędne koszty (CPU). W modelowaniu białek stosuje się prostą regułę: delta t nie powinien być mniejszy niż 1/10 okresu najszybszego modelowanego zjawiska periodycznego. W praktyce są to oscylacje układu atomów C-H, a stosowany krok czasowy to 1 fs, lub niekiedy 2 fs. W symulacjach na ćwiczeniach warto osobiście sprawdzić jak obliczone trajektorie zależą od kroku czasowego. Jeśli procedura symulacyjna nagle „pada” – obliczenia przerywają się bez istotnej przyczyny, zaś w pliku wynikowym obserwujemy gwałtowny wzrost energii badanego układu, warto wówczas powtórzyć obliczenia z krokiem mniejszym o połowę.

Chcemy prowadzić symulacje białcząsteczki w ruchu. Ruch atomów odpowiada pewnej temperaturze równowagowej systemu i otoczenia. Informacja o temperaturze zawarta jest w prędkościach poszczególnych atomów. Możemy tutaj mówić o wielkości statystycznej jaką jest temperatura stosunkowo bezpiecznie, bowiem badane układy (białko wraz z wodą) zawierają tysiące atomów. Najpierw wybiera się (losuje) prędkości początkowe, tak by całkowity pęd układu był równy zero:

$$\mathbf{P} = \sum_{i=1}^N m_i \mathbf{v}_i = 0$$

Kierunki wektorów prędkości są losowe. Wartości składowych prędkości atomów są generowane losowo z rozkładu Maxwella-Boltzmanna przy zadanej temperaturze – T :

$$p(v_{xi}) = \sqrt{\frac{m_i}{2\pi k_B T}} \exp\left[-\frac{m_i v_{xi}^2}{2k_B T}\right]$$

$$T = \frac{1}{3N} \sum_{i=1}^N \frac{|p_i|^2}{m_i}$$

Algorytm „położeniowy” Verleta korzysta z położenia (i przyspieszeń) atomów zarejestrowanych w chwilach t i $t - \delta t$ po to by wyznaczyć nowe położenia w chwili $t + \delta t$. Wzór na nowe położenia można

uzyskać bardzo prosto, rozwijając w szereg Taylora względem czasu do przodu i „do tyłu” wyrażenie na położenie $r(t)$:

$$\left. \begin{aligned} r(t + \delta t) &= r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \\ r(t - \delta t) &= r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \end{aligned} \right\} +$$

Po obcięciu rozwinięcia do II rzędu i dodaniu stronami obu wzorów otrzymujemy prosty przepis na propagację położenia atomów:

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2$$

Algorytm „Leap-frog”

Popularnym wariantem „położeniowego” Verleta jest tzw. algorytm „leap-frog” (przeskakiwanka). W algorytmie *leap-frog* **prędkości** są liczone dla czasu $t + \delta t/2$ i są następnie wykorzystywane do obliczania **położeń** w chwili $t + \delta t$.

$$\left\{ \begin{aligned} v\left(t + \frac{1}{2}\delta t\right) &= v\left(t - \frac{1}{2}\delta t\right) + a(t)\delta t \\ r(t + \delta t) &= r(t) + v\left(t + \frac{1}{2}\delta t\right)\delta t \end{aligned} \right.$$

Zaletą tego algorytmu jest możliwość śledzenia prędkości atomów (mamy już przestrzeń fazową, a nie tylko konfiguracyjną), jednak zauważmy, że nie są one wyznaczone w tej samej chwili co położenia (przesunięcie o pół kroku czasowego) oblicza się je lepiej ze wzoru:

$$v(t) = \frac{1}{2} \left[v\left(t - \frac{1}{2}\delta t\right) + v\left(t + \frac{1}{2}\delta t\right) \right]$$

Znany jest też tzw. wariant „prędkościowy” algorytm Verleta:

$$\left\{ \begin{aligned} r(t + \delta t) &= r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \\ v(t + \delta t) &= v(t) + \frac{1}{2}[a(t) + a(t + \delta t)]\delta t \end{aligned} \right.$$

Praktyczne aspekty symulacji dynamiki molekularnej. Badanie wytrzymałości biomateriału nanokompozytowego.

Temperatura w symulacjach MD

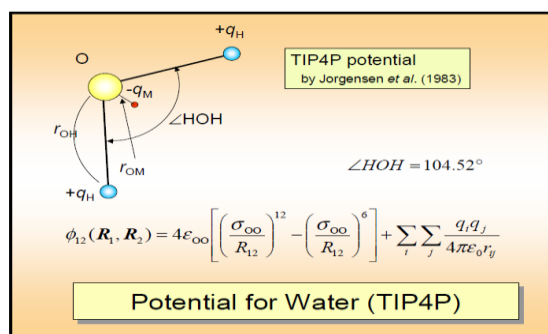
Ważnym problem w czasie całkowania równań Newtona jest zachowanie odpowiedniego zespołu statystycznego. Możemy żądać, m.in., zachowania energii całkowitej układu (zespół kanoniczny) lub np. zachowania stałej temperatury (tzw. zespół mikrokanoniczny, zakładamy kontakt z rezerwuarem termicznym). Omówimy ten ostatni przypadek, bowiem na ćwiczeniach takie obliczenia będziemy realizować. W czasie obliczeń numerycznych, z powodu przybliżeń, a nawet natury problemu, temperatura zwykle się zmienia. Zachowanie stałej temperatury można zapewnić, m.in., poprzez sprzężenie z termostatem na poziomie równań (metoda Langevina, taki protokół jest proponowany na ćwiczenia, omówienie teorii tego sposobu jest zbyt skomplikowane jak na czas tego wykładu). Inny sposób to proste skalowanie prędkości. Z zasady ekwipartycji energii wiemy, że na każdy stopień swobody układu mechanicznego będącego w równowadze termicznej z otoczeniem przypada energia równa $1/2kT$, gdzie k – stała Boltzmanna. Jeśli mamy N atomów w układzie, to zachodzi zależność:

$$E_{kin} = \frac{3}{2} NkT = \sum_{i=1}^N \frac{m_i v_i^2}{2}$$

Widzimy od razu, że możemy regulować temperaturę systemu poprzez odpowiednie skalowanie wartości prędkości atomów (zwiększanie lub zmniejszanie v).

Problem otoczenia

Symulacje biopolimerów przeprowadza się zwykle dla środowiska wodnego. W fizyce funkcjonuje kilkanaście różnych komputerowych modeli przybliżonych wody, żaden nie jest idealny (http://en.wikipedia.org/wiki/Water_model , Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K, Pekka Mark and Lennart Nilsson , *J. Phys. Chem. A*, 2001, 105 (43), pp 9954–9960). Cząsteczka ta jest szczególnie trudna do opisu klasycznego. Na ćwiczeniach stosować będziemy popularny model TIP3P, przedstawiony poniżej model TIP4P (rys. z Internetu: u-tokyo/~murayama)) jest podobny:



Rys. 74. Zasada budowy modelu komputerowego wody.

Można zanurzyć badany układ w „kropki” komputerowej wody, ale można też zasymulować środowisko kryształu poprzez odpowiedni dobór periodycznych warunków brzegowych (PBC). Przestrzeń dzieli się np. na 27 przylegających sześcianów (jak w kostce Rubika), centralny sześcian zawiera symulowany układ i wodne otoczenie, zaś 26 pozostałych jest jego kopiami uzyskanymi przez odpowiednie operacje symetrii. Jeśli w wyniku symulacji jakiś atom wychodzi poza ścianki głównej skrzynki, to wchodzi on do niej po przeciwnej stronie. Atomy „wychodzą” z odpowiedniej skrzynki obrazowej. W ten sposób nic w modelu nie ginie, zaś cząsteczka białka siedząca w środku centralnej części „czuje” otoczenie kwasu-kryształu.

Głównym problemem technicznym dynamiki molekularnej są długie czasu symulacji. Istniejące kody zużywają ponad 70 proc. czasu CPU na obliczaniu oddziaływań niewiązanych (VdW i elektrostatycznych). Problem ten się nasila jak N^2 ze wzrostem wielkości układu. Ponieważ te oddziaływania słabną dość szybko z rosnącą odległością między atomami, w praktyce stosuje się zwykle tzw. promień odcięcia: oddziaływania atomów oddalonych więcej niż zadany próg (cut-off) po prostu się ignoruje. Oczywiście trzeba zadbać o to, by stosowane obcięte potencjały miały dobrą asymptotykę na dużych odległościach. W typowych symulacjach cut-off na poziomie 12-16 Ang jest wystarczający i znakomicie poprawia czasy symulacji.

Ogólny schemat MD

Ogólny schemat realizacji obliczeń MD (przykładowy) jest następujący:

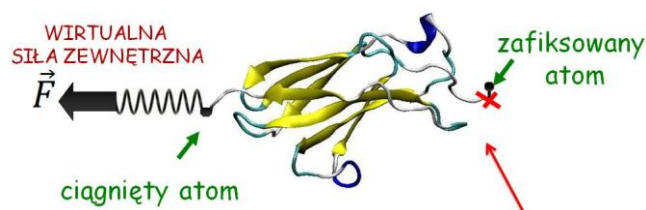
1. Wybór modelu (trzeba wiedzieć co chce się badać i po co, oraz skąd wziąć startowe współrzędne atomów badanego systemu)
 - a) Dobrym źródłem startowych struktur białek jest światowa baza: pdb.org.
 - b) Białko, po uzupełnieniu specjalnymi programami brakujących atomów, np. H, lub innych zanurza się w zrównoważonej skrzynce „komputerowej” wody, np. TIP3P.
 - c) Usuwa się cząsteczki wody znajdujące się zbyt blisko atomów białka (np. < 2.3 Ang).
 - d) Przeszukuje się kieszenie wypełnione wodą wewnątrz białka, zachowuje się jedynie te napotkane cząsteczki wody, które mają 3 wiązania wodorowe do atomów białka.
 - e) Określa się stopień jonizacji naładowanych aminokwasów (Glu, Asp, Arg, Lys, His) – pH
 - f) Sprawdza się, czy to „uprotonowanie” jest zgodne z założonym pH (kwasowością).
2. Określa się pole siłowe stosowane do symulacji (inne np. dla białek, inne dla DNA), sprawdza się czy są dostępne parametry pola siłowego dla grup specjalnych i niestandardowych (zwykle nie ma!).
3. Wybiera się wartości parametrów odcięcia oraz rozmiary pierwotnej skrzynki dla PBC. Określa się częstość regeneracji listy oddziaływań niewiązanych.
4. Dla ustalonych (zafiksowanych specjalnym potencjałem) położeń atomów białka wykonuje się ekwilibrację otoczenia wodnego (np. 500 ps- 1 ns, stopniowe podgrzewanie wody).
5. Podgrzewa się cały układ+woda, zdejmując stopniowo więzy i krokowo (np. co 10 K) podnosząc temperaturę.
6. Kontroluje się poprawność wyjściowej struktury na grafice.
7. Wykonuje się zadana liczbę kroków (czas kroku – zwykle 1 fs) dla celów zrównoważenia termicznego badanego układu, może to być np. 1 ns MD w temp. 300K.

8. Zbiera się dane z symulacji właściwej, zapisując na twardym dysku „ramki” (snapshots) z trajektorii w zadanych interwałach, np. co 100 lub 1000 fs. Kontroluje się parametry badanego układu w locie (temp. ciśnienie, energię i jej składowe).
9. Zapisane trajektorie (czasem nawet 20 Gb danych) ogląda się „na grafice” (np. programami VMD, Maestro, PyMol, Yasara) i analizuje (czasem wiele miesięcy) przy pomocy dostępnych programów (np. VMD) albo własnych specjalistycznych skryptów (np. w Pythonie czy języku Tcl).
10. Powtarza się obliczenia badanego systemu innym programem i/lub polem siłowym aby się (nieco) upewnić, czy zaobserwowane efekty nie są aby artefaktami metody i przybliżonego pola siłowego.

Badanie wytrzymałości biomateriału – włókien chruścika.

Niektóre owady, np. jedwabniki, pająki, chruściki potrafią w tajemniczy sposób wytwarzać bardzo silne nici. Służą one do zdobywania pożywienia (sieci łowne) lub do schronienia. Skład tych materiałów nie jest dobrze poznany. Zawierają one b. dużo białek, są też składniki mineralne. Jest to na ogół materiał kompozytowy o bardzo ciekawych właściwościach mechanicznych. Wytrzymałość niektórych lekkich nici pająków jest wyższa od kevlaru (kamizelki kuloodporne). Bionanotechnologia stosuje często podejście biomimetyczne polegające na podglądaniu natury. Tutaj byśmy chcieli zrozumieć dlaczego materiały naturalne mają tak doskonałe właściwości mechaniczne.

Możemy skorzystać z modelowania klasycznego, wariantu MD zwanego Sterowaną Dynamiką Molekularną (SMD). Idea jest bardzo prosta: do wybranego atomu białka (lub grupy atomów) dodaje się potencjał „zatrzymujący” te atomy w przestrzeni. Z kolei do innego atomu (np. na przeciwległym końcu białka) dodaje się dodatkowy atom „holujący” połączony z nim sprężyną o określonej stałej sprężystości k . Gdy atom „holujący” oddala się od atomu, do którego jest przywiązany, sprężyna się wydłuża i zgodnie z prawem Hooke’a i III zasadą dynamiki Newtona działa na ten atom z rosnącą siłą. Siła ta powoduje stopniowe rozplatanie białka, a my obserwujemy ten proces na komputerze. Możemy zbadać wytrzymałości mechaniczne różnych białek lub, w ramach tego samego modelu, fragmentów danego białka. Takie badanie będzie przedmiotem ćwiczeń.



Rys. 75. Idea SMD (rys. K. Mikulska).

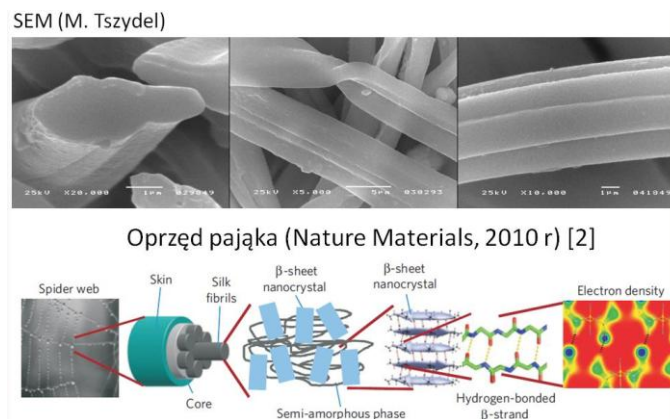
Badać będziemy fragment białka chruścika (Trichoptera), który potrafi sklejać pod wodą różne materiały. Być może zrozumienie konstrukcji tego materiału pomoże stworzyć nowe rodzaje opatrunków medycznych. Włókna te były badane doświadczalnie w ZBiFM IF UMK, m.in. przez WN i dr J. Strzeleckiego. Symulacje wykonywali mgr K.Mikulska i W.N.

(jest publikacja dla ciekawych: J. W. Strzelecki, J. Strzelecka, K. Mikulska, M. Trzydel, A. Balter, W. Nowak, *Nanomechanics of new materials – AFM and computer modelling studies of Trichoptera silk*, Cent. Eur. J. Phys. 9 (2011) 482 - 491.)

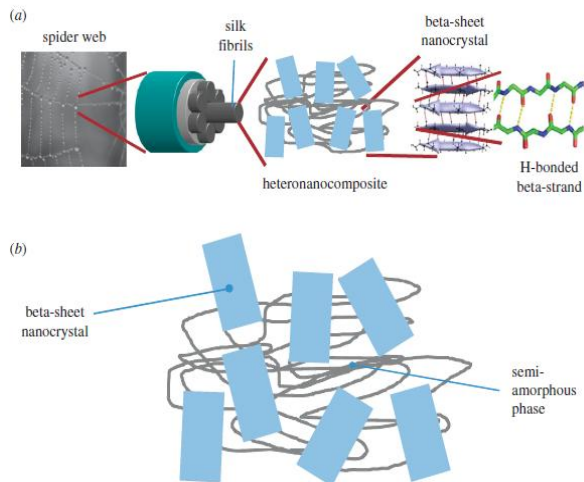


Rys. 76. Trichoptera potrafią wytwarzać nici spajające kamienie.

Warto popatrzeć na to, co co wiemy o budowie oprzędu chrzączki i pająka:



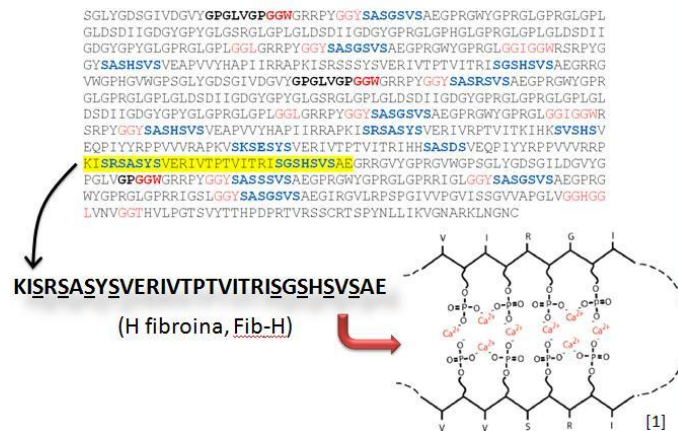
1710 *Nanomechanics of spider silk proteins* S. Keten and M. J. Buehler



Rys. 77. Oprzęd chrzączki (czy pająka) jest biomateriałem nanokompozytowym.

Do symulacji wybieramy tylko niewielki, ale bardzo charakterystyczny fragment fibroiny, głównego białka włókien chruścików (czy jedwabników):

Sekwencja H fibroiny

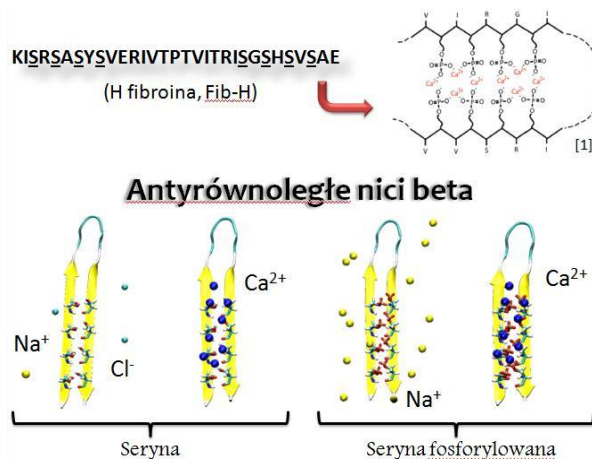


Rys. 78. Fragment sekwencji białka H-fibroiny, które będzie przedmiotem symulacji w ramach ćwiczeń.

Zwróćmy uwagę na skład tego fragmentu H-f: zawiera on 8 aminokwasów oznaczonych literą S – Seryna. Seryna ma grupę boczną OH i ulega fosforylacji, tzn. pewne enzymy dołączają do niej resztę kwasu fosforowego (to są te grupy PO₄ na rysunku). Reszta ta w wodzie ma ładunek ujemny. Jeśli chruścik zapewne sobie dostawę jonów wapnia Ca²⁺, to te dwuwartościowe jony mogą elektrostatycznie spajać przedstawione na rycinie fragmenty beta fibroiny. Spodziewamy się, że powinno to raczej wzmacniać materiał nici. Hipotezę tę będziemy testować w ramach ćwiczeń. Zastosujemy metodę SMD i spróbujemy rozzerwać nić w przynajmniej dwóch przypadkach:

- kiedy seryny są fosforyzowane i zawierają jony wapnia
- kiedy jest fosforylacja, ale wapnia nie ma.

Oczywiście powinniśmy dla jasności sytuacji wykonać jeszcze symulacje SMD w sytuacji, gdy seryny nie uległy fosforylacji, a wapń jest (3) albo go nie ma (4). Wszystkie te przypadki były przez nas przebadane, na ćwiczeniach wykonamy tylko wybrane obliczenia (4-6 h oznacza brak czasu na wszystkie przypadki):

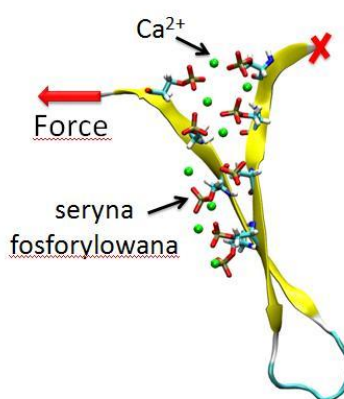


Rys. 79. Cztery modele fragmentu beta H-fibroiny, badanie roli fosforylacji i Ca²⁺ w nanomechanice nici.

Wybrać musimy parametry naszej symulacji SMD, pewne podpowiedzi są w materiałach do ćwiczeń:

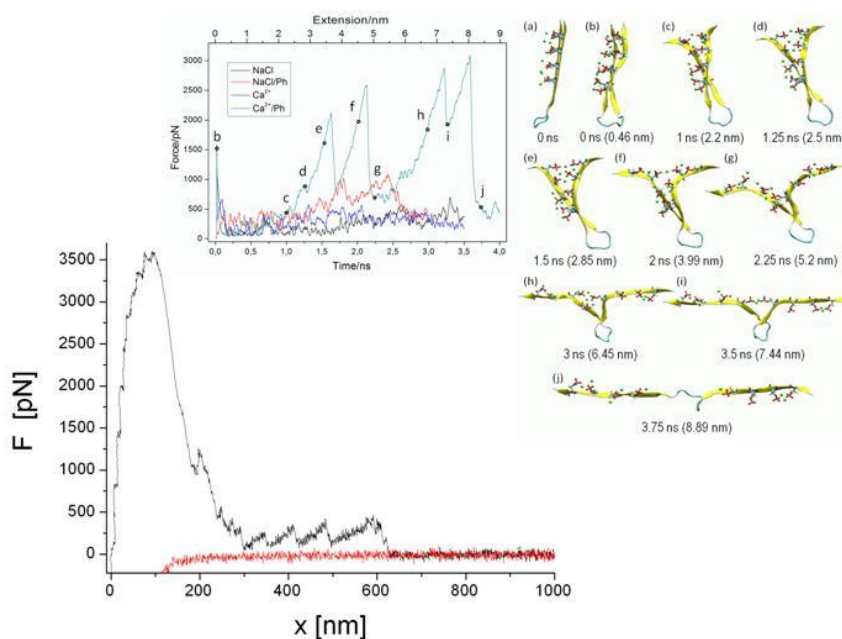
Wybieramy:

- atomy
- prędkość ciągnięcia
- stałą sprężystości
- kierunek ciągnięcia



Rys. 80. Sugerowane ustawienie atomów nieruchomych oraz kierunku siły.

Wyniki wcześniejszych symulacji trochę psują „suspens” naszych badań, ale lepiej jest zawsze wiedzieć co „powinno” wyjść z eksperymentu komputerowego:



Rys. 81. Wyniki niektórych symulacji SMD fragmentu H-f wykonanych przez KM i WN, warto zwrócić uwagę na to jakie są rejestrowane siły i jak silnie maksimum siły zależy od obecności (lub nieobecności) wapnia i fosforylacji seryn.

Nasze wnioski były następujące: Fosforylacja nieco (100%), a fosforylacja z wapniem znacząco (800%-1100% !) **podnosi** wytrzymałość mechaniczną badanego fragmentu białka. Ciekawe czy wyniki te uda się potwierdzić, albo odrzucić? Powodzenia w obliczeniach 😊.