



Title	Techniques for Parallel Simulation of Large Scale Heterogeneous Biophysical Systems
Author(s)	Heien, Eric Martin
Citation	大阪大学, 2010, 博士論文
Version Type	VoR
URL	<a href="https://hdl.handle.net/11094/1598">https://hdl.handle.net/11094/1598</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

【9】

氏 名	エリック マーティン ハイエン Eric Martin Heien
博士の専攻分野の名称	博 士（情報科学）
学 位 記 番 号	第 2 3 9 1 3 号
学 位 授 与 年 月 日	平成 22 年 3 月 23 日
学 位 授 与 の 要 件	学位規則第 4 条第 1 項該当 情報科学研究科コンピュータサイエンス専攻
学 位 論 文 名	Techniques for Parallel Simulation of Large Scale Heterogeneous Biophysical Systems (大規模な異種混在生体モデルの並列シミュレーションに関する研究)
論 文 審 査 委 員	(主査) 教 授 萩原 兼一 (副査) 教 授 増澤 利光 教 授 松田 秀雄 教 授 野村 泰伸 (大学院基礎工学研究科)

論 文 内 容 の 要 旨

Recently several multidisciplinary projects have begun to model and simulate human biophysical and physiological systems. These projects aim to create databases of models which can be combined to generate larger and more complex biophysical models. These large scale models can then be used to perform in silico (computer based) simulations to accurately predict the effects of drugs over a range of circumstances and patients. The ultimate goal of these projects is to realize a system for predictive medicine. This uses patient specific models to determine treatment for an ailment in a particular patient or predict side effects of a drug on patients with certain genetic traits.

Though this research area has advanced in recent years, there is still a lack of tools to effectively perform these types of simulations, particularly in regards to simulating large scale models on parallel computing platforms. These types of large scale simulations are required to effectively solve the problems posed by predictive medicine. Although many software packages exist to perform biophysical simulations, few of these support parallel simulation of large scale models, and the ones that do generally support only a particular type of model. This dissertation discusses techniques and tools that are applicable to performing parallel simulations of these kinds of large scale models. This subject is addressed on multiple scales. First, we examine techniques for performing parallel simulations by converting models into C++ source code which is compiled and executed. This method is common because of its simplicity and ease of simulation generation. Also, there are tools such as MATLAB and Mathematica which may be used as a back end to perform the computation. However, these types of simulations are difficult to parallelize because of the complexity of the models and difficulty of writing metacode for parallel simulations. In this case, we use model analysis and redundant computation to simplify the parallel simulation yet maintain good performance and calculate the same results. However, this technique is limited because of the time required to compile individual simulations and the complexity as the model becomes more heterogeneous.

Next we describe insilicoSim, an extendable simulation engine for performing parallel large scale biophysical simulations. Rather than creating source code for each model, this engine imports models, converts them to internal data structures and performs simulations. This solves the difficulties of source based simulations in dealing with large heterogeneous models and allows new types of models and simulations. We present three key components of the simulator for improving extensibility and performance. First, we demonstrate how a

standardized plugin interface allows for easy extension of the simulator to new types of input, output and simulation methods. We detail a technique for improving simulation performance by simplifying and compiling simulation related calculations into a byte code representation for fast evaluation. Finally, we describe the simulation object manager which allows for shared object access between simulation interfaces while transparently performing parallel synchronization. We demonstrate the effectiveness of these methods by simulating several models on both serial and parallel computing platforms.

Finally, we demonstrate a method for utilizing large scale unreliable computing resources to perform parallel computations such as those used in modeling biomolecular dynamics. We propose algorithms for computing batches of medium grained tasks with deadlines in pull-style volunteer computing environments. First we develop models of unreliable workers based on analysis of trace data from an actual volunteer computing project. These models are used to develop algorithms for task distribution in volunteer computing systems with a high probability of meeting batch deadlines. We develop algorithms for perfectly reliable workers, computation-reliable workers and unreliable workers. Finally, we demonstrate the effectiveness of the algorithms through simulations using traces from actual volunteer computing environments.

## 論文審査の結果の要旨

最近、生体学を中心としたモデル作成の研究が世界中で注目されている。これらの研究目的は、生体モデルのデータベースを構築し、大規模かつヘテロな生体モデルを作成し、シミュレーションを行うことにより、患者にとって適切な治療と薬物を予測することである。ただし、このような生体シミュレーションを行うためのツール開発が十分でない。特に大規模なモデルはシミュレーション時間が長くなり、シミュレーション時間の高速化が求められている。シミュレーション高速化のためには、シミュレーションの並列処理が不可欠な技術である。生体シミュレーションを並列処理するシステムはいくつか存在するが、特定臓器を対象とするものが多く、人体全体を対象とする一般的な生体モデルの並列シミュレータは殆どない。本論文では、一般的生体モデルの並列シミュレータを開発し、その成果を次の3つの観点から述べたものである。

(1) 並列シミュレータを実装するため、生体モデルをC++ソースコードに変換してコンパイルし、それを並列環境で実行するシミュレータを開発した。単純にモデルのシミュレーションを並列実行すると、通信のために時間がかかり、全体の処理が遅くなる可能性が高い。この問題点に対して、生体モデルを解析して、通信を回避するためにあえて冗長計算（同一計算の複製）を行い、並列環境で実行するようにした。この方法の短所は、計算対象の生体モデルが大きくなれば、コンパイル時間が長くなることである。

(2) 異種の生体モデルが新たに追加されても、その拡張が容易に可能な並列大規模シミュレータinsilicoSimを開発した。insilicoSimは、生体モデルをソースコードに変換するのではなく、シミュレーション用のデータ構造に変換し、シミュレーションを行う。上記(1)のシミュレータの問題点を解決でき、新しいモデルや複数のモデル計算方法を容易にシミュレータに組み込むことが可能になる。insilicoSim実装の重要な点は、次の3つである。標準化された入出力インタフェースを用いてシミュレーションを拡張できること、シミュレーションを高速にするために、データ構造をバイトコードに変更すること、そして並列同期を隠蔽するために、共有のデータマネージャーを用いることである。insilicoSimを複数の生成モデルに対して、複数の並列環境で実行し、効率が良いことを実証した。

(3) 生体分子シミュレーション実行のために、大規模計算資源を利用するボランティア並列計算環境向けのスケジューリング方法を提案した。まず、実際の大規模なトレースデータを解析し、信頼性の低い計算機のモデルを提案した。次に、この計算機モデルにおいて高い確率でタスク計算締切期限を満たすためのタスク分配アルゴリズムを開発した。実際の並列環境から得たトレースデータでシミュレーションを行い、このアルゴリズムの有効性を示した。

これらの研究成果は、生体シミュレーション研究をおおいに発展させるシステムを構築したものであり、博士(情報科学)の学位論文として価値のあるものと認める。