

Резюмета на научните публикации
на доц.д-р Емануил Атанасов, ИИКТ-БАН

за участие в конкурс за академична длъжност „професор“, по професионално направление Математика, специалност специалност „Изчислителна математика (високопроизводителни изчисления)“, за нуждите на секция „Грид технологии и приложения“, с ново наименование „Високопроизводителни системи, мрежи и алгоритми“, обявен в ДВ бр. 41/21.05.2019 г.

[1] Misev, A., Atanassov, E., Performance analysis of GRID middleware using process mining, (2008) Lecture Notes in Computer Science, 5101 LNCS (PART 1), pp. 203-212, DOI: 10.1007/978-3-540-69384-0_26, ISSN: 0302-9743

Анализът на производителността на GRID мидълуера (междинния софтуер), използван в продукционна среда, може да осигури ценна информация както на GRID потребителите, така и на разработчиците. Нов подход към този въпрос е използването на технологични техники за process mining. Анализирайки логовете на дейностите на мидълуера, извършени в пилотната продукционна инфраструктура на SEE-GRID, може да се получи обективна количествена и качествена информация за това какво всъщност се случва. Използвайки подходящи инструменти като ProM за прилагане на алгоритмите за process mining, могат да бъдат направени много интересни констатации и заключения. В тази статия се описва подхода за извършването на анализа и получените заключения.

[2] Stoykov, S., Atanassov, E., Margenov, S.. Efficient sparse matrix-matrix multiplication for computing periodic responses by shooting method on Intel Xeon Phi. AIP Conference Proceedings, 1773, 110012, AIP Publishing, 2016, ISBN:978-073541431-0, ISSN:0094-243X, DOI:10.1063/1.4965016, 110012-110012

Много от научните приложения включват операции с плътни или разреждени матрици, като например решаване на линейни системи, умножение на матрици продукти и т.н. интерес. Методът "shooting" се използва широко за получаване на периодични решения на нелинейни системи. Методът включва едновременно операции с разреждени и плътни матрици. Една от скъпоструващите изчислителни операции в метода е умножението на разредените с плътни матрици. В настоящата работа е представен нов алгоритъм за пресмятане на такива произведения. Алгоритъмът отчита структурата на разредената матрица, която се получава от пространствената дискретизация на нелинейното уравнение на движението на плоскостта на Миндлин чрез метода на крайните елементи. Алгоритъмът е разработен за използване на векторните възможности на ускорителите Intel Xeon Phi. Той се сравнява със стандартния алгоритъм и разработения от Intel за библиотеката MKL и е показано, че се получава значително подобрение.

- [3] Nedjalkov, M., Vasileska, D., Atanassov, E., Palankovski, V., Ultrafast Wigner transport in quantum wires, (2007) *Journal of Computational Electronics*, 6 (1-3), pp. 235-238, DOI: 10.1007/s10825-006-0101-y, ISSN: 1569-8025

Изследвани са два квантово-кинетични модела, които моделират преноса на първоначално високо неравновесно разпределени носители, генерирани локално в нано-проводна линия. Процесите на дисипация, дължащи се на фононите, управляват релаксацията на носителя, която в ранните етапи на еволюцията се характеризира с нарушаване на запазването на енергията в сблъсъците. Моделите се анализират и приближават числено чрез обратен метод Монте Карло. Основната разлика между тях е в начина на третиране на времето за крайна продължителност на сблъсъка. Последното въвежда квантови ефекти от разширяването и забавянето, свръхбързия пространствен трансфер и модификация на класическите траектории, които се демонстрират в представените резултати от симулациите.

- [4] Shterev, K., Atanassov, E., Stefanov, S., GPU calculations of unsteady viscous compressible and heat conductive gas flow at supersonic speed, (2014) *Lecture Notes in Computer Science*, 8353, pp. 549-556, DOI: 10.1007/978-3-662-43880-0_63, ISSN: 0302-9743

Неотдавнашната тенденция за използване на графични процесорни единици (GPU) за високопроизводителни изчисления се движи от високото съотношение на производителността към цените за тези устройства. Такива устройства все повече се разгръщат не само като ускорители за суперкомпютърни инсталации, но и в GPU-съвместими възли в Grid и Cloud инсталации. На пръв поглед изчислителните задачи на флуидната динамика (CFD) съвпадат перфектно с възможностите на графичния процесор, защото те правят интензивни изчисления и използват сравнително малко памет. Въпреки това, налице са оскъдни резултати за практическото използване на това сериозно предимство на GPU над CPU, особено за изчисления на вискозни, свиваеми, топлопроводими газови потоци с двойна точност. В тази работа е представено изчисляване на нестационарен, вискозен, свиваем и топлопроводим газ с двойна точност, използвайки GPU-поддръжка на версията на алгоритъма SIMPLE-TS, написана на стандарт OpenCL. Като тестов модел се използва потока покрай квадрат в микроканал при свръхзвукова скорост с число на Махат $M = 2,43$ на AMD Radeon HD 7950 GPU и е постигната скорост от 90 GFlops, което е 46 пъти по-бързо от серийния код на процесора на Intel Xeon X5560.

- [5] Ivanov, P., Atanassov, E., Jaime, C., Computational study on the conformations of CD38 and inclusion complexes of some lower-size large-ring cyclodextrins, (2014) *Journal of molecular structure*, 1056, pp. 238-245, DOI: 10.1016/j.molstruc.2013.10.048, ISSN: 0022-2860

Конформациите на CD38 са изследвани чрез конформационно търсене със симулации на молекулярната динамика, използвайки силовото поле Glycam04. Резултатите са сравнени с предишните за CD26, най-големият циклодекстрин, за

който са налични определени данни. При последващата обработка на симулационните траектории е приложен анализ на основните компоненти (РСА). Ограниченият брой режими определят общите деформации на макроринга на CD38. По-дългият периметър на макрорежима позволява образуването на форма, която не се наблюдава досега - триходова спирала, оформена като къса тръба. По аналогия с CD26, значителното участие беше наблюдавано за конформации на CD38 с еднопосочни спирали на противоположните страни на макролинията, свързани помежду си от „дъното” и от „върха” с удължени разделители на моста. Най-интересното поведение представя комплекса на CD26 с адамантан, в който случай малката молекула действа като "център на нуклеация" за образуване на втори спирален оборот около молекулата на субстрата.

[6] Balaž, A., Prnjat, O., Vudragović, D., Slavnić, V., Liabotis, I., Atanassov, E., Jakimovski, B., Savić, M., Development of Grid e-Infrastructure in South-Eastern Europe, (2011) Journal of Grid Computing, 9 (2), pp. 135-154, DOI: 10.1007/s10723-011-9185-0, ISSN: 1570-7873

За период от шест години и три фази програмата SEE-GRID създаде силна регионална човешка мрежа в областта на разпределените научни изчисления и създаде мощна регионална грид-инфраструктура. Тя привлече редица потребителски общности и приложения от различни области от страни в Югоизточна Европа. От гледна точка на инфраструктурата, първата фаза на проекта създаде пилотна Грид инфраструктура с повече от 20 ресурсни центъра в 11 страни. По време на следващите две фази на проекта инфраструктурата се разрасна до 55 ресурсни центъра с повече от 6600 процесора и 750 ТВ дисково пространство, разпределени в 16 участващи страни. Включване на нови ресурсни центрове в съществуващата инфраструктура, както и подкрепа за нови потребителски общности, изисква създаването на регионално разпределени основни услуги, разработване на нови мониторингови и оперативни инструменти и тясно сътрудничество на всички партньорски институции в управлението на такава сложна инфраструктура. В тази статия е направен преглед на развитието и актуалното състояние на регионалната инфраструктура на SEE-GRID и нейните услуги, както и прехода ѝ към модела, базиран на NGI в EGI, със силното регионално сътрудничество в ЮИЕ.

[7] Atanassov, E., Gurov, T., Karaivanova, A., Energy aware performance study for a class of computationally intensive Monte Carlo algorithms, (2015) Computers & mathematics with applications, 70(11), pp. 2719-2725, DOI: 10.1016/j.camwa.2015.07.014, ISSN: 0898-1221

Най-новите разработки в областта на НРС доведоха до внедряването на сложни системи от екстремни мащаби, базирани на различни изчислителни устройства (CPU, GPU, ускорители), като по този начин поставиха въпроса за скалируемост в светлината не само на паралелната ефективност, но и на енергийната ефективност.

В тази статия са предложени нови показатели за оценка на енергийната ефективност въз основа на опита на авторите и на анализ на съществуващите показатели. Изследвана е работата на интензивни изчислителни приложения от тип Монте Карло, разгърнати на хетерогенни HPC системи, с акцент върху енергийната ефективност и разходите за оборудване. Сравняват се резултатите за енергийна ефективност на CPU и GPU варианти на тестваните алгоритми по отношение на въведените мерки и метрики. Резултатите от проучването показват важността на отчитането не само на мащабируемостта на HPC приложенията, но и на енергийната ефективност и разходите за оборудване. Те също така показват как да се оптимизира изборът на изчислителни процесори или сървъри с GPGPU. Резултатите могат да се използват от разработчици на приложения / потребители и от доставчици на ресурси.

[8] Atanassov, E., Gurov, T., Karaivanova, A., Energy Performance Evaluation of Quasi-Monte Carlo Algorithms on Hybrid HPC, (2015) Large-scale scientific computing, 9374, pp.172-181, DOI: 10.1007/978-3-319-26520-9_18, ISSN: 0302-9743

Нарастващите изисквания на научните приложения и нарастващият капацитет на съвременните компютърни системи водят до необходимостта от оценка на потреблението на енергия и съответно до разработването на енергийно ефективни алгоритми. В тази статия се изучават енергийните характеристики на един клас квази-Монте Карло алгоритми на хибридни HPC системи. Тези алгоритми се прилагат за решаване на квантови кинетични интегрални уравнения, като се използват редиците на Соболев и Холтън. Резултатите за енергоемкостта се сравняват между компютърни платформи, базирани на процесори, и компютърни платформи с ускорители като GPU карти и процесори Intel Xeon Phi по отношение на няколко показателя.

[9] Atanassov, E., Dimitrov, D., Gurov, T., Evaluation of Stochastic Algorithms for Financial Mathematics Problems from Point of View of Energy-efficiency, (2015) Application of mathematics in technical and natural sciences, 1684, DOI: 10.1063/1.4934300, ISSN: 0094-243X

Използването на различни типове ускорители като графични процесори, Intel Xeon Phi стана основна и много алгоритми и приложения бяха пренесени, за да ги използват, когато има такива. Във финансовата математика на въпроса за оптималното използване на изчислителните ресурси трябва да се вземат предвид и ограниченията на пространството, тъй като в много случаи на използване сървърите се разполагат в близост до източниците на данни (exchanges). В тази работа се оценяват различни алгоритми за ценообразуване на опции, които се прилагат за различни целеви архитектури по отношение на тяхната енергийна ефективност. Тъй като е установено, че редиците с нисък дискрепанс могат да

бъдат по-добри от псевдослучайните числа за тези типове алгоритми, също така се изследват и редиците на Соболев и Холтън. Представени са както необработените резултати, така и изчислените показатели и изводите от тестовете.

[10] Özturan, C., Kotroni, V., Atanassov, E., Development of virtual organizations, applications and services for earth science on grid e-Infrastructures, (2010) Earth Science Informatics, 3 (4), pp. 197-198, DOI: 10.1007/s12145-010-0074-z, ISSN: 1865-0473

Гридовете се появиха като платформа за споделяне на изчислителни ресурси в началото на 2000-те години. Концепцията за Грид беше обобщена, за да включва споделянето не само на изчислителни ресурси, но и на други ресурси, като данни, устройства и приложения в така наречените виртуални организации (VOs). Гридовете помагат да преминем към така наречената ера на е-науката (E-Science). E-Science се отнася до генерирането на компютърни знания, използвайки различни комбинации от масивни изчислителни ресурси, данни и устройства в силно разпределени Грид среди. Електронната наука е следващата естествена стъпка отвъд уеб-базираните платформи от последното десетилетие, която предлага разпространение на научни данни и работа с приложения чрез уеб интерфейси на клиент-сървър архитектури. В статията са разгледани недостатъците на традиционните уеб-базирани платформи и предимствата на използването на Грид за E-Science.

[11] Atanassov, E., Karaivanova, A., Gurov, T., Ivanovska, S., Durchova, M., Dimitrov, D.S., Quasi-Monte Carlo integration on the grid for sensitivity studies, (2010) Earth Science Informatics, 3 (4), pp. 289-296, DOI: 10.1007/s12145-010-0069-9, ISSN: 1865-0473

В тази статия е представен анализ на грешките и производителността на квази-Монте Карло алгоритми за решаване на многомерни интегрални (в размерност до 100) в Грид среда с помощта на MPI. Взема се под внимание фактът, че Грид е потенциално хетерогенна изчислителна среда, в която потребителят не познава спецификата на целевата архитектура. Следователно паралелните алгоритми трябва да могат да се адаптират към тази хетерогенност, като осигуряват автоматизирано балансиране на натоварването. Алгоритмите Монте Карло могат да бъдат пригодени към такива среди, при условие че са налични подходящи паралелни генератори на псевдослучайни числа. Използването на квази-Монте Карло алгоритми създава повече трудности. И в двата случая ефективната реализация на алгоритмите зависи от функционалността на съответните пакети за генериране на псевдослучайни или квазииндивидуални числа. В статията се предлага ефективна паралелна реализация на редиците на Соболев за Грид среда и се демонстрират числени експерименти върху хетерогенен Грид. За да се постигне висока паралелна ефективност, се използва новоразработена специална Грид услуга наречена услуга за проследяване на задачите, която осигурява ефективно управление на наличните компютърни ресурси чрез резервации.

- [12] Georgiev, D., Atanassov, E., Extensible framework for execution of distributed genetic algorithms on grid clusters, (2014) 37th international convention on information and communication technology, electronics and microelectronics (MIPRO), pp. 301-306, DOI: 10.1109/MIPRO.2014.6859581, ISBN: 978-953-233-077-9

Генетичните алгоритми са ефективни мета-евристични методи за оптимизация, основани на принципите на естествения подбор и генетиката. Въпреки че са в състояние да намерят адекватни решения в приемливо време за малки проблеми, времето им на изпълнение се увеличава бързо за проблеми с голямо пространство за търсене и сложна оценъчна функция. Изчислителните гридове осигуряват подходяща платформа за изпълнение на паралелни генетични алгоритми. Въпреки това, когато се използват няколко кълстера за една компютърна задача, трябва да се спазват редица технически и административни ограничения.

В тази статия е описана нова реализация на GA, написана на C ++ 11, която осигурява механизъм за паралелно изпълнение и разширяем интерфейс за програмиране. Предложената рамка използва MPI, ZeroMQ и най-новите средства за програмиране на

C ++, което позволява изпълнението на разпределени GA на множество грид кълстери. Проведени са различни тестове на кълстери от европейската грид инфраструктура, които доказват ефективността на новия подход.

- [13] Atanassov, E., Gurov, T., Karaivanova, A., Message Oriented Framework with Low Overhead for Efficient High-Performance Monte Carlo Simulations, (2013) 36th international convention on information and communication technology, electronics and microelectronics (MIPRO), pp. 169-171, ISSN:1847-3946

През последните години България придоби значителни високопроизводителни ресурси от различни видове. Най-голям ресурс в този момент е суперкомпютърът BlueGene / P на SAITC с 8192 процесорни ядра, докато Българската академия на науките има два HPC кълстера с Intel CPU и Infiniband взаимно свързване, което е общо над 1000 логически ядра. В допълнение някои сървъри, оборудвани с мощни графични процесори, са на разположение за приложения, които могат да се възползват от тях. Координираното използване на такива ресурси от едно приложение е изправено пред значителни предизвикателства, дължащи се на разнородността на ресурсите и ограниченията в мрежите и сигурността.

За да се улесни координираното използване на всички тези ресурси, където всеки ресурс се използва за частите на приложението, където той е най-ефективен, е разработена рамка, която позволява на изследователя да свърже ресурсите на

горните типове с минимални разходи. В тази статия е описана архитектурата на системата и е демонстрирана нейната ефективност за приложение за моделиране на полупроводници, показващо числени и времеви резултати.

[14] Misev, A., Atanassov, E., User level Grid Quality of service, (2010) Lecture Notes in Computer Science, 5910 LNCS, pp. 507-514, DOI: 10.1007/978-3-642-12535-5_60, ISSN: 0302-9743

Подобряването на качеството на обслужването (Quality of Service) на грид-инфраструктурата е един от най-важните текущи въпроси в грид-общността. Той има много последици - от по-широкото приемане от страна на потребителите на технологията до пренасочването на използването на грид от научния свят към бизнеса. Докато QoS базиран най-добрите усилия (best-effort) може да бъдат приемлив в научната общност, бизнес приложенията трябва да имат ясни и добре дефинирани нива на услуги, базирани на детерминистични QoS метрики. В тази статия е представен различен тип Grid QoS, под контрола на потребителите. Чрез него потребителите могат да имат по-добър контрол върху нивото на услугите, които използват от грид-инфраструктурата.

[15] Atanassov, E., Karaivanova, A., Ivanovska, S., Tuning the generation of Sobol sequence with Owen scrambling, (2010) Lecture Notes in Computer Science, 5910 LNCS, pp. 459-466, DOI: 10.1007/978-3-642-12535-5_54, ISSN: 0302-9743

Редиците на Соболев е най-широко използваната редица с нисък дискрепанс при числено решаване на множество интеграла и други квази-Монте Карло изчисления. Оуен първи предложи скрамблиране на тази редица чрез пермутации по начин, който поддържа ниския дискрепанс и подобрява сходимостта за определени класове функции. Реализацията му е необходима не само за получаване на оценка на грешката, но и за паралелни изчисления. Доброто разбъркване е особено важно за GRID приложенията. Въпреки това, често е трудно да се осъществи scrambling. В тази статия е предложен метод за бързо генериране на редици на Соболев със скрамблиране по Оуен, настроен за специфичен хардуер. Представени са и са дискутирани резултатите от числени експерименти и времеви резултати, които демонстрират предимствата на предложения подход.

[16] Misev, A., Atanassov, E., ULMon - Grid monitoring from user point of view, (2009) Proceedings of the International Conference on Information Technology Interfaces, ITI, pp. 621-626, DOI: 10.1109/ITI.2009.5196158, ISSN: 1330-1012

Използването на сложна инфраструктура, като например компютърните гридове, изисква подходящо ниво на мониторинг. Тъй като това е технология в процес на разработка, сегашният изчислителен мидълуер на Grid предлага много малко възможности за наблюдение от гледна точка на потребителите. Приближаването на

мрежата до потребителите изисква потребителите да имат по-задълбочено разбиране за това какво се случва с изчислителните задачи, които представят. В тази статия е представен нов инструмент за наблюдение на изпълнението на задачите в Грид инфраструктурата от гледна точка на потребителите.

[17] Atanassov, E., Dimov, I.T., What Monte Carlo models can do and cannot do efficiently?, (2008) Applied Mathematical Modelling, 32 (8), pp. 1477-1500, DOI: 10.1016/j.apm.2007.04.010, ISSN: 0307-904X

За някои функционални пространства, които определят регулярността на входните данни, се обсъжда въпросът "какво могат да правят моделите на Монте Карло и какво не може да се направи ефективно". Разглеждат се класове данни, които са важни за практическите изчисления: класовете функции с ограничени производни и с условия от тип на Хьолдер, както и пространствата, подобни на разглежданите от Коробов.

Даден е теоретичен анализ на работата на някои алгоритми с неподобряема скорост на сходимост. Представени са оценки на изчислителната сложност на два класа алгоритми – детерминистични и рандомизирани за двата проблема - числено интегриране на многомерни функции и изчисляване на линейни функционали на решението на клас интегрални уравнения.

[18] Atanassov, E., Georgiev, D., Manev, N., Number Theory Algorithms on GPU Clusters, (2014) High-performance computing infrastructure for south east europe's research communities: results of the hp-see user forum 2012, 2, pp. 131-138, DOI: 10.1007/978-3-319-01520-0_16, ISSN: 2196-7334

Много алгоритми от теорията на числата и тяхното внедряване в софтуер са от голямо практическо значение, тъй като са изграждащи примитиви на много протоколи за криптиране на данни и удостоверяване на интернет връзки. Алгоритмите от теория на числата също са основната част от криптоаналитичните процедури. Много от тези алгоритми могат да бъдат паралелизирани по естествен начин. В тази статия е описан софтуерен пакет, който изпълнява различни алгоритми от теория на числата на кълъстерите с поддръжка на графични процесори и нова реализация на целочислената факторизация, използвайки NVIDIA CUDA върху кълъстери, оборудвани с NVIDIA GPUs. Също така се докладват и резултатите от експерименти използващи новата реализация.

[19] Georgiev, D., Atanassov, E., Alexandrov, V., A Framework for Parallel Genetic Algorithms for Distributed Memory Architectures, (2014) 5th workshop on latest advances in

scalable algorithms for large-scale systems (SCALA), pp. 47-53, DOI: 10.1109/ScalA.2014.13, ISBN: 978-1-4799-7562-4

Генетичните алгоритми са мета-евристични методи за търсене, основани на принципите на биологичната еволюция и генетика. Използването на паралелни реализации на генетични алгоритми с цел овладяване на мощта на съвременните компютърни платформи е мощен подход за смекчаване на този проблем. В тази статия се правят няколко паралелни реализации, вариращи от MPI до хибридни MPI / OpenMP и MPI / OmpS. Тези реализации са оптимизирани за изпълнение върху плътно свързани системи с разпределена памет. Разглеждат се въпроси, които възникват при работа с разпределен генетичен алгоритъм и е представена адаптивна миграционна схема. Направено е и сравнение на тяхната ефективност.

[20] Atanassov, E., Durchova, M., Generation of the Scrambled Halton Sequence Using Accelerators, (2013) 36th international convention on information and communication technology, electronics and microelectronics (MIPRO), pp. 177-181, ISSN: 1847-3946

Редицата на Холтън е една от най-популярните редици с нисък дискрепанс. За да се задоволят някои практически изисквания, оригиналната редица обикновено се модифицира по някакъв начин. Алгоритъмът за разбъркване (scrambling), предложен от Оуен, има няколко теоретични предимства, но от друга страна е труден за прилагане на практика поради компромиса между висока памет и високи изчислителни изисквания. В нашата работа се концентрираме върху случая, когато броят на координатите е сравнително висок. Използването на изчислителни ускорители и особено на графичните процесори е все по-важно за такива практически приложения, тъй като все повече и повече ресурси, достъпни чрез инфраструктурата на грид и Cloud, осигуряват достъп до такива ускорители, при условие че софтуерът може да ги използва. В тази статия е представен нов алгоритъм за генериране на последователността на Холтън със scrambling на Оуен, реализирано с помощта на NVIDIA CUDA. Също така са показани числени резултати, постигнати на възли, които са оборудвани с NVIDIA M2090 карти.

[21] Atanassov, E., Ivanovska, S., Computation and Analysis of Sobol Coefficients for Air Pollution Concentrations over the Territory of Bulgaria, (2013) 36th international convention on information and communication technology, electronics and microelectronics (MIPRO), pp 234-237, ISSN: 1847-3946

Един от основните инструменти за моделиране на замърсяването на въздуха на територията на България е системата на САЩ EPA Models-3. Основните компоненти на системата са MM5 / WRF метеорологичен препроцесор, препроцесор SMOKE емисии и CMAQ химически транспортен модел. Като въглеродни емисии се използва инвентаризацията на емисиите на TNO. Опцията

Модел-3 "Интегриран анализ на скоростта на процеса" се прилага, за да се разграничи ролята на различните динамични и химични процеси за замърсяването за всички SNAP категории. В тази работа оценяваме влиянието на различните входни параметри като концентрация в различни SNAP категории при постоянни метеорологични условия към изходните концентрации над България по методология на Соболев-Салтели. За да се получат надеждни оценки на коефициентите Sobol, ние изпълняваме големи MPI задачи, използвайки кластерите в Югоизточния регион. Използвайки тези коефициенти, ние оценяваме относителната важност на различните входни параметри и техните взаимодействия.