

Kraków, 15.06.2022r.

**Recenzja rozprawy doktorskiej mgr. Tomasza Tarkowskiego zatytułowanej
„Przewidywanie struktury krystalicznej nanodrutów z użyciem obliczeń ewolucyjnych”**

Rozprawa doktorska Pana Tomasza Tarkowskiego pt. „Przewidywanie struktury krystalicznej nanodrutów z użyciem obliczeń ewolucyjnych” liczy sobie 33 strony i składa się z pięciu rozdziałów, wstępu oraz zakończenia. Do pracy dołączono 12 załączników elektronicznych w tym bibliografię. W samej pracy odniesienia literaturowe realizowane są w postaci przypisów na dole strony. Zestawienie wszystkich odniesień literaturowych liczące 114 pozycji znajduje się jedynie w załączniku elektronicznym. Nieliczne odniesienia zrealizowane w postaci przypisów mają formę niekompletną w porównaniu do wersji zawartej w załączniku elektronicznym (np. Holland str. 8, Baker, str. 10).

We wstępie autor przytacza historię powstania algorytmów genetycznych, opisuje poprzez kilka przykładów stan świata w końcu lat 80-tych XX wieku oraz wspomina o pierwszych pracach wykorzystujących algorytmy ewolucyjne do przewidywania struktury krystalicznej materiałów.

Rozdział pierwszy składa się z trzech stron i zawiera skrótową prezentację teorii funkcjonu gęstości. Autor wychodząc z założenia minimalizacji całkowitej energii układu sieci krystalicznej, opisuje stosowane przybliżenia i doprowadza do ostatecznej postaci równania Schrödingera, użytej w pakiecie obliczeniowym Quantum ESPRESSO wykorzystywanym przez autora do wyliczania funkcji przystosowania w algorytmie ewolucyjnym. Treść rozdziału jest poprawna, ale samo zagadnienie potraktowane zostało bardzo pobieżnie.

W rozdziale drugim, liczącym 8 stron, autor – w języku logiki formalnej i teorii mnogości - definiuje różne pojęcia wykorzystywane w algorytmach genetycznych (gen, populacja, funkcja przystosowania, operatory mutacji, funkcja selekcji w tym proporcjonalna i rankingowa, warunek zakończenia). Dodatkowo pod koniec rozdziału znalazł się opis mutacji samoadaptacyjnej. Treść rozdziału ma charakter zbioru definicji w związku z tym nie budzi zastrzeżeń, poza stwierdzeniem zawartym w ostatnim akapicie na stronie 5, że zdefiniowanie funkcji przystosowania dla fenotypu zamiast genotypu nie ma szczególnego znaczenia, z uwagi na konieczność zdefiniowania mapowania, która w ostateczności sprowadza ewolucję do przestrzeni genotypów. Jest to sformułowanie zbyt daleko idące. Po pierwsze ewolucja w przestrzeni fenotypów nie jest tożsama z ewolucją w przestrzeni genotypów. Oczywiście zawsze ewolucji podlegają geny, ale sam przebieg ewolucji może zależeć od tego w jakiej przestrzeni zdefiniowano funkcję przystosowania. Sytuacja taka będzie szczególnie dobrze widoczna, gdy przestrzeń fenotypu i przestrzeń genotypu będą się różnić metrykami. Po drugie stwierdzenie to pomija genotypy diploidalne, dla których w odróżnieniu od genotypów haploidalnych funkcja dostosowania zawsze musi być zdefiniowana w przestrzeni fenotypu.

Rozdział trzeci, liczący 7 stron, rozpoczyna się od opisu historii powstania języka C++, a następnie jego ewolucji do zbliżającego się standardu C++23. Później autor wspomina o czterech istniejących bibliotekach numerycznych wspomagających tworzenie algorytmów ewolucyjnych w tym algorytmów genetycznych, po czym decyduje się na stworzenie kolejnej biblioteki służącej temu celowi, którą nazywa Quile. Z tekstu pracy nie wynika dlaczego autor zdecydował się na stworzenie kolejnej biblioteki programistycznej, poza chęcią napisania jej w najnowszej implementacji języka C++. Dalej autor wymienia szereg narzędzi programistycznych wykorzystanych podczas pracy nad implementacją biblioteki, a następnie przechodzi do omówienia przykładu użycia biblioteki do znalezienia maksimum stosunkowo prostej funkcji dwuwymiarowej. W pracy zamieszczono pełen kod programu oraz szczegółowe instrukcje kompilacji tegoż przykładu w dwóch różnych kompilatorach języka C++. W dalszej części następuje zestawienie funkcji i operatorów zaimplementowanych w bibliotece, po czym autor przechodzi do kolejnego przykładu jej zastosowania, tym razem do znanego problemu szachujących się hetmanów. Recenzent nie zdecydował się na analizę implementacji biblioteki uznając brak swoich kompetencji w dziedzinie nowoczesnego programowania. Mimo to w zakresie posiadanych kompetencji, pozwala sobie na stwierdzenie, że zaimplementowanie bazy danych genotypów, do której programista odwołuje się przez obiekt pośredniczący jest rozwiązaniem oryginalnym. Natomiast wykorzystanie wzorca puli wątków choć nie jest oryginalne, pozwala znacznie uprościć pracę z populacją.

Rozdział czwarty, mający 7 stron, poświęcony jest zagadnieniu strojenia algorytmu genetycznego. Jest to standardowa procedura podczas pracy z algorytmami genetycznymi, gdzie korzystając ze znanych i sprawdzonych funkcji testowych poszukuje się ich minimum zmieniając parametry stworzonego algorytmu genetycznego. Autor arbitralnie wybrał 16 jednocelowych funkcji testowych. Do opisu efektywności oraz wydajności algorytmu wykorzystano popularne parametry SR (*ang. success rate*) oraz AUS (*ang. average number of unique individuals to get a solution*) wraz z odchyleniem standardowym tego parametru. Podczas strojenia algorytmu zmieniano operatory rekombinacji (arytmetyczna i arytmetyczna pojedyncza) i mutacji (Gausa i random-reset). Wyniki zostały zebrane w tabeli 4.2. Na podstawie przeprowadzonych testów autor wysnuwa trzy wnioski, z czego dwa dotyczące pozornie optymalnego zestawu parametrów optymalizacyjnych oraz tego, że skuteczność optymalizacji zwiększa się ze zmniejszaniem rozmiaru problemu optymalizacyjnego są oczywiste i znane w teorii algorytmów genetycznych niemal od początku ich rozwoju. Trzeci wniosek ma charakter praktyczny i dotyczy wyboru parametrów optymalizacyjnych do zastosowania w dalszej części pracy.

Ostatni, piąty rozdział pracy, na 6 stronach przedstawia zastosowanie opracowanego algorytmu genetycznego do wyznaczania struktury przestrzennej nanodrutów borowych w wersji płaskiej i trójwymiarowej. Autor założył periodyczność modelowanych struktur wzdłuż osi z oraz ustanowił ograniczenia na położenie atomów w taki sposób, aby nie mogły powstawać wiązania zbyt krótkie lub zbyt długie. Oddziaływanie atomów boru uwzględnia warunki periodyczności. Struktury zostały zakodowane w postaci pozycji atomów boru w komórce elementarnej. W przypadku 2D zastosowano układ kartezjański w przypadku 3D cylindryczny. Funkcja dostosowania wyliczana jest poprzez wykonywanie obliczeń DFT w zdyskretyzowanej (1x1x8) siatce k – punktów Monkhorsta-Packa przy czym liczbę iteracji w

procedurze samouzgodnionej ograniczono do 100. Przy użyciu rozkładu płaskiego dla wartości poszczególnych genów wygenerowano losowo osobniki pierwszego pokolenia. Warto przy tym zwrócić uwagę, że w przypadku 2D procedura ta zapewnia równomierny rozkład gęstości centrów atomów na płaszczyźnie, natomiast w przypadku 3D, z uwagi na zastosowanie układu cylindrycznego, rozkład gęstości przestrzennej punktów będących centrami położenia atomów nie będzie równomierny. Gęstość będzie rosła z malejącym promieniem. Z uwagi na małą liczbę atomów użytych w symulacji efekt ten nie będzie zauważalny. Nie wiadomo jednak, czy jest to efekt niedopatrzenia czy świadomej decyzji. Korzystając z parametrów ustalonych w rozdziale czwartym autor uruchomił algorytm genetyczny pięciokrotnie dla liczby atomów od 1 do 8 dla przypadków 2D i 3D. Najlepsze rozwiązania dla każdego przypadku zostały następnie poddane obliczeniom samouzgodnionym DFT jednak na dokładniejszej siatce k – punktów ($1 \times 1 \times 16$). Limitem iteracji była w tym przypadku liczba 200, jednak jak napisano w pracy nigdy limit ten nie został osiągnięty. Autor stwierdził, że w modelowanych warunkach bor ma tendencję do tworzenia motywów trójkątnych, kwadratowych lub posiadających dziury. Ani w pracy, ani w załączniku elektronicznym nie podano jednak energii wiązania na atom dla struktur znalezionych przez algorytmy genetyczne. Podano jedynie odpowiednią energię dla struktur zrelaksowanych metodą DFT. Trudno zatem ocenić jakość dopasowania genetycznego. Nawet ocena wizualna znalezionych struktur jest mocno utrudniona, ponieważ jak sam autor konkluduje w ostatnim akapicie rozdziału zastosowanie różnej siatki dyskretyzacji (8 vs. 16) w algorytmie genetycznym i w późniejszej relaksacji w zasadzie ten proces uniemożliwia.

W zakończeniu autor porównuje stworzony algorytm z genotypami organizmów żywych, podsumowując to porównanie stwierdzeniem, że „Zaprezentowane procesy ewolucji struktur krystalicznych są właściwie niczym w porównaniu z ewolucją naturalną...”. Warto jednak przy tym pamiętać, że ewolucja naturalna miała również nieporównanie więcej czasu i znacznie większą liczbę osobników do dyspozycji niż autor w ramach swojej pracy. Na zakończenie autor zamieszcza sześć propozycji rozwinięcia lub udoskonalenia zaproponowanego algorytmu.

Przedstawiona do recenzji praca ma charakter informatyczno-programistyczny. Głównym osiągnięciem pracy jest implementacja biblioteki programistycznej wspomagającej programowanie algorytmów genetycznych. O informatycznym charakterze pracy świadczy jej struktura: 22 strony pracy poświęcone są zagadnieniom informatycznym włącznie ze szczegółami implementacji i kompilacji, 3 strony opisowi zagadnienia fizycznego i 6 stron prezentujących przykład aplikacji opracowanej biblioteki programistycznej do zagadnienia fizycznego. Tezę o informatycznym charakterze pracy wspiera także analiza cytowań. Ponad 80 cytowań odnosi się do publikacji z zakresu informatyki (w tym metody numeryczne i algorytmy genetyczne) a tylko 26 cytowań odwołuje się do publikacji z zakresu fizyki. Recenzentowi trudno ocenić oryginalność implementacji biblioteki programistycznej, niemniej analiza pracy pozwala stwierdzić, że niezależnie od jej oryginalności jej efekty lokują się w zakresie dyscypliny informatyka techniczna i telekomunikacja w dziedzinie nauk inżyniersko-technicznych.

Niezrozumiała jest decyzja o podjęciu się pisania nowej implementacji biblioteki programistycznej. Jeżeli założonym celem pracy miało być „Przewidywanie struktury krystalicznej nanodrutów z użyciem obliczeń ewolucyjnych”, to rodzi się pytanie dlaczego autor nie wykorzystał jednej z istniejących bibliotek i nie skoncentrował swoich wysiłków na aspektach fizycznych zagadnienia? Jeśli istniały jakieś przesłanki – choć recenzent ich nie dostrzega – uniemożliwiające zastosowanie istniejących bibliotek do wspomnianego w tytule pracy problemu, to autor powinien to wyjaśnić. Jeżeli jednak jedyną przesłanką była chęć stworzenia biblioteki w najnowszej implementacji języka C++ o czym autor wspomina w rozdziale trzecim, to ponownie osiągnięcie to lokuje się w dyscyplinie informatyka techniczna i telekomunikacja.

W końcówce rozdziału czwartego autor konstatuje: „nie można stwierdzić, że algorytm genetyczny jest na chwilę obecną najlepszym sposobem na rozwiązanie problemu przewidywania struktury krystalicznej.” Po czy w pierwszym zdaniu rozdziału piątego pisze: „Użyty będzie jednak, zgodnie z tytułem pracy, algorytm genetyczny”. Również ta decyzja pozostaje niezrozumiała. Jeżeli celem pracy było znalezienie struktur nanodrutów, to należało dostosować używane metody do celu pracy, nawet kosztem zmiany tytułu, co jest praktykowane. Jeśli celem było zastosowanie algorytmów genetycznych, to ponownie jest to zagadnienie z zakresu informatyki. Nawet jednak w tym przypadku, autor nie musiał korzystać z algorytmu genetycznego, jeśli uznał jego niską skuteczność. W tytule pracy jest mowa o znacznie szerszej klasie „obliczeń ewolucyjnych”, dla których algorytmy genetyczne są tylko jednym z przykładów.

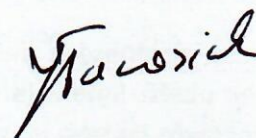
Zastrzeżenia budzi także warsztat badawczy autora. Zastosowanie algorytmów genetycznych w obliczeniach DFT jest obecnie niezwykle popularne. Po wpisaniu w wyszukiwarkę artykułów tylko na samym Science Direct słów kluczowych *genetic algorithms* i *DFT* otrzymujemy ponad 3000 artykułów za ostatnie 10 lat. W pracy nie ma jednak przeglądu literatury w tym zakresie. Liczba wszystkich cytowań z zakresu fizyki jak wspomniano wyżej to 26 pozycji, z czego za ostatnie 10 lat jest to 5 pozycji, z czego jedno to autocytowanie, a jedno to cytaty pracy promotora doktoratu. Świadczy to o braku zorientowania autora w aktualnych trendach w wykorzystaniu algorytmów genetycznych w obliczeniach DFT. A szkoda, bo analiza stosowanych obecnie metod kodowania, selekcji i operatorów mutacji mogłaby nakierować autora na inną ścieżkę rozwoju swojej aplikacji oraz oszczędziłaby czasu podczas testowania i strojenia algorytmu. Również sam proces strojenia świadczy o podejściu bardziej informatycznym niż typowym dla osób prowadzących badania i modelowanie w zakresie fizyki. Zastosowanie standardowych funkcji testowych jest typową procedurą informatyczną w testowaniu algorytmów genetycznych. Niewiele jednak wnosi w kontekście problemu postawionego w tytule pracy. Nie wiemy czy i w jakim stopniu, którakolwiek z funkcji testowych ma podobny, zbliżony charakter do krajobrazu rozwiązań dla problemu minimalizacji energii struktur krystalicznych. A przecież istnieją naturalne funkcje testowe dokładnie dla tego problemu. Znamy bardzo dokładnie struktury tysięcy związków, bądź to z obliczeń teoretycznych dla prostych struktur bądź z obliczeń ab initio dla bardziej skomplikowanych, w tym wielu monoatomowych. Wystarczyło użyć tych struktur jako funkcji testowych dla stworzonego algorytmu i dla nich przeprowadzić strojenie. Byłoby to podejście uzasadnione od strony fizycznej, nie tylko informatycznej. W kwestii warsztatu należy również

wspomnieć o ustawieniu różnej rozdzielczości siatek k – punktów podczas obliczania wartości funkcji dostosowania oraz w obliczeniach relaksacyjnych, co skutecznie uniemożliwiło porównanie wyników otrzymanych z algorytmu genetycznego oraz z relaksacji DFT. Z punktu widzenia tytułu pracy to porównanie jest bardzo ważne ponieważ pozwalałoby przynajmniej częściowo ocenić użyteczność zastosowania algorytmów genetycznych w badanym problemie.

Podsumowując, przedstawiona do oceny rozprawa doktorska:

1. prezentuje osiągnięcie w postaci stworzenia biblioteki programistycznej wspierającej programowanie algorytmów genetycznych, które nie mieści się w zakresie dyscypliny nauki fizyczne;
2. nie dowodzi wiedzy autora w zakresie współczesnych metod i osiągnięć wykorzystania algorytmów genetycznych w obliczeniach DFT;
3. z uwagi na wymienione mankamenty w warsztacie badawczym autora, nie dowodzi umiejętności samodzielnego prowadzenia pracy naukowej;

i jako taka, nie może być podstawą nadania stopnia doktora w dyscyplinie nauki fizyczne.



dr hab. inż, prof. AGH

Jacek Tarasiuk

Katedra Fizyki Materii Skondensowanej
Wydział Fizyki i Informatyki Stosowanej