

Uniwersytet Mikołaja Kopernika
Wydział Matematyki i Informatyki

Piotr Tadeusz Różański
nr albumu: 217724

Praca magisterska
na kierunku informatyka

Obliczenia własności ekscytonowych
o quasi-liniowej złożoności czasowej
dla nanostruktur półprzewodnikowych
zawierających
ponad pół miliona atomów

Opiekun pracy dyplomowej
dr Michał Zieliński
Instytut Fizyki UMK

Toruń 2012

Pracę przyjmuję i akceptuję

Potwierdzam złożenie pracy dyplomowej

.....
data i podpis opiekuna pracy

.....
data i podpis pracownika dziekanatu

*Uniwersytet Mikołaja Kopernika zastrzega sobie prawo własności niniejszej pracy
magisterskiej w celu udostępniania dla potrzeb działalności naukowo-badawczej lub
dydaktycznej.*

Spis treści

Część 1. Wprowadzenie	1
Rozdział 1. Nanostruktury	3
1. Wybrane rodzaje nanostruktur	3
2. Optyczne własności kropek kwantowych	4
Rozdział 2. Obliczenia atomistyczne	7
1. Metoda ciasnego wiązania	7
2. Elementy macierzowe operatora Coulomba	8
3. Formalizm macierzowy	9
4. Podsumowanie metody atomistycznej	10
Część 2. Metoda	11
Rozdział 3. Rekonstrukcja funkcji falowej	13
1. Wybór bazy funkcyjnej	13
2. Ustalenie parametrów rekonstrukcji	18
3. Normalizacja funkcji falowej	21
Rozdział 4. Dyskretyzacja oddziaływania	23
1. Oddziaływanie kulombowskie	23
2. Ekranowanie dielektryczne	24
3. Zasięg oddziaływania	25
Rozdział 5. Obliczenia elementów macierzowych	27
1. Twierdzenie o splocie	27
2. Zastosowanie twierdzenia o splocie	29
Rozdział 6. Obliczenia dla wielu całek	31
1. Podstawowe symetrie	31
2. Degeneracja Kramersa	32
3. Zastosowanie dla określonych całek	33
Część 3. Implementacja	35
Rozdział 7. Opis pakietu oprogramowania	37

1. Informacje podstawowe	37
2. Interfejs użytkownika	38
3. Obliczenia z wykorzystaniem systemu PL-Grid	39
4. Składnia plików danych	40
Rozdział 8. Paralelizacja	43
1. MPI	43
2. OpenMP	43
3. CUDA	44
Rozdział 9. Elementy składowe pakietu	45
1. Program pqr-atomist	45
2. Program pqr-coulomb	45
3. Program pqr-listing	46
4. Program pqr-mapper	47
5. Program pqr-normalize	49
6. Program pqr-overlap	49
Część 4. Wyniki	51
Rozdział 10. Wyniki obliczeń	53
1. Charakterystyka badanego układu	53
2. Wartości elementów macierzowych	53
3. Studium promienia odcięcia	54
4. Studium zasięgu oddziaływania	56
Podsumowanie	59
Bibliografia	61

Część 1

Wprowadzenie

ROZDZIAŁ 1

Nanostruktury

Pojęciem nanostruktur określa się układy atomów o wymiarach przestrzennych rzędu od kilku do kilkuset nanometrów¹. W tej skali rozmiarów obserwowane są kwantowe efekty wynikające z ograniczenia przestrzennego (ang. *quantum confinement*) elektronów, które nie mogą swobodnie poruszać się w jednym lub więcej wymiarów. W odróżnieniu od kryształów makroskopowych, gdzie występuje symetria translacyjna w trzech kierunkach przestrzennych, nanostruktury określa się mianem struktur niskowymiarowych.

Szczególną uwagę poświęca się zerowymiarowym (ograniczonym we wszystkich trzech wymiarach) nanostrukturom półprzewodnikowym, zwanym kropkami kwantowymi. Ich unikalne własności elektronowe i optyczne dają szansę wykorzystania tych struktur w mikro- i optoelektronice. Rozważane jest m.in. wykorzystanie kropek kwantowych do budowy wydajnych źródeł pojedynczych i splątanych par fotonów, dla celów kwantowej informatyki, w tym kryptografii (Singh and Bester, 2009).

1. Wybrane rodzaje nanostruktur

1.1. Nanokryształy. Pojęciem nanokryształów określane są nanostruktury jednorodne pod względem składu chemicznego, posiadające dobrze określoną sieć krystaliczną (Fahlman, 2007). Nanokryształy otrzymuje się metodą chemicznej syntezy (Klimov, 2009). Możliwa jest kontrola rozmiarów wytwarzanych nanokryształów, co pozwala na dość dokładną kontrolę ich własności widmowych.

1.2. Samorosnące kropki kwantowe. Wykorzystywany w tworzeniu kropek kwantowych proces epitaksji z wiązek atomowych (Jacak et al., 1998) polega na nanoszeniu atomów jednego rodzaju (np. InAs) na powierzchnię innego rodzaju (np. GaAs). Na skutek niedopasowania struktur krystalicznych obu materiałów, na powierzchni otrzymanej struktury pojawiają się zagęszczenia materiału nanoszonego. Przy właściwym doborze materiałów ze względu na ich własności energetyczne, uzyskuje się efekt przestrzennego ograniczenia elektronów w powstałych zgęstnieniach, określanych mianem samorosnących kropek kwantowych.

¹1 nm = 10⁻⁹ m

1.3. Kropki kwantowe w nanodrutach. Zastosowanie metody epitaksji na odpowiednio spreparowanym podłożu oraz wykorzystanie nanodrobin złota jako katalizatorów wzrostu (Borgström et al., 2005) pozwala na uzyskanie jednowymiarowych struktur, tzw. nanodrutów, również szeroko wykorzystywanych w nanotechnice. Jeżeli w trakcie epitaksji na krótko zmieni się rodzaj napylanej substancji, możliwe jest uzyskanie wewnątrz nanodrutu krótkiego odcinka innego materiału, który wykazuje, z uwagi na przestrzenne ograniczenie, własności kropki kwantowej.

2. Optyczne własności kropek kwantowych

Na skutek ograniczenia przestrzennego we wszystkich trzech wymiarach, widmo energetyczne kropek kwantowych różni się znacznie od widma energetycznego kryształu objętościowego. W odróżnieniu od występujących w kryształach ciągłych pasm walencyjnego i przewodnictwa, w kropkach kwantowych obserwujemy dyskretne poziomy energetyczne. W stanie podstawowym kropki, poziomy znajdujące się poniżej poziomu Fermiego będą obsadzone, zaś te znajdujące się powyżej — nieobsadzone. Różnica pomiędzy najwyższym poziomem obsadzonym (HOMO) a najniższym poziomem nieobsadzonym (LUMO) określa się mianem (jednocząstkowej) przerwy energetycznej (ang. *band gap*).

Jeżeli pojedynczy elektron uzyskuje energię z zewnątrz (np. na skutek oddziaływania z promieniowaniem) zostaje on wzbudzony do jednego z poziomów znajdujących się powyżej przerwy energetycznej. Wprowadza się również pojęcie *dziury*, która jest kwazicząstką „zajmującą” nieobsadzony stan poniżej poziomu Fermiego. Na skutek tego, stany powyżej przerwy energetycznej określa się mianem stanów elektronowych, a te poniżej — stanów dziurowych.

Przeniesienie elektronu z poziomu dziurowego na elektronowy powoduje powstanie dwóch kwazicząstek: elektronu i dziury. Para elektronowo-dziurowa określana jest mianem *ekscytonu*, zaś energia ekscytonu równa jest sumie energii jednocząstkowych elektronu i dziury pomniejszonej o energię oddziaływania kulombowskiego pomiędzy nimi.

Wyjście poza przybliżenie jednokonfiguracyjne, a więc uwzględnienie efektów korelacyjnych dodatkowo wprowadza określoną (relatywnie niewielką) poprawkę do energii, ma jednak istotny jakościowy wpływ na optyczne widmo emisji ekscytonowej (rekombinacji pary elektron-dziura) oraz występowanie tzw. struktury subtelnej widma ekscytonowego (jasnych i ciemnych linii widmowych). Znajomość własności ekscytonowych, jak i własności kompleksów ekscytonowych (biekscytonów, trionów) jest niezwykle istotna z punktu widzenia optycznych zastosowań kropek kwantowych.

2.1. Struktura subtelna ekscytonu. Na skutek efektów mieszania konfiguracji, w widmie ekscytonowym kropki kwantowej, zarówno jasna jak i ciemna linia ekscytonowa może ulec rozszczepieniu. Powstała w ten sposób strukturę widmową nazywa się strukturą subtelną ekscytonu (ang. *fine structure splitting*).

Występowanie rozszczepienia jasnej linii ekscytonowej wiąże się ściśle z występowaniem asymetrii kształtu i/lub niskiej symetrii sieci krystalicznej w kropce kwantowej — dla kropek o wysokiej symetrii rozszczepienie nie będzie obserwowane. Dzięki możliwości doświadczalnego wyznaczenia widma ekscytonowego z dużą dokładnością, analiza struktury subtelnej w kropkach kwantowych może posłużyć do wywnioskowania informacji o kształcie lub symetrii badanej kropki.

Badanie jasnej linii ekscytonowej motywowane jest także postulowaną możliwością wykorzystania kropek kwantowych jako źródła splątanych par fotonów (Singh and Bester, 2009). Jak się okazuje, obecność rozszczepienia jasnej linii ekscytonowej znacznie utrudnia, a nawet uniemożliwia efektywne generowanie par splątanych. Fundamentalne znaczenie mają więc ilościowe badania, które pozwolą na modelowanie struktur o minimalnym rozszczepieniu jasnej linii, stanowiących potencjalnych kandydatów do generacji splątanych par fotonów.

Aby badania były skuteczne, muszą być oparte o metodę obliczeniową, która strukturę subtelną ekscytonu modeluje we właściwy, czyli zgodny z wynikami doświadczalnymi sposób. Metody fizyki ciała stałego wykorzystujące przybliżenie ciągłego ośrodka (np. przybliżenie masy efektywnej *EMA*) nie pozwalają na właściwe modelowanie rozszczepienia linii widmowych ekscytonu, bezpośrednio na skutek założeń poczynionych przy konstrukcji metody.

Z drugiej strony, metody oparte o teorię funkcjonałów gęstości (DFT) nie mogą być zastosowane do kropek kwantowych z uwagi na zbyt dużą złożoność obliczeniową. Rozwiązaniem pośrednim, pozwalającym na uzyskanie zgodnych z doświadczeniem wyników przy osiągalnym koszcie obliczeniowym, jest podejście atomistyczne oparte o metodę ciasnego wiązania.

ROZDZIAŁ 2

Obliczenia atomistyczne

Atomistyczna metoda obliczeń (Zieliński et al., 2010) pozwala na obliczenie własności ekscytonowych układu na podstawie jego geometrii i składu chemicznego. Obliczenia składają się z kilku etapów, wśród których wyszczególnić można:

- optymalizację geometrii układu metodą *Valence Force Field*,
- budowę macierzy hamiltonianu uwzględniającego efekty odkształceń,
- częściową lub całkowitą diagonalizację macierzy hamiltonianu,
- obliczenia elementów macierzowych operatora kulombowskiego (ang. *Coulomb matrix elements*),
- uwzględnienie oddziaływania konfiguracji.

Większa część niniejszej pracy poświęcona jest wyznaczaniu elementów macierzowych operatora Coulomba, czyli m.in. całek kulombowskich i wymiany. Etap ten jest bardzo kosztowny pod względem czasu obliczeń, więc stworzenie efektywnych programów obliczeniowych jest z całą pewnością istotne z punktu widzenia dalszych badań. Ponadto, analiza wpływu wprowadzanych na tym etapie parametrów na wyniki obliczeń może przyczynić się do głębszego zrozumienia ekscytonowych własności nanostruktur.

1. Metoda ciasnego wiązania

Centralną część obliczeń atomistycznych stanowi metoda ciasnego wiązania (ang. *tight binding method*, w skrócie TB) dla układów aperiodycznych. Pozwala ona na otrzymanie jednocząstkowych energii oraz funkcji falowych w przybliżeniu LCAO (ang. *Linear Combination of Atomic Orbitals*), czyli w postaci liniowej kombinacji orbitali atomowych:

$$(1) \quad \Psi(\vec{r}) = \sum_n \sum_\alpha c_{n\alpha} \psi_\alpha(\vec{r} - \vec{R}_n).$$

Funkcje $\psi_\alpha(\vec{r})$, pełniące rolę orbitali atomowych, nie są do końca sprecyzowane w metodzie ciasnego wiązania. Zamiast tego, na podstawie wymiaru bazy $\{\psi_\alpha\}$ oraz ustalonych symetrii poszczególnych orbitali, elementy macierzy hamiltonianu ustalane są półempirycznie. Są one dopasowywane do wyników doświadczalnych, tj. przerw energii wzbronionej oraz mas efektywnych kryształów objętościowych materiałów, z których zbudowana jest nanostruktura.

Tytułowe założenie metody ciasnego wiązania oparte jest na tym, że orbitale atomowe mają niewielką rozciągłość przestrzenną, elektrony są więc silnie („ciasno”) związane z atomami. Zakłada się także, że orbitale bazowe są parami ortonormalne, co jest o tyle istotne, że prowadzi o zwykłego, a nie uogólnionego (dużo bardziej złożonego numerycznie) zagadnienia własnego dla macierzy hamiltonianu.

Ponadto przyjmuje się, że orbitale bazowe mają określoną symetrię przestrzenną ($s, p, d, \dots$), co upraszcza istotnie budowę macierzy hamiltonianu (Slater and Koster, 1954). Dodatkowo stosuje się przybliżenie najbliższych sąsiadów, które efektywnie prowadzi do tego, że hamiltonian układu jest opisywany przez macierz rzadką. Dzięki powyższym założeniom możliwe jest efektywne wyznaczanie energii i stanów własnych hamiltonianu ciasnego wiązania za pomocą metod przestrzeni Kryłowa (algoritmy Lanczosa i Arnoldiego), przy czym obliczenia dotyczą jedynie stanów z najbliższego otoczenia przerwy energetycznej. W głównej mierze to właśnie te stany odpowiedzialne są za własności optyczne kropek kwantowych.

2. Elementy macierzowe operatora Coulomba

W rezultacie zakończonych obliczeń metodą ciasnego wiązania, otrzymuje się jednocząstkowe energie i funkcje własne. W dalszej kolejności, na użytek obliczeń wielocząstkowych, należy wyznaczyć wartości całek kulombowskich i wymiennych. W analogii do nomenklatury anglojęzycznej, w przypadku ogólnym, nazywamy te całki elementami macierzowymi operatora Coulomba (ang. *Coulomb matrix elements*), zdefiniowanymi jako

$$(2) \quad \langle i j k l \rangle = \iint \frac{\Psi_i(\vec{r}_1)^* \Psi_j(\vec{r}_2)^* \Psi_k(\vec{r}_2) \Psi_l(\vec{r}_1)}{4\pi\epsilon|\vec{r}_1 - \vec{r}_2|} dV_1 dV_2,$$

przy czym $\{\Psi_i(\vec{r})\}$ to zbiór jednocząstkowych funkcji falowych, $*$ oznacza operację sprzężenia zespolonego, zaś ϵ jest stałą dielektryczną materiału. W ogólnym przypadku można tu także zastosować funkcję dielektryczną $\epsilon(r) \equiv \epsilon(|\vec{r}_1 - \vec{r}_2|)$.

Szczególne przypadki elementów macierzowych postaci $J_{ij} = \langle i j j i \rangle$ określa się mianem całek kulombowskich (mają one klasyczną interpretację oddziaływania dwóch ładunków przestrzennych), zaś elementy $K_{ij} = \langle i j i j \rangle$ noszą nazwę całek wymiany.

Podstawienie funkcji falowych w postaci LCAO (1) do równania (2) prowadzi do otrzymania wyrażenia zależnego od współczynników ciasnego wiązania, czyli

$$(3) \quad \langle i j k l \rangle = \sum_{n_i} \sum_{\alpha_i} \sum_{m_j} \sum_{\beta_j} \sum_{m_k} \sum_{\beta_k} \sum_{n_l} \sum_{\alpha_l} \left(c_{n_i \alpha_i}^{i*} c_{m_j \beta_j}^{j*} c_{m_k \beta_k}^k c_{n_l \alpha_l}^l \right. \\ \left. \iint \frac{\psi_{\alpha_i}(\vec{r}_1 - \vec{R}_{n_i})^* \psi_{\beta_j}(\vec{r}_2 - \vec{R}_{m_j})^* \psi_{\beta_k}(\vec{r}_2 - \vec{R}_{m_k}) \psi_{\alpha_l}(\vec{r}_1 - \vec{R}_{n_l})}{4\pi\epsilon|\vec{r}_1 - \vec{r}_2|} dV_1 dV_2 \right).$$

Obliczenie powyższego wyrażenia jest bardzo kosztowne numerycznie (złożoność $O(N^4)$, gdzie N jest liczbą atomów struktury). Dla celów praktycznych obliczeń (Zieliński et al., 2010) stosuje się szereg przybliżeń, w szczególności zaniedbuje się wkłady trój- i czterocentrowe, uwzględniając jedynie wkłady jednocentrowe ($n_i = m_j = m_k = n_l$) oraz kulombowskie dwucentrowe ($n_j = n_l, m_j = m_k$). Z uwagi na ortogonalność i dobrą lokalizację przestrzenną orbitali atomowych, wkłady dwucentrowe zdominowane będą przez oddziaływania monopol-monopol ($\alpha_i = \alpha_l, \beta_j = \beta_k$).

Powyższe spostrzeżenia pozwalają sparametryzować wkłady jednocentrowe przy pomocy stałej nawęzłowej U_n zależnej od pierwiastka

$$(4) \quad \iint \frac{\psi_{\alpha_i}(\vec{r}_1 - \vec{R}_n)^* \psi_{\beta_j}(\vec{r}_2 - \vec{R}_n)^* \psi_{\beta_k}(\vec{r}_2 - \vec{R}_n) \psi_{\alpha_l}(\vec{r}_1 - \vec{R}_n)}{4\pi\epsilon|\vec{r}_1 - \vec{r}_2|} = \delta_{\alpha_j\alpha_k} \delta_{\beta_j\beta_k} U_n,$$

zaś wkłady dwucentrowe przybliżyć w postaci oddziaływania ładunków punktowych (monopol-monopol), czyli

$$(5) \quad \iint \frac{\psi_{\alpha_i}(\vec{r}_1 - \vec{R}_n)^* \psi_{\beta_j}(\vec{r}_2 - \vec{R}_m)^* \psi_{\beta_k}(\vec{r}_2 - \vec{R}_m) \psi_{\alpha_l}(\vec{r}_1 - \vec{R}_n)}{4\pi\epsilon|\vec{r}_1 - \vec{r}_2|} = \frac{\delta_{\alpha_j\alpha_k} \delta_{\beta_j\beta_k}}{4\pi\epsilon|\vec{R}_n - \vec{R}_m|}.$$

Przy zastosowaniu powyższych przybliżeń, elementy macierzowe operatora Coulomba przyjmują postać

$$(6) \quad \langle i j k l \rangle = \sum_n \sum_m \left(\sum_{\alpha} c_{n\alpha}^{i*} c_{n\alpha}^l \right) V_{nm} \left(\sum_{\beta} c_{m\beta}^{j*} c_{m\beta}^k \right),$$

gdzie

$$(7) \quad V_{nm} = \begin{cases} U_n & : n = m, \\ \frac{1}{4\pi\epsilon|\vec{R}_n - \vec{R}_m|} & : n \neq m. \end{cases}$$

3. Formalizm macierzowy

Wprowadźmy macierz C , zdefiniowaną w taki sposób, że jej kolumny odpowiadają kolejnym parom stanów jednoelektronowych, zaś wiersze — kolejnym atomom ($1 \dots N$) układu. Wprowadźmy podwójne indeksy $v \equiv (j, k)$, $u \equiv (i, l)$. Elementy macierzy C określimy jako

$$(8) \quad C_{mv} = \sum_{\beta} c_{m\beta}^{j*} c_{m\beta}^k.$$

Równanie (6) można teraz zapisać w postaci

$$(9) \quad \langle i j k l \rangle = \sum_n \sum_m C_{nu} V_{nm} C_{mv},$$

co sprowadza się do mnożenia trzech macierzy

$$(10) \quad \left\langle \underbrace{i \quad \overbrace{j \quad k}^v}_{u} \quad l \right\rangle = (C^T V C)_{uv}.$$

Powyższa konstrukcja pozwala na sprowadzenie obliczeń atomistycznych do prostych działań na macierzach, dla których istnieją efektywne implementacje (np. BLAS). W tym miejscu warto jednak zauważyć, że dla układu o N atomach, macierz V ma rozmiar $N \times N$, więc już przy $N \sim 10^5$ przechowywanie tej macierzy w całości staje się bardzo trudne.

Jedną z możliwości rozwiązania tego problemu jest generowanie macierzy V na bieżąco, w miarę obliczania kolejnych wierszy iloczynu VC . Zakładając, że w każdym momencie przechowywany będzie tylko jeden wiersz macierzy V , złożoność pamięciowa staje się liniowa (nie zaś kwadratowa) ze względu na liczbę atomów struktury.

4. Podsumowanie metody atomistycznej

Obliczenia atomistyczne pozwalają na dość proste obliczenie wartości elementów macierzowych operatora kulombowskiego. Jednakże, złożoność obliczeniowa tej metody wynosi, przy optymalnej implementacji, $O(NB + N^2)$, gdzie N jest liczbą atomów, a B wymiarem bazy orbitali atomowych. W rzeczywistych obliczeniach $B \ll N$, jednakże nawet złożoność $O(N^2)$ bardzo utrudnia obliczenia dla dużych struktur (rzędu 10^6 atomów).

Należy przy tym pamiętać, że dla dalszego wykorzystania metody oddziaływania konfiguracji, wymagane jest obliczenie znacznej ilości (ok. 10^4) różnych elementów macierzowych. Obliczenia będą miały przeto charakter obliczeń wielkoskalowych.

Pożądane byłoby zatem opracowanie metody, która pozwoliłaby na obniżenie czasowej złożoności obliczeń, jednocześnie utrzymując zgodność z doświadczeniem na poziomie metody atomistycznej. Dodatkowo, jeżeli metodę tą udało się skonstruować bez stosowania wszystkich przybliżeń wykorzystywanych w metodzie atomistycznej, mogłoby to posłużyć do weryfikacji stosowności tych przybliżeń dla różnych układów.

W dalszej części pracy zaproponowana zostanie alternatywna metoda wyznaczania kulombowskich elementów macierzowych, spełniająca powyższe wymagania. Przedstawiona metoda stanowić będzie rozwinięcie metody ciasnego wiązania o obliczenia na regularnej, trójwymiarowej siatce przestrzennej.

Część 2

Metoda

ROZDZIAŁ 3

Rekonstrukcja funkcji falowej

Pierwszy etap obliczeń stanowi „rekonstrukcja” funkcji falowej, czyli odwzorowanie przestrzennego jej przebiegu na regularnej, trójwymiarowej siatce (na podstawie współczynników ciasnego wiązania). W celu uwzględnienia faktu posiadania spinu przez elektrony, należy osobno rekonstruować obie składowe spinowe funkcji falowej ($\Psi_{\uparrow}$ i $\Psi_{\downarrow}$). Całkowitą funkcję falową można w tym przypadku zapisać w uogólnionej postaci

$$(11) \quad \Psi(\vec{r}, s) = \begin{cases} \Psi_{\uparrow}(\vec{r}) & : s = +\frac{1}{2} \\ \Psi_{\downarrow}(\vec{r}) & : s = -\frac{1}{2} \end{cases},$$

przy czym współrzędna spinowa s przyjmuje wartości $\pm\frac{1}{2}$.

1. Wybór bazy funkcyjnej

Metoda ciasnego wiązania (str. 7) wprowadza kilka założeń (symetria, ortonormalność) dotyczących orbitali atomowych, jednak nie precyzuje przestrzennej (radialnej) postaci orbitali. Nie jest to konieczne, ponieważ elementy macierzy hamiltonianu dane są jako parametry empiryczne. Problem pojawia się, gdy konieczne jest obliczenie elementów macierzowych innych operatorów (np. operatora Coulomba). Jednym z rozwiązań stosowanych w literaturze (Zieliński et al., 2010) jest wprowadzenie *a posteriori* bazy funkcyjnej, spełniającej (choćby w przybliżeniu) założenia metody ciasnego wiązania. Funkcjami takimi są na przykład opisane krótko w dalszej części pracy orbitale Slatera i Hermana-Skillmana.

Na podstawie założeń metody ciasnego wiązania związanych z symetrią orbitali, funkcję falową orbitala atomowego można rozłożyć na czynnik radialny (o symetrii sferycznej) oraz czynnik kątowy:

$$(12) \quad \psi(\vec{r}) = T(r) Y(\theta, \phi).$$

Ponieważ orbitale atomowe powinny stanowić funkcje własne operatora kwadratu momentu pędu (Edmonds, 1960), kątowa część orbitali atomowych musi być kombinacją liniową harmonik sferycznych

$$(13) \quad Y_l^m(\theta, \phi) = (-1)^m \left[\frac{(2l+1)(l-m)!}{4\pi(l+m)!} \right]^{1/2} P_l^m(\cos \theta) e^{im\phi},$$

przy czym $|m| \leq l$, zaś dla $l = 0, 1, 2$ przyjmuje się oznaczenia s, p, d ¹. Notacją $P_l^m(x)$ oznaczone są *stowarzyszone funkcje Legendre'a*, zdefiniowane jako

$$(14) \quad P_l^m(x) = (1-x^2)^{m/2} \frac{d^m}{dx^m} \left(\frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2-1)^l \right).$$

Powiązana z orbitalem atomowym gęstość prawdopodobieństwa znalezienia elektronu w punkcie $\vec{r}$ wynosi $|\psi(\vec{r})|^2$. Po podstawieniu rozkładu (12) do warunku normalizacyjnego gęstości prawdopodobieństwa, otrzymamy warunek normalizacji kątowej części orbitala atomowego:

$$(15) \quad \int_0^{2\pi} \int_0^\pi Y(\theta, \phi)^2 \sin \theta \, d\theta \, d\phi = 1$$

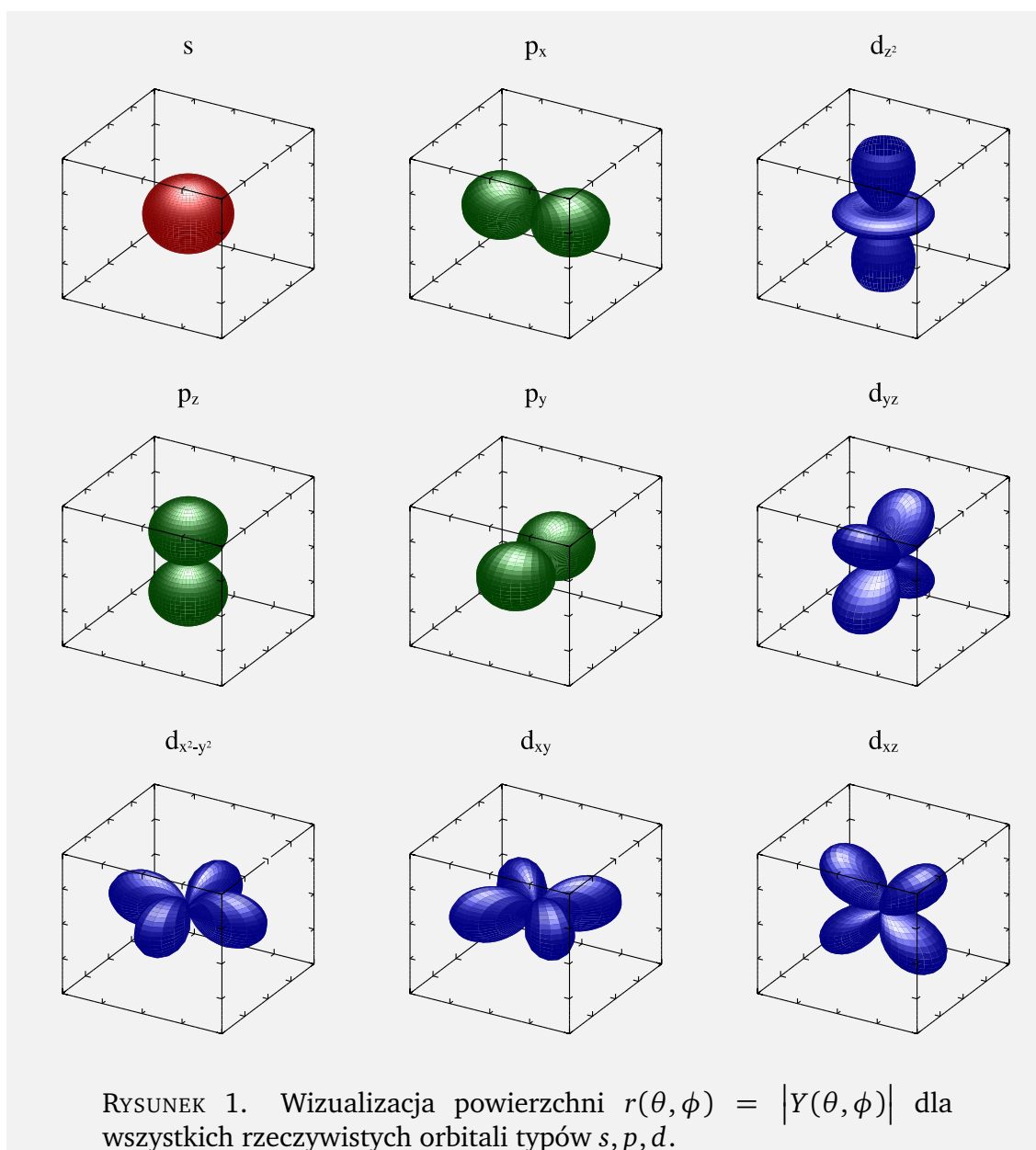
W celu otrzymania orbitali rzeczywistych (o zerowej części urojonej), a jednocześnie znormalizowanych, można stworzyć następujące kombinacje liniowe:

$$(16) \quad \begin{aligned} Y_s &= Y_0^0 & Y_d^{yz} &= \frac{i}{\sqrt{2}}(Y_2^{-1} + Y_2^1) \\ Y_p^x &= \frac{1}{\sqrt{2}}(Y_1^{-1} - Y_1^1) & Y_d^{xz} &= \frac{1}{\sqrt{2}}(Y_2^{-1} - Y_2^1) \\ Y_p^y &= \frac{i}{\sqrt{2}}(Y_1^{-1} + Y_1^1) & Y_d^{xy} &= \frac{i}{\sqrt{2}}(Y_2^{-2} - Y_2^2) \\ Y_p^z &= Y_1^0 & Y_d^{x^2-y^2} &= \frac{1}{\sqrt{2}}(Y_2^{-2} + Y_2^2) \\ & & Y_d^{z^2} &= Y_0^2. \end{aligned}$$

Analityczną postać orbitali rzeczywistych da się wyrazić się w prosty sposób przez współrzędne kartezjańskie x, y, z (oznaczając dodatkowo $r = \sqrt{x^2 + y^2 + z^2}$):

$$(17) \quad \begin{aligned} Y_s &= (4\pi)^{-\frac{1}{2}} & Y_d^{yz} &= \left(\frac{4}{15}\pi\right)^{-\frac{1}{2}} \frac{yz}{r^2} \\ Y_p^x &= \left(\frac{4}{3}\pi\right)^{-\frac{1}{2}} \frac{x}{r} & Y_d^{xz} &= \left(\frac{4}{15}\pi\right)^{-\frac{1}{2}} \frac{xz}{r^2} \\ Y_p^y &= \left(\frac{4}{3}\pi\right)^{-\frac{1}{2}} \frac{y}{r} & Y_d^{xy} &= \left(\frac{4}{15}\pi\right)^{-\frac{1}{2}} \frac{xy}{r^2} \\ Y_p^z &= \left(\frac{4}{3}\pi\right)^{-\frac{1}{2}} \frac{z}{r} & Y_d^{x^2-y^2} &= \left(\frac{16}{15}\pi\right)^{-\frac{1}{2}} \frac{x^2-y^2}{r^2} \\ & & Y_d^{z^2} &= \left(\frac{16}{5}\pi\right)^{-\frac{1}{2}} \left(\frac{3z^2}{r^2} - 1\right). \end{aligned}$$

¹notacja pochodzi ze spektroskopii



Powyższe rozważania nie precyzują jednak radialnych części $T(r)$ orbitali atomowych, dlatego do ich wyznaczenia niezbędne jest wykorzystanie dodatkowych metod obliczeniowych. W ogólności, radialna część orbitali będzie od rodzaju pierwiastka, może być także różna dla poszczególnych symetrii orbitalnych ($s, p, d, \dots$).

1.1. Orbitale Slatera. Zaproponowane przez Johna C. Slatera w 1930 roku orbitale mają część radialną

$$(18) \quad T(r) = N r^{\nu-1} e^{-\beta r},$$

gdzie ν i β są parametrami, zaś N czynnikiem normalizującym, równym²

$$(19) \quad N = \frac{(2\beta)^{\nu+\frac{1}{2}}}{\sqrt{\Gamma(2\nu+1)}},$$

co gwarantuje warunek normalizacji dla części radialnej orbitala atomowego

$$(20) \quad \int_0^\infty T(r)^2 r^2 dr = 1.$$

Współczynniki ν i β mogą zostać wyznaczone w oparciu o tzw. *reguły Slatera* na podstawie konfiguracji elektronowej danego pierwiastka. Parametr ν wyznacza się w oparciu o główną liczbę kwantową n jako

n	1	2	3	4	5	6
ν	1.0	2.0	3.0	3.7	4.0	4.2

zaś parametr β obliczany jest jako suma ważona liczby elektronów na poszczególnych powłokach (Slater, 1930).

Jeżeli potrzebny jest przebieg orbitala atomowego dla podpowłoki, która w stanie podstawowym jest nieobsadzona, należy dokonać przeniesienia elektronu z najbliższego zajętego orbitala. Dla przykładu, w celu uzyskania radialnej części orbitala atomowego d dla fosforu, którego podstawowa konfiguracja elektronowa wynosi

$$1s^2 2s^2 2p^6 3s^2 3p^3,$$

należy dokonać obliczeń dla konfiguracji powstałej poprzez przeniesienie pojedynczego elektronu z orbitala p , czyli

$$1s^2 2s^2 2p^6 3s^2 3p^2 3d^1.$$

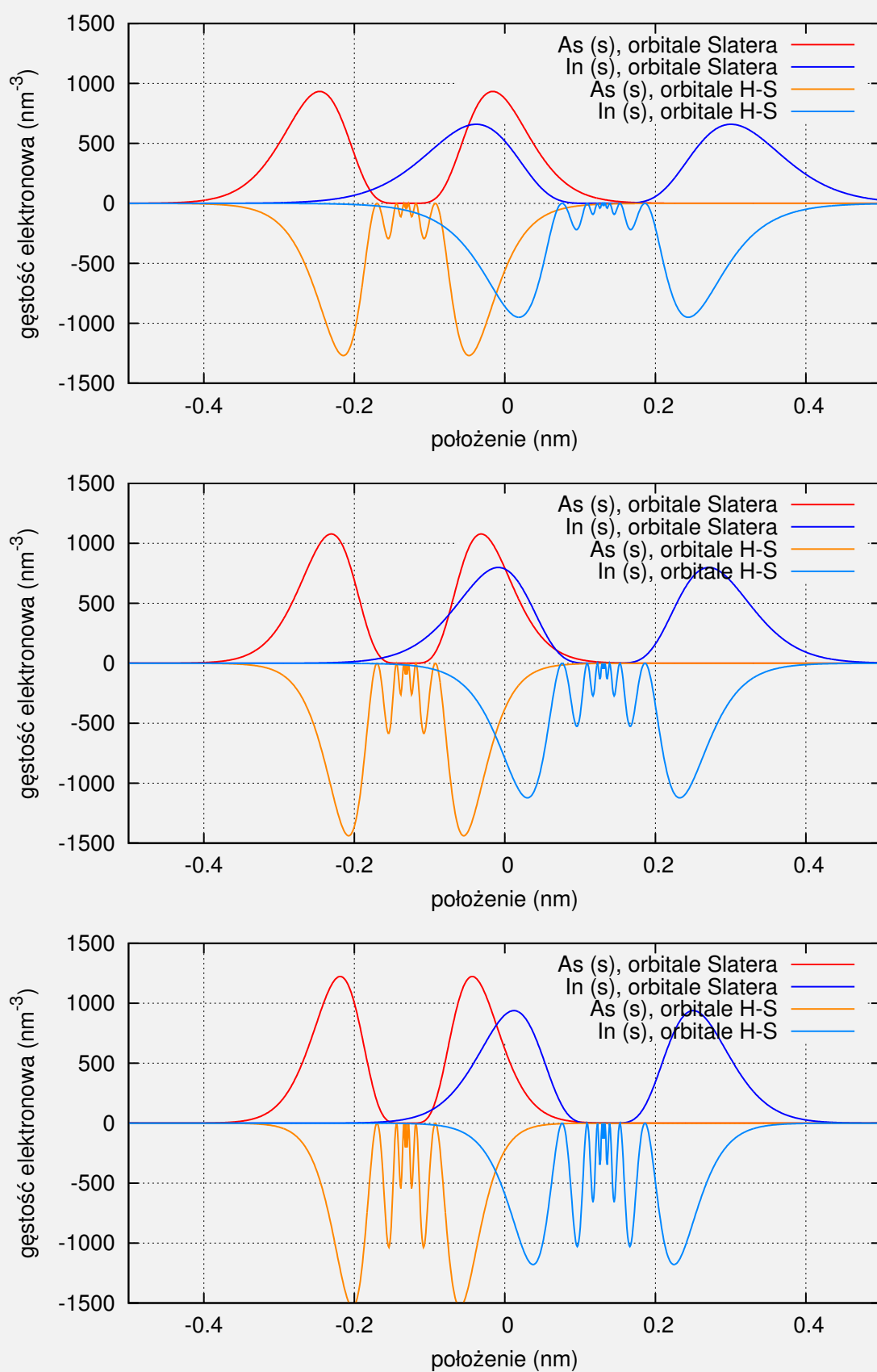
1.2. Orbitale Herman-Skillman. Orbitale atomowe można także otrzymać bezpośrednio z równań Hartree-Focka, rozwiązując je numerycznie metodą pola samouzgodnionego (ang. *self-consistent field*). Na tej zasadzie działa program Herman-Skillman³ (od nazwisk pierwszych autorów), rozwijany z przerwami od roku 1960.

Orbitale otrzymane tym sposobem są m.in. znacznie mniej przestrzennie rozciągnięte niż orbitale Slatera, a więc w większym stopniu spełniają założenia metody TB dotyczące ortogonalności bazy atomowej. Ekstrema widoczne w przestrzennym przebiegu orbitali (wykres 2, str. 17) wynikają z oddziaływania elektronu walencyjnego z elektronami o niższej energii (znajdującymi się na niższych powłokach).

1.3. Ściskanie orbitali. Występujące w metodzie TB założenie dotyczące ortogonalności bazy atomowej, jest w przypadku orbitali Slatera i Herman-Skillman

² Γ oznacza tu funkcję gamma, zdefiniowaną jako $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$.

³<http://hermes.phys.uwm.edu/projects/electruct/hermsk/HS.TOC.html>



RYSUNEK 2. Porównanie orbitali Slatera z orbitalami H-S na przykładzie najbliższych sąsiadów w strukturze InAs dla różnych wartości współczynnika ściskającego κ . Kolejno od góry: $\kappa = 0$, $\kappa = 5 \text{ nm}^{-1}$, $\kappa = 10 \text{ nm}^{-1}$.

spełnione jedynie w przybliżeniu. Rozwiązaniem tego problemu byłaby ortogonalizacja bazy np. metodą Grama-Schmidta. Skomplikowałoby to jednak proces tworzenia macierzy hamiltonianu ze względu na inną niż atomową symetrię bazy po procesie ortogonalizacji. Poniżej proponujemy proste „robocze” rozwiązanie tego problemu.

Orbitale wyznaczone w poprzednich akapitach odpowiadają elektronom należącym do pojedynczych atomów w pustej przestrzeni. W kryształach objętościowym, średnie odległości elektronów od jądra atomu są zbliżone do odległości pomiędzy najbliższymi atomami. Obecność otaczającej atom sieci krystalicznej powoduje efektywnie dodatkową, przestrzenną kompresję orbitali, która powinna zostać uwzględniona także i w przypadku nanostruktur.

Dodatkowe ścisnienie orbitali można uwzględnić, wprowadzając do części radialnej czynnik multiplikatywny $e^{-\kappa r}$ zależny od parametru $\kappa > 0$. Uwzględniając konieczność ponownej normalizacji orbitali, otrzymujemy

$$(21) \quad \tilde{T}(r) = N T(r) e^{-\kappa r},$$

gdzie czynnik normalizacyjny N wybrany jest tak, aby był spełniony warunek normalizacji (20).

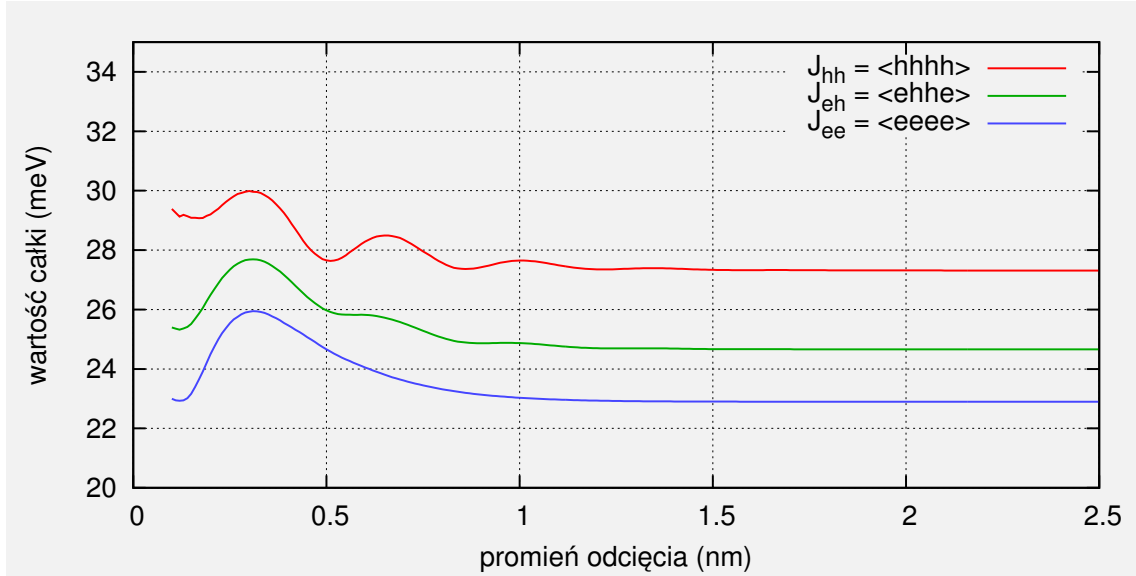
Wykres 2 (str. 17) przedstawia przestrzenne części radialne orbitali najbliższych sąsiadów sieci krystalicznej InAs w postaci $\tilde{T}(r)^2 r^2$ w zależności od współczynnika ściskającego κ .

2. Ustalenie parametrów rekonstrukcji

Po sprecyzowaniu postaci orbitali atomowych, można przystąpić do odwzorowania funkcji falowej w postaci LCAO (1) na dyskretnej siatce. Należy w tym celu sprecyzować parametry określające położenia punktów siatki oraz szczegóły rekonstrukcji.

2.1. Promień odcięcia. Jak zauważyć można na przykładzie opisanych wyżej orbitali bazowych, wartość orbitalnej funkcji falowej maleje silnie już w odległości kilku Å od środka atomu (jest to zgodne z istotą metody ciasnego wiązania). Na skutek tego, rekonstrukcja wkładu do funkcji falowej pochodzącego od danego atomu nie musi być prowadzona na całym obszarze siatki, lecz tylko na niewielkim jego otoczeniu. Jest to niezmiernie istotne dla uzyskania (quasi-)liniowej złożoności czasowej metody także i na tym etapie obliczeń.

Otoczenie wokół każdego atomu modelować można w postaci kuli o określonym promieniu odcięcia R_c . Przykładową zależność wartości całek kulombowskich od R_c przedstawia wykres 3 (str. 19) — na podstawie analogicznego studium można określić minimalną wartość promienia odcięcia dla zadanej struktury, w celu uzyskania żądanej dokładności numerycznej.



RYSUNEK 3. Wartości całek kulombowskich względem promienia odcięcia dla przykładowej samorosnącej kropki kwantowej InAs/GaAs.

Testy numeryczne wykonane nad różnymi rodzajami struktur pozwoliły zauważyć, że wartość promienia odcięcia wymagana w celu uzyskania względnej dokładności numerycznej na poziomie 0.1% zawiera się w przedziale 1.5 – 2.0 nm dla orbitali HS (2.0 – 2.5 nm dla orbitali Słatera) i tylko w niewielkim stopniu zależy od rozmiaru i rodzaju badanej struktury.

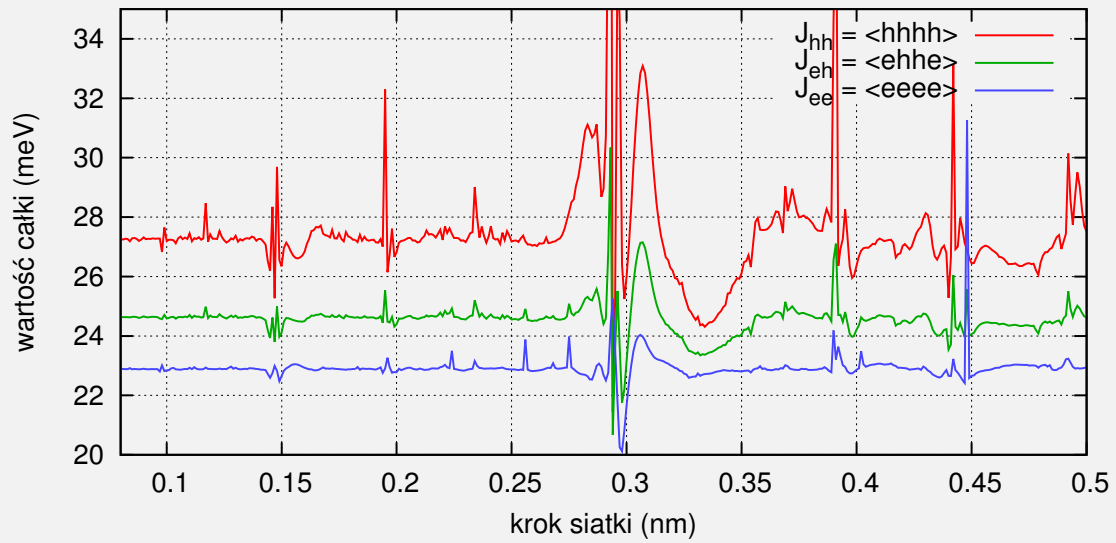
2.2. Rozmiar i krok siatki. Dokładność odwzorowania funkcji falowej na siatce jest ściśle związana z *krokiem siatki*, czyli odległością (we wszystkich trzech wymiarach) pomiędzy sąsiednimi punktami siatki. Fizyczny zakres położenia punktów siatki wyznaczony jest natomiast przez położenia atomów badanej struktury w taki sposób, aby wszystkie atomy znalazły się we wnętrzu obszaru rekonstrukcji.

Oznaczmy przez $x_{min}, y_{min}, z_{min}$ (analogicznie $x_{max}, y_{max}, z_{max}$) najmniejszą (analogicznie największą) wartość poszczególnych współrzędnych kartezjańskich atomów struktury. Przy kroku siatki równym h , oraz zakładając że punkt $[0, 0, 0]$ jest jednym z punktów siatki, będzie ona składała się z punktów o współrzędnych

$$(22) \quad \vec{r} = [i_x h, i_y h, i_z h],$$

gdzie

$$(23) \quad \begin{aligned} \left\lfloor \frac{x_{min}}{h} \right\rfloor &\leq i_x \leq \left\lceil \frac{x_{max}}{h} \right\rceil \\ \left\lfloor \frac{y_{min}}{h} \right\rfloor &\leq i_y \leq \left\lceil \frac{y_{max}}{h} \right\rceil \\ \left\lfloor \frac{z_{min}}{h} \right\rfloor &\leq i_z \leq \left\lceil \frac{z_{max}}{h} \right\rceil. \end{aligned}$$



RYSUNEK 4. Wartości całek kulombowskich względem kroku siatki dla przykładowej samorosnącej kropki kwantowej InAs/GaAs.

W powyższy sposób definiowany jest prostopadłościan wyznaczający obszar rekonstrukcji, w którego wnętrzu znajdują się wszystkie atomy analizowanej struktury. Można zauważyć, że zwiększenie (przy ustalonym kroku siatki) rozmiaru struktury w każdym wymiarze o stały czynnik, spowoduje taki sam wzrost liczby atomów jak i liczby komórek siatki. W związku z tym można stwierdzić, że liczba komórek siatki jest liniowo zależna od liczby atomów układu.

Jak pokazuje wykres 4 (str. 20), wpływ kroku siatki na wyniki obliczeń jest dość znaczny. Przyczyną chaotycznych wahań wartości całek jest systematyczny błąd dyskretyzacji spowodowany nakładaniem się regularnej siatki na regularności sieci krystalicznej materiału. Na podstawie testów można stwierdzić, że uzyskanie wartościowych wyników będzie możliwe tylko przy przyjęciu kroku siatki poniżej 0.1 nm.

2.3. Odstęp przestrzenny. Na skutek ograniczenia obszaru rekonstrukcji według (23), część wkładu do funkcji falowej pochodzącego od atomów skrajnych nie jest brana pod uwagę. Dla niektórych układów (np. nanokryształów), dla których wkłady pochodzące od skrajnych atomów mają znaczący wpływ na funkcję falową, sytuacja taka może zaowocować błędnymi wynikami. W celu właściwego uwzględnienia wkładów od wszystkich atomów, korzystne jest wprowadzenie dodatkowego *marginiesu* R_m czyli dodatkowego rozszerzenia obszaru rekonstrukcji.

Wartość marginesu może zawierać się w przedziale $0 \leq R_m \leq R_c$. Po jego uwzględnieniu, równania (23) przybiorą postać

$$(24) \quad \begin{aligned} \left\lfloor \frac{x_{\min} - R_m}{h} \right\rfloor &\leq i_x \leq \left\lceil \frac{x_{\max} + R_m}{h} \right\rceil \\ \left\lfloor \frac{y_{\min} - R_m}{h} \right\rfloor &\leq i_y \leq \left\lceil \frac{y_{\max} + R_m}{h} \right\rceil \\ \left\lfloor \frac{z_{\min} - R_m}{h} \right\rfloor &\leq i_z \leq \left\lceil \frac{z_{\max} + R_m}{h} \right\rceil . \end{aligned}$$

W praktyce, przy ustalonych h oraz R_m uzasadnione może być dodatkowe zwiększenie liczby punktów siatki w taki sposób, aby rozmiar siatki we wszystkich trzech wymiarach rozkładał się na względnie małe (jednocyfrowe) czynniki pierwsze. Jest to pożądane w przypadku późniejszego stosowania algorytmu FFT, którego wydajność może zależeć od istnienia takiego rozkładu.

3. Normalizacja funkcji falowej

Zgodnie z tzw. kopenhaską interpretacją w mechanice kwantowej, kwadrat modułu funkcji falowej $|\psi^2|$ stanowi stowarzyszoną z elektronem gęstość prawdopodobieństwa. W związku z tym, całkowita gęstość prawdopodobieństwa powinna być równa jedności, czyli

$$(25) \quad \int |\Psi(\vec{r})|^2 dV = 1 .$$

Z uwagi na degenerację spinową, przy normalizacji należy wziąć pod uwagę obie spinowe części funkcji falowej, a warunek normalizacyjny przedstawia się jako

$$(26) \quad \int |\Psi_{\uparrow}(\vec{r})|^2 dV + \int |\Psi_{\downarrow}(\vec{r})|^2 dV = 1 ,$$

co w przypadku dyskretnej siatki o kroku h przechodzi w

$$(27) \quad \sum_{\vec{r}} |\Psi_{\uparrow}(\vec{r})|^2 + \sum_{\vec{r}} |\Psi_{\downarrow}(\vec{r})|^2 = h^{-3} .$$

Niewątpliwa zaleta powyższego sposobu normalizacji (z uwzględnieniem kroku siatki h) polega na możliwości porównywania wartości funkcji falowych uzyskanych na siatkach o różnej dokładności.

Dyskretyzacja oddziaływania

Elementy macierzowe operatora Coulomba (2), będące w centrum zainteresowania niniejszej pracy, stanowią miarę oddziaływania pomiędzy produktami funkcji falowych. Skoro funkcje falowe zostają odwzorowane na regularnej siatce, a więc przedstawione na skończonym zbiorze punktów, to oddziaływania również sprowadzają się do interakcji pomiędzy komórkami trójwymiarowej siatki.

1. Oddziaływanie kulombowskie

Analityczną postać elementów macierzowych operatora kulombowskiego (2) można zapisać w postaci

$$(28) \quad \langle i j k l \rangle = \iint \Psi_i(\vec{r}_1)^* \Psi_j(\vec{r}_2)^* G(\vec{r}_1 - \vec{r}_2) \Psi_k(\vec{r}_2) \Psi_l(\vec{r}_1) dV_1 dV_2,$$

gdzie wprowadzono oznaczenie

$$(29) \quad G(\vec{r}) = \frac{e^2}{4\pi\epsilon} \frac{1}{|\vec{r}|}.$$

W przypadku funkcji falowych odwzorowanych na siatce o kroku h , całka (28) przybliżona zostanie przez sumę

$$(30) \quad \langle i j k l \rangle = \sum_{\vec{r}_1} \sum_{\vec{r}_2} \Psi_i(\vec{r}_1)^* \Psi_j(\vec{r}_2)^* G(\vec{r}_1 - \vec{r}_2) \Psi_k(\vec{r}_2) \Psi_l(\vec{r}_1) h^6.$$

Funkcja $G(\vec{r})$ wyrażająca oddziaływanie jest, z uwagi na dyskretyzację funkcji falowej, zdefiniowana również tylko na regularnej siatce o kroku równym h . Jeżeli komórki siatki traktowane są jako ładunki punktowe, postać funkcji oddziaływania jest taka jak w przypadku ciągłym (29). W takiej sytuacji pojawia się jednak problem z uwzględnieniem oddziaływania w przypadku $\vec{r}_1 = \vec{r}_2$.

Aby uniknąć osobliwości dla $\vec{r}_1 = \vec{r}_2$, należy komórki siatki traktować jako kostki o boku h , wewnątrz których wartość funkcji falowych jest stała. W takiej sytuacji $G(\vec{r})$ wyrażać będzie oddziaływanie pomiędzy dwoma jednorodnie naładowanymi kostkami o boku h , przesuniętymi względem siebie o $\vec{r}$. Oddziaływania takie analizowali Mura i Handy (1995), obliczając analitycznie całki tożsame z

$$(31) \quad I(\vec{r}) = \int_{[0;1]^6} \frac{1}{\|(\vec{r}_1 - \vec{r}_2) + \vec{r}\|} dV_1 dV_2.$$

x	y	z	$I([x\ y\ z])$	$\ [x\ y\ z] \ ^{-1}$
0	0	0	1.882313	—
1	0	0	0.980885	1.000000
1	1	0	0.708495	0.707107
1	1	1	0.578797	0.577350
2	0	0	0.499140	0.500000
2	1	0	0.447100	0.447214
2	2	0	0.353591	0.353553
2	2	1	0.333392	0.333333
2	2	2	0.288715	0.288675

TABELA 1. Porównanie oddziaływania w modelu jednorodnych kostek z oddziaływaniem ładunków punktowych (na podstawie Mura and Handy, 1995).

Wyniki uzyskane przez autorów przedstawia tabela 1 (str. 24). Rozbieżności pomiędzy modelem oddziaływania jednorodnych kostek a oddziaływaniem ładunków punktowych są dość niewielkie (poniżej 2%) i istotne tylko dla oddziaływania pomiędzy pobliskimi komórkami. Zaletą tego modelu jest jednak możliwość uwzględnienia oddziaływania dla $\vec{r} = 0$ jako

$$(32) \quad G(\vec{0}) = \frac{e^2}{4\pi\epsilon} \frac{I(\vec{0})}{h}.$$

Analityczna postać stałej $I(\vec{0})$ wynosi

$$(33) \quad \begin{aligned} I(\vec{0}) &= \log\left(\frac{\sqrt{2}+1}{\sqrt{2}-1}\right) + 2 \log\left(\frac{\sqrt{3}+1}{\sqrt{3}-1}\right) - \frac{4}{5}\sqrt{3} + \frac{2}{5}(\sqrt{2}+1) - \frac{2}{3}\pi \\ &= 1.8823126443896601601\dots \end{aligned}$$

2. Ekranowanie dielektryczne

Na gruncie fizyki klasycznej można stwierdzić, że pole elektryczne ulega osłabieniu wewnątrz dielektryka (Griffiths, 1999). Zjawisko to określa się mianem ekranowania i jest spowodowane lokalną polaryzacją ładunku, na skutek której powstaje lokalne pole elektryczne równoważące częściowo pole zewnętrzne.

W przypadku oddziaływania dwóch ładunków, pole pochodzące od jednego z nich ulegnie osłabieniu na skutek ekranowania, oddziaływanie między nimi będzie zatem mniejsze o ten sam czynnik. Ekranowanie wpłynie zatem na wartość elementów macierzowych operatora Coulomba, stanowią one bowiem miarę oddziaływania elektrostatycznego pomiędzy quasi-ładunkami (równanie 51, str. 29).

2.1. Stała dielektryczna. Najprostszy model ekranowania opiera się na założeniu, że ekranowane pole elektryczne zawsze ulegnie zmniejszeniu o niezmienny czynnik ϵ_r , określany mianem względnej stałej dielektrycznej i stanowiący własność materiałową. Jeżeli tak jest, to funkcję oddziaływania $G(\vec{r})$ (29) można zapisać jako

$$(34) \quad G(\vec{r}) = \frac{e^2}{4\pi \underbrace{\epsilon_0 \epsilon_r}_{\epsilon}} \frac{1}{|\vec{r}|}.$$

2.2. Model Thomasa-Fermiego. Resta wykazał (Resta, 1977), że model Thomasa-Fermiego opisujący strukturę elektronową układów wielociałowych może być zastosowany do opisu zjawiska ekranowania w półprzewodnikach. Model ten opisuje ekranowanie przy pomocy zależnej od odległości funkcji dielektrycznej $\epsilon(r)$ zdefiniowanej jako

$$(35) \quad \epsilon(r) = \begin{cases} \frac{\epsilon_0 q_{TF} r_{TF}}{\sinh[(r_{TF}-r)q_{TF}] + q_{TF} r} & : r \leq r_{TF} \\ \epsilon_0 & : r > r_{TF}, \end{cases}$$

przy czym

$$(36) \quad q_{TF} = \frac{2}{a_0 \sqrt{\pi}} \sqrt[3]{96\pi^2}.$$

Wielkość a_0 jest stałą sieci krystalicznej, ϵ_0 stałą dielektryczną w kryształach makroskopowych, zaś r_{TF} (tzw. promień Thomasa-Fermiego) jest zdefiniowany jako rozwiązanie równania

$$(37) \quad \sinh(r_{TF} q_{TF}) = \epsilon_0 r_{TF} q_{TF}.$$

Dla dużych odległości ($r > r_{TF}$), model Thomasa-Fermiego jest tożsamy z modelem stałej dielektrycznej. Różnice widoczne są natomiast dla odległości mniejszych, a szczególnie dla $r \rightarrow 0$, dla którego model T-F przewiduje $\epsilon \rightarrow 1$.

Warto zauważyć, że przy zastosowanej metodzie obliczeń, wprowadzenie funkcji dielektrycznej nie komplikuje w żaden sposób złożoności obliczeń. Wyznaczanie wartości tej funkcji wymagane jest tylko raz, podczas dyskretyzacji oddziaływania jako

$$(38) \quad G(\vec{r}) = \frac{e^2}{4\pi \epsilon(|\vec{r}|)} \frac{1}{|\vec{r}|}.$$

3. Zasięg oddziaływania

Kulombowskie oddziaływanie ładunków ma naturę długozasięgową, jego efektywny zasięg jest więc porównywalny z rozmiarami badanej kropki kwantowej. W niektórych przypadkach, szczególnie przy przestrzennie niejednorodnym rozkładzie funkcji falowych, może pojawić się potrzeba wyodrębnienia i zmierzenia oddziaływania pomiędzy pobliskimi lub odległymi wkładami do kwaziładunków.

W przypadku obliczeń atomistycznej, można tego dokonać, modyfikując macierz V_{nm} w taki sposób, aby była ona równa zero dla par atomów położonych w odległości przekraczających ustalony próg:

$$(39) \quad V_{nm} = \begin{cases} U_n & : n = m, \\ \frac{1}{4\pi\epsilon|\vec{R}_n - \vec{R}_m|} & : n \neq m \quad \wedge \quad |\vec{R}_n - \vec{R}_m| < r_{\max}, \\ 0 & : n \neq m \quad \wedge \quad |\vec{R}_n - \vec{R}_m| > r_{\max}. \end{cases}$$

W przypadku obliczeń z wykorzystaniem regularnej siatki, zastosowanie powyższego wyrażenia mogłoby prowadzić do pojawienia się niepożądanych efektów numerycznych związanych z ziarnistością domeny obliczeniowej. Należy więc w metodzie odwzorowania wykorzystać „wygładzoną” funkcję skoku (funkcję sigmoidalną), jak na przykład

$$(40) \quad G(\vec{r}) = \frac{e^2}{4\pi\epsilon} \frac{1}{|\vec{r}|} \left(1 + e^{\frac{1}{\lambda}(|\vec{r}| - r_{\max})} \right)^{-1}.$$

Parametr $\lambda > 0$ pełni rolę parametru wygładzania oddziaływania. W granicy $\lambda \rightarrow 0$ na sferze o promieniu $r_{\max}$ pojawia się nieciągłość funkcji $G(\vec{r})$.

ROZDZIAŁ 5

Obliczenia elementów macierzowych

Wyrażenie (30) z poprzedniego rozdziału można wykorzystać do obliczenia wartości elementów macierzowych operatora kulombowskiego, jednak wymaga to całkowania (sumowania) w sześciu wymiarach przestrzennych. Jest to równoważne kwadratowej złożoności czasowej metody atomistycznej (strona 10), sposób ten nie jest więc praktyczny dla dużych układów ($\sim 10^6$ atomów).

Jednakże, korzystając z regularności siatki obliczeniowej, możliwe jest wprowadzenie efektywnej metody obliczeń opartej o twierdzenie o splocie i wykorzystującej Szybką Transformatę Fouriera.

1. Twierdzenie o splocie

W celu wprowadzenia twierdzenia o splocie, warto przypomnieć kilka definicji; równoważne definicje można odnaleźć w wielu podręcznikach ze wstępu do przetwarzania sygnałów.

Definicja 5.1. Jeżeli x i y są sygnałami dyskretnymi, to splotem (zwykłym) x i y nazywamy taki sygnał $(x \star y)$, że

$$(41) \quad (x \star y)[n] = \sum_{i=-\infty}^{+\infty} x[n-i]y[i].$$

Definicja 5.2. Jeżeli x i y są sygnałami dyskretnymi, to splotem cyklicznym x i y nazywamy taki sygnał $(x_T \star y)$, że

$$(42) \quad (x_T \star y)[n] = \sum_{i=-\infty}^{+\infty} x_T[n-i]y[i],$$

przy czym x_T jest zdefiniowane jako

$$(43) \quad x_T[i] = \sum_{k=-\infty}^{+\infty} x[i-kT].$$

Twierdzenie 5.3 (twierdzenie o splocie cyklicznym). *Jeżeli x i y są sygnałami dyskretnymi, to*

$$DFT(x_T \star y) = DFT(x_T)DFT(y_T),$$

przy czym $DFT(\cdot)$ oznacza Dyskretną Transformatę Fouriera.

Można zauważyć, że jeśli $y[i]$ jest sygnałem o zwartym nośniku ($0 \leq i < N$), to równanie (42) jest równoważne

$$(44) \quad (x_T \star y)[n] = \sum_{i=0}^{N-1} x_T[n-i]y[i].$$

Ustalając $T = 2N$ oraz ograniczając się do $n \in [0 \dots N-1]$, powyższą sumę można rozłożyć na dwa składniki, korzystając przy tym z okresowości sygnału x_T :

$$(45) \quad \begin{aligned} (x_T \star y)[n] &= \sum_{i=0}^n x_T[n-i]y[i] + \sum_{i=n+1}^{N-1} x_T[n-i]y[i] \\ &= \sum_{i=0}^n x_T[\underbrace{n-i}_{\in [0 \dots N-1]}]y[i] + \sum_{i=n+1}^{N-1} x_T[\underbrace{n-i+2N}_{\in [N+1 \dots 2N-1]}]y[i]. \end{aligned}$$

Do tej pory nie wprowadziliśmy żadnych założeń dotyczących sygnału x . Zdefiniujmy go więc w taki sposób, że

$$(46) \quad x[i] = \begin{cases} 0 & : i < 0 \\ g[i] & : i \in [0 \dots N-1] \\ g[i-2N] & : i \in [N \dots 2N-1] \\ 0 & : i \geq 2N \end{cases}$$

dla pewnego sygnału dyskretnego $g[i]$ zdefiniowanego na nośniku $[-N \dots N]$. Wtedy $x_{2N}[i] = x[i]$ dla każdego i z przedziału $[0 \dots 2N-1]$.

Można zauważyć, że

$$(47) \quad \begin{aligned} (x_T \star y)[n] &= \sum_{i=0}^n g[n-i]y[i] + \sum_{i=n+1}^{N-1} g[n-i]y[i] \\ &= \sum_{i=0}^{N-1} g[n-i]y[i] \\ &= \sum_{i=-\infty}^{+\infty} g[n-i]y[i] = (g \star y)[n], \end{aligned}$$

co w połączeniu z twierdzeniem 5.3 pozwala na sformułowanie bardzo praktycznego twierdzenia.

Twierdzenie 5.4 (szczególne twierdzenie o splocie). *Jeżeli g jest sygnałem dyskretnym na nośniku $[-N \dots N]$, zaś y jest sygnałem dyskretnym na nośniku $[0 \dots N-1]$, to*

$$DFT(g \star y) = DFT(x_{2N})DFT(y_{2N}),$$

przy czym $DFT(\cdot)$ oznacza Dyskretną Transformatę Fouriera, a x jest zdefiniowany zgodnie z równaniem (46).

Praktyczny wniosek płynący z twierdzenia 5.4 jest taki, że spłot dowolnych sygnałów g i y spełniających założenia twierdzenia można policzyć z wykorzystaniem DFT (a więc i FFT). Transformaty obliczane są wtedy na sygnałach o długości $2N$, a sygnał y_{2N} może być otrzymany z sygnału y poprzez wypełnienie zerami na przedziale $[N \dots 2N - 1]$.

Powyższe rozważania można przeprowadzić w analogiczny sposób dla sygnałów wielowymiarowych. Warto jednak zauważyć, że obliczenie spłotu sygnałów wielowymiarowych wymagać będzie dwukrotnego zwiększenia długości sygnału w każdym wymiarze, zatem przy D wymiarach objętość sygnału y wzrośnie o czynnik 2^D .

2. Zastosowanie twierdzenia o splocie

Kulombowskie elementy macierzowe (28) można zapisać, grupując funkcje falowe z tym samym argumentem, jako

$$(48) \quad \langle i j k l \rangle = \iint (\Psi_i^* \Psi_l)(\vec{r}_1) G(\vec{r}_1 - \vec{r}_2) (\Psi_j^* \Psi_k)(\vec{r}_2) dV_2 dV_1,$$

gdzie $G(\vec{r})$ wyraża rodzaj oddziaływania, a w szczególnym przypadku oddziaływania kulombowskiego ma postać (29).

Jeżeli funkcje falowe mają dwie składowe spinowe, wyrażenie powyższe należy nie tylko wyciąkować po sześciu wymiarach przestrzennych, ale także po dwóch (dyskretnych) wymiarach spinowych, co sprowadza się do

$$(49) \quad \langle i j k l \rangle = \iint (\Psi_{i\uparrow}^* \Psi_{l\uparrow} + \Psi_{i\downarrow}^* \Psi_{l\downarrow})(\vec{r}_1) G(\vec{r}_1 - \vec{r}_2) (\Psi_{j\uparrow}^* \Psi_{k\uparrow} + \Psi_{j\downarrow}^* \Psi_{k\downarrow})(\vec{r}_2) dV_2 dV_1.$$

Wprowadzając oznaczenia

$$(50) \quad \begin{aligned} \rho_{il} &= \Psi_{i\uparrow}^* \Psi_{l\uparrow} + \Psi_{i\downarrow}^* \Psi_{l\downarrow} \\ \rho_{jk} &= \Psi_{j\uparrow}^* \Psi_{k\uparrow} + \Psi_{j\downarrow}^* \Psi_{k\downarrow} \end{aligned}$$

można zapisać kulombowski element macierzowy $\langle i j k l \rangle$ w postaci

$$(51) \quad \langle i j k l \rangle = \iint \rho_{il}(\vec{r}_1) G(\vec{r}_1 - \vec{r}_2) \rho_{jk}(\vec{r}_2) dV_2 dV_1,$$

co ma analogiczną postać do energii oddziaływania dwóch ładunków ciągłych o gęstościach ρ_{il} i ρ_{jk} .

Istotnie, w przypadku $i = l$ oraz $j = k$, wielkości ρ_{il} i ρ_{jk} (nazwijmy je *kwazigęstościami*) są rzeczywiste i wyrażają gęstości prawdopodobieństwa związane z poszczególnymi funkcjami falowymi. W ogólności jednak kwazigęstości, będąc wielkościami zespolonymi, nie muszą posiadać klasycznej interpretacji.

Zdefiniujmy następnie potencjał V_{jk} pochodzący od kwazigęstości ρ_{jk} , czyli

$$(52) \quad V_{jk}(\vec{r}_1) = \int G(\vec{r}_1 - \vec{r}_2) \rho_{jk}(\vec{r}_2) dV_2.$$

Przechodząc do przypadku dyskretnego, otrzymamy

$$(53) \quad V_{jk}[\vec{r}_1] = \sum_{\vec{r}_2} G[\vec{r}_1 - \vec{r}_2] \rho_{jk}[\vec{r}_2] h^3 = (G \star \rho_{jk})[\vec{r}_1] h^3$$

zaś poszukiwana wielkość $\langle i j k l \rangle$ wyrazi się jako

$$(54) \quad \langle i j k l \rangle = \sum_{\vec{r}_1} \rho_{il}[\vec{r}_1] V_{jk}[\vec{r}_1] h^3.$$

Na podstawie twierdzenia 5.4, można wykorzystać Szybką Transformatę Fouriera do wyznaczenia pola potencjału V_{jk} w czasie $O(\Gamma \log \Gamma)$, gdzie Γ jest ilością punktów siatki obliczeniowej. Podobnie, obliczenie wartości $\langle i j k l \rangle$ wymaga tylko pojedynczego sumowania po całej siatce, a więc zaledwie $O(\Gamma)$ operacji. Skoro dla struktur niezdegenerowanych $\Gamma \sim N$, gdzie N oznacza liczbę atomów struktury (rozdział 2.2), obliczenie wartości kulombowskich elementów macierzowych może być dokonane w czasie $O(N \log N)$.

ROZDZIAŁ 6

Obliczenia dla wielu całek

W przypadku metody oddziaływania konfiguracji, do wyznaczenia własności ekscytonowych, a szczególnie kompleksów ekscytonów, wymagane jest obliczenie ok. 10^4 różnych elementów macierzowych operatora Coulomba. W celu znaczącego zredukowania ilości wykonywanych obliczeń, można wykorzystać dobrze określone symetrie zachodzące pomiędzy całkami.

1. Podstawowe symetrie

Na podstawie analitycznej postaci kulombowskich elementów macierzowych (51) można zauważyć, że

$$\begin{aligned}
 \langle i j k l \rangle &= \iint \rho_{il}(\vec{r}_1) G(\vec{r}_1 - \vec{r}_2) \rho_{jk}(\vec{r}_2) dV_2 dV_1 \\
 (55) \qquad &= \iint \rho_{jk}(\vec{r}_1) G(\vec{r}_1 - \vec{r}_2) \rho_{il}(\vec{r}_2) dV_2 dV_1 = \langle j i l k \rangle.
 \end{aligned}$$

Co więcej, ponieważ

$$\begin{aligned}
 \rho_{il} &= \Psi_{i\uparrow}^* \Psi_{l\uparrow} + \Psi_{i\downarrow}^* \Psi_{l\downarrow} = (\Psi_{l\uparrow}^* \Psi_{i\uparrow} + \Psi_{l\downarrow}^* \Psi_{i\downarrow})^* = \rho_{li}^* \\
 (56) \qquad \rho_{jk} &= \Psi_{j\uparrow}^* \Psi_{k\uparrow} + \Psi_{j\downarrow}^* \Psi_{k\downarrow} = (\Psi_{k\uparrow}^* \Psi_{j\uparrow} + \Psi_{k\downarrow}^* \Psi_{j\downarrow})^* = \rho_{kj}^*,
 \end{aligned}$$

a wartości funkcji oddziaływania $G(\vec{r}_1 - \vec{r}_2)$ są rzeczywiste, więc zachodzi również

$$\begin{aligned}
 \langle i j k l \rangle &= \iint \rho_{il}(\vec{r}_1) G(\vec{r}_1 - \vec{r}_2) \rho_{jk}(\vec{r}_2) dV_2 dV_1 \\
 (57) \qquad &= \iint \rho_{li}(\vec{r}_1)^* G(\vec{r}_1 - \vec{r}_2)^* \rho_{kj}(\vec{r}_2)^* dV_2 dV_1 = \langle l k j i \rangle^*.
 \end{aligned}$$

Połączenie symetrii (55) oraz (57) pozwala na czterokrotne (w przybliżeniu) zmniejszenie liczby niezależnych kulombowskich elementów macierzowych, według poniższego diagramu:

$$\begin{aligned}
 \langle i j k l \rangle &= \langle j i l k \rangle \\
 (58) \qquad \parallel &\qquad \parallel \\
 \langle l k j i \rangle^* &= \langle k l i j \rangle^*
 \end{aligned}$$

Na podstawie diagramu można również zauważyć, że skoro

$$(59) \quad \begin{aligned} \langle i j k l \rangle &= \langle l k j i \rangle^* \\ \langle i j k l \rangle &= \langle k l i j \rangle^*, \end{aligned}$$

to całki postaci $\langle i j j i \rangle$ oraz $\langle i j i j \rangle$ są rzeczywiste ($z = z^* \iff z \in \mathbb{R}$).

2. Degeneracja Kramersa

W przypadku układów złożonych z nieparzystej liczby fermionów, stany jedno-ciałowe (elektronowe i dziurowe) układu są podwójnie zdegenerowane, o ile układ nie jest poddany działaniu pola magnetycznego. Zjawisko to nosi nazwę *degeneracji Kramersa* i wynika z niezmienniczości hamiltonianu względem operacji odwrócenia czasu ($t \mapsto -t$).

W metodzie ciasnego wiązania, degeneracja Kramersa wyraża się w tym, iż każdemu wektorowi własnemu Ψ odpowiada stan Ψ^D (nazwijmy go stanem dubletowym) o identycznej energii, taki, że

$$(60) \quad \begin{aligned} \Psi_{\uparrow}^D &= \Psi_{\downarrow}^* \\ \Psi_{\downarrow}^D &= -\Psi_{\uparrow}^*. \end{aligned}$$

Na skutek powyższego, dla dowolnych stanów Ψ_j, Ψ_k zachodzi

$$(61) \quad \rho_{j^D k^D} = \Psi_{j\uparrow}^{D*} \Psi_{k\uparrow}^D + \Psi_{j\downarrow}^{D*} \Psi_{k\downarrow}^D = \Psi_{k\uparrow}^* \Psi_{j\uparrow} + \Psi_{k\downarrow}^* \Psi_{j\downarrow} = \rho_{kj},$$

na podstawie czego można stworzyć diagram symetrii dla degeneracji Kramersa

$$(62) \quad \begin{array}{ccc} \langle i j k l \rangle & = & \langle i k^D j^D l \rangle \\ \parallel & & \parallel \\ \langle l^D j k i^D \rangle & = & \langle l^D j^D k^D i^D \rangle, \end{array}$$

a dzięki obserwacji, iż $(\Psi^D)^D = -\Psi$, diagram ten może być zastosowany dla dowolnych stanów $\Psi_i, \Psi_j, \Psi_k, \Psi_l$ także wtedy, gdy są one stanami dubletowymi. W połączeniu z (58), pozwala to na znaczne (nawet szesnastokrotne) zredukowanie liczby całek potrzebnych w obliczeniach.

Ponadto, kwazigęstości postaci $\rho_{i^D i}$ (oraz analogicznie $\rho_{i i^D}$) są, na podstawie (60), równe

$$(63) \quad \rho_{i^D i} = \Psi_{i\downarrow} \Psi_{i\uparrow} - \Psi_{i\uparrow} \Psi_{i\downarrow} = 0,$$

dzięki czemu całki $\langle i j k l \rangle$ takie że $\Psi_i = \pm \Psi_l^D$ lub $\Psi_j = \pm \Psi_k^D$ są tożsamościowo równe zeru.

3. Zastosowanie dla określonych całek

Metoda oddziaływania konfiguracji wymaga obliczenia elementów macierzowych w postaci $\langle e e e e \rangle$, $\langle h h h h \rangle$, $\langle e h h e \rangle$ oraz $\langle e h e h \rangle$. Każda z powyższych notacji oznacza cały podzbiór elementów macierzowych; litera e na danym miejscu oznacza, że odpowiedni stan jest jednym ze stanów elektronowych, zaś h — jednym ze stanów dziurowych. Można zauważyć, że w przypadku M stanów elektronowych i dziurowych, każdy z powyższych podzbiorów zawiera $O(M^4)$ różnych elementów macierzowych operatora Coulomba.

Diagramy (58) oraz (62) są prawdziwe dla dowolnych funkcji falowych. Można zauważyć, że dowolna zamiana stanów zgodna z diagramami nie spowoduje przechożenia pomiędzy wymienionymi powyżej zbiorami, może jednak spowodować otrzymanie całki należącej do innej grupy niż powyższe. Z związku z tym, dla każdej ze zdefiniowanych grup całek, wynikający z symetrii efektywny zysk obliczeniowy będzie inny.

3.1. Całki elektronowe i dziurowe. W przypadku całek elektronowych i dziurowych, czyli $\langle e e e e \rangle$ i $\langle h h h h \rangle$, dowolna zamiana stanów nie spowoduje zmiany rodzaju całki. Dzięki temu, możliwe jest wykorzystanie maksymalnej liczby tożsamości.

Korzystając tylko z symetrii podstawowych można na podstawie diagramu (58) wypisać cztery równoważne całki

$$(64) \quad \langle i j k l \rangle = \langle j i l k \rangle = \langle l k j i \rangle^* = \langle k l i j \rangle^*,$$

zaś przy dodatkowym wykorzystaniu degeneracji Kramersa aż szesnaście:

$$(65) \quad \begin{aligned} \langle i j k l \rangle &= \langle j i l k \rangle = \langle l k j i \rangle^* = \langle k l i j \rangle^* \\ &= \langle i k^D j^D l \rangle = \langle j l^D i^D k \rangle = \langle l j^D k^D i \rangle^* = \langle k i^D l^D j \rangle^* \\ &= \langle l^D j k i^D \rangle = \langle k^D i l j^D \rangle = \langle i^D k j l^D \rangle^* = \langle j^D l i k^D \rangle^* \\ &= \langle l^D k^D j^D i^D \rangle = \langle k^D l^D i^D j^D \rangle = \langle i^D j^D k^D l^D \rangle^* = \langle j^D i^D l^D k^D \rangle^*. \end{aligned}$$

3.2. Całki postaci $\langle e h h e \rangle$. Dla całek postaci $\langle e h h e \rangle$, dozwolone są już tylko niektóre permutacje indeksów. Na skutek tego, podstawowe symetrie pozwalają już tylko na jedną tożsamość

$$(66) \quad \langle i j k l \rangle = \langle l k j i \rangle^*,$$

zaś w przypadku degeneracji Kramersa osiem:

$$\begin{aligned}
 (67) \quad \langle i j k l \rangle &= \langle l k j i \rangle^* \\
 &= \langle i k^D j^D l \rangle = \langle l j^D k^D i \rangle^* \\
 &= \langle l^D j k i^D \rangle = \langle i^D k j l^D \rangle^* \\
 &= \langle l^D k^D j^D i^D \rangle = \langle i^D j^D k^D l^D \rangle^*.
 \end{aligned}$$

3.3. Całki postaci $\langle e h e h \rangle$. Dla całek postaci $\langle e h e h \rangle$, podobnie jak dla $\langle e h h e \rangle$, na podstawie symetrii podstawowych można skorzystać tylko z jednej permutacji indeksów

$$(68) \quad \langle i j k l \rangle = \langle k l i j \rangle^*,$$

a w przypadku degeneracji Kramersa z czterech tożsamości pomiędzy całkami:

$$\begin{aligned}
 (69) \quad \langle i j k l \rangle &= \langle k l i j \rangle^* \\
 &= \langle k^D l^D i^D j^D \rangle = \langle i^D j^D k^D l^D \rangle^*.
 \end{aligned}$$

Część 3

Implementacja

Opis pakietu oprogramowania

Metody opisane w niniejszej pracy zostały zaimplementowane w postaci autonomicznego pakietu oprogramowania, zaprojektowanego z myślą o wielkoskalowych obliczeniach. Pakiet stanowi zbiór programów, z których każdy wypełnia określone zadanie obliczeniowe, oraz wspólne dla wszystkich programów klasy i funkcje pomocnicze. Poszczególne programy mogą być uruchamiane zarówno na zwykłych stacjach roboczych, jak i na klastrach obliczeniowych, także z wykorzystaniem systemu PL-Grid.

1. Informacje podstawowe

1.1. Język programowania. Kod źródłowy stworzony został w pełni obiektowo w języku C++, w dialekcie C++98. Za wyborem tego języka programowania przemawiała zarówno wydajność kompilowanego w nim kodu, jak i dobre wsparcie dla programowania równoległego, w postaci interfejsów MPI, dyrektyw OpenMP oraz możliwości dołączania kodu napisanego dla CUDA.

Zrezygnowano z wykorzystania możliwości aktualnego standardu C++11 z uwagi na brak powszechnego wsparcia dla tego dialektu na klastrach obliczeniowych. Zrezygnowano również z wyodrębniania części kodu jako dynamicznie dołączanej biblioteki, na rzecz umożliwienia swobodnego przenoszenia binarnych plików wykonywalnych między systemami.

1.2. Wykorzystane biblioteki. Przy tworzeniu oprogramowania celowo zrezygnowano z wykorzystania niestandardowych, zewnętrznych bibliotek, w celu uniknięcia konieczności ręcznej instalacji biblioteki w przypadku ewentualnej jej niedostępności na klastrze obliczeniowym.

Wyjątek uczyniono dla biblioteki numerycznej GSL (*GNU Scientific Library*), na której opiera się obsługa elementów algebry liniowej, oraz dla implementacji Szybkiej Transformaty Fouriera o nazwie FFTW w wersji 3, wykorzystywanej przy operacjach splotu.

Warto wspomnieć, że kompilacja programów jest możliwa także bez wykorzystania biblioteki GSL, wpływa to jednak na efektywność obliczeń (szczególnie dla operacji na wektorach i macierzach). W przypadku braku biblioteki FFTW, możliwa jest kompilacja wszystkich programów za wyjątkiem pqr-coulomb.

FLAGI		
Klucz	Opis	
--enable-mpi	włącz obsługę standardu MPI	
--enable-cuda	włącz obsługę modułów CUDA	
--disable-fftw	nie wykorzystuj biblioteki FFTW	
--disable-gsl	nie wykorzystuj biblioteki GSL	
--disable-openmp	wyłącz obsługę standardu OpenMP	
PARAMETRY		
Klucz	Typ	Opis
--fftw-lib	tekst	ścieżka zawierająca dynamiczną bibliotekę FFTW
--fftw-include	tekst	ścieżka zawierająca pliki nagłówkowe FFTW
--gsl-lib	tekst	ścieżka zawierająca dynamiczną bibliotekę GSL
--gsl-include	tekst	ścieżka zawierająca pliki nagłówkowe GSL

TABELA 1. Opcjonalne parametry konfiguracji pakietu pqr.

1.3. Kompilacja i instalacja. Proces kompilacji oprogramowania należy rozpocząć od wywołania z opcjonalnymi parametrami skryptu `configure`, którego zadaniem jest przygotowanie pliku `Makefile`. Możliwe parametry konfiguracji podane są w tabeli 1, żaden z tych parametrów nie jest obowiązkowy.

Po konfiguracji pakietu poleceniem `./configure`, kompilacja sprowadza się do podania polecenia `make`, które skompiluje wszystkie pliki wykonywalne i umieści je w podkatalogu `bin`. Jeśli programy mają być zainstalowane w systemie na stałe, należy je po prostu przenieść do wybranego katalogu, np. poleceniem

```
cp bin/* /usr/local/bin/
```

2. Interfejs użytkownika

2.1. Parametry wiersza poleceń. Wszelkie parametry dla programów przekazywane są przy pomocy wiersza poleceń. Programy rozróżniają trzy rodzaje danych wejściowych:

- flagi (niezawierające wartości), symbolizujące włączenie (lub wyłączenie) określonej funkcjonalności, np. `-N` lub `--overwrite`,
- parametry (z przypisaną wartością), np. `--output-file="out.dat"`, oznaczające nazwane parametry przekazywane do programu
- argumenty (wszystkie pozostałe), zawierające obowiązkowe dane wejściowe

Każdy z programów, uruchomiony z flagą `-h` lub `--help`, wyświetla na standardowe wyjście pełny opis składni. Jeżeli składnia jest błędna (np. program został uruchomiony bez wymaganych parametrów), specyfikacja składni wypisywana jest na standardowy strumień diagnostyczny. Dzięki temu, nawet gdy program uruchamiany

jest w tle i/lub z przekierowaniem standardowego wyjścia do pliku (jak często bywa w przypadku zadań obliczeniowych), użytkownik zostanie powiadomiony o błędnym wywołaniu programu.

Nazwy poszczególnych flag wiersza poleceń zależą od funkcjonalności danego programu i są szczegółowo omówione w rozdziale 9. Warto natomiast wspomnieć, że nazwy flag lub parametrów pokrywają się w różnych programach wtedy i tylko wtedy, gdy opisują dokładnie tę samą funkcjonalność.

2.2. Komunikaty. Komunikaty dla użytkownika wypisywane są przez programy na standardowe wyjście. Szczególną klasę komunikatów stanowią informacje o przebiegu kolejnych etapów działania programu, jak na przykład

Reading files... (2.721 s)

Na standardowe wyjście wypisywane są również wszelkie komunikaty o błędach programu, na przykład

ERROR: parameter 'output-file' not specified

W razie wypisania komunikatu o błędzie, program zazwyczaj kończy swoje działanie.

3. Obliczenia z wykorzystaniem systemu PL-Grid

Dzięki dobrze zdefiniowanemu interfejsowi wejścia/wyjścia oraz możliwości kompilacji programów do niewielkich, niezależnych plików wykonywalnych, opisywana implementacja doskonale nadaje się do obliczeń gridowych z wykorzystaniem systemu PL-Grid. Zarówno przy pomocy interfejsu gLite, jak i środowiska UNICORE możliwe jest przygotowanie zadania obliczeniowego i uruchomienie go na wskazanym systemie.

Najwygodniejszy sposób przygotowania zadania umożliwia środowiska graficznego UNICORE Rich Client¹, oparte o framework Eclipse. Przy tworzeniu nowego zadania na wybranym systemie (opcja *create job*), należy jako typ zadania wybrać *Script*. Po przejściu do graficznej edycji zadania, należy w głównym polu tekstowym stworzyć skrypt startowy żadanego programu, zawierający nazwę programu (np. *pqr-mapper*) wraz z parametrami wywołania. Odwołania do wszystkich plików powinny być skonstruowane tak, jakby znajdowały się one w bieżącym katalogu, np.

```
./pqr-coulomb --dielectric=InAs listing
```

Następnie, wszystkie pliki, do których odwołuje się skrypt (łącznie z plikiem wykonywalnym) należy dodać do listy importowanych plików w zakładce *Files*. Możliwe jest dodanie pliku bezpośrednio z lokalnego komputera, jak również wskazanie pliku przeniesionego uprzednio do zdalnego magazynu (wtedy w polu *Source Type* należy wybrać *UNICORE_Storage* zamiast *Local_File*).

¹<http://www.unicore.eu/download/unicore6/>

W celu zebrania wyników działania programu, należy do listy eksportowanych plików (również w zakładce *Files*) podać w polu *File(s) in Job Directory* nazwę pliku wynikowego, który ma zostać wyeksportowany. W polu *Source Type* należy wybrać opcję *UNICORE_Storage*, aby plik został wyeksportowany do magazynu UNICORE po zakończeniu obliczeń.

Oczywiście, nic nie stoi na przeszkodzie, aby połączyć działanie kilku następujących po sobie programów. Można zrealizować to w standardowy sposób, wywołując kolejno kilka plików wykonywalnych wewnątrz skryptu, lub też skorzystać z funkcjonalności „Workflow”.

4. Składnia plików danych

Poza informacjami pobranymi od użytkownika poprzez wiersz poleceń, większość programów wymaga dostępu do określonych danych fizycznych, na podstawie których wykonane mają zostać obliczenia. Aby zagwarantować zarówno efektywny, jak i (co ważniejsze) bezbłędny zapis i odczyt tych danych, składnia zawierających je plików powinna być dobrze zdefiniowana.

Dla jak najlepszej kompatybilności z istniejącym oprogramowaniem, np. programami do optymalizacji geometrii lub pakietem do obliczeń metodą ciasnego wiązania (dr M. Zieliński), wszędzie tam, gdzie było to możliwe, zapożyczono format plików stosowany w dotychczas wykorzystywanych programach.

4.1. Pliki struktury atomowej. Każdy z plików tego rodzaju opisuje położenia i pierwiastki poszczególnych atomów tworzących strukturę badanego układu. Pliki mają bardzo prosty format tekstowy, po jednej linii na każdy atom. Pojedyncza linia pliku zawiera oddzielonych odstępami siedem liczb, oznaczających kolejno

- współrzędne x , y , z położenia atomu (trzy liczby rzeczywiste),
- dwa pola zawierające zera, obecnie nieużywane,
- identyfikator związku chemicznego A (liczba naturalna),
- identyfikator pierwiastka B (liczba naturalna).

Dla zachowania kompatybilności wstecznej, wszystkie powyższe liczby zapisywane są jako zmiennoprzecinkowe. Identyfikator związku chemicznego jest liczbą, oznaczającą pozycję na liście materiałów podawanych przez użytkownika w wierszu poleceń jako wartość `--materials`. Identyfikator pierwiastka wyraża natomiast kolejność pierwiastka w sumarycznym wzorze chemicznym; kolejność ta liczona jest od końca, czyli od anionu. Przykład identyfikacji podany jest w tabeli 2.

4.2. Pliki współczynników ciasnego wiązania. Metoda ciasnego wiązania pozwala na wyznaczenie współczynników $c_{i\alpha}$ (równanie 1, str. 7) dla wszystkich (pełna

A	B	pierwiastek
1	1	As
1	2	In
2	1	As
2	2	Ga

TABELA 2. Przykładowa identyfikacja atomów w pliku struktury dla parametru `--materials` o wartości `InAs,GaAs`.

diagonalizacja hamiltonianu) lub dla wybranych stanów jednoelektronowych. Wyznaczone współczynniki poszczególnych stanów zapisywane są w osobnych plikach tekstowych o ustalonej składni.

Pierwsza linia pliku zawiera kilka pól w odpowiednim porządku, oddzielonych pojedynczym odstępem:

- część rzeczywistą energii własnej (liczba rzeczywista),
- część urojoną energii własnej² (liczba rzeczywista),
- słowo `Orbitals`,
- liczbę spinorbitali³ bazy funkcyjnej (liczba naturalna),

po których następuje pewna liczba innych pól, współcześnie niewykorzystywanych. Dla zachowania kompatybilności wstecznej, wszystkie liczby oprócz wymiaru bazy funkcyjnej zapisywane są jako zmiennoprzecinkowe.

Po powyższym nagłówku następuje, dla każdego atomu i z układu, blok danych zawierający dla każdego spinorbitala α po jednej linii, składającej się z oddzielonych pojedynczym odstępem części rzeczywistej i urojonej współczynnika $c_{i\alpha}$.

4.3. Pliki orbitali atomowych. Do przeprowadzenia rekonstrukcji funkcji falowej, konieczna jest znajomość analitycznej (lub numerycznej) postaci orbitali atomowych. Część kątowna orbitali (rysunek 1, str. 15) jest dobrze określona i wymaga jedynie ustalenia obowiązujących symetrii bazy funkcyjnej, część radialna musi natomiast zostać wybrana arbitralnie. Aby przy każdej kolejnej rekonstrukcji nie było konieczności generowania orbitali bazowych, są one przechowywane w plikach tekstowych o wyjątkowo prostej składni.

Pliki orbitali danego rodzaju zgrupowane są w katalogi o ściśle ustalonej hierarchii. W ustalonym katalogu znajdują się podkatalogi o nazwach odpowiadających symbolom pierwiastków. W każdym z tych podkatalogów znajdują się pliki o nazwach: `s`, `p`, `d`, `ss` zawierające przebiegi poszczególnych orbitali.

²równą zero z dokładnością do błędów numerycznych

³liczba spinorbitali równa jest liczbie orbitali (np. 10 dla $sp^3d^5s^*$) pomnożonej przez dwa (dwie orientacje spinowe)

Pojedynczy plik z przebiegiem orbitala rozpoczyna się od linii nagłówkowej postaci np.

```
# count=50001 step=1.0000000000000000e-03
```

gdzie wartość „count” oznacza liczbę pozostałych linii w pliku, a „step” oznacza krok dyskretyzacji wartości orbitala (w angstromach).

Każda z pozostałych linii zawiera dwie liczby oddzielone pojedynczym odstępem:

- r , czyli odległość od centrum (atomu) [Å],
- $T(r)r$, czyli wartość orbitala pomnożoną przez czynnik skalujący.

Ponadto, odległości r wyrażone w kolejnych liniach zawsze zaczynają się od 0 i rosną liniowo z krokiem określonym w linii nagłówkowej. Ta pozorna nadmiarowość danych pozwala jednak na łatwą wizualizację orbitali, np. przy pomocy programu gnuplot. Jednocześnie, w łatwy sposób można zweryfikować, czy orbitale są poprawnie znormalizowane — powinno bowiem zachodzić na podstawie równania 20 (str. 16)

$$(70) \quad \sum_r (T(r)r)^2 \Delta r \approx \int T(r)^2 r^2 dr = 1.$$

Do generowania orbitali Slatera służy program pqr-orbslat, zaś do generowania orbitali Herman-Skillman program pqr-orbhers.

ROZDZIAŁ 8

Paralelizacja

1. MPI

Z uwagi na możliwość zastosowania projektu na różnego rodzaju klastrach obliczeniowych, dwa główne programy (pqr-mapper i pqr-coulomb) zostały zaprojektowane z wykorzystaniem interfejsu MPI (ang. *Message Passing Interface*) w wersji 2. Programy skompilowane z obsługą MPI mogą być uruchamiane zarówno przy pomocy polecenia `mpiexec`, jak i przez standardowe wywołanie pliku wykonywalnego z linii poleceń, wtedy program wykonany zostanie przez pojedynczy proces.

Ponadto, dzięki wykorzystaniu dyrektyw preprocesora w procesie konfiguracji i kompilacji, możliwa jest kompilacja programów również bez obsługi interfejsu MPI, także wtedy, gdy biblioteki MPI nie są dostępne. Obsługa MPI korzysta ze stworzonego na tę okazję niewielkiego interfejsu C++ stworzonego na bazie szablonów. Dzięki temu, możliwe było łatwe i czytelne wykorzystanie zaawansowanych funkcjonalności (`MPI_Scatterv`, `MPI_Gatherv`), które nie są wspierane przez upowszechnione interfejsy obiektowe do MPI (np. Boost.MPI). Programy poprawnie współpracują zarówno z implementacją MPICH2 jak i z OpenMPI.

Aby skompilować pakiet z obsługą MPI, należy do skryptu konfiguracyjnego podać flagę `--enable-mpi`.

2. OpenMP

Niezależnie od interfejsu MPI, obsługiwana jest również paralelizacja z wykorzystaniem standardu OpenMP. Dyrektywy `#pragma omp` wprowadzone są w kodzie dość powszechnie, wszędzie tam, gdzie tylko miało to uzasadnienie. Możliwe jest jednocześnie korzystanie z MPI i OpenMP, wtedy równoległe wątki tworzone są w obrębie każdego procesu.

W celu optymalnego dopasowania do architektury, możliwe jest ustalenie liczby wątków przypadających na proces przy uruchomieniu programu; pozwala na to parametr wiersza poleceń `--threads-per-node`. Jeżeli parametr nie zostanie podany, liczba wątków będzie ustalana automatycznie.

Domyślnie, programy kompilowane są z obsługą OpenMP. Jeżeli programy mają zostać skompilowane bez tej funkcjonalności, należy do skryptu konfiguracyjnego podać flagę `--disable-openmp`.

3. CUDA

Wraz z rozwojem mocy obliczeniowej kart graficznych i udostępnieniem interfejsów programistycznych (CUDA, OpenCL), obliczenia oparte na GPU (ang. *Graphics Processing Unit*) stają się korzystną alternatywą dużych serwerów obliczeniowych. Ten model programowania równoległego nadaje się głównie do rozwiązywania problemów, które dają się rozdzielić na dużą ilość analogicznych, choć niezależnych obliczeń. Ponadto, najlepsze wyniki można uzyskać wtedy, gdy czas obliczeń jest dużo większy od czasu przesyłania danych.

Obliczenia atomistyczne (strona 7), charakteryzujące się kwadratową złożonością czasową oraz względnie prostym schematem obliczeń (który dodatkowo można przedstawić w postaci operacji macierzowych), doskonale nadają się do implementacji przy pomocy GPU. Obsługa interfejsu CUDA została więc zaimplementowana w programie pqr-atomist służącym do masowego obliczania elementów macierzowych na podstawie współczynników TB.

Dla maksymalnej elastyczności, program skompilowany z obsługą CUDA może wykonywać obliczenia zarówno na GPU, jak i na CPU, a wybór jednostki obliczeniowej dokonywany jest dopiero przy uruchomieniu programu. Aby skorzystać z karty graficznej, należy na wierszu poleceń podać flagę `--use-cuda`. Dodatkowo, aby dokonać wyboru GPU, na którym ma zostać uruchomiony program, należy podać flagę z opcjonalną liczbą naturalną, np. `--use-cuda=1`, oznaczającą identyfikator żądanej karty graficznej. Jeżeli karta nie zostanie wskazana *explicite*, zostanie wybrana automatycznie.

Moduł CUDA jest kompilowany przy pomocy kompilatora `nvc`, wymaga zatem zainstalowanego środowiska NVIDIA SDK. Aby program został skompilowany z obsługą GPU, należy podać do skryptu konfiguracyjnego flagę `--enable-cuda`. W przeciwnym razie, będzie możliwe uruchamianie programu jedynie na CPU.

Elementy składowe pakietu

1. Program pqr-atomist

Program pqr-atomist pozwala na przeprowadzenie obliczeń atomistycznych, czyli obliczenia wartości wybranych elementów macierzowych operatora kulombowskiego na podstawie struktury atomowej i współczynników ciasnego wiązania. Program został zaprojektowany z użyciem wymiennego modułu GPU/CPU realizującego główną część obliczeń, dzięki czemu możliwy jest swobodny wybór jednostki obliczeniowej w czasie wykonania programu.

Praca programu rozpoczyna się od przetworzenia parametrów wiersza poleceń i wczytaniu zadanych plików wejściowych: pliku struktury atomowej i plików ze współczynnikami TB poszczególnych stanów jednoelektronowych. Następnie, wykorzystując formalizm macierzowy obliczeń atomistycznych (str. 9), tworzona jest macierz C zawierająca kwazigęstości ładunku pochodzące od wszystkich par stanów.

Dalsza część obliczeń wykonywana jest przy pomocy wymiennego modułu GPU/CPU. Moduł realizuje dwie funkcje:

- pomnożenia macierzy C przez dynamicznie tworzoną macierz V ,
- pomnożenia dwóch zadanych macierzy (C^T oraz VC).

W wersji CPU, moduł składa się z funkcji obliczających bezpośrednio żądane wielkości, zaś wersja GPU zawiera interfejsy do wywołania kerneli CUDA. W obu przypadkach zostały wykorzystane implementacje BLAS.

Po zebraniu wyników, zostają one uporządkowane i zapisane do zadanych plików wyjściowych. Warto wspomnieć, że programy pqr-atomist i pqr-coulomb stosują identyczny format plików wynikowych, dzięki czemu możliwe jest łatwe porównanie i prezentacja wyników. Wszystkie możliwe opcje uruchomienia programu zawiera tabela 1 (str. 46).

2. Program pqr-coulomb

Program pqr-coulomb ma za zadanie obliczanie wartości wskazanych elementów macierzowych operatora kulombowskiego na podstawie funkcji falowych zrekonstruowanych uprzednio za pomocą programu pqr-mapper. Program zaprojektowany jest z obsługą interfejsu MPI, a wszystkie możliwe opcje uruchomienia zawiera tabela 2.

FLAGI		
Klucz	Opis	
--use-cuda	wybierz automatycznie kartę GPU	
-w / --overwrite	zezwól na nadpisanie plików wynikowych	
PARAMETRY		
Klucz	Typ	Opis
--atoms	tekst	ścieżka do pliku struktury atomowej
--dielectric	$\mathbb{R}$ / materiał	względna stała dielektryczna ϵ_r , domyślnie: 1
--materials	tekst	lista wzorów sumarycznych kolejnych materiałów struktury
--output-dir	tekst	ścieżka do zapisania plików wynikowych
--use-cuda	$\mathbb{N}$	identyfikator wybranej karty GPU
--trunc-radius	$\mathbb{R}$ [Å]	zasięg oddziaływania, domyślnie: bez ograniczeń
ARGUMENTY		
	Krotność	Typ Opis
	1	tekst ścieżka do pliku listingowego

TABELA 1. Opcje uruchomienia programu pqr-atomist.

TABELA 1. Opcje uruchomienia programu pqr-atomist.

Po uruchomieniu programu, proces główny przetwarza parametry przekazane w wierszu polecenia, po czym wczytuje wskazany plik listingowy oraz pliki wymienionych w nim funkcji falowych. Następnie, niezbędne dane (w tym funkcje falowe) rozsyłane są do wszystkich procesów.

Na podstawie listy wskazanych elementów macierzowych oraz obowiązujących pomiędzy nimi symetrii, proces główny wyznacza równoważny zbiór niezależnych elementów, minimalizując przy tym liczbę przeprowadzanych operacji splotu. Wyznaczona lista elementów zostaje następnie rozdzielona, w miarę równomiernie, pomiędzy wszystkie procesy.

Każdy proces, otrzymawszy listę elementów do obliczenia, rozpoczyna pracę, wyznaczając kolejne elementy macierzowe z wykorzystaniem twierdzenia o splocie i FFT. Po obliczeniu wszystkich całek, dane zbierane są przez proces główny, który na tej podstawie odtwarza wartości wszystkich elementów określonych przez symetrię, w szczególności wymienionych na wierszu poleceń. Dane wyjściowe zapisywane są następnie do plików tekstowych w podanym przez użytkownika katalogu.

3. Program pqr-listing

Zadaniem programu pqr-listing jest wygenerowanie, na podstawie współczynników ciasnego wiązania, pliku listingowego zawierającego wszystkie wskazane stany

FLAGI		
Klucz	Opis	
-S / --no-symmetries	nie wykorzystuj żadnych symetrii pomiędzy CME	
-w / --overwrite	zezwól na nadpisanie plików wynikowych	
PARAMETRY		
Klucz	Typ	Opis
--dielectric	$\mathbb{R}$ / materiał	względna stała dielektryczna ϵ_r , domyślnie: 1
--integrals	tekst	lista żądanych do obliczenia CME, domyślnie: eeee, hhhh, ehhe, eheh
--output-dir	tekst	ścieżka do zapisania plików wynikowych
--threads-per-node	$\mathbb{N}$	liczba wątków OpenMP na proces, domyślnie: wybór automatyczny
--tf-lattice	$\mathbb{R}$ [Å] / materiał	stała sieci krystalicznej dla modelu TF
--trunc-radius	$\mathbb{R}$ [Å]	zasięg oddziaływania, domyślnie: bez ograniczeń
--trunc-smooth	$\mathbb{R}$ [Å]	wygładzanie granicy zasięgu, domyślnie: 0 Å= brak wygładzania
ARGUMENTY		
Krotność	Typ	Opis
1	tekst	ścieżka do pliku listingowego

TABELA 2. Opcje uruchomienia programu pqr-coulomb.

TABELA 2. Opcje uruchomienia programu pqr-coulomb.

jednoelektronowe. Przy uruchomieniu z flagą -d, program identyfikuje zależności wynikające z degeneracji Kramersa i oznacza w generowanym pliku stany dubletowe.

Wszystkie dozwolone parametry wiersza polecenia zawiera tabela 3.

4. Program pqr-mapper

Program pqr-mapper służy do przestrzennej rekonstrukcji funkcji falowej na podstawie współczynników ciasnego wiązania. Program zaprojektowany jest z obsługą interfejsu MPI, zatem obliczenia są rozdzielone pomiędzy określoną liczbę komunikujących się między sobą procesów.

Po uruchomieniu, proces główny analizuje parametry podane na wierszu polecenia (tabela 4, str. 49) i wczytuje z dysku potrzebne dane: przebiegi orbitali bazowych, położenia atomów oraz listę funkcji falowych do odwzorowania. Na podstawie położenia atomów oraz zadanych przez użytkownika parametrów rekonstrukcji (margines odwzorowania, krok siatki) wyznaczone zostają wymiary siatki obliczeniowej i

FLAGI		
Klucz	Opis	
-d / --doublets	zidentyfikuj stany dubletowe	
-w / --overwrite	zezwól na nadpisanie plików wynikowych	
PARAMETRY		
Klucz	Typ	Opis
--fermi-level	$\mathbb{R}$ [eV]	energetyczny poziom Fermiego, domyślnie: 0 eV
--output-file	tekst	ścieżka do zapisania pliku listingowego
ARGUMENTY		
Krotność	Typ	Opis
dowolna	tekst	ścieżki do plików ze współczynnikami TB

TABELA 3. Opcje uruchomienia programu pqr-listing.

TABELA 3. Opcje uruchomienia programu pqr-listing.

położenie początku układu współrzędnych. Następnie, struktura atomowa zostaje podzielona i rozesłana pomiędzy procesy w sposób możliwie równomierny. Rozesłane zostają także wszelkie parametry konieczne do właściwego przeprowadzenia rekonstrukcji, po czym każdy z procesów alokuje potrzebne zasoby (siatkę obliczeniową).

Dla każdej zadanej funkcji falowej, proces główny wczytuje współczynniki ciąsnego wiązania, a następnie rozdziela je pomiędzy wszystkie procesy w taki sposób, aby każdy proces otrzymał współczynniki dotyczące atomów uprzednio jemu przypisanych. Każdy z procesów rekonstruuje wkład do funkcji falowej pochodzący tylko i wyłącznie od odpowiednich atomów. Po zakończeniu rekonstrukcji, domeny obliczeniowe wszystkich procesów zostają zsumowane i przesłane do procesu głównego, gdzie dokonywana jest normalizacja funkcji falowej. Następnie, przebieg zrekonstruowanej funkcji zapisywany jest do plików wyjściowych przez proces główny.

W celu usprawnienia procesu rekonstrukcji, zarówno część radialna orbitali (wczytywana z pliku), jak również część kątowna, są dyskretyzowane na bardzo drobnej siatce, dzięki czemu nie jest konieczne analityczne wyznaczanie ich wartości w poszczególnych punktach siatki obliczeniowej. Dyskretyzacja jest na tyle dokładna, że dla zaoszczędzenia czasu obliczeń stosowana może być prosta interpolacja liniowa, a w przypadku części kątownej — dwuliniowa.

FLAGI		
Klucz	Opis	
-N / --do-not-normalize	nie normalizuj funkcji falowych	
-w / --overwrite	zezwól na nadpisanie plików wynikowych	
PARAMETRY		
Klucz	Typ	Opis
--atoms	tekst	ścieżka do pliku struktury atomowej
--cutoff	$\mathbb{R}$ [Å]	promień odwzorowania funkcji falowej
--materials	tekst	lista wzorów sumarycznych kolejnych materiałów struktury
--orbital-dir	tekst	ścieżka zawierająca katalogi orbitali dla poszczególnych pierwiastków
--orbital-stretch	$\mathbb{R}$ [1/Å]	współczynnik ściskający orbitale
--output-dir	tekst	ścieżka do zapisania plików wynikowych
--step	$\mathbb{R}$ [Å]	krok siatki odwzorowania
--threads-per-node	$\mathbb{N}$	liczba wątków OpenMP na proces, domyślnie: wybór automatyczny
ARGUMENTY		
Krotność	Typ	Opis
1	tekst	ścieżka do pliku listingowego

TABELA 4. Opcje uruchomienia programu pqr-mapper.

TABELA 4. Opcje uruchomienia programu pqr-mapper.

ARGUMENTY		
Krotność	Typ	Opis
1	tekst	ścieżka do pliku listingowego

TABELA 5. Opcje uruchomienia programu pqr-normalize.

5. Program pqr-normalize

Program pqr-normalize ma za zadanie dokonać normalizacji wszystkich zrekonstruowanych uprzednio funkcji falowych, zawartych we wskazanym przez użytkownika pliku listingowym. Obliczenia wykonywane są sekwencyjnie dla kolejnych funkcji falowych, są za to zrównoleglone przy pomocy dyrektyw OpenMP.

Dozwolone parametry wiersza polecenia zawarte są w tabeli 5.

6. Program pqr-overlap

Program pqr-overlap pozwala na obliczanie całek nakrywania orbitali, czyli

$$(71) \quad O_{12} = \int \psi_1(\vec{r} - \vec{R}_1)^* \psi_2(\vec{r} - \vec{R}_2) dV$$

FLAGI		
Klucz	Opis	
-w / --overwrite	zezwól na nadpisanie plików wynikowych	
PARAMETRY		
Klucz	Typ	Opis
--cutoff	$\mathbb{R}$ [Å]	promień odwzorowania funkcji falowej
--offset	$\mathbb{R}^3$ [Å]	względne przesunięcie orbitali, np. 0:0:2
--orbital-dir	tekst	ścieżka zawierająca katalogi orbitali dla poszczególnych pierwiastków
--orbital-stretch	$\mathbb{R}$ [1/Å]	współczynnik ściskający orbitale
--output-file	tekst	ścieżka do zapisania wyniku obliczeń
--step	$\mathbb{R}$ [Å]	krok siatki odwzorowania
ARGUMENTY		
Krotność	Typ	Opis
2	tekst	symbole obu pierwiastków

TABELA 6. Opcje uruchomienia programu pqr-overlap.

TABELA 6. Opcje uruchomienia programu pqr-overlap.

w zależności od przesunięcia $\vec{R}_1 - \vec{R}_2$ i rodzaju orbitali. Wyniki obliczeń mogą posłużyć do oceny zgodności orbitali z założeniami metody TB, lub też ilościowego oszacowania wpływu nieortogonalności bazy na własności ekscytonowe.

Działanie programu jest analogiczne do programu pqr-mapper, z tą jedynie różnicą, że nie jest konieczne alokacja całej domeny obliczeniowej, dzięki czemu obliczenia mogą być wykonane bardzo dokładnie (program charakteryzuje się stałą złożonością pamięciową).

Dozwolone parametry wiersza polecenia zawarte są w tabeli 5. Po ich przetworzeniu, wyznaczane są granice przestrzennego obszaru obliczeń, obejmujące wspólną część obszarów odwzorowania obu atomów. Po ukończeniu obliczeń, wyniki wypisywane do zadanego pliku wynikowego są w postaci tabeli, zawierającej wielkości O_{12} dla wszystkich par orbitali bazowych.

Część 4

Wyniki

Wyniki obliczeń

1. Charakterystyka badanego układu

Badanym układem była cylindryczna kropka kwantowa InAs o podstawie eliptrycznej (półoś wielka: 12.5 nm, półoś mała: 11 nm) i wysokości 3 nm, zanurzona w cylindrycznej matrycy InP o średnicy 34 nm i wysokości 15 nm. Cały układ składał się z 533709 atomów. Do obliczeń wzięte zostały dwa stany jednoelektronowe odpowiadające energiom LUMO i HOMO, przy czym każdy z nich był dwukrotnie zdegenerowany (degeneracja Kramersa). Część obliczeń została wykonana z pomocą infrastruktury PL-Grid, w szczególności z wykorzystaniem poznańskiego klastra REEF.

Obliczenia rozpoczęto od analizy energii i współczynników TB przy pomocy programu pqr-listing, w celu stworzenia pliku listingowego. Obecność degeneracji Kramersa została poprawnie zidentyfikowana przez program, dzięki czemu wymagana była rekonstrukcja tylko dwóch z czterech funkcji falowych. Rekonstrukcja wykonana została przy pomocy programu pqr-mapper, przyjąwszy promień odwzorowania $R_c = 2$ nm oraz krok siatki $h = 0.08$ nm. Funkcje falowe układu odwzorowane zostały zarówno dla orbitali Slatera, jak i orbitali H-S.

2. Wartości elementów macierzowych

Na podstawie zrekonstruowanych funkcji falowych, obliczone zostały przy pomocy programu pqr-coulomb wartości¹ wybranych elementów macierzowych operatora Coulomba:

- całki kulombowskiej $J_{ee} = \langle e^1 e^1 e^1 e^1 \rangle$ opisującej odpychanie pomiędzy elektronami w najniższym stanie elektronowym,
- całki kulombowskiej $J_{hh} = \langle h^1 h^1 h^1 h^1 \rangle$ opisującej odpychanie pomiędzy dziurami w najwyższym stanie dziurowym,
- całki kulombowskiej $J_{eh} = \langle e^1 h^1 h^1 e^1 \rangle$ opisującej przyciąganie elektronu i dziury pomiędzy stanami HOMO i LUMO,
- całki $\langle e^1 h^2 e^2 h^1 \rangle$ opisującej rozszczepienie jasnego dubletu ekscytonowego,
- całki $\langle e^1 h^1 e^2 h^2 \rangle$ opisującej rozszczepienie ciemnego dubletu ekscytonowego.

¹w przypadku wartości zespolonych, podane zostały moduły

element	wartość (energia)				
	orbitale Slatera		orbitale H-S		atomistycznie
	ϵ_{InAs}	T-F	ϵ_{InAs}	T-F	
$\langle e^1 e^1 e^1 e^1 \rangle$ [meV]	15.875	15.899	15.237	15.263	15.466
$\langle e^1 h^1 h^1 e^1 \rangle$ [meV]	15.472	15.496	16.684	16.720	17.125
$\langle h^1 h^1 h^1 h^1 \rangle$ [meV]	15.184	15.226	18.790	18.906	19.663
$\langle e^1 h^2 e^2 h^1 \rangle$ [μ eV]	47.346	47.484	8.451	8.055	4.548
$\langle e^1 h^1 e^2 h^2 \rangle$ [μ eV]	0.334	0.337	0.160	0.152	0.027

TABELA 1. Wartości wybranych całek kulombowskich i wymiennych dla badanego układu cylindrycznej kropki kwantowej InAs/GaAs.

Dla obu rodzajów orbitali, powyższe wartości wyznaczone zostały zarówno dla modelu stałej dielektrycznej ($\epsilon_{InAs} = 15.15$), jak i dla ekranowania Thomasa-Fermiego.

Ponadto, przy pomocy programu pqr-atomist uruchomionego na GPU², powyższe elementy macierzowe wyznaczone zostały metodą atomistyczną. Nie uwzględniano przy tym ekranowania wg modelu Thomasa-Fermiego, ponieważ wpływa on na oddziaływanie dopiero na odległościach mniejszych niż odległości pomiędzy najbliższymi sąsiadami. Porównanie wyników otrzymanych obiema metodami zawiera tabela 1.

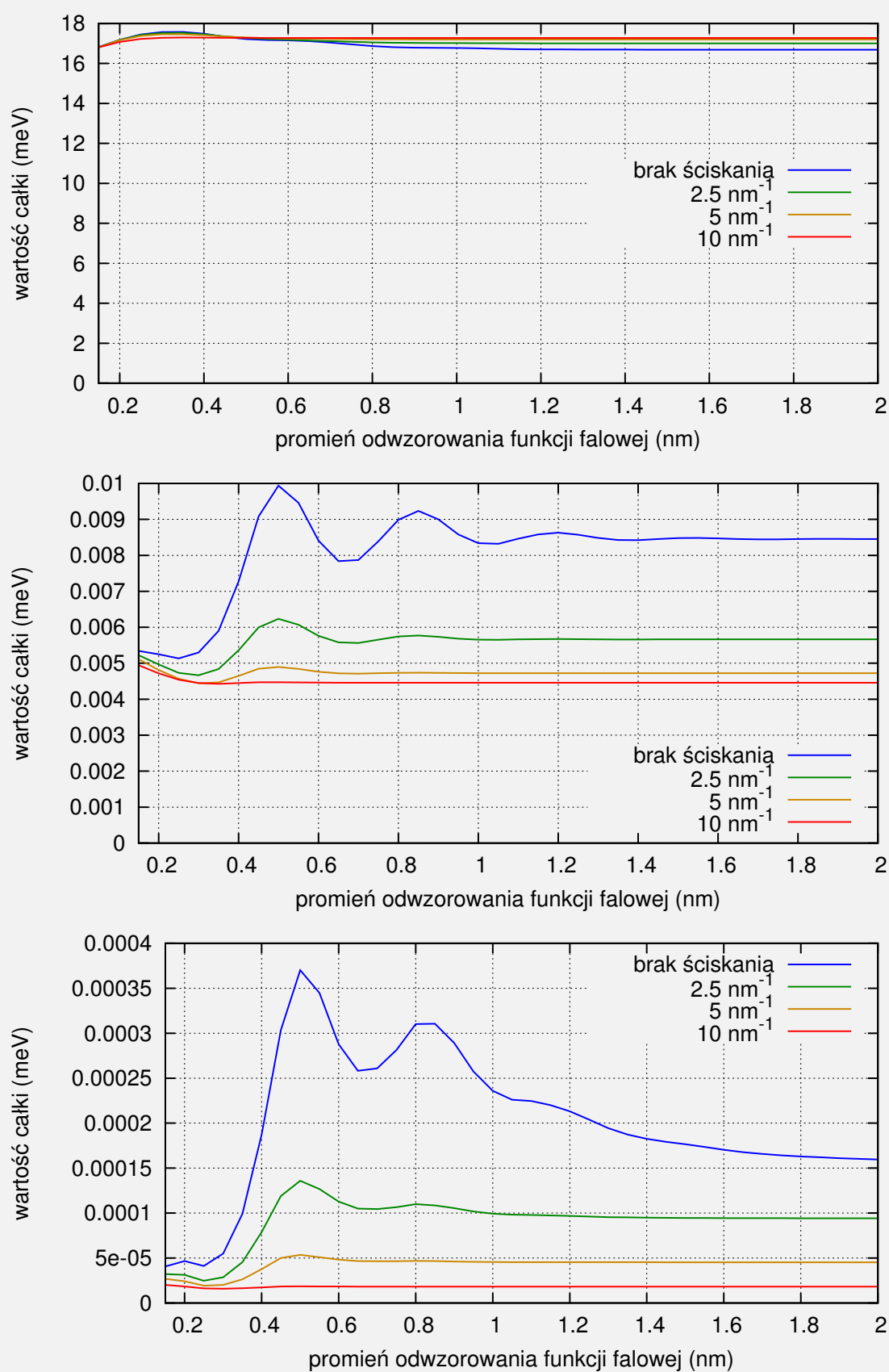
Zgodność otrzymanych trzema metodami wyników dla całek kulombowskich jest dość dobra, co oznacza, że dla tych całek założenia metody atomistycznej (str. 9) są spełnione w zadowalającym stopniu.

Wyraźne rozbieżności widoczne są natomiast w przypadku pozostałych dwu całek (całek wymiany). Jest to szczególnie istotne, ponieważ całki te opisują własności ekscytonowe (strukturę subtelną) badanego układu. Należy zauważyć, że różnice są mniejsze w przypadku orbitali H-S, charakteryzujących się mniejszą rozciągłością przestrzenną. Można na tej podstawie przypuszczać, że różnice wynikają właśnie z przestrzennej rozciągłości orbitali, nieuwzględnianej w metodzie atomistycznej; wymaga to jednak dalszych, szczegółowych badań.

3. Studium promienia odcięcia

W celu przeanalizowania wpływu rozciągłości orbitali bazowych na wyniki obliczeń, zbadano wpływ promienia odwzorowania funkcji falowej (promienia odcięcia) dla rekonstrukcji orbitalami H-S. Obliczenia przeprowadzono przy pomocy programu pqr-coulomb, obliczając wartości wybranych elementów macierzowych przy R_c zmieniającego się od 0 do 2 nm. Ponadto, obliczenia zostały przeprowadzone dla kilku różnych wartości współczynnika κ ściskającego orbitale bazowe.

²do obliczeń została użyta karta Tesla C2075



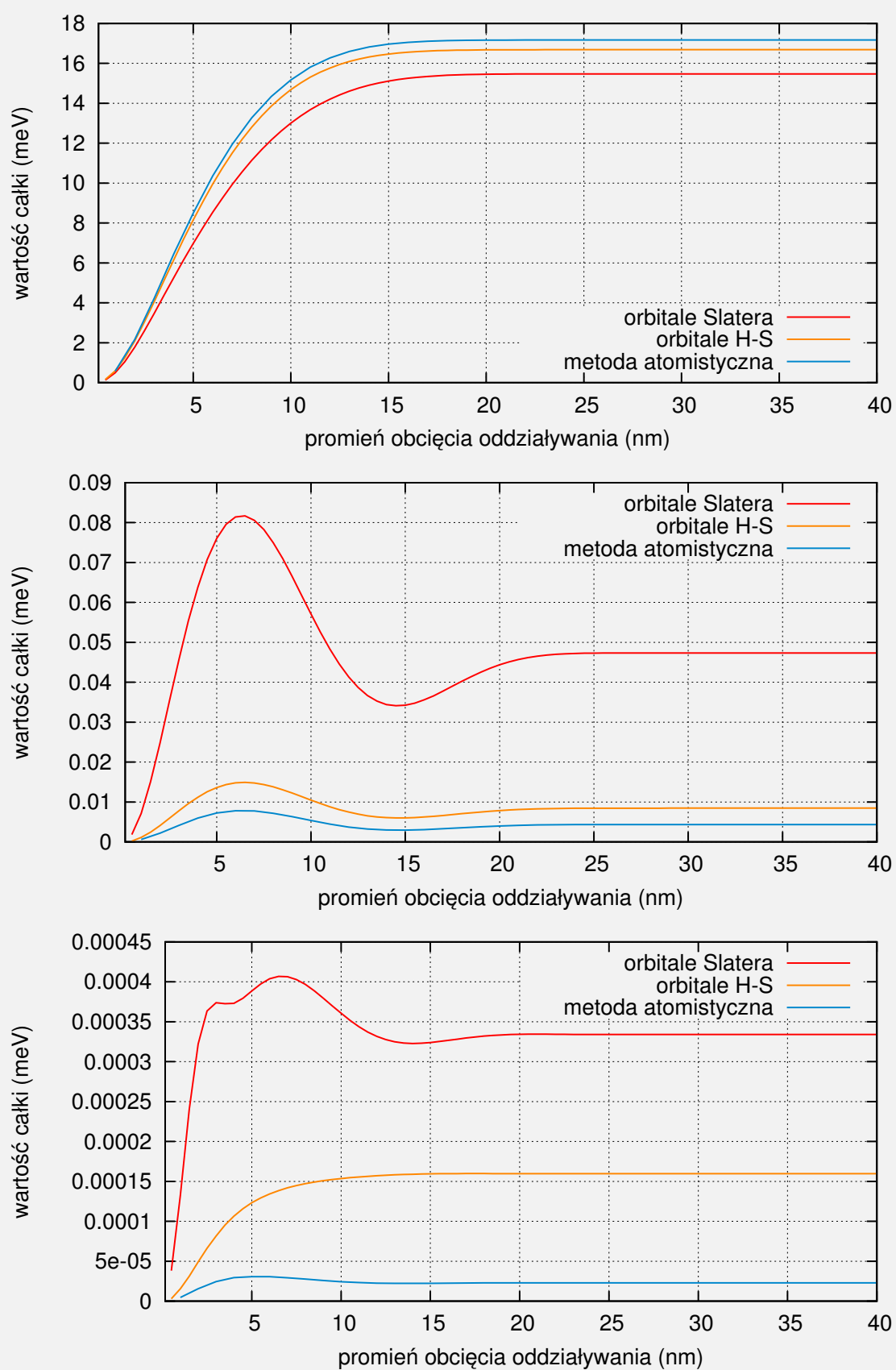
RYSUNEK 1. Wartości elementów macierzowych operatora Coulomba w zależności od promienia odcięcia R_c , dla różnych wartości współczynnika ściskającego κ . Od góry: $\langle e^1 h^1 h^1 e^1 \rangle$, $\langle e^1 h^2 e^2 h^1 \rangle$, $\langle e^1 h^1 e^2 h^2 \rangle$.

Wyniki obliczeń (wykres 1) pokazują, zgodnie z intuicją, że wraz ze wzrostem parametru κ coraz mniejsza wartość R_c wymagana jest do osiągnięcia zbieżności numerycznej. Widoczny jest również, na co warto zwrócić uwagę, silny wpływ współczynnika ściskającego na wartości całek wymiany, przy dość niewielkich zmianach całki kulombowskiej. Pozwala to stwierdzić, że całki wymiany zależą silnie od efektów związanych z nakładaniem się orbitali atomowych położonych na najbliższych (lub nawet dalszych) sąsiadach w sieci krystalicznej.

4. Studium zasięgu oddziaływania

Powyższe wyniki nie dają jednoznacznej odpowiedzi na pytanie, czy na wartości całek dominujący wpływ ma wkład długo- czy krótkozasięgowy. Można to jednak sprawdzić za pomocą ograniczania zasięgu oddziaływania (str. 25).

Przeprowadzono obliczenia, w których zasięg oddziaływania $r_{\max}$ zmieniał się w zakresie od 0 do 40 nm, przy czym badano zmiany wartości wybranych elementów macierzowych operatora kulombowskiego. Wyniki (wykres 2) pokazują wyraźną nie-monotoniczność w przypadku całek wymiany, zaś dla całek kulombowskich efektu tego nie obserwuje się. Co istotne, analogiczne zachowanie obserwuje się zarówno dla metody rekonstrukcji, jak i dla metody atomistycznej, nie może więc być ono związane z parametrami rekonstrukcji funkcji falowej (np. bazą). Można na tej podstawie postulować, że badane całki wymiany można rozdzielić na dwa przeciwstawne wkłady: dominujący wkład o zasięgu 10-20 nm oraz mniejszy co do wartości bezwzględnej wkład o przeciwnie skierowanej fazie i zasięgu 20-30 nm.



RYСУNEK 2. Wartości elementów macierzowych operatora Coulomba w zależności od zasięgu oddziaływania r_{max} . Od góry: $\langle e^1 h^1 h^1 e^1 \rangle$, $\langle e^1 h^2 e^2 h^1 \rangle$, $\langle e^1 h^1 e^2 h^2 \rangle$.

Podsumowanie

Nanostruktury półprzewodnikowe, w szczególności kropki kwantowe, są obiektem zainteresowania badaczy z uwagi na wyjątkowe własności elektronowe i optyczne. Postulowana możliwość wykorzystania kropek kwantowych do celów kwantowej kryptografii poprzez generowanie splątanych par fotonów (Singh and Bester, 2009), wymaga ilościowych badań efektu rozszczepienia struktury subtelnej (ang. *fine structure splitting*) w kropkach kwantowych.

Do badania rozszczepienia widmowych linii ekscytonowych z powodzeniem stosowane jest podejście atomistyczne (Zieliński et al., 2010) oparte o metodę ciasnego wiązania (TB, ang. *tight binding*). Obliczenia elementów macierzowych operatora kulombowskiego na podstawie wyników metody TB charakteryzują się złożonością obliczeniową $O(N^2)$ (w funkcji liczby atomów), co nie pozwala na efektywne badanie dużych ($N \sim 10^6$) struktur, szczególnie w przypadku wielkoskalowych obliczeń wielu całek.

W niniejszej pracy zaproponowano metodę obliczania elementów macierzowych operatora Coulomba opartą o rekonstrukcję funkcji falowej na podstawie wyników metody ciasnego wiązania. Przedstawiono algorytm o quasi-liniowej ($O(N \log N)$) złożoności czasowej, wykorzystujący twierdzenie o splocie i Szybką Transformatę Fouriera, umożliwiającą także uwzględnienie efektów ekranowania dielektrycznego. Dokonano studium możliwych optymalizacji przy obliczeniach wielu całek, w celu pełnego wykorzystania istniejących między nimi symetrii.

Przedstawiona została także efektywna implementacja opracowanej metody, umożliwiającą obliczenia równoległe na klastrach obliczeniowych (w tym, z użyciem infrastruktury PL-Grid) oraz przy użyciu kart graficznych CUDA. Zaprezentowano wyniki obliczeń na przykładzie cylindrycznej asymetrycznej kropki kwantowej InAs/InP układu zawierającego ponad pół miliona atomów.

Niniejsza praca magisterska została sfinansowana ze środków Fundacji na rzecz Nauki Polskiej w ramach grantu HOMING PLUS dr Michała Zielińskiego pt. *Control of exciton levels in quantum dots: atomistic theory as a step towards entangled photon pairs generation program*. W oparciu o wyniki pracy magisterskiej, w przygotowaniu są dwie publikacje naukowe w *Physical Review B* oraz *Journal of Applied Physics*.

Praca stanowić też będzie podstawę i inspirację do dalszych badań, w szczególności zaimplementowane metody i oprogramowanie będzie wykorzystane do badań nad strukturą subtelą linii ekscytonowych w półprzewodnikowych kropkach kwantowych.

Bibliografia

- Borgström, M. T., Zwiller, V., Müller, E., and Imamoglu, A. (2005). Optically bright quantum dots in single nanowires. *Nano Letters*, 5(7):1439–1443. PMID: 16178254.
- Edmonds, A. R. (1960). *Angular Momentum in Quantum Mechanics*. Princeton University Press.
- Fahlman, B. (2007). *Materials Chemistry*. Springer.
- Griffiths, D. (1999). *Introduction to electrodynamics*. Prentice Hall.
- Jacak, L., Hawrylak, P., and Wójs, A. (1998). *Quantum dots*. NanoScience and Technology Series. Springer.
- Klimov, V. (2009). *Nanocrystal Quantum Dots, Second Edition*. Taylor & Francis.
- Mura, M. E. and Handy, N. C. (1995). Cuboidal basis functions. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 90:145–165. 10.1007/BF01113845.
- Resta, R. (1977). Thomas-fermi dielectric screening in semiconductors. *Phys. Rev. B*, 16:2717–2722.
- Singh, R. and Bester, G. (2009). Nanowire Quantum Dots as an Ideal Source of Entangled Photon Pairs. *Physical Review Letters*, 103(6):063601.
- Slater, J. C. (1930). Atomic Shielding Constants. *Physical Review*, 36:57–64.
- Slater, J. C. and Koster, G. F. (1954). Simplified LCAO method for the periodic potential problem. *Phys. Rev.*, 94:1498–1524.
- Zieliński, M., Korkusiński, M., and Hawrylak, P. (2010). Atomistic tight-binding theory of multiexciton complexes in a self-assembled inas quantum dot. *Phys. Rev. B*, 81:085301.