R E V I E W

on the procedure for defense for a dissertation work on the topic:
Porous Medium Flow Simulations using Massively Parallel MLMC algorithm
for obtaining educational and scientific degree Doctor

candidate: Nikolay Georgiev Shegunov,
Areas of higher education: 4. Nature of science, mathematics and computer science
Professional direction: 4.6. Informatics and Computer Science
Doctoral program: Information systems, cathedra: Computer informatics,
Sofia University “St. Kliment Ohridski “, Faculty of Mathematics and Computer Science.

This review was prepared by: Prof. Dr. Vasil Georgiev Tsunizhev, FMI - Sofia University "St. Cl. Ohridski" as a member of the scientific jury, according to Order № RD-38-258 / 7. 06. 2021 of the Rector of Sofia University.

1. General characteristics of the dissertation and the presented materials

The dissertation presented for review with the above title and with the author's translation of the title "Simulation of flows in porous media by massively parallel Monte Carlo multilevel algorithm" is a monograph in English, dedicated to solving a scientifically applied problem in the field of simulation modeling. of interaction of materials in liquid and solid phase. Regardless of the field of application, however, the author essentially set himself the task of optimizing the process of machine numerical processing of such a class of simulation models by numerically solving the corresponding system of equilibrium equations by the method of random processes known as Monte Carlo method. The two main features of the chosen approach are

(1) applicability of the "multilevel" Monte Carlo method for modeling the problem area related to the convection-reaction-diffusion equation and also

(2) applicability and results of the parallel numerical processing of the model.

In view of the results shown, it is a matter of machine and hence software parallelism, which can actually be described as massive, since in some of the reported experiments the scalability was measured at values of parallelism 9600.

In essence, the contribution of the dissertation is precisely a model for highly parallel numerical processing of the "multilevel" Monte Carlo algorithm. Of course, in the course of deployment and experimental study of this model, a number of concomitant issues have been resolved, including those related to the initial field of application of the simulation numerical model, namely the porosity of solid media.

The main paper is presented on 129 pages, containing 5 chapters and a short conclusion, which includes a list of contributions and a list of five publications on the dissertation. This text also includes several appendices on the abbreviations used, a list of figures and tables, a brief description of the hardware infrastructure used (2 different shared clusters), and a list of sources with 62 titles.

To this main text is added an Abstract in Bulgarian of 32 pages and additions to them, presenting mainly the list of sources, the list of author's publications and the list of contributions.

In general, the main work and the accompanying documents create the impression of competence, deep knowledge of the researched issues and free entry into interdisciplinary areas. The description is precise and concise, with properly selected sub-topics for discussion in the individual chapters. Some weaknesses and opportunities for improvement can be identified, which I will discuss in item 6.

2. Data and personal impressions of the candidate

I know the dissertation Nikolay Shegunov on the occasion of his research and teaching work at FMI. Its preparation and activity are carried out under the supervision of colleagues from the Department of Computer Informatics, of which they are members together with their supervisor Associate Professor Armyanov. What characterizes and is known in his professional

community colleague Nikolai is the in-depth work, high knowledge in the problem areas of interest, his successful and sought-after participation in the educational process, his integration into international and multidisciplinary research teams and his loyalty. not to be detached from the main research and educational unit.

3. Content analysis of the scientific and scientific-applied achievements of the candidate, contained in the presented dissertation and the publications to it, included in the procedure

The scientific and scientific-applied contributions of the dissertation are correctly reflected in the attached documents of the procedure.

The dissertation begins with an overview and comparative analysis of the existing solutions to the problems. The advantages and disadvantages of the existing solutions for generating stochastic fields and sampling algorithms are evaluated. Part of Chapter 1. and almost the entire Chapter 2. of the dissertation are devoted to this review. At the end of Section 1.3, the different approaches used to simulate flows in porous media are discussed. The advantages and disadvantages for each of them are shown. Section 1.4 discusses technical tools and software libraries that can be used to solve the problems posed. In point 2.1 the ways of generating stochastic fields for the purposes of the set task are considered and compared. Section 2.2 discusses the finite volume method used to model the problem. Section 2.4 describes in detail the model used in the dissertation based on the "multilevel" Monte Carlo method. In Chapter 3, different approaches to the stochastic level approximation for the Laplace problem are compared and analyzed.

Further in Chapter 3. the approach for renormalization of the stochastic field for the purposes of the "multilevel" Monte Carlo method scalar elliptic stochastic partial differential equations is explained and substantiated. This is evident from the comparison of the "multilevel" Monte Carlo method using renormalization over the classical one, which show the advantage of the considered approach are placed in the dissertation of figures: 3.6, 3.7 and 3.8 (pages 41-43). They show, respectively, the number of experiments performed at different levels and the reduction of calculation time (Figure 3.8). Tables 3.2 and 3.3 show the advantages of the simplified renormalization applied by the author over the arithmetic mean.

Chapter 4 describes the application of the "multilevel" Monte Carlo method to the convection-reaction-diffusion equation. An adaptive algorithm has been developed for allocating resources to the different levels of the "multilevel" Monte Carlo algorithm. Results with different simulation parameters are applied - different numbers of Hell and Damkoler. The results are presented in the dissertation in Table 4.1 and Table 4.2.

In terms of approbation and application of the examinations made in Chapter 5. there are also good results. A review, analysis and comparison of 6 strategies for parallelization is made as a dynamic strategy for parallelization between different levels is considered and for the proposed strategy are analyzed three approaches for parallelization of the two layers - local interruption, global interruption and "tail" parallelism.

The described model of the parallel processing of the numerical algorithm is investigated experimentally through program realization and processing in the conditions of systems with massive parallelism. The reported acceleration test results (normalized by the magnitude efficiency) show completely satisfactory results in the conditions of parallelism of the order of several thousand. The use of the global interrupt optimization proposed by the author leads to an efficiency improvement in 4800 processor cores by 15%, in 7200 cores by 12% and in 9600 cores by 7%. Of course, we cannot expect infinite scalability from any parallel algorithm. Suffice it to say that for the specific scope of parallelism, we see that optimization "works". The author explains the decrease in relative efficiency with increasing parallelism - and with good reason - by two factors: fixed computational complexity of the task with increasing processing capacity and increasing communication delays in the larger server cluster.

4. Approbation of the results

The five publications on the dissertation reported by the candidate, as well as the dissertation itself and the prepared abstract, are characterized by the following features:

(1) fully, accurately and correctly reflect the presented dissertation;

(2) all dissertation publications are indexed in Scopus and therefore it can be calculated that the requirements for publication activity are met almost twice: more precisely, for

the general index of the author RNSh by group of indicators D of the LRA, the following scientometric calculations can be made (according to the [indices] of the list of publications): R_{NSh} = 1 / 3x30 t. [1] + 0.8 x30 t. [2] + 1 / 3x30 t. [3] + 1 / 2x30 t. [5] = 59 >> 30. Publication [4] is neglected due to the large number of co-authors. The citations in Scopus have not been established so far, if I do not count 3 self-citations and the author has not attached information for a possible citation, but this indicator is not required under the current procedure;

(3) all publications have not been reported under a previous procedure;

(4) the publications of the candidate are free from plagiarism considering the performed checks in the peer-reviewed and indexed editions as well as the machine check of the dissertation work and the author's abstract.

5. Qualities of the abstract

As I noted in 4. (1), the abstract fully, accurately and correctly reflects the dissertation and the results obtained, taking into account the language difference between the two documents.

6. Critical remarks and recommendations

The wording of the title and essentially the whole study is not properly focused in order to present the content of the dissertation: instead of presenting a comprehensive study of a formulated model on the studied class of computational problems, the title focuses on solving an applied problem in porosity. . It corresponds more to a report of an applied research than to a dissertation, which assumes modeling of the problem area and study of the characteristics of the [multidimensional] model space thus obtained. I would prefer as a more community-relevant title of the type [Model of] massive parallelism in the processing of MLMC [class of] algorithms. And test application cases can be left in a subtitle or other level of presentation, albeit in a prominent place - especially in the final part with the experimental results. Incidentally, the initial motive to develop the calculation program proposed in the dissertation, even if it was modeling the porosity, for the purposes of the dissertation, it can easily be omitted, as the topic and the claimed contributions are not in the field of materials science. Of

course, the presence of an applied field is an important advantage and its citation, even emphasis, is strictly necessary, but rather in the approbation part of the dissertation.

Fortunately, this does not significantly affect the depth and quality of the study: for example, the research model can be read implicitly formulated in points 5.1 to 5.4, and in p. 5.5 and 5.6 give a research plan for experimental results with respect to two typical research cases.

The dissertation is presented in English, which masks the traditional problems with terminology in the main text, but they appear in the abstract. The direct translation of multilevel as multilevel is a word that sounds set and ambiguous. This is not the place and occasion for terminological discussion, but in this particular case I would suggest to the attention of Monte Carlo experts the simulation of the term multiphase or multistage: first, intermediate results must be generated, on the basis of which [from the next phase and possibly from another algorithm] results from another level or step are generated. This is also the case for calculations: "The expected value at the finest level is obtained as the sum of the calculated value for that level plus adjustments obtained for the values at the coarser levels".

Another reason for discussion is the use by the dissertation of the term granularity and its degrees from coarse (coarse - I would prefer large) to fine. The meaning of the use of granularity here is in the direction of precision or discretization of the numerical model of the corresponding structure of the solid. The advantage of precision and discretization is that they do not create a semantic conflict with each other - very important in parallel processing with data decomposition - a quantity that is also referred to as granularity [of decomposition]. This conflict remains hidden in the peer-reviewed work, but only because the problem with the granularity of the decomposition is not clearly addressed in the research plan.

The description of the executive platform / s for test experiments - hardware clusters of parallel machines, as well as the supporting intermediate platform - are given very schematically. In the meantime, individual parameters of this platform could participate in the High Parallel Processing Model and test their role in increasing efficiency - in particular, I mean adaptability to the cache of individual cores, as well as the "cost" of synchronization between local and remote processes. - ie in general of the topology of the processes and its relation to the topology of the nodes.

It is also known that in the field of Monte Carlo simulation modeling and numerical methods for its parallel processing in Bulgaria we have a whole school, but unfortunately this is not evident from the list of sources in the dissertation. On the other hand, I do not see how such

specialized text benefits from sources with a predominantly reference profile, such as [22, 23, 24].

Efficiency is correctly translated as efficacy, but is found in the abstract and as efficacy (in Fig. 4.7), a variant more suitable for the field of drugs.

7. Conclusion

After getting acquainted with the dissertation presented in the procedure and the accompanying scientific papers and based on the analysis of their significance and the scientific and scientific-applied contributions contained in them, I confirm that the presented dissertation and scientific publications to it, as well as the quality and originality of the results and achievements presented in them, meet the requirements of ZRASRB, the Regulations for its application and the respective Regulations of Sofia University "St. Kliment Ohridski "for acquiring by the candidate of the educational and scientific degree Doctor in the scientific field 4. Natural sciences, mathematics and informatics and professional field 4.6. Informatics and computer science. I confirm that the candidate meets the minimum national requirements in the professional field and no plagiarism has been established in the scientific papers submitted at the competition.

Based on the above, I recommend the scientific jury to award Nikolay Georgiev Shegunov educational and scientific degree Doctor in Science 4. Natural Sciences, Mathematics and Informatics and professional field 4.6. Informatics and computer science and I wish Nikolai a successful and motivating career as a young scientist.

August 8, 2021 Reviewer: .....

(Prof. Dr. Vasil Tsunizhev)