Supercomputing Frontiers and Innovations

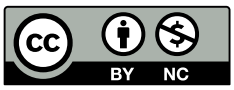
2018, Vol. 5, No. 3

Scope

- Enabling technologies for high performance computing
- Future generation supercomputer architectures
- Extreme-scale concepts beyond conventional practices including exascale
- Parallel programming models, interfaces, languages, libraries, and tools
- Supercomputer applications and algorithms
- Distributed operating systems, kernels, supervisors, and virtualization for highly scalable computing
- Scalable runtime systems software
- Methods and means of supercomputer system management, administration, and monitoring
- Mass storage systems, protocols, and allocation
- Energy and power minimization for very large deployed computers
- Resilience, reliability, and fault tolerance for future generation highly parallel computing systems
- Parallel performance and correctness debugging
- Scientific visualization for massive data and computing both external and in situ
- Education in high performance computing and computational science

Editorial Board

Contents



DOI: 10.14529/jsfi180301

Continuum Computing – on a New Performance Trajectory beyond Exascale

Thomas Sterling[1]***, Maciej Brodowicz***[1]***, Matthew Anderson***[1]

The end of Moore's Law is a cliché that none the less is a hard barrier to future scaling of high performance computing systems. A factor of about 4× in device density is all that is left of this form of improved throughput with a 5× gain required just to get to the milestone of exascale. The remaining sources of performance improvement are better delivered efficiency of more than 10× and alternative architectures to make better use of chip real estate. This paper will discuss the set of principles guiding a potential future of non-von Neumann architectures as adopted by the experimental class of Continuum Computer Architecture (CCA). It is being explored by the Semantic Memory Architecture Research Team (SMART) at Indiana University. CCA comprises a homogeneous aggregation of cellular components (function cells) which are orders of magnitude smaller than lightweight cores and individually is unable to accomplish a computation but in combination can do so with extreme cost efficiency and unprecedented scalability. It will be seen that a path exists based on such unconventional methods like neuromorphic computing or dataflow that not only will meet the likely exascale milestone in the same time with much better power, cost, and size but also will set a new performance trajectory leading to Zettaflops capability before 2030. The remainder of this paper is organized as follows. Section 1 describes a new class of high performance architectures and discusses the rationale for its introduction. The scaling analysis is presented in section 2. Future performance projections, including steps leading to Zettaflops, are outlined in section 3. Finally, the principal outcomes of this study are summarized in the conclusions section.

Keywords: high performance computing, parallel computing, exascale, non-von Neumann architecture.

Introduction

The fastest computer in the world measured by the HPL or Linpack benchmark [5] is the Summit [19], at Oak Ridge National Laboratory, in the United States. As recently measured, its performance is 188 Petaflops peak performance and 122 Petaflops R_{max}. Number 1 on the June 2018 Top-500 List [2], it is followed by two Chinese systems with TaihuLight coming in at Number 2 and 93 Petaflops R_{max} positioned there for two and a half years. Following these are 498 HPC systems with the 500^{th} still measured in the hundreds of Teraflops. These unprecedented advances reflect an exponential progress over a period of more than two dozen years with a total growth of more than a factor of a million over that period and of more than ten trillion since the beginning of the age of the modern digital electronic stored program computer more than 70 years ago. Much of this is a combination of von Neumann architecture [17] derivatives and the sustained improvement of device components often regarded as Moore's Law [16]. The current era could perhaps be referred to as Pax Semiconductor as the constant tick-tock of produce development and delivery has provided the confidence of continued improvements into the indefinite future. However, as the enabling technology approaches that of nano-scale measured in number of atoms of thickness, the previous trends that have gained us so much are meeting their terminus. Feature sizes, power consumption, and clock rates are all reaching their flat asymptotes while parallelism both at ILP and concurrent processes are struggling to grow. Even as HPC capabilities approach exascale, admittedly at enormous cost, the future embracing passed conventional practices is uncertain and of questionable utility.

[1]Center for Research in Extreme Scale Technologies, Indiana University

Although the age of digital computing exhibits the semblance of a continuity of performance growth and therefore the appearance of sameness, the truth is far more complex. There have been at least six phases of computing paradigms to exploit the evolving capabilities of enabling technologies. In each case, both computer architectures and programming models with supporting software have changed distinctly in response to driving technology advances to take advantage of their opportunities and to address the challenges they impose. One possible delineation of these epochs with very fuzzy overlapping borders is presented as:

I. Genesis – vacuum tubes, magnetic storage, von Neumann architecture (e.g., EDSAC, UNIVAC).

II. Transistor – semiconductor, micro-control, operating system (e.g., IBM 7090, CDC 6800).

III. SSI (small scale integration) – first generation integrated circuits (e.g., DEC PDP-8, IBM 360).

IV. MSI (medium scale integration) – vector pipeline FPU, vector register (e.g., Cray-1).

V. LSI (large scale integration) – SIMD (e.g., TMC CM-2, MASPAR-1).

VI. VLSI (very large scale integration) – MPP, commodity clusters, networks (e.g., Beowulf, Intel Touchstone Delta, Cray T3E, Red Storm).

VII. Multi-core – multiple processors per socket (e.g., IBM Roadrunner, Sunway TaihuLight).

VIII. Heterogeneous – integrated specialized computing components (e.g., IBM/NVIDIA Summit).

As indicated, each of these epochs comprise a distinction in form and function in response to technology. With the new challenges of the end of Moore's Law, it is important to determine the next set of innovations that will force future computing methods and means if HPC is to extend beyond the boundaries of exascale by the beginning of the next decade. Even now forwarding looking concepts are being researched within academia and industry in areas such as quantum computing and neuromorphic computing. Quantum Computing directly exploits quantum mechanical phenomena to store the equivalent of many bit values in a single storage element called a "qubit" [20] and process the many possible solutions at the same time, delivering one of the results at the end. Cooled to a fraction of a Kelvin these devices employ quantum superposition and entanglement to achieve this. For some algorithms, exponential scaling should in principle be achievable with polynomial scaling for others. It is recognized that a conventional computer would take longer than the lifetime of the universe (i.e., 13.82 billion years) to accomplish what a future quantum computer may be able to do in finite time.

A second paradigm shift being pursued internationally is variably called "neuromorphic computing" [15] or "brain-inspired computing" [14] for employing techniques motivated by neural structures of the mammalian brain. Some techniques modeled after complex topologies of components representative of neurons with their spiking behavior and thousands of interconnection points. With some, such structures are emulated with multi lightweight core chips. For others, actual circuits designed to reflect neuronal-like devices are employed. The hope is to be able to achieve large complex associative and pattern matching processing that humans appear to do effortlessly and at very low power consumption. As interesting and possibly promising

both quantum computing and neuromorphic computing are, much work in both methods will be required. But the challenges facing the future of computing are immediate and alternative techniques are required.

The remainder of this paper discusses an alternative conceptual strategy based on near-term semiconductor technologies but, perhaps more than in previous epochs, embracing innovations in computing model, architecture, and programming. Like all previous generation systems this addresses the fundamental factors that determine ultimate performance. These are: starvation, latency, overhead, and contention with additional concerns of energy consumption and reliability. But it relaxes the constraints inherent to essentially all previous widely used HPC systems. Specifically, it employs dynamic adaptive parallel task scheduling and resource management as opposed to conventional static system and application management. Of equal or more significance, is the abrogation of premises implicit in all von Neumann derivative architectures (exclusively populating the Top 500 list [1]: Fig. 1) to liberate computer architecture from the constraints of legacy machines and opening up new possibilities through potential non-von Neumann architecture, still within the realm of known enabling technologies and chip fabrication methods and facilities. The conclusion of this paper as will be shown is that far from being incarcerated in performance by the end of Moore's Law, opening a new path to non-von Neumann yields new opportunities to future performance growth. Early analysis suggests more than three orders of magnitude peak performance in less than a decade.

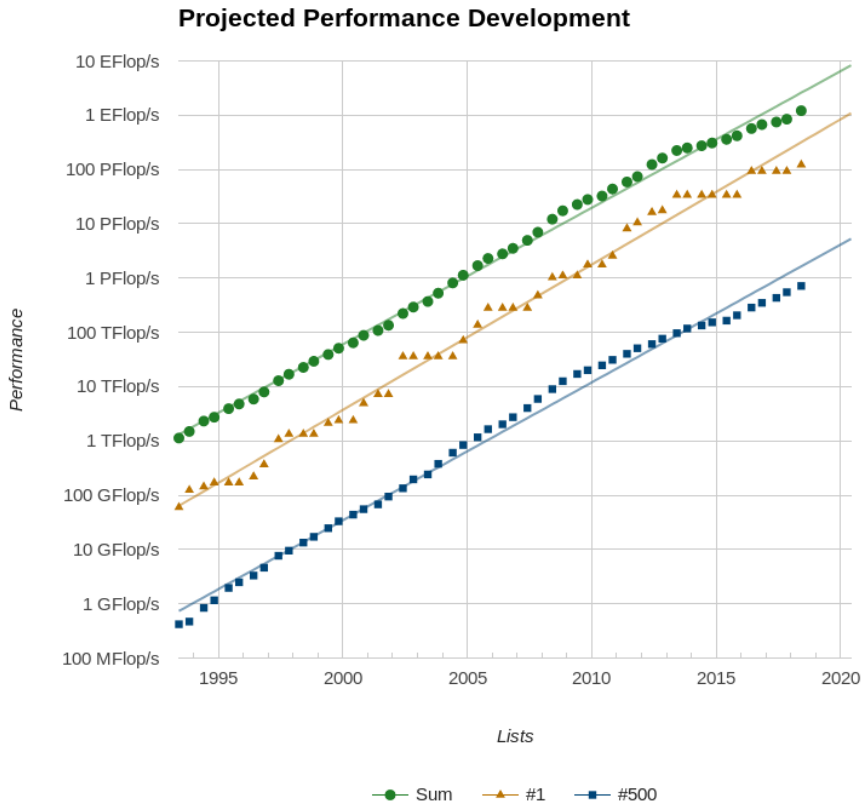


Figure 1. Supercomputer performance trends as of June 2018

1. Foundation Concepts for CCA

HPC applications are expanding to encompass not just conventional static numeric intensive problems but also those incorporating complex data structures typified by time-varying and irregular organizations referred to generally as dynamic graphs. Rapidly expanding market domains such as data analytics, machine learning, facial and voice recognition, autonomous vehicles, crypto-currency, the internet of things (also referred to as sensor nets), and 3-D virtual reality, are among other new computational application targets. There are a number of competitive approaches to satisfying user demands related to these and other scientific and commercially valuable problems including conventional multi-core SMP and MPP configurations and heterogeneous system architectures mixing general purpose processors with GPU accelerators. For some user applications, these types of systems are effective. But two major challenges are facing the future of fast computing and in so doing are opening new opportunities that, with insightful innovations, may be satisfied exposing the potential for new system classes and capabilities. These are the limited future scalability due to the end of Moore's Law and constrained power and the computing data access patterns with minimal data reuse. The proposed innovations present the opportunity for a new class of computing that address these and other challenges while exploiting largely untapped improvements. These are described in detail here.

It is ironic that the single most dramatic technology in terms of rate of growth is also very conservative. Over more than six decades, in spite of the multiple transitions in enabling technologies from vacuum tubes to multi-core sockets, the von Neumann architecture has dominated with the apparently diverse computing classes (e.g., vector versus communicating sequential processes (CSP) [9]) almost all von Neumann derivatives. The trade-offs implicit to the von Neumann architecture, once entirely justified, are no longer optimal; possibly the opposite as related to todays enabling technology. One of the staples of computing has been the compiler determined static management of resources and task scheduling. This traditional method has been effective for regular static algorithms. But for those many problems with behavior determined by intermediate results, runtime information is required to guide both task scheduling and resource management, especially for distributed systems with large memory hierarchies and widely varying latencies of access. The challenges of starvation, latency, overhead, and contention for shared resources must be addressed by any computing system. The proposed innovations provide innovative techniques for doing so.

The proposed strategy combines an innovative non-von Neumann architecture to employ current and future enabling semiconductor technologies with a control methodology based on the ParalleX execution model [6, 11] for dynamic adaptive resource management and task scheduling to deliver superior efficiency and scalability. Together they address their mutual needs in synergy to provide potential customers with low cost solutions to their computing requirements. These are described in detail below with the resulting value delivered following.

1.1. ParalleX Dynamic Control

The non-von Neumann architecture (and its justification of innovation) described below, alone, does not explain associated resource management and task scheduling. While these critical requirements are not fully devised in detail at this time, the ParalleX execution model (Fig. 2) provides a comprehensive strategy for addressing these needs. Also, in advocating the innovation enabled by ParalleX, its critical need is drastic reduction of overheads, availability of hardware

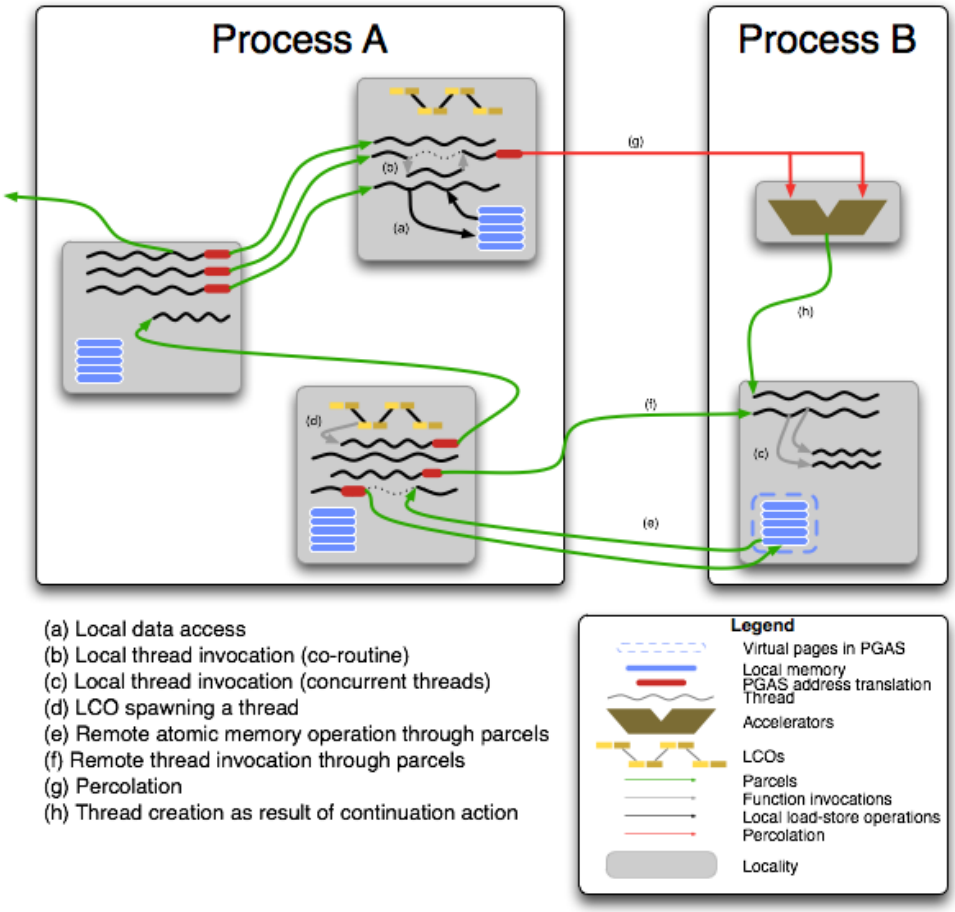


Figure 2. Components of the ParalleX execution model

fine grain parallelism for scalability, and latency hiding. The hardware architecture delivers all of these mechanisms to the performance of the ParalleX execution model.

ParalleX replaces the CSP execution model that has served and dominated large scale HPC for the last three decades providing a new dynamic methodology in place of the old static techniques. ParalleX integrates four key properties to enable dynamic adaptive execution, greater scalability through exposing and exploiting greater parallelism, and addressing as well as exploiting asynchrony of execution. These key constructs are:

- Global address space that enhances user productivity and performance portability.
- Local dataflow parallel tasks.
- Message-driven computation.
- Asynchronous Futures/dataflow synchronization constructs.

All elements are ephemeral and except for the parcels are first class objects which means that the can be addressed by the user applications or the system software dedicated to runtime control.

ParalleX addresses the critical challenges reflected by the SLOWER equation [22] of performance factors. Computing systems derived from this formulation can benefit from these optimizations. It mitigates starvation by exposing and exploiting more forms and more amounts of parallelism including lightweight threads with rapid context switching for non-blocking. By reducing or eliminating the parallelism control it yields finer granularity and therefore exposes more parallelism. ParalleX incorporates first class objects that allow it to engage in runtime parallelism discovery through data-directed (meta-data) execution within graph data structures.

And, it overlaps successive phases of computation by avoiding global barrier synchronization for more effective parallelism. Latency is another source of performance degradation and ParalleX delivers methods of mitigating both latency and the asynchrony to which it contributes. As previously identified, ParalleX supports lightweight thread context switching for non-blocking and overlaps computation and communication to hide latency to limit its effects. It employs massage-driven computation to move the work to the data rather than always moving the data to the work thus reducing actual latency under some conditions through the reduction of number and size of global messages. Perhaps the most insidious of the SLOWER factors is overhead work. It not only imposes useless work adding time to the computation and energy consumption as well. But for a fixed amount of work (strong scaling) it establishes an upper bound on the total performance achievable independent of the quantity of hardware dedicated to it. ParalleX-based systems achieve this by mostly eliminating global barriers, employing synchronization objects exhibiting high semantic power with respect to amount of work required. It reduces context switching time. And for specific actions, thread instantiation is not required thus avoiding its overheads completely in the case of very lightweight tasks that are most sensitive to overheads. Finally, contention is a source of bottlenecks most recognized in the forms of memory access bandwidth and network bandwidth. Ironically, it is ALU limitations that are the source of largely unacknowledged contention that is addressed in brute force fashion through scaling. ParalleX class systems approach this through adaptive resource allocation with a multiplicity of resources and by eliminating polling. It also facilitates adaptive routing to work around hot spots.

Multiple early implementations of the ParalleX execution model have been developed in the form of runtime systems software in support of applications that will benefit from these capabilities such as adaptive mesh refinement, fast multipole methods, particle in cell codes, and deep learning. But limitations in efficiency were measured due to overheads of key mechanisms for some applications more suitable to conventional practices. Nonetheless, the ParalleX model was demonstrated as an effective abstraction for capturing and controlling the interface between parallel applications and parallel architectures such as the non-von Neumann one here.

An implementation of the ParalleX execution model was the HPX-5 runtime system software package and employed for a wide range of applications. Among the many measurements taken of the SLOWER performance factors were the major ways in which overheads intruded into the overall operation of the parallel system. This work was performed by Daniel Kogler, a graduate student at Indiana University, exposed a surprising result. It was shown that while the minimum of overhead times was in the hundreds of nanoseconds, there was a very wide distribution of times; as much as a factor of two in some cases and approaching an order of magnitude in others. This variability which was unexpected creates a new problem and is caused by as well contributes to the uncertainty of asynchrony.

There are many sources of operational asynchrony. Principal among these is system scale. As the scale of a system grows, the distribution of interacting components expands as well contributing to the variability of the time to completion of any specific operation. One of these is caused by the integrating network of the system with its increasing range of latencies and also the opportunities for packet collisions. More complex topologies in some systems allow dynamic routing. All these network related factors contribute to the variability of operation and the uncertainties due to exaggerated asynchrony. Another major property is the adoption of deeper and wider memory hierarchies. Intended to reduce average latency of data access, a side effect is the variability of possible access times from the shortest to L1 data caches to the

longest main memory blocks across the extent of the distributed system. The sensitivity is not just with the inherent distances involved but to data placement and localities, both temporal and spatial. Another source of asynchronous behavior is due to scheduling conflicts for shared resources by multiple concurrent threads. The order of scheduling is undetermined in part due to other asynchronies therefore increasing the variability of tasks, but time to initiation and time to completion. To manage power consumption based on resource control, systems will actively and sometimes automatically vary voltage and clock rates causing variability of time for operations to be performed, adding to the asynchrony. Other sources also exist that extend the uncertainties of asynchrony even more. Therefore, systems need to be designed that are able to respond and adjust to asynchronous operation and where possible reduce both the times and variability of actions. One important implication is that new hardware mechanisms be incorporated in future architectures to mitigate these overheads and their uncertainty. Processor core architecture that includes such mechanisms to support parallel processing overheads for this purpose is discussed in section 1.3.

1.2. The von Neumann Cul-de-Sac

Figure 3. John von Neumann (left) and Robert Oppenheimer (right) in front of the IAS computer

The von Neumann architecture concept introduced in 1945 by Eckhart, Mauchly, and von Neumann (Fig. 3) established the principles of the digital electronic stored-program computer that has dominated almost all commercial computing designs since then and from which they have been derived. Even advanced HPC systems such as vector, SIMD, SMP, MPP, and clusters are all at their core von Neumann in origin. When first prescribed, this breakthrough concept was strongly influenced by the available enabling technologies, their costs, and their representative trade-offs. The choices made, for example the dominant investment in the ALU, were reasonable not just at the time but for three decades onwards. But todays enabling technologies are both very different from the original catalyzing devices and demand very different trade-offs for optimization of system structures with respect to critical metrics such as performance, time, energy, size, and cost. For decades the continued exponential improvement of feature size and clock rate demanded only those architecture changes that incrementally retained the same relationship to user codes as the previous generation while gaining in performance and improving in market

share. As a result, complicated core designs emerged with many layers of memory hierarchy (caches) transparent to the user, speculative execution such as branch prediction and memory access, associative mechanisms such as TLBs and cache line selection, and a plethora of other specialized functions to incrementally extend the capabilities just a bit farther than the previous tick-tock product cycle. As all these specializations accrued, the relative size, cost, and power of the ALU itself became minor in comparison. The other major change was that of the enabling technologies themselves and their interrelationships. Initially, major classes of components (i.e., memory, logic and control, and communication) were distinct (e.g., vacuum tube logic versus magnetic core memory) and by their intrinsic nature were physically separated. But now and for many years all three of these component categories are realized in semiconductor technology. In principle, they need not be divided, but due to legacy of the von Neumann model, are forced to be so to this day.

As a result, there are a number of inherent underlying assumptions intrinsic to commercial computing systems derived from the von Neumann model that forces poor design choices compared to what would be possible with contemporary semiconductor technologies through architecture structures and control semantics if not constrained by von Neumann precepts. A few but not all of these can be identified as the following:

1. FPU utilization is still prescribed as primary metric of efficiency dictated by von Neumann when the FPU is now a very small part of the core die area and power consumption. Yet the majority remaining area of the core architecture serves little other purpose than to keep the FPU busy; clearly a false objective function.
2. The processor core logic including the FPU and the main memory of the computing system are historically separated on different integrated circuits with intervening communication paths, constraining data access bandwidths, imposing significant data access latencies, increasing energy costs of data access, and demanding substantial interface and control logic for memory access. This is known as the "von Neumann bottleneck." With both logic and main memory both implemented with semiconductor technology, in principle the logic and memory could be mutually integrated on the same die largely eliminating these degradation factors.
3. To maintain the above requirement of running the ALU at fastest rate which is two orders of magnitude greater than DRAM, layers of cache memory taking up space and power are interposed based on the hope of temporal and spatial locality. As the caches are transparent to the user, this is supposed to make it easier to use. Ironically, for HPC the obvious is the case as the cache operation policies conflict with the needs of the user; yet, the user cannot control the cache.
4. The von Neumann model imposes sequential execution through the use of the special program counter buffer and incrementing of this instruction pointer. Yet, for high speed computing, parallel execution, not sequential control, is required. This is especially true with data flow or graph meta-data control that would expose greater amounts of parallelism. A machine architecture that is intrinsically parallel is needed for future scalable computing; not serial mechanisms forced through high overhead software to emulate parallel processing.
5. Sequential consistency is a policy implicit in von Neumann architecture-based computers that over constrain concurrency of memory access thus reducing parallelism and scalability. Various weak-consistency memory access models have been explored with some success but not large uniform adoption. As a result, substantial cache consistency mechanisms are

required to keep the potentially many copies of a data set to retain synchrony with costs in time, die space, and energy.

6. To further retain the appearance of sequential execution even with parallel operations, instruction level parallelism (ILP), execution pipelining, reservation stations, and speculative execution such as branch prediction all require additional and sometimes substantial mechanism hardware (and compiler instruction ordering software) to deal with these as well as pipeline hazard resolution. Die area, time, and energy are all sacrificed to a degree to assert von Neumann dictates.
7. Registers were an early addition and extension of the original von Neumann buffers to increase bandwidth and reduce latency to logic and are a mainstay of todays processor cores with both large sets of registers and sometimes multiple such sets. The registers are independent of the main memory and enforce a load/store protocol. Registers are so intuitively obvious that they are never questioned (except in GPU accelerators) and their presence assumed. Nonetheless, they are an additional cost in terms of the usual metrics.
8. Virtual address TLBs and page tables are a means to exhibit a uniform memory image required by the von Neumann model but across the complex main memory and secondary storage hierarchy. This method has never been sufficiently mastered in terms of time effectiveness but nonetheless is required with substantial costs in hardware mechanisms (not to mention OS services).

As a result of the above constraints and costs imposed by the von Neumann model and the now rapidly approaching end of Moore's Law and clock rates, future performance growth is most likely only achievable to any great degree through innovations of computer architecture. The often-successful use of GPU accelerators demonstrates one possible approach by augmenting conventional architectures with special purpose devices in a generalized approach referred to as heterogeneous computing. Where a particularly important algorithm with commercial market value can be accelerated through special purpose architectures such as the D.E. Shaw Anton architecture [21] optimized for molecular dynamics simulation. Here we propose a new non-von Neumann architecture that is more general than SPDs, addresses new market niches in graph processing, data analytics, machine learning, and other domains. The proposed architecture is based on its value to these customer problem domains by exploiting the described and fundamental disadvantages of the von Neumann architecture through an innovative class of physical structures combined with a dynamic adaptive strategy of parallel computing based on the ParalleX model. It is upon this significant innovation that we are launching our new business enterprise.

1.3. Continuum Computer non-von Neumann Architecture

The Continuum Computer architecture is a vast array of simple logic cells that are much smaller than todays lightweight cores (e.g., ARM [3]). Such structures are sometimes called "cellular," the earliest of which was devised by von Neumann himself in about 1950 when he proved some of them Turing equivalent [18]. Historically, complexity of operation has been achieved through concomitant complexity of design. Following this strategy would make it improbable that a small technical team such as SMART would be able to contribute significant innovation to the field. But another strategy of achieving complexity of operation through simplicity of design of such cells combined with highly replicated and tessellated structures of them permits two simple design steps: the design of a local simple cell (complexity of about 100K logic gate equiv-

alents) and the design of a global simple structure of replicated cells. This novel approach with respect to conventional practices makes the objectives of this project feasible within the scope of resources and period of performance. The global structure is as simple as adjacent connectivity to neighboring cells which is straight forward. To extend the system to multiple chip sockets again is accomplished through nearest neighbor socket to socket interconnects. Beyond Phase 1, a last level of interconnect between nearest neighbor PC boards will provide total design capacity of data storage active memory capacity and peak operational capability. The investigators are considering a second hierarchical network spanning the global system emphasizing low-latency lightweight messaging but for the active memory accelerator, prescribed target market, and customer segment this extensibility will be unnecessary and beyond the scope of the proposed work and product offering.

Continuum computer architecture based on the active memory system will engage ancillary functionality for support activities. Such services include bootstrapping, user program loading, host to accelerator I/O (in this case with PCI-Express), configuration on-board switches and output lights, and JTAG. These features will be supported by an FPGA controller, also to be designed, and integrated on the same PCB as the cell array chips. Not only does this controller provide flexibility in accelerator design, it permits debugging and upgrades of end user products. The FPGA controller is supported by an adjacent SRAM for available scratchpad memory and Flash NVRAM for FPGA controller internal configuration and system initialization. Associated with each cell array socket is a DRAM chip of service memory which is not part of the global virtual address space nor part of the user memory but rather dedicated local storage for copies of codes, tables, lightweight operating system support, overflow from active memory, and other secondary purposes. This service memory can be accessed by any cell on the specific socket.

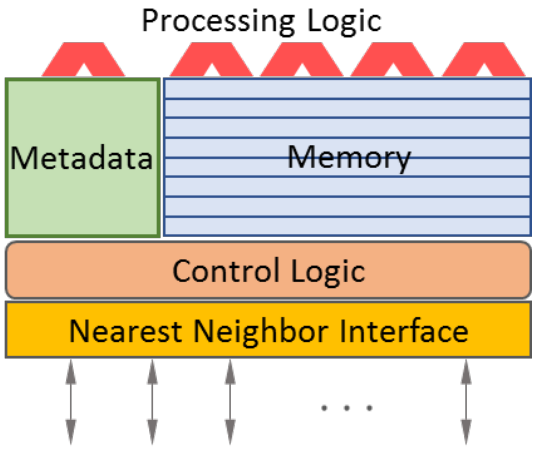


Figure 4. Generalized CCA cell architecture

The details of the architecture and design of a specific computing cell (Fig. 4), the primitive element making up the computing system, has yet to be determined and is under development by the current IARPA sponsored seedling project. However, the major properties of the component cell are understood even now, although the precise design trade-offs will be investigated. The single most important property to be realized is the integration of the three usually separate functions into a single function element. By merging data storage, data transformation, and data transfer into the unified computing element, the possible designs eliminate or significantly mitigate the legacy shortcomings of the conventional von Neumann architecture derivatives.

Critical among these is the bandwidth and latency between the active memory and logic which are essentially the same component. The data path is not word wide but line width to permit operations on multiple fields simultaneously. There are no separate registers but rather the block of multiple active memory lines. The technology used for active storage is a variation of SRAM not unlike that employed for level-2 caches. However, each block line includes some internal logic for primitive operations that can be performed in parallel. Examples include shift and roll operations, increment/decrement over masked ranges, GF-2 [12] for government customers, associative searches, and necessary primitives for dynamic graph processing.

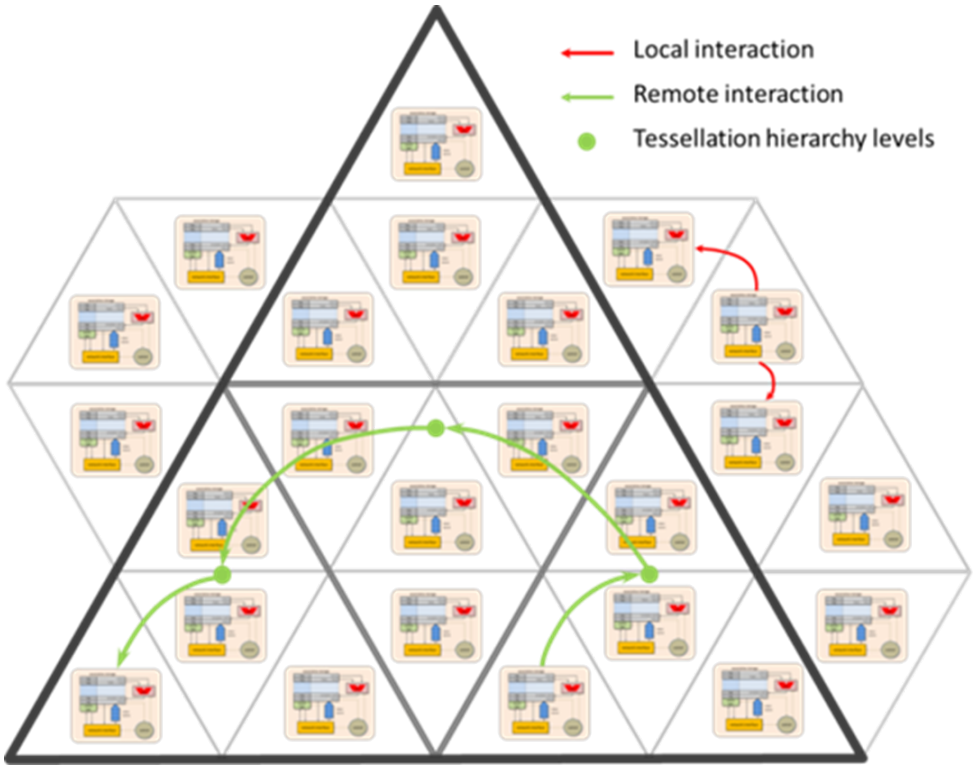


Figure 5. Example tessellation of silicon die area implementing cell array

The second class of functionality integrated in the computing cell is the combination of addressing and routing for virtual address management and data transfers. This is particularly important for rapid pointer tracing for parallel graph processing which discovers and exploits meta-data parallelism. Communication is primarily nearest neighbor. The cell to be designed is triangular (Fig. 5) for simplicity and to minimize hardware communication overhead with respect to storage and logic. It also provides very tight packing on the silicon die. But several other geometries are possible as well with different trade-offs.

The third class of functionality is the logical operations that can be performed in a single cycle within the cell. Some of these were briefly mentioned above. Where the logic is more complex than that of an accumulator, some extra logic is shared among block lines while still being relatively straight forward. But complex operations including floating point require more components than is anticipated for any single cell. In conventional cores, an FPU is a small part of the real estate. Cells are so much smaller by intent that an FPU would consume a significant part of the die. Instead, several cells in a contiguous pattern through pipelining implement such complex operations. Many other complex operations fall into this category as well. Beyond complex operations are compound functions represented by compute complexes of the ParalleX

model. Again, contiguous assemblies of cells with instructions interspersed with data provide locality, atomic operation sequences, and open-ended flexibility of functionality. Such structures also permit graph tracing.

1.4. Graceful Degradation

The ability to achieve fault tolerance, that is non-stop computing, is challenging and complicated, rarely achieved, with the fall back of venerable checkpoint-restart often applied for large scale systems and applications. For special purpose cases such as systems control (e.g., flight control), a multiplicity of systems running the same codes are employed concurrently to continue operation even in the presence of one failing providing both error detection and continued operation. But this is an expensive brute force approach aggravating cost, power, size, and weight. The proposed target market, that of crypto-currency, suffers from failures that result in potentially significant profit reductions due to periodic down times of forced maintenance with the need for the replacement of component modules prior to operation again.

Full fault tolerance would require a number of steps available and automatically performed by the hardware software system. These include error detection, failure diagnosis, replacement of component, roll back of computation to known good point, restart. Slightly different methodologies are also employed. But for some computations and systems such as the one being proposed a simpler and less costly methodology is sufficient for the customer purposes. Graceful degradation (not a new concept) incorporates means by which a systems capabilities diminish over time with incurred hardware faults but continues to operate, albeit at reduced performance. Such techniques can be made more complex due to the need to safely restart a partially completed task. However, in the case where a myriad of disconnected tasks is being performed such as transaction systems (e.g., a Google query), the task associated with the failed device can be dropped, simplifying the recovery.

The failed device must be isolated from the remainder of the structure. For many conventional architectures, this is either not possible or the granularity of the failed module is so large that the degree of performance diminution is too substantial and costly. But in the case of the proposed non-von Neumann architectures the natural module to isolate is the cell in which the failure has occurred. The design to be tested will include the means to detect the error event and trigger the isolation action. For any single such resulting hole in the array of computing cells, dynamic adaptive routing enables traffic to move around the broken cell for other tasks. Simple tests are to be incorporated in the cell logic design including some well-known techniques. Additional tests will come from the adjacent cells such as heartbeat and other normal behaviors being at variance with anticipated operation. There are special cases including particular patterns of failed cells in which working components are no longer useful. Such conditions decrease performance at worse than linear rates. But with potentially tens of thousands of cells per socket, the loss of any one is unlikely to be significant. Graceful degradation will allow continuous operation of the system without costly experience of downtime. This provides an important value proposition for the crypto-currency market and many others.

1.5. Detailed Value Achieved through Innovations

Many advantages are achieved through the innovative non-von Neumann architecture described above with additional ones not yet discussed. Here is summarized many of the specific values garnered through this unique structure and semantics:

- Reliability is achieved through turning off cells that are defective and using adaptive routing to work around these holes. Graceful degradation avoids many single point failure modes dramatically reducing or even eliminating the need for checkpoint-restart overheads.
- Higher yield to lower cost of chip manufacturer by up to 2× which also reduces testing cost for 1.5× customer price reduction. This results from the same value above where the random faults encountered in die fabrication are simply cut off from the rest of the chip and adaptively routed around. This will make almost all dies of the wafer usable, thus reducing cost of production and price to customer.
- Increase of computational throughput normalized by size compared to conventional practices by 100×, at least peak performance, results from the elimination of most of the core die area of typical chips allowing far greater logic density and resulting number of cells served by much greater memory bandwidth.
- Enabling asynchronous graph processing with built in synchronization, futures in vertices.
- Enabling scalable parallel graph processing with hardware support for direct manipulation of meta-data (pointer type) to discover, expose, and exploit parallelism and convert serial issue to parallel (logarithmic) issue from source vertex to many destination vertices.
- Eliminate distance barrier between data storage (formerly main memory) and data transformation (formerly separated ALU/FPU logic) by merging the two (PIM [7] is another variant) to minimize latency between state and logic (von Neumann bottleneck).
- Power reduction of 95% normalized to throughput.
- Active memory replaces main memory, scratch pad memory, caches, and registers. Simplified programming model and greater portability.
- ALUs emphasize availability, not utilization. Greatly enhances overall efficiency and programmability.
- Maximize local active memory bandwidth through merger of logic and storage as well as wide word (struct) processing. Lower power, faster processing, greater efficiency. Customer applications good even with low data reuse enabling to do different types of science that were less well suited to conventional methods.
- Maximize local bandwidth through nearest neighbor direct point to point adjacency interconnect and between sockets to lower communication power and increases efficiency for problems like using stencils.
- ParalleX dynamic adaptive execution model for scalability and simplification of programming.
- Global address space.
- Message driven computation.
- Futures synchronization.
- Locality management.
- Ease of adoption through industry grade standard hardware and software interfaces; transparency to user.

2. Scaling Opportunities derived from CCA Class Systems

2.1. CCA as a non von Neumann Strategy

The true importance of the opportunities delivered by the CCA class of non-von Neumann systems is in the dual properties of efficiency and scalability. At one time, it was conventional practice to measure efficiency in terms of FPU utilization; that is the ratio of sustained floating-point performance to peak performance. This had been extended to use the HPL benchmark as the metric of the sustained performance. This was a favorable parameter to industry vendors of large scale systems whose designs optimized their core and system architectures to local floating-point operations. But other measures of efficiency are far more important as enabling technologies follow current trends. These emphasize different objective functions; among these: cost, energy, availability, and size. The ParalleX model was based on the intent to use runtime information and control to make better use of resources, thus exhibiting superior efficiency than compiler-driven methods alone. It was seen to achieve this in certain cases such as computations that embrace dynamic adaptive methods such as adaptive mesh refinement or finite multiple methods. But in other cases, static methods proved to be adequate, even superior depending on details of the runtime implementation. Key to enhancing effectiveness of runtime techniques was the reduction or elimination of parallel control overheads through hardware mechanisms. New architectures, including CCA, expose such opportunities. More broadly, for a given problem efficiency is a function of the time to solution and the cost including deployment and operational. Within this cost is energy. What becomes clear is that utilization is not a direct indicator of efficiency. Heterogeneous computing alternatives exploit this fact where overall utilization may be lower but time to solution is also shorter.

2.2. Principles of CCA Scaling

Scalability is a higher impact property of computing and one that is severely challenged by the end of Moore's Law. Even improvements in efficiency are bounded by the roofline of any machine. Only scalability can open the pathways to unprecedented capability and the value this brings to myriad domains of exploration. But with tens of millions of processor cores and the supporting infrastructure required to constitute a working system anywhere near an exascale platform, the density of processing itself is restricted by technology trends. This statement is only true if it is assumed, as it has always been in the commercial market, that the future supercomputers are comprised of von Neumann cores. The fundamental premise of this paper is that they are not. In the limit for a finite space of (e.g., die area), as the size of a computing cell is diminished, the number of the cells increases, in theory to infinity with the capability of such cells asymptotically approaching zero for a zero-area computing cell. It is when a finite area of infinitesimal computing cells is combined that measurable work is performed. So, what's the point? It seems like it's just a different way to divide up the working resources. True. But the potential impact and value is derived by viewing this from the perspective of SLOWER. As the cells diminish in size several things happen. First, the distance from bit storage to data transformation logic diminishes, narrowing the latency gap. Second, as more of the bit storage is directly exposed to the data transformation logic, the aggregate system wide bandwidth increases reducing contention. Third, reliability is enhanced because the loss of any one computing cell at a time is a tiny portion of the total capability and graceful degradation can be employed to maintain operation albeit at a slightly diminished capacity. This greatly, if not perfectly,

expands availability and reliability. Fourth, system bandwidth dramatically grows with every cell having immediate access to state of its adjacent neighboring computing cells. Fifth, energy is reduced as only those cells undertaking constructive work during any single system cycle has to consume energy; this could be considered a form of active dark silicon. Sixth, there is a marked change to the creation of such system from the common practice of achieving complexity of operation through complexity of design of processor cores to the alternative of achieving complexity of operation through simplicity of design of computing cells and their high replication. A consequence is that a computing cell can be developed by a small group of engineers in a small amount of time and cost less for the NRE. It is these and other properties that determine the nature and value of the class of "continuum computer architecture."

2.3. CCA Scaling Analysis

The details of the internal design of a computing cell are not presented in this paper. There are many possible solutions and their variations to this problem and will be the subject of future disclosure and discussion. But such specifics are not needed to explore the remarkable potential scalability of CCA systems even within the scope of conventional semiconductor enabling technologies and their immediate extensions in the nano-scale regime. The following analysis does not presume any inherent breakthroughs than those already discussed which is largely a change of perspective and frankly a willingness to consider a design space beyond that of legacy approaches. Ironically, once the transformative principles of CCA are embraced, the scalability of systems that are so derived is essentially one of incrementalism. This is illustrative in the typical form of a hierarchy of system components from individual computing cells, chips containing on the order of 10,000 such cells, one or more such chips in a socket, a multiplicity of sockets on a 16" by 10" PC board, one to four boards to a module. For sake of clarity and as determined from design studies, the range of modules structures can all fit within a 1-U rack mounted package using air-cooled with one to two boards and liquid cooling for three to four boards. The module can be used as an attached accelerator to a conventional user SMP system or as one of a plethora of units interconnected by a system area network (Fig. 6).

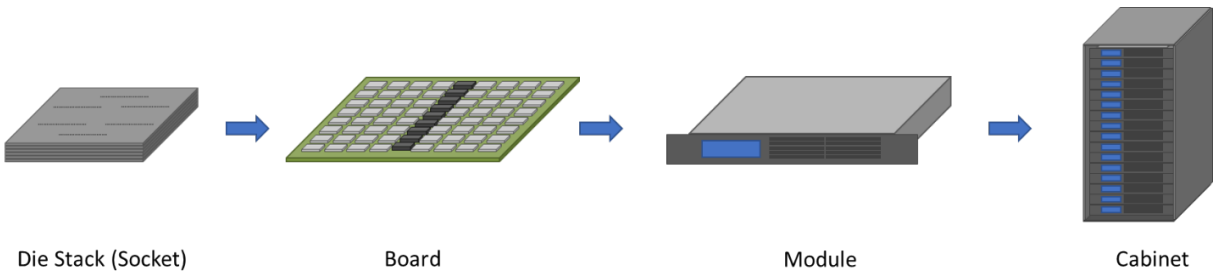


Figure 6. System capability scaling

Using available technologies from which to construct nodes, part densities, capabilities, and power are determined. It is estimated that a computing cell as described in previous sections will comprise 100 to 200 thousand transistors.

System structures capacities are shown in Tab. 1 from the single chip level to a full up exascale system which requires 1K modules in 24 racks. Table 2 presents the details of the projected capabilities of the same units of the system in terms of peak performance which start at about 16 Mflops per computing cell to 1 Exaflops per large scale system. An important

Table 1. System component counts

Level	Cells	Chips	Sockets	Boards	Modules
Chip	11 k	1	-	-	-
Socket	43.8 k	4	1	-	-
Board	16.3 M	1,488	372	1	-
Module	65.2 M	5,952	1,488	4	1
Rack	2.74 G	250 k	62.5 k	168	42
System	65.7 G	6 M	1.5 M	4,032	1,008

milestone is the module of 1 Petaflops of peak performance (liquid cooled) from which many different configurations are possible in scale both as active memory accelerators and as full self-hosted standalone supercomputers. A rendering of an Exaflops computer is shown in Fig. 7 1 as a cylindrical configuration of conventional industrial grade racks. An important metric is footprint for any supercomputer. This system, using only available enabling technologies, requires less than 400 square feet which is almost two orders of magnitude smaller than using common practices. Although of small interest, for practical reasons weight is another consideration. A fully populated rack comes in just under the industry standard limits of 1,200 Kilograms in a foot print of less than 1 square meter. The total Exaflops system is about 30,000 Kilograms.

Table 2. Capacities of system components

Level	Peak OPS	Peak Flops	Active memory [bytes]	DRAM [bytes]	Peak memory bandwidth [bytes/s]	Peak neighbor bandwidth [bytes/s]
Cell	128 M	16 M	1 k	-	3.07 G	4.1 G
Chip	1.4 T	175 G	11.2 M	-	33.6 T	44.9 T
Socket	5.61 T	701 G	44.9 M	1.07 G	135 T	179 T
Board	2.09 P	261 T	16.7 G	399 G	50.1 P	66.8 P
Module	8.35 P	1.04 P	66.8 G	1.6 T	200 P	267 P
Rack	351 P	43.8 P	2.8 T	67.1 T	8.41 E	11.2 E
System	8.41 E	1.05 P	67.3 T	1.61 P	202 E	269 E

2.4. Power Consumption

A practical CCA system will take advantage of power mitigation solutions provided by the state-of-the-art high-density CMOS fabrication technologies. Among those, application of FinFET [8] transistors appears to be the preferred approach for leakage current reduction over thin-body Silicon-On-Insulator [23] implementations, despite requiring a more involved lithography process. Assuming geometries of 14 nm Global Foundries FinFET process [24] and typical power derating factors for random data bit distribution and estimates of fraction of actively switching logic per cycle, a single CCA die is expected to dissipate approximately 1.2 Watts. In this figure, roughly a quarter is spent to power the processing logic, less than a quarter attributed to leakage, and the remainder used for data movement between memory and processing structures inside the cells as well as for nearest-neighbor and off-chip communication. Taking

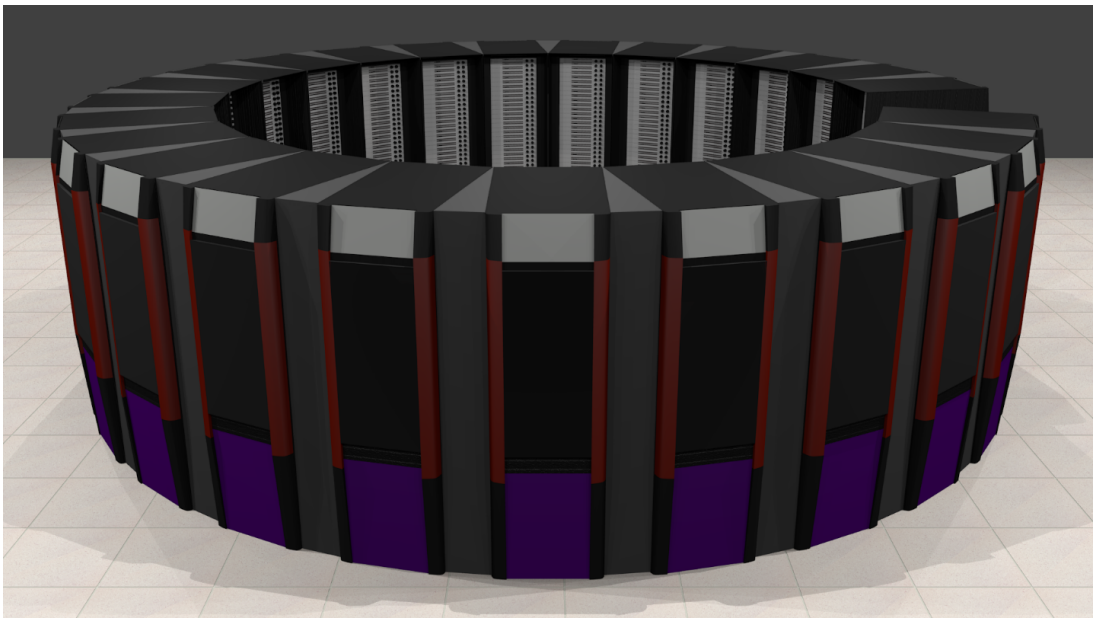


Figure 7. Rendering of an Exaflops system composed of 24 racks in a circular arrangement. A rack-wide opening is left for maintenance access to the back of cabinets

into account the die count in the Petaflops module and additional power requirements of the system level interconnect, module's power draw is estimated at 7.2 kW, fully populated rack's at 316 kW, and the exascale system at 7.6 MW.

3. Scaling to Zettaflops

With the advent of the venerable (by computer years) TaihuLight and the genesis of the Summit 200 Petaflops system, it is anticipated that extensions of current methodologies will empower the eventual deployment of exascale computing by the beginning of the next decade. There is a major international thrust by many countries to do just that. With the will to invest hundreds of millions of dollars, large machine rooms, and large power consumption, the exascale era is upon us. Adopting an entirely new form and function of computing concept, even if superior by key metrics, is probably unjustifiable if it were targeted to this singular operating point. But such is not the case. A set of incremental advances is identified that in some, as yet to be identified order, can traverse the pan-exascale performance regime. It appears this can be accomplished in less than a decade with full deployment of Zettaflops scale systems delivered on or before 2027. The following is a brief description of these possible steps beyond the analysis for 1 Exaflops CCA.

1. Switching from a circular foot-pad to the conventional rows of racks and doubling in both length and width will yield a 4× peak gain.
2. The original clock rate used for all analysis was 128 MegaHertz. It is planned that in incremental steps of doubling that the final clock rate will be the modest 500 MHz for another 4× is peak speed.
3. Chip stacking [4] is a well understood fabrication technology of socket form. The 1 Exaflops system already presumes 4-dies per socket. The expansion of the CCA to system will ride the advance of this packaging technique to 8-dies per socket for another 2× peak performance gain.

4. Over the span up to 2027 and perhaps significantly before, the path to nanoscale (and sadly the end of Moore's Law) will likely permit feature sizes to move from 16 nm which the original exascale analysis used to 8 nm which is already working in the laboratories of semiconductor manufacturers. This provides a doubling of density in chip length and width for a 4× gain in number of computing cells per die and peak performance gain.
5. ILP (instruction level parallelism) was ignored for the original analysis assuming that a computing cell either does nothing during a particular cycle or does one action. But the local structure permits many micro-operations per cycle and a conservative estimate of performance advantage through responsible ILP will deliver at least a 4× peak performance gain.
6. An assumed $1\,\mathrm{cm}^2$ die size was employed throughout the exascale CCA analysis in the preceding sections. But this is limiting. The die even in todays technology could be twice the size in each dimension for a 2 cm by 2 cm die. This does not increase the cell density. But it does have a secondary effect of improving the PC board area utilization by 50% resulting in a 1.5× performance increase.
7. The last technique to extend system capability is to break with one of the long-held assumptions about digital computing, that its hardware is base-2. In fact, Boolean logic is not restricted to base-2 and it is possible to build circuits in memory and logic to work at base-4 (and above). It is recognized that this is a radical departure from the norm. But it is entirely feasible saving space, time, and energy. It is included here to achieve one last doubling for another 2× of peak performance.

In each case, the factors of gain reflect various degrees of risk from trivial and obvious to more speculative. But these seven dimensions in the design space if combined and acknowledging that this is entirely in terms of peak performance will provide a roadmap to a gain of 1,500× which is equivalent to Zettaflops. This does not constitute a detailed analysis but only a framework for considering the future roadmap for CCA class systems in the next decade. But it is sufficient to have confidence in a future path.

There is one final point that while speculative in the extreme, is nonetheless technically viable with more than two decades of experience in the basic physics and design. If the assumption that conventional semiconductor devices is relaxed and an alternative is considered in its place, there is an opportunity to be pursued and applied. It is superconducting technology. Admittedly, this is at best an acquired taste and for most of the last two to three decades has not received wide attention or use. SFQ or single flux quantum gates [13] (of which there are multiple designs) stores state in SQUIDs (superconducting quantum interference devices) or loops of two Josephson Junctions (JJ) [10] as a constant flux. Researchers have developed entire logic circuit families around this simple well-known component. SFQ has two exceptionally important properties that may dramatically advance supercomputing if it were to be applied. The first is its speed. Although most ongoing work seems to be limited to a mere 50 GHz or so, the best in breed laboratory experiments have shown capabilities of more than 700 GHz. The second property is its energy-free operation. It is able to operate many orders of magnitude less power than conventional electronics. Both properties can be exceptionally important to supercomputing progress. With the expected optimization of CCA designs beyond that already presented and the implementation of the computing cell with SFQ rather than semiconductor logic, the possibility of achieving Yottaflops peak performance is not beyond reach. If work were started this year, we could hit Yottaflops by 2030.

Conclusions

This paper is an early presentation on the concept of Continuum Computer Architecture, a family of non-von Neumann architectures that relax many of the constraints and bottlenecks of conventional computers. This treatise projects the possibility of exascale computer platforms much less expensive, and lower power than extrapolations of conventional practices. The combination of the ParalleX execution model and the CCA is considered to enable pan-exascale performance through practical evolution in the near future. It is concluded from this study that Zettaflops computing is conceivable within the next decade.

References

1. Top 500 performance development (Jun 2018), `https://www.top500.org/statistics/perfdevel/`, accessed: 2018-07-17

2. Top 500. The List. (Jun 2018), `https://www.top500.org/`, accessed: 2018-07-17

3. ARM Holdings Architecture Group: ARMv8 Instruction Set Overview (Nov 2011), PRD03-GENC-010197

4. Black, B., et al.: Die stacking (3D) microarchitecture. In: Proceedings of the 39^{th} Annual IEEE/ACM International Symposium on Microarchitecture, MICRO'06. pp. 469–479 (December 2006), DOI:10.1109/MICRO.2006.18

5. Dongarra, J.: Performance of various computers using standard linear equations software. Tech Report CS-89-85, University of Tennesse Computer Science (2014), `http://www.netlib.org/benchmark/performance.pdf`, accessed: 2017-07-17

6. Gao, G.R., Sterling, T., Stevens, R., Hereld, M., Zhu, W.: Parallex: A study of a new parallel computation model. In: 2007 IEEE International Parallel and Distributed Processing Symposium. pp. 1–6 (March 2007), DOI: 10.1109/IPDPS.2007.370484

7. Hall, M., Kogge, P., Koller, J., Diniz, P., Chame, J., Draper, J., LaCoss, J., Granacki, J., Brockman, J., Srivastava, A., Athas, W., Freeh, V., Shin, J., Park, J.: Mapping irregular applications to DIVA, a PIM-based data-intensive architecture. In: Proceedings of the 1999 ACM/IEEE Conference on Supercomputing. SC '99, ACM, New York, NY, USA (1999), DOI: 10.1145/331532.331589

8. Hisamoto, D., Lee, W.C., Kedzierski, J., Takeuchi, H., Asano, K., Kuo, C., Anderson, E., King, T.J., Bokor, J., Hu, C.: FinFET – a self-aligned double-gate MOSFET scalable to 20 nm. IEEE Transactions on Electron Devices 47(12), 2320–2325 (Dec 2000), DOI: 10.1109/16.887014

9. Hoare, C.A.R.: Communicating sequential processes. Commun. ACM 21(8), 666–677 (Aug 1978), DOI: 10.1145/359576.359585

10. Josephson, B.: Possible new effects in superconductive tunnelling. Physics Letters 1(7), 251–253 (1962), DOI: 10.1016/0031-9163(62)91369-0

11. Kaiser, H., Brodowicz, M., Sterling, T.: ParalleX an advanced parallel execution model for scaling-impaired applications. In: Parallel Processing Workshops, 2009. ICPPW '09. International Conference on. pp. 394–401 (Sep 2009), DOI: 10.1109/ICPPW.2009.14

12. Lidl, R., Niederreiter, H.: Finite Fields. Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2 edn. (1996), DOI: 10.1017/CBO9780511525926

13. Likharev, K.K., Semenov, V.K.: RSFQ logic/memory family: a new Josephson-junction technology for sub-terahertz-clock-frequency digital systems. IEEE Transactions on Applied Superconductivity 1(1), 3–28 (March 1991), DOI: 10.1109/77.80745

14. Modha, D.S.: Brain-inspired computing. In: 2015 International Conference on Parallel Architecture and Compilation (PACT). pp. 253–253 (Oct 2015), DOI: 10.1109/PACT.2015.49

15. Monroe, D.: Neuromorphic computing gets ready for the (really) big time. Comm. ACM 57(6), 13–15 (Jun 2014), DOI: 10.1145/2601069

16. Moore, G.E.: Cramming more components onto integrated circuits. Electronics pp. 33–35 (Apr 1965), DOI: 10.1109/N-SSC.2006.4785860

17. von Neumann, J.: First draft of a report on the EDVAC. Tech. rep., Moore School of Electrical Engineering, University of Pennsylvania (Jun 1945)

18. von Neumann, J., Taub, A.W., Taub, A.H.: The Collected Works of John von Neumann: 6-Volume Set. Reader's Digest Young Families (1963)

19. Oak Ridge National Laboratory: Summit: America's newest and smartest supercomputer (2018), `https://www.olcf.ornl.gov/summit/`, accessed: 2018-07-17

20. Schumacher, B.: Quantum coding. PhysRevA 51, 2738–2747 (Apr 1995), DOI: 10.1103/PhysRevA.51.2738

21. Shaw, D.E., Deneroff, M.M., Dror, R.O., Kuskin, J.S., Larson, R.H., Salmon, J.K., Young, C., Batson, B., Bowers, K.J., Chao, J.C., Eastwood, M.P., Gagliardo, J., Grossman, J.P., Ho, C.R., Ierardi, D.J., Kolossváry, I., Klepeis, J.L., Layman, T., McLeavey, C., Moraes, M.A., Mueller, R., Priest, E.C., Shan, Y., Spengler, J., Theobald, M., Towles, B., Wang, S.C.: Anton, a special-purpose machine for molecular dynamics simulation. Commun. ACM 51(7), 91–97 (Jul 2008), DOI: 10.1145/1364782.1364802

22. Sterling, T., Kogler, D., Anderson, M., Brodowicz, M.: SLOWER: A performance model for exascale computing. Supercomputing frontiers and innovations 1(2), 42–57 (2014), DOI: 10.14529/jsfi140203

23. Stern, J.M., Ivey, P.A., Davidson, S., Walker, S.N.: Silicon-on-insulator (SOI): A high performance ASIC technology. In: 1992 Proceedings of the IEEE Custom Integrated Circuits Conference. pp. 9.2.1–9.2.4 (May 1992), DOI: 10.1109/CICC.1992.591170

24. WikiChip: 14 nm lithography process, `https://en.wikichip.org/wiki/14_nm_lithography_process`, accessed: 2018-07-17

DOI: 10.14529/jsfi180302

Three-dimensional Inversion of Electromagnetic Geophysical Data with Parallel Computational Code on Supercomputer Complex "Lomonosov"

Sergey V. Zaytsev[1,2]***, Viktor A. Kulikov***[1,2]***, Andrei G. Yakovlev***[1,2]***, Denis V. Yakovlev***[2]

Usage of 2D inversion of magnetotelluric data for real geological objects can cause distortion, but it is more often used in commercial projects, because of its effectiveness and great experience. Whereas in the case of 3D inversion is not such a great experience and there are a number of global problems. When switching to 3D inversion of MT data, the requirement for computer technology is significantly increased. In this paper we will discuss a few examples of 3D inversion of electromagnetic geophysical field data with the usage of "Lomonosov" supercomputer and show its effectiveness on several geological objects. Each object is associated with a variety of problems: from search for shallow ore to regional hydrocarbon exploration. But all these objects contain a large volume of measurements obtaining qualitative results for which requires a huge amount of time. So that the use of 3D inversion with a high-performance computational complex makes it possible to obtain a qualitative result of solving a wide range of problems.

Keywords: magnetotelluric, three-dimensional inversion, supercomputer, geophysics.

Introduction

In most cases, two-dimensional (2D) inversion of magnetotelluric (MT) data is used to solve practical problems and three-dimensional (3D) inversion is not frequently used in commercial projects. This is due to the fact that most of the commercial studies are done on single profiles, and with already well-developed 2D inversion algorithms, what allows you to get results on a very fine grid. There is some skepticism in commercial companies associated with the use of 3D inversion on real geological objects. Is it possible to get a good resolution of the resulting model, which would suit interpreters? Is the time spent on 3D inversion comparable with the time for which the 2D inversion will be executed? Where are the limits of the applicability of the 3D inversion, in terms of solved problems?

To answer these questions, it is necessary to accumulate a base for solving the inverse problem on real objects, which would demonstrate the effectiveness of the 3D inversion and its capabilities. When selecting such objects, several rules should be considered:

1. The ability to check the results and compare them with other methods or geology;
2. The objects should be aimed at solving various problems for understanding the possible limitations of the applied technology.

For the study, two objects were selected that satisfy these conditions: an MT project on shallow ore deposits (Bystrynskoe field) and an Oil & Gas project in Bolivia. They were chosen from different areas and aimed at solving various problems. We will try to show the effectiveness and possibilities of applying 3D inversion. The article is organized as follows. Section 1 is devoted to methods that were used. In section 2, we describe the implementation of 3D inversion on the ore deposit. Section 3 contains our regional study in Bolivia aimed at hydrocarbon deposits. Conclusion summarizes the study.

[1]Faculty of Geology, Moscow State University, Moscow, Russia
[2]Nord-West Ltd., Moscow, Russia

1. Methods

In this paper we are solving regularized EM inverse problems, i.e. minimization of a penalty functional of the form:

$$\Phi(m,d) = (d - f(m))^T C_d^{-1}(d - f(m)) + v(m - m_0)^T C_m^{-1}(m - m_0) \quad (1)$$

to recover an Earth conductivity model parameter vector m, which provides fit to a data vector d. In (1), C_d is the covariance of data errors, $f(m)$ defines the forward mapping, m_0 is a prior or first guess model parameter, v is a trade-off parameter, and C_m defines the model covariance or regularization term. The forward mapping requires solution of the frequency domain EM partial differential equation – finding electrical and magnetic fields with known conductivity of model with finite-difference methods $f(m) = d$. In MT methods forward problem is calculated independently for two modes of electrical field polarization: TE and TM modes. So, for both modes we need to calculate two electrical fields for each frequency and project it on each site. Then transform to magnetic field $h = Te$, where T is transform operator. Penalty functional can be directly minimized using a gradient-based optimization algorithm such as non-linear conjugate gradient (NLCG). The NLCG approach requires calculating gradient of (1) with respect to m:

$$\frac{d\Phi}{dm} = -2J^T r + 2vm_n. \quad (2)$$

In equation (2) J is sensivity matrix which depends on model parameters and field data. So in each iteration to solve inverse problem (finding conductivity of the earth with known field data) $m = f^{-1}(d)$ it is necessary to calculate a forward problem $f(m)$ and sensivity matrix J. To control results of inversion we use nRMS – normalized on error floor root-mean-square deviation.

For our tasks we use ModEM code [1] to provide 3D inversion. The main advantage of the ModEM code is the ability to use parallel approaches to the calculation of forward and inverse problems using the Message Passing Interface (MPI) technology, which significantly reduces the time spent on finding the solutions of the inverse and forward problem, since the forward problem for each period is calculated independently.

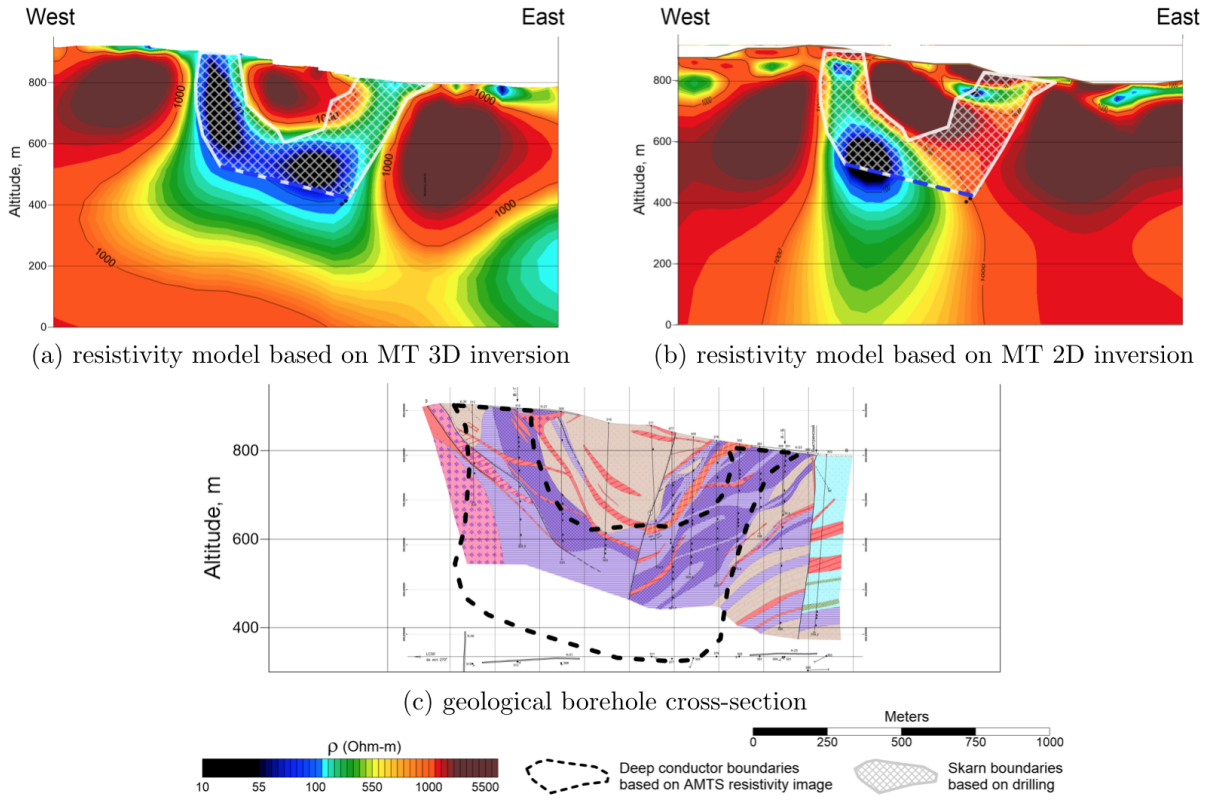
Another important segment of the study is application of the supercomputer complex "Lomonosov" [3]. Its peak performance is 2,962.3 TFlop/s, which is the No. 1 of Russia for today. A more important parameter for computing on a supercomputer for us is the number of CPUs used. Since on all investigated objects the number of periods was no more than 30, all calculations were performed on 64 CPUs.

The standard approach for determining the weights of data under inversion was used: percentage of $\sqrt{Z_{xy} * Z_{yx}}$. Therefore, in the description below, the method for determining the weights will be omitted.

Following article [2] the smoothing was determined in the ModEM as the covariance matrix selected in the directions $X = Y = 0.3$ and $Z = 0.4$. All starting models were homogeneous half space with resistivity changes depending on prior information. In this paper, only the final results are demonstrated. Of course, the choice of the "most successful" result depends on the experience of the geophysicist. We tried different approaches and show what we think to be the best result.

2. Bystrinskoe Skarn Cu-Fe Ore Deposit, Russia

The first example is the application of a 3D inversion for a shallow mineral exploration in the Bystrinskoye ore field. This area was chosen for having a well-known geological structure, which allows evaluating the quality of the results obtained. The inversion was carried out for 546 stations. Therefore, only the main components of the impedance tensor were inverted. The error floor of the main components was chosen as 5%. The horizontal size of the grid was 50x50 m, which allows obtaining a good horizontal resolution. Since the task consisted of detecting bodies at depths of the first hundreds of meters, the vertical size of the first cell was 10 meters. Then it increased with a coefficient of 1.1 to 1.5 km, then to 40 km with a coefficient of 2. Total size of the grid is 60x85x40 cells. It took 46 hours to calculate 51 iterations with misfit (nRMS) 1.87. The comparison of the resistivity models (Fig. 1) obtained on the MT inversion and drilling results indicate that ore boundaries can be determined with high resolution right up to a depth of 800–1000m but on 2D inversion result we cannot successfully determine the ore body. Successful application of MT methods potentially may reduce the amount of closely spaced expensive drilling.



(a) resistivity model based on MT 3D inversion

(b) resistivity model based on MT 2D inversion

(c) geological borehole cross-section

Figure 1. Comparison of MT inversion result and borehole data along the AMTS Line No. 1

3. Oil & Gas Regional Studies, Bolivia

A more complicated example of 3D inversion is a regional research aimed at finding oil and gas traps in Bolivian mountain, as well as clarifying the deep geological structure of the region. Figure 2 shows a comparison of the results of 3D inversion and 2D inversion in the area with strong 3D distortions in the deeper part at the profile. At depths greater than 4 km, it can be

seen that the 2D inversion "drags" the conductor to a depth, whereas the 3D inversion clearly restores the position of the conducting syncline. To obtain a qualitative result (nRMS < 1.8) of a 3D inversion on 360 stations for 21 periods, 49 hours of calculations were required. At the same time, 2D inversion was performed in several stages for each profile to eliminate 3D distortions and took more than a few days.

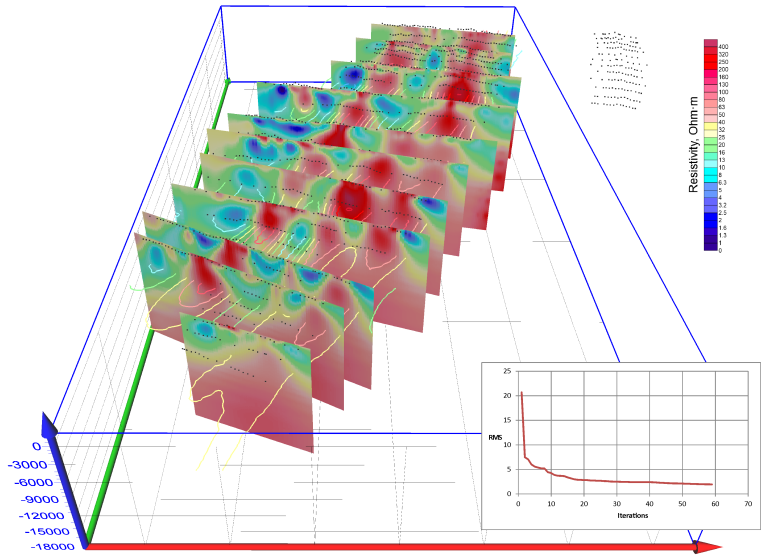


Figure 2. Result of 3D inversion of Bolivian Subandian project (depth scale in meters); on the right shown RMS vs Iteration graph

Conclusions

The performed work shows that with the use of modern 3D inversion algorithms and high-performance computing system, it is possible for a short period of time achieve acceptable results for a huge amount of MT stations. We tried to find the universal parameters of 3D inversion for each object and there was a good correlation with a geological priori data for different types of projects. The effectiveness of using 3D inversion and good correlation of the obtained results with known geology were demonstrated on field examples.

Acknowledgements

I express my gratitude to the staff of the company Nord-West for the data provided. We used a licensed copy of the ModEM software provided by Nord-West Ltd. The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011.

References

1. Kelbert, A., Meqbel, N., Egbert, G.D., Tandon, K.: Modem: A modular system for inversion of electromagnetic geophysical data. Computers & Geosciences 66, 40–53 (2014), DOI: 10.1016/j.cageo.2014.01.010

2. Miensopust, M.P.: Application of 3-d electromagnetic inversion in practice: Challenges, pitfalls and solution approaches. Surveys in Geophysics 38(5), 869–933 (2017), DOI: 10.1007/s10712-017-9435-1

3. Sadovnichy, V., Tikhonravov, A., Voevodin, Vl., Opanasenko, V.: Lomonosov: Supercomputing at Moscow State University. In: Contemporary High Performance Computing, pp. 283–307. Chapman and Hall/CRC (2013)

DOI: 10.14529/jsfi180303

Simulating the Long-timescale Structural Behavior of Bacterial and Influenza Neuraminidases with Different HPC Resources

Yana A. Sharapova[1], *Dmitry A. Suplatov*[2], *Vytas K. Švedas*[1,2]

Understanding the conformational dynamics which affects ligand binding by Neuraminidases is needed to improve the *in silico* selection of novel drug candidates targeting these pathogenicity factors and to adequately estimate the efficacy of potential drugs. Conventional molecular dynamics (MD) is a powerful tool to study conformational sampling, drug-target recognition and binding, but requires significant computational effort to reach timescales relevant for biology. In this work the advances in a computer power and specialized architectures were evaluated at simulating long MD trajectories of the structural behavior of Neuraminidases. We conclude that modern GPU accelerators enable calculations at the timescales that would previously have been intractable, providing routine access to microsecond-long trajectories in a daily laboratory practice. This opens an opportunity to move away from the "static" affinity-driven strategies in drug design towards a deeper understanding of ligand-specific conformational adaptation of target sites in protein structures, leading to a better selection of efficient drug candidates *in silico*. However, the performance of modern GPUs is yet far behind the deeply-specialized supercomputers co-designed for MD. Further development of affordable specialized architectures is needed to move towards the much-desired millisecond timescale to simulate large proteins at a daily routine.

Keywords: neuraminidases, molecular dynamics, long-timescale trajectories, GPU, co-design.

Introduction

Co-infection of the human lower respiratory tract with viruses and pathogenic bacteria can lead to life-threatening complications to influenza including pneumonia. The mechanism that facilitates this synergism is a major focus of scientific research, and thus Neuraminidases, which are key virulence factors of both pathogens, are considered as important drug targets. Despite intensive studies, no inhibitors targeting bacterial Neuraminidases have been approved for clinical use so far. Four drugs targeting the influenza Neuraminidases are available, but their efficacy is challenged by evolution of the pathogen and emerging resistance [1]. It is therefore important to design new drugs before the currently available ones become useless. Probably, the cause for the poor success rate of delivering approved drugs to the market is that drug discovery is usually driven by optimizing the binding affinity and selectivity of the respective drug candidate in a static crystallographic structure of a target protein. Previous studies indicated that structures of bacterial and influenza Neuraminidases contain flexible regions which can be involved in a conformation-specific accommodation of ligands [2, 5, 7]. A deeper understanding of conformational flexibility and dynamics in Neuraminidases that can affect the binding of ligands, can help to improve the selection of novel drug candidates *in silico*.

The conventional molecular dynamics (MD) is a powerful computational method which can take on the challenge of accounting for protein flexibilities at studying drug recognition and binding to a target site [3]. The key problem is that computational cost of a simulation is very high even for a small protein, and much higher for the large multi-domain/chain Neuraminidases, thus the average MD trajectory so far has been limited to tens of nanoseconds. E.g., in one of the most recent studies of the influenza Neuraminidase ten MD trajectories, each of at most 100 ns, were calculated [2]. These computations can help to understand protein flexibility and function,

[1]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia
[2]Belozersky Institute of Physicochemical Biology, Lomonosov Moscow State University, Moscow, Russia

but overall are far below timescales relevant for the majority of functionally important structural rearrangements. The use of CPU-based computing clusters can provide some speed-up, but the most promising solution so far is the emergence of specialized architectures. The Anton-1 and Anton-2 are special-purpose supercomputers for MD built by a privately held biochemistry company D. E. Shaw Research [6]. Antons run calculations entirely on specialized application-specific integrated circuits and can accommodate only custom-designed MD software, presenting a successful example of a deep co-design of the computer software and hardware to execute it. Although some access to this unique and apparently very expensive hardware is provided to scientific community in the US, its overall availability so far has been very limited. Alternatively, the graphical processing units (GPU) are serially produced specialized architectures which offer an affordable speed-up to MD.

In this work the advances in computer technologies were evaluated at simulating long MD trajectories of Neuraminidases. The article is organized as follows. Section 1 describes the molecular modeling protocol, hardware and software setup. In section 2, we report the results and their analysis. The Conclusion section summarizes the study.

1. Methods

1.1. Hardware and Software

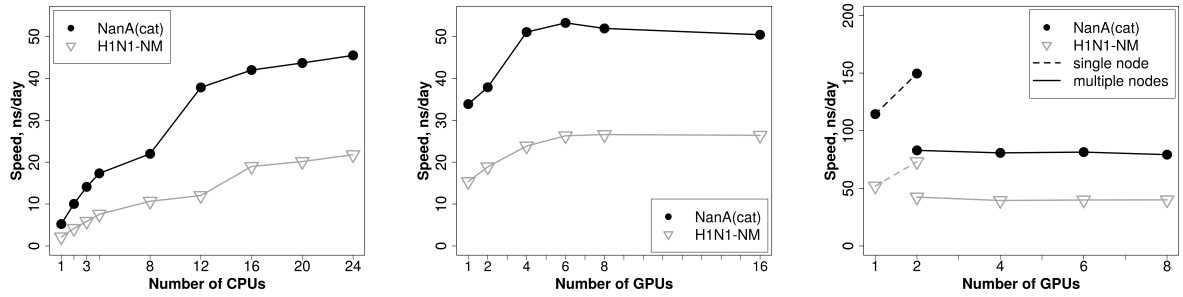
The computations were carried out on the "Lomonosov-2" supercomputer [4] using nodes equipped with (1) one Tesla K40 GPU, one Intel Xeon E5-2697 v3 CPU, and 64 GB RAM; and (2) two Tesla P100 GPUs, one Intel Xeon Gold 6126 CPU, and 96 GB RAM; and connected by Infiniband FDR. The AMBER16 package was compiled with Intel collection v.15.0.3, GNU collection v.4.8.5, MKL v. 11.1.3, CUDA v. 8.0, and OpenMPI v. 1.10.7.

1.2. Structure Preparation and the Molecular Modeling Protocol

The molecular systems of H1N1 influenza virus Neuraminidase (PDB 3B7E) and Neuraminidase A (NanA) from *Streptococcus pneumoniae* bacteria (PDB 2YA8) were prepared and simulated in the FF15IPQ force field with SPC/E_b water model as recently described [5]. The homotetrameric structure of the H1N1 Neuraminidase was reconstructed based on the BIOMT record in PDB. Only the catalytic domain (including the insertion domain) of the NanA Neuraminidase was selected for this study as the key catalytically competent part of the structure [5]. The size of influenza and bacterial Neuraminidases in a water box was ∼150000 and ∼65000 atoms, respectively. To compare the speed of MD on contemporary GPUs with the special-purpose Anton family of supercomputers, the molecular system of Human Dihydrofolate Reductase (PDB 2C2T) in water worth of ∼24000 atoms was constructed, which was equivalent to one previously used to benchmark Antons [6].

2. Results

Speed of MD simulation of H1N1 influenza virus Neuraminidase and the bacterial NanA in a water box was compared between classical Intel Xeon Gold CPUs and two types of GPUs — Tesla K40 based on Kepler architecture, and the recently introduced Tesla P100 based on Pascal architecture. The maximum speed-up reached on CPUs and Tesla K40 GPUs were comparable; however, it took at most 4–6 GPU-nodes to gain the performance of 20+ CPU-nodes (Fig. 1,



(a) Intel Xeon Gold 6126 (b) Tesla K40 (c) Tesla P100

Figure 1. Efficiency and scalability of the MD implemented in AMBER on different hardware

a and b). Therefore, the CPU-based acceleration of MD is possible, but not efficient. The computational yield of Pascal-based cards, although superior in absolute values, was qualitatively different compared to its Kepler-based predecessor. The speed of MD on a single Tesla P100 was 1.95x–2.14x times higher than the top output from 4–6 nodes equipped with Tesla K40 (Fig. 1, c). However, the performance of Tesla P100 did not improve when executed on multiple nodes, instead it degraded by 22–28%. The use of multiple Tesla P100 to run MD in parallel was efficient only within a single node in the peer-to-peer mode, when GPUs communicate directly via the PCI-E bus, providing the speed-up of 1.31x–1.41x with the yield of 73–150 ns/day. Therefore, a microsecond-long trajectory can be reached in at most 1–2 weeks.

The performance of the above-mentioned hardware was compared with special-purpose supercomputers by using a previously described benchmark based on a small protein system (see Section 1.2 of Methods). The peak performances of Anton-1 and Anton-2 machines were ∼40x and ∼228x times superior to two Tesla P100 GPUs in the peer-to-peer mode (Fig. 2).

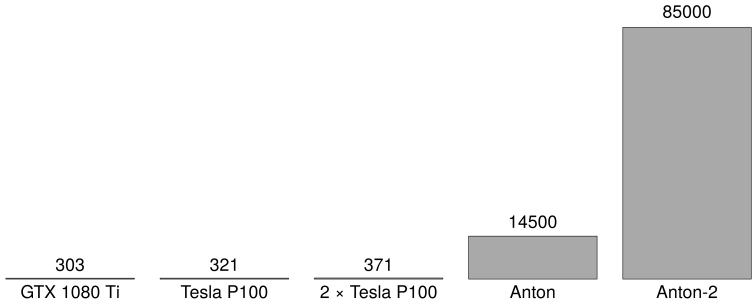


Figure 2. Speed of the MD at different GPU setups and Anton supercomputers (in ns/day)

Discussion and Conclusions

Molecular dynamics has a potential of becoming a particularly important tool in drug discovery to evaluate ligand recognition and binding to a flexible target site, but its use at a daily routine has been limited by a significant computational cost. We have shown that MD simulation of the structural dynamics of large Neuraminidases on modern GPUs can reach the microsecond timescale. This opens an opportunity to move away from the "static" affinity-driven strategies towards a deeper understanding of conformational dynamics in structures of proteins and mechanisms of association/dissociation of a drug candidate from a target site, leading to a better selection of promising compounds *in silico* to be further profiled experimentally. With the de-

velopment of more powerful video cards, the use of more than one GPU-node per simulation is becoming inefficient, leaving GPUs yet far behind the deeply-specialized Anton supercomputers, and thus supporting the recent trend for co-design of software and hardware for a particular purpose, of which Antons are an example, as a strategy to accelerate calculations. Further development of affordable specialized architectures is needed to move towards the much-desired millisecond timescale to simulate structural behavior of large proteins at a daily routine.

Acknowledgments

This work was carried out with the financial support of the Russian Science Foundation grant No. 15-14-00069-P. Supercomputer modeling was performed using equipment of the shared research facilities of HPC computing resources at the Lomonosov Moscow State University supported by the project RFMEFI62117X0011 [4].

References

1. Abed, Y., Boivin, G.: A review of clinical influenza A and B infections with reduced susceptibility to both Oseltamivir and Zanamivir. Open Forum Infectious Diseases 4(3), ofx105 (2017), DOI: 10.1093/ofid/ofx105

2. Durrant, J.D., Bush, R.M., Amaro, R.E.: Microsecond molecular dynamics simulations of influenza neuraminidase suggest a mechanism for the increased virulence of stalk-deletion mutants. The Journal of Physical Chemistry B 120(33), 8590–8599 (2016), DOI: 10.1021/acs.jpcb.6b02655

3. Ganesan, A., Coote, M.L., Barakat, K.: Molecular dynamics-driven drug discovery: leaping forward with confidence. Drug Discovery Today 22(2), 249–269 (2017), DOI: 10.1016/j.drudis.2016.11.001

4. Sadovnichy, V., Tikhonravov, A., Voevodin, V., Opanasenko, V.: Lomonosov: Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale Toward Exascale (Chapman & Hall/CRC Computational Science). pp. 283–307. CRC Press Boca Raton, Fla, USA (2013), DOI: 10.1201/2F9781351104005

5. Sharapova, Y., Suplatov, D., Švedas, V.: Neuraminidase a from streptococcus pneumoniae has a modular organization of catalytic and lectin domains separated by a flexible linker. The FEBS Journal 285(13), 2428–2445 (2018), DOI: 10.1111/febs.14486

6. Shaw, D.E., et al.: Anton 2: Raising the Bar for Performance and Programmability in a Special-purpose Molecular Dynamics Supercomputer. In: International Conference for High Performance Computing, Networking, Storage and Analysis, 2014, 16–21 Nov. 2014, New Orleans, LA, USA. pp. 41–53. IEEE Press (2014), DOI: 10.1109/SC.2014.9

7. Xu, Z., et al.: Sequence diversity of nana manifests in distinct enzyme kinetics and inhibitor susceptibility. Scientific Reports 6, 25169 (2016), DOI: 10.1038/srep25169

DOI: 10.14529/jsfi180304

Parallel GPU-based Implementation of One-Way Wave Equation Migration

Alexander L. Pleshkevich[1]***, Vadim V. Lisitsa***[2]***, Dmitry M. Vishnevsky***[2]***, Vadim D. Levchenko***[3]***, Boris M. Moroz***[1]

We present an original algorithm for seismic imaging, based on the depth wavefield extrapolation by the one-way wave equation. Parallel implementation of the algorithm is based on the several levels of parallelism. The input data parallelism allows processing full coverage for some area (up to one square km); thus, data are divided into several subsets and each subset is processed by a single MPI process. The mathematical approach allows dealing with each frequency independently and treating solution layer-by-layer; thus, a set of 2D cross-sections instead of the initial 3D common-offset vector gathers are processed simultaneously. This part of the algorithm is implemented suing GPU. Next, each common-offset vector image can be stacked, processed and stored independently. As a result, we designed and implemented the parallel algorithm based on the use of CPU-GPU architecture which allows computing common-offset vector images using one-way wave equation-based amplitude preserving migration. The algorithm was used to compute seismic images from real seismic land data.

Keywords: GPU, nested OMP, MPI, seismic imaging.

Introduction

Seismic imaging is a procedure which allows construction of the subsurface images from seismic data. Eventually, a seismic image is a convolution of a recorded signal and two Green's functions (one for source position, the other one for the receiver position). This procedure should be applied to all possible source-receiver pairs. Thus, the most computationally intense part is the computation of the Green's functions. To simplify this step, depth extrapolation of the wavefield was suggested in [1], where the wavefield is computed by solving the initial value problem for the one-way wave equation (OWE). A pseudo-spectral method to solve one-way wave equation was suggested recently in [2, 3].

OWE-based migration has several computational features, which make it suitable for the processing of large common-offset datasets. First, Greens functions for different sources-receivers positions can be computed independently and then combined to construct the image in (ω-x) domain. Second, Greens functions and thus images are extrapolated layer-by-layer, thus at each step of the algorithm one deals with 2D cross-sections of the solution defined on plane $z = const$, and we can process a high number of solutions simultaneously. As a result, one can consider any combination of the Greens functions for a fixed frequency at a fixed depth and construct an image; i.e. common-shot, common-receiver, common-azimuth, etc.

1. Mathematical Background

Consider initial value problem for the one-way wave equation stated in $(\omega - x)$ domain:

$$\frac{\partial u}{\partial z} - i\sqrt{\frac{\omega^2}{c^2(x,y,z)} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}}[u] = 0, \quad u(\omega, x, y, 0) = u_0(\omega, x, y). \tag{1}$$

[1]Central Geophysical Expedition JSC, Rosgeo, Moscow, Russian Federation
[2]Institute of Petroleum Geology and Geophysics SB RAS, Novosibirsk, Russian Federation
[3]Keldysh Institute of Applied Mathematics RAS, Moscow, Russian Federation

To solve equation (1), we suggest using the pseudo-spectral method presented in [2]. This approach allows to perform explicit stepping in vertical direction; i.e. assuming the solution to be known at a plane $z = z_0$, we can compute it at a plane $z = z_0 + \Delta z$ by the following rule:

$$u(\omega, x, y, z + \Delta z) = \sum_{j=n}^{n+1} \alpha_j F^{-1} \left[\exp\left(i\omega \Delta z \sqrt{\frac{1}{v_j^2} - \frac{k_x^2}{\omega^2} - \frac{k_y^2}{\omega^2}} \right) F[u(\omega, x, y, z)] \right] + \alpha_0 u(\omega, x, y, z), \tag{2}$$

where k_x and k_y are the spatial frequencies, $F[]$ and $F^{-1}[]$ denote the forward and inverse Fourier transforms with respect to the directions x and y, v_j are the nodal velocities, so that for each point $v(x, y, z) \in [v_n, v_{n+1}]$, parameters α_j are the coefficients of interpolation, as described in [2].

Using delta-functions as initial data one can compute Green's functions, corresponding to the positions of the sources and the receivers, and construct an image:

$$I(x, y, z) = \int_{\vec{x}_s \in D_S} \int_{\vec{x}_r \in D_R} \int_{\omega_{min}}^{\omega_{max}} G(\omega, x, y, z, \vec{x}_s) G(\omega, x, y, z, \vec{x}_r) \bar{f}(\omega, \vec{x}_s, \vec{x}_r) d\vec{x}_s d\vec{x}_r d\omega, \tag{3}$$

where $\vec{x}_s = (x_s, y_s)$ is a vector of sources coordinates at the daylight, and $G(\omega, x, y, z, \vec{x}_s)$ is the Green's function corresponding to the source position $\vec{x}_s$. Same notations are used for the receivers. Function f is the impulse recorded by the receiver at point $\vec{x}_r$ and emitted by the source at point $\vec{x}_s$. Choice of the domains of integration D_s and D_r define the type of constructed images (common-shot, common-receiver, common-offset etc.).

Note, that a coarse grid (50 m) in a vertical direction is used to solve OWE, after that the solution is interpolated to the grid with the step of 5 m to construct an image. In this work, we use a linear frequency-dependent interpolation, which can be applied directly to the image, rather than to the Green's function, which drastically reduces the number of computations.

2. Parallel Implementation

To account for data parallelism, we divide the input data into subsets, to be processed by a single MPI process. We use the coordinates of the mid-points to parametrize the imaging domain, so we apply the 2D domain decomposition and consider all the sources and receivers corresponding to the mid-points from a subdomain as a single dataset or "stencil". Data division leads to the loss of the algorithm scaling because some of the sources/receivers belong to several datasets. Thus, we need to compute the Green's functions for these sources/receivers several times; however, we achieve the number of the Green's functions recomputations as low as 2.3 on average.

Parallel implementation of the computational part of the algorithms includes two aspects. First, for each time-frequency, we use a loop with respect to depth; i.e. we compute the Green's functions for all sources and receivers positions, combine them and multiply by the corresponding signal. Thus, at fixed depth level, we construct all possible images $I^k(\omega_m x, y, z_l)$. The GPUs do the computation of the Green's functions and construction of images. After that, a set of 2D cross-sections of the images $I^k(\omega_m x, y, z_l)$ are uploaded to RAM. Next, we apply all-to-all MPI communications to exchange the images, so that each MPI process accumulates single common-offset images for all stencils. Second, we sum up images from different stencils (integrate over sources/receivers positions) and interpolate the result within a slab $[z, z + \Delta z]$. CPU performs these operations in parallel with the computation of the Green's functions described in the

previous paragraph. The output data of each MPI process is a single common-offset vector image for all stencils. The block-scheme of the algorithm is presented in Fig. 1.

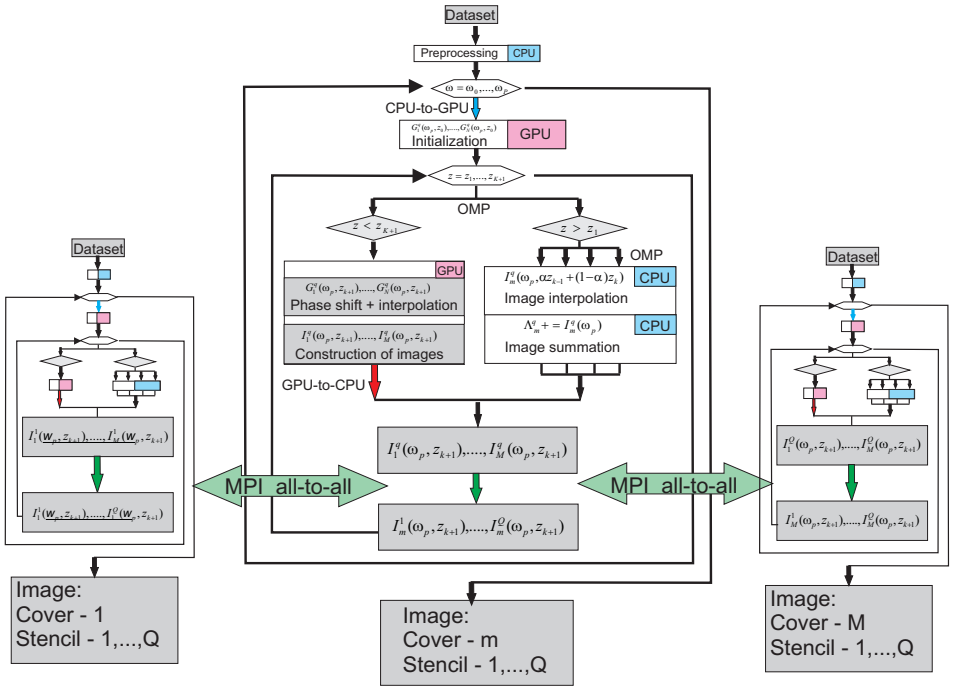


Figure 1. The block-scheme of the parallel algorithm

3. Numerical Experiments

3.1. Weak Scaling

To estimate a weak scaling of the algorithm, we consider the SEG Salt model (open-source model), for which we compute 17 images, having 2600 datasets. We perform simulations using 17, 34, 51, and 68 MPI processes, so, that each node deals with one dataset at a time. We provide the times needed for simulation in Tab. 1. The loss of the efficiency is caused by the MPI exchanges, which are implemented with enforced synchronization.

Table 1. Computation time for weak scaling estimation

	17 proc	34 proc	51 proc	68 proc
Time (hours)	1.91	2.22	2.64	3.02
W scaling (%)	-	86	72	63

3.2. Real-Data Example

We use the algorithm to construct the seismic image using real onshore seismic data. The size of the model is 15525 m InLine (x-direction) and 11250 m CrLine (y direction), depth is 5000 m. We compute 40 common-offset vector images, using 320 datasets, and 40 computational nodes. The total wall-clock time is about 320 hours; thus, the computational time is 11520 node-hours to construct the set of 40 images. A slalom-line section of the obtained 3D seismic image is presented in Fig. 2.

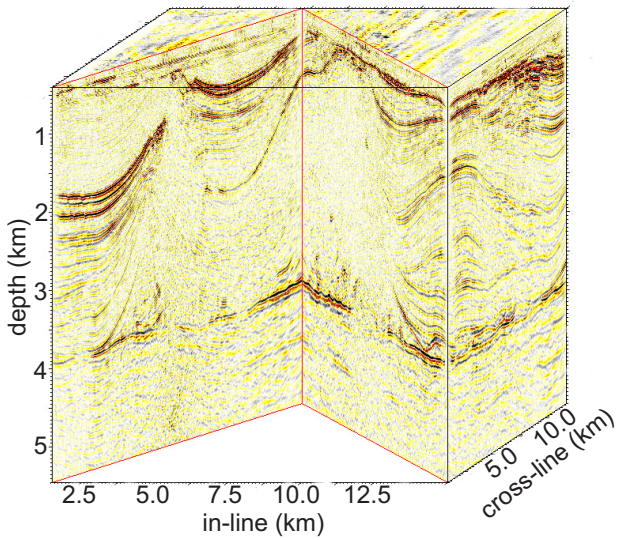


Figure 2. Slalom-line section of the obtained 3D seismic image

Conclusions

We presented an original algorithm of seismic migration, based on the solution of the one-way wave equation. The algorithm combines MPI, OMP, and CUDA technologies. Dataflow is parallelized via MPI so that each node deals with a single dataset. Computations of the Green's functions and the images are performed by GPU. After that, the images are passed between the nodes using MPI. Additional, computations and I/O are implemented via OMP technology.

Acknowledgments

This research was initiated and sponsored by Central Geophysical Expedition JSC of Rosgeo. V. Lisitsa and D. Vishnevsky are also thankful to Russian Foundation for Basic Research for partial financial support of this work, grants No. 18-05-00031, 18-01-00579, 16-05-00800. The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 and cluster NKS-30T+GPU of the Siberian Supercomputer Center.

References

1. Gazdag, J.: Wave equation migration with the phaseshift method. Geophysics 43(7), 1342–1351 (1978), DOI: 10.1190/1.1440899

2. Pleshkevich, A., Vishnevskiy, D., Lisitsa, V.: Development of pseudospectral amplitude-preserving 3d depth migration. Russian Geophysics (S) 0, 94–101 (2017)

3. Pleshkevich, A., Vishnevskiy, D., Lisitsa, V.: Explicit additive pseudospectral schemes of wavefield continuation with high-order approximation, pp. 5546–5550 (2017), DOI: 10.1190/segam2017-17633779.1

DOI: 10.14529/jsfi180305

High-performance Computational Modeling of Chromosome Structure

Yuri A. Eidelman[1], *Svetlana V. Slanina*[1], *Oleg A. Gusev*[2], *Sergey G. Andreev*[1,3]

We present a polymer modeling approach to generate the ensemble of 3D chromosome conformations at different time points of mitosis-interphase transition. Dynamics of structure during mitosis-G1 transition indicates quick and slow stages of chromosome shape alterations. At intermediate and late time scale the changes in chromosome compaction are small. To assess time dependence of contact map establishment during G1 we calculate contact maps at different times after mitotic decondensation. We demonstrate that the patterns of contacts observed soon after mitotic decondensation remain similar during G1. Whole contact map for mouse chromosome 18 at late G1 time correlates with the experimental chromosome conformation capture data. The simulations reproduce the main experimental findings, contact map persistence during G1 as well as specific pattern of long-range interactions in interphase chromosome. Our results suggest that spatial compartmentalization of an interphase chromosome is driven by interactions between different types of megabase sized chromatin domains during the formation of globular chromosome state at the end of mitotis to G1 transition.

Keywords: chromosome conformation capture, chromosome structure, computational modeling, mouse chromosome 18.

Introduction

Whole-genome chromosome conformation capture (Hi-C)-based experiments provide unique information about frequencies of contacts of any genetic loci within and between chromosomes in a cell. In general, the method represents a physico-chemical approach which uses chromatin fragmentation and proximity ligation and gives a snapshot of crosslinked chromosomal contacts in the population of fixed cells in the form of contact maps [5, 8]. To infer 3D chromosome structures from Hi-C data, computational approaches are used [3]. Methods based on polymer physics can predict ensembles of 3D conformations and contact maps, which are sensitive to principles of organization. Studies aimed at polymer modeling of interphase chromosomes for Hi-C analysis use condensation algorithms [1, 4]. The condensation conditions for long polymer chromatin chain due to bridging proteins [4] or attracting potentials [1, 10] are applied, and large scale transition from extended chromatin fiber, as rod or coil, to compact state, associated with interphase chromosome is simulated. These assumptions, however, contradict common knowledge that interphase chromosomes are formed in the course of decondensation of mitotic chromosomes.

To explore complexity of chromosome folding, we have mapped genetic loci contacts and predicted positions of megabase domains in interphase chromosome at different time of mitosis-interphase transition by Monte Carlo simulations. Our results support a mechanism of chromosome folding underlying contact patterns formation via interactions between different types of domains during establishment of globular state at the end of mitotis-G1 (M-G1) transition. This mechanism consistently reproduces a complex chess-like type of contact map and early G1 establishment of long-range contacts, as well as spatial compartmentalization of chromosome.

[1]Institute of Biochemical Physics IBCP, Russian Academy of Sciences, Moscow, Russia
[2]Translational Genomics Unit, RIKEN, Yokohama, Japan
[3]National Research Nuclear University MEPhI, Moscow, Russia

1. Chromosome Structure and Dynamics Modeling

Here we link the mitotic decondensation algorithm [2, 6] with several subunit type interactions. Any pair of elements (i, j) interacts through a Lennard-Jones potential. Since the chain is a heteropolymer, different types of elements have their own pairwise interaction potentials. To avoid complexity, subunit diameter d=350 nm was selected for all elements. Interphase structure is formed as a result of decondensation from an ultra-compact rodlike structure representing a mitotic-like chromosome. The Monte Carlo (MC) simulations were performed for times up to 500,000 MC steps at which the macroscopic parameter, chain gyration radius, ceases to change. At different times all macroscopic and microscopic characteristics for individual chromosomes and statistical ensembles (500 structures per each contact map calculation) are obtained.

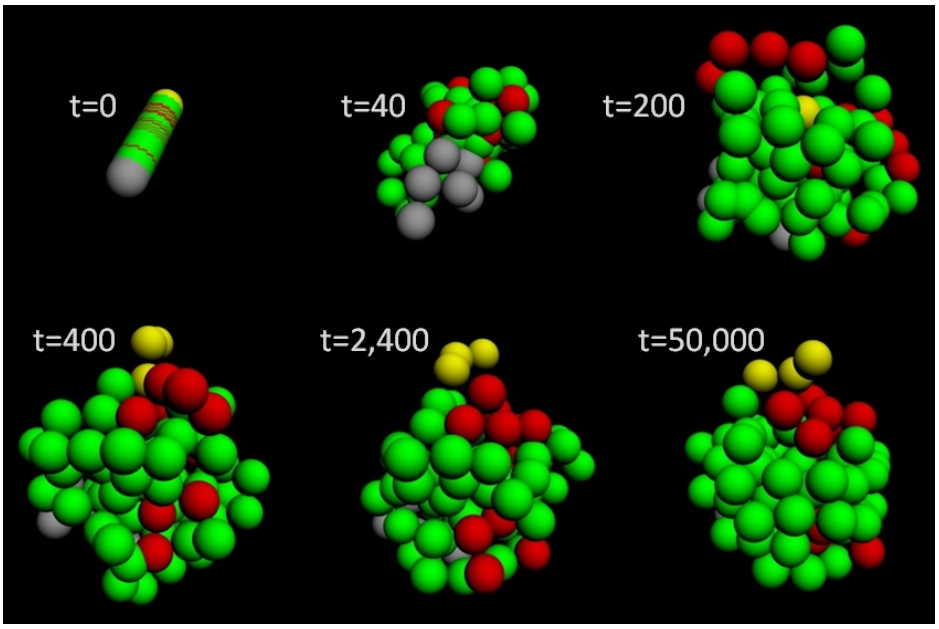
2. Results

Using MC simulations, we observe establishment of interphase chromosome organization in the course of mitosis-G1 transition. At t=0 decondensation begins and the chromosome is extended to the chain of subunits representing megabase-sized domains [5]. Domains interactions shape 3D conformations during mitosis-G1 decondensation and in interphase. The simulated dynamics of mouse chromosome 18 structure is shown in Fig. 1a. Different colors correspond to different subunit type. Dynamics of structure indicates that quick growth of chromosomal size gives way to slow decline. To assess time dependence of contact map establishment during G1 we calculate contact maps at different times after mitotic decondensation: at early (2,400 and 5,000 MC steps), intermediate (10,000 and 50,000 MC steps) and late times (500,000 MC steps) (Fig. 1b–f). We associate these points with different times in G1, from early to late subphases. The boundary of G1 is not determined here, and onset of S phase is not considered in the model.

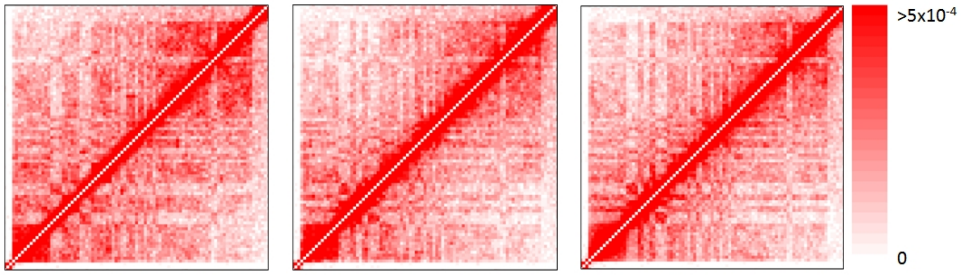
Map calculations for different times, Fig. 1b–f, show that the chess-like contact pattern is blurry at 2,400 MC steps and becomes clearer at 5,000 and 10,000 MC steps. But positions of characteristic dark/light blocks at early time remain unchanged at intermediate and late times. After 50,000 steps the map visually ceases to change. These data reveal that early G1 contact maps (at 2,400 and 5,000 MC steps) are highly correlated with the later time contact maps (Pearson correlation R=0.938 for 2,400 vs 50,000, R=0.923 for 2,400 vs 500,000, R=0.963 for 5,000 vs 50,000, R=0.952 for 5,000 vs 500,000 MC steps). Quality of the contact map in G1 generated by the polymer modeling is demonstrated by its correlation (R=0.897) with the experimental Hi-C map for ch12lx cells [9], Fig. 1g. Thus, main patterns of the experimental map [9] are well reproduced by polymer modeling of interphase chromosome 18, established following mitosis-G1 transition.

3. Discussion

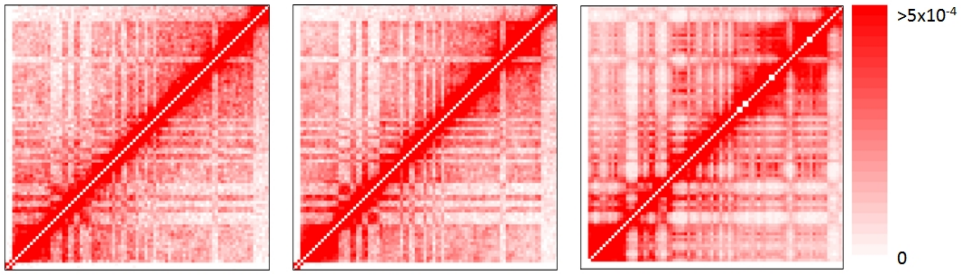
The modeling study of mitosis-G1 transition reveals two-stage dynamics, fast decondensation followed by slow structural rearrangements in G1 accompanied by weak subcondensation. The maps at early vs late times in G1 are highly correlated and look similar. Our polymer approach simulates mitosis-G1 transition as a biologically realistic pathway of establishment of interphase chromosome organization and contact maps during G1 phase of cell cycle. The chess-like contact pattern seen as alternating megabase-sized regions of high vs low intensities on the contact map and spatial compartmentalization ofchromosomal subunits are quantitatively re-



(a) Evolution of the typical conformation in the course of mitosis-G1 transition



(b) 2,400 MC steps (c) 5,000 MC steps (d) 10,000 MC steps



(e) 50,000 MC steps (f) 500,000 MC steps (g) experiment

Figure 1. Structure of mouse chromosome 18 predicted as a result of mitosis-G1 transition algorithm. (a): conformations; (b)–(g): contact maps

produced by mitotic decondensation modeling algorithm. The simulations agree with the main experimental findings, map persistence during G1, as well as specific pattern of long-range contacts in interphase chromosome.

Conclusions

We demonstrated how 3D interphase chromosome structure can be established during M-G1 transition. We identified the physical mechanism of interphase structure formation which incorporates differential interactions between megabase chromatin domains and explains main

features of experimental Hi-C contact maps. The results support the globular model of interphase chromosomes [1].

Acknowledgements

S.A. acknowledges support from the MEPhI Academic Excellence Project (Contract No. 02.a03.21.0005) and from the Russian Foundation for Basic Research grant 14-01-00825. The paper is supported by IBCP state contract 1201253309. The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011.

References

1. Andreev, S.G., Edelman, Iu.A.: Globular model of interphase chromosome and intrachromosomal exchange aberrations. Radiats. Biol. Radioecol. 39(1), 10–20 (1999)

2. Andreev, S.G., Edel'man, Iu.A., Talyzina, T.A.: Structural chromosome organisation and radiation-induced interchromosomal aberrations. Radiats. Biol. Radioecol. 46(1), 16–19 (2006)

3. Ay, F., Noble, W.S.: Analysis methods for studying the 3D architecture of the genome. Genome Biol. 16, 183 (2015), DOI: 10.1186/s13059-015-0745-7

4. Chiariello, A.M., Annunziatella, C., Bianco, S. et al.: Polymer physics of chromosome large-scale 3D organisation. Sci. Rep. 6, 29775 (2016), DOI: 10.1038/srep29775

5. Dixon, J.R., Selvaraj, S., Yue, F. et al.: Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485(7398), 376–380 (2012), DOI: 10.1038/nature11082

6. Eidelman, Y.A., Slanina, S.V., Aleshchenko, A.V., Andreev, S.G.: Chromosome interactome inferred from mitosis-G1 transition. BioRxiv preprint (2016), DOI: 10.1101/084608

7. Fritsch, C.C., Langowski, J.: Chromosome dynamics, molecular crowding, and diffusion in the interphase cell nucleus: a Monte Carlo lattice simulation study. Chromosome Res. 19(1), 63–81 (2011), DOI: 10.1007/s10577-010-9168-1

8. Lieberman-Aiden, E., van Berkum, N.L., Williams, L. et al.: Comprehensive mapping of long range interactions reveals folding principles of the human genome. Science 326(5950), 289–293 (2009), DOI: 10.1126/science.1181369

9. Rao, S.S., Huntley, M.H., Durand, N.C. et al.: A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159(7), 1665–1680 (2014), DOI: 10.1016/j.cell.2014.11.021

10. Zhang, B., Wolynes, P.G.: Topology, structures, and energy landscapes of human chromosomes. PNAS 112(19), 6062–6067 (2015), DOI: 10.1073/pnas.1506257112

DOI: 10.14529/jsfi180306

Numerical Simulations of Structural Chromosomal Instability

Yuri A. Eidelman[1], *Svetlana V. Slanina*[1], *Valentina S. Pyatenko*[1,2], *Sergey G. Andreev*[1,3]

The origin of dose-response curves for radiation-induced chromosomal instability (CI) is studied using the mechanistic CI model. The model takes into account DNA damage generation and repair in the progeny of irradiated cells and cell passage through mitotic cycle. We consider the formation of DNA double-strand breaks (DSBs) *de novo* in the S phase, where predominantly chromatid-type aberrations are formed. Among them sister chromatid exchanges of the "isochromatid deletion" type, or "chromatid dicentrics" are of primary interest. When the cell enters mitosis, the fate of chromosomal aberrations depends on their types. Chromosomal and chromatid fragments, having entered mitosis, either are transmitted into one of the daughter cells, or are lost. A chromatid dicentric in mitosis forms an anaphase bridge. These mechanistic assumptions were used to demonstrate that the dose-response curves are closely related to the dynamic curves for CI. The principles underlying this relationship are analyzed.

Keywords: chromosomal instability, ionizing radiation, delayed chromosomal damage, prediction, dose response.

Introduction

Chromosomal instability (CI) is defined as the increased frequency of chromosomal rearrangements (chromosomal aberrations) in offspring of irradiated cells [1]. As a theoretical basis, the theory of targets is not applicable here [2]. The first attempts at computer simulation of CI brought about promising results [3, 4], but led to a multitude of questions. Some of these issues are resolved in this paper. The influence of various factors on the main characteristics of gamma-induced CI dose-response and dynamic curves are analyzed. The delayed aberrations in the form of dicentrics are modeled here as an end-point of radiation-induced CI. The sensitivity of dose-response curves for dicentrics to the variation of the parameters of the CI model is analyzed.

1. Modeling

The mechanistic model of CI following ionizing radiation exposure incorporates DNA / chromosome damage interaction pathways determining outcomes of factors involved in genome destabilization. The modeling technique described previously [4] is used here with some modifications. The main points of the model are as follows.

DNA double-strand breaks (DSBs) can be formed in both irradiated cells and their progeny in the G1 phase, as well as in the S phase. The DSB formation in the G2 phase upon irradiation is neglected. For most cell types, when the asynchronous population is irradiated, the fraction of G2 cells is small, 10–15%. Besides, irradiation of G2-phase cells does not lead to the formation of dicentrics. In addition to DSBs, DNA single-strand breaks (SSBs) and oxidative base damage (BD), as well as complex lesions (SSB + BD) are also induced. They are repaired by BER pathway. Unrepaired SSBs + BDs alone do not lead to the formation of aberrations, but can turn into DSBs either due to nuclease attack of opposite DNA chain in G1 or during replication. The

[1]Institute of Biochemical Physics IBCP, Russian Academy of Sciences, Moscow, Russia
[2]Medical Radiological Research Center, Obninsk, Russia
[3]National Research Nuclear University MEPhI, Moscow, Russia

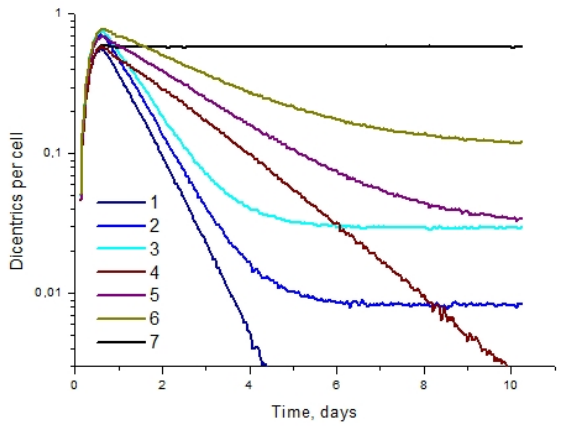
DSBs can be repaired by NHEJ or HR pathways, misrepaired with the formation of aberration of chromosome or chromatid type depending on the phase of the cell cycle, or lose their reactivity, forming a blunt-end aberration, or fragment. In the progeny of irradiated cells, as in [4], we consider the formation of DSBs *de novo* in the S phase, where predominantly chromatid-type aberrations are formed, of which sister chromatid exchanges of the "isochromatid deletion" type, or "chromatid dicentrics" are of primary interest. When the cell enters mitosis, the fate of CAs depends on their types. Chromosomal and chromatid fragments, having entered mitosis, either are transmitted into one of the daughter cells, or are lost. A chromatid dicentric in mitosis forms an anaphase bridge. Since both kinetochores in this aberration belong to the same chromatid, it cannot segregate normally. The mitotic spindle pulls it to opposite poles. An anaphase bridge either leads to cell death, or breaks (part of the so called "BFB cycle" [4]), and each of the daughter cells gets a centric chromosomal fragment with a sticky end. The chromosomal dicentric, in contrast to the chromatid dicentric, has four kinetochores, two on each chromatid, and therefore can either segregate normally, or form a double anaphase bridge. With normal segregation, each of the daughter cells receives a dicentric, and with the formation of a double bridge, the same outcomes are possible as in the case of a single bridge: the death and bridge breakage as part of the BFB cycle. As a result of breakage of anaphase bridges, in G1 centric fragments with sticky ends appear: single or double, depending on the type of the bridge that is broken. The reactive ends can interact with each other as well as with the DSBs generated in the S phase to form dicentrics and other types of aberrations. The cycle of formation, breakage and fusion of anaphase bridges is called BFB (Breakage and Fusion of Bridges [4]).

DSBs of non-radiation nature in the progeny are formed through three channels, Ndsb = N1 + N2 + N3. N1 is an autonomous channel, i.e. these DSBs are formed in each cell regardless of external factors (intercellular signals). N2 is a non-autonomous channel, determined by intercellular interactions through gap junctions. It depends on the density of the cells, both having and not having DNA damage. N3 is also a non-autonomous channel, it is determined by intercellular interactions through soluble factors in the medium. For simplicity, DSBs of all three types are considered structurally indistinguishable. The cell passage through cell cycle is taken into account in the same way as in [4].

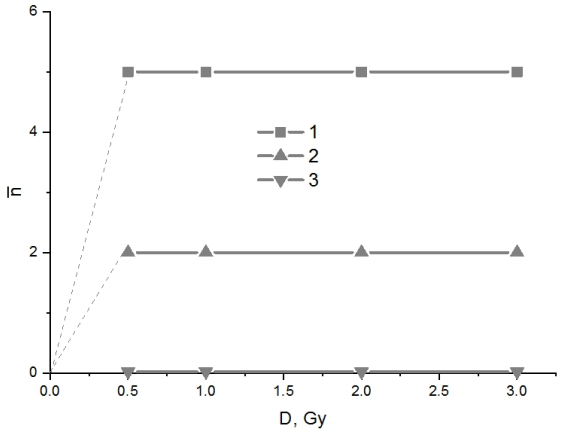
2. Results and Discussion

We studied the impact of the parameters of spontaneous DSB generation in the progeny of irradiated cells, as well as the parameters of the breakage-fusion of the anaphase bridge (BFB cycle) on the shape of the dynamic and dose-response curves of radiation-induced CI. Figure 1a shows dynamic CI curves, i.e. dependence of dicentric frequency on time after irradiation (3 Gy) for different probabilities of anaphase bridge breakage, p, and different levels of DSB generation in the S phase, $\bar{n}$. Corresponding dose dependencies of DSB generation $\bar{n}$ are shown in Fig. 1b. Figure 1c shows the set of dynamic curves for different doses with parameters corresponding to Fig. 1a, curve 2. The resulting dose dependence at 10 days is presented in Fig. 1d.

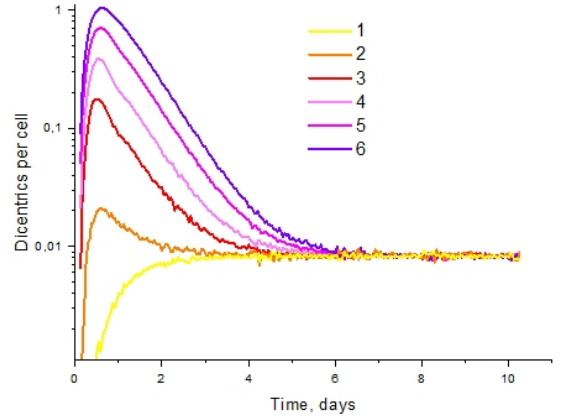
Figure 1 and Fig. 2 demonstrate that, depending on the combination of parameters, two main types of dynamic curves for chromosomal aberrations in offspring of irradiated cells can be distinguished: with (Fig. 1c) and without the plateau (Fig. 2a). All curves are observed for the rate of DSB generation independent of the dose in the range of medium and large doses (Fig. 1b).



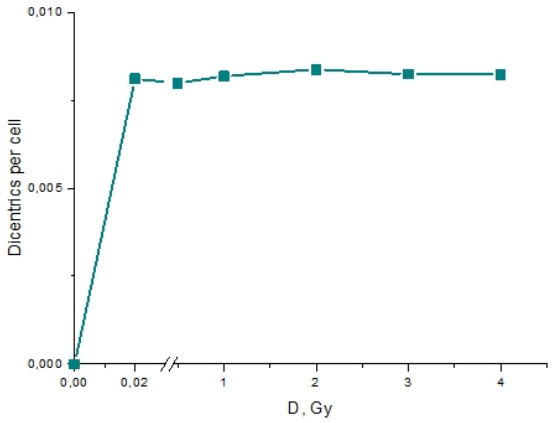
(a) dynamic CI curves for different model parameters. 1–3 - p=0.6; 4–6 - p=0.3; 7 - p=1.0. 1,4,7 - $\bar{n}$=0; 2,5 - $\bar{n}$=2; 3,6 - $\bar{n}$=5



(b) dose dependence of $\bar{n}$. 1 - for curves 3 and 6 in (a); 2 - for curves 2 and 5 in (a); 3 - for curves 1, 4 and 7 in (a)



(c) dynamic curves for different doses (curves 1–6: D=0.02, 0.5, 1, 2, 3, 4 Gy)



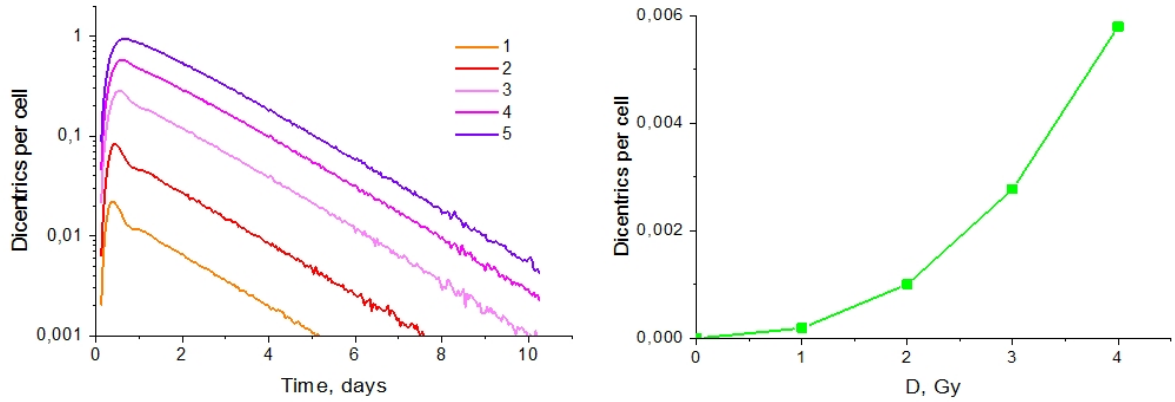
(d) dose curves at t=10 days, corresponding to the dynamic curves in (c)

Figure 1. Impact of the DSB generation and BFB parameters on the shape of CI dynamic and dose-response curves

These two types of CI dynamic curves are manifested in two different types of dose-response curves, plateau in a broad dose range (Fig. 1d) and pronounced dose dependence (Fig. 2b).

Conclusions

In conclusion, the basic properties of the developed model of radiation induced CI can be formulated as follows: (*) persistent induction of DNA DSBs and their repair impact the dynamic and dose characteristics of CI; (**) the shape of dose dependence of CI is determined by the quantitative relationships between accumulation and elimination of chromosomal aberrations at any time after irradiation of dividing cell population. Thus, the phenomenon of CI dose dependence-independence is of dynamic origin.



(a) dynamic curves for different doses with parameters corresponding to curve 4 in Fig. 1a (curves 1–5: D=0.5, 1, 2, 3, 4 Gy)

(b) the dose curve at t = 10 days corresponding to the dynamic curves in panel (a)

Figure 2. dose dependence of CI arises from the shape of dynamic curves at late times

Acknowledgements

The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011. S.A. acknowledges support from the MEPhI Academic Excellence Project (Contract No. 02.a03.21.0005). The work was supported in part by the Russian Foundation for Basic Research grant 14-01-00825 to S.A. The paper is supported by IBCP state contract (No. 201253309).

References

1. UNSCEAR: Radiation-induced genomic instability. In: United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). Effects of Ionizing Radiation. UNSCEAR 2006 Report. Vol II, Annex C: Non-targeted and Delayed Effects of Exposure to Ionizing Radiation, pp. 5–22. United Nations, New York (2009), http://www.unscear.org/docs/publications/2006/UNSCEAR_2006_Annex-C.pdf

2. Kadhim, M., Salomaa, S., Wright, E., Hildebrandt, G., Belyakov, O.V., Prise, K.M., Little, M.P.: Non-targeted effects of ionising radiation – implications for low dose risk. Mutat. Res. 752(2), 84–98 (2013), DOI: 10.1016/j.mrrev.2012.12.001

3. Andreev, S.G., Eidelman, Y.A.: Dose-response prediction for radiation-induced chromosomal instability. Radiat. Prot. Dosim. 143(1–4), 270–273 (2011), DOI: 10.1093/rpd/ncq509

4. Andreev, S.G., Eidelman, Y.A., Salnikov, I.V., Slanina, S.V.: Modeling study of dose-response relationships for radiation-induced chromosomal instability. Dokl. Biochem. Biophys. 451(1), 171–175 (2013), DOI: 10.1134/S1607672913040017

DOI: 10.14529/jsfi180307

GPU-based Implementation of Discrete Element Method for Simulation of the Geological Fault Geometry and Position

Vadim V. Lisitsa[1], *Vladimir A. Tcheverda*[1], *Victoria V. Volianskaia*[2]

We present an algorithm for numerical simulation of the geological fault formation. The approach is based on the discrete elements method, which allows modeling of the deformations and structural discontinuity of the Upper part of the Earth crust. In the discrete elements method, the medium is represented as an combination of discrete particles which interact as elastic or visco-elastic bodies. Additionally, external potential forces, for example gravitational forces, may be introduced. At each time step the full set of forces acting at each particle is computed, after that the position of the particle is evaluated on the base of Newtonian mechanics. We implement the algorithm using CUDA technology to simulate single statistical realization of the model, whereas MPI is used to parallelize with respect to different statistical realizations. Obtained numerical results show that for low dip angles of the tectonic displacements relatively narrow faults form, whereas high dip angles of the tectonic displacements lead to a wide V-shaped deformation zones.

Keywords: Discrete Element Method, geological faults, CUDA, statistical simulation.

Introduction

A classical definition of the geological faults is that they are discontinuities of sedimentary, metamorphic or magmatic rock bodies. Thus, no physical properties are assigned to a fault; however, real geological faults have a complex structure which includes main fault body (fault core) and fractured or damage zones around [5]. Usually, a fault is the result of tectonic movements. The properties of the fault and near-fault damage zone may significantly differ from the intact rocks and can be a fluid nature active flow channel even in tight formations. Very often, to do field observations or laboratory studies of the real fault is difficult or impossible due to some natural reasons. Thus, numerical simulation is a reliable and efficient way to investigate the peculiarities of the structures forming and tectonic movement process. There are numerous techniques to simulate finite deformations in geological formations including finite elements, finite differences, finite volumes, discrete elements; a review is presented in [3].

In this paper, we present an algorithm based on the Discrete Elements Method (DEM). This approach is based on the media representation by a set of discrete particles. These particles interact as stiff elastic bodies according to the mechanical rules; i.e., elastic and frictional forces affect each particle, that leads to the particle movement according to the Newton mechanics [2]. Computation of the forces affecting a particle includes a high number of floating point and logical operations; thus, it is computationally intense and hard to implement on CPU, using vectorization, etc. As a result, the efficiency of the CPU based realizations of DEM is low, and computation time to solve even a 2D problem may be as long as several thousand node-hours. On the contrary, GPU architecture is more appropriate for DEM implementation, because it can efficiently handle a big number of flops with a small amount of memory involved in computations.

[1]Institute of Petroleum Geology and Geophysics SB RAS, Novosibirsk, Russian Federation
[2] G&G Expert, Moscow, Russian Federation

1. Discrete Element Method

Discrete element method is a meshless approach where the media is represented as agglomeration of independent particles (typically spheres of circles). For each particle total force field is computed, which includes potential forces (gravity field), normal elastic forces, frictional forces, and dissipative forces. Consider two particles with the numbers i and j, with the coordinates $\vec{x}^i$ and $\vec{x}^j$ and radii R^i and R^j respectively. Particle j affects particle i with the normal forces:

$$\vec{F}_n^{ji} = \begin{cases} K_r^-(R^i + R^j - \|\vec{X}^{ji}\|)\vec{n}^{ji}, & R^i + R^j - \|\vec{X}^{ji}\| > 0, \quad repulsion, \\ K_r^+(R^i + R^j - \|\vec{X}^{ji}\|)\vec{n}^{ji}, & 0 \le R^i + R^j - \|\vec{X}^{ji}\| \le r_0, \quad active \;\; bond, \\ 0, & R^i + R^j - \|\vec{X}^{ji}\| > r_0, \quad no \;\; bond, \end{cases} \tag{1}$$

and tangential forces (friction):

$$\vec{F}_t^{ji} = \begin{cases} -K_s\delta_t\vec{t}^{ji}, & K_s\delta_t \le \mu^s\|\vec{F}_n^{ji}\|, \quad static \;\; friction, \\ -\mu^d\|\vec{F}_n^{ji}\|\vec{t}^{ji}, & K_s\delta_t > \mu^s\|\vec{F}_n^{ji}\|, \quad dynamic \;\; friction. \end{cases} \tag{2}$$

In these notations, $K_r^{\pm}$ are the repulsion and attraction Bulk moduli, μ^d and μ^s are the dynamic and static friction coefficients respectively, $\vec{X}^{ji} = \vec{x}^i - \vec{x}^j$, $\vec{n}^{ji}$ is a unit vector directed from the center of the j-th particle to the center of the i-th particle, δ_s is tangential displacement of j-th particle over i-th particle, $\vec{t}^{ji}$ is the self-normalized projection of the relative velocity vector to the plane normal to vector $\vec{n}^{ji}$; i.e., it is tangential to both particles.

Having computed all external forces acting at j-th particle one may recompute its position using classical mechanics principles:

$$M^j\frac{d^2\vec{x}^j}{dt^2} = \sum_{i\in J(i)}\left(\vec{F}_{ji}^n(\vec{x}^j,\vec{x}^i) + \vec{F}_{ji}^t(\vec{x}^j,\vec{x}^i)\right) - \nu\frac{d\vec{x}^j}{dt}, \tag{3}$$

where M^j is the mass of the particle, ν is the artificial viscosity. To integrate equation (3) numerically, we use the Verlet scheme [4].

2. Implementation of the Algorithm

According to the general formulation of the particle-based methods, one has to compute the forces affecting each particle due to the interaction with all other particles. However, in geomechanical modeling by the discrete element method, for each particle only a small number of neighboring particles directly contact the considered one. The adjacency matrix is sparse, but it can evolve. Thus, two related problems should be solved. First, organizing the process of adjacency matrix construction (approximation). Second, computing forces and applying the time stepping.

To construct the adjacency matrix, we suggest using the lattice method. As it follows from the equations (1) and (2), only directly contacting particles affect each other; thus, for each particle, the domain of dependence does not exceed $2R_{max} + r_0$, where R_{max} is the maximal radius of the particles. Also, due to the stability criterion of the Verlet scheme, a single particle cannot move more than $0.1R_{min}$ per a single time step, where R_{min} is the minimal radius over all particles. Thus, we can introduce a grid with the lattice size equal to $2R_{max}+r_0$, so that each particle and all its neighbors belong to the same lattice of directly adjoint lattice. Now we can state the rule of adjacency matrix approximation - for each particle, all the particles belonging

to the same or directly adjoint lattices are neighbors. In this case, we overestimate the number of connected particles but strictly simplify the process of the matrix construction.

The initial assignment of the particles to the lattices is performed by a sequential code by CPU. It is implemented particle-by-particle so that we determine the lattice number for considered particle and add the particle number to the list of particles for this lattice. This procedure is inapplicable under OMP of CUDA parallelization. Thus, the GPU implementation of the reassignment of the particles to the lattices is done lattice-by-lattice. The lattices are large enough, so that after one time step a particle may either stay in the same lattice or move to a directly adjoint lattice. Thus, to update the list of particles for each lattice, we need to check the particles which previously belonged to this lattice or the directly adjusted one. Similar ideas are used in the molecular dynamics and lattice Boltzmann methods but with different principles of lattices construction [1].

Computation of the forces and the numerical solution of the equation of motion is implemented on GPU. The parallelization is applied particle-by-particle, so that a GPU core computes forces for one particle at a time.

3. Numerical Experiments

We use the algorithm to simulate geological fault position for different directions of tectonic displacements in 2D. The computational domain was 4000 m by 500 m. The boundary conditions introduced the displacements. We consider five scenarios with the displacement direction equal to 0° (vertical displacement), 15°, 30°, 45°, and 60° from vertical direction pointing downward. So, in all cases, these are normal dip-slip faults. The vertical displacement was 100 m. Radii of the elements were homogeneously distributed from 0.75 to 1.5 m. The dynamic friction coefficient $\mu_d = 0.3$. We provide the results of simulation – distribution of the horizontal strains ε_{xx} in Fig. 1. Note, that the sub-vertical displacements lead to the formation of a wide damage zone, which has a form of a wedge. The closer the displacements to horizontal direction, the narrower is the wedge with higher deformations in the fault body. Moreover, the main direction of the fault; i.e., the dip angle tends to about 30° which is governed by the properties of the media, rather than displacement direction.

Conclusions

We presented an algorithm for numerical simulation of the geological fault geometry and position. The approach is based on the Discrete Elements Method which allows modeling of finite deformations and structural discontinuities in the Earth's crust. We implement the algorithm using CUDA technology to simulate single statistical realization of the model, whereas MPI is used to parallelize for different statistical realizations. The presented numerical experiments illustrate the formation of the V-shaped fault damage zone. The obtained strains distributions can be used further to estimate the mechanical and transport properties of the fault damage zone.

Acknowledgments

V. Lisitsa and V. Tcheverda proposed and justified this approach to geological modeling of tectonic motions and faults formation. V. Lisitsa developed the algorithm and implemented the

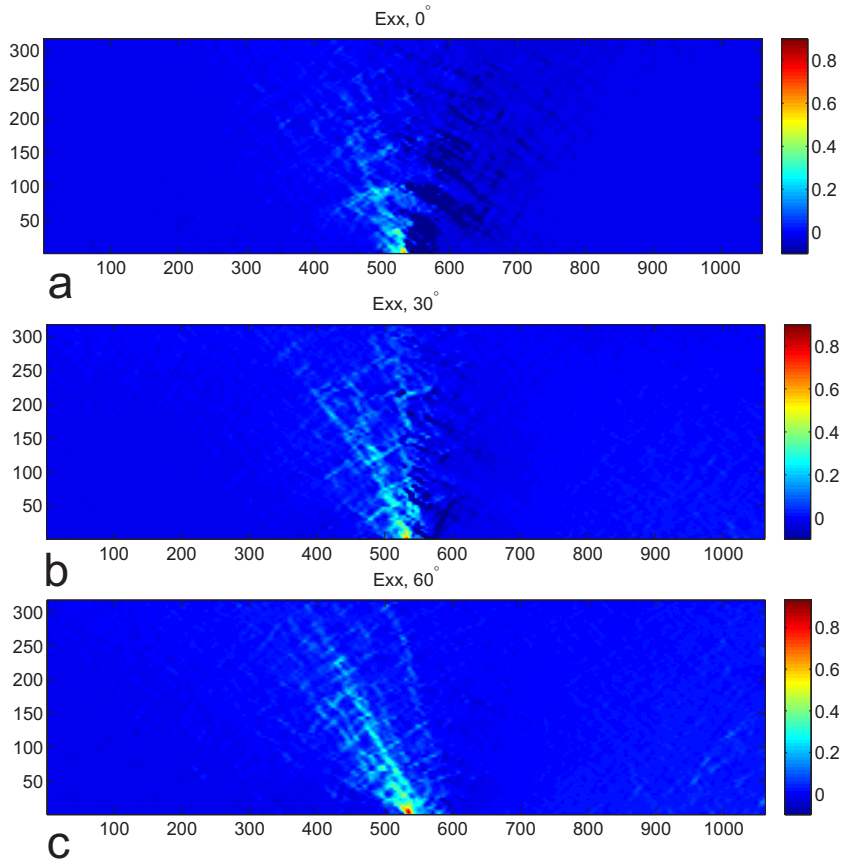


Figure 1. Distribution of horizontal strains ε_{xx} for the displacement directions 0° (a), 30° (b), and 60° (c)

series of numerical experiments. V. Volianskaia did the geological explanation of the simulation results. V. Lisitsa and V. Tcheverda were sponsored by the Russian Science Foundation grant 17-17-01128. The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 and cluster NKS-30T+GPU of the Siberian supercomputer center.

References

1. Alpak, F.O., Gray, F., Saxena, N., Dietderich, J., Hofmann, R., Berg, S.: A distributed parallel multiple-relaxation-time lattice boltzmann method on general-purpose graphics processing units for the rapid and scalable computation of absolute permeability from high-resolution 3D micro-ct images. Computational Geosciences 22(3), 815–832 (2018), DOI: 10.1007/s10596-018-9727-7,

2. Hardy, S., Finch, E.: Discrete-element modelling of detachment folding. Basin Research 17(4), 507–520 (2005), DOI: 10.1111/j.1365-2117.2005.00280.x

3. Lisjak, A., Grasselli, G.: A review of discrete modeling techniques for fracturing processes in discontinuous rock masses. Journal of Rock Mechanics and Geotechnical Engineering 6(4),

301–314 (2014), DOI: 10.1016/j.jrmge.2013.12.007

4. Mora, P., Place, D.: Simulation of the frictional stick-slip instability. Pure and Applied Geophysics 143(1), 61–87 (1994), DOI: 10.1007/BF00874324

5. Peacock, D.C.P., Dimmen, V., Rotevatn, A., Sanderson, D.J.: A broader classification of damage zones. Journal of Structural Geology 102, 179–192 (2017), DOI: 10.1016/j.jsg.2017.08.004

DOI: 10.14529/jsfi180308

Modelling of Quantum Qubit Behaviour for Future Quantum Computers

Andrey N. Chibisov[1], *Mary A. Chibisova*[1]

This work deals with quantum qubit modelling based on a silicon material with embedded phosphorus atoms because a future quantum computer can be built on the basis of this qubit. The building of atomic models of bulk crystalline silicon and silicene, as well as calculation of their total energies, were performed using the Quantum ESPRESSO software package, using highperformance computing (HPC). For silicon and phosphorus atoms the generalized gradient approximation (GGA) was used in terms of the spin-orbit non-collinear interaction by means of the Quantum ESPRESSO package. The equilibrium orientations of the phosphorus qubit spins and localization of the wave functions in the 2D and bulk crystalline silicon phases were theoretically investigated by means of quantum-mechanical calculations. The existence of an exchange interaction between qubits has been confirmed, which leads to a change in the wave function's localization and spin orientation, and in the case of silicene, this interaction was stronger.

Keywords: quantum qubit, quantum computer, quantum-mechanical calculations, spin.

Introduction

The problem of practical implementation of quantum computers is an important scientific and technological task at present time. The quantum computer will be able to instantly solve such important challenges as designing modern materials with specified properties, creation of new types of drugs, new types of cryptographic encryption, and also to significantly improve the work of machine learning and artificial intelligence systems. Over the past 10 years, various research groups have actively attempted to study and construct a quantum point qubit based on the spin in silicon [8]. Some groups have used the technological scheme proposed by Kane in 1998 [6]. According to this scheme, qubits are formed from long-lived nuclear spins of single phosphorus impurities in the Si crystal and are controlled by external surface radio-frequency, magnetic and electric fields. The spin states of an electron bound to one phosphorus donor in silicon show a long time of consistency and relaxation. Then, for the first time, Marello with co-workers [7] reported the possibility of the spin orientation readout in silicon. This allowed giving a further powerful impetus for an active study of spin orientations in silicon [9]. After that, a numerical model was developed for investigating electron-spin transfer through donors in silicon [5]. The authors showed that there are surprising effects that arise due to the quantization of phosphorus nuclei donors. They showed how it is possible to display the electron-exchange interaction using simple transport measurements. Thus, in Ref. [10], a behaviour of one- and two-qubit operations in a quantum point system using the exchange interaction was demonstrated. According to the latest data, researchers are trying to achieve a controlled overlap of the wave function for high-accuracy readout of spins on each qubit. Thereby, in Ref. [2] anticorrelated spin states between two donor spin-qubits in silicon located at a distance of 16 $\pm$ 1 nm were obtained. The authors have shown that the exchange interaction can be turned "on" and "off" by means of an electric field controlling the two phosphorus atoms.

This work deals with quantum qubit modelling based on a silicon material with embedded phosphorus atoms because a future quantum computer can be built on the basis of this qubit. To form the physical foundations of the quantum computer behaviour, it is necessary to study

[1]Computing Center, Far Eastern Branch, Russian Academy of Sciences, Khabarovsk, Russia

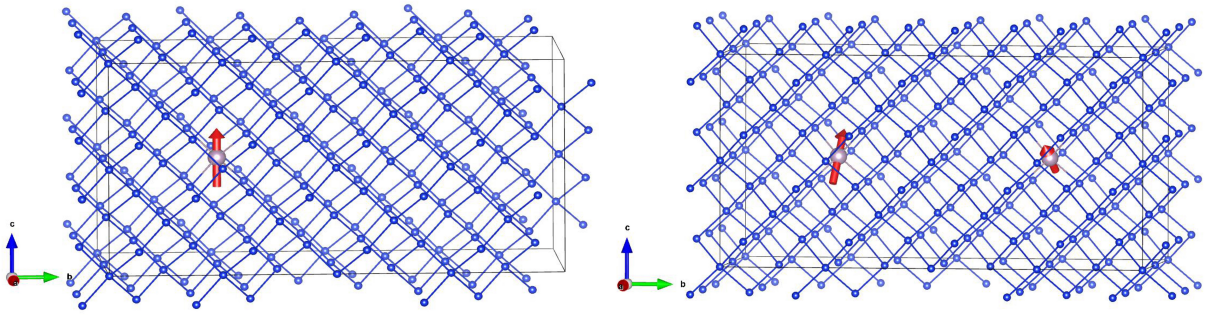
a model system based on a silicon material with embedded phosphorus atoms. Moreover, it is necessary to accurately describe the phosphorus impurity behaviour in a silicon matrix. The results will provide accurate data on the quantization and transmission of quantum information by means of spins.

1. Methods and Approaches

The building of atomic models of bulk silicon and silicene, as well as calculation of their total energies, were performed using the Quantum ESPRESSO software package [4], using high-performance computing (HPC). PBE (Perdew–Burke–Ernzerhof) pseudopotentials for silicon and phosphorus atoms in the generalized gradient approximation (GGA) in terms of the spin–orbit non-collinear interaction are taken from the Quantum ESPRESSO package. Calculation of a unit cell of bulk silicon was performed taking into account the $8 \times 8 \times 8$ k-points. For silicene, a special set of $9 \times 9 \times 1$ k-points was used with the 476.2 eV cut off energy of plane waves. The equilibrium silicene model was obtained by free relaxation of all atoms in the structure. For this purpose, the 32-atom model was placed in a cell with a size of $15.467 \times 15.467 \times 12.490$ Å^3. Atomic relaxation was carried out to the interatomic forces value of about 0.026 eV/Å.

2. Results and Discussion

During an implantation of one phosphorus atom (we denote it as 1P) into the bulk silicon lattice, the orientation "↑" is observed for the spin of the P atom's excess electron (see Fig. 1a). Adding a second impurity phosphorus atom (designated as 2P), at a distance of about 11 Å from the first one, leads to a change in the spin orientation of the 1P qubit (see Fig. 1b). It rotates to the $(\theta, \varphi) = (24°, -160°)$ angles (according to the notations in the Bloch sphere [1]). The second phosphorus atom has an orientation equal to $(\theta, \varphi) = (42°, -22°)$. Figure 2a shows the localization probability $|\psi|^2$ of the excess two electrons for the simultaneous presence case of 1P and 2P qubits in the silicon lattice.

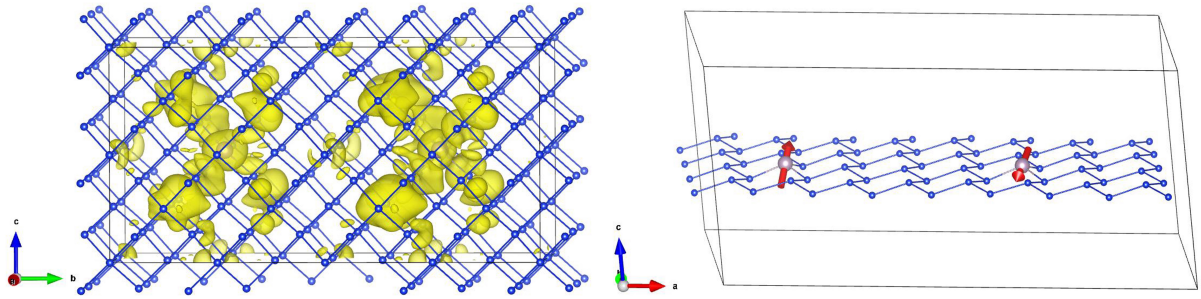


(a) for the case 1P

(b) for the case of the 1P (left) and 2P (right) atoms

Figure 1. The excess electron-spin orientation of the P atom

Then, we have calculated the equilibrium structure of pure silicene. The defectless silicene has the following geometric parameters (for the diatomic basis): $a = b = 3.867$ Å, the Si–Si interatomic distance is 2.270 Å, and the corrugation parameter is $d = 0.410$ Å. These parameters are in good agreement with the literature data: $a = b = 3.884$ Å; d(Si–Si) = 2.280 Å [3]. The phosphorus atom implantation into the silicene structure (at the 1/32 concentration) leads to its



(a) the silicon lattice (b) the silicene

Figure 2. The $|\psi|^2$ density in the silicon lattice and the spin orientations in silicene

strong distortion. The bond lengths slightly increase compared to pure silicene and, on average are Si–Si ≈ 2.275 Å, Si–P ≈ 2.285 Å. The θ_{SiPSi} internal angle for the phosphorus-doped silicene significantly decreased to approximately 109.6°. In pure silicene, the θ_{SiSiSi} angle was 116.8° on the average. During the one phosphorus atom implantation (1P) into the silicene lattice for the P electron spin, the "↑" orientation is also observed. The addition of a second phosphorus atom (2P) at a distance of about 15.467 Å from the first one leads to the spin orientation change of the 1P qubit as in bulk material. In this case, the 1P qubit spin turns to $(\theta, \varphi) = (43°, 95°)$ angles (see Fig. 2b). For the second phosphorus atom the orientation is $(\theta, \varphi) = (120°, –85°)$.

Thus, there is an exchange interaction between the two 1P and 2P qubits, leading to a localization change of the wave functions, spin orientation, and local charge density distribution. Besides, in case of the 2D silicon phase (silicene), this interaction is stronger because there is a greater deviation of the qubit spins' directions from the equilibrium orientations. We believe that the obtained results have a prospective significance for their use in the qubits design technology for future quantum computers.

Conclusion

We have used first-principles calculations to investigate the equilibrium orientation of the phosphorus qubit spins in 2D and bulk silicon phases. The localization of the wave functions and spin orientations were studied in detail. The results indicate that there is an exchange interaction between the two 1P and 2P qubits, leading to a localization change of the wave functions, spin orientation and local charge density distribution. Besides, in case of the 2D silicon phase (silicene), this interaction is stronger.

Acknowledgements

This work was supported by a grant from the Khabarovsk Territory Ministry of Education and Science (No. 117/2018D). The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 and partly using a cluster at the Shared Facility Centre "Data Centre of FEB RAS" (Khabarovsk, Russia). The authors would like to thank the Irkutsk Supercomputer Centre of SB RAS for providing access to HPC-cluster "Akademik V.M. Matrosov" (Irkutsk Supercomputer Centre of SB RAS, Irkutsk: ISDCT SB RAS; http://hpc.icc.ru, accessed 27.08.2018).

References

1. Bergou, J.A., Hillery, M.: Introduction to the Theory of Quantum Information Processing. Springer New York (2013), DOI: 10.1007/978-1-4614-7092-2

2. Broome, M.A., Gorman, S.K., House, M.G., Hile, S.J., et al.: Two-electron spin correlations in precision placed donors in silicon. Nature Communications 9(1) (mar 2018), DOI: 10.1038/s41467-018-02982-x

3. Dzade, N.Y., Obodo, K.O., Adjokatse, S.K., Ashu, A.C., Amankwah, E., Atiso, C.D., Bello, A.A., Igumbor, E., Nzabarinda, S.B., Obodo, J.T., Ogbuu, A.O., Femi, O.E., Udeigwe, J.O., Waghmare, U.V.: Silicene and transition metal based materials: prediction of a two-dimensional piezomagnet. Journal of Physics: Condensed Matter 22(37), 375502 (aug 2010), DOI: 10.1088/0953-8984/22/37/375502

4. Giannozzi, P., Baroni, S., Bonini, N., Calandra, M., et al.: QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. Journal of Physics: Condensed Matter 21(39), 395502 (sep 2009), DOI: 10.1088/0953-8984/21/39/395502

5. Gorman, S.K., Broome, M.A., Baker, W.J., Simmons, M.Y.: Impact of nuclear spin dynamics on electron transport through donors. Physical Review B 92(12) (sep 2015), DOI: 10.1103/physrevb.92.125413

6. Kane, B.E.: A silicon-based nuclear spin quantum computer. Nature 393(6681), 133–137 (may 1998), DOI: 10.1038/30156

7. Morello, A., Pla, J.J., Zwanenburg, F.A., Chan, K.W., Tan, K.Y., Huebl, H., Möttönen, M., Nugroho, C.D., Yang, C., van Donkelaar, J.A., Alves, A.D.C., Jamieson, D.N., Escott, C.C., Hollenberg, L.C.L., Clark, R.G., Dzurak, A.S.: Single-shot readout of an electron spin in silicon. Nature 467(7316), 687–691 (sep 2010), DOI: 10.1038/nature09392

8. Oberbeck, L., Hallam, T., Curson, N.J., Simmons, M.Y., Clark, R.G.: STM investigation of epitaxial si growth for the fabrication of a si-based quantum computer. Applied Surface Science 212-213, 319–324 (may 2003), DOI: 10.1016/s0169-4332(03)00370-2

9. Pla, J.J., Tan, K.Y., Dehollain, J.P., Lim, W.H., Morton, J.J.L., Jamieson, D.N., Dzurak, A.S., Morello, A.: A single-atom electron spin qubit in silicon. Nature 489(7417), 541–545 (sep 2012), DOI: 10.1038/nature11449

10. Veldhorst, M., Yang, C.H., Hwang, J.C.C., Huang, W., Dehollain, J.P., Muhonen, J.T., Simmons, S., Laucht, A., Hudson, F.E., Itoh, K.M., Morello, A., Dzurak, A.S.: A two-qubit logic gate in silicon. Nature 526(7573), 410–414 (oct 2015), DOI: 10.1038/nature15263

DOI: 10.14529/jsfi180309

Multiscale Simulations Approach: Crosslinked Polymer Matrices

***Pavel V. Komarov*[1,2], *Daria V. Guseva*[3], *Vladimir Yu. Rudyak*[3], *Alexander V. Chertovich*[3]**

Atomistic molecular dynamics simulations can usually cover only a very limited range in space and time. Thus, the materials like polymer resin networks, the properties of which are formed on macroscopic scale, are hard to study thoroughly using only molecular dynamics. Our work presents a multiscale simulation methodology to overcome this shortcoming. To demonstrate its effectiveness, we conducted a study of thermal and mechanical properties of complex polymer matrices and establish a direct correspondence between simulations and experimental results. We believe this methodology can be successfully used for predictive simulations of a broad range of polymer matrices in glassy state.

Keywords: polymers, networks, atomistic molecular dynamics, mesoscale simulations, multiscale simulations.

Introduction

In recent years, exciting opportunities associated with the design and fabrication of novel polymer matrices have appeared with the promise to produce a new generation of polymer-based materials having unique mechanical and thermal properties. Crosslinked polymer matrices are complex to design since their precise structure and network topology are generally unknown. Moreover, to optimize technological curing process, it is important to know how the mechanical properties vary during crosslinking. Thus, the development of specific simulation methods for the analysis of processes of polymer network formation is in high demand. Since complex nature of crosslinked polymer matrices with irregular local structures at the atomistic level ultimately manifest itself, in macroscopic properties, to simulate these matrices on a large range of length and time scale, a combination of modelling techniques is required. Mesoscale chemistry (MSC) techniques [1] consider the presence of a discrete mesoscale-level structure in a material and are primarily used to predict structure and topological properties of polymer networks from chemical constitution of initial comonomers. Atomistic molecular dynamics (MD) technique, on the other hand, is able to predict a variety of physical properties of composite materials for a wide range of engineering applications. These two types of techniques must be combined to a multiscale approach that is capable to predict the structure and macroscopic properties of polymer-matrix materials.

During the last decade, we developed a multiscale simulation methodology that makes it possible to predict various physical properties of highly crosslinked polymer materials [2–6]. In this work, we summarize our simulations methodology combining coarse-grained MSC technique with full atomistic modelling. Here, we apply it to generate fully atomistic model of highly crosslinked networks of phthalonitrile resin, based on the monomer bis(3-(3,4-dicyanophenoxy) phenyl) phenyl phosphate (DPPPP) and the initiator 1,3-bis(4-aminophenoxy)benzene (APB) [5], and to predict their physical properties (see Fig. 1a). The selected resin is one of the most promising examples of thermostable material to retain mechanical properties at elevated temperatures up

[1]Nesmeyanov Institute of Organoelement Compounds RAS, Moscow, Russian Federation.
[2]Tver State University, Tver, Russian Federation.
[3]Lomonosov Moscow State University, Moscow, Russian Federation.

to 450 °C and can be already considered as substitution for aluminum and titanium alloys in such applications as jet engine blades, skin of hypersonic aircrafts, etc. [7].

1. Methodology and Implementation

The first stage of our methodology involves the concept of the so-called "mesoscale chemistry" [1, 3, 6], which is used to study the complex polymerization process of appropriate monomers and to construct and equilibrate polymer networks at the mesoscale level. We introduce all molecular structures in terms of the simplified bead-and-spring model. Coarse-grained (CG) representation includes reactive beads that can form new chemical bonds during simulation of curing process. We use dissipative particle dynamics (DPD) [3, 8] as simulation engine at this level in order to model matrix curing through polycondensation or radical polymerization processes.

On the second stage we perform extensive topological analysis of the networks, first proposed by Khalatur in [3]. This stage is important to refine the adaptive model: (a) to optimize the set of DPD parameters for generation of polymer matrix with correct topology [5], and (b) to modify the network by removing "wrong" bonds leading to internal stresses.

On the third stage, the mechanical properties of constructed CG matrices are analyzed by DPD simulations. A uniaxial deformation of the simulation box along one of the axes is applied keeping the volume of the box constant (NVT ensemble), and stress-strain response curves are measured [6]. The dimensionless units in DPD are rescaled to GPa by the simple linear scaling procedure, the scaling factor depends on CG representation of molecular species.

Then, we construct fully atomistic models of crosslinked samples using an automatic reverse mapping procedure. This procedure extracts the coordinates of all DPD beads and their topological connections, and replaces the beads with the corresponding molecular fragments according to coarse-graining scheme (see details in [5]). The restored matrix may contain the so-called "spearing" monomers. They may be formed due to spearing of benzene rings by interatomic bonds of the closely situated fragments. Such misalignments are identified and eliminated using special Monte Carlo based procedure of fragment relocation, developed in our group.

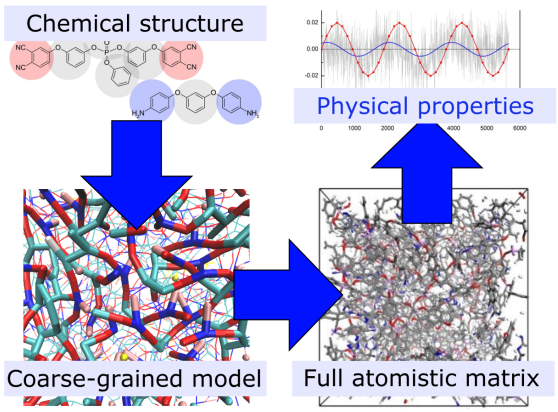
Finally, the physical properties (the glass transition temperature, T_g, mechanical properties, etc.) of the recovered atomistic samples are analyzed through molecular dynamics simulations using standard GROMACS and/or LAMMPS packages with pcff [9] force field for interatomic interactions. The mechanical properties of these samples are investigated by applying uniaxial deformation of small amplitude (either with constant strain rate or cyclic) and measuring its mechanical response. The details on equilibration procedure and the methodology of calculation of T_g and mechanical properties are given in works [5, 6].

The hybrid computational scheme described above was tested for different polymer matrices. Below, its application to phthalonitrile resin is demonstrated (see [5, 6] for details). CG simulations of matrix curing required 10–100 CPU hours depending on system size. CG mechanical analysis required 100–1000 CPU hours, and MD simulations were the most time consuming and required $\sim$ 12700 CPU hours on Lomonosov-2 supercomputer (installed at the Lomonosov Moscow State University) per one material sample. The described scheme was validated by the comparison of the predicted physical properties to the corresponding experimental measurements. The experimental studies were done simultaneously with simulations and the comparison was performed only at the final stage (without refinement of the model to fit experimental data).

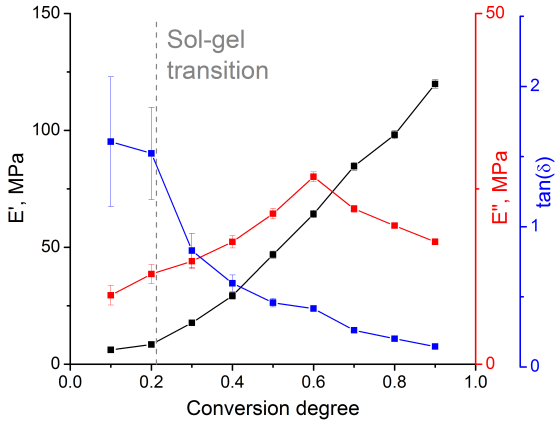
2. Results and Discussion

Figure 1b–d presents our results obtained for CG systems and fully atomistic samples with conversion degrees from 0.1 to 0.9. The dynamic properties (storage, $E^{'}$, and loss, $E^{''}$, moduli, and the mechanical loss coefficient, $\tan\delta = E^{'}/E^{''}$) demonstrate behavior typical for crosslinking systems: the transition from viscous behavior at small conversion degrees to elastic behavior at large conversion degrees (Fig. 1b), in full accordance with classical rheological literature data [10]. In addition, one can easily observe that there is good qualitative matching between both simulation techniques: system shows zero moduli for poorly crosslinked system, and elasticity (Young's modulus, E) exponentially grows up soon after sol-gel transition (Fig. 1c).

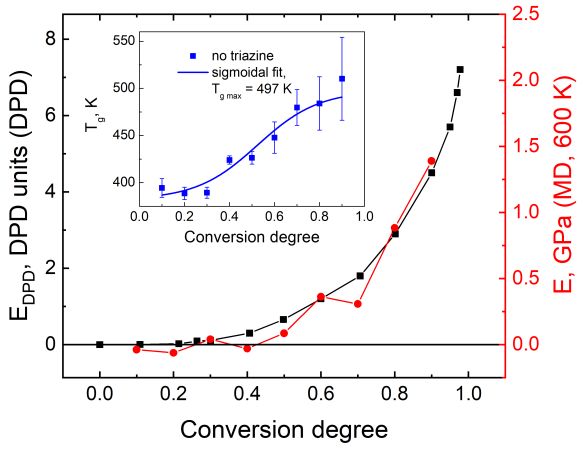
Figure 1d illustrates the similarity of data from MD simulations and experimental results [5, 6] for thermo-mechanical properties ($E^{'}$ *vs.* T_g) of phthalonitrile networks at various conversion degrees and temperatures. Thus, the suggested scheme of simulation allows us to make reasonable predictions of physical properties for highly crosslinked polymer matrices.



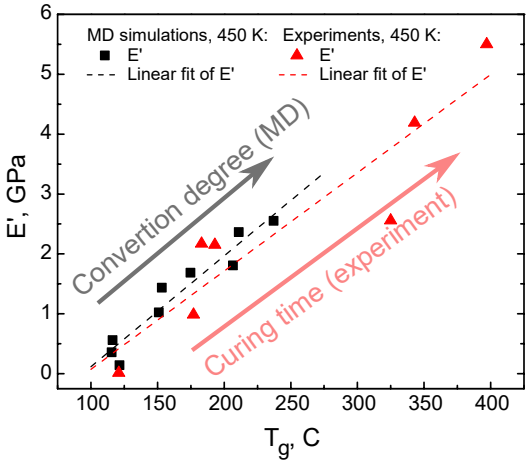
(a) Overview of the suggested scheme of simulations



(b) Dynamic moduli and mechanical loss coefficient of phthalonitrile matrices (DPD, rescaled to real units)



(c) Elasticity measured in DPD and MD simulations (T=600 K, deformation rate 5×10^{-5} nm/ps). The insert shows T_g as a function of the conversion degree



(d) The storage modulus $E^{'}$ as a function of T_g, MD simulations ($f \approx 500$ MHz) and experimental data ($f = 1$ Hz, see [6]), both at $T = 450$ K

Figure 1. Scheme of simulations and results on elasticity and dynamic moduli of the samples, simulations and experiments

Conclusions

This work demonstrates the efficiency of our *in silico* methodology for predictions of the physical properties of crosslinked polymer matrices on the challenging example of phthalonitrile thermosets. The results show good quantitative agreement with the experimental measurements at temperatures below and near glass transition. We claim that the presented methodology could be used without significant changes as the universal manual to perform computational studies of crosslinked polymer matrices and related nanocomposites constructed from a wide range of compounds and through various crosslinking mechanisms.

Acknowledgments

The reported study was funded by RFBR according to the research project No. 16-33-60215 "mol_a_dk". The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011.

References

1. Antonietti, M., Ozin, G.A.: Promises and problems of mesoscale materials chemistry or why meso? Chem. Eur. J. 10 (1), 28–41 (2004), DOI: 10.1002/chem.200305009

2. Komarov, P.V., Chiu, Y.-T., Chen, S.-M., Khalatur, P.G., Reineker, P.: Highly cross-linked epoxy resins: an atomistic molecular dynamics simulation combined with a mapping/reverse mapping procedure. Macromolecules 40(22), 8104–8113 (2007), DOI: 10.1021/ma070702+

3. Gavrilov, A.A., Komarov, P.V., Khalatur, P.G.: Thermal properties and topology of epoxy networks: a multiscale simulation methodology. Macromolecules 48(1), 206–212 (2015), DOI: 10.1021/ma502220k

4. Rudyak, V.Y., Gavrilov, A.A., Guseva, D.V., Chertovich, A.V.: Complex curing pathways and their influence on the phthalonitrile resin hardening and elasticity. Macromol. Theory Simul. 26(4), 1700015 (2017), DOI: 10.1002/mats.201700015

5. Guseva, D.V., Rudyak, V.Y., Komarov, P.V., Sulimov, A.V., Bulgakov, B.A., Chertovich, A.V.: Crosslinking mechanisms, structure and glass transition in phthalonitrile resins: insight from computer multiscale simulations and experiments. J. Polym. Sci. Part B. Polym. Phys. 56(5), 362–374 (2018), DOI: 10.1002/polb.24548

6. Guseva, D.V., Rudyak, V.Yu., Komarov, P.V., Bulgakov, B.A., Babkin, A.V., Chertovich, A.V.: Dynamic and static mechanical properties of crosslinked polymer matrices: multiscale simulations and experiments. Polymers 10(7), 792 (2018), DOI: 10.3390/polym10070792

7. Bulgakov, B.A., Sulimov, A.V., Babkin, A.V., Timoshkin, I.A., Solopchenko, A.V., Kepman, A.V., Avdeev, V.V.: Phthalonitrile-carbon fiber composites produced by vacuum infusion process. J. Compos. Mater. 51(30), 4157–4164 (2017), DOI: 10.1177/0021998317699452

8. Hoogerbrugge, P.J., Koelman, J.M.V.A.: Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. Europhys. Lett. 19, 155–160 (1992), DOI: 10.1209/0295-5075/19/3/001

9. Sun, H.: Ab initio calculations and force field development for computer simulation of polysilanes. Macromolecules 28(3), 701–712 (1995), DOI: 10.1021/ma00107a006

10. Menard, K.P.: DMA: Introduction to the Technique, Its Applications and Theory, CRC Press (1999)

DOI: 10.14529/jsfi180310

Application of High Performance Computing for Comparison of Two Highly Branched Lysine Molecules of Different Topology

***Igor M. Neelov*[1], *Oleg V. Shavykin*[1], *Maxim Y. Ilyash*[1], *Valeriy V. Bezrodnyi*[1], *Sofia E. Mikhtaniuk*[1], *Anna A. Marchenko*[1], *Emil I. Fatullaev*[1], *Anatolii A. Darinskii*[2], *Frans A. M. Leermakers*[3]**

High performance computations were performed for comparison of size and other properties of big heavily charged biocompatible molecules of complex topology in water. Lysine dendrimer and short dendritic brush of the same molecular weight were studied by molecular dynamics simulation method and GROMACS software package. The size and structural properties of these two systems were compared. It was shown that dendritic brush has smaller size and more dense core than the dendrimer. Radial density profile for both molecules is not monotonous and has minimum near core of molecules. This minimum is wider and deeper for dendrimer than for dendritic brush. Thus dendrimer has larger region of low density than dendritic brush and is more suitable for use for encapsulation and delivery of hydrophobic drugs.

Keywords: high performance computing, dendrimer, dendritic brush, Poly-L-lysine.

Introduction

High performance computing are widely used for molecular simulation of complex biological and biocompatible molecules. Lysine dendrimers were studied in several papers both by computer simulation using molecular dynamics method [6, 7, 10, 13, 15] and theoretically using numerical SCF approach [14]. At the same time there is almost no papers on theory and simulation of lysine dendritic brushes and comparison of properties of these molecules with properties of lysine dendrimers. The goal of the present paper is to compare the size and internal structure of these two types of molecules.

1. Model and Method

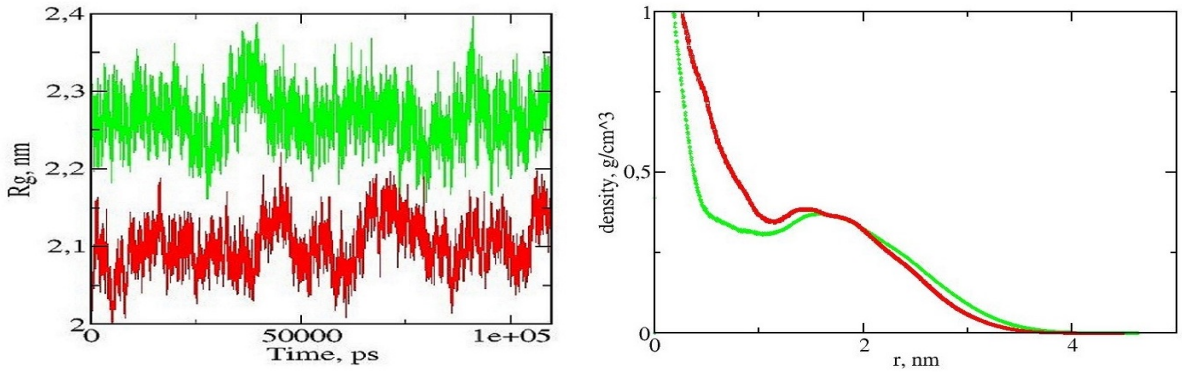
We studied dendrimer and dendritic brush of the same molecular weight but with different core structure. The dendrimer of the fifth generation has point-like core (one lysine residue) and the short dendritic brush has a linear core consisting of 8 lysine residues in main chain and 8 lysine dendrons of 2nd generation [6, 15]. The molecular weight M=16496 and 16553, total number of atoms N_a=2810 and 2819 and number of charged terminal NH3+ groups N_t=128 and 128 were for dendrimer and brush, correspondingly.

The computer simulation was performed by using the molecular dynamics method on the base of the Gromacs-4.5.6 package and the full atomic AMBER99SB-ildn force field. The simulation box contained one dendrimer or dendritic brush, water molecules and chlorine counterions. We also used several computer programs from our previous simulation of linear polymers [1, 11], dendrimers [8, 9, 17, 18], polymer brushes [12] and polyelectrolytes [2–5]. More simulation details could be found in [6].

[1]St. Petersburg National Research University of Information Technologies, Mechanics and Optics (ITMO University), St. Petersburg, Russia
[2]Institute of Macromolecular Compounds, Russian Academy of Sciences, St. Petersburg, Russia
[3]Physical Chemistry and Soft Matter, Wageningen University, Wageningen, The Netherland

2. Results

2.1. Large-Scale Properties



(a) (b)

Figure 1. (a) the gyration radius R_g as function of time and (b) the density profiles of simulated molecules – dendrimer (green color) and dendritic brush (red color)

The dependence of instant size (gyration radius R_g) on time t during production run 120 ns was calculated (see Fig. 1a). The values of R_g for both systems fluctuate but their average values practically does not change with time and R_g for dendrimer (green line in Fig. 1a) is always greater than R_g for dendritic brush (red line in Fig. 1a). The average values of R_g are equal 22.6 and 20.8 for dendrimer and dendritic brush correspondingly. The asphericity parameter of both molecules is smaller than 0.04. It means that the shape of both molecules is very close to spherical.

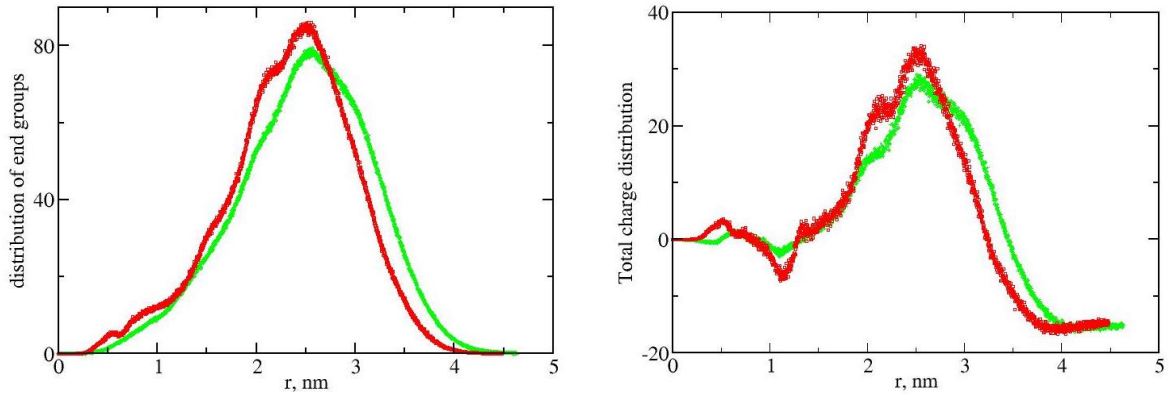
2.2. Internal Structure

Internal structure of both molecules could be characterized by the radial density profile $\varphi(r)$ around center of mass. The density profiles for both molecules (see Fig. 1b) decrease from the core to the periphery for both molecules. For the dendrimer a wide minimum of density near core of molecule is observed. For dendritic brush there is similar minimum but it is not so wide and less deep.

Figure 2a demonstrates that terminal monomers distributed almost over all possible distances r from their center of mass. This phenomenon ("backfolding") is well known for dendrimers. However such behavior was not known earlier for short dendritic brushes (Fig. 2a). The total charge distribution function shows that the charge of positive NH3+ are almost neutralized by the negative chlorine counterions at small r (i.e. it is suitable for encapsulation of hydrophobic drugs). Non-compensated NH3+ groups are concentrated near the outer layer (see positive maximum on Fig. 2b) with chlorine counterions forming spherical layer.

Conclusions

It was shown that both molecules have spherical shape but the dendrimer has greater gyration radius R_g and more deep density minimum (void) near core than dendritic brush. Therefore,



(a) (b)

Figure 2. The radial distributions of number of terminal NH3+ groups (a) and of the total charge (b) for dendrimer (green color) and brush (red color).

dendrimer is more suitable for use for encapsulation and delivery of drugs than dendritic brush with the same molecular weight.

Acknowledgement

The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 [16]. I.M.N. and O.V.S. are supported by Government of Russian Federation (grant 08-08).

References

1. Darinsky, A., Lyulin, A., Neelov, I.: Computer simulations of molecular motion in liquid crystals by the method of Brownian dynamics. Macromol. Chem. Theory Simul. 2, 523–530 (1993), DOI: 10.1002/mats.1993.040020402

2. Ennari, J., Elomaa, M., Neelov, I.: Modeling of water-free and water containing solid polyelectrolytes. Polymer 41, 985–990 (2000), DOI: 10.1016/S0032-3861(99)00235-9

3. Ennari, J., Neelov, I., Sundholm, F.: Molecular dynamics simulation of the PEO sulfonic acid anion in water. Computational and Theoretical Polymer Science 10, 403–410 (2000), DOI: 10.1016/S1089-3156(00)00006-4

4. Ennari, J., Neelov, I., Sundholm, F.: Simulation of a PEO based solid polyelectrolyte, comparison of the CMM and the Ewald summation method. Polymer 41, 2149–2155 (2000), DOI: 10.1016/S0032-3861(99)00382-1

5. Ennari, J., Neelov, I., Sundholm, F.: Estimation of the ion conductivity of a PEO-based polyelectrolyte system by molecular modeling. Polymer 42, 8043–8050 (2001), DOI: 10.1016/S0032-3861(01)00311-1

6. Falkovich, S., Markelov, D., Neelov, I., Darinskii, A.: Are structural properties of dendrimers sensitive to the symmetry of branching? computer simulation of lysine dendrimers. J. Chem. Phys. 139, 064903 (2013), DOI: 10.1063/1.4817337

7. Markelov, D., Falkovich, S., Neelov, I., Ilyash, M., Matveev, V., Lahderanta, E., Ingman, P., Darinskii, A.: Molecular dynamics simulation of spin-lattice NMR relaxation in Poly-L-Lysine dendrimers: Manifestation of the semiflexibility effect. Phys. Chem. Chem. Phys. 17, 3214–3226 (2015), DOI: 10.1039/C4CP04825C

8. Mazo, M., Shamaev, M., Balabaev, N., Darinskii, A.: Conformational mobility of carbosilane dendrimer: Molecular dynamics simulation. Phys. Chem. Chem. Phys. 6, 1285–1289 (2004), DOI: 10.1039/B311131H

9. Neelov, I., Adolf, D.: Brownian dynamics simulations of dendrimers under elongational flow: Bead-rod model with hydro-dynamic interactions. Macromolecules 36, 6914–6924 (2003), DOI: 10.1021/ma030088b

10. Neelov, I., Markelov, D., Falkovich, S., Ilyash, M., Okrugin, B., Darinskii, A.: Mathematical modeling of lysine dendrimers: Temperature dependencies. Polym. Sci. Ser. C 55, 154–161 (2013), DOI: 10.1134/S1811238213050032

11. Neelov, I., Adolf, D., McLeish, T., Paci, E.: Molecular dynamics simulation of dextran extension by constant force in single molecule AFM. Biophys. J. 91, 3579–3588 (2006), DOI: 10.1529/biophysj.105.079236

12. Neelov, I., Binder, K.: Brownian dynamics of grafted polymer chains-time-dependent properties. Macromol. Theor. Simul. 4, 1063–1084 (1995), DOI: 10.1002/mats.1995.040040605

13. Neelov, I., Janaszewska, A., Klajnert, B., Bryszewska, M., Makova, N., Hicks, D., Pearson, H., Vlasov, G., Ilyash, M., Vasilev, D., Dubrovskaya, N., Tumanova, N., Zhuravin, I., Turner, A., Nalivaeva, N.: Molecular properties of lysine dendrimers and their interactions with $\alpha\beta$-Peptides and neuronal cells. Curr. Med. Chem. 20, 134–143 (2013), DOI: 10.2174/0929867311302010013

14. Okrugin, B., Neelov, I., Leermakers, F.M., Borisov, O.: Structure of asymmetrical peptide dendrimers: Insights given by self-consistent field theory. Polymer 125, 292–302 (2017), DOI: 10.1016/j.polymer.2017.07.060

15. Okrugin, B., Ilyash, M., Markelov, D., Neelov, I.: Lysine dendrigraft nanocontainers. influence of topology on their size and internal structure. Pharmaceutics 10, 128 (2018), DOI: 10.3390/pharmaceutics10030129

16. Sadovnichy, V., Tikhonravov, A., Voevodin, V., Opanasenko, V.: Contemporary High Performance Computing: From Petascale Toward Exascale, pp. 283–307. CRC Press, Boca Raton, FL (2013), DOI: 10.1201/9781351104005

17. Shavykin, O., Mikhailov, I., Neelov, I.M., Darinskii, A., Leermakers, F.: Effect of an asymmetry of branching on structural characteristics of dendrimers revealed by Brownian dynamics simulations. Polymer 146, 256–266 (2018), DOI: 10.1016/j.polymer.2018.04.055

18. Shavykin, O., Neelov, I., Darinskii, A.: Is the manifestation of the local dynamics in the spin-lattice NMR relaxation in dendrimers sensitive to excluded volume interactions? Phys. Chem. Chem. Phys. 18, 24307–24317 (2016), DOI: 10.1039/C6CP01520D

DOI: 10.14529/jsfi180311

Developing Efficient Implementations of Bellman–Ford and Forward-Backward Graph Algorithms for NEC SX-ACE

Ilya V. Afanasyev[1], *Alexander S. Antonov*[1], *Dmitry A. Nikitenko*[1], *Vadim V. Voevodin*[1], *Vladimir V. Voevodin*[1], *Kazuhiko Komatsu*[2], *Osamu Watanabe*[2], *Akihiro Musa*[2], *Hiroaki Kobayashi*[2]

The main goal of this work is to demonstrate that the development of data-intensive applications for vector systems is not only important and interesting, but is also very possible. In this paper we describe possible implementations of two fundamental graph-processing algorithms for an NEC SX-ACE vector computer: the Bellman–Ford algorithm for single source shortest paths computation and the Forward-Backward algorithm for strongly connected components detection. The proposed implementations have been developed and optimised in accordance with features and properties of the target architecture, which allowed them to achieve performance comparable to other traditional platforms, such as Intel Skylake, Intel Knight Landing or IBM Power processors.

Keywords: graph algorithms, NEC SX-ACE, vector computing, data-intensive applications.

Introduction

With a modern variety of hardware and software solutions in supercomputing, it is very important to study fundamental properties of algorithms in order to understand which architectures are more suitable for various groups of algorithms. One very important and challenging example is a group of data-intensive algorithms: those algorithms usually stress target platform memory subsystem, as a result demonstrating low performance, high latency and poor data-cache usage due to low data locality. Graph algorithms represent data-intensive applications extremely well since they tend to have an enormous amount of random memory accesses, paired with low computational complexity. Moreover, graph processing is extremely relevant nowadays since it is used in various important real-world applications, such as web-graphs and social-networks processing. On the hardware side, vectorisation is an important feature of modern processors. NEC company has a long history and a lot of experience in developing vector systems with its computers having a large amount of unique features, such as extremely long vector length. Moreover, NEC computers are equipped with high-performance nodes and high-bandwidth memory, what makes them a good candidate for development of data-intensive applications. Since the area of data-intensive application development (and particularly graph algorithms) for vector systems is currently not studied enough, this paper describes implementation details of two fundamental graph processing problems for a NEC SX-ACE computer: single source shortest paths computation and strongly connected components search.

1. Target Platform and State of the Art

In this work we investigate implementations for a NEC SX-ACE computer. A single SX-ACE node includes 4 vector cores, each one with 1 GHz clock frequency. A single vector processing unit (VPU) of each core is capable of processing vectors of 256 length. The peak vector performance

[1]Lomonosov Moscow State University, Moscow, Russian Federation
[2]Tohoku University, Tohoku, Japan

of a single NEC SX-ACE socket is 256 Gflop/s, while peak scalar performance is only 4 Gflop/s per socket. Each vector core also supports out of order execution for vector load and store operations, with advanced data forwarding in vector pipes chaining. Each node is equipped with 256 GB/s high-bandwidth memory. Therefore, the bytes per flop ratio (the ratio of the peak performance to the memory bandwidth) is equal to 1, which is somewhat higher compared to other scalar and accelerator computers.

The approaches described in this paper are suitable only for shared-memory architectures (e.g. servers or individual supercomputer nodes). Graph algorithms implementations for shared-memory architectures are generally much more efficient compared to implementations for distributed-memory systems, since graph-processing problems tend to have an enormous amount of inter-node communications, what can easily become a bottleneck for overall performance. Moreover, modern studies demonstrate that platforms with shared-memory architecture are widely used to process various real-world graphs (social networks, road graphs and many others). There are not many currently existing studies and researches, which cover the development of graph algorithms or other data-intensive applications on NEC systems. However, there are several existing approaches describing the development of graph algorithms for other systems focused on vectorisation, such as Intel Knight Landing processors [1, 3].

2. Algorithms

In this paper two fundamental graph processing problems are investigated: single source shortest paths (SSSP) and strongly connected components (SCC). The SSSP problem implies finding the shortest paths in an undirected weighted graph from the selected source vertex to other vertices. The SCC problem implies partitioning a directed unweighted graph into disjoint sets of vertices, each one representing a strongly connected component where each vertex is reachable from another. In order to solve the shortest paths problem, the Bellman–Ford algorithm is used, while the Forward-Backward algorithm with a trim step is used for solving the strongly connected components problem. Those algorithms, together with possible approaches for their parallel implementation are described in detail in [2, 4]. From the computational point of view the Bellman–Ford algorithm requires both floating point and integer arithmetics together with frequent indirect memory accesses. In the meantime, the Forward-backward algorithm requires only integer arithmetics with frequent indirected memory accesses, and has a more complex nested parallelism potential. Furthermore, the Bellman–Ford algorithm has a slightly higher computational complexity than the Forward-Backward algorithm, since it operates with undirected graphs. Another important property of both algorithms is that they can be implemented without atomic operations usage, which tend to strongly impede vectorisation.

3. Implementation Details

An important point in any graph algorithm development is the selection of a data structure. This point is even more important for vector systems, since the selected data structures have to support efficient (bank-conflict free) memory accesses. In order to achieve efficient vectorisation, input graphs are represented in the format of a list of edges. This format allows a very simple and very efficient vectorisation of the loops, where the whole graph is traversed. Moreover, if the list of edges format is used, the edges can be stored in any order; some particular order

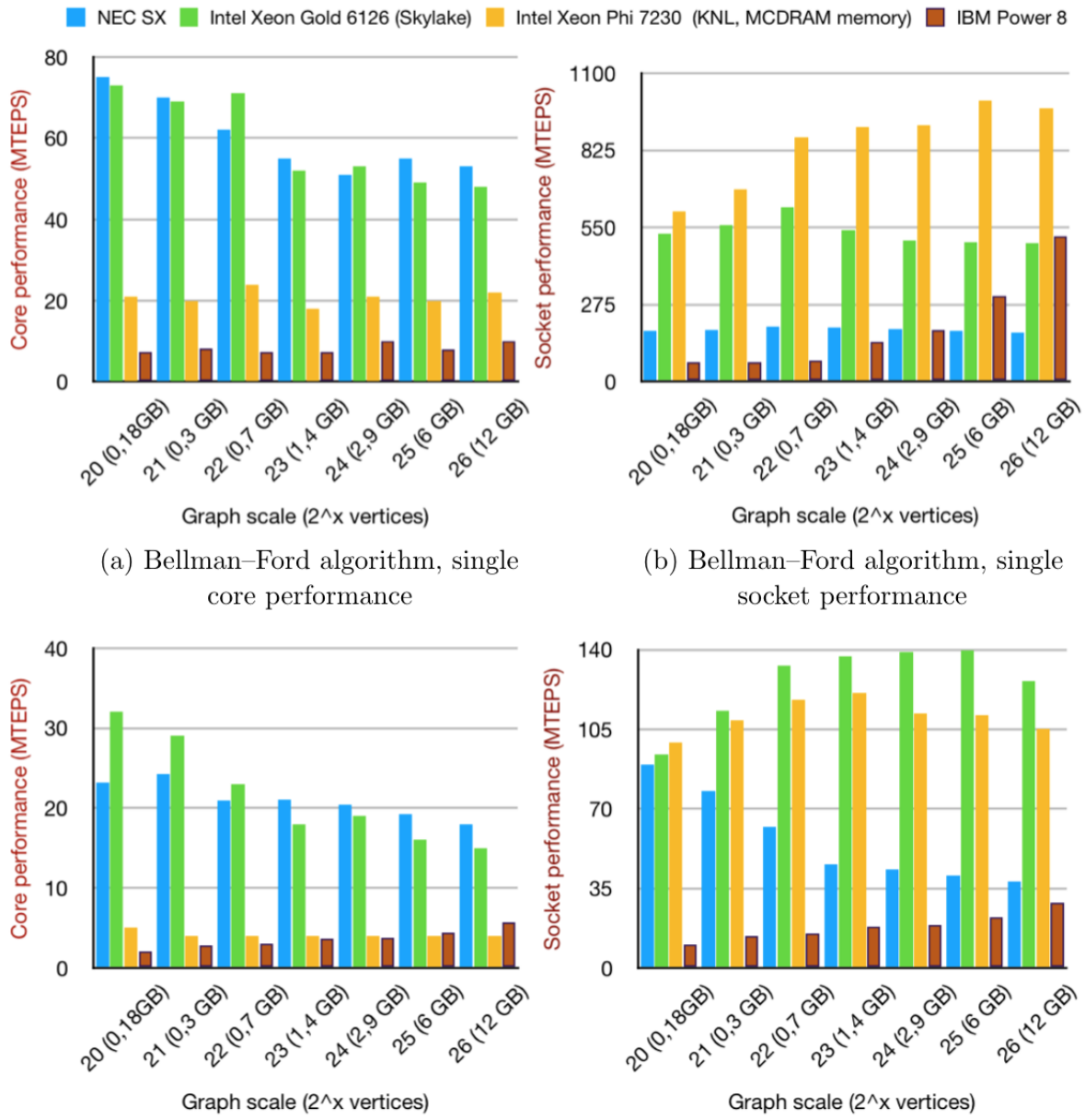
can greatly help to improve data locality and avoid bank conflicts, which results in significantly better performance.

In order to improve data locality, the following sorting strategy is used: input graph edges are sorted in a way that edges, stored in adjacent memory cells, point to adjacent cells in reachability (or distances) arrays. This reordering results into gather and scatter vector operations being much more efficient for adjacent edges (since information is gathered from adjacent cells of memory in distances or reachability array). It is also possible to remove loops and multiple arcs during this pre-processing stage. The proposed optimisation of storage formats provides almost up to a 20x performance improvement compared to the implementation with randomly sorted edges of the input graph.

To achieve efficient parallelisation and vectorisation inside a single NEC SX-ACE socket, several OpenMP and compiler-specific directives are inserted. The *#pragma parallel for* OpenMP directive is used with the *schedule(guided,1024)* OpenMP clause, in order to reduce overheads of synchronisations between threads; *#pragma cdir nodep* directive is used to provide the information to the compiler that there is no dependency (caused by indirect memory accesses) among loop iterations, which can prevent vectorisation. Moreover, *#pragma cdir vprefetch* directive, *#pragma cdir vovertake* and *#pragma cdir vob* directives are used to effectively issue vector gather and scatter instructions, which can easily become bottlenecks, as early as possible by overtaking other vector instructions. To effectively apply the overtakes of the vector gather and scatter instructions, loop unrolling is applied manually. Both implemented algorithms demonstrate high (above 99%) vector operation ratio (percentage of vector operations with maximum length inside the generated code) with an average vector length of 256, which means that the whole code for both algorithms was successfully vectorised.

4. Performance Evaluation

In this paper we compare the developed programs for the NEC SX-ACE computer with implementations for other architectures - Intel Knight Landing, Intel Skylake and IBM Power8. All those architectures support vector data-processing: modern Intel architectures include AVX-512 support, while IBM architecture has VMX instructions support. All implementations for other architectures are also highly optimised, with data locality improvements and vectorisation support. In order to compare various graph algorithms on different platforms, Traversed Edges Per Second (TEPS) metrics is used, which is equal to the number of edges in the input graph, divided on the algorithm execution time. As a result, the performance defined in TEPS is independent from the size and internal structure of an input graph. Figure 1 represent the sustained performance (in TEPS) of both implementations for a single socket and a single core of NEC SX-ACE system, compared to other platforms with vectorisation support: Intel Xeon Gold 6126 (Skylake family) processor, Intel Xeon Phi 7230 (Knight Landing family) and IBM Power 8 processors (RMAT graphs are used). The provided comparison demonstrates that NEC SX-ACE architecture allows to achieve the highest per-core performance among other architectures, but unfortunately lacks per-socket performance, possibly only due to insufficient number of cores. Moreover, the developed implementations demonstrate high parallel efficiency and vector operation ratio, which proves that at least these two graph algorithms can be efficiently vectorised.



(a) Bellman–Ford algorithm, single core performance

(b) Bellman–Ford algorithm, single socket performance

(c) Forward–Backward algorithm, single core performance

(d) Forward–Backward algorithm, single socket performance

Figure 1. Performance evaluation: Bellman–Ford algorithm, single core (a) and single socket (b); Forward-Backward algorithm, single core (c) and single socket (d)

Conclusion

In this paper efficient implementations of two important graph-processing algorithms for NEC SX-ACE platform have been described. The performance values have been compared to the performance of other implementations, obtained on various Intel and IBM platforms capable of vector data processing. The provided comparison demonstrates that NEC SX-ACE architecture allows to achieve the highest per-core performance among other architectures, but unfortunately lacks per-socket performance.

Acknowledgments

The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 and the computational resources of Cyberscience Center at Tohoku University supported by Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures in Japan (Project ID: jh170049-ISJ). The reported study was supported by the Russian Foundation for Basic Research, project No. 18-57-50005.

References

1. Besta, M., Marending, F., Solomonik, E., Hoefler, T.: SlimSell: A vectorizable graph representation for breadth-first search. In: 2017 IEEE International Parallel and Distributed Processing Symposium (IPDPS). IEEE (may 2017), DOI: 10.1109/ipdps.2017.93

2. Fleischer, L.K., Hendrickson, B., Pınar, A.: On identifying strongly connected components in parallel. In: Lecture Notes in Computer Science, pp. 505–511. Springer Berlin Heidelberg (2000), DOI: 10.1007/3-540-45591-4_68

3. Jiang, L., Chen, L., Qiu, J.: Performance characterization of multi-threaded graph processing applications on many-integrated-core architecture. In: 2018 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS). IEEE (apr 2018), DOI: 10.1109/ispass.2018.00033

4. Nepomniaschaya, A.S.: An associative version of the bellman-ford algorithm for finding the shortest paths in directed graphs. In: Lecture Notes in Computer Science, pp. 285–292. Springer Berlin Heidelberg (2001), DOI: 10.1007/3-540-44743-1_28

DOI: 10.14529/jsfi180312

Applications of High Performance Computing: Born–Oppenheimer Molecular Dynamics of Complex Formation in Aqueous Solutions

Dmitry P. Kapusta[1], ***Yulia I. Meteleshko***[1], ***Ilya V. Babchuk***[1], ***Maria G. Khrenova***[1,2]

The progress of supercomputer technologies initiated the development of methods of computational chemistry and their applications, particularly molecular dynamic simulations with *ab initio* potentials. These new methods allow to solve important problems of chemistry and technology. Particularly, solvent extraction and separation techniques are widely used to decrease the amount of radioactive wastes, especially radioactive caesium isotopes present in liquid phases. We demonstrated that the calculated binding constants between the alkali cation and calix[4]arene differ 10^3 times for Cs^+ and Na^+ ions, that is in good agreement with the experimental value. We report the results of benchmark calculations of our model system composed of 929 atoms described in the density functional theory approximation with the GGA-type functional PBE with the empirical dispersion correction D3 and combined basis of Gaussian functions and plane waves DZVP with Goedecker-Teter-Hutter pseudopotentials. We demonstrate that efficiency of calculations decrease to about half if the amount of nodes is 16 on the Lomonosov-2 supercomputer.

Keywords: Born-Oppenheimer molecular dynamic, free binding energy, parallel algorithms, calixarene.

Introduction

The progress of supercomputer technologies initiated the development of methods of computational chemistry and their applications. In this paper we focus on the calculations of the free binding energies of the complex formation between alkali metal cations and calix[4]arene ligand. Calix[4]arene and its derivatives are successfully used in solvent extraction and separation techniques to decrease the amount of radioactive wastes containing caesium isotopes. In the end of 20^{th} century computational works aiming to study these interactions were based on either classical molecular dynamics simulations [6], or *ab initio* calculations in the gas phase [1, 2]. The first approach suffers from the shortcomings of the classical force fields especially when analyzing interactions of the charged particles like metal cations. The second group of methods can better describe the interaction, although they do not account for the entropic contributions important for the binding process and do not explicitly consider solvent molecules.

In the present study we perform molecular dynamics simulations with the DFT (density functional theory) potentials to study the binding processes of sodium and caesium cations with the calix[4]arene ligand considering explicitly the solvation water box. We also perform benchmark tests to evaluate the computational cost of these simulations and suggest the optimal amount of the computational resources for this type of calculations.

1. Models and Methods

The model system comprised 291 water molecules, calix[4]arene with one of the four alcohol groups being deprotonated, and the alkali metal cation (Na^+ or Cs^+), 929 atoms in total. It was

[1]Department of Chemistry, Lomonosov Moscow State University, Moscow, Russian Federation
[2] Federal Research Centre Fundamentals of Biotechnology of the Russian Academy of Sciences, Moscow, Russian Federation

preliminaryly equilibrated to the constant volume of $20.8 \times 19.4 \times 22.3$ Å^3 (Fig. 1) in classical molecular dynamic (MD) simulations. The Born-Oppenheimer MD was performed in the CP2K program package [8].

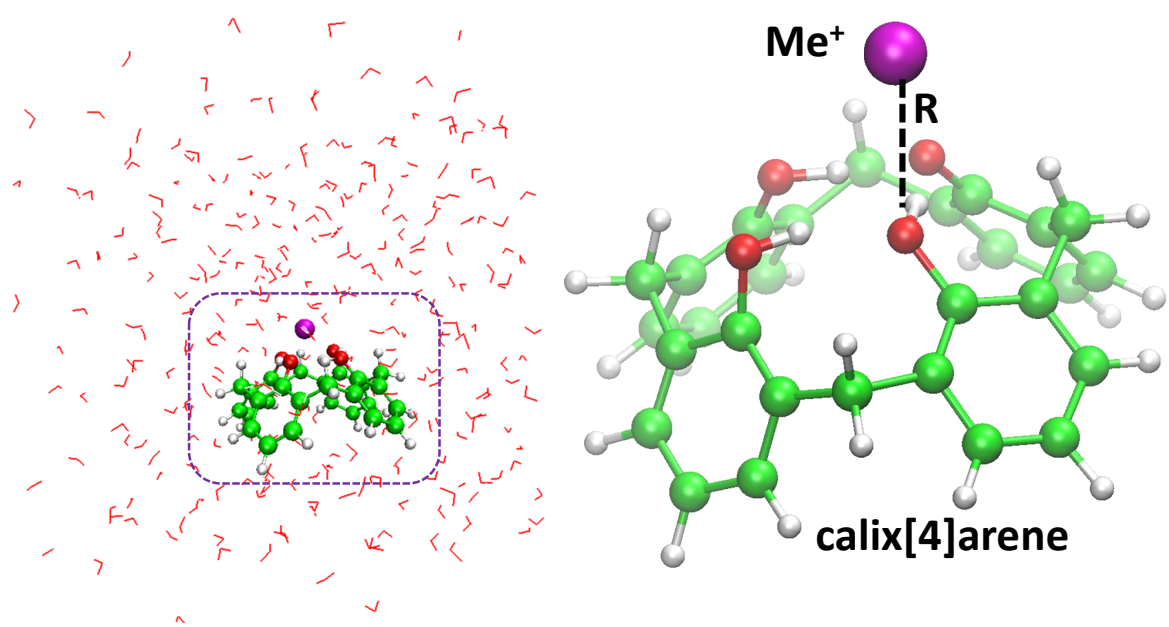


Figure 1. The model system (left) and the reaction coordinate between the alkali metal cation and oxygen atom of calix[4]arene (right). Carbon atoms are shown in green, oxygen in red, metal cation in magenta and hydrogen in white

The entire system was described in the density functional theory approximation with the GGA-type functional PBE [7] with the empirical dispersion correction D3 [5] and combined basis of Gaussian functions and plane waves DZVP with Goedecker-Teter-Hutter pseudopotentials [3]. The calculations were performed with the biasing harmonic potentials centered at the certain values of the reaction coordinate being the distance between the alkali metal cation and one of the oxygen atoms of calix[4]arene. For the system containing caesium cation 16 simulations with the different binding potentials were chosen, and for the system with the sodium cation – 8. Each trajectory was 10 ps with the integration time step of 0.5 fs at T=300 K. The trajectory analysis was performed with the WHAM (weighted histogram analysis method) and UI (umbrella integration) approaches.

2. Results and Discussion

We begin with the benchmark calculations of our model system on the Lomonosov-2 supercomputer (1696 nodes, single-socket Intel Haswell-EP E5-2697v3, 64Gb RAM, NVidia Tesla K40M) [8] (Tab. 1). It is more important to focus on the strong scaling efficiency of the single MD step rather than single SCF step as MD simulations are performed. If the amount of nodes is 16, the efficiency decreases to about half therefore we did not use more than 12 nodes in our calculations (see Tab. 1).

Figure 2 demonstrates the standard Gibbs energy profiles calculated for the sodium and caesium cations. The equilibrium distances between the oxygen atom of the alkali cation and the oxygen atom of calix[4]arene are 2.6 Å and 3.5 Å for sodium and caesium cations, respectively.

Table 1. Strong scaling efficiency of the single SCF (self-consistent field) and MD step calculations of the model system containing caesium cation relative to 1 node

Number of nodes	Time, s		Scaling efficiency	
	SCF step	MD step	SCF step	MD step
1	16	152	-	-
2	8.4	81	95%	94%
4	5.0	46	80%	83%
8	2.8	33	71%	63%
16	2.2	30	61%	63%
16	1.8	22	56%	53%
24	1.6	20	42%	45%
32	1.4	18	36%	40%

It is in line with the ionic radius difference being around 0.7 Å. The calculated Gibbs free energy of complex formation is approximately 6 kcal/mol larger in case of caesium cation resulting in approximately 5×10^4 difference in dissociation constants, that is in a good agreement with the experimental value of 5×10^3 [4].

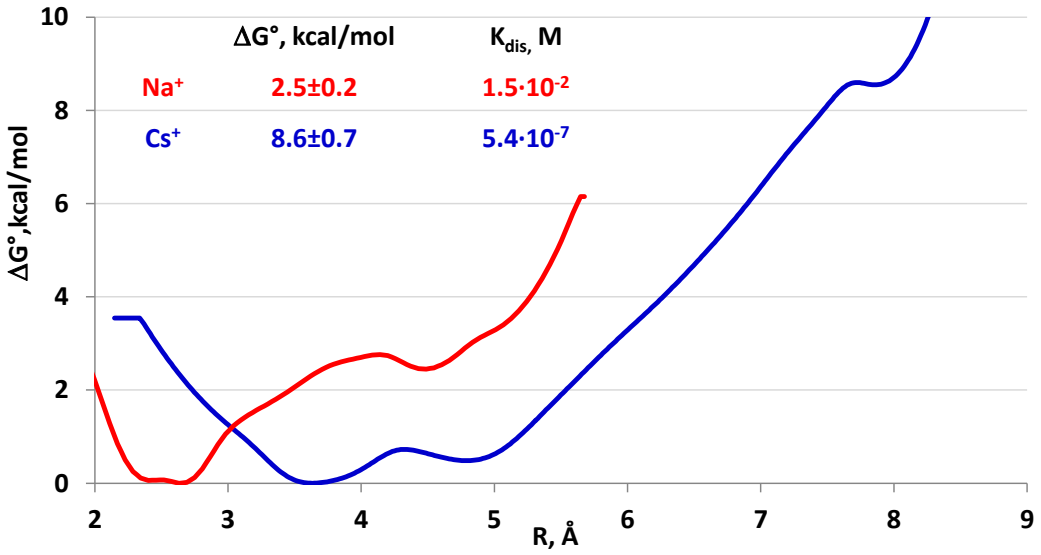


Figure 2. Gibbs free energy profiles of the sodium (red) and caesium (blue) cations binding to the calix[4]arene

Conclusion

We demonstrate that modern supercomputer facilities allow calculating free energy profiles of metal cation – ligand binding in aqueous solution. MD simulations with the DFT potentials implemented in the CP2K program package demonstrate good scalability and correct description of the model systems.

Acknowledgments

This work was supported by the Russian Foundation for Basic Research (project No. 18-03-00605). The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011.

References

1. Bernardino, R., Costa Cabral, B.: Complexation of calix[4]arene with alkali metal cations: Conformational binding selectivity and cation-π driven inclusion. Supramolecular Chemistry 14(1), 57–66 (2002), DOI: 10.1080/10610270290006574

2. Chinta, J.P., Ramanujam, B., Rao, C.P.: Structural aspects of the metal ion complexes of the conjugates of calix[4]arene: Crystal atructures and computational models. Coordination Chemistry Reviews 256(23-24), 2762–2794 (2012), DOI: 10.1016/j.ccr.2012.09.001

3. Goedecker, S., Teter, M., Hutter, J.: Separable dual-space gaussian pseudopotentials. Physical Review B 54(3), 1703–1710 (1996), DOI: 10.1103/PhysRevB.54.1703

4. Gorshkov, V., Ivanov, V., Staina, I.: Selectivity of phenolformaldehyde resins and separation of rare alkali metals. Reactive and Functional Polymers 38(2), 157–176 (1998), DOI: 10.1016/S1381-5148(97)00165-X

5. Grimme, S., Antony, J., Ehrlich, S., Krieg, H.: A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. The Journal of Chemical Physics 132(15), 154104 (2010), DOI: 10.1063/1.3382344

6. Guilbaud, P., Varnek, A., Wipff, G.: Molecular dynamics study of p-tretbutylcalix[4]arenetetraamide and its complexes with neutral and cationic guests. influence of solvation on structures and stabilities. Journal of the American Chemical Society 115(18), 8298–8312 (1993), DOI: 10.1021/ja00071a044

7. Perdew, J.P., Burke, K., Ernzerhof, M.: Generalized gradient approximation made simple. Phys. Rev. Lett. 77, 3865–3868 (Oct 1996), DOI: 10.1103/PhysRevLett.77.3865

8. Sadovnichy, V., Tikhonravov, A., Voevodin, Vl., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale toward Exascale, pp. 283–307 (2013)

DOI: 10.14529/jsfi180313

Analysis of the Effect of Dispersion Forces on the Dielectric Film Properties Using Parallel Computing

Kirill A. Emelyanenko[1], ***Ludmila B. Boinovich***[1], ***Alexandre M. Emelyanenko***[1]

The paper presents the analysis of dispersion forces effect on local properties in thin free films. Using a Coupled Fluctuated Dipole Method with developed methods for numerical calculations of dielectric properties, the films with different lateral sizes and thicknesses were studied. In particular, the molecular polarizabilities at different distance from the film interface were analyzed. It was shown that dispersion interaction between the molecules, even for the case of nonpolar liquid with weak intermolecular interactions, causes a notable variation in dielectric properties of thin film, which is associated with the boundary layer formation. This variation, in turn, causes a strong dependence of polarizability accuracy on the cut-off radius. It is demonstrated that parallel computing algorithms can be effectively applied for obtaining the reliable data on properties of liquids in wetting films and boundary layers even under resource-imposed constraint on the size of ensemble of molecules to be handled in the numerical studies.

Keywords: CFDM, thin film properties, dispersion forces, local polarizability.

Introduction

The properties of contacting phases vary continuously upon transition from one phase to another. This variation of different properties of both phases caused the appearance and the development of a concept of boundary layer, as the layer adjacent to the interface, the anomalous properties of which result from the manifestation of long-range effects from the confining body [1]. The investigations of rheological and optical properties of boundary layers of liquids in the vicinity of interfaces as well as in wetting films indicated special structure and anisotropy of boundary layers extending from the interface into a liquid phase far beyond the monomolecular dimensions. The mechanisms of such effects are poorly studied so far. As discussed in [1] the most significant effect of deviation of boundary layer properties from those in the bulk is expected for polar liquids strongly interacting with the confining phase. In this study, we show that the formation of boundary layer is a universal phenomenon induced by the van der Waals interaction between the molecules. Using a nonpolar hexane we show the significant variation of its molecular polarizability near interface with air. Besides, it is demonstrated that parallel computing algorithms can be effectively applied for obtaining reliable data on optical properties of liquids in wetting films and in boundary layers near single interfaces even under resource-imposed constraint on the size of ensemble of molecules to be handled in the numerical studies.

1. Coupled Fluctuated Dipoles Method

Coupled fluctuated dipoles method (CFDM) is a microscopic method for calculation of the dispersion interactions. Introduced by Renne and Nijboer [4] in 1967, it considers a system of interacting bodies as an aggregation of three-dimensional oscillators; each oscillator can be associated with an individual atom, a molecule or an elementary unit of the corresponding material. At the time of invention, the applicability of CFDM was hindered by lack of computational power. However, it was revived in series of articles by Kim and Velegol in the recent decade.

[1]A.N. Frumkin Institute of Physical Chemistry and Electrochemistry RAS, Moscow, Russian Federation

Under the action of an external field $\bar{\boldsymbol{E}}$ on the harmonic oscillator with radius-vector $\bar{r}_i$ a dipole moment $\bar{\boldsymbol{p}}_i$ is induced:

$$\bar{\boldsymbol{p}}_i = \alpha_i \bar{\boldsymbol{E}}(\bar{r}_i), \tag{1}$$

where α_i is external field frequency dependent polarizability of i-th oscillator.

This induced dipole moment creates electric field by itself which affects the neighbouring atoms. Thus, for every oscillator in condensed material full external electric field will consist of a macroscopic field and local fields of other oscillators:

$$\bar{\boldsymbol{E}}(r_i) = \bar{\boldsymbol{E}}_{\boldsymbol{0}}(\bar{r}_i) + \sum_{j=1}^{N} \boldsymbol{T}_{ij} \bar{\boldsymbol{p}}_j; \quad \boldsymbol{T}_{ij} = \frac{3\bar{\boldsymbol{n}}_{ij}\bar{\boldsymbol{n}}_{ij} - \boldsymbol{I}}{r_{ij}^3}, \tag{2}$$

where $\bar{\boldsymbol{n}}_{ij}$ and r_{ij} are unit normal vector and distance between elementary units i and j.

Considering all oscillators in the system we get a system of equations for each i-th oscillator:

$$\bar{\boldsymbol{p}}_i - \alpha_i \sum_{j=1}^{N} \boldsymbol{T}_{ij} \bar{\boldsymbol{p}}_j = \alpha_i \bar{\boldsymbol{E}}_{\boldsymbol{0}}(\bar{r}_i). \tag{3}$$

Particular solution of this system of inhomogeneous equations can be found if we note that for static electric field α_i should be substituted with $\alpha_{(0)i}$. This solution corresponds to static dipole moments induced by an external field. These individual dipole moments can be used for finding local polarizability and other local dielectric properties.

1.1. Numerical Methods

Basic computational routines required by CFDM for calculation of both the dipole moments and the frequency spectra are standard and can be found in `LAPACK` and `ScaLAPACK` libraries. However, since systems of interest for microscopic dispersion interactions consist of tens and hundreds of thousands of atoms, a set of special numerical methods should be used.

Studying the system infinite in one or more directions can be done much more effectively, that is, with substantially less amount of computation, by using periodic boundary conditions. It eliminates the impact of spatial confinement in undesired dimension; for instance, for film studies it eliminates effects of lateral constraints, while periodicity length effectively becames a cut-off radius, that is, a distance beyond which interactions are neglected. Another approach which substantially decreases the size of equations system for calculating individual dipole moments, is exploiting symmetries in the systems. Since for some oscillators, a translation symmetry leads to equal moments, these oscillators can initially be separated into equivalence classes with decreasing number of equations from *one-per-oscillator* to *one-per-class*. It becomes an especially effective technique in combination with the abovementioned periodic boundary conditions. Finally, iterative methods of finding eigenfrequencies by calculating traces of interaction matrix in corresponding exponent enable trade-off between accuracy and performance.

2. Results and Discussion

In this paper, we discuss approaches for numerical studying of influence of surface forces caused by dispersion interactions on distribution of dielectric properties inside the films. As a model system, free hexane film was chosen, since it was shown [2] that for this system CFDM predicts quantitatively correct properties, in particular, the values of Hamaker constants. As

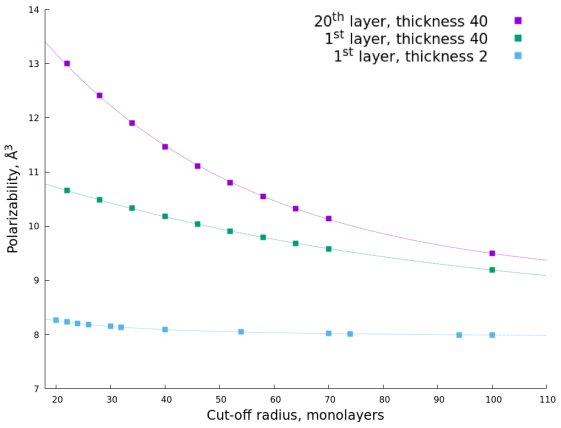


Figure 1. Characteristic dependence on cut-off radius for polarizability of 1^{st} layer in bilayer (blue) and 40 layer (green) system; polarizability of 20^{th} layer in film 40 monolayers thick (violet)

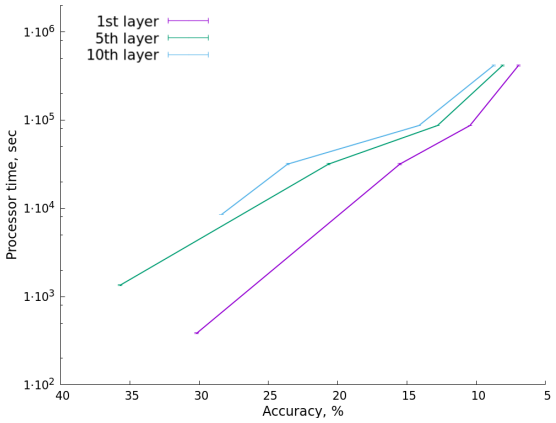


Figure 2. Processor time required for obtaining given accuracy of local polarizability values for different layers of the free film. Calculations were performed on 256 cores at Lomonosov supercomputer

demonstrated previously [3], in absence of periodical boundary conditions, long-range nature of dipole-dipole interactions hinders the applicability of microscopic approaches for evaluating properties of bulk materials or systems infinite in one or more dimensions. Thus, for our case when laterally infinite films were simulated by periodical boundary conditions, it was important to study how cut-off radius affects the results, and which periodicity length is optimal under trade-off between the computing performance, defined by the size of the system, and the accuracy.

Typical dependences of local polarizability (in the layer adjacent to the interface and in the film midplane) on cut-off radius for films with different thicknesses are shown in Fig. 1. The notable difference detected in the values for the first monolayer adjacent to hexane/air interface in the bilayer film and for the film of 40 monolayers thick convincingly illustrates the effect of boundary layers overlapping on the local polarizabilities. The dispersion interactions cut-off associated with setting the periodic boundary conditions for the finite lateral size of film result in notable variation of the numerically obtained values of local polarizability. As expected, films with higher thicknesses require much larger lateral sizes to reach properties corresponding to laterally infinite films. It was found that approaching to infinite values is exponential with convergence rate dependent both on film thickness and layer depth.

Since scaling properties of CFDM method trivially reflect scaling properties of underlying linear algebra routines, we do not present here dependence of computation time on number of processors. However, from practical point of view it is important to understand parameters of perfomance/accuracy trade-off. Processor time required for reaching given accuracy of polarizability values is presented in Fig. 2 for different locations inside the film. Thus, for studying the properties of relatively thick films it is necessary to use supercomputer clusters with parallel linear algebra routines. In our work, we have performed calculation on HPC supercomputer cluster at Lomonosov Moscow State University [5] with `ScaLAPACK` routines. We calculated the variation of polarizability, characteristic of the individual molecular layers across the film (Fig. 3). It can be seen that the influence of surface forces on the film properties is not limited by layers adjacent to the interface, but propagates deep inside the film causing significant deviation of the local polarizability from the bulk values.

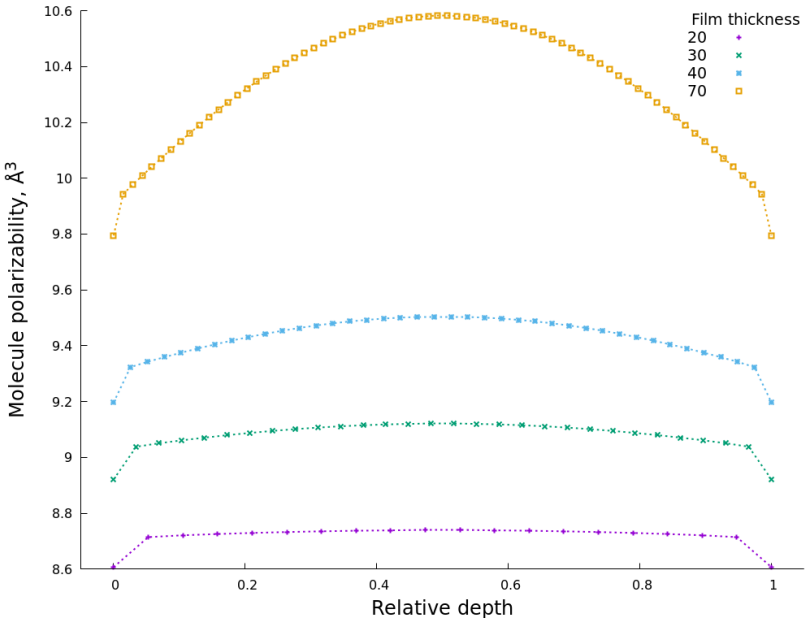


Figure 3. Polarizability distribution in the free hexane films with different thicknesses. Cut-off radius is 100 monolayers

Conclusions

The developed method of numeric calculations of the film properties based on the values of polarizability of individual molecules allows to obtain reliable results for the films of different lateral sizes and thicknesses. Using this approach we have shown that dispersion interaction between the molecules causes notable variation in dielectric properties of thin liquid film even in case of nonpolar liquid with weak intermolecular interactions.

Acknowledgments

The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011.

References

1. Derjaguin, B., Churaev, N., Muller, V.: Surface Forces (Consultants Bureau, New York, 1987). Springer, Boston, MA (1987), DOI: 10.1007/978-1-4757-6639-4

2. Emelyanenko, K.: Influence of Dicrete Nature of Charge and Material on the Surface Forces in Nanosystems (rus.). PhD dissertation (2018)

3. Kwaadgras, B.W., Verdult, M., Dijkstra, M., van Roij, R.: Polarizability and alignment of dielectric nanoparticles in an external electric field: Bowls, dumbbells, and cuboids. The Journal of Chemical Physics 135(13), 134105 (2011), DOI: 10.1063/1.3637046

4. Renne, M., Nijboer, B.: Microscopic Derivation of Macroscopic Van der Waals Forces. Chemical Physics Letters 1(8), 317–320 (1967), DOI: 10.1016/0009-2614(67)80004-6

5. Sadovnichy, V., Tikhonravov, A., Voevodin, V., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale toward Exascale. pp. 283–307. Chapman & Hall/CRC Computational Science, CRC Press, Boca Raton, United States (2013)

DOI: 10.14529/jsfi180314

Quantum Chemistry Research of Interaction between 3D-Transition Metal Ions and a Defective Graphene on the Supercomputer Base

Nikolai V. Khokhriakov[1], *Santiago Melchor*[2]

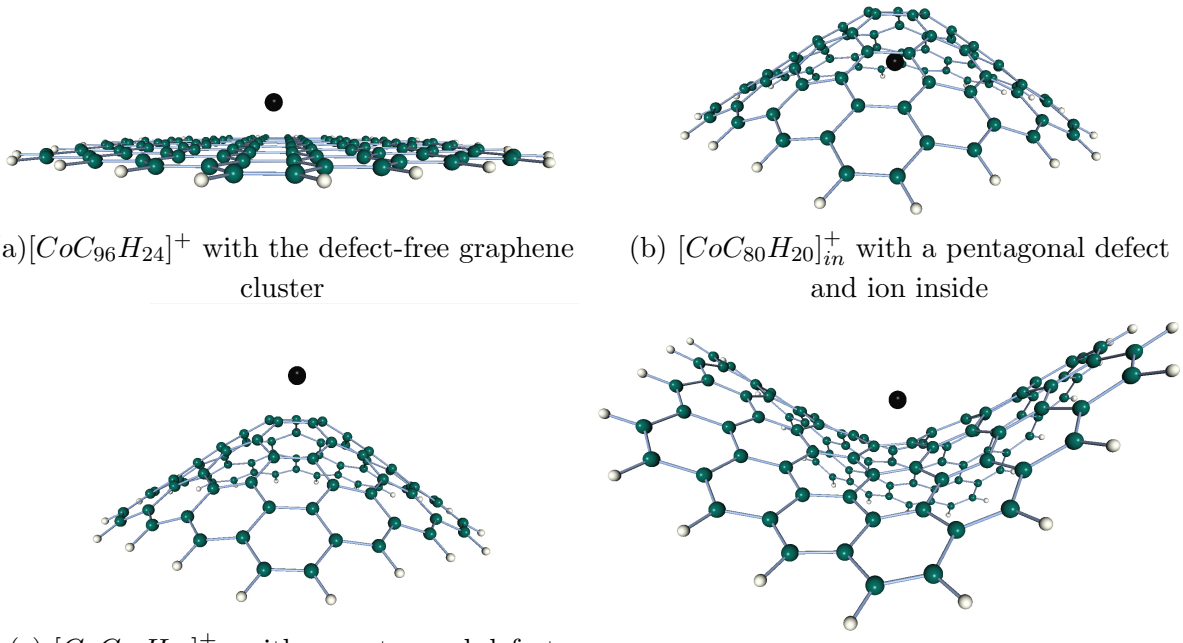
Quantum chemistry research is presented in the article, and it concerns the interaction within the complexes formed by the defective graphene clusters and ions of 3d-transition metals $V, Cr, Mn, Fe, Co, Ni, Cu$. The charges of all regarded ions were +1. All calculations were made at UDFT B3LYP/6-31G level of theory with the BSSE error taken into account. The strongest interaction with the defective clusters is observed in the case of Co^+ ion. At the same time, this ion has demonstrated rather weak interaction with the defect-free graphene. Thus, the presence of Co^+ in the reaction media increases probability of defect formation with the further forming of short nanotubes and curved carbon clusters with complex topology of their own.

Keywords: graphene, transition metals, quantum chemistry, density functional.

In the last thirty years, properties of nanoparticles and nanostructured media have been attracting unrelenting interest in scientific community. Computer experiments are necessary to search for synthesis conditions to obtain the particles with predetermined structures and properties because of their complexity and possible formation of short-lived states influencing a synthesis process. The characteristic size of a nanoparticle can reach tens of nanometers, and the model system may include millions of atoms in order to reflect correctly its geometrical features and chemical behavior. The computing complexity of the classical algorithms of quantum chemistry based on calculation of nanoparticles electronic Hamiltonian eigenfunctions begins from $O(N^3)$ for the simplest semi-empirical tight-binding model (where N is the quantity of atomic orbitals in a basis set). Computing demands increase dramatically with improving of quantum chemistry model. The problem becomes NP-complete at accurate accounting of electronic exchange and correlation. Since the end of the 20th century quantum chemistry algorithms with complexity $O(N)$ have been developing, but the coefficient of proportionality for similar algorithms is rather large. Thus, the advantage of linear-scaling algorithms is reached only for the systems with a large amount of atoms. Therefore, the quantum chemistry research studies in the field of nanotechnologies are possible only when using the capacities of supercomputers.

The interaction of transition metal ions with a defective graphene is considered in the article. An ideal graphene is a plane packed by hexagons with the carbon atoms in their corners. Replacement of one of the hexagons with a pentagon during the growth of the graphene plane leads to formation of conic surface, and replacing with a heptagon results in formation of saddle surface. The system of topological defects can be constructed to obtain a junction of several carbon nanotubes of various types, or a closed shell of fullerene. The defective areas in a graphene structure demonstrate higher chemical activity; therefore, the defective graphene nanostructures are perspective in chemical technologies. Understanding of interaction in complexes formed by the graphene and transition metal ions is of special interest for some practical applications. Such situation is implemented, for instance, when using graphene for cleaning water off ions of

[1]Izhevsk State Agricultural Academy, Izhevsk, Russian Federation
[2]Universidad de Granada, Granada, Spain



(a)$[CoC_{96}H_{24}]^+$ with the defect-free graphene cluster

(b) $[CoC_{80}H_{20}]^+_{in}$ with a pentagonal defect and ion inside

(c) $[CoC_{80}H_{20}]^+_{out}$ with a pentagonal defect and ion outside

(d) $[CoC_{112}H_{28}]^+$ with a heptagonal defect

Figure 1. Optimized complexes of cobalt ion with graphene clusters

transition metals [9]; when using graphene clusters as sensors for definition of transition metals in water environment [2]; during synthesis of carbon nanostructures in molten salts of transition metals [5]; at contact of a graphene with oxide of transition metal in the anode of a lithium ion battery [10]. And, probably what is of most interest, the matrix of a defective graphene with metal atoms fixed on the defects is considered as promising single-atom catalyst [3].

All calculations were made at the UDFT B3LYP/6-31G level of theory. Minimal basic set m6-31G was used adapted for compounds of transition metals [6]. The method was examined by calculation of chemically similar systems formed by transition metal ions and benzene [4]. Calculation results for the interaction energies between ions and a molecule of benzene are in complete accordance with the available experimental data.

An infinite graphene leaf with a single defect polygon was modelled by a finite size cluster constructed from the defect polygon in the center, the latter surrounded by three strips of carbon hexagons. Dangling bonds on the boundaries were compensated by hydrogen atoms. Further on, such graphene clusters have been considered. Including the cluster with pentagonal defect of a cone form, a saddle-like cluster with heptagonal defect, and a plane defect-free cluster with a carbon hexagon in the center. The complexes were considered formed by each graphene cluster and ions of 3d-transition metals $V, Cr, Mn, Fe, Co, Ni, Cu$. The charges of all ions were +1.

Initially, every ion was placed above the center of every graphene cluster. For a cluster with a pentagonal defect, two situations had been considered: a case when the ion of metal lies inside the cone (such systems are designated further by the index “in”), and the case when the ion is located outside (designated further by the index “out”). Further, the geometry optimization of the complex was completed. Optimized complexes with the cobalt ion are shown in Fig. 1. Green spheres designate carbon atoms, the light ones hydrogens, and cobalt ion is black.

Interaction energy E_{int} between metal ion and a carbon cluster has been estimated by the formula

$$E_{int} = E_{Me} + E_{Gr} - E_{MeGr}, \tag{1}$$

where E_{MeGr} is the energy of the complex after optimization, E_{Me} is the energy of the isolated ion, E_{Gr} is the energy of the isolated carbon cluster. E_{int} was calculated taking into account the BSSE correction.

Table 1. Interaction energy between transition metal ions and graphene clusters (in kJ/mol)

Transition metal ion	Spin multiplicity	$C_{96}H_{24}$	$[C_{80}H_{20}]_{out}$	$[C_{80}H_{20}]_{in}$	$C_{112}H_{28}$
Co^+	3	222	403	281	407
Cr^+	6	216	254	185	262
Cu^+	1	273	346	218	348
Fe^+	4	327	342	300	386
Mn^+	7	206	242	158	241
Ni^+	2	253	346	248	345
V^+	5	238	270	216	291

The computed results for interaction energies between transition metal ions and carbon clusters are given in Tab. 1. Almost for all atoms, the weakest interaction is observed when placing an ion inside a conic cluster with the pentagonal defect. The strongest interaction is observed for complexes including carbon clusters with a heptagonal defect. The interaction between conic graphene clusters and an ion of metal located outside of the cone is also strong. The negative charge usually has the maximum density near the vertex outside the cone, and that explains the raised attraction of a positive ion. At the same time, inside the cone the density of the negative charge rapidly decreases while departing the vertex. The metal ion located inside the conic cluster is not able to be close to the vertex because of high density of carbon atoms in the internal area and repulsion of their internal electron shells. Therefore, the ion arrangement inside the conic cluster is energetically unfavorable. In case of the heptagonal defect, the saddle surface is being formed with a number of areas, thus having positive and some of them negative charge. Geometrically, the ion of metal is less constrained than inside the cone; therefore, it acquires wide opportunities for the movement in space, thus resulting in energy decrease. In case of a planar defectless graphene cluster weak interaction with a metal ion can be observed and explained by high stability of the graphite structure, that in this case remains safe.

The calculations show that the strongest interaction with the defective graphene clusters is observed for cobalt ion. At the same time, this ion rather weakly interacts with a defect-free graphene fragment. Therefore, presence of cobalt ions in the reaction media can increase probability of the defect formation. As a result, shorter nanotubes are formed as a result of synthesis, and the graphene leaf can get complex and curved topology.

Acknowledgements

The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 [7] and the program FireFly QC [1] developed on the basis of GAMESS (US) [8].

References

1. Granovsky, A.A.: Firefly version 8, `http://classic.chem.msu.su/gran/firefly/index.html`, accessed: 2018-09-30

2. Huang, H., Liao, L., Xu, X., Zou, M., Liu, F., Li, N.: The electron-transfer based interaction between transition metal ions and photoluminescent graphene quantum dots (GQDs): A platform for metal ion sensing. Talanta 117, 152–157 (2013), DOI: 10.1016/j.talanta.2013.08.055

3. Jiang, K., Wang, H.: Electrocatalysis over graphene-defect-coordinated transition-metal single-atom catalysts. Chem 4(2), 194–195 (2015), DOI: 10.1016/j.chempr.2018.01.013

4. Khokhriakov, N.V., Kodolov, V.I., Karpova, V.S.: Quantum chemistry research of transition metals complexes with aromatic hydrocarbons. Chemical physics and mesoscopy 16(4), 622–626 (2014)

5. Kodolov, V.I., Kuznetsov, A.P., Nikolaeva, O.A., Shayakhmetova, E.Sh., Makarova, L.G., Shabanova, I.N., Khokhriakov, N.V., Volkova, E.G.: Investigation of metal-carbon tubulenes by x-ray photoelectron spectroscopy and electron microscopy. Surface and interface analysis 32(1), 10–14 (2001), DOI: 10.1002/sia.996

6. Mitin, A.V., Baker, J., Pulay, P.: An Improved 6-31G* Basis Set for first-row transition metals. The journal of chemical physics 118(17), 7775–7782 (2003), DOI: 10.1063/1.1563619

7. Sadovnichy, V., Tikhonravov, A., Voevodin, Vl., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In contemporary high performance computing: from petascale toward exascale. pp. 283–307. Chapman & Hall/CRC Computational Science, CRC Press, Boca Raton, United States (2013)

8. Schmidt, M.W., Baldridge, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.H., Koseki, S., Matsunaga, N., Nguyen, K.A., Su, S., Windus, T.L., Dupuis, M., Montgomery, J.A.: General atomic and molecular electronic structure system. Journal of computational chemistry 14(11), 1347–1363 (1993), DOI: 10.1002/jcc.540141112

9. Yu, J.-G., Yu, L.-Y., Yang, H., Liu, Q., Chen, X.-H., Jiang, X.-Y., Chen, X.-Q., Jiao, F.-P.: Graphene nanosheets as novel adsorbents in adsorption, preconcentration and removal of gases, organic compounds and metal ions. Science of the total environment 502, 70–79 (2015), DOI: 10.1016/j.scitotenv.2014.08.077

10. Zhu, J., Duan, R., Zhang, S., Jiang, N., Zhang, Y., Zhu, J.: The application of graphene in lithium ion battery electrode materials. SpringerPlus 3, 585 (2014), DOI: 10.1186/2193-1801-3-585

DOI: 10.14529/jsfi180315

A Flux Splitting Method for the SHTC Model for High-performance Simulations of Two-phase Flows

Nadezhda S. Smirnova[1], ***Michael Dumbser***[1,2], ***Mikhail N. Petrov***[1], ***Alexander V. Chikitkin***[1], ***Evgeniy I. Romenski***[3]

In this paper we propose a new flux splitting approach for the symmetric hyperbolic thermodynamically compatible (SHTC) equations of compressible two-phase flow which can be used in finite-volume methods. The approach is based on splitting the entire model into acoustic and pseudo-convective submodels. The associated acoustic system is numerically solved applying HLLC-type Riemann solver for its Lagrangian form. The convective part of the pseudo-convective submodel is solved by a standart upwind scheme. For other parts of the pseudo-convective submodel we apply the FORCE method. A comparison is carried out with unsplit methods. Numerical results are obtained on several test problems. Results show good agreement with exact solutions and reference calculations.

Keywords: flux splitting, two-phase compressible flow, complete Riemann solver, finite-volume method, hyperbolic equations, supercomputer computations.

Introduction

Modeling of two-phase compressible flows finds many applications in various engineering spheres. However, the research of two-phase models is still a challenging area of computational fluid dynamics. The numerical investigation of these problems requires powerful computing resources and therefore parallel calculations. Nowadays mathematical models of this class of problems and their computational methods are actively developed. The most widely used two-phase models are the Baer-Nunziato model [1], the Kapila model [3] and the SHTC model [4]. The main advantage of these models is the hyperbolicity of the governing equations that allows to apply well-known methods for this type of equations. The key disadvantage of the Baer-Nunziato and the Kapila models is that they are of non-conservative form, while the SHTC equations can be written in the conservation-law form. This shortcoming of first two models leads to difficulties in the definition of the discontinuous solutions and in the development of high order numerical methods. In this paper we introduce a new method for solving the SHTC equations of compressible two-phase flow with one common entropy. The present method is based on the original method for the Kapila equations [2]. The aim of the present work is to develop a method which allows efficient parallelization and provides reliable numerical solutions.

1. Numerical Method

In this paper we consider the governing partial differential SHTC equations of compressible two-phase flow in one-dimensional case [4]. The approach with one common entropy S for two phases is applied for the description of thermal effects. We study flows of water-gas and gas-gas mixtures, for which the equations of states are presented in [4].

We apply a splitting-based method for the considered system. First, we identify the corresponding acoustic system, which includes the same equations for mixture mass, mixture mo-

[1]Moscow Institute of Physics and Technology, Moscow, Russia
[2]University of Trento, Trento, Italy
[3]Sobolev Institute of Mathematics, Novosibirsk, Russia

mentum, mixture energy as presented in the acoustic system in [2] with adding two equations: $\partial_t(\alpha_1) = 0$ and $\partial_t(c_1) = 0$, and the pseudo-convective system

$$\begin{aligned}
&\partial_t\rho + u\partial_x\rho = 0, \quad \partial_t(\rho u) + \partial_x(\rho w E_w) + u\partial_x(\rho u) = 0,\\
&\partial_t(\rho E) + \partial_x(\rho E_w(uw + E_{c_1})) + u\partial_x(\rho E) = 0,\\
&\partial_t(\rho\alpha_1) + u\partial_x(\rho\alpha_1) = -\lambda(p_2 - p_1),\\
&\partial_t(\rho c_1) + \partial_x(\rho E_w) + u\partial_x(\rho c_1) = 0,\\
&\partial_t w + \partial_x(E_{c_1}) + u\partial_x w = -\chi w.
\end{aligned} \tag{1}$$

The intermediate time $(n+1-)$ update formulae of the acoustic system in Eulerian variables is the following

$$\begin{aligned}
&R_j\rho_j^{n+1-} = \rho_j^n, \quad R_j(\rho u)_j^{n+1-} = (\rho u)_j^n - \frac{\Delta t}{\Delta x}(p^*_{j+1/2} - p^*_{j-1/2}),\\
&R_j(\rho E)_j^{n+1-} = (\rho E)_j^n - \frac{\Delta t}{\Delta x}(p^*_{j+1/2}u^*_{j+1/2} - p^*_{j-1/2}u^*_{j-1/2}),\\
&(\alpha_1)_j^{n+1-} = (\alpha_1)_j^n, \quad (c_1)_j^{n+1-} = (c_1)_j^n, \quad w_j^{n+1-} = w_j^n - w_j^n\frac{\Delta t}{\Delta x}(u^*_{j+1/2} - u^*_{j-1/2}),
\end{aligned} \tag{2}$$

where $R_j = 1 + \frac{\Delta t}{\Delta x}(u^*_{j+1/2} - u^*_{j-1/2})$, p^* u^* – pressure and velocity in the Star Region [6], which were constructed using HLLC-type Riemann solver [7] for the acoustic system in Lagrangian coordinates. More detailed description of solving similar system is given in [2].

The resulting conservative vector of the next time level $(\phi_j^{n+1})^T = (\rho, \rho u, \rho E, \rho\alpha_1, \rho c_1, w)_j^{n+1}$ is obtained from the discretization of the pseudo-convective system using finite-volume method:

$$\begin{aligned}
\phi_j^{n+1} = \phi_j^{n+1-} - \frac{\Delta t}{\Delta x}(u^*_{j+1/2}\phi^{n+1-}_{j+1/2} - u^*_{j-1/2}\phi^{n+1-}_{j-1/2}) +\\
+\frac{\Delta t}{\Delta x}\phi_j^{n+1-}(u^*_{j+1/2} - u^*_{j-1/2}) - \frac{\Delta t}{\Delta x}(F_{j+1/2} - F_{j-1/2}),
\end{aligned} \tag{3}$$

where the numerical intercell flux $F_{i+1/2}$ is obtained applying the FORCE method [6] and the convective terms of the equations (1) are approximated with the upwind rule using corresponding velocity u^* to find the direction of the convective flux.

2. Numerical Results

In this section we test the performance of the presented flux splitting method in order to verify accuracy and correctness on several Riemann problems. The flux splitting method is compared with direct methods (Rusanov [6] in tests $1-3$ and GFORCE [8] in test 3). The numerical results for first phase volume fraction for tests $1-2$ are shown in Fig. 1, and for both first phase volume fraction and mixture density for test 3 are presented in Fig. 2. The numerical solutions are computed in the spatial domain $0 \le x \le 1$ using the mesh of $M = 200$ cells. Transmissive boundary conditions are applied. The calculations have been run on Lomonosov-2 system of Moscow State University using up to 72 CPU cores.

In the first test [6] we consider two perfect gases with the same properties. The initial data is set to the left and to the right of the discontinuity position x_0 as $(\alpha_1, \rho_1, \rho_2, u_1, u_2, S)^T_L = (0.8, 1.0, 1.0, -2.0, -2.0, -654.23158)$ and $(\alpha_1, \rho_1, \rho_2, u_1, u_2, S)^T_R = (0.5, 1.0, 1.0, 2.0, 2.0, -654.23158)$. The phase parameters are taken as follows: $\rho_{01} = \rho_{02} = 1.0$, $\gamma_1 = \gamma_2 = 1.4$, $C_1 = C_2 = 1.18322$, $c_{v1} = c_{v2} = 714$. The solution of this test consists of two

symmetric rarefaction waves and trivial stationary contact wave. The Star Region between two rarefaction waves is close to vacuum, hence this problem is appropriate for accessing relevant numerical method for low-density flows. The flux splitting scheme allows to resolve volume fraction significantly better than the Rusanov scheme.

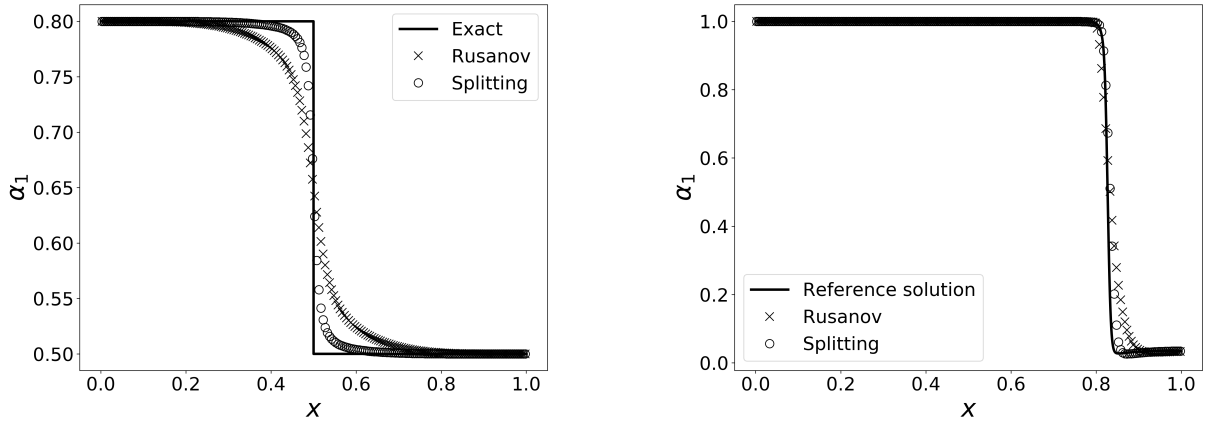


Figure 1. Test1 (left) and test2 (right). Comparison of numerical solutions computed by the Rusanov and the present flux splitting method with the exact solution at time $t = 0.15$ and $x_0 = 0.5$ for test1, and with the reference solution at time $t = 229 \cdot 10^{-4}$ and $x_0 = 0.7$ for test2

In the second test we study water-air flow. The formulation of this test is close to diffuse interface problems, which have one of phases volume fraction nearly to unity and the another to zero. The initial data are: $(\alpha_1, \rho_1, \rho_2, u_1, u_2, S)^T_L = (0.995, 1000.0, 50.0, 0.0, 0.0, 932.76862)$ and $(\alpha_1, \rho_1, \rho_2, u_1, u_2, S)^T_R = (0.005, 1000.0, 50.0, 0.0, 0.0, 4309.77059)$. The phase parameters are given by $\rho_{01} = 1000$, $\rho_{02} = 1.0$, $\gamma_1 = 4.4$, $\gamma_2 = 1.4$, $C_1 = 1624.80768$, $C_2 = 1.18322$, $c_{v1} = 951$, $c_{v2} = 714$, $p_{01} = 0$. The reference solution is obtained by using the Rusanov solver on 1000 mesh cells. The flux splitting scheme shows sharper resolution of shock wave structure than Rusanov scheme.

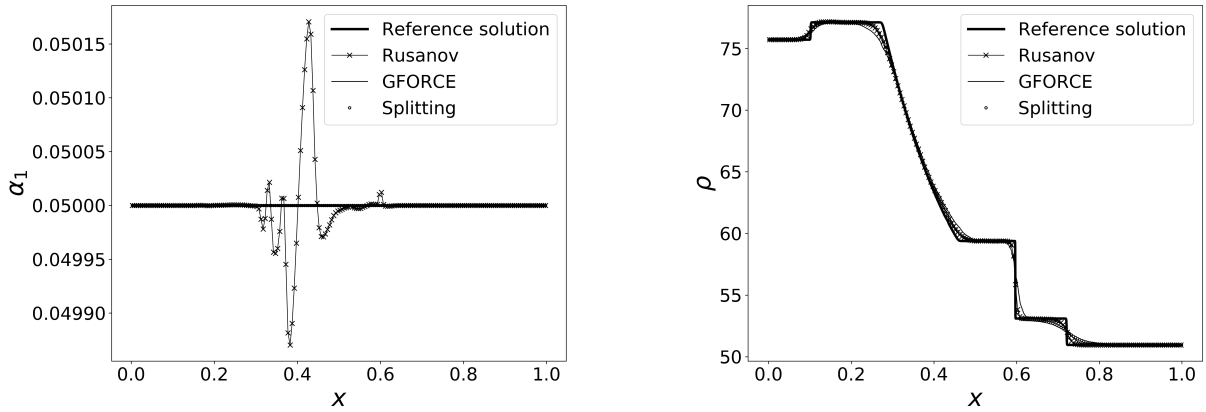


Figure 2. Test3. Comparison of numerical solutions computed by the Rusanov, the GFORCE and the present splitting with the reference solution at time $t = 2 \cdot 10^{-4}$ and $x_0 = 0.5$

In the third test, called sonic point test problem [4], we also investigate water-air flow. This test is suitable for analysing of the entropy satisfaction property of numerical methods. We consider the isentropic model, when the entropy S is constant and equal to zero. The initial data are set as $(\alpha_1, \rho_1, \rho_2, u_1, u_2)^T_L = (0.05, 1004.18441, 26.84394, 100.0, 100.0)$ and $(\alpha_1, \rho_1, \rho_2, u_1, u_2)^T_R = (0.05, 1000.04200, 1.00063, 0.0, 0.0)$. The chosen phase parameters are:

$\rho_{01} = 1000$, $\rho_{02} = 1.0$, $\gamma_1 = 2.8$, $\gamma_2 = 1.4$, $C_1 = 1543$, $C_2 = 374$, $p_{01} = 0$. We ignore source terms in the considered system [4] in order to compare results obtained by the flux splitting, the Rusanov and the GFORCE methods with the reference solution, which corresponds to the GFORCE flux computed for 4000 mesh cells in the article [4]. Here we use second-order flux splitting and Rusanov methods by applying second order reconstruction of variables with the slope limiter function minmod [6]. All schemes give physically correct solution of density. The Rusanov and the flux splitting methods perform density distribution a little better than the GFORCE method. The Rusanov method provides more accurate resolution of right shock structure in density distribution, but produces small oscillations of volume fraction.

Conclusion

We have shown that the proposed flux splitting method for the one-dimensional SHTC equations provides good agreement with reference and exact solutions. Future work will concern the extension of the method to solid-water and solid-gas flows, as well as for multi-dimensional problems. We are going to improve the method for solving diffusive interface problems.

Acknowledgments

The research was made possible by the Government of the Russian Federation (Agreement 074-02-2018-286). The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University [5] supported by the project RFMEFI62117X0011. We thank Dr. V.A. Titarev for helping to run calculations on Lomonosov-2 supercomputer.

References

1. Baer, M., Nunziato, J.: A Two-phase mixture theory for the deflagration-to-detonation transition (DDT) in reactive granular materials. International Journal of Multiphase Flow 12(6), 861–889 (1986), DOI: 10.1016/0301-9322(86)90033-9

2. ten Eikelder, M.F.P., Daude, F., Koren, B., Tijsseling, A.S.: An acoustic-convective splitting-based approach for the Kapila two-phase flow model. Journal of Computational Physics 331, 188–208 (2017), DOI: 10.1016/j.jcp.2016.11.031

3. Kapila, A., Menikoff, R., Bdzil, J., Son, S., Stewart, D.S.: Two-phase modeling of deflagration-to-detonation transition in granular materials: Reduced equations. Physics of Fluids 13(10), 3002–3024 (2001), DOI: 10.1063/1.1398042

4. Romenski, E., Drikakis, D., Toro, E.F.: Conservative models and numerical methods for compressible two-phase flow. Journal of Scientific Computing 42(1), 68–95 (2010), DOI: 10.1007/s10915-009-9316-y

5. Sadovnichy, V., Tikhonravov, A., Voevodin, Vl., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From

Petascale toward Exascale. pp. 283–307. Chapman & Hall/CRC Computational Science, CRC Press, Boca Raton, United States, (2013)

6. Toro, E.F.: Riemann Solvers and Numerical Methods for Fluid Dynamics. Springer (2009), DOI: 10.1007/b79761

7. Toro, E.F., Spruce, M., Speares, W.: Restoration of the contact surface in the HLL-Riemann solver. Shock Waves 4(1), 25–34 (1994), DOI: 10.1007/bf01414629

8. Toro, E.F., Titarev, V.A.: MUSTA fluxes for systems of conservation laws. Journal of Computational Physics 216(2), 403–429 (2006), DOI: 10.1016/j.jcp.2005.12.012

DOI: 10.14529/jsfi180316

Magnetic Properties of $LaAlO_3/SrTiO_3$ Heterostructure Modelled on a Supercomputer

Irina Piyanzina[1,2], *Volker Eyert*[3], *Thilo Kopp*[4], *Dmitrii Tayurskii*[1]

The oxide heterostructure composed of $LaAlO_3$ (LAO) thin film on top of $SrTiO_3$ (STO) substrate is the best known example of a system where a metallic state is formed in the STO layers next to the interface [1]. In the frame of present work we analyze an impact of oxygen vacancies and hydrogen dopants located in the AlO_2 surface layer and in the TiO_2 interfacial plane of LAO/STO heterostructure onto the magnetic properties by performing spin-polarized calculations based on density functional theory (DFT). We found stable local magnetic moments formed within atomically thin magnetic layers at the interface. We confirmed that magnetism can be generated by oxygen vacancies located either at the surface or at the interface. In addition, we demonstrate magnetic moments formation by hydrogen dopants located at the interface. Finally, the case of two defects combination was investigated, when negligibly small magnetic moment induction was found to take place.

Keywords: DFT, $LaAlO_3/SrTiO_3$ heterostructure, defects, magnetic properties.

Introduction

The arising magnetic order in the LAO/STO system is a matter of intensive discussion [2, 3]. It was found from *ab initio* calculations that the bare heterostructure is non-magnetic, and magnetic ordering was related to defects formation, in particular to oxygen vacancies at the interface and/or the surface [2]. Other scenarios were also suggested, for instance, the formation of Ti^{3+}-on-Al^{3+} defects in LAO near the interface or the Zener exchange between an insulating interface layer and the nearest TiO_2 plane [3]. No doubt, that defect states play an important role in structural, electronic and magnetic properties, especially if they are electron donor states. In spite of the fact that experimentalists seek to get rid of any impurities and defects, especially dealing with electronic devices, it is almost impossible to avoid any possible contaminations. Obviously, defects affect the functionalization of devises based on oxide heterostructures. The facts that oxygen vacancies and hydrogen dopants have relative low defect formation energies, that they are omnipresent in experiments, and both are electron-donor defects motivate us to investigate their impact on magnetic properties.

1. Results

The *ab initio* calculations were based on DFT [4] in the framework of the GGA+U method, with additional local correlations of $U = 2\,\text{eV}$ to the Ti $3d$ orbitals and $U = 8\,\text{eV}$ to the La $4f$ orbitals [5]. We used Vienna Ab-Initio Simulation Package (VASP) [6], which is part of the MedeA® software of Materials Design. To study the heterostructure with defects, we used 2×2 in-plane supercells of the bare heterostructures [7] with introduced oxygen vacancies and hydrogen dopants at the surface, or in an interfacial layer on both sides while preserving the inversion symmetry of the slabs. More details about computation parameters and slab geometry can be found in [5, 7, 8].

[1]Kazan Federal University, Kazan, Russia
[2]Zavoisky Physical-Technical Institute, FIC KazanSC of RAS, Kazan, Russia
[3]Materials Design SARL, Montrouge, France
[4]University of Augsburg, Augsburg, Germany

While investigating defects profile in the LAO/STO heterostructure we have found that for sufficiently large concentrations of oxygen vacancies the formation energy drops at the interface [8]. We started with a bare insulative and non-magnetic heterostructure containing three LAO overlayers. The presence of an oxygen vacancy caused an upshift of the Fermi energy and magnetic moment induction. We considered two cases of oxygen vacancy location: either in the surface AlO_2 layer or in the interfacial TiO_2 layer. Quantitative and qualitative comparison of obtained magnetic moments are presented in Fig. 1 and Tab. 1, respectively. Both cases of a surface and interface oxygen vacancy show that about 85% of the total magnetization corresponds to the layer comprising the vacancy.

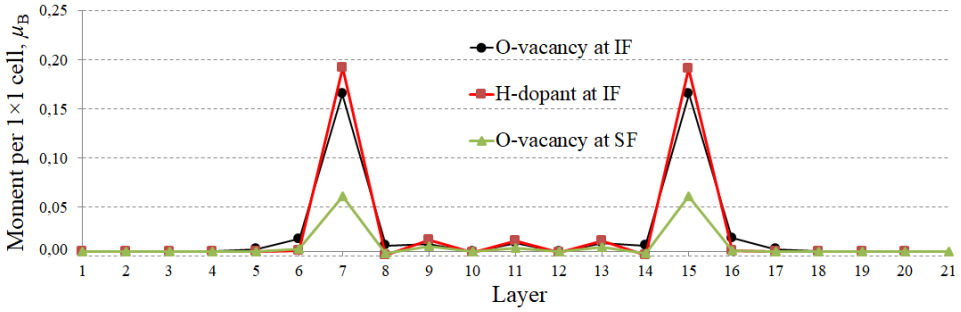


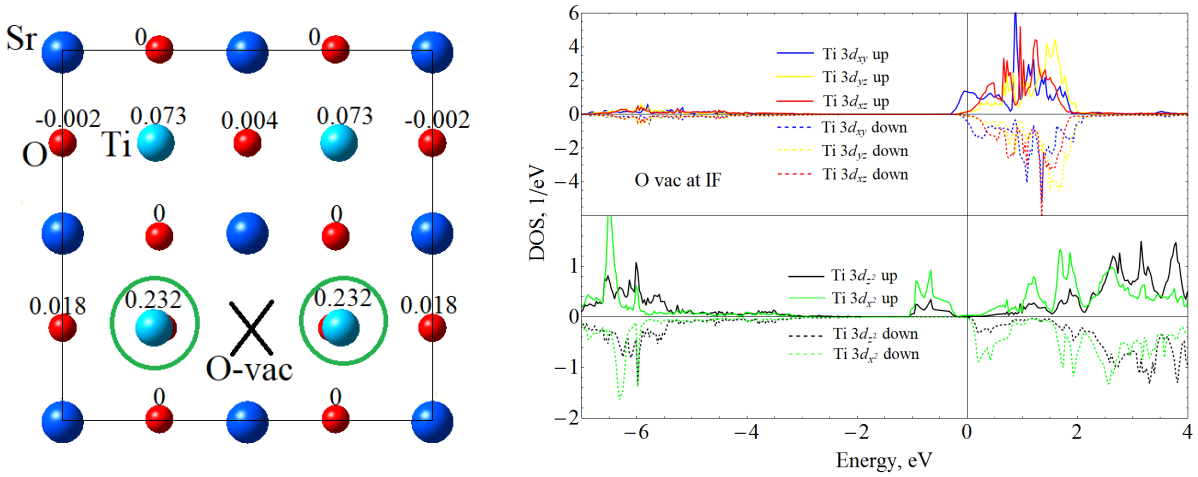
Figure 1. Magnetic moments per 1×1 cell of the 3 LAO/4.5 STO/3 LAO heterostructures with hydrogen dopant atom located at the interface, oxygen vacancy at the interface and surface

Table 1. Total and per interface (total/per IF) magnetic moments in Bohr magnetons per 1×1 cell of heterostructure with defects located whether at the surface (SF) or at the interface (IF)

Defect location	O-vacancy	H-dopant	O-vac and H-dop
at SF	0.12/0.06	0/0	0/0
at IF	0.38/0.16	0.41/0.19	0/0

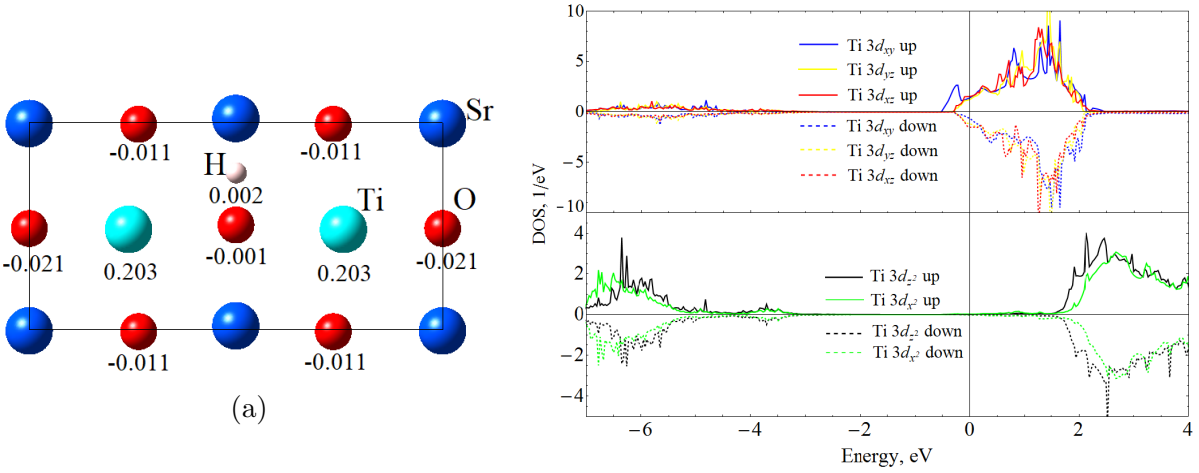
The most astounding result was found considering the densities of states (DOS) of interfacial Ti atoms. The contribution from a distant Ti atom and a Ti atom next to the vacancy differ significantly. In particular, we found the downshift of the e_g states of the latter atom as compared to those of the former by about 2 eV [8], which leads to a finite occupation of these orbitals as well as a finite contribution to the local magnetic moment, which even exceeds that of the t_{2g} states (Fig. 2 b)). In contrast, the t_{2g} partial DOS of both atoms are very similar. The calculated local magnetic moments per interface atom are given in Fig. 2 a). Thus, oxygen vacancies at the interface induce atomically thin magnetic layers with a rather uniform background of magnetic moments generated by the Ti $3d$ t_{2g} states, which are complemented by well localized magnetic moments due to e_g states of the Ti atoms neighboring the vacancy [8].

Spin-polarized calculations were also performed for a 2×1 supercell of the 3LAO/4.5STO/3LAO heterostructure with one hydrogen dopant located either in the AlO_2 surface layer or in the TiO_2 interface layer. These cases lead to the same electrons concentrations as the case of one vacancy per 2×2 cell. We found that hydrogen dopants located at the surface do not produce magnetization, whereas located in the interfacial layer induce sizable magnetization. The calculated local magnetic moments per interfacial atom are given in Fig. 3 a).



(a) (b)

Figure 2. (a) Local moments and (b) spin-resolved Ti $3d$ partial DOS of a 2×2 3 LAO/4.5 STO/3 LAO heterostructure with one oxygen vacancy



(a) (b)

Figure 3. (a) Local moments and (b) spin-resolved Ti $3d$ partial DOS of a 2×1 3 LAO/4.5 STO/3 LAO heterostructure with one hydrogen dopant

Generated total and per 1×1 interface layer magnetic moments amplitudes are listed in Tab. 1. Layer-distribution is shown in Fig. 1. As in the previous case we found strong confinement of magnetization within interface layers. Hydrogen dopants located in the interface plane produce even larger magnetic moment, however, the formation energy does not have a minimum at the interface as in the case of oxygen vacancy [8]. The spin-resolved Ti $3d$ partial DOS are shown in Fig. 3 (b). Here, no downshift of the Ti $3d$ e_g orbitals is observed since the octahedral environment of the metal atoms is not affected in contrast to the case of oxygen vacancy. Moreover, in Fig. 3 (a), the hydrogen dopant forms an almost rectangular triangle with the neighboring Ti atoms in the same plane, its s-orbital hybridizes mainly with the Ti d_{xy} orbitals. As a consequence, electron transfer from the H $1s$ orbital is predominantly to the d_{xy} orbitals of the $3d$ t_{2g} manifolds centered at the neighboring Ti ions, which thus hold the large part of the local magnetic moment.

Conclusions

In the present work, first principles electronic structure calculations as based on DFT and including local electronic correlations within the GGA+U approach have been employed to study the impact of oxygen vacancies and hydrogen dopant atoms on the magnetic properties of insulating 3LAO/STO/3LAO heterostructure slabs. We confirmed the local-moment formation induced by vacancies located either at the surface or in the TiO_2 interface layer and found strong confinement of the magnetization within these layers. Hydrogen dopant atoms in the TiO_2 interface layer of the LAO/STO heterostructure give rise to magnetism with even larger magnetic moments, which are mainly carried by the Ti $3d_{xy}$ states and again confined within the interface layer.

Acknowledgment

The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011. The reported study was funded by RFBR according to the research project 18-32-00595. The work of D. A. T. was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities (project 3.9779.2017/8.9).

References

1. Ohtomo A., Hwang H.Y.: A high-mobility electron gas at the $LaAlO_3/SrTiO_3$ Heterointerface. Nature 427, 423–426 (2004), DOI: 10.1038/nature02308

2. Pavlenko, N., Kopp, T., Tsymbal, E.Y., Mannhart, J., Sawatzky, G.A.: Oxygen vacancies at titanate interfaces: Two-dimensional magnetism and orbital reconstruction. Physical Review B 86(6), 064431 (2012), DOI: 10.1103/PhysRevB.86.064431

3. Michaeli, K., Potter, A.C., Lee, P.A.: Superconducting and ferromagnetic phases in $SrTiO_3/LaAlO_3$ oxide interface structures: Possibility of finite momentum pairing. Physical Review Letters 108(11), 117003 (2012), DOI: 10.1103/PhysRevLett.108.117003

4. Hohenberg, P., Kohn, W.: Inhomogeneous electron gas. Physical Review 136(3B), B864 (1964), DOI: 10.1103/PhysRev.136.B864

5. Piyanzina, I.I., Kopp T., Lysogorskiy, Y.V., Tayurskii D.A., Eyert V.: Electronic properties of $LaAlO_3/SrTiO_3$ n-type interfaces: a GGA+U study. Journal of Physics: Condensed Matter 29, 095501 (2017), DOI: 10.1088/1361-648X/aa57ac

6. Kresse, G., Furthmller, J.: Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Computational Materials Science 6(1), 15–50 (1996), DOI: 10.1016/0927-0256(96)00008-0

7. Piyanzina, I.I., Lysogorskiy, Y.V., Varlamova, I.I., Kiiamov A.G., Kopp T., Eyert V., Nedopekin O.V., Tayurskii D.A.: Analysis of electronic and structural properties of surfaces

and interfaces based on $LaAlO_3$ and $SrTiO_3$. Journal of Low Temperature Physics 185, 597–602 (2016), DOI: 10.1007/s10909-016-1483-2

8. Piyanzina, I.I., Eyert V., Lysogosrkiy Yu.V., Tayurskii D.A, and Kopp T.: Oxygen Vacancies and Hydrogen Doping in $LaAlO_3/SrTiO_3$ Heterostructures: Electronic Properties and Impact on Surface and Interface Reconstruction. arXiv:1803.01382 [cond-mat.str-el] (2018), `https://arxiv.org/abs/1803.01382`

DOI: 10.14529/jsfi180317

Optimization of BWB Aircraft Using Parallel Computing

***Kirill S. Anisimov*[1], *Andrey A. Savelyev*[1], *Innocentiy A. Kursakov*[1], *Alexander V. Lysenkov*[1], *Prajwal S. Prakasha*[2]**

Nacelle shape optimization for Blended Wing Body (BWB) is performed. The optimization procedure is based on numerical calculations of the Reynolds–averaged Navier–Stokes equations. For the Top Level Aircraft Requirements, formulated in AGILE project, the propulsion system was designed. The optimization procedure was divided in two steps. At first step, the isolated nacelle was designed and optimized for cruise regimes. This step is listed in paragraph 3. At second step the nacelles positions over airframe were optimized. To find the optimum solution, surrogate–based Efficient Global Optimization algorithm is used. An automatic structural computational mesh creation is realized for the effective optimization algorithm working. This whole procedure is considered in the context of the third generation multidisciplinary optimization techniques, developed within AGILE project. During the project, new techniques should be implemented for the novel aircraft configurations, chosen as test cases for application of AGILE technologies. It is shown that the optimization technology meets all requirements and is suitable for using in the AGILE project.

Keywords: optimization, CFD, Blended Wing Body, nacelle, power plant.

Introduction

The AGILE EU Project [3] is dedicated to the development of distributed multidisciplinary optimization methodology. The project is based on the key technologies developed over the last 10 years in the DLR: such as, for example, common data format CPACS [6] and RCE environment. The main purpose of AGILE project is to reduce the time of the convergence process in the aircraft optimization by 20%, and for the multidisciplinary optimization in a team of various experts by 40% by the end of 2018. The main objective for TsAGI in the current project is to optimize the external aerodynamics of the power plant. This task is possible within the framework of the project, because the project ideology at each step of the global optimization permits both the disciplinary analysis and the disciplinary optimization. At that, a number of specific requirements are made to the optimization. One of such requirements is the optimization speed, because it is necessary to optimize the external aerodynamics of outer nacelle at each step of global optimization.

1. Task Formulation

In the project beginning, the Top Level Aircraft Requirements were formulated [5]. The initial shape of airframe was also designed (Fig. 1). The requirements for airplane was reformulated to the initial parameter of optimization: cruise Mach Number = 0.8; operation altitude = 10668m (35000ft); operation weight = 300000 kg (it means Cy = 0.5185 for the BWB with mentioned above parameters and wing area equal to 900 m^2).

All the calculations were performed using solver EWT–TsAGI [2] based on the full 3D non–stationary Reynolds equation system closed by Spalart–Allmaras turbulence model. In the present work automatic algorithm [1] for structured computational mesh rebuilding is developed. All operations are made in program Grid_Creator developed in TsAGI (Russia). Free library

[1]Central Aerohydrodynamic Institute, TsAGI, Russian Federation
[2]Institute of System Architectures in Aeronautics, DLR, Hamburg, Germany

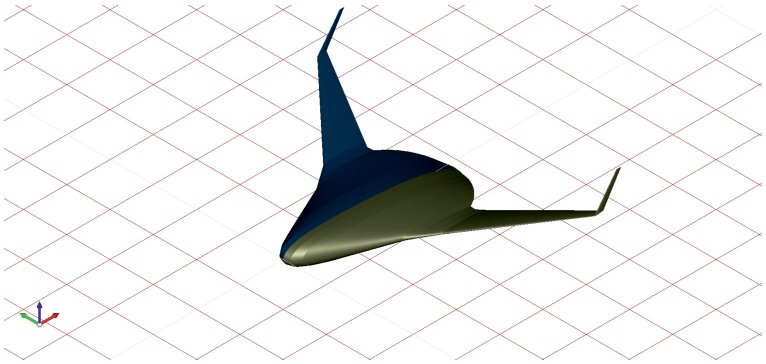


Figure 1. Reference airframe for engine design

cgnslib version 3.1.3 is used in the program Grid_Creator for operation with CGNS format. In addition Grid Creator has a number of additional functions: usage of additional possibility of EWT–TsAGI [2] solvers (families, turbulence model parameters, etc.), setting of irregular flows on the computational region boundary, cluster load optimization. The solver EWT–TsAGI could be efficiently paralyzed up to 500 cores for a task with number of blocks in structured mesh greater than 2000. The optimization efficiency additionally increased by running computations for a number of geometry variants simultaneously. The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 [7].

During the optimization, the value of effective thrust losses (1) for isolated nacelle at the cruise regime has been used as an objective function.

$$dP_{eff} = P_{id} - P_{eff}, \tag{1}$$

where P_{id} – the ideal engine trust; $P_{eff} = P - F_x$ – the effective engine trust (thrust–minus–drag); P_{id} – the engine thrust determined with the use of the internal parameters; F_x – projection of the total force of external drag on the engine axis.

As an optimizer code, SEGOMOE, developed by ONERA [4] is used. SEGOMOE is very efficient for the tasks with expensive problem, in terms of computing resources, with moderate noise pollution (depending on the used calculation method and the grid detalization), and non–zero probability of finding local extremes. It means that it is possibly necessary to use global nongradient–based optimization methods with the purpose to reduce the noise influence and try not to get into local extremum. Initial DOE (Design Of Experiment) points are computed simultaneously, but after DOE points are running one by one.

2. Isolated Nacelle Optimization

Based on the authors experience, which has been obtained in the optimization of nacelle turbofan engine with high bypass ratio, and based on the results of calculations performed in the preliminary design stage, it has been concluded that there is weak interference between the nozzle and the inlet. Therefore, the initial problem of designing the aerodynamic contours of nacelle has been divided into two independent nozzle and inlet design optimization problems.

Nacelle geometry has been divided into two parts at mid–section. At that, mid–section diameter and position are nozzle parameters. For this reason, the nozzle has been designed at

the first stage. At the second stage, the inlet has been designed for mid–section diameter and position chosen at the first stage.

After the designing a shape of the axisymmetric inlet, setting of the inlet takes place: the inlet axis rotates around the OZ axis at an angle with respect to the engine axis. The convergence of effective thrust losses received by SEGOMOE is shown in Fig. 2. For this task 30 DOE points for 18 parameters and 2 constraints are used. Gas mass flow rates throw core and fan nozzles are used as constraints. After 90 points, SEGOMOE found the optimal solution with satisfying of constraints. The further calculations (Fig. 2) showed that the discovered solution was optimal. It is a very good result for an optimizer with this kind of task.

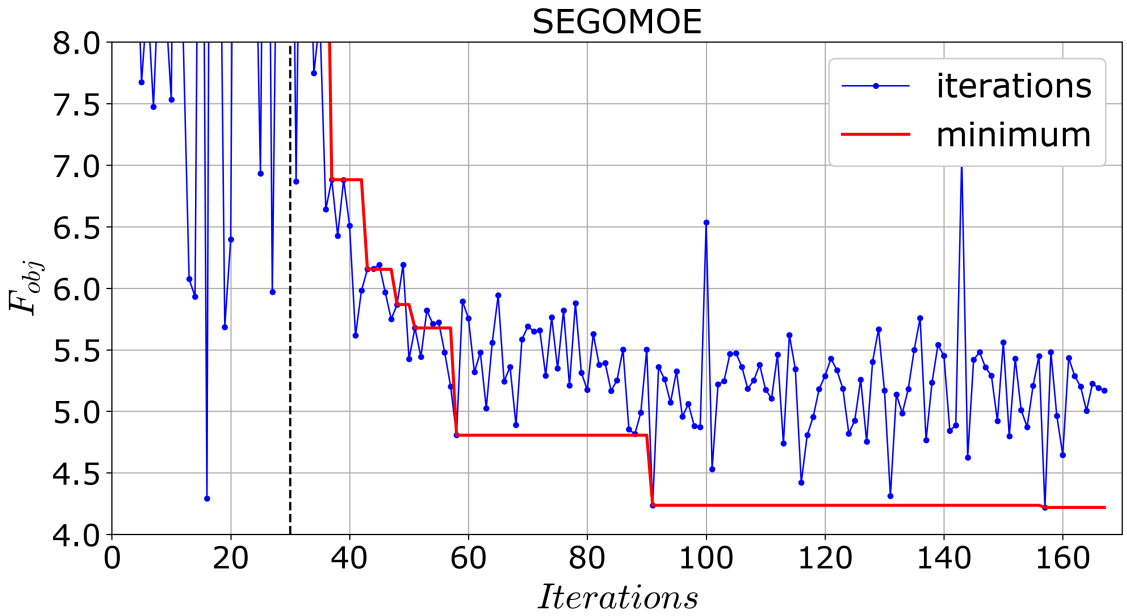


Figure 2. Convergence of effective thrust losses by SEGOMOE

3. Engine/Airframe Integration

The position optimization was proceeded like the isolated nacelle optimization. In this case, the design parameters were: x and z coordinates of central nacelle; x, y and z coordinates of side nacelle; angle of attack of nacelles (the angle was the same for all nacelles); sweep angle for side nacelle; aircraft angle of attack. This parameters are necessary for supplying of constraint. As a constraint the lift force was used for satisfying (equivalent of Cy = 0.5185) of Top Level Aircraft Requirements for Operation Weight. After finishing optimization procedure the optimal parameters were received. The received lift coefficient for overall configuration was Cy = 0.5185. This means satisfying of requirements. The final configuration is presented in Fig. 3.

The received data analysis permits to talk about huge negative interference between wing and fuselage. During the optimization, all parameters try to increase the distance between two nacelles, as well as nacelles and the fuselage to reduce the interference. But the constraint and range of variety force them to be together. If we will analyze the thrust of different engines we can find the thrust of central engine equal to $P_c = 75244.52$ and side engine $P_s = 77721.41$. This showed bigger efficiency of side engines because of low interference. This fact points out the conclusion about negative interference.

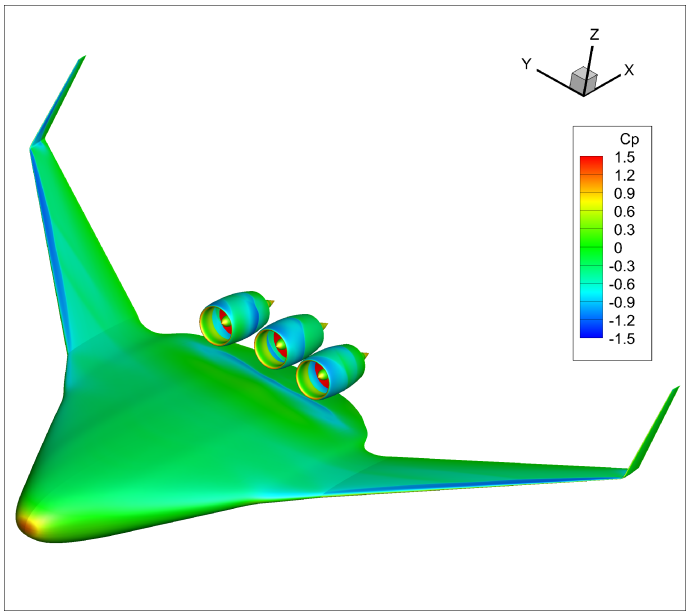


Figure 3. Pressure coefficient distribution over BWB aircraft with three optimal engines

Conclusions

The task aerodynamic design of propulsion system was successfully made for BWB configuration. The airplane with three optimal engines satisfy the Top Level Aircraft Requirements. This configuration is appropriate for further investigation in this area and possible to be used for multidisciplinary optimization of overall aircraft. During further investigation huge attention should be paid to the aerodynamic interference between airframe and fuselage.

Acknowledgments

The research presented in this paper has been performed in the framework of the AGILE project and has received funding from the European Union Horizon 2020 Programme H2020-MG-2014-2015 under grant agreement N 636202. The authors are grateful to the partners of the AGILE consortium for their contribution and feedback.

References

1. Anisimov, K., Savelyev, A., Kursakov, I., Lysenkov, A., Mirzoyan, A., Prakasha, P.: Propulsion System - Airframe Integration and Optimization of Civil Aircraft- AGILE EU Project. ICAS Conference Proceedings, Belo Horizonte, Brazil (Sept 2018)

2. Bosnyakov, S., Kursakov, I., Lysenkov, A., Matyash, S., Mikhailov, S., Vlasenko, V., Quest, J.: Computational tools for supporting the testing of civil aircraft configurations in wind tunnels. Progress in Aerospace Sciences 44(2), 67–120 (2008), DOI: 10.1016/j.paerosci.2007.10.003

3. Ciampa, P.D., Nagel, B.: chap. The AGILE Paradigm: the Next Generation of Collaborative

MDO. AIAA AVIATION Forum, American Institute of Aeronautics and Astronautics (Jun 2017), DOI: 10.2514/6.2017-4137

4. Lefebvre, T., Bartoli, N., Dubreuil, S., Panzeri, M., Lombardi, R., D'Ippolito, R., Della Vecchia, P., Nicolosi, F., Ciampa, P.D.: chap. Methodological Enhancements in MDO Process Investigated in the AGILE European Project. AIAA AVIATION Forum, American Institute of Aeronautics and Astronautics (Jun 2017), DOI: 10.2514/6.2017-4140

5. Prakasha, P., Ciampa, P., Della Vecchia, P., Ciliberti, D., Voskuijl, M., Charbonnier, D., Jungo, A., Fioriti, M., Anisimov, K., Mirzoyan, A.: Multidisciplinary Design Analysis of Blended Wing Body Through Collaborative Design Approach: AGILE EU Project. ICAS Conference Proceedings, Belo Horizonte, Brazil (Sept 2018)

6. Rizzi, A., Zhang, M., Nagel, B., Boehnke, D., Saquet, P.: Towards a unified framework using cpacs for geometry management in aircraft design. In: 50th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition. pp. AIAA 2012–0549– (2012), DOI: 10.2514/6.2012-549

7. Sadovnichy, V., Tikhonravov, A., Voevodin, V., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale toward Exascale. pp. 283–307. Chapman & Hall/CRC Computational Science, CRC Press, Boca Raton, United States (2013)

DOI: 10.14529/jsfi180318

Supercomputer Simulations of Nondestructive Tomographic Imaging with Rotating Transducers

Sergey Y. Romanov[1]

A method of nondestructive ultrasound tomographic imaging employing a rotating transducer system is proposed. The rotating transducer system increases the number of emitters and detectors in a tomographic scheme by several times and makes it possible to neutralize image artifacts resulting from incomplete-data tomography. The inverse problem of tomographic reconstructing the velocity structure inside the inspected object is considered as a nonlinear coefficient inverse problem for a scalar wave equation. Scalable iterative algorithms for reconstructing the longitudinal wave velocity inside the object are discussed. The methods are based on the explicit representation for the gradient of the residual functional. The algorithms employ parallelizing the computations over emitter positions. Numerical simulations performed on the "Lomonosov-2" supercomputer showed that the tomographic methods developed can not only detect boundaries of defects, but also determine the wave velocity distribution inside the defects with high accuracy provided that both reflected and transmitted waves are registered.

Keywords: supercomputer, ultrasound tomography, nondestructive testing, inverse problems.

Introduction

This paper considers ultrasound tomography methods as applied to nondestructive testing (NDT). Typical NDT tasks include ultrasonic inspection of welds, non-destructive testing of concrete structures, products made of plastics and composite materials [4]. However, commonly used ultrasonic inspection methods are not tomographic, as the object is usually sounded from a single side. There are many NDT methods that can detect the boundaries of defects inside the object by measuring reflected ultrasonic waves. One example is "topological imaging" [1, 5]. Synthetic Aperture Focusing Technique (SAFT) [3, 10] is also widely used.

In the practice of NDT, inspected objects often have areas with different acoustic properties, which can be unknown. An example of such a problem is inspection of a welded joint. The speed of sound in the joint and in the base metal is different. Topological imaging and SAFT methods cannot determine the acoustic parameters, such as the sound speed distribution, but correct application of these methods is possible only if the acoustic properties of the object are known. The tomographic methods used in this study use both reflected and transmitted waves for image reconstruction and make it possible not only to detect the boundaries of various regions inside the object, but also to determine the wave velocity in these regions [7, 9, 11, 12].

In [2], a method of ultrasound tomography for NDT with fixed transducer arrays was investigated. A rotating transducer system for sounding an object from different angles allows us to improve the quality of the reconstructed image in this paper. In addition, the use of a rotating transducer array increases the number of emitter positions by several times, thus compensating for a small number of detectors in typical linear transducers used in NDT. However, increasing the number of emitters also increases the computational complexity linearly. Scalable algorithms for general-purpose multi-CPU supercomputers have been developed to tackle this problem. The numerical simulations were performed on the "Lomonosov-2" supercomputer [14].

[1]Lomonosov Moscow State University, Moscow, Russian Federation

From the mathematical point of view, inverse problems of ultrasound tomography are complex non-linear coefficient inverse problems. Solving them requires huge computational resources, and implementation of solution algorithms is not possible without the use of high-performance computing systems. In recent years, significant progress has been made in developing efficient numerical methods for solving such inverse problems using supercomputers [6, 13].

1. Inverse Problem and the Solution Method

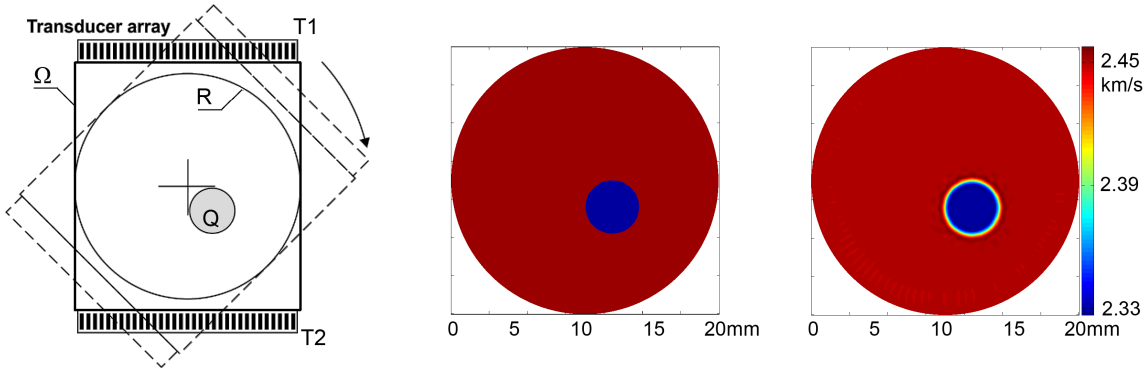
A distinctive feature of solids is that both longitudinal and transverse waves can propagate through them. However, the velocities of these waves differ by a factor of two or more, which allows the separation the longitudinal waves by pulse arrival time. In this study, the inverse problem is considered in terms of scalar wave model for the longitudinal waves $u(\boldsymbol{r}, t)$

$$c(\boldsymbol{r})u_{tt}(\boldsymbol{r}, t) - \Delta u(\boldsymbol{r}, t) = \delta(\boldsymbol{r} - \boldsymbol{r_0}) \cdot g(t), \quad u(\boldsymbol{r}, t = 0) = 0, \quad u_t(\boldsymbol{r}, t = 0) = 0. \tag{1}$$

Here $c(\boldsymbol{r}) = 1/v^2(\boldsymbol{r})$, $v(\boldsymbol{r})$ is the velocity of the longitudinal wave in the medium. Computing the $u(\boldsymbol{r}, t)$ wave field for a given $c(\boldsymbol{r})$ coefficient using the equation (1) constitutes a direct problem.

The inverse problem of ultrasonic tomography under the scalar wave model consists in reconstructing the unknown wave velocity $v(\boldsymbol{r})$ inside the region of interest using the measured wave field $u(\boldsymbol{r}, t)$ at the detectors. This inverse problem is nonlinear, and we formulate it as a problem of minimizing the residual functional between the measured and computed wave fields. We use an iterative gradient method to minimize the functional [8, 13].

Figure 1(a) shows the scheme of a tomographic examination with rotating transducer arrays. We consider the simplest 2D problem, in which the inspected region is a circle R containing an inhomogeneity Q. Two linear transducer arrays T1, T2 are located on opposite sides of the region R. Each array consists of 24 elements at a pitch of 0.6 mm which can both emit and receive ultrasound waves. The transducers can be rotated around the inspected object. The wave velocity v_1 in the Q is unknown and differs from the known v_0 in the surrounding medium.



(a) the tomographic imaging scheme (b) the phantom (c) the reconstructed image

Figure 1. A tomographic examination with rotating transducer arrays

The experiment is carried out as follows. Each of the elements of transducers T1 and T2 successively emits a sounding pulse, while all the elements of both transducers act as detectors. Then the transducers are rotated around the region R by a fixed angle. The process is repeated.

The software was designed for HPC systems under the Linux OS and implemented in C++. The MPI was used for inter-process communication. The computing nodes contained Intel Haswell-EP E5-2697v3 14-core processors at 2.6 GHz, 64 GB of RAM, and Infiniband FDR

network. The computations were parallelized so that one computing core was allocated for every emitter position. This method is natural for the problem considered, since the computations for each emitter are practically independent. This approach was proven to be effective and the algorithm practically linearly scales up to several hundreds of CPU cores. The data exchange overhead amounted to $\approx 2\%$ of the total computation time. The total of 384 CPU cores were used. Spatial domain decomposition was not used, since parallelization over emitter showed much greater efficiency. Single-precision floating point arithmetic was used in computations.

2. Numerical Simulations

The numerical experiment consisted of solving direct and inverse problems. First, for each emitter position we solve the direct problem of wave propagation in a square computational domain Ω encompassing the region of interest R (Fig. 1(a)). The transducers T1, T2 are then rotated along with the computational grid by angle 22.5°. The wave field is registered by the detectors for each rotation step and used as simulated experimental data for the inverse problem.

The central wavelength of the sounding pulse was 0.466 mm (5 MHz frequency). The beam width of each element was $\approx 50°$. A inter-element pitch larger than the wavelength and a narrow beam resulting from a large element size are a typical case in NDT applications.

Figure 1(b) shows the phantom. The sound speed in the medium was set to $v_0 = 2450$ m·s^{-1} (ebonite material), and in the inhomogeneity Q — to $v_1 = 2330$ m·s^{-1} (rexolite material). Figure 1(c) shows the reconstructed sound speed The initial sound speed approximation for the iterative method was chosen as $v_0 = const = 2450$ m·s^{-1}. As is evident from Fig. 1(c), a rotating transducer system allows not only to reconstruct the boundaries of the object, but also determine the sound speed inside the object with high accuracy. No artifacts, which are typical for incomplete-data tomographic reconstruction [2], are present in the image.

The size of the computational domain was 20×20 mm, the finite difference grid contained 700×700 points. The total of 384 emitter positions were used in the simulation. The computing time for 40 gradient descent iterations on 384 CPU cores in parallel amounted to 30 minutes.

Conclusion

In this paper, we propose a method of ultrasonic tomographic imaging involving a rotating transducer, and discuss parallel algorithms for reconstructing the longitudinal wave velocity inside the inspected object. The use of a rotating transducer system yields reconstructed images without artifacts typically present in incomplete-data reconstructions. The computational complexity increases several times, however, parallelizing the computations over emitter positions leads to scalable CPU algorithms on a supercomputer.

Acknowledgments

This research was supported by Russian Science Foundation (project No. 17-11-01065). The research is carried out at the Lomonosov Moscow State University. The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011.

References

1. Bachmann, E., Jacob, X., Rodriguez, S., Gibiat, V.: Three-dimensional and real-time two-dimensional topological imaging using parallel computing. J. Acoust. Soc. Am. 138(3), 1796–1796 (2015), DOI: 10.1121/1.4933696

2. Bazulin, E.G., Goncharsky, A.V., Romanov, S.Y., Seryozhnikov, S.Y.: Parallel CPU- and GPU-algorithms for inverse problems in nondestructive testing. Lobachevskii J. Math. 39(4), 486–493 (2018), DOI: 10.1134/S1995080218040030

3. Bazulin, E.G., Sadykov, M.S.: Determining the speed of longitudinal waves in an isotropic homogeneous welded joint using echo signals measured by two antenna arrays. Russ. J. Nondestruct Test 54(5), 303–315 (2018), DOI: 10.1134/S1061830918050029

4. Blitz, J., Simpson, G.: Ultrasonic Methods of Non-destructive Testing. Springer (1995)

5. Dominguez, N., Gibiat, V.: Non-destructive imaging using the time domain topological energy method. Ultrasonics 50, 367–372 (2010), DOI: 10.1016/j.ultras.2009.08.014

6. Goncharsky, A.V., Seryozhnikov, S.Y.: The architecture of specialized GPU clusters used for solving the inverse problems of 3D low-frequency ultrasonic tomography. Communications in Computer and Information Science 793, 363–395 (2017), DOI: 10.1007/978-3-319-71255-0_29

7. Goncharsky, A.V., Romanov, S.Y.: Supercomputer technologies in inverse problems of ultrasound tomography. Inverse Probl. 29(7), 075004 (2013), DOI: 10.1088/0266-5611/29/7/075004

8. Goncharsky, A.V., Romanov, S.Y.: Iterative methods for solving coefficient inverse problems of wave tomography in models with attenuation. Inverse Probl. 33(2), 025003 (2017), DOI: 10.1088/1361-6420/33/2/025003

9. Goncharsky, A.V., Romanov, S.Y, Seryozhnikov, S.Y.: Inverse problems of 3D ultrasonic tomography with complete and incomplete range data. Wave Motion 51(3), 389–404 (2014), DOI: 10.1016/j.wavemoti.2013.10.001

10. Hall, T.E., Doctor, S.R., Reid, L.D., Littlield, R.J., Gilber, R.W.: Implementation of real-time ultrasonic SAFT system for inspection of nuclear reactor components. Acoustical Imaging 15, 253–266 (1987), DOI: 10.1007/978-1-4684-5320-1_23

11. Klibanov, M.V., Kolesov, A.E., Nguyen, L., Sullivan, A.: Globally strictly convex cost functional for a 1-D inverse medium scattering problem with experimental data. SIAM Journal on Applied Mathematics 77(5), 1733–1755 (2017), DOI: 10.1137/17M1122487

12. Natterer, F.: Possibilities and Limitations of Time Domain Wave Equation Imaging. In: AMS: Tomography and Inverse Transport Theory, vol. 559, pp. 151–162 (2011)

13. Romanov, S.: Optimization of numerical algorithms for solving inverse problems of ultrasonic tomography on a supercomputer. Communications in Computer and Information Science 793, 67–79 (2017), DOI: 10.1007/978-3-319-71255-0_6

14. Sadovnichy, V., Tikhonravov, A., Voevodin, Vl., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale toward Exascale, pp. 283–307 (2013)

DOI: 10.14529/jsfi180319

Reverse Mapping Algorithm for Multi-scale Numerical Simulation of Polylactic Acid

Mikhail K. Glagolev[1], *Valentina V. Vasilevskaya*[1,2]

An algorithm is proposed to convert the coarse-grained A-graft-B model of polylactic acid into the atomistic representation. In the A-graft-B model the atoms of the backbone are mapped onto A beads, which form the linear backbone of the coarse-grained macromolecule, the methyl groups are mapped onto B side pendants. The algorithm restores atomic positions based on positions of coarse-grained beads with the help of pre-defined chain fragments, called templates. The dimensions of the templates are adjusted by affine transformation to ensure coincidence of the backbone in coarse-grained and atomistic representation. The transition between coarse-grained and atomistic models conserves information about the fine structure of polymer chains. The restored configurations are suitable for further molecular-dynamic simulations. Both atomistic and coarse-grained representations require standard GROMACS software. The algorithm can be used for reverse mapping of other A-graft-B polymer models.

Keywords: molecular dynamics, multiscale simulation, reverse mapping, poly(lactic acid).

Introduction

Polylactic acid (PLA) is a biocompatible, biodegradable polymer which can be cheaply produced in industrial quantities from renewable resources [7]. To bring it on equal footing with the currently used petroleum-based plastics, its mechanical and barrier properties need to be tuned [2]. In pure PLA and in mixtures with its oligomers (OLA) these depend on stereosequence and molecular weight of the chains. Identical structure of the monomer units in short and long PLA chains complicates the experimental analysis of the materials and makes computer simulations an indispensable tool for their guided development. The simulations of deformation, diffusion, and crystallization require large timescales and determine the applicability of the multi-scale approach [4, 6]. The latter includes coarse-grained simulations using specially tailored models and requires efficient algorithms for switching between the coarse-grained and atomistic representations. In the previous paper [3] we proposed an A-graft-B coarse-grained model of PLA, with each monomer unit represented by an A-B dumbbell. The assignment of separate side pendant for the methyl group allowed explicit treatment of the PLA stereosequence. The macroscopic and microscopic properties of the systems calculated using the coarse-grained model closely resemble those calculated using atomistic simulations for both polymer and oligomer melt in a wide range of temperatures [3].

1. A-graft-B Model

The mapping scheme of the coarse-grained model [3] is shown in Fig. 1. It determines the mutual correspondence of atom groups and coarse-grained beads. The atoms of the backbone, including the attached hydrogens and ester groups, are mapped onto A beads, which form the linear backbone of the coarse-grained macromolecule. Each methyl group is mapped onto a B bead, which is connected to the A bead as side pendant. To make the A beads closer to the

[1]A. N. Nesmeyanov Institute of Organoelement Compounds of Russian Academy of Sciences, Moscow, Russia
[2]Chemistry Department, Moscow State University, Moscow, Russia

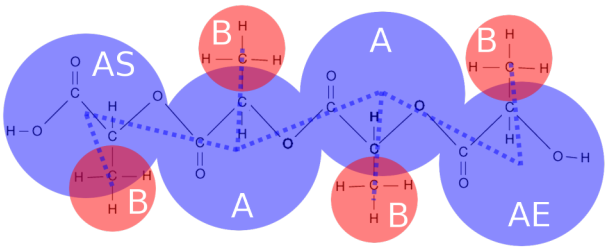


Figure 1. Mapping scheme of A-graft-B model of PLA

corresponding B beads, the backbone is chunked into A beads across the ester groups. The extra atoms at the ends of the chains are incorporated into the terminal A beads. The latter are assigned special types AS (start) and AE (end). The force-field of the coarse-grained model was derived to fit the parameters of atomistic simulations of OLA melt [5]. The simulations using the coarse-grained model were performed using an unmodified version of GROMACS [1] on "Lomonosov-2" supercomputer [8]. It was shown [3] that the coarse-grained model provides a 17-fold speedup compared with atomistic calculations.

2. Reverse Mapping

2.1. Structure Restoration

The atomistic structure was restored based on positions of coarse-grained beads with the help of monomer-scale structure fragments (templates). The templates included atomic coordinates and the positions of A and B coarse-grained beads calculated as centers of masses of the corresponding atoms according to the forward mapping procedure [3].

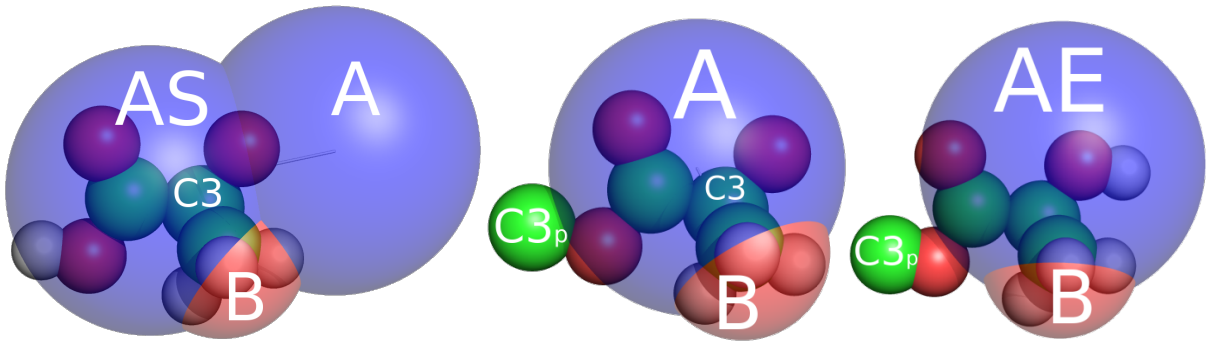


Figure 2. Template for start unit (left), middle unit (center), and end unit (right)

The templates for the start, middle, and end monomer units are shown in Fig. 2. The starting template also included the position of the next (second) coarse-grained A bead required for alignment of the atomistic structure. The reverse mapping was carried out for one chain at a time. For each of the chains, the restoration of the atomistic structure was performed sequentially and began with the start unit. The AS bead of the starting template (Fig. 2, left) was superimposed on AS bead of the coarse-grained configuration and served as a fixed anchor. The template was then rotated around AS bead so that the AS-A axis of the template was aligned with the corresponding axis of the coarse-grained configuration. The dimensions of the template were adjusted by affine transformation along the AS-A axis to ensure coincidence of A beads. After that, the template was rotated around the AS-A axis to orient the AS-B bond of the template in accordance to the AS-B bond of the coarse-grained configuration. As a result

of the procedure, the positions of the atoms of the first monomer unit of PLA were established. The templates for middle and end monomer units additionally contained the carbon atom of the previous monomer unit (denoted as C3p). As the first step of monomer restoration, the C3p atom of the template was superposed directly onto the C3 (shown in green) atom of the previous, already restored monomer. The procedures described above for the start unit (fitting, affine transformation, and AS-B orientation) were repeated for all monomer units with C3p/C3 atom serving as an anchor. The procedure ensured positioning of coarse-grained A beads of the templates precisely at positions of A beads of the coarse-grained system and correct orientation of methyl groups. The algorithm was applied in turn for each of the macromolecules.

2.2. Eqilibration

The restored atomistic structures underwent steep descent energy minimization using GROMACS with a step size of 10^{-5} to reduce the maximum force to 200 kJ/mol/nm. After that, the system was equilibrated in several steps first at constant volume and then at the constant pressure of 1 bar. The timestep of the molecular dynamics simulation was gradually increased from 10^{-6} ps to 10^{-3} ps. The total equilibration time of the system was around 100 ns.

2.3. Verification

To check the validity of the reverse mapping algorithm, we compared the macroscopic and microscopic parameters of the system containing 113 OLA molecules during atomistic and coarse-grained simulations. By the end of the equilibration the properties of the system, such as melt density, the energies of excluded volume and electrostatic interactions, as well as that of bonds, angles and dihedrals were within the margin of error of the corresponding values in the original atomistic simulation run. The gyration radius of the OLA chains, which was slightly different in the coarse-grained model, was restored to the original value observed in atomistic simulations (see Tab. 1).

Table 1. Properties of OLA melt in original atomistic, CG, and restored atomistic simulations

Parameter	Original AA	CG	Restored AA
Melt density	968.6 ± 1.0	968.2 ± 0.5	967.8 ± 1.2
Gyration radius	0.889 ± 0.004	0.899 ± 0.001	0.887 ± 0.004

Conclusions

We completed the stack of algorithms for multiscale simulation of PLA by presenting a reverse mapping procedure for translation of coarse-grained A-graft-B model of PLA into atomistic representation. Both atomistic and coarse-grained representations require standard GROMACS software, and the simulations can be carried out in parallel on supercomputers. The equilibration time of the restored atomistic samples is 0.1 μs which is several orders of magnitude lower than the characteristic times of the processes which can be simulated in coarse-grained representation. The restored atomistic configurations are suitable for further molecular-dynamic simulations and detailed analysis.

Acknowledgments

The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 [8].

The research was supported by Russian Foundation for Basic Research, project No. 17-03-00742.

References

1. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E.: Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1-2, 19–25 (2015), DOI: 10.1016/j.softx.2015.06.001

2. Burgos, N., Tolaguera, D., Fiori, S., Jimenez, A.: Synthesis and characterization of lactic acid oligomers: Evaluation of performance as poly(lactic acid) plasticizers. Journal of Polymers and the Environment 22(2), 227–235 (2014), DOI: 10.1007/s10924-013-0628-5

3. Glagolev, M., Glova, A., Mezhenskaia, D., Falkovich, S., Larin, S., Vasilevskaya, V., Lyulin, S.: Coarse-grained a-graft-b model of poly(lactic acid) for molecular dynamics simulations. Journal of Polymer Science Part B: Polymer Physics 56(7), 604–612 (2018), DOI: 10.1002/polb.24567

4. Glagolev, M.K., Lazutin, A.A., Vasilevskaya, V.V., Khokhlov, A.R.: Influence of cross-linking rate on the structure of hypercrosslinked networks: Multiscale computer simulation. Polymer 86, 168–175 (2016), DOI: 10.1016/j.polymer.2016.01.040

5. Glova, A.D., Falkovich, S.G., Larin, S.V., Mezhenskaia, D.A., Lukasheva, N.V., Nazarychev, V.M., Tolmachev, D.A., Mercurieva, A.A., Kenny, J.M., Lyulin, S.V.: Poly(lactic acid)-based nanocomposites filled with cellulose nanocrystals with modified surface: All-atom molecular dynamics simulations. Polymer International 65(8), 892–898 (2016), DOI: 10.1002/pi.5102

6. Lazutin, A.A., Glagolev, M.K., Vasilevskaya, V.V., Khokhlov, A.R.: Hypercrosslinked polystyrene networks: An atomistic molecular dynamics simulation combined with a mapping/reverse mapping procedure. Journal of Chemical Physics 140(13) (2014), DOI: 10.1063/1.4869695

7. Maria Laura Di Lorenzo, R.A. (ed.): Synthesis, Structure and Properties of Poly(lactic acid). Advances in Polymer Science, Springer (2018), DOI: 10.1007/978-3-319-64230-7

8. Sadovnichy, V., Tikhonravov, A., Voevodin, V., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale toward Exascale. pp. 283–307. Chapman & Hall/CRC Computational Science, CRC Press, Boca Raton, United States (2013)

DOI: 10.14529/jsfi180320

Supercomputer Technologies as a Tool for High-resolution Atmospheric Modelling towards the Climatological Timescales

Vladimir S. Platonov[1], *Mikhail I. Varentsov*[1]

Estimation of the recent and future climate changes is the most important challenge in the modern Earth sciences. Numerical climate models are an essential tool in this field of research. However, modelling results are highly sensitive to the spatial resolution of the model. The most of the climate change studies utilize the global atmospheric models with a grid cell size of tens of kilometres or more. High-resolution mesoscale models are much more detailed, but require significantly more computational resources. Applications of such high-resolution models in climate studies are usually limited by regional simulations and by relatively short timespan. In this paper we consider the experience of the long-term regional climate studies based on the mesoscale modelling. On the examples of urban climate studies and extreme wind assessments, we demonstrate the principle advantage of long-term high-resolution simulations, which were carried out on the modern supercomputers.

Keywords: regional climate model, long-term simulations, supercomputer technologies, extreme wind, urban climate, urban precipitation, COSMO.

Introduction

The current state in the numerical weather and climate modelling field reached a level, which often allows to consider this technique as an alternative to traditional climate-related data sets, based on the in-situ observations. Climate models are the parallelized programs that solve the system of differential equations describing the hydrothermodynamics of the atmosphere, ocean, soil and other components of environment. Such system does not have an analytical solution and has to be solved by finite-difference numerical methods, applied for a prescribed model grid. The spacing between horizontal and vertical grid points is a key parameter of the atmospheric models and determines their resolution the scale of the processes, which could be represented explicitly by the basic model equations. Other processes are called as subgrid processes and should be parameterized, i.e. expressed through an additional physical model or empirical approaches.

Decreasing of the grid spacing makes the simulation results more detailed and physical. However, such improvements lead to a huge increase of required computational resources, because a stable numerical solution for a smaller grid step demands a smaller time step according to the Courant criterion. Therefore, the grid spacing of the global climate models is still limited by the first tens of kilometres. It means that only the synoptic-scale processes could be resolved, while the most of the mesoscale processes – e.g. polar lows, sea/lake breezes, convective systems, which are responsible for many severe weather events, – could not be resolved. This is done by atmospheric models with a grid step within a range of 1–10 km (mesoscale models), which are usually applied for limited-area simulations using the dynamical downscaling approach [1], short timespan (case study, one season or few years) and require significant computational resources.

In this short communication, we present a review of our recent studies, based on the long-term simulations with a regional (limited-area) mesoscale climate model COSMO-CLM, carried out on the modern supercomputer complex of Lomonosov Moscow State University [2]. We focus on the specific research opportunities, which are opened by the long-term regional climate

[1] Lomonosov Moscow State University, Moscow, Russian Federation

simulations based on two examples: the first one is an urban climate study for Moscow megacity, and the second one is an extreme wind assessment over maritime region in the Far East of Russia.

1. Model Description and Experiments Design

We used the COSMO-CLM model (ver. 5.0) as the main tool in our studies. COSMO-CLM is the regional climate model developed by German Weather Service (DWD) and CLM-Community ($http://www.clm-community.eu$; [3, 4]). The dynamical downscaling technique for the chain of nested domains was applied in our simulations using ERA-Interim reanalysis data [5] ($\sim 0.75^0$ resolution) as forcing, and the "spectral nudging" technique was applied in order to link the model behaviour to the real large-scale atmospheric dynamics [6, 7].

2. Experiment Results

2.1. Research of the Urban-induced Climate Anomalies of Moscow Megacity

Urban climate effects, such as urban heat island (UHI), are well-studied for surface layer of the atmosphere. However, the most of urban climate studies are based only on episodic observations or short-term numerical simulations. Recent developments of supercomputer technologies have opened new opportunities to study the urban-induced climate features towards the climatological timescale. The investigation of the urban-induced climate features of Moscow megacity based on the long-term mesoscale simulations is presented in details in [8].

In the study for Moscow, the COSMO-CLM model was used for the dynamic downscaling of the reanalysis data for three nested domains (Tab. 1). An urban canopy scheme TERRA_URB [9] was applied to parametrize the urban surface features in simulations for the final domain D3_Mos with 1 km horizontal grid step. Simulations with and without the TERRA_URB scheme (URB/noURB runs) were conducted for 10 summer seasons (2007–2016). Supercomputer "Lomonosov-2" of Lomonosov Moscow State University [2] was used for calculations ($177{\cdot}10^3$ CPU hours in total, see details in Tab. 1).

Table 1. Description of the model domains and the computational resources, used in the modelling studies for Moscow (May–August) (D_Mos) and the Far East of Russia (D_Sak)

Name	Horizontal grid size, cells/km	Horizontal grid step, km	Time step, sec	Nodes used	Computing time for year run, hours
D1_Mos	140x140	12	120	64	5.1
D2_Mos	200x200	3	40	144	14.5
D3_Mos	180x180	1	10	196	39
D1_Sak	145x355	13.2	120	288	20
D2_Sak	228x525	6.6	40	288	60
D3_Sak	300x500	2.2	20	196	-

Comparison between URB and noURB runs confirm the existence and significance on seasonal scales of the various megacity-induced effects, including the UHI extended to the atmospheric boundary layer, the urban breeze circulations and positive urban precipitation anomaly. We have considered these effects in more details due to high-resolution simulations. The difference in summer precipitation sum between URB and noURB runs for a one season is charac-

terized by a chaotic spatial structure [8] caused by the stochastic nature of summer convective showers, but averaging over ten summer seasons reveals a substantial urban-induced anomaly of summer precipitation (by 10% in average over the city). The difference between these simulations could explain a significant uncertainty of urban precipitation effects, indicated in [10].

2.2. Extreme Winds Assessment in the Far East of Russia

An investigation of extreme weather events and wind speeds could be made using the long-term (for decades) simulations and robust statistical estimates of outputs. We have run the long-term experiment similar to the previous one for the Sea of Okhotsk and Sakhalin Island region [11] for the 1985-2014 period (see Tab. 1).

Long-term simulations for domains D1_Sak and D2_Sak allowed to select the cases with the highest winds (more than 25 m/s) according to the "independent storms" technique [12]. For the selected cases a detailed synoptic analysis was performed in order to investigate the typical synoptic patterns favorable for the genesis of extreme wind speeds [11]. Moreover, long-term simulations allowed us to discover a significant ($p > 0.95$) positive trend for the frequency of the wind speed exceeding 20 m/s in the area under study (Fig. 1).

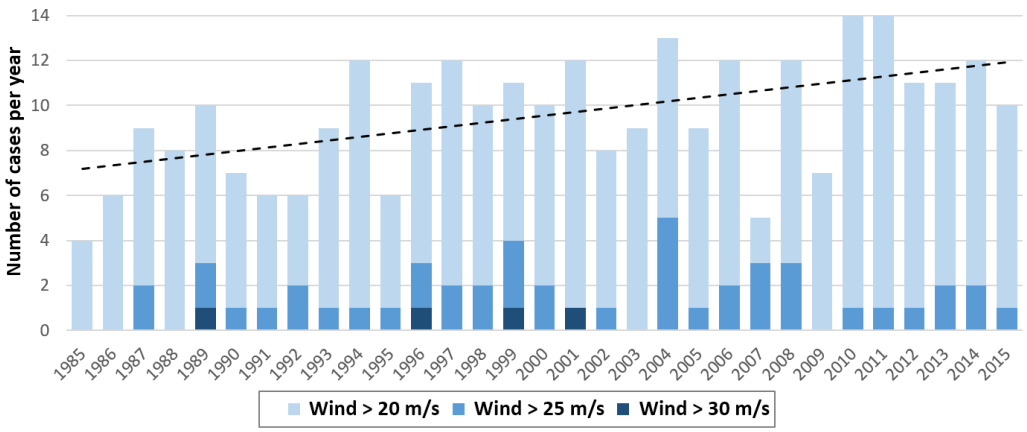


Figure 1. Number of cases with wind speed above thresholds 20, 25 and 30 m/s for 1984–2015 period. The linear trend refers to 20 m/s threshold

Conclusion

Long-term detailed regional climate simulations are opening up a lot of opportunities to Earth science researchers, e.g. estimations of modern and future climate trends, environmental consequences. The main advance of this approach is a suppression of stochastic processes in climate system and a great potential to statistical analysis of weather and climate extremes. Ultimately, namely the development of supercomputer technologies is a prerequisite of successful climate and Earth system research.

Acknowledgements

The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011. COSMO-CLM model runs and data analysis for Moscow region were funded by the grant program of Russian Science Foundation (project 17-77-20070) and by Russian Foundation for Basic Research (project 18-35-00604).

References

1. Rummukainen M.: State-of-the-art with regional climate models. Wiley Interdiscip. Rev. Chang. 1(1) 82–96 (2016), DOI: 10.1002/wcc.8

2. Sadovnichy V., Tikhonravov A., Voevodin Vl., and Opanasenko V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale toward Exascale, Chapman & Hall/CRC Computational Science, CRC Press, Boca Raton, USA, pp. 283–307 (2013)

3. Böhm U. et al.: CLM the climate version of LM: Brief description and long-term applications. COSMO Newsletters 6, 225–235 (2006)

4. Rockel B., Will A., Hense A.: The regional climate model COSMO-CLM (CCLM). Met. Zeit. 17(4), 347–348, (2008), DOI: 10.1127/0941-2948/2008/0309

5. Dee D.P. et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quart. J. RMS 137(656), 553–597 (2011), DOI: 10.1002/qj.828

6. Von Storch H., Langenberg H., Feser F. A spectral nudging technique for dynamical downscaling purposes. Mon. Wea. Rev. 128(10), 3664–3673 (2000), DOI: 10.1175/1520-0493(2000)128<3664:ASNTFD>2.0.CO;2

7. Varentsov M.I., Verezemskaya P.S., Zabolotskih E.V., Repina I.A. Evaluation of the quality of polar low reconstruction using reanalysis and regional climate modelling. Sovrem. Probl. Dist. Zond. Zemli iz Kosmosa 4, 168–191 (2016). (in Russian) DOI: 10.21046/2070-7401-2016-13-8-168-191

8. Varentsov M., Wouters H., Platonov V., Konstantinov P. Megacity-induced mesoclimatic effects in the lower atmosphere: A modeling study for multiple summers over Moscow, Russia. Atmosphere (Basel) 9(2), 50–73 (2018), DOI: 10.3390/atmos9020050

9. Wouters H. et al. The efficient urban canopy dependency parametrization (SURY) v1.0 for atmospheric modelling: Description and application with the COSMO-CLM model for a Belgian summer. Geosci. Model Dev. 9(9), 3027–3054 (2016), DOI: 10.5194/gmd-9-3027-2016

10. Han J.Y., Baik J.J., Lee H. Urban impacts on precipitation. Asia-Pacific J. Atmos. Sci. 50(1), 17–30 (2014), DOI: 10.1007/s13143-014-0016-7

11. Kislov A.V. et al. Mesoscale atmospheric modeling of extreme velocities over the sea of Okhotsk and Sakhalin. Izvestiya, Atm. and Ocean. Phys., Pleiades Publishing, Ltd 4(54), 322–326 (2018), DOI: 10.1134/S0001433818040242

12. Cook N.J. Towards better estimation of wind speeds. J. Wind Eng. Ind. Aer. 9, 295–323 (1982), DOI: 10.1016/0167-6105(82)90021-6

DOI: 10.14529/jsfi180321

Supercomputer Simulations in Design of Ultrasound Tomography Devices

Alexander V. Goncharsky[1]***, Sergey Y. Seryozhnikov***[1]

The paper considers the use of supercomputers in design of medical ultrasound tomography devices. The mathematical models describing the wave propagation in ultrasound tomography should take into account such physical phenomena as diffraction, multiple scattering, and so on. The inverse problem of wave tomography is posed as a coefficient inverse problem with respect to the wave propagation velocity and the absorption factor. Numerous simulations made it possible to determine the optimal parameters of an ultrasound tomograph in order to obtain a spatial resolution of 1.5 mm suitable for early-stage breast cancer diagnosis. The developed methods were tested both on model problems and on real data obtained at the experimental test bench for tomographic studies. The computations were performed on GPU devices of Lomonosov-2 supercomputer at Lomonosov Moscow State University.

Keywords: ultrasound tomography, coefficient inverse problem, spatial resolution, supercomputer, GPU.

Introduction

Modern medical tomographs are complex and expensive devices that can not be designed without extensive mathematical modeling. In ultrasound tomography, the mathematical models used to describe the wave propagation process should take into account such physical phenomena as diffraction, refraction, multiple scattering, and so on. Tomographic image reconstruction involves solving nonlinear large-dimensional inverse problems. The methods developed in the 1970s–1990s for solving inverse problems [8, 9] are the most striking mathematical results of the last century. The inverse problem considered in this paper is a coefficient inverse problem [3]. The solution algorithms rely on the processing power of modern GPU clusters.

The concept of resolving power is widely used in the practice of tomographic studies [1]. In quantitative ultrasound tomography [10], both the spatial resolution and the reconstruction accuracy are important. The resolving power of an ultrasound tomograph depends on a large number of parameters, such as the wavelength, the number of sound sources, the number of detectors and distance between them, the frequency spectrum of sounding pulses. For a given set of these parameters, the resolving power depends on the wavefield measurement precision. In this paper we assess the reconstructed image resolution using mathematical modeling, as well as real data obtained in physical experiments.

The images reconstructed from experimental data showed that a spatial resolution of $\approx$ 1.5 mm is attainable in a low-frequency setup with a central wavelength of ≈ 3 mm, which can be implemented in practice. The algorithms developed by the authors in [3, 5] for layer-by-layer ultrasound tomography schemes were used to solve the inverse problems.

1. Formulation of the Inverse Problem of Wave Tomography

A scalar wave model based on a second-order hyperbolic equation is a simple mathematical model that takes into account the effects of ultrasound diffraction and absorption. According to

[1]Lomonosov Moscow State University, Moscow, Russian Federation

this model, acoustic pressure $u(\boldsymbol{r}, t)$ satisfies the equation:

$$c(\boldsymbol{r})u_{tt}(\boldsymbol{r}, t) + a(\boldsymbol{r})u_t(\boldsymbol{r}, t) - \Delta u(\boldsymbol{r}, t) = \delta(\boldsymbol{r} - \boldsymbol{q})f(t); \quad \partial_n u(\boldsymbol{r}, t)|_{ST} = p(\boldsymbol{r}, t). \tag{1}$$

Here, $c(\boldsymbol{r}) = 1/v^2(\boldsymbol{r})$, where $v(\boldsymbol{r})$ is the speed of sound; $\boldsymbol{r} \in \mathbb{R}^2$ is the point in the imaging plane; $a(\boldsymbol{r})$ is the absorption factor; $f(t)$ describes the sounding pulse emitted from point $\boldsymbol{q}$; Δ is the Laplace operator with respect to $\boldsymbol{r}$. The initial conditions are zero. $\partial_n u(\boldsymbol{r}, t)|_{ST}$ is the normal derivative to the surface S of the domain Ω (Fig. 1), where $(\boldsymbol{r}, t) \in S \times (0, T)$; function $p(\boldsymbol{r}, t)$ is known. It is assumed that $v(\boldsymbol{r}) = v_0 = const$, $a(\boldsymbol{r}) = 0$ outside of the studied object. This wave propagation model can be used to describe ultrasound waves in soft tissues.

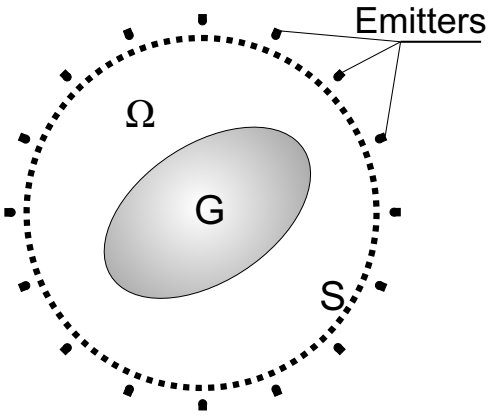


Figure 1. Tomographic examination scheme

In this study, we use a layer-by-layer tomography scheme shown in Fig. 1. Object G is insonified using the ultrasound emitters that successively produce sounding pulses. Acoustic pressure $U(\boldsymbol{s}, t)$ is measured at points $\boldsymbol{s}$ of the boundary S for the time interval $(0; T)$.

The inverse problem of reconstructing the unknown coefficients $c(\boldsymbol{r})$ and $a(\boldsymbol{r})$ in equation (1). This inverse problem is ill-posed, and thus we formulate it as a problem of minimizing the residual functional between the measured and numerically simulated wavefields with respect to its argument $\{c, a\}$:

$$\Phi(u(c, a)) = \frac{1}{2} \int_0^T \int_S (U(\boldsymbol{s}, t) - u(\boldsymbol{s}, t))^2 \, \mathrm{d}\boldsymbol{s} \, \mathrm{d}t. \tag{2}$$

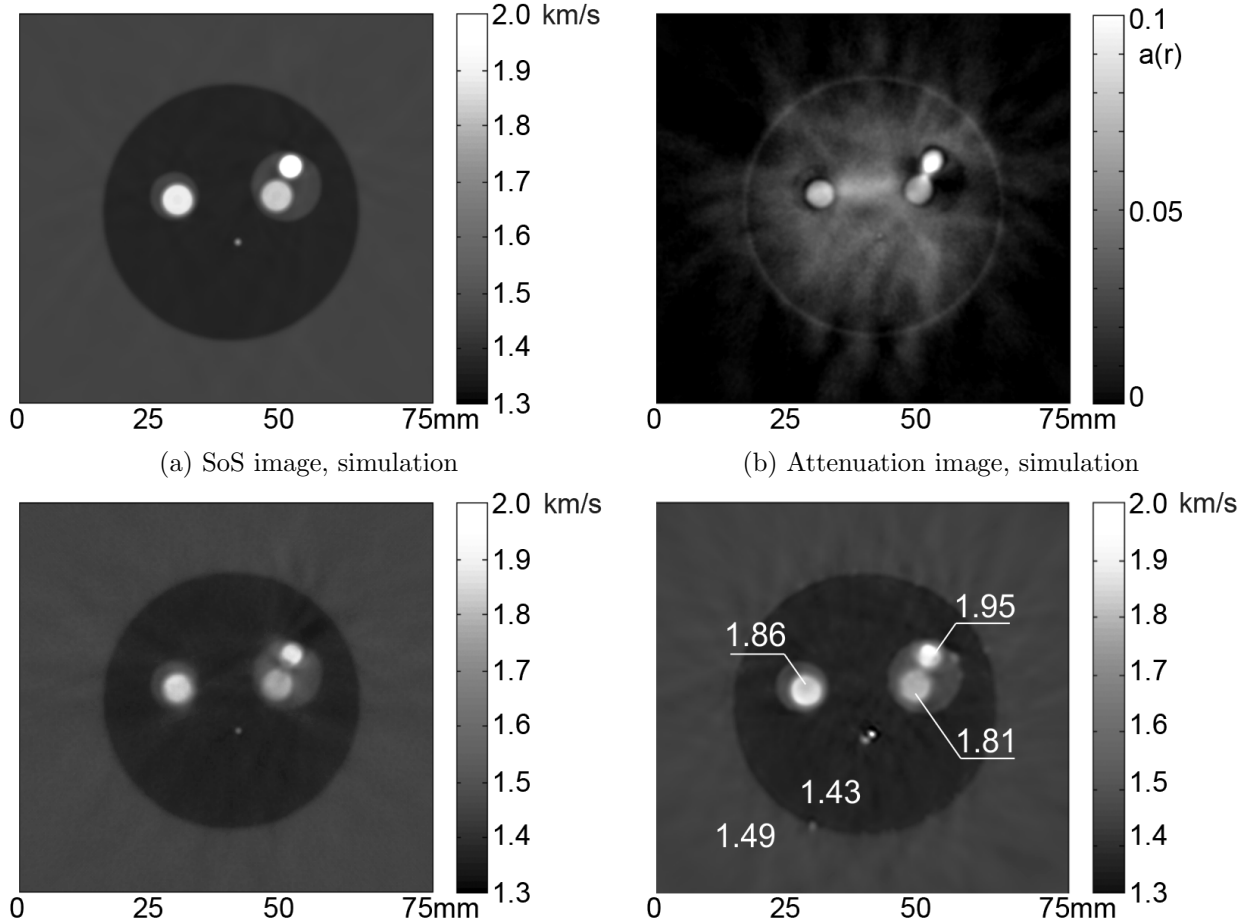
Here, $u(\boldsymbol{s}, t)$ is the solution of the direct problem (1). We use the iterative gradient method [4] to minimize the functional. Representations of the gradient $\Phi'_c(u(c, a))$, $\Phi'_a(u(c, a))$ were obtained in [3, 5]. Finite difference time-domain method [6] was used to compute the wavefields.

2. Numerical Simulations and Experimental Results

By means of numerical simulation, we studied the dependence of the resolving power on the wavelength, the number of emitters and detectors, and other factors. Optimal parameters of an ultrasound tomograph producing a resolving power of ≈ 1.5 mm were determined. Such resolution is sufficient for early-stage breast cancer diagnosis [10].

Figures 2(a) and 2(b), respectively, show the reconstructed speed of sound (SoS) and attenuation images obtained in a numerical simulation. The parameters of the simulated phantom were chosen to match the silicone phantom used in a physical experiment. The minimum distance between the objects is ≈ 1 mm, and the sound speed difference between objects is $\approx 5\%$. The

speed of sound is reconstructed better than the absorption factor. This is a natural consequence of the fact that at sufficiently high frequencies the coefficient of the second derivative in equation (1) is reconstructed better than the coefficient of the first derivative. Figure 2(c) shows the SoS image obtained in a numerical experiment with simulated measurement errors. Figure 2(d) shows the SoS image reconstructed from experimental data.



(a) SoS image, simulation

(b) Attenuation image, simulation

(c) SoS image, simulation with error

(d) SoS image, real data

Figure 2. Reconstructed speed of sound (SoS) and attenuation images

The boundaries of each object are clearly visible. The phantom used in the physical experiment contained a steel needle 1 mm in diameter. The size of the needle in the reconstructed image does not exceed 1.5 mm, and thus we estimate the achieved resolution as 1.5 mm. The developed tomographic methods have high reconstruction accuracy. The numbers in Fig. 2(d) indicate the reconstructed speed of sound inside the respective objects.

The computations were performed on NVidia Tesla K40s devices of Lomonosov-2 supercomputer [7]. GPU computing has proven to be the most promising technology for solving inverse problems of wave tomography [2]. It takes ≈ 15 minutes to obtain images for a single plane on a single GPU device. The finite difference grid contained 1000×1000 points at a pitch of 0.33 mm. The simulation time step was 0.15 μs for a total of 2400 steps. The frequency range of sounding waves was 100–600 kHz with a central wavelength of 3 mm.

Conclusion

It is impossible to design ultrasound tomographs without extensive mathematical modeling. At the design stage it is necessary to perform reconstructions for various phantoms and for a large number of physical parameters of the device. Numerous simulations and processing of experimental data on a supercomputer made it possible to determine the optimal parameters that provide high spatial resolution of ≈ 1.5 mm and high reconstruction accuracy. The physical experiments showed that a total of 20–30 ultrasound emitters are sufficient to achieve such resolution, while the distance between detectors should be approximately half the wavelength.

Acknowledgments

This research is supported by Russian Science Foundation (project No. 17-11-01065). The research is carried out at Lomonosov Moscow State University. The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011.

References

1. Birk, M., Dapp, R., Ruiter, N.V., Becker, J.: GPU-based iterative transmission reconstruction in 3D ultrasound computer tomography. J. Parallel Distrib. Comput. 74, 1730–1743 (2014), DOI: 10.1016/j.jpdc.2013.09.007

2. Goncharsky, A.V., Seryozhnikov, S.Y.: The Architecture of Specialized GPU Clusters Used for Solving the Inverse Problems of 3D Low-Frequency Ultrasonic Tomography. In: Voevodin, V., Sobolev, S. (eds.) Supercomputing. RuSCDays 2017. Communications in Computer and Information Science. vol. 793, pp. 363–395. Springer (2017), DOI: 10.1007/978-3-319-71255-0_29

3. Goncharsky, A.V., Romanov, S.Y.: Inverse problems of ultrasound tomography in models with attenuation. Phys. Med. Biol. 59(8), 1979–2004 (2014), DOI: 10.1088/0031-9155/59/8/1979

4. Goncharsky, A.V., Romanov, S.Y.: Iterative methods for solving coefficient inverse problems of wave tomography in models with attenuation. Inverse Probl. 33(2), 025003 (2017), DOI: 10.1088/1361-6420/33/2/025003

5. Goncharsky, A., Romanov, S., Seryozhnikov, S.: A computer simulation study of soft tissue characterization using low-frequency ultrasonic tomography. Ultrasonics 67, 136–150 (2016), DOI: 10.1016/j.ultras.2016.01.008

6. Romanov, S.: Optimization of Numerical Algorithms for Solving Inverse Problems of Ultrasonic Tomography on a Supercomputer. In: Voevodin, V., Sobolev, S. (eds.) Supercomputing. RuSCDays 2017. Communications in Computer and Information Science. vol. 793, pp. 67–79. Springer (2017), DOI: 10.1007/978-3-319-71255-0_6

7. Sadovnichy, V., Tikhonravov, A., Voevodin, Vl., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale toward Exascale. pp. 287–307. CRC Press, Boca Raton, USA (2013)

8. Tikhonov, A.N.: Solution of incorrectly formulated problems and the regularization method. Soviet Math. Dokl. 4, 1035–1038 (1963)

9. Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V., Yagola, A.G.: Numerical Methods for the Solution of Ill-Posed Problems. Springer Netherlands (1995), DOI: 10.1007/978-94-015-8480-7

10. Wiskin, J., Borup, D., Andre, M., Johnson, S., Greenleaf, J., Parisky, Y., Klock, J.: Three-dimensional nonlinear inverse scattering: quantitative transmission algorithms, refraction corrected reflection, scanner design, and clinical results. J. Acoust. Soc. Am. 133(5), 3229–3229 (2013), DOI: 10.1121/1.4805138

DOI: 10.14529/jsfi180322

Recent Progress on Supercomputer Modelling of High-Speed Rarefied Gas Flows Using Kinetic Equations

Anna A. Frolova[1], *Vladimir A. Titarev*[2]

Numerical solution of the Boltzmann equation for stationary high-speed flows around complex three-dimensional bodies is an extremely difficult computational problem. This is because of high dimension of the equation and lack of efficient implicit methods for the calculation of the collision integral on arbitrary non-uniform velocity grids. Therefore, the use of the so-called model (approximate) kinetic equations appears to be more appropriate and attractive. This article uses the numerical methodology recently developed by the second author which includes an implicit method for solving the approximating kinetic equation of E.M. Shakhov (S-model)on arbitrary unstructured grids in both velocity and physical spaces.

Since most of model equations have a well-known drawback associated with the velocity-independent collision frequency it is important to determine the deviations of solutions of these equations from the solution of the complete Boltzmann equation or DSMC for high-speed gas flows. Our recent comparison of the DSMC and S-model solutions for monatomic gases with a soft interaction potential shows good agreement of surface coefficients of the pressure, heat transfer and friction, which are most important for industrial applications.

In this paper, we compare the solution of model equations and the Boltzmann equation for the problem of supersonic gas flow around a cylinder when molecules interact according to the law of hard spheres. Since this law of molecular interaction is the most rigid, the difference in solutions can show the maximum error that can be obtained by using model equations instead of the exact Boltzmann equation in such problems. Our high-fidelity computations show that the use of model kinetic equations with adaptation in phase space is very promising for industrial applications.

Keywords: Boltzmann kinetic equation, S-model, rarefied, high-speed, unstructured.

Introduction

At present there is a large number of studies devoted to the analysis of highly non-equilibrium external rarefied gas flows at high-speed ($M_\infty \geq 10$) regimes. Since for high-speed flows over convex body there appears strong non-equilibrium boundary layer, it is important to use the so-called kinetic approaches in the analysis. However, this class of flows requires the use of considerable computing resources in case of three-dimensional geometries if one uses the direct simulation Monte-Carlo methods (see e.g. [1]), and especially the Boltzmann kinetic equation with the exact collision integral (BKE). Therefore, for engineering applications the use of the approximate (so-called model) kinetic equations appears to be an attractive alternative from the computational point of view.

It is important to estimate possible error of such an approach, by comparing the numerical solutions of the model kinetic equations with the corresponding solutions of BKE or the direct simulation Monte-Carlo (DSMC) solutions. Calculations presented in [8] demonstrate that the pressure, friction and heat transfer coefficients on the surface of the body in case of monatomic gas at super- and high-speed flow regimes are very close to the DSMC results for the S-model kinetic equation of E.M. Shakhov [5] and relatively "soft" viscosity law.

In this work we compare the numerical solutions of the model kinetic equations and BKE for the supersonic flow over the circular cylinder using the viscosity law, which corresponds to

[1]Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

the hard-sphere intermolecular interaction. Since such a viscosity law is the most "rigid", the difference between solutions will show the maximum possible error which can occur due to the usage of the model kinetic equations instead of BKE.

1. Formulation of the Problem and Governing Equations

Let us consider the supersonic ($M_\infty = 10$) flow of argon over the circular cylinder of $r_{\text{cyl}} = 6$ inches (15.24 cm) radius. The free stream density, velocity and temperature are $\rho_\infty = 1.127 \times 10^{-6}\ kg/m^3$, $u_\infty = 2624\ m/s$, $T_\infty = 200$ K, respectively. The viscosity law is $\mu = \mu_\infty \sqrt{T/T_\infty}$ (hard-sphere intermolecular interaction), where the free-stream viscosity coefficient $\mu_\infty = 1.554409 \times 10^{-5}$ Pa $\cdot$ s [3]. The fixed surface temperature $T_w = 2.5T_\infty$ is used, at which we assume the diffusive boundary condition with the complete temperature accommodation. The rarefaction parameter is $\delta = (p_\infty r_{\text{cyl}})/(\mu_\infty \sqrt{2RT_\infty}) = 1.59$, which corresponds to the Knudsen number $\text{Kn} = \lambda_\infty/\text{r}_{\text{cyl}} = 0.56$. Therefore, the flow regime is rarefied and requires the use of the kinetic approaches.

The state of rarefied gas is determined by the velocity distribution function $f(t, \boldsymbol{x}, \boldsymbol{\xi})$, where t is physical time, $\boldsymbol{\xi} = (\xi_x, \xi_y, \xi_z)$ are the components of the molecular velocity vector in the spatial directions $\boldsymbol{x} = (x_1, x_2, x_3) = (x, y, z)$. The macroscopic quantities are defined as three-dimensional integrals of the velocity distribution function over the molecular velocity space as

$$\rho = m \int f d\xi, \quad \rho u_k = m \int \xi_k f d\xi, \quad \frac{3}{2} p + \frac{1}{2} \rho u_{\alpha\alpha} = \frac{1}{2} m \int \xi^2 f d\xi.$$

The distribution function is found by solving the kinetic equation with appropriate boundary conditions at free stream and surface of the body:

$$\frac{\partial f}{\partial t} + \boldsymbol{\xi} \nabla f = J(f).$$

For the collision integral $J(f)$ we consider both the exact expression (BKE) and approximate form by E.M. Shakhov [5] (the so-called S-model kinetic equation). BKE collision integral involves complicated five-dimensional integration over the velocity space and two impact parameters, whereas for the S-model it is a much simpler relaxation term.

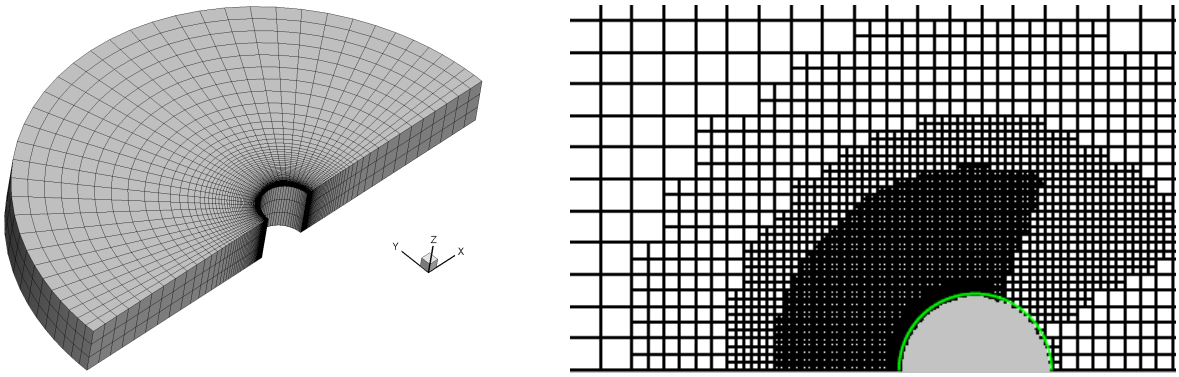
To compare the numerical results of BKE and S-model equation, we use the non-dimensional surface pressure, friction and heat transfer coefficients, defined as follows:

$$c_p = \frac{p_n - p_\infty}{\rho u_\infty^2/2}, \quad c_f = \frac{p_\tau}{\rho u_\infty^2/2}, \quad c_h = \frac{E_n}{\rho u_\infty^3/2}, \quad (p_n, p_\tau, E_n) = m \int \xi_n (\xi_n, \xi_\tau, \frac{1}{2}\xi^2) f d\xi,$$

where ξ_n and ξ_τ are normal and tangential projections of the molecular velocity vector $\boldsymbol{\xi}$.

2. Kinetic Solvers and Details of the Calculations

In our calculations we use two different numerical methods and software packages. The first package is the parallel computational code "Nesvetay" [7, 8] developed by the second author over the recent years. It solves model kinetic equations of the monatomic gas in three space dimensions. The planar flows are modelled using special arrangements of the spatial mesh. The method of solution is an original version of the discrete velocity approach, which combines explicit and implicit TVD schemes with the conservative evaluation of the collision integral on arbitrary unstructured meshes in both physical and velocity meshes. In particular, fully implicit



(a) "Nesvetay"mesh (b) UFS mesh (part of the domain)

Figure 1. Spatial meshes, used in computations

method and high CFL numbers up to 10^5 are used to construct the steady-state solution of the problem, which speeds up convergence at rarefied regimes by 3 orders of magnitude. For present calculations the physical mesh in $x-y$ plane contains 115×40 and 3 cells in the z direction, see Fig. 1. The first cell size near the surface is $h/r_{\text{cyl}} = 10^{-4}$. For this simple flow it is possible to use a relatively fine three-dimensional molecular velocity mesh with $\Delta\xi/\sqrt{2RT_\infty} = 0.5$, consisting of 247656 nodes. The total number of cells in the 6-dimensional mesh in the phase space is thus up to ≈ 3.5 **billions**.

The second computation package used in the present study is the kinetic module of the Unifed Flow Solver (UFS) [2] which can solve both BKE and model kinetic equations. The numerical solution is found by the discrete ordinate method on the uniform molecular velocity mesh. In the physical space the solver uses a cubical mesh with adaptive mesh refinement and a finite-volume TVD scheme. The surface of the body is approximated using the cut cell approach. The steady-state solution is found by marching in time using an explicit time evolution method. The number of cells in the planar physical domain changes during calculation from 2000 to 6000 due to mesh refinement. The first cell size near the surface $h/r_{\text{cyl}} = 3.2 \times 10^{-3}$ is much smaller than the local Knudsen number $\lambda_{\text{local}}/r_{\text{cyl}} = 10^{-2}$. To compute the exact collision integral, the Korobov sequences for the velocity nodes, and conservative two-point projection method for post-collision velocities are used [6]. The number of velocity cells N_v is kept constant during calculation, with the cell size of $\Delta\xi/\sqrt{2RT_\infty} \approx 1$. For the model kinetic equation reduced $\xi_x - \xi_y$ velocity domain contains $N_v = 5000$ nodes, whereas for BKE we have $N_v \approx 30000$ (the half-sphere domain is used). The number of collisions $N_{col} \approx 10^6$. The total number of cells for BKE is up to ≈ 0.2 billions.

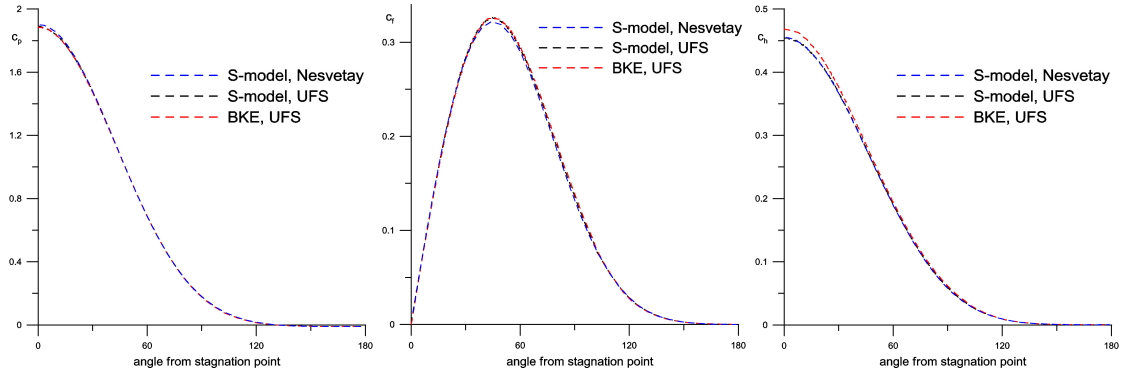
It should be emphasised that despite much smaller number of cells, the numerical solution of BKE is extremely computationally intensive due to the five-dimensional BKE collision integral and the use of the explicit time evolution.

3. Results

Since a typical calculation of a high-speed ($M_\infty \geq 10$) flow can utilize billions of cells in the computation mesh, an efficient parallel implementation of the solution methods is essential. "Nesvetay"employs two-level MPI+OpenMP parallel implementation with mesh decomposition in both physical and velocity spaces. The code was shown in [8] to scale to tens of thousands of hyperthreads for three-dimensional steady calculations. In the present work all calculations

by "Nesvetay" were carried out on "Lomonosov-2" supercomputer at Lomonosov Moscow State University [4], using up to 512 CPU cores.

The parallel implementation of the UFS is single-level and uses the decomposition in the physical space. An optimal balance between processors is archived by using the space filling curves. This approach is especially efficient for hieratical adaptive meshes, e.g. octree in 3D. The calculations were run at Joint Supercomputing Center of Russian Academy of Sciences using up to 512 CPU cores.



(a) pressure coefficient c_p (b) friction coefficient c_f (c) heat transfer coefficient c_h

Figure 2. Surface distributions as predicted by BKE and S-model equations. Blue line – S-model, "Nesvetay"code; black line – S-model, UFS code; red line – BKE, UFS code

Fig. 2 shows comparison of numerical results from UFS solvers (BKE and S-model) and "Nesvetay" solver (S-model) for $M_\infty = 10$, $\delta = 1.59$ (Kn $= 0.56$). Firstly, we see that the S-model equation results from both solvers agree perfectly even though these solvers use very different meshes and numerical methods. Secondly, BKE and S-model equation results coincide for pressure and skin friction coefficients and are very close for the most sensitive heat transfer coefficient. The difference for c_h is around 2% at the stagnation point.

Conclusions

We have computed supersonic rarefied flow of argon over circular cylinder, using the Boltzmann kinetic equation with the exact collision integral and Shakhov model collision integral. Fine spatial and velocity meshes have been used to obtain accurate results. Our calculations have good to excellent agreement between solutions of exact and approximate (model) kinetic equations for this supersonic external flow. Especially important is the agreement of heat transfer coefficient, which is a very sensitive quantity to compute. This result allows us to conclude that model kinetic equations, such as S-model equation [5], can potentially complement DSMC approach for high-speed flows of rarefied gas in industrial applications, such as modelling of aerodynamics and heat transfer of existing and future spacecraft.

Acknowledgments

The work was supported by the Russian Foundation for Basic Research, projects 18-08-00501, 18-07-01500. The research is carried out using the equipment of the shared research

facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 and Joint Supercomputing Center of Russian Academy of Sciences.

References

1. Kashkovsky, A., Bondar, Y., Zhukova, G., Ivanov, M., Gimelshein, S.: Object-oriented software design of rReal gas effects for the DSMC method. In: 24th International Symposium on Rarefied Gas Dynamics. AIP Conference Proceedings. vol. 762, pp. 583–588 (2004). DOI: 10.1063/1.1941599

2. Kolobov, V., Arslanbekov, R., Aristov, V., Frolova, A., Zabelok, S.: Unified solver for rarefied and continuum flows with adaptive mesh and algorithm refinement. J. Comput. Phys. 223, 589–608 (2007). DOI: 10.1016/j.jcp.2006.09.021

3. Lofthouse, A.: Nonequilibrium hypersonic aerothermodynamics using the direct simulation Monte Carlo and Navier-Stokes models. In: Ph.D. thesis. The University of Michigan (2008)

4. Sadovnichy, V., Tikhonravov, A., Voevodin, Vl., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale toward Exascale. pp. 283–307. Chapman & Hall/CRC Computational Science, CRC Press, Boca Raton, USA (2013)

5. Shakhov, E.: Generalization of the krook kinetic relaxation equation. Fluid Dynamics 3(5), 95–96 (1968), DOI: 10.1007/bf01029546

6. Tcheremissine, F.G.: Solution to the Boltzmann kinetic equation for high-speed flows. Computational Mathematics and Mathematical Physics 46(2), 315–329 (2006). DOI: 10.1134/s0965542506020138

7. Titarev, V.A.: Efficient deterministic modelling of three-dimensional rarefied gas flows. Commun. Comput. Phys. 12(1), 161–192 (2012). DOI: 10.4208/cicp.220111.140711a

8. Titarev, V.A.: Application of model kinetic equations to hypersonic rarefied gas flows. Computers & Fluids, Special issue "Nonlinear flow and transport" 169, 62–70 (2018). DOI: 10.1016/j.compfluid.2017.06.019

DOI: 10.14529/jsfi180323

Supercomputer Modeling of Parachute Flight Dynamics

Andrey A. Aparinov[1], *Vladimir A. Aparinov*[2], *Alexey V. Setukha*[3]

In this article the authors present parallel implementation of numerical method for computer modeling of dynamics of a parachute with filled canopy. To solve the 3D problem of parachute free motion numerically, authors formulate tied problem of dynamics and aerodynamics where aerodynamic characteristics are found with discrete vortices method on each step of integration in time, and to find motion law the corresponding motion equations have to be solved. The solution of such problems requires high computational resources because it is important to model parachute motion during a long physical time period. Herewith the behavior of vortex wake behind the parachute is important and has to be modeled. In the approach applied by the authors the wake is modeled as a set of flexible vortex elements. So to increase computational efficiency, the authors used methods of low-rank matrix approximations, as well as parallel implementations of algorithms. Short description of numerical method is presented, as well as the examples of numerical modeling.

Keywords: parallel algorithms, numerical simulation methods, fluid dynamics, vortex methods, parachute aerodynamics.

Introduction

Parachute is a complex aeroelastic system the geometrical form of which appears as a result of aerodynamic and elastic forces interaction. That is why the first problem in computer modeling of a parachute is to simulate the overall process of canopy geometry generation in steady flow. For its solution authors developed mathematical model based on simultaneous application of bars method and lumped-mass method to describe canopy deformations [1], and discrete vortices method to model the flow past canopy [2]. Modeling a real parachute flight requires additional development of this model to simulate the dynamics of the parachute.

In this article the authors describe an approach based on solution of the above problem in two steps. On the first step we find the aeroelastic form of canopy in the assumption of steady flow past it using the vortex method and lumped-mass method. During aerodynamic computation the canopy shape and stress-strain behavior characteristics are adjusted on each step of iterational process based on the pressure distribution got from aerodynamic calculation. Then, final canopy shape obtained on first step of algorithm is used to solve the tied problem of aerodynamics and dynamics of parachute flight where canopy is supposed to be stiff.

It is notable that the solution of such problems requires high computational resources because it is important to model parachute motion during a long physical time period. Herewith the behavior of vortex wake behind the parachute is important and has to be modeled. In the approach applied by the authors the wake is modeled as a set of flexible vortex elements. So to increase computational efficiency, the authors used methods of low-rank matrix approximations, as well as parallel implementations of algorithms.

[1]Central Aerohydrodynamic Institute, Zhukovsky, Russian Federation
[2]Parachute Building Institute, Moscow, Russian Federation
[3]Lomonosov Moscow State University, Moscow, Russian Federation

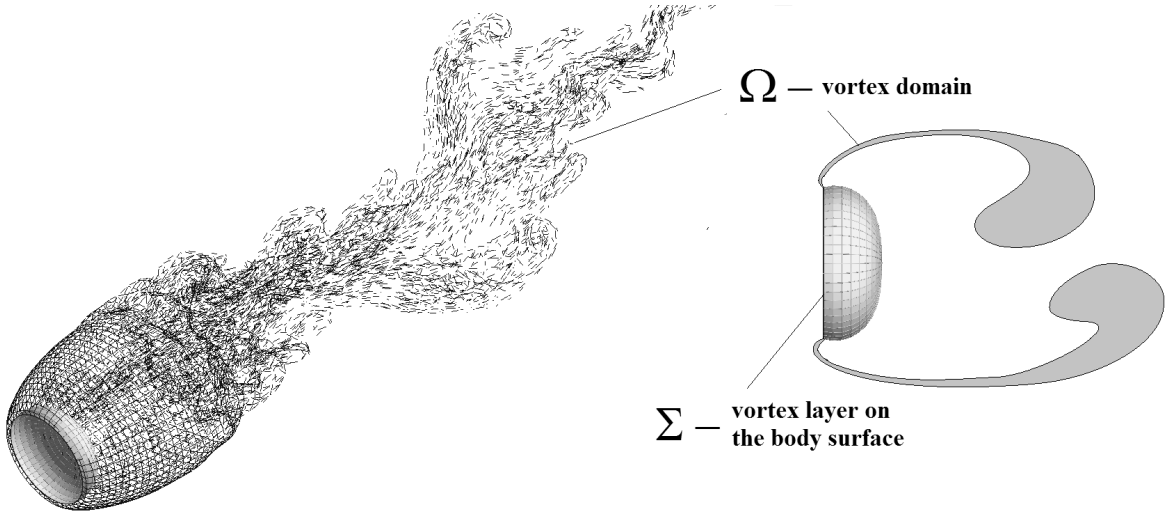


Figure 1. Vortex method

1. About Numerical Method

The authors consider the 3D problem of separated flow past bodies within the model of inviscid non-compressible fluid. The main idea of vortex methods is to utilise integral representation of velocity field $\vec{w}$ through vorticity $\vec{\omega} = rot\vec{w}$. With this it is assumed that vorticity is concentrated locally in bounded domains Ω and thin layer on the body surface Σ (Fig. 2). Then velocity field can be represented as

$$\vec{w} = \vec{w}_\infty + \vec{w}_1 + \vec{w}_2, \quad \vec{w}_1(x,t) = \int_\Omega \vec{\omega}(y,t) \times \vec{V}(x-y)\, dy, \quad \vec{w}_2(x) = \int_\Sigma \vec{\gamma}(y,t) \times \vec{V}(x-y)\, d\sigma_y,$$

where $\vec{V}(x-y) = (y-x)(1/(4\pi|x-y|^3))$, function γ is surface density of vorticity distribution, $\vec{w}_\infty$ – flow velocity on infinity.

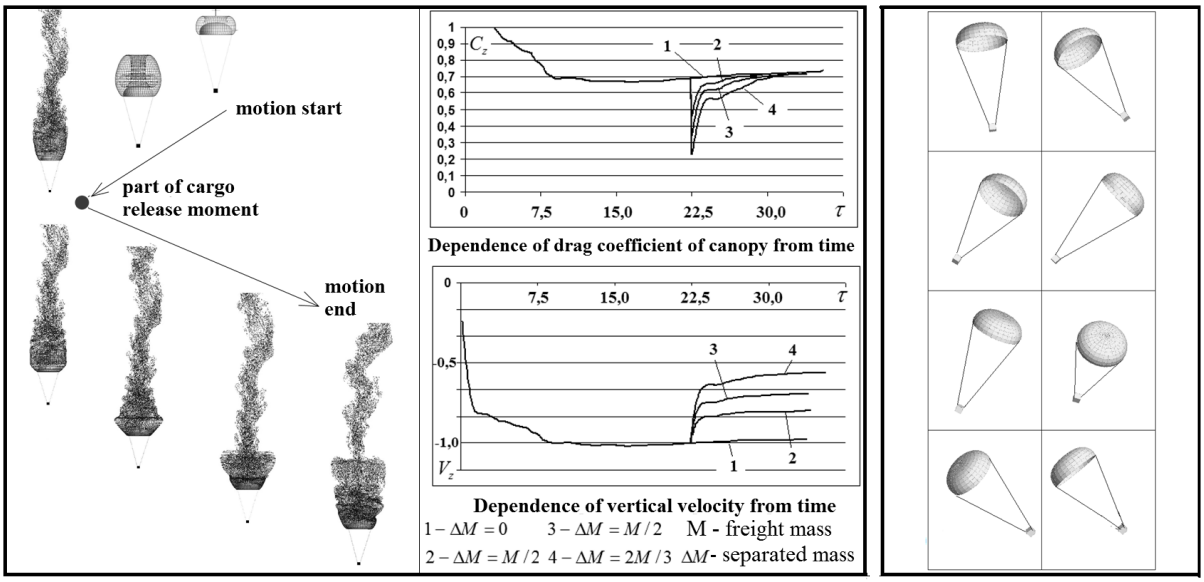
The authors use numerical model (see [2]) where vortex domain Ω is approximated with the system of discrete vortex segments l_i, each of which has the beginning in point x_i^-, and the end in point x_i^+, and has some vortex intensity $\vec{\omega}_i$, $i = 1, ..., N_\Omega$. Body surface Σ is approximated with set of cells σ_i, $i = 1, ..., N$, each cell is bounded with closed vortex line with intensity $g_i = g_i(t)$. Velocity field is approximated with expression:

$$\vec{w}_1(x) \approx \sum_{i=1,N_\Omega} \vec{\omega}_i \left[(x_i^+ - x_i^-) \times \vec{V}(x - x_i)\right], \quad \vec{w}_2(x) \approx \sum_{i=1,N} g_i \int_{\partial\sigma_i} \vec{dl}_y \times \vec{V}(x-y),$$

$x_i = (x_i^- + x_i^+)/2$. Approximation of vorticity transport laws is assured if the following motion equations for the ends of vortex segments are fulfilled in domain Ω:

$$\frac{dx_i^\pm}{dt} = \vec{w}(x_i^\pm, t), i = 1, ..., N_\Omega, \quad \frac{\vec{\omega}_i}{dt} = 0. \tag{1}$$

In the numerical solution of the problem on each time integration step we assume known the positions of vortex segments in the domain approximating vorticity area and their intensities ω_i. Unknown intensities g_i on the body surface are calculated from the system of linear equations representing non-penetration condition of fluid on body surface, which is checked in collocation



(a) Separation of part of freight (b) Self-rotation

Figure 2. Simulation examples

points $x_i \in \sigma_i$, $i = 1, ..., N$:

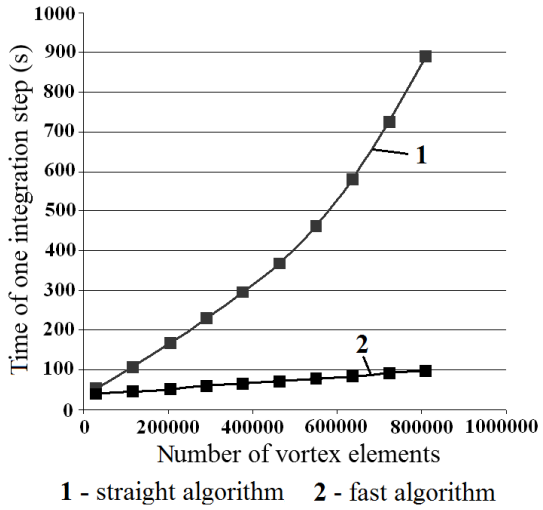
$$\sum_{j=1,N} g_j(\vec{W}_j(x_i)\vec{n}(x_i)) = -(\vec{w}_\infty + \vec{w}_2(x_i))\vec{n}(x_i). \tag{2}$$

Then we shift the ends of vortex segments approximating vortex domain in accordance to equations (1) – here Euler scheme of the first order by the time is used. Then the process of new vortex elements birth is modelled on body surface and appending of vortex domain by them. These segments appear on predefined separation lines or on the full surface of the body.

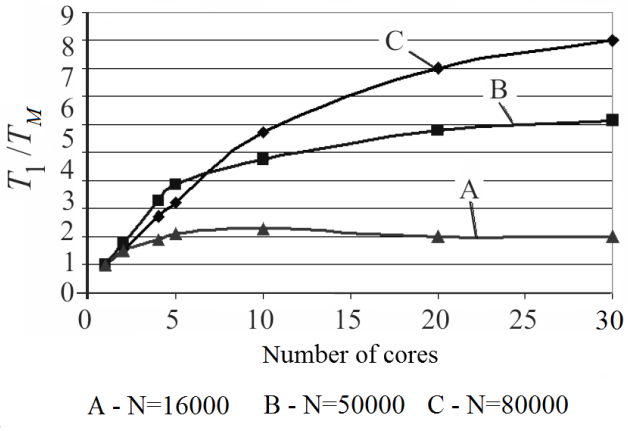
In case when body makes free motion, the solid body dynamics equations also have to be written.

2. Computational Examples and Discussion

Considered numerical modeling of vertical flight with separation of part of freight. It is known from practice that at the moment of freight part separation the dramatic decrease of loads on canopy occurs. This leads to the breath effect on canopy and in some cases may lead to canopy collapse. So numerical modeling of the separation of part of freight is the important problem which helps to make right evaluation of freight part that can be securely separated. From the descriptions above it follows that at the moment of part of cargo release the parachute slows down and part of vortex wake that was behind canopy overtakes it. Herewith the dramatic decrease of pressure drop coefficient happens over all canopy surface, and consequently, the decrease of drag coefficient. Figure 2a (left side) shows the shape of vortex structures at different moments of time. On the right the dependency plot of parachute drag coefficient and vertical velocity from measureless time is shown. Let us note that measureless time τ, measureless velocity V and drag coefficient C_x were introduced as $\tau = t\ V_0/D$, $V = V^*/V_0$, $C_x = 2\ F_x/(\rho V_0^2 S)$, where t is physical time, V – physical velocity, D – diameter of filled canopy, V_0 – steady vertical velocity of parachute with cargo, F_x – drag force of air, ρ – air density, S – canopy square. The results obtained in numerical experiments showed good qualitative correspondence with observations.



(a) Comparison of fast and straight algorithms



(b) Acceleration of one time-integration step for various number of cores

Figure 3. Algorithm efficiency

For the parachute with fixed canopy authors made computations to model parachute motion from angled initial position as the next example. The transition to self-rotation mode is shown on the Fig. 2b.

To solve the problem in the example, the authors were to use more than 800000 approximating vortex elements. Figure 3a shows time comparison of one integration step made with fast (methods of low-rank matrix approximations) and straight algorithms. In both cases 128 processors were used, calculations were made on supercomputer Lomonosov in Lomonosov Moscow State University supercomputer center. The supercomputer node utilised in calculations had the following characteristics: 2 x Xeon 5570/2.93 GHz, memory 3 Gb per core. It is notable that straight algorithm gives quadratic dependence of time growth from element number growth, and fast algorithms shows linear time growth (theoretical estimate $O(Nlog^4N)$). Figure 3b shows dependence of vortex structures transformation time (fast algorithm) from the number of processors used in calculations and the number of vortex elements in the wake (M – cores quantity, T_1 – time of one integration step on one core, T_M – time of one integration step on M cores, N – number of vortex elements). Details of parallel algorithm for vortex method are presented in [3, 4].

Conclusion

The numerical method for the solution of tied aerodynamics and dynamics problems in parachute flight was presented. Numerical modelling of the separated flow past the moving parachute canopy was carried out with the vortex method. So the vortex wake was considered as an ensemble of moving vortex particles. The simultaneous application of the low-rank matrix approximation method together with the parallel implementation of the algorithm, made it possible to significantly increase the complexity of problems that can be solved. For example, in calculations on 128 cores, the number of vortex elements reached 10^6. Therefore the results obtained numerically in the considered problems showed good qualitative correspondence with observations.

In practice such approach can be applied to solve the following problems: investigate the behavior of a parachute with fixed thimble vectored from equilibrium state, and to find stable balance positions (it is important, for example, for brake parachute); modeling of development of slight disturbances in steady motion of ballistic parachute; modeling of dynamic of parachute motion with changing freight mass (here the most interesting case is separation of part of cargo during steady parachute descent which may lead to the system rollover).

Acknowledgments

The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011.

References

1. Leonov, S.V., Morozov, V.I., Ponomarev, A.T.: Shape modeling and strenght analysis of parachutes. Mechanics of Solids. 46(2), 311–324 (2011), DOI: 10.3103/S0025654411020208

2. Aparinov, A.A., Setukha, A.V.: Application of mosaic-skeleton approximations in the simulation of three-dimensional vortex flows by vortex segments. Computational Mathematics and Mathematical Physics 50(5), 890–899 (2010), DOI: 10.1134/S096554251005012X

3. Aparinov, A.A., Setukha, A.V.: Parallelization in the vortex method for solving aerodynamic problems. Vychisl. Metody Programmir. 14, 406–418 (2013). `http://num-meth.srcc.msu.ru/zhurnal/tom_2013/pdf/v14r147.pdf` (in Russian)

4. Aparinov, A.A., Setukha, A.V., Stavtsev, S.L.: Parallel implementation for some applications of integral equations method. Lobachevskii Journal of Mathematics 39(4), 477–485 (2018), DOI: 10.1134/S1995080218040029

DOI: 10.14529/jsfi180324

Supercomputer Simulation of MATIS-H Problem

Mikhail A. Zaitsev[1], *Vasilij M. Goloviznin*[1], *Sergej A. Karabasov*[2]

A supercomputer simulation of the benchmark MATiS-H problem is considered. A high-resolution CABARET code is applied for solving Navier-Stokes equations in the framework of the Monotonically Integrated LES approach for the MATiS-H problem. The code is based on a generalisation of low-dissipative, low-dispersive and non-oscillatory CABARET scheme to hybrid topology meshes in the supercomputing framework. The solutions for the time-averaged fields are reported. These show a relatively small sensitivity to the grid density. Comparison with the experiment data available is provided.

Keywords: Navier-Stokes equations, parallel computation, modelling, Cabaret method.

Introduction

For modelling of unsteady flow mixing in the MATiS-H rod-bundle configuration with a spacer grid [1], the Monotonically Integrated LES (MILES) approach of Fureby and Grinstein [2] is used. Similar to other Implicit LES techniques, this approach does not include any explicit sub-grid-scale closure. The sub-grid-scales are modelled implicitly on the assumption that the shock-capturing and the good linear wave properties of the numerical scheme are most important for accurate transport of localised vortical structures embedded in a weak background vorticity that has direct relevance for the correct representation of turbulence energy cascade. Indeed, the capability of capturing sharp velocity gradients for smallest resolved scales leads to the emulation of high-wavenumber end of the inertial subrange region characterised by thin filaments of intense vorticity. The above assumption was indirectly confirmed in the work of Fureby and Grinstein who showed that a MILES model of sub-grid-scale closure can be rearranged to a non-linear dissipation filter form typical of the classical LES approach and leads to encouragingly accurate results for the solution of several shear and channel flow problems in comparison with the LES solutions obtained based on explicit sub-grid-scale modelling. For the MILES approach, the use of high-resolution shock-capturing numerical methods that are also low dissipative and low dispersive is important. The CABARET scheme is one example of such methods. CABARET stands for Compact Accurately Boundary Adjusting high-REsolution Technique which is an original computational method for convection/advection dominated flow modeling that has been developed over the years, starting from a 1D advection equation to a highly parallelized 3D unstructured Navier-Stokes code that allows one to conduct eddy-resolving simulations with correct capture of inertial range of turbulent kinetic energy spectra up to very high frequencies. Characteristic features of CABARET include its low dispersion and low dissipation. For linear advection the dispersion error of the CABARET method is as low as that of fourth order finite-difference schemes, yet it is shock-capturing, robust for non-uniform grids, fast and as simple for implementation as conventional shock-capturing finite-volume schemes, e.g., Total Variation Diminishing (TVD) Schemes. In comparison with the latter, the dispersion and dissipation errors of CABARET scheme are a few orders of magnitude lower. Up to present, CABARET algorithm has been found very efficient for a number of Computational Fluid Dynamics (CFD) problems including the modeling of forced turbulence, high-speed jet and airfoil flows and also

[1]Nuclear Safety Institute, Moscow, Russian Federation
[2]Queen Mary University of London, London, UK

turbulent flow mixing in a T-Junction configuration corresponding to a OECD/NRS blind test exercise [3–6]. Another important property of the CABARET method is the compactness of its computational stencil. This makes CABARET scheme very affordable for supercomputing because of the stencil simplicity and low communication overheads. In the current work, for most flexible treatment of the complex spacer grid geometry, the previous CABARET code that has been available only for the hexagonal grid topology has been extended to hybrid meshes.

1. The Governing Problem Formulation and Parallel Numerical Solution Procedure

The unsteady flow in the MATiS-H configuration is modelled by the system of slightly compressible Navier-Stokes equations:

$$\begin{aligned}
&\frac{\partial p}{\partial t} + \rho_0 c^2 (\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}) = 0 \\
&\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} + \frac{1}{\rho_0}\frac{\partial p}{\partial x} = \nu(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}) \\
&\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} + \frac{1}{\rho_0}\frac{\partial p}{\partial y} = \nu(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}) \\
&\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z} + \frac{1}{\rho_0}\frac{\partial p}{\partial z} = \nu(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}).
\end{aligned}$$

Here ρ_0 – density, c – sound speed, x, y, z – coordinates, t – time, u, v, w – velocity vector components, ν – kinematic viscosity, p – hydrostatic pressure.

Laminar inflow boundary conditions are specified at about 10 hydraulic diameters upstream of the space grid. The meanflow parameters for the inflow input, as well as the initial conditions are obtained from a precursor k-e solution of Reynolds-Averaged Navier-Stokes (RANS) equations on the same grid. The numerical solution procedure of the unsteady flow equations consists of 3 phases [3]. Phase 1 and phase 3 correspond to the conservation predictor and corrector steps of the CABARET scheme and stand for the conservation balance laws. Conservation flux of the CABARET scheme is calculated at phase 1 and phase 3. Phase 2 is a characteristic decomposition step. During this stage, the one-dimensional Riemann variables are defined for each face normal projection. The characteristic values at the new time level are obtained from a second-order extrapolation according to the characteristic directions.

The values of the Riemann variables computed are truncated if found to lie outiside the allowable min/max in accordance with the discrete maximum principle. The algorithm is implemented in a highly scalable hybrid mesh framework based on a custom-made MPI parallelisation with a very good efficiency of up to 2,000 computational cores for problem sizes up to several million cells.

2. Parallel Decomposition and I/O Details for Supercomputing

Because of the large-eddy-resolving nature of simulations typical of the MILES approach, LOMONOSOV Supercomputing Facility of Moscow State University has been used. In parallel, several calculations are performed on a dedicated computer cluster at the Moscow Institute of Nuclear Safety (IBRAE) of Russian Academy of Science. The use of a massively parallel computational environment in the framework of the complex grid geometry of MATiS-H required an additional upgrade of the CABARET code as outlined below. The hybrid mesh model generated with a commercial grid generation package has been exported to OpenFOAM. The latter is a free-source software package that includes a grid decomposition kit that is fairly reliable for

parallel computations. Despite the general robustness of this software, we had some problems during the processing of the models, the size which exceeds 20×10^6 cells as discussed below. During the mesh decomposition over 1000 processors, 8 files in each of 1000 directories have been created, and because of the time-transient calculations during the post-processing stage this can quickly end up with 1,000,000 files which is far above the admissible limit of the number files for a supercomputer file system. Hence, we implemented special Fortran procedures to relax the load on the file system and simplify the post processing. The parallel decomposition for the 4.5×10^6 and 13.1×10^6 models has been conducted with METIS. For the model of 47×10^6 cells, ManualFile is used. For this decomposition, the scalability of the 47×10^6 model up to 1800 computational cores is about 75% of the linear speed-up rate, which is acceptable in our application. Figure 1 (left) shows a cross-section of the computational core enumeration for the ManualFile method. For calculations with 1800 cores, the number of cores along the stream is 72.

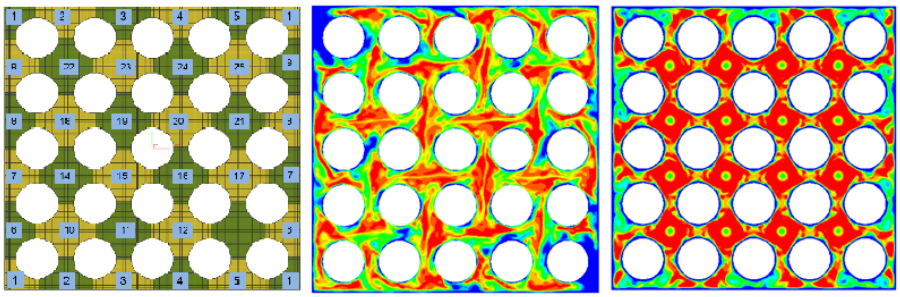


Figure 1. Core enumeration in the cross-section and snapshot of the axial velocity (z=4.0dh)

3. Numerical Result Mesh Comparison for the MAIS-H Test

Figure 1 (center and right) shows the axial velocity contours in the cross-sectional snapshot z=4.0dh for the split-type and swirl-type geometries. Numerical comparison of the predicted velocity field distribution with the experiment is shown in Fig. 2. Figure 2 also shows the results of the mesh sensitivity study.

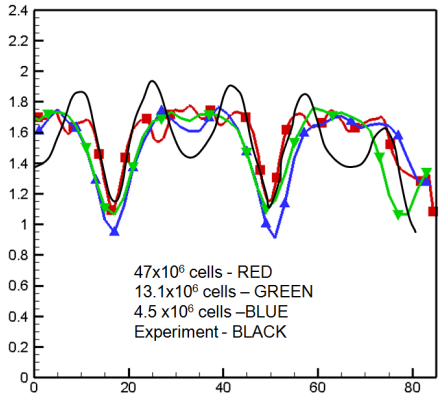


Figure 2. Axial velocity comparison for swirl, y/p=0.5 and z=0.5Dh with different meshes (grids=4.5, 13.1, and 47x106 cells)

The numerical solution generally shows an encouraging agreement with the experiment. For a more correct representation of mixing effects, resolving the scales could also require a fine computational grid further downstream of the mixer grid, as well as in its vicinity.

Conclusion

A hybrid-mesh modification of the MILES CABARET method has been applied for the MATiS-H problem. The computational results show only a small grid sensitivity for the grid resolutions studied ($4.5 - 47 \times 10^6$) in terms of the means velocity distributions. Lomonosov supercomputer facility has good scalability of up to 1800 computational cores.

Acknowledgments

The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011.

References

1. Chang S.K., Kim S., Song C.H.: OECD/NEA-KAERI Rod Bundle CFD Benchmark Exercise Test, Proc. Conf. CFD4NRS-4, Daejeon, Korea, September 10–12, 2012, Korea Atomic Energy Research Institute (2012)

2. Fureby, C., Grinstein F.F.: Large eddy simulation of high-reynolds-number free and wall-bounded flows. Journal of Computational Physics 181(1), 68–97 (2002), DOI: 10.1006/jcph.2002.7119

3. Karabasov, S.A., Goloviznin, V.M.: Compact Accurately Boundary Adjusting high-REsolution Technique for fluid dynamics. J. Comput. Phys. 228(19), 7426–7451 (2009), DOI: 10.1016/j.jcp.2009.06.037

4. Berloff, P., Karabasov, S., Farrar, J.T., Kamenkovich, I.: On latency of multiple zonal jets in the oceans. Journal of Fluid Mechanics 686, 534–567 (2011), DOI: 10.1017/jfm.2011.345

5. Faranosov, G.A., Goloviznin, V.M., Karabasov, S.A., Kondakov, V.G., Kopiev, V.F., Zaitsev, M.A.: CABARET method on unstructured hexahedral grids for jet noise computation. Computers and fluids 88, 165–179 (2013), DOI: 10.1016/j.compfluid.2013.08.011

6. Obabko, A.V., Fischer, P.F., Tautges, T.J., Goloviznin, V.M., Zaytsev, M.A., Chudanov, V.V., Pervichko, V.A., Aksenova, A.E., Karabasov, S.A.: Large eddy simulation of thermo-hydraulic mixing in a T-Junction. In: Guillen, D.P. (ed.) Nuclear Reactor Thermal Hydraulics and Other Applications. IntechOpen, Rijeka (2013). DOI: 10.5772/53143

7. Goloviznin, V.M., Zaitsev, M.A., Karabasov, S.A.: Towards Empiricism-Free Large Eddy Simulations for T-Junction Benchmark Problems, CFD FOR NUCLEAR REACTOR SAFETY APPLICATION (CFD4NRS-3) WORKSHOP, USA, Washington D.C., 14–16 September 2010, p. 104 (2010)

DOI: 10.14529/jsfi180325

High-performance Full-atomistic Simulation of Optical Thin Films

Fedor Grigoriev[1], *Vladimir Sulimov*[1], *Alexander Tikhonravov*[1]

The experimental study of the dependence of thin film properties on the deposition conditions may be still a great challenge. Today the progress in high performance computing allows one to perform the investigation of these dependencies on the atomistic level using the classical molecular dynamics (MD) simulation. In the present work the computational cost and efficiency of classical full-atomistic simulation of thin film deposition process using the Lonmonosov-2 supercomputer facilities is discussed. It is demonstrated that using 512 computational cores of the Lomonosov-2 supercomputer ensures the simulation of thin film cluster with technologically meaningful thickness of an optical film. Because of a relatively slow growth of the simulation time with the increase of film thickness we guess that simulations clusters with thicknesses that are several times higher than the currently achieved thicknesses about one hundred nanometers is quite realistic if the number of available computational cores will be increased up to several thousands.

Keywords: thin film structure, deposition process, molecular dynamic simulation, silicon dioxide.

Introduction

Optical thin films and coatings are important components of almost all modern optical and optoelectronic instrumentation. The detailed knowledge of the dependence of thin film structural properties on parameters of film deposition process is a key to a further progress in thin film technology. Experimental study of thin film microscopic structure may be still a great challenge. At the same time the progress in high performance computing allows one to perform the investigation of thin films properties on the atomistic level using the classical molecular dynamics (MD) simulation [1]. The MD-based approach for the simulation of deposition of thin films clusters with thicknesses up to hundreds of nanometers was recently developed and was successfully applied to the detailed investigation of silicon dioxide films [2–5].

In the present paper the current status and further prospects of the high-performance investigation of structural and mechanical properties of optical thin films are briefly discussed.

1. Simulation Method

The simulation of the silicon dioxide thin film deposition is organized as a step-by-step procedure [2–5]. At every deposition step, SiO_2 groups are injected at the top of the simulation box and are directed to the film surface. Duration of every deposition step is enough to ensure the formation of chemical bonds between the film and injected atoms. All atoms inside the simulation box move in agreement with the Newton laws. Their interactions are described in the frame of the DESIL force filed [2] that was specially developed for the simulation of the deposition process. The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University. Simulation of the deposition process is performed using parallel KUVALDA program [2], for MD part of the algorithm the GROMACS package [6] is used.

[1]Research Computing Center of M.V. Lomonosov Moscow State University, Moscow, Russian Federation

2. Results and Discussion

The time of MD simulation of one injection cycle depends linearly on the number of deposited atoms N_{at} and the total time of the deposition process simulation is proportional to the square of the N_{at} [2]. This is a reason for a relatively high computational cost of the simulation procedure as compared to the MD simulation of clusters consisting of fixed numbers of atoms.

Dependencies of the simulation time on the number of the deposited atoms are shown in Fig. 1.

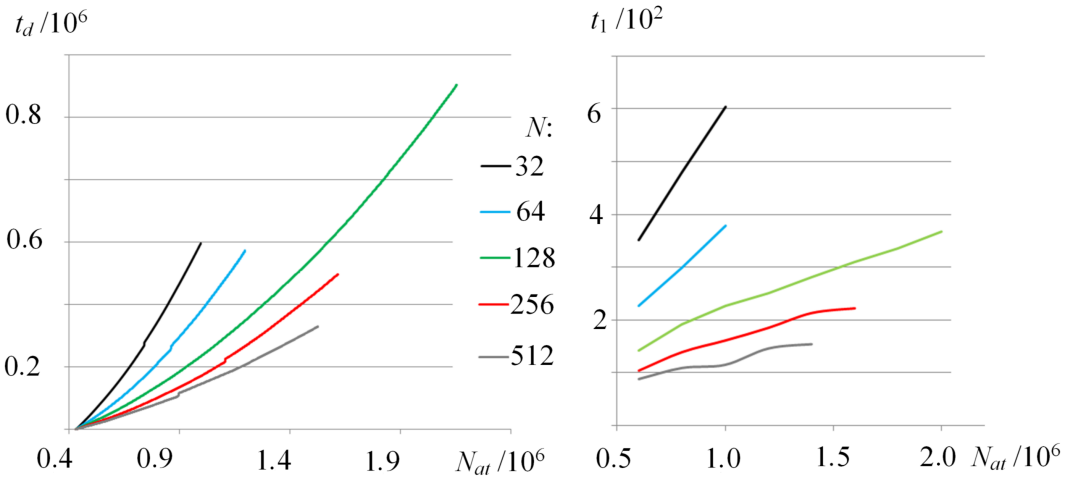


Figure 1. Dependencies of the full deposition time $t_d(s)$ and duration of one deposition cycle $t_1(s)$ on the number of deposited atoms N_{at} for different numbers N of computational cores

It is seen from Fig. 1 that using 512 computational cores of the Lomonosov-2 supercomputer ensures the simulation of thin film cluster with approximately 1.5×10^6 atoms which corresponds to the technologically meaningful thickness of an optical film. The simulation time in this case is about three days. But this makes possible a systematic investigation of thin film properties. It is interesting that the duration of one deposition cycle increases slowly with the growth of the number of deposited atoms (right part of Fig. 1). Because of a relatively slow growth of the simulation time with the increase of film thickness, we believe that simulations of film clusters with thicknesses that are several times higher than the currently achieved thicknesses is quite realistic if the number of available computational cores will be increased up to several thousands.

When the deposition process is over, film properties such as density profiles, point defects concentration, surface roughness, mechanical stress, Young modulus, etc. can be investigated [7]. Investigation of these properties requires MD simulations with the durations of hundreds of picoseconds. The results of the investigation of the computational cost of these simulations are shown in Fig. 2. Multi-core efficiency is calculated as:

$$a = (t_{32}/t_N)(N/32), \tag{1}$$

where t_{32} and t_N are the simulation times for the same cluster when 32 and N cores are used respectively. Although the efficiency is reduced with the growth of N (Fig. 2), the parallel computations with $N = 512$ and 1024 ensure a systematic study of film structural properties for the clusters with thicknesses of about 100 nm. It is clearly seen from the comparison of the results presented in Fig. 1 and Fig. 2 that the simulation of deposition process is the most time-consuming part of the investigation of thin film properties.

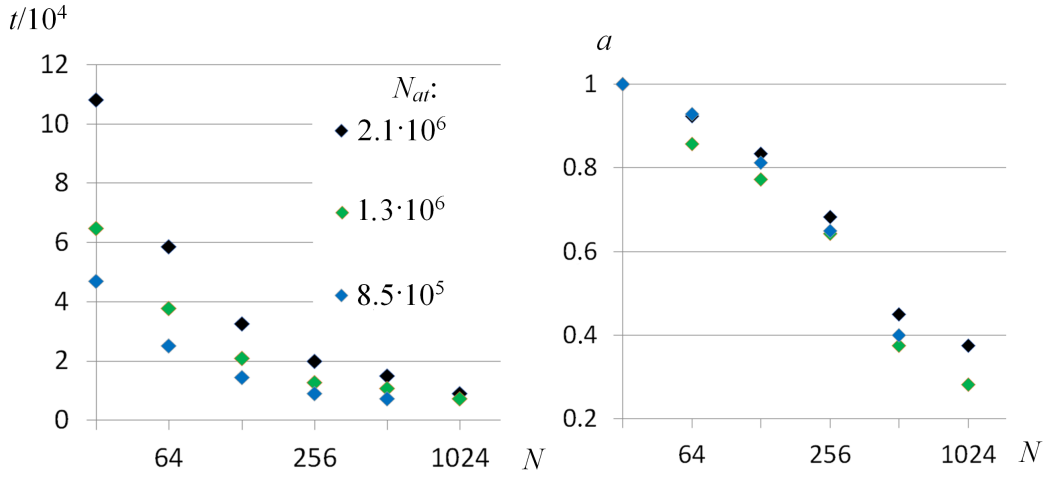


Figure 2. Dependencies of the simulation time $t(s)$ and multi-core efficiency a on the number of computational cores N; N_{at} is the number of atoms in the deposited cluster. Duration of the MD trajectory is 1 ns

Conclusions

Currently available computational resources of the Lomonosov-2 supercomputer are sufficient to the systematic full-atomistic investigation of structural and mechanical properties of optical films with technologically meaningful thicknesses of about a quarter of visible light wavelength. Simulation of the deposition process is the most time-consuming part of the investigation. Increasing the number of computational cores up to several thousand will make possible a full-atomistic simulation of coatings with several thin films.

Acknowledgments

The research was financially supported by Russian Scientific Foundation (grant number 14-11-00409). The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 [8].

References

1. Turowski, M., Jup, M., Ehlers, H., Melzig, T., Pflug, A., Ristau, D.: Simulation in Thin Film Technology. In: Lequime, M., Macleod, H.A., Ristau, D. (ed.) SPIE Optical Systems Design 2015, 7-10 September 2015, Jena, Germany. Optical Systems Design 2015: Advances in Optical Thin Films V, vol. 9627, pp. 9627-1–9627-10 (2015), DOI: 10.1117/12.2191693

2. Grigoriev, F.V., Sulimov, A.V., Kochikov, I.V., Kondakova, O.A., Sulimov, V.B., Tikhonravov, A.V.: High-performance atomistic modeling of optical thin films deposited by energetic processes. Int. J. of High Perform. Comp. Appl. 29, 184–192 (2015),

DOI: 10.1177/1094342014560591

3. Grigoriev, F.V., Sulimov, A.V., Katkova, E.V., Kochikov, I.V., Kondakova, O.A., Sulimov, V.B., Tikhonravov, A.V.: Computational experiments on atomistic modeling of thin film deposition. Applied Optics. 56, C87–C90 (2017), DOI: 10.1364/AO.56.000C87

4. Grigoriev, F.V., Sulimov, A.V., Katkova, E.V., Kochikov, I.V., Kondakova, O.A., Sulimov, V.B., Tikhonravov, A.V.: Full-atomistic nanoscale modeling of the ion beam sputtering deposition of SiO_2 thin films. J. of N.-Cryst. Sol. 448, 1–5 (2016), DOI: 10.1016/j.jnoncrysol.2016.06.032

5. Grigoriev, F.V., Sulimov, A.V., Katkova, E.V., Kochikov, I.V., Kondakova, O.A., Sulimov, V.B., Tikhonravov, A.V.: Annealing of deposited SiO_2 thin films: Full-atomistic simulation results. Opt. Mat. Express. 6, 3960–3966 (2016), DOI: 10.1364/OME.6.003960

6. Abraham, M.J., Murtola, T., Schulz, R., Pll, S., Smith, J.C., Hess, B., Lindahl, E.: GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 12, 19–25 (2015), DOI: 10.1016/j.softx.2015.06.001

7. Grigoriev, F., Sulimov, V., Tikhonravov, A.: Simulation of the optical coating deposition. Adv. Opt. Techn. 7(12), 13–22 (2018), DOI: 10.1515/aot-2017-0079

8. Sadovnichy, V., Tikhonravov, A., Voevodin, Vl., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale toward Exascale. pp. 283–307. Chapman & Hall/CRC Computational Science, CRC Press, Boca Raton, United States (2013)

DOI: 10.14529/jsfi180326

Supercomputer Docking: Investigation of Low Energy Minima of Protein-Ligand Complexes

***Danil C. Kutov*[1,2], *Alexey V. Sulimov*[1,2], *Vladimir B. Sulimov*[1,2]**

It is shown that the global energy minimum of a protein-ligand complex, when the energy is calculated by the PM7 quantum-chemical semiempirical method with the COSMO implicit solvent model, can be determined as follows. First, the low energy minima are found by a docking program when the protein-ligand energy is calculated with the MMFF94 force field in vacuum. Second, energies of all these minima are recalculated with the PM7 method and the COSMO implicit solvent model. Third, among these recalculated energies the minimal energy is determined and the respective minimum is the global energy minimum when the energy is calculated with the PM7 method and the COSMO implicit solvent model. The optimal width of the spectrum of low energy minima found with MMFF94 in vacuum is determined to perform minimal quantity of quantum-chemical recalculations. The proposed approach allows to perform docking in solvent with the quantum-chemical method and to increase the docking positioning accuracy.

Keywords: docking, protein-ligand, global minimum, force field, quantum-chemical method.

Introduction

Docking is the most demanded software for the development of new medicine. Docking performs positioning of ligands (small molecules) in protein and computes the protein-ligand binding energy. For the effective application of docking to the drug discovery, the docking accuracy must be improved. The satisfactory positioning accuracy demonstrated by many docking programs is insufficient to obtain the high enough accuracy of the binding energy calculation.

To overcome this problem, supercomputer docking programs have been recently developed [4, 9]. These programs do not use most simplifications worsening the accuracy of many existing docking programs, and they are based on the docking paradigm [4] assuming that the global minimum (GM) of the energy of the protein-ligand complex corresponds to the experimentally detected ligand pose in the protein. The docking accuracy depends on the energy calculation method, and best results [8] are obtained with the PM7 quantum-chemical semiempirical method [7] and the COSMO implicit solvent model [3]. Unfortunately the PM7 method demands much larger computational resources than classical force fields, as well as docking in vacuum is faster than docking in solvent. Nevertheless, docking with PM7 can be realized using the quasi-docking procedure [8] when a sufficiently broad spectrum of low energy minima is found using a force field, and then energies of all these minima are recalculated with PM7 and COSMO. By this procedure we are able to identify the ligand pose corresponding to the global energy minimum for PM7 with solvent, and the goal of docking is reached. The quasi-docking procedure saves a great deal of computational resources because it is necessary to fulfil many hundreds of thousands of local optimizations for the reliable detection of low energy minima [4], but for the energy recalculation several thousand low energy minima found with the force field are enough for the detection of the global minimum with the PM7+COSMO energy [8]. However, for practical application of the quasi-docking procedure we should answer the following question. How many low energy minima found with a given force field must be saved and then

[1]Research Computer Center of Lomonosov Moscow State University, Moscow, Russian Federation
[2]Dimonta Ltd., Moscow, Russian Federation

recalculated with PM7 to detect the global PM7 energy minimum? The answer is found, and respective results are presented below.

1. Materials and Methods

The FLM docking program [4] is used to search for low energy minima of protein-ligand complexes. FLM-0.05 utilizes the MMFF94 force field [2] in vacuum, and docking is performed by randomly tossing the ligand into the protein followed by the local optimization of the energy of the protein-ligand system with variations of Cartesian coordinates of all ligand atoms and keeping fixed all protein atoms. FLM finds and saves a pool of a given number of unique low energy minima. This pool consists of the global minimum and every successive minimum above it. FLM performs parallel processing until the set of low energy minima stops to get renewed. FLM-0.10 selects low energy minima on the base of their MMFF94 energy in water with the PCM implicit solvent model [10]. FLM-0.05 is much faster than FLM-0.10 because the latter consumes additional computing resources for PCM computations. Energy calculations by the PM7 method with the COSMO solvent model [3] are performed by the MOPAC program [6]. 16 protein-ligand complexes [8] with good structures taken from Protein Data Bank [1] are utilized.

2. Results

The quasi-docking procedure is performed on the base of 8192 low energy minima found by FLM-0.10 for every test protein-ligand complex [8]. This set of minima is designated as {2}MMFF94+PCM. Then, energies of all these minima are recalculated with PM7+COSMO. The ligand pose corresponding to the global PM7+COSMO energy minimum is near the crystallized native ligand pose for almost all test complexes [8], and therefore, the docking paradigm is fulfilled for them.

Is it possible to accelerate the quasi-docking procedure and to find the pool of low energy minima with MMFF94 in vacuum using the much faster FLM-0.05 program? In order to answer this question, pools of 8192 low energy minima are found (we designate them as {1}MMFF94) by the FLM-0.05 program. Next, for every complex we perform the search in {1}MMFF94 for the minimum corresponding to the ligand pose which is closest (by RMSD) to the pose corresponding to the global minimum of PM7+COSMO energy in {2}MMFF94+PCM. Values of such minimal RMSD are presented in Fig. 1. The PDB id is the identification of the 3D-structure from Protein Data Bank [1]. Values in bars are differences of PM7+COSMO energies in respective minima.

As we can see, for 10 complexes (out of 16) among {1}MMFF94 minima it is possible to identify a minimum which corresponds to the ligand pose in the close proximity of the ligand pose corresponding to the global minimum of PM7+COSMO energy from the {2}MMFF94+PCM set. It can also be seen that the close proximity can be defined by the condition RMSD $<$ 0.1Å, because the respective difference in PM7+COSMO energies is sufficiently small less than 0.5 kcal/mol. However, for other 6 complexes the respective global minimum of PM7+COSMO energy is not revealed among {1}MMFF94 minima: for these complexes RMSD $>$ 0.1Å, and the respective difference of PM7+COSMO energies is larger than 1 kcal/mol. One reason for this can be related to the insufficient quantity of low energy minima stored by FLM-0.05.

To support this explanation, the MMFF94 in vacuum energy bands are presented in Fig. 2: the energy band (green) occupied by 4096 unique low energy minima from the {1}MMFF94 set and the energy band (beige) between the energy of the ligand pose corresponding to the global

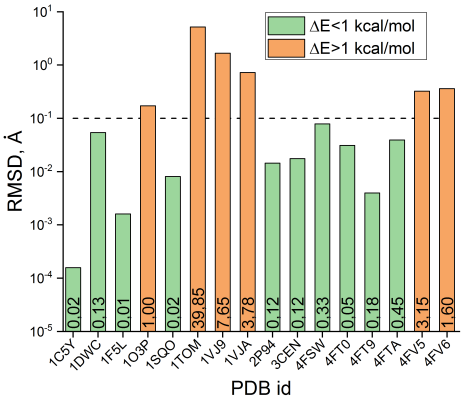


Figure 1. The minimal RMSD of the PM7+COSMO GM from MMFF94 minima

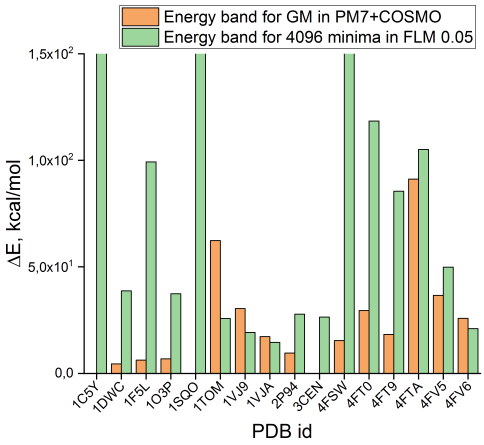


Figure 2. Energy bands calculated with MMFF94 in vacuum

PM7+COSMO energy minimum (from the {2}MMFF94+PCM set) and the energy of the global minimum from the {1}MMFF94 set.

As we can see for four complexes (1TOM, 1VJ9, 1VJA and 4FV6), the beige bars are higher than the green ones. This means that the energy bands occupied by minima found by FLM-0.05 with MMFF94 in vacuum do not cover the energy range needed to detect the global PM7+COSMO energy minimum from the {2}MMFF94+PCM set. For all these 4 complexes RMSD $>$ 0.1Å in Fig. 1 and the global minimum of PM7+COSMO energy is not detected by FLM-0.05. For the remaining two complexes (1O3P, 4FV5) with RMSD $>$ 0.1Å in Fig. 1 the possible reason can be related to incompleteness of the low energy minima set found by the FLM-0.05 program due to casual nature of the minima search algorithm. It is worth to note that for three complexes (1C5Y, 1SQO and 3CEN) the global minimum of the energy computed with MMFF94 in vacuum coincides with the global minimum of the PM7+COSMO energy, and the respective energy band is zero.

Conclusion

The presented results show that the quasi-docking procedure with the PM7 quantum-chemical method in solvent described by the COSMO model can be performed on the base of low energy minima found by docking when the energy is calculated with the MMFF94 force field in vacuum. For reliable finding of the global minimum of the protein-ligand energy calculated with PM7 and COSMO methods, sufficiently large number of unique low energy minima computed with the MMFF94 force field in vacuum must be identified and stored by docking. This number is specific for a given complex, it is determined by the energy band occupied by these minima, and this number is less than 4096 for most of test complexes. The energy band (MMFF94 in vacuum) occupied by these minima can reach several dozen kcal/mol for some complexes.

Acknowledgments

The work was financially supported by the Russian Science Foundation, Agreement No. 15-11-00025-П. The research is carried out using the equipment of the shared research

facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 [5].

References

1. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.: The protein data bank. Nucleic Acids Research 28(1), 235–242 (2000), DOI: 10.1093/nar/28.1.235

2. Halgren, T.A.: Merck molecular force field. Journal of Computational Chemistry 17(5-6), 490–519 (1996), DOI: 10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P

3. Klamt, A., Schuurmann, G.: COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J. Chem. Soc-Perkin Trans. 2 (5), 799–805 (1993), DOI: 10.1039/P29930000799

4. Oferkin, I.V., Katkova, E.V., Sulimov, A.V., Kutov, D.C., Sobolev, S.I., Voevodin, V.V., Sulimov, V.B.: Evaluation of Docking Target Functions by the Comprehensive Investigation of Protein-ligand Energy Minima. Adv Bioinformatics 2015, 126858 (2015)

5. Sadovnichy, V., Tikhonravov, A., Voevodin, V., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale toward Exascale. pp. 283–307. Chapman & Hall/CRC Computational Science, CRC Press, Boca Raton, United States (2013)

6. Stewart, J.J.: MOPAC2016. `http://OpenMOPAC.net` (2016), accessed: 2018-09-06

7. Stewart, J.: Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters. J. Mol. Model. 19(1), 1–32 (2013), DOI: 10.1007/s00894-012-1667-x

8. Sulimov, A.V., Kutov, D.C., Katkova, E.V., Ilin, I.S., Sulimov, V.B.: New generation of docking programs: Supercomputer validation of force fields and quantum-chemical methods for docking. J. Mol. Graph. Model. 78, 139–147 (2017), DOI: 10.1016/j.jmgm.2017.10.007

9. Sulimov, A.V., Zheltkov, D.A., Oferkin, I.V., Kutov, D.C., Katkova, E.V., Tyrtyshnikov, E.E., Sulimov, V.B.: Evaluation of the novel algorithm of flexible ligand docking with moveable target-protein atoms. Comput. Struct. Biotechnol. J. 15, 275–285 (2017), DOI: 10.1016/j.csbj.2017.02.004

10. Sulimov, V.B., Mikhalev, A.Y., Oferkin, I.V., Oseledets, I.V., Sulimov, A.V., Kutov, D.C., Katkova, E.V., Tyrtyshnikov, E.E.: Polarized continuum solvent model: Considerable acceleration with the multicharge matrix approximation. Int. J. Appl. Eng. Res. 10(24), 44815–44830 (2015)