

Enabling Large Scale Individual-Based Modelling through High Performance Computing

Maciej Cytowski¹ and Zuzanna Szymańska^{1,a}

¹ICM, University of Warsaw, ul.Pawińskiego 5A, 02-106 Warszawa, Poland

Abstract. We present a novel large scale parallel computational model allowing 3-D simulations of cell colonies growing and interacting with variable environment. The cells are modelled as individual objects located in the lattice-free 3-D space. The model incorporates cellular environment described in a continuous manner. Discrete and continuous formulations are efficiently coupled in one model allowing considerations of cell colonies composed of up to 10^9 individual cells. This large scale computational approach enables simulations to be carried out over realistic spatial scales up to 1cm^3 in size i.e. the tissue scale.

1 Introduction

Most of individual-based models used in computational biology focus on conceptual modelling neglecting the efficient implementation usually at the expense of a realistic scale of the problem. We did the opposite, i.e. we put emphasis on the efficient implementation of a generic model of cell colonies dynamics rather than modelling particular biological process. The ultimate goal was to enable simulations in the field of mathematical biology at previously unavailable scales through the use of novel HPC algorithms, solutions and technology. In order to achieve high sustained performance on current peta- and future exascale systems we have identified key constraints that our implementation addresses. This led us to appropriate design decisions for the algorithm and its implementation used in the simulation code.

Timothy (<http://timothy.icm.edu.pl>) is a novel open-source and large scale parallel computational framework allowing 3-D simulations of cell colonies growing and interacting with a variable environment. The cells are modelled as individuals located in lattice-free 3-D space. The model incorporates the cellular environment represented in a continuous manner. The mathematical description, based on partial differential equations, is formulated for selected important components of the environment. Discrete and continuous formulations are efficiently coupled in one model and allow considerations of different scales: sub-cellular, cellular and tissue scale. The high parallel scalability of the implementation allows for simulation of up to 10^9 individual cells. This large scale computational approach allows for simulations to be carried out over realistic spatial scales up to 1cm^3 in size i.e. the tissue scale. Recently the Timothy model has been extended to include blood vessels supplying nutrients to the tissue, which greatly increased the computational complexity however makes a very important step towards realistic simulations of complex processes occurring in tissue.

^ae-mail: z.szymanska@icm.edu.pl

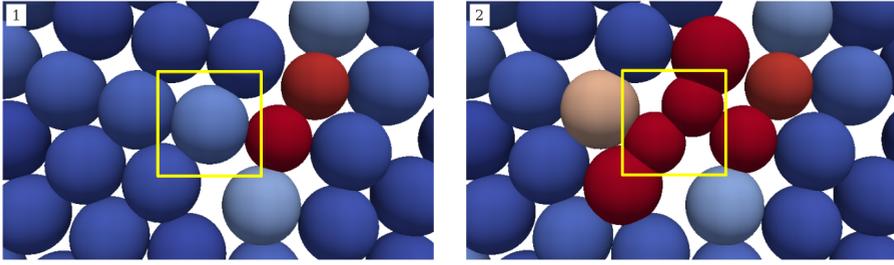


Figure 1. Process of mitosis of a given cell (indicated by a yellow frame) dividing into two daughter cells. Colours indicate cell density. Snapshot from a 2-D simulation.

The mathematical model and implementation details of the Timothy framework have been previously described in [1–3]. Here we describe only its most important characteristics.

2 Modelling approach

We consider cells as free spheres that reside in three-dimensional space. Both the position and size of a cell can change in time. Cells interact with each other. For every cell its maximum inter-cellular interaction distance is defined by a neighbourhood sphere.

The growth of a population of cells is based on the growth and division of individual cells. In order to divide, a cell must undergo an ordered series of reactions called the cell-cycle. The eukaryotic cell cycle consists of four phases: M - phase, when the division of the cell occurs, G_1 - phase, which is the growth phase between cell division and the initiation of nuclear DNA replication, S - phase, when nuclear DNA is duplicated, and G_2 - phase, which is the interval between the completion of nuclear DNA replication and mitosis. After the G_1 phase, a cell takes a decisive step to initiate DNA replication and complete the division cycle. If conditions are particularly unfavourable, instead of entering S phase, a cell can enter a resting state - G_0 phase, where it remains until conditions improve and it restarts the cell cycle or it dies. The G_2 phase provides a safety gap, allowing the cell to ensure that DNA replication is complete before mitosis. The splitting of the mother cell into two new daughter cells is presented on Figure 1.

The interactions between cells in the system are described by a modified Hertz model [4]. A decreasing distance between the centres of cells results in an attractive interaction related to adhesive forces. Experiments suggest that cells have only limited compressibility which gives rise to repulsive interaction. The potential between two adjacent cells in the model is a combination of repulsive and attractive forces. The potential function for each cell is given by the sum of potential values related to interactions with all neighbours. Cells can change their position after each simulation step. The displacement velocity vector is computed by deriving the potential function.

Living cells continuously interact with their environment. Therefore, the discrete model of cellular dynamics is extended by a continuous description of the cellular environment, e.g. nutrients. In the model, at each time step, the cell interacts with its environment. If the conditions are appropriate the cell continues its life cycle until the next time step. If the conditions become unfavourable and cell is not able to pass through the cell cycle, the cell ceases proliferation and becomes quiescent. A further deterioration of conditions leads to cell death. A general equation governing the external nutrient in the environment is given by:

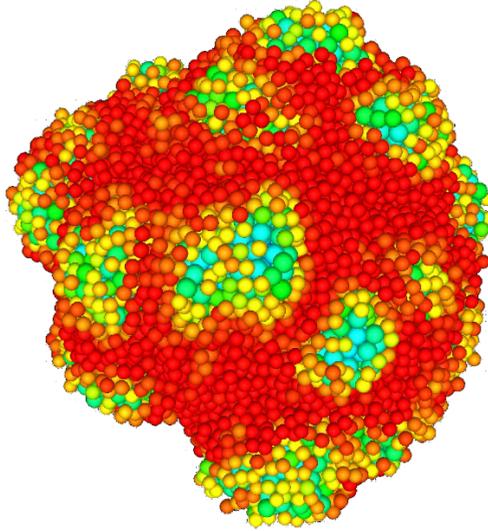


Figure 2. 3-D solid tumour developed within healthy tissue in Timothy. Only tumor cells are visible, which enables us to analyse the topology of the 3-D structure.

$$\frac{\partial u(x; t)}{\partial t} = D_u \nabla^2 u(x; t) - \beta u(x; t) + f(x; t) - g(x; t). \quad (1)$$

Here $u(x; t)$ denotes concentration of nutrient whereas $f(x; t)$ and $g(x; t)$ are source and sink/uptake functions of external nutrient by individual cells. Both functions f and g are calculated based on cell density approximated at each node of the discretization grid. In this way we obtain a coupling of the discrete model describing the cellular dynamics and the continuous model describing the environment. An example of a simulation conducted with the use of Timothy framework is presented on Figure 2.

3 Algorithms and implementation

The Timothy framework is currently composed of two modules related to cellular dynamics (discrete) and cellular environment (continuous).

Cellular dynamics module

Each iteration of the simulation begins with a domain decomposition which distributes cells across available parallel processes. We have selected the most appropriate decomposition method based on the analysis of resulting load balancing. We decided to use a dynamic decomposition algorithm based on the concept of Hilbert space filling curves (HSFC), which is particularly useful in the case of systems of free moving objects.

In consecutive iterations each cell identify its neighbours. This process can be very time consuming, especially when a naive algorithm is used. We use an efficient method based on tree algorithms, with complexity of the order of $O(N \log N)$, where N is the total number of cells. The tree is built on each parallel process by an iterative procedure. It starts from a cubical root node containing all cells assigned to the process. This root node is repeatedly subdivided into eight daughter nodes of

half the side length each, until one ends up with leaf nodes containing single cells. We consider only short-distance interactions since each cell has a finite neighbourhood sphere. Neighbours of each cell are found by traversing the tree starting from the leaf upwards. In this way we compute the most important characteristics of the simulation: the potential and density functions.

Cellular environment module

For the environment computations we employ the Hypre parallel library [5]. The discretization of the PDEs is achieved by use of an implicit in time finite difference scheme. The system is then defined in the ParCSR parallel format. The decomposition of data in Hypre is achieved by assigning computational grid blocks to different processes. We are using a conjugate gradient solver with a preconditioner to solve the system.

Coupling of the discrete and continuous approaches

Solving environment fields together with the discrete cellular system is a challenging task from the point of view of parallel processing. The problem lies in the block data decomposition required by the Hypre library, which is different to the HSFC method used in tree based computations. In each iteration of the simulation the source and sink/uptake functions for the environment fields need to be calculated from cellular data. On the other hand individual cells need information on their environment in order to incorporate this information into their cycle. To address this problem we have developed a parallel interpolation algorithm which efficiently computes those interactions.

Technical details

Timothy is written in the C programming language. Parallelization is based on the hybrid MPI/OpenMP model. Communication between parallel processes is based mostly on non-blocking MPI communication mechanisms, therefore we are able to overlap computations and communication. High scalability of the model has been demonstrated on a single rack IBM Blue Gene/Q (more than 16 000 computer cores) available at ICM UW. Timothy uses an MPI based parallel I/O scheme to handle output files generation (various output data formats). We have also implemented a check-point/restart mechanism which can be used to save the current state of a cell colony development.

4 Summary and future work

Timothy software provides an exceptional benefit for the scientific community. It is the first and to our knowledge the only individual-based cellular biosystems modelling tool capable of simulating processes at the tissue scale. It provides a research tool complementary to the traditional, heuristic experimental approaches. The model has been already successfully used to simulate complex biological processes. Timothy has enormous potential for future developments including quantitative, predictive modelling of cancer growth, wound healing and embryogenesis. Simulations planned to be executed based on the Timothy model will take multi-scale mathematical modelling of cellular systems to a completely new level. In our future work we are also planning to further improve MPI/OpenMP scalability as well as port the application to the Intel Xeon Phi architecture and the OpenPOWER architecture including NVidia GPUs.

Acknowledgments

This work was supported by the OCEAN project, co-financed by the European Regional Development Fund. The authors were partially supported by the National Science Centre Poland Grant 2011/01/D/ST1/04133, National Science Centre Poland Grant 2014/15/B/ST6/05082 and The National Centre for Research and Development Grant STRATEGMED1/233224/10/NCBR/2014.

References

- [1] M. Cytowski, Z. Szymańska, B. Borucki, 2-D Large-scale individual based model of solid tumour growth, GAKUTO International Series Mathematical Sciences and Applications, Nonlinear Analysis in Interdisciplinary Sciences **36**, 43-56 (2013)
- [2] M. Cytowski, Z. Szymańska, Large Scale Parallel Simulations of 3-D Cell Colony Dynamics, IEEE Computing in Science & Engineering **16**(5), 86-95 (2014)
- [3] M. Cytowski, Z. Szymańska, Large Scale Parallel Simulations of 3-D Cell Colony Dynamics. II. Coupling with continuous description of cellular environment, IEEE Computing in Science & Engineering **17**(5), 44-48 (2015)
- [4] J. Galle, M. Loeffler, D. Drasdo, Modelling the Effect of Deregulated Proliferation and Apoptosis on the Growth Dynamics of Epithelial Cell Populations In Vitro, Biophysical Journal **88**, 62-75 (2005)
- [5] A. H. Baker, R. D. Falgout, T. V. Kolev, U. M. Yang, Scaling hypre's multigrid solvers to 100,000 cores, in High-Performance Scientific Computing, Eds. M. W. Berry, K. A. Gallivan, E. Gallopoulos, A. Grama, B. Philippe, Y. Saad, and F. Saied, Springer London, 261-279 (2012).