Journal of Computational Science 9 (2015) 70–75

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Massively parallel simulations of hemodynamics in the primary large arteries of the human vasculature

Amanda Randles[a,*], Erik W. Draeger[a], Peter E. Bailey[b]

[a] Lawrence Livermore National Laboratory, Livermore, CA, USA
[b] University of Arizona, Tucson, AZ, USA

ARTICLE INFO

Article history:
Available online 17 April 2015

Keywords:
Lattice Boltzmann
Computational fluid dynamics
High performance computing
Patient-specific hemodynamics
Strong scaling

ABSTRACT

We present a computational model of three-dimensional and unsteady hemodynamics within the primary large arteries in the human on 1,572,864 cores of the IBM Blue Gene/Q. Models of large regions of the circulatory system are needed to study the impact of local factors on global hemodynamics and to inform next generation drug delivery mechanisms. The HARVEY code successfully addresses key challenges that can hinder effective solution of image-based hemodynamics on contemporary supercomputers, such as limited memory capacity and bandwidth, flexible load balancing, and scalability. This work is the first demonstration of large fluid dynamics simulations of the aortofemoral region of the circulatory system at resolutions as small as 10 μm.

Published by Elsevier B.V.

1. Introduction

A longstanding goal within the field of computational biomechanics has been to understand the principles that govern vascular disease localization and progression [20,5,28]. Such image-based simulations can yield insight into the impact of local factors on global hemodynamics, direct the design of next-generation drug delivery mechanisms, and inform surgical planning. Although important progress towards this goal has been made using various algorithmic methods [11,29,12,21,13,32], the computational demands of these simulations have historically restricted the resolution and size of the circulatory system that can be modeled.

In recent years, there has been a great deal of work in the area of computational hemodynamics. These studies are typically limited to small regions of the body or use a one-dimensional setting to describe the human arterial network [26,1,30]. Xiao et al. presented the first model of full unsteady and three-dimensional hemodynamics in the primary large arteries from head to foot. While this was a significant advance in computational fluid dynamics, the goal was to demonstrate the feasibility of the 3D framework. However, the resolution presented was insufficient to reach grid independence [32]. In that work, the finite element mesh consisted of 14,438,720 linear tetrahedra and 2,674,545 nodes. High resolution studies based on 3-D reconstructions of patient-specific data are typically focused on either the cerebral vasculature [9,13,7] or the cardiovascular region [3,10,19]. The current state of the art in the numerical investigation of hemodynamics in patient-specific geometries are those by Bernaschi et al. which studies the coronary arteries in a bounding box of 300 billion grid points containing one billion fluid nodes [3,4]. Work presented here goes beyond these scales, simulating a vertical section of the aortofemoral section of the circulatory system spanning 614 cm at 10 μm resolution, consisting of more than 128 billion fluid nodes. To the best of our knowledge, our work is the first large-scale study of blood flow in a region of this size and level of detail.

Building realistic models of transport phenomena in the human circulatory system presents a formidable computational challenge due to the geometric complexity of the system, memory requirements associated with high-resolution grids, and load balancing issues associated with the processor core counts required. Our proposed solution extends the design and parallel efficiency of HARVEY [22], a computational fluid dynamics code based on the lattice Boltzmann method (LBM). One fundamental hurdle for high-resolution fluid simulations is the size of the underlying data grid and associated memory requirements. In order to study key macroscopic risk factors in patient-specific data, a resolution of at least 20 μm is required [18]. For full body simulations, this resolution corresponds to 18.4 billion fluid nodes in a bounding box of 8.8 trillion total grid points. These data sizes create an additional challenge to load balance, as work must be assigned to over one million tasks without computing or storing global data. We present a

* Corresponding author.
E-mail addresses: randles2@llnl.gov (A. Randles), draeger1@llnl.gov (E.W. Draeger), pbailey@cs.arizona.edu (P.E. Bailey).

http://dx.doi.org/10.1016/j.jocs.2015.04.003
1877-7503/Published by Elsevier B.V.

multi-step iterative load balance algorithm that allows for the distribution of large, complex arterial geometries on a 3D process grid. Efficiently modeling the hemodynamics in the large primary arteries also required data-reordering techniques to increase spatial locality, both optimized data structures and access patterns to reduce the overall memory footprint and efficient communication layout to overcome bandwidth limitations.

In this work we make the following contributions: increasing the number of fluid nodes that can be simulated by an order of magnitude (thereby increasing potential system size and/or resolution), incorporating preprocessing to reduce storage and I/O burdens, and enabling an unprecedented scale of hemodynamic simulation demonstrated by the 10 μm resolution simulation of the aortofemoral region of the circulatory system.

2. Methodology

This work relies on the lattice Boltzmann method (LBM), an alternative to the conventional Navier–Stokes equation, introduced by both teams of McNamera and Zanetti [17] and Higuera and Jimenez [15]. LBM comes from kinetic theory and is a minimal form of the Boltzmann equation based on the collective dynamics of fictitious particles that represent a local ensemble of molecules moving between the points of a regular Cartesian lattice. The time advancement is explicit and the computational stencil is formed by local neighbors of each computational node, making it particularly well-suited for massively parallel simulations (c.f. [6,31,23]).

The governing equation describes the evolution of the distribution function denoted by $f_i(\vec{x}, t)$, describing the probability of finding a particle at grid point $\vec{x}$, at time t, with discrete velocity $\vec{c}_i$. In this work, we use the 19-speed cubic stencil, with the Bhatnager-Gross-Krook (BGK) collision formulation with a single relaxation time. The grid spacing is defined by Δx, where discrete velocities connect grid points to first and second neighbors on the 19-point stencil. The fluid populations are advanced in a timestep Δt through:

$$f_i(\vec{x} + \vec{c}_i \Delta t, t + \Delta t) = f_i(\vec{x}, t) - \omega \Delta t [f_i(\vec{x}, t) - f_i^{eq}(\vec{x}, t)] \quad (1)$$

The local equilibrium, $f_i^{eq}(\vec{x}, t)$, is the result of a second-order expansion in the fluid velocity of a local Maxwellian with speed $\vec{u}$ and is defined by:

$$f_i^{eq} = w_i \rho \left[1 + \frac{\vec{c}_i \cdot \vec{u}}{c_s^2} + \frac{1}{2} \left(\frac{(\vec{c}_i \cdot \vec{u})^2}{(c_s^2)^2} - \frac{u^2}{c_s^2} \right) \right] \quad (2)$$

where ρ denotes the density, $\vec{u}$ the average fluid speed, $c_s = 1/\sqrt{3}$ the speed of sound in the lattice, and w_i the weights attributed to each discretized velocity as determined by the lattice structure. Due to the use of explicit time-stepping, LBM requires small time-steps that scale with Δx^2. In the case of the 10 μm simulations discussed in this work, approximately 3 million time-steps would be required to simulate one cardiac cycle.

We implement the Zou-He boundary conditions [33], in which a pulsating velocity is imposed at the inlet through a *plug profile* at the entrance to the vessel and a constant pressure is imposed at the outlets. While the inlet condition does not assert the known parabolic profile that drops to zero close to the wall, it allows a total flow to be imposed at a set value. In a short distance past the inlet, the parabolic profile is recovered. This method uses information streamed from the bulk fluid nodes alongside a completion scheme for the unknown particle populations whose neighbors are outside the fluid domain. This method can be executed with second-order accuracy [16]. In this paper, the modification introduced by Hecht and Harting [14] is used in which the velocity conditions are specified on-site, thus removing the constraint that all nodes of a given inlet or outlet must be aligned on a plane that is perpendicular to one of the three main axes. Furthermore, this addition allows the boundary conditions to be applied locally. A no-slip boundary condition is imposed at the walls via the full bounce-back method. For more details regarding the lattice Boltzmann method, see Ref. [27].

3. HARVEY implementation details

All simulations presented here were carried out using the HARVEY code. Despite the excellent scalability reported previously[24], significant restructuring of the code had to be done to enable the resolution and scale of the systems studied in this work. Details of the original implementation can be found in Ref. [22]. All simulations were run on the Sequoia machine at Lawrence Livermore National Laboratory, a 98,304 node IBM Blue Gene/Q machine (1,572,864 cores).

In order to simulate the hemodynamics in the aortofemoral geometry at a high-resolution, we had to overcome the following challenges:

- Memory footprint. Large numbers of grid points are required to reach convergence of macroscopic quantities of interest. These requirements impose a high demand in terms of on-node memory requirements.
- I/O bandwidth. Setting up the large, high-resolution grids through the existing preprocessing and initialization stages involves I/O operations on petabytes of data, causing the simulation setup to actually overwhelm the overall simulation time even for the large number of time steps, $\sim 10^6$, needed to model a full cardiac cycle.
- Scalability. As each Blue Gene/Q core has only 1 GB of available total memory, the scope of this problem requires use of the entire LLNL Sequoia system of 1,572,864 cores. Effective utilization of such a large core count means that traditional parallelization tools like global communication tables are not feasible options.
- Load imbalance. The geometry of the human vasculature is incredibly complex. The bounding box holds grid points representing fluid, inlets, outlets, walls, and exterior points. Distributing the workload across hundreds of thousands to millions of cores requires careful attention to load balance.

We address these challenges by extending the capability of HARVEY through (i) embedding of preprocessing and use of buffered meshing to avoid global bottlenecks and reduce I/O stress, (ii) the introduction of indirect addressing to reduce the memory footprint, (iii) development of a multi-step load balancing scheme that prioritizes locality and memory reduction, (iv) removal of global communication tables to improve scalability.

3.1. Parallel preprocessing

The original implementation of HARVEY used multiple preprocessing steps to construct the 3D spatial grid from the surface mesh and set up neighbor lists and communication tables. This strategy becomes infeasible at the target scales of this work, as the full 3D mesh must be read from disk and distributed across tasks, creating both an I/O and memory bottleneck. Instead, we have integrated these routines into HARVEY so that only the surface mesh is used in the initial load balance and the volume grid (Fig. 1) can be generated in place on local MPI tasks.

Communication tables can be generated during setup using only local information due to the use of a structured process grid, i.e. tasks need only talk to their process grid neighbors to discover who owns fluid nodes in their stencil.