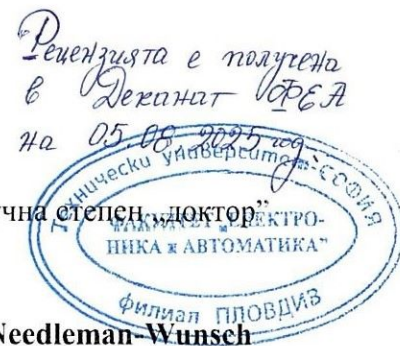


Процедура № ФБА56-НС1-028



РЕЦЕНЗИЯ

върху дисертационен труд за придобиване на образователна и научна степен „доктор“

Автор на дисертационния труд: **маг. Емилия Хаим Пардо**

Тема на дисертационния труд: **“Разпаралеляване на алгоритъма Needleman-Wunsch върху графични ускорители“**

Рецензент: **проф. д-р Димитър Михайлов Токмаков**

1. Актуалност на разработвания в дисертационния труд проблем в научно и научноприложно отношение. Степен и нива на актуалността на проблема и конкретните задачи, разработени в дисертацията.

Дисертацията разглежда една от най-ключовите теми в съвременната биоинформатика – ускоряване на алгоритми за подравняване на ДНК/РНК последователности чрез GPU изчисления с CUDA.

Алгоритъмът Needleman-Wunsch е класически метод за глобално подравняване, но е изчислително тежък. Поради експоненциалния ръст на геномни данни, оптимизацията му за паралелна обработка е стратегически важна за развитието на персонализираната медицина, фармакогеномиката и молекулярната биология.

Използването на CUDA и графични ускорители (GPU) е в крак с глобалните тенденции в суперкомпютрите и хетерогенното програмиране в науката и индустрията.

Проблемът има висока степен на актуалност както в научен, така и в приложен контекст. Подходът за решаването му е в пълно съответствие с глобалните научни тенденции в биоинформатиката, изкуствения интелект и високопроизводителните изчисления.

2. Степен на познаване състоянието на проблема и творческа интерпретация на литературния материал.

Дисертацията демонстрира много добро и широко обхватно запознаване със съществуващото състояние на научния проблем, което личи от:

- Подробен литературен обзор (Глава 1) на 91 литературни източника с включени ключови теми от:



- структура на ДНК и РНК;
- алгоритми за подравняване (Needleman-Wunsch, Smith-Waterman, Dot Matrix и др.);
- биологични бази данни и техники за търсене;
- класификация на подравняващи алгоритми и свристични методи (BLAST, FASTA и др.);

Творческият подход се изразява в междусекторно обединяване на теоретични подходи като сравнение между биологични и компютърни процеси (ДНК ↔ софтуерен код, трансляция ↔ компилация);

Извършено е аналитично сравнение на подравняващи алгоритми и разглеждане на тяхната ефективност в контекста на реални биологични данни.

Обоснован е изборът на алгоритъма Needleman-Wunsch за паралелна имплементация на GPU – с ясно разграничение между глобално и локално подравняване.

Авторът показва висока осведоменост по ключовите концепции, алгоритми и системи, свързани с темата.

3. Съответствие на избраната методика на изследване с поставената цел и задачи на дисертационния труд.

Целта на дисертационния труд е да се разгледа и анализира възможността за оптимизиране на алгоритъма Needleman-Wunsch чрез неговото изпълнение върху графичен ускорител (GPU) от типа NVIDIA, използвайки програмния модел CUDA C.

Поставените задачи в дисертацията са ясно формулирани, конкретни, технически измерими и съответстват на общата цел на изследването. Формулирани са на правилното ниво на абстракция.

Дисертационният труд демонстрира висока степен на съответствие между формулираната цел, поставените задачи и избраната методология на изследване.

Целта за оптимизация на алгоритъма Needleman-Wunsch чрез паралелна имплементация с CUDA е конкретна, приложима и научно значима, а методите, използвани за нейното постигане, са адекватно подбрани и коректно реализирани. Изследването стъпва на стабилна теоретична основа, комбинира алгоритмично моделиране, GPU-програмиране,

разработка на приложен софтуер и експериментална проверка. Добавянето на компонент с изкуствен интелект (LLaMA 2) подсилва иновативността на методологията и показва междудисциплинарен подход, който обогатява класическите биоинформатични методи.



4. Кратка аналитична характеристика на естеството и оценка на достоверността на материала, върху който се градят приносите на дисертационния труд.

Дисертационният труд е с обем от 141 страници и съдържа 4 глави, множество фигури и таблици.

Първа глава е посветена на литературния обзор. Втора глава разглежда CUDA архитектурата и паралелното програмиране. Трета глава описва софтуерния инструмент BioPoolSelect. В четвърта глава е разгледан алгоритъма Needleman-Wunsch и техники за неговото разпаралеляване и оптимизация.

Дисертационният труд стъпва на два основни типа материал:

- Теоретичен материал:

-91 литературни източника, включващи рецензирани научни публикации, авторитетни учебници и техническа документация от NVIDIA;

-Систематизиран обзор на алгоритми за подравняване на последователности (Needleman-Wunsch, Smith-Waterman, BLAST и др.);

-Детайлно описание на архитектурата CUDA и принципите на паралелно програмиране.

- Емпиричен материал:

-Авторска реализация на алгоритъма Needleman-Wunsch с CUDA;

-Разработен собствен софтуерен инструмент BioPoolSelect;

-Извършени измервания на производителност с различни входни данни и хардуерни конфигурации;

-Интегриране на AI модел (LLaMA 2) за тестване на нов хибриден подход.

Материалите, върху които се базират приносите в дисертационния труд, са научно обосновани, актуални и проверими.



5. Научни и/или научноприложни приноси на дисертационния труд

В автореферата и дисертационния труд са формулирани три научно-приложни и два приложни приноса.

Научни приноси:

- Създадена е оптимизация на паралелна версия на алгоритъма на Needleman-Wunsch за графични процесори (GPU).
- Въведен е индекс на изравняване на последователности.
- Разработено е ръководство по дисциплината “Програмиране на съвременни хетерогенни архитектури”.

Приложни приноси:

- Разработен е софтуерен инструмент за геномен анализ (BioPoolSelect). Създаден е нов софтуер, който позволява селектиране на специфични геномни характеристики за анализ, като намалява размера на първичните файлове и улеснява обработката на данните.
- Представен е сравнителен анализ на геноми.

Приемам приносите и считам че те са коректно дефинирани.

6. Оценка за степента на личното участие на дисертанта в приносите.

Анализът на съдържанието на дисертацията показва, че дисертантът има висока степен на лично участие, особено в следните ключови направления:

-Създаване на CUDA-базирана паралелна реализация на алгоритъма Needleman-Wunsch.

-Разработка на софтуерен инструмент BioPoolSelect – функционалността и интерфейсът на инструмента са подробно описани и резултатите от използването му са обвързани с целите на изследването.

-Анализ на производителност, параметри и конфигурации на тестовите, сравнителните таблици и интерпретацията на резултатите сочат, че дисертантът е извършил самостоятелно експерименталната част.



Не познавам лично докторанта маг. Емилия Пардо, но считам че нейния личен принос е доказано безспорен.

7. Преценка на публикациите по дисертационния труд: брой, характер на изданията, в които са отпечатани. Отражение в науката – използване и цитиране от други автори, в други лаборатории, страни и пр.

Списъкът на публикациите включва четири научни публикации на международни конференции, като три от тях са индексирани в Scopus и едно публикувано университетско учебно пособие.

Публикациите на докторанта отразяват основните резултати и приноси получени при изготвяне на дисертационния труд. Количеството и качеството на публикациите отговарят на изискванията за получаване на ОНС „доктор.

8. Оценка на съответствието на автореферата с изискванията за изготвянето му, както и на адекватността на отразяване на основните положения и приносите на дисертационния труд.

Авторефератът е изготвен според изискванията на съответните правилници, и отразява основните резултати, постигнати в дисертацията.

9. Мнения, препоръки и бележки.

Дисертационният труд демонстрира висока научна и приложна стойност с ясно изразено лично участие на дисертанта в алгоритмичната реализация и софтуерната разработка. Методологията е адекватна, резултатите са експериментално валидирани, а приносите – оригинални и практически приложими.

Препоръчва се разширяване на сравнителния анализ с други паралелни решения и по-ясно разграничаване на ограниченията при използването на AI.

Препоръчвам също така в бъдещо развитие на темата, да се състави формален модел на сложност, който да дава теоретичен анализ на алгоритъма по отношение на изчислителните ресурси, които той изисква – най-често време (броя на стъпки) и памет (пространство), изразени като функция на входните данни.

10. Заключение с ясна положителна или отрицателна оценка на дисертационния труд.

Дисертационният труд на маг. Емилия Пардо съдържа научно-приложни и приложни резултати, които представляват оригинален принос в науката и напълно отговаря на изискванията на Закона за Развитие на Академичния Състав на Република България, и на съответния правилник на ТУ-София. Приложеният към него автореферат отразява същността на изследването и коректно представя приносите от него.

Като се имат предвид достойнствата, актуалността, значимостта на представеният дисертационен труд, му давам **положителна** оценка и предлагам на научното жури да **присъди на Емилия Пардо образователната и научна степен „ДОКТОР“** в научна област 5. Технически науки, Професионално направление: 5.3. Комуникационна и компютърна техника, специалност: “Компютърни системи, комплекси и мрежи“



Дата: 04.08.2025г.

РЕЦЕНЗЕНТ:

/проф. д-р инж. Димитър Токмаков/

REVIEW

of a dissertation for the award of the educational and scientific degree "Doctor"



Author of the dissertation: MSc Emilia Haim Pardo

Title of the dissertation: *"Parallelization of the Needleman-Wunsch Algorithm on Graphics Processing Units"*

Member of the scientific jury: Prof. Dr. Dimitar Mihaylov Tokmakov

1. Relevance of the problem developed in the dissertation from a scientific and applied science perspective. Degree and levels of relevance of the problem and specific tasks addressed in the dissertation.

The dissertation addresses one of the key topics in modern bioinformatics—acceleration of DNA/RNA sequence alignment algorithms using GPU computation with CUDA. The Needleman-Wunsch algorithm is a classical method for global alignment but is computationally intensive. Due to the exponential growth of genomic data, optimizing this algorithm for parallel processing is strategically important for the advancement of personalized medicine, pharmacogenomics, and molecular biology. The use of CUDA and GPUs aligns with global trends in supercomputing and heterogeneous programming in both science and industry.

The problem is highly relevant both scientifically and in terms of practical application. The approach used to solve it fully corresponds with global scientific trends in bioinformatics, artificial intelligence, and high-performance computing.

2. Degree of knowledge of the state of the problem and creative interpretation of the literary material.

The dissertation demonstrates a very good and comprehensive familiarity with the current state of the scientific problem, evident through:

- A detailed literature review (Chapter 1) referencing 91 sources and covering key topics such as:
 - DNA and RNA structure
 - Alignment algorithms (Needleman-Wunsch, Smith-Waterman, Dot Matrix, etc.)
 - Biological databases and search techniques



- Classification of alignment algorithms and heuristic methods (BLAST, FASTA, etc.)

The creative approach is evident in the cross-sector integration of theoretical concepts, such as analogies between biological and computer processes (DNA ↔ software code, translation ↔ compilation).

An analytical comparison of alignment algorithms is conducted, evaluating their effectiveness in the context of real biological data.

The choice of the Needleman-Wunsch algorithm for GPU-based parallel implementation is well-justified, with a clear distinction between global and local alignment. The author demonstrates deep awareness of key concepts, algorithms, and systems relevant to the topic.

3. Alignment of the chosen research methodology with the stated goal and objectives of the dissertation.

The aim of the dissertation is to analyze the possibilities of optimizing the Needleman-Wunsch algorithm by executing it on a graphics processing unit (GPU) of the NVIDIA type, using the CUDA C programming model.

The tasks are clearly formulated, specific, technically measurable, and aligned with the overall research objective. They are defined at an appropriate level of abstraction. The dissertation shows a high degree of consistency between the stated aim, objectives, and chosen research methodology.

The goal of optimizing the Needleman-Wunsch algorithm through parallel CUDA implementation is specific, applicable, and scientifically significant. The methods used to achieve it are appropriately selected and correctly implemented. The research builds on a solid theoretical foundation and combines algorithmic modeling, GPU programming, application software development, and experimental validation.

The addition of an artificial intelligence component (LLaMA 2) enhances the innovativeness of the methodology and demonstrates an interdisciplinary approach that enriches classical bioinformatics methods.

4. Brief analytical characterization and evaluation of the reliability of the material upon which the dissertation's contributions are based.



The dissertation consists of 141 pages and includes four chapters, numerous figures, and tables:

Chapter 1: Literature review;

Chapter 2: CUDA architecture and parallel programming;

Chapter 3: Description of the software tool BioPoolSelect;

Chapter 4: Needleman-Wunsch algorithm and techniques for its parallelization and optimization.

The dissertation is based on two main types of material:

-Theoretical material:

- 91 literature sources including peer-reviewed scientific publications, authoritative textbooks, and technical documentation from NVIDIA
- A systematized review of sequence alignment algorithms (Needleman-Wunsch, Smith-Waterman, BLAST, etc.)
- A detailed explanation of CUDA architecture and principles of parallel programming

-Empirical material:

- Author's implementation of the Needleman-Wunsch algorithm using CUDA
- Developed custom software tool – BioPoolSelect
- Performance measurements with various input data and hardware configurations
- Integration of the LLaMA 2 AI model for testing a new hybrid approach

The materials upon which the dissertation's contributions are based are scientifically sound, up-to-date, and verifiable.

5. Scientific and/or applied contributions of the dissertation.

Three scientific-applied and two applied contributions are stated in the dissertation and its abstract.

Scientific contributions:

- An optimized parallel version of the Needleman-Wunsch algorithm for GPUs has been created
- A sequence alignment index has been introduced
- A manual has been developed for the course “Programming of Modern Heterogeneous Architectures”

Applied contributions:

- A genomic analysis software tool (BioPoolSelect) has been developed. This software enables the selection of specific genomic features for analysis, reducing file size and easing data processing
 - A comparative genome analysis has been presented
- I accept the contributions as correctly defined.

6. Assessment of the degree of personal involvement of the doctoral candidate in the contributions.

The content analysis of the dissertation indicates a high degree of personal involvement by the doctoral candidate, especially in the following key areas:

- CUDA-based parallel implementation of the Needleman-Wunsch algorithm
- Development of the BioPoolSelect software tool—its functionality and interface are thoroughly described, and the results are clearly linked to the research objectives
- Performance analysis, parameter testing, and system configurations—the tests, comparative tables, and result interpretations show that the experimental work was conducted independently by the candidate

I do not personally know the PhD candidate Emilia Pardo, but I believe her individual contribution is clearly and indisputably demonstrated.

7. Evaluation of the publications related to the dissertation: number, type of publications, and impact in science—use and citations by other authors, laboratories, countries, etc.

The list of publications includes four scientific papers presented at international conferences, three of which are indexed in Scopus, and one university teaching material. The publications reflect the core results and contributions achieved in the dissertation. Both the



quantity and quality of the publications meet the requirements for the award of the educational and scientific degree “Doctor.”



8. Assessment of the compliance of the abstract with the formal requirements and adequacy in reflecting the dissertation’s main statements and contributions.

The abstract is prepared according to the applicable regulations and adequately reflects the main results achieved in the dissertation.

9. Comments, recommendations, and notes.

The dissertation demonstrates high scientific and practical value, with clearly defined personal contributions in algorithm development and software implementation. The methodology is sound, the results are experimentally validated, and the contributions are original and applicable. It is recommended to expand the comparative analysis with other parallel solutions and to more clearly outline the limitations of AI use. Additionally, it is advisable to construct a formal model of algorithm complexity in future work, providing theoretical analysis of computational resource requirements—typically time (steps) and memory (space), expressed as functions of input data.

10. Conclusion with a clear positive or negative assessment of the dissertation.

The dissertation by MSc Emilia Pardo presents scientific and applied results that constitute an original contribution to science and fully meet the requirements of the Law on the Development of the Academic Staff of the Republic of Bulgaria and the regulations of the Technical University of Sofia. The accompanying abstract reflects the essence of the research and correctly presents its contributions.

Considering the merits, relevance, and significance of the dissertation, I give it a **positive** assessment and recommend that the scientific jury **award Emilia Pardo the educational and scientific degree “DOCTOR”** in scientific field 5. Technical Sciences, Professional Direction 5.3. Communication and Computer Engineering, Specialty: “Computer Systems, Complexes and Networks.”

Date: August 4, 2025

REVIEWER:.....
/ Prof. Dr. Eng. Dimitar Tokmakov/